Leibniz Universität Hannover

Institut für Hydrologie und Wasserwirtschaft

Mitteilungen Heft 108

J. Dietrich

ISSN 0343-8090

Wasserwirtschaft unter globalem Wandel – Herausforderungen und Lösungswege für die Simulation von Flussgebieten

Wasserwirtschaft unter globalem Wandel – Herausforderungen und Lösungswege für die Simulation von Flussgebieten

Von der Fakultät für Bauingenieurwesen und Geodäsie der

Gottfried Wilhelm Leibniz Universität Hannover

genehmigte Habilitationsschrift von

Dr.- Ing. Jörg Dietrich

geboren am 08.07.1970 in Frankfurt am Main

2019

Tag der Habilitation: 21.01.2019

Referentin und Referenten:

Prof. Dr.-Ing. Uwe Haberlandt, Institut für Hydrologie und Wasserwirtschaft, Leibniz Universität Hannover

Prof. Dr. Nicola Fohrer, Insitut für Natur- und Ressourcenschutz, Abteilung für Hydrologie und Wasserwirtschaft, Christian-Albrechts-Universität zu Kiel

Prof. Dr.-Ing. Markus Disse, Lehrstuhl für Hydrologie und Flussgebietsmanagement, Technische Universität München

Wissenschaftliche Originalpublikationen der kumulativen Habilitationsschrift

I. Dietrich, J.; Trepte, S.; Wang, Y.; Schumann, A. H.; Voß, F.; Hesser, F. B.; Denhard, M. (2008): Combination of different types of ensembles for the adaptive simulation of probabilistic flood forecasts. Hindcasts for the Mulde 2002 extreme event. Nonlin. Processes Geophys. 15 (2), 275–286.

II. Dietrich, J.; Schumann, A. H.; Redetzky, M.; Walther, J.; Denhard, M.; Wang, Y. et al. (2009): Assessing uncertainties in flood forecasts for decision making. Prototype of an operational flood management system integrating ensemble predictions. Nat. Hazards Earth Syst. Sci. 9 (4), 1529–1540.

III. Dietrich, J.; Denhard, M.; Schumann, A. H. (2009): Can ensemble forecasts improve the reliability of flood alerts? Journal of Flood Risk Management 2 (4), 232–242.

IV. Wallner, M.; Haberlandt, U.; Dietrich, J. (2012): Evaluation of different calibration strategies for large scale continuous hydrological modelling. Advances in Geosciences 31, 67–74.

V. Petry, U.; Dietrich, J.; Förster, K.; Wallner, M.; Berndt, C.; Meon, G.; Haberlandt, U. (2015): Ein Ansatz zur Validierung von Klimamodelldaten als Basis für die Interpretation von wasserwirtschaftlichen Klimafolgenabschätzungen in Niedersachsen. Hydrologie und Wasserbewirtschaftung 59 (4), 155–173.

VI. Maier, N.; Dietrich, J. (2016): Using SWAT for Strategic Planning of Basin Scale Irrigation Control Policies: a Case Study from a Humid Region in Northern Germany. Water Resources Management 30(9), 3285-3298.

VII. Uniyal, B.; Dietrich, J.; Vasilakos, C.; Tzoraki, O. (2017): Evaluation of SWAT simulated soil moisture at catchment scale by field measurements and Landsat derived indices. Agricultural Water Management 193, 55-70.

VIII. Ojwang, R.; Dietrich, J.; Kasargodu Anebagilu, P.; Beyer, M.; Rottensteiner, F. (2017): Rooftop Rainwater Harvesting for Mombasa: Scenario Development with Image Classification and Water Resources Simulation. Water 9(5), 359.

IX. Nguyen, V. T.; Dietrich, J. (2018): Modification of the SWAT Model to Simulate Regional Groundwater Flow Using A Multi-Cell Aquifer. Hydrological Processes 32(7), 939-953, DOI: 10.1002/hyp.11466.

Abstract

Global change is a new challenge in water resources management, in particular coping with extreme climate and weather situations and fulfilling the rising water demand for the food production for a growing world population.

The analysis and the management of water resources puts more emphasis on the regional scale of hydrological meso-scale river catchments. Simulation models in quantitative water resources planning and management need to deal with water availability and water demand within process adequate temporal and spatial scales. For the operation of structures and farms, there are highly specialized models on smaller scales available and established. For long-term analysis of water availability and food supply on larger scales, global models have been applied but identified as rather uncertain on regional scale.

The work presented here deals with the application and improvement of hydrological catchment models on meso-scale for water resources management in the context of global change. This cumulative habilitation thesis summarizes results from several research activities of the author in flood forecasting and in water resources management, followed by a critical review realted with the current state of scientific knowledge.

The dissemination of flood warnings can be improved with the availability of probabilistic meteorological ensemble forecasts. A seamless forecast chain of different ensembles with different lead times was developed. The uncertainty of the hydro-meteorological future development could be framed and produced better-informed decisions about warnings. The model ArcEGMO was enabled for operational flood forecasting. For estimating the characteristics and the impact of climate change, climate projections have been published. The use of new climatic datasets for water resources management requires a tailored validation. The systematic bias of precipitation, temperature and global radiation has high impact on the results of quatitative water resources studies. In catchment modelling, water demand models like WEAP are less established because of the simplification of the hydrological processes. However, their strength is scenario development under high uncertainty of external drivers like climate and socio-economy. WEAP was first applied and extended for the strategic planning of larger scale roof top rain water harvesting in a semi-arid area. The simulation of agricultural catchments with SWAT requires special consideration of the quality of soil moisture simulation and irrigation operations. The horizontal connectivity of the aquifer in SWAT was added by a new groundwater component.

Future research should focus on the improvement of flood warnings and the further development of hydrological connectivity and irrigation control in catchment models.

Key words (3): River Basin Modelling, Flood Forecasting, Water Scarcity

Zusammenfassung

Der globale Wandel stellt die Bewirtschaftung von Wasserressourcen vor neue Herausforderungen, zu denen die Bewältigung von extremen Klima- und Wettersituationen gehören wie auch der steigende Wasserbedarf für die Ernährungssicherung der Bevölkerung.

Die Betrachtung und Bewirtschaftung von Wasserressourcen erfolgt oft auf regionaler Ebene bzw. auf der Skala von Flusseinzugsgebieten. Die Anwendung von Simulationsmodellen als quantitative Planungswerkzeuge in der Wasserressourcenbewirtschaftung muss sowohl Wasserverfügbarkeit als auch Wassserbedarf abbilden und in der räumlichen und zeitlichen Auflösung prozessadäquat sein. Auf kleineren Skalenebenen (Feld) bestehen zur Planung und operationellen Steuerung von Anlagen oder landwirtschaftlichen Betrieben hoch aufgelöste Spezialmodelle. Fragen des langfristigen Wasserdargebotes und der Welternährung werden hingegen oft mit stark vereinfachenden globalen Modellen behandelt.

Diese Arbeit beschäftigt sich mit der Anwendung und Verbesserung von hydrologischen Flussgebietsmodellen zur Behandlung wasserwirtschaftlicher Herausforderungen des globalen Wandels. Forschungsergebnisse mehrerer Forschungsprojekte zu den Themen Hochwasser-vorhersage und Wassermengenbewirtschaftung unter Beteiligung des Autors werden in der hier vorliegenden kumulativen Habilitationsschrift zusammengefasst und mit Bezug zum aktuellen Forschungsstand bewertet.

Die Herausgabe von Hochwasserwarnungen kann durch neue Entwicklungen der probabilistischen Wettervorhersage mit Ensembles mit längerer Vorlaufzeit und mit Information über die erwartete hydro-meteorologische Unsicherheit erfolgen. Es wurde basierend auf verschiedenen Ensembleprodukten eine nahtlose Vorhersagekette entwickelt, in deren Zuge das Modell ArcEGMO zur operationellen Simulation von Hochwasserereignissen ertüchtigt wurde. Zur Abschätzung der Ausprägung und Folgen des Klimawandels sind Klimaprojektionen verfügbar, welche auch in der Wasserbewirtschaftung angewendet werden. Die Verwendung neuer Klimadaten erfordert deren Validierung zur Nutzung im vorgesehenen Anwendungskontext, wobei die Behandlung systematischer Fehler der Größen Niederschlag, Temperatur und Globalstrahlung für die Wasserressourcenbewirtschaftung von hoher Bedeutung sind. In der Flussgebietsmodellierung sind Wasserbedarfsmodelle wie WEAP aufgrund hydrologischer Vereinfachungen weniger etalbliert. Deren Stärke liegt vor allem in der Szenarioentwicklung und damit auch der Planung unter Unsicherheit der zukünftigen Entwicklung. WEAP wurde hierbei für die strategische Planung der großräumigen Nutzung von Regenwasser in semi-ariden Gebieten erweitert. In der Simulation landwirtschaftlicher Flusseinzugsgebiete mit SWAT wurde Entwicklungsbedarf bezüglich der Abbildung der Bodenfeuchte sowie der Steuerung der Bewässerung identifiziert. Die horizontale Vernetzung des Grundwasserleiters im Modell SWAT wurde durch eine neue Grundwasserkomponente hergestellt.

Weiterer Forschungsbedarf wird u.a. gesehen in der Verbesserung der Hochwasserwarnung sowie in der Verbesserung der Bewässerungssteuerung sowie der hydrologischen Konnektivität in Flussgebietsmodellen.

Schlagworte (3): Flussgebietsmodellierung, Hochwasservorhersage, Wassermangel

Inhaltsverzeichnis

1 Einleitung

1.1 Motivation

Die letzten beiden Dekaden brachten das Auftreten extremer Hochwasserereignisse (insbesondere an Elbe und Donau) sowie von Trockenheiten (Hitzewelle 2003) in das Bewusstsein der Öffentlichkeit zurück. Seit Mitte der 1990-er Jahre haben die globalen Durchschnittstemperaturen ihre langjährige Schwankungsbreite nach oben verlassen. Der Temperaturanstieg gilt als sicheres Anzeichen einer derzeit stattfindenden globalen Klimaveränderung, welche im geologischen Maßstab gesehen sehr schnell verläuft. Das Auftreten einzelner Wetterextreme kann im Rahmen der natürlichen Variabilität des Wetters liegen, die Häufung solcher Ereignisse kann jedoch ein Anzeichen des Klimawandels mit entsprechenden Auswirkungen auf den Wasserkreislauf und damit die Bewirtschaftung von Wasserressourcen sein.

Die Wasserwirtschaft, spezieller die Wasserressourcenbewirtschaftung, ist eine angewandte Wissenschaft, welche den Ausgleich zwischen dem natürlichen Wasserdargebot und der Wassernachfrage durch den Menschen behandelt. Zu den Aufgabenbereichen gehören folgende Fragen:

- zu viel Wasser: Schutz der Menschen und ihrer Investitionen vor Hochwasser;
- zu wenig Wasser: Sicherstellung der Wasserversorgung auch zu Mangelzeiten, einschließlich der Sicherung der Nahrungsmittelproduktion und der Energieproduktion aus Wasserkraft;
- sauberes Wasser: Erhalt und Sanierung der Wasserqualität;
- wasserabhängige Ökosysteme: Feuchtländer, Flussauen, Struktur und Durchgängigkeit von Fließgewässern als wesentliche Elemente eines guten Umweltzustandes.

Da das Wasserdargebot stark von Wetter (kurzfristig) und Klima (langfristig) abhängt, gehört die Berücksichtigung der äußeren wetter- und klimabedingten Einflüsse zu den wesentlichen Herausforderungen der wasserwirtschaftlichen Vorhersage (kurzfristig, operationell) und Planung (langfristig, strategisch).

Insbesondere die Überschwemmungen und Schäden der Hochwasserereignisse von 1999, 2002 und weiteren haben zu hohen Investitionen zur Verbesserung des Hochwasserrisiko-managements sowie zu Forschungsinitiativen und zur Verbesserung der Wetter- und Hochwasservorhersage geführt. Für den Umgang mit Trockenheiten und Hitzewellen haben keine vergleichbaren Aktivitäten im deutschen Raum stattgefunden. Im mediterranen Raum und in semi-ariden Regionen hingegen können steigende Temperaturen und ausbleibende Niederschläge gesellschaftlich relevanten Wassermangel hervorrufen oder verstärken. Insbesondere die Versorgung der Landwirtschaft mit Wasser für die Bewässerung, also letztlich die regionale Nahrungsmittelproduktion, sind von extremen klimatischen Situationen bedroht. Auf EU-Ebene und global stellen der Schutz der Menschen vor Wetter- und Klimaextremen sowie die Sicherstellung der Nahrungsmittelproduktion unter Klimawandel und Bevölkerungs-wachstum große Herausforderungen der Zukunft dar.

1.2 Entwicklung der Flussgebietsmodellierung

Systemanalytische Methoden und Werkzeuge sind eine wichtige Grundlage quantitativer Ansätze in der Bewirtschaftung der Ressource Wasser. Zu den klassischen systemanalytischen Methoden gehören Simulation, Optimierung und Entscheidungsunterstützung. Die Simulation wasserwirtschaftlicher Fragestellungen in Flussgebieten basiert im Wesentlichen auf hydrologischen Modellen, da hydrologische Prozesse die naturwissenschaftliche Basis für die Berechnung der räumlichen und zeitlichen Verfügbarkeit von Wasser darstellen. Wesentliche Antriebe für natürliche hydrologische Prozesse sind meteorologische Prozesse und die Schwerkraft. Für die Bewirtschaftung von Wasserressourcen spielen hydrologische Prozesse auf der Landoberfläche (vor allem der Niederschlag-Abfluss-Prozess) und im Untergrund (vor allem die Grundwasserneubildung) eine wichtige Rolle. Schneedecken und Gletscher sind wichtige hydrologische Speicher, so dass auch die Umwandlungen von Aggregatszuständen sowie Verweilzeiten in Kompartimenten in der Flussgebietssimulation berücksichtigt werden müssen. Während das Erkenntnisinteresse der Flussgebietssimulation in der Hydrologie vorwiegend die hydrologischen Prozesse und der Wasserhaushalt sind, steht bei der Bewirtschaftung der Wassserressourcen das für die menschliche Nutzung verfügbare Wasserdargebot im Vordergrund. Da die Aufgaben der Wasserwirtschaft auch den Gewässerschutz, allgemeiner den Schutz wasserabhängiger Ökosysteme umfassen, sind die Auswirkungen der Wassernutzung auf Wassermenge und Wassergüte häufig Gegenstand von Simulationen. Ferner zählt auch der Schutz der Menschen und ihrer Investitionen vor Hochwasser zu den wasserwirtschaftlichen Aufgaben.

Die genannten vielfältigen Anwendungsmöglichkeiten für Flussgebietsmodelle führen zu folgenden Anforderungen in der räumlichen und zeitlichen Auflösung:

- Raumskala: Kleine Einzugsgebiete ab ca. 10 km^2 (kleinste Auflösung der Einzugsgebiete der EU Wasserrahmenrichtlinie, Hochwasserentstehungsgebiete) bis zu großen Flussgebieten von mehreren 100.000 km^2. Operationelle Messstellen beschreiben häufig Einzugsgebiete ab wenigen 100 km^2, so dass Flussgebietsmodelle in der Regel mindestens eine Größenordnung von 100 km^2 abdecken, damit eine gute Modellkalibrierung anhand der Pegelzeitreihen erfolgen kann.
- Raumauflösung: Eine Abbildung der Landbewirtschaftung erfordert Gruppen von landwirtschaftlichen Feldern ähnlicher oder gleicher Vegetation zu berücksichtigen (ha bis km^2), während die Hydrologie in hügeligen und gebirgigen Lagen für die zugrundeliegenden Geländemodelle Auflösungen von 50 m und höher erfordert. Die Raumauflösung von Flussgebietsmodellen kann durch die Verfügbarkeit von Daten und durch die Rechenkapazitäten limitiert sein, insbesondere, wenn große Flussgebiete simuliert werden.
- Zeitskala: Für planerische Anwendungen und Bemessungsfälle sollten in der Hydrologie mindestens 30 Jahre kontinuierlich simuliert werden, um die meteorologische und hydrologische Charakteristik eines Einzugsgebietes beschreiben zu können. Für Klimaimpaktstudien werden Zeiträume von 100 Jahren und mehr verwendet. Für operationelle Vorhersagen sind hingegen oft Ereignissimulationen von wenigen Stunden bis Tagen ausreichend, z.B. in der Hochwasservorhersage und in der Bewässerungs-planung.
- Zeitauflösung: Die Hochwasservorhersage benötigt Auflösungen von 1 Stunde oder höher, während Wasserhaushaltssimulationen in der Regel eine Auflösung von 1 Tag oder sogar

von 1 Monat haben. Es sei angemerkt, dass weder 1 Tag noch 1 Monat prozessadäquat sind. Die hydrologische Antwortzeit von Flussgebieten auf Niederschlag und auch die Schneeschmelze liegt in der Regel deutlich unter einem Monat. Der Tag-Nacht-Zyklus von Temperatur und Strahlung ist von hoher Relevanz für die Bodenfeuchte und die Evapotranspiration sowie Schneeprozesse. Aufgrund limitierter Datenverfügbarkeit wurden Ansätze in Modelle eingebaut, welche mit meteorologischen Tageswerten arbeiten aber subkalige Prozesse vereinfacht abbilden (z.B. Sinusverläufe oder Korrekturfaktoren). Kaum angewendet wird die Simulation mit einer zeitlichen Auflösung von 6 Stunden, obwohl diese einen prozessnahen Kompromiss zwischen Stunden- und Tageswerten darstellt.

- Meteorologischer Antrieb: Flussgebietsmodelle benötigen Gebietswerte meteorologischer Größen. Beobachtungswerte von Stationen können als Antrieb dienen, müssen jedoch in der Regel durch geostatistische Methoden flächenhaft verteilt werden. Zur Simulation schlecht beobachter Gebiete können durch meteorologische Modelle erzeugte Re-Analysen der Vergangenheit verwendet werden. Für Vorhersagen stehen Daten numerischer Wettervorhersagemodelle zur Verfügung, während für Zukunftsprojektionen des Klimas entsprechende Klimaprojektionen verwendet werden können. Zunehmend werden auch Fernerkundungsdaten genutzt.

Im Maßstab Flussgebiet kommen vereinfachte hydrologische Modellansätze zur Anwendung. Anstelle von hydraulischen Modellen für die Wasserströmung in Boden, Grundwasser und Gerinne werden häufig Einzellinearspeicher eingesetzt, welche eine Füllung und Entleerung natürlicher Speicher abbilden und durch entsprechende Auslaufkonstanten Retentionseffekte simulieren können. Neben Materialeigenschaften (z.B. hydraulischen Kennwerten) werden oft Modellparameter angewendet, deren Werte erst im Rahmen einer Modellkalibrierung ermittelt werden und zumeist größeren Unsicherheiten unterliegen. Zur Untersuchung der Parametersensitivität und zur Parameteroptimierung wurden daher automatisierte Verfahren entwickelt (van Griensven et al., 2006; Baroni & Tarantola, 2014; Pianosi et al., 2016). Das Prinzip der Equifinalität (Beven & Binley, 1992) besagt allerdings, dass unterschiedliche Modellparameter dieselbe Modellgüte erzielen können und damit auch physikalisch unplausible Parameter ermittelt werden können, was insbesondere bei automatisierten Verfahren zur Kalibrierung berücksichtigt werden muss und zu hybriden Ansätzen aus physikalisch basierten und datenbasierten Verfahren zur Verbesserung der Physikalität automatisierter Kalibrierung geführt hat (Wallner et al., 2013). Die differenzierte Abbildung der räumlichen Verteilung von Klimagrößen und Landnutzung erfordert räumlich verteilte Modelle (Fenicia et al., 2016). Flussgebietsmodelle verwenden für die räumliche Diskretisierung entweder Raster oder homogene Flächen (z. B. Elementarflächen, Hydrotope). Dabei werden einzelne Parameter als Blockparameter für die gesamte Modelldomäne verwendet. Die vertikale Vernetzung erfolgt über mehrere Schichten des Bodens und der Grundwasserleiter. In der Horizontalen sind Flussgebietsmodelle oft nur über das Gewässernetz vernetzt. Grundwasserflüsse werden selten dargestellt, sind aber über die Kopplung mit speziellen Grundwassermodellen simulierbar. Für Langzeitsimulationen erfolgt in vielen Modellen eine „Aufwärmphase“, in welcher die hydrologischen Speicher gefüllt werden. Für die operationelle Vorhersage ist es erforderlich den Anfangszustand des Modells (Speicherfüllungen, z.B. Abfluss, Bodenfeuchte) zu schätzen, z.B. aus kontinuierlicher Simulation oder durch Datenassimilation. Dabei besteht das Problem, dass für die meisten Speicher von Flussgebietsmodellen keine Beobachtungsdaten vorliegen.

Die Anwendung von Flussgebietsmodellen konzentrierte sich in den letzten 10 bis 15 Jahren auf das Flussgebietsmanagement in der EU mit der planerischen Umsetzung der Wasserrahmenrichtlinie sowie der Hochwasserrisikomanagement-Richlinie. In dem Zeitraum wurde auch der Begriff der „Ökohydrologie“ geprägt (Rodriguez-Iturbe, 2000), wobei hier vor allem Klima, Boden und Vegetation in der Flussgebietsmodellierung besser berücksichtigt werden sollten. Für die Flussgebietsbewirtschaftung stellen dabei Nährstoffkreisläufe (Drewry et al., 2006) sowie das Management von Nährstoffeinträgen aus Punktquellen und diffusen Quellen wichtige Aspekte dar (Horn et al., 2004). Ein weltweit häufig angewendetes und ständig weiterentwickeltes ökohydrologisches Modell stellt das Soil and Water Assessment Tool (SWAT) dar (Arnold et al., 1998; Gassman et al., 2007; Bieger et al., 2017).

Die Kopplung mit ökologischen Modellen im engeren Sinne wurde nicht unter „Ökohydrologie“ gefasst. Flüsse und allgemeiner alle wasserbeeinflussten Lebensräume finden seit der EU Wasserrahmenrichtlinie, welche einen guten ökologischen Zustand der Gewässer fordert und anhand von biologischen Indikatoren bewertet, eine größere Beachtung als in der Anfangszeit der Flussgebietsmodellierung. Eine Modellkopplung zwischen physischen Habitateigenschaften und der Biozönose stellt aufgrund der Komplexität ökologischer Systeme eine große Herausforderung dar (Guse et al., 2015) und es bestehen Wissenslücken in der für die Flussgebietsbewirtschaftung relevanten Abbildung von Maßnahmen zur Verbesserung des ökologischen Zustands und deren Wirkungen. Fortschrittte sind beispielsweise in der Kalibrierung von hydrologischen Modellen auf artenrelevante hydrologische Indikatoren erfolgt (Kiesel et al., 2017). Auch ökologisch orientierte Bewertungsverfahren (ecological flow, ecosystem services) und naturbasierte Lösungen (nature based solutions, Keesstra et al., 2018) sind in jüngerer Vergangenheit erneut in den Fokus des Interesses gerückt.

Flussgebietsmodelle im Kontext der Wasserbewirtschaftung müssen relevante menschliche Tätigkeiten zeitvariant abbilden, um sowohl die Auswirkungen dieser Tätigkeiten als auch die Änderungen aufgrund von Bewirtschaftungsmaßnahmen simulieren zu können. In Hinblick auf die Wassermengenwirtschaft müssen vor allem Wasserentnahmen und Einleitungen erfasst werden, in Hinblick auf die Wassergütewirtschaft Tätigkeiten mit punktuellen und diffusen Einträgen von Stoffen, z.B. Kläranlageneinleitungen und Düngerausbringungen (Dietrich & Funke, 2007). Modelle stellen in der Flussgebietsbewirtschaftung Werkzeuge zur Planung und Entscheidungsunterstützung dar. Sie können als partizipative Modelle in partizipative Entscheidungsprozesse eingebunden werden und damit die Effektivität der Modellieurng erhöhen (Falconi & Palmer, 2017). Im Unterschied zu hydrologischen Modellen simulieren Wasserbedarfsmodelle bzw. Wasserbewirtschaftungsmodelle die Deckung des Wasserbedarfs unterschiedlicher Nutzer aus dem Wasserdargebot. Über Szenariorechnungen werden verschiedene Varianten der Bewirtschaftung unter Berücksichtigung äußerer Rahmenbedingungen simuliert.

Im globalen Maßstab wie auch regional fanden zahlreiche Untersuchungen über die Auswirkungen des Klimawandels auf die Wasserwirtschaft statt. Flussgebietsmodelle sind ein oft angewandtes Hilfsmittel bei der Klimafolgenabschätzung (z.B. Barthel et al., 2012). Hierbei kommt der Simulation vergletscherter Gebiete eine besondere Rolle zu (Chen et al., 2017).

Zu den modelltechnischen Verbesserungen der letzten Jahre gehört die Verbesserung der hydrologischen Konnektivität der Modelle (Bracken & Croke, 2007), wobei hier insbesondere die

Anbindung von Feuchtgebieten und Gewässerrandstreifen Beachtung fand (Hattermann et al., 2006; Ameli & Creed, 2017). In Bezug auf das Wasserressourcenmanagement wurden neue Ansätze zur Einbeziehung kleiner Speicher in ariden Regionen entwickelt (Mamede et al., 2018). Experimentell werden Methoden der Isoptopen- und Tracerhydrologie zum Nachweis hydrologischer Konnektivität verwendet, wobei zunehmend die Nutzbarmachung besser verfügbarer Güteparameter für das Verständnis von hydrologischen Prozessen in Flussgebieten (fingerprinting) erschlossen wird und damit zur weiteren Verbesserung entsprechender Flussgebietsmodelle beitragen kann (Birkel & Soulsby, 2015; Birkel et al., 2015; Collins et al., 2017).

1.3 Ensemble-Vorhersagen

Vor dem Aufkommen numerischer Wettervorhersagemodelle basierte die Wettervorhersage auf Beobachtungsdaten und Expertenwissen bzw. Erfahrung und tradierten Regeln der Bevölkerung („Bauernregeln“ oder auch „indigenes Wissen“). Deterministische numerische Wettervorhersagen (numerical weather prediction, NWP) erlauben bei entsprechender Vorlaufzeit oberhalb der Konzentrationszeit des betreffenden Flussgebietes die Simulation deterministischer Abflussvorhersagen. Aufgrund der hohen Nichtlinearität des atmosphärischen Systems und der unvollständigen Kenntnis der Anfangsbedingungen ist eine exakte Wettervorhersage allerdings physikalisch nicht möglich (Lorenz, 1969). In der operationellen Meteorologie fand ein Paradigmenwechsel statt, infolgedessen die Unsicherheit der Wettervorhersage anerkannt wurde und Methoden zur Quantifizierung und Kommunikation der Unsicherheit entwickelt wurden. Vor allem in der ersten Dekade des 21. Jahrhunderts wurden die deterministischen numerischen Wettervorhersagen der operationellen Dienste um Ensemblevorhersagesysteme (ensemble prediction systems, EPS) ergänzt. Ensembles stellen eine Menge alternativer simulierter Wettervorhersagen da, welche alle für den gleichen Ort und die gleiche Zeit gelten und bei guter Konstruktion des Ensembles in ihrer Gesamtheit die Unsicherheit der zukünftigen Wetterentwicklung einrahmen können, zumindest jedoch verschiedene plausible Szenarien anbieten. Durch Nutzung der Mitglieder des EPS als Antrieb für hydrologische Modelle kann damit eine Ensemble-Hochwasservorhersage simuliert werden. Cloke & Pappenberger (2009) geben einen Überblick über erste Anwendungen von Ensembles in der Hochwasservorhersage.

Folgende Generierungsmechanismen bzw. Typen von Ensembles wurden entwickelt:

- Lagged Average Forecast (LAF, Hoffman & Kalnay, 1983): deterministische Vorhersagen unterschiedlicher Modellstartzeitpunkte desselben Modells werden kombiniert. Dahinter steht die Annahme, dass Wettervorhersagen bei der Annäherung an das vorhergesagte Ereignis nicht zwangsläufig besser werden.
- Multi-Modell-Ensemble (Georgakakos et al., 2004; Diomede et al., 2008): deterministische Vorhersagen unterschiedlicher Modelle zum selben Zeitpunkt werden kombiniert. Es wird angenommen, dass es kein generell bestes Modell gibt und die Vorhersagen vieler Modelle im Mittel bessser sind als eine einzelne Vorhersage (Najafi & Moradkhani, 2016).
- Single-Modell-Ensemble (Palmer, 2000): Störung von Anfangs- und Randbedingungen oder Parametern eines Modells oder Verwendung unterschiedlicher Konvektionsschemata (physikalisches Ensemble).

Je nach räumlicher Abdeckung und Vorlaufzeit kommen unterschiedliche Modelltypen zur Anwendung, wobei für die Kürzestfristvorhersage (Vorlaufzeiten unter einem Tag) räumlich hoch aufgelöste Modelle mit physikalischer Auflösung der Konvektion angewendet werden. Während die hier dargestellten Typen von Ensembles für die Mittel- bis Kürzestfrist mit Vorlaufzeiten von rund zwei Wochen bis zu wenigen Stunden entwickelt wurden, konnten neue Entwicklungen die Vorhersagezeiträume in Richtung Echtzeitvorhersage (nowcasting) und sub-saisonale bis saisonale Vorhersage erweitern mit dem Ziel einer nahtlosen Abdeckung verschiedener Raum- und Zeitauflösungen sowie verschiedener Anwendungen durch unterschiedliche Modelle („seamless prediction“, WMO, 2015).

Ensemblevorhersagen geben typischerweise keine Wahrscheinlichkeiten für die einzelnen Mitglieder des Ensembles aus, sondern nehmen an, dass alle Mitglieder gleich wahrscheinlich sind und die Gesamtheit der möglichen Entwicklungen abdecken. Damit erlauben Ensembles lediglich eine frequentistische Auswertung: sagen beispielsweise 8 von 10 Mitgliedern eine Niederschlagsmenge von mehr als 10 mm voraus, so wird üblicherweise für die Schwelle von 10 mm Niederschlag eine Überschreitungswahrscheinlichkeit von 80 % angegeben. Es ist jedoch zu erwarten und erst bei Vorliegen von Beobachtungsdaten zu quantifizieren, dass die einzelnen Mitglieder unterschiedlich gut sind. Ferner kommt es häufig vor, dass das Ensemble die Unsicherheit nicht vollständig einrahmt. Insofern sind die zuvor genannten 80 % keine probabilistische Information im engeren Sinne. Für die probabilistische Bewertung von Ensembles haben sich daher verschiedene Techniken etabliert, welche auf der Assimilation von Beobachtungsdaten sowie einem statistischen Postprocessing basieren. Alternativ zu und kombinierbar mit Ensemblevorhersagen haben sich weitere Techniken der probabilistischen Vorhersage etabliert, welche auf Monte-Carlo-Simulationen und Bayes‘ scher Statistik basieren (Krzysztofowicz, 1999; Duan et al., 2007; Han & Coulibaly, 2017). Eine Herausforderung in der probabilistischen Hochwasservorhersage stellt die Überlagerung und Fortpflanzung von Fehlern in der operationellen Vorhersagekette dar, wobei sich die Unsicherheit der Niederschlagsvorhersage als dominierende Unsicherheitsquelle gegenüber hydrologischen Modellparametern und Anfangszuständen herausgestellt hat (Zappa et al., 2011). Alternative wahrscheinlichkeitstheoretische Ansätze aus dem Bereich „imprecise probabilities“ sind vielversprechend, haben aber bisher gerige Anwendung in Hydrologie und Wasserwirtschaft gefunden, z. B. die Theorie nach Dempster-Shafer in Limbourg und de Rocquigny (2010).

Da Entscheidungen über die Herausgabe von Warnungen oder über andere operationelle Maßnahmen oft binären Charakter haben, müssen Ensembles in der Praxis oft reduziert werden auf eine der deterministischen Vorhersage vergleichbare Form. Dies kann z. B. durch Auswahl des besten Mitglieds sowie durch abgeleitete Größen aus der Kombination der Ensemblemitglieder unterstützt werden (Diks und Vrugt, 2010). Raftery et al. (2005) entwickelten dafür den „Bayesian Model Average (BMA)“, welcher von Ajami et al. (2007) in ein Framework für die Unsicherheitsquantifizierung der hydrologischen Vorhersage integriert wurde. Li et al. (2018) zeigten den Mehrwert der Verwendung mehrer hydrologischer Modelle in diesem Framework. Probleme bei der Anwendung des BMA entstehen, wenn nur eine geringe Datenbasis von Ensemblevorhersagen für das Training verfügbar ist. Dietrich et al. (2009) schlugen die Bildung so genannter Super- und Subensembles vor. Hemri et al. (2013) entwickelten das ursprünglich univariate BMA weiter zu einer multivariaten Version, welche zur Kombination von Ensembles verschiedener Vorlaufzeiten zur Anwendung kam. Weitere Probleme bei der Anwendung des

BMA in der hydrologischen Vorhersage sind die notwendige Transformation der Verteilung der Abflusswerte in spezielle Verteilungsfunktionen sowie die Voraussetzung, dass die simulierten Abflusswerte keinen systematischen Fehler (bias) aufweisen. Zur Vermeidung dieser Probleme integrierten Madadgar & Moradkhani (2014) Copulas in BMA und demonstrierten die Anwendung mit einem aus sieben hydrologischen Modellen bestehenden Multi-Modell-Ensemble in den USA. Bogner et al. (2017) untersuchten neben BMA auch „nonhomogeneous Gaussian regression (NGR)“ und „Beta-transformed linear pooling (BLP)“ als Methoden zur Kombination von Ensemblemitgliedern. Diese Studie bestätigte sowohl den Mehrwert der Kombination von Mitgliedern für die Entscheidungsunterstützung mit Vorteilen für BLP, zeigte aber auch die zeitlich starke Variation der Gewichte und die Notwendigkeit einer guten Datenbasis für die Kalibrierung der Kombinationsmethoden.

Neben Kombinationsverfahren wurden auch Methoden der künstlichen Intelligenz angewendet, um aus Modellvorhersagen Entscheidungen abzuleiten. Mediero et al. (2007) nutzten Bayes’sche Netzwerke, um Schlussfolgerungen für die Steuerung von Speichern aus simulierten Abflüssen zu ziehen. Doycheva et al. (2017) verwendeten Support Vector Machines, Multilayer Perceptron und Rotation Forest Klassifikation zur Einengung der Unnsicherheitsbandbreiten von Ensembles.

Der ökonomische Mehrwert der Unsicherheitsquantifizierung für die Entscheidungsunterstützung in der Hochwasservorhersage bzw. der Frühwarnung wurde von Thiboult et al. (2017) untersucht. Die Autoren kommen zu dem Schluss, dass Frühwarnungssysteme generell einen Mehrwert darstellen und dass Systeme mit Einbeziehung der Vorhersageunsicherheit die Schadenserwartung am besten senken können.

Aktuell sind für Mitteleuropa folgende meteorologische Ensemble-Vorhersagesysteme mit Relevanz für Anwendungen in der Hochwasservorhersage im operationellen Einsatz:

- COSMO-DE-EPS (Theis et al., 2012), gerechnet am DWD, operationell seit 2012, Deutschland und Nachbargebiete, Gitterweite 2,8 km (experimentell 1 km), Vorhersage für die nächsten 27 Stunden, Single-Model-Ensemble, Antrieb durch ICON seit 2017. Anwendung in der Hochwasservorhersage z. B. in Rheinland-Pfalz (Bartels et al., 2017)
- ICON, operationell seit 2015, Vorgänger GME und COSMO-EU
- SRNWP-PEPS, zusammengestellt am DWD, Europa, Vorhersage für die nächsten 2 Tage, multi-model ensemble.
- ECMWF ENS, Europäisches Zentrum für Mittelfristige Wettervorhersage, global, Vorhersage für die nächsten 15 Tage, single-model ensemble.
- COSMO-LEPS, geliefert von ARPA-SIMC (Bologna), im Rahmen des COSMO Konsortiums. Gebiet über Zentral- und Südeuropa, Vorhersage für die nächsten 5 Tage, single-model ensemble.

Erste wissenschaftlich publizierte Fallstudien für die operationelle Anwendung von Ensemble-Hochwasservorhersagen in Europa wurden im Rahmen des Projektes MAP D-PHASE in den Alpen und Zentraleuropa durchgeführt (Zappa et al., 2008), für den Rhein (Reggiani & Weerts, 2008) sowie für die Mulde (Dietrich et al., 2008). Auf europäischer Ebene wurde das European Flood Awareness System (EFAS) entwickelt (Thielen et al., 2009). Wetterhall et al. (2013) geben einen Überblick über operationelle hydrologische Ensemblevorhersagen.

Die Anwendung hydrologischer Modelle in der operationellen Hochwasservorhersage nahe Echtzeit erfordert u.a. eine hohe Recheneffizienz für die Simulation von Ensembles für Flussgebiete sowie einen hohen Automatisierungsgrad für die Bereitstellung der Ergebnisse mit minimaler Interaktion durch Fachleute in Hochwasservorhersagezentralen. Ferner ist abweichend von der oft üblichen kontinuierlichen Simulation die Berücksichtigung und Aktualisierung hydrologischer Anfangsbedingungen und Systemzustände von Bedeutung. Insbesondere konzeptionelle hydrologische Modelle, welche eine Kalibrierung von Parametern erfordern, müssen für die ereignisbasierte Berechung von Hochwasserwellen auf andere Indikatoren hin kalibriert werden als für eine kontinuierliche Niederschlag-Abfluss-Modellierung.

1.4 Simulation bewässerter Flussgebiete

Die landwirtschaftliche Bewässerung stellt global die wichtigste Wassernutzung dar. Der globale Wandel umfasst sowohl sozio-ökonomische Veränderungen als auch den Klimawandel. Das Bevölkerungswachstum sowie das Wachstum des Wohlstands mit veränderten Konsumgewohnheiten werden zu einem größeren Bedarf an Nahrungsmitteln führen. Ferner führt der Klimawandel zu einem erhöhten Strahlungsinput und einer erhöhten Temperatur, was wahrscheinlich zu einer erhöhten Evapotranspiration und damit zu höherem Pflanzenwasserbedarf führen wird. Beide Faktoren des globalen Wandels zusammen zeigen die hohe Relevanz der zukünftigen Entwicklung der Bewässerung sowohl bezüglich Wasserquantität als auch Wasserqualität bei einer angenommenen Verdoppelung des Bedarfs an Nahrungsmitteln (Assouline et al., 2015). Insbesondere in Teilen Asiens erfordert die zukünftige Entwicklung eine Anpassung der Nahrungsproduktion (Biemans et al., 2013).

Bewässerung wird auf unterschiedlichen Skalen simuliert. Einige der bekannten globalen hydrologischen Modelle simulieren den Bewässerungsbedarf (z. B. WaterGAP, Döll & Siebert, 2002 sowie Wisser et al., 2008). Wriedt et al. (2008) berechneten den Bewässerungsbedarf für das Gebiet der Europäischen Union. Auf kontinentaler bis globaler Skala wird die am Standort vorliegende Kombination aus Bodeneigenschaften und Pflanze genausowenig aufgelöst wie die konkrete Steuerung der Bewässerung. Es interessiert vielmehr der Wasserbedarf der Landwirtschaft als ein wichtiger Wassernutzer. Aufgrund der starken Vereinfachung und der reduzierten Auflösung solcher Modelle sind Modellunsicherheiten unvermeidbar und die Ergebnisse können regional große Fehler aufweisen. Haddeland et al. (2014) simulierten daher den anthropogenen Einfluss auf die Wasserressourcen unter Berücksichtigung des Klimawandels mit einem Ensemble aus sieben globalen Modellen. Ferner spielt die landwirtschaftliche Wassernutzung (und damit vor allem die Bewässerung) eine wichtige Rolle bei der Berechnung des „Wasser-Fußabdrucks“, so dass entsprechende Berechnungen auf globaler Ebene für umweltpolitische und planerische Entscheidungen hilfreich sind (Liu & Yang, 2010; Ercin & Hoekstra, 2014).

Auf der Feldskala hingegen erfolgt die Simulation der Bewässerung oft mit 1D-Modellen, welche die Wassserbewegung in der ungesättigten Zone physikalisch simulieren (z. B. SWAP mit der Richards-Gleichung, van Dam et al., 1997). Stärken dieser Art Modelle liegen in der standortgerechten Berücksichtigung des Pflanzenwachstums und der Möglichkeit, die benötigten Bewässerungsmengen und die Zeitpunkte der Wassergaben ebenfalls standortgerecht zu ermitteln. Bastiaanssen et al. (2007) geben einen Überblick über 25 Jahre der Modellierung von

bewässerten und entwässerten Böden und kommen zu dem Schluss, dass die Komplexität der Wasserbewegung in der ungesättigten Zone sowie des Pflanzenwachstums zwar von Modellen gut beschrieben werden kann, jedoch nicht von den Anwendern solcher Werkzeuge beherrscht wird, welche sich oft in ariden und semi-ariden Regionen und damit in Entwicklungsländern befinden. Daher besteht immer noch ein Defizit in der praktischen Anwendung der hydrologischen Modellierung in der Bewässerungsplanung und Steuerung auf der operativen Ebene. Die Autoren sehen weitere wichtige Entwicklungen in Richtung wassersparender Bewässerungssysteme sowie Produktivität der landwirtschaftlichen Produktionssysteme (Ertragssteigerung). Modelle dieser Art können vor allem in der landwirtschaftlichen Beratung und in der Forschung eingesetzt werden.

Für eine regionale Betrachtung werden zusammenhängende Bewässerungsgebiete (Gruppen von Feldern auch mit unterschiedlicher Vegetation) und Flussgebiete betrachtet. Insbesondere für letztere ist es erforderlich und möglich, die Bewässerung in die Wasserbilanz des Einzugsgebietes einzubeziehen (sowohl bezüglich Wasserverfügbarkeit als auch Wasserbedarf), die Quellen des Bewässerungswassers sowie dessen Verbleib und generell Zusammenhänge im Flussgebiet mit zu simulieren (Kite & Droogers, 1999). Im Flussgebietsmaßstab können auch Fragen der Grundwasserabsenkung und des Einflusses der Bewässerung auf den Abfluss (Huang et al., 2012) simuliert werden. Aufgrund der unterschiedlichen Stärken und Schwächen der verschiedenen Arten von Modellen wurden multiskalige Studien durchgeführt. Droogers and Kite (2001) simulierten auf der Feldskala mit SWAP, aggregierten die Ergebnisse für Bewässerungsgebiete (Droogers et al., 2000) und simulierten das Flussgebiet mit dem hydrologischen Modell SLURP (Kite, 2000). Andere Flussgebietsmodelle mit Berücksichtigung der Bewässerung sind WASIM (Ahrends et al., 2008; Uribe et al., 2009), SHE (Lohani et al., 1993; Mishra et al., 2005), WEAP (Raskin et al., 1992; Mehta et al., 2013; Varela-Ortega et al., 2016), SWAT und andere. Hydrologische Modelle werden häufig nur am (räumlich-zeitlich integrierenden) Abfluss aus dem Untersuchungsgebiet kalibriert. Die Verfügbarkeit von flächendeckenden Fernerkundungsdaten, z.B. für die Bodenfeuchte und Evopotranspiration, ermöglicht eine flächendifferenzierte Modellkalibrierung unter Berücksichtigung der für die Bewässerung wichtigen räumlich-zeitlichen Verteilung von Bodenfeuchte und/oder Evapotranspiration. Immerzeel et al. (2008) sowie Gampe et al. (2016) kalibrierten beispielsweise Flussgebietsmodelle anhand der aus Fernerkundungsdaten ermittelten Evapotranspiration.

Das ökohydrologische bzw. agrohydrologische Modell SWAT ermöglicht die Simulation hydrologisch relevanter landwirtschaftlicher Aktivitäten, zu denen auch die Bewässerung gehört. Ficklin et al. (2009) simulierten die Klimasensitivität eines intensiv landwirtschaftlich genutzten Einzugsgebietes mit SWAT. Abbaspour et al. (2015) wendeten das Modell SWAT für eine europaweite Simulation von Hydrologie und Wasserqualität unter Einbeziehung landwirtschaftlicher Tätigkeit an. Luan et al. (2018) berechneten den regionalen Wasser-Fussabdruck der Landwirtschaft. In mehreren Studien wurde SWAT zur Entwicklung von Bewässerungsstrategien planerisch eingesetzt (z.B. Maier & Dietrich, 2016 in Niedersachsen; Udias et al., 2018 auf Kreta). Die Bewässerungssteuerung in SWAT wurde in den letzten Jahren verbessert. Xie & Cui (2011) erweiterten SWAT zur Simulation des Reisanbaus. Dechmi et al. (2012) modifizierten SWAT, so dass für intensiv bewässerte Felder Bewässerungsmengen oberhalb der Feldkapazität der Böden aufgebracht und anschließend in der Wasserbilanz

berücksichtigt werden können. Chen et al. (2018) verbesserten die automatische Bewässerungssteuerung in SWAT.

2 Zielstellung

In vorliegenden kumulativen Habilitationsschrift soll die Simulation von Flussgebieten als systemanalytisches Werkzeug der Wasserressourcenbewirtschaftung weiterentwickelt werden, insbesondere für Untersuchungen über die Auswirkungen von Wetter und Klima auf die planerische und operationelle Wasserwirtschaft. Folgende Ziele stehen im Vordergrund:

- Entwicklung von Techniken zur Simulation und Auswertung von Ensembles im Bereich der operationellen Hochwasservorhersage unter Unsicherheit;
- Untersuchungen zur Validierung und Nutzung von hydrologischen und meteorologischen Daten in Flussgebietssimulationen, speziell Klimaprojektionen für die Klimafolgenabschätzung sowie Fernerkundungsdaten;
- Weiterentwicklung von Flussgebietsmodellen zur besseren Simulation wasserwirtschaftlich relevanter hydrologischer Prozesse, speziell Ergänzung des Wasserbewirtschaftungsmodells WEAP für die Anwendung bei der Regenwassergewinnung von Dachflächen sowie Verbesserung der Grundwasserkomponente des agrohydrologischen Modells SWAT durch bessere horizontale Vernetzung.
- Arbeiten zur Evaluaierung und Verbesserung der Simulation bewässerter Flusseinzugsgebiete mit dem Modell SWAT.

Die als Unterkapitel eingefügten neuen Originalarbeiten wurden in begutachteten Zeitschriften veröffentlicht und behandeln verschiedene in Abb. 1 dargestellte methodische Aspekte der Flussgebietsmodellierung. Diese kumulative Habilitation deckt bewusst mehrere Felder der Bewirtschaftung von Wasserressourcen ab, um dem interdisziplinären Charakter dieses Faches gerecht zu werden.

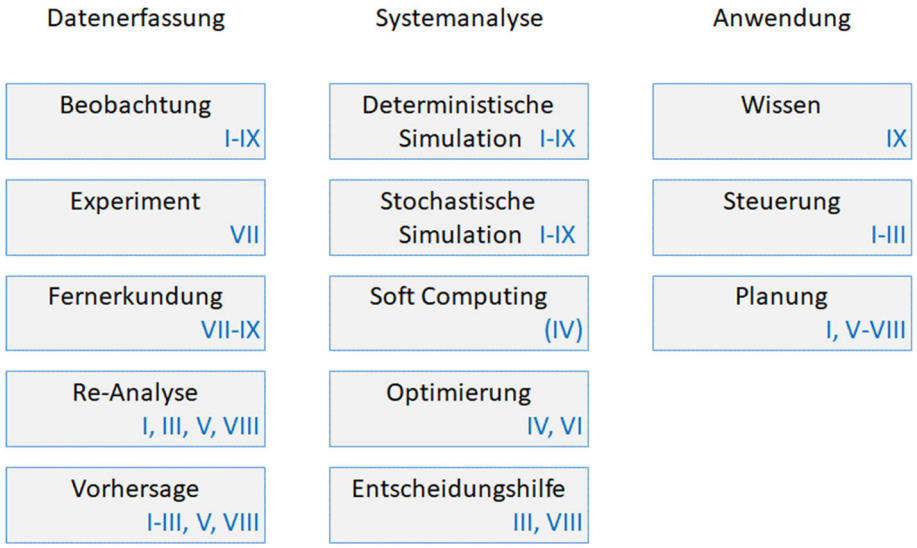


Abb. 1: Übersicht der in diesem Habilitationsvorhaben angewendeten Methoden mit Verweis auf diejenigen Artikel, in denen die jeweilgen Methoden ein Untersuchungsgegenstand waren.

3 Originalarbeiten

3.1 Zusammenfassung

In diesem Unterkapitel folgt eine Zusammenfassung der wesentlichen Inhalte und Erkenntnisse der Fachartikel dieser kumulativen Habilitationsschrift, während die eigentlichen Artikel im vollständigen Wortlaut der angenommenenen und publizierten Manuskripte im zweiten Unterkapitel dieses Kapitels dokumentiert werden. Die Fachartikel werden mit römischen Zahlen nummeriert und innerhalb der Fachartikel erfolgt die Nummerierung der Kapitel entsprechend der Originalveröffentlichung.

Die Rolle des Verfassers dieser Habilitationsschrift wird in drei Klassen für die Artikel angegeben:

Erstautorschaft: Hauptverantwortliche Koordinierung der Arbeiten und wesentlicher Anteil an der textlichen und graphischen Ausarbeitung des Artikels, korrespondierender Autor für die Einreichung und den Begutachtungsprozess.

Mitautorschaft aus wissenschaftlicher Zusammenarbeit: Beteiligung an den im Artikel beschriebenen Forschungsergebnissen durch selbständige eigene Arbeitspakete oder durch koordinierende und beratende Tätigkeit, Beteiligung an der Ausarbeitung und Revision des Artikels.

Mitautorschaft als Themensteller und Betreuer: Der Erstautor oder die Erstautorin wurde als Promovierende oder Promovierender vom Verfasser dieser Habilitationsschrift bei der Durchführung der Forschungsarbeiten und bei der Verfassung des Artikels hauptverantwortlich angeleitet, einschließlich der Themenstellung.

I. Dietrich, J.; Trepte, S.; Wang, Y.; Schumann, A. H.; Voß, F.; Hesser, F. B.; Denhard, M. (2008): Combination of different types of ensembles for the adaptive simulation of probabilistic flood forecasts. Hindcasts for the Mulde 2002 extreme event. ***Nonlin. Processes Geophys.*** **15 (2), 275–286. DOI: 10.5194/npg-15-275-2008.**

Rolle: Erstautorschaft als PostDoc in einem BMBF-Verbundvorhaben RIMAX

Die zukünftige Entwicklung des Wetters kann nicht exakt vorhergesagt werden. Ensemble-Vorhersagen liefern eine größere Zahl an möglichen Entwicklungen. Es gibt verschiedene Typen von Ensembles, beispielsweise Multi-Modell-Ensembles, Ensembles mit Variation der Anfangsparameter oder der Modellgleichungen sowie „Lagged Average"-Ensembles mit Kombination verschiederner Rechenläufe eines Modells. Ein gutes Ensemble rahmt die Unsicherheit der Vorhersage ein. Meteorologische Ensemblevorhersagen können mit Niederschlag-Abfluss-Modellen zur Anwendung in der operationellen Hochwasservorhersage gekoppelt werden.

In dieser Studie erfolgte die Entwicklung einer nahtlosen Vorhersagekette mit Vorlaufzeiten von fünf Tagen bis zu einem Tag durch Verwendung dreier verschiedener Ensemble-Vorhersagen. Diese Technik wurde später als „seamless prediction" bezeichnet und mit neueren Ensemble-

Vorhersagen weiterentwickelt. Das hydrologische Modell ArcEGMO wurde von einem Wasserhaushaltsmodell zu einem kontinuierlichen und ereignisbasierten Simulationsmodell entwickelt, wobei für letzteres die Ermittlung und Nachführung der Anfangszustände des hydrologischen Systems im Vordergrund standen (Bodenfeuchte, Füllung der Grundwasserspeicher und aktueller Abfluss).

Als Fallbeispiel wurden Hindcast-Simulationen des extremen Hochwasserereignisses im August 2002 im Erzgebirge mit den Ensembles COSMO-LEPS und COSMO-DE gerechnet, wobei letzteres hier als „Lagged Average" Ensemble angewendet wurde und inzwischen als physikalisches Ensemble operationell in Deutschland auch in der Hochwasservorhersage eingesetzt wird. Abflussimulationen erfolgen für Teileinzugsgebiete der Mulde im Erzgebirge.

Die Funktionsfähigkeit einer gekoppelten meteorologisch-hydrologischen Ensemble-vorhersage konnte demonstriert werden. Die Ensemblesimulationen konnten das beobachtete Extremeregnis 2002 weitgehend einrahmen und stellten damit einen Mehrwert gegenüber der deterministischen Vorhersage dar. Die Qualität der Ensemble-Vorhersage stellte sich für eine verschiedene Anzahl von Vorlauftagen unterschiedlich dar, wobei mit drei Tagen Vorlauf eine Frühwarnung vor dem Ereignis aufgrund der vorhergesagten hohen Wahrscheinlichkeit der Überschreitung extremer Abflussmengen > HQ 100 möglich gewesen wäre. Einschränkend ist festzustellen, dass nachträglich berechnete Extremsituationen mit Wissen und Modellen angefertigt werden, welche gegenüber dem realen Zeitpunkt des Ereignisses noch nicht vorhanden waren. Zum Zeitpunkt der Studie befanden sich operationelle Ensemblesysteme der Wetterdienste in einem frühen Entwicklungsstadium, so dass der verfügbare Datenumfang gering war.

II. **Dietrich, J.; Schumann, A. H.; Redetzky, M.; Walther, J.; Denhard, M.; Wang, Y. et al. (2009): Assessing uncertainties in flood forecasts for decision making. Prototype of an operational flood management system integrating ensemble predictions.** ***Nat. Hazards Earth Syst. Sci.*** **9 (4), 1529–1540. DOI: 10.5194/nhess-9-1529-2009.**

Rolle: Erstautorschaft als PostDoc in einem BMBF-Verbundvorhaben RIMAX

Die Einführung von Ensemblevorhersagen soll eine probabilistische Vorhersage ermöglichen, welche beispilesweise die Angabe von Überschreitungswahrscheinlichkeiten für wichtige Schwellenwerte angibt, konkret Alarm- oder Warnstufen für Hochwasser. Ensembles verfügen jedoch aufgrund der Rechenkapazitäten nur über eine relativ geringe Anzahl an Mitgliedern (bis zu wenige Dutzend), für welche wiederum angenommen wird, dass sie gleich wahrscheinlich sind. In dieser Studie erfolgt eine Diskussion der frequentistischen Natur „roher" Ensembles versus den Anspruch einer probabilistischen Vorhersage mit Angabe von echten Überschreitungs-wahrscheinlichkeiten.

Es wird in dieser Studie ein Ansatz zur Erzeugung einer probabilistischen Vorhersage durch Wichtung von Ensemblemitgliedern nach Training an Beobachtungsdaten vorgestellt. Dabei erfolgte zunächst eine Kombination von ausgewählten Ensemblemitgliedern unterschiedlicher Systeme zu einem Superensemble und daraus die Ableitung eines probabilistischen Subensembles durch statistisches Postprocessing mit Bayesian Model Averaging (BMA).

Die in Artikel I. vorgestellte nahtlose Vorhersagekette wurde für die Integration in ein operationelles Hochwasservorhersagesystem weiterentwickelt. Dies beinhaltete die Entwicklung eines technischen Systemrahmens zur Kombination von meteorologischen Ensembledaten mit dem für dieses Vorhaben weiterentwickelten Niederschlag-Abfluss-Modell ArcEGMO und GIS-Daten zu Überschwemmungsgebieten. Für die ressourceneffiziente Berechnung von Überflutungsflächen wurde ein Ersatzmodell entwickelt, welches unter Verwendung vorab durchgeführter hydraulischer Simulationen für die jeweiligen Mitglieder des simulierten Abfluss-Ensembles die zugehörigen Überflutungsflächen zuweist. Durch die Entwicklung hydrologischer Parameter-Ensembles konnte die Unsicherheitspropagation auf die ereignisbasierte Niederschlag-Abfluss-Simulation erweitert werden.

Die Demonstration des technischen Vorhersagesystems erfolgte am Fallbeispiel des Extremhochwassers 2002 und als probabilistische Vorhersage am Fallbeispiel einer einjährigen Periode mit Verfügbarkeit mehrerer Ensembles für 2007/2008.

Es konnte eine geschlossene Vorhersagekette von der Niederschlagsvorhersage über die Abflussvorhersage hin zur Überflutungsvorhersage für die Mittel- bis Kurzfrist in mesoskaligen Flusseinzugsgebieten dokumentiert werden, wobei der Niederschlag-Abfluss-Prozess probabilistisch simuliert wurde. Ferner wurde die technische Umsetzbarkeit in ein Vorhersagesystem für den Einsatz in der operationellen Hochwasservorhersage für mesoskalige Flussgebiete demonstriert. Es zeigte sich in der Fallstudie, dass der Trainingsdatensatz für die BMA-Methode sehr kurz war und vor allem keine Extremereignisse umfasste. Die erzielten Gewichte der Ensemble-Mitglieder konnten daher weder als allgemeingültig noch robust angesehen werden. Die probabilistische Vorhersage führte nicht immer zu einer Verbesserung gegenüber der frequentistischen Betrachtung des Ensembles. Die hydrologische Ensemblebildung konnte erfolgreich getestet werden. Aufgrund der Überlagerung von Parameter- und Datenunsicherheit während der Kalibrierung kann das erzielte Parameterensemble durch systematische oder zufällige Fehler in den Eingangsdaten beeinflusst sein, was dessen Anwendung auf unbeobachtete Situationen einschränkt bzw. eine häufige Re-Kalibrierung nach Erkenntnisgewinn durch neue Beobachtungsdaten notwendig macht. Die Arbeiten zeigten damit auch die weiterhin bestehende Notwendigkeit der schnellen und zuverlässigen Verfügbarkeit hochwertiger Messdaten.

III. **Dietrich, J.; Denhard, M.; Schumann, A. H. (2009): Can ensemble forecasts improve the reliability of flood alerts?** ***Journal of Flood Risk Management*** **2 (4), 232–242. DOI: 10.1111/j.1753-318X.2009.01039.x.**

Rolle: Erstautorschaft als PostDoc in einem BMBF-Verbundvorhaben RIMAX

In dieser Studie erfolgte eine Erweiterung der Hindcast-Simulationen mit den drei Ensemblesystemen COSMO-LEPS, SRNW-PEPS und COSMO-DE/LAF auf den Zeitraum 2002 bis 2008. Die Bandbreite der Ensembles war für eine Vorlaufzeit von drei bis fünf Tagen sehr hoch und engte sich für die letzten beiden Tage sichtbar ein. Erstmals wurde die Zuverlässigkeit der probabilistischen kategorischen Vorhersage der Überschreitungswahrscheinlichkeit der Hochwasserwarnstufen mit Reliability-Diagrammen bewertet. Auch wenn der geringe Umfang an Ereignissen und die Beschränkung auf tatsächlich stattgefundene Ereignisse (Auslassung einiger

Fehlalarme) noch keine generellen Schlüsse erlauben, so war zu erkennen, dass das Multi-Modell-Ensemble SRNWP-PEPS eine höhere Zuverlässigkeit zeigte als COSMO-LEPS, wobei letzteres eine längere Vorlaufzeit bietet.

Es zeigte sich, dass das untere Quartil des Ensembles ein wichtiger Indikator für die Bewertung einzelner extremer Mitglieder sein kann. Ferner bestätigte sich, dass der Median oder ein bestimmtes Quantil der Vorhersage anstelle einer deterministischen Vorhersage für die Auslösung von Warnungen verwendet werden könnte. Aufgrund des immer noch geringen Datenumfangs konnten präskriptive Entscheidungsregeln zur Herausgabe von Hochwasserwarnungen noch nicht abgeleitet werden.

IV. Wallner, M.; Haberlandt, U.; Dietrich, J. (2012): Evaluation of different calibration strategies for large scale continuous hydrological modelling. *Advances in Geosciences* 31, 67–74.

Rolle: Mitautorschaft aus wissenschaftlicher Zusammenarbeit im Rahmen des Forschungsverbundes zur Klimafolgenforschung Niedersachsen KLIFF

Flussgebietsmodelle können für die Simulation der hydrologischen und wasserwirtschaftlichen Auswirkungen des Klimawandels angewendet werden. Für Untersuchungen zur Veränderung von Hochwasserabflüssen, wie sie beispielsweise für die Anpassung von Bemessungswerten an erwartete zukünftige Klimaverhältnisse vorgenommen wird, werden hydrologische Modelle in hoher zeitlicher Auflösung gerechnet. Zur Abbildung der Gebietseigenschaften werden zumeist räumlich verteilte Modelle verwendet, welche über eine hohe Zahl an Modelleinheiten verfügen. Die Kalibrierung von Flussgebietsmodellen erfordert daher intelligente Strategien, um eine große Zahl alternativer Parametersätze zu testen und zu optimieren sowie bei der Wahl der Parameter die physikalischen Eigenschaften des Flussgebietes zu berücksichtigen. Letzteres ist von Bedeutung für die Robustheit des Modells bei dessen Anwendung mit in der Kalibrierung nicht verwendeten Klimadaten.

In dieser Studie wurden vier unterschiedliche Kalibrierungsstrategien und drei Optimierungsalgorithmen angewendet und die Güte und Robustheit eines halb-verteilten hydrologischen Modells bewertet. Als Kalibrierungsstrategien wurden angewendet: a) Blockparameter, b) 1-Faktor, c) verteilte Parameter und d) Parameter-Regionalisierung. Als Optimierungsalgorithmen wurden PEST, DDS und SCE angewendet. Die Fallstudie wurde mit dem Modell HEC-HMS für fünf Teileinzugsgebiete des Aller-Leine-Flusseinzugsgebietes durchgeführt. Die Kalibrierung wurde nicht nur anhand der simulierten Zeitreihen ausgewertet, sondern auch räumlich durch Übertragung der kalibrierten Parameter auf unkalibrierte Flussgebiete und Bewertung der Güte des unkalibrierten Modells.

Für die einfachen Kalibrierungsstrategien „Blockparameter“ und „1-Faktor“ konnte der lokale Optimierer PEST gute Lösungen finden. Für die beiden komplexeren Strategien zeigten die globalen Optimierer DDS und SCE eine bessere Performanz mit Vorteilen für DDS. Die Kalibrierungsstrategie der Parameter-Regionalisierung zeigte zwar die geringste Güte bezüglich der kalibrierten Abflusszeitreihen in den drei Kalibrierungsgebieten, aber die beste Güte für die beiden Validierungsgebiete. Ferner zeigte die Parameter-Regionalisierung die größte Robustheit

in der räumlichen Validierung (die Güte für Kalibrierung und Validierung lagen am nächsten beieinander). Auch die Blockparameter zeigten überraschend gute Performanz verglichen mit den beiden anderen Kalibrierungsstrategien.

V. **Petry, U.; Dietrich, J.; Förster, K.; Wallner, M.; Berndt, C.; Meon, G.; Haberlandt, U. (2015): Ein Ansatz zur Validierung von Klimamodelldaten als Basis für die Interpretation von wasserwirtschaftlichen Klimafolgenabschätzungen in Niedersachsen. *Hydrologie und Wasserbewirtschaftung* 59 (4), 155–173.**

Rolle: Mitautorschaft aus wissenschaftlicher Zusammenarbeit im Rahmen des Forschungsverbundes zur Klimafolgenforschung Niedersachsen KLIFF

Zur Abschätzung der möglichen Folgen des Klimawandels werden in verschiedenen Fachbereichen wie auch in der Wasserwirtschaft Klimaprojektionen eingesetzt, um die prognostizierte Veränderung relevanter Größen direkt zu analysieren oder um Fachmodelle anzutreiben, mit welchen die jeweils interessierenden Größen simuliert werden, zum Beispiel Abfluss und Evapotranspiration in der Hydrologie. Dafür wesentliche Eingangsdaten sind Temperatur, Globalstrahlung und vor allem Niederschlag.

Es hat sich gezeigt, dass bestimmte Größen in Klimamodellen mit systematischen Fehlern (*bias*) behaftet sind. Dies trifft auch auf den für die Hydrologie wichtigen Niederschlag zu. Vor einer Verwendung der Klimaprojektionen für eine Klimafolgenabschätzung muss daher deren Kontrollauf des vergangenen Jahrhunderts auf die Übereinstimmung mit verfügbaren Beobachtungsdaten untersucht werden. In dieser Studie wurde darauf verzichtet eine Bias-Korrektur im Postprocessing vorzunehmen, da die Konsistenz der physikalischen Größen für die einzelnen Zeitschritte aufgehoben werden kann und die Fortschreibung des Bias auf die Zukunftsprojektionen lediglich eine Annahme darstellt.

In dieser Studie werden daher nur die Rohdaten validiert. Es wurden Simulationen der Temperatur und des Niederschlags eines statistischen regionalen Klimamodells (WETTREG2006) sowie von zwei dynamischen regionalen Klimamodellen (REMO und CLM), alle angetrieben durch das Globalmodell ECHAM5/MPI-OM, mit statistischen Methoden auf ihre Güte hin untersucht. Untersuchungsgebiet war das Flusseinzugsgebiet von Aller und Leine in Niedersachsen. Die Auswertungen erfolgten für das Gesamtgebiet von 15.733 km^2 und für neun Teilgebiete ab einer Größe von 129 km^2 anhand von interpolierten und auf Niederschlags-Messfehler korrigierten Beobachtungsdaten des Zeitraums 1961-2000.

Bei den dynamischen Modellen zeigten sich Abweichungen gegenüber dem Beobachtungsdatensatz insbesondere bei der Temperatur und den Trockenwetter-Indizes. WETTREG hingegen zeigte bei allen Größen erwartungsgemäß eine bessere Übereinstimmung. Die Verwendung der Klimamodelldaten in der hydrologischen Modellierung der Untersuchungsgebiete ergab für die simulierten Abflüsse im Mittel vergleichbare Modellgüten wie bei den korrespondierenden Kenngrößen des Niederschlags. Hohe wie niedrige Abflussverhältnisse wiesen jedoch eine vergrößerte Bandbreite der Güte auf. Mit Hilfe des Wilcoxon-Mann-Whitney-Tests konnte für ein Signifikanzniveau von 95% gezeigt werden, dass die betrachteten Modellketten geeignet erscheinen, um wasserwirtschaftliche

Klimafolgenabschätzungen vorzunehmen. Die Darstellung extremer Verhältnisse, vor allem bei Niedrigwasser, sollte jedoch stets in Abhängigkeit des Modells bewertet werden, um mögliche Entscheidungen über Anpassungsmaßnahmen abwägen zu können.

VI. **Maier, N.; Dietrich, J. (2016): Using SWAT for Strategic Planning of Basin Scale Irrigation Control Policies: a Case Study from a Humid Region in Northern Germany. *Water Resources Management* 30(9), 3285-3298, DOI: 10.1007/s11269-016-1348-0.**

Rolle: Mitautorschaft als Themensteller und Betreuer

Die Steuerung der Bewässerung in der Landwirtschaft wird üblicherweise auf der Feld- oder Betriebsskala betrachtet. Da die Bewässerung jedoch ein wesentlicher Wasserverbraucher ist, muss diese auch bei der Bewirtschaftung von Flussgebieten angemessen berücksichtigt werden. Damit kommt ihr bei der hydrologischen Simulation und der Bewirtschaftungsplanung von Flusseinzugsgebieten eine große Bedeutung zu. Insbesondere bei Klimafolgenabschätzungen spielt die Bewässerung eine große Rolle, da durch den Temperaturanstieg eine höhrere Verdunstung und damit auch ein höherer Bewässerungsbedarf erwartet werden.

Das ökohydrologische bzw. agrarhydrologische Modell SWAT wird weltweit für die Simulation kleiner bis sehr großer landwirtschaftlich genutzter Flussgebiete angewendet. Es kann das Pflanzenwachstum sowie die Landbewirtschaftung, z.B. die Bewässerung, in vereinfachter Form abbilden. Da auch in humiden Regionen auf Böden mit geringer Feldkapazität bewässert wird und auch hier der Klimawandel einen Anstieg des Bewässerungsbedarfs erwarten lässt, wird in dieser Studie untersucht, welche Auswirkungen eine wassersparende Bewässerung auf den Wasserverbrauch und den landwirtschaftlichen Ertrag im Flusseinzugsgebiet der unteren Fuhse in Niedersachsen haben könnte. Dazu wurden im Modell unterschiedliche Szenarien der automatisierten Bewässerungssteuerung nach Plflanzenwasserbedarf und nach Bodenfeuchte in der Wurzelzone entwickelt. Das kalibrierte SWAT-Modell zeigt plausible Änderungen der Bewässerungsmengen bei Änderung der Trigger für den Einsatz der Bewässerung, wobei keine Messwerte der realen örtlichen Bewässerungsmengen verfügbar sind.

Bei der Anwendung von Defizitstrategien bei der Bewässerung, d.h. bei verspäteter Wassergabe gegenüber dem üblichen Einsatz der Bewässerung bei 50 % der Feldkapazität, konnten die Wassermengen bei geringen Ertragsrückgängen deutlich reduziert werden. Die relativen Unterschiede der simulierten Bewässerungsmengen zwischen verschiedenen Feldfrüchten und Bodentypen sind überwiegend plausibel. Insgesamt überschätzte SWAT die Bewässerungsmengen systematisch und zeigte eine sehr hohe Variabilität zwischen den Jahren, wobei die Gesamtmengen als akzeptabel angesehen wurden. Die Steuerung der Bewässerung in SWAT soltle weiterentwickelt werden, um eine zuverlässigere Anwendung in der Flussgebietsmodellierung und Klimafolgenabschätzung für die Landwirtschaft zu ermöglichen.

VII. **Uniyal, B.; Dietrich, J.; Vasilakos, C.; Tzoraki, O. (2017): Evaluation of SWAT simulated soil moisture at catchment scale by field measurements and Landsat derived indices. *Agricultural Water Management* 193, 55-70.**

Rolle: Mitautorschaft als Themensteller und Betreuer

Die operationelle Messung der Bodenfeuchte ist im regionalen Maßstab mangels geeigneter automatisierter Messnetze nicht möglich. Daher erfolgt die Nutzung von Fernerkundungsdaten oder von Simulationen. Die Beobachtung der Bodenfeuchte über Fernerkundung ist eine indirekte Methode und aufgrund der groben räumlichen Auflösung nur mit Einschränkungen für hydrologische Studien und kaum für die Feldskala geeignet.

In dieser Studie erfolgte die Simulation der regionalen Verteilung der Bodenfeuchte im Flussgebiet der Ilmenau in der Lüneburger Heide im Tages-Zeitschritt mit dem ökohydrologischen Modell SWAT, welches für den Abfluss der Teileinzugsgebiete der Wipperau und Gerdau unter Berücksichtigung der Bewässerung mit befriedigender Güte kalibriert wurde. Für das Untersuchungsjahr 2016 wurde an sieben Tagen an zehn Stellen im Untersuchungsgebiet die Bodenfeuchte im ersten Horizont mit TDR und gravimentrisch bestimmt. Zur Schätzung der räumlichen Verteilung der Bodenfeuchte erfolgte die Entwicklung und Erprobung einer Methode zur Ableitung der Bodenfeuchte aus LANDSAT Satellitendaten mit 30 m horizontaler Auflösung. Dabei wurde der „Thermal Vegetation Difference Index“ TVDI aus dem „Normalized Difference Vegetation Index“ NDVI und der „Brightness Temperature“ abgeleitet. Unter Verwendung der selbst erhobenen Messdaten wurde eine Regressionsbeziehung zur Berechnung der Bodenfeuchte trainiert.

Für die sieben Feldkampagnen während der Bewässerungssaison 2016 wurde die von SWAT simulierte Bodenfeuchte mit den Ergebnissen der LANDSAT-Auswertungen verglichen. Es konnte eine gute Übereinstimmung der regionalen Verteilung und der saisonalen Dynamik der Bodenfeuchte erzielt werden, während es auf Ebene einzelner Modelleinheiten zu deutlichen Abweichungen kam. Dies kann teilweise damit erklärt werden, dass die jeweils in der Saison angebauten Feldfrüchte nicht räumlich exakt kartiert werden konnten. Die Bodentypen prägten Muster in der Verteilung der Bodenfeuchte aus, was auch Rückschlüsse auf die Parametrisierung von Böden entsprechend der Bodenkarte BÜK 200 erlaubte. Die simulierte und abgeleitete Bodenfeuchte stimmten während trockener Tage deutlich besser überein als während nasser Tage. Die mittlere absolute Abweichung zwischen der Bodenfeuchte lag für die meisten Bodentypen zwischen 0,9 und 10 % und nach Neuklassifizierung eines Bodentypen bei 1,06 bis 6 %.

Problematisch ist das Erfordernis klarer Sicht auf die Landoberfläche bei der Anwendung der Methode in Norddeutschland. Während im Untersuchungszeitraum 2016 beim Überflug von Landsat 7 und 8 alle acht Tage häufiger klare Sicht herrschte, wäre 2017 nur selten eine brauchbare Bildinformation verfügbar gewesen. Allerdings bestand aufgrund der feuchten Witterung 2017 auch kaum Bewässerungsbedarf. Dennoch ist die Methode daher vorzugsweise in Regionen mit häufig klarem Himmel anwendbar, besispielsweise in mediterranen und semi-ariden Regionen. Aufgrund der häufigen Ausfälle der Daten und der aufwändigen Nachbearbeitung der Stellitenbilder eignet sich die Methode zwar für die Validierung und Kalibrierung von Flussgebietsmodellen zusätzlich zu Messdaten, nicht jedoch für operationelle Zwecke zum Beispiel in der Bewässerungsberatung. Auch ein Einsatz in unbeobachteten Gebieten ist nicht möglich, da die Ableitung der Bodenfeuchte aus LANDSAT-Daten mit Feld-Messdaten kalibriert werden muss.

VIII. **Ojwang, R.; Dietrich, J.; Kasargodu Anebagilu, P.; Beyer, M.; Rottensteiner, F. (2017): Rooftop Rainwater Harvesting for Mombasa: Scenario Development with Image Classification and Water Resources Simulation.** ***Water 9(5), 359, doi:10.3390/w9050359*****.**

Rolle: Mitautorschaft aus wissenschaftlicher Zusammenarbeit und Erstbetreuung im Rahmen eines interdisziplinären Masterprojektes

Mombasa in Kenya leidet unter Wasserknappheit vor allem während der Trockenzeit. Die vorhandene Versorgungsstruktur kann die Nachfrage nicht decken. In der wasserwirtschaftlichen Entwicklungsplanung für Mombasa ist vorgesehen, in neuen Stadtvierteln die Nutzung von Regenwasser der Dachflächen in die Wasserversorgung einzubeziehen (*roof top rainwater harvesting*). In dieser Studie erfolge eine Entwicklung langfristiger Szenarien der Wasserverfügbarkeit und des Versorgungsgrades unter Berücksichtigung des Bevölkerungswachstums und des Klimawandels bis 2035. Hierfür wurde das Wasserbewirtschaftungsmodell WEAP (Water Evaluation and Planning System) um eine Gleichung zur Berücksichtigung der Wassserspende der Dächer erweitert. Für die Bestimmung der Dachflächen wurde auf die Klassifizierung von Luftbildern zurückgegriffen. WEAP simuliert den Deckungsgrad des Wasserbedarfs durch das Wasser aus allen verfügbaren Quellen.

Es wurden vier verschiedene Bewirtschaftungsstrategien kombiniert mit Szenarien über das prognostizierte Bevölkerungswachstum, Annahmen über steigenden Wohlstand sowie Klimaszenarien.

WEAP zeigte sich gut geeignet zur Berechnung der Szenarien. Die Klassifizierung von Luftbildern lieferte für ein vom Boden ausgewertetes Testgebiet akzeptable Ergebnisse über die verfügbare Dachfläche. Je nach Ausdehnung der Dachflächen in den verschiedenen Szenarien können 2,3 bis 23 Millionen m^3 Wasser gewonnen werden. Allerdings reicht selbst im besten Falle die Einbeziehung der Dachflächen nicht alleine dazu aus, den künftigen Wasserbedarf von Mombasa zu erfüllen. Ferner ist zu bedenken und Gegenstand künftiger Untersuchungen, dass das aufgefangene Wasser teilweise der Grundwasserneubildung verlorengeht. In Anbetracht der Küstenlage Mombasas könnte eine Absenkung des Grundwasserspiegels die Intrusion von salzhaltigem Wasser in die küstennahen Aquifere begünstigen.

IX. **Nguyen, V. T.; Dietrich, J. (2018): Modification of the SWAT Model to Simulate Regional Groundwater Flow Using A Multi-Cell Aquifer. Hydrological Processes 32(7), 939-953, DOI: 10.1002/hyp.11466.**

Rolle: Mitautorschaft als Themensteller und Betreuer

In der Flussgebietsmodellierung werden häufig verteilte Modelle wie SWAT angewendet, welche aufgrund der starken Vereinfachung des Grundwasserleiters keine lateralen Grundwasserflüsse simulieren können und auch keine Grundwasserstände und deren Schwankungen im Jahresverlauf abbilden können. Dadurch sind regionale Grundwasserbewegungen, zum Beispiel von den Höhenzügen in die Flusstäler, sowie auch der nur aus einem einfachen Linearspeicher kommende Basisabfluss nicht befriedigend darstellbar.

In dieser Studie wurde ein neues „Multi-Cell Aquifer“-Modul (SWAT-MCA) zur Simulation des Grundwassers in SWAT entwickelt. Der lokale Aquifer entspricht dem „shallow aquifer“ im originalen SWAT, während der regionale Aquifer den tiefen Grundwasserleiter „deep aquifer“ von SWAT ersetzt und in den hydrologischen Kreislauf einbezogen wird. Der regionale Aquifer wird durch Thiessen-Polygone um die verfügbaren Beobachtungsstellen räumlich eingeteilt. Die Nachbarschaftsbeziehungen werden als Topologie angelegt. Die Grundwasserneubildung wird in beide Aquifere aufgeteilt. Die Grundwasserströmung im regionalen Aquifer wird entsprechend den hydraulischen Gradienten zu den Nachbarzellen mit einer an das Gesetz von Darcy angelehnten Gleichung simuliert. Der Basisabfluss wird als nicht-lineare Speicher-Ausfluss-Beziehung aus beiden Aquiferen simuliert.

Die Anwendung des Modells konnte für zwei Teileinzugsgebiete der Wipperau und Neetze im Flussgebiet der Ilmenau demonstriert werden. Die Simulation des Niedrigwasserabflusses verbesserte sich im Gebiet der Wipperau deutlich. In beiden Gebieten konnten saisonale Schwankungen der Grundwasserstände unter Berücksichtigung des Maßstabes und der geringen Rechenanforderungen gut abgebildet werden. Damit stellt das SWAT-MCA-Modell einen guten Kompromiss aus dem stark vereinfachten, nicht horizontal vernetzten Grundwassermodul im originalen SWAT und einer deutlich komplexeren Kopplung mit einem speziellen Grundwassermodell wie MODFLOW dar. Eine Verallgemeinerung dieses für Porenaquifere entwickelten Modells wird daraus noch nicht propagiert, da weitere Fallstudien in anders charakterisierten Gebieten durchgeführt werden sollten.

3.2 Fachartikel

I. Combination of different types of ensembles for the adaptive simulation of probabilistic flood forecasts: hindcasts for the Mulde 2002 extreme event

Dietrich, J.[1], Trepte, S.[2], Wang, Y.[1], Schumann, A. H.[1], Voß, F.[1,*], Hesser, F. B.[1,**], Denhard, M.[2]

(1) Institute of Hydrology, Water Resources Management and Environmental Engineering, Ruhr University Bochum, Germany
(2) Deutscher Wetterdienst DWD (German National Weather Service), Offenbach, Germany
(*) now at: Center for Environmental Systems Research, University of Kassel, Germany
(**) now at: Bundesanstalt für Wasserbau (Federal Waterways Engineering and Research Institute, Hamburg, Germany

Abstract

Flood forecasts are essential to issue reliable flood warnings and to initiate flood control measures on time. The accuracy and the lead time of the predictions for head waters primarily depend on the meteorological forecasts. Ensemble forecasts are a means of framing the uncertainty of the potential future development of the hydro-meteorological situation.

This contribution presents a flood management strategy based on probabilistic hydrological forecasts driven by operational meteorological ensemble prediction systems. The meteorological ensemble forecasts are transformed into discharge ensemble forecasts by a rainfall-runoff model. Exceedance probabilities for critical discharge values and probabilistic maps of inundation areas can be computed and presented to decision makers. These results can support decision makers in issuing flood alerts. The flood management system integrates ensemble forecasts with different spatial resolution and different lead times. The hydrological models are controlled in an adaptive way, mainly depending on the lead time of the forecast, the expected magnitude of the flood event and the availability of measured data.

The aforementioned flood forecast techniques have been applied to a case study. The Mulde River Basin (South-Eastern Germany, Czech Republic) has often been affected by severe flood events including local flash floods. Hindcasts for the large scale extreme flood in August 2002 have been computed using meteorological predictions from both the COSMO-LEPS ensemble prediction system and the deterministic COSMO-DE local model. The temporal evolution of a) the meteorological forecast uncertainty and b) the probability of exceeding flood alert levels is discussed. Results from the hindcast simulations demonstrate, that the systems would have predicted a high probability of an extreme flood event, if they would already have been operational in 2002. COSMO-LEPS showed a reasonably good performance within a lead time of 2 to 3 days. Some of the deterministic very short-range forecast initializations were able to predict the dynamics of the event, but others underpredicted rainfall. Thus a lagged average ensemble approach is suggested. The findings from the case study support the often proposed added value of ensemble forecasts and their probabilistic evaluation for flood management decisions.

1 Introduction

Uncertainties in flood forecasting mainly result from incomplete knowledge of the further meteorological development and from uncertainties of hydrological and hydraulic modelling. Known and knowable sources of uncertainty are the availability and quality of input data, respectively initial and boundary conditions for the models, as well as model parameters and model structure. Inaccurate human interaction and technical problems may also affect the output of a flood prediction chain. The highly nonlinear behaviour of the atmospheric system and the land-atmosphere interaction adds unknowable sources of uncertainty. Thus a perfect weather forecast is impossible (Lorenz, 1963). Resulting from these uncertainties it is not possible to issue a perfect flood forecast.

During the last decades modelling and forecasting techniques evolved from a deterministic towards a probabilistic paradigm. Uncertainty estimation in forecasting aims at framing the possible future development, admitting and communicating the imperfection of the forecast. Exceedance probabilities for threshold values (e. g. critical discharge levels causing inundation) can be provided to flood managers and decision makers.

Ensemble techniques are suitable methods for producing probabilistic forecasts (Anderson, 1996; Kalnay, 2002; Toth et al., 2003). In the context of flood management, ensembles are a group of alternative scenarios of possible future development of the hydro-meteorological situation. Different types of ensembles can be classified according to the generating mechanisms (for meteorological as well as for hydrological applications):

- single system ensembles: perturbation of initial and boundary conditions, different convection schemes (physically based ensembles), perturbation of model parameters;
- multiple systems or multi-model ensembles ("poor man ensembles"): combination of simulations from different models (e. g. Georgakakos et al, 2004);
- lagged average ensembles: combination of current forecasts with forecasts from earlier model runs (Hoffman & Kalnay, 1983).

Meteorological ensemble prediction systems (EPS) have become operational. Buizza et al. (2005) compare three global EPS. The development of hydrological applications of ensemble forecasts has started in the late 1990-ies and is subject of ongoing research (e. g. de Roo et al., 2003; Mullusky et al., 2004;Gouweleeuw et al., 2005; Verbunt et al., 2006; Reed et al., 2007).

Beven and Binley (1992) published seminal work about uncertainty estimation in hydrological modelling. Most of the techniques applied in hydrologic modelling need hundreds or even thousands of model runs and often produce a lot of information, which is not useful for decision oriented operational flood management. Ensemble techniques aiming at the representation of hydrological uncertainty with only a few (in the dimension of 10) members still have to be developed. Methods of combining the simulations of multiple models are subject of current research (e. g. Bayesian Model Averaging, Hoeting et al., 1999; Raftery et. al., 2005; Sloughter et al., 2006). Recently Ajami et al. (2007) presented a study about the computation and combination of hydrological multi-model ensembles to account for model structure uncertainty. They list relevant sources in this field. Interdisciplinary studies dealing with the probabilistic assessment of the flood forecast chain have been published e. g. by Krzysztofowicz (2002), Apel et al. (2004) and Pappenberger et al. (2005).

The participatory HEPEX project (Hydrological Ensemble Prediction Experiment, Schaake et al., 2007) integrates meteorologists, hydrologists and users in order to promote the development of ensemble streamflow forecast systems. In Europe, the probabilistic Flood Alert System (EFAS) is under development (Thielen et al., 2008). EFAS aims to provide flood information for the medium to long-range at large scale river basins being relevant for decisions at national or EU level.

Probabilistic forecast systems require the computation of a large number of model runs within a time frame of a few hours. Here computational resources still constrain the possibilities of probabilistic forecasts in an operational real-time environment. However the application of numerical models on personal computers can profit from recent developments in parallelization. Flood forecasts for large river basins can now be simulated on workstations, which can be operated by local water management authorities.

The aim of this paper is to present an ensemble based operational flood management strategy, which integrates methodological developments in ensemble techniques. A prototype of a corresponding flood management system demonstrates the adaptive coupling and control of forecast models. In the subsequent sections of this paper we summarize the main characteristics of the operational meteorological ensembles from different sources, which are used as input for a rainfall-runoff-model. Whereas the conceptualization of the presented flood management strategy covers the complete flood management chain from rainfall to risk, our case study presented in this paper deals with the probabilistic forecast of the rainfall-runoff process in head waters. As an example we present and discuss results of ensemble hindcasts for the 2002 extreme flood event in the Mulde river basin, which is one of the fastest reacting river systems in the Central European low mountain range.

2 Conceptualization of an ensemble-based operational flood forecast

A flood forecast system for operational application is a complex set of coupled models, sensors and databases. When designing such a system, a compromise between computational efficiency, availability of data, predictive capability of the models and the cognitive burden for the flood manager has to be negotiated. On the one hand, data flow and control activities must be automated to the highest achievable level. On the other hand, the complex nature of the problem requires options for flood managers and decision makers to take control over the simulation process, e. g. when sources of information are identified as unreliable or when parts of the model chain fail.

We have designed and partially tested an ensemble based strategy for medium to very short-range flood forecasts. In Figs. 1 and 2, we give an overview of the pathways to combine the components in an adaptive manner in order to manage uncertainty. Two main perspectives are adopted: on Fig. 1 in view of the components of the meteorological forecast and on Fig. 2 in view of the main activities in hydrological modelling and flood management.

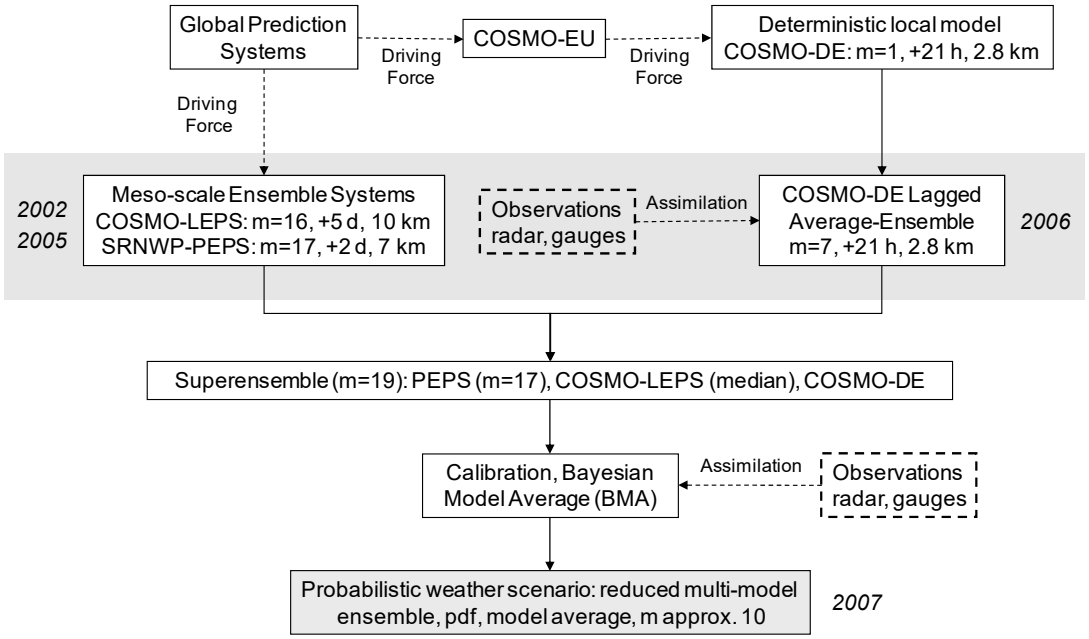


Fig. 1. Meteorological ensembles from different sources used for an ensemble-based operational flood risk management strategy. The number of ensemble members is denoted by m, the other values indicate the lead time and the horizontal spatial resolution of the ensemble prediction systems.

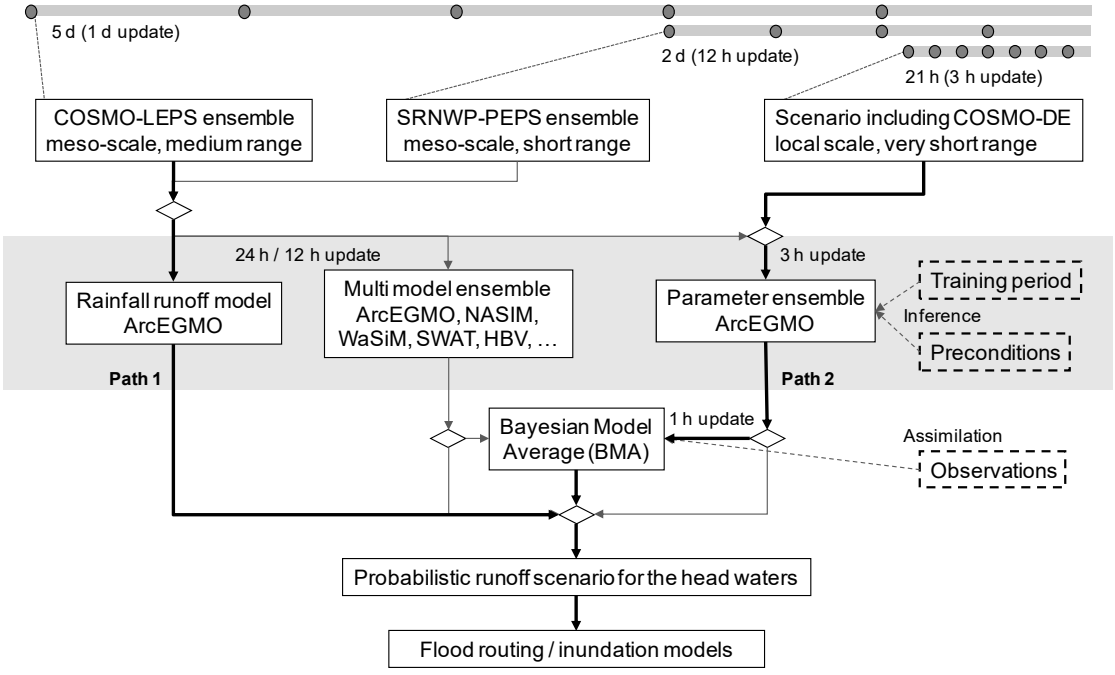


Fig. 2. Simulation of hydrological ensembles within an adaptive flood management strategy. Model averaging and ensemble updating are optional branches in the flow of actions.

The proposed strategy combines meteorological medium-range forecasts (3 to 5 days lead time), short-range forecasts (1 to 2 days lead time) and very short-range forecasts (< 1 d lead time) from different operational prediction systems. From these systems a superensemble for the very short-range can be generated and calibrated with the Bayesian Model Average method to produce a high resolution probabilistic weather scenario (Fig. 1, detailed description of the prediction systems in Section 3). Hydrological and hydraulic models are driven by meteorological input scenarios provided by the different forecast models. Rainfall-runoff models can simulate

ensemble forecasts of river flow at several points of interest like gauges and vulnerable sites. We decided to choose a conceptual hydrological model as the default component for transformation of climatic input into runoff. Conceptual rainfall-runoff models describe the complex natural hydrological processes in a simplified manner. These models are widely used for the meso- and macro-scale due to their significant advantages compared to physical models regarding parameter estimation and computation time (Carpenter and Georgakakos, 2006; Smith et al., 2004; Ajami et al., 2004).

The adaptive ensemble simulation strategy shown in Fig. 2 has two default pathways, which can be followed for the entire catchment depending on the hydro-meteorological situation and the needs of flood managers. Path 1, shown on the left, is operated at lead times of 3 to 5 days. The lead time could be extended to 10 days using predictions from global systems, but the spatial resolution of these systems restricts their applicability in meso-scale flood forecasting. Medium-range forecasts provide the basis for decisions about reservoir management and early warning previous to a potential extreme flood event. Here one hydrological model (ArcEGMO in the case study) is computed with a default parameter set, which proved to be efficient for historic flood events. Hydrological uncertainty is considered low compared to meteorological uncertainty. The latter is mainly represented by the spread of the ensemble. In case of a predicted extreme event, hydrological uncertainty becomes more crucial. Due to limited predictive capabilities of the conceptual model in extrapolation, a second branch of the workflow offers a multi-model computation for the head waters to compute an ensemble of runoff generation. A multi-model ensemble is computationally and cognitive very demanding. Thus it can only be implemented for a few sub-catchments in practice. For the two Mulde sub-catchments "Upper Zwickauer Mulde/Eibenstock reservoir" and "Schwarze Pockau" models from different research and consulting activities are available. These models range from more physically based (WaSiM-ETH), conceptual with strong physical meaning (SWAT, ArcEGMO) to more conceptual models (NASIM, HBV).

Short to very short-range forecasts (path 2 in Fig. 2) are used for issuing flood alerts and planning of tailored flood defence measures. Here the attention of flood managers is focused on a variety of aspects and the computational and cognitive resources are limited. Nevertheless decision makers need more detailed information about relevant criteria like peak time, peak discharge and possible inundation areas. Here only one conceptual model is used and updated in 1 - 3 hourly intervals including assimilation of observed rainfall and discharge. If we assume that the chosen model structure has sufficient predictive capabilities and input uncertainty is expressed by the meteorological ensemble, two additional sources of hydrological uncertainty are regarded: the initial state of the model and the selection of model parameters.

Uncertainty in initial conditions (e.g. soil moisture) is reduced by an iterative trial and error system state update procedure, which adjusts the storage representing soil moisture and upper groundwater until the hydrograph of the seven days preceding the expected rainfall is best represented by the model. This initial state update violates continuity, but it is important to tackle uncertainty from prior simulation of the continuous model without performing a recalibration during real-time application of the system. The updated state will only be kept for the event simulation to reduce error in flood forecast. After the event took place, the calibration of the continuous model is revised. The Ensemble Kalman filter (Evensen, 1994) will be implemented for automation of state updating.

For framing parameter uncertainty of the hydrological model we propose an ensemble approach. Hydrological parameter ensembles are generated by combination of model parameter sets, which proved to be efficient for simulating flood events in the calibration and test periods. We classified the historic flood events into four types of hydrological response, mainly depending on maximum intensity and total amount of rainfall as well as snow cover. A parameter ensemble combines up to 20 simulations using different parameter sets, which are chosen from the best performing sets for each group obtained by a priori Monte-Carlo simulations. These ensemble members are weighted according to the a priori expected type of event. If no decision about the type of event is possible, these parameter sets are equally weighted. The combination of the different types of ensembles adds up to a discharge ensemble with about 100 members.

A Bayesian inference mechanism adjusts the weights when new data become available during the event. The Bayesian updating procedures can, but not necessarily must reduce the number of members. We aim at successively reducing the number of parameter ensemble members for a proper relation between meteorological and hydrological contributions to total (knowable) uncertainty. Ideally, this procedure sequentially reduces uncertainty by using new information when available. Note that we do not recalibrate the parameters online.

Each of the two pathways can be simulated in two different modes: hindcast and forecast. Path 1 has been used for hindcast simulations as shown in the case study (Sections 5 and 6), when the hydrological model can be calibrated on observations. The hindcast mode is also used for test and further improvement of the system including an update of the knowledge base of efficient parameter sets.

3 Meteorological ensemble prediction systems for Central Europe

Our flood management system aims to combine early warnings with dynamical refinement strategies of precipitation forecasts in the short and very short-range. For medium-range forecasting (3 to 5 days) and early warnings the COSMO-LEPS is used, which performs a dynamical downscaling of the global EPS operated by ECMWF (European Center for Medium-Range Weather Forecasting). The short-range SRNWP-PEPS combines 23 deterministic forecasts from 21 national meteorological services of which 17 cover the Mulde catchment in a poor man's EPS. It concentrates the most sophisticated knowledge in limited area numerical modelling on the meso-scale and has a lead time of 48 h. For the incorporation of forecast refinements with very high horizontal resolution and improved quantitative precipitation forecasts we use the COSMO-DE model of the National German Weather Service (DWD). This model is among the new generation of convection resolving models and is running operationally every three hours with 2.8 km horizontal resolution.

3.1 COSMO-LEPS

The COSMO-LEPS is a limited area EPS, developed within the COSMO (Consortium for Small-scale Modelling) to improve the predictability of extreme weather events, especially when orographic and meso-scale-related processes play a crucial role. COSMO is a consortium involving Germany, Italy, Switzerland, Greece and Poland which aims to develop, improve and maintain the non-hydrostatic limited-area COSMO model. The LEPS methodology (Limited area Ensemble Prediction System) combines the members from the 51-member ECMWF global EPS into a small

number of clusters. A representative member of each cluster provides initial and boundary conditions to run the COSMO model (Molteni et al., 2001; Montani et al., 2003). The added value of the system resides in joining the skill of a global ensemble system to depict the possible evolution scenarios with the capability of a limited area model to improve the descriptions of local meteorological processes. The current COSMO-LEPS suite uses 16 representative members driving 16 COSMO model integrations which generate probabilistic output with lead time 12 UTC +132 hours. The horizontal resolution is approximately 10 km. Marsigli et al. (2005) show that the COSMO-LEPS system significantly enhances the skill of the ECMWF-EPS in predicting local amounts of precipitation.

3.2 SRNWP-PEPS

The EUMETNET Short-Range Numerical Weather Prediction Programme (SRNWP) incorporates the four modelling consortia HIRLAM, ALADIN, COSMO and the UK Met Office. The weather services associated to these consortia produce a reasonable variety of operational forecasts on different domains with different grid resolutions using different model parameterisations or releases and data assimilation techniques. Since 2005 the DWD has been joining all available high resolution numerical forecasts in a poor man's EPS (PEPS, Denhard & Trepte, 2006). This multi-model ensemble currently generates quasi-operational probabilistic forecasts for Europe using 23 different deterministic forecasts provided by 21 meteorological services (http://www.dwd.de/PEPS). The single model forecasts are interpreted on a horizontal reference grid with a spacing of 0.0625° (~ 7 km) like the COSMO-EU model of DWD. Since the individual members have different spatial resolutions and integration areas, the ensemble size depends on location. For selected meteorological parameters ensemble mean and exceedance probabilities of certain thresholds are calculated at each PEPS grid point from the ensemble members using a nearest neighbour approach. All ensemble members are equally weighted and the probability is calculated by counting the corresponding members at each grid point. The results of the SRNWP-PEPS are distributed to the contributing members on an operational basis four times a day. It can be seen from case studies and probabilistic verification for Germany (Trepte et al. 2006) that this ensemble is a valuable tool for severe weather forecasting. Of course its performance is limited by the contributing models. If these are not capable to forecast a given weather pattern, the SRNWP-PEPS will not be capable as well. A major benefit of this multi-model EPS is the possibility to compare the behaviour of all operational European limited area models.

3.3 COSMO-DE

Since Spring 2007, the newly developed numerical weather prediction system COSMO-DE (prior name LMK) for very short-range forecasts up to 21 hours and with a resolution on the meso-γ scale is added to the operational meteorological model chain at the DWD (Doms et al., 2004; Steppeler et al., 2003). The emphasis of this model system lies in the prediction of severe weather events related to deep moist convection and to interactions of the flow with small scale topography. COSMO-DE couples to the models COSMO-EU (meso-β) and GME (global). One of the most important changes from meso-β to meso-γ resolution is the abandoning of a parameterisation of deep convection. Such a model needs special requirements concerning data assimilation: at this scale highly resolved, rapidly updated observations are needed, which are delivered by precipitation radar data with a horizontal resolution of roughly 1 km. They are assimilated by the latent heat nudging approach (Klink & Stephan, 2005). The assimilation cycle

leads to new forecasts every 3 hours. There is not much experience with the skill of COSMO-DE in forecasting extreme precipitation events on the convection permitting scale. A hindcast of the 2002 event is presented in section 6.2. Under the assumption that the synoptic scale weather patterns, which force the COSMO-DE at the boundaries, are quite stable with respect to this high frequency of new initialisations, it is reasonable to join the time lagged deterministic runs of COSMO-DE in a lagged average forecast ensemble. Recently, DWD started the development of a physical single model ensemble based on COSMO-DE, the COSMO-DE-EPS, which will be operational in 2011.

4 The rainfall-runoff model ArcEGMO

For the Mulde case study presented in this paper, the hydrological model ArcEGMO (Becker et al., 2002) was adapted to the needs of operational flood forecasting. ArcEGMO is a GIS-based modular modelling system containing several sub-models. It can be characterized as a conceptual model, whose parameters have a physical meaning.

The catchment is partitioned into homogenous units (hydrotopes), for which the vertical and horizontal processes are simulated. For a consideration of the spatial variability of land use and soil related properties, areal distribution functions for essential parameters like saturated conductivity, field capacity, etc. are applied. Runoff generation is calculated with an empirical formulation for the infiltration capacity (Holtan, 1961) based on the saturated conductivity for the upper most soil layer and the relative soil moisture deficit. Parameterization is therefore directly related to mapped soil characteristics and related attributes. The unsaturated zone in the soil is basically described by two virtual moisture layers which are filled by infiltration and depleted by evapotranspiration. The total thickness of the unsaturated zone is defined as the minimum either of the root depth, the groundwater depth or the soil thickness. Runoff concentration at the surface is modelled with a kinematic wave approach. For the description of the lateral subsurface processes the hydrotopes are reaggregated to different types of hydrological behaviour: e. g. uplands with a deep aquifer supplying mainly stream base flow, hillslopes with dominating lateral interflow and shallow groundwater areas with a more complex groundwater-surface water interaction (e. g. riparian zones, wetlands). Subsurface flow is modelled with a system of linear reservoirs for each of the groups of hydrotopes. A second fast reservoir can be activated when the water level in the main reservoir exceeds a threshold value (Becker et al., 2002). This reservoir system allows the spatially distributed definition of storage constants for slow and fast runoff components, which can be calibrated from measured discharge records via recession curve analysis. Subsurface flow builds up about 90 % of the hydrograph, mainly depending on the preconditions of the river basin and the characteristics of the rainfall event. A sensitivity analysis (Wang et al., 2007) revealed that the model parameters for the fast reservoirs are most relevant for the calibration of runoff in the headwaters during the summer season. These parameters can be used as lumped or distributed parameters. Channel routing can be simulated with the Kalinin-Miljukov method or a linear reservoir cascade.

The system states of a continuously operated model with daily time steps can be used as initial conditions for a flood event model operating on sub-daily (commonly hourly) time steps. This event model can be rerun from that state with several parameter sets and different meteorological input (e. g. from ensemble forecasts). So the computational demand is reduced to one single model

initialization run and a large number of semi-automated executions of the process computation for a limited number of time-steps. The model can be continuously run in sub-daily resolution within the operational flood forecast system. For the hindcast simulations of historic events, which provide essential knowledge about model parameterization in case of extreme rainfall, only a limited amount of precipitation data is available in sub-daily resolution. Here a switching between different temporal resolutions is advantageous. This strategy allows an efficient integration of ArcEGMO into the computation and optimization of a probabilistic flood forecast chain.

5 Study region

The upper Mulde catchment is situated in the Ore Mountains (Germany and Czech Republic, 7400 km^2 total catchment area). It is formed by several parallel sub-basins, draining from South to North. Narrow and steep valleys cause a fast reaction of the watershed and critical superimposition of flood waves. Several cities are located in the flood plain of the lower Mulde river basin (Fig. 3). During west-cyclonic rainfall events, which caused several extreme flood events in the past, the uncertainty of precipitation forecasts in location, time and volume is crucial. Thus the reliability of flood alerts is an issue of concern.

In the last 100 years six local short-time extreme events with catastrophic effects (sub-catchments up to 200 km^2) and several large scale flood events have been observed in the Mulde river basin. The most recent local extreme event was the Marienberg flood 1999 caused by 144.6 mm rainfall in 90 minutes over a small sub-catchment (Büttner et al., 2001). The recurrence period of the recorded rainfall intensity is considerably higher than 100 years (63.4 mm for 90-minute rainfall at Marienberg according to Bartels et al. 2005). After two extreme floods in 1954 and 1974, the Mulde river basin was again affected by an extreme flood event in 2002 (Elbe flood). The flood was caused by a so called “Vb” atmospheric circulation pattern, where a low-pressure system moves from the Atlantic southeast to the Mediterranean and then turns northeast across the Alps toward Central Europe. There it causes heavy and intense rainfall, especially in the Central European mountain ranges (Becker & Grünewald, 2003). In Zinnwald-Georgenfeld, close to the upper Freiberger Mulde (Fig. 3), the largest 24-hour rainfall ever recorded in Germany was 312 mm on August 12, 2002.

At the occurrence of the 2002 extreme flood event, the ensemble prediction systems used in this case study have not yet been operational. Hindcasts have been computed by rerun of the COSMO-LEPS and the COSMO-DE models. Within the next section we present first results.

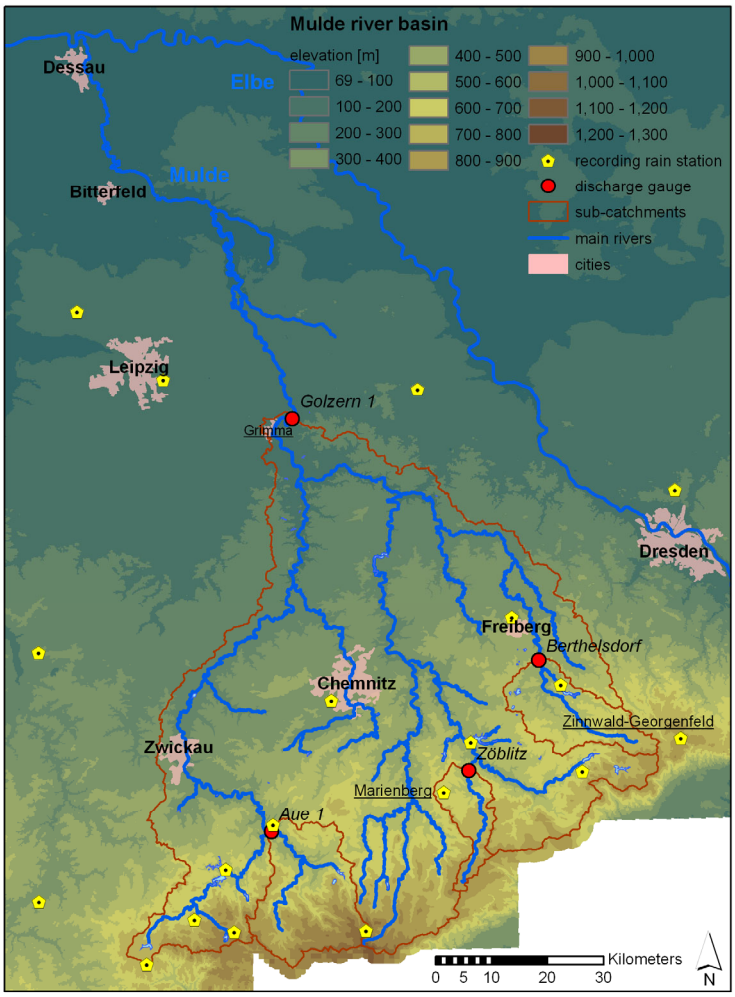


Fig. 3. Topographic map of the Mulde river basin showing the sub-catchments and points of interest mentioned in the text.

6 Results

6.1 Rainfall-runoff model

For the setup of the Mulde catchment model, a digital elevation model, soil data, land use data and discharge time series were provided by local authorities. Climate time series with daily time steps were used to calibrate and test a continuous model for the period 1950-2006. A single best parameter set has been obtained by calibration with iterations between manual adjustment of parameters and Monte Carlo simulations. We paid specific attention to low flow periods, peak flows and recession curves to minimize errors in initial conditions and to provide a reliable first guess for default parameters. System states are saved by the long-range mode model and loaded as starting values by the short-range mode model. Six summer flood events were calibrated on observed rainfall from 10 to 19 recording climate stations (Fig. 3) in high temporal resolution (hourly time step), namely 2002 (recurrence period T up to > 500 a), 1995 (T approx. 10 a) and four flood events with T < 5 a (1995 – 1998, Wang et al., 2007). Event specific parameter sets and a common “best” compromise parameter set for all calibrated events have been obtained by

iterations between expert and Monte Carlo simulations. A test of the event model has been performed on four flood events in the 1950-ies and 1980-ies.

In this paper we focus on processing of the meteorological ensembles by the hydrological model, but not on hydrological uncertainty. We use the calibration for the 2002 flood event to simulate ensembles of discharge driven by meteorological ensembles. Figure 4 shows the calibration results for three head waters of different hydrological characteristics for the 2002 flood event. Nash-Sutcliffe model efficiency is 0.80 for Aue 1 gauge (Schwarzwasser, 362 km^2), 0.93 for Zöblitz gauge (Schwarze Pockau, 129 km^2), and 0.95 for Berthelsdorf gauge (Freiberger Mulde, 243 km^2). Deviation of peak discharge is 10 % for Aue 1 and below 5 % for the other two gauges.

The computational efficiency of the adaptive modelling strategy outlined in Sect. 2 could be proven. For the 5442 km^2 catchment area of the Golzern gauge (Fig. 3) and a flood event with 220 hourly time-steps the computation time was only 14 seconds on a standard personal computer with a 3 GHz Pentium processor.

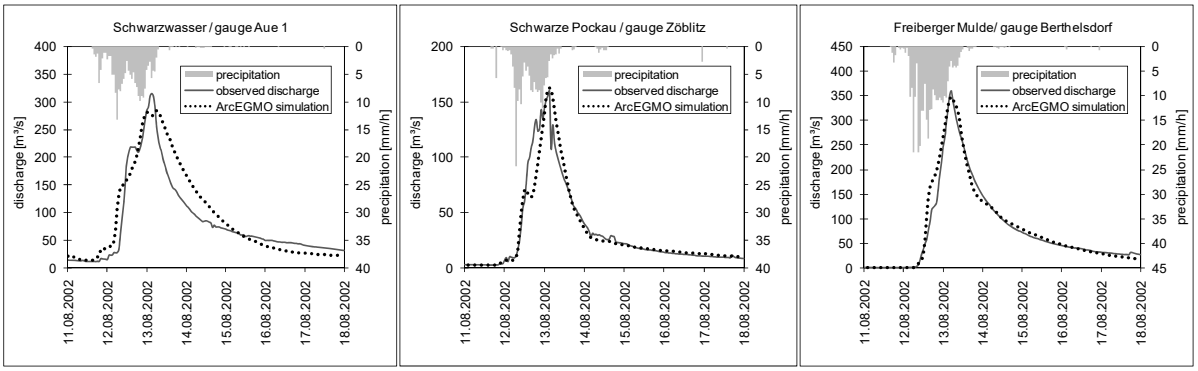


Fig. 4. Calibration results for the rainfall-runoff model for sub-catchments located in the western, middle and eastern part of the Ore Mountains (model time step 1 h).

6.2 Stream flow ensemble hindcasts of the 2002 extreme flood in the Mulde catchment

Within the framework of the European Project PREVIEW (http://www.preview–risk.com), a rerun of the period between 20/07/ and 31/08/2002 of ECMWF EPS and COSMO-LEPS (10 members) was performed. COSMO-LEPS hindcasts with daily initialization between 07/08/2002 and 12/08/2002 have been used to drive a single ArcEGMO model according to path 1 in Fig. 2 (Sect. 2). The results of the hydrological simulations are ensembles of possible future development of discharge at several gauges within the river basin. Fig. 5 shows a sequence of discharge ensembles for the Zöblitz gauge. The hydrographs integrate uncertainty from the meteorological forecast as well as from the hydrological model (cf. Fig. 4). The latter are not explicitly analyzed for this case study, but have been reduced by calibrating the model on observed discharge (which is of course only possible when simulating hindcasts). Between 08/08/ and 10/08/ the median and the spread of the ensemble (shown as interquartile range IQR = 0.75-quantile – 0.25-quantile) enlarged. The forecast initialized at 11/08 indicated that the situation was less tense again. The outlier from the 10/08/ initialization was closest to the discharge peak level observed 2.5 days later. For the Schwarze Pockau catchment (Zöblitz gauge) the event was generally underpredicted (Fig. 5). Simulations for the Aue 1 gauge (left in Fig. 6) in the Western Ore Mountains show a “good” ensemble embracing the observed runoff, which was

close to a 100 year recurrence level. The observed peak level is near the upper limit of the IQR confidence band and relatively close to the median of the ensemble members, but the spread is higher. For the Freiberger Mulde (Berthelsdorf gauge, right in Fig. 6) the flood event was underpredicted. Table 1 lists the probabilities of exceeding flood alert levels at the Berthelsdorf gauge. All COSMO-LEPS members have been equally weighted. The significant increase in probability of exceeding levels 3 and 4 between 09/08/ and 10/08/ is remarkable, as well as the decrease in 11/08/. Even though the forecasts underpredicted peak flow, a probability of 90 % for significant inundation (alert level 3) and still 60 % for severe damage and danger of life (alert level 4) would have shown a clear evidence of a large flood event two and a half days in advance.

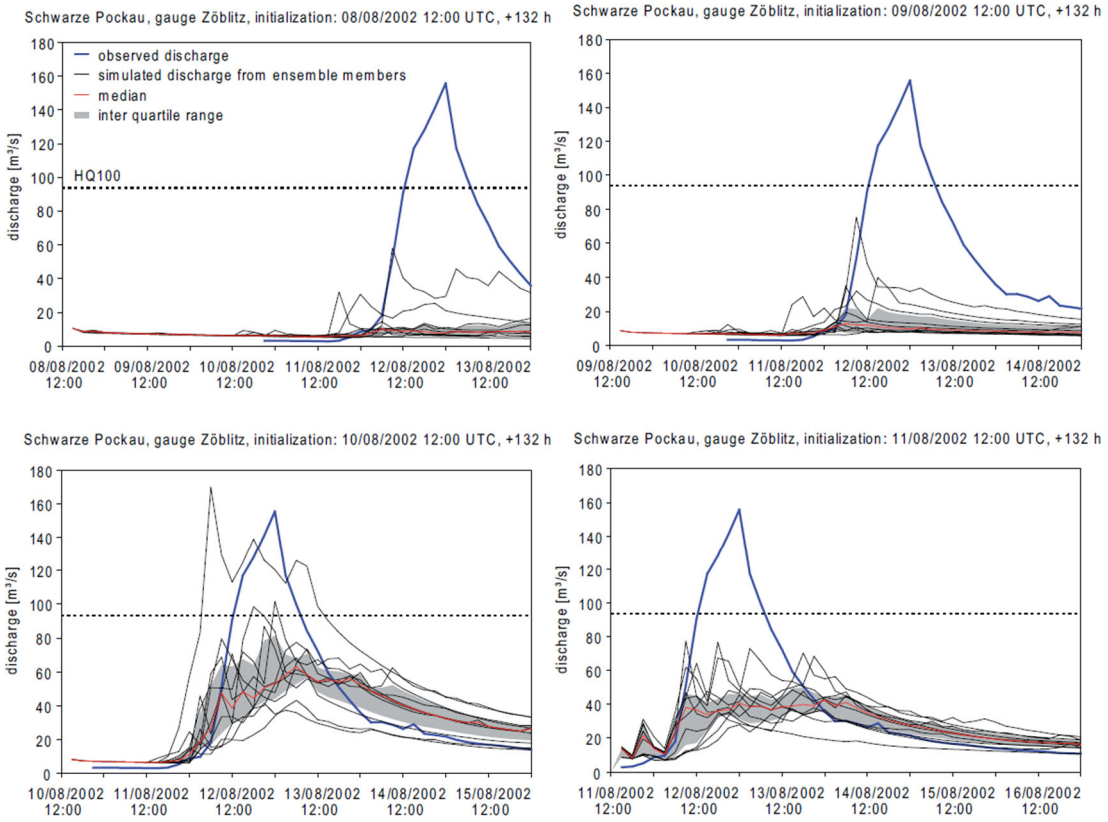


Fig. 5. Sequence of discharge forecasts before the disastrous 2002 flood for the Schwarze Pockau catchment (gauge Zöblitz). COSMO-LEPS hindcasts were initialized at 08/08, 09/08, 10/08 and 11/08/2002 at 12:00 UTC and processed by the hydrological model ArcEGMO (model time step 3 h).

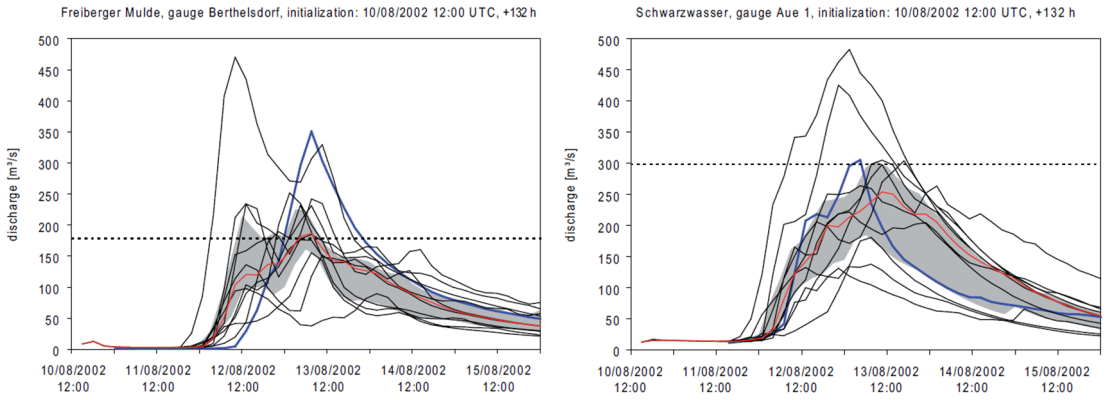


Fig. 6. Discharge ensembles for different sub-catchments initialized from COSMO-LEPS hindcasts on 10/08/2002 12:00 UTC with two and a half days lead time to the peak discharge.

Table 1: Temporal evolution of forecasted exceedance probabilities for alert discharge levels at Berthelsdorf gauge (Freiberger Mulde). All COSMO-LEPS (C-L) members are equal weighted.

Flood alert levels and activities	C-L 7.8. +5.5d	C-L 8.8. +4.5d	C-L 9.8. +3.5d	C-L 10.8. +2.5d	C-L 11.8. +1.5d	C-L 12.8. +0.5d
1: report	0.5	0.4	0.6	1	1	1
2: control	0.3	0.2	0.5	1	1	1
3: prepare	0.2	0.1	0.4	0.9	1	1
4: defend	0	0	0	0.6	0.2	1

The COSMO-DE model provides deterministic weather forecasts for the very short-range. A rerun for the period around the 2002 flood event has been initialized in a 3 hourly interval with a lead time of 21 hours. Fig. 7 shows a sequence of discharge simulations driven by different initialization times of COSMO-DE. This combination of forecasts is known as lagged average forecast ensemble (LAF). From the beginning of the event to the initialization of the current forecast, the hydrograph is simulated using combined information from rain gauges and radar observations provided by Haberlandt (2007) for the 2002 case study. When a new forecast is available, the latest system state of the model (driven by observed rainfall) provides hydrological initial conditions for the forecast mode.

The hydrological simulations driven by COSMO-DE forecasts underestimate discharge at the Zöblitz gauge with respect to the amplitude and the duration. In particular, the precipitation forecasts generate discharges of varying quality indicating that the latest COSMO-DE forecast not necessarily gives the best estimation of the observed discharge. This strengthens the need for a LAF approach as shown in Fig. 7.

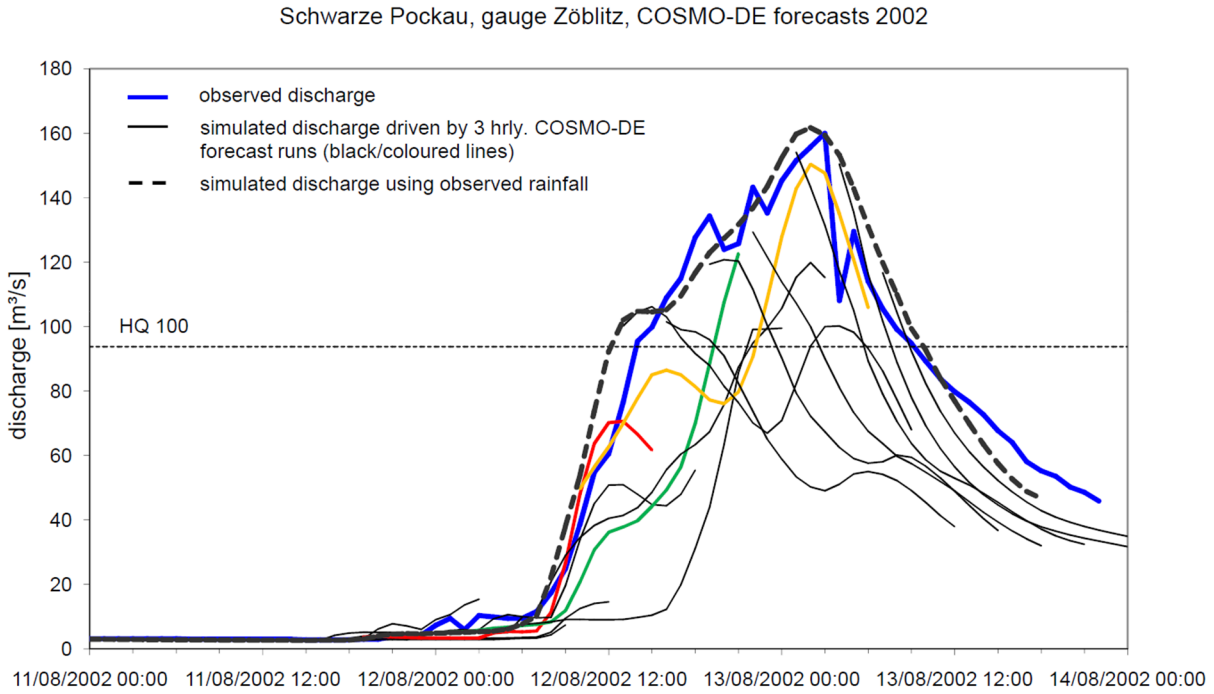


Fig. 7: Simulated discharge for the Zöblitz gauge using observed rainfall and COSMO-DE predictions of precipitation (model time step 1h).

The COSMO-DE modelling chain predicts two major discharge events. The first peak is strongly underestimated by most runs, while for the second peak at least four runs indicate a strong increase of discharge. Fig. 8 shows precipitation patterns from two selected COSMO-DE runs displaying the two-peak structure consistently forecasted by all COSMO-DE runs. The first precipitation event lasts approximately four hours showing strong localized precipitation maxima, while the second event incorporates 7 hours of heavy precipitation distributed over a much wider area. Due to the more spacious character of the forecasted patterns of the second event, the Zöblitz discharge forecasts are much more reliable. The time-lag between the forecasted two rainfall events was very small (3 hours). Thus the hydrographs do not show a noticeable two-peak structure because of the superimposition of the different components of the hydrological system. When comparing COSMO-DE precipitation forecasts with hourly observations from the Haberlandt study (not shown), the observed precipitation patterns show a similar spacious character than the COSMO-DE forecasts of the second peak. In contrast to the COSMO-DE forecasts this spacious character dominates the whole period of more than 20 hours of observed precipitation generating the highest ever recorded flood peak at the Zöblitz gauge.

Table 2 shows the persistence of threshold exceedances for flood alert levels at the Berthelsdorf gauge for the deterministic COSMO-DE forecasts. At 12/08/2002 06:00 UTC COSMO-DE gave the first evidence of a potential severe flood event exceeding alert level 3 19 hours later (13/08/ 01:00). In the following initializations of the modelling chain the predicted time of exceeding level 3 moved backwards to 12/08/ 15:00 (which is partly due to hydrologic uncertainty, see rising limb of the calibrated hydrograph shown in Fig. 4). The model initializations from 12/08/ 18:00 on gave strong evidence of an extreme flood event. Nevertheless there was always at least one overlapping model run, which did not forecast the exceedance of level 4 (210 m^2/s). The observed flood peak was 360 m^2/s at 13/08/ 03:00.

Table 2: Exceedance of flood alert levels at Berthelsdorf as forecasted by COSMO-DE. The head row shows the forecast hours from 11/08/2002 20:00. The other rows show the alert levels, which are exceeded by the forecasted discharge for each time step. The two bottom rows show the number of overlapping forecasts, which exceed alert level 3 resp. 4. If the level has been exceeded, the cells are marked as grey.

init. / forecast hr.	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
11/08/2002 18:00	0	0	1	1	1	0																																	
11/08/2002 21:00	0	0	0	0	0	0	0	0	0																														
12/08/2002 00:00	0	0	0	0	0	0	0	0	1	1	1	2																											
12/08/2002 03:00	0	0	0	0	0	0	0	0	0	1	1	1	1	1																									
12/08/2002 06:00	0	0	0	0	0	0	0	1	1	1	1	1	2	2	2	2	3	3																					
12/08/2002 09:00	0	0	1	1	1	1	2	2	2	3	3	3	3	4	4	4	4	4	4	4																			
12/08/2002 12:00				1	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4															
12/08/2002 15:00							3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3												
12/08/2002 18:00										4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3									
12/08/2002 21:00													4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3						
13/08/2002 00:00																4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3				
13/08/2002 03:00																			4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3		
13/08/2002 06:00																						4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3		
13/08/2002 09:00																									4	4	4	4	4	4	4	4	4	4	3	3	3		
13/08/2002 12:00																												4	4	4	4	4	4	4	4	3	3		
13/08/2002 15:00																															4	4	4	4	4	3	3		
13/08/2002 18:00																																		4	4	3	3		
exceed. of level 3				0	0	0	1	2	2	4	4	4	5	5	5	6	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7		
exceed. of level 4				0	0	0	0	0	0	1	1	1	2	2	3	4	4	4	5	6	6	6	6	6	6	5	5	5	5	5	6	6	5	5	4	1		

Compared to the results of the forecasts driven by COSMO-LEPS, the performance of the COSMO-DE initializations before 12/08/ 18:00 was surprisingly weak. The very short-range forecasts did not provide more reliable exceedance probabilities than the medium-range forecasts did. However the very short-range forecasts performed better in estimating the peak discharge during the event (cf. results for the Zöblitz gauge shown above).

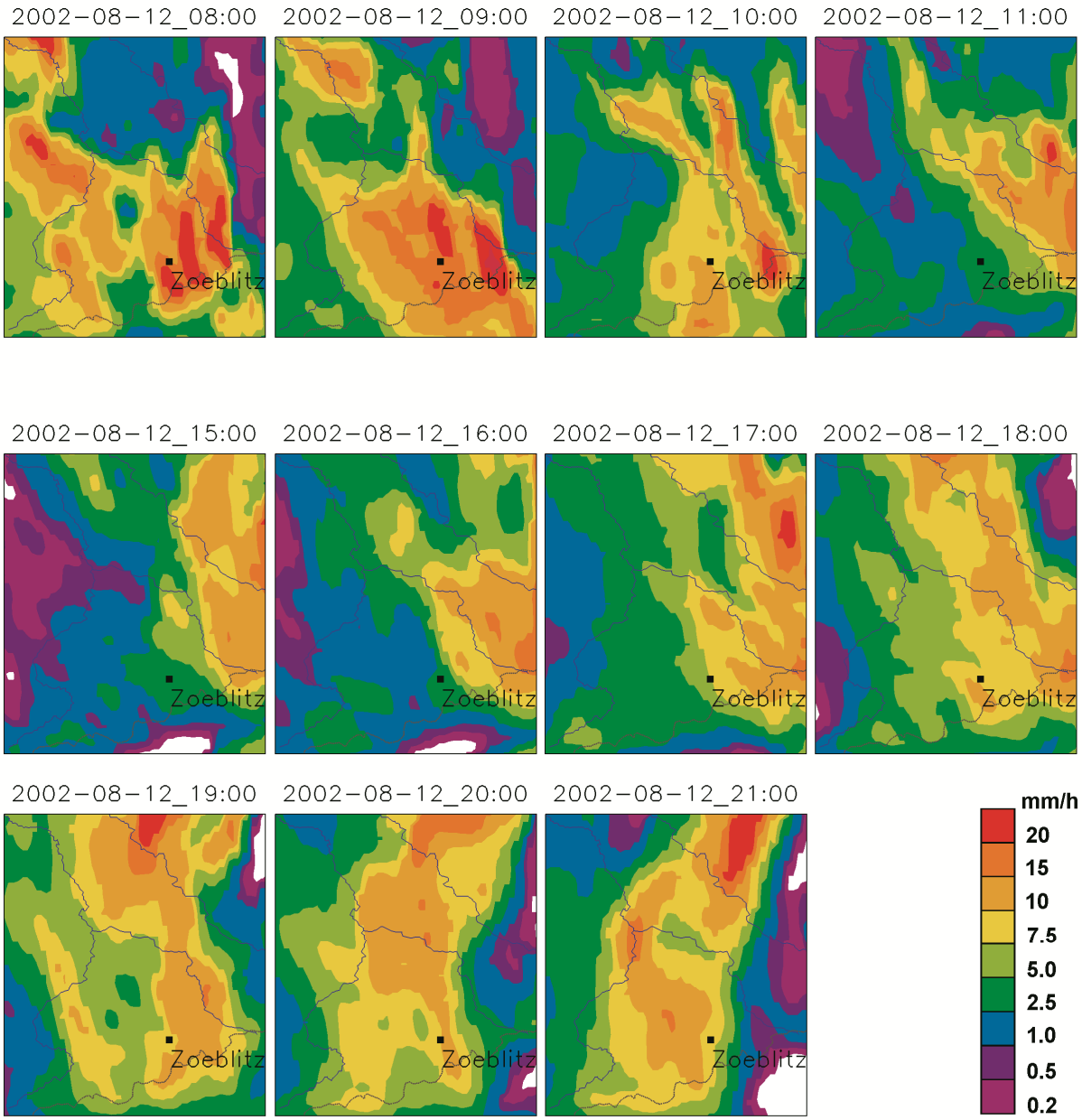


Fig. 8. COSMO-DE forecasts for the Mulde catchment with forecast initialisations at 11/08/2007 18:00 UTC (first row, red forecast in Fig. 7) and 12/08/2007 0:00 UTC (second and third rows, green forecasts in Fig. 7) showing the first and second precipitation maximum in the Zöblitz catchment respectively.

7 Conclusions

A flood forecast strategy combining meteorological ensembles with different spatial and temporal resolution and a calibrated rainfall-runoff model enable the simulation of probabilistic discharge forecasts for meso-scale catchments. These forecasts can account for different sources of uncertainties, e.g. uncertainties of the initial conditions, model structure and model parameters from both the meteorological and the hydrological models. Data assimilation allows an update of the probabilistic assessment of the ensembles in an operational real-time environment. There is a chance, but no guarantee to reduce uncertainty successively during the event with an adaptive model control strategy. Compared with deterministic forecasts, probabilistic forecasts of

discharge provide additional information to flood managers, who are addressed by the flood management strategy presented in Sect. 2 of this paper.

A part of the proposed flood management strategy has been successfully applied for the simulation of hindcasts for an extreme flood event. The COSMO-LEPS and COSMO-DE hindcasts consistently underpredicted the rainfall rate of the 2002 extreme event. Nevertheless both systems evaluated within this case study produced evidences for a rainfall event which could cause a flood in the dimension of a 100 year recurrence period. If such a probabilistic forecast would already have been operationally available in August 2002, a reliable flood forecast would have been possible with a lead time of 2 to 3 days. Flood management would have been significantly improved compared to the deterministic forecast, which had been issued at that time. Both the COSMO-LEPS and COSMO-DE simulations did not always improve with more recent forecasts.

The derivation of decision rules, e. g. which exceedance probabilities should be used for issuing alerts or initiating flood defence measures, is subject of ongoing research. Within the study area, ensemble hindcasts were only available for one single extreme event. General conclusions about the skill of the models cannot be drawn from this case study. Currently the authors evaluate false alarms to extend their knowledge about the predictive skills of these models. In future work related to the forecast skill of the meteorological models, the study area will be extended in order to evaluate a larger number of extreme rainfall events. Further work of the authors is focused on the hydrological ensemble generation mechanisms. A detailed description and results from evaluation of this method will be published in a subsequent paper. The prototype of an operational system will be extended to the lower Mulde catchment including the adaptive control of flood routing and inundation models at vulnerable hot spots. The probabilistic approach will then enable an explicit consideration of risk and can thus further improve the basis for decision making in operational flood management.

The presented ensemble based flood forecast strategy does not directly address long-term strategic flood management. Nevertheless it is aimed to be an integral part of a sustainable flood management policy. The longer forecasting horizon, the possible reduction of the number of false alerts and a detailed prediction of expected flood damage can improve the preparedness of affected people and thus reduce potential damage.

Acknowledgements

The research reported in this paper was carried out with support from the German Ministry for Education and Research (BMBF) under the initiative “Risk Management of Extreme Flood Events (RIMAX)”. The authors thankfully acknowledge funding. Bernd Pfützner (Büro für Angewandte Hydrologie BAH Berlin) provided source code and support of the ArcEGMO model. Data and practical experience have been contributed by the State Flood Center of Saxony (Dresden), the Saxonian Reservoir authority (Pirna), the Saxonian land survey authority and the National German Weather Service (DWD).

References

Ajami, N.K., Duan, Q. and Sorooshian, S.: An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, Water Resour. Res., 43, W01403, doi:10.1029/2005WR004745, 2007.

Ajami, N.K., Gupta, H., Wagener, T. and Sorooshian, S.: Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system, Journal of Hydrology, 298 (1-4), 112-135, 2004.

Anderson, J. L.: A method for producing and evaluating probabilistic forecasts from ensemble model integrations, Journal of Climate, 9, 1518-1530, 1996.

Apel, H., Thieken, A.H., Merz, B. and Blöschl, G.: Flood risk assessment and associated uncertainty, Natural Hazards and Earth System Sciences, 4, 295-308, 2004.

Bartels, H., Dietzer, B., Malitz, G., Albrecht, F. M., Guttenberger, J.: Starkniederschlagshöhen für Deutschland, KOSTRA-DWD-2000. Fortschreibungsbericht. Deutscher Wetterdienst – Hydrometeorologie, Offenbach, 2005.

Becker, A. and Grünewald, U.: Disaster management: flood risk in Central Europe, Science, 300 (5622), 1099, 2003.

Becker, A., Klöcking, B., Lahmer, W. and Pfützner, B.: The hydrological modelling system ARC/EGMO, in: Mathematical models of large watershed hydrology (Eds.: Singh, V.P. and Frevert, D.K.), Water Resources Publications, Littleton/Colorado, 2002.

Beven, K. J., and Binley, A.: The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes, 6, 279-298, 1992.

Buizza, R., Houtekamer, P. L., Toth, Z., Pellerin, G., Wei, M. and Zhu, Y.: A comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. Mon. Wea. Rev., 133, 1076–1097, 2005.

Büttner, U., Fügner, D. and Winkler, U.: das Hochwasser am 5./6. Juli 1999 im Raum Marienberg im Erzgebirge, Hydrologie und Wasserbewirtschaftung 45 (3), 102-111, 2001 [The flood on 5 and 6 July 1999 in the region of Marienberg in the Ore Mountains, Saxonia, in German with English abstract and captions].

Carpenter, T. M. and Georgakakos, K. P.: Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales, Journal of Hydrology 329 (1-2), 174-185, doi:10.1016/j.jhydrol.2006.02.013, 2006.

De Roo, A., Gouweleeuw, B., Thielen, J., Bates, P., Hollingsworth, a. et al.: Development of a European Flood Forecasting System, I. R. B. M., 1(1), 49-59, 2003.

Denhard, M. and Trepte, S.: Calibration of the European multi-model ensemble SRNWP-PEPS, Second THORPEX international science symposium, WMO/TD No. 1355, WWRP/THORPEX No. 7, 2006.

Doms, G. and Förstner, J.: Development of a kilometre-scale NWP-system: LMK. In Doms, G., Schättler, U. and Montani, A. (editors), COSMO Newsletter No. 4, 168-176, 2004.

Evensen, G.: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99, 10143–10162, 1994.

Georgakakos, K. P., Seo, D.-J., Gupta, H., Schaake, J. and Butts, M. B.: Towards the characterization of streamflow simulation uncertainty through multimodel ensembles, Journal of Hydrology, 298 (1-4), 222-241, 2004.

Gouweleeuw, B., Thielen, J., Franchello, G., De Roo, A. and Buizza, R.: Flood forecasting using medium-range probabilistic weather prediction. Hydrol. Earth Syst. Sci., 9(4), 365-380, 2005.

Haberlandt, U.: Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event, Journal of Hydrology, 332, 144-157, 2007.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T.: Bayesian model averaging: A tutorial with discussion. Statistical Science, 14, 382-417, 1999. Printing errors corrected in version available at http://www.stat.washington.edu/www/research/online/hoeting1999.pdf.

Hoffman, R. N. and Kalnay, E.: Lagged average forecasting, an alternative to Monte Carlo forecasting, Tellus, 35A, 100-118, 1983.

Holtan, H.N.: A concept for infiltration estimates in watershed engineering, Agr. Res. Service Pub., U.S. Dept. Agr., 41-51, 1961.

Kalnay, E.: Atmospheric modelling, data assimilation and predictability, Cambridge University Press, 2002.

Klink, S. and Stephan, K.: Latent heat nudging and prognostic precipitation. COSMO Newsletter, 5, 124-131, 2005.

Krzysztofowicz, R.: Bayesian system for probabilistic river stage forecasting, J. Hydrol., 268, 16-40, 2002.

Lorenz, E. N.: Deterministic Nonperiodic Flow. J. Atmosph. Sci. 20, 130-141, 1963.

Marsigli, C., Boccanera, F., Montani, A. and Paccagnella, T.: The COSMO-LEPS mesoscale ensemble system: validation of the methodology and verification, Nonlin. Proc. Geophys., 12, 527–536, 2005.

Molteni, F., Buizza, R., Marsigli, C., Montani, A., Nerozzi, F. and Paccagnella, T.: A strategy for high-resolution ensemble prediction, part I: definition of representative members and global model experiments, Quart. J. Roy. Meteor. Soc., 127, 2069-2094, 2001.

Montani, A., Marsigli, C., Nerozzi, F., Paccagnella, T., Tibaldi, S. and Buizza, R.: The Soverato flood in Southern Italy: performance of global and limited-area ensemble forecasts. Nonlin. Proc. Geophys., 10, 261-274, 2003.

Mullusky, M., Demargne, J., Welles, E., Wu, L. and Schaake, J.: Hydrologic applications of short and medium range ensemble forecasts in the NWS, Advanced Hydrologic Prediction Services (AHPS), Extended Abstract, 84, AMS Meeting, Seattle, WA. Jan. 10.-16., 2004.

Pappenberger, F., Beven, K. J., Hunter, N. M., Bates, P. D., Gouweleeuw, B. T., Thielen, J. and de Roo, A. P. J.: Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European Flood Forecasting System (EFFS), Hydrology and Earth System Sciences, 9 (4), 381-393, 2005.

Raftery, A. E., Gneiting, T., Balabdaoui, F. and Polakowski, M.: Using Bayesian Model Averaging to calibrate forecast ensembles. Mon. Wea. Rev., 133, 1155-1174, 2005.

Reed, S., Schaake, J. and Zhang, Z.: A distributed hydrologic model and threshold frequency-besed method for flash flood forecasting at ungauged locations, J. Hydrology, 337, 402-420, 2007.

Schaake, J. C., Hamill, T. M., Buizza, R. and Clark, M.: HEPEX: The Hydrological Ensemble Prediction Experiment. Bull. Amer. Meteor. Soc., 88 (10), 1541-1547, 2007.

Sloughter, J. M., Raftery, A. E. and Gneiting, T.: Probabilistic quantitative precipitation forecasting using Bayesian Model Averaging. Tech. report no. 496, Department of Statistics, University of Washington, 2006.

Smith, M.B., Seo, D.-J., Koren, V.I., Reed, S.M., Zhang, Z., Duan, Q., Moreda, F. and Cong, S.: The distributed model intercomparison project (DMIP): motivation and experiment design. Journal of Hydrology, 298 (1-4): 4-26, 2004.

Steppeler, J., Doms, G., Schättler, U., Bitzer, H.W., Gassmann, A., Damrath, U. and Gregoric, G.: Meso-gamma scale forecasts using the nonhydrostatic model LM. Meteorol. Atmos. Phys., 82, 75-96, 2003.

Thielen, J., Bartholmes, J., Ramos, M.-H. and de Roo, A.: The European Flood Alert System – Part 1: Concept and development. Hydrol. Earth Syst. Sci. Discuss., 5, 257-287, 2008.

Toth, Z., Talagrand, O., Candille, G. and Zhu, Y.: Probability and ensemble forecasts, in: Joliffe, I. T. and Stephenson, D. B. (eds.): Forecast verification: a practitioner's guide in atmospheric science, John Wiley & Sons, 2003.

Trepte, S., Denhard, M., Göber, M. and Anger, B.: SRNWP-PEPS: some results of verification, Second THORPEX international science symposium, WMO/TD No. 1355, WWRP/THORPEX No. 7, 2006.

Verbunt, M., Zappa, M., Gurtz, J., and Kaufmann, P.: Verification of a coupled hydrometeorological modelling approach for alpine tributaries in the Rhine basin, J. Hydrol., 324, 224-238, 2006.

Wang, Y., Dietrich, J., Voss, F. and Pahlow, M.: Identifying and reducing model structure uncertainty based on analysis of parameter interaction, Adv. Geosci., 11, 117-122, 2007.

This chapter is an edited version of the following original scientific article:

Dietrich, J.; Trepte, S.; Wang, Y.; Schumann, A. H.; Voß, F.; Hesser, F. B.; Denhard, M. (2008): Combination of different types of ensembles for the adaptive simulation of probabilistic flood forecasts. Hindcasts for the Mulde 2002 extreme event. Nonlin. Processes Geophys. 15 (2), 275–286.

Publisher: Copernicus Publications

II. Assessing uncertainties in flood forecasts for decision making: prototype of an operational flood management system integrating ensemble predictions

Dietrich, J.[1,*], Schumann, A. H.[1], Redetzky, M.[2], Walther, J. [2], Denhard, M.[3], Wang, Y.[1], Pfützner, B.[4], Büttner, U.[5]

(1) Institute of Hydrology, Water Resources Management and Environmental Engineering, Ruhr University Bochum, Germany
(2) DHI-WASY GmbH, Dresden, Germany
(3) Deutscher Wetterdienst DWD (German National Weather Service), Offenbach, Germany
(4) Büro für Angewandte Hydrologie, Berlin, Germany
(5) Sächsisches Landesamt für Umwelt, Landwirtschaft und Geologie, Dresden, Germany
(*) now at: Institute of Water Resources Management, Hydrology and Agricultural Hydraulic Engineering, Leibniz University, Hanover, Germany

Abstract

Ensemble forecasts aim at framing the uncertainties of the potential future development of the hydro-meteorological situation. A probabilistic evaluation can be used to communicate forecast uncertainty to decision makers. Here an operational system for ensemble based flood forecasting is presented, which combines forecasts from the European COSMO-LEPS, SRNWP-PEPS and COSMO-DE prediction systems. A multi-model lagged average super-ensemble is generated by recombining members from different runs of these meteorological forecast systems. A subset of the super-ensemble is selected based on a priori model weights, which are obtained from ensemble calibration. Flood forecasts are simulated by the conceptual rainfall-runoff-model ArcEGMO. Parameter uncertainty of the model is represented by a parameter ensemble, which is a priori generated from a comprehensive uncertainty analysis during model calibration. The use of a computationally efficient hydrological model within a flood management system allows us to compute the hydro-meteorological model chain for all members of the sub-ensemble. The model chain is not re-computed before new ensemble forecasts are available, but the probabilistic assessment of the output is updated when new information from deterministic short range forecasts or from assimilation of measured data becomes available. For hydraulic modelling, with the desired result of a probabilistic inundation map with high spatial resolution, a replacement model can help to overcome computational limitations. A prototype of the developed framework has been applied for a case study in the Mulde river basin. However these techniques, in particular the probabilistic assessment and the derivation of decision rules are still in their infancy. Further research is necessary and promising.

1 Introduction

Due to the intrinsic uncertainty of meteorological forecasts, flood forecasts are also affected by uncertainty. Furthermore, as hydrological models are used for transformation of rainfall into runoff, their structural parameter uncertainty should be considered in the forecasts as well. Inaccurate human interaction and technical problems may also affect the output of a flood forecast chain. Thus uncertainty is an issue of concern when dealing with flood forecasts. On the one hand,

an underestimation or even missing of a flood warning may hinder the affected people from preparing for a flood. As a consequence damage may increase or even casualties may occur. On the other hand, "crying wolf" too often may encourage people to ignore warnings in the future.

During the last decades, modelling evolved from a deterministic towards a probabilistic paradigm. Nowadays forecasters have information about uncertainty. The imperfection of forecasts is more and more accepted and the communication of uncertainties does not automatically make the users losing confidence in the forecast. Numerous approaches for uncertainty estimation and probabilistic evaluation were developed. Interdisciplinary studies dealing with the probabilistic assessment of the flood forecast chain were published e. g. by Krzysztofowicz (2002), Apel et al. (2004) and Pappenberger et al. (2005).

The predictive uncertainty characterises differences between observed values and forecasted values (Fig. 1). The forecasts are derived from a single model with defined parameters under event-specific initial and boundary conditions. Forecasts of many different events can be used to characterise the forecasts uncertainty empirically.

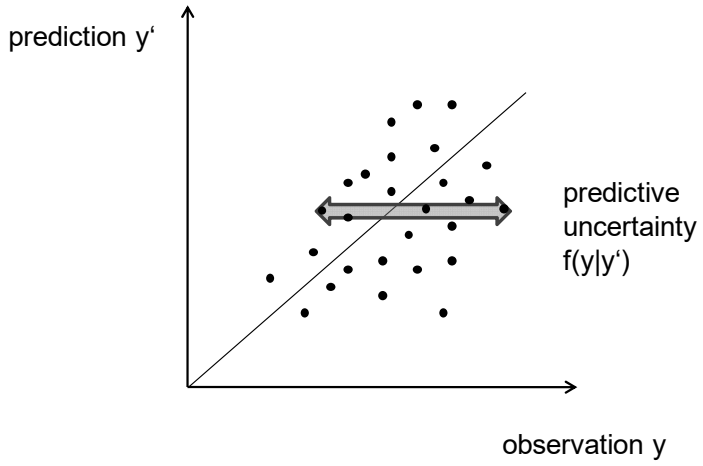


Fig. 1. Predictive uncertainty, describing the differences of observations from a prediction.

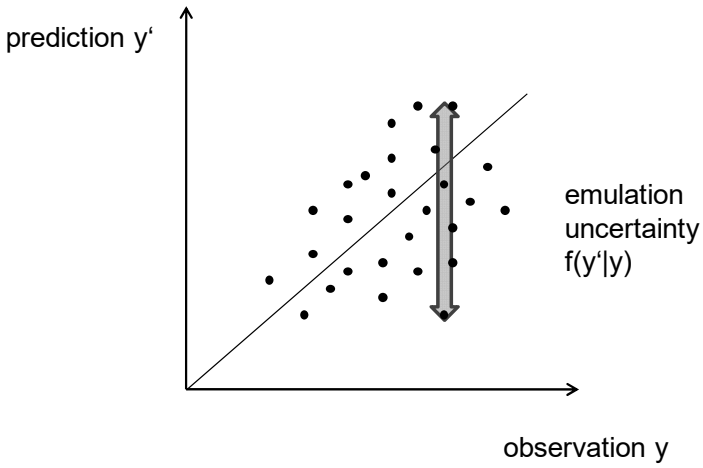


Fig. 2. Emulation uncertainty (Todini, 2010).

Ensemble techniques were developed in order to frame the uncertainty with a relatively low number of simulations (Anderson, 1996; Kalnay, 2002; Toth et al., 2003). Within the context of

operational forecasting, this makes ensemble techniques superior to many other ways of generating a probability distribution instead of a single solution. Thus ensembles specify the differences between several forecasts (Fig. 2). Different types of ensembles can be classified according to the generating mechanisms (for meteorological as well as for hydrological applications):

- single system ensembles: variation of initial and boundary conditions, different model components, e. g. convection schemes (physically based ensembles), variation of model parameters;
- multiple systems or multi-model ensembles ("poor man ensembles"): combination of simulations from different models (e. g. Georgakakos et al, 2004; Ajami et al., 2007);
- lagged average ensembles: combination of current forecasts with forecasts from earlier model runs (Hoffman & Kalnay, 1983).

Ensembles characterise the so-called "emulation uncertainty" (Todini, 2010). In informatics, the term "emulation" describes the imitation of the behaviour of a computer or other electronic system with the help of another type of computer of system. The differences between forecasts within an ensemble can be used to discuss the effects of the mechanisms used to construct this ensemble.

However, the probabilities within ensembles are not aleatoric ones, because many epistemic uncertainties are included in their estimations. Frequentists use Monte Carlo simulation to account for uncertainty associated with the parameters of a probability model that Bayesian methods handle natively. Ensembles are based on a lower number of data points and make many assumptions about the models and other characteristics. It should be considered that the existing ensemble forecasting systems do not completely represent the uncertainties of models. Hence a probabilistic evaluation of the outcome is needed. Let's say there may be more uncertainty in uncertainty propagation itself. Thus it is challenging to develop ensemble generation mechanisms which do not only result in a bunch of model outcomes, but also represent the probabilistic assessment of the variables under consideration. The basic assumption is that each result has the same probability. Neither the total variation (which would result from a large amount of possible combinations of uncertain aspects of modelling) nor a differentiation in more or less probable forecasts can be represented in this way.

Nowadays meteorological ensemble prediction systems (EPS) are operational at global scale (e.g. ECMWF, MSC, NCEP, Buizza et al., 2005) and regional scale. The COSMO-LEPS (Molteni et al., 2001) is a limited area physically based EPS for the medium range (3 to 5 days lead time). It was developed within the COSMO (Consortium for Small-scale Modelling) to improve the predictability of extreme weather events in Central and Southern Europe. The added value of the system resides in joining the skill of the ECMWF EPS to depict the possible evolution scenarios with the capability of the COSMO limited area model to improve the descriptions of local meteorological processes. Further regional EPS in Europe are MOGREPS, NORLAMEPS, ARPEGE/ALADIN and PEARP. The short-range SRNWP-PEPS (Denhard and Trepte, 2006) combines up to 23 deterministic forecasts from 21 national meteorological services with a lead time of two days. It can be seen from case studies and probabilistic verification for Germany (Trepte et al. 2006) that this ensemble is a valuable tool for severe weather forecasting. A major benefit of this multi-model EPS is the possibility to compare the behaviour of all operational European limited area models.

The development of hydrological applications of ensemble forecasts has started in the late 1990-ies and is subject of ongoing research (e. g. Verbunt et al., 2006; Komma et al., 2007; Reed et al., 2007; Diomede et al., 2008). The participatory HEPEX project (Hydrological Ensemble Prediction Experiment, Schaake et al., 2007) integrates meteorologists, hydrologists and users in order to promote the development of ensemble stream flow forecast systems. In Europe, the probabilistic Flood Alert System (EFAS) is under development (Thielen et al., 2009). EFAS aims to provide flood information for the medium to long-range at large scale river basins being relevant for decisions at national or EU level.

For the setup and near real-time operation of an uncertainty aware flood management system, a compromise between predictive capability of the models, computational efficiency and cognitive burden for the flood managers is still unavoidable. On the one hand, data flow and control activities must be automated to the highest achievable level. On the other hand, the complex nature of the problem requires options for flood managers and decision makers to take control over the simulation process, e. g. when sources of information are identified as unreliable or when parts of the model chain fail. This contribution presents a framework for ensemble based flood forecasting, which is based on hydrological forecasts driven by operational meteorological EPS with different spatial resolution and different lead times. The hydrological models are controlled in an adaptive way, mainly depending on the lead time of the forecast, the expected magnitude of the flood event and the availability of measured data.

In the following section we introduce the respective workflow and a corresponding software prototype for an operational flood management system (OFMS). In Sect. 3 we demonstrate the combination of ensemble members from different meteorological forecast systems and different model runs to generate a multi-model super-ensemble with weighted members. Furthermore we introduce an approach for the a priori framing of parameter uncertainty with a hydrological ensemble. Results from a case study in the Mulde river basin including hindcast simulations of historic flood events were shown by Dietrich et al. (2008) and Dietrich et al. (2009). Here we focus on methodological developments.

2 Methods

2.1 Concept and workflow of an operational flood management system

The operational flood management system (OFMS) presented in this paper is designed for ensemble-based flood forecasting in meso- to macro-scale river basins (100 km^2 to 10.000 km^2 contributing catchment area). The OFMS integrates meteorological forecasts from different operational limited area prediction systems and supports the flood manager in managing uncertainty. Typically the OFMS is installed in a regional or national flood management centre.

The OFMS combines meteorological medium-range forecasts from COSMO-LEPS (3 to 5 days lead time), short-range forecasts from SRNWP-PEPS (1 to 2 days lead time) and very short-range forecasts (< 1 d lead time) in an adaptive manner (Fig. 3). For the latter the convection resolving, deterministic COSMO-DE model with 21 hours lead time and horizontal resolution of 2.8 km is used (Doms and Förstner, 2004). The COSMO-DE model forecasts from different initialization times are combined to generate a lagged average forecast (LAF) ensemble, which proofed to be an improvement over the use of the latest deterministic model run in the Mulde case study

(Dietrich et al., 2008; Dietrich et al., 2009). The lead time of the OFMS could be extended to 10 ore more days by way of using predictions from global systems, but the spatial resolution of these systems restricts their applicability in meso-scale flood forecasting.

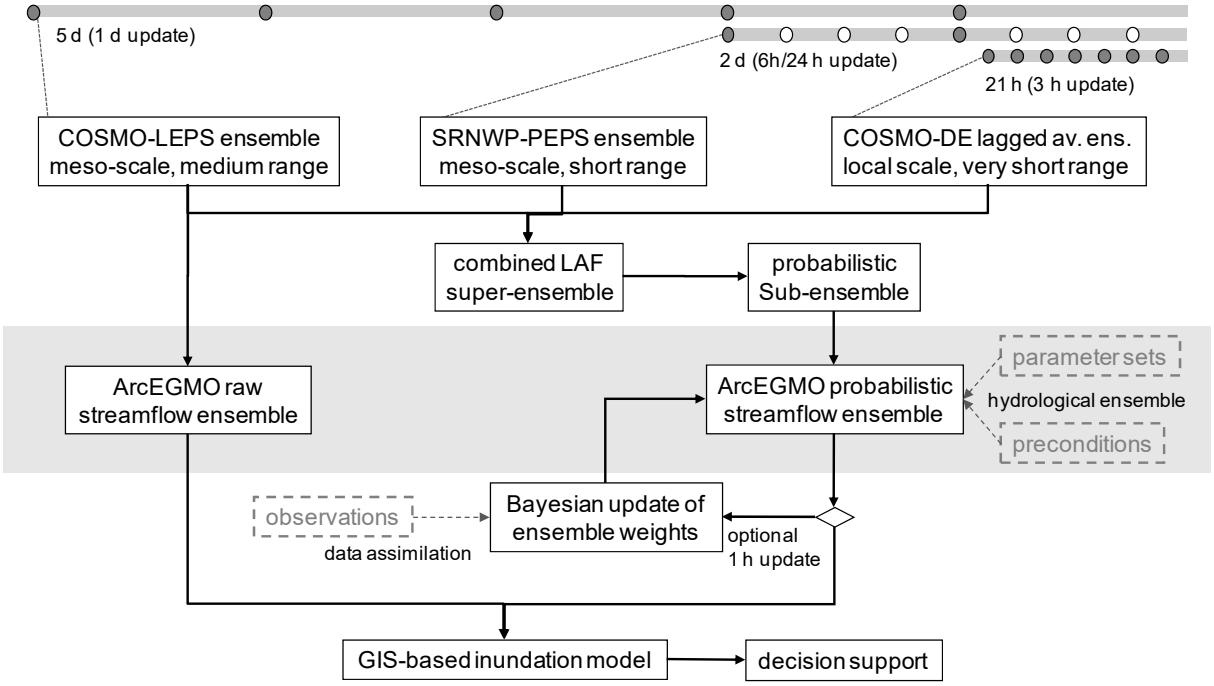


Fig. 3. Flow chart of the flood forecast chain.

Each of the three meteorological prediction systems has individual advantages for flood forecasters. Medium-range forecasts provide the basis for decisions about reservoir management and early warnings previous to a potential extreme flood event. Short to very short-range forecasts are used for issuing flood alerts and planning of local flood defence measures. One would expect that the shorter range forecasts are more accurate than the medium range forecasts. As it has been shown in the Mulde case study mentioned above, this expectation does not always hold.

In a real time situation, the attention of flood managers is focused on a variety of aspects and the computational and cognitive resources are limited. There is not time to speculate which result would be the best this time. Decision makers need detailed information about relevant criteria like peak time, peak discharge and possible inundation areas. An approach for combining the maximum information from a priori knowledge with the actual forecasts is the generation of a calibrated super-ensemble for the short-range, which also includes information from medium range forecasts and from earlier model runs. Because such a multi-model lagged average super-ensemble has too many members for real time computation of stream flow forecasts, a reduced sub-ensemble is generated as a representative subset of the members of the super-ensemble. The sub-ensemble aims at producing a high resolution probabilistic weather scenario as input for flood forecast models (Fig. 3, detailed description in the following Section).

A hydrological rainfall-runoff model simulates ensemble forecasts of stream flow at several points of interest like gauges and vulnerable sites. We decided to choose the conceptual hydrological model ArcEGMO (Becker et al., 2002) as the default OFMS component for transformation of quantitative weather forecasts into stream flow forecasts for the Mulde case study. Conceptual rainfall-runoff models describe the complex natural hydrological processes in

a simplified manner. These models are widely used for the meso- and macro-scale due to their significant advantages compared to physical models regarding parameter estimation and computational efficiency (Carpenter and Georgakakos, 2006; Smith et al., 2004).

The OFMS simulates stream flow forecasts for all members and all initialization times of the meteorological prediction systems, using uncalibrated "raw" ensembles (assuming that a bias correction has been performed by the data providers before, if necessary). Based on the generation rules for the meteorological sub-ensemble, the simulations can be recombined and weighted to build a stream flow ensemble. Here, the hydrological model is computed with a default parameter set, which proved to be efficient for historic flood events. If we assume that the chosen model structure has sufficient predictive capabilities and input uncertainty is expressed by the meteorological sub-ensemble, two additional sources of hydrological uncertainty are regarded: the initial state of the model and the selection of model parameters (Sect. 2.3). While the m-member super-ensemble is simulated with a single "best" set of hydrological model parameters, the smaller n-member sub-ensemble is simulated with p sets of parameters building up an n*p-member probabilistic stream flow ensemble.

A Bayesian inference mechanism adjusts the weights of the ensemble members when new observed rainfall and discharge data become available during the event. The Bayesian updating procedures can, but not necessarily must reduce the number of members. We aim at successively reducing the number of parameter ensemble members for a proper relation between meteorological and hydrological contributions to total (knowable) uncertainty. Ideally, this procedure sequentially reduces uncertainty by using new information when available. Note that we do not recalibrate the parameters online.

2.2 Meteorological super-ensemble and sub-ensemble

The meteorological ensemble prediction systems COSMO-LEPS, SRNWP-PEPS and COSMO-DE/LAF try to frame the future development of the weather situation by variation of initial states and boundary conditions and/or modifications of the model physics. Per se all forecasts have the same probability. However the probabilistic assessment of ensemble members can be updated by statistical post-processing when observations become available. Nevertheless there is a need to aggregate this information a) to reduce the number of simulations of flood forecast models (or other model chains driven by meteorological input) and b) to compute a kind of best estimation (e.g. the median or a weighted average), which can be used like a deterministic forecast.

The Bayesian model averaging method (BMA, Raftery et al., 2005) assigns a weight and an error distribution to each model by means of iterative optimization. The predicted probability distribution of the complete ensemble is computed by overlay of the error distributions of the single models for the respective forecasted variables. There is a training period necessary for calibrating the BMA. The longest possible training period is still very short (two years) and includes only a few heavy rainfall events. Unfortunately there are no re-forecasts available for the three systems under consideration. For local extreme rainfall events a training of BMA seems not to be possible at all because of the unique nature of such events. There is a danger of over fitting the calibration parameters of the BMA, in particular in the context of flood forecasting.

Thus we developed an alternative ensemble post-processing method. An eventually missing spread of the ensemble (observations are not fully framed) is not compensated by stamping an

error distribution on the ensemble but by integrating previous COSMO-LEPS model runs into the ensemble. Even though old forecasts have in general less skill the COSMO-LEPS system is designed to give optimal spread estimates in the medium range (forecast day 3-5). This information can be used to improve short range probability forecasts.

The older COSMO-LEPS model runs normally have a higher root mean square error than the short range forecasts, but their higher spread helps to frame reality, which is an improvement of the basic ensemble, if proper weights are assigned to the lagged forecasts. This is done by an iterative multiple linear regression approach (IMLR) between observations and forecasts. Starting from the full regression with all super-ensemble members only the members with a regression coefficient greater than zero survive the first iteration. This procedure is repeated with the remaining members until all regression coefficients are positive.

The result is a weighting scheme for the remaining members including a bias correction. In Fig. 4 (bottom), *f* is the weighted mean of the *M* selected forecasts from the super-ensemble of size *K*. The term *b* is the bias correction and the positive coefficients *w* are normalized such that they sum up to one.

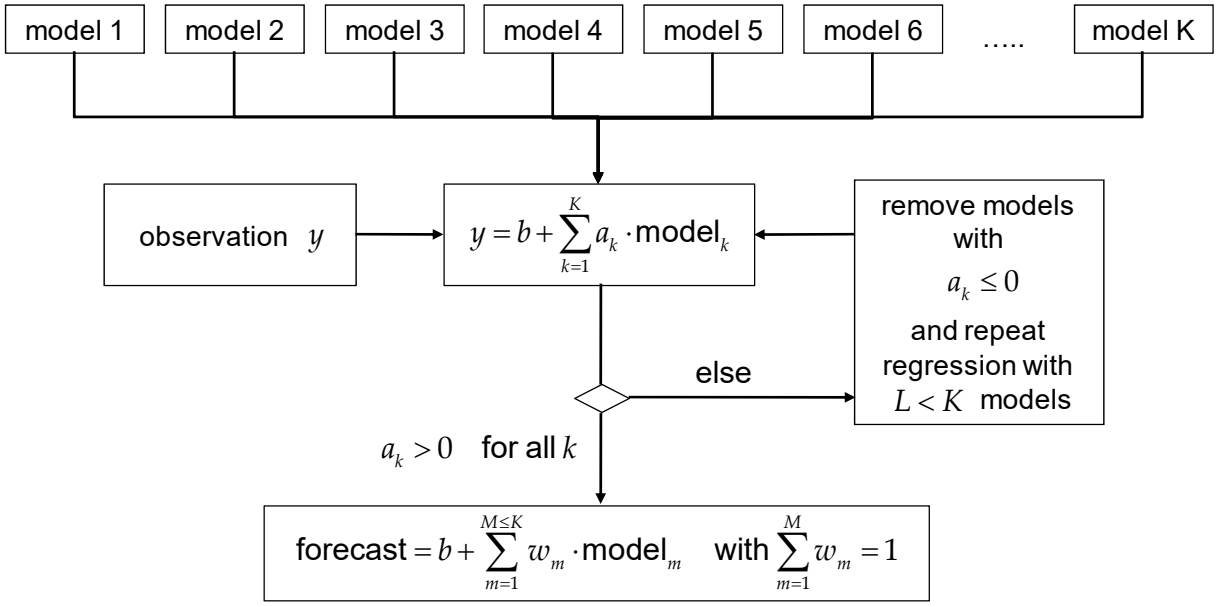


Fig. 4. Workflow of the sub-ensemble calibration procedure

2.3 Combined hydro-meteorological ensembles

The ArcEGMO rainfall-runoff-model was chosen for the OFMS because of its computational efficiency and its process oriented conceptualisation. On the one hand the first estimation of the model parameters can be performed by GIS analyses, which is beneficial when experts perform this task and experience from previous work is available. On the other hand the model has more parameters than less process based conceptual models have. In ArcEGMO, there is an interaction between parameters. Also some parameters are less sensitive in most but not all situations. The simulated runoff is an overlay of several processes, which can not be calibrated separately in case of missing observations for the different runoff components. In the mathematical sense the model is over parameterized. Thus calibration results in a large number of equifinal sets of parameters (i. e. sets of parameters with different values produce model outputs with similar efficiency).

Furthermore, the number of flood events available for calibration is limited. The event specific response of the hydrological system may not have been sufficiently observed. Thus parameter uncertainty is an issue of concern.

For operational use of the model near real time, the following main steps are performed:

1. compute initial state from continuous simulation (daily or hourly time steps)
2. update initial state for the preceding seven days (hourly time step)
3. simulate raw ensembles to build up the super-ensemble flood forecast
4. simulate parameter ensembles driven by the members of the reduced meteorological sub-ensemble

Uncertainty in initial conditions (e.g. soil moisture, fast groundwater storage) is reduced by an iterative trial and error system state update procedure, which adjusts the storage representing soil moisture and upper groundwater until the hydrograph of the seven days preceding the expected rainfall is best represented by the model. This initial state update violates continuity, but it is important to tackle uncertainty from prior simulation of the continuous model without performing a recalibration during real-time application of the system. Alternatively, an automated state update is possible with the Ensemble Kalman Filter (Evensen, 2003), which is also implemented in the OFMS. Only in specific situations (mainly very heavy rainfall at the beginning of the event) the model is very sensitive even against the updated system state. Here an additional state ensemble can be used (this has not been done in the case study).

For framing parameter uncertainty of the hydrological model we propose an ensemble approach. Hydrological parameter ensembles are generated by combination of model parameter sets, which proved to be efficient for simulating flood events in the calibration and validation periods. For the derivation of the generating rules of the parameter ensemble, we combine optimization, stochastic methods and expert knowledge.

In the first step we used multi-criteria optimization (Yapo et al., 1998; Vrugt et al., 2003) to find efficient parameter sets for single historic flood events. The selection and weighting of objective functions allows finding a large set of numerically efficient solutions. Hereby it is possible to differentiate between characteristic types of rainfall events. Among the numerically efficient solutions might be some parameter sets, which have a less realistic composition of runoff components. Here a plausibility check is performed by model experts in order to exclude such parameter sets. After theses analyses the non dominating process parameters can be fixed and the dominating parameters can be limited within plausible ranges.

A large number of parameter sets is generated by variation of the dominating parameters with a Monte Carlo experiment. Structural uncertainty of the model produces mixed quality of model efficiency for the different flood events. Thus the parameter ensemble contains a priori defined sets of parameters, which are efficient for different types of past flood events. For a multi-criteria selection of the parameter sets to be integrated in the ensemble we used compromise programming. The ensemble members are weighted according to the a priori expected type of event. If no decision about the type of event is possible, these parameter sets are equally weighted. The parameter ensemble concept is designed for real-time application with the possibility to update the weights of the ensemble members when new information becomes available. The restricted number of possible simulations in real time (we already have to simulate the meteorological ensembles!) restricts the number of members of the parameter ensemble.

In operational forecasting it is possible to assimilate observed stream flow data. These data can be useful for the reduction of model uncertainty. Four general types of update procedure are established (Refsgaard, 1997):

1. correction of input data,
2. update of internal state variables,
3. recalibration of model parameters,
4. direct correction of model output.

By using meteorological ensembles from external providers, the first approach is not recommended (the meteorologist should use data assimilation and apply a bias correction instead, if required). Within the event, the correction of internal state variables violates mass continuity and may produce unreliable model results in the near future. The online recalibration of model parameters during the event is critical for similar reasons. Thus we apply approach 4. Based on the method described above we generate a parameter ensemble with 20 sets of parameters, which cover the range of observed events. With a single driving meteorological input, we perform 20 simulations of the rainfall-runoff model, which 20-folds the computational demand of the ensemble simulations.

The resulting stream flow ensemble is evaluated against observed stream flow. A likelihood value according to the principle of Bayesian inference (Box & Tiao, 1992) is computed for the time series of each ensemble member (respectively this is the likelihood for each parameter set). Thus the model output produced by the 20 sets of parameters can be weighted according to their likelihood values. From the ensemble members with higher likelihood values one can construct an uncertainty band. Additionally it is possible to compute a weighted average, which can be communicated like a deterministic forecast. After new rainfall observations are available, new runoff simulations are performed with the 20 sets of parameters. After new stream flow observations are available, the likelihood values of these simulations are evaluated for the generation of new weights of the members of the stream flow ensemble.

2.4 OFMS prototype implementation

The software prototype of an operational flood management system (OFMS) integrates data, models and GIS-based tools (Fig. 5). The core components of the OFMS are based on the ESRI ArcGIS software architecture, particularly the DHI-WASY “GeoFES” spatial decision support system for emergency management. The OFMS interface assists the user in the intelligent control of hydrological and hydraulic simulations as well as spatial analyses and the visualization of model output. The ArcEGMO model code was not completely integrated in the OFMS for reason of computational efficiency, but the existing ArcEGMO software was redesigned to work as a set of three dynamic link libraries controlled by the OFMS interface. This allows the simulation of forecasts from an a priori defined system state without re-initialization, but with different climatic inputs. Based on the meteorological ensembles, which are transferred from the operational forecasts database to the hydrological model, the OFMS starts the required number of model runs and manages analysis, aggregation, visualization and archiving of results. The OFMS GIS tools support further analyses like the computation of flood areas with a replacement model for hydraulic simulation (Sect. 3.3) and risk analyses.

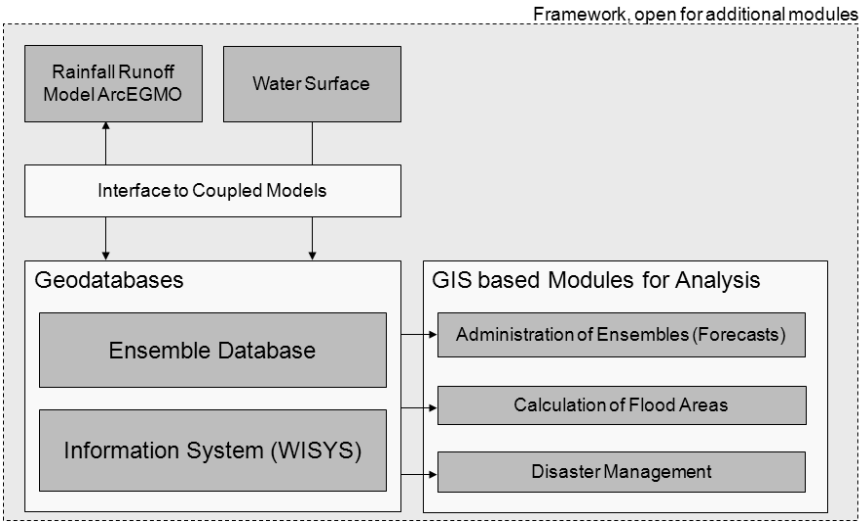


Fig. 5. Methodological concept and components of the operational flood management system (OFMS)

In the OFMS, required data for flood management is maintained in two geodatabases: the information system and the ensemble database. Based on the UML object model of the WISYS software (Becker et al., 2006), the information system comprises spatial base data and further geodata originating mainly from the flood protection concepts of the respective rivers. Dynamic data for the model applications and their results is maintained in the ensemble database. Amongst other things, it includes discharge values calculated with the ArcEGMO simulation model, derived water levels and flood plains as well as the measured precipitation and discharges which are retrievable from respective databases (of the State Flood Center of Saxony in the Mulde case study).

To connect the simulation model ArcEGMO to the geodatabases, an interface was developed. It guarantees the data pre-processing and data provision necessary for the model, the model initialization and the storage of the simulation results in the ensemble database. Data pre-processing and provision is carried out in two steps. Firstly, elementary areas are defined by intersecting land use, soil data and catchment area boundaries under consideration of elevation. These elementary areas are combined to hydrological response units and then intersected with the grid cells of the respective meteorological ensemble prediction systems. The area proportion of the EPS grid cells in the hydrological response units is then calculated and saved. The initialization of ArcEGMO and the storage of the simulation results occur for all simulations selected in the user interface.

A post-processing tool converts simulated discharges to water levels. It uses the data tables stored in the information system that, for gauge sites, correspond to the valid stage-discharge relation. The predicted water levels are linked to the water levels of the flood alert level of the respective gauges. This enables a forecast of the flood alert level at the gauges for all simulated time steps. As mentioned before, it is still difficult to make a probabilistic judgement of ensemble members. The OFMS prototype offers frequency analysis of the ensemble forecasts. Alert persistence charts (Fig. 6) can be generated for the gauges, aggregating the ample results of an ensemble forecast and reducing them to the information essential for the user. From the alert persistence chart the following information can be deduced:

- The duration (columns, coloured background) of the predicted alert level for each ensemble available for the query time steps (rows),
- for each time step for which the alert levels are determined, the number of ensembles exceeding the alert levels 3 and 4 respectively,
- the time between the time step for which the alert levels are determined and the query time step.

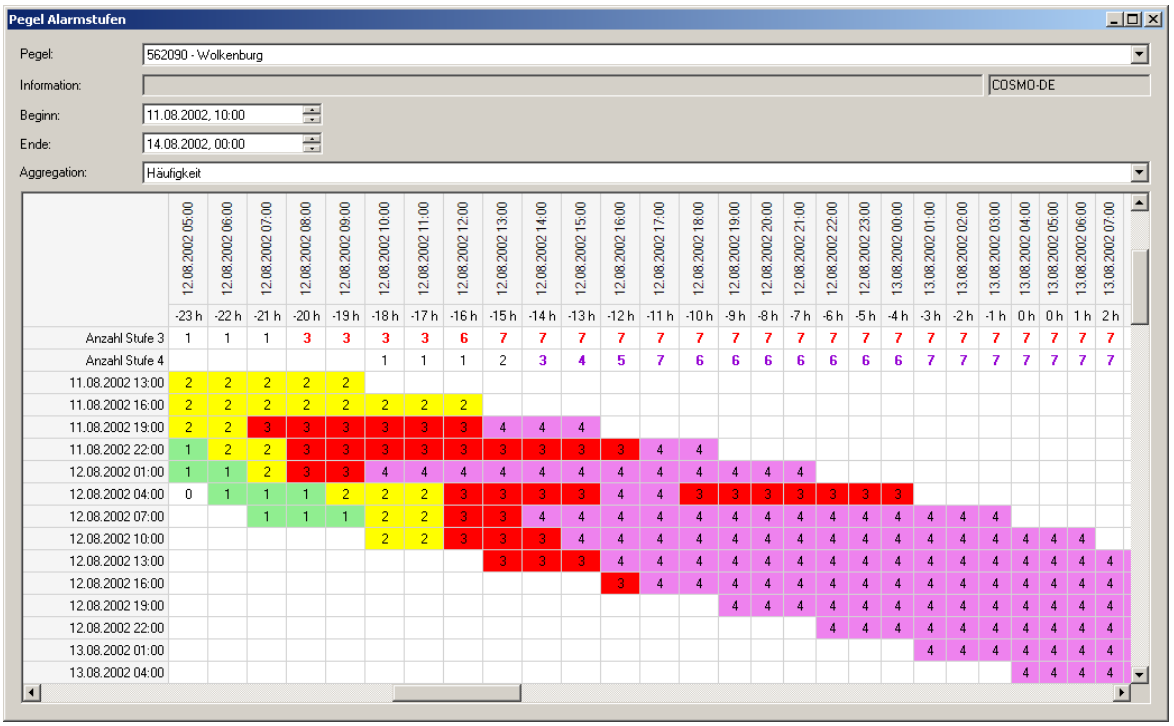


Fig. 6. Persistence chart as presented by the OFMS user interface.

Alert persistence charts are very demonstrative because of their compact and temporal overlapping presentation of the available results at query time. The considerable uncertainty, especially persisting during the rise of a flood, and the stability (persistence) with increasing forecast duration are illustrated clearly. Alert persistence charts are, consequently, an adequate instrument for decision support in ensemble-based flood forecasting.

3 Experimental Application

3.1 Meteorological Super-Ensemble/Sub-Ensemble demonstration for 2007/2008

For the period from May 2007 to April 2008 a multi-model super-ensemble of accumulated 12 h (6 to 18 UTC) rainfall forecast for Germany was generated. In the first step the 17 members of the SRNWP-PEPS forecast were combined with the 0 UTC COSMO-DE forecast and the median of the COSMO-LEPS forecast from the preceding day to build up a 19 member ensemble with equal weights. After a bias correction of the single models, the COSMO-DE is one of the best performing models. The COSMO-LEPS median is often the best model regarding the outliers. Thus the 19 member super-ensemble is considered an improvement over the single systems.

To fully implement the idea of adding lagged average forecasts, in a second step, we combined a lagged average multi-model super-ensemble from 4 lagged COSMO-DE, runs (3, 0, 21, 18 UTC),

up to 5 COSMO-LEPS runs of the preceding 5 days and the SRNWP-PEPS run of 0 UTC. The super-ensemble combines up to 101 forecasts (17 SRNWP-PEPS, 4 COSMO-DE and 80 COSMO-LEPS from 5 model runs with 16 members each).

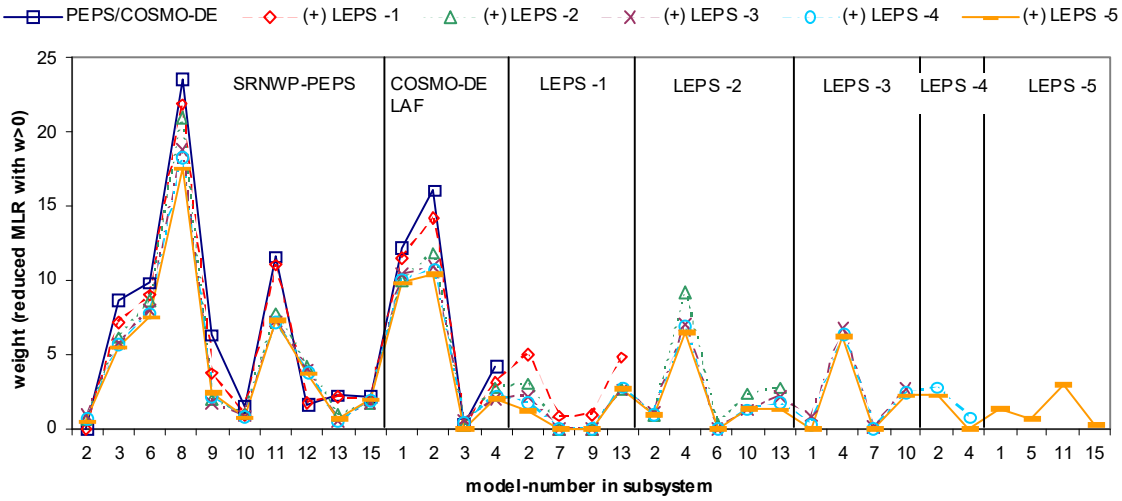


Fig. 7. Weights of the selected members from different super-ensembles by the iterative regression method (IMLR) described in the text. All super-ensembles contain the SRNWP-PEPS (PEPS) forecasts initialized at 0 UTC and the COSMO-DE lagged average forecasts from 3 and 0 UTC on the day of the validation as well as 21 and 18 UTC from the day before. The model numbers (1-4) in the COSMO-DE LAF ensemble increase with increasing lag-time. Time-lagged COSMO-LEPS forecasts are added (+) to this "basic" super-ensemble, where e.g. -1 indicates the LEPS run from the day before the validation time period.

For summer 2007, the ensemble could be reduced by application of the multiple linear regression approach to a sub-ensemble built up by 26 members (Fig. 7). With an independent validation dataset, the forecast quality of the sub-ensemble was better than with other post-processing methods (Schumann and Dietrich, 2009). The calibration of the ensemble improved the forecast in the case study.

3.2 Hydrological parameter ensemble demonstration for the 2002 extreme flood

The parameter ensemble updating procedure (Sect. 2.3) aims at the reduction of forecast uncertainty based on data assimilation. Meteorological uncertainty is propagated through the hydrological model and cannot be completely separated from hydrological uncertainty in real time. As driving meteorological input we used a) the observed precipitation from rain gauges up to the forecasting point and b) afterwards the COSMO-DE very short range forecast, which has a lead time of 21 hours and is initialized every 3 hours. The selected forecasting point at 13/08/2002 01:00 UTC is at the end of the rainfall event, but 10 hours before the flood peak passes the Wechselburg 1 gauge. The likelihood values of the members of the parameter ensemble are shown in Fig. 8 a). It is obvious that the error in the stream flow forecast is mainly caused by the underestimation of precipitation in the forecast (Fig. 8 b and c). The uncertainty of rainfall input dominates the uncertainty of the model parameters in this example.

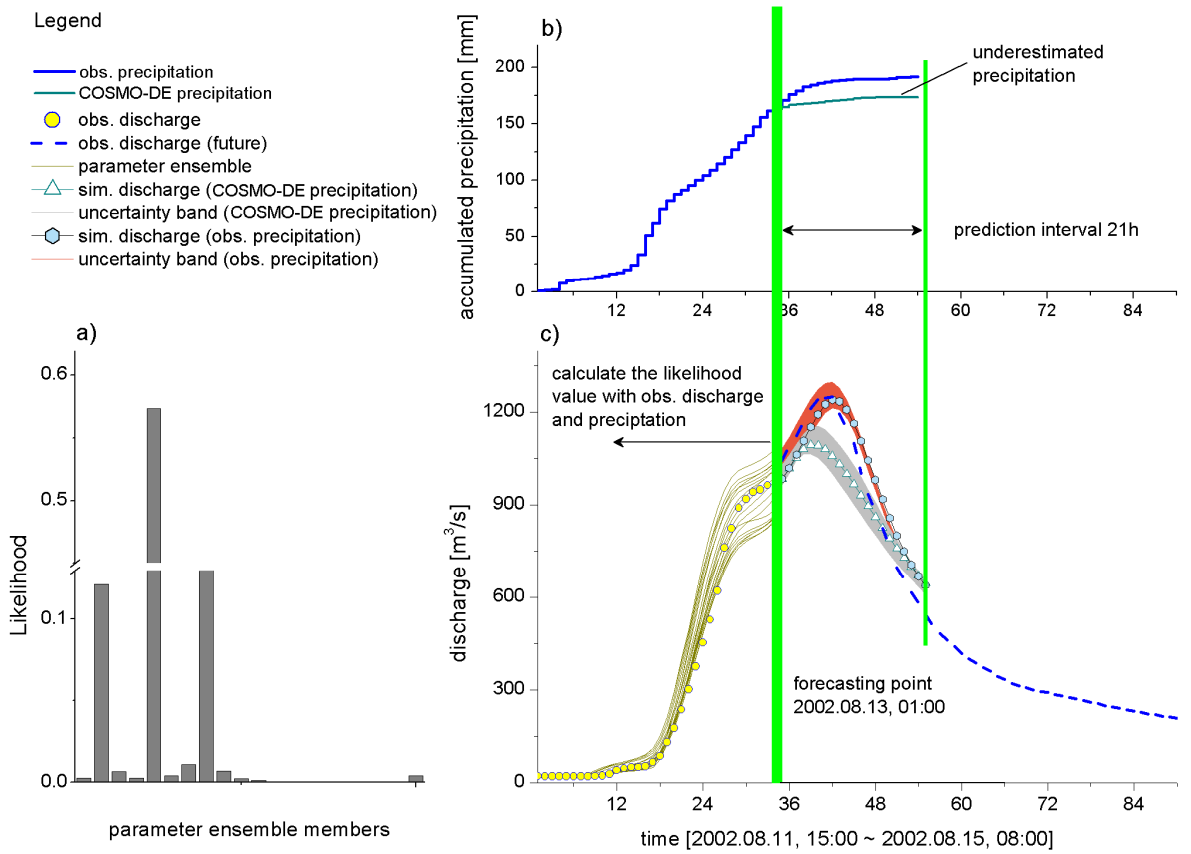


Fig. 8. Parameter ensemble updating procedure shown by example of the 2002 extreme flood event in the Mulde river (Wechselburg 1 gauge). Likelihoods for the parameter sets of the hydrological model are shown in a), b) compares accumulated precipitation from observation and deterministic forecast, c) compares the bands of hydrological parameter uncertainty. Evolving likelihoods reduce the band width of forecasts. The differences resulting from utilisations of observed or forecasted precipitation are also shown in c).

The combination of the meteorological sub-ensemble and the hydrological parameter ensemble adds up to a stream flow ensemble with 520 members, which can be completely updated every 24 or 12 hours (when all systems deliver new data) and partly updated every 3 hours (when the COSMO-DE lagged average ensemble receives one new member). Due to the ArcEGMO model code optimization described above, the complete update computation of the stream flow ensemble takes about one hour at a personal computer workstation.

3.3 Replacement model for GIS-based inundation forecast

To be able to calculate flood plains and to carry out risk analysis, further data tables (water levels, discharge) are calculated for a sample of points in the watercourses and archived in the information system. This is done with the help of the hydrological models developed in the OFMS. Simulated water levels are the primary data for the calculation of flood plains (Fig. 9).

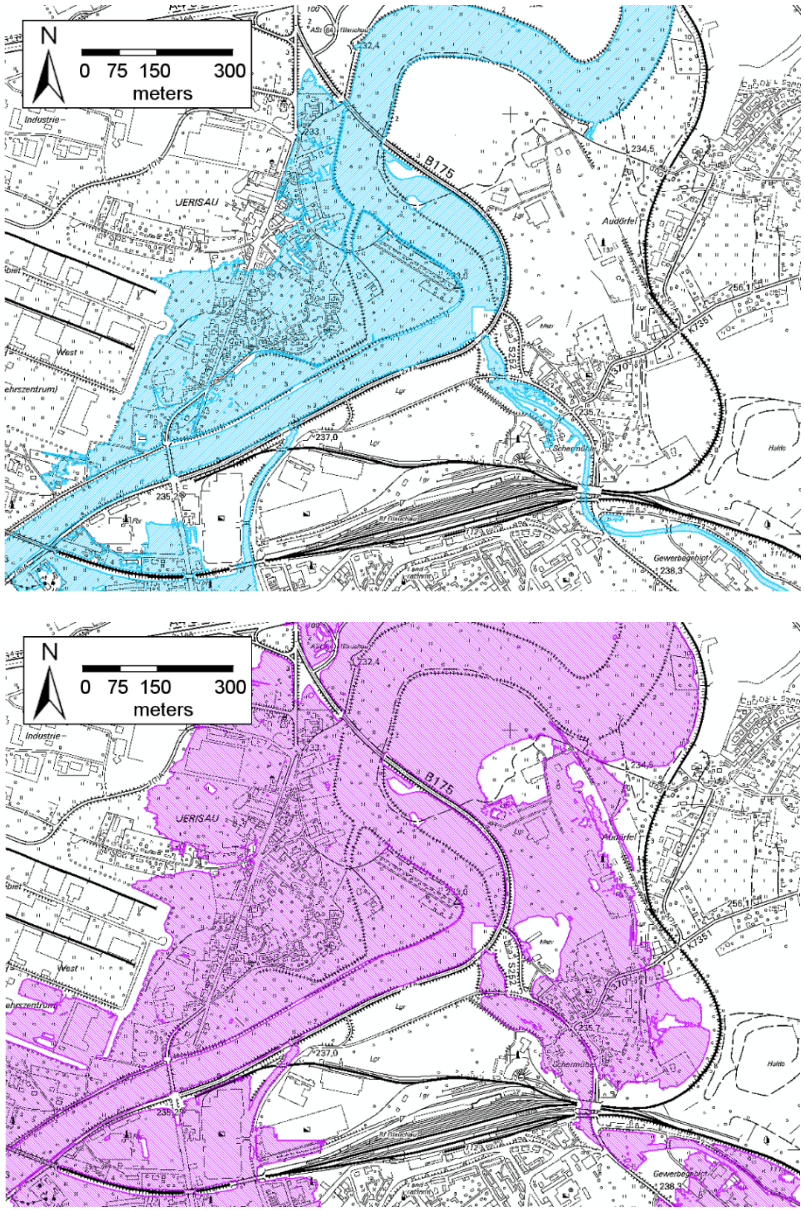


Fig. 9. Calculated flood plain in the area of the mouthing of the Lungwitzbach into the Zwickauer Mulde in Glauchau on 12/08/2002, 6 am (top) and 6 pm (bottom)

The determination of flood plains using 1D-hydrological calculations consists of the interpolation of a water surface and the intersection with a digital elevation model (DEM). A triangulated irregular network (TIN) is created, based on a network of base points located in the potential flood plain. The distance between the base points and the resulting resolution of the TIN determine the degree of detail of the resulting flood plain. When defining the network base points, the DEM applied (being the base data) and the detail requirements of the calculation should be taken into consideration. Interpolating the values of the routed water course, a kilometre specification is assigned to all chosen base points in the flood plain. The generation of the base point network is executed only once.

As the previously selected sample points in the watercourse are situated along the routed watercourse, they also have a kilometre specification, as well as calculated water levels. Joining sample points and network base points through their kilometre specification, water levels can be interpolated and assigned to each base point. The result is a water surface as a TIN, which can

then be converted into a GRID. Cells without a direct connection to the water course are identified and marked as „not flooded“. After intersecting this GRID with the DEM, an inundation model is now available as a first result. The contour of this inundation model, in turn, corresponds to the flood extent.

This tool also allows taking culvert constructions and barriers into account. However, within the OFMS this function is not made use of.

The computational demand of the calculation of flood plains is directly related to the resolution of the elevation model as well as to the number of base points used in the calculation. The number of sample points in the watercourses in the test system/prototype with stage-discharge relations adds up to 1600, with 18000 base points in total. With a spatial resolution of 2x2 m, the calculation time amounts to about 10 minutes per forecast in the case study. When using a raster of 20x20 m, the calculation time can be reduced by a factor of 10. One must consider that a fine-tuning of the prototype was not conducted.

In the OFMS Mulde, one can chose between two modes: On one hand, it is possible to calculate the flood areas for a specific time step in the forecast period. If this procedure is repeated for various time steps, the forecasted temporal progression of the flood can be visualized. The second mode allows the determination of the maximum extent of the flood plain within the complete forecast period by calculating the maximum water level for each sample point in the watercourse. These two modes can be utilized in the OFMS according to the task in question. With the definition of the maximum flood plain extent during the forecast period a fast overview of all potentially flooded areas is possible. A time step oriented analysis can be applied when generating evacuation plans or assigning roads for provision delivery.

4 Conclusions and discussion

In this study and in the two preceding studies (Dietrich et al., 2008; Dietrich et al., 2009) we demonstrated the use of medium range to very short range ensembles in flood forecasting for meso-scale river basins. In most of the hindcasts the ensembles were able to frame uncertainty, even at the extreme 2002 flood event (Dietrich et al., 2008). The hindcasts have shown an improvement of the spatial and temporal resolution of flood forecasts compared to conventional forecasts. The extension of lead times is possible, but at the cost of higher uncertainty, which is in turn reflected in the ensembles.

A prototype of an operational flood management system (OFMS) was developed, which combines forecasts from three meteorological systems with different characteristics and different predictive capability. The OFMS supports flood managers and decision makers in managing uncertainty. The prototype demonstrates the applicability of stream flow ensemble forecasts in an operational environment, e.g. at local authorities.

A major drawback of the probabilistic assessment of ensembles is the limited availability of hindcasts (resp. re-forecasts), that means forecasts for already observed events, which have been computed with the operational forecast system afterwards. For the three systems under consideration there was only a two year period from the pre-operational setup of the models available. From this time series it is impossible to perform an a priori assessment of meteorological uncertainty in extreme weather situations. With a longer time series we suppose

to be able to assign more reliable weights to the members of the combined super-ensemble, which are valid for the specific large scale weather situation. Despite these limitations we have developed an approach for the probabilistic assessment of the super-ensemble based on a one year time series to demonstrate the potential value of ensemble post-processing for flood forecasting. We are not yet able to validate the method under the conditions of heavy or extreme rainfall. Further work must be related to this issue. Here we see options to integrate likelihoods, derived from observed data into real-time assessments of ensemble forecasts. For meteorological data this could be done e.g. with radar data to consider the spatial distribution and the movement of convective cells during extreme rainfall events.

In general the integration of observations in the flood management strategy is very important. Observations deliver evidences about the quality of the forecast. Further more they allow the adaptation of the (uncertain) probabilistic assessment of the forecast ensembles by processing all available information. This could support local human forecasters which have to judge unusual situations and modify forecasts based on their familiarity with models and the flood situation in the area of interest (Blöschl, 2008).

The hydrological parameter ensemble is a promising approach for regarding uncertainty of the hydrological model. However, the interaction between parameter uncertainty and input uncertainty cannot be completely resolved during calibration. One can expect that some of the optimized parameter sets perform an indirect bias correction of the input (e.g. by lowering the fast groundwater storage activation level of ArcEGMO). If input uncertainty is completely framed by the meteorological ensemble forecasts, the parameter ensemble generated from historic data (with uncertain rainfall input) may even over estimate the spread of the hydrological uncertainty. The probability distribution of the combined output may be unnecessarily flattened.

The GIS-based inundation model proved to be able to simulate peak inundation reasonably well. The dynamics of inundation is simplified compared to non-stationary hydraulic models (assuming that retention effects are negligible). On the one hand, this adds structural uncertainty. On the other hand, in the context of the OFMS presented here, the forecaster may be more interested in peak inundation at the lead times under consideration. The further development of adaptive schemes a) to improve the combination and assessment of forecasts and b) to reduce the number of ensemble members is promising. Future work should also deal with the spatial heterogeneity of the forecasted variables and of forecast uncertainty.

Acknowledgements

The authors thank their colleagues and project partners for continuing support and discussion. The research reported in this paper was carried out with support from the German Ministry for Education and Research (BMBF) under the initiative “Risk Management of Extreme Flood Events (RIMAX)”. The authors thankfully acknowledge funding. Data and practical experience have been contributed by the State Flood Center of Saxony (Dresden), the Saxonian Reservoir authority (Pirna), the Saxonian land survey authority and the German National Meteorological Service (DWD). The editor thanks the anonymous referees for assisting in evaluating this paper.

References

Ajami, N.K., Duan, Q. and Sorooshian, S.: An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, Water Resour. Res., 43, W01403, doi:10.1029/2005WR004745, 2007.

Anderson, J. L.: A method for producing and evaluating probabilistic forecasts from ensemble model integrations, Journal of Climate, 9, 1518-1530, 1996.

Apel, H., Thieken, A.H., Merz, B. and Blöschl, G.: Flood risk assessment and associated uncertainty, Natural Hazards and Earth System Sciences, 4, 295-308, 2004.

Becker, A., Klöcking, B., Lahmer, W. and Pfützner, B.: The hydrological modelling system ARC/EGMO, in: Mathematical models of large watershed hydrology (Eds.: Singh, V.P. and Frevert, D.K.), Water Resources Publications, Littleton/Colorado, 2002.

Becker, A., Michels, I., McCurdy, E., Düwel, H., Müller, R. and Timmermann, R.: Informationssystem, in: Dietrich, J. & Schumann, A. (Eds..): Werkzeuge für das integrierte Flussgebietsmanagement. Ergebnisse der Fallstudie Werra – Konzepte für die nachhaltige Entwicklung einer Flusslandschaft, No.7, Weißensee-Verlag Berlin, 2006 (in German).

Blöschl, G.: Flood warning – on the value of local information. Intl. J. River Basin Management, 6 (1), 41–50, 2008.

Box, G. E. P. and Tiao, G. C.: Bayesian inference in statistical analysis, Wiley Classics Library Edition, Wiley-Interscience Publication, John Wiley and Sons 1992.

Buizza, R., Houtekamer, P.L., Toth, Z., Pellerin, G., Wei, M. and Zhu, Y.: A comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. Monthly Weather Review 133, 1076–1097, 2005.

Carpenter, T. M. and Georgakakos, K. P.: Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales, Journal of Hydrology 329 (1-2), 174-185, doi:10.1016/j.jhydrol.2006.02.013, 2006.

Denhard, M. and Trepte, S.: Calibration of the European multi-model ensemble SRNWP-PEPS. Second THORPEX international science symposium, WMO/TD No. 1355, WWRP/THORPEX No. 7, 2006.

Dietrich, J., Trepte, S., Wang, Y., Schumann, A.H., Voß, F., Hesser, F.B. and Denhard, M.: Combination of different types of ensembles for the adaptive simulation of probabilistic flood forecasts: hindcasts for the Mulde 2002 extreme event. Nonlinear Processes in Geophysics 15, 275-286, 2008.

Dietrich, J., Denhard, M. and Schumann, A.H.: Can ensemble forecasts improve the reliability of flood alerts? Journal of Flood Risk Management, 2009 (accepted).

Diomede, T., Davolio, S., Marsigli, C., Miglietta, M. M., Moscatello, A., Papetti, P., Paccagnella, T., Buzzi and A., Malguzzi, P.: Discharge prediction based on multi-model precipitation forecasts. Meteorol. Atmos. Phys. 101, 245–265, 2008.

Doms, G. and Förstner, J.: Development of a kilometre-scale NWP-system: LMK, Doms, G., Schättler, U., and Montani, A. (editors), COSMO Newsletter No. 4, 168–176, 2004.

Evensen, G.: The Ensemble Kalman Filter: Theoretical formulation and practical implementation, Ocean Dynamics 53, 343–367, 2003.

Georgakakos, K. P., Seo, D.-J., Gupta, H., Schaake, J. and Butts, M. B.: Towards the characterization of streamflow simulation uncertainty through multimodel ensembles, Journal of Hydrology, 298 (1-4), 222-241, 2004.

Hoffman, R. N. and Kalnay, E.: Lagged average forecasting, an alternative to Monte Carlo forecasting, Tellus, 35A, 100–118, 1983.

Kalnay, E.: Atmospheric modelling, data assimilation and predictability, Cambridge University Press, 2002.

Komma, J., Reszler, C., Blöschl, G. and Haiden, T.: Ensemble prediction of floods – catchment non-linearity and forecast probabilities. Nat. Hazards Earth Syst. Sci., 7, 431-444, 2007.

Krzysztofowicz, R.: Bayesian system for probabilistic river stage forecasting, J. Hydrol., 268, 16-40, 2002.

Molteni, F., Buizza, R., Marsigli, C., Montani, A., Nerozzi, F. and Paccagnella, T.: A strategy for high-resolution ensemble prediction, part I: definition of representative members and global model experiments. Quart. J. Roy. Meteor. Soc. 127, 2069-2094, 2001.

Pappenberger, F., Beven, K. J., Hunter, N. M., Bates, P. D., Gouweleeuw, B. T., Thielen, J. and de Roo, A. P. J.: Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European Flood Forecasting System (EFFS), Hydrology and Earth System Sciences, 9 (4), 381-393, 2005.

Raftery, A. E., Gneiting, T., Balabdaoui, F. and Polakowski, M.: Using Bayesian Model Averaging to calibrate forecast ensembles. Mon. Wea. Rev., 133, 1155-1174, 2005.

Reed, S., Schaake, J. and Zhang, Z.: A distributed hydrologic model and threshold frequency-besed method for flash flood forecasting at ungauged locations, J. Hydrology, 337, 402-420, 2007.

Refsgaard, J. C.: Validation and intercomparison of different updating procedures for real-time forecasting, Nordic hydrology 28, 65-84, 1997.

Schaake, J. C., Hamill, T. M., Buizza, R. and Clark, M.: HEPEX: The Hydrological Ensemble Prediction Experiment. Bull. Amer. Meteor. Soc., 88 (10), 1541-1547, 2007.

Schumann, A. H. and Dietrich, J. (eds.): Entwicklung integrativer Lösungen für das operationelle Hochwassermanagement am Beispiel der Mulde, Schriftenreihe Hydrologie und Wasserwirtschaft No. 23, Ruhr-Universität Bochum, 2009 (in German).

Smith, M. B., Seo, D.-J., Koren, V. I., Reed, S. M., Zhang, Z., Duan, Q., Moreda, F., and Cong, S.: The distributed model intercomparison project (DMIP): motivation and experiment design, J. Hydrol., 298 (1–4), 4–26, 2004.

Thielen, J., Bartholmes, J., Ramos, M.-H. and de Roo, A.: The European Flood Alert System – Part 1: Concept and development. Hydrol. Earth Syst. Sci., 13, 125-140, 2009.

Todini, E.: Predictive Uncertainty in Flood Forecasting and Emergency Management, Proc. 17th Congress of the Asia and Pacific Division of the IAHR, Auckland, New Zealand, February 24–27, 2010, accepted paper, personal communication, 2009.

Toth, Z., Talagrand, O., Candille, G. and Zhu, Y.: Probability and ensemble forecasts, in: Joliffe, I. T. and Stephenson, D. B. (eds.): Forecast verification: a practitioner's guide in atmospheric science, John Wiley & Sons, 2003.

Trepte, S., Denhard, M., Göber, M., and Anger, B.: SRNWP-PEPS: some results of verification, Second THORPEX international science symposium, WMO/TD No. 1355, WWRP/THORPEX No. 7, 2006.

Verbunt, M., Zappa, M., Gurtz, J., and Kaufmann, P.: Verification of a coupled hydrometeorological modelling approach for alpine tributaries in the Rhine basin, J. Hydrol., 324, 224-238, 2006.

Yapo, P. O., Gupta, H. V. and Sorooshian, S.: Multi-objective global optimization for hydrologic models. J. Hydrol., 204(1-4), 83-97, 1998.

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W. and Sorooshian, S.: Effective and efficient algorithm for multi-objective optimization of hydrologic models. Water Resources Research, 39(8), 1214, doi:10.1029/2002WR001746, 2003.

This chapter is an edited version of the following original scientific article:

Dietrich, J.; Schumann, A. H.; Redetzky, M.; Walther, J.; Denhard, M.; Wang, Y. et al. (2009): Assessing uncertainties in flood forecasts for decision making. Prototype of an operational flood management system integrating ensemble predictions. Nat. Hazards Earth Syst. Sci. 9 (4), 1529–1540.

Publisher: Copernicus Publications

III. Can ensemble forecasts improve the reliability of flood alerts?

Dietrich, J.[1], Denhard, M.[2], Schumann, A. H.[3]

(1) Institute of Water Resources Management, Hydrology and Agricultural Hydraulic Engineering, Leibniz University, Hanover, Germany
(2) Deutscher Wetterdienst DWD (German National Weather Service), Offenbach, Germany
(3) Institute of Hydrology, Water Resources Management and Environmental Engineering, Ruhr University Bochum, Germany

Abstract

A probabilistic evaluation of ensemble forecasts can be used to communicate uncertainty to decision makers. We present a flood forecast scheme, which combines forecasts from the European COSMO-LEPS, SRNWP-PEPS and COSMO-DE (lagged average) ensemble prediction systems with a rainfall-runoff-model. The methodology was demonstrated with a case study for the Central European Mulde River basin. In this paper we summarize results from hindcast simulations for seven events from 2002-2008. The ensemble spread resulting from uncertainty in rainfall forecast was very high at 2-5 days lead time. The median of the medium and short range forecasts and a lagged average ensemble of the very short range forecasts proved to be reliable regarding the probability of exceeding flood alert levels. However, the limited number of observed events does not allow for the postulation of prescriptive binary decision rules. Flood managers have to adapt their decisions when new information becomes available.

Introduction

Because of the highly nonlinear behavior of the atmospheric system and the land-atmosphere interaction, the knowledge of the future meteorological and hydrological development is inherently incomplete and uncertain. Particularly the quantitative forecast of rainfall events is subject to uncertainty, which often (but not necessarily) increases with a longer lead time of the forecast. Within a flood forecast chain the meteorological forecasts can force hydrological and hydraulic models. These models add additional sources of uncertainty, e. g. the availability and quality of input data, the initial and boundary conditions for the models, model parameters and model structure. Inaccurate human interaction and technical problems may also affect the output of a flood forecast chain. Resulting from these uncertainties it is not possible to issue a perfect flood forecast.

During the last decades the traditional deterministic modelling paradigm has more and more been superseded by the probabilistic paradigm, which admits the imperfectness respectively the uncertainty of the forecast. Ensemble forecasts aim at framing the uncertainty of the potential future evolution of the hydro-meteorological situation (Anderson, 1996; Kalnay, 2002; Toth et al., 2003). An ensemble based forecast produces a set of values instead of a single value of the variable under consideration. The assessment and the aggregation of information are necessary for decision support, e.g. the computation of exceedance probabilities for threshold values of critical discharge levels causing inundation.

Among the numerous ensemble generation methods are physics ensembles (perturbation of model parameters or use of different schemes within one model), multi-model ensembles (combination of different models) and lagged average ensembles (combination of different model runs of a single model). Different Meteorological ensemble prediction systems (EPS) became operational at global scale (Buizza et al., 2005). During the last years the spatial resolution of the EPS has been continuously improved. Nowadays many weather services offer EPS at regional scale with different horizontal resolutions and different lead times.

EPS can force hydrological models, which simulate the rainfall-runoff process, river routing and inundation. A chain of meteorological and hydrological models can compute flood forecasts. The development of hydrological applications of ensemble forecasts has been demonstrated by several studies (e.g. de Roo et al., 2003; Gouweleeuw et al., 2005; Verbunt et al., 2006; Reed et al., 2007; Dietrich et al., 2008; Diomede et al., 2008).

Different duties and responsibilities in water resources management, particularly in the field of flood management can potentially benefit from ensemble forecasts: reservoir control, issuing flood alerts, initiating flood defence measures. An ensemble based operational flood management system (OFMS) should fulfil different requirements regarding the relevance of uncertainty (accuracy needed) and the lead times of the forecast. This study proposes an adaptive flood management strategy, which aims at an efficient, use case driven processing and evaluation of the available forecasts and observations. The authors developed a scheme for the combination of ensembles from different sources as a basis for an operational flood management system (Dietrich et al., 2008). The OFMS integrates ensemble forecasts from three systems, which are operationally provided by meteorological services:

- COSMO-LEPS (Molteni et al., 2001) is a regional scale physics EPS with 16 (10) members, approx. 10 km horizontal resolution and five days lead time.
- SRNWP-PEPS (Denhard and Trepte, 2006) combines up to 23 deterministic forecasts from 21 national meteorological services. This multi-model-ensemble has approx. 7 km horizontal resolution and two days lead time.
- COSMO-DE (Steppeler et al., 2003) is a deterministic local model with 2.8 km horizontal resolution and 21 hours lead time. The most recent model run can be combined with earlier model runs to build a lagged average ensemble.

Water managers from local authorities were involved in the case study reported within this paper. Practitioners wondered about the reliability of probabilistic forecasts, which can be derived from the forecast ensembles. This paper addresses the following questions related to flood management:

- Can the additional knowledge about uncertainty provide added value for decision making?
- Are the probabilistic flood alert level predictions reliable?

At this time, the authors can only rely on a limited amount of data in particular for heavy and extreme events (which are of special interest for flood managers). Thus it will not be possible to draw general conclusions from this case study alone. Nevertheless we provide examples for the use of raw ensemble forecasts in one important use cases of flood managers: issuing of flood alerts. Further work will deal with refinements of the OFMS by integration of more advanced ensemble post-processing methods and data assimilation, which is of special importance for application near real-time, when flood defence measures have to be initiated and when

unexpected events (e.g. local extremes, dike breeches, loss of sensors) have to be tackled by the system.

The next sections of the paper give a short introduction into the ensemble based flood forecast framework. We evaluate and discuss a) implications from knowledge about uncertainty for issuing flood alerts and b) the reliability of flood alerts derived from the different ensemble prediction systems.

Combination of different types of ensembles for operational flood forecasting

When designing an operational flood management system (OFMS), a compromise between computational efficiency, availability of data, predictive capability of the models and the cognitive burden for the flood manager has to be found. An OFMS is typically built from components, which are among the generic components of decision support systems (DSS), namely a knowledge system, a problem processing system and a user interface (language system and presentation system, Dos Santos and Holsapple, 1989). Thus the OFMS can be seen as a specific type of DSS. As mentioned in the introductory section, there is a variety of ensemble approaches. For the specific forecast situation, only a subset of the information may be accessible and useful at the same time. Furthermore, only a subset of the problem processing tools may be needed to obtain the desired outcome of the computation. An adaptive systems approach can support the efficient combination of the sources of information, which are available and useful for the current situation and the current user of the system. The OFMS presented here follows an adaptive approach in the wider sense. Adaptive DSS in the narrower sense as defined by Holsapple et al. (1993) include additional components for unsupervised machine learning. This means that the system is able to learn and adapt itself in order to improve the quality of the outcome by using the same input data. Extending the adaptation capabilities of the OFMS is subject of further research.

The OFMS prototype described in this study combines medium-range forecasts (3 to 5 days lead time), short-range forecasts (1 to 2 days lead time) and very short-range forecasts (< 1 d lead time) from different operational meteorological prediction systems with hydrological models (Fig. 1). Medium-range flood forecasts forced by COSMO-LEPS provide the basis for decisions about reservoir management and early warnings previous to a potential large or extreme flood event. Additional short range forecasts from SRNWP-PEPS can be used for issuing flood alerts and first planning of flood defence measures. For the incorporation of forecast refinements with 2.8 km horizontal resolution and 3-hourly update we use the COSMO-DE model.

The meteorological ensembles are fed into hydrological models to simulate stream flow ensembles. Note that we do not process meteorological probability forecasts within the OFMS. We use the complete ensemble to force the hydrological models. Thus a stream flow ensemble has at least as many members as the forcing meteorological ensemble. The skill of the hydrological forecast strongly depends on the skill of the precipitation and temperature forecast, because these two climate variables dominate the generation of fast runoff processes and snow melt. The OFMS prototype presented in this paper includes the distributed conceptual rainfall-runoff model ArcEGMO (Becker et al., 2002). Most of its parameters have a physical meaning and can be derived from catchment characteristics. ArcEGMO is a modular modelling system, whose modelling kernel can be controlled via an external flood management application in a computationally efficient way.

This allows for the simulation of a large number of forecasts near real time. Thus the rainfall-runoff model can simulate ensemble forecasts of stream flow at several points of interest like gauges and vulnerable sites.

The OFMS can be operated in two modes: standard and extended. The standard mode is designed for the fast computation of raw (uncalibrated) stream flow ensembles (outside the grey area in Fig. 1). The stream flow simulations are updated whenever one of the three operational ensemble systems provides a new meteorological forecast via file transfer or database connection. The update interval is between 3 h for COSMO-DE and 1 d for COSMO-LEPS. Fig. 1 shows the update sequence in advance of a time step of interest (black dots). The SRNWP-PEPS is updated every 6 hours, but not all participating systems deliver data within that interval. The ArcEGMO model is run in a continuous mode forced by observed climatic input and in a forecast mode forced by the meteorological ensembles. The OFMS can save system states of the hydrological model at every time step. In forecast mode, the model is restarted for every single member of the meteorological forecast using the system states of the continuous model as consistent initial states.

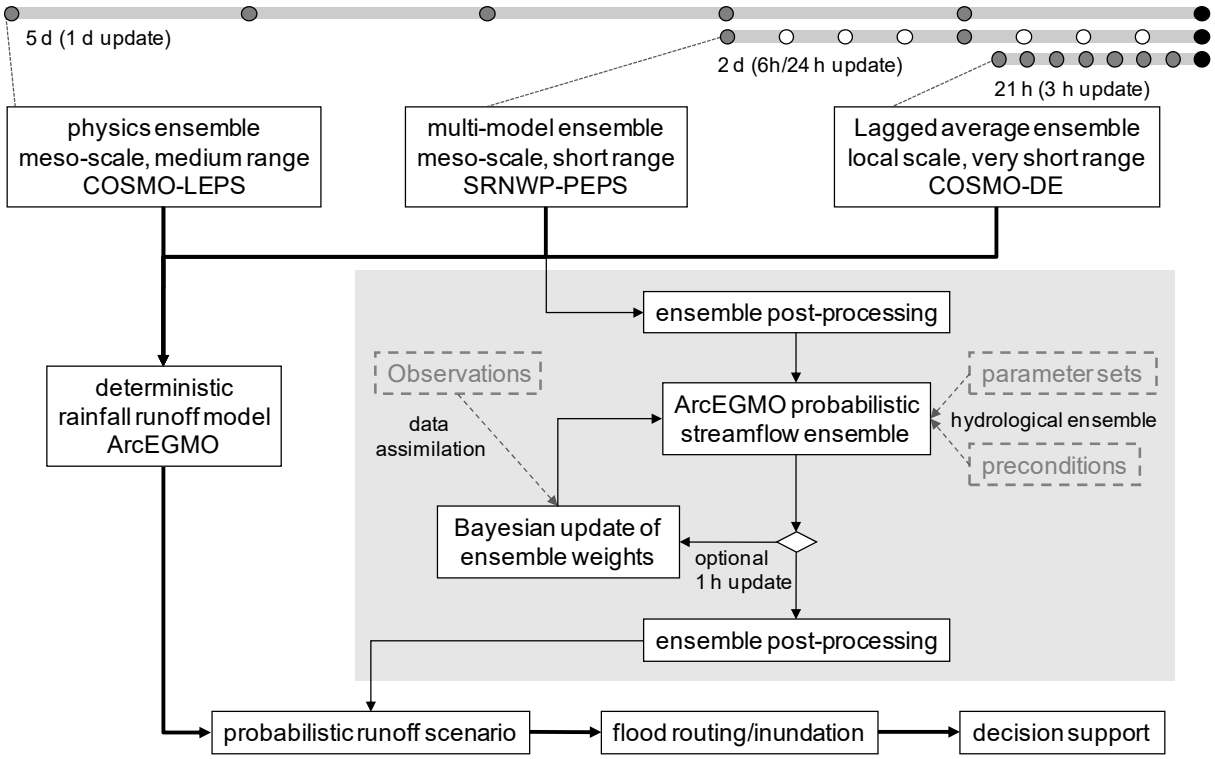


Figure 1: Scheme of an operational flood forecast system with an adaptive combination of meteorological ensemble forecasts from different sources and hydrological models (changed from Dietrich et al., 2008).

Currently the standard mode does not include ensemble post-processing. Thus the probabilistic evaluation is merely based on a relative frequency approach here. There are more advanced techniques available, which may improve the probabilistic evaluation of the ensemble forecasts and therefore provide a better picture of uncertainty. The most important methods are the generation of closed probability distributions (e.g. by using probability density functions or kernel density estimates) and the calibration of ensembles (e.g. by computing weights of the ensemble members and/or by correcting the bias). The shortcoming of these methods is the introduction of more parameters and the need of training against observations. In the case of

heavy or extreme events, there are not a sufficient number of situations with forecast and observation available.

The extended modules of the OFMS are subject of ongoing research and will be published in subsequent papers. These modules will support the generation of hydrological ensembles, the calibration and post-processing of ensembles and a more advanced probabilistic evaluation (grey area in Fig. 1). Hydrological ensembles were designed to regard the uncertainty of the initial state and the parameters of the model (Dietrich et al., 2008). The members of the hydrological ensemble, respectively alternative system states and parameter sets of the model, are selected by inference of antecedent precipitation, the type of the expected event and efficient parameter sets obtained by model calibration against past flood events. In the extended mode the forecast can be updated in 1 - 3 hourly intervals including assimilation of observed rainfall and discharge data. More details and a case study for this option will be presented in a subsequent paper.

Case study

The upper Mulde river basin is situated in the Ore Mountains (Germany and Czech Republic, Fig. 2). Narrow and steep valleys cause a fast reaction of the watershed and critical superimposition of flood waves. Several cities are located in the flood plain of the lower Mulde river basin. During west-cyclonic rainfall events, which caused several extreme flood events in the past, the uncertainty of precipitation forecasts in location, time and volume is crucial. Thus the reliability of flood alerts is an issue of concern. After a disastrous flood event in August 2002 local authorities reconsidered and redirected flood protection and related disaster management in Saxony (Socher and Böhme-Korn, 2008). The development of the flood forecast scheme presented here was accompanied by local flood managers. The result will be partially included into the operational flood forecast system of the State Flood Center of Saxony.

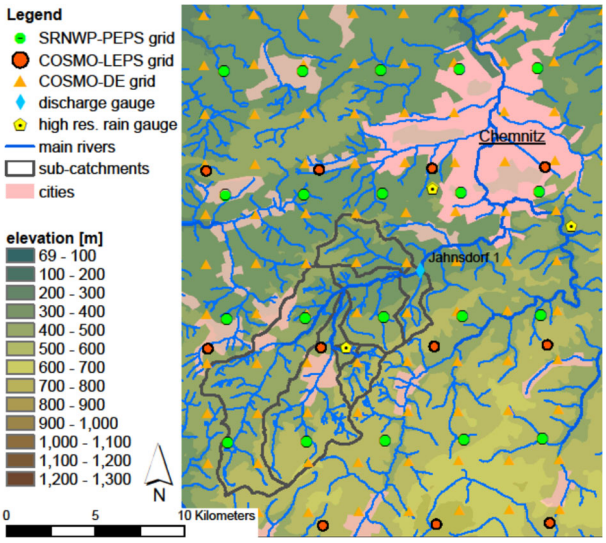


Figure 2: Study area. The upper Würschnitz catchment (Jahnsdorf 1 gauge) and the Chemnitz city downstream are highlighted.

The time period covered by regular operation of the ensemble forecast systems is small (COSMO-LEPS operational from 2005, SRNWP-PEPS operational from 2004, COSMO-DE operational from 2007). Hindcast simulations are necessary. A hindcast (reforecast) is a

prediction for a date in the past using the prediction system that is currently operational. Hindcasts were performed with the COSMO-LEPS and COSMO-DE systems for the disastrous 2002 Elbe flood. The 2006 COSMO-DE forecasts are obtained from the pre-operational system. Within this paper we present results from simulations using the OFMS prototype for seven periods within the years 2002 to 2008. A complete list of the events can be found in Table 2 below.

The hydrological model ArcEGMO was calibrated against observed data for the case study. For the Würschnitz/Chemnitz case study presented in this paper, ArcEGMO proved to be very robust with a default parameter set, which was obtained from calibration and validation of historic flood events from 1954 to 2008. Thus we focused on the impact of the uncertainty of quantitative precipitation forecast on flood alerts (OFMS standard mode). The probabilities for exceeding flood alert levels were obtained by evaluating the frequency of the exceedance in the stream flow forecast ensemble.

COSMO-LEPS hindcasts of extreme events

First we compare four medium range COSMO-LEPS forecasts of heavy or extreme events: the disastrous flood in August 2002, two false alerts in summer 2005 and a very uncertain situation in August 2006. The very short range COSMO-DE forecast for the 2002 flood has been shown and discussed in Dietrich et al. (2008).

COSMO-LEPS hindcasts (10 ensemble members) from 07/08/2002 to 12/08/2002 (one forecast model run per day) have been used to force the hydrological ArcEGMO model. The results of the hydrological simulations are ensembles of possible future development of discharge at several gauges within the river basin for the next 5 days. Fig. 3 shows a sequence of discharge ensembles for the Jahnsdorf gauge (Würschnitz sub-catchment, 103 km^2).

Between 08/08/ and 11/08/2002 the median and the interquartile range (IQR = 0.75-quantile – 0.25-quantile) significantly enlarged. The flood event was overpredicted for the Würschnitz catchment from 10/08 on, but the ensemble framed reality. Note that the 0.25-quantile over topped alert level 1 at 09/08 and alert level 3 at 10/08 and 11/08. For other catchments in the Eastern part of the Ore Mountains this event was underpredicted (Dietrich et al., 2008).

Within the time period from 2005 to 2008, there were two forecasts in summer 2005, which could have caused false alerts (Fig. 4). In July and August 2005 the ensemble median was close to alert level 1, but spread was very high with single discharge forecasts significantly exceeding the threshold value for the highest flood alert level 4, but a very low 0.25-quantile (Fig. 4). These forecasts were not an issue of concern for local flood managers, because they have not been confirmed by deterministic forecasts for the meso-scale. Within these days, several local storm events occurred, which could not be predicted by the COSMO-LEPS system due to spatial resolution and the missing ability to simulate advection. Unfortunately, COSMO-DE hindcasts were not available for that period.

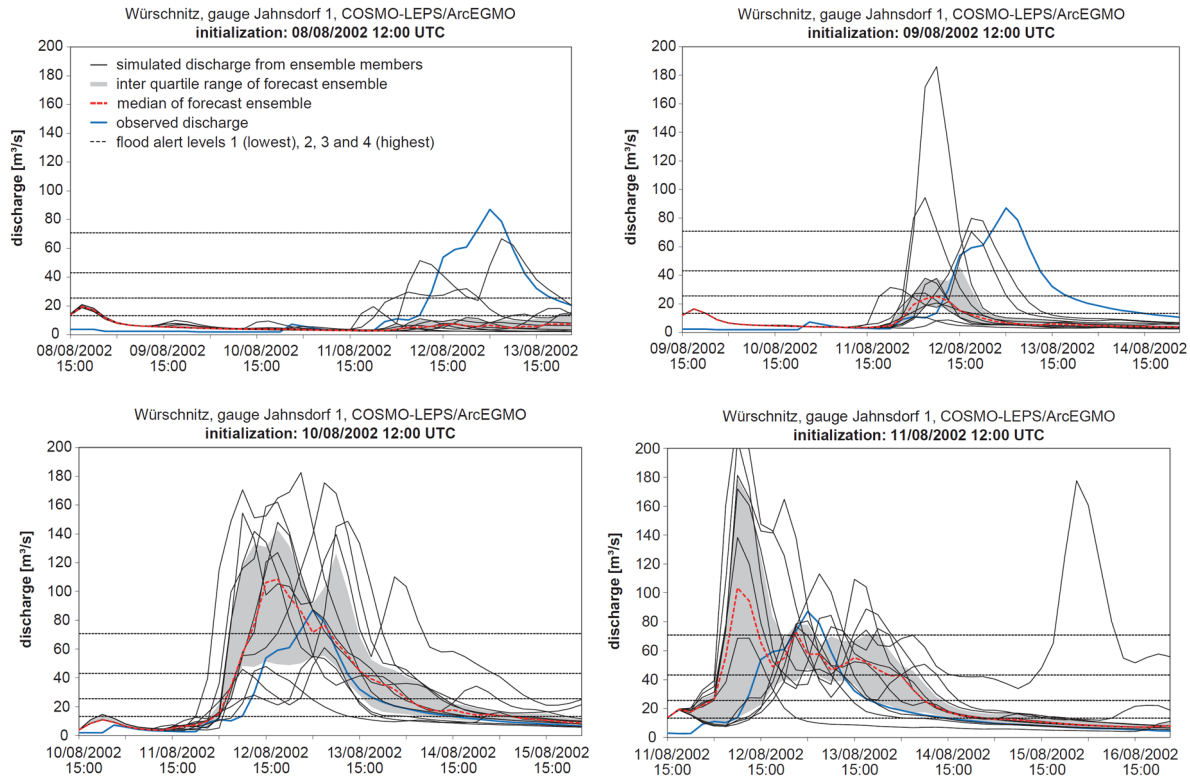


Figure 3: Sequence of discharge forecasts before the disastrous 2002 flood for the Würschnitz catchment (gauge Jahnsdorf). COSMO-LEPS hindcasts were initialized at 08/08, 09/08, 10/08 and 11/08/2002 at 12:00 UTC and processed by the hydrological model ArcEGMO.

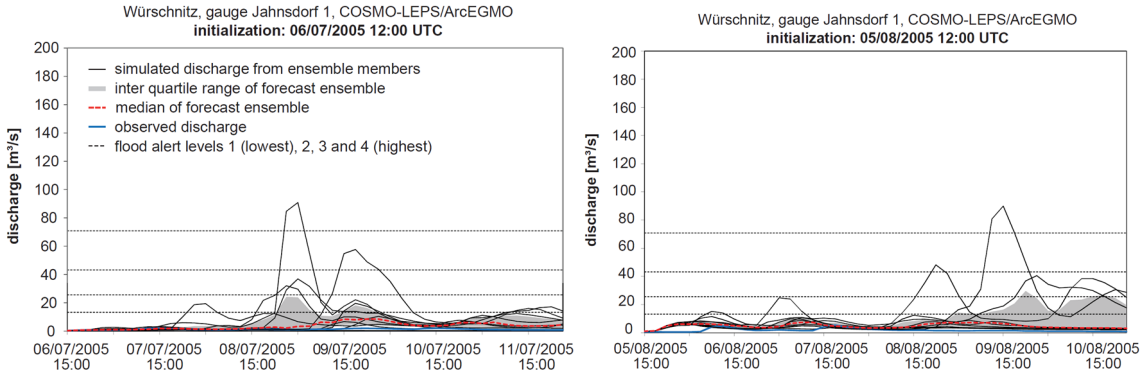


Figure 4: COSMO-LEPS false alarms from July 2005 and August 2005 (legend from Fig. 3, but showing flood alert level 1). In both cases alert level 1 has not been exceeded in reality.

The COSMO-LEPS forecasts issued at 03/08/2006 gave relatively strong evidence for a major flood event within the following two days. Ensemble spread was high again, but many of the COSMO-LEPS members indicated a flood peak between alert level 3 and 4 (Fig. 5). The synoptic forecasters at that time even issued a heavy rainfall warning with a "probability of 100 mm/day rainfall" within two consecutive days for the upper Mulde river basin. In reality the total areal rainfall amount of the 2006 event was approx. 70 mm. The resulting flood peak was below alert level 1 and below a two year return period. The hydrologic reaction was influenced by very low antecedent precipitation. Local flood managers were worried about the dimension of the expected flood and initiated water release from a drinking water reservoir with additional flood protection function in the upper Mulde basin. This decision was right seen from the deterministic forecast as well as from the ensemble forecast: there was considerable uncertainty with the potential of an

extreme flood and reservoir release must start several days in advance. At 05/08 this early flood defence measure was stopped after the forecasted rainfall event did not occur. The rainfall at 06/08 helped to reduce the loss of drinking water. Thus the costs of the measure were very low compared to the potential damage of the predicted flood and the local stakeholders were not affected by this measure. Issuing a flood pre-alert was over cautious at this time. The ensembles gave some evidence, that the situation may be over estimated. The false alert adds to the number of false alerts, which are accepted by the stakeholders before they lose confidence in the flood warning systems.

Comparing the three COSMO-LEPS forecasts with 2 – 5 days lead time shown in Fig. 3, 4 and 5, the 0.25 quantile (Q25) is significantly higher in the 2002 forecasts than in all false alerts. The median of the discharge ensemble would have been a good predictor for the 2002 flood (the exceedance of flood alert level 4 was predicted), but not for the 2006 false alarm (though it gave better results than the synoptic forecast, Table 1). The limited number of available events is not sufficient to draw decision rules. But there is evidence that the number of ensemble members with low precipitation (e.g. Q25) may give more information about uncertainty than the relatively often occurring outliers do.

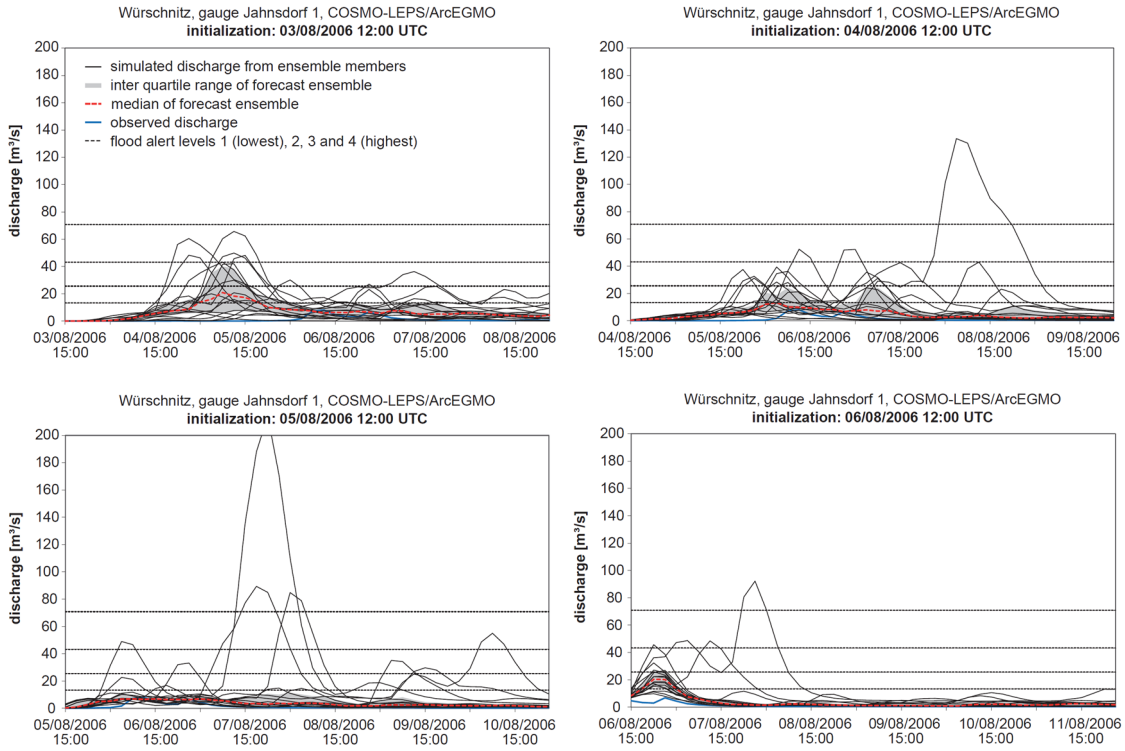


Figure 5: Sequence of discharge forecasts from August 2006 for the Würschnitz catchment (gauge Jahnsdorf). COSMO-LEPS forecasts were initialized at 03/08, 04/08, 05/08 and 06/08/2006 at 12:00 UTC and processed by the hydrological model ArcEGMO (model time step 3 h).

Table 1: Cumulated rainfall (in mm) for the two events in Aug. 2002 and Aug. 2006 from observation (areal rainfall for the Golzern gauge), synoptic forecast for the Mulde river basin and COSMO-LEPS forecast for the area of the Mulde river basin.

event period	observed	synoptic forecast	COSMO-LEPS
2002 (10.8.-13.8.)	192	75 (median), 130 (Q90) at 12/08.	143 (median), 194 (Q90) at 10/08
2006 (3.8.-6.8.)	66	135 (median), 290 (Q90/worst) at 04/08	100 (median), 148 (Q90) at 03/08

Reliability of flood alert level forecasts forced by COSMO-LEPS, SRNWP-PEPS and COSMO-DE

The OFMS scheme offers an adaptive forecast strategy combining medium-range forecasts from COSMO-LEPS as shown above with short-range forecasts from SRNWP-PEPS and deterministic very short-range forecasts from COSMODE. Table 2 summarizes the daily model runs from the three systems (one run of COSMO-LEPS and SRNWP-PEPS, eight runs of COSMO-DE each day) for seven flood event periods from 2002 to 2008. For each system we show the maximum observed discharge (maxObs) at gauge Jahnsdorf within the complete lead time of the meteorological forecasts (121,148 and1132 h, respectively). For COSMO-DE we show the maximum discharge forecasted by the deterministic model run (maxDet). Furthermore we compute a lagged average ensemble and evaluate the 121-h maximum of the median of the seven forecasts available for each hourly time step (maxMed). For SRNWP-PEPS and COSMO-LEPS the maximum of the 0.25 quantile (maxQ25), the maximum of the median (maxMed) and the maximum of the 0.75 quantile (maxQ75) are shown. The colours indicate the observed (first column of each system) and the flood alert levels forecasted by the quantile predictors.

Compared to the results of the ArcEGMO flood forecasts forced by COSMO-LEPS, the performance of the forecasts forced by COSMO-DE at 12/08/2002 was surprisingly weak. The very short-range forecasts did not provide more reliable exceedance probabilities than the medium-range forecasts did. Only the maximum of all available COSMO-DE runs predicted flood alert level 4. The median of a COSMO-DE LAF ensemble only predicted alert level 3 (discharge 56.3 m^3/s compared with observation of 89.9 m^3/s). However, the very short-range forecasts performed better in estimating the peak discharge during the event (COSMO-DE peak error 6 % compared with COSMO-LEPS median 27 %, cf. results for the Zöblitz gauge shown in Dietrich et al., 2008).

At all events from 2005 to 2008, one can see an improvement of the prediction of alert levels with decreasing lead time. SRNWP-PEPS has less often overpredicted the alert levels than COSMO-LEPS. In 2006 the COSMO-DE/ArcEGMO forecasts gave a good estimation of the hydrometeorological situation: only the maximum forecast was above alert level 1. If the lead time of +21 h or +48 h (minus operational time for model run and data delivery) is sufficient for flood managers, they can use COSMO-LEPS to get an early impression of the meteorological situation and evaluate the short time forecasts for confirmation, before they issue an alert.

In all cases the forecast ensembles framed the observations, but not at all lead times. In 2002 and 2008 the COSMO-LEPS with +132 h lead time underestimated the event, but performed reasonably well later. Alert level 1 was only hit with 1 day lead time in 2008. Here the reason is supposed to be the high rainfall intensity of the event, which cannot be sufficiently represented by the 3 h time step of the COSMO-LEPS/ArcEGMO simulations. For all lead times of the COSMO-LEPS forecasts, the 0.35 quantile gave a good estimation of the expected alert levels.

A reliability diagram can be used to assess the probability forecast of a binary event. This diagram plots the observed relative frequency of event occurrence. A complete overview of the reliability of a categorical forecast with a single diagram is not very convenient. Furthermore, a large sample size is needed to produce a meaningful reliability diagram. In the case of severe or extreme flood alerts, there is only a very small sample size if not only one single event. Hamill

(1997) introduced multicategory reliability diagrams (MCRD) to tackle these limitations. The MCRD plots the average percentage of observations below specified quantiles. Like the conventional reliability diagram this graph provides information about the reliability or calibration of a probabilistic forecast.

Table 2: Comparison of the ensemble hindcast simulations for seven time periods using three meteorological prediction systems. The numbers show the peak discharge within the forecast period, the colors show the respective flood alert levels.

EPS	COSMO-DE + 21h			SRNWP-PEPS + 42 h				COSMO-LEPS + 132 h			
Initialization	maxObs	maxDet	maxMed	maxObs	maxQ25	maxMed	maxQ75	maxObs	maxQ25	maxMed	maxQ75
07.08.2002	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	73.5	16.0	16.8	48.9
08.08.2002	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	87.1	19.0	19.4	19.8
09.08.2002	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	87.1	18.3	25.4	46.0
10.08.2002	7.3	10.5	8.2	n.a.	n.a.	n.a.	n.a.	87.1	55.3	108.5	142.7
11.08.2002	26.8	41.8	20.8	n.a.	n.a.	n.a.	n.a.	87.1	47.4	103.5	179.0
12.08.2002	89.9	84.8	56.3	n.a.	n.a.	n.a.	n.a.	87.1	98.8	110.6	116.7
04.07.2005	n.a.	n.a.	n.a.	3.3	2.3	3.5	5.2	0.7	3.1	6.9	8.9
05.07.2005	n.a.	n.a.	n.a.	3.3	3.1	4.0	4.9	0.7	2.9	6.7	17.6
06.07.2005	n.a.	n.a.	n.a.	1.1	1.2	1.5	2.0	0.7	4.7	8.5	24.2
07.07.2005	n.a.	n.a.	n.a.	0.7	1.2	1.4	1.8	0.7	5.3	10.4	18.9
08.07.2005	n.a.	n.a.	n.a.	0.7	1.2	2.1	3.7	0.6	4.1	5.7	9.3
09.07.2005	n.a.	n.a.	n.a.	0.4	1.7	3.2	4.3	0.7	4.4	9.4	15.2
31.07.2005	n.a.	n.a.	n.a.	0.5	1.1	1.1	1.1	4.4	1.0	1.5	5.6
01.08.2005	n.a.	n.a.	n.a.	0.4	0.6	1.0	1.1	4.8	3.5	5.5	8.8
02.08.2005	n.a.	n.a.	n.a.	4.8	1.8	3.1	5.6	5.4	5.9	7.4	14.6
03.08.2005	n.a.	n.a.	n.a.	4.8	4.7	6.6	8.4	5.4	7.0	7.3	7.4
04.08.2005	n.a.	n.a.	n.a.	2.2	8.2	8.4	8.7	5.4	6.0	6.1	9.0
05.08.2005	n.a.	n.a.	n.a.	8.2	7.8	7.8	7.8	5.4	6.0	7.4	29.7
02.08.2006	n.a.	n.a.	n.a.	0.1	0.2	0.4	0.8	8.1	4.0	6.4	10.2
03.08.2006	0.1	0.9	0.1	0.1	0.2	0.4	0.7	8.1	8.3	20.9	40.0
04.08.2006	0.1	1.7	0.2	0.2	5.1	8.6	53.4	8.1	7.4	13.3	24.3
05.08.2006	9.6	15.5	2.6	9.6	2.5	8.2	35.1	8.1	6.5	6.9	11.5
06.08.2006	9.3	21.8	6.2	9.6	3.7	11.2	30.3	6.8	15.9	20.3	26.2
07.08.2006	n.a.	n.a.	n.a.	3.5	2.7	2.7	2.7	1.0	1.1	1.5	2.6
25.05.2007	n.a.	n.a.	n.a.	0.3	0.6	0.9	1.1	3.4	1.3	2.8	6.3
26.05.2007	n.a.	n.a.	n.a.	0.3	2.3	2.3	2.3	3.4	2.5	3.4	13.1
27.05.2007	n.a.	n.a.	n.a.	0.7	2.9	5.0	7.9	3.4	2.5	7.1	17.4
05.11.2007	n.a.	n.a.	n.a.	2.2	3.9	4.5	6.4	11.0	15.4	21.7	27.5
06.11.2007	n.a.	n.a.	n.a.	3.6	5.5	7.9	9.7	19.4	15.4	17.0	20.5
07.11.2007	n.a.	n.a.	n.a.	10.3	6.7	8.0	10.4	19.4	12.3	18.3	26.4
08.11.2007	n.a.	n.a.	n.a.	11.5	13.8	13.8	13.9	19.4	15.8	22.1	26.7
09.11.2007	n.a.	n.a.	n.a.	11.5	10.1	12.2	13.2	19.4	22.0	30.4	33.9
10.11.2007	n.a.	n.a.	n.a.	19.7	12.1	19.7	24.6	19.4	17.7	25.9	29.1
11.11.2007	n.a.	n.a.	n.a.	19.7	17.8	19.4	24.7	19.4	16.3	16.4	16.6
12.11.2007	n.a.	n.a.	n.a.	11.5	10.8	10.8	11.9	10.9	15.3	15.7	17.3
08.04.2008	n.a.	n.a.	n.a.	3.0	3.2	3.2	3.2	17.9	3.6	6.3	11.1
09.04.2008	n.a.	n.a.	n.a.	2.4	2.5	3.1	3.5	17.9	5.5	7.2	13.1
10.04.2008	n.a.	n.a.	n.a.	2.3	2.4	2.4	3.2	17.9	8.0	11.8	16.2
11.04.2008	n.a.	n.a.	n.a.	20.4	6.7	12.2	16.5	17.9	15.5	21.4	27.4
12.04.2008	n.a.	n.a.	n.a.	20.4	23.6	24.4	24.9	8.8	10.2	10.2	10.2
13.04.2008	n.a.	n.a.	n.a.	5.7	7.8	7.8	7.8	4.4	6.0	10.1	15.3
14.04.2008	n.a.	n.a.	n.a.	3.8	6.5	8.2	10.4	3.4	11.0	13.3	16.6
15.04.2008	n.a.	n.a.	n.a.	3.1	6.6	7.6	9.7	2.9	4.3	4.8	5.7

The MCRD in Fig. 6 show the reliability of the probabilistic forecast of the flood alert levels. The latter build up five mutually exclusive and collectively exhaustive categories: no alert, alert 1 (observation), alert 2 (alarm), alert 3 (flood defence, inundation of settlement area) and alert 4 (flood defence, risk of high damage and fatalities). We evaluated observed flood alert levels versus flood alert levels simulated by a hydrological model forced by the respective meteorological EPS. Note that the MCRD does not show information about the reliability of the alert level forecast for a single day. It is integrated over the complete forecast period. We assume that the flood manager is interested if the alert levels are predicted for any time step within the forecast period. That may

explain, why the COSMO-LEPS forecasts with more than 2 days lead time do not become significantly more unreliable with increasing lead time (Fig. 6). However, the +24h COSMO-LEPS is more reliable than the other lead times, but still not as reliable as the SRNWP-PEPS, which has only a small bias.

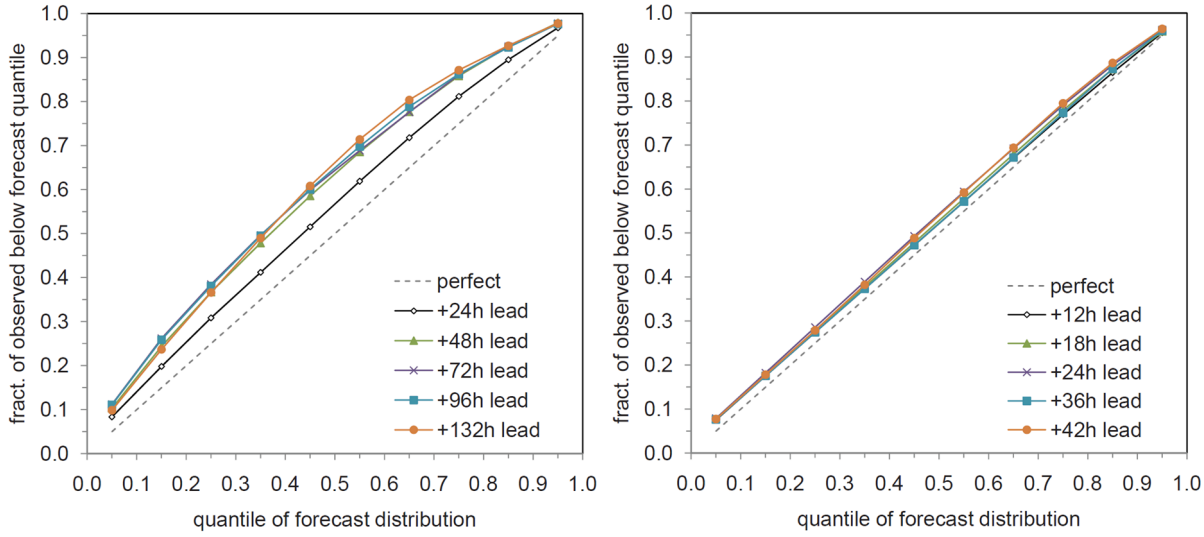


Figure 6: Multicategory reliability diagrams for the flood alert levels simulated with ArcEGMO forced by COSMO-LEPS (left) and the SRNWP-PEPS (right) with different lead times.

The desired persistence of the decision recommendation over time adds another aspect of reliability. Persistence charts can provide a graphical representation of the exceedance probabilities for each time step and for each initialization of the forecast system as predicted by the OFMS (Thielen et al., 2008). Figure 7 shows an example for a relatively stable hydro-meteorological situation: From 6/11/2007 there were evidences for the potential exceedance of alert level 1, which strengthened at 7/11 (probability more than 50 %) and remained until the event occurred at 11/11/2007 (blue box). Nevertheless there have been false signals at the beginning of the period shown in Figure 7.

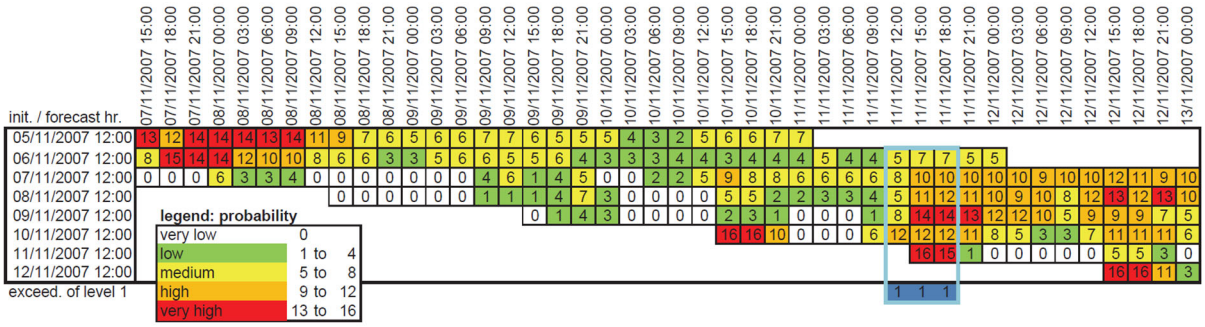


Figure 7: Persistence chart for the COSMO-LEPS/ArcEGMO stream flow forecast in November 2007. The vertical axis displays the initializations (daily), while the horizontal axis displays the time steps of the forecast (3 hourly). The numbers denote the number of ensemble members overtopping alert level 1. The colors show the respective exceedance probabilities (white: very low, green: low, yellow: medium, orange: high, red: very high).

Conclusions

A flood forecast framework combining meteorological ensembles of different spatial and temporal resolution with a calibrated rainfall-runoff model enables the simulation of probabilistic discharge forecasts for meso-scale catchments. Compared with deterministic forecasts, probabilistic forecasts of discharge provide additional information to flood managers, who are addressed by the operational flood management scheme presented in this paper.

The COSMO-LEPS and COSMO-DE hindcasts consistently underpredicted the rainfall rate of the 2002 extreme event. Nevertheless both systems produced evidences for a rainfall event that could cause a flood in the dimension of a 100 year recurrence period. If such a probabilistic forecast would already have been operationally available in August 2002, a reliable flood forecast would have been possible with a lead time of 2-3 days. In August 2006 the combination of three forecast systems with the hydrological model gave evidences for not issuing a flood alert, however uncertainty was very high. Flood management would have been significantly improved compared to the deterministic forecast, which had been issued at that time. The presented ensemble-based framework adds additional value to flood forecasts.

However, the derivation of decision rules, e. g. which exceedance probabilities should be used for issuing alerts or initiating flood defence measures, remains subject of ongoing research because of a lack of hindcasts for severe flood events. From this case study there are some findings to be proved with future case studies:

1. The forecasts from all systems did not always improve with more recent forecasts.
2. For the 'no alert' and 'alert level 1' situations the ensemble flood forecasts proved to be very reliable, even with the raw ensembles.
3. Ensemble means can be useful predictors for flood alert levels. The 0.35 quantile of the COSMO-LEPS was a good predictor for flood alerts in the case study.
4. If the 0.25 quantile over tops alert levels, flood managers should be very cautious and even consider the extreme ensemble members. The persistence of the signal should be observed.
5. There have been only two observed events with alert level 2 and above, including the disastrous 2002 flood. The latter would have been forecasted reasonably well. From this low number of forecasts, the conclusions to be drawn from the MCRD reliability analysis are still rather week for extreme rainfall events. One cannot conclude that the next extreme can be detected as well.

Acknowledgements

The research reported in this paper was carried out with support from the German Ministry for Education and Research (BMBF) under the initiative "Risk Management of Extreme Flood Events (RIMAX)". The authors thankfully acknowledge funding. Bernd Pfützner (Büro für Angewandte Hydrologie BAH Berlin) provided source code and support of the ArcEGMO model. DHI-WASY (Dresden office) was involved in the development of the prototype of the flood management system. Data and practical experience have been contributed by the State Flood Center of Saxony (Dresden), the Saxonian Reservoir authority (Pirna), the Saxonian land survey authority and the German National Meteorological Service (DWD).

References

Anderson, J.L. (1996) A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate* **9**, 1518-1530.

Becker, A., Klöcking, B., Lahmer, W. and Pfützner, B. (2002) The hydrological modelling system ARC/EGMO, in: *Mathematical models of large watershed hydrology* (Eds.: Singh, V.P. and Frevert, D.K.). Water Resources Publications, Littleton/Colorado.

Buizza, R., Houtekamer, P.L., Toth, Z., Pellerin, G., Wei, M. and Zhu, Y. (2005) A comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. *Monthly Weather Review* **133**, 1076–1097.

De Roo, A., Gouweleeuw, B., Thielen, J., Bates, P., Hollingsworth, A. et al. (2003) Development of a European Flood Forecasting System. *International Journal of River Basin Management* **1**(1), 49-59.

Denhard, M. and Trepte, S. (2006) Calibration of the European multi-model ensemble SRNWP-PEPS. *Second THORPEX international science symposium*, WMO/TD No. 1355, WWRP/THORPEX No. 7.

Dietrich, J., Trepte, S., Wang, Y., Schumann, A.H., Voß, F., Hesser, F.B. and Denhard, M. (2008) Combination of different types of ensembles for the adaptive simulation of probabilistic flood forecasts: hindcasts for the Mulde 2002 extreme event. *Nonlinear Processes in Geophysics* **15**, 275-286.

Diomede, T., Davolio, S., Marsigli, C., Miglietta, M. M., Moscatello, A., Papetti, P., Paccagnella, T., Buzzi and A., Malguzzi, P. (2008) Discharge prediction based on multi-model precipitation forecasts. *Meteorol. Atmos. Phys.* **101**, 245–265.

Dos Santos, B. L. and Holsapple, C. W. (1989) A Framework for Designing Adaptive DSS Interfaces. *Decision Support Systems* **5**, 1-11.

Gouweleeuw, B., Thielen, J., Franchello, G., De Roo, A. and Buizza, R. (2005) Flood forecasting using medium-range probabilistic weather prediction. *Hydrol. Earth Syst. Sci.* **9**(4), 365-380.

Hamill, T. M. (1997) Reliability Diagrams for Multicategory Probabilistic Forecasts. *Weather and Forecasting* **12**, 736-741.

Holsapple, C. W., Pakath, R., Jacob, V. S. and Zaveri, J. S. (1993) Learning by problem processors: adaptive decision support systems. *Decision Support Systems* **10**, 85-108.

Kalnay, E. (2002) *Atmospheric modelling, data assimilation and predictability*, Cambridge University Press.

Molteni, F., Buizza, R., Marsigli, C., Montani, A., Nerozzi, F. and Paccagnella, T. (2001) A strategy for high-resolution ensemble prediction, part I: definition of representative members and global model experiments. *Quart. J. Roy. Meteor. Soc.* **127**, 2069-2094.

Reed, S., Schaake, J. and Zhang, Z. (2007) A distributed hydrologic model and threshold frequency-besed method for flash flood forecasting at ungauged locations. *Journal of Hydrology* **337**, 402-420.

Socher, M. and Böhme-Korn, G. (2008) Central European Floods 2002: lessons learned in Saxony. *J Flood Risk Management* **1**, 123–129.

Steppeler, J., Doms, G., Schättler, U., Bitzer, H.W., Gassmann, A., Damrath, U. and Gregoric, G. (2003) Meso-gamma scale forecasts using the nonhydrostatic model LM. *Meteorol. Atmos. Phys.* **82**, 75-96.

Thielen, J., Bartholmes, J., Ramos, M.-H. and de Roo, A. (2008) The European Flood Alert System – Part 1: Concept and development. *Hydrol. Earth Syst. Sci. Discuss.* **5**, 257–287. http://www.hydrol-earth-syst-sci-discuss.net/5/257/2008/.

Toth, Z., Talagrand, O., Candille, G. and Zhu, Y. (2003) Probability and ensemble forecasts, in: Joliffe, I. T. and Stephenson, D. B. (eds.): *Forecast verification: a practitioner's guide in atmospheric science*, John Wiley & Sons.

Verbunt, M., Zappa, M., Gurtz, J., and Kaufmann, P. (2006) Verification of a coupled hydrometeorological modelling approach for alpine tributaries in the Rhine basin. *Journal of Hydrology* **324**, 224-238.

This chapter is an edited version of the following original scientific article:

Dietrich, J.; Denhard, M.; Schumann, A. H. (2009): Can ensemble forecasts improve the reliability of flood alerts? Journal of Flood Risk Management 2 (4), 232–242.

Publisher: John Wiley and Sons

IV. Evaluation of different calibration strategies for large scale continuous hydrological modelling

Wallner, M. [1], Haberlandt, U. [1], Dietrich, J.[1]

(1) Institute of Water Resource Management, Hydrology and Agricultural Hydraulic Engineering, Faculty of Civil Engineering, University Hannover, Germany

Abstract

For the analysis of climate impact on flood flows and flood frequency in macroscale river basins, hydrological models can be forced by several sets of hourly long-term climate time series. Considering the large number of model units, the small time step and the required recalibrations for different model forcing an efficient calibration strategy and optimisation algorithm are essential.

This study investigates the impact of different calibration strategies and different optimisation algorithms on the performance and robustness of a semi-distributed model. The different calibration strategies were (a) Lumped, (b) 1-Factor, (c) Distributed and (d) Regionalisation. The latter uses catchment characteristics and estimates parameter values via transfer functions. These methods were applied in combination with three different optimisation algorithms: PEST, DDS, and SCE. In addition to the standard temporal evaluation of the calibration strategies, a spatial evaluation was applied. This was done by transferring the parameters from calibrated catchments to uncalibrated ones and validating the model performance of these uncalibrated catchments. The study was carried out for five sub-catchments of the Aller-Leine River Basin in Northern Germany.

The best result for temporal evaluation was achieved by using the combination of the DDS optimisation with the Distributed strategy. The Regionalisation method obtained the weakest performance for temporal evaluation. However, for spatial evaluation the Regionalisation indicated more robust models, closely followed by the Lumped method. The 1-Factor and the Distributed strategy showed clear disadvantages regarding spatial parameter transferability. For the parameter estimation based on catchment descriptors as required for ungauged basins, the Regionalisation strategy seems to be a promising tool particularly in climate impact analysis and for hydrological modelling in general.

1 Introduction

The use of hydrological models for answering questions in water resources management is nowadays the technical standard, not only in sciences. In many cases the modeller has to handle catchments on large scales ranging from 100 km^2 to more than 10,000 km^2. On this scales it is not possible to describe the hydrological cycle of this subsystem (catchment) in physical detail (Chow et al., 1988). There is a number of different process oriented models (some are called physically based), whose parameters are closely related to the physical properties of the catchment. These parameters are inherently uncertain. Moreover, the availability of the required data to estimate the parameters is often a problem considering the scales hydrologists are working on. For example it is hardly possible to get a sufficiently detailed description of the soils. In practice this

means that at least some of the parameters have to be estimated via calibration (Beven, 2001) what is done automatically in this study. Taking all parameters of a distributed- or semi-distributed-model into account the dimension Ψ of the parameter search space can be described as:

$$\Psi = N_s \cdot N_p, \tag{1}$$

where N_s gives the number of catchment units (normally subbasins or grid cells) and N_p stands for the number of parameter per unit (Pokhrel and Gupta, 2010; Samaniego et al., 2010). The high dimensional parameter search space leads to the problem of equifinality which means, that there are many different parameter sets which will result in comparable solutions with respect to the model performance (Beven, 2001). Generally the dimension of the search space for the calibration should be reduced, for example by eliminating N_s in Equation 1, in order to make the parameter estimation more robust. Pokhrel and Gupta (2010) achieved good results by reducing the dimension of the calibration problem. Other studies are dealing with the Regionalisation of model parameters, which may lead as well to a more robust model performance. One common approach of Regionalisation is to relate parameters of the model to catchment characteristics. Usually, the hydrological model is calibrated for some selected catchments independently and then the parameters are related to catchment descriptors via (multiple-) regression analysis (Haberlandt et al., 2001; Merz and Bloeschl, 2004). This procedure does not consider the parameter equifinality problem. Another idea is to calibrate parameters for different catchments simultaneously via optimisation of transfer functions which relate the parameters to the catchment characteristics. By reducing the risk of equifinality this approach showed good results in some studies (Hundecha and Bárdossy, 2004; Samaniego et al., 2010). Besides the calibration strategy the selected optimisation method plays an important role in the calibration process of hydrological models. Local search algorithms are fast but are prone to trap into local minima (Abbaspour et al., 2001). One often applied local optimisation tool is PEST (Kim et al., 2007). Sophisticated global search algorithms will more likely find a global optimum (Blasone et al., 2007). Typical candidates are the Shuffled-Complex-Evolution (SCE) method (Duan et al., 1993a) and the Dynamically Dimensioned Search (DDS) algorithm (Tolson and Shoemaker, 2007).

This research is part of a climate impact study. The focus is on long-term continuous hydrological modelling with the well known conceptual model HEC-HMS (Feldman, 2000) at an hourly time step for subsequent flood frequency analyses. For a number of catchments within a macroscale region, several sets of climate data will be used as forcing. Considering the large spatial scale, the small time step and the required recalibrations for different model forcing an efficient calibration strategy and optimisation algorithm are essential. Using a subset of catchments and data the main objective of this study is to find such a strategy together with a suitable optimisation algorithm. In such a framework not only the model performance for single catchments is of interest, but especially the estimation of robust models. Here, the model was considered as robust if its parameters can be transferred to ungauged catchments using a relation between catchment descriptors and model parameters without a significant decrease in model performance. Four different calibration strategies were combined with three different optimisation algorithms. First, a temporal evaluation of all model performances was done via a split sampling of the time series. In a further step a spatial evaluation was carried out, regionalising and applying the calibrated parameters for uncalibrated catchments.

2 Methods

2.1 Hydrological model

The hydrological model used in this investigation was the Hydrological-Modelling-System Version 3.3 from the Hydrologic Engineering Center (HEC-HMS) of the US Army Corps of Engineers (Feldman, 2000). HEC-HMS is a semi-distributed model with horizontal structure realized via subbasins. Figure 1 shows the structure of HEC-HMS (Bennett and Peters, 2000). For runoff generation a Soil Moisture Accounting scheme is used. The runoff concentration of the surface runoff is calculated with Clark´s Unit Hydrograph, the interflow and baseflow are calculated with linear reservoirs. All flows in the channel are routed with the Muskingum method. The calculation of the potential evapotranspiration is carried out with the Priestley-Taylor method, the snow melt with the Temperature-Index method. All meteorological input data were interpolated using Ordinary-Kriging for precipitation, External-Drift-Kriging (elevation as secondary information) for temperature and Inverse-Distance for net radiation. The model was run in a continuous mode on an hourly time step.

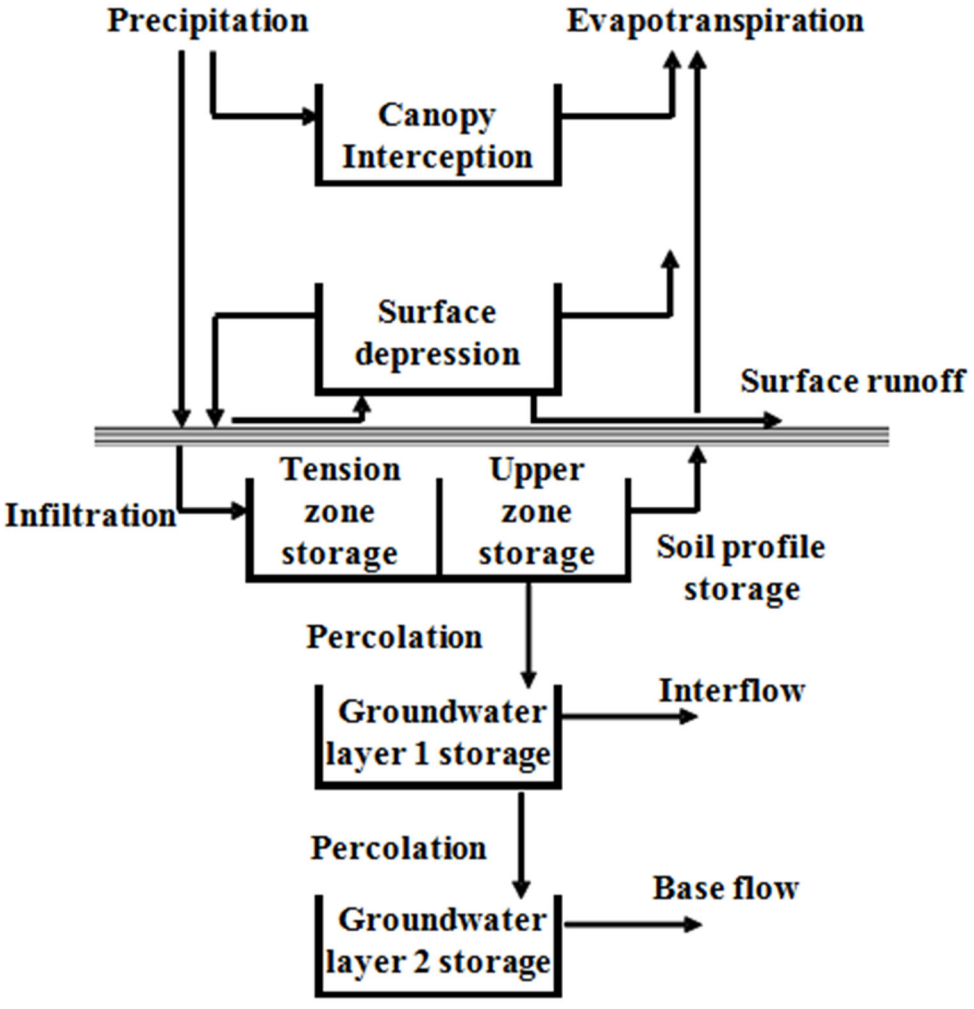


Fig. 1. Structure of the HEC-HMS Model (Bennett and Peters, 2000).

2.2 Optimisation algorithm

Optimisation algorithms can be generally distinguished between local search algorithms and global search algorithms. In this study both types were used, where PEST (Doherty, 1994) is a local optimisation and SCE (Duan et al., 1994) and DDS (Tolson and Shoemaker, 2007) are global optimisation algorithms. PEST uses the Gauss-Marquardt-Levenberg method for for nonlinear parameter estimation. One condition for the application of PEST is that the adjustable parameters must be continuously differentiable (Doherty, 1994). The SCE algorithm builds on four well proved concepts: first the combination of random and deterministic, second the concept of clustering, third the concept of systematic evolution and fourth the concept of competitive evolution. The SCE algorithm was created to solve a broad class of problems such as finding solutions for a multi parameter space which is not even continuous (Duan et al., 1993b). The DDS is a stochastic based global search algorithm. For each iteration a set of parameters is selected which are then perturbed by values randomly sampled from a normal distribution. The number of dimensions decreases with an increasing number of iterations, so that the solver searches more globally at the beginning and more locally at the end of the optimisation procedure (Tolson and Shoemaker, 2007).

2.3 Calibration strategies and validation

There are many different approaches for automatic calibration of watershed models. The simplest one is the use of the same parameter value for all subbasins. This strategy is named Lumped (LUM) parameter estimation here. In Fig. 2 a sketch of a semi-distributed model is shown. For each subbasin the parameters $\Phi_{i,j}$ need to be estimated, where the index i stands for the parameter and the index j for the subbasin. For the Lumped method all particular parameter values are equal:

$$\Phi_{i,1} = \Phi_{i,2} = \Phi_{i,3} = \Phi_{i,4} \;;\; i = 1,\ldots,N_p . \tag{2}$$

This implicates that the dimension of the search space depends only on the number of different parameters N_p. By eliminating the number of subbasins N_s in Eq. (1) the dimension Ψ becomes:

$$\Psi = N_p . \tag{3}$$

Employing the Lumped method as we do here, the spatial variability of the climate forcing is still represented due to the attribution of time series to the subbasins, but the spatial variability of the parameters is lost. To avoid this, the 1-Factor (1-F) method is introduced. Initial parameter values are defined in a pre-processing step based on basin characteristics such as soil, landuse and topography. As a result spatially variable initial values for the parameters are obtained:

$$\Phi_{i,1} \neq \Phi_{i,2} \neq \Phi_{i,3} \neq \Phi_{i,4} \;;\; i = 1,\ldots,N_p . \tag{4}$$

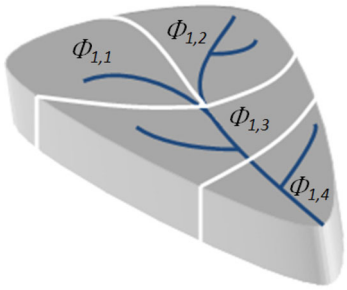


Fig. 2. Typical parameter set up for a semi-distributed model. Every $\Phi_{i,j}$ stands for one parameter value (i = parameter, j = subbasin).

Due to the conceptuality of the hydrological model used in this study the prior parameter estimation does not provide reliable values but is more an indicator of the spatial variability of the parameters. In the next step one factor is created for each parameter. These factors are calibrated and then multiplied with the particular parameter values (Pokhrel and Gupta, 2010). The dimension of the search space is independent of the number of subbasins, like for the Lumped strategy, and Eq. (3) is valid. Both strategies are limiting the variability of the parameter values. Either there is no spatial variability (LUM) or the variability is given by pre-processing which depends on the modeller's subjective assessment of the catchment.

Hence a further Distributed calibration strategy (DIS) is applied which considers all parameters for all subcatchments independently. Since each single parameter value is calibrated, the dimension of the parameter search space is the product of the number of subbasins Ns and the number of parameters N_p (Eq. 1) and is generally much larger than for the other methods.

As last calibration strategy the Regionalisation method (REG) is introduced, assuming that the parameters of the hydrological model can be related to basin characteristics. Each parameter value is calculated via transfer functions which have to be pre-defined and should include the main catchment characteristics affecting the particular parameter. The following structure of such a function is used here:

$$\Phi_{i,j} = \sum_{n=1}^{k} \alpha_{i,n} \cdot S_{j,n} + \sum_{m=1}^{u} \beta_{i,m} \cdot L_{j,m} + \gamma_i \cdot Area_j + \ \ldots ; \ i = 1,\ldots,N_p; \ \ j = 1,\ldots,N_s , \qquad (5)$$

where $S_{j,n}$ and $L_{j,m}$ are relative areas of soil- and landuse classes, respectively and k and u are the number of different soil/- landuse classes defined for each subbasin. It is important to note, that not the parameters themselves are calibrated but the coefficients of the transfer function (($\alpha_{i,n}$, $\beta_{i,m}$, γ_i, …). Further details of this strategy can be found in Hundecha and Bárdossy (2004). The soil classes $S_{j,n}$ were described via fuzzy numbers (Fig. 3). In a first step the main physical descriptor has to be estimated for each parameter (e.g. parameter: maximum infiltration → descriptor: hydraulic conductivity). The estimation of the specific values for these physical descriptors for each soil is based on (Finnern et al., 1994). With these values the fuzzy numbers were set up. For each single soil a membership to the fuzzy numbers can be defined. The upscaling of this information to the subbasin was simply done by averaging. The Regionalisation method allows calibrating an arbitrary number of catchments simultaneously with constant dimension of the parameter search space. It uses physical basin characteristics and hydrological information

from different sites. In addition, this procedure alleviates the equifinality problem in model parameter regionalisation.

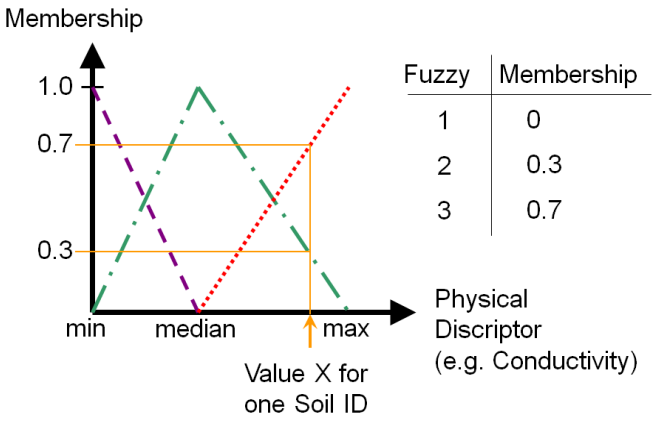


Fig. 3. Example for the definition of the fuzzy numbers for one physical basin descriptor (fuzzy number 1 dashed line; fuzzy number 2 dotted dashed line; fuzzy number 3 dotted line).

Beside the standard evaluation of the calibration strategies via a temporal split sampling, a spatial split sampling was introduced (only for the DDS calibration). This was done by transferring the parameters from calibrated donor catchments to uncalibrated ones and validating the model performance for these uncalibrated catchments. For the Lumped, 1-Factor and the Distributed method, multiple regressions with the structure of the transfer functions (Eq. 5) were fitted on the calibrated catchments. Then for the uncalibrated catchments the parameters were estimated using these regressions and the catchment characteristics (cf. Eq. 5).

To evaluate the model performance three different objective functions were used. The squared error (SqE):

$$SqE = \sum_{t=1}^{N} \left(Q_{Obs}(t) - Q_{Sim}(t)\right)^2 \quad \rightarrow \quad Min\,, \tag{6}$$

the Nash-Sutcliffe efficiency coefficient (NSE) (Nash and Sutcliffe, 1970):

$$NSE = 1 - \frac{\sum_{t=1}^{N} \left(Q_{Obs}(t) - Q_{Sim}(t)\right)^2}{\sum_{t=1}^{N} \left(Q_{Obs}(t) - \overline{Q_{Obs}}\right)^2} \quad \rightarrow \quad Max\,, \tag{7}$$

and the volume-error (VoE):

$$VoE = \frac{\sum_{t=1}^{N} Q_{Obs}(t) - \sum_{t=1}^{N} Q_{Sim}(t)}{\sum_{t=1}^{N} Q_{Obs}(t)} \quad \rightarrow \quad Min\,, \tag{8}$$

where $Q_{Obs}(t)$ and $Q_{Sim}(t)$ are observed and simulated discharge for each time step, respectively, $\overline{Q_{Obs}}$ is the mean observed discharge and N is the number of time steps. To combine the NSE and the VoE a multi criteria objective function (ObjFunc) was implemented:

$$ObjFunc = \sqrt{(1 \cdot (1 - NSE))^2 + (1.4 \cdot (0 - VoE))^2} \quad \rightarrow \quad Min, \tag{9}$$

giving more weight on the volume error.

3 Study area and data

The Aller-Leine river basin covers most of the south-eastern part of Lower Saxony in Germany (Fig. 4). For this investigation five subcatchments with a size between 300 km^2 and 1000 km^2 and with subbasin sizes from 20 km^2 to 40 km^2 were chosen. Some of the basin characteristics are given in Table 1. The climate data for the Aller-Leine catchment (provided by the German Weather Service DWD and Meteomedia) include 100 precipitation stations with a high temporal resolution (≤ 1hr), 244 precipitation stations with a daily resolution, 38 temperature stations with a high temporal resolution (≤ 1hr) and 20 other stations for which net radiation was calculated on a daily time step. However, from the large number of recording precipitation stations only 11 stations had an observation period of more than 10 years. The soil map (BüK 1000) with a scale of 1:1,000,000 was provided by the Federal Institute for Geosciences and Natural Resources (BGR), as land cover map CORINE 2000 was used and a digital elevation model (DEM) with a resolution of 10 x 10 meters was provided by the Lower Saxony Water Management, Coastal Defence and Nature Conservation Agency (NLWKN). Discharge time series were obtained for streamflow gauges at the outlets of the five catchments in high temporal resolution (≤ 1hr). Three of the catchments with gauges: Derneburg, Reckershausen and Wieckenberg, were used for temporal split sampling and all five, including the other two with gauges: Leineturm, Glentorf, for spatial split sampling. The whole study was based on continuous rainfall runoff simulations with an hourly resolution. The calibration period was three years (2004/2007/2008), the validation period two years (2005/2006). The first year (2003) was used as spin-up period to determine the initial conditions for the model.

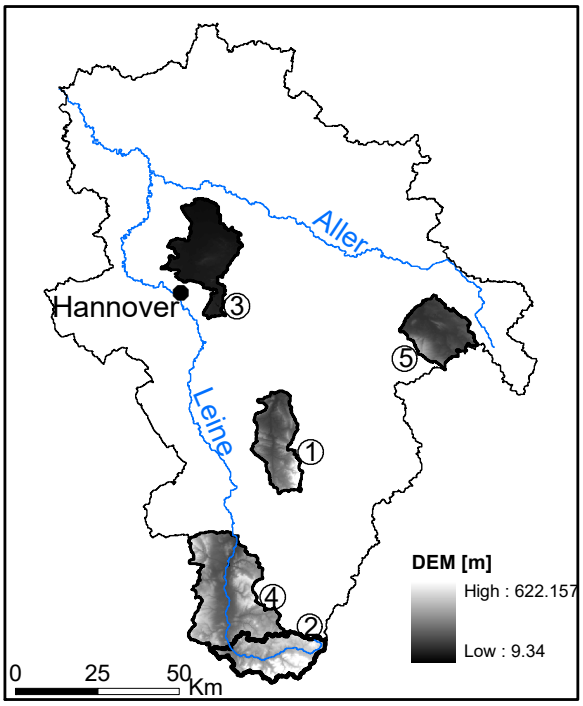


Fig. 4. Aller-Leine River Basin with catchments used for hydrological simulations. The indicated numbers correspond to the catchment ID´s in Table 1.

Table 1. Characteristics of the catchments (mean values for precipitation, temperature and runoff).

ID	Catchment /Gauge	Area [km^2]	Elevation [m]	Pcp. [mm/yr]	Temp. [°C]	Runoff [m^3/s]	No. of Subbasins
1	Nette/Derneburg	309	206	872	9.29	2.8	10
2	Leine/Reckershausen	319	340	771	8.61	2.3	10
3	Wietze/Wieckenberg	411	50	722	10.20	2.0	11
4	Leine/Leineturm	990	272	752	8.95	7.8	36
5	Schunter/Glentorf	290	137	727	9.56	1.5	10

4 Analysis and results

4.1 Evaluation by temporal split sampling

In addition to the evaluation of the model performance, the efficiency of the different combinations of calibration strategies and optimisation algorithms were investigated. This was done by limiting the number of iterations to 1000 for each catchment and comparing the model performance after applying the different optimisation algorithms. In a first step the squared error (Eq. 6) was used as objective function for the calibration. For the Regionalisation method the standardised objective function (Eq. 9) was implemented. In Fig. 5 the Nash-Sutcliffe efficiency and the volume error (VoE) are illustrated for three catchments, calibrated with the Lumped calibration strategy in combination with the PEST optimisation algorithm. Comparing calibration and validation performances, for all catchments a small decrease of the NSE can be recognised. But the VoE for Derneburg stays nearly the same, and for Reckershausen it even increases. To evaluate these results for all combinations of calibration methods the average of the objective function (Eq. 9) for all three catchments, Derneburg, Reckershausen and Wieckenberg, was calculated and subtracted from 1.00. The closer the results are to 1.00, the better is the model performance.

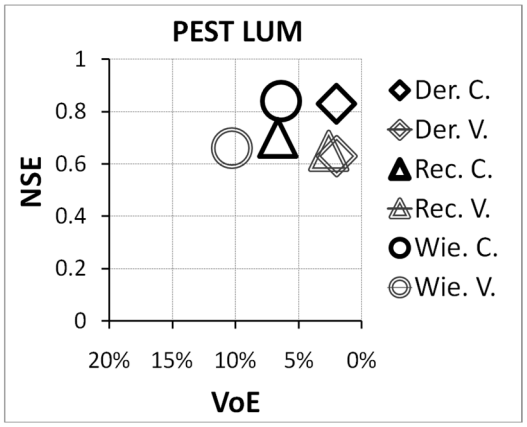


Fig. 5. Model performance for the Lumped calibration strategy in combination with the PEST optimisation algorithm. The ideal point with an NSE = 1 and a VoE = 0 is shown in the top right corner. (Der. = Derneburg; Rec. =Reckershausen; Wie.= Wieckenberg; C. = calibration (years 2004,2007-2008); V. = validation (years 2005-2006)).

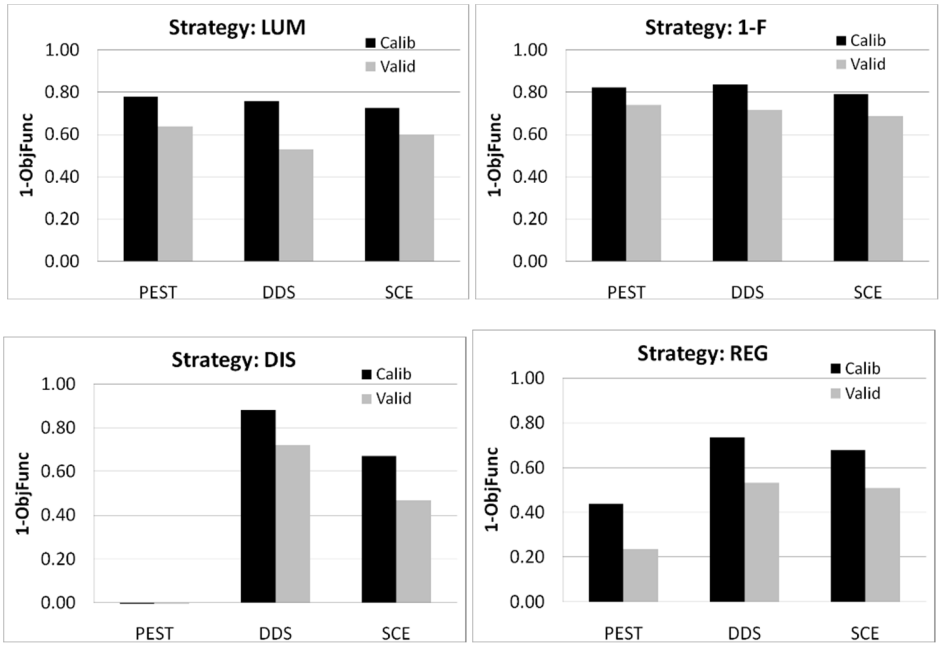


Fig. 6. Mean model performance calculated as 1.00 minus the average over the objective function values (Eq. 9) for the three catchments, Derneburg, Reckershausen and Wieckenberg for all combinations of calibration strategies and optimisation algorithms. The optimum result would be 1.00.

Figure 6 illustrates the results for all combinations. In all cases the performance of the validation period decreased compared to the calibration period. The Lumped method as well as the 1-Factor method showed good results for all three optimisation algorithms in the calibration period. Considering that both methods handle the same dimension of the parameter search space (42 dimensions for all three catchments) the slightly better performance of the 1-Factor method might be due to a good estimation of the spatial variability of the initial parameter set. For the Lumped and 1-Factor method there were no significant differences in performance between the different optimisation algorithms. However, considering the Distributed and Regionalisation methods it seemed that the PEST algorithm is not able to solve these more complex problems. In this study it was not possible to obtain acceptable results using PEST with those methods. In both cases the DDS algorithm outperformed the SCE algorithm for calibration and validation. An

indicator of the robustness of the model is the difference in performance between calibration and validation, where the 1-Factor strategy showed the best results. The overall best performance was obtained using the Distributed strategy, calibrated with the DDS algorithm. So far the Regionalisation method led to the least performance of all calibration strategies. This may be attributed to the following reasons: the development of the transfer function used for the Regionalisation is just in its first phase and different catchments were calibrated simultaneously which sets more restrictions on the parameter estimation. The parameter dimensions of the Regionalisation strategy were 33 whereas the Lumped and the 1-Factor method had 42 dimensions and the Distributed method even had 416 dimensions (as sum over all calibrated catchments). All results are listed in Table 2.

Table 2. Nash-Sutcliff efficiency and volume error for the calibration and validation of the different catchments and the different optimisation algorithms (gauges: Der. = Derneburg; Rec. = Reckershausen; Wie. = Wieckenberg). The calibration years are 2004, 2007 & 2008. The validation was done for 2005 & 2006.

		Lumped				1-Factor				Distributed				Regionalisation			
		Calib.		Valid.		Calib.		Valid.		Calib.		Valid.		Calib.		Valid.	
	Gauge	NSE	VoE	NSE	VoE	NSE	VoE	NSE	VoE	NSE	VoE	NSE	VoE	NSE	VoE	NSE	VoE
PEST	Der.	0.83	0.02	0.63	0.02	0.82	0.09	0.70	0.02	0.75	0.19	0.56	0.17	0.49	0.38	-0.22	0.17
	Rec.	0.70	0.07	0.65	0.03	0.78	0.01	0.71	0.00	-2.71	0.17	-1.65	0.06	0.50	0.14	0.51	0.09
	Wie.	0.84	0.06	0.66	0.10	0.90	0.03	0.84	0.09	-0.11	0.00	-0.41	0.07	0.52	0.05	0.45	0.01
DDS	Der.	0.77	0.06	0.31	0.05	0.81	0.01	0.71	0.01	0.86	0.05	0.69	0.05	0.69	0.10	0.37	0.18
	Rec.	0.69	0.02	0.61	0.02	0.79	0.02	0.69	0.00	0.86	0.03	0.67	0.02	0.71	0.01	0.62	0.02
	Wie.	0.84	0.07	0.69	0.09	0.91	0.01	0.82	0.15	0.94	0.01	0.82	0.07	0.83	0.00	0.65	0.07
SCE	Der.	0.76	0.08	0.68	0.14	0.75	0.12	0.64	0.10	0.70	0.24	0.41	0.22	0.70	0.12	0.57	0.21
	Rec.	0.67	0.02	0.60	0.02	0.75	0.01	0.67	0.03	0.64	0.07	0.57	0.08	0.65	0.01	0.53	0.02
	Wie.	0.77	0.04	0.59	0.13	0.91	0.00	0.83	0.13	0.80	0.04	0.53	0.17	0.72	0.04	0.50	0.06

4.2 Evaluation by spatial split sampling

Beside the common temporal evaluation of model performance criteria a special focus was given to the spatial transferability of parameter values, described in Sect. 2.3. Therefore the parameters of the three donor catchments, calibrated with the DDS algorithm, were transferred to the two validation catchments. Due to some changes in the transfer functions the results of the Regionalisation method in this chapter are not directly comparable with those from Sect. 4.1. The upper diagram of Fig. 7 illustrates the model performance, averaged for the three calibration respectively the two validation catchments, in the calibration period. Although the Regionalisation method showed the poorest performance for the calibration catchments, this method outperformed the other ones considering the validation catchments. Surprisingly, the Lumped strategy showed good results as well. Similar results were obtained for the validation period (lower diagram in Fig. 7). The Lumped as well as the Regionalisation method showed a better model performance for the validation catchments than the other two strategies. As mentioned before the transferability of parameters is a clear indicator of model robustness. Comparing the 1-Factor and the Lumped method the results differed considerably from each other. This might be explained by the transfer of a locally estimated spatial distribution of the parameter set for the 1-Factor method, which is not necessarily representative for other catchments.

In the spatial evaluation, the performance of the Distributed method noticeably decreased compared to the temporal evaluation to. This can be explained by the high dimensional parameter search space, which allowed a most flexible fitting of the parameters to the hydrograph of the calibration catchment, but is prone to the problem of equifinality.

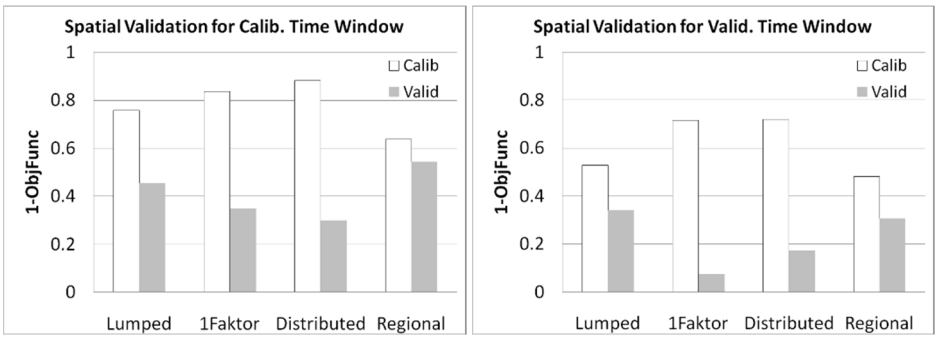


Fig. 7. Mean model performance calculated as 1.00 minus the average over the objective function values (Eq. 9) of the respective catchments for the spatial evaluation in the calibration period (left figure) and in the validation period (right figure). The white bars show the mean performance for the three calibrated catchments (gauges: Derneburg, Reckershausen, Wieckenberg) and the grey bars represent the mean performance for the two validation catchments (gauges: Leineturm, Glentorf).

5 Conclusions

This study focused on the comparison of different calibration strategies in combination with different optimisation algorithms. It was shown, that depending on the complexity of the optimisation problem, the performance of different optimisation algorithms can vary significantly. The local search algorithm PEST found good solutions for the simple calibration strategies, Lumped and 1-Factor. With increasing complexity, the performance of PEST significantly decreased. The DDS algorithm slightly outperformed the SCE algorithm. The restriction of 1000 iterations per catchment certainly plays an important role for this result. Both global optimisation algorithms showed that, even with this tough restriction, they were able to solve complex problems with an adequate performance.

Comparing the calibration strategies purely for temporal evaluations, the Regionalisation was the method with the least performance. Taking the spatial evaluation into account the Regionalisation method indicated the most robust models, closely followed by the Lumped method.

It is important to find a general way to define parameters for hydrological models. This can be done via catchment classification, which can assist the regionalisation of the parameters. Therefore the dominant controls of catchment structures must be understood (Wagener et al., 2007). Further work is necessary regarding the improvement of transfer functions, the optimal selection of catchment descriptors, the suitable definition of multiple criteria objective functions and applications to a larger number of catchments.

Acknowledgements. The authors thank all supporters who gave comments to this work. Particularly we want to thank our student assistant Felipe Teixeira, furthermore Dr. Brian Tolson for some advices in using the DDS and Dr. William Scharffenberg for his support using the HEC-HMS model. We want to thank DWD and Meteomedia for providing precipitation time series and NLWKN for providing runoff time series. The

study was supported by the Ministry for Science and Culture of Lower Saxony within the network KLIFF – climate impact and adaptation research in Lower Saxony.

References

Abbaspour, K. C., Schulin, R., and Genuchten, M. T. V.: Estimating unsaturated soil hydraulic parameters using ant colony optimization: Adv. Water Resour., 24, 827–841, 2001.

Bennett, T. H. and Peters, J. C.: Continuous Soil Moisture Accounting in the Hydrologic Engineering Center Hydrologic Modeling System (HEC-HMS), edited by: Rollin, H. H. and Michael, G., Vol. 104, ASCE, 149–159, 2000.

Beven, K. J.: Rainfall-Runoff Modelling: The Primer, John Wiley & Sons, Ltd, 360 pp., 2001.

Blasone, R.-S., Madsen, H., and Rosbjerg, D.: Parameter estimation in distributed hydrological modelling: comparison of global and local optimisation techniques, Nord. Hydrol., 38, 451–476, 2007.

Chow, V. T., Maidment, D. R., and Mays, L. W.: Applied Hydrology, McGraw-Hill Book Company, 572 pp., 1988.

Doherty, J.: PEST – Model-Independent Parameter Estimation User Manual: 5th Edition User Manual: 5th Edition, Watermark Numerical Computing, 1994.

Duan, Q. Y., Gupta, V. K., and Sorooshian, S.: Shuffled Complex Evolution Approach for Effective and Efficient Global Minimization, J. Optimiz. Theory App., 76, 501–521, 1993a.

Duan, Q., Sorooshian, S., and Gupta, V. K.: Optimal use of the SCEUA global optimization method for calibrating watershed models, J. Hydrol., 158, 265–284, 1994.

Feldman, A. D.: Hydrological Modeling System HEC-HMS – Technical Reference Manual, US Army Corps of Engineers, 2000.

Finnern, H., Grottenthaler, W., K¨uhne, D., Pl¨achen, W., Schraps, W.-G., and Sponagel, H.: Bodenkundliche Kartieranleitung, Bundesanstalt f¨ur Geowissenschaften und Rohstoffe und den Geologischen Landesämtern in der Bundesrepublik Deutschland, 1994.

Haberlandt, U., Klöcking, B., Krysanova, V., and Becker, A.: Regionalisation of the base flow index from dynamically simulated flow components – a case study in the Elbe River Basin, J. Hydrol., 248, 35–53, 2001.

Hundecha, Y. and B´ardossy, A.: Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model, J. Hydrol., 292, 287–295, 2004.

Kim, S. M., Benham, B. L., Brannan, K. M., Zeckoski, R. W., and Doherty, J.: Comparison of hydrologic calibration of HSPF using automatic and manual methods, Water Resour. Res., 43, W01402, doi:10.1029/2006WR004883, 2007.

Merz, R. and Bloeschl, G.: Regionalisation of catchment model parameters, J. Hydrol., 287, 95–123, 2004.

Nash, J. E. and Sutcliffe, I. V.: River Flow Forecasting through Conceptual Models - Part I - A Discussion of Principles, J. Hydrol., 10, 282–290, 1970.

Pokhrel, P. and Gupta, H. V.: On the use of spatial regularization strategies to improve calibration of distributed watershed models, Water Resour. Res., 46, W01505, doi:10.1029/2009WR008066, 2010.

Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resour. Res., 46, W05523, doi:10.1029/2008WR007327, 2010.

Tolson, B. A. and Shoemaker, C. A.: Dynamically dimensioned search algorithm for computationally efficient watershed model calibration: Water Resour. Res., 43, W01413, doi:10.1029/2005WR004723, 2007.

Wagener, T., Sivapalan, M., Troch, P., and Woods, R., 2007, Catchment Classification and Hydrologic Similarity: Geography Compass, v. 1/4, p. 901–931.

This chapter is an edited version of the following original scientific article:

Wallner, M.; Haberlandt, U.; Dietrich, J. (2012): Evaluation of different calibration strategies for large scale continuous hydrological modelling. Advances in Geosciences 31, 67–74.

Publisher: Copernicus Publications

V. Validierung ausgewählter Klimamodelldaten als Basis für die Interpretation von wasserwirtschaftlichen Klimafolgenabschätzungen in Niedersachsen

Petry, U.[1], Dietrich, J.[2], Förster, K.[3], Wallner, M.[4], Berndt, C.[2], Meon, G.[5], Haberlandt, U.[2]

(1) Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz, Hildesheim
(2) Institut für Wasserwirtschaft, Hydrologie und landwirtschaftlichen Wasserbau, Leibniz Universität Hannover
(3) alpS – Centre for Climate Change Adaptation, Innsbruck, Österreich
(4) Bundesanstalt für Geowissenschaften und Rohstoffe (BGR), Hannover
(5) Leichtweiß-Institut für Wasserbau, Abt. Hydrologie, Technische Universität Braunschweig

Kurzfassung

Klimamodelldaten werden als Eingangsdaten für Impaktmodelle eingesetzt, um die Folgen eines sich wandelnden Klimas zu simulieren. Die Bewertung der Verlässlichkeit dieser Ergebnisse ist ein wichtiger Punkt, da sie eine Grundlage für Anpassungsmaßnahmen darstellen. Ein erster Schritt zur Identifikation möglicher Unsicherheiten ist die Validierung der Klimamodelldaten. In dieser Studie wurde über einen einfachen Ansatz mittels statistischer Verfahren die Güte von Klimamodelldaten abgeschätzt und ein Maß für die Beurteilung der Eignung dieser Daten für ausgewählte hydrologische Fragestellungen in der Impaktmodellierung vorgestellt. Hierzu wurden beispielhaft die Daten eines statistischen regionalen Klimamodells (WETTREG2006) sowie von zwei dynamischen regionalen Klimamodellen (REMO und CLM), alle angetrieben durch das Globalmodell ECHAM5/MPI-OM, hinsichtlich ihrer Fähigkeit untersucht, die beobachteten Temperatur- und Niederschlagsverhältnisse nachzubilden. Die Validierung erfolgte auf Basis interpolierter Gebietsmittel von Tagesmittelwerten für den Zeitraum von 1961-2000. Das Untersuchungsgebiet umfasste das Einzugsgebiet von Aller und Leine (Niedersachsen) sowie neun mesoskalige Teileinzugsgebiete innerhalb des Gesamtgebietes. Die Auswertung erfolgte anhand verschiedener statistischer Kennwerte und Gütekriterien. Bei den dynamischen Modellen zeigten sich Abweichungen gegenüber dem Beobachtungsdatensatz insbesondere bei der Temperatur und den Trockenwetter-Indizes. WETTREG hingegen zeigte bei allen Kennwerten erwartungsgemäß eine tendenziell gute Übereinstimmung. Die Verwendung der Klimamodelldaten in der hydrologischen Modellierung der Untersuchungsgebiete ergab für die simulierten Abflüsse im Mittel vergleichbare Modellgüten wie bei den korrespondierenden Kenngrößen des Niederschlags. Hohe wie niedrige Abflussverhältnisse wiesen jedoch eine vergrößerte Bandbreite der Güte auf. Mit Hilfe des Wilcoxon-Mann-Whitney-Tests konnte für ein Signifikanzniveau von 95% gezeigt werden, dass die betrachteten Modellketten geeignet erscheinen, um wasserwirtschaftliche Klimafolgenabschätzungen vorzunehmen. Die Darstellung extremer Verhältnisse, vor allem bei Niedrigwasser, sollte jedoch stets in Abhängigkeit des Modells bewertet werden, um mögliche Entscheidungen über Anpassungsmaßnahmen abwägen zu können.

Abstract

Climate projections are used as input for impact models to simulate the effects of a changing climate for various applications. Assessing the reliability of these results is important because of their use for climate change adaptation. A first step to identify possible uncertainties is the validation of the climate model data. In this study, a simple approach is presented which uses statistical methods to estimate the accuracy of climate model data and provides a measure to assess the adequacy of the data for hydrological impact modeling. As an example the data of three regional climate models - including the statistical model WETTREG2006 and the dynamical models REMO and CLM, all driven by the global climate model ECHAM5/MPI-OM - are evaluated with respect to their ability to reproduce observed temperature and precipitation. The validation is carried out for interpolated areal means for the period from 1961–2000 based on daily values using different indices and efficiency criteria. The study area is the whole catchment of the rivers Aller and Leine (Lower Saxony, Germany), including nine subbasins of the catchment. The results show deviations for the dynamical climate models according to the observed temeperatures as well as the drought indices. As regards WETTREG, however, all indices were in more or less good agreement. The hydrological model was able to adequately simulate discharge when driven by meteorological data and climate model data from the 20th century control run. However, the bandwidth of the goodness-of-fit of simulated historical high and low discharge was, however, larger than the bandwidth obtained for the corresponding extreme climate conditions. For a significance level of 95 % it could be evidenced by means of the Wilcoxon-Mann-Whitney test that the examined model chains are suitable for performing climate impact assessment. Albeit extreme conditions, especially low flows, should invariably be evaluated depending on the model chain in order to provide a more reliable basis for decisions and adaptation measures.

1 Einleitung

Veränderungen klimatischer Verhältnisse sowie deren sektorale Folgen werden anhand von Änderungssignalen eingeschätzt, die sich aus dem Vergleich von Simulationen der Vergangenheit und der Zukunft ergeben. Impaktmodelle, zu denen auch hydrologische Modelle zählen, nutzen dabei Klimamodelldaten als Eingangsdaten, um die Folgen eines sich wandelnden Klimas unter bestimmten Voraussetzungen (Szenarien) und Fragestellungen auf regionaler Ebene zu simulieren (z.B. Fowler et al. 2007, Graham et al. 2007, Kay et al. 2009). Zukünftige Veränderungen extremer Abflussverhältnisse (Hoch- und Niedrigwasser) haben eine hohe wasserwirtschaftliche Relevanz, weshalb diese im Fokus vieler Untersuchungen standen (Blöschl et al. 2011, Huang et al. 2013a, 2013b).

Klimaprojektionen weisen stets Unsicherheiten auf, die verschiedene Ursachen haben. Neben dem unvollkommenen Wissen über die Prozesse und Wechselwirkungen im Klimasystem (Klimavariabilität) sowie der Frage nach der zukünftigen Entwicklung der Treibhausgas-Emissionen (Szenarienannahme) sind hier vor allem strukturelle Modellunsicherheiten zu nennen, also die Umsetzung des begrenzten Wissens über systemimmanente Prozesse in mathematische Algorithmen (Modellbeschränkung) (Linke et al. 2014). Zudem können Klimamodelle systematische Fehler (Bias) aufweisen, welche

wiederum zu systematischen Fehlern in den Ergebnissen der Impaktmodellierung führen und somit konkrete Aussagen bezüglich der Klimafolgen erschweren (Mudelsee et al. 2010, Muerth et al. 2013). So kommen Kay et al. (2006) in ihrer Untersuchung von Hochwasserstatistiken in zwei Einzugsgebieten in Großbritannien zu dem Ergebnis, dass die Struktur globaler und regionaler Klimamodelle in diesem Zusammenhang die größte Unsicherheitsquelle darstellt, hydrologische Modelle für sich dagegen weniger Fehler produzieren. Zu ähnlichen Erkenntnissen kommen auch Prudhomme und Davis (2009) sowie Chen et al. (2013).

Grundsätzlich stellt sich für Planer und Entscheidungsträger die Frage, wie sicher und robust die Aussagen von Impaktmodellen bzw. der zugrunde liegenden Modellkette für die Zukunft sind und ob diese Ergebnisse geeignet erscheinen, um darauf basierend meist langfristige und kostenintensive Anpassungsmaßnahmen umzusetzen. Eine integrative Bewertung von Klimamodelldaten schließt grundsätzlich mehrere Schritte ein, deren verschiedene Ansätze in der Literatur diskutiert wurden (z.B. Merz et al. 2012, Müller 2010, Bronstert et al. 2007). Ein Punkt ist die Beurteilung der Modellphysik, d.h. der Abbildung des physikalischen Prozessverständnisses über das Klimasystem innerhalb des Modells. Ein weiterer Aspekt ist die Konsistenz bzw. Robustheit der Modellergebnisse innerhalb eines Ensembles verschiedener Modelle, d.h. stimmen die zukünftigen Änderungssignale der jeweiligen Modelle grundsätzlich überein oder weisen sie eine große Bandbreite auf. Ein erster Schritt, die Klimamodelle hinsichtlich ihrer Plausibilität zu untersuchen, ist die Validierung der simulierten Zeitreihen anhand historischer Beobachtungen. Generell werden dazu statistische Größen, welche für die spätere Impaktmodellierung relevant sind, aus Klimamodell- und Beobachtungsdaten abgeleitet und miteinander verglichen. Hierbei ist jedoch nicht sicher, ob sich die Modellunsicherheiten bei einem geänderten Klima in gleicher Weise fortsetzen werden. Refsgaard et al. (2014) schlagen in diesem Zusammenhang ein Testverfahren mit Proxy-Daten zukünftiger Klimabedingungen vor.

Prein et al. (2011) weisen darauf hin, dass verschiedene Klimamodelle die betrachteten Klimagrößen unterschiedlich gut abbilden können. Fowler et al. (2007) fassen in ihrem Review gegenwärtiger Literatur zum Thema des „Downscaling“, also der Verbesserung der räumlichen Auflösung von Klimamodelldaten, für die hydrologische Modellierung zusammen, dass Klimamodelle das Winterklima besser wiedergeben als den Sommer und nasse Bedingungen besser als trockene. Bronstert et al. (2007) verglichen die drei regionalen Klimadatensätze REMO, WETTREG und STAR für Süddeutschland. Sie zeigten, dass Prozesse wie Schneeschmelze oder Evapotranspiration, welche stark von der Temperatur abhängen, allgemein besser nachgebildet werden konnten als Niederschlag dominierte Prozesse wie z. B. Sommerhochwasser. In ihrer Studie lag WETTREG im Mittel, sowohl für die Temperatur, als auch für den Niederschlag, deutlich näher an der Beobachtung als REMO. Letzteres überschätzte sowohl die Temperatur als auch den Niederschlag. Ein ähnliches Muster wurde für REMO bei einem Vergleich zwischen 10 regionalen Klimamodellen von Jacob et al. (2007) beobachtet, wobei darauf verwiesen wurde, dass der systematische Fehler (Bias) stark durch die Randbedingungen, also das antreibende globale Klimamodell, beeinflusst sein kann. Tendenzen der Überschätzung der Temperatur und/oder des Niederschlags von REMO wurden auch in weiteren Studien festgestellt (z.B. Kotlarski et al. 2005, Wegehenkel et al. 2010). WETTREG2006 zeigt in Spekat et al. (2007) für den Nordwesten Deutschlands

dagegen eine leichte Unterschätzung der Niederschläge. Schoetter et al. (2012), die in ihrer Studie für die Metropolregion Hamburg (Projekt KLIMZUG-NORD) neben REMO auch das Regionalmodell CLM betrachten, stellen für CLM eine leichte Unterschätzung der Temperaturen im Sommer und Herbst fest sowie eine Überschätzung der Niederschläge durch beide Modelle, speziell im Sommer. Sie betonen in diesem Zusammenhang, dass die Abweichungen beim Niederschlag auch erheblich vom Referenzdatensatz der Beobachtung abhängen. Auch im Zuge des Projektes KLIMA (Weber 2009) werden bei REMO und CLM für den Süden Deutschlands deutliche Abweichungen beid er Temperatur und dem Niederschlag gegenüber den Beobachtungen aufgezeigt. WETTREG zeigt für beide Kenngrößen dagegen gute Übereinstimmungen.

Diese Studie hat zum Ziel, einen einfachen Ansatz zur Abschätzung der Qualität von Klimamodelldaten vorzustellen. Diese Herangehensweise soll es der Impaktmodellierung in der Praxis ermöglichen, eine erste fallspezifische Beurteilung der Daten bzgl. ihrer Eignung für ausgewählte Fragestellungen vornehmen zu können. Das Verfahren wird angewendet am Beispiel der Kontrollläufe der regionalen Klimamodelldatensätze WETTREG2006 (Spekat et al. 2007), REMO (Jacob et al. 2008, Jacob et al. 2009) sowie CLM (Rockel et al. 2008), alle angetrieben durch das Globalmodell ECHAM5/MPI-OM. Die Evaluierung erfolgt anhand von interpolierten Beobachtungsdaten des Deutschen Wetterdienstes für einen ausgewählten Referenzzeitraum (1961-2000) auf Einzugsgebietsebene in Niedersachsen.

2 Untersuchungsgebiet und Daten

Im Zuge der Forschungsprojekte KliBiW (Globaler Klimawandel – Wasserwirtschaftliche Folgenabschätzung für das Binnenland; Hölscher et al. 2012) und KLIFF/KLIFWA (Auswirkungen von Klimaänderungen auf Wasserdargebot, Hochwasserrisiko und Gewässerbelastung in Niedersachsen; Haberlandt et al. 2014) sollten Projektionen der zukünftigen hydrologischen Verhältnisse im Einzugsgebiet von Aller, Leine und Oker im südöstlichen Niedersachsen mittels eines Ensembles von ausgewählten regionalen Klimamodelldaten und einem Impaktmodell simuliert werden. Das Gebiet schließt mit dem Harz und dem Weser- und Leine-Bergland im Süden, dem Weser-Aller-Flachland im zentralen Bereich sowie der Lüneburger Heide im Norden unterschiedliche Naturräume des niedersächsischen Binnenlandes ein (Abb. 1). Um den möglichen Einfluss der räumlichen Skala und der Topographie auf die Güte der Modelldaten zu beurteilen, erfolgte die Validierung der Klimamodelldaten für das Gesamtgebiet sowie für neun weitere, darin befindliche Teileinzugsgebiete unterschiedlicher Größe und Lage (Abb. 1, Tab. 1). Die Ausdehnung der Gebiete reicht dabei von etwa 100 km^2 bis 15.000 km^2. Der mittlere jährliche Niederschlag und die mittlere jährliche Temperatur wurden für den Beobachtungszeitraum (1961–2000) anhand von Daten des Deutschen Wetterdienstes ermittelt. Am Gebietsauslass eines jeden Einzugsgebietes befindet sich zudem ein Abflusspegel.

Der Referenzdatensatz für die Bewertung der Modelldaten bestand aus geprüften Stationszeitreihen des Deutschen Wetterdienstes (DWD). Diese wurden bzgl. des Niederschlagsmessfehlers (Richter 1995) korrigiert (vgl. Kapitel 3). Für das Untersuchungsgebiet lagen Daten von 247 Niederschlagsstationen sowie 43 Klimastationen

auf Tageswertbasis vor (Abb. 2). Hinzu kamen Daten von weiteren Stationen außerhalb des Gebietes.

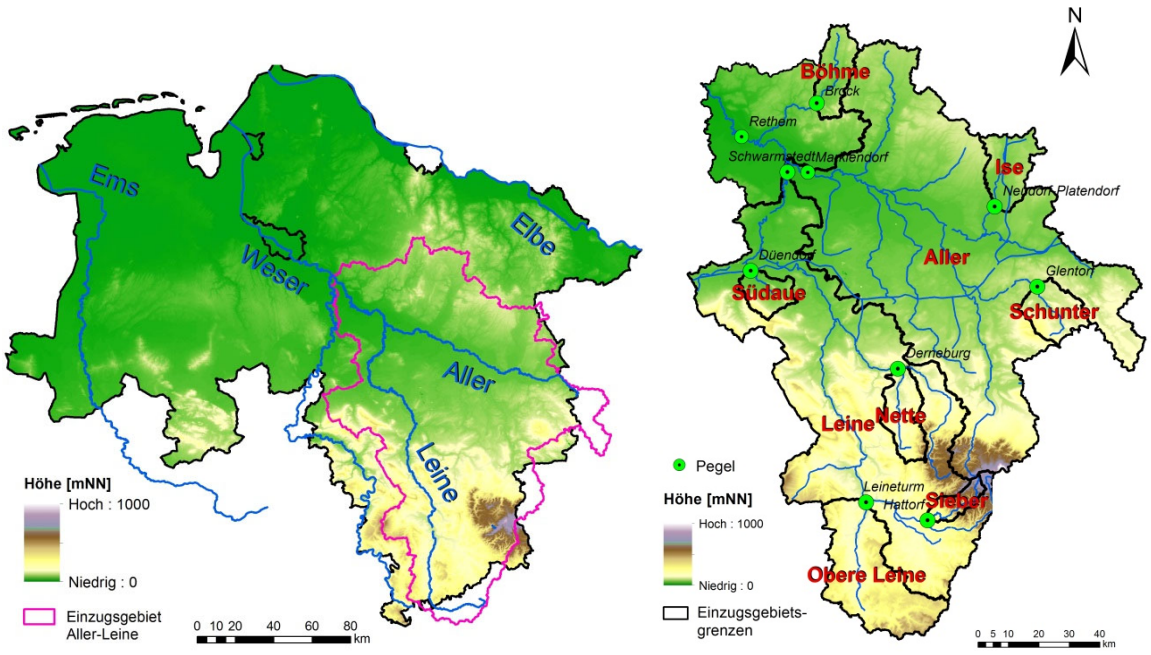


Abbildung 1: Lage des Einzugsgebietes von Aller und Leine in Niedersachsen (links); Lage der Teileinzugsgebiete (mit Abflusspegeln) innerhalb des Betrachtungsraumes von Aller und Leine (rechts)
Location of the catchment of the rivers Aller and Leine in Lower Saxony, Germany (left); Location of the subbasins within the catchment of the rivers of Aller and Leine (right)

Tabelle 1: Untersuchungsgebiete und ausgewählte Charakteristika (1961 – 2000)
Subbasins and selected characteristics (1961 – 2000)

Gebiet	Größe [km^2]	Ø Höhe [m]	min. Höhe [m]	max. Höhe [m]	Ø Niederschlag [mm]	Ø Temperatur [°C]	Pegel	Ø Abfluss MQ [m^3/s]
Aller-Leine (ALO)	15.733	139	5	1141	794	8,7	Rethem	112,1
Aller	7.428	92	9	1141	750	8,8	Marklendorf	42,3
Leine	6.448	203	5	929	838	8,5	Schwarmstedt	59,2
Obere Leine	993	250	114	518	793	8,2	Leineturm	8,2
Ise	341	76	53	122	735	8,7	Neudorf-Platendorf	2,1
Nette	304	208	88	622	881	8,5	Derneburg	3,0
Schunter	292	137	86	314	717	8,8	Glentorf	1,5
Böhme	288	77	41	149	868	8,5	Brock	3,1
Südaue	191	93	44	400	788	9,0	Düendorf	1,4
Sieber	129	480	181	929	1175	6,8	Hattorf	2,6

Anhand der Beobachtungsdaten wurden verschiedene regionale Klimamodelldatensätze evaluiert. Grundsätzlich sollte bei der Betrachtung von Klimafolgen ein möglichst großes Ensemble von Modellen herangezogen werden, um die systemimmanenten Unsicherheiten, wie z.B. Klimavariabilität und Modellbeschränkungen, besser quantifizieren zu können (Linke et al. 2014). In der Praxis scheitert dieser Ansatz jedoch häufig an dem Aufwand, die gesamte Modellkette Globalmodell → Regionalmodell → Impaktmodell für ein solches Ensemble zu durchlaufen. Daher hat man sich im Zuge des Projektes KliBiW zunächst auf die Betrachtung von in Deutschland entwickelten regionalen Klimamodellen beschränkt, die den Zeitraum von 1961–2000 umfassen. Neben Daten aus den dynamischen Modellen REMO und CLM wurden auch Daten des statistischen Modells WETTREG in der Version 2006 für die Analysen verwendet. Alle Modelle werden angetrieben durch das Globalmodell ECHAM5/MPI-OM (Roeckner et al. 2003), welches eine horizontale Auflösung der Atmosphäre von etwa 200x200 km besitzt.

Dynamische Regionalisierungsverfahren beschreiben, ähnlich wie Globale Klimamodelle, die physikalischen Prozesse des Klimasystems durch mathematische Algorithmen innerhalb eines dreidimensionalen Gitters. Als Randbedingungen nutzen sie dabei die Vorgaben der Globalmodelle und rechnen anschließend auf einer räumlich höheren Auflösung. Im Fall von REMO beträgt der Abstand zwischen den Gitterknoten, für die die Ergebnisse simuliert werden, 0,088° (ca. 10 km). Insgesamt liegen 163 REMO-Rasterpunkte im Untersuchungsgebiet (Abb. 2). Für die Untersuchung wurden zwei Versionen von REMO betrachtet; zum einen REMO-UBA (Jacob et al. 2008), angetrieben durch den Lauf 1 des Globalmodells ECHAM5/MPI-OM, und zum anderen die Version REMO-BfG (Jacob et al. 2009), angetrieben vom Lauf 2 des gleichen Globalmodells. Im Fall von CLM (Rockel et al. 2008) beträgt der Abstand der Gitterknoten 0,2° (ca. 20 km), von denen 52 im Untersuchungsgebiet liegen. Auch das CLM-Modell lag in zwei Versionen vor, die von ECHAM5/MPI-OM Lauf 1 und Lauf 2 angetrieben wurden.

Der dritte Datensatz, WETTREG in der Version 2006 (Spekat et al. 2007), basiert ebenfalls auf dem Lauf 1 von ECHAM5/MPI-OM. WETTREG stellt gegenüber REMO ein statistisches Regionalisierungsverfahren dar. Hierbei wird ein statistischer Zusammenhang zwischen großräumigen Wetterlagen, abgeleitet aus ERA40-Reanalysedaten, und lokalen Veränderungen von Klimavariablen aus den Beobachtungsdaten an Stationen für den Zeitraum 1971 – 2000 ermittelt. Anschließend werden anhand dieser Statistik und den Großwetterlagen aus dem Globalmodell im Zukunftszeitraum synthetisch lokale Wetterereignisse erzeugt. Für WETTREG lagen 60 Niederschlagsstationen und 14 Klimastationen auf Tageswertbasis im Untersuchungsgebiet vor (Abb. 2).

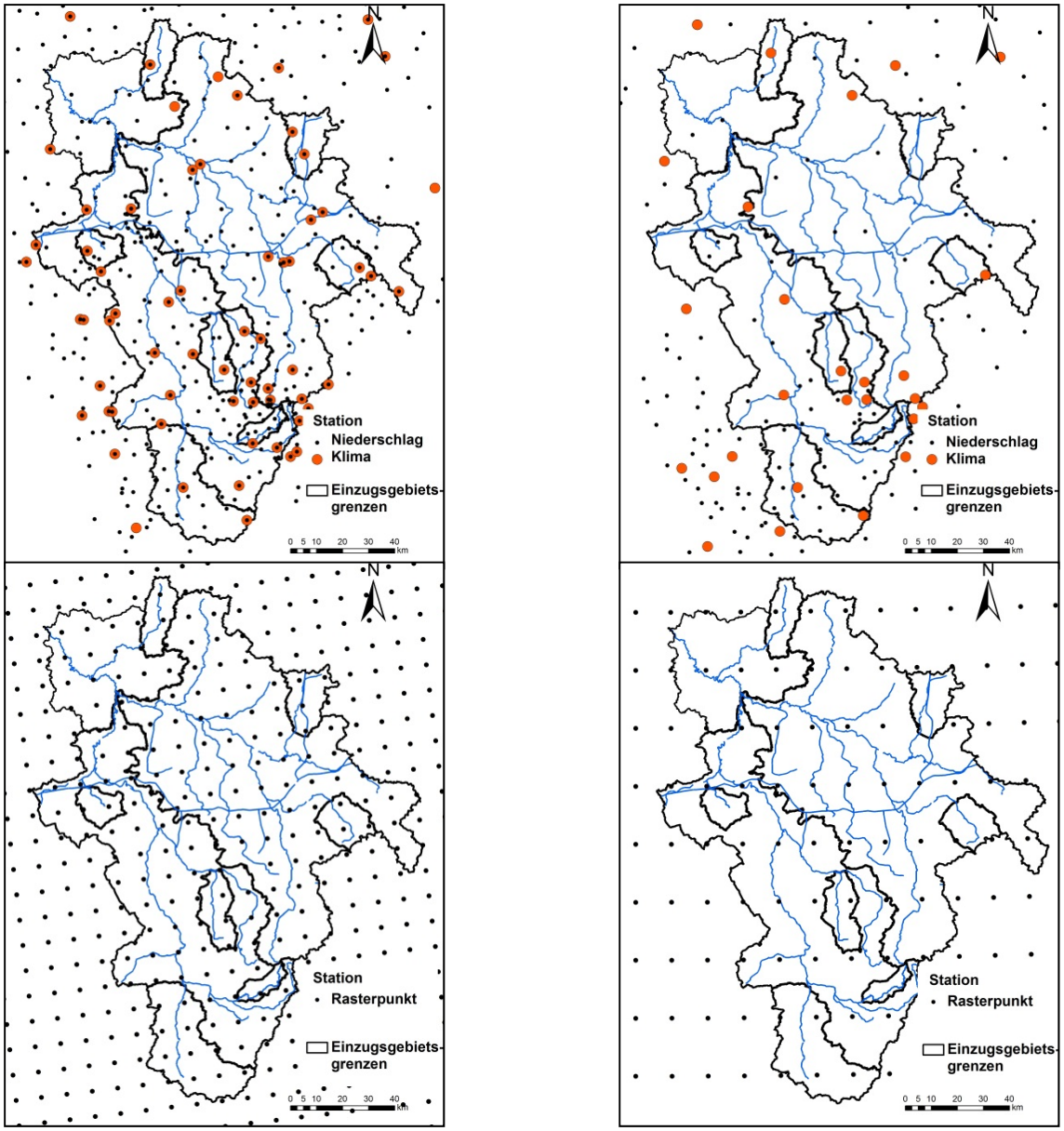


Abbildung 2: Stationen bzw. Gitterpunkte mit Zeitreihen der Klimavariablen im Bereich des Einzugsgebietes von Aller und Leine (links oben: Beobachtung, rechts oben: WETTREG2006, links unten: REMO, rechts unten: CLM)
Stations or grid points with time series of climate variables within the catchment of the rivers Aller and Leine (left top: observation, right top: WETTREG2006, left bottom: REMO, right bottom: CLM)

Die Validierung erfolgte auf Basis von Tageswerten für den Zeitraum 1961–2000, der auch als Kontroll-Lauf (20C) der Modelldaten bezeichnet wird. Die Beobachtungsdaten sowie die Datensätze von WETTREG und CLM lagen bereits in dieser zeitlichen Auflösung vor. Die Daten von REMO, standardmäßig in stündlicher Auflösung, wurden zu Tageswerten aggregiert.

Die weiteren Ausführungen zu den genannten Regionalmodellen beziehen sich grundsätzlich immer auf die gesamte Modellkette, also z.B. REMO-UBA: ECHAM5/MPI-OM Lauf1 / REMO-UBA.

3 Methodik

3.1 Aufbereitung der Basisdaten

In dieser Studie sollten die Ergebnisse einer hydrologischen Klimafolgenabschätzung hinsichtlich der mit ihnen verbundenen Unsicherheit beurteilt werden. Diese Unsicherheiten resultieren zum einen aus den gewählten Elementen der betrachteten Modellkette, also aus zwei Realisierungen eines Globalmodells (ECHAM5/MPI-OM, Lauf 1 und 2), zwei dynamischen (REMO, CLM) und einem statistischen (WETTREG) Regionalisierungsverfahren sowie einem Impaktmodell. Zum anderen ergeben sich Unsicherheiten auch aus der Weiterverarbeitung der originären Daten, wie sie im Folgenden beschrieben wird. Andere Unsicherheitsquellen, wie beispielsweise andere Global- oder Regionalmodelle, weitere Impaktmodelle oder verschiedene Referenzdatensätze der Beobachtung wurden nicht betrachtet.

Als klimatische Größen, die wesentlichen Einfluss bei der hydrologischen Impaktmodellierung haben, wurden der Niederschlag und die mittlere Temperatur ausgewählt. Diese wurden im Zuge der Studie bewertet. Fehler, die bei der Messung dieser Daten auftreten, die als Referenz bei der Validierung der Modelldaten dienen, können neben der genannten Modellunsicherheit eine weitere Unsicherheitsquelle darstellen. Ein gängiges Verfahren zur Korrektur des systematischen Fehlers bei Niederschlagsmessungen durch den Hellmann-Niederschlagsmesser ist das Verfahren nach Richter (1995). Diese Methode wurde bei allen Stationen der Beobachtung und für WETTREG, welches auf diesen Beobachtungsdaten beruht, angewendet. Sie erfolgte einerseits auf Basis der Niederschlagsart in Abhängigkeit von der Temperatur, andererseits auf Basis der Horizontabschirmung der Stationen (Haberlandt et al. 2013). Auf die Korrektur eines möglichen Bias bei den Niederschlägen der dynamischen Modelle REMO und CLM wurde verzichtet, auch aufgrund der damit verbundenen Gefahr physikalischer Inkonsistenzen sowie der Frage, ob sich der vorhandene Bias in gleicher Form in der Zukunft fortsetzen wird (Ehret et al. 2012).

Die Evaluierung erfolgte anhand von Gebietsmittelzeitreihen der o.g. zehn (Teil-) Einzugsgebiete. Hierfür wurden die Stationsdaten (Beobachtung und WETTREG) bzw. Gitterpunktzeitreihen (REMO und CLM) in einem ersten Schritt auf ein gleichmäßiges Raster interpoliert (Haberlandt et al. 2013). Die Größe der Rasterzellen betrug bei der Beobachtung 1x1 km, bei WETTREG 5x5 km und bei REMO bzw. CLM 10x10 km. Ein Vorteil dieser Herangehensweise ist die Möglichkeit, alle Daten in denselben räumlichen Bezug zu überführen und damit vergleichbar zu machen. Die Interpolation von Daten unterschiedlicher räumlicher Auflösung kann jedoch zu regional unterschiedlich ausgeprägten Fehlern führen. Dieser Aspekt stellt eine weitere Unsicherheit in der Modellkette dar und sollte bei der Interpretation der Ergebnisse berücksichtigt werden. Die angewendeten Verfahren zur Interpolation wurden für jede Klimagröße und jeden Modelldatensatz mittels Kreuzvalidierung anhand verschiedener Gütekriterien bewertet und das jeweils optimale Interpolationsverfahren für das Untersuchungsgebiet ermittelt. Für die Interpolation der Stationsdaten wurde für die Temperatur das External Drift Kriging (EDK) mit einem digitalen Höhenmodell als Zusatzinformation verwendet. Bei der Interpolation der Niederschläge

zeigte das Ordinary Kriging (OK) die gleiche Güte wie das EDK, so dass in diesem Fall das einfachere OK genutzt wurde. Da die REMO- und CLM-Daten bereits in Rasterform vorlagen, wurden diese lediglich mittels Inverser Distanz-Gewichtung (IDW) in Form des Quadrantenverfahrens an die Raster der anderen Daten angepasst. Eine detaillierte Beschreibung der Vorgehensweise und Ergebnisse sind in Haberlandt et al. (2013) zu finden.

3.2 Festlegung der zu analysierenden Kennwerte

Für den Niederschlag und die Temperatur wurden verschiedene Indizes festgelegt (Tab. 2), die sowohl die mittleren Verhältnisse, als auch die hydro-meteorologisch relevanten nassen bzw. trockenen Bedingungen bei Hoch- bzw. Niedrigwasserführung charakterisieren. Diese Indizes wurden als der Größe nach aufsteigend geordnete Jahreswerte betrachtet.

Tabelle 2: Ausgewählte Kenngrößen von Temperatur und Niederschlag zur Einschätzung der Modellgüte bei hydrologisch relevanten Klimafolgenabschätzungen
Selected indices of temperature and precipitation for the evaluation of model efficiency in the context of hydrological climate impact assessment

Abkürzung	Einheit	Beschreibung
Temperatur		
Tmav	°C	Mittelwert der Tagesmitteltemperatur Tm
Tmq90	°C	90%-Quantil der Tagesmitteltemperatur
Tncd	d	Anzahl der Tage mit Tm ≤ 0 °C (Kältetage)
Niederschlag		
Pav1	mm/d	Mittelwert des Niederschlags für Tage mit N ≥ 1mm
Px1d	mm/d	Größte Tagesniederschlagssumme
Px3d	mm/3d	Größte 3-Tagesniederschlagssumme
Pndd	d	Anzahl der Tage mit N < 1mm (Trockentage)
Pxcdd	d	Maximale Anzahl aufeinanderfolgender Tage mit N < 1mm

Vor dem Hintergrund wasserwirtschaftlicher Anwendungen wurden diese Daten als Gebietswerte untersucht, welche eine besondere Bedeutung für die Betrachtung des Wasserhaushaltes von Flusseinzugsgebieten haben und als Eingangsdaten für entsprechende Impaktmodellierungen Verwendung finden. Es war nicht Ziel der Studie, die Güte der Klimaprojektionen ohne Kontext zu Anwendungen in der Impaktforschung zu bewerten. Hierzu würde man beispielsweise das stationsbasierte WETTREG an den Beobachtungen der entsprechenden Stationen validieren.

3.3 Auswahl der Gütekriterien

Die Evaluierung der Klimadatensätze erfolgte anhand von ausgewählten Gütekriterien. In der Literatur findet sich eine Vielzahl verschiedener Gütemaße (Krause et al. 2005, Legates & McCabe 1999, Moriasi et al. 2007). In dieser Studie wurden drei Verfahren angewendet, um die Effizienz der Klimamodelldaten gesamtheitlich zu beurteilen.

Ein häufig verwendetes Kriterium zur Beurteilung der Güte von Modellzeitreihen ist der Nash Sutcliffe Effizienzkoeffizient (*NSE*) nach Nash & Sutcliffe (1970):

$$NSE = 1 - \frac{\sum_{i=1}^{N}\left(X_{obs}^{i} - X_{sim}^{i}\right)^2}{\sum_{i=1}^{N}\left(X_{obs}^{i} - \bar{X}_{obs}\right)^2} \quad (1)$$

Hierbei beschreibt N die Anzahl der Stichprobenelemente, X_{obs} die beobachteten Zeitreihenwerte und X_{sim} die vom Modell entsprechend simulierten Werte. Als Zähler i werden hier die der Größe nach sortierten Jahreswerte der Zeitreihe bezeichnet. Der Wertebereich des *NSE* reicht von 1 bis -∞. Ein Wert von 1 bedeutet eine optimale Wiedergabe der beobachteten Werte durch das Modell, während bei einem $NSE < 0$ der Mittelwert der Beobachtung eine bessere Schätzung als die simulierten Werte darstellt. Einen objektiven Grenzwert, ab dem ein Modell als „gut" oder „schlecht" bezeichnet werden kann, bietet der *NSE* im engeren Sinne nicht.

Um den Wert des *NSE* besser beurteilen zu können, wurde daher zusätzlich mittels des Mann-Whitney-Wilcoxon Tests (Mann & Whitney 1947, Wilcoxon 1945), auch als U-Test bekannt, analysiert, ob sich die Werte der Beobachtung signifikant von den Werten der Modelle unterscheiden. Das Prinzip des Testverfahrens wird an dieser Stelle nur kurz erläutert. Für eine detaillierte Darstellung wird auf entsprechende Literatur verwiesen (z.B. Sachs 2004). Der U-Test ist ein parameterfreier Signifikanztest, der die Wahrscheinlichkeit überprüft, ob die zentrale Tendenz von zwei verschiedenen Stichproben (SP) unterschiedlich ist. Hierfür werden die SP zunächst der Größe nach geordnet und die aus beiden SP gemeinsam resultierenden Rangplätze ermittelt. Die Berechnung der Testgröße erfolgt dann über

$$U_1 = n_1 * n_2 + \frac{n_1*(n_1+1)}{2} - R_1 \qquad U_2 = n_1 * n_2 + \frac{n_2*(n_2+1)}{2} - R_2 \quad (2)$$

$$U = Min(U_1, U_2) \quad (3)$$

Hierbei beschreiben n_1 und n_2 die SP-Umfänge, R_1 und R_2 die jeweiligen Summen der Rangplätze der SP und U_1, U_2 und U sind Hilfsgrößen. Bei größeren SP-Umfängen ($n > 10$) ergibt sich die Testgröße *z(U)* anhand

$$z(U) = \frac{\left|U - \frac{n_1*n_2}{2}\right|}{\sqrt{\frac{n_1*n_2*(n_1+n_2+1)}{12}}} \quad (4)$$

Kommen bei den nach Größe sortierten SP Werte mehrfach vor - man spricht hier von Bindungen - dann erhalten die gleichen Einzelwerte jeweils die mittlere Rangzahl. Die korrigierte Formel der Testgröße für diesen Fall lautet

$$z(U) = \frac{\left|U - \frac{n_1*n_2}{2}\right|}{\sqrt{\left[\frac{n_1*n_2}{S*(S-1)}\right]*\left[\frac{S^3-S}{12} - \sum_{i=1}^{i=r}\frac{t_i^3-t_i}{12}\right]}} \quad (5)$$

Der resultierende Wert für *z(U)* wird anhand der kritischen Größe z_α der Standardnormalverteilung beurteilt. Für $z(U) > z_\alpha$ wird auf dem gewählten Signifikanzniveau entschieden, dass sich die beiden SP, in diesem Fall also die Modell- und Beobachtungsdaten, unterscheiden. In dieser Studie wurde ein Signifikanzniveau von 95%, also α=0,05, gewählt.

Schließlich sollte durch die Gütekriterien auch die grundsätzliche Richtung der Abweichung identifiziert werden, also ob das Modell die Beobachtung tendenziell eher über- oder unterschätzt. Dies erfolgt je nach Kennwert durch den absoluten Fehler (*BIAS*) bzw. den relativen oder prozentualen Fehler (*PBIAS*):

$$BIAS = \sum_{i=1}^{N}\left(X_{sim}^{i} - X_{obs}^{i}\right) \tag{6}$$

$$PBIAS = 100 * \frac{\sum_{i=1}^{N}\left(X_{sim}^{i} - X_{obs}^{i}\right)}{\sum_{i=1}^{N} X_{obs}^{i}} \tag{7}$$

Die o.g. Gütekriterien wurden als Maßzahlen für die Validierung der Klimagrößen Temperatur und Niederschlag der Modellketten ECHAM5/MPI-OM Lauf 1 / REMO-UBA, ECHAM5/MPI-OM Lauf 2 / REMO-BfG, ECHAM5/MPI-OM Lauf 1 / CLM, ECHAM5/MPI-OM Lauf 2 / CLM und ECHAM5/MPI-OM Lauf 1 / WETTREG2006 in dieser Studie benutzt.

Am folgenden Beispiel soll die Anwendung und Aussagekraft der einzelnen Kriterien kurz erläutert werden. In Abb. 3 werden die mittleren Niederschläge aus den genannten Modellen den beobachteten mittleren Niederschlägen für das Teileinzugsgebiet der Schunter als Gebietsmittel gegenübergestellt. Dazu wurden die Jahreswerte der Datensätze jeweils der Größe nach geordnet und anschließend als Datenpaare, Beobachtung vs. Modell, neu kombiniert. Im Fokus stand hierbei die Abbildungsgüte bezüglich der Häufigkeit der jeweiligen Ausprägungen der Klimavariablen innerhalb des Betrachtungszeitraumes. Würden die Modelle die beobachteten Werte exakt wiedergeben, lägen die jeweiligen Punkte auf der schwarzen Linie. Die Abweichungen der modellierten Niederschläge werden beschrieben durch die Gütekriterien *NSE*, *z(U)* und *PBIAS* in Tab. 3. Hierbei wird für WETTREG jeweils der Mittelwert aller Realisationen bewertet.

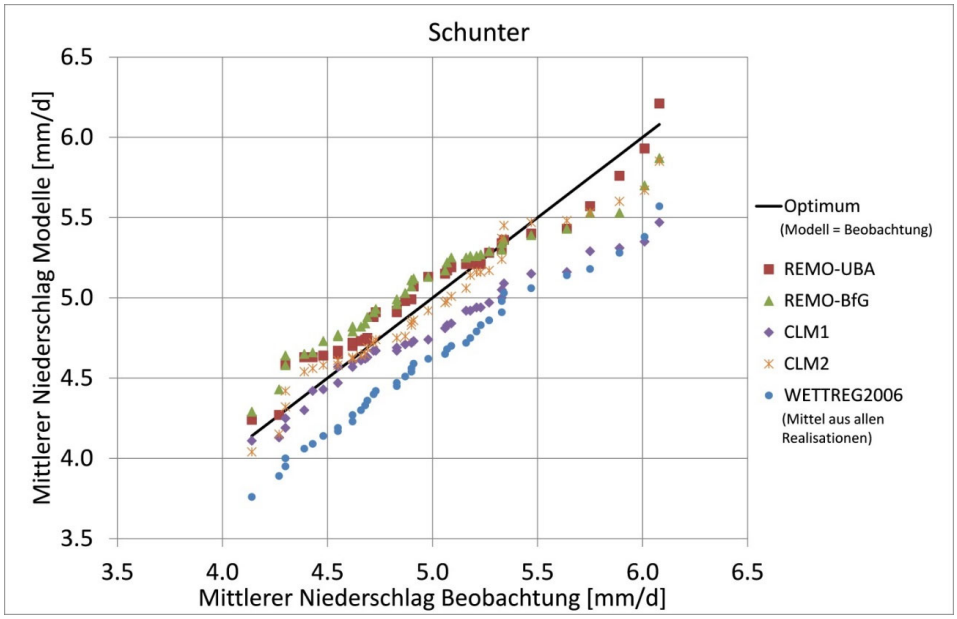


Abbildung 3: Gegenüberstellung von beobachteten und modellierten mittleren Niederschlägen Pav1 (1961–2000) aus den Gebietsmittelzeitreihen als empirische Verteilung am Beispiel des Einzugsgebiets der Schunter

Comparison of observed and simulated model data of mean daily rainfall Pav1 (1961 – 2000) as order-ranked areal means for the subbasin of the river Schunter

Tabelle 3: Modellgüte für die mittleren Niederschläge (Pav1) aus den Gebietsmittelzeitreihen der Klimamodelle für das Einzugsgebiet der Schunter, die als empirische Verteilung den beobachteten Daten gegenübergestellt wurden
Model efficiency of the climate models for the mean daily rainfall (Pav1), based on order-ranked areal means, for the subbasin of the river Schunter

	NSE [-]	*z(U)* [-]	*PBIAS* [%]
ECHAM5/MPI-OM Lauf 1/ REMO-UBA	0,93	0,92	1,39
ECHAM5/MPI-OM Lauf 2 / REMO-BfG	0,85	1,39	1,95
ECHAM5/MPI-OM Lauf 1 / CLM	0,70	1,75	-4,12
ECHAM5/MPI-OM Lauf 2 / CLM	0,94	0,32	-0,98
ECHAM5/MPI-OM Lauf 1 / WETTREG2006	0,26	3,43	-7,89

Der *NSE* zeigt bei WETTREG einen schlechteren Wert als bei den dynamischen Modellen REMO und CLM. Betrachtet man zusätzlich *z(U)* bei einem angenommenen Signifikanzniveau von 95%, d.h. $z_{\alpha=0,05} \approx 1,645$, so wird deutlich, dass neben WETTREG auch CLM Lauf 1 mit einem *NSE* von 0,70 noch signifikant von der Beobachtung abweicht. Würde das Signifikanzniveau auf 99% erhöht werden (mit $z_{\alpha=0,01} \approx 1,96$), so wären die Abweichungen von CLM Lauf 1 noch „akzeptabel". Mit Hilfe des *PBIAS* kann die grundsätzliche Richtung der Abweichungen identifiziert werden. Im Fall von REMO wird der beobachtete mittlere Niederschlag (Pav1) in beiden Läufen leicht überschätzt, während CLM und WETTREG die Beobachtung im Mittel unterschätzen.

4 Analyse der Modellgüte

4.1 Gebietswerte der Temperatur

In einem ersten Schritt wurde die von den Klimamodellen simulierte Temperatur anhand der Beobachtung validiert. Abbildung 4 zeigt, dass REMO-BfG und WETTREG2006 die beobachtete mittlere Temperatur (Tmav) über alle Untersuchungsgebiete in etwa gleich gut simulieren (*NSE* zumeist > 0,75). REMO-UBA zeigt hier eine schlechtere, die beiden CLM-Läufe eine wesentlich schlechtere Modellgüte. Die Prüfgröße *z(U)* des U-Tests bestätigt, dass sich die Werte von REMO-UBA, CLM1 und CLM2 signifikant ($\alpha = 0,05$) von den Beobachtungsdaten unterscheiden. Unter Verwendung des absoluten *BIAS* ist erkennbar, dass beide REMO Varianten die mittlere Temperatur gegenüber der Beobachtung als zu warm einschätzen. WETTREG zeigt im Mittel keine besonderen Abweichungen. Die CLM-Läufe unterschätzen die mittlere Temperatur um etwa 0,5 bis 1,0 °C.

Bei der Betrachtung der Anzahl der Kältetage (Tncd) zeigen alle Modelle durchweg eine relativ gute Modellgüte (*NSE* meist > 0,75) (vgl. Abb. 4). Die Prüfgröße *z(U)* zeigt, dass lediglich REMO-UBA hier z.T. signifikant von der Beobachtung abweicht. Der *BIAS* bei beiden REMO Varianten zeigt eine Unterschätzung der Kältetage (-10 bis -20%), während die CLM-Läufe eine Überschätzung aufweisen (im Mittel über +10%). WETTREG zeigt im Mittel eine nur geringe Überschätzung.

Für das 90%-Quantil der Temperaturen (Tmq90) zeigen die Modelle sehr unterschiedliche Modellgüten, ähnlich wie bei Tmav. Während der *NSE* bei REMO und WETTREG im Mittel oberhalb von 0,75 liegt, bewegen sich die Werte der CLM-Läufe deutlich unter Null. Dementsprechend unterscheiden diese sich nach *z(U)* deutlich von der

Beobachtung. Der *BIAS* zeigt leichte Unterschätzungen der beiden REMO Varianten (bis -0,5 °C) und deutliche Unterschätzungen bei den beiden CLM-Läufen (etwa -1 bis -1,9 °C), während WETTREG kaum Abweichungen aufweist. Zur Veranschaulichung der beschriebenen Eigenschaften der Modellketten erfolgte in einem nächsten Schritt die Gegenüberstellung zur Beobachtung als empirische Verteilung, gemittelt über alle Untersuchungsgebiete, für jede Temperaturkenngröße und jeden Klimamodell-Datensatz (vgl. Kapitel 3.3, Abb. 3). Aufbauend auf dieser Vorgehensweise zeigt Abb. 5 die Charakteristika der Modelle, wie sie zuvor durch die Gütekriterien beschrieben wurden.

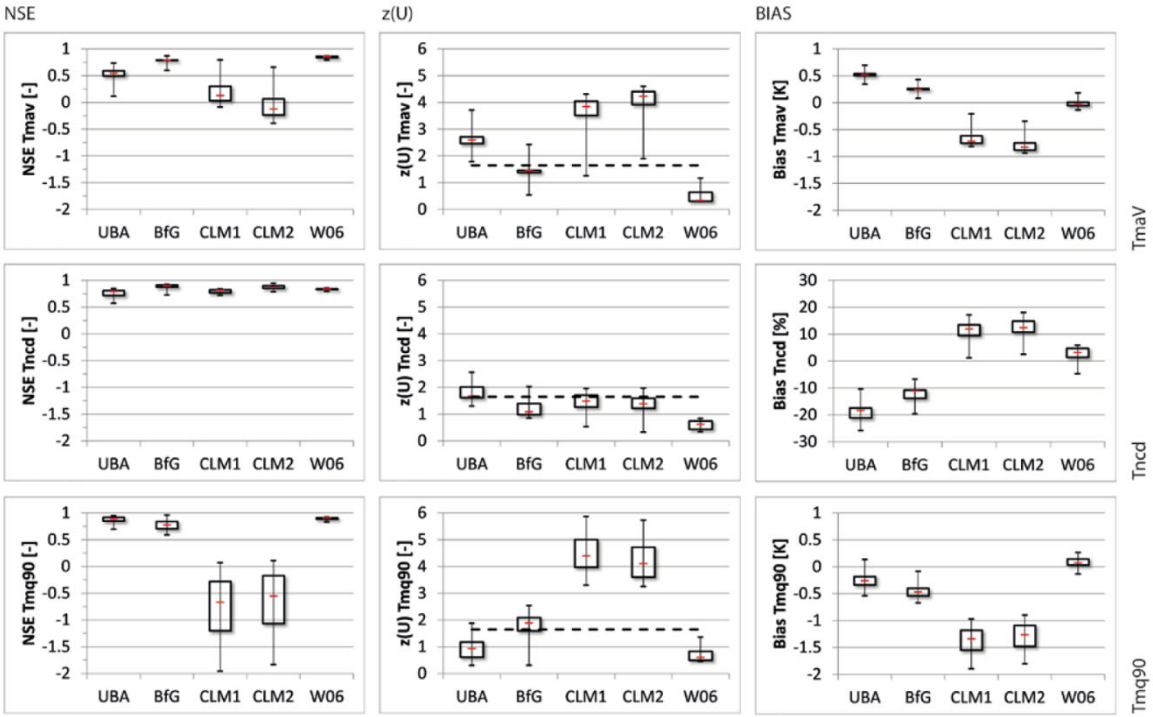


Abbildung 4: Gütekriterien und deren Bandbreite über alle 10 Teilgebiete für die Kenngrößen mittlere Temperatur (Tmav), Anzahl Kältetage (Tncd) und 90%-Quantil der Temperatur (Tmq90) für die Klimamodelldaten von REMO-UBA, REMO-BfG, CLM1, CLM2 und WETTREG2006; die gestrichelte Linie bei *z(U)* markiert die Signifikanzgrenze (bei einer Irrtumswahrscheinlichkeit von α = 0,05)
Band width of goodness-of-fit measures over all 10 subbasins for the indices mean temperature (Tmav), number of cold days (Tncd) and 90%-quantile of temperature (Tmq90) for REMO-UBA, REMO-BfG , CLM1, CLM2 and WETTREG2006; the dashed line at z(U) indicates the margin of significance (α = 0,05)

Es zeigt sich, dass WETTREG2006 bei den kleineren (kühleren) Werten der Temperatur-Indizes zu einer leichten Überschätzung tendiert, während größere (wärmere) Werte eher unterschätzt werden. Die Bandbreite der beobachteten Temperaturdynamik wird tendenziell also unterschätzt. Insgesamt liegt WETTREG aber relativ nahe an den Beobachtungsdaten. Beide REMO Varianten schätzen die mittlere Temperatur über fast den gesamten Wertebereich als zu warm ein. Die Anzahl der Kältetage und die hohen Temperaturen werden dagegen eher unterschätzt. REMO-UBA zeigt insgesamt etwas größere Differenzen zur Beobachtung als REMO-BfG, mit Ausnahme des Tmq90. Beide CLM-Läufe unterschätzen die Temperaturen deutlicher als die REMO-Läufe. Die Anzahl der Kältetage wird von CLM tendenziell überschätzt.

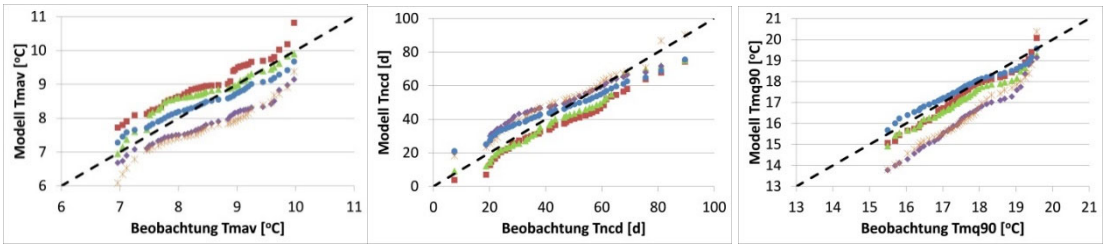


Abbildung 5: Gegenüberstellung von Beobachtungs- und Modelldaten (REMO-UBA (rot), REMO-BfG (grün), CLM1 (lila), CLM2 (organge) und WETTREG2006 (blau)) für die mittlere Temperatur (Tmav), die Anzahl der Kältetage (Tncd) und das 90%-Quantil der Temperaturen (Tmq90), als empirische Verteilung jeweils gemittelt über alle Teilgebiete; die gestrichelte Linie markiert den Bereich, in dem die Modellwerte genau der Beobachtung entsprächen
Comparison of observed and simulated data (REMO-UBA (red), REMO-BfG (green), CLM1 (purple), CLM2 (orange) and WETTREG2006 (blue)) for mean temperature (Tmav), number of cold days (Tncd) and 90%-quantile of temperature (Tmq90), as order-ranked areal means averaged over all subbasins; the dashed line indicates the area of perfect fit

4.2 Gebietswerte des Niederschlags

Die Betrachtung der Modellgüte für die Niederschlagskenngrößen wurde auf die gleiche Weise wie die Bewertung der Temperaturkenngrößen durchgeführt.

Die Beurteilung der Modellgüte für den mittleren Niederschlag (Pav1) anhand des in Abbildung 6 dargestellten *NSE* zeigt im Mittel eine relativ gleichmäßige Werteverteilung. Alle Modelle liegen im Wesentlichen oberhalb von 0,5, im Mittel um 0,7 bis 0,75. Lediglich die CLM-Läufe weisen hier für das kleinste Teilgebiet (Sieber) Werte unter Null auf. Die Prüfgröße *z(U)* zeigt, dass in der Mehrzahl der Gebiete die Modelldaten nicht signifikant von der Beobachtung abweichen, mit Ausnahme von WETTREG2006, dessen Werte aber nur leicht oberhalb der Signifikanzgrenze liegen. Der prozentuale Fehler (*PBIAS*) liegt bei allen Modellen zumeist im Bereich von 0 bis -10%.

Der Vergleich des größten 1-Tages-Niederschlages (Px1d) zeigt für alle Datensätze bei fast allen Gebieten mittlere bis gute Modellgüten für den *NSE* (Abbildung 6). Lediglich die REMO Varianten und CLM2 zeigen beim Einzugsgebiet der Ise und der Sieber z.T. erheblich schwächere Werte. Der Prüfwert *z(U)* zeigt entsprechend, dass alle Modelldaten für die Mehrzahl der Gebiete mit den Kennwerten der Beobachtung übereinstimmen. Die dynamischen Modelle überschätzen die Maximalniederschläge um etwa 10% (CLM etwas mehr als REMO), und dies in kleineren Einzugsgebieten stärker als in größeren. WETTREG hingegen unterschätzt diesen Kennwert um etwa 10%.

Die Untersuchung der maximalen 3-Tages-Niederschlagssumme (Px3d) zeigt insgesamt ähnliche Tendenzen der Modelle wie die Analyse des 1-Tages-Maximums. Der *NSE* lag hier bei allen Modellen im Mittel etwas besser als beim Px1d, mit Ausnahme von CLM2. Auch *z(U)* belegt in diesem Zusammenhang insgesamt eine bessere Abbildung der Beobachtungswerte durch die Modelle als für die 1-Tages-Maxima. Dementsprechend fallen auch die Abweichungen (*PBIAS*) etwas geringer aus. Dies zeigt, dass die Modelle die beobachteten hohen Niederschläge für längere Zeiträume anscheinend etwas besser reproduzieren können als für kürzere. Dies macht sich vor allem in den kleineren Teilgebieten bemerkbar.

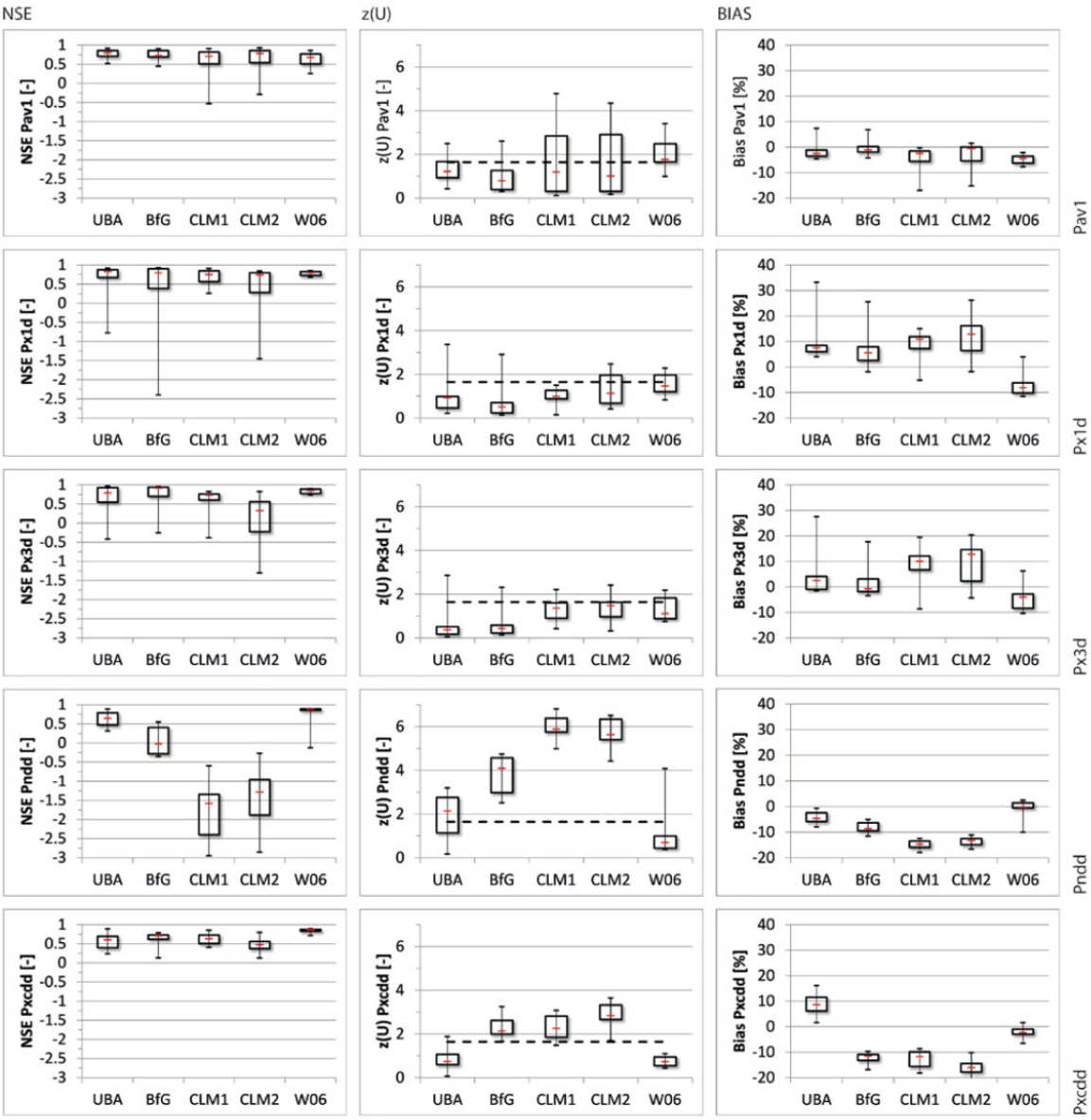


Abbildung 6: Gütekriterien und deren Bandbreite über alle 10 Teilgebiete für die Kenngrößen mittlerer Niederschlag (Pav1), maximaler Tagesniederschlag (Px1d), maximaler 3-Tages-Niederschlag (Px3d), Anzahl Trockentage (Pndd) und maximale Anzahl aufeinanderfolgender Trockentage (Pxcdd) für die Klimamodelldaten von REMO-UBA , REMO-BfG , CLM1, CLM2 und WETTREG2006; die gestrichelte Linie bei *z(U)* markiert die Signifikanzgrenze (bei einer Irrtumswahrscheinlichkeit von α = 0,05)

Band width of goodness-of-fit measures over all 10 subbasins for the indices mean precipitation (Pav1), maximum 1-day precipitation (Px1d), maximum 3-day precipitation (Px3d), number of dry days (Pndd) and maximum number of cumulative dry days (Pxcdd) for REMO-UBA, REMO-BfG, CLM1, CLM2 and WETTREG2006; the dashed line at z(U) indicates the margin of significance (α = 0,05)

Die maximale Anzahl aufeinanderfolgender Trockentage (Pxcdd) wird von WETTREG am besten abgebildet (*NSE* im Mittel um 0,8). Die anderen Modelle liegen schlechter, weisen aber häufig noch einen Wert oberhalb von 0,5 auf. Die Prüfgröße *z(U)* belegt deutlich, dass außer WETTREG2006 und REMO-UBA keines der Modelle in der Lage ist die Beobachtung, bezogen auf diesen Kennwert, hinreichend genau wiederzugeben. Der *PBIAS* zeigt, dass REMO-UBA die maximale Dauer der Trockenphasen um rund 10% überschätzt, während REMO-BfG sie

in etwa gleichem Maße unterschätzt. Die CLM-Läufe liegen hier zwischen -10 und -20%. WETTREG weist kaum nennenswerte Abweichungen auf.

Die Untersuchung der Anzahl der Trockentage mit einem Grenzwert von N < 1mm (Pndd), zeigt ein ähnliches Ranking der Modelle wie zuvor die maximal aufeinanderfolgenden Trockentage. D.h. die Trockentage werden von WETTREG verhältnismäßig gut abgebildet, REMO UBA liegt etwas schlechter. Die Modellgüte von REMO-BfG sowie den beiden CLM-Läufen liefert hingegen deutlich schlechtere Werte. Dies macht sich auch in dem Abstand dieser drei Modelle zur Signifikanzgrenze bei *z(U)* bemerkbar.

Auch für die Kenngrößen des Niederschlags wurden die mittleren Werte der jeweiligen empirischen Verteilung über alle Teilgebiete berechnet und für jeden Klimamodell-Datensatz der Beobachtung gegenübergestellt (vgl. Abb. 7).

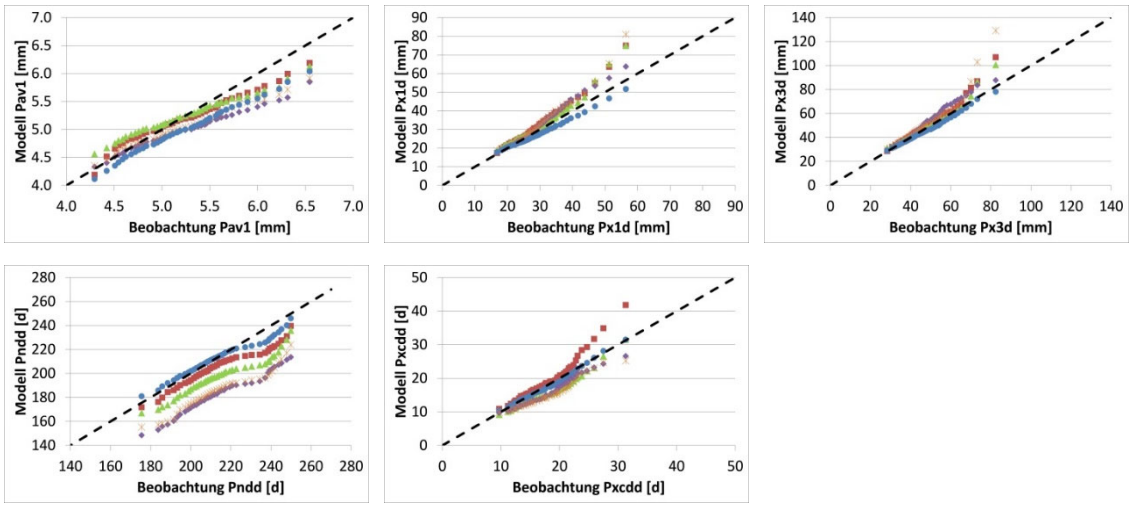


Abbildung 7: Gegenüberstellung von Beobachtungs- und Modelldaten (REMO-UBA (rot), REMO-BfG (grün), CLM1 (lila), CLM2 (organge) und WETTREG2006 (blau)) für den mittleren Niederschlag (Pav1), den maximalen Tagesniederschlag (Px1d) , den maximalen 3-Tages-Niederschlag (Px3d), die Anzahl der Trockentage (Pndd) und die maximale Anzahl aufeinanderfolgender Trockentage (Pxcdd), als empirische Verteilung jeweils gemittelt über alle Teilgebiete; die gestrichelte Linie markiert den Bereich, in dem die Modellwerte genau der Beobachtung entsprächen
Comparison of observed and simulated data (REMO-UBA (red), REMO-BfG (green), CLM1 (purple), CLM2 (organge) and *WETTREG2006 (blue)) for mean precipitation (Pav1), maximum 1-day precipitation (Px1d) , maximum 3-day precipitation (Px3d), number of dry days (Pndd) and maximum number of cumulative dry days (Pxcdd), as order-ranked areal means averaged over all subbasins; the dashed line indicates the area of perfect fit*

WETTREG2006 zeigt bei allen Kenngrößen relativ gute Übereinstimmungen mit der Beobachtung, mit Ausnahme der mittleren Niederschläge (Pav1), die grundsätzlich unterschätzt werden, vor allem in nasseren Jahren. Hohe Niederschläge und Jahre mit vielen Trockentagen werden tendenziell leicht unterschätzt, aber insgesamt besser abgebildet als durch die anderen Modelle. REMO-UBA zeigt Tendenzen, die extremen (großen) Ausprägungen der betrachteten Kenngrößen schwach abzubilden. Dies zeigt sich in einer deutlichen Überschätzung der hohen Niederschläge und Trockenphasen bzw. einer Unterschätzung der Anzahl von Trockentagen. REMO-BfG zeigt ein ähnliches Verhalten. Allerdings wird die Anzahl der Trockentage deutlicher unterschätzt und Trockenphasen

werden grundsätzlich unterschätzt. Die Abweichungen zur Beobachtung fallen bei den CLM-Läufen meist noch etwas größer aus als bei REMO. Die mittleren Niederschläge werden während nasser Jahre noch deutlicher unterschätzt, hohe Niederschläge z.T. deutlicher überschätzt und Trockentage bzw. Trockenphasen werden ebenfalls deutlich unterschätzt.

4.3 Stationarität der Modellgüte

Eine wichtige Frage, die sich im Zusammenhang mit der Bewertung von Klimamodelldaten auf Basis eines Vergleichs mit Beobachtungsdaten in einem ausgewählten Zeitraum stellt, ist, ob sich die identifizierte Modellqualität näherungsweise stationär verhält oder ob sie, je nach Betrachtungszeitraum, variiert. Diese Erkenntnisse können Hinweise darauf geben, ob die identifizierte Güte auch auf zukünftige Zeiträume übertragbar erscheint. Um die Eigenschaften der Modelldaten in dieser Studie hinsichtlich dieser Fragestellung einzuschätzen, wurde die jeweilige Modellgüte für ein gleitendes Zeitfenster innerhalb des Zeitraumes 1961–2000 analysiert. Aufgrund der natürlichen Klimavariabilität wird nach Linke et al. (2014) empfohlen, Klimadaten über einen Zeitraum von wenigsten 30 Jahren zu betrachten. Dieser Empfehlung folgend wurde die Übereinstimmung von Beobachtungs- und Modelldaten für elf 30-Jahres-Zeitfenster (1961-1990, 1962-1991, …, 1971-2000) anhand der Prüfgröße *z(U)* des U-Tests analysiert. Die Ergebnisse werden in Abbildung 8 für die Temperatur-Kennwerte und in Abbildung 9 für die Niederschlags-Kennwerte dargestellt.

Es zeigt sich, dass die Modellgüte von REMO für Tmq90 relativ konstant bleibt, während sie sich für die anderen beiden Temperatur-Kenngrößen tendenziell sogar leicht verbessert. Die CLM-Läufe scheinen die Anzahl der Kältetage (Tncd) mit der Zeit etwas schlechter abzubilden, ansonsten bleibt auch bei ihnen die Modellgüte relativ konstant. Gleiches gilt auch für die Daten von WETTREG2006. Als Ursache für die mögliche Veränderung der Modellgüten kann die Anzahl der Beobachtungsstationen, die für die Interpolation des Referenzdatensatzes genutzt wurden, nahezu ausgeschlossen werden, da diese sich im Zeitraum von 1961 bis 2000 für ganz Niedersachsen nur von 60 auf 68 verändert hat. Als Fazit kann festgehalten werden, dass Aussagen über die zukünftige Entwicklung der Anzahl der Kältetage (Tncd) anhand des betrachteten Modell-Ensembles in dieser Hinsicht tendenziell eher mit Unsicherheiten behaftet sind als die Entwicklung der mittleren bzw. hohen Temperaturen.

Beim Niederschlag zeigen die Betrachtungen aller Kennwerte ein relativ konstantes Verhalten bei allen Modellketten. Eine Ausnahme bildet der mittlere Niederschlag (Pav1), bei dem sich die Modellgüte bei den dynamischen Modellen REMO und CLM mit der Zeit leicht verbessert. Insgesamt scheinen Aussagen über die zukünftige Entwicklung dieser Werte aber im Hinblick auf die Veränderung der Modellgüte mit relativ geringen Unsicherheiten behaftet zu sein.

Diese Ergebnisse geben keine Garantie dafür, dass sich die identifizierte Modellgüte für den Referenzzeitraum auch auf den Zukunftszeitraum übertragen lässt. Sie zeigen aber, dass diese Annahme nicht von vorneherein ausgeschlossen werden muss.

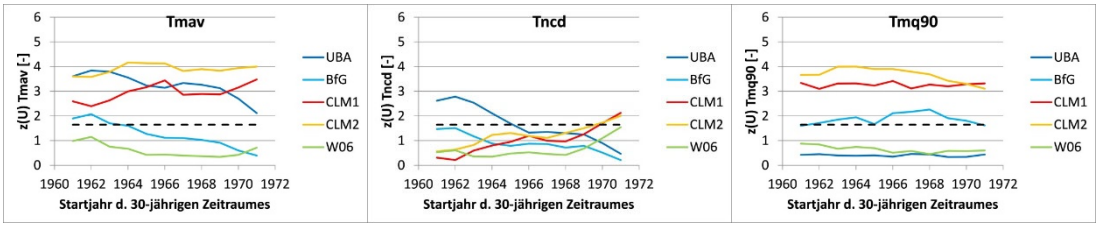


Abbildung 8: zeitlicher Verlauf der Modellgüte für die Temperatur-Kennwerte (Mittelwert über alle 10 Teilgebiete), gemessen anhand der Prüfgröße des U-Tests (z(U)), für ein gleitendes 30jähriges Fenster über den Zeitraum 1961-2000; die gestrichelte Linie markiert die Signifikanzgrenze (bei einer Irrtumswahrscheinlichkeit von α = 0,05)

Trend of model efficiency for indices of temperature (mean of all 10 subbasins), on the basis of the result of the U-test for a moving window of 30 years over 1961 – 2000); the dashed line indicates the margin of significance

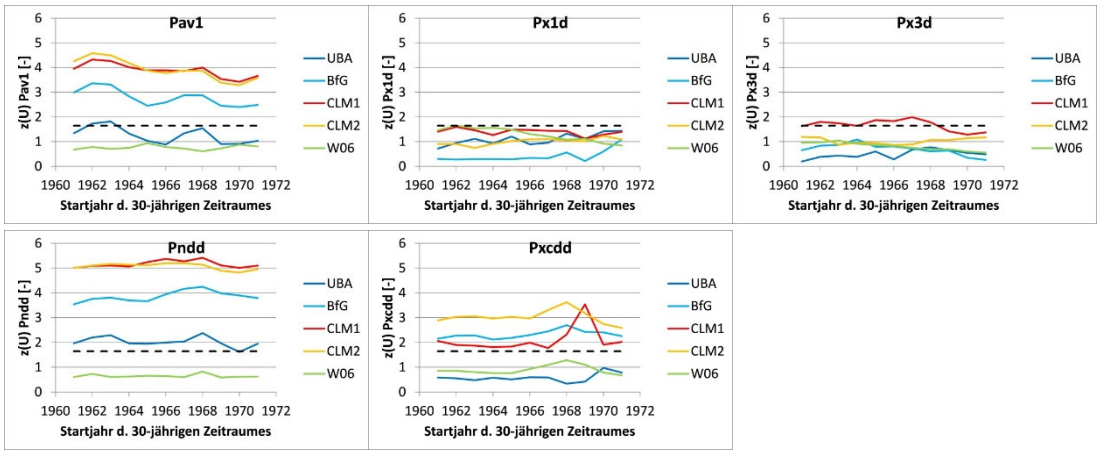


Abbildung 9: zeitlicher Verlauf der Modellgüte für die Niederschlags-Kennwerte (Mittelwert über alle 10 Teilgebiete), gemessen anhand der Prüfgröße des U-Tests (z(U)), für ein gleitendes 30jähriges Fenster über den Zeitraum 1961-2000; die gestrichelte Linie markiert die Signifikanzgrenze (bei einer Irrtumswahrscheinlichkeit von α = 0,05)

Trend of model efficiency for indices of precipitation (mean of all 10 subbasins), on the basis of the result of the U-test for a moving window of 30 years over 1961 – 2000); the dashed line at indicates the margin of significance

4.4 Diskussion der Ergebnisse

Abbildung 10 und Abbildung 11 geben einen zusammenfassenden Überblick bzgl. der Übereinstimmung von Beobachtung und Modellen anhand der Prüfgröße *z(U)* des U-Tests. Als Nullhypothese wird in diesem Zusammenhang angenommen, dass die Werte der zwei Stichproben, also die 40 Jahreswerte von Beobachtung bzw. Modell, statistisch gesehen der gleichen Grundgesamtheit angehören. Die Prüfgröße *z(U)* des U-Tests zeigt, mit welcher Irrtumswahrscheinlichkeit diese Hypothese abgelehnt werden kann, d.h. wie sicher gesagt werden kann, ob sich Beobachtung und Modell signifikant voneinander unterscheiden. Die Farben der Felder in den Abbildungen symbolisieren die unterschiedlichen Signifikanzniveaus, bei denen eine Übereinstimmung von Beobachtung und Modell abgelehnt würde. Die Anordnung der Gebiete von links nach rechts in der Grafik erfolgt mit absteigender Gebietsgröße (gerundet).

Es ist erkennbar, dass die Modellketten der dynamischen Modelle (ECHAM5 mit REMO bzw. mit CLM) bei der Nachbildung der beobachteten Temperatur-Kennwerte meist eine schlechtere Übereinstimmung aufweisen als das statistische Modell WETTREG2006 mit ECHAM5. REMO-UBA zeigt hierbei vor allem Schwächen bei der mittleren Temperatur (Tmav), teilweise auch bei der Anzahl der Kältetage (Tncd), während REMO-BfG tendenziell eher signifikante Abweichungen bei Tmav und bei den hohen Temperaturen (Tmq90) aufweist. Die CLM-Läufe zeigen im Vergleich zu den anderen Modellen die Abweichungen mit der höchsten Signifikanz, vor allem bei Tmav und Tmq90. Untereinander weisen sie keine wesentlichen Unterschiede auf. Räumliche bzw. skalenabhängige Schwerpunkte können bei den Temperatur-Kennwerten kaum identifiziert werden. Lediglich das Gebiet der Sieber wird bzgl. Tmav und Tncd am besten, und bzgl. Tmq90 am schlechtesten von allen Teilgebieten durch alle Modelle abgebildet.

Bei den Kennwerten des Niederschlags zeigt WETTREG2006 bei der Anzahl der Trockentage (Pndd) und der Dauer der maximalen Trockenphasen (Pxcdd) die besten Übereinstimmungen mit der Beobachtung von allen Modellen. Die dynamischen Modellketten zeigen hier größere Schwächen, wobei REMO-UBA noch die besten Werte erzielt. Die hohen Niederschläge (Px1d und Px3d) werden von REMO relativ gut abgebildet. CLM und WETTREG zeigen hier z.T. Schwächen. Der mittlere Niederschlag (Pav1) wird von allen Modellen in den größeren Teilgebieten tendenziell besser wiedergegeben, wobei WETTREG die geringsten Übereinstimmungen mit der Beobachtung aufweist. Insgesamt weisen die Modelle auch bei den Niederschlags-Kennwerten praktisch keine skalenabhängigen Muster bzgl. der Güte auf. Eine Ausnahme bildet der mittlere Niederschlag, der in kleineren Gebieten (< 1.000 km^2) eher schlechter abgebildet wird, und das Teilgebiet der Sieber, das als kleinstes und topographisch stark gegliedertes Gebiet stets die schlechteste Übereinstimmung zwischen Beobachtung und Modellen aufweist.

Die Modellgüten für die betrachteten Kennwerte verhalten sich innerhalb des Betrachtungszeitraumes relativ konstant, so dass Aussagen über die zukünftige Entwicklung dieser Werte im Hinblick auf die Veränderung der Modellgüte mit relativ geringen Unsicherheiten behaftet sind.

Diese Studie beschäftigt sich mit der Validierung abgeleiteter Gebietswerte aus Klimamodellketten als Eingangsgrößen für hydrologische Simulationen. Hierbei beeinflussen unterschiedliche Faktoren die Qualität der Modellergebnisse. Neben systemimmanenten Unsicherheiten in den jeweiligen Modellketten führen Limitierungen in der Auflösung und der räumlichen Verteilung der Klimaprojektionen in Zusammenwirken mit der Interpolation auf Gebietswerte zu regional unterschiedlich guten Ergebnissen im Vergleich zur hoch auflösenden Beobachtung. Während das statistische Modell WETTREG2006 für einzelne Stationen beispielsweise nahezu keinen Bias aufweisen sollte, ist ein solcher für die interpolierten Gebietswerte auf Basis von ECHAM5/MPI-OM vorhanden (Tabelle 3). Die dynamischen Modelle REMO und CLM zeigen bei der Nachbildung der beobachteten Temperatur-Kenngrößen signifikante Schwächen, wobei REMO die Werte überschätzt während CLM sie unterschätzt. Hier zeigt WETTREG2006 eine klar bessere Übereinstimmung.

Tmav	ALO	Aller	Leine	Ob. Leine	Ise	Nette	Schunter	Böhme	Südaue	Sieber
Größe (km²)	15700	7500	6500	1000	350	300	300	300	200	130
REMO-UBA										
REMO-BfG										
CLM1										
CLM2										
W06										

Tncd	ALO	Aller	Leine	Ob. Leine	Ise	Nette	Schunter	Böhme	Südaue	Sieber
Größe (km²)	15700	7500	6500	1000	350	300	300	300	200	130
REMO-UBA										
REMO-BfG										
CLM1										
CLM2										
W06										

Tmq90	ALO	Aller	Leine	Ob. Leine	Ise	Nette	Schunter	Böhme	Südaue	Sieber
Größe (km²)	15700	7500	6500	1000	350	300	300	300	200	130
REMO-UBA										
REMO-BfG										
CLM1										
CLM2										
W06										

Abbildung 10: Darstellung der Modellgüte anhand der Prüfgröße *z(U)* des U-Tests für alle Modelle und Kennwerte der Temperatur, differenziert nach Teilgebieten; die Farben symbolisieren die Signifikanz des Unterschiedes zwischen Beobachtung und Modell (der Unterschied ist nicht signifikant mit α>0,2 (grün), schwach signifikant mit α=0,2 (gelb), signifikant mit α=0,1 (orange), hoch signifikant mit α=0,05 (rot)).

Illustration of the model efficiency on the basis of the results of the U-test (z(U)) for all climate models, indices of temperature and subbasins; the color indicates the significance of the difference between observation and model (difference is not significant (α>0,2, green), slightly significant (α=0,2, yellow), significant (α=0,1, orange), very significant (α=0,05, red)).

Pav1	ALO	Aller	Leine	Ob. Leine	Ise	Nette	Schunter	Böhme	Südaue	Sieber
Größe (km²)	15700	7500	6500	1000	350	300	300	300	200	130
REMO-UBA										
REMO-BfG										
CLM1										
CLM2										
W06										

Px1d	ALO	Aller	Leine	Ob. Leine	Ise	Nette	Schunter	Böhme	Südaue	Sieber
Größe (km²)	15700	7500	6500	1000	350	300	300	300	200	130
REMO-UBA										
REMO-BfG										
CLM1										
CLM2										
W06										

Px3d	ALO	Aller	Leine	Ob. Leine	Ise	Nette	Schunter	Böhme	Südaue	Sieber
Größe (km²)	15700	7500	6500	1000	350	300	300	300	200	130
REMO-UBA										
REMO-BfG										
CLM1										
CLM2										
W06										

Pndd	ALO	Aller	Leine	Ob. Leine	Ise	Nette	Schunter	Böhme	Südaue	Sieber
Größe (km²)	15700	7500	6500	1000	350	300	300	300	200	130
REMO-UBA										
REMO-BfG										
CLM1										
CLM2										
W06										

Pxcdd	ALO	Aller	Leine	Ob. Leine	Ise	Nette	Schunter	Böhme	Südaue	Sieber
Größe (km²)	15700	7500	6500	1000	350	300	300	300	200	130
REMO-UBA										
REMO-BfG										
CLM1										
CLM2										
W06										

Abbildung 11: Darstellung der Modellgüte anhand der Prüfgröße *z(U)* des U-Tests für alle Modelle und Kennwerte des Niederschlags, differenziert nach Teilgebieten; die Farben symbolisieren die Signifikanz des Unterschiedes zwischen Beobachtung und Modell (der Unterschied ist nicht signifikant mit α>0,2 (grün), schwach signifikant mit α=0,2 (gelb), signifikant mit α=0,1 (orange), hoch signifikant mit α=0,05 (rot)).
Illustration of the model efficiency on the basis of the results of the U-test (z(U)) for all climate models, indices of precipitation and subbasins; the color indicates the significance of the difference between observation and model (difference is not significant (α>0,2, green), slightly significant (α=0,2, yellow), significant (α=0,1, orange), very significant (α=0,05, red)).

Die vergleichsweise schlechten Modellergebnisse von WETTREG2006 für die Niederschlags-Kenngrößen werden z.T. auf die in Abbildung 2 ersichtliche geringe Stationsdichte von WETTREG2006, speziell in kleineren Teilgebieten, zurückgeführt, z.T. auch auf das antreibende Globalmodell. Die dynamischen Modelle überschätzen die Niederschläge insgesamt und zeigen deutliche Schwächen bei den Trockenwetter-Kennwerten. Die Niederschlagsverhältnisse in kleineren Betrachtungsräumen scheinen subjektiv etwas schlechter von den Modellen abgebildet zu werden (Abb. 11). Ein Grund hierfür kann die hohe räumliche Variabilität des Niederschlags sein, die durch die geringe Gitterpunktdichte bzw. Stationsanzahl der Modelle gegenüber der Beobachtung nicht hinreichend abgebildet werden kann. Dies fällt in besonderem Maße bei CLM mit einer Auflösung von 18x18 km auf.. Ein vermuteter Zusammenhang zwischen Betrachtungsskala und Modellgüte auf Basis einer Korrelationsanalyse konnte jedoch für keine der Modellketten und für keine der Kenngrößen eindeutig nachgewiesen werden.

5 Auswirkung auf die hydrologische Impaktmodellierung

Für die Interpretation von Klimafolgenabschätzungen stellt sich die Frage, ob sich die aufgezeigten Charakteristika der Klimamodelldaten bei deren Nutzung als Eingabe für die Impaktmodellierung in den Ergebnissen in gleichem Maße fortsetzen oder ob sie im Verlauf der Modellkette verändert werden. Diese Unsicherheit muss identifiziert und soweit möglich quantifiziert werden. Dabei spielt es eine wichtige Rolle, wie sensitiv das Impaktmodell gegenüber Variationen bzw. Unsicherheiten einzelner meteorologischer Eingabe-Größen reagiert. Um diese Frage zu beantworten, wurden die aus den betrachteten Klimamodelldaten für die jeweiligen Teilgebiete resultierenden Abflüsse mittels des Wasserhaushaltsmodells PANTA RHEI modelliert.

5.1 Das hydrologische Modell PANTA RHEI

PANTA RHEI ist ein deterministisches hydrologisches Modellsystem (LWI-HYWAG und IfW 2012). Es wurde von der Abteilung Hydrologie, Wasserwirtschaft und Gewässerschutz am Leichtweiß-Institut (LWI) für Wasserbau der TU Braunschweig und der Institut für Wassermanagement GmbH, Braunschweig entwickelt. Das Modellsystem wird laufend für verschiedene Fragestellungen in der wasserwirtschaftlichen Praxis und der Forschung weiterentwickelt. Beispielsweise wird PANTA RHEI für die Hochwasservorhersage in Niedersachsen eingesetzt (Meyer et al. 2013).

In PANTA RHEI wird ein Einzugsgebiet in Teileinzugsgebiete unterteilt, die sich wiederum aus Hydrotopen oder Hydrological Response Units (HRUs) zusammensetzen. In einem Hydrotop werden Flächen gleicher hydrologischer Eigenschaften hinsichtlich Bodenarten, Landnutzung und ggf. weiterer Gebietseigenschaften innerhalb eines Teileinzugsgebietes zusammengefasst. Für die hydrologischen Prozesse, wie etwa Schneeakkumulation- und –schmelze (Förster et al. 2014, Förster und Meon 2012), Interzeption, Verdunstung, Bodenwasserhaushalt, Abflusskonzentration sowie Gewässerrouting, stehen je nach Fragestellung unterschiedlich komplexe Verfahren zur Verfügung. Zudem erlaubt PANTA RHEI auch die Berücksichtigung von hydrologisch relevanten wasserwirtschaftlichen Bauwerken wie Talsperren und Hochwasserrückhaltebecken.

Für das Untersuchungsgebiet von Aller und Leine wurde ein sehr detailliertes hydrologisches Modell mit PANTA RHEI erstellt. Das in Abschnitt 2 beschriebene Untersuchungsgebiet wurde dazu in ca. 4200 Teilflächen mit insgesamt ca. 62.000 Hydrotopen untergliedert. Dabei wurden auch 154 Pegelstationen, 6 Talsperren und zahlreiche weitere Bauwerke wie Düker und Überleitungen im Modell berücksichtigt. Das Modell wurde anhand von beobachteten Klima- und Pegeldaten für den Zeitraum 1961-1990 kalibriert und in einem unabhängigen Zeitraum von 1991 bis 2000 validiert. Je nach Pegel lag dabei die Modelleffizienz, bewertet anhand des *NSE* für die Tageswerte der Abflusszeitreihen, zwischen 0,76 und 0,88 (Hölscher et al. 2012).

5.2 Analyse ausgewählter hydrologischer Kennwerte

Die Abflüsse, basierend auf den Klimamodelldaten, wurden mit den modellierten Abflüssen der beobachteten Klimadaten aus dem Wasserhaushaltsmodell an den Referenzpegeln der Teilgebiete (vgl. Tabelle 1) für den gleichen Zeitraum (1961–2000)

verglichen und in gleicher Weise hinsichtlich ihrer Güte bewertet (vgl. Kapitel 3.3 und Abbildung 3). Als Kennwerte dienten die in Tabelle 4 angegebenen hydrologischen Indizes.

Tabelle 4: Ausgewählte Kenngrößen des Abflusses zur Bewertung der hydrologischen Modellierung
Selected indices of discharge for the evaluation of hydrological modelling

Abkürzung	Einheit	Beschreibung
MQ	m^3/s	Mittelwasserabfluss
HQ	m^3/s	Jahreshöchstabfluss
NM7Q	m^3/s	niedrigstes arithmetisches Mittel des Abflusses über 7 Tage

Für die Gegenüberstellung der Güte der Klimamodelle (hier Niederschlagsdaten) und der Güte der resultierenden hydrologischen Modellabflüsse wurde die Prüfgröße z(U) des U-Tests herangezogen. Die Bandbreite der Ergebnisse je Modell und Kennwert wird in Abbildung 12 dargestellt. Dabei wurde angenommen, dass die betrachteten Klimakenngrößen jeweils einen wesentlichen Einflussfaktor für die ausgewählten Abflusskenngrößen darstellen. Der Mittelwasserabfluss (MQ) würde demnach vor allem durch die mittleren Niederschläge (Pav1) beeinflusst, der Jahreshöchstabfluss durch die Maximalniederschläge (Px1d) und der Niedrigwasserabfluss (NM7Q) durch die maximale Dauer der Trockenperioden (Pxcdd). Weitere klimatische Größen, die in PANTA RHEI als Eingabedaten dienen, können einen zusätzlichen Einfluss auf die resultierenden Abflussgrößen besitzen, werden an dieser Stelle aber nicht weiter betrachtet. Auf die hydrologische Modellierung der Klimamodelldaten von CLM wurde in diesem Abschnitt verzichtet. Grund hierfür war die durchweg schwache bis sehr schwache Modellgüte der CLM-Läufe für die relevanten Niederschlags-Kenngrößen (vgl. Abbildung 11).

Der Vergleich der Bandbreite der Gütewerte von Modell-Niederschlag und resultierenden modellierten Abflüssen zeigt, dass eine mögliche Veränderung der Modellgüte im Zuge der hydrologischen Klimafolgenabschätzung tendenziell von dem betrachteten Kennwert abhängt sowie der Sensitivität des hydrologischen Modells gegenüber den klimatischen Kenngrößen, die als Eingabe dienen. So ändert sich beispielsweise die Güte beim MQ gegenüber der beim Pav1 kaum. Dies ist möglicherweise ein Hinweis darauf, dass das hydrologische Modell in diesem Fallbeispiel bei der Abbildung mittlerer Abflussverhältnisse auf kleinere Unsicherheiten bei den klimatischen Eingangsdaten weniger sensitiv reagiert. Bei den Kennwerten, die extremere hydrologische Verhältnisse widerspiegeln (HQ und NM7Q) kommt es dagegen eher zu einer Vergrößerung der Bandbreiten der Abweichungen und damit von *z(U)* gegenüber Px1d und Pxcdd. Die Unsicherheiten nehmen also tendenziell eher zu. Gründe hierfür können zum einen sein, dass neben den betrachteten Kenngrößen des Niederschlags noch andere Faktoren eine Rolle bei der Entstehung dieser Abflusskenngrößen spielen, etwa die Bodenfeuchte, Grundwasser oder die saisonale Schneedecke, die in der Gegenüberstellung in Abbildung 12 nicht berücksichtigt wurden. Zum anderen können auch die nicht-linearen Zusammenhänge beim Niederschlag-Abfluss-Prozess, bzw. deren Art der Abbildung im Modell, ein Grund für diese Veränderungen sein. Daher reagiert das hydrologische Modell auf Abweichungen bei klimatischen Kenngrößen, die extreme Abflussbedingungen verursachen, sensitiver.

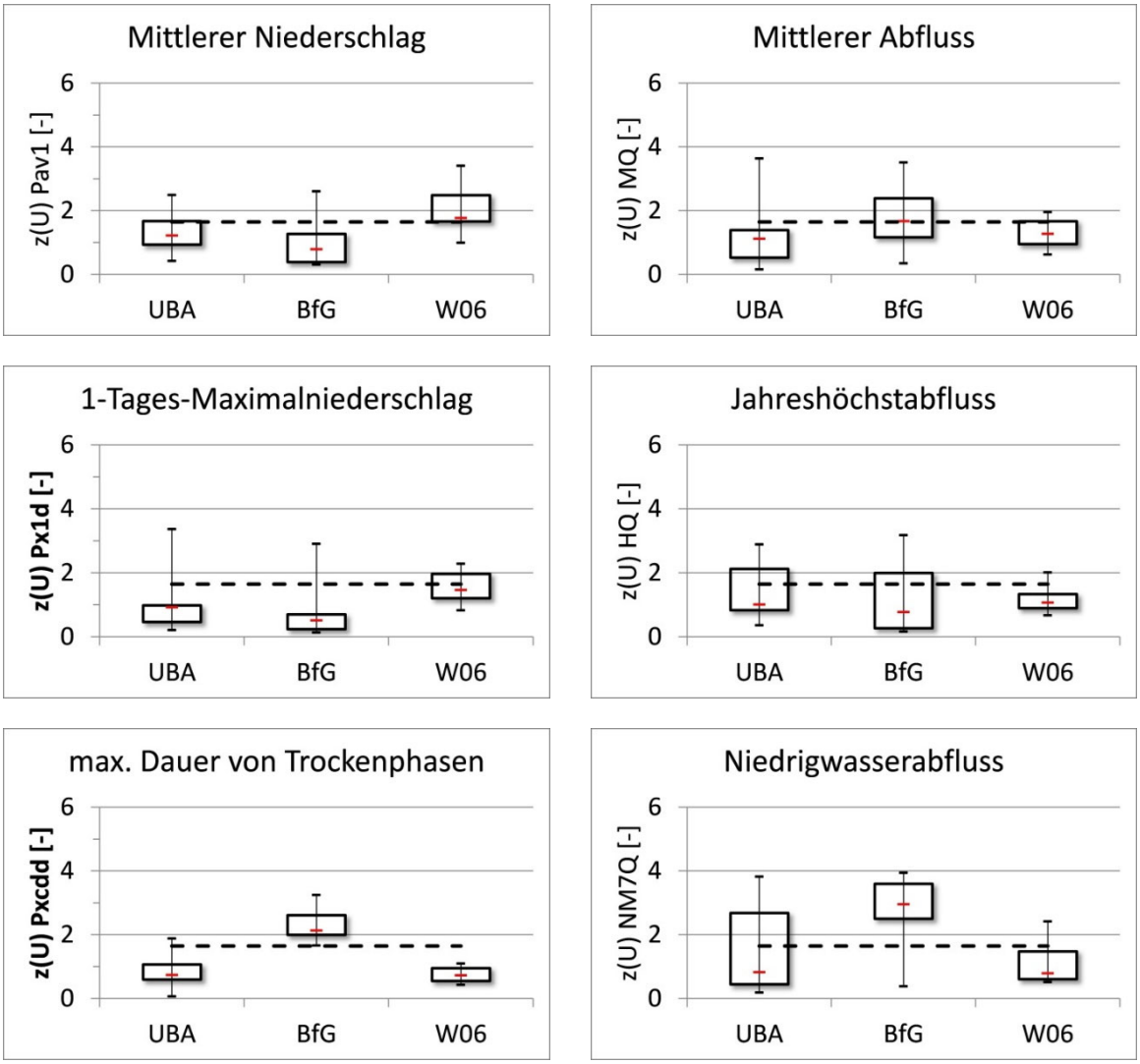


Abbildung 12: Bandbreiten der Modellgüte über alle 10 Teilgebiete bzw. Pegel (bewertet anhand der Prüfgröße *z(U)* des U-Tests gegenüber der Beobachtung) für ausgewählte Kennwerte des Niederschlags der Klimamodelldaten (links) und der resultierenden modellierten Abflüsse aus PANTA RHEI (rechts) für den Zeitraum 1961–2000; die gestrichelte Linie markiert die Signifikanzgrenze (bei einer Irrtumswahrscheinlichkeit von $\alpha = 0{,}05$)

Band width of model efficiency for all 10 subbasins and gauges (test criteria z(U) of U-test in comparison to the observation) for selected indices of precipitation of climate models (left) and simulated discharge of PANTA RHEI (right) for time period 1961-2000; the dashed line indicates the margin of significance

6 Zusammenfassung und Schlussfolgerung

Die vorliegende Studie beschäftigte sich mit der Analyse der Güte von regionalen Klimamodelldatensätzen. Das Ziel war ein einfacher, übertragbarer Ansatz, der es den Anwendern in der Praxis ermöglichen soll, eine erste fallspezifische Beurteilung von Klimamodelldaten bzgl. ihrer Eignung für ausgewählte Fragestellungen in der Impaktmodellierung vorzunehmen. Die Methodik wird anhand der Datensätze von REMO, CLM und WETTREG2006, alle angetrieben durch das Globalmodell ECHAM5/MPI-OM, am Fallbeispiel des Einzugsgebietes von Aller und Leine vorgestellt. Der Ansatz ist nicht

modellspezifisch entwickelt worden und kann daher auch für die Einschätzung anderer Modellkonfigurationen von globalen und regionalen Klimamodellen herangezogen werden.

Die Bewertung anhand von Beobachtungsdaten für den Zeitraum 1961-2000 (20C) mittels verschiedener Kennwerte und Gütekriterien zeigte für das Untersuchungsgebiet, dass die Qualität der betrachteten Modellketten und damit die Aussagen darauf basierender Klimafolgenabschätzungen je nach Kenngröße und Modell differenziert beurteilt werden müssen. Die Bewertung anhand von Beobachtungsdaten mittels verschiedener Gütekriterien zeigt, dass die Aussagen der Klimamodelle und die Aussagen von Klimafolgenabschätzungen je nach Kenngröße und Modell differenziert beurteilt werden müssen. Abweichungen der Klimamodelldaten gegenüber der Beobachtung gab es auf Basis von Gebietskennwerten vor allem bei

- extremen Ausprägungen einer Kenngröße, wie z.B. trockenen Jahren, großen Maximalniederschlägen, langen Trockenphasen
- Temperatur- und Trockenwetter-Indizes
- mittleren und maximalen Niederschlägen in kleineren Betrachtungsräumen (< 1.000 km^2)

Während die dynamischen Modelle Schwächen bei der Abbildung der beobachteten Temperatur-Kennwerte und den Trockenwetter-Indizes aufwiesen, vor allem CLM, zeigten sie bei den mittleren und maximalen Niederschlägen relativ gute Übereinstimmungen. Das statistische Modell WETTREG2006 hatte bei diesen Klimagrößen tendenziell größere Probleme, speziell bei den mittleren Niederschlägen mit mehr als 1mm. Die Temperaturen hingegen wurden sehr gut wiedergegeben. Die Güte der Modelle verhält sich innerhalb des Referenzzeitraumes relativ stationär, so dass eine Übertragung der Erkenntnisse bzgl. der Modellgüte von der Gegenwart auf die Zukunft plausibel erscheint.

Die Qualität der Klimamodelldaten spiegelt sich auch in den simulierten Abflüssen der anschließenden Impaktmodellierung mithilfe des hydrologischen Modellsystems PANTA RHEI wider. Im Detail war die Güte der simulierten versus der beobachteten Mittelwasserabflüsse vergleichbar mit der Güte des mittleren Niederschlags. Bei extremen Abflussverhältnissen, wie Hoch- und Niedrigwasser, nahm die Bandbreite der Modellgüte im Vergleich zu den korrespondierenden Werten von hohen Niederschlägen bzw. langen Trockenphasen zu. Anhand des Wilcoxon-Mann-Whitney-Tests (auch U-Test genannt) wurde für ein Signifikanzniveau von 95% nachgewiesen, dass die Modellketten ECHAM5/MPI-OM → REMO bzw. WETTREG2006 in Kombination mit einem hydrologischen Impaktmodell (PANTA RHEI) geeignet erscheinen, um die wasserwirtschaftlichen Kenngrößen mittlerer Abfluss (MQ) und Jahreshöchstabfluss (HQ) im Referenzzeitraum zufriedenstellend abzubilden. Beim Niedrigwasserabfluss zeigte REMO z.T. deutliche Schwächen und unterschied sich signifikant von der Beobachtung. Dieser Effekt ist mit sich überlagernden Teilprozessen in der hydrologischen Modellierung erklärbar sowie deren Nichtlinearität und Persistenz, wonach meteorologische Extreme nicht notwendigerweise entsprechende hydrologische Extreme zur Folge haben.

Die möglichen Ursachen für die aufgezeigten Abweichungen zwischen Modell- und Beobachtungsdaten sind vielfältig. Neben der Modellunsicherheit in der Modellkette

Globalmodell (ECHAM5/MPI-OM) → Regionalmodell (REMO, CLM bzw. WETTREG2006) → Impaktmodell (PANTA RHEI) kommt auch das „Pre-processing“ der Daten in Frage. Dies bezieht sich zum einen auf die Korrektur möglicher Messfehler beim Niederschlag als auch zum anderen auf die Interpolation zu Gebietsmittelwerten, basierend auf stationsbasierten bzw. gerasterten Originaldaten unterschiedlicher Datendichte. Die Validierung von Gebietskennwerten von Klimavariablen liefert in diesem Zusammenhang zusätzliche Erkenntnisse für flächenbezogene Anwendungen wie in der Einzugsgebietshydrologie und ergänzt somit die direkte Validierung der Klimadaten.

Weitere Unsicherheiten, die in dieser Studie nicht betrachtet wurden, erschweren zusätzlich die Interpretation der Erkenntnisse bei Klimafolgenabschätzungen. Die Unschärfe der internen Klimavariabilität, wie sie in dieser Studie anhand von zwei Läufen eines Globalmodells dargestellt wurde, ist in diesem Zusammenhang nur eine Möglichkeit. Weitere Unsicherheiten ergeben sich aus anderen Global-, Regional- und Impaktmodell-Kombinationen, anderen Referenzdaten, einer möglichen Bias-Adjustierung der Modelldaten oder der Wahl von Emissionsszenarien. Diese Unsicherheiten rufen Bandbreiten in den Projektionen möglicher zukünftiger Klimafolgen hervor. Diese müssen kommuniziert und als Entscheidungsgrundlage für Anpassungsmaßnahmen, z.B. Low-regret-Maßnahmen, herangezogen werden. Sie können ergänzt werden durch qualitative Einschätzungen über die generelle Tendenz (Richtung) von Klimasignalen. Eine exakte quantitative Bewertung von Klimafolgen, wie es von Politikern oder Planern häufig gewünscht wird, erscheint dagegen kaum vertretbar.

Summary and conclusion

This study analysed the quality of the meteorological data fields obtained from regional climate models (RCM). The objective was to provide a simple approach to assess climate model data with respect to its adequacy for hydrological impact modeling. The method is presented using the example of three regional climate models - including the statistical model WETTREG2006 and the dynamical models REMO and CLM, all driven by the global climate model ECHAM5/MPI-OM. Therefor the catchment of the rivers Aller und Leine (located in Lower Saxony, Germany) was taken as a study area. The approach can also be used for the assessment of any other combination of different global and regional climate model chains.

The evaluation based on data observed in the period 1961–2000 by means of different indices and efficiency criteria for the study area, shows that the results of climate and impact modelling have to be assessed carefully. RCM data may differ from the observation based on areal values especially in terms of

- extreme manifestation of indices, e.g. dry years, high maximum precipitation, long lasting dry spells
- indices of temperature and dry conditions
- mean and maximum precipitation in small catchments (< 1.000 km^2)

While displaying deviations in reproducing observed temperature and dry condition indices, especially CLM, the dynamical models showed relatively good conformance regarding mean and maximum precipitation. Due to its statistical nature, the WETTREG model showed

good performance for all indices. Deviations in mean daily precipitation of more than 1 mm rainfall might be related to the varying number of precipitation stations contained in the modelling and monitoring data sets as well as the driving global model. In the other hand, WETTREG provided excellent temperature reproductions. The accuracy of the models remains relatively stationary for a moving 30-year window within the period 1961–2000, hence transferring present findings respectively quality of the model chains to the future seems conceivable.

The range of quality of the RCM data can also be recognized in the results of the following impact modelling with the help of the hydrological modeling system PANTA RHEI. The goodness of the simulated discharge during mean flow periods was comparable to the goodness of the corresponding rainfall from climate modeling. The bandwidth of goodness concerning simulated extreme discharge, such as floods or low flow, however, is increasing in comparison to the corresponding rainfall and dry conditions from climate models. By means of the Wilcoxon-Mann-Whitney test (also referred to as U-test), it has been evidenced for a significant level of 95%, that the model chains ECHAM5/MPI-OM → REMO resp. WETTREG2006 combined with a hydrological impact model (PANTA RHEI) seem suitable for the satisfactory reproduction of the water management indices mean discharge (MQ) and annual peak discharge (HQ). In terms of low flow conditions, however, REMO occasionally displayed significant deviations in comparison to observations. These effects can be caused by the more complex, in some cases overlapping hydrological processes and their descriptions required in modeling extreme flow conditions.

There are various reasons for the deviations shown between observed and modelled data. Apart from the model uncertainty within the model chain global climate model (ECHAM5/MPI-OM) → regional climate model (REMO, CLM and WETTREG2006) → impact model (PANTA RHEI), this study also indicates the preprocessing of data. On the one hand, there is the correction of the observational error, on the other hand, one can note the interpolation to areal values, and the different spatial resolution of the datasets as possible causes. Therefore the validation of climate model data as areal indices provides additional knowledge for spatial applications such catchment hydrology, in addition to the direct validation of the original climate model data.

Further uncertainties cause difficulties in evaluating the results. Climate variability, shown in this survey by two runs of a global climate model, is one reason. Additional uncertainties may be caused by other model chains, different reference datasets, bias adjustment of model data or different emission scenarios. These uncertainties prevent an accurate prediction of climate change as well as of climate impact. Hence, they have to be communicated and taken into account in terms of climate change adaptation, e.g. by low-regret-measures. A quantitative accurate evaluation of climate impact, as requested by politics and planners would like to obtain, seems to be hardly reasonable.

Danksagung

Die vorliegenden Untersuchungen erfolgten im Rahmen des Forschungsprojektes KliBiW (Globaler Klimawandel – wasserwirtschaftliche Folgenabschätzungen für das Binnenland), finanziert durch das Niedersächsische Ministerium für Umwelt, Energie und Klimaschutz,

sowie im Rahmen des Forschungsverbundes KLIFF im Teilprojekt KLIFWA (Auswirkungen von Klimaänderungen auf Wasserdargebot, Hochwasserrisiko und Gewässerbelastung in Niedersachsen), finanziert durch das Niedersächsische Ministerium für Wissenschaft und Kultur. Die Bereitstellung der beobachteten Klimadaten erfolgte mit freundlicher Unterstützung des Deutschen Wetterdienstes (DWD).

Literaturverzeichnis

Blöschl, G., Viglione, A., Merz, R., Parajka, J., Salinas, J.L. und Schöner, W. (2011): Auswirkungen des Klimawandels auf Hochwasser und Niederwasser: Springer Verlag, S. 21-30.

Bronstert, A., Kolokotronis, V., Schwandt, D. and Straub, H. (2007): Comparison and evaluation of regional climate scenarios for hydrological impact analysis: General scheme and application example. Int. J. Climatol. 27: 1579 – 1594 (2007), DOI: 10.1002/joc.1621

Chen, J., Brissette. F.P., Chaumont, D. and Braun, M. (2013): Performance and uncertainty evaluation of empirical downscaling methods in quantifying the climate change impacts on hydrology over two North American river basins. Journal of Hydrology, 479, p. 200 – 214.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K. and Liebert, J. (2012): „Should we apply bias correction to global and regional climate model data?". Hydrol. Earth Syst. Sci., 16, 3391-3404. www.hydrol-earth-syst-sci.net/16/3391/2012/ doi: 10.5194/hess-16-3391-2012

Förster, K., Meon, G., Marke, T., Strasser, U. (2014): Effect of meteorological forcing and snow model complexity on hydrological simulations in the Sieber catchment (Harz Mountains, Germany). Hydrology and Earth System Sciences Volume 18, Issue 11, Pages 4703-4720.

Förster, K., Meon, G. (2012): Assessing the model performance of snow water resources simulated by a coupled mesoscale atmospheric and hydrologic model. EGU General Assembly 2012, 22 - 27 April 2012, Vienna. http://presentations.copernicus.org/ EGU2012-1731_presentation.pdf

Fowler, H. J., Blenkinsop, S. and Tebaldi, C. (2007): Linking climate change modeling to impacts studies: recent advances in downscaling techniques for hydrological modeling. International Journal of Climatology, 27, p. 1547 – 1578.

Graham, L.P., Hagemann, S., Jaun, S. and Beniston M. (2007): On interpreting hydrological change from regional climate models. Climatic Change, 81(Suppl. 1), 97-122.

Haberlandt, U., Wallner, M., Dietrich, J., Meon, G., Gelleszun, M., Förster, K., Gocht, M., Ptak, T., Sauter, M., Herold, M., Lange, S., Heuer, A., Krause, F., van der Heijden, S., Maier, N., Rosenwinkel, K.-H., Spering, V., Verworn, F., Stein, K., Hölscher, J., Anhalt, M., Forberg, C., Haas, A., Restemeyer, B. und Gerkensmeier, B. (2014): Klimafolgenforschung in Niedersachsen – Forschungsthema 6: Binnengewässer – Auswirkungen von Klimaänderungen auf Wasserdargebot, Hochwasserrisiko und Gewässerbelastung in Niedersachsen (KLIFWA). Abschlussbericht, Hannover.

Haberlandt, U., Heijden, S.v.d., Verworn, A., Berndt, C., Dietrich, J., Wallner, M. und Krause, F. (2013): Regionalisierung von Klimabeobachtungsdaten und WETTREG-Szenarien für Niedersachsen als Grundlage für mittel- bis großskalige Modellierungen. Institut für Wasserwirtschaft, Hydrologie und landwirtschaftlichen Wasserbau, Hannover.

Hölscher, J., Petry, U., Bertram, M., Anhalt, M., Schmidtke, S., Haberlandt, U., Müller, H., van der Heijden, S., Berndt, C., Verworn, A., Wallner, M., Belli, A., Dietrich, J., Meon, G., Förster, K., Gelleszun, M., Riedel, G., Lange, A., und Eggelsmann, F. (2012): Globaler Klimawandel - Wasserwirtschaftliche Folgenabschätzung für das Binnenland (KliBiW). Oberirdische Gewässer, Band 33, NLWKN, Norden.

Huang, S., Krysanova, V.and Hattermann, F.F. (2013a): Projection of low flow conditions in Germany under climate change by combining three RCMs and a regional hydrological model. Acta Geophysica 61(1), 151-193.

Huang, S., Hattermann, F.F., Krysanova, V. and Bronstert, A. (2013b): Projections of climate change impacts on river flood conditions in Germany by combining three different RCMs with a regional eco-hydrological model. Climatic Change 116(3-4), 631-663.

Jacob, D., Nilson, E., Tomassini, L. & Bülow, K. (2009): REMO climate of the 20th century run and A1B scenario run, BfG project, 0.088 degree resolution. World Data Center for Climate. CERA-Database. http://cera-www.dkrz.de/WDCC/ui/Project.jsp?proj=REMO-BFG

Jacob, D., Göttel, H., Kotlarski, S., Lorenz, P. und Sieck, K. (2008): Klimaauswirkungen und Anpassung in Deutschland - Phase 1: Erstellung regionaler Klimaszenarien für Deutschland. UBA Forschungsbericht 204 41 138, v. UBA-FB 000969, p. 1 – 159.

Jacob, D., Bärring, L., Christensen, O., Christensen, J., de Castro, M., Déqué, M., Giorgi, F., Hagemann, S., Hirschi, M., Jones, R., Kjellström, E., Lenderink, G., Rockel, B., Sánchez, E., Schär, C., Seneviratne, S., Somot, S., van Ulden, A. and van den Hurk, B. (2007): An inter-comparison of regional climate models for Europe: model performance in present-day climate: Climatic Change, v. 81, p. 31-52.

Kay, A. L., Davies, H. N., Bell, V. A. and Jones, R. G. (2009): Comparison of uncertainty sources for climate change impacts: flood frequency in England. Climatic Change, 92 (1-2), p. 41-63.

Kay, A., Bell, V. and Davies H. (2006): Model Quality and Uncertainty for Climate Change Impact. Centre for Ecology and Hydrology, Wallingford.

Kotlarski, S., Block, A., Böhm, U., Jacob, D., Keuler, K., Knoche, R., Rechid, D. and A.Walter (2005): Regional climate model simulations as input for hydrological applications: evaluation of uncertainties: Advances in Geosciences, v. 5, p. 119-125.

Krause, P., Boyle, D. P. and Bäse, F. (2005): Comparison of different efficiency criteria for hydrological model assessment. Advances in Geosciences, 5, p. 89 – 97.

Legates, D. R. and McCabe, G. J. (1999): Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resources Research, 35 (1), p. 233 – 241.

Linke, C. et al. (2014): Leitlinien zur Interpretation regionaler Klimamodelldaten des Bund-Länder-Fachgespräches „Interpretation regionaler Klimamodelldaten". Bremen. http://klimawandel.hlug.de/?id=448 (30.07.2014)

LWI-HYWAG und IfW (2012): Panta Rhei Benutzerhandbuch - Programmdokumentation zur hydrologischen Modellsoftware (unveröffentlicht). Abteilung Hydrologie, Wasserwirtschaft und Gewässerschutz am Leichtweiß-Institut für Wasserbau der TU Braunschweig in Kooperation mit dem Institut für Wassermanagement IfW GmbH, Braunschweig.

Mann, H.B., Whitney, D.R. (1947): On a test of wether one of two random variables is stochastically larger than the other. Annals of Mathematical Statistics 18, 50 – 60.

Meyer, S., Riedel, G., Lichtenberg, T., Meon, G., Lange, S. (2013): Operationelle Hochwasservorhersage in Niedersachsen mit dem Wasserhaushaltsmodell Panta Rhei. In: Casper, M. und Gronz, O. (Hrsg.): Simulation hydrologischer Systeme - Wie nah kommen wir der Realität? Beiträge zum 3. Trierer Workshop zur Niederschlag-Abfluss-Modellierung am 17. und 18. September 2012 in Trier. Forum für Hydrologie und Wasserbewirtschaftung, Heft 33.13, 59-66.

Merz, B., Maurer, T. und Kaiser, K. (2012): Wie gut können wir vergangene und zukünftige Veränderungen des Wasserhaushalts quantifizieren? Hydrologie und Wasserbewirtschaftung, 56, S. 244 – 255.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Binger, R. L., Harmel, R. D. and Veith, T. L. (2007): Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. American Society of Agriculture and Biological Engineers, 50 (3), p. 885 – 900.

Mudelsee, M., Chirila, D., Deutschländer, T., Döring, C. Haerter, J.O., Hagemann, S., Hoffmann, H., Jacob, D., Krahe, P., Lohmann, G., Moseley, C., Nilson, E., Panferov, O., Rath, T., Tinz, B. (2010): Climate Model Bias Correction und die Deutsche Anpassungsstrategie. Mitteilungen DMG 3, 2-7.

Muerth, M. J., Gauvin St-Denis, B., Ricard, S., Velázquez, J. A., Schmid, J., Minville, M., Caya, D., Chaumont, D., Ludwig, R., and Turcotte, R. (2013): On the need for bias correction in 25 regional climate scenarios to assess climate change impacts on river runoff, Hydrol. Earth Syst. Sci., 17, 1189–1204.

Müller P. (2010): Constructing climate knowledge with computer models. WIREs Climate Change 1 (4), p. 565 – 580.

Nash, J.E. and Sutcliffe, I.V. (1970): River Flow Forecasting through Conceptual Models - Part I - A Discussion of Principles. Journal of Hydrology, v. 10, p. 282-290.

Prein, A.F., Truhetz, H., Suklitsch, M., and Gobiet, A. (2011): NHCM-1: Non-hydrostatic climate modelling, Part III: Evaluation of the Local Climate Model Intercomparison Project (LocMIP) simulations, University of Graz.

Prudhomme, Ch.; Davies, H. (2009): Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 2: future climate. Climatic Change, 93 (1-2), p. 197-222.

Refsgaard, J. C., Madsen, H., Andréassian, V., Arnbjerg-Nielsen, K., Davidson, T. A., Drews, M., Hamilton, D. P., Jeppesen, E., Kjellström, E., Olesen, J. E., Sonnenborg, T. O., Trolle, D., Willems, P. and Christensen, J. H. (2014): A framwork for testing the ability of models to project climate change and its impacts. Climatic Change, 122(1-2), 271-282

Richter, D. (1995): Ergebnisse methodischer Untersuchungen zur Korrektur des systematischen Meßfehlers des Hellmann-Niederschlagsmessers. Berichte des Deutschen Wetterdienstes, Volume 194, Main, Deutscher Wetterdienst, p. 1-47.

Rockel, B., Will, A., Hense, A. (2008): The Regionale Climate Model COSMO-CLM. Meteorologische Zeitschrift, Vol. 17, No. 4, 347 - 348

Roeckner, E., Bäuml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, M. Hagemann, S., Kirchner, I., Kornblueh, L., Manzini, E., Rhodin, A., Schlese, U., Schulzweida, U., Tompkins, A. (2003): The atmospheric general circulation model ECHAM 5. PART I: Model description, ISSN 0937 - 1060

Sachs, L. (2004): Angewandte Statistik: Anwendung statistischer Methoden. Springer-Verlag, ISBN 3540405550, 889 Seiten.

Schoetter R, Hoffmann P, Rechid D, Schlunzen K.H. (2012): Evaluation and bias correction of regional climate model results using model evaluation measures. J. Appl. Meteor. Climatol., 51, 1670–1684

Spekat, A., Enke, W., und Kreienkamp, F. (2007): Neuentwicklung von regional hoch aufgelösten Wetterlagen für Deutschland und Bereitstellung regionaler Klimaszenarios auf der Basis von globalen Klimasimulationen mit dem Regionalisierungsmodell WETTREG auf der Basis von globalen Klimasimulationen mit ECHAM5/MPI-OM T63L31 2010 bis 2100 für die SRES Szenarios B1, A1B und A2. UBA Forschungsbericht 204 41 138, p. 1-149.

Weber, J. (2009): Auswertung/Verwendung regionaler Klimaszenarien im Vorhaben KLIWA. – In: Fachvorträge vom 4. KLIWA-Symposium „Klimaänderung und Konsequenzen für die Wasserwirtschaft“ in Mainz. – Hrsg. KLIWA Arbeitskreis, KLIWA-Berichte 15, 71-84

Wegehenkel, M., Heinrich, U., Jochheim, H., Kersebaum, K.C. and Röber, B. (2010): Evaluation of three different regional climate change scenarios for the application of a water balance model in a mesoscale catchment in Northeast Germany: Adv. Geosci., 27, 57-64, 2010, www.adv-geosci.net/27/57/2010/ doi:10.5194/adgeo-27-57-2010.

Wilcoxon, F. (1945): Individual comparisons by ranking methods. Biometrics Bulletin 1, 80 – 83.

Dieses Kapitel ist eine editierte Fassung der folgenden wissenschaftlichen Originalveröffentlichung:

Petry, U.; Dietrich, J.; Förster, K.; Wallner, M.; Berndt, C.; Meon, G.; Haberlandt, U. (2015): Ein Ansatz zur Validierung von Klimamodelldaten als Basis für die Interpretation von wasserwirtschaftlichen Klimafolgenabschätzungen in Niedersachsen. Hydrologie und Wasserbewirtschaftung 59 (4), 155–173.

Herausgeber: Bundesanstalt für Gewässerkunde

Lizenz: Die Genehmigung zur Zweitveröffentlichung in der Habiliationsschrift wurde erteilt (der Originalartikel ist frei zugänglich)

VI. Using SWAT for Strategic Planning of Basin Scale Irrigation Control Policies: a Case Study from a Humid Region in Northern Germany

Maier, N.[1], Dietrich, J.[2]

(1) Institute for Landscape Ecology and Resources Management, Research Centre for BioSystems, Land Use and Nutrition (IFZ), Justus Liebig University Giessen, Germany
(2) Institute of Water Resources Management, Hydrology and Agricultural Hydraulic Engineering, University of Hannover, Germany

Abstract

The eco-hydrological model SWAT is used worldwide for simulating hydrology and water quality of agricultural catchments. One of the main water uses is irrigation, predominantly in arid and semi-arid regions. Climate impact simulations show that a future increase of irrigation demand can be expected for humid regions. Options for adaptation include the improvement of irrigation techniques and the modification of crop patterns. In our study we investigate the application of SWAT for the development of water saving irrigation control strategies in a humid river catchment in Northern Germany. We developed different scenarios using both soil moisture deficit control and plant water demand control. The results show plausible changes of irrigation amounts when changing the trigger points of both control methods. By deficit control strategies, the water consumption could be reduced with only a moderate decrease of crop yield. Differences between soil characteristics were well shown in the SWAT simulations, but the model consistently overestimated irrigation values. Furthermore we found a high variability of the model errors between the different years, even if the long term average values are considered acceptable. Future research is needed to improve the model accuracy in automatic irrigation control.

1 Introduction

Plants need water for their growth. Hence, agriculture depends on natural availability of water or supports production by irrigation. Irrigation requires strategic planning regarding the investment in structures, which might cover large areas, e.g. when canals are built or large quantities of water are abstracted from aquifers. During the growing season, operational aspects like irrigation scheduling come into play. Farmers need to optimize multi criteria like water use and crop productivity, while under water scarcity also social criteria of water sharing are important (Moghimi and Sepaskhah 2016). Climate change will most likely exert pressure on farmers, in particular when small-scale farms run out of economic profitability due to reduced water availability (Fernández et al. 2016). Strategic and operational aspects can be combined for the development of water management policies e.g. for climate change adaptation (Brookfield and Gnau 2016).

The irrigation water requirement of the crop is the quantity of water that must be applied in order to fill the deficit between the crop water requirement and the natural availability of water. Thus, it mainly depends on evapotranspiration and on soil properties. Due to the non-stationary nature of soil moisture and plant growth, the amount and timing of irrigation has influence on crop yield and quality (Ali and Talukder 2001). Deficit irrigation strategies can save water (Zhang et al. 2002).

In central and northern Europe irrigation is used to improve yield in times of dry summers or water limitations, especially during the initial crop stage (Wriedt et al. 2009). Due to the humid climate of these regions irrigation is mainly applied on intensively cultivated farmland (Mostafa and Derbala 2013). Climate change impact projections reveal a growing future demand for irrigation in humid regions (Kreins et al. 2015). In Germany in the year 2002, 142 Mm^3 water were used for irrigation. About two-third was applied to farmland in the federal state of Lower Saxony (Statistisches Bundesamt 2004). Detailed field records of the scheduling of irrigation operations are only available for experimental sites. In Lower Saxony, recommendations for farmers are given regularly based on soil moisture and weather forecast (Fricke and Riedel 2013).

In order to simulate irrigation water demand, models must be capable of simulating water transfer through soil, vegetation and atmosphere (Santhi et al. 2005). A wide range of models has been developed to calculate irrigation requirements for crops at the field scale, e.g. CROPWAT (Smith 1992) and SWAP (van Dam et al. 2008), which solves the Richards' equation for vertical water movements. Eco-hydrological catchment models like the semi-distributed SWAT model (Arnold et al. 1998) implement irrigation schemes, but use simplified approach for soil water dynamics. Gayley (2013) analyzed the effective annual irrigation rates based on plant needs with SWAT for two catchments in South Central Texas and showed an over-application of water by 5% to 14%.

In this study, we demonstrate and discuss the application of SWAT for the large scale planning of water saving irrigation strategies in an intensively irrigated river catchment in a humid region. We investigate different irrigation control scenarios for different combinations of soils and crops and their consequences on crop yields. We use the automatic irrigation operations within SWAT.

2 Methodology

2.1 Study Area and Data

The Fuhse river catchment in Northwest Germany covers an area of about 991 km^2 (Fig. 1). The investigation area in the lower part is dominated by glacial sediments of the Northern German moraines (Geest), i.e. sands with very low water holding capacities. The most dominant soils according to the soil map BÜK 1000 are:

- Fluvisols/Gleysols (B10, available water capacity AWC=80 mm/m),
- Spodo-Stagnic Cambisols/Stagnic Podzoluvisols (B28, AWC=154 mm/m),
- Cambic Podzols/Spodic Arenosols (B31, AWC=86 mm/m),
- Haplic Podzols/Dystric Regosols (B33, AWC=128 mm/m) and
- Haplic Luvisols/Eutric Podzoluvisols/Stagnic Gleysols (B42, AWC=201 mm/m).

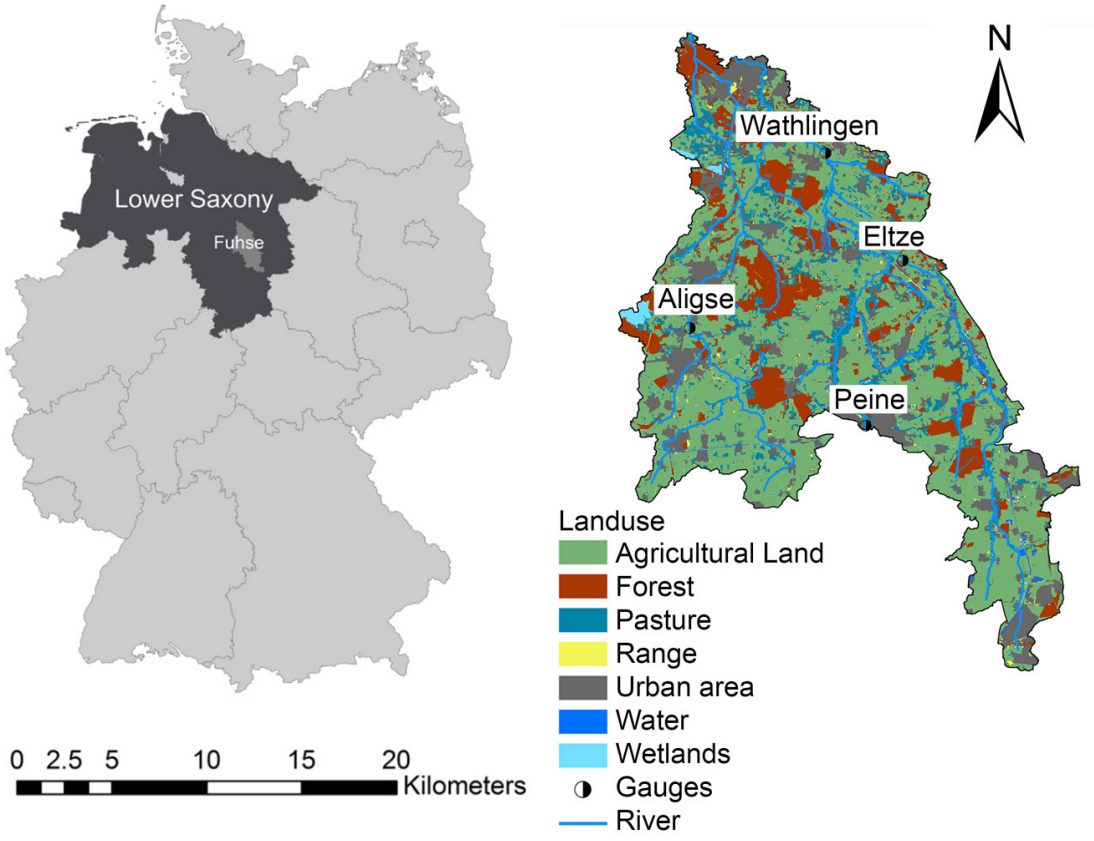


Fig. 1. Geographic location (left) and land use map of the lower Fuhse catchment (right)

The mean annual temperature is between 8.5 and 9.0°C (winter: 0–1°C, summer: 16–17°C). The annual precipitation varies between 600 and 700 mm. The climate variables temperature, precipitation, relative humidity, sunshine duration and wind velocity from the German Weather Service (Deutscher Wetterdienst) were interpolated on a 1 km × 1 km grid (Petry et al. 2015). Discharge values of four gauging stations in the catchment were available on a daily time step.

The land use of the catchment is dominated by arable land (55%), forest (16%), urban area (14%), pasture (12%), water (2%) and range land (1%) according to CORINE Land Cover (CLC2006, Federal Environment Agency, DLR-DFD 2009). For the agricultural land, crop distributions were available from the agricultural statistics of Lower Saxony (Agrarstrukturerhebung, tables Z6070421, K6070411, K6070412 and K6080014 with data from 1977 to 2007 reported every four years, Niedersächsisches Landesamt für Statistik).

2.2 Simulating Irrigation Operations with SWAT

The process-based continuous time model SWAT can simulate water fluxes, plant growth and land management operations in agricultural catchments (Arnold et al. 1998). Gassman et al. (2007) provide a comprehensive overview of model studies done with SWAT. In this paper, we will focus on irrigation in SWAT. Irrigation can be scheduled in SWAT either by fixed schemes ("manually") or automatically. The source of irrigation has to be defined as reach, reservoir, shallow aquifer, deep aquifer or a source outside the watershed. Automatic irrigation adds water until field capacity is reached and the excess water is returned to the source. Automatic irrigation can be triggered by either a water stress threshold or a soil water deficit threshold. If irrigation

gets triggered by water stress based on plant water demand, water stress threshold is a fraction of potential plant growth. Water stress is simulated in SWAT by a comparison of actual and potential plant transpiration:

$$wstr = 1 - \frac{E_{t,act}}{E_t} = 1 - \frac{w_{actualup}}{E_t}, \quad (1)$$

where wstr is the water stress, E_t is the maximum plant transpiration, $E_{t,act}$ is the actual amount of transpiration and $w_{actualup}$ is the total plant water uptake. The plant water uptake is a function of the maximum plant transpiration, the water-use distribution parameter, the depth of the soil layer and the depth of root development. The water stress factor thus represents soil and crop conditions. Irrigation is triggered when the water stress factor falls below a predefined value and fills the soil up to field capacity, given the source of irrigation provides enough water. Larger predefined trigger points indicate wetter soils when irrigation starts and vice versa.

With regard to the soil water deficit, irrigation is triggered for each hydrologic response unit (HRU, unique combination of soil type, land use and slope within each sub-basin) when the soil water in the total soil column "falls below field capacity by more than the soil water deficit threshold" (Neitsch et al. 2011). One should differentiate between the total field capacity and the water that is available for the plant in the rooting zone (AWC). Considering that soil types have different total field capacities, and crops have different rooting depths, the AWC differs for crop–soil combinations. Therefore individual maximum depletion values of soil water, before irrigation starts, have to be calculated according to the soil and crop type. Consequently the irrigation depth for different crops on the same soil can vary depending on their rooting depth. Thompson et al. (2007) showed the uncertainties and the influence of the selected soil depth used to calculate the AWC. Commonly the upper 60 cm of the soil layer are used as a reference to initiate irrigation (Liu et al. 2010). Other field experiments start irrigation when soil moisture content at 10 cm depth is below 25 vol.% (Mostafa and Derbala 2013).

2.3 Model Setup, Calibration and Irrigation Scenarios

The SWAT model consists of 52 sub-basins and 9178 HRUs (average area of one HRU = 10.81 ha). The agricultural HRUs were separated into 50 different classes with a size of approximately 5 ha, which were equally distributed over the catchment, in order to allow a non-stationary distribution of crops via the SWAT management operations. With the Hungarian method (Kuhn 1955), possible crop rotations based on the federal state's data were developed and implemented on the agricultural areas. Therewith some crops are overestimated by the crop generator: winter wheat by 14%, triticale by 2%, rye with about 2.4%, winter barley by 1% and rape by 1.6%. On the contrary the remaining crops are underestimated by 0.11% (barley), 1.6% (oats), 2.7% (potatoes), 14% (sugar beet) and by 0.2% (corn sillage).

Table 1. Calibration parameters for the catchment

SMFMX	Snow Melt Factor on June 21st	5.5
SMFMN	Snow Melt Factor on December 21st	5.5
ESCO	Soil Evaporation Compensation Factor	0.8
EPCO	Plant Uptake Compensation Factor	0.5
SURLAG	Surface runoff lag Coefficient	1
MSK_CO1	Muskingum Calib-Coeff to control the impact of storage time constant for Normal Flow	0.25

MSK_CO2	Muskingum Calib-Coeff to control the impact of storage time constant for low Flow	0.3
MSK_X	Muskingum weighing factor that controls the relative importance of inflow and outflow for reach storage	0.25
OV_N	Manning's value for overland flow	0.01-2.4
DEP_IMP	Depth of Impervious Layer	3450*, 4200**, 5900***
CH_N1	Manning's roughness coefficient for Tributary Channel	0.036*, 0.065**, 0.06***
CH_K1	Effective Hydraulic conductivity for Tributary Channel	0.02*, 0.02**, 0.01***
CN2	Initial SCS curve number for soil moisture condition	-25%
CH_N2	Manning's roughness coefficient for Main Channel	0.058
CH_K2	Effective Hydraulic conductivity for Tributary Channel	1.38*, 1.15**, 0.3***
GWQMN	Shallow aquifer threshold depth for return flow	45
GW_REVAP	Ground Water Revap Coefficient	0.075*, 0.02**, 0.05***
REVAPMN	Shallow aquifer threshold depth for percolation	2
ALPHA_BF	Alpha Factor Base flow	0.07*, 0.005**, 0.035***
GW_DELAY	Ground Water Delay time	180*, 30**, 60***
DRAIN_D	Depth to subsurface drain	800
DRAIN_T	Time to drain soil to field capacity	48
DRAIN_G	Drain tile lag time	48

*subbasins contributing to Aligse Gauging station, **subbasins contributing to Eltze and Wathlingen gauging stations, ***subbasins contributing to Wathlingen gauging station

For the case study the catchment of the lower Fuhse was calibrated and validated manually. The inflow from upstream of the Peine gauging station was taken from previous simulation. Calibration parameters are shown in Table 1. During the calibration process the model considered only basic agricultural management operations such as planting and harvesting. Agricultural management options were refined after the calibration process. Furthermore, drainage was implemented on agricultural land with a slope of 0% to 2% (Battermann and Theuvsen n.Y., Tetzlaff et al. 2009). A Nash-Sutcliffe efficiency (NSE) of 0.68 was achieved for the calibration period (1984–1994 for the gauging station Wathlingen). For the validation period (1995–2004) a NSE of 0.70 was achieved (Fig. 2). Through the implementation of the different irrigation scenarios, the percentage bias of –5% of the calibration period improved between 1% and 2%. For the validation period the bias changed from –2% to 0.3%.

The schedules of the planting and harvesting operations were implemented by date according to Kutschera et al. (2009) for the different crops in the catchment of the case study. Hence it was ensured, first, that the winter crops are planted in fall and harvested in late summer of the next year and, second, that there is no overlapping between the harvesting time and the planting time of the following crop. Automatic fertilizer application was activated to ensure that the growth of the crops is not limited by nitrogen stress.

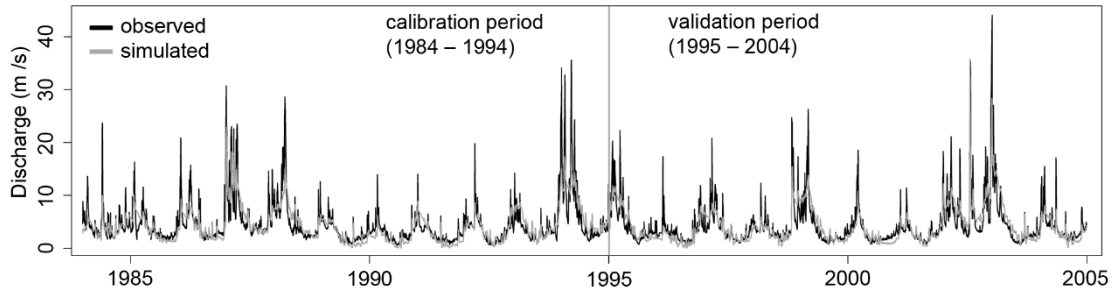


Fig. 2. Observed and simulated daily discharge for the calibration period (1984–1994) and the validation period (1995–2004)

Irrigation was implemented as an automatic scheme triggered by both water stress identifiers, plant water demand and soil water content. The threshold values for irrigation triggered by soil water content (AUTO_WSTRS) define the depletion of the soil water when irrigation starts. Field experiments in the area of the case study contemplate in practice two trigger points: "optimal irrigation" starting at 50% depletion of AWC (S1) and "reduced irrigation" starting at 65% depletion of AWC (S2). For plant water demand, scenarios with threshold values of 0.9 (S3), 0.8 (S4) and 0.7 (S5) for the trigger parameter AUTO_WSTRS were applied. As further scenarios, to quantify the effect of irrigation on yield, AUTO_WSTRS was set from 30% to 70% depletion of AWC (in 5% steps). The soil was filled up to 80% of the AWC (IRR_MX). In all scenarios the efficiency of the irrigation (IRR_EFF) was set to 75%, which is comparable to sprinkler irrigation (Brouwer et al. 1989), because in the region of Peine, on more than 90% of the irrigated area sprinkler irrigation is used (Landesamt für Statistik 2010). Surface runoff was set to 10% (IRR_ASQ). As a source for the irrigation, the shallow groundwater table of each subbasin was used (IRR_NOA). In the following, we give gross irrigation values (including the water losses of the irrigation system).

3 Results and Discussion

3.1 Evapotranspiration for the Catchment

The SWAT model was calibrated against stream flow, which is the best observed hydrological variable in the case study area. For irrigation demand simulations, evapotranspiration (ET) is important. Measured ET values were not available for the study. Thus, we compare the SWAT simulated ET of selected locations of the catchment (A = HRU with permanent grass, B = HRU with winter wheat, winter barley and grass in annual rotations) with daily grids of actual evapotranspiration obtained from the DWD (Löpmeier 1994). This data set represents simulated evapotranspiration over grass and sandy loam with daily update and serves as a proxy for observations. HRUs, which contain the soil B28 are selected from the SWAT model. For the HRU with permanent grass (A) a NSE of 0.64 in average is calculated and a bias of 15 %. For the HRU with crop rotations crop coefficients according to the crop and the crop state were applied (winter wheat and winter barley: $K_{c\,ini}$= 0.7, $K_{c\,mid}$ = 1.15, $K_{c\,end}$ = 0.25, sugar beet: $K_{c\,ini}$= 0.4, $K_{c\,mid}$ = 1.25, $K_{c\,end}$ = 0.7). In average a NSE of 0.34 is calculated for the HRU (B), and in specific a (modified) NSE of 0.44 for winter wheat and of 0.28 for winter barley. Winter wheat and winter barley are mostly overestimated at high evaporation values. Simulation with no irrigation results in the highest NSE (0.39) and simulation for plant water triggered irrigation scheduling (S3) in the lowest NSE (0.27). It should be noted that the DWD dataset does not regard irrigation. Thus it is plausible to obtain better performance for the non-irrigated scenarios.

3.2 Irrigation depths for the catchment

For the different irrigation schemes the irrigation depths for the whole catchment were simulated for the years 1971 to 2008 (Fig. 3). The earlier the soil water content falls below the trigger point, the higher the irrigation depth in total (50% depletion of rooting zone (S1) > 65% depletion of rooting zone (S2)). The mean average irrigation depth for the optimal irrigation schemes S1 is 107.02 mm (standard deviation SD: 45.74 mm). The mean average irrigation depth

for scenario S2 with reduced irrigation is 82.69 mm (SD: 43.31 mm). With regard to a moving average over 7 years, the difference between the deficit irrigation and the optimal irrigation is about 24 mm. The correlation coefficient between the soil water deficit irrigation scenarios (S1 and S2) and between the plant water demand scenarios (e.g. S3 and S4) is r>0.95, whereas the correlation between the plant water demand irrigation scheme and the soil water deficit irrigation schemes (e.g. S1 and S3) is between 0.6 and 0.7. The SWAT User Manual recommends a trigger point of 0.9 for irrigation triggered by plant water demand. Using this trigger point, the irrigation depth is on average 128 mm (min 60 mm, max 238 mm) higher than in the irrigation scheme triggered by soil water. Changing the trigger point for the plant water demand from 0.9 to 0.8 causes an average yearly decrease in irrigation depth of 45 mm and a trigger point change from 0.8 to 0.7 an average yearly decrease of further 36 mm. The results from the soil water deficit irrigation scenario recommend a trigger point of 0.7. The yields of both types of irrigation control are comparable, with exception scenario S5, which shows a difference of up to 7% (year 2003) to the scenario S1.

For Farmers a local water abstraction of 80 mm in a 7-year average is allowed by the authorities (Battermann and Theuvsen n.Y.). All scenarios exceed the 80 mm limit between 1971 and 2008 on a 7-year average. S2 exceeds the limiting value in 37% of the simulated time period (years 1977 to 1987). S1 is below 80 mm only from 1980 to 1984.

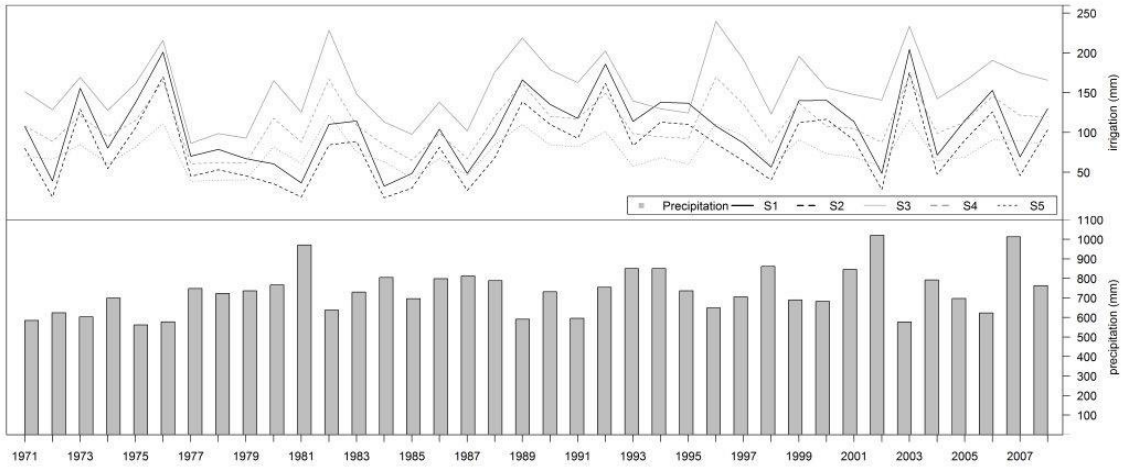


Fig. 3. Yearly irrigation depth and yearly precipitation from 1971 to 2008 for the different irrigation scenarios

3.3 Irrigation Depths for Different Crops

Notably is the difference in irrigation depths between the irrigation schemes (S1 – S5) when differentiating by crops and soil types. For low-rooting crops (potatoes, winter wheat, winter barley), the mean annual irrigation depth (1971 – 2008) simulated by the plant water demand is higher than for the soil water scenarios. For deep-rooting crops (barley, sugar beet) the irrigation depth simulated by soil water deficit is higher than that simulated by plant water demand (Fig. 4).

The different available water capacities of the soils have an impact on the water storage for plant use and thus the water amounts needed for irrigation. For winter wheat, sugar beet and winter barley the average irrigation depths triggered by plant water demand and soil water deficit show high variability, depending on the soil. For potatoes the difference in the irrigation depths

for scenarios triggered by soil water and plant water is larger in the soils B31, B10 and B33 with low AWC than in the soils B42 and B28 with higher AWC. Despite the differences in AWC, the simulated average irrigation depths for the different soils vary in S1 and S2 only between 86 mm and 127 mm, whereas the variation in S3 to S5 is between 88 and 499 mm. The difference in the irrigation depths for the crops in the different soils is much smaller for the scenarios triggered by soil water deficit (winter wheat: 46 mm, winter barley: 31 mm, potatoes: 56 mm, sugar beet: 37 mm, barley: 33 mm) than for those triggered by plant water demand (winter wheat: 250 mm, winter barley: 337 mm, potatoes: 410 mm, sugar beet: 218 mm, barley: 118 mm).

From an experimental site (Hamerstorf, 52°54'30.00''N, 10°28'2.00''E, approx.. 60 km north of the case study area outlet), annual irrigation amounts for the years 2006 to 2008 for winter barley, winter wheat and potatoes are available (Fricke and Riedel 2013) and can be compared with the simulated irrigation amounts from the case study on the same soil type (Table 2). Simulated irrigation amounts are higher than reported irrigation amounts from the experimental site (on average 50 mm for winter barley, 22 mm for winter wheat and 10 mm for potatoes).

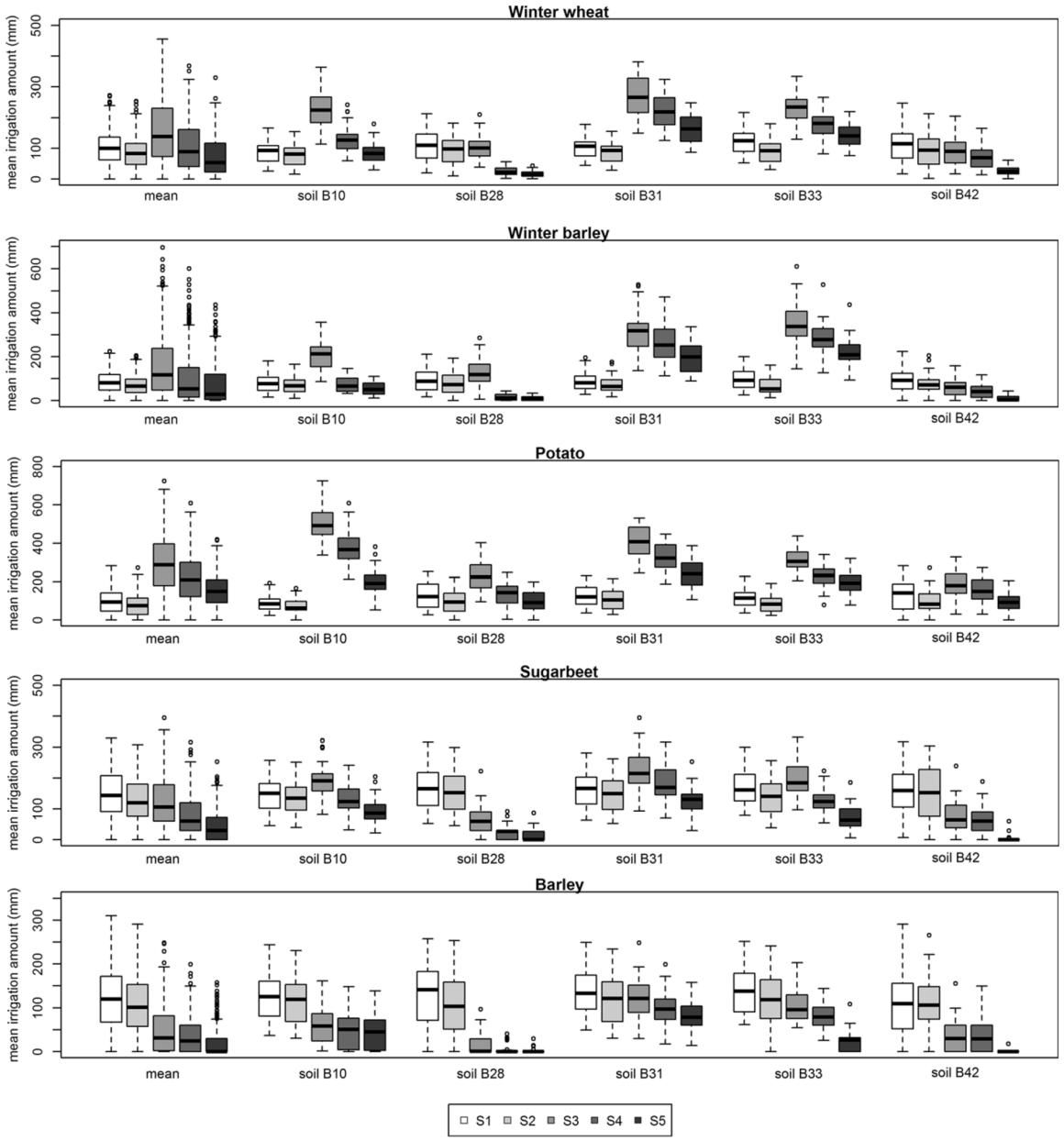


Fig. 4. Average irrigation depth (mm/year) for the years 1971-2008 of different crops for different simulation scenarios

3.4 Crop yields

Yields from the catchment area can be found in statistical reports for the years 1991 to 2008 (Landesamt für Statistik 2010). SWAT overestimates the yield of winter wheat in average by 8%, winter barley by 2 %, sugar beet by 27% and barley by 46%. Potatoes are underestimated in average by 10.4 %. The yield depends on the soil: winter wheat shows differences of 69 to 76 % between the five main soil types in the catchment for non-irrigated scenarios and scenarios S1-S5 (Table 3). Differences for winter barley are between 47 and 50% and for potatoes between 12 and 21%. Sugar beet and barley show differences of 18 and 20% for the non-irrigated scenario. In the irrigated scenarios (S1-S5) the differences minimizes to 7-9% for sugar beet and 2-6% for barley. Crops achieve the highest yield on soil B42, which shows the highest AWC (201 mm/m). Only for barley the highest yields are achieved on soil type B28, followed by soil B42. B28 represents the soil with the second highest AWC (154 mm/m). To evaluate the best soil water deficit irrigation scheme, irrigation was triggered at soil water depletions of 70%, 65%, 60%, 55%, 50%, 45%, 40%, 35% and 30% of AWC and refilled up to 80% of the AWC. Figure 5 shows the corresponding change in yield in the different scenarios compared to an exclusively rain-fed scenario for the five main crops and main soil types for the years 2000 to 2008. For all crops the applicable law is, if the depletion of the AWC is smaller when irrigation is triggered, the irrigation depth and the yield increase. The greatest changes in yield due to irrigation are simulated for potatoes. On soil B28 the yield increases even up to 28% from the no irrigation scenario to the 30% depletion scenario, whereas for all other crops an increase of less than 5% on this soil is simulated. On soil B42 the increase in the yield of potatoes is simulated by about 8% to 10%, whereas the yields of all other crops increase just less than 5%. On Soil B31, the yield of all crops increases, due to irrigation, more than 11%.

Table 2. Comparisons of the irrigation depth and yield for the last years (2006-2008) and for winter wheat, winter barley and potatoes with the experimental site Hamerstorf

year		Winter wheat			potatoes			Winter barley		
		2006	2007	2008	2006	2007	2008	2006	2007	2008
Irrigation amount (mm) Experimental site Hamerstorf	*S1*	118	85	167	201	75	141	60	55	139
	S2	58	25	86	118	25	58	30	25	58
Irrigation amount (mm) SWAT	*S1*	148	71	123	137	49	105	130	77	133
	S2	119	46	95	112	16	68	101	57	107
Deviation of irrigation amount (%)	*S1*	26	-16	-26	-32	-35	-26	117	41	-5
	S2	106	84	11	-5	-36	18	236	129	84
Yield (dt/h) Experimental site Hamerstorf	*S1*	6.3	6.8	4.9	13.8	15.2	14.4	6.7	6.7	5.6
	S2	7.1	7.1	7.1	15.0	15.0	15.0	6.8	6.8	6.8
	Non irrigated	5.7	5.6	4.4	8.7	14.3	12.0	4.8	4.8	4.0
Yield (dt/h) SWAT	*S1*	8.3	8.4	8.3	9.8	8.7	8.2	6.0	6.0	5.9
	S2	8.3	8.4	8.3	10.0	8.7	8.3	6.0	6.0	6.0

	Non irrigated	5.9	6.1	5.9	6.5	6.3	6.4	6.0	6.0	5.9
Deviation of yield (%)	*S1*	30	23	68	-29	-43	-43	-10	-10	7
	S2	18	19	18	-33	-42	-45	-11	-12	-12
	Non irrigated	3	9	35	-26	-56	-47	25	25	46

There are different correlations between the yield increases from the no irrigation scenario to an irrigated scenario for different soil types. Whereas a linear correlation between the irrigation depth and the change in yield due to irrigation is simulated for the soils B10, B28 and B42, a more curve linear relation for the other soil types is simulated. A linear relationship between yield and applied water was also shown for corn in a Mediterranean semiarid climate by Irmak et al. (2000). Yazar et al. (2002) found a linear and a curvilinear relationship between water use and yield for cotton on a clay-textured soil in an arid region. Both studies considered only one year of experiments (1995 and 1999). Cetin and Bilgel (2001) found, in the same region as Yazar et al. (2002), quadratic relationships between applied water and cotton yield for three different irrigation methods (furrow, sprinkler, drip) as a mean for the years 1991 to 1994.

These different curve shapes show that an increase in irrigation depth does not necessarily produce the same increase in yield. For winter wheat, winter barley, sugar beet and barley on soils B28 and B42 the simulated yield increase is almost the same for the different irrigation depths. The relation between the yield increase and the irrigation depths for crops on soil B10 is almost linear, so that here it can be stated that larger irrigation amounts produce greater yields. Hence, the relevance of yield increase and water saving has to be considered and balanced.

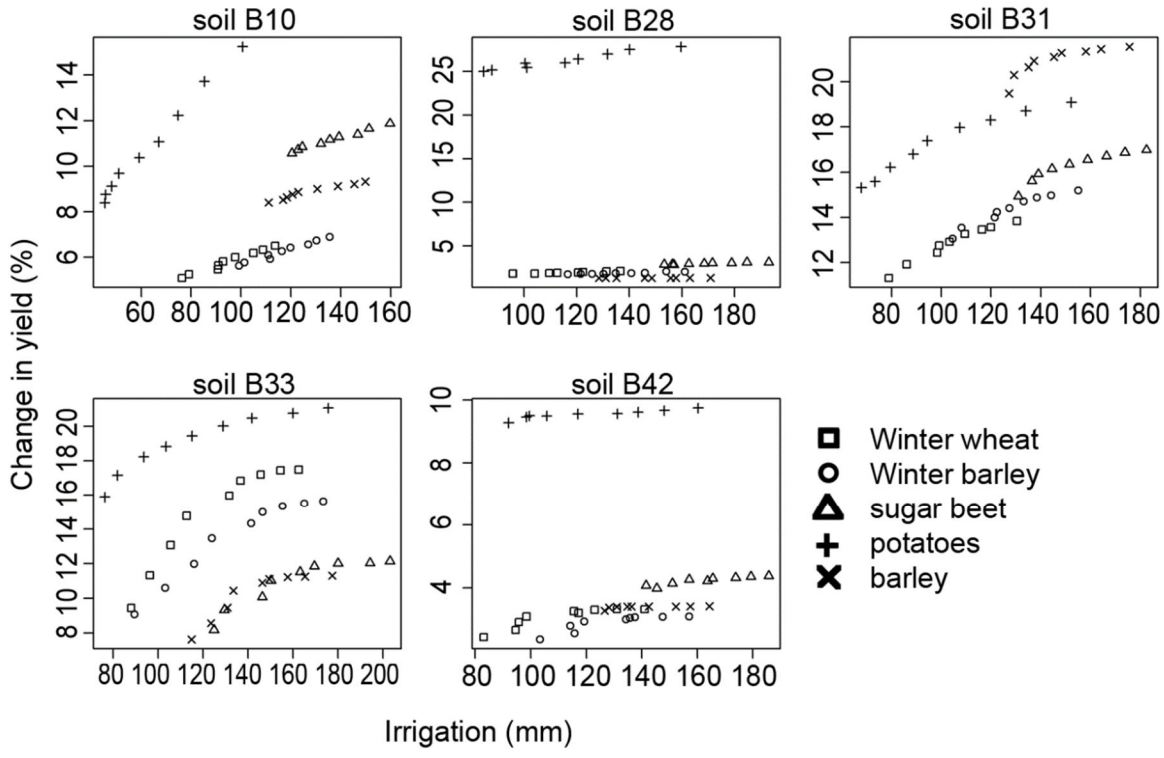


Fig. 5. Change in yield from non-irrigated scenario against irrigation depth for the most important crops

3.5 Optimized irrigation strategies

As previously stated, all simulation results show an increase in yield due to irrigation. On average the increase in yield from an exclusively rain-fed scenario (no irrigation) to the "reduced" irrigation scenario (depletion of 65% AWC) is 8%, whereas the further increase from the reduced to the "optimal" irrigation scenario (depletion of 50% AWC) is only about 1%.

Based on the average simulation results in yield change due to irrigation over the years 2000 to 2008, two different irrigation scenarios were set up. The trigger points were fixed to 65% depletion of AWC and 50% depletion of AWC, but only those crops were irrigated for which an increase of more than 5% in yield (Scenario R5) or more than 10% in yield (Scenario R10) compared to the rain-fed scenario was simulated on the different soils. Overall, for the scenario R5 all crops on soils B07 and B40 and in Scenario R10 crops on soils B06, B07, B11, B17, B40, B42 and B48 were no longer irrigated. All crops on soil B49 in the scenario R10 and on soils B31, B33, B49 and B70 in the scenario R5 stayed fully irrigated.

From the fully irrigated scenario to the scenario R5, an average yield reduction of 1.03% (S1) and of 1.17% (S2) was simulated. In scenario R10 a decrease in yield of 1.56% (S1) and 1.60% (S2) was simulated. Further a reduction in irrigation depth of 36 mm/year (30.9 %) (S1) and of 34 mm/year (33.5%) (S2) for scenario R5 and 54 mm/year (45.8%) (S1) and 48 mm/year (47.9%) (S2) for scenario R10 were simulated (Table 4). Winter wheat experiences the largest reductions in irrigation depth. The irrigation depth decreases by 58% to 70%, but the yield only by 2% to 2.3%. Irrigation depth for winter barley is reduced by about 39% and yield by 1.2%, whereas the irrigation depth reduction for potatoes of 11% leads to a yield reduction of 1%. Thus much more water can be saved when reducing the irrigation depth for barley than for potatoes with the same reduction in yield.

Table 4. Percent change in yield and irrigation depth from fully irrigated scenario to scenario R5 and R10 for two different irrigation practices

		catchment	barley	potatoes	sugar beet	winter barley	winter wheat
R5 scenario	Yield (S1)	–1.03	–0.13	–0.05	–0.01	–0.85	–1.93
	Irrigation (S1)	–30.9	–3.2	0.1	–0.4	–30.1	–58.7
	Yield (S2)	–1.17	–0.42	–0.35	–0.37	–0.98	–1.94
	Irrigation (S2)	–33.5	–6.3	–6.3	–2.1	–33.9	–61.1
R10 scenario	Yield (S1)	–1.56	–0.58	–2.00	–0.45	–1.41	–2.30
	Irrigation (S1)	–45.8	–25.8	–16.2	–22.6	–44.3	–67.5
	Yield (S2)	–1.60	–0.79	–1.63	–0.69	–1.45	–2.26
	Irrigation (S2)	–47.9	–27.9	–19.2	–22.1	–47.4	–69.5

4 Conclusions

Major problems with respect to irrigation management with SWAT include the dependency of irrigation on the phenological state of the crop. Hansen et al. (2006) states that the timing of irrigation has a greater influence on the crop yield than the quantity of irrigation. For example,

irrigation can be considered as unnecessary during the maturation phase of the crops (Riediger et al. 2014). As irrigation in the SWAT model was implemented on the basis of heat units, irrigation was not applied from the time of seeding, because a specific heat accumulation has to have been achieved beforehand, and irrigation operations were applied until the harvest day. Thus irrigation amounts and yields do not necessarily represent the local conditions. The study was performed to demonstrate the application of an eco-hydrological model for irrigation planning at catchment scale only. The model shows in average an over-application of irrigation water compared to an experimental site close to the catchment, which confirms the findings of Gayley (2013). Notable is, that the deviation of irrigation amount is much larger than the deviation in yield. In two optimized scenarios, between 34 and 54 mm/year water can be saved. It was considered in the study that the water needed for irrigation is abstracted from the shallow aquifer. A comparison of the discharge and recharge of the shallow aquifer indicates that the loss in aquifer volume due to water extraction for irrigation can be reduced by 17 to 35%, when irrigation scheduling is optimized as proposed in this study.

Further work will be related with the parameterization of local crops to reduce the deviation of simulated yields, with investigations about the actual irrigation practice of farmers, and with studies about climate change adaptation and sustainable use of water resources within the catchment.

Acknowledgements

We acknowledge the German National Weather Service (DWD), NLWKN Niedersachsen and the German Federal Institute for Geosciences and Natural Resources for providing data. Furthermore we would like to thank Florian Krause and Seema Samad for their support, Sarah Collins for proof reading and two anonymous reviewers for their valuable comments which helped to improve the manuscript.

References

Ali MH, Talukder MSU (2001) Methods or Approaches of Irrigation Scheduling – An Overview. J. of the Institute of Engineers, Bangladesh 1(28):11–23

Arnold JG, Srinivasan R, Muttiah RS, Williams JR (1998) Large area hydrologic modeling and assessment part I: Model development. Journal of the American Water Resources Association 34(1):73–89

Battermann HW, Theuvsen L: Feldberegnung in Nordost-Niedersachsen. Regionale Bedeutung und Auswirkung differenzierter Wasserentnahmeerlaubnisse. Endbericht. Studie im Auftrag des Fachverbandes Feldberegnung 2007-2009. Georg-August-Universität Göttingen, Department für Agrarökonomie und Rurale Entwicklung. http://www.lwk-niedersachsen.de/index.cfm/portal/6/nav/203/article/14250.html. Accessed 29 March 2016

Brookfield AE, Gnau C (2016) Optimizing Water Management for Irrigation Under Climate Uncertainty: Evaluating Operational and Structural Alternatives in the Lower Republican River Basin, Kansas, USA. Water Resources Management 30(2):607–622. doi: 10.1007/s11269-015-1180-y

Brouwer C, Prins K, Heinloem M (1989) Irrigation Water Management: Irrigation Scheduling. Annex I: Irrigation efficiencies. Training manual No. 4. FAO. Rome, Italy

Cetin O, Bilgel L (2001) Effects of different irrigation methods on shedding and yield of cotton. Agricultural Water Management 54:1–15. doi: 10.1016/S0378-3774(01)00138-X

Fernández FJ, Ponce RD, Blanco M, Rivera D, Vásquez F (2016) Water Variability and the Economic Impacts on Small-Scale Farmers. A Farm Risk-Based Integrated Modelling Approach. Water Resources Management 30(4): 1357–1373. doi: 10.1007/s11269-016-1227-8

Fricke E, Riedel A (2013) Feldtag Beregnung. Edited by Landwirtschaftskammer Niedersachsen und Fachverband Feldberegnung e.V. Hamerstorf. Available online at http://fachverband-feldberegnung.de/pdf/FeldfuehrerBeregnungsversuche2013.pdf. Accessed 29 March 2016

Gassman PW, Reyes MR, Green CH, Arnold JG (2007) The Soil and Water Assessment Tool: Historical Development, Applications and Future Research Directions. Transactions of the ASABE 50(4): 1211–1250

Gayley A (2013): SWAT-Based Evapotranspirative Water Conservation Analysis Performed on Irrigated Croplands to Determine Potential Regional Water Savings. J. Irrig. Drain Eng. 139(6):456–462. doi:10.1061/(ASCE)IR.1943-4774.0000562

Hansen JH, Challinor A, Ines A, Wheeler T, Moron V (2006) Translating climate forecasts into agricultural terms: advances and challenges. Climate Research 33:27–41. doi: 10.3354/cr033027

Irmak S, Haman DZ, Bastug R (2000) Determination of Crop Water Stress Index for Irrigation Timing and Yield Estimation of Corn. Agron J. 92:1221–1227. Doi: 10.2134/agronj2000.9261221x

Kreins P, Henseler M, Anter J, Herrmann F, Wendland F (2015) Quantification of Climate Change Impact on Regional Agricultural Irrigation and Groundwater Demand. Water Resources Management 29(10): 3585–3600. doi:10.1007/s11269-015-1017-8

Kuhn HW (1955) The Hungarian method for the assignment problem. Naval Research Logistics 2(1-2): 83–97. doi: 10.1002/nav.3800020109

Kutschera L, Lichtenegger E, Sobotik M (2009) Wurzelatlas der Kulturpflanzen gemäßigter Gebiete mit Arten des Feldgemüsebaus, 7. Band der Wurzelatlas Reihe. DLG-Verlag

Landesamt für Statistik (2010) Statistische Berichte Niedersachsen, Landwirtschaftszählung 2010, Heft 03, Bodennutzung, Rechtsform der Betriebe, Ökologischer Landbau, Zwischenfruchtbau, Bewässerung

Liu L, Luo Y, He C, Lai J, Li X (2010) Roles of the combined irrigation, drainage, and storage of the canal network in improving water reuse in the irrigation districts along the lower Yellow River, China. Journal of Hydrology 391(1-2):157–174. doi: 10.1016/j.jhydrol.2010.07.015

Löpmeier, F.-J. (1994) Berechnung der Bodenfeuchte und Verdunstung mittels agrarmeteorologischer Modelle. Zeitschrift f. Bewässerungswirtschaft, 29: 157–167.

Moghimi MM, Sepaskhah AR (2016) Effect of Various On-Farm Water Management Scenarios on Equity and Productivity in Irrigation Networks. Water Resources Management. doi: 10.1007/s11269-016-1295-9

Mostafa H, Derbala A (2013) Performance of supplementary irrigation systems for corn silage in the sub-humid areas. Agric Eng Int: CIGR Journal 15(4):9–15

Neitsch SL, Arnold JG, Kiniry JR, Williams JR (2011) Soil and Water Assessment Tool - Theoretical Documentation Version 2009

Petry U, Dietrich J, Förster K, Wallner M, Berndt C, Meon G, Haberlandt U (2015) Ein Ansatz zur Validierung von Klimamodelldaten als Basis für die Interpretation von wasserwirtschaftlichen Klimafolgenabschätzungen in Niedersachsen. Hydrologie und Wasserbewirtschaftung 59(4):155-173. doi: 10.5675/HyWa_2015,4_3

Riediger J, Breckling B, Nuske RS, Schröder W (2014) Will climate change increase irrigation requirements in agriculture of Central Europe? A simulation study for Northern Germany. Environ Sci Eur 26(1):18. doi: 10.1186/s12302-014-0018-1

Santhi C, Muttiah RS, Arnold JG, Srinivasan R (2005) A GIS-based regional planning tool for irrigation demand assessment and savings using SWAT. Transactions of the ASAE 48(1):137–147. doi: 10.13031/2013.17957

Statistisches Bundesamt (2004): Statistik der Wasserversorgung in der Landwirtschaft. Wiesbaden

Smith M (1992) CROPWAT: A Computer Program for Irrigation Planning and Management. Food and Agriculture Organization of the United Nations (FAO irrigation and drainage paper)

Tetzlaff B, Kuhr P, Wendland F (2009) A new method for creating maps of artificially drained areas in large river basins based on aerial photographs and geodata. Irrig. and Drain. 58(5):569–585. doi:10.1002/ird.426

Thompson RB, Gallardo M, Valdez LC, Fernández MD (2007) Using plant water status to define threshold values for irrigation management of vegetable crops using soil moisture sensors. Agricultural Water Management 88(1-3):147–158. doi:10.1016/j.agwat.2006.10.007

van Dam JC, Groenendijk P, Hendriks RFA, Kroes JG (2008) Advances of modeling water flow in variably saturated soils with SWAP. Vadose Zone J. 7(2)640–653. doi: 10.2136/vzj2007.0060

Wriedt G, van der Velde M, Aloe A, Bouraoui F (2009) A European irrigation map for spatially distributed agricultural modelling. Agricultural Water Management 96(5):771–789. doi:10.1016/j.agwat.2008.10.012

Yazar A, Sezen SM, Sesveren S (2002) LEPA and trickle irrigation of cotton in the Southeast Anatolia Project (GAP) area in Turkey. Agricultural Water Management 54:189–203. doi:10.1016/S0378-3774(01)00179-2

Zhang X, Pei D, Li Z, Li J, Wang Y (2002) Management of supplemental irrigation of winter wheat for maximum profit. Deficit irrigation Practices. Water Reports 22:57-65. Food and Agriculture Organization of the United Nations, Rome

This chapter is an edited version of the following original scientific article:

Maier, N.; Dietrich, J. (2016): Using SWAT for Strategic Planning of Basin Scale Irrigation Control Policies: a Case Study from a Humid Region in Northern Germany. Water Resources Management 30(9), 3285-3298.

Publisher: Springer Nature

VII. Evaluation of SWAT simulated soil moisture at catchment scale by field measurements and Landsat derived indices

Bhumika Uniyal[1], Jörg Dietrich[1], Christos Vasilakos[2] and Ourania Tzoraki[3]

[1] Institute of Hydrology and Water Resources Management, Leibniz University Hannover, Germany
[2] Department of Geography, University of the Aegean, Mytilene, Greece
[3] Marine Sciences Department, School of Environment, University of the Aegean, Mytilene, Greece

Abstract

The quantification of soil moisture under different soils and crops at regional scale is a challenging task. Such studies are limited by the availability of ground based measurements. The current study evaluates the spatial and temporal patterns of daily soil moisture simulated by the Soil and Water Assessment Tool (SWAT) for the upper 30 cm of the soil profile with indirect soil moisture estimates from Landsat for 2016. The Thermal Vegetation Difference Index (TVDI),was calculated based on the Normalized Difference Vegetation Index (NDVI) and the brightness temperature (BT) using Landsat images, from which regression models were trained by using field measurements from Time Domain Reflectometer (TDR) to calculate soil moisture. Two agricultural catchments namely, Gerdau and Wipperau in Germany were satisfactorily calibrated using SWAT for daily streamflow (1975-2000) with NSE (Nash-Sutcliffe-Efficiency) > 0.55 and PBIAS (Percent bias) < 5.5%. The parameter uncertainty assessment during the irrigation season (Mar-Sept, 2016) for soil moisture revealed that the uncertainty band is narrow (*p-factor*=0.57-0.83; *r-factor*=0.52-1.3). Spatial and temporal patterns of soil moisture from Landsat and SWAT were evaluated by using boxplots and absolute soil moisture error maps. Results revealed that the overall spatial and temporal patterns of boxplots matched better for the dry period (correlation, $r \geq 0.90$) compared to the wet period ($r \geq 0.57$). The mean absolute error between soil moisture from Landsat and SWAT ranged between 0.9-10% for most soils. In addition to it, the soil map was refined to match soil moisture patterns shown in Landsat images for one sandy soil, which further improved the mean absolute difference (1.06-6%). The current study provides an approach to use remotely sensed soil moisture for verifying hydrological modeling results and for optimizing the parameterization of soils, which may bridge the gap between global, regional and field studies in agricultural water management.

1. Introduction

Soil water availability plays a vital role in crop productivity. Agricultural and hydrological models often use the depletion of soil water as the trigger for irrigation operations. Thus, soil moisture is a key hydrological state variable, which is of great interest amongst agriculturists, meteorologist and hydrologists (Brocca et al., 2010; Schmugge et al., 1980; Walker et al., 2001; Zucco et al., 2014). The amount and vertical distribution of soil water depends non-linearly on soil physical properties (Timm et al., 2006), topography (Western et al., 1999), type and stage of crop (Chen et al., 2007), and previous and current weather conditions (Seneviratne et al., 2010). The soil moisture present in the upper soil layer shows maximum spatio-temporal variation compared to the lower soil layers (Chen et al., 2010; Li et al., 2016). It also plays a major role in checking the

water use intensity in agricultural catchments. Therefore, a good insight of soil moisture variability will bring researchers a step closer in understanding the catchment's hydrology, crop processes, irrigation control and the management of green water in a better way.

Soil moisture estimation can be performed by direct and indirect methods. One of the direct method is the gravimetric technique (oven-drying technique), which is widely used because of its simplicity, reliability and accuracy (Schmugge et al., 1980). Most of the direct methods are labor intensive, time consuming and prohibitive on monetary basis for continuous application in large catchments. On the other hand, indirect methods are simple and can be implemented for continuous applications. These include neutron scattering (Gardner & Kirkham, 1952), gamma ray attenuation (Gurr, 1962), electromagnetic techniques (Topp et al., 1980), tensiometric techniques (Richards and Gardner, 1936), hygrometric techniques, soil water and hydrological models (Tavakoli and Smedt, 2013) and remote sensing techniques (including ground based drones, Vivoni et al., 2014).

Satellite-based estimations of surface soil moisture have received a considerable attention in hydrology and water resources management because antecedent soil moisture is a critical variable in rainfall-runoff modeling (Han et al., 2012) and model based irrigation studies (Phogat et al., 2012). The European Space Agency's (ESA) Soil Moisture and Ocean Salinity (SMOS) mission (2009); Sentinel-2A (2015) (http://apps.sentinel-hub.com/sentinel-playground) and National Aeronautics and Space Administration's (NASA) latest Soil Moisture Active/Passive (SMAP) mission (2014) are designed to better monitor the soil moisture on a global extent with increasing spatial resolution (Entekhabi et al., 2010; Gruhier et al., 2010; O'Neill et al., 2010). The technique of indirect soil moisture estimation in terms of soil moisture index from remote sensing data on a catchment scale dates back to the 1980's (Schultz et al., 1988). The evaluation of remotely sensed soil moisture using eco-hydrological catchment models is the subject of current research (Narasimhan et al., 2005; Li et al., 2010; Rajib et al., 2016). Many approaches have been applied in order to model the relationship between soil moisture and soil reflectivity mainly based on Normalized Difference Vegetation Index (NDVI) and the surface temperature (Ts) (Zhang & Zhou, 2016). The slope of the Ts/NDVI curve can provide valuable information regarding soil moisture and vegetation conditions (Goetz, 1997). Sandholt et. al. (2002) developed the Thermal Vegetation Dryness Index (TVDI). It is based on an empirical parameterization of the relationship between Ts and NDVI resulting from the triangular or trapezoidal shape of the Ts/NDVI scatter plots (Carlson et al., 1994; Moran et al., 1994; Xin et al., 2006). Such methodologies have been evaluated by many researchers for validating indirect methods of estimating soil moisture in large catchments (Carlson et al., 1995; Schultz, 1988; Muller & Decamps, 2001).

The Soil Water Assessment Tool (SWAT) has been used by many researchers for evaluating soil moisture on catchment scale. Muttiah and Wurbs (2002) investigated the change in the water balance components, specifically evapotranspiration, soil water storage and water yield corresponding to two soil maps with different spatial resolution. It has been revealed from the results that the aforementioned change in water balance components are more sensitive to watersheds under wet climate and heterogeneous soils. Deliberty and Legates (2003) and Mapfumo et al. (2004) studied the spatial and temporal variability of soil moisture in US and Canada. No seasonal variation in the temporal autocorrelation was found in the first study whereas, the simulated soil moisture in the second study was under and over predicted by the model in dry and wet conditions, respectively. Narasimhan et al. (2005) used SWAT to produce a

long-term soil moisture dataset for drought monitoring and crop yield prediction in the United States. The results indicated that NDVI can be considered as a good indicator for evaluating crop stress and for determining the onset of agricultural drought in semi-arid areas. Milzow et al. (2011) combined three datasets namely, surface soil moisture from radar, radar altimetry by Envisat and the temporal change in total water storage recorded by GRACE for calibrating a data scarce catchment (Okavango river) using rainfall input from three different sources (ECMWF ERA-Interim, TRMM 3B42, FEWS-Net RFE). The results revealed that the error in simulated model outflow can be reduced by reducing error in the precipitation. Park et al. (2014) evaluated the soil moisture simulated by SWAT using MODIS NDVI and land surface temperature (LST) for a forest in Spain. The results revealed that, except in frequent rainy years SWAT simulated soil moisture showed a higher correlation with MODIS LST and NDVI during forest leaf growing and falling period respectively. Li et al. (2016) assessed the spatio-temporal variation in soil moisture and other water balance variables under different precipitation gradients in the Yellow river basin, China. Results revealed that soil moisture showed a non-linear relationship to precipitation and evapotranspiration, however all variables exhibited an annual decreasing trend. Rajib et al. (2016) evaluated the spatially distributed surface and root zone soil moisture in two watersheds of Indiana, USA for improving the hydrologic predictability of SWAT. It was indicated from the results that root zone soil moisture may play an important role in model calibration.

The outcome of most of the soil moisture estimation methods is at point or field scale, whereas hydrological simulation and remote sensing techniques can be used to quantify soil moisture not only at point/field, but also catchment or global scale. Soil moisture can also be estimated for previous years with models and remotely sensed data, which is impossible to derive in case of experimental measurements. Therefore, the availability of model simulated results and remotely sensed data can fill temporal and spatial data gaps and improve long term catchment studies of soil moisture and related agricultural water management problems. The spatio-temporal evaluation of soil moisture at catchment scale was done in few studies only. Therefore, an attempt has been made through this study to check the applicability of Landsat data to be used for monitoring the soil moisture at a catchment scale. The basic objective of the current study is to evaluate the spatial and temporal behavior of SWAT simulated soil moisture with direct and indirect estimates of soil moisture obtained from field and remotely sensed data. It also includes examining the effect of parameter uncertainty on the simulated soil moisture under different soil and land-use conditions. The following main innovations were done in this study:

- Evaluation of SWAT simulated soil moisture at HRU scale using high-resolution indirect estimation from Landsat data (30 m) and field measurements.
- Creation of soil moisture parameter uncertainty bands for specific combinations of soil and crops.
- Development of absolute soil moisture error maps using soil moisture from Landsat and SWAT.
- Adjustment of SWAT soil parameters to match soil moisture patterns in SWAT using Landsat image.

2. Materials and methods

2.1. Model description

The Soil and Water Assessment Tool (SWAT) is a semi-distributed catchment scale hydrological and water quality model (Arnold et al., 1998; Arnold et al., 2012). It is often used in simulating hydrology, chemicals, sediments, crop growing, agriculture management, etc. within agricultural watersheds. The water balance equation used by the model is well explained in the SWAT manual (Neitsch et al., 2011; Arnold et al., 1998). The soil water movement in different soil layers present in the root zone follows a cascade model. Based on this, the model provides net water to the first layer, after canopy interception and evaporation loss. Then according to its field capacity and hydraulic conductivity, storage, runoff excess and infiltration takes place. Excess water further seeps into the next layers. In this process, if the last layer is saturated and still there is excess water then the model assigns it to the first layer. The soil moisture in each subsurface layer is updated by using the following equation in the model:

$$sol_{st_i} = sol_{st} - sepday_i - latlyr_i - lyrtile_i \quad (1)$$

where sol_{st_i} (mm) is the amount of water stored in the soil layer on the i^{th} day; $sepday_i$ (mm) is the percolation from the soil layer on the i^{th} day; $latlyr$ is the lateral subsurface flow in the layer on the i^{th} day; and $lyrtile_i$ (mm) is tile drainage in the soil layer on the i^{th} day. In addition to this, the lateral flow in each subsurface soil layer is calculated with Sloan's kinetic storage model (Sloan and Moore, 1984; Sun et al., 2016).

2.2. Study area and data

This study is conducted in two sub-catchments of the Ilmenau River basin in Northern Germany (52-54°N, 10-11°E), namely Gerdau (Hansen gauging station, 308 km^2 catchment area) and Wipperau (Oetzmühle, 201 km^2) (Fig. 1). The average annual precipitation from 1975 to Sept 2016 was 793 mm and 721 mm in Gerdau and Wipperau, respectively. Both catchments are dominated by agricultural land use with more than 50% of its total area under agriculture. Most of the soils in the catchments can be categorized as medium to fine sand with more than 80% of sand material and more than 90% sand in the lower layers, resulting in low water holding capacity and fast drainage (Tables 1 and 2). As a result, irrigation is highly required for carrying out agriculture even under humid climatic conditions (Maier and Dietrich, 2016). Sources for irrigation water are the shallow quaternary aquifer (Wittenberg, 2003), a navigation canal and reuse of processed water from the sugar industry. Crops grown in this area are wheat, potatoes, corn silage and sugar beets (Agrarstrukturerhebung, tables Z6070421, K6070411, K6070412 and K6080014 with data from 1977 to 2007 reported every four years, Niedersächsisches Landesamt für Statistik). These crops are irrigated mostly by sprinklers but few fields in this region are nowadays using pivots.

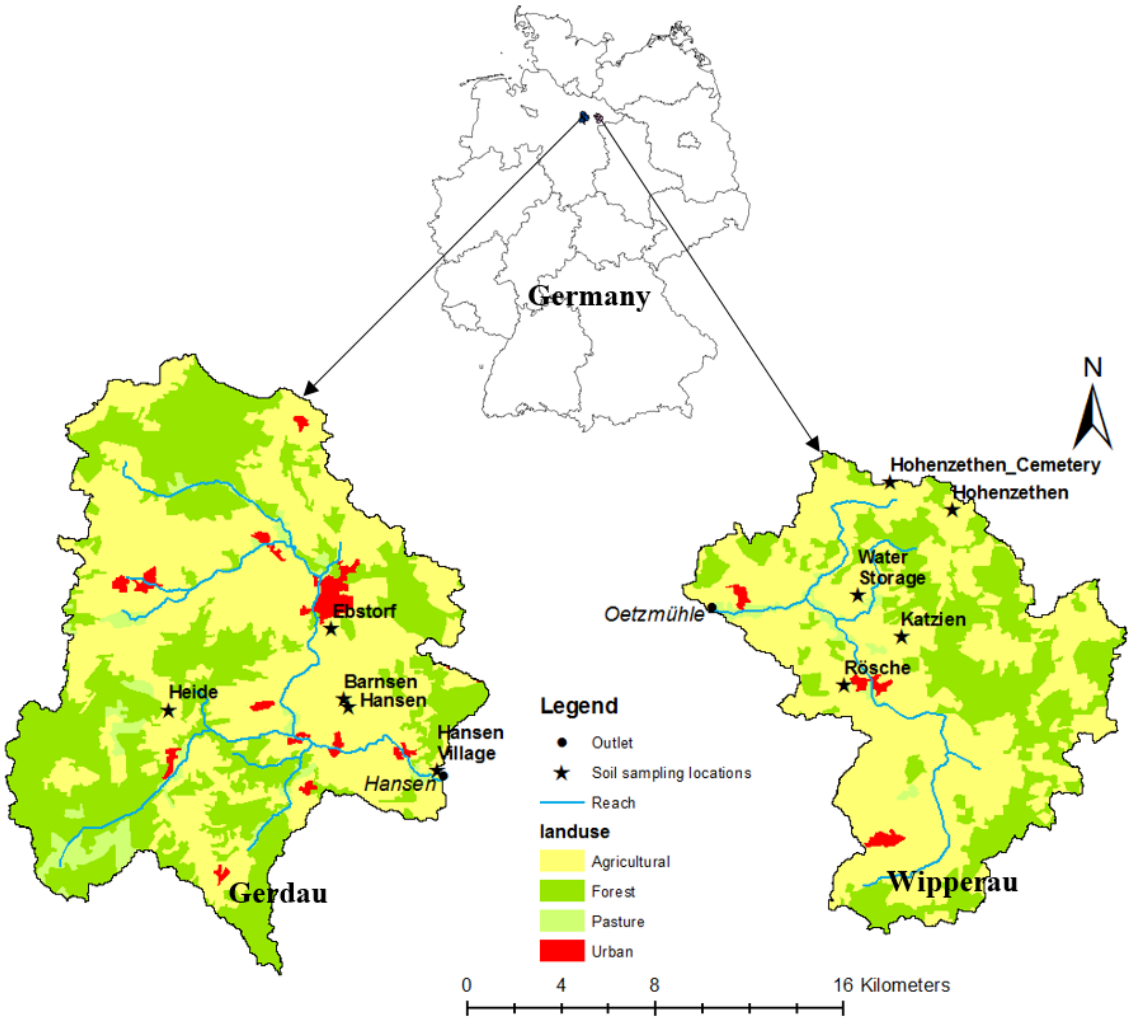


Fig. 1. Location of the catchments along with the soil sampling locations.

The watershed was delineated from a 20 m resolution Digital Elevation Model (DEM), which was aggregated from the original 5 m resolution DEM of NLWKN (Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz). Soils were represented by BÜK200 (Bundesanstalt für Geowissenschaften und Rohstoffe, 2016 https://www.bgr.bund.de/DE/Themen/Boden/Produkte/Karten/Downloads/Flyer BUEK200.pdf? blob=publicationFile&v=5) using soil hydraulic parameters from Wessolek et al. (2009). The Land cover map was generated by combining CORINE (CORINE Land Cover CLC2006, Federal Environment Agency, DLR-DFD 2009) with crop distribution statistics of Lower Saxony for the year 2007 (Niedersächsisches Landesamt für Statistik, 2010). Based on these data, crops reported at community level were distributed in each sub-basin. SWAT hydrological response units (HRUs) were generated by overlay of soil and land use, then the agricultural land use was split by using the crop fractions within each sub-basin. In this process, the net area of crops within a sub-basin is secured but the actual spatial locations of the HRUs are not retained. Apart from the spatial data, temporal weather data for Gerdau and Wipperau catchments were available from 1/1/1975 to 30/09/2016. Daily precipitation data from 6 rain gauge stations in Gerdau and 12 stations in Wipperau were used. Precipitation data was corrected for measurement errors using Richter (1995) depending on temperature. Daily data on temperature, relative humidity and wind velocity were interpolated from four weather stations. These aforementioned weather

parameters were interpolated for all sub-basins using inverse distance interpolation. Solar radiation data was used from the lysimeter station at Hohenzethen (Landesamt für Bergbau,Energie und Geologie, LBEG) for 2001-2013. This time series was extended until Sept, 2016 by using solar radiation values from DWD Braunschweig station. The solar radiation values of Braunschweig were multiplied by a factor representing the bias in the past data between both stations, which differ in location and hence exhibit a systematic difference in radiation. Daily streamflow data were procured from NLWKN for the respective catchment outlet marked in the map (Fig. 1).

2.3. Soil moisture estimation from field experiment

The soil moisture field measurements were coincided with the Landsat satellite crossing date over the region. The field measured soil moisture was used for calibrating the indirect soil moisture estimation from Landsat images. Out of the total field measurement days, only 7 days in case of Gerdau and 6 in case of Wipperau could be used for the irrigation season of 2016 (Mar-Sept). This was due to unclear sky at 10:15 am (UTC) over the catchments for the rest of the days. The clear sky at image recording time was a prerequisite for interpreting Landsat images. Soil moisture was estimated by using Time Domain Reflectometer (TDR) at 10 sites (Fig.1), which were spatially distributed in the two catchments. Gravimetric measurements were conducted on seven days for assessing the quality of soil moisture measured from TDR. TDR provides a representative measure of the volume of soil moisture present in the entire length of the rod (16 cm) placed in the soil profile. The instrument is specially designed for field use and calibrated for universal soils (https://imko.de/en/products /soilmoisture/soil-moisture-sensors/trimepico64). Tables 1 and 2 show the type of crops grown (in the year 2016) and soil physical characteristics of the specific field locations present in Gerdau and Wipperau, respectively. Grain size analysis of all the ten sites was performed according to DIN EN 933-1 for the upper layer of the soil. This has helped to crosscheck the soil database and get a better understanding of the spatial variability of soil physical properties.

Table 1. Specifications of the soil moisture measuring sites in Gerdau

Sl. No.	Name & Crop grown	Location	Soil Name	Soil Physical Properties	
				Texture (%) (*sand/silt/clay*)	Bulk density (g/cm^3)
1.	Hansen (Onion)	10.48° E 52.95° N	128	73/20/7	1.365
2.	Hansen Village* (Winter Wheat)	10.48° E 52.95° N	449	73/20/7	1.365
3.	Barnsen (Corn)	10.42° E 52.98° N	129	73/20/7	1.365
4.	Ebstorf *(Sugar beet)	10.41° E 53.01° N	161	93/5/2	1.045
5.	Heide (Corn)	10.31° E 52.98° N	165	94/5/1	1.41

**soil sampling locations used for further analysis.*

Table 2. Specifications of the soil moisture measuring sites in Wipperau

Sl. No.	Name & Crop grown	Location	Soil Name	Soil Physical Properties	
				Texture (%) (*sand/silt/clay*)	Bulk density (g/cm^3)

1.	Rösche (Herb)	10.74° E 52.98° N	161	93/5/2	1.41
2.	Katzien (Corn)	10.77° E 53.00° N	165	94/5/1	1.045
3.	Water Storage* (Corn)	10.75° E 53.02° N	165	93/5/2	1.41
4.	HZ (Potato) *	10.81° E 53.05° N	129	73/20/7	1.365
5.	HZ_Cemetery (Fodder)	10.77° E 53.06° N	165	93/5/2	1.41

**soil sampling locations used for further analysis.*

2.4. Soil Moisture Estimation from Landsat

The satellite images from Landsat 7 and 8 were acquired from United States Geological Survey (https://espa.cr.usgs.gov/index/). The Landsat surface reflectance high-level data products i.e., *NDVI* based on the surface reflectance and the brightness temperature (BT, at top of atmosphere) were used in order to calculate the *TVDI* (Masek et al, 2006; Vermote et al. 2016). The cloud mask product was used for masking anything else except clear land. The scatter plot of T_s and T_s/*NDVI* space usually follows the trapezoidal form shown in Fig. 2. This distribution is defined by the upper edge which is the dry edge and the lower wet edge. The *TVDI* is based on the empirical parametrization of the dry edge (T_{smax}) and the wet edge (T_{smin}) given by (Sandholt et al., 2002):

$$T_{s\max} = (a_1 * NDVI) + b_1 \quad (2)$$

$$T_{s\min} = (a_2 * NDVI) + b_2 \quad (3)$$

Then, *TVDI* is calculated by using the formula:

$$TDVI = \frac{T_{s\max} - T_s}{T_{s\min} - T_{s\min}} \quad (4)$$

Where, T_{smax} *and* T_{smin} are the dry and the wet edge, respectively; *NDVI* is the observed normalized difference vegetation index and T_s is the observed surface temperature. For small intervals of *NDVI*, the maximum and minimum temperatures are observed in order to estimate a_1, b_1, a_2 and b_2 coefficients of T_{smax} and T_{smin} through linear regression. Most approaches are based on the estimation of T_{smax} and T_{smin} for each image that is being studied. However, this approach may result in incomparable results between different dates since the *TVDI* relies on the maximum and minimum temperatures for each image per day. Furthermore, the cloud fraction influences the pixels that are masked i.e., pixels that are not plotted. In order to cope with this drawback in the current research, dry and wet edges were estimated using universal scatter plot where the T_s/*NDVI* values for all images were plotted at the same time. Hence, the *TVDI* for all images are based on common dry and wet edges. Next step was the calculation of correlation between the soil moisture values from field experiments and the *TVDI*, *NDVI* and brightness temperature (*BT*) values of the corresponding pixels. In order to explore this correlation, linear regression and curve estimation methods were applied to get the best fit.

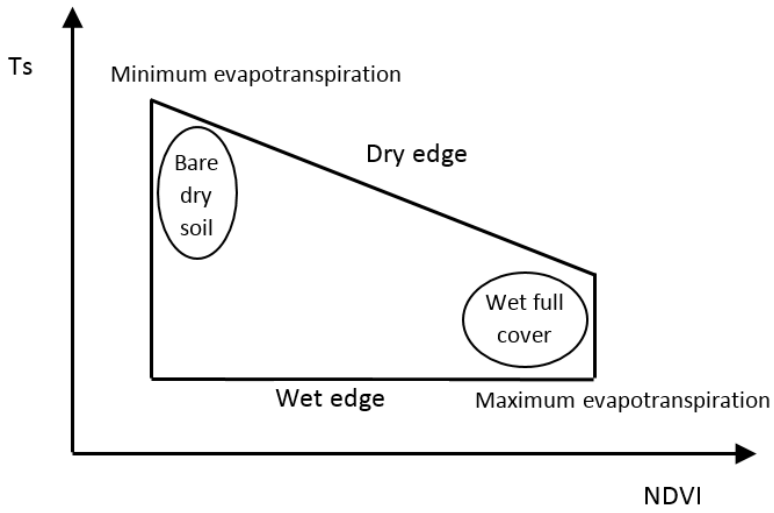


Fig. 2. Scatter plot of T_s/NDVI space.

2.5. Model set up, Calibration and Soil Moisture Extraction

For this study, Gerdau was divided into 18 sub-basins and 1382 HRUs, whereas, Wipperau was divided into 8 sub-basins and 567 HRUs. This is considered as a satisfactory representation of the basin heterogeneity. Surface runoff was estimated by using the SCS Curve Number method, flow in the catchments was routed by using Muskingum's routing equation. The Penman-Monteith equation was used for estimating the evapotranspiration of the study area. Auto-irrigation was activated in the catchments based on plant water demand. Crops were scheduled according to the planting and harvesting dates provided by the agricultural chamber of Lower Saxony. The eco-hydrological model SWAT was calibrated for streamflow from 1975 to 2000 at the respective outlets of the Gerdau and Wipperau catchments with a five year warm up period. Semi-automatic calibration was performed for Gerdau and Wipperau using both manual and automatic calibration by SWAT-CUP (Abbaspour et al., 2012). A base flow filter program was also used to have an initial estimation of the groundwater flow parameters as well as the percentage of base flow occurring in the catchments (Arnold et al., 1995). The ranges of the sensitive parameters were chosen from previous studies performed in German catchments (Lam et al., 2011) as well as from the SWAT database (Neitsch et al., 2011). The sensitive parameters used for calibration are shown in Table 3. Model performance was evaluated by matching observed and simulated hydrographs visually as well as by using statistical indicators. In addition to this, the parameter uncertainty evaluation was also provided by SWAT-CUP in terms of 95% parameter uncertainty (95PPU) band, *r-factor* and *p-factor* for evaluating the model performance. The *p-factor* denotes the percentage of measured data bracketed by the 95 % prediction uncertainty (95PPU), whereas the *r-factor* denotes the average thickness of the 95PPU band divided by the standard deviation of the measured data (Abbaspour, 2011). After attaining an acceptable model performance during the calibration period (1980-2000), model validation was performed (2001-2014). Soil moisture at HRU level was extracted from the final model output. In the current study, the first layer of the soil was evaluated, which corresponds to the top 30 cm of the soil profile. The model simulated soil water content (in mm) was first converted into relative soil moisture (percentage) and then into total moisture present in the respective soil layer. Moisture content at wilting point (WP) corresponding to different soil texture and bulk density was extracted from Wessolek et al.,

(2009). The aforementioned database was used for converting the SWAT output i.e., plant available water content (SW) to total moisture present (SW + WP) in the selected soil layer.

2.6. Soil moisture data analysis

The spatial and temporal analysis of the soil moisture derived from three different sources i.e., Landsat, SWAT and TDR were conducted. There is a scale gap between soil moisture extracted from remote sensing (Landsat resolution 30 m) and soil moisture data from SWAT (Average HRU area: 22.32 ha for Gerdau and 35.35 ha for Wipperau) due to their different resolutions. Landsat provides precise spatial estimates of soil moisture under varying soil and crop presence in the catchment on a given day. However, in SWAT crops were not explicitly incorporated in terms of spatial extent, they were distributed only on the basis of actual land use statistics. Therefore, boxplots were drawn to represent the overall distribution of soil moisture when a clear image from the Landsat was available. Apart from this, the seasonal dynamics of soil moisture during the entire irrigation season extracted from different sources (TDR, SWAT and Landsat) was also evaluated separately using the boxplots for Gerdau and Wipperau. In addition to the graphical measures, one-on-one comparison of the soil moisture extracted from different sources has been performed for the entire irrigation period of 2016. In addition to the overall soil moisture distribution using boxplots, soil and land use (agricultural land use) specific soil moisture maps were also developed. The absolute soil moisture error maps for four days in the entire irrigation period (7 March, 12 May, 24 Aug and 9 Sept) were created for agricultural soils in the two catchments. The aforementioned soil moisture difference maps were developed by creating the soil moisture raster maps from SWAT and Landsat, respectively. Furthermore, the soil specific maps were extracted from SWAT and Landsat soil moisture raster maps corresponding to the agricultural land use and analysis was performed using these maps. For space constraints, soil moisture difference maps and spatial statistics for the Wipperau catchment are shown in the results section.

2.7. Quantification of SWAT soil parameter uncertainty

Parameter uncertainty is one of the key uncertainties present in modeling studies. There is a huge concern in hydrological modeling about equifinality, which means that different combinations of parameters can lead to the same model result (Beven and Binley, 1992). Therefore, to keep a check on the possible non-uniqueness of model results corresponding to the model parameters, parameter uncertainty is usually quantified in hydrological modelling for stream flow. In this study, a parameter uncertainty band for simulated soil moisture was created for specific combinations of crops and soils on HRU basis to represent the soil moisture dynamics in the catchment. The selections of the HRUs were based on two soils (sandy and sandy loam) covering the major portion of the two catchments along with the major grown crops in the area, corn, potato, sugar beet and winter wheat. First, SWAT was run with different combinations of parameters by using the SWAT-CUP software. Then, the parameter sets from SWAT-CUP yielding in good model efficiency (Nash-Sutcliffe efficiency higher than 0.5) were selected. SWAT was run for these parameter sets and the corresponding simulated soil moisture data for each parameter set were extracted. After that the soil moisture parameter uncertainty band for different combinations of soil type and crop was created from the maximum and minimum values of soil moisture for each time step. The overall spatio-temporal variation (1980-2014) in soil moisture was evaluated by using boxplots for Gerdau and Wipperau, respectively. In addition to this, soil

moisture parameter uncertainty bands during the irrigation period 2016 were shown for the four experimental locations along with the observed soil moisture values in Wipperau and Gerdau, respectively (*marked **in Table 1 and 2). Both statistical as well as graphical indicators were used to evaluate the parameter uncertainty band for the chosen locations in the two catchments.

3. Results and Discussion

3.1. Hydrological model performance

Model calibration is performed using both surface and groundwater parameters for calibrating the developed models for streamflow. Table 3 shows the sensitivity ranking of the different parameters and their final values for the two catchments. As shown in this table, the sensitivity ranking of a particular parameter is different in different catchments. Groundwater recharge to the deep aquifer was found to be most sensitive in both catchments, which is in accordance with the study performed by Wittenberg (2015). The relatively high value of recharge to the deep aquifer (RCHRG_DP) in Wipperau (0.7) can be justified by high infiltration rates, low density of surface water bodies and ground water streams to the north (Elbe basin) makes it a losing catchment. The graphical results are shown for calibration (Figs. 3-6), whereas the statistical model evaluation is provided for both calibration and validation periods for the two catchments (Table 4). Apart from the streamflow hydrographs (Figs. 3&5), flow duration curves (Figs. 4&6) of observed and simulated stream flow are drawn on a logarithmic axis for the better representation of low flow. The low and mean flows are predicted relatively well in both the cases (Figs. 4&6), whereas few peaks are underestimated. Both models are under predict the very low flow (<0.05 m^3/s in case of Wipperau & <1 m^3/s in case of Gerdau) (Figs. 4&6) [NSE (Gerdau) = 0.57 & NSE (Wipperau) = 0.67; PBIAS (Gerdau & Wipperau) = 5.2 %]. This might be due to the high groundwater loss and disconnection of the deep aquifer in SWAT. However, according to the statistical model evaluation (Table 4), the performance of both of the models ranged between good to satisfactory for different statistical indicators (Moriasi et al. 2007). It can be seen from the aforementioned statistics that there is more uncertainty in case of Gerdau (*r-factor* = 1.41) as compared to that of Wipperau (*r-factor* = 0.87), as the width of 95PPU band (*r-factor*) is more. This is justifies the higher p-factor in case of Gerdau (p-factor = 0.76) as compared to the Wipperau catchment (p-factor = 0.72) during calibration period.

Table 3. Sensitive parameters along with the final uncertainty range

Sl. No.	Parameters	Gerdau		Wipperau	
		Sensi-tivity Ranking	Final Parameter range & Best calibrated parameter value	Sensi-tivity Ranking	Final Parameter range & Best calibrated parameter value
1	r__CN2.mgt (SCS curve number)	*8*	(–0.20, –0.15), -0.17	*3*	(–0.20, –0.15), -0.17
2	v__ALPHA_BF.gw (Base flow recession constant, days)	*4*	(0.0, 0.42), 0.37	*7*	(0.43, 0.84), 0.8
3	v__GW_DELAY.gw (Groundwater delay, days)	*5*	(372.64, 704.98), 586.33	*4*	(117.16, 213.97), 185.17
4	v__RCHRG_DP.gw (Recharge to deep aquifer)	*1*	(0.12, 0.5), 0.4	*1*	(0.60, 0.81), 0.67

5	r__SOL_AWC().sol (Avaiable soil water capacity, mmH_20/mm of soil)	*3*	(–0.18, –0.02), –0.17	*2*	(–0.18, –0.02), –0.17
6	r__SOL_K().sol (Soil hydraulic conductivity, mm/h)	*6*	(–0.016, –0.11), –0.05	*6*	(–0.016, –0.11), –0.05
7	v__ESCO.hru (Soil evaporation compensation factor)	*2*	(0.5, 1.0), 0.97	*5*	(0.04, 0.4), 0.14
8	v__GW_REVAP.gw (Groundwater re-evaporation coefficient)	*7*	(0.009, 0.046), 0.024	-	-

**v means replace and r means relative change.*

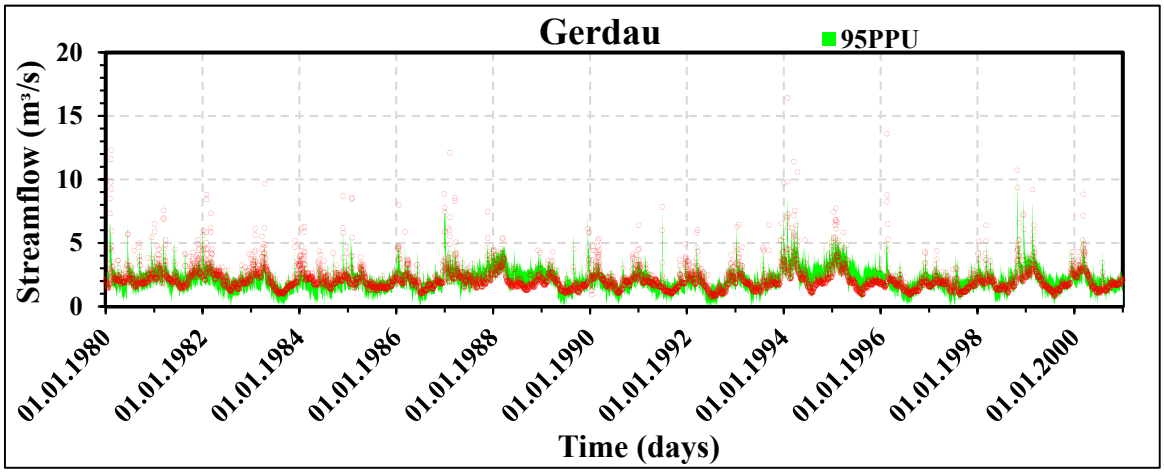


Fig. 3. Comparison of daily streamflow hydrographs for calibration period (1980-2000) for the Gerdau catchment.

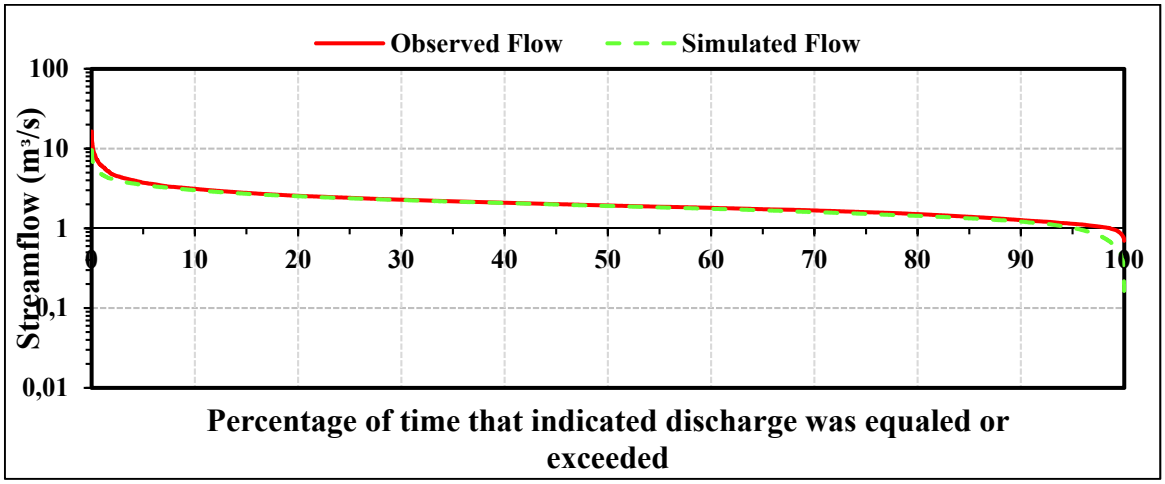


Fig. 4. Flow duration curve for the Gerdau during calibration period.

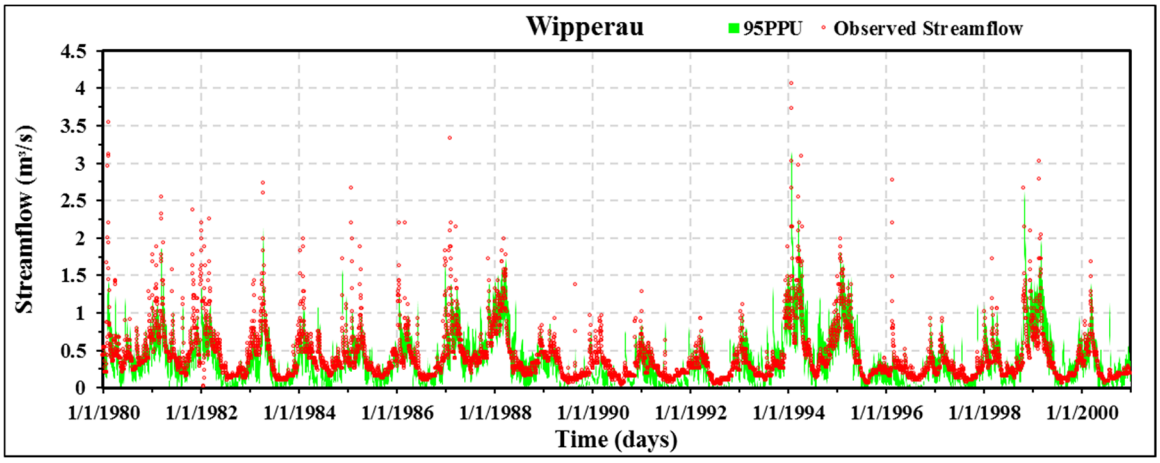


Fig. 5. Comparison of daily streamflow hydrographs during calibration period (1980-2000) for the Wipperau catchment.

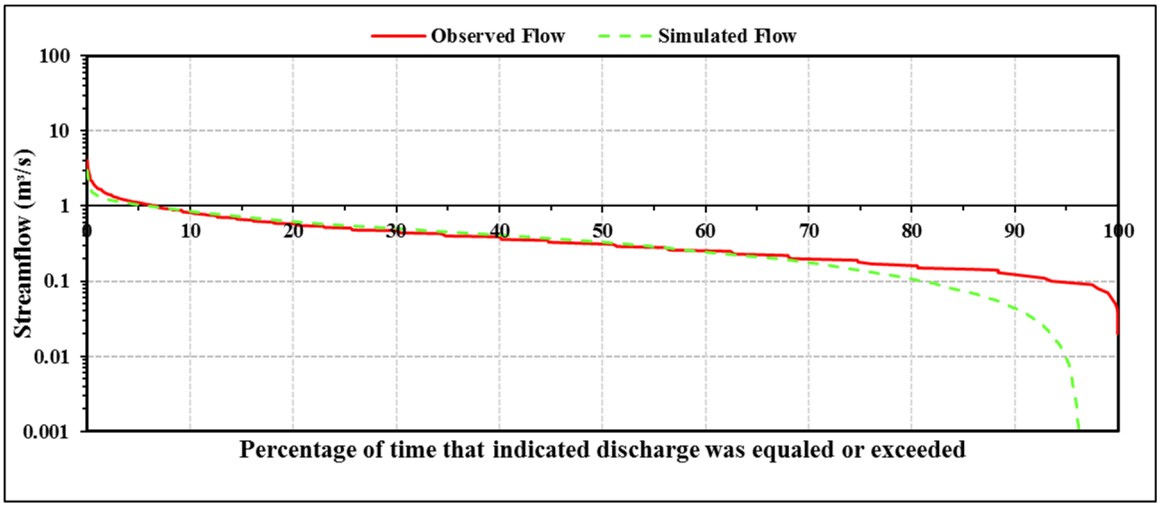


Fig. 6. Flow duration curve for the Wipperau during calibration period.

Table 4. Model Evaluation Statistics.

Statistical Indicator	Gerdau		Wipperau	
	Calibration (1980-2000)	Validation (2001-2014)	Calibration (1980-2000)	Validation (2001-2014)
p-factor	0.76	0.87	0.72	0.61
r-factor	1.41	1.39	0.87	0.65
R^2 (Coefficient of Determination)	0.59	0.52	0.68	0.67
NSE (Nash-Sutcliffe Efficiency)	0.57	0.45	0.67	0.65
PBIAS(%, Percent Bias)	5.2	5.3	5.2	1.4
KGE (Kling-Gupta Efficieny) (Gupta et al., 2009)	0.71	0.71	0.81	0.81

3.2. Uncertainty in SWAT soil moisture simulation

For analizing the parameter uncertainty of SWAT simulated soil moisture, a combination of four relevant crops and two characteristic soil types were investigated. Corn silage (CSIL), potato

(POTA), sugar beet (SGBT) and winter wheat (WWHT) are the main crops growing in both of the catchments. Soil 128 is sandy loam, whereas soil 165 is medium sand (Tables 1 and 2). SGBT and WWHT were analyzed for Gerdau, whereas CSIL and POTA were used in case of Wipperau to show the overall dynamics of soil moisture under two different selected soils. Figs. 7 and 8 show the spatial and temporal distribution of SWAT simulated soil moisture during the model simulation period (1980-2014) for the whole Wipperau and Gerdau, respectively. It can be seen from Fig. 7 that there is a clear difference between the mean soil moisture values in 165 (sandy) and 128 (sandy loam) soils. It can also be concluded from the results that the soil texture plays a more important role in soil moisture storage than the crops (Fig. 7, SGBT & WWHT; Fig. 8 CSIL & POTA). In addition to this, one can also see a clear difference in overall variability of soil moisture in the same soils with different crops. Even though the median is nearly same for the same soil with different crops in all the four cases, the overall spread is different (including upper quantile, lower quantile and outliers). This deviation in the soil moisture band corresponding to different crops growing (SGBT & WWHT) in same soil (Fig. 7) can be explained by the different rate of evapotranspiration of root and grain crops and also different planting and maturing periods. At some time-steps, the simulated soil moisture exceeds field capacity. This can be justified by the oversaturation of soil after intensive rainfall and frost periods during the winter season. When the soil surface temperature increases, the melting of frozen water starts in the upper layers. If the lower soil layer is still frozen, then water accumulates in the top layer resulting in oversaturation of the upper soil layer.

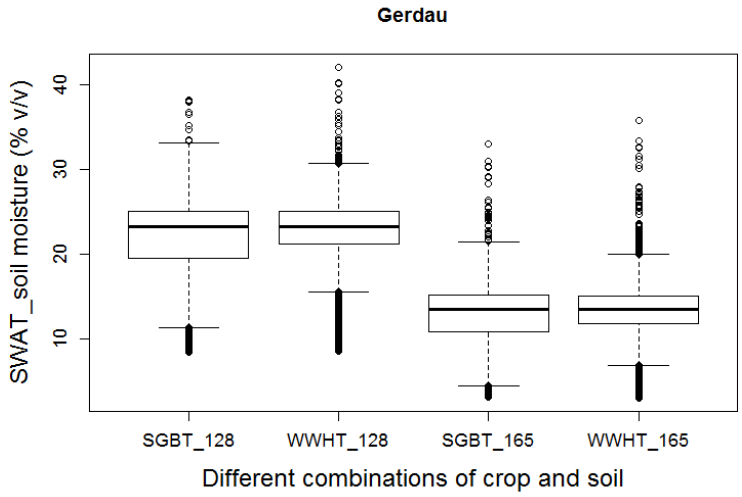


Fig. 7. Overall variation in SWAT simulated soil moisture of the top layer (30 cm) for 128 and 165 soils with sugar beet (SGBT) and winter wheat (WWHT) for the Gerdau catchment.

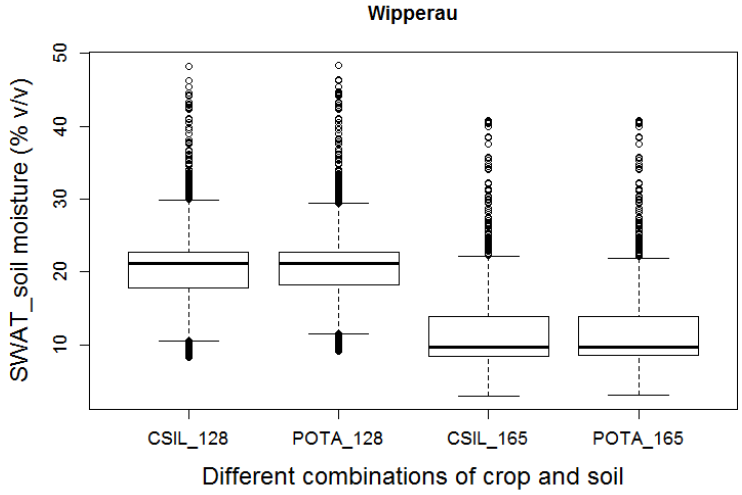
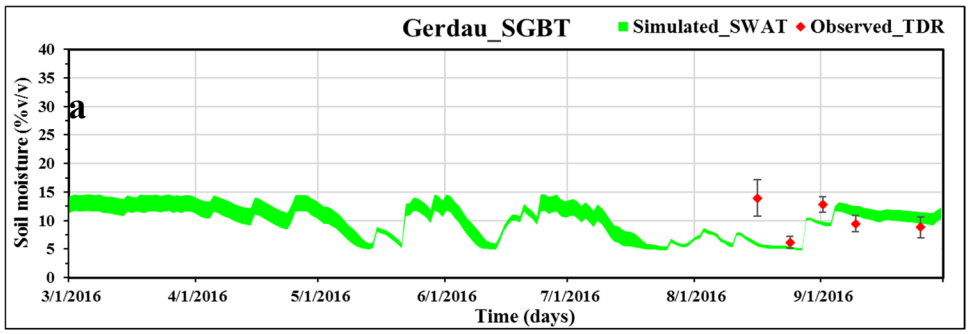


Fig. 8. Overall variation in SWAT simulated soil moisture of the top layer (30 cm) for 128 and 165 soils with corn silage (CSIL) and potato (POTA) for the Wipperau catchment.

Apart from the boxplots, soil moisture uncertainty bands were created specifically for 2016 along with the observed range of soil moisture values from field measurements for Gerdau and Wipperau [Figs. 9(a,b) and 10(a,b)]. In this case, one specific HRU corresponding to the sampling site is considered (Marked * in Table 1 and 2), unlike in the previous case where all the HRUs comprising of a particular crop and soil in the catchment were considered for creating the boxplots. It can be seen from Figs. 9(a,b) and 10(a,b) that most of the observed range of soil moisture measurements are within or close to the soil moisture parameter uncertainty band simulated by SWAT. The range represented as bar depicts the spatial variability of soil moisture at the sampling location obtained from 10 samplings within one field. It can also be seen from the observed TDR soil moisture values that the overall seasonal variability is met relatively well by SWAT except during the harvesting period. This may be explained by the activation of auto-irrigation based on soil water demand. The model provides irrigation water, whenever the soil moisture falls below the threshold value of soil moisture for triggering irrigation, whereas in actual practice farmers stop irrigating their crops a few weeks before the harvest. Another aspect to note would be, the field measurements might not be well representative for the HRU as the area of measurement is relatively low as compared to the area of the HRU. Table 5 shows the parameter uncertainty estimators for the soil moisture uncertainty band at the selected sampling locations. The *p-factor,* which is nearly or more than 0.6 in all the cases, suggests that an acceptable number of observed data lies within the uncertainty band. An *r-factor* (uncertainty width) close to one is usually acceptable which is in accordance with the current case. However, it should be noted that the uncertainty band in the current research only includes parameter uncertainty.



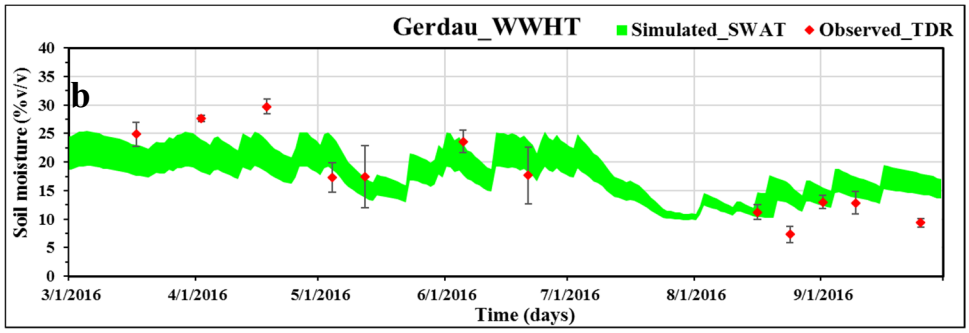


Fig. 10(a,b). Soil moisture parameter uncertainty band for SGBT and WWHT with the respective range of observed soil moisture at different soil sampling locations (*marked) in the Gerdau catchment.

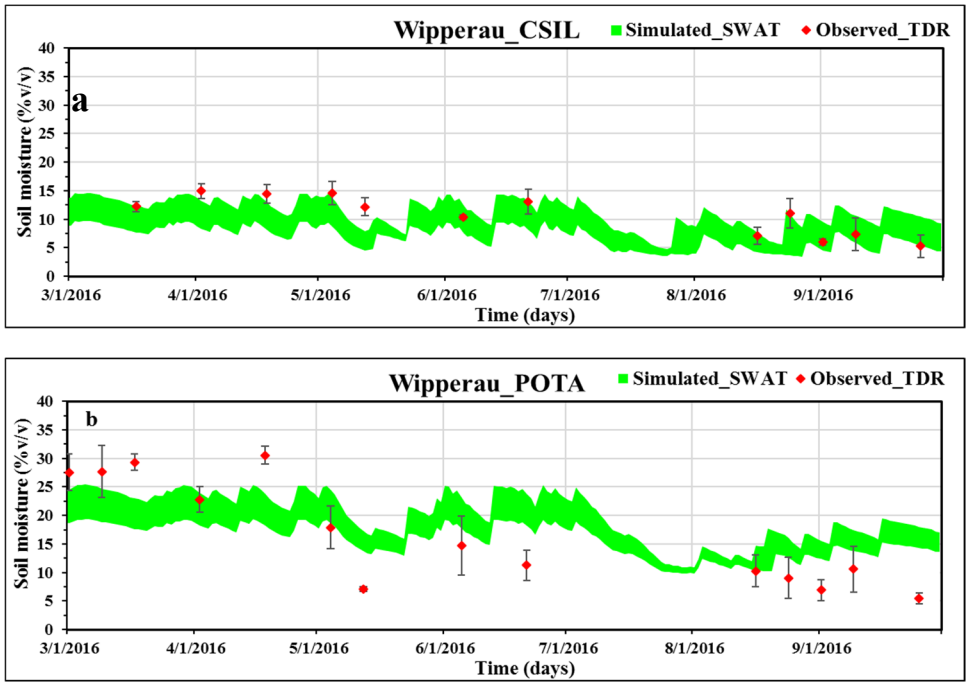


Fig. 11(a,b). Soil moisture parameter uncertainty band for CSIL and POTA with the respective range of observed soil moisture at different soil sampling locations (*marked) in the Wipperau catchment.

Table 5. Parameter Uncertainty Estimators.

Statistical Indicator	Gerdau		Wipperau	
	SGBT	WWHT	CSIL	POTA
p-factor	0.60	0.66	0.83	0.57
r-factor	0.72	0.65	1.31	0.52

3.3. Remote sensing moisture modelling

The scatter plot of the temperature and *NDVI* for all images is presented in Fig. 11. The dry and wet edges are defined by the following equations:

$$T_{s\max} = (-0.0101 * \mathrm{NDVI}) + 3100.4 \quad (5)$$

$$T_{s\min} = (-0.0018 * \mathrm{NDVI}) + 2765.6 \quad (6)$$

Then, the TVDI value for each sampling pixel was calculated based on its corresponding NDVI and BT values retrieved from satellite images. The whole dataset was randomly divided into 70%

training and 30% validation. Six different models were used to train the TVDI and NDVI values for calculating the respective soil moisture using Landsat imagery. The models and the performance results in training (R2 and RMSE) as well as in validation (RMSE) sets are shown in Table 6.

The overall spatial variation of soil moisture on 24 August 2016 is evaluated corresponding to the observed values from TDR. Fig. 12 represents the spatial comparison of soil moisture from 6 different models (M1-M6) for 24 August 2016 explained in Table 6. The aforementioned day is selected because soil is assumed to be dry during this period and also from the review that the overall error in remotely sensed data is relatively less in dry periods as compared to the wet days. It can be easily accessed from the figure that in spite of having comparable model performance in training and testing (Table 6), the overall model spread is quite different in all the models. Therefore, out of all the selected models, M5 is selected for further analysis as its spread and median is closer to the observed soil moisture data.

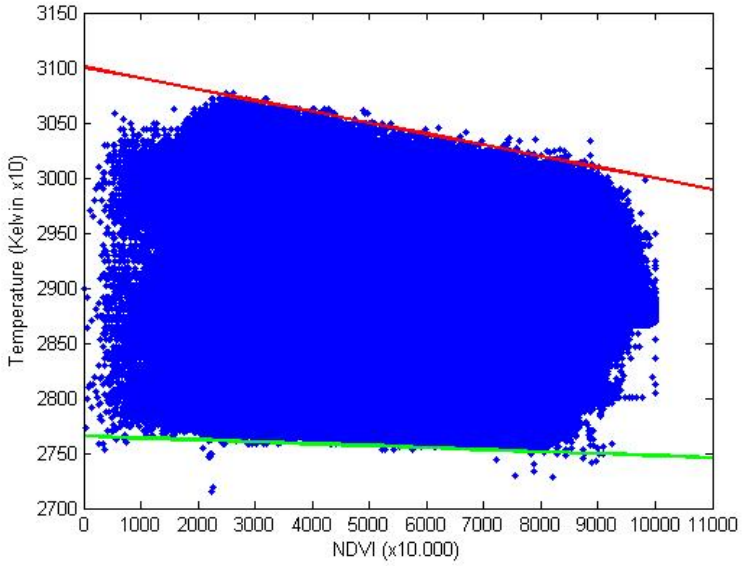


Fig. 11. T_s/NDVI scatter plot of all satellite images and the wet (shown as green line) and dry (shown as red line) edges. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

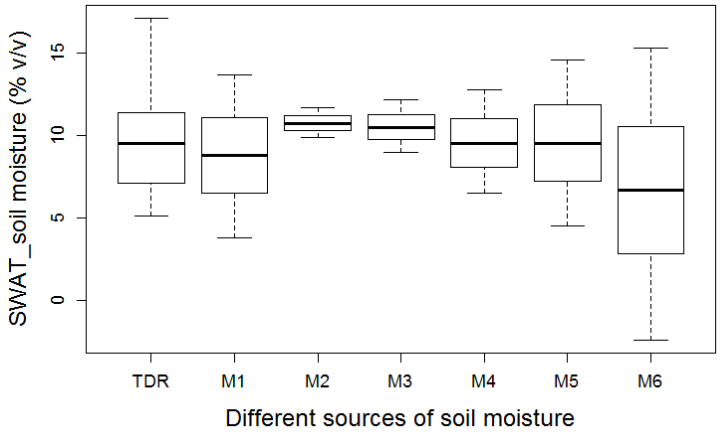


Fig. 12. Comparison of observed soil moisture (TDR) with soil moisture calculated from different regression models using NDVI/Ts/TVDI on August 24th, 2016.

Table 5. TDR modelling results of linear regression and curvature estimation

Model	Goodness of fit		Goodness of validation
	R^2	RMSE	RMSE
(M1) *TDR = 22.68*TVDI+4.203 TDR*	0.64	4.1	4.1
(M2) *TDR = 35.83*TVDI²– 14.03*TVDI + 11.3*	0.69	3.9	3.9
(M3) *TDR = 23.29 * TVDI^2.327 + 9.065*	0.69	3.9	4.0
(M4) *TDR = 6.722 * exp(1.528 * TVDI)*	0.68	3.9	4.1
(M5) *TDR = 225.4 –0.0008178*NDVI–0.07095*T*	0.64	4.3	4.1
(M6) *TDR= 25.46+0.04966*NDVI – 0.005673*T – 3.531e –0.7 NDVI²–1.591e –(0.5*NDVI*T)*	0.67	3.9	4.9

3.4. Soil moisture comparison

In this section, soil moisture values from three different sources (TDR, SWAT, Landsat) are compared for two dates close to the beginning and end of the irrigation season, i.e., 17 March and 24 August, 2016. Practically the irrigation ended in September but there is no clear Landsat image available after 24 August, 2016. It can be seen from the boxplots that the median values of soil moisture from Landsat (RS), SWAT simulated soil moisture for the top 30 cm (SM_30cm) and soil moisture measured by using TDR (TDR) are not in the same range for 17 March, 2016. However, median and spread of soil moisture extracted from all the sources match relatively well for August (Fig.12). In addition to it, the overall boxplots comparison is worse for March in both the models. Although the comparison of boxplots between the two catchments [Gerdau, Fig. 13 & Wipperau Fig. 14] revealed that the soil moisture distribution is better in case of Wipperau as compared to Gerdau. It can also be seen that there are outliers in the soil moisture simulated by SWAT for the upper 30 cm of the soil profile. Few unexplainable values of soil moisture from SWAT (≥ 40 % v/v) may be explained by the complete filling of pore space in the first layer. Table 6 shows the correlation between the soil moisture from TDR, RS and the SM_30cm. Soil moisture correlation varies from 0.68 to 0.88, except at one point in Wipperau where the correlation between TDR and SM_30cm is only 0.54. As mentioned earlier, there is a scale gap between the methods, so this single point evaluation is not expected to give very high correlations, but results in Table 7 show a good overall correlation of the three methods.

Table 7. Soil moisture correlation matrix.

Gerdau

Sources of soil moisture	TDR	RS	SM_30 cm
TDR	1.00		
RS	0.87	1.00	
SM_30 cm	0.78	0.82	1.00

Wipperau

Sources of soil moisture	TDR	RS	SM_30 cm
TDR	1.00		
RS	0.88	1.00	
SM_30 cm	0.54	0.68	1.00

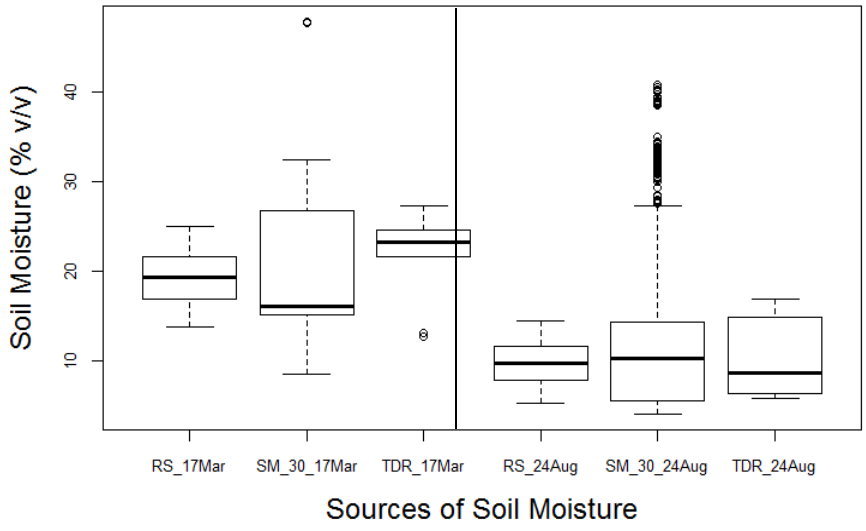


Fig. 13. Comparison of remotely sensed (RS), SWAT simulated (SM) and observed (TDR) soil moisture on March, 17 and August, 24 in the Gerdau catchment.

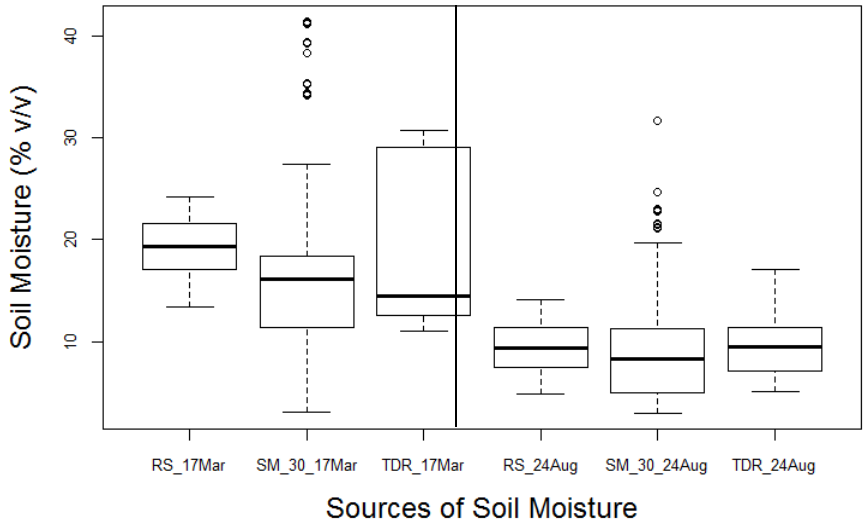
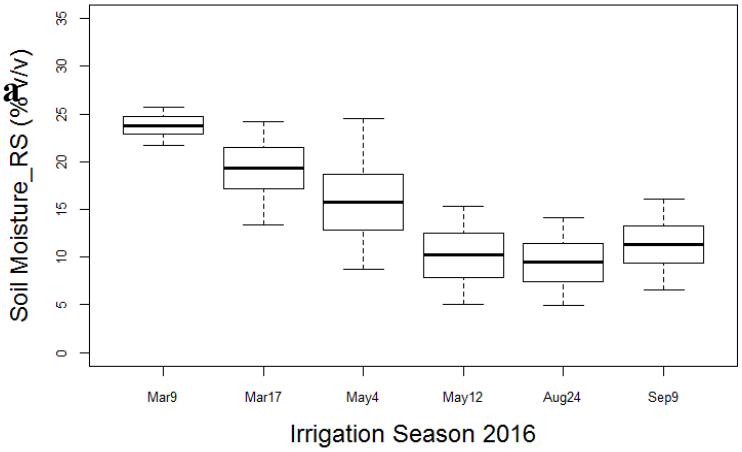
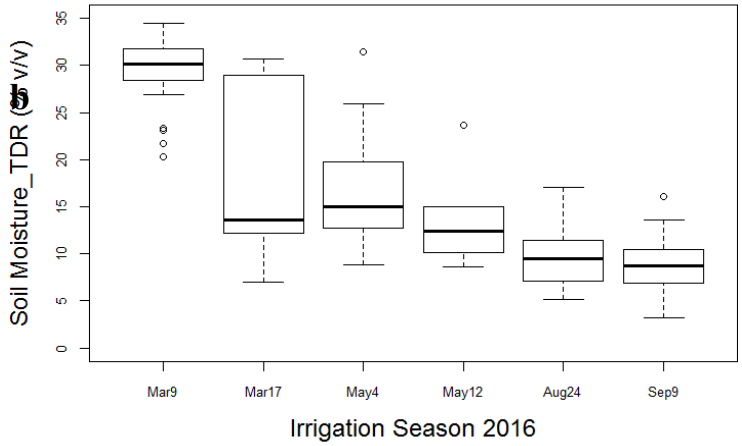


Fig. 14. Comparison of remotely sensed (RS), SWAT simulated (SM) and observed (TDR) soil moisture on March, 17 and August, 24 in the Wipperau catchment.

In addition to this, the temporal dynamics of soil moisture for Gerdau and Wipperau is also evaluated. The boxplots in case of Wipperau [Fig. 15(a-c)] for the six days of the irrigation season 2016, where clear Landsat images are available, show that although median and spread of the data from all the sources do not match well with each other (on a particular day), still the overall temporal dynamics of soil moisture is relatively consistent during the irrigation season.





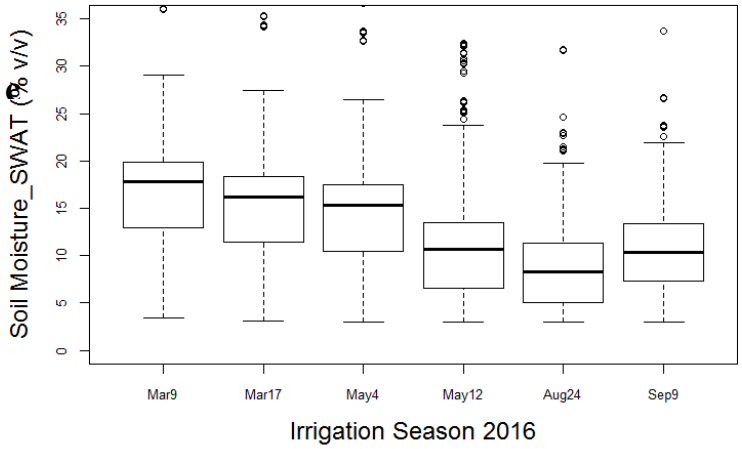


Fig. 15(a-c). Temporal dynamics of soil moisture in the Wipperau catchment for the field sampling dated during the irrigation season 2016.

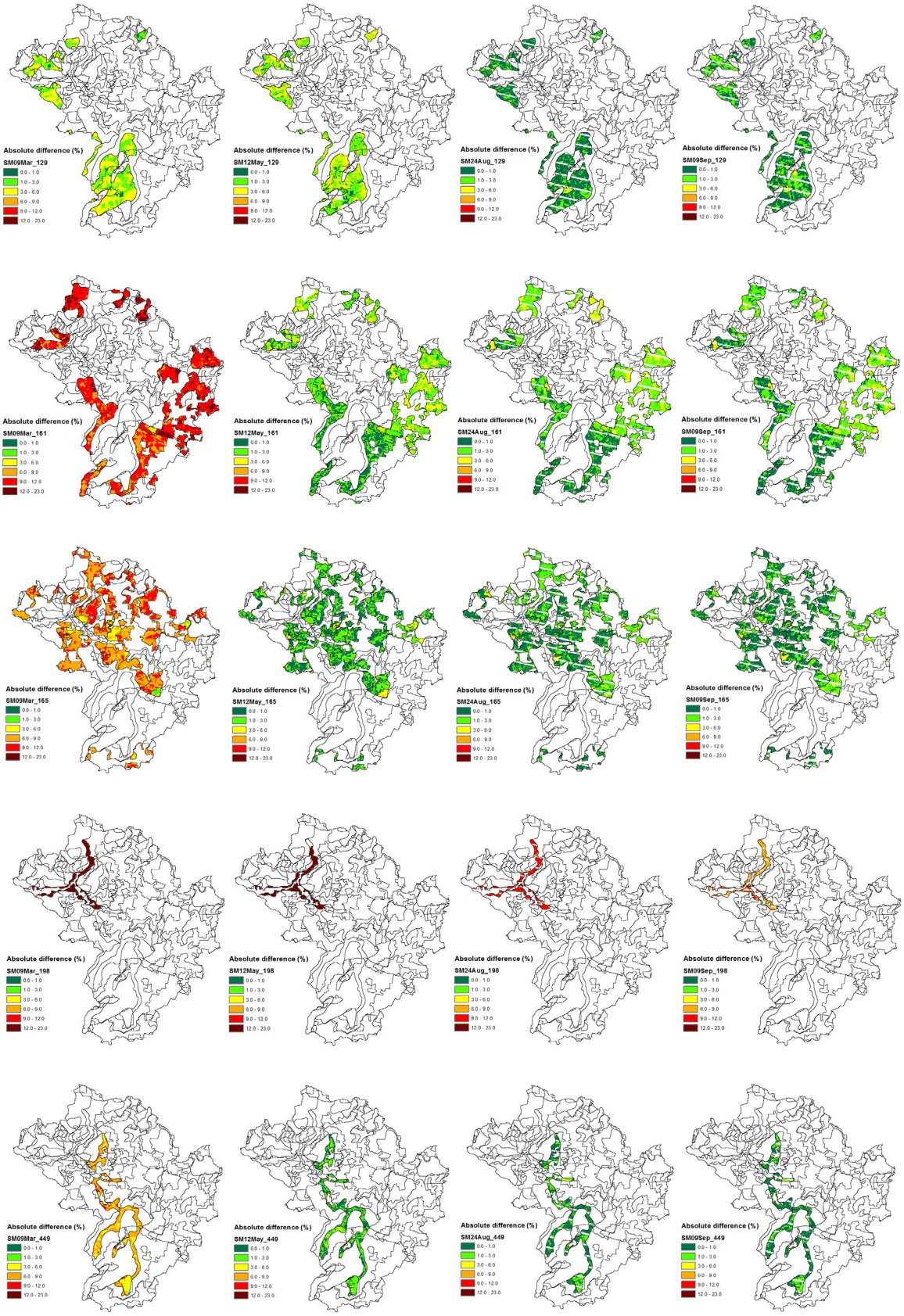


Fig. 16. Soil moisture absolute error maps for Wipperau under major agricultural soils in the area.

Soil moisture difference maps are used to provide a better visual representation of the overall spatial and temporal dynamics of soil moisture for specific combinations of soils under

agricultural land use. Fig. 16 shows the major soils in the Wipperau catchment which are considered for creating the absolute soil moisture difference maps from Landsat and SWAT. It can be assessed that the absolute difference in most of the spatial maps is low (less than 10%, shown in green color). In addition to this, the range of absolute difference in case of soil 198 for all the considered days (9 March, 12 May, 24 Aug, and 9 Sept, 2016) is usually higher in all the cases (6–23%). Landsat fails to represent soil 198 which has higher field capacity equal to 0.40 mm/mm (Wessolek et al. 2009), which was also confirmed by field investigations. However, soil 198 was not included into the calibration of regression models. The absolute soil moisture difference map show very low difference (<3%) in the most of the maps (SM24Aug_129) with relatively high difference in SM09Mar_161 and SM09Mar_165 maps. The soil moisture difference maps are also backed by the average spatial statistics using Table 8. It can be evaluated from the table that the overall mean absolute difference (4 - 15.69 %) and standard deviation (1.74 - 4.06) is high in the 198 soil type. The mean absolute difference of sandy soils (161 and 165) are more than that of sandy loam soil (128). It can also be seen that the mean absolute difference is reducing from March to Sept, which confirms that the Landsat extracted soil moisture can be better used as an indicator for irrigation planning and management during dry periods.

Table 8. Spatial statistics for Wipperau Catchment

Sl.No.	Date	Soil_type	Mean absoulte difference	Standard Deviation
1	9-Mar-2016	165	8.17	1.79
2	9-Mar-2016	449	6.87	1.76
3	9-Mar-2016	198	15.69	2.66
4	9-Mar-2016	129	3.33	1.64
5	9-Mar-2016	161	10.15	1.58
1	12-May-2016	165	1.38	1.47
2	12-May-2016	449	1.45	1.45
3	12-May-2016	198	18.79	4.06
4	12-May-2016	129	2.86	1.59
5	12-May-2016	161	2.08	1.4
1	24-Aug-2016	165	1.28	1.09
2	24-Aug-2016	449	0.91	0.78
3	24-Aug-2016	198	10.04	1.74
4	24-Aug-2016	129	0.809	0.72
5	24-Aug-2016	161	1.97	1.2
1	9-Sep-2016	165	1.17	1.11
2	9-Sep-2016	449	0.89	0.86
3	9-Sep-2016	198	7.45	1.86
4	9-Sep-2016	129	1.1	0.94
5	9-Sep-2016	161	1.77	1.23

3.5. Adjustment of SWAT soil parameters

Based on the absolute soil moisture difference maps shown in Fig. 16, it can be concluded that the soil 161 shows a typical behavior, which can be explained by using image SM24Aug 161. It can be seen from the figure that the absolute difference is less in the lower section (south western) as compared to the upper section (north eastern). The aforementioned pattern is common in all the

cases with soil type 161, which prompted for refinement of the soil parameters. The original database of profiles of the soil map BÜK 200 provides alternative profiles for a single soil type, along with the frequency of occurrence. Backed by field/lab investigations of the grain size distribution on several locations, a sub-type of soil 161 (161 1; Table 9) with new parameters from a less frequent profile is created. After re-simulating the SWAT model, new spatial maps are drawn for all the four days considered for this analysis. Fig. 17 shows the soil moisture absolute difference maps after adjusting the model parameters. It can be seen from the modified difference maps that the soil moisture absolute difference range is reduced in all the four days, which are considered for creating difference maps. In addition to this, the overall absolute difference statistics have also reduced in all the cases (Table 10). It is concluded from this analysis that Landsat plays a significant role in improving the overall distribution and also reducing the mean absolute soil moisture difference. Therefore, this may be noted as a considerable advantage of Landsat data in hydrological modeling as it not only provides the indirect soil moisture estimations but has an upperhand in improving the overall spatial soil moisture patterns.

Table 9. Physical properties of sub-type (161_1) of 161 soil

Soil Depth (mm)	**Soil Texture (%)** (Sand/Silt/Clay)	**Hydraulic Conductivity** (mm/h)	**USLE_K**	**Available soil water capacity** (mm of soil/mm of water)	**Bulk density** (kg/m^3)
0-300	85/13/2	65.42	0.24	0.18	1.397
300-530	85/13/2	36.67	0.23	0.16	1.627
530-2000	94/5/1	75	0.20	0.10	1.64

Table 10. Spatial statistics for adjusted 161 soil in Wipperau Catchment

Sl.No.	**Date**	**Soil_type**	**Mean absoulte difference**	**Standard Deviation**
1	9-Mar-2016	161_1	6.00 (*10.15*)	3.41 (*1.58*)
2	12-May-2016	161_1	2.24 (*2.08*)	1.3 (*1.4*)
3	24-Aug-2016	161_1	1.09 (*1.97*)	0.87 (*1.2*)
4	9-Sep-2016	161_1	1.06 (*1.77*)	0.88 (*1.23*)

**Values in italics are the previous mean absolute error and standard deviation before the soil adjustment.*

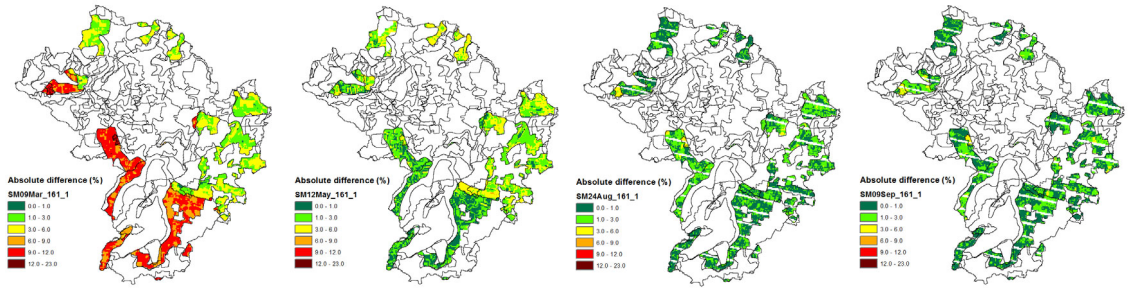


Fig. 17. Absolute soil moisture error maps for Wipperau under new soil subtype for 161 (161_1) soil.

4. Conclusions

The present study demonstrates the application of remote sensing and field data in evaluating SWAT simulated soil moisture at regional scale for two catchments in Northern Germany. The

calibrated and validated model was used to derive soil moisture uncertainty bands under different soils and crops of the irrigation season 2016 by considering the uncertainty of soil related model parameters. The results reveal that parameter uncertainty varies with different soils and different crops. The results showed that parameter uncertainty almost frames the observed soil moisture values. However, there is considerable uncertainty of the model structure, which can be related to the simplified soil water equations of SWAT, using a cascade of tipping buckets approach. In addition to this, there is also substantial variability in the observed soil moisture data. Temperature, NDVI and TVDI were calculated from Landsat images, which were converted into soil moisture by using several regression models. Regression models were trained by using the TDR measurements and the one, whose median and overall data spread matched relatively well with the TDR, was used for further analysis. The current study reveals that soil moisture extracted from Landsat could be used as a good indicator to evaluate the spatial and temporal dynamics of soil moisture extracted from hydrological model when field based soil moisture data is not enough. The field based soil moisture data is necessarily required to calibrate the spatial maps obtained from the Landsat. For both aspects, only few data are available due to the lack of soil moisture monitoring stations and due to the need of a clear sky for the remote sensing method. The spatial and temporal resolution of soil moisture procured from three sources are not exactly comparable. In addition to this, the soil moisture extracted from SWAT provides the average value of soil moisture of the upper 30 cm of the soil profile, whereas TDR gives the average soil moisture values from the upper ~16 cm of the soil profile. However, Landsat provides soil moisture of the top few centimeters of the soil profile only, without knowing the precise depth, from which the image is influenced. Therefore, further research could be done to improve the level of precision. This would involve extended studies at the field scale, which is beyond the focus of the current study. In the current research, the SWAT soil parameters were modified by investigating a consistent behavior of spatial soil moisture patterns from Land-sat images. The modification was confirmed by a higher resolution soil map and by field investigations. However, using the results of this analysis, the hydrological model can be rectified and applied with higher confidence in simulating soil moisture. SWAT provides daily estimates of soil moisture at finer resolution, which can be used for continuous simulation and forecasting of soil moisture. Possible fields of application are studies in planning and design of large scale irrigation systems and irrigation control schemes, and investigations about the impact of climate change on soil moisture and irrigation scheduling. However, further research is required to study the behavior of SWAT in simulating crop processes. This research helped to improve knowledge about large scale spatio-temporal dynamics of soil moisture at finer resolution in a humid country. Arid or semi-arid countries have an advantage of clear sky throughout the year, and irrigation demand is there during the entire growing season. As the spread and behavior of soil moisture is better matching during the dry season from all the sources in this study from a humid region, we expect that this method can be even better applied under semi-arid and arid conditions. Additionally, the Landsat extracted soil moisture can also be used for recalibrating the hydrological model with the aim of reducing uncertainty of simulated soil moisture. The findings of this study and the follow up studies can also be used for predicting and monitoring agricultural droughts in the catchments around the globe. Future research will be done in comparing multiple hydrological models and in simulating the effect of the quality of soil moisture simulation on irrigation control. The latter one brings uncertainty to water balance and water management studies in agricultural catchments, which is much less investigated compared to hydrological model parameter uncertainty.

Acknowledgements

Landsat Surface Reflectance products courtesy of the U.S. Geological Survey. Landsat Spectral Indices products courtesy of the U.S. Geological Survey Earth Resources Observation and Science Center. The authors are highly thankful to Univ.-Prof. Dr.-Ing. Martin Achmus and Michel-André Schröder of the Institute of Geotechnics at Leibniz University Hannover for providing the laboratory and test facilities and also the editor and anonymous reviewers for providing constructive comments for improving the overall manuscript. In addition to this, we thank Ms. Prajna K.A. for proof reading the revised manuscript. Dr. Tzoraki O. is supported by DAAD for the short research stay at the Leibniz University of Hannover.

References

Abbaspour, K.C., 2011. SWAT Calibration and Uncertainty Programs: A User Manual. Swiss Federal Institute of Aquatic Science and Technology, Eawag, Dubendorf.

Arnold, J.G., Allen, P.M., Muttiah, R., Bernhardt, G., 1995. Automated base flow separation and recession analysis techniques. Ground Water 33 (6), 1010–1018.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment part I: model development. J. Am. Water Resour. Assoc. 34 (1), 73–89.

Arnold, J.G., Moriasi, D.N., Gassman, P.W., Abbaspour, K.C., White, M.J., Srinivasan, R., Santhi, C., Harmel, R.D., Van Griensven, A., Van Liew, M.W., Kannan, N.,2012. SWAT: Model use, calibration, and validation. Trans. ASABE 55 (4), 1491–1508.

Beven, K., Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrol. Processes 6 (3), 279-298.

Brocca, L., Melone, F., Moramarco, T., Morbidelli, R., 2010. Spatial-temporal variability of soil moisture and its estimation across scales. Water Resour. Res.46 (2), 1–14.

Carlson, T.N., Gillies, R.R., Perry, E.M., 1994. A method to make use of thermal infrared temperature and NDVI measurements to infer surface soil water content and fractional vegetation cover. Remote Sensing 9, 161–173.

Carlson, T.N., Gillies, R.R., Schmugge, T.J., 1995. An interpretation of methodologies for indirect measurement of soil water content. Agric. Forest Meteorol. 77 (3),191–205.

Chen, L., Huang, Z., Gong, J., Fu, B., Huang, Y., 2007. The effect of landcover/vegetation on soil water dynamic in the hilly area of the loess plateau, China. Catena 70 (2), 200–208.

Chen, H., Zhang, W., Wang, K., Fu, W., 2010. Soil moisture dynamics under different land uses on karst hillslope in northwest Guangxi, China. Environ. Earth Sci. 61(6), 1105–1111.

DeLiberty, T.L., Legates, D.R., 2003. Interannual and seasonal variability of modelled soil moisture in Oklahoma. Int. J. Climatol. 23 (9), 1057–1086.

Entekhabi, D., Njoku, E.G., O'Neill, P.E., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T.J., Johnson, J., Kimball, J., 2010. The soil moisture active passive (SMAP) mission. Proc. IEEE 98 (5), 704–716.

Gardner, W., Kirkham, D., 1952. Determination of soil moisture by neutron scattering. Soil Sci. 73 (5), 391–402.

Goetz, S.J., 1997. Multi sensor analysis of NDVI, surface temperature and biophysical variables at a mixed grassland site. Int. J. Remote Sens. 18 (1),71–94.

Gruhier, C., Rosnay, P.D., Hasenauer, S., Holmes, T., Jeu, R.D., Kerr, Y., Mougin, E., Njoku, E., Timouk, F., Wagner, W., Zribi, M., 2010. Soil moisture active and passive microwave products: intercomparison and evaluation over a Sahelian site. Hydrol. Earth Syst. Sci. 14 (1), 141–156.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J. Hydrol. 377 (1), 80–91.

Gurr, C.G., 1962. Use of gamma rays in measuring water content and permeability in unsaturated columns of soil. Soil Sci. 94 (4), 224–229.

Han, E., Merwade, V., Heathman, G.C., 2012. Implementation of surface soil moisture data assimilation with watershed scale distributed hydrological model. J. Hydrol. 416, 98–117.

Lam, Q.D., Schmalz, B., Fohrer, N., 2011. The impact of agricultural best management practices on water quality in a North German lowland catchment. Environ. Monit. Assess. 183 (1-4), 351–379.

Landesamt für Statistik (2010). Statistische Berichte Niedersachsen, Landwirtschaftszählung 2010, heft 03, Bodennutzung. Rechtsform der Betriebe, Ökologischer Landbau, Zwischenfruchtbau, Bewässerung.

Li, M., Ma, Z., Du, J., 2010. Regional soil moisture simulation for Shaanxi Province using SWAT model validation and trend analysis. Sci. China Earth Sci. 53 (4),575–590.

Li, S., Liang, W., Zhang, W., Liu, Q., 2016. Response of soil moisture to hydro-meteorological variables under different precipitation gradients in the yellow river basin. Water Resour. Manage. 30 (6), 1867–1884.

Maier, N., Dietrich, J., 2016. Using SWAT for strategic planning of basin scale irrigation control policies: a case study from a humid region in Northern Germany. Water Resour. Manage. 30, 3285–3298.

Mapfumo, E., Chanasyk, D.S., Willms, W.D., 2004. Simulating daily soil water under foothills fescue grazing with the soil and water assessment tool model(Alberta, Canada). Hydrol. Processes 18 (15), 2787–2800.

Masek, J.G., Vermote, E.F., Saleous, N., Wolfe, R., Hall, F.G., Huemmrich, F., Gao, F.,Kutler, J., Lim, T.K., 2006. A landsat surface reflectance data set for north america, 1990-100. IEEE Geosci. Remote Sens. Lett. 3, 68–72.

Milzow, C., Krogh, P.E., Bauer-Gottwein, P., 2011. Combining satellite radar altimetry, SAR surface soil moisture and GRACE total storage changes for hydrological model calibration in a large poorly gauged catchment. Hydrol. Earth Syst. Sci. 15 (6), 1729–1743.

Moran, M.S., Clarke, T.R., Inoue, Y., Vidal, A., 1994. Estimating crop water deficit using the relation between surface–Air temperature and spectral vegetation index. Remote Sens. Environ. 49, 246–263.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50 (3), 885–900.

Muller, E., Decamps, H., 2001. Modeling soil moisture–reflectance. Remote Sens. Environ. 76 (2), 173–180.

Muttiah, R.S., Wurbs, R.A., 2002. Scale-dependent soil and climate variability effects on watershed water balance of the SWAT model. J. Hydrol. 256 (3), 264–285.

Narasimhan, B., Srinivasan, R., Arnold, J.G., Di Luzio, M., 2005. Estimation of long-term soil moisture using a distributed parameter hydrologic model and verification using remotely sensed data. Trans. ASAE 48 (3), 1101–1113.

Neitsch, S.L., Arnold, J.G., Kiniry, J.R., William, J.R., 2011. Soil and water assessment tool theoretical documentation version 2009. In: Texas Water Resources Institute, Technical Report No. 406, Texas A&M University, System College Station, Texas.

O'Neill, P., Entekhabi, D., Njoku, E., Kellogg, K., 2010. The NASA soil moisture active passive (SMAP) mission: overview. Geoscience and Remote Sensing Symposium (IGARSS), IEEE International Conference, July 2010, 3236–3239.

Park, J.Y., Ahn, S.R., Hwang, S.J., Jang, C.H., Park, G.A., Kim, S.J., 2014. Evaluation of MODIS NDVI and LST for indicating soil moisture of forest areas based on SWAT modeling. Paddy Water Environ. 12 (1), 77–88.

Phogat, V., Mahadevan, M., Skewes, M., Cox, J.W., 2012. Modelling soil water and salt dynamics under pulsed and continuous surface drip irrigation of almond and implications of system design. Irrigation Sci. 30 (4), 315–333.

Rajib, M.A., Merwade, V., Yu, Z., 2016. Multi-objective calibration of a hydrologic model using spatially distributed remotely sensed/in-situ soil moisture. J.Hydrol. 536, 192–207.

Richards, L.A., Gardner, W., 1936. Tensiometers for measuring the capillary tension of soil water. J. Am. Soc. Agronomy 28 (1), 352–358.

Richter, D., 1995. Ergebnisse methodischer Untersuchungen zur Korrektur dessystematischen Messfehlers des Hellmann-Niederschlagsmessers. Selbstverslag des Deutschen Wetterdienstes. Offenbach am Main.

Sandholt, I., Rasmussen, K., Andersen, J., 2002. A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. Remote Sens. Environ. 79, 213–224.

Schmugge, T.J., Jackson, T.J., McKim, H.L., 1980. Survey of methods for soil moisture determination. Water Resour. Res. 16 (6), 961–979.

Schultz, G.A., 1988. Remote sensing in hydrology. J. Hydrol. 100 (1-3), 239–265.

Seneviratne, S.I., Corti, T., Davin, E.L., Hirschi, M., Jaeger, E.B., Lehner, I., Orlowsky, B., Teuling, A.J., 2010. Investigating soil moisture–climate interactions in a changing climate: a review. Earth Sci. Rev. 99 (3), 125–161.

Sloan, P.G., Moore, I.D., 1984. Modeling subsurface stormflow on steeply sloping forested watersheds. Water Resour. Res. 2 (12), 1815–1822.

Sun, L., Seidou, O., Nistor, I., Goita, K., Magagi, R., 2016. Simultaneous assimilation of in situ soil moisture and streamflow in the SWAT model using the Extended Kalman Filter. J. Hydrol. 543, 671–685.

Tavakoli, M., De Smedt, F., 2013. Validation of soil moisture simulation with a distributed hydrologic model (WetSpa)? Environ. Earth Sci. 69 (3), 739–747.

Timm, L.C., Pires, L.F., Roveratti, R., Arthur, R.C.J., Reichardt, K., Oliveira, J.C.M.D., Bacchi, O.O.S., 2006. Field spatial and temporal patterns of soil water content and bulk density changes. Scientia Agricola 63 (1), 55–64.

Topp, G.C., Davis, J.L., Annan, A.P., 1980. Electromagnetic determination of soil water content: measurements in coaxial transmission lines. Water Resour. Res.16, 574–582.

Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sens. Environ. 185, 46–56, http://dx.doi.org/10.1016/j.rse.2016.04.008.

Vivoni, E.R., Rango, A., Anderson, C.A., Pierini, N.A., Schreiner-McGraw, A.P., Saripalli, S., Laliberte, A.S., 2014. Ecohydrology with unmanned aerial vehicles. Ecosphere 5 (10), 1–14.

Walker, J.P., Willgoose, G.R., Kalma, J.D., 2001. One-dimensional soil moisture profile retrieval by assimilation of near-surface observations: a comparison of retrieval algorithms. Adv. Water Resour. 24, 631–650.

Wessolek, G., Kaupenjohann, M., Renger, M. (Eds.), 2009. Technische Universität Berlin, Bodenökologie und Bodengenese, Heft 40.

Western, A.W., Grayson, R.B., Blöschl, G., Willgoose, G.R., McMahon, T.A., 1999. Observed spatial organization of soil moisture and its relation to terrain indices. Water Resour. Res. 35 (3), 797–810.

Wittenberg, H., 2003. Effects of season and man-made changes on baseflow and flow recession: case studies. Hydrol. Processes 17 (11), 2113–2123.

Wittenberg, H., 2015. Groundwater abstraction for irrigation and its impacts on low flows in a watershed in northwest Germany. Resources 4 (3), 566–576.

Xin, J.F., Tian, G.L., Liu, Q.H., Chen, L.F., 2006. Combining vegetation index and remotely sensed temperature for estimation of soil moisture in China. Int. J.Remote Sens. 27, 2071–2075.

Zhang, D., Zhou, G., 2016. Estimation of soil moisture from optical and thermal remote sensing: a review. Sensors 16 (8), 1308.

Zucco, G., Brocca, L., Moramarco, T., Morbidelli, R., 2014. Influence of land use on soil moisture spatial-temporal variability and monitoring. J. Hydrol. 516, 193–199, http://dx.doi.org/10.1016/j.jhydrol.2014.01.043.

This chapter is an edited version of the following original scientific article:

Uniyal, B.; Dietrich, J.; Vasilakos, C.; Tzoraki, O. (2017): Evaluation of SWAT simulated soil moisture at catchment scale by field measurements and Landsat derived indices. Agricultural Water Management 193, 55-70.

Publisher: Elsevier

VIII. Rooftop Rainwater Harvesting for Mombasa: Scenario Development with Image Classification and Water Resources Simulation

Robert O. Ojwang [1], Jörg Dietrich[2], Prajna Kasargodu Anebagilu [2], Matthias Beyer[3] and Franz Rottensteiner[4]

[1] Coast Water Services Board, Mombasa, Kenya
[2] Institute of Hydrology and Water Resources Management, Leibniz University Hannover, Germany
[3] Federal Institute for Geosciences and Natural Resources (BGR), Hannover, Germany
[4] Institute of Photogrammetry and GeoInformation, Leibniz University Hannover, Germany

Abstract

Mombasa faces severe water scarcity problems. The existing supply is unable to satisfy the demand. This article demonstrates the combination of satellite image analysis and modelling as tools for the development of an urban rainwater harvesting policy. For developing a sustainable remedy policy, rooftop rainwater harvesting (RRWH) strategies were implemented into the water supply and demand model WEAP (Water Evaluation and Planning System). Roof areas were detected using supervised image classification. Future population growth, improved living standards, and climate change predictions until 2035 were combined with four management strategies. Image classification techniques were able to detect roof areas with acceptable accuracy. The simulated annual yield of RRWH ranged from 2.3 to 23 million cubic meters (MCM) depending on the extent of the roof area. Apart from potential RRWH, additional sources of water are required for full demand coverage.

1 Introduction

Rainwater harvesting is a technique used to collect and store rainwater e.g., from buildings, rock catchments, and land or road surfaces. The authors of [1,2,3] describe rainwater harvesting to be a dominant contributor for sufficing urban water demand. Rooftop rainwater harvesting (RRWH) refers to the collection and storage of water from rooftops [4]. The level of expertise required is low and ownership can be at a household level, making it easily acceptable to many people [1,5,6]. RRWH can support the water supply in almost any place either as a sole source or by reducing stress on other sources through water savings. The authors of [7] observed that the most important feature of RRWH at a domestic level is its ability to deliver water to households “without walking”. This is particularly important in developing countries where women and children have to walk over long distances to fetch water. RRWH can be one aspect of the adaptation of water supply systems to climate change [8]. In Sub-Saharan Africa, the reliability of RRWH systems for domestic water supply can be improved by the consideration of rainfall characteristics, e.g., the number of events above a certain threshold, wet spells, etc., and improved technical design [9].

The quantity and quality of the harvested rainwater greatly depend on the type of roofing material used. Hard surfaces like iron, concrete, and tiles produce the highest amount of collected water because they have high runoff coefficients. A study by [10] showed that

galvanized steel yielded the best quality rainwater that met the WHO (World Health Organization) drinking water guidelines for chemical, physical, and biological parameters. Furthermore, the slope of the roof has considerable influence on the roof runoff [11]. The "first flush" is a criteria often used for the design of RRWH systems. It describes the amount of initial rainfall, which is needed after a dry period to remove contaminants such as particles, dirt, bird droppings, and insect bodies from the roof and the gutter. First flush diverters need to be incorporated into the system in order to protect the water quality in the collection tank from contamination [1,12]. Due to first flush diversion, the quantitative yield of RRWH systems is reduced.

The amount of water harvested can be calculated using the rational method commonly used for very small urban catchments if the roof area, amount of rainfall, and the roof runoff coefficient are known. The runoff coefficient represents losses due to evaporation and leakages [1,13]. A study in Jordan showed that the potential of rainwater harvesting is about 15.5 MCM per year, which allowed potable water savings ranging from 0.3% to 19.7% in 12 administrative units [1]. In the UK, an average water saving efficiency (ET) of 87% over a period of 8 months was reported for the total toilet flushing demand of an office building using RRWH. ET is the percentage ratio of rainwater supplied to the total estimated demand [14]. For Iran, [15] reported supply rates of 75% but for different durations of 40–75% of the time, depending on climatic conditions.

The use of collected rainwater for domestic purposes (tertiary uses like gardening) is a major component of water supply in the rural areas of South Africa, with 96% of 34,000 RWH tanks being located in rural regions [8]. When harvested water is used as potable water, the quality becomes paramount. The authors of [6] showed that most people in Zambia expressed interest in RRWH, but were concerned about its quality. Several other studies have shown that rainwater usually meets the WHO standards for physical and chemical parameters but may fail regarding the biological parameters, i.e., fecal and total coliform counts [10,16,17,18].

The key parameters for estimating RRWH water yield are roof area and rainfall. Another major factor is the estimation of roof areas, which is difficult especially for unplanned city areas. The authors of [10] list several methods:

- Sampling: representative samples of rooftops are obtained and extrapolated to the total area. This method is suitable for estimating roof areas for large areas;
- Multivariate sampling: correlations are drawn between additional variables (e.g., population) and roof area;
- Complete census: gives the most accurate results but involves the computation of the entire area of the rooftops in the area of interest by using statistical information like floor area, number of floors, and number of housing units;
- Digitization or image classification tools can be used from remotely sensed high-resolution images to compute the roof areas with a Geographical Information System (GIS).

The authors of [1] used a complete census to estimate roof areas in Jordan. Available information on the dwelling units such as different types of units, number of units per type, and average area per type were used to estimate the roof area in each of the governorates.

The same approach was applied in another study in Seoul, South Korea to determine the city's rooftop photovoltaic (PV) potential [19]. The results were validated by using automated vector detection software. In a study of the informal settlement of Diepsloot near Johannesburg, South Africa, an automatic feature extraction (image classification) method was applied using aerial satellite imagery to extract roof areas with 80% accuracy [20]. Similarly, in the Kibera slums, Nairobi, feature extraction was successfully used to estimate rooftop areas, which was then used to derive the population in the slum [21].

The rooftop area is a crucial input for the incorporation of RRWH systems into water resource system models, which can be applied for the quantitative planning of water resources and the development of water policies. WEAP (Water Evaluation and Planning System), developed by the Stockholm Environment Institute (SEI, Stockholm, Sweden), is a planning tool extensively used in integrated water resources management. It is both a model for simulating water systems in an integrated manner (natural and manmade components/infrastructures) and a policy oriented decision-support system (DSS) [22]. Demand and supply sites are considered concurrently. WEAP follows the principle of the "scenario-based gaming approach" that has been developed to reduce water demand-supply conflicts within the area of interest [23]. The model uses scenarios to promote stakeholder involvement in the entire water resources planning and decision making process. The water resources system in WEAP is represented by demand sites, supply sites, catchments, withdrawal points, transmission links, wastewater treatment, environmental needs, and the generation of pollution. Depending on the need and data availability, WEAP simulates several aspects such as sectoral water demands, water allocation rights and priorities, ground and surface water flows, reservoir operations, and the assessment of vulnerability and cost-benefit analysis, amongst others.

The authors of [24] tested WEAP's demand management scenario evaluation in the water-stressed Olifants river basin, South Africa, and [25] applied WEAP in the study of the Upper Ewaso Ng'iro North Basin, Kenya to balance the water requirements of competing users against the available water resources in the basin. The study found that the use of WEAP improved the complex system of demand-supply of the basin. Applications of WEAP in the development of RRWH strategies have not been documented so far.

In this study, WEAP is applied to develop a variety of future scenarios of RRWH for the city of Mombasa by using projections of population growth and climate change. This research contributes to closing the gaps in the current methods of investigating RRWH by combining remote sensing with water resources modelling. The main objectives of the study are as follows:

- Determination and discrimination of rooftop areas and different roof types from high resolution satellite images;
- Setup and parameterization of an extended WEAP model with an implemented simple RRWH scheme for large scale planning;
- Implementation of future scenarios in WEAP and evaluation of their implications and potential for long-term management of the urban water supply.

2 Materials and Methods

2.1 Study Area and Data

Mombasa City is the second largest city in Kenya with an estimated 1.1 million inhabitants and a land area of 229.7 km2. The city is located in the southern part of Kenya and is divided into four main areas: Island, South Mainland, West Mainland, and North Mainland (Figure 1).

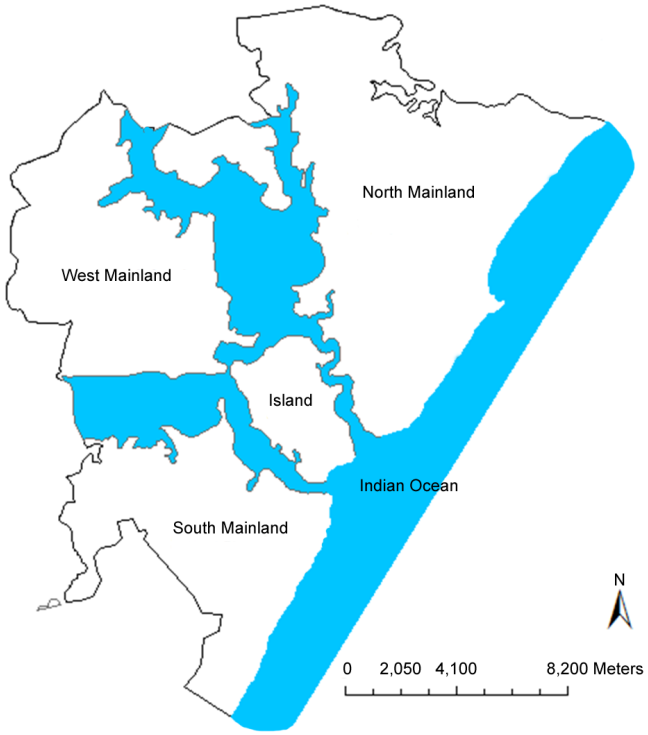


Figure 1. Map of Mombasa City showing the four main zones in the study area.

The city experiences a tropical climate, which is hot and humid throughout the year with a mean daily minimum and maximum temperature of 22 °C and 30 °C, respectively. Annual precipitation is 1024 mm (data from 1984 to 2013 by the Kenya Meteorological Department). There are two rainy seasons. For the long rainy season between April and July, the average monthly rainfall is 134 mm. Between October and December, there are short rains with 100 mm average monthly rainfall. During the dry season, the average monthly rainfall is 37 mm with several months without rain. These variations in climatic conditions are attributable to the S-E and N-E monsoon winds and oceanic factors. The average monthly areal precipitation over the catchment was computed as the arithmetic mean of 10 rain gauge stations within the catchment. Annual potential evaporation exceeds the rainfall by magnitudes, which results in freshwater deficits during the dry periods. Reference evapotranspiration (ET0) data were obtained from the FAO (Food and Agriculture Organization of the United Nations) CLIMWAT 2.0 database using Moi International Airport, Mombasa station located at 4.03° S and 39.61° E. According to the Kenya Integrated Water Resources Management and Water Efficiency Plan, the country's annual average water availability in 2002 was 647 m3 per capita. This ranks Kenya as a water scarce country according to [26], where below 1700 m3/capita/year means water stress, below 1000 m3/capita/year means water scarcity, and below 500

m3/capita/year means absolute water scarcity. The situation in Kenya is predicted to worsen to 235 m3 per capita by the year 2020 [27].

For future climate projections, downscaled precipitation and temperature data from the CIMP5 Global Climate Models (GCM) were used for the RCP 4.5 and RCP 8.5 pathways. The Intergovernmental Panel on Climate Change (IPCC) in its Climate Change 2014 Synthesis Report uses Representative Concentration Pathways (RCPs) to describe four different pathways of greenhouse gas concentration in the atmosphere, named after the change in radiative forcing in W/m^2 compared to pre-industrial times. In the case of RCP 4.5, the global mean surface temperature is likely to increase from 1.1 K to 2.6 K for 2081–2100 relative to 1986–2005.

The CIP datasets provide time series data (observed and downscaled projections) for different climatic variables for weather stations across Africa. The projections are compiled of downscaled products of 11 GCMs, summarized in Table 1. From the long-term historical (1971–2000) monthly ET0 data, future values can be estimated based on temperature changes. The authors of [28] suggested that for each degree rise in temperature, there is a corresponding 5% increase in ET0. The historical ET0 values were then corrected for both RCP 4.5 and RCP 8.5 to obtain future estimations.

In order to address model uncertainty, multi-model ensemble averages can be used; however, spatiotemporal information, especially of extreme events, can be lost [29,30]. The evaluation of bias (Table 1) and root mean square error (RMSE, not shown) between the downscaled RCP 4.5 and RCP 8.5 data (monthly series) and the observed historical data for the period from 1984 to 2013 showed that the bias for the GFDL-ESM2G model is lower than the ensemble bias, but the RMSE is lowest for the ensemble. Furthermore, it is not clear if the bias is non-stationary. Thus, using the model ensemble is preferable over using one single model. Multi-model means were used to generate future monthly rainfall series in this study.

Only domestic water demand was considered in this study. In order to estimate the per capita water consumption (water use rate), the recommended amounts for different categories according to the practice manual for water supply services [31] together with poverty level data for Mombasa City [32] were used (Table 2). The estimated water use rate is around 116 L per capita per day (LCPD).

Table 1. Details of the GCM models used and their bias (%) in the height of precipitation for the historical period from 1984–2013.

Model Name	RCP 4.5	RCP 8.5	Climate Modelling Institution/Centre
MIROC-ESM	−12.70	−15.70	National Institute for Environmental Studies, Japan
CNRM-CM5	−8.00	−8.10	Centre National de Recherches Meteorologiques, France
CAN-ESM2	−3.80	−2.30	Canadian Centre for Climate Modelling and Analysis
FGOALS-S2	−13.80	−19.70	Institute of Atmospheric Physics, Chinese Academy of Sciences
BNU-ESM	−16.00	−15.00	Beijing Normal University
MIROC5	9.20	8.90	National Institute for Environmental Studies, Japan
GFDL-ESM2G	0.40	2.70	Geophysical Fluid Dynamics Laboratory, USA
MIROC-ESM-CHEM	−15.50	−15.40	National Institute for Environmental Studies, Japan
GFDL-ESM2M	−1.20	−1.70	Geophysical Fluid Dynamics Laboratory, USA
MRI-CGCM3	26.90	22.20	Japan Meteorological Agency
BCC-CSM1-1	−10.90	−10.70	Beijing Climate Centre
Ensemble Average	−4.10	−5.00	-

Table 2. Domestic water use rates for Mombasa City.

Category	Persons	Water Use Rate	Remarks
	Percent	LCPD	
High Class Houses (HCH)	5.00%	250	Poverty level was 37.6% in 2013, remaining 62.4% assumed as 5% for HCH and 57.4% for MCH
Medium Class Houses (MCH)	57.40%	150	
Low Class Houses with individual connections (LCH_IC)	18.80%	75	
Low Class Houses without individual connections (LCH_WIC)	18.80%	20	
Weighted		116	-

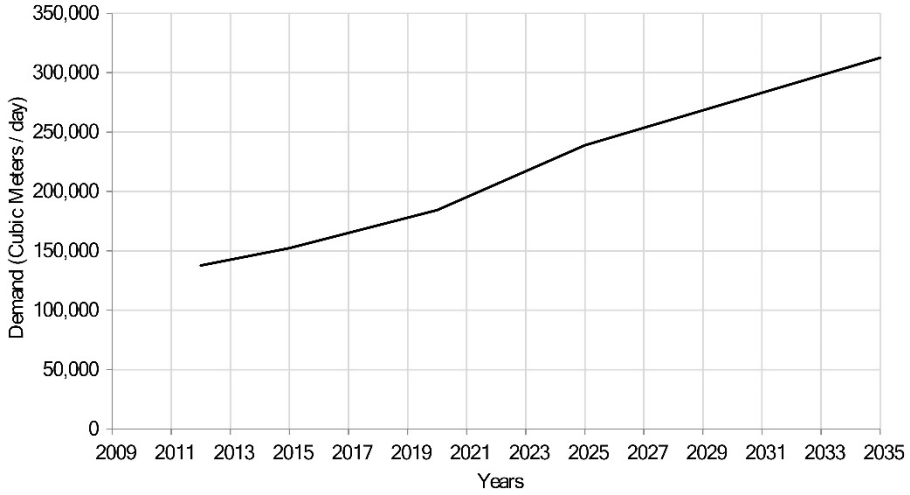


Figure 2. Water demand projections for Mombasa City [34].

Assuming the population will continue to grow at the average rate of 3.2% experienced between 1999 and 2009 [33], the population is expected to rise to 2.2 million people by 2035, leading to an increased water demand from 150,000 m3/day to 320,000 m3/day (Figure 2). The existing water supply to the city is 102,000 m3/day from Mzima springs, Tiwi boreholes, Marere Springs, and the Baricho Wellfield managed by the Coast Water Services Board (CWSB). Considering system losses of around 47%, the current demand coverage is low [34,35].

The water supply master plan [34] identified several projects to help bridge the gap, mostly by the expansion of the Baricho Wellfield and Mzima Springs to full capacity, the construction of the 228,000 m3/day Mwache Dam, and the acquisition of the Mkurumudzi dam. The plan considered RRWH, desalination, and wastewater reuse as alternative sources of water even though no estimates of quantities are available at this point. The biggest challenge is the large amount of investment capital required. It is further doubtful if the projects will fully cover the rising demand upon completion or if continued unsustainable extraction may lead to depletion. As a stopgap measure, the Mombasa Water and Sanitation Company (MOWASCO, Kenya) adopted water rationing of 6 h of supply per day [35]. Consequently, around 13,000 households use individual boreholes and hand-dug wells to supplement the conventional piped water supply [32]. Seawater intrusion due to over extraction may soon be a problem.

2.2 Overview of the Methodology

This study involved the following key steps: (i) data collection; (ii) image classification to estimate roof areas within Mombasa City; (iii) using estimated areas to calculate the rainwater harvesting potential and (iv) future scenario analysis with WEAP. In the scenario analysis, impacts of climate change on rainwater harvesting were also incorporated by using an ensemble of projections as shown in Table 1. The overall procedure and workflow is shown in Figure 3.

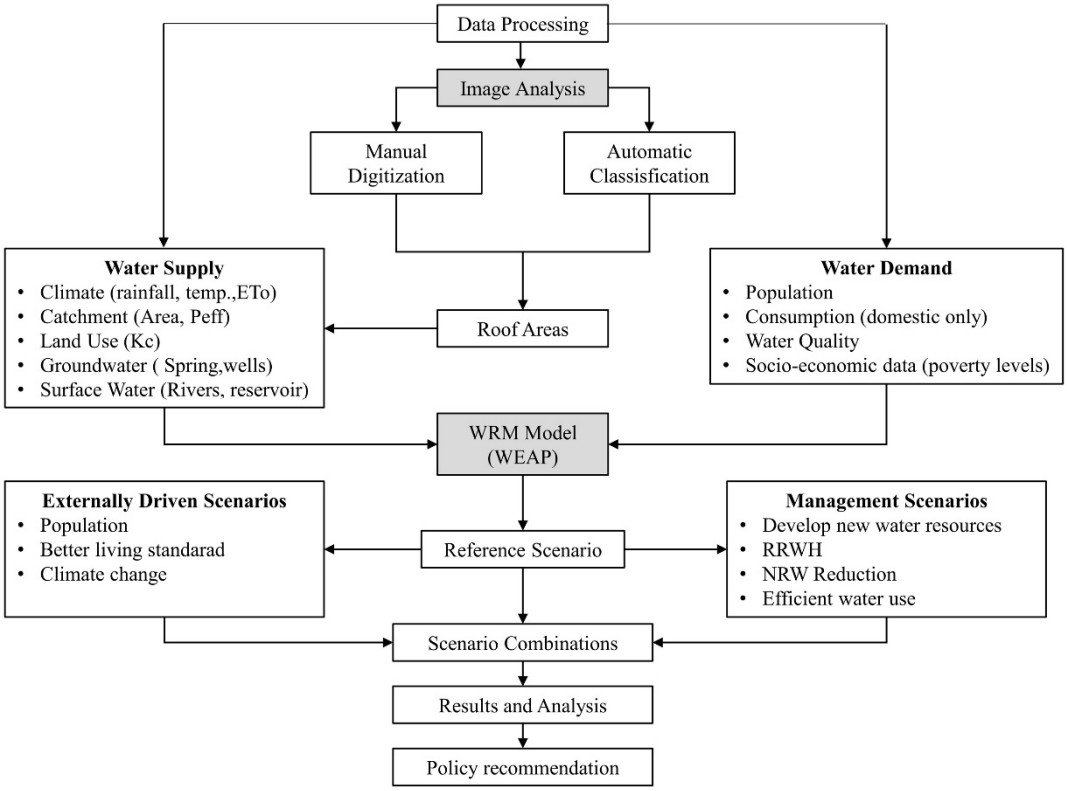


Figure 3. Methodological framework and workflow.

2.3 Roof Area Estimation

Roof areas were estimated manually by digitization and via automatic image classification using the ArcGIS® software by ESRI (Environmental Systems Research Institute, Redlands, CA, USA). The commonly used roofing materials in the city (tile, iron, and concrete making up 90% of the roof area) were considered. The following criteria were used to select the satellite images for this study from the Digital Globe foundation data for 2010 to 2014: (i) spatial resolution (high); (ii) cloud cover (low); and (iii) spatial coverage (extent of the study area covered). The best images in terms of spatial resolution were available from the satellite WorldView2 (WV2). Two WV2 images, which covered the whole study area, were selected for further processing and analysis (Table 3).

For manual digitization, only planned areas within the study area were targeted (referred to subsequently as the "control area") due to the ease of roof type identification for digitization. The other reason was that unplanned congested areas might not be suitable for

RRWH since they lack the required space. Furthermore, the inhabitants may not be able to afford the system. Figure 4 shows the difference between planned and unplanned areas within the city.

Table 3. Selected images for processing (digitization and classification). The spatial resolution is after pan-sharpening.

Sensor Name	Acquisition Date	Cloud Cover (%)	Multispectral Bands	Off-Nadir (°)	Spatial Resolution (m)
WV-2	29/11/2013	0.6	8	23.10	0.5
WV-2	15/08/2013	8.9	8	6.60	0.5

(a) (b)

Figure 4. Planned areas (a) and unplanned areas (b) (informal settlements).

Automatic image classification was used to estimate roof areas for the whole city. The performance of the automatic classification was assessed using the results from the manual digitization for the common areas covered by both techniques. In this study, supervised classification using the Gaussian maximum likelihood (ML) method was adopted [36]. Six classes were used: tiled roofs, iron roofs, concrete roofs, vegetation, roads, and ground. In ML classification, the user has to provide training areas for each class. In the training phase, the parameters of a Gaussian mean and covariance matrix are estimated from the image feature vectors of each class. The feature vector of an image pixel collects all the spectral values observed at that pixel. In the classification phase, for each pixel to be classified, the feature vector of that pixel is used to compute the Gaussian probability density for each class using the parameters estimated in the training phase. This value is referred to as the likelihood for the feature vector to belong to the respective class. The pixel is assigned to the class of maximum likelihood.

2.4 The WEAP Model for Mombasa City

2.4.1 Conceptual Model Scheme

The main water infrastructure of the city of Mombasa was implemented into WEAP as a conceptual model (Figure 5). Subsequently, a description of the demand and supply elements was incorporated into the model:

a) Demand site: Even though six different demand sites have been shown in the model (Mombasa City, Malindi Town, Kilifi Town, Kwale Town, Mariakani Town, and Voi Town), the study is focused only on Mombasa City and the rest are used to provide a complete picture of the sharing of water resources in the Coastal region.

b) Water sources:

I. Current situation:

- The city receives water from Mzima Springs, Baricho boreholes, Marere Springs, Tiwi-Likoni boreholes, and individual dug-out wells.
- The rivers of Marere, Mwache, Sabaki, and Rare are some of the rivers that flow around Mombasa City. However, currently there is no abstraction from these rivers.

II. Future (presented in the model):

- The head flow generated from the Mwache catchment feeds the Mwache River. The Mwache Reservoir is expected to supply water from 2020.
- The rooftop areas are implemented as five catchment nodes, corresponding to the roof areas for each of the four zones in Mombasa, namely North Mainland (NML), South Mainland (SML), West Mainland (WML), Island, and new buildings to be constructed in the future. The water, which is collected from the rooftops of these five catchments, is directed into one reservoir "RRWH", which is modelled as a local reservoir.
- Operation of Mkurumudzi Dam in supplying water to Mombasa is expected to start from 2030 [34].
- Return flows are not considered in the WEAP model because the city mainly depends on onsite wastewater disposal methods such as pit latrines, cesspits, and septic tanks that do not allow any return flows to the rivers, and the sewer system of the city drains to the Indian Ocean. The two wastewater treatment plants, Kizingo and West Mainland, serve a very small population and also discharge directly to the Indian Ocean.

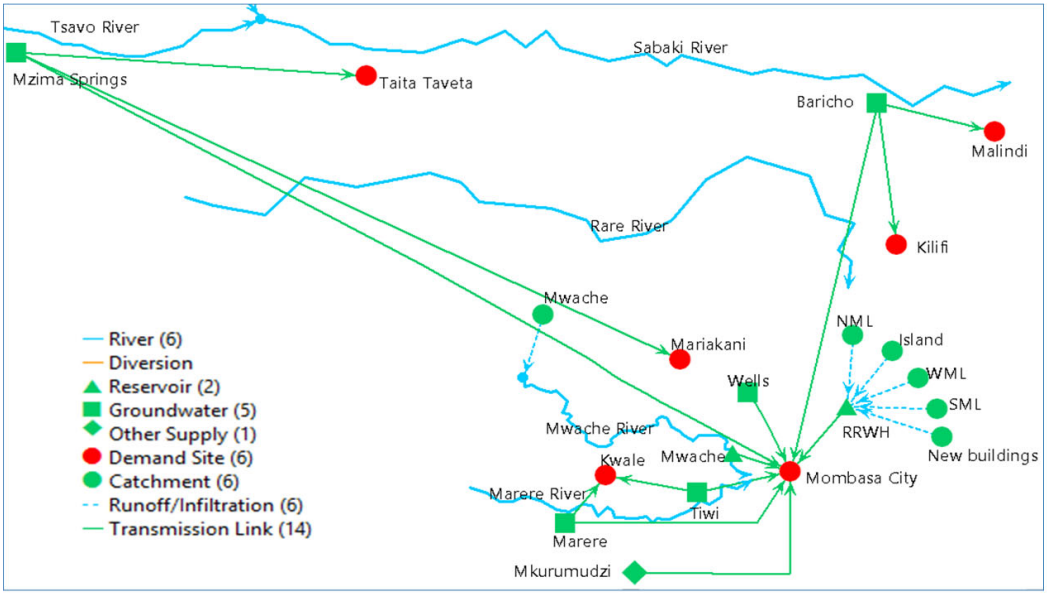


Figure 5. Conceptual model for Mombasa city (not drawn to scale).

2.4.2. Catchment and RRWH Implementation

The FAO rainfall runoff (simplified coefficient) method as implemented in WEAP was used in this study to simulate both natural and RRWH runoff generation from the catchments. This simple method calculates runoff by subtracting evapotranspiration from precipitation. The effective precipitation parameter ranges between 0% and 100% with 0% indicating that all precipitation produces runoff while 100% means that all precipitation is available for evapotranspiration.

The Mwache catchment covers approximately 2250 km2 and the main vegetation in the catchment are deciduous forest and dry grasslands covering 38% and 62% of the area, respectively [37]. The FAO simplified method uses a crop coefficient Kc, which is relative to the reference crop for a particular land class type. The FAO Paper 56 for Irrigation and Drainage recommends Kc values for deciduous trees on grass between 0.8 and 0.9 and for deciduous trees with bare ground between 0.3 and 0.4. Assuming that dry grasslands are similar to bare ground, the Kc value was set to 0.5, while the effective precipitation was set to 65%. For the Mwache Reservoir, a storage capacity of 118.7 MCM was used, while the volume elevation curve and estimated monthly average reservoir evaporation were obtained from [37].

The five rooftop catchments represent the total rooftop area of each of the four zones of Mombasa with existing buildings (summarized as one catchment each) and one zone with new buildings. The size of the RRWH catchment area was determined from the roof identification by image classification depending on the scenario of roof usage considered. The precipitation of the RRWH catchments was corrected for the first flush loss occurring at the beginning of rainfall. According to the Texas Manual on Rainwater Harvesting [38], the first flush loss amounts to 1 Gallon (1 Gallon = 3.785 Liters) for every 100 Square feet (1 Square foot = 0.092 Square meters) of roof area, which is equivalent to 0.4 mm of rainfall. The average monthly number of rainy days from the observed data from 1984 to 2013 was used to estimate the average monthly first flush loss, which was then subtracted from the precipitation input. The runoff coefficients for the different roof materials were taken from [7], using 0.9 for iron and 0.8 for tile and concrete. Mombasa roofs have different slopes, hence there is no characteristic pitch. The roof pitches could not be obtained from the satellite images used in this study. Stereo images were not available, and an exhaustive ground observation was not possible. Similar to [1] and other studies, the roof angle could not be considered and average runoff coefficients were used based on the material alone.

The water collected from the rooftops of all five catchments is diverted into one reservoir in the WEAP model as a conceptual representation of the RRWH storage systems. This simplification was introduced because the potential of RRWH for additional water supply is investigated on a larger scale here. In reality, there can be centralized and/or decentralized solutions for the storage of the harvested water. The efficiency of different storage techniques, however, is not a subject of this study. The RRWH system storage was modelled using a single reservoir, which receives runoff from the five RRWH catchments (NML, SML, WML, Island, and new buildings) through runoff links. Since the target is to collect all of the rainwater, the storage capacity is set as unlimited. There is no evaporation from the RRWH reservoir,

because it is simply a conceptual representation of closed storage tanks in WEAP. Thus, a fictitious volume elevation curve is used (1 m rise in level for each additional 1 m3). Operation-wise, the top of the conservation zone is set to the storage capacity and no dead storage is provided since all of the water in the tank is assumed to be available for use depending on the demand requirements.

Flow from the supply sources to the demand sites was implemented using transmission links with the consideration of losses due to leakages. Losses were taken as 47% based on the latest Impact Report No. 7 [35] and 20% was assumed for the RRWH structures, representing the losses occurring mainly in the gutters, downpipes, and storage tanks.

2.4.3. Baseline Scenario

Due to the availability of suitable satellite images for roof area estimation and observed precipitation data, the year 2013 was chosen as the base year (Current Accounts Year in WEAP). For the last year of the scenarios, the year 2035 was selected in order to coincide with the last year of the Coast Water Services Board planning horizon and to be after the end of the Government of Kenya's Vision 2030 development blueprint. The temporal resolution of the model was set to yearly with monthly time steps. Parameters that might be subject to changes under different scenarios such as population growth rates, water use rates, non-revenue water (NRW) levels, or crop coefficients (K_c) were further implemented as "Key Assumptions". The data used in the current accounts is compiled in Table 4.

Table 4. Input data for current accounts.

Source	Parameters	Value	Reference
Roof	Crop coefficient, K_c	0.1	Lower than bare soil (0.3 from FAO Paper 56)
	Effective rainfall, Peff	Iron-10% Tile and concrete-20%	Reasonable assumptions
Groundwater sources: Baricho, Mzima, Tiwi, Marere, Ind. Wells	Storage Capacity Initial Storage (MCM) Max Withdrawal (MCM) Recharge	Unlimited 80, 82, 7.3, 7.3, 16 Same as initial storage 83, 405, 21, 15, 23	Reasonable assumptions Mumma and Lane, 2010; TAHAL/Bhundia Consultants, 2013; JBG Gauff Ingenieure, 1995; Samez Consultants, 2008; Sincat/Atkins Consultants, 1994; Fichtner/Wanjohi Consultants, 2014
Mwache Dam	Storage capacity (MCM) Evaporation Rate Effective rainfall, Peff Crop coefficient, K_c	118.7 Monthly rates 65% 0.5	CES Consultants, 2013 -
Transmission	Loss in transmission links Loss in RRWH transmission	47% 20%	WASREB, 2014 Reasonable assumptions

2.4.4. Future Scenarios for Mombasa

The Reference Scenario inherits all of the information and data set up under the Current Accounts year (2013) and extends it over the entire timeframe (2014–2035) with no interventions to improve demand coverage. Here, the water supply remains at 102,000 m3/day from the main sources and 29,143 m3/day from the individual wells, with losses (NRW) of 47%. The per capita water consumption rate and population growth rate remain at 116 LCPD and 3.2% throughout the period, respectively.

Future scenarios for Mombasa are created to investigate the combined influence of (i) possible future changes of external factors, which are out of direct control of the water managers and which are uncertain, such as population growth, socio-economic dynamics, and climate change; and (ii) management decisions such as construction and expansion of more water sources, reduction of NRW, and implementation of RRWH. The growth assumptions for future scenarios are summarized in Table 5.

Table 5. Growth assumptions for Future Scenarios (External Factors).

Parameters	Value	Reference
High Population Growth (HPG) rate	4.2%	Kenya National Bureau of Statistics (2009), BCEOM/Mangat (2011) and Mombasa County (2014)
Low Population Growth (LPG) rate	1.9%,	
Increased water consumption due to better standard of living	116 LPCD to 155 LPCD	

Table 6. Planned future flows to Mombasa City in m^3/day (TAHAL, 2013).

Source	Capacity	Current	Phase I		Phase II	Phase III	
		2014	2017	2020	2025	2030	2035
Baricho	175,000	60,000	82,000	55,805	106,594	80,395	80,395
Mzima	105,000	24,000	24,000	15,292	13,370	59,050	59,050
Marere	12,000	8000	8000	7135	6051	3173	3173
Tiwi	13,000	10,000	10,000	10,000	10,000	8662	8662
Mwache	228,000	0	0	95,595	102,859	145,838	145,838
Mkurumudzi	20,000	0	0	0	0	15,191	15,191

Table 7. Scenario combinations under the normal population growth rate.

Scenario Combination	Description
NWS/RRWH_4	Existing system with new water sources developed (NWS) and RRWH_4 (using all new roofs) implemented
NWS/NRWS	Existing system with NWS and non-revenue water (NRWS) reduction strategy implemented
RRWH_4/NRWS	Existing system with both RRWH_4 and NRWS strategy implemented
NWS/EWU	Existing system with new water sources developed and water use efficiency (EWU) improved
RRWH_4/NRWS/EWU	Existing system without new water sources developed but RRWH_4, NRWS, and EWU implemented
NWS/NRWS/EWU	Existing system with new water sources developed and NWRS and EWU implemented, but no RRWH_4
NWS/RRWH_4/NRWS/EWU	All strategies implemented (new water sources, RRWH, non-revenue water reduction, and efficient water use)

The scenarios driven by management decisions were: (i) Development of New Water Sources (NWS); (ii) Reduction of Non-Revenue Water Strategy (NRWS) where the NRW levels decrease from 47% to 20% to meet the Water Services Regulatory Board (WASREB) sector target; (iii) Efficient Water Use (EWU) where per capita consumption decreases from 116 LPCD to 93 LPCD; and (iv) RRWH scenarios where rainwater harvesting is practiced. The estimated allocated future flows to different parts of the city are shown in Table 6. The following five RRWH scenarios are investigated: (i) All existing roofs are used (RRWH_1); (ii) only new roofs are used (RRWH_2); (iii) selected existing buildings are used (RRWH_3); (iv)

selected existing roofs and all new roofs are used (RRWH_4); and (v) all existing roofs and all new roofs to be used (RRWH_5). The different combinations of scenarios used in the model are shown in Table 7.

3 Results and Discussion

3.1 Determination of Rooftop Area

Figure 6 shows a visual comparison of a part of the classified image with the ground truth obtained by digitization (only roof classes shown). The visual assessment reveals that the classifier performed well in separating the different classes of roof material. The extent or degree of accuracy however cannot objectively be assessed by visual interpretation alone.

The results of manual digitization suggest that the selected control area has around 3 km2 of suitable roofs for RRWH (based on non-congested planned areas). Table 8 summarizes the results of the manual digitization for the selected areas. Apart from representing one of the scenarios, the manual digitization was further used to assess the performance of the automatic classification technique. The total roof area within the city was found to be approximately 28 km2 based on automatic image classification with 18.1 km2 of iron, 5.6 km2 tile, and 4.3 km2 concrete (Table 9).

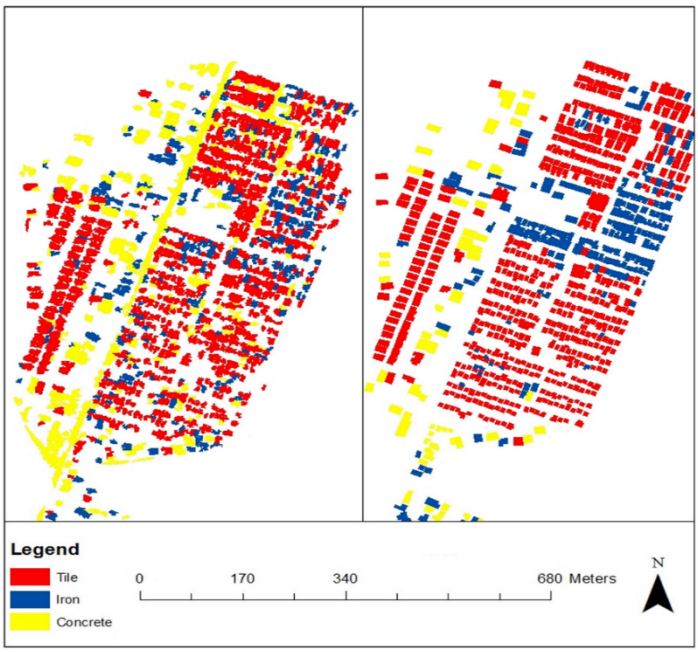


Figure 6. Comparison between manually digitized and automatic classification for a portion of the classified area (only roofing materials shown).

Table 8. Area of different roof materials (in m^2) resulting from the manual digitization for the selected control areas.

Zone	Tile	Iron	Concrete	Total
Island	610,339	114,854	298,186	1,023,379
North Mainland	1,139,521	313,385	418,285	1,871,192
South Mainland	42,367	34,158	2011	78,536
West Mainland	76,738	80,726	8439	165,902
Total	1,868,966	543,123	726,921	3,139,009

Table 9. Area of different roof materials (in m^2) resulting from automatic image classification for the whole investigation area.

Area	Tile	Iron	Concrete	Total
Island	1,686,856	2,874,256	1,300,597	5,861,709
North Mainland	2,115,315	6,550,999	1,934,685	10,600,999
South Mainland	66,500	3,644,231	116,438	3,827,169
West Mainland	1,759,221	5,044,778	944,178	7,748,177
Total	5,627,892	18,114,264	4,295,898	28,038,054

One common way to assess the accuracy of classification is to construct a confusion matrix [39]. The matrix indicates how the classifier confuses between the different classes. Three types of accuracies can be derived from the confusion matrix: overall accuracy (OA), user's accuracy (correctness, UA) and producer's accuracy (completeness, PA). OA is the percentage of pixels assigned to the correct class. For each class, UA gives the percentage of pixels assigned to that class that also belong to that class in the reference, whereas PA gives the percentage of pixels of that class in the reference that were also assigned to that class by the classification procedure.

Table 10 shows the confusion matrix generated from the image classification process. As suspected, the classifier has a significant problem in differentiating concrete from the background, for example, 3.6% of the background pixels are incorrectly labelled as concrete. Error analysis showed that the OA of the image classification was 88.6%. However, this high level of OA could be misleading because the background class constituted a larger percentage of the total area compared to the other three classes. It was therefore necessary to check the producer's accuracy and user's accuracy for each class. The PA were 92.9%, 58.5%, 37.4%, and 54.3% for the background, tile, iron, and concrete, respectively. The UA on the other hand were 95.7%, 59.1%, 51.3%, and 22.2% for the background, tile, iron, and concrete, respectively. The high OA was due to the high values of PA and OA obtained for the background. When computing the confusion matrix for the differentiation of the background and roof (all materials), the PA is 93.0% for the background and 63.8% for the roof, and the UA is 95.6% for the background and 51.5% for the roof. Moreover, this reveals that the classifier had difficulties in differentiating some classes, especially tiles or concrete, from the background. This is largely because concrete roofs and roads have similar spectral properties. An idea to overcome this problem could be to use a Digital Surface Model, which can be derived from stereo images [40] by techniques such as semi-global matching [41]. Despite some further potential to improve the image classification, the OA of 88.6% indicates that the automatic classification was acceptable. Therefore, the results were used as input for the WEAP model.

Table 10. Confusion matrix based on selected pixels in the classified image.

		Class in Results (Automatic Classification) (All values in %)					
		Background	**Tile**	**Iron**	**Concrete**	**Sum**	**Completeness**
Class in Reference (Digitized)	Background	83.3	1.7	1.0	3.6	89.6	92.9
	Tile	1.8	2.8	0.1	0.2	4.8	58.5
	Iron	1.0	0.2	1.2	0.8	3.2	37.4
	Concrete	1.0	0.1	0.0	1.3	2.4	54.3
	Sum	87.0	4.8	2.3	5.9	-	-
	Correctness	95.7	59.1	51.3	22.2	-	88.6

3.2. WEAP Scenarios

The projected water demand up to the year 2035 for Mombasa City under the different scenarios, namely the reference scenario, better living standards (BLS), high population growth (HPG), low population growth (LPG), and efficient water use (EWU), are shown in Figure 7.

The simulated demand coverage in 2035 for Mombasa City with no interventions decreases from 54% in 2014 to 28% for the reference scenario. The demand coverage is lower under the HPG with BLS scenario (17%), BLS (21%), and HPG (23%) scenarios, but slightly improves in the EWU (35%) and LPG (37%) scenarios. As water demand coverage for Mombasa City continues to dwindle, it was necessary to investigate the impact of RRWH if practiced under different implementation scenarios. Figure 8 illustrates the amount of rainwater that can potentially be harvested under the various rainwater harvesting scenarios.

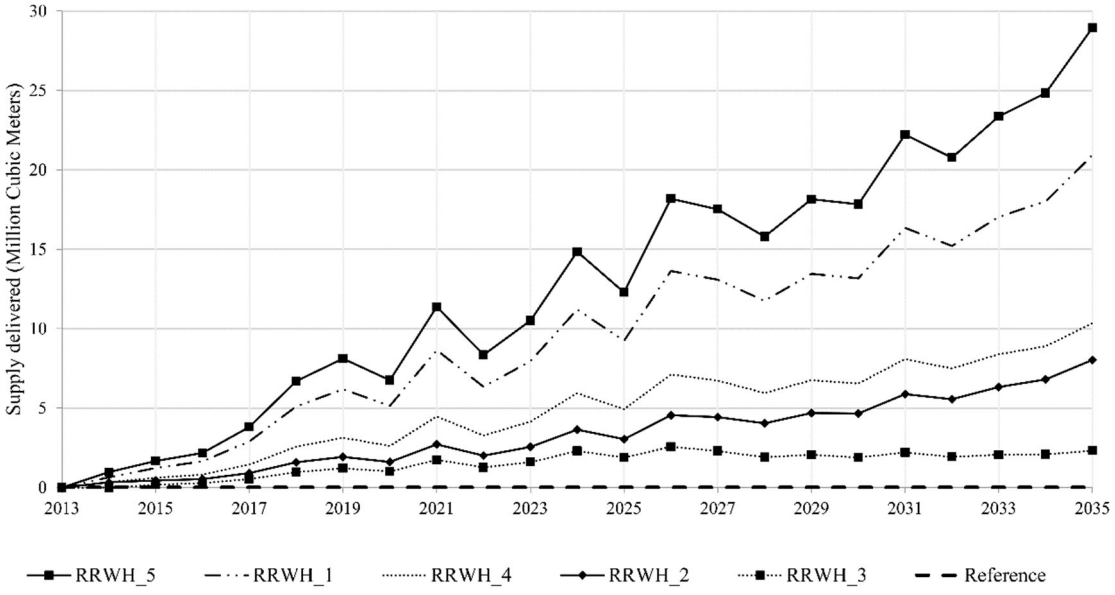


Figure 7. Annual Water demand projections in Million Cubic Metres (MCM) for different scenarios. Abbreviations: better living standards (BLS), high population growth (HPG), low population growth (LPG), efficient water use (EWU).

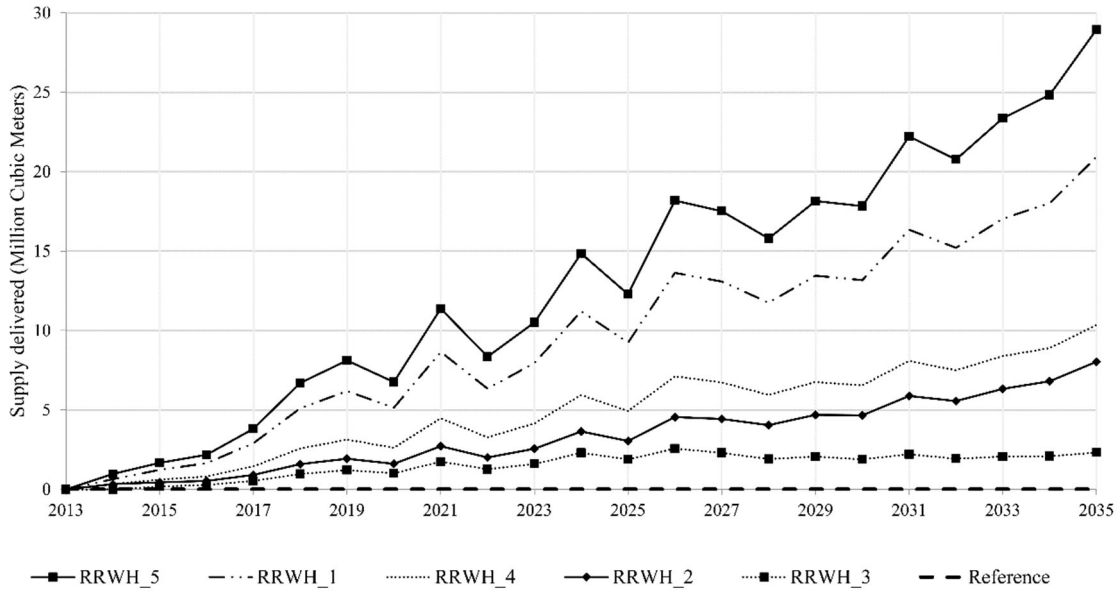


Figure 8. Additional water supply delivered under different RRWH scenarios for Mombasa City.

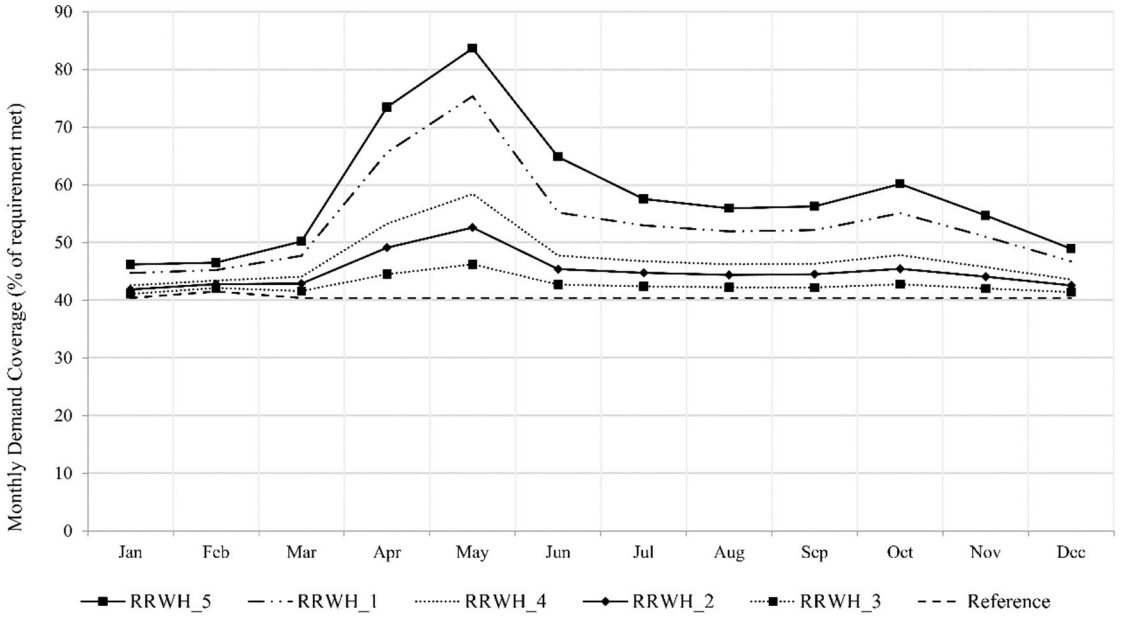


Figure 9. Average monthly demand coverage of RRWH combined with the existing system (2014–2035).

The results show that the potential of RRWH varies greatly between the different management scenarios. The inter-annual variability within all the scenarios is due to variations in precipitation over the years. The highest amount of rainwater can be provided under the scenario RRWH_5 (supply of over 28 MCM by 2035). The other scenarios RRWH_1 to RRWH_4 yield between 2.3 MCM and 20 MCM by 2035. These results suggest that the potential of RRWH in the city greatly depends on the strategy adopted by the city water management.

The average monthly demand coverage using the existing supply and RRWH as the only additional strategy is shown in Figure 9. For the reference scenario where no intervention was made to the existing system, the average monthly demand coverage for the whole 2014–2035 period is around 40%. The results clearly show that implementing RRWH improves the water supply situation. Based on the bimodal rainfall pattern in Mombasa City, higher demand coverage is achieved within the rainy months every year. In the month of May, which records the highest rainfall, the average monthly demand coverage increases to 46%, 53%, 58%, 75%, and 84% for RRWH_3, RRWH_2, RRWH_4, RRWH_1, and RRWH_5, respectively.

Apart from the RRWH strategies, the other possible management scenarios considered were the development and expansion of new water sources (NWS), efficient water use (EWU), and reduction of non-revenue water (NRWS). In terms of the water supply delivered for the whole period from 2013 to 2035, the responses of the system are presented in Table 11. The NWS scenario, which involves the expansion of existing water sources and the development of new ones, provides the best strategy, which increases the supply by 470 MCM throughout the whole study timeframe. The other strategies result in increases of the supply ranging between 34 MCM to 295 MCM over the same period. The UWE scenario does not increase the supply but reduces the demand, hence its effect can only be seen under demand coverage or unmet demand. Figure 10 indicates that no single strategy will completely solve the water scarcity in the city. Thus, a combination of different strategies is recommended.

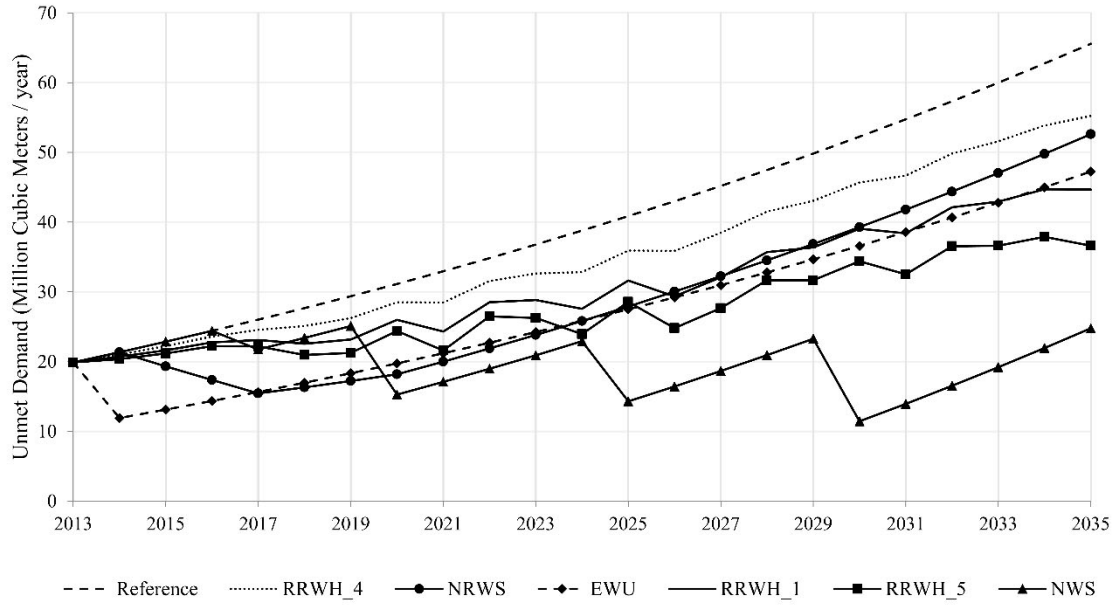


Figure 10. Unmet demand for Mombasa City for different management strategies (see Table 11 for acronym definitions).

For the combined scenarios, only the most feasible RRWH_4 is used, where only selected existing buildings and all new buildings implement RRWH. The results (Figure 11) show that before the year 2017, none of the combined strategies will meet the demand. Subsequently, only two scenario combinations, namely NWS/NRWS/EWU and NWS/RRWH_4/NRWS/EWU

cover the demand fully from 2017 to the end of the simulation period. Consequently, meeting the demand before 2017 is not achievable by any of the investigated management strategies.

Table 11. Total supply delivered under different management scenarios (2013–2035).

Supply/Demand Strategy	Total Supplied (MCM)	Supply Increase (MCM)
New water sources (NWS)	1055	470
All existing and new buildings (RRWH_5)	880	295
Non-Revenue Water Reduction (NRWS)	837	252
All existing buildings (RRWH_1)	804	219
Selected existing and all new buildings (RRWH_4)	696	111
Only new buildings (RRWH_2)	661	76
Selected existing buildings (RRWH_3)	619	34
Efficient Water Use (EWU)	585	0
Reference	585	0

The availability of water resources greatly depends on climatic conditions and the success of RRWH depends largely on the available precipitation. In this study, the reference scenario and other scenarios were based on the RCP 4.5 stabilization scenario with RCP 8.5 being used to understand how the system responds to a different climate change forcing scenario. The results show that the effect of predicted climate change considering RRWH_4 and RRWH_5 is not very appreciable (Figure 12). Considering RRWH_4, the average annual amount of rainwater harvested between 2014 and 2035 for RCP 4.5 and RCP 8.5 are 4.8 MCM and 4.7 MCM, respectively. The reduction of 2.5% is small compared to the uncertainty of the different GCM predictions. The two-sample t-test gives a p-value of 0.93, indicating no significant difference in the means at the 5% significance level. For RRWH_5, the average annual amount of rainwater delivered is 12.7 MCM for RCP 4.5 and 12.4 MCM for RCP 8.5, with a p-value of 0.94. Based on the results it can be concluded that the predicted effect of climate change on RRWH is negligible for Mombasa City.

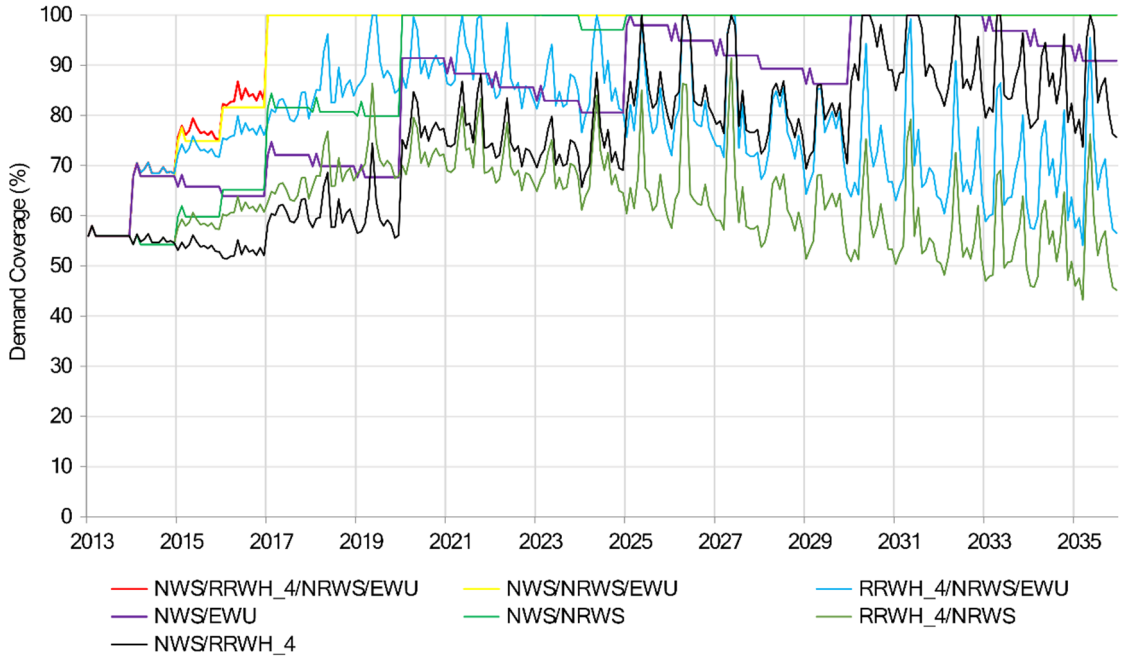


Figure 11. Demand coverage under different scenario combinations for Mombasa City.

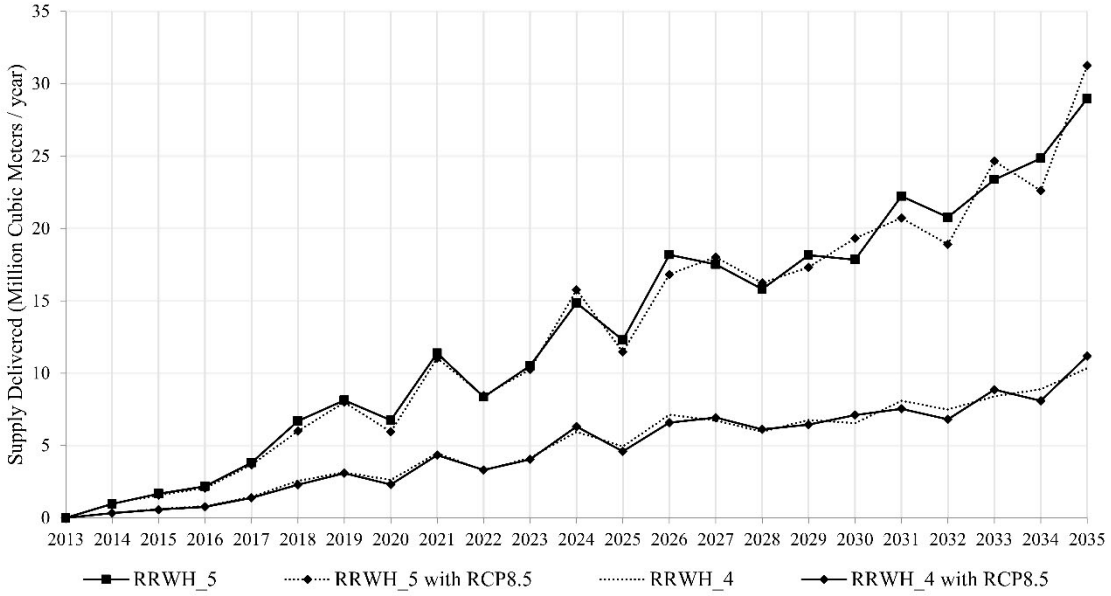


Figure 12. Effect of climate change on the RRWH_4 and RRWH_5 scenarios.

4 Conclusions

Overall, the results from this study show that the water demand for Mombasa City is expected to rise within 2014–2035 mainly due to socio-economic factors. Five possible RRWH implementation scenarios were established and investigated using a WEAP model. Using RCP 4.5 future climate data, the results showed that the average demand coverage improves significantly, mostly during the rainy season.

Comparing all of the management strategies, the development of new water sources (NWS) would lead to the highest demand coverage over the planning period up to 2035. None of the proposed strategies implemented in isolation will lead to full demand coverage. This means that NWS identified under the water supply master plan (2013) can only meet the demand if implemented alongside other strategies.

By combining the various strategies (only RRWH_4 used in the combinations) under a normal population growth rate, it was found that no combination of methods can cover the full demand for the entire 2014–2035 period. Under the highest water demand scenario, i.e., high population growth with better living standards, the city's water demand can never be met even under the most favorable climatic conditions. Any RRWH strategy is able to remedy but not completely solve the problem. The challenge to water managers is to make use of a combination of water supply sources, and to develop other strategies beyond those considered here e.g., river abstractions, desalination, or water reuse.

The results show that climate change will most likely not have an impact on the quantity of water delivered by RRWH systems in Mombasa. This means that RRWH can serve as a robust strategy against climate change effects.

The combination of image classification and water resource modelling proved to be a suitable tool for the development of roof rainwater harvesting strategies under changing

water availability and demand. The efficiency of automatic image classification can be further improved by including height information obtained from stereo images.

This study mostly addressed the first part of RRWH supply systems at a larger scale and from a more technical perspective. Conveyance and storage systems have not been investigated in detail, and for the practical implementation of RRWH, socio-economic aspects and water quality issues should be considered as well. Additional studies are recommended on building conditions, roof rainwater quality, tank optimization/design, costs and fundraising, awareness creation, and sensitization of the city residents regarding RRWH. Another interesting area that could be investigated in the future is hydrological aspects, such as the benefit of RRWH for flood risk reduction and the effect of RRWH on groundwater recharge.

Acknowledgments

The work of Robert Ojwang was funded by Deutscher Akademischer Austauschdienst (DAAD) under the Entwicklungsbezogene Postgraduiertenstudiengänge (EPOS) program. The Digital Globe Foundation granted images. The Kenya Meteorological Department, Coast Water Services Board, and Mombasa Water and Sanitation Services Company provided data. We acknowledge the support of the aforementioned organizations. We thank the four anonymous reviewers for their comments, which helped to improve the manuscript. The publication of this article was funded by the Open Access fund of Leibniz Universität Hannover.

Author Contributions

Jörg Dietrich and Franz Rottensteiner developed the idea for the study. Matthias Beyer and Jörg Dietrich designed the model experiments; Robert Ojwang performed the model experiments; Robert Ojwang and Prajna Kasargodu Anebagilu analyzed the data; Robert Ojwang, Jörg Dietrich, Prajna Kasargodu Anebagilu, Matthias Beyer, and Franz Rottensteiner wrote the paper.

Conflicts of Interest

The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Abdulla, F.; Al-Shareef, A. Roof rainwater harvesting systems for household water supply in Jordan. Desalination 2009, 243, 195–207.
2. Siegert, K. Introduction to Water Harvesting: Some Basic Principles for Planning, Design and Monitoring; Water Reports FAO: Rome, Italy, 1994.
3. Angrill, S.; Segura-Castillo, L.; Petit-Boix, A.; Rieradevall, J.; Gabarrell, X.; Josa, A. Environmental performance of rainwater harvesting strategies in Mediterranean buildings. Int. J. Life Cycle Assess. 2016.
4. Gould, J.; Nissen-Petersen, E. Rainwater Catchment Systems for Domestic Supply: Design, Construction and Implementation; Intermediate Technology Publications: London, UK, 1999.

5. Domènech, L.; Saurí, D. A comparative appraisal of the use of rainwater harvesting in single and multi-family buildings of the Metropolitan Area of Barcelona (Spain): Social experience, drinking water savings and economic costs. J. Clean. Prod. 2011, 19, 598–608.
6. Handia, L.; Tembo, J.M.; Mwiindwa, C. Potential of rainwater harvesting in urban Zambia. Phys. Chem. Earth 2003, 28, 893–896.
7. Thomas, T.H.; Martinson, D.B. Roof Water Harvesting: A Handbook for Practitioners; IRC International Water and Sanitation Centre: Delft, The Netherlands, 2007.
8. Kahinda, J.M.; Taigbenu, A.E.; Boroto, R.J. Domestic rainwater harvesting as an adaptation measure to climate change in South Africa. Phys. Chem. Earth 2010, 35, 742–751.
9. Taffere, G.R.; Beyene, A.; Vuai, S.A.H.; Gasana, J.; Seleshi, Y. Reliability analysis of roof rainwater harvesting systems in a semi-arid region of sub-Saharan Africa: Case study of Mekelle, Ethiopia. Hydrol. Sci. J. 2016, 61, 1135–1140.
10. Lee, J.Y.; Bak, G.; Han, M. Quality of roof-harvested rainwater—Comparison of different roofing materials. Environ. Pollut. 2012, 162, 422–429.
11. Farreny, R.; Morales-Pinzo, T.; Guisasola, A.; Taya, C.; Rieradevall, J.; Gabarrell, X. Roof selection for rainwater harvesting: Quantity and quality assessments in Spain. Water Res. 2011, 45, 3245–3254.
12. Gikas, G.D.; Tsihrintzis, V.A. Assessment of water quality of first-flush roof runoff and harvested rainwater. J. Hydrol. 2012, 466–467, 115–126.
13. UNEP; CEHI. A Handbook on Rainwater Harvesting in the Caribbean; The United Nations Environment Programme (UNEP): Washington, DC, USA; The Caribbean Environmental Health Institute (CEHI): Castries, Saint Lucia, 2009.
14. Ward, S.; Memon, F.A.; Butler, D. Performance of a large building rainwater harvesting system. Water Res. 2012, 46, 5127–5134.
15. Mehrabadi, M.H.R.; Saghafian, B.; Haghighi Fashi, F. Assessment of residential rainwater harvesting efficiency for meeting non-potable water demands in three climate conditions. Resour. Conserv. Recycl. 2013, 73, 86–93.
16. Meera, V.; Mansoor Ahammed, M. Water quality of rooftop rainwater harvesting systems: A review. J. Water Supply Res. Technol. AQUA 2006, 55, 257–268.
17. Sazakli, E.; Alexopoulos, A.; Leotsinidis, M. Rainwater harvesting, quality assessment and utilization in Kefalonia Island, Greece. Water Res. 2007, 41, 2039–2047.
18. Mendez, C.B.; Klenzendorf, J.B.; Afshar, B.R.; Simmons, M.T.; Barrett, M.E.; Kinney, K.A.; Kirisits, M.J. The effect of roofing material on the quality of harvested rainwater. Water Res. 2011, 45, 2049–2059.
19. Byrne, J.; Taminiau, J.; Kurdgelashvili, L.; Kim, K.N. A review of the solar city concept and methods to assess rooftop solar electric potential, with an illustrative application to the city of Seoul. Renew. Sustain. Energy Rev. 2015, 41, 830–844.
20. Williams, N.; Quincey, D.; Stillwell, J. Automatic classification of roof objects from aerial imagery of informal settlements in Johannesburg. Appl. Spat. Anal. Policy 2015, 9, 269–281.
21. Veljanovski, T.U.; Kanjir, U.; Pehani, P.; Oštir, K.; Kovačič, P. Object-based image analysis of VHR satellite imagery for population estimation in informal settlement Kibera-Nairobi, Kenya. In Remote Sensing-Applications; InTech Europe: Rijeka, Croatia, 2012; pp. 407–434.
22. Yates, D.; Sieber, J.; Purkey, D.; Huber-Lee, A. WEAP21—A demand-priority and preference-driven water planning model Part 1: Model characteristics. Water Int. 2005, 30, 487–500.
23. Yates, D.; Purkey, D.; Sieber, J.; Huber-Lee, A.; Galbraith, H. WEAP21—A demand-priority and preference-driven water planning model. Part 2: Aiding freshwater ecosystem service evaluation. Water Int. 2005, 30, 501–512.
24. Lévite, H.; Sally, H.; Cour, J. Testing water demand management scenarios in a water-stressed basin in South Africa: Application of the WEAP model. Phys. Chem. Earth 2003, 28, 779–786.
25. Mutiga, J.K.; Mavengano, S.T.; Zhongbo, S.; Woldai, T.; Becht, R. Water allocation as a planning tool to minimize water use conflicts in the Upper Ewaso Ng'iro North Basin, Kenya. Water Resour. Manag. 2010, 24, 3939–3959.

26. Falkenmark, M.; Lundquist, J.; Widstrand, C. Macro-scale Water Scarcity Requires Micro-scale Approaches: Aspects of Vulnerability in Semi-arid Development. Nat. Resour. Forum 1989, 13, 258–267.
27. WRMA. Integrated Water Resources Management and Water Efficiency Plan for Kenya; Water Resources Management Authority (WRMA): Nairobi, Kenya, 2009.
28. Droogers, P.; Butterfield, R.; Dyszynski, J. Climate Change and Hydropower, Impact and Adaptation Costs: Case Study Kenya; Future Water: Wageningen, The Netherlands, 2009.
29. Taylor, K.E.; Stouffer, R.J.; Meehl, G.A. An overview of CMIP5 and the experiment design. Bull. Am. Meteorol. Soc. 2012, 93, 485–498.
30. Overland, J.E.; Wang, M.; Bond, N.A.; Walsh, J.E.; Kattsov, V.M.; Chapman, W.L. Considerations in the selection of global climate models for regional climate projections: The Arctic as a case study. J. Clim. 2011, 24, 1583–1597.
31. MWI. Practice Manual for Water Supply Services in Kenya; Ministry of Water and Irrigation (MWI): Nairobi, Kenya, 2005.
32. Mombasa County. Mombasa County Government First County Integrated Development Plan (2013–2017); Government of Kenya: Mombasa, Kenya, 2014.
33. Kenya National Bureau of Statistics. Kenya 2009 Population and Housing Census; Government of Kenya: Mombasa, Kenya, 2009.
34. TAHAL/Bhundia Consultants. Water Supply Master Plan for Mombasa and Other Towns within Coast Province; Coast Water Services Board: Mombasa, Kenya, 2013.
35. WASREB. A Performance Review of Kenya's Water Services Sector 2012–2013 (Impact Report Issue No. 7); Water Services Regulatory Board (WASREB): Nairobi, Kenya, 2014.
36. Lillesand, T.M.; Kiefer, R.W. Remote Sensing and Image Interpretation; John Wiley & Sons: New York, NY, USA, 1994.
37. CES Consultants. Feasibility Study, Preliminary and Detailed Engineering Designs of Development of Mwache Multi-Purpose Dam Project along Mwache River—Hydrology Report; Ministry of Regional Development: Nairobi, Kenya, 2013.
38. TWDB. The Texas Manual on Rainwater Harvesting; Texas Water Development Board (TWDB): Austin, TX, USA, 2005.
39. Stehman, S.V. Selecting and interpreting measures of thematic classification accuracy. Remote Sens. Environ. 1997, 62, 77–89.
40. Grote, A.; Rottensteiner, F. Assessing the impact of digital surface models on road extraction in suburban areas by region-based road subgraph extraction. Int. Arch. Photogramm. Remote Sens. and Spat. Inf. Sci. 2009, 38, 27–34.
41. Hirschmüller, H. Stereo processing by semiglobal matching and mutual information. IEEE Trans. Pattern Anal. Mach. Intell. 2008, 30, 328–341.

This chapter is an edited version of the following original scientific article:

Ojwang, R.; Dietrich, J.; Kasargodu Anebagilu, P.; Beyer, M.; Rottensteiner, F. (2017): Rooftop Rainwater Harvesting for Mombasa: Scenario Development with Image Classification and Water Resources Simulation. Water 9(5), 359.

Publisher: MDPI

IX. Modification of the SWAT Model to Simulate Regional Groundwater Flow Using A Multi-Cell Aquifer

Van Tam Nguyen[1], Jörg Dietrich[1]

[1] Institute of Hydrology and Water Resources Management, Leibniz University Hannover, Germany

Abstract

The Soil and Water Assessment Tool (SWAT) has been widely used and thoroughly tested in many places in the world. The application of the SWAT model has pointed out that two of the major weaknesses of SWAT are related to the non-spatial reference of the Hydrologic Response Unit (HRU) concept and to the simplified groundwater concept, which contribute to its low performance in baseflow simulation and its inability to simulate regional groundwater flow. This study modified the groundwater module of SWAT to overcome the above limitations. The modified groundwater module has two aquifers. The local aquifer, which is the shallow aquifer in the original SWAT, represents a local groundwater flow system. The regional aquifer, which replaces the deep aquifer of the original SWAT, represents intermediate and regional groundwater flow systems. Groundwater recharge is partitioned into local and regional aquifer recharges. The regional aquifer is represented by a Multi-Cell Aquifer (MCA) model. The regional aquifer is discretized into cells using the Thiessen polygon method, where centers of the cells are locations of groundwater observation wells. Groundwater flow between cells is modeled using Darcy's Law. Return flow from cell to stream is conceptualized using a non-linear storage-discharge relationship. The SWAT model with the modified aquifer module, the so-called SWAT-MCA, was tested in two basins (Wipperau and Neetze) with porous aquifers in a lowland area in Lower Saxony, Germany. Results from the Wipperau basin show that the SWAT-MCA model is able (1) to simulate baseflow in a lowland area (where baseflow is a dominant source of streamflow) better than the original model, and (2) to simulate regional groundwater flow, shown by the simulated groundwater levels in cells, quite well.

1 Introduction

Hydrological models are often used to gain understanding of the surface and subsurface hydrologic processes within the interested region. Furthermore, they can be applied as systems analytic tools in water resources management. Hydrological models can be classified into lumped, semi-distributed, and fully distributed models (Jajarmizadeh et al., 2012; Pechlivanidis et al., 2011). Lumped models consider a basin as a single unit. Fully distributed models, especially the ones with physically based approach, discretize the basin into grid cells and simulate surface and subsurface flow between these grid cells. However, the use of fully distributed models at the basin scale is often restricted due to extensive data requirements and high computational cost. Semi-distributed models appear as compromise models between lumped models and fully distributed models even though semi-distributed models often neglect groundwater flow between subbasin units, for example, HBV (Bergström, 1992), SLURP (Kite, 1997), TOPMODEL (Beven and Kirkby, 1979).

The Soil and Water Assessment Tool (SWAT) is a semi-distributed hydrological model used to assess the impact of land management practices on water, sediment and agricultural chemical yields (Neitsch et al., 2011). The SWAT model (a) operates at a basin scale, (b) is continuous in time, (c) is computationally efficient, (d) uses readily available inputs and (e) is a free and open-source software. SWAT can simulate most aspects of the hydrologic cycle in agricultural catchments and has been widely used and thoroughly tested worldwide (Arnold and Fohrer, 2005; Gassman et al., 2007; Krysanova and White, 2007; Francesconi et al., 2016; Strauch et al., 2013; Wagner et al., 2016; Bieger et al., 2017). While there exist numerous semi-distributed models, few of them can simulate groundwater flow between subbasin units (Rozos et al., 2004; Efstratiadis et al., 2008; Joodavi et al., 2016; Lindström et al., 2010). However, it is hardly to find a model, which has all of the aforementioned advantages like the SWAT model in agricultural water management at catchment scale. Therefore, the SWAT model was selected for further modification in this study in order to improve its horizontal subsurface hydrological connectivity.

Applications of the SWAT model showed that one of the major weaknesses of SWAT is related to the non-spatial reference of the Hydrologic Response Unit (HRU) concept (Arnold and Fohrer, 2005; Gassman et al., 2007; Bosch et al., 2010; Bieger et al., 2017). HRUs are created by lumping all areas having the same combination of land use, soil type and slope within a subbasin (Leavesley et al., 1983). The HRU concept is computationally efficient while representing the above mentioned landscape heterogeneity. However, the HRU has no explicit spatial information, which is required to simulate flows between HRUs.

Simulation of groundwater flows between HRUs, subbasins, or different landscape units (LUs) is considered as an important part of water resources management (Refsgaard et al., 2010). In many places, groundwater is considered as one of the main sources for domestic uses and/or irrigation (e.g., Hutson et al., 2004; Siebert et al., 2010; Wittenberg et al., 2003; Wittenberg, 2015). Groundwater flow can alter the overall water budget or affect the subsurface and surface water quality of a region.

In recent years, some studies have been carried out to simulate groundwater flow between different LUs in the SWAT model. Volk et al. (2007) and Arnold et al. (2010) developed a SWAT landscape model by using a catena approach (Lane and Nearing, 1989; Kirkby, 1998) to delineate the basin into three different LUs, the so-called divide, hillslope, and floodplain (each LU can be further delineated into HRUs). Within this approach, groundwater is routed from the divide through the hillslope, the floodplain, and ultimately to the stream as flow through a series of linear storage elements (e.g., Brutsaert, 2005). Rathjens et al. (2015) developed a grid-based version of the SWAT landscape model. In this model, groundwater flow between grid cells is also modelled as flow through a series of linear storage elements. The model performance in streamflow simulation significantly depends on the grid size and the routing algorithm implemented. In addition, this model requires significant computation time (Pignotti et al., 2017). Sun et al. (2016) further developed the SWAT landscape model by delineating the floodplain into three smaller LUs corresponding to flooded areas with different flood return periods. Groundwater flow between LUs within the floodplain and the groundwater-surface water interaction were modeled using Darcy's Law. The model is capable of simulating groundwater levels in the floodplain LUs. Both the models (Volk et al., 2007 and Sun et al., 2016), however, only considered groundwater flow driven by topographic gradient (from the divide through the hillslope, the floodplain, and ultimately to the stream) and neglected groundwater flow at larger scales (at inter-subbasin and

inter-basin scales). Groundwater flow at larger scales might not follow local topographic lows and highs (Tóth, 1963). In addition, in low-land areas, it is difficult to differentiate (a) between the divide, the hillslope, and the flood plain, and (b) between groundwater and soil water.

The SWAT model also has a poor performance in baseflow simulation, which is normally associated with the return flow from groundwater (Eckhardt et al., 2002; Kalin and Hantush, 2006; Bosch et al., 2010; Luo et al., 2012; Lv et al., 2014; Guse et al., 2014; Pfannerstill et al., 2014a). This is not only due to the unaccounted effect of the flows between HRUs but also due to the simplified representation of the aquifer system and the return flow from groundwater. SWAT uses a two-layer aquifer model, shallow or unconfined aquifer (typically 2 - 20 m) and deep or confined aquifer (> 20 m), to represent the aquifer system in each HRU (Neitsch et al., 2011; Luzio et al., 2004). The deep aquifer in SWAT is inactive, which means that water entering the deep aquifer is considered a loss to the hydrological system (Neitsch et al., 2011). The contribution of groundwater from the shallow aquifer to streamflow is modeled using a linear aquifer storage-discharge relation approach.

Modifications of the SWAT model regarding the aquifer structure and the interaction between aquifer and stream have been made to improve baseflow simulation. For example, Luo et al. (2012) allowed groundwater from the deep aquifer to contribute to streamflow in a similar manner as of the shallow aquifer. Gan and Luo (2013) added another aquifer layer and used a nonlinear aquifer storage-discharge relation approach to simulate return flow. Pfannerstill et al. (2014a) divided the shallow aquifers into active fast and active slow shallow aquifers to account for fast and slow response components, respectively, of the return flow. Wang and Brubaker (2014) used a non-linear aquifer storage-discharge relation approach instead of a linear aquifer storage-discharge relation approach to simulate return flow. Results from the above mentioned studies showed that the modified SWAT models are able to simulate baseflow better than the original SWAT model. However, these studies have not considered lateral groundwater flow.

Another approach, which enables a reasonable representation of streamflow, groundwater flow, and groundwater-surface water interaction, integrates SWAT with the Modular Three-Dimensional Finite-Difference Groundwater Flow (MODFLOW) model (McDonald and Harbaugh, 1988). MODFLOW is a fully distributed physically-based groundwater model. There have been several versions of integrated SWAT and MODLFOW models, for example, SWATMOD (Sophocleous et al., 1999), SWAT-MODFLOW (Kim et al., 2008; Bailey et al., 2016), and SWATmf (Guzman et al., 2015). Results showed that these models are able to simulate streamflow, groundwater flow, and groundwater-surface water interactions at a basin scale (Sophocleous et al., 1999; Kim et al., 2008; Guzman et al., 2015) with satisfactory results. However, the integrated models are not always applicable because they require extensive computational time and input data. In addition, the aforementioned models are not easy to use for modelers, who are not familiar with the MODFLOW model.

There is also an ongoing development and testing of the next generation of the SWAT model called SWAT+ (Bieger et al., 2017). SWAT+ allows users to discretize the aquifer in a flexible way. Aquifers could be discretized into a number of aquifer units (also called spatial objects), which are no longer tied to HRUs. Although there is an option for flow routing between these aquifer units (as flow through a series of linear storage elements) with SWAT+, however, users have to manually define a connection between these aquifer units and also a connection between these

aquifer units with other spatial objects (e.g., HRU, channel, LUs, etc.). This could be problematic with many spatial objects in large basins. In addition, the groundwater height in these aquifer units has the same meaning with the groundwater height of the original SWAT model, which is not referenced to a physical datum.

The research objective of the present study is to develop a simple semi-distributed, physically-based groundwater module for SWAT. The developed model is expected to be capable of simulating streamflow (especially baseflow) and groundwater flow at a basin/regional scale with fast computational time, less input data requirement, and satisfactory results for porous aquifers. In order to achieve the above-mentioned objective, we applied a different discretization technique to delineate the aquifer while keeping the surface discretization as in the original SWAT model. The Multi-Cell Aquifer (MCA, Bear, 1979; Bear and Cheng, 2010) model was used as a replacement of the deep aquifer in the original SWAT model. The modified SWAT was applied in two lowland basins, where contribution of groundwater to streamflow is dominant.

2 Methodology

2.1 The original groundwater module of SWAT

As mentioned in the previous section, SWAT divides the aquifer system in each HRU into two types of aquifers, shallow and deep aquifers. Percolated water from the unsaturated zone is partitioned into the shallow and deep aquifer recharges via a parameter, which is typically calibrated. Water entering the deep aquifer is considered a loss to the system (Neitsch et al., 2011). This makes SWAT more flexible to calibrate, especially when there is a substantial amount of groundwater storage in the shallow aquifer, which results in an over-estimation of baseflow. Groundwater in the shallow aquifer can be 'revaporated' (w_{revap}) or contribute to streamflow as return flow (Q_{gw}) or be removed by pumping ($w_{pump,sh}$). The volumetric water balance equation for the shallow aquifer in the original SWAT model on each day (e.g., day i) is as follows (Neitsch et al., 2011):

$$aq_{sh,i} = aq_{sh,i-1} + w_{rchrg,sh} - Q_{gw} - w_{revap} - w_{pump,sh} \quad (1)$$

where $aq_{sh,i}$ and $aq_{sh,i-1}$ is the amount of water stored in the shallow aquifer on day i and on day $i-1$, respectively, $w_{rchrg,sh}$ is the groundwater recharge, other variables were as described previously. SWAT can simulate groundwater height in the shallow aquifer in each HRU (Neitsch et al., 2011). However, this variable provides only information about the groundwater storage content and has limited physical meaning because (1) it is not referenced to any physical datum, hence it does not represent the groundwater level (Vazquez-Amábile and Engel, 2005) and (2) it does not account for groundwater flow between adjacent HRUs. Moreover, this variable has a unique value for each HRU while a single HRU can be scattered over different areas having different altitudes. It should be noted that there is no groundwater flow between HRUs and return flow from each HRU is routed directly to the stream outlet of the respective subbasin. The groundwater module in SWAT was documented in more detail in Neitsch et al. (2011). Hereinafter, when the SWAT model is mentioned, we refer to the one documented in Neitsch et al. (2011) unless otherwise stated.

2.2 The modified groundwater module

In this section, a brief review of the natural groundwater flow systems and groundwater flow dynamics at the basin scale is first provided. Tóth (1963) suggested that groundwater flow can be classified into local, intermediate, and regional groundwater flows. The local groundwater flow originates from the recharge area (local topographic high) and ends in the discharge area (adjacent local topographic low). Therefore, local groundwater flow is strongly driven by the topographic gradient. The intermediate and regional groundwater flows have recharge and discharge areas separated by several topographic highs and lows. Tóth (1963) also pointed out that the aforementioned types of groundwater flow systems might not exist simultaneously in all types of aquifers and the boundaries between them might not be well defined.

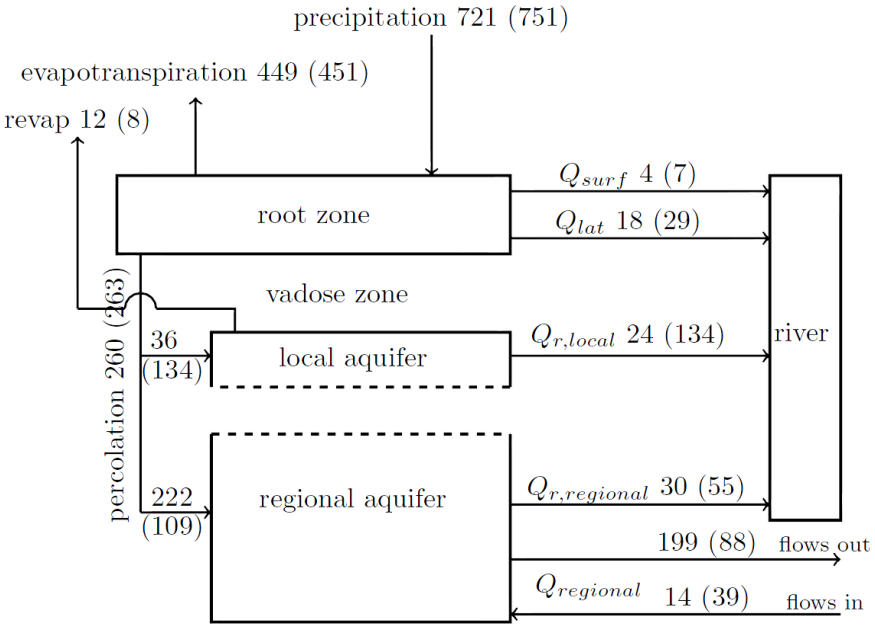


Figure 1. Schematic representation of the vertical structure of the modified groundwater module in SWAT. Q_{surf}, Q_{lat}, $Q_{r,\,local}$, $Q_{r,\,regional}$, and $Q_{regional}$ are the surface run-off, lateral flow, return flow from local aquifer, return flow from regional aquifer, and regional groundwater flow, respectively. The numbers indicate the average annual water balance components in millimeters per year in the Wipperau (numbers outside parentheses) and the Neetze (numbers inside parentheses) basins

To mimic the groundwater flow systems as described above in the modified SWAT model, we modified the concept of the two-aquifer layer in the original SWAT model (Figure 1). The local aquifer layer in this study represents local groundwater flow while the regional aquifer layer represents intermediate and regional groundwater flow. To be more precise, local groundwater flow in the modified SWAT model is restricted within a subbasin, which has a subbasin divide (topographic high) as recharge area and a stream of that subbasin (adjacent topographic low) as discharge area. In addition, the area between the subbasin divide and the stream is also considered as recharge area for the local groundwater flow system. Modelling local groundwater flow between local aquifers will not be the focus of our model. Therefore, we only model recharge, discharge and use a function to route local groundwater from recharge area to discharge area. Considering the above concept, the shallow aquifer in the original SWAT model will be considered as the local aquifer in our modified model. Therefore, discretization of the local aquifers is identical to that of HRUs and all simulations applied to the shallow aquifer in the original SWAT

will be applied to the local aquifers in our modified model. The main differences between the modified SWAT model presented here with the original SWAT model lie in the regional aquifer layer. We used the MCA model to replace the deep aquifer in the original SWAT model and to represent intermediate and regional groundwater flows. The resulting model hereafter is referred to as the SWAT-MCA model.

Delineation of the regional aquifer layer differs from that of the local aquifer because intermediate and regional groundwater flows are not restricted within a subbasin. The regional aquifer layer is discretized into cells (polygons) using Thiessen (1911) polygons constructed around the locations of the groundwater observation wells. Hence, one cell can be fully or partially overlaid by a single subbasin or several subbasins (Figure 2). It should be noted that the selected groundwater observation well must be representative for that cell and the screening depth must be within the unconfined aquifer. Groundwater flow between cells is modeled based on Darcy's Law. Considering the fact that there could be no clear boundary between local, intermediate and regional groundwater flows, and the local aquifer might only be filled after rainfall, flow in the regional aquifers in the SWAT-MCA model is assumed to be unconfined. This assumption is valid if the study area has discontinuous clay layers (lenses) or no distinct stratigraphic layers within the aquifer at the regional scale. Therefore, the volumetric water balance equation for each cell (e.g., cell i) has the following form:

$$\sum_{j=1}^{k} Q_{ij} + W = SY_i \frac{\Delta h_i}{\Delta t} A_i \tag{2}$$

where Q_{ij} [L^3T^{-1}] is the total inflow (+) and outflow (-), which is calculated by using Darcy's Law, k is the number of the neighboring cells of cell i, W [L^3T^{-1}] represents sources (+) and sinks (-), SY_i [-] is the specific yield of cell i, Δh_i [L] is the change in groundwater level at cell i, Δt [T] is the time step size, which is one day in this study, A_i [L^2] is the surface area of cell i. In finite-difference form with implicit time-stepping scheme, Equation (1) has the following form:

$$\sum_{j=1}^{k} K_{ij} a_{ij} \frac{h_j^t - h_i^t}{L_{ij}} + W^t = SY_i \frac{h_i^t - h_i^{t-1}}{\Delta t} A_i \tag{3}$$

where K_{ij} [LT-1] is the harmonic mean hydraulic conductivity between cells i and j, a_{ij} [L2] is the average saturated area between cells i and j, a_{ij} is a function of h_i^t [L] and h_j^t [L], which are the hydraulic heads of cells i and j, respectively, on the current day, L_{ij} [L] is the distance between centers of cells i and j, h_i^{t-1} [L] is the hydraulic head of cell i on the previous day.

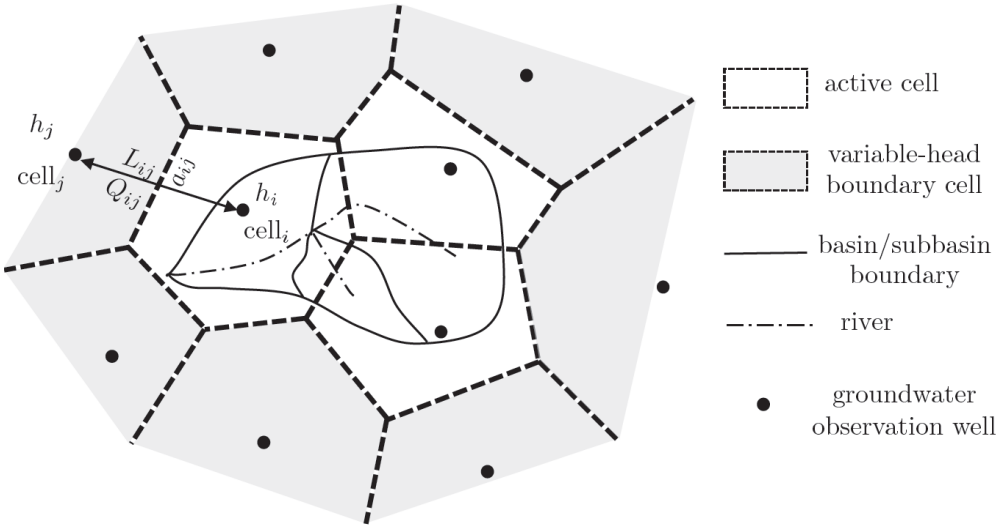


Figure 2. Schematic representation of the multicell aquifer model

Sources (+) and sinks (-) can be a portion of recharge from the unsaturated soil zone (+), return flow (-), and pumping for irrigation (-). We applied the HRU-cell conversion technique used by Kim et al. (2008) to assign a spatial linkage between HRUs and cells. The amount of recharge from an HRU to a cell is proportional to the overlapping area between that HRU and that cell. To calculate recharge from an HRU to a cell, the overlapping area matrix between HRUs and cells is needed (Table 1). The total amount of recharge to a cell, which is not overlaid completely by the basin, is assumed to be area proportional to the recharge from the overlapping area.

Table 1. The overlapping area matrix between HRUs and cells.

	hru_1	hru_2		hru_m
cell_1	$\text{cell}_1\_\text{hru}_1$	$\text{cell}_1\_\text{hru}_2$	-	$\text{cell}_1\_\text{hru}_m$
cell_2	$\text{cell}_2\_\text{hru}_1$	$\text{cell}_2\_\text{hru}_2$	-	$\text{cell}_2\_\text{hru}_m$
-	-	-	-	-
cell_n	$\text{cell}_n\_\text{hru}_1$	celln_hru_2	-	$\text{cell}_n\_\text{hru}_m$

Note. $\text{cell}_i\_\text{hru}_j$ (-) is the overlapping area between cell_i and HRU_j (in m^2)
n and m are the total number of cells and HRUs, respectively.

Return flow from cell to stream is conceptualized using the nonlinear storage-discharge relation suggested by Wang and Brubaker (2014). The total amount of return flow from each cell is calculated as follows:

$$Q_{r,regional} = \left(\alpha\left(S - S_{\min}\right)\right)^{1/\beta} \tag{4}$$

where $Q_{r,regional}$ [L^3T^{-1}] is the total amount of return flow, S [L^3] is the total groundwater storage, $S_{\min}$ [L^3] is the minimum groundwater storage for return flow to occur (if $S < S_{\min}$ then $Q_{r,regional} = 0$), α [$T^{\beta}L^{3(1-\beta)}$] and β [-] are coefficient parameters. Return flow from one cell can contribute to streamflow in several subbasins. Conversely, one subbasin can receive return flow from several cells. This concept represents the recharge and discharge area of intermediate and regional groundwater flows. The amount of return flow from a cell to a subbasin is proportional to the overlapping area between the respective cell and subbasin (Table 1).

2.3 Input Data Preparation

In addition to the input data as described by Arnold et al. (2013), input data for the SWAT-MCA model must be prepared in ASCII text files. They are named 'parameters.txt', 'hru.txt', and 'cell.txt'. The 'parameters.txt' file contains information regarding the cell's number, cell's geometry (coordinates of the cell's center, bottom and top elevations of the cell, and ID of the neighboring cells), cell's hydrogeological characteristics (saturated hydraulic conductivity and specific yield), cell's initial condition (initial groundwater level), and the parameters to control the amount of return flow (α, β, and $S_{\min}$). If the cell is used as a variable-head boundary cell, the name of the file, which contains observed daily groundwater levels of that cell during the simulation time, must be provided. Other types of boundary conditions could be also implemented with minor modification of the code. The 'hru.txt' and 'cell.txt' files are used to assign a spatial linkage between HRUs and cells and to calculate the overlapping area matrix between HRUs and cells (Table 1). The 'hru.txt' and 'cell.txt' files can be created using GIS tools.

2.4 Model Integration Framework

The MCA model was integrated into SWAT (SWAT2012 rev. 664) by using two main subroutines. The first subroutine, 'read_process', reads and processes input data for MCA. This subroutine is called only once (at the beginning of the simulation). The second subroutine, 'simulate_gw', simulates groundwater flow between cells, groundwater levels, and return flow from the regional aquifer (Equations (3) and (4)). This subroutine is called each timestep after the subroutine 'subbasin', which simulates all land phase processes for each HRU, and before the subroutines 'route', 'add', and other subroutines, which simulate the routing phase of the hydrologic cycle. After each simulation day, results are printed out. For more information regarding the subroutines 'subbasin', 'route', and 'add', one could refer to Arnold et al. (2013). Minor modifications in the SWAT code were also made to transfer variable values between subroutines of SWAT and MCA and to print out daily results, resulting in only one executable program (Figure 3). The source code of the SWAT-MCA model could be available upon request.

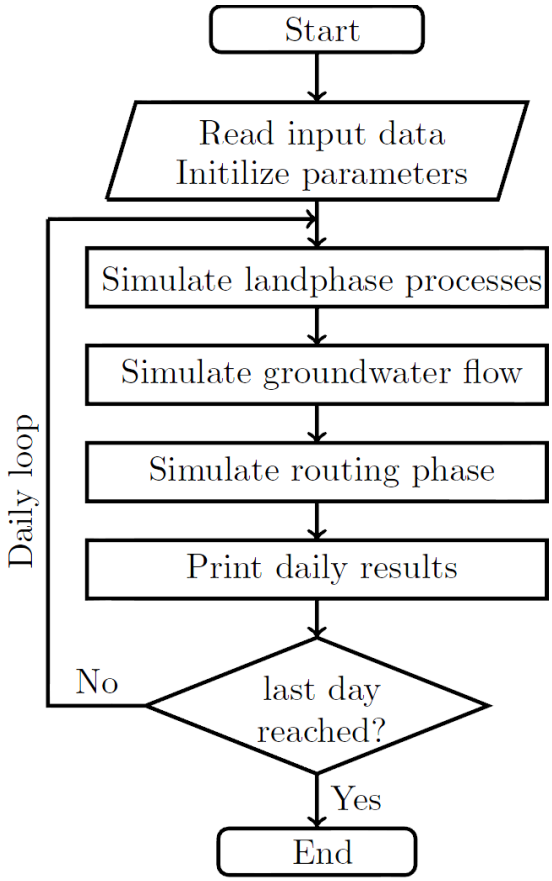


Figure 3. Integration framework of the soil and water assessment tool–multicell aquifer model

3 Case Study

3.1 Study Areas and Data

The study areas, the Wipperau and the Neetze basins, are located in a lowland area within the Lüneburger Heide region in Lower Saxony, Germany (Figure 4). The two basins are part of the Ilmenau basin, which drains to the Elbe River in the North-West. The Digital Elevation Model (DEM) of 10 m resolution was provided by the Lower Saxony Water Management, Coastal Defence and Nature Conservation Agency (NLWKN). The study areas are relatively flat with elevation ranges from 19 m to 139 m above mean sea level (a.m.s.l.) (Figure 4). Daily weather data (precipitation, wind speed, sunshine hours, relative humidity) from 1976 to 2007 were obtained from the German Weather Service (DWD). Weather data were interpolated for all subbasins using the inverse distance weighting method. Daily sunshine hours were converted to solar radiation using the formulae suggested by Ångström (1924). The average annual precipitation values of the two basins from interpolated data are of similar magnitude (Table 2). Daily streamflow records at the outlets of the Wipperau basin (Oetzmühle gauging station) and the Neetze basin (Süttorf gauging station) from 1976 to 2007 were obtained from NLWKN. Streamflow data show that the average annual water yields of the two basins are significantly different (Table 2). Baseflow analysis using the baseflow filter program (Arnold et al., 1995) indicates that (1) return flow from groundwater is a dominant source of streamflow in both basins, and (2) return flow from the Neetze basin is much higher than return flow from the Wipperau basin (Table 2).

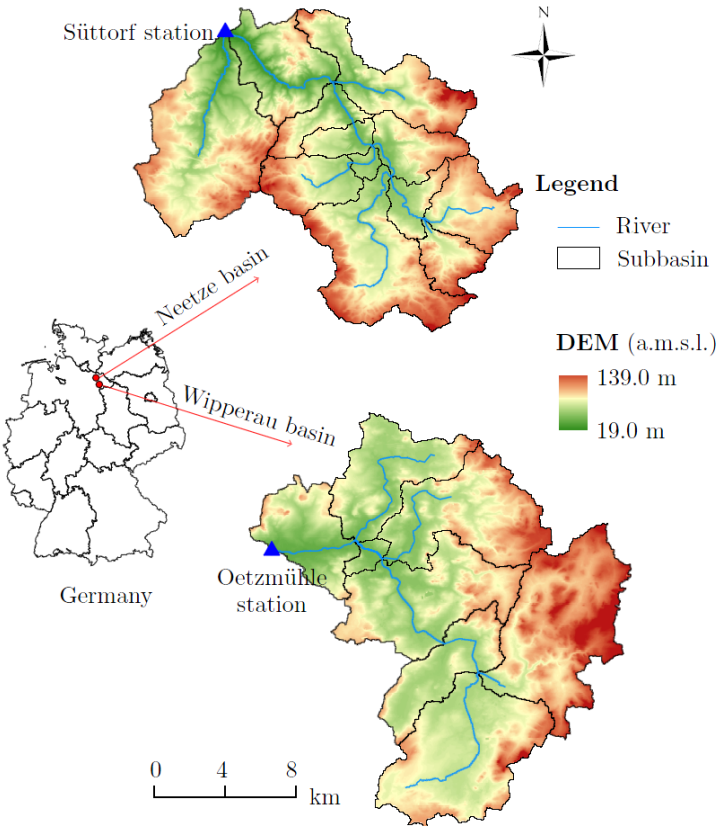


Figure 4. Location and digital elevation model (DEM) of the Neetze and Wipperau basins

Land use/Land cover map and soil map 1:200000 (BÜK200, https://geoviewer.bgr.de/mapapps/resources/apps/geoviewer/index.html?lang=de&tab=boden&layers=boden_buek200_ags) were taken from the CORINE Land Cover project and the Federal Institute for Geosciences and Natural Resources (BGR), respectively. The dominant land cover types in both basins are agricultural land and forest (Table 2). The majority of the agricultural land (about 90%) is irrigated with an average annual amount of groundwater extraction for irrigation of around 73 mm (Riediger et al., 2014; Wittenberg, 2015). The dominant soils in the Wipperau basin are (1) podzols and brown podzolic soils, 43.4% of the total area, and (2) brown earth and podzolic-brown earth soils, 32.4% of the total area. In the Neetze basin, the dominant soils are (1) podzols and brown podzolic soils, 47.2% of the total area, and (2) brown earth soil, 23.2% of the total area. Soil hydraulic properties were derived by using the pedotransfer functions/tables from Wessolek et al. (2009).

Table 2: Characteristics of the study areas.

	Wipperau basin	Neetze basin
Total area (km²)	200.4	172.8
Agricultural land (% total area)	65.6	56.0
Forest (% total area)	31.2	39.3
Average annual precipitation (mm)	721	751
Average annual water yield (mm)	66.3	173.4
Average annual baseflow (mm)	49.4	151.8
(% average annual water yield)	75	88

Average annual groundwater recharge recorded at the lysimeters at Hohenzethen, which are operated by the Landesamt für Bergbau, Energie und Geologie (LBEG), in the Wipperau catchment

from 2001 to 2015 varied between 328 mm and 347 mm. However, simulated annual groundwater recharge from Lemke and Elbracht (2008) was around 225 mm. These values along with the estimated return flow from groundwater (Table 2) indicate that in both basins, the majority of groundwater recharge does not become return flow.

Hydrogeological cross-sections of the area published by LBEG indicate that aquifers and aquitards are sandwiched and the aquitards are discontinuous at regional scale. Thus, the assumption that the regional aquifer is unconfined in the SWAT-MCA model is valid in this case. The dominant hydrogeological layers of the underlying aquifer are aquifers of class L3 to class L6 according to the classification given by Manhenke et al. (2001). This means that the hydraulic conductivity of the underlying aquifers varies between 0.86 m/day and 86.4 m/day.

Observed groundwater levels from 1980 to 2007 within and nearby the study areas (Figure 5) were obtained from NLWKN. Groundwater levels were observed at different time steps, from monthly to daily time steps. Linear-in-time interpolation was applied to get daily groundwater levels at the boundary cells.

3.2 Spatial Discretization

Figure 5 shows the spatial discretization of the surface and subsurface of the Wipperau and Neetze basins. The Wipperau basin was divided into 8 subbasins and 221 HRUs while the Neetze basin was discretized into 12 subbasins and 213 HRUs. Discretization of the subbasins into HRUs was not shown because HRUs are not easy to recognize visually. The local aquifers within the two basins were delineated identically with the delineation of their HRUs, resulting in 221 local aquifer units in the Wipperau basin and 213 local aquifer units in the Neetze basin. The regional aquifers of the Wipperau and the Neetze basins were delineated into 33 and 26 cells, respectively, using Thiessen polygon method with locations of the groundwater observation wells as cell's centre.

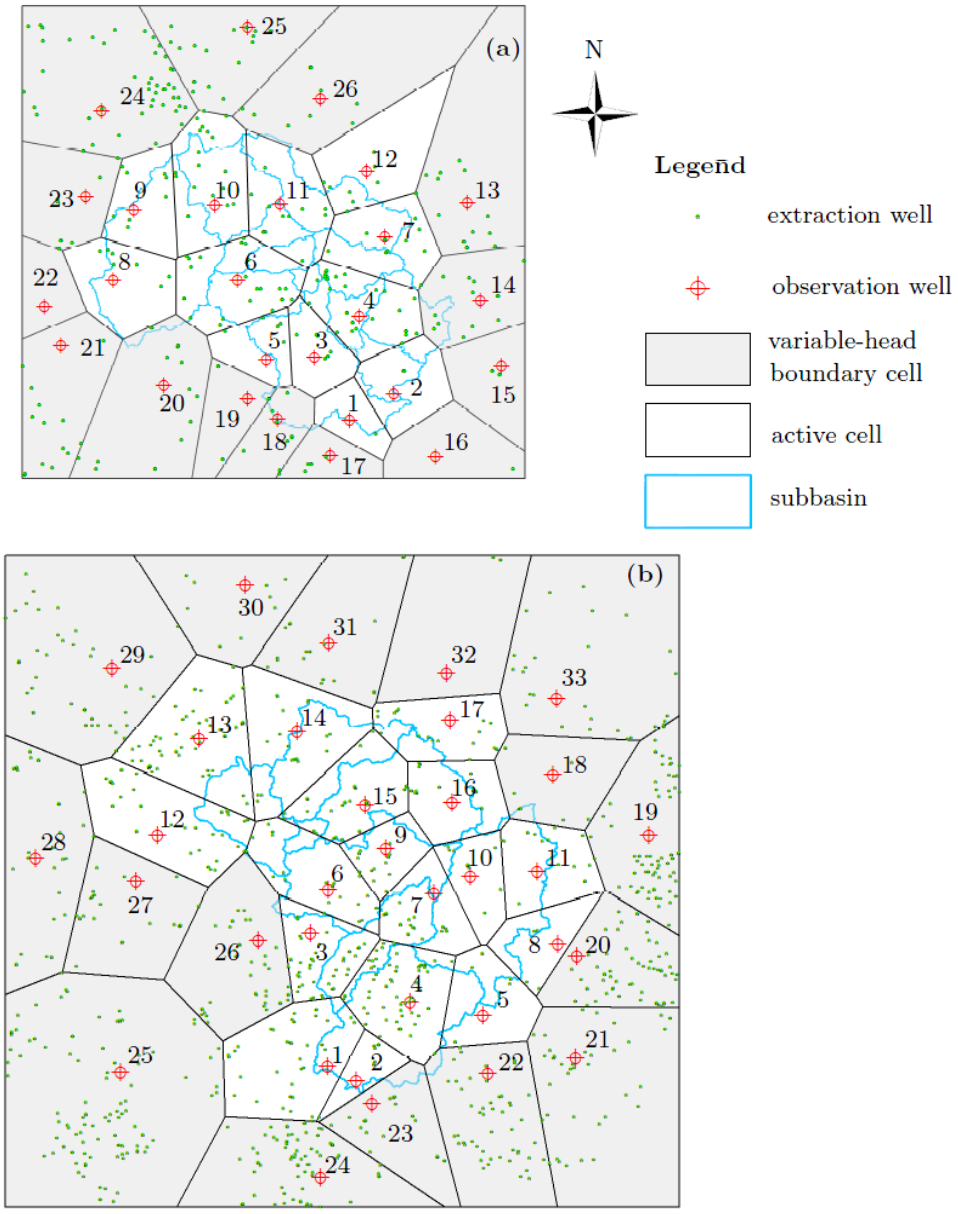


Figure 5. Delineation of the subbasin and the regional groundwater aquifer of (a) the Neetze basin and (b) the Wipperau basin

3.3 Calibration and validation strategy

The SWAT and SWAT-MCA models were run for 32 years (1976 - 2007) with 4 years for warm-up (1976 - 1979), 14 years for calibration (1980 - 1993), and 14 years for validation (1994 - 2007). The regional groundwater module was activated after the warm-up period with initial groundwater levels obtained from observed groundwater levels. By doing that, the groundwater recharge is expected to be reasonably estimated when the regional groundwater module is activated. Similar with Guzman et al. (2015), we first calibrated and validated the original SWAT model against the observed streamflow to ensure that the surface processes are adequately represented and the amount of groundwater recharge is reasonably estimated. Parameters used for calibrating the original SWAT were selected based on the dominant conditions in the basin, sensitivity analysis, and literature review of the most commonly used parameters (Arnold et al., 2013). The selected parameters are shown in Table 3. Assuming the aquifer recharge is

satisfactorily estimated by the original SWAT, the best calibrated parameter values from the original SWAT model will be kept unchanged during the calibration of the SWAT-MCA model, except the RCHRG_DP, GW_DELAY, and ALPHA_BF. In the SWAT-MCA model, RCHRG_DP represents the portion of groundwater recharge percolating to the regional aquifer. Parameters used for calibrating the SWAT-MCA model are RCHRG_DP, GW_DELAY, ALPHA_BF, K, SY, α, β, and $S_{\min}$.

Table 3. Best calibrated parameter values of the original SWAT and SWAT-MCA models.

Variables	Wipperau basin		Neetze basin	
	SWAT	SWAT-MCA	SWAT	SWAT-MCA
r_CN2	- 15%	-	-10%	-
v_SURLAG (days)	0.1	-	0.1	-
r_SOL_AWC (mm H_2O)	-20%	-	-15%	-
r_SOL_K (mm/hr)	-3%	-	-10%	-
v_RCHRG_DP	0.75	0.86	0.15	0.45
v_ALPHA_BF (days)	0.15	0.25	0.2	0.2
v_GW_DELAY (days)	175	75	750	500

Note. "r_" and "v_" mean relative change and replacement of the default values, CN2 is the SCS curve number for moisture condition II, SURLAG is the surface runoff lag coefficient in the HRU, SOL_AWC is available soil water content, SOL_K is soil saturated hydraulic conductivity, RCHRG_DP is deep aquifer percolation coefficient, ALPHA_BF is baseflow recession constant, and GW_DELAY is delay time for aquifer recharge.

Daily streamflow at the basin outlet was used to calibrate and validate the original SWAT model. To calibrate and validate the SWAT-MCA model, however, both time series of observed daily streamflow at the basin outlets and observed groundwater levels at active cells were used. As the numbers of subbasins and cells are small, calibration of both models was performed manually by adjusting one parameter at a time until satisfactory model performance is achieved.

The common parameters of the SWAT and SWAT-MCA models (Table 3) were adjusted in the ranges suggested by Arnold et al. (2013) and other studies in lowland basins in Germany (e.g, Uniyal et al., 2017; Kiesel et al., 2010), while K was varied within the range mentioned in the previous section, from 0.864 m/d to 86.4 m/d. SY was varied between 0.5 and 0.3, which is the range for various geologic materials (Morris and Johnson, 1967). The constant value of β=2 was taken for all cells to represent the nonlinear behavior of the baseflow as pointed out by other studies (e.g., Pfannerstill et al., 2014a). $S_{\min}$ and α were varied cell to cell. It should be noted that for each active cell, five parameters need to be calibrated (K, SY, $S_{\min}$, α, and β). However, only one parameter (K) needs to be calibrated for each boundary cell. With these parameters, there could be a problem of equifinality because K and SY were varied in a wide range and the actual return flow from the deep aquifer to stream is unknown.

Model performance in terms of simulated streamflow and simulated groundwater levels was evaluated qualitatively and quantitatively. Qualitative evaluation of simulated streamflow was based on time series plots and the flow-duration curve, while qualitative evaluation of simulated groundwater levels was done based on time series plots. Quantitative evaluation of both simulated streamflow and groundwater levels was based on three statistical indices: the Nash-

Sutcliffe efficiency (NSE), the Percent Bias (PBIAS), and the ratio of the root mean square error to the standard deviation of measured data (RSR) (e.g., Moriasi et al., 2007; Sun et al., 2016). In addition, the boxplot of the absolute differences between observed and simulated groundwater levels was used to qualitatively evaluate simulated groundwater levels. The logarithmic Nash-Sutcliffe efficiency (lnNSE) index was used to quantitatively evaluate the quality of simulated low flow (e.g., Pushpalatha et al., 2012). To enable the calculation of lnNSE, all time steps with zero values of streamflow were excluded.

4 Results and discussion

4.1 Calibrated Parameter Values

The best calibrated parameter values of the original SWAT and SWAT-MCA models are listed in Table 3. It is seen from this table that RCHRG_DP was increased from 0.75 and 0.15 in the original SWAT model to 0.86 and 0.45 in the SWAT-MCA model. This is because in the original SWAT model, more water needs to be stored in the shallow aquifer to sustain adequate baseflow. In the SWAT-MCA model, however, baseflow is sustained by both local and regional aquifers. Therefore, less water needs to be stored in the local aquifer. In contrast, the best calibrated GW_DELAY value was decreased from 175 days and 750 days in the original SWAT model to 75 days and 500 days in the SWAT-MCA model. The reason is that the GW_DELAY value in the original SWAT model should represent both fast flow and slow flow responses of the baseflow to precipitation. In the SWAT-MCA model, the GW_DELAY value was reduced to represent the fast flow response of the baseflow. The slow flow response of the baseflow is controlled by α and β of the regional aquifer. Additional parameters, which were used to calibrate the new groundwater module of the SWAT-MCA model and their best calibrated values were varied within the ranges given in section 3.3, except the best calibrated K values were varied in a smaller range, from 1 m/d to 40 m/d.

4.2 Overall water balance

Figure 1 provides a schematic representation of the overall water balance from the calibrated SWAT-MCA model for the Wipperau and the Neetze basins. It clearly shows that there is a substantial amount of groundwater flow out of the two study areas. Results from both calibrated models show that (1) the simulated average annual groundwater recharge in the Wipperau and the Neetze basins is 258 mm and 243 mm, respectively, (2) a considerable amount of groundwater recharge, about 222 mm, in the Wipperau basin percolates to the deep aquifer compared to only 109 mm in the Neetze basin, and (3) the contribution of groundwater to streamflow (return flow) is dominant. About 71% and 84% of streamflow of the Wipperau basin and Neetze basin, respectively, are baseflow. The simulated groundwater recharge in this study is lower than the observed percolation at the lysimeter, but higher than the simulated value from Lemke and Elbracht (2008). The model currently underestimated irrigation demand compared with the reported values from Riediger et al. (2014). The incorporation of crop species could be done to improve the estimation of irrigation water demand. The amount of baseflow is close to the estimated values from the baseflow filter program (Table 2). In the original SWAT model, all baseflow is from the shallow aquifer, while in the SWAT-MCA model, about 44% of the baseflow of the Wipperau basin and 71% of the baseflow of the Neetze basin are from the local aquifer and the rest is from the regional aquifer.

4.3 Stream discharge

Figure 6a,b shows the observed and simulated daily streamflow at the outlet of the Wipperau basin with the original SWAT model and the SWAT-MCA model, respectively. The streamflow hydrographs were plotted in log scale to emphasize the quality of simulated low flow since improvement in high-flow simulation is not the focus of this present study. Visual assessment shows that low flows are poorly simulated with the original SWAT model, while with the SWAT-MCA model, low flows are better simulated. For example, simulated streamflow with the original SWAT sometimes dropped to zero, while observed low flows during the respective periods were much higher (Figure 6a), which is not observed with the SWAT-MCA model (Figure 6b). Better simulation of low flows with the SWAT-MCA model is also shown in the flow duration curve, where low flows are underestimated with the original SWAT model (Figure 7), and in the lnNSE efficiency (Table 4), where there is a significant improvement in the lnNSE with the SWAT-MCA. The changes in NSE, PBIAS and RSR indices between the two models are minor (Table 4) because these indices are more sensitive to high flows, which were underestimated in both models (Figure 7). The medium range of streamflow values was overestimated in both models (Figure 7). However, all statistical index values (Table 4) are within the 'satisfactory' range according to Moriasi et al., 2007, meaning that the SWAT and SWAT-MCA models for the Wipperau basin were successfully calibrated against streamflow and the simulated groundwater recharge can be considered as reasonable.

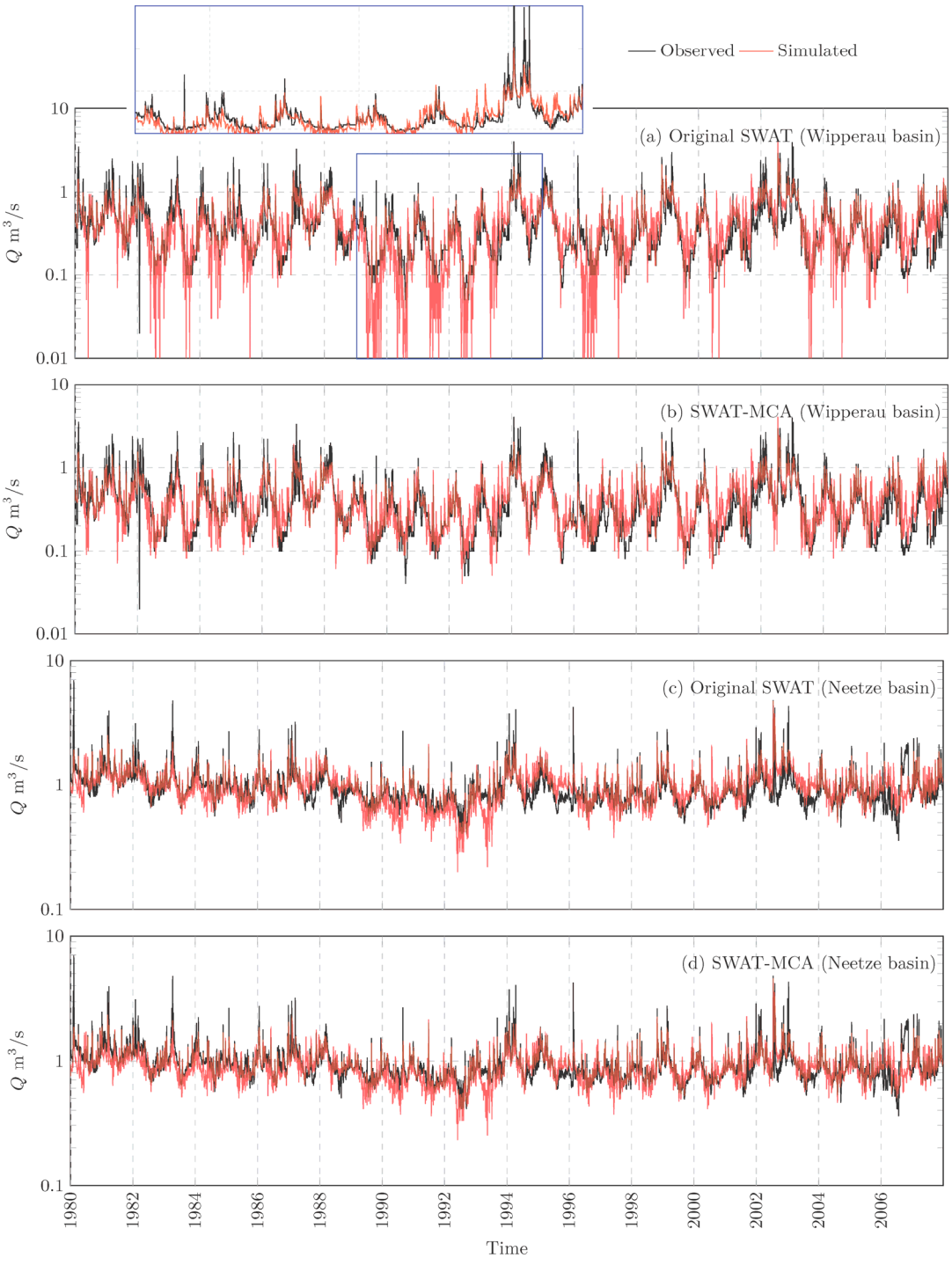


Figure 6. Observed and simulated streamflows from (a, c) the original SWAT models and (b, d) the SWAT-MCA model at the Oetzmühle (Wipperau basin) and Süttorf (Neetze basin) gauging stations during 1980–2007. A portion of the semi-log plot was shown in normal plot (a) to show that it is difficult to see the quality of the simulated low flow with normal plot. SWAT = soil and water assessment tool; MCA = multicell aquifer

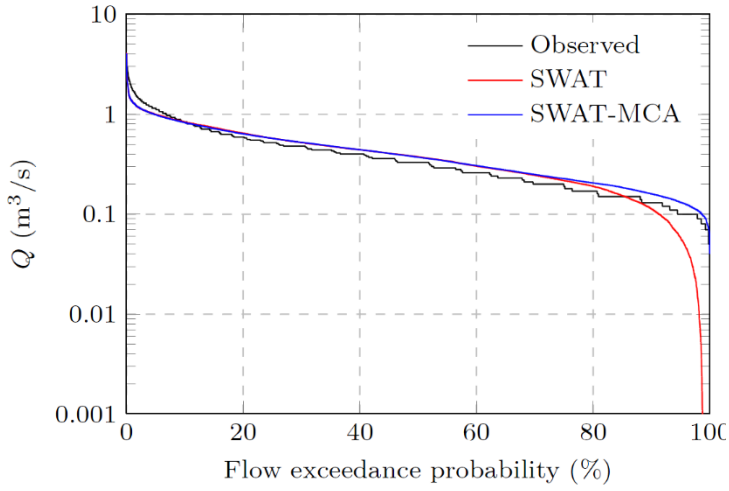


Figure 7. Flow duration curves of the observed and simulated streamflows at the Oetzmühle gauging station from the SWAT and SWAT-MCA models during 1980–2007. SWAT = soil and water assessment tool; MCA = multicell aquifer

Table 4. Performance of the calibrated SWAT and SWAT-MCA model (in terms of streamflow at the Oetzmühle and Süttorf gauging stations).

	Oetzmühle station (Wipperau basin)		Süttorf station (Neetze basin)	
	SWAT	SWAT-MCA	SWAT	SWAT-MCA
NSE	0.62 (0.65)	0.64 (0.65)	0.50 (**0.27**)	0.46 (**0.42**)
lnNSE	-0.07 (0.38)	0.63 (0.60)	0.38 (0.29)	0.39(0.46)
PBIAS	9.5 (-13.8)	2.8 (-11.8)	4.4 (-10.5)	8.7 (-3.8)
RSR	0.61 (0.59)	0.59 (0.60)	**0.70 (0.86)**	**0.73 (0.76)**

According to Moriasi et al., 2007, model performance is considered as satisfactory if NSE > 0.50, PBIAS < ±25%, and RSR < 0.7. Numbers outside parentheses indicate values of the calibration period while numbers inside parentheses indicate values of the validation period.

Figure 6c,d shows the observed and simulated daily streamflow of the Neetze basin with the original SWAT model and the SWAT-MCA model, respectively. Visual assessment shows that there is no significant difference between simulated streamflow from the original SWAT and the SWAT-MCA models. NSE and lnNSE values show that there are only minor improvements of simulated streamflow with the SWAT-MCA model (Table 4). However, NSE, lnNSE, and RSR indices show that the SWAT and SWAT-MCA models were less well calibrated against observed streamflow compared to Wipperau. This could indicate that the surface processes were not sufficiently represented. Nevertheless, the simulated groundwater recharge is considered as acceptable because of the following reasons: (1) the annual simulated groundwater recharge of the Neetze basin is similar with that of the Wipperau basin, (2) the Neetze and Wipperau basins are located close to each other and have similar climate and physical characteristics (Table 2), (3) the contribution of baseflow to streamflow in the Neetze basin is much larger than in the that in the Wipperau basin. This could be an explanation why the simulated streamflow in the Neetze basin is less satisfactorily calibrated than that in the Wipperau basin.

4.3 Groundwater levels

Figure 8a-c shows the time series plots of simulated and observed groundwater levels with the NSE, PBIAS, and RSR indices and the box plots of the absolute differences between observed and simulated groundwater levels. In these figures, observed groundwater levels fluctuated quite

smoothly without sudden increases and decreases, indicating that observed groundwater levels in these cells were not or only minor affected by extraction wells located nearby (Figure 5). Observed groundwater levels in other cells (Figure 9), however, were disturbed by extraction wells located nearby (Figure 5). This is shown by the sudden increases and decreases in observed groundwater levels (Figure 9). Therefore, only time series plots were used to evaluate simulated groundwater levels in these cells.

Figure 8a illustrates that in cells where there are no sharp increases and decreases of the observed groundwater levels, the simulated groundwater levels match well with the observed. Especially the long-term fluctuations are well reproduced by the SWAT-MCA model. It is interesting to note that the quality of the simulated groundwater levels in cells 7 and 9 in the Neetze basin, which are located in the middle of the modelling domain, is comparable with the simulated groundwater levels at other cells, which are neighbors of boundary cells. The boxplots (which were attached to the right of the time series plots) show that the third quartile varies from less than 0.2 m to about 0.4 m, meaning that 75% of the simulated groundwater level errors are within this range. The statistical indices NSE, PBIAS, and RSR, vary from -0.88 to 0.7, from -0.5% to 0.7%, and from 0.6 to 1.4, respectively.

Figure 8b presents cells, where the observed groundwater levels fluctuate quite smoothly similar to Figure 8a. Although the SWAT-MCA model here reproduces the long-term fluctuations as well, the time of occurrence of the simulated low and high groundwater levels mismatch with that of the observed. This problem could be due to the delay time for groundwater recharge to reach the regional aquifer, which is controlled by the GW_DELAY parameter. The GW_DELAY parameter was assigned as the same value for both the local and regional aquifers. The boxplots (except the boxplots of cells 10 and 11 in the Neetze basin) and the statistical indices (NSE, PBIAS, and RSR) also indicate higher errors compared with that of the cells shown in Figure 8a. The boxplots of cells 10 and 11 in the Neetze basin show that the third quartile is quite small mainly because the simulated and observed groundwater levels in these cells fluctuate in a narrow range (less than 0.5 m).

Figure 8c shows a group of cells where variations of observed groundwater levels were in a wider range and a shorter time compared with that of cells in Figure 8a and b. However, results show that the SWAT-MCA model can reproduce both long-term (multi-annual) and short-term (annual) groundwater fluctuations quite well. For example, it is seen that although observed groundwater levels in cell 12 of the Wipperau basin varied more than 2 m, the time series plot of simulated groundwater levels is almost identical to the observed and the third quartile is less than 0.2 m with NSE, PBIAS, and RSR are 0.9, 0%, and 0.3, respectively.

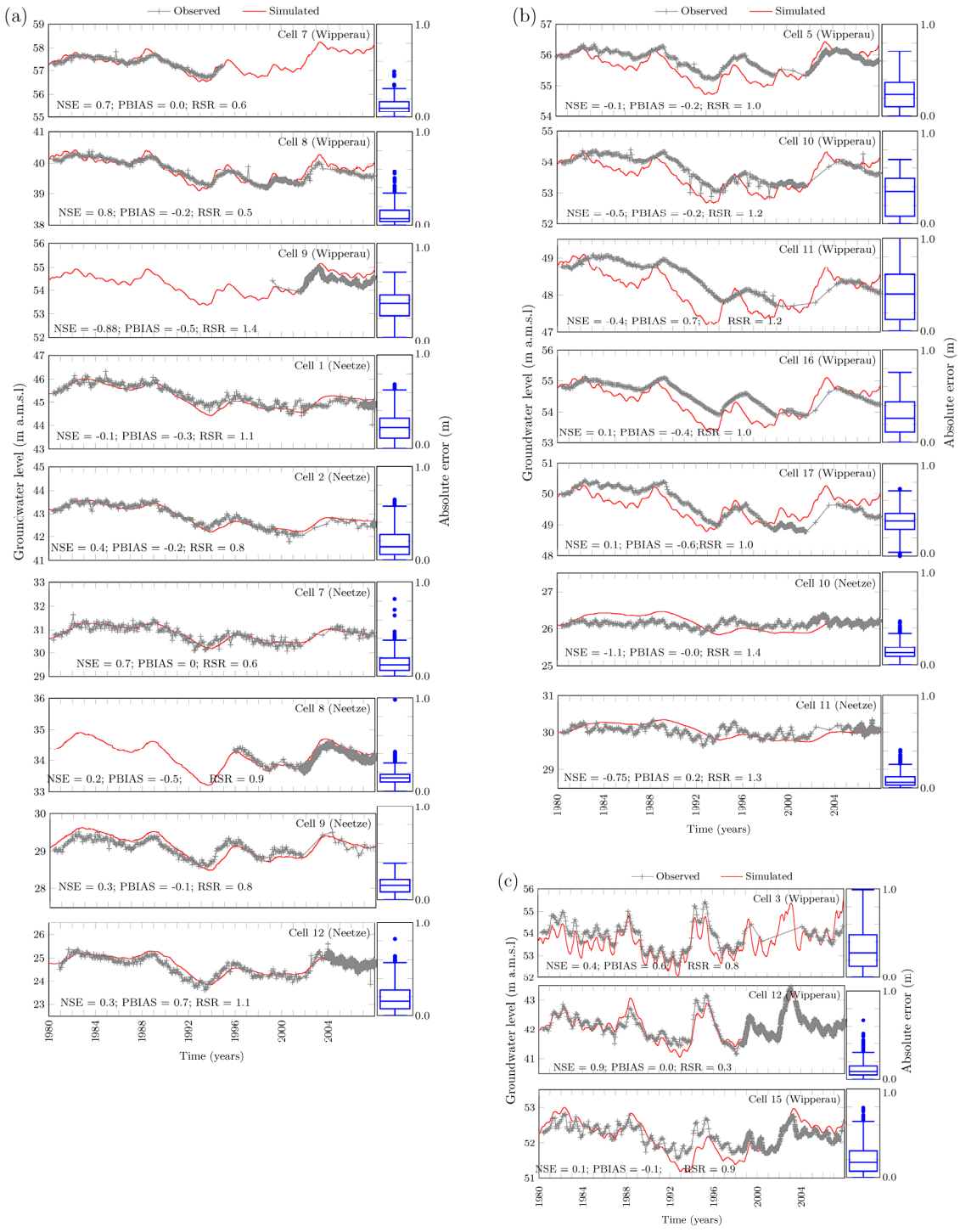


Figure 8. Time series plots of the observed and simulated groundwater levels during 1980–2007 and statistical indices (Nash–Sutcliffe efficiency [NSE], percent bias [PBIAS], and ratio of the root mean square error to the standard deviation of measured data [RSR]): (a) good match between observed and simulated groundwater levels, (b) mismatch between the time of occurrences of the high and low groundwater levels between observed and simulated groundwater levels, and (c) well reproduced of short- and longterm groundwater fluctuations. Boxplots of the absolute differences between simulated and observed groundwater levels are attached to the right of the time series plots

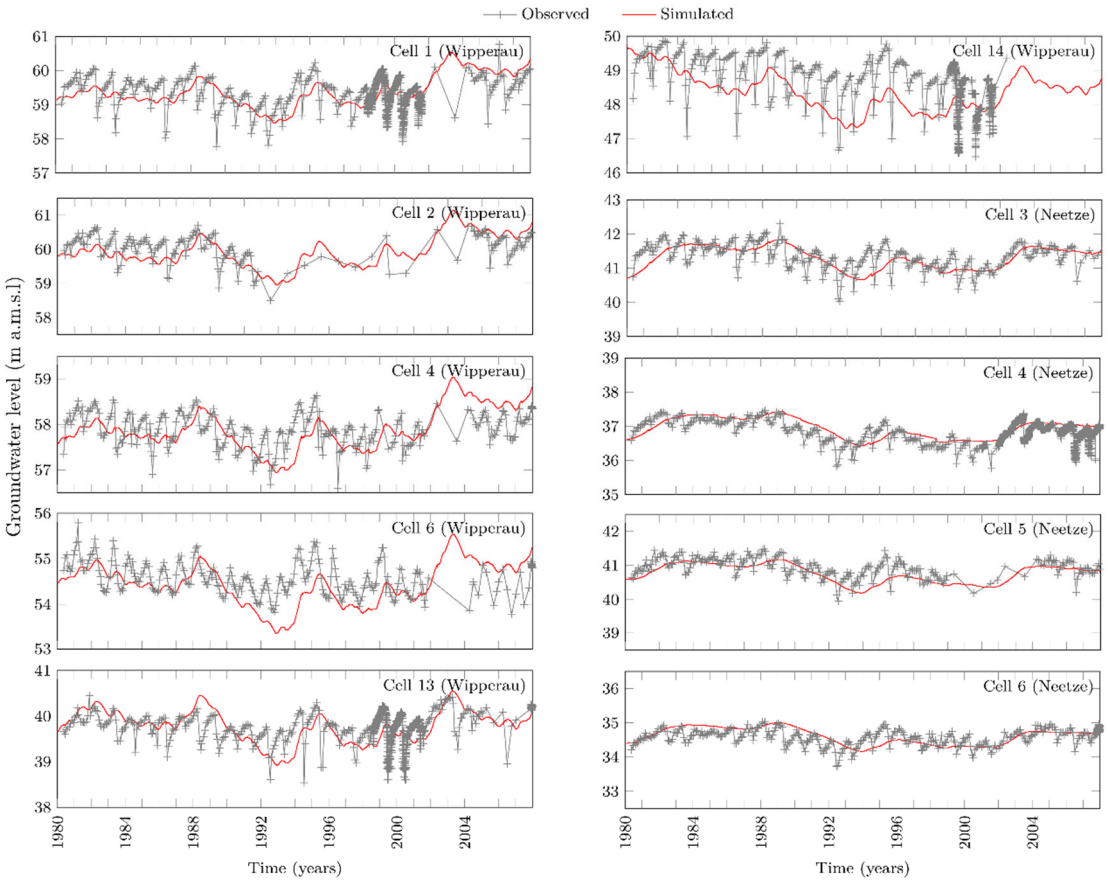


Figure 9. Time series plots of the observed and simulated groundwater levels during 1980–2007 in cells that are strongly affected by the cones of depression of the extractions wells located nearby

Figure 9 shows a group of cells, where observed groundwater levels in these cells decrease and increase sharply within a short time period. This can be explained by the effect of extraction wells for irrigation located nearby (Figure 5), which have not been incorporated directly into the model. Depression cones cannot be simulated by the SWAT-MCA model because of the model concept. The groundwater level is simulated for polygons, which are assumed to be homogenous, and cover an area larger than a depression cone from a single well. It is seen in these wells that the long-term fluctuations were reproduced by the model, however, the short-terms fluctuations cannot be reproduced. Although the extracted amount of groundwater was incorporated into the model indirectly by activating auto irrigation, however, the effect of extracted water for irrigation on the simulated groundwater level is weakened because the cell size is large.

5 Conclusion and recommendations

In this study, the SWAT model was modified to account for regional groundwater flow by replacing the deep aquifer with a multi cell aquifer (MCA), resulting in the SWAT-MCA model. SWAT-MCA was tested in two basins in Niedersachsen, Germany. Results showed that SWAT-MCA is able to simulate groundwater levels (which was previously impossible with the original SWAT model) and baseflow better than the original SWAT model. The SWAT-MCA model (1) requires less input data regarding the subsurface than fully distributed groundwater models (e.g., MODFLOW), (2) is computationally efficient due to its semi-distributed characteristic (using big

cell size), and (3) is simple and easy to use. The delineation of the regional aquifer and the flow direction between these aquifers units with the SWAT-MCA model are flexible. The delineation of the regional aquifer does not necessarily have to follow HRU or basin delineation and flow direction between the aquifer units is automatically defined by hydraulic gradient. If a dense network of groundwater observation wells is not available, SWAT-MCA can work with even a single cell with given boundary conditions, but the advantage over the original SWAT is minimal. The SWAT source code is open and available for download via its official website (http://swat.tamu.edu/). Additional source codes from the SWAT-MCA model could be also provided upon request.

In this study, model calibration was done against observed daily stream flow and groundwater levels. More advanced calibration and evaluation techniques could be applied to improve the model performance by including more constraints and criteria (Yilmaz et al., 2008), e.g. satellite-based soil moisture and evapotranspiration products (López et al., 2017), and by applying signature metrics (Pfannerstill et al., 2014b; Pokhrel et al., 2012).

Further works with the SWAT-MCA model are suggested to overcome its limitations. At the present, the SWAT-MCA model is only applied to areas with porous and unconfined aquifers. Discretization of cells with the MCA model can be more flexible. Cells could have a rectangular shape while representing any subsurface reservoir with equivalent storage and hydraulic conductivity. Although the SWAT-MCA model cannot capture drawdown cones at individual pumping wells because of the large cell size, reducing the cell size and having more data about the pumping well could solve this problem. However, it should be noted that this is not the objective of the MCA model, which is developed to use a small number of cells in order to make a clear difference with fully distributed subsurface models (Bear, 1979). In addition, the interaction between the regional aquifer and the river in the SWAT-MCA was simplified using a conceptual nonlinear storage-discharge relationship. A more physically based approach, which is based on the hydraulic head in the river and the underlying aquifer, could be used to model the aquifer-river interactions. However, it should be noted that increasing the model complexity could require more computational time, input data, and a more complex model does not always guarantee better performance.

Acknowledgments

We thank the editor and two anonymous reviewers, whose constructive comments helped to improve this manuscript. We also thank Bhumika Uniyal for proofreading the original manuscript. Furthermore, we thank the providers of data.

References

Ångström, A. (1924). Solar and terrestrial radiation. Report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation. Q*. J. Roy. Meteor. So.,* 50 (210), 121-126.

Arnold, J. G., & Fohrer, N. (2005). SWAT2000: current capabilities and research opportunities in applied watershed modelling. *Renew. Hydrol. Process.,* 19 (3), 563-572.

Arnold, J. G., Allen, P. M., & Bernhardt, G. (1993). A comprehensive surface-groundwater flow model. *Renew. J. Hydrology 142 (1-4)*, 47-69.

Arnold, J. G., Allen, P. M., Muttiah, R., & Bernhardt. G. (1995). Automated base flow separation and recession analysis techniques. *GroundWater,* 33, 1010-1018.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area hydrologic modeling and assessment. Part I: model development. *J. Am. Water Resour. Assoc.,* 34 (1), 73-89.

Arnold, J. G, Allen, P. M., Volk, M., Williams, J. R., & Bosch, D. D. (2010). Assessment of different representations of spatial variability on SWAT model performance. *Trans. ASABE,* 53 (5): 1433-1443.

Arnold, J. G., Kiniry, J. R., Srinivasan, R., Williams, J. R., Haney, E. B., & Neitsch, S. L. (2013). Soil and Water Assessment Tool Input/Output Documentation Version 2012. Tex. Water Resour. Inst., TR-439.

Bailey, R. T., Wible, T. C., Arabi, M., Records, R. M., & Ditty, J. (2016). Assessing regional-scale spatio-temporal patterns of groundwater–surface water interactions using a coupled SWAT-MODFLOW model. *Hydrol. Process.*, 30, 4420-4433.

Bear, J. (1979). *Hydraulics of Groundwater*. McGraw-Hill, New York.

Bear, J., & Cheng, A. H.-D. (2010). Modeling Groundwater Flow and Contaminant Transport. Springer, Dordrecht, Netherlands.

Bergström, S. (1992). *The HBV model: Its structure and applications*. Report Hydrology No. 4, Swedish Meteorological and Hydrological Institute.

Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrolo. Sci. J.*, 24 (1), 43-69.

Bieger, K., Arnold, J. G., Rathjens, H., White, M. J., Bosch, D. D., Allen, P. M., Volk, M., & Srinivasan, R. (2017). Introduction to SWAT+, A Completely Restructured Version of the Soil and Water Assessment Tool. *J. Am. Water Resour. Assoc.*, 53 (1), 115-130.

Bosch, D. D., Arnold, J. G., Volk, M., & Allen, P. M. (2010). Simulation of a low-gradient coastal plain watershed using the SWAT landscape model. *Trans. ASABE,* 53 (5), 1445-1456.

Brutsaert, W. (2005). Hydrology: An introduction. Cambridge University Press, Cambridge, UK.

Eckhardt, K., Haverkamp, S., Fohrer, N., & Frede, H.-G. (2002). SWAT-G, a version of SWAT99.2 modified for application to low mountain range catchments. *Phys. Chem. Earth, Parts A/B/C*, 27 (9), 641-644.

Efstratiadis, A., Nalbantis, I., Koukouvinos, A., Rozos, E., & Koutsoyiannis, D. (2008). HYDROGEIOS: a semi-distributed GIS-based hydrological model for modified river basins. *Hydrol. Earth Syst. Sci.*, 12, 989-1006.

Francesconi, W., Srinivasan, R., Pérez-Miñana, E., Willcock, S. P., & Quintero, M. (2016). Using the Soil and Water Assessment Tool (SWAT) to model ecosystem services: A systematic review. J. Hydrol. 535, 625-636.

Gan, R., Luo, Y. (2013). Using the nonlinear aquifer storage–discharge relationship to simulate the base flow of glacier- and snowmelt-dominated basins in northwest China. Hydrol. Earth Syst. Sci., 17, 3577-3586.

Gassman, P. W., Reyes, M., Green, C. H., & Arnold, J. G. (2007). The Soil and Water Assessment Tool: historical development, applications, and future research directions. *Trans. ASABE*, 50, 1211-1250.

Guse, B., Reusser D. E., & Fohrer, N. (2014). How to Improve the Representation of Hydrological Processes in SWAT for a Lowland Catchment – Temporal Analysis of Parameter Sensitivity and Model Performance. *Hydrol. Process.*, 28, 2651-2670.

Guzman, J. A., Moriasi, D. N., Gowda, P. H., Steiner, J. L., Starks, P. J., Arnold, J. G., & Srinivasan, R. (2015). A model integration framework for linking SWAT and MODFLOW. *Environ. Model. & Software,* 73, 103-116.

Hutson, S. S., Barber, N.L., Kenny, J. F., Linsey, K. S., Lumia, D. S., & Maupin, M. A. (2004). Estimated Use of Water in the United States in 2000. US Geological Survey Circular 1268, Reston, United States.

Jajarmizadeh, M., Harun, S., & Salarpour, M. (2012). A review on theoretical consideration and types of models in hydrology. *J. Environ. Sci. Technol.*, 5 (5), 249-261.

Joodavi, A., Zare, M., Raeisi, E., & Ahmadi M. B. (2016). A multi-compartment hydrologic model to estimate groundwater recharge in an alluvial-karst system. Arab. J. Geosci., 9(3), 195.

Kalin, L., & Hantush, M. H. (2006). Hydrologic modeling of an eastern Pennsylvania watershed with NEXRAD and rain gauge data. *J. Hydrol. Eng.,* 11, 555-569.

Kiesel, J., Fohrer, N., Schmalz, B., & White, M. J. (2010). Incorporating landscape depressions and tile drainages of a northern German lowland catchment into a semi-distributed model. *Hydrol. Process.*, 24, 1472–1486.

Kim, N. W., Chung, I. M., Won, Y. S., & Arnold, J. G. (2008). Development and application of the integrated SWAT–MODFLOW model. J. Hydrol. 356, 1-16.

Kirkby, M. J. (1998). Modelling across scales: the MEDALUS family of models. In: Boardman, J., Favis-Mortlock, D., Modelling Soil Erosion by Water. Springer, Berlin, 161-173.

Kite, G. W. (1997). *Manual for the SLURP hydrological model V. 11*. National Hydrology Research Institute, Saskatoon, Canada.

Krysanova, V., & White, M. (2007). Advances in water resources assessment with SWAT-an overview. *Hydrolo. Sci. J.*, 60 (5), 771-783.

Lane, L. J., & Nearing, M. A. (1989). USDA water erosion prediction project: hillslope profile model documentation. USDA-ARS National Soil Erosion Research Laboratory, NSERL Report No. 2.

Leavesley, G. H., Lichty, R. W., Troutman, B., & Saindon, L. G. (1983). Precipitation-runoff modelling system. User's manual. USGS Water Resour. Inst., Report 83-4238.

Lemke, D., & Elbracht, J. (2008). Grundwasserneubildung in Niedersachsen: eine vergleich der Methoden Dörhöfer & Josopait und GROWA06V2. GeoBerichte 10. Landesamt für Bergbau, Energie und Geologie, Hannover.

Lindström, G., Pers, C.P., Rosberg, R., Strömqvist, J., & Arheimer, B. (2010). Development and test of the HYPE (Hydrological Predictions for the Environment) model – A water quality model for different spatial scales. Hydrol. Res. 41 (3-4), 295-319.

López, P. L., Sutanudjaja, E. H., Schellekens, J., Sterk, G., & Bierkens, M. (2017). Calibration of a large-scale hydrological model using satellite-based soil moisture and evapotranspiration products. Hydrol. Earth Syst. Sci., 21, 3125--144.

Luo, Y., Arnold, J. G., Allen, P., & Chen, X. (2012). Baseflow simulation using SWAT model in an inland river basin in Tianshan mountains, Northwest China. *Hydrol. Earth Syst. Sci.*, 16, 1259-1267.

Luzio, M. D., Srinivasan, R., & Arnold, J. G. (2004). A GIS-coupled hydrological model system for the watershed assessment of agricultural nonpoint and point sources of pollution. *Trans. GIS*, 8 (1), 113-136.

Lv. M,, Hu, T., & Dan, L. (2014). Daily streamflow simulation in a small-scale farmland catchment using modified SWAT model. *Trans. ASABE*, 57 (1), 31-41.

Manhenke, V. O., Reutter, E., Hübschmann, M., Limberg, A., Lückstedt, M., Nommensen, B., Peters, A., Schlimm, W., Taugs, R., & Voigt, H. J. (2001). Hydrostratigrafische Gliederung des nord-und mitteldeutschen känozoischen Lockergesteinsgebietes. *Z. Angew. Geol.*, 47 (3-4), 146-152.

McDonald, M. G., Harbaugh, & A. W. (1988). A Modular Three-dimensional Finite-difference Ground-water Flow Model. US Geological Survey, Techniques of Water Resources Investigation Book 6/A1.

Moriasi, D. N., Arnold, J. G., Van Liew M. W., Bingner, R. L., Harmel, R. D., & Veith T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE*, 50 (3), 885-900.

Morris, D. A., & Johnson, A. I. (1967). Summary of hydrologic and physical properties of rock and soil materials as analyzed by the Hydrologic Laboratory of the U.S. Geological Survey. U.S. Geological Survey Water-Supply Paper 1839-D.

Neitsch, S. L., Arnold, J. G., Kiniry, J. R., & Williams, J. R. (2011). Soil and Water Assessment Tool Theoretical Documentation Version 2009. Tex. Water Resour. Inst., TR-406.

Pechlivanidis, I. G., Jackson, B. M., Mcintyre, N. R., & Wheater, H. S. (2011). Catchment scale hydrological modelling: a review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. *Global NEST J.*, 13 (3), 193-214.

Pfannerstill, M., Guse, B., & Fohrer, N. (2014a). A multi-storage groundwater concept for the SWAT model to emphasize nonlinear groundwater dynamics in lowland catchments. *Hydrol. Process.*, 28 (22), 5599-5612.

Pfannerstill, M., Guse, B., & Fohrer, N. (2014b). Smart low flow signature metrics for an improved overall performance evaluation of hydrological models. *J. Hydrol.*, *510*, 447-458.

Pignotti, G., Rathjens, H., Cibin, R., Chaubey, I., & Crawford, M. (2017). Comparative Analysis of HRU and Grid-Based SWAT Models. *Water*, 9(4), 272.

Pokhrel, P., Yilmaz, K. K., & Gupta, H. V. (2012). Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures. *J. Hydrol.*, *418-419*, 49-60.

Pushpalatha, R., Perrin, C., Le Moine, N., & Andréassian, V. (2012). A review of efficiency criteria suitable for evaluating low-flow simulations. *J. Hydrol*., 420-421, 171-182.

Rathjens, H., Oppelt, N., Bosch, D.D., Arnold, J. G., & Volk, M. (2015). Development of a Grid-Based Version of the SWAT Landscape Model. Hydrol. *Process*., 29, 900–914.

Refsgaard, J. C., Hjberg, A. L., Møjller, I., Hansen, M., & Søndergaard, V. (2010). Groundwater Modeling in Integrated Water Resources Management—Visions for 2020. *GroundWater*, 48 (5), 633-648.

Riediger, J., Breckling, B., Nuske, R. S., & Schröder, W. (2014). Will climate change increase irrigation requirements in agriculture of Central Europe? A simulation study for Northern Germany. Env. Sci. Eur. 26:18.

Rozos, E., Efstratiadis, A., Nalbantis, I., & Koutsoyiannis, D. (2004). Calibration of a semi-distributed model for conjunctive simulation of surface and groundwater flows / Calage d'un modèle semi-distribué pour la simulation conjointe d'écoulements superficiels et souterrains. Hydrol. Sci., 49 (5), 819-842.

Rozos, E., & Koutsoyiannis, D. (2010). Error analysis of a multi-cell groundwater model. J. Hydrol., 392 (1-2), 22-30.

Siebert, S., Burke, J., Faures, J. M., Frenken, K., Hoogeveen, J., Döll, P., & Portmann, F. T. (2010). Groundwater use for irrigation – a global inventory. *Hydrol. Earth Syst. Sci.*, 14, 1863-1880.

Sophocleous, M. A., Koelliker, J. K., Govindaraju, R. S., Birdie, T., Ramireddygari, S. R., & Perkins, S. P. (1999). Integrated numerical modeling for basin-wide water management: the case of the Rattlesnake Creek basin in south-central Kansas. *J. Hydrol*., 214, 179-196.

Strauch, M., Lima, J. E. F. W., Volk, M., Lorz, C., & Makeschin, F. (2013). The impact of Best Management Practices on simulated streamflow and sediment load in a Central Brazilian catchment. *J. Environ. Manage.*, 127, S24-S36.

Sun, X., Bernard-Jannin, L., Garneau, C., Volk, M., Arnold, J. G., Srinivasan, R., Sauvage, S., & Sánchez-Pérez, J. M. (2016). Improved simulation of river water and groundwater exchange in an alluvial plain using the SWAT model. *Hydrol. Process*., 30 (2), 187-202.

Thiessen, A. H. (1911). Precipitation averages for large areas. *Mon. Weather Rev*., 39 (7), 1082-1089.

Tóth, J. (1963). A theoretical analysis of groundwater flow in small drainage basins. *J. Geophys. Res*., 68, 4795-4812.

Uniyal, B., Dietrich, J., Vasilakos, C., & Tzoraki, O. (2017). Evaluation of SWAT simulated soil moisture at catchment scale by field measurements and Landsat derived indices. *Agric. Water Manage.,* 193, 55-70.

Vazquez-Amábile G. G., & Engel B. A. (2005). Use of SWAT to compute groundwater table depth and streamflow in the Muscatatuck River watershed. *Trans. ASABE,* 48 (3), 991-1003.

Volk, M., Arnold, J. G., Bosch, D. D., Allen, P. M., & Green, C. H. (2007). Watershed configuration and simulation of landscape processes with the SWAT model. In: Oxley, L., Kulasiri, D. (Eds.), MODSIM 2007 International Congress on Modelling and Simulation. Christchurch, Modelling and Simulation Society of Australia and New Zealand, 2382-2389.

Wagner, P. D., Bhallamudi, S. M., Narasimhan, B., Kantakumar, L. N., Sudheer, K. P., Kumar, S., Schneider, K., & Fiener, P. (2016). Dynamic Integration of Land Use Changes in a Hydrologic Assessment of a Rapidly Developing Indian Catchment. *Sci. Total Environ.,* 539, 153–164.

Wang, Y., & Brubaker, K. (2014). Implementing a nonlinear groundwater model in the Soil and Water Assessment Tool (SWAT). *Hydrol. Process.,* 28 (9), 3388-3403.

Wessolek, G., Kaupenjohann, M., & Renger, M. (2009). Bodenphysikalische Kennwerte und Berechnungsverfahren für die Praxis. Bodenökologie und Bodengenese, Heft 40, Technische Universität Berlin.

Wittenberg, H. (2015). Groundwater abstraction for irrigation and its impacts on low flows in a watershed in northwest Germany. *Resour.,* 4(3), 566-576.

Wittenberg, H. (2003). Effects of season and man-made changes on baseflow and flow recession: case studies. *Hydrol. Process.,* 17, 2113-2123.

Yilmaz, K. K., Gupta, H. V., & Wagener, T. (2008). A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resour. Res.*, 44(9).

This chapter is an edited version of the following original scientific article:

Nguyen, V. T.; Dietrich, J. (2018): Modification of the SWAT Model to Simulate Regional Groundwater Flow Using A Multi-Cell Aquifer. Hydrological Processes 32(7), 939-953, DOI: 10.1002/hyp.11466.

Publisher: John Wiley and Sons

License: Permit for re-use in dissertation/thesis given by publisher

4 Diskussion

4.1 Zusammenfassende kritische Bewertung der Arbeiten

Im Rahmen der hier vorgestellten Arbeiten erfolgte ein früher Ansatz der angepassten Nutzung von Ensemble-Produkten unterschiedlicher Vorlaufzeiten (später von anderen Forschern als „seamless prediction“ sowohl in Richtung Now-casting als auch sub-saisonale Vorhersage weiterentwickelt) sowie eine der ersten publizierten Evaluierungen von Ensemble-Hochwasservorhersagen bezüglich einer probabilistischen kategorischen Vorhersage der Warnstufe. Zum Zeitpunkt der Durchführung der Studien waren allerdings die verfügbaren Daten auf wenige Hochwasserereignisse limitiert und vorwiegend nachträglich mit den operationellen Systemen gerechnete Vorhersagen verfügbar (Hindcasts). Daher konnten die Ergebnisse nicht generalisiert werden. Die prozessorientierte Flussgebietsmodellierung hat sich in den letzten zehn Jahren in der Hochwasservorhersage als quantitatives systemanalytisches Werkzeug der Wasserwirtschaft etabliert, verbunden mit einem enormen Kenntnisgewinn der operationellen Dienste (Philipp & Kerl, 2018). In Sachsen erfolgte mit dem Aufbau einer Hochwasser-Vorhersagezentrale ein Ausbau der Ombrometer-Messnetze zur Gewinnung hochaufgelöster Niederschlagsdaten sowie eine Systemintegration von Daten, Modellen und Auswertungswerkzeugen. Es erfolgte ferner ein deutlicher Ausbau der probabilistischen (im engeren Sinne frequentistischen) Vorhersage in der Meterorologie (z. B. COSMO-DE- Ensemble des DWD). Die Angabe von Unsicherheitsbandbreiten der vorhergesagten Abflüsse wird zunehmend vorgenommen, z. B. auch in der operationellen Hochwasservorhersage des Landes Niedersachsen, während eine echte probabilistische Vorhersage auch in weit entwickelten operationellen Systemen noch nicht in die Praxisanwendung gelangt ist (Philipp & Kerl, 2018). Für die Frühwarnung mit einer für kleinere Gebiete als ausreichend angesehenen Vorlaufzeit von einem Tag wird in Sachsen vielmehr auf einfachere Ersatzmodellansätze gesetzt. Auf europäischer Ebene existiert ein operationelles Abfluss-Vorhersagesystem mit hohem fachlichen Entwicklungsgrad aber aufgrund der Skala begrenzter operationeller Anwendbarkeit in kleineren Flussgebieten (EFAS, Thielen et al., 2009).

Zu den Erkenntnissen dieser Arbeit gehört in Übereinstimmung mit den Erkenntnissen anderer Arbeiten eine gute technische Nutzbarkeit von Ensembles für die operationelle Hochwasservorhersage, da eine Vorhersage unter Einbeziehung der Unsicherheit der deterministischen Vorhersage in der Regel überlegen ist. Sinnbildlich erfolgt hier die Anwendung des Prinzips der kollektiven Intelligenz auf Simulationen. Je nach Typ des Ensembles werden unterschiedliche Modelle, unterschiedliche Parametrisierungen, unterschiedliche Anfangszustände so kombiniert, dass das Ensemble oder daraus abgeleitete Größen im Mittel besser liegen können als eine einzelne Vorhersage. Vor allem die Fehlerquoten oder die Totalversagen der Vorhersage sollten dadurch reduziert werden. Allerdings stellt die Information über die Unsicherheit hohe Anforderungen an die Entscheidungsfindung. Hier sind bisher vergleichsweise wenige Ergebnisse publiziert worden und es besteht sowohl Forschungsbedarf als auch geringe Erfahrung in der Praxis (Leskens et al., 2014).

In der Klimafolgenforschung hat sich die Validierung der Klimadatensätze für den Beobachtungszeitraum und die Anwendung von Methoden des statistischen Postprocessing als bedeutend gezeigt und ist von der wasserwirtschaftlichen Praxis aufgegriffen worden (Hölscher et al., 2012). Bias-Korrekturen mögen dahingehend umstritten sein, dass sie nachträglich ein Modellergebnis korrigieren und damit die simulierten Verhältnisse der Variablen stören. In der Impaktmodellierung zeigten sich aber deutlich bessere Ergebnisse, wenn systematische Fehler relevanter Größen korrigiert werden. Das Erkenntnisinteresse in der Hydrologie und Wasserwirtschaft bezieht sich häufig auf die Betrachtung der Veränderung einzelner Variablen (z. B. Trends) und weniger auf die Abbildung realistischer Momentaufnahmen des meteorologischen Gesamtsystems. Die Notwendigkeit der Korrektur wichtiger Variablen reduziert allerdings das Vertrauen in die Zuverlässigkeit von Klimafolgenstudien und verlangt nach einer weiteren Verbesserung der Klimamodelle vor allem auf der regionalen Ebene. Die Analyse des „Tripels“ aus a) Klimabeobachtung der Vergangenheit, b) Kontrolllauf des Klimamodells für die Vergangenheit und c) Vorhersagelauf des Klimamodells (Klimaprojektion) ist unerlässlich, da die Unsicherheit des Klimamodells bei der Ableitung von Schlussfolgerungen berücksichtigt werden muss. Ergänzend kann mit dem Klimamodell auch eine Re-Analyse der Vergangenheit gerechnet werden.

Die horizontale hydrologische Vernetzung erfolgt in Flussgebietsmodellen bisher überwiegend über das Gewässernetz der Oberflächengewässer, welche in einer im Modell verankerten Netzwerktopologie die Abflussanteile der Teileinzugsgebiete sammeln und mit Anwendung hydrologischer Wellenablauf-Gleichungen (z.B. Muskingum, Kalinin-Miljukov) zum Auslass des Flussgebietes leiten. Die horizontale Vernetzung der Grundwasser-komponente des Flussgebietsmodells SWAT wurde für Porengrundwasserleiter unter Anwendung der Darcy-Gleichung erfolgreich umgesetzt, ohne dazu ein spezielles Grundwassermodell koppeln zu müssen. Damit wurde ein Kompromiss zwischen der leichten Anwendbarkeit eines konzeptionellen Modells wie SWAT und der besseren Prozessabbildung in spezialisierten Modellen geschaffen. Die simulierten Grundwasserstände erweitern den Anwendungsspielraum für SWAT auf Studien mit unterirdischen Wasserflüssen zwischen Teileinzugsgebieten und auf Untersuchungen zu großräumigen Auswirkungen von Grundwasserentnahmen, z. B. für die Bewässerung. Kleinräumige Betrachtungen, z. B. die Simulation von Absenktrichtern, sind mit dieser Art von Modell allerdings nicht möglich und bleiben hoch aufgelösten hydraulischen Modlelen vorbehalten. Auch ist die entwickelte Lösung bisher nur in Gebieten mit guter Datenlage umsetzbar und nicht ohne weiteres auf andere Aquifertypen übertragbar.

Die Simulation der Bewässerung im Flussgebietsmaßstab zeigte, dass die automatische bodenfeuchteabhängige Bewässerungssteuerung in SWAT plausible Bewässerungsmengen simuliert, während der durch die Evapotranspiration und den Aufbau der Biomasse bedingte Pflanzenwasserbedarf stark unterschätzt wird und damit die Steuerung der Bewässerung über die Evopotranspiration nicht zufriedenstellend funktioniert. Ferner berücksichtigt die Bodenfeuchte-Methode nicht die Wachstumsstadien der Pflanzen, wobei hier Chen et al. (2018) Weiterentwicklungen vorschlagen. Die Anwendung von SWAT für planerische wasserwirtschaftliche Fragen der Bewässerung im Flussgbietsmaßstab konnte erfolgreich demonstriert werden. Hierbei ist insbesondere die Güte der Simulation der Bodenfeuchte

relevant. Da kein Monitoringnetzwerk für Bodenfeuchte zur Verfügung steht, werden Flussgebietsmodelle selten auf die Bodenfeuchte kalibriert sondern lediglich auf den Abfluss und in jüngerer Zeit auch auf Fernerkundungsdaten der Evapotranspiration. Die Notwendigkeit und der Nutzen flächendeckender hoch ausgelöster Bodenfeuchtedaten wurden gezeigt. Die Ableitung der Bodenfeuchte aus Landsat-Daten, wie in dieser Arbeit vorgenommen, hat zwar den Vorteil einer hohen Raumauflösung (30 m), ist jedoch in vielen Regionen aufgrund von Bewölkung und aufgrund des hohen Aufwandes der Kalibirerung anhand von Bodendaten nicht praktikabel. Es ist festzustellen, dass die Bodenfeuchte in vielen Regionen der Welt (einschließlich Deutschland) nicht zufriedenstellend erfasst wird, das heißt regelmäßig in ausreichender räumlicher Dichte sowie für unterschiedliche Bodenarten und Vegetationsarten. Neue, höher aufgelöste Fernerkundungsdaten werden die Datenlage im Bereich Bodenfeuchte deutlich verbessern.

Die Anwendungen eines Wasserbewirtschaftungsmodells (WEAP) auf Fallstudien haben gezeigt, dass die hydrologische Komponente dieses Modells sehr stark vereinfacht ist und die Anwendung für wassermengenwirtschaftliche Fragestellungen einschränkt. Allerdings gibt es für die Simulation des Wasserbedarfs unter globalem Wandel in der Regel keine entsprechenden Funktionalitäten in hydrologischen Flussgebietsmodellen. Für eine praxisbezogene Anwendung unter hoher Unsicherheit der Eingangsdaten oder bei Datenknappheit bietet sich dennoch WEAP an, weshalb es häufig in Entwicklungsländern angewendet wird.

4.2 Ausblick

Im Bereich der Ensemblesimulation zeichnet sich eine Ausweitung der wasserwirtschaftlichen Anwendungsfälle auf subsaisonale Vorhersagen für die Planung länger andauernder Wettersituationen ab, dabei auch extreme Trockenheiten. Zu den offenen Fragen mit weiterem Forschungsbedarf gehört die Ermittlung von Entscheidungsregeln für die Entscheidung unter Unsicherheit, wobei eine Reduzierung der Ensembleinformation auf zumeist binäre Entscheidungen und Handlungen erfolgen muss. Die künftige Forschung in der wasserwirtschaftlichen Nutzung von Ensembles sollte verstärkt auf die direkte Ableitung von kategorischen Warnungen aus der probabilistischen meteorologischen Vorhersage und relevanten hydrologischen Gebietszuständen setzen, z. B. über Bayes'sche Netzwerke oder andere datenbasierte Ansätze, welche nach Verfügbarkeit längerer Zeiträume mit Vorhersagen auch extremer (und damit seltener) Situationen besser trainiert werden können. Die Umsetzung von Frühwarnsystemen in die Praxis bleibt insbesondere in Entwicklungsländern eine große Herausforderung, wo interdisziplinäre Arbeiten zusammen mit Sozialwissenschaftlern und transdisziplinäre Arbeiten zusammen mit zuständigen operationellen Betreibern und Nutzern der Systeme erfolgen sollten. An der Schnittstelle Computer – Mensch können auch religiöse Fragen (Beziehung extremer Naturereignisse auf Götter), Vertrauen in Autoritäten (Umgang mit unvermeidlichen Fehlwarnungen) und die Nutzung indigenen bzw. lokalen Wissens über die Kompetenz technisch spezialisierter Fachexperten hinausgehen, aber von Bedeutung für die Praxisumsetzung von modellgestützten Risikomangementstrategien sein. Dies stellt eine große Herausforderung an die Empathie der betreffenden Forscher dar sowie deren Liebe zum eigentlichen Objekt

und nicht nur der Methode, welche zumindest in der Grundlagenforschung zumeist im Vordergrund steht. Auch an die Zielgruppen einer bevölkerungsnahen Wissenschaft werden hohe Anforderungen gestellt bezüglich des Vertrauens in die Objektivität und das Wissen der Forscher, aber auch in Bezug auf das Verständnis der Komplexität und die Bereitschaft zur Veränderung bzw. die Anpassung eigener Verhaltensweisen. Eine effektive Hochwasserwarnung geht weit über die technischen Möglichkeiten hinaus und öffnet zahlreiche Forschungsfelder unterschiedlciher Disziplinen.

In der Flussgebietsmodellierung sollten die aus verschiedenen Erkenntnisinteressen entwickelten Techniken zur Nutzung neuartiger Eingangsdaten aus Fernerkundung und meteorologischer Simulation, zur Parametrisierung, zur Unsicherheitsanalyse und letztlich zur Verbesserung der Prozessbeschreibung auf dieser Maßstabsebene fortgeführt und zu neuen Synergien gebracht werden. Die Parameter-Regionalisierung zeigte vielversprechende Ergebnisse und wurde in Folgearbeiten mit hybriden Modellansätzen durch Anwendung künstlicher Intelligenz (Wallner et al., 2013) methodisch verbessert. Im Bereich der wassermengenwirtschaftlichen Anwendung erfolgen im Unterschied zur hydrologischen Simulation nur selten Studien zur Parameterunsicherheit, zu strukturierten Methoden der Parameteroptimierung sowie zur Vorhersageunsicherheit. Hier sollten die jüngst erzielten Erkenntnisse in Simulation und Risikomanagement aus „Extremen mit viel Wasser“ (Hochwasser) auf „Extreme mit wenig Wasser“ (Trockenheit) sowie auf Anwendungsfälle des landwirtschaftlichen Wassermanagements übertragen werden, wobei ein Mangel an großflächig verfügbaren zuverlässigen Daten über real aufgebrachte Wassermengen die wissenschaftliche Arbeit in letzterem Bereich erschwert.

Die hydrologische Konnektivität ist Gegenstand naturwissenschaftlicher Grundlagenforschung, spielt aber auch in der Wasserressourcenbewirtschaftung eine wichtige Rolle. Weitere Forschungstätigkeiten beschäftigen sich daher mit der horizontalen Vernetzung von Karstgrundwasserleitern (Nguyen & Dietrich, 2018), mit vulkanischen Grundwasserleitern sowie mit alluvialen Grundwasserkörpern in semi-ariden Gebieten. Erweiterungen von SWAT für hoch durchlässige Aquifere mit unterirdischen Wasserflüssen zwischen Teileinzugsgebieten sind in Arbeit.

Die Abbildung der Bewässerungssteuerung in Flussgebietsmodellen sollte verbessert werden, damit Prognosen des zukünftigen Bewässerungsbedarfs unter globalem Wandel eine höhere Präzision erhalten als mit grob gerasterten globalen Modellen möglich. Multiskalige Betrachtungen der Bewässerung in Modellen können neben der Nutzung experimenteller Daten von Versuchsfeldern (wie in Niedersachsen von der Landwirtschaftskammer in Hamerstorf betrieben) zur Übertragung des Wissens aus der Feldskala auf die Flussgebietsskala beitragen (Dietrich et al., 2018). Entsprechende Studien mit Feld- und Flussgebietsmodellen sollen durchgeführt werden und die Überarbeitung der Bewässerungskomponente von SWAT ermöglichen. Da SWAT de facto Standard in der Simulation landwirtschaftlicher Flussgebiete ist und es in der Wissenschaft eine breite Anwendung gefunden hat, kommt dem Modell bei Untersuchungen in Zusammenhang mit dem globalen Wandel auf der Skala von Flussgebieten weltweit eine weiterhin große Bedeutung zu (Uniyal & Dietrich, 2018). Erkenntnisse in verschiedenen Bereichen der

Hydrologie und Landwitschaft tragen regelmäßig zur Modernisierung des frei zugänglichen Modellcodes bei.

Die Nutzung neuer Fernerkundungsdaten (z.B. Sentinel) insbesondere zu Bodenfeuchte und Evapotranspiration könnte die Simulation bewässerter Flussgebiete verbessern und auch die Grundlage für eine aufgrund fehlender Daten problematische Weiterentwicklung der wasserwirtschaftlichen Routinen in Flussgebietsmodellen liefern. Operationell verfügbare hoch aufgelöste Bodenfeuchtedaten hätten einen multiplen Nutzen für mehrere der betrachteten Anwendungsfälle von Flussgebietsmodellen. Zum einen können sie zur besseren Kalibrierung räumlich verteilter Modelle dienen. Zum anderen können Bodenfeuchtedaten zur Zustandsnachführung hydrologischer Modelle verwendet werden. Dies kann zur Verbesserung der Vorhersagegüte operationeller Modelle beitragen, da simulierte Anfangsbedingungen (Systemzustände) fehlerhaft sein können. Die Assimilation von Bodenfeuchtedaten kann in der Hochwasservorhersage als auch in der modellgestützten operationellen Bewässerungsplanung und -steuerung zur Anwendung kommen.

Die integrative Bewirtschaftung von Wasserressourcen (IWRM) hat in den vergangenen Jahren zu einer engeren Integration natur-, sozial- und ingenieurwissenschaftlicher Methoden geführt. Die IAHS hat ihre wissenschaftliche Dekade 2013 bis 2022 „Panta Rhei – Change in Hydrology and Society“ genannt. In dem Zusammenhang wurde auch der Begriff der „Sozio-Hydrologie“ eingeführt (Sivapalan et al. 2012). Für die Modellierung von Flussgebieten ergibt sich ein wieder aufkommendes Interesse an der Kopplung hydrologischer und sozialer Modelle, z. B. der agentenbasierten Modellierung (ABM), um Rückkopplungen zwischen Hydrologie und Gesellschaft sowohl besser analysieren als auch vorhersagen zu können. Für den Bereich der Frühwarnung wurde bereits von Molina und Blasco (2004) das Verhalten der an der Warnkette beteiligten Personen in Form eines ABM simuliert. ABM oder ähnlich gelagerte Ansätze könnten künftig die systemanalytischen Werkzeuge der Wasserwirtschaft ergänzen. Schwächen von ABM liegen in der zumeist fehlenden Validierungsmöglichkeit und in der generellen Problematik der Modellierung sich frei entscheidender Agenten, d.h. letztlich in der Prognose menschlichen Verhaltens. An diesen wichtigen Schnittstellen zwischen Natur- und Ingenieruwissenschaften sowie Sozialwissenschaften sollte weitere Forschung erfolgen.

Literaturverzeichnis

An dieser Stelle wird nur die außerhalb der beigefügten Fachartikel zitierte Literatur angegeben. Zur Dokumentation des Fortschritts der Praxisanwendung der beschriebenen Methoden sowie zur Dokumentation laufender Forschungstätigkeiten werden hier auch nicht referierte Veröffentlichungen von Behörden und Konferenzbeiträge aufgeführt.

Abbaspour, K.C., Rouholahnejad, E., Vaghefi, S., Srinivasan, R., Yang, H., Kløve, B. (2015): A continental-scale hydrology and water quality model for Europe. Calibration and uncertainty of a high-resolution large-scale SWAT model. Journal of Hydrology 524, pp. 733–752. DOI: 10.1016/j.jhydrol.2015.03.027.

Ahrends, H., Mast, M., Rodgers, C., Kunstmann, H. (2008): Coupled hydrological–economic modelling for optimised irrigated cultivation in a semi-arid catchment of West Africa. Environmental Modelling & Software 23 (4), pp. 385–395. DOI: 10.1016/j.envsoft.2007.08.002.

Ajami, N.K., Duan, Q., Sorooshian, S. (2007): An integrated hydrologic Bayesian multimodel combination framework. Confronting input, parameter, and model structural uncertainty in hydrologic prediction. Water Resour. Res. 43 (1), 655. DOI: 10.1029/2005WR004745.

Ameli, A.A., Creed, I.F. (2017): Quantifying hydrologic connectivity of wetlands to surface water systems. Hydrol. Earth Syst. Sci. 21 (3), pp. 1791–1808. DOI: 10.5194/hess-21-1791-2017.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., Williams, J. R. (1998): Large Area Hydrologic Modeling and Assessment Part I. Model Development. J Am Water Resources Assoc 34 (1), pp. 73–89. DOI: 10.1111/j.1752-1688.1998.tb05961.x.

Assouline, S., Russo, D., Silber, A., Or, D. (2015): Balancing water scarcity and quality for sustainable irrigated agriculture. Water Resour. Res., 51, 3419–3436, doi:10.1002/2015WR017071.

Baroni, G., Tarantola, S. (2014): A General Probabilistic Framework for uncertainty and global sensitivity analysis of deterministic models. A hydrological case study. Environmental Modelling & Software 51, pp. 26–34. DOI: 10.1016/j.envsoft.2013.09.022.

Bartels, J., Bliefernicht, J., Seidel, J., Bárdossy, A., Kunstmann, H., Johst, M., Demuth, N. (2017): Bewertung von Ensemble-Abflussvorhersagen für die operationelle Hochwasserwarnung. Hydrologie und Wasserbewirtschaftung / BfG – Jahrgang: 61.2017,5 - ISSN 1439-1783. DOI: 10.5675/HyWa_2017,5_1.

Barthel, R., Reichenau, T.G., Krimly, T., Dabbert, S., Schneider, K., Mauser, W. (2012): Integrated Modeling of Global Change Impacts on Agriculture and Groundwater Resources. Water Resour Manage 26 (7), pp. 1929–1951. DOI: 10.1007/s11269-012-0001-9.

Bastiaanssen, W.G.M., Allen, R.G., Droogers, P., D'Urso, G., Steduto, P. (2007): Twenty-five years modeling irrigated and drained soils: State of the art. Agricultural Water Management 92, 111-125.

Beven, K., Binley, A. (1992): The future of distributed models. Model calibration and uncertainty prediction. Hydrol. Process. 6 (3), pp. 279–298. DOI: 10.1002/hyp.3360060305.

Biemans, H., Speelman, L.H., Ludwig, F., Moors, E.J., Wiltshire, A.J., Kumar, P., Gerten, D., Kabat, P. (2013): Future water resources for food production in five South Asian river basins and potential for adaptation — A modeling study. Science of the Total Environment 468–469, S117–S131.

Birkel, C., Soulsby, C. (2015): Advancing tracer-aided rainfall-runoff modelling. A review of progress, problems and unrealised potential. Hydrol. Process. 29 (25), pp. 5227–5240. DOI: 10.1002/hyp.10594.

Birkel, C., Soulsby, C., Tetzlaff, D. (2015): Conceptual modelling to assess how the interplay of hydrological connectivity, catchment storage and tracer dynamics controls nonstationary water age estimates. In Hydrol. Process. 29 (13), pp. 2956–2969. DOI: 10.1002/hyp.10414.

Bogner, K., Liechti, K., Zappa, M. (2017): Technical note. Combining quantile forecasts and predictive distributions of streamflows. Hydrol. Earth Syst. Sci. 21 (11), 5493–5502. DOI: 10.5194/hess-21-5493-2017.

Bracken, L.J., Croke, J. (2007): The concept of hydrological connectivity and its contribution to understanding runoff-dominated geomorphic systems. Hydrol. Process. 21 (13), pp. 1749–1763. DOI: 10.1002/hyp.6313.

Chen, Y., Li, W., Fang, G., Li, Z. (2017): Review article: Hydrological modeling in glacierized catchments of central Asia – status and challenges. Hydrol. Earth Syst. Sci., 21, 669–684.

Chen, Y., Marek, G.W., Marek, T.H., Brauer, D.K., Srinivasan, R. (2018): Improving SWAT auto-irrigation functions for simulating agricultural irrigation management using long-term lysimeter field data. Environmental Modelling & Software 99, 25-38.

Cloke, H., Pappenberger, F. (2009): Ensemble flood forecasting: A review. Journal of Hydrology 375, 613–626.

Collins, A.L., Pulley, S., Foster, I.D.L., Gellis, A., Porto, P., Horowitz, A.J. (2017): Sediment source fingerprinting as an aid to catchment management. A review of the current state of knowledge and a methodological decision-tree for end-users. Journal of Environmental Management 194, pp. 86–108. DOI: 10.1016/j.jenvman.2016.09.075.

Dechmi, F., Burguete, J., Skhiri, A. (2012): SWAT application in intensive irrigation systems: Model modification, calibration and validation. Journal of Hydrology 470-471, 227-238.

Dietrich, J., Funke, M. (2009): Integrated catchment modelling within a strategic planning and decision making process. Werra case study. Physics and Chemistry of the Earth, Parts A/B/C 34 (8-9), pp. 580–588. DOI: 10.1016/j.pce.2008.11.001.

Dietrich, J.; Trepte, S.; Wang, Y.; Schumann, A. H.; Voß, F.; Hesser, F. B.; Denhard, M. (2008): Combination of different types of ensembles for the adaptive simulation of probabilistic flood forecasts. Hindcasts for the Mulde 2002 extreme event. Nonlin. Processes Geophys. 15 (2), 275–286. DOI: 10.5194/npg-15-275-2008.

Dietrich, J.; Schumann, A. H.; Redetzky, M.; Walther, J.; Denhard, M.; Wang, Y. et al. (2009): Assessing uncertainties in flood forecasts for decision making. Prototype of an operational flood management system integrating ensemble predictions. Nat. Hazards Earth Syst. Sci. 9 (4), 1529–1540. DOI: 10.5194/nhess-9-1529-2009.

Dietrich, J., Uniyal, B., Themer, C. (2018): Evaluating expert based irrigation recommendations by the simulation of irrigation water demand in a humid investigation area in Northern Germany with SWAP. EGU General Assembly, 8 – 13 April 2018, Vienna, Geophysical Research Abstracts Vol. 20, EGU2018-16196-1.

Diks, C.G.H., Vrugt, J.A. (2010): Comparison of point forecast accuracy of model averaging methods in hydrologic applications. Stoch Environ Res Risk Assess 24 (6), 809–820. DOI: 10.1007/s00477-010-0378-z.

Diomede, T.; Davolio, S.; Marsigli, C.; Miglietta, M. M.; Moscatello, A.; Papetti, P. et al. (2008): Discharge prediction based on multi-model precipitation forecasts. Meteorol Atmos Phys 101 (3-4), 245–265. DOI: 10.1007/s00703-007-0285-0.

Döll, P., Siebert, S. (2002): Global modeling of irrigation water requirements. Water Resources Research 38(4), 1037.

Doycheva, K., Horn, G., Koch, C., Schumann, A., König, M. (2017): Assessment and weighting of meteorological ensemble forecast members based on supervised machine learning with application to runoff simulations and flood warning. Advanced Engineering Informatics 33, 427–439. DOI: 10.1016/j.aei.2016.11.001.

Drewry, J.J., Newham, L.T.H., Greene, R.S.B., Jakeman, A.J., Croke, B.F.W. (2006): A review of nitrogen and phosphorus export to waterways. Context for catchment modelling. Mar. Freshwater Res. 57 (8), p. 757. DOI: 10.1071/MF05166.

Droogers, P., Bastiaanssen, W.G.M, Beyazgül, M., Kayam, Y., Kite, G.W., Murray-Rust, H. (2000): Distributed agro-hydrological modeling of an irrigation system in western Turkey. Agricultural Water Management 43 (2), pp. 183–202. DOI: 10.1016/S0378-3774(99)00055-4.

Droogers, P., Kite, G. (2001): Simuation Modeling at Different Scales to Evaluate the Productivity of Water. Phys. Chem. Earth 26(11-12), 877-880.

Ercin, A., Hoekstra, A.Y. (2014): Water footprint scenarios for 2050: a global analysis. Environment international 64, pp. 71–82. DOI: 10.1016/j.envint.2013.11.019.

Falconi, S.M., Palmer, R.N. (2017): An interdisciplinary framework for participatory modeling design and evaluation – What makes models effective participatory decision tools? Water Resour. Res., 53, 1625–1645, doi:10.1002/2016WR019373.

Fenicia, F., Kavetski, D., Savenije, H.H.G., Pfister, L. (2016): From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions. Water Resour. Res., 52, 954–989, doi:10.1002/2015WR017398.

Gampe, D., Ludwig, R., Qahman, K., Afifi, S. (2016): Applying the Triangle Method for the parameterization of irrigated areas as input for spatially distributed hydrological modeling - Assessing future drought risk in the Gaza Strip (Palestine). The Science of the total environment 543 (Pt B), pp. 877–888. DOI: 10.1016/j.scitotenv.2015.07.098.

Gassman, P.W., Reyes, M.R., Green, C.H., Arnold, J.G. (2007): The Soil and Water Assessment Tool. Historical Development, Applications, and Future Research Directions. Transactions of the ASABE 50 (4), pp. 1211–1250. DOI: 10.13031/2013.23637.

Georgakakos, K. P., Seo, D.-J., Gupta, H., Schaake, J., and Butts, M. B. (2004): Towards the characterization of streamflow simulation uncertainty through multimodel ensembles, Journal of Hydrology 298(1–4), 222–241.

Guse, B., Kail, J., Radinger, J., Schröder, M., Kiesel, J., Hering, D., Wolter, C., Fohrer, N. (2015): Eco-hydrologic model cascades: Simulating land use and climate change impacts on hydrology, hydraulics and habitats for fish and macroinvertebrates. Science of the Total Environment 533, 542-556.

Haddeland, I., Heinke, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N. et al. (2014): Global water resources affected by human interventions and climate change. Proceedings of the National Academy of Sciences of the United States of America 111 (9), pp. 3251–3256. DOI: 10.1073/pnas.1222475110.

Han, S., Coulibaly, P. (2017): Bayesian flood forecasting methods. A review. Journal of Hydrology 551, 340–351. DOI: 10.1016/j.jhydrol.2017.06.004.

Hölscher, J., Petry, U., Bertram, M., Anhalt, M., Schmidtke, S., Haberlandt, U., Müller, H., v.d.Heijden, S., Berndt, C., Verworn, A., Wallner, M., Belli A., Dietrich, J., Meon, G., Förster, K., Gelleszun, M., Riedel, G., Lange, A., Eggelsmann, F. (2012): Globaler Klimawandel – Wasserwirtschaftliche Folgenabschätzung für das Binnenland - Oberirdische Gewässer Band 33, Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz, Norden

Hoffman, R. N., Kalnay, E. (1983): Lagged average forecasting, an alternative to Monte Carlo forecasting, Tellus, 35A, 100–118.

Horn, A.L., Rueda, F.J., Hörmann, G., Fohrer, N. (2004): Implementing river water quality modelling issues in mesoscale watershed models for water policy demands––an overview on current concepts, deficits, and future tasks. Physics and Chemistry of the Earth, Parts A/B/C 29 (11-12), pp. 725–737. DOI: 10.1016/j.pce.2004.05.001.

Huang, Y., Li, Y.P., Chen, X., Ma, Y.G. (2012): Optimization of the irrigation water resources for agricultural sustainability in Tarim River Basin, China. Agricultural Water Management 107, pp. 74–85. DOI: 10.1016/j.agwat.2012.01.012.

Immerzeel, W.W., Gaur, A., Zwart, S.J. (2008): Integrating remote sensing and a process-based hydrological model to evaluate water use and productivity in a south Indian catchment. Agricultural Water Management 95 (1), pp. 11–24. DOI: 10.1016/j.agwat.2007.08.006.

Keesstra, S., Nunes, J., Novara, A., Finger, D., Avelar, D., Kalantari, Z., Cerdà, A. (2018): The superior effect of nature based solutions in land management for enhancing ecosystem services. Science of the Total Environment 610–611, 997–1009.

Kiesel, J., Guse, B., Pfannerstill, M., Kakoueia, K., Jähnig, S.C., Fohrer, N. (2017): Improving hydrological model optimization for riverine species. Ecological Indicators 80, 376-385.

Kite, G., Droogers, P. (1999): Irrigation Modelling in the Context of Basin Water Resources. International Journal of Water Resources Development 15 (1-2), pp. 43–54. DOI: 10.1080/07900629948925.

Kite, G. (2000): Manual for the SLURP hydrological model. International Irrigation Management Institute, Colombo, Sri Lanka.

Krzysztofowicz, R. (1999): Bayesian theory of probabilistic forecasting via deterministic hydrologic model. Water Resour. Res. 35 (9), 2739–2750. DOI: 10.1029/1999WR900099.

Leskens, J.G., Brugnach, M., Hoekstra, A.Y., Schuurmans, W. (2014): Why are decisions in flood disaster management so poorly supported by information from flood models? Environmental Modelling & Software 53, pp. 53–61. DOI: 10.1016/j.envsoft.2013.11.003.

Li, B., He, Y., Ren, L. (2018): Multisource hydrologic modeling uncertainty analysis using the IBUNE framework in a humid catchment. Stoch Environ Res Risk Assess 32 (1), 37–50. DOI: 10.1007/s00477-017-1424-x.

Limbourg, P., de Rocquigny, E. (2010): Uncertainty analysis using evidence theory – confronting level-1 and level-2 approaches with data availability and computational constraints. Reliability Engineering & System Safety 95 (5), 550–564. DOI: 10.1016/j.ress.2010.01.005.

Liu, J., Yang, H. (2010): Spatially explicit assessment of global consumptive water uses in cropland: Green and blue water. Journal of Hydrology 384 (3–4), pp. 187–197. DOI: 10.1016/j.jhydrol.2009.11.024.

Lohani, V.K., Refsgaard, J.C., Clausen, T., Erlich, M., Storm, B. (1993): Application of SHE for Irrigation-Command-Area Studies in India. J. Irrig. Drain Eng. 119 (1), pp. 34–49. DOI: 10.1061/(ASCE)0733-9437(1993)119:1(34).

Lorenz, E. (1969): The predictability of a flow which contains many scales of motion. Tellus A 21, 289–307.

Luan, X., Wu, P., Sun, S., Wang, Y., Gao, X. (2018): Quantitative study of the crop production water footprint using the SWAT model. Ecological Indicators 89, pp. 1–10. DOI: 10.1016/j.ecolind.2018.01.046.

Madadgar, S., Moradkhani, H. (2014): Improved Bayesian multimodeling. Integration of copulas and Bayesian model averaging. Water Resour. Res. 50 (12), 9586–9603. DOI: 10.1002/2014WR015965.

Maier, N., Dietrich, J. (2016): Using SWAT for Strategic Planning of Basin Scale Irrigation Control Policies. A Case Study from a Humid Region in Northern Germany. Water Resour Manage 30 (9), pp. 3285–3298. DOI: 10.1007/s11269-016-1348-0.

Mamede, G.L., Guentner, A., Medeiros, P.H.A., de Araújo, J.C., Bronstert, A. (2018): Modeling the effects of multiple reservoirs on water and sediment dynamics in a semiarid catchment in Brazil. Journal of Hydrologic Engineering (accepted)

Mediero, L., Garrote, L., Martin-Carrasco, F. (2007): A probabilistic model to support reservoir operation decisions during flash floods. Hydrological Sciences Journal 52 (3), 523–537. DOI: 10.1623/hysj.52.3.523.

Mehta, V.K., van Haden, R., Joyce, B.A., Purkey, D.R., Jackson, L.E. (2013): Irrigation demand and supply, given projections of climate and land-use change, in Yolo County, California. Agricultural Water Management 117, pp. 70–82. DOI: 10.1016/j.agwat.2012.10.021.

Mishra, A., Singh, R., Raghuwanshi, N.S. (2005): Development and Application of an Integrated Optimization-Simulation Model for Major Irrigation Projects. J. Irrig. Drain Eng. 131 (6), pp. 504–513. DOI: 10.1061/(ASCE)0733-9437(2005)131:6(504).

Molina, M., Blasco, G. (2004): A multi-agent system for emergency decision support. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 2690.

Nguyen, T., Dietrich, J. (2018): Hydrological modeling in non-conservative catchments in karst-dominated regions with the modified Soil and Water Assessment Tool. EGU General Assembly, 8 – 13 April 2018, Vienna, Geophysical Research Abstracts Vol. 20, EGU2018-6885-1.

Palmer, T. (2000): Predicting uncertainty in forecasts of weather and climate, Rep. Prog. Phys., 63, 71–116.

Philipp, A., Kerl, F. (2018): Hochwasserfrühwarnung für kleine Einzugsgebiete. Möglichkeiten und Grenzen im Lichte operationeller Anforderungen am Beispiel Sachsens. Sächsisches Landesamt für Umwelt, Landwirtschaft und Geologie, Dresden. https://publikationen.sachsen.de/bdb/artikel/30155 (Zugriff am 30.05.2018).

Pianosi, F., Beven, K., Freer, J., Hall, J,W., Rougier, J., Stephenson, D.B., Wagener, T. (2016): Sensitivity analysis of environmental models. A systematic review with practical workflow. Environmental Modelling & Software 79, pp. 214–232. DOI: 10.1016/j.envsoft.2016.02.008.

Raftery, A. E., Gneiting, T., Balabdaoui, F., Polakowski, M. (2005): Using Bayesian Model Averaging to Calibrate Forecast Ensembles. Mon. Wea. Rev. 133 (5), 1155–1174. DOI: 10.1175/MWR2906.1.

Raskin, P., Hansen, E., Zhu, Z., Stavisky, D. (2009): Simulation of Water Supply and Demand in the Aral Sea Region. Water International 17 (2), pp. 55–67. DOI: 10.1080/02508069208686127.

Reggiani, P.; Weerts, A. H. (2008): A Bayesian approach to decision-making under uncertainty. An application to real-time forecasting in the river Rhine. Journal of Hydrology 356 (1-2), 56–69. DOI: 10.1016/j.jhydrol.2008.03.027.

Rodriguez-Iturbe, I. (2000): Ecohydrology. A hydrologic perspective of climate-soil-vegetation dynamies. Water Resour. Res. 36 (1), pp. 3–9. DOI: 10.1029/1999WR900210.

Sivapalan, M., Savenije, H.H.G., Blöschl, G. (2012): Socio-hydrology: A new science of people and water. Hydrological Processes 26(8), 1270-1276.

Theis, S., Gebhardt, C. & Bouallègue, Z.B. (2012): Beschreibung des COSMO-DE-EPS und seiner Ausgabe in die Datenbanken des DWD, Version 1.0. Offenbach: Deutscher Wetterdienst.

Thielen, J., Bartholmes, J., Ramos, M.-H., and de Roo, A. (2009): The European Flood Alert System – Part 1: Concept and development, Hydrol. Earth Syst. Sci., 13, 125–140, doi:10.5194/hess-13-125-2009.

Udias, A., Pastori, M., Malago, A., Vigiak, O., Nikolaidis, N.P., Bouraoui, F. (2018): Identifying efficient agricultural irrigation strategies in Crete. The Science of the total environment 633, pp. 271–284. DOI: 10.1016/j.scitotenv.2018.03.152.

Uniyal, B., Dietrich, J. (2018): Simulating Regional PlantWater Availability and Irrigation Demand using SWAT in Different Agro-Climatic Zones. EGU General Assembly, 8 – 13 April 2018, Vienna, Geophysical Research Abstracts Vol. 20, EGU2018-7860.

Uribe-C., H., Arnold, T., Arumí, J., Berger, T., Rivera, D. (2009): Modification of the hydrological model WaSiM-ETH to improve its application in irrigated areas [Modificacion del modelo hidrológico WaSiM-ETH para mejorar su aplicación en áreas regadas]. Ingenieria Hidraulica en Mexico 24 (2), pp. 23–36.

van Dam, J.C., Huygen, J., Wesseling, J.G., Feddes, R.A., Kabat, P., van Walsum, P.E.V., Groenendijk, P., van Diepen, C.A. (1997): Theory of SWAP version 2.0, Technical Document 45, Wageningen Agricultural University.

van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., Srinivasan, R. (2006): A global sensitivity analysis tool for the parameters of multi-variable catchment models. Journal of Hydrology 324 (1-4), pp. 10–23. DOI: 10.1016/j.jhydrol.2005.09.008.

Varela-Ortega, C., Blanco-Gutiérrez, I., Esteve, P., Bharwani, S., Fronzek, S., Downing, T.E. (2016): How can irrigated agriculture adapt to climate change? Insights from the Guadiana Basin in Spain. Reg Environ Change 16 (1), pp. 59–70. DOI: 10.1007/s10113-014-0720-y.

Wallner, M., Haberlandt, U., Dietrich, J. (2013): A one-step similarity approach for the regionalization of hydrological model parameters based on Self-Organizing Maps. Journal of Hydrology 494, pp. 59–71. DOI: 10.1016/j.jhydrol.2013.04.022.

Wetterhall, F.; Pappenberger, F.; Alfieri, L.; Cloke, H. L.; Thielen-del Pozo, J.; Balabanova, S. et al. (2013): HESS Opinions "Forecaster priorities for improving probabilistic flood forecasts". Hydrol. Earth Syst. Sci. 17 (11), 4389–4399. DOI: 10.5194/hess-17-4389-2013.

Wisser, D., Frolking, S., Douglas, E.M., Fekete, B.M., Vörösmarty, C.J., Schumann, A.H. (2008): Global irrigation water demand: Variability and uncertainties arising from agricultural and climate data sets. Geophysical Research Letters 35, L24408.

WMO (2015): Seamless Prediction of the Earth System: from Minutes to Months. World Meteorological Organization, Genf, https://library.wmo.int/pmb_ged/wmo_1156_en.pdf (Zugriff 18.4.2018).

Wriedt, G., van der Velde, M., Aloe, A., Bouraoui, F. (2008): Water Requirements for Irrigation in the European Union. A model based assessment of irrigation water requirements and regional water demands in Europe. European Commission Joint Research Centre, Institute for Environment and Sustainability.

Xie, X., Cui, Y. (2011): Development and test of SWAT for modeling hydrological processes in irrigation districts with paddy rice. In Journal of Hydrology 396 (1-2), pp. 61–71. DOI: 10.1016/j.jhydrol.2010.10.032.

Zappa, M., Rotach, M.W., Arpagaus, M., Dorninger, M., Hegg, C., Montani, A. et al. (2008): MAP D-PHASE. Real-time demonstration of hydrological ensemble prediction systems. Atmos. Sci. Lett. 9 (2), 80–87. DOI: 10.1002/asl.183.

Zappa, M., Jaun, S., Germann, U., Walser, A., Fundel, F. (2011): Superposition of three sources of uncertainties in operational flood forecasting chains. In Atmospheric Research 100 (2-3), pp. 246–262. DOI: 10.1016/j.atmosres.2010.12.005.

Herausgegeben im Selbstverlag
des Institutes für Hydrologie und Wasserwirtschaft
Gottfried Wilhelm Leibniz Universität Hannover

Appelstraße 9a; D-30167 Hannover
Tel.: 0511/762-2237
Fax: 0511/762-3731
E-Mail: info@iww.uni-hannover.de

2019