ELSEVIER

Atmospheric Research 63 (2002) 303–324

ATMOSPHERIC
RESEARCH

www.elsevier.com/locate/atmos

Quantitative precipitation forecasts: a statistical adaptation of model outputs through an analogues sorting approach

Charles Obled [a,*], Guillaume Bontron [a], Rémy Garçon [b]

[a] *Laboratoire d'étude des Transferts en Hydrologie et Environnement, Institut National Polytechnique de Grenoble, B.P. 53, 38041 Grenoble, Cedex 9, France*
[b] *Division Technique Générale, Electricité de France, Grenoble, France*

Received 15 November 2001; received in revised form 19 March 2002; accepted 12 April 2002

Abstract

Medium-term quantitative precipitation forecasts (QPFs) up to several days ahead are required to issue early flood warnings and to allow optimum operation of hydraulic structures or reservoirs. This paper describes an approach which can be seen as an adaptation of deterministic meteorological model outputs. It involves searching for a sample of past situations similar to the current one from a long meteorological archive. The analogy is considered in terms of general circulation patterns over a window covering western Europe. For this restricted sample of days similar to the day at hand, the corresponding sample of observed daily precipitation is extracted for each catchment. The rainfall to be observed during the current day is assumed to follow the same distribution, known from this empirical sample. This provides a probabilistic forecast expressed, for example, by a central quantile and a confidence range. This paper describes the many choices underlying the optimisation of this approach: choice of predictor variables to characterise a meteorological situation, choice of similarity criterion between two situations, criterion for performance evaluation between two versions of the algorithm, etc. This method was calibrated over about 50 catchments located in France, Italy and Spain, using a meteorological and hydrological archive running from 1953 to 1996. Comparisons carried out over a validation sample (1995–1996) with three poor-man methods prove the interest of this approach, in a perfect prognosis context. In real-time operation, the use of forecast instead of observed predictor variables, essentially geopotential fields, produces only a minor decrease in performance. The use of the single-valued central quantile supplemented by the confidence interval

* Corresponding author.
E-mail address: charles.obled@hmg.inpg.fr (C. Obled).

PII: S0169-8095(02)00038-8

provided a QPF that has proved effective and informative on the potential for extreme values.

Keywords: Quantitative precipitation forecast; QPF; Probabilistic forecast; Model output statistic; Analogues sorting

1. Introduction

In a recent bulletin of the American Meteorological Society, the report of a meeting concerning the prospective U.S. Weather Research Program (Emanuel et al., 1995) stressed the need for "an optimal blend of numerical and statistical approaches" to forecast "basin integrated precipitation". It also stressed the strong demand for medium-term precipitation forecasts, allowing early warnings and timely alerts to be issued. This need is deeply felt by the hydrological community, since the operational management of flood warnings is often limited by the shortage of time. For many catchments subject to flash floods, the concentration time is too short to allow effective action to ensure public safety or protect property based only on the rainfall that has already fallen. Many systems relying only on such observed rainfall, or even on nowcasting 1 or 2 h ahead, are far from effective in triggering an alert in due time: according to Zevin (1994), about 50% of the warnings issued by the National Weather Service were received after the flood had occurred. Ideally, the lead-times required for an alert should be at least in the range of 12 to 24 h before the event, while an appropriate early warning for operational or safety services would be 2 or 3 days ahead. Furthermore, the management of flood-retarding structures, like the partial emptying of a reservoir, may take several days to avoid damage downstream or to the structure itself.

This paper describes an approach which addresses these problems by providing quantitative precipitation forecasts at basin scale for up to 6 days ahead.

Currently, meteorological forecasts are provided by deterministic models. As far as synoptic circulation variables are concerned, they perform reasonably well up to 3 days or more ahead (e.g. Atger, 2000). However, focusing on quantitative precipitation forecasts (QPFs), they display a significant drop in accuracy beyond the first 24 h. Furthermore, such operational models often use spatial resolutions that are too large for hydrological needs: grid spacings of the order of 15×15 km^2 at the most do not allow the irregular limits of medium-sized mountain catchments to be followed nor site-specific features to be taken into account. Only research models, or in operational use, limited area models, are able to work at resolutions of about 3×3 km^2. With the latter, the need for initial and lateral boundary conditions usually taken from larger scale models counterbalances the gain in resolution and description of physical processes and partly erodes their medium-term performance (Warner et al., 1997).

Lastly, all such models provide a deterministic, single-valued forecast which does not display any uncertainty or residual variance on the predicted variable. In the future, ensemble forecasting should partly overcome this limitation, and allow the use of probabilistic forecasts that are better suited to the use of optimal decision-making techniques by end-users (Krzysztofowicz, 1998).

However, these requirements can already be fulfilled by alternative forecasting techniques known as «statistical adaptation of model outputs», or MOS (Glahn and Lowry, 1972). Most often, such methods directly address the variable of interest, e.g. the amount of precipitation provided by the model, and try to correct the raw model value. Usually, they are based on an antecedent set of observed and model output values, which can be related through regression analyses or more sophisticated techniques. However, such methods are often limited by the frequent changes occurring in a given model, which do not allow sufficiently long periods of stationary state to occur to base statistics upon them.

Another, although less frequent, choice considers that the adaptation does not need to be made on the end variable of interest, but may start earlier in the modelling chain. Our purpose here is to describe one of these approaches which has been used operationally for more than 30 years by the French power company, Electricité de France, for managing its reservoirs, in particular during flood events.

In a nutshell, one considers that deterministic models provide reliable forecasts of large-scale hydrodynamic variables, like geopotential fields. But it is commonly considered that the physics they require to manage atmospheric moisture and proceed further in time is less reliable and can be by-passed by statistical analysis to provide the end variable of interest, namely basin-wide precipitation. The method described here uses an analog sorting approach. It provides a forecast based on a set of past historical situations that are as similar as possible to the situation at hand for which a forecast is needed. A complete illustration of its performance in real-time, as well as a comparison with precipitation outputs from an operational model during the MAP experiment, are given in another paper (Djerboua and Obled, in press).

This method has been in use for 30 years but was extensively reviewed in 1998, partly to participate in the MAP experiment. This paper recalls the general principles of this forecasting method, but also the new data processing and algorithmic choices made on that occasion. The optimisation and calibration of certain parameters are described as well as the performance obtained off-line by validation on independent test periods and in "perfect prognosis" conditions. Finally, some perspectives will be outlined, given the availability of new data.

2. General principles of statistical adaptation by analog search

Historically, the method was initiated in the early 1970s by Duband (1970, 1980) as a pattern recognition technique. It tries to mimic the forecasting and learning process implicitly used by human forecasters but also to make it more objective and reproducible. Considering that «similar» general circulation patterns should provide «similar» local effects, the search for past situations similar to the one at hand should provide hints on what could happen locally in terms of, e.g. amounts of precipitation for this situation. This approach, summarised by Lorenz (1969), assumes (i) that there have been synoptic situations in the past that were not necessarily identical but similar to the current one for which a forecast is required, (ii) that during such situations, local variables, such as precipitation over a medium-sized mountain catchment, react partly in response to the synoptic situation, but also to more local features (e.g. orography, wind channelling, etc.) and (iii) that for this given day, the part explained by regional circulation will be similar to that observed in analogous situations.

So if one considers a given day for which a precipitation forecast is needed over specified catchments, the approach proceeds in two steps. First, given the synoptic patterns characterising this day (either observed or taken from a meteorological model), the most similar past situations are looked up in a long meteorological archive, thus extracting a subset of analogs (Fig. 1). Then, for every day of this analog subset, the rainfall amounts collected over the specific catchment for these particular days are extracted from a hydrological archive. This provides an empirical sample of the amounts of rain that can be collected in this type of situation. Then, an incomplete gamma function is fitted over this sample by the classical moment method, taking into account the frequency of zero values (Fig. 2). Finally, this conditional distribution can be considered as the probabilistic forecast for the day considered and expressed in different ways (such as quantiles for non-exceedance).

The variables used to characterise the synoptic patterns were obtained by screening the variables used at that time by human forecasters, combined with some trade-off between past data availability and consideration of the forecasting capacity of such predictor variables with a view to extending lead-times. The resulting selection is described in the next section.

The assumptions expressed above show why the forecast must be issued in probabilistic form. If we were able to compare two states of the atmosphere precisely, taking into account all their components exhaustively, then the best forecast would be based on the analogy with the single most similar situation extracted, i.e. the precipitation value observed on the best analog. However, for practical reasons, and also because of our

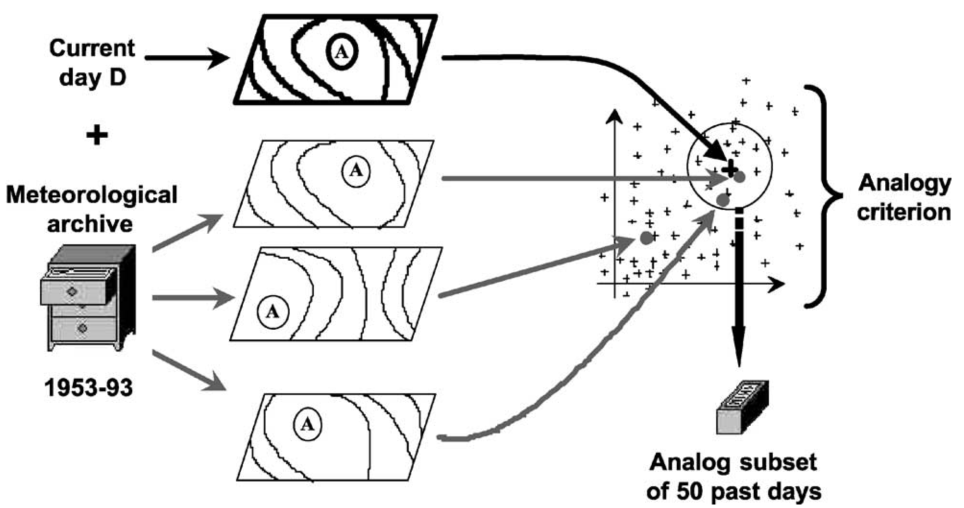


Fig. 1. First step of the forecast: selection of a subset of analog dates.

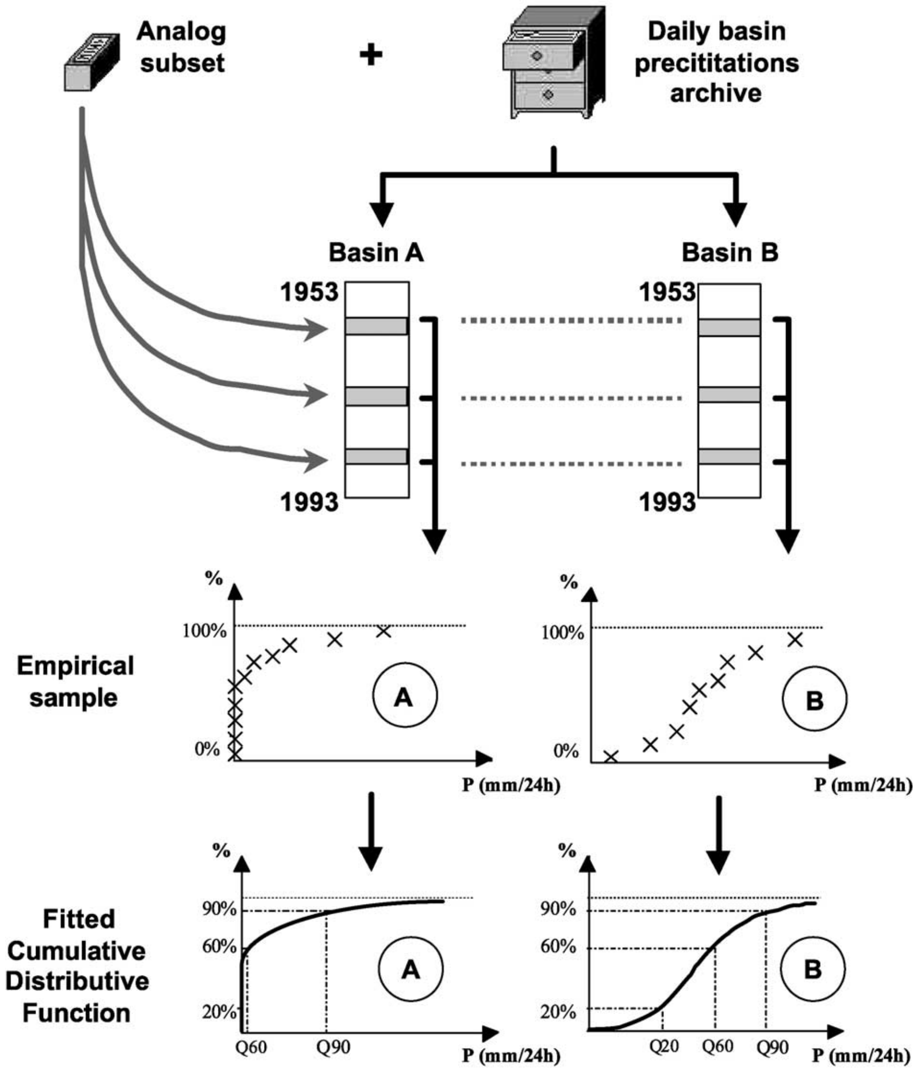


Fig. 2. Second step of the forecast: elaboration of basin-specific conditional distributions.

limited knowledge, the analogy used relies only on a subset of synoptic variables $\{V_{SYN}\}$ while ignoring those $\{V_{LOC}\}$ that are more representative of local conditions (and are either not monitored or poorly modelled). If, in a given situation, the observed rainfall is P_{OBS}, then a certain part P_{SYN} depends on the synoptic or regional variables, while the other part $P_{LOC} = P_{OBS} - P_{SYN}$ has been controlled by local variables. Considering now an analogous situation, with a different observed precipitation $P_{OBS}^* = P_{SYN}^* + P_{LOC}^*$, it is assumed that the synoptic part P_{SYN}^* will be close to P_{SYN}, while P_{LOC}^* is just one sample among others of what can be caused by the variability of local conditions. If an infinite meteorological archive were available, it would be possible to find a sub-infinity of situations rigorously that were similar in terms of synoptic patterns (therefore satisfying $P_{SYN} = P_{SYN}*$), while the observed precipitation $P_{OBS}*$ would provide the conditional distribution of P_{OBS} (P_{SYN} fixed) $= P_{SYN}$ (fixed) $+ P_{LOC}*$ and reflect the variability of the effect of local variables on basin precipitation for this predefined synoptic pattern.

In practice, the archive is never infinite, so that absolutely similar analogs can hardly ever be obtained, even at the synoptic scale. Therefore, a finite number of situations, analogous enough to the day at hand, are collected. The more we select, the better the sampling, but the less this subset is conditioned by the situation on the day of interest, because of the decreasing analogy. This is why some trade-off must be accepted, depending on the size of the available archive. This can be easily understood through two extreme cases:

• If only the single best analog is kept, then the forecast comes down to the single precipitation value observed for this analog. However, it is almost certain that the day of interest differs from its best analog in terms of local variables $\{V_{LOC}\}$, so that its observed rainfall will differ from the previous one.

• Conversely, if the constraints of analogy are significantly relaxed, considering that all the situations contained in the archive are similar to the day of interest, then the marginal, i.e. climatological distribution is obtained. This will certainly include the new situation, but it will be of little informative interest in terms of forecasting!

In practice, implementation requires several ingredients: a meteorological archive (Section 3.1), a hydrological one (Section 3.2) and a search algorithm (Section 4).

3. Available data

As explained before, two distinct archives are needed: on the one hand, a meteorological archive of synoptic variables to select analogous situations, and on the other hand, a hydrological archive of basin rainfall values for the different catchments of interest in order to determine their respective rainfall distributions.

3.1. Meteorological archive

The building of a meteorological archive is a crucial step in implementing the method. It requires a trade-off between (i) the length of the archive, which should enable a sufficient number of good analogs to any situation to be selected, (ii) the amount of information, which means a variety of pertinent variables to discriminate better between rain-generating situations and (iii) the homogeneity of the archive over the entire period it covers, to avoid artefacts due to changes in sensors or data processing.

The first point implies the use of several decades of data. A 10-year-long archive, for example, might in fact not include a single 10-year return rainfall over the target catchment. Until very recently (before the release of reanalysis data), the only way to collect a long homogeneous archive was to restrict oneself to standard variables such as geopotential or temperature fields. More elaborate variables would have been more model-dependent, while the models used to process the data have changed many times over the years. Even for reliable and basic variables such as pressure and temperature, the monitoring network of radio-sounding stations has also changed significantly. Besides, if we want to extend the forecasting scheme in the medium-term future, the variable used must allow reliable forecasts, which are again restricted to those that are not too sensitive to the meteorological model used.

This explains why, historically, the archive has been based «only» on the geopotential heights (Z) for 1000 and 700 hPa measured at 0000 UTC. Raw data are taken at 37 radio-sounding stations covering Western Europe and the near Atlantic. In the recent updating of the method, since some stations have been closed, they have first been reconstituted and then interpolated by a spline technique over a gridded domain of 418 nodes (22×19). This grid, which is regular in a Mercator projection, covers a domain ranging from 15.7°W to 24.3°E and from 33.4°N to 57.0°N (Fig. 3). The period collected, which is updated every 2 years or so, is now 1953–1998, i.e. 45 years, (although the developments made in 1998 and presented hereunder were based on 1953–1993, i.e. 41 years, or 14 975 different days). In operational use, the synoptic forecasts required to characterise the coming days (Z 1000 and Z 700 hPa) have been taken from different models, up to 72 h ahead since the end of the 1970s. They are now taken from the ARPEGE model of Meteo–France (Geleyn et al., 1994) up to 72 h ahead, and from the European model

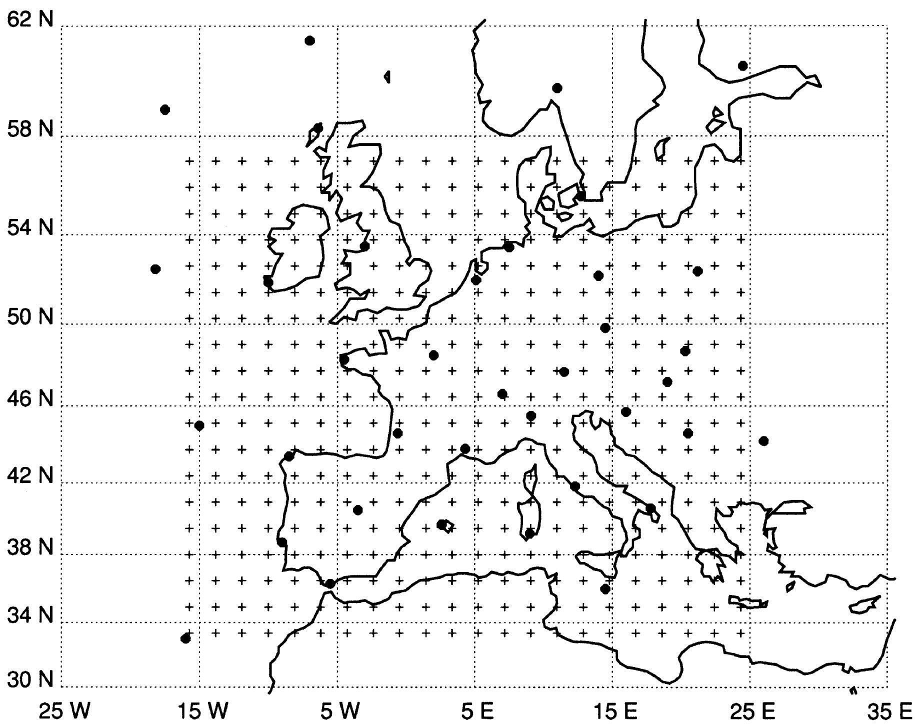


Fig. 3. Location of the nodes (crosses) of the grid used to store geopotential data issued by the 37 radio-sounding stations (dots).

ECMWF (Simmons et al., 1989) for 72 up to 192 h. Before use, these forecasts are themselves interpolated over the same grid as the historical archive.

Of course, these steps of reconstitution and interpolation are not optimal, but they have made it possible to take advantage of the filing work initiated in the 1970s.

3.2. Hydrological archive

The target variable, or predictand, is the amount of rainfall collected on any given catchment. For each catchment of interest, basin precipitation was computed from gauge data over the same period as the meteorological archive (1953–1998), so that for each analog date, we know the basin-integrated rainfall collected over these catchments. Since we need to go back in time as far as 1953, the only readily available rainfall data are the daily totals collected manually every day at 0600 UTC. This is why we focused on

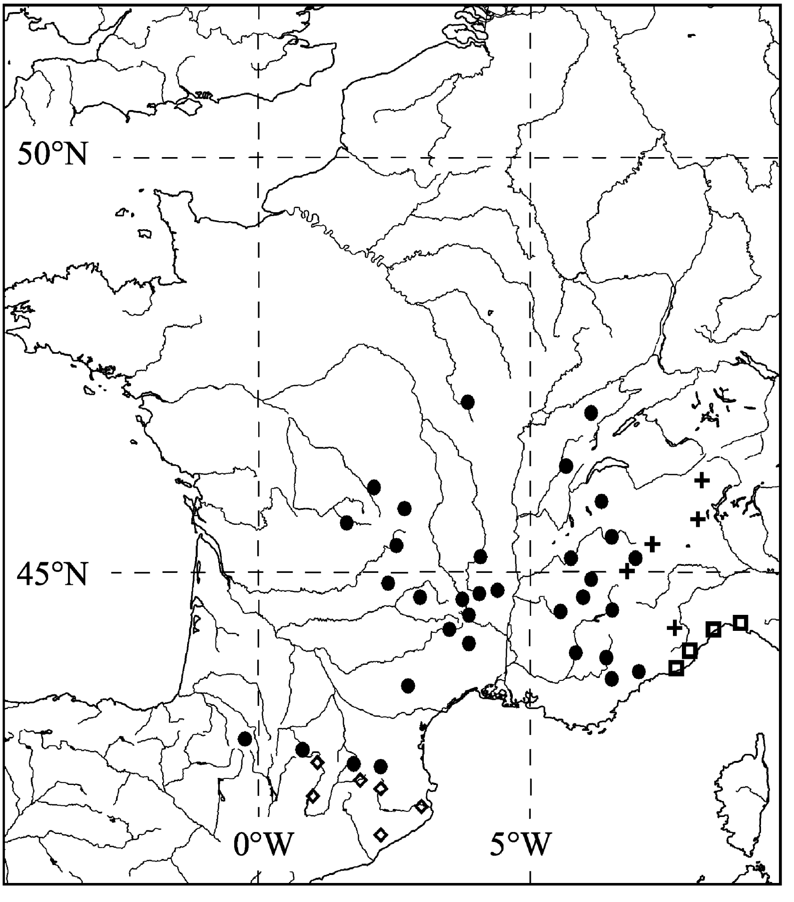


Fig. 4. Location of the French (black dots) and Catalan (diamonds) target catchments. The Italian target catchments are split according to their region: Piedmont (crosses) and Liguria (squares).

forecasts at a daily time scale. Moreover, we chose to consider catchment essentially in the 200–2000 km^2 range. These are homogenous enough in terms of rainfall climatology at daily time-steps and furthermore cover most hydrological requirements, at least for a power company.

Here again, the quality of these estimates strongly depends on the density of the monitoring network (Creutin and Obled, 1982), which fluctuated throughout the year, and on the quality of individual sensors (in particular automatic ones in the case of snow). So the choice of the target catchments was based not only on their hydrological interest but also on the availability of a good homogeneous archive. Given that, 33 catchments were initially selected. These have since been increased to more than 50, covering a large part of southeast France, essentially in the mountains. For research purposes, a few more located in Catalonia (Spain) and in Piedmont and Liguria (Italy), have been added (Fig. 4).

4. Algorithmic tools

4.1. Analogy criterion

The selection of analogs is based on a comparison, for at least one synoptic variable (e.g. the 700 hPa geopotential field at 0000 UTC), of the pattern expected for the day at hand with all observed patterns in the historical archive. This means that a quantitative criterion is needed to quantify the similarities between the patterns. It is then possible to sort all situations in the archive as a function of their similarities with the situation considered, depending on how they score against the analogy criterion.

There are many such criteria, such as Euclidian distance, or correlation between the two sets of grid points, etc. But with regard to the similarity between geopotential fields, what we look for in fact is an analogy in circulation patterns bringing precipitable water over the catchment. A Euclidian distance, which here would consist of an average squared difference between the two gridded fields, is of limited interest: in fact, two fields can be reasonably close in absolute departure, while undulating in opposition of phase with different centres of action and therefore completely different circulation patterns.

This is why the Teweless–Wobus score is used (Teweless and Wobus, 1954), as it emphasises analogy in shape, and therefore in circulation. It has already been proposed in the literature to select analogs for temperature forecasts (Woodcock, 1980). Developed initially to evaluate the quality of geopotential forecasts, this criterion focuses on synoptic circulation by considering, for any grid node, not the absolute values but rather their south–north and west–east gradients. This is tantamount to comparing the geostrophic winds generated by these pressure fields. The exact formula is:

$$\mathrm{TW} = \frac{\sum_i |e_G^i| + \sum_j |e_G^j|}{\sum_i |G_L^i| + \sum_j |G_L^j|} \tag{1}$$

where e_G is the difference, around a given node, between the geopotential gradients selected respectively for one field (e.g. the observed one) and for the other (e.g. the forecast one) while G_L is the maximum of these two gradients in the direction considered. The sum of all pairs of adjacent points in the south–north (index i) and west–east (index j) directions of the gridded domain is then calculated. In the present case, this may be either the full grid of 418 points or any sub-domain. Dividing by G_L, it is possible to scale within the field (since a given difference is less pronounced in a strongly varying field than in a smooth one), and also to remove seasonal effects. When multiplied by 100, the score varies between 0 (for perfectly parallel fields) and 200 (when undulating in complete phase opposition), while a score of 100 corresponds to totally independent fields.

4.2. Seasonal effects

Another constraint was introduced to cope with seasonal effects. For a given day, the analog selection is restricted to a period of 4 months centred around this particular date. By doing so, it is hoped, following Lorenz (1969), that candidate situations will present similar characteristics in terms of solar energy amounts, surface fluxes, etc.

4.3. Deriving the quantitative precipitation forecast

Once a set of M analogs has been selected, the hydrological archive is consulted. The basin-integrated precipitation collected for each date of the analog set is extracted for each individual catchment. These M samples provide an empirical conditional distribution of rainfall over the catchment for given synoptic conditions. This distribution constitutes the forecast of the analog method. It can be fitted to a probabilistic model (either a truncated exponential or gamma distribution, since it often contains many zero values) and expressed in quantile values (or values with fixed probabilities of being exceeded).

A refined version involves taking into account the fact that the analogs selected may have different values of the analogy criterion, i.e. correspond to synoptic situations of varying similarity to the target situation. This may be taken into account by weighting the rainfall value of each analog accordingly to favour the closest ones. The forecast will then differ slightly from the empirical distribution observed over the sample.

4.4. Optimisation criterion

Different forecasting schemes may be devised, changing the variables, the domain, or the analogy criterion, etc. Therefore, a single objective function must be defined to optimise the scheme, i.e. to decide if one scheme is better than another one.

However, the present forecasts are issued in probabilistic form, so that the a posteriori observed single value has to be compared with an a priori given distribution. This is a troublesome problem (see e.g. Wilson et al., 1996) and in such cases, it is better to consider the usefulness of the forecast rather than its quality or accuracy. Usefulness suggests that we have in mind some particular use for this forecast and the evaluation criterion will more or less reflect this particular goal. Consequently, its choice will certainly be crucial in the way the algorithm is going to be optimised, with certain applications in mind underlying this choice.

The Ranked Probability Score (RPS) (Epstein, 1969; Murphy, 1971) was eventually chosen. It is designed to evaluate a probabilistic forecast issued in the form of probabilities of belonging to predefined ordered classes. In the present case, these are rainfall amounts and the score is expressed as:

$$\mathrm{RPS} = \frac{1}{N}\sum_{n=1}^{N}\sum_{g=1}^{G}\left[\sum_{k=1}^{g}(p_{kn} - \delta_{kn})\right]^2 \tag{2}$$

where N is the number of forecasts issued, G the prescribed number of classes, p_{kn} the probability of falling in class k allocated in the nth forecast, with δ_{kn} defined as $\delta_{kn}=1$ if the observed value corresponding to the nth forecast falls into class k, and $\delta_{kn}=0$ otherwise. This score has the advantage of preserving the ordered structure of a variable such as precipitation, i.e. it considers that forecasting a class just before or just after the observed one should be less penalised than forecasting a more remote one. The closer the criterion to 0, the more useful the forecast. For example, if the forecast is categorical about a single class, then the score is 0 if the forecast proves to be true, and $G-1$ if it is false.

Hopefully, in the present case, the number of classes and the class boundaries should be defined catchment per catchment, with respect to the regional climatology (certain high values are fairly common in some Mediterranean catchments but extremely rare in others) but also basin sizes (some high basin rainfall values may be reached in small mountain catchments but become unrealistic for large catchments). Nevertheless, given the number of catchments to consider and the burden of heuristic optimisation, as will be seen later, a single division into classes was selected and applied to all catchments. It consists of eight classes: 0 (no rain), 0–1, 1–5, 5–10, 10–25, 25–50, 50–100 mm, and more than 100 mm. For each individual forecast and each single catchment, the probability of falling into a given class k is defined, for example without weighting, as m_k/M if there are m_k analogs with their associated amounts of rainfall falling in class k among the ensemble of M analogs selected.

5. Optimisation and calibration of the analog method

Several critical choices as well as a number of parameters have to be decided upon to optimise the method. Among them are those already discussed (analogy criterion, performance score, processing of the analog, etc.), but many others have to be screened, such as the choice of synoptic variables (field type, level, time in the day) as well as the domains in which they should be considered (size, position), the way they are to be pooled if several are to be selected, and the processing of analog rainfall values, etc. All these call for some exploration and calibration in order to optimise forecasting capacity.

5.1. Exploratory and calibration analysis

In the current state of the method as implemented operationally (leaving aside further research carried out since), optimisation was performed essentially over the 50 French catchments considered as a whole. This meant looking for a single set of rules and parameters for all these target catchments, implying that the same situations are considered as analogs for all of them. We shall see later that this constraint was relaxed for Spanish and Italian catchments.

The current calibration focused essentially on autumn situations (September to November), which are the most prone to flooding in Mediterranean regions. The historical file was therefore restricted to 15th August–15th November. The optimisation was performed in a context of "perfect prognosis", i.e. the values characterising a given day were the best available, i.e. the observed or analysed values. In order to test a set of model parameters, the forecast for every day of the historical archive (~ 4000 days) was reprocessed off-line. Obviously, to avoid the best analog of a given day being the day itself or the day just before or after, analogs could not be drawn from the same year. All days in the archive are scanned, thus providing a set of 3731 test forecasts drawing their analogs from among 3640 days.

Given that analysed synoptic variables were used (i.e. perfect prognoses instead of actual forecasts), the performances will be overestimated compared with operational ones. Furthermore, the fact that optimisation is done by a kind of jackknife method also suggests that the performances measured on the learning sample could be even more over-optimistic. However, they will not be considered in the absolute sense, but only as an objective function for optimisation. More realistic evaluations were made on a separate validation sample, first with perfect prognoses and next in realistic forecast conditions.

5.2. Selection of synoptic variables

The historical archive available (geopotential heights 1000 and 700 hPa at 0000 UTC) allows only a limited number of combinations. However, it is worth remembering that the target variable considered is the amount of precipitation from 0600 UTC on day D until 0600 UTC on day $D+1$ so that the geopotential of day $D+1$ at 0000 UTC, (or day D at 2400 UTC) may also be an acceptable candidate, within the time window, to characterise day D.

Four fields were therefore tested, firstly one after another (700 hPa at 0000 UTC, 700 hPa at 2400 UTC, 1000 hPa at 0000 UTC and 1000 hPa at 2400 UTC) all other parameters being kept identical. The best performances were obtained by finding the analogy for the 700 hPa field at 2400 UTC (Fig. 5). The 1000 hPa field at 2400 UTC was the second choice for a single field, showing that being as centred as possible on the rainfall time window gives the best results. The level itself appears as secondary compared to taking this time of 2400 UTC.

If the analogy is to be based on several fields, i.e. different levels or times, this raises the problem of pooling the analogy between each field two by two. The overall criterion taken to select the analog dates is just the sum of the individual criteria computed field by field. Several combinations were tested involving one, two, three or four fields (Fig. 5). Combinations with two fields always prove to be better than any one taken separately, and the best combination of two fields involves two different times and levels. Indeed, since the predictand (rainfall amount) is cumulated over a rather long period (24 h), it seems as if the analogy must involve both the change in time as well as the three-dimensional structure of the lower atmosphere, through the use of different observation times and

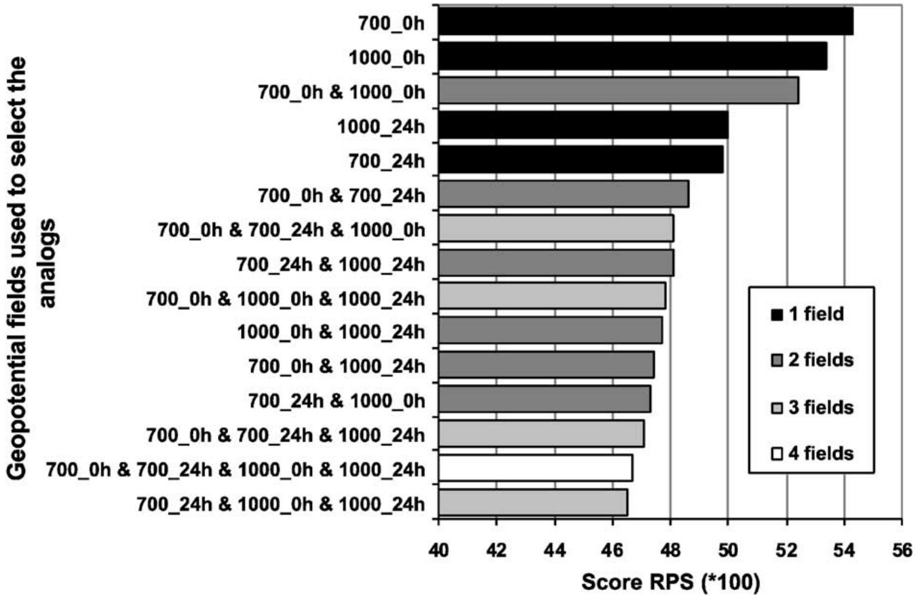


Fig. 5. Utility scores obtained in a context of perfect prognosis using different combinations of geopotential fields to select analogous dates: 700 hPa at 0000 UTC (700_0h), 700 hPa at 2400 UTC (700_24h), 1000 hPa at 0000 UTC (1000_0h) and 1000 hPa at 2400 UTC (1000_24h). These results were obtained considering sets of 50 analogs and grid domain number 5 (from 15.7°W to 18.6°E and from 33.4°N to 53.7°N).

levels. However, beyond two, combinations of three or four fields (among the few available here, e.g. using information on day *D*-1 or *D*-2) only produce minor improvements, or even lower the performance.

Nevertheless, it was decided to use four fields in the operational version, i.e. the 1000 and 700 hPa level at 0000 and 2400 UTC (to explain rainfalls from 0600 to 3000 UTC). This is slightly sub-optimal but the loss in performance is slight, and the symmetry in the variables seems to make further interpretation by a human forecaster easier. The analogy could also have been weighted between observations at 0000 and 2400 UTC because the latter is more centred and presumably more representative for the precipitation period. Tests using different weightings for the Teweless–Wobus scores of the four fields respectively when quantifying the overall analogy did not prove decisive.

5.3. Analogy domain

To specify the grid over which the analogy must be satisfied, several characteristics must be fixed: its size, its location, and possibly its mesh resolution.

Only the first two were tested. In previous works, Nieminen, 1983 showed that the Teweless–Wobus criterion is relatively insensitive to the mesh resolution as long as this remains reasonably "small" compared with the longest wavelengths of atmospheric circulation (in order not to filter them). So the mesh of the archive grid ($\sim 2 \times 1.5°$) was kept.

5.3.1. Domain size

The domain initially selected was in fact chosen from the maps that human forecasters used to analyse. However, there was no strong reason to believe that similarity in rainfall response over one catchment required similarity in synoptic circulation over this entire domain. So, from the initial "overall" grid of 418 nodes, over which geopotential heights are archived, six different sub-domains were defined, all centred on the set of target catchments (Fig. 6). The utility, or performance scores obtained by calibration for each configuration (using the two fields 1000 and 700 hPa at 0000 and 2400 UTC selected previously) are respectively (from grid no. 1 to grid no. 6): 51.6, 47.7, 45.0, 45.5, 46.7 and 48.1. If we remember that the lower the RPS score the better the performance, it appears that a medium-sized grid, covering the zone from 6.2°W to 12.9°E and from 38.0°N to 50.3°N is appropriate. There seems no point, for the set of catchments considered, in searching for an analogy over a larger domain. This is consistent with results obtained by Van Den Dool (1989) who spoke of "limited area analogs" and showed that for a short-term forecast (1 day), analogy within a circle of around 1000 km centred on the target zone was sufficient.

This might seem somewhat limited in domain size to a human forecaster, but the point is that we used two observation times, 0000 and 2400 UTC. If the observation at 2400 UTC is dropped, the information required to explain the rainfall falling by the end of the period 0600–3000 UTC should come from further away, outside the domain. And this is what appears when one restricts the analogy to 0000 UTC only; the optimal domain shifts essentially to the west and must be twice as large. This looks like a crude application of the Taylor hypothesis!

5.3.2. Location of the grid domain

Keeping this optimal size (sub-domain no. 3) and again considering an analogy based on the four geopotential fields, 1000 and 700 hPa at 0000 and 2400 UTC, we system-

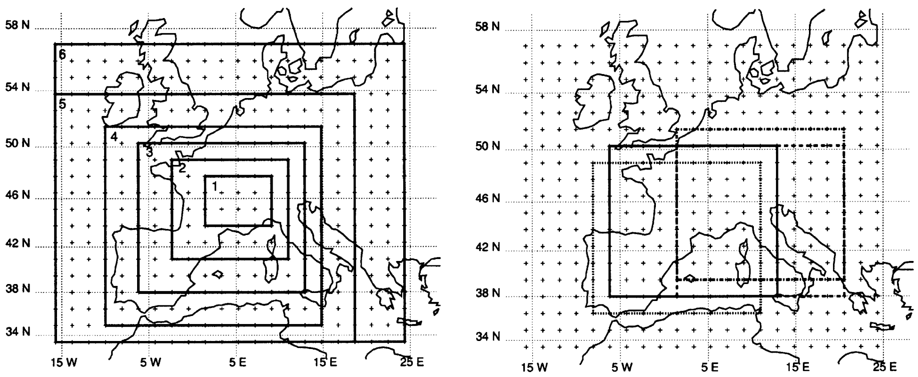


Fig. 6. The six grid domains tested (left) and the optimal grids (right) obtained for the different sets of target catchments: France (solid line), Catalonia (dotted line), Piedmont (dash-dotted line) and Liguria (dashed line).

atically tested all possible north–south and east–west shifts of one node point at a time to determine the optimal location of the sub-domain within the complete domain. It appeared that the best one is the initial location centred over the set of 50 target catchments. However, when we have considered the Spanish and Italian catchments separately, with the latter further subdivided into two groups according to their geographical location—Piedmont and Liguria—the same set of trials as above showed that optimal grid size no. 3 is best located when centred over the set of catchments considered (Fig. 6). It even appeared fairly sensitive, since for example with the two sets of Italian catchments, a shift between their respective domains for synoptic analogy proves effective in terms of rainfall forecasts.

It may be noticed that the shift suggested by these blind searches are again consistent with the climatology and could be expected again by applying the Taylor hypothesis. This also suggests that using a single sub-domain for all French catchments spread over 1000 km is probably not optimal and that some splitting of the catchment set should be considered, but this was not retained in order to keep implementation simple.

5.4. Derivation of a probabilistic forecast

It has already been explained that this forecast is based on the conditional distribution of rainfalls observed in the set of analog dates, for each catchment, respectively. However, every day in the archive has a certain degree of analogy with any specific day considered, and we have to decide where we stop considering 2 days as sufficiently "analogous" to be informative concerning the rainfall potential of the situation at hand. The performances achieved when selecting 10 to 100 analogs were therefore compared, and the relatively flat optimum suggests a value of around 50.

It must be stressed that this conclusion is highly dependent on the size of the learning sample: the larger it is, the richer it is in pattern diversity and the more "good" analogs we can expect to find close to the day of interest. This value of 50, obtained by trial and error, is therefore a trade-off for this particular archive in its present state. The result also depends on the score selected to measure performance and utility. For example, the RPS score used here tends to favour rather spread forecasts (displaying a rather large standard deviation) against "sharper" forecasts. Another score, more favourable to fine distributions, could have suggested a smaller number of analogs to narrow the distribution spectrum.

Next, once this number has been fixed, all selected dates do not have the same degree of analogy. A given situation may not have many analogs in the archive, just for sampling reasons. Considering a 10-year return rainfall, there is little chance of observing it 50 times in 45 years...! So absolute conditioning by the current situation and its representativeness changes from day to day; they are measured by the distribution of the TW scores. This cannot be modified other than by increasing the archive size, but at least it can be used as a forecast confidence index. The only way this effect can be evaluated is by running the analog selection over smaller archives (10, 20, 30 years) and trying to extrapolate. It can be seen that the performance increases regularly, and may differ for two independent archives 20 years long. So the longer the archive, the better the analogs and therefore the forecast.

However, differences in the accuracy of the analogy between the 50 selected dates can be taken into account. For example, different weightings of the analog rainfalls according to their TW scores were tested, giving more weight to the observation corresponding to "good" analogs. These tests proved to be scarcely sensitive on average, because in the most frequent situations, rainfall amounts have low or zero values and are not sensitive to the weighting. However, it was shown that this could significantly refine the forecasts of extreme values. Nevertheless, this has not been implemented so far (but will be in future versions).

6. Results

6.1. Calibration retained

Finally, the method was implemented with the following choices:

- analogy based on four geopotential fields: 1000 and 700 hPa at 0000 and 2400 UTC, equally weighted (Section 5.2)
- domain used corresponding to grid no. 3 (Section 5.3)
- selection of the 50 best analogs (Section 5.4)
- forecast conditional distribution of rainfalls observed for these 50 analogs (Section 4.3).

All the results presented hereafter are averaged over the entire set of French catchments.

6.2. Performances in perfect prognosis conditions

These results obtained by calibration (period 1953–1993) and then by validation over the independent period 1995–1996 are presented in Table 1. The context of perfect prognosis means that the predictors, and in particular the fields of day *D* at 2400 UTC, are analysed fields and not output forecasts from a meteorological model (as they should be in a real-time context).

In order to interpret the results more easily, it is worth having some reference levels. We considered three rather simplified or "poor-man forecasting" methods:

- Climatology: the forecast is the same every day and consists of the marginal distribution of rainfall during the autumn season, ignoring the specific conditions of the day at hand
- Persistence: i.e. the forecast for the day at hand (day *D*) is simply the value observed on the day before (day *D*-1)
- Random: i.e. the 50 analogs are taken at random from the archive.

Since the interpretation using the RPS score is hardly intuitive, a "success index" (referred to as SI) in terms of Rain/No rain is also given. This corresponds to the percentage of success of rainfall occurrence. Starting from the conditional distribution, the forecast is considered as "Rain" if the median of the precipitation distribution is larger

Table 1
Performances reached by the Analog method (ANALOG) and by three poor-man methods (Climatology, Persistence and Random drawing) over the 1953–1993 period (Calibration) and the 1995–1996 period (Validation)

	SI (%)		RPS (* 100)	
	Calibration	Validation	Calibration	Validation
Climatology	52.4	48.6	75.4	69.9
Persistence	72.8	71.3	98.6	93.7
Random	53.4	49.9	76.6	72.3
ANALOG	80.1	77.6	45.1	48.7

These performances are expressed in term of Success Index (SI) and Ranked Probability Score (RPS).

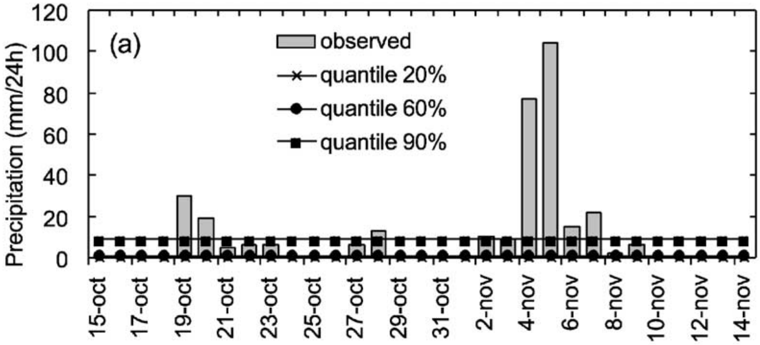
than 0 mm, and "No Rain" otherwise. However, this index completely ignores the quantitative aspect of the forecast and is given only as an addition.

The results obtained during calibration (Table 1) show that Climatology and Random sampling have very similar scores, which is quite obvious since random sampling draws a random sample of 50 values every day from the entire sample of around 4000 constituting the climatology. The analog approach also samples the climatology, but in a non-random way, since it gathers samples with similar synoptic conditions only. This conditioning proves very efficient and performs much better (i.e. with larger SI and lower RPS scores) than climatology. As expected, persistence appears superior to climatology in terms of SI (Rain/No rain), because the dry or rainy periods last longer than 1 day. But persistence drops considerably when considering the quantitative aspect, and always remains worse than the analog approach.

Looking now at results during validation on an independent test sample, Fig. 7 shows the forecasts simulated day by day for the period 15 October–15 November 1994 on the Tanaro River, a tributary of the Po River. Its upper catchment (~ 1700 km^2) is located close to the French–Italian border.

It must be stressed that the analogs used in this example are taken only from the period 1953–1993. For the sake of clarity, the forecast is displayed only as the three quantiles corresponding to non-exceedance probabilities of 20%, 60% and 90%, respectively (referred to as Q_{20}, Q_{60} and Q_{90}). This provides a confidence interval (Q_{90}–Q_{20}) within which there is a 70% chance of getting the observed rainfall. Instead of being centred around the median, this interval has been slightly shifted to allow for the large number of zero-values present in this variable.

Fig. 7a is trivial and shows what climatological quantiles would provide, with the lowest two (Q_{20} and Q_{60}) blocked at 0 mm. Fig. 7b shows how the analog method reacts day by day: the "signal", either the 60% quantile or the whole "tube" consisting of the temporal sequence of daily confidence intervals, very closely follows and almost always comprises the observed rainfall. There are a few false alarms, where rainfall is expected but not observed (on November 8th and 10th for example), but no failure to give an alarm. In particular, the catastrophic flood of November 4th and 5th (Lionetti, 1996) is well predicted. The observed rainfalls fit quite well in the confidence interval proposed by the forecast.



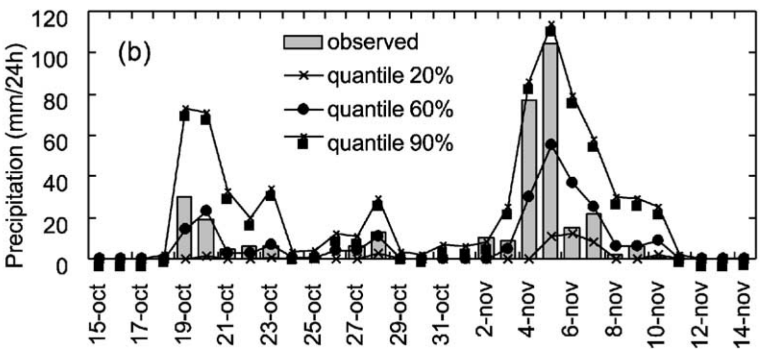


Fig. 7. Comparison between observed and forecast rainfall during the period 15 October–15 November 1994 over the Tanaro catchment (Italy) given by Climatology (a) and by the Analog method in a perfect prognosis framework (b).

In general, over a long period, some statistical bias for very large observed rainfall values is all that may be pointed out: the forecasts seem to underestimate the amount of rainfall in these cases. This is certainly due to the limitation of the historical archive: as already said, 50 values of 10-year return precipitation cannot be extracted from a 50-year-long archive!

6.3. Implementation for operational forecasting

The version of the analog approach described in this paper has now been running in real-time since 1999. Forecasts are issued every day at 0600 UTC for the next 6 days, i.e. the next seven periods of 24 h. Unlike the perfect prognosis case described in Section 6.2, all the geopotential fields used in real-time are not analysed ones: only those at 0000 UTC are analysed; the others (24 h, 48 h, etc. until 168 h ahead) are forecast. These are taken from the French model used by Meteo–France until 72 h, and from the European model used by ECMWF beyond. So the uncertainty inherent to the method itself (linked to the effect of local variables), as measured in perfect prognosis conditions, is further increased by the uncertainty in the geopotential forecasts. If this uncertainty were always the same, as it is with analysed fields, the performance of the analog method would be the same for every lead-time as well. However, the uncertainty of forecast geopotential increases with lead-time, thus diminishing the performance of the analog method.

As an example, Fig. 8 compares the QPFs proposed for the month of November 1996 on the upper Loire catchment depending on whether perfect prognosis (Fig. 8a) or operational conditions (Fig. 8b) are considered. Only the forecasts for the first lead-time (0–24 h) are displayed. It appears that the use of forecasts instead of analysed values for geopotential fields at 24 h generates only a small decrease. Generally speaking, over the entire set of French catchments, the performance or utility criteria computed for the 1995–1996 test period show that for the first 24 h, the use of actual forecasts for the 24-h geopotential instead of analysed values lowers the success index SI from 77.6% to 77.4% and increases the rank probability score RPS (* 100) from 48.7 to 49.4. Therefore, for this first lead-time of 24 h, the uncertainty resulting from the uncertainty in geopotential field forecasts is very limited, and small compared with the intrinsic uncertainty of the analog method itself.

However, this may be more sensitive at longer lead-times and also depends on the meteorological model used to get the geopotential forecasts. Recent experiments performed by Djerboua and Obled (in press) during the MAP (Mesoscale Alpine Program—

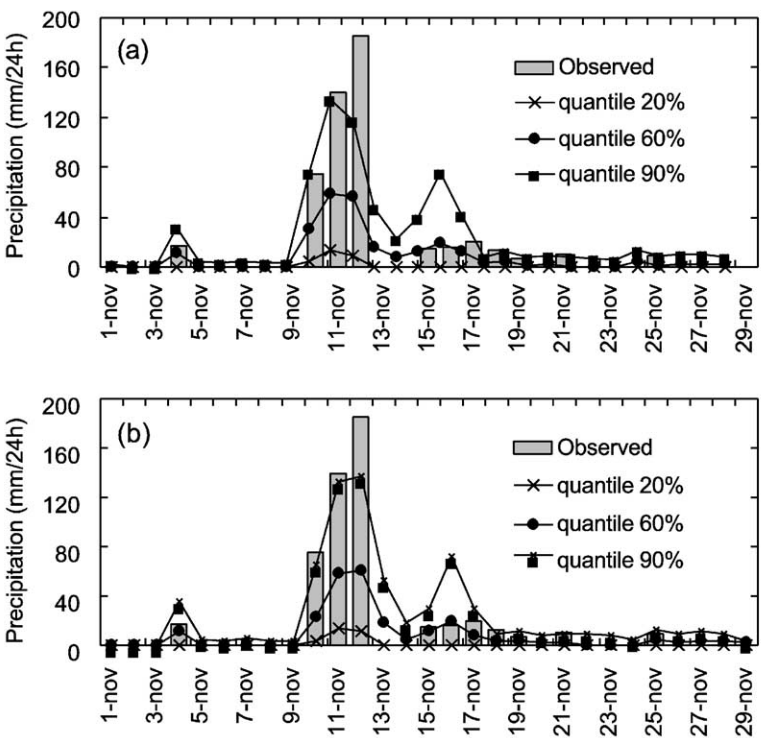


Fig. 8. Comparison between observed and forecast rainfall during November 1996 over the Upper Loire catchment in a perfect prognosis (a) and real-time framework (b).

see Bougeault et al., 2001) showed that, in an operational context, the decay in performance of the geopotential forecasts (Atger, 2000) with longer lead-times, of up to 7 days ahead, exceeds the intrinsic uncertainty of this adaptation method.

7. Discussions and conclusions

Over the years, and still more so in the recent ones with its updated version, this adaptation method involving analog sorting has proved a very efficient alternative in quantitative precipitation forecasts for medium-sized basins (200–2000 km^2).

Obviously, it cannot be considered as fully optimised yet. The set of predictors selected is certainly too limited and the analogy criterion may still be improved, as well as the performance score, which controls optimisation of the space domain and analog processing. However, the single-valued central quantile (here Q_{60}) might already be considered as a good deterministic QPF, while the confidence interval provided (essentially the Q_{60}–Q_{90} range) is informative with regard to the potential for extreme values. Moreover, beyond the purely technical presentation given above, a great amount of know-how has been acquired through lengthy use, which cannot be presented fully here. For example, each day is characterised by a set of analogs resembling the day at hand to varying degrees. This analogy is measured by the Teweless–Wobus score. However, by looking at the distribution of the TW score, one quickly gets an index of the quality of the forecasts, depending on whether they are small or large (i.e. excellent or poor analogy), concentrated or spread (i.e. sharp or fuzzy forecasts), etc.

Such an approach is easy to understand and to interpret for a forecaster. It builds upon an objective learning process and a reasonable assumption: if we find synoptic situations in the past that are similar enough to the one at hand, it is likely that the catchment, with all its specific features (orography, altitude range, aspects, etc.) will react in a similar way in terms of rainfall. Obviously, in anticyclonic conditions (over the space domain considered), pressure fields are usually fairly smooth, and it is easy to find similar fields and to see that rainfall is essentially zero with a good level of confidence. However, in cyclonic conditions, pressure fields may be much more varied, and the similar situations that can be extracted are usually less analogous and have a more widespread range of observed rainfall. But as already said, even the spread of past observed rainfalls provides information on the rainfall potential of this situation. Furthermore, this approach is cheap and quick to set up in operational practice.

The major drawback until recently was the burden of setting up a long meteorological archive and updating it over time. In the present case, for example, it was initiated in the early 1970s through the manual collection of radio-sounding data. At that time, the screening of candidate predictors (e.g. should a temperature field, or a humidity field, be included in the analogy?) meant collecting and handling a lot a data that may prove useless afterwards. This is why the set of predictors was limited to the geopotential fields 1000 and 700 hPa and remained unchanged until recently. There are even some concerns about the homogeneity of this archive, since over time, radio-sounding stations have been closed and data processing techniques (assimilation schemes) have been constantly evolving. However, there has been a major change of perspective recently with the advent of reanalyses, for example those proposed by the NCEP/NCAR (Kalnay et al., 1996). They offer more than 50 years of atmospheric data over the globe, all processed with the same standard model, which does not constitute a complete guarantee of homogeneity but at least a great improvement. Furthermore, they allow a lot of new candidate predictors (such as humidity or instability index) to be screened and tested, without the burden of collecting them as before. This exploratory work has already started and is very encouraging. Surprisingly now, it is the hydrological archive of past rainfalls that may become the limiting factor. Although the daily basis may look rather coarse, this is not an intrinsic limitation and interesting results have been obtained on a 12 h basis, although there is little hope of extending them very much because of the limited number of recording rain gauges in the decades 1950–1970.

In conclusion, this method introduces an alternative way to issue medium-term forecasts, in addition to nowcasting (1–2 h lead-time) and short-term forecasting (6–36 h) for limited areas or fine-mesh models. It extends and capitalises on the outputs from larger-scale meteorological models, which are known to be exploitable in terms of synoptic variables but not reliable in terms of basin rainfalls at lead-times beyond 1 or 2 days. This approach is attractive in practical applications for water resources management and advanced flood warning systems. It is also an interesting tool for the scientific community in that it provides a kind of "basic" reference approach for one of the most challenging variables to forecast: the amount of precipitation. This kind of approach can help in understanding the physical processes of deterministic models by discriminating more clearly between the local and synoptic parts of rainfall events.

Meanwhile, this approach itself offers good prospects for improvement thanks to improvements in the model themselves, enhanced by the reanalysis data sets and also by the extension in the lead-times of dynamic models.

Acknowledgements

Dr. Daniel Duband, who initiated this approach in the early 1970s must be thanked for his continuous support in developing and improving this method. He has always managed to make all the necessary data available to us and agreed open-mindedly to most of our suggestions for improving the method in daily operational use. His followers, in particular Dr. Patrick Tourasse, have done their best for this challenging cooperation to continue and develop.

The "Servizio dei rischi idrogeologici" of the Piedmont Region, especially Renata Pelosini and Enrico Minguzzi, and Maria-Carmen Llasat from the Department of Astronomy and Meteorology at the University of Barcelona have kindly provided the data for the Italian and Catalonian catchments.

References

Atger, F., 2000. La prévision du temps à moyenne échéance en France. Meteorologie 8 (30), 61–86.

Bougeault, P., Binder, P., Buzzi, A., Dirks, R., Houze, R., Kuettner, J., Smith, R.B., Steinacker, R., Volkert, H., 2001. The MAP special observing period. Bull. Am. Meteorol. Soc. 82 (3), 433–462.

Creutin, J.-D., Obled, Ch., 1982. Objective analyses and mapping techniques for rainfall fields: an objective comparison. Water Resour. Res. 18 (2), 413–431.

Duband, D., 1970. Reconnaissance dynamique de la forme des situations météorologiques. Application à la prévision quantitative des précipitations. PhD Thesis, Faculté des Sciences de Paris, France.

Duband, D., 1980. Dynamic selection of analogue flow patterns to produce quantitative precipitation forecasts. WMO Symposium on Probabilistic and Statistical Methods in Weather Forecasting, Nice, France, September 1980, World Meteorological Organization, Geneva, pp. 487–492.

Djerboua, A., Obled, Ch., 2002. Medium-term QPFs by an anlog-based approach: objective evaluations and comparisons during the MAP experiment. Q. J. R. Meteorol. Soc. MAP-SOP special issue (in press).

Emanuel, K., Raymond, D., Betts, A., Bosart, L., Bretherton, C., Droegemeier, K., Farrell, B., Fritsch, M., Houze, R., Le Mone, M., Lilly, D., Rotunno, R., Shapiro, M., Smith, R., Thorpe, A., 1995. Report of the first prospectus development team of the U.S. Weather Research Program to NOAA and the NSF. Bull. Am. Meteorol. Soc. 76 (7), 1194–1208.

Epstein, E.S., 1969. A scoring system for probabilities of ranked categories. J. Appl. Meteorol. 8, 985–987.

Geleyn, J.-F., Bazile, E., Bougeault, P., Déqué, M., Ivanovici, V., Joly, A., Labbé, L., Piédelièvre, J.-P., Piriou, J.-M., Royer, J.-F., 1994. Atmospheric Parameterization Schemes in Météo–France's Arpege NWP Model. ECMWF Seminar Proceedings on «Parametrization of Sub-Grid Scale Physical Processes», Reading, UK, 5–9 September 1994, ECMWF, Reading, UK, pp. 385–402.

Glahn, H.R., Lowry, D.A., 1972. The use of model output statistics (MOS) in objective weather forecasting. J. Appl. Meteorol. 11, 1203–1211.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, M., Iredell, M., Saha, S., White, G., Woolen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., 1996. The NCEP/NCAR 40-year reanalysis project. Bull. Am. Meteorol. Soc. 77 (3), 437–471.

Krzysztofowicz, R., 1998. Probabilistic hydro-meteorological forecasts: toward a new era in operational forecasting. Bull. Am. Meteorol. Soc. 79 (2), 243–251.

Lionetti, M., 1996. The Italian floods of 4–6 November 1994. Weather 1, 18–27.

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

Murphy, A.H., 1971. A note on the ranked probability score. J. Appl. Meteorol. 10, 155–156.

Nieminen, R., 1983. Operational verification of ECMWF forecast fields and results for 1980–1981. ECMWF Technical Report no. 36, 11 pp.

Simmons, A.J., Burridge, D.M., Jarraud, M., Girard, C., Wergen, W., 1989. The ECMWF medium-range prediction models. Development of the numerical formulations and the impact of increased resolution. Meteorol. Atmos. Phys. 40, 28–60.

Teweless, J., Wobus, H., 1954. Verification of prognosis charts. Bull. Am. Meteorol. Soc. 35 (10), 455–463.

Van Den Dool, H., 1989. A new look at weather forecasting through analogues. Mon. Weather Rev. 117, 2230–2247.

Warner, T.T., Peterson, R.A., Treadon, R.E., 1997. A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. Bull. Am. Meteorol. Soc. 78 (11), 2599–2617.

Wilson, L.J., Strauss, B., Lanzinger, A., 1996. On the verification of weather element forecast from an ensemble prediction system, Preprints, 15th Conference on weather analysis and forecasting, Norfolk, VA, Am. Meteorol. Soc., Boston, MA.

Woodcock, F., 1980. On the use of analogues to improve regression forecasts. Mon. Weather Rev. 108, 292–297.

Zevin, F.S., 1994. Steps toward an integrated approach to hydrometeorological forecasting services. Bull. Am. Meteorol. Soc. 75 (7), 1267–1276.