Comparison of Three Radar-Based Precipitation Nowcasts for the Extreme July 2021 Flooding Event in Germany

Mohamed Saadi, Carina Furusho-Percot, Alexandre Belleflamme, Silke Trömel, Stefan Kollet, Ricardo Reinoso-Rondinel

Comparison of Three Radar-Based Precipitation Nowcasts for the Extreme July 2021 Flooding Event in Germany

MOHAMED SAADI,[a,b,c] CARINA FURUSHO-PERCOT,[a,b,d] ALEXANDRE BELLEFLAMME,[a,b] SILKE TRÖMEL,[e,f] STEFAN KOLLET,[a,b] AND RICARDO REINOSO-RONDINEL[e,g,h]

[a] *Institute of Bio- and Geosciences (IBG-3, Agrosphere), Forschungszentrum Jülich, Jülich, Germany*
[b] *Centre for High-Performance Scientific Computing in Terrestrial Systems, Geoverbund ABC/J, Jülich, Germany*
[c] *Institut de Mécanique des Fluides de Toulouse, Université de Toulouse, CNRS-INPT-UPS, Toulouse, France*
[d] *U.S. 1116 AGROCLIM, INRAE Centre de Recherche PACA, Avignon, France*
[e] *Institute for Geosciences, Department of Meteorology, Universität Bonn, Bonn, Germany*
[f] *Laboratory for Clouds and Precipitation Exploration, Geoverbund ABC/J, Bonn, Germany*
[g] *Faculty of Engineering Science, Department of Civil Engineering, Katholieke Universiteit Leuven, Leuven, Belgium*
[h] *Department of Meteorological Observations and Research, Royal Meteorological Institute of Belgium, Brussels, Belgium*

(Manuscript received 9 July 2022, in final form 14 April 2023, accepted 14 April 2023)

ABSTRACT: Quantitative precipitation nowcasts (QPN) can improve the accuracy of flood forecasts, especially for lead times up to 12 h, but their evaluation depends on a variety of factors, namely, the choice of the hydrological model and the benchmark. We tested three precipitation nowcasting techniques based on radar observations for the disastrous mid-July 2021 event in seven German catchments (140–1670 km^2). Two deterministic [advection-based and spectral prognosis (S-PROG)] and one probabilistic [Short-Term Ensemble Prediction System (STEPS)] QPN with a maximum lead time of 3 h were used as input to two hydrological models: a physically based, 3D-distributed model (ParFlowCLM) and a conceptual, lumped model (GR4H). We quantified the hydrological added value of QPN compared with hydrological persistence and zero-precipitation nowcasts as benchmarks. For the 14 July 2021 event, we obtained the following key results. 1) According to the quality of the forecasted hydrographs, exploiting QPN improved the lead times by up to 4 h (8 h) compared with adopting zero-precipitation nowcasts (hydrological persistence) as a benchmark. Using a skill-based approach, obtained improvements were up to 7–12 h depending on the benchmark. 2) The three QPN techniques obtained similar performances regardless of the applied hydrological model. 3) Using zero-precipitation nowcasts instead of hydrological persistence as benchmark reduced the added value of QPN. These results highlight the need for combining a skill-based approach with an analysis of the quality of forecasted hydrographs to rigorously estimate the added value of QPN.

KEYWORDS: Extreme events; Ensembles; Nowcasting; Hydrologic models; Model evaluation/performance; Flood events

1. Introduction

Precipitation extremes are intensifying due to human-driven climate change (Fowler et al. 2021). This means more severe and more frequent flooding events, which will lead to costlier damages to infrastructures and heavier human losses (Dottori et al. 2018; Dougherty and Rasmussen 2020). To mitigate these damages, operational and efficient flood warning systems are needed more than ever (Pappenberger et al. 2015a). These provide flood forecasts by relying on hydrological models fed with meteorological forecasts from numerical weather prediction (NWP) systems (Alfieri et al. 2012; Cloke and Pappenberger 2009). With ensemble modeling, data assimilation, and improved representation of physical processes enabled by the development of convection-permitting schemes (Speight et al. 2021; Clark et al. 2016), the skill of NWP has significantly increased during the last decades (Bauer et al. 2015), making it the best input for flood forecasting at the regional scale and for long horizons (>6 h; Lin et al. 2005). However, their use for short lead times (<6 h) in small-scale applications (enabled by using convection-permitting NWP) is hindered by the time needed for their spinup and their too coarse spatial resolution for hydrological needs.

Statistical extrapolation of the up-to-date weather radar observations (or nowcasting) can fill this gap by providing quantitative precipitation nowcasts (QPN) at high spatial and temporal resolutions (up to 1 km^2 and 5 min, respectively; see, e.g., Reinoso-Rondinel et al. 2022), which can outperform the NWP for short lead times (Berenguer et al. 2012). This level of detail is particularly useful to forecast flash floods from convective precipitation events especially in urban areas and rapidly responding catchments (Berenguer et al. 2005; Foresti et al. 2016; Ochoa-Rodriguez et al. 2015). Most QPN are generated by 1) estimating the motion field from remote sensing products, such as radar or satellite images, and 2) applying this motion field to displace the most recently observed precipitation field (Ayzel et al. 2019). These two steps form the core of most deterministic nowcasting techniques such as TREC (Tracking Radar Echo with Correlations; Rinehart and Garvey 1978), MAPLE (McGill Algorithm for Precipitation nowcasting by Lagrangian Extrapolation; Germann and Zawadzki 2002), S-PROG (Spectral Prognosis; Seed 2003),

Corresponding author: Mohamed Saadi, mohamed.saadi@toulouse-inp.fr

DOI: 10.1175/JHM-D-22-0121.1

TABLE 1. Summary of applications using deterministic and probabilistic precipitation nowcasting methods with and without hydrological evaluation.

Reference	QPN method	Location	Hydrological model
Berenguer et al. (2011)	SBMcast (probabilistic)	Barcelona, Spain	—
Atencia and Zawadzki (2014, 2015)	Two probabilistic nowcasting methods	United States	—
Mejsnar et al. (2018)	COTREC (deterministic; Li et al. 1995)	Czech Republic	—
Imhoff et al. (2020)	Four deterministic and probabilistic methods, namely, Sparse, DenseRotation, S-PROG, and STEPS, implemented within Rainymotion (Ayzel et al. 2019) and pySTEPS (Pulkkinen et al. 2019)	12 catchments in the Netherlands	—
Reinoso-Rondinel et al. (2022)	S-PROG (deterministic; Seed 2003)	Germany	—
Berenguer et al. (2005)	S-PROG (deterministic; Seed 2003), Lagrangian advection and Eulerian persistence (deterministic)	Barcelona, Spain	DiCHiTop (distributed)
Šálek et al. (2006)	COTREC (deterministic; Li et al. 1995)	Czech Republic	HYDROG (distributed)
Vivoni et al. (2006)	STNM algorithm (deterministic; Wolfson et al. 1999)	Midwestern United States	tRIBS (physically based, distributed)
Xuan et al. (2014)	STEPS (probabilistic; Bowler et al. 2006)	1 catchment in the United Kingdom	PDM (lumped)
Heuvelink et al. (2020)	Lagrangian persistence/COTREC (deterministic; Li et al. 1995) and SBMcast (probabilistic; Berenguer et al. 2011)	3 catchments in the Netherlands	WALRUS (lumped)
Lovat et al. (2022)	AROME-NWC (deterministic, NWP-based; Auger et al. 2015) and PIAF (combination of radar nowcasts and NWP; Moisselin et al. 2019)	19 catchments in southeastern France	ISBA-TOP (distributed)
Imhoff et al. (2022)	Four deterministic and probabilistic methods, namely, Sparse, DenseRotation, S-PROG, and STEPS, implemented within Rainymotion (Ayzel et al. 2019) and pySTEPS (Pulkkinen et al. 2019)	12 catchments in the Netherlands	SOBEK (semidistributed) and WALRUS (lumped)

and SWIRLS (Short-range Warning of Intense Rainstorms in Localized Systems; Woo and Wong 2017). To account for uncertainties in the motion field as well as in the evolution of the precipitation cells, many techniques adopt a stochastic approach by adding random perturbations based on corresponding spatiotemporal properties to produce an equally likely ensemble of QPN. Examples of these probabilistic techniques include STEPS (Short-Term Ensemble Prediction System; Bowler et al. 2006), STEPS-BE (STEPS system for Belgium; Foresti et al. 2016), SBMcast (String of Beads Model; Berenguer et al. 2011), and ENS (Sokol et al. 2017).

To evaluate the skill of QPN, several studies quantified the ability of nowcasting techniques to provide accurate short-term predictions of observed precipitation (see Table 1; Berenguer et al. 2011; Atencia and Zawadzki 2014, 2015; Mejsnar et al. 2018; Imhoff et al. 2020; Reinoso-Rondinel et al. 2022). Their approach compares the predicted precipitation from QPN for a given lead time with quantitative precipitation estimates (QPE) obtained from radar observations. These studies focused on improving the nowcasting methods to account for uncertainties in the prediction of precipitation fields and highlighted the limits of the applied methods in the case of warm-season and convective events (Mejsnar et al. 2018). To characterize and enhance the hydrological predictability of associated flash floods, Imhoff et al. (2020) analyzed the effect of catchment properties and event characteristics (such as the size and location) on the nowcasting skill. Toward a nationwide nowcasting system, Reinoso-Rondinel et al. (2022) improved the S-PROG technique by introducing spatially localized parameters for the inherent autoregressive model and evaluated the skill with respect to radar-based QPE for 10 observed rain events in Germany.

An alternative evaluation framework exploits (ensemble) QPN to serve as input to a hydrological model (see Table 1; Šálek et al. 2006; Berenguer et al. 2005; Vivoni et al. 2006; Xuan et al. 2014; Heuvelink et al. 2020; Lovat et al. 2022; Imhoff et al. 2022). The resulting simulated discharge time series are then compared to a reference discharge time series, which can be either the observed discharge, if available, or the simulated discharge by the hydrological model with QPE (i.e., observed precipitation) as input precipitation. This framework is

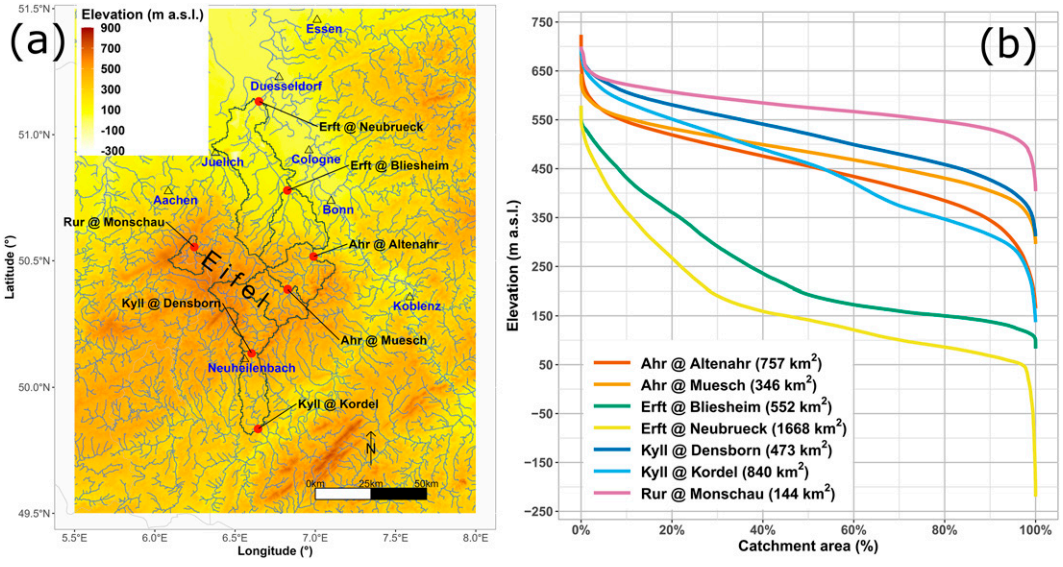

FIG. 1. (a) Location, topography, and hydrographic network of study catchments, where contours indicate the catchment polygons, and (b) hypsometric curves of the catchment set. Negative elevations are due to the existence of open-pit mines in the region.

more relevant for flood forecasting applications since it quantifies the added value of QPN, with respect to, e.g., zero-precipitation nowcasts, in improving the lead time of hydrological forecasts. All studies found that radar-based QPN enhanced the forecasting skill achieved by the hydrological model, especially when blended with NWP forecasts (Lovat et al. 2022). Moreover, the forecasting skill depended on the physical properties of the catchment (such as size and topography), the type of the event (convective versus stratiform), and the season (rain versus snow).

Despite these findings, previous studies did not focus on the evaluation methodology of the nowcasting techniques. Namely, all of the listed studies (except Imhoff et al. 2022) adopted a single modeling approach (i.e., either lumped or distributed, conceptual or physically based), which did not allow for considering how the choice of the hydrological model structure impacted the evaluation of the nowcasting techniques (such as done by Poméon et al. 2020). In addition, the impact of the benchmark nowcasting model (such as zero-precipitation nowcasts or hydrological persistence) on the forecasting skill remains poorly investigated while it can have significant impact on the estimated added value of QPN (Pappenberger et al. 2015b).

To tackle these gaps, we evaluated one probabilistic (STEPS) and two deterministic nowcasting techniques (advection-based and S-PROG) by measuring their ability in forecasting simulated hydrographs with QPE. Our study focuses on the disastrous mid-July 2021 events in seven catchments located in western Germany. These events caused more than 220 deaths and cost up to €32.05 billion in total losses in Germany alone, making them one of the most severe natural disasters caused by heavy rain and flooding in Germany (Mohr et al. 2023). We adopted a novel multimodeling approach by evaluating QPN as inputs to a conceptual, lumped model (GR4H) and to a physically based, 3D-distributed model (ParFlowCLM). Thus, the aim of this study is to investigate whether a more detailed representation of hydrological processes leads to a better discrimination of QPN compared to a simpler, lumped one. Moreover, we checked whether different choices of skill evaluation and benchmarks impact the estimation of the added value of the nowcasting techniques.

This paper is organized as follows. Section 2 presents the case study, the catchment set, and the QPE product used to produce the QPN. Section 3 introduces the tested nowcasting techniques, the hydrological models, and the evaluation framework, while sections 4 and 5 comment and discuss the results. Finally, section 6 concludes our study.

2. Catchment set and QPE product

In July 2021, sustained stratiform rain connected to a cutoff low pressure system (Junghänel et al. 2021) led to record-breaking precipitation amounts and disastrous floods (Kreienkamp et al. 2021; Mohr et al. 2023), especially over relatively high altitudes at the Eifel range on the left bank of the Rhine River and the Bergisches Land on the right bank (Figs. 1 and 2). On 14 July 2021, observed total precipitation sums exceeded 160 mm at some rain gauges (Fig. 2c), which is equivalent to 2–3 months of accumulated precipitation based on the annual averages (i.e., by dividing 160 mm by the annual averages listed in Table 2). Since rain gauges do not provide a detailed description of the spatial variability of precipitation, measurements from four polarimetric C-band radars (located at Essen, Flechtdorf, Neuheilenbach, and Offenthal; Fig. 2a), operated by the German Weather Service (DWD), were exploited to derive a gridded QPE product for the 14 July 2021 (Fig. 2b) with 1-km horizontal resolution and 5-min temporal resolution. This hybrid product

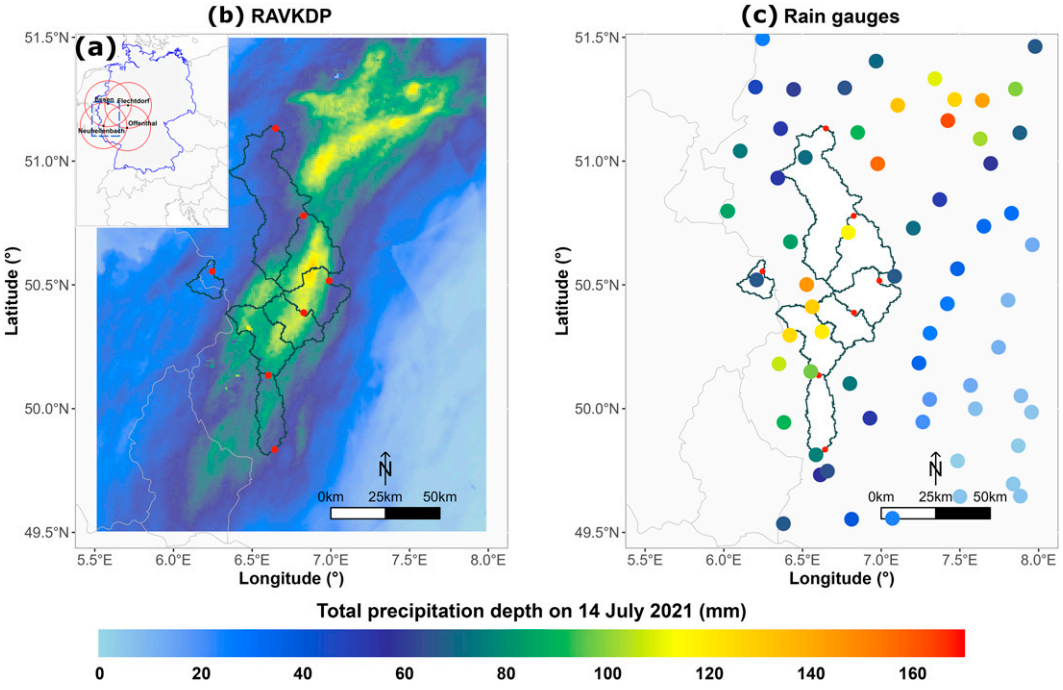

FIG. 2. (a) Location of the four C-band radars (Essen, Flechtdorf, Neuheilenbach, and Offenthal) operated by the German Weather Service (DWD) and used to derive the QPE product RAVKDP. (b) Total precipitation amounts on 14 Jul 2021 (from 0000 UTC 14 Jul 2021 to 0000 UTC 15 Jul 2021) estimated from the radar-based QPE RAVKDP and (c) from 63 rain gauges. For the 63 rain gauges, the ratio of total precipitation from RAVKDP to that from the rain gauge varied between 0.27 and 3.00, with a median value of 0.72. The light gray contours indicate the borders with the neighboring countries (The Netherlands, Belgium, Luxembourg, and France).

combines precipitation estimates derived from specific attenuation at vertical polarization A_V, $R(A_V)$, with retrievals of specific differential phase K_{DP} for horizontal reflectivity rates Z_h higher than 40 dBZ, $R(K_{DP})$. This product clearly outperformed retrievals based on horizontal reflectivity alone (Chen et al. 2021; Saadi et al. 2023). The hybrid QPE product, referred to as RAVKDP in the following, served as input for QPN algorithms.

To hydrologically evaluate the benefits of QPN, we selected a set of seven catchments located in western Germany that drain parts of the Eifel mountain range (Fig. 1a), characterized by a rolling plateau at elevations up to 750 m MSL (Fig. 1b). These catchments have areas ranging between 140 and 1670 km^2 (Table 2). Three of the seven catchments are located in the federal state of North Rhine–Westphalia and are drained by the Erft and the Rur Rivers. The remaining four catchments are located in the federal state of Rhineland–Palatinate and are drained by the Ahr and the Kyll Rivers. The region is characterized by a temperate climate under maritime influence, which is reflected by the range of the average annual precipitation amounts (710–1070 mm yr^{-1}) and the values of the aridity index as defined by the United Nations Environment Programme (1.13–1.92; UNEP 1992). The land cover of the catchments is mainly occupied by forest and agricultural areas according to the CORINE Land Cover database of 2018 (Langanke et al. 2016). Soils are dominated by sand (34%–41%) and silt contents (29%–38%; Panagos 2006).

To estimate total precipitation amounts on 14 July 2021 at the catchment scale, we applied the Thiessen polygon method on measurements from rain gauges and on RAVKDP. Estimated precipitation amounts from rain gauges varied between 66 and 121 mm across our catchment set (Table 2), reflecting the severity of the event and its variability from one catchment to another. Based on RAVKDP, obtained estimates totaled only 34–90 mm (Table 2), indicating an underestimation with respect to estimated amounts from rain gauges. This underestimation is partly attributed to collision–coalescence processes that took place close to the surface, i.e., below the heights monitored by the radars (Saadi et al. 2023; Chen et al. 2022).

3. Evaluation of the added value of QPN

a. *Tested nowcasting techniques*

Based on the QPE product RAVKDP, we computed 3-h-long QPN with 1-km spatial and 5-min temporal resolution. In this study, three nowcasting strategies following Reinoso-Rondinel et al. (2022) have been applied:

1) The deterministic method based on Lagrangian persistence (advection-based) assumes a constant precipitation field (i.e., with no growth or decay) advected using a static motion field. First, the motion field was estimated from the RAVKDP product using the optimal-flow method

TABLE 2. Summary of catchment characteristics. Catchment-average, total precipitation amount on 14 Jul 2021 (from 0000 UTC 14 Jul 2021 to 0000 UTC 15 Jul 2021) are extracted from RAVKDP, the radar-based QPE product, and from rain gauges using Thiessen polygons. In the far-right column, the total number of rain gauges used for the 14 Jul 2021 for each catchment is provided in parentheses. For the columns labeled artificial, agricultural, and forest, these metrics were computed based on the CORINE Land Cover classification of the Copernicus Land Monitoring Service (Langanke et al. 2016); they correspond to the proportion of the catchment that is occupied by the classes belonging to 1) "artificial surfaces" for artificial, 2) "agricultural areas" for agricultural, and 3) "forest and seminatural areas" for forest. See https://land.copernicus.eu/user-corner/technical-library/corine-land-cover-nomenclature-guidelines/html (last accessed 14 Apr 2023).

River[a]	Area (km^2)	Average precipitation ($mm\ yr^{-1}$)	Aridity index[b] (—)	Average discharge ($mm\ yr^{-1}$)	Artificial (%)	Agricultural (%)	Forest (%)	Total precipitation amount on 14 Jul 2021 (mm) from RAVKDP/ rain gauges (No. of rain gauges)
Erft at Neubrueck	1668	740 (2006–21)	1.16	180 (2000–20)	17.7	64.3	17.8	66/99 (13)
Kyll at Kordel	840	830 (2006–21)	1.41	370 (1967–2021)	5.4	51.9	42.7	80/103 (10)
Ahr at Altenahr	757	750 (2006–21)	1.27	280 (1945–2021)	3.5	39.5	57	89/108 (7)
Erft at Bliesheim	552	710 (2006–21)	1.13	130 (2000–20)	12.6	59.1	28.2	88/108 (7)
Kyll at Densborn	473	890 (2006–21)	1.54	450 (1972–2021)	4	47.7	48.2	87/115 (7)
Ahr at Muesch	346	790 (2006–21)	1.34	280 (1972–2021)	4	52.9	43.1	90/121 (6)
Rur at Monschau	144	1070 (2006–21)	1.92	760 (2000–21)	6.1	25.4	62.9	34/66 (1)

[a] All catchments contain at least one reservoir (lake or dam) according to the database at https://dewiki.de/Lexikon/Liste_von_Talsperren_in_Deutschland (in German, last accessed 14 Apr 2023).
[b] The aridity index was computed as the ratio of average annual precipitation to average annual atmospheric evaporative demand, which we expressed as the average annual potential evapotranspiration (UNEP 1992). Potential evapotranspiration was computed using a temperature-based formula (Oudin et al. 2005).

DARTS (Ruzanski et al. 2011). Then, the latest observed precipitation field is advected along the estimated motion trajectories for the next 3 h using a semi-Lagrangian backward scheme (Germann and Zawadzki 2002).

2) The deterministic method S-PROG (Seed 2003) assumes that the spatial scale of precipitation features is on par with its lifetime and thus its predictability. This means that larger precipitation features tend to last longer and can be predicted with larger lead times. First, the precipitation field is decomposed into a multiplicative cascade of spatial scales. Second, an autoregressive model (AR) is used to model and forecast the temporal evolution and to advect each cascade level. Finally, the nowcasted field is computed as the aggregation of the advected cascade levels. This leads to a smoothing of the precipitation field as the small-scale, high-frequency features tend to vanish with time according to the AR model. Compared with Seed (2003), we kept the order of the AR model at 1 instead of 2, and we fixed the number of levels of the multiplicative cascade at 6, resulting in the following spatial scales of 900, 56, 20, 7, 3, and 1 km. Moreover, we used the precipitation field instead of the reflectivity field. Since the precipitation field does not follow a Gaussian distribution, the above processes were applied to the log-transformed values of precipitation, which we assumed to have a near-Gaussian distribution. After extrapolation, an inverse transformation was applied to the nowcasted precipitation field. These choices follow the study by Reinoso-Rondinel et al. (2022).
3) The probabilistic method STEPS (Bowler et al. 2006) builds on S-PROG by adding stochastic perturbations to account for the uncertainties in the estimated motion field and the evolution of the precipitation cells. More precisely, each cascade level is perturbed by Gaussian white noise that is correlated with the spatial properties of the last observed precipitation field (Seed et al. 2013), which leads to an ensemble of QPN. In our study, we considered an ensemble of 20 members.

For each QPN method, we generated 3-h long time series of nowcasted precipitation every 5 min (i.e., at 0000, 0105, 0110 UTC 14 July 2021, etc.) with a temporal resolution of 5 min. Since we chose to feed these QPN to hourly hydrological models, we kept only QPN that were issued at round hours (i.e., at 0100 UTC, at 0200 UTC, …, and at 1800 UTC 14 July 2021) and discarded the remaining ones. In addition, we aggregated the 5-min QPN time series to obtain hourly accumulations of precipitations.

b. *Hydrological models*

We analyzed the impact of the hydrological model on the evaluation of QPN by selecting two contrasting modeling approaches, for which the implementation is described in Table 3. As a physically based, distributed model, we used ParFlow with its internal land surface module CLM (Common Land Model), hereafter ParFlowCLM (Kollet and Maxwell 2006; Kuffour et al. 2020; Maxwell 2013). CLM estimates the actual evapotranspiration, infiltration, and net precipitation (i.e., the part that gives rise to runoff) by resolving the energy budget at the land surface and the water exchange at the interface between the atmosphere, the land, and the soil. ParFlow solves the 3D Richards' equation for variably saturated subsurface and groundwater flow and the kinematic wave equation for the overland flow routing. These two equations are coupled at the land surface by

TABLE 3. Details of ParFlowCLM and GR4H implementation: resolution, parameter estimation, and sources of hydroclimatic data needed for each model. ASTER = Advanced Spaceborne Thermal Emission and Reflection Radiometer; DEM = digital elevation model; MERIT = Multi-Error-Removed Improved Terrain; USDA = United States Department of Agriculture; IHME1500 = International Hydrogeological Map of Europe at the scale of 1:1 500 000; CORINE = Coordination of Information on the Environment; CLMS = Copernicus Land Monitoring Service; IGBP = International Geosphere-Biosphere Programme; RADOLAN = Radar-Online-Aneichung; DWD = Deutscher Wetterdienst (German Weather Service).

Model	Spatial and temporal resolution	Parameter estimation	Hydroclimatic data
ParFlowCLM	~611-m horizontal resolution with a geometrically varying vertical resolution, hourly	Topography: ASTER DEM (Abrams et al. 2020; https://lpdaac.usgs.gov/products/astgtmv003) combined with MERIT Hydro (Yamazaki et al. 2019). Soil and geology: SoilGrids250m (Hengl et al. 2017), reclassified into 12 USDA texture types, and IHME1500 (Duscher et al. 2015) for the typology below the depth to bedrock; ROSETTA model (Schaap et al. 2001) to obtain hydraulic parameters (hydraulic conductivity, residual and saturated water content, and van Genuchten parameters) depending on soil types. Land cover: CORINE Land Cover database of the CLMS for the year 2018 (https://land.copernicus.eu/pan-european/corine-land-cover/clc2018; Langanke et al. 2016), reclassified into 18 IGBP categories. A uniform Manning's coefficient at 0.2 $s \cdot m^{-1/3}$ (Schalge et al. 2019) was used for the whole domain. **Only one parameter set for each catchment** (Belleflamme et al. 2023)	Precipitation: RADOLAN of the DWD (Winterrath et al. 2018), which is a Germany-wide, radar-based near-real time precipitation product available at 1-km resolution and hourly time steps obtained using relationships between horizontal reflectivity and precipitation rates, and then adjusted to rain gauges (i.e., RADOLAN-RW, https://opendata.dwd.de/, last accessed 14 Apr 2023). RAVKDP was used for precipitation on 14 Jul 2021 (Chen et al. 2021). 2-m air temperature, surface pressure, downward solar and thermal radiation, specific humidity, and eastward and northward components of the 10-m wind: ERA5-Land dataset (Muñoz-Sabater et al. 2021), available at 9-km resolution.
GR4H	Lumped, hourly	Four catchment-scale parameters representing the maximum retention capacity of the soil, the exchange between surface water and groundwater, the surface flow dynamics and the baseflow dynamics. These parameters are calibrated on historical discharge measurements using a gradient-descent based algorithm (Coron et al. 2017; Edijatno et al. 1999). **12 optimal sets of 4 parameters for each catchment** (Saadi et al. 2023)	Catchment-averaged precipitation: RADOLAN of the DWD (Winterrath et al. 2018), available at 1-km and hourly resolutions, estimated based on horizontal reflectivity and adjusted to rain gauges (i.e., RADOLAN-RW, https://opendata.dwd.de/, last accessed 14 Apr 2023). RAVKDP was used for precipitation on 14 Jul 2021 (Chen et al. 2021). Thiessen polygons were used to estimate the catchment-average precipitation at each hour. Catchment-averaged potential evapotranspiration: obtained from catchment-average, 2-m air temperature using a temperature-based formula (Oudin et al. 2005). Discharge: used for model calibration, available at daily resolution (https://www.elwasweb.nrw.de; https://wasserportal.rlp-umwelt.de, last accessed 20 Sep 2021).

estimating the boundary fluxes for the kinematic wave model from Richards' equation, and vice versa. Thanks to this coupling, the model represents a variety of runoff processes (Hortonian versus Dunne runoff) as well as the reinfiltration and exfiltration processes along the hydraulic pathway. We implemented ParFlowCLM at a resolution of 611 m with 15 vertical layers down to 60 m below the surface (Belleflamme et al. 2023). It was forced with gridded weather inputs over a spinup period starting from 2007, with only one parameter set for each catchment based on landscape properties, as detailed in Table 3.

As a conceptual, lumped model, we chose GR4H (Ficchì et al. 2019). This model estimates net precipitation and actual evapotranspiration using a soil-moisture accounting reservoir. The net precipitation gives rise to runoff through two routing

branches. The quick flow branch transfers 10% of net precipitation via a unit hydrograph, while the slow flow branch transfers the remaining 90% via a unit hydrograph and a nonlinear reservoir. Over both branches, an exchange between surface flow and groundwater is enabled. GR4H uses catchment-average weather inputs to simulate the discharge at the catchment outlet. Model parameters were estimated using historical discharge measurements over the period 2007–21, with a spinup over the year 2006 to limit the effect of model initialization on calibration. We tested several choices of calibration combining the subperiod of calibration, the cost function, and the weights attributed to discharge measurements to emphasize high values, yielding 12 optimal parameter sets for each catchment (Saadi et al. 2023).

c. Comparison and evaluation framework of QPN

Following Berenguer et al. (2005), we evaluated the skill of QPN on two levels. On the first level, we analyzed how QPN succeeded in matching QPE for each lead time, first at the gridcell scale, then at the catchment scale by averaging the precipitation fields using the catchment polygon. At the gridcell scale, we adopted the mean absolute error (MAE; mm h^{-1}) and the root-mean-square error (RMSE; mm h^{-1}) as evaluation metrics, expressed as

$$\mathrm{MAE}(L) = \frac{1}{N_t N_c} \sum_{i=1}^{N_c} \sum_{t=1}^{N_t} |\widehat{P_{t+L|t}}(i) - P_{t+L}(i)|, \tag{1}$$

$$\mathrm{RMSE}(L) = \sqrt{\frac{1}{N_t N_c} \sum_{i=1}^{N_c} \sum_{t=1}^{N_t} [\widehat{P_{t+L|t}}(i) - P_{t+L}(i)]^2}, \tag{2}$$

where N_t is the number of initialization time steps (i.e., hours or 5-min time steps between 0100 and 1800 UTC 14 July 2021), N_c is the number of grid cells, L is the lead time, $\widehat{P_{t+L|t}}(i)$ is the QPN intensity for the time step $t + L$ issued at time step t for the grid cell i, and $P_{t+L}(i)$ is the QPE intensity at time step $t + L$ for the grid cell i. Both RMSE (mm h^{-1}) and MAE (mm h^{-1}) vary between 0 (perfect match) and $+\infty$. Note that for the computation of the spatial average of MAE and RMSE, we excluded the grid cells for which the total precipitation amount on 14 July 2021 (according to RAVKDP) was equal to zero. By this choice, we aimed at limiting the number of grid cells for which the errors are equal or very close to zero, the inclusion of which would artificially decrease the two accuracy measures. At the catchment scale, we first averaged the precipitation time series using the catchment polygon, then we computed the MAE between the resulting catchment-scale QPE time series and catchment-scale QPN time series.

For the probabilistic STEPS method, since each member served as input to the hydrological models, both MAE and RMSE scores were estimated for each of the 20 members, then for a deterministic nowcast STEPS-m taken as the ensemble mean at each grid cell and at each time step. Following the approach by Foresti et al. (2016), we also aimed at analyzing the spread of the ensemble with respect to the errors of the deterministic forecast (i.e., S-PROG or STEPS-m) in order to qualify whether the ensemble was underdispersive (i.e., underestimating the uncertainty in the evolution of the precipitation field) or overdispersive (i.e., overestimating the uncertainty in the evolution of the precipitation field; see Foresti et al. 2016). To this aim, we estimated the spread of the ensemble at the gridcell scale at each lead time using the following equation (Foresti et al. 2015):

$$\mathrm{spread}(L) = \sqrt{\frac{1}{N_t N_c} \sum_{i=1}^{N_c} \sum_{t=1}^{N_t} \frac{1}{M-1} \sum_{m=1}^{M} [\widehat{P_{t+L|t}}(i,m) - \overline{\widehat{P_{t+L|t}}}(i)]^2}, \tag{3}$$

where $M = 20$ is the total number of members, $\widehat{P_{t+L|t}}(i, m)$ is the QPN intensity for the time step $t + L$ issued at time step t for the grid cell i by the STEPS member m, and $\overline{\widehat{P_{t+L|t}}}(i)$ is the intensity of the STEPS ensemble mean nowcast at time step $t + L$ issued at time step t for the grid cell i. Ideally, the spread should be of the same order of variability of the QPE around the ensemble mean, measured in our case by the RMSE of the ensemble mean nowcast STEPS-m. When the spread is higher than this RMSE, the ensemble is overdispersive, otherwise the ensemble is underdispersive (Foresti et al. 2016).

On the second level, QPN were used to extend the precipitation input to the hydrological models. First, both models were run prior to 0100 UTC 14 July 2021 with the version of RADOLAN that was adjusted to rain gauges (i.e., RADOLAN-RW; Winterrath et al. 2018) as input precipitation (see Table 3). These runs started from January 2021 for ParFlowCLM and from 2007 for GR4H. Starting from 0100 UTC 14 July 2021, the QPE product RAVKDP was used instead of RADOLAN for our study region. At each initialization hour (e.g., 0100 UTC 14 July 2021), the QPE was replaced by the 3-h QPN (e.g., at 0200, 0300, and 0400 UTC) followed by zero precipitation (e.g., from 0500 UTC onward). Then, the resulting forecasted hydrographs were compared to the simulated hydrograph with the QPE product RAVKDP as input for 14 July 2021 and RADOLAN as input for the remaining days (i.e., the hindcasted hydrograph).

In a first step, we evaluated the quality of the hydrological forecasts obtained by the use of QPN and the benchmarks using the Nash–Sutcliffe efficiency score (NSE; Nash and Sutcliffe 1970), computed as

$$\mathrm{NSE}(L) = 1 - \frac{\sum_{t=1}^{N_t} (Q_{t+L} - \widehat{Q_{t+L|t}})^2}{\sum_{t=1}^{N_t} (Q_{t+L} - \overline{Q_{t+L}})^2}, \tag{4}$$

where $\widehat{Q_{t+L|t}}$ is the forecasted discharge values at the time step $t + L$ initialized at the time step t, Q_{t+L} the hindcasted discharge values (i.e., simulated hydrographs using QPE) at time step $t + L$ and $\overline{Q_{t+L}}$ their average. NSE varies between $-\infty$ and 1, the latter being the ideal value. As the lead time increases, NSE is expected to decrease. Figure 8 of Berenguer et al. (2005) and Fig. 2 of Heuvelink et al. (2020) illustrate the application of this evaluation method.

The added value of each QPN can be estimated by comparison with a benchmark/reference option (Pappenberger et al. 2015b). To measure this added value, we computed the gain in lead time defined as (Berenguer et al. 2005)

$$\text{gain in lead time} = L_{\text{QPN}}(\text{NSE}_{\text{th}}) - L_{\text{Ref}}(\text{NSE}_{\text{th}}), \quad (5)$$

where $L_{\text{QPN}}(\text{NSE}_{\text{th}})$ is the lead time at which the obtained NSE with the QPN as input to the hydrological model equals NSE_{th} for the first time, and $L_{\text{Ref}}(\text{NSE}_{\text{th}})$ is the lead time at which the obtained NSE with the benchmark Ref equals NSE_{th} for the first time. Following Heuvelink et al. (2020), we adopted a threshold of $\text{NSE}_{\text{th}} = 0.9$. To analyze the impact of this threshold, we computed the gain for an additional threshold of $\text{NSE}_{\text{th}} = 0.5$.

In a second step, we applied the average of the continuous ranked probability score (CRPS) (Hersbach 2000), expressed for each lead time L as

$$\text{CRPS}(L) = \frac{1}{N_t}\sum_{t=1}^{N_t}\int_0^{+\infty}[F_{\widehat{Q_{t+L|t}}}(x) - \mathbb{1}\{Q_{t+L} \le x\}]^2 dx \quad (6)$$

where $F_{\widehat{Q_{t+L|t}}}$ is the cumulative distribution function of the forecasted discharge values $\widehat{Q_{t+L|t}}$ initialized at the time step t for the time step $t + L$, and Q_{t+L} is the value at time step $t + L$ of the simulated hydrograph using QPE (i.e., the hindcasted hydrograph). The term $\mathbb{1}\{y \le x\}$ is the Heaviside step function that equals 1 if $y \le x$ and 0 otherwise. CRPS was chosen because it helps undistinguishably evaluate both the probabilistic and the deterministic nowcasting methods. For a deterministic forecast, it is equivalent to MAE.

To evaluate the added value of the QPN methods with respect to a benchmark, a skill score based on the CRPS was computed as follows (Chen et al. 2017):

$$\text{Skill}_{\text{CRPS, Ref}} = \frac{\text{CRPS(Ref)}}{\text{CRPS(QPN)} + \text{CRPS(Ref)}}, \quad (7)$$

which is a bounded (between 0 and 1) and a scale-independent metric. A skill higher than 0.5 indicates that the forecasts obtained with QPN are better than the ones obtained with the benchmark [i.e., CRPS (QPN) < CRPS (Ref)], and vice versa.

To investigate the effect of the benchmark choice on the evaluation of QPN, we evaluated the skill of QPN with respect to 1) a hydrological persistence model (Berthet et al. 2009) that forecasts the future discharge to be constant and equal to the hindcasted discharge at the hour of initialization ($\text{Skill}_{\text{CRPS},Q}$), and 2) a forecasted hydrograph using zero precipitation nowcasts (ZNC; Heuvelink et al. 2020; Berenguer et al. 2005) as QPN ($\text{Skill}_{\text{CRPS,ZNC}}$). The latter is costlier than the former because it involves running the hydrological model for the ZNC. Finally, for the CRPS-based skill in Eq. (7), we retrieved the lead time up to which QPN is considered to be "useful" with respect to the benchmark using two skill thresholds: the theoretical one at 0.5, and a more demanding one [$2/3 \approx 0.67$, which is equivalent to $\text{CRPS(QPN)} < \text{CRPS(Ref)}/2$].

To qualitatively analyze the effect of catchment properties on the added value of QPN with respect to the benchmark, the gains in lead time based on NSE [Eq. (4)] and based on the skills [Eq. (7)] were ranked first with respect to catchment area, and second with respect to the Gravelius index of the catchment, defined as (Bendjoudi and Hubert 2002)

$$K = \frac{P}{2\sqrt{\pi A}}, \quad (8)$$

where P is the perimeter of the catchment polygon (in km) and A the catchment area (in km^2). Catchments with lower K tend to have compact or circular shapes, which would generally result in flashier hydrological responses for a given precipitation event covering the whole catchment.

4. Results

a. *Evaluation of QPN with respect to QPE*

Aggregating the QPN time series to the hourly time step reduced the differences between the three methods and modified their ranking, as can be seen in Fig. 3. At 5-min resolution (Figs. 3a,b), QPN obtained by advection and STEPS had similar MAE and RMSE scores over the domain during the lead times. As the lead time increased, S-PROG clearly outperformed the other two QPN methods. At 1-h resolution (Figs. 3c,d), the three methods obtained lower MAE and RMSE values compared with the 5-min resolution, S-PROG preserved its ranking with respect to advection and STEPS, whereas advection slightly outperformed the STEPS ensemble, suggesting that changing the accumulation window can modify the ranking of the QPN methods. For both time resolutions, the STEPS ensemble mean (STEPS-m) outperformed all the other members for all time steps, suggesting that the stochastic perturbations of the S-PROG method (materialized by STEPS members) got penalized for this event. Finally, there were very small differences between the different STEPS members in terms of MAE and RMSE (hardly visible in Fig. 3), which is somewhat expected from averaging the errors in space (over the domain) and time (across the initialization time steps) for members generated randomly and independently for each initialization. The small spread of STEPS members compared with the RMSE of the ensemble mean STEPS-m or the RMSE of the deterministic S-PROG method suggests that the ensemble nowcasts were underdispersive (Foresti et al. 2016), i.e., that they underestimated the uncertainty in the nowcasted precipitation field for this event.

At the hourly time step, the three QPN methods showed comparable performances in reproducing the observed precipitation at the gridcell scale, with a slightly higher performance for S-PROG, as shown in Fig. 4. The spatial pattern of MAE followed that of the precipitation sums for the event (Fig. 2b), with a slight shift for the part of the event cell located over the catchment set to the southwest. Unsurprisingly, the errors were minimal for the shortest lead time (i.e., 1 h) and increased with increasing lead time. For the 1-h lead time, domain-average MAE values were around 0.36–0.37 mm h^{-1} for the advection and S-PROG methods, whereas they reached 0.39 mm h^{-1} on average for STEPS members, indicating a slightly deteriorated accuracy for the probabilistic QPN. For the 3-h lead time, these errors more than doubled and reached 0.77 mm h^{-1} for advection, 0.7 mm h^{-1} for S-PROG, and 0.79 mm h^{-1} on average for STEPS, indicating

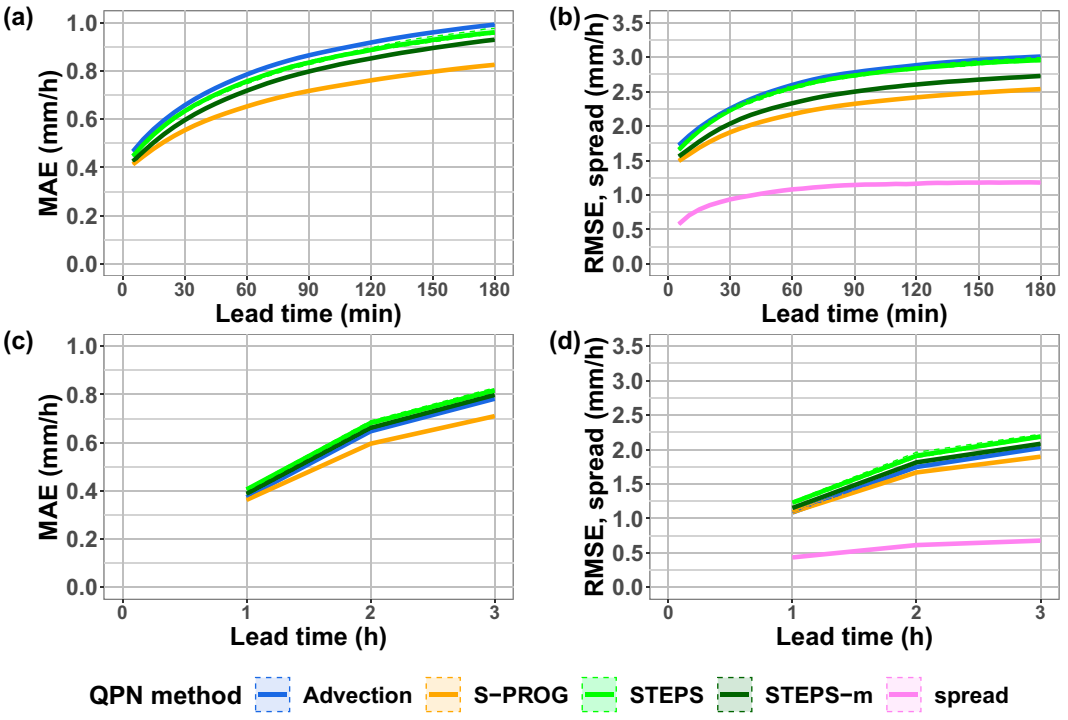

FIG. 3. Evolution of the spatial average of mean absolute errors (MAE) and root-mean-square errors (RMSE) of QPN with respect to precipitation rates from the QPE product RAVKDP for different lead times at (a),(b) 5-min resolution updated each 5 min and (c),(d) 1-h resolution updated each hour. In (b) and (d), "spread" indicates the spread of the STEPS ensemble [Eq. (3)]. STEPS-m indicates the ensemble mean, i.e., the nowcast made by taking the average of the nowcasted depths from the 20 STEPS members at each grid cell and each time step. The spatial average was computed on all domain grid cells except the ones with zero-precipitation amounts on 14 Jul 2021 according to the QPE product RAVKDP. The ensemble of MAE and RMSE errors for the STEPS method is hardly visible due to very small differences between the members.

a better performance of the S-PROG method over the domain.

At the catchment scale, the advection method obtained slightly better results than S-PROG and STEPS, as shown in Fig. 5. The change in the ranking of the QPN methods with respect to Fig. 3 may be explained by the catchment-scale aggregation of the precipitation fields prior to the computation of the errors, or the fact that the catchments do not cover the whole domain on which MAE values of Fig. 3 were computed (see Figs. 2b and 4). Moreover, advection does not change QPE intensities across the lead times, whereas S-PROG filters the observed QPE field, leading to smoother QPN field and to an underestimation of precipitation for persistent and heavy events. This results in advection mimicking better the QPE than S-PROG, especially over our catchment set where the July 2021 event was persistent and heavy. Overall, QPN had better success in reproducing the average precipitation for the catchments drained by the Rur at Monschau and the Erft than for the catchments drained by the Kyll and the Ahr. The variability in the ensemble errors of STEPS increased with increasing lead time. In addition, the errors of the STEPS method bracketed those of the two deterministic methods except for some cases where the advection showed a lower error than the whole STEPS ensemble. The evolution of the errors does not indicate a dependency on catchment size, although the largest catchment (Erft at Neubrueck) showed lower MAE errors for a lead time of 3 h. The variability of errors across the catchments reflects the effect of their location with respect to the precipitation field.

b. *Added hydrological value of QPN*

To investigate the added value of the tested QPN methods from a hydrological point of view, we first show in section 4b(1) the hindcasted hydrographs using RADOLAN and the QPE product RAVKDP for the 14 July 2021 event (Fig. 6) based on which the quality of the forecasted hydrographs is estimated using NSE (Fig. 7). Second, in section 4b(2), we show the skill of the QPN methods computed using the CRPS between the corresponding forecasted hydrographs and the hindcasted hydrograph using QPE, with respect to the benchmarks of the hydrological persistence or the zero-precipitation nowcasts (Figs. 8 and 9). Finally, in section 4b(3), we show the gains in lead time obtained using either the efficiency-based approach (with NSE) or the skill-based approach (with CRPS) and depending on the benchmark (hydrological persistence or zero-precipitation nowcasts; Fig. 10).

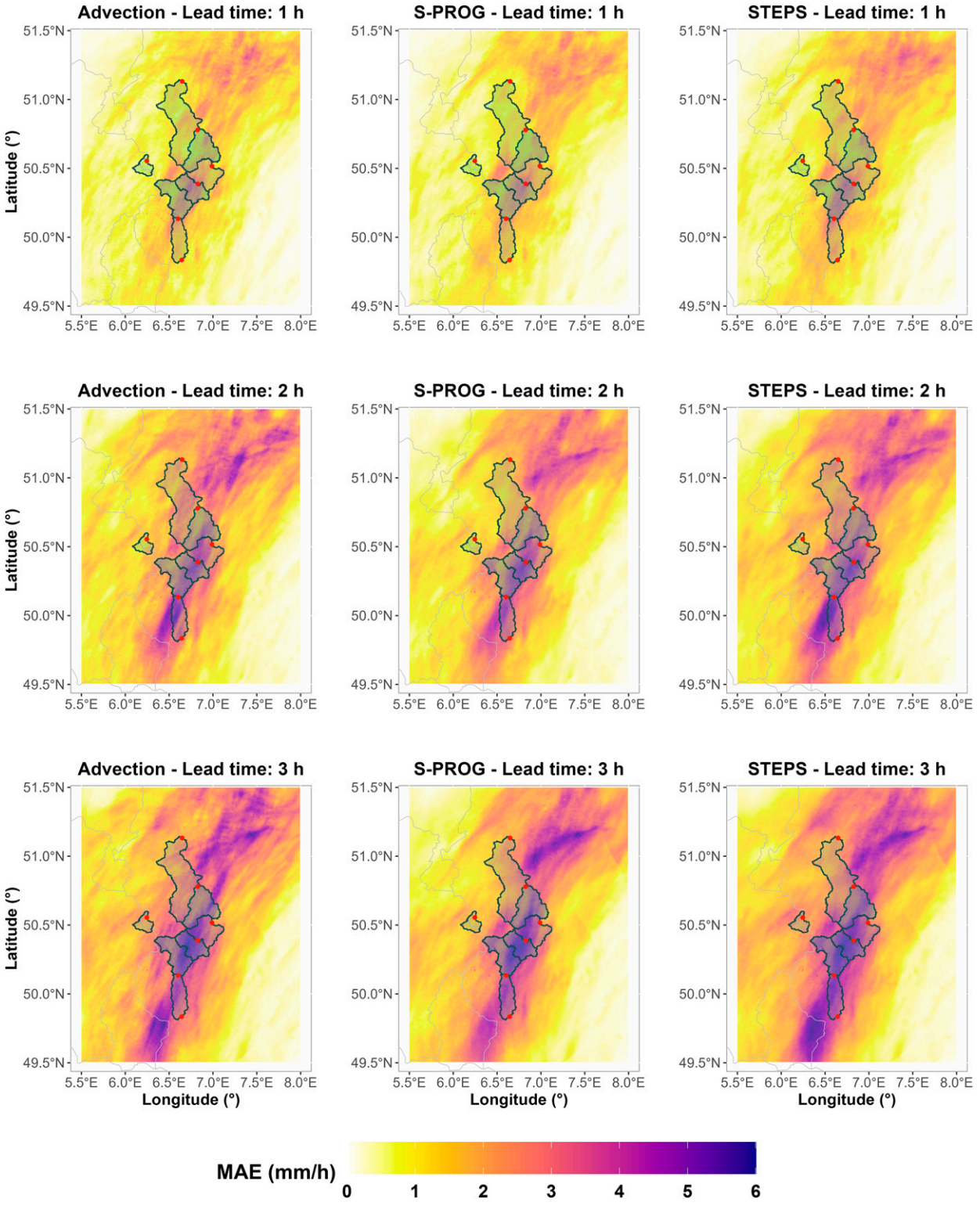

FIG. 4. Mean absolute errors (MAE) of QPN obtained using (left) advection, (center) S-PROG, and (right) STEPS for the (top) 1-, (middle) 2-, and (bottom) 3-h lead time. MAE values were computed with respect to observed precipitation rates from RAVKDP. For STEPS, the median errors over the 20 members are shown.

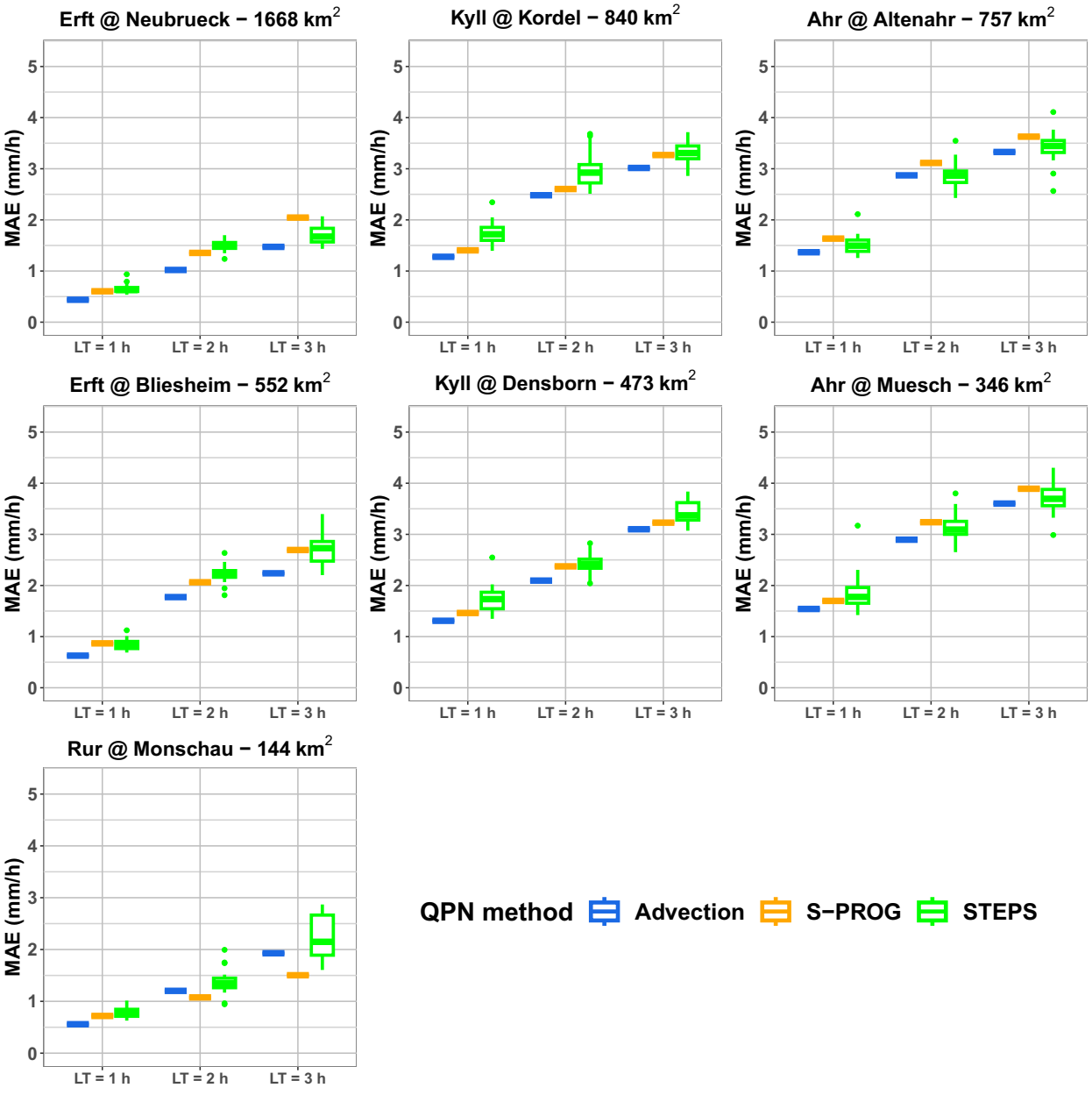

FIG. 5. Mean absolute errors between observed precipitation (QPE) and QPN estimated using advection, S-PROG, and STEPS at the scale of each catchment. LT refers to lead time.

1) HINDCASTED HYDROGRAPHS AND QUALITY OF THE FORECASTED HYDROGRAPHS

To illustrate the dynamics of the catchment responses to the extreme rainfall event of 14 July 2021, simulated hydrographs using RADOLAN (prior to 14 July 2021) and the QPE product RAVKDP (for 14 July 2021) by GR4H and ParFlowCLM are presented in Fig. 6. They indicate that the highest recorded peak flow prior to July 2021 (in orange dashed lines) was surpassed by model simulations at least once in all the catchments except the Rur at Monschau. However, where available, the last measured peak flow before the unavailability of records was surpassed by model simulations only for the catchments drained by the Ahr river. Qualitatively, GR4H and ParFlowCLM agreed for the catchments drained by the Ahr and the Kyll, whereas they significantly disagreed over the Erft and the Rur. Finally, the spread in the GR4H simulations reflects the large uncertainty in simulated hydrographs due to parameter uncertainty (Saadi et al. 2023).

From the hydrological viewpoint, the three QPN methods yielded very similar hydrological forecasts across the seven catchments, as suggested by their NSE scores in Fig. 7. At the threshold of $NSE_{th} = 0.9$, the three methods yielded

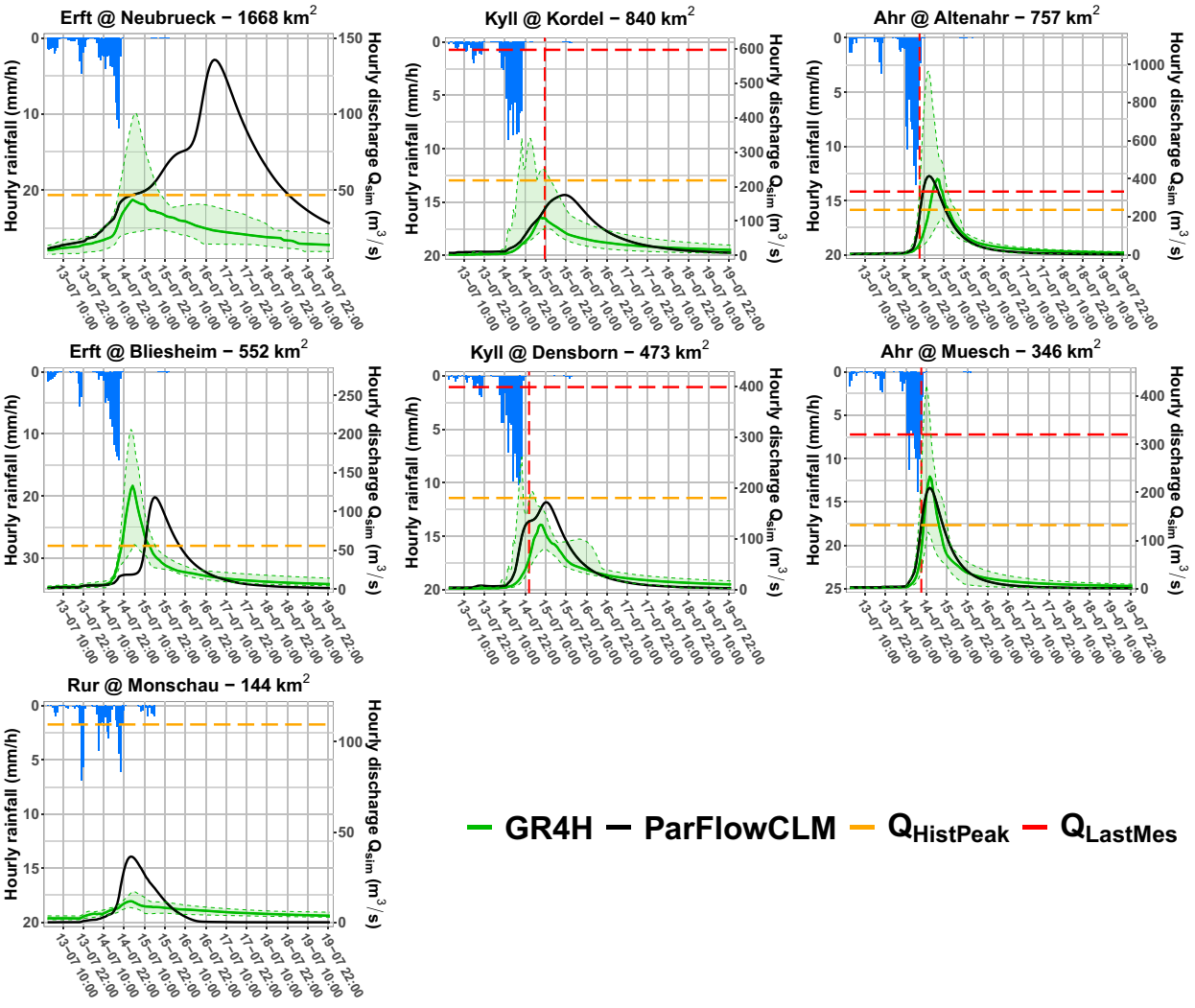

FIG. 6. Simulated hydrographs for the July 2021 events using 12 parameter sets with GR4H (shaded area in green) and one parameter set with ParFlowCLM (in black). Orange dashed lines indicate the highest recorded peak flow (Q_{HistPeak}) prior to July 2021. Subject to availability, red dashed horizontal lines indicate the reported last measured peak flows before the failure of the monitoring devices (Q_{LastMes}), and red dashed vertical lines are their timings. Hydrographs are simulated using the QPE product RAVKDP on 14 Jul 2021 and RADOLAN for the remaining time steps. For GR4H, the shaded area is delimited by the minimum–maximum of the simulations at each time step using 12 parameter sets for each catchment.

satisfactory hydrological forecasts for lead times ranging from 1 up to 5 h (GR4H for the Erft at Neubrueck, ParFlowCLM for the Rur at Monschau). The benchmark of the hydrological persistence (Q) obtained the fastest decreasing NSE curves, which is expected given its limits for a highly variable catchment response during the event. However, the benchmark of the zero-precipitation nowcasts (ZNC) succeeded in yielding better hydrological forecasts using ParFlowCLM for the catchments drained by the Kyll. For these two catchments, the use of the QPN products led to early increases of the forecasted hydrographs with respect to the hindcasted hydrograph, resulting in deteriorated NSE values compared with the ZNC benchmark in the early lead times. Finally, the QPN methods led to more satisfactory hydrological forecasts when using GR4H than when using ParFlowCLM, except for the Rur at Monschau.

2) SKILL OF THE QPN METHODS

The three QPN methods were also similar in terms of their skill with respect to the benchmark of hydrological persistence, as can be seen in Fig. 8. The skill indicates that using QPN yielded better forecasts than the persistence model for lead times higher than 30 h, except for the Rur at Monschau where the skill dropped after only 16 h. Note that this should be interpreted in the light of the quality of the forecasted hydrographs by both the QPN methods and the benchmark of the hydrological persistence, which in all cases had negative NSE values after a lead time of 12 h (see Fig. 7). The evolution of the

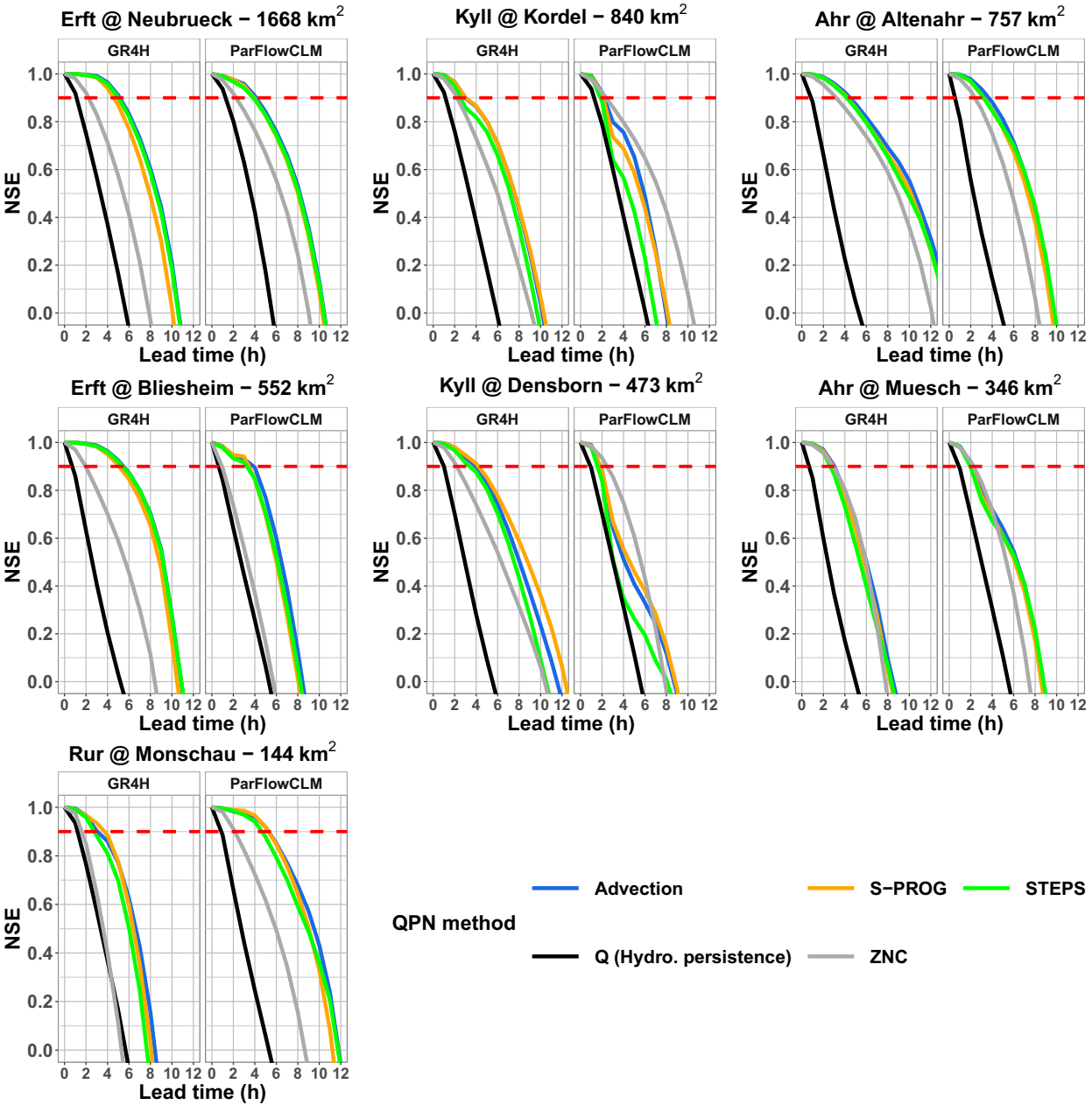

FIG. 7. Evolution of the Nash–Sutcliffe efficiency (NSE) of the forecasted hydrographs using the QPN methods and the benchmarks (hydrological persistence Q, zero-precipitation nowcasts ZNC) with respect to lead time. Red dashed lines indicate the NSE threshold $NSE_{th} = 0.9$. For GR4H, the curves represent the median score from the 12 simulations. For STEPS, the curves represent the median score from the 20 members.

skill was somewhat distinct for each catchment, but similar for the catchments drained by the same river, which indicates a dependency on location with respect to the precipitation event. The rebound in the skill curve for the catchments drained by the Ahr River may reflect the change in the forecasted part of the hydrograph from the rising to the falling limb. Finally, both models showed similar evaluation of the three QPN methods, with GR4H showing slightly higher skill scores than ParFlowCLM at the very short lead times.

Changing the benchmark to zero-precipitation nowcasts (ZNC) had a limited impact on the skill of the tested QPN, as suggested by Fig. 9. Namely, the skill slightly decreased compared with the hydrological persistence in Fig. 8, indicating that the ZNC is a more challenging benchmark to beat than the hydrological persistence. This is somewhat expected given the better performances of ZNC compared with hydrological persistence, as can be seen in Fig. 7. This is not, however, the case for all catchments. The Rur at Monschau indicates that

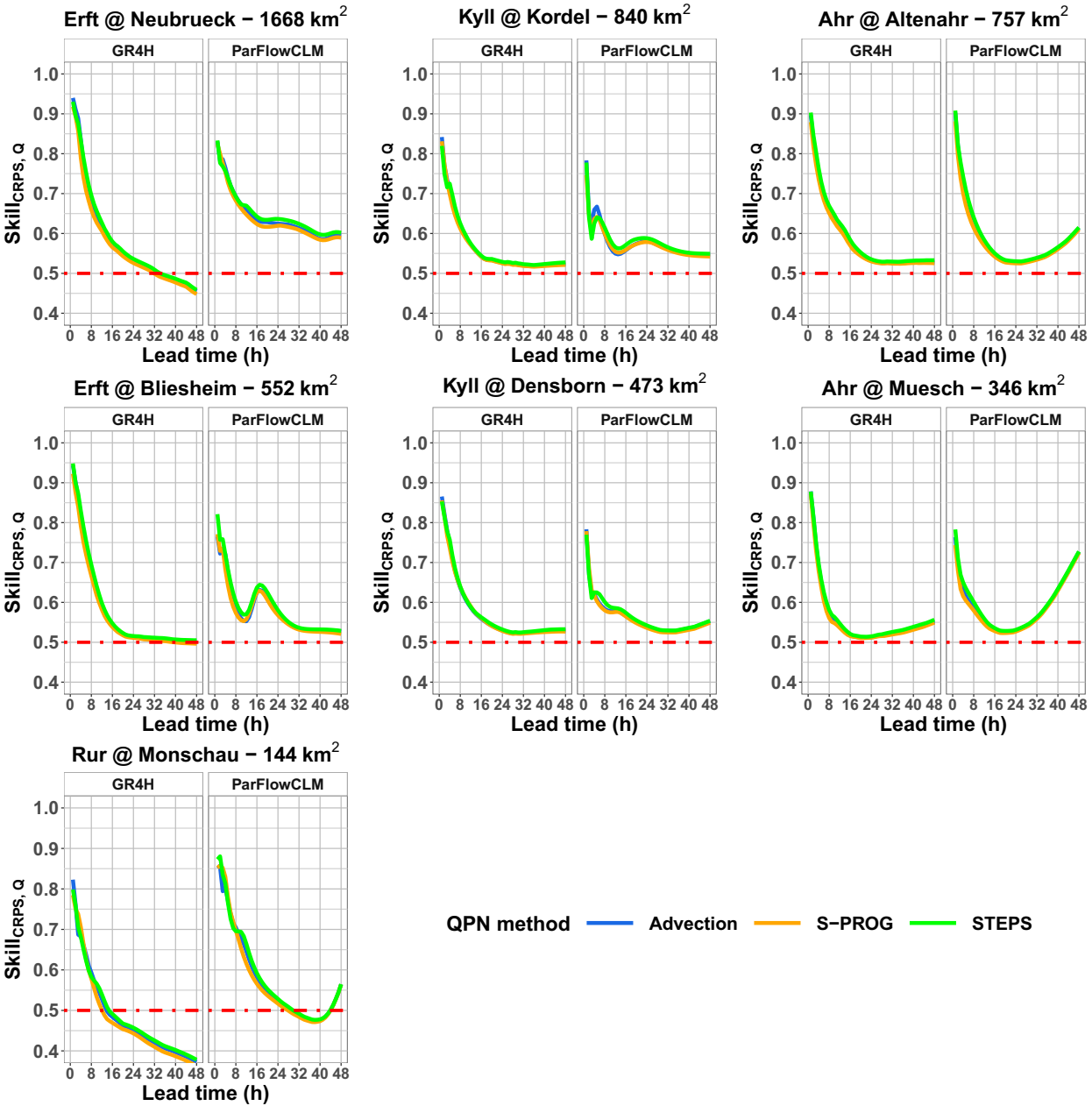

FIG. 8. Evolution of the skill of the forecasted hydrographs using QPN with respect to a hydrological persistence benchmark ($Skill_{CPRS,Q}$) over the seven catchments. Dashed red lines indicate a skill of 0.5, i.e., using QPN is as good as the hydrological persistence model. For GR4H, the curves represent the median score from the 12 simulations.

the ZNC benchmark is easier to outperform than the hydrological persistence. The rebound effect observed in Fig. 8 for the Ahr catchments disappeared with the ZNC benchmark, as all QPN tend to be equal to ZNC after the end of the event.

3) GAINS IN LEAD TIME WITH RESPECT TO THE BENCHMARKS

The gain in lead time reflects the dependency on the evaluation method, the benchmark and the chosen threshold for NSE or for the skill, as shown in Fig. 10. Based on NSE and for a threshold at $NSE_{th} = 0.9$ (Figs. 10a,b), gains in lead time ranged between 1 and 5 h with GR4H (on average) and between 0 and 5 h with ParFlowCLM with respect to the hydrological persistence. With respect to ZNC, gains ranged between 0 and 4 h with GR4H (on average) and in some cases there were losses with ParFlowCLM, specifically in the catchments drained by the Kyll and the Ahr at Muesch. Changing the threshold from $NSE_{th} = 0.9$ to 0.5 led to increases in the gains only with respect to the hydrological persistence (Figs. 10c,d). In this case, the gains ranged between 2 and 8 h with GR4H

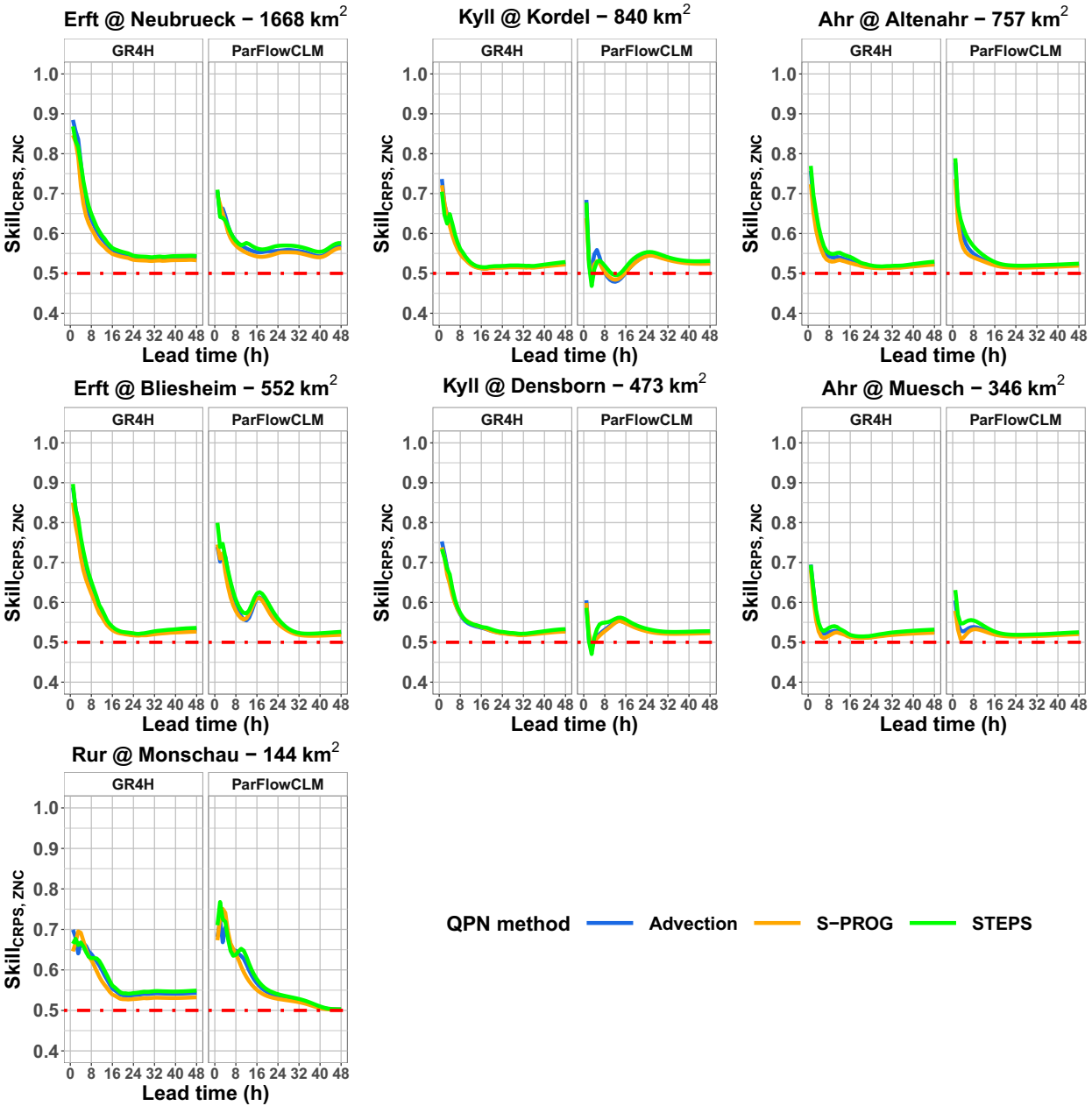

FIG. 9. Evolution of the skill of the forecasted hydrographs using QPN with respect to forecasted hydrographs with zero-precipitation nowcasts (ZNC). Red dashed lines indicate a skill of 0.5, i.e., using QPN is as good as feeding the models with ZNC. For GR4H, the curves represent the median score from the 12 simulations.

and between 0 and 7 h with ParFlowCLM. With respect to ZNC, changing the NSE threshold from 0.9 to 0.5 resulted in poorer gains, especially for ParFlowCLM (range: from −3 to 4 h). This is caused by faster decreases in the quality of the forecasted hydrographs with the QPN methods compared with those forecasted with ZNC (Fig. 7)

Using the skill-based approach, choosing the default threshold (0.5, Figs. 10e,f) yielded much larger gains in lead time compared with a more demanding threshold (0.67, Figs. 10g,h). With a threshold of 0.5 (Figs. 10e,f), the improvements were up to 48 h, the maximum range to which we limited our analysis, which should be interpreted in the light of the poor performances of the benchmarks (Fig. 7). With a threshold of 0.67, the use of QPN improved the forecast lead time by 4 h up to 9 h with GR4H, and by 1 h up to 12 h with ParFlowCLM (Erft at Neubrueck, the largest catchment) with respect to the hydrological persistence as benchmark (Fig. 10g). With respect to ZNC (Fig. 10h), the improvements ranged between 1 and 7 h with GR4H and between 0 and 5 h with ParFlowCLM.

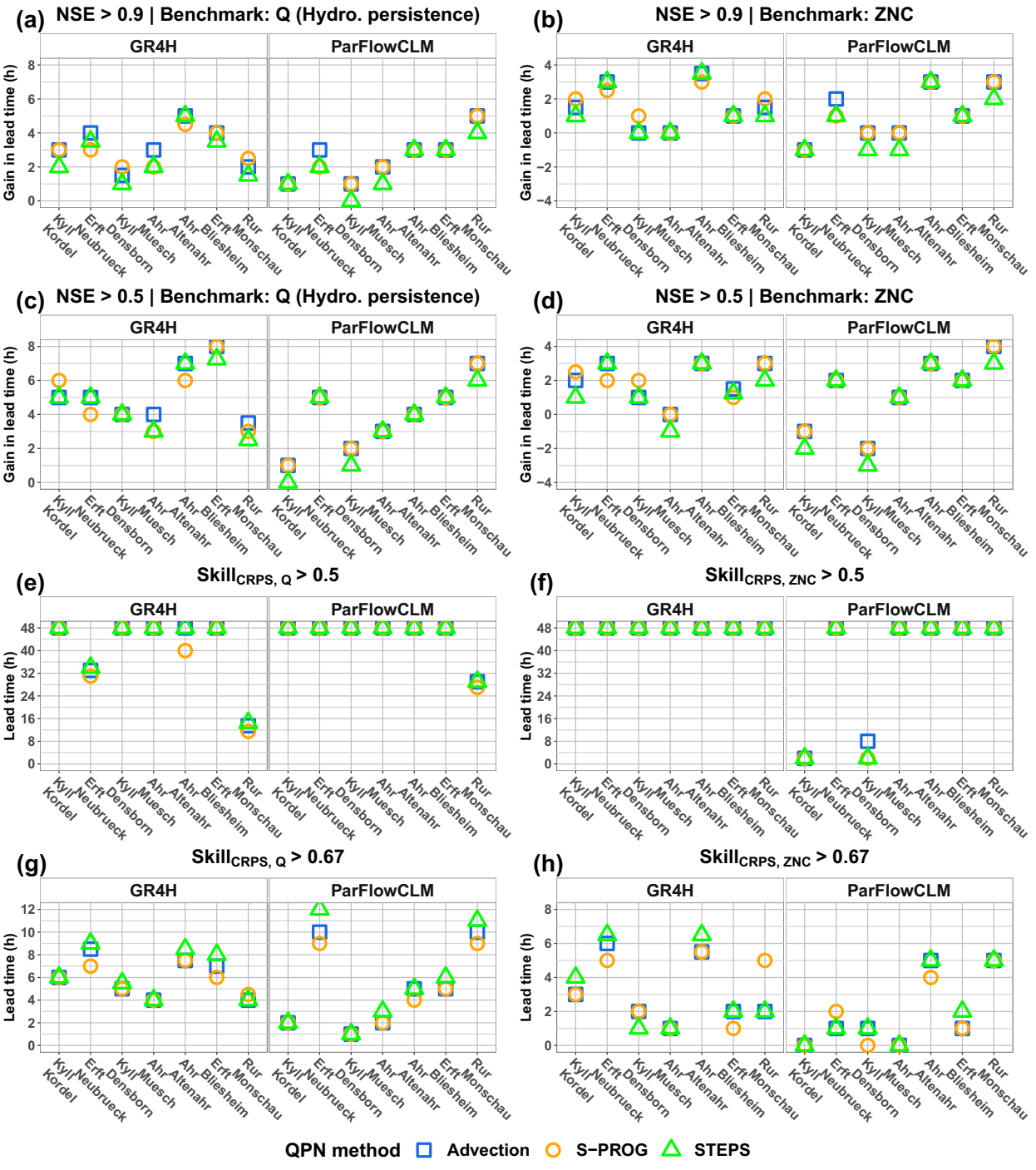

FIG. 10. Gain in lead time based on Nash–Sutcliffe score with respect to (left) hydrological persistence and (right) zero-precipitation nowcasts as benchmarks with a score threshold at (a),(b) 0.9 and (c),(d) 0.5. Lead time (h) up to which using QPN was better than the benchmark (hydrological persistence on the left, zero-precipitation nowcasts on the right) according to a skill threshold at (e),(f) 0.5 and (g),(h) 0.67. On the *x* axis, catchments are ranked with decreasing Gravelius index from left to right. Note that the evolution of the skills was computed up to a maximum of 48 h.

Finally, we found no consistent pattern in the gains in lead time with respect to catchment area (not shown here). However, when the catchments are ranked by their Gravelius index [Eq. (8)], the added value of the QPN methods with respect to the hydrological persistence using ParFlowCLM increased as the Gravelius index decreased (Figs. 10a,c,g). This suggests that with ParFlowCLM, using the QPN methods was more beneficial for catchments with a more compact shape.

This dependency on the catchment shape was not detectable with GR4H.

5. Discussion

For the major flooding event of July 2021 in our study region, the three tested QPN products obtained very similar performances in terms of both the reproduction of observed precipitation and hydrological forecasting. The highly similar performances of the three methods can be attributed to two factors. First, the aggregation to the hourly time step may have filtered out the differences between the three methods, which are reported by previous studies to be in the order of a few to tens of minutes (Ayzel et al. 2019; Berenguer et al. 2005; Heuvelink et al. 2020; Imhoff et al. 2020). Figure 3 shows, however, that the methods are already similar at 5-min resolution, with a slightly better performance for S-PROG. Second, the persistent nature of the event (see event hyetographs in Fig. 6) might have made it as easy to be forecasted by simple nowcasting methods (advection) as by more sophisticated ones (S-PROG and STEPS). The poor performances of the STEPS ensemble compared with the ensemble mean STEPS-m suggests that the perturbations of the deterministic forecast got penalized for this event (Heuvelink et al. 2020). In addition, comparing the spread in the STEPS ensemble with the deterministic STEPS-m or S-PROG forecasts suggests that the STEPS ensemble might have underestimated the uncertainty in the evolution of the precipitation field for this event (Foresti et al. 2016). Nevertheless, the similarity in terms of performances between STEPS and the deterministic methods should not undermine its benefits in providing probabilistic nowcasts, which are of greater value for decision makers than deterministic ones (Fundel et al. 2019; Merz et al. 2020).

Overall, our results show that the QPN methods improved the hydrological forecasts compared with hydrological persistence or with the zero-precipitation nowcasts. Previous studies such as Heuvelink et al. (2020), Berenguer et al. (2005), and Vivoni et al. (2006) reported improvements of 2–6 h for catchments of $\sim 10^3$ km^2 of size. First, some of the differences may be explained by the hydroclimatic settings of the studied catchments and/or their characteristic response times (or concentration times). Berenguer et al. (2005) studied catchments located in the Mediterranean region (northeast of Spain), for which the fast response may explain the absence of significant improvements using QPN beyond 2 h. On the contrary, Heuvelink et al. (2020) showed that significant improvements in the hydrological forecasts can be obtained when using 3-h-long QPN for catchments under humid, temperate climate (Netherlands) and characterized by slower responses than Mediterranean catchments. In our case, the improvements are in general limited to 4 h with respect to a benchmark of zero-precipitation nowcasts, with a high variability from one catchment to another (Figs. 10b,d). For some catchments (drained by the Kyll and the Ahr at Muesch), the QPN methods showed worse performances compared with the zero-precipitation nowcasts, especially with ParFlowCLM. The variability in the performances of the QPN from one catchment to another can be either explained by the location of each catchment within the precipitation field, or their properties that modulate the delay between the precipitation and the catchment response.

Second, some of the differences can be attributed to methodological choices. Using the same approach as in Heuvelink et al. (2020) and Berenguer et al. (2005) based on the quality of the forecasted hydrographs measured by NSE, we obtained similar improvements up to 4–5 h with an NSE threshold at 0.9, depending on the benchmark (Figs. 10a,b). However, this NSE-based approach suffers from the arbitrary selection of an efficiency threshold, which, according to Figs. 10a–d, impacted the estimation of the added value of QPN, especially with respect to the hydrological persistence as benchmark. We attempted to circumvent this issue by following a skill-based approach, which provides an a priori objective threshold [i.e., 0.5, see Eq. (7)], but leads to too optimistic results and suggests that the skill of QPN lasts for much longer lead times (Figs. 8 and 9, except for the Rur at Monschau in Fig. 8 and the catchments drained by the Kyll in Fig. 9). By adopting a more demanding threshold such as 0.67, our skill-based approach leads to results that agree with previous studies (Figs. 10g,h; Heuvelink et al. 2020; Berenguer et al. 2005). The combination of a skill-based approach with an analysis of the quality of forecasted hydrographs helped objectively estimate the added value of the QPN products and avoid a distorted evaluation when the benchmark performs poorly. In all cases, the obtained improvements may look small, but they can still be of high value for emergency management and the fire services involved in event response (Speight et al. 2021). However, if observed hydrographs of the event were available and used, they would probably have led to lower added value of QPN because of additional errors resulting from the disagreement between the QPE product and corresponding hindcasted hydrographs with the observed precipitation and observed hydrographs, respectively.

Several studies reported a dependency of improvements on the catchment size and the event type (convective versus stratiform; Berenguer et al. 2005; Heuvelink et al. 2020; Imhoff et al. 2020). We showed that, in addition, there is also a dependency on methodological choices, namely, the chosen benchmark and the applied hydrological model. The dependency on the benchmark used to estimate the skill was visible (albeit to a limited extent) from comparing Figs. 8 and 9, which warns that choosing a simple model (such as hydrological persistence) may lead to overly optimistic interpretations of the improvements (Pappenberger et al. 2015b). The dependency on catchment size was hardly visible over our catchment set for the studied event, except for the largest catchment (Erft at Neubrueck, 1668 km^2) which showed a slow decrease in the skill with respect to lead time. The low sensitivity of the skills to catchment size is perhaps a result of working on a single event, which emphasizes the impact of the differences of precipitation amounts registered in the study catchments. Nevertheless, we noticed a dependency on catchment shape when the hydrological persistence is chosen as benchmark and with ParFlowCLM as a hydrological model (Figs. 10a,c,g), but not with GR4H. The fact that GR4H did not mirror the effect of catchment shape can be explained by

the sensitivity of GR4H parameters to the anthropogenic effects through calibration on historical observations (Saadi et al. 2020), which are behind the large differences between model simulations for the Erft at Neubrueck and the Rur at Monschau in Fig. 6. Accounting for the anthropogenic effects by GR4H (even implicitly) may have buffered the effect of catchment shape on the catchment response.

The dependency of the skill on the catchment response, i.e., the event hydrograph, was accentuated by the adoption of the hydrological persistence as benchmark. Since the reference event hydrograph is the one simulated by the hydrological models using the QPE, the NSE curves in Fig. 7 suggest that the flashier and the higher the hindcasted peak flow, the higher the benefit of the use of QPN. This can be seen from comparing the hindcasted hydrographs by GR4H and ParFlowCLM for the Erft at Bliesheim (Fig. 6), where GR4H hindcasted an earlier peak flow than ParFlowCLM, mirrored by better NSE values for the three QPN methods when GR4H was applied for this catchment (Fig. 7). Conversely, the simulated hydrograph by GR4H for the Rur at Monschau was smoother than the one simulated by ParFlowCLM, which was accompanied by better gains in lead times for QPN with ParFlowCLM (Fig. 7). The smoother hydrographs led to less gains because the use of QPN led to earlier rises in the forecasted hydrographs, which penalized their use. These differences also reflect the errors of QPN with respect to the QPE product, which are highlighted by the distributed ParFlowCLM for the catchments drained by the Kyll, where the zero-precipitation nowcasts showed similar or better NSE scores at the early time steps (Fig. 7).

Nevertheless, the choice of the adopted hydrological model did not much alter the conclusions regarding the similarity of the tested QPN methods. The agreement between QPN according to the distributed ParFlowCLM model suggests that the methods agreed also in the spatial distribution of precipitation for this particular event at least from a hydrological point of view, in line with the MAE patterns in Fig. 4. Effects of uncertainties in parameter estimation of the hydrological models were not included here, but they would be relatively low given the general agreement of the three methods in terms of predicting the observed QPE. Effects of uncertainties in initial moisture conditions were minimized by the long spinup period of both GR4H and ParFlowCLM models prior to the event.

Focusing only on one event limits our investigation of other factors that could have impacted the skill of the tested nowcasting methods, such as the type of the event and the season (Imhoff et al. 2020). The absence of observed discharge values limits the evaluation of the accuracy of model simulations, but that should not undermine the obtained improvements by the use of QPN. Quantifying the added value of the tested QPN with respect to an NWP-based benchmark for this event would give more convincing results from an operational point of view, given the relative poorness of the adopted benchmarks in our study. Finally, the relatively heavy cost of model simulations with ParFlowCLM (especially when applied with the probabilistic STEPS nowcasts) hampers its test with more parameter sets, which could have an impact on the evaluation of the skill, especially with respect to the hydrological persistence as a benchmark.

6. Conclusions and future work

We investigated the usefulness of using precipitation nowcasts to improve the skill of two hydrological models in forecasting the response of seven catchments located in the west of Germany for the disastrous July 2021 event. We evaluated three precipitation nowcasting techniques, namely, the Lagrangian advection, S-PROG, and the probabilistic method STEPS. Our evaluation consisted of analyzing their ability in forecasting the observed precipitation at 5-min and hourly time steps, then in improving the ability of two contrasting hydrological models, GR4H and ParFlowCLM, in reproducing the simulated hydrographs by the hydrological models using observed precipitation (or hindcasted hydrographs). For the July 2021 events in our study region, our main conclusions are as follows:

1) The three methods improved the forecasting skill of the hydrological models with respect to two benchmarks, the hydrological persistence and the zero-precipitation nowcasts. These improvements varied from one catchment to another, and reached up to 4–5 h according to an NSE at 0.9 and up to 12 h according to the CRPS skill at a threshold of 0.67 (i.e., the use of QPN halved the forecasting errors of the benchmarks).
2) The three methods obtained very similar performances in terms of both precipitation and discharge forecasting. In particular, the deterministic methods (advection and S-PROG) performed as good as the average/median probabilistic one (STEPS).
3) The use of a conceptual, lumped model (GR4H) led to similar conclusions as with a physically based, 3D-distributed model (ParFlowCLM). However, the gains in lead time were on average lower with ParFlowCLM than with GR4H. The differences between the two models can be attributed to the anthropogenic influences in the catchments, which are implicitly accounted for by GR4H through its calibrated parameters on historical observations.
4) The choice of the evaluation method, the benchmark and the skill threshold impacted the estimation of the added value of the QPN methods.

As future work, more robust conclusions would be obtained by considering a large sample of events with a variety of seasons and typologies. Increasing the horizon of the input precipitation to the models with quantitative precipitation forecasts that make use of (convection-permitting) NWP outputs through blending approaches (Lovat et al. 2022; Speight et al. 2021; Clark et al. 2016) would shed more light on the ability of the current hydrometeorological chains in hedging the damages by issuing useful and timely flood warnings. Quantifying the economic gains from including precipitation nowcasts (Le Bihan et al. 2017; Pappenberger et al. 2015a) would provide more convincing arguments about their usefulness. Finally, the added value of the nowcasting techniques presented in this work motivates exploiting their benefit in

generating nationwide and useful short-time forecasts for better disaster preparedness (Reinoso-Rondinel et al. 2022).

Acknowledgments. This study is part of the RealPEP (Near-Realtime Quantitative Precipitation Estimation and Prediction, https://www2.meteo.uni-bonn.de/realpep/doku.php) P4 project (Evaluation of QPE and QPN improvements in a flash flood nowcasting framework with data assimilation), funded by the Deutsche Forschungsgemeinschaft (German Research Foundation). The authors gratefully acknowledge the Earth System Modeling Project (ESM) for funding this work by providing computing time for ParFlowCLM runs on the ESM partition of the supercomputer JUWELS at the Jülich Supercomputing Centre (JSC). The authors declare they have no conflict of interest.

Data availability statement. All original data are public, except for the QPN and the QPE products generated for the study, which can be made available upon reasonable request from the authors.

REFERENCES

Abrams, M., R. Crippen, and H. Fujisada, 2020: ASTER Global Digital Elevation Model (GDEM) and ASTER Global Water Body Dataset (ASTWBD). *Remote Sens.*, **12**, 1156, https://doi.org/10.3390/rs12071156.

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.

Atencia, A., and I. Zawadzki, 2014: A comparison of two techniques for generating nowcasting ensembles. Part I: Lagrangian ensemble technique. *Mon. Wea. Rev.*, **142**, 4036–4052, https://doi.org/10.1175/MWR-D-13-00117.1.

——, and ——, 2015: A comparison of two techniques for generating nowcasting ensembles. Part II: Analogs selection and comparison of techniques. *Mon. Wea. Rev.*, **143**, 2890–2908, https://doi.org/10.1175/MWR-D-14-00342.1.

Auger, L., O. Dupont, S. Hagelin, P. Brousseau, and P. Brovelli, 2015: AROME–NWC: A new nowcasting tool based on an operational mesoscale forecasting system. *Quart. J. Roy. Meteor. Soc.*, **141**, 1603–1611, https://doi.org/10.1002/qj.2463.

Ayzel, G., M. Heistermann, and T. Winterrath, 2019: Optical flow models as an open benchmark for radar-based precipitation nowcasting (rainymotion v0.1). *Geosci. Model Dev.*, **12**, 1387–1402, https://doi.org/10.5194/gmd-12-1387-2019.

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55, https://doi.org/10.1038/nature14956.

Belleflamme, A., and Coauthors, 2023: Hydrological forecasting at impact scale: the integrated ParFlow hydrological model at 0.6 km for climate resilient water resource management over Germany. *Front. Water*, **5**, 1183642, https://doi.org/10.3389/frwa.2023.1183642.

Bendjoudi, H., and P. Hubert, 2002: The Gravelius compactness coefficient: Critical analysis of a shape index for drainage basins. *Hydrol. Sci. J.*, **47**, 921–930, https://doi.org/10.1080/02626660209493000.

Berenguer, M., C. Corral, R. Sánchez-Diezma, and D. Sempere-Torres, 2005: Hydrological validation of a radar-based nowcasting technique. *J. Hydrometeor.*, **6**, 532–549, https://doi.org/10.1175/JHM433.1.

——, 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.

——, M. Surcel, I. Zawadzki, M. Xue, and F. Kong, 2012: The diurnal cycle of precipitation from continental radar mosaics and numerical weather prediction models. Part II: Intercomparison among numerical models and with nowcasting. *Mon. Wea. Rev.*, **140**, 2689–2705, https://doi.org/10.1175/MWR-D-11-00181.1.

Berthet, L., V. Andréassian, C. Perrin, and P. Javelle, 2009: How crucial is it to account for the antecedent moisture conditions in flood forecasting? Comparison of event-based and continuous approaches on 178 catchments. *Hydrol. Earth Syst. Sci.*, **13**, 819–831, https://doi.org/10.5194/hess-13-819-2009.

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.

Chen, C., J. Twycross, and J. M. Garibaldi, 2017: A new accuracy measure based on bounded relative error for time series forecasting. *PLOS ONE*, **12**, e0174202, https://doi.org/10.1371/journal.pone.0174202.

Chen, J.-Y., S. Trömel, A. Ryzhkov, and C. Simmer, 2021: Assessing the benefits of specific attenuation for quantitative precipitation estimation with a C-band radar network. *J. Hydrometeor.*, **22**, 2617–2631, https://doi.org/10.1175/JHM-D-20-0299.1.

——, R. Reinoso-Rondinel, S. Trömel, C. Simmer, and A. Ryzhkov, 2022: A radar-based quantitative precipitation estimation algorithm to overcome the impact of vertical gradients of warm-rain precipitation: The flood in western Germany on 14 July 2021. *J. Hydrometeor.*, **24**, 521–536, https://doi.org/10.1175/JHM-D-22-0111.1.

Clark, P., N. Roberts, H. Lean, S. P. Ballard, and C. Charlton-Perez, 2016: Convection-permitting models: A step-change in rainfall forecasting. *Meteor. Appl.*, **23**, 165–181, https://doi.org/10.1002/met.1538.

Cloke, H. L., and F. Pappenberger, 2009: Ensemble flood forecasting: A review. *J. Hydrol.*, **375**, 613–626, https://doi.org/10.1016/j.jhydrol.2009.06.005.

Coron, L., G. Thirel, O. Delaigue, C. Perrin, and V. Andréassian, 2017: The suite of lumped GR hydrological models in an R package. *Environ. Modell. Software*, **94**, 166–171, https://doi.org/10.1016/j.envsoft.2017.05.002.

Duscher, K., A. Günther, A. Richts, P. Clos, U. Philipp, and W. Struckmeier, 2015: The GIS layers of the "International Hydrogeological Map of Europe 1:1,500,000" in a vector format. *Hydrogeol. J.*, **23**, 1867–1875, https://doi.org/10.1007/s10040-015-1296-4.

Dottori, F., and Coauthors, 2018: Increased human and economic losses from river flooding with anthropogenic warming. *Nat. Climate Change*, **8**, 781–786, https://doi.org/10.1038/s41558-018-0257-z.

Dougherty, E., and K. L. Rasmussen, 2020: Changes in future flash flood–producing storms in the United States. *J. Hydrometeor.*, **21**, 2221–2236, https://doi.org/10.1175/JHM-D-20-0014.1.

Edijatno, N. de Oliveira Nascimento, X. Yang, Z. Makhlouf, and C. Michel, 1999: GR3J: A daily watershed model with three free parameters. *Hydrol. Sci. J.*, **44**, 263–277, https://doi.org/10.1080/02626669909492221.

Ficchì, A., C. Perrin, and V. Andréassian, 2019: Hydrological modelling at multiple sub-daily time steps: Model improvement via flux-matching. *J. Hydrol.*, **575**, 1308–1327, https://doi.org/10.1016/j.jhydrol.2019.05.084.

Foresti, L., L. Panziera, P. V. Mandapaka, U. Germann, and A. Seed, 2015: Retrieval of analogue radar images for ensemble nowcasting of orographic rainfall. *Meteor. Appl.*, **22**, 141–155, https://doi.org/10.1002/met.1416.

——, 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.

Fowler, H. J., and Coauthors, 2021: Anthropogenic intensification of short-duration rainfall extremes. *Nat. Rev. Earth Environ.*, **2**, 107–122, https://doi.org/10.1038/s43017-020-00128-6.

Fundel, V. J., N. Fleischhut, S. M. Herzog, M. Göber, and R. Hagedorn, 2019: Promoting the use of probabilistic weather forecasts through a dialogue between scientists, developers and end-users. *Quart. J. Roy. Meteor. Soc.*, **145**, 210–231, https://doi.org/10.1002/qj.3482.

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.

Hengl, T., and Coauthors, 2017: SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, **12**, e0169748, https://doi.org/10.1371/journal.pone.0169748.

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, 2020: Spatial and temporal evaluation of radar rainfall nowcasting techniques on 1,533 events. *Water Resour. Res.*, **56**, e2019WR026723, https://doi.org/10.1029/2019WR026723.

——, ——, 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.

Junghänel, T., and Coauthors, 2021: Hydro-klimatologische Einordnung der Stark- und Dauerniederschläge in Teilen Deutschlands im Zusammenhang mit dem Tiefdruckgebiet "Bernd" vom 12. bis 19. Juli 2021. Deutscher Wetterdienst, 16 pp., https://www.dwd.de/DE/leistungen/besondereereignisse/niederschlag/20210721_bericht_starkniederschlaege_tief_bernd.pdf;jsessionid=1B3CCDDBF6FF5C13C36F83E0CFB9345F.live11053?__blob=publicationFile&v=10.

Kollet, S. J., and R. M. Maxwell, 2006: Integrated surface–groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. *Adv. Water Resour.*, **29**, 945–958, https://doi.org/10.1016/j.advwatres.2005.08.006.

Kreienkamp, F., and Coauthors, 2021: Rapid attribution of heavy rainfall events leading to the severe flooding in Western Europe during July 2021. World Weather Attribution, https://www.worldweatherattribution.org/heavy-rainfall-which-led-to-severe-flooding-in-western-europe-made-more-likely-by-climate-change.

Kuffour, B. N. O., N. B. Engdahl, C. S. Woodward, L. E. Condon, S. Kollet, and R. M. Maxwell, 2020: Simulating coupled surface–subsurface flows with ParFlow v3.5.0: Capabilities, applications, and ongoing development of an open-source, massively parallel, integrated hydrologic model. *Geosci. Model Dev.*, **13**, 1373–1397, https://doi.org/10.5194/gmd-13-1373-2020.

Langanke, T., M. Steidl, C. Schleicher, and C. Sannier, 2016: Copernicus land monitoring service—High resolution layer imperviousness: Product specifications document. European Environment Agency, 39 pp., https://land.copernicus.eu/user-corner/technical-library/hrl-imperviousness-technical-document-prod-2015.

Le Bihan, G., O. Payrastre, E. Gaume, D. Moncoulon, and F. Pons, 2017: The challenge of forecasting impacts of flash floods: Test of a simplified hydraulic approach and validation based on insurance claim data. *Hydrol. Earth Syst. Sci.*, **21**, 5911–5928, https://doi.org/10.5194/hess-21-5911-2017.

Li, L., W. Schmid, and J. Joss, 1995: Nowcasting of motion and growth of precipitation with radar over a complex orography. *J. Appl. Meteor.*, **34**, 1286–1300, https://doi.org/10.1175/1520-0450(1995)034<1286:NOMAGO>2.0.CO;2.

Lin, C., S. Vasić, A. Kilambi, B. Turner, and I. Zawadzki, 2005: Precipitation forecast skill of numerical weather prediction models and radar nowcasts. *Geophys. Res. Lett.*, **32**, L14801, https://doi.org/10.1029/2005GL023451.

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.

Maxwell, R. M., 2013: A terrain-following grid transform and preconditioner for parallel, large-scale, integrated hydrologic modeling. *Adv. Water Resour.*, **53**, 109–117, https://doi.org/10.1016/j.advwatres.2012.10.001.

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., and Coauthors, 2020: Impact forecasting to support emergency management of natural hazards. *Rev. Geophys.*, **58**, e2020RG000704, https://doi.org/10.1029/2020RG000704.

Mohr, S., and Coauthors, 2023: A multi-disciplinary analysis of the exceptional flood event of July 2021 in central Europe—Part 1: Event description and analysis. *Nat. Hazards Earth Syst. Sci.*, **23**, 525–551, https://doi.org/10.5194/nhess-23-525-2023.

Moisselin, J.-M., P. Cau, C. Jauffret, I. Bouissières, and R. Tzanos, 2019: Seamless approach for precipitations within the 0–3 hours forecast-interval. *Third European Nowcasting Conf.*, Madrid, Spain, Agencia Estatal de Meteorología, 1–17, https://hdl.handle.net/20.500.11765/10588.

Muñoz-Sabater, J., and Coauthors, 2021: ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data*, **13**, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021.

Nash, J. E., and J. V. Sutcliffe, 1970: River flow forecasting through conceptual models Part I—A discussion of principles.

J. Hydrol., **10**, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6.

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.

Oudin, L., F. Hervieu, C. Michel, C. Perrin, V. Andréassian, F. Anctil, and C. Loumagne, 2005: Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *J. Hydrol.*, **303**, 290–306, https://doi.org/10.1016/j.jhydrol.2004.08.026.

Panagos, P., 2006: European soil database. *GeoConnexion*, **5**, 32–33.

Pappenberger, F., H. L. Cloke, D. J. Parker, F. Wetterhall, D. S. Richardson, and J. Thielen, 2015a: The monetary benefit of early flood warnings in Europe. *Environ. Sci. Policy*, **51**, 278–291, https://doi.org/10.1016/j.envsci.2015.04.016.

——, M. H. Ramos, H. L. Cloke, F. Wetterhall, L. Alfieri, K. Bogner, A. Mueller, and P. Salamon, 2015b: How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction. *J. Hydrol.*, **522**, 697–713, https://doi.org/10.1016/j.jhydrol.2015.01.024.

Poméon, T., N. Wagner, C. Furusho, S. Kollet, and R. Reinoso-Rondinel, 2020: Performance of a PDE-based hydrologic model in a flash flood modeling framework in sparsely-gauged catchments. *Water*, **12**, 2157, https://doi.org/10.3390/w12082157.

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.

Reinoso-Rondinel, R., M. Rempel, M. Schultze, and S. Trömel, 2022: Nationwide radar-based precipitation nowcasting—A localization filtering approach and its application for Germany. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **15**, 1670–1691, https://doi.org/10.1109/JSTARS.2022.3144342.

Rinehart, R. E., and E. T. Garvey, 1978: Three-dimensional storm motion detection by conventional weather radar. *Nature*, **273**, 287–289, https://doi.org/10.1038/273287a0.

Ruzanski, E., V. Chandrasekar, and Y. Wang, 2011: The CASA nowcasting system. *J. Atmos. Oceanic Technol.*, **28**, 640–655, https://doi.org/10.1175/2011JTECHA1496.1.

Saadi, M., L. Oudin, and P. Ribstein, 2020: Crossing the rural–urban boundary in hydrological modelling: How do conceptual rainfall–runoff models handle the specificities of urbanized catchments? *Hydrol. Processes*, **34**, 3331–3346, https://doi.org/10.1002/hyp.13808.

——, C. Furusho-Percot, A. Belleflamme, J.-Y. Chen, S. Trömel, and S. Kollet, 2023: How uncertain are precipitation and peak flow estimates for the July 2021 flooding event? *Nat. Hazards Earth Syst. Sci.*, **23**, 159–177, https://doi.org/10.5194/nhess-23-159-2023.

Šálek, M., L. Brezková, and P. Novák, 2006: The use of radar in hydrological modeling in the Czech Republic—Case studies of flash floods. *Nat. Hazards Earth Syst. Sci.*, **6**, 229–236, https://doi.org/10.5194/nhess-6-229-2006.

Schaap, M. G., F. J. Leij, and M. Th. van Genuchten, 2001: ROSETTA: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *J. Hydrol.*, **251**, 163–176, https://doi.org/10.1016/S0022-1694(01)00466-8.

Schalge, B., V. Haefliger, S. Kollet, and C. Simmer, 2019: Improvement of surface run-off in the hydrological model ParFlow by a scale-consistent river parameterization. *Hydrol. Processes*, **33**, 2006–2019, https://doi.org/10.1002/hyp.13448.

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.

Sokol, Z., J. Mejsnar, L. Pop, and V. Bližňák, 2017: Probabilistic precipitation nowcasting based on an extrapolation of radar reflectivity and an ensemble approach. *Atmos. Res.*, **194**, 245–257, https://doi.org/10.1016/j.atmosres.2017.05.003.

Speight, L. J., M. D. Cranston, C. J. White, and L. Kelly, 2021: Operational and emerging capabilities for surface water flood forecasting. *Wiley Interdiscip. Rev.: Water*, **8**, e1517, https://doi.org/10.1002/wat2.1517.

UNEP, 1992: *World Atlas of Desertification.* Edward Arnold, 69 pp.

Vivoni, E. R., D. Entekhabi, R. L. Bras, V. Y. Ivanov, M. P. V. Horne, C. Grassotti, and R. N. Hoffman, 2006: Extending the predictability of hydrometeorological flood events using radar rainfall nowcasting. *J. Hydrometeor.*, **7**, 660–677, https://doi.org/10.1175/JHM514.1.

Winterrath, T., and Coauthors, 2018: Radar climatology (RADKLIM) version 2017.002; gridded precipitation data for Germany: Radar-based gauge-adjusted one-hour precipitation sum (RW). Deutscher Wetterdienst, accessed 14 April 2023, https://doi.org/10.5676/DWD/RADKLIM_RW_V2017.002.

Wolfson, M. M., B. E. Forman, R. G. Hallowell, and M. P. Moore, 1999: The growth and decay storm tracker. *8th Conf. on Aviation, Range, and Aerospace Meteorology*, Dallas, TX, Amer. Meteor. Soc., 4.3., https://ams.confex.com/ams/99annual/abstracts/1494.htm.

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.

Xuan, Y., D. Zhu, P. Triballi, and I. Cluckie, 2014: Forecast uncertainty of a lumped hydrological model coupled with the STEPS radar rainfall nowcasts. *Ninth Weather Radar and Hydrology (WRaH) Int. Symp.*, Washington, DC, WMO, https://cronfa.swan.ac.uk/Record/cronfa19214.

Yamazaki, D., D. Ikeshima, J. Sosa, P. D. Bates, G. H. Allen, and T. M. Pavelsky, 2019: MERIT hydro: A high-resolution global hydrography map based on latest topography dataset. *Water Resour. Res.*, **55**, 5053–5073, https://doi.org/10.1029/2019WR024873.