



**HAL**  
open science

# Daily quantitative precipitation forecasts based on the analogue method: Improvements and application to a French large river basin

A. Ben Daoud, Eric Sauquet, G. Bontron, C. Obled, M. Lang

## ► To cite this version:

A. Ben Daoud, Eric Sauquet, G. Bontron, C. Obled, M. Lang. Daily quantitative precipitation forecasts based on the analogue method: Improvements and application to a French large river basin. Atmospheric Research, 2016, 169 (A), pp.147-159. 10.1016/j.atmosres.2015.09.015 . hal-01804766

**HAL Id: hal-01804766**

**<https://hal.science/hal-01804766>**

Submitted on 1 Jun 2018

**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.

## Daily quantitative precipitation forecasts based on the analogue method: improvements and application to a French large river basin

Aurélien Ben Daoud<sup>a</sup>, Eric Sauquet<sup>b</sup>, Guillaume Bontron<sup>a</sup>, Charles Obled<sup>c</sup>, Michel Lang<sup>b</sup>

<sup>a</sup> Compagnie Nationale du Rhône, 2 rue André Bonin, F-69316 Lyon Cedex 04, France

<sup>b</sup> Irstea, Hydrology-Hydraulics Research Unit, 5 rue de la Doua CS70077, F-69626  
Villeurbanne Cedex, France

<sup>c</sup> Grenoble-INP/UJF-Grenoble 1/CNRS/IRD, LTHE UMR 5564 Grenoble, France

### Citation :

Ben Daoud A., Sauquet E., Bontron G., Obled C., Lang M. (2016). Daily quantitative precipitation forecasts based on the analogue method: Improvements and application to a French large river basin, Atmospheric Research, 169, Part A, 1, 147–159, doi:10.1016/j.atmosres.2015.09.015

### Abstract:

**This paper presents some improvements of a probabilistic quantitative precipitation forecasting method based on analogues, formerly developed on small basins located in South-Eastern France. The scope is extended to large scale basins mainly influenced by frontal systems, considering a case study area related to the Saône river, a large basin in eastern France. For a given target situation, this method consists in searching for the most similar situations observed in a historical meteorological archive. Precipitation amounts observed during analogous situations are then collected to derive an empirical predictive distribution function, i.e. the probabilistic estimation of the precipitation amount expected for the target day. The former version of this forecasting method (Bontron, 2004) has been improved by introducing two innovative variables: temperature, that allows taking seasonal effects into account and vertical velocity, which enables a better characterization of the vertical atmospheric motion. The new algorithm is first applied in a perfect prognosis context (target situations come from a meteorological reanalysis) and then in an operational forecasting context (target situations come from weather forecasts) for a three years period. Results show that this approach yields useful forecasts, with a lower false alarm rate and improved performances from the present day  $D$  to day  $D+2$ .**

**Keywords:** Probabilistic quantitative precipitation forecasting, Statistical adaptation,

## Analogues, ERA-40 reanalyses, Saône river basin

### 1. Introduction

Many water-related stakeholders need quantitative precipitation forecasts as reliable as possible to anticipate discharges in river basins several hours or days ahead. For example, French operational flood forecasting services require precipitation forecasts several days ahead to anticipate flood risks (Lacaze et al., 2008), and hydroelectricity power producers need accurate and reliable precipitation forecasts to anticipate the discharge evolution along the regulated rivers and their tributaries in order to perform their activities and to provide safety control (Bompart et al., 2009).

Nowadays forecast uncertainties related to meteorological predictions and implied by modelling processes are more and more taken into account, leading to probabilistic quantitative precipitation forecasts (PQPFs) that provide ensemble forecasts. In particular, this kind of forecast enables to evaluate the risk of extreme events. At least, two approaches for producing PQPFs are commonly used: (i) regional ensemble weather forecasts based on dynamical approaches (e.g. COSMO-LEPS, Marsigli et al., 2005; PEARP, Thirel et al., 2008; EFAS, Thielen et al., 2009; ECMWF-ENS, Miller et al., 2010), (ii) statistical approaches based on a search for analogues (e.g. Obled et al., 2002; Hamill and Whitaker, 2006; Messner and Mayr, 2010; Marty et al., 2012).

The principle of analogue methods (AMs) is to use deterministic meteorological outputs of a target day  $D$  as an input for a statistical search of similar past days in terms of general circulation patterns. Similarity is measured on a set of relevant predictor variables computed by Numerical Weather Prediction (NWP) models. Finally the analogue situations selected are used to produce a sample of observed daily precipitations, which provides a probabilistic forecast for the day  $D$ .

The AMs assume a relationship between large-scale variables (predictors) and a local-scale target variable (predictand). Different predictors can be found in the literature. Table 1 presents a short description of a selection of representative models developed to predict precipitation fields under various climates. Note that this table does not include methods using combination of predictors in linear regression to estimate predictands. One can note that the number of predictors may differ from one region to another. Timbal et al. (2008) suggest defining combinations of predictors adapted for each season and each climatic region.

The predictors can be divided into two categories. Most AM developments deal with one or more low flow atmospheric fields, *i.e.* descriptors of the meteorological state of the atmosphere at

synoptic scale and a set of geopotential heights is usually chosen. The number of geopotential heights and their corresponding pressure level varies from one study to another: Altava-Ortiz et al. (2006) used three geopotential heights at 500, 850 and 1000 hPa to estimate precipitation amounts for one major heavy rainfall event in Catalonia (Spain) while Diomede et al. (2008) selected the geopotential at 500 hPa to estimate precipitation in Northern Italy. Following Bontron and Obled (2005), Horton et al. (2012) and Chardon et al. (2014) selected analogous situations based on geopotential heights at 500 and 1000 hPa-level in Switzerland and in France, respectively.

The second category of predictors is made of moisture related variables with among them, the large-scale precipitation outputs PRCP from NWP models (e.g. Diomede et al., 2014; Turco et al., 2011). Themeßl et al. (2011) found PRCP the most important predictors for local precipitation far beyond the other variables related to the state of the atmosphere. However Dayon et al. (2015) have compared different versions of AM and pointed out large possible biases in an application to France, when PRCP is used. In Catalonia, Barrera et al. (2007) added the humidity at 1000 hPa to the set of predictors suggested formerly by Altava-Ortiz et al. (2006). Similarly Bliefernicht and Bardossy (2007) selected three predictors (two geopotential heights and the moisture flux at 700 hPa defined by the product of the specific humidity at 700 hPa and the westerly wind) to predict rainfall in the Rhine basin in Germany. Bontron (2004) and later Gibergans-Báguena and Llasat (2007) and Diomede et al. (2008) tested different combinations of variables as predictors. Diomede et al. (2008) showed that geopotential at 500 hPa and vertical velocity at 700 hPa are the most relevant set of predictors to estimate precipitation in Northern Italy while Bontron (2004) highlighted that the best moisture related variable to be used in addition to synoptic scale variables to estimate precipitation in South-Eastern France is a combination of relative humidity and total column precipitable water. More exhaustively, Gibergans-Báguena and Llasat (2007) have identified 7 variables among a list of 33 thermodynamic descriptors and the AM combining these 7 variables to geopotential fields as predictors outperforms the other tested AMs. These works show that improved analogues sorting technique can be expected by incorporating new levels of analogy (i.e. incorporating new predictors and related similarity criteria). Note that analogy between meteorological situations can be also used in a post-processing of precipitation forecasts to improve the prediction skill of NWP models (e.g. Wang and Fu, 2009).

We will focus here on the application of AMs, and more specifically on one of the reference PQPF methods already used in France in an operational context and suggested by Bontron and Obled (2005) following the work of Bontron (2004). This former method was developed considering small basins located in South-Eastern France, subject to flash-flood events. The aim of this paper is to

extend the scope of this AM, considering large French river basins under oceanic influences. This study presents then two improvements of the reference method and compares the relative performances of the latter with two alternatives, on the large French basin of the Saône river. We consider two contexts: (i) a “perfect prognosis context” where the meteorological fields of the target day  $D$  are directly built with observed data or re-analysed by a NWP model; (ii) an “operational forecast” context where the meteorological fields are provided by an operational NWP model, including forecast errors.

Section 2 presents the study area and the datasets used, and section 3 introduces the scores selected for the evaluation of probabilistic forecasts. Section 4 gives the main principles of the benchmark method for searching analogues, based on a two-step procedure, plus two alternatives adding supplementary explanatory variables, air temperature and vertical air motion. The main results of the comparison of the methods are analysed in the fifth section, considering the “perfect prognosis” and the “operational forecast” contexts. The last section draws general conclusions.

## **2. Study area and data**

### **2.1 Study area**

This application of the AM focuses on the Saône River basin located in eastern France. With a response time of 12 to 24 h, the Saône River is one of the most important tributaries of the Rhône River. This basin is under oceanic influences and westerly fluxes generating rainfall when large fronts pass but it can be also affected by heavy rainfall events extending from the Mediterranean region. For this application, the total drainage area ( $\sim 30000 \text{ km}^2$ ) is divided into three sub-catchments defined by three gauging stations (Figure 1). This partition is considered in the operational forecasting system developed by the French electricity producer CNR (Compagnie Nationale du Rhône).

### **2.2 Meteorological archive**

Three well-known re-analysis archives have originally been identified for this study: (i) the first National Centers for Environmental Prediction/National Center for Atmospheric Research re-analysis (NNR) (Kalnay et al., 1996); (ii) the 45-Year European Centre for Medium-Range Weather Forecasts Re-Analysis (ERA), available from 1957 to 2002 (Uppala et al., 2005); (iii) the new generation of European reanalyses now available, called ERA-Interim (Dee et al., 2011). Finally, based on the results of previous works (Ben Daoud et al., 2008), ERA was chosen.

.

### 2.3 Precipitation

A 39-year precipitation data set over the 1970-2008 period was extracted from the 8-km-resolution atmospheric re-analysis over France performed with the Safran gauge-based system (Quintana-Seguí et al. 2008, Vidal et al., 2010). However the years 1970 and 1971 were only available after the optimisation phase in perfect prognosis context (section 4). These additional data were then used for the analysis of AMs performances in both real-time and perfect prognosis contexts (section 5). Basin scale precipitation data at daily time step (from 6 UTC the current day to 6 UTC the day after) were obtained by superimposing the delineation of the three sub-basins taken from Sauquet (2006) on the Safran grid. Table 2 shows the main characteristics of the Saône sub-catchments: proportion of dry days  $F0$ , daily rainfall amounts with return periods of 5 and 10 years  $PJXA5$  and  $PJXA10$  given by the Gumbel distribution fitted to the annual maximum series, and maximum values  $Pmax$  observed over the period.

These statistics show that only 20% of the days are dry, a rather low value given the climatological regime of the region (semi-continental climate). That may be explained by the size of the catchment leading to the assimilation of several rain gauges among which at least one of them reported one non-zero value. This effect of spatialisation should be emphasised by the aggregation over the sub basins. Regarding heavy events, the influence of the orography is illustrated by the high values for the Doubs basin, which is characterised by the Jura mountain. However, the downstream section of the Saône basin is also characterised by high values, even higher than for the Doubs basin ( $Pmax$ ). This fact is closely related to southerly heavy rainfall events extending from the Mediterranean region, especially in autumn.

Besides, since i) climatological regimes and precipitation distributions of the sub-catchments are different and ii) precipitation amount is an asymmetric variable, a set of transformed precipitation data was built for the development step of the AM. The data were transformed following this equation:

$$P = \sqrt{\frac{P_{raw}}{PJXA10}} \quad (1)$$

Where  $P$  is the transformed value,  $P_{raw}$  is the raw precipitation value and  $PJXA10$  the value with return period 10 years.

### 3. Numerical scores for performance evaluation

Performance evaluation of any forecasting system relies on numerical scores to assess the accuracy of the forecasts (see Jolliffe and Stephenson (2003) for an extended list of scores). In this study,

several scores are used:

- i) A global performance criterion on the overall forecasts (i.e. considering no rain situations as well as precipitation event). The Continuous Ranked Probability Skill Score was chosen to perform the optimisation process;
- ii) End-user oriented criteria based on contingency tables (i.e considering the ability of the system to detect specific event), used to evaluate specifically the optimised AMs. Several scores from contingency tables with prescribed threshold have been used.

### 3.1 Continuous Ranked Probability Skill Score (CRPSS)

The *CRPSS* score is based on the Continuous Ranked Probability Score (*CRPS*; Brown, 1974; Matheson and Winkler, 1976; Hersbach, 2000) which is a continuous version of the widely used Ranked Probability Score. *CRPS* is equivalent to the mean absolute error of a deterministic forecast. It is especially recommended for the verification of continuous variables probabilistic forecasts when performances are evaluated for all the range of possible values (no threshold is considered). The *CRPS* is defined as:

$$CRPS = \frac{1}{N} \sum_{i=1}^N \int_{-\infty}^{+\infty} [F(x) - H_{x_0}(x)]^2 dx \quad (2)$$

where  $N$  is the number of forecasts,  $x_0$  is the observed value,  $F$  is the distribution function of the forecasted variable  $x$ , and  $H_{x_0}$  is the Heaviside function (set to 0 for values below  $x_0$  and 1 for values above  $x_0$ ). The minimum value of *CRPS* is 0, when the forecast system is perfect and there is no theoretical upper bound for *CRPS*. Thus, the lower the *CRPS* the more useful the forecast. Furthermore, the *CRPS* takes into account i) the forecast resolution part, which represents the uncertainty of the probabilistic forecast, and ii) the forecast reliability part, that could be interpreted as the error between the observed value and the median value of the  $F(x)$  distribution (Hersbach, 2000; Candille and Talagrand, 2005). In this work, we followed Bontron (2004) who defined the *CRPS* as a score of utility, which is easily decomposed into a resolution part (*Reso*) and a reliability part (*Reli*):

$$CRPS = Reso + Reli$$

$$\text{where } Reso = \int_{-\infty}^{+\infty} [F(x) - H_{Q_{50\%}}(x)]^2 dx \text{ and } Reli = \int_{x_0}^{Q_{50\%}} [1 - 2F(x)] dx \quad (3)$$

and  $Q_{50\%}$  is the 50% quantile of the distribution function  $F$ .

For the interpretation of the results in terms of gain, it is widespread to use a reference forecast.

Thus, the relative *CRPS* skill score (*CRPSS*) could be used instead:

$$CRPSS = 1 - (CRPS_M / CRPS_{ref}) \quad (4)$$

where  $CRPS_M$  and  $CRPS_{ref}$  are the scores of method  $M$  and of a reference method, respectively. In this paper the reference method will be given by the climatological distribution of the predictand  $x$ . Other references such as persistence or random forecast may be used (Wetterhall et al., 2005; Diomede et al., 2008). If *CRPSS* is 0, there is no improvement from method  $M$  with regard to the reference. If *CRPSS* is higher than 0,  $M$  performs better than the reference method. If *CRPSS* is 1, forecasts obtained with  $M$  are perfect (i.e.  $CRPS_M = 0$ ). Note that this skill score may be applied to the resolution and reliability parts of the *CRPS* (Eq. 3).

Several methods are compared in the following section and it is more appropriate to evaluate the gain  $G$  of performance of any improvement in AMs. When one method  $B$  is compared to one method  $A$ ,  $G$  is defined by the relative skill score of the *CRPS* of the method  $B$  where the method  $A$  is considered as the reference method:

$$G(B/A) = 1 - (CRPS_B / CRPS_A) \quad (5)$$

The perfect score  $G$  is 1, whereas a null value of  $G$  means no improvement in forecast skill provided by the method  $B$ .

### 3.2 Scores from contingency tables

An operational stakeholder is mainly interested in a forecasting system which performances are satisfactory with regard to his expectations. For instance, irrigation decision management needs information on the probability of additional rainfall provided by short-term weather forecasts, while flood warning services need to estimate if heavy rainfall (over a specific threshold) is expected before providing flood warning alerts. In this context, additional scores are computed from contingency tables.

To build a contingency table, each forecast is considered as a forecast of the occurrence or non-occurrence of an event, that will or will not effectively occur. In the precipitation context, an event is defined by a precipitation threshold  $T$ . Exceedance over specific thresholds  $T$  is examined using various quantiles  $P_f$  taken from the empirical distribution of precipitation amounts derived from analogous days. Thus, within a period of evaluation of  $n$  days, given a precipitation threshold  $T$ , the contingency table (see table 3) is fulfilled as followed, considering for each day the observed and forecasted precipitation values  $P_o$  and  $P_f$ :

- $N_{CA}$  = number of correct alarms ( $P_f > T$  and  $P_o > T$ );
- $N_{FA}$  = number of false alarms ( $P_f > T$  and  $P_o \leq T$ );
- $N_{MA}$  = number of missed alarms ( $P_f \leq T$  and  $P_o > T$ );
- $N_{CR}$  = number of correct rejections ( $P_f \leq T$  and  $P_o \leq T$ ).

According to the methodology, the sum ( $N_{CA} + N_{FA} + N_{MA} + N_{CR}$ ) is the total number  $n$  of days considered in the evaluation period. If the forecast system is perfect, then it would produce only correct alarms and correct rejections, no false alarm and no missed alarm. Once the contingency table is completed, it is possible to compute a panel of scores (Stephenson, 2000). In this study, three scores were calculated:

$$\begin{aligned} POD &= N_{CA} / (N_{CA} + N_{MA}) \quad ; \quad FAR = N_{FA} / (N_{CA} + N_{FA}) \\ TSS &= N_{CA} / (N_{CA} + N_{MA}) - N_{FA} / (N_{FA} + N_{CR}) \end{aligned} \quad (6)$$

The Probability Of Detection ( $POD$ ) measures the probability that an observed event is well forecasted by the system.  $POD$  is between 0 and 1 (1 for a perfect forecast). The False Alarm Rate ( $FAR$ ) gives the probability that an expected event did not occur.  $FAR$  is between 0 and 1 (0 for a perfect forecast) but could not be calculated if no event is forecasted, what might occur for a high threshold  $T$ . Since no system is perfect, the better way is to strike a balance between a  $POD$  high enough and a  $FAR$  low enough. The True Skill Statistic ( $TSS$ ), also known as Hanssen-Kuipers score, takes into account both correct alarms and false alarms. It corresponds to the difference between  $POD$  and the Probability Of False Detection, as defined by Wilks (2006). It gives an overview of forecast system skill, since it is based on all indices of the contingency table.  $TSS$  has the disadvantage to be not so useful for extreme events since it could converge to  $POD$  ( $N_{CR}$  becoming very large). This score has a range of -1 to +1 (1 for a perfect forecast, 0 indicating no skill).

#### 4. Development of the analogue method

Duband (1970) initiated the development of the AM in France, which was followed by several academic works by Guilbaud (1997), Bontron (2004) and Ben Daoud (2010). This paper relates to the latter work, focussed on the case of large basin under oceanic influences, such as the Saône river basin. We will show how the introduction of additional variables enables to provide a better forecast skill.

Each time a new variable is introduced, a new step of analogues selection is created in the algorithm. According to this methodology, it is necessary to define several parameters and variables: predictors, spatial domain on which the analogy is searched for, number of analogues for each successive step, and also a criterion to measure the quality of the analogy. These parameters are obtained by optimisation.

During this optimisation phase, the whole period where predictors from the meteorological archive and predictand from the hydrological archive are available (1972-2002) was divided into two independent sub-periods: the calibration period over which the method was optimised (~25 years), and the validation period over which the results were then validated (~5 years). Splitting into calibration-validation periods was not undertaken by considering two successive sub-periods, as it could introduce a climatological bias (succession of alternating wet and dry periods, heterogeneity in the meteorological archive, stationarity issues related to climate change), but rather by removing five evenly spaced years from the whole period (1978, 1983, 1988, 1993 and 1998).

In this study, the choice of the variables was mainly guided by their meteorological meaning related to the mechanism responsible of a precipitation event. Several variables expected to be useful for precipitation forecasting were tested in the analogues selection algorithm (equivalent potential temperature as an indicator of air masses transition, potential vorticity as a predictor for lifting processes at synoptic scale, etc.).

Finally, two additional variables were introduced: temperature and vertical velocity. The main results obtained from the optimisation phase are presented in the following sub-sections after a brief description of the starting AM.

#### **4.1 Method A: The benchmark method for searching analogues**

Bontron (2004) developed a version of the analogue method (called hereafter method A) adapted for small to medium-sized sub-catchments. The selection of analogous days is divided into two successive steps: the first step relies on an analogy on the general synoptic circulation based on geopotential heights fields (Guilbaud and Obled, 1998), whereas the second one involves humidity predictors at local scale.

Let's consider a target day  $D$  for which meteorological forecasts are generated. Before applying the algorithm of analogues selection, the past candidate situations are pre-selected by taking seasonal effects into account. Indeed, according to Lorenz (1969), factors such as vegetation, sea temperature or snowpack, depending on the period of the year (i.e. the season), should have impacts on precipitation formation processes. Consequently, the dates of candidate situations should be

restricted to a temporal window around the target date, which period was optimised and fixed to 4 months by Bontron (2004).

Method A is divided in two steps (Figure 2):

- A first step of selection (L1-A) is applied in order to search for  $N1$  situations that are analogous to the target one, based on geopotential heights at 1000 and 500 hPa-levels ( $Z1000$  &  $Z500$ ) considered at 12 UTC the day  $D$  and 00 UTC the day  $D+1$  respectively. The analogy criterion is the  $SI$  score (Teweles and Wobus, 1954), which emphasises analogy in fields shape (already used by Guilbaud and Obled, 1998). The optimised spatial domain of analogy is large enough to catch the main features that are responsible for precipitation;
- A second step (L2-A) is applied in order to search, among the  $N1$  dates, for  $N2$  situations that are analogous to the target one, based on a humidity variable defined by the product of relative humidity at 850 hPa-level ( $RH850$ ) and the total column of water ( $TCW$ ), considered both at 12 UTC the day  $D$  and 00 UTC the day  $D+1$ . The analogy criterion is the root mean square error ( $RMSE$ ). The spatial domain of analogy is small, reduced to a 2x2 grid points centred on the target catchment, meaning that humidity should be considered at a local scale.

Considering the study area and the archives containing predictors and precipitation, the parameters of the two steps of analogue selection (numbers  $N1$  and  $N2$  and the two spatial domains) were optimised over the whole period.  $N1$  and  $N2$  were fitted by testing different couples ( $N1, N2$ ), where  $N1$  is a multiple of 10 set between 10 and 200 and  $N2$  a multiple of 5 set between 5 and  $N1$ . The spatial domains were optimised by increasing them successively in each direction and by retaining at each step the best domain among the four expanded domains. One can note that an extended version of this procedure was proposed by Radanovics et al (2013). In this study, the optimised numbers  $N1$  and  $N2$  are 70 and 25 respectively, and the optimised spatial domains are  $[-5^{\circ}\text{E}$  to  $15^{\circ}\text{E}$ ;  $40^{\circ}\text{N}$  to  $50^{\circ}\text{N}]$  and  $[5^{\circ}\text{E}$  to  $7.5^{\circ}\text{E}$ ;  $45^{\circ}\text{N}$  to  $47.5^{\circ}\text{N}]$  for geopotential analogy (step L1-A) and moisture related analogy (step L2-A), respectively.

Method A has been implemented at the Laboratoire des Transferts en Hydrologie et Environnement in Grenoble since 2004. It was used in real-time by the Observatoire Hydro-météorologique Méditerranéen Cévennes Vivarais (<http://www.ohmcv.fr/>) to anticipate extreme rainfall events in south-eastern France until January 2015.

## 4.2 Method B: Temperature as a seasonal indicator

In method A, a large sample of past situations contained in a moving temporal window of 4 months is considered as candidate situations for the preliminary step of the selection. For instance, to forecast precipitation expected for the 1st of January, candidate situations for analogy are considered between the 1st of November and the 1st of March of every year of the period covered by the reanalyses archive. This constraint on seasonality is also applied by other authors (see for example Gibergans-Baguena and Llasat, 2007; Diomedea et al., 2008). One main drawback is that selecting candidate situations in a limited temporal window may be too restricting, removing sometimes very informative analogues. The suggested solution to overcome this limitation is to make a pre-selection in method B based on a variable with a pronounced seasonal pattern more appropriate than the date itself. The search for such a variable leads naturally to the air temperature (*Tair*). Values of *Tair* between 925 hPa and 500 hPa were tested by creating a preliminary step (L0-B) before (L1-A). Several times within the target day *D* were also considered: from -12h to +42h by step of 6h (-12h is corresponding to 12 UTC the previous day of the day *D*).

The optimal predictor was determined through an optimisation procedure. To limit calculation time, before starting the optimisation procedure several parameters of this new analogy step were carefully fixed *a priori*:

- the spatial domain considered for this preliminary step L0-B is reduced to the grid point nearest to the target catchment gravity centre, here the Saône River basin (5°E; 47,5°N) since the seasonal effect should be monitored at the most local scale;
- the numerical criterion that measures analogy is *RMSE* (another measure of distance could be used, however the *SI* score that accounts for the shape of the temperature field should not be relevant *a priori*);
- the number of preselected situations is assumed to be roughly equivalent to the number of situations automatically chosen by method A (i.e. about 3000 for a historical period of 30 years);
- the subsequent steps of method A are applied without any change (L1-B=L1-A and L2-B=L2-A).

A first testing was carried out by analysing *Tair* at only one pressure level and one time. Figure 3 shows that *Tair* should be analysed in very low troposphere (925 hPa) at +36h, leading to a gain of +2.7% towards method A. The fact that temperature should be taken beyond the standard period of total rainfall, which is from +6h to +30h, was rather unexpected.

We assume that this time lag is related to the propagation of the frontal rainfall systems. Let us

consider a front that defines the boundary between two air masses with contrasting thermal states. It is well-known that precipitations usually occur along the front. Searching analogy after 30h allows detecting changes in the air mass since temperature is an indirect way to follow the movement of the fronts. The link to the passage of frontal systems is partly confirmed by a further analysis discriminating wet days from dry days. For each of the two samples, the range for pressure level and time defined below is screened and distinct combinations are finally obtained for dry days and for wet days (*i.e.* 700 hPa at +12h and 925 hPa at +36h, respectively).

A further step in the optimisation consisted in measuring the benefit of introducing several predictors related to *Tair* with distinct pressure levels and times. All possible combinations with different numbers of pressure levels and times (1 level and 1 time, 1 level and 2 times, 2 levels and 1 time for each level) were investigated, leading to the following: *Tair* should be taken at 925 hPa at +36h and at 600 hPa at +12h. Finally, the other parameters formerly chosen based on expert knowledge were revised considering this optimal set of *Tair* predictors:

- the optimal number of preselected candidates (2900) is finally close to the initial number obtained through the temporal window (*i.e.* 3000);
- performances decrease when the size of the domain increases. The elementary grid domain (only one grid point) initially considered is found optimal;
- different numerical criteria for analogy were tested (mean absolute error, correlation coefficient, *SI* score) and the best performance in terms of *CRPSS* is obtained with *RMSE*.

The gain of method *B* towards method *A* over the calibration period and over the validation period arises to +4.0% and +4.5%, respectively, showing that the introduction of temperature in a preliminary analogue selection step (L0-B) leads to better performance than with a moving temporal window.

### **4.3 Method C: Accounting for the vertical motion**

This section aims at introducing a new predictor which accounts for vertical motions, especially frontal lifting, that may be responsible of well spatially-organised rainfall fields over large areas.

#### *4.3.1 Choice of the predictor*

Here several predictors extracted from ERA were tested, among them:

- Equivalent potential temperature (*EPT*; Rossby, 1932), frequently analysed in weather

forecasting, which allows to characterise humidity and temperature gradients of the air mass. An important gradient of *EPT* in low layers of the troposphere (usually at 850 hPa pressure level) is observed when a frontal zone crosses one region;

- Horizontal divergence (*DIV*), which fields enable the forecaster to detect convergence zones and thereafter the presence of a front when *DIV* is analysed in low troposphere;
- Vertical velocity (*VV*), which represents the vertical motion that occurs in low troposphere when two different air masses are confronted.

A correlation analysis was performed between the predictors already used in method *B* and the set of variables (*EPT*, *VV* and *DIV*), for the nearest grid point to the Saône River basin (5°E; 47,5°N). Results showed a significant relationship between *EPT* at 850 hPa and geopotential height at 500 hPa, with a correlation coefficient of 0.72, which is partly explained by thermodynamic laws. Conversely there is no obvious link between variables used in method *B* and *DIV* or *VV* (the coefficients of correlation range from 0 to 0.25). Based on this analysis, the two variables *DIV* and *VV* were tested in the AM.

Since the geopotential fields mostly control vertical motions, the selection of analogues with *DIV* or *VV* was introduced after L1-B but with two options (before or after considering the moisture information, i.e. between L1-B and L2-B, or after L2-B) using prescribed parameters for this additional step. The first option provides the better result and scores obtained with *VV* were better than with *DIV* (*G* value maxima of +1.75% for *DIV* at 925 hPa against +2.24% for *VV* at 850 hPa, towards method *A*). This leads to the choice of *VV* to account for vertical motion processes through a selection step between L1-B and L2-B. This version of the AM is called hereafter method *C*, where L1-C=L1-B, L2-C is additional selection step, and L3-C=L2-B.

One should note that *VV* is usually considered as a good predictor for small scale convective systems such as thunderstorms. Indeed, this kind of convection, which is frequently due to topography (mountains), is associated with strong vertical velocities. However, given the rough spatial resolution of the reanalysis, and due to the poor representation of the topography, proposed *VV* values cannot be strictly interpreted in terms of low level orographic lifting but variability in *VV* results rather from the frontal lifting due to contrasting air masses.

#### 4.3.2 Optimisation

As for the pre-selection with temperature, it is necessary to determine beforehand some parameters

to limit calculation time during the optimisation process. Then the methodology followed for the optimisation of this new selection step is based on assumptions. The parameters that may be easily fixed, for which the impact on performances is assumed to be low, are (i) the number of analogues (set at 50) and (ii) the times of the predictor observation (set at +12h and +24h).

The analogy criterion used for the preliminary testing with *VV* was the *RMSE* score. Mean absolute error, correlation coefficient, Teweles-Wobus score were also considered and the associated analogy domain optimised for each criterion. Results showed that gains of about +2% towards method *B* are obtained with *MAE* and *RMSE*.

Secondly, the optimisation of the numbers of analogues retained at L1-C (predictor is geopotential height) and L2-C (predictor is *VV*) was carried out. At the preliminary step L0-C (predictor is temperature), there is no reason for re-optimising the number of analogues, whereas at the last step the number of 25 analogues was kept to allow comparison with the scores obtained through the method *B*. Finally, the optimum numbers are 170 for L1-C and 70 for L2-C, based on the additional predictor *VV*. The related gain *G* towards method *B* arises +4.52%, against only +2.24% before re-optimisation of these numbers. This improvement highlights the importance of the re-optimisation when new predictors are introduced.

Lastly, the predictor itself was optimised, that is the choice of the pressure level and times at which *VV* should be considered. We followed the same methodology applied for temperature. A first testing was performed by screening all combinations of only one level pressure and one time (Figure 4), and for each combination, the spatial domain was optimised. Positive values of the gain *G* towards method *B* are obtained within the standard period of total rainfall, and *G* is maximal when *VV* is taken in low troposphere (850 to 700 hPa) at +18h. Thus, the pressure level was fixed to 850 hPa, and all combinations of 1 to 5 observation times among the 10 possible times were successively considered. Again, the analogy domain was optimised consistently. Figure 5a displays optimal combinations of times and Figure 5b gives the related value of gain *G* towards method *B*. Figure 5a shows a symmetry in the repartition of the times. This result highlights that the search for analogy on vertical motions does not give priority to any particular time. Further, the gain towards method *B* increases with the number of times involved, arising more than +6% when *VV* is analysed at all times of the rainfall period: *VV* should be used at every time of the rainfall period. Although the result is not shown here, the optimal analogy domain is identical for all combinations but one. This domain, formed by  $3 \times 2$  grid points, is centred on the Saône catchment. Besides, it is interesting to note that this domain sets between the domains applied for L1-C and L3-C, indicating that the way this intermediate selection is introduced is consistent with a regular decrease of domain

size with the step of analogy (Figure 6).

However, a further testing for the choice of the predictor was carried out by discriminating wet days from dry days. Results showed that for wet days, the highest gain (+4% towards method *B*) is obtained for four times instead of five to put emphasise on wet days forecast. The following parameters of the newly introduced selection step L2-C are therefore:

- *RMSE* is the analogy criterion;
- the predictors are *VV* values at 850 hPa and at +6h, +12h, +18h, and +24h;
- the size of domain stands between previous step and next step (for the Saône catchment, 2.5°E to 7.5°E and 45°N to 47.5°N);
- $N2 = 70$  analogue days are chosen among the  $N1 = 170$  ones selected according to the geopotential heights.

Considering these evolutions, the performances were then computed by applying the updated method over the calibration and validation periods. The corresponding gains towards method *B* over the two periods reach +5.94% and +5.91%, respectively.

Lastly, a slight improvement of the AM was obtained through the re-optimisation of the predictor of the last selection step (L3-C). Indeed, new pressure levels are available in the ERA-40 reanalyses: henceforward, two variables describing the relative humidity are taken both at 925 and 700 hPa-levels instead of a single one at 850 hPa-level, and again at the times +12h and +24h. The related gain reaches +2.2% over both the calibration and the validation periods.

The new method C is organised in four steps (Figure 7): a preliminary selection step L0-C (identical to L0-B) through temperature; a second one L1-C (identical to L1-B and L1-A) through geopotential heights, a third one L2-C searching for the analogy on vertical motions, and the last one through humidity L3-C (similar to L2-B and L2-A).

## **5. Accuracy of perfect prognosis and real time forecasts**

### **5.1 Perfect prognosis forecasts**

During the development phase of the algorithm, parameters were optimised in a perfect prognosis mode over the calibration period, and improvements were confirmed over the validation period. In this section, the total length of available record was considered for searching analogues dates, *i.e.*

from 01/08/1970 to 30/08/2002. That means an extension of 7 years, compared to the previous section (5 years of the validation period defined for section 4, plus the 2 years 1970 and 1971 that were then available in the SAFRAN archive, see section 2.3). Since the numbers of dates selected at each step strongly depend on the length of the available learning period, all numbers were re-optimised leading to  $N0=3900$  instead of 2900;  $N1=330$  instead of 170;  $N2=90$  instead of 70; whereas  $N3=25$  remains identical.

For each of the three methods, PQPFs were provided over the whole period of 32 years. The *CRPSS* was computed on the three sub-catchments, in addition to the gain  $G$  towards method *A*, and averaged. The score *CRPS* was also decomposed into a resolution part and a reliability part (see section 3.1). Following the same methodology, the gains  $G$  towards method *A* were obtained in terms of resolution and reliability (Table 4). The decomposition of the *CRPS* score shows that most improvements are related to a gain in resolution. Nevertheless, PQPFs provided by methods *B* and *C* are also more reliable than those from method *A*, with positive values of  $G$ . It is interesting to note that the gain of skill with method *C* is rather similar to the one obtained by Bontron (2004) when humidity variables (*TCW* and *RH850*) were introduced in the AM.

The *TSS* score was computed with each method for different probability thresholds derived from the predictive cumulative distribution. These values are defined by quantiles used to characterise the distribution: a central value (60% quantile) bounded by two values that form an asymmetric confidence interval (20% and 90% quantiles). Different events defined by precipitation thresholds  $T$  were examined: 0 mm (threshold of rain/no rain), and others corresponding to quantiles from the marginal probability distribution of non-zero rainfall amounts.

Figure 8 shows that performances depend on both the intensity of the event (precipitation threshold) and on the predictive quantile: when the forecast value is defined by the 20% quantile of the predictive cumulative distribution, the highest skill is obtained for dry days, whereas the highest score is obtained for high precipitation thresholds (here the 80% quantile of distribution of non-zero values) when the forecast is defined by the 90% quantile. This latter result highlights the under-estimation often observed with an AM; i.e. heavy rain events are anticipated using a high predictive quantile. The shapes of the curves look very similar demonstrating no major discrepancy between method performances. However the method *C* seems to outperform the two other methods when the forecast is defined by a small to medium predictive quantile. For the highest precipitation thresholds, performances obtained with methods *B* and *C* are equivalent. Furthermore, by comparing the score differences between methods, the source of improvement is obviously the introduction of *VV*, especially for medium to high predictive quantiles.

The previous results naturally lead to the following question: are the gains due to a better ability to detect events? Are they due to a decrease of false alarms? Figure 9 shows *POD* and *FAR* scores obtained considering the 60% predictive quantile, function of the precipitation threshold. Values for *POD* are roughly similar between the three methods, with a slight advantage for method *B*. By contrast, the graph of *FAR* scores reveals noticeable differences. A reduction of the *FAR* score of 5 to 10% for medium to high precipitation thresholds is obtained when vertical velocity is used as predictor, whereas the use of temperature to select candidate situations does not improve the false alarm rate (method *B* equivalent to method *A*). In conclusion, in a perfect prognosis mode, the improved skill of the method *C* is mainly due to a reduced false alarm risk. Given the intermediate performances of method *B* and the small differences in terms of *FAR* score, only results of methods *A* and *C* are presented thereafter.

## 5.2 Real time forecasts

In the previous sections, the AMs were optimised and evaluated in a perfect prognosis context (predictors are extracted from a reanalysis). During this calibration phase, predictand was defined by the transformed value of precipitation (see section 2.3) and predictors came from re-analysis. In this section, we suggest assessing the performance of the AMs in near real time conditions, i.e. the prediction scheme is kept unchanged, but predictors are now provided by a NWP model embracing meteorological uncertainties due to modelling errors and biases while the predictand is now defined by the raw value of precipitation.

Predictors extracted from forecasts available in the Meteorological Archive Retrieval System of the European Centre for Medium-Range Weather Forecasts (ECMWF) are the inputs to the AMs. Performances are compared over the 01/10/2001-01/10/2004 period. The choice of the period was partly conditioned by the frequent changes in operational forecasting system (modifications were made in 2001 and later in 2006). A period of three years was considered long enough to reasonably allow an objective evaluation of PQPFs. The initialisation of each forecast run is at 00h UTC, and the maximum lead-time for PQPFs is eight days. An adjustment was made since temperature is not available at 600 hPa-level over the selected period. Different pressure levels close to 600 hPa-level were tested for method *C* in a perfect prognosis context before retaining the temperature at 700 hPa. Since raw values of precipitation are used as predictand here, and given the differences between the climatological regimes of the three sub-catchments of the Saône basin, in this section *CRPSS* scores were computed for each sub-catchment separately. The *CRPSS* score was evaluated at each lead-time (from the current day  $D$  to the day  $D+7$ ) over the 3-year period with  $CRPS_{ref}$  given by forecasts

based on climatology. The evolution of the *CRPSS* scores with the lead-time, obtained for the Doubs sub-catchment with methods *A* and *C*, is shown in Figure 10. Note that results are quite similar for the two other sub-catchments. Method *C* outperforms method *A* up to the lead-time  $D+2$  and the skill of both methods decreases gradually with the lead-time, becoming close from  $t=D+3$  up to  $t=D+7$ .

As noted in the previous section, introducing the air temperature in the preliminary selection step does not lead to substantial improvement for this basin (comparison between results from methods *A* and *B*). The higher efficiency obtained for method *C* for the three first lead-times (from days  $D$  to  $D+2$ ) results mainly from the use of vertical velocity as additional predictor. We may suspect that starting from day  $D+3$  the better efficiency initially shown by method *C* is neutralised by the inability of the ECMWF model to forecast *VV* for longer lead-times. The poor performances of AM predictions beyond 6 days may similarly be related to the poor reliability of the geopotential forecasts. Thus, the first step of analogue selection may not be pertinent enough, leading to irrelevant predictions at following steps. Otherwise, whatever the method the *CRPSS* becomes negative by day  $D+6$ , meaning that beyond, the forecast based on climatology brings more information about rainfall than the AMs. Then, with the current NWP models, forecasting precipitation by using the AM beyond six or seven days ahead may not be of interest.

### 5.3 Real time mode *versus* perfect prog mode

The drop of forecast efficiency with the lead-time is probably due to the increasing inability of the NWP model to forecast the predictors. In order to check if this assumption is correct, forecasts were run over the same period in a context of perfect prognosis, i.e. predictors come from the ECMWF analyses. Figure 11 displays scores obtained with methods *A* and *C*, both in a perfect prognosis context (PERF) and in real conditions (OPER), for the Doubs sub-catchment, function of the lead-time, with: a) the *CRPSS* scores, b) the resolution and c) the reliability parts of the *CRPS* score.

The accuracy of the weather forecasts is quite high for the first lead-time  $D$  since for each version of the AM the curves OPER and PERF merge. Starting from the day ahead ( $D+1$ ), according to the graphs a) (*CRPSS*) and c) (skill of reliability) the accuracy of operational forecasts decreases for each AM version, while the graph b) (skill of resolution) shows no strong tendency with the lead-time. Resolution of the PQPFs provided by the two versions of the AM is not affected by the accuracy of the NWP forecasts (Figure 11 b). Consequently, this result indicates that the resolution of the PQPFs does depend mainly on the algorithm of selection of analogues. The comparison of the curves of methods *A* and *C* shows that PQPFs from the method *C* have a much better resolution than those from the method *A*. Then the predictor *VV* tends to reduce forecast uncertainties. One

should note another important result from Figure 11c: even though for the first lead-time the skill of reliability is higher for method *C* than for method *A*, method *C* becomes less reliable than method *A* from day  $D+2$ . Variables *VV* are thus less robust predictors compared to the others (geopotential heights and humidity): *VV* forecasts quality is sharply downgraded with the lead-time. However, the discrepancy between the two curves is tiny from day  $D+4$ , showing that the errors in other predictor forecasts become predominant.

## 6. Conclusions

The aim of this work was the adaptation of an analogue method, initially developed in the 2000s for small to medium-sized mountain catchments, to large river basins mainly influenced by frontal systems at large scale. In this study, we focused on the Saône basin located in eastern France.

Hence several tests were performed in order to improve the selection of analogues over large river basins, by proposing new predictors that are useful for a weather forecaster. Especially, different predictors that are representative of the large-scale lifting process were tested. As outcomes of this exploration, two main successive evolutions of the method were found. In the previous version of the AM (Bontron, 2004), seasonal effects were taken into account by considering only past dates included in the same 4-month temporal windows. This fixed criterion of pre-selection has been replaced by a preliminary selection based on temperature. Yet one should note that this novel evolution is not specific to large river basins. In contrast, the second evolution of the AM enables to characterise and localize large-scale fronts that are the main generator of rainfall over the large river basins in France. Among the predictors tested, the vertical velocity proved the most pertinent. It should be noted that given the climatology of rainfall events and the topography of the zone of interest in this study, local lifting due to orography is negligible and then not taken into account. Here vertical velocity is only considered as a characteristic of large scale air vertical lifting. This new step of analogue selection was inserted between the two steps of the previous version of the AM.

All parameters of these new analogue selection steps (number of analogues, times and pressure levels of the predictor, spatial domain of analogy, optimization criterion) were optimised over a calibration (learning) period in a perfect prognosis context. Results of the optimisation show a substantial increase of QPF skill, which was preserved over an independent validation period. Thus the gain of skill is of the same order of magnitude as the one obtained by Bontron (2004) when humidity variables were introduced as predictors in the AM. This upgraded analogues sorting technique was recently applied to the whole of France (Radanovics et al., 2013) through SANDHY

### “Stepwise ANalogue Downscaling for HYdrology”.

An analysis was performed to deal with the study of the AM performances in depth, both in perfect conditions, where predictors are observations, and in real conditions, where predictors come from weather forecasts provided by the ECMWF model. Results show that:

- The main improvement of the overall skill is due to a lower false alarm rate, as the detection of precipitation events is not very different between the updated AM and the previous one.
- In real conditions, the overall increase in skill obtained with the new AM is perceptible only until the lead-time  $D+2$ . As a consequence, by  $D+3$  the better efficiency due to the analogy on vertical motions is neutralised by the inability of the ECMWF model to forecast correctly the vertical velocity. Vertical velocity even downgrades PQPFs skill at longer lead-times.
- On the one hand, from a certain lead-time PQPF reliability does not depend only on the quality of VV forecasts, but also on the quality of the predictors used to select the first sample of analogues (geopotential heights). On the other hand, PQPF resolution does not depend on NWP forecasts but only on the algorithm of selection of analogues itself.

The three versions of the AM presented here are associated with different levels of complexity. Comparing methods allow assessing the relative interest of replacing the calendar criterion with an analogy on temperature (method *A* versus method *B*) and of the introduction of a new analogy step based on vertical velocity (method *C* versus method *B*). The gain obtained with the introduction of the air temperature is less than the gain obtained with a new step. Finally, method *B* leads to intermediate prediction skills, between methods *A* and *C*. One should note that air temperature is certainly a more universal predictor than vertical velocity (adapted to localize large frontal systems in France) and thus that method *B* may perform better than method *C* under other climates. One should also note that method *B* requires less computational resources than method *C*, which could be an advantage for an application in an operational context.

In conclusion, there are still several ways to improve the AM: in the context of precipitation forecasting particularly, the AM needs to be adapted to the zone of interest by selecting other predictors closely linked to physical processes that are responsible of precipitation events. This can be done especially as new generations of reanalyses are still emerging, containing new predictors at higher and higher spatial resolution. In the case of French large river basins, large scale vertical motions may be exploited in the AM to improve PQPF skill. However, when a new predictor is introduced in the algorithm, one must pay attention to the capacity and reliability of the NWP models in providing forecasts for this predictor, especially at long lead-time. As verified in the past,

AM performances for precipitation forecasts will probably increase in parallel with the improvement of NWP models to more distant lead-times.

## Acknowledgements

The financial support provided by CNR and Cemagref-Irstea for the PhD research of A. Ben Daoud is gratefully acknowledged. The NCEP/NCAR and ECMWF ERA-40 re-analysis data were obtained from the NOAA web site (<http://www.cdc.noaa.gov/>) and from the ECMWF data server ([http://data.ecmwf.int/data/d/era40\\_daily/](http://data.ecmwf.int/data/d/era40_daily/)), respectively. The authors acknowledge Météo-France for providing the ECMWF forecasts and the Safran data set for research purposes. The authors acknowledge the anonymous reviewers for their very helpful comments.

## References

- Altava-Ortiz, V., Barrera, A., Llasat, M.C., Prat, M.A., Gibergans-Báguena, J., and Barnolas, M.: Application of the MM5 and the analogous method to heavy rainfall events. The case of 16–18th October 2003 in Catalonia (NE Spain). *Adv. Geosci.*, 7, 313–319, 2006.
- Barrera, A., Altava-Ortiz, V., Llasat, M.C. and Barnolas, M.: Heavy rain prediction using deterministic and probabilistic models – the flash flood cases of 11–13 October 2005 in Catalonia (NE Spain). *Adv. Geosci.*, 12, 121–126, 2007.
- Ben Daoud, A., Sauquet, E., Obled, Ch., Lang, M., and Bontron, G.: A flood forecasting system based on probabilistic precipitation scenarios: performance evaluation and sensitivity analysis to the re-analysis archive. *Proceedings of the International Interdisciplinary Conference HydroPredict'2008 on Predictions for Hydrology, Ecology, and Water Resources Management: Using Data and Models to Benefit Society*, Prague, Czech Republik, 207-210, 15-18 Sept. 2008.
- Ben Daoud, A.: Améliorations et développements d'une méthode de prévision probabiliste des pluies par analogie. Application à la prévision hydrologique sur les grands bassins fluviaux de la Saône et de la Seine (in French), PhD Thesis, Université de Grenoble, 2010.
- Bliefernicht, J. and Bárdossy, A.: Probabilistic forecast of daily areal precipitation focusing on extreme events. *Nat. Hazards Earth Syst. Sci.*, 7, 263-269, doi:10.5194/nhess-7-263-2007, 2007.
- Bompart, P., Bontron, G., Celie, S., and Haond, M.: An operational hydrometeorological forecasting chain for CNR's hydroelectric production needs. *La Houille Blanche*, 5, 54-60, 2009.
- Bontron, G. and Obled, Ch.: A probabilistic adaptation of meteorological model outputs to hydrological forecasting. *La Houille Blanche*, 1, 23-28, doi: <http://dx.doi.org/10.1051/lhb:200501002>, 2005.
- Bontron, G.: Prévision quantitative des précipitations: adaptation probabiliste par recherche

- d'analogues. Utilisation des réanalyses NCEP-NCAR et application aux précipitations du Sud-Est de la France (in French), PhD Thesis, Institut National Polytechnique de Grenoble, 2004.
- Brown, T.A.: Admissible scoring systems for continuous distributions. The Rand Paper Series P-5235, The Rand Corporation, Santa Monica, California, 1974.
- Candille, G. and Talagrand, O.: Evaluation of probabilistic prediction systems for a scalar variable. *Journal of the Royal Meteorological Society*, 1-20, doi:10.1256/qj.04.71, 2005.
- Cannon, A.J.: Nonlinear analog predictor analysis: a coupled neural network/analog model for climate downscaling. *Neural Networks*, 20, 444-453; doi:10.1016/j.neunet.2007.04.002, 2007.
- Chardon, J., Hingray, B., Favre, A.-C., Autin, P., Gailhard, J., Zin, I. and Obled, Ch.: Spatial Similarity and Transferability of Analog Dates for Precipitation Downscaling over France. *Journal of Climate*, 27, 5056–5074. doi: <http://dx.doi.org/10.1175/JCLI-D-13-00464.1>, 2014.
- Dayon, G., Boé, J., and Martin, E.: Transferability in the future climate of a statistical downscaling method for precipitation in France. *Journal of Geophysical Research, Atmospheres* 02/2015; DOI: 10.1002/2014JD022236, 2015.
- Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E. V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P.,Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. Roy. Meteorol. Soc.*, 137, 553–597, doi:10.1002/qj.828, 2011.
- Diomede, T., Marsigli C., Montani, A., Nerozzi, F., and Paccagnella, T.: Calibration of Limited-Area Ensemble Precipitation Forecasts for Hydrological Predictions. *Mon. Wea. Rev.*, 142, 2176–2197. doi: <http://dx.doi.org/10.1175/MWR-D-13-00071.1>, 2014.
- Diomede, T., Nerozzi, F., Paccagnella, T., and Todini, E.: The use of meteorological analogues to account for LAM QPF uncertainty. *Hydrology and Earth System Sciences*, 12, 141-157, 2008.
- Duband, D.: Reconnaissance dynamique de la forme des situations météorologiques. Application à la prévision quantitative des précipitations (in French), PhD Thesis, Faculté des Sciences de Paris, 1970.
- Gibergans-Báguena, J., and Llasat C.L.: Improvement of the analog forecasting method by using local thermodynamic data. Application to autumn precipitation in Catalonia, *Atmospheric Research*, 86, 173-193, 2007.
- Guilbaud, S. : Prévision quantitative des précipitations journalières par une méthode statistico-dynamique de recherche d'analogues (in French). PhD Thesis, Institut National Polytechnique de

Grenoble, 1997.

- Guilbaud, S., and Obled, Ch.: L'approche par analogues en prévision météorologique. *La Météorologie*, 8, 24, 21-35, 1998.
- Hamill, T.M., and Whitaker, J.S. : Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application. *Mon. Wea. Rev.*, 134, 3209–3229, 2006.
- Hersbach, H.: Decomposition of the ranked probability score for ensemble prediction systems, *Weather Forecasting*, 15, 559-570, 2000.
- Horton, P., Jaboyedoff, M., Metzger, R., Obled, Ch., and Marty, R.: Spatial relationship between the atmospheric circulation and the precipitation measured in the western Swiss Alps by means of the analogue method. *Natural Hazards and Earth System Sciences*, 12:777–784, 2012.
- Jolliffe, I.T., and Stephenson, D.B.: *Forecast verification. A practitioner's guide in atmospheric science*, John Wiley & Sons Edition, 2003.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S, White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*, 77, 437-471, 1996.
- Lacaze, Y., Javelle, P., Goutx, D., and Berthet, L.: Sensitivity of flood forecasting models towards the quality of future rainfalls: a case study of sensitivity analysis and assessment of the operational efficiency on the river Yonne and the river Loing, *Colloque SHF Prévisions Hydrométéorologiques*, 18-19 Nov., Lyon, 2008.
- Lorenz, E.N.: Atmospheric predictability as revealed by natural occurring analogues, *Journal of the Atmospheric Sciences*, 26, 636-646, 1969.
- Marsigli, C., Boccanera, F., Montani, A., and Paccagnella, T.: The COSMO-LEPS mesoscale ensemble system: validation of the methodology and verification. *Nonlinear Processes in Geophysics*, 12, 527–536, 2005.
- Marty, R., Zin, I., Obled, C., Bontron, G., and Djerboua, A.: Toward real-time daily PQPF by an analog sorting approach: Application to flash-flood catchments, *J. Appl. Meteorol. Clim.*, 51, 505–520, doi:10.1175/JAMC-D-11-011.1, 2012.
- Matheson, J. E. and Winkler, R. L.: Scoring rules for continuous probability distributions. *Manage Sci.*, 22, 1087–1095, 1976.
- Mehrotra, R., Evans, J. P., Sharma, A., and Sivakumar, B.: Evaluation of downscaled daily rainfall hindcasts over Sydney, Australia using statistical and dynamical downscaling approaches. *Hydrology Research*, 45(2), 226, DOI: 10.2166/nh.2013.094, 2014.

- Messner, J. W. and Mayr, G. J.: Probabilistic forecasts using analogs in the idealized Lorenz96 setting. *Monthly Weather Review*, 139, 1960-1971, 2010.
- Miller, M., Buizza, R., Haseler, J., Hortal, M., Janssen, P., and Untch, A.; Increased resolution in the ECMWF deterministic and ensemble prediction systems. *ECMWF Newsletter*, 124, 10–16, 2010.
- Obled, Ch., Bontron, G., and Garçon, R.: Quantitative precipitation forecasts: a statistical adaptation of model outputs through an analogues sorting approach, *Atmospheric Research*, 63, 303-324, 2002.
- Osca, J., Romero, R., and Alonso, S.: Precipitation projections for Spain by means of a weather typing statistical method. *Global and Planetary Change*. 10/2013, DOI: 10.1016/j.gloplacha.2013.08.001, 2013.
- Palatella, L., Miglietta, M.M., Paradisi, P., and Lionello, P.: Climate change assessment for Mediterranean agricultural areas by statistical downscaling. *Nat. Hazard Earth Syst Sci*, 10? 1647–1661, 2010.
- Quintana-Segui, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L., and Morel, S.: Analysis of near surface atmospheric variables: Validation of the SAFRAN analysis over France. *Journal of Applied Meteorology and Climatology*, 47, 92-107, 2008.
- Radanovics, S., Vidal, J.-P., Sauquet, E., A. Ben Daoud, A. and Bontron G.: Optimising predictor domains for spatially coherent precipitation downscaling. *Hydrology and Earth System Science*, 17, 4189–4208, doi:10.5194/hess-17-4189-2013, 2013.
- Ribalaygua, J., Torres, L., Pórtoles, J., Monjo, R., Gaitán, E., and Pino, M. R.: Description and validation of a two-step analogue/regression downscaling method. *Theoretical and Applied Climatology*, 114(1-2), 253-269, doi: 10.1007/s00704-013-0836-x, 2013.
- Rossby, C.G.: Thermodynamics applied to air mass analysis. *Massachusetts Institute of Technology, Papers in Physical Oceanography and Meteorology*, 1, n°3, 1932.
- Salathé, E.P.: Comparison of various precipitation downscaling methods for the simulation of streamflow in a rainshadow river basin. *International Journal of Climatology*, 23(8), 887-901, 2003.
- Sauquet, E.: Mapping mean annual river discharges: geostatistical developments for incorporating river network dependencies. *Journal of Hydrology*, 331, 300-314, 2006.
- Schenk, F. and Zorita, E.: Reconstruction of high resolution atmospheric fields for Northern Europe using analog-upscaling, *Clim. Past*, 8, 1681-1703, doi:10.5194/cp-8-1681-2012, 2012.
- Schmidli, J., Goodess, C. M., Frei, C., Haylock, M. R., Hundecha, Y., Ribalaygua, J., and Schmith

- T.: Statistical and dynamical downscaling of precipitation: An evaluation and comparison of scenarios for the European Alps. *J. Geophys. Res.*, 112, D04105, doi:10.1029/2005JD007026, 2007.
- Stephenson, D.B.: Use of the odds ratio for diagnosing forecast skill. *Weather and Forecasting* No. 15, 221-232, 2000.
- Teng J., Chiew, F.H.S., Timbal, B., Wang, Y., Vaze, J., and Wang., B.: Assessment of an analogue downscaling method for modelling climate change impacts on runoff, *Journal of Hydrology*, 472–473, 23, 111-125, <http://dx.doi.org/10.1016/j.jhydrol.2012.09.024>, 2012.
- Teweles, J., and Wobus, H.: Verification of prognosis charts. *Bulletin of the American Meteorological Society*, 35, 455-463, 1954.
- Themeßl J.M., Gobiet, A., and Leuprecht, A.: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *Int. J. Climatol.*, 31, 1530–1544. doi: 10.1002/joc.2168, 2011.
- Thielen, J., Bartholmes, J., Ramos, M.H., and Roo, A.: The European Flood Alert System – Part 1: Concept and development. *Hydrology and Earth System Sciences*, 13(2), 125-140, 2009.
- Thirel, G., Rousset-Regimbeau, F., Martin, E., and Habets, F.: On the Impact of Short-Range Meteorological Forecasts for Ensemble Streamflow Predictions, *Journal of HydroMeteorology*, 9, 1301–1317, 2008.
- Timbal, B., Dufour, A., and McAvaney, B.J.: An estimate of climate change for Western France using a statistical downscaling technique. *Clim. Dyn.*, 17, 947-963, 2003.
- Timbal, B., Li, Z., and Fernandez, E.: The Bureau of Meteorology Statistical Downscaling Model Graphical User Interface: user manual and software documentation. CAWCR technical report, 4, 90pp. <http://www.cawcr.gov.au/publications/technicalreports.php>, 2008.
- Turco, M., Quintana-Seguí, P., Llasat, M. C., Herrera, S., and Gutiérrez, J. M.: Testing MOS precipitation downscaling for ENSEMBLES regional climate models over Spain, *J. Geophys. Res.*, 116, D18109, doi:10.1029/2011JD016166, 2011.
- Uppala, S.M., Kållberg, P.W., Simmons, A.J., Andrae, U., da Costa Bechtold, V., Fiorino, M., Gibson, J.K., Haseler, J., Hernandez, A., Kelly, G.A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R.P., Andersson, E., Arpe, K., Balmaseda, M.A., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B.J., Isaksen, L., Janssen, P.A.E.M., Jenne, R., McNally, A.P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N.A., Saunders, R.W., Simon, P., Sterl, A., Trenberth, K.E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-

- analysis, Quarterly Journal of the Royal Meteorological Society, 131, 2961-3012, 2005.
- Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M., and Soubeyroux, J.-M.: A 50-year high-resolution atmospheric reanalysis over France with the Safran system, International Journal of Climatology, DOI: 10.1002/joc.2003, 2010.
- Wang, H., and Fan, K.: A New Scheme for Improving the Seasonal Prediction of Summer Precipitation Anomalies. Weather Forecasting, 24, 548–554, <http://dx.doi.org/10.1175/2008WAF2222171.1>, 2009.
- Wetterhall, F., Halldin, S., and Xu, C-Y.: Statistical precipitation downscaling in central Sweden with the analogue method. Journal of Hydrology, 306, 174-190, 2005.
- Wilks, D.: Statistical Methods in Atmospheric Sciences, Elsevier, second edition, 627 pp., 2006.
- Wu, W., Liu, Y., Ge M., Rostkier-Edelstein D., Descombes, G., Kunin, P., Warner T., Swerdlin S., Givati, A., Hopson, T., and Yates D.: Statistical downscaling of climate forecast system seasonal predictions for the Southeastern Mediterranean. Atmospheric Research, 118(15), 346-356, <http://dx.doi.org/10.1016/j.atmosres.2012.07.019>, 2012.
- Zorita, E. and von Storch, H.: The analog method as a simple statistical downscaling technique: comparison with more complicated methods. Journal of Climate, 12(8):2474–2489, 1999.

**Table 1.** Predicators and study areas for a selection of AM applications (TCW: total column of water; RH: relative humidity, SLP: Sea Level Pressure; Z: geopotential height; T: temperature, SH: specific humidity; PRCP: Total precipitation; VV: vertical velocity; additional numbers indicate the related atmospheric level in hPa)

Reference	Study area	Predictors
Altava-Ortiz et al. (2006)	Catalonia (Spain)	Z850, Z1000
Barrera et al. (2007)	Catalonia (Spain)	Z500, Z850, Z1000, humidity at 1000 hPa
Bliefernicht and Bardossy (2007)	Western Germany	Z700, Z1000, SH700, westerly wind at 700 hPa
Bontron and Obled (2005)	South-Eastern France	Z500, Z10000, TCW, RH850
Cannon (2007)	British Columbia (Canada)	SLP, RH700, horizontal wind components at 700 hPa, Z500
Chardon et al. (2014)	France	Z500, Z1000
Dayon et al. (2015)	France	SLP, 2 m air temperature TAS, lifted condensation level, moisture flux at 850 hPa, totals total index, SH850
Diomede et al. (2008)	Reno river basin (northern Italy)	Z500, VV700
Diomede et al. (2014)	Northern Italy, Germany, Switzerland	PRCP
Gibergans-Báguena and Llasat (2007)	Catalonia (Spain)	Z700, Z1000, precipitable water mass in 700–500 hPa stratum, relative temperature at 850 hPa, instability indices (lifted index and K index), level at which the temperature is 0 °C, Convective Available Potential Energy, potential temperature gradient in 850–700 hPa stratum, equivalent potential temperature gradient of the surface - 950 hPa stratum.
Horton et al. (2012)	Swiss Alps (Switzerland)	Z500, Z1000
Marty et al. (2012)	Southern France	Z500, Z1000, TCW, RH850
Mehrotra et al. (2014)	Sydney region (Australia)	T700, T850, meridional gradient of Z850, VV500, VV850, Z850
Oscá et al. (2013)	Spain	SLP, T850
Palatella et al. (2010)	Apulia region and Po Valley (Italy), Ebro river basin (Spain), Antalya province (Turkey)	SLP, T500
Ribalaygua et al. (2013)	Aragon (Spain)	Speed and direction of the geostrophic wind at 500hPa and at 1000 hPa
Salathé (2003)	Washington and Oregon states (USA)	Z1000
Schenk and Zorita (2012)	Northern Europe and the Baltic Sea region	SLP
Schmidli et al. (2007)	Alps (France, Germany, Swiss, Italy, Austria)	Geostrophic wind (direction and velocity) at 1000 and 500 hPa
Teng et al. (2012)	South-Eastern Australia	SLP, PRCP, SH850, Surface max temperature Tmax, T850, zonal and meridional wind components at 850 hPa
Thiemeßl et al. (2011)	Austria	27 variables
Timbal et al. (2003)	Western France	SLP, Z1000, Z500, T850, Z1000-Z500, TCW
Timbal et al. (2008)	Australia	SLP, PRCP, SH850, RH850, RH700, Surface max temperature Tmax, T850, zonal wind components at 850 hPa, meridional wind components at 700 hPa and 850 hPa
Turco et al. (2011)	Spain	PRCP
Wetterhall et al. (2005)	South-central Sweden	SLP
Wu et al. (2012)	Southeastern Mediterranean	SLP;H700; T700;U850; V850
Zorita and von Storch (1999)	Spain, Portugal, Southern France, Northern Morocco, Northern Algeria	SLP

**Table 2.** Characteristics of the Saône basin pluviometric regime (frequency  $F0$  of 0 mm days (%), daily rainfall amounts (mm) with a return period of 5 ( $PJXA5$ ) and 10 years ( $PJXA10$ ), maximum amount (mm) ( $Pmax$ ) observed during the 39 year period (Aug. 1970 -July 2008).

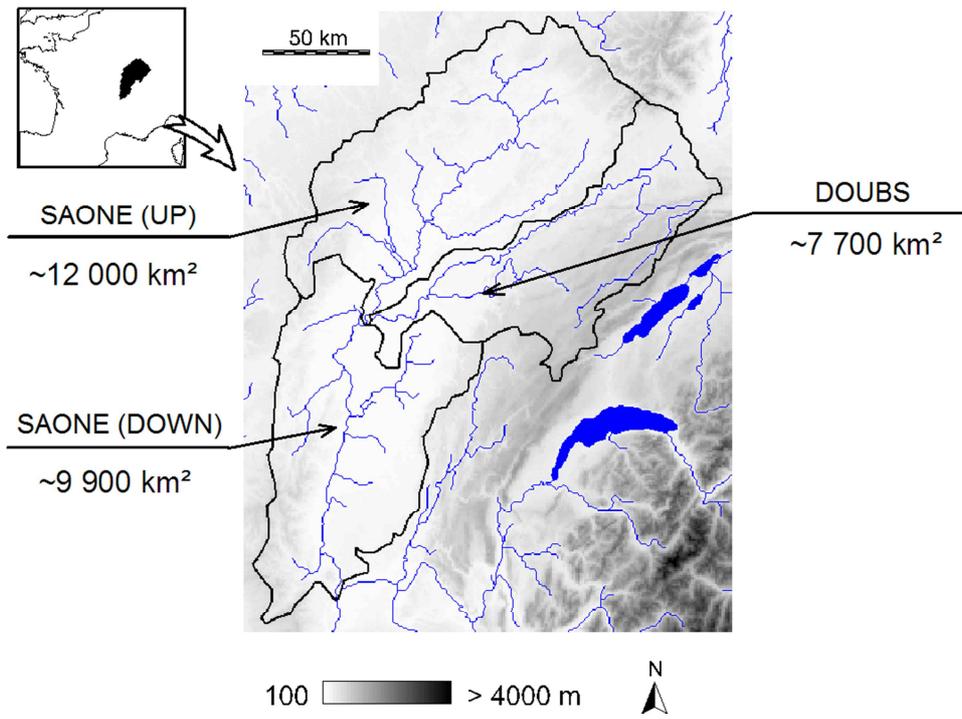
Catchment	$F0$	$PJXA5$	$PJXA10$	$Pmax$
Doubs	20,5	46,3	51,5	58,1
Saône (Up)	20	35,9	39,7	58,5
Saône (Down)	23	45,5	51,8	71,1

**Table 3.** Contingency table for a precipitation threshold  $T$  ( $N_{CA}$  = number of correct alarms;  $N_{FA}$  = number of false alarms;  $N_{MA}$  = number of missed alarms; and  $N_{CR}$  = number of correct rejections).

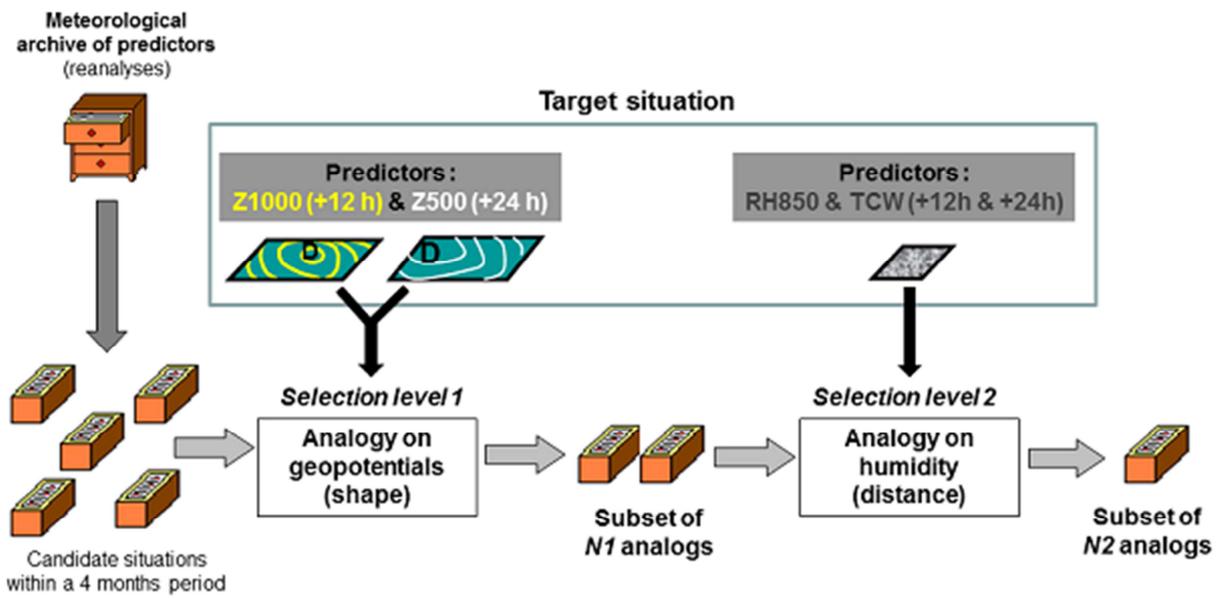
		Observation ( $P_o$ )	
		$P_o > T$	$P_o \leq T$
Forecast ( $P_f$ )	$P_f > T$	$N_{CA}$	$N_{FA}$
	$P_f \leq T$	$N_{MA}$	$N_{CR}$

**Table 4.** Gains  $G$  (%) of performance towards method  $A$  obtained by applying method  $B$  and method  $C$  (total gain and gain decomposed into resolution and reliability).

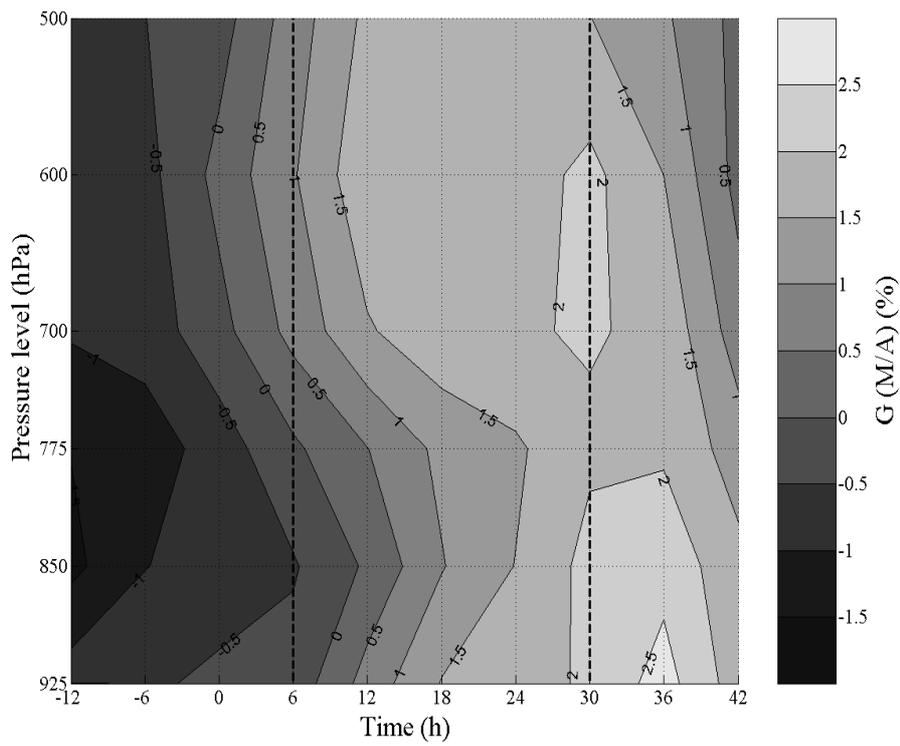
	Total	Reliability	Resolution
Method $B$	+3.7%	+2.7%	+4.7%
Method $C$	+12.3%	+7.8%	+17.0%



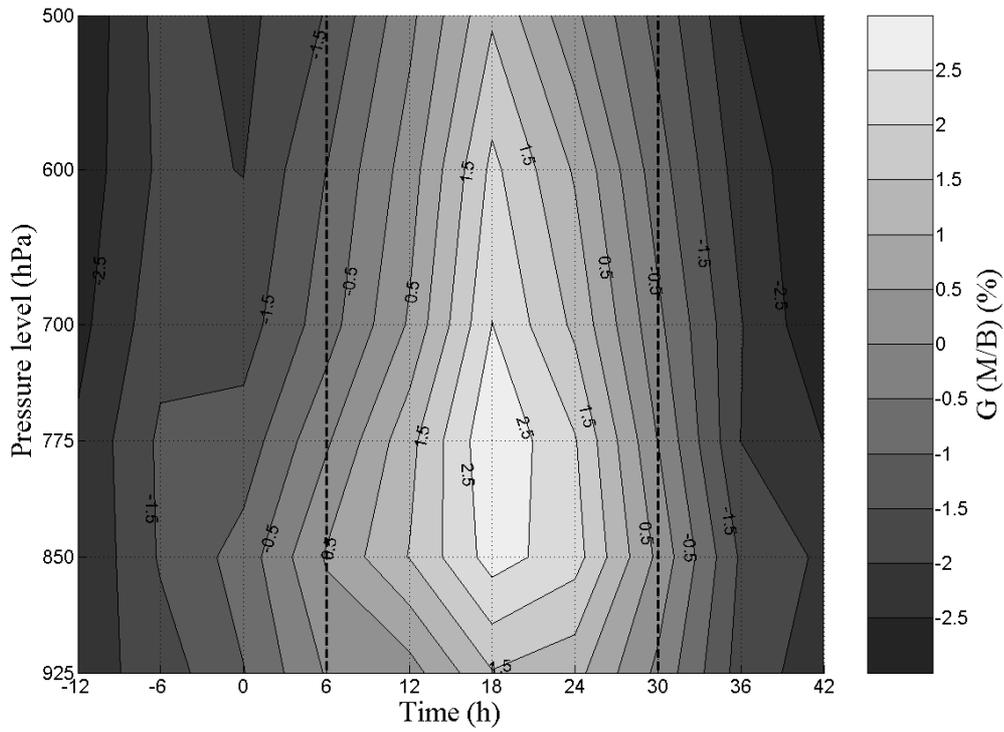
**Figure 1.** Location of the Saône river basin divided into three sub-catchments.



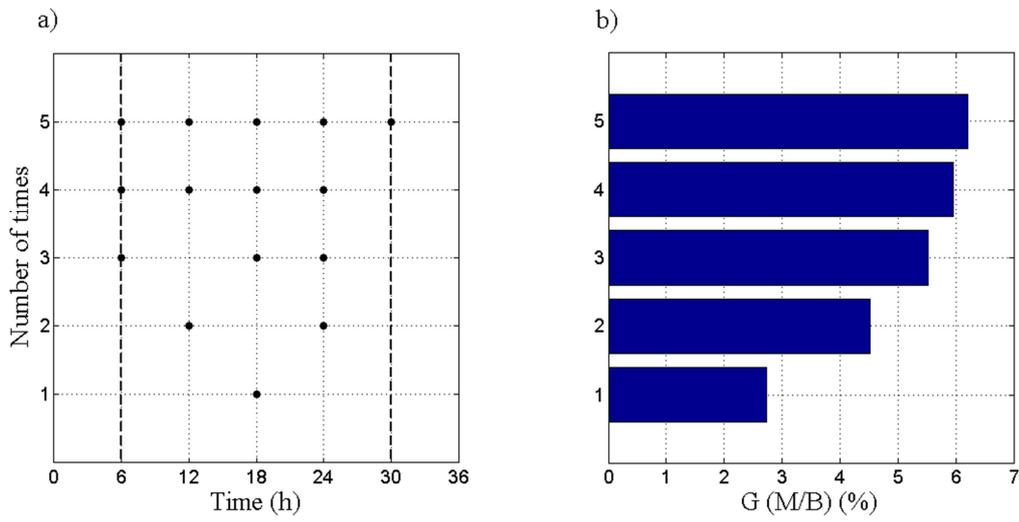
**Figure 2.** Schema of the analogues searching method (method A) developed by Bontron (2004).



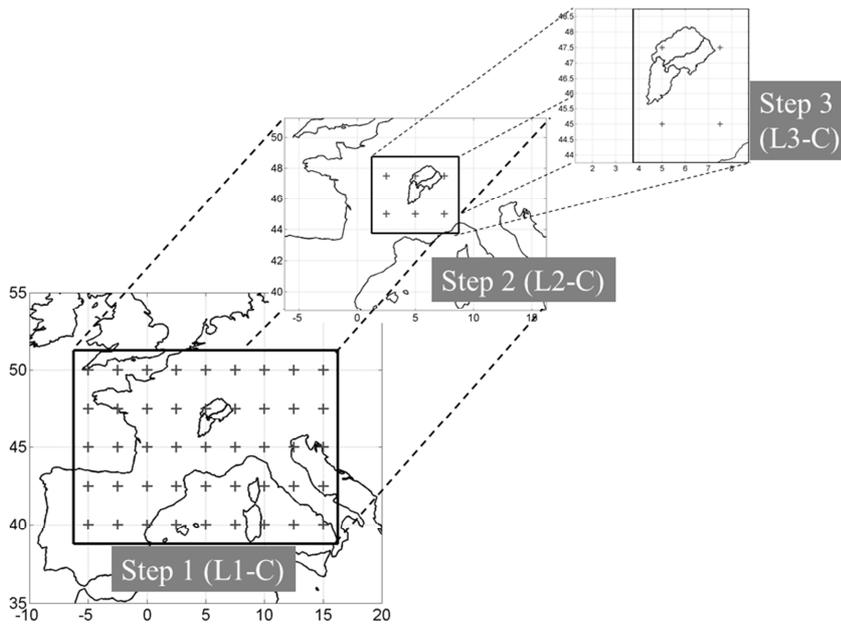
**Figure 3.** Gain of performance ( $G$  in %) expressed towards method  $A$ , function of the time and the pressure level of the temperature analysis. Vertical dashed lines represent the standard period of the total rainfall.



**Figure 4.** Gain of performance ( $G$  in %) expressed towards method  $B$ , function of the time and the pressure level of the vertical velocity analysis. Vertical dashed lines represent the standard period of the total rainfall.



**Figure 5.** a) Combination of times and the number of times and b) gain of performance  $G$  (in %) function of the number of times.



**Figure 6.** Schema showing the fitting of the analogy domains successively applied at step 1 (L1-C – identical to L1-B), new step 2 (L2-C - step with *VV*), and step 3 (L3-C – identical to L2-B). Crosses represent the mesh centres (grid points).

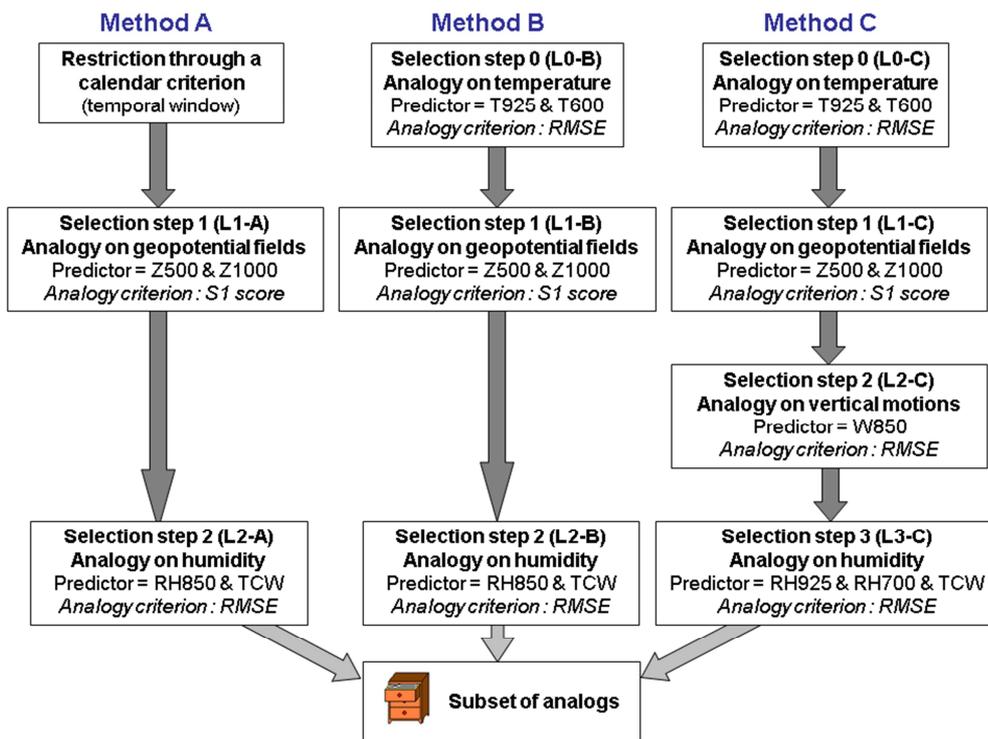
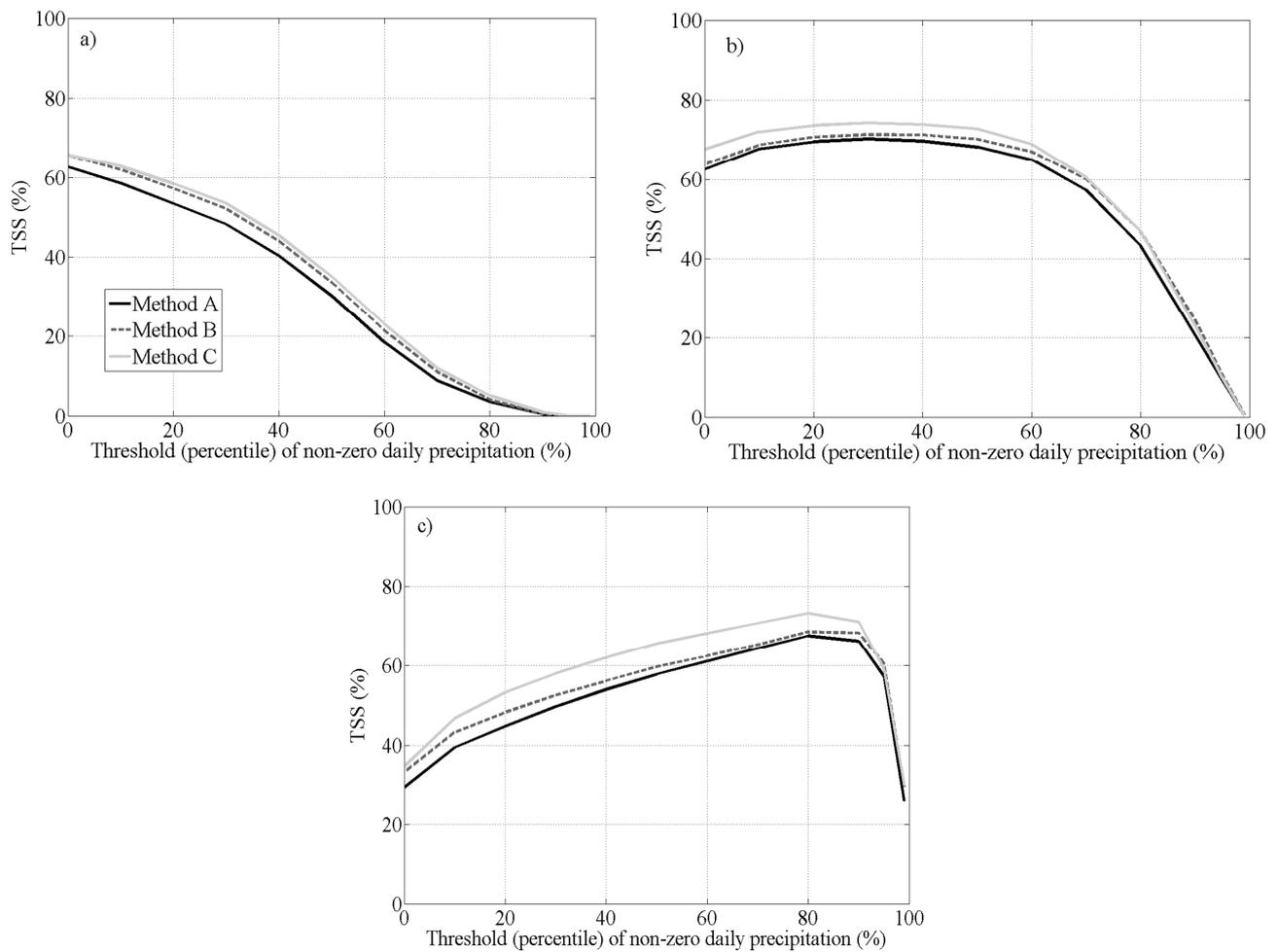
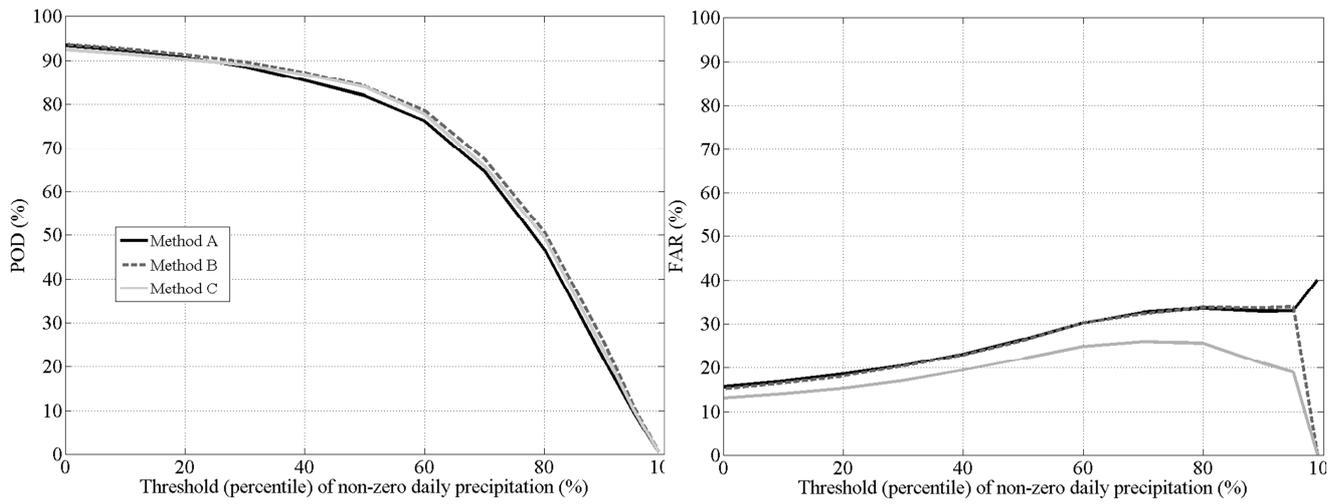


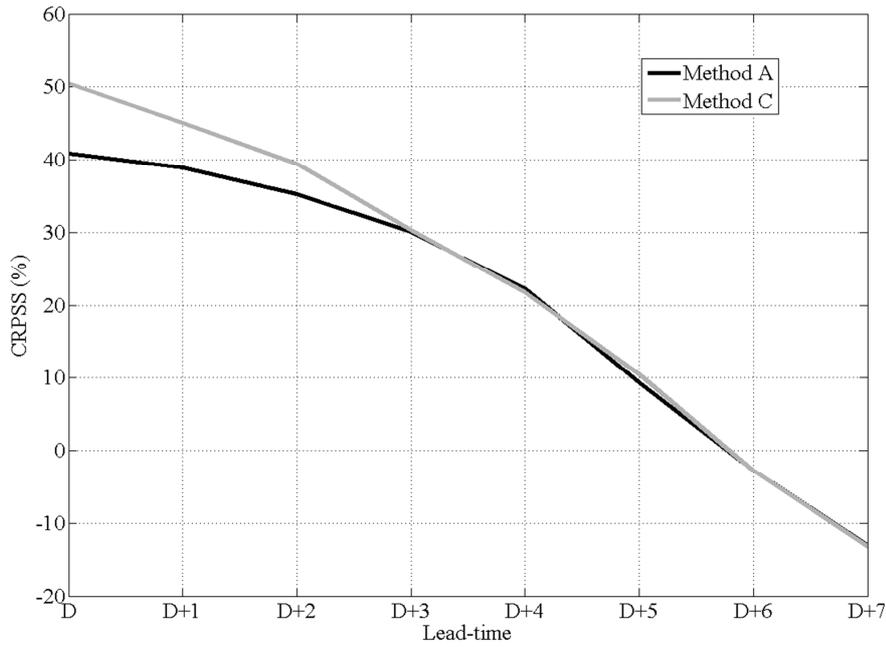
Figure 7. Schema of the methods of analogues searching.



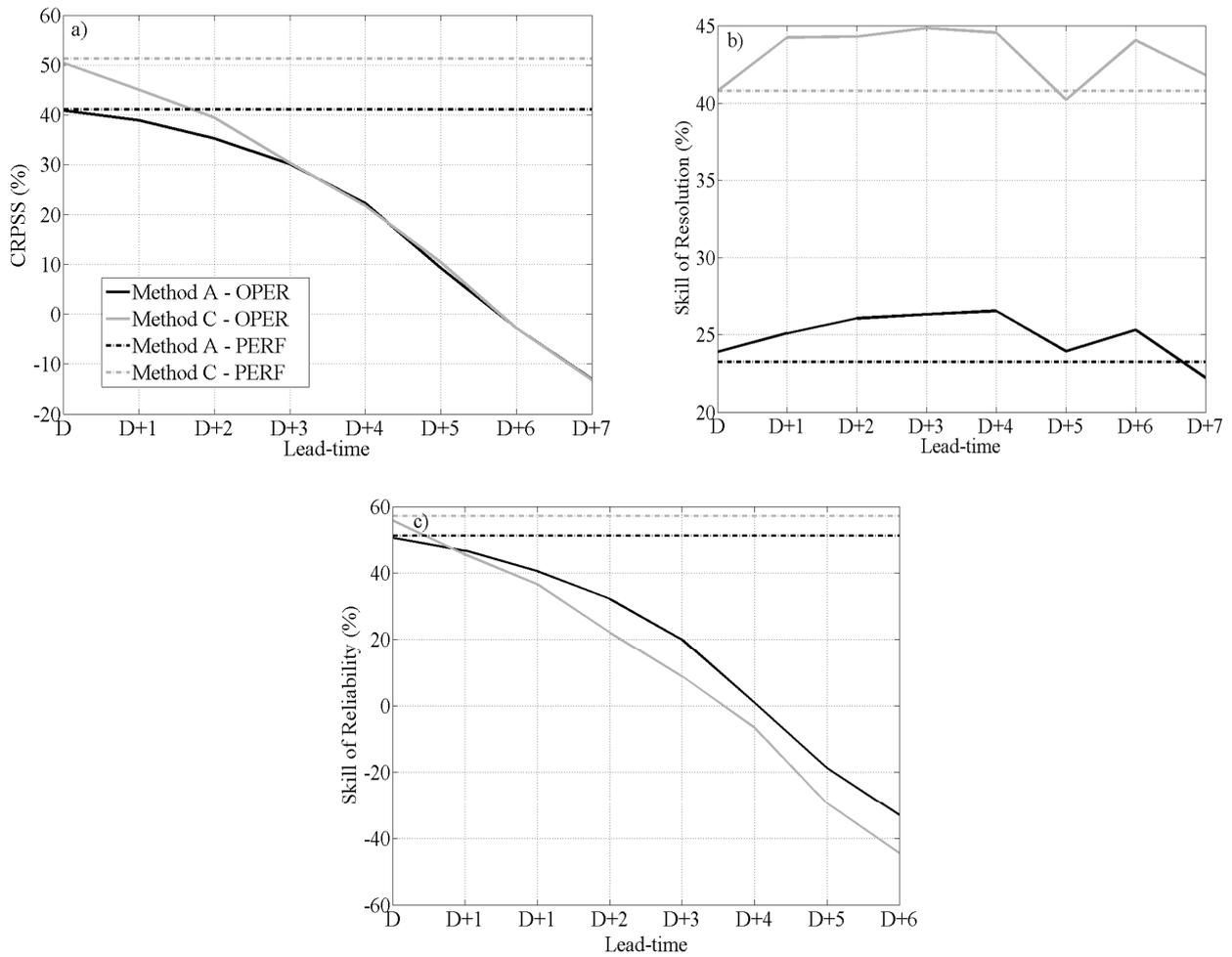
**Figure 8.** TSS evolution, according to precipitation thresholds defined by percentiles of the climatological distribution of non-zero, obtained in perfect prognosis with methods A, B and C. Forecast value is represented by a) the 20% quantile, b) the 60% quantile and c) the 90% quantile of the predictive distribution.



**Figure 9.** *POD* (left) and *FAR* (right) scores evolution, according to different precipitation thresholds defined by percentiles of the climatological distribution of non-zero, obtained in perfect prognosis with methods *A*, *B* and *C*. Forecast value is represented by the 60% quantile of the predictive distribution.



**Figure 10.** CRPSS score as function of the lead-time, obtained in real conditions (operational forecasts) with methods A and C for the Doubs sub-catchment. The reference forecast is the climatological areal rainfall distribution



**Figure 11.** Scores as function of the lead-time, obtained in both perfect prognosis context (dash-dot lines) and real conditions (solid lines), with methods A and, for the Doubs sub-catchment. a) *CRPSS*, b) Skill of Resolution part of the *CRPS* score, and c) Skill of Reliability part of the *CRPS* score. The reference forecast is the climatological areal rainfall distribution.