Verification of radar rainfall nowcasting for urban hydrological applications in the Netherlands

Nathalie Rombeek

January 2022

Abstract

Radar rainfall nowcasts can be used for real-time control (RTC) of the sewage system instead of numerical weather predictions (NWP), as nowcasts have a higher spatial (1 km^2) and temporal (5 min) resolution than NWP. In this study the potential of nowcasts for RTC was investigated. Two nowcasting algorithms were tested in the urban areas of Helmond and Laarbeek in the Netherlands: Pysteps Deterministic (PS-D) and Rainymotion DenseRotation (RM-DR). An analysis of 32 rainfall events was performed to determine the potential of nowcasting in: forecasting rainfall and RTC of the sewage system. In addition, the nowcasts were compared with the NWP model HARMONIE on both their resulting rainfall forecasts and subsequent sewer overflow forecasts. The maximum skilful lead time for 1-hour duration events is around 19 min or 37 min and for 24-hour duration events it is 56 min and 108 min for respectively RM-DR and PS-D. HARMONIE was not skilled for lead times up to 3 hours. Stratiform winter events result in higher skill than convective summer events. RM-DR can better capture high intensity rainfall as PS-D leads to more dissipation. Both PS-D and RM-DR outperform HARMONIE, but still underestimate the total sewage overflow volume, which partly results from the operational radar product. Skill was highest for events in spring, while lowest for events in summer. Events in 2016 resulted in a gain in anticipation time between 75-90 min for RM-DR, while PS-D did not predict the sewage overflow of these events. Results show much variability between events and seasons, indicating that nowcasts are not reliable enough for RTC based on forecasting a sewage overflow, however, it is more promising when based on forecasted rainfall that enters the system.

Contents

1 | Introduction

1.1 Context and motivation

Urban areas are vulnerable to extreme rainfall due to high-density built-up areas and the large proportion of impervious surfaces, consequently leading to a fast hydrological response (Fletcher et al., 2013; de Vos et al., 2017). Extreme rainfall will increase as climate change projections expect more frequent and intense rainfall events (Westra et al., 2014). Climate change, together with the rapid urbanization trend will raise the vulnerability of cities (UN, 2018). Consequently, flood forecasting and real-time available warning systems are in demand as they are able to forecast a flood hours in advance (Codo and Rico-Ramirez, 2018), making real-time water management possible. These urban flood forecasting systems require rainfall data at higher temporal and spatial resolutions due to the fast response (Berenguer et al., 2005; Thorndahl et al., 2017), and high heterogeneity in urban areas (Berne et al., 2004; Tilford et al., 2002). Due to the spatial and temporal variability of rainfall fields, current rainfall estimates and especially forecasts are uncertain, leading to uncertain hydrological forecasts (Berenguer et al., 2005; Codo and Rico-Ramirez, 2018; Thorndahl et al., 2017).

One of the methods that have been developed for rainfall forecasting is numerical weather prediction (NWP) models. Multiple countries, such as the Netherlands, use these models in their current operational weather forecasting system (Bengtsson et al., 2017). The dynamics and physics of the atmosphere are simulated in NWP models. Current NWP models can give relatively reliable forecasts on large spatial scales up to two to three days ahead (Liguori et al., 2012). NWP models can be useful in predicting stratiform long-duration storms. However, for short-term predictions up to 6 hours NWP models are not reliable enough for early warning systems (Lin et al., 2005). These models are restricted in predicting small-scale processes, such as convective showers, due to their coarse spatial resolution and low update frequency. Therefore, it is not suited for urban flood forecasting systems (Liguori et al., 2012).

To overcome the restrictions of NWP models, radar rainfall products have improved significantly over the last decades to get more reliable forecasts with higher spatial and temporal resolution and shorter lead times. Nowadays, radar images are used to forecast rainfall events in the near future, which is called radar rainfall nowcasting. Nowcasts are extrapolated real-time radar rainfall fields. There are two techniques to make nowcasts, namely field- and object-based nowcasting. Field-based nowcasting advects the most recent observed precipitation field, and is for example applied in Rainymotion (Ayzel et al., 2019) and Pysteps (Pulkkinen et al., 2019), while object-based nowcasting identifies individual storms and tracks it, and is for example applied in the Thunderstorm Identification, Tracking, Analysis, and Nowcasting (TITAN) algorithm (Dixon and Wiener, 1993).

The high spatial and temporal resolution (1 km and 5 min in the Netherlands) of nowcasts can be a valuable addition to hydrological modelling and real-time flood warning systems. Reliability and accuracy depends on many factors such as rainfall type and intensity (Liguori et al., 2012). However, uncertainties from the rainfall forecasts are not necessarily resulting in the same uncertainties in catchment flow predictions (Achleitner et al., 2009; Brauer et al., 2016). Multiple studies already quantified the skill of nowcasts in catchments and showed promising results in hydrological applications (Berenguer et al., 2005; Heuvelink et al., 2020).

Despite progress in catchments, research about the potential of nowcasts in urban flood warning systems is still in an early stage of development (Thorndahl et al., 2017). Achleitner et al. (2009) analysed the value of nowcasts for better control of urban drainage systems. Foresti et al. (2016) investigated ensemble precipitation forecasts for urban hydrology in Belgium and concluded that different event types vary in maximum skilful lead time. Thorndahl et al. (2012) made a comparison of NWP forecast and nowcasts for a small urban catchment in Denmark. This study concluded that there is potential for nowcasts with a lead time of 0 to 2 h, while NWP is reasonable for lead times of 6 to 24 h. The same study setup, but for a larger urban catchment was done by Thorndahl et al. (2013).

However, these studies only contained a small sample of precipitation events (6 or less events), or in the case of Thorndahl et al. (2012) also a very small urban area (< 0.8 km^2). Therefore robust conclusions about the usefulness of nowcasts in urban hydrology can not be made from these studies. Earlier research already showed that the skilfulness of nowcasts depends on factors such

as event duration, storm type, and seasonality (Imhoff et al., 2020). However, due to the limited number of precipitation events studied in urban areas, no conclusions could be made on the dependence of forecast skill on meteorological factors, such as event duration, type, or seasonality. Furthermore, in most studies, only one nowcast algorithm was used. In addition, previous studies did not assess the added value of nowcasts compared with NWP in urban areas.

The sewerage systems in urban areas are a good example of a hydrological system that is worthwhile investigating (Schellart et al., 2014). The traditional sewerage systems consist of a combined system, which is designed to collect rainwater, industrial wastewater, and domestic sewage in the same system. In order to prevent flooding in urban areas, rainwater is drained away. Part of the pump capacity in the sewer pumping station is used to discharge collected rainfall. This part is referred to as pump capacity (POC). However, in the case of extreme rainfall, not all water can always be disposed immediately, which can result in a combined sewer overflow. These overflows consist of a mixture of rainwater, untreated human waste and industrial waste, which are loaded with toxic materials (Passerat et al., 2011). This can result in disastrous effects for the receiving water quality. Radar rainfall nowcasting could be used as real-time available warning system to forecast the sewage overflow and to prevent and mitigate damage.

This study will analyse 32 rainfall events that took place over two urban areas, using two field-based nowcasting algorithms: Rainymotion (Ayzel et al., 2019) and Pysteps (Pulkkinen et al., 2019). The nowcasting algorithms will be compared with the current forecasting system in the Netherlands, namely HARMONIE. In addition, the added value of nowcasting for hydrological forecasting in an urban area in the Netherlands will be investigated. This will be done by using a simple model that predicts the overflow from the sewerage systems into water bodies based on the receiving precipitation.

1.2 Research questions

The main objective of this research is to determine the added value of radar-based rainfall nowcasting for real-time control (RTC) in sewerage systems in the urban areas Helmond and Laarbeek in the Netherlands. Different rainfall events in the period 2008-2018 will be investigated for this. The rainfall events will be simulated using two nowcast algorithms. Next, a simple sewer overflow model will be used to investigate the potential of radar-based rainfall nowcasting for RTC.

In order to reach the objective stated in the previous section, the following questions should be answered:

1. What is the performance of radar rainfall nowcasts in forecasting rainfall in the urban area of Helmond and Laarbeek?

2. What is the performance of radar rainfall nowcasts as input for urban hydrological models in the urban area of Helmond and Laarbeek?

   2.1. To what extent would radar rainfall nowcasts have been able to forecast the extreme precipitation events (in May and June 2016) and the subsequent inundated areas in the urban area of Helmond and Laarbeek?

For both questions the following sub-questions will be investigated:

1. What is the influence of different rainfall characteristics and seasonal differences on the performance?

2. What is the influence of different nowcast algorithms on the performance?

3. How do radar rainfall nowcasts compare to rainfall forecasts from numerical weather predictions?

1.3 Thesis contents

Chapter 2 contains the description of the study area and the data that are used in this research. Chapter 3 describes the methods used in this study. This is divided in two parts, one part to verify the radar nowcasts and HARMONIE and the other part to verify the hydrological model output. This is followed by the results (chapter 4) and discussion (chapter 5) and ends with a conclusion (chapter 6).

2 | Field site and data

2.1 Field site

For this study, two urban areas were used, namely Helmond and Laarbeek. These urban areas were chosen because both are covered by an existing database of nowcasts. Both urban areas are located in the south-east of the Netherlands, in the province Noord-Brabant.

Helmond has an area of approximately 55 km^2 (Fig. 2.1). The area consists of approximately 46% of built-up area and 44% of nature area (either agriculture or forest) (CBS, 2015). It has a population of just over 91,500 people. The water system in Helmond is complex, as the Aa, Astense Aa and Bakelse Aa all debouch in the Zuid-Willemsvaart (Waterschap Aa en Maas, 2016). This is a canal between Maastricht and 's-Hertogenbosch which crosses Helmond and Laarbeek.

Laarbeek has an area of around 56 km^2 (Fig. 2.1), which is approximately the same as Helmond. However, the area consists for a smaller part of built-up (only 16%) area and 79% of nature area (either agriculture or forest) (CBS, 2015). Consequently, with approximately 22300 people, the population of Laarbeek is smaller than Helmond.

In the period 2008-2018 on average 744 mm of precipitation fell per year over Helmond. On average, April is the driest month with 37 mm of precipitation, while July is on average the wettest month, with 82 mm. However, July had much variation between the years, the driest month over this period was July 2018, with only 3.6 mm. The wettest month was June 2016, with 200 mm of rainfall. In four days already more than 100 mm of rain fell. At the end of May and beginning of June 2016, much rain fell in a few days, which resulted in floods and consequently much damage in and around Helmond.

During the period 2008-2018 the average reference evaporation was 619 mm per year (at the meteorological station in Eindhoven) (KNMI, https://www.knmi.nl/nederland-nu/klimatologie/gegevens/monv).

The municipality of Helmond and Laarbeek measure the sewage overflow already at different locations in the city. However, the measurements, especially in Helmond, are according to water authority Aa en Maas not very accurate.
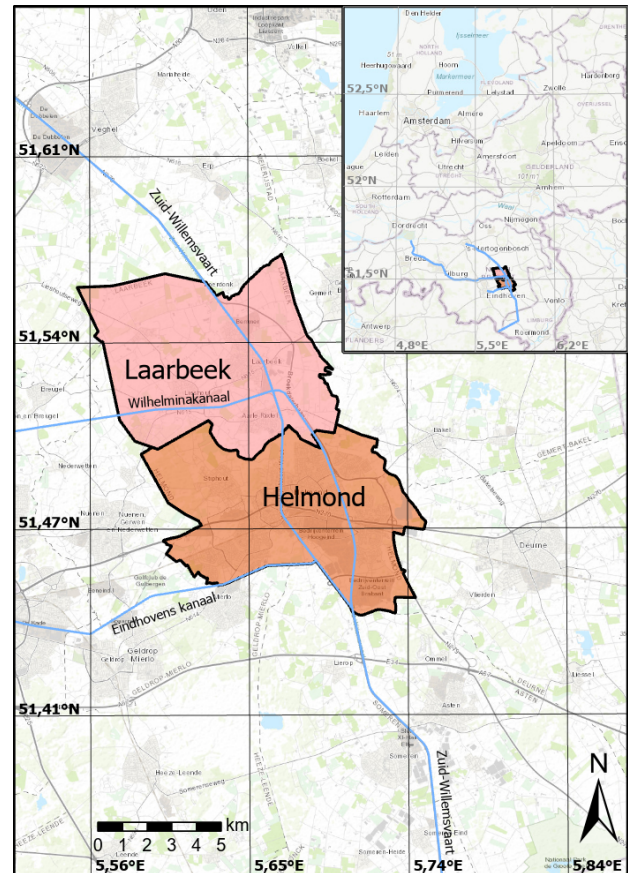

Figure 2.1: Map of Helmond, located in the Netherlands. The areas show the field sites, which will be the extent of the radar rainfall composite. The blue lines indicate the canals which crosses Helmond and Laarbeek.

2.2 Events for radar nowcasts

The used precipitation events in this research were based on the available nowcasts from Imhoff et al. (2020), which contain precipitation events in the period 2008-2018. In this study, a total of 128 rainfall events were used that cover the Aa catchment (4 seasons, 8 per season, and 4 durations). The selected events contain different types of precipitation events.

A selection of these 128 rainfall events was made based on the rainfall amounts and structure, and based on whether they produced a significant hydrological response in the urban areas. This was done by focussing on events which resulted in high sums during 1 and 24 hours. In addition, the seasonal dependency was investigated by looking at all four seasons. Per season four events were investigated. In total 32 rainfall events were used in this research, see Table 2.1.

Table 2.1: Chosen rainfall events and the accumulated rainfall averaged over the study area of Helmond and Laarbeek. Note that the end time of the rainfall event is given here.

| | 1 hour duration | | | 24 hours duration | | |
| | | Accumulated rainfall [mm] | | | Accumulated rainfall [mm] | |
	End [UTC]	Helmond	Laarbeek	End [UTC]	Helmond	Laarbeek
Autumn	12-09-2008, 10:40	11.95	9.82	08-10-2009, 07:00	37.93	35.54
	07-10-2009, 18:30	7.74	8.85	24-11-2009, 01:00	47.98	54.45
	22-10-2013, 20:40	9.52	9.25	13-11-2010, 15:40	37.63	41.8
	15-09-2016, 21:25	7.98	5.51	23-09-2015, 11:00	24.40	21.08
Spring	15-05-2008, 19:00	17.69	17.03	16-05-2008, 09:00	34.56	35.06
	30-05-2016, 15:50	9.22	16.16	12-05-2010, 06:00	25.11	26.38
	10-04-2018, 21:50	11.01	10.04	31-05-2016, 00:00	33.37	46.91
	30-04-2018, 00:15	9.25	9.75	30-04-2018, 17:50	30.57	33.32
Summer	12-07-2010, 09:50	12.94	17.35	23-08-2011, 13:00	45.78	50.26
	14-07-2010, 16:50	12.22	11.81	02-06-2016, 03:00	23.59	8.49
	27-07-2013, 09:40	6.94	8.31	24-06-2016, 06:00	53.85	38.85
	30-08-2017, 17:55	8.28	9.91	12-07-2017, 12:00	44.69	43.74
Winter	05-02-2008, 23:45	6.39	11.34	10-02-2009, 16:40	25.81	26.13
	10-12-2009, 03:10	7.16	6.68	23-12-2012, 11:00	28.26	33.33
	03-01-2012, 16:55	8.80	5.4	13-01-2017, 14:00	24.78	25.64
	08-12-2018, 19:50	6.38	9.64	23-02-2017, 04:00	31.48	39.07

2.3 Radar nowcasts

In this study, existing nowcasts made for the 128 rainfall events mentioned in the previous section were obtained from Imhoff et al. (2020). These nowcasts have a 5-min temporal and 1 km^2 spatial resolution and a forecast horizon of 6 hours. Two radar rainfall nowcast algorithms from this research were used, based on two models, namely Pysteps (Pulkkinen et al., 2019) and Rainymotion (Ayzel et al., 2019). The radar images used to make these nowcasts are operational available unadjusted radar data. The nowcasts are corrected with a bias reduction factor as a preprocessing step.

2.3.1 PS-D

Pysteps is an open-source Python framework, that focusses on the development of probabilistic and deterministic nowcasting methods. These methods are based on extrapolating radar precipitation fields.

For this study only the deterministic setup (Pysteps deterministic (PS-D)) was used. This model uses radar images to determine the motion field. Those radar images are extrapolated into the near future by using a backward semi-Lagrangian advection method (Germann

and Zawadzki, 2002). This method allows rotational movements. A second order autoregression is used on different spatial scales. PS-D is based on the S-PROG algorithm. The resulting forecast produced by S-PROG has a smooth field. When the forecast time is longer, the radar precipitation fields become more smooth. This ensures that the smallest scales of the field are filtered out when they become unpredictable (Berenguer et al., 2005). As a last step, post-processing is applied, so the nowcasts received the same statistical properties as the latest available observations.

2.3.2 RM-DR

Rainymotion has different models to make precipitation forecasts. These models use global or local optical flow algorithms (proposed by Farnebäck, 2003) in order to determine advection of rainfall fields.

For this research the global optical flow model Rainymotion DenseRotation (RM-DR) was used. This model uses radar images to estimate the velocity of each image pixel. Extrapolation is based on a forward semi-Lagrangian advection scheme. This scheme allows large-scale rotational movements. After extrapolation, Inverse Distance Weighting is used to interpolate the predicted

Table 2.2: Chosen forecasts for HARMONIE with given issue times. For every issue time, lead times from 1 to 6 hours are used.

Issue time [UTC]
2018-01-03 06:00
2018-04-11 00:00
2018-04-30 00:00
2018-04-30 06:00
2018-05-22 18:00
2018-05-23 00:00
2018-05-23 06:00
2018-05-23 12:00
2018-12-09 00:00

pixel values to their original radar grid. Contrary to PS-D, RM-DR does not use a second order autoregression which filters out the smallest scales

2.4 HARMONIE

Currently the Royal Netherlands Meteorological Institute (KNMI) uses the numerical weather prediction model HARMONIE 36 (HIRLAM ALADIN Research on Mesoscale Operational NWP in Euromed) for rainfall forecasting. This is a non-hydrostatic model that uses the current state of the atmosphere. The rainfall forecasts are used by the Dutch water authority Aa en Maas. The data has a spatial resolution of 2.5 km^2, a time step of 1 hour, and an update frequency of 6 hours. The forecasts are issued at 00:00, 06:00, 12:00, and 18:00 UTC. However, the model takes some time to run (up to a few hours), consequently, the forecasts are only available after a few hours. The forecast horizon is 48 hours.

In this study, HARMONIE was compared with radar nowcasts. However, historic HARMONIE data is only available since October 2017 until now, so only nowcasts between October 2017 and 2018 were compared with HARMONIE. Because HARMONIE has a different time step, different events than shown in Table 2.1 were used in this part. The used forecasts were based on the available nowcasts in the period October 2017 and 2018 and the issue times of HARMONIE forecasts. The selected forecasts were based on the latest NWP forecast that would have been available in real-time (with lead times varying between 1 and 6 hours). This resulted in 9 issued forecasts (shown in Table 2.2).

2.5 Reference data

Gauge-adjusted radar data from a rainfall dataset by KNMI was used as rainfall reference for the two types of forecast. It is an accurate radar product as it is corrected with rain gauge data. Radar composites have a high spatio-temporal resolution, however, weather radars often underestimate the rainfall. In contrast, rain gauges produce accurate point measurements, but the network density is too low to cover the spatial rainfall variability. In order to get the best quality of the radar products, the rain gauge data is merged with radar rainfall. Two adjustments, using a dense manual gauge network, are applied to the unadjusted radar-based precipitation depths. This gauge-adjusted radar data set has a 5-min temporal resolution and a 1 km^2 spatial resolution. (Overeem et al., 2009).
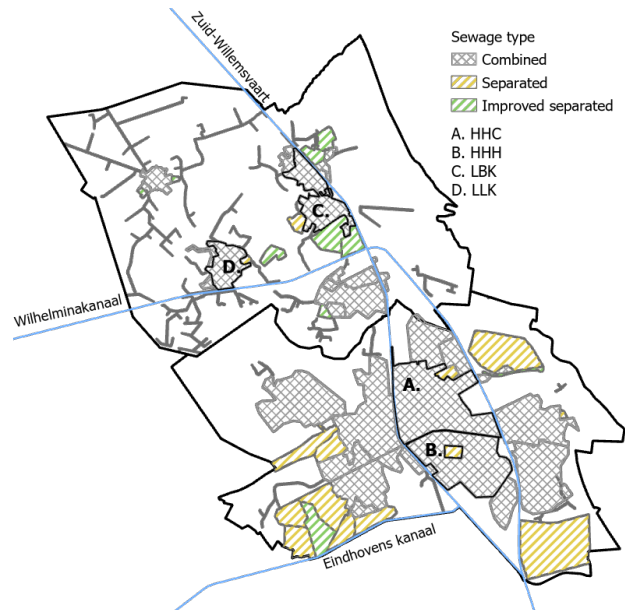

Figure 2.2: Map of Helmond and Laarbeek, with the layout of sewerage network, type of sewerage system and used sub-areas for this research. Layout of the sewerage network is obtained from Aa en Maas.

2.6 Hydrological model

In this research, a simple model that predicts the overflow from the sewerage system into surface water bodies was used. This model is developed and used by the water authority Aa en Maas. Consequently, model parameters were already calibrated. The model uses rainfall (mm per 5 minutes) as input and the output is the sewage overflow at different sub-areas in the city. The input is the area averaged rainfall, which is determined by the grid

```
22        BERGING_MINUS = 1  # Surface storage area constant
23        DAMP_MINUS = 0.1  # Infiltration evaporation constant
24
25        neerslag = recent_rain.last_hour()
26        prev_2h_mm = recent_rain.previous_2_hours()
27        # Calculate inloop.
28        minus = BERGING_MINUS if prev_2h_mm < 1 else DAMP_MINUS
29        inloop_mm = max(rnd(neerslag - minus), 0)
30
31        foo = prev_berging + inloop_mm  # Calculates the new storage
32        alt_poc = max_poc if prev_berging > 0.5 else max_poc * 0.5 # determines the POC
33        poc = min(foo, alt_poc)  # Calculates POC
34        berging = min(max_berging, max(0, foo - poc))  # Calculate storage.
35        prev_berging = rnd(berging)
36        # Only overstort if max_berging is exceeded.
37        overstort = max(0, foo - poc - max_berging)  # Calculate overflow.
38        inloop_m3 = rnd(overstort * area * 10)  # Calculate inloop in m^3.
39
40
```

Figure 2.3: Small piece of the python code of the simple sewage overflow model from the water authority Aa en Maas. Code shows the threshold based on the precipitation 2 hours prior to the last hour, line 28. The threshold of the maximum storage also determines if a sewage overflow will take place or not (line 37).

Table 2.3: Chosen sewerage sub-areas and the specifics per area. Locations with the same prefix (e.g. HEL-HEL-CEN) have the same rainfall input. Information in this table is obtained from Aa en Maas.

Location code	Sewage type	POC (mm/h)	Storage (mm)	Area (ha)	Shapefile code
HEL-HEL-CEN-1 (HHC1)	Improved combined	0.7	21.96	100.00	HEL-HEL-CEN
HEL-HEL-CEN-2 (HHC2)	Improved combined	0.7	12.97	44.69	HEL-HEL-CEN
HEL-HEL-HOO-8 (HHH8)	Improved combined	0.7	10.59	65.29	HEL-HEL-HOO
HEL-HEL-HOO-9 (HHH9)	Combined	0.7	7.80	43.53	HEL-HEL-HOO
LAA-BEE-KOM-8 (LBK8)	Improved combined	0.7	7.77	58.94	LAA-BEE-KOM
LAA-BEE-KOM-11 (LBK11)	Combined	0.7	6.43	4.06	LAA-BEE-KOM
LAA-LIE-KOM-15 (LLK15)	Improved combined	0.7	8.92	11.68	LAA-LIE-KOM
LAA-LIE-KOM-16 (LLK16)	Improved combined	0.7	7.85	17.52	LAA-LIE-KOM

cells that overlap the area of the sewerage sub-system indicated in Fig. 2.2.

The different sewerage sub-areas are shown in Fig. 2.2. These sewerage systems consist of different types, namely (improved) separated sewerage and (improved) combined sewerage. Combined sewerage collects both wastewater and rainfall in the same pipe, while a separated system has one pipe for the wastewater and another pipe for the rainwater. For every sewerage sub-area the maximum pump capacity (POC, mm h^{-1}), storage (mm) and service area of the location (ha) were given (Table 2.3). The parameters are lumped for each sub-area. The POC is the part of the pump capacity

that is used to discharge collected rainfall. The output of the model is the actual POC, rainfall that fell during the hour (mm) and overflow (m^3 h^{-1}). The output is given per hour per sub-area. The overflow is the volume of water that can not be discharged to the wastewater treatment plant as the maximum capacity is reached, instead it is discharged into the environment, often the nearest water body.

Figure 2.3 shows the most important steps of the model. In line 25 the amount of rain that fell during the last hour is determined, called *neerslag*. Next, in line 26 the total amount of rainfall that fell during the two hours prior to the last hour is determined and called

prev_2h_mm. The amount that will be subtracted from the *neerslag* is determined in line 28 and is based on two constants (1 or 0.1, see line 22 and 23) and on *prev_2h_mm*. In line 29 the amount of rainfall that actually reaches the system is calculated and is called *inloop*.

With the *inloop* and previous storage, the new storage, before subtracting the POC, can be calculated (line 31). The POC is based on the previous storage or the previous storage plus the *inloop* (line 31, 32, and 33). This shows that when there is more than 0.5 mm of storage in the system, the POC will be on its maximum. In line 34 and 35 the new storage is determined by either the maximum storage, or on the previous storage plus *inloop* (line 31) minus the POC (line 33). In line 37 the sewage overflow (in mm per hour) is based on the previous storage plus *inloop* (line 31) minus the current POC (line 33) and maximum storage. The overflow is multiplied with the surface area (ha) and 10 to calculate the overflow volume (m^3 h^{-1}) (line 38).

The used sub-areas in this research can be found in Table 2.3 and are also indicated in Fig. 2.2, in which the locations with the same prefix (e.g. HEL-HEL-CEN) have the same precipitation input. Those used sub-areas only consist of combined or improved combined sewerage systems, as those systems result in the most environmental damage during an overflow. An improved combined sewerage system has extra capacity to reduce the overflow from wastewater. In addition, close to the pump, a storage settling basin is built to settle waste. This will result in less waste emission during an overflow.

Access to this model was granted by water authority Aa en Maas.

3 | Methods

This chapter explains the methods used in this research. The methods are divided in two parts. First, the radar nowcasts and HARMONIE were verified with observed gauge-adjusted radar data (Fig. 3.1A). Note that RM-DR and PS-D use observed unadjusted radar data to make forecasts, while the verification data consists of observed gauge-adjusted radar data. The nowcasts are later corrected with a bias reduction factor. Secondly, the rainfall forecasts were used as an input for a hydrological model (a simple sewer overflow model) and compared with the reference run (Fig. 3.1B).

3.1   Data preparation

The radar nowcasts were corrected with a bias reduction factor. This was done by using Climatology-based Adjustments for Radar Rainfall in an OperaTional Setting (CARROTS), obtained from Imhoff et al. (2021). This is a set of bias reduction factors, which varies per day of the year and per grid cell in the Netherlands.

As a pre-processing step, The precipitation fields of HARMONIE were re-projected to the precipitation fields of the gauge-adjusted radar data. Additionally, the cell size of HARMONIE was downscaled, in which the new cells keep the same value as the overlaying cell.

3.2   Verifying nowcasts

First, radar nowcast developed by two algorithms (PS-D and RM-DR) were compared with gauge-adjusted radar data and subsequently mutually. The gauge-adjusted radar data were only used for verifying the nowcasts, which was done both area-averaged and grid-based. In this way, the quality of the nowcast algorithms was investigated by taking into account multiple aspects. The area-averaged precipitation was obtained by using the shapefiles of the study area and calculating the average rainfall depth (mm) over this area.

Different performance metrics were used. The false alarm rate (FA) and hit rate (HR) were calculated based on contingency tables (Wilks, 2011). Pearson's correlation, the root mean square error (RMSE), and bias were used to quantify the accuracy of nowcasts. This analysis was done over the surface area of Helmond and Laarbeek, shown in Fig. 2.1. Verification of the radar

nowcasts with the gauge-adjusted radar data were done for all 32 selected events, shown in Table 2.1.

3.2.1   Contingency tables

The hit rate (HR) and false alarm rate (FA) use predefined thresholds for rainfall intensity or rainfall sum (Liguori et al., 2012). The threshold in this study was later on determined, by taking into account which intensities led to inundation in the past. Mostly small intense rainfall events have a considerable impact in urban areas. The HR and FA were calculated with the accumulated rainfall (mm) during the specific event. This was done for each 5-min lead time up to maximum 2 hours.

A HR gives the proportion of events that were correctly forecasted for a predefined threshold, and is calculated as:

$$HR = \frac{a}{a + c} \tag{3.1}$$

where a is the number of correct forecasts and c is the number of events that were not forecasted while the observation actually exceeded the threshold.

A FA gives the proportion of events that were not forecasted for a predefined threshold, and is calculated as:

$$FA = \frac{b}{b + d} \tag{3.2}$$

where b is the number of incorrect forecasts (forecast exceeds the threshold, but the observation does not) and d the number of correct negative forecasts (both observation and forecast do not exceed the threshold). FA varies between 0 and 1, in which 0 is the best possible forecast and 1 the worst one possible.

The HR and FA were calculated at a given grid point.

3.2.2   Pearson's correlation

For decision making it is important to recognize the threat as soon as possible. Therefore a maximum lead time that is still reliable is necessary, also referred to as maximum skilful lead time. Pearson's correlation was used for this. It measures the correlation between two sets of data (in this case the gauge-adjusted radar data and forecasted rainfall). Per event and lead time Pearson's correlation was calculated. A threshold of $1/e$
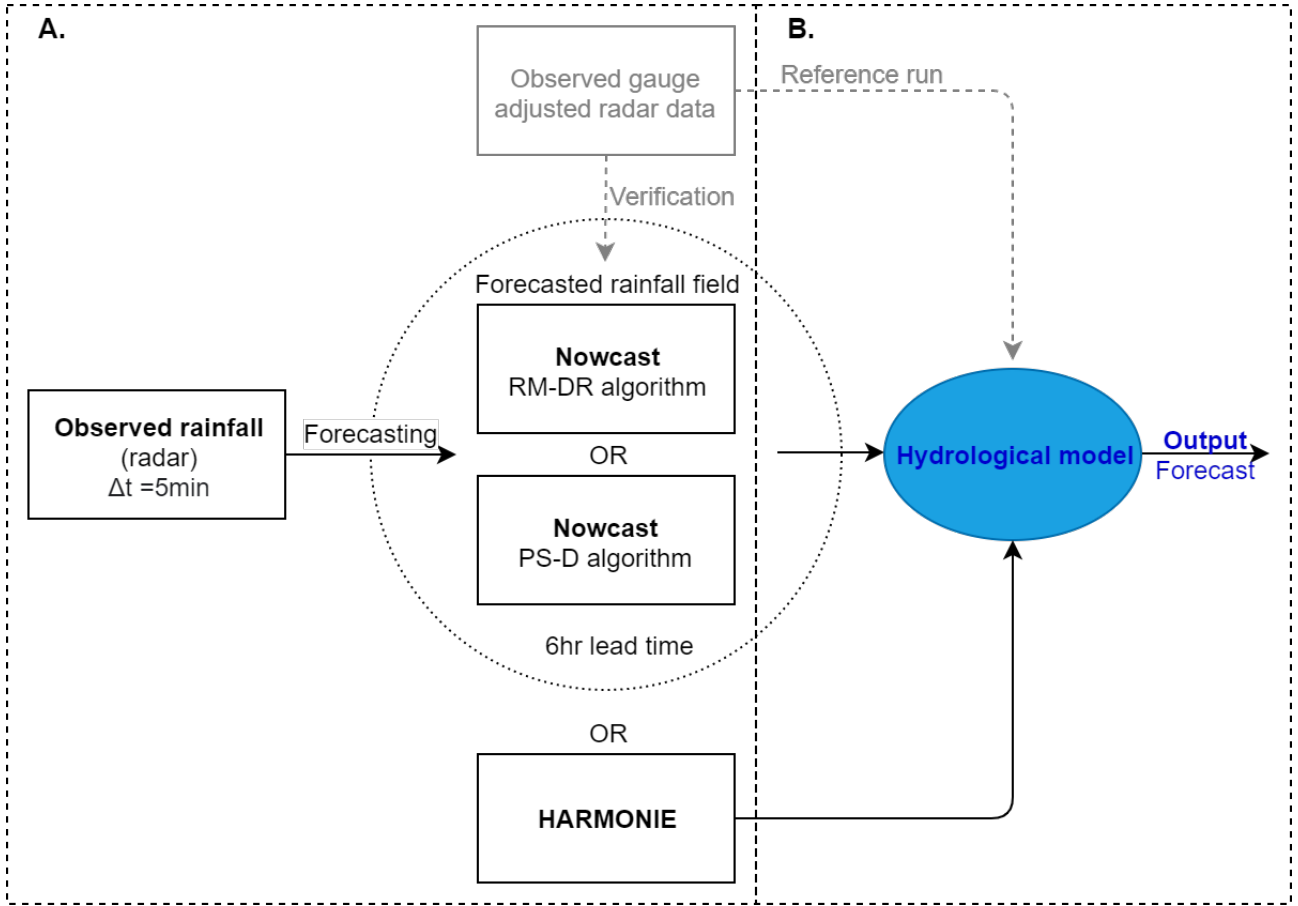
Figure 3.1: Schematization of the rainfall forecasting chain for a single simulated event. Gauge-adjusted radar data is used for verification of the nowcast or HARMONIE (A) and the hydrological model output (B).

($\rho \approx 0.37$) was used in order to say if there is still a correlation or not (Berenguer et al., 2011). Below this threshold, the lead time is not skilful any more. The Pearson's correlation coefficient (ρ) is calculated as:

$$\rho = \frac{1}{N} \sum_{i=1}^{N} \frac{(F_i - \mu_F)(O_i - \mu_O)}{\sigma_F \sigma_O} \qquad (3.3)$$

with F_i the forecasted and O_i observed rainfall amounts, in which i indicates a given grid cell. N is the amount of forecasts with lead time t, μ_F, and μ_O are the mean of the forecasted and observed rainfall and σ_F and σ_O are the standard deviation for the forecasted and observed rainfall.

Pearson's correlation coefficient was calculated in two ways. It was first calculated over the area-averaged precipitation, per lead time, and per event.

Second, Pearson's correlation coefficient was calculated per grid cell, per lead time. In this way, the spatial dependency was calculated. This was done over the whole study area (both Helmond and Laarbeek).

3.2.3 RMSE

The RMSE was calculated per lead time as:

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(F_i - O_i)^2}{N}} \qquad (3.4)$$

The RMSE was calculated over the area-averaged precipitation, per lead time, per event.

3.2.4 Bias

The bias was used to quantify overestimations and underestimations for the rainfall forecasts compared with the observed rainfall. The bias is calculated with:

$$bias = \frac{Fc_i}{Obs_i} \qquad (3.5)$$

in which Fc is the forecasted accumulated rainfall during one certain event at a given lead time and Obs is the observed accumulated rainfall and i indicated a certain event. The accumulated rainfall is calculated over the area-averaged rainfall. The bias is calculated for multiple lead times up to two hours.

A bias of 1 indicates that the total rainfall sum of the forecast is in perfect agreement with the observation, a bias smaller than 1 indicates underestimation and a bias greater than 1 indicates overestimation of the forecasts.

3.3 Comparison nowcasts and HARMONIE

The nowcasts were also compared to HARMONIE, so the skill of nowcasts could be computed relative to HARMONIE. This was done by using an area-averaged and grid-based approach. HARMONIE, PS-D and RM-DR were verified with the observed gauge-adjusted radar data. For this analysis Pearson's correlation coefficient (section 3.2.2) was calculated for lead times up to 6 hours, contingency tables (section 3.2.1) were calculated for different thresholds and lead times up to 6 hours and lastly, the RMSE (section 3.2.3) was calculated per grid cell. The results of the verification process for the three forecast methods were compared mutually.

Comparison of the nowcasts with HARMONIE was only done for a few events that fell between the period October 2017 and 2018 (see Table 2.2), because HARMONIE is only available for this period. The events used for this comparison were based on the issue times of HARMONIE and lead times up to 6 hours. Note that these selected events did not necessarily contain the most extreme ones. During some events, no precipitation was observed in Helmond and Laarbeek.

The spatial resolution of HARMONIE (2.5 km^2) differs from the observed gauge-adjusted radar data and the nowcasts (1 km^2), so HARMONIE was downscaled by averaging the coarser-resolution to the finer spatial resolution of the radar reference data. The temporal resolution of HARMONIE is 1 hour. For that reason the accumulated rainfall (mm) in one hour was calculated for the observed radar data and nowcasts.

3.4 Verifying model output

The urban hydrological model, described in section 2.6, was used to forecast sewage overflow volume. Only a selection of the sub-sewerage systems shown in Fig. 2.2 were investigated (see Table 2.3). Different model outputs were made by using either PS-D nowcast, RM-DR nowcast, or HARMONIE as input. The observed gauge-adjusted radar data were used as a reference run. The model output was compared mutually, and subsequently

with the reference run. The dependence of different rainfall characteristics on the performance of the urban hydrological model was investigated. These are the same as described in section 2.2.

The focus of this study is on quantifying the added value of nowcasts for operational water management. For that reason, the output of the model was not compared with measurements but with simulations using reference rainfall as input. In this way, uncertainties caused by the hydrological model will have limited effect on the conclusions since they impact all model runs. Evaluation of the performance of the forecasts as model input was done by computing the skill relative to the hydrological forecasts based on the observed gauge-adjusted radar data. To verify the output of the urban hydrological model, the bias, HR, and FA were calculated for each 5 min issue times between 5 and 65 min. E.g. for the event between 9.45-10.40 all issued forecasts between 8.40 and 9.40 were used. For the 24 hour events, a new issue time was used after a new hour of the event has passed. E.g. for the event starting at 07-10-2009 07:05 and ending at 08-10-2009 07:00, an issue time will be taken for 07:05 until 8:00, a new one for 08:05-09:00, a new one for 09:05-10:00 and so on.

The initial conditions were determined by running the model 9 days prior to the start of the selected precipitation event. The 9 days were used as Aa en Maas did the same in the model. This model run was performed with the gauge-adjusted radar data as input. After obtaining the initial conditions, the 1- or 24-hour event was used as input by using either output from one of the forecast methods or the reference data. After the event the model ran for another 12-hours with no precipitation input, to see the effect of an event on the overflow and the actual pump capacity in time.

In order to prevent flooding in urban areas, rainwater is drained away. The part of the pump capacity that is used to discharge collected rainfall (pump capacity (POC)) should be used optimally in case of a sewage overflow. In this way, the overflow can be prevented or reduced. First, the HR and FA were calculated over the total sewage overflow volume during one event, to see if the forecasts were able to predict the overflow. The used criterion was the occurrence of a sewage overflow. Secondly, the bias was calculated to investigate to what extent the overflow volume was predicted. Lastly, the RMSE was calculated over the POC for events where an overflow occurred and was forecasted.

For a fair comparison between HARMONIE and the

radar nowcasts, these model output comparisons were done by using nowcasts with the same issue time as HARMONIE.

3.5 Real-time control

Before investigating what the maximum anticipation time is, which is the time between forecast and occurrence, the reference runs were investigated. In this way, it was investigated if there was space for better use of the POC and how much sewage overflow volume could be reduced. The model ran with the observed radar rainfall as input, in order to check if there was a sewage overflow caused by the specific event. In the case of an overflow, a closer look at the POC and storage (mm) is necessary. Optimal use of the POC can avoid or reduce the amount of sewage overflow.

First, the reference run was analysed to see what the effect is of using the storage and pump capacity more efficiently. Second, the maximum skilful issue time and the amount of water (mm) that could be discharged during this time were analysed. If the forecast time is one hour, there is already one hour in time that can be gained. Without any forecast, the user can only measure the values after the event has passed. Based on the maximum issue time and the forecast time, the gain in anticipation time was determined.

Next, the extreme events in May and June 2016 and subsequent inundated areas were looked into. The potential of real-time control was investigated to see if the storage capacity and pump capacity could be used better in order to prevent or reduce sewage overflow.

Lastly, the performance of the nowcasts and HARMONIE, and output of the hydrological model were compared in order to analyse if there are any relations. This was done by comparing the bias and contingency tables, calculated for either the forecasts or forecasted hydrological output.

4 | Results

4.1 Performance nowcasts

In this section, the most important findings related to the performance of the radar rainfall nowcasts is presented. The performance of the nowcasts was interpreted by using a set of events and verifying it with the reference rainfall (observed gauge-adjusted radar data). The main focus was to identify the difference between the rainfall characteristics, seasons and between the algorithms in forecasting rainfall.

Figure 4.2a serves as an example of radar rainfall scans and two nowcast algorithms (PS-D and RM-DR) at different time steps. The event shown here is only used to give a good example of the difference between radar rainfall measurements and rainfall forecasts made by the algorithms. The nowcasts were both issued at 14:50 UTC on 05 May 2016.

From Fig. 4.2 it is visible that forecasting rainfall has some flaws. Both algorithms are able to capture the event and its movement to a certain extent (Fig. 4.2a). However, both algorithms are not able to predict the average amount of rainfall that fell over the urban area of Helmond in one hour, which was 9.22 mm. RM-DR only measured 3.13 mm during this event and PS-D measured a higher value of 5.16 mm (for the nowcast issued at 14:50) (Fig. 4.2b). RM-DR is meant to be able to capture the high-intensity rainfall centres (bright yellow pixels in Fig. 4.2a), while PS-D is characterized by capturing the mean of a large-scale field of rainfall. Consequently, PS-D results in more dissipation of the high-intensity fields. In Fig. 4.2a it is also visible that the nowcasting algorithms have more difficulty predicting the right location and intensity of the rainfall fields at higher lead times.

The used urban areas in this research only cover a small part of the radar image in Fig. 4.2a. The urban areas of Helmond and Laarbeek are small. Consequently, the hydrological systems of these areas are characterized by a fast response time. In addition, smaller areas are more sensitive for mislocation of the rainfall because it is easier to miss rainfall if it mislocated. For those reasons, predicting the right location of the high-intensity rainfall centres is very important.
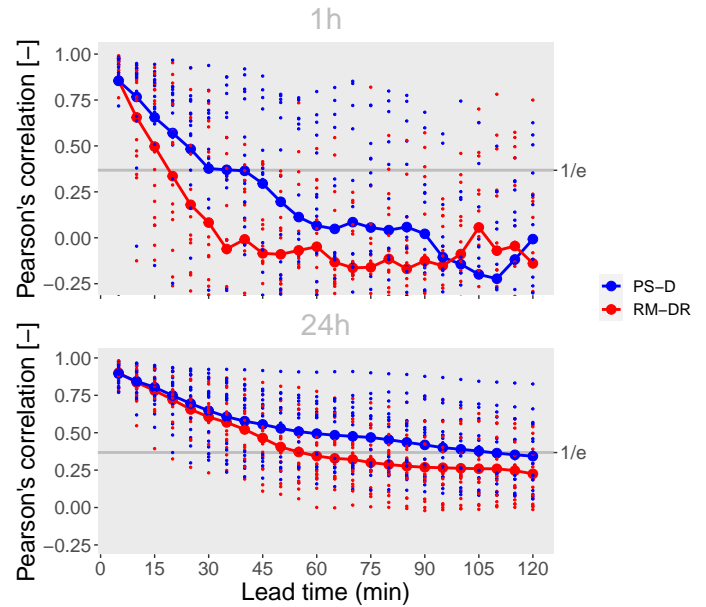

Figure 4.1: Pearson's correlation calculated over the catchment averaged rainfall of Helmond per lead time, averaged over all events for 1-hour and 24-hour duration events. Pearson's correlation is calculated over both algorithms. The grey line is the threshold ($1/e$), indicating the minimum correlation for a nowcast to be skilful. The dots indicate the variability per event.

4.1.1 Dependency on rainfall characteristics

The lead time for which the radar nowcasts are still skilful varies between rainfall duration and algorithm (Fig. 4.1). Pearson's correlation coefficient was calculated over the catchment averaged precipitation in Helmond, per event, per lead time. Second, the average of all events was calculated and shown in Fig. 4.1. This resulted in a maximum skilful lead time for 1-hour duration events of around 19 min or 37 min and for 24-hour durations, it is around 56 min or 108 min for respectively RM-DR and PS-D. For both durations, PS-D results in the longest skilful lead time and is almost twice as long as RM-DR. It can be stated that increasing lead times decreases the quality of the rainfall forecast. Calculating Pearson's correlation over Laarbeek resulted in similar results and can be found in Appendix A.1.

There is much variation in the correlation between the events, indicated by the dots in Fig. 4.1. The correlation of 1-hour duration events, which often consists of convective showers (short, but high intense precipita-
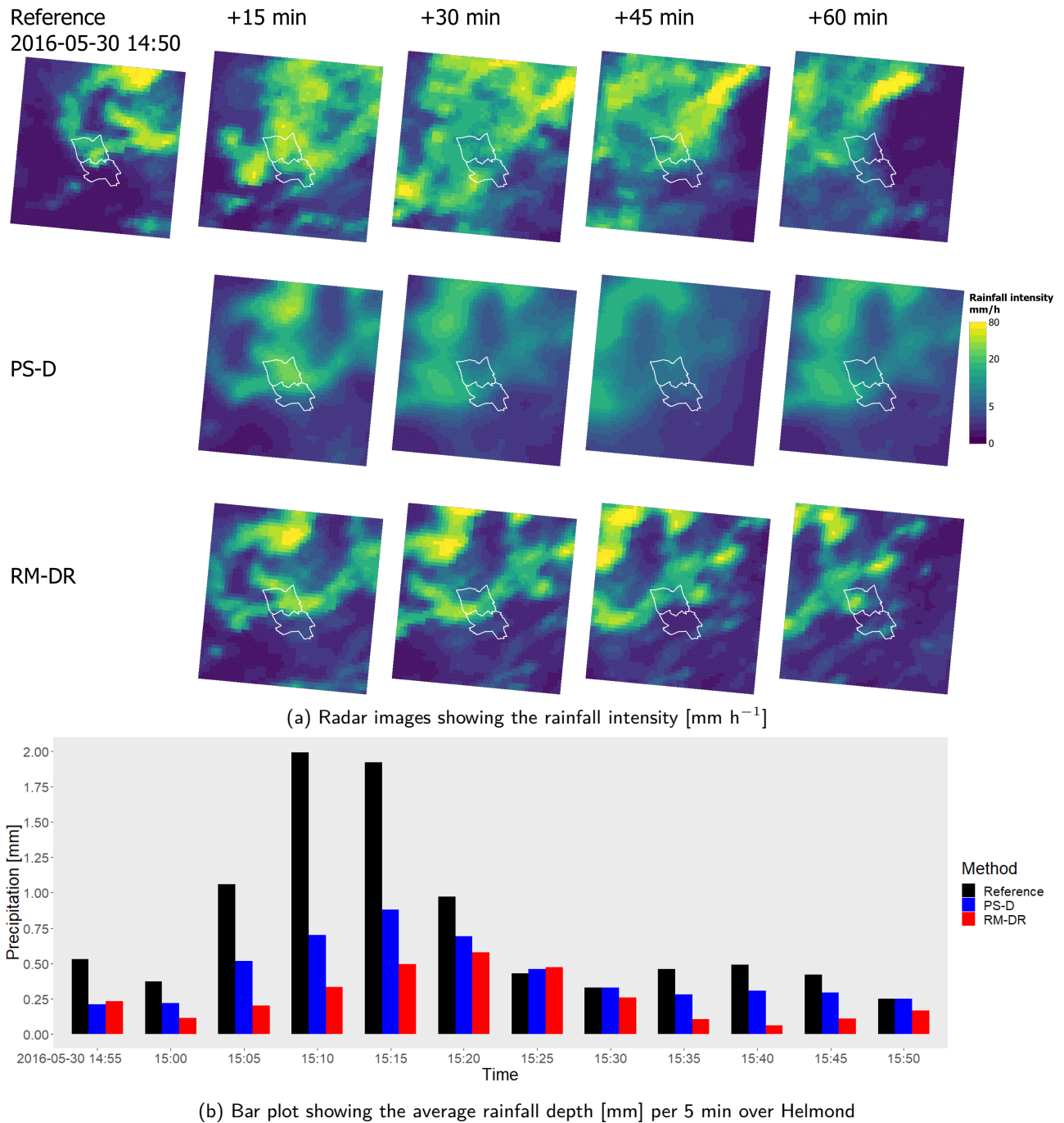
(a) Radar images showing the rainfall intensity [mm h^{-1}]

(b) Bar plot showing the average rainfall depth [mm] per 5 min over Helmond

Figure 4.2: Example of the reference radar, PS-D and RM-DR nowcast for the Aa catchment. (a) Rainfall intensity (mm h^{-1}) based on 5-min accumulation is shown here and (b) the average rainfall depth [mm] per 5 min over Helmond. The event took place on 2016-05-30 (1-hour duration). The issue time shown here is 14:50 UTC. An average of 9.22 mm fell in one hour over Helmond, while RM-DR forecasted 3.13 mm and PS-D 5.13 mm in one hour. White contour lines indicate the study area (see also Fig. 2.1)

tion events), varies more than the 24-hour events (often characterized by long-duration storms). Convective showers are small scale processes, leading to more variation in the forecasts of these type of events. The events with a longer duration are often stratiform events, which are larger-scale events that are more persistent. Consequently, the predictive skill for longer duration events is expected to be higher.

Figure 4.3 shows the distribution of the RMSE, based on the comparison of precipitation forecasts against measured radar fields. The RMSE was calculated per season over the whole study area of Helmond, for all events. The average of lead times between 5-30 min, 35-60 min, 65-90 min and 95-120 min were

Table 4.1: Bias calculated over the accumulated rainfall (mm) during the event duration (1-hour or 24-hours), averaged over all events, averaged over lead times of 5-30 min, 35-60 min, 65-90 min and 95-120 min.

Algorithm	Duration	Bias Helmond				Bias Laarbeek			
		5-30 min	35-60 min	65-90 min	95-120 min	5-30 min	35-60 min	65-90 min	95-120 min
PS-D	1 hours	0.69	0.44	0.3	0.21	0.77	0.46	0.28	0.19
	24 hours	0.76	0.6	0.53	0.44	0.8	0.6	0.51	0.43
RM-DR	1 hours	0.76	0.38	0.21	0.16	0.81	0.45	0.25	0.16
	24 hours	0.86	0.76	0.69	0.66	0.89	0.76	0.72	0.72

taken. The RMSE increases over longer lead times. A substantial difference in the performance of the forecasts between seasons is clearly visible. The largest error is during summer, with an average RMSE of 0.72 mm and 0.61 mm for respectively RM-DR and PS-D (Fig. 4.3). During winter the RMSE is considerably lower with a value of 0.39 mm and 0.36 mm for respectively RM-DR and PS-D. At the same time, there is more spread between the events that took place in the summer than in winter. The events during spring contain the most outliers (black dots in Fig. 4.3).

In the Netherlands, there is a variation between the precipitation types during the seasons. Summers are characterized by high intense rainfall of short duration, while winters have a lower rainfall intensity but the duration is often longer. Consequently, it is expected that capturing precipitation events in winter is more accurate.

Table 4.1 shows that in all cases the bias has a value lower than 1, indicating that both nowcast algorithms have a tendency to underestimate the total accumulated rainfall that fell during one event. The bias was averaged over all events, calculated per algorithm and event duration (1 hour or 24 hours) for both Helmond and Laarbeek. In all cases, the underestimation of the nowcasts increases with longer lead times. The bias for events with a duration of 24 hours is closer to 1, indicating that the nowcasts are more representative for 24-hour events than for 1-hour events. For shorter lead times, PS-D underestimates the total accumulated rainfall more than RM-DR. However, for events with a duration of 1 hour, PS-D becomes better than RM-DR with increasing lead times. For events with a duration of 24 hours, RM-DR scores better for all lead times than PS-D.

4.1.2 Spatial dependency of nowcast algorithms

For urban areas it is important to have a high spatial accuracy of rainfall. For that reason, Pearson's correlation
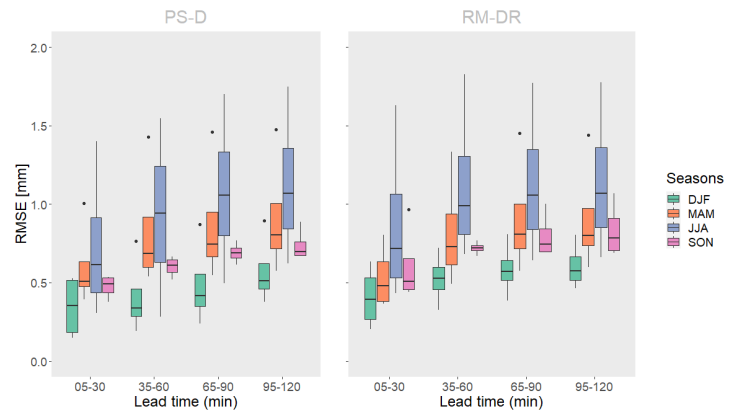

Figure 4.3: RMSE (calculated per 5-min time step) per season for all events with a duration of 1 hour over the catchment averaged rainfall of Helmond, averaged over lead times of 5-30 min, 35-60 min, 65-90 min and 95-120 min. Black dots indicate outliers.

is calculated per grid cell for different events.

In Fig. 4.4 the spatial dependency is made visible in terms of Pearson's correlation coefficient, averaged for lead times of 5-30 min and 35-60 min. Pearson's correlation was calculated over the grids in and around the study area. Both events shown have a 1-hour duration. In Appendix Figs. A.3 and A.4 the corresponding radar images of the events can be found.

For the averaged lead times of 5-30 min and 35-60 min both PS-D and RM-DR are in agreement about the spatial dependency of the area shown in Fig. 4.4a, while for Fig. 4.4b there is much difference between PS-D and RM-DR. In addition, comparing the two events, Pearson's correlation varies much per location. For both events, RM-DR results in a lower Pearson's correlation coefficient than PS-D, which is also in agreement with the results of Fig. 4.1. Over longer lead times, Pearson's correlation coefficient decreases. However, the extent in which Pearson's correlation decreases over longer lead times differs substantially between the events.

For most grid cells for the event at 2017-08-30 17:55 (Fig. 4.4a), Pearson's correlation coefficient is less than the threshold, and thus not seen as skilful. In contrary,

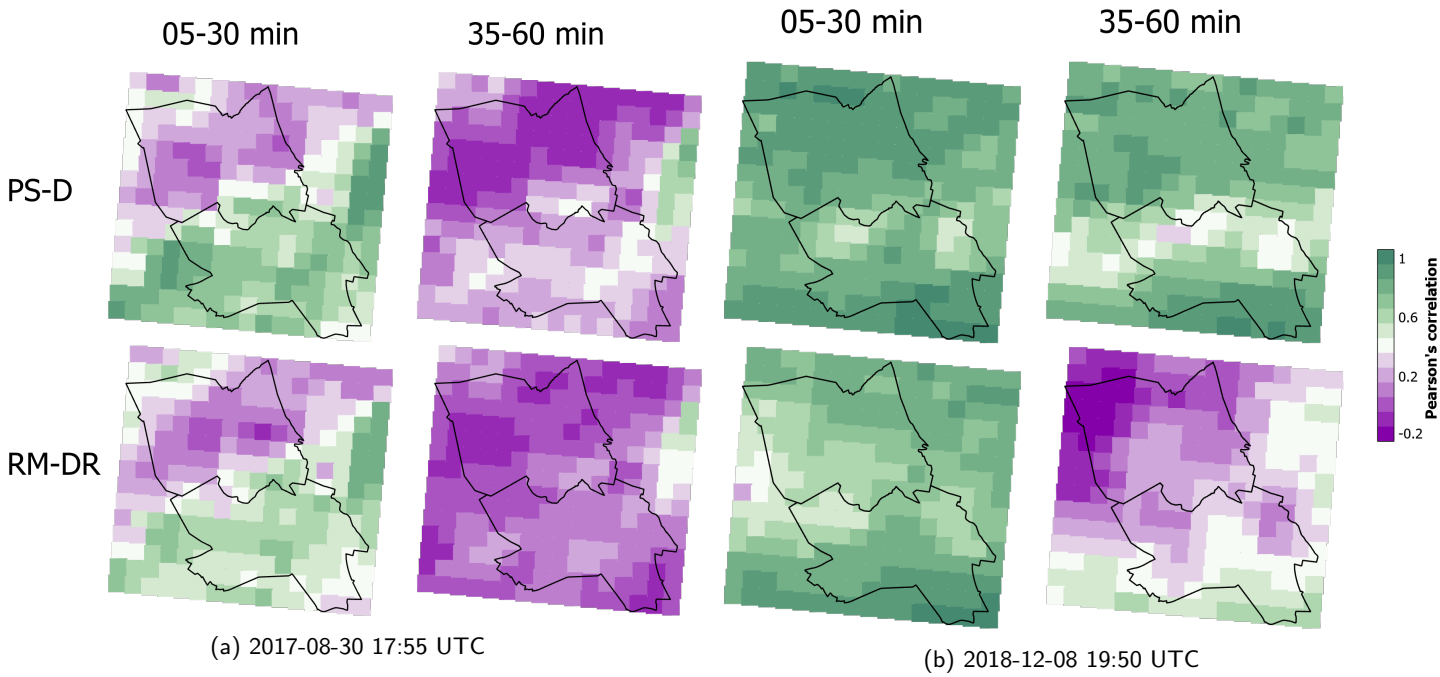(a) 2017-08-30 17:55 UTC

(b) 2018-12-08 19:50 UTC

Figure 4.4: Pearson's correlation per grid cell for the 1-hour event ending at 2017-08-30 17:55 UTC (a) and event ending 2018-12-08 19:50 UTC (b), averaged over lead times of 5-30 min and 35-60 min.
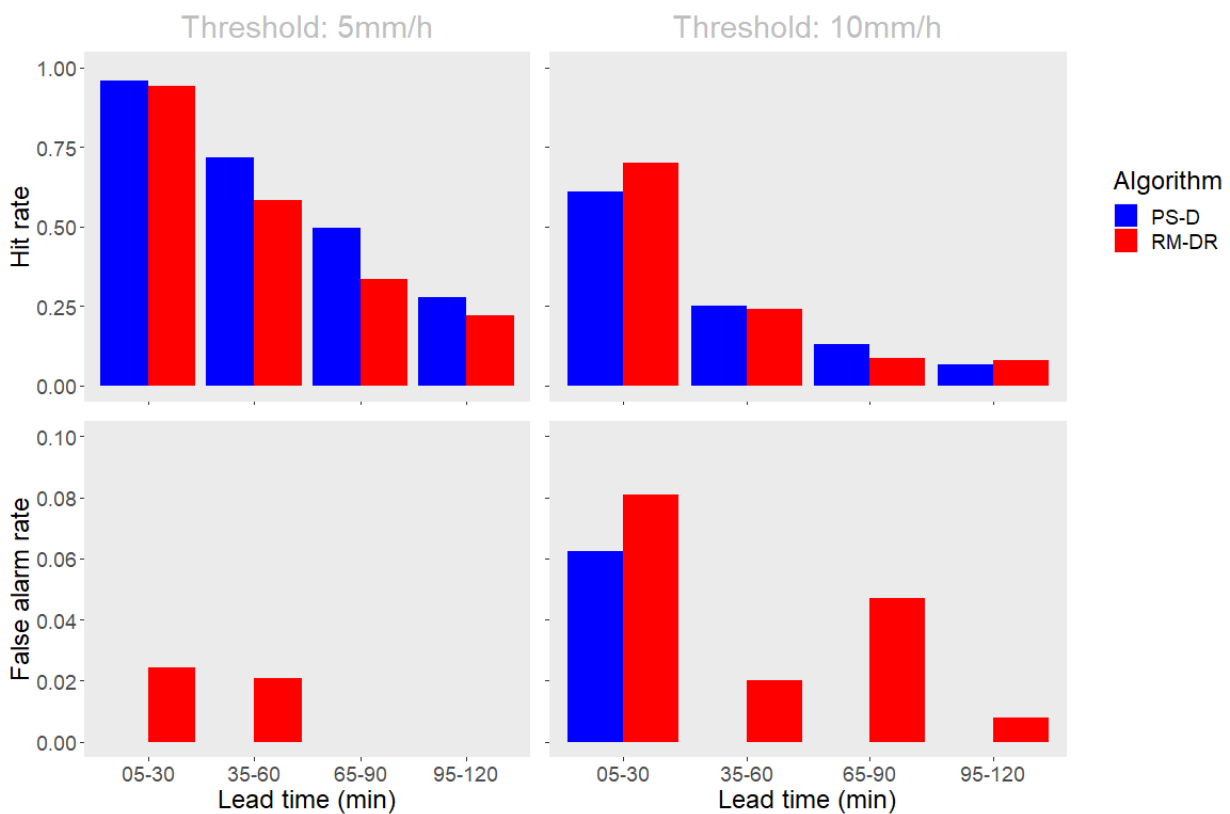

Figure 4.5: Hit rate (HR) and false alarm rate (FA) calculated over the urban area of Helmond. The HR and FA are calculated over the accumulated rainfall in 1 hour, by using a threshold of 5 mm h^{-1} or 10 mm h^{-1} (see Appendix A.4 Table A.1 for number of times that these thresholds were reached by the reference rainfall). All events were taken together, and averaged over lead times of 5-30 min, 35-60 min, 65-90 min and 95-120 min. Note the difference in scale on y-axis between the HR and FA.
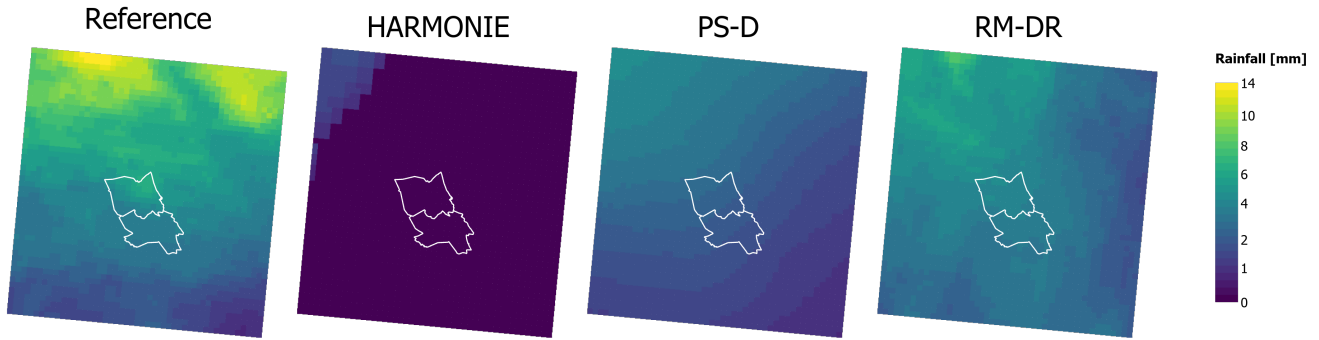
Figure 4.6: Radar image, with the total amount of precipitation (mm) that fell in one hour, for the event of 2018-04-30, ending at 01:00 (issued at 00:00 UTC). The reference radar, HARMONIE, PS-D and RM-DR nowcasts for the Aa catchment are shown. An average amount of 3.24 mm feel in one hour over Helmond, while HARMONIE forecasted 0 mm, PS-D 1.77 and RM-DR 3.20 mm. White contour lines indicate the study area (see also Fig. 2.1).

the event at 2018-12-08 19:50 (Fig. 4.4b) is in the case of PS-D seen as skilful for both lead times between 5-30 min and 35-60 min. However, in the case of RM-DR, the forecast is not seen as skilful for every grid cell, especially for lead times between 35-60 min Pearson's correlation is smaller than 0. This indicates that per event there is much variation in the skill of the nowcasts.

Accurate rainfall intensity forecasting is essential for predicting the maximum discharge and consequently real-time control (RTC). Especially high intense precipitation events can cause urban floods when the drainage capacity is insufficient to deal with the rainfall. For that reason, the hit rate and false alarm rate were calculated over events with a duration of 1 hour (Fig. 4.5). The used thresholds are 5 mm h^{-1} and 10 mm h^{-1}. These thresholds were determined by taking into account which rainfall depths (mm) led to an overflow according to the hydrological model described in section 2.6. See Appendix A.4 Table A.1 for the number of times that these thresholds were reached by the reference rainfall.

Figure 4.5 shows the HR and FA for both PS-D and RM-DR, determined for a threshold of 5 mm h^{-1} and 10 mm h^{-1}. Both nowcasts are able to predict the rainfall depth to a certain extent. There is a substantial difference between the HR with a threshold of 5 mm h^{-1} and 10 mm h^{-1}. The HR is lower in the case of a threshold of 10 mm h^{-1}. This indicates that both RM-DR and PS-D are able to predict rainfall, but are less skilled to predict the intensity of the rainfall events accurately. PS-D is more skilled (has a higher HR) than RM-DR for a threshold of 5 mm h^{-1}.

For larger lead times the HR decreases for both nowcasts. This is also in agreement with Table 4.1, which shows that with larger lead times the precipitation is more underestimated.

For thresholds of 5 mm h^{-1} the FA for PS-D is 0, indicating that there are no false alarms present in these nowcasts. RM-DR has a FA of around 0.02 for lead times of 5-30 min and 35-60 min. This is not desirable as it sends out false warnings. However, the FA is very low in this case, indicating that only in a few cases a false warning was send out. For higher thresholds (10 mm h^{-1}), the number of false alarms increases.
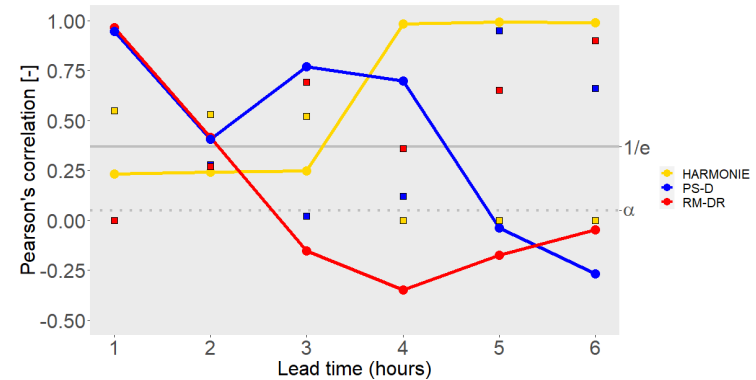

Figure 4.7: Pearson's correlation calculated over the catchment averaged rainfall of Helmond per lead time, averaged over 9 events. Pearson's correlation is calculated over both nowcasting algorithms and HARMONIE. The grey line is the threshold (1/e), indicating the minimum correlation for a nowcast to be skilful. The dotted grey line indicates the significance level of 0.05, the squares show the p-value per forecast per lead time.

4.2 Relative comparison between forecasts

This section covers the most important results of the comparison in performance between the forecast methods (HARMONIE, PS-D and RM-DR). A different set
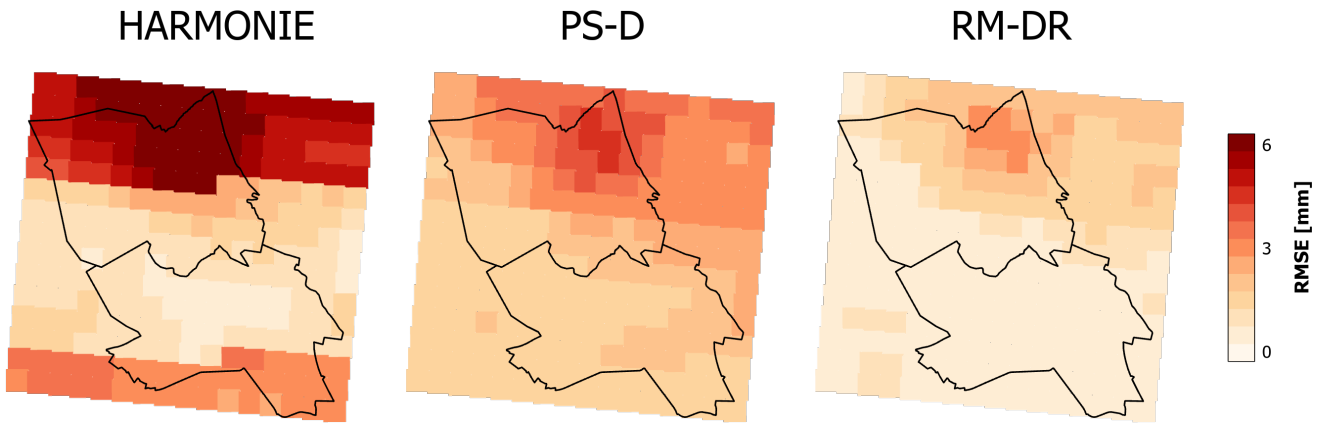
HARMONIE   PS-D   RM-DR

Figure 4.8: RMSE per grid, calculated over the total amount of precipitation (mm) that fell in one hour. The event of 2018-04-30 ending at 01:00 UTC is taken as an example here. Corresponding radar images containing the rainfall depth can be seen in Fig. 4.6.

of events is used than described in section 4.1. The forecasts were verified with the reference rainfall.

Figure 4.6 serves as an example to compare a radar rainfall scan with forecasts made by HARMONIE, PS-D and RM-DR. The event of 2018-04-30 ending at 01:00 (issued at 2018-04-30 00:00 UTC) is shown here. It is visible that PS-D and RM-DR are both able to capture the event to some extent, while HARMONIE is not able to capture the rainfall event at all. The underestimation of the total rainfall forecasted by HARMONIE is too severe, consequently, the spatial variation of rainfall is not visible.

The average amount of rainfall that fell over the urban area of Helmond was 3.24 mm. HARMONIE did not forecast any rainfall. PS-D underestimated the average rainfall, and only forecasted 1.77 mm. However, RM-DR forecasted 3.20 mm of rainfall, which is very close to the reference. All three models are not able to predict the core of the rainfall event (upper part of Fig. 4.6).

Pearson's correlation coefficient was calculated for HARMONIE, PS-D and RM-DR over the area-averaged precipitation in Helmond for lead times up to 6 hours (Fig. 4.7). The lines are not smooth; some bumps are visible. It is expected that this is due to the small number of events (only 9) that are used for this analysis. Consequently, no clear relation between Pearson's correlation coefficient and lead times for the different forecast methods is visible. In addition, not all correlation coefficients were significant, as the p-value was in most cases higher than 0.05 (see squares in Fig. 4.7).

While PS-D and RM-DR seem skilful up to lead times around 2 hours, HARMONIE is not. With longer lead times, HARMONIE becomes more skilful (after

3 hours), while RM-DR and PS-D lose their skilfulness over time. RM-DR is not seen as skilful for lead times of more than 2 hours, PS-D is not seen as skilful at lead times of 5 and 6 hours.

Note that this comparison was only done for 9 forecasts and not during all forecasts (extreme or heavy) precipitation was observed. It is expected that it is easier to forecast no precipitation, so that can cause the differences in skilful lead times for RM-DR and PS-D in Fig. 4.7 compared to Fig. 4.1.

The spatial dependency of forecasting precipitation in terms of RMSE for the three forecast methods is clearly visible in Fig. 4.8. The RMSE based on the forecast issued at 2018-04-30 00:00 UTC is shown here. All methods are in agreement about the spatial dependency, with the upper part in Fig 4.8 having the highest RMSE. At the upper part the highest amount of rain fell during the hour (Fig. 4.6). All forecast methods have problems with capturing the right amount of rainfall. Over the whole area, RM-DR results in the lowest RMSE compared to PS-D and HARMONIE.

In Fig. 4.9 the hit rate (HR) and false alarm rate (FA) are shown for the three forecast methods (HARMONIE, PS-D and RM-DR) over different lead times up to 6 hours. Thresholds of 0.5 mm h^{-1} and 1 mm h^{-1} were used, as the 9 events used in this analysis contained lower rainfall intensities. For PS-D it shows that the proportion of events that were correctly forecasted decreases for longer lead times. For RM-DR and HARMONIE the pattern is less clear over different lead times. Both PS-D and RM-DR have a higher HR for lead times up to 4 hours for both thresholds than HARMONIE (Fig. 4.9).
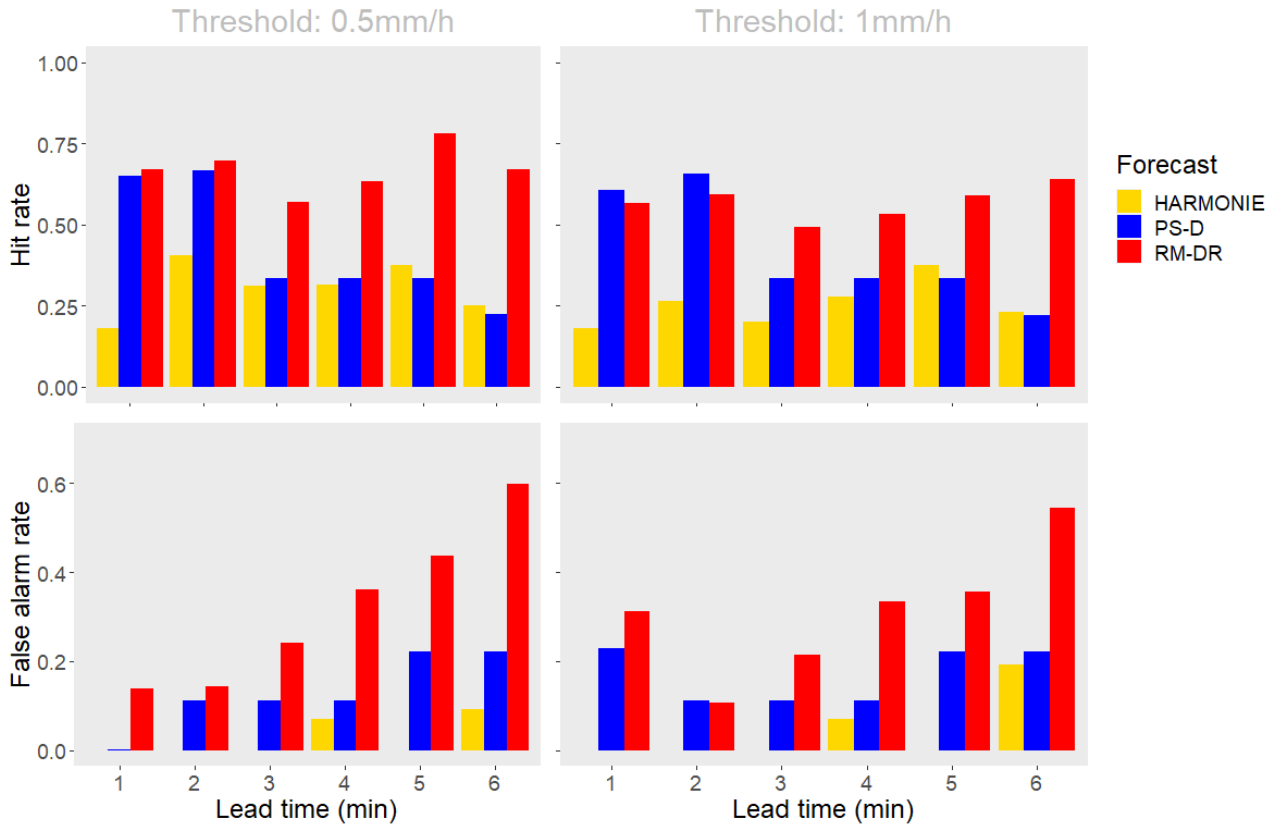
Figure 4.9: Hit rate (HR) and false alarm rate (FA), calculated over the urban area of Helmond. A threshold of 0.5 mm h^{-1} and 1 mm h^{-1} was used. See Appendix A.4 Table A.2 for number of times that these thresholds were reached by the reference rainfall. All events are taken together. Note the difference in scale on y-axis between the HR and FA.

The difference in proportion of events that were correctly forecasted between PS-D and RM-DR increases with longer lead time.

The proportion of events that were forecasted, but not measured, increases with longer lead times for a threshold of 0.5 mm h^{-1} for both PS-D and RM-DR. For HARMONIE only at lead times of 4 or 6 hours false alarms were present. For all lead times, HARMONIE results in a lower FA than RM-DR and PS-D.

4.3 Performance nowcasts in a simple sewage overflow model

In this section the differences between the sewerage sub-areas are presented. Next, most important findings related to the performance of the radar rainfall nowcasts as input for a simple sewage overflow model are shown. Results are shown for different locations on the sewerage network around the municipalities of Laarbeek and Helmond (See Table 2.3 for the locations).

4.3.1 Sewerage sub-areas

The chosen sewerage sub-areas for this research differ in storage capacity, service area and type of sewerage system (Table 2.3). Consequently, the frequency and volume of a sewage overflow differs per location. In addition, the waste emission during an overflow differs per sewerage system, because improved combined sewerage systems have a storage settlement basin. However, the lower waste emission is not visible in the output of the model.

The locations HH9, LBK8, LBK11, LLK15 and LLK16 (see Fig. 2.2 for sewerage locations and Table 2.3 for characteristics) have similar maximum storage capacity (varies between 6.43 and 8.92 mm). Consequently, it is expected that the frequency of the sewage overflow will be around the same. However, the service area has a larger variety (between 4.06 and 58.94 ha). This will affect the volume of the sewage overflow. HHC1, HHC2 and HHH8 have larger storage capacities (varying between 10.59 and 21.96 mm). Consequently, less sewage overflow will occur. Especially for the events of 1-hour,
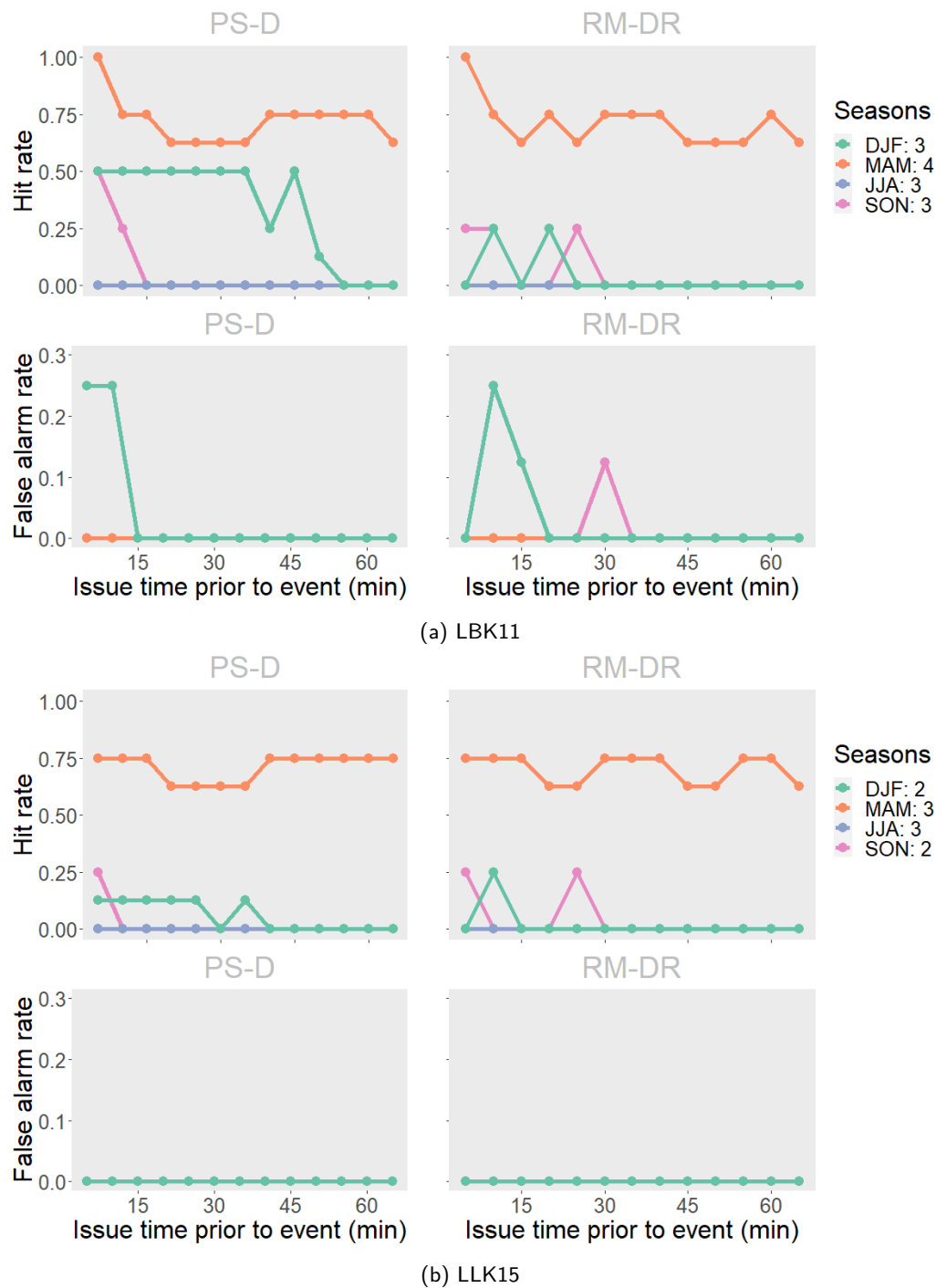
(a) LBK11

(b) LLK15

Figure 4.10: Hit rate (HR) and false alarm rate (FA) per season for all events over the sewerage sub-area LBK11 (a) and LLK15 (b), averaged over all events for 1-hour duration event. HR and FA are shown for issue times prior to the start of the event. Number behind the season indicates the number of occurrences of sewage overflow that took place in the reference run. Note the difference in scale on y-axis between the HR and FA.

the frequency of a sewage overflow varies much between the locations (Table 4.2). HHC1 has the largest storage capacity, resulting in a lower frequency of overflows (2 times for 1-hour events). However, the service area is larger, resulting in larger volumes of overflow during the occurrence (a total overflow volume of 9833 m^3 during the 1-hour events). On the other hand, LBK11 has the lowest storage capacity, and also the lowest service area. The frequency is larger, but the volume of an overflow will be smaller (13 times for 1-hour events, with a total overflow volume of 3712 m^3).
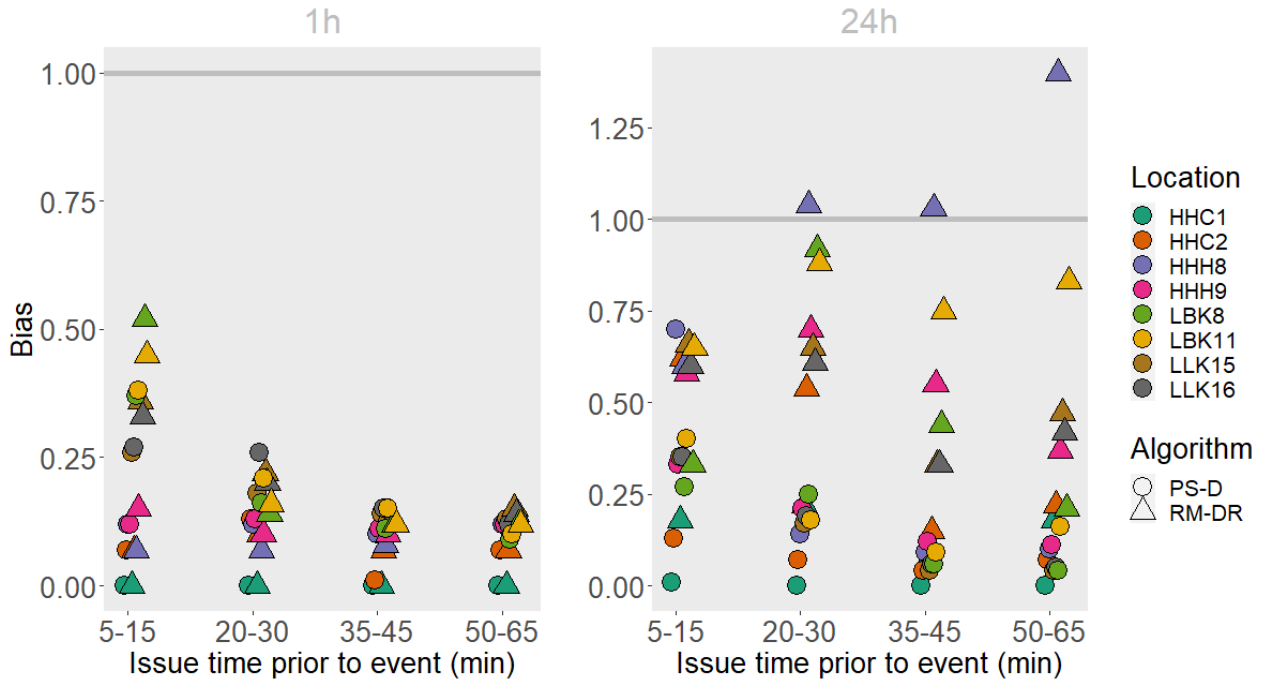
Figure 4.11: Bias calculated over the total amount of sewage overflow during the event duration (1 hours or 24 hours), averaged over all events and averaged over issue times of 5-15 min, 20-30 min, 35-45 min and 50-65 min prior to the start of the event. The sewerage sub-areas shown are HHC1, HHC2, HHH8, HHH9, LLK15, LLK16, LBK8 and LBK11. Grey line indicates a bias of 1. Note that the scale on the y-axis is different.

4.3.2 Dependency on rainfall characteristics

The hit rate (HR) and false alarm rate (FA) were calculated per season for the sub-areas LBK11 and LLK15 Fig. 4.10. Those areas were chosen because the storage capacity is small, consequently more sewage overflow took place here. The used criterion is a sewage overflow, so the volume of overflow that took place will be ignored.

From Fig. 4.10 it can be seen that there is a substantial difference between the seasons. The highest HR is during spring, in which the HR is around 0.75 for all issue times. This is followed by winter and autumn. However, for longer issue times the HR is 0 for winter and autumn, indicating that the sewage overflow is not predicted at all. The forecasts in summer result in the lowest HR with a value of 0 for all issue times, while three out of four events in summer resulted actually in a sewage overflow (for LBK11). During winter PS-D has a higher skill than RM-DR, with PS-D having a HR of 0.5 for issue times until 35 min (for LBK11). For the location LLK15, PS-D also has a higher skill than RM-DR for the winter events for most issue times.

For both locations PS-D has a FA of 0 for autumn, spring and summer, indicating that the forecasts

do not result in any false alarms. PS-D only results in a false alarm for one event in winter, for issue times until 10 min, while RM-DR has two events in winter and one event in autumn that results in a false alarm. The events that occur in spring have both the highest HR, and one of the lowest FA.

Per sub-area and event, the bias was calculated over the total amount of sewage overflow that took place (Fig. 4.11). The bias was averaged over all events and averaged over issue times of 5-15 min, 20-30 min, 35-45 min and 50-65 min prior to the event.

In most cases the bias is below 1 (Fig. 4.11), indicating that in general the radar nowcasts underestimate the total amount of overflow that took place. Only some of the 24-hour events predicted by RM-DR resulted in an overestimation.

In general, RM-DR results in an overflow volume that is closer to the reference than PS-D for issue times of 5-15 min. For longer issue times, PS-D becomes relatively more skilful than RM-DR. For events with a duration of 24 hours the difference between RM-DR and PS-D increases, with RM-DR being more skilful than PS-D for all issue times. For events with a duration of 1 hour, RM-DR and PS-D have a similar bias for issue

times larger than 15 min.

Between the sewerage sub-areas there is a large difference in bias (e.g. varying between 0 and 0.52 for issue times ranging from 5 to 15 min). HHC1 results in the lowest values (Fig. 4.11). A zero indicates that an overflow occurred but the forecasted output did not simulate any overflow at all. HHC1 has the largest storage capacity, so it takes longer before this maximum storage is reached. Consequently, underestimating the rainfall depth (mm) results quicker in a discrepancy between the measured and forecasted overflow. In addition, note that the sub-systems with a different prefix (e.g. HHC) have different precipitation input, for which the skill of the forecasts can vary substantial spatially (see also Figs. 4.4a and 4.4b for the spatial difference in forecasting skill).

Figure 4.11 indicated that in most cases the overflow volume is underestimated. Consequently, this can influence the predicted pump capacity that is used to discharge collected rainfall (POC (mm/h)). For that reason, the RMSE was calculated over the POC for the events where an overflow occurred (Fig. 4.12).

For events with a duration of 1 hour, RM-DR has a higher RMSE for all issue times, while for events with a duration of 24 hours, the difference between PS-D and RM-DR is smaller, with PS-D having a higher RMSE.

There is a larger increase in RMSE for longer issue times for the 1-hour events than for the 24-hour events. The variability between the events is larger for 1-hour events than for 24-hour events. For the 1-hour events, the RMSE varies between 0 and $\pm$ 0.6, while the 24-hour events have a maximum RMSE of 0.35 and 0.45 for respectively PS-D and RM-DR. Note that there is a maximum pump capacity of 0.7 mm h^{-1}, which limits the RMSE to a maximum value of 0.7.

4.4 Comparison between forecasts in a hydrological model

In this section the forecasted sewage overflow, resulted from the sewage overflow model, forced by HARMONIE, PS-D or RM-DR as input, are compared relative to each other.

The number of overflows that took place per lead time was little (see Appendix A.5 Table A.3 for the number of overflows that were predicted by the reference run per location). For that reason, the hits and misses were determined instead of the hit rate and false alarm rate.
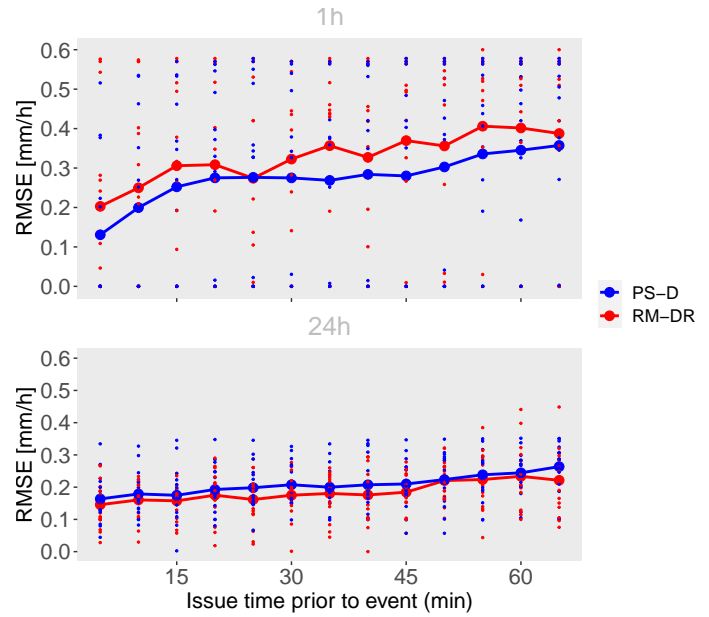

Figure 4.12: RMSE calculated over the POC (mm/h) of LBK11 per issue time prior to the start of the event, averaged over all events where a sewage overflow occurred for 1-hour and 24-hour duration. RMSE is calculated over both algorithms. The dots indicate the variability per event.

There is a large difference between the forecasts in bias, hits and misses (Fig. 4.13). There is also a substantial difference in the bias between the locations (e.g. varying between 0.63 and 15.81 for HARMONIE). HARMONIE is not able to forecast any of the overflows for any of the events for lead times up to 3 hours (Fig. 4.13). For longer lead times, the number of misses decreases substantially. For lead times of 4 hours, HARMONIE mostly overestimates the sewage overflow volume, while for lead times of 6 hours the volume is underestimated. In contrary, PS-D and RM-DR are able to forecast the occurrence of an overflow for lead times of 1 hour. PS-D underestimates the overflow volume for all locations, while RM-DR overestimates the overflow volume. For longer lead times PS-D and RM-DR only results in misses. For lead times of 1 hour, RM-DR results in a bias close to 1 for some locations, indicating that the forecasted model output and the reference are in good agreement. The number of false alarms is higher for RM-DR than for the other two methods.

4.5 Potential for real-time control

This section focusses on the initial conditions of the sewerage systems, prior to a sewage overflow, to investi-
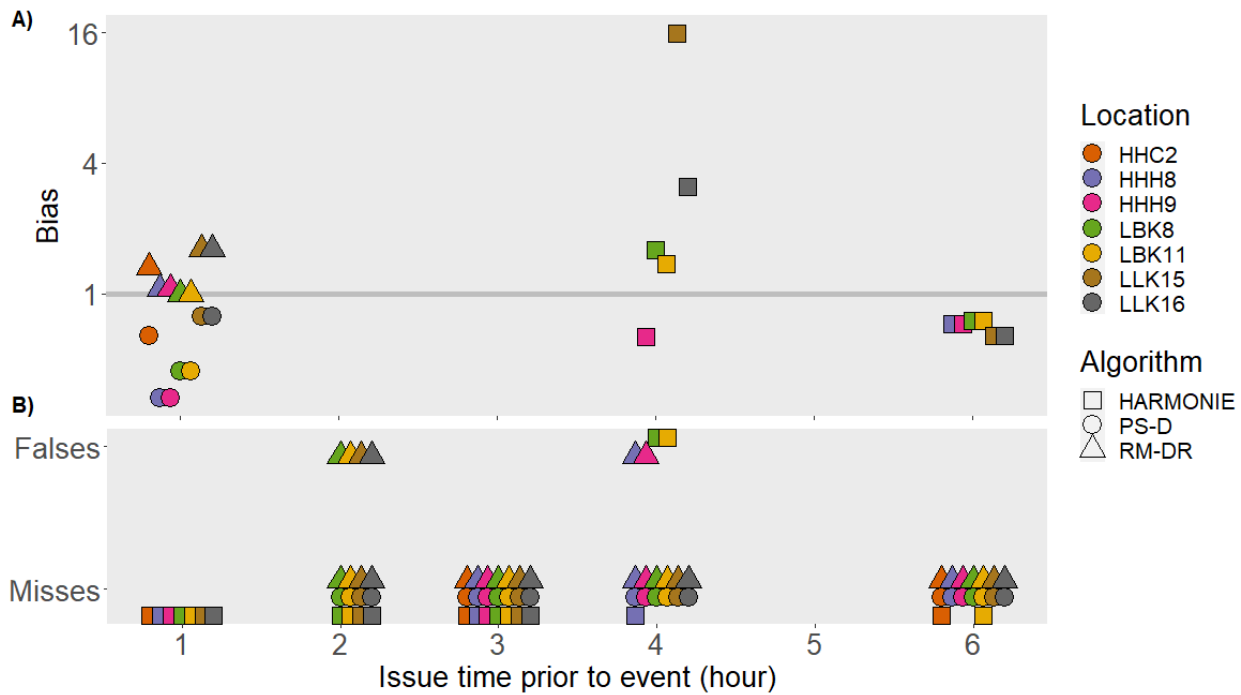
Figure 4.13: A) Average bias, B) misses and false alarms for forecasting sewage overflow amounts for all events, calculated over lead times up to 6 hours for HARMONIE, PS-D and RM-DR. See Appendix A.5 Table A.3 for the amount of overflows that were predicted by the reference run per location

gate the possibilities for RTC. In addition, some specific high-intensity rainfall events in May and June 2016 were analysed. These events resulted in floods in and around Helmond, which caused damage in the built-up areas. In addition, a 1-hour event in spring is analysed, as results showed that events in spring have the highest hit rate. Those events were investigated, to see if RTC could have prevented the part of the floods which resulted from sewage overflow.

Table 4.2 shows the initial conditions per event and sewerage sub-area. In some cases the POC prior to the event was 0.35 (due to the used criterion of the model, see section 2.6) and there was still storage available to discharge. This indicates that better use of the POC is possible, and could reduce the amount of water stored in the sewerage system. However, this is often a maximum of 0.35 mm due to the used thresholds in the model (Fig. 2.3). In addition, the maximum POC is often reached before the overflow happened. Or, in some cases, the rainfall depth (mm) in one hour is higher than the maximum storage capacity, resulting in immediate sewage overflow. In these cases RTC does not have any effect.

The 1-hour event of 30-05-2016 ending at 15:50 UTC resulted in a sewage overflow for all sub-locations.

Prior to the event there was already some rainfall. Consequently, the POC was at its maximum and part of the storage capacity of the sewerage system was already used. The POC started working and reached its maximum 9 hours prior to the specific event, indicating that for this specific event it was not possible to prevent or reduce the sewer overflow.

The 24-hour event of 31-05-2016 ending at 00:00 UTC, resulted also in a sewage overflow for all locations. However, prior to the event there was no rainfall, resulting in initial conditions of 0 for both POC and storage. During the event itself much rainfall fell which caused the sewage overflow. For some locations the rainfall depth that fell during 1 hour was higher than the maximum storage capacity, resulting immediately in a sewage overflow. Consequently, RTC could not have prevented an overflow. For HHC1 a better use of the POC could have prevented the sewage overflow. The POC could have been 0.7 mm h^{-1} instead of 0.35 mm h^{-1}, 10 hours before the overflow occurred. Precipitation nowcasts as input of the model did not result in any forecasted overflow, but could still be helpful in pumping the water away on time, as the rainfall nowcasts forecasted some rainfall.

The 24-hour event of 2016-06-02 ending at 03:00 UTC resulted in a sewage overflow at 19:00, 20:00 and

Table 4.2: The initial conditions (pump capacity (POC) and storage (S) [mm]) per event and sewerage sub-location for events with a duration of 1 hour. No information indicates that there was no overflow for that specific event. Red numbers indicate that the maximum storage is reached prior to the event.

	Event (1hour)	HHC1		HHC2		HHH8		HHH9		LBK8		LBK11		LLK15		LLK16	
	End [UTC]	POC	S	POC	S	POC	S	POC	S	POC	S	POC	S	POC	S	POC	S
Autumn	12-09-2008, 10:40					0	0	0	0	0	0	0	0				
	07-10-2009, 18:30							0.7	3.31	0.7	2.38	0.7	2.38	0.7	3.33	0.7	3.33
	22-10-2013, 20:40							0	0	0	0	0	0	0	0	0	0
	15-09-2016, 21:25																
Spring	15-05-2008, 19:00	0.35	0.09	0.35	0.09	0	0	0	0	0.35	0.2	0.35	0.2				
	30-05-2016, 15:50	0.7	14.86	0.7	12.25	0.7	9.94	0.7	7.14	0.7	7.77	0.7	6.43	0.7	8.76	0.7	7.69
	10-04-2018, 21:50			0.35	0.94	0.35	1.62	0.35	1.62	0.35	0.37	0.35	0.37	0.35	0.38	0.35	0.38
	30-04-2018, 00:15			0.7	9.77	0.7	8.94	0.7	7.8	0.7	7.77	0.7	6.43	0.7	8.92	0.7	7.85
Summer	12-07-2010, 09:50					0	0	0	0	0	0	0	0	0	0	0	0
	14-07-2010, 16:50					0	0	0	0	0	0	0	0	0	0	0	0
	27-07-2013, 09:40																
	30-08-2017, 17:55									0	0	0	0	0	0	0	0
Winter	05-02-2008, 23:45									0	0	0	0	0.35	0.14	0.35	0.14
	10-12-2009, 03:10									0.35	1.36	0.35	1.36			0.35	1.68
	03-01-2012, 16:55							0	0								
	08-12-2018, 19:50									0	0	0	0	0	0	0	0

01:00 UTC (Fig. 4.14). The storage was 0 at 16:00. Only a better use of the POC at 17:00 could have reduced the sewer overflow with 152.36 m^3 (0.35 (POC) × 43.53 (service area, ha) × 10 = 350 m^3) for HHH9. However, this is two hours prior to the first overflow (19:00). RM-DR predicted an overflow at 18:00 and 19:00 UTC for issue times of 15 and 30 minutes. If RTC is solely based on a forecasted overflow, than an early warning signal could have been sent out due to the forecasted overflow at 18:00. Consequently, the gain in anticipation time here could have been 90 min (issue time of 30 min prior to the event and forecasting time of 60 min for the model output). PS-D did not predict any overflow at all and the POC is not even close to the reference POC, and would not been useful in this case. On the other hand, PS-D and RM-DR both underestimated the rainfall around 08:00 UTC, resulting in a POC of 0, while for the reference run the POC was 0.7. However, this could be adjusted by real-time measurements of the storage.

For the 24-hour event of 24-06-2016 ending at 06:00 UTC, the storage prior to the event was 0. Still the event resulted in a sewage overflow. During the event itself much rain fell (e.g. 18.27 mm in one hour). Consequently, the sewage overflow could not have been prevented, only reduced. One hour prior to the sewage overflow the POC could have been 0.7 instead of 0.35. This would have reduced the overflow with 14.21 m^3 (0.35 (POC) × 4.06 (service area, ha) × 10 = 14.21 m^3) for LBK11. In Fig. 4.15 the event is shown with the

corresponding overflow and POC. An overflow occurred at 20:00, 21:00 and 22:00 UTC. PS-D was not able to predict the first overflow at the right moment. The first predicted overflow by PS-D was at 21:00 for the issue time of 15 min. For earlier issue times, the overflow was predicted later. Consequently, PS-D does not result in a gain in anticipation time. RM-DR was able to predict the overflow at the right moment for issue times of 15 and 30 min. For longer issue times, RM-DR was not able to predict the overflow at the right moment in time. RM-DR results in a gain in anticipation time of 90 min. From the red bar, it can be seen that RM-DR overestimates the precipitation more often, resulting in higher volumes of overflow or wrong warning signals.

Figure 4.10 already indicated that events in summer result in a HR of zero for events of 1 hour, while events in spring result in a much higher HR. For that reason the event of 10-04-2018 ending at 21:50 UTC was also investigated for LBK11 (Fig. 4.16) to see what the gain in anticipation time is when using radar rainfall nowcasting. An overflow occurred at 22:00. Before the overflow there was already storage present in the system. The overflow was simulated by both RM-DR and PS-D for all issue times up to 60 min. The forecasted POC is also in perfect agreement with the reference POC. The gain in anticipation time is at least 120 min. Consequently, the sewage overflow volume could have been reduced with at least 14.21 m^3 (0.35 (POC) × 4.06 (service area, ha) × 10 = 14.21 m^3). Although for this specific event
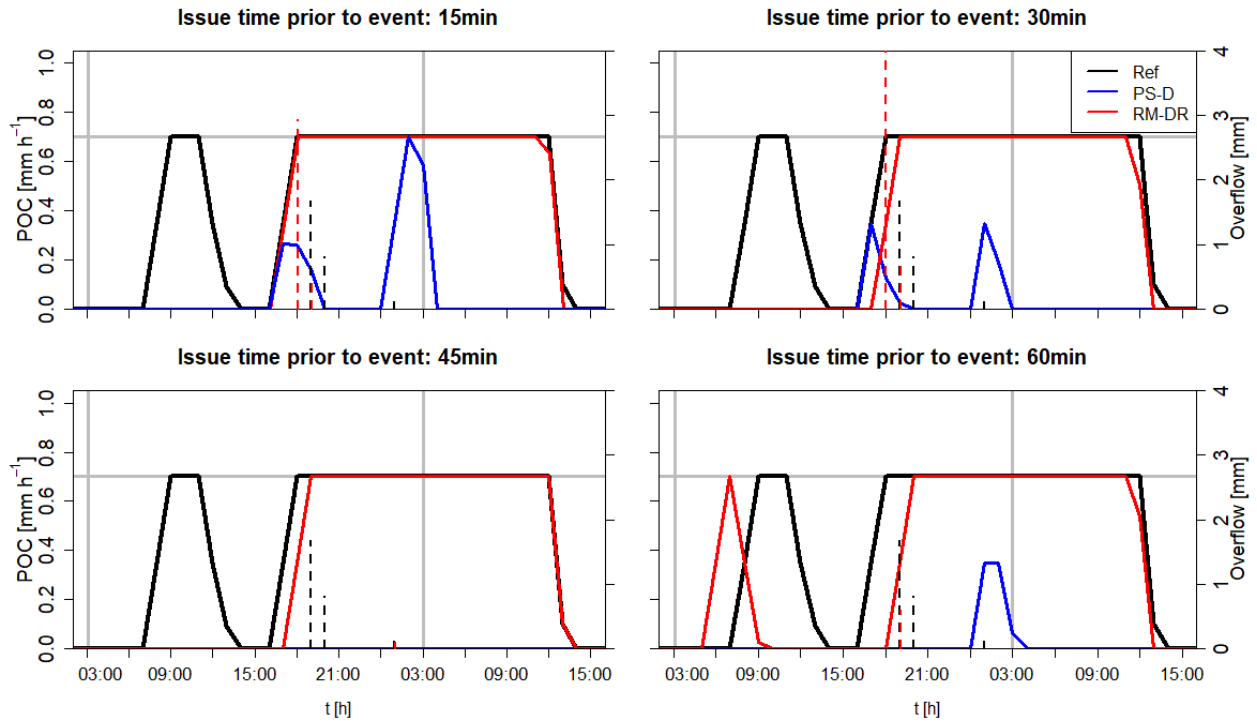
Figure 4.14: The 24-hour event of 02-06-2016 ending at 03:00 UTC, lines showing the POC based on the reference run, RM-DR and PS-D for the location HHH9. New nowcasts are issued for every hour that has passed (thus for 03:00, 04:00, 05:00 and so on, for issue times of 15, 30, 45, 60 min prior to the new hour). The vertical dashed bars show the sewage overflow [mm] that occurred. Grey horizontal line indicates the maximum POC. Grey vertical lines indicate the start and end of the specific event.

the forecasts are in perfect agreement with the reference, there is a substantial amount of variation between the events (e.g. seen in Fig. 4.12).
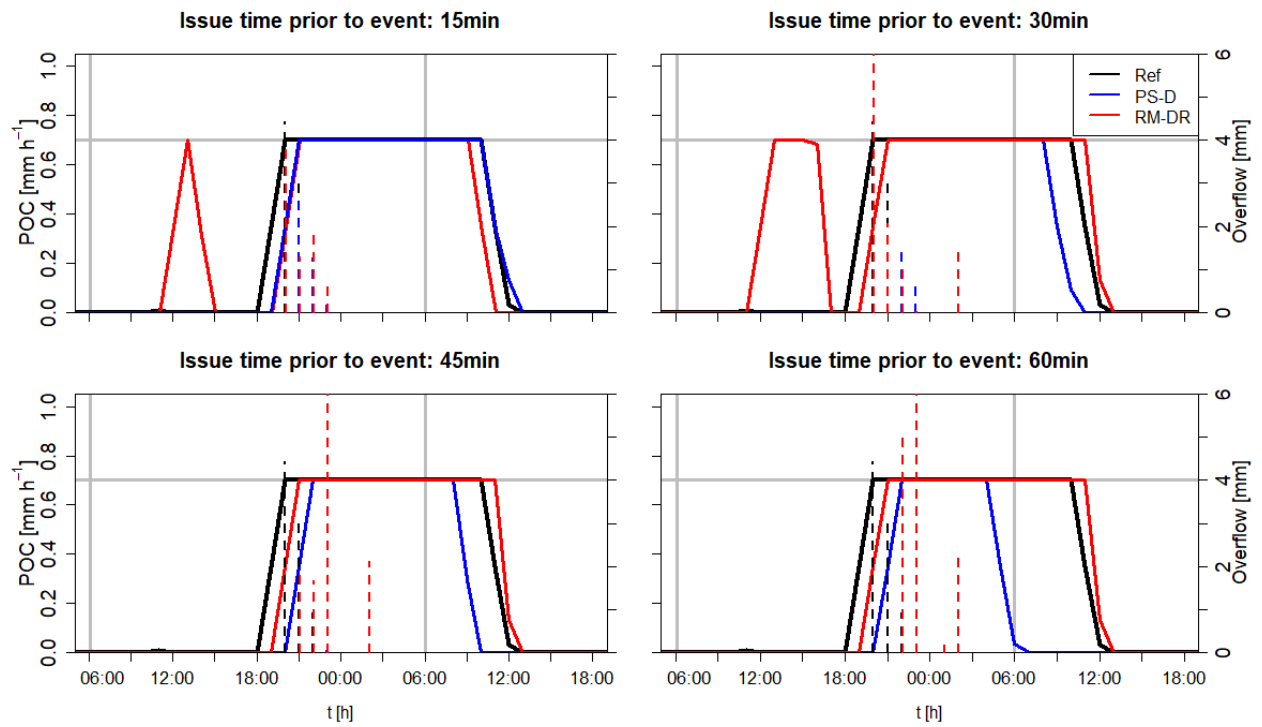
Figure 4.15: The 24-hour event of 24-06-2016 ending at 06:00 UTC, lines showing the POC based on the reference run, RM-DR and PS-D for the location LBK11. New nowcasts are issued for every hour that has passed (thus for 03:00, 04:00, 05:00 and so on, for issue times of 15, 30, 45, 60 min prior to the new hour). The vertical dashed bars show the sewage overflow [mm] that occurred. Grey horizontal line indicates the maximum POC. Grey vertical lines indicate the start and end of the specific event.
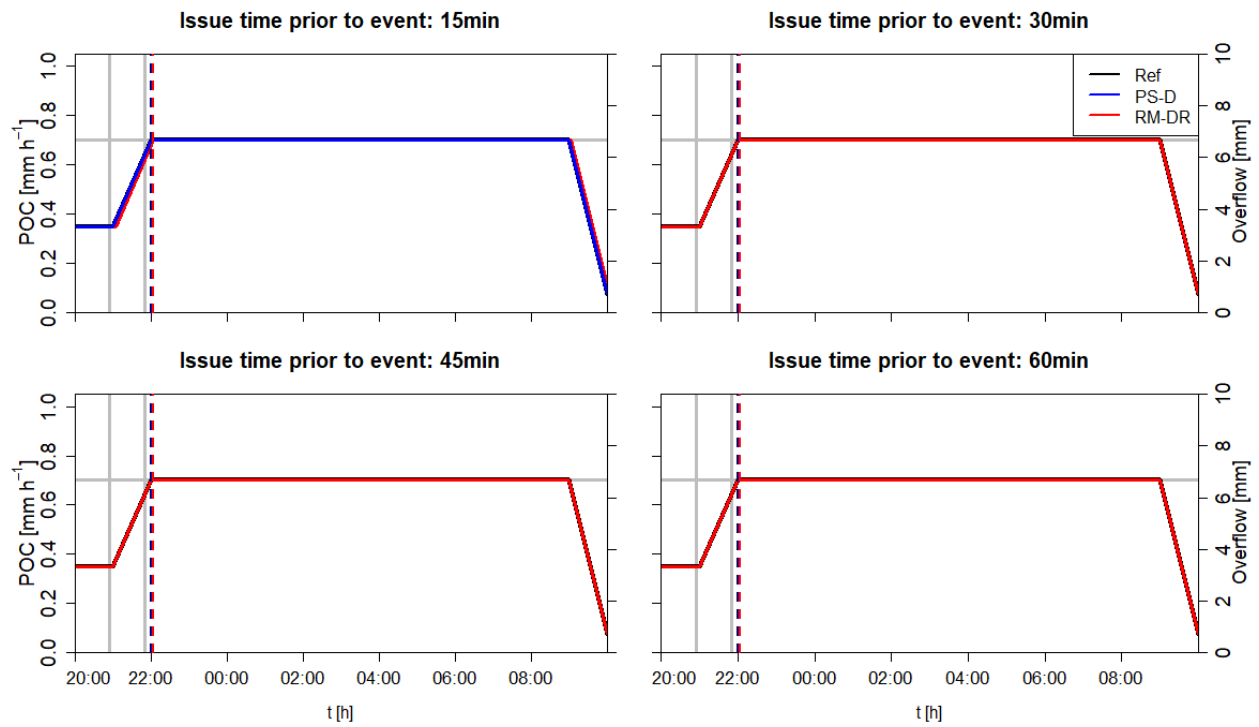
Figure 4.16: The 1-hour event of 10-04-2018 ending at 21:50 UTC, lines showing the POC based on the reference run, RM-DR and PS-D for the location LBK11. Different issue times are plotted (15, 30, 45 and 60 min prior to the start of the event, thus nowcasts are issued at 20:40, 20:25, 20:10 and 19:55 respectively). The vertical dashed bars show the sewage overflow [mm] that occurred. Grey horizontal line indicates the maximum POC. Grey vertical lines indicate the start and end of the specific event.

5 | Discussion

In this chapter, the methods that resulted in uncertainty in the output of the nowcasts and hydrological model are discussed. Next, the most important results are discussed and compared with other research. In addition, the limitations of this research are indicated. Lastly, the performance of the forecasts is compared with the performance of the forecasts as the input for the sewage overflow model.

5.1 Events

In this research, the events were based on an existing nowcast database containing heavy precipitation events. Therefore, this study does not give any insight about wrong warning signals made by the radar nowcasts, which could result in unnecessary discharging of water. However, in the case of sewerage systems, discharging of water is only possible if there is water available in the system. Consequently, wrong warning signals do not always have an impact on the sewerage systems. Only when sewage water is present, wrong warning signals affect the optimization of the water treatment process. Additionally, wrong warning signals can compromise the credibility of the forecasts. It is more likely that RM-DR will cause some wrong warning signals, as results showed that the false alarm rate is higher for RM-DR than PS-D.

Furthermore, the existing nowcast database did not contain enough events that could be used for the comparison with HARMONIE. Only overlapping events with HARMONIE were used for this analysis. The events used in this part were not selected on rainfall characteristics. In addition, the events did not always contain much precipitation. Consequently, no large statistical analysis could be done.

5.2 Radar rainfall nowcasts

5.2.1 Methodology

The radar rainfall nowcasts in this study were compared with gauge-adjusted radar data (reference data), which is a very accurate radar product from the KNMI. However, the radar rainfall nowcasts are made by extrapolating operational available unadjusted radar data. This radar data contains significant biases, with a tendency of underestimation, which influences the use of those nowcast products (Imhoff et al., 2021). To reduce the bias, the used nowcasts in this study were multiplied with a correction factor from Imhoff et al. (2021). Consequently, verifying the radar nowcasts with the reference data will always lead to a bias, as both products use different correction factors, i.e. comparing the initial precipitation conditions (t=0) would already result in a deviation.

5.2.2 Results

The used metrics in this study show different results in skill for the radar nowcasts, indicating the strengths of the nowcasts. Although Pearson's correlation and RMSE indicate that PS-D is more skilful than RM-DR, the bias indicated that RM-DR underestimates less (for lead times up to 30 min). In addition, RM-DR is better in predicting higher intensity rainfalls for lead times up to 30 min (See Fig. 4.5, with the used threshold of 10 mm h^{-1}). PS-D often predicts too much dissipation, losing high-intensity rainfall centres. As a result, RM-DR nowcasts perform better in predicting higher rainfall intensities, and thus a higher HR for larger thresholds. However, RM-DR mislocates the rainfall fields often, especially for longer lead times (See Fig. 4.2a an Appendix Figs. A.3 and A.4 for comparison of the radar images of the reference en nowcasts). The mislocation of rainfall fields is more important for smaller catchments, i.e. in this case urban areas. Mislocation can result in lower Pearson correlation coefficients. Even though PS-D leads to more dissipation, it ensures that it has the same statistical properties as the latest observation.

From Figs. 4.4a and 4.4b it became clear that there is much spatial variation in the correlation of the radar nowcasts. It is expected that spatial variation is caused by the movement of an event and the location of the core of the specific event (Appendix Figs. A.3 and A.4), as intensity dynamics are the most difficult to forecast (Ayzel, 2021). Especially growth and dissipation processes are hard to forecast, as those are not or only partially accounted for.

It should be mentioned that the used Pysteps product in this study (PS-D) is not the most recent one. Already much improvement is done, in which improved products, e.g. individual ensemble members from Pysteps probabilistic (PS-P), result in less dissipation and

expecting to have a better overall performance. The study from Imhoff et al. (2020) already indicated that PS-P has a higher performance than PS-D.

5.2.3 Comparison with other studies

Results showed that the accuracy of the nowcasts decreases with lead time. In addition, results indicated that the nowcasts are less able to predict higher rainfall intensities. This was also found by Liguori et al. (2012) and Codo and Rico-Ramirez (2018). The spatial coverage of the radar can influence the nowcasts in such a manner that with longer lead times the underestimation increases. Golding (1998) indicated already that growth and decay of the precipitation processes are not taken into account, only the initial precipitation state. Consequently, rainfall that was measured just outside of the radar domain will not be accounted for with increasing lead times.

This study shows that PS-D has the highest maximum skilful lead time for 1-hour and 24-hour duration events with lead times of 37 min and 108 min, respectively. These results show similarities with earlier research from Imhoff et al. (2020). This same paper also showed that forecasting winter events result in more skill than summer events, which is in line with the findings of this research. Additionally, these results indicate that stratiform events are better forecasted than convective events, which was also found by Liguori et al. (2012); Foresti et al. (2016).

5.3 HARMONIE

5.3.1 Methodology

The precipitation events used for this part are solely based on the issue times and lead times which were also available for the radar nowcasts. Hence, the selection of events is not based on the most extreme ones. In addition, per lead time only 9 forecasts were compared. Consequently, it turned out that the results from Fig. 4.7 are not significant. In order to draw more statistical meaningful conclusions, a study containing more rainfall events should take place.

5.3.2 Results

HARMONIE data was compared with radar nowcasts made by RM-DR and PS-D. The radar nowcasts have lead times up to 6 hours. To make the most fair
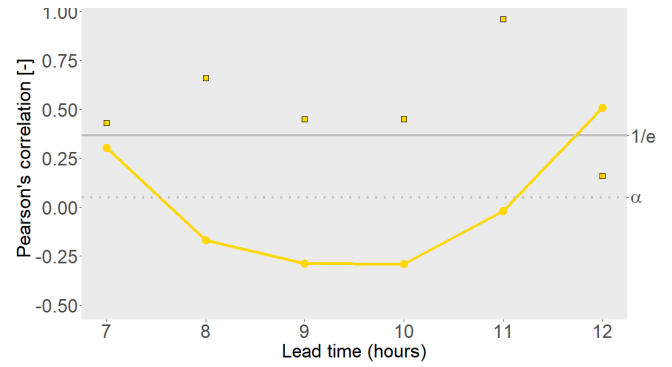

Figure 5.1: Pearson's correlation, indicated by the yellow line, calculated over the catchment averaged rainfall of Helmond per lead time, averaged over 9 events. Pearson's correlation is calculated over HARMONIE. The grey line is the threshold ($1/e$), indicating the minimum correlation for a nowcast to be skilful. The dotted grey line indicates the significance level of 0.05, the squares show the p-value per forecast per lead time.

comparison, the same lead times for HARMONIE were used. However, for operational water management HARMONIE cannot be used for lead times up to 3 hours because it takes some time (around 3 hours) to run the model. Consequently, the user will obtain HARMONIE a few hours after the forecast is issued.

Another large disadvantage of HARMONIE is that it gives hourly volumes of rainfall instead of every 5 min for PS-D and RM-DR. Within those 6 hours, the atmospheric conditions can change significantly, especially during summers, when more convective events take place.

Results indicated that precipitation forecasts made by HARMONIE were not skilful for lead times up to 3 hours. This was also seen in the comparison of the sewerage overflow model output. Pearson's correlation coefficients indicated that for lead times longer than 3 hours HARMONIE became skilful. However, the low HR indicated that HARMONIE is not skilful in predicting the right amount of precipitation, even for low thresholds. Earlier research also stated that NWP models are not useful for short term predictions up to 6 hours (Lin et al., 2005). For that reason the same events are investigated, but with forecasts that are issued 6 hours earlier. Figure 5.1 contains Pearson's correlation coefficient calculated for HARMONIE, with lead times varying between 7 and 12 hours. However, results show that lead times of 7 till 11 hours are not skilled. The results should be questioned as the p-value is higher than the significance level. It is recommended that further research should use more events and also taking into account different

type of events.

Radar rainfall nowcasts are more skilled in predicting the rainfall for lead times up to 2 hours than HARMONIE. Consequently, using nowcasting compared to HARMONIE, results in a gain of 2 hours. Comparison of the output of the sewage overflow model indicated that HARMONIE is not skilled in predicting an overflow for lead times of 1 hour, while PS-D and RM-DR are. This results in a gain of 1 hour compared to the benchmark HARMONIE. However, note that even though PS-D and RM-DR outperform HARMONIE for smaller lead times, both still result in an underestimation or overestimation for PS-D and RM-DR respectively (based on the results in Fig. 4.13).

5.4 Climate change

Earlier research already stated that climate change projections expect more frequent and intense rainfall (Westra et al., 2014). A more recent report about the climate scenarios for the Netherlands, made by the KNMI (KNMI, 2021), confirms these climate projections. The climate scenarios in this report project that the extreme precipitation events in summer will become more severe. Between 1951 and 2020 an increase of around 10% in extreme precipitation events is observed for the inland of the Netherlands. In addition, the intensity per hour is increased by approximately 10-15% over the last 30 years for events with a recurrence time of 10 years.

This research is based on extreme precipitation events between 2008 and 2018. However, with the observed increasing trend in extreme precipitation events, it should be questioned how valid and accurate the results of this research are for future precipitation events. Especially, as the results of this study show that both radar nowcasts are less skilful in predicting higher rainfall intensities (Fig. 4.5).

5.5 Hydrological forecast

5.5.1 Sewage overflow model

The model used in this research has some limitations. It is a simple conceptual model that only uses precipitation as input. The input is solely based on the temporal and spatial resolution of the radar rainfall nowcasts. However, urban areas are characterized by a high spatial variability in land use, which influences the fraction of precipitation that actually reaches the system.

The variability in the degree of impervious areas, slope and soil properties also affects the hydrological response (Singh, 1997). In addition, previous research showed that there is a correlation between the peak flow and the level of urbanization (Isidoro et al., 2012). The same study indicated that wind-driven rain in combination with increased urbanization also affects the hydrological response, such as a higher peak discharge. For those reasons, and also because measurement data were not available, the model output was not compared with measurements, in order to reduce uncertainties caused by the model itself. In this way the focus will only be on the uncertainties of the radar nowcasts compared to the reference. But for future research it is recommended to compare it with overflow measurements, to get a more complete picture of the situation and find out which uncertainties are caused by the model.

A small difference in forecasted and observed precipitation can result in a large difference in sewage overflow due to the used thresholds in the model. One of the thresholds is the amount of precipitation that fell the past two hours, prior to the last hour. When the precipitation is underestimated, this can result, on top of the precipitation difference itself, in a maximum of 0.9 mm difference in storage (See Fig. 2.3 for a small part of the model with the threshold). Another threshold is the storage, which leads to an overflow if the maximum capacity is reached. Consequently, underestimating the precipitation can lead to a substantial difference in the occurrence and volume of an overflow between the locations (see Fig. 4.11).

5.5.2 Results

Results showed that there is much variation between the locations (Fig. 4.10 and Fig. 4.11). It is expected that the sub-areas with larger maximum storage capacities result in lower skill for the nowcast because it is more likely that only the intense rainfall intensities result in an overflow in those areas, and those are the most difficult the forecast.

This study showed that real-time control (RTC), based on forecasted sewage overflows, in order to prevent sewage overflows, has limited use for the selected sewerage areas in this study. However, this is mainly because there can not be much gained in terms of reducing or preventing the amount of sewage overflow. Because either the rainfall depth is higher than the maximum

storage capacity, or the POC is already at its maximum. Due to the used thresholds in the model (see Fig. 2.3), the POC never immediately starts at its maximum (0.7) in the model. Alternatively, the model user can change the threshold of the model in such way that it is not only based on the precipitation of the last 2 hours, prior to the event, but also use the rainfall forecast in the model. If a significant amount of rainfall is forecasted in the next hour, the POC can already start earlier on its maximum. RM-DR would be more skilful for this as RM-DR results in less underestimation than PS-D, especially for events with a duration of 24 hours. However, note that the calculated bias is based on a larger area than the input of the sewer overflow model is.

If RTC is based on the forecasted rainfall that enters the system, thus not only on forecasted sewage overflows, and the user does not mind that all water will be immediately pumped away, PS-D and RM-DR would be skilled for longer issue times than indicated in the results described in section 4.5. The maximum skilful issue time depends on the chosen threshold of the user.

Results indicated that using radar rainfall nowcasting for RTC in order to prevent the overflow in May and June 2016 was not successful. Only for a few events a small amount of sewage overflow (0.35 mm) could have been reduced. Using RM-DR would have resulted in an anticipation time of 90 min. However, RM-DR was not always able to predict the correct amount of precipitation. which resulted in a lower POC than the reference. In this research it was not investigated what the effect is of too low estimations of the POC on the storage. However, real-time measurements in the sewerage system can adjust the POC to the actual POC that is necessary. So it is expected that underestimating the POC will not have a large impact. Real-time control is not yet implemented in Laarbeek and Helmond, however there is a future vision from both municipalities to implement RTC (Arcadis, 2018a,b).

There is also a possibility that hydrological forecasting can be used for other things. Namely, the supply of sewage water varies a lot during the day, while supplying water evenly to the water treatment plants improves the water treatment process. Knowing in advance if a sewage overflow will happen or not, could be beneficial for using the POC optimally and consequently reducing costs. The municipalities of Laarbeek, 's-Hertogenbosch and Heusden already investigated the possibilities in steering the sewage water (https://www.winnovatie.nl/innovatie/optimaliseren-

Table 5.1: Bias of the operational radar product, calculated over the total amount of sewage overflow during the event duration (1 hours or 24 hours), averaged over all events. Bias is shown for all used sub-areas.

Location	Bias	
	1 hours	24 hours
HHC1	0.80	0.59
HHC2	0.58	0.57
HHH8	0.73	0.97
HHH9	0.52	0.76
LLK15	0.60	0.75
LLK16	0.64	0.72
LBK8	0.68	0.84
LBK11	0.69	0.92

aanvoer-afvalwater-naar-de-zuivering-op-droge-dagen). However, Fig. 4.10 already indicated that the HR is very low for most seasons. Consequently, using radar nowcasts will lead to lower POC than necessary, which may result in a larger sewer overflow. The results of this study are only based on heavy precipitation events. Further studies should take all kind of precipitation events into account to investigate what the possibilities is for steering the sewage water.

5.5.3 Operational radar product

In order to distinguish the effect of precipitation underestimation in the radar product and the effect of the nowcasts itself, unadjusted radar data was used as input for the sewage overflow model. However, first, the unadjusted radar product was corrected with the same bias reduction factor (CARROTS from Imhoff et al. (2021)) that was used for the nowcasts. The HR, FA, and bias were calculated by comparing the model output generated by the unadjusted radar data with the reference run.

In Fig. 5.2 the HR and FA are shown per season for the locations LBK11 and LLK15. Results indicate that there is a difference between the seasons, in which events in spring result in the highest HR. For both locations and almost all seasons the HR is below 1, indicating that the underestimation of the radar product is too severe, consequently no overflow is measured. The same figure also indicates that in some cases the bias reduction factor is too high, resulting in a FA. Winter and autumn also
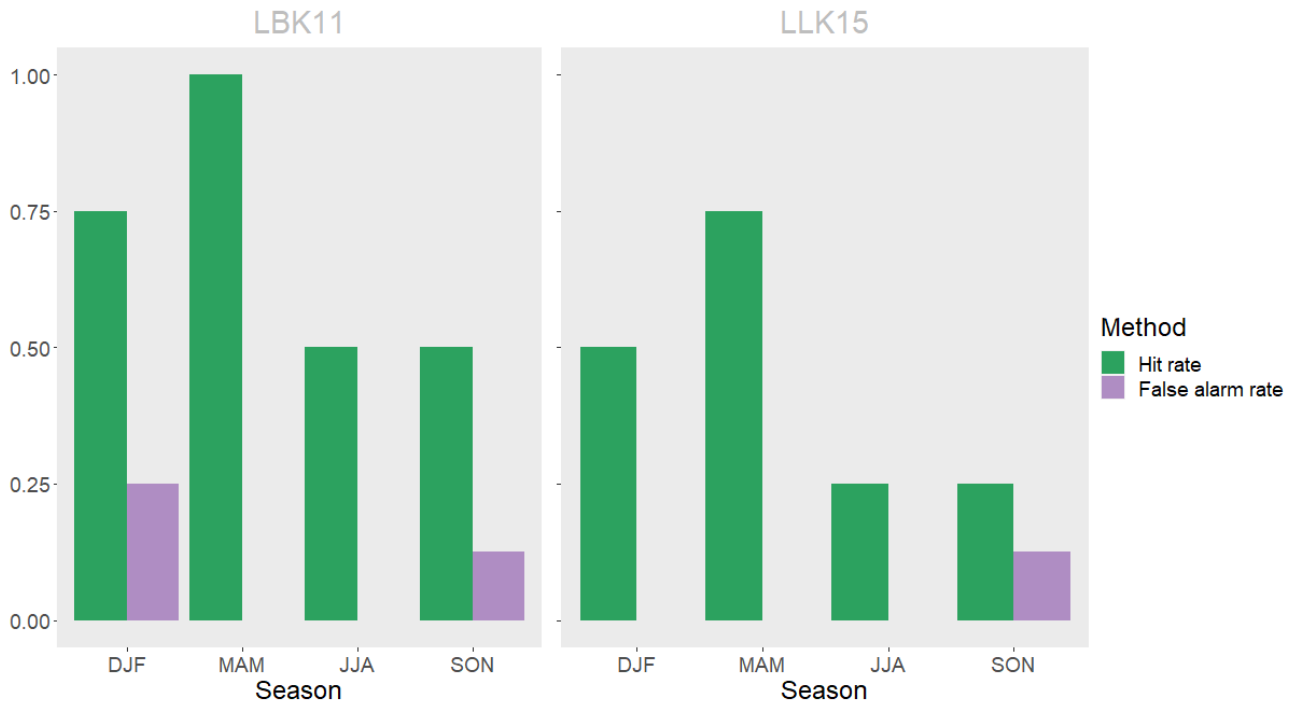
Figure 5.2: Hit rate (HR) and False alarm rate (FA) per season for all events over the sewerage sub-area LBK11 and LLK15, averaged over all events for 1-hour duration event. The shown HR and FA are based on the operational radar product.

resulted in false alarms in Fig. 4.10.

Results indicate that in the unadjusted radar data, after correcting with a bias reduction factor, still results in a bias for all sewerage sub-areas and both event durations (Table 5.1). For most sub-areas the events with a duration of 1 hour have a larger underestimation than the events with a duration of 24 hours. Comparing Tables 5.1 and 4.1 makes it clear that still part of the underestimation results from the nowcasts itself.

5.5.4 Comparison with other studies

Previous research showed that using rainfall information has a positive effect on reducing the combined sewer overflow. However, the effect of RTC is limited by operational constraints such as the treatment capacity of the wastewater treatment plant (Rouault et al., 2008). Another research indicated that forecast horizons exceeding 90 minutes result in uncertainties that become unacceptable with regard to RTC in the sewer system (Achleitner et al., 2009). Or in the case of Codo and Rico-Ramirez (2018), lead times longer than 1 hour do not result in accurate forecasts. These findings seem more promising than the overall results from this research.

However, note that results from other research are

not immediately transferable as different hydrological models are used, with different thresholds, which has an effect on the results.

5.6 Comparison rainfall forecasts and hydrological forecasts

This section compares the results from the performance of the nowcasts and how they perform as an input for a hydrological model.

The bias was calculated for the radar nowcasts (Table 4.1) and for the total amount of sewage overflow that took place during an event (Fig. 4.11). Those results were compared in order to see if there are any relations between them. There is more underestimation in the model output than in the radar nowcasts itself. Table 4.1 and Fig. 4.11 both showed that in general RM-DR scores better (lower underestimation) than PS-D over shorter lead times for 1-hour events. For longer lead times PS-D becomes more skilful. However, there is some variability between the sub-systems if RM-DR or PS-D is more skilful in predicting the right volume (Fig. 4.11). The bias calculated over the total sewage overflow volume varies between the sub-areas. Consequently, no clear relation between the bias from Table 4.1 and Fig. 4.11 can be

seen. Additionally, Table 4.1 shows the bias for different lead times, while the bias in Fig. 4.11 is calculated over different issue times, making it more difficult to see any relation.

Note that the bias for the rainfall forecasts are calculated over larger areas (containing 60 grid cells) than the input of the hydrological forecasts (containing 1 to 3 grid cells). This can influence the bias, as the rainfall forecasts are averaged over a larger area and outliers will have a smaller effect on the results. Therefore, it is recommended for future studies to analyse the precipitation in the sub-areas.

Results showed that there is a clear seasonal dependency on the performance of the radar rainfall nowcasts (Fig. 4.3), in which events during winter resulted in the lowest RMSE. However, the output from the hydrological analysis indicated that spring resulted in the highest HR over all issue times (Figs. 4.10a and 4.10b). However, those are two different metrics. RMSE measures the differences between the predicted values and the observed values, while the HR uses a threshold and is more sensitive to higher rainfall intensities.

According to this research, the maximum skilful lead time for PS-D and RM-DR was respectively 37 and 19 min for events of 1-hour duration, and 108 and 56 min for events of 24-hour duration a (Fig. 4.1). However, the hydrological analysis in this research showed different results. The hydrological analysis indicated that RM-DR is more skilled in predicting the overflow volume than PS-D. Additionally, results from the specific events in May and June 2016 indicated more promising results for RM-DR than for PS-D. Achleitner et al. (2009) also indicated that uncertainties in the forecasts does not immediately lead to the same uncertainties in flow dynamics.

6 | Conclusion

The main goal was to find the added value of radar rainfall nowcasting for real-time control (RTC) in sewerage systems in the urban areas Helmond and Laarbeek. In this study, existent nowcasts made with Pysteps Deterministic (PS-D) or Rainymotion DenseRotation (RM-DR) were used as an input for a simple urban hydrological model. In total 32 extreme precipitation events were analysed to find the performance of the radar nowcasts in: forecasting rainfall and as an input for a simple sewage overflow model (made by the water board Aa en Maas). The nowcasts were also compared to HARMONIE, the operational numerical weather prediction model in the Netherlands.

In general, the longer the rainfall is forecasted prior to the event, the lower the skill becomes, and the higher the underestimation. The maximum skilful lead time for 1-hour duration events is around 19 min or 37 min and for 24-hour durations events it is around 56 min and 108 min for respectively RM-DR and PS-D, while HARMONIE was not skilled for lead times up to 3 hours. HARMONIE results in a lower HR than the radar nowcasts, indicating that HARMONIE results in more underestimation. RM-DR is better in predicting high-intensity rainfall for lead times up to 30 min as PS-D results in more dissipation. At the same time, RM-DR often mislocates the core of the rainfall. A clear difference between more persistent stratiform and convective events was observed, in which stratiform events have a maximum skilful lead time that is almost 3 times higher. Additionally, events in winter have a higher predictability as they resulted in a lower RMSE than events in summer.

Verification of the hydrological model output showed that for longer issue times the forecast skill decreases. PS-D results in more underestimation than RM-DR. Part of the underestimation is already present in the operational radar data. The difference in hit rate (HR) between PS-D and RM-DR is small. Only for the events in winter it is clear that PS-D outperforms RM-DR. Results show that events with a duration of 1 hour result in more underestimation of the total sewer volume than 24-hour events. This is because 1-hour duration events consist often of convective showers having small scale processes, leading to more variation in the forecasts. Events in spring resulted in the highest HR, and summer in the lowest.

The specific events in June and May 2016 indi-
cated that it was possible to reduce the amount of sewer overflow volume for some events. RM-DR was able to forecast the sewer overflow to some extent, resulting in a gain in anticipation time varying between 75-90 min compared to the benchmark HARMONIE or not using a forecast. The underestimation in precipitation by PS-D is often too severe, resulting in no overflow and in a gain in anticipation time of less than 75 min. However, looking at an event in 2018 showed that both PS-D and RM-DR were able to perfectly forecast the overflow and POC, resulting in a gain in anticipation time of at least 120 min. Together with the results from the used metrics and the restrictions of the used sewerage systems, indicate that the performance of RM-DR and PS-D is highly variable in predicting an overflow, and for that reason not yet reliable enough to use for RTC.

Interpretation of the performance of the nowcasts as input for the sewage overflow model highly depends on the goal of the user. If the user does not mind that all water will immediately be pumped away, RTC can be based on forecasted rainfall that enters the system instead of forecasting a sewage overflow. In this case, PS-D and RM-DR would be skilled for longer issue times than mentioned before. However, the performance of the radar nowcasts depends on the chosen threshold (amount of precipitation [mm]) of the user.

The hydrological model output indicated that for lead times up to 3 hours HARMONIE only resulted in misses, while PS-D and RM-DR were able to forecast the sewer overflow for lead times of 1 hour. However, there is still much bias present in the nowcasts. For that reason, it is recommended to further improve the radar rainfall nowcasts to reduce these uncertainties. Further research should also focus on blending radar nowcasts with numerical weather prediction models, in order to improve the forecast skill.

Acknowledgements

First and most of all, I would like to thank my supervisors Claudia Brauer and Ruben Imhoff for the help and meetings during my MSc thesis and the very useful constructive feedback. I especially want to thank Ruben for sending me all the data, helping me out with the re-projection part and always answering immediately. I also want to thank the water board Aa en Maas for sending me the hydrological model and corresponding data that I used for my thesis. Lastly, I want to thank my fellow students who gave useful feedback during the thesis ring.

Bibliography

Achleitner, S., Fach, S., Einfalt, T., Rauch, W., 2009. Nowcasting of rainfall and of combined sewage flow in urban drainage systems. Water Science and Technology 59 (6), 1145–1151.

Arcadis, 2018a. GRP 2019-2023, onweerstaanbaar Helmond.
URL https://www.helmond.nl/bis/2021/Beleid%20en%20onderzoek/Helmond_GRP2019-2023.pdf

Arcadis, 2018b. GRP 2019-2023, onweerstaanbaar Laarbeek.
URL https://www.laarbeek.nl/sites/default/files/GRP-2019-2023.pdf

Ayzel, G., 2021. Advancing radar-based precipitation nowcasting: an open benchmark and the potential of deep learning. Ph.D. thesis, Universität Potsdam.

Ayzel, G., Heistermann, M., Winterrath, T., 2019. Optical flow models as an open benchmark for radar-based precipitation nowcasting (rainymotion v0. 1). Geoscientific Model Development 12 (4), 1387–1402.

Bengtsson, L., Andrae, U., Aspelien, T., Batrak, Y., Calvo, J., de Rooy, W., Gleeson, E., Hansen-Sass, B., Homleid, M., Hortal, M., et al., 2017. The harmonie–arome model configuration in the ALADIN–HIRLAM NWP system. Monthly Weather Review 145 (5), 1919–1935.

Berenguer, M., Corral, C., Sánchez-Diezma, R., Sempere-Torres, D., 2005. Hydrological validation of a radar-based nowcasting technique. Journal of Hydrometeorology 6 (4), 532–549.

Berenguer, M., Sempere-Torres, D., Pegram, G. G., 2011. SBMcast–an ensemble nowcasting technique to assess the uncertainty in rainfall forecasts by lagrangian extrapolation. Journal of Hydrology 404 (3-4), 226–240.

Berne, A., Delrieu, G., Creutin, J.-D., Obled, C., 2004. Temporal and spatial resolution of rainfall measurements required for urban hydrology. Journal of Hydrology 299 (3-4), 166–179.

Brauer, C. C., Overeem, A., Leijnse, H., Uijlenhoet, R., 2016. The effect of differences between rainfall measurement techniques on groundwater and discharge simulations in a lowland catchment. Hydrological Processes 30 (21), 3885–3900.

CBS, 2015. Bodemgebruik; uitgebreide gebruiksvorm, per gemeente.
URL https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70262ned/table?dl=7EDE

Codo, M., Rico-Ramirez, M. A., 2018. Ensemble radar-based rainfall forecasts for urban hydrological applications. Geosciences 8 (8), 297.

de Vos, L., Leijnse, H., Overeem, A., Uijlenhoet, R., 2017. The potential of urban rainfall monitoring with crowdsourced automatic weather stations in Amsterdam. Hydrology and Earth System Sciences 21 (2), 765–777.

Dixon, M., Wiener, G., 1993. Titan: Thunderstorm identification, tracking, analysis, and nowcasting—A radar-based methodology. Journal of atmospheric and oceanic technology 10 (6), 785–797.

Farnebäck, G., 2003. Two-frame motion estimation based on polynomial expansion.

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.

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. Hydrology and Earth System Sciences 20 (1), 505–527.

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 (12), 2859–2873.

Golding, B., 1998. Nimrod: A system for generating automated very short range forecasts. Meteorological Applications 5 (1), 1–16.

Heuvelink, D., Berenguer, M., Brauer, C. C., Uijlenhoet, R., 2020. Hydrological application of radar rainfall nowcasting in the Netherlands. Environment international 136, 105431.

Imhoff, R., Brauer, C., Overeem, A., Weerts, A., Uijlenhoet, R., 2020. Spatial and temporal evaluation of radar rainfall nowcasting techniques on 1,533 events. Water Resources Research 56 (8), e2019WR026723.

Imhoff, R., Brauer, C., van Heeringen, K.-J., Leijnse, H., Overeem, A., Weerts, A., Uijlenhoet, R., 2021. A climatological benchmark for operational radar rainfall bias reduction. Hydrology and Earth System Sciences Discussions, 1–22.

Isidoro, J. M., de Lima, J. L., Leandro, J., 2012. Influence of wind-driven rain on the rainfall-runoff process for urban areas: Scale model of high-rise buildings. Urban Water Journal 9 (3), 199–210.

KNMI, 2021. KNMI klimaatsignaal'21: hoe het klimaat in Nederland snel verandert, KNMI, De Bilt, 72 pp. URL https://cdn.knmi.nl/knmi/asc/klimaatsignaal21/KNMI_Klimaatsignaal21.pdf

Liguori, S., Rico-Ramirez, M., Schellart, A., Saul, A., 2012. Using probabilistic radar rainfall nowcasts and NWP forecasts for flow prediction in urban catchments. Atmospheric Research 103, 80–95.

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 (14).

Overeem, A., Holleman, I., Buishand, A., 2009. Derivation of a 10-year radar-based climatology of rainfall. Journal of Applied Meteorology and Climatology 48 (7), 1448–1463.

Passerat, J., Ouattara, N. K., Mouchel, J.-M., Rocher, V., Servais, P., 2011. Impact of an intense combined sewer overflow event on the microbiological water quality of the Seine River. Water research 45 (2), 893–903.

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). Geoscientific Model Development 12 (10), 4185–4219.

Rouault, P., Schroeder, K., Pawlowsky-Reusing, E., Reimer, E., 2008. Consideration of online rainfall measurement and nowcasting for RTC of the combined sewage system. Water Science and Technology 57 (11), 1799–1804.

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 (7), 1418–1436.

Singh, V., 1997. Effect of spatial and temporal variability in rainfall and watershed characteristics on stream flow hydrograph. Hydrological processes 11 (12), 1649–1669.

Thorndahl, S., Einfalt, T., Willems, P., Nielsen, J. E., Veldhuis, M.-C. t., Arnbjerg-Nielsen, K., Rasmussen, M. R., Molnar, P., 2017. Weather radar rainfall data in urban hydrology. Hydrology and Earth System Sciences 21 (3), 1359–1380.

Thorndahl, S., Poulsen, T. S., Bøvith, T., Borup, M., Ahm, M., Nielsen, J. E., Grum, M., Rasmussen, M. R., Gill, R., Mikkelsen, P. S., 2013. Comparison of short-term rainfall forecasts for model-based flow prediction in urban drainage systems. Water science and technology 68 (2), 472–478.

Thorndahl, S. L., Bøvith, T., Rasmussen, M. R., Gill, R., 2012. On comparing NWP and radar nowcast models for forecasting of urban runoff. In: International Symposium on Weather Radar and Hydrology. IAHS Press, pp. 620–625.

Tilford, K., Fox, N., Collier, C., 2002. Application of weather radar data for urban hydrology. Meteorological Applications 9 (1), 95–104.

UN, 2018. Revision of world urbanization prospects. UN Department of Economic and Social Affairs 16.

Waterschap Aa en Maas, 2016. Feitenrapport wateroverlast mei - juni 2016. URL https://edepot.wur.nl/403128

Westra, S., Fowler, H., Evans, J., Alexander, L., Berg, P., Johnson, F., Kendon, E., Lenderink, G., Roberts, N., 2014. Future changes to the intensity and frequency of short-duration extreme rainfall. Reviews of geophysics 52 (3), 522–555.

Wilks, D. S., 2011. Statistical methods in the atmospheric sciences (3rd ed). Vol. 100. Academic press, Elsevier.

A | Additional figures
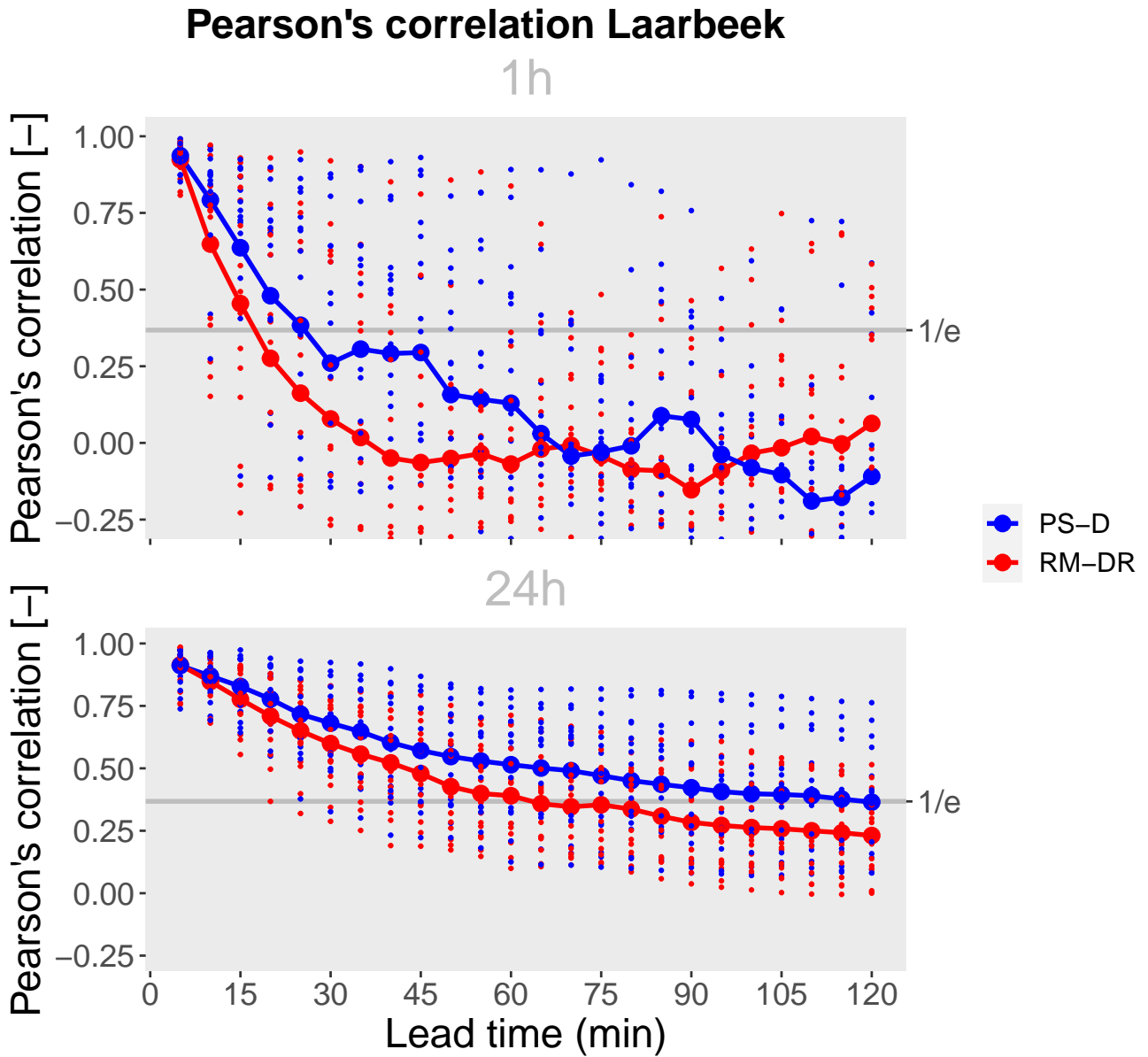
A.1 Pearson's correlation Laarbeek

Figure A.1: Pearson's correlation calculated over the catchment averaged rainfall over Laarbeek per lead time, averaged over all events for 1-hour and 24-hour duration. Pearson's correlation is calculated over both algorithms. The grey line is the threshold $(1/e)$, indicating the minimum correlation for a nowcast to be skilful. The dots indicate the variability per event.
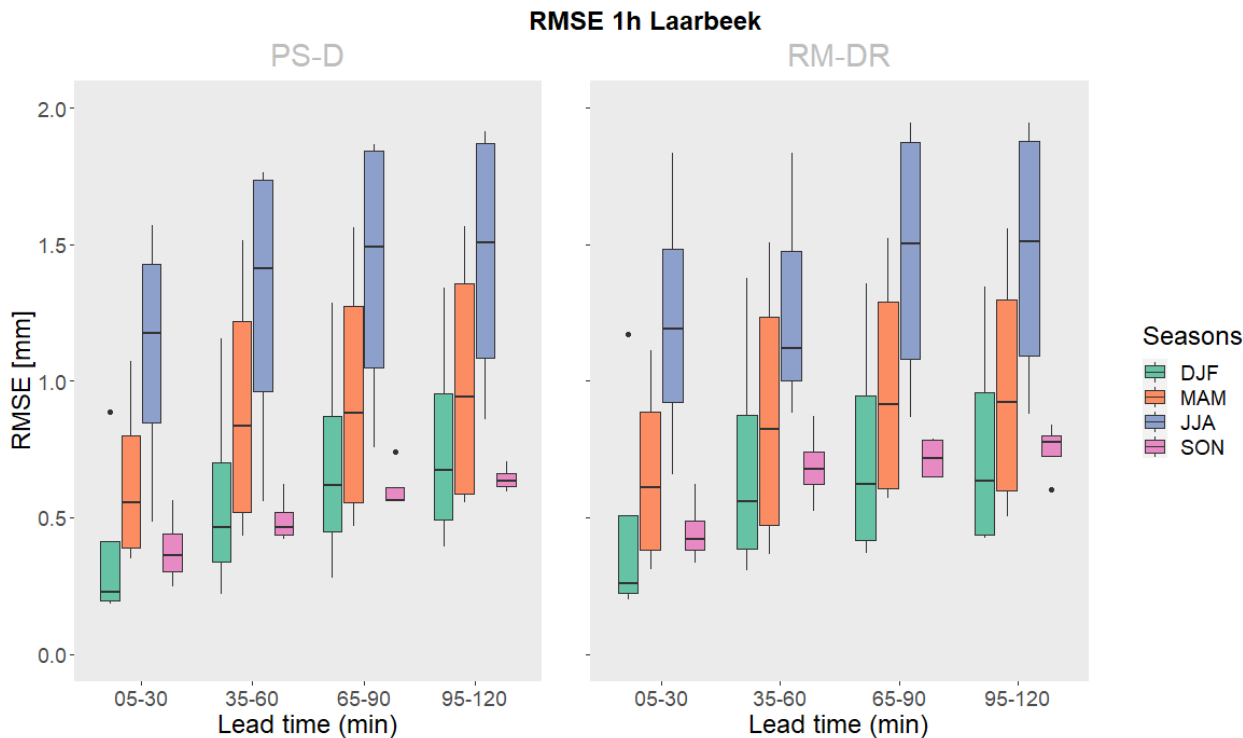
A.2 RMSE 1h Laarbeek

Figure A.2: RMSE per season for all events with a duration of 1 hour over the catchment averaged rainfall of Laarbeek, averaged over lead times of 5-30 min, 35-60 min, 65-90 min and 95-120 min.
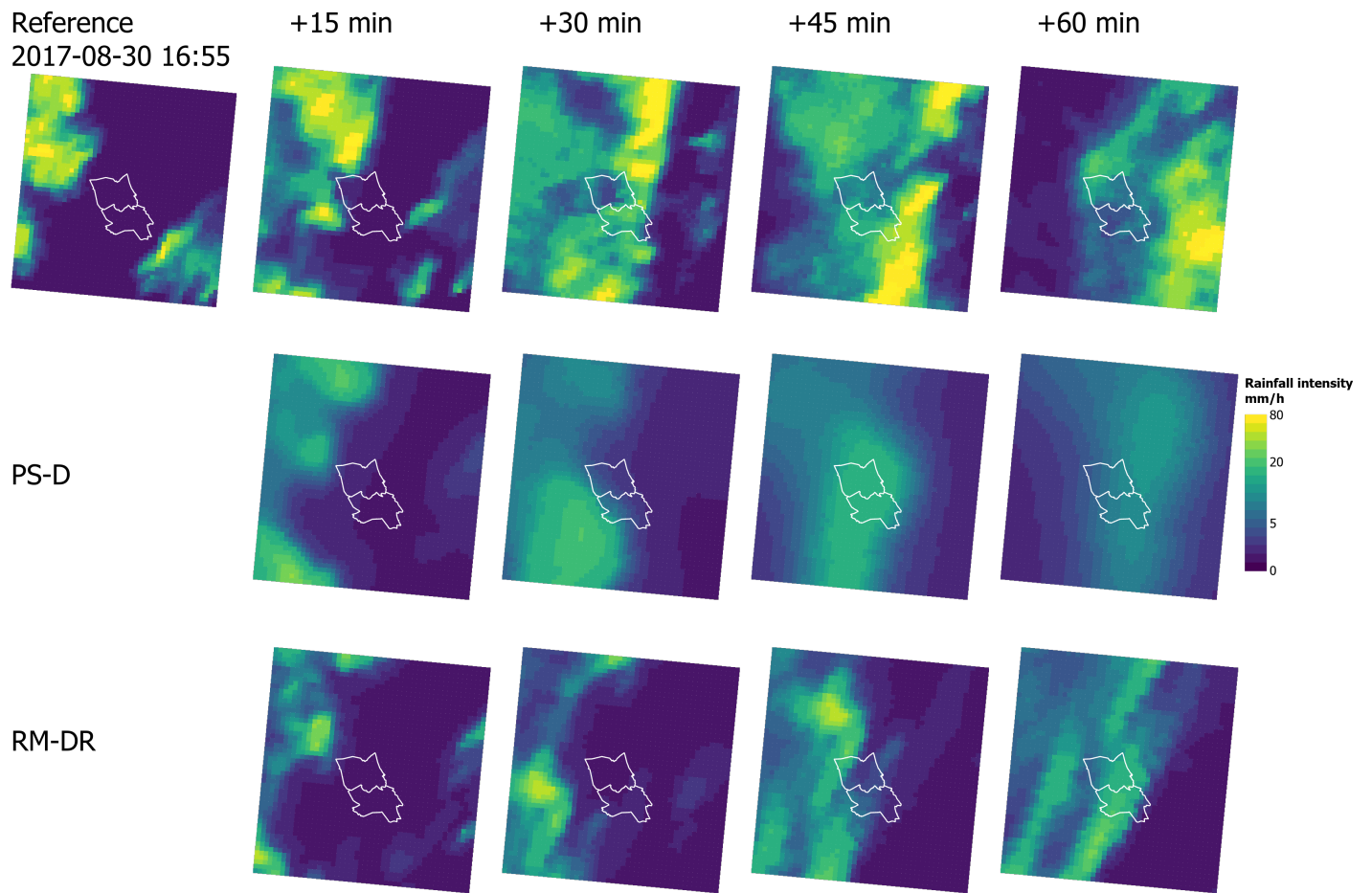
A.3 Radar images

Figure A.3: Radar image of the event of 2017-08-30, ending on 17:55. The reference radar, PS-D and RM-DR nowcasts for the Aa catchment are shown. Rainfall intensity (mm h^{-1}) based on 5-min accumulation is shown here. White contour lines indicate the study area (see also Fig. 2.1)
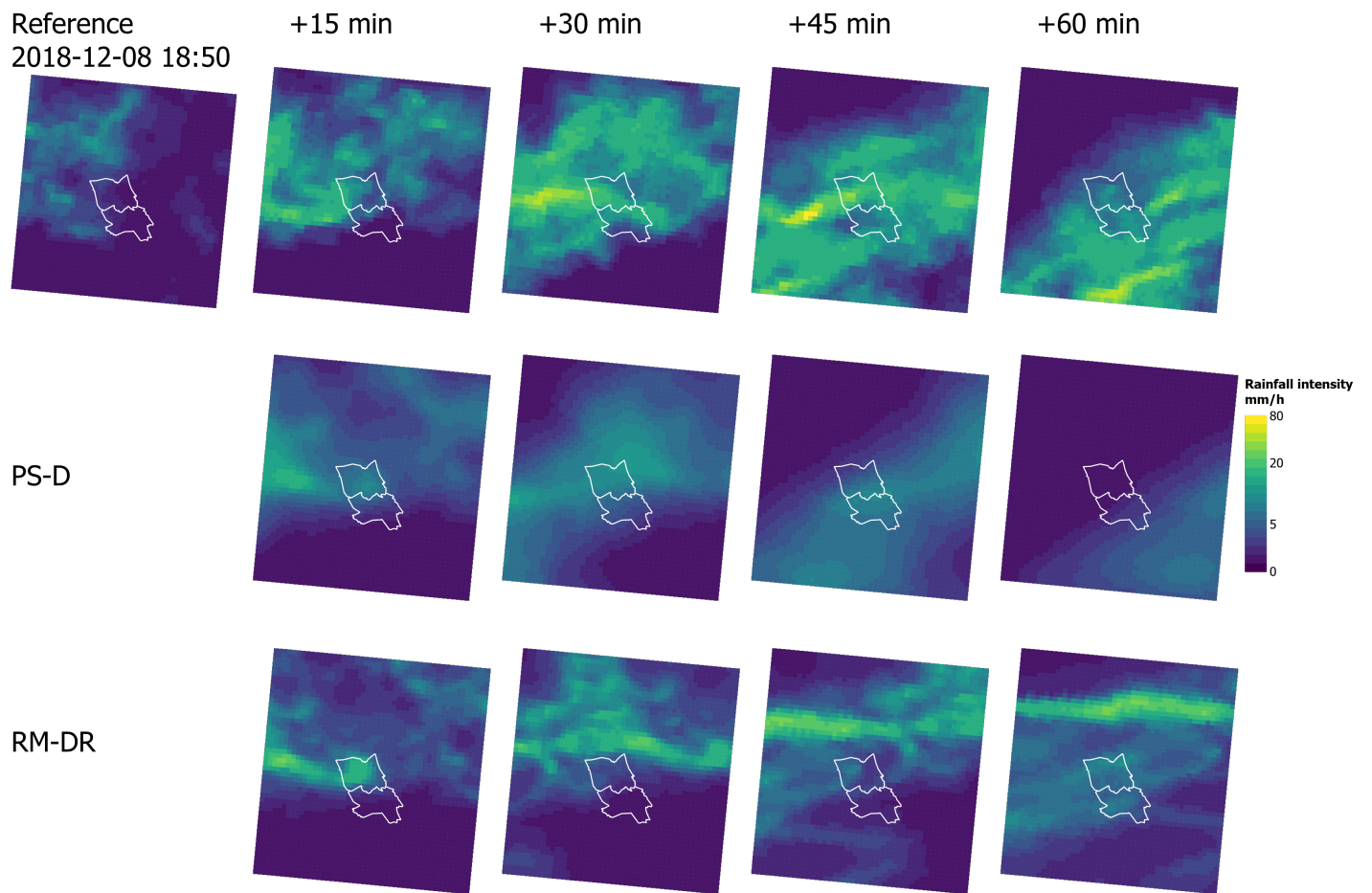.

Figure A.4: Radar image of the event of 2018-12-08 ending on 19:50. The reference radar, PS-D and RM-DR nowcasts for the Aa catchment are shown. Rainfall intensity (mm h^{-1}) based on 5-min accumulation is shown here. White contour lines indicate the study area (see also Fig. 2.1)
.

A.4   Contingency tables

Table A.1: Amount of times that the threshold of 5 and 10 mm h^{-1} was reached by the rain gauge-adjusted radar data, calculated for every grid cell for Helmond and Laarbeek per event. Both areas contain a maximum of 60 grid cells.

|	Helmond		Laarbeek	
End [UTC]	5 mm/h	10 mm/h	5 mm/h	10 mm/h
12-09-2008, 10:40	60	50	60	25
07-10-2009, 18:30	60	3	60	7
22-10-2013, 20:40	60	13	60	9
15-09-2016, 21:25	47	10	31	0
15-05-2008, 19:00	47	39	55	53
30-05-2016, 15:50	56	21	60	60
10-04-2018, 21:50	60	50	60	26
30-04-2018, 00:15	60	17	60	30
12-07-2010, 09:50	60	52	60	55
14-07-2010, 16:50	60	45	60	36
27-07-2013, 09:40	60	0	60	10
30-08-2017, 17:55	60	12	60	30
05-02-2008, 23:45	60	4	60	43
10-12-2009, 03:10	60	0	60	0
03-01-2012, 16:55	60	16	43	0
08-12-2018, 19:50	53	0	60	27
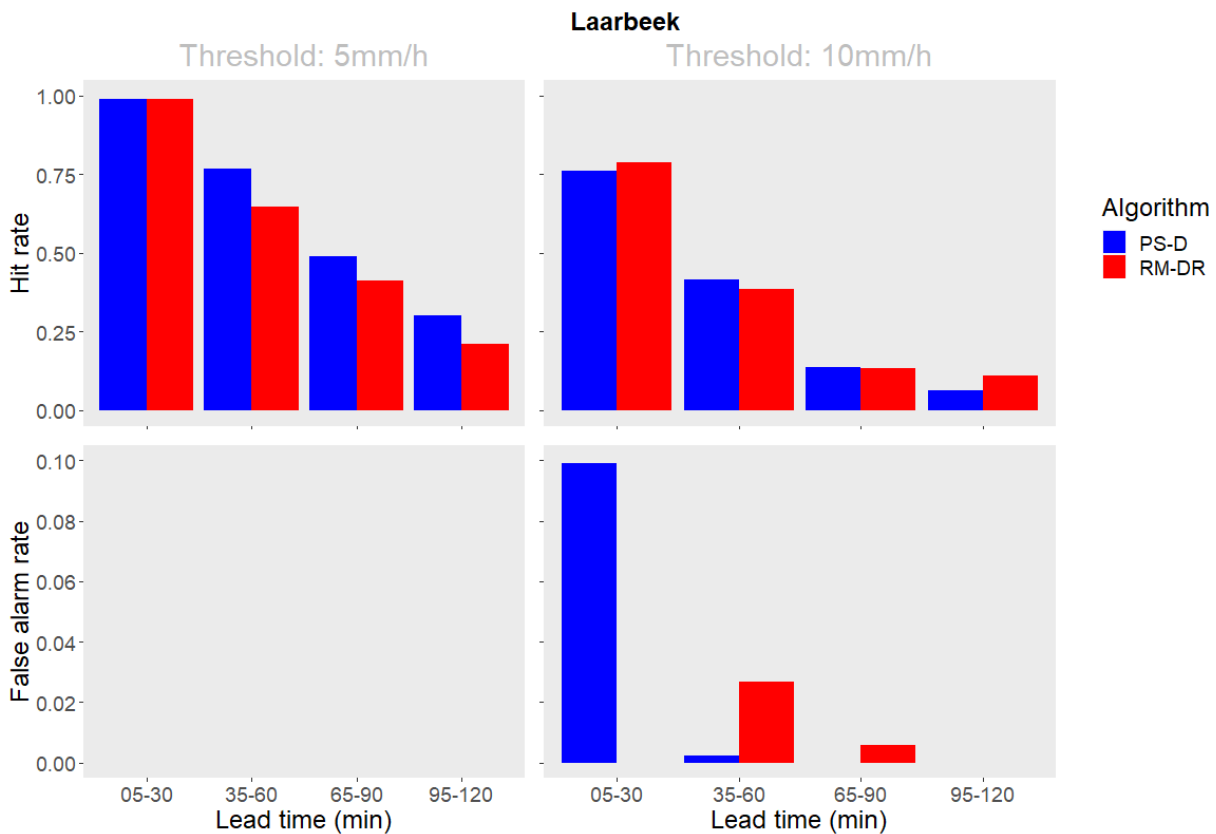
Figure A.5: Hit rate (HR) and false alarm rate (FA) calculated over the urban area of Laarbeek. The HR and FA are calculated over the accumulated rainfall in 1 hour, by using a threshold of 5 mm h^{-1} or 10 mm h^{-1} (See Table A.1 for number of times that these thresholds were reached by the reference rainfall). All events were taken together, and averaged over lead times of 5-30 min, 35-60 min, 65-90 min and 95-120 min. Note the difference in scale on y-axis between the HR and FA.

Table A.2: Amount of times that the threshold of 0.5 and 1 mm h^{-1} was reached by the rain gauge-adjusted radar data, calculated for every grid cell for Helmond per event for the analysis with HARMONIE. Both areas contain a maximum of 60 grid cells.

Issue time [UTC]	1		2		3		4		5		6	
	0.5 mm/h	1 mm/h	0.5 mm/h	1 mm/h	0.5 mm/h	1 mm/h	0.5 mm/h	1 mm/h	0.5 mm/h	1 mm/h	0.5 mm/h	1 mm/h
2018-01-03 06:00	0	0	60	60	60	51	60	60	0	0	0	0
2018-04-11 00:00	0	0	60	60	60	60	60	60	60	60	40	40
2018-04-30 00:00	60	37	60	60	60	60	60	0	60	60	60	60
2018-04-30 06:00	60	60	40	5	60	40	0	0	34	0	53	40
2018-05-22 18:00	0	0	0	0	0	0	0	0	0	0	0	0
2018-05-23 00:00	0	0	60	60	60	60	60	51	18	0	13	4
2018-05-23 06:00	0	0	0	0	0	0	0	0	0	0	1	0
2018-05-23 12:00	7	0	0	0	0	0	0	0	0	0	0	0
2018-12-09 00:00	60	60	60	60	0	0	0	0	55	15	27	6

A.5 Model output

Table A.3: Sewerage sub-areas and the number of overflows that took place per lead time in a specific sub-area. Calculated for lead times up to 6 hours.

Amount of overflows predicted by the reference run						
Location code	1	2	3	4	5	6
HHC1						
HHC2	1		1			1
HHH8	1		1	1		1
HHH9	1		1	1		1
LBK8	1	1	1	1		1
LBK11	1	1	1	1		2
LLK15	1	1	1	1		1
LLK16	1	1	1	1		1