Institut für Hydrologie und Wasserwirtschaft

Mitteilungen Heft 111

Improving the rainfall nowcast for fine temporal and spatial scales
suitable for urban hydrology.

IMPROVING THE RAINFALL NOWCAST FOR FINE TEMPORAL AND SPATIAL SCALES SUITABLE FOR URBAN HYDROLOGY

Von der Fakultät für Bauingenieurwesen und Geodäsie
der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des akademischen Grades

Doktor-Ingenieurin

- Dr.-Ing. -

genehmigte Dissertation

von

Bora Shehu, M.Sc.
geboren am 14. March 1990 in Lezhe (Albanien)

Hannover, May 2020

Acknowledgements

First and foremost, I would like to express my gratitude to my doctor thesis supervisor Prof. Haberlandt, for his mentoring, motivation and patience. Thank you for supporting me during this study, for giving me the freedom to develop new ideas, but most importantly for directing me at the right direction.

Secondly, I would like to thanks my colleagues who have made this journey quite enjoyable. The doctoral thesis has its' highs and lows, but we were through this together: not only did we share a nice time but you were there always for me. Particularly, I'm thankful to Luisa, who was there to help me not only with this manuscript, but with so many other things during my PhD. Moreover, I'm thankful to Ana, who introduced me to the research and who motivated me to be the better version of myself. In addition to my colleagues in the institute, I would like to thank my colleagues from the EVUS project. We were such a nice team, and we had so many great discussions together.

Thank you to Ross Pidoto, Anne Fangmann, Ela Merkohitaj, Carolina Aguirre, Christian Berndt and Ana Callau for the valuable help with this manuscript. Particularly I'm thankful to Malkin Gerchow, who introduced me to the fascinating world of neural networks and who built the basis of the CNN code used in this study.

I am grateful for my circle of friends, that are like my second family here in Germany; thank you for having my back! Special thanks for my partner Michael, who has seen me go through some dark times in finalizing this thesis, but who was always there, taking care of me and supporting me at each moment.

Lastly, my outermost gratitude goes for my parents and my sister. Thank you for believing in me, for pushing me even further than I thought I could and for giving me the great gift of freedom. You have loved me despite of my choices, and despite of the distance, we have walked this journey together.

Declaration

I, Bora Shehu, hereby declare that:

1. I know the regulations for the doctoral candidates at Leibniz University of Hannover and I have met all the requirements. I agree with an examination under the provisions of the doctoral regulations.

2. I have completed the thesis independently, and where I have consulted other works, to the best of my knowledge, this is always clearly attributed.

3. I have not paid any third party to contribute to the content of my dissertation. The scientific work has not been acquired or conveyed, either in part or in full, by a third party for payment or any other consideration.

4. This works contains no material which has been accepted for the award of any other degree or diploma at any other university or tertiary institution.

5. The same or partly similar work has not been submitted at any other university or academic institution.

6. I agree that my thesis can be used for verifying the compliance with scientific standards in particular using electronic data processing programs.

---------------------------   ------------------------

Date and Time   M.Sc. Bora Shehu

Abstract

Accurate Quantitative Precipitation Forecasts (QPF) at high spatial and temporal resolution are crucial for urban flood prediction. Typically, Lagrangian persistence based on radar data is used to nowcast rainfall intensities with up to 3 hours lead time, but nevertheless is not able to deliver reliable QPFs past 20 min lead time (known as well as the predictability limit). Especially, for extreme events causing pluvial floods, accurate QPFs cannot be achieved past 5 min lead time. Furthermore when compared to gauge recordings, the QPFs are not useful at all. There is an essential need to provide better QPFs by improving the rainfall field supplied to the nowcast and by employing non-linear processes for the extrapolation of rainfall into the future. This study is focused on these two main problems, and it investigates different geostatistical and data-driven methods for the improvement of the QPFs at fine scales.

The study was conducted within the Hannover radar range where observations between 2000 to 2018 were available. The skill of the nowcast models was assessed on the point (1 km^2 and 5 min) and storm scale, based on continuous criteria comparing both radar and gauge observations. A total of 100 gauge measurements inside the study area were as well employed for the assessment. From the period 2000-2012, 93 events of different properties were distinguished and used as a basis for the method development and assessment. Two state-of-the-art nowcast models (HyRaTrac and Lucas-Kanade) were chosen as reference and used as benchmarks for improvement.

To improve the rainfall field, a real time merging between radar and gauge data was investigated. Among different merging techniques (mean field bias, quantile bias correction and kriging interpolation), conditional merging (CM) yielded the best rainfall field. When fed to the reference nowcast models, it led to improvements of up to 1 hour of the predictability limit and of the agreement between radar based QPFs and gauge data. To improve the QPF accuracy even further, two different data driven techniques were developed in order to learn non-linear behaviours from past observed rainfall. First, a nearest neighbour approach (k-NN) was developed and employed instead of Lagrangian Persistence on the HyRaTrac nowcast model. The k-NN method accounts for the non-linearity of the storm evolution by consulting k-similar past storms. A deterministic nowcast issued by averaging the behaviours from the 3 most similar storms yielded the best results, extending the predictability limit at the storm scale to 2-3 hours. Second, an ensemble nowcast accounting for the 10 closest neighbours was generated in order to estimate the uncertainty of the QPF. Third, a deep convolution neural

network (CNN) was trained on past merged data, in order to learn the non-linearity of the rainfall process. The network based on the last 15 min of observed radar images proved to successfully capture death and decay and partly birth processes, and extended the rainfall predictability limit at the point scale to 3 hours.

Lastly, the methods were tested on 17 convective extreme events, extracted from the period 2013-2018, to compare the tested methods for an urban flood nowcast application. The CNN based on merged data outperformed both reference methods as well as the k-NN based nowcast, with the predictability limit reaching 30 – 40 min. The k-NN, although better than the Lagrangian persistence, suffered greatly from the shortcomings of the storm tracking algorithm present under fast moving and extreme storms. To conclude, even though clear improvements were achieved, there is a clear limit to the data-driven methods that cannot be overcome, unless coupled with the convection initialization from Numerical Weather Prediction (NWP) models. Nevertheless, complex relationships learned from past observed data, together with a better rainfall field as input, were proven to be useful in increasing the QPF accuracy and predictability limits for urban hydrology application.

Kurzfassung

Quantitative Niederschlagsvorhersagen (QPF) in hoher räumlicher und zeitlicher Auflösung sind entscheidend für die Prognose urbaner Sturzfluten. Der auf Radardaten basierende Lagrange Ansatz wird typischerweise für Regenintensitätsvorhersagen mit einem Horizont von 3 Stunden verwendet. Zuverlässig ist dieser allerdings nur bis 20 Minuten (bekanntes Prognoselimit). Bei extremen Niederschlagsereignissen, die urbane Sturzfluten verursachen, ist das Limit sogar bereits bei 5 Minuten erreicht. Außerdem kommt es zu deutlichen Abweichungen zwischen der QPF und den Messdaten an Niederschlagsstationen. Eine Verbesserung der QPF ist demnach zwingend erforderlich. Eine solche Verbesserung kann durch die Anpassung des Eingabe-Niederschlagsfeldes und durch die Anwendung nichtlinearer Prozesse für die Extrapolation des Niederschlags erreicht werden. Die vorliegende Studie konzentriert sich auf diese beiden Hauptprobleme und untersucht verschiedene geostatistische und Data-Mining Methoden zur Verbesserung der QPF auf solchen Skalen.

Die Studie wurde im Radarbereich von Hannover durchgeführt, wo Beobachtungsdaten von 2000 bis 2018 verfügbar sind. Die Güte der Nowcast-Modelle wurde auf der Punkteskala (1 km^2 und 5 min.) anhand kontinuierlicher Kriterien evaluiert und in Relation zu Radar- und Stationsbeobachtungen gesetzt. Hierfür wurden insgesamt 100 Stationsmessungen innerhalb des Untersuchungsgebietes verwendet. Aus dem Zeitraum 2000 bis 2012 wurden 93 Ereignisse mit unterschiedlichen Eigenschaften als Grundlage für die Methodenentwicklung und -beurteilung ausgewertet. Zwei gängige Nowcast-Modelle (HyRaTrac und Lucas-Kanade) wurden als Referenzmodelle ausgewählt und als Maßstab für Verbesserungen eingesetzt.

Um das Niederschlagsfeld zu verbessern, wurden Radar- und Stationsdaten in Echtzeit zusammengeführt. Unter den verschiedenen Methoden (Mean Field Bias, Quantile Mapping Bias, Kriging-Interpolation) ergab das Conditional Merging (CM) das optimalste Niederschlagsfeld. Als Input für die beiden Referenzmodelle verwendet, führte das CM zu einer Verlängerung des Prognoselimits auf bis zu eine Stunde. Auch die Übereinstimmung der radargestützten QPF mit den Stationsdaten verbesserte sich. Um das Prognoselimit noch weiter auszudehnen, wurden zwei verschiedene Data-Mining Techniken entwickelt, um die nichtlinearen Verhaltensweisen aus vergangenen Regenfällen zu erlernen: Zunächst wurde ein Nächster-Nachbar-Ansatz (k-NN) entwickelt und anstelle der Lagrange Persistenz im

HyRaTrac-Nowcast-Modell eingesetzt. Die k-NN-Methode berücksichtigt die Nichtlinearität der Regensturmentwicklung, indem k-ähnliche vergangene Stürme herangezogen werden. Ein deterministischer Nowcast, der durch Mittelwertbildung der Verhaltensweisen der drei ähnlichsten Stürme erstellt wurde, lieferte die besten Ergebnisse und verlängerte das Prognoselimit auf bis zu zwei-drei Stunden. Ein Ensemble-Nowcast, bei dem die zehn nächsten Nachbarn berücksichtigt wurden, wurde ebenfalls erstellt, um die Unsicherheit des QPF abzuschätzen. Zudem wurde ein künstliches neuronales Netz (CNN) basierend auf vergangenen Daten entwickelt, um die Nichtlinearität des Niederschlagsprozesses zu berücksichtigen. Das neuronale Netz, das mit den beobachteten Radarbildern der letzten 15 Minuten gefüttert wurde, erwies sich als erfolgreich in der Erfassung von Todes-, Zerfalls- und Geburtsprozessen von Stürmen und konnte das Prognoselimit auf bis zu drei Stunden erweitern.

Um die Wirksamkeit der entwickelten Methoden für die Vorhersage urbaner Sturzfluten zu untersuchen, wurden sie auf 17 konvektive Extremereignisse aus dem Zeitraum 2013 bis 2018 angewendet. Der k-NN Ansatz war zwar besser als die Lagrange Persistenz, litt aber stark unter den Fehlern des Sturmverfolgungs-Algorithmus bei schnellen und extremen Stürmen. Das CNN übertraf sowohl die Referenzmethoden als auch den k-NN-basierten Nowcast. Das Prognoselimit konnte so von 5 auf 30 bis 40 Minuten erweitert werden. Für eine weitere Verbesserung zeichnete sich letztlich eine klare Grenze ab, die nur mit der Konvektionsinitialisierung aus Numerischen Wettervorhersagemodellen (NWP-Modellen) überwunden werden kann. Im Vergleich mit den ausgewählten Referenzmodellen, können, durch die hier entwickelten Methoden, die Genauigkeit und das Prognoselimit der QPF in der städtischen Hydrologie erheblich verbessert werden.

Schlagwörter: Niederschlagsvorhersage, Prognoselimit, Data-Mining Methoden, Merging-Methoden, urbane Sturzfluten

Table of Contents

Table of Figures

Index of Tables

Abbreviations

ANN	Artificial neural network
BGR	German Federal Agency for Cartography and Geodesy
BM	Bayesian merging
BPNN	Back propagation neural network
CDF	Cumulative probability distribution
CM	Conditional merging
CNN	Convolutional neural network
CoK	Co-Kriging
Cosmo-DE	DWD numerical weather prediction model
CPU	Computational processing units
DARTS	The dynamic and adaptive radar tracking of storms
DispNet	Disparity learning with convolutional network
DSD	Dry spell duration
DWD	German Weather Service
E-TITAN	Enhanced TITAN nowcast model
ELU	Exponential linear unit
ETS	Equitable threat score
FAR	False alarm ratio
Flownet	Learning optical flow with convolutional networks
GEV	Generalized extreme value distribution
GPU	Graphical processing units
GSLIB	Geostatistics Library in Fortran
GSTAT	Geostatistics Library in R
HAR	Hit alarm ratio
HS	Horn and Schunck optical field
HyRaTrac	Hydrological radar tracking algorithm
IDK	Indicator kriging
K	Differential phase shift in dual-polarized radar data

KED	Kriging with external drift
k-NN	K-nearest neighbour method
LK	Lucas and Kanade optical flow
lmomco	Statistics library in R
LT	Lead time
MetOffice	UK Meteorological Office
MFB	Mean field bias correction
NSE	Nash-Sutcliffe criteria
NWP	Numerical weather prediction models
OK	Ordinary kriging
PI	Rainfall Intensity
PIC	Partial information content
pysteps	Implementation of STEPS in python
QPF	Quantitative precipitation forecast
QQ	Quantile mapping based bias correction
RIA	Relative importance analysis
ReLU	Rectified linear unit activation function
RMSE	Root mean squared error
ROC	Receiver operating characteristic curve
RR	Raw radar data
SSIM	Structural similarity index
STEPS	Short-term ensemble prediction system
Ta	Return period of a phenomena
TINT	The 2D implementation of TITAN
TITAN	Thunderstorm identification, tracking, analysis and nowcasting
TREC	Tracking radar echoes by correlation
TS	Time step of 5 minutes
WAR	Wet area ratio in a radar image
WSA	Wet spell amount
WSD	Wet spell duration
WSI	Wet spell intensity
Z	Reflectivity measured by radar

1. INTRODUCTION

Urban floods are caused mainly by convective rainfall events that develop quite rapidly and are characterized by short lived and intense rainfall. Recently, there is a growing concern towards these floods as on one hand the rainfall extreme events are becoming more intense, short and frequent [Einfalt et al., 2009][van Dijk et al., 2013], and on the other hand the urban areas are expanding due to population growth [Jacobson, 2011][Worldbank, 2019]. As the infiltration capacity in cities decreases, and more water is coming in at shorter duration, the rainfall-runoff process becomes much faster, leading to large inundation areas. These types of floods, typically called pluvial floods, may happen everywhere, and are expected to increase in the future [Lau et al., 2010]. Examples of such floods are for instance a) Dortmund (Germany) July 2008 with a rainfall volume of 200mm over 3 hours (return period Ta was over 100 years) and a damage of 17.2 million Euro [Grünewald, 2009], b) Hull (UK) June 2007 with rainfall volume of 100 mm in 24 hours (return period Ta ~ 150 years) that damaged 8600 residential buildings and 1300 businesses [Coulthard and Frostick, 2010], Tokyo (Japan) August 2008 with more than 50 mm within 30 min that caused 5 casualties [Kato and Maki, 2009]. In order to decrease human and economic losses associated with these floods, forecasting and modelling are very important.

To model the dynamics of the pluvial floods, urban models are built. Given the rainfall as an input, they are able to predict the surface water levels and surcharge flows [van Dijk et al., 2013][Russo et al., 2015]. Since these models simulate processes that are fast (minutes to a few hours) and at specific locations (at each street or manhole), rainfall intensities at fine scales (1-5 min and $100m^2$–$1km^2$) are a perquisite [Berne et al., 2004][Einfalt, 2005][Ochoa-Rodriguez et al., 2015]. Therefore, accurate and fine resolution rainfall forecasts (referred to as Quantitative Precipitation Forecast – QPF) are necessary to model urban floods and to issue warnings. Nevertheless, issuing accurate QPF is quite challenging due to the erratic nature of the rainfall itself at such scales.

The large and synoptic scale, can be modelled and predicted well enough by Numerical Weather Prediction (NWP) Models for days and even weeks ahead, and produce reliable rainfall intensities at spatial scales larger than $10km^2$ and accumulations longer than an hour [Golding, 2009][Surcel et al., 2015]. For finer scales, however, these models are not able to predict the exact location and the sub-hourly peaks of the rainfall intensities, and hence are

not suitable for urban flood application [Kato et al., 2017][Šálek et al., 2006][Surcel et al., 2015]. Radar data on the other hand can capture spatio-temporal structures of rainfall intensities at the desired fine scales (5min and 1km^2) and give the basis for the identification and forecasting of the rainfall storms for forecast horizons (referred here as lead times) up to two hours (referred here as nowcasting) [Berenguer et al., 2012][Jensen et al., 2015][Lin et al., 2005][Zahraei et al., 2012]. Even though the radar can deliver fine resolution rainfall nowcast, it can only deliver useful results up to 20 minutes (at maximum 25 min) lead time. This is mainly for two reasons:

First, the radar measures the rainfall intensity indirectly by capturing the reflected energy from the water droplets in the atmosphere, and as such is prone to numerous sources of errors. When compared to gauge data (which are typically considered the ground truth and used for urban model calibration), the radar data show a systematic underestimation of rainfall intensities. Second, the radar based nowcast usually employs a Lagrangian persistence to extrapolate the rainfall fields into the future. Once a rainfall pattern (storm) is recognized, the Lagrangian persistence considers its movement as constant in time, with no changes at its inner structure (the intensities are kept constant). Thus, such persistence is not able to predict the birth (predicting the storm initiation before it has been observed), the death (predicting when the storm has stopped) and the growth or decay (predicting how the storm is changing between birth and death). Consequently, the nowcast is limited to lead times shorter than the average lifetime of the rainfall storms.

Attempts have been made to improve radar based nowcast by blending the Lagrangian persistence with a NWP (in order to capture better the storm birth processes) and with a stochastic generator (in order to model better the growth and decay processes). A famous example of such algorithm, is the STEPS nowcast used at the UK MetOffice [Bowler et al., 2006]. The performance of STEPS was also investigated for urban flood forecast application, and even though it improved the nowcast skill compared to the Lagrangian persistence, it could not yield useful nowcasts for lead times longer than 30 minutes [Foresti et al., 2016].

Ideally a one- or two-hour duration rainfall nowcast is required as an input for the urban model (as typically the concentration times in an urban catchment are one to two hours). The current radar based nowcasts cannot satisfy this condition, as apparently there is limit (20-30 minutes) that the Lagrangian persistence and the stochastic noise generator cannot overcome (also referred here as the predictability limit). Obviously, there is a need to deliver better radar based nowcasts by either improving the radar rainfall field or by including other methods that account for the erratic and intermittent nature of the rainfall process. In this thesis, this is achieved by studying different methods in the Hannover Radar region in Lower Saxony, Germany.

First, different correction procedures based on gauge observations are applied on the radar data in order to achieve a better rainfall field as input for the nowcast. The assumption here is that a better rainfall field will yield better nowcasts in terms of higher predictability

limit and better agreement with the gauge observations. Here it is investigated for the method that is more reliable when applied online and that brings the biggest additional value to the nowcast. Of course, the selected method should yield better results than the uncorrected radar data or the interpolation of the gauge data.

Second, two data-driven methods are developed and applied to the radar data in order to understand and to improve the predictability limits at fine scales. One method is the Nearest Neighbour (k-NN) method that aims to identify storms that are behaving the same, and to nowcast based on similar observed patterns in the past. The assumption here is that similar storms behave similarly and display similar growth/decay rates and death times. Consequently, it is important to investigate what are the right predictors in recognizing similar storms. The other method is a deep Convolution Neural Network (CNN) trained independently for each lead time. As neural networks have gained popularity at the recent years, due to their ability to learn complex relationships, it is important to investigate their potential at improving the rainfall nowcast. The assumption here is that certain growth, decay and death patterns can be learned from past observed storms and can be applied to future storms.

Third, since the sources of uncertainty are many in a radar based nowcast (the input merged field, the choice of the nowcast model etc.) and the models at the moment are not able to predict the exact location and intensity of the storm birth, it is important to issue probabilistic nowcast (ensembles) instead of a deterministic one. The assumption here is that a set of ensembles nowcasts can provide better prediction, as in addition to the nowcasted rainfall intensities, they provide useful information about the uncertainty of such intensities.

Lastly, as the ultimate objective is the pluvial flood forecast, the rainfall nowcast should be able to give reliable predictions for extreme rainfall events that can potentially cause such floods. For this purpose, the developed nowcast methods are tested under extreme convective events, to prove an improvement of the predictability limit even for these types of events.

1.1 Objectives of the thesis

The three main problems for rainfall nowcasting that the thesis is focused on are:

1. identifying the 'true' rainfall field – accurate rainfall intensities at fine temporal and spatial resolution ($1km^2$ and 5min) which will be an input for the nowcast model.

2. improving the predictability of rainfall mostly of convective events at such fine scales to more than 30 min, by including storm decay, birth, growth and death.

3. tackling the problem of prediction uncertainty and quantifying it, by considering an ensemble approach.

Based on these three problems, the main objectives of this work are as follow:

1. deriving a more accurate field of spatially distributed rainfall suitable for operational use in the urban flood forecast by using different merging techniques between radar and gauge data

2. developing a new nowcast method based on data-driven techniques that would include the non-stationary nature of the rainfall, and lead to improved performance and predictability towards the existing reference methods.

3. integrate ensemble prediction on the new nowcast method in order to consider different sources of uncertainty in the data and the model, and to cover a wide range of possible nowcasts for given extreme events.

1.2 Research Questions

The main research questions that this thesis tries to answer are as follows:

1. What is the best way to describe predictability of storms and predictability for the urban flood application?

2. What are the gaps of the current methods used for nowcasting, that could be improved with data-driven method?

3. How can merging of radar and gauge rainfall information be optimized in order to have an accurate and high-resolution rainfall field?

4. Can a better rainfall input field improve the predictability of storms?

5. Do similar storms behave similarly? And if so what would be the best way or features to identify similar storms, that exhibit similar relationship?

6. Can data-driven methods be useful in improving the actual rainfall predictability?

7. Can ensemble nowcast yield better and more reliable nowcast than a deterministic one?

8. Can the final product yield reliable results even for extreme convective events that can cause urban floods?

1.3 Organization of the thesis

An overview of target objectives (blue boxes), research questions (bold writing) and approaches to reach them (red arrows) is given in the following Figure 1.1. The structure of the thesis and their content is indicated by the green boxes.
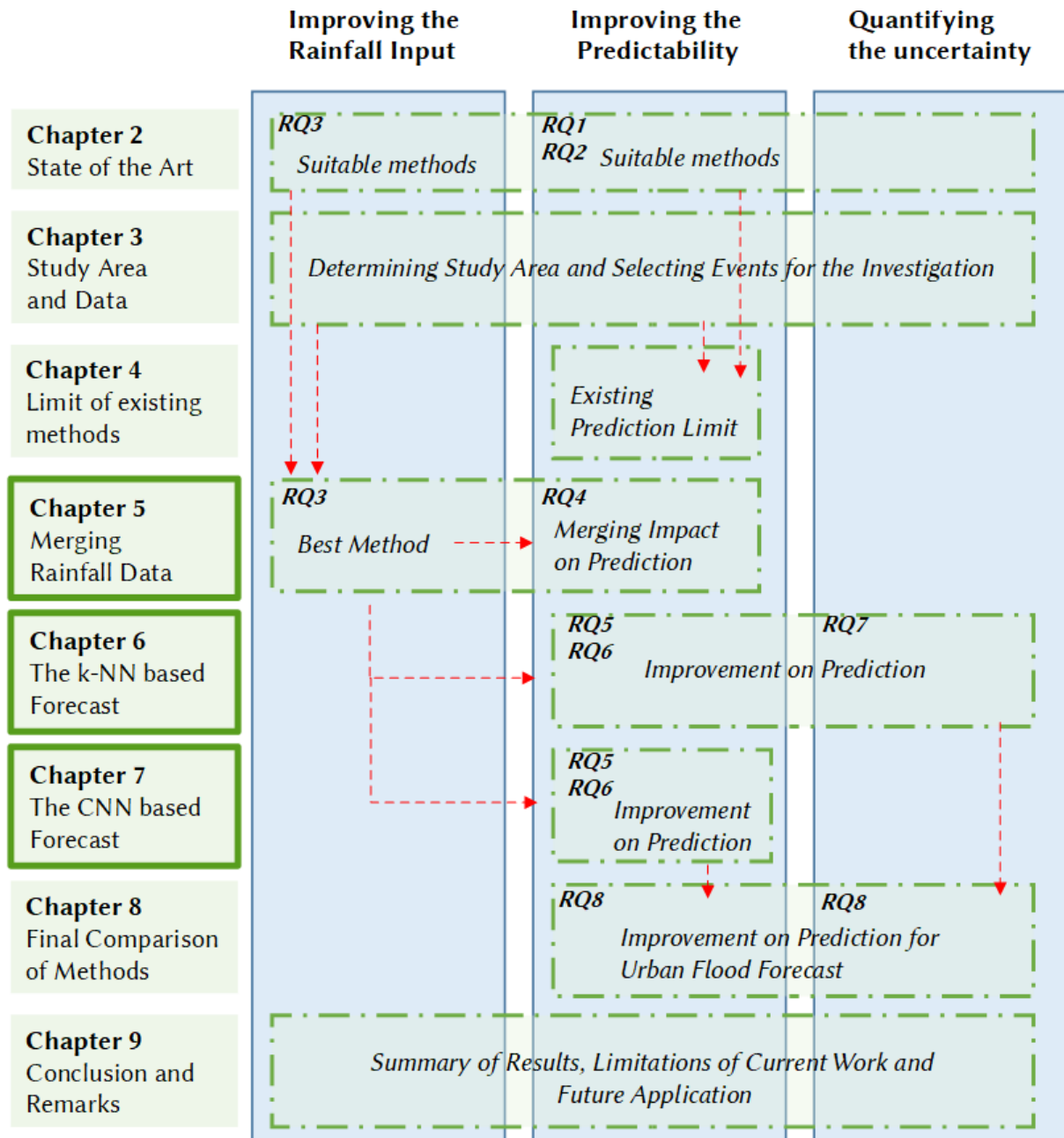
Figure 1.1: *The overall organization of the thesis: chapters (in green) and their focus on the three main researches objectives (the blue columns). The chapters where the innovative methods are explained and employed are illustrated with a bold green box.*

First, some basic background information about the topic and the selection of the reference methods are given in Chapter 2. The study area, the available data and the selection of relevant events are described in Chapter 3. The predictability limits of the reference nowcast methods under the selected events are drawn in Chapter 4 and form the benchmark for improvement. Three innovative methods are employed to improve these limits and are explained respectively in Chapter 5, 6, and 7. Chapter 5 focuses on the improvement of the rainfall field as an input for the nowcast, and the impact it has on the predictability limits. The obtained optimal rainfall field is then used as basis for the development of the two new nowcast models: the nearest neighbour method (described in Chapter 6) and the convolutional neural network (described in Chapter 7). Finally, a comparison among the developed and reference methods under extreme convective conditions is performed in Chapter 8, in order to investigate the improvement of predictability limits for events causing pluvial floods. The thesis is closed with conclusion, remarks and future works in Chapter 9.

2. STATE OF THE ART

2.1 Hydro-Meteorological Processes

2.1.1 Meteorological Rainfall Types and Features

Precipitation is one of the most important components of the water cycle on earth; it has large impacts on the human society and the ecosystem. As the main driver for runoff, it affects greatly the water supply, the surface hydrology, and can lead to rather severe disasters causing human, economic and ecological damages. Due to its high variability in time and space, it has been and it is essential to understand the processes governing its formation. The main field that studies the formation of precipitation is the cloud physics, which summarizes the equations and the physical processes that describe the occurrence of different droplet types, based on the medium (water, ice or mixed clouds) at the micro scale[1].

For specific conditions of the weather systems, the precipitation cloud systems are organized and structured, according to the dominance of two mechanisms: convective and stratiform movements. These two mechanisms exhibit different types of droplets and rain formation that are initiated by air motions with different magnitudes, and consequently are characterized by different rainfall duration and intensities. The convective precipitation is caused by strong local vertical air motions (for instance created by strong solar radiation) that can initiate in a very short time (in the range of minutes) the formation of big droplets at the base of the cloud formation. On the other hand, the stratiform precipitation is developed usually under a longer time (in the range of hours or even days), as the vertical air motion is very weak, causing small droplets to be formed at the top of the cloud formation. At the presence of mountains, at both stratiform or convective structures, a vertical movement of the air masses can occur due to the orographic lifting. Based on the current state of the weather and the governing mechanism, three major classes of precipitation storm types are distinguished in the mid-latitude of the northern hemisphere, according to [Maidment, 1993], whose properties are discussed below and summarized in Table 2.1.

---

1   Scale is defined here as the attribute of a process that describes the spatial or temporal extension of it. Regarding the atmospheric scale of the rainfall formation we distinguish: micro-scale ($<1km^2$), meso-γ scale (10^2–10^4km^2), meso-β scale (10^4–10^5km^2) and macro-α scale ($>10^8km^2$).

a. Extratropical Cyclones

The synoptic systems can either be cyclones (low air pressure) or anticyclones (high air pressure) caused by well-established and stable air flow patterns around the earth, mainly in regions outside of the tropics. While the anticyclones are associated with a dry period, the cyclones are associated with precipitation formation caused by the encounter of two air masses with different temperature and humidity (depending on the region of their origin). As warmer air has lower density compared to the cold one, it rises above, forming fronts of precipitation that persist[2] temporally for several hours and spatially from tens to hundreds of kilometres, and that are caused by a strong gradient in temperature, pressure and humidity. Typically, the vertical lifting in such systems is lower than 0.1 km/h, thus the predominant mechanism is the stratiform type [Maidment, 1993].

The created front can either be *a cold front* – cold air moves underneath a slower warm air mass, or a *warm front* – warm air moves above a slower colder air mass. Warmer fronts typically last longer (more than 24 hours) and are wide spread, with no specific storm shapes and with low intensities that rarely exceed 20 mm/h. The cold fronts last only for some hours, and form narrow, stripe shaped storms that move rather uniformly and rapidly, accompanied by intensities higher than the warmer front but hardly exceeding 50-60mm/h [Ehret, 2003]. While the synoptic systems occur at the macro-α scale, the fronts occur at the meso-β scale. A typical formation of the cold and warm front is illustrated in Figure 2.1.

a)   b)   c)
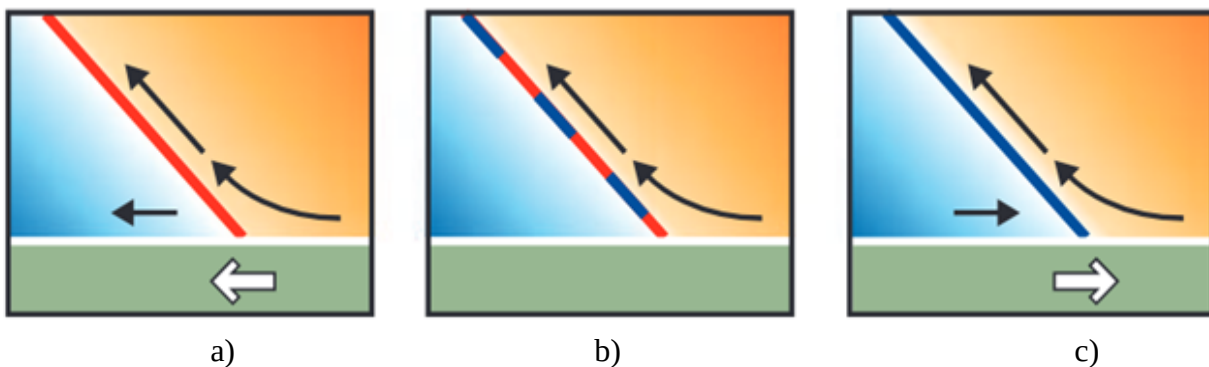
Figure 2.1: *Illustration of the air motions when fronts occur: the warm air is shown with red, the cold with blue, the major movement of the front is indicated by the thick white arrow, while the air masses movements by the black arrows. Here can be distinguished a) warm front as the warm air rises above the cold air, b) stationary front – with no clear frontal movement, and c) cold front – as cold air moves underneath the warm air [Wallace and Hobbs, 1977].*

b. Mid-latitude Thunderstorms

Thunderstorms are created by the convective mechanisms, in air masses with uniform temperature, low-level moisture and little wind shear, which are destabilized by strong radiation, causing a strong upward convective current. Warm and moisture air, reaches the

---

2   Persistence is referred here as the duration of a process, which continues to exists with little or no change in the features that describe it.

height of the troposphere rather rapidly, causing fast and short-lived rainfall storms. Figure Error: Reference source not found illustrates the formation of such a storm. Usually the convective storms happen at the meso-γ scale (10^2–10^4km²) and are characterized by high rainfall intensities (typically from 100 mm/h up to 400 mm/h) that are limited in space, thus exhibit a great spatial variability. Such storms are associated with thick and high clouds, big rain droplets, and at some cases also with hail formation. The convective storms are characterized by oval shapes that develop very erratically both in space and time. Under very dynamic atmospheric conditions (with sufficient humidity), super cells may be formed and are more persistent that the normal convective ones and exhibit a more distinctive motion and longer lifetimes [Ehret, 2003][Maidment, 1993].

Organized thunderstorms happening at the meso-β (10^4–10^5km²) scale are mainly caused by distinctive horizontal air advection, and are classified in two categories: a) meso-scale convective systems – which exhibit circular and more regular storm shapes that can cover up to 100,000 km² and persist for several hours with a rather uniform advection movement, and b) squall lines – which are often created together with a cold front, either in front of it or parallel to it. As the name suggests, these convective storms are organized in a linear form, characterized by intense, persistent and long rainfall. Both convective and stratiform mechanism are present in this type, and rainfall burst mainly at the gusts of the wind. These two types are of great importance as they are the major driver of hydrological floods (both in urban and rural catchments) in the mid latitudes [Ehret, 2003].
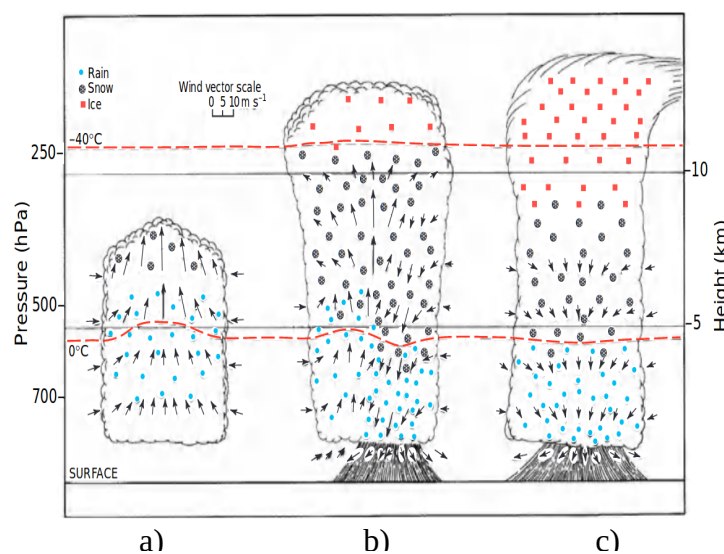
Figure 2.2: *Illustration of a convective storm formation stages a) cummulus, b) mature, and c) dissipation stage. The 0ºC and the –40ºC isotherms are indicated in red, while the horizontal scale is 30% of the vertical one [Wallace and Hobbs, 1977].*

c. Orographic Rainfall

The orography can influence any of the above-mentioned storms; as the warm and humid air is forced to move up a terrain obstacle (either a hill or a mountain), it produces convective instabilities, resulting in cooling or condensation of the moistured air that can enhance or

cause precipitation. The storm created or enhanced, is mainly falling around the obstacle (both on the leeward and windward side), and is characterized by different intensities, duration and persistence. Their importance depends on the terrain feature and the storm type [Maidment, 1993].

However, in this study the focus is mainly on two types of storms a) the fronts – which will be referred to as stratiform events and b) the thunderstorms – which will be referred as convective events. The orographic mechanism is left aside for three main reasons: i) the majority of storms are mainly caused either by convective or stratiform mechanisms and later emphasized by the orography; ii) since the creation of rainfall due to orography is accompanied by convective instabilities, in a sort of way they are convective events (and will be regarded as such), only that the main driver is not the strong solar radiation but instead the orographic force; iii) as the rainfall will be looked at the fine scale, the change due to the orography can be neglected. The following table (Table 2.1) shows some features of the considered rainfall structures, and forms the basis for the event type identification and classification presented in section 3.4.

Table 2.1: *Rainfall structures and characteristics in typical mid-latitude storms in the northern hemisphere [Ehret, 2003].*

Storm structure	Warm fronts	Cold fronts	Meso-scale Convection	Super cells	Convective cells
Atmospheric scale	meso-β	meso-β	meso-β	meso-γ	meso-γ
Horizontal scale [km²]	10^3-10^4	10^3-10^4	$10-10^3$	10-50	10-50
Time scale [h]	4-24	4-24	0.5-4	0.5-4	0.5-1.5
Air motion Mechanism	stratiform	stratiform	mixed	mixed	convective
Shape	front	front	Cells along a front	Irregular cells	Irregular cells, oval shapes
Maximum Intensity [mm/h]	20	60	200	300	400
Average intensity [mm/h]	5	30	40	100	100
Main motion	With fronts	With fronts	With prevailing air motion	With prevailing air motion	Erratic

2.1.2   Urban Hydrology

Today more than half of the world's population lives in cities [Worldbank, 2019], a rising trend that has been observed from 1950 (mainly in Europe and America) and is expected to continue in the future; rising up to 68% by year 2050 [UN, 2018]. As the area of the urban settlements increases, dramatic changes are caused in the catchment hydrology that lead to faster response times and water quality degradation. As the generation of the urban flow is more complex and unknown than in rural hydrology, and the vulnerability of the population is as well higher in urban settlements, in the last decades, urban hydrology has become an important and active field of research both in terms of fundamental understanding and management approaches.

There are two mains reasons why there is a bigger and faster runoff response to the rainfall in urban areas. First, as the impervious areas are increasing (due to streets, pavements, buildings, top soil compaction etc.), the infiltration of the water in the ground is limited, increasing greatly the volume of the surface runoff (the faster component of the runoff) and causing immediate and faster peaks or flood levels. Secondly, the water flow is facilitated through the construction of sewer and drainage systems, which accelerates the travel time. Resulting, thus, in a response within minutes or few hours at maximum, instead of several hours or days that are common for rural catchments. Moreover, the sewer and drainage systems are designed on rainfall intensity statistics at a specific duration (often one or two hours). Nonetheless, it happens that when higher intensities are recorded at shorter duration (i.e. 15min), the pipes will be filled at a faster rate and cause overflow [Jacobson, 2011].

To predict the implications of the rainfall in urban areas and infrastructure; the surface water levels, flow velocities and discharge in the pipe network or in the inundation area, hydraulic models are built. Typically, these models consists of a coupling between the pipe network where 1D shallow water equations are used, and a 2D open flow model that computes the water levels based on the topography and shallow water equations. The two parts of the model are connected via the manholes; once the pipe network simulates surcharge at a specific manhole, the surface flow computation is initialized. Logically, the model input is the rainfall (either continuous or a single event) in order to produce the response discharge and water levels [van Dijk et al., 2013][Russo et al., 2015].

Since the models need to simulate processes that occur fast (at the scale of minutes or seldom hours) and at specific locations, they are often run at short time steps (1-6 min) [Fletcher et al., 2013]. Hence, the rainfall input provided to urban models should be at the adequate resolution to ensure an accurate runoff estimation. Different studies have investigated the sensitivity of the urban flow towards the rainfall resolution, and they agree that the required temporal resolution of rainfall should be within 1-5 minutes, and the spatial resolution between 100-500m^2 [Berne et al., 2004][Einfalt, 2005][Ochoa-Rodriguez et al., 2015] [Schilling, 1991]. The true spatio-temporal rainfall field at such scales is still unknown, but

however there are two rainfall measurements techniques that can provide rainfall information at similar scales:

a) Rain Gauge Networks

The traditional method for recording rainfall volumes is through networks of rain gauges; the typical one being a tipping bucket that records the rainfall at 0.1 mm resolution and at 1 min time steps. Due to the wind, wetting and the evapotranspiration effects, the gauge measurement are not error free [Lanza and Vuerich, 2009]. They manifest a standard error that changes with the severity of the storm; for instance an error of 7% is expected for a rainfall intensity of 25 mm/h falling within 5 min [Ciach, 2003]. However, the rain gauges represent rather correctly the rainfall volume at very fine scales close to the earth surface, therefore they are mainly used as an input for small urban catchment [Codo and Rico-Ramirez, 2018][Quirmbach and Schultz, 2002].

Nevertheless, rain gauges are point measurements and alone cannot describe the spatial structure of rainfall. Given a specific network, the point measurements can be interpolated at different spatial resolutions for each time step in order to get the spatial information. The relevance of such interpolation depends strongly on the network density. According to Berne et al. (2004), the required density depends on the catchment and on the city under consideration, varying from 1 gauge per $7km^2$ (case study Barcelona) to 1 per $18km^2$ (case study Lyon). But generally, a high density should be present in order to capture the strong spatial variability of the rainfall at the scales required by an urban model. Such dense networks don't exist because they are expensive and difficult to build and maintain [Berne et al., 2004].

b) Radar Networks

Weather radar measures indirectly the rainfall field through reflectivity; microwaves signals are transmitted by the radar via pulses, and the energy reflected by hydro-meteors, is captured back and is converted into intensity. As the radar does a 360° scanning around itself, it builds up rainfall fields usually at $1 km^2$ and 5 min resolutions. Due to the ability of the radar to capture the structure of the rainfall at such fine scales, its' potential for urban hydrology use, has been long recognized by urban hydrologist [Berne and Krajewski, 2013].

However, since the radar data doesn't measure the rainfall directly, it is subjected to several sources of errors, compromising so the accuracy of the rain rates. Many attempts have been made to recognize and to mitigate these sources of errors over the past decades, nevertheless since the true rainfall field is not known (radar rainfall rates at such scales don't necessarily meet the rain gauge measurement [Ochoa-Rodriguez et al., 2019]), no final conclusion can be reached regarding its accuracy [Steiner et al., 1999][Vasiloff et al., 2009]. But the use of radar data in urban hydrology is still a common practice, because they can capture small rainfall features (which the gauge could miss) that might have considerable impact on the urban floods [Gires et al., 2012][Ochoa-Rodriguez et al., 2015][Schellart et al.,

2014]. Moreover, given their ability to capture the spatial structure of the rainfall, they are perfect candidates to track and nowcast rainfall storms. A more detailed description of the use of radar in urban hydrology is given in the Section 2.2.2.

Regarding different types of storms (as mentioned in the previous section), urban catchments are particularly vulnerable to convective events (both at meso-γ and meso-β scale). Because these storms live very shortly (less than 4 hours) and deliver high amount of rainfall volume, the urban catchments respond very fast, causing the manholes to overflow and hence resulting in surface inundation [Thorndahl et al., 2016]. Even though in most of the cases, the water resides in a few hours, water could be trapped in depression, causing mobility interruption, economic losses (basements are commonly flooded), potential ground and groundwater contamination, and in severe cases even human casualties. Although there is some work being done already for increasing the resilience of the cities (by increasing the infiltration area, green roofs or retention volumes) [Löwe et al., 2017], it is still of great importance to develop fast and accurate pluvial floods forecasts, in order to inform the residents and to take measures in time for the mitigation of the damages [Jha et al., 2012][van Dijk et al., 2013]. While some issues related to urban models are still open for research; like the calibration of the model, the fast simulation time, the structure of the model etc., the main deficiencies today for such forecasts are related with the poor rain nowcast at fine resolutions [Thorndahl et al., 2016][Schellart et al., 2014]. There is an incomprehensible need to improve the rainfall nowcast for these events at such scales.

2.2 Weather Radar for Hydrology

2.2.1　　Radar Theory and Precipitation Measurements

The term Radar stands for "Radar Detection and Ranging", as it recognizes the presence of objects by capturing the electromagnetic radiation reflected by their surface. Based on the energy reflected, one can identify the location, the extent and the material of the object. Water droplets, as objects in the atmosphere, can be detected by microwaves, if located within the detectable range of the radar. Typically, the radar device consists of three parts; a) the transmitter that generates the microwave pulse, b) an antenna that shoots the waves in the atmosphere and receives the reflected energy, and c) a receiver that detects, amplifies and transforms the incoming signal. The schematic diagram of a radar measurement principle is illustrated in Figure 2.3. First the radar dish emits very short pulses of microwaves (the pulses are emitted at 3 milliseconds apart, for a microsecond) in a conical beam with a fixed angle, referred here as the elevation angle of the antenna (approximately 1.5°), into the atmosphere. Once the electromagnetic wave encounters an object (water droplets in this case), a part of the wave is absorbed or scattered, but another part is reflected back, which eventually reaches the radar where it is recorded by the receiver. Table 2.2 gives a summary of different bands, wavelengths and frequencies that are typically used in the radar systems.
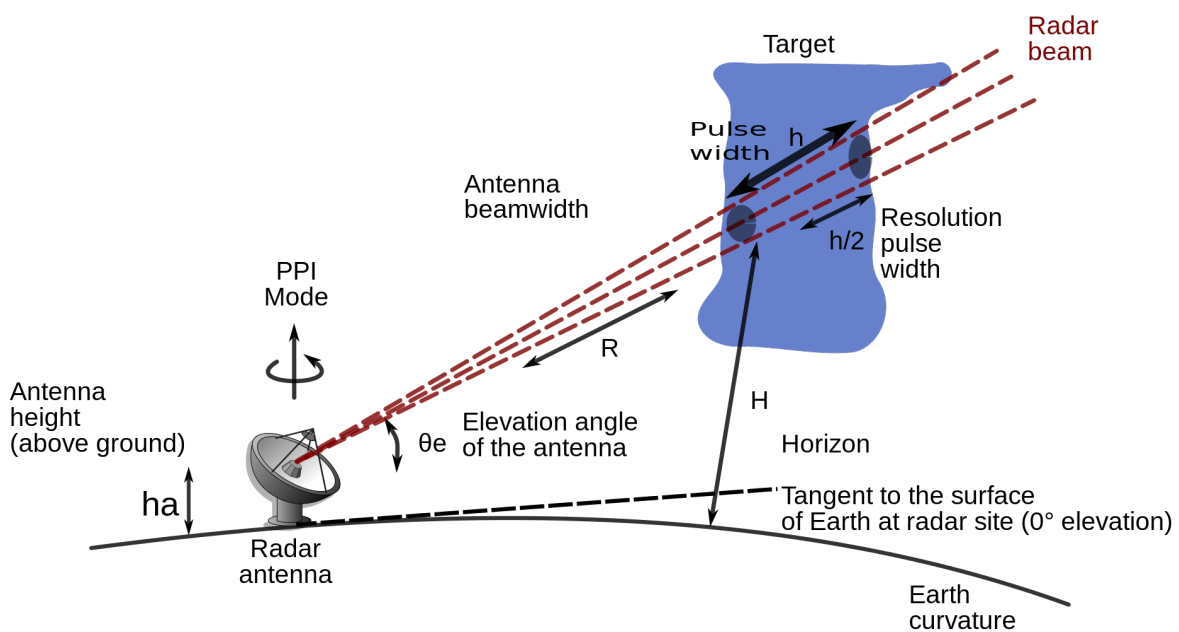
Figure 2.3: *A schematic representation of radar instrument measurement principle [Pierre, 2011]*

Table 2.2: *Different bands, wavelengths and frequencies of typically used radar [Sauvageot, 1992].*

Band	Frequency [GHz]	Wavelength [cm]
L	1 - 2	30 – 15
S	2 - 4	15 – 7.5
C	4 - 8	7.5 – 3.75
X	8 – 12.5	3.75 – 2.4
K_u	12.5 - 18	2.4 – 1.67
K	18 – 26.5	1.67 – 1.13

The time of the reflected energy (reflected time) is measured, and together with the azimuth and the elevation angle noted when the signal was emitted, one can conclude the exact position of the obstacle (water droplet) as indicated by Equation 2.1.

$$r = \frac{c}{f_r} = \frac{t \cdot c}{2}$$

where:
r – target distance [m]
c – speed of light [m/s]
t – time to target [s]
f_r – reflected wave frequency [GHz]

(2.1)

Once the distance is found, one can use the radar equation (Equation 2.2) to find out the reflectivity (Z) of the water particles [in mm6/m3] based on the measured back-scattered radiation (P_b[W]), the measured target distance from the reflected time (r[m]) and based on two empirical constants; i) C -the radar coefficient and ii) parameter k - the hydro-meteor reflection factor. The k parameter value for the hydro-meteors is found empirically, and typically it lays within 0.964 for rain and 0.456 for snow.

$$Z = \frac{P_B \cdot r^2}{C \cdot k^2}$$

where:
Z – reflectivity [mm^6/m^3]
P_b – back-scattered radiation [W]
r – target distance [m]
C – radar coefficient [Wm^5/mm^6]
k – hydro-meteor reflection factor [-]

(2.2)

Until today three types of radar systems are recognized [Krämer, 2008]; a) the conventional radars (X and C band), which are historically the most used and investigated ones. They emit a radiation beam only in the horizontal direction and measure only the reflected energy; b) Doppler radars, which emit a horizontal beam of energy, however they don't only record the reflected energy but also the phase shift between the emitted and the reflected waves (the so called Doppler effect/shift). By recognizing the shift, this type of radar enables us to compute the velocity of the reflected surface (either it is moving towards the radar, or away from it), that is a valuable information when identifying the object reflected; c) Polarized radars, which in contrast to the conventional and the Doppler radar do not only send a horizontal beam, but as the name suggests, their signal is polarized also in the vertical direction. The two emitted beams are perpendicular to each other, and thus perform a much better scan of the weather, recognizing as well the height of the objects (due to the difference between the vertical and horizontal reflection). This is particularly important in recognizing and classifying hydro-meteors. Dual-Polarised radars are new and have just started to become an active area of

research. Since the conventional radars are frequently used and have longer observed data, they are usually used for hydrological studies. In this thesis, conventional radar data (namely C-Band) is used, thus the subsequent literature review is focused on the error measurement and the hydrological importance of C-band radar.

As the C-band radar does a 360° scan around its' axes, it captures the reflected energy at a certain angle accuracy (usually at 0.5 or 1°) depending on the rotation speed, while in the radial direction the discretization of the measured reflectivity is determined by the pulse length (τ) according to Equation 2.3; for instance, for a pulse length of 1 µs the discretization would be 150m [Krämer, 2008].

$$\Delta r = \frac{c \cdot \tau}{2}$$

where:
Δr – radial resolution [m]
c – light speed [m/s]
τ – pulse length [s]

(2.3)

Thus, the reflectivity is measured in polar coordinates, and is usually converted in the Cartesian coordinates (typically with 1 km² resolution) to facilitate the interpretation of the data. The convectional radar systems, issue a set of pulses at a given opening angle, thus the volume scanned by the radar is increasing with the distance from the radar (as shown in Figure 2.3). Consequently, the radar receives the signal of the water droplets from different heights, that in the end are averaged together without any grouping based on the height information. Figure 2.4 illustrates the relationship between different opening angles, the height of the radar beam and the distance from the radar [Seed et al., 2002]. Therefore, one has to keep in mind that the radar reflectivity field given by the C-band radar at a specific spatial and temporal resolution, is actually representing instant scans of the atmosphere height, that is increasing with the distance from the radar.
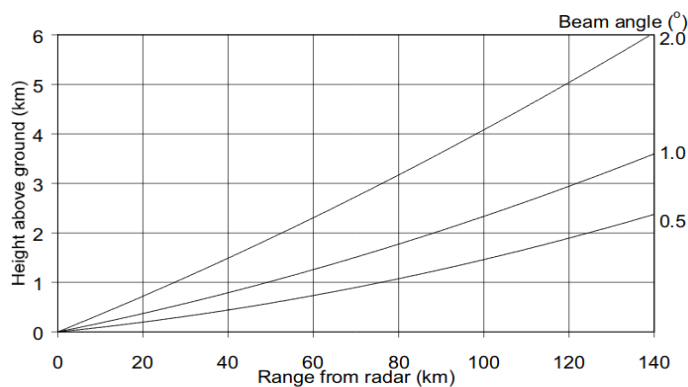
Figure 2.4: *Height of the radar beam at different ranges from the radar (km) and at different opening angles [Seed et al., 2002].*

Once the spatio-temporal distribution of the reflectivity is determined, rainfall rate can be estimated from the empirical relationship with the measured reflectivity. This relationship is determined based on observed data from disdrometer and tipping bucket measurement. Marshall and Palmer were one of the firsts to study this relationship and concluded that the

relationship between reflectivity and rain rate is of a power type, and the rain rate is to be calculated as indicated by the Equation 2.4 - where *a* and *b* are empirical constants.

$$z = a \cdot R^b$$

where:
Z – the reflectivity measured in [dBz]
a, b – Marshall Palmer Coefficients [–]
R – Rainfall rate [mm/5min]

(2.4)

They suggested the values a = 200 and b = 1.6 for these constants. However, the values of the empirical parameters *a* and *b* vary with different meteorological conditions. Table 2.3 gives typical values that are used for different type of storms. Even though choosing the right constant is still an open research field, in this thesis the Germans' Weather Service (DWD) constants for the Marshall-Palmer relationship are used.

Table 2.3: *Typically used a, b coefficients for the Marshall-Palmer Relationship [Ehret, 2003].*

Meteorological condition	a	b
convection	300–500	1.4-1.5
stratiform	200–250	1.5-1.6
DWD values	256	1.42
Original Marshall and Palmer	200	1.6

As already mentioned, because the radar is measuring indirectly the rainfall rate, it is subjected to a range of errors that reduce considerably its' ability to deliver accurate rainfall rates. While some of the errors can be detected and corrected, others might be difficult to quantify. These errors are summarized below into 4 groups:

a) Errors due to the ground clutters and objects that obstruct the beam

While the radar is performing its' scan of the atmosphere, microwaves of not-watery objects may be reflected or blocked. For instance, a ground obstacle (a mountain/hill), here referred to as ground clutter, can stop the radar beam, and cause a void in the field behind the obstacle. These are usually static clutters, that don't change their position with time, and hence can be identified easily from accumulating the radar data reflectivity over a long time period (one or more years) [Berndt et al., 2013]. Once they are identified, interpolation methods from the surrounding grid cells can be used, or the recording from another radar can be integrated to fill the missing values. There are also other objects in the atmosphere (near to the earth surface) that are time-invariant, here referred to as dynamic clutter, whose signal may be captured by the radar; this can be birds flocks, dust, flying object etc. As these objects are time-invariant, it is quite difficult to clear the radar data from these noises (or errors).

b) Errors due to the attenuation of the waves

One of the main problems with the conventional radar data is that, depending on the size and the density of the water droplets, the signal can be strongly attenuated with the distance from the radar location. This is specially the case when hail or an intense convective storm is located at the proximity of the radar; the storm will scatter most of the energy emitted by the radar, hence all the energy beyond the storm or the hail is very weak (causing very low

reflectivity) or in-existent. Thus, a decrease in reflectivity is observed for convective storms or even for wide spread stratiform storms, that is dependent on the intensity of the storm, the frequency of the waves and the distance from the radar (as indicated in the Figure 2.5).
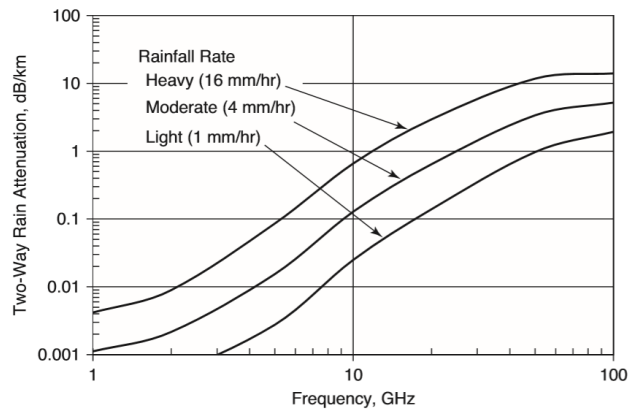
Figure 2.5: *The loss of the rainfall information due to an increase of the attenuation [in dB/km] for different storm intensities in relationship with different frequencies of the electromagnetic waves, according to [Curry, 2012].*

Another very important source of error is that the radar beam doesn't sample exactly the given area of interest. It is desired to have scans of the near-surface atmosphere, so one could conclude how much rain is falling on the surface. However, as the horizontal beam is propagating, the scan area increases with the distance due to the opening angle (as shown in Figure 2.3) and due to the earths' curvature; a horizontal signal issued at the earth surface, after 100 km will be located at a height of 780 m. As the precipitation is very chaotic in its' nature, with high spatial variability (both in horizontal and vertical direction), such measurements do not represent the true rainfall volume on the ground surface. Altogether a 20% error in the rainfall estimation can be introduced to the conversion just due to the measurement of the wrong rainfall volume [Seo, 1997].

c) Errors due to the indirect measurement of the rainfall

As mentioned above, the rainfall rate is calculated through two steps a) the reflectivity is calculated based on the radar equation, and b) based on the power relationship the reflectivity is converted in rainfall. Both of these conversions are not perfectly representing the relationship between the back-scattered radiation and rain rate, thus are introducing additional errors into the transformation. For instance, the radar equation [Crozier, 1986] is considered to be accurate under the following assumptions:

i) the particles are homogeneous in volume and are of spherical shape,
ii) the reflectivity Z is uniform within the volume scanned and the pulse interval,
iii) the water particles are all in the same phase (either ice or water),
iv) the volume of the pulse is filled with randomly distributed water particles.

Such assumptions are most of the time not fulfilled, because the droplets may not be symmetrical or homogeneous in the volume, neither can they be all in the same phase (either

ice or water). Moreover, the governing relationship by Marshall and Palmer between reflectivity and rain rate is a general case, as it has already been proven its' dependency on the rainfall event type and the type of the hydro-meteor. Lastly, the measured data are converted from a polar grid to Cartesian one causing unwanted smoothing of the data [DWD, 1998] [Sharif et al., 2001].

d) Errors related with the radar system itself

Seo (1997) suggests that the radar system can introduce bias into the measurements due to the system losses with time and if the antenna gain in not precisely known. He suggested that these errors could be identified once the radar measurements are compared with the measurements of another one.

The main challenge when trying to recognize and to mitigate these errors, is the missing true rainfall field which could be used as a basis for the evaluation of the radar rainfall estimates. Usually gauge data are considered as a true observation of the rainfall at the ground level and hence are often used as a reference to judge the radar data reliability and to evaluate the error-reducing methods [Kneis and Kneis, 2009]. For fine temporal accumulation (for instance under 1 hour) the measuring scales of radar and gauge are very different, and as such, one should not expect that the volume captured by the radar in one scan to be exactly the same as the volume accumulated at the gauge within 5min [Ochoa-Rodriguez et al., 2019]. However, for longer temporal accumulations (hourly, daily, monthly) a direct comparison of the radar with the gauge data is possible and is commonly done [Heistermann et al., 2013]. While using the longer accumulation durations can tell the hydrologist about the quality of the radar data and the applied correction method, it is still not very clear how representative the fine temporal structure of the radar measurements is.

Nevertheless, even though the radar data at such resolutions is prone to numerous errors, it has still been proved useful for urban hydrology [Berne and Krajewski, 2013]; as the urban system is sensitive to the rainfall intensities at fine scale, the radar data are well suited for the modelling or forecasting of the urban floods and for urban water management. Thus, the next section gives first an overview of methods used to mitigate radar errors, and then a summary of radar based forecast most often used.

2.2.2   Radar Data as Input for Urban Modelling

The rainfall process is characterized by a high intricate spatial and temporal variability (a combination of continuity and intermittence). The 1 min recording gauges can give a good insight about the temporal variability of the rainfall, however since they are point measurements, can't provide the same for the spatial variability. As for urban hydrology the spatial resolution of the rainfall events is a must [Berne et al., 2004][Gires et al., 2012] [Ochoa-Rodriguez et al., 2015][Schellart et al., 2012] a need thus emerged to either interpolate gauge data or to use the radar information.

A rainfall interpolation of fine temporal scales (~ 5 min) from the gauge data can be done either by geostatistics[3] (Ordinary Kriging for instance) [Berndt et al., 2013] or other interpolation techniques (multi-regression, nearest neighbour, inverse distance etc.). However, these interpolations are highly dependent on the study area, precipitation type and time scales [Hu et al., 2019]. At fine temporal scales, these methods don't exhibit a high spatial variability (characteristics of the rainfall field); even the Ordinary Kriging (OK) that tries to model the observed spatial consistency through a variogram[4], has been proven to smooth too much the rainfall intensities in space. This means that especially for convective events, the rainfall would be overestimated at some locations, and underestimated at locations between gauges where a storm has passed but was not recorded. Thus, the performance of such interpolation methods is dependent of the station density available; a higher station density yields better results than a lower station density. However the high density needed is usually not provided [Berne et al., 2004].

That's why the hydrologist turned their sight towards radar data. Radar data, as already mentioned, can produce a high spatial variability of the rainfall but however at the cost of the accuracy. A lot of research is going on to reduce these errors, nevertheless one has to be realistic that it is impossible to reach error-free radar precipitation estimates. Overall there are three ways in the literature to obtain the so-desired high spatial and temporal rainfall information by using radar data; a) by trying to minimize the radar error, b) by integrating the radar information to the interpolation of the gauges, and c) merging of both radar and gauge information to obtain a final field.

2.2.2.1   Minimizing radar error

The radar error for the rainfall intensity estimation is commonly corrected based on the tipping bucket recording data. Here the radar volumes at specific time intervals are corrected by considering the accumulations of gauge data as true observations. A correction factor is calculated at each gauge location for a specific accumulation time, and then these factors are transferred to other locations. These are usually referred to as bias adjustment methods.

---

3   Geostatistics includes the techniques to describe and to model the spatio-temporal relationship between the data.

4   Variogram describes the spatial dependency of the variance between pairs of data based on their distance.

The simplest application of such method is the Mean Field Bias adjustment (MFB), which assumes that there is a spatial uniform bias throughout the radar image which can or not change with time (static bias – one bias calculated for a long time period, dynamic bias – bias calculated for antecedent periods). Thus, a correction factor is found for the whole radar field (either static or changing in time), and is multiplied with each radar grid cell. Such methods can work for higher accumulations (monthly, daily, annually) [Borga et al., 2000][Borga et al., 2002][Goudenhoofdt and Delobbe, 2009] however are not yielding very good results for sub-daily durations due to the high spatio-temporal variability of rainfall at such scales [Ochoa-Rodriguez et al., 2019]. Nevertheless for nowcasting application dynamic MFB (typically for hourly accumulations) is commonly used [Seo et al., 1999][Chumchean et al., 2006].

Other studies have been performed in order to find correction factors at each station location and then regionalize or interpolate the correction factors at other radar grid cells with no gauge information [Seo and Breidenbach, 2002]. However, one of the main drawbacks of MFB method is that, if for instance a local convective storm has been captured by a gauge (hence high rainfall accumulation), and due to strong attenuation the radar accumulation at that location is low, the MFB will produce a high correction factor. This factor when applied to other regions of the radar, may cause very high unrealistic rainfall accumulation from the radar data, i.e. at a place where another convective event is registered but the attenuation is not that strong.

Contrary to MFB, Quantile Mapping Bias correction technique (QQ) has been developed to adjust the bias both in space and time, while keeping the probability distribution of intensities intact [Rabiei and Haberlandt, 2015]. The radar intensities at each grid cell are substituted according to their respective quantile value from the gauge probability distribution. Thus, it causes more reasonable radar intensities while keeping the similar structure from the radar data. However for such method to work, one needs enough gauge recorded intensities to build a reliable cumulative distribution function. This could be achieved by either increasing spatially (interpolating the rain gauge values) or temporally (by extending the time steps considered) the sample size. Whilst the first one can introduce new uncertainties due to the smoothing of the rainfall field (as in the case of the MFB), the latter one may increase the duration so much that different events might be sampled, introducing thus an additional bias at a certain time step [Rabiei and Haberlandt, 2015]. As both of the methods presented here have their advantages and disadvantages it is in the aim of this study to find out which of the bias correction would work better for radar based nowcasting for the urban hydrology application.

2.2.2.2   *Integrating radar information to gauge interpolation*

As mentioned before Ordinary Kriging[5] (OK) interpolation of gauges smooths the spatial information of the rainfall. Thus a Kriging with External Drift (KED) is used instead, by

---

5   Ordinary Kriging models the spatial relationship of the given data based on their variogram by minimizing the error variance.

introducing a new variable which constrains the interpolation of the gauge data: the radar rainfall rate. Another application of KED is with the elevation as an external drift variable [Haberlandt, 2007]. While this affects time accumulations of days or higher, the elevation is not highly correlated with the intensities at 5 min time steps [Bárdossy and Pegram, 2013], indicating that the elevation can be left out. A different application of KED is the indicator KED (IDK) which constrains even further the interpolation on specific thresholds of intensities. However, Berndt et al. (2013) found that for very short time steps, the IDK does not introduce additional accuracy to the interpolation.

Overall it appears that KED with radar information might perform well also in 5 min time steps interpolation of rainfall information, that is why it is worth considering and investigating its' usage for radar based nowcasting.

2.2.2.3   *Merging radar and gauge information*

As the name suggest, the methods under this category try to merge together the two data-sources, without giving more priority to one or to the other. Three main methods are distinguished in the literature: a) Co-Kriging (CoK), b) Bayesian Merging (BM) and c) Conditional Merging (CM) of radar data (known also as kriging with radar-based correction).

The Co-Kriging is very similar to the KED but unlike it, the outcome field is calculated based on a linear relationship between the gauge and radar intensities. This method has been used by Krajewski (1987), which requires as input into the kriging system both the individual covariance of rain and gauge data and as well their cross variance. In comparison to the KED, the CoK is computationally expensive and may lead to numerical instabilities [Ochoa-Rodriguez et al., 2019]. A reduced form of the CoK was performed by Schuurmans et al. (2007) and Velasco-Forero (2009), where only the radar information at the target locations is used, simplifying thus the CoK matrix and leading to more stable and faster systems. However, they concluded that the KED performs better than CoK respectively for daily and hourly resolution. Thus, one can conclude that KED is a superior method due to its simplicity and reliable results.

Bayesian merging performs the merging of the rain-radar intensities by considering their uncertainties; a) the gauge uncertainties are computed from the error covariance of an Ordinary Kriging field, b) the radar uncertainties are computed by direct comparison to the OK of gauge data. Once the uncertainties have been computed, a Kalman filter[6] is employed to combine both sources of data proportional to their uncertainties. Different configuration of the Bayesian Merging were studied by Mazzetti and Todini (2004), Wang et al. (2013) and Ochoa-Rodriguez et al. (2015a) (the latter ones performed on a 5 min and 1 km^2 resolution typically needed for urban applications). Ochoa-Rodriguez et al. (2015a) compared a

---

6   A Kalman filter is an algorithm that keeps track of both the current state of a variable (x), and the variance or the uncertainty of the variable (Px). In radar nowcasting, the uncertainty of the radar derived intensities is calculated based on the discrepancies between radar and gauge. For each time step t, the intensity at lead time $+LT$ is found based on the last radar intensity ($x_{t|t-1}$) and uncertainty observation ($Px_{t|t-1}$).

modified version of the Bayesian theory towards the KED and found out that both of them behave similarly.

Lastly, conditional merging (CM) was first done by Ehret (2003) and later was implemented and tested also in Berndt et al. (2013). The method consists in merging together the OK interpolated field of the gauges with the variance of the radar data. Berndt et al. (2013) found that CM outperforms KED (slightly) at fine temporal scales up to 10min. However other studies done by Jewell and Gaussiat (2015), Nanding et al. (2015) and Kumar et al. (2016) at coarse resolutions found out that KED outperforms the CM.

Overall it looks like that at fine temporal resolution the KED is the most popular method which gives the most reliable results, as indicated by Ochoa-Rodriguez et al. (2019). Regarding the CM, even though Berndt et al. (2013) showed that CM was superior to the KED, other studies are not consistent with this result. While the MFB has been proved to be outperformed by KED, nothing is said related to the QQ: there is no direct comparison of the methods to such fine resolution. The BM is shown to perform as good as KED, thus one can expect that if one of these methods outperforms the KED then consequently it will outperform the Bayesian Merging as well.

Therefore at this point a question arises on which of the methods is more adequate for correcting radar data on fine temporal and spatial scales; Quantile Mapping Bias Correction, Mean Field Bias Correction, Conditional Merging, or Kriging with External Drift? Since not many of the literature check the merging results towards the Ordinary Kriging of the stations, it is logical to ask whether at high density, the merging techniques still provide better information than the single data sources (either OK interpolation of stations or radar data)? Moreover, only a few of these corrections have been performed for the nowcasting application of the radar data, consequently the questions arise: how suitable are these methods when run online (only past information is available)? Can they outperform the MFB which is usually employed for radar based nowcasting? Up to what degree does the merged data affect the radar nowcast and the prediction accuracy? These are the main questions treated in Chapter 5.

As indicated as well at Ochoa-Rodriguez et al. (2019), only very few cases focus on the performance of such integration methods at fine scales suitable for the urban model. This is mainly due to the fact that many hydrologists are very sceptical at the integration of these two completely different precipitation measurements: radar represents an instant scan of the atmosphere at different heights, while station measurements represent a rainfall accumulation at a single point at the ground surface; both are representing different scales, thus cannot be merged together! However, one has to keep in mind that at the moment this is the best option available to perform nowcast for the urban catchment and to validate models. First, it has been proven that running a model with gauge data, yields best results, thus many references for the urban model are flood levels simulated with gauge data. But when radar based flood nowcast are compared to the gauge based flood forecast, they are starting with a clear disadvantage. Thus, in order to minimize as much as possible the discrepancies between radar

and gauge and in order to improve the predictability of floods simulated by models calibrated by gauge data, it makes sense to use the merging methods at such fine temporal and spatial scales.

2.2.3   Use of Weather Radar in Urban Flood Nowcast

While the sub seasonal and synoptic scale rainfall forecasts[7] are quite predictable, knowing the exact precipitation volume that will fall in the next hours (here referred to as nowcasting) at the meso-γ or -β scales, is a difficult task to perform [Keil et al., 2014][Surcel et al., 2015]. This is mainly because the atmospheric processes and the current state of the weather system are quite complex to describe and to model [Germann et al., 2006][Germann and Zawadzki, 2004]. Numerical Weather Prediction (here referred to as NWP) models try to describe and to simulate such processes, with the aim to predict different variables, and among them the precipitation volumes expected to fall in the near future, by solving a handful of equations at finite discretization of the atmosphere.

NWP models have undergone a lot of improvement on the recent years and they can produce quite well the forecast at large scale (scales larger than $10km^2$ and longer than 1 hour accumulations for lead time higher than 6 hours – here referred as short-range forecast) [Ahrens and Jaun. 2007][Ballard et al., 2012][Bowler et al., 2006][Golding, 2009][Surcel et al., 2015]. However for the fine scales suitable for the urban catchment, NWP models are not able to provide good rainfall nowcast; they can give good estimates on the severity of the storms and intensities [Kato et al., 2017], but fail to produce the exact time and location of such intensities that are crucial for the urban floods [Kato et al., 2017][Šálek et al., 2006] [Surcel et al., 2015]. This may happen for various reasons; either because the simulation becomes computationally very expensive, or the physical process associated with the rainfall formation at such scales are not fully understood [Grecu and Krajewski, 2000][Germann and Zawadzki, 2004] or high uncertainties are present due to the spin-up problem [Golding, 2009] [Kato et al., 2017][Surcel et al., 2015]. The spin-up problem, also called the initial condition problem, rises from our inability to measure and to know the true rainfall field. As described by Lorenz (1973), even if we could build up the perfect forecast model, a small error at the initial condition will grow up so much with longer lead times[8], that soon the model will have large errors and zero ability to predict the rainfall. Mainly for these reasons, NWP models still need some time to provide reliable rainfall nowcast at fine scales required for the urban flood forecast.

On the other hand, radar information offers a useful tool to identify and nowcast the rainfall rates at fines scales. Leaving aside the errors associated with the radar ability to measure accurately the rainfall field, rainfall structure from radar data are easy to be

---

7   Depending on the forecast horizon, the forecasts distinguished here in this thesis are: a) long range – with lead times > 1 day, b) medium range – with lead times < 1 day, c) short range – with lead times between 6 and 12 hours, and d) nowcasting – with lead times < 6 hours.

8   Lead time is the horizon of the forecast issued; the duration between the time the forecast is initialized and the time of the target forecast.

identified, tracked and extrapolated into the future. Many studies have shown that for short lead time (1 to 2 hours) radar based nowcast outperforms the NWP [Berenguer et al., 2012] [Jensen et al., 2015][Lin et al., 2005][Zahraei et al., 2012]. Contrarily, for longer lead times (more than 2 hours) the performance of radar based nowcast decreases rapidly (depending on the storm type and the size of radar field available), and hence a blending between NWP and radar based nowcast is proposed [Bowler et al., 2006][Jensen et al., 2015] to improve the nowcast skill at these lead times. However, one has to keep in mind that the accuracy of the blended product at short lead times is as high as the accuracy of the radar nowcast. It is the aim of this thesis to investigate the radar based nowcast, with the assumption that an improved radar nowcast, would lead to a better blended product for the overall prediction of rainfall structures at fine temporal and spatial scales. Therefore, a blending with NWP is not included. Also, following the study from Luk et al. (2001) that the nowcast based on solely a dense gauge network is not yielding good results for lead times longer than 10-15min, the following literature review and the development of methods are focused only on radar based nowcasting techniques.

Typically there are three mains steps used in radar based nowcasting, illustrated as well in Figure 2.6: a) identifying the rainfall in the area of study at time t_0 (the nowcast is initialized), b) track the rainfall to determine main advection[9] velocity and direction from time t_{+1} and previous time step t_0, and c) given an advection vector, the rainfall is extrapolated in the future at different lead times t_{+LT}. The rainfall is either considered as an object – individual storms are identified and tracked, also called the object oriented approach (Figure 2.6), or as a field – the whole radar imaged is tracked with no special consideration for singular storms, also referred to as a field based approach (Figure 2.7). While the object based approach is considered more accurate especially for convective events, as it considers storms independently, there are many issues regarding splitting and merging of the storms [Pierce et al., 2004]. Thus, the field based approach has been widely used in literature, however they best perform when the field has a distinctive movement (i.e. stratiform events) [Han et al., 2009]. A review of the existing tracking methods employed in some of the most used or cited nowcast models in the literature, is given in the following section.

a) Storm Identification   b) Storm Tracking   c) Storm Extrapolation
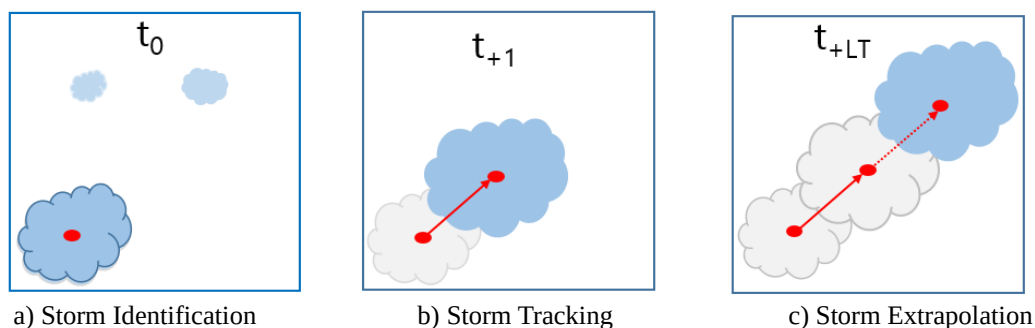
Figure 2.6: *The main steps of an object based radar nowcast. Blue indicates the current instance of the storm at any time t, and grey indicates the past instances of the storm (here at t_{-1}). The advection vector is shown with red.*

---

9   Advection here is considered as the horizontal movement of the storms or rainfall structures as captured by radar data.

| t₀ | t₊₁ | t₊LT |

a) Rain detection   b) Field Tracking   c) Field Extrapolation
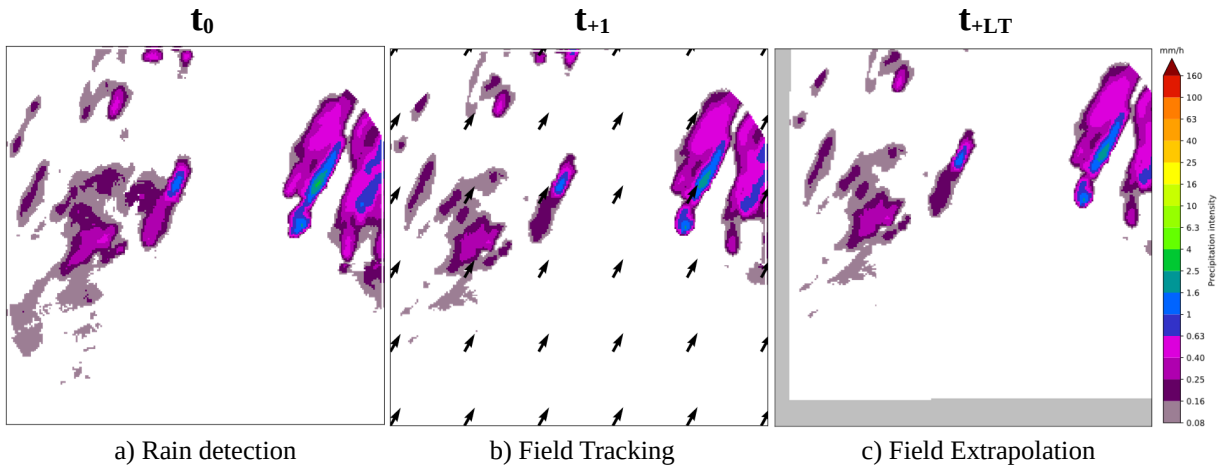
Figure 2.7: *The main steps of a field based radar nowcast. Where coloured pixels indicate the rainfall volume according to the colour bar in mm/h, the field advection vectors are shown in black vectors and the grey plot region at figure (c) indicates the gap created due to the movement of the field with a specific advection vector at different lead times.*

2.2.3.1   *Object oriented approach*

The methodology of this approach is shown in Figure 2.6. A storm here is defined as a group of radar grid cells (the group must be bigger than a threshold X set beforehand) that have intensities higher than a threshold (Y in mm/h). If a storm is identified both at the time step t_0 and t_{+1}, they are checked whether or not are the continuation of the same storm – also referred here as matching of the storms. The storms are matched, if the centre of intensity of storm t_{+1} falls within the boundary box of the storm t_0. Once the storms are matched, a movement vector (referred to here as advection) is found by the cross-correlation method. In this method, the storm at time t_0 is shifted with different advection vectors (different magnitudes and directions), and is cross-correlated with the storm at time t_1. The advection vector with the highest correlation is then said to be the advection velocity of the storm.

Other methods calculate the advection speed directly based on the displacement of the centre of intensity. However at such methods, because some storms are very variant in time, the location of centre of intensity may change drastically from time step to time step, causing very high and unreasonable advection speeds. Thus, the cross-correlation method is thought to provide more reliable results. A gradient search is often included in the cross-correlation in order to decrease the run time of the tracking algorithm.

It is usually the case, that two storms merge together at a certain time, or a single storm split between several daughter storms. Thus, special attention should be paid to these splitting and merging of the storms. This evolution can be identified by checking two criteria: a) the minimum distance between the storms that have splatted or merged should be smaller than the perimeter of the merged or that-is-splitting storm, and b) the position of the centre of intensity of former storms should be within the boundaries of the latter storm. A brief illustration of such splitting and merging criteria is given in Figure 2.8.
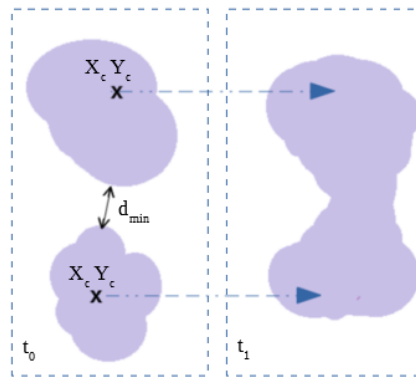
Figure 2.8: *The combination of two individual cells at time t_0 to a single one at time t_1 is done only if the $d_{min} < d_{limit}$ and if centroids of t_0 storms (X_c, Y_c) fall within the contour storm at time t_1 – shown with arrows (adapted from Krämer, 2008).*

Examples of object-oriented approach in the literature are for instance a) HyRaTrac (Hydrological Radar Tracking) developed by Krämer (2008) with cross-correlation based advection, and b) TITAN (Thunderstorm Identification, Tracking and Nowcasting) developed by Dixon and Weiner (1993) with centroid based advection on a 3D Cartesian coordinate system (later called TINT for the nowcast on the 2D data). Problems with the TITAN application mainly on the advection estimation for some strong convective events have been already identified [Pierce et al., 2004] and attempts have been made to improve such deficiencies [Han et al., 2009] by integrating a cross-correlation field approach. At a certain degree the Enhanced TITAN (E-TITAN) performance improved with the cross-correlation approach, indicating that these methods perform better than the centroid based method. Krämer (2008) implemented the cross-correlation method to determine directly the local storm advection, and unlike E-TITAN doesn't have the same restrictions for the maximum advection speeds. The restrictions present on E-TITAN cause an underestimation of advection for very severe storms that move very fast, while HyRaTrac may be able to track them due to the maximum found cross-correlation. However, one has to keep in mind that HyRaTrac just like E-TITAN can be prone to false merging or splitting of cell, and hence introduce errors in the advection speed and in the nowcast of the storms. Since the HyRaTrac was applied successfully to the real-time nowcast of a sewer system flow (with NSE8 = 0.7 for lead time of 15min) [Krämer et al., 2007] and it includes the same basic principle as TITAN, it is here selected as a reference model for the object based radar nowcast.

2.2.3.2 Field based approach

The methodology of this approach is shown in Figure 2.7: like the objected oriented approach the advection vector is determined by the cross correlation method, but instead of finding local advections for each storms, the field based approach works on the entire domain – each grid cell identified with rain will move according to one advection vector. The traditional method to calculate such field vector is the cross-correlation method. According to Johnson et al. (1998) this method provides satisfactory results about the image advection, and can work both on stratiform and convective events. An example of such implementation is the Tracking Radar Echoes by Correlation (TREC) method developed by Rinehart and Garvey

(1978). However, one of the main drawbacks of this method is that since there is one motion vector for the radar echo, it can not represent well enough the individual storms that might co-exist inside the echo but exhibit different speeds and even different directions [Han et al., 2009].

Thus, another approach was created first by Horn and Schunck (1981) that later served as a basis of many studies [Germann and Zawadzki, 2002][Grecu and Krajewski, 2000][Liu et al., 2015][Zahraei et al., 2012]. It calculates the motion field by investigating the slope of the intensity change between two consecutive radar images – that today is referred to as optical flow method. The basic assumption of the optical flow theory, is that, for a short time period dt, the intensity doesn't change from time t_0 to time t_0+dt, but instead it is just displaced with a vector dx and dy. Therefore it is expected that the intensity (I) at coordinates (x, y, t) is the same as the intensity at coordinates ($x+dx, y+dy, t+dt$) as indicated in Equation 2.5 [Marzban and Scott, 2010]. For each grid cell a given dx and dy is determined, and the set of all dx and dy (representing all the grid cells in the image) makes the optical flow field.

$$I\big(x,y,t\big) \approx I\big(x+dx, y+dy, t+dt\big) \qquad (2.5)$$

The Equation 2.5 can be expanded by the Taylor Series and can be solved by finite differences between the two images. But since there are two unknowns (dx and dy) some constrains have to be made. Based on the assumption of such constrains different solution methods are distinguished: a) [Horn and Schunck, 1981] – uses a constrain where nearby pixels have similar displacements, b) [Lucas and Kanade, 1981]– assumes that the displacement within a window/region of size W is constant, and c) [Ruzanski et al., 2011] through Fourier transformation, the radar images are transformed and the optical flow equation in solved in the spectral domain (DARTS). The former two methods require a minimum of 2 images to determine the optical flow (thus considering only linear movement), while the latter requires 4 images to determine the optical flow (thus considering rotational movements as well).

Liu et al. (2015) compared the two former methods Lucas-Kanade (LK) and Horn and Schunck (HS), and as the LK outperforms the HS, they suggested to use a local LK method for the rainfall nowcast from infrared satellite images. Ruzanski et al. (2011) tested the DARTS method against the traditionally used cross-correlation method on radar data and found that DARTS improves the nowcast, nevertheless concluded that the advection field estimation affects little the nowcast performance. Moreover, DARTS advection field seems to be more sensitive than LK for individual storms, an example of such a case is illustrated in Figure 2.9: where the LK predicts more uniform flow vectors than the DARTS method. However, the solutions of DARTS require the determination of an empirical value based on prior knowledge. Since no prior information is available at the moment about what value would fit the best to the case study, only LK is considered here. Thus, the optical flow in LK will be calculated from the last 15 minutes according to the results from Ayezel et al. (2019) indicating that the rotational movements performs better than a linear one.

a) Lucas – Kanade   b) DARTS
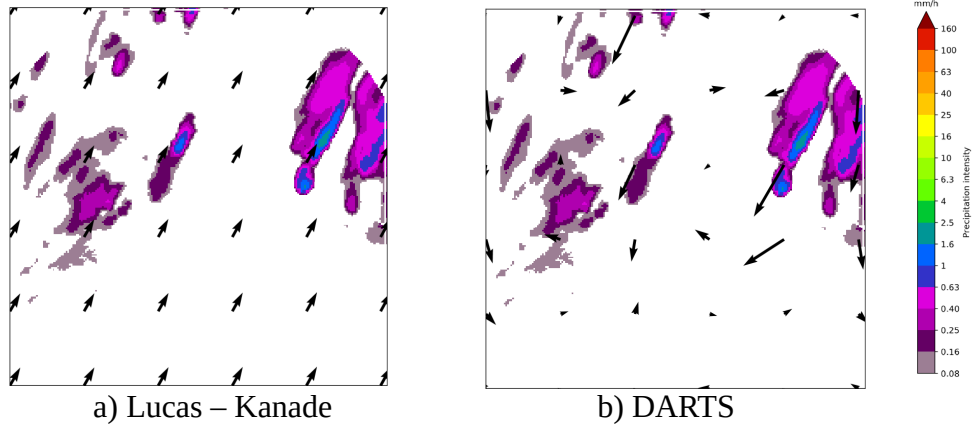
Figure 2.9: *The estimated optical flow field from Lucas – Kanade and DARTS method for a 5 min time step of a radar image. The black arrows indicate the motion field.*

2.2.3.3   *Extrapolation methods*

Once the advection vector is determined (either for an object or for a field), it can be used to extrapolate the rainfall in the future at a specific lead time. Persistence as a term is generally used to describe the assumption that the rainfall structure in the near future is not changing, or partially changing. The two used persistences in radar nowcasting are: a) Eulerian persistence – the storms are kept stationary as observed at the last time step. There is no change in intensity nor in the coordinates as dictated by Equation 2.6;

$$I(x,y,t+LT)=I(x,y,t)$$

where:
LT – lead time [min]   *(2.6)*

or b) Lagrangian persistence – the rainfall intensities are kept constant as last observed but given the vectors U and V (each for x and y direction) of advection, the coordinates are changed according to Equation 2.7. While for Eulerian persistence no change of the radar image is done, the Lagrangian persistence is illustrated in Figure 2.6 (c) and Figure 2.7 (c).

$$I(x,y,t+LT)=I(x\mp U\cdot LT,y\mp V\cdot LT,t)$$

(2.7)

where:
LT – lead time   [min]
U – x component of advection [m/s]
Y – y component of advection [m/s]

As mentioned in Grecu and Krajewski (2000), the radar structures at larger scales (for instance 4×4 km^2 and hourly accumulation) exhibit better persistence than at smaller scales. For large scales the Eulerian and Lagrangian persistence can produce reliable results, but as the resolution gets finer, the temporal distribution of rainfall structures is changing rapidly, and these persistence-s lose their ability to represent accurately the rainfall. The change of the nowcast performance from different models at different scales was investigated by Surcel et al. (2015), and the Figure 2.10-left gives an overview of their results. Nevertheless, for very short lead time (at the range of few minutes), under the assumptions that the rainfall doesn't exhibit a strong variation, the two persistence-s could be useful even for fine scales. Lorenz (1973)

concept about the prediction skills of different models based on lead time is presented in Figure 2.10-right: he argued that for very short lead times Eulerian persistence can produce good results, but shortly after the nowcast initialization it is outperformed by the Lagrangian persistence. However like the Eulerian, the Lagrangian persistence does not consider any birth or death of the storms, and as such it loses its' nowcast skill quite fast. For longer lead times more complex approaches are required that can model storm birth, decay and death of the storms.
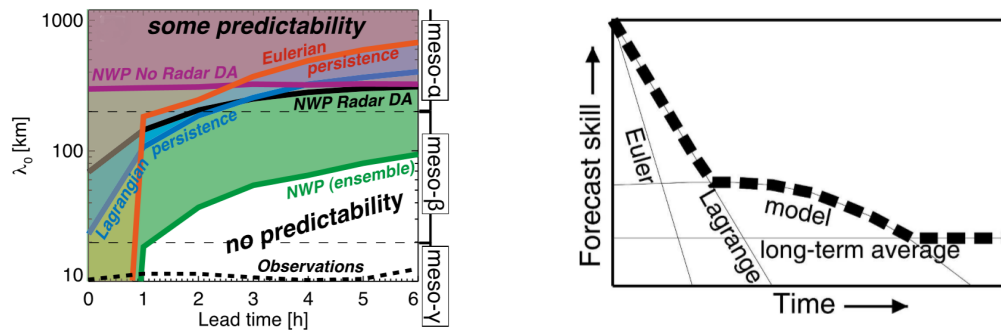
Figure 2.10: *The dependency of the forecast skill of different models based on the scales at consideration and different lead times left) according to Surcel et al., (2015) and right) according to Lorenz (1973).*

Thus, complex models that account for the non-linear extrapolation of the rainfall field in the future are studied in this thesis, with the aim to investigate the effect that the non-linear consideration has on the nowcast skill at different lead times.

2.3 Predictability Limits

As the Lagrangian persistence does not include the storm death or birth, the nowcast skill of this model is strongly dependent on the life time of the storm. Consequently, for stratiform events, the Lagrangian based nowcast can yield reliable results for longer lead times than for convective events. For instance, for a storm that lives typically one hour long, the Lagrangian based nowcast losses its' nowcast skill shortly before one hour. Depending on the storm dynamics, the persistence can lose its' skill even before this, due to non-consideration of growth or decay of the storms. Table 2.4 gives a summary of up to what lead time one can expect the Lagrangian persistence to provide useful nowcasts (at fine scales) for different type of events.

Table 2.4: *Typical limits of predictability for different types of rainfall events [Krämer, 2008]*

Type of events	Reliable Nowcasts
Individual Convective cells	5 - 20 min
Thunderstorms	10 – 60 min
Meso-scale Convection	60 – 120 min
Stratiform	120 – 180 min

Here it is important to understand how to determine the lead time up to which the nowcast can be useful, which in literature is referred to as predictability limit. Even if the perfect nowcast model existed, it can yield reliable results only until a specific lead time, as due to our inability to measure and observe the true rainfall field, initial errors are introduced in the nowcast system. Lorenz (1973) defined this limit as the intrinsic predictability limit, which can be considered as well as the theoretical predictability limit - the limit that we can never cross, no matter how good the nowcast model is. However in real life, we are far away from creating a perfect nowcast model, thus other errors due to the model structure are introduced in the nowcasted field. This limit, that often is calculated in studies, is referred to as the practical predictability of storms. A good nowcast model is thus considered one that can beat the predictability limit of the Lagrangian persistence and hence provide a better nowcast skill for longer lead times. In literature there are mainly three ways to calculate directly the practical predictability limit of a nowcast system:

1) **The de-correlation time** (τ) - as presented in Grecu and Krajewski (2000), is the lead time when the nowcast field becomes completely de-correlated with the observed field. The fields are treated as completely de-correlated, when the correlation becomes lower than 0.5. However, in a later study by Germann and Zawadzki (2002), the value of 1/e ~ 0.37 was proposed as a good indicator for the predictability limit.

2) **Equitable Threat Score** (*ETS*) – also presented in Germann and Zawadzki (2002) is a categorical criteria that can be calculated as per Equation 2.8 for detection of rainfall above a given threshold [mm/h]. A value lower than 0.3 indicates that the nowcast is not useful at all.

$$ETS = \frac{A - A_R}{A+B+C+D} \quad \text{and} \quad A_R = \frac{(A+B)(A+C)}{A+B+C+D} \tag{2.8}$$

where:
A – represents the case when the rainfall was forecasted and observed (Hit Alarm)
B – represents the case when the rainfall was forecasted but not observed (False Alarm)
C – represents the case when the rainfall was observed but not forecasted (Miss Alarm)
D – represents the case when the rainfall was not observed and not forecasted (No Alarm)

3) **Receiver Operating Characteristic Curve** (*ROC*) – as presented in Codo and Rico-Ramirez (2018) is a categorical criteria for ensemble nowcasts that indicates the relationship between Hit Alarm Ratio (as per Equation 2.9) and False Alarm Ratio (as per Equation 2.10) on different probability of rainfall detection above a given threshold (mm/h). A typical ROC curve is illustrated in Figure 2.11. A nowcast model is said to have no nowcast skill if the area under the curve is less than 0.5 – the False Alarm Ratio is higher than the Hit Alarm Ratio.

$$HAR = \frac{A}{A+C} \tag{2.9}$$

where:
A – Hit Alarms
B – False Alarms
C – Miss Alarms
D – No Alarms

$$FAR = \frac{B}{B+D} \tag{2.10}$$
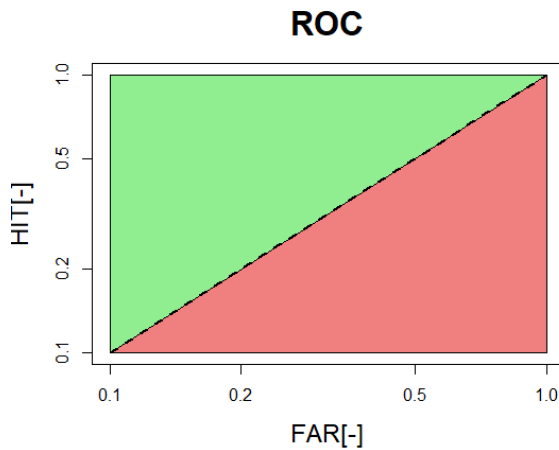
Figure 2.11: *Example of the ROC Curve; if the nowcast falls within the green area then it is considered useful, otherwise is considered as unreliable.*

It has been the object of many studies to investigate the predictability limits of the Lagrangian persistence and how to push through these limits. For small scale applications the typical predictability limits found by different studies are summarized in Table 2.5. Generally for small scale applications (1 km² and 5 min), the Lagrangian persistence has been proven to yield reliable results up to 20 – 25 min, after this lead time it is very difficult to predict the rainfall structures. Especially for high rainfall intensity, the predictability limit is quite short (up to 7 min for 50 mm/h).

Other attempts have been made to improve the predictability of rainfall by considering blending the radar based nowcast with NWP [Codo and Rico-Ramirez, 2018][Foresti et al., 2016][Jasper-Tönnies et al., 2018]. An example for this which is also quite famous is STEPS [Bowler et al., 2006]. This method uses the radar based nowcast from Seed (2003) up to 2 hours, and later on it uses downscaled inputs from NWP to extend the lead time up to 3-6 hours. As it has been proven that the large-scale precipitation is more persistent and influences the movement of the small scales [Surcel et al., 2015], the STEPS nowcast employees a cascade approach – estimating the temporal evolution of rainfall at different scales, and uses the Lagrangian persistence to predict the motion of rainfall features at such scales. A stochastic noise generator is then used to provide ensemble predictions and to account for different conditions of the storm birth. Such method has been proven to improve the rainfall predictability by some minutes (up to 30 min) [Foresti et al., 2016]. Even though the results are slightly better than the convectional single scale persistence, it is clear that there is a limit to the rainfall predictability that may not be exceeded at the moment by stochastic generators.

Table 2.5: *A summary of predictability limits on different spatial and temporal scales investigated by different literature studies with different nowcasting techniques.*

Study	Nowcast Resolution		Predictability Measure	Nowcast Type	Threshold [mm/h]	Predictability Limit [min]
	Spatial [km²]	Temporal [min]				
Grecu & Krajewski (2000)	4 km² 32 km²	15min	Decorrelation Time (τ ~ 0.5)	Lagrangian	0.1	30 min 60 min
Kato et al. (2017)	250 m² 1km²	5 min	Equitable Threshold Score (ETS ~ 0.3)	Semi - Lagrangian	1 5 10 30 50	35 min 22 min 17 min 10 min 7 min
Ruzanski et al. (2011)	0.5km²	1min	Equitable Threshold Score (ETS ~ 0.3)	Semi - Lagrangian	25 dBz	21 –23 m
Germann & Zawadzki (2002)	4 km²	15min	Equitable Threshold Score (ETS ~ 0.3)	Semi - Lagrangian	0.1 1 10	2.5 − 6h 1-3h 30 min
Foresti et al. (2016)	1 km²	5min	Equitable Threshold Score (ETS ~ 0.3)	Steps	0.5 5	60 − 90 min 30 min
Codo & Rico-Ramirez (2018)	1 km²	5min	ROC Curve (Area ~ 0.5)	Steps	0.1 1 3	2 − 3 h 2 − 3 h 1 − 2 h
Jasper -Tönnies et al. (2018)	1 km²	5min	ROC Curve (Area ~ 0.5)	Cosmo-DE Blending	3 7	<1h <1h

To summarize, the sources for errors (uncertainties) that are restricting the predictability limit are: a) a storm has to be identified in order to nowcast its' movement, hence no birth or storm pop up is possible. With stochastic generators (ensemble predictions) it is possible to model such births, but still their ability to nowcast the exact time, location and intensity is low. Also due to the image boundary of the radar data, no information is continuously flowing from this boundary at longer lead times (see Figure 2.7-c); b) uncertainties related with the tracking and evaluation of the advection vectors, and lastly c) uncertainties related to the extrapolation of the observed rainfall storms/field in the future due to the in-stationary nature of the rainfall. Uncertainty of the type (a), the inability to predict the birth of the storms limits greatly the prediction skill of the radar to about two hours, followed up by the uncertainty of the type (c), the inability of the nowcast to predict the death or decay of the cells, limiting the prediction skill of the radar up to the duration of storm itself.

In this thesis the focus is mainly at reducing the uncertainty of the type (c). More complex methods are developed to improve the simulation of storm growth/decay and death processes and consequently extend the practical predictability limit. As the archived observed dataset is becoming larger, new techniques rise: heuristic methods that learn from the available database and can predict based on previously observed motions and behaviours. Thus, here two data-driven methods are investigated on their ability to yield better nowcasts than Lagrangian persistence: the Neural Networks and the Nearest Neighbour approach.

2.3.1   Neural Networks

The Artificial Neural Networks (ANN) have already been used for rainfall prediction but not directly in rainfall nowcasting [Asanjan et al., 2018][Chaing et al., 2007]. Given a sample of input-output pairs, the ANN tries to model the relationship between them by employing a number of units (called neurons) that have non-linearly relationship between them (dictated by activate functions). The number of neurons and the way how the information flows between them (neurons having the same input make up a layer) build up the ANN architecture (see Figure 2.12). The weights at each neuron control how the data is transformed from input to output and can be optimized by minimizing a cost function (for instance reducing the bias between observed and simulated output), which is typically referred to as the training of the network. Depending on the right predictors (the input), the ANN can learn complex relationship and predict them successfully. The only main drawback of such predictions is that, since the ANN is a black-box model, there is no physical meaning to the weights and transformation learned.
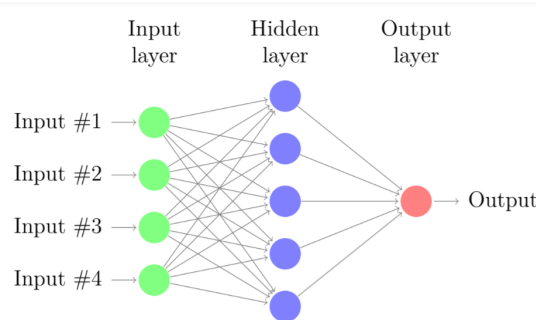
Figure 2.12: *Architecture of an artificial neural network with one hidden layer (retrieved from https://artineurons.com/introduction-to-neuron-network/)*

Grecu and Krajewski (2002) were one of the firsts that trained a neural network to nowcast rainfall intensities from the radar data on a 4 km^2 grid and 15 min time steps. They used a Back Propagation Neural Network (BPNN) and fed into it reflectivity-changes time series at each grid cell. That is for a specific region (a training quadrant), the predictors of each grid cell within the region (the reflectivity-change time series) were pooled to train the ANN and to keep a certain spatial consistency of the nowcasted field. The results were disappointing because there was not a significant gain in the predictability of the rainfall compared to the Lagrangian persistence.

Another type of ANN structure that is commonly used in pattern recognition, is the deep convolution neural networks (CNN). These are mainly used in image processing, where hidden layers consist of a group of neurons that convolve with a multiplication or another dot product. For instance, in the stereo-depth training, two images of a camera (mimicking left and right eyesight) with a specific resolution *(M X N)* are combined together in order to predict the true location of the objects inside them. A matrix with dimensions *MxNx2* is given as an input to the Neural Network and a filter of a fixed size *k* convolves in the matrix (one convolution over *kxk* pixels represents the input of a single neuron). The convolution is

repeated several times for the complete matrix, thus increasing the resolution of the image each time/level (each time/level of convolution represents a neuron layer). Through the training of the network, the weights inside the k -filter are changed and adapted such that, the predicted output and the observed output have a minimized cost function [Gerchow, 2018].

DispNet Architecture was developed by Mayer (2016) with the same flow as explained above, however the training for the data is not only done at the end level (reduce the error on the output image), but instead it proposes the minimization of cost function at each level when the k filter is applied. Thus ensuring that the coarser features are maintained and that no spatial inconsistencies are produced by the CNN. Such architecture was implemented successfully by Gerchow (2018) for the estimation of stereo depth from two images (left and right) for the application in micro laser surgery.

It would be interesting to investigate if such an architecture can be applied for rainfall nowcast where the inputs are the radar images at time t_0 and t_{-1}. Given that the CNN is able to learn from and keep the structure at different resolutions, it would be a suitable candidate for the rainfall nowcast, as it has already been observed that the coarse rainfall scales influence or model the fine scales. Thus, one of the objectives of this thesis is to check if such an architecture can extend the predictability of the rainfall. Is the CNN able to learn birth and death of the storms?

2.3.2   Nearest Neighbour Bootstrap Method

On the other hand, other methods like nearest neighbour has been used in hydrology for prediction (i.e. stream forecast based on rainfall, temperature and discharge [Galieti, 1990]). The idea of nearest neighbour is that similar events are caused by similar predictors, thus if one successfully identifies the predictors, then the hypothesis is that the events characterized by the same predictors have the same (or at least similar) behaviour. As described by Lall and Sharma (1996) a nearest neighbour approach, given an event to be forecasted, is applied in two steps; a) first from a group of past events, the k-most similar events are selected. Similarity of events is determined by the Euclidean distance of the predictors as per Equation 2.11; and b) the response/behaviour of the event at hand, is then the average of the k-most similar events as per Equation 2.12. The k – number giving the best value is learned by investigating a cost minimization function for different k.

$$E_d = \sqrt{\left(X_1 - Y_1\right)^2 + \left(X_2 - Y_2\right)^2 + \ldots + \left(X_N - Y_N\right)^2} \qquad (2.11)$$

where:
E_d – is the Euclidean distance
1,2,....N   – refers to the predictors used for the similarity estimation
$X_1, X_2, \ldots X_N$ – are the predictors of the event at hand
$Y_1, Y_2, \ldots Y_N$ – are the predictors from in-past observed event

$$R_{new} = \sum_{i=1}^{k} w_i R_i$$

where:
R_{new} – is the target response that one wants to forecast
k – is the number of similar neighbours selected
R_i – is the response from the ith neighbour
w_i – is the weight for the ith neighbour

(2.12)

The application of k-Nearest Neighbours (k-NN) seems valid when using the object oriented approach; it can be that similar storms are developing similarly and having more or less same life time. Thus, it can be an important tool in predicting the storm death and decay – decreasing so the high false alarm ratio associated with the Lagrangian persistence. One of the advantages of the k-NN is that, since different neighbours can be similar to the storm at hand, an ensemble prediction can be made by considering all of them instead of averaging the k-responses. Moreover, the errors due to the death or decay can be better identified and quantified. Hence, one of the objectives of this thesis is to check if such a storm based k-NN can be applied as an extension to the object based nowcast (namely HyRaTrac) to increase the predictability of the rainfall. If yes, what would be the best predictor set to recognize similar events? And what would be the best way to determine if events are behaving similarly?

Galieti (1990) successfully forecasted daily discharge data from the previous daily discharge, the rainfall daily volume and mean temperature. In such case the application of k-NN is quite fruitful because there is a clear (with physical background) connection between the input variables and the target discharge. However, when such information is not available, it can be challenging to identify the right predictors and hence have a good performance from the k-NN. The identification of the right predictors is bounded to the predictor set that one chooses to investigate. This task is an open research topic not only in the field of rainfall nowcast but overall in hydrology.

In non-parametric forecast models, the most common way of identifying a predictor is the presence of a strong correlation between the predictor and the target variable. Mutual Information [Fraser and Swinney, 1986] is another method widely used, that can help to express the statistical dependence of two variables (either if they have a linear or non-linear relationship). However, both of them consider the effect of only one predictor, and not the interaction between them (the partial dependence). As it might be that some predictors are left out just because they don't have a direct relationship with the target variable, but might improve the forecast if used in relationship with another predictor.

Thus Sharma and Mehrotra (2014) proposed a modified version of the Mutual Information, called Partial Information Content (PIC), that can measure partial dependency of the predictors set. The applied PIC metric is a number of synthetic generated data that included linear and non-linear relationship. The PIC metric (also referred to weights) gives values from 0 to 1 – one representing high dependency and 0 no dependency at all. The results look promising and such weights are advised to be used as well in a k-NN based forecast approach [Sharma et al., 2016].

Relative Importance Analysis (RIA), on the other hand, represents another way to understand the interaction between different predictors and the target variable, by studying how the variance of the target variable is explained based on the predictors' variance [Tonidandel and LeBreton, 2011]. RIA offers a global analysis to determine the variable importance given the prior hypothesis that the predictors and the target exhibit a linear relationship. Unlike the multi-regression weights, RIA considers the interaction among the predictors, but due to a transformation in an orthogonal space, the effect of co-linearity is removed.

Both RIA and PIC metric, can give direct weights of predictors importance but nonetheless since the RIA assumes a linear relationship, only the PIC weights can be used directly as input for the estimation of the Euclidean distance in the k-NN nowcast. However, RIA can provide useful information on how the predictors interact with each other, what are the most important ones, and which one can be discarded. Thus, in this thesis the information from the RIA will be integrated with the PIC, to estimate the right predictors and the respective weights. Additionally, the correlation weights will be used, to check whether the applied methods can provide better estimators than the most commonly used one.

3.   STUDY AREA AND DATA

This chapter gives an overview about the study area – Section 3.1, and available data – Section 3.2 that are used in this thesis to investigate the predictability of the storms by different nowcast methods. Section 3.3 summarizes the procedure employed for the identification of relevant rainfall events that would be suitable for the training and the validation of these methods. As the nowcast methods are expected to behave differently based on the event types, a simple classification scheme of the selected events is described in Section 3.4.

3.1 Study Area

The study area is located in northern Germany, as illustrated in Figure 3.1-a, and it lies within the Hannover Radar Range as shown in Figure 3.1-b. The Hannover radar is situated at the Hannover Airport and it covers an area with a radius of 115 km, including most of the Lower Saxony Region, the city of Bremen, the north-east part of Northrein-Westphalia, and eastern part of Saxony-Anhalt. As illustrated by the topography map provided by the German Federal Agency for Cartography and Geodesy (BGR), the area is mainly flat, with elevation barely reaching 200 m in the centre and northern part, while the south-west is characterized by few hills and the south-east by the Harz mountains that reach up to 1140 m a.s.l.

As the study area is in the proximity of the North Sea, the majority of the region (mainly the northern flats) is dominated by the maritime climate - characterized by cool summers and not so cold winters, with high humidity around the year and low sunshine duration in winter. In the southern part of the region, where the hills and the mountains are located, the climate is rather continental with dry summers and cold rainy winters.

The precipitation is mainly of a) stratiform type in winter - caused by the Westerlies Air Masses (from the Atlantic) and characterized by long rainfall with low intensities; b) convective type in summer - caused by the radiation instabilities which occur at local scales and for a short duration of time; and c) orographic type throughout the year - due to the presence of the mountains in the south-east part. Figure 3.2 illustrates the spatial variability of

the seasonal rainfall volume (averaged from 1983 to 2010) for the study region. Due to the presence of the orography, it is clear that the Harz mountains are receiving the most rainfall for both summer and winter, while the other regions (the northern flats) receive more rainfall in summer (due to convective processes) than in winter.
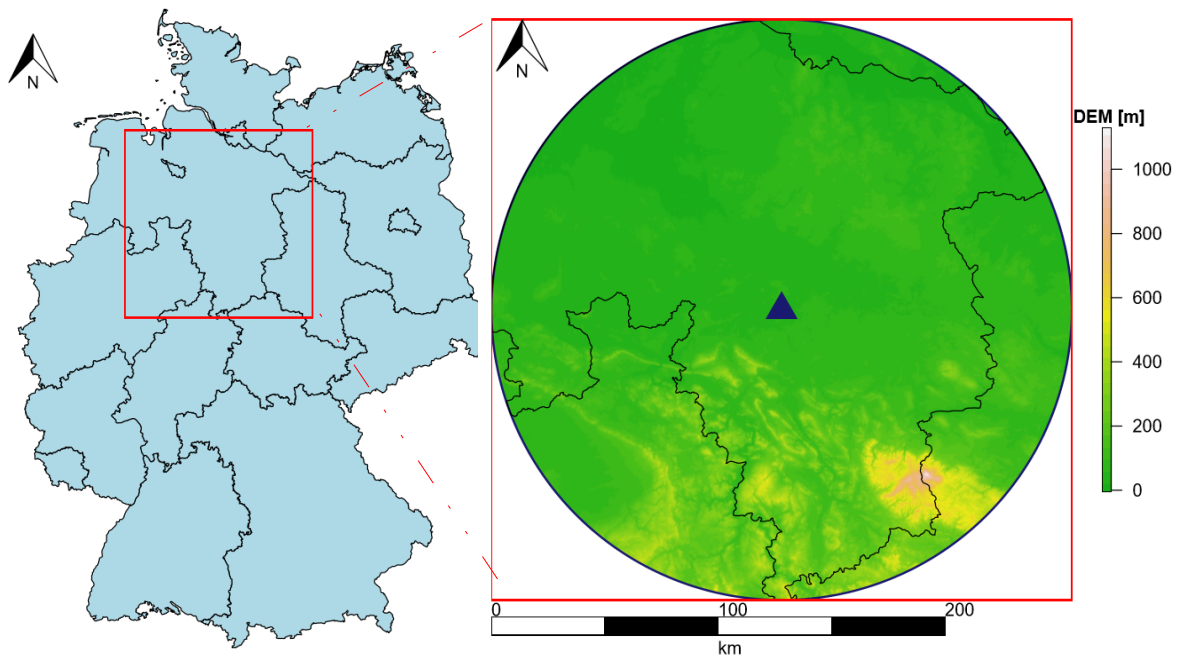
Figure 3.1: *The location of the study area a) within Germany and b) with the corresponding elevation and boundaries.*
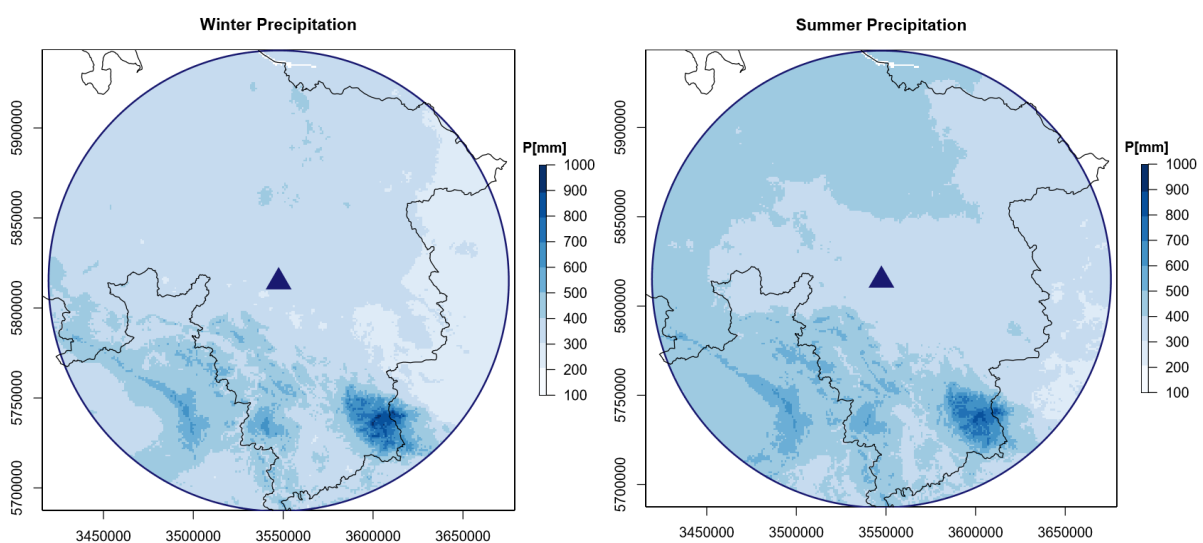
Figure 3.2: *The spatial variability of the long term (1983-2010) seasonal rainfall volume inside the study area according to Deutscher Wetterdienst (DWD) data.*

3.2 Data Available

As indicated in Section 2.1.2 the two types of rainfall measurements adequate for urban flood forecast and that are used on this study are gauge recordings (discussed in Section 3.2.1) and radar scans (discussed in Section 3.2.2). Both measurements are managed by the Deutscher Wetterdienst (DWD) and are available for the public since 2017 under the server: https://opendata.dwd.de/climate_environment/CDC/. For the purpose of this study as illustrated by Figure 3.3 were employed 100 recording stations (shown in dark red) and 1 radar recording (shown in dark blue).
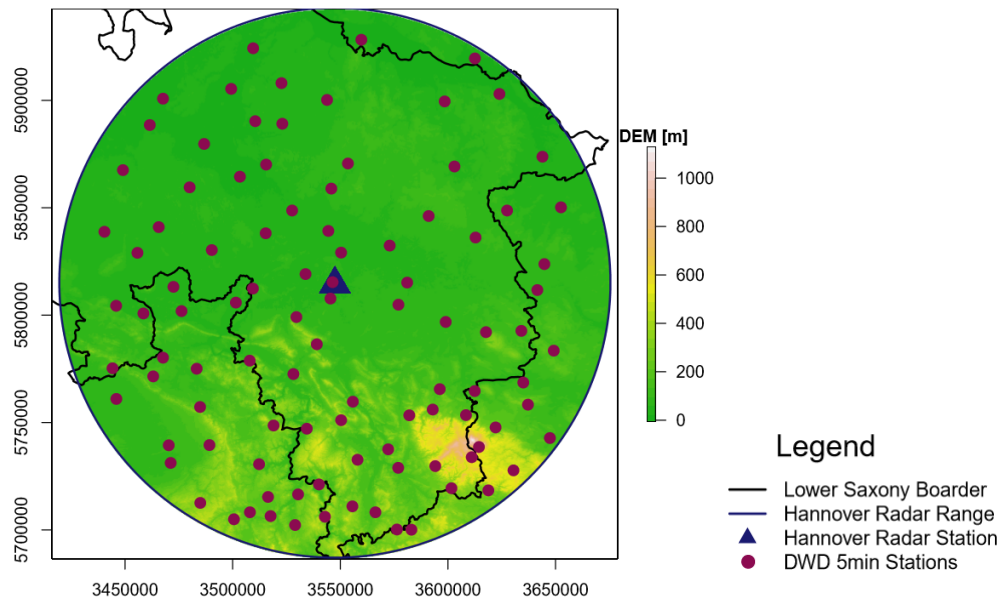
Figure 3.3: *The available recording stations (red) and radar (blue) inside the study area.*

3.2.1   Station Data

The gauge data used in this study are tipping bucket devices that record the rainfall at 0.1 mm accuracy at 1 min resolution. Since the data provided by the DWD are raw data, and only an automatic check was performed prior to data delivery, the 1 min data are first aggregated to 5 min time steps and then are further checked for plausible values. Attention is given that the extreme rainfall at different periods doesn't exceed the highest recorded in the region. In dry periods where it is clear that there is no sudden high rainfall, the values are set to 0, while for the other cases are set to missing values.

3.2.2   Radar Data

The radar used in this study is of C-band that gives reflectivity measurements at an azimuth resolution of 1° and every 5 min scans. The reflectivity measurements were converted to intensity according to the Marshall-Palmer relationship with the constant coefficients of a= 256 and b=1.42 according to the Equation 2.4. For the correction of the radar data, a simple

static clutter correction according to Berndt et al. (2013) was employed. The procedure is as follows:

> **Step 1:** Random years of radar measured intensity were summed up (2 or 3 years) in order to identify the static clutters. Here static clutters are considered as zones where the rainfall is either too low (zones with rainfall lower than 500 mm) or too high (zones with rainfall higher than 2800 mm). The location of the erroneous beam was estimated visually as well by the accumulated volume and is treated as a clutter.

> **Step 2:** Once the clutters were identified, the inverse distance interpolation method was employed to fill in the rainfall information from the surrounding grid cells.

> **Step 3:** The corrected intensities from step 2 are converted to a Cartesian Coordinate system with a 1 km^2 resolution.

The radar data obtained are used as a basis for the later procedures and for the merging with gauge data, and are referred to as raw radar product (RR). No further correction is done to the radar data, as it is assumed, that the merging with the gauge information, will handle the other types of errors present in the radar measurements.

3.3 Event Selection

Since the radar data are available only from 2000 and on, the observation period considered in this study is from 2000 to 2018. As the precipitation nowcast is case dependent (depending highly on the storm type and its' distinctive movement and transformation), it is important to consider separate events, instead of continuous time series. Thus, the aim is to extract events that are relevant for the study and give a good overview of convective and stratiform mechanisms. To select the relevant events, first the gauge data were consulted and then the radar one.

The procedure for the extraction of the events is illustrated in Figure 3.4. First, for each of the 100 recording stations the extreme rainfall volumes for different durations were extracted. As indicated in Figure 3.4 upper-left, a sliding window with the span of the duration under investigation (in the example the duration is 15 min), is used to calculate the rainfall volumes accumulated under the duration. For each rainfall event, the maximum volume (PE) is selected and used to build the partial extreme time series. Rainfall extremes are recognized from one another by a dry spell duration (DSD$_{min}$) longer than the sliding window. The partial extreme time series is not only useful because it helps to recognize the most extreme storms, but also by fitting a GEV (Generalized Extreme Value) in the partial series, it gives a tool to measure the severity of storms given a specific duration (as shown in Figure 3.4. lower-left).

Nevertheless, the extreme value analysis indicates extreme events happening at a point location, but since the interest is in spatial events, it is important to recognize events that are happening around the whole radar region. To account for the region, areal averages (for the whole area) were computed for each 5 min time steps based on the 100 stations. The average time series was then used to identify the wet period (WSD) – time steps where rainfall intensity (PI) is bigger than a minimum intensity threshold (WSImin=0.05 mm); and the dry periods – time steps with PI less than 0.05mm, in the region as shown in Figure 3.4.-right. The wet periods were distinguished from one another if a dry period (DSDmin) of more than 4 hours is present [DWA, 2012].

Once the areal events were identified, only the ones with the pre-recognized extreme events at local scale were selected for the further evaluation. In the final step, the radar data were employed to check whether or not a rainfall event was recognized by the radar, and hence if spatial information was available. Thus an event, starts when the average areal intensity is higher than 0.05 mm and ends when there is no rainfall present in the region for more than 4 hours.

Based on the procedure a total of 100 events were selected for the further evaluation of the methods. A detailed description of the events, with the duration and extremity, is illustrated in Appendix A – Table A.1. Since, as discussed in Section 2.3 the predictability limit is expected to vary based on the event type, it is important to distinguish between them when analysing the results. For this purpose, a classification scheme was used to distinguish between types of events as described in Section 3.4. The set of 100 events is then separated into two datasets: i) 2000-2012 dataset used for estimating the limits of rainfall nowcast predictability and developing the new nowcast methods (see Section 3.4.1), and ii) 2013-2018 dataset used for the final comparison of the nowcast methods under severe convective events for the urban flood nowcast application (see Section 3.4.2).
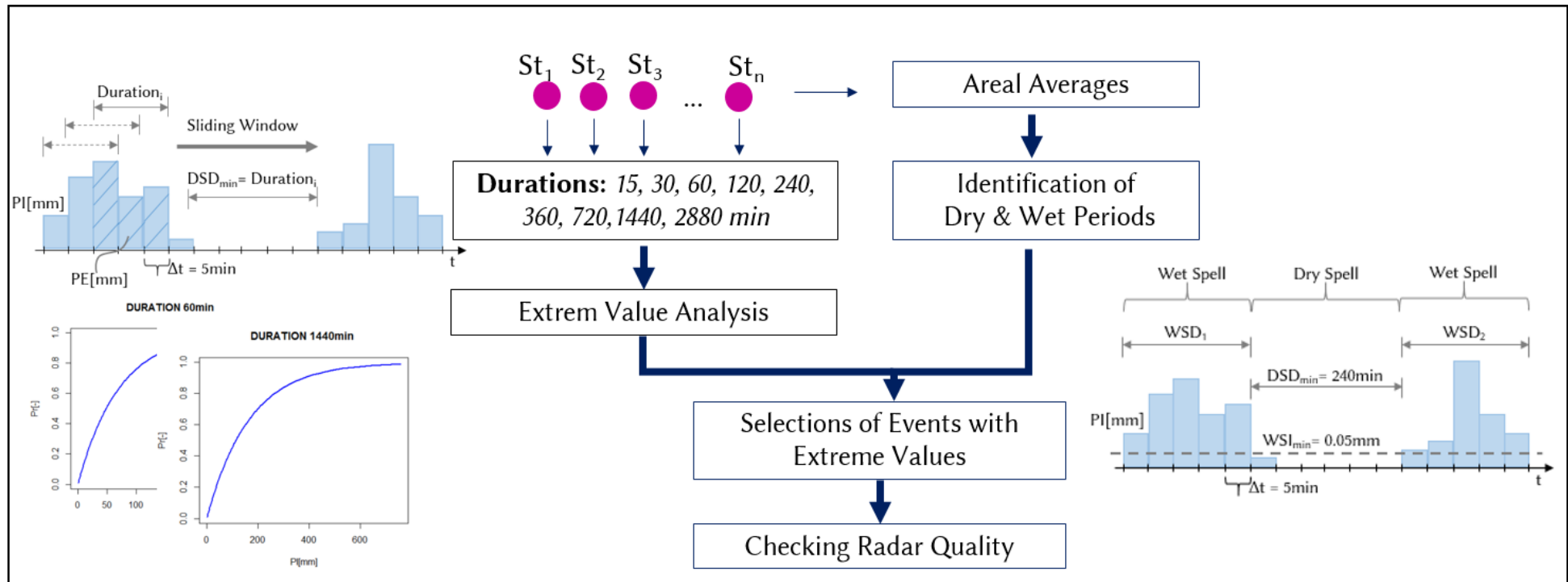
Figure 3.4: *The procedure for the selection of relevant events; a) left - the selection of extreme events per each station is illustrated. Cumulative Probability Distributions of partial extremes are built for each station and duration to know the severity of the storms at each station; right) the procedure for the selection of the areal events is illustrated. For each recognized areal event, the extreme events at each station for different duration are consulted, in order to select only the most relevant ones. Once the relevant events are selected, the radar data is consulted; if the radar data are available for the given periods and if the quality of the measurements is good enough for the nowcast.*

3.4 Event Classification

In the classification scheme proposed here (illustrated in Figure 3.5) four variables were employed. The variables were calculated for the event – the start and the end of the wet period considering the whole radar range, and individual storms – which here is referred to as a group of rainy grid cells recognized by the radar observed on different time steps. Hence one event includes many storms, with different duration and areal extents.

1- ***Maximum Rainfall Intensity*** (PI_{max}) observed in the ***station data*** [mm/5min] during each event – According to Ehret (2003) cold and warm fronts can reach intensities up to 60 mm/h and not higher. However, since the fine resolution peaks are generally higher than the coarse resolution ones, we consider 7 mm/5min as a threshold for the separation of the convective and stratiform events.

2- ***Wet Area Ratio (WAR)***: Percent of event time steps where the fraction of the ***radar image*** that has rainfall intensity higher than 0.05 mm/5min, is more than 45% - According to Ehret (2003) warm and cold front can be distinguished by a WAR higher than 45%, while a convective one by a WAR lower than 8%. However, it can be than the convective events are spread on the whole radar range, thus the WAR itself cannot provide a full classification, that is why the duration and the maximum extend of singular storms were considered as well.

3- ***Maximum Storm Area*** in each event from ***radar data*** [km^2] – According to Ehret (2003) convection and meso-scale convection cannot reach extents higher than 10^3 km^2, thus this can be used as a limit to distinguish between convective and stratiform events. Nevertheless, since the investigation is done on a single radar station, the extent of the storm is limited by the radar boundaries.

4- ***The Longest Duration*** of all singular storms observed in each event from ***radar data*** – According to Table 2.1, the convective storms typically live between 30 min and 4 hours. They seldom live longer. That is why the 4 hours (360min) duration was used as a threshold to identify storms that are convective or stratiform.

Based on these characteristics, events are distinguished into 4 groups: a) stratiform, b) simple convection, meso-γ supercell convection, meso-β wide spread convection, and mixed events. As depending of the event duration considered, it may be that both stratiform and convective storms are present during an event, thus a clear separation is not possible. Therefore a new group is introduced – the mixed events. The classification of the events was then done according to 5 rules as explained below and as illustrated in Figure 3.5.

Rule 1: IF WAR is bigger than 45% for at least half of the event duration, the maximum storm duration within the event is longer than 4 hours, and the maximum intensity observed by the stations is lower than 7mm/5min, the event is classified as stratiform (colour code green).

Rule 2: IF maximum storm duration is shorter than 4 hours and the maximum storm area is lower than 10^2 km², the event is classified as convective (colour code light blue).

Rule 3: IF the maximum storm duration is shorter than 4 hours, the maximum area is higher than 10^2 km², and the maximum observed intensity is higher than 7 mm/5min, the event is classified as meso-γ scale convective (colour code blue).

Rule 4: IF the maximum storm duration is shorter than 4 hours, the maximum area is higher than 10^2 km², and the maximum observed intensity is lower than 7 mm/5min, the event is classified as meso-β scale convective (colour code dark blue).

Rule 5: IF the maximum storm duration is longer than 4 hours, and the maximum observed intensity is higher than 7 mm/5min, the event is classified as a mixed event (colour code turquoise).
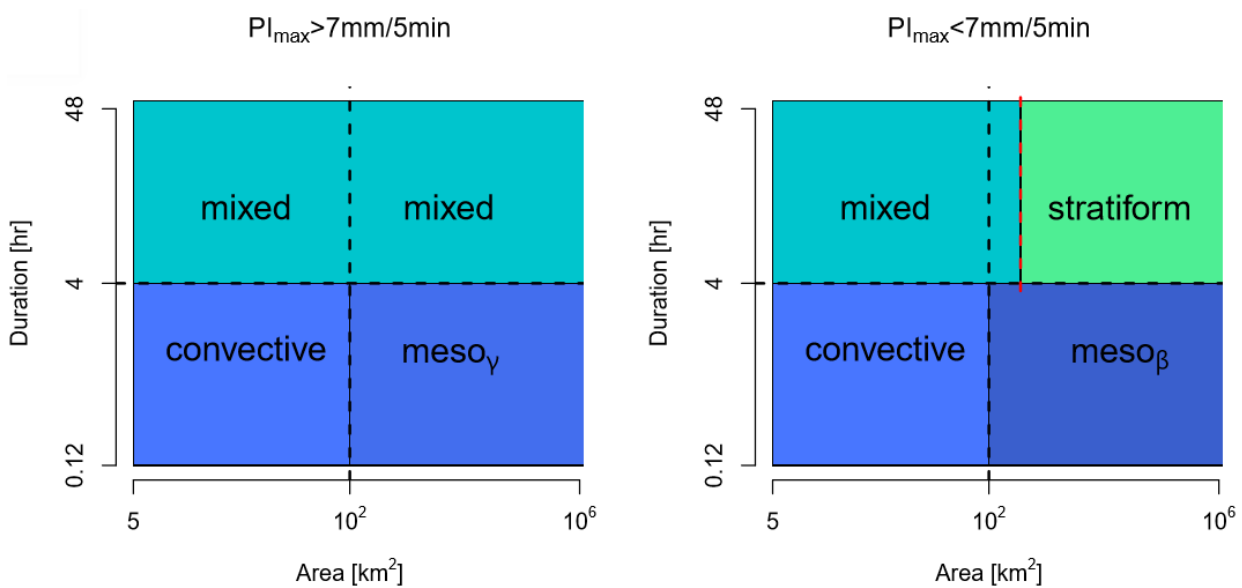
Figure 3.5: *The schematic diagram for the event type classification for events with maximum observed intensity higher (left) and lower (right) than 7mm/5min based on Maximum Storm Area (x-axis), Maximum Storm Duration (y-axis) and Coverage of Wet Area Ration (the red line).*

3.4.1   Events for the Evaluation of Rainfall Nowcast

In total 93 events were recognized by the procedure in 3.3 for the period from 2000 to 2012. These events are to be used for the training, validation and evaluation of the rainfall nowcast algorithms. From the 93 selected events – 13 were classified as stratiform, - 14 as convective, –14 as meso-γ convective, –25 as meso-β convective and – 27 as mixed events. The properties of each group of events are shown in Figure 3.6. As expected the most severe events (regarding urban floods) are within the convective and meso-γ convective events, while the

mixed events can also exhibit high intensities, but they share as well characteristics with the stratiform events.
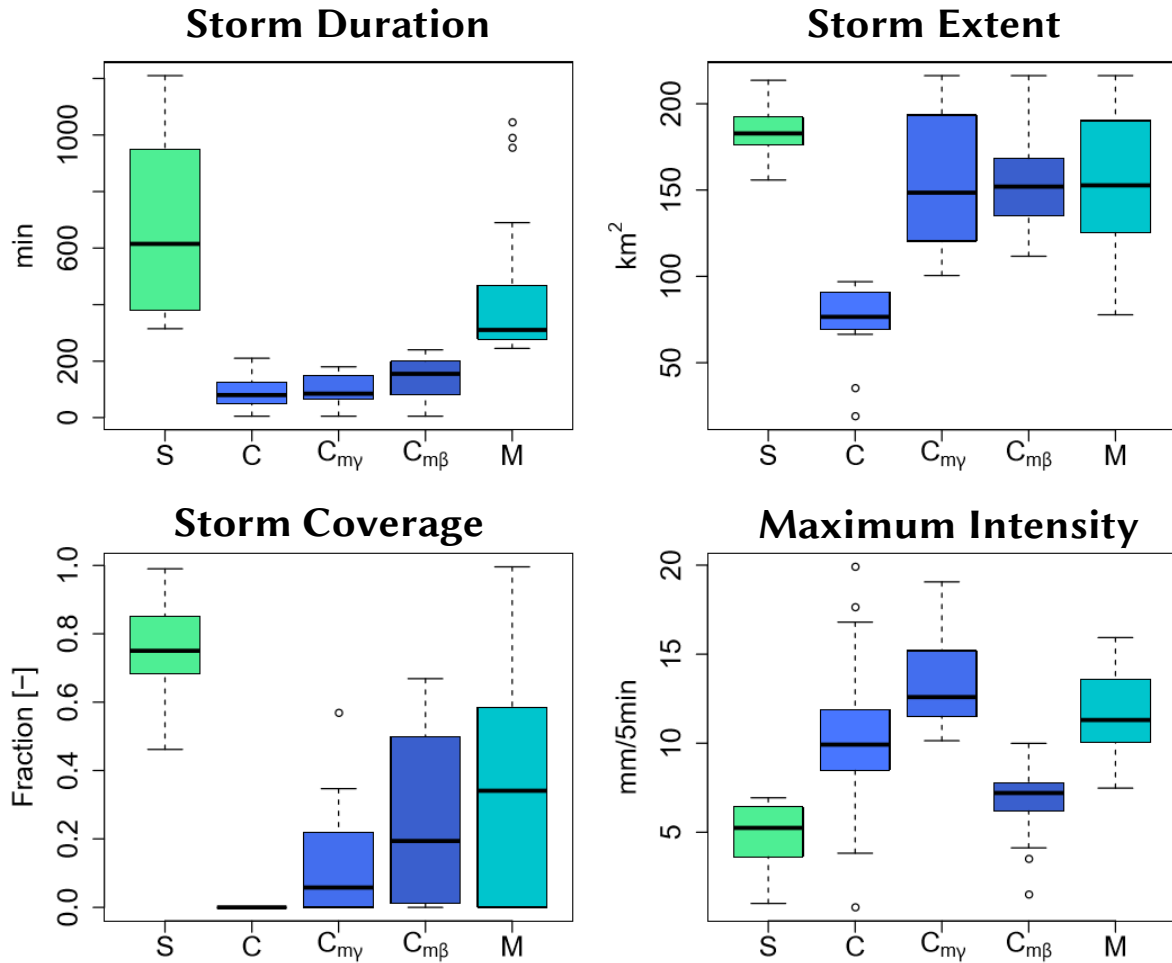
Figure 3.6: *The maximum storm duration (up-left), maxim storm area (up-right), WAR event coverage (lower-left) and maximum gauge intensity (lower-right) of the selected events grouped in a) stratiform (green), b) simple convective (light blue), meso-γ convective (blue), meso-β convective (dark blue) and c) mixed (turquoise).*

3.4.2 Events for the Evaluation of Urban Flood Nowcast

As discussed in Section 2.1.2 meso-scale convective events are the main cause for the urban floods. Even though some of these events have been selected to test the predictability and to train new data-driven nowcasts, it is important to test these methods under distinctive events that can cause urban pluvial flood. For this purpose, a total of 17 events were selected from the period 2013-2018 that exhibited extreme rainfall volumes for the short duration (60 to 120 min corresponding to the concentration time typical of the urban catchment). In total 17 extreme events were recognized - 5 as simple convective, –7 as meso-γ convective, and – 5 as mixed events. The properties of each group of events are shown in Figure 3.7.
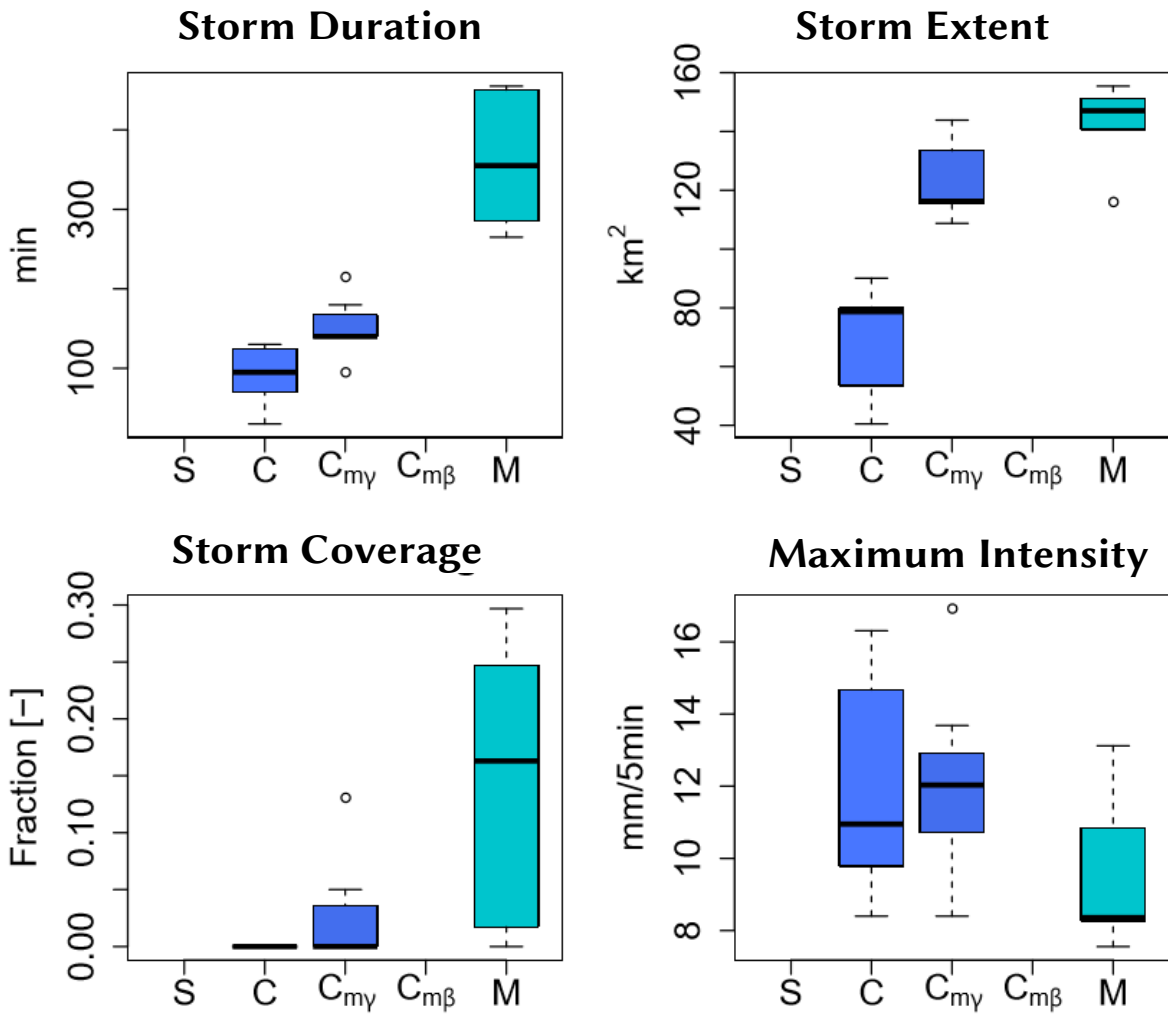
Figure 3.7: *The maximum storm duration (up-left), maxim storm area (up-right), WAR event coverage (lower-left) and maximum gauge intensity (lower-right) of the selected events for the urban flood simulation grouped in a) stratiform (green), b) simple convective (light blue), meso-γ convective (blue), meso-β convective (dark blue) and c) mixed (turquoise).*

4.   REFERENCE METHODS

This chapter gives an overview about the performance and the predictability limits of the existing nowcast methods for the Hannover study area and for different event types. The selected nowcast reference methods, introduced in the literature review, are described shortly in Section 4.1, followed up by the predictability and performance measurements for different thresholds in Section 4.2. The existing gap between radar and gauge data is illustrated in Section 4.3 and finally in Section 4.4 an overview of the predictability limits and performance of the reference methods for different storms is given, and conclusions are derived about the need and room for improvement of the existing nowcast methods.

4.1 Existing Nowcast Methods

The Lagrangian persistence is used as a reference for the estimation and improvement of the predictability. Eulerian persistence is not considered here, because as mentioned in the literature review, it is outperformed by the Lagrangian one. The two types of Lagrangian radar based nowcast considered here are; i) the object based approach represented by the HyRaTrac model [Krämer, 2008] and the field motion approach represented by Lucas-Kanade optical flow method [Lucas and Kanade, 1981]. The algorithms of the methods are illustrated shortly in Figure 2.6 and Figure 2.7 from the Section 2.2.3 of the literature review.

4.1.1   HyRaTrac Settings

As already mentioned HyRaTrac performs the nowcast in three steps; I) identification of the individual storms, ii) tracking of the individual storms, and iii) extrapolation of the individual storms by the Lagrangian persistence. The nowcast is initialized when a wet time step is present; when a group of 64 radar grid cells has an intensity higher than the initial threshold of Z = 20 dBz. Afterwards in each wet time step, an individual storm is recognized if a group of grid cells exceeds a certain threshold. To account for both types of storms, two thresholds are set; 20 dBz for more than 128 radar pixels (stratiform), and 25 dBz for more than 16 radar pixels (convective).

Once a storm is recognized and tracked, the advection vectors are estimated by maximizing the cross-correlation between the two radar images; the current image at time t, and the

previous image at time t-1. First global velocities are calculated for the whole radar image, to have a first estimate about the direction of advection, then local velocities for each storm region[10] are calculated. If a storm is just identified (so no previous storm state is known) then the local advection is set to the global one. To account for missing time steps, a 15min window is given; this means that if a previous time step or the current time step is missing, the information used for the advection in the last 15 min is still used as a basis for the nowcast.

Once the advection is determined, the extrapolation of the storm in the future is done based on the Lagrangian persistence as shown in the following Figure 4.1. The advection is not applied at the centroid of the storm, but at the storm region. This means that the corner coordinates of the storm are displaced by the advection vector. The nowcast is given every 5 min for the total lead time of 3 hours.

a) time t_0 b) time t_{+LT}
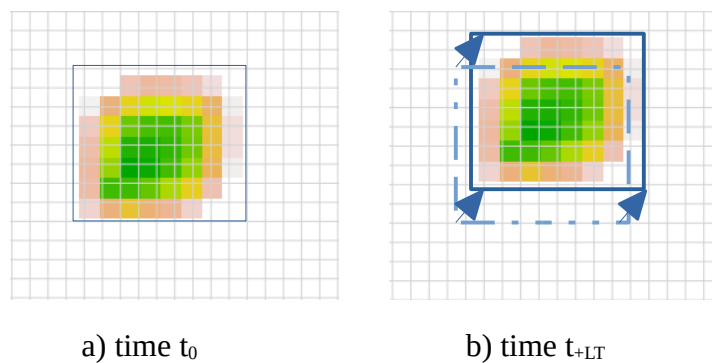
Figure 4.1: *The extrapolation of a storm region (blue box) as last observed at time t_0 (when the nowcast is issued), and in the future with a specific lead time (t_{+LT}) based on the local advection vectors (blue arrows). The storm region belonging to the previous time step is indicated in a dashed line.*

4.1.2   Lucas-Kanade Flow Settings

As with HyRaTrac a threshold of 20dBz was used to recognize the rainfall structures of the radar images. Unlike HyRaTrac there is no limit about the group of cells that are higher than this threshold. Instead all the radar pixels that are lower than this threshold are set to 0. The optical flow algorithm according to Lucas-Kanade method is then employed to determine the flow field based on the last 15 min; current time step t, and previous time steps t-1 and t-2. The extrapolation on the rainfall field in the future is done based on the Lagrangian persistence for every 5 min until a lead time of 3 hours. For the implementation of the Lucas-Kanade method, the pysteps package in python was used [Pulkkinen et al., 2019].

---

10   The storm region is the region within the minimum and maximum X and Y extent of the storm.

4.2 Measurements of Predictability

As discussed in Section 2.3, the predictability can be distinguished between the theoretical and the practical one. In this study the focus is only on the practical predictability, as the models at hands are far from perfect and furthermore in the existing nowcast methods the nowcast uncertainty due to the initial conditions cannot be estimated. Thus, an overview of the theoretical predictability is not possible. However, the practical predictability can still be investigated by comparing the nowcast either to the gauge or to the radar observations.

As the final scope is the urban flood nowcast, and urban flood models are calibrated typically with gauge data, a comparison of the nowcast with the gauge observation will give an overview on how good are the radar based nowcast for the point scale application. This predictability will be referred here as the gauge (practical) predictability. As the gauge predictability is greatly affected by the existing gap between radar and gauge measuring scales, it is important to determine as well the predictability of the model itself independent from the error between gauge and radar data. This predictability is referred here as the radar (model) predictability and it is calculated based on the input radar data fed to the nowcast model.

Theoretically, an adequate estimation of the predictability would be possible if both gauge and radar data are available continuously in space and time. For the radar data this is the case, however for the gauge observations this is not possible. One can interpolate the gauge observation to a continuous space, but this might introduce a new uncertainty in the estimation of the nowcast predictability. It is for this reason that the predictability of the storms is computed only at the gauge location (as illustrated in Figure 4.2) for both gauge and radar predictability. The radar predictability can be calculated in a continuous space, but to ensure a fair comparison between the two predictabilities, the calculation is kept the same as for the gauge one.
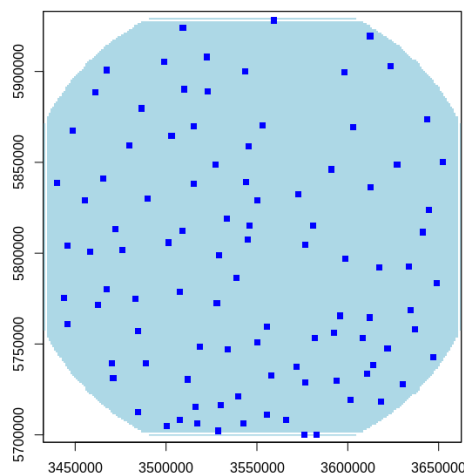
Figure 4.2: *The location of the reference points (corresponding to the gauge location) for the performance assessment of the radar products and the nowcast models (shown in dark blue). The extend of the radar measurement is illustrated in light blue.*

Based on the literature review, three measurements were proposed for the estimation of the predictability limit of the nowcast for specific lead time, when the cross-correlation falls below 0.37, when the Equitable Threat Score (ETS) falls below 0.3, and the area under the ROC curve is smaller than 0.5. The ROC curve works for probability nowcasts and as such is left outside at the moment because none of the reference methods can produce ensemble nowcast. The ETS gives a good overview of the predictability where certain events are at focus; for instance, there is a specific intensity that triggers an alarm. However, at the moment it is not clear what 5 minute intensity can trigger an urban flood (or other accumulation volumes), and since this is site specific, the ETS is also left outside of the investigation. The decorrelation time seems reasonable for the urban flood purpose as, in the urban model the rainfall distribution is quite important. Thus, here in this study the decorrelation time as explained below is used as a measurement for the predictability limit.

a) **_Decorrelation time_** (τ_{LT})– the lead time when the cross-correlation between nowcast and observation data becomes lower than 0.37.

$$\tau \; when \; \rho_{For,Obs} = \frac{\sum_{i=1}^{n} \left(For_i - \overline{For} \right) \left(Obs_i - \overline{Obs} \right)}{\sqrt{\sum_{i=1}^{n} \left(For_i - \overline{For} \right)^2} \sqrt{\sum_{i=1}^{n} \left(Obs_i - \overline{Obs} \right)^2}} \leq \frac{1}{e} \approx 0.37 \tag{4.1}$$

where:
n   – number of observed values
Obs_i – the respective i^{th} observed value
For_i – the respective i^{th} nowcasted value
$\overline{For}$ – mean of the nowcasted values
$\overline{Obs}$ – mean of the observed values

The decorrelation time gives the information up to which lead time the nowcast issued is useful. Typically for the nowcast assessment, in the literature, the cross-correlation in space is calculated. For instance, at a given time step, the cross-correlation of the nowcast field with the observed field is calculated, and if one desires to have the cross-correlation value for an event with a specific duration, the space cross-correlation is averaged over all the time steps of the event as illustrated in Figure 4.3-a and in Equation 4.2. The cross-correlation computed in space and then averaged over time, indicates how well the nowcast can capture the spatial variability of the rainfall.

$$Performance_{spatially} = \frac{\sum_{ts=1}^{N} f_{ts} \left(For_{st=1}^{100}, Obs_{st=1}^{100}, \right)}{N} \tag{4.2}$$

where:
N   – number of time steps in an event
st   – the station reference from 1 to 100
Obs – the observed value
For – the nowcast value
f_{ts} – the performance function computed spatially for a time step
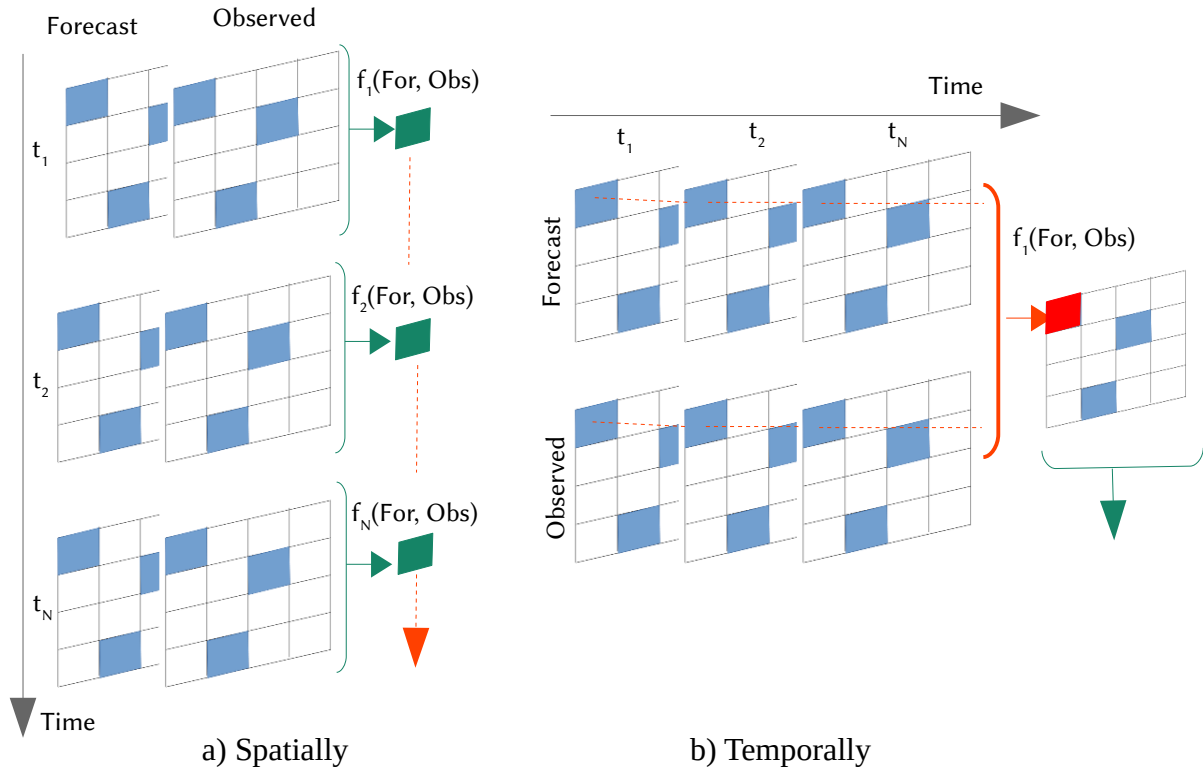
Figure 4.3: *Nowcast performance assessment a) spatially - forecast and observed are compared spatially (in green) over all grids (in blue) then averaged temporally (in red) over the event time steps; b) temporally - at each grid (blue) the forecast and observed are compared temporally (in red) over all event time steps and then averaged spatially (in green) over all the grids.*

Nevertheless, the cross-correlation in space doesn't give any information how well the nowcast can model the temporal variability of the rainfall at a point scale (i.e. the temporal variability of the gauge). To have an idea about this, the cross-correlation can be calculated at each gauge location for each lead time between the nowcasted and observed time series (over the whole duration of the event) as illustrated in Figure 4.3-b. Here in this study this is referred to as the temporal correlation, and can be averaged over the whole group of gauges as indicated by Equation 4.3.

$$Performance_{temporally} = \frac{\sum_{st=1}^{100} f_{st}\left(For_{ts=1}^{N}, Obs_{ts=1}^{N},\right)}{100}$$

(4.3)

where:

N   – number of time steps in an event
st   – the station reference from 1 to 100
Obs – the observed value
For – the nowcast value
f_{st} – the performance function computed temporally for a station

Moreover, for the urban flood nowcast the high intensities are particularly of great interest because they are the ones causing the surcharge. Typically for the nowcast assessment, low rainfall thresholds have been set, like 1, 5 or 10 mm/h, which for 5 minute lead to intensities

less than 1 mm. Additionally to these limits, here it is also investigated the 60 mm/h (5mm/5min) intensities when calculating the space and temporal cross-correlation.

Even though the cross-correlation can give an idea about how good the nowcast is, and until what lead time the nowcast is useful, it doesn't give enough information about the rainfall nowcast error that is fed to the urban model when nowcasting urban floods. Thus, evaluation of continuous errors produced by the nowcast for the whole time series at a point scale is necessary. This can be illustrated for instance by the root mean square error of the complete time series or the error in the time and volume of the peak:

b) **Root Mean Square Error (RMSE)** for the whole time series

$$RMSE\,[mm/5\,min]=\sqrt{\frac{\sum_{i=1}^{n}(For_i-Obs_i)^2}{n}}$$

where:
n   – number of time steps
Obs – the observed time series at a location
For – the nowcasted time series at a location

(4.4)

c) **Peak Relative Bias (Bias)** for the maximum rainfall volume observed in 30 min.

$$Bias\,[\%]=\frac{\sum_{i=1}^{n}(\sum_{DUR=30\,min}^{max}For_i-\sum_{DUR=30\,min}^{max}Obs_i)}{\sum_{DUR=30\,min}^{max}Obs_i}$$

(4.5)

where:
N   – number of time steps in an event
st   – the station specific value
Obs – the nowcast for the respective observed value
For – the mean of the nowcasted values

d) **Peak Time Shift (TP$_{Shift}$)** – the time difference between the maximum 30 min peak from nowcast and from observation.

$$TP_{Shift}\,[5\,min]=t_{\sum_{DUR=30\,min}^{max}For_i}-t_{\sum_{DUR=30\,min}^{max}Obs_i}$$

(4.6)

where:
i   – number of time steps in an event
t   – the time of peak for a station
Obs – the observed time series
For   – the nowcasted time series

d) **The Variance Ratio (RVAR)** – the ratio between the variance of the nowcasted time series and the variance of the observed time series.

(4.7)

where:
st   – the station specific value
Obs$_{st}$ – the observed time series at a specific station
For$_{st}$ – the nowcasted time series at a specific station

$$RVAR\,[\,]=\frac{Var(For_{st})}{Var(Obs_{st})}$$

4.3 Discrepancy between Radar and Gauge

The radar based nowcast can be as good as the radar input fed to the nowcast algorithm. Therefore before looking at the nowcast performance, it is important to understand what is the input data error; hence the radar to gauge discrepancy. To investigate this, both spatial and temporal performance criteria for all events types are derived. Figure 4.4 illustrates the cross-correlation calculated both temporally (upper graph) and spatially (lower graph) for the five given thresholds and grouped according to the five event types.
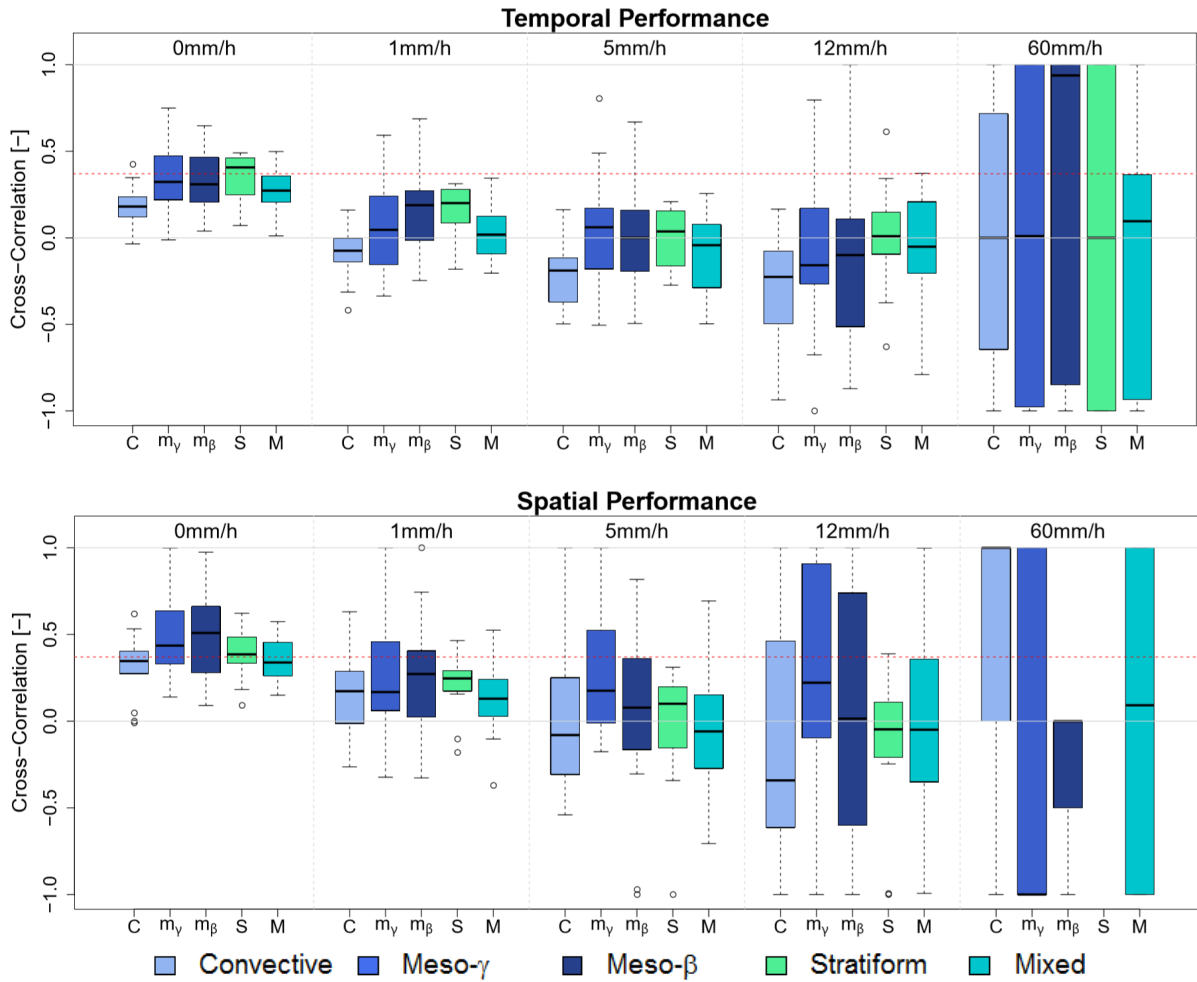
Figure 4.4: *The cross-correlation between the radar and gauge measurements calculated for each threshold: temporally (averaged over the 100 stations) and spatially (averaged over the event time steps) for each event. The red dashed line indicates the value 0.37 taken as a reference for the predictability limit.*

It is clear that the cross-correlation tends to decrease with a higher threshold, as the discrepancy between radar and gauge becomes higher. This is the case for the temporal correlation. For thresholds higher than 1 mm/h the discrepancies are so big, that the correlation for all the events is mostly below the predictability limit (0.37). Hence when comparing the nowcast with the gauge data, one cannot expect a predictability at all, as the input data cross-correlation is lower than the limit. Moreover, as expected the correlation is higher for the stratiform events and the lowest for the simple convective events. This may be

caused by the wind drift and the mismatch between the measuring scales of radar and gauges, as also documented by Berndt et al. (2013). Overall it should be expected that the stratiform events yield higher predictability than the convective ones, at least for the lowest threshold. What is somehow surprising are the results for the 60 mm/h threshold. For some of the convective events it seems that the timing and the variability of the high thresholds are met, and one may expect good predictability results. However one has to keep in mind, that the time steps and stations were the threshold is higher than 60 mm/h are much fewer than the other thresholds, thus affecting the performance calculation.

Contrary to the temporal performance, the cross-correlation in space indicates that the radar data can capture better the spatial rainfall variability, as generally the values exhibited by all the event types are higher than the one exhibited by the temporal performance. At least for the lowest threshold the values of most events are above the predictability limit, suggesting that the nowcast may show good skills when calculating the spatial performance. Overall, the meso-scale convective events seem to be better represented spatially by the radar. On the other hand, stratiform, mixed and simple convective events apparently suffer from higher errors. This can either be attributed to the wrong coefficients of the Z-R relationship (mainly in the case of the convective type), or attenuation (in the case of the stratiform events) or either to the storm distance with the radar and the effect of the earth curvature (for both event types).
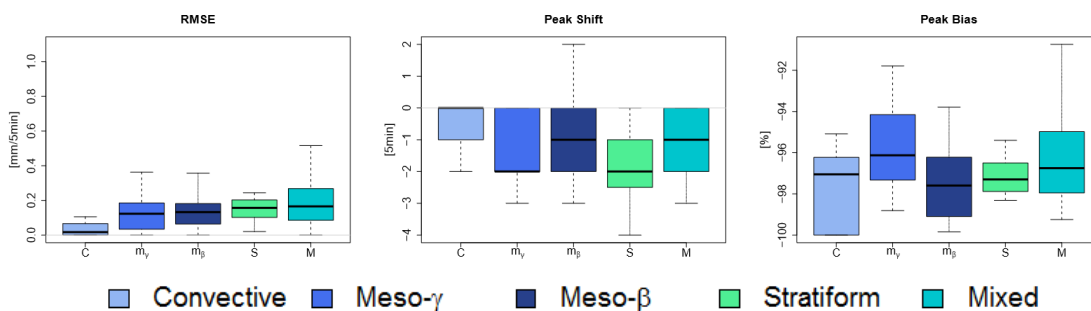
Figure 4.5: *The temporal performance: RMSE (left), shift in peak time (middle) and the peak bias (right) averaged over the all stations for each event type. No threshold is set for the calculation of these criteria.*

To investigate how suitable are the radar data to be fed in an urban flood nowcast, the temporal performance at the point scale, grouped for each event type, is illustrated in Figure 4.5. It seems that the RMSE is quite small and acceptable when no threshold is set (> 0 mm/h). When looking to the peak shift, it is clear that a consistent bias on the time of peak is present (approximately 10 minutes delay). This agrees with the time discrepancy as reported by the other studies due to the presence of the wind. The shift in peak looks consistent throughout different events and thresholds, but is mainly bigger in stratiform events. Moreover, when looking at the volume of the peak, the radar data is underestimating strongly the rainfall intensities, causing a systematic underestimation up to 100% of the peak volume. As expected the RMSE is increasing with the threshold (Appendix C- Figure C.3),

and for the 60 mm/h threshold even though the correlation was above the correlation limit, the RMSE is quite big and with little difference among the event types. It is evident that when nowcasting urban floods, the radar data error will propagate in the urban model, most probably causing a big underestimating of flood volume, since the rainfall peak is completely missed.

Overall, it is clear that without the smoothing of the radar data and their correction based on the gauge intensities, the gauge predictability will be almost zero; for the lowest threshold (0mm/h) it may exhibit a certain predictability degree, however for intensities higher than 1 mm/h the correlation between radar and gauge is lower than 0.37, hence the nowcast will have no predictability at all. The results suggest that a smoothing in time and space of the radar data should be done before the merging in order to ensure a better estimation of the peak. The smoothing should be done for a window of at least 10 min, in order to overcome the measuring gap (due to the wind drift) between the gauge and radar data. Such smoothing will help with a better prediction of the peak time and peak volume. However, the smoothing itself will not be enough to overcome the errors of the radar measurements. Thus, posterior to the smoothing, a proper merging technique should be selected such that the radar and the gauge discrepancies are minimized, resulting in a higher temporal performance. Nevertheless, such methods may come to the expense of the spatial performance, but since the focus is urban flood nowcast, priority is given to the temporal performance.

4.4 Predictability Limit

The raw radar data are fed to the reference nowcast models HyRaTrac and Lucas-Kanade, and the results for each are shown respectively in Sections 4.4.1 and 4.4.2. A direct comparison of the two models is given in Section 4.4.3, in order to understand, which of the method may give better results when fed with not corrected radar data.

4.4.1   HyRaTrac Predictability

Figure 4.6 illustrates the cross-correlation computed both temporally (upper) and spatially (lower) between the HyRaTrac nowcasts at different lead times and the input radar (left) or gauge data (right) for each event type.
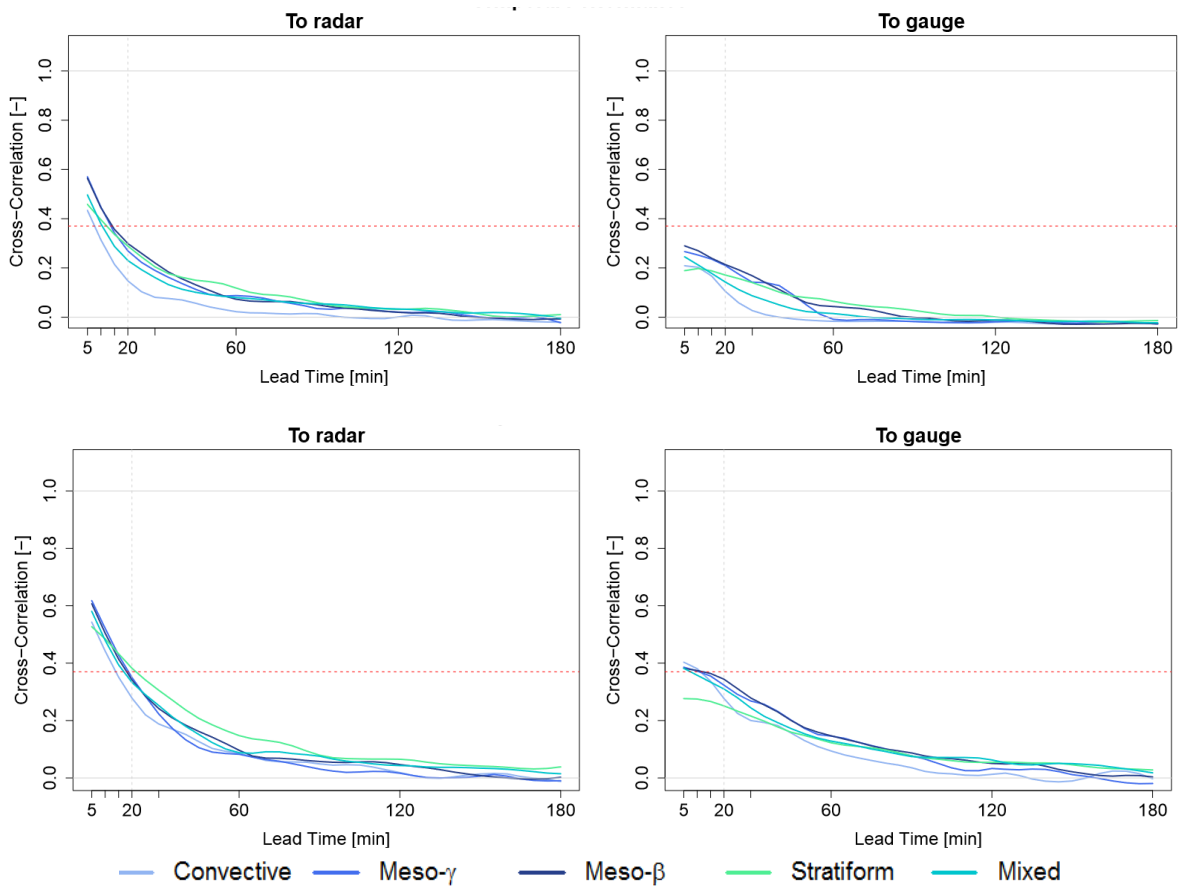
Figure 4.6: *The event median cross correlation for each event type as nowcasted by HyRaTrac for different lead times, calculated temporally – upper row (averaged over 100 station per event) and spatially – lower row (averaged over the time steps per event). The red dashed line indicates the correlation value of 0.37 for the predictability limit, while the grey dashed line is a reference to the 20 min predictability limit as suggested by the literature. A threshold of 0 mm/h was used to calculate the cross correlation. Performances for higher rainfall intensities are given in Appendix B – Figure B.1 and B.2.*

The temporal radar predictability limit here is 15 min for stratiform and meso-scale convective events, and 10 min for the other event types. It seems that the predictability limit for the meso-scale convective events is higher than for the convective, mixed and stratiform events. For the small convective events, this is to be expected as they have relatively short life, while for the stratiform since they have a large coverage, the errors due to the lack of information at the boundary of the radar may become so big that it affects the correlation. Whilst the meso-scale events, that are not so large but with higher intensities, can be distinguished and tracked better by HyRaTrac. This is consistent also with the results from Grecu and Krajewski (2000) that claim that the nowcast is better with higher intensity storms.

The spatial radar predictability is higher than the temporal one, reaching up to 20 min for all the events, except the simple convective one (whose limit is approximately 15min). This is as well consistent with the results from Section 4.3, as the spatial cross-correlation between radar and gauge is higher than the temporal one, thus it makes sense that the decorrelation time is longer for the spatial performance. Moreover, the exhibited radar predictability agrees well with the one in the literature, and suggests that the HyRaTrac is at the same line as the state of the art methods. The rate of decrement in the spatial performance is smaller than the temporal one, suggesting that the spatial patterns can be modelled better by the HyRaTrac. Particularly for the stratiform events, the rate is smaller than the others, indicating that these events can be best nowcasted by HyRaTrac, even though the spatial cross-correlation between radar input and gauge is not the highest among the event types. Consequently, the stratiform events, as they are more persistent in time, are easier to be tracked and nowcasted.

Whilst when compared to the input data, the nowcast model exhibits some predictability, this is not the case for the gauge one. As expected, since the radar to gauge correlation is quite low, the gauge predictability is almost non-existent, for both spatial and temporal performance. That suggests again that the radar based nowcasts are not suitable at all for the urban flood nowcast application.

To get a complete overview on the temporal performance of the nowcast on the point scale, the temporal RMSE, peak shift and the peak bias of the nowcasted time series at different lead times compared to both radar and gauge data, for all of the event types, is illustrated in Figure 4.7. The RMSE based on both radar and gauge data displays no big difference along different lead times for all the events; even for the short lead times (for example 5 or 10 minutes) the RMSE is not considerably smaller than for the long ones. The RMSE compared to the radar data, as expected, are lower than the ones compared to the gauge data. There is a difference between the performance of the events, for instance the RMSE of the stratiform nowcast to radar is lower than the one of the convective, however this has to do more with the event characteristics (like duration, intensity and persistence) rather than with the nowcast model. However, what is interesting is that, the storm types having the lowest RMSE compared to the radar are not the same ones having the lowest RMSE when compared to the gauge. This suggests that the radar to gauge errors are more dominant than the model errors, causing no gauge predictability at all.

Figure 4.7: *The median of the temporal performance: RMSE (upper), Shift in Peak Time (middle) and Peak Bias (lower) for each of the event types of the HyRaTrac nowcast with different lead times, compared to the input radar data (left) and gauge data (right). The red dashed line indicates the limit performance value for each of the criteria, while the grey dashed line is a reference to the 20 min predictability limit as suggested in literature. Here no intensity threshold was used.*

Unlike the RMSE, the peak shift for both radar- and gauge-compared performance is getting worse with longer lead times. Surprisingly for the convective events when comparing the peak with the gauge observation the shift is zero, indicating that the nowcast can capture the peak for such events but when looking at the peak bias it is completely underestimated. Actually this is the case for all the event types, no matter the lead time, the underestimation of the peak volume when compared to the gauge data is always 100%. This suggests once

again, that even though the nowcast to radar may be useful for the first lead times, the radar to gauge errors are so big that the nowcast is useless. At this point the correction of the radar data is a must!

When considering the radar predictability both is space and time (see Figure 4.6), stratiform events appears to be better predicted than the others, however when considering the temporal performance of the peak at the point scale, the nowcast of stratiform events exhibits a very high peak shift in comparison to other events. This suggests that the predictability limit (especially the spatial one) may not be always a good measure for the temporal behaviour at point scale. Generally speaking, there is a link between the temporal cross-correlation and the temporal performance. However, for the first lead times, even though the temporal cross-correlation is higher than the limit, the temporal peak bias and the peak shift are already very high, suggesting that the nowcast model is not appropriate for the point scale. If this can be improved by feeding merged data to the model or by choosing another nowcast model, is yet to be seen.

4.4.2   Lucas-Kanade Predictability

Figure 4.8 illustrates the cross-correlation computed both temporally (upper) and spatially (lower) between the Lucas-Kanade (LK) nowcasts at different lead times and the input radar (left) or gauge data (right) for each event type. As seen by HyRaTrac, the nowcast predictability when compared to the gauge is way below the correlation limit. It is clear that as radar data is underestimating systematically the rainfall intensities, when high intensity threshold are set, the correlation becomes zero with little difference between the event types. This emphasizes again the need for radar correction to gauge data.

Regarding the comparison with the radar input, as seen in HyRaTrac, the temporal radar predictability limit is 15 min. However, there is a slight difference on event predictability. At LK it seems that the events with the highest correlation are the small convective and the stratiform events. As the LK is a field motion approach it is expected to yield better results for the stratiform events which are wide spread. For the small convective events, as they are characterized by smaller intensities and maybe by lower speed and rate of change, the nowcast works well for a short period and then it losses its predictability quite fast; which can be related to the life span of the small convective cells. It is still unclear why the temporal cross-correlation after the first hour of nowcast starts increasing for both convective and meso-β convective events.

Similar results are also exhibited by the spatial cross-correlation, with a slight increment of radar predictability up to 20 min for the lowest threshold. Convective and stratiform events, here as well, are the ones with the highest correlation, although they have different decrement rates over the lead times. For the convective events, as expected, the drop in predictability is quite fast, due to the storm death which is not considered by the model. Whilst for the stratiform events the drop in predictability is slower, suggesting that as the stratiform is more

persistent in time, it is better nowcasted by the optical flow. Nevertheless, the spatial cross-correlation is dropping at 20 min lead time below the predictability limit. This suggests that even for the stratiform events (not only the convective ones at different scales) the storm growth or decay is important to capture. Moreover, one has to keep in mind that the drop in the stratiform predictability can also be caused by the radar boundaries as there is no continuous flow of rainfall information (as illustrated in Figure 2.7).
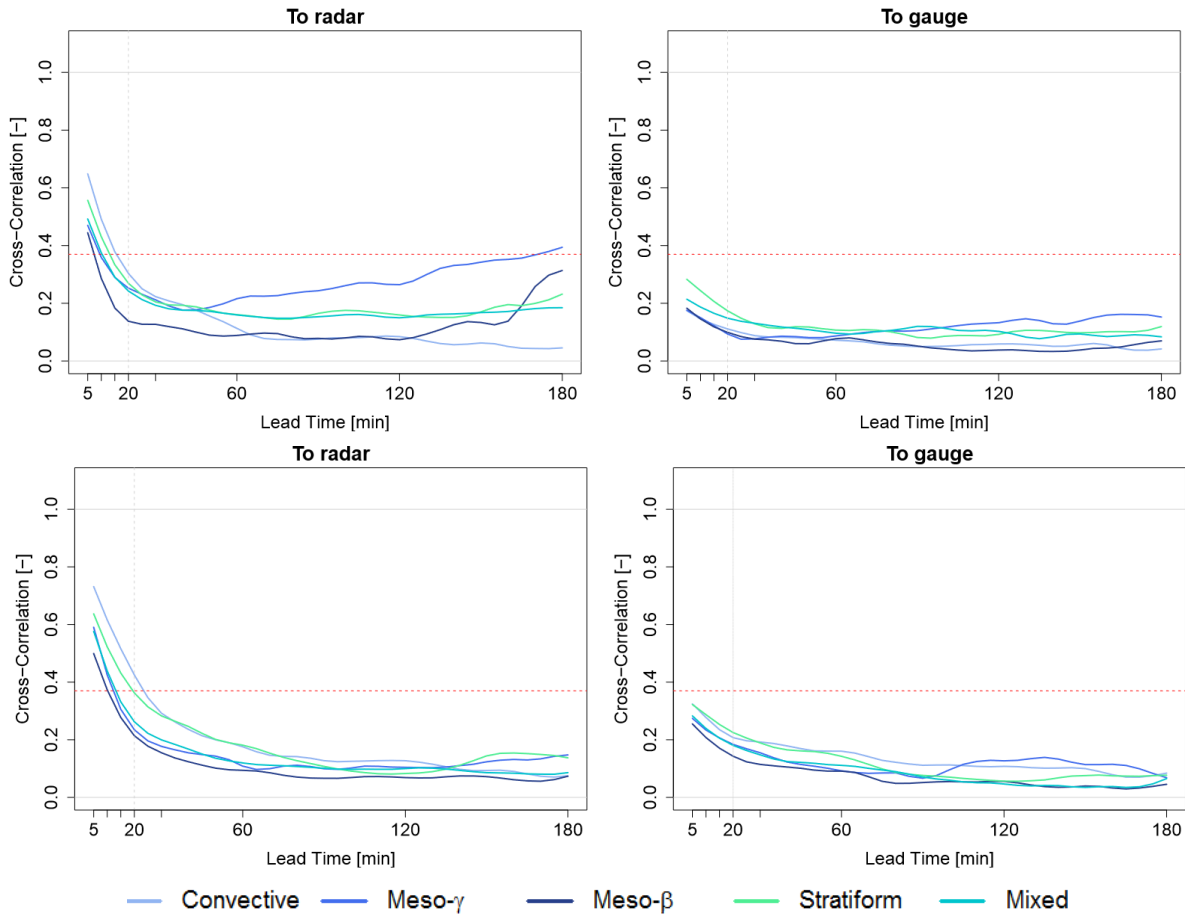
Figure 4.8: *The event median cross correlation for each event type as nowcasted by LK algorithm for different lead times, calculated temporally – upper row (averaged over 100 station per event) and spatially – lower row (averaged over the time steps per event). The red dashed line indicates the correlation value of 0.37 for the predictability limit, while the grey dashed line is a reference to the 20 min predictability limit as suggested by the literature. A threshold of 0 mm/h was used to calculate the cross correlation. For performance of higher intensity thresholds see Appendix B – Figure B.3 and B.4.*

To get a complete overview on the temporal performance of the nowcast on the point scale, the temporal RMSE, peak shift and the peak bias of the nowcasted time series at different lead times compared to both radar and gauge data, for all of the event types, are illustrated in Figure 4.9. As seen with HyRaTrac, the RMSE based on both radar and gauge data displays no big difference along different lead times for all the events. The RMSE compared to the radar data as expected are lower than the ones compared to the gauge data, but generally both are higher than the HyRaTrac error. As in HyRaTrac, the storm types

having the lowest RMSE compared to the radar are not the same ones having the lowest RMSE when compared to the gauge. At LK as well, the radar to gauge errors are dominant to the model errors, causing no gauge predictability.
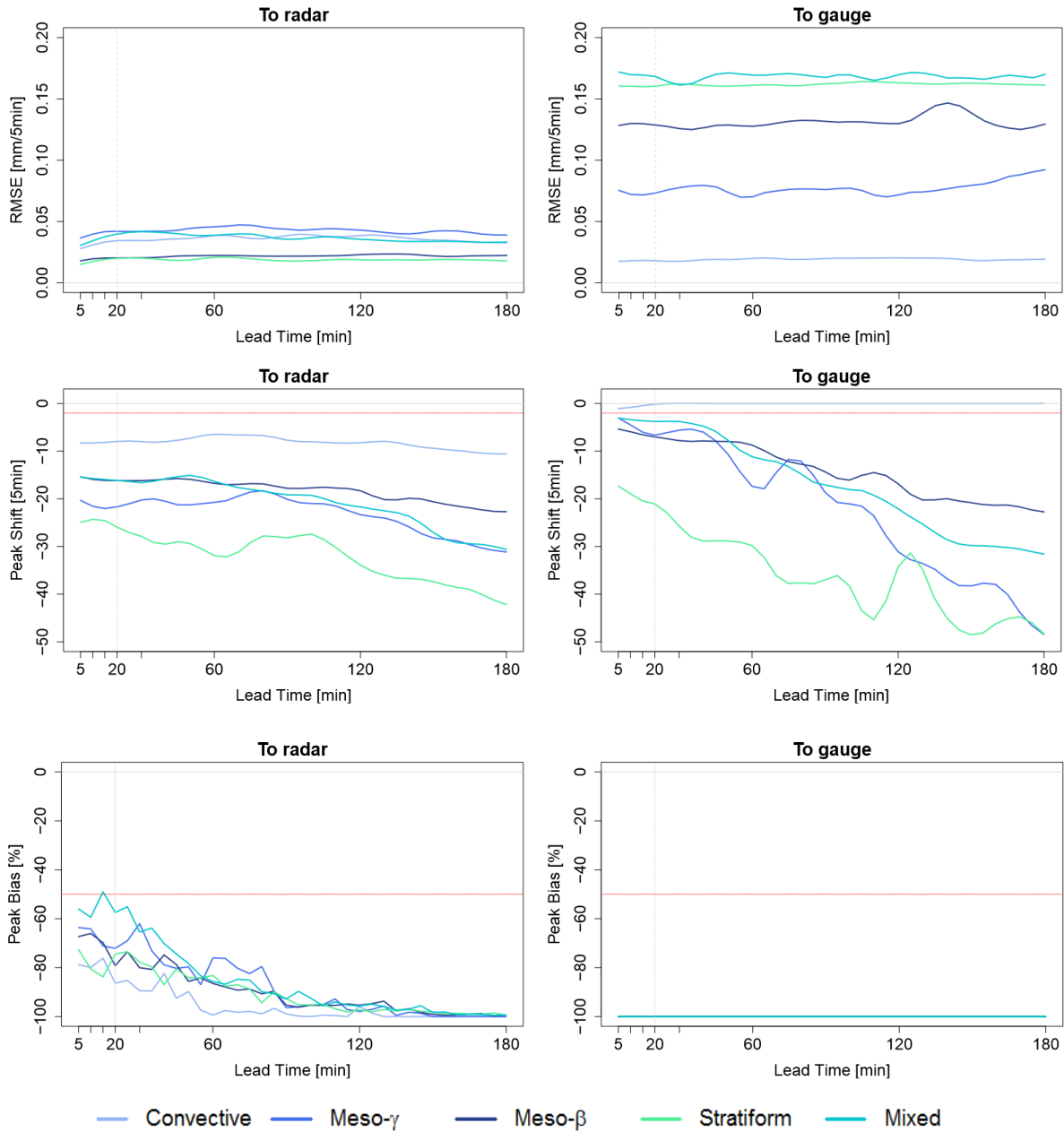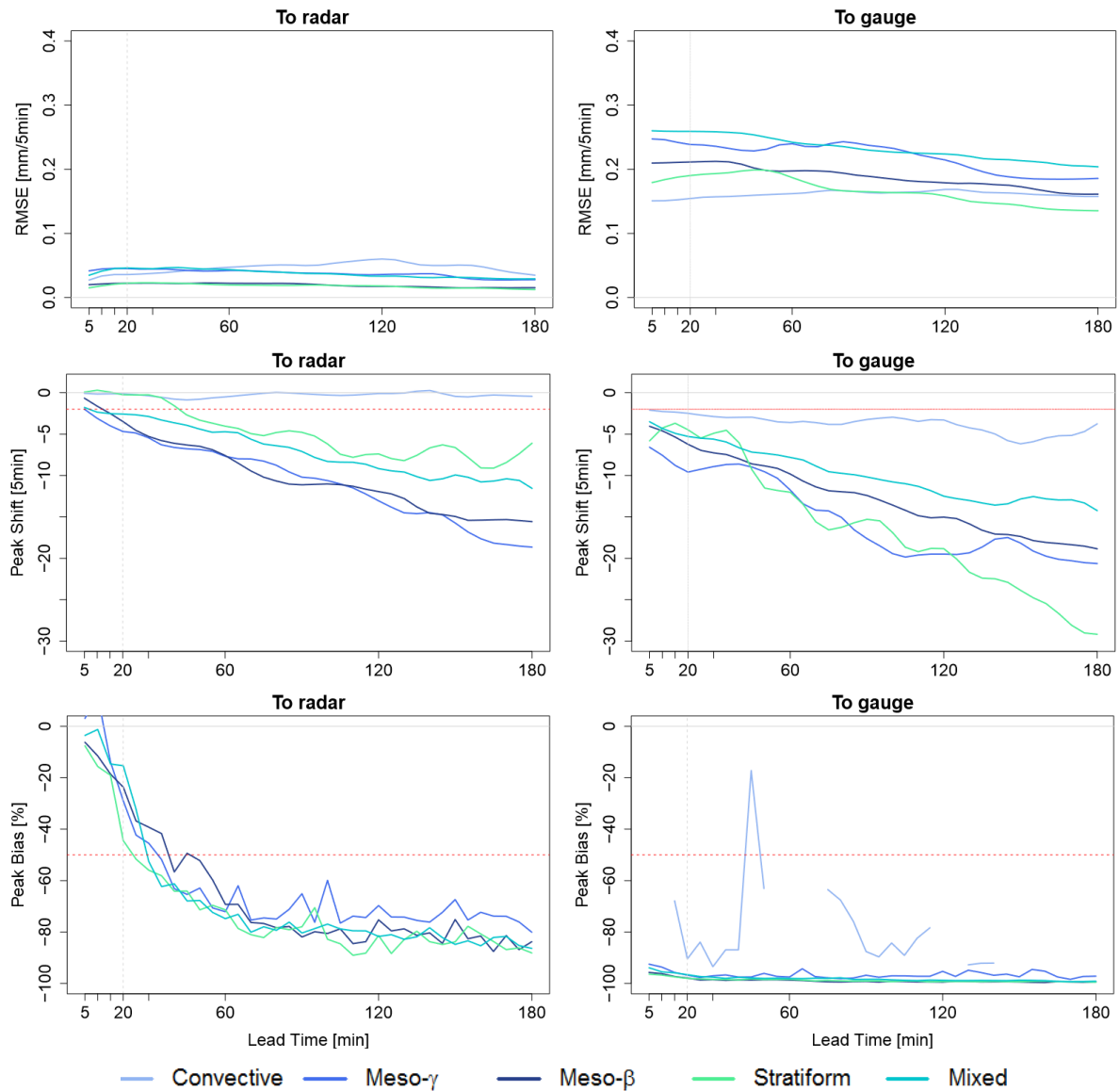

Figure 4.9: *The median of the temporal performance: RMSE (upper), Shift in Peak Time (middle) and Peak Bias (lower) for each of the event types of the LK method with different lead times, compared to the input radar data (left) and gauge data (right). The red dashed line indicates the limit performance value for each of the criteria, while the grey dashed line is a reference to the 20 min predictability limit as suggested in literature. Here no intensity threshold was used.*

Unlike HyRaTrac, the peak shift is increasing clearly with the lead time (for both nowcasts compared with radar and gauge), but the first lead times the peak shift is nearly 0. It seems that the peaks for the convective and stratiform events are better captured and reproduced, at least for lead times up to 40 minutes. Apparently the information from three past time steps used for advection (the case of LK Method) is more useful than the two past time steps information (the case of HyRaTrac) for the estimation of the temporal patterns at point scale.

Moreover, the peak bias compared to radar, doesn't start from a 50% underestimation like in HyRaTrac, but instead it starts from 0-10% underestimation. However, this advantage seems valid only until 30 min lead time, afterwards the behaviour is the same as in HyRaTrac. This is to be expected, as the advection estimation plays a role only on the short lead times, while at the longer ones, the inclusion of growth, decay and death is necessary for a better performance.

4.4.3   Comparison between HyRaTrac and Lucas-Kanade

Figure 4.10 and Figure 4.11 give a direct comparison of the temporal and spatial predictability limits of the two nowcast methods (HyRaTrac and Lucas-Kanade) when compared to radar and to the respective gauge information averaged over all the events. The performance of the nowcasts are quite similar, mainly due to the poor quality of the radar data. For the short lead times, it seems that HyRaTrac has slightly better temporal cross-correlation than the LK nowcast, however the performance of HyRaTrac decreases more rapidly than the LK method. At longer lead times, the LK yields better results than HyRaTrac, but the temporal correlation is still below the predictability limit. The same behaviour is also seen at the spatial cross-correlation. Overall the cross-correlation value for both temporal and spatially performance is decreasing rapidly with the increasing threshold. For both methods the radar predictability limit seems to be 10-15 min temporally and 15-20 min spatially. Regarding the gauge predictability limit, at both types, is lower than 5 minutes.

Even though the HyRaTrac shows better predictability performance (both spatially and temporally) than LK method, the latter one has still the advantage of modelling better the temporal variability of rainfall at the point scale. For this reason, the LK is still included in the study, to see whether or not, an improvement on the radar input can improve as well the predictability of both methods, and if so until what lead time. It is clear that both of the methods are limited due to the implementation of the Lagrangian persistence and as dictated by the literature, can provide useful results up to 20 min. An inclusion of storm birth, growth, decay and death is necessary in order to increase the radar predictability limit even further on.
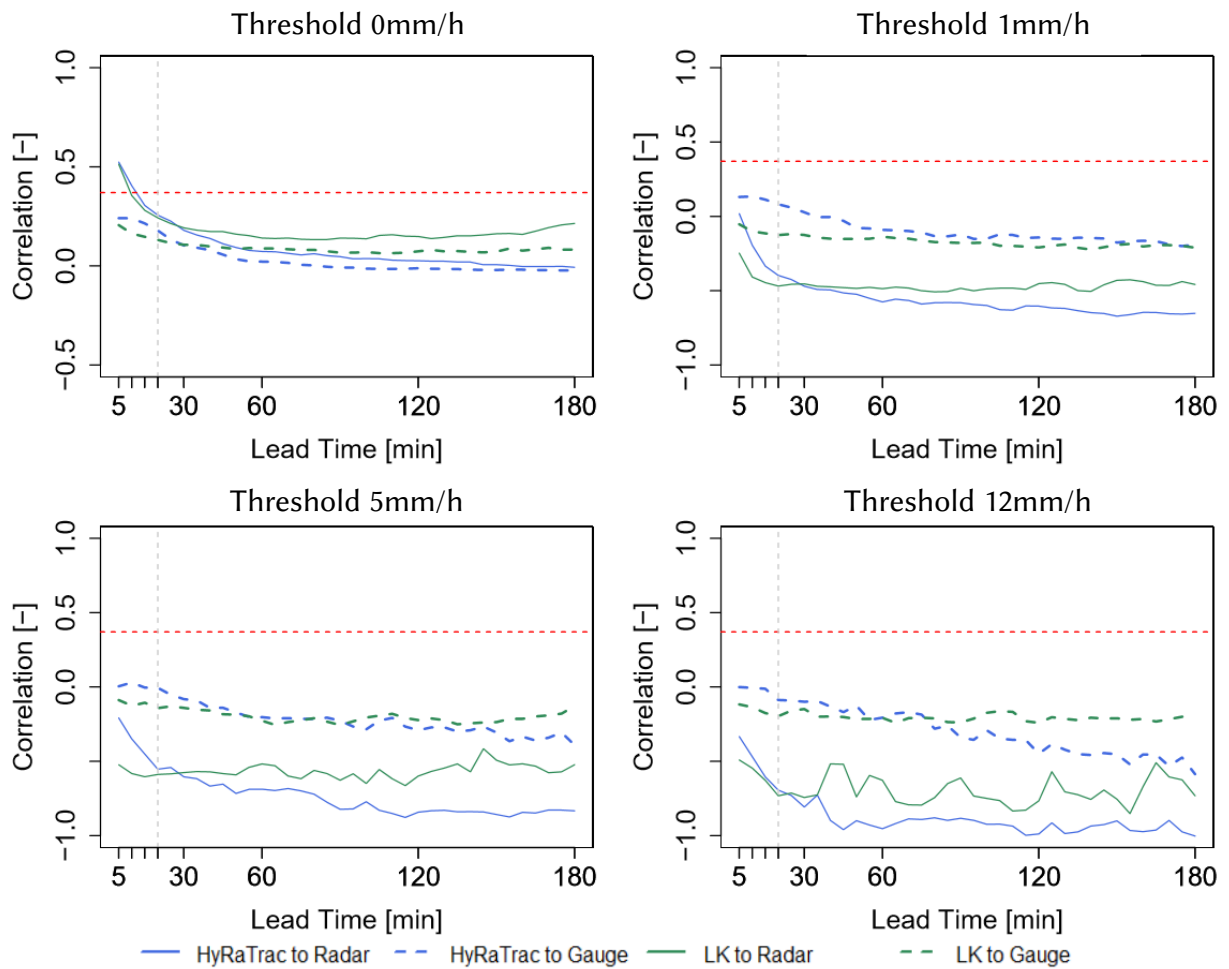
Figure 4.10: *The average temporal cross-correlation over all the events for the HyRaTrac and LK Nowcast for the lead times from 5 min to 180 min (every 5min) for intensies higher than 0 mm/h, 1 mm/h, 5mm/h and 12 mm/h calculated based on raw radar input (solid lines) and gauge reference (dashed line). The red dashed line indicates the correlation 0.37 considered as the predictability limit, while the grey dashed line indicates the limit predictability as shown in the literature (20min).*
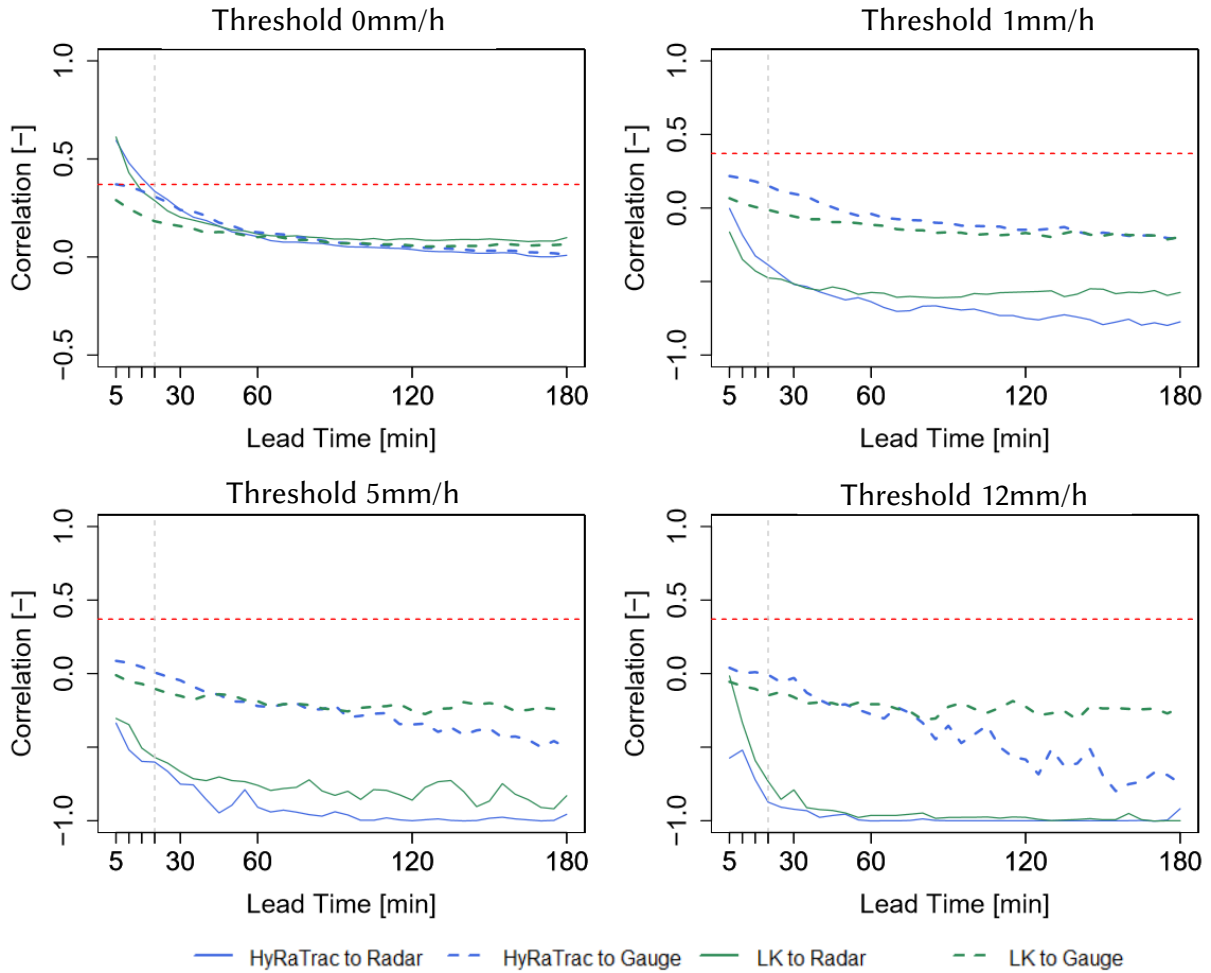
Figure 4.11: *The average spatial cross-correlation over all the events for the HyRaTrac and LK Nowcast for the lead times from 5 min to 180 min (every 5min) for intensities higher than 0 mm/h, 1 mm/h, 5mm/h and 12 mm/h calculated based on raw radar input (solid lines) and gauge reference (dashed line). The red dashed line indicates the correlation 0.37 considered as the predictability limit, while the grey dashed line indicates the limit predictability as shown in the literature (20min).*

5.   MERGING DIFFERENT DATA SOURCES

As already mentioned in the literature review, radar data are prone to numerous sources of errors and hence are not as accurate as the gauge recordings. On the other hand, the gauges provide point information, and even though one can get the spatial information by interpolation, the rainfall structures are smoothed out and high peaks are typically underestimated. It seems that the single source products (either radar or gauge) are not enough to provide a reliable rainfall field. Thus, a merging approach, integrating both data sources, is required in order to have an accurate and highly resolved rainfall field. However, since the true rainfall field is unknown, the only way to validate the methods is through either split sampling or cross validation on gauge data, and this may favour the direct interpolation of gauge data. For instance, Ordinary Kriging interpolation of station data has already been proven to yield good results based on cross validation when a dense station network is at hand, thus the question arises – can the radar data through a merging method provide an additional value to the interpolation of the gauges even on a dense network? And if yes, what would be the best merging method and setting that would bring the most value? Would the performance depend on the event type? Would a better evaluation of rainfall field improve the predictability of the storms with the current existing methods?

These are the main questions treated in this chapter. Section 5.1 gives an overview about the single products: on one hand radar data and on the other hand the interpolation of gauge information through Ordinary Kriging. Next as discussed in the literature review, the merging methods; Mean Field Bias correction, Quantile Mapping based bias correction, Conditional Merging and Kriging with External Drift are explained in Section 5.2. The validation method, performance criteria and the operational application of these methods are described in Section 5.3. The results and the answers of the above mentioned questions are discussed in Section 5.4. As the resulting best merging method will be used as an input for the nowcasting of the rainfall, it is important to prove that this method can bring additional value to the nowcast – meaning that when validating the nowcast with the station data, the nowcast fed with the merged product should yield significantly better results than the one fed with the raw radar data. This is treated in Section 5.5, where it is also discussed up to which lead time the radar input quality influences the nowcast.

5.1 Single Products

5.1.1 Radar Data

As introduced in Section 2.2.2 and later illustrated in Section 4.3, there is a consistent spatial and temporal bias in the radar data when compared to the gauge data, mainly because of the scale mismatch between the two types of measurements. Due to different advection or wind speed and direction, the rainfall volume measured by the radar data at a specific time, may be captured with a time delay or even a spatial shift by the gauge instrument. Such a mismatch in time and space can be recovered partly if the radar data are smoothed in space and time. This was first suggested by Bárdossy and Pegram (2013), as they indicated that the prevailing wind direction is affecting the rainfall mismatch between the two data sources. Later, Berndt et al. (2013) used spatial and temporal smoothing of the radar data, and found out that not only does it improve the agreement between radar and gauge, but when used as a basis for the combination of the two data sources, it improves the performance considerably. In their results, they stated clearly that a strong spatial smoothing (with 25 adjacent grid cells- Figure 5.1 right), yields better results that a light spatial smoothing (with 9 adjacent grid cells - Figure 5.1 left). Following their results, here a strong spatial smoothing with 25 adjacent grid cells (Figure 5.1 - right) is performed before the temporal smoothing.
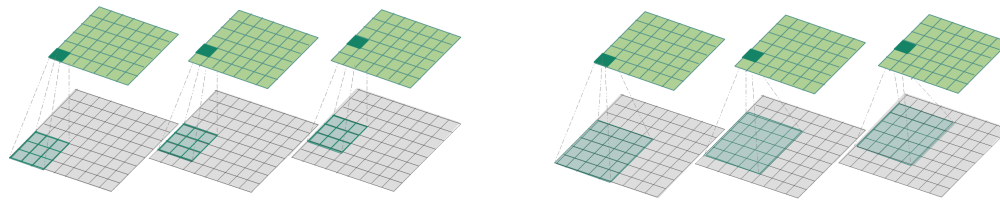
Figure 5.1: *Spatial smoothing applied to the radar data with two kernels; left – light smoothing with 3x3 kernel, right – strong smoothing with 5x5 kernel.*

Regarding the temporal smoothing, Berndt et al. (2013) suggest a temporal window of 15 minutes. However, as seen in Figure 4.5, the rainfall peaks at most events are exhibiting a temporal shift of 5 to 10 minutes. Based on the event results, and the fact that the investigation performed by Berndt et al. (2013) was done on a continuous time period, rather than specific events, it is important to choose with care the temporal window that would fit best for the nowcast application. For this purpose, three temporal windows are considered (10, 15 and 20 min) and applied as indicated (for the case of 15min) in the Equation 5.1. As the application is for nowcast, the temporal smoothing of a time step can be done by using only past information (see Figure 5.6 in Section 5.3). Spatial smoothing was done before the temporal smoothing for each case. The results of the best smoothing technique are indicated in Section 5.4.1. The smoothing window with the best correlation and lowest peak bias will be used as an input for the merging methods.

$$PI_{(t)} = \omega_2 \, PI_{(t-2)} + \omega_1 \, PI_{(t-1)} + \omega_0 \, PI(t)$$

where:
t   –   time step   [5min]
PI -   intensity at time t [mm/5min]
ω   -   weight for each time step [-]

(5.1)

5.1.2   Ordinary Kriging

Ordinary Kriging (OK) is one of the most used and simple methods among geostatistics interpolation techniques. It considers only one data source (uni-variate method) when estimating unobserved locations from adjacent known locations based on a Gaussian process governed by prior covariances. Geostatistics (also Ordinary kriging) regard spatial data as a result of a random process, thus each measurement at each location is a realization of such a random process. Ideally each location (*u*) within the study area could be described by a Gaussian probability distribution, and the set of all these probability distributions gives the random function that can describe and predict the random process. However, in practice such information is lacking and only a few realizations of the random process are available at some specific locations. In order to overcome this lack in information, all the known realizations of the random process at specific location are pooled together to build up the random function. Nevertheless, such implementation is possible if the second order hypothesis (as described below) is true:

1. The second order hypothesis states that the expected value of a random process in a specific study area is constant (Equation 5.2.) and that the covariance of two random variables (*C*) in the study area is dependent only on the distance between the two (*h*) and not on the exact location (neither *u* or *u+h*) as illustrated as well by Equation 5.3. The second condition is not always met, that is why the second order hypothesis is substituted by the intrinsic hypothesis which has slightly weaker constrains than the second order one.

$$E[Z(u+h)] = E[Z(u)] = m \qquad (5.2)$$

$$E[(Z(u+h)-m)(Z(u)-m)] = C(h) \qquad (5.3)$$

where:
h   – distance between two locations
Z(u)   – value at location u
Z(u+h) – value at location u+h
C(h)   – the covariance function
E[]   – the expected value

2. The intrinsic hypothesis states that the increment on the variance of two random variables at location *u* and *u+h*, depends only on the distance between the two locations *h* as illustrated in Equation 5.4. As for the second order hypothesis the equation 5.2 is still valid.

$$Var[Z(u+h)-Z(u)] = E[(Z(u+h)-Z(u))^2] = 2\gamma(h) \qquad (5.4)$$

where:
γ   – the semi-variogram function

The function that can describe the change of variance at a specific time step with the distance between any two pair locations, it's called a semi-variogram (here referred for simplicity as variogram). The empirical variogram is estimated as shown in Equation 5.5.

$$\gamma(h) = \frac{1}{2N(h)} \sum_{u_i - uj = h} (Z(u_i) - Z(u_j))^2 \qquad (5.5)$$

where:
u_i and u_j – locations at distance h
N   – the available number of pairs

As the empirical variograms are not continuous functions, theoretical functions need to be fitted to the observed values. Several theoretical variogram models may be used to describe the spatial variance of the data. Typically for the rainfall interpolation, the exponential variogram is used as indicated by Equation 5.6 [Berndt et al., 2013][Rabiei and Haberlandt, 2015].

$$\gamma_h = c_o + c\left[1 - \exp\left(\frac{-h}{a}\right)\right]$$

where:
c_o – the nugget
c – the sill
a – the range

(5.6)

Ideally the variograms are fitted for each time step, but in practice the variograms are commonly normalized and averaged over a period of time according to Equation 5.7. Verworn and Haberlandt (2011) state that the choice of variogram, either automatically fitted for each time step or averaged per each event, has little impact on the performance of the interpolation. Thus, the averaged variograms as per Equation 5.7 are used here. However, instead of variograms averaged per event, here they are averaged per season under the assumption that in summer (from 1st May to 31st of October) the dominant rainfall mechanism is of convective type and in winter (1st of November to 30th of April) of stratiform type. The reason behind this is that, in a nowcast application, the whole event information is not available, thus a seasonal variogram may be better suited.

$$\gamma_{avg}(h) = \frac{1}{n} \cdot \sum_{t=1}^{n} \frac{\gamma_t(h)}{Var(z_t)}$$

(5.7)

where:
γ_t – the variogram at time step t
$Var(z_t)$ – variance of z at time step t
n – number of time steps

For the calculation of the empirical variogram, radar data were used instead of the station data. As they provide more spatial information, they can produce better variograms than the station data [Germann and Jürg, 2001]. The semi-variance at the $1km^2$ data was averaged from two winter and two summers separately (years considered 2007-2008, 2009-2010 and 2011-2012). Only time steps with areal rainfall higher than 0.1 mm/5min were considered for the empirical variogram. The fitting of the theoretical variogram was done visually and based on the least square error optimization. The parameters for both winter and summer models are given in Table 5.1 and the variograms are illustrated in Figure 5.2

a)   b)
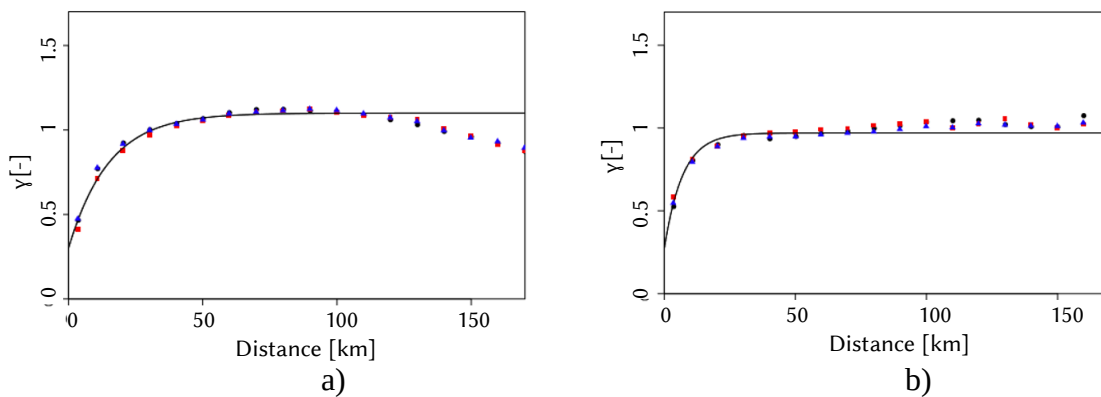
Figure 5.2: *The experimental variogram for a) winter and b) summer averaged over 2007-2008 (black), 2009-2010 (red) and 2011-2012 (blue). The fitted exponential variogram is shown as a black line.*

Table 5.1: *Parameters of the theoretical exponential variogram models used for the OK interpolation.*

	c_o	c	a_{eff}
Winter	0.3	0.8	45 km
Summer	0.27	0.7	21 km

Once a variogram has been determined, it can be used as a basis for the rainfall interpolation of 5min time steps on a 1km^2 grid (corresponding to the radar data). As any other interpolation technique, Ordinary Kriging estimates the rainfall at an-unknown location (Z') by averaging the rainfall from nearby known-location (Z) by certain weights as indicated in the Equation 5.8.

$$Z'(u_o) = \sum_{i=0}^{n} \lambda_i \cdot Z(u_i)$$

where:
λ_i – the weight from location u_i
$Z(u_i)$ – known variable at location u_i
$Z'(u_0)$ – unknown variable at location u_o
n – total neighbouring measurements

(5.8)

The weights of each nearby known location (λ_i) are determined based on the variogram and on the distance between the pair known-unknown location. The two equations (the kriging system) given in Equation 5.9 and 5.10 are used for finding the weights (λi) of Equation 5.8.

$$\sum_{j=1}^{n} \lambda_j \gamma_{(u_i - u_j)} + \mu = \gamma_{(u_i - u_j)} \, for \, i = 1, \dots, n$$

where:
$\gamma_{(ui-uj)}$ – the variogram for the known pair u_i, u_j
λ_j – the weight from location u_j
μ – the Lagrange multiplier
n – total neighbouring measurements

(5.9)

$$\sum_{j=1}^{n} \lambda_j = 1$$

(5.10)

This kriging system is solved for each point of the grid, in a way that for each time step the weights are estimated and later on the rainfall is calculated according to the Equation 5.8. The geostatistics software library (GSLIB) and the R library (GSTAT) were used to solve the kriging systems [Deutsch and Journel, 1998][Pebesma, 2004]. Following the hypothesis of stationarity, the OK is one of the best methods to achieve a spatial unbiased interpolation. However, the rainfall will be smoothed in space and many of the peaks will be underestimated and no particular structure (no storm recognition like in radar) is displayed. Thus, even though the spatial bias is expected to be zero, most probably the temporal bias is higher/lower than 0, since the peaks will be underestimated and the dry periods will be overestimated.

5.2 Merging Methods

5.2.1   Kriging with External Drift

In the Kriging with External Drift (KED) a secondary variable can be introduced that influences the interpolation. Contrary to OK, the intrinsic hypothesis is left out. Instead, the KED assumes that the expected value at a location *u* depends linearly on m additional variables as shown by the Equation 5.11.

$$E[Z(u)|Y_1(u),Y_2(u),....,Y_m(u)]=b_0+\sum_{k=1}^{m} b_k Y_k(u) \tag{5.11}$$

where:
$Z(u)$ – target variable at location u
m   – number of additional variables used
$Y_{1...m}(u)$ – *m* additional variables at location u
$b_o,b_1.,b_k$ – unknown constants of the linear relationship

Thus, theoretically speaking, the variogram for KED should not be computed from the absolute values of the target variable (as in OK), but from the residuals between the target variable and the external drift one. However, the results from Delrieu (2014) indicate that the OK variograms can be used instead of the residual variograms and produce similar results. Thus in this study, as in Berndt et al (2013), the OK variograms were used for the KED interpolation. Nevertheless, the assumption that the expected values of two locations are only dependent on the distance between the two (*h*) and not to their exact location still holds. The KED system is then described by the following equations

$$\sum_{j=1}^{n} \lambda_j \gamma_{(u_i-u_j)}+\sum_{k=1}^{m} \mu_k Y_k(u_i)+\mu_0=\gamma_{(u_i-u_0)} for\, i=1,...,n \tag{5.12}$$

$$\sum_{j=1}^{n} \lambda_j=1 \tag{5.13}$$

$$\sum_{j=1}^{n} \lambda_j Y(u_j)=Y(u) \tag{5.14}$$

In order to solve the system the same way as with OK, the new external variable (*Y*) has to be known for all of the locations of the grid. The GSTAT R-package is used for the application of the KED on each time step [Pebesma, 2004]. As suggested by Berndt et al. (2013) the KED implemented here only uses one external drift – the radar data. The KED does not depend on an assimilation time window to average the results, but only on the radar information at the given time step; either on the raw radar data or the temporally and spatially smoothed radar data as mention on Section 5.1.1. Thus, three radar products are investigated for the best application of the KED method as indicated in Table 5.2; the raw radar data, the smoothing with 10 minutes (as the best smoothing method from Section 5.4.1) and the smoothing with 15 minutes as suggested by Berndt et al. (2013). Depending on

which application yields the best results (given in Section 5.4.3), the final application is chosen and compared with the single data sources performances.

5.2.2 Mean Field Bias Correction

As seen in Figure 4.5 there is a consistent bias for all event types in radar data when compared to gauge measurements. A scaling with a factor can be useful to reduce the bias between the two data sources. These factors can either be singular values for each station or areal values for the whole radar field for a single time step. The single values, since they are point values, need to be interpolated in space in order to have factors for the whole radar region. On the other side, areal values can correct the whole radar image. For simplicity the areal values (or factors) for the whole radar region are implemented here. This is referred as Mean Field Bias correction (MFB) and is done according to Equation 5.15.

$$B_t = \frac{\sum\limits_{i=1}^{n} G_{i,t}}{\sum\limits_{i=1}^{n} R_{i,t}}$$

where:
t – the current time step to be corrected
n – number of available gauges
t – the current time step
$G_{i,t}$ – gauge observation at the i station and t time
$R_{i,t}$ – radar observation at the i station location and t time
B_t – the bias correction factor for the time t

(5.15)

The factor calculated only from one time step, due to the wind drift and scale mismatch between radar and gauge, may not be very useful for the correction. Instead another approach is used; the rainfall is accumulated at specific time periods (assimilation periods) and the correction factor is calculated based on this accumulation volume, as shown in Equation 5.16.

$$B_t = \frac{\sum\limits_{j=0}^{w} \sum\limits_{i=1}^{n} G_{i,t-j}}{\sum\limits_{j=0}^{w} \sum\limits_{i=1}^{n} R_{i,t-j}}$$

where:
t - the current time step to be corrected
n – number of available gauges
w – number of time steps as a basis for the accumulation
$G_{i,j,t}$ – gauge observation at the i station and time step j
$R_{i,j,t}$ – radar observation at the i station location and time step j
B_t – the bias correction factor for the time t

(5.16)

For MFB hourly accumulations are typically used. Thus, for a current time step, for instance t, the accumulations from $t_0 - 1h$ to t_0 can be used to correct the current radar image fed to the nowcast algorithm. Nevertheless, the hourly accumulation can work for some type of events and not work for others. For this reason, the accumulations over periods of 0.5, 1, and 2 hours are investigated to find which accumulation is suited better for the study area and nowcast purpose. As radar input for all three methods will be the 10 min smoothed radar data, as the best radar product (see Section 5.4.1).

5.2.3 Quantile Mapping based Bias Correction

The main idea of the Quantile Mapping Bias correction is the correction of a dataset by manipulating its probability distribution based on one from another dataset that is considered the true observation. Here the dataset to be corrected is the radar data and the dataset

considered as the true rainfall is the gauge information. The procedure of the quantile based bias correction is shown in Figure 5.3 and is mainly done in three steps.
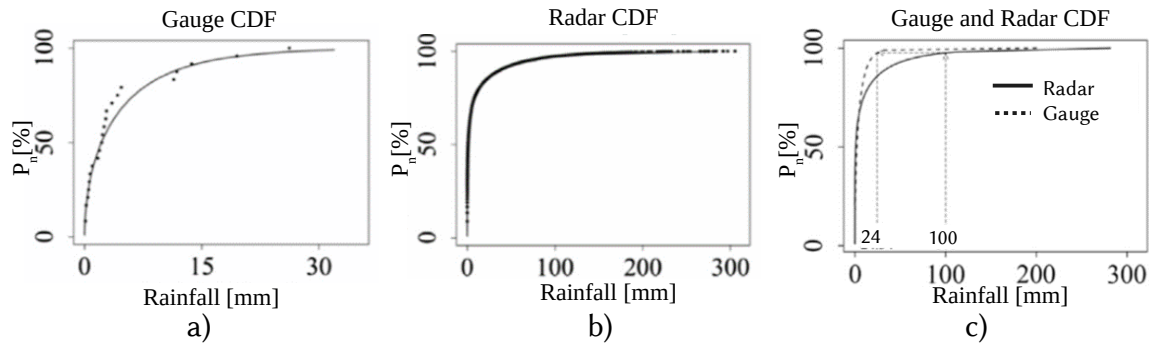
Figure 5.3: *Different steps in the bias correction of the radar data based on quantile mapping: a) fitting a Gamma distribution to gauge recordings, and b) to radar intensities. Based on the two distribution, for same quantiles, the radar intensities are substituted by the gauge ones (c) [Rabiei and Haberlandt, 2015].*

First the empirical cumulative probability distributions (CDF) for both radar and gauge data are built up as in (a) and (b). A theoretical Gamma distribution is fitted to the each of the empirical distributions as indicated by the Equation 5.17. Here the R package lmomco [Asquith, 2011] is used to estimate the parameters of the Gamma distribution based on the method of L-moments.

$$(5.17)$$

$$F(Z(x,t);r,\lambda)=\frac{\lambda^r Z(x,t)^{r-1}\exp(-\lambda Z(x,t))}{\Gamma(\alpha)};Z(x,t)>0$$

where:
F — the non-exceeding probability for a data type
$Z(x,t)$ — the rainfall value at location x and time t
r,λ — parameter of the Gamma distribution
$\Gamma(\alpha)$ — the Gamma function

Then the correction is applied; for each grid point of the radar data which has to be corrected, the quantile based on the respective CDF is computed, and the radar value is substituted with the value of the gauge data corresponding to the same quantile as illustrated in Equation 5.18. An example of such application is given by the Figure 5.3 (c); A value of 100 mm/h is recorded by the radar, which based on the radar CDF corresponds to the quantile 0.9. The gauge intensity corresponding to the same quantile is 24 mm/h, hence the value of 100 mm/h in the radar is substituted with the one of 24 mm/h.

$$(5.18)$$

$$Z'_R(x,t)=F^{-1}_{obs,t}(F_{rad,t}(Z_R(x,t)))$$

where:
$Z'_R(x,t)$ — the new value of radar
$F_{rad,t}$ — the quantile of radar observed value at x and t ($Z_R(x,t)$)
$F^{-1}_{rad,t}$ — the gauge value corresponding to the quantile $F_{rad,t}$

As indicated in the Equation 5.17, only intensities higher than a specific threshold can be used for the empirical distribution intensities. As the measurement accuracy of radar and

gauge is 0.01 mm/5min, values that are equal or higher than this threshold are used to built the CDF of both datasets. Actually, Rabiei and Haberlandt (2015) investigated various threshold (0, 0.02, 0.05, 0.07 and 0.1 mm/h) for the CDF calculation of radar data, and concluded that their empirical CDF should be build by intensities higher than 0.5mm/h. Nevertheless, in this application, the 0.01 mm/5min yielded stable and satisfactory results.

Theoretically, the quantile bias correction has to be done on each time step. However, care has to be taken that enough data are available to fit the Gamma CDFs, otherwise the parameter estimation is not possible. In the best case scenario, 100 stations will be available for the CDF construction. A hundred stations are not enough to provide a good correction of the radar CDF where around 40 000 points are used. Two solutions can be used to deal with this lack of data.

i) The first one is to interpolate the station data to the same grid like the radar and use the interpolated field to build up the CDF. Nevertheless, such an application will introduce additional uncertainty to the correction as the interpolated field comes with errors as well. Moreover, as OK is smoothing in space, it will cause an underestimation of the peak intensities at the radar.

ii) The second one is to pool together radar and gauge observations from a fixed duration (also referred here as temporal window). However, one has to be careful with determining the temporal window from which to take the information, as a too long window may mix information from different rainfall mechanisms (hence affect the radar correction), and a too short window can cause instabilities when estimating the parameters of the Gamma distribution.

To investigate what duration yields best results for all the types of events, three durations are considered and investigated: 1, 3, and 6 hours. As radar input for all three methods will be the 10 min smoothed radar data, as the best radar product resulting from Section 5.4.1.

5.2.4   Conditional Merging

The Conditional Merging (CM) is the combination of two data source: the radar data and the interpolated gauge data with Ordinary Kriging. The combination is done on 5 steps as illustrated in Figure 5.4. For a specific time step, the gauge information is available on specific points (a) and radar data (b) is available continuously in space (displaying a high variance). The first step is the OK interpolation of the gauge data as shown in (c). Second, the information from the radar grid points matching with the gauge location, is interpolated via OK with the same variogram as in the gauge interpolation (d). Then a deviation grid between the radar data (b) and the interpolated radar field (d) is derived (e). This deviation grid is added to the interpolation of gauge information (f). The final product of conditional merging is the one indicated in part (g) and it includes an unbiased estimator in space (due to OK) and a

high spatial variability (due to the radar data). Moreover, two conditions are set to the final product in order to avoid negative rainfall and to keep the storm structures as dictated by the radar data, so that they can be easier tracked by nowcast algorithms.

1. if the OK field is smaller than the deviation field then the resulting grid is set to zero.
2. if there is no rainfall recognized by the radar, then the resulting grid is set to zero.
3. if the radar data are missing, then the OK of gauge is used.

Unlike the MFB or the QQ correction methods, the CM doesn't depend on a time window to average the results, but depends only on the radar information at the given time step. Thus, the CM can be run either on the raw radar data or the temporally and spatially smoothed radar data as mention on Section 5.1.1. As in the KED application, three radar products are investigated for the best application of the CM method as indicated in Table 5.2; the raw radar data, the smoothing with 10 min (as the best smoothing method from Section 5.4.1) and the smoothing with 15 min as suggested by Berndt et al. (2013). Depending on which yields the best results (given in Section 5.4.3), the final application will be chosen and compared with the single data sources performance.
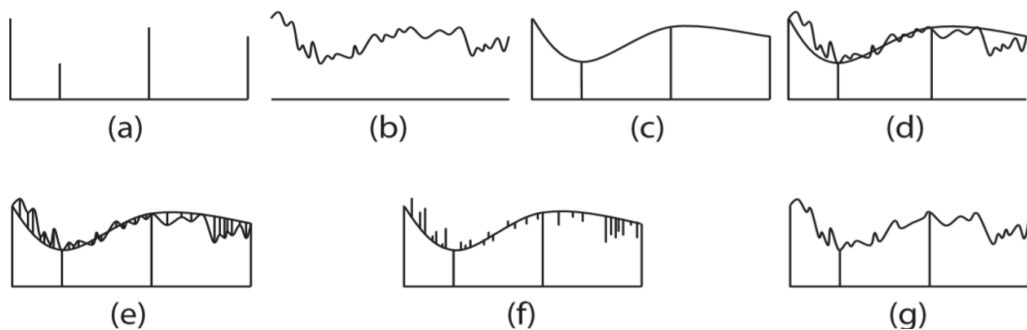
Figure 5.4: *The procedure for the conditional merging: a) represents the observations at the gauges which are discrete, b) represents the observations in radar which are continuous in space, c) the observation by the gauge are interpolated with OK, d) the radar observations corresponding to the gauge location are interpolated with OK, e) the derivation grid is computed between the original radar and radar interpolation, f) the deviation grid is added to the gauge interpolation data, and g) the final data as the product from gauge interpolation and radar variation grid [Berndt et al., 2013].*

5.3 Method Evaluation and Application

5.3.1   Merging Validation and Application

For estimating the best merging method, the performance criteria as explained in Section 4.2 are used. The true information for the rainfall considered here, is the gauge information. Thus, the performance assessment is done at the point scale, where the gauge measurements (true rainfall) are available. For the validation of the methods a split-sampling validation is used; the 100 stations are separated into two groups a) 68 calibration stations – the stations that are used in the merging application, and b) 32 validation stations – the stations where the performance criteria are calculated. The stations in each group are selected randomly and are illustrated in the Figure 5.5.
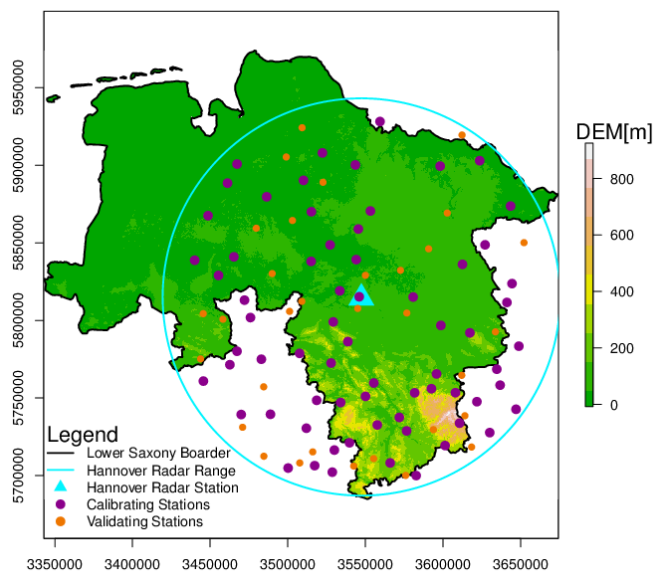
Figure 5.5: *The stations used for the calibration (in magenta) and for the validation (in yellow) of the merging methods.*

Since the methods are used to merge radar and gauge data for the nowcast application, only rainfall information from the past of the current time step can be used. That means that the methods are applied online. Offline application, in contrary, can use the information from the past and the future of the current time step. The illustration of the two modes online and offline application is shown in Figure 5.6, where the current time step is t=0 and the assimilation window is receptively the online and offline filter.
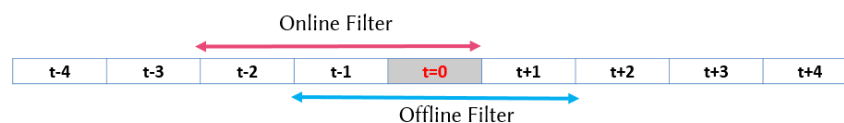
Figure 5.6: *The application of the merging methods online (pink) and offline (blue).*

The summary of the merging methods investigated, for different temporal assimilation windows and for different radar inputs, is indicated in the following Table 5.2:

Table 5.2: *Summary of the methods investigated, the temporal window applied, the input product and the reference name for each of them.*

	Window [min]			Input			Product Name		
Smoothing	10	15	20	RR_{raw}			RR10	RR15	RR20
OK				G				OK	
MFB	30	60	120	RR10			MFB30	MFB1	MFB2
QQ	60	180	360	RR10			QQ1	QQ3	QQ6
KED				RR_{raw}	RR10	RR15	KED_{raw}	KED10	KED15
CM				RR_{raw}	RR10	RR15	CM_{raw}	CM10	CM15

5.3.2   Application of Best Merging Product at the Nowcast

The best merging method achieved (from Table 5.2) will be used as an input for the nowcast algorithms and all the 100 stations (as illustrated in Figure 4.3) are used as input for the best merging method. Based on these 100 stations the performance criteria listed in Section 4.2 are used to determine the impact of the best merging method to the nowcast performance. To have a direct comparison between the nowcast fed with merged data and the nowcast fed with raw radar data, a percent of improvement is calculated for each performance criteria as indicated by the following Equations 5.19 - 5.22. A positive improvement indicates that the nowcast fed with merged data had better performance than the one fed with raw radar data, and a negative improvement indicates the opposite.

$$\rho_{impr}[\%]=100\frac{(\rho_{merge}-\rho_{raw})}{\rho_{raw}} \tag{5.19}$$

$$RMSE_{impr}[\%]=100\frac{(RMSE_{raw}-RMSE_{merge})}{RMSE_{raw}} \tag{5.20}$$

$$BIAS_{impr}[\%]=100\frac{(|BIAS_{raw}|-|BIAS_{merge}|)}{|BIAS_{raw}|} \tag{5.21}$$

$$TP_{shift}[\%]=100\frac{(|TP_{raw}|-|TP_{merge}|)}{|TP_{raw}|} \tag{5.22}$$

where:
ρ	– the correlation displayed by the radar product	merge	– criteria displayed by the merged radar
RMSE	– the root mean square error for each time step	raw	– criteria displayed by the raw radar
BIAS	– the absolute error in volume estimation	impr	– the improvement in criteria due to the
TP	– the shift in time of rainfall peak		merged radar

5.4 Merging Results

5.4.1   Radar Smoothing

Figure 5.7 shows the temporal performance of the three temporal smoothing windows on radar data compared to the use of the raw data. For all smoothing windows, the temporal correlation is improved considerably, for all the thresholds, as well the temporal RMSE and the Peak Bias. Among the different durations, the 10min smoothing window exhibits the best results. Of course the longer the window, the bigger is the smoothing (the lower the variance) causing lower correlations. However, for the 5 and 12 mm/h intensities it seems than the 15 min window performs a bit better in terms of temporal correlation. In terms of RMSE there is no big difference between the smoothing windows. The 20 min window seems to produce slightly lower RMSE, but that is to be expected as the smoothing is stronger. Regarding the peak bias, the 10 and 15min windows yield the best results, with the 10 min being slightly better. It is clear that the temporal smoothing improves the match of radar and gauge observation, nevertheless one has to be careful not to use a strong smoothing as the variation and the peak volume will decrease considerably. Based on the results, the 10 min window is chosen as the best smoothing window for the event based application.
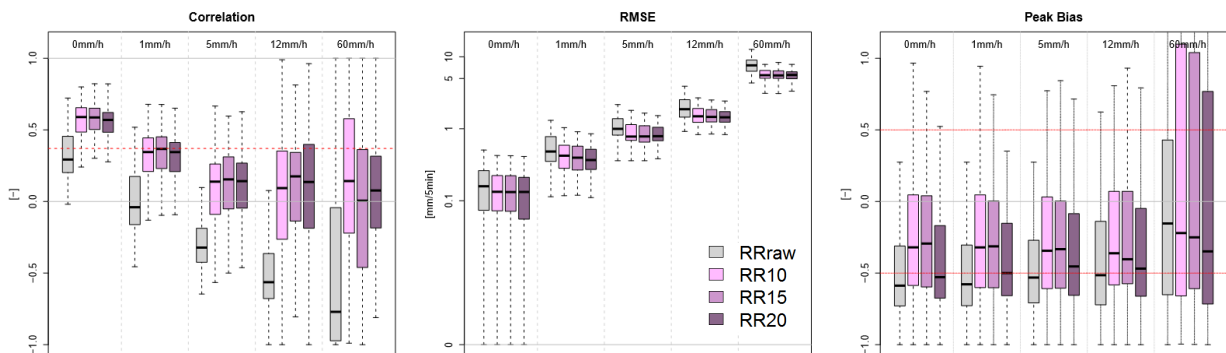
Figure 5.7: *The temporal performance (correlation -left, RMSE -middle, Peak Bias - right) for the application of each temporal smoothing window of the radar data. The grey dashed line are guidelines while the red dashed one indicate critical values as dictated by the literature.*

Figure 5.8 shows the temporal and spatial cross-correlation between the 10 min smoothed radar data for each event type. Overall there is an improvement of the correlation for all the storms. As in the raw radar, the spatial cross-correlation is higher than the temporal one for the low thresholds, however for the high thresholds the temporal one is higher. Especially, for the 60 mm/h intensity threshold, the stratiform spatial structure is very well captured. Nevertheless, the cross-correlation, both spatial and temporal, is higher than 0.37 (predictability limit) only for the lowest threshold. Thus, it is still hard to expect a high predictability skill from the nowcast model at the high intensities.

For the temporal cross-correlation, the smoothing seems to be better for the stratiform and the meso-β events. This is also the case for the spatial cross-correlation. These two types of

events are widespread and not with very high intensities, therefore smoothing in time and space yields better results. Nevertheless, it was expected that the smoothing in time and space could improve particularly the results for the convective events – where due to the wind drift, the gap between the radar and the gauge is bigger. If compared with the Figure 4.4 from the raw radar data, it is clear that all types of events (also convective ones) agree better with the gauge data when radar data are smoothed.
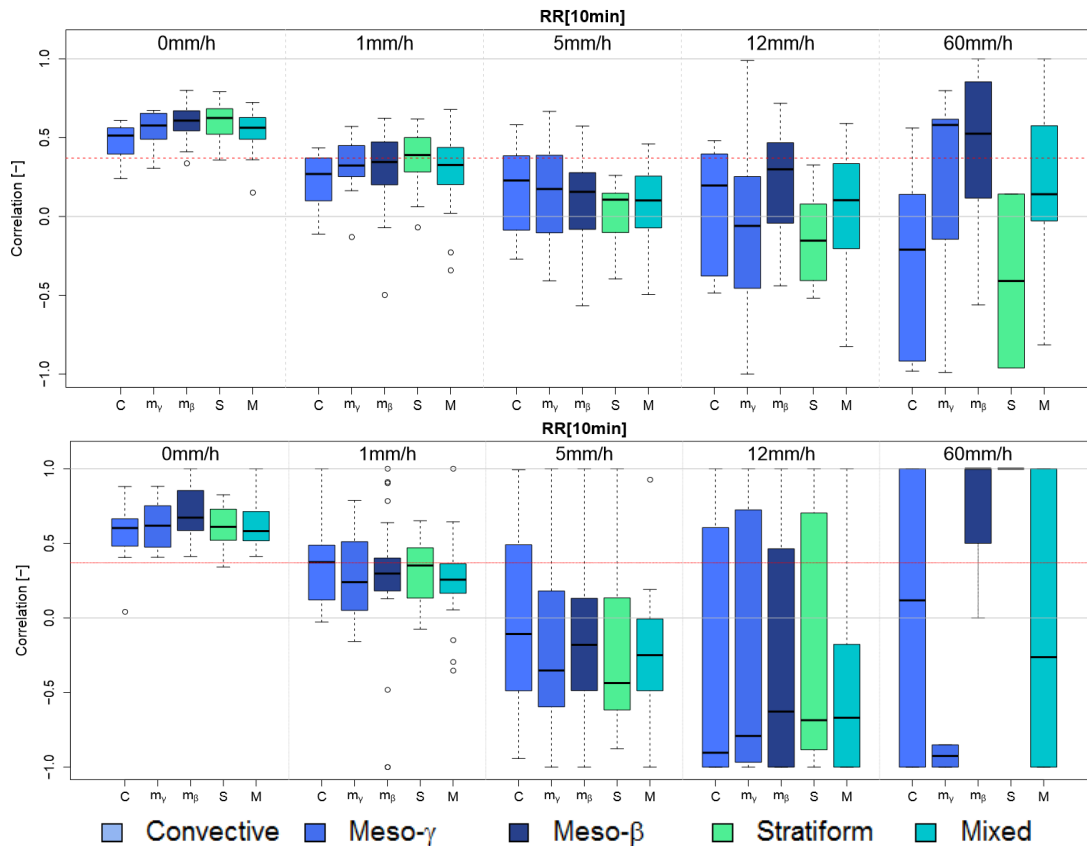
Figure 5.8: *The temporal (upper) and the spatial (lower) cross-correlation of the radar data smoothed with a 10 min window for the 5 types of the events. The grey dashed lines are guidelines while the red dashed lines illustrate the predictability limit value as shown by the literature.*

However, it still seems that the convective (simple and meso-γ) events fall behind the others. This may be caused by the so-called advection errors; as the radar measurement is only a time shot at periodic times (it captures an instant state of the storm at different times see Figure 5.9-a and -b), it may miss certain rainfall location due to the distance between the instant states of the storms as illustrated in Figure 5.9-c. For slowly moving storms, like stratiform events, this gap is recovered by the spatial and temporal smoothing, however for the fast moving storms, the smoothing is not that effective. This may be solved in future works, by an advection correction as suggested by Seo and Krajewski (2015).
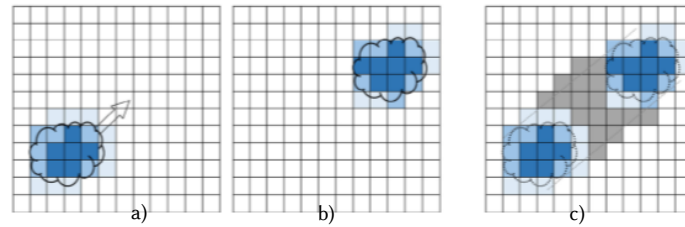
Figure 5.9: *Schematic representation of the instant radar rainfall estimation at t (a) and t$_{+5min}$ (b) on 1km² grid, and rainfall accumulation between the two time steps (c). Blue grids indicate the rainfall intensities captured by the radar, while the grey ones illustrate the locations where the radar did not capture any rainfall [taken from Seo and Krajewski, 2015].*

5.4.2   Gauge OK Interpolation

Figure 5.10 illustrates the temporal performance criteria for the interpolation of the gauge data with OK for different intensity thresholds. As expected the temporal correlation is quite low (lower than the predictability limit value) and it gets worse with the higher intensities. Even though OK is unbiased in space, this is not the case for the bias in time. The OK interpolation is clearly underestimating the peak volume especially for the high intensities. It is apparent that the OK cannot be used as a single input to the nowcast algorithm, because the temporal performance criteria are quite bad when compared to the smoothed radar data. Thus, merging methods are needed!
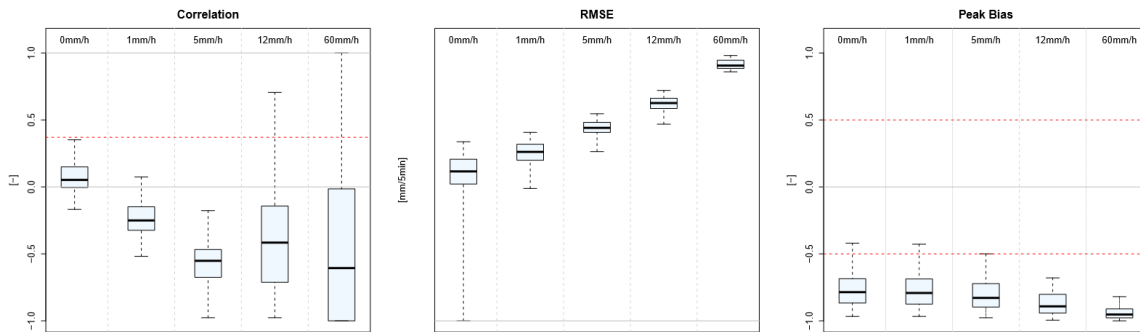
Figure 5.10: *The temporal performance (correlation -left, RMSE -middle, Peak Bias - right) for the OK interpolation of the gauge data.*

Figure 5.11 and Figure 5.12 illustrate the temporal and spatial cross-correlation for different thresholds and event types for the OK interpolation of gauges. Among all the event types, it seems to capture best the structure of the stratiform events. This is expected as those events are quite extended in space, and many radar grids have captured rainfall. However, since it is smoothing in space, the high intensities are not well captured (hence low correlation for intensities higher than 60 mm/h). Regarding the convective events, even though a summer variogram was implemented, with clearly shorter range, it is not enough to model such events correctly.
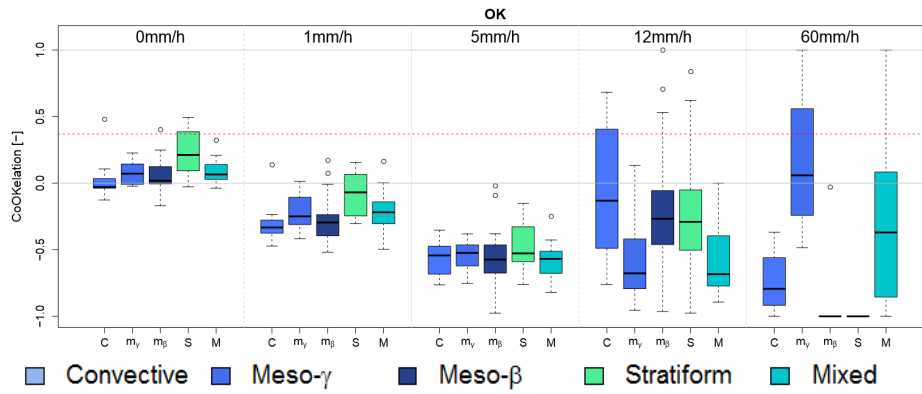
Figure 5.11: *The temporal correlation of the OK gauge interpolation for the 5 types of the events. The red dashed line illustrates the predictability limit value as shown by the literature, while the grey lines are for guidance.*
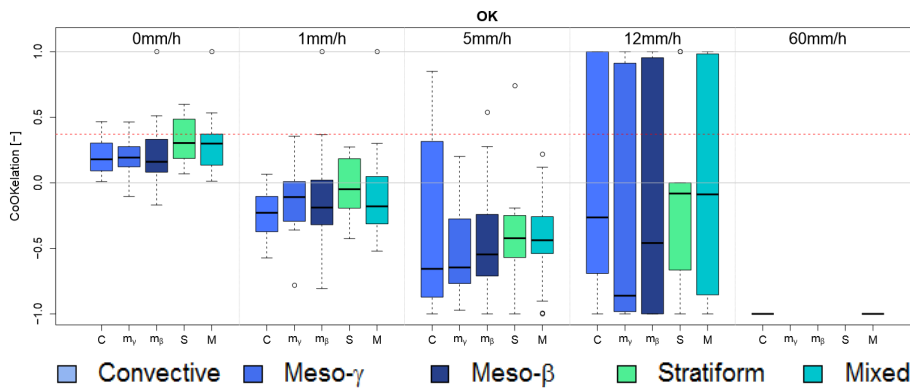
Figure 5.12: *The spatial cross-correlation of the OK gauge interpolation for the 5 types of the events. The red dashed line illustrates the predictability limit value as shown by the literature, while the grey lines are for guidance.*

5.4.3   Best Application for each Method

For each merging method, different applications (as shown in Table 5.2) were investigated in order to find the best application for each method. The evaluation of the best application is done based on the temporal performance criteria over the 32 validation stations. The results for all merging methods are shown in Figure 5.13.
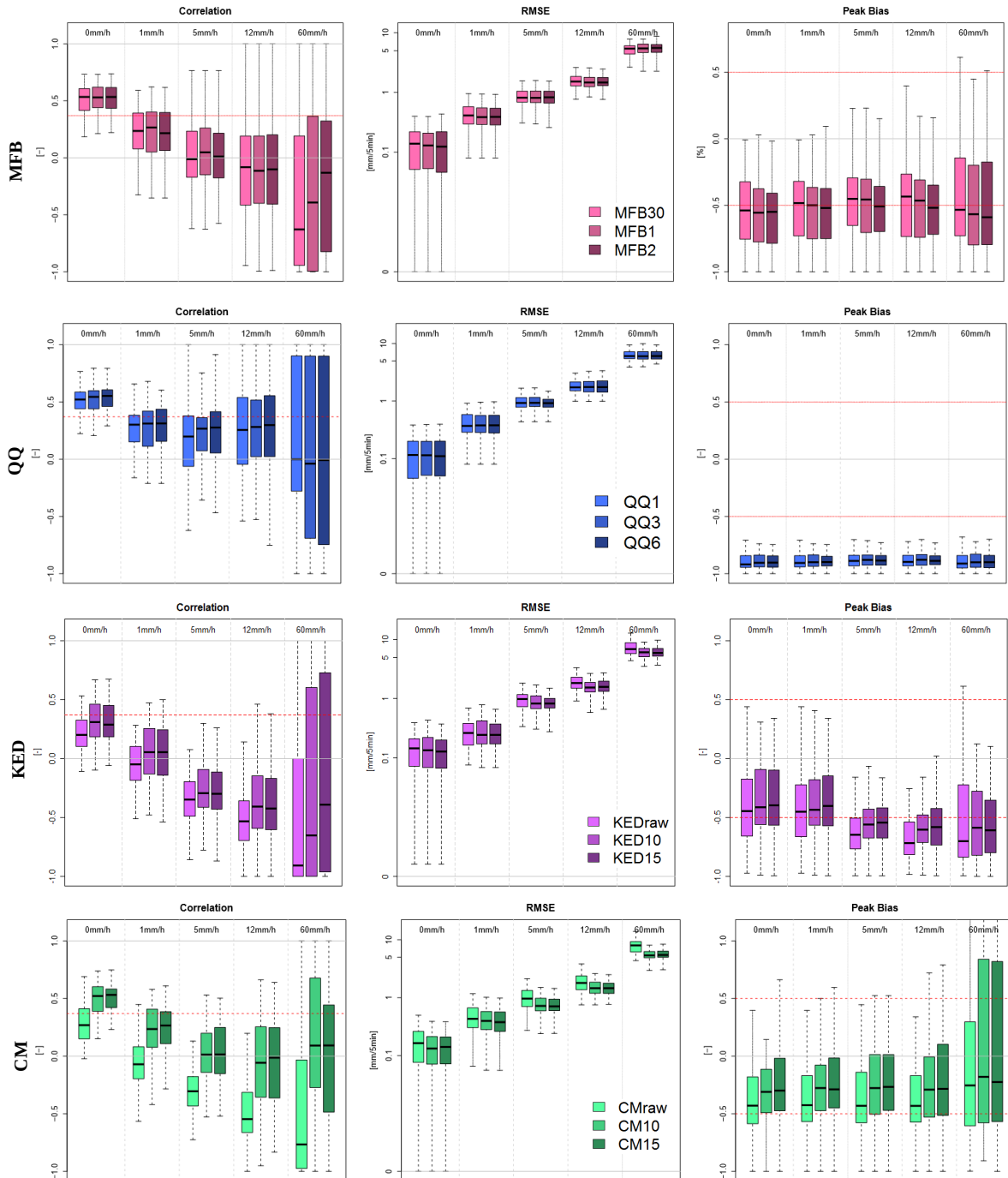
Figure 5.13: *The temporal performance (correlation -left, RMSE -middle, Peak Bias - right) for the application of each merging method (MFB, QQ, KED and CM) with different temporal windows used for the assimilation of gauge and radar information. The grey lines are for guidance while the red dashed lines show the critical values as indicated by the literature.*

The performance of three different implementation of MFB corrections on radar data are very similar to one another, with very few differences. The temporal correlation for the low intensity thresholds shows the same range for all three applications. A small difference is apparent at the high intensity threshold of 60 mm/h, where the MFB on two hours accumulation has slightly higher values. However, this application has the highest peak bias when compared to the other one. Contrary to this, the MFB on 30 minutes accumulations has the lowest bias when compared to the others, but also the lowest correlation for the high intensities. It appears that the hourly MFB has a more balanced trade between the correlation and the peak bias than the other applications, thus it is selected as the representative one. This is not only in agreement with the literature review, but also with the RMSE results, as the hourly MFB displays slightly lower RMSE than the others.

The performance of three different implementation of QQ corrections on radar data are also similar to one another, however show a higher variability than the MFB implementations. The temporal correlation looks better for the 3- and 6-hours implementations; only at the highest intensity threshold (60mm/h) the hourly implementation is slightly better. Nevertheless, when looking at the peak bias, QQ for each implementation displays a constant and high bias in time. Even though the QQ (together with MFB) is optimized by reducing the bias in space, apparently this is not enough to reduce the bias in time. When advised with the Figure 5.11 and Figure 5.12, it seems that this feature of QQ is inherited from the gauge information, as OK also underestimates considerably the temporal peaks. From the three implementations, the 3 hours one shows slightly better results for the peak bias, and it is selected as the representative one.

The performance of three different implementation of KED merging of radar and gauge data are quite different from one another. For starter, it is clear that the spatio-temporal smoothing of the radar data, improves clearly the temporal performance of the KED when compared to the one using raw radar data as an external drift. Both KED10 and KED15 show a higher correlation and smaller RMSE and peak bias when compared to the KED_{raw}. Among the two smoothed radar products fed to the KED, there is not a very big difference. From one side the KED15 shows better temporal correlation for the high intensities but on the other side it shows also higher peak bias. It is for this reason, the lower peak bias, that the KED10 is chosen as a representative of this method.

The performance of three different implementation of CM are, as in the KED case, quite different from one another. Also, here it is clear that the spatio-temporal smoothing of radar data improves clearly the temporal performance when compared to CM_{raw}. Both CM10 and CM15 show a higher correlation and smaller RMSE and peak bias when compared to the $CM_{raw.}$ Unlike the KED implementation, here it is clearer that the CM10 runs have slightly better results at the highest intensity threshold (60 mm/h), than the CM15; better correlation and peak bias median. For the other intensity thresholds the performance of both products seems very close to one another. Also in terms of RMSE, their performance is quite similar, with CM10 having slightly lower values for the 60mm/h intensity. Since the CM10 exhibits

better results for the highest intensity, it is selected as the representative from the CM implementations. Since both KED and CM are fed with 10 min smoothed radar data, a direct comparison between them is possible.

5.4.4   Best Merging Method

A direct comparison of the temporal and spatial performance of the merging methods (MFB1, QQ3, KED10 and CM10) and the reference single source products (RR_{raw} and OK) for the different intensity thresholds is illustrated by Figure 5.14 and Figure 5.15. Regarding the temporal correlation, it seems that all of the merging methods have a better performance than the OK of stations. Among all the merging methods, CM10, MFB1 and the QQ3 outperform the KED10, with CM10 and QQ3 being ranked among the best. The MFB1 has similar performance to the CM10, however for the highest intensity the superiority of CM10 is clear. QQ3 shows a bit higher correlation than CM10 at the mid-range intensities, but losses this advantage at the higher intensities.

It is surprising to see that KED10 is the worst merging method in terms of temporal correlation for all the intensities. When looking at the temporal correlation of KED10 for each event type (see Figure 5.16), it exhibits quite high correlation for the stratiform events, however very low values for the convective events. As a derivative of OK, KED10 seems to inherit the bad ability of the OK to capture the convective events. CM10 on the other side, inherits the high correlation of the radar data and seems not to be so dependent on the OK performance. Therefore as the RR10, the CM10 displays very good correlation for the wide range events (like stratiform, mixed and meso-β convective scale), but cannot capture well the temporal pattern of the simple and meso-γ convective events. On the opposite to CM10, QQ3 seems to improve the performance especially for these two types of events: simple and meso-γ convective events, while for the other events the correlation is lower than the CM10 method.

In terms of temporal peak shift and bias (Figure 5.14), the methods are again better than the OK of the station data. KED10 seems again to inherit the high peak shift from the OK, although the bias in the peak is considerably better than OK. In contrast to the temporal correlation, the supremacy of CM10 in the peak capturing is quite obvious; first the CM10 has a median peak shift of 0 minutes consistent throughout all the thresholds, which is not the case for the other methods; secondly it has a peak bias lower than 50%. Of course, the underestimation is still considerable, but given the merging methods, it is the best in reducing this bias. In terms of RMSE, all merging methods improve the performance by having lower RMSE than OK and RR data.

Regarding the spatial cross-correlation (Figure 5.15) QQ3 looks like the best performing method, followed up from the CM10 and MFB1. The KED10. even though it shows a higher spatial than temporal correlation, it still falls behind the other methods. Regarding the space variance in Figure 5.15, QQ3 seems to inherit the low variance of the gauge OK (see the difference between the OK and the QQ3), unlike the CM10 which exhibits a higher variation

from the radar data. KED10 has a slightly better variation coefficient than OK, but still very low – indicating that the spatial variance of the storms is underestimated. MFB1 at the other hand has a RVAR within the 1 range, but the median is worse than the CM. Regarding the spatial RMSE, all merging methods exhibit lower value than the RR and OK, especially when high intensity thresholds are considered.
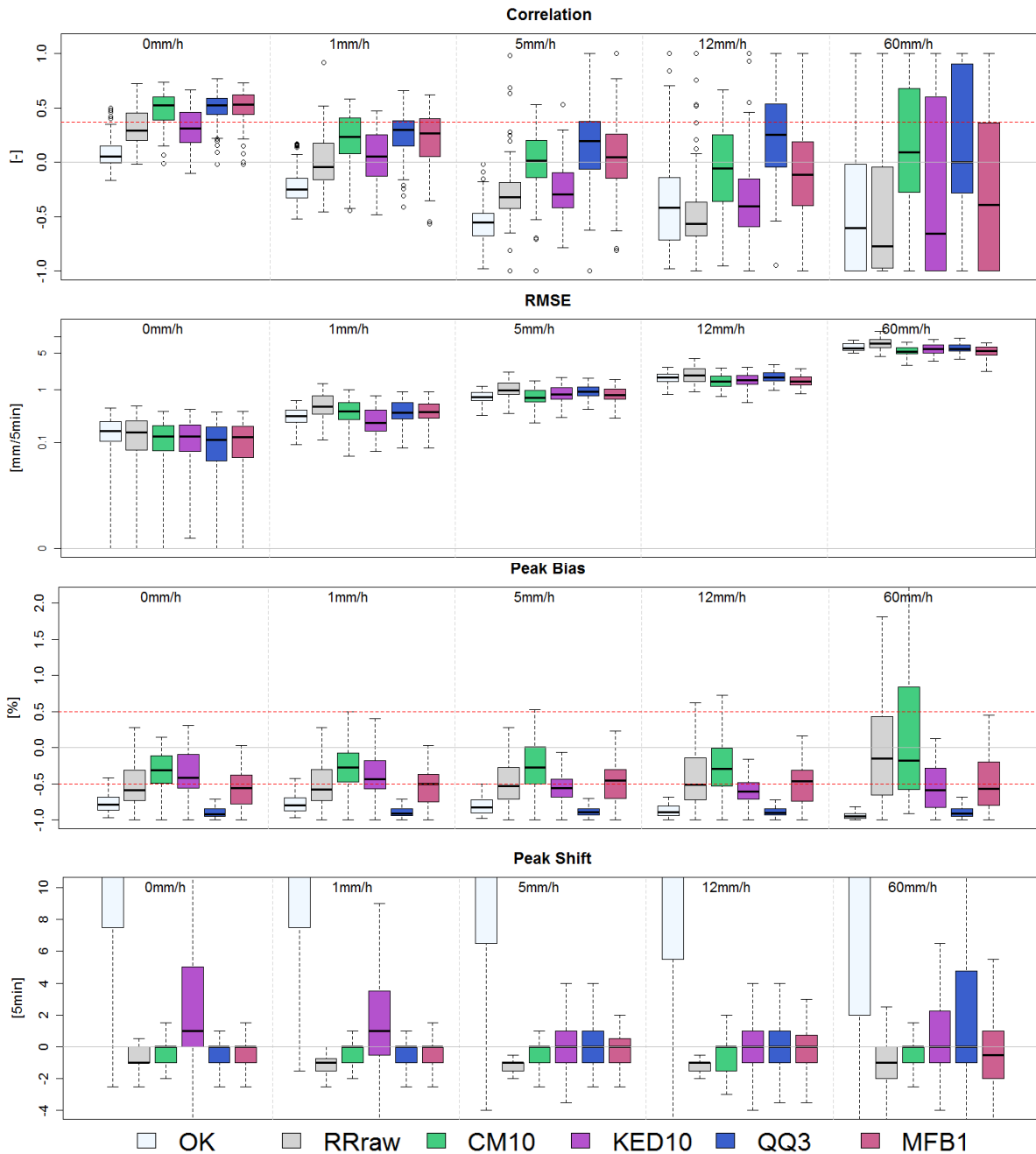
Figure 5.14: *The temporal performance criteria for each of the single (OK and RR) and merging products (MFB1, QQ3, KED10 and CM10) for different types of thresholds. The grey lines are for guidance, while the red dashed ones indicate the critical values.*

Unlike in the temporal correlation, the QQ3 can capture better the spatial variability of the convective events, rather than the stratiform events. The CM10 on the other hand, keeps still the same behaviour as in the temporal correlation; it can capture better the wide range events (stratiform, mixed and meso-β events) than the small scale ones (simple and meso-γ

convective events). KED10 displays higher cross-correlation than OK for all the event types, but as OK it captures better the structure of the wide range events, (still not as good as the CM10). Lastly MFB1 shows very good correlation for all the type of events, and surprisingly it improves the cross-correlation mainly for the small scale events rather than the large scales, displaying a slightly higher correlation than the CM10 for the low intensity threshold.

Overall, the merging methods have their own advantage and disadvantage. Some can capture very well the temporal and spatial structure of the rainfall but in the cost of the peak bias or the RMSE. From all the methods, CM10 seems to be more consistent with the results displaying good correlation, RMSE, peak bias, peak shift and RVAR. It is for this reason that CM10 is chosen as the merging method for correcting the input radar data of the nowcast. QQ3 seems a second promising method, however its application is yet to be improved. Lastly, the KED with only RR doesn't seem that promising for urban application, unless other additional variables are used as external drifts.
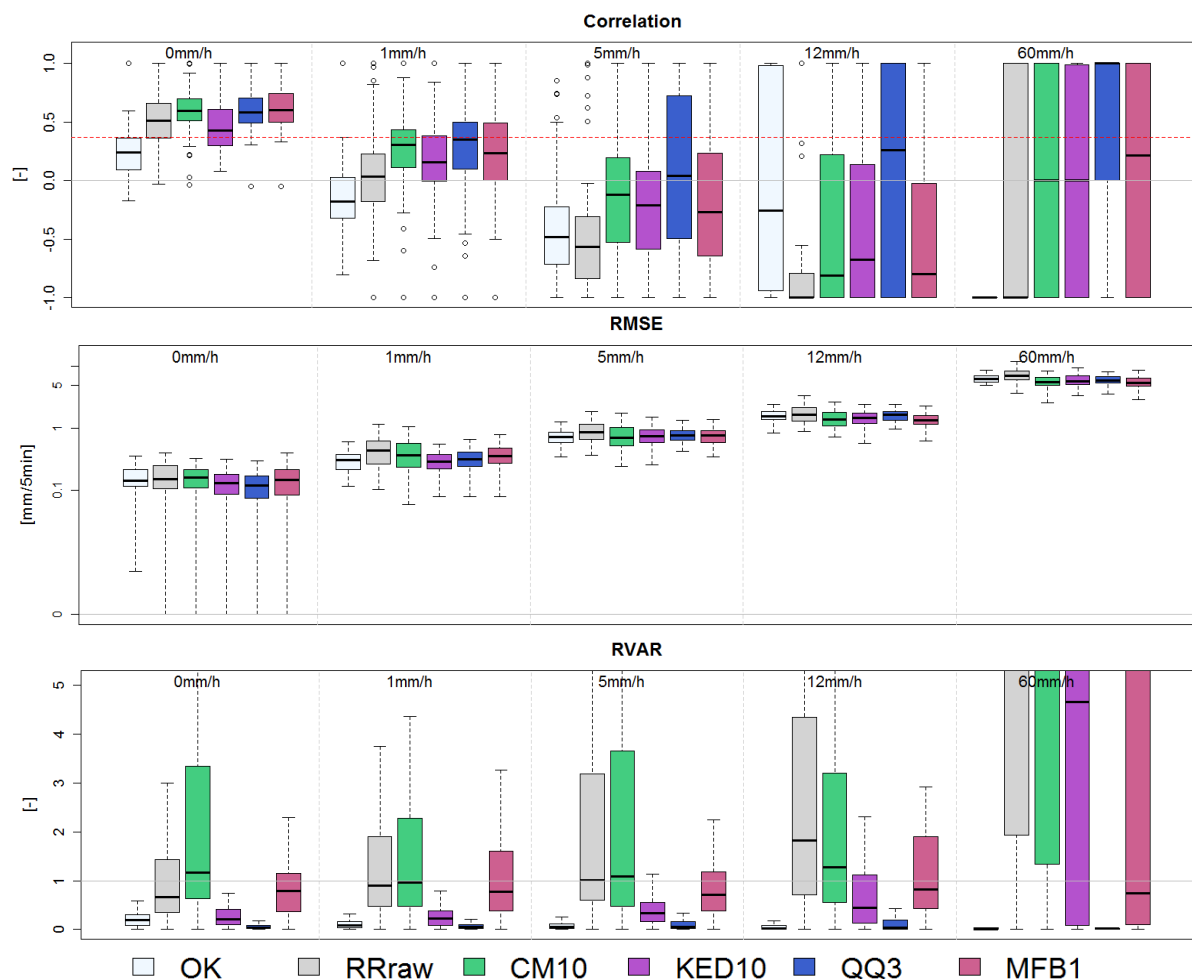
Figure 5.15: *The spatial performance criteria for each of the single (OK and RR) and merging products (MFB1, QQ3, KED10 and CM10) for different types of intensity thresholds. The grey lines are for guidance, while the red dashed one indicates the critical value as suggested by the literature.*
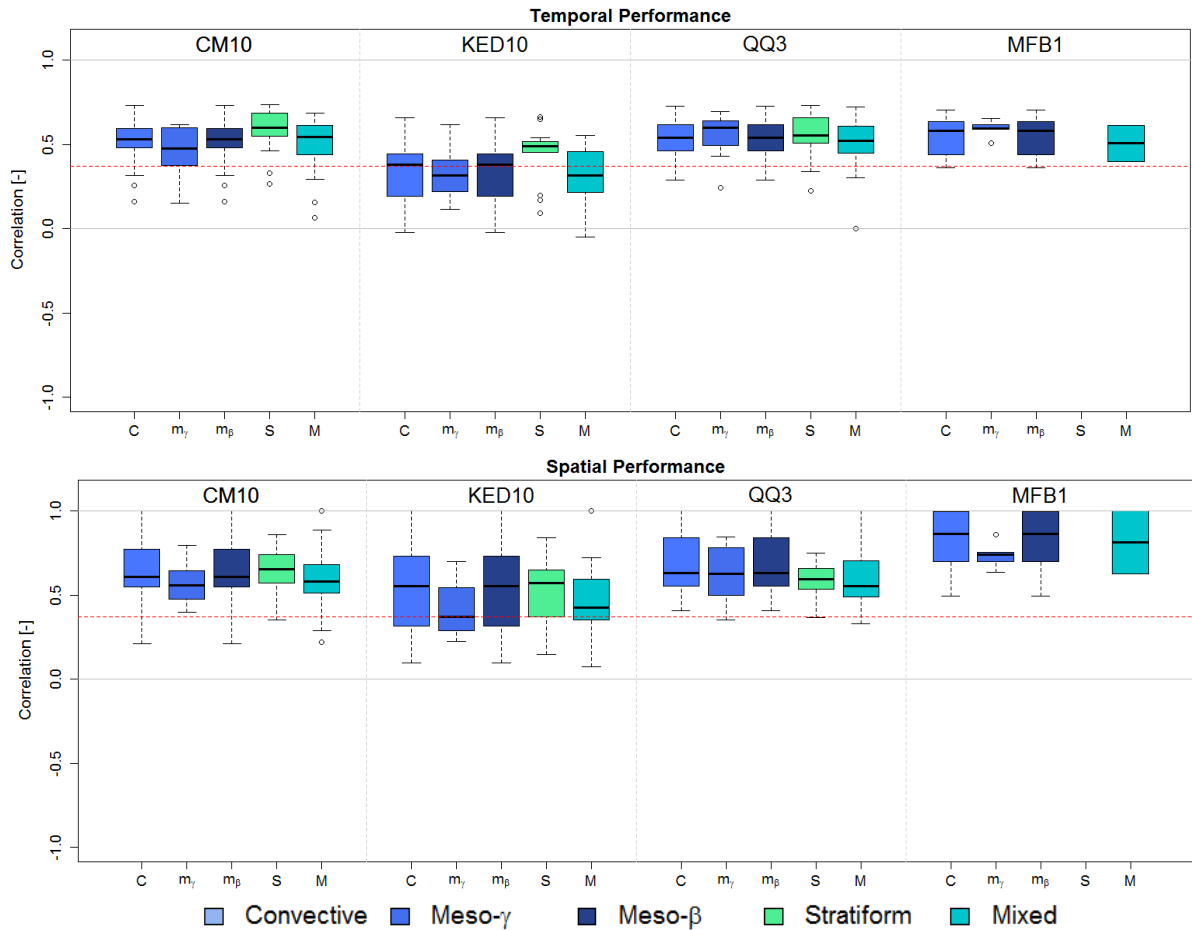
Figure 5.16: *The temporal (upper) and the spatial cross-correlation (lower) of each of the merging methods for the 5 types of the events. The red dashed lines illustrate the predictability limit value as shown by the literature, while the grey lines are for guidance. Here no intensity threshold was used. For higher intensity thresholds see Appendix C – Figure C.1 and C.2*

Some spatial visualization of merging techniques for 5 randomly selected events are given in Appendix E – from Figure E.1 to E.5.

5.5 Nowcast Results with Merged Data

The best merged product, as per Section 5.4.4, the conditional merging between the temporal smoothed radar data with a 10 minutes window and gauge data (from here on referred to as CM instead of CM10) is used as an input for the nowcast with HyRaTrac and Lucas-Kanade algorithms.

5.5.1 Nowcast with HyRaTrac

Figure 5.17 illustrates the cross-correlation computed both temporally (upper) and spatially (lower) between the CM fed HyRaTrac nowcasts and either the input radar (left) or gauge data (right) in respect with different lead times and different types of events. The first noticeable result, is that both gauge and radar predictabilities have increased for all the event types (also both spatial and temporal performances when compared to Figure 4.6). Especially for the stratiform events the improvement is clearly visible. Through the merged input, the gauge predictability of the stratiform events is increased to more than 30 minutes, and the radar predictability to about 60 minutes. The stratiform events are followed up by the mixed events, whose gauge and radar predictability are respectively higher than 20 and 30 minutes.
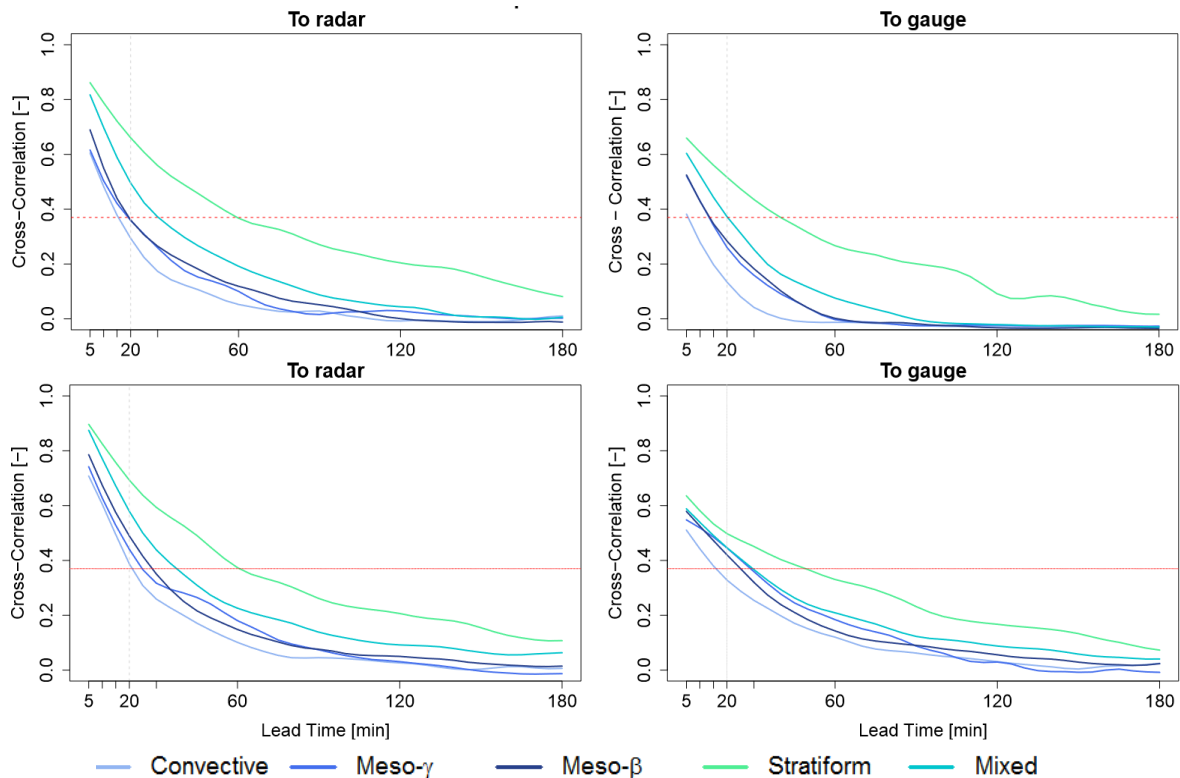
Figure 5.17: *The event median cross correlation for each event type as nowcasted by HyRaTrac for different lead times, calculated temporally upper row (averaged over 100 station per event) and spatially – lower row (averaged over the time steps per event). The red dashed line indicates the correlation value of 0.37 for the predictability limit, while the grey dashed line is a reference to the 20 min predictability limit as suggested by the literature. No intensity threshold was used here. For higher intensity thresholds see Appendix B – Figure B.5 and B.6.*

For the other remaining convective types, the radar predictability is at least 20 minutes, while the gauge one is about 15 minutes. The local convective type is the worst in performance, even though when advised with Figure 3.6, most of convective events live longer than 30 minutes. This might suggest a shortcoming of the HyRaTrac to nowcast the convective events. Overall, there is a pattern in the results; i) storms that are bigger and more persistent have a higher predictability; and ii) as seen also with the raw radar data as an input to the nowcast, the spatial cross-correlation is higher than the temporal one. However there seems to be an exception in the stratiform case where the spatial radar predictability behaves more or less the same as the temporal one.
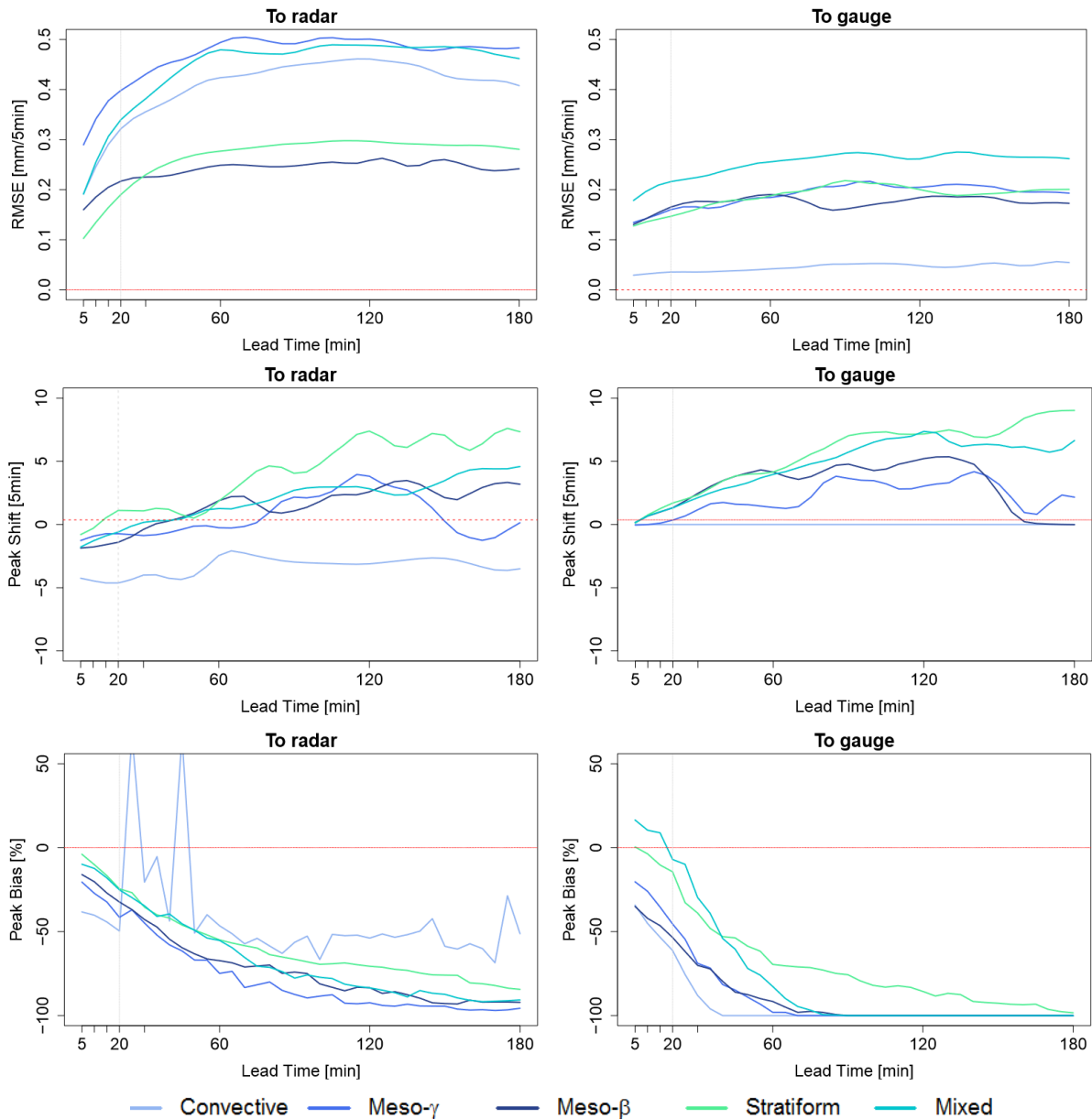
Figure 5.18: *The median of the temporal performance: RMSE (upper), Shift in Peak Time (middle) and Peak Bias (lower) for each of the event types of the HyRaTrac forecast with different lead times, compared to the input radar data (left) and gauge data (right). The red dashed line indicates the limit performance value of the criteria, while the grey dashed line is a reference to the 20 min predictability limit as suggested in literature. No intensity threshold was used here.*

Moreover, Figure 5.18 illustrates the additional temporal performance of the CM fed nowcast for the different event types. When compared to the Figure 4.7 from the RR fed nowcast, a great improvement of the peak shift is observed for both results and especially for the longer lead times. This is not the case though for RMSE, as for all the events it seems that the RMSE for both inputs has increased. Regarding the peak bias the improvement is mainly seen at the comparison to the gauge observation; the RR fed nowcast has a consistent underestimation of 100% for the peak volume. However, this is not the case for the CM fed HyRaTrac at the first hour of the nowcast. Whilst at the predictability a ranking of the events was displayed clearly, at the temporal performance it is not true that the best performance is always reached by the stratiform events.

5.5.2   Nowcast with Lucas-Kanade

Figure 5.19 illustrates the cross-correlation computed both temporally (upper) and spatially (lower) between the CM fed LK nowcasts and either the input radar (left) or gauge data (right) in respect with different lead times and different types of events. As with HyRaTrac there is clearly an improvement in both spatial and temporal predictability (compared with Figure 4.8). The biggest improvement seems to be for the temporal radar predictability.
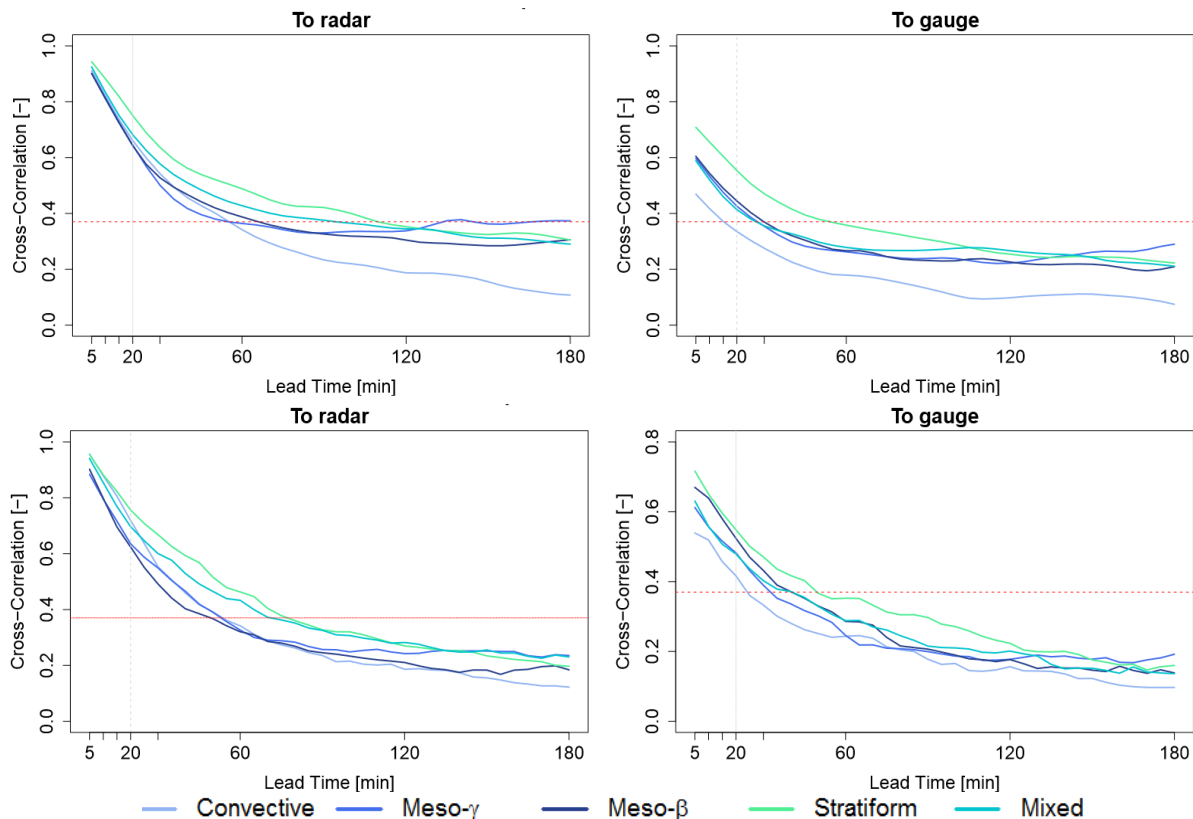
Figure 5.19: *The event median cross correlation for each event type as nowcasted by Lucas Kanade for different lead times, calculated temporally – upper row (averaged over 100 station per event) and spatially – lower row (averaged over the time steps per event). The red dashed line indicates the correlation value of 0.37, while the grey dashed line the 20 min predictability limit as in the literature. No intensity threshold was used here. For higher intensity thresholds see Appendix B-Figure B.7-B.8.*

Here all the events display a predictability higher than 50 minutes, even for the convective events! This is surprising as the convective events are very short lived, and thus are not expected to have predictability higher than 30minutes. However, when consulting with the storm characteristics diagram (see Figure 3.6 from Section 3.4.1), we see that the median of the convective storm duration is 100 minutes (more than one hour). Thus, since the plots shows the median values for each event type, it is still reasonable that the convective events display a predictability skill up to 50 minutes lead times.
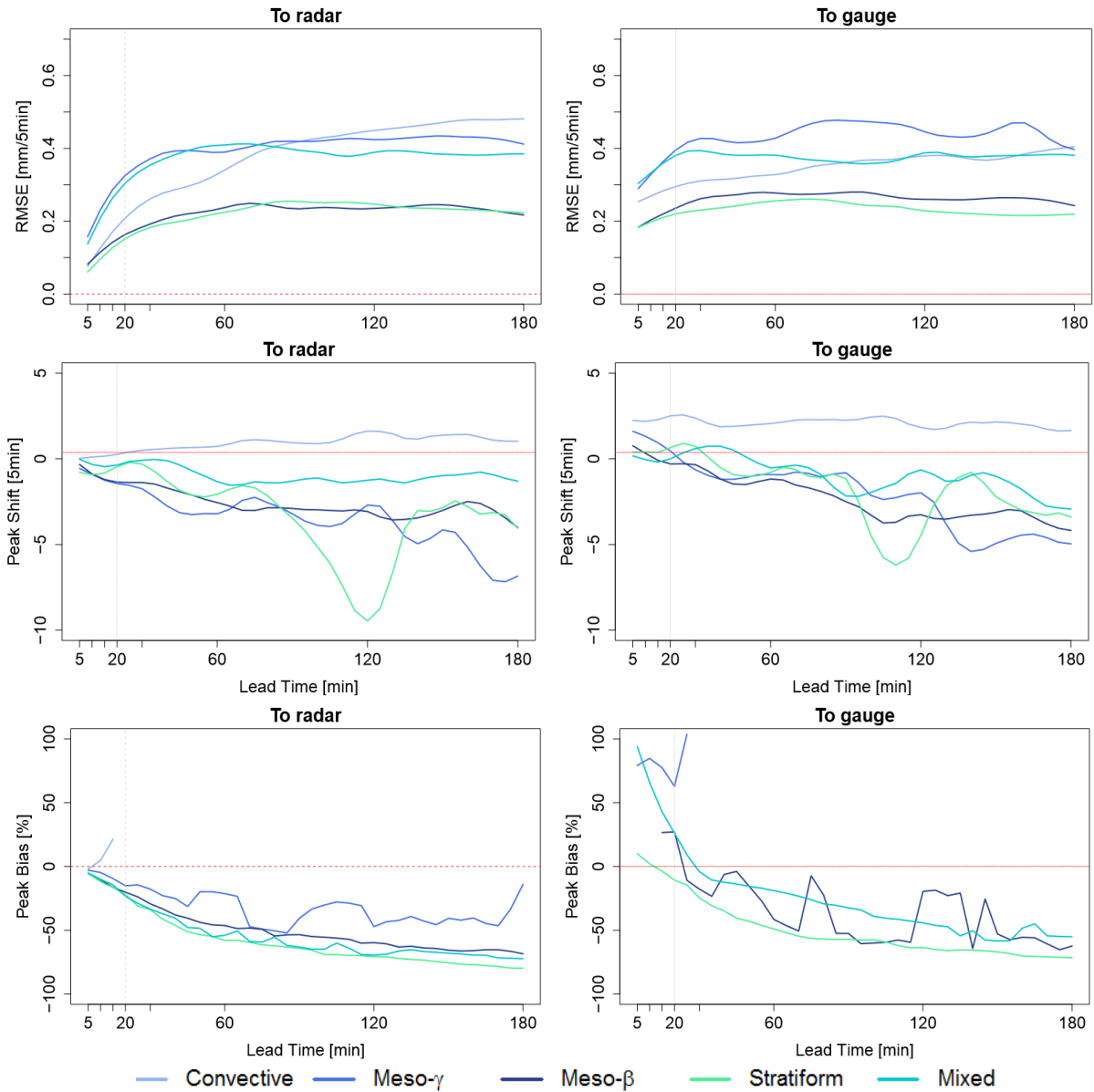
Figure 5.20: *The median of the temporal performance: RMSE (upper), Shift in Peak Time (middle) and Peak Bias (lower) for each of the event types of the LK method with different lead times, compared to the input radar data (left) and gauge data (right). No intensity threshold was used here.*

The gauge predictability has also increased considerably when compared to the one from Figure 4.8 of Section 4.4.2. Mainly the spatial performance, for all the storms. outperform the predictability limit of 20 min from the literature review. Overall, unlike HyRaTrac here only

one pattern is observed; that the bigger and longer the storm the longer the predictability. Which predictability, either the spatial or the temporal one, can be modelled better is hard to say. For the radar predictability the temporal one shows a better performance, however for the gauge one (as in the case with the RR or with HyRaTrac) the spatial one shows a better performance. Figure 5.20 illustrates the temporal performance of the CM fed LK nowcast for different types of events. When compared to Figure 4.9 of Section 4.4.2 as in HyRaTrac there is an increase in the temporal RMSE for all the lead times, and a slight improvement of the temporal peak shift and peak bias.

5.5.3 Comparison between HyRaTrac and Lucas-Kanade

Figure 5.21 and Figure 5.22 give a direct comparison of the temporal and spatial predictability limits of the two nowcast methods (HyRaTrac and Lucas-Kanade) when compared to radar and to the respective gauge information averaged over all the events. Unlike the case with the RR as input (see Figure 4.10 and Figure 4.11 from Section 4.4.3), the performance of the two methods is quite different from one another: the LK predictability seems to be higher than the one from HyRaTrac, especially for the temporal one. While for the spatial predictability (at least for high intensity thresholds) the gap between the two methods is not quite big, for the temporal predictability the LK nowcast is clearly the superior method.

Apparently, it seems that CM enhances the performance of both methods, but particularly it gives more advantage to the field motion based nowcast. Since CM makes the rainfall patterns more persistent in time and space (due to smoothing), the optical flow can estimate better the advection velocities. Moreover, it has to be mentioned that the LK method uses three time steps to calculate the motion field, while HyRaTrac only uses two time steps, which may not be enough to evaluate correctly the advection vectors. Another reason that may disadvantage the HyRaTrac, is the matching of the storms. For a storm to be matched, it needs to be in a certain vicinity of the pre-existing state of the storm, if this is not the case, then the storms will not be matched and thus have global or not updated advection velocities. Despite of the true nature of the error, it can be concluded that a proper estimation of motion field can increase the predictability up to 10 or 30 minutes (event 1 hour) depending on the intensity threshold and the predictability considered.
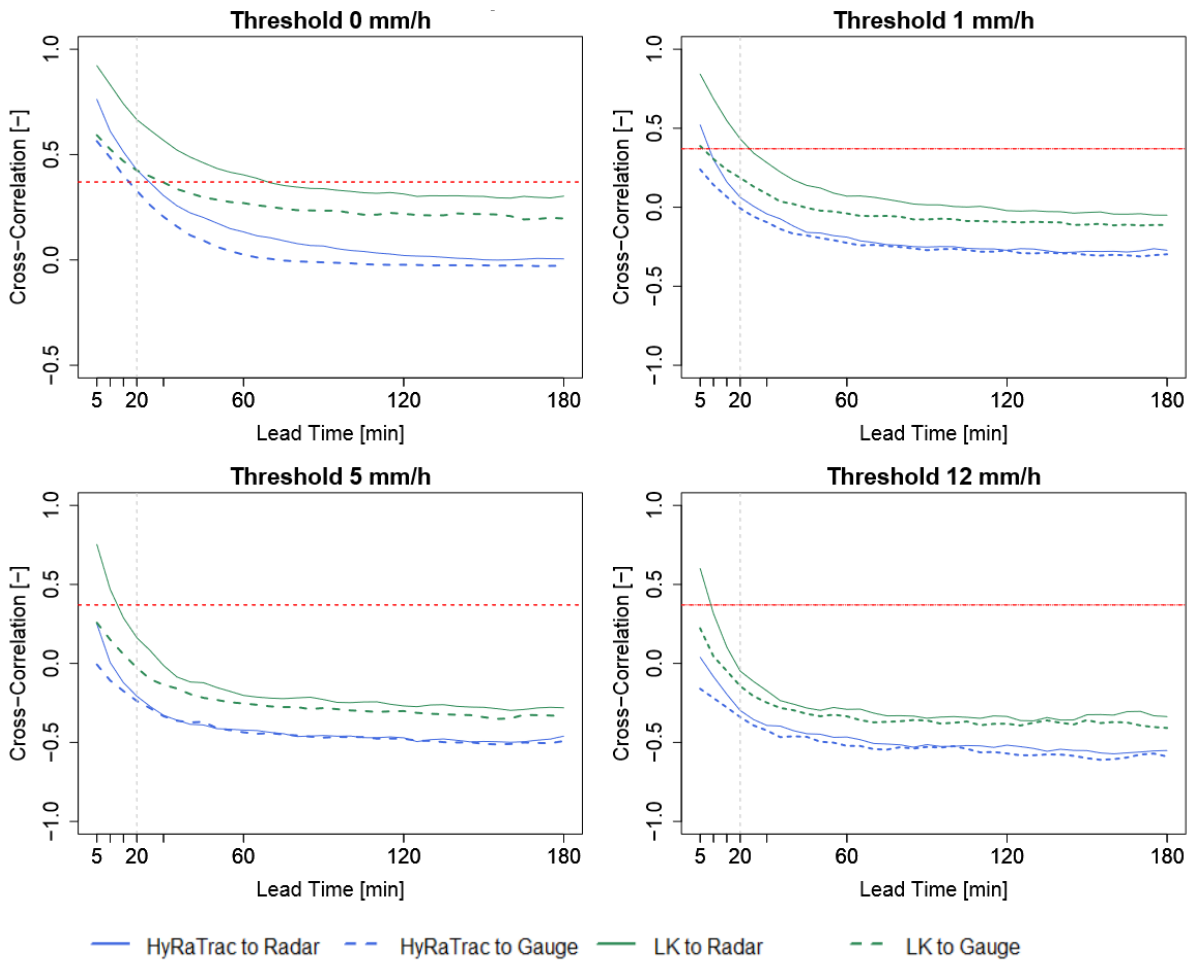
Figure 5.21: *The average temporal cross-correlation over all the events for the HyRaTrac and Lucas-Kanade Nowcast for the lead times from 5 min to 180 min (every 5min) for intensities higher than 0.12 mm/h, 1 mm/h, 5mm/h and 12 mm/h calculated based on raw radar input (solid lines) and gauge reference (dashed line). The red dashed line indicates the correlation 0.37 considered as the predictability limit, while the grey dashed line indicates the limit predictability as shown in the literature (20min).*
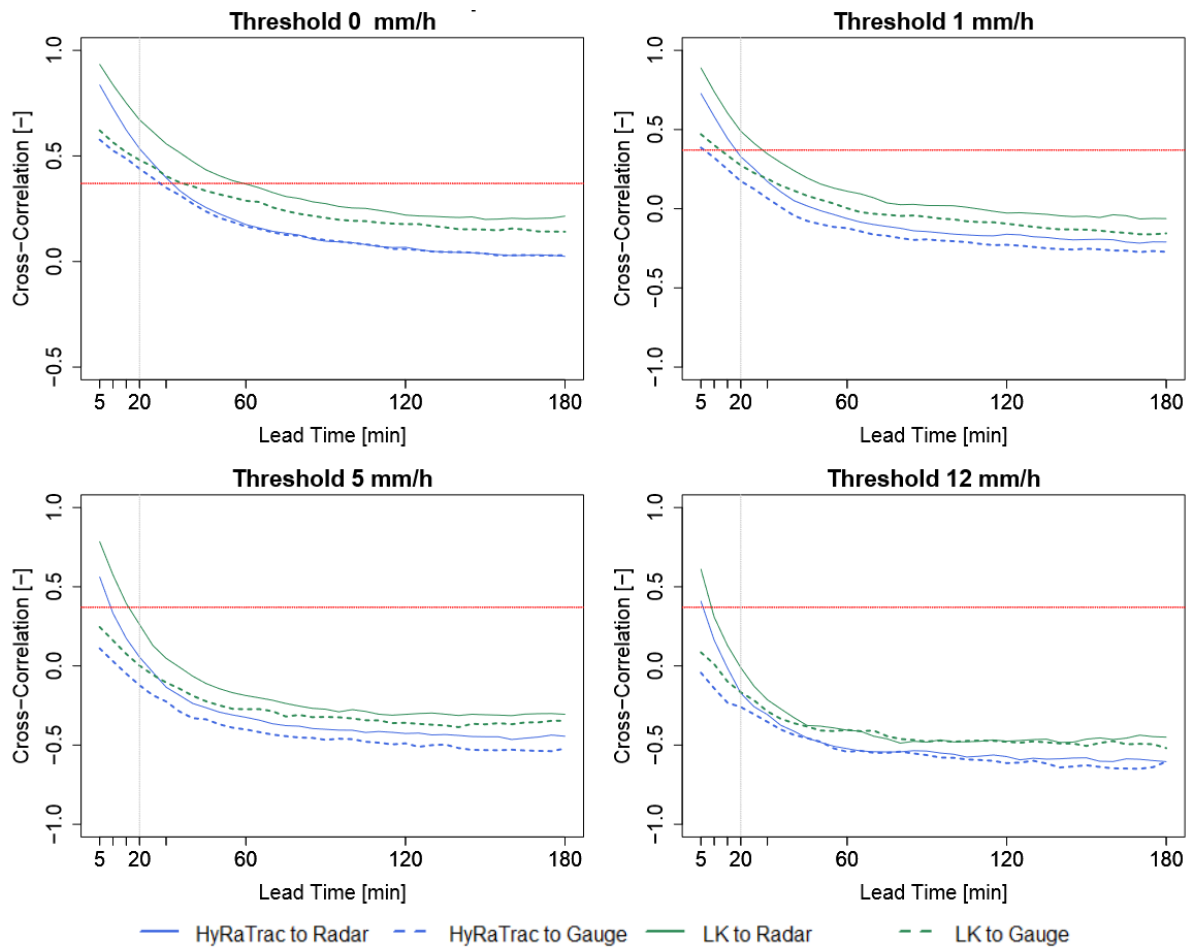
Figure 5.22: *The average spatial cross-correlation over all the events for the HyRaTrac and Lucas-Kanade Nowcast for the lead times from 5 min to 180 min (every 5min) for intensities higher than 0.12 mm/h, 1 mm/h, 5mm/h and 12 mm/h calculated based on raw radar input (solid lines) and gauge reference (dashed line). The red dashed line indicates the correlation 0.37 considered as the predictability limit, while the grey dashed line indicates the limit predictability as shown in the literature (20min).*

5.5.4   Improvement of Rainfall Predictability

Figure 5.23 illustrates the direct comparison for all events and performance criteria between both nowcast methods fed with RR and CM data. Here instead of absolute values of predictability, the percent of improvement for each performance criteria is plotted for the nowcast compared to the radar input and the gauge reference. A plus value (the green area) indicates that the nowcast fed with CM performs better than the one fed with RR data. For the negative values (the red area) the opposite is true; the RR fed nowcast has better performance.

For the low threshold it seems that for all the lead times the CM as input data improves the nowcast results, not only compared to the gauge data (which is expected due to the merging technique) but also when compared to the radar data. This supports the assumptions that a conditionally merged product is more persistent in time and easier to be tracked than the structures from the raw radar data. The consistent underestimation of rainfall intensities and the discrete nature of the radar measurements, may cause such bad performance of the RR fed nowcast even when compared to the input radar data.

Nevertheless, it is not clear to say which of the nowcasting methods is more favoured by the use of the CM data. Regarding the spatial and temporal predictability, the LK method is improving the most, however in terms of peak bias and shift HyRaTrac seems to have the most benefits. In terms of spatial and temporal RMSE, both methods have very similar results, with very little change at the high lead times. Moreover, while this pattern is true for the predictability throughout all thresholds set, this is not the case for the peak performance and the RMSE (see Appendix D – from Figure D.1 to D.4 for more detailed information regarding the other thresholds) for the high intensity thresholds. There are times (when compared to the gauge for higher intensities than 5mm/h) that the RR based nowcast yields better results. Thus, there is a limit to the advantage of using the CM data which is varying with the intensity threshold, event type, the lead time and the performance at hand.

To conclude, a better estimation of the rainfall field can improve the predictability of the storms both when nowcasted with an object-oriented or a field-oriented approach. The quality of the radar input can affect the predictability skill of the method from 10 min up to 1 hour lead time depending on the intensity threshold set. However, for lead times longer than an hour, the rainfall field cannot bring any additional value to the predictability. This is because at such lead times, the rainfall structure has changed so much that the Lagrangian persistence is not valid any-more. Processes like birth, growth, decay and death must be included in order to increase the predictability of the storms even more.
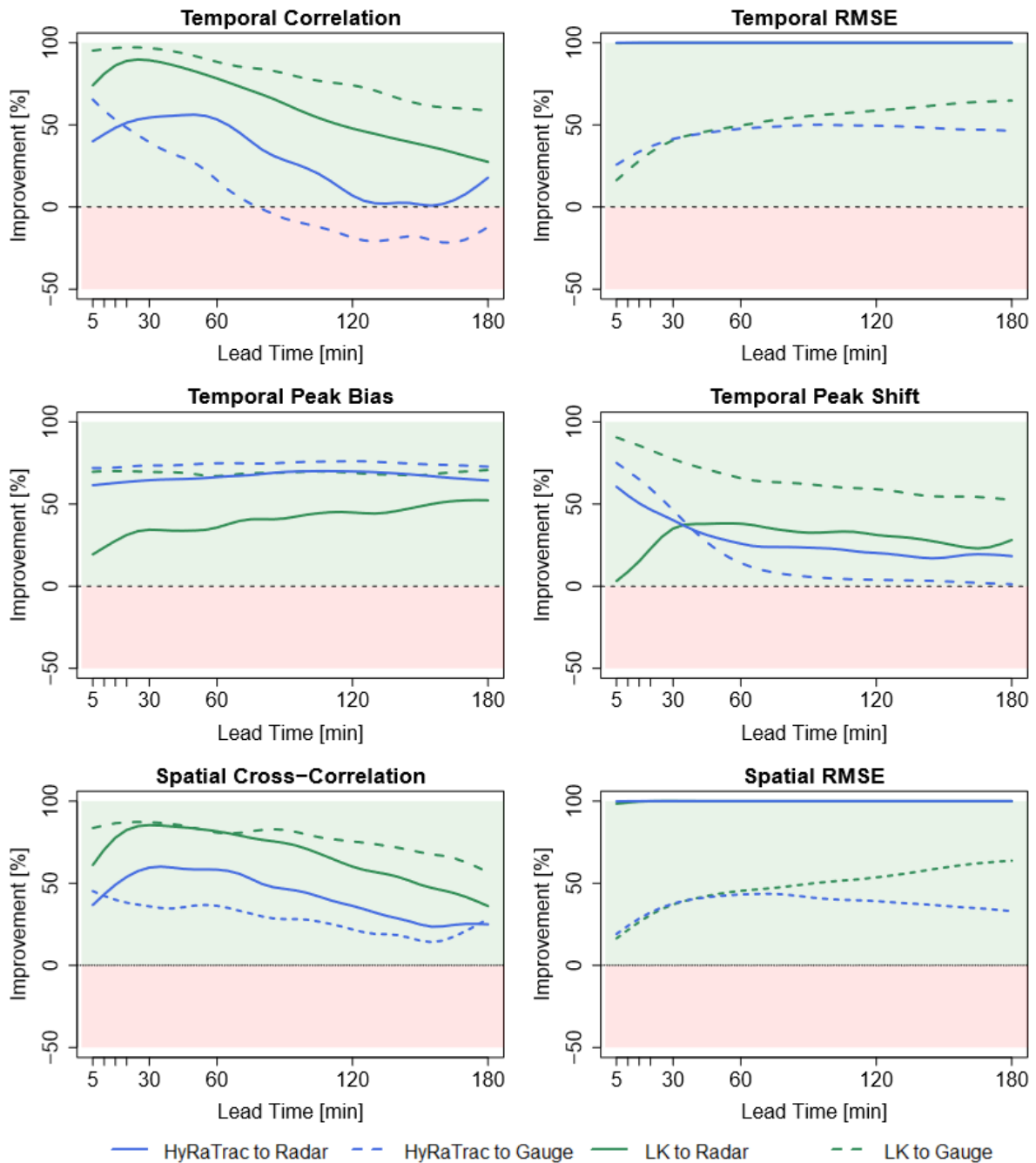
Figure 5.23: *The average improvement of Spatial and Temporal Criteria for each lead time (from 5 to 180 min) when using CM data instead of RR as an input for both HyRaTrac and LK nowcast method. The green area indicates a better performance by the CM fed nowcast, and the red area indicates a better performance by the RR fed nowcast. No intensity threshold was used here. For higher intensity threshold see Appendix D – from Figure D.1 to D.4.*

6.   k-NN RESAMPLING OBJECT BASED NOWCAST

As seen in Chapter 5, the use of merged radar data can improve the predictability of the nowcast up to 30-60 min lead times depending on the storm type. Past this lead time, the nowcasts issued by HyRaTrac were useless. Of course a part of the error is attributed to the matching and tracking of the storm, nevertheless a main deficit of the HyRaTrac method is that, by implementing the Lagrangian Persistence, it doesn't consider the intermittence nature of the rainfall. Whilst the birth of the storms is mostly related to meteorological conditions (thus information from NWP is necessary), the changing rate and death time may be learned from past observed storms. For instance from previous observed convective storms, one knows how long they typically last and what is the rate of decay. Such information can be useful when a convective storm is identified and a nowcast must be issued. Here a nearest neighbour approach is implemented to perform such task; when a new storm is identified, similar past storms (neighbours) are selected and their behaviour is assigned to the nowcast of the storm at hand. Naturally the first questions that rise are: How to evaluate similar storms and do similar storms actually behave similarly? In other words, how can the similarity of the storms be channelled through features? And once similar storms are identified, could their behaviour improve the HyRaTrac nowcast skill?

These are the main questions that are treated in this chapter. Here, a k-NN is applied at the storms identified from the HyRaTrac algorithm, in order to investigate if past information can bring additional value to the HyRaTrac nowcast, especially regarding the decay and death of storms. Section 6.1 gives a brief introduction to the k-NN theory and the methods chosen for the predictor estimation, followed up by Section 6.2 where the structure of the k-NN implemented here as an extension of HyRaTrac is illustrated. The model set up and optimization is described in Section 6.3 and the respective results for such implementation are discussed in Section 6.4.

6.1 Basics of the k-NN Approach

The nearest neighbour approach is mainly used in regression or classification purposes. The nowcast application, since it aims in simulating future behaviours, is in a way a regression problem. The assumption here is that events that exhibit similar features (also referred as predictors) are expected to behave similarly (have similar responses). Hence, each event can be described by N features (x_1, x_2, ..., x_N) and has a corresponding behaviour (y). When a new event, with the respective features x^{new}_1, x^{new}_2, x^{new}_N, is observed and whose response is unknown, the k-responses (y^1, y^2, ..., y^k) from the most similar past observed events, are used to predict the new response (y^{new}). The response prediction from the k-nearest neighbours (hence the name k-NN) is done in two steps.

In the first step, the most similar events are selected based on the distance of the new event with the past ones on the feature space. Typically, Euclidean Distance is used as per Equation 6.1. Given the prior knowledge on the importance of the predictors, weights can be assigned at each of them for the distance calculation. Based on the distance, the events are then ranked from the most similar to the least similar.

$$E^i_d = \sqrt{\alpha_1\left(X^{new}_1 - X^i_1\right)^2 + \alpha_2\left(X_2 - X^i_2\right)^2 + \ldots + \alpha_N\left(X_N - X^i_N\right)^2} \qquad (6.1)$$

where:
N   – number of features or predictors
E^i_d – is the Euclidean distance between the ith past and new event
X_i – the feature describing the ith past event
α_N – the importance weight of the Nth features/ predictor

Once the most k-similar events are recognized, their response is averaged according to the respective weights for the estimation of the new response as per Equation 6.2.

$$y^{new} = \sum_{i=1}^{k} w^i y^i \qquad (6.2)$$

where:
y^{new} – is the new event behaviour
k   – is the number of similar neighbours selected
y^i – is the behaviour from the ith neighbour
w^i – is the weight for the ith neighbour

The weights for each neighbour (w^i) can be calculated based on the rank probability as per Equation 6.3. So, depending on the number of neighbours (k), the first one will have the highest weight, followed up by the second one and so on.

$$w^i = \frac{\dfrac{1}{k}}{\sum_{(i=1)}^{k} 1 + \dfrac{1}{k}} \qquad (6.3)$$

where:
w^i – is the weight for the ith neighbour
k   – is the rank of the k- most similar neighbours

However, one of the main drawbacks of the rank probability is that, it gives importance also to the neighbours that might have a significantly high distance from the new one, but are still within the k-selected neighbours. Thus, a distance probability is thought to be more

accurate, as it gives the highest weight to the neighbours in the close vicinity as described in Equation 6.4

$$w^i = \frac{\dfrac{1}{d_i}}{\displaystyle\sum_{(i=1)}^{k} 1 + \dfrac{1}{d_i}}$$

where:
w^i – is the weight for the ith neighbour
k – is the number of neighbours
d_i – is the distance from the ith neighbour

The best k-number for a certain application is learned by optimizing a loss function either on a split sampling or on a leave one out cross validation. The weighted average of the k-past behaviours, creates a new one, not previously observed, which satisfies the statement that nature doesn't repeat itself. Nevertheless, since the new behaviour is derived from past observed one, the approach itself is unable to predict extreme behaviours; the new behaviour will be within the range of the past ones. Even though the k-number of neighbours can be learned, the performance of the k-NN approach is largely depending on two factors; i) the right features or predictors have been selected to describe the events, and ii) the Euclidean Distance can describe adequately the similarity between events.

6.1.1   Predictors Selection

Finding the right predictors is an active field of research in hydrology. Typically, most of the predictors selected have a physical relationship, for instance when forecasting the discharge at a location, the rainfall volume over the last days can be selected as a predictor [Haberlandt, 2015] as the two processes (rainfall-runoff) are connected to each other. However, having such linked predictors is not always the case. In such cases, the presence of a strong correlation (for instance Pearson correlation) between the predictors and the response (target variable) is used to identify the most important predictors. Nonetheless, a correlation can only measure a linear relationship between two variables, thus it doesn't account for the interaction between predictors. For instance, two predictors A and B may have low correlation with target C, but due to the interaction of A and B, both of them can predict satisfactory the value of C. Thus, another approach should be considered, in order to account as well for the interaction between the predictors.

Since the k-NN approach selects its' prediction based on the Euclidean Distance, only the events characterized by predictors in the close vicinity of the new event are selected. Often a "local" linear approximation is assumed. Hence multiple-regression models can be fitted into all the selected predictors and targets (by the least square minimization approach) in order to recognize the important predictors based on the standardized regression coefficients as per Equation 6.5:

$$Y = \beta_1 X_1 + \beta_2 X_2 + ... + \beta_N X_N + \varepsilon$$

where:
N – the number of predictors
β_i – the regression weight for the ith predictor
X_i – the ith predictor
Y – the response

Nevertheless, the regression coefficients are dependent on the rank on which the predictors are introduced into the regression, and moreover they are susceptible to the co-linearity among the predictors. Alternatively, a Relative Importance Analysis (RIA) could be employed in multiple-regression models to determine the variable importance [Tonidandel and LeBreton, 2011]. Unlike the standardized regression weights, RIA creates a new set of predictors that are orthogonal to each other, in order to avoid co-linearity and to recognize the most important ones.
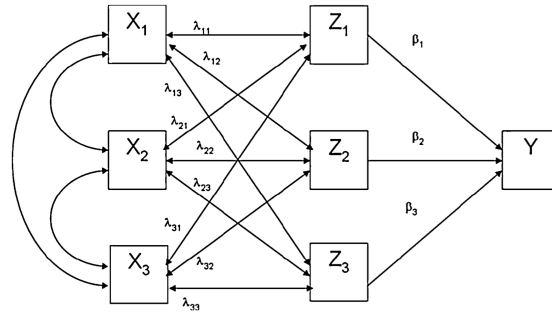
Figure 6.1: *Calculation of the relative weights for a regression model with three predictors (X_1, X_2, X_3) transformed in a set of orthogonal predictors (Z_1, Z_2, Z_3) for the estimation of the final target variable (Y) [Tonidandel and LeBreton, 2011].*

For instance, for a set of predictors (X_1, X_2 and X_3) whose response is Y, three new predictors Z_1, Z_2 and Z_3, that are orthogonal to each other and maximally related to the original predictors set, are created (as illustrated in Figure 6.1). Two types of standardized regression weights are determined; $ß_N$ representing the relationship between the orthogonal predictors and the target variable Y, and $\lambda_{N,N}$ representing the relationship between the orthogonal predictors and the actual ones. The final relative weights (αj) for each original predictor are then calculated by adding up the products of squared standardized regression coefficients as indicated by the Equation 6.6.

$$\alpha_j = \beta_1^2 \lambda_{j1}^2 + \beta_2^2 \lambda_{j2}^2 + \beta_3^2 \lambda_{j3}^2$$

where:
α_j – is the weight for the jth predictor
β_k – the regression weight for the kth orthogonal predictor
λ_{jk} – the standardized regression weight for the jth predictor X
to orthogonal predictor Z.

(6.6)

The relative weights found based on RIA can give useful information about the importance and the ranking of variables when given a certain regression model. Moreover, to account for the statistical importance of the weights, a bootstrapping procedure is possible as described by Tonidandel et al. (2009). Nevertheless, for the computation of the relative weights, RIA assumes that the regression model at hand is the correct one. Thus, the weights are not to be used as direct weights in the k-NN approach, but rather as an additional information to understand which of the predictors play an important role in the explanation of the target variance. For this reason, a multivariate-regression will be fitted to the predictors in this thesis, only if the coefficient of determination (R^2) is high enough to indicate a linear relationship between the predictors and target variables, and the respective

relative weights from RIA will be used additional to the Pearson correlation to recognize the most important predictors.

However, the relationship between the predictors and target variables may be of non-linear nature. Hence, an additional method is used here, where no prior assumption about the system type is required. Such a method was proposed by Sharma and Mehrotra (2014), which is specifically designed for nearest neighbour approach given the prior knowledge of the most important predictor. The method is based on a metric called the Partial Information Content (PIC) and is computed from the Partial Information (PI) as in Equation 6.7.

$$PIC = \sqrt{(1 - \exp(-2PI))}$$

where:
PI – Partial Information as extension to the MI
PIC – Partial Information Correlation

(6.7)

The Partial Information itself is a modification of the Mutual Information (MI – as in Equation 6.8) in order to measure partial dependency between the predictors and the target variable, by adding predictors one at a time (step-wise procedure). Thus, when X predictors are available, their partial dependency on the target Y is as described by Equation 6.9

(6.8)

$$MI(X,P) = \int f_{X,P}(x,p) \log \left[\frac{f_{X,P}(x,p)}{f_X(x)f_P(p)} \right] dxdp$$

where:
MI – Mutual Information between variables
X,P – The two variables considered
$f_X(x)$ – The marginal probability density function of X
$f_P(p)$ – The marginal probability density function of P
$f_{X,P}(x,p)$ – The joint probability density function of X and P

$$PI(Y, X_{new} | X) = MI(Y, X_{new}, X) - MI(Y, X) - MI(X_{new}, x) + MI(x)$$

(6.9)

where:
MI – Mutual Information between variables
X   – The pre-existing identified predictor
X_{new} – The new predictor added step-wise
Y   – The target variable

The evaluation of PIC needs a pre-existing identified predictor from which the computation can start. If the pre-defined predictor is correctly selected, then through the Equation 6.9, the method is able to recognize and leave out the new predictors which are not related to the response and which don't bring additional value to the existing relationship between the current predictors and target variable. As proven by Sharma and Mehrotra (2014), for such cases the PI would be 0, thus consequently causing a PIC of zero in Equation 6.7. Contrarily, the stronger the relationship between the predictors, the higher the PI, and the closest is PIC to the value 1.

As in the case of a linear regression model where regression weights are assigned based on a least square method, relative weights can be derived as well for the k-NN regression application, as a relationship between the PIC metric and the associated partial correlation, as

indicated by Equation 6.10. These weights can be used as part of the Euclidean distance (Equation 6.1) when selecting the most similar responses.

$$\alpha_j = PIC_{X,X_j|X(-j)} \frac{S_{Y|X_{(-j)}}}{S_{X_j|X_{(-j)}}}$$

where:
α_j – predictor coefficient
X – the predictor
Y – the target response
$SY_{|X(-j)}$– scaled conditional standard deviations between the first predictor and the target
$S_{Xj|X(-j)}$– scaled conditional standard deviations between the additional predictor and the first one

In this thesis the NPRED package in R [Sharma et al., 2016] was used for the step-wise estimation of the predictors and the respective weights. As the implementation of the PIC requires a pre-determined predictor, the most important one as indicated by both Pearson Correlation and RIA, will be used as a start for the analysis. Then the other predictors, ranked important by the two mentioned methods, will be added one by one in order to determine their weights for the similarity estimation.

6.2 Developing the k-NN Model

The k-NN approach is applied on the storm scale. From the respective radar data for each of the 93 events used for the development of the nowcast methods, storms have been recognized and tracked in hind-cast mode by the HyRaTrac algorithm (as described from Section 2.2.3.1 and 4.1.1). Thus, a dataset with several types of storm is built and saved. The storms are saved with an ID based on the starting time and location, and for each time step of the storm evolution (noted as well on the ID), the spatial information is saved. Here the spatial rainfall intensities of a storm at a particular time step (in 5min) of the storms' life, is referred to as the "instance" of the storm (see Figure 4.1). A storm that has been observed for 15 minutes, consists of three "instances" each occurring at a 5 min time step.

Theoretically, since the aim of the nowcast is to have a cross-correlation above the predictability limit (0.37), a correlation distance can be employed to evaluate similar storms. Nevertheless, when thousands of storms are available, looking for the most cross-correlated[11] storm in the past can be very time consuming mainly for two reasons: i) The storms are located at different regions of the radar data. As one may expect that storms behave similarly no matter the location, one has to shift the instances of the past storms accordingly to match at best the storm-instance at hand. Hence, this should include a cross-correlation optimizer tool, which increases the computational time. ii) Additionally, the present instance of the storm should be cross-correlated (and optimized) to each time step (instance) of each of the previous storms to find the most similar neighbours, increasing considerably the computation time of the k-NN nowcast.

---

11 Cross-correlation is here referred to the correlation between the spatial rainfall intensities of any two storm regions (instances), indicating thus, if these two storms have similar spatial rainfall distributions at a particular instance of their storm life.

To overcome such high computation times, a four step approach is proposed for the application of the k-NN method on the object-oriented nowcast as illustrated in the Figure 6.2 and explained step by step from Sections 6.2.1 to 6.2.4.

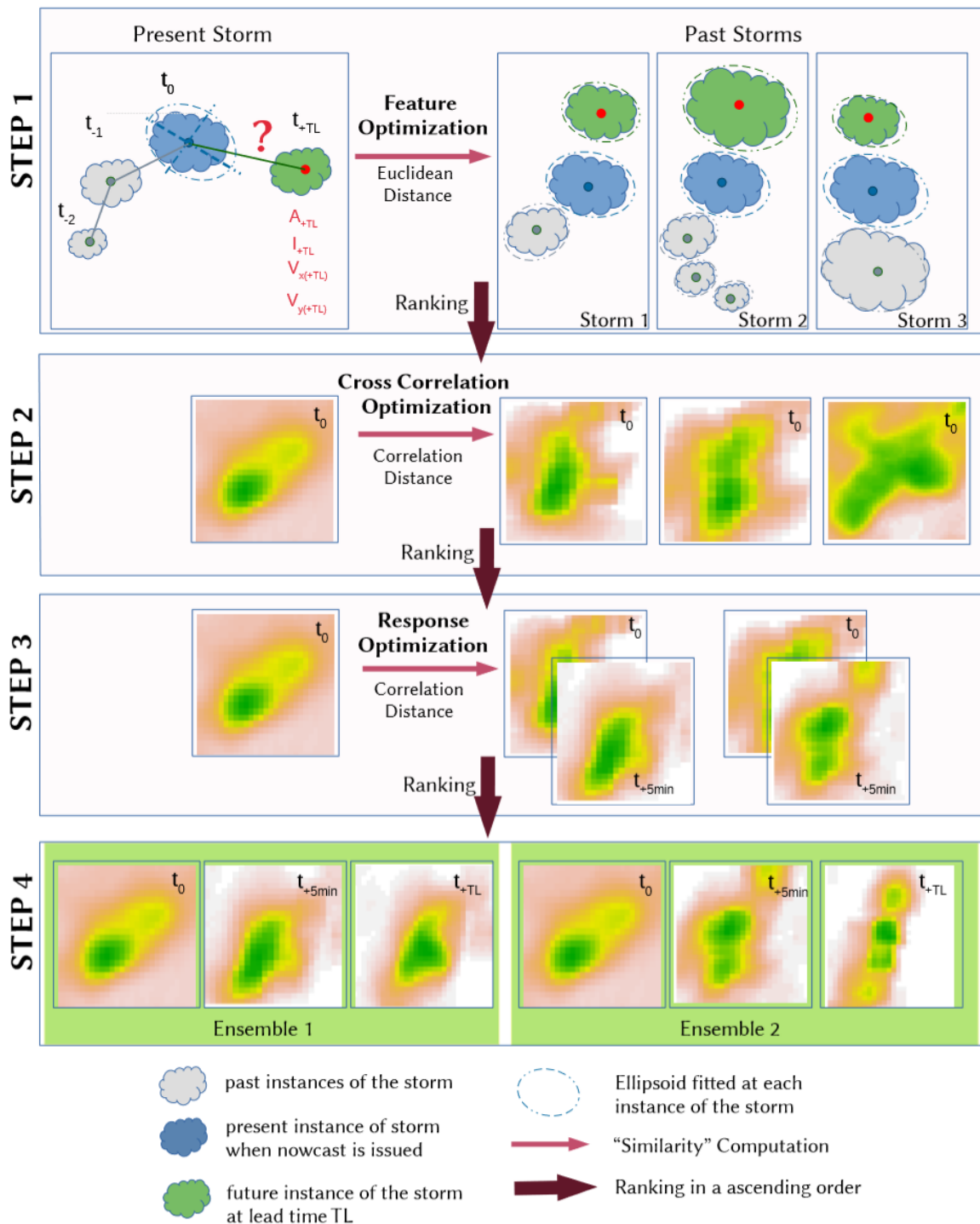**Figure 6.2:** *The main steps involved in the k-NN based nowcast with the estimation of similar storms (step 1 and step 2) for a new observed storm, and the future predictions according to the most similar storms (step 3 and step 4).*

For any given storm that has to be nowcasted, in Figure 6.2, the blue-instance represents the current state of the storm at the initialization of the nowcast (time t_0), the grey-instances

represent the previous states of the storm (t_{-1}, t_{-2}), and the green-instances represent the future state of the storm at a specific lead time (t_{+LT}) that is to be nowcasted. These instances shown in Figure 6.2 and 6.3 as clouds, represent the spatial intensities at a 5 min time step inside the storm region. At first the storm-instances (for both the past storm database and the present to-be-nowcasted storm) are treated as objects with certain features (serving as predictors) that describe them at each time step of the storms' life (Step 1 - Figure 6.2). In **Step 1**, through the Euclidean Distance on these features, the most-similar instances of the past storms (indicated in the Figure 6.2 -right in blue), that have similar features and that might exhibit similar patterns in the future, are selected (see Section 6.2.1). Once a similar instance of a past storm has been recognized (the blue-instance in Figure 6.2-right), the future instances of that storm (the green-instances in Figure 6.2-right, each for a specific lead time from the occurrence of the selected similar blue-instance), are treated as future instances (the green-instances in Figure 6.2-left) of the to-be-nowcasted storm.

After recognizing the most similar storms (in terms of predictors), in **Step 2**, a cross-correlation optimization is performed between each of the blue-instances of the past selected storms (Step 2: Figure 6.2-right) with the current blue-instance of the to-be-nowcasted storm (Step 2: Figure 6.2-left). The past selected storms are then ranked based on the cross-correlation between the past-present blue-instance pairs (see Section 6.2.2). Furthermore, to ensure that the nowcast fields, derived from each of the neighbours, will be representative of the to-be-nowcasted storm, in **Step 3**, a cross-correlation distance is computed between each of the past green-instances (at time t_{+5min} from the neighbours) with the current blue-instance of the to-be-nowcasted storm. The storms are then ranked according to descending cross-correlation values. Only k-neighbours are selected, and their respective future behaviours (the green-instances) are set as the future responses of the current blue-instance of the to-be-nowcasted storm (see Section 6.2.3). At last, in **Step 4**, either a probabilistic or a deterministic nowcast is issued. If a deterministic nowcast is selected, then the green-instances of the k-neighbours are averaged according to their rank for each lead time (see Section 6.2.4). Contrary, if a probabilistic nowcast is selected, k-ensembles are issued independently; to each neighbour a probability is assigned according to their distance with the to-be-nowcasted storm (see Section 6.2.4).

6.2.1   Step 1 – Feature Optimization

At first storms are treated like objects that exhibit certain features (predictors) like area, intensity, lifetime etc., at each instance (state) of the storm life until the storm dies (and the predictors are all set to zero). The features of the objects are categorized in two parts a) present features and persistence features, as illustrated in Figure 6.3 (shown respectively in blue and grey). The present features describe the current state of the storm at the time of nowcast (denoted with t_0 in Figure 6.3), and are calculated from one instance of the storm. To compute certain features, an ellipsoid is fitted to each instance of the storm. The persistence features, on the other hand, describe the past instances of the storm (denoted with t_{-1}, t_{-2} in

Figure 6.3) and how these features change over the past life of the storm (as illustrated in Table 6.1). For example, the change of area from time t_{-2} to t_{-1} is a persistence feature. If the change of area is averaged from the last 30 minutes of the storm, then the average area changing rate is obtained. While the present features help to recognize storms that are similar at the given instance when the nowcast is issued, the persistence ones give the information about the past evolution of the storm. The features used here to describe the storms as objects, and hence tested as predictors, are indicated in Table 6.1.
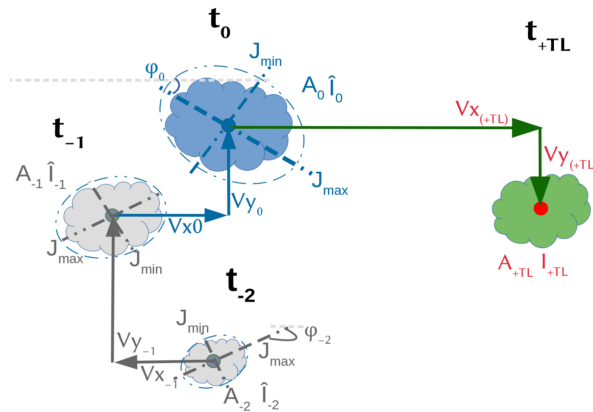
Figure 6.3: *The features describing the past (grey) and present (blue) instances of the storm used as predictors to nowcast the future instance of the storm (green) at a specific lead time (T_{+LT}) described by the 4 target variables (in red). The present instance (blue) indicates the state of the storm when the nowcast is issued.*

The aim of these features is to recognize the states of previously observed storms (illustrated in blue in Figure 6.2 and Figure 6.3) that are most similar to the current blue-instance of the to-be-nowcasted storm. Once the most similar past-blue-instances are recognized, the green-instances of these past-blue-instances at different lead times can be assigned as the future behaviour of the blue-instance of the to-be-nowcasted storms. The similarity is assessed based on the Euclidean Distance of the most important predictors. Nevertheless, to recognize the most important predictors, a sensitivity analysis has to be performed. Thus, specific target variables are needed in order to recognize the most important predictors and to conclude if storms with similar predictors do actually behave the same (have similar response or target variable values).

Since the storms are regarded as objects with specific features, similarity at different lead times is determined by the area, intensity and velocity of the storm instances. Thus, for each storm-instance, the area (A_{+LT}), mean intensity (I_{+LT}) and velocity in X (Vx_{+LT}) and Y (Vy_{+LT}) direction, for its' future states at +5min to +180min (every 5 min) lead times, are saved together with the past and present features of that storm-instance. In case the storm is dying before reaching up the 180min, then all the target variables are set to zero. The features and the target variables are saved for each of the storms recognized by HyRaTrac at each event. Before training and validating of the k-NN method, an importance analysis is performed for each of the target variables in order to recognize the most important predictors(see Section 6.4.1).

Table 6.1: *List of all the past and present features of the storm object that are investigated for their importance as predictors.*

	Features	Symbol
Present Features	area of the storm	**A**
	average spatial intensity of the storm	**Iave**
	maximum spatial intensity	Imax
	standard deviation of the spatial intensities	**Isd**
	global velocity	**Vg**
	x component of the local velocity	**Vx**
	y component of the local velocity	**Vy**
	ratio of the major and minor axis of the ellipsoid fitted at the instance of the storm	J
	orientation angle of the major axis of the ellipsoid fitted at the instance of the storm	φ
Past Features	change in area over the last two instances of the storm	$\Delta A_{10} = A_{-1} - A_0$
	change in average intensity over the last two instances	$\mathbf{\Delta Iave_{10}} = Iave_{-1} - Iave_0$
	change in maximum intensity over the last two instances	$\Delta Imax_{10} = Imax_{-1} - Imax_0$
	change in the standard deviation of intensity over the last two instances	$\mathbf{\Delta Isd_{10}} = Isd_{-1} - Isd_0$
	change in the global velocity over the last two instances	$\Delta Vg_{10} = Vg_{-1} - Vg_0$
	change in the horizontal local velocity over the last two instances	$\mathbf{\Delta Vx_{10}} = Vx_{-1} - Vx_0$
	change in the vertical local velocity over the last two instances	$\mathbf{\Delta Vy_{10}} = Vy_{-1} - Vy_0$
	change in the ellipsoid axis ration over the last two instances	$\mathbf{\Delta J_{10}} = J_{-1} - Jg_0$
	change in the ellipsoid major axis orientation over the last two instances	$\Delta\varphi_{10} = \varphi_{-1} - \varphi_0$
	average change in area over last 15 and 30 min	$\Delta A_{15,}$ $\mathbf{\Delta A_{30}}$
	average change in average intensity over last 15, 30 min	$\Delta Iave_{15,}$ $\Delta Iave_{30}$
	average change in maximum intensity over last 15, 30 min	$\Delta Imax_{15,}$ $\Delta Imax_{30}$
	average change in standard deviation of intensity over last 15 and 30 min	$\Delta Isd_{15,}$ ΔIsd_{30}
	average change in global velocity over last 15 and 30 min	$\Delta Vg_{15,}$ ΔVg_{30}
	average change in horizontal local velocity over last 15 and 30 min	$\Delta Vx_{15,}$ ΔVx_{30}
	average change in vertical local velocity over last 15 and 30 min	$\Delta Vy_{15,}$ ΔVy_{30}
	average change in ellipsoid axis ratio over last 15 and 30 min	$\Delta J_{15,}$ ΔJg_{30}
	average change in ellipsoid major axis orientation ratio over last 15 and 30 min	$\Delta\varphi_{15,}$ $\Delta\varphi_{30m}$

6.2.2   Step 2 – Cross-Correlation Optimization

Step 1 aims at finding out which instances of the past storms are most similar to the blue-instance of the to-be-nowcasted storm. Once these past blue-instances have been recognized, it is important to select only the ones that display similar spatial rainfall variability with the current to-be-nowcasted storm. Since these instances are most probably at different regions of the radar image, a transposition (spatial shifting) of the past blue-instances is done in order to maximize the similarity between the past and the to-be-nowcasted blue-instances. This is done by the cross-correlation optimization. Given the coordinates in x and y, the past blue-instance is shifted with different lags in both x and y direction in order to maximize the cross-correlation between the to-be-nowcasted and selected past storm blue-instances. The optimization includes only the non-zero values of both storms and is expressed by the following equation. Once the maximum cross-correlation is found, the x and y lags together with the maximum cross-correlation value are saved. The 100 storms selected from Step 1, are then ranked according to the highest displayed cross-correlation.

$$x_i, y_i \in D \rightarrow max \int_{x_i, y_i}^{D} \rho_{For_{(x+x_i)(y+y_i)}, Obs} = \frac{\sum_{i=1}^{n} \left(For_{(x+x_i)(y+y_i)} - \overline{For_{(x+x_i)(y+y_i)}} \right)\left(Obs_i - \overline{Obs} \right)}{\sqrt{\sum_{i=1}^{n} \left(For_{(x+x_i)(y+y_i)} - \overline{For_{(x+x_i)(y+y_i)}} \right)^2} \sqrt{\sum_{i=1}^{n} \left(Obs_i - \overline{Obs} \right)^2}}$$

(6.11)

where:

x_i, y_i	–	the x and y lags
Obs	–	the observed field of the blue instance of the to-be-nowcasted storm
$For_{(x+xi,y+yi)}$	–	the blue instance field from the kth neighbour shifted with x_i, y_i lag
$\overline{For_{(x+xi,y+yi)}}$	–	mean of the the blue instance field (from the kth neighbour) shifted with x_i, y_i lag
$\overline{Obs}$	–	mean of the observed field of the blue instance of the to-be-nowcasted storm

6.2.3   Step 3 – Response Optimization

The future states of the selected past blue-instances, are then assigned as response (green-instances) for the blue-instance of the to-be-nowcasted storm. Nevertheless, it may happen, that even though at a time step, the two storms display a high spatial cross-correlation, the past green-instances are not highly correlated to the blue one of the to-be-nowcasted storm. Since it has been shown that storms are persistent for 5 min (as Eulerian Persistence works well for 5 min lead times), a further ranking of the storms is done based on the cross-correlation between the past green-instances at time +5min with the blue-instance of the to-be-nowcasted storms.

$$max \int_{k}^{100} \rho_{For_{k,+5min}, Obs} = \frac{\sum_{i=1}^{n} \left(For_{k,+5min} - \overline{For_{k,+5min}} \right)\left(Obs_i - \overline{Obs} \right)}{\sqrt{\sum_{i=1}^{n} \left(For_{k,+5min} - \overline{For_{k,+5min}} \right)^2} \sqrt{\sum_{i=1}^{n} \left(Obs_i - \overline{Obs} \right)^2}}$$

(6.12)

where:

Obs	–	the observed field of the blue instance of the to-be-nowcasted storm
$For_{(k,+5min)}$	–	the green-instance field at t_{LT}=+5min from the kth neighbour
$\overline{For_{(k,+5min)}}$	–	the green-instance field at t_{LT}=+5min from the kth neighbour
$\overline{Obs}$	–	mean of the observed field of the blue instance of the to-be-nowcasted storm

6.2.4   Step 4 – Deterministic and Probabilistic Nowcast

For the ensemble nowcast mode, the probability of occurrence for each of the neighbours is determined based on the rank probability as determined in Step 3. A distance-based probability was also tested, nevertheless the results from the rank probability were more stable and could converge to a specific number of k-neighbours when used for the deterministic approach. Only neighbours that display a distance higher than 0.37 (as dictated by the predictability limit in the literature) are selected. In case this condition is not satisfied, then the number of ensembles is restricted to 10 for each blue-instance of the to-be-nowcasted storms. Since two or more storms may be observed at the same time step when issuing a nowcast, a combination of all of the ensembles at the storm scale from k-NN is build up. Depending on their probability of occurrence described in Equation 6.13, the maximum number of ensembles is fixed to 150.

$$Pr_i = \frac{(1/k_i)}{\sum_{k_i=1}^{k} (1/k_i)} \qquad (6.13)$$

where:
k   – is the number of similar neighbours selected
k_i – is the descending rank from the ith neighbour
Pr_i – is the rank probability of the ith neighbour

For the deterministic approach, an averaging of the most-similar ensemble storms is done for each lead time according to the rank probability as expressed in Equation 6.14. The response here is referred to the spatial distribution of the rainfall intensities at 1 km^2 and 5 min time steps. For time steps when in the radar image more than one storm needs to be nowcasted, responses as per Equation 6.14 are derived for each storm, and then later on summed together.

$$I_{new}(x,y) = \sum_{i=1}^{k} Pr_i I_i(x,y) \qquad (6.14)$$

where:
I_{new} – is the target intensity at a specific lead time at location x and y of the radar image
k   – is the number of similar neighbours selected
I_i – is the intensity at location x and y from the ith neighbour
Pr_i – is the weight for the ith neighbour

Lastly, since the performance of the k-NN is highly dependent on the number of k – neighbours used for the averaging, a prior training is done in order to select the right k-neighbours that yields the best performance (see Section 6.4.2).

6.3 Model Training and Optimization

From the period 2002- 2012 (over the 93 events) a total of 12 428 storms were identified with different durations and properties as describes in Figure 3.5 from Section 3.4.1. For the selection of the best k-number and as well of the most important predictors, a training of the k-NN model is performed on a split sampling validation. For this purpose, the whole dataset was separated randomly between the "past" dataset and the "new" one according to the respective 70 %and 30% ratio.

The past dataset is used for the predictors analysis, which results are discussed in Section 6.4.1, and for the training of the k-neighbours for the deterministic nowcast as shown in Section 6.4.2. The training of the k-neighbours is done with the objective to find the response that is spatially most correlated with the future observed behaviours at different lead times. The k-number with the best correlation performance for most of the lead times is selected as a representative for the deterministic approach.

Once the k and the most important predictors are analysed, for each of the *new* storms, a deterministic and ensemble nowcast is issued at different stages of the storms' life. This is done in order to comprehend better when the k-NN behaves good and how long should the information from storms be saved in order to have a good nowcast. Since the storms are of different variations, five representative stages of storm life are considered here: the first, the second, the third, the 12th (1-hour after initialization) and the 36th (3 hours after initialization) time step of the storm life. The performance is assessed only on the storm scale (no point scale considered) and only spatial criteria is computed. The two criteria used here for the validation of the k-NN approach are:

a) the spatial cross-correlation:

$$\rho(k)_{LT} = \frac{\sum_{i=1}^{n}\left(For(k)_i - \overline{For(k)}\right)\left(Obs_i - \overline{Obs}\right)}{\sqrt{\sum_{i=1}^{n}\left(For(k)_i - \overline{For(k)}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Obs_i - \overline{Obs}\right)^2}} \qquad (6.15)$$

where:
n   – number of grid cells with intensity > 0 mm/5min
Obs_i – the respective observed value at the ith grid cell
$For(k)_i$ – the respective nowcast value of the k response at the ith grid cell
$\overline{For}$ – mean of the nowcasted values at a specific lead time (LT)
$\overline{Obs}$ – mean of the observed values at a specific lead time (LT)

b) the spatial root mean square error:

$$RMSE(k)_{LT}[mm/5\,min] = \sqrt{\frac{\sum\limits_{i=1}^{n}(For(k)_i - Obs_i)^2}{n}}$$

(6.16)

where:
n   – number of grid cells
Obs   – the observed intensity at the ith grid cell
For(k) – the nowcasted intensity from the k response at the ith grid cell

As the nowcast is evaluated at the storm scale, it is possible to calculate the errors associated with the death or growth/decay of the storms. These errors give a better insight on how important these non-linear processes are in the nowcast model, and what is the gained improvement towards the Lagrangian persistence. Figure 6.4 illustrates for an observed storm (top row) how the death and growth/decay are considered for the error calculation. For a given storm, the birth is referred to as the time step when the storm is first recognized by radar (in Figure 6.4 is the time t-5min). The storm then changes itself continuously until t+15min – this is referred to as the growth/decay of the storm. At t+20min the storm stops existing – hence the death of the storm.
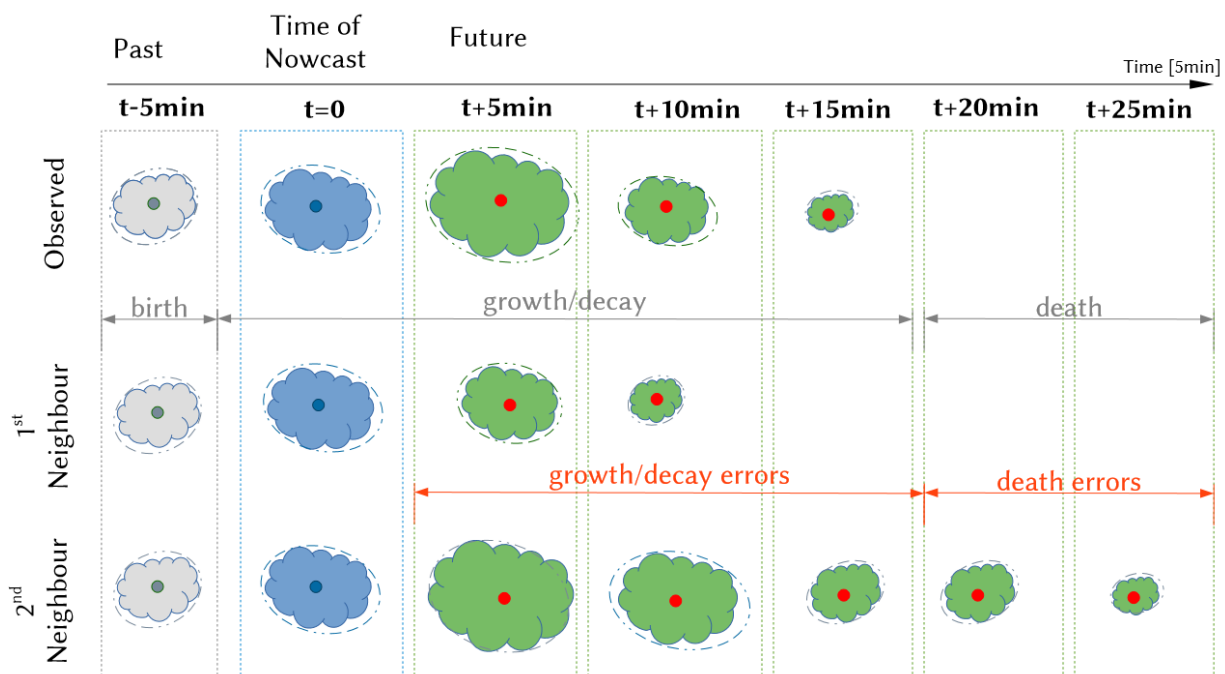
Figure 6.4: *The visualization of the growth/decay and death errors at the storm scale for the k-NN nowcast.*

In this example, a 2-NN nowcast is issued at time t=0 (respectively middle and lower row). The RMSE is then calculated separately for two time steps groups: a) the time steps between the t+5min and the death of the storm (t+15min) – representing the **growth/decay error** and b) the time steps between t+20min (after observed storm death) and t+25min (death of the nowcasted storm) – representing the **death error**. In this case, the death error displayed

by the 1st neighbour is zero, while for the 2nd one is higher than zero. Since a storm has to be recognized in order to be nowcasted, the birth process is not included in the k-NN nowcast and hence the birth errors cannot be calculated.

Lastly, as all the storms from all the events are pooled together and split-sampled for the training and the validation of the k-NN, the consideration of the prevailing mechanism is not possible any-more. Moreover, even though an event has been classified as convective, it may happen that stratiform mechanism are present. Thus, the results are pooled together for all types of storms. For some visualization examples of the k-NN application please refer to the nowcast plots at Appendix G – from Figure G.6 to Figure G.10.

6.4 Results

6.4.1   Predictor Selection

In order to have reliable results for the k-NN application, the most important predictors have to be identified. First the Pearson correlations are calculated between each of the features selected in Table 6.1 and the target variables. To simplify the selection, the correlations computed only for the target variables at lead times +5min, +1h and +3h are shown in Figure 6.5. As indicated by the colour shades, the area target variable seems to be most affected by the area of the instances, followed by intensity and the advection velocities. The x and y component of the local velocities seem to be mainly correlated only to their history. For the mean intensity, the three features of the intensity (mean, maximum and standard deviation) seem highly correlated with the target variable (at least for lead times up to 1 hour). However, one has to keep in mind that some of the features are highly correlated. For instance, as seen in the correlation plot of the features with one another (in Appendix F – Figure F.1), the features describing the present intensity are highly correlated to one another, consequently the intensity target will be highly correlated to all features describing the intensity.

Figure 6.5: *The absolute Pearson Correlation between each of the features (as named in Table 6.1) with the four target values (in this order Area, Intensity, horizontal and vertical velocity) at three different lead times +5min, +1h and +3h.*

Additionally, to the correlation between all the features, a multi-regression (based on all the mentioned features) was fitted to each of the target variables. Since the regression was satisfactory, with r^2 on average higher than 0.85, a linear relationship may be assumed between the predictors and the target variables. Therefore, the relative importance weights were computed for each of the target variables and are illustrated in Figure 6.6. The weights from the RIA analysis match somehow with the correlation values, however some features are enhanced. For instance, the change in the ellipse axis ratio is displaying more importance than in the Figure 6.5. The area target variable is not only dependent on the observed area,

but as well the mean intensity. Note that here, only mean intensity is important and not the other intensity features. Through the orthogonal new predictors, the RIA has removed the effect of the co-linearity which was present in the correlation weights. Moreover, in the intensity target variable, it seems like the maximum and the standard deviation intensity are important for the +5min intensity but not for the longer lead times; there the mean intensity is more important. It is surprising to see, that the velocity target variables are dependent on the axis ration changes and on the average intensity. This was a relationship not displayed by the correlation.

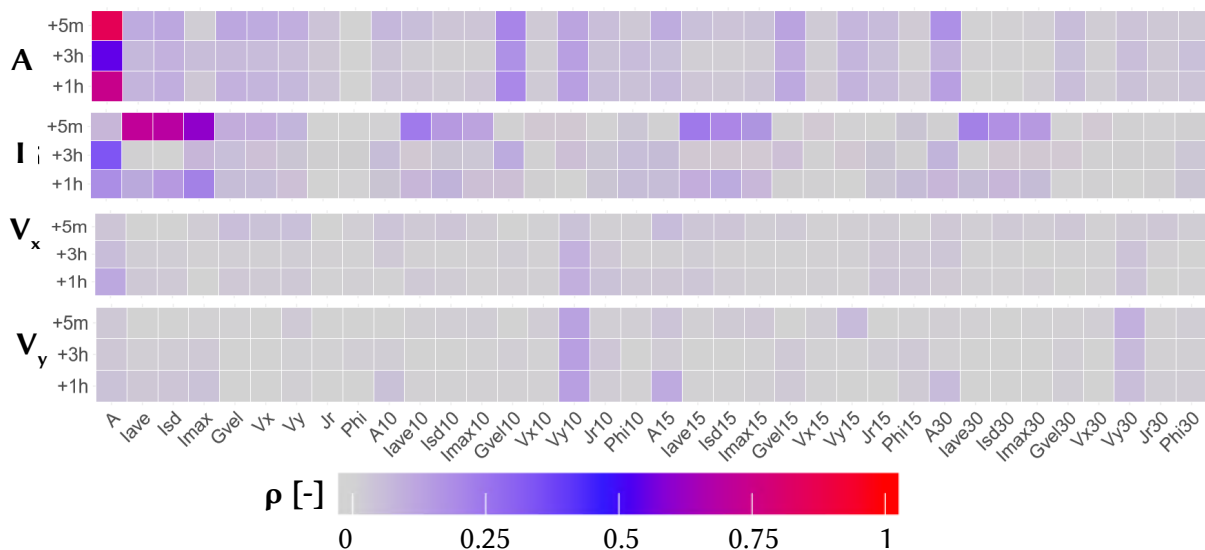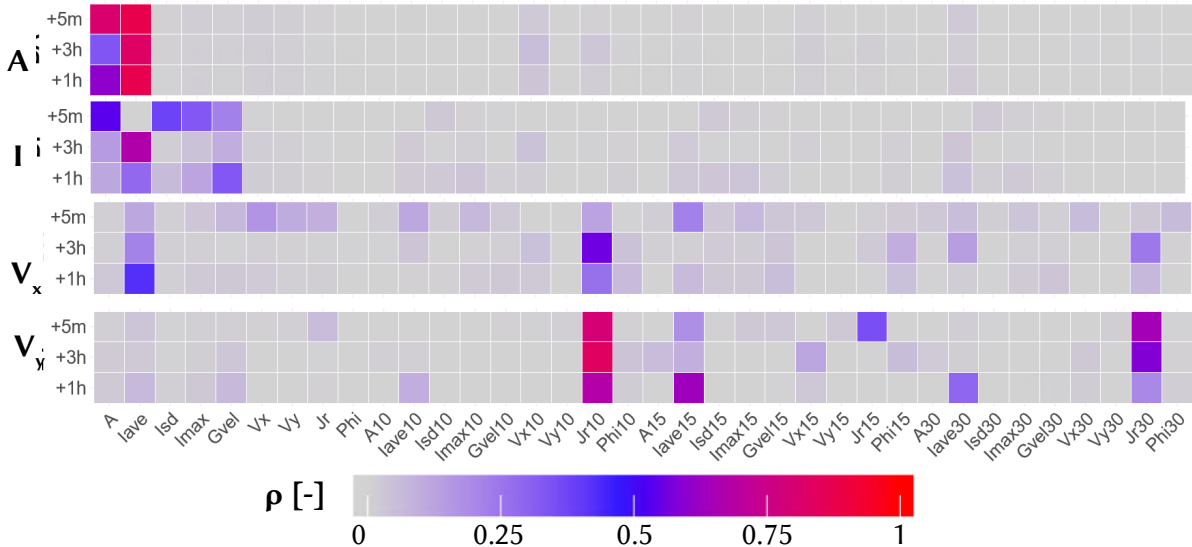**Figure 6.6:** *The RIA weights between each of the features (as named in Table 6.1) with the four target values (in this order Area, Intensity, horizontal and vertical velocity) at three different lead times +5min, +1h and +3h.*

Based on the results of both analyses, the Pearson Correlation and the RIA weights, 12 features were selected as important predictors for the nowcast. These predictors are shown in bold in Table 6.1 and are respectively: A, Iave, Isd, Vg, Vx, Vy, Iave10, Isd10, Vx10, Vy10, J10 and A30. For the given selected predictors, the step wise PIC was computed for each of the target variables in order to get an overview about the weights to be used in the Euclidean Distance. As already mentioned, an important predictor has to be recognized prior to the step wise PIC procedure. Thus the ones with the highest values from RIA weights for each of the target variables were used, respectively for Area the A, for the Intensity the I_{ave}, for Velocity in x,y direction the J_{r10}. Nevertheless, the step wise PIC didn't yield satisfactory results. For the Area and Intensity, where the first predictors showed a strong correlation (see Appendix F- Figure F.1), the analysis yield only these two predictors as the most important ones with a weight of 80%. For the V_x and V_y, it was unable to find any meaningful predictors. Even though the PIC is thought to recognize important predictors especially for the k-NN approach, it appears that it is not able to work well for very large dataset with several predictors. This either shows a shortcoming of the method or that the current predictors are either equally important or equally non-important.

In the end, the 12 selected predictors, normalized according to the observed range, were chosen for the Euclidean Distance computation without any weights applied, to check whether or not a k-NN nowcast based on these predictors is able to provide reliable results in terms of similar storms. The Euclidean Distance, as shown per Equation 6.1, is first used to calculate the similarity between the to-be-nowcasted storm and the past storms and to rank all the storms in an ascending order. Only the 100 most similar storms are being selected for the next step of the k-NN to optimize the cross-correlation.

6.4.2   Training k-Neighbours for Deterministic Nowcast

The cross-correlation results of the deterministic nowcast with different k-neighbours used for different lead times are shown in Figure 6.7. Neighbours up to 50 were investigated to find out which of the k-neighbour yields the best spatial cross-correlation results when issuing the nowcast at different life stages of the storm cell. The first thing noticeable, is that the cross-correlations displayed by nowcasts issued at the first-time step of the storm evolution are quite low when compared to the other time steps. Apparently, the uncertainty at the first time steps is quite high and the k-NN approach is not able to select the closest neighbours. This can be attributed to the lack of the past persistence predictors. As many storm instances have similar present features, a defined number of 100 neighbours is not enough to find the most cross-correlated storm instance in the past.

This, however, is changing when the nowcast is issued at later time steps of the storm evolution. For example, for nowcast issued at the third time step, the spatial cross-correlation exhibited by the k-NN neighbours with various neighbour numbers, reaches values up to 0.90. The highest cross-correlation values are reached between the neighbours 2 to 10. As expected the first neighbour is not the best one, reinforcing thus the theory, that an averaging from k-neighbours should be done, instead of taking the first neighbour. Moreover, it seems that for higher lead times, the responses of neighbour storms are more similar than for the short lead times. This suggest that similar storms might exhibit similar patterns on the long term rather than the short term.

Similar results are also visible for the nowcast issued at the 1st and 3rd hour of the storm life. In contrast to the nowcast issued at the 3rd time step, here the highest cross-correlation values are clearly centred between 2 to 10 neighbours, more precisely 2 and 3 neighbours yielded the best results. Here as well the same spatial cross-correlation pattern is observed, the longer lead times display higher correlation than the short lead times. As the results exhibited here are quite good and they represent mainly storms that have longer duration (thus storm of the meso-scale and stratiform type), one can conclude that the area persistence over the last 30 minutes together with the persistence from the last time step predictors are enough to recognize similar patterns between storms.

Overall, for nowcasts issued at higher stages of storms' life (higher than 1) the k-NN indicates promising results in selecting the closest neighbours. Moreover, the probability

selected for the averaging of the storms seems also adequate as the cross-correlation values reached are quite high. Other probability-weights based on the distance were investigated, however they didn't exhibit so high cross-correlation values. A representative number of neighbours as 3 was selected to average the results for the deterministic k-NN approach. It seems that for all the life stages of the storm, three neighbours will yield reasonable and good results. Nevertheless, for the 1st time step of the storm evolution, the approach of k-NN might not work due to the high uncertainties in neighbour selection.
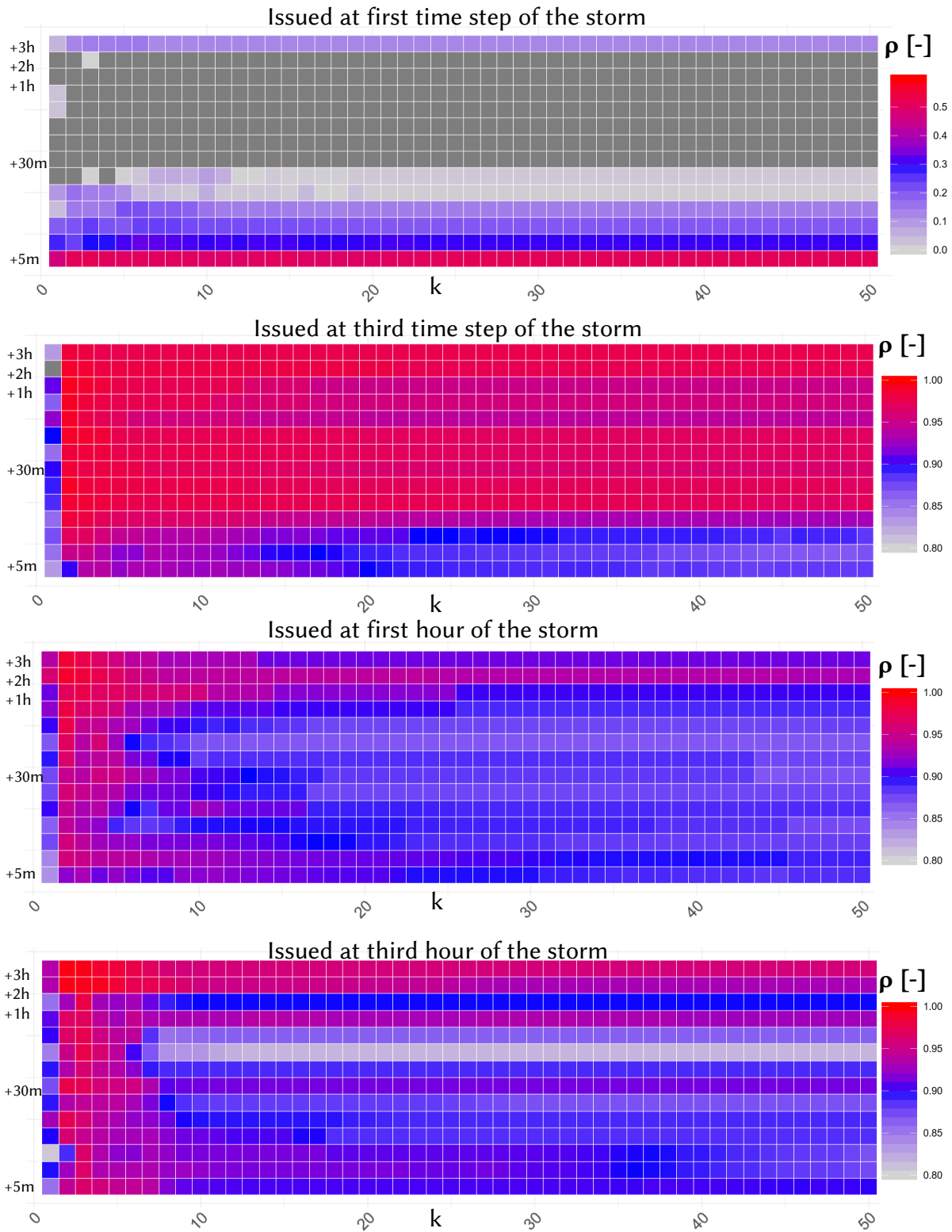
Figure 6.7: *The performance (cross-correlation) of the deterministic k-NN for different neighbours (k) considered, at different lead times and at different life stages of the storm when the nowcast is issued.*

6.4.3 Performance of k3-NN Deterministic Nowcast

The spatial performance criteria (both cross-correlation and average RMSE) of the k3-NN nowcast issued at different time steps of the storm life is illustrated in Figure 6.8. As it is expected from the training of the k-NN, the cross-correlation displayed when the nowcast is issued at the first-time step is quite low. Within the first 10 min the nowcasts loses its predictability skill. This is not the case for the nowcasts issued at the other time steps, where the predictability exhibited is above the limit and doesn't change much with the lead times. This proves once again that the k3-NN is able to find the right neighbours and the respective behaviours can model well the spatial distribution of the new storms' rainfall.
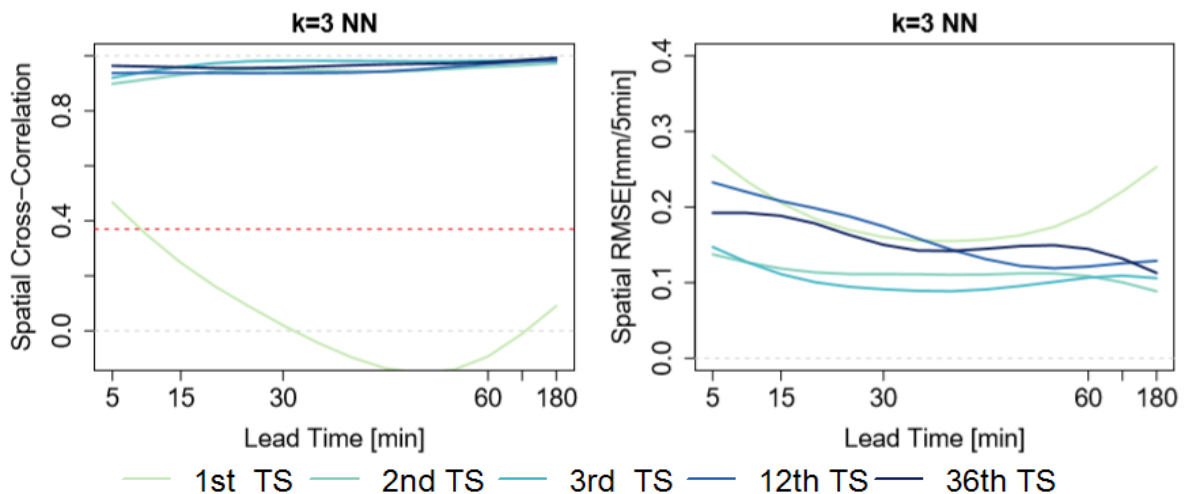

Figure 6.8: *The median spatial performance criteria of the k=3 NN nowcast for different lead times and for nowcast issued at different life time of the storm cells. TS is short for time step.*

Also when consulted with spatial RMSE performance, the first time step nowcast has the highest error, followed up by the 3rd and 1st hour error, and in the end by the 2nd and 3rd time step where the error is the lowest. While at the first time step the RMSE is getting higher for lead times longer than 1 hour, for the 1st and 3rd hour of nowcast the errors are getting smaller. This suggest that at short lead time, the k3-NN is not able to simulate well the growth/decay or the death of the storms. Since many storms are available and grouped together, no matter their type, the response from different types of storms may be averaged together resulting in a high error. Nevertheless, when the nowcast is issued at the high lead times, which means that most of past storm will be decaying and dying, the RMSE is getting smaller.

To get a better view of how the k3-NN can simulate the processes of grow/decay and death, the RMSE of the k-NN and Lagrangian persistence during the storm life (decay errors) and after the storm death (death errors) are illustrated in Figure 6.9. For the 1st time step nowcast, the decay errors are bigger than the death errors, suggesting that at the beginning of the storm life the k3-NN cannot distinguish well similar storms and storms of the same type that can consequently have similar life time. This may happen due to the matching procedure in HyRaTrac, many storms are left unmatched and thus exhibit only one time steps. This can

affect the k3-NN in two ways: a) since the majority of the storms are dying within the first time step, it causes the prediction to stop the life of the storm, thus increasing the decay errors, b) since the data set is so large and the predictors describe only the instant features of the storm, the k3-NN is unable to detect the most similar storms.
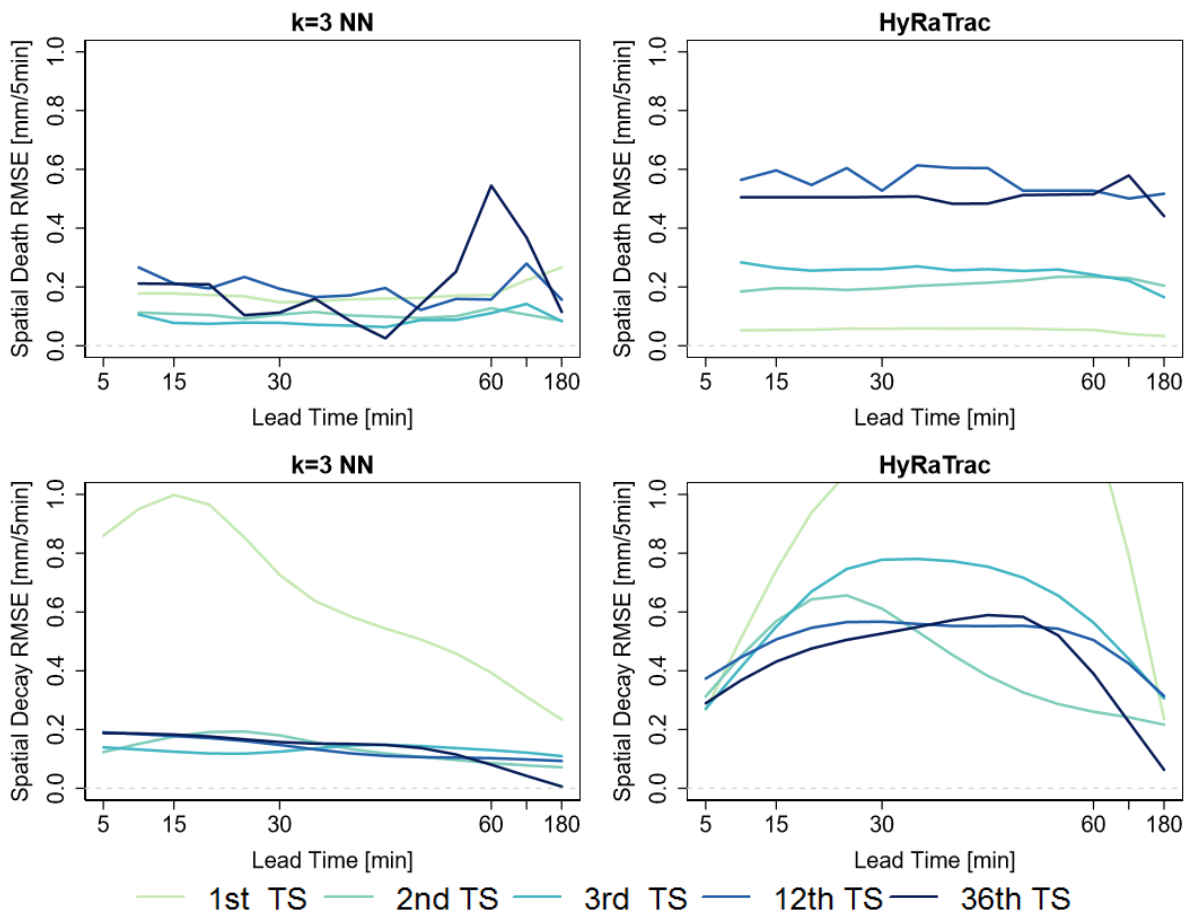

Figure 6.9: *The spatial RMSE observed after and before storm death at both k=3 NN and HyRaTrac nowcasts for nowcasts issued at different stages of the storm cell life. TS is short for time step (5 min).*

For the nowcast issued at the other beginning time steps of the storm (up to the first hour), the k3-NN performs well enough as both decay and death errors remain at the same range for all lead times. This suggests that at these time steps, the k3-NN has enough information to make a suitable identification of the storms and there are enough storms in the past to enable a good prediction of the future instances of the storm at hand. The nowcast at longer time steps (for example the nowcast at the 3rd hour of storm life) suffers mainly from the overestimation of the death time – the storms are simulated to live longer than they actually do. This can as well be attributed to the number of storms available for the k-NN which, since the long events have many time steps from where the neighbours can be sampled, it causes an overestimation of the death time.

6.4.4   Ensemble Nowcast Performance

Apart of the deterministic nowcast, ensemble nowcast were issued at each of the time steps and each lead time. Since the number of ensembles issued was not a fixed number but rather controlled by the condition that each of the ensemble should exhibit a cross-correlation higher than 0.37 with the present instance of the storm at hand. Figure 6.10 illustrates at left - the number of ensemble issued at different time steps where the cross-correlation between the nowcast fields at different lead times and the observed field is higher than the predictability limit. As seen in the case of the k3-NN nowcast, the ensembles issued at the first-time step are useless, as almost none of them have a cross-correlation higher than 0.37. Regarding the other times steps, there is an increase of useful ensembles (that exhibit cross-correlation higher than 0.37) with the lead times. Especially for the nowcast issued at the 3rd time step of the storm life, the increase is rather steep, reaching the highest number of useful ensemble members.
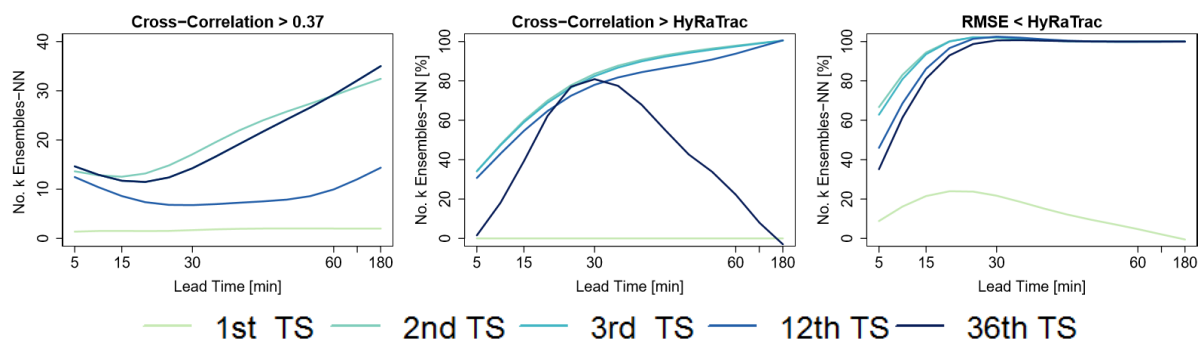

Figure 6.10: *Number of ensembles that exhibit a cross-correlation higher than the predictability limit (left), percent of ensembles that exhibit higher spatial cross-correlation and RMSE than the HyRaTrac for different lead times and for nowcasts issued at different time step of the storm life. TS is short for time step (5 min).*

To understand how many of the ensembles add value to the nowcast, Figure 6.10 middle indicates the percent of the ensemble members that have on average a higher predictability skill than the Lagrangian persistence. As expected no ensemble members are better than HyRaTrac for nowcasts issued at the first-time step of the storm life. Nevertheless, with increasing time step and lead time the percent of members which are better than HyRaTrac is increasing. The peak is mainly for the early life stage of the cells and for the long lead time. There almost 100% of the issued ensembles are better than the Lagrangian persistence. This is also supported by the spatial RMSE performance. On general the fields nowcasted by k-NN, have smaller RMSE than the ones issued by HyRaTrac. Especially for lead times longer than 30 min, for the nowcasts issued from the 2nd time step of the storm life, all of the members issued have smaller errors than the nowcast issued by HyRaTrac.

Overall both the deterministic and the ensemble approach of the k-NN method increase the spatial predictability of the storms for all of the lead times. The deterministic approach can capture the death and growth/decay rates on average 50% better than the Lagrangian persistence. For this reason, the predictability limit is extended up to three hours for all of the events. Ensemble nowcasts have also been shown to improve the predictability limit for most

of lead times and several members are available to provide a wide range of rainfall nowcasts. Nevertheless, even though the k-NN approach improves considerably the nowcast at the storm scale, it is yet to be seen (Chapter 8) how well can the k-NN members and deterministic approach capture the temporal predictability and performance of the rainfall at the local scale.

6.4.5   Improvement of Rainfall Predictability

Figure 6.11 illustrates the spatial performance of the k3-NN when compared with the Lagrangian persistence of the HyRaTrac and the added value that it brings to the nowcast. It is clear that the k3-NN improves considerably the performance of the nowcast particularly at the longer lead times and for storm time steps longer than 10 min. As it is visible from the graph, the nowcast of the first-time step of storm life, is as bad as or even worse nowcasted than HyRaTrac. Figure 6.12 illustrates the improvement in the death and decay/growth rate when using the k3-NN approach instead of the Lagrangian persistence. The k3-NN can capture the processes better than the Lagrangian persistence, particularly the death at the short at medium range lead times. Nevertheless, here as well there is no added value to the nowcast for the nowcast issued at the first time step of the storm life. Thus, it is not advised to issue k3-NN nowcasts at the first time step of the storm.

Figure 6.11: *The improvement of the spatial cross-correlation and RMSE displayed by the k3-NN deterministic nowcast when compared with the Lagrangian persistence implemented in HyRaTrac. The results are shown for nowcast issued at different time steps of the storms life. The green area indicates that the k3-NN performs better than the persistence, and the red area indicates the vice-versa. TS is short for time step (5 min).*
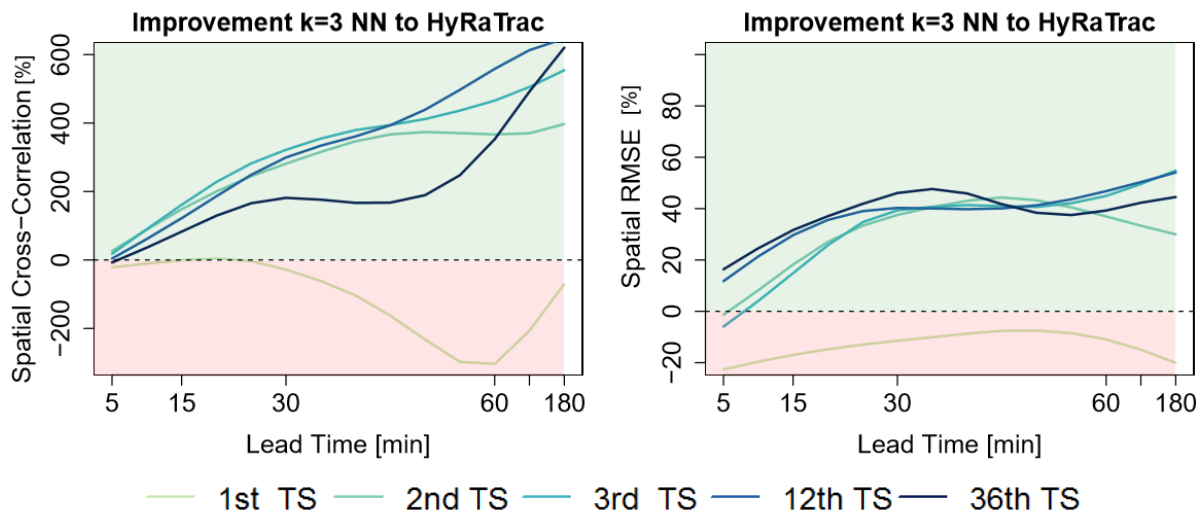
Figure 6.12: *The improvement of the spatial death and growth/decay RMSE displayed by the k3-NN deterministic nowcast when compared with the Lagrangian persistence implemented in HyRaTrac. The results are shown for nowcast issued at different time steps of the storms life. The green area indicates that the k3-NN performs better than the persistence, and the red area indicates the vice-versa. TS is short for time step (5 min).*
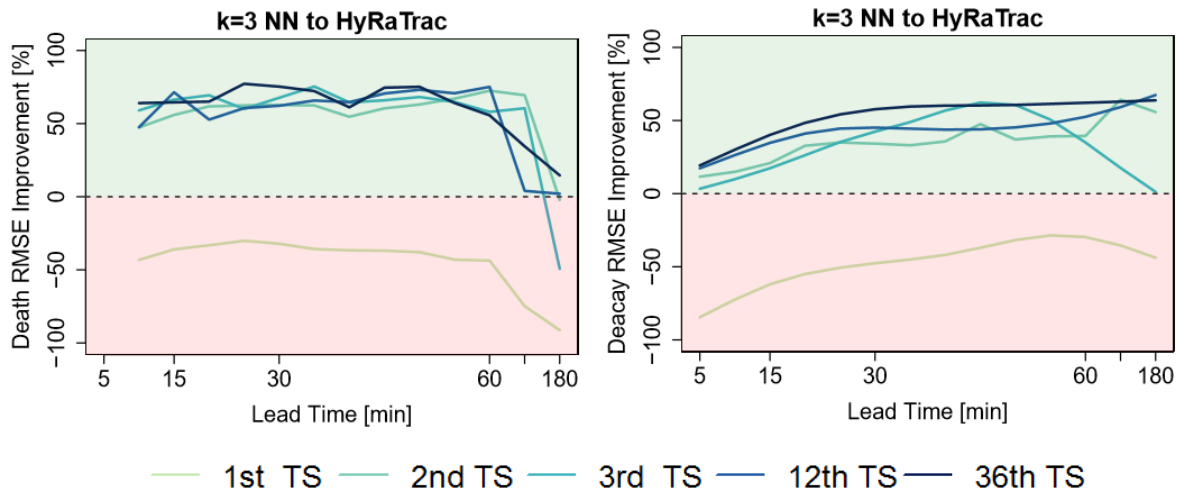
7.   CNN FIELD ORIENTED NOWCAST

As already seen in Chapter 4 and 5, the current field-motion-based nowcast employs only the Lagrangian persistence, which underestimates the intermittent nature of the rainfall process. Even though such method, depending on the storm type, can yield reasonable results up to 1-hour lead time, after this nowcast horizon, the nowcasts are useless. This is mainly related to the inability of the Lagrangian persistence to model several processes like the initialization of the rainfall, the end of rainfall, the rainfall continuous motion at the boundaries of the radar, and lastly the growth/decay rate of the rainfall. At the point scale, (temporal performance) one may expect the errors to be smaller than the area scale (spatial performance) due to movement of storms. Nevertheless, results from Chapter 4 and 5 indicate that the spatial performance is better than the temporal one, suggesting first that the point scale changes are bigger than the area scale, and second that improvements have to be done to capture the intermittent nature of the rainfall process at the point scale. One way to learn how the rainfall changes its' structure can be by exploring the past dataset and investigate complex functions that can model such behaviour. Convolution Neural Network (CNN) are quite good in learning complex relationships between variables and are built to take advantage of the 2D structure of the images, thus these questions rise: Can the CNN learn how the rainfall changes itself in space and time? If yes, can it provide more information than the Lagrangian persistence in terms of capturing better the initialization, growth/decay, death and the continuation of the rainfall at the radar boundaries? If yes, until what limit can the CNN bring an additional value to the nowcast? Lastly, does the CNN improve only the spatial performance or the temporal one too?

These are the main questions that are treated in this chapter. Here, a CNN is applied at the merged radar data and trained separately for each lead time in order to assess the potential of deep learning algorithms for rainfall nowcast. Section 7.1 gives a brief introduction to the basics of neural network theory, particularly focusing on the structure of CNN, followed up by the explanation on the model setting and implementation in this study in Section 7.2. The results of such network are given in Section 7.3. where first the training and validating performance are illustrated (Section 7.3.1 and 7.3.2), followed up by the predictability exhibited by storms and the improvement achieved towards the Lagrangian persistence (Section 7.3.3).

7.1 CNN Basic Theory

7.1.1   Basics of Neural Network Theory

As indicated in Section 2.3.1 of the Literature Review (Figure 2.12), an Artificial Neural Network (ANN) is composed of one input layer, one or several hidden layers and the output layer. Typically, the transformation from one layer to the other is done by transforming the input of each layer as in the following Equation 7.1:

$$Output = \int_{act} \left\{ b + \sum_{i=1}^{N} w_i \cdot Input_i \right\} \qquad (7.1)$$

where:

N – number of inputs
w_i – weight for each input
b – a bias added to the input
$\int_{act}$ – the activation function

The activation functions are quite important in the theory of neural networks, as they add the non-linearity into the model, and it's due to them that the ANNs are particularly good at modelling complex relationships. The typical used activation functions in ANN are illustrated in Figure 7.1. At the moment the Leaky Rectified Linear Unit (Leaky ReLU) and the Exponential Linear Unit (ELU) are the most desired ones, as they can deal with negative inputs (unlike the ReLU), are computationally fast, and have no problem with saturation[12] when training the network. It is for this reason that these two activation functions are used in the CNN employed for the radar based nowcast.
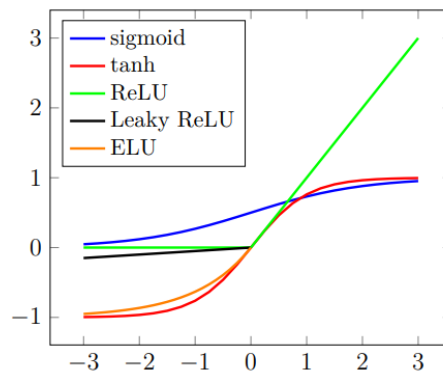

Figure 7.1: *Typically used activation functions in ANN to include non-linearity in the model, in x-axis is the normalized value of the input, and in y-axis the corresponding activation function value. Retrieved from* [Gerchow, 2018]*.*

The training (also referred to as learning) of the specific weights and b – coefficient can be done by supervised (the output is available) or unsupervised (output is not available) learning. The unsupervised learning needs some prior assumptions in order to train the data, and it is commonly used in pattern detection or stereo matching for depth estimation. The supervised learning has the advantage of being properly trained and validated, since the output is known. For this study, as the nowcast will be issued in hind-cast mode (the events are already observed), the supervised learning will be used.

---

12   The saturation is usually caused at the tails of the activation function, as at values very close to the tails (for example in the sigmoid function) the gradient becomes very close to zero, delaying or stopping the learning process of the ANN.

In supervised learning, one can try several weights and choose the set which gives the lowest difference between the modelled and the observed output that is commonly defined by a loss function (also here referred as performance criteria). Typically, the weights are learned by a backward propagation, which focuses in minimizing the gradient of the loss function with respect to the output. Normally a gradient descent method[13] is employed for this purpose. The gradients from single-layer networks can be computed easily, nevertheless when several layers are present (as in deep network applications) the automatic differentiation procedure is much faster. Due to its feasibility and fast computation rates, the so called Reverse Automatic Differentiation is mainly used in deep neural networks (the reader is directed to Zhang et al., (2020) for more information about the method).

Moreover, when looking for an optimum with the gradient descent method, the whole input-output training dataset is used, which again causes long training times. Therefore, Bottou (2010) proposed a stochastic approach rather than a deterministic one, where several small randomly picked training samples (x_i, y_i) are used to estimate the gradients (*J*) and hence the best weights (θ), as shown by Equation 7.2. The learning rate ($\boldsymbol{\alpha}$) is set to a constant number, which depending on the gradient rate (the converge $\boldsymbol{\Delta_\theta}$) is halved. The closer the weight set is to the optimum, the smaller will the $\boldsymbol{\alpha}$ be.

$$\theta = \theta - \alpha \, \Delta_\theta J\left(\theta, x^i, y^i\right)$$

where:
x^i, y^i – the pair from the training data set
θ – the model weights set
$J(\theta)$ – the respective loss function
α – the learning rate
$\Delta(\theta)$ - the gradient rate

(7.2)

7.1.2 The Convolutional Operation

The convolution is a mathematical operator between two matrices (input data and filter/kernel) that produces another matrix whose shape and values are determined by the modifications that the kernel causes on the input data. The following Figure 7.2 illustrates an example of such operation. The input is a 2D matrix of shape 3x3 and the filter is a 2D matrix of shape 2x2. In the case (a), the filter moves over the input by one cell at a time (first a horizontal sliding, then one vertical), which commonly is referred to as a stride[14] of 1. For each location of the filter over the input matrix, the matrices are multiplied and summed to produce the entry value for the output matrix (as indicated in the Figure 7.2-a). For this case the output matrix is reduced in dimensions due to the size of the filter. In case one would like to avoid such reduction in dimensions, then padding is introduced. The padding as in Figure 7.2-b, adds zero rows and columns at the input matrix borders (here shown in grey), in order to have similar size or bigger size (as in the indicated case) output matrices. Figure 7.2-c illustrates another example with padding, but in this case with a stride of 2. Here the output matrix is always downsized due to the presence of a stride higher than 1.

---

13 Gradient descent finds the best value of a loss function (that is either a maximum or a minimum) by following the change in its gradient with the purpose to find the location where the gradient is zero (hence the location for the maxima or minima).

14 A stride is the number of pixels that the kernel is being shifted after each iteration over the input matrix.
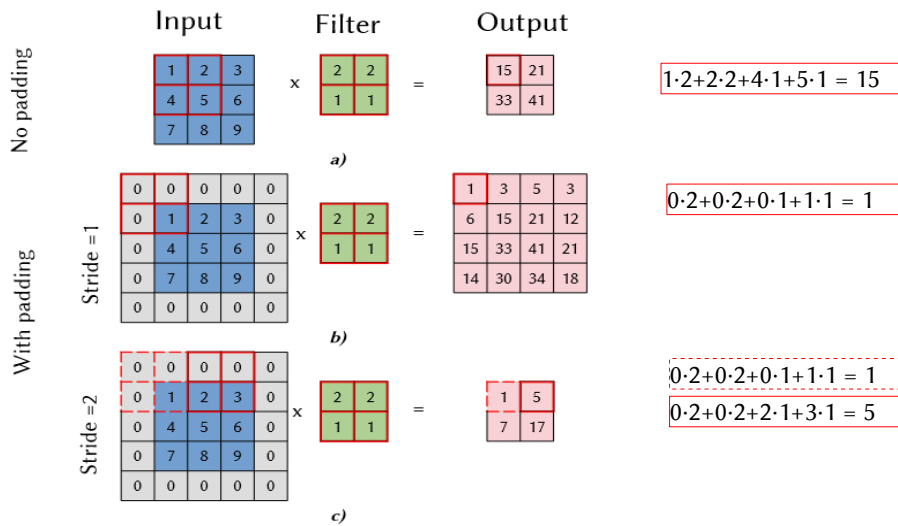
Figure 7.2: *The convolution of a filter (green) over an input 2D matrix (blue) resulting in a 2D output matrix (red). The convolution is done in (a) with no padding and stride 1, in (b) with padding and stride 1, and in (c) with padding and stride 2.*

7.1.3 Convolutional Neural Networks

Convolutional Neural Networks can take advantage of the convolutional operators to learn and predict relationship from 2D images. Related to the Equation 7.1, in the CNN application, the input layer can be represented by one (or multiple) 2D matrix, the sum unit of Equation 7.1 can be done by a filter with a fixed dimension (k), and then the output can be obtained by using an activation function to transform the convoluted output. The result of such operation builds up a so-called "Feature Map" as indicated in red in Figure 7.3. When more than on image (matrix) is used as an input, the filter convolves to each of them and the outputs are cross-correlated together. The cross-correlation[15] as an operator here, means that the corresponding cells from the convoluted outputs are summed together. By applying the filter at some input images and adding them up, the CNN keeps the locality of the features, such that features in a local region are independent and not affected by the features at greater distances.
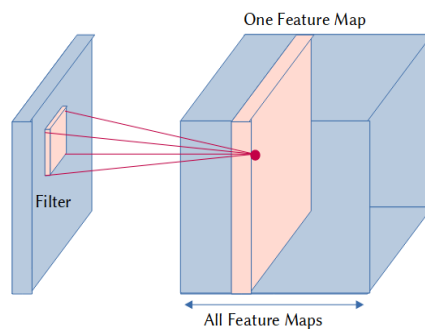

Figure 7.3: *Convolution kxk filters c times over an input image resulting in a 3D matrix. Retrieved from [Gerchow, 2018].*

---

15 The cross-correlation as an operator by the CNN should not be confused with the cross-correlation performance criteria when validating the nowcast results.

The number of input matrices in CNN is usually denoted by the term channels and it is considered differently than the x and y dimensions (that is why CNN is referred to as a 2D and not 3D matrix). The channels represent instances of a specific variable continuous in x and y. The features dictated by the channels (for instance in radar images the feature is the rain intensity) display behaviours that are independent on their location (x and y). This is particularly important when training the CNN, as responses from same objects that are situated in different regions of the input matrix should be the same (also referred to as translation invariance).

When multiple channels are present, it may be desired to have the same number of channels (or more) in the output product. This can be achieved by having multiple filters of the same size (k) but different weights, and produce one feature map for each filter convolving at the input data. The product of all output channels is indicated in Figure 7.3 as the blue volume. The output size is dependent on the input resolution and channels, as well as on the number and size of the filters applied. Having multiple output channels (more than the input channels) is essential when employing a deep CNN because it enables a trade off between the spatial resolution and more channel depth. Depending on the number of hidden layers at the CNN, it can be computationally very expensive to train the weights of full-sized images. Thus, decreasing the resolution and increasing the number of channels allows for fast computations and meaningful feature analysis. That is why in CNN architectures, it is quite popular to decrease the resolution and increase the channel size with deeper layers as illustrated on Figure 7.4. This shown example of the EV-Flownet, will be used as a reference to explain the other main parts of the deep CNN, which served as a basis for the CNN employed in this thesis.
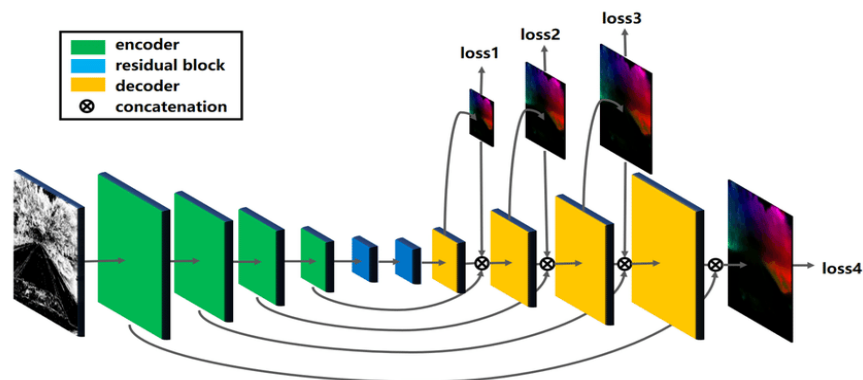

Figure 7.4: *The overall structure of a deep CNN (the EV-Flownet architecture) consist of input matrices decreasing in size [Zhu et al., 2018].*

As the resolution decreases with the hidden layers, it becomes easier to learn global (large scale) features, however in the case of nowcasting, it is desired to have the output at the same resolution as in the input data. For this reason, the CNN can be built on two components: the encoder and the decoder as illustrated in Figure 7.4. Whilst the encoder is based on the convolution operation and reduces the image size to a certain point (referred here to upscaling), the decoder does the opposite; it applies a kernel to the input image in order to increase its resolution (referred here as down-scaling). This is done by transposed convolution

(also referred here as deconvolution[16]) with same kernel size as the convolution. To explain shortly a specific convolution and the respective transposed convolution, three cases are illustrated in Figure 7.5. Given an input matrix of size i, kernel of size k, padding of size p and stride size s, the output with size o is created by the convolution operation. To come back at the same size of matrix input i, when the output size o is given ($i_T=0$), the transposed convolution is actually a convolution operation with $k_T=k$, $p_T=k-p-1$, $s_T=1$, and in case the $s>1$, $s-1$ empty rows and columns are added inside the i_T matrix (as illustrated in case b and c of Figure 7.5). Note that the kernel size and the locations of the border cells are kept the same from the convolution (encoding) to the transposed one (decoding). However, although the resolution is restored, the transposed convolution doesn't restore the full structure of the feature.

Since the decoder is restoring the resolution, but not the precise structure of the feature, in CNN decoder, additional solutions need to be integrated to keep the structure of the feature when increasing the resolution. This is done by introducing concatenation of layers as illustrated in Figure 7.4. The convoluted planes in the encoder, for each specific layer, are concatenated (added) to the planes that need to be decoded. Thus, when the next layer is decoded, the predicted information from the previous layer is present together with the feature structure from the encoder. The presence of the concatenation removes the influence from the background loads and yields to better predictions [Wu and Wang, 2019].

Convolution　　　　　　　　　　**Transposed Convolution**

$i = 4$, $k = 3$, $s = 1$, $p=0$, $o = 2$　　　$i^T=o+(o-1)(s-1)= 2$, $k^T = 3$, $s = 1$, $p^T=k-p-1=2$, $o^T=i=4$

$i = 5$, $k = 3$, $s = 2$, $p=0$, $o = 2$　　　$i^T=o+(o-1)(s-1)= 3$, $k^T = 3$, $s = 1$, $p^T=k-p-1=2$, $o^T=i=5$

$i = 5$, $k = 3$, $s = 2$, $p=1$, $o = 3$　　　$i^T=o+(o-1)(s-1)= 5$, $k^T = 3$, $s = 1$, $p^T=k-p-1=1$, $o^T=i=5$
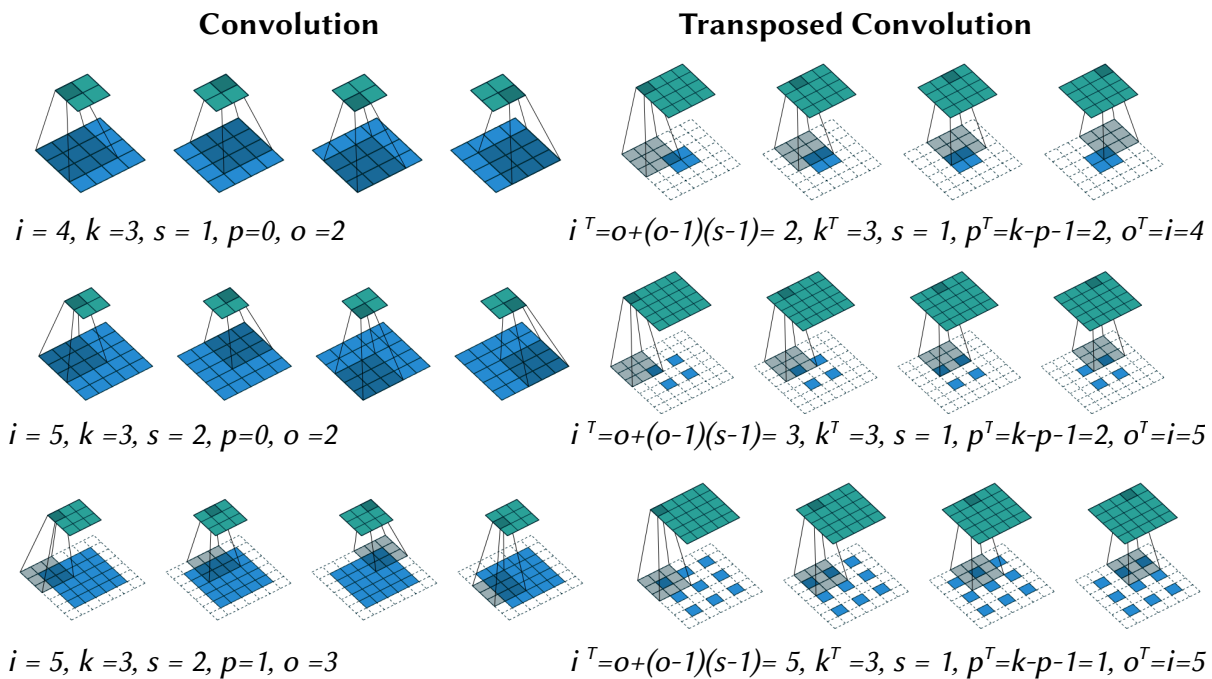
Figure 7.5: For each convolution operator, there is a transposed operator which uses the same kernel to store the same resolution as in the convolution input. Here three cases are shown: - left the convolution, and - right the respective transpose convolution. Blue indicates the input matrices, green the output matrix, grey the filter and the dashed blue lines the padding effect for each operation [Dumoulin and Visin, 2016].

16　The term deconvolution is not exactly the transposed convolution, however for simplicity for encoder-decoder devices, we use here the pair of operations convolution-deconvolution.

As shown in Figure 7.4, the encoder and the decoder components are connected together by residual blocks which are responsible for the deep learning. Since the image is shrank down by the encoder, only the global features are passed to the deep layers to be learned. This not only saves time and makes the learning more efficient, but is in accordance with the finding from Grecu and Krajewski (2000), that the large scale patterns are easier to predict and model than the small scale precipitation. Theoretically, the more layers are present in these blocks, the better will the learning be. Nevertheless, in practice this is seldom the case, as the training error starts to increase with deeper layers. To overcome this problem, He et al., (2015) proposed to learn the residuals rather than the feature itself. Moreover, instead of fitting the results of each layer to a specific objective, they found that the deep network efficiency and accuracy is increasing considerably when introducing the so called "shortcut connections", as indicated in Figure 7.6. The residuals are optimized only at the shortcut connections (every 2 layer), and the network is trained as usually with the end-to-end Stochastic Gradient Method with back-propagation.
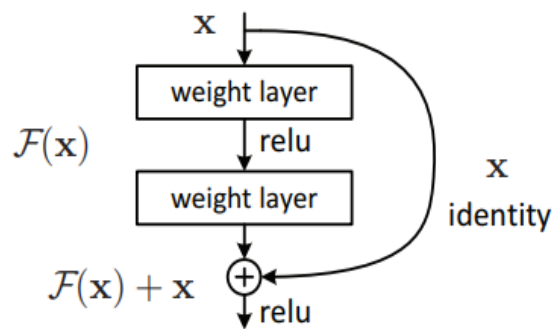

Figure 7.6: The building block from the residual learning (from He et al., 2015); where the feature or identity is x, the Residual is F(x), one connection is skipped between the two layers. The residual is optimized based on the feature (x) and the x + F(x) at the second connector [Gerchow, 2018].

Typically, in supervised learning, the performance criteria (the loss) is calculated at the end of the network comparing the final output with the reference observed image. However, in CNN, since the image is down-scaled by the decoder at different scales, it would make sense to have several locations within the decoder where the performance criteria is calculated. By having several loss calculations as illustrated in Figure 7.4, we make sure that the decoder is working properly at different scales.

Such implementation of CNN as indicates so far and illustrated in Figure 7.4 have shown good learning skills and fast computation for image recognition and depth estimation. Since the computations are using the GPU instead of the CPU, they are quite fast in network training. Due to the results displayed at these studies, the CNN is potentially an adequate tool to nowcast rainfall based on the radar data mainly for the three following reasons:

i) the nowcast is first done on the global scale and then is down-scaled to a higher resolution. This architecture is in accordance with the literature that the larger-scale of the rainfall models the smaller-scale.

ii) the input data can have different channels and this enables the use of several past observed radar images. As already seen by Lagrangian persistence implemented in LK method, the use of past radar images improves the nowcast. Moreover, due to the translation invariance and the presence of several channels, similar features will behave the same regardless on their location.

iii) lastly, the loss function is computed at different scales, assuring a good nowcast first at the global scales and then to the point one.

7.2 CNN Training and Application

The CNN employed in this study follows the structure of the Flownet, and the whole model components, input and output data, are illustrated in Figure 7.8. To make a fair comparison with the Lucas-Kanade method, 3-time steps of radar information (the last 15 minutes observed) are used as an input to the network. The encoder consists of 7 convolution operations with different kernel sizes (7x7, 5x5 and 3x3), which reduce the resolution to 2x2 km^2 and 512 channels. The decoder on the other side, uses as well 7 deconvolution operations, but the kernel is always kept the same as 3x3. The concatenation of the encoding planes to the decoding ones is indicated by the black arrows. The loss function (performance criteria) as per Equation 7.3 is computed at the end output and at three positions (as indicated by purple arrows) inside the decoder chain.

$$RMSE_{train} = \frac{\sum_{i=1}^{N} \sqrt{(For_i - Obs_i)^2}}{N} \quad \text{where:}$$

For_i – the forecasted image
Obs_i – the observed image
N – number of pixels

(7.3)

As the CNN works better on normalized data, the rainfall field is normalized with a maximum value of 80 mm/5min. Moreover, as the CNN cannot handle missing values nor circular matrices (in the case of the radar data), the radar image is cropped to the range illustrated in Figure 7.9. Thus, the training and the validation of the model is done only in this region. It has been observed that the training of the CNN is sensitive to the input data, especially to the presence of the zero. If many time steps with a considerable amount of zero are present, then the model will learn to produce mostly zeros. Thus, it was necessary to include a global intensity threshold (as in the case of HyRaTrac and LK method). Here only time steps where at least 25 radar grids have intensities higher than 20 dBz are used for the training of the CNN.

Within the training of the CNN a cross-validation is performed, to ensure that the optimized data provide a decreasing loss also in another "unknown" dataset. For the training, all the 93 events from 2000-2012 were considered, 80 random train-validation initializations were performed one after the other. Here these samples are referred to as epochs. At each epoch random input-output data were used to train and validate the weights, where 70% and 30% of the data correspond respectively for the training and validation dataset. The weights

learned during an epoch, are then used and further optimized by the consequent one. To avoid the over-fitting of the weights to the data sample, the learning rate (as describe in Equation 7.2) is reduced with each epoch, until it becomes zero and the weights ca not be further trained. The training was done for each of those lead times +5, +10, +15, +20, +25, +30, +60, +120 and +180 minutes. As the CNN training is time consuming, only these lead times were considered. For some visualization examples of the CNN application, see Appendix G – from Figure G.1 to G.5.
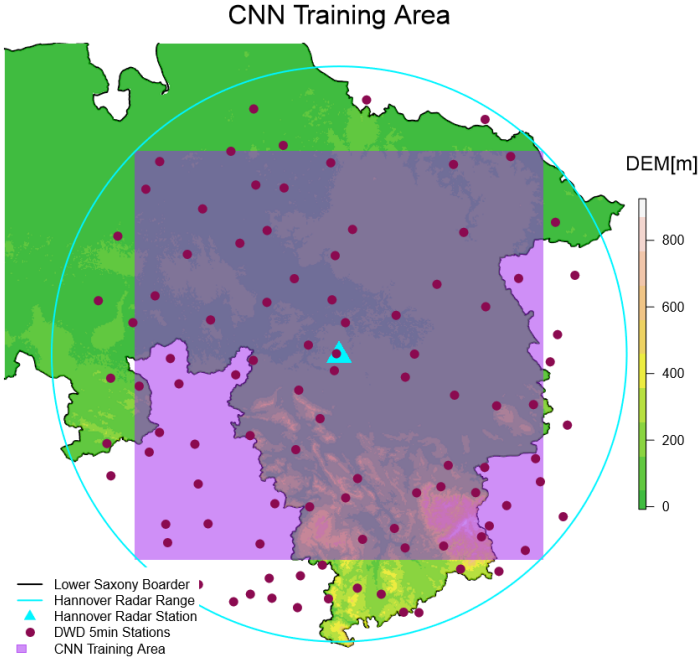

Figure 7.7: *The area inside the radar range used for the training of the CNN nowcast.*
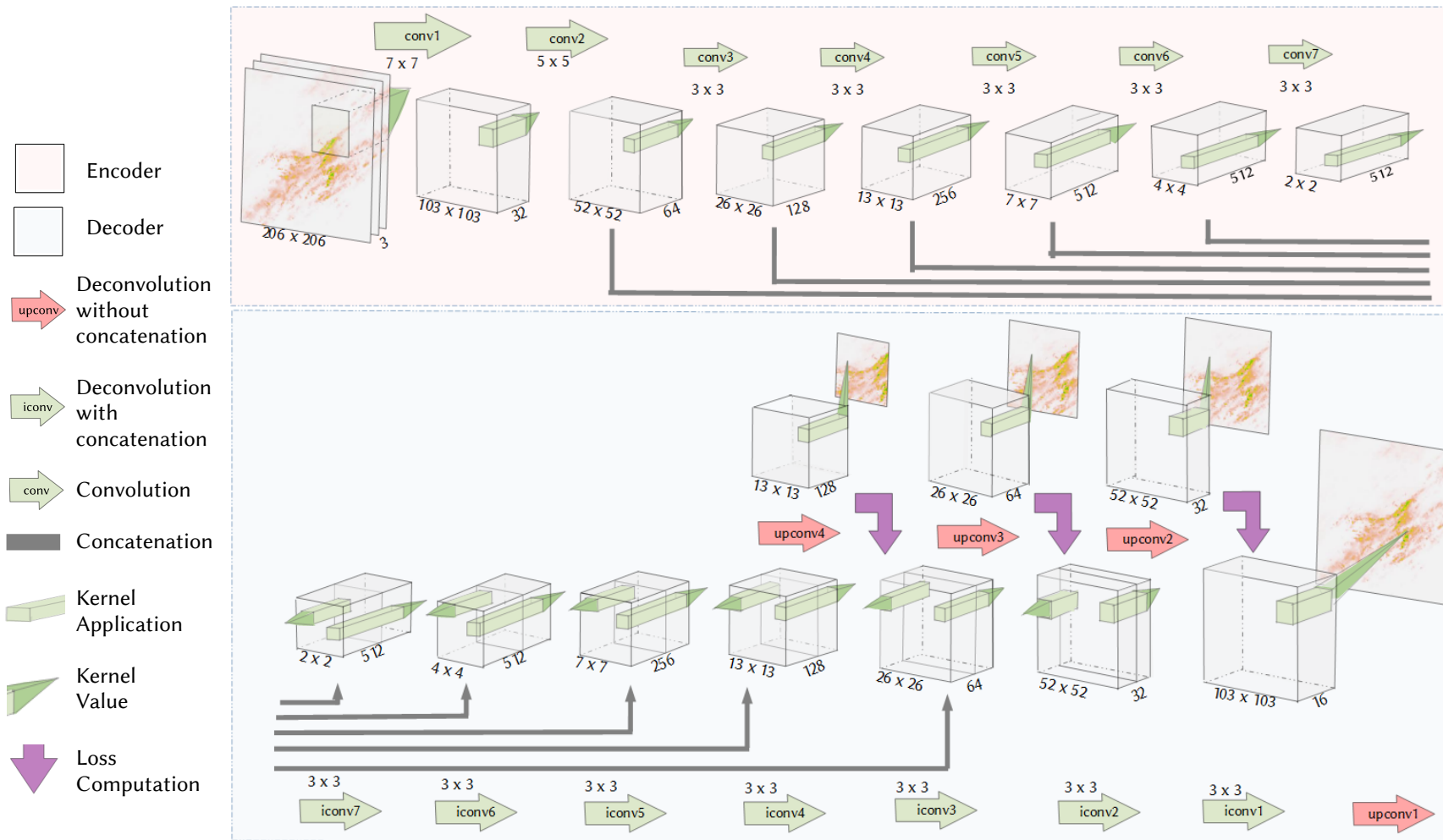
Figure 7.8: The structure of the deep convolution neural network used in this study for the nowcast of the rainfall at different lead times. The input consists of three radar time steps (the last 15 min observed rainfall) and the output is a radar image at lead time t+LT. The red area indicates the encoding component of the CNN, while the blue the decoding part.

7.3 Results

7.3.1   Training Performance

The results of the computed training loss at the final output of the CNN for each epoch are illustrated in Figure 7.9 and the respective validation losses are given in Figure 7.10. The first thing noticeable is that the LT +05min nowcast is converging quite fast to a point below 0.1. However, there is a drastic decrease of the +05minutes loss, indicating that even though the values are converging, most probably the weights chosen will not be very representative. This is slightly the case also for the LT+10,+15,+20 and +25 minutes, although the training curves have smoother shapes. Surprisingly for the high lead times +60, +120 and +180min, the training curve is gradually receding, indicating that the CNN is learning well with each iteration and decreasing properly the loss value. Nevertheless, the learning doesn't fully converge and the number of epochs considered should be higher.
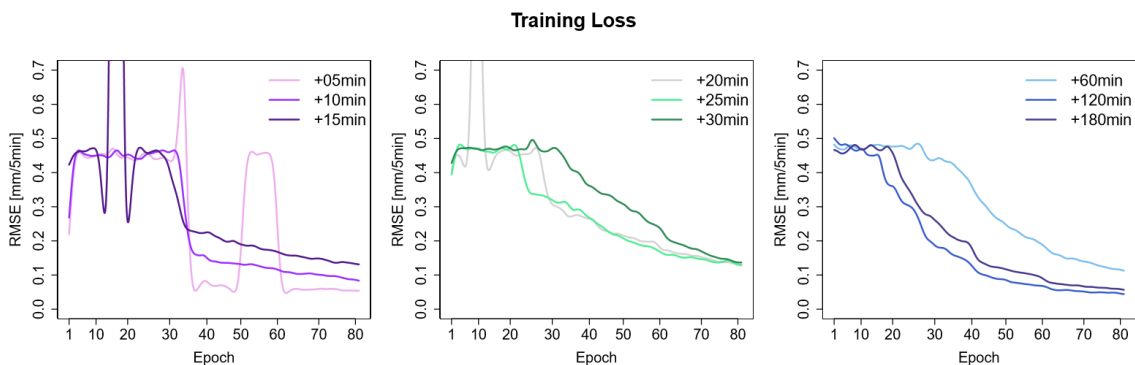

Figure 7.9: *The training loss function (RMSE) for the lead times +5, +10, +15 min (left), +20, +25, +30 min (middle) and +60, +120, +180 min (right).*
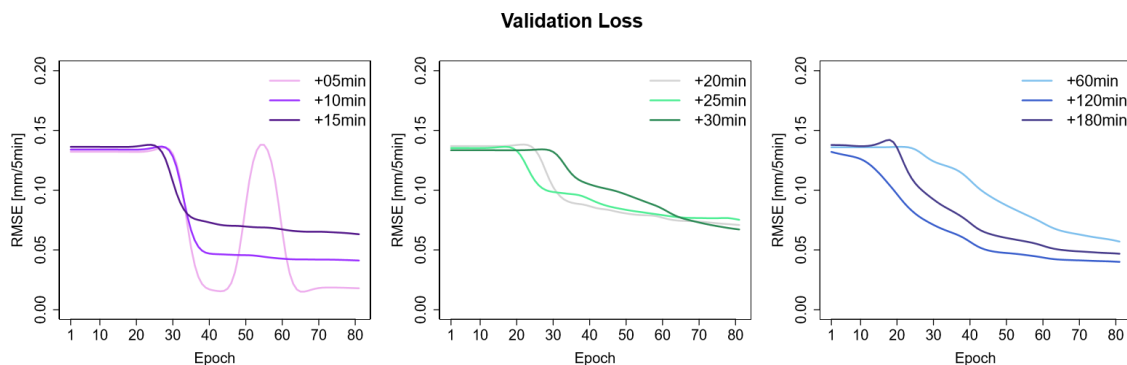

Figure 7.10: *The validation loss function (RMSE) for the lead times +5, +10, +15 min (left), +20, +25, +30 min (middle) and +60, +120, +180 min (right).*

The validation loss follows the same pattern as the one from the training one. Overall the simulated loss is quite low, however it looks like for some lead times the performance is better. One would expect the short lead times to have lower RMSE, but this is not the case when comparing for assistance +120min with +20 min. It appears that the CNN can learn better the longer lead times, as most probably the large-scale prediction plays an important role here.

Instead in short lead times, the large-scale volumes are similar, leaving the accuracy of the CNN to be dependent mainly on the weights used for the down-scaling (deconvolution). Even though the CNN would yield better results for longer epochs, the results are acceptable and indicate that the CNN can perform well for the nowcast.

7.3.2 Rainfall Nowcast Performance

The trained weights for each of the lead times, were run once again for each of the 93 events. Note that the ultimate validation of the CNN nowcast is done in Chapter 8 with the 17 events of period 2013-2018 ,which are completely unknown to the training of the CNN.

Nevertheless the aim here is, to explore the full potential and dynamics of the CNN for nowcast application; meaning that the performance over these 93 events exhibits the best performance of the CNN (as the weights are trained on these events). Moreover, since the training is done on random samples of the 93 events, it can be that some events are better nowcasted than others, suggesting that some processes are better simulated by the CNN.
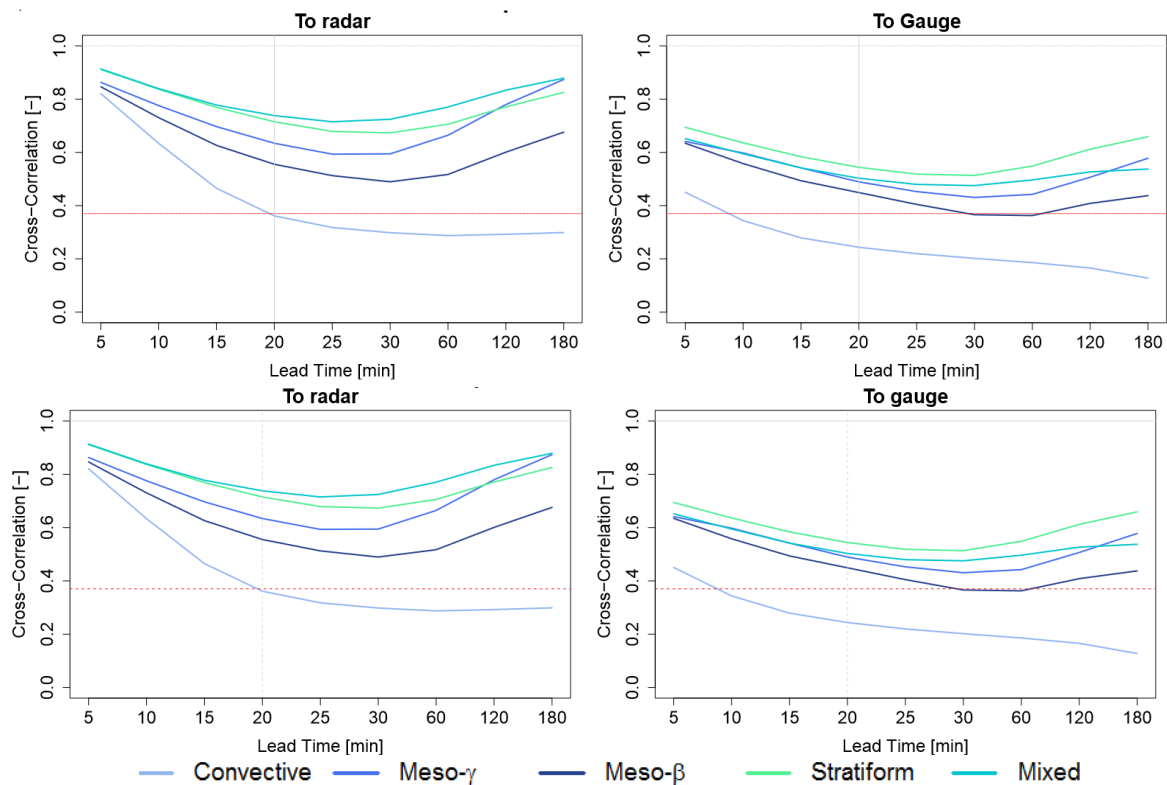

Figure 7.11: *The event median cross correlation for each event type as nowcasted by CNN for different lead times, calculated temporally (upper row) and spatially (lower row). The red dashed line indicates the correlation value of 0.37 for the predictability limit, while the grey dashed line is a reference to the 20 min model predictability limit as suggested by the literature. No intensity threshold was used here. For higher intensity thresholds see Appendix B – Figure B.9 and B.10.*

The cross-correlation of the CNN nowcasts averaged for each event type is illustrated in Figure 7.11 for both the temporal (upper) and the spatial (lower) cross-correlation. The first noticeable result, is that the performance of the convective events is considerably worse than

the ones from the other events. While the radar and the gauge performances are higher than the value of the predictability limit in the literature for all the events, for the convective ones this is not the case. Although the radar predictability is 20min, as suggested by the literature, this is not satisfactory. This can either be caused by the low impact of convective events in the training of the CNN (convective events are shorter in duration and fewer in numbers) or the inability of the CNN to learn processes specific to the convective type. Lastly, when compared to each other, unlike the behaviour of the reference nowcasting methods, the spatial and temporal cross-correlation are quite similar.
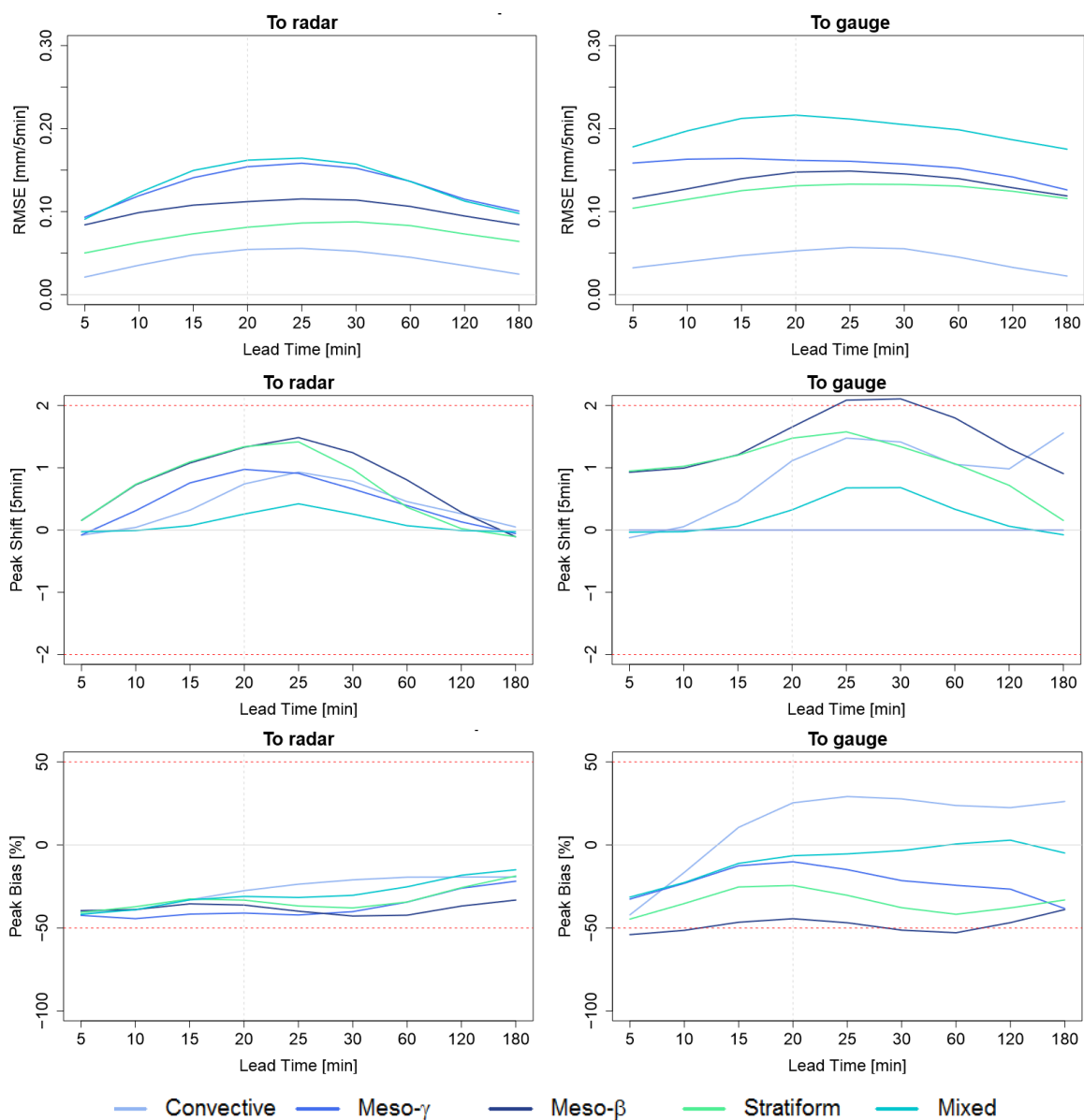

Figure 7.12: *The median of the temporal performance: RMSE (upper), Shift in Peak Time (middle) and Peak Bias (lower) for each of the event types of the CNN nowcast method with different lead times, compared to the input radar data (left) and gauge data (right). No intensity threshold was used here.*

The meso-γ convective events, on the other hand, have similar duration with the convective events (Figure 3.6-Section 3.4.1) and same number of events. Nonetheless their performance is quite good, with cross-correlation always above the 0.37 limit. Thus, the number of events and duration cannot be the reason why the convective events are simulated so poorly. One main difference between these two types of events lies at the storm extent; the meso-γ events have extents much higher than the convective ones. Judging from the results of the CNN, it seems that the evolution of the meso-γ convective events is controlled by the global scale, thus the performance is better. On the other side, for simple convective events, since their extent is quite small and the CNN is not able to predict their evolution good enough, it seems that the global scale cannot model the storm evolution. This is understandable as the small-extent convective events are initialized quite rapidly by temperature instabilities and are not affected much by large scale processes.

The additional temporal performance criteria of the CNN nowcasts, averaged for each event type, are shown in Figure 7.12. As expected the RMSE (even though the temporal and not the spatial one) is quite low, as the CNN was optimized by reducing the RMSE value. It is surprising to see, that for all types of events the CNN can capture quite well the time of the peak. Of course, there is a time shift of 5 to 10 minutes, but this stands within an acceptable range. Nevertheless, CNN displays a constant bias in peak volume, especially when compared to the input radar data regardless of the lead time. This performance is a bit disappointing, nevertheless as the time of peak is well captured, the volume can be later corrected either with a bias adjustment method or perhaps by employing an extra layer of maximum pooling in the CNN structure.

Overall, for all the events there seems to be a pattern over the lead times considered; the longer lead times (+60, +120 and +180 minutes) seem to perform better. This is in accordance with the training and validation loss functions as per Figure 7.9 and Figure 7.10. Even though the CNN model would clearly perform better with higher epochs (as the convergence is not yet reached), the training yielded lower losses than the other lead times. On the other side the performance of the short lead times is also good, they display low temporal errors and high correlation. Nevertheless, the underestimation of the peak volume is higher than the ones displayed by HyRaTrac and Lucas-Kanade nowcast. This suggest that probably at these lead times a saturation is reached (indicated also by the strong gradient decrease and then constant in the training loss plot). Regarding the other lead times (20-30 minutes) the poor performance is directly related to the training loss curve achieved (Figure 7.9-middle). For longer training epochs, results might improve.

7.3.3   Improvement of Rainfall Predictability

The CNN nowcast for the 93 events was compared directly to the Lucas Kanade (LK) nowcast, to see the extent of the CNN potential to add value in the predictability of storms. Note that a leave-one-out cross-validation was not used for the CNN, and the following results express just the maximum improvements towards the LK method. The improvements are calculated as per Equations 5.19 -5.22 from Section 5.3.2, and the average values over all the 93 events of the selected performance criteria are illustrated in Figure 7.13. As the CNN is trained on the minimization of RMSE, both temporal and spatial one, display a big improvement (for all the lead times) when compared to the LK nowcast. The same behaviour is seen as well for the temporal peak shift. Meanwhile for the temporal peak bias (as expected) the improvement is mainly at the long lead times.

Regarding the temporal and spatial cross-correlation the improvement is seen for lead times higher than 20 minutes, although the improvement to gauge observation is higher than the one to radar observation. This is in accordance with the literature, stating that the Lagrange persistence works quite good for the short lead times. Apparently, for these lead times, the non-linearity doesn't introduce additional skill to the nowcast method, the linear motion is more than enough. However, for high lead times, the benefit of the non-linearity transformation is clearly visible.

The results shown here, are valid for a 0 mm/h threshold set on the data. The results for the other thresholds (1, 5, 12 and 60 mm/h) are given in Appendix D – Figure D.5 to D.8. With a higher threshold, the nowcast CNN gives better results than the LK for shorter lead times. Although in general it still gives the best performance for the long lead times (+60, +120 and +180 min). Overall, the CNN seems able to capture the intermittent nature of rainfall by simulating the growth and decay rate and even death of storms. While the former is visible at all events, the latter is mainly visible at the meso-scale events, whose duration is on average 100 minutes, but the lead times of +120 and +180 minutes show the best results. Even though theoretically speaking the CNN is not able to predict storm births, because as in LK nowcast rainfall has to be detected in order to initialize the nowcast. However, in cases when the nowcast has been initialized, the CNN may predict the birth of storms in the vicinity of pre-existing storms.
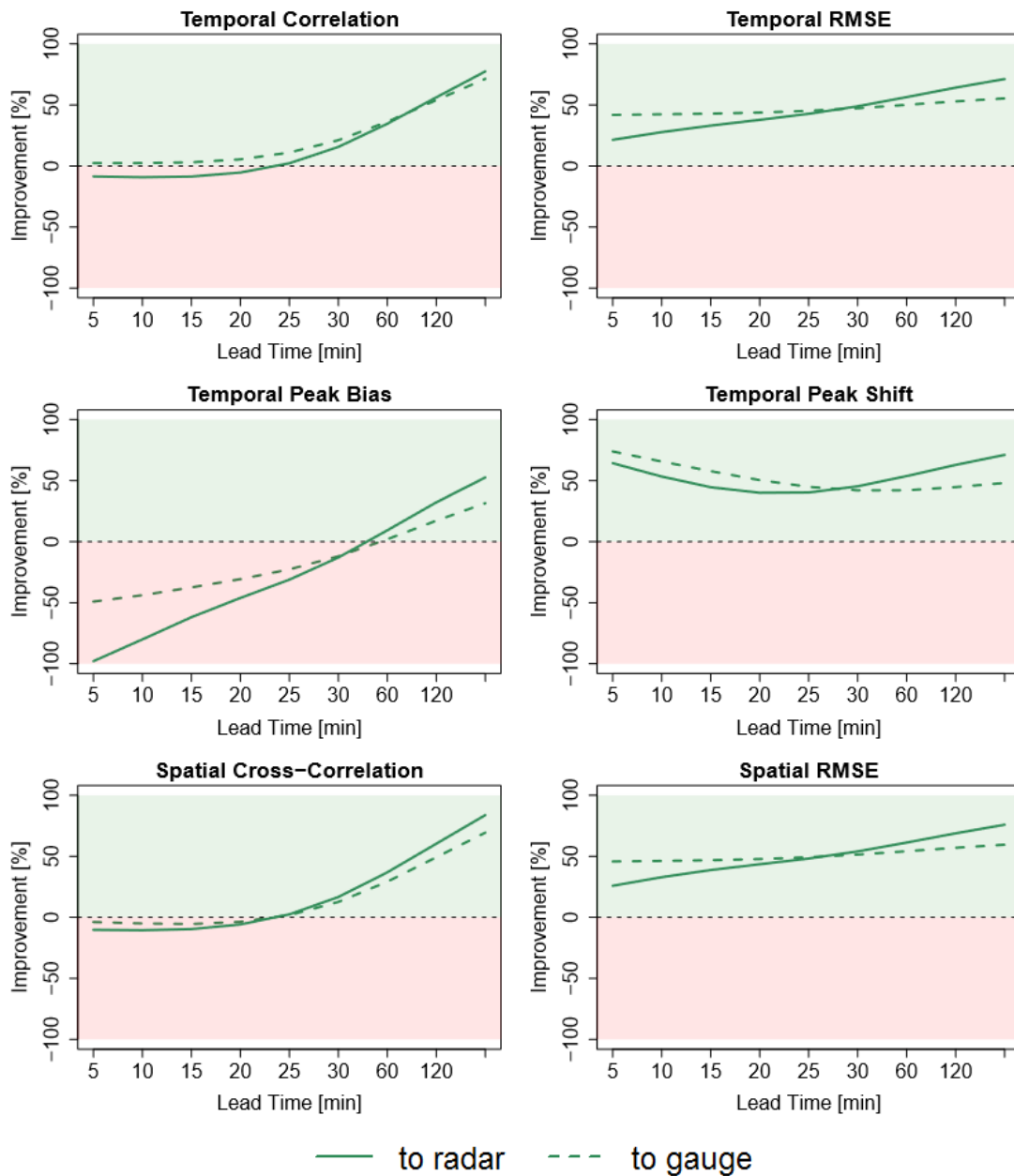
Figure 7.13: *The average improvement of Spatial and Temporal Criteria for each lead time (from 5 to 180 min) of the CNN nowcast compared to the LK one. The green area indicates a better performance by the CNN, and the red area indicates the contrary. No intensity thresholds was used here. For higher intensity thresholds see Appendix D – from Figure D.5 to D.8.*

8.   FINAL COMPARISON OF THE METHODS

Three new methods were trained and validated on 93 events in order to improve the nowcast skill and the predictability of storms at fine scales. These methods are; the conditional merging of radar with gauge data (in order to have a better rainfall field), the k-NN nowcast approach (in order to have a better estimation of the storm evolution in time and space), and a CNN model (in order to have a better predictability of the rainfall field). The three methods proved to improve the nowcasting skills for all the types of events used in the training dataset. Nevertheless, only a few of these events are characterized by extreme rainfall intensities at short durations. In order to test the improvement of the nowcast achieved with these methods specifically for the urban flood application, 17 additional events that display very high rainfall intensity at short duration, are used for the final assessment of the methods. The aim here is to test, whether the developed methods trained with a variety of event types, can be still useful in improving the predictability of extreme events that are causing urban pluvial floods. Of course, the ideal case would be to train the methods only for events causing urban floods, nonetheless this is very difficult to achieve. First because there are not enough recorded extreme events (for instance events with return period higher than 100 years) at a specific location to train and to validate the models. Secondly, even if enough extreme events are at hand, one would need to know in advance that an extreme event is about to happen, and initiate the nowcast. Even though sometimes, extreme events may be expected due to extreme meteorological conditions and NWP predictions, issuing a nowcast trained only on extreme events may tend to overestimate the intensities and to issue false alarms. Thus, methods trained in a variety of events, are used here to predict the extreme events and to answer the following questions: Does the CM improve the rainfall field also in cases of extreme short events? Up to which limit can the existing nowcast method HyRaTrac and LK yield reasonable results for such extreme events? Can the developed k3-NN and the CNN extend the predictability of such storms at the point scale? Can any of the ensemble members issued by k-NN provide any additional value to the nowcast? These are the main questions investigated in this chapter.

First at Section 8.1 the performance of the CM input field is discussed, followed up by the direct comparison of the nowcast methods in Section 8.2. The 17 events used in this chapter are explained in Section 3.4.2 and the performance criteria assessment is described in 4.2. To

have a direct and fair comparison of the methods, only the stations as illustrated in Figure 7.7 due to the restriction window of the CNN, are considered for the performance assessment.

8.1 Improvement of Rainfall Field

Figure 8.1 illustrates both temporal and spatial performance criteria of the raw radar data (RR) and conditionally merged radar data (CM) calculated for each of the 5 intensity thresholds considered in this study and averaged for each of the events. Here the true reference is considered the gauge information.
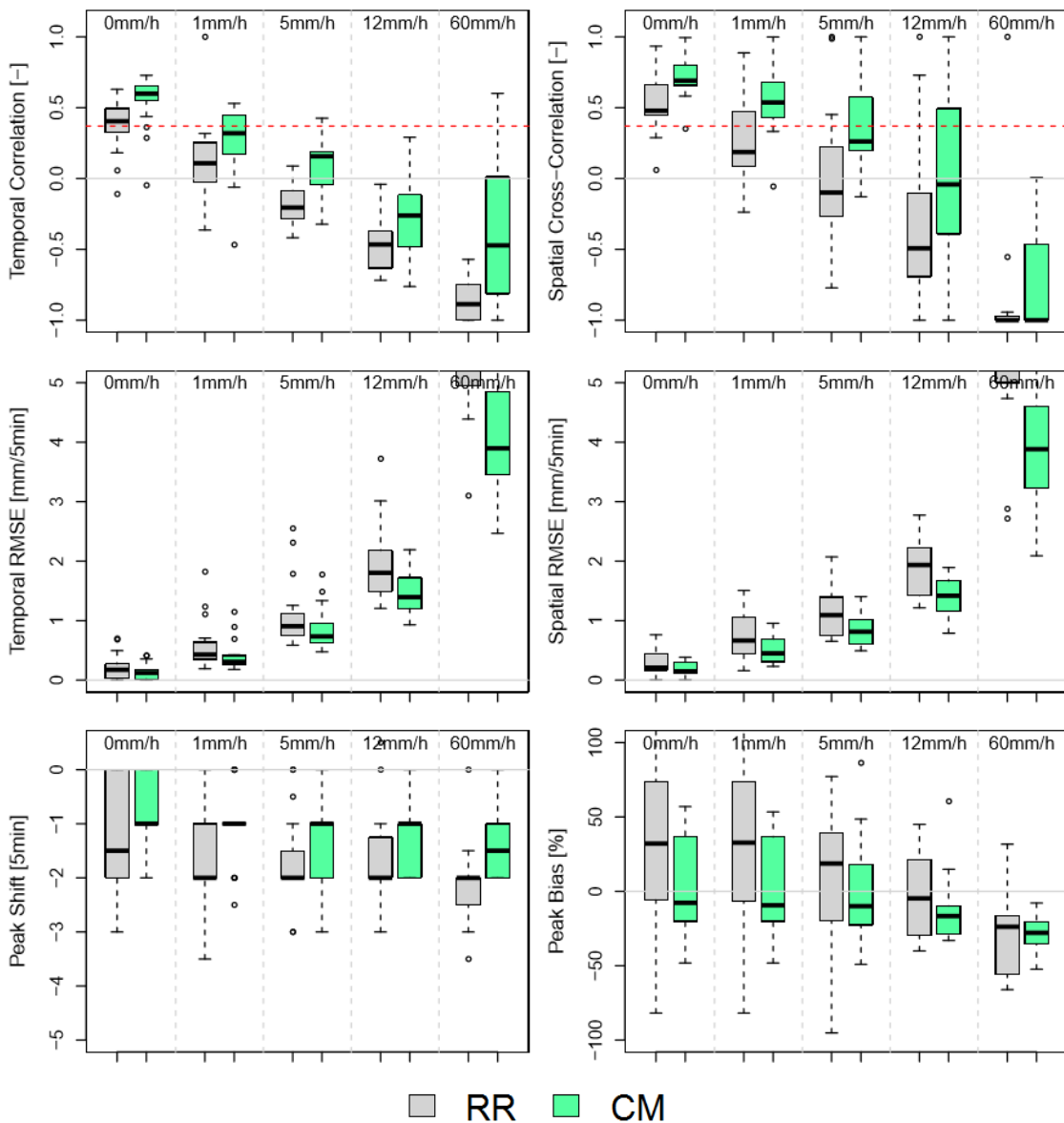

Figure 8.1: *The temporal and spatial performance criteria of the two rainfall products raw radar data (in grey) and conditional merging (in green) averaged per event for each of the intensity thresholds considered. The red dashed line indicates the limit of predictability, while the grey lines are for guidance.*

As seen in the training events in Section 5.4.4, also for the extreme events, the CM outperforms the RR data for all of the intensity thresholds. Both temporal and spatial cross-correlation and RMSE are improved considerably for all the thresholds. Regarding the peak performance, the peak time is improved to a delay of 5 minutes and an underestimation of mainly 10%. The spatial cross-correlation performance is higher than the temporal one indicating that the spatial patterns may be easier to predict than the temporal rainfall patterns at a point location. Moreover, the CM data display a cross-correlation higher than the predictability limit only for the lower intensity thresholds (0 and 1 mm/h), thus no predictability can be expected from the nowcast methods for the higher intensity thresholds (5, 12 and 60 mm/h). For this reason, only the low intensity is shown in the following section, to illustrate the predictability skill of the nowcasting methods for the extreme events.

8.2 Improvement of Rainfall Nowcast

For the comparison of the methods, apart from the HyRaTrac and LK methods used as reference, the deterministic nowcast of k3-NN and the CNN method are considered. To investigate as well the value of the ensembles, the first 20 ensemble members issued for each time step are considered. These ensembles are mainly useful to visualize the uncertainty in the spatial performance rather than in the temporal ones. In order to visualize the uncertainty for different temporal performances, then the combinations of all the ensembles issued at each time step should be regarded. Due to the complexity and the computational time, the results of the ensembles are shown here only for the spatial performances. As the CM was proven to outperform RR, it was used as an input for all the nowcast methods.

Figure 8.2 and 8.3 illustrate respectively the temporal and spatial performance of the selected nowcast methods both compared to the radar input (model performance) and to the gauge reference (practical performance). In terms of predictability, HyRaTrac displays the worst temporal and spatial predictability for both model and practical one. It seems that for such extreme events, the identification and tracking of the storm cells is not enough to display some predictability skill at any lead time. This, of course, affects greatly the performance of the k-NN method.

As the k-NN method is a continuation of the tracking done by HyRaTrac, substituting only the extrapolation in the future with the Lagrangian persistence, the results cannot be much better than the HyRaTrac. This is particularly true for the temporal performances. As seen in Figure 8.2, the k-NN approach behaves slightly better than the Lagrangian persistence. Nevertheless, the main improvements of k3-NN are observed at the spatial performance for lead times up to 30 minutes. Clearly the k3-NN performs better than the Lagrangian persistence, but due to the tracking of individual storms, very short model (+10 min) or practical predictability (+5min) is observed. Nonetheless some of growth and decay processes are captured better by the k3-NN, while the death ones (see for lead times higher than an hour), the performance of the k3-NN is quite similar with the Lagrangian one.
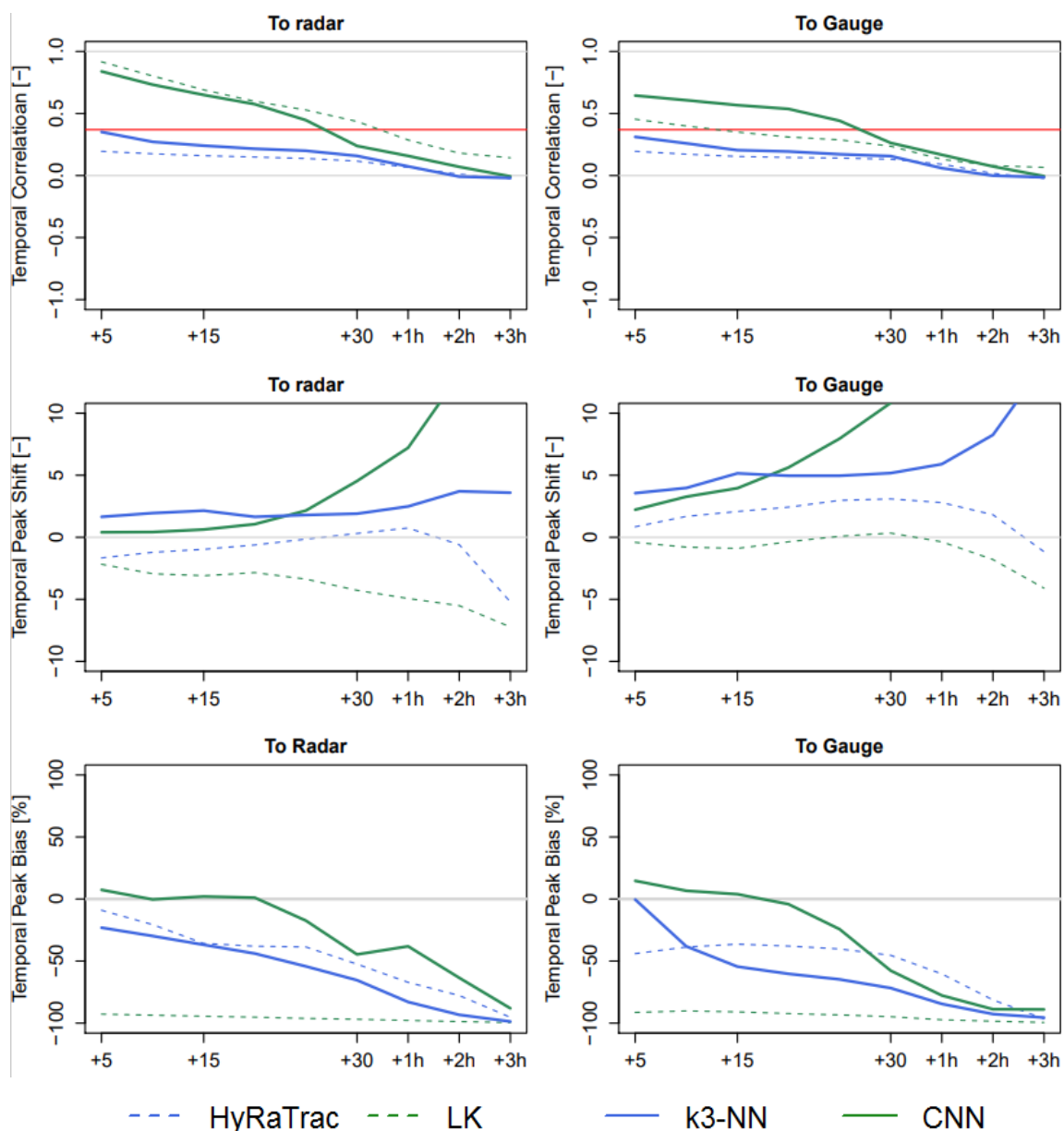
Figure 8.2: *The temporal performance criteria of the selected nowcast methods both compared to the CM radar input (left) and the respective gauge information (right) averaged over all the 17 events for all the lead times up to 3 hours (x-axis). No intensity threshold was used here. The red dashed line indicates the predictability limit while the grey lines are for guidance only.*

LK on the other hand displays good predictability of the storms, with spatial radar (model) predictability reaching up to 1 hour and spatial gauge (practical) predictability up to 30 min. The performance of the LK is improved greatly by the use of the CM product as input data. As in the case of the training events, the Lagrangian persistence based on optical field can yield reliable spatial results up to 1 hour when fed with better and more accurate rainfall data. Past this lead time, the predictability of the LK decreases quite rapidly. Regarding the temporal predictability, the radar (model) one is extended to about 30 minutes, while the

gauge (practical) one can barely reach the 15 min lead time. This is the same pattern as in the training events: the LK can capture quite well the spatial variability of the rainfall, but not so good the temporal variability at the point scale. This is also seen by the high underestimation that the LK causes to the peak volume both when compared to the radar or the gauge input. Even though the temporal shift of the peak is improved (a delay of 5 to 10 minutes is exhibited for the short lead times), there is a constant underestimation of the rainfall intensities.

The results from the 20 ensembles of the k-NN approach are illustrated as well in Figure 8.3. As expected all of the ensembles bring an additional value to the HyRaTrac nowcast. For the first 15 min lead times, all of these ensembles deliver better performances than the Lagrangian persistence. Moreover for these lead times, it seems that the k3-NN is representing the best solutions from the similar storms. For the high lead times, the importance of the ensemble nowcast is better seen. For instance, for lead times of 1 hour or even longer, the performance exhibited by the ensembles is better than both of the deterministic approaches. Apparently, they help in capturing better the decay and the death processes of the storms. Lastly, the performance band obtained from the ensembles, can help to quantify the uncertainty of nowcasts past the 15-min lead time. At the present, due to the 20 number of ensembles chosen, the uncertainty band is quite narrow (but still better than the reference), nevertheless wider bands can be obtained with a higher number of ensembles considered.
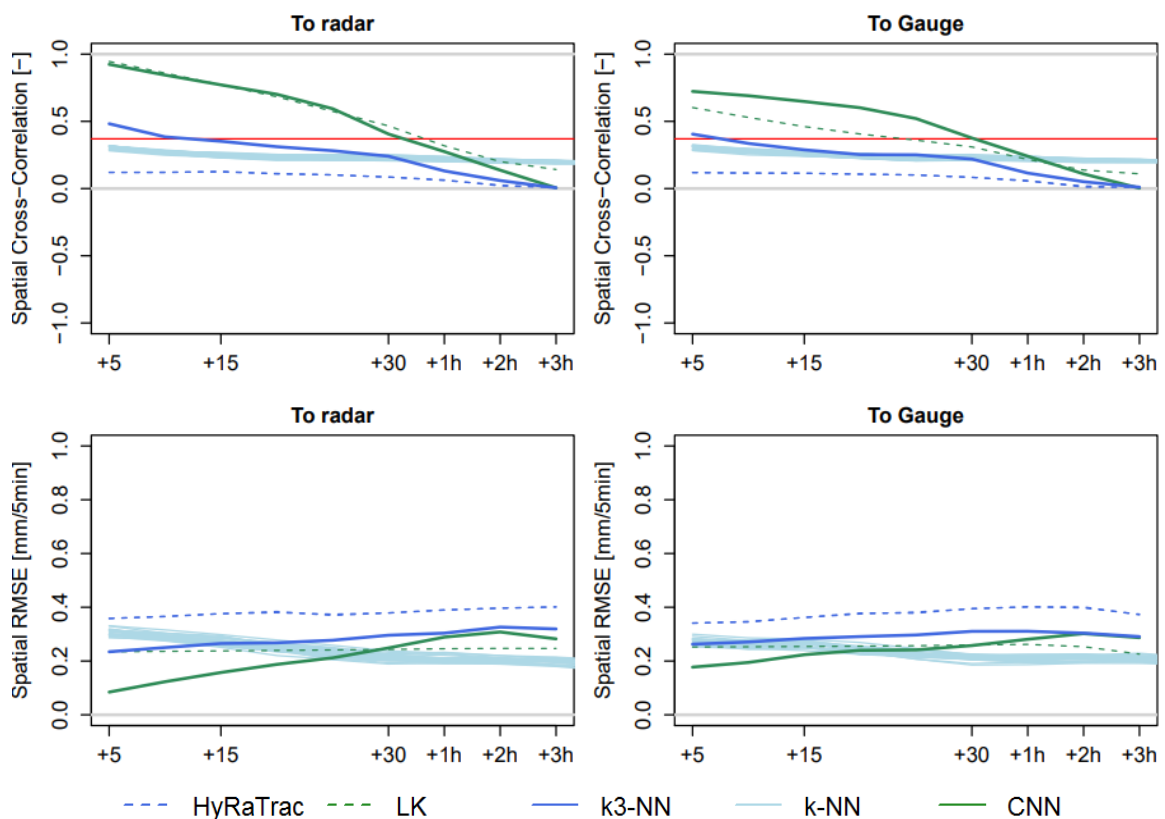

Figure 8.3: *The spatial performance criteria of the selected nowcast methods both compared to the CM radar input (left) and the respective gauge information (right) averaged over all the 17 events for all the lead times up to 3 hours. The red dashed line indicates the predictability limit while the grey lines are for guidance only. No intensity threshold was used here.*

The CNN on the other side exhibits the highest predictability skill for both temporal and spatial performance. For the radar (model) predictability (both spatial and temporal one) the CNN displays similar limits and behaviour as the LK nowcast. Here the spatial predictability is extended to 1-hour, and temporal one to round about 30 minutes. Regarding the temporal radar (model) predictability, it seems that LK performs better than the CNN, but nevertheless one has to keep in mind, that the line for the CNN is discrete as the training has been done only for lead time +30min, + 1hours and not for the other lead times in between. Thus it may happen that the CNN exhibits higher predictability for the durations between these two lead times, and thus displays an extended predictability.

The main advantage of the CNN is visible in estimating the gauge (practical) predictability of the nowcast for both temporal and spatial performance. The temporal practical predictability is extended to 30 minutes, while the spatial one is extended to almost one hour. Also, it has to be noted, that for the short lead times the CNN performs much better than the LK method. Even though during the training of the CNN, the short lead times were considered saturated, thus not so good results were expected of them, their performance on the extreme events is quite satisfactory. Regarding the volume peak, it seems that the CNN can capture well enough the peak shift and volume until 30 minute lead time for both nowcasts compared to radar and to gauge data. Of course, the performance is getting worse with the lead time, but for the first 20 minutes the peak bias is around 0%, while from 20-30 minutes it increases gradually to 40%. Moreover unlike the LK, the performance of the CNN seems to be similar for both spatial and temporal performance and consistent for both model and practical performances.

9.   CONCLUSIONS AND OUTLOOK

9.1 Summary

Accurate rainfall nowcasts at fine temporal (5 min) and spatial (1 km^2) scales are essential for the prediction of urban floods, nevertheless are very challenging to achieve. The intermittence nature of rainfall at such small scales and our inability to measure its true distribution in space and time, make it quite difficult for the simple persistence models to work reliably past the 20 min lead time. Consequently, a predictability limit is built up which cannot be surpassed by the models at hand. This is especially true for the urban flood application, where the rapid high intensity peaks of rainfall field are crucial. Thus, the scope of this thesis was to investigate how different event types through the state-of-the-art persistence models exhibit different predictability limits and how such limits can be pushed further by employing more complex models.

For this purpose, the area with a radius of 115 km around the Hannover City, Germany, was considered. Here only a radar scan was employed to nowcast rainfall intensities at the 5 min time steps and 1 km^2 spatial scales. From the period 2000-2018 of available radar data, 110 events with different characteristics and severities were selected as a basis for the investigation. First, 93 events were used for the overall predictability limit investigation and the training of the complex nowcast models, while 17 events, with high importance for urban flood generation, were selected to investigate the predictability limit of the extreme events. Two state-of-the-art approaches based on radar data were selected as a reference for the evaluation: HyRaTrac (an object-oriented approach) and Lucas Kanade Optical Flow (a field-oriented approach).

Even though the rainfall nowcasting remains a difficult task to perform, a framework was proposed and implemented successfully in this thesis to improve the rainfall nowcast predictability, particularly for the point scale. The main milestones of the framework were: I) integrating gauge measurements in the nowcast, ii) integrating knowledge from the past observed storms in the nowcast, iii) utilizing ensemble prediction rather than deterministic ones. The main conclusions derived from this thesis are discussed in the following section as answers to the research questions stated in Section 1.2.

9.2 Conclusions

Research Question 1: The decorrelation time from Germann and Zawadzki (2002) was used as a tool to estimate the predictability limit at area spatial scales. To match the purpose for urban modelling, this concept was modified to illustrate the decorrelation time of the temporal correlation on a point scale. Moreover, other point temporal performances like peak bias and peak shift were used as additional tools to estimate the nowcast skill of the methods.

Research Question 2: Since both of the reference methods employ the Lagrangian persistence for the extrapolation of the rainfall information into the future, they fail to model correctly the transformation processes like birth, growth, decay and death, restricting the predictability limit of rainfall up to 15 or 20 minutes. In particular, the convective events with high intensity displayed very short predictability limits (especially at the point scale, the predictability is limited to 10 min), indicating that the intermittent nature of the rainfall is more important at the point scale than at the large one. Thus, there is a need to improve not only the predictability limit at the large scale, but as well as on the point scale.

Research Question 3: The gauge information was added to the observed radar field based on two assumptions; first as a point information when integrated with the radar data it would improve the point predictability of rainfall, second it leads to better urban flood nowcasts as the urban models are typically calibrated on gauge data. In total 100 gauge recordings inside the radar range were merged together with radar observations on 5 min time steps. Five different merging techniques were compared, namely Mean Field Bias (MFB), Quantile Mapping based Bias Correction (QQ), Kriging with External Drift (KED), and Conditional Merging (CM). The latter one, CM on spatial and temporal smoothed radar data, reached the best agreement between the accuracy of gauge measurements and the high spatial variability of the radar data. Despite the fact that KED is the most preferred as a merging method in the literature, the results show clearly than the Kriging with radar data as an external drift cannot yield as good results as the CM. The QQ on the other hand has not been used so far (to the knowledge of the author) for the online merging of both data sources, thus it was surprising to see that it performs similar to the CM. Unfortunately, even though the large scale is represented very good by the QQ, it is unable to represent the point scale of the rainfall, as it displays high underestimation of the rainfall peaks. Nevertheless, QQ seems promising and further investigations are advised in order to reach a better rainfall product. Overall the CM on temporally and spatially smoothed radar data yielded the best results both at point and large scale, and thus was used as an input to the nowcast algorithms.

Research Question 4: When fed to both nowcast algorithms, the CM improves considerably the predictability of all the event types for the short lead times (up to one hour). The Lucas-Kanade nowcast seems to take the most advantage of this; displaying a nowcasting skill up to one hour lead time. The method seems to favour more the large scale

events (such as stratiform or mixed). Even though improvements are seen as well for the convective events, especially for the object based approach, an advection correction should be implemented in order to achieve even better results. Moreover, with the application of the merging, the local predictability behaves as good as the larger scale, emphasizing again the importance of two products in the nowcast skill. Overall, the CM is advised for the pluvial flood application as it is almost bias free when estimating the amount and time of the local rainfall peaks; it improves the point predictability of the rainfall and the agreement with the gauge measurements. This is not only true for regular events, but has been proven as correct as well for the extreme events that have high importance for urban floods.

Nevertheless, even though the CM can improve the predictability of rainfall, there is a limit to the additional value that the CM can bring to the nowcast. At this limit, the errors associated to the persistence nowcast are greater than the errors of the input. Thus, to extend the predictability even further, the temporal and spatial patterns are learned from previously observed storms. Two data-driven approaches were proposed in this thesis for further improvement; a nearest neighbour storm (for the object-oriented approach) and a convolutional neural network (for the field-based approach).

Research Question 5: The nearest neighbour approach (k-NN) works as an extension of the HyRaTrac algorithm. Once a storm is identified and tracked by HyRaTrac, the extrapolation in the future is done based on the 3 most similar neighbours from the past dataset. The procedure applied here seems to yield very good results for the area scale, causing spatial cross-correlation higher than the limit value for all of the lead times. Once similar neighbours are successfully identified, the response from the neighbours improves greatly the predictability. This supports the assumption that similar storms do behave similarly and past information can supply with additional useful information. Nevertheless, the similar neighbours are not always successfully identified. For specific time steps the performance was unsatisfactory, because the k-NN failed to identify very similar storms. Moreover, for the first time step of the storm life, the selected predictors do not describe correctly the similarity of storms. Thus, it is advised to use the k-NN as an extension to HyRaTrac only when the following criteria are fulfilled: i) similar storms are identified, meaning that the cross-correlation optimization reaches values of higher than 0.5, and ii) issuing nowcasts from the 2nd time step of the storm life. When any of these criteria are not fulfilled, the Lagrangian persistence should be used instead.

Research Question 6: Even though the spatial performance of the k-NN was good, it failed to capture the temporal variability of the rainfall. Clearly a spatial optimization at each time step is altering the temporal variability of the rainfall and causing underestimation of the rainfall. Thus, the k-NN could bring additional value to the HyRaTrac nowcast, extending mainly the spatial point predictability of rainfall. Nevertheless, the performance of the k-NN method is highly dependent on the matching and tracking of the storms by HyRaTrac. This was seen for the extreme events, when the object-based approach did not display any predictability at all. The k-NN could improve the performance when compared to the

Lagrangian persistence, but nonetheless could not reach values above the predictability limits.

Research Question 6: The CNN was trained independently at each lead time based on a split sampling of the training data set. The CNN seemed to perform particularly well for the stratiform and meso-scale convective events, where patterns at coarser scales dictate the overall movements of the rainfall structures. Overall, the CNN has the potential to extend the predictability limit up to 3 hours lead time. For extreme events, the CNN seems to yield reasonable results, despite the fact that it was not trained on extreme convective rainfall events. For these events, the CNN improves the predictability of the storms up to 30 − 40 minutes lead time. Especially for the practical predictability, the CNN brings the most added value to the nowcast. Moreover, unlike the other methods, the performance displayed by CNN both at temporal and spatial performance are very similar and constant between the radar (model) and gauge (practical) predictability.

Research Question 7: Ensemble predictions were issued only by the k-NN nowcast method, where several neighbours were selected for the nowcast. The number of neighbours were not fixed, but were adapted accordingly to only members that exhibit a cross-correlation higher than 0.37 with the last observed radar image. For the short lead times the deterministic approach yielded better results than the ensemble nowcast. However, the number of useful ensembles increased with the lead time, suggesting that the ensemble nowcast is particularly useful for the long lead times. This makes sense, as for higher lead times, the uncertainty is much higher due to the death and decay of storm cells. This was also the case for the extreme events, where the ensemble nowcast outperformed the deterministic one. Thus, ensemble nowcasts are suggested to be used whenever the uncertainties are expected to be large. For predictable storms and short lead time, the deterministic approach may be used instead.

Research Question 8: To conclude, the field based nowcasts are more suitable for the nowcasting of extreme rainfall at fine spatial and temporal scales. The object based nowcast are highly dependent on the storm matching algorithm, which for severe convective events, do not represent the temporal evolution of the storm. Thus, for the urban flood nowcast, field based approaches are recommended. More precisely the CNN trained on merged radar data should be used as a basis for the flood nowcast, as it represents the most robust method for rainfall estimation at point scales. Contrary, the k-NN method should be used when rainfall estimation at the storm scales are required, or when the spatial evolution of storms is needed.

9.3 Outlook

The k-NN method seems a promising method in terms of nowcasting based on past observed storms. Nevertheless, more work has to be done at estimating the right predictors for similarity estimation. Monte-Carlo simulations with various dataset and predictors may

be used to explain the variance of the target variables (i.e., Sobol indexes). However, as the target variables are many (and continuous in time) it may be difficult to explain simultaneously the variance of several predictors. Estimating the right predictors, could help to classify the storms early on in different behaviour groups, which can serve as a basis not only for nowcasting purposes but also for other hydrological purposes. For example, determining the impact of certain storm types on different urban or rural catchments, use different storm statistics for generating area synthetic rainfall data for areal Intensity-Frequency-Duration curves or other design purposes, recognition of dynamic clutter and application of specific Z-R relationship for each storm type in radar data correction, etc.

Another problem with the application of the k-NN is the limitation of the data set. For instance, the radar used for this thesis has only 19 years of observations, thus when issuing nowcasts, the method may not always find the right neighbour, especially in the case of extreme events. A regionalization may be required here. Using the storm information from other regions can increase considerably the data set, thus improve the performance of the method. Once the right predictors have been identified to recognize similar storms, they can be used as well to classify storms in similar groups. In addition to these predictors, one can reduce the data set by using certain location descriptors that enable the selection of storms that have similar meteorological and morphological conditions to the target location.

Also, the CNN can benefit from a bigger data base by considering radar observations at other regions. Even though the CNN showed the best performance for the events at hand, further improvements can be achieved if the database is larger. If more extreme events are available, a special tuning of CNN can be done, in order to learn separately weights specifically for these extreme events. Information from NWP can then be used as an additional input when initializing the weights from the special tuning, and thus to predict better the evolution of the extreme events. Further improvements of the CNN may be achieved if longer past time steps are considered in the training of the weights, or if better spatial criteria are used for the training. For instance, instead of RMSE, a cross-correlation tool or even the structural similarity index (SSIM) might be applied to improve the nowcasts. Moreover, probabilistic nowcasts may be issued by integrating the model uncertainty; Neurons may be switched randomly in order to capture the structure uncertainty, different activation function may be employed to capture the non-linearity uncertainty and training with different data set may be done to capture the input uncertainty.

Further works may include the use of the both methods (k-NN and CNN) on polarized radar data. For the k-NN application the Differential Phase Shift (K) from polarized data can be used as an additional predictor to recognized similar storm types. As storms caused by different mechanisms display different hydro-meteors, the K-predictor seems promising in distinguishing successfully different storm types. The K-information may be used as input as well for the CNN nowcast in order to learn weights according to the hydro-meteor type. Furthermore, a better product may be achieved if the field- and object- based nowcast are integrated together. As the CNN is smoothing the rainfall structure in space, information from

a k-NN nowcast may be included in order to maintain similar storm shapes. Finally, as discussed in the literature review, the best approach to extend the predictability limit past 2 or 3 hours lead times would be to couple the radar based nowcast with the NWP forecasts and to regard the storm birth uncertainty by stochastic random generators.

Bibliography

Ahrens, B., Jaun, S. (2007). On evaluation of ensemble precipitation forecasts with observation-based ensembles. Advances in Geosciences, 10, 139-144.

Akbari Asanjan, A., Yang, T., Hsu, K., Sorooshian, S., Lin, J., Peng, Q. (2018). Short-Term Precipitation Forecast Based on the PERSIANN System and LSTM Recurrent Neural Networks. Journal of Geophysical Research: Atmospheres, 123 (22), 12,543-12,563.

Asquith, W.H. (2011). Distributional analysis with L-moment statistics using the R environment for statistical computing: First edition. Lubbock, Texas, Create Space Independent Publishing Platform, 2nd printing, ISBN 978–1463508418.

Ballard, S., Li, Z., Simonin, D., Buttery, H., Charlton-Perez, C., Gaussiat, N., Hawkness-Smith, L. (2012). Use of radar data in NWP-Based nowcasting in the Met Office. IAHS-AISH Publication, 351, 336-341.

Bárdossy, A., Pegram, G. (2013). Interpolation of precipitation under topographic influence at different time scales. Water Resources Research, 49.

Berenguer, M., Surcel, M., Zawadzki, I., Kong, M., (2012). The Diurnal Cycle of Precipitation from Continental Radar Mosaics and Numerical Weather Prediction Models. Part II: Inter-comparison among Numerical Models and with Nowcasting. Monthly Weather Review, 140(8), 2689-2705.

Berndt, C., Rabiei, E., Haberlandt, U. (2013). Geostatistical merging of rain gauge and radar data for high temporal resolutions and various station density scenarios. Journal of Hydrology, 508, 88-101.

Berne, A., Delrieu, G., Creutin, J., Obled, C. (2004). Temporal and Spatial Resolution of Rainfall Measurements Required for Urban Hydrology. Journal of Hydrology, 299, 166-179.

Berne, A., Krajewski, W. (2013). Radar for hydrology: Unfulfilled promise or unrecognized potential? Advances in Water Resources, 51, 357-366.

Borga, M., Anagnostou, E. N., Frank, E. (2000).On the use of real-time radar rainfall estimates for flood prediction in mountainous basins. Journal of Geophysical Research Atmospheres, 105(D2), 2269-2280.

Bowler, N. E., Pierce, C. E., Seed, A. W. (2006). STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP. Quarterly Journal of the Royal Meteorological Society,132, 2127-2155.

Pierce, C.E., Ebert, E., Seed, A.W., Sleigh, M., Collier, C. G., Fox, N. I., Donaldson, N., Wilson, J. W., Roberts, R., Mueller, C. K. (2004). The Nowcasting of Precipitation during Sydney 2000: An Appraisal of the QPF Algorithms. Weather and Forecasting, 19, 7-21.

Chiang, Y., Chang, F., Jou, B. J., Lin, P.F. (2007). Dynamic ANN for precipitation estimation and forecasting from radar observations. Journal of Hydrology, 334, 250-261.

Chumchean, S., Seed, A., Sharma, A.(2006). Correcting of real-time radar rainfall bias using a Kalman filtering approach. Journal of Hydrology, 317, 123-137.

Ciach, G.J. (2010). Local Random Errors in Tipping-Bucket Rain Gauge Measurements. Journal of Atmospheric and Oceanic Technology, 20, 752-759.

Codo, M., Rico-Ramirez, M.A. (2018). Ensemble Radar-Based Rainfall Forecasts for Urban Hydrological Applications. Geosciences, 8, 297, 2076-3263.

Coulthard, T.J., Frostick, L.E. (2010). The Hull floods of 2007: implications for the governance and management of urban drainage systems. Journal of Flood Risk Management, 3, 223-231.

Crozier, C. (1986). Weather Radar Operations Manual and User Guide. Toronto Weather Radar Research Station.

Curry, G. R. (2012).Radar essentials: A concise handbook for radar design and performance analysis.

Deutsch, C. V., Journel, A. G. (1998).GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press.

Dumoulin, V., Visin, F. (2016).A guide to convolution arithmetic for deep learning.

DWA Regelwerk (2012). Starkregen in Abhängigkeit von Wiederkehrzeit und Dauer.

DWD (1998). RADOLAN: Routineverfahren zur Online-Aneichung der Radarniederschlagsdaten mit Hilfe von automatischen Bodenniederschlagsstationen (Ombrometer) .Geschäftsfeld Hydrometeorologie.Deutscher Wetterdients

Ehret, U. (2003). Rainfall and Flood Nowcasting in Small Catchments Using Weather Radar. Institut für Wasserbau. Universität Stuttgart.

Einfalt, T. (2005). A hydrologists' guide to radar use in various applications.

Einfalt, T., Hatzfeld, F., Wagner, A., Seltmann, J., Castro, D., Frerichs, S. (2009). URBAS: forecasting and management of flash floods in urban areas. Urban Water Journal,6,369-374,1573-062X

Fletcher, T. D., Andrieu, H., Hamel, P. (2013). Understanding, management and modelling of urban hydrology and its consequences for receiving waters: A state of the art. Advances in Water Resources,51(),261-279, 03 091 708

Foresti, L., Reyniers, M., Seed, A., Delobbe, L. (2016). Development and verification of a real-time stochastic precipitation nowcasting system for urban hydrology in Belgium. Hydrol. Earth Syst. Sci.,20,505-527,1607-7938

Fraser, A. M., Swinney, H. L. (1986). Independent coordinates for strange attractors from mutual information. Physical Review A,33(),1134-1140,1050-2947

Galieti, G. (1990). A comparison of parametric and non-parametric methods for runoff forecasting. Journal of Hydrological Sciences, 35 (1). pg 79-94.

Gerchow, M. (2018). Deep learning based stereo depth estimation with application to laser microsurgery. Institut für Mechatronische Systeme. Leibniz Universität Hannover.

Germann, U., Jürg, J. (2001). Variograms of radar reflectivity to describe the spatial continuity of Alpine precipitation. Journal of Applied Meteorology,40.

Germann, U., Zawadzki, I. (2002). Scale-Dependence of the Predictability of Precipitation from Continental Radar Images. Part I: Description of the Methodology. Monthly Weather Review,130(),2859-2873.

Germann, U., Zawadzki, I., Turner, B. (2004). Scale Dependence of the Predictability of Precipitation from Continental Radar Images. Part II: Probability Forecasts. Journal of Applied Meteorology,43(),74-89.

Germann, U., Zawadzki, I., Turner, B. (2006). Predictability of Precipitation from Continental Radar Images. Part IV: Limits to Prediction. Journal of the Atmospheric Sciences,63,2092-2108,

Gires, A., Onof, C., Maksimovic, C., Schertzer, D., Tchiguirinskaia, I., Simoes, N.(2012). Quantifying the impact of small scale unmeasured rainfall variability on urban runoff through multifractal downscaling: A case study. Journal of Hydrology,442-443(),117-128,0022-1694

Golding, B. W. (2009). Long lead time flood warnings: Reality or fantasy? Meteorological Applications,16(),3-12,13504 827

Goudenhoofdt, E., Delobbe, L. (2009). Evaluation of radar-gauge merging methods for quantitative precipitation estimates. Hydrology and Earth System Sciences, 13, 195-203.

Grecu, M., Krajewski, W. F. (2000). A large-sample investigation of statistical procedures for radar based short-term quantitative precipitation forecasting. Journal of Hydrology, 239, 69-84.

Grünewald, U. (2009). Gutachten zu Entstehung und Verlauf des extremen Niederschlag-Abfluss-Ereignisses am 26.07.2008 im Stadtgebiet von Dortmund.

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

Haberlandt, U. (2015). Stochastische Simulation täglicher Abflüsse für die Speicherplanung unter Berücksichtigung von Klimaänderungen. Hydrologie und Wasserbewirtschaftung, 59, 247–254.

Han,L., Fu,S., Zhao, L., Zheng, Y.,Wang, H., Lin, Y. (2009). 3D Convective Storm Identification, Tracking, and Forecasting—An Enhanced TITAN Algorithm. Journal of Atmospheric and Oceanic Technology, 26, 719-732.

Heistermann, M., Jacobi, S., Pfaff, T. (2013). Technical Note: An open source library for processing weather radar data. Hydrol. Earth Syst. Sci.,17, 863-871.

Horn, B. K., Schunck, B. G. (1981). Determining optical flow. Artificial Intelligence, 17, 185-203.

Hu, Q., Li, Z.,Wang, L., Huang, Y., Wang, Y., Li, L. (2019). Rainfall Spatial Estimations: A Review from Spatial Interpolation to Multi-Source Data Merging. Water, 11.

Jacobson, C. R. (2011). Identification and quantification of the hydrological impacts of imperviousness in urban catchments: A review. Journal of Environmental Management, 92, 1438-1448.

Jasper-Tönnies, A., Hellmers, S., Einfalt, T., Strehz, .A, Fröhle, Peter (2018). Ensembles of radar nowcasts and COSMO-DE-EPS for urban flood management. Water Science and Technology,2017(),27-35,0273-1223

Jensen, D,G., Petersen, C., Rasmussen, M. R. (2015). Assimilation of radar-based nowcast into a HIRLAM NWP model. Meteorological Applications, 22, 485-494.

Jha, A., Lamond, J., Proverbs, D., Bhattacharya-Mis, N., Lopez, A., Papachristodoulou, N., Bird, A., Bloch, R., Davies, J., Barker, R. (2012). Cities and Flooding: A guide to integrated urban flood risk management for the 21st Century. Journal of Regional Science, *52 (5), 885-887.*

Kato, A., Maki, M. (2009). Localized Heavy Rainfall Near Zoshigaya, Tokyo, Japan on 5 August 2008 Observed by X-band Polarimetric Radar — Preliminary Analysis —Sola, 5, 89-92.

Kato, R., Shimizu, S., Shimose, K., Maesaka, T., Iwanami, K., Nakagaki, H. (2017). Predictability of meso-γ-scale, localized, extreme heavy rainfall during the warm season in Japan using high-resolution precipitation nowcasts. Quarterly Journal of the Royal Meteorological Society,143, 1406-1420.

Keil, C., Heinlein, F., Craig, G.C. (2014). The convective adjustment time-scale as indicator of predictability of convective precipitation. Quarterly Journal of the Royal Meteorological Society, 140, 480-490.

Kneis, D., Kneis, M. (2009). Quality assessment of radar-based precipitation estimates with the example of a small catchment. Hydrologie Und Wasserbewirtschaftung,160-171.

Krämer, S. (2008). Quantitative Radardatenaufbereitung für die Niederschlagsvorhersage und die Siedlungsentwässerung. Mitteilungen des Institutes für Wasserwirtschaft, Hydrologie und landw. Wasserbau. Leibniz Universität Hannover.

Krämer, S., Fuchs, L., Verworn, H.R. (2007). Aspects of Radar Rainfall Forecasts and their Effectiveness for Real Time Control -The Example of the Sewer System of the City of Vienna. Water Practice and Technology,2(),1751-231X.

Lanza, L.G., Vuerich, E. (2009). The WMO Field Intercomparison of Rain Intensity Gauges. Atmospheric Research,94(),534-543,0169-8095

Lau, C.L., Smythe, L.D., Craig, S.B., Weinstein, P. (2010). Climate change flooding, urbanisation and leptospirosis: Fuelling the fire? 104(),631-638.

Lin, C., Vasić, S., Kilambi, A., Turner, B., Zawadzki, I. (2005). Precipitation forecast skill of numerical weather prediction models and radar nowcasts. Geophysical Research Letters,32, 0094-8276

Liu, Y., Xi, D.G., Li, Z.L., Hong, Y. (2015). A New Methodology for Pixel-Quantitative Precipitation Nowcasting using a Pyramid Lucas Kanade Optical Flow Approach. Journal of Hydrology,529.

Löwe, R., Urich, C., Sto. Domingo, N., Mark, O., Deletic, A., Arnbjerg-Nielsen, K. (2017). Assessment of urban pluvial flood risk and efficiency of adaptation options through simulations – A new generation of urban planning tools. Journal of Hydrology,550(),355-367,0022-1694

Lucas, B., Kanade, T. (1981). An Iterative Image Registration Technique with an Application to Stereo Vision (IJCAI),81.

Maidment D. R. (1993). Handbook of Hydrology. McGraw-Hill.

Marzban, C., Scott S. (2010).Optical Flow for Verification. Weather and Forecasting,25(),1479-1494,

Ochoa-Rodriguez, S., Wang, L. P., Willems, P., Onof, C. (2019). A Review of Radar-Rain Gauge Data Merging Methods and Their Potential for Urban Hydrological Applications. Water Resources Research,.

Ochoa-Rodriguez, S., Wang, L.P., Gires, A., Pina, R.D., Reinoso-Rondinel, R., Bruni, G., A., Gaitan, S., Cristiano, E., van Assel, J., Kroll, S., Murlà-Tuyls, D., Tisserand, B., Schertzer, D., Tchiguirinskaia, I, Onof, C., Willems, P., ten Veldhuis, M.C. (2015). Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation. Journal of Hydrology,531(),389-407,0022-1694.

Pebesma, E. J. (2004). Multivariable geostatistics in S: the gstat package. Computers and Geosciences,30(),683-691,

Pierre, The radar beam path with height, 2011

Pulkkinen, S., Nerini, D., Pérez Hortal, A. A., Velasco-Forero, C., Seed, A., Germann, U., Foresti, L. (2019). Pysteps: an open-source Python library for probabilistic precipitation nowcasting (v1.0). Geosci. Model Dev.,12,4185-4219,1991-9603

Quirmbach, M., Schultz, G. A. (2002). Comparison of rain gauge and radar data as input to an urban rainfall-runoff model. Water Science and Technology,45(),27-33,

Rabiei, E., Haberlandt, U. (2015). Applying bias correction for merging rain gauge and radar data. Journal of Hydrology,522(),544-557.

Russo, B., Sunyer, D., Velasco, M., Djordjević, S. (2015). Analysis of extreme flooding events through a calibrated 1D/2D coupled model: The case of Barcelona (Spain). Journal of Hydroinformatics,17(),473.

Ruzanski, E., Chandrasekar, V., Yanting W. (2011). The CASA Nowcasting System. Journal of Atmospheric and Oceanic Technology,28(),640-655.

Sauvageot, H.(1992).Radar Hydrology. Artech House.

Schellart, A., Liguori, S., Krämer, S., Saul, A., Rico-Ramirez, M. A. (2014). Comparing quantitative precipitation forecast methods for prediction of sewer flows in a small urban area. Hydrological Sciences Journal, 59,1418-1436.

Schellart, A., Shepherd, W.J., Saul, A.J. (2012). Influence of rainfall estimation error and spatial variability on sewer flow prediction at a small urban scale. Advances in Water Resources, 45, 65-75.

Schilling, W. (1991). Rainfall data for urban hydrology: what do we need? Atmospheric Research, 27, 5-21.

Seed, A., Siriwardena, L., Sun, X., Jordan, P., Elliot, J. (2002). On the calibration of the Australian weather radars. Cooperative Research Centre for Catchment Hydrology.

Seo, D. J. (1997). Rainfall estimation in the Vexrad Era-Operational Experience, Issues, and Ongoing Efforts in the U.S National Weather Service.

Seo, D. J., Breidenbach, J. P., Johnson, E. R. (1999). Real-time estimation of mean field bias in radar rainfall data. Journal of Hydrology, 223, 131-147.

Seo, D. J., Breidenbach, J. P. (2002). Real-time correction of spatially non-uniform bias in radar rainfall data using rain gauge measurements. Journal of Hydrometeorology, 3, 93-111.

Sharif, H., Ogden, F., Krajewski, W. F., Xue, M., (2001). Using Radar Rainfall Estimates at Ground Level: Beam, Storm, and Watershed Geometric Interaction.

Sharma, A., Mehrotra, R., Li, J., Jha, S. (2016). A programming tool for non-parametric system prediction using Partial Informational Correlation and Partial Weights. Environmental Modelling & Software, 83, 271-275.

Steiner, M., Smith, J. A., Burges, S. J., Alonso, C. V., Darden, R. W. (1999). Effect of bias adjustment and rain gauge data quality control on radar rainfall estimation. Water Resources Research, 35, 2487-2503.

Surcel, M., Zawadzki, I., Yau, M.K. (2015).A Study on the Scale Dependence of the Predictability of Precipitation Patterns. Journal of the Atmospheric Sciences, 72, 216-235.

Thorndahl, S., Nielsen, J., Jensen, D., (2016). Urban pluvial flood prediction: A case study evaluating radar rainfall nowcasts and numerical weather prediction models as model inputs. Water Science and Technology, 74.

Tonidandel, S., LeBreton, J.M. (2011). Relative importance analysis: A useful supplement to regression analysis. Journal of Business and Psychology, 26(01), 1-9.

Tonidandel, S., LeBreton, J.M., Johnson, J.W. (2009). Determining the statistical significance of relative weights. Phycological Methods, 14, 387–399.

United Nations(2018). 2018 Revision of World Urbanization.

van Dijk, E., van der Meulen, J., Kluck, J., Straatman, J. H. M. (2013). Comparing modelling techniques for analysing urban pluvial flooding. Water Science and Technology, 69, 305-311.

Vasiloff, S. V., Howard, K. W., Zhang, J. (2009). Difficulties with correcting radar rainfall estimates based on rain gauge data: A case study of severe weather in Montana on 16-17 June 2007. Weather and Forecasting, 24, 1334-1344.

Wallace, J.M., Hobbs, P.V. (1977). Atmospheric Science: An Introductory Survey.

Wang, L.P., Ochoa-Rodriguez, S., Simões, N., Onof, C., Maksimovic, Č. (2013). Radar-raingauge data combination techniques: A revision and analysis of their suitability for urban hydrology. Water science and technology: a journal of the International Association on Water Pollution Research, 68, 737-747.

Worldbank (2019). World Urbanization Prospects: 2018 Revision.

Wu, Qi., Wang, F. (2019). Concatenate Convolutional Neural Networks for Non-Intrusive Load Monitoring across Complex Background. Energies, 12, 1572.

Zahraei, A., Hsu, K., Sorooshian, S., Gourley, J. J., Lakshmanan, V., Hong, Y., Bellerby, T, (2012). Quantitative Precipitation Nowcasting: A Lagrangian Pixel-Based Approach. Atmospheric Research,118(),418-434,0169-8095

Zhu, A., Yuan, L., Chaney, K., Daniilidis, K. (2018). EV-FlowNet: Self-Supervised Optical Flow Estimation for Event-based Cameras.

Šálek, M., Brezková, L., Novák, P. (2006). The use of radar in hydrological modelling in the Czech Republic – case studies of flash floods. Nat. Hazards Earth Syst. Sci, 6, 229-236.

APPENDIX A

Information about the 110 events chosen for the training and validation of the nowcast methods. First are listed the 93 events for the period 2000-2012 that have been used for the training of the nowcast methods, and then the 17 extreme events of the period 2013-2018 that have been used for the validation of the developed methods for the purpose of the urban flood application.

Table A.1: *General Information about the 110 events investigated in this thesis.*

Start Time [UTC]	End Time [UTC]	Duration [ts]	PImax [mm/5min]	WR >45%	Storm Duration [ts]	Storm Type
2001/6/30/8/20	2001/6/30/22/5	166	10.7	53	24	Meso-γ
2002/7/10/13/40	2002/7/10/19/55	76	7.37	42	26	Meso-β
2002/7/16/21/25	2002/7/18/18/5	537	7.9	335	138	Mixed
2003/6/3/13/0	2003/6/3/20/55	96	7.1	7	14	Meso-β
2003/6/8/10/30	2003/6/8/18/55	102	10.5	58	34	meso-γ
2003/7/21/22/45	2003/7/22/7/15	103	5.8	20	16	Meso-β
2003/8/18/11/25	2003/8/18/20/20	108	7	5	9	Meso-β
2003/9/10/13/35	2003/9/11/16/5	319	1.5	209	13	Meso-β
2004/7/18/12/10	2004/7/18/17/55	70	12.6	15	13	Meso-γ
2004/7/21/16/30	2004/7/21/18/25	24	12.87	1	8	Meso-γ
2005/6/21/14/5	2005/6/21/19/20	64	5.9	0	7	Meso-β
2005/7/20/12/50	2005/7/20/18/5	64	3.5	7	8	Meso-β
2005/7/29/14/30	2005/7/30/10/5	236	17.88	33	20	Meso-γ
2005/8/14/4/25	2005/8/15/1/40	256	6.2	128	41	Meso-β
2005/8/20/6/45	2005/8/20/12/25	69	5.25	4	22	Meso-β
2006/6/15/23/50	2006/6/16/8/35	106	12.8	0	1	Meso-γ
2006/6/19/11/35	2006/6/19/20/35	109	13.48	0	62	Mixed
2006/6/25/15/45	2006/6/25/23/50	98	12.49	34	2	Meso-γ
2006/7/7/11/15	2006/7/7/23/15	145	11.86	0	25	Convective
2006/7/20/14/45	2006/7/20/20/5	65	7.2	30	33	Meso-β
2006/7/27/10/0	2006/7/27/17/25	90	15.2	0	16	Meso-γ
2006/8/1/13/40	2006/8/1/20/45	86	8.92	0	43	Meso-β
2006/8/14/11/25	2006/8/15/4/45	209	2.88	145	63	Stratiform
2006/8/18/7/0	2006/8/18/15/15	100	7.73	63	40	Meso-β
2006/8/26/4/10	2006/8/26/15/25	136	6.95	1	22	Meso-β
2007/1/18/2/15	2007/1/18/22/5	239	7.48	238	191	Mixed
2007/5/7/6/40	2007/5/8/4/10	259	5.6	177	223	Stratiform
2007/5/25/5/40	2007/5/25/16/20	129	10.46	14	58	Mixed
2007/5/29/3/45	2007/5/29/19/30	190	1.85	180	190	Stratiform

2007/6/9/15/10	2007/6/9/21/40	79	8.48	0	13	Convective
2007/6/12/13/30	2007/6/12/15/30	25	8.54	0	1	Convective
2007/6/14/14/40	2007/6/15/0/50	123	12.58	27	18	Meso-γ
2007/6/15/10/20	2007/6/16/4/30	219	14.6	101	57	Mixed
2007/6/21/4/45	2007/6/21/18/15	163	6.43	134	123	Stratiform
2007/8/9/14/45	2007/8/9/21/5	77	10.85	0	77	Mixed
2007/8/10/11/50	2007/8/10/21/50	121	8.84	0	42	Convective
2007/8/15/20/55	2007/8/16/1/30	56	9.19	34	49	Mixed
2007/8/21/5/30	2007/8/21/19/50	173	5.24	104	158	Stratiform
2007/8/24/1/25	2007/8/24/5/15	47	7.23	0	47	Meso-β
2007/9/28/15/0	2007/9/29/20/35	356	6.93	317	236	Stratiform
2007/12/23/18/45	2007/12/24/9/30	178	19.91	0	9	Convective
2008/6/3/13/5	2008/6/3/23/35	127	11	0	39	Convective
2008/6/22/13/55	2008/6/22/21/15	89	15.93	26	56	Mixed
2008/7/3/12/15	2008/7/4/9/35	257	13.55	166	209	Mixed
2008/7/10/16/10	2008/7/11/2/45	128	15.47	1	15	Meso-γ
2008/7/29/12/25	2008/7/29/18/15	71	15.78	27	71	Mixed
2008/8/1/7/55	2008/8/1/23/30	188	13.17	0	54	Mixed
2008/8/3/20/50	2008/8/4/4/35	94	8.58	53	92	Mixed
2008/8/7/14/30	2008/8/7/17/20	35	11.15	0	11	Convective
2008/8/22/11/30	2008/8/23/3/55	198	11.3	155	198	Mixed
2008/9/30/11/20	2008/10/1/15/20	337	14.12	227	132	Mixed
2009/7/3/15/45	2009/7/3/23/0	88	11.07	0	52	Mixed
2009/7/17/22/50	2009/7/18/4/55	74	3.62	63	74	Stratiform
2009/7/21/18/15	2009/7/22/0/20	74	7.64	23	21	Meso-β
2009/7/23/8/35	2009/7/23/15/15	81	19.06	6	36	Meso-γ
2009/8/8/12/55	2009/8/8/17/45	59	11.93	0	53	Mixed
2009/10/7/10/0	2009/10/8/7/50	263	13.67	131	94	Mixed
2010/5/30/9/25	2010/5/30/19/55	127	4.5	89	127	Stratiform
2010/6/9/13/55	2010/6/10/4/25	175	10.1	66	93	Mixed
2010/7/3/16/25	2010/7/3/18/15	23	8.56	0	23	Convective
2010/7/9/23/0	2010/7/10/22/55	288	0.78	0	3	Convective
2010/7/16/20/50	2010/7/17/0/35	46	11.52	0	13	Meso-γ
2010/7/26/10/50	2010/7/26/16/10	65	10.09	42	65	Mixed
2010/8/1/14/55	2010/8/1/20/20	66	14.48	2	59	Mixed
2010/8/12/3/15	2010/8/12/13/25	123	6.24	0	36	Meso-β
2010/8/26/2/15	2010/8/27/16/40	462	8.11	309	1	Meso-β
2010/9/12/14/30	2010/9/12/20/55	78	6.61	41	76	Stratiform
2010/9/25/23/0	2010/9/26/22/55	288	1	133	242	Stratiform
2010/11/12/21/5	2010/11/13/13/30	198	4.12	46	48	Meso-β
2011/6/5/15/30	2011/6/5/19/35	50	17.64	0	17	Convective

2011/6/22/11/40	2011/6/22/19/10	91	14.14	31	58	Mixed
2011/6/29/13/10	2011/6/30/4/30	185	13.24	104	56	Mixed
2011/8/3/23/0	2011/8/4/22/55	288	3.82	0	38	Convective
2011/8/12/10/10	2011/8/12/17/45	92	7.75	1	47	Meso-β
2011/8/18/18/25	2011/8/19/9/40	184	8.6	29	36	Meso-β
2011/8/23/2/10	2011/8/23/13/50	141	8.96	14	52	Mixed
2011/8/24/17/20	2011/8/24/22/25	62	13.02	0	53	Mixed
2011/8/26/13/20	2011/8/27/1/25	146	9.13	40	39	Meso-β
2011/9/4/13/0	2011/9/4/23/0	121	11.71	0	31	Meso-γ
2011/9/11/13/15	2011/9/11/18/55	69	6.17	52	69	Stratiform
2012/5/2/16/25	2012/5/3/1/0	104	8.4	0	70	Mixed
2012/5/10/19/20	2012/5/10/23/35	52	6.49	0	31	Meso-β
2012/5/31/13/15	2012/5/31/21/25	99	5.21	98	90	Stratiform
2012/6/29/12/40	2012/6/29/17/5	54	9.99	31	42	Meso-β
2012/6/30/22/40	2012/7/1/1/5	30	10.14	0	30	Meso-γ
2012/7/5/16/5	2012/7/6/2/55	131	10	0	95	Mixed
2012/7/7/11/15	2012/7/7/16/45	67	11.4	0	15	Convective
2012/7/19/4/45	2012/7/19/15/20	128	6.9	96	98	Stratiform
2012/7/27/16/5	2012/7/28/7/0	180	16.8	0	10	Convective
2012/8/6/0/0	2012/8/6/4/25	54	8.85	19	23	Meso-β
2012/8/24/22/45	2012/8/25/1/45	37	7.64	19	33	Meso-β
2012/8/31/12/25	2012/8/31/16/35	51	4.8	0	20	Convective
2012/9/11/12/20	2012/9/11/18/15	72	10.63	3	55	Mixed
2014/7/29/15/50	2014/7/30/4/30	153	8.4	20	31	Meso-γ
2015/8/14/15/0	2015/8/15/1/10	123	8.4	0	25	Mixed
2015/8/16/9/30	2015/8/17/0/55	186	13.12	46	57	Mixed
2016/7/22/11/35	2016/7/22/15/40	50	12.17	0	28	Meso-γ
2017/5/18/12/10	2017/5/18/17/55	70	10.49	0	28	Meso-γ
2017/6/3/16/25	2017/6/3/23/55	178	10.97	2	43	Meso-γ
2017/6/22/9/45	2017/6/22/21/20	141	13.68	7	28	Meso-γ
2017/7/10/3/05	2017/7/10/14/15	136	8.36	22	71	Mixed
2014/6/10/19/10	2014/6/10/21/50	34	9.78	0	19	Convective
2014/7/8/15/0	2014/7/9/19/55	349	10.84	6	90	Mixed
2014/8/5/2/25	2014/8/5/9/55	92	7.56	0	91	Mixed
2014/9/19/15/30	2014/9/19/18/55	43	10.95	0	26	Convective
2016/5/28/14/15	2016/5/28/15/30	17	16.31	0	6	Convective
2016/6/24/13/10	2016/6/24/17/30	54	12.03	0	19	Meso-γ
2016/7/21/14/05	2016/7/21/19/10	63	14.67	0	14	Convective
2017/7/19/16/30	2017/7/19/21/45	65	8.27	19	53	Mixed
2018/7/28/11/40	2018/7/28/21/25	119	16.92	0	36	Meso-γ

APPENDIX B

The performance criteria with raw radar input for the high intensity thresholds of existing nowcast models a) HyRaTrac Figure B.1 -B.2, and b) Lucas-Kanade Figure B.3- B.4.
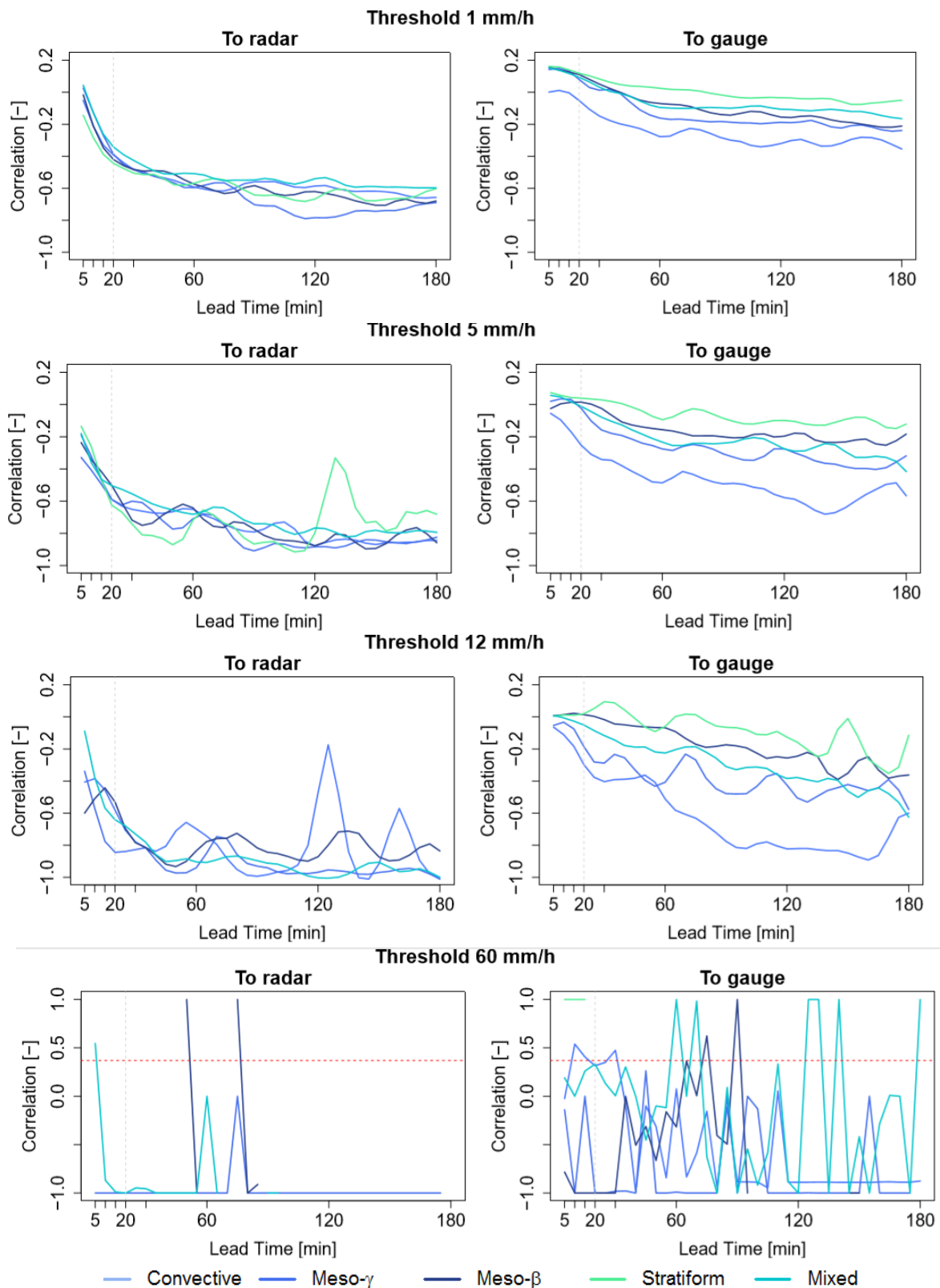
Figure B.1: *The event median temporal correlation for each event type as nowcasted by HyRaTrac on raw radar data for different lead times and different thresholds.*
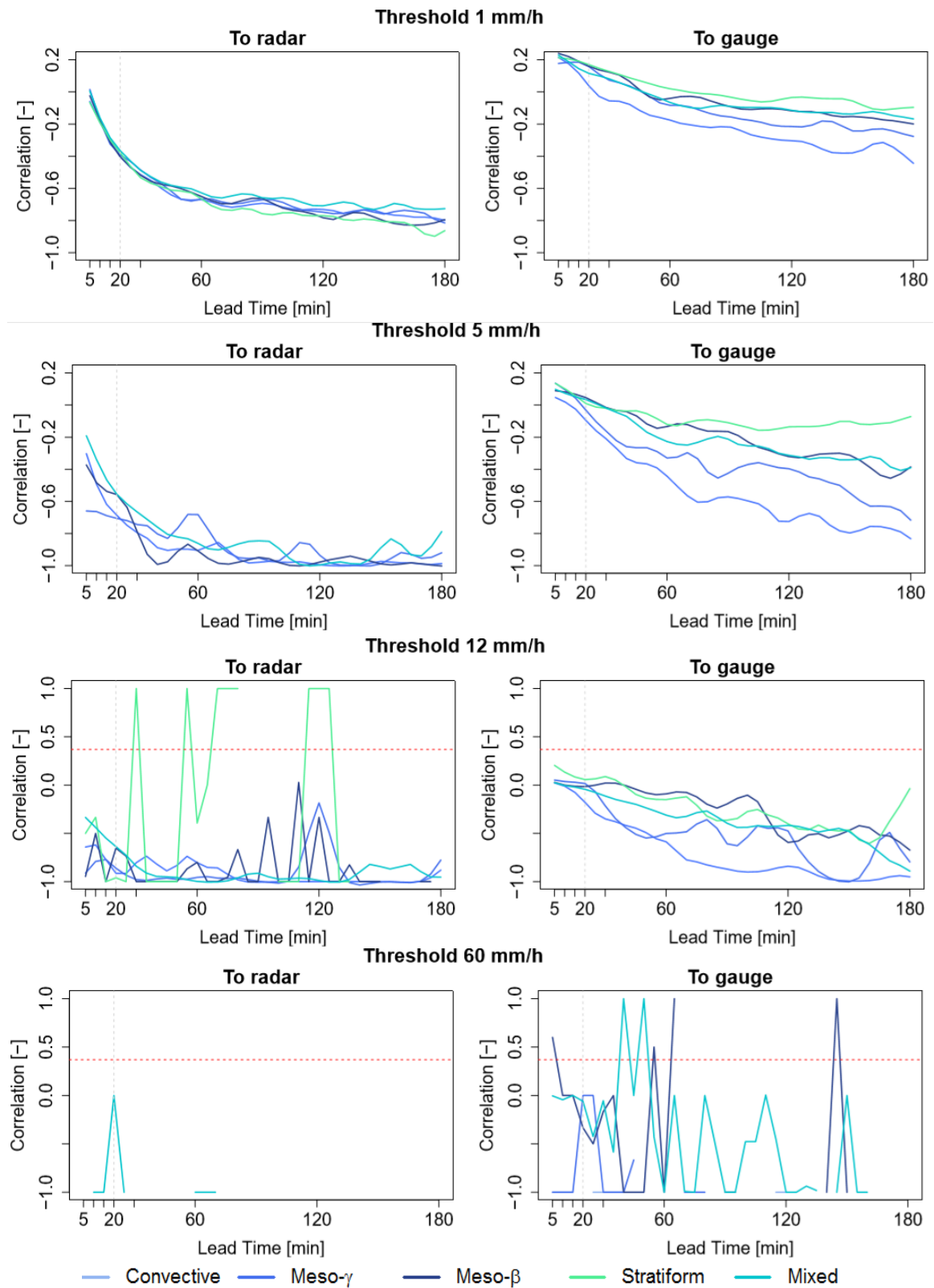
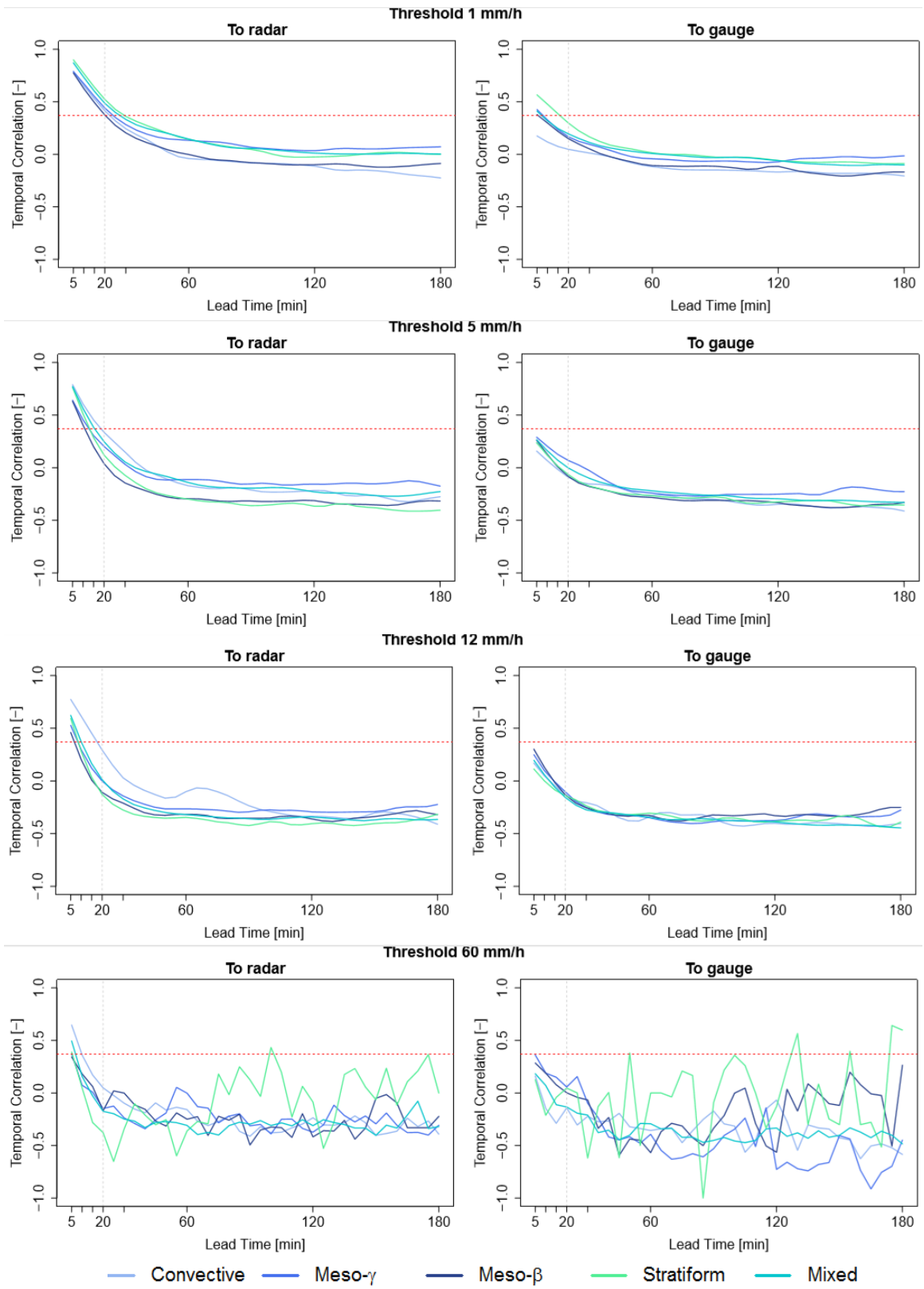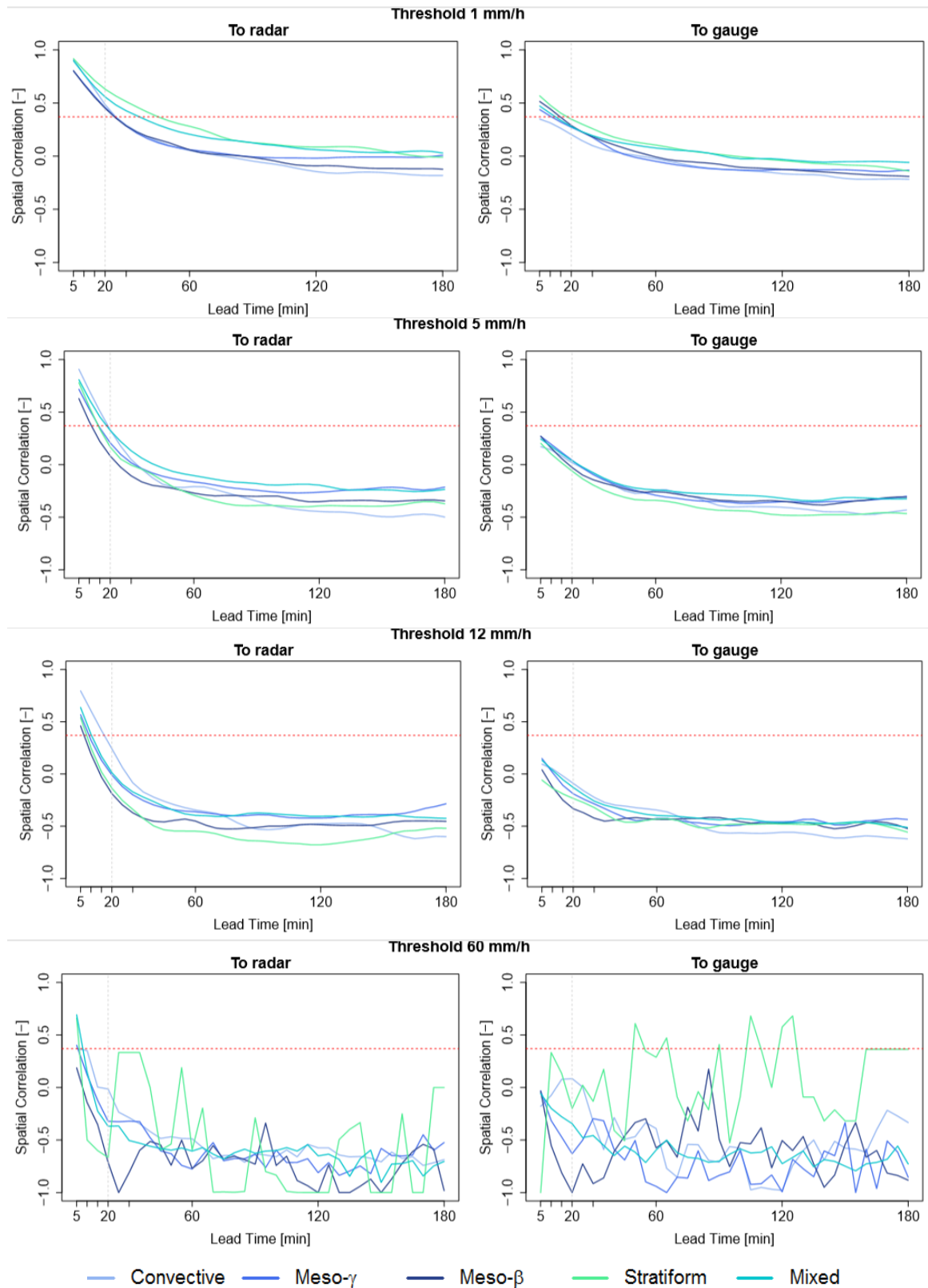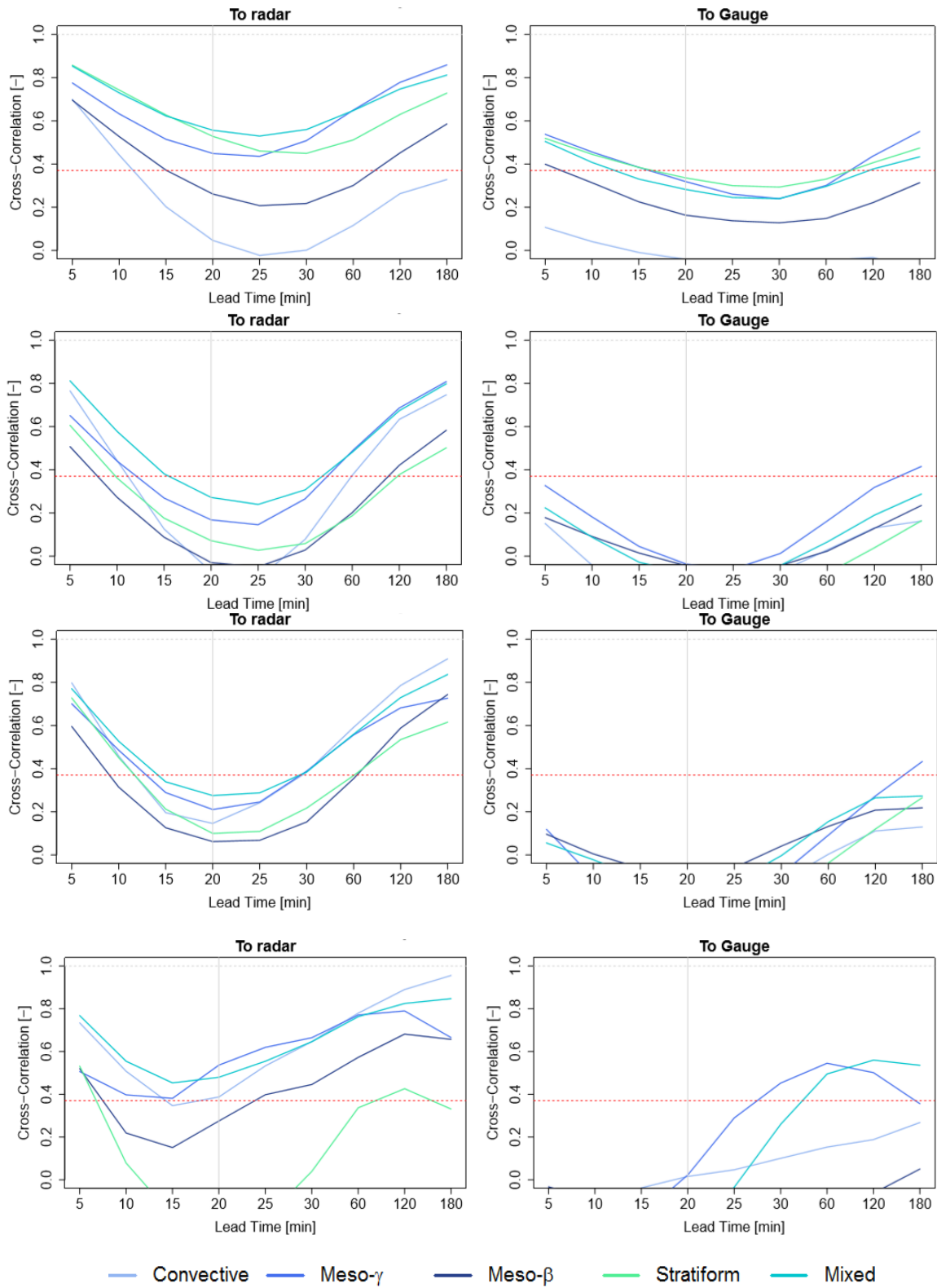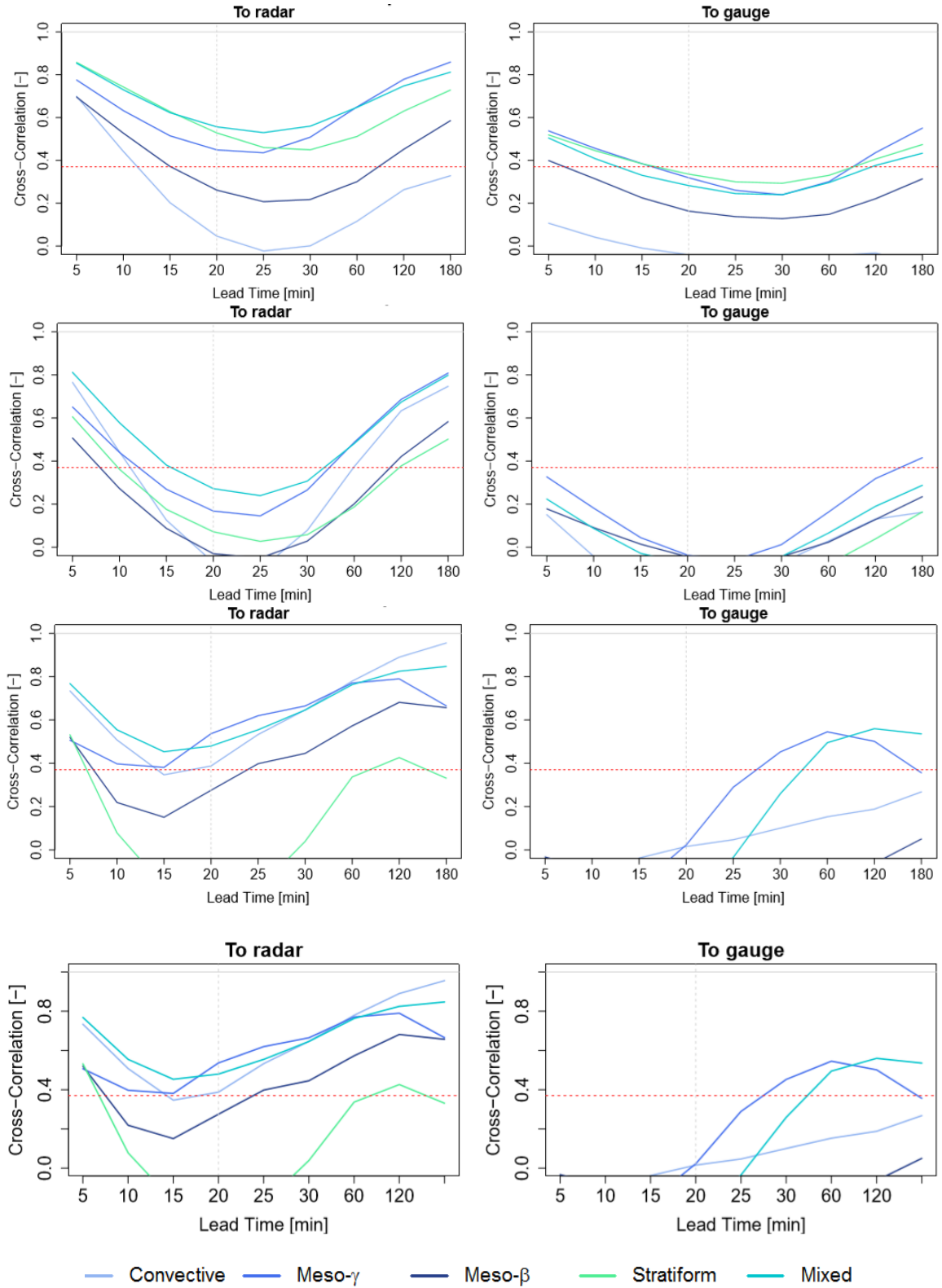**Figure B.2:** *The event median spatial cross-correlation for each event type as nowcasted by HyRaTrac on raw radar data for different lead times and different thresholds.*
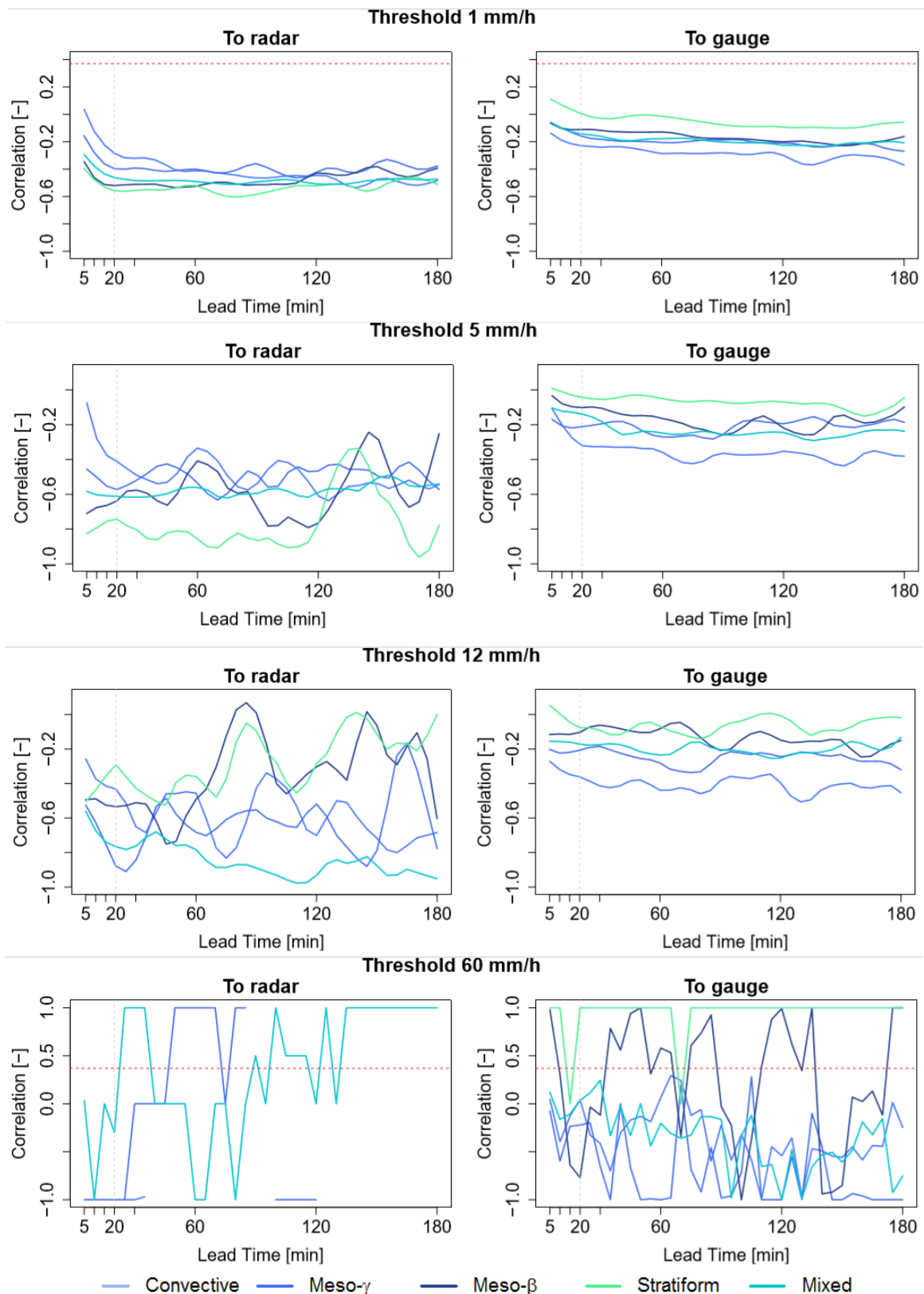
Figure B.3: *The event median temporal correlation for each event type as nowcasted by LK on raw radar data for different lead times and different thresholds.*
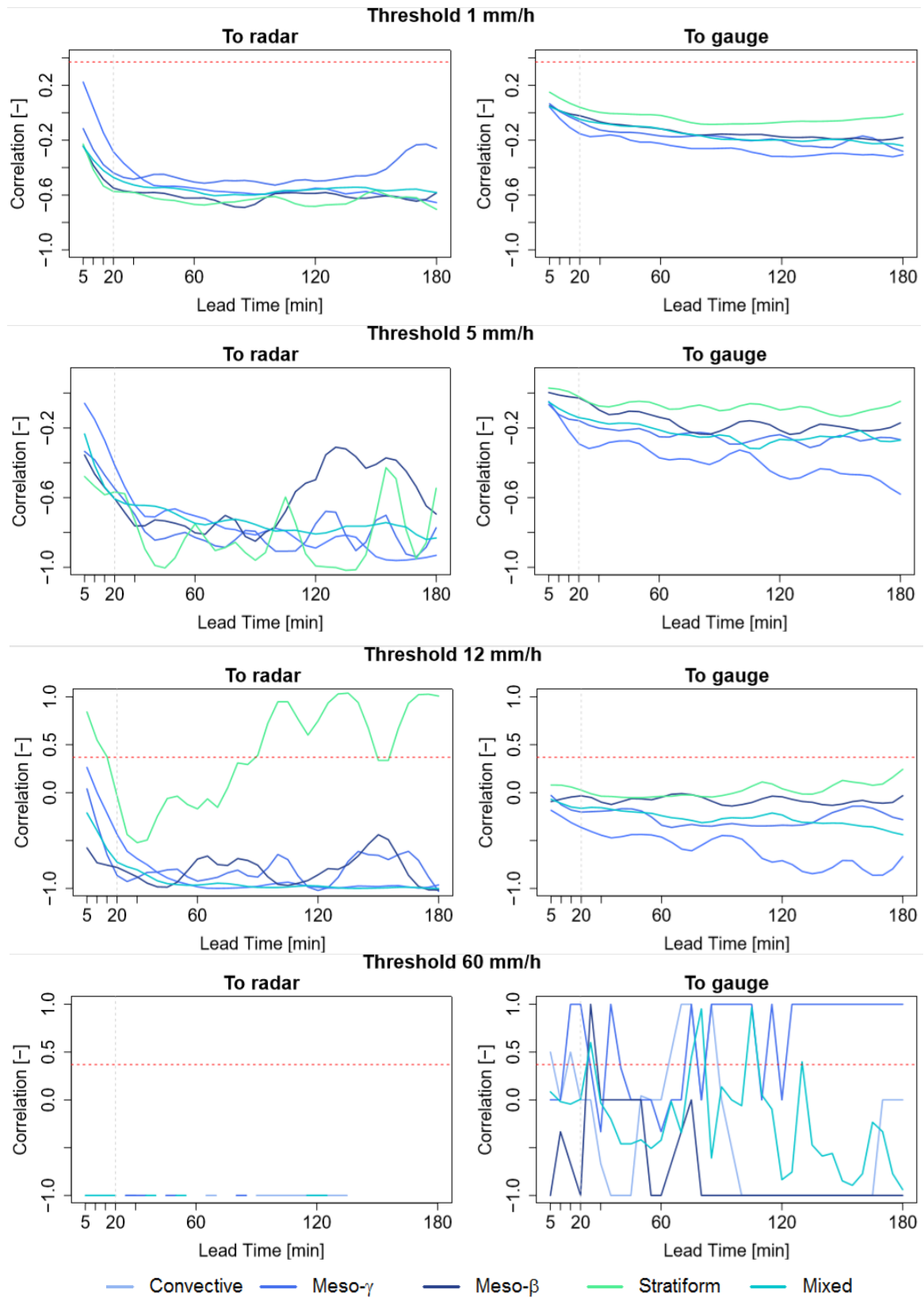
Figure B.4: *The event median spatial cross-correlation for each event type as nowcasted by LK on raw radar data for different lead times and different thresholds.*

The performance criteria with merged radar input for the high intensity thresholds of existing nowcast models a) HyRaTrac Figure B.5 -B.6, and b) Lucas-Kanade Figure B.7- B.8.
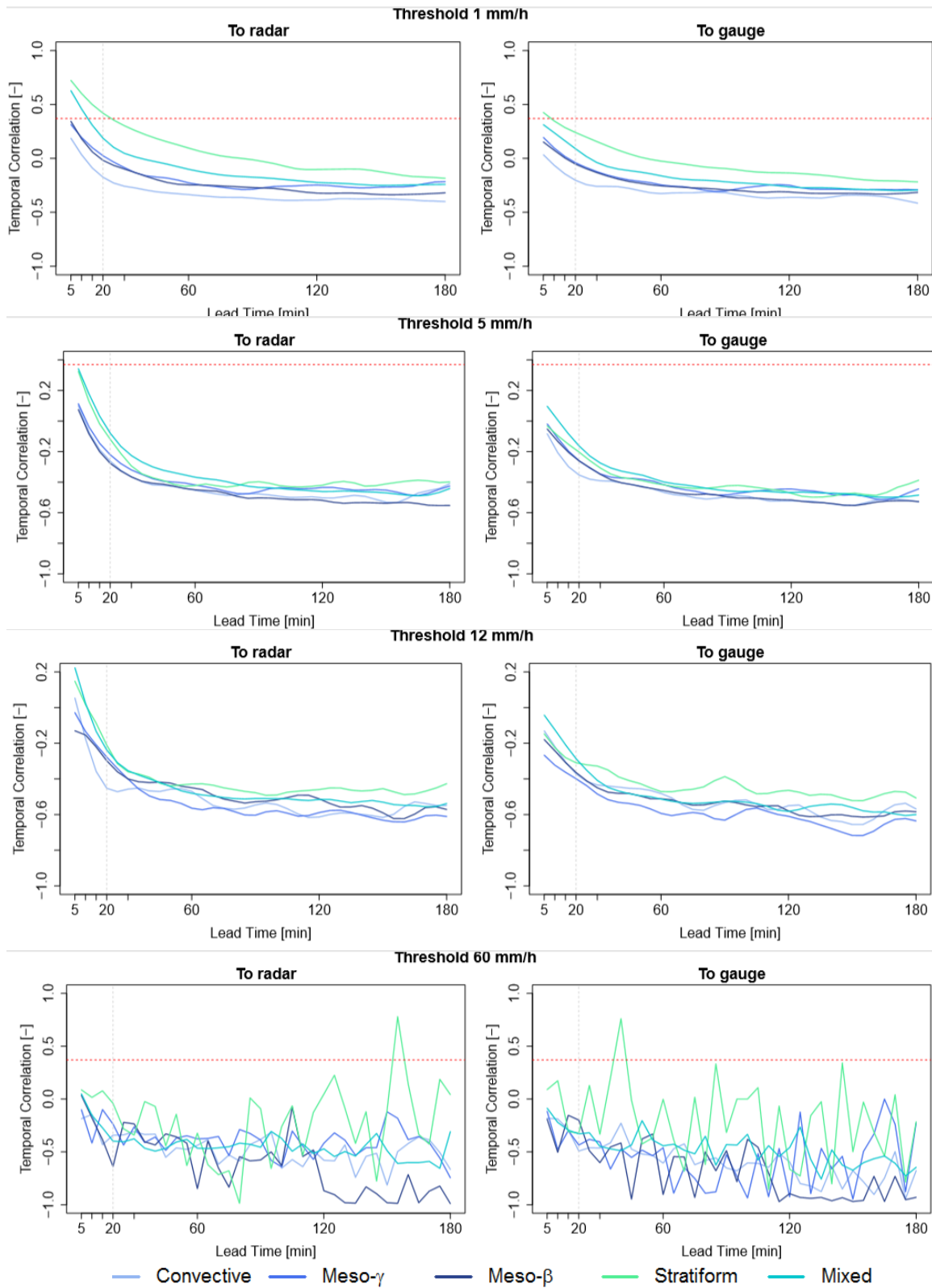
Figure B.5: *The event median temporal correlation for each event type as nowcasted by HyRaTrac on merged radar data for different lead times and different thresholds.*
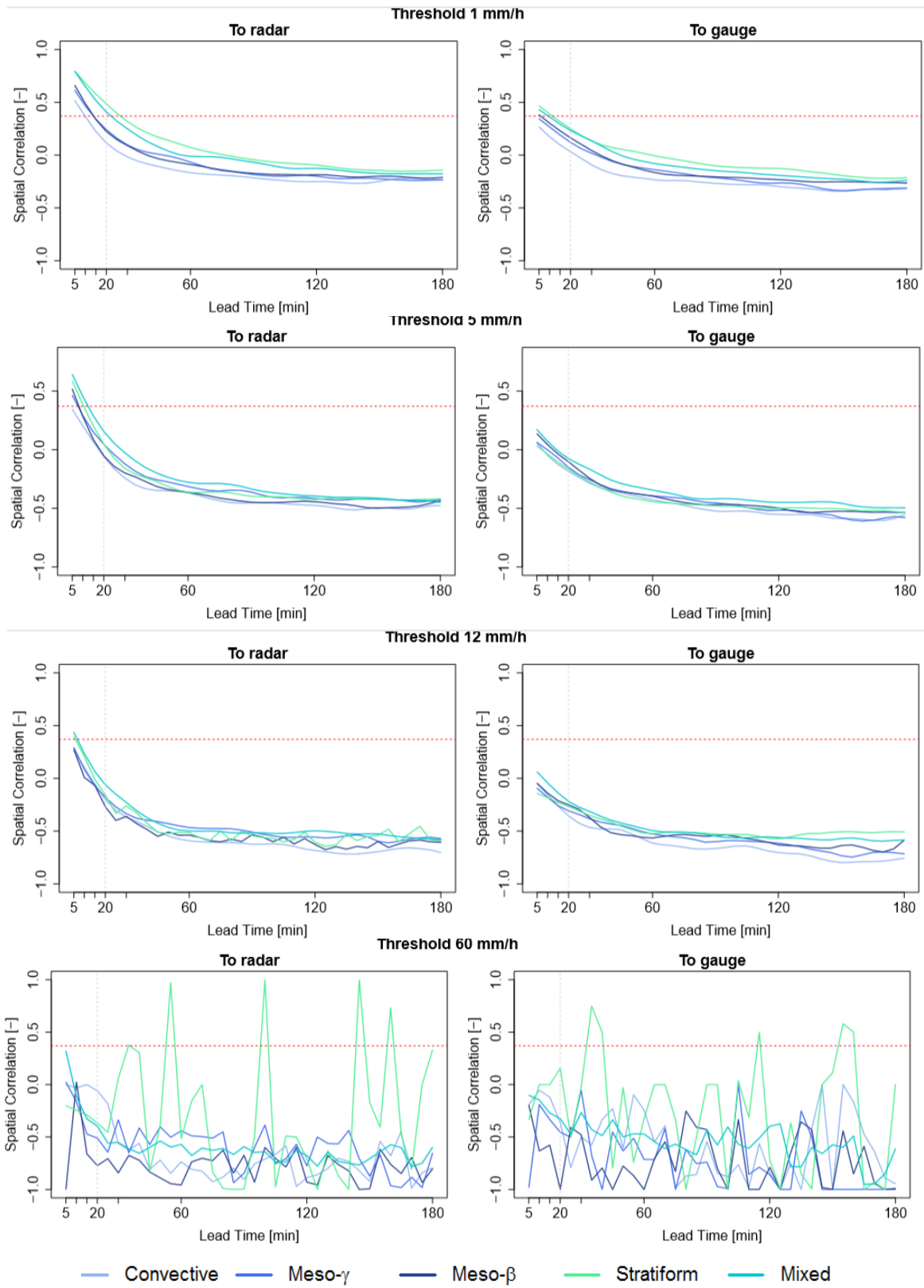
Figure B.6: *The event median spatial cross-correlation for each event type as nowcasted by HyRaTrac on merged radar data for different lead times and different thresholds.*

Figure B.7: *The event median temporal correlation for each event type as nowcasted by LK on merged radar data for different lead times and different thresholds.*

Figure B.8: *The event median spatial cross-correlation for each event type as nowcasted by LK on merged radar data for different lead times and different thresholds.*

The performance criteria with merged radar input for the high intensity thresholds of new CNN nowcast model (Figure B.9 to Figure B.10)

Figure B.9: *The event median temporal correlation for each event type as nowcasted by CNN on merged radar data for different lead times and different thresholds.*

Figure B.10: *The event median spatial cross-correlation for each event type as nowcasted by CNN on merged radar data for different lead times and different thresholds.*

APPENDIX C

The performance criteria of all the merging methods for the high intensity thresholds for the 93 events (Figure C.1 to Figure C.2).
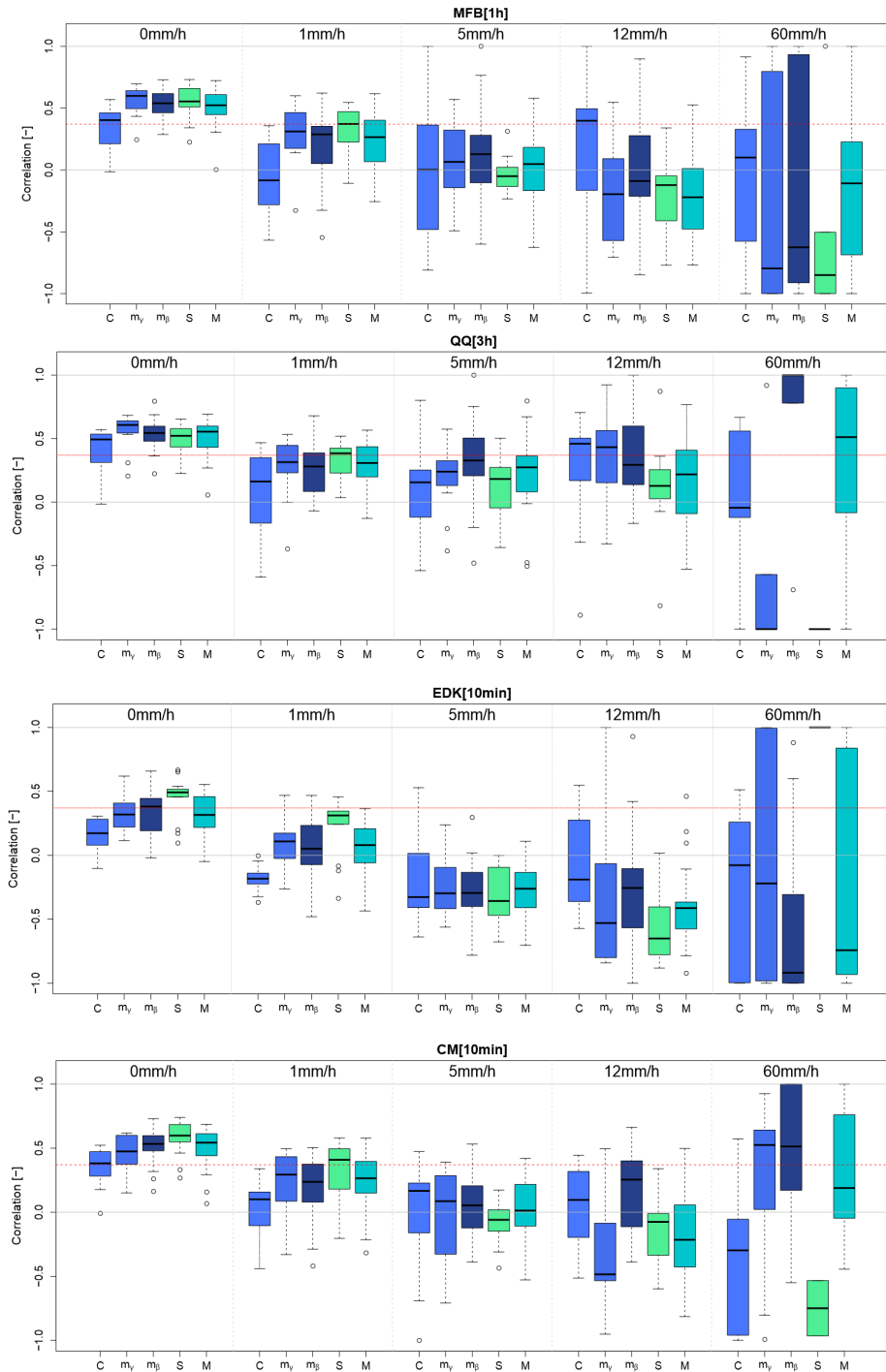
Figure C.1: *The Temporal Cross-Correlation of the best-implemented merging methods for the 5 types of the events. The red dashed line illustrated the predictability limit value as shown by the literature.*
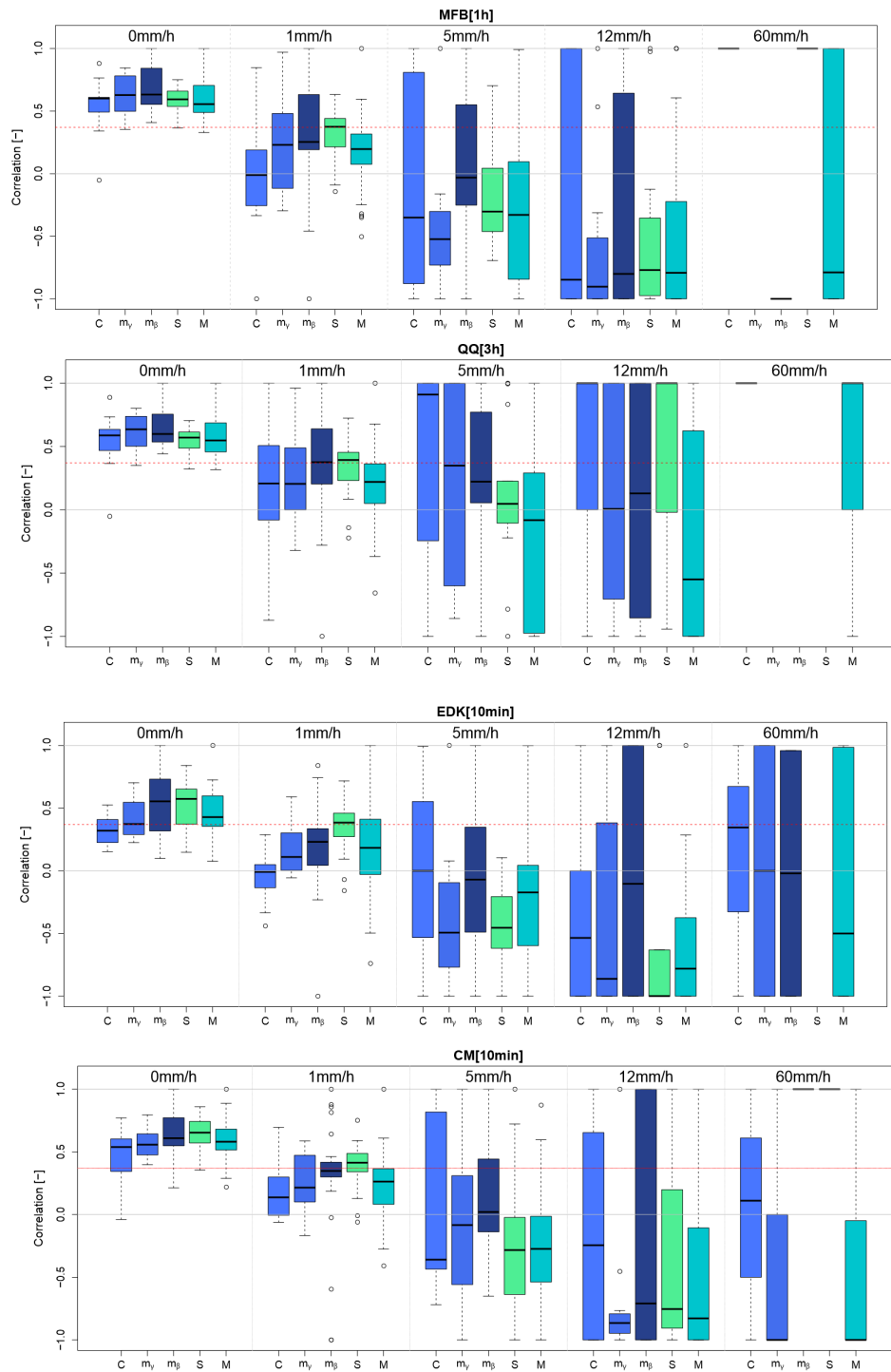
Figure C.2: *The Spatial Cross-Correlation of the best-implemented merging methods for the 5 types of the events. The red dashed line illustrated the predictability limit value as shown by the literature.*
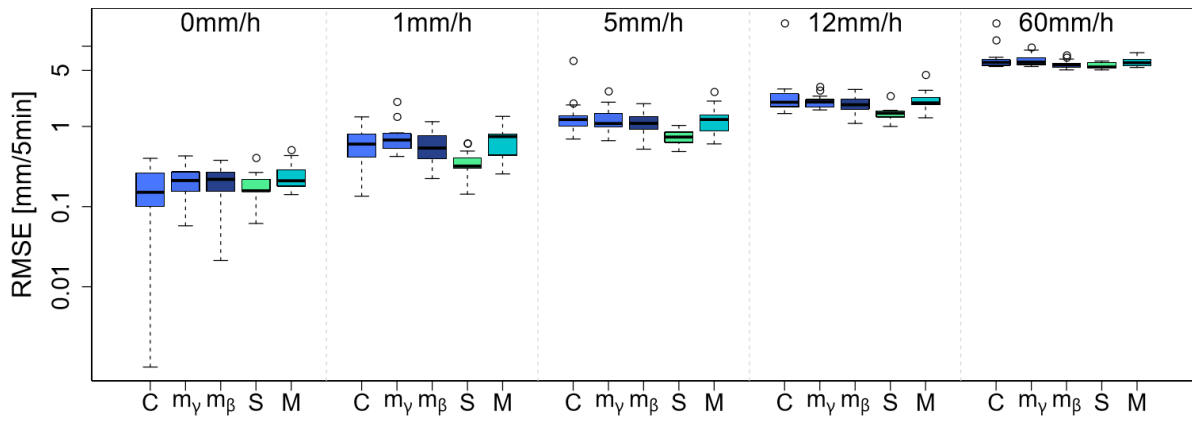
Figure C.3: *The Spatial RMSE of the raw radar data for the 5 types of the events and for different intensity thresholds.*

APPENDIX D

The improvements due to the merged radar data as an input to the existing nowcast models for high intensity thresholds.
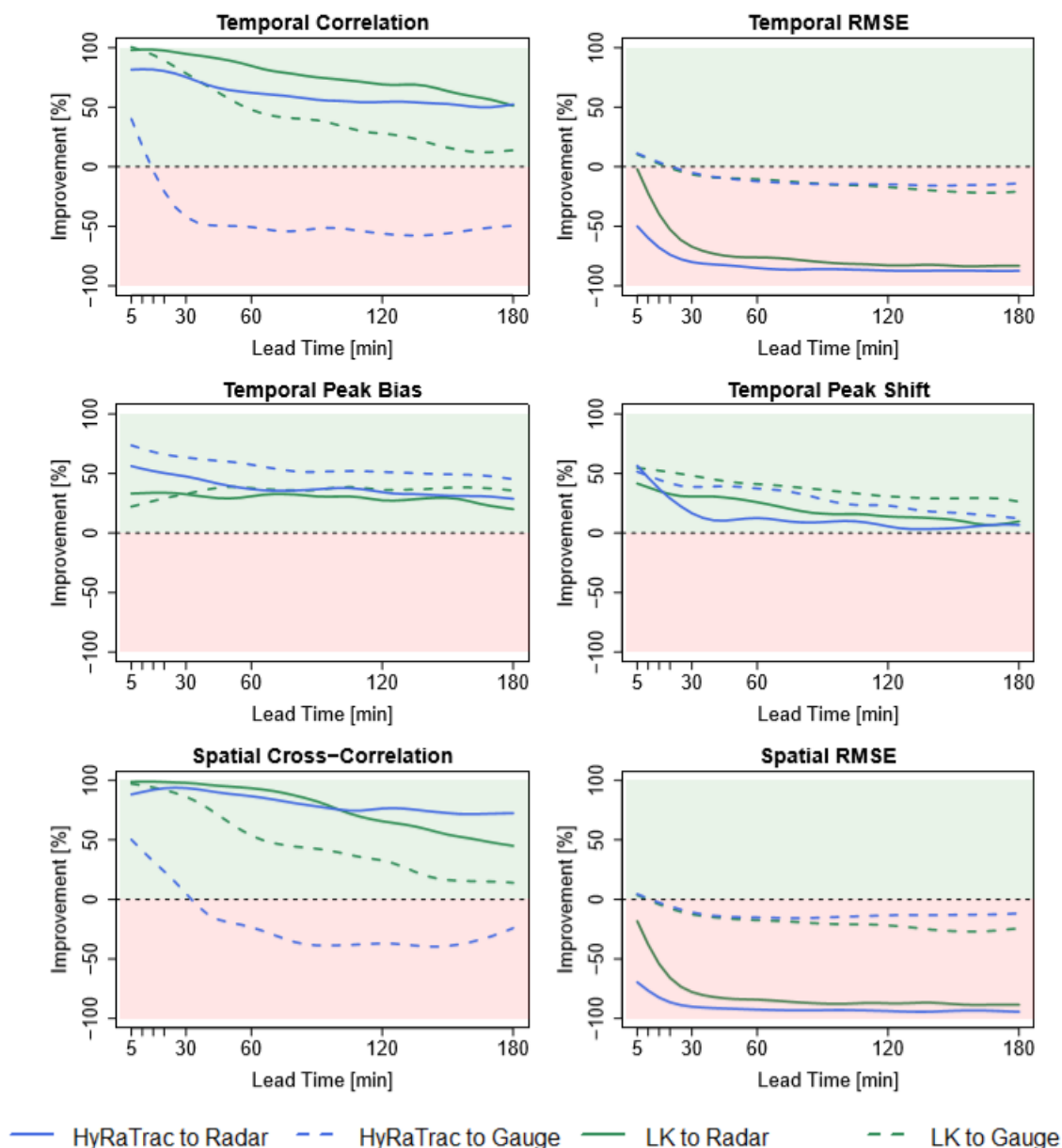
Figure D.1: *The average improvement of spatial and temporal criteria for each lead time when using CM data instead of RR as Input for both HyRaTrac and LK nowcast methods. Here an intensity threshold of 1 mm/h was used.*
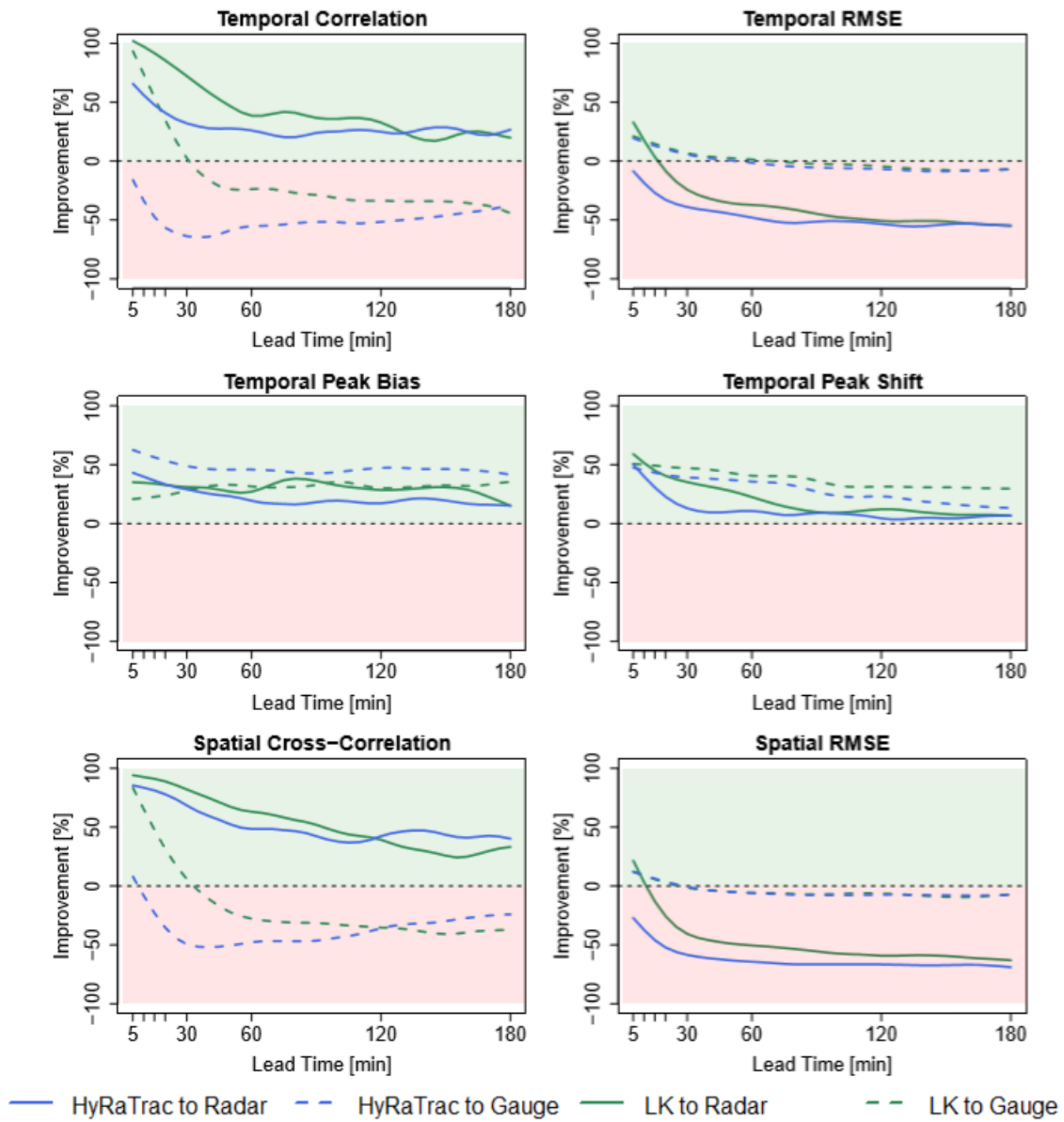
Figure D.2: *The average improvement of spatial and temporal criteria for each lead time when using CM data instead of RR as Input for both HyRaTrac and LK nowcast methods. Here an intensity threshold of 5 mm/h was used.*
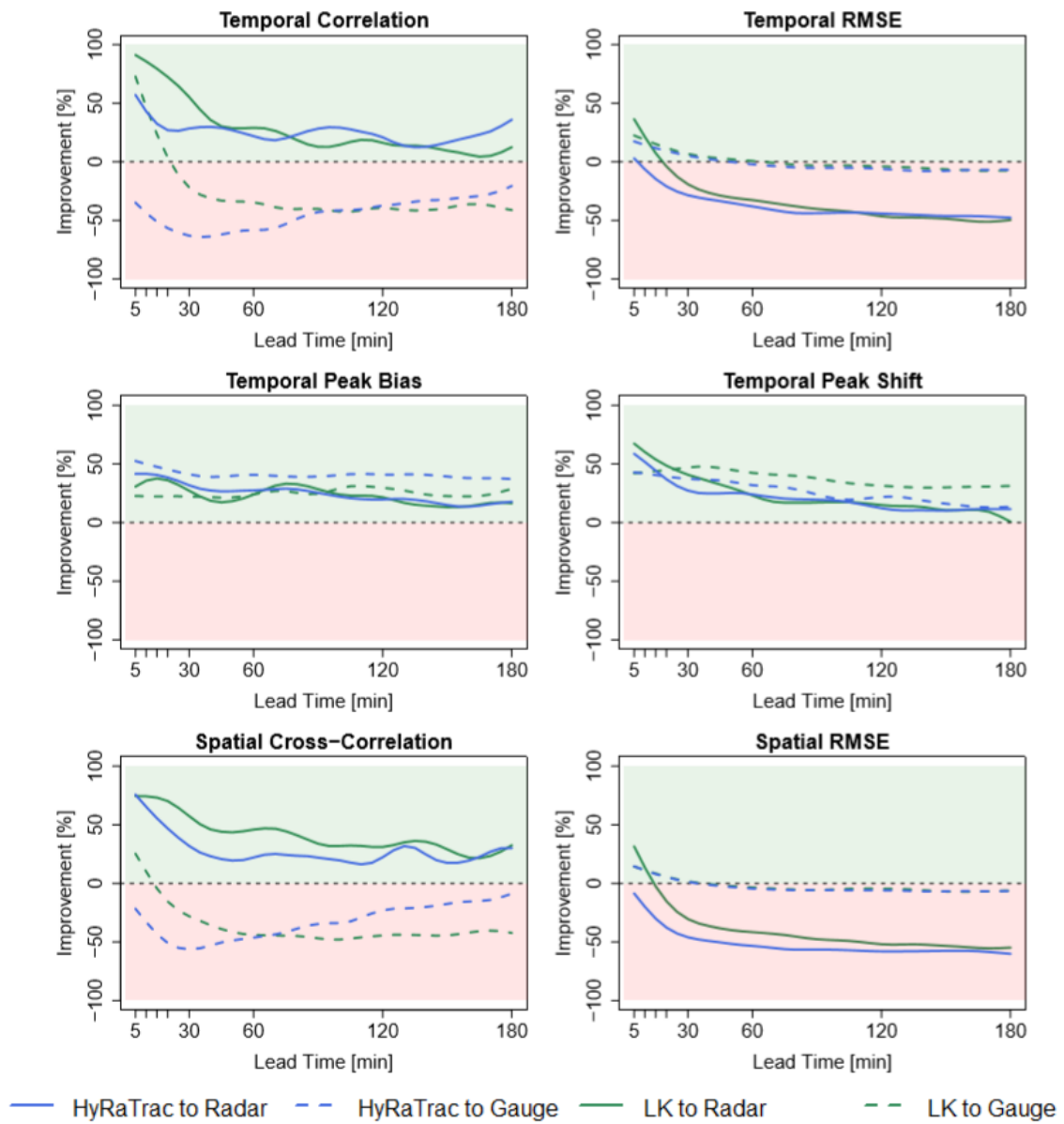
Figure D.3: *The average improvement of spatial and temporal criteria for each lead time when using CM data instead of RR as Input for both HyRaTrac and LK nowcast methods. Here an intensity threshold of 12 mm/h was used.*
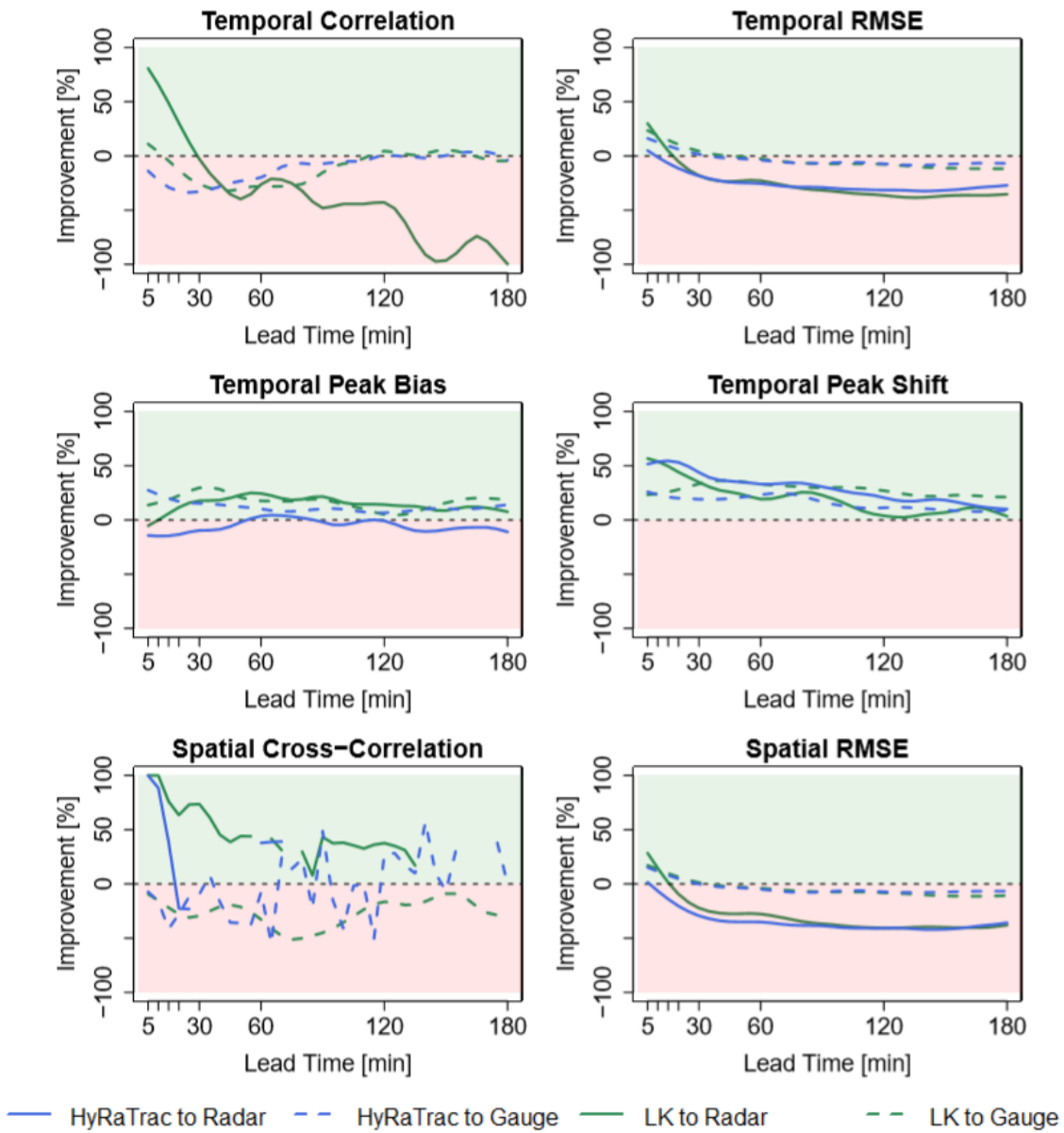
Figure D.4: The average improvement of spatial and temporal criteria for each lead time when using CM data instead of RR as Input for both HyRaTrac and LK nowcast methods. Here an intensity threshold of 60 mm/h was used.

The improvements of the nowcast skill due to the use of the CNN model towards LK nowcast approach for high intensity thresholds.
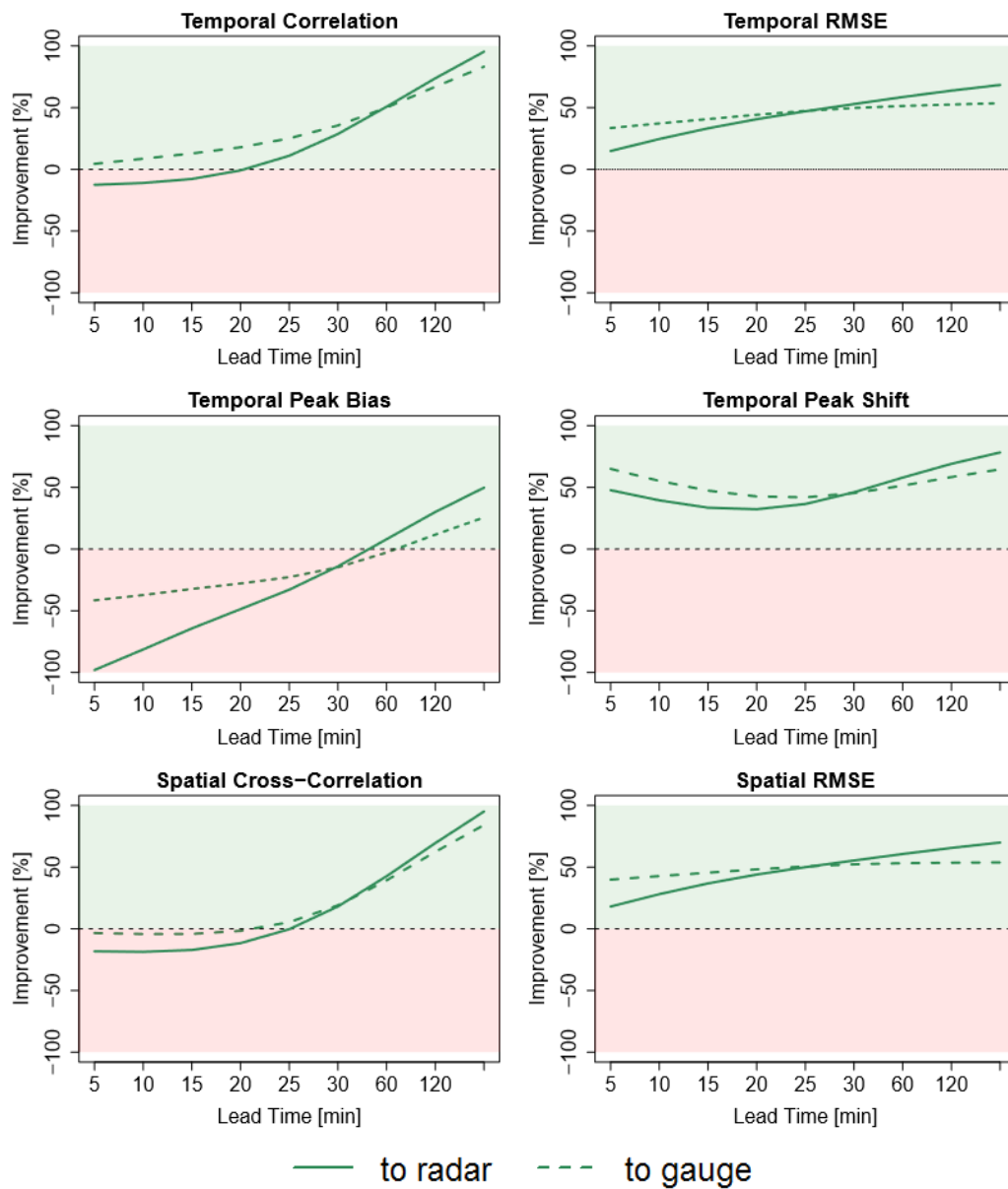
Figure D.5: *The average improvement of spatial and temporal criteria for each lead time when using CNN instead of LK nowcast methods. Here an intensity threshold of 1 mm/h was used.*
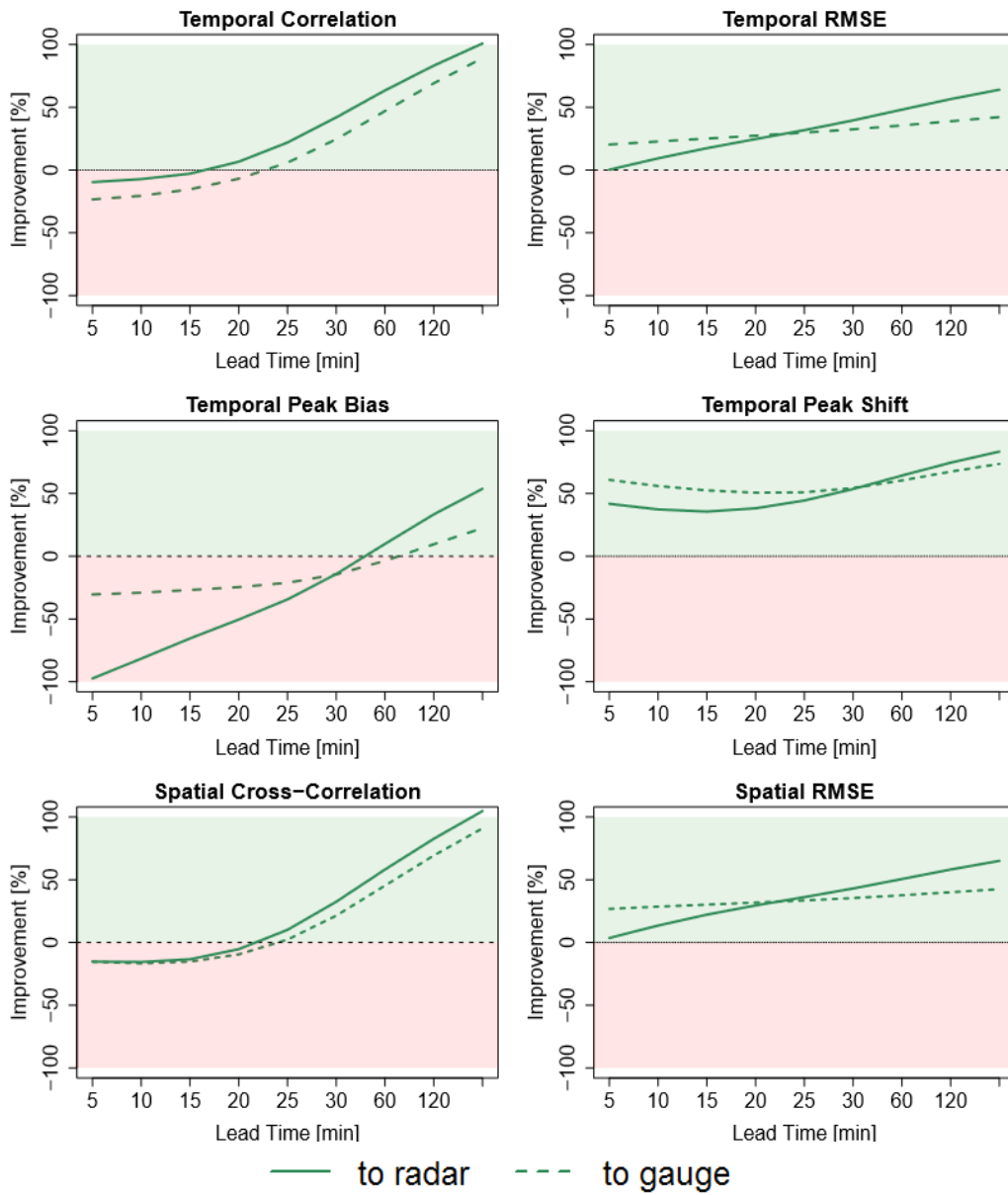
Figure D.6: *The average improvement of spatial and temporal criteria for each lead time when using CNN instead of LK nowcast methods. Here an intensity threshold of 5 mm/h was used.*
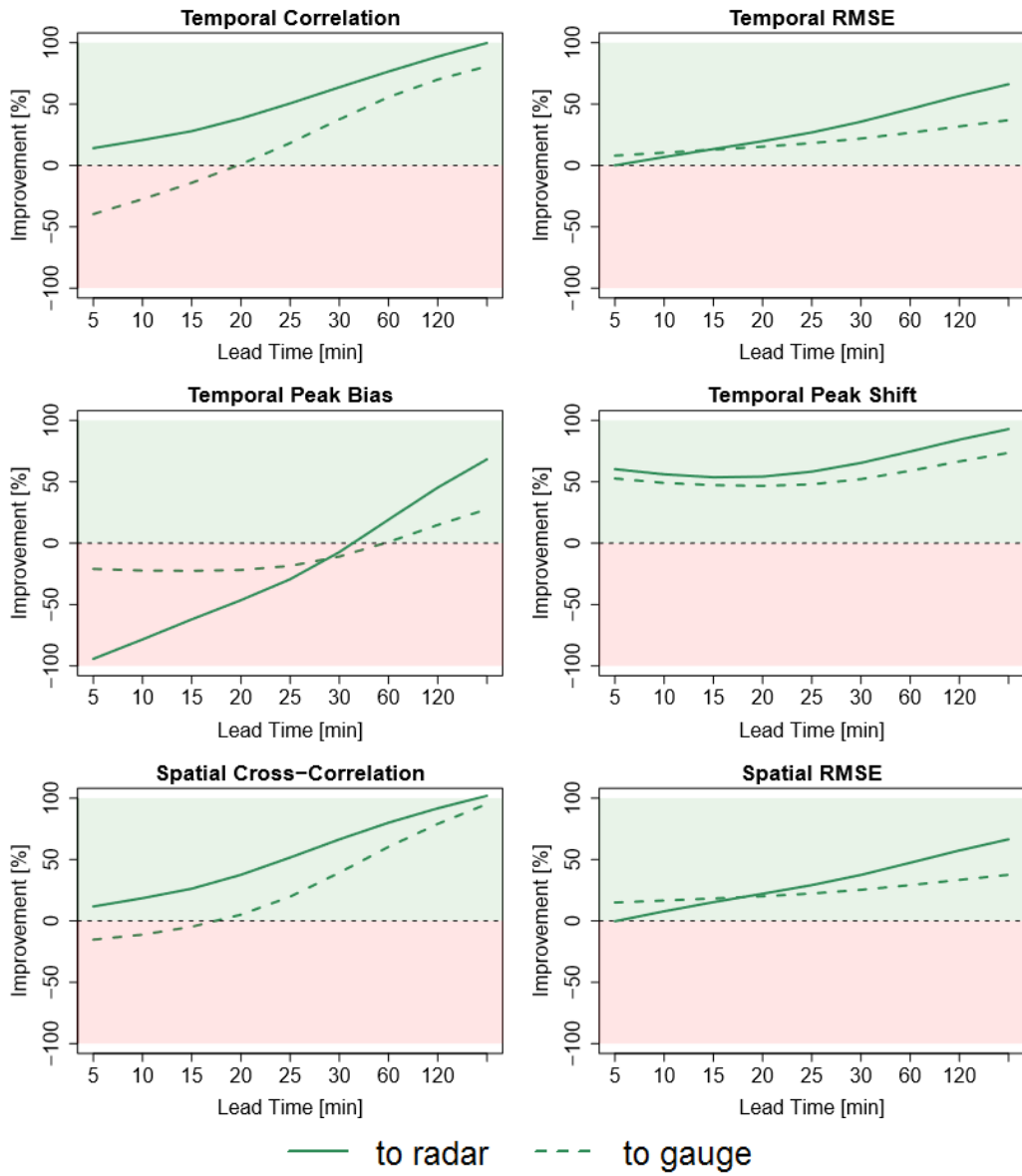
Figure D.7: *The average improvement of spatial and temporal criteria for each lead time when using CNN instead of LK nowcast methods. Here an intensity threshold of 12 mm/h was used.*
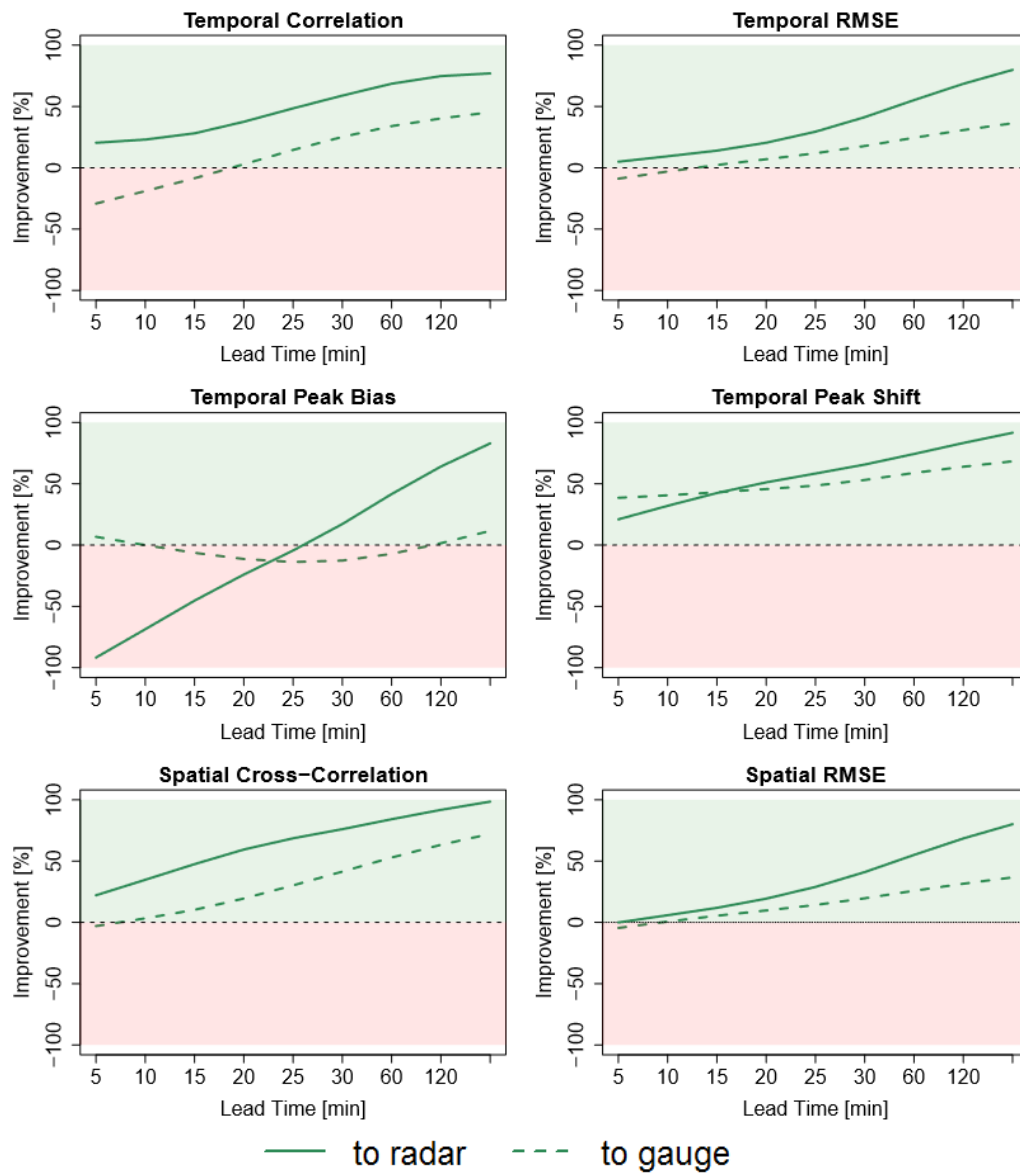
Figure D.8: *The average improvement of spatial and temporal criteria for each lead time when using CNN instead of LK nowcast methods. Here an intensity threshold of 60 mm/h was used.*

APPENDIX E

The spatial distribution of the rainfall sum of 5 selected events for all the merging methods investigated Figure E.1 to Figure E.5.
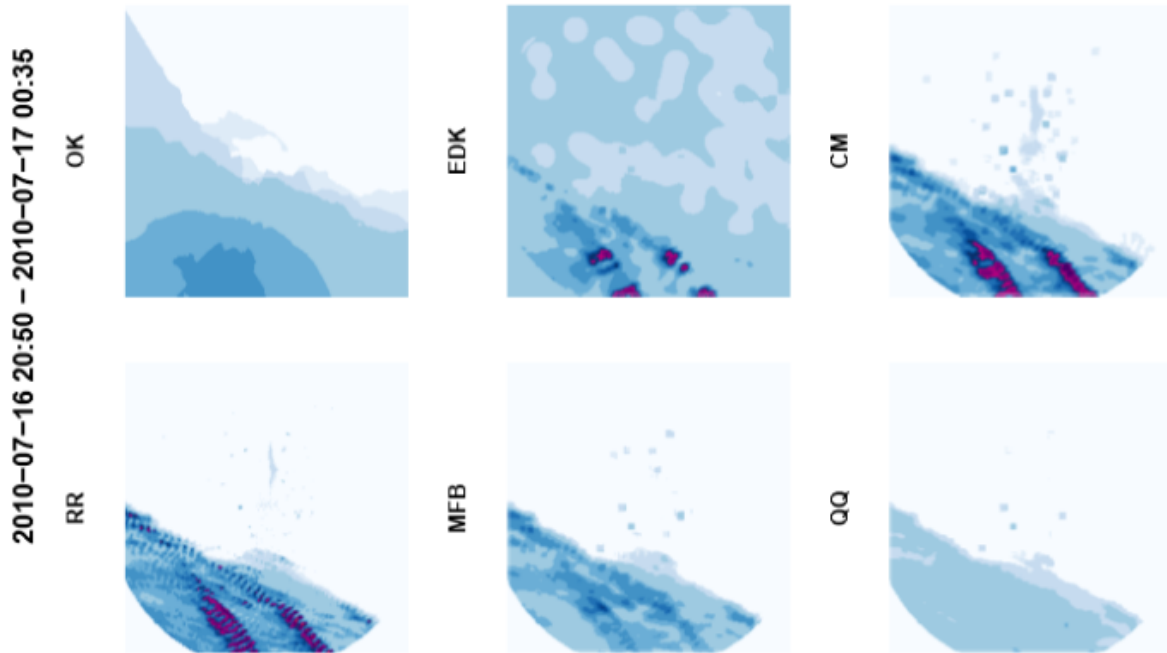
Figure E.1: *The spatial distribution of the rainfall sum of all the merging methods for a storm classified as a meso-γ type.*
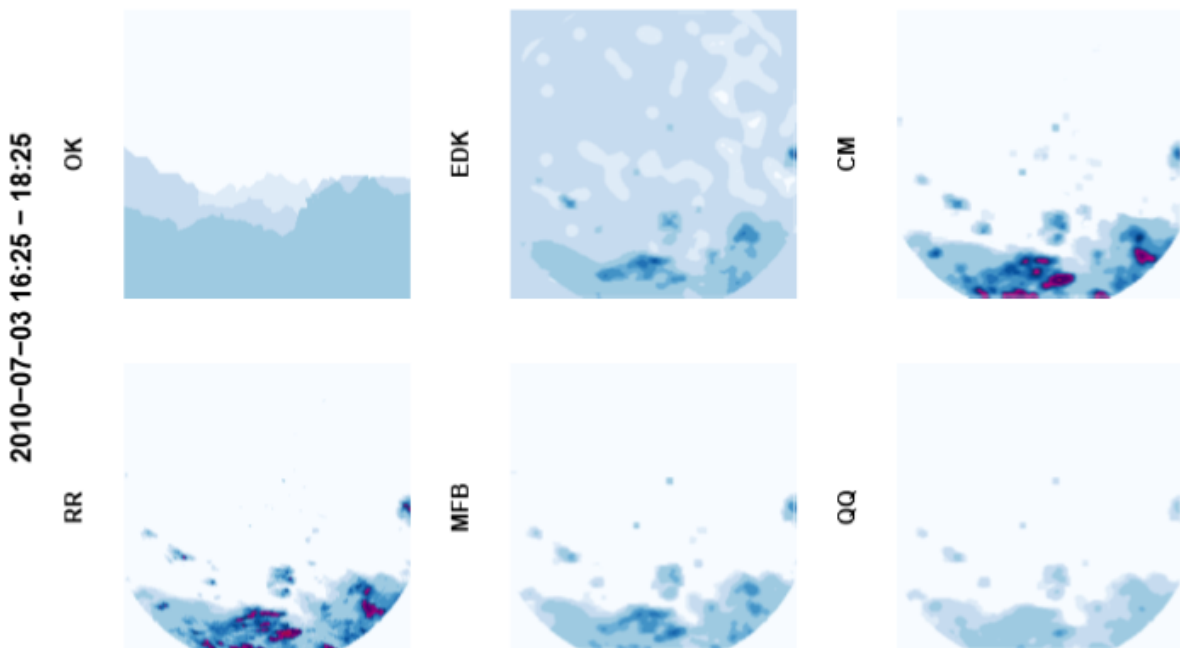
Figure E.2: *The spatial distribution of the rainfall sum of all the merging methods for a storm classified as a convective type.*
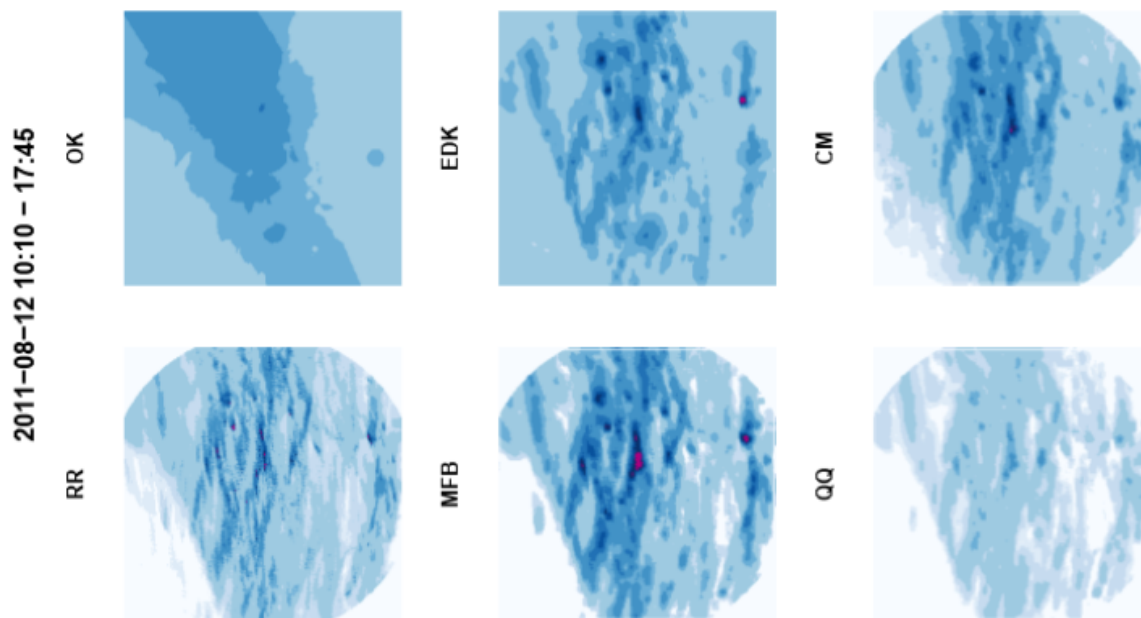
Figure E.3: *The spatial distribution of the rainfall sum of all the merging methods for a storm classified as a meso-β type.*
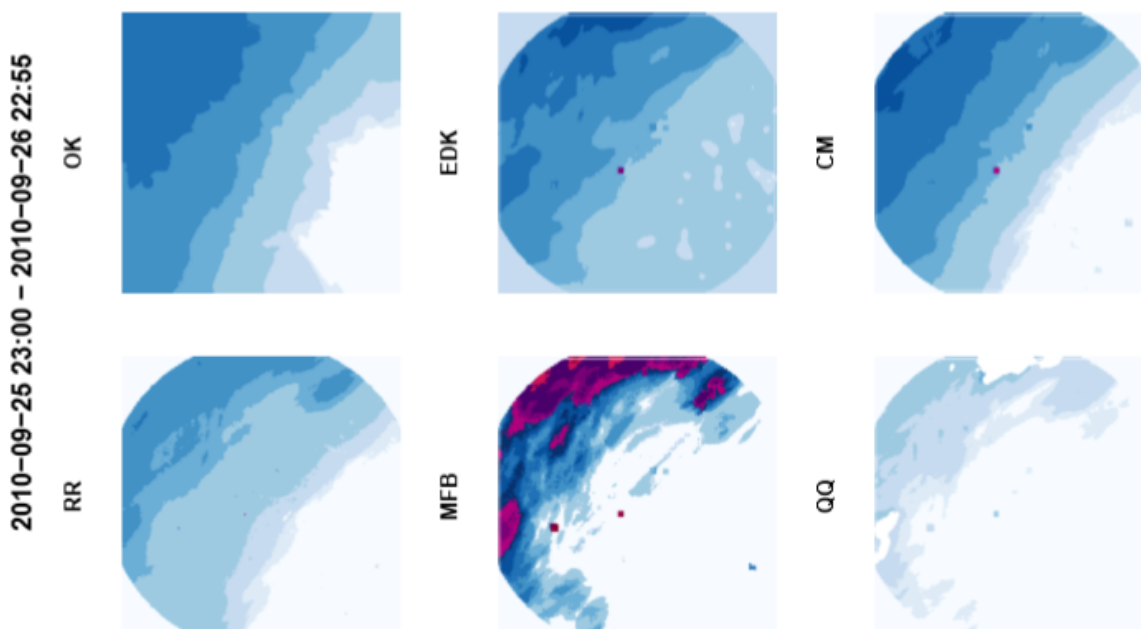
Figure E.4: *The spatial distribution of the rainfall sum of all the merging methods for a storm classified as a stratiform type.*
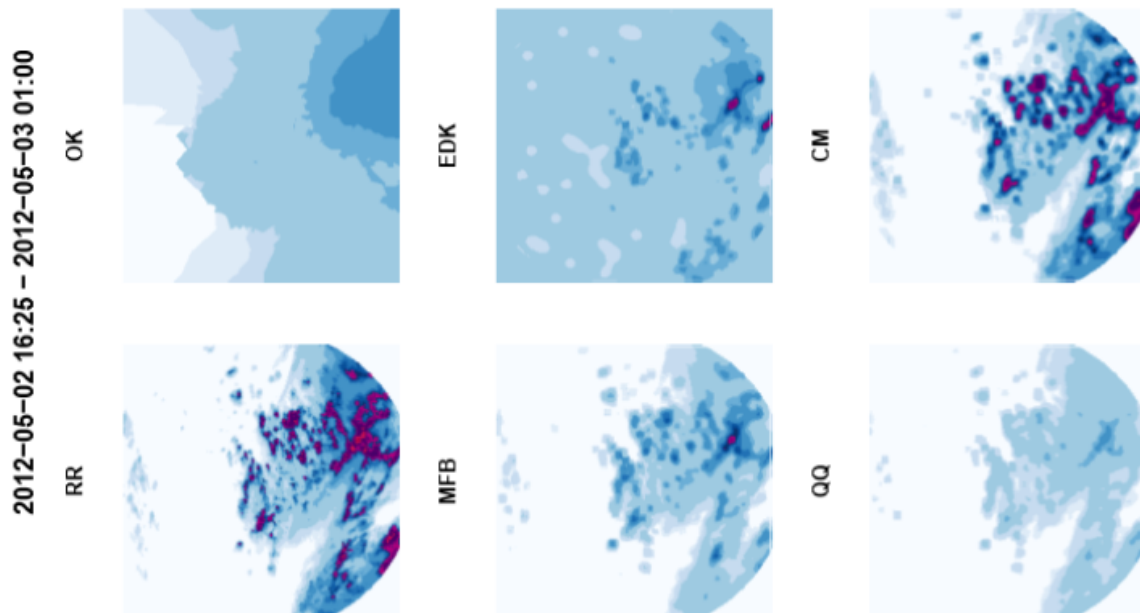
Figure E.5: The spatial distribution of the rainfall sum of all the merging methods for a storm classified as a mixed type.

APPENDIX F

The correlation matrix for the selected storm predictors of the k-NN approach for the step 1 of neighbour selection (Figure F.1)
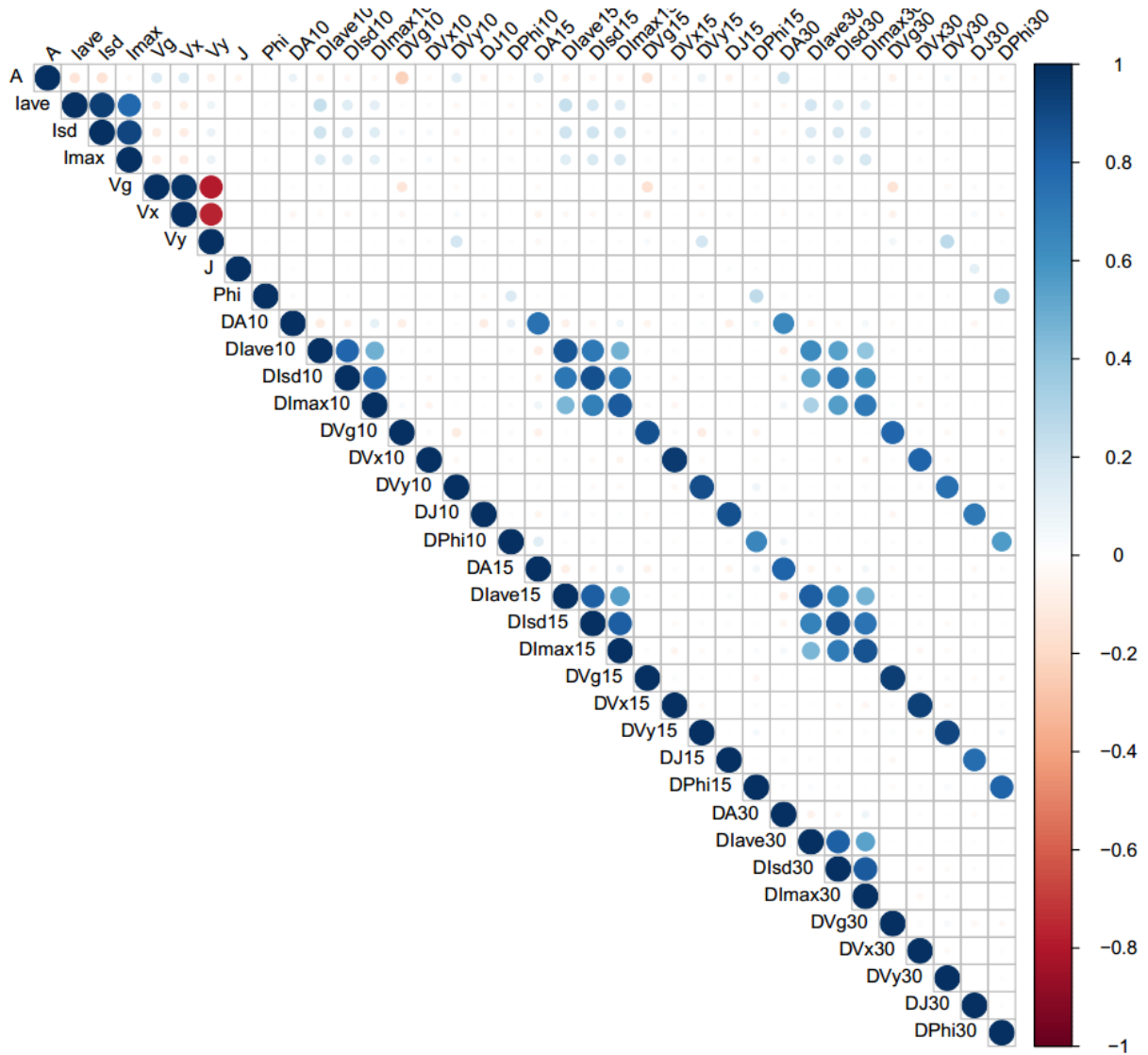
Figure F.1: *The correlation among the selected predictors for the recognition of similar storm.*

APPENDIX G

The spatial visualisation for the nowcast issued from 5 randomly selected rainfall events at +5. +30 and +60 min lead time are given below for:

I) field based nowcast approaches : Lucas-Kanade (middle row) and CNN (lower row) models from Figure G.1 to G.5.

II) object based nowcast approaches: HyRaTrac (second row), 3-NN (third row), and an ensemble member from k-NN (forth row) models from Figure G.6 to G.10.

The observed rainfall fields at each time steps and lead times is given for all the figures on the first row.
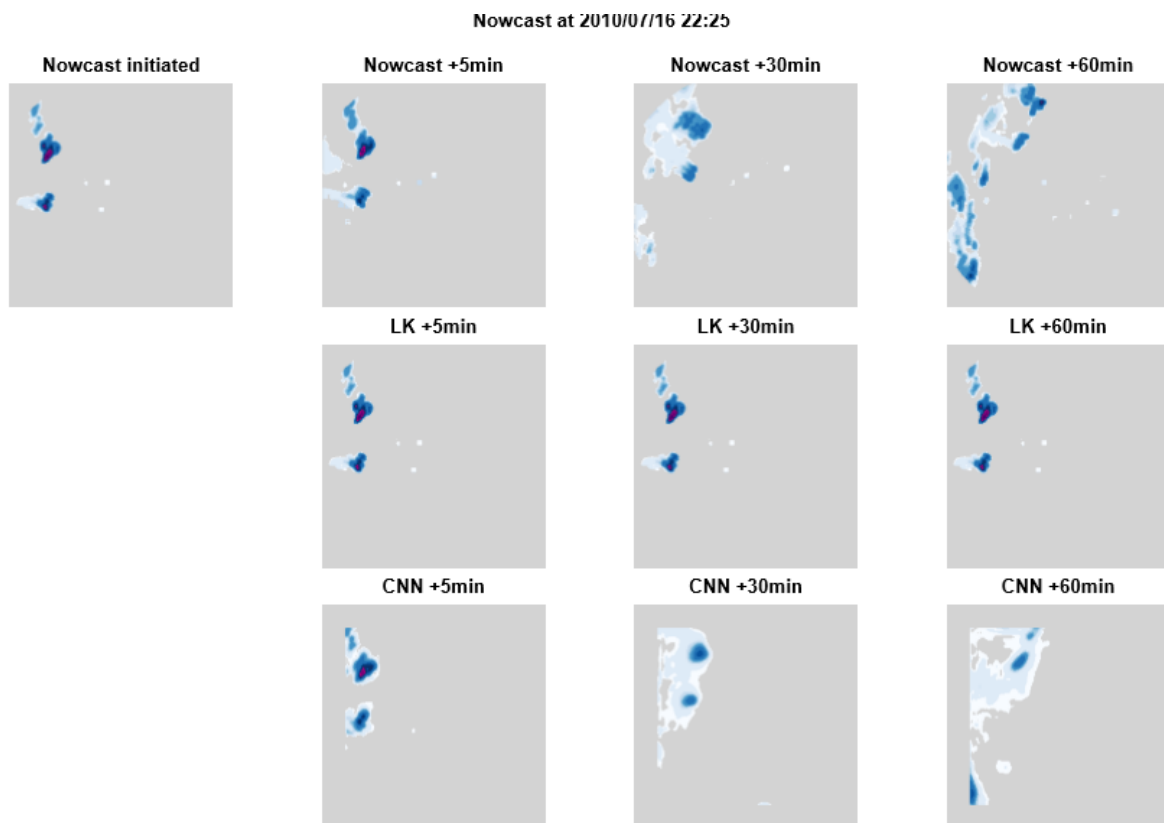
Figure G.1: *The field based nowcast approach example for a meso-γ convective event.*
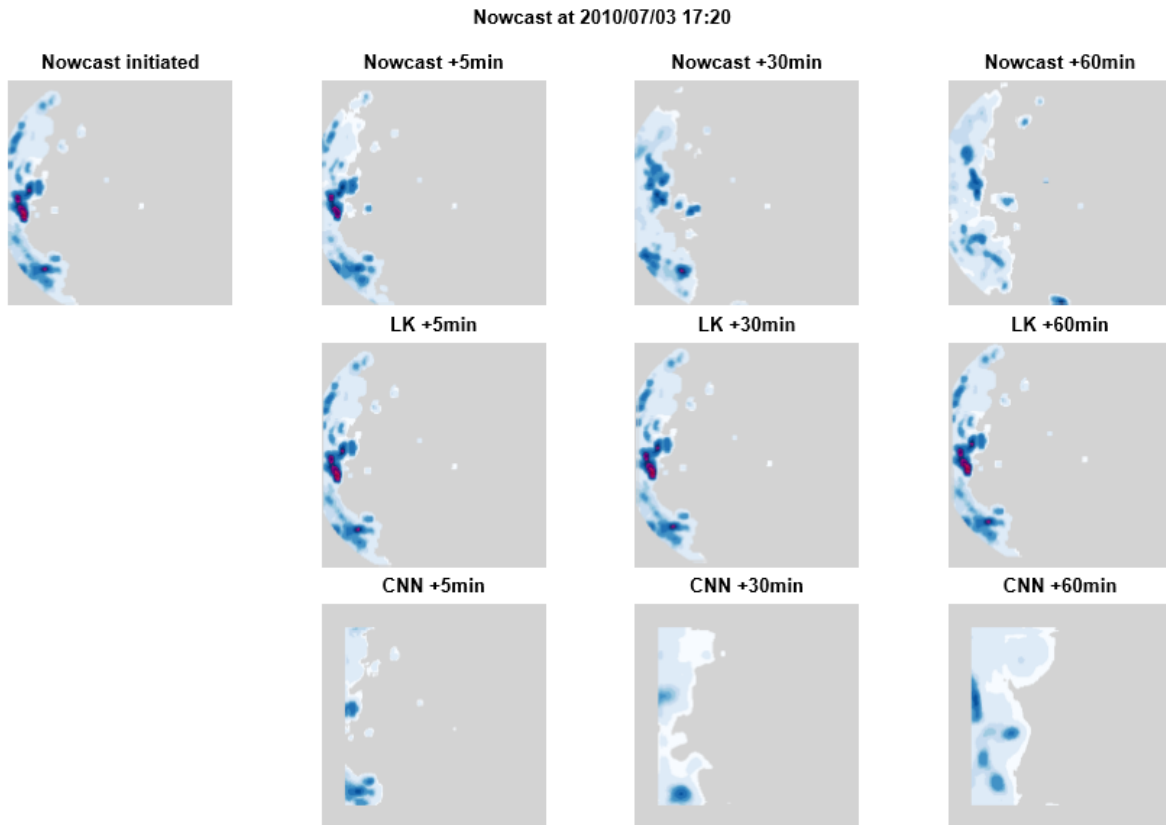
Figure G.2: *The field based nowcast approach example for a convective event.*
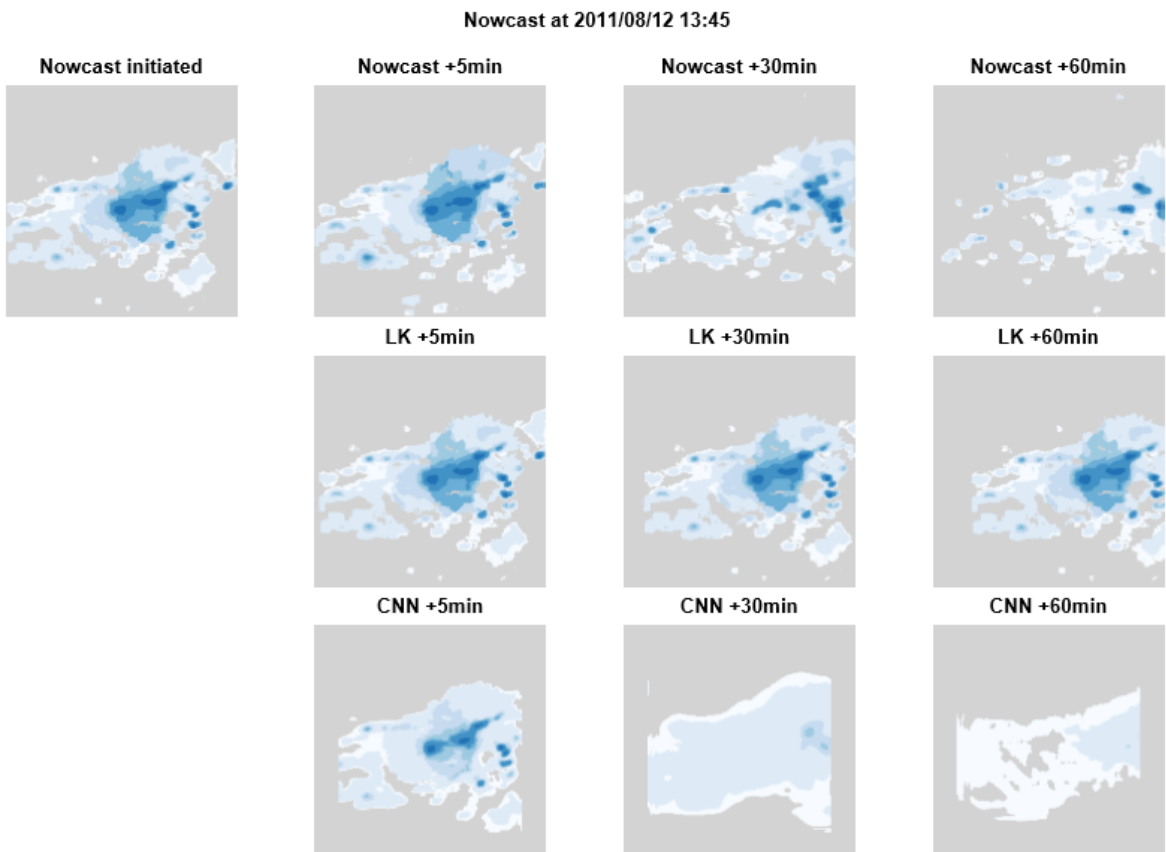
Figure G.3: *The field based nowcast approach example for a meso-β convective event.*
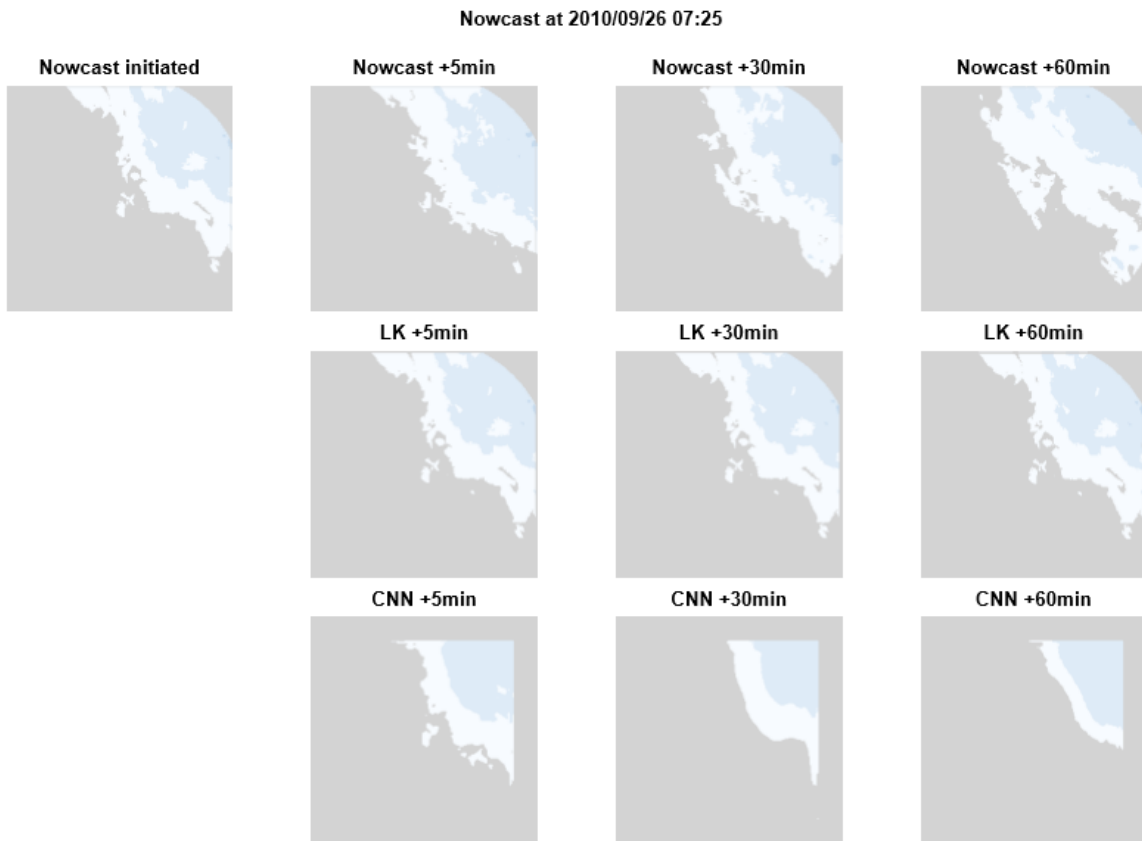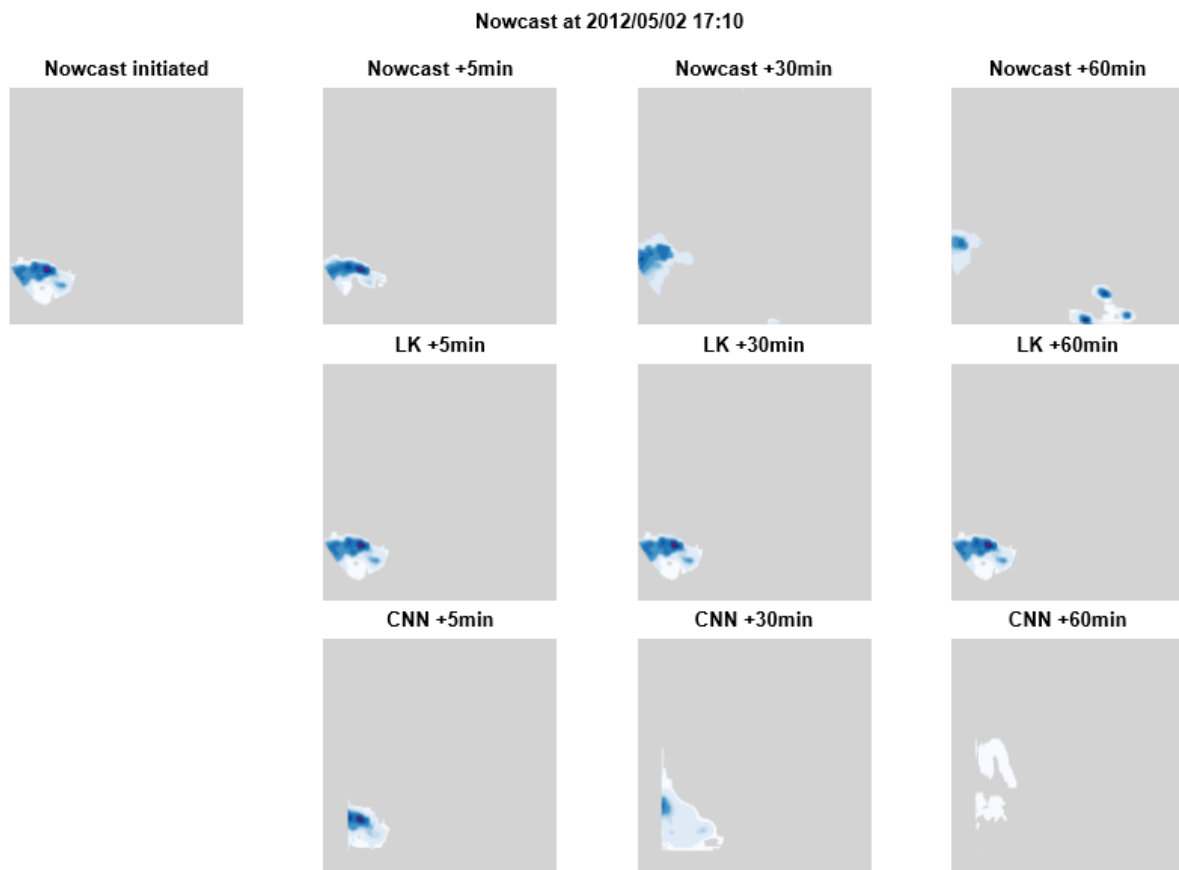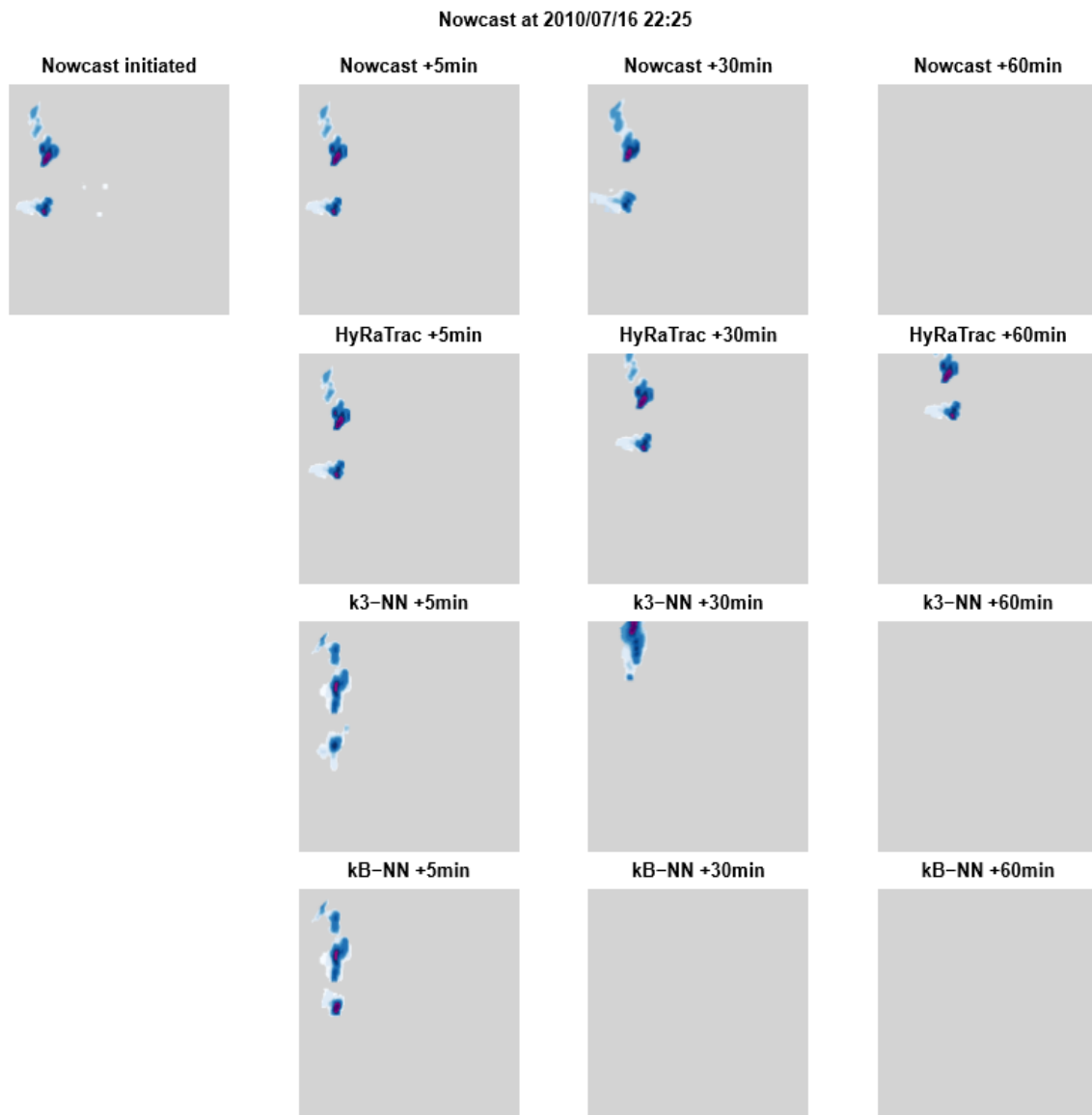
Nowcast at 2010/09/26 07:25

Nowcast initiated | Nowcast +5min | Nowcast +30min | Nowcast +60min

LK +5min | LK +30min | LK +60min

CNN +5min | CNN +30min | CNN +60min

Figure G.4: *The field based nowcast approach example for a stratiform event.*

Nowcast at 2012/05/02 17:10

Nowcast initiated | Nowcast +5min | Nowcast +30min | Nowcast +60min

LK +5min | LK +30min | LK +60min

CNN +5min | CNN +30min | CNN +60min

Figure G.5: *The field based nowcast approach example for a mixed event.*

Figure G.6: *The object based nowcast approach for a meso-γ convective event.*

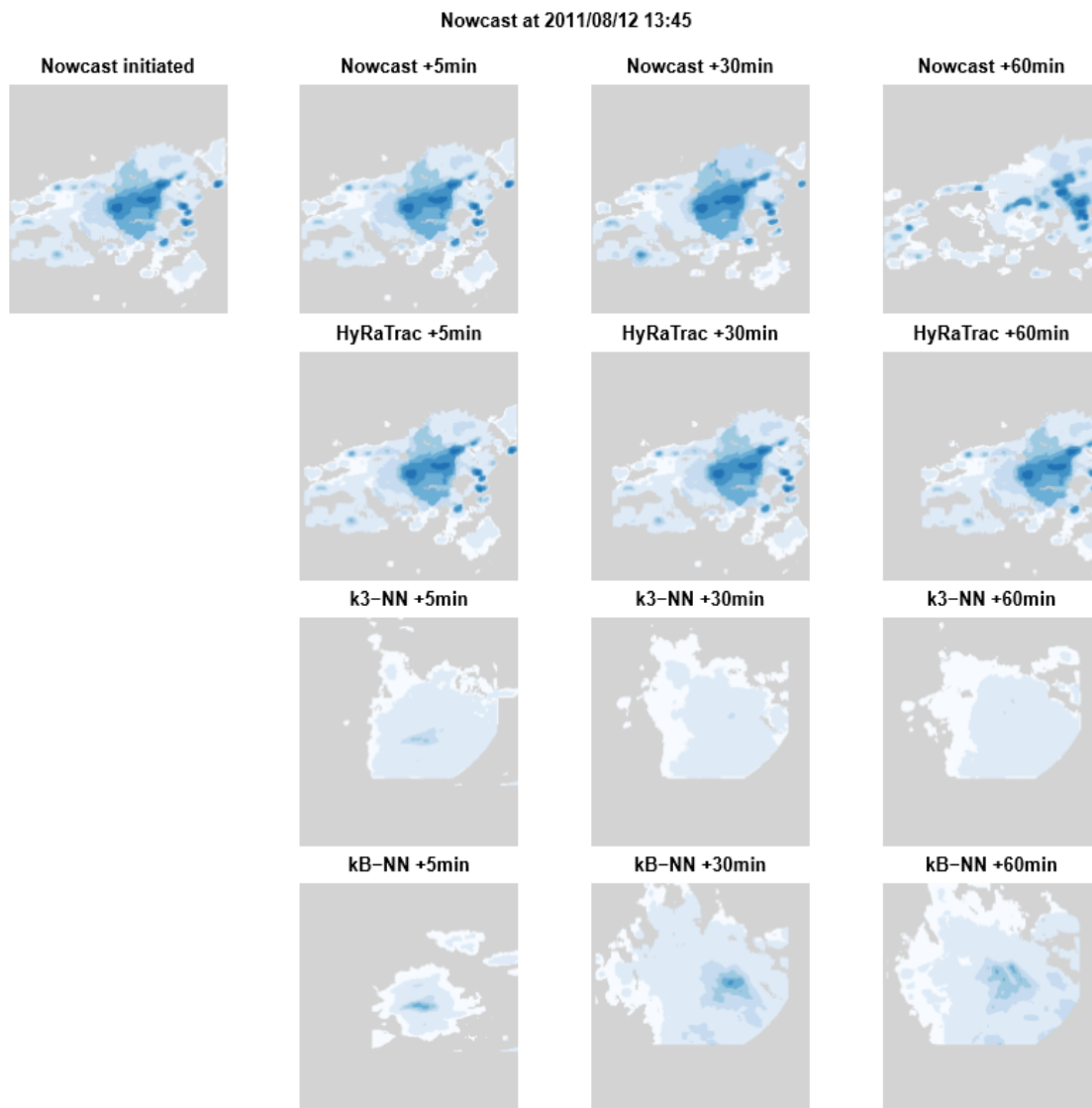**Figure G.7:** *The object based nowcast approach for a convective event.*

Figure G.8: *The object based nowcast approach for a meso-β convective event.*

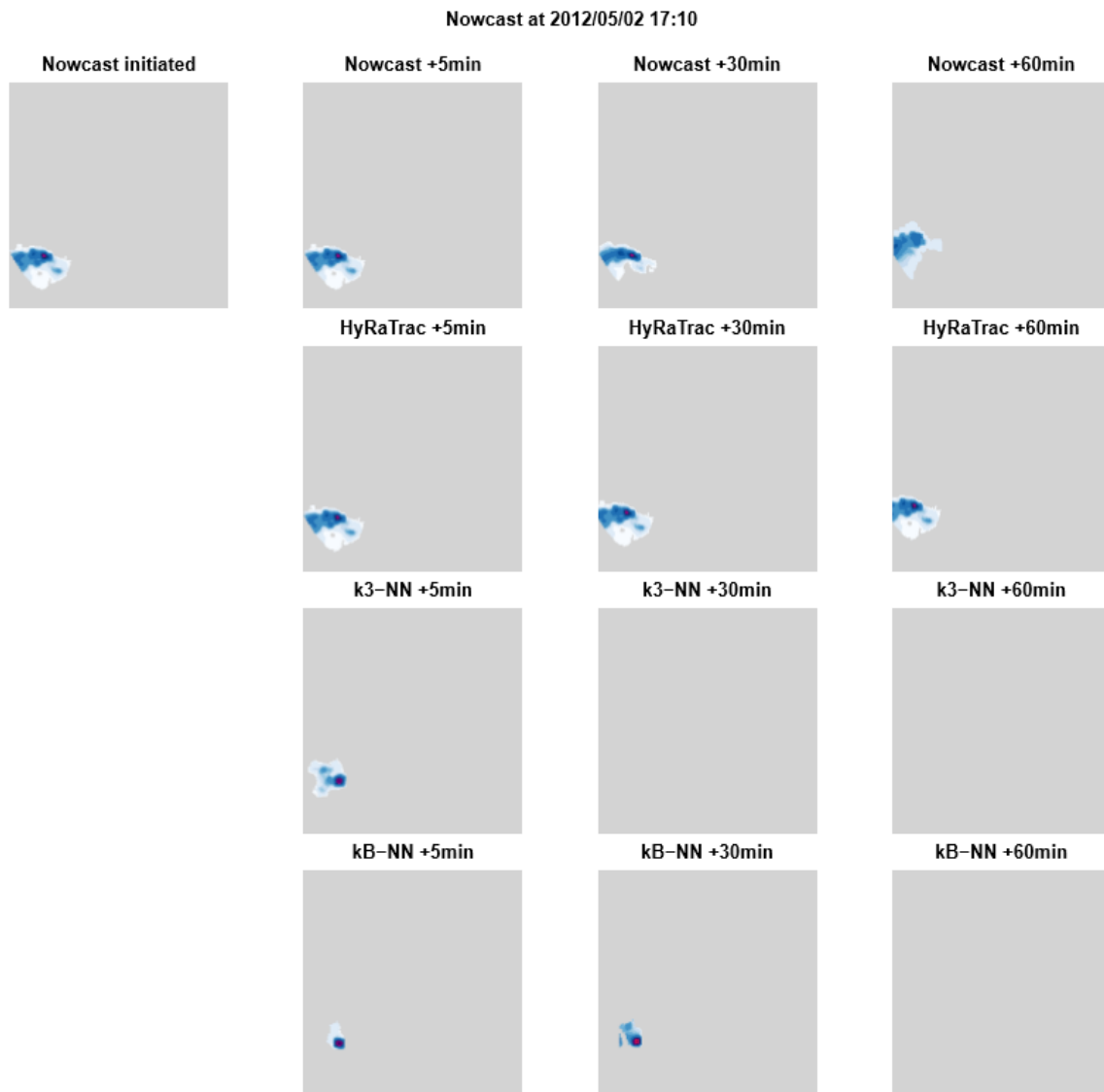**Figure G.9:** *The object based nowcast approach for a stratiform event.*

Figure G.10: *The object based nowcast approach for a mixed event.*