HAL
open science

# Towards Real-Time Daily PQPF by an Analog Sorting Approach. Application to Flash Flood Catchments

R. Marty, I. Zin, C. Obled, G. Bontron, A. Djerboua

## To cite this version:

R. Marty, I. Zin, C. Obled, G. Bontron, A. Djerboua. Towards Real-Time Daily PQPF by an Analog Sorting Approach. Application to Flash Flood Catchments. Journal of Applied Meteorology and Climatology, 2011, 51, 10.1175/JAMC-D-11-011.1 . insu-00648774

**HAL Id: insu-00648774**

**https://insu.hal.science/insu-00648774v1**

Submitted on 10 Mar 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Toward Real-Time Daily PQPF by an Analog Sorting Approach: Application to Flash-Flood Catchments

RENAUD MARTY, ISABELLA ZIN, AND CHARLES OBLED

*Grenoble-INP/UJF-Grenoble 1/CNRS/IRD, LTHE UMR 5564, Grenoble, France*

GUILLAUME BONTRON

*Compagnie Nationale du Rhône, Direction de l'Energie, Lyon, France*

ABDELATIF DJERBOUA

*RHEA, Nanterre, France*

(Manuscript received 3 January 2011, in final form 10 October 2011)

ABSTRACT

Heavy-rainfall events are common in southern France and frequently result in devastating flash floods. Thus, an appropriate anticipation of future rainfall is required: for early flood warning, at least 12–24 h in advance; for alerting operational services, at least 2–3 days ahead. Precipitation forecasts are generally provided by numerical weather prediction models (NWP), and their associated uncertainty is generally estimated through an ensemble approach. Precipitation forecasts also have to be adapted to hydrological scales. This study describes an alternative approach to commonly used limited-area models. Probabilistic quantitative precipitation forecasts (PQPFs) are provided through an analog sorting technique, which directly links synoptic-scale NWP output to catchment-scale rainfall probability distributions. One issue concerns the latest developments in implementing a daily version of this technique into operational conditions. It is shown that the obtained PQPFs depend on the meteorological forecasts used for selecting analogous days and that the method has to be reoptimized when changing the source of synoptic forecasts, because of the NWP output uncertainties. Second, an evaluation of the PQPFs demonstrates that the analog technique performs well for early warning of heavy-rainfall events and provides useful information as potential input to a hydrological ensemble prediction system. It is shown that the obtained daily rainfall distributions can be unreliable. A statistical correction of the observed bias is proposed as a function of the no-rain frequency values, leading to a significant improvement in PQPF sharpness.

## 1. Introduction

Catchments of southern France are regularly subject to flash floods generated by intense-rainfall events—generally in the autumn (i.e., from September to December). Thus, flood forecasting requires an appropriate anticipation of future rainfall: to issue an early flood warning, at least 12–24 h before the event; to alert operational or safety services, at least 2–3 days ahead. Such quantitative precipitation forecasts (QPF) are generally provided by numerical weather prediction (NWP) models or systems. The chaotic atmospheric processes imply that uncertainty is intrinsic to meteorological forecasting, but NWP systems also introduce uncertainties associated with the dynamical and physical representation of the atmosphere, especially related to subgrid parameterizations. Another source of uncertainty arises from the characteristics of our hydrological targets: in our flash-flood context, we are concerned with small to midsized catchments (100–1000 km$^2$) in mountainous or Mediterranean regions. Medium-range NWP systems are generally run at coarse vertical resolution, smoothing the topography, and provide QPFs on grids at horizontal resolutions from 15 up to 30 km, which is often inadequate for such catchments. Thus, it is preferable for precipitation forecasts to be probabilistic (PQPF), so as to express related uncertainty. These PQPFs can be given either in the form of probabilities of exceeding a given precipitation

*Corresponding author address:* Isabella Zin, LTHE, BP 53, 38041 Grenoble CEDEX 9, France.
E-mail: isabella.zin@ujf-grenoble.fr

DOI: 10.1175/JAMC-D-11-011.1

© 2012 American Meteorological Society

Unauthenticated | Downloaded 03/10/21 09:06 AM UTC

threshold or as precipitation amounts by percentiles. The latter form is considered here, and PQPFs are given in terms of cumulative distribution functions of precipitation amount.

The most common source of PQPFs comes from ensemble prediction systems that expand the classical deterministic approach by taking into account the initial conditions and the model uncertainties (Kalnay 2003; Leutbecher and Palmer 2008). Such systems are operational at the Canadian Meteorological Center (Houtekamer et al. 1996), at the National Centers for Environmental Prediction (NCEP; Toth and Kalnay 1997), and at the European Centre for Medium-Range Weather Forecasts (ECMWF; Molteni et al. 1996; Buizza et al. 2007). They can eventually be coupled with a limited-area model (LAM) to provide meteorological ensemble forecasts at lower spatiotemporal scales; in general, only a few representative members of ensemble prediction systems are put into a LAM (Marsigli et al. 2001, 2005; Brankovic et al. 2008). Such an approach adds additional sources of uncertainty due to the coupling of two meteorological models and because of the LAM's own uncertainty. Nevertheless, such ensemble prediction systems require a large amount of numerical resources, and, even if the ensemble members provide useful information about future rainfall and its uncertainty, ensemble forecasts are frequently biased and underdispersive, especially for precipitation.

Local-scale forecasts can also be obtained by statistical postprocessing of NWP output, with the advantage that these statistical methods generally correct a substantial part of the bias. Multiple linear regression is a widely used method for generating high-resolution forecasts (e.g., Murphy 1999; Wilby et al. 2003). Other methods aim at reconstructing the subgrid spatiotemporal variability from NWP by providing PQPFs at spatiotemporal scales required by hydrological modeling. These include the Schaake shuffle of Clark et al. (2004) and the analog sorting technique proposed by Obled et al. (2002). We have recently seen a growing interest in this latter technique, which adapts well-forecast synoptic variables issued by meteorological models to provide a conditional distribution for more local variables like the expected rainfall (Bliefernicht and Bárdossy 2007; Gibergans-Báguena and Llasat 2007; Diomede et al. 2008). Also, Hamill et al. (2006, 2008) highlight that statistical calibration of NWP output can also be done to improving NWP predictive skill by using a long set of reforecasts.

In this paper, the focus will be on an improved version of the analog technique proposed by Bontron and Obled (2005), adapted for midsized catchments of southern France and described in section 2. Section 3 details further developments required to implement the most-recent version into an operational context. Then, in section 4, PQPFs provided in real time are assessed as early-warning predictors, as well as potential input for hydrological ensemble prediction systems. Section 5 is devoted to reliability evaluation and correction of the internal bias of the analog method.

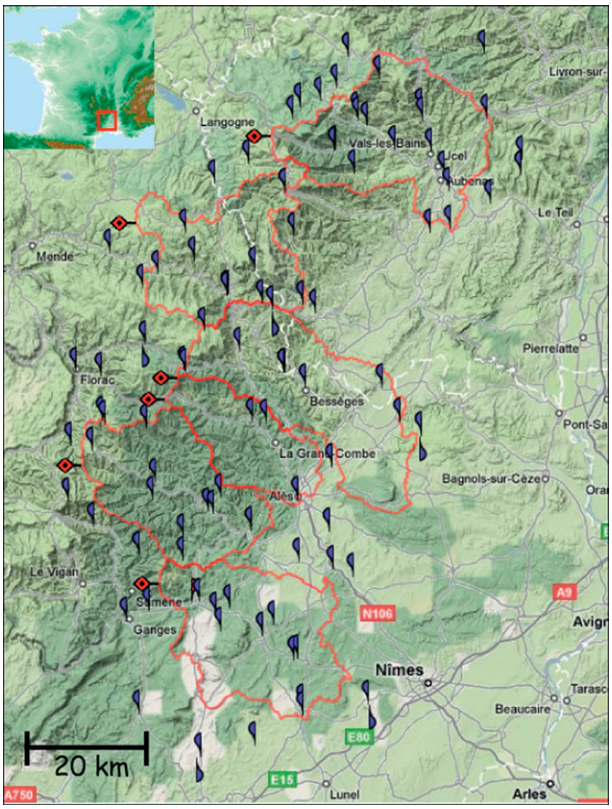


FIG. 1. Considered flash-flood catchments in the Cévennes-Vivarais region. Rain gauges are shown by a blue flag, and catchment boundaries are in red. Shown from north to south are Ardèche, Chassezac, Cèze, Gardon at Alès, Gardon at Anduze, and Vidourle (from SEVnOL; Boudevillain et al. 2011).

## 2. An analog approach adapted to flash-flood catchments

### a. *Flash-flood catchments in the Cévennes-Vivarais region*

Although the approach proposed hereinafter is general, this study will focus on six flash-flood-prone catchments in the Cévennes-Vivarais region (see details in Fig. 1 and Table 1). These catchments are part of the Cévennes-Vivarais Mediterranean Hydrometeorological Observatory (CVMHO; see online at http://www.ohmcv.fr; Delrieu et al. 2005), which collects the observed data and supports research to improve understanding of intense events.

Unauthenticated | Downloaded 03/10/21 09:06 AM UTC

TABLE 1. Study catchment descriptions.

| Catchment | Area ($km^2$) | Alt (m) |
|---|---|---|
| Ardèche at Vogüé | 635 | 140–1467 |
| Cèze at Tharaux | 665 | 115–1580 |
| Chassezac at Gravières | 498 | 172–1680 |
| Gardon at Alès | 316 | 123–1347 |
| Gardon at Anduze | 545 | 132–1567 |
| Vidourle at Sommières | 618 | 27–960 |

### b. *General principles of the analog sorting approach*

The analog sorting approach considers a given targeted situation, characterized by some synoptic meteorological fields, and attempts to provide a probability distribution of some local variables, such as catchment-averaged rainfall, conditional to the given synoptic situation. The targeted situation may be a past situation, for which the synoptic situation is known but the rainfall is lacking (reconstruction or gap filling of past data), or, as in our case, it is a future situation, for which a more or less reliable synoptic forecast is available but the corresponding rainfall is unknown and has to be predicted. To do that, the method assumes that 1) synoptic situations similar to the targeted one have been observed in the past, 2) local variables are linked to synoptic ones, and 3) variability due to the local effects is contained in the observed rainfall archive. The selected synoptic variables have to be well forecast by meteorological models and, of course, physically interrelated with the local variable that is to be forecast, here catchment-averaged rainfall. The forecast synoptic fields of the targeted situations are first compared with the past ones. The similarity is determined by an analogy criterion applied over a given analogy domain. Thus, a cluster of the best similar situations is extracted. Precipitation values collected during those past situations over each catchment of interest provide a sample that is directly defined at the catchment hydrological scale. This allows one to derive the expected conditional rainfall distribution for the targeted situation. Last, a classical probability model can be fitted on the different samples (here, a gamma law has been chosen). The method requires that both a long meteorological archive and the concurrent rainfall archive be available.

### c. *Optimized algorithms and performance in perfect-prognosis conditions*

Bontron and Obled (2005) have developed the analog method for midsized catchments in mountainous regions and hilly landscapes. Only results corresponding to flash-flood catchments in the Cévennes-Vivarais region are detailed in this paper. The optimization of the method was performed in perfect-prognosis conditions. This means that target and analog situations are extracted from the same archive, here made up of NCEP–National Center for Atmospheric Research (NCAR) reanalyses (Kalnay et al. 1996), laying down the constraint that an analog situation cannot belong to the same year as the target situation to avoid selecting as the best analog of a situation the situation itself. The meteorological archive considered here covers the period 1953–2001. Two algorithms have been elaborated by Bontron and Obled (2005) and are summarized in Fig. 2. The first one, called ana24-M2, uses a single level of analogy that is based on geopotential fields only. The second one, called ana24-M3,

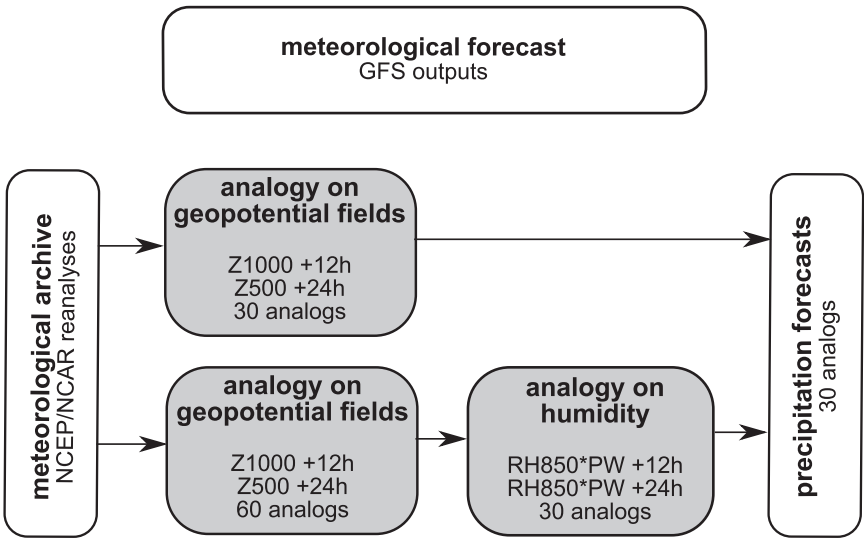


FIG. 2. Schema of one- and two-level algorithms of the analog approach.

Unauthenticated | Downloaded 03/10/21 09:06 AM UTC

completes the selection of analog situations extracted at the first level with a second level of analogy based on more local information from humidity fields, namely the product of precipitable water available (PW) and relative humidity at 850 hPa (RH850). Precipitable water gives information about the absolute amount of water vapor content in the atmospheric column, and the relative humidity adds information on how close the air is to saturation and then on how water vapor tends to condense. The combination of these two meteorological variables has been shown by Bontron and Obled (2005) to be a better explanatory variable than PW itself.

The first algorithm, ana24-M2, considers a domain that has been optimized over southern Europe (37.5°–47.5°N, 9°W–15°E) with a 2.5° grid. It selects the most similar situations according to the Teweless–Wobus or S1 score (Jolliffe and Stephenson 2003), which evaluates the similarity in shape of geopotential fields:

$$\mathrm{S1} = 100 \times \frac{\sum_i |\Delta\hat{x}_i - \Delta x_i|}{\sum_i \max(|\Delta\hat{x}_i|, |\Delta x_i|)}, \tag{1}$$

where the sums are determined over the analogy grid. In Eq. (1), $\Delta\hat{x}_i$ and $\Delta x_i$ are, respectively, the forecast pressure difference and the candidate analog pressure difference for the $i$th pair of adjacent points. The S1 value is equal to 0 for identical fields and is 200 for totally opposite fields; small values of S1 are desirable. In perfect-prognosis conditions, the optimum number of analogs to retain by the ana24-M2 algorithm has been optimized to 30.

The second algorithm, ana24-M3, first selects the best 60 analog situations following the same algorithm as in ana24-M2. Then, it applies a second level of analogy on humidity fields over a smaller analogy domain (42.5°–45°N, 2.5°–5°E), where again the 30 most-similar situations are extracted from the 60 best situations retained at the first level. This time, the chosen similarity criterion for humidity is the root-mean-square error (RMSE) (Jolliffe and Stephenson 2003) and is applied locally on the four nearest grid points surrounding the targeted catchments:

$$\mathrm{RMSE} = \left[\frac{1}{4}\sum_{i=1}^{4}(\hat{x}_i - x_i)^2\right]^{1/2}, \tag{2}$$

where $\hat{x}_i$ and $x_i$ are respectively the forecast and the candidate analog values at the $i$th point of the humidity field.

For both algorithms, the search for similar situations is restricted to a 4-month period around the target forecast date to cope with seasonal effects. For each selected analog situation, the corresponding catchment-averaged rainfall is extracted from the hydrological archive, and the resulting sample allows one to build an empirical rainfall distribution that constitutes the PQPF provided by the analog method. In fact, PQPFs are given in terms of a transformed variable $R$, defined as

$$R = (P/P_{10})^{1/2}, \tag{3}$$

where the $P$ are the analog precipitation values, $P_{10}$ is the catchment-averaged 10-yr return-period rainfall amount estimated by extreme values analysis (EVA) on the hydrological archive data by classical methods (e.g., Gumbel 1958; Castillo 1988), and the $R$ are the transformed precipitation values. The division by $P_{10}$ is a form of adimensional scaling that allows comparison between catchments that are different in size and/or climatological characteristics, and the square root reduces the skewness of the rainfall distribution.

Last, the empirical PQPFs are fitted by a gamma law with two parameters ($\lambda$ and $\rho$) adjusted on the strictly positive transformed rainfall values, conditional to the empirical no-rain frequency value:

$$F(r) = F(0) + [1 - F(0)]F_+(r, \lambda, \rho), \tag{4}$$

with

$$F_+(r, \lambda, \rho) = \frac{1}{\Gamma(\lambda)}\int_0^r e^{-z/\rho}\left(\frac{z}{\rho}\right)^{\lambda-1} dz. \tag{5}$$

In the previous two equations, $F(r)$ is the cumulative transformed rainfall distribution, $F(0)$ is the frequency of no rain, $F_+$ is the cumulative distribution of strictly positive transformed rainfall values, $r$ are the transformed rainfall values, $\lambda$ and $\rho$ are the adjusted gamma-law parameters, and $\Gamma$ is the gamma function of $\lambda$.

All of the ingredients in this analog approach—that is, the retained predictive variables, the shape and the position of the synoptic-fields analogy windows for the synoptic fields, and the number of retained analog situations—result from an intensive calibration by learning from the past archive. For that, the performance has been evaluated by the continuous ranked probability skill score (CRPSS) derived from the continuous ranked probability score (CRPS) as shown in Eqs. (6) and (7) below. As described by Jolliffe and Stephenson (2003), this score assesses the proposed rainfall distribution or PQPF as a whole in comparison with the observed rainfall. The CRPSS value represents the prediction gain provided by the analog-based PQPFs relative to a reference probabilistic forecast, taken here as the climatological rainfall distribution given on each considered catchment:

Unauthenticated | Downloaded 03/10/21 09:06 AM UTC

$$\mathrm{CRPS} = \frac{1}{N}\sum_{j=1}^{N}\int_0^{\infty}[F_j(x) - H(x, x_j^o)]^2\,dx, \qquad (6)$$

where $N$ is the number of issued forecasts, $F_j$ is the probability of occurrence of the transformed precipitation $x$ value, $x_j^o$ is the transformed rainfall observation value, and $H(x, x_j^o)$ is the Heaviside function, equal to 1 if $x \geq x_j^o$ and 0 otherwise. So,

$$\mathrm{CRPSS} = 1 - \frac{\mathrm{CRPS}}{\mathrm{CRPS}_{\mathrm{clim}}}, \qquad (7)$$

where $\mathrm{CRPS}_{\mathrm{clim}}$ is the CRPS obtained by taking for $F(x)$ the climatological rainfall distribution calculated on the basis of the observed rainfall archive values [once again, after transforming the observed values $P$ into transformed values $R$, following Eq. (3)].

In perfect-prognosis conditions and for the whole period 1953–2001 (taking into account all days), the CRPSS, averaged over all catchments, is 39.0% for the analog method ana24-M2; the ana24-M3 algorithm increases the performance up to 42.7%. These results outline that the analog approach is a very useful adaptation method for estimating precipitation at catchment scale in perfect-prognosis conditions. The next section is devoted to the assessment of the analog-based statistical adaptation's skillfulness in predicting rainfall distributions in a real-time context.

## 3. The "RainFAST" system: Toward operational PQPFs

### *a. Adaptation to real-time forecasts*

Up to now, the analog approach has been developed and optimized in perfect-prognosis conditions. When moving to real-time forecasts, the first choice to be made is to select the NWP model from which the meteorological forecasts will be collected. In the ideal case, these forecasts should be issued by the same model as the one used to provide the reanalyses that constitute the meteorological archive, at the same spatial and temporal resolutions. This would guarantee that the statistical structures of both the reanalyzed situations in the archive and the forecast ones would be similar. The NWP model will have different forecast performance depending on the choice of the lead time and the considered synoptic variables. For example, the model may be reasonably good at predicting the geopotential fields at lead times of 5 or 7 days whereas it may lose its capacity to predict relative humidity fields beyond 2 or 3 days only. These two aspects—1) the consistency between the synoptic fields provided by the forecast and those in the meteorological archive and 2) the decrease in the performance of the NWP model with increasing lead time—may influence the efficiency of the analog approach for successive lead times. Therefore, some further optimization and learning in true forecast conditions are necessary.

Long archives of forecasts are not readily accessible, however, not to mention that NWP model versions are changing very often. In our case, the most readily available archive of meteorological forecasts was that proposed by ECMWF. It was derived from the Ensemble Prediction System (EPS), which is made of 51 members, including a deterministic or control forecast (Molteni et al. 1996). We have collected an EPS-based forecast archive for a 5-yr period (1997–2001), in which the forecasts associated with each targeted day are considered up to 10 days ahead (from 0 to 240 h ahead), with a 12-h time step. The extracted variables are those selected in the previous optimization process, that is, the 1000- and 500-hPa geopotential fields, relative humidity at 850 hPa, and total column precipitable water (Thévenot 2004).

#### 1) HOW DO THE FORECAST PREDICTIVE VARIABLES DEPEND ON THE CHOSEN NWP MODEL AND ON THE LEAD TIME?

EPS is proposed by ECMWF at a finer spatial resolution than NCEP–NCAR reanalyses (1.125° during 1997–2000 and 0.7° during 2000–01 vs 2.5°, respectively). Therefore, because the continental topography is better resolved in the ECMWF simulations, EPS is potentially more informative than NCEP–NCAR reanalyses are, even after having aggregated them at the same spatial resolution (i.e., 2.5°), as we did. Indeed, a comparison of the NCEP–NCAR reanalyses and the forecast EPS control member issued on the same day has been performed to detect a possible dependency of the analog-based PQPFs on the source of meteorological forecasts.

The comparison has been made over the whole period 1997–2001. For a given date (e.g., 1 January 2000), we have the EPS situations, either analyzed at 0000 UTC or forecast for 1, 2, or more days in advance. In the ideal situation, when we search the NCEP–NCAR reanalyses for the most similar situation to that analyzed by the EPS, we get the same date (here 1 January 2000). When we move to what EPS has forecast for the 1-day lead time, and if we look at the most similar situation in the NCEP–NCAR archive, we expect to find the day immediately after (i.e., 2 January 2000), and this is most generally the case, because at 1-day lead time the EPS forecast is very good and is consistent with the true situation observed (and reanalyzed) as it appears in the NCEP–NCAR archive for 2 January 2002. When we

Unauthenticated | Downloaded 03/10/21 09:06 AM UTC

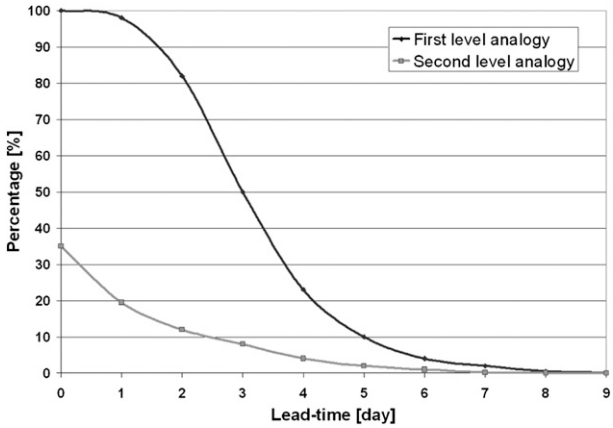


FIG. 3. Percentage evolution with the lead time of the matching rate between the date proposed by the EPS forecast's control member and the date of its best observed analog found in the NCEP–NCAR archive.

move to longer lead times (e.g., the 6-day lead time), then what has been forecast by the EPS model on 1 January 2000 for 7 January 2000 may not match what has been observed, as it appears in the NCEP–NCAR archive for 7 January 2000, and we may consider that its most similar observed situation corresponds to another date (e.g., 17 February 1999). Figure 3 shows this matching rate when only the geopotential fields are considered (first-level analogy curve) and when the relative humidity is added as a predictive variable (second-level analogy curve). When considering only geopotential fields, there is no difference between the NCEP–NCAR reanalyses and the EPS analyses with a lead time equal to 0 (day $D$): the best analog is always the day $D$ itself. This behavior remains similar when 1-day lead-time EPS control members are compared with NCEP–NCAR reanalyses (98% agreement), but the performance decreases with increasing lead time (50% agreement for day $D + 3$).

The analogy on 500-hPa fields apparently performs better than on 1000-hPa fields (not shown in Fig. 3). This is probably due to the more-precise topography used by ECMWF. When humidity fields are used for the second level of analogy, however, the differences between EPS control members and NCEP–NCAR reanalyses become immediately more significant. Only 35% of the days during 1997–2001 prove to be analogs of themselves in the two meteorological archives, even at lead time 0. In fact, differences in the humidity field statistics explain these results. For example, the mean value of RH850 calculated over the whole considered period is 66% with the EPS forecast but declines to 62% with the NCEP–NCAR reanalyses, and the standard deviations are respectively 21.5% versus 22.5%. Moreover, the correlation squared $R^2$ between the two sources is only 0.8.

TABLE 2. Evolution of the best number of analogs to select with lead time.

| | | Lead time (days) | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| One-level analogy | | 30 | 30 | 40 | 60 | 120 | 680 | 820 |
| Two-level analogy | First level | 60 | 60 | 80 | 120 | 230 | 500 | 860 |
| | Second level | 30 | 30 | 40 | 60 | 170 | 350 | 860 |

This shows that the implementation of the analog approach is dependent on the selected NWP model, first because of a possible departure in the consistency between the same fields, either reanalyzed or forecast by the model (differences in resolutions and/or parameterization), and second because of its own decay in performance with lead time.

### 2) REOPTIMIZATION OF THE ANALOG APPROACH FOR REAL-TIME APPLICATIONS

The available forecast archive is too short for reoptimizing the whole analog approach, that is, to reconsider which are the best predictive variables, the best analogy windows, and the best hours of observations. Thus, these parameters were kept as selected in the perfect-prognosis learning context and only the decay in performance with extending lead times has been addressed here by reoptimizing the optimal number of analogs to select. To assess the impact of lead time on performance, the analog approach was performed each day $D$ and for each lead time, from $D$ up to $D + 10$. For each lead time, the number of analogs to select has been reoptimized, either for the single-level forecast (ana24-M2, using geopotential only) or the two-level forecast (ana24-M3, using also relative humidity information).

Table 2 highlights that the decay in performance of the NWP model must be compensated by a progressive and nonlinear increase in the number of analogs to be retained. Forecasts are less and less accurate with lead time, and therefore the number of analogs should increase to account for uncertainty, and the forecast distribution comes progressively closer and closer to the climatological case: for the first lead time, 30 analogs are selected with the one-level analogy algorithm, representing about 1/200th of the climatological sample; when a 6-day lead time is considered, up to 860 analogs (one-eighth of the climatological sample) have to be taken into account for the two-level analogy. In the case of longer lead times, forecasts would provide no information and the best PQPF becomes therefore the climatological distribution of the catchment-averaged precipitation, which can be empirically approached by selecting all of the members of the meteorological archive as analogs. Next, although this will not be detailed here, it appears that the inclusion of relative humidity information in the second level of analogy

Unauthenticated | Downloaded 03/10/21 09:06 AM UTC

is no longer useful beyond 3 days: although this information has proven to be valuable in perfect-prognosis conditions, its forecasts by NWP become too poor to be useful beyond this lead time.

### b. Operational implementation

The operational version of the analog approach, called RainFAST, has been implemented at the Laboratoire des Études des Transferts en Hydrologie et Environnement (LTHE) to provide PQPFs by the analog technique in real time since autumn of 2002. The meteorological fields used for the adaptation are downloaded by a file transfer protocol (ftp) connection to the National Oceanic and Atmospheric Administration (NOAA) Operational Model Archive and Distribution System (NOMADS) servers (Rutledge et al. 2006). The fields are extracted from the 0000 UTC run of the Global Forecast System (GFS) NWP model, over a grid with a horizontal resolution of 1°, and then are regridded to match the 2.5° resolution of the NCEP–NCAR reanalysis that constitutes our meteorological archive. Next, the two ana24-M2 and ana24-M3 algorithms are activated to provide daily rainfall distributions given from 0600 UTC for 24 h ahead (day $D$), for the next 24 h (day $D + 1$), and so on, until day $D + 6$, for each of our targeted catchments.

Some of these PQPFs are displayed on the Cévennes-Vivarais Mediterranean Hydrometeorological Observatory (CVMHO) Internet site (http://www.ohmcv.fr/P751_analogues.php). This prototype runs automatically every day at about 0700 UTC since September 2002. The real-time PQPFs archive unfortunately contains some misses that may be due to the unavailability of GFS output, to failure of the ftp connection, and/or to a default in the local computer network. Nevertheless, all of the misses since January 2005 have been filled in with data from the NOMADS server at the National Climatic Data Center (NCDC; http://nomads.ncdc.noaa.gov/), allowing a continuous record and a complete evaluation in time.

## 4. Evaluation of daily PQPF

Hydrological operational services need to anticipate future precipitation, which can be provided by PQPFs. The PQPFs can be useful for detecting the coming storm events in two respects: early warning is generally based on the exceedance of a given rainfall amount threshold, and the whole rainfall distribution is needed by ensemble hydrological prediction systems aimed at forecasting future discharges at some particular outlets. Thus, an evaluation attempting to bring to forecasters an informative measure of the PQPF quality needs to take into account these two different aspects: early-warning verification and rainfall-distribution verification. It also has to provide an objective way to improve the whole forecasting system, notably if any bias is detected. These two points are discussed in sections 4a and 4b.

TABLE 3. Contingency table.

| | Obs | |
|---|---|---|
| Forecast | Yes | No |
| Yes | $a$ | $b$ |
| No | $c$ | $d$ |

### a. The reference: Observed rainfall dataset

The observed daily precipitation dataset is extracted from the CVMHO database by means of the System for Data Extraction and Visualization On Line (SEVnOL) interface (Boudevillain et al. 2011). The database is made up of data from 103 hourly rain gauges displayed in Fig. 1 over the catchments detailed in section 2a. This dense rain gauge network of about one station per 15–20 km$^2$ allows one to compute catchment-averaged rainfall by kriging hourly rainfall observations and then aggregating the obtained hourly rainfall fields into daily ones. The archive is only available for autumn months (September–December) over the period 2000–08. Thus, the forecast evaluation has been accomplished over the autumns of 2005–08, which is the common period of avalability of observed rainfall fields and real-time PQPFs.

### b. Early-warning verification

#### 1) CONTINGENCY TABLES AND SCORES

Early-warning verification is based on the determination of the best quantile to be used to detect a rainfall event as defined by a given threshold on the transformed precipitation $R$. Verification scores derived from contingency tables (Table 3) are commonly applied to assess categorical forecasts of an event. Here, three different scores are used to evaluate operational PQPFs provided by the analog approach (Joliffe and Stephenson 2003): the probability of detection (POD), the specificity (SPE; often called PODn, i.e., probability of detecting a no event, which is also equal to 1 − POFD, where POFD is the probability of false detection), and the true skill statistic score (TSS; also called the Hanssen–Kuipers discriminant or Peirce skill score):

$$\mathrm{POD} = \frac{a}{a + c}, \tag{8}$$

$$\mathrm{SPE} = \frac{d}{b + d}, \quad \text{and} \tag{9}$$

$$\mathrm{TSS} = \frac{ad - bc}{(a + c)(b + d)} = \mathrm{POD} - (1 - \mathrm{SPE}), \tag{10}$$

Unauthenticated | Downloaded 03/10/21 09:06 AM UTC

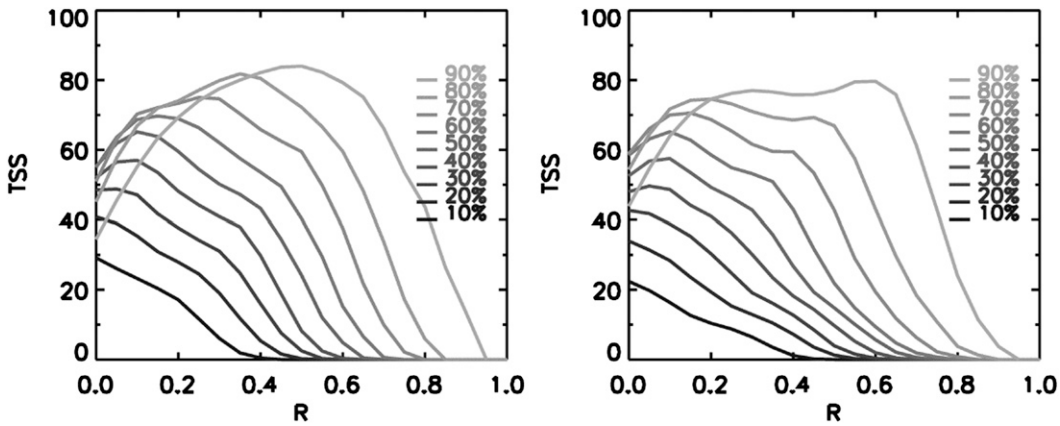


FIG. 4. TSS (%) evolution, according to a considered threshold on the transformed rainfall $R$, for the (left) ana24-M2 and (right) ana24-M3 algorithms. The considered lead time is $D$ (1 day).

with $a$, $b$, $c$, and $d$ defined in Table 3 and $a + b + c + d$ equal to the total number of evaluated forecasts.

POD is the proportion of events that are correctly forecast. SPE corresponds to the probability of correctly issuing a nonevent conditional to the event not being observed. TSS summarizes the previous scores by assessing the correct alert and the correct rejection: it is equal to 1 when the forecasts are perfect. In the following, TSS is used to determine which quantiles are the best quantiles to be used for rainfall-event detection—that is, to determine the ones that maximize the probability of event detection and minimize the probability of false alerts.

2) WHAT ARE THE BEST QUANTILES FOR EARLY WARNING?

Figure 4 displays the evolution of TSS scores for the ana24-M2 and ana24-M3 algorithms as a function of the transformed precipitation $R$ threshold, for several PQPF quantiles and a lead time of 24 h ahead (day $D$). It emerges that the upper quantile value $Q_{60\%}$ is the best indicator for forecasting rain/no rain. Indeed, this quantile is a good compromise between event detection and false-alert issuance. TSS scores for the upper quantile value $Q_{60\%}$ reach 55% for ana24-M2 and 60% for ana24-M3. The upper quantile value $Q_{90\%}$ is more convenient for intense-rainfall events, here defined as events with $R \geq 0.5$ (i.e., $P \geq 0.25P_{10}$, or one-quarter of the 10-yr rainfall). The corresponding TSS values are respectively 81% and 78% for ana24-M2 and ana24-M3, showing that the analog method performs better for intense-rainfall-event detection and that the relevance of using relative humidity information from EPS remains debatable. Figure 5 shows the evolution of TSS for $Q_{60\%}$ and $Q_{90\%}$, respectively, with lead times from 1 day ahead ($D$) to 7 days ahead ($D + 6$). As expected, TSS drops away with lead time: $D$ and $D + 1$ have similar performance for both quantiles; the TSS scores then decrease slightly for $D + 2$ and $D + 3$ and significantly fall beyond $D + 4$.

To summarize, forecasters should look at $Q_{60\%}$ to detect rain/no rain and at $Q_{90\%}$ to forecast events with high rainfall amounts. The global performance of the analog-based PQPFs decreases with lead time and especially after day 5 ($D + 4$; i.e., from +96 to +120 h).

### c. Daily rainfall distribution verification

Hydrological ensemble prediction systems require ideally the complete rainfall distributions as input (e.g., Marty et al. 2008). Thus, PQPFs have to be assessed as a whole in comparison with the observed rainfall. The verification score applied here is the CRPSS already used for optimizing the method. Reference forecasts are taken again as the climatological observed rainfall distributions for each considered catchment.

When all the days of the archive are considered for the verification, the two analog algorithms (ana24-M2 and ana24-M3) have similar performance with respect to the reference climatological conditions (Fig. 6). The maximum skill scores are obtained for day $D$, where CRPSS values are 42%. Then, the CRPSS decreases with lead time and reaches only 5% on day $D + 6$. On the contrary, when a 25% threshold is considered on the observed $R$ (i.e., when only the more intense rainy days are considered), the performances of the two algorithms differ. For day $D$, CRPSS values are 62% and 55% for ana24-M2 and ana24-M3, respectively. From this lead-time $D$ to day $D + 4$, the CRPSS departures between the two algorithms are about 10%. The CRPSSs become equal on day $D + 6$, when the score is about 20% for both methods.

Once again, one can conclude that PQPFs provided by the analog approach are more efficient with high rainfall amounts. This could be expected from the calibration process itself. Indeed, the criterion used for optimizing

Unauthenticated | Downloaded 03/10/21 09:06 AM UTC

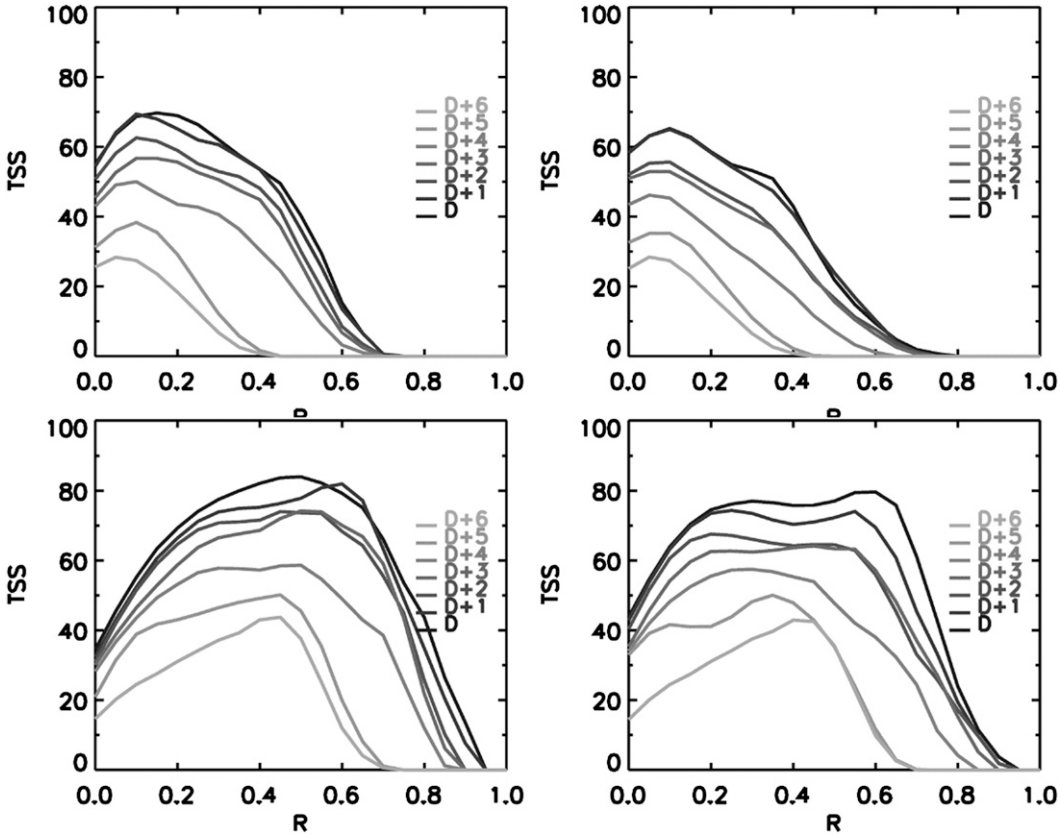


FIG. 5. TSS (%) evolution, according to a considered threshold on the transformed rainfall $R$, for lead times $D$ (1 day) up to $D + 6$ for (left) ana24-M2 and (right) ana24-M3, and for the upper (top) $Q_{60\%}$ and (bottom) $Q_{90\%}$ quantiles.

the method was to minimize over the whole learning period the average of each daily CRPS. These values can be much higher for rainy days, with a widespread distribution, rather than for days without rain or days that are slightly rainy, where the distribution is sharp, close to zero, and often close to the observed value. Thus, the optimization algorithm tends to favor the reduction of the criterion mainly on rainy days, therefore giving a better performance when only these days are looked at.

Note also the performance calculated for days $D$ and $D + 1$ over the whole evaluation period (i.e., without fixing a threshold on $R$), which is higher than that of the perfect-prognosis conditions. This is certainly due to the evaluation sample, which is based on the rainiest season of the Cévennes-Vivarais region (autumn).

## 5. Reliability evaluation and bias correction

### a. Reliability definition

To feel confident about a probabilistic forecasting system, forecasts have to be statistically consistent with the observations. This implies that the forecast probabilities are statistically reliable (i.e., that they match the observed ones).

In general, the reliability of continuous-type probability forecasts is evaluated by considering different forecast classes $i$: the forecasting system is called reliable if the forecast frequency distributions $F_i$ are consistent with the corresponding frequency distributions of observations $O_i$ for all of the considered classes $i$. The great variety of rainfall distributions issued by the analog approach makes such an analysis difficult, as highlighted by Jolliffe and Stephenson (2003): ''For example, when evaluating the reliability of continuous-type probability forecasts one has to decide when two forecast distributions are considered as the same. Grouping (pooling) more diverse forecast cases into the same category will increase sample size but can potentially reduce useful forecast verification information.''

Here, reliability evaluation is based on a more accessible process that is applied on quantiles rather than on the whole distribution: each quantile frequency and the relative observed frequency are compared. Quantiles are considered at regular intervals of 10% and are

Unauthenticated | Downloaded 03/10/21 09:06 AM UTC

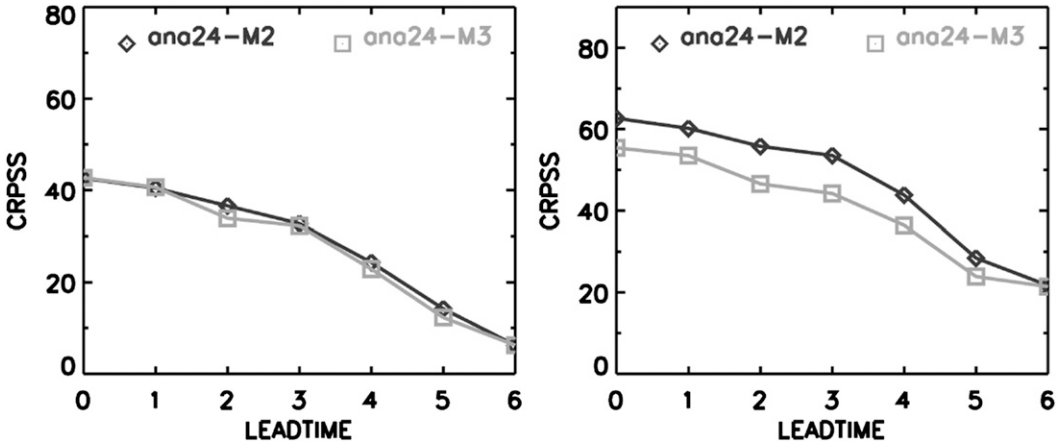


FIG. 6. CRPSS (%) of real-time PQPFs provided by the ana24-M2 and ana24-M3 algorithms: (left) without threshold and (right) with a threshold on the transformed precipitation $R$: $R \geq 0.25$. The considered reference is the climatological rainfall distribution issued from the available catchment-averaged precipitation archive. Lead time is expressed in days.

extracted from the forecast distributions of transformed rainfall $R$, namely $F$, by considering the two terms $F(0)$ (the cumulative distribution of no rain) and $F_+$ (the cumulative distribution of strictly positive transformed rainfall values), according to Eq. (5). Three different reliability analyses can then be made: on $F(0)$, on $F_+$, and on the cumulative distribution of strictly positive transformed rainfall values, conditioned to the no-rain frequency: $F_+ \mid F(0)$.

### b. Reliability of real-time PQPFs

Figure 7, applied to day $D$ (24 h ahead), shows that both $F(0)$ and $F_+$ are biased. The no-rain forecast probability $F(0)$ is globally underestimated with regard to that

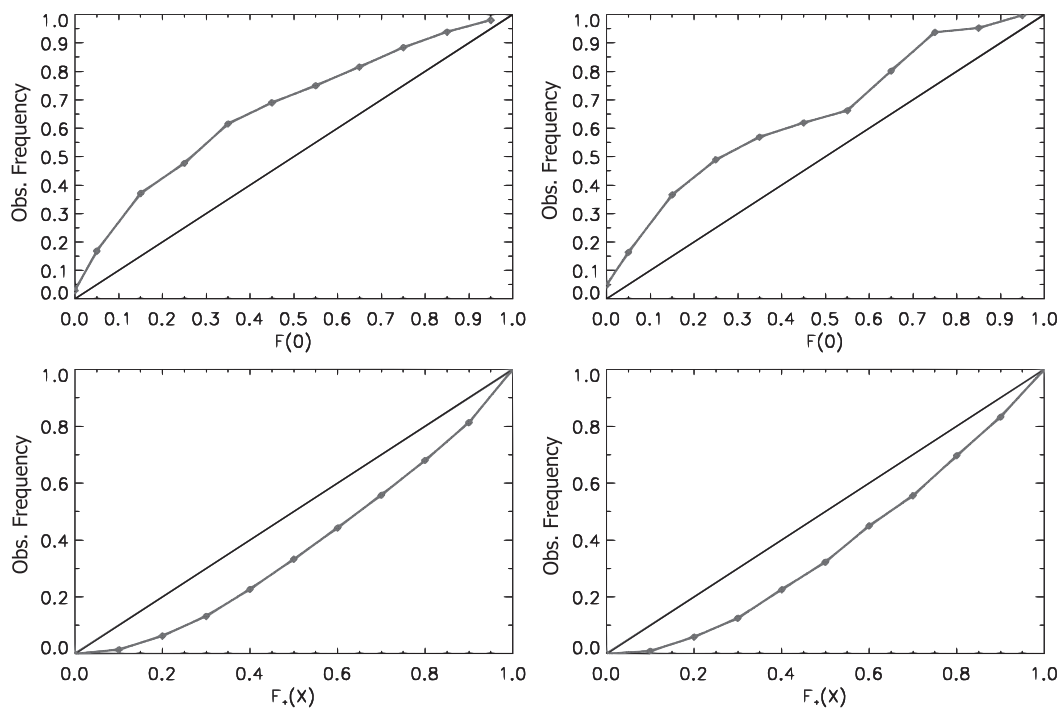


FIG. 7. Reliability evaluation of (top) $F(0)$ and (bottom) $F_+$ for (left) ana24-M2 and (right) ana24-M3 in real-time conditions. Lead time is set to $D$ (1 day), and all catchments are considered.

Unauthenticated | Downloaded 03/10/21 09:06 AM UTC

observed. For example, the observed probabilities associated with the forecast 0% quantile are 3% and 5% for the ana-M2 and the ana24-M3 algorithms, respectively. This bias is more obvious with greater values of $F(0)$. About 70% (for ana24-M2) and 65% (for ana24-M3) of days corresponding to the forecast 50% quantile of $F(0)$ are days with no observed rain.

On the contrary, the forecast positive rainfall values probabilities $F_+$ are overestimated. Indeed, the observed probabilities associated with a given forecast quantile are always lower than expected. For instance, observed precipitation is lower than or equal to the forecast 50% quantile over about 32% of time for both the ana24-M2 and ana24-M3 algorithms. This again confirms that the optimization of the method has favored forecasts of high rainfall rather than a good estimation of the rain/no-rain frequency $F(0)$.

In conclusion, even if PQPFs have proven to be useful with respect to climatological values (see section 4), they are not reliable (i.e., they are not consistent with the observations). Correction of the observed bias requires long and homogenous real-time PQPFs archives that are difficult to collect because of the perpetual evolution of NWP models. Indeed, statistical correction of reliability can only be applied here on PQPFs in perfect-prognosis conditions. This means that only the internal part of the bias (i.e., the bias directly due to the analog approach, not the one that is dependent on the NWP model) will be corrected. The external part, which is due to the NWP model used for generating meteorological forecasts, cannot be assessed here.

### c. *Reliability of PQPF in perfect-prognosis conditions*

#### 1) DIAGNOSIS IN PERFECT-PROGNOSIS CONDITIONS

The evaluation is then performed in perfect-prognosis conditions, on the NCEP–NCAR reanalyses archive, over the period 1953–2001. Unlike the real-time context, both components of the $F$ distribution are mostly unbiased (Fig. 8): the $F(0)$ and $F_+$ curves are close to the bisector. This means, for example, that when $F(0)$ is expected to be equal to 50%, then one-half of the corresponding days in the observed catchment-averaged rainfall archive are actually days with no rain.

The analysis has been completed by evaluating the term $F_+$ conditional on the $F(0)$ value. This additional step reveals that the overall lack of bias of $F_+$ is due to a compensation of bias, depending on the $F(0)$ value, as shown at the bottom of Fig. 8. As an example, when $F(0)$ is equal to 0 (i.e., the day is expected to be rainy), observed rainfall is lower than or equal to the 50% quantile only about 40% of the time, instead of 50%, as expected, suggesting that forecasts underestimate the precipitation. To adjust this frequency to 50%, the forecast 50% quantile has to be increased. In a general way, when the targeted day is expected to be wet [i.e., $F(0) \rightarrow 0$], then the $F_+$ quantiles seem to be underestimated; on the other hand, when the targeted day is expected to have no or little rain [i.e., $F(0) \rightarrow 1$], the $F_+$ quantiles appear to be overestimated. In the following, we have taken advantage of this finding by proposing a correction factor for the $F_+$ quantiles, depending on the dry or wet character of the targeted day.

#### 2) CORRECTION OF THE INTERNAL BIAS OF THE ANALOG APPROACH

A first attempt at correcting this bias is to assume that a perfectly reliable a posteriori distribution $F'_+$ exists and that it can be determined from the a priori distribution $F_+$ by transforming the latter into a new a posteriori distribution, through a given function $\Theta$:

$$F'_+ = \Theta(F_+). \tag{11}$$

Such a function is difficult to obtain because of the wide spectrum of a priori functions $F_+$. Thus, another possible way (followed here) is to correct each quantile $Q_{x\%}$ by a coefficient $\alpha_{F_0,x\%}$ function of $F(0)$ and of the associated $F_+$ probability $x\%$ (Déqué 2007):

$$\forall x \in [0, \ldots, 100]\, Q'_{x\%} = \alpha_{F_0,x\%} Q_{x\%}. \tag{12}$$

Then, the a posteriori distribution $F'_+$ is fitted with a gamma law by the L-moments method.

The obtained remedial factors confirm the previous comment on $F_+$ bias. As an example, Fig. 9 gives the Gardon at Anduze catchment's factors. When the targeted day is expected to be wet [$F(0) \rightarrow 0$], these are greater than 1, corroborating the a priori underestimation of the forecast $F_+$ quantiles. On the other hand, for $F(0) \rightarrow 1$, factors are lower than 1, showing the a priori overestimation of the $F_+$ quantiles. The particular behavior for $Q \geq 80\%$ could be explained by the low dispersion of PQPFs when the day is expected to be without rain.

The a posteriori PQPFs obtained after the quantiles correction have in turn to be evaluated to verify their reliability. The reliability charts show that the a posteriori distributions are more reliable than the a priori ones, for all of the study catchments (Fig. 10). For $F(0) \leq 0.40$ (thus, $F_+ \geq 0.60$), the a posteriori distributions are almost perfectly reliable. When $F(0)$ increases (i.e., $F_+$ decreases in Fig. 10), this comment becomes less and less true for the lower part of the distributions, because of the gamma-law

Unauthenticated | Downloaded 03/10/21 09:06 AM UTC

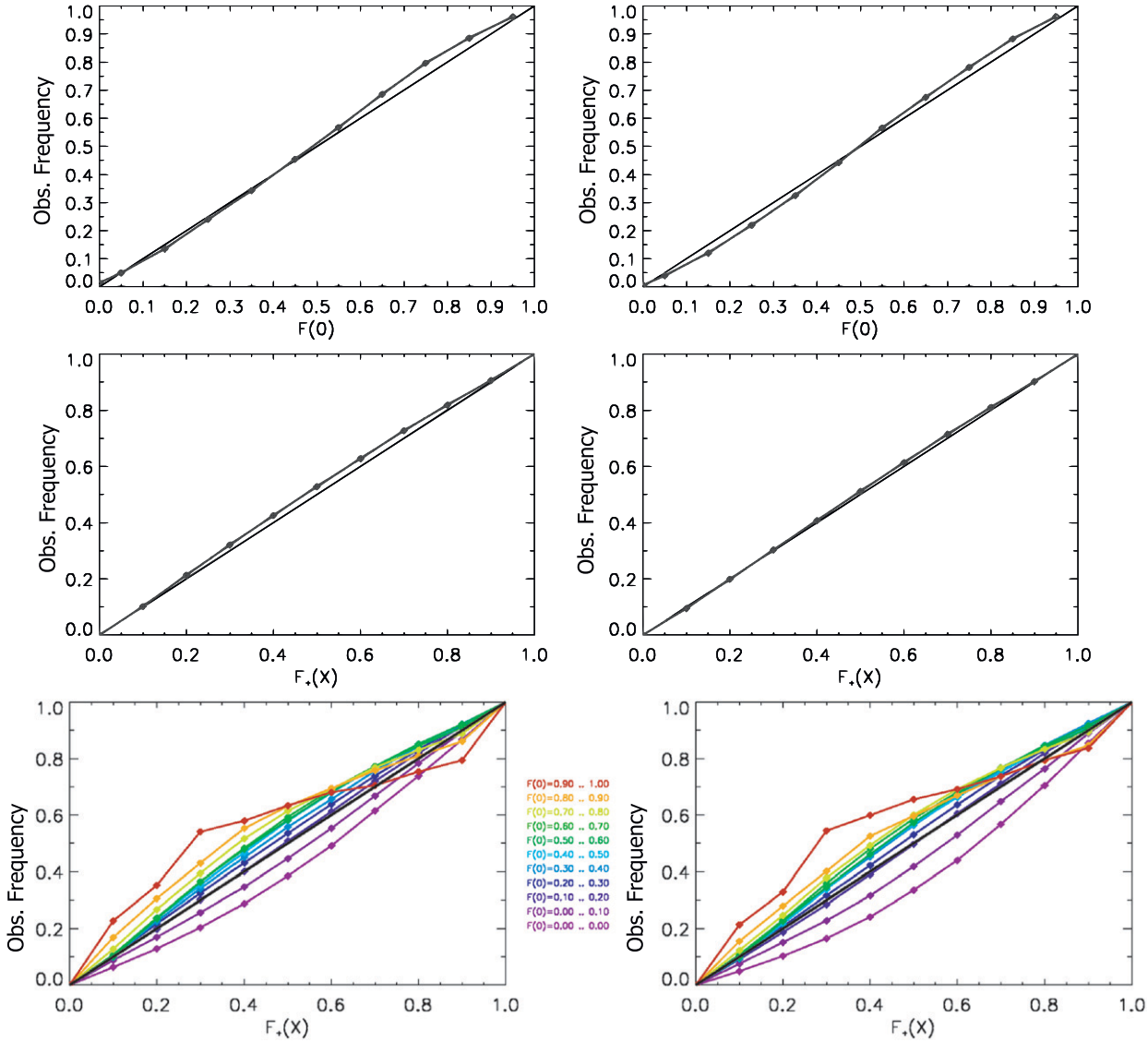


FIG. 8. Reliability evaluation on (top) $F(0)$, (middle) $F_+$, and (bottom) $F_+$ conditional on $F(0)$ for (left) ana24-M2 and (right) ana24-M3 in perfect-prognosis conditions and considering all of the catchments.

fitting by L moments, which introduces some residual errors for lower frequencies.

### 3) IMPACT ON DAILY RAINFALL DISTRIBUTIONS

As an example, Fig. 11 illustrates the impact of such a statistical correction for the PQPF provided in perfect-prognosis conditions around 4 November 1994 for the initial ana24-M3 version (continuous lines with boxes). The corrected PQPF, ana24-M3cb (continuous lines), corresponds to distributions corrected with catchment-specific remedial factors and shows a reduced interquantile range. The a priori 60% quantile is similar to the 20% a posteriori quantile. This example shows that the reliability correction improves also the distribution sharpness. Another version, ana24-M3ce (not shown here), corresponds to distributions using remedial factors common to all catchments and appears to be slightly less efficient.

### 4) INFLUENCE ON REAL-TIME PQPF PERFORMANCE

The previous statistical correction also improves the real-time PQPF efficiency. Figure 12 displays the CRPSS gain between the original versions (ana24-M2 or ana24-M3) and the corrected ones (ana24-M2cb and ana24-M2ce or ana24-M3cb and ana24-M3ce) for the autumns of 2005–08. Algorithms ana24-M2cb and ana24-M3cb correspond to PQPFs corrected by catchment-specific

Unauthenticated | Downloaded 03/10/21 09:06 AM UTC

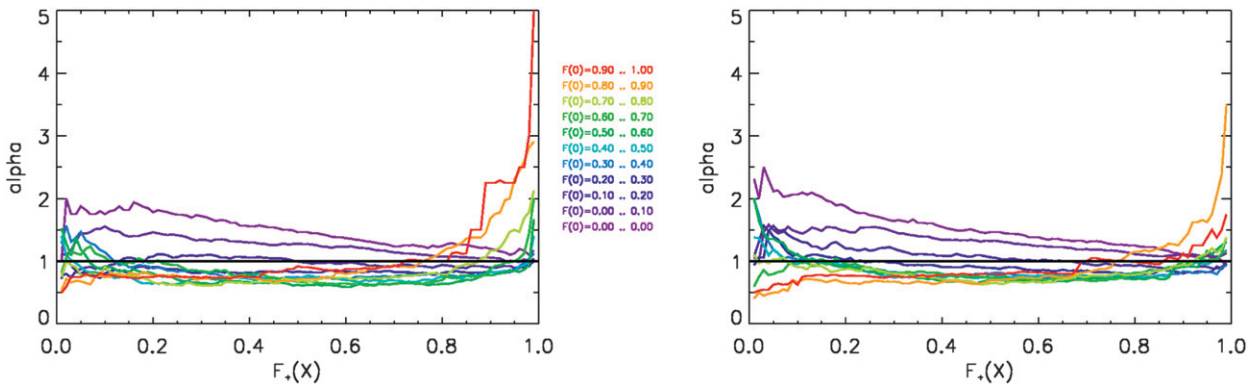


FIG. 9. Remedial factors obtained for the Gardon at Anduze catchment for (left) ana24-M2 and (right) ana24-M3.

remedial factors. A posteriori PQPFs provided by ana24-M2ce and ana24-M3ce are issued after correction with factors common to all of the study catchments. The gain in CRPSS is very slightly positive for all lead times, whatever the correction factors used (either common or catchment specific). The improvement is more significant with higher rainfall amounts, however. With a threshold on $R$ (i.e., $R \geq 0.25$), the gain reaches 5% with catchment-specific factors and reaches 3% with common factors. This gain decreases after day $D + 1$ and becomes negative beyond day $D + 4$, however.

The PQPF sharpness can be assessed through the CRPSS value obtained by substituting the medium quantile for the observed rainfall. Figure 12 outlines that the CRPSS gain is related to a sharpness gain in the skill score. CRPSS gains are about 8% for the first lead times and increase with lead time to reach 13% for day $D + 6$ at the cost of a small decrease in accuracy. To summarize, the statistical correction improves the reliability of PQPFs provided by the analog technique and also increases the sharpness of the rainfall distributions.

## 6. Discussion and conclusions

The analog approach has proven to be a very useful adaptation method for estimating precipitation at the catchment scale, by complementing usefully the information provided by NWP models. Our effort was mainly devoted to the implementation of a daily version into operational real-time conditions on the basis of two different algorithms [making use (ana24-M3) or not (ana24-M2) of relative humidity forecasts] to provide probabilistic quantitative precipitation forecasts. In perfect-prognosis conditions, relative humidity at 850 hPa did appear to be a useful and significant predictor beyond the geopotential fields. When moving to the real-time operational context, M2 showed itself to be more skillful at forecasting high precipitation amounts, whereas M3 seemed to be more powerful for forecasting of no rain or low precipitation amounts. The poorer performance of M3 in the case of heavy-precipitation events, which was initially unexpected, was explained by the fact that relative humidity and precipitable water fields forecast by NWP are less

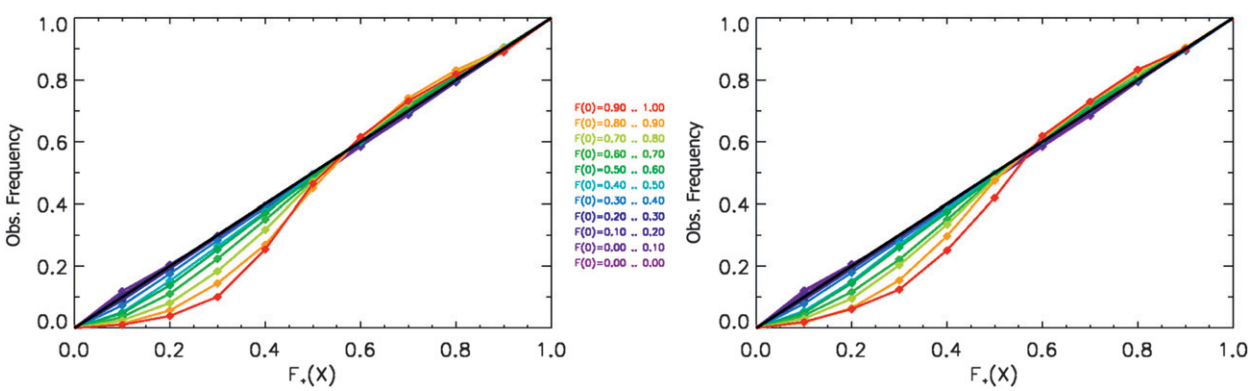


FIG. 10. Reliability evaluation on a posteriori $F_+$ conditional on $F(0)$ for (left) ana24-M2 and (right) ana24-M3. All catchments are considered.

Unauthenticated | Downloaded 03/10/21 09:06 AM UTC

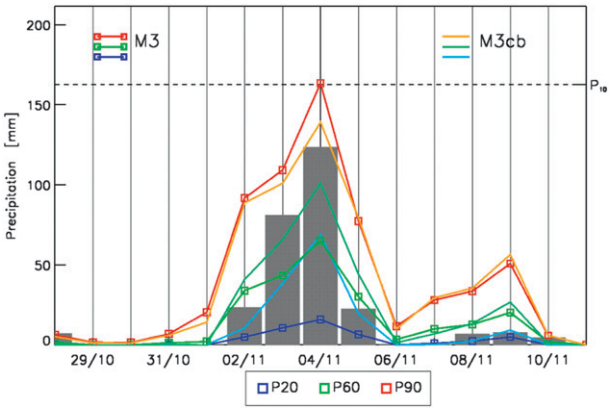


FIG. 11. PQPF provided around 4 Nov 1994 in perfect-prognosis conditions by the ana-M3 (continuous lines with boxes) and the corrected ana24-M3cb (continuous lines) versions. The gray histogram represents observed rainfall. The forecast is given for the Gardon at Anduze catchment.

reliable, and are even useless beyond 3-day lead time, than are the geopotential forecasts (these remain powerful up to 8–10 days ahead). Indeed, both the ana24-M2 and ana24-M3 precipitation distributions were kept as guidance, especially for the first few days of lead time, because of the complementary strengths and weaknesses of the two algorithms.

First, we have examined the dependence of the analog approach on the NWP model performance. A comparison of predictive variables issued from NCEP–NCAR reanalyses by NOAA and by ECMWF EPS control forecasts shows the dependence of analogs on the selected NWP, because of possible differences in terms of truncation, parameterization, and analysis techniques from one NWP model to another. This dependence proves to be more sensitive for thermodynamic variables like relative humidity than for synoptic variables like geopotential fields. Accordingly, the implementation of the analog approach under real-time conditions has required a reoptimization of the numbers of analogs to retain as a function of lead time, because of the growing uncertainty of the NWP output.

Second, the evaluation of the two operational algorithms reveals that this approach is efficient for issuing early warnings by looking at the quantiles $Q_{60\%}$ and $Q_{90\%}$ to detect rain/no rain and heavy rainfall, respectively. PQPFs assessed through the continuous ranked probability score prove to be more efficient in comparison with the climatological distribution, especially for higher rainfall amounts. The performance of PQPFs used as early-warning tools or as input into hydrological ensemble prediction systems decreases, as expected, with lead time. PQPFs provided by this method unfortunately appear still to be biased. In fact, it is suspected that the CRPS and the choice of the optimization procedure could explain this bias in favor of large rainfall amounts.

Last, the internal part of this bias (i.e., the part directly related to the analog approach and not dependent on the NWP model) has been statistically corrected. This postprocessing improves the PQPF reliability by reducing the rainfall distribution dispersion, because corrected PQPFs are sharper than the original ones. Thus, there are favorable grounds for applying such an approach that runs already in real-time conditions at Electricité de France, at the Compagnie Nationale du Rhône, and in several French flood-forecasting services (Service de Prévision des Crues) in the Alps and in the Loire catchments.

Note that we have looked here at the average performance of our analog algorithms, using classical scores for forecast verification, to quantify simply the performance we have qualitatively assessed from our own experience with real-time forecasts. Thus, this study constitutes a first step before more detailed analyses: further works using more recent techniques could improve our understanding of the analog-approach behavior but would require an extensive statistical analysis for hypothesis verification. Notably, we have not considered any confidence interval on our results, nor have we used hypothesis testing, as suggested by Jolliffe (2007) and Gilleland (2010). Also, an EVA-based score such as the extreme dependency score of Stephenson et al. (2008) could have been used to evaluate the performance of the forecasts in terms of extremes, as a complement to our analysis on upper quantiles.

Among the often-raised questions on analog techniques, common ones are that the targeted situation has no exact equivalent in the archive and that no rainfall larger than the maximum contained in the archive can be forecast. This last question is wrong, since our hypothesis is that we provide a conditional distribution based here on a sample of 30. After the fitting of a probability model to this sample, we can issue for instance the 99% quantile, which has never been observed but which has a 1% chance to appear, in much the same way that one derives the 100- or even the 500-yr rainfall by EVA methods from a 50-yr sample of the annual maximum. The first question seems to be more problematic, however, in that our analog rainfall samples are not fully similar to a classical sample of the theoretical expected distribution, since they come from analogs that are individually more or less similar to the targeted situation. This means that only the most similar situations, below a certain threshold in terms of the Teweless–Wobus S1 score, should be kept or that a weighting should be used according to the similarity criterion of each selected analog.

Unauthenticated | Downloaded 03/10/21 09:06 AM UTC

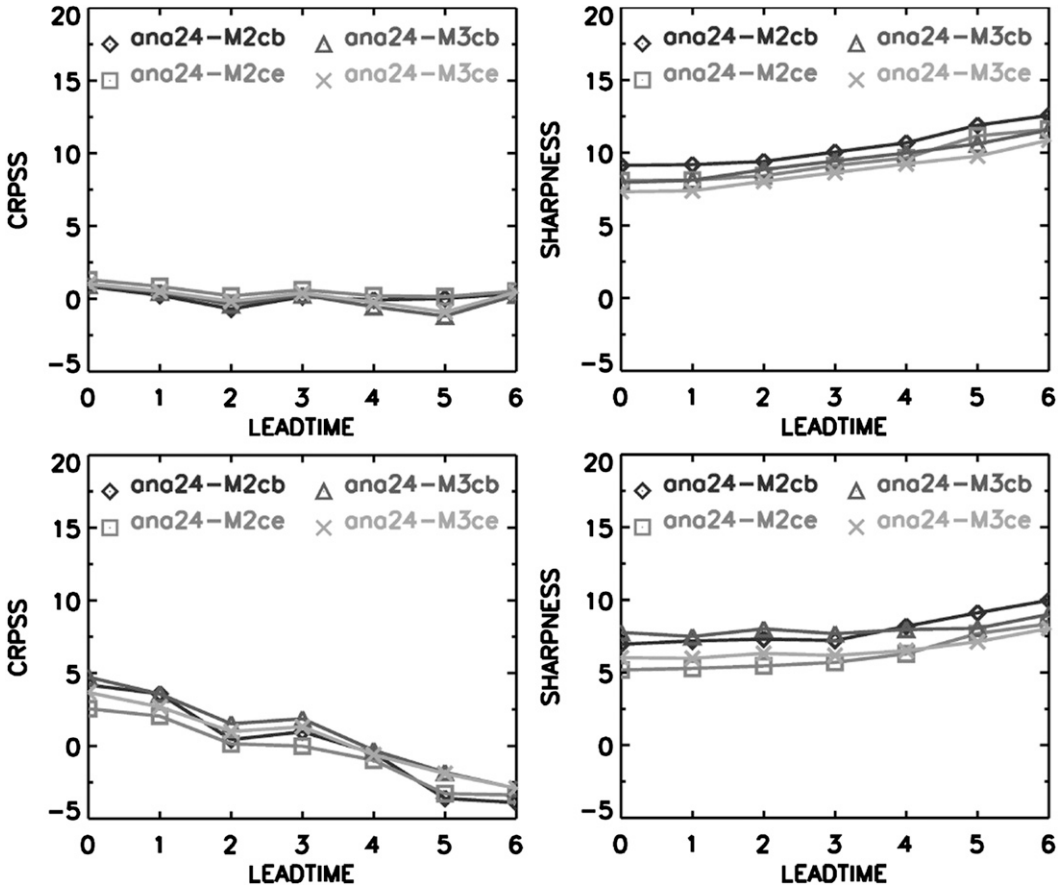


FIG. 12. Gain of (left) CRPSS (%) and (right) sharpness obtained by the statistical correction with catchment-specific factors (cb) or common factors (ce) of the analog algorithms without (M2) and with (M3) a relative humidity analogy. (top) No threshold and (bottom) a 25% threshold on the transformed rainfall $R$ ($R \geq 0.25$) are considered. The references used in CRPSS are respectively the a priori ana24-M2 for ana24-M2cb and ana24-M2ce as well as the a priori ana24-M3 for ana24-M3cb and ana24-M3ce. Lead time is expressed in days.

Several other ways could also be explored for this alternative approach to provide more efficient PQPFs. Up to now, the targeted synoptic situation has always been defined on a fixed 0–24 h (UTC) temporal window and the analogs have only been searched among situations defined also on the same 0–24-h window. Relaxing this constraint by allowing a moving temporal window during the analog searching procedure would allow one to select, for example, a 6–30-h situation, with a better match to the targeted one. This should also improve the similarity of the retained analog situations. Nevertheless, this would require being able to propose the corresponding 24-h rainfall on a moving temporal window and therefore also having a 6- or 12-h rainfall archive, which is difficult to build up far back in the past (e.g., over a 50-yr period).

Another question raised by the bias in the estimation of the $F(0)$ rain/no-rain frequency is whether the approach can be optimized for the correct estimation of both high rainfall amounts and rain/no-rain probability. It could not be the same variables that control these two different phenomena, and therefore one may think of a splitting into a two-level procedure, using different predictive variables that are optimized separately depending on the forecast variable. Therefore, on the basis of the above discussion, our major conclusion is that many further aspects remain to be explored and that there is promising potential in the development and the evaluation of the analog approach.

*Acknowledgments.* This work has been supported by the French Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations (SCHAPI). It has also benefited from free access to the NOMADS servers hosting GFS output (NOAA, NCDC, and NCEP) and the NCEP–NCAR reanalysis. ECMWF is thanked for the access to the MARS forecasting archive. The operational precipitation dataset was provided by the CVMHO

Unauthenticated | Downloaded 03/10/21 09:06 AM UTC

through the SEVnOL interface, and N. Thevenot is acknowledged for his contribution to the adaptation to real time of the analogy-based PQPF. The authors thank the three anonymous reviewers for their valuable comments and suggestions that improved the final form of the paper.

## REFERENCES

Bliefernicht, J., and A. Bárdossy, 2007: Probabilistic forecast of daily areal precipitation focusing on extreme events. *Nat. Hazards Earth Syst. Sci.,* **7,** 263–269.

Bontron, G., and Ch. Obled, 2005: A probabilistic adaptation of meteorological model outputs to hydrological forecasting. *Houille Blanche-Rev. Int. Eau,* **1,** 23–28.

Boudevillain, B., G. Delrieu, B. Galabertier, L. Bonnifait, L. Bouilloud, P.-E. Kirstetter, and M.-L. Mosini, 2011: The Cévennes-Vivarais Mediterranean Hydrometeorological Observatory database. *Water Resour. Res.,* **47,** 6.

Brankovic, C., B. Matjacic, S. Ivatek-Sahdan, and R. Buizza, 2008: Downscaling of ECMWF ensemble forecasts for cases of severe weather: Ensemble statistics and cluster analysis. *Mon. Wea. Rev.,* **136,** 3323–3342.

Buizza, R., J. R. Bidlot, N. Wedi, M. Fuentes, M. Hamrud, G. Holt, and F. Vitart, 2007: The new ECMWF VAREPS (Variable Resolution Ensemble Prediction System). *Quart. J. Roy. Meteor. Soc.,* **133,** 681–695.

Castillo, E., 1988: *Extreme Value Theory in Engineering*. Academic Press, 389 pp.

Clark, M., S. Gangopadhyay, L. Hay, B. Rajagopalan, and R. Wilby, 2004: The Schaake shuffle: A method for reconstructing space–time variability in forecasted precipitation and temperature fields. *J. Hydrometeor.,* **5,** 243–262.

Delrieu, G., and Coauthors, 2005: The catastrophic flash-flood event of 8–9 September 2002 in the Gard region, France: A first case study for the Cevennes-Vivarais Mediterranean Hydrometeorological Observatory. *J. Hydrometeor.,* **6,** 34–52.

Déqué, M., 2007: Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values. *Global Planet. Change,* **57,** 16–26.

Diomede, T., F. Nerozzi, T. Paccagnella, and E. Todini, 2008: The use of meteorological analogues to account for LAM QPF uncertainty. *Hydrol. Earth Syst. Sci.,* **12,** 141–157.

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

Gilleland, E., 2010: Confidence intervals for forecast verification. NCAR Tech. Note NCAR/TN-479+STR, 78 pp.

Gumbel, E. J., 1958: *Statistics of Extremes*. Columbia University Press, 1988 pp.

Hamill, T. M., J. S. Whitaker, and S. L. Mullen, 2006: Reforecasts—An important dataset for improving weather predictions. *Bull. Amer. Meteor. Soc.,* **87,** 33–46.

——, R. Hagedorn, and J. S. Whitaker, 2008: Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part II: Precipitation. *Mon. Wea. Rev.,* **136,** 2620–2632.

Houtekamer, P. L., L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.,* **124,** 1225–1242.

Jolliffe, I. T., 2007: Uncertainty and inference for verification measures. *Wea. Forecasting,* **22,** 637–650.

——, and D. Stephenson, 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley and Sons, 254 pp.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 341 pp.

——, and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Leutbecher, M., and T. N. Palmer, 2008: Ensemble forecasting. *J. Comput. Phys.,* **227,** 3515–3539.

Marsigli, C., A. Montani, F. Nerozzi, T. Paccagnella, S. Tibaldi, F. Molteni, and R. Buizza, 2001: A strategy for high-resolution ensemble prediction. II: Limited-area experiments in four Alpine flood events. *Quart. J. Roy. Meteor. Soc.,* **127,** 2095–2115.

——, F. Boccanera, A. Montani, and T. Paccagnella, 2005: The COSMO–LEPS mesoscale ensemble system: Validation of the methodology and verification. *Nonlinear Processes Geophys.,* **12,** 527–536.

Marty, R., I. Zin, and Ch. Obled, 2008: On scaling PQPFs to fit hydrological needs: The case of flash flood forecasting. *Atmos. Sci. Lett.,* **9,** 73–79.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF Ensemble Prediction System: Methodology and validation. *Quart. J. Roy. Meteor. Soc.,* **122,** 73–119.

Murphy, J., 1999: An evaluation of statistical and dynamical techniques for downscaling local climate. *J. Climate,* **12,** 2256–2284.

Obled, Ch., G. Bontron, and R. Garçon, 2002: Quantitative precipitation forecasts: A statistical adaptation of model outputs through an analogues sorting approach. *Atmos. Res.,* **63** (3–4), 303–324.

Rutledge, G. K., J. Alpert, and W. Ebuisaki, 2006: NOMADS: A climate and weather model archive at the National Oceanic and Atmospheric Administration. *Bull. Amer. Meteor. Soc.,* **87,** 327–341.

Stephenson, D. B., B. Casati, C. A. T. Ferro, and C. A. Wilson, 2008: The extreme dependency score: A non-vanishing measure for forecasts of rare events. *Meteor. Appl.,* **15,** 41–50.

Thévenot, N., 2004: Prévision quantitative des précipitations par une méthode d'analogie: Utilisation de la prévision d'ensemble du CEPMMT. LTHE, Institut National Polytechnique de Grenoble, M.Sc. thesis, 58 pp.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.,* **125,** 3297–3319.

Wilby, R., O. Tomlinson, and C. Dawson, 2003: Multi-site simulation of precipitation by conditional resampling. *Climate Res.,* **23,** 183–199.

Unauthenticated | Downloaded 03/10/21 09:06 AM UTC