

Hydrological Verification of Two Rainfall Short-Term Forecasting Methods with Floods Anticipation Perspective

MARIA LAURA POLETTI,^a MARTINA LAGASIO,^a ANTONIO PARODI,^a MASSIMO MILELLI,^a VINCENZO MAZZARELLA,^a STEFANO FEDERICO,^c LORENZO CAMPO,^a MARCO FALZACAPPA,^b AND FRANCESCO SILVESTRO^a

^a CIMA Research Foundation, Savona, Italy

^b Italian Civil Protection Presidency of the Council of Ministers, Rome, Italy

^c ISAC CNR, Rome, Italy

(Manuscript received 1 August 2023, in final form 7 February 2024, accepted 8 February 2024)

ABSTRACT: Flood forecasting remains a significant challenge, particularly when dealing with basins characterized by small drainage areas (i.e., 10^3 km² or lower with response time in the range 0.5–10 h) especially because of the rainfall prediction uncertainties. This study aims to investigate the performances of streamflow predictions using two short-term rainfall forecast methods. These methods utilize a combination of a nowcasting extrapolation algorithm and numerical weather predictions by employing a three-dimensional variational assimilation system, and nudging assimilation techniques, meteorological radar, and lightning data that are frequently updated, allowing new forecasts with high temporal frequency (i.e., 1–3 h). A distributed hydrological model is used to convert rainfall forecasts into streamflow prediction. The potential of assimilating radar and lightning data, or radar data alone, is also discussed. A hindcast experiment on two rainy periods in the northwest region of Italy was designed. The selected skill scores were analyzed to assess their degradation with increasing lead time, and the results were further aggregated based on basin dimensions to investigate the catchment integration effect. The findings indicate that both rainfall forecast methods yield good performance, with neither definitively outperforming the other. Furthermore, the results demonstrate that, on average, assimilating both radar and lightning data enhances the performance.

KEYWORDS: Forecasting; Nowcasting; Operational forecasting; Short-range prediction; Hydrologic models

1. Introduction

In the last two decades, significant efforts have been made to predict flash floods and their consequences. These events are characterized by intense rainfall occurring in small geographical areas or subcatchments (Rebora et al. 2013; Braud et al. 2016; Edouard et al. 2018; Faccini et al. 2021; Guastavino et al. 2022). Despite notable advancements in hydrometeorological forecast systems in recent years (Buzzi et al. 2014; Davolio et al. 2015; Lagasio et al. 2019; Ramos Filho et al. 2021; Wang et al. 2021; Lagasio et al. 2022b), flash floods are still widely recognized for their inherent challenges in terms of predicting their location, intensity, and spatiotemporal structure of precipitation.

The idea of using meteorological data, obtained from both ground-based and remote sources, together with various three-dimensional schemes of rainfall prediction systems, has emerged within the hydrometeorological community. There is a consistent focus on enhancing and testing nowcasting radar extrapolation techniques (Panziera et al. 2011; Foresti et al. 2016; Imhoff et al. 2020, 2022), machine learning techniques from ground stations data (Pirone et al. 2023), and assimilation methods in numerical weather prediction models (Lagasio et al. 2019; Gastaldo et al.

2022) for improving results. Recently, Zhang et al. (2023) introduced a method combining physical evolution schemes and conditional learning methods to improve nowcasting skill. However, improving extreme precipitation nowcasts is still a challenge. In general, the prevailing approach is to update forecasts typically focusing on short time ranges frequently. Bowler et al. (2006) developed a short-term rainfall forecast system by combining radar extrapolation techniques and the NWP, assigning appropriate weights based on the extrapolation and NWP forecasts. Hwang et al. (2015) defined a merged rainfall field by using a method that preserves the features from multiple images considering saliency (or importance) as the locations of strong cells. Atencia et al. (2020) presented a technique that allows for keeping the spatial correlation of the errors to merge the forecast from Lagrangian extrapolation and mesoscale models with locally different weights to enhance rainfall predictions, while Franch et al. (2020) explored a deep learning approach to improve extreme rainfall events. The work of Sideris et al. (2020) developed a blending method that combines extrapolation techniques and the NWP using a linear weighting function calibrated based on the performance of NWP. This weighting function varies in time instead of being constant as in simpler approaches.

Many efforts have been made to harness nowcasting and short-term rainfall predictions to enhance hydrological forecasting, yielding promising results. Berenguer et al. (2005) validated a radar-based nowcasting technique in terms of hydrological simulations. Silvestro and Rebora (2012) pursued a similar approach but employed a probabilistic method. Liguori et al. (2012) investigated the feasibility of combining an extrapolation radar rainfall forecast with a high-resolution NWP rainfall forecast to evaluate

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

Corresponding author: Francesco Silvestro, francesco.silvestro@cimafoundation.org

DOI: 10.1175/JHM-D-23-0125.1

© 2024 American Meteorological Society. This published article is licensed under the terms of the default AMS reuse license. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Unauthenticated | Downloaded 02/04/25 02:05 PM UTC

the flow predictions in urban areas. More recently, [Poletti et al. \(2019\)](#) and [Charpentier-Noyer et al. \(2022\)](#) have conducted an analysis on the hydrological response anticipation driven by various rainfall fields products, which include blending methodologies and data assimilation frameworks. These studies have also demonstrated the benefits of utilizing both extrapolation and NWP in a blended rainfall product compared to using extrapolation alone. [Heuvelink et al. \(2020\)](#) employed precipitation nowcasting to predict streamflow in the Netherlands and found that uncertainty increases when catchments dimension decreases. [Lovat et al. \(2022\)](#) presented and discussed the hydrometeorological verification of systems in relation to heavy rainfall events in the Mediterranean environment.

In this work, a further investigation into the aforementioned topics is carried out with the aim of enhancing short-term forecasts and frequently updating streamflow predictions by utilizing recent and state-of-the-art systems. Two methodologies to perform rainfall short-term forecasts are considered. The first methodology is an improved version of the blending method described in [Poletti et al. \(2019\)](#). A more robust assimilation technique is used in the NWP system ([Lagasio et al. 2019](#)), and the blending function used for linear combination is made more flexible and time dependent ([Sideris et al. 2020](#)). The second methodology, described in [Lagasio et al. \(2022a\)](#), also utilizes the same components (NWP and nowcasting method). However, it takes advantage of the overlapping forecast time windows from different NWP runs when applying a frequent assimilation approach.

Eventually, a distributed hydrological model is employed to forecast streamflow and serves as rainfall integrator. This approach addresses the “double penalty” problem (<https://www.ecmwf.int/en/about/media-centre/science-blog/2023/verifying-high-resolution-forecasts>, last accessed 29 January 2024) encountered in many meteorological verification approaches. By utilizing the distributed hydrological model, the results are evaluated in terms of catchment-scale processes.

A hindcast experiment is then conducted, encompassing a significant number of basins (59) and two periods of approximately 10 days each, characterized by a high frequency of events that predominantly affected northwest region of Italy. Due to a combination of factors such as (i) the frequent update of forecasts, (ii) the duration of the hindcast periods, and (iii) the inclusion of multiple basins, a substantial dataset was available for assessing how the set of evaluation scores used to measure performances varies with the forecast lead time. The initial analysis focuses on comparing the potential assimilation of both radar and lightning data versus radar data alone using only the first nowcasting method described in [Poletti et al. \(2019\)](#). Subsequently, a comparison is made between the two nowcasting methods, namely, the one proposed by [Poletti et al. \(2019\)](#) and the method presented in [Lagasio et al. \(2022a\)](#).

One interesting aspect is the fact that the two short-term rainfall forecast systems exploit the comparison of forecast and observed rainfall in order to dynamically adjust their setting and parameterization. The objective of this research is to determine whether one strategy is better than the other when predicting streamflow. Another novel aspect is the investigation of the

effects resulting from the utilization of different datasets in the assimilation process. Both of these topics are approached in terms of streamflow prediction.

2. Materials and methods

a. Study area and data

The study area is located in the northwestern part of Italy and includes watersheds belonging mainly to the Liguria region, as well as areas of southern Piedmont and northwestern Tuscany ([Fig. 1](#)). This region is characterized by a Mediterranean climate, with the main rainy season occurring during autumn. The area is prone to highly intense rainfall events that often result in floods and flash floods, occurring with a certain frequency ([Silvestro et al. 2016](#); [Parodi et al. 2020](#)). The morphology of the study area is characterized by basins ranging in size from approximately 10 to 10⁴ km². These basins exhibit steep slopes, and a significant portion of the territory is covered by forests and brush. The region features mountains having altitudes reaching up to 2000–2500 m. Cities and towns are predominantly situated along the coast or adjacent to the rivers ([Brandolini and Spotorno 1997](#); [Brandolini and Sbardella 2001](#); [Roccati et al. 2018](#); [Faccini et al. 2021](#)). As a result, the catchments' response time typically ranges from 1 to 12 h ([Delrieu et al. 2005](#); [Silvestro et al. 2018](#)). During the early autumn, the temperatures along the coast commonly range between 20° and 25°C, which gradually drop to 8°–12°C in the late autumn. Generally, the snow accumulations are negligible in this season, except for some small areas situated at altitudes higher than 2000 m.

Radar data used in this study are obtained from the Italian National Radar Network ([Vulpiani et al. 2012, 2016](#)) and provided by National Civil Protection (<https://dpc-radar.readthedocs.io/it/latest/>, last accessed 26 January 2024). The following products are used:

- 1) surface radar intensity (SRI): this product serves as input to the nowcasting system (frequency of use 10 min). The National Civil Protection delivers an adjusted version of SRI (SRIadj), which incorporates rain gauges observations. The SRIadj is the product used in the experiments presented.
- 2) constant-altitude plan position indicator (CAPPI) of radar reflectivity (CAPPIZ): CAPPIZ at three different heights (2000, 3000, 5000 m) is used in the data assimilation process every 3 h for the three-dimensional variational assimilation system (3DVAR) application.

Data obtained from the Italian National Meteorological Network are utilized to generate the required inputs for streamflow simulations. Hourly temperature, air relative humidity, solar radiation, and wind velocity data are employed after undergoing spatial interpolation ([DPC 2022](#)). Hourly rainfall data, along with radar data, are combined using a merging algorithm derived from [Sinclair and Pegram \(2005\)](#) and described in [Bruno et al. \(2021\)](#).

Lightning data, along with radar CAPPI, are utilized for assimilation purposes. They are assimilated using a 3DVAR

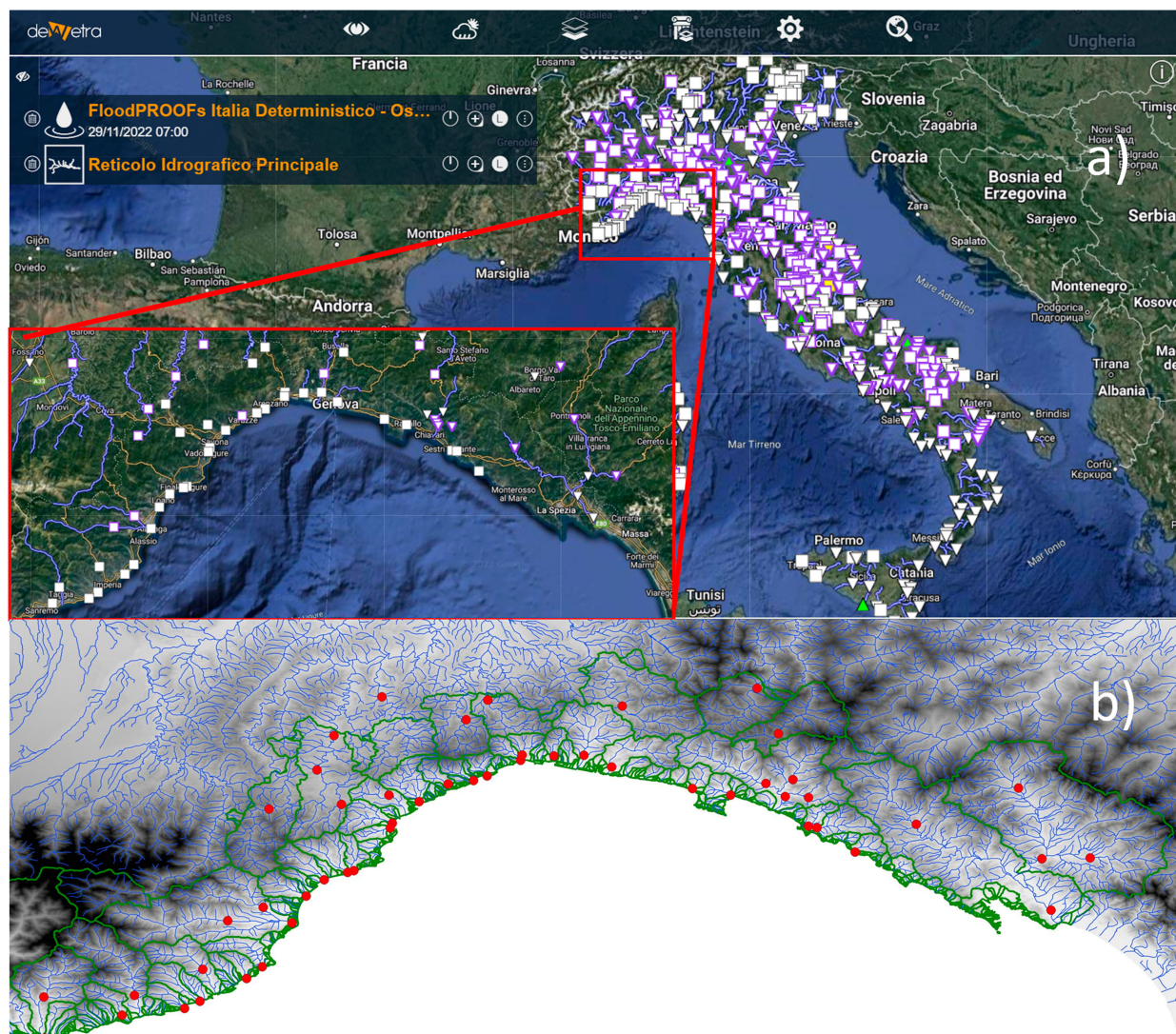


FIG. 1. (a) Hydrological streamflow simulation system operationally implemented at Italian National Scale available on the Dewetra Early Warning Support system. Markers with violet borders are the hydrological sections where streamflow observations are available. Squares represent sections where streamflow has no trend, and triangles represent the ones where trend is increasing or decreasing (trend is estimated on a time window of 12 h centered on the time of most recent run of the model). The zoom shows the area of interest for the analysis. (b) Higher detail of the 59 sections involved in the analysis (red markers), the hydrographic network (blue lines), and the main basins (green polygons). The digital elevation model shown in grayscale highlights the morphology.

and a nudging technique at the frequency outlined in section 2d. For this purpose, SFLOC (term derived from the words spheric and location) data of the lightning flashes are obtained from the LAMPINET Aeronautic Meteorological Service, which is based on Vaisala technology (De Leonibus et al. 2010). The LAMPINET network consists of 15 sensors with improved accuracy from combined technology, offering enhanced sensitivity and performance, uniformly distributed across the Italian national territory (De Leonibus et al. 2010). The network employs both MDF (magnetic direction finding) and TOA (time of arrival) techniques (De Leonibus et al. 2008). Network demonstrates a detection efficiency of 90% for peak currents exceeding 50 kA with a location accuracy of about 500 m throughout Italy. The

sensors used are of the IMPACT-ESP (Improved Accuracy through Combined Technology Enhanced Sensitivity and Performance Sensor) type. The sensors detect radio frequency energy from both cloud-to-ground (CG) and intra-cloud (IC) discharges within a bandwidth ranging from 1 to 350 kHz. In addition to location accuracy, the sensors also measure azimuth angle, time of signal arrival, peak signal strength, and discharge width.

b. Streamflow short-term forecast framework

The short-term forecast framework is made by a cascade of models that use observations as input or for their initialization:

- 1) an NWP with frequent cycles of the 3DVAR assimilation process, described in [section 2d](#);
- 2) a nowcasting algorithm based on rainfall radar fields, described in [section 2e](#);
- 3) two algorithms for combining rainfall forecasts derived by point 1 and point 2, described in [sections 2f](#) and [2g](#); and
- 4) a hydrological model used to transform rainfall forecasts into streamflow forecasts, described in [section 2c](#).

For a general scheme one can refer to Fig. 5 of [Poletti et al. \(2019\)](#).

c. Hydrological model

The continuous and distributed hydrological model employed in this experiment is extensively described in [Silvestro et al. \(2013, 2015, 2021\)](#) and is referred to as Continuum. The specific configuration adopted in this study is widely documented in [Davalio et al. \(2017\)](#) and [Silvestro et al. \(2018\)](#), which also describe the calibration strategy. Furthermore, [Bruno et al. \(2021\)](#) presents the implementation of the hydrometeorological system at the Italian National Scale for civil protection purposes. The model utilized in this work has a spatial resolution of 0.05° (around 480 m) with a time resolution of 1 h. The system is implemented nationwide and incorporates over 350 control sections ([Fig. 1](#)). Streamflow time series from these control sections are utilized to monitor and forecast the intense rainfall events impacts. The results are published on myDewetra, one of the official Early Warning Support Systems employed by Italian Civil Protection ([Pagliara et al. 2011](#); https://www.mydewetra.org/wiki/index.php/FloodPROOFs_Italia_Deterministico_-_Osservazioni).

As described in [section 3](#), the analysis periods under consideration primarily focus on the northwest region of Italy. Consequently, 59 sections within this area are selected to conduct the experiments and the analysis. The drainage areas of these selected sections range from 10 to 10^3 km². To initialize the state variables of the Continuum model at the beginning of the two analysis periods, a long-term run is performed using observations as input.

d. WRF and assimilation approach

The Weather Research and Forecasting (WRF) Model v3.8.1 ([Skamarock et al. 2008](#)) was chosen as the numerical weather model for this study. The WRF Model is a fully compressible nonhydrostatic model that utilizes mass-based terrain-following coordinates. It was developed by the National Center for Atmospheric Research (NCAR) in collaboration with several institutes and universities. The model serves both for operational weather forecasting and research purposes. For this work, three two-way nested domains ([Fig. 2](#)) were selected. The horizontal grid spacing for each domain ([Fig. 2](#)) is 22.5 km (216×191 grid points), 7.5 km (523×448), and 2.5 km (430×469). All domains consist of 50 vertical levels and extend up to 50 hPa.

All the simulations in this study employ the same set of physical parameterizations which have been previously validated and proven successful in other studies focusing on extreme events in Italy ([Fiori et al. 2017](#); [Lagasio et al. 2017](#),

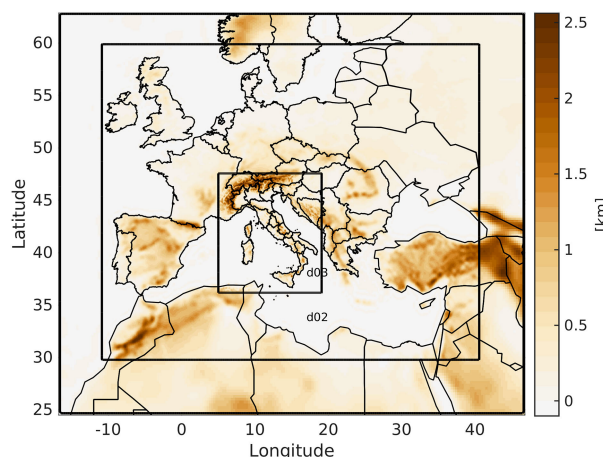


FIG. 2. WRF Model configuration. The three boxes represent the horizontal grid spacing domains: 22.5 km (external domain), 7.5 km, and 2.5 km (smaller domain that includes mainly Italy).

[2019, 2022b](#); [Mazzarella et al. 2022](#)). The following parameterizations are adopted:

- Surface layer: the MM5 scheme is adopted ([Paulson 1970](#); [Dyer and Hicks 1970](#); [Webb 1970](#)).
- Surface fluxes: a parameterization scheme following [Beljaars \(1995\)](#) is used to enhance surface fluxes of heat and moisture.
- Soil model: the Rapid Update Cycle (RUC) scheme is used as a multilevel soil model (6 levels), with higher-resolution levels distributed in the upper soil parts (0, 5, 20, 40, 160, and 300 cm). This soil model solves the heat diffusion and Richards moisture transfer equations (with a layer approach) and, in the cold season, considers phase changes in soil water ([Smirnova et al. 1997, 2000](#)).
- Planetary boundary layer (PBL): The dynamics are parameterized with the diagnostic nonlocal Yonsei University (YSU) PBL scheme ([Hong et al. 2006](#)).
- Microphysics: The WSM6 six-class scheme is adopted ([Hong and Lim 2006](#)).
- Radiative processes: They are parameterized by means of the longwave and shortwave RRTMG schemes ([Iacono et al. 2008](#)).
- Convective parameterization: the new simplified Arakawa–Schubert scheme ([Han and Pan 2011](#)) is adopted, except in the inner domain where convection is explicitly resolved.

1) ASSIMILATING RADAR REFLECTIVITY DATA

The forecasts with data assimilation (WRF-DA setup) were conducted using a 3-h cycle 3DVAR technique of radar reflectivity observations as indicated in [Fig. 3](#). The assimilation is denoted as “DA hh REFL,” where hh represents the assimilation time instant. The 3DVAR assimilation purpose is to minimize the difference between background forecast and the observations by optimizing a cost function ([Ide et al. 1997](#)). The WRF data assimilation package WRFDA ([Barker et al. 2012](#)) v3.9.1 is employed. In this study, the reflectivity modified direct operator as presented in [Lagasio et al. \(2019\)](#),

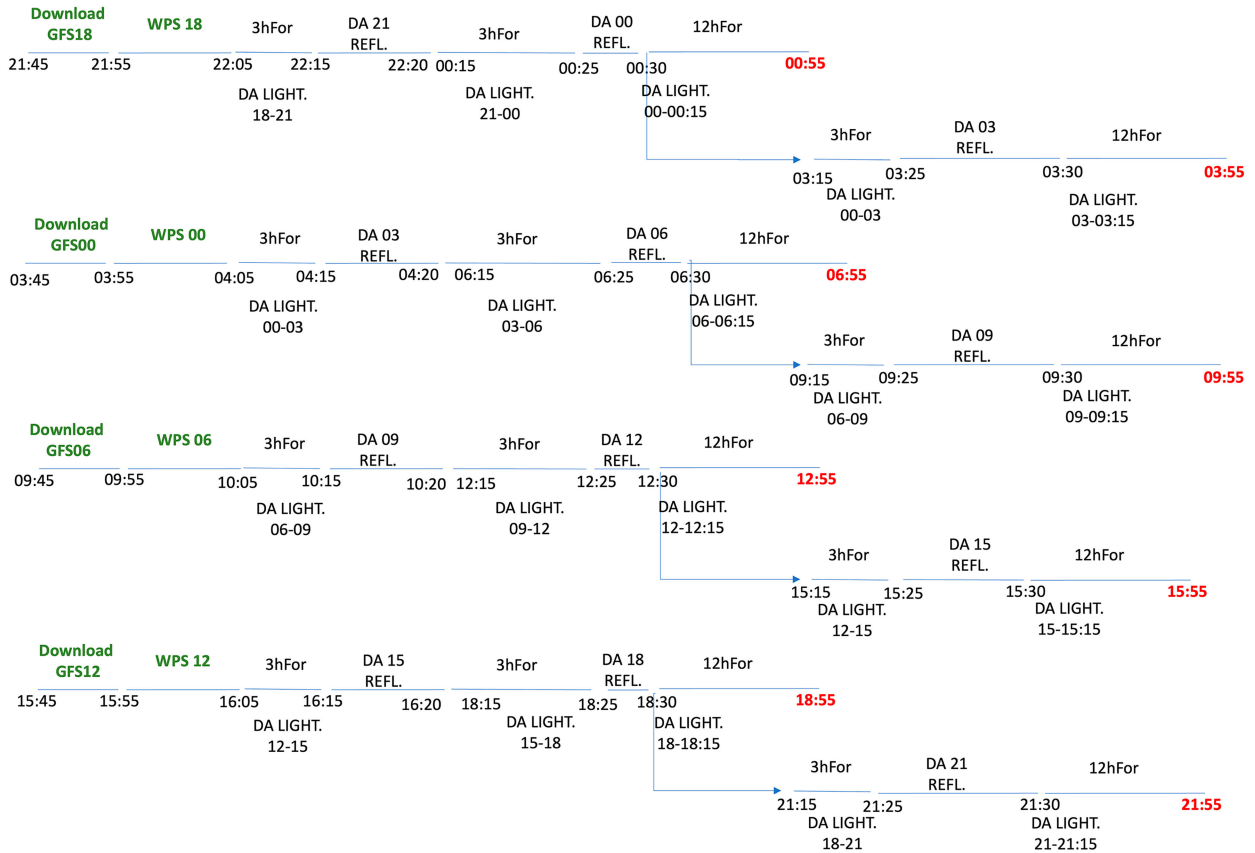


FIG. 3. Assimilation timing scheme for radar and lightning data.

is used. The control variable option selected is 5 (CV5), which refers to the **B** matrix (background error covariance matrix) calculation using the National Meteorological Center (NMC) method (Wang et al. 2014).

The NMC method was applied over the entire month of October 2015, with a 24-h lead time for the forecasts starting at 0000 UTC and a 12-h lead time for the ones initialized at 1200 UTC on the same day. The differences between the two forecasts ($t + 24$ and $t + 12$) valid for the same reference time were utilized to calculate the domain-specific error statistics.

2) ASSIMILATING LIGHTNING DATA

The method proposed by Fierro et al. (2012) is utilized for lightning data assimilation (LDA). This method involves nudging a water vapor mixing ratio profile where and when lightning is observed. The nudged water vapor mixing ratio profile, denoted as q_v , is computed using the following equation:

$$q_v = Aq_s + B \tanh(CX)[1 - \tanh(Dq_g^\alpha)]. \quad (1)$$

The coefficients in the equation are set as follows: $A = 0.86$, $B = 0.15$, $C = 0.30$, $D = 0.25$, $\alpha = 2.2$. Here, q_s represents the saturation mixing ratio at the model atmospheric temperature, and q_g denotes the graupel mixing ratio (g kg^{-1}). The

mixing ratio described by Eq. (1) increases with the flash rate (X), which indicates the number of flashes accumulated in a grid box over a 15-min period. Conversely, it decreases with the graupel mixing ratio.

The mixing ratio q_v described in Eq. (1) is computed at each LDA time step for grid points where flashes are observed, indicated by X greater than zero, within the layer between the 0° and -25°C isotherms. The computed mixing ratio from Eq. (1) is then compared with the predicted mixing ratio by the model. If the computed mixing ratio of Eq. (1) is larger than the simulated one, the modeled mixing ratio is replaced with the value obtained from Eq. (1). On the other hand, if the modeled mixing ratio is already larger, it remains unchanged. It is important to note that this method can only introduce additional water vapor to the forecast.

In this study, the lightning assimilation is conducted according to the timing presented in Fig. 3, denoted as “DA LIGHT.hh1-hh2,” represents the assimilation time coverage, starting from hh1 and extending until hh2.

e. PhaSt nowcasting algorithm

PhaSt is a probabilistic nowcasting algorithm that utilizes a spectral approach. It takes two consecutive rainfall fields retrieved from radar as input and evolves them within the spectral space through a nonlinear transformation. A detailed

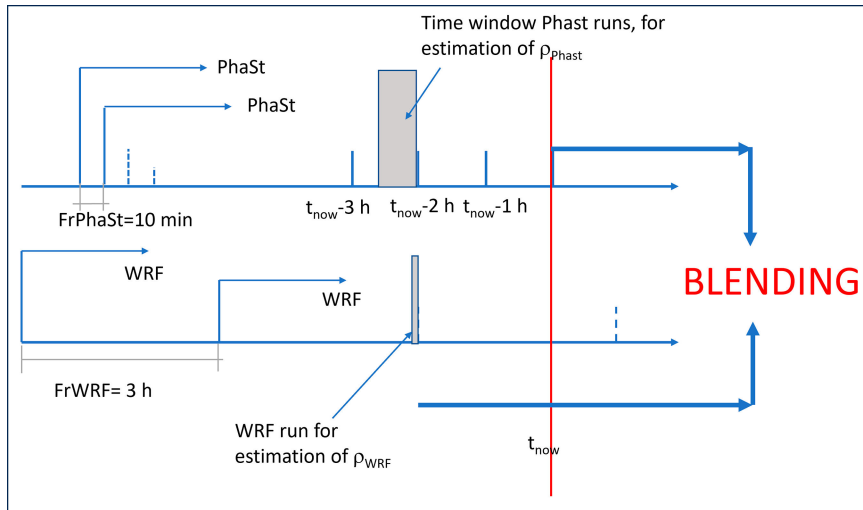


FIG. 4. Blending framework scheme and forecasts set used to estimate the correlations between WRF, PhaSt, and observations. PhaSt is updated every 10 min, WRF with data assimilation is updated every 3 h, and blending is updated every hour.

algorithm description can be found in [Metta et al. \(2009\)](#) and [Silvestro et al. \(2015\)](#), while [Poletti et al. \(2019\)](#) provides information on the specific algorithm version used as the starting point in this research. Here, we explore a deterministic approach, resulting in the removal of the probabilistic component. Consequently, the set of equations used in this research is as follows:

$$d\Phi_{k_s} = \omega_{k_s} dt, \quad (2)$$

$$d\omega_{k_s} = -(\omega_{k_s} - \omega') \frac{dt}{T}, \quad (3)$$

where Φ_{k_s} represents the spectral phase for a particular angular velocity k_s , which is dependent on the wavenumbers k_x and k_y ; ω_{k_s} denotes the angular frequency for k_s . Each wavenumber moves with its corresponding spectral phase; ω' represents the initial angular frequency estimated from the last two observed fields, and T is a parameter defined as the decorrelation time. It is a time scale of relaxation to the initial angular frequency.

To overcome the method limitations, which maintain constant precipitation volumes based on the most recent observation, the rainfall volume correction method described in [Poletti et al. \(2019\)](#) is implemented in this study. This correction method involves determining the percentage increasing or decreasing of the precipitation total volume for each time step relative to the time of the most recent radar observation available, using information derived from NWP. This percentage adjustment is then applied to the PhaSt nowcasted fields (refer to Fig. 3 in [Poletti et al. 2019](#)). This approach allows us to utilize information derived by the NWP but while still leveraging the predictive capabilities of the extrapolation techniques, particularly during the first 1–3 h of forecast.

f. Blending approach

The blending strategy proposed in this study falls under the category of linear combinations of NWP and nowcasting

rainfall fields ([Hwang et al. 2015](#)). Building upon the methodology presented in [Poletti et al. \(2019\)](#), the technique enhanced by incorporating the approach utilized by [Sideris et al. \(2020\)](#). This approach involves utilizing the comparison between models and observations to parameterize the blending function. As a result, the nowcasting and the most recent NWP run are dynamically weighted based on their performance, allowing for adaptive blending.

The weight expression for the nowcasting rainfall fields (PhaSt) as a function of lead time t is defined as follows:

$$w_{\text{PhaSt}}(t) = 1, \text{ for } t \leq 1 \text{ h}, \quad (4)$$

$$w_{\text{PhaSt}}(t) = \exp[-a(t-1)^{2.5}], \text{ for } t > 1 \text{ h}. \quad (5)$$

While the NWP (WRF) weight is calculated as

$$w_{\text{WRF}}(t) = 1 - w_{\text{PhaSt}}(t). \quad (6)$$

The final rainfall R (mm) field is obtained by combining the PhaSt and WRF rainfall fields as follows:

$$R(x, y, t) = R_{\text{PhaSt}}(x, y, t)w_{\text{PhaSt}}(t) + R_{\text{WRF}}(x, y, t)w_{\text{WRF}}(t). \quad (7)$$

Inspired by [Sideris et al. \(2020\)](#), the parameter a (—) of Eq. (5) is estimated based on the correlation between modeled and observed rainfall fields:

$$a = (1 - \rho_{\text{PhaSt}-p})^k. \quad (8)$$

Here, k is set to 1 for simplicity, while $\rho_{\text{PhaSt}-p}$, the relative correlation coefficient for the nowcasting, is calculated as

$$\rho_{\text{PhaSt}-p} = \frac{\rho_{\text{PhaSt}}}{\rho_{\text{PhaSt}} + \rho_{\text{WRF}}}. \quad (9)$$

The two correlations ρ_{PhaSt} and ρ_{WRF} are calculated by comparing the accumulated rainfall fields after the first hour of

TABLE 1. Basin classes defined basing on drainage area together with an approximated indication of response time.

Basin class	Drainage area (km ²)	t_r (h)
Class 1	Area ≤ 50	≤ 2
Class 2	$50 < \text{Area} < 150$	$2 < t_r < 4$
Class 3	Area ≥ 150	$t_r \geq 4$

forecast (Sideris et al. 2020). For shorter lead times within the range of 0–1 h, the nowcasting is considered to be generally better on average (Metta et al. 2009), and the NWP contribution is determined by the precipitation volume trend (Poletti et al. 2019). The correlation ρ_{WRF} is calculated based on a 1-h accumulation with lead time of +2 and +3 h, relative to the NWP running time t_r . To simulate the operational conditions, let t_{now} represent the time when the blending procedure is applied. The correlations are estimated to the closest time t_r , which can be $t_r = t_{\text{now}} - 2$ h or $t_r = t_{\text{now}} - 3$ h. If both correlations can be estimated for both time instants, an average value is considered to estimate ρ_{WRF} .

A similar approach is applied to estimate ρ_{Phast} , but since PhaSt can be run with higher frequency (FrPhaSt = 10 min), the correlation is estimated for all the predictions made in the time interval from $t_r = t_{\text{now}} - 2.5$ h to $t_r = t_{\text{now}} - 2$ h. The correlation ρ_{Phast} is then estimated as average of those values.

In this analysis, the blended final rainfall field is generated with an update frequency of 1 h. Figure 4 provides a schematic representation of the blending framework. In the presented study, NWP is updated every 3 h, but nowcasting can be updated with high frequency as well as the a parameter estimation.

g. SWING method

The postprocessing algorithm SWING (score-weighted improved nowcasting; Lagasio et al. 2022a) is developed to account for the temporal and spatial uncertainty in simulating convective fields. SWING is employed considering the three forecasts that span the same 6-h time window, repeated every 3 h (refer to Fig. 3 in Lagasio et al. 2022a). It combines the PhaSt output and the WRF Model simulations to generate an updated forecast every 3 h. However, in this particular case, SWING is not used to produce the rainfall hazard scenarios presented in Lagasio et al. (2022b). Instead, it provides hourly weighted rainfall every 3 h obtained from the three simulations that cover the same 6-h period of forecast (Fig. 3 in Lagasio et al. 2022a) to provide a useful input to the hydrological model. The forecast weighing is determined by evaluating the simulation performance during the 3 h preceding the time of interest using an object-based approach. This evaluation involves comparing the modeled rainfall objects with the observed one obtained by the radar and rain gauges rainfall data on the national map (Bruno et al. 2021). The rainfall objects are identified by isolating the regions with rainfall values exceeding 2 mm. The comparison between the observed and modeled rainfall objects involves evaluating various objects' characteristics, such as rainfall volumes, areas extension, distance between center of gravity, objects' orientation, and objects' intersection. These

characteristics are used to calculate scores and determine the correspondence between the observed and modeled objects. The scores are then summarized into an overall field score (OFS; Lagasio et al. 2022a). In addition to the OFS, a missed object score (MOS) is calculated to account for observed objects that are not present in the model. The MOS ranges from 0 to 1 (Lagasio et al. 2022a). These scores, along with the OFS, are combined to obtain a final reliability score (RS; Lagasio et al. 2022a). The RS score is used to weight the three considered forecasts, resulting in the mean rainfall fields that can be utilized as input for the Continuum hydrological model.

Basin class	$Q_f \geq Q_t$	$Q_f < Q_t$
$Q_b \geq Q_t$	Hit (A)	Miss (B)
$Q_b < Q_t$	False alarm (C)	Correct negative (D)

characteristics are used to calculate scores and determine the correspondence between the observed and modeled objects. The scores are then summarized into an overall field score (OFS; Lagasio et al. 2022a). In addition to the OFS, a missed object score (MOS) is calculated to account for observed objects that are not present in the model. The MOS ranges from 0 to 1 (Lagasio et al. 2022a). These scores, along with the OFS, are combined to obtain a final reliability score (RS; Lagasio et al. 2022a). The RS score is used to weight the three considered forecasts, resulting in the mean rainfall fields that can be utilized as input for the Continuum hydrological model.

3. Hindcast experiment and hydrological verification

a. The experiment

The hindcast experiment involved simulating the entire forecast system for two distinct periods. The first period spanned from 14 to 22 October 2019, while the second period encompassed 14–21 November 2019 [refer to Lagasio et al. (2022a) for detailed information on the two periods]. For sake of simplicity, they will be named October and November. According to official event analysis documents (<https://www.arpal.liguria.it/tematiche/meteo/publicazioni-bis/rapporti-di-eventi-significativi/rapportieventisignificativi-2019.html>), both periods were characterized by nearly continuous rainy weather and various flood events. During the October period, the events were mainly concentrated in small areas, with local rainfall accumulations of up to 100–120 mm in 1 h and 250–350 mm in 6 h. On the other hand, the events in November were more spatially distributed (Molini et al. 2009) with significant rainfall accumulations also on larger areas (50–100 mm in 24 h and 10³ km²), as well as local storms with hourly accumulation of 40–60 mm and local accumulations of 300–400 mm in 24–36 h.

Each period consists of approximately 7–8 days, and the two presented methods have different update frequencies:

TABLE 3. Number of forecast samples used for building the graphs in Figs. 5–7 for the blending analysis.

Basin class	October	November
Class 1	4368	4316
Class 2	2856	2822
Class 3	2688	2490

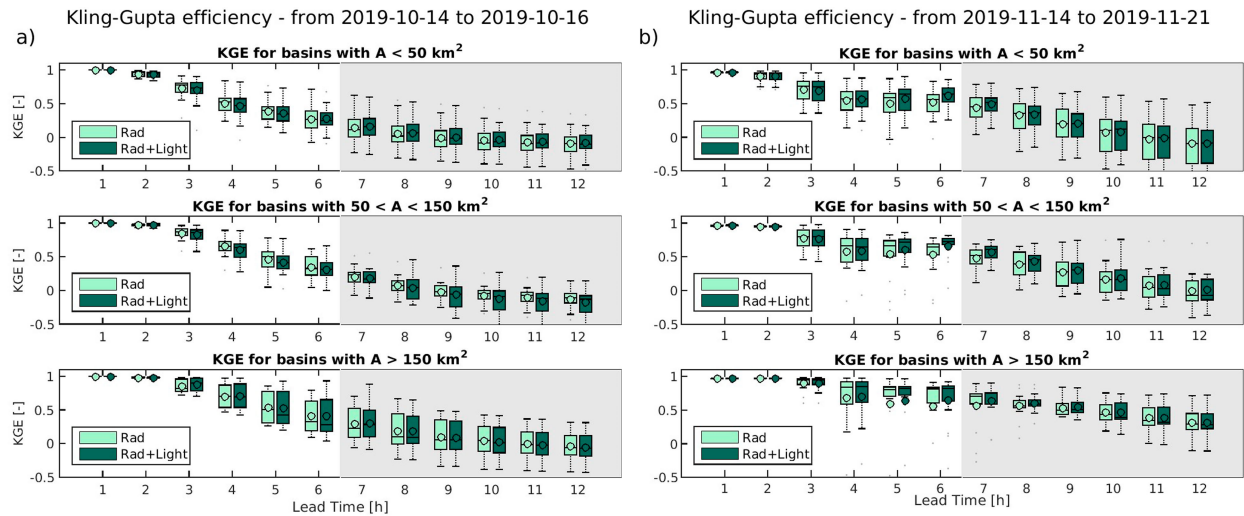


FIG. 5. KGE scores as a forecast lead time function for the blending method for the October period. The x axis reports lead time, and the y axis reports the boxplot of score values and the average value (dots). Rad indicates the radar data assimilation, while Rad+Light is the assimilation of both radar and lightning data. The “A” is the drainage area of the basins. The light gray area indicates that from the seventh hour rainfall input is not available and only the effects of hydrological model routing are shown.

- The blending-driven method, which utilizes the most recent WRF run, is updated every hour.
- The SWING-driven method is updated every 3 h.

The maximum lead time of rainfall forecast is 6 h. However, for hydrological analysis, from hour 7 to 12 zero rainfall is considered. This is because the larger basins involved in the experiment have drainage area on the order of 10^3 km², with concentration times around 10 h.

Therefore, a large number of forecasts is available to conduct a robust analysis. However, for the blending approach, there is a larger number of samples compared to the SWING approach. As a result, an initial investigation is conducted to compare the simulation using both radar and lightning assimilation or radar-only assimilation using all the available blending

forecasts. The comparison with the SWING method is performed only when both methods are available. Each method is applied using the two assimilation configurations: the first one uses radar data only, while the second uses both radar and lightning data.

The benchmark used for the comparison is the hydrological model run using observed meteorological variables as input. This approach, which has been employed in previous studies (Borga 2002; Vieux and Bedient 2004), allows for the inclusion of multiple basins in the analysis while also eliminating the potential effects of hydrological model calibration errors. To evaluate the performance scores for different forecast lead times, the method described in Berenguer et al. (2005) is utilized.

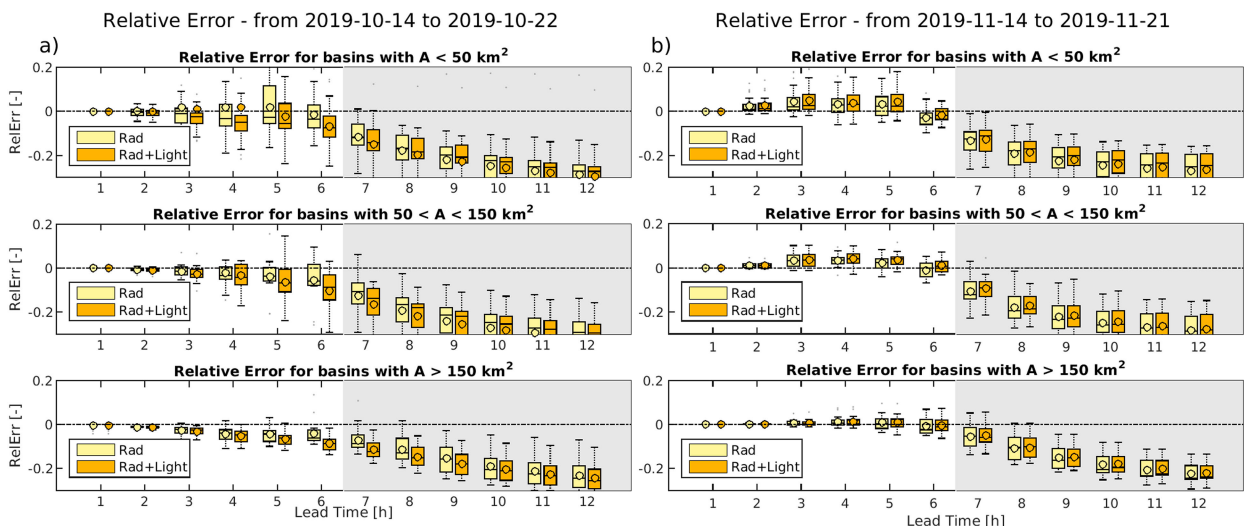
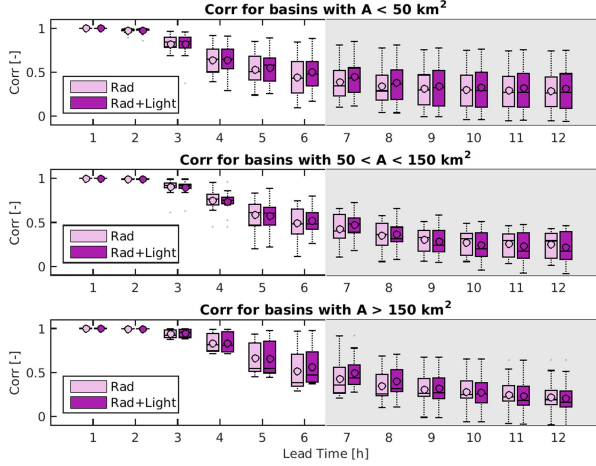
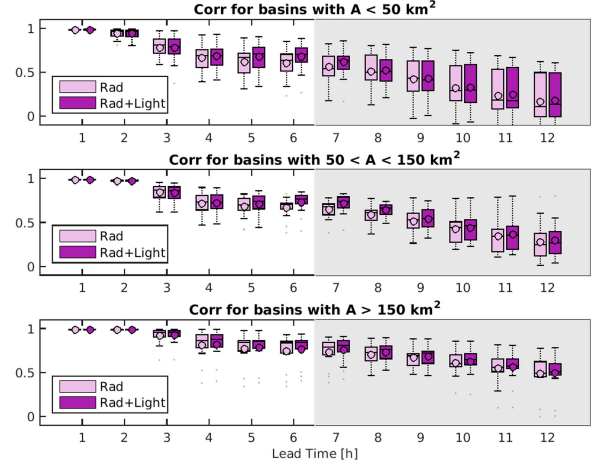


FIG. 6. RelErr score as a forecast lead time function for the blending method.

a) Correlation coefficient - from 2019-10-14 to 2019-10-16



b) Correlation coefficient - from 2019-11-14 to 2019-11-21

FIG. 7. Correlation coefficient ρ score as a forecast lead time function for the blending method.

b. Performance indicators

Three scores that measure the goodness-of-fit are considered in the analysis. One is the Kling–Gupta efficiency (KGE),

$$\text{KGE} = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (1 - \beta)^2}, \quad (10)$$

where r is the linear correlation between benchmark and forecast, α is a measure of the flow variability error, and β is a bias term (Knoben et al. 2019; Sofokleous et al. 2022):

$$\alpha = \frac{Q_{fm}}{Q_{bm}}, \quad (11)$$

$$\beta = \frac{\sigma_f}{\sigma_b}, \quad (12)$$

where Q_{fm} is the average forecast streamflow and Q_{bm} is the average benchmark streamflow (i.e., streamflow simulated with observed variables). The σ_f and σ_b are the standard

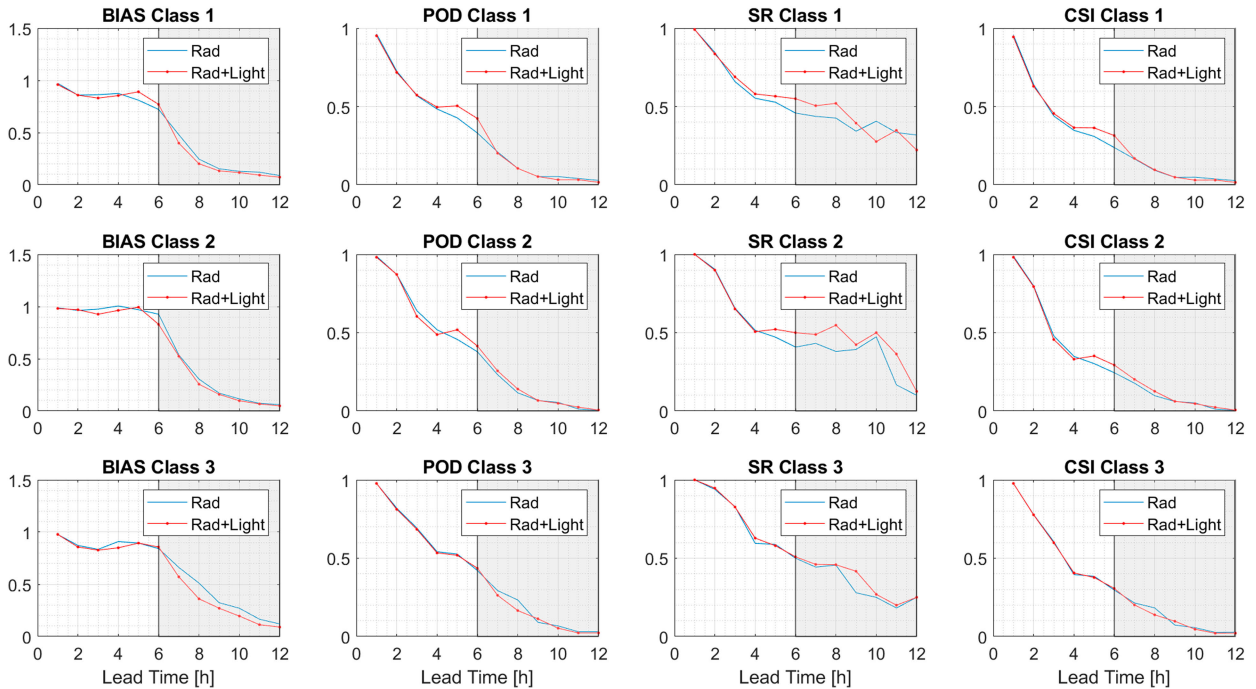


FIG. 8. Critical success index (CSI), BIAS score, success ratio (SR), and probability of detection (POD) as a function of forecast lead time. Results are shown for each basin class. Rad indicates the radar data assimilation, while Rad+Light is the assimilation of both radar and lightning data.

TABLE 4. Number of forecast samples used for building the graphs in Figs. 9–14 for the blending–SWING comparison.

Basin class	October	November
Class 1	1456	702
Class 2	952	459
Class 3	896	405

deviations of the forecast and benchmark. KGE varies in the range $[-\infty, 1]$.

The relative error is

$$\text{RelErr} = \frac{1}{n} \sum_{i=1}^n \frac{(Q_{if} - Q_{ib})}{Q_{ib}}, \quad (13)$$

where Q_{if} is the forecast streamflow, Q_{ib} is the benchmark, and n is the length of simulation in terms of time steps number.

The Pearson correlation coefficient is

$$\text{Corr} = \frac{\sum_{i=1}^n (Q_{if} - Q_{fm})(Q_{ib} - Q_{bm})}{\sqrt{\sum_{i=1}^n (Q_{if} - Q_{fm})^2} \sqrt{\sum_{i=1}^n (Q_{ib} - Q_{bm})^2}}. \quad (14)$$

To account for variations in the response time to rainfall and the different ability of catchments to integrate rainfall, three classes of basins were considered. These classes were defined based on the drainage area, following the approach outlined

in Poletti et al. (2019). This pooling strategy was employed to ensure a sufficiently large sample size while also distinguishing the basins according to their size, thereby avoiding the masking of significant heterogeneity in the verification results.

Table 1 provides an overview of the three basin classes, along with a rough estimation of the range of concentration times for each class.

A further analysis was carried out to assess the capability of predicting a predefined streamflow threshold, following an approach similar to Mason (1982) and Charpentier-Noyer et al. (2022). A contingency table (Table 2) is used, and four scores are considered (CAWCR 2015): critical success index (CSI), probability of detection (POD), success ratio (SR), and bias score (BIAS). Analysis is performed on the blending-driven method, where the sample of data is larger, and data from both periods are considered together. The chosen threshold is the 70th percentile of each benchmark streamflow time series, allowing for the selection of the main occurring peak events.

The score formulas (referred to Table 2) are reported here, together with the range of variation:

$$\text{CSI} = \frac{A}{A + B + C}, \quad [0, 1] \quad (\text{perfect score : } 1), \quad (15)$$

$$\text{POD} = \frac{A}{A + B}, \quad [0, 1] \quad (\text{perfect score : } 1), \quad (16)$$

$$\text{BIAS} = \frac{A + C}{A + B}, \quad [0, \infty] \quad (\text{perfect score : } 1), \quad (17)$$

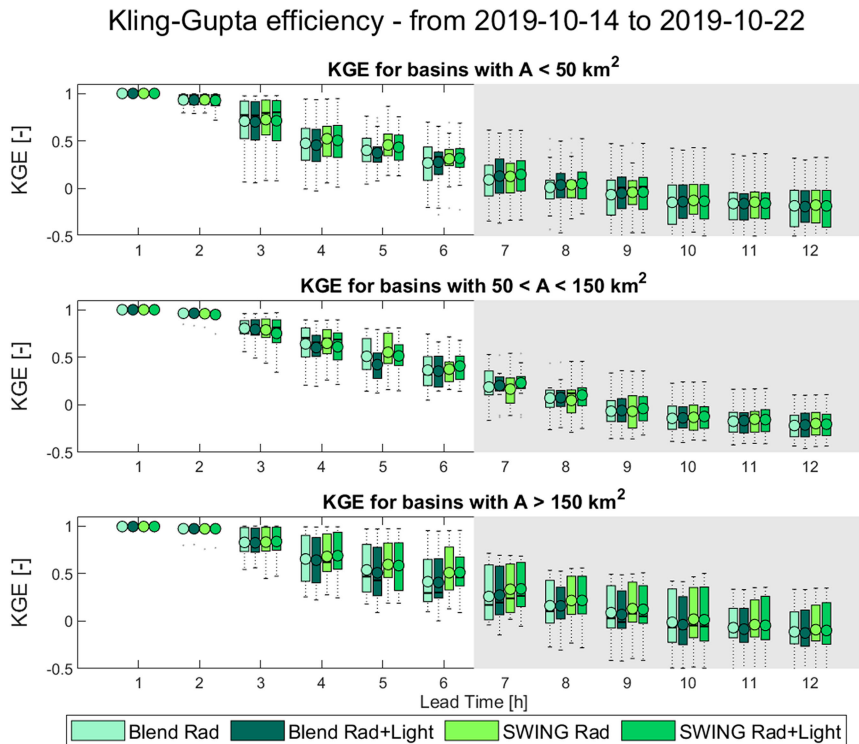


FIG. 9. Comparison between the blending and SWING methods. KGE scores as a forecast lead time function for the October period.

Kling-Gupta efficiency - from 2019-11-14 to 2019-11-21

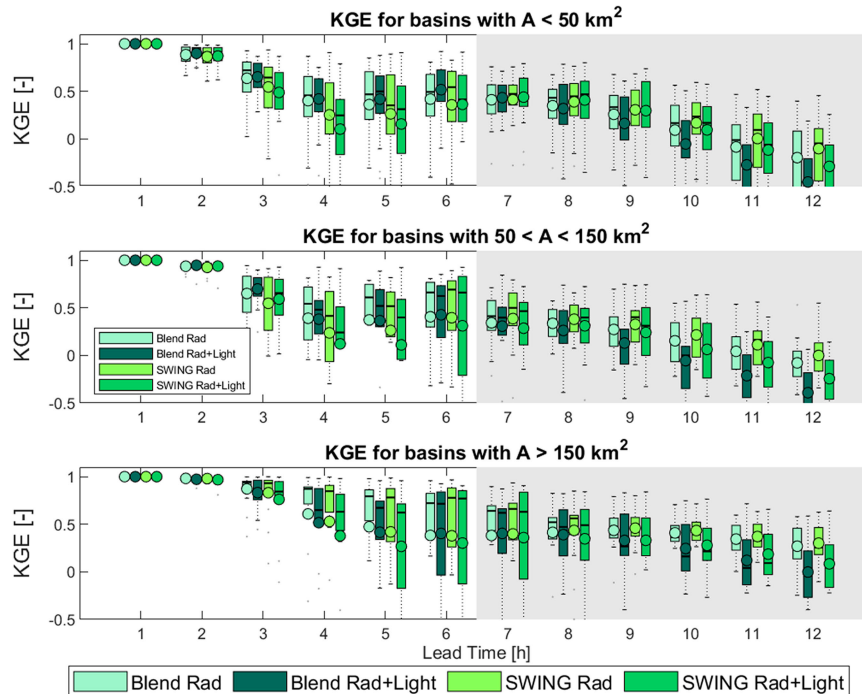


FIG. 10. As in Fig. 9, but for the November period.

$$SR = \frac{A}{A + C}, \quad [0, 1] \quad (\text{perfect score : 1}). \quad (18)$$

4. Results

a. Blending system (with 1-h updating)

The first verification is conducted exclusively using the blending system only with both assimilation methods in WRF. This method allows for updates every hour, resulting in a large sample of forecasts that contributes to robust statistical analysis.

To assess the performance, we constructed graphs depicting the forecast lead time on the x axis and the scores' boxplot on the y axis. The boxplots present the interquartile range (25th and 75th percentiles) while the whiskers provide insight into the variability of the results relative to the median, and the mean is represented by the dot. These visualizations aid in understanding the distribution and variability of the scores across different forecast lead times.

The predicted rain impact on streamflow varies according to the basin size, which serves as an input integrator. Table 3 provides the number of samples used to construct the graphs for each basin class, categorized by the two periods. The number of samples depends on the sections involved and on the number of forecasts available.

Figure 5 shows the KGE for the two periods, October and November 2019. As suggested by Knoben et al. (2019), a KGE value of $1 - \sqrt{2}$ (about -0.41) indicates that the mean flow performs better than the simulation as a predictor. Therefore, the y -axis limits were set accordingly. The label “Rad”

represents assimilation of radar only, while “Rad+Light” indicates assimilation of both radar and lightning data. All plots have maximum value on the x axis larger than the rainfall forecast time window of 6 h, but a differentiation in the background of the graphs is introduced (light gray area) in order to indicate that from the seventh hour rainfall input is not available and only the effects of hydrological model routing are shown.

Both periods exhibit satisfactory KGE values, with the KGE remaining largely over 0 in both assimilation configurations within the 6-h rainfall forecast time window. This holds true for all the three classes of basins, although the performance tends to decrease at a slower rate with increasing lead time as the drainage area increases. Additionally, it is worth noting that events during the November period are generally more predictable compared to those in October.

During the October period, the KGE performances are quite similar for both assimilation systems. For drainage areas in the range 50–150 km², Rad performs better for lead times between 3 and 6 h, while for areas larger than 150 km², KGE is slightly better when Rad+Light is considered. In the November period, the Rad+Light configuration generally performs better for lead times ranging from 4 to 8 h. This variation in performance could be attributed to the fact that, depending on the affected area and lead time, the rainfall field produced with the Rad configuration may be better than Rad+Light in reproducing observed rainfall, and vice versa.

In both periods, there are no noticeable differences in the first 2 h of forecast, where the combination of nowcasting and routing of observed rainfall plays a significant role.