Nowcasting of High-Intensity Rainfall for Urban Applications in the Netherlands

GUO-SHIUAN LIN,[a] RUBEN IMHOFF,[b] MARC SCHLEISS,[c] AND REMKO UIJLENHOET[a]

[a] *Department of Water Management, Delft University of Technology, Delft, Netherlands*
[b] *Department of Operational Water Management and Early Warning, Unit of Inland Water Systems, Deltares, Delft, Netherlands*
[c] *Department of Geoscience and Remote Sensing, Delft University of Technology, Delft, Netherlands*

(Manuscript received 8 November 2023, in final form 6 February 2024, accepted 8 February 2024)

ABSTRACT: Radar rainfall nowcasting has mostly been applied to relatively large (often rural) domains (e.g., river basins), although rainfall nowcasting in small urban areas is expected to be more challenging. Here, we selected 80 events with high rainfall intensities (at least one 1-km^2 grid cell experiences precipitation >15 mm h^{-1} for 1-h events or 30 mm day^{-1} for 24-h events) in five urban areas (Maastricht, Eindhoven, The Hague, Amsterdam, and Groningen) in the Netherlands. We evaluated the performance of 9060 probabilistic nowcasts with 20 ensemble members by applying the short-term ensemble prediction system (STEPS) from Pysteps to every 10-min issue time for the selected events. We found that nowcast errors increased with decreasing (urban) areas especially when below 100 km^2. In addition, at 30-min lead time, the underestimation of nowcasts was 38% larger and the discrimination ability was 11% lower for 1-h events than for 24-h events. A set of gridded correction factors for the Netherlands, CARROTS (Climatology-based Adjustments for Radar Rainfall in an Operational Setting) could adjust the bias in real-time QPE and nowcasts by 70%. Yet, nowcasts were still found to underestimate rainfall more than 50% above 40-min lead time relative to the reference, which indicates that this error originates from the nowcasting model itself. Also, CARROTS did not adjust the rainfall spatial distribution in urban areas much. In summary, radar-based nowcasting for urban areas (between 67 and 213 km^2) in the Netherlands exhibits a short skillful lead time of about 20 min, which can only be used for last-minute warning and preparation.

KEYWORDS: Radars/radar observations; Forecast verification/skill; Nowcasting

1. Introduction

The most recent IPCC report revealed a significant increase of extreme precipitation in many regions in the world since the 1950s (IPCC 2021). High-intensity rainfall is expected to increase flooding in Europe and around the globe (IPCC 2012; Madsen et al. 2014; Ralph et al. 2014; Tabari 2020). Flooding can harm the environment, economic, and human life by destroying ecosystems, inundating infrastructures, and disrupting socioeconomic networks (Merz et al. 2010; Allaire 2018; Kasmalkar et al. 2020). Therefore, accurate weather forecasts of high-intensity rainfall events are important for the environment and human society (Alfieri et al. 2012; Imhoff et al. 2022).

Currently, numerical weather prediction (NWP) models are widely used as operational weather forecasting algorithms around the world. However, because of its coarse spatial resolution and relatively low update frequency, NWP is currently unable to provide accurate and timely (within 6-h lead time) precipitation forecasts with high temporal-spatial resolution for high-intensity rainfall events (Davolio et al. 2015; Silvestro et al. 2016). Thus, its ability to forecast rapid rainfall-induced events like flash floods and pluvial floods is limited. This limitation is even more pronounced in areas with fast hydrological responses such as small urban areas.

Because of their higher fraction of impervious surfaces, urban areas have faster hydrological responses and thus quicker runoff formation than rural areas (Berne et al. 2004; Sharif et al. 2006; Tingsanchali 2012; Cristiano et al. 2017). This leads to shorter anticipation times for flooding events. Besides, urban areas are usually characterized by more complex terrains than rural areas, which causes higher small-scale rainfall variability (Schellart et al. 2014; Cristiano et al. 2017; Maier et al. 2020). Meanwhile, flooding in urban areas often leads to substantial economic loss and safety concerns (The Guardian 2016; CNN 2021; Koks et al. 2022). To enhance precipitation forecast accuracy in urban areas, previous studies found that increasing forecast lead time and spatial and temporal resolution is crucial (Kotroni and Lagouvardos 2004; Rafieeinasab et al. 2015; Ochoa-Rodriguez et al. 2015; Cristiano et al. 2017).

Rainfall nowcasting, the process of (statistically) extrapolating the real-time quantitative precipitation estimates (QPEs) from (generally) weather radar(s), has the potential to provide skillful forecasts up to 3 h in advance (Bowler et al. 2006; Berenguer et al. 2011; Olsson et al. 2014; Jensen et al. 2015; Shehu and Haberlandt 2021), as such a form of "very-short-range" rainfall forecasting (Golding 1998; Sun et al. 2014). Besides, current operational QPE products usually have high temporal (5 min)

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JHM-D-23-0194.s1.

Lin's current affiliation: Ecole polytechnique fédérale de Lausanne, School of Architecture, Civil and Environmental Engineering, Institute of Architecture, Laboratory of Urban and Environmental Systems, Lausanne, Switzerland.

Corresponding author: Remko Uijlenhoet, r.uijlenhoet@tudelft.nl

DOI: 10.1175/JHM-D-23-0194.1

and spatial resolution (1 km) (Zhang et al. 2011; Winterrath et al. 2018; Chang et al. 2021; Overeem et al. 2021). By using the QPE as input, the resulting nowcasts have the same high spatial and temporal resolution. However, QPE products are subject to error and uncertainty (Uijlenhoet and Berne 2008; Krajewski et al. 2010; Hazenberg et al. 2011; van de Beek et al. 2016), especially during high-intensity rainfall (Schleiss et al. 2020) and far away from the weather radars (Imhoff et al. 2021). Thus, verifying QPE products is critical for radar-based nowcasting techniques, especially when volumes matter, such as for hydrological purposes. To our knowledge, no previous research has attempted to distinguish the error in the forecasts caused by the rainfall product employed as a basis for the forecast (as compared with a reference) and those caused by the applied nowcasting method.

Motivated by these considerations, extensive research focusing on nowcasting performance in urban areas has been conducted in recent years. However, most research focused only on a few events (e.g., Achleitner et al. 2009; Berenguer et al. 2011; Foresti et al. 2016; Thorndahl et al. 2016; Heuvelink et al. 2020) or nowcasts for large, mostly rural, areas, notably at the country or continental scale (Berenguer et al. 2011; Kato et al. 2017; Mejsnar et al. 2018; Ravuri et al. 2021; Shehu and Haberlandt 2021). Imhoff et al. (2020a) conducted the first large-sample analysis of over 1500 rainfall events to systematically verify nowcasts for 12 mostly rural catchments (from 6.5 to 957 km^2) in the Netherlands. They showed that nowcasting error is larger for smaller catchments, summer convective rainfall, and catchments located in the upwind direction. However, they focused on the catchment scale, so the results may not be directly applicable to urban areas, especially for the larger catchments considered.

Urban topography, geometry, heat, and aerosols can modify rainfall development and evolution (e.g., Schmid and Niyogi 2017; Liu and Niyogi 2019; Lalonde et al. 2023), which may impact nowcasting performance. In addition, remote sensing of precipitation over urban areas is more challenging because of 1) more obstacles and interference from other reflective sources relative to rural areas, and 2) most of the rain gauges used to bias correct the radar are outside of urban areas. Although radar-based rainfall nowcasts are increasingly being used as inputs for urban flood and sewer system modeling, the quality of the rainfall nowcasts remains an important source of uncertainty (van der Werf et al. 2023) and thus has to be verified by evaluating more events (Liguori et al. 2012). The aim of this research is to complement the analysis in Imhoff et al. (2020a) by focusing on (smaller) urban areas instead of (larger) rural catchments. By focusing on urban areas, we significantly reduce the spatial scale of the area of interest, which should affect the quality of nowcasting for the urban area and highlight focus points for future research. We do so by choosing five major urbanized municipalities in the Netherlands (Amsterdam, the Hague, Groningen, Maastricht, and Eindhoven). We selected 80 events from the municipalities and analyzed 9060 probabilistic nowcasts for the events. These are fewer events than in the study by Imhoff et al. (2020a) because here we focus on high-intensity rainfall events, we choose from fewer event durations, and we do not investigate the impacts of seasons, which were already studied extensively. In this study, we try to answer three research questions:

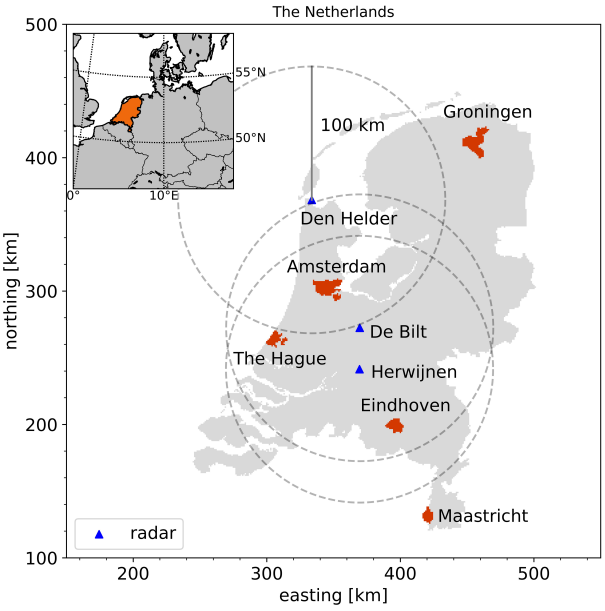

FIG. 1. Locations of the five urban areas (dark red polygons) and the operational C-band radars from KNMI (blue triangles) employed in this study. The radar at De Bilt was replaced by the radar at Herwijnen in 2017. The QPE composites were based on the radar data from Den Helder and either De Bilt (before 2017) or Herwijnen (after 2017). A range of 100 km around each radar station is shown. The x and y values are the coordinates in the KNMI radar projection.

1) What is the performance (in terms of accuracy of forecast rainfall intensity, skillful lead time, and discrimination) of nowcasting during high-intensity rain over urban areas? 2) What are the main factors (e.g., precipitation intensity, event duration, and city size) that influence the nowcasting performance? 3) What is the effect of the quality of the employed QPE product on the nowcasting performance, in terms of forecast rainfall volumes and spatial correspondence?

This paper is structured as follows. The study areas, radar rainfall products, and methods are detailed in section 2. The nowcast results are analyzed in section 3 and discussed in section 4, and our conclusions can be found in section 5.

2. Material and methods

a. Study area

We selected five urban municipalities with different sizes and locations spread over the Netherlands, namely Maastricht, Eindhoven, The Hague, Amsterdam, and Groningen, as shown in Fig. 1. A detailed description of the cities is listed in Table 1. The municipal data and their boundaries were determined by the district and neighborhood map 2020, version 2 (Statistics Netherlands 2020).

b. Radar rainfall products

Three high-resolution radar rainfall datasets were used in this study. These rainfall datasets represent quantitative precipitation estimates (QPEs). The QPEs are obtained from observed radar

TABLE 1. Details of the studied urban areas.

Municipality	Area as the no. of grid cells (1 km^2 each) on the KNMI radar-projection map (Fig. 1)	Population (Statistics Netherlands 2021)	Urban surface ratio (RIONED 2020)
Maastricht	67	120 227	0.70
Eindhoven	96	235 691	0.90
The Hague	96	548 320	0.89
Amsterdam	211	873 338	0.86
Groningen	213	233 273	0.54

reflectivity by the two operational C-band weather radars of the Royal Netherlands Meteorological Institute (KNMI). The reflectivity is converted to rainfall rate using the standard Marshall–Palmer relationship (Marshall et al. 1955):

$$Z = 200R^{1.6}. \quad (1)$$

Here, Z is the radar reflectivity (mm^6 m^{-3}) and R is the rainfall intensity (mm h^{-1}). The obtained rainfall intensities range from 0.1 to 100 mm h^{-1}. The dataset has a temporal resolution of 5 min and a spatial resolution of 1 km. From 2008 to January 2017, the data came from two single-polarized C-band radars in De Bilt and Den Helder. From February 2017 onward, the radar in De Bilt was replaced by the radar at Herwijnen and the radar in Den Helder was replaced by a new radar at the same location. Both new radars are dual-polarized C-band radars. The radars perform three azimuthal scans around a vertical axis to construct pseudo constant-altitude plan position indicators (pseudo CAPPI), which are used for the QPE products. For more information about the radar rainfall products, readers are referred to Beekhuis and Holleman (2008), Overeem et al. (2011), and Beekhuis and Mathijssen (2018).

1) CLIMATOLOGICAL QPE

The first dataset consists of quality-controlled 5-min precipitation accumulations from the climatological gauge-adjusted QPE product (Overeem et al. 2009a, 2011; Royal Netherlands Meteorological Institute 2022a). Doppler and statistical filtering are applied and non-meteorological echoes are removed applying a satellite cloud mask. Radar reflectivities below 7 dBZ are discarded to avoid noise and reflectivity values over 55 dBZ are fixed at 55 dBZ to avoid the influences from hail and strong clutter. An hourly mean-field bias (MFB) adjustment using the ~32 automatic rain gauges from the operational KNMI automatic weather station network and a daily spatial adjustment using the ~320 manual rain gauges are performed. Because of the adjustment using quality-controlled manual rain gauge observations, the dataset is only updated monthly. Although attenuation, bright-band effect, and rainfall advection are not explicitly accounted for in the adjustment, the spatial adjustment corrects for most of the errors (Overeem et al. 2009a). For a more complete description of the rainfall radar products and adjustment methods, readers are referred to Overeem et al. (2009a,b, 2011). The product is considered the best quantitative precipitation estimate product in the Netherlands before January 2023 [a better product is published from February 2023; see Royal Netherlands Meteorological Institute (2023)], so this dataset was used as "true" rainfall and is referred to as climatological QPE here. By using the climatological QPE product as reference, we avoid comparing nowcasts with the (unknown) ground truth, so it also implies that a good verification result in this study does not guarantee that the ground truth is forecast correctly.

2) REAL-TIME QPE

As the climatological QPE is not available in real-time, it cannot be used for nowcasting in practice. Instead, a second QPE product consisting of 5-min nonadjusted real-time precipitation accumulations (Royal Netherlands Meteorological Institute 2022b) was used. The dataset has the same temporal and spatial resolution as the previous dataset, but it is available in real-time. The dataset is referred to as real-time rainfall in this study. Its disadvantage is the errors and uncertainty with respect to the actual rainfall (Uijlenhoet and Berne 2008; Krajewski et al. 2010; Hazenberg et al. 2011; van de Beek et al. 2016; Schleiss et al. 2020; Imhoff et al. 2021). The three main causes of these errors are erroneous reflectivity measurements, wrong conversion of reflectivity to rainfall rate, and sampling errors (Ochoa-Rodriguez et al. 2019).

3) CARROTS QPE

To mitigate errors in real-time radar QPE, a set of correction factors, CARROTS (Climatology-based Adjustments for Radar Rainfall in an Operational Setting), was proposed by Imhoff et al. (2021). CARROTS was calculated based on a 10-yr comparison between the real-time and climatological QPE. It covers the entire Netherlands with correction factors (ranging from 0.7 to 4.6) for each grid cell and each day of the year. For a more complete description of CARROTS, readers are referred to Imhoff et al. (2021). The CARROTS-adjusted real-time QPE is referred to as CARROTS QPE later.

c. *Methods*

1) RAINFALL EVENT SELECTION AND CHARACTERISTICS

To evaluate the nowcasting performance in each urban area, high-intensity events were selected for each municipality using RadarTools (RIONED 2020). RadarTools can list the rainfall events based on the highest climatological rainfall accumulation at any grid cell within each municipality in the Netherlands, from 2008 until present. For the selection criteria, RadarTools sets different thresholds to select rainfall events of different periods. For example, the selection criteria

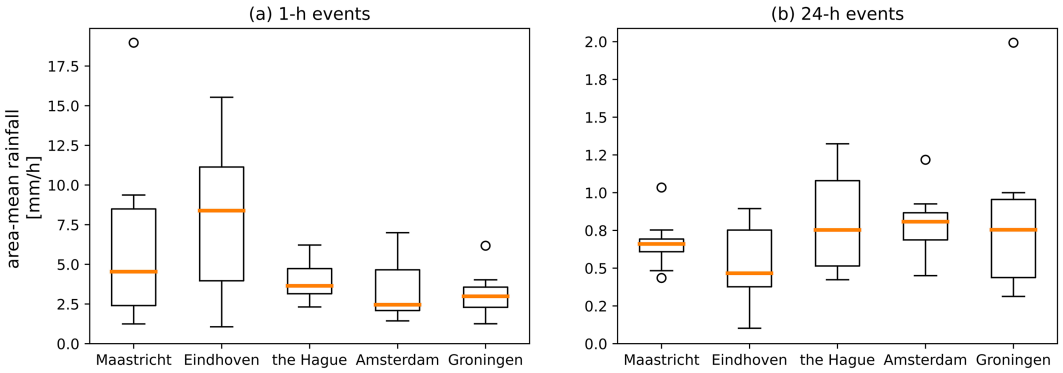

FIG. 2. Rainfall characteristics averaged over the urban areas of all the (a) 1- and (b) 24-h rainfall events in each urban area. The orange lines are the medians. Each of the boxes ranges from the first quartile to the third quartile of the data. The whiskers extend to the data that is within 1.5× interquartile range from the first and third quartiles. The circles are the outliers.

for a 1-h period is that the precipitation sum in any grid cell is above 15 mm. For a 24-h period, the threshold is 30 mm. Rainfall is not necessarily continuous during the period. For a more detailed description of the event selection criteria on RadarTools, see RIONED (2020).

Event durations of 1 and 24 h were selected because nowcasting skills were found to depend on event durations (Turner et al. 2004; Berenguer et al. 2011; Liguori et al. 2012; Olsson et al. 2014; Mejsnar et al. 2018; Imhoff et al. 2020a). Long durations are usually associated with stratiform rainfall (lower intensities and less variability), whereas short durations are typical for convective rainfall (higher intensities and more variability). We selected the eight highest 1-h and eight highest 24-h events in each municipality in between 2008 and 2021. When selecting the 24-h events, the events that covered the period of the previously selected 1-h events were excluded per urban area to ensure the independence of the events. Thus, the events considered per city are independent, although the events in different cities may be caused by the same precipitation system. This selection procedure ultimately led to 5 (cities) × 2 (durations) × 8 (events) = 80 events in this study.

Figure 2 shows the statistics of the selected events. The average rainfall intensity is 4.9 mm h^{-1} for the 1-h events and 0.7 mm h^{-1} (17.3 mm day^{-1}) for the 24-h events, although the highest grid-cell rainfall for each event is over 15 mm h^{-1} (1-h events) or 30 mm day^{-1} (24-h events). To compare with extreme rainfall statistics in the Netherlands, the rainfall intensity for a half-year return period rainfall event is 10 mm h^{-1} for a 1-h duration and 29 mm day^{-1} for a 24-h duration within a radar area of approximately 100 km^2 (Beersma et al. 2019). Note that the return values in the five cities should be adjusted based on urban area sizes (67–213 km^2 in this study) and the chosen event durations (Beersma et al. 2019; Overeem et al. 2010).

For the 1-h events, Fig. 2a shows that Eindhoven and Maastricht experience higher rainfall rates than the other three urban areas, which is partly caused by their smaller domain sizes, which results in less spatially aggregated rainfall fields over the urban areas. For the 24-h events, the area-mean rainfall rates in all urban areas become lower and more similar, as shown in Fig. 2b. For a more complete visualization, rainfall accumulation maps for all events are shown in Fig. S1 in the online supplemental material. Also, supplemental Figs. S1 and S2 show that 1-h rainfall events sometimes just cover small parts of the urban areas. On the other hand, 24-h events cover the entire urban areas, which is typical for stratiform rainfall.

2) NOWCASTING MODEL

Pysteps (Pulkkinen et al. 2019), an open-source Python nowcasting framework, was used in this study. Pysteps includes multiple nowcasting methods. In the study, we used the same setup as in Imhoff et al. (2020a), which consisted of the STEPS (short-term ensemble prediction system) nowcasting method (Seed 2003; Bowler et al. 2006; Seed et al. 2013) using a semi-Lagrangian advection method and the Lucas–Kanade optical flow method (using the QPE from time $t - 3$ to t). An autoregressive model of order 2 was used, and eight cascade levels were set to decompose the rainfall field with decreasing spatial scale. Nowcasts were run for a 4-h lead time. For each time, nowcasts were made with 20 ensemble members to capture the uncertainty related to growth and dissipation of rainfall, as this is not explicitly captured in the STEPS and most other nowcasting methods.

3) VERIFICATION METRICS

Verification was only performed for the nowcasts within the rainfall duration and within the borders of the urban areas. Besides, metrics were only computed for the time steps at which the observed area-averaged rainfall intensities of the input QPE products were higher than 0.1 mm h^{-1}. We employed various metrics because there is no one-size-fit-all metric to quantify nowcast performance, and evaluating nowcast dependency on different factors require diverse metrics.

(i) Continuous ranked probability score

To measure the accuracy of probabilistic nowcasts per grid cell, the continuous ranked probability score (CRPS) (Hersbach 2000) was calculated using

$$\text{CRPS} = \frac{1}{N_f}\sum_{i=1}^{N_f}\int_{-\infty}^{\infty}[P_{F_i}(x) - P_{O_i}(x)]^2\,dx. \quad (2)$$

Here, N_f is the number of forecasts per lead time, P_{F_i} is the nonexceedance probability of nowcast rainfall, and P_{O_i} is the counterpart for the observed rainfall. Although originally designed for probabilistic forecasts, Hersbach (2000) introduced a decomposition of the CRPS for finite ensembles, which is very widely used nowadays, such as by ECMWF (Leutbecher and Haiden 2021) and others (Poletti et al. 2019; Imhoff et al. 2020a; Ravuri et al. 2021; Shehu and Haberlandt 2022), to compare ensemble forecasts with observations. CRPS calculates the area between the cumulative distribution curves of the entire ensemble forecast and the observation. That is, CRPS verify all ensemble members, unlike most of the other metrics outlined below which were only used to verify the mean of the ensembles. Small values of CRPS indicate that the difference between nowcast and observation is small. For Eulerian persistence (and any other deterministic forecast), the CRPS reduces to the mean absolute error (MAE). To compare CRPS across different regions, Reduction CRPS (RCRPS) is used (Trinh et al. 2013). RCRPS is defined as CRPS divided by the standard deviation (σ) of observed rainfall intensity during the event:

$$\text{RCRPS} = \text{CRPS}/\sigma. \quad (3)$$

RCRPS is independent of rainfall magnitude, which facilitates comparison of nowcast accuracy between events and between regions (Trinh et al. 2013; Ye et al. 2014).

(ii) Relative bias

For (hydrological) applications using the forecast rainfall amount, the forecast volume generally matters. The relative bias metric is used to measure the accuracy of nowcast rainfall volume over the whole area, as

$$\text{relative bias} = \left(\frac{F_i}{O_i} - 1\right) \times 100\%. \quad (4)$$

Here, F_i and O_i are forecast and observed 5-min rainfall sums over all areas. The closer the relative bias is to 0, the more accurate is the nowcast.

Likewise, bias also exists between real-time QPE_r and climatological QPE_c as described in section 2b. So, we also define a relative bias in QPE to measure the accuracy of real-time QPE:

$$\text{relative bias}_{\text{QPE}} = \left(\frac{\text{QPE}_r}{\text{QPE}_c} - 1\right) \times 100\%. \quad (5)$$

(iii) Fraction skill score

Besides accuracy, skillful lead times were also estimated using the fraction skill score (FSS) within the urban areas. To compute the FSS, rainfall fields of both nowcast and observation need to be converted to binary fields according to a user-defined threshold. In the study it was set at 1 mm h^{-1} to exclude light rainfall [as the characterization used in van de Beek et al. (2010) and Cristiano et al. (2017)] with the aim to analyze nowcast skills for high-intensity rainfall. Subsequently, a square area surrounding the grid cell is created containing a number of grid cells that is defined by the length scale. Four length scales (corresponding to the areas used for verification) were tested in the study: 1, 5, 10, and 15 km because they represent sizes from an individual grid cell (1 km^2) to the largest urban area in this study (213 km^2). The number of grid cells within the square area that exceed the thresholds are accumulated for the nowcast rainfall and the observed rainfall separately. Finally, the FSS is calculated as the mean squared error (MSE) between the two accumulated numbers. The skillful lead time of the nowcast is then derived as

$$\text{FSS}_{\text{skillful}} \geq 0.5 + (f_0/2). \quad (6)$$

Here, the right-hand side of Eq. (6) is the random forecast skill ($\text{FSS}_{\text{random}}$) and f_0 is the fraction of grid cells in the observation that are higher than the threshold. For a more detailed description of the metric, readers are referred to Roberts and Lean (2008).

(iv) Pearson correlation

Pearson correlation provides another method to measure skillful lead time. It is calculated per lead time:

$$\rho = \frac{1}{N}\sum_{i=1}^{N}\frac{(F_i - \mu_F)(O_i - \mu_o)}{\sigma_F\sigma_o}. \quad (7)$$

Here, F_i and O_i denote the ensemble-mean nowcast and observed rainfall at a given grid cell; μ is the mean rainfall, and σ is the standard deviation at a given time over the area; N is the number of grid cells in the area, so the Pearson correlation derived is the average for the area. The e-folding time, $1/e$ ($\cong 0.37$), was used to determine the decorrelation time between nowcast and observed rainfall. Once the average Pearson's correlation coefficient drops below 0.37, the forecast is not considered skillful anymore (Mejsnar et al. 2018; Imhoff et al. 2020a; Choi and Kim 2022). We used the Pearson correlation coefficient including subtraction of the mean, because Imhoff et al. (2020a) concluded that without subtracting the mean, as was introduced by Germann and Zawadzki (2002), the results bias toward (too) long skillful lead times in the Netherlands (Imhoff et al. 2020a).

(v) Receiver operating characteristic curves

We assessed nowcast discrimination power by using the receiver operating characteristic (ROC). A ROC curve plots the hit rate against the false alarm rate. The hit rate (HR) and false alarm rate (FAR) were calculated for three thresholds (0.1, 1, and 5 mm h^{-1}), as

$$\text{HR} = \frac{\text{TP}}{\text{FN} + \text{TP}}, \quad (8)$$

$$\text{FAR} = \frac{\text{FP}}{\text{FP} + \text{TN}}. \quad (9)$$

Here, FN (false negative), TP (true positive), FP (false positive forecasts), and TN (true negative forecasts) were calculated

per grid cell. The choice of the three thresholds was to assess the nowcasts on their ability to forecast different rainfall intensities. A higher ratio between HR and FAR means higher discrimination power. If the ratio between HR and FAR is below 1, the nowcast is considered to have no skill.

4) EXPERIMENTAL SETUP

To evaluate nowcast quality with lead time, nowcasts were constructed with Pysteps and issued for every 10 min instead of 5 min from 4 h before the start of the event up until 10 min prior to the end of the selected event. In this way, we saved computational time and storage, and our nowcasts have the same frequency of issue time as nowcasts operationally produced by the Dutch water authorities. The nowcasts had a temporal resolution of 5 min, and a forecast horizon of 4 h. The experimental setup for a 1-h event is illustrated in Fig. 3. Only the forecast lead times of the nowcasts that fall within the event duration are verified. Eulerian persistence was used as a baseline to evaluate the additional value of nowcasting in comparison with a "poor man's forecast." Eulerian persistence uses the most recent QPE as forecast for future time steps. For instance, the Eulerian persistence issued at 1350 UTC uses the most recent QPE at 1345 UTC as the rainfall forecast for the next 4 h until 1750 UTC.

The nowcasts were computed on the supercomputer at the Delft University of Technology, Delft Blue [Delft High Performance Computing Centre (DHPC)]. One probabilistic nowcast (20 ensemble members) of 48 timesteps took about 2.5 min to run on one core on an Intel Xeon Gold 6248R processor.

In this study, the real-time QPE was used for nowcast generation and verification in sections 3b and 3c, so the evaluation of nowcasting performance is not affected by QPE errors. Climatological and CARROTS QPEs were employed for selected events and analyzed in sections 3a and 3d to verify the "true" accuracy of nowcasts in an operational setting as compared with the reference (the climatological QPE). Note that only the 1-h events before 2020 (23 events) were run with climatological and CARROTS QPE because the format of the climatological radar products changed at the end of 2019 that made the comparison difficult.

To comprehensively verify nowcast performance in the urban areas, we used all five metrics. Because rainfall intensity and volume are different for events of different durations, we used relative bias to analyze the underestimation of nowcast rainfall volumes and compare the results for 1- and 24-h events. To analyze nowcast accuracy for urban areas of various sizes, RCRPS was used because it is calculated at a higher spatial resolution (namely, for every grid cell) than relative bias (for the entire area considered). FSS and Pearson correlation were used to measure the skillful lead time of nowcasts. Finally, we used ROC to compare the discrimination ability of nowcasts.

Previous studies found that the skillful lead time of nowcasts depends on the size of the region of interest (Pulkkinen et al. 2019; Heuvelink et al. 2020; Imhoff et al. 2020a). To further analyze the effect of city size on nowcasting performance, six stationary and concentric square subareas of 900, 400, 100, 64, 16, and 4 km^2 were defined for all urban areas (see Fig. 4). Nowcasting results averaged over these square areas were compared to determine nowcasting dependence on area size.

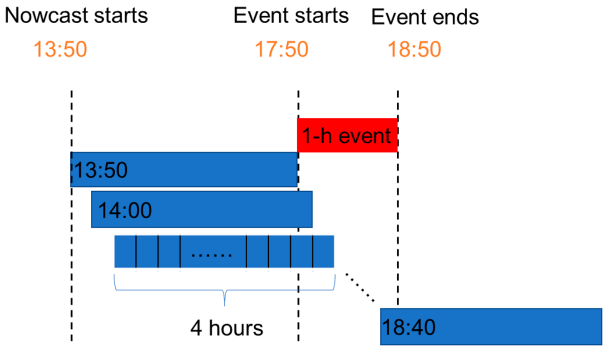

FIG. 3. Illustration of the issue time and forecast horizon for the 1-h event starting at 1750 UTC. The blue bars show a forecast horizon of 4 h with 5-min resolution at different issue times. The red bar shows the event duration. Only the nowcasts that fall within the event duration are verified with observations.

3. Results

We start by analyzing one rainfall event as an example to demonstrate the performance and challenges of nowcasting in section 3a. Following that, all nowcasts are analyzed concerning three main aspects. First, the nowcast model performance for the five urban areas is compared in section 3b. Second, the dependency of nowcast model performance on rainfall duration and area sizes are assessed in section 3c. Third, different approaches to radar QPE and their corresponding nowcasts are compared in section 3d.

a. Case study

In Fig. 5, we illustrate a 1-h event that started in the Hague at 0155 UTC 23 June 2016 to demonstrate the difference in the QPE products and challenges of nowcasting. The climatological, CARROTS, and real-time QPE and nowcasts (using the real-time QPE as inputs) at 30-, 60-, and 90-min lead times are shown in the upper 2 rows in Fig. 5. First, we compare the three radar rainfall products. In the Hague, the rainfall accumulation using real-time QPE is only 31% of the corresponding value from the climatological QPE. The CARROTS adjustments improve the ratio to 55%. Although underestimation still exists, CARROTS QPE shows better agreement with the climatological QPE.

Second, we compare nowcasts at the three lead times. Both the first ensemble member and ensemble mean correctly identify the areas of high rainfall intensity (the center-left area). However, underestimation increases as the lead time increases. Moreover, nowcasting performance strongly depends on the location and size of the target area; for example, rainfall is largely missed when zooming in to the urban area of the Hague.

The probability of exceedance is plotted by using the 20 ensemble members of the nowcasts. The bottom 2 rows in Fig. 5 show the rainfall exceedance probability maps according to

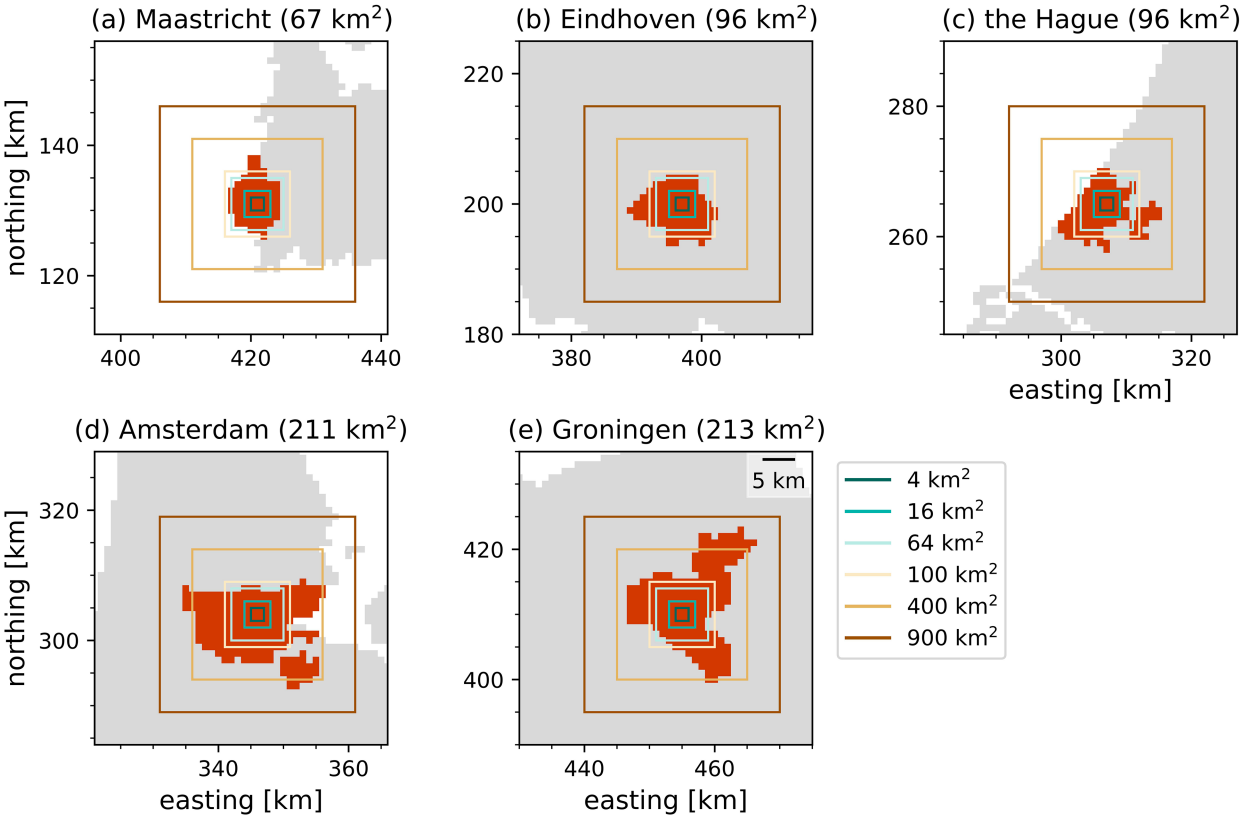

FIG. 4. Map of the six concentric square subareas from 900 to 4 km^2 for each urban area. The gray areas are radar projections of the five municipal areas from Statistics Netherlands (2020). The x and y values are the coordinates in the KNMI radar projection (see Fig. 1).

thresholds of 1 and 5 mm h^{-1} respectively. In the observed QPE maps, areas where rainfall is higher than the threshold are given a probability of 1 while the other regions are given 0. Similar to previous results, exceedance probability is underestimated as lead time increases for the nowcasts. In addition, the probability of rainfall over 5 mm h^{-1} is underestimated more than the probability of lower-intensity rainfall. This suggests that it is more difficult to nowcast high rainfall intensities correctly, as also noted by, for example, Liguori et al. (2012), Foresti et al. (2016), and Imhoff et al. (2020a).

b. *Investigating nowcast model performance in the urban areas*

We proceed to the analysis of all 80 events. In this section, nowcast accuracy and skill for different urban areas are compared and related to the area sizes and rainfall characteristics. We use real-time QPE as the input and reference for assessing strictly the performance of the nowcast model itself, regardless of the accuracy of the QPE products, which will be investigated in section 3d.

1) NOWCAST MODEL ACCURACY

We start with comparing the nowcast accuracy for the urban areas. The average RCRPS of the eight 1-h events in each urban area is shown in Fig. 6a. Lower forecast accuracy for both 1- and 24-h events is found in Maastricht and Eindhoven, probably due to their smaller sizes. This dependence is later supported by the strong correlation between RCRPS and area sizes, as shown in Fig. 12. Although rainfall intensity is higher in Maastricht and Eindhoven, which increases their CRPS (Fig. 7a), a comparison of nowcast performance across regions was made possible by employing RCRPS, which does not correlate with rainfall magnitude (Fig. 7b).

For 1-h events, the average RCRPS increases strongly between the lead times from 0 to 30 min ahead. The RCRPS levels off at around 30-min lead time, and the nowcasts show similar skill as Eulerian persistence when the lead time is beyond 60 min. Yet, nowcasts for the 24-h events almost always show better skill than Eulerian persistence throughout all lead times from 0 to 4 h. This indicates that nowcasting is more accurate and skillful for a longer time for more persistent events [similar to the findings in, e.g., Imhoff et al. (2020a)].

2) THE SKILLFUL LEAD TIME OF THE NOWCASTS

Figure 6 shows that nowcast accuracy deteriorates quickly with increasing lead time, especially for 1-h events. Following this finding, we further used FSS to compare skillful lead times in the urban areas. The results are shown in Fig. 8.

Figure 8 shows that the FSS is larger when the length scale is longer, as expected because rainfall statistics over larger

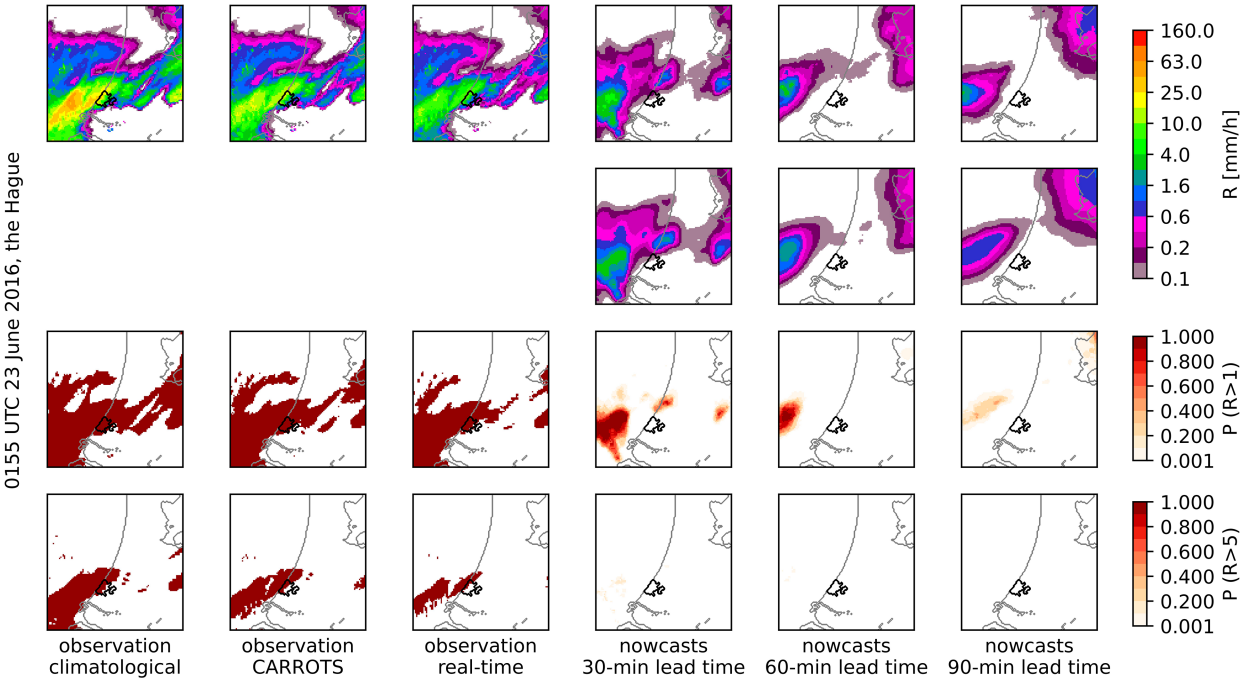

FIG. 5. Climatological, CARROTS-adjusted, and real-time radar QPE and nowcasts with 30-, 60-, and 90-min lead time for the rainfall event in the Hague from 0155 to 0255 UTC 23 Jun 2016. (top),(top middle) Rainfall intensity, showing the average intensity during the 1-h event. Nowcast rainfall intensity is averaged from the nowcasts with the indicated lead times prior to the observation. Nowcasts in the top row use individual ensemble member number 1. The second row shows the nowcasts as ensemble mean. (bottom middle),(bottom) Rainfall exceedance probability for thresholds of 1.0 and 5.0 mm day^{-1}, respectively, using the full ensemble. The black borders indicate the Hague, and the gray borders show the Dutch coastline.

areas are less sensitive to displacement errors in the forecast. Notably, the increase in FSS is higher from 1 to 5 km than from 5 to 10 km. A length scale of 10 km already leads to a spatial scale of 100 km^2, which is larger than Maastricht, the Hague, and Eindhoven. Further upscaling to 15 km does not enhance FSS much because a spatial resolution of 225 km^2 is larger than all the urban areas selected (Table 1).

The skillful lead times in the urban areas at all length scales vary between 12 (1-km length scale in Amsterdam) and 26 min (15-km length scale in Eindhoven). However, the difference in skillful lead times between the urban areas is marginal when using the same length scale. Thus, no urban area shows particularly longer skillful lead times. The FSS at a length scale of 5 km is averaged for each urban area and plotted in the bottom-right figure. It shows that the average skillful lead time from FSS is just around 18 min, although nowcasts can show better skills than Eulerian persistence up to a lead time of more than one hour (Fig. 6a). Changing the threshold of FSS

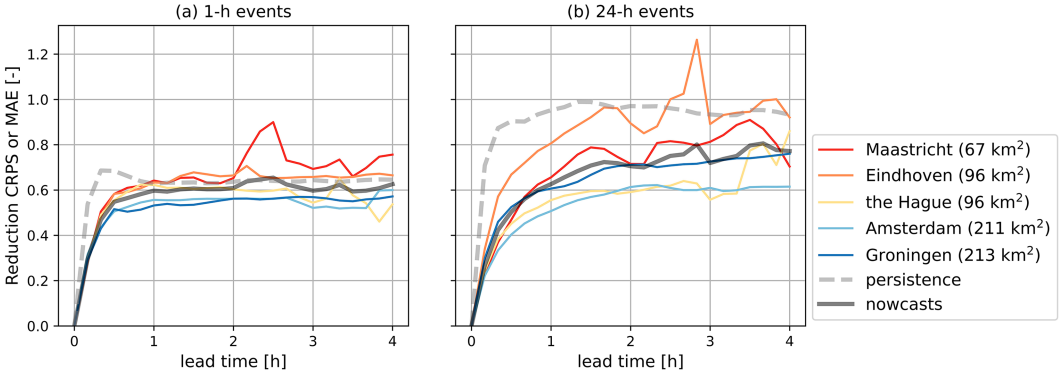

FIG. 6. Average RCRPS for each urban area for the (a) 1-h events and (b) 24-h events. The RCRPS is resampled to 10 min. The black lines show the average RCRPS for all nowcasts in all urban areas. The dashed lines show the average MAE of Eulerian persistence in all urban areas.

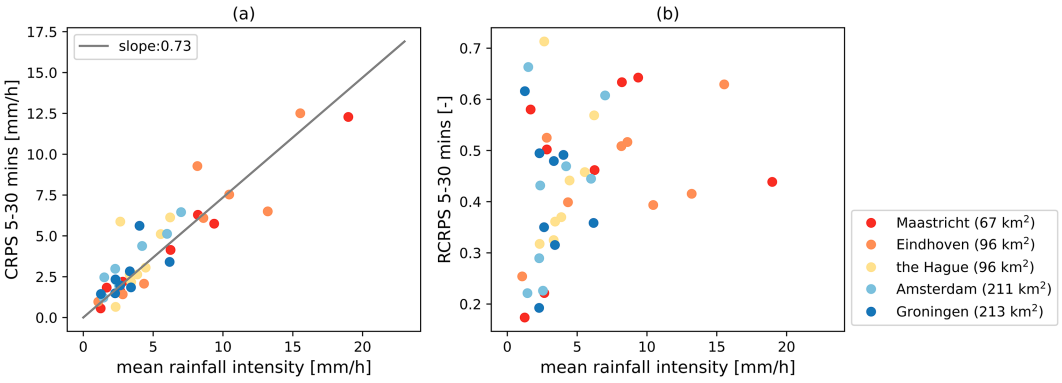

FIG. 7. ECRPS and RCRPS (average between 5- and 30-min lead time) against mean rainfall intensity and rainfall spatial standard deviation in the five urban areas during the 1-h events. Each dot is the result from one rainfall event.

calculation to 0.1 mm h^{-1} does not extend the skillful lead time much (Fig. S3 in the online supplemental material), probably because FSS_{random} increases when a lower threshold is used.

3) NOWCAST DISCRIMINATION ABILITY

After measuring the accuracy and skillful lead time of nowcasts, we continue to quantify the discrimination ability. The ROC curve of each urban area for the 1- and 24-h events is shown in Fig. 9. We chose to analyze the ROC curve at 30-min lead time because this lead time is a challenge for nowcasting high-intensity rainfall and/or for nowcasting in urban areas (Olsson et al. 2014; Foresti et al. 2016; Mejsnar et al. 2018; Imhoff et al. 2020a; Shehu and Haberlandt 2021). We noticed that AUC sometimes tends to be higher when the observed rainfall (and nowcast) covers larger areas (relative

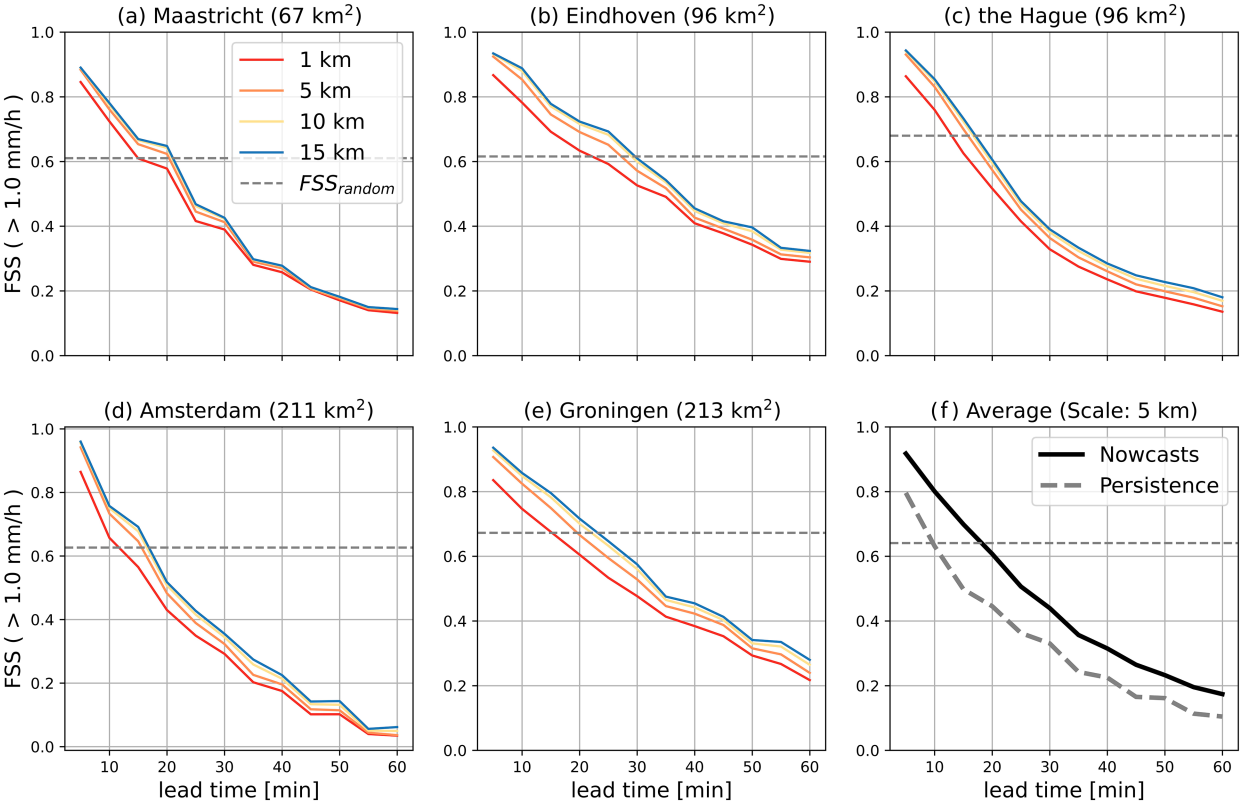

FIG. 8. (a)–(e) Fraction skill score for the 1-h events in each urban area. (f) The average FSS of all five urban areas at a length scale of 5 km (black line). The horizontal gray dashed line shows the random fraction skill score in each urban area [Eq. (6)]. Values of FSS lower than the gray horizontal lines imply that nowcasts are not skillful anymore.

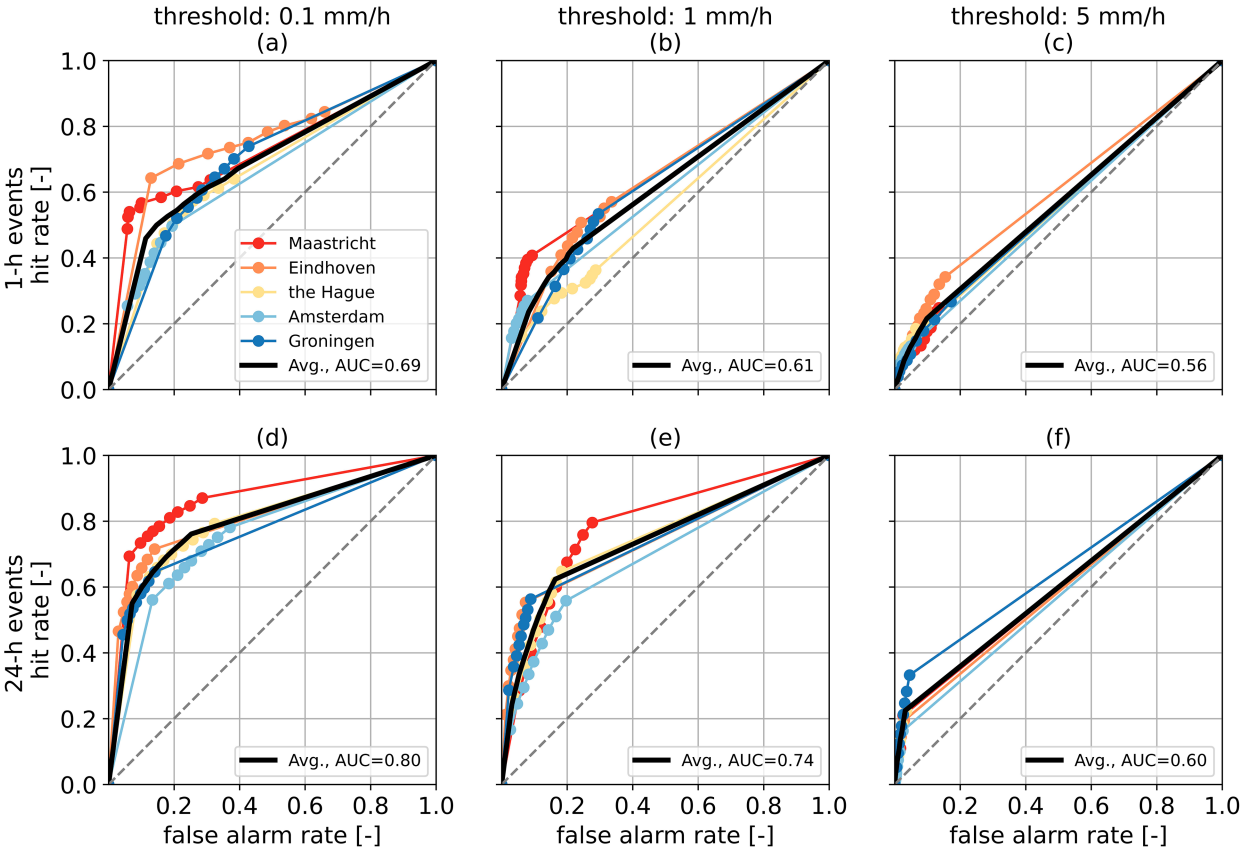

FIG. 9. ROC curve of the (a)–(c) 1-h and (d)–(f) 24-h events for each urban area at three thresholds: (left) 0.1, (center) 1, and (right) 5 mm h^{-1}. Hit rate and false alarm rates are averaged from nowcasts at 30-min lead time. The black lines are the averages of the five urban areas. AUC is the area under the black curve. The gray line means hit rate is equal to false alarm rate, below which nowcasts are not skillful. Higher above the line means better discrimination skill. There are nine dots on each line indicating the forecast probability of exceeding the threshold from 0.1 (rightmost) to 0.9 (leftmost).

to the size of the urban areas), probably by reducing the false positive rate [Eq. (9)] to very close to zero. Thus, smaller cities like Eindhoven and Maastricht have higher AUC for some rainfall events (Fig. S4 in the online supplemental material) and on average higher AUC than the other cities (Fig. 9). For Amsterdam and Groningen, parts of the urban areas are not covered by some of the rainfall events (as shown in Fig. S1 in the online supplemental material), so wrong forecast rainfall in those areas reduces AUC.

Besides, the AUC is smaller for shorter events or when the threshold is higher, indicating nowcasts have poorer discrimination under such conditions. The is discussed in more detail in section 3c(1).

To summarize nowcast performance in the five urban areas, first, urban areas with higher rainfall intensity and spatial variation have lower nowcast accuracy. Second, discrimination for forecasting low rainfall (0.1–1 mm h^{-1}) is generally higher in Eindhoven and Maastricht probably because their higher rainfall spatial coverage causes higher hit rates and reduces false alarm rates. Finally, although differences are discovered in nowcast accuracy and discrimination ability, FSS-based skillful lead times are similar in the urban areas (about 20 min).

c. *Factors that affect the nowcast model accuracy*

Section 3b shows that nowcast model accuracy and skills vary between the urban areas. Here, we continue with the analysis to test some rainfall and environmental characteristics that could explain the differences between the urban areas.

1) RAINFALL DURATION

The average relative biases between nowcast rainfall and real-time QPE are shown in Fig. 10a. For the 1-h events, relative biases of area-mean rainfall drop below −50% within 30-min lead time. Nowcasts reduce the overestimation of Eulerian persistence within 30-min lead time. Eulerian persistence often leads to overestimating rainfall in some grid cells, which can increase the relative bias significantly if the grid cells are within the verification area (Fig. S3 in the online supplemental material). Despite this improvement at short lead times, nowcasts show similar relative bias as Eulerian persistence at lead times beyond 50 min.

Unlike the large underestimation of the 1-h events, nowcasts show a better estimation of rainfall volume for the 24-h events. Figure 10b shows that nowcasts can estimate the total rainfall volume accurately (relative bias between 20%) with a

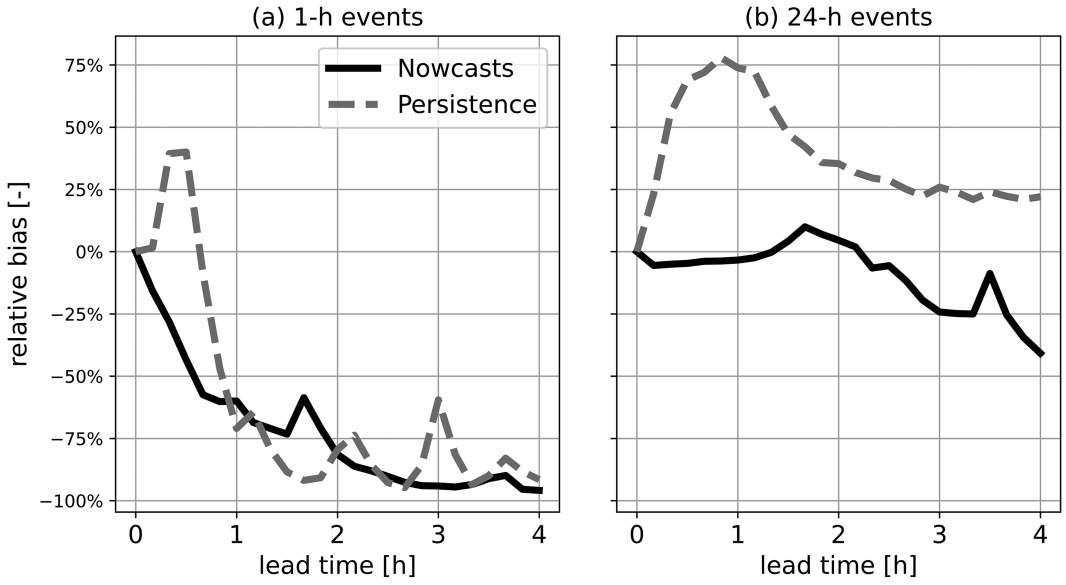

FIG. 10. Average relative bias of nowcasts for (a) all 40 1-h events and (b) all 40 24-h events. The relative bias is resampled to 10 min. The solid line and the dashed line are the average relative bias of nowcasts and Eulerian persistence, respectively.

lead time up to 2.5 h for the 24-h events. Such performance also surpasses Eulerian persistence, which constantly overestimates rainfall intensity.

Figure 8 shows that the FSS-based skillful lead time of nowcasts for 1-h events at a length scale of 5 km is about 18 min. Here, we apply Pearson correlation to compare skillful lead time of nowcasts for different urban areas and event durations. The average Pearson correlation coefficient between nowcast and observed rainfall in each urban area is shown in Fig. 11. The urban areas have similar skillful lead times for both event durations. The average skillful lead times for both event durations are very close: 20 min (1-h events) and 24 min (24-h events). Therefore, nowcast skillful lead times appear to be quite independent of the tested event durations. Although the skillful lead time seems limited, it is 2 times the skillful lead time given by Eulerian persistence. The skillful lead time is compared with other previous research in section 4a, and its application is discussed in section 4b.

Following the result in section 3b(3) that nowcast discrimination ability varies per city, we further compare the discrimination for events of different durations. Figure 9 shows that the ROC curves for the 24-h events display better discrimination than the 1-h events for all three thresholds. The reason is that more persistent events tend to have higher autocorrelations, so a nowcasting model can forecast rainfall more correctly from previous time steps. At low threshold (0.1 mm), the difference between their areas under the ROC curves is larger (15%), whereas the difference is smaller at high thresholds (10% at

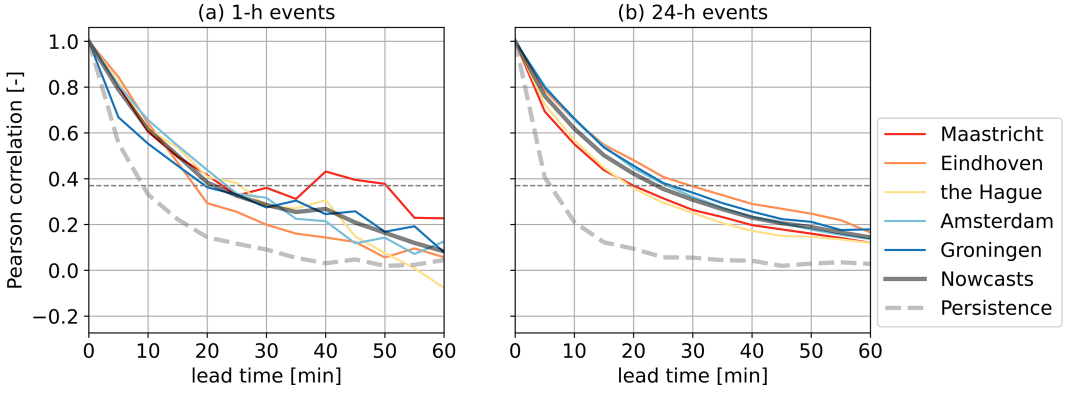

FIG. 11. Pearson correlation coefficient averaged by the eight (a) 1- or (b) 24-h events for each urban area. The horizontal gray dashed line marks a correlation of 0.37 ($1/e$) below which nowcasts are considered as not skillful. The solid black line and the gray dashed line are the average Pearson correlation of nowcasts and Eulerian persistence, respectively.

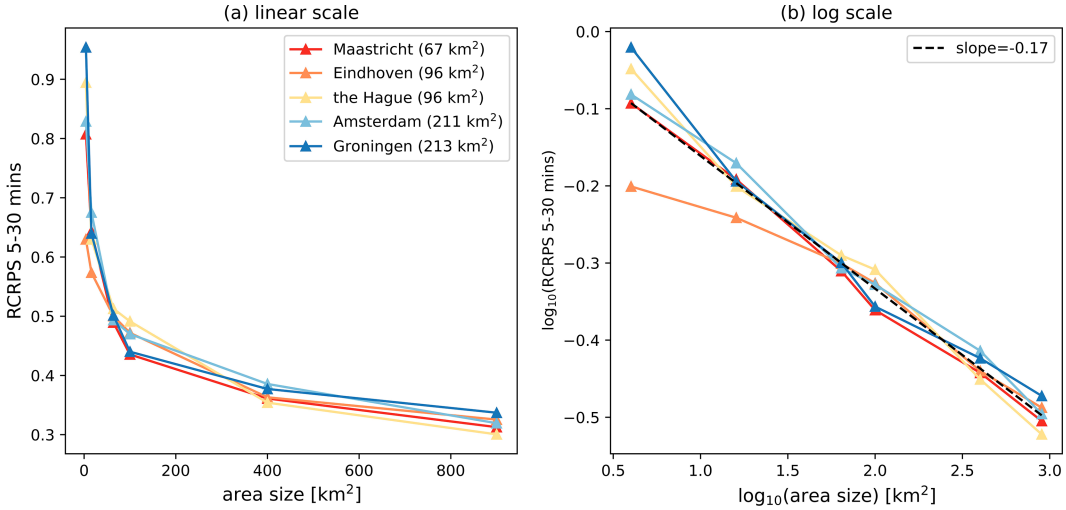

FIG. 12. Average RCRPS (averaged over nowcasts between 5- and 30-min lead times) of the eight 1-h events in each urban area against verification areas from 4 to 900 km^2.

1 mm and 7% at 5 mm). In fact, for both event durations, the AUC drops to values close to the random forecast skill for a threshold above 5 mm h^{-1}. This finding agrees with the poor prediction of high-intensity rainfall shown in Fig. 5.

Concluding, we found that nowcasts improve the underestimation of total rainfall volume for 24-h events relative to Eulerian persistence, whereas the improvement is not apparent for the 1-h events. In addition, the discrimination ability of nowcasts is higher for 24-h events. Despite these differences, skillful lead times do not differ much between 1- and 24-h event durations for the urban areas considered.

2) URBAN SIZES

After analyzing the nowcast for different rainfall durations, we analyze its dependency on the area size by using the square subareas defined in section 2c. Figure 12 indicates that smaller areas tend to have a higher RCRPS, meaning a larger nowcast error. In all urban areas considered, the RCRPS increases nonlinearly as the area reduces from 900 to 4 km^2. Especially when the size reduces to under 100 km^2, the error increases significantly. The enlarging error with reducing areas shows the difficulty of accurately nowcasting rainfall in small urban areas. Note that, although area-mean rainfall generally reduces as the square subareas increase, RCRPS is independent of rainfall magnitude. Skillful lead time also slightly prolongs when area size increases. Pearson correlation of nowcasts over the six subareas around Groningen along with lead time is shown as an example (Fig. S5 in the online supplemental material).

In summary, nowcasts show lower RCRPS and higher spatial correlation to the observed rainfall with increasing (urban) area size. In other words, rainfall at coarser scales (or bigger areas) is nowcasted better than at finer scales. This is probably because larger areas are less prone to rainfall misplacement error, large-scale precipitation features are more persistent (Surcel et al. 2015), and their predictability is higher than small-scale rainfall events (Grecu and Krajewski 2000).

d. *Dependency of nowcast skills on radar QPE products*

As explained in section 2b, the real-time radar QPE can deviate strongly from actual rainfall under certain circumstances, which can result in an additional error source when radar QPE and QPF are used in for instance hydrological applications. Here, we assess the accuracy of the real-time QPE products by using the climatological QPE as "true" rainfall. Second, we evaluate the nowcasts produced from the three different radar rainfall products by comparing them with the climatological QPE as reference.

1) COMPARING THE QPE PRODUCTS

The area-mean rainfall from the three QPE products and the average relative QPE bias during the events before 2020 in each urban area are shown in Fig. 13. It shows that real-time rainfall volumes in all urban areas are lower than in the climatological QPE. The largest underestimation exists in Maastricht (−69%), which is the furthest from the weather radars (Fig. 1). This result is similar to the finding in Imhoff et al. (2021). Note that the 24-h events have larger relative QPE bias in all cities because the rainfall rate is lower so the relative difference between real-time and climatological QPE becomes larger, although the actual underestimation is larger for the 1-h events (4 mm h^{-1} on average) than for the 24-h events (1 mm h^{-1} on average).

Figures 13c and 13d show that CARROTS improves the real-time QPE in terms of rainfall volume. The relative bias of real-time QPE improves from −69% to within −36% in the five urban areas after adjusting with CARROTS. Overall, CARROTS QPE is on average 77% (1-h events) and 92% (24-h events) higher than the real-time QPE. Hence, Using CARROTS QPE to run nowcasts should yield higher rainfall,

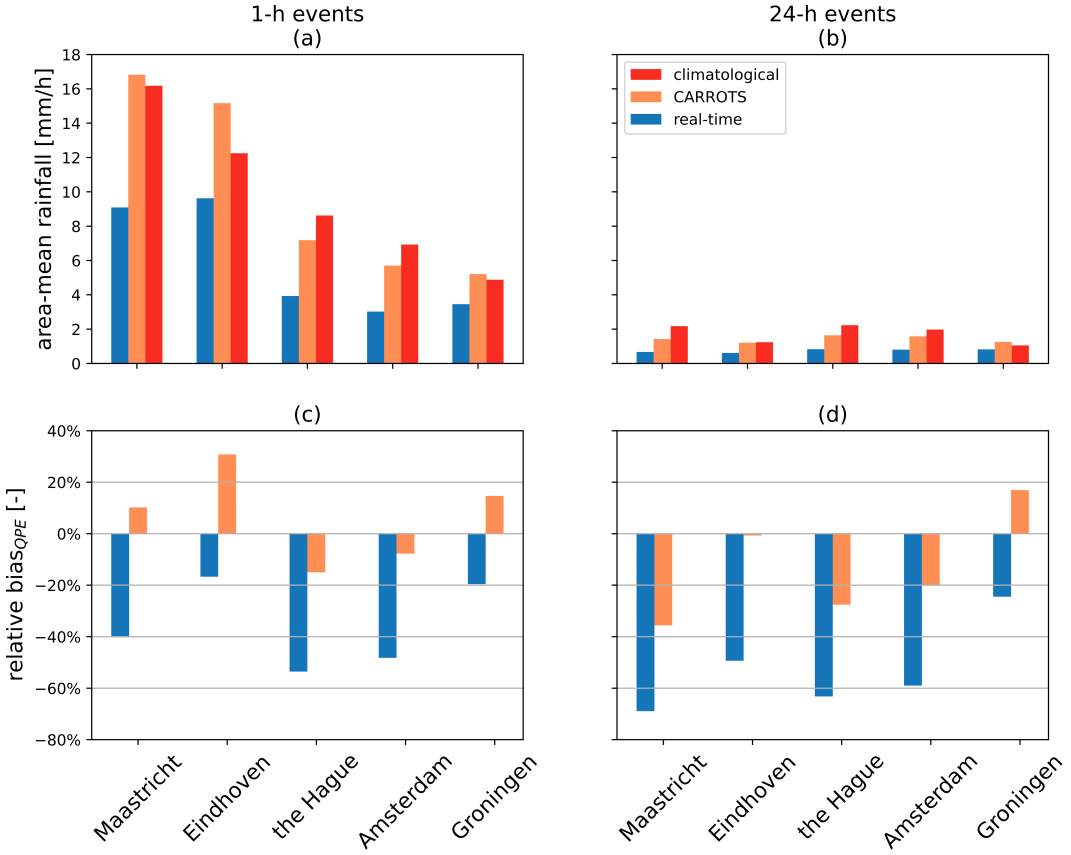

FIG. 13. Average rainfall intensity over the urban areas for (a) 1-h events and (b) 24-h events. Also shown is average relative bias between climatological and real-time radar QPE with and without adjustment by the CARROTS factors for (c) 1-h events and (d) 24-h events. Only the 23 1-h events happening before 2020 are considered.

which is closer to the true rainfall amount, than using unadjusted real-time QPE.

2) EFFECT OF THE QPE PRODUCTS ON THE RESULTING NOWCASTS

The difference in QPE products results in different nowcast outputs. Figure 14 shows the relative bias of area-mean rainfall between nowcasts using the three QPE products and the reference product. It shows that nowcasts using real-time QPE give the largest underestimation in all the urban areas. CARROTS increases the nowcast rainfall volume by 70% on average relative to nowcasts using the real-time QPE.

CARROTS adjusts the nowcast rainfall by correcting the observed rainfall, but it cannot adjust the error caused by the nowcast model. Thus, Fig. 14 shows that nowcasts underestimate rainfall more than 50% above 1-h lead time no matter which radar product is used. Another error is that CARROTS nowcasts sometimes overestimate the rainfall, particularly in Eindhoven and Maastricht within 30-min lead time. This implies that the CARROTS factors are slightly too high in these areas (as also seen in Figs. 13c,d).

Averaging over the five urban areas (Fig. 14f), it may seem like CARROTS QPE is the best input product to use, but this is because the overestimation of the CARROTS QPE (particularly in Eindhoven and Maastricht) compensates for the underestimation resulting from the nowcast model (i.e., better results for the wrong reasons).

CARROTS only corrects the rainfall volumes and not the rainfall spatial distribution. From an analysis of nowcast skill using Pearson correlation (see Fig. S6 in the online supplemental material), we see that the correlation does not change for the three tested products. This indicates that the shape and location of the rainfall fields is hardly altered by the three different QPE products, while the volume is.

In summary, using CARROTS increases the real-time QPE and nowcast rainfall volume by more than 70%, showing better correspondence to climatological QPE. Yet, underestimation of nowcast is still larger than 50% above 40-min lead time. In addition, CARROTS does not improve the spatial correlation of nowcasts with climatological observations.

4. Discussion

In this study, we systematically verified nowcasts of high-intensity rainfall for different urban areas, rainfall characteristics, area sizes, and QPE inputs. In section 4a, we compare the results in the study with previous research. Then, we explain

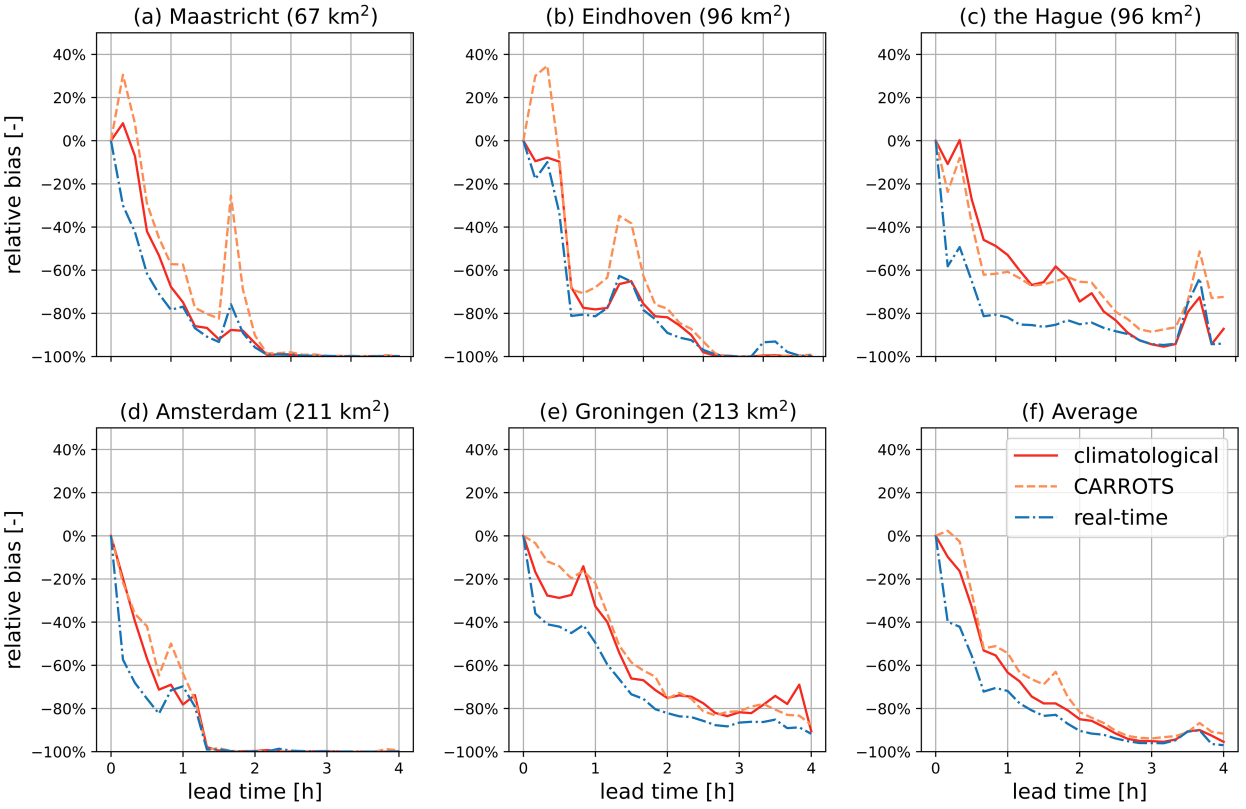

FIG. 14. Relative bias of the nowcasts run with climatological, CARROTS, and real-time radar rainfall from the 23 1-h events before 2020. The relative bias is resampled to 10 min. The relative bias is calculated between the nowcasts and the reference product (the climatological QPE).

the application of the results for operational radar rainfall nowcasting in section 4b. Last, we summarize future improvement possibilities in section 4c.

a. Relation to previous work

We found that the skillful lead time for 1-h high-intensity events over urban areas is on average about 20 min [section 3c(1)], which is similar to previous studies at both urban and rural catchment scales. For instance, skillful lead times are shorter than 20 min at a catchment scale in Germany (Shehu and Haberlandt 2022). The 20-min skillful lead time we found is nearly the same as the 25-min skillful lead time for 384 1-h events in 12 catchments in the Netherlands (Imhoff et al. 2020a). Therefore, radar-based nowcasts have limited skillful lead time to forecast short high-intensity events, for both urban and rural catchment areas. However, note that different methods to derive the decorrelation time, particularly the choice of whether to subtract the mean when calculating the Pearson correlation coefficient [Eq. (7)], can lead to divergence in the resultant skillful lead time (Mejsnar et al. 2018).

We also showed that nowcast discrimination ability reduces as rainfall duration shortens or the threshold of interest rises [section 3b(3)]. Our result is similar to several previous studies. For instance, a study in Belgium showed that nowcast discrimination ability is higher for a threshold of 0.5 mm h^{-1} than for 5 mm h^{-1} (Foresti et al. 2016). Since nowcast discrimination power is generally poorer for more intensive rainfall, nowcasts should be judged meticulously before applying to uses such as issuing flood early warnings.

Unlike previous studies showing that skillful lead time prolongs with rainfall duration, for example, 116 min for 24-h events [derived with the same equation as Eq. (7)] (Imhoff et al. 2020a), the skillful lead time barely changes in this study. The skillful lead time for 24-h events averaged across the five urban areas is only 24 min. There are three main reasons causing the skillful lead time for 24-h events to be almost the same for 1-h events. First, the urban areas in the study are small (67–213 km^2). As shown in sections 3a and 3c, verification of nowcasts in smaller areas is more sensitive to displacement errors of forecast rainfall, so their Pearson correlation becomes lower. Second, this study considers events only based on the highest rainfall accumulation in any grid cell. So, the events in this study have high spatial variability because parts of the areas do not experience rainfall during the events (as shown in Fig. S1 in the online supplemental material). Third, most events in this study are from summers (70 of 80 events are between May and September), so they are generally more convective than the rainfall events chosen by Imhoff et al. (2020a), which reduces the skillful lead time of nowcasts.

b. Operational radar rainfall nowcasting for urban areas

In an operational setting, nowcasts can provide effective early warning insights for (small) urban areas even if their skillful lead time is only around 20 min. The short lead time might be too late to evacuate people, but it can inform people to adopt better shelter-in-place strategies to reduce casualties (Haynes et al. 2009). It can identify areas of hazardous rainfall intensity and warn the community by disseminating real-time warning information on webpages, mobile applications, and social media (Acosta-Coll et al. 2018). Also, warning messages can be sent to local authorities (for instance the fire brigade) to better prepare for emergency actions. In addition, urban discharge takes time to respond to rainfall [e.g., Cristiano et al. (2017) reviews the literature on hydrological response times and Berne et al. (2004) found the response times range between 11.5 and 57 min for six French urban catchments of 0.38 to 104 km^2], which extends the skillful lead time of nowcasts for flow estimation (Sharif et al. 2006; van der Werf et al. 2023). Therefore, nowcasting can still be a timely and useful early warning tool for urban applications.

There are two sources causing underestimation of nowcast rainfall and the resulting short skillful lead time. The first source is the underestimation in the real-time QPE products, as shown in section 3d. Previous studies show that further increasing the resolution of QPE (e.g., to 1 min in time and 200 m in space) enhances the accuracy of the QPE product (Nielsen et al. 2014; Thorndahl et al. 2016), and so the corresponding nowcasts for urban hydrological applications (Thorndahl et al. 2016). Alternative sources such as the QPE converted from signal level data of commercial microwave links outperformed real-time QPE as inputs for nowcasts for high rainfall rates (Imhoff et al. 2020b). In this study, we used a climatology-based adjustment of the QPE and showed that it reduces the QPE error in operational radar rainfall nowcasting (section 3d). However, although correcting the rainfall volumes, these adjustments do not change the spatial distribution of the forecast rainfall fields much. The average CARROTS factor in the considered urban areas during the 1-h events before 2020 was 1.73 (unitless), but their average spatial standard deviation within the urban areas was only 0.04 (about 2% of the average CARROTS factor). Thus, although the CARROTS factors are provided at a high resolution of 1 km^2, they are basically the same for the whole (small) urban area. As a result, the rainfall spatial structure and Pearson correlation remain mostly unchanged, so skillful lead time is not improved (section 3d), only the total rainfall volumes are.

The second source is the nowcasting model itself. From sections 3a to 3d, we identified the challenging rainfall characteristics for the STEPS model: intensive and/or short-lived events occurring over small areas. Similar limitations of operational rainfall nowcasting are also found in other regions in the world. The nowcast model developed by the Czech Hydrometeorological Institute, COTREC (Novák et al. 2009), show skillful lead time shorter than 25 min for local convective storms (Mejsnar et al. 2018). In the Netherlands, NWP can outperform nowcasts after a lead time of 1–2 h for high-intensity convective rainfall events (Imhoff et al. 2023). Likewise, the nowcasting system in Hong Kong (SWIRLS) was also found to have trouble forecasting the growth and decay of rainfall at longer lead times well (Woo and Wong 2017). PIAF, the nowcasting system in France (Moisselin et al. 2019), shows that the quality of nowcasts in France deteriorates quickly and becomes lower than NWP within a lead time of 75 to 90 min for the same reason (Lovat et al. 2022). Comparing nowcast performance with NWP as a function of lead time for rainfall in urban areas would be of interest for future work, because most previous studies focused on national or catchment scales (e.g., Kober et al. 2012. Poletti et al. 2019).

Furthermore, identifying the dominant type of error (i.e., false alarms, misplacement of rainfall, underestimation/overestimation of rainfall volume) based on event type, duration, intensity, and urban area would be an insightful aspect for future work. One main reason for underestimating intensive events is that localized convective rainfall can develop in tens of minutes (Germann et al. 2006; Mejsnar et al. 2018). Most radar-based nowcasts do not take the growth of new rainfall into account (Bowler et al. 2006; James et al. 2018; Pulkkinen et al. 2019; Shehu and Haberlandt 2021) because they do not consider information on atmospheric stability, such as convective available potential energy (CAPE) and moisture convergence, which can lead to convective rainfall (Steinheimer and Haiden 2007; Sun et al. 2014). Thus, small-area and short-lived rainfall that is not represented in the original QPE cannot be forecast by current radar-based nowcasting models and, as a result, underestimation of rainfall in the forecast is unavoidable.

c. Future perspectives

Some innovative nowcasting procedures outlined below have been proposed to tackle the mentioned current limits of operational rainfall nowcasting. However, it is still not clear whether they will lead to dramatic improvements during strong and convective events that are intrinsically chaotic. Last, more analysis that can supplement this research is proposed.

1) RECENT DEVELOPMENT IN NOWCASTING TECHNIQUES

Because of the difficulty to parameterize the nonlinear growth and decay of extreme rainfall, recent research has started to integrate machine learning techniques into precipitation nowcasting (Marrocu and Massidda 2020; Li et al. 2021; Ravuri et al. 2021; Ehsani et al. 2022). They show that nowcasting models based on machine learning outperform traditional optical-flow methods from certain perspectives. Machine learning models forecast rainfall evolution by training on previous events and avoid the heavy reliance on the advection equations used in optical flow methods. Thus, they have potential to forecast convective growth which is not considered by deterministic nowcasting models that assume a constant rainfall rate. With the above literature, machine learning methods are proven to be valuable tools in extreme rainfall nowcasting and should be

considered in future work on enhancing nowcasting skills for urban areas.

Blending numerical weather prediction (NWP) with nowcasting models is also shown to improve nowcasting skills. NWP numerically solves atmospheric processes, so blending NWP with radar-based nowcasts can improve the forecast on rainfall evolution that cannot be achieved solely by radar-based nowcasting models (Kober et al. 2012; Kato et al. 2017; Poletti et al. 2019; Sideris et al. 2020; Imhoff et al. 2023). With the significantly improved efficiency in computation and resolution of NWP in the last decade, blending NWP in the nowcasting process has a lot of potential to advance nowcast skills (Sun et al. 2014; Poletti et al. 2019). Thus, using the STEPS procedure that blends QPE extrapolation and NWP might be a potential next step for advancing nowcasting skill (Bowler et al. 2006; Seed et al. 2013).

Another potential method to better simulate rainfall growth is three-dimensional nowcasting. 3D scanned temperature, humidity, and wind data can be used to indicate convective stability to complement radar-based nowcasting results (Steinheimer and Haiden 2007). Using 3D reflectivity, it is possible to nowcast extreme convective events like thunderstorms even with a lifetime below 60 min (Yoshikawa et al. 2012; Ushio et al. 2015). In essence, extrapolating 3D radar reflectivity outperforms 2D nowcasting because it better recognizes the vertical motion of rainfall convection (Otsuka et al. 2016; Sun et al. 2022). Furthermore, combining a convolutional neural network with 3D radar reflectivity may further improve nowcast quality (Kim et al. 2021).

2) Effects of urban features on nowcasting accuracy

This study discusses nowcasting accuracy in different urban areas, by focusing on rainfall characteristics as well as sizes and locations of the urban areas. However, other urban features may change nowcast accuracy as well. For instance, the five urban areas considered have different land use and building structures, which might impose additional uncertainty to nowcasts. Urban roughness can create flow anomalies by changing the intensity and position of moisture convergence (Yang et al. 2021). Also, urban heat and aerosols can modify rainfall in and around the cities (Shepherd 2005; Schmid and Niyogi 2017; Sarangi et al. 2018), with precipitation amounts that can be 16% higher above a city (Liu and Niyogi 2019). Furthermore, city shapes are also found to have notable influences on rainfall accumulation (Zhang et al. 2022). Such phenomena may add higher difficulty to precipitation nowcasting around urban areas than in rural regions. Therefore, future analysis could systematically compare nowcast performance for urban and rural areas while controlling other variables like domain sizes and distances to the nearest weather radar.

5. Conclusions

With this study, we aim to provide insight into the potential of radar nowcasting for rainfall warnings in small urban areas. We analyzed the quality of ensemble radar rainfall nowcasting at the urban scale using the short-term ensemble prediction system (STEPS) from Pysteps. We chose 80 high-intensity rainfall events in five urban areas in the Netherlands. We analyzed the corresponding 9060 nowcasts (with a lead time of 4 h and 20 ensemble members) and focused on nowcast performance for different urban areas, the factors that affect nowcast performance, and the effects of using different QPE products.

Overall, the nowcast accuracy varied in the five urban areas mainly due to different area sizes. We found nowcast RCRPS increases nonlinearly when (urban) area size decreases from 900 to 4 km^2, because smaller areas are more sensitive to misplaced rainfall in the forecast. Because of the nonlinear relationship, nowcast errors increase more strongly for areas below 100 km^2. These findings imply the limitation on spatial accuracy for the employed STEPS method and radar resolution (1 km and 5 min).

Nowcasting performance also depended on the rainfall duration. The rainfall underestimation of the nowcasts was more severe for 1-h events than for 24-h events. Nowcasts estimated the total rainfall volume accurately (error within 20%) with a lead time up to 2.5 h for the 24-h events, but the error was larger than 50% within 30-min lead time for the 1-h events. The discrimination skill of nowcasts at 30-min lead time was also higher for 24-h events than for 1-h events by 11% (threshold = 0.1 mm). For both event durations, AUC reduced as the threshold of interest increased.

To also distinguish the nowcasts error from the QPE errors in an operational setting, different quantitative precipitation estimates (QPEs) in the Netherlands were compared and run with STEPS. We found that real-time QPE underestimates the climatological QPE (best reference in the Netherlands) by 17%–69% depending on the regions and rainfall duration. Climatology-based Adjustments for Radar Rainfall in an Operational Setting (CARROTS) reduced the error in real-time QPE to within 36%. Accordingly, using CARROTS-adjusted QPE to run nowcasts also reduced the rainfall underestimation of real-time nowcasts by 70%, relative to the real-time QPE product, during the 4-h lead time tested. Yet, underestimation of rainfall is still above 50% when lead times are longer than 40 min, which can be attributed to the error in the forecast from the nowcasting model. Besides, the average skillful lead time found for the 1-h events was limited to 20 min regardless of the QPE product used. The reason is that CARROTS adjusts the rainfall volume but does not alter the spatial distribution of the rainfall fields much in small areas.

This study shows that radar-based nowcasting techniques based on optical flow methods can provide skillful forecasts for very short lead times of up to 24 min in urban areas, which can be precious for emergency responses. Such a lead time is a common obstacle that many operational nowcasting models are challenged with when forecasting high-intensity rainfall in urban areas. To advance nowcasting skills, it is recommended to use higher-quality QPE products (e.g., via real-time adjustments or higher resolution) and employ machine learning, 3D nowcasting, or blending nowcasts with numerical weather prediction (NWP) to better model the growth and dissipation of rainfall.

Acknowledgments. We acknowledge Aart Overeem from Royal Netherlands Meteorological Institute (KNMI) for

providing information about the KNMI radar rainfall products. We are grateful for the free use of RadarTools offered by the RIONED foundation. We are thankful for the radar rainfall data provided by KNMI and the municipal data provided by Statistics Netherlands. RI was supported by Deltares's strategic research under the Dutch Subsidy for Institutes for Applied Research. RU acknowledges support from the 4TU program HERITAGE (HEat Robustness In relation To AGEing cities), funded by the High Tech for a Sustainable Future (HTSF) program of 4TU, the federation of the four technical universities in The Netherlands. The authors acknowledge the use of computational resources of DelftBlue supercomputer, provided by Delft High Performance Computing Centre (https://www.tudelft.nl/dhpc).

Data availability statement. Nowcasts analyzed during the current study are available on 4TU Research Data (https://doi.org/10.4121/22246273.v1). These datasets were derived from the 5-min real-time radar rainfall accumulation data (https://dataplatform.knmi.nl/dataset/nl-rdr-data-rtcor-5m-1-0) and 5-min gauge-adjusted climatological radar rainfall accumulation (https://dataplatform.knmi.nl/dataset/rad-nl25-rac-mfbs-em-5min-netcdf4-2-0). The daily CARROTS factors used to adjust the real-time QPE are available online (https://doi.org/10.4121/13573814; Imhoff et al. 2021).

REFERENCES

Achleitner, S., S. Fach, T. Einfalt, and W. Rauch, 2009: Nowcasting of rainfall and of combined sewage flow in urban drainage systems. *Water Sci. Technol.*, **59**, 1145–1151, https://doi.org/10.2166/wst.2009.098.

Acosta-Coll, M., F. Ballester-Merelo, and M. Martínez-Peiró, 2018: Early warning system for detection of urban pluvial flooding hazard levels in an ungauged basin. *Nat. Hazards*, **92**, 1237–1265, https://doi.org/10.1007/s11069-018-3249-4.

Alfieri, L., P. Salamon, F. Pappenberger, F. Wetterhall, and J. Thielen, 2012: Operational early warning systems for water-related hazards in Europe. *Environ. Sci. Policy*, **21**, 35–49, https://doi.org/10.1016/j.envsci.2012.01.008.

Allaire, M., 2018: Socio-economic impacts of flooding: A review of the empirical literature. *Water Secur.*, **3**, 18–26, https://doi.org/10.1016/j.wasec.2018.09.002.

Beekhuis, H., and I. Holleman, 2008: From pulse to product, highlights of the digital-IF upgrade of the Dutch national radar network. *Proc. Fifth European Conf. on Radar in Meteorology and Hydrology (ERAD 2008)*, Helsinki, Finland, Finnish Meteorological Institute, 1–3, https://cdn.knmi.nl/system/data_center_publications/files/000/068/061/original/erad2008drup_0120.pdf?1495621011.

——, and T. Mathijssen, 2018: From pulse to product, highlights of the upgrade project of the Dutch national weather radar network. *Proc. 10th European Conf. on Radar in Meteorology and Hydrology (ERAD)*, Ede-Wageningen, The Netherlands, Wageningen University & Research, 960–965, https://edepot.wur.nl/454537.

Beersma, J., H. Hakvoort, R. Jilderda, A. Overeem, and R. Versteeg, 2019: Neerslagstatistiek En -Reeksen Voor Het Waterbeheer, 202 pp., https://www.stowa.nl/sites/default/files/assets/PUBLICATIES/Publicaties%202019/STOWA%202019-19%20neerslagstatistieken.pdf.

Berenguer, M., D. Sempere-Torres, and G. G. S. Pegram, 2011: SBMcast – An ensemble nowcasting technique to assess the uncertainty in rainfall forecasts by Lagrangian extrapolation. *J. Hydrol.*, **404**, 226–240, https://doi.org/10.1016/j.jhydrol.2011.04.033.

Berne, A., G. Delrieu, J.-D. Creutin, and C. Obled, 2004: Temporal and spatial resolution of rainfall measurements required for urban hydrology. *J. Hydrol.*, **299**, 166–179, https://doi.org/10.1016/S0022-1694(04)00363-4.

Bowler, N. E., C. E. Pierce, and A. W. Seed, 2006: STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP. *Quart. J. Roy. Meteor. Soc.*, **132**, 2127–2155, https://doi.org/10.1256/qj.04.100.

Chang, P.-L., and Coauthors, 2021: An operational multi-radar multi-sensor QPE system in Taiwan. *Bull. Amer. Meteor. Soc.*, **102**, E555–E577, https://doi.org/10.1175/BAMS-D-20-0043.1.

Choi, S., and Y. Kim, 2022: Rad-cGAN v1.0: Radar-based precipitation nowcasting model with conditional Generative adversarial networks for multiple dam domains. *Geosci. Model Dev.*, **15**, 5967–5985, https://doi.org/10.5194/gmd-15-5967-2022.

CNN, 2021: Germany's deadly floods were up to 9 times more likely because of climate change, study estimates. CNN, accessed 21 August 2022, https://edition.cnn.com/2021/08/23/europe/germany-floods-belgium-climate-change-intl/index.html.

Cristiano, E., M.-C. ten Veldhuis, and N. van de Giesen, 2017: Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas—A review. *Hydrol. Earth Syst. Sci.*, **21**, 3859–3878, https://doi.org/10.5194/hess-21-3859-2017.

Davolio, S., F. Silvestro, and P. Malguzzi, 2015: Effects of increasing horizontal resolution in a convection-permitting model on flood forecasting: The 2011 dramatic events in Liguria, Italy. *J. Hydrometeor.*, **16**, 1843–1856, https://doi.org/10.1175/JHM-D-14-0094.1.

Ehsani, M. R., A. Zarei, H. V. Gupta, K. Barnard, E. Lyons, and A. Behrangi, 2022: NowCasting-Nets: Representation learning to mitigate latency gap of satellite precipitation products using convolutional and recurrent neural networks. *IEEE Trans. Geosci. Remote Sens.*, **60**, 1–21, https://doi.org/10.1109/TGRS.2022.3158888.

Foresti, L., M. Reyniers, A. Seed, and L. Delobbe, 2016: Development and verification of a real-time stochastic precipitation nowcasting system for urban hydrology in Belgium. *Hydrol. Earth Syst. Sci.*, **20**, 505–527, https://doi.org/10.5194/hess-20-505-2016.

Germann, U., and I. Zawadzki, 2002: Scale-dependence of the predictability of precipitation from continental radar images. Part I: Description of the methodology. *Mon. Wea. Rev.*, **130**, 2859–2873, https://doi.org/10.1175/1520-0493(2002)130<2859:SDOTPO>2.0.CO;2.

——, ——, and B. Turner, 2006: Predictability of precipitation from continental radar images. Part IV: Limits to prediction. *J. Atmos. Sci.*, **63**, 2092–2108, https://doi.org/10.1175/JAS3735.1.

Golding, B. W., 1998: Nimrod: A system for generating automated very short range forecasts. *Meteor. Appl.*, **5** (1), 1–16, https://doi.org/10.1017/S1350482798000577.

Grecu, M., and W. F. Krajewski, 2000: A large-sample investigation of statistical procedures for radar-based short-term

quantitative precipitation forecasting. *J. Hydrol.*, **239**, 69–84, https://doi.org/10.1016/S0022-1694(00)00360-7.

Haynes, K., L. Coates, R. Leigh, J. Handmer, J. Whittaker, A. Gissing, J. Mcaneney, and S. Opper, 2009: 'Shelter-in-place' vs. evacuation in flash floods. *Environ. Hazards*, **8**, 291–303, https://doi.org/10.3763/ehaz.2009.0022.

Hazenberg, P., H. Leijnse, and R. Uijlenhoet, 2011: Radar rainfall estimation of stratiform winter precipitation in the Belgian Ardennes. *Water Resour. Res.*, **47**, W02507, https://doi.org/10.1029/2010WR009068.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559–570, https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Heuvelink, D., M. Berenguer, C. C. Brauer, and R. Uijlenhoet, 2020: Hydrological application of radar rainfall nowcasting in the Netherlands. *Environ. Int.*, **136**, 105431, https://doi.org/10.1016/j.envint.2019.105431.

Imhoff, R. O., C. C. Brauer, A. Overeem, A. H. Weerts, and R. Uijlenhoet, 2020a: Spatial and temporal evaluation of radar rainfall nowcasting techniques on 1,533 events. *Water Resour. Res.*, **56**, e2019WR026723, https://doi.org/10.1029/2019WR026723.

——, A. Overeem, C. C. Brauer, H. Leijnse, A. H. Weerts, and R. Uijlenhoet, 2020b: Rainfall nowcasting using commercial microwave links. *Geophys. Res. Lett.*, **47**, e2020GL089365, https://doi.org/10.1029/2020GL089365.

——, C. Brauer, K.-J. van Heeringen, H. Leijnse, A. Overeem, A. Weerts, and R. Uijlenhoet, 2021: A climatological benchmark for operational radar rainfall bias reduction. *Hydrol. Earth Syst. Sci.*, **25**, 4061–4080, https://doi.org/10.5194/hess-25-4061-2021.

——, C. C. Brauer, K.-J. van Heeringen, R. Uijlenhoet, and A. H. Weerts, 2022: Large-sample evaluation of radar rainfall nowcasting for flood early warning. *Water Resour. Res.*, **58**, e2021WR031591, https://doi.org/10.1029/2021WR031591.

——, and Coauthors, 2023: Scale-dependent blending of ensemble rainfall nowcasts and numerical weather prediction in the open-source pysteps library. *Quart. J. Roy. Meteor. Soc.*, **149**, 1335–1364, https://doi.org/10.1002/qj.4461.

IPCC, 2012: *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. C. B. Field et al., Eds., Cambridge University Press, 582 pp.

——, 2021: *Climate Change 2021: The Physical Science Basis*. V. Masson-Delmotte et al., Eds., Cambridge University Press, 2391 pp.

James, P. M., B. K. Reichert, and D. Heizenreder, 2018: NowCastMIX: Automatic integrated warnings for severe convection on nowcasting time scales at the German Weather Service. *Wea. Forecasting*, **33**, 1413–1433, https://doi.org/10.1175/WAF-D-18-0038.1.

Jensen, D. G., C. Petersen, and M. R. Rasmussen, 2015: Assimilation of radar-based nowcast into a HIRLAM NWP model. *Meteor. Appl.*, **22**, 485–494, https://doi.org/10.1002/met.1479.

Kasmalkar, I. G., K. A. Serafin, Y. Miao, I. Avery Bick, L. Ortolano, D. Ouyang, and J. Suckale, 2020: When floods hit the road: Resilience to flood-related traffic disruption in the San Francisco Bay area and beyond. *Sci. Adv.*, **6**, eaba2423, https://doi.org/10.1126/sciadv.aba2423.

Kato, R., S. Shimizu, K.-i. Shimose, T. Maesaka, K. Iwanami, and H. Nakagaki, 2017: Predictability of meso-γ-scale, localized, extreme heavy rainfall during the warm season in Japan using high-resolution precipitation nowcasts. *Quart. J. Roy. Meteor. Soc.*, **143**, 1406–1420, https://doi.org/10.1002/qj.3013.

Kim, D. K., T. Suezawa, T. Mega, H. Kikuchi, E. Yoshikawa, P. Baron, and T. Ushio, 2021: Improving precipitation nowcasting using a three-dimensional convolutional neural network model from multi parameter phased array weather radar observations. *Atmos. Res.*, **262**, 105774, https://doi.org/10.1016/j.atmosres.2021.105774.

Kober, K., G. C. Craig, C. Keil, and A. Dörnbrack, 2012: Blending a probabilistic nowcasting method with a high-resolution numerical weather prediction ensemble for convective precipitation forecasts. *Quart. J. Roy. Meteor. Soc.*, **138**, 755–768, https://doi.org/10.1002/qj.939.

Koks, E. E., K. C. H. van Ginkel, M. J. E. van Marle, and A. Lemnitzer, 2022: Brief communication: Critical infrastructure impacts of the 2021 mid-July western European flood event. *Nat. Hazards Earth Syst. Sci.*, **22**, 3831–3838, https://doi.org/10.5194/nhess-22-3831-2022.

Kotroni, V., and K. Lagouvardos, 2004: Evaluation of MM5 high-resolution real-time forecasts over the urban area of Athens, Greece. *J. Appl. Meteor.*, **43**, 1666–1678, https://doi.org/10.1175/JAM2170.1.

Krajewski, W. F., G. Villarini, and J. A. Smith, 2010: Radar-rainfall uncertainties: Where are we after thirty years of effort? *Bull. Amer. Meteor. Soc.*, **91**, 87–94, https://doi.org/10.1175/2009BAMS2747.1.

Lalonde, M., L. Oudin, and S. Bastin, 2023: Urban effects on precipitation: Do the diversity of research strategies and urban characteristics preclude general conclusions? *Urban Climate*, **51**, 101605, https://doi.org/10.1016/j.uclim.2023.101605.

Leutbecher, M., and T. Haiden, 2021: Understanding changes of the continuous ranked probability score using a homogeneous Gaussian approximation. *Quart. J. Roy. Meteor. Soc.*, **147**, 425–442, https://doi.org/10.1002/qj.3926.

Li, D., Y. Liu, and C. Chen, 2021: MSDM v1.0: A machine learning model for precipitation nowcasting over eastern China using multisource data. *Geosci. Model Dev.*, **14**, 4019–4034, https://doi.org/10.5194/gmd-14-4019-2021.

Liguori, S., M. A. Rico-Ramirez, A. N. A. Schellart, and A. J. Saul, 2012: Using probabilistic radar rainfall nowcasts and NWP forecasts for flow prediction in urban catchments. *Atmos. Res.*, **103**, 80–95, https://doi.org/10.1016/j.atmosres.2011.05.004.

Liu, J., and D. Niyogi, 2019: Meta-analysis of urbanization impact on rainfall modification. *Sci. Rep.*, **9**, 7301, https://doi.org/10.1038/s41598-019-42494-2.

Lovat, A., B. Vincendon, and V. Ducrocq, 2022: Hydrometeorological evaluation of two nowcasting systems for Mediterranean heavy precipitation events with operational considerations. *Hydrol. Earth Syst. Sci.*, **26**, 2697–2714, https://doi.org/10.5194/hess-26-2697-2022.

Madsen, H., D. Lawrence, M. Lang, M. Martinkova, and T. R. Kjeldsen, 2014: Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. *J. Hydrol.*, **519**, 3634–3650, https://doi.org/10.1016/j.jhydrol.2014.11.003.

Maier, R., G. Krebs, M. Pichler, D. Muschalla, and G. Gruber, 2020: Spatial rainfall variability in urban environments—High-density precipitation measurements on a city-scale. *Water*, **12**, 1157, https://doi.org/10.3390/w12041157.

Marrocu, M., and L. Massidda, 2020: Performance comparison between deep learning and optical flow-based techniques for

nowcast precipitation from radar images. *Forecasting*, **2**, 194–210, https://doi.org/10.3390/forecast2020011.

Marshall, J. S., W. Hitschfeld, and K. L. S. Gunn, 1955: Advances in radar weather. *Advances in Geophysics*, Vol. 2, Academic Press, 1–56, https://doi.org/10.1016/S0065-2687(08)60310-6.

Mejsnar, J., Z. Sokol, and J. Minářová, 2018: Limits of precipitation nowcasting by extrapolation of radar reflectivity for warm season in Central Europe. *Atmos. Res.*, **213**, 288–301, https://doi.org/10.1016/j.atmosres.2018.06.005.

Merz, B., H. Kreibich, R. Schwarze, and A. Thieken, 2010: Review article "Assessment of economic flood damage". *Nat. Hazards Earth Syst. Sci.*, **10**, 1697–1724, https://doi.org/10.5194/nhess-10-1697-2010.

Moisselin, J.-M., P. Cau, C. Jauffret, I. Bouissières, and R. Tzanos, 2019: Seamless approach for precipitations within the 0–3 hours forecast-interval. *European Nowcasting Conf.*, Madrid, Spain, EUMETNET, https://repositorio.aemet.es/bitstream/20.500.11765/10588/3/SPDA2_Moisselin_3ENC2019.pdf.

Nielsen, J. E., S. Thorndahl, and M. R. Rasmussen, 2014: A numerical method to generate high temporal resolution precipitation time series by combining weather radar measurements with a nowcast model. *Atmos. Res.*, **138**, 1–12, https://doi.org/10.1016/j.atmosres.2013.10.015.

Novák, P., L. Březková, and P. Frolík, 2009: Quantitative precipitation forecast using radar echo extrapolation. *Atmos. Res.*, **93**, 328–334, https://doi.org/10.1016/j.atmosres.2008.10.014.

Ochoa-Rodriguez, S., and Coauthors, 2015: Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation. *J. Hydrol.*, **531**, 389–407, https://doi.org/10.1016/j.jhydrol.2015.05.035.

——, L. P. Wang, P. Willems, and C. Onof, 2019: A review of radar-rain gauge data merging methods and their potential for urban hydrological applications. *Water Resour. Res.*, **55**, 6356–6391, https://doi.org/10.1029/2018WR023332.

Olsson, J., L. Simonsson, and M. Ridal, 2014: Rainfall nowcasting: Predictability of short-term extremes in Sweden. *Urban Water J.*, **11**, 605–615, https://doi.org/10.1080/1573062X.2013.847465.

Otsuka, S., and Coauthors, 2016: Precipitation nowcasting with three-dimensional space–time extrapolation of dense and frequent phased-array weather radar observations. *Wea. Forecasting*, **31**, 329–340, https://doi.org/10.1175/WAF-D-15-0063.1.

Overeem, A., I. Holleman, and A. Buishand, 2009a: Derivation of a 10-year radar-based climatology of rainfall. *J. Appl. Meteor. Climatol.*, **48**, 1448–1463, https://doi.org/10.1175/2009JAMC1954.1.

——, T. A. Buishand, and I. Holleman, 2009b: Extreme rainfall analysis and estimation of depth-duration-frequency curves using weather radar. *Water Resour. Res.*, **45**, W10424, https://doi.org/10.1029/2009WR007869.

——, ——, ——, and R. Uijlenhoet, 2010: Extreme value modeling of areal rainfall from weather radar. *Water Resour. Res.*, **46**, W09514, https://doi.org/10.1029/2009WR008517.

——, H. Leijnse, and R. Uijlenhoet, 2011: Measuring urban rainfall using microwave links from commercial cellular communication networks. *Water Resour. Res.*, **47**, W12505, https://doi.org/10.1029/2010WR010350.

——, H. de Vries, H. Al Sakka, R. Uijlenhoet, and H. Leijnse, 2021: Rainfall-induced attenuation correction for two operational dual-polarization C-band radars in the Netherlands. *J. Atmos. Oceanic Technol.*, **38**, 1125–1142, https://doi.org/10.1175/JTECH-D-20-0113.1.

Poletti, M. L., F. Silvestro, S. Davolio, F. Pignone, and N. Rebora, 2019: Using nowcasting technique and data assimilation in a meteorological model to improve very short range hydrological forecasts. *Hydrol. Earth Syst. Sci.*, **23**, 3823–3841, https://doi.org/10.5194/hess-23-3823-2019.

Pulkkinen, S., D. Nerini, A. A. Pérez Hortal, C. Velasco-Forero, A. Seed, U. Germann, and L. Foresti, 2019: Pysteps: An open-source Python library for probabilistic precipitation nowcasting (v1.0). *Geosci. Model Dev.*, **12**, 4185–4219, https://doi.org/10.5194/gmd-12-4185-2019.

Rafieeinasab, A., and Coauthors, 2015: Toward high-resolution flash flood prediction in large urban areas–Analysis of sensitivity to spatiotemporal resolution of rainfall input and hydrologic model. *J. Hydrol.*, **531**, 370–388, https://doi.org/10.1016/j.jhydrol.2015.08.045.

Ralph, F. M., and Coauthors, 2014: A vision for future observations for western US extreme precipitation and flooding. *J. Contemp. Water Res. Educ.*, **153**, 16–32, https://doi.org/10.1111/j.1936-704X.2014.03176.x.

Ravuri, S., and Coauthors, 2021: Skilful precipitation nowcasting using deep generative models of radar. *Nature*, **597**, 672–677, https://doi.org/10.1038/s41586-021-03854-z.

RIONED, 2020: RadarTools. Accessed 10 March 2022, https://radartools.nl.

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, https://doi.org/10.1175/2007MWR2123.1.

Royal Netherlands Meteorological Institute, 2022a: Precipitation—5 minute precipitation accumulations from climatological gauge-adjusted radar dataset for The Netherlands (1 km, extended mask) in KNMI HDF5 format, version 2.0. KNMI, accessed 1 May 2022, https://dataplatform.knmi.nl/dataset/rad-nl25-rac-mfbs-em-5min-2-0.

——, 2022b: Precipitation—radar/gauge 5 minute real-time accumulations over the Netherlands, version 1.0. KNMI, accessed 1 May 2022, https://dataplatform.knmi.nl/dataset/nl-rdr-data-rtcor-5m-1-0.

——, 2023: Precipitation—radar/gauge 5 minute final reanalysis accumulations over the Netherlands, version 1.0. KNMI, accessed 4 November 2023, https://dataplatform.knmi.nl/dataset/nl-rdr-data-rfcor-5m-1-0.

Sarangi, C., S. N. Tripathi, Y. Qian, S. Kumar, and L. Ruby Leung, 2018: Aerosol and urban land use effect on rainfall around cities in Indo-Gangetic Basin from observations and cloud resolving model simulations. *J. Geophys. Res. Atmos.*, **123**, 3645–3667, https://doi.org/10.1002/2017JD028004.

Schellart, A., S. Liguori, S. Krämer, A. Saul, and M. A. Rico-Ramirez, 2014: Comparing quantitative precipitation forecast methods for prediction of sewer flows in a small urban area. *Hydrol. Sci. J.*, **59**, 1418–1436, https://doi.org/10.1080/02626667.2014.920505.

Schleiss, M., and Coauthors, 2020: The accuracy of weather radar in heavy rain: A comparative study for Denmark, the Netherlands, Finland and Sweden. *Hydrol. Earth Syst. Sci.*, **24**, 3157–3188, https://doi.org/10.5194/hess-24-3157-2020.

Schmid, P. E., and D. Niyogi, 2017: Modeling urban precipitation modification by spatially heterogeneous aerosols. *J. Appl. Meteor. Climatol.*, **56**, 2141–2153, https://doi.org/10.1175/JAMC-D-16-0320.1.

Seed, A. W., 2003: A dynamic and spatial scaling approach to advection forecasting. *J. Appl. Meteor.*, **42**, 381–388, https://doi.org/10.1175/1520-0450(2003)042<0381:ADASSA>2.0.CO;2.

——, C. E. Pierce, and K. Norman, 2013: Formulation and evaluation of a scale decomposition-based stochastic precipitation nowcast scheme. *Water Resour. Res.*, **49**, 6624–6641, https://doi.org/10.1002/wrcr.20536.

Sharif, H. O., D. Yates, R. Roberts, and C. Mueller, 2006: The use of an automated nowcasting system to forecast flash floods in an urban watershed. *J. Hydrometeor.*, **7**, 190–202, https://doi.org/10.1175/JHM482.1.

Shehu, B., and U. Haberlandt, 2021: Relevance of merging radar and rainfall gauge data for rainfall nowcasting in urban hydrology. *J. Hydrol.*, **594**, 125931, https://doi.org/10.1016/j.jhydrol.2020.125931.

——, and ——, 2022: Improving radar-based rainfall nowcasting by a nearest-neighbour approach–Part 1: Storm characteristics. *Hydrol. Earth Syst. Sci.*, **26**, 1631–1658, https://doi.org/10.5194/hess-26-1631-2022.

Shepherd, J. M., 2005: A review of current investigations of urban-induced rainfall and recommendations for the future. *Earth Interact.*, **9**, https://doi.org/10.1175/EI156.1.

Sideris, I. V., L. Foresti, D. Nerini, and U. Germann, 2020: NowPrecip: Localized precipitation nowcasting in the complex terrain of Switzerland. *Quart. J. Roy. Meteor. Soc.*, **146**, 1768–1800, https://doi.org/10.1002/qj.3766.

Silvestro, F., N. Rebora, F. Giannoni, A. Cavallo, and L. Ferraris, 2016: The flash flood of the Bisagno Creek on 9th October 2014: An "unfortunate" combination of spatial and temporal scales. *J. Hydrol.*, **541**, 50–62, https://doi.org/10.1016/j.jhydrol.2015.08.004.

Statistics Netherlands, 2020: District and neighborhood map 2020, version 3. Statistics Netherlands, accessed 1 March 2022, https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/wijk-en-buurtkaart-2020.

——, 2021: Regional key figures for the Netherlands. Statistics Netherlands, accessed 18 March 2022, https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70072ned/table?dl=5A35F.

Steinheimer, M., and T. Haiden, 2007: Improved nowcasting of precipitation based on convective analysis fields. *Adv. Geosci.*, **10**, 125–131, https://doi.org/10.5194/adgeo-10-125-2007.

Sun, J., and Coauthors, 2014: Use of NWP for nowcasting convective precipitation: Recent progress and challenges. *Bull. Amer. Meteor. Soc.*, **95**, 409–426, https://doi.org/10.1175/BAMS-D-11-00263.1.

Sun, N., Z. Zhou, Q. Li, and J. Jing, 2022: Three-dimensional gridded radar echo extrapolation for convective storm nowcasting based on 3D-ConvLSTM model. *Remote Sens.*, **14**, 4256, https://doi.org/10.3390/rs14174256.

Surcel, M., I. Zawadzki, and M. K. Yau, 2015: A study on the scale dependence of the predictability of precipitation patterns. *J. Atmos. Sci.*, **72**, 216–235, https://doi.org/10.1175/JAS-D-14-0071.1.

Tabari, H., 2020: Climate change impact on flood and extreme precipitation increases with water availability. *Sci. Rep.*, **10**, 13768, https://doi.org/10.1038/s41598-020-70816-2.

The Guardian, 2016: Paris floods: 'There's something terrifying about it.' *Guardian*, 4 June, https://www.theguardian.com/world/2016/jun/03/paris-river-seine-floods.

Thorndahl, S., J. E. Nielsen, and D. G. Jensen, 2016: Urban pluvial flood prediction: A case study evaluating radar rainfall nowcasts and numerical weather prediction models as model inputs. *Water Sci. Technol.*, **74**, 2599–2610, https://doi.org/10.2166/wst.2016.474.

Tingsanchali, T., 2012: Urban flood disaster management. *Procedia Eng.*, **32**, 25–37, https://doi.org/10.1016/j.proeng.2012.01.1233.

Trinh, B. N., J. Thielen-del Pozo, and G. Thirel, 2013: The reduction continuous rank probability score for evaluating discharge forecasts from hydrological ensemble prediction systems. *Atmos. Sci. Lett.*, **14**, 61–65, https://doi.org/10.1002/asl2.417.

Turner, B. J., I. Zawadzki, and U. Germann, 2004: Predictability of precipitation from continental radar images. Part III: Operational nowcasting implementation (MAPLE). *J. Appl. Meteor.*, **43**, 231–248, https://doi.org/10.1175/1520-0450(2004)043<0231:POPFCR>2.0.CO;2.

Uijlenhoet, R., and A. Berne, 2008: Stochastic simulation experiment to assess radar rainfall retrieval uncertainties associated with attenuation and its correction. *Hydrol. Earth Syst. Sci.*, **12**, 587–601, https://doi.org/10.5194/hess-12-587-2008.

Ushio, T., T. Wu, and S. Yoshida, 2015: Review of recent progress in lightning and thunderstorm detection techniques in Asia. *Atmos. Res.*, **154**, 89–102, https://doi.org/10.1016/j.atmosres.2014.10.001.

van de Beek, C. Z., H. Leijnse, J. N. M. Stricker, R. Uijlenhoet, and H. W. J. Russchenberg, 2010: Performance of high-resolution X-band radar for rainfall measurement in The Netherlands. *Hydrol. Earth Syst. Sci.*, **14**, 205–221, https://doi.org/10.5194/hess-14-205-2010.

——, ——, P. Hazenberg, and R. Uijlenhoet, 2016: Close-range radar rainfall estimation and error analysis. *Atmos. Meas. Tech.*, **9**, 3837–3850, https://doi.org/10.5194/amt-9-3837-2016.

van der Werf, J. A., Z. Kapelan, and J. G. Langeveld, 2023: Predictive heuristic control: Inferring risks from heterogeneous nowcast accuracy. *Water Sci. Technol.*, **87**, 1009–1028, https://doi.org/10.2166/wst.2023.027.

Winterrath, T., and Coauthors, 2018: Radar climatology (RADKLIM) version 2017.002: Reprocessed gauge-adjusted radar data, one-hour precipitation sums (RW). DWD, accessed 2 December 2022, https://doi.org/10.5676/dwd/radklim_rw_v2017.002.

Woo, W.-c., and W.-K. Wong, 2017: Operational application of optical flow techniques to radar-based rainfall nowcasting. *Atmosphere*, **8**, 48, https://doi.org/10.3390/atmos8030048.

Yang, L., G. Ni, F. Tian, and D. Niyogi, 2021: Urbanization exacerbated rainfall over European suburbs under a warming climate. *Geophys. Res. Lett.*, **48**, e2021GL095987, https://doi.org/10.1029/2021GL095987.

Ye, J., Y. He, F. Pappenberger, H. L. Cloke, D. Y. Manful, and Z. Li, 2014: Evaluation of ECMWF medium-range ensemble forecasts of precipitation for river basins. *Quart. J. Roy. Meteor. Soc.*, **140**, 1615–1628, https://doi.org/10.1002/qj.2243.

Yoshikawa, E., T. Ushio, Z. Kawasaki, S. Yoshida, T. Morimoto, F. Mizutani, and M. Wada, 2012: MMSE beam forming on fast-scanning phased array weather radar. *IEEE Trans. Geosci. Remote Sens.*, **51**, 3077–3088, https://doi.org/10.1109/TGRS.2012.2211607.

Zhang, J., and Coauthors, 2011: National Mosaic and Multi-Sensor QPE (NMQ) system: Description, results, and future plans. *Bull. Amer. Meteor. Soc.*, **92**, 1321–1338, https://doi.org/10.1175/2011BAMS-D-11-00047.1.

Zhang, W., J. Yang, L. Yang, and D. Niyogi, 2022: Impacts of city shape on rainfall in inland and coastal environments. *Earth's Future*, **10**, e2022EF002654, https://doi.org/10.1029/2022EF002654.