Weather and Climate Extremes 50 (2025) 100839

Contents lists available at ScienceDirect

Weather and Climate Extremes

journal homepage: www.elsevier.com/locate/wace

Improving extreme precipitation forecasts in Catalonia (NE Iberian Peninsula) using analog methods: A comparison with the GFS model

Carlo Guzzon [a],*, Raül Marcos-Matamoros [a,b], Maria Carmen Llasat [a,c], Montserrat Llasat-Botija [a,d]

[a] Department of Applied Physics, Universitat de Barcelona, Martí Franqués 1, 08028, Barcelona, Spain
[b] Earth Sciences Department, Barcelona Supercomputing Center - CNS, Pl. d'Eusebi Güell, 1-3, Barcelona 08034, Spain
[c] UBICS, Universitat de Barcelona, Martí Franqués 1, 08028, Barcelona, Spain
[d] Water Research Institute, Universitat de Barcelona, Martí Franqués 1, 08028, Barcelona, Spain

ARTICLE INFO

Dataset link: https://github.com/cguzzon1998/analog_algorithm.git, https://www.ncei.noaa.gov/products/weather-climate-models/global-forecast, https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=download

Keywords:
Analog methods
Extreme precipitation
Flash flood
Numerical models

ABSTRACT

Flood forecasting in the Mediterranean region remains particularly challenging due to the localized and convective nature of extreme precipitation events. This study evaluates the potential of analog-based methods (AMs) to enhance 24-hour precipitation forecasts for Catalonia (northeastern Iberian Peninsula), with the broader objective of supporting flood risk management and early warning systems. The tested AMs use geopotential height fields at 500 and 1000 hPa as predictors and differ in complexity, combining Weather-Type classification (WT), Seasonal Standardization (S), and the Perfect Prognosis (PP) framework, a novel configuration in analog-based forecasting. Model performance was assessed against operational Global Forecast System (GFS) forecasts using fifth-generation ECMWF Re-Analysis (ERA5) as reference, for both moderate and extreme precipitation events associated with historical floods. Results show that AMs integrating Seasonal Standardization and the Perfect Prognosis framework markedly improve 24-hour precipitation forecasts relative to GFS, particularly in reproducing the intensity and spatial distribution of extreme events. These findings highlight the operational potential of enhanced AMs as efficient, data-driven complements to numerical weather prediction models, offering improved skill for flash-flood forecasting and impact-based risk management.

1. Introduction

Floods are the most frequent weather-related natural disasters worldwide and represent a major source of economic and human losses. Reflecting this prominence, data from 1990 to 2021 show that floods accounted for 42.6% of all documented natural disasters globally (EM-DAT, 2024; Damien Delforge, 2023). The risks associated with floods are particularly relevant for the Mediterranean region, where disastrous flash floods are much more frequent than in other parts of Europe (Gaume et al., 2009; Llasat et al., 2010). For instance, in Spain between 1971 and 2020, the Insurance Compensation Consortium (*Consorcio de Compensación de Seguros*, CCS) paid over €7 billion (adjusted to 2020 values) in compensation for flood-related damages, representing more than 60% of all insurance claims during that period (CCS, 2020).

The impact of floods is related to three main factors: hazard conditions (frequency and magnitude), vulnerability, and exposure (Merz et al., 2021). Regarding hazard conditions, floods in the Mediterranean region are often triggered by intense, short-duration convective rainfall events. These storms, which can produce precipitation rates up to 180 mm/h in just five minutes, tend to have a limited total rainfall accumulation, generally less than 100 mm (Gaume et al., 2016). Their impact is usually confined to small areas, leading to flash floods in small river basins with rapid hydrological responses. In other cases, mesoscale convective systems can produce sustained rainfall over several hours, leading to significantly higher accumulations, sometimes exceeding 200 mm in just a few hours (Gaume et al., 2016). These events can cover larger areas and are particularly common during the fall season along the northwestern Mediterranean coast. Finally, extreme rainfall may be part of a large-scale synoptic low lasting several days, leading to even more substantial accumulations. In such cases, local rainfall totals may reach 1,500 mm over four days, with records of nearly 1,000 mm in just 24 h (Gaume et al., 2016). Minimizing the risks associated with extreme events requires reducing both vulnerability and exposure. For this reason, it is essential to implement efficient early warning systems that increase public awareness and encourage participation in preventive actions (WMO, 2022). However, forecasting flood

* Corresponding author.
E-mail address: cguzzon@meteo.ub.edu (C. Guzzon).

https://doi.org/10.1016/j.wace.2025.100839
Received 13 June 2025; Received in revised form 20 October 2025; Accepted 21 November 2025
Available online 22 November 2025

events remains challenging, particularly in the case of flash floods, as they are characterized by their rapid onset (usually within less than six hours of rainfall), leaving very limited opportunity for an effective response (Hapuarachchi et al., 2011). Additionally, the highly convective and localized nature of these precipitation events makes their accurate prediction through numerical weather prediction models difficult, especially when their scale is smaller than the spatial resolution of the model (Hohenegger and Schär, 2007; Weusthoff et al., 2010). As a result, accurately locating these events in time and space with sufficient lead time remains difficult, and the limited reliability of precipitation forecasts constrains the accurate simulation of the hydrological response at the basin scale. Operational global forecasting systems, such as GFS and ECMWF, often exhibit inaccuracies in timing, location, and intensity, and tend to underestimate intense, localized precipitation associated with convective storms. Therefore, real-time flood forecasting critically depends on obtaining reliable precipitation forecasts with enough lead time to support timely and effective early warning systems. The classical forecasting chain involves the use of mesoscale numerical models forced by global numerical weather prediction (NWP) models with lower resolution (Cole et al., 2007). However, this methodology requires substantial computational resources, and the final forecast still presents uncertainties and biases (Horton et al., 2017).

A well-established alternative, extensively explored in literature, is forecasting based on analog methods (AMs) (Lorenz, 1956, 1969; Kruizinga and Murphy, 1983; van den Dool, 1989), particularly for precipitation prediction (Daoud et al., 2016; Horton et al., 2017; Panziera et al., 2011; Zhou and Zhai, 2016).

This weather forecasting technique is based on the definition of analogs by Lorenz (1956, 1969), i.e., two states of the atmosphere that closely resemble each other. The main hypothesis of AMs forecasting is that similar atmospheric conditions are likely to evolve in similar ways (Lorenz, 1969). One of the main advantages of this technique is its capacity to generate realistic weather predictions without relying on simplifications of atmospheric physics. On the other side, one of its limitations is that it cannot account for situations that are not present in the historic analog pool (Radinović, 1975; van den Dool, 1989). Various versions of AMs have been applied and tested in the literature, particularly for precipitation forecasting. The most explored AMs often rely on two levels of parameterization of atmospheric variables used as predictors to assess the degree of similarity among analogs: the first one focuses on analogies in atmospheric circulation, through the analysis of geopotential height fields, while the second introduces an additional level of analogy based on moisture variables (Obled et al., 2002; Bontron and Obled, 2005). In this regard, Bontron and Obled (2005) showed that geopotential heights at 500 hPa and 1000 hPa are the most effective predictors for identifying analogies in atmospheric circulation. For a broader synthesis of analog method variants and their applications, the reader is referred to Daoud et al. (2016).

Building on these established methods, this study introduces key innovations to the analog framework. Analog selection is guided by daily Weather Types (WT) to reduce computational cost and enhance operational applicability. Seasonal Standardization (S) is applied to account for intra-annual variability, while the Perfect Prognosis (PP) framework is implemented to link NWP forecasts with the fifth-generation ECMWF Re-Analysis (ERA5), thereby expanding the analog pool, improving the representation of extreme precipitation, and enhancing probabilistic skill. Previous studies in Catalonia and the broader Mediterranean region (Miró et al., 2020; Gibergans-Báguena and Llasat, 2007) have demonstrated the value of classical AMs; however, the integration of Seasonal Standardization and Perfect Prognosis represents a novel extension tested here for operational flash flood forecasting. AMs can complement NWP guidance in this context because they rely on historical situations with similar large-scale circulation patterns to transfer the corresponding observed precipitation fields to the forecast case. In doing so, AMs provide empirically grounded predictions that can better capture localized extremes, naturally produce probabilistic information from the ensemble of selected analogs, and require substantially fewer computational resources than high-resolution dynamical downscaling.

Expanding upon earlier studies of analog forecasting techniques, this research aims to develop an analog-based methodology to improve the prediction of precipitation fields during extreme precipitation events, with a specific focus on operational flash flood forecasting in the Catalonia region. To this end, this study investigates whether AMs can improve the timing, location, and intensity of 24-hour extreme precipitation forecasts over Catalonia compared with operational GFS output, and whether they can provide operationally useful probabilistic guidance for flash flood early warning.

In operational early warning chains for flash floods, implementing AMs can provide not only improved meteorological warnings but also enable impact-based forecasting (IBF). This is feasible because the impacts associated with the retrieved historical analogs can be transferred to the forecast case. Similar IBF approaches have been explored in recent studies about extreme rainfall and floods (Singhal et al., 2022; Rözer et al., 2021) and multi-hazards (Boult et al., 2022).

To provide a comprehensive analysis of the methodology and its performance, the remainder of the paper is structured as follows: Section 2 presents the data and methods used to implement the analog-based approaches. Section 3 presents the main results, while the discussion about the main findings of the study is presented in Section 4. Finally, Section 5 offers the conclusions, outlines the main limitations, and suggests directions for future work.

2. Data and methods

2.1. Spatial domain

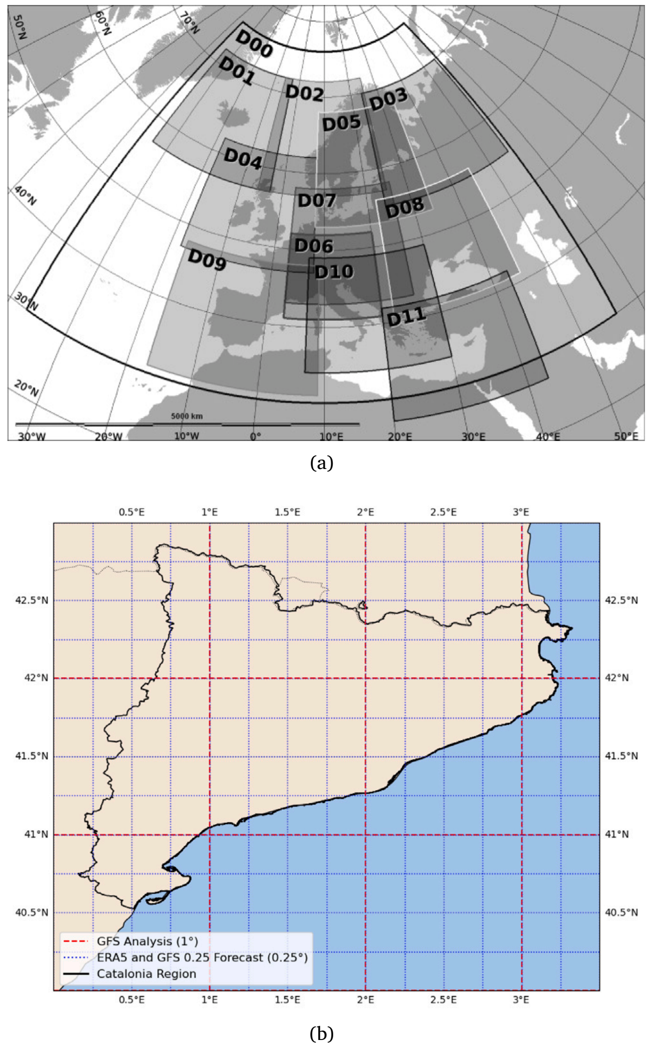
The selection of analogs has been based on the evaluation of geopotential fields at the synoptic scale. The spatial domain selected for this analysis is centered on the Iberian Peninsula and has been defined following the guidelines of COST Action 733: Harmonization and Application of Weather Type Classification for European Regions (Tveito et al., 2016; Philipp et al., 2010). Specifically, the D09 domain, corresponding to the Western Mediterranean region, has been selected. This domain spans from 17° W to 9° E in longitude and from 31° N to 48° N in latitude, and it is shown in Fig. 1(a). The validation of the precipitation fields produced by the different analog methods has been carried out over the Catalonia region. For this purpose, a rectangular subdomain centered on Catalonia has also been defined, bounded by 0° W to 3.5° E in longitude and 40° N to 43° N in latitude (Fig. 1(b)).

2.2. Reanalysis data

The verification of the AMs has been carried out using the precipitation field retrieved from the ERA5 Reanalysis, as ground truth for all the analyzed events. ERA5, developed by the European Centre for Medium-Range Weather Forecasts (ECMWF), is the fifth-generation global atmospheric reanalysis, covering the period from 1940 to the present (Hersbach et al., 2020). The data are provided on a regular latitude–longitude grid with a spatial resolution of 0.25° and a temporal resolution of one hour. In this study, we use the dataset 'ERA5 hourly data on single levels from 1940 to present', which provides hourly accumulated precipitation. The 24-hour accumulated precipitation fields have been computed by summing the hourly values. The resolution of the data over the study region is illustrated in Fig. 1(b).

2.3. GFS data

The Global Forecast System (GFS) is a global numerical weather prediction system operated by the U.S. National Weather Service (NWS). It is run 4 times daily, at 00, 06, 12, and 18 UTC. The forecasting component is based on the Finite Volume Cubed-Sphere (FV3) dynamical core with horizontal resolution of approximately 13 km and 127 vertical



(a)

(b)

Fig. 1. Spatial domains used in the study: (a) D09 Western Mediterranean, used to evaluate the analogs. (b) Catalonia region used for the verification of the AM's Precipitation Fields. Blue and Red grid show the spatial resolution of the GFS and ERA5-Reanalysis.
Source: Fig. taken from Philipp et al. (2010).

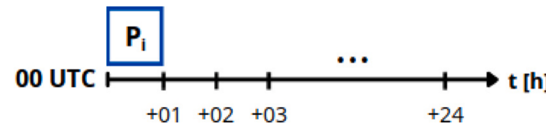


Fig. 2. Computation scheme of GFS-24h cumulated precipitation field.

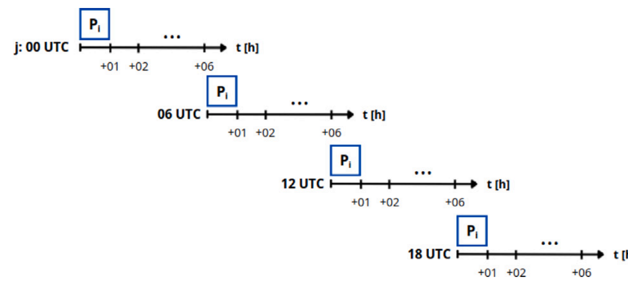


Fig. 3. Computation scheme of GFS-6h cumulated precipitation field.

layers. The atmospheric model is coupled to the NCEP Global Wave Model. It provides hourly forecast outputs up to 120 h (first 5 days), followed by 3-hourly outputs from hour 123 to hour 384, covering a total forecast range of up to 16 days (NOAA, 2024).

Two types of GFS products have been used in this study:

- GFS 0.25 Degree Forecast: this dataset consists of operational GFS model outputs with a spatial resolution of 0.25°. However, only five years of data (2015 to 2020) are available online. These data were used to compare the performance of the GFS against the AMs in predicting daily precipitation. The 24-hour accumulated precipitation field was obtained by summing the predicted hourly precipitation values from the model initialized at 00 UTC, following the scheme shown in Fig. 2 and described in Eq. (1). In the verification process, this product is referred to as GFS-24h. It represents the operational 24-h forecast typically used in real-time flood forecasting, where the model is run once daily at 00 UTC, and forecast lead time is a critical factor (Jain et al., 2017)

$$P_{24}(\text{GFS-24h}) = \sum_{i=01}^{24} P_i \tag{1}$$

- GFS Analysis (GFS-ANL Grid 3): This dataset contains historical outputs from the GFS model, covering the period from March 2004 to April 2020. It offers a spatial resolution of 1° x 1° in both latitude and longitude with model runs initialized four times daily at 00, 06, 12, and 18 UTC. The 24-hour cumulative precipitation field is derived by summing the four successive 6-hour forecasts from each day's model cycles. As a result, it provides more accurate 24-hour precipitation estimates and is therefore used in this study as a benchmark dataset, an idealized but non-operational reference for quality assessment. This dataset was used for two main purposes: (a) the analog search, based on geopotential height fields at 500 hPa and 1000 hPa at 00 UTC; and (b) the verification of the AMs, by computing 24-hour cumulative precipitation from the hourly forecasts of the four daily model runs. Specifically, forecasts from +01 to +06 h were extracted from each initialization and aggregated following the scheme in Fig. 3 and Eq. (2). For clarity, this precipitation product is referred to as GFS-6h in the verification process

$$P_{24}(\text{GFS-06h}) = \sum_{j}^{(00,06,12,18)} \sum_{i=1}^{6} P_i \tag{2}$$

2.4. Selection of historical events

To assess the effectiveness and quality of the implemented AMs, a verification process was conducted comparing the precipitation fields estimated by the analogs with the corresponding observed values. This comparison was performed separately for two distinct pools of historical events, allowing assessment of the AMs' performance under different conditions.

(1) Precipitation Days pool: it includes all days within the study period (2004–2020) during which precipitation was recorded in the study region. A day is classified as precipitation event if at least 1 mm/h of precipitation was recorded during any of the 24 hourly intervals, in at least one grid point of the spatial domain. The *Precipitation Days* pool comprises 3746 days. It was used to evaluate the performance of the AM in estimating precipitation fields during a wide range of precipitation events, from light to heavy.

(2) INUNGAMA Days pool: this pool comprises 230 days on which flood events of varying intensity were recorded across Catalonia. It was

derived from the INUNGAMA database (Barnolas and Llasat, 2007), an impact-based archive of flood events in the region covering the period 1900–2020. For the purpose of this study, however, only events between 1940 and 2020 were considered, consistent with the availability of ERA5 reanalysis data; earlier events were excluded. This pool of dates is used to evaluate the performance of the AM in reproducing extreme precipitation events.

2.5. Overview of the AMs tested

This section presents the methodology used to define the analogs. Several Analog Methods were tested iteratively to identify the configurations that deliver the best performance for both light and extreme precipitation events. The objective of the study is to assess how methodological modifications and increased levels of complexity can lead to improved skill in the specific task of forecasting rain and heavy rain events in the Catalonia region.

The following methodologies were implemented: (1) AM without weather types (No-WT); (2) AM with weather type classification (WT); (3) AM with weather types and seasonal anomaly standardization (WT-S); (4) AM with weather types and Seasonal Standardization under the Perfect Prognosis assumption (WT-PP-S).

Basic AM and ranking metrics

The Analog Method without weather type classification was selected as the benchmark for evaluating the other tested configurations. This version represents the simplest and most established form of the method. For the identification of analogs, the selected predictors were the geopotential height fields at 500 hPa (Z500) and 1000 hPa (Z1000), chosen to capture both mid-tropospheric and near-surface atmospheric conditions. Previous studies support this choice: Bontron and Obled (2005) identified Z500 and Z1000 as the most effective predictors within the NCEP/NCAR reanalysis framework, while (Horton et al., 2012) employed them to characterize the spatial relationship between atmospheric circulation and precipitation using analog techniques. Based on this evidence, Z500 and Z1000 were adopted in this study under the assumption that they are the most reliable predictors for identifying analogous meteorological situations in GFS model outputs.

The analog search was conducted by comparing the geopotential fields simulated by the GFS model for the target day with the fields from GFS analysis (GFS-ANL) at 00 UTC for the past days. For each analyzed event, analogs were searched within a pool of days included in a two-month window centered around the target day, to account for seasonality and thus consider days with a similar amount of solar radiation and comparable availability of energy in the atmosphere. To ensure the comparison was not biased by the event itself, days from the same year as the target date were excluded from the pool of available candidates for the analog search.

The analogs were then ranked from the most similar to the least similar to the target date by comparing the respective geopotential height fields. In total, three ranking methods were tested to assess their performances, ensuring that the most suitable approach was selected. The best-ranking method was chosen by computing the Root Mean Square Error (RMSE) between the geopotential field of the target day and each field of the best analogs, aggregating all the computed metrics to obtain a single value to minimize.

The first ranking metric tested was the Euclidean distance (E_D). Euclidean distance is a scale-dependent metric that measures the overall magnitude of the differences between corresponding values in the two spatial matrices. Euclidean distance is defined by Eq. (3). In our specific case, a_{ij} and b_{ij} represent geopotential height grid values belonging to the analog field (A) and the target day field (B), respectively.

The second tested metric was the Pearson spatial correlation (r). This metric accounts for the spatial correlation between the analyzed fields and measures the strength and direction of the linear relationship between the two compared fields, regardless of the scale of the values. Values of r close to 1 indicate that the matrices have a strong positive linear relationship. r is defined by Eq. (4). In this case, $\overline{a}$ and $\overline{b}$ are the mean values of the analog field (A) and the target day field (B), respectively. To obtain a dissimilarity metric to minimize, similar to the Euclidean distance, the Pearson correlation distance, defined as $P_D = 1 - r$ was considered.

Finally, the third ranking method implemented considers the combination of E_D and P_D metrics. The combination was carried out first by the normalization of each metric to ensure independence from their absolute values and then by calculating their arithmetic mean. This approach ensures that both the absolute differences in geopotential values between the two fields and their spatial correlation are equally considered. The smaller the value of the combined metric, the more similar the geopotential field of the analog is to the reference date. This process was carried out both for Z1000 and Z500 fields, and the corresponding metrics were then combined through simple addition, thereby ensuring equal weight and significance to both atmospheric levels. The aggregated score metric has to be minimized to identify the best analogs. It was found that the combination of the Euclidean distance and Pearson correlation coefficient performs better than applying the two metrics separately. For this reason, this new combined metric (E_D+P_D) was chosen for the ranking of the best analogs. Once the most efficient analogs ranking method was selected, the top 10 best analogs were considered for each target data analyzed.

$$E_D = \sqrt{\sum_{i,j}(a_{ij} - b_{ij})^2} \tag{3}$$

$$r = \frac{\sum_{i,j}(a_{ij} - \overline{a})(b_{ij} - \overline{b})}{\sqrt{\sum_{i,j}(a_{ij} - \overline{a})^2 \sum_{i,j}(b_{ij} - \overline{b})^2}} \tag{4}$$

WT AM

As part of efforts to enhance the analog selection process, a refinement step based on the classification of Weather Types (WT) was integrated into the methodology. This stratification aims to improve the robustness and accuracy of the analog search by constraining it to meteorologically homogeneous conditions. Weather types were computed and assigned to each day in the dataset based on geopotential height fields at 500 hPa (Z500) and 1000 hPa (Z1000), extracted from the GFS Analysis at 00 UTC. The classification method adopted follows the approach developed by Beck (2000), a threshold-based technique included among the standard classification methods compiled during COST Action 733 (Tveito et al., 2016; Philipp et al., 2010). The method, described in detail by Beck et al. (2007), assigns a specific weather type to each synoptic situation by evaluating its similarity to three idealized flow patterns: westerly, meridional, and cyclonic. It defines 18 classes that account for both flow direction and cyclonic characteristics. A summary of these synoptic types is provided in Table 1. This approach involves searching for analogs within a pool of past events whose atmospheric configurations are, by definition, similar to that of the target event being examined. The analog's ranking metric selected is the same as presented before ($E_D + P_D$).

WT-S AM

The analog method based on weather-type (WT) classification typically incorporates a temporal window to ensure that analogs are selected from similar calendar periods, thereby accounting for seasonality. However, this temporal constraint can overly restrict the pool of potential analogs, limiting the ability to find suitable matches, especially for rare or extreme events. To overcome this limitation, we remove the seasonal constraint and instead apply a post-hoc Seasonal Standardization (S) to the precipitation field. This approach allows us to bypass the rigid link between calendar date and severity, adapting the precipitation values to their seasonal context while broadening the

Table 1
Circulation types and their corresponding class codes.

Circulation type	Class
West Cyclonic	WC
South-West Cyclonic	SWC
North-West Cyclonic	NWC
North Cyclonic	NC
North-East Cyclonic	NEC
East Cyclonic	EC
South-East Cyclonic	SEC
West Anticyclonic	WA
South-West Anticyclonic	SWA
North-West Anticyclonic	NWA
North Anticyclonic	NA
North-East Anticyclonic	NEA
East Anticyclonic	EA
South-East Anticyclonic	SEA
South Anticyclonic	SA
Low Pressure (Cyclone)	C
High Pressure (Anticyclone)	A

pool of candidate analogs throughout the year. The seasonal correction relies on two fundamental assumptions:

(1) The daily precipitation field can be decomposed into two distinct components: a climatological seasonal signal, which reflects the expected precipitation for that specific time of year (from days above 1 mm/day), and an anomalous component, representing deviations from the seasonal norm.

(2) The anomalous component is assumed to be statistically independent of the seasonal cycle. By isolating and standardizing this component, we can compare precipitation values across different times of the year, thus enabling analog selection without the need for strict temporal alignment. This allows the analog method to capture similar dynamical situations while accommodating the seasonal context through a post-hoc correction.

Given these assumptions, each analog precipitation field (k = 1–10) is adjusted by recombining the climatological seasonal value corresponding to the target date with the anomalous component derived from the analog date. The anomaly field, denoted as $\hat{P}$, is obtained by seasonally standardizing the analog precipitation field, removing the seasonal mean and scaling by the seasonal variability, so that only the non-seasonal component remains. This operation is applied cell by cell using Eq. (5). In this equation, each combination of i and j represents the cells of the precipitation field. k is the number of the analog, and it varies from 1 to 10. d is the day in the year of the analog date (from 1 to 365). $\mu^{d}_{i,j}$ is the mean precipitation field computed from ERA5 Reanalysis using 30 years of data, from 1991 to 2020, for each day of the year d and for each cell of the spatial domain i,j. $\sigma^{d}_{i,j}$ is the standard deviation computed similarly to μ The corrected precipitation value ($\overline{P}$) for the analog date is reconstructed, for each cell i,j of the field, adding the climatological seasonal value of the target date to the anomaly value of the analog ($\hat{P}^{k,d}_{i,j}$), following Eq. (6). In this case, t represents the day in the year (from 1 to 365) of the target day to reconstruct the corrected precipitation field of the analog, considering the climatology of the precipitation associated with the target day.

$$\hat{P}^{k,d}_{i,j} = \frac{P^{k,d}_{i,j} - \mu^{d}_{i,j}}{\sigma^{d}_{i,j}} \quad (5)$$

$$\overline{P^{k}_{i,j}} = \sigma^{t}_{i,j} \cdot \hat{P}^{k,d}_{i,j} + \mu^{t}_{i,j} \quad (6)$$

WT-PP-S AM

The last AM tested is based on the definition of WTs, under the hypothesis of Perfect Prognostic (PP) between the forecast and the observations, with Seasonal Standardization (WT-PP-S). Under Perfect Prognosis hypothesis, the output of the model is considered perfect, hence the name Perfect Prognostic (AMS-Glossary, 2024). PP condition was applied between the GFS model output and the observations (ERA5 Reanalysis). Under this hypothesis, the search for analogs is conducted by comparing the geopotential fields of the target day with the observed fields of past events. This method allows to search for analogs in a pool of candidates covering 83 years, i.e. the entire ERA5 reanalysis database from 1940 to 2023, instead of the GFS forecast pool, consisting of only 17 years of data, from 2004 to 2020. This approach significantly extends the temporal domain for the analog search, thereby increasing the pool of potential candidate days. Seasonal Standardization and the analog ranking method presented earlier were also applied to the Perfect Prognosis AM. An overview of the tested AMs is provided in Table 2.

Average Analog's precipitation field

For each of the analog methods presented, the final analog precipitation field corresponding to each target date (P^{an}), used for comparison with both observational data and the output of the GFS numerical weather prediction model, was computed as the mean of the fields of the ten best analogs identified. This was done using the Eq. (7) for each grid cell i,j within the spatial domain. The mean of the best analogs mitigates the influence of outliers by accounting for inherent uncertainties and variability in atmospheric patterns, leading to more stable and representative precipitation forecasts. Fernández-Ferrero et al. (2010), Horton et al. (2017), Bellier et al. (2017)

$$P^{an}_{i,j} = \frac{1}{k}\sum_{k=1}^{10} P^{k}_{i,j} \quad (7)$$

2.6. Verification questions and metrics

The performances in predicting precipitation of the AMs are compared with the GFS model ones, using ERA5 Reanalysis as reference. In the verification process, we want to answer the following questions, both for general precipitation events (*Precipitation Days* Pool) and for extreme precipitation events (*INUNGAMA Days* Pool):

1. Which is the forecasting capability of the different AMs tested in predicting the maximum amount of precipitation that fell in the region during 24 h, regardless of the location of this maximum within the spatial domain?
2. What are the deterministic forecasting skills of the AMs in predicting the cumulated 24-hour precipitation field?
3. What is the ability of the different AMs to reproduce the spatial distribution of the precipitation field in the study region?

In general, the objective of the verification process is to assess whether the tested analog methods can improve the prediction of precipitation events of varying intensities compared to those of the global GFS model, for a forecast lead time of +24 h. The aim is to determine which of the tested methods delivers the best performances, especially for the prediction of extreme precipitation events, to identify the one that could be operationally applied in an early warning system for flood events in the Catalonia region.

To answer the questions presented above, the following quality metrics have been used for the deterministic verification of the AMs:

- Mean Error (ME): measures the systematic bias in the forecasts
- Mean Absolute Error (MAE): reflects the average magnitude of the errors
- Root Mean Square Error (RMSE): emphasizes larger errors by squaring differences, making it sensitive to outliers
- Probability Of Detection (POD): measures how well the forecast predicts precipitation above a threshold, ranging from 0 (never) to 1 (always) correctly predicted
- False Alarm Ratio (FAR): is the fraction of forecasted precipitation events that do not occur, with lower values indicating better forecasts

Table 2
Summary of AMs tested and their features.

Method	Weather type (WT)	Seasonal standardization (SS)	Perfect prognosis (PP)
No-WT AM	No	No	No
WT AM	Yes	No	No
WT-S AM	Yes	Yes	No
WT-PP-S AM	Yes	Yes	Yes

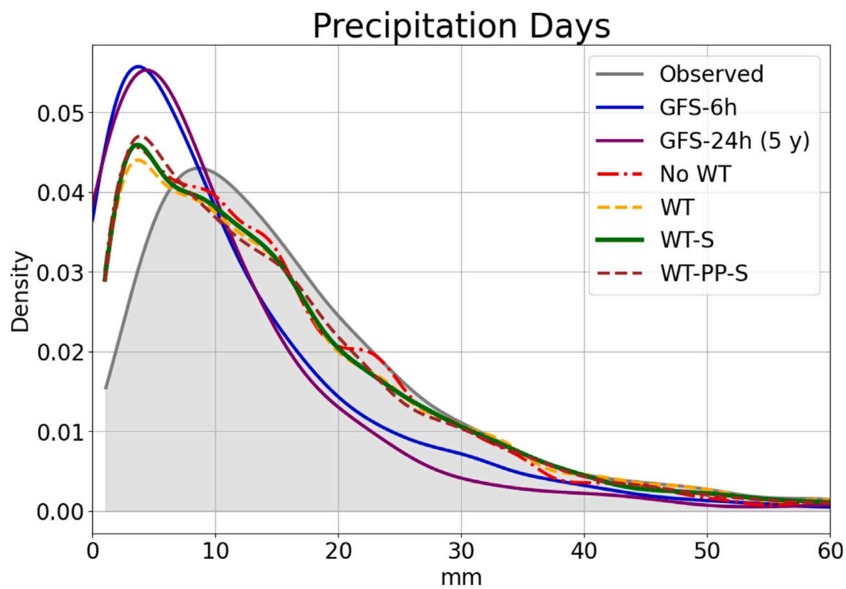


Fig. 4. Comparison of the PDFs of the maximum value of the 24-hour cumulated precipitation between Observations, GFS forecasts, and AMs predictions for the *Precipitation Days* pool. All AMs reduce the underestimation of heavy precipitation relative to GFS-24h, with no significant differences found among the tested methods.

- Critical Success Index (CSI): measures the fraction of correctly forecasted events out of all forecasted and observed occurrences, with values from 0 (worst) to 1 (best)
- Pearson correlation coefficient (r): measures the strength and direction of the linear relationship between predicted and observed precipitation fields

The POD, FAR, and CSI are computed using three threshold values of 24-hour accumulated precipitation for each dataset. The occurrence of threshold exceedance in precipitation amount is evaluated, pixel-by-pixel, across the entire spatial domain. An event is considered to have occurred if at least one grid cell within the study region records a 24-hour accumulated precipitation value exceeding the specified threshold. The threshold values are defined based on the distribution of the observed maximum 24-hour precipitation for each event (Figs. 4 and 7). The selected thresholds are 5, 10, and 30 mm/24 h for the *Precipitation Days* Pool and 25, 50, and 100 mm/24 h for the *INUNGAMA Days* Pool.

3. Results

3.1. Precipitation days verification

Maximum precipitation peak forecast

The first verification of the AMs focuses on their ability to predict the maximum precipitation peak observed during an event within the study region, and compares their performance against the GFS model. This predictand was selected because intense precipitation concentrated over a limited area is one of the most critical factors in flood forecasting.

Fig. 4 shows the comparison between the observed and the forecasted Probability Density Function (PDF) of this predictand, for the different AMs tested and for the GFS model. The gray distribution corresponds to the observed values, while the blue and purple distributions represent the values forecasted by the GFS-6h and the GFS-24h, respectively. The remaining distributions show the values of maximum precipitation value forecasted by the different AMs tested for each date in the analyzed pool of days.

The GFS models display a higher density in the lower precipitation range (5–10 mm) and a faster drop-off for heavier precipitation compared to observations, indicating difficulties of the model in predicting the likelihood of higher precipitation events. The tail of the distribution for the AMs closely follows the tail of the observations, with a slight underestimation bias, smaller than the one of the GFS model. The AM model without the definition of weather types (No-WT, in red) demonstrates predictions that closely align with the observed values for intermediate precipitation amounts, between 10 and 30 mm/24 h, while the models with definition of WTs (WT, in yellow; WT-S, in green; WT-PP-S, in brown) better reproduce the distribution of observations for heavy rainfall values ($\geq$ 30 mm/24 h). In general, the differences among the tested models are minimal, and no significant improvements are observed by applying Seasonal Standardization (WT-S), or analogs searching under Perfect Prognosis hypothesis between the GFS model and ERA5 observations (WT-PP-S).

Performance measurements

The verification of the AMs was extended to the full precipitation field to evaluate their ability to reproduce the overall spatial distribution of precipitation across the region. The deterministic verification metrics presented in Section 2.6 have been applied to the tested forecasting methods.

ME, MAE, and RMSE of each forecasting method are displayed in Fig. 5. The error metrics, all expressed in mm of rainfall, were calculated by comparing the predicted and the observed field for each event. The error values were then aggregated for all analyzed events, both for

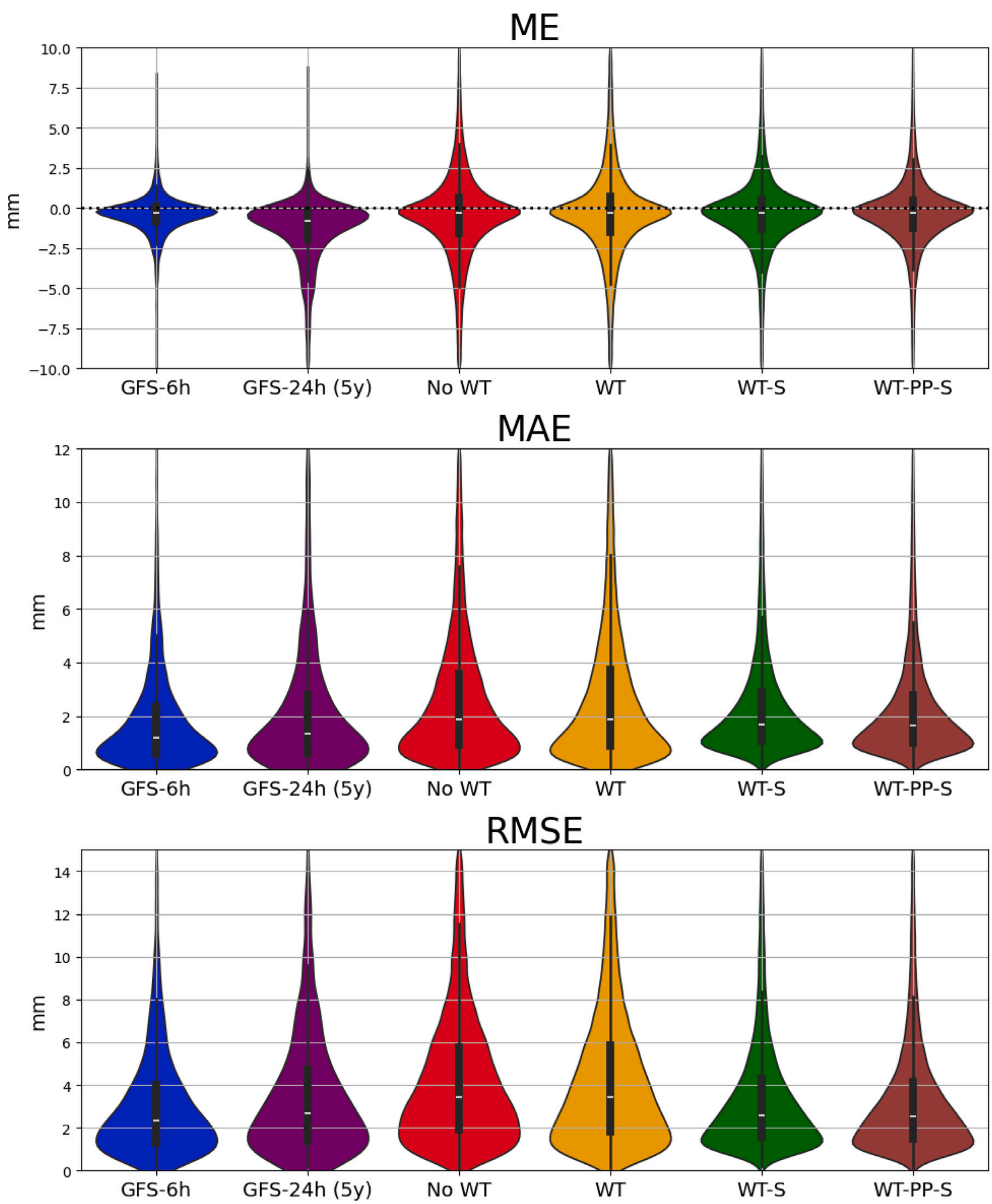


Fig. 5. Distribution of ME, MAE, and RMSE of the tested forecasting models for the *Precipitation Days* pool. Each violin represents the kernel-density estimate of the corresponding error metric across the *Precipitation Days*. All error values are expressed in millimetres of rainfall. The central marker denotes the median, while the thick bar indicates the interquartile range, providing a concise summary of the central tendency and spread. The violin shape illustrates the event-to-event variability and the presence of outliers or particularly large errors.

the GFS model and the AMs, in order to visualize their distribution in the violin plots. In terms of ME (Fig. 5 top), the GFS models (GFS-6h and GFS-24h) display distributions with a higher concentration of negative values, where both the mode and median are positioned below zero, indicating a negative bias and a tendency to underestimate precipitation. This is particularly pronounced in the GFS-24h (purple violin), which has a median ME value of approximately −1 mm. The AMs exhibit more symmetrical distributions around zero, with both the mode and median still slightly negative across all models, but they manage to reduce the negative bias observed in the GFS forecasts. It appears that Seasonal Standardization (WT-S and WT-PP-S) brings a slight improvement, shifting the distribution of ME closer to zero. The GFS-6h distribution has the smallest spread, while GFS-24h shows notable density down to ME values as low as −6 mm. The AMs display a larger spread than GFS-6h, both in the positive and negative directions. The MAE (Fig. 5 middle) demonstrates that WT-S and WT-PP-S achieve performance levels closer to the GFS-6h, as these three models show the smallest spread and lower densities at higher MAE values. The GFS-24h forecast shows a larger spread compared to the three mentioned models, and the same is true for No-WT and the WT models. Finally,

Table 3
Forecasting skill scores (POD, FAR, and CSI) for the GFS model and the best AM (WT-P-SS) for the *Precipitation Days* pool. WT-P-SS model seldom outperforms GFS-6h but consistently outperforms GFS-24h.

(a) POD			
Methods	**5 mm**	**10 mm**	**30 mm**
GFS-6h	0.742	0.634	0.488
GFS-24h	0.643	0.476	0.235
wt-pp-s	0.749	0.615	0.261
(b) FAR			
Methods	**5 mm**	**10 mm**	**30 mm**
GFS-6h	0.023	0.059	0.300
GFS-24h	0.062	0.140	0.515
wt-pp-s	0.083	0.242	0.669
(c) CSI			
Methods	**5 mm**	**10 mm**	**30 mm**
GFS-6h	0.729	0.610	0.404
GFS-24h	0.617	0.442	0.168
wt-pp-s	0.703	0.514	0.171

similar considerations can be made for the RMSE distributions (Fig. 5 bottom), where WT-S and WT-PP-S outperform GFS-24h, and they show similar performance to those of GFS-6h.

Table 3 shows the results of the verification of the binary forecast in terms of POD (3a), FAR (3b), and CSI (3c), computed for 3 different values of 24-hours cumulated rainfall (i.e. 5, 10, and 30 mm), to compare performances between GFS and AMs. Performance metrics were computed for all tested analog methods, but here we only show the results for WT-PP-S, as it consistently achieved the best performance across all precipitation thresholds. For lighter precipitation (< 5 mm), the performances of the WT-PP-S model are comparable, in terms of FAR and CSI, to GFS-6h, with even higher skills in POD . However, GFS-6h excels at reducing false alarms, showing a lower FAR of 0.023, compared to the 0.083 for WT-PP-S. Consequently, GFS-6h has a slightly better overall performance, as indicated by its higher CSI (0.729 versus 0.703). Despite this, the AMs outperform GFS-24h in terms of POD and CSI at the 5 mm threshold, although GFS-24h performs slightly better in terms of FAR. At the 10 mm threshold, GFS-6h still holds the lead in performance. WT-PP-S exhibits POD values very close to those of the GFS-6h, though it tends to produce more false alarms (higher FAR), which results in a worse CSI compared to GFS-6h. However, it outperforms the GFS-24h. For heavier precipitation events (≥ 30 mm), GFS-6h significantly outperforms the AMs. Although WT-PP-S model does not match the performance of the GFS-6h for any of the three metrics analyzed, it still outperforms the GFS-24h in terms of POD and CSI.

We then proceed to characterize the ability of the AMs to reproduce the spatial pattern of each precipitation event using Pearson spatial correlation coefficient (r). A higher r-value suggests a greater similarity between the spatial patterns of the two precipitation fields. The results are displayed in Fig. 6, which presents the PDFs of r computed by aggregating all the events of the verification pool analyzed, comparing the performances of the GFS and the AMs. GFS-6h's PDF is mostly concentrated on positive values of r and presents a maximum density for correlation values around 0.75, demonstrating a good spatial predictive ability of the precipitation field. The performances of the GFS significantly decrease considering the GFS-24h, with higher density for negative values of r and a maximum density for r-values about 0.3.

A single initialization of the numerical model and a longer forecast time reduce the accuracy of the GFS model in predicting the precipitation field, as expected. The tested AMs are unable to outperform the GFS-6h in terms of spatial distribution of the precipitation field; however, they do outperform the GFS-24h, particularly for WT-S (green) and WT-PP-S (brown) models. Conversely, the No-WT and WT models show worse performance and fail to outperform the GFS-24h.

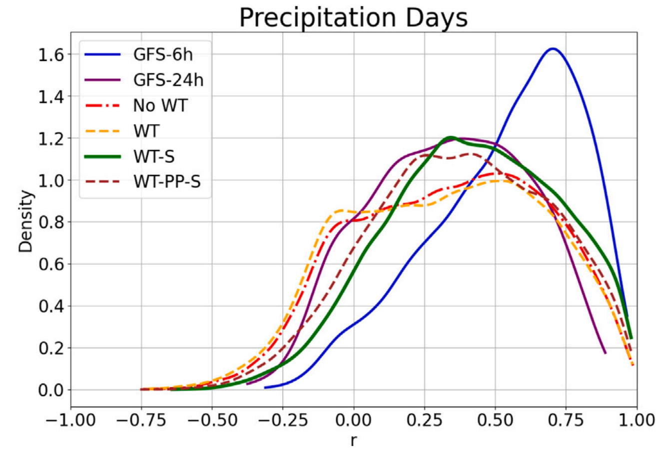


Fig. 6. PDFs of Pearson spatial correlation coefficient (r) of the tested forecasting models for the *Precipitation Days* pool. GFS-6h shows the best spatial agreement with observations, while among the analog methods, WT-S achieves the best performance.

3.2. *INUNGAMA days verification*

The same verification process has been applied to the *INUNGAMA Days* pool, to assess the performances of the AMs in predicting extreme precipitation events.

Maximum precipitation peak forecast

Fig. 7 shows the PDFs distribution of the maximum precipitation peak recorded on the domain for all 230 events of the *INUNGAMA Days* pool. All forecasting methods (both GFS and AMs) tend to underestimate the maximum precipitation peak for values exceeding 40 mm/24 h, underscoring the challenge of accurately predicting extreme precipitation events. However, some of the tested AMs bring some improvements compared to GFS-24h forecast, suggesting that AMs may offer a better representation of high-intensity precipitation under certain conditions.

Among the AMs, the No-WT and WT models provide a reasonable approximation of the observed distribution for moderate precipitation amounts (up to 40 mm/24 h), though they still exhibit a tendency to underestimate higher values. For precipitation exceeding this threshold, GFS-24h shows a distribution slightly closer to the observations than these baseline AMs, but it suffers from substantial underestimation as well. In contrast, the WT-S and WT-PP-S models demonstrate a marked improvement. Their distributions are more concentrated at higher precipitation levels, reducing the underestimation bias and outperforming GFS-24h across much of the extreme range. While GFS-6h remains the model whose distribution most closely matches the observations overall, the WT-S and WT-PP-S models significantly narrow the performance gap, indicating the added value of the seasonal correction in capturing high-intensity events.

Despite the improvements observed, all AMs still struggle to fully capture the extreme observed precipitation values.

Performance measurements

The results of the deterministic verification metrics (ME, MAE, RMSE) for the *INUNGAMA Days* pool are shown in Fig. 8. In general, error values for the *INUNGAMA Days* pool are larger than those computed for the *Precipitation Days* pool. This highlights the challenges faced by both GFS model and AMs in accurately predicting these extreme precipitation events, as they tend to underestimate the rainfall field.

In terms of Mean Error (ME), the analog methods incorporating Seasonal Standardization (WT-S and WT-PP-S) show the best performance. Compared to the No-WT and WT models, their ME distributions have

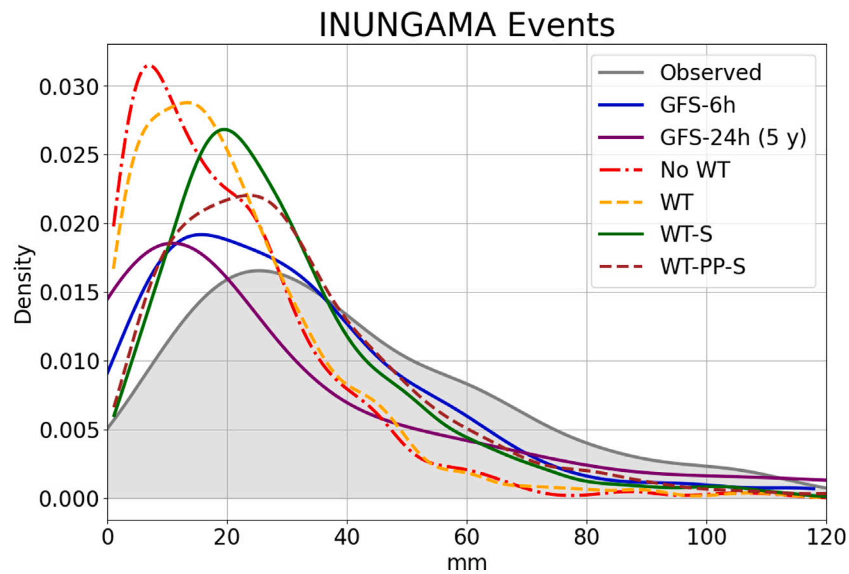


Fig. 7. Comparison of the PDFs of the maximum 24-hour accumulated precipitation between observations, GFS forecasts, and AMs predictions for the *INUNGAMA Days* verification pool. WT-S and WT-PP-S show the closest agreement with observations for heavy precipitation values, outperforming GFS-24h and approaching the performance of GFS-6h.

a reduced negative bias, with mode and median values closer to zero and a shorter left tail. Notably, the WT-PP-S method, based on Perfect Prognosis conditions, displays a distribution with a higher frequency of ME values near zero, indicating improved accuracy in capturing the overall precipitation magnitude. An analysis of the MAE and RMSE distributions further confirms that GFS-6h still provides the best overall performance in forecasting precipitation fields, even for these extreme events. However, WT-S and WT-PP-S consistently outperform GFS-24h, exhibiting lower median error values and fewer instances of large errors.

Table 4 shows the results of the verification of the binary forecast in terms of POD (4a), FAR (4b), and CSI (4c), computed for 3 thresholds of 24-hour cumulated rainfall (i.e. 20, 40, and 80 mm). The table reports only the results for the best-performing AM for simplicity. For the 20 mm threshold, the GFS-6h forecast achieves the highest POD (0.717), followed closely by the WT-PP-S model (POD = 0.591). GFS-6h shows a FAR which remains low (0.073), indicating strong predictive accuracy. On the other hand, the GFS-24h forecast shows a notable drop in detection skill (POD = 0.502) and a higher FAR (0.086), indicating a trade-off in the performance of the GFS model with one initialization per day in predicting the precipitation field. For higher precipitation thresholds (40 mm, 80 mm) the WT-PP-S method consistently outperforms the GFS-24h. Additionally, while the AM shows higher FARs compared to GFS-6h, it remains competitive with GFS-24h, exhibiting a lower FAR at 40 mm (0.366 vs 0.336). Looking at the CSI, the GFS-6h model consistently provides the highest values across all thresholds, particularly at 20 mm (CSI = 0.679). The WT-PP-S method, however, achieves a CSI of 0.457 at 20 mm and performs better than GFS-24h at higher thresholds, showing a CSI of 0.205 (40 mm) and 0.144 (80 mm), outperforming GFS-24h at both 40 mm and 80 mm precipitation thresholds. In conclusion, while GFS-6h provides the most reliable results across all thresholds, WT-PP-S model shows competitive performance and outperforms GFS-24h.

Fig. 9 shows the PDFs of Pearson spatial correlation (r) for the *INUNGAMA Days* Pool computed for all the tested forecasting methods. The GFS-6h (blue) shows the highest density distribution in the positive values region of r, with the peak around r = 0.7, suggesting that GFS-6h forecasts are more closely aligned with observations. GFS-24h (purple) exhibits a broader distribution with a lower peak located at values of r close to 0.25, highlighting weaker correlations and more variable performance in forecasting the precipitation field with a 24-hour lead time. Among the analog methods, WT-S (green) demonstrates a robust performance with a well-centered and relatively narrow distribution, indicating consistent and relatively high correlations with the observed precipitation fields. WT-PP-S (brown dashed line) also performs competitively, though with a slightly broader spread. The No-WT (red dotted line) and WT (orange dashed line) methods show lower and broader distributions, reflecting weaker correlation coefficients compared to the other methods.

While averaging the best 10 analogs can yield more stable and representative forecasts, it may also introduce excessive smoothing, potentially weakening the signal of heavy precipitation. This effect is particularly relevant when verifying extreme precipitation events (*INUNGAMA Days*). To address this concern, we evaluated how the

Table 4
Forecasting skill scores (POD, FAR, and CSI) for the GFS model and the best AM (WT-PP-S) for the *INUNGAMA Days* pool. WT-PP-S model substantially improving POD and CSI compared with GFS-24h across thresholds. However, it exhibits elevated FAR at the most extreme threshold.

(a) POD

Methods	20 mm	40 mm	80 mm
GFS-6h	0.717	0.677	0.375
GFS-24h	0.502	0.309	0.195
wt-pp-s	**0.591**	**0.332**	**0.200**

(b) FAR

Methods	20 mm	40 mm	80 mm
GFS-6h	0.073	0.045	0.400
GFS-24h	0.086	0.336	0.575
wt-pp-s	0.132	0.366	0.636

(c) CSI

Methods	20 mm	40 mm	80 mm
GFS-6h	0.679	0.456	0.300
GFS-24h	0.433	0.313	0.154
wt-pp-s	0.457	0.205	0.144

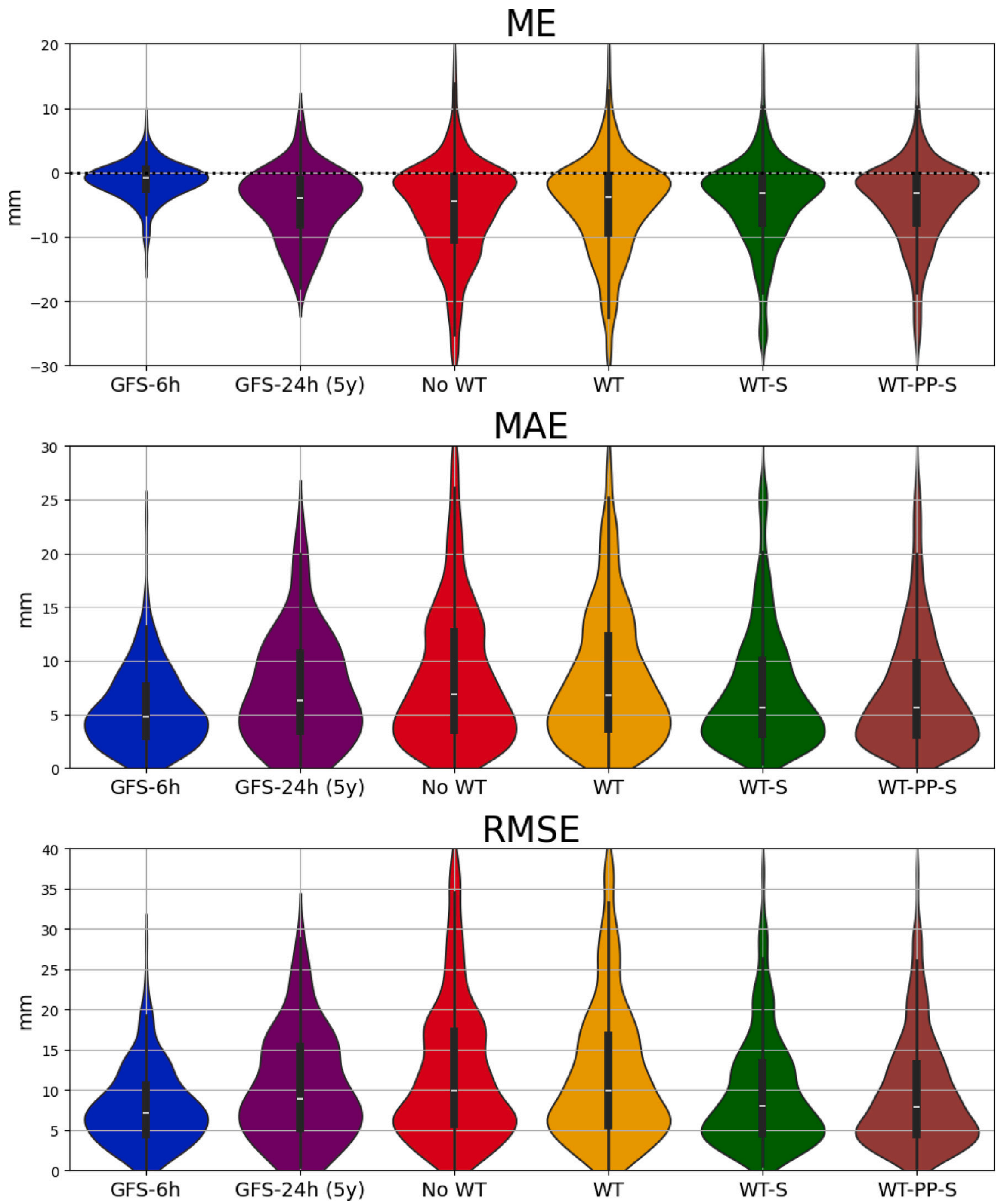


Fig. 8. Distribution of ME, MAE, and RMSE for the tested forecasting models in the *INUNGAMA Days* pool. Same violin-plot notation and units as in Fig. 5.

RMSE and the Pearson spatial correlation coefficient vary with the number of analogs used to compute the final analog field, ranging from the single best analog to the average of all 10. Fig. 10 summarizes the results of this analysis. RMSE values exhibit minor oscillations but remain largely stable when averaging up to five analogs, particularly for the best-performing methods (WT-S and WT-PP-S). Beyond this threshold, RMSE shows a slight increase, suggesting a modest reduction in forecast skill. Similarly, the Pearson spatial correlation coefficient decreases only marginally as more analogs are averaged . Overall, these variations are very small, on the order of a few decimal points, indicating that the analog search pool is sufficiently large and diverse to ensure high-quality analogs even among the lower-ranked members (8–10).

4. Discussion

We demonstrated that introducing weather-type classification, Seasonal Standardization, and the Perfect Prognosis framework significantly improved the performance of the tested analog methods, allowing them to outperform the predictive skill of the numerical GFS model at a 24-hour forecast lead time.

Including weather types (WTs) in the definition of AMs significantly enhances their operational applicability. By initially selecting a large pool of historical events, the method ensures that relevant atmospheric configurations closely resembling the target event are included. Once the WTs for the target event are identified, the WTs associated with the target event refine the analog pool, reducing its size to only events with

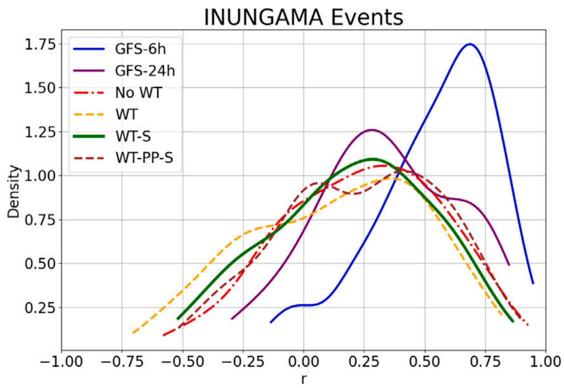


Fig. 9. PDFs of the Pearson spatial correlation coefficient (r) for the tested forecasting models in the *INUNGAMA Days* pool. GFS-6h exhibits the highest spatial agreement with observations, while among the analog methods, WT-S and WT-PP-S achieve the best performance.

matching WTs. This refinement greatly reduces computational demands while simultaneously improving the quality and efficiency of the analog search process.

The verification conducted on the *Precipitation Days* Pool demonstrated that the tested AMs outperformed the GFS-24h forecast, particularly at low and intermediate precipitation thresholds. Among the AMs, those incorporating weather type definition (WT), seasonally standardized (WT-S), and Perfect Prognosis hypothesis (WT-PP-S), showed notably competitive skill, outperforming GFS-24h. AM performances, computed for lighter precipitation events (≥ 5 mm), get closer to the GFS-6h ones, even outperforming the benchmark model in key metrics such as the Probability of Detection (POD) and the Critical Success Index (CSI). Despite these strengths, the GFS-6h remained the top-performing reference for higher precipitation thresholds (≥ 30 mm), particularly in reducing the false alarm ratio. In terms of spatial verification (r), although the AMs did not match the GFS-6h performances, they consistently outperformed the GFS-24h forecast, indicating better spatial coherence relative to the standard operational baseline.

The prediction of extreme precipitation events remains particularly challenging. Even the GFS-6h configuration, despite its theoretical nature and frequent model updates, struggles to reproduce both the maximum daily precipitation peak (Fig. 7) and the spatial structure of the precipitation field, resulting in higher error values than those observed for moderate precipitation cases (Fig. 8). This is especially pronounced for events exceeding 40 mm in 24 h. The AMs also tend to underestimate precipitation in extreme cases. However, we can clearly observe the advantages introduced by WT definition, Seasonal Standardization and PP conditions. In particular, WT-S and WT-PP-S align more closely with the observed distribution of peak precipitation (Fig. 7), especially for intermediate intensities up to 40 mm/24 h. WT-PP-S model demonstrates the best performance for extremes. The extended reference period increases the likelihood of identifying analogs for rare, high-impact events with multi-decadal return periods, thereby enhancing the skill of the analog approach in forecasting extreme precipitation, outperforming the GFS-24h predictions.

The final analysis examined the sensitivity of analog-based forecasts to the number of analogs averaged to generate the final precipitation field. The results show only minor variations in both RMSE and Pearson correlation coefficients as the number of averaged analogs increases. These variations, on the order of a few decimal points, confirm that the analog pool is sufficiently large to provide high-quality matches even beyond the top-ranked analogs. This supports the use of a 10-analog ensemble mean, which preserves the signal of extreme precipitation while enhancing forecasts robustness and stability. From an operational perspective, this configuration is also well-suited for early warning applications, where both the ensemble mean (average of the 10 best analogs) and the individual analog fields can be used as ensemble members to generate probabilistic forecasts, similarly to those produced by GFS and other NWP systems.

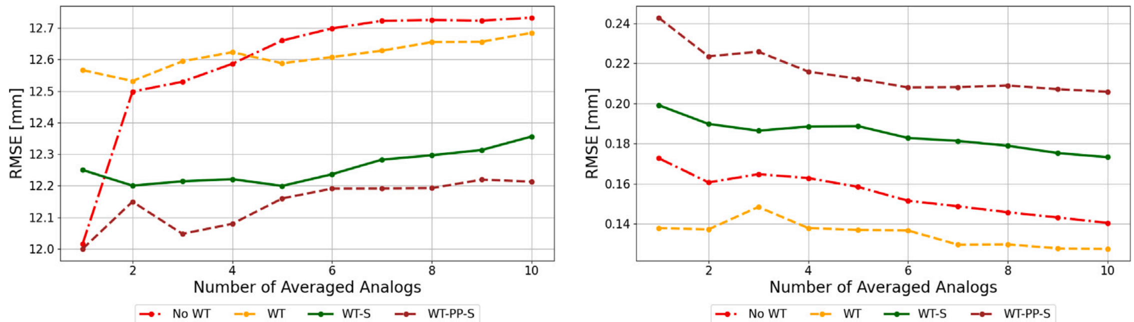
These results represent a preliminary, theoretical step towards a broader early warning system based on impact-based forecasting. The best-performing analog method identified in this study will be implemented operationally, with its precipitation outputs serving as inputs for hydrological simulations in the case-study basins and integrated into the Delft-FEWS system (Werner et al., 2013) to support impact-based flood forecasting. The hydrological applicability of the tested AMs has already been explored in a pilot study for the Francolíbasin located in the Catalonia region (Carril-Rojas et al., 2025), which demonstrated the potential of analog-based forecasts to enhance real-time flood risk management.

5. Conclusion

This study assessed several analog-based forecasting methods for their ability to predict heavy and extreme precipitation in Catalonia (NE Iberian Peninsula), with the goal of identifying the most effective configurations for operational use to complement the GFS predictions. The analog approaches, built using the 500 and 1000 hPa geopotential height fields as predictors and verified against ERA5 reanalysis, were compared to GFS forecasts using both theoretical (GFS-6h) and operational (GFS-24h) benchmarks. Performance was evaluated over a large set of general precipitation days and a targeted pool of flood events, to evaluate models' performances for extreme precipitation events. Methodological refinements of the analog methods tested included the use of weather types (WT), Seasonal Standardization (S), and a *Perfect Prognosis* approach (PP) to extend the analog search period and enhance forecast skill. The WT-S and WT-PP-S analog models showed the highest overall skill in forecasting precipitation across the study region, particularly under extreme conditions where other analog configurations and the GFS-24h forecast exhibited notable limitations. Among them, the WT-PP-S model proved to be the most robust, making it the strongest candidate for operational implementation. These findings underscore the added value of integrating Seasonal Standardization and the Perfect Prognosis framework into the analog methodology.

However, it is important to acknowledge certain limitations associated with the Perfect Prognosis approach used in the WT-PP-S model. This framework compares atmospheric fields from two distinct sources, the GFS forecast model and the ERA5 reanalysis, whose differences in physical parameterizations, spatial resolution, and model dynamics can introduce systematic inconsistencies during the analog selection process. Moreover, the approach implicitly assumes that GFS forecasts provide a near-perfect representation of the atmospheric state, thereby overlooking potential biases or errors inherent in the model itself. Despite these constraints, the principal strength of the WT-PP-S configuration lies in its access to an extended pool of analogs derived from a long historical archive (1940–2023), which significantly enhances the likelihood of identifying close matches for rare and intense events. This advantage is especially relevant for extreme precipitation scenarios with multi-decadal return periods and contributes to the overall operational robustness of the method.

The ERA5 reanalysis is used as the reference dataset (ground-truth) in this study; however, it is not free from uncertainties. In regions with complex terrain, such as Catalonia, ERA5 precipitation estimates may exhibit biases and errors related to its spatial resolution and the simplified representation of convective processes. These uncertainties propagate into the verification of both GFS forecasts and the analog methods, and should be considered when interpreting the results. Another limitation concerns the verification framework. The GFS-24h forecasts are only available for a relatively short temporal window (2015–2019), whereas the analog methods were tested over a much longer historical record. This discrepancy in dataset length introduces



(a) RMSE (b) Pearson *r*

Fig. 10. Impact of the number of averaged analogs on forecast performance for extreme precipitation events (*INUNGAMA Days*). Both RMSE (a) and Pearson *r* (b) remain largely stable when averaging up to 5 analogs, with minor degradation when including all 10 analogs.

some uncertainty into the direct comparison between AMs and GFS-24h. Nevertheless, benchmarking the AMs against GFS-24h remains essential for assessing their potential operational use, as it provides a realistic and widely adopted reference for forecast evaluation.

Beyond the methodological and verification aspects discussed above, analog-based forecasting relies on a fundamental assumption: that historical atmospheric situations are representative of present and future conditions. This dependence on historical similarity may be increasingly challenged by non-stationarity under climate change, particularly given projections of rising frequency and intensity of extreme precipitation events in the Mediterranean region (IPCC, 2023). Consequently, rare or unprecedented events may not have sufficiently close analogs in the historical archive, potentially reducing the predictive long-term skill of the analog methodology.

To further assess the operational relevance of analog-based forecasting, future work will go beyond meteorological validation. Specifically, the predictive value of analog-derived precipitation fields will be tested as inputs to hydrological models. These simulations will be compared with those driven by GFS-derived precipitation forecasts to evaluate whether analog-based methods can achieve comparable or superior performance. Ultimately, this analysis will help determine the extent to which analog forecasting approaches can support decision-making in flood forecasting and water resource management.

CRediT authorship contribution statement

Carlo Guzzon: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Raül Marcos-Matamoros:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization. **Maria Carmen Llasat:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization. **Montserrat Llasat-Botija:** Writing – review & editing, Project administration, Funding acquisition, Conceptualization.

Funding

This study is part of the project PLEC2022-009403, Flood2Now – Improvement of early warning systems for flood risk based on past information and citizen science data, funded by MCIN/AEI/10.13039-/501100011033 and the European Union NextGenerationEU/PRTR.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study is part of the project *PLEC2022-009403, Flood2Now – Improvement of early warning systems for flood risk based on past information and citizen science data*, funded by MCIN/AEI/10.13039/501100011033 and the European Union NextGenerationEU/PRTR. Furthermore, the project was carried out thanks to Inund-IA, funded by the *Agència Catalana de l'Aigua de la Generalitat de Catalunya*. One of the coauthors, Raül Marcos-Matamoros is a Serra Húnter fellow.

Data availability

The operational version of the algorithm code based on this analog methods study is available and can be downloaded here: https://github.com/cguzzon1998/analog_algorithm.git.

The GFS model data used for the verification of the tested methods can be downloaded from the NOAA website: https://www.ncei.noaa.gov/products/weather-climate-models/global-forecast, while the ERA5 reanalysis data can be obtained here: https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=download.

References

AMS-Glossary, 2024. Perfect prognostic (perfect prog).

Barnolas, M., Llasat, M.C., 2007. A flood geodatabase and its climatological applications: the case of catalonia for the last century. Nat. Hazards Earth Syst. Sci. 7 (2), 271–281.

Beck, C., 2000. Zirkulationsdynamische Variabilität im Bereich Nordatlantik-Europa seit 1780.

Beck, C., Jacobeit, J., Jones, P.D., 2007. Frequency and within-type variations of large-scale circulation types and their effects on low-frequency climate variability in central Europe since 1780. Int. J. Climatol. 27 (4), 473–491.

Bellier, J., Bontron, G., Zin, I., 2017. Using meteorological analogues for reordering postprocessed precipitation ensembles in hydrological forecasting. Water Resour. Res. 53 (12), 10085–10107.

Bontron, G., Obled, C., 2005. A probabilistic adaptation of meteorological model outputs to hydrological forecasting. Houille Blanche-Revue Int. L Eau 1.

Boult, V.L., Black, E., Saado Abdillahi, H., Bailey, M., Harris, C., Kilavi, M., Kniveton, D., MacLeod, D., Mwangi, E., Otieno, G., Rees, E., Rowhani, P., Taylor, O., Todd, M.C., 2022. Towards drought impact-based forecasting in a multi-hazard context. Clim. Risk Manag. 35, 100402.

Carril-Rojas, D., Guzzon, C., Mediero, L., Fernández-Fidalgo, J., Garrote, L., Llasat, M.C., Marcos-Matamoros, R., 2025. A flood forecasting method in the francolíriver basin (Spain) using a distributed hydrological model and an analog-based precipitation forecast. Hydrology 12 (8).

CCS, 2020. CCS: La cobertura de los riesgos extraordinarios en España, Consorcio de Compensacion de Seguros, Ministerio de Economìa y Hacienda, Madrid. CCS, Madrid, Spain.

Cole, S., Moore, R., Roberts, N., 2007. Using high resolution numerical weather prediction models to reduce and estimate uncertainty in flood forecasting. AGU Fall Meet. Abstr. -1, 03.

Damien Delforge, R.B., 2023. EM-DAT: the emergency events database.

Daoud, A.B., Sauquet, E., Bontron, G., Obled, C., Lang, M., 2016. Daily quantitative precipitation forecasts based on the analogue method: Improvements and application to a French large river basin. Atmos. Res. 169, 147–159.

EM-DAT, 2024. EM-DAT: The International Disaster Database. CRED/UCLouvain, Brussels, Belgium, (Accessed 10 September 2024).

Fernández-Ferrero, A., Sáenz, J., Ibarra-Berastegi, G., 2010. Comparison of the performance of different analog-based Bayesian probabilistic precipitation forecasts over Bilbao, Spain. Mon. Weather Rev. 138 (8), 3107–3119.

Gaume, E., Bain, V., Bernardara, P., Newinger, O., Barbuc, M., Bateman, A., Blaškovičová, L., Blöschl, G., Borga, M., Dumitrescu, A., Daliakopoulos, I., Garcia, J., Irimescu, A., Kohnova, S., Koutroulis, A., Marchi, L., Matreata, S., Medina, V., Preciso, E., Sempere-Torres, D., Stancalie, G., Szolgay, J., Tsanis, I., Velasco, D., Viglione, A., 2009. A compilation of data on European flash floods. J. Hydrol. 367 (1), 70–78.

Gaume, E., Borga, M., Llassat, M.C., Maouche, S., Lang, M., et al., 2016. Mediterranean extreme floods and flash floods. In: Editions, I. (Ed.), The Mediterranean Region under Climate Change. a Scientific Update. In: Coll. Synthèses, HAL open science, pp. 133–144.

Gibergans-Báguena, J., Llasat, M., 2007. Improvement of the analog forecasting method by using local thermodynamic data. Application to autumn precipitation in Catalonia. Atmos. Res. 86 (3), 173–193.

Hapuarachchi, H.A.P., Wang, Q.J., Pagano, T.C., 2011. A review of advances in flash flood forecasting. Hydrol. Process. 25 (18), 2771–2784.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.-N., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146 (730), 1999–2049.

Hohenegger, C., Schär, C., 2007. Atmospheric predictability at synoptic versus cloud-resolving scales. Bull. Am. Meteorol. Soc. 88 (11), 1783–1794.

Horton, P., Jaboyedoff, M., Metzger, R., Obled, C., Marty, R., 2012. Spatial relationship between the atmospheric circulation and the precipitation measured in the western swiss alps by means of the analogue method. Nat. Hazards Earth Syst. Sci. 12 (3), 777–784.

Horton, P., Jaboyedoff, M., Obled, C., 2017. Global optimization of an analog method by means of genetic algorithms. Mon. Weather Rev. 145, 1275–1294.

IPCC, 2023. In: Lee, H., Romero, J. (Eds.), Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Intergovernmental Panel on Climate Change (IPCC), Geneva, Switzerland, pp. 35–115.

Jain, S., Mani, P., Jain, S., Prakash, P., Singh, V., Tullos, D., Kumar, S., Agarwal, S., Dimri, A.P., 2017. A brief review of flood forecasting techniques and their applications. Int. J. River Basin Manag. 1–36.

Kruizinga, S., Murphy, A.H., 1983. Use of an analogue procedure to formulate objective probabilistic temperature forecasts in The Netherlands. Mon. Weather Rev. 111 (11), 2244–2254.

Llasat, M.C., Llasat-Botija, M., Prat, M.A., Porcú, F., Price, C., Mugnai, A., Lagouvardos, K., Kotroni, V., Katsanos, D., Michaelides, S., Yair, Y., Savvidou, K., Nicolaides, K., 2010. High-impact floods and flash floods in Mediterranean countries: the FLASH preliminary database. Adv. Geosci. 23, 47–55.

Lorenz, E.N., 1956. Empirical Orthogonal Functions and Statistical Weather Prediction Techincal Report. Technical Report, Massachusetts Institute of Technology, Department of Meteorology.

Lorenz, E.N., 1969. Atmospheric predictability as revealed by naturally occurring analogues. J. Atmos. Sci. 26 (4), 636–646.

Merz, B., Blöschl, G., Vorogushyn, S., Dottori, F., Aerts, J.C.J.H., Bates, P., Bertola, M., Kemter, M., Kreibich, H., Lall, U., Macdonald, E., 2021. Causes, impacts and patterns of disastrous river floods. Nat. Rev. Earth Environ. 2 (9), 592–609.

Miró, J., Pepin, N., Peña, J., Martin-Vide, J., 2020. Daily atmospheric circulation patterns for Catalonia (northeast Iberian Peninsula) using a modified version of Jenkinson and Collison method. Atmos. Res. 231, 104674.

NOAA, 2024. Global forecast system (GFS). (Accessed 17 September 2024).

Obled, C., Bontron, G., Garçon, R., 2002. Quantitative precipitation forecasts: a statistical adaptation of model outputs through an analogues sorting approach. Atmos. Res. 63, 303–324.

Panziera, L., Germann, U., Gabella, M., Mandapaka, P.V., 2011. NORA–nowcasting of orographic rainfall by means of analogues. Q. J. R. Meteorol. Soc. 137 (661), 2106–2123.

Philipp, A., Bartholy, J., Beck, C., Erpicum, M., Esteban, P., Fettweis, X., Huth, R., James, P., Jourdain, S., Kreienkamp, F., Krennert, T., Lykoudis, S., Michalides, S.C., Pianko-Kluczyńska, K., Post, P., Álvarez, D.F.R., Schiemann, R.K.H., Spekat, A., Tymvios, F., 2010. Cost733cat - A database of weather and circulation type classifications. Phys. Chem. Earth 35, 360–373.

Radinović, D., 1975. An analogue method for weather forecasting using the 500/1000 mb relative topography. Mon. Weather Rev. 103 (7), 639–649.

Rözer, V., Peche, A., Berkhahn, S., Feng, Y., Fuchs, L., Graf, T., Haberlandt, U., Kreibich, H., Sämann, R., Sester, M., Shehu, B., Wahl, J., Neuweiler, I., 2021. Impact-based forecasting for pluvial floods. Earth's Futur. 9 (2), 2020EF001851. 2020EF001851 e2020EF001851.

Singhal, A., Raman, A., Jha, S.K., 2022. Potential use of extreme rainfall forecast and socio-economic data for impact-based forecasting at the district level in northern India. Front. Earth Sci. 10 - 2022.

Tveito, O.E., Huth, R., Philipp, A., Post, P., Pasqui, M., Esteban, P., Beck, C., Demuzere, M., Prudhomme, C., 2016. COST Action 733: Harmonization and Application of Weather Type Classifications for European Regions; Final Scientific Report. Technical Report, Fakultät für Angewandte Informatik.

van den Dool, H.M., 1989. A new look at weather forecasting through analogues. Mon. Weather Rev. 117 (10), 2230–2247.

Werner, M., Schellekens, J., Gijsbers, P., van Dijk, M., van den Akker, O., Heynert, K., 2013. The delft-FEWS flow forecasting system. Environ. Model. Softw. 40, 65–77.

Weusthoff, T., Ament, F., Arpagaus, M., Rotach, M.W., 2010. Assessing the benefits of convection-permitting models by neighborhood verification: Examples from MAP D-PHASE. Mon. Weather Rev. 138 (9), 3418–3433.

WMO, 2022. Early Warnings for All: The UN Global Early Warning Initiative for the Implementation of Climate Adaptation - Executive Action Plan 2023–2027. World Meteorological Organization (WMO), Geneva, p. 56, Digital format, Action Plan.

Zhou, B., Zhai, P., 2016. A new forecast model based on the analog method for persistent extreme precipitation. Weather. Forecast. 31 (4), 1325–1341.