

# Building long-term and high spatio-temporal resolution precipitation and air temperature reanalyses by mixing local observations and global atmospheric reanalyses: the ANATEM model

A. Kuentz<sup>1,2,\*</sup>, T. Mathevet<sup>1</sup>, B. Hingray<sup>3,4</sup>, and J. Gailhard<sup>1</sup>

<sup>1</sup>Électricité de France, DTG, BP 41, 38040 Grenoble CEDEX 09, France

<sup>2</sup>Irstea, Hydrosystems and Bioprocesses Research Unit (HBAN), 1 rue Pierre-Gilles de Gennes, CS 10030, 92761 Antony CEDEX, France

<sup>3</sup>CNRS, LTHE UMR 5564, Grenoble, F-38000, France

<sup>4</sup>Univ. Grenoble Alpes, LTHE UMR 5564, Grenoble, F-38000, France

\*presently at: Swedish Meteorological and Hydrological Institute, 601 76 Norrköping, Sweden

Correspondence to: A. Kuentz (anna.kuentz@gmail.com)

## Abstract

Efforts to improve the understanding of past climatic or hydrologic variability have received a great deal of attention in various fields of geosciences such as glaciology, dendrochronology, sedimentology and hydrology. Based on different proxies, each research community produces different kinds of climatic or hydrologic reanalyses at different spatio-temporal scales and resolutions. When considering climate or hydrology, many studies have been devoted to characterising variability, trends or breaks using observed time series representing different regions or climates of the world. However, in hydrology, these studies have usually been limited to short temporal scales (mainly a few decades and more rarely a century) because they require observed time series (which suffer from a limited spatio-temporal density).

This paper introduces ANATEM, a new method that combines local observations and large-scale climatic information (such as the 20CR Reanalysis) to build long-term air temperature and precipitation time series with a high spatio-temporal resolution (one day and a few km<sup>2</sup>). ANATEM was tested on the air temperature and precipitation time series of 22 watersheds situated in the Durance River basin, in the French Alps. Based on a multi-criteria and multi-scale diagnosis, the results show that ANATEM improves the performance of classical statistical models. The ANATEM model has been validated for the regional scale, improving the spatial homogeneity of performances on independent long-term time series. It was able to capture regional low-frequency variability over more than a century (1883–2010).

## 1 Introduction

Multi-decadal variations of climate variables, intrinsically arising from the chaotic and non-linear nature the climate system, have long been observed for a number of large as well as local scale climate features (Madden, 1976).

In a non-stationary climate, multi-decadal variations can remain substantially above or below the long-term trend. In climate projections for the coming decades, they often lead to large uncertainties (e.g. Hawkins and Sutton, 2009; Deser et al., 2012). For precipitation or hydrometeorological variables such as streamflow, these uncertainties can even surpass uncertainties due to climate models (e.g. Terray and Boé, 2013; Lafaysse et al., 2014).

Unfortunately, most climate change impact studies still fail to account for them. For example, projected climatic and hydrological scenarios for a given future lead time are classically compared to a so-called reference period (around 30 years of data) expected to be representative of the recent climate context. As shown by Hänggi and Weingartner (2011) with a 200-years runoff time series of the Rhine at Basel, the hydrological reference features are however likely to highly depend on the period used for their estimation. In such a case, the relevance of conclusions and/or adaptation recommendations formulated on the basis of such a study may be questionable. In our opinion, they suffer from a certain lack of large historical perspective, which would at least require characterising the multi-scale variability of climate variables.

Today, characterising the multi-scale variability of climate variables would appear to be important (if not mandatory) in order to put future climate projections into perspective. Numerous studies worldwide have investigated past variability of climate and related variables. In hydrology for instance, the following studies could be considered as representative for France (Renard, 2006), Spain (Lorenzo-Lacruz et al., 2012), Germany (Renner and Bernhofer, 2011), Europe (Stahl et al., 2010), Canada (Zhang et al., 2001), western North America (Rood et al., 2005) or Australia (CSIRO, 2010). They are based on a set of observed time series available for the region of interest. However, given that the density of observations was significantly lower before 1960 (Hannah et al., 2011), most time series usually cover a few decades only, which is obviously not sufficient for a relevant analysis of multi-decadal variations (Mathevet and Garçon, 2010; Hannaford et al., 2013). Long-term historical time series (covering a period longer than 100 years) are of course the ideal material for such an analysis. Such historical series have been used for the Loire river in France (Renard, 2006), the Columbia and Missouri rivers in the USA (Rood et al., 2005), the Murray-Darling

basin in Australia (CSIRO, 2010) and, more recently, for a larger set of French stations by Boé and Habets (2013). Long-term streamflow time series are rare, with a typically very low spatial density. Some could still be recovered from various national and regional archive sources but the recovery process is long and requires demanding digitising and quality check phases. Finally, the temporal homogeneity of data is often questionable (e.g. because of the evolution of measurement practices as shown in Kuentz et al., 2012, 2014, anthropogenic influences, etc.), hindering the use of some series for variability analysis.

Characterising the low-frequency variability of climate and related variables from observations is therefore seldom possible. An alternative is to reconstruct the past temporal variations of the variable of interest. A number of reconstruction approaches have been presented for numerous fields of geosciences. They use environmental markers such as tree rings (Frank and Esper, 2005) or lake sediments (Wilhelm et al., 2013, 2012), narrative evidence of droughts (Pfister et al., 2006) or geochemical tracers in ice cores from glaciers (Jouzel et al., 2007). Simulations provide an efficient way to reconstruct past flow variations. Simulated discharge times series are obtained using a hydrological model forced with past variations of meteorological variables available for the region. When meteorological observations required for such analysis do not cover the whole target period, they can also be reconstructed. A classical reconstruction is obtained using external data (proxy data) from long-term series of observations available from one or several neighbouring stations. The most popular reconstruction approach is based on linear (multiple-)regression models but a variety of other approaches have been proposed, including non-linear multiple regression (e.g. neural networks), kriging methods and copula based methods (Coulibaly and Evora, 2007; Teegavarapu, 2012; Bárdossy and Pegram, 2014).

Local meteorological data can alternatively be reconstructed from past climate variations. The recent release of two major atmospheric reanalyses covering the entire 20th century (from 1871 for the NOAA 20CR, Compo et al., 2011 and from 1900 for the ECMWF ERA-20C, Poli et al., 2013) provides a great opportunity for such a reconstruction. Unfortunately, their spatial and temporal resolutions rarely fit the resolution needs (typically sub-daily time

step, up to 1000 km<sup>2</sup>) of hydrological applications. In such a case, the required local meteorological data can be obtained through downscaling.

This study compares 3 different statistical approaches for the reconstruction of high-resolution precipitation and temperature data. Reconstructions are respectively obtained from observations available at a neighbouring station, from large-scale (meso-scale) atmospheric variables extracted from the 20CR reanalysis and from a combination of both. Although the first two approaches have already been applied in similar studies, the last is an original approach in that makes use of both local observations and large-scale atmospheric information. The principle of reconstructions obtained with the 3 approaches is illustrated in Fig. 1. Reconstructions are built at a daily time step for the 22 subcatchments of the Upper Durance River basin, a mesoscale catchment located in the south-eastern Alps. They have been produced for hydrological reconstructions covering the past 140 years. An exhaustive evaluation of the whole hydrological reconstruction process can be found in Kuentz (2013).

The Upper Durance River basin as well as the meteorological and atmospheric data are presented in Sect. 2. The three reconstruction models are presented in Sect. 3 and evaluated and compared in Sect. 4. Section 5 briefly discusses the low-frequency climatic variability reconstructed over the 1870–2010 period. Finally, conclusions and perspectives emanating from this work are given in Sect. 6.

## 2 Data

### 2.1 Case study location and spatial climatic inputs

The three methods have been applied for the reconstruction of mean areal temperature and precipitations of 22 sub-basins of the Durance River basin, a mesoscale alpine watershed located in south-eastern France (Fig. 2). The main characteristics of the watersheds are detailed in Table 1.

Bounded in the north by the Écrins mountain range of the Alps and in the south by the Mediterranean Sea, the various subcatchments display very different climates. Upstream

hydrological regimes are dominated by snow with high snowmelt flows in late spring and early summer. When moving downstream, they become more Mediterranean with additional autumn floods due to large rainfall amounts in that period.

For each watershed, daily mean areal air temperature and precipitation data over the 1948–2010 period have been taken from the SPAZM meteorological analysis produced by Gottardi et al. (2012). In the following, the 1948–2010 period will be referred to as the “observation period” and the SPAZM series will be referred to as “observations” even though they are not direct recordings, but rather mean areal air temperature and precipitation series aggregated at the watershed scale from local observations of temperature and precipitation (Gottardi et al., 2012).

## 2.2 Long local reference series

To reconstruct the mean areal air temperatures and precipitation of the 22 watersheds, it was first necessary to search for the longest observed series on or near the Durance watershed. In a technical report published in 1892, Imbeaux (1892) reported 4 air temperature and 40 precipitation measurement stations in the watershed and its neighbourhood. Unfortunately, most of the data from these stations have been lost and only very few and incomplete series remain available today. For precipitation, it was possible to rebuild a 1883–2010 series for the Gap location by merging two sources of data, provided respectively by Électricité de France (EDF) and Météo-France. For air temperature, the nearest daily series found, provided by Météo-France, was for Marseille and covers the period 1868–2010.

For a qualitative assessment of the reconstructed series, 5 monthly time series from the HISTALP project database (monthly series, Auer et al., 2007) were also used. They also go back to the 1870’s. For air temperature, the selected stations are located around the south-eastern part of the Alps, at Genova University, Milano-Brera, Montpellier, Nice airport and Nîmes airport. For precipitation, they are closer to the Durance watershed, located in the cities of Aix-en-Provence, Nice (Cap-Ferrat), Orange, Saint-Paul-les-Durance and Toulon.

## 2.3 Large-scale climatic data

Large-scale atmospheric data (describing meso-scale circulation) were extracted from the “20th Century Reanalysis” (“20CR”, Compo et al., 2011) from the project of the same name, supported by the US Department of Energy and the Climate Program Office of the National Oceanic and Atmospheric Administration (NOAA). This reanalysis was produced by assimilating only sea level pressure data, making it possible to go back to the end of the 19th century. The reanalysis covers the 1871–2010 period. In the present work, the large-scale variables used for the reconstruction are the fields of 700 and 1000 hPa geopotential heights in the rectangular spatial domain situated between longitudes  $8^{\circ}$  W and  $12^{\circ}$  E and latitudes  $38^{\circ}$  N and  $50^{\circ}$  N.

## 3 Methodology: combining two sources of information

In climatology and hydrology, the reconstruction of past climatic data is usually necessary to estimate missing values, assess data quality or build long-term climatic reanalyses. Different methods are classically used to reconstruct climatic observations. Some of them are solely based on the series being reconstructed (long-term average or regime methods, temporal interpolation techniques...), others are based on external data (proxy data) used to calibrate and run a reconstruction model. For climatic reconstructions, proxy data could be either observations of the same variable as the one to be reconstructed or observations of different variables assumed to be linked to it.

In the following section, the three methods used for the reconstruction are presented. The first one uses local neighbour observations of a similar proxy (respectively, air temperature or precipitation observation). The second is basically a downscaling approach using regional large-scale information of a different proxy (geopotential fields). The third approach uses both proxies.

As in most reconstruction works, these methods rely on a period over which both data at the reconstruction point and proxy information are available (see Fig. 1). This period will

be referred to as the observation period. The reconstruction period is the period over which the reconstruction model is applied, corresponding to the period where proxy information is available but data is missing at the reconstruction point (in the following, the reconstruction are also presented for the observation period).

### 3.1 Local information

A classical method used for climatic reconstructions is based on regression-like models, where predictors should be well correlated with the data to be reconstructed. This model is calibrated against observations during the observation period.

In the following, the principle of the local model (LM) is to reconstruct the target series (referred to as  $T_g$ ) from a local neighbour series (referred to as  $T_{Ne}$ ) using a classical linear regression model.

#### 3.1.1 Air temperature reconstruction

For air temperature reconstruction, the LM model classically uses an additive correction, assumed to be constant over time and mainly influenced by the altitude difference between the target and neighbour series. However, even when the target and neighbour series are very well correlated, residuals of such a model usually exhibit a strong seasonal pattern. In this case, the LM model can be slightly improved by applying an additive correction that varies over time. In the present case, it is represented by a daily harmonic function, calibrated on the interannual mean monthly residuals of the differences between the target series and the neighbour series.

The local model for air temperature reconstruction can thus be written as:

$$\text{LM} : \hat{T}_{\text{LM}}(d) = T_{\text{Ne}}(d) + \beta(d) + \varepsilon(d) \quad (1)$$

where  $\hat{T}_{\text{LM}}(d)$  is the estimate of the target air temperature for day  $d$ ,  $T_{\text{Ne}}(d)$  is the value of the neighbour series temperature for the same day,  $\beta(d)$  is the correction, depending on the calendar day of the year, and  $\varepsilon(d)$  is a residual assumed to have zero mean.



In the present study this model has been used in a deterministic way, that is without considering the residual term. Uncertainty is accounted for in the mixed model as explained in Sect. 3.3.

### 3.1.2 Precipitation reconstruction

For precipitation reconstruction, the LM model classically uses a multiplicative correction, assumed to be constant over time. This multiplicative correction is compatible with the asymmetrical distribution of precipitation values (never negative). The correction factor is taken to be constant throughout the year. The improvement obtained by using a variable correction has been assessed and shown to be negligible (Kuentz, 2013). The constant multiplicative correction factor is calibrated over the observation period as:

$$\alpha = \frac{\overline{P}_{Tg}}{\overline{P}_{Ne}} \quad (2)$$

where  $\overline{P}_{Ne}$  is the mean value of the neighbour series and  $\overline{P}_{Tg}$  is the mean value of the target series, both calculated over the observation period.

The local model used for precipitation reconstruction reads:

$$LM : \widehat{P}_{LM}(d) = \alpha \cdot P_{Ne}(d) + \varepsilon(d) \quad (3)$$

where  $\widehat{P}_{LM}(d)$  is the estimate of the target precipitation for day  $d$ ,  $P_{Ne}(d)$  is the value of the neighbour series precipitation for this same day and  $\varepsilon(d)$  is a residual with zero mean. Once again, a simple version of this model with a residual term equal to zero is used in the present work.

### 3.2 Large-scale climatic information: the analogue method

The second reconstruction model is based on the analogue method introduced by Lorenz (1969). Currently, this method is largely used to produce meteorological scenarios in the

context of weather forecasting (Van Den Dool, 1989; Horton et al., 2012) or climate projections (Teng et al., 2012; Bourqui et al., 2011; Hingray et al., 2013). The method is seldom applied to the reconstruction of climatic series as done by Timbal et al. (2006). Nevertheless, the release of the long atmospheric reanalyses for the 20th century opens the door to more such uses, allowing the reconstruction of long climatic series covering the entire 20th century.

The analogue method is based on the fact that local meteorological variables are strongly influenced by the state of the atmosphere and its meso-scale circulation. Provided that a sufficiently long archive with concomitant local and large-scale observations is available, it is therefore possible to produce local meteorological scenarios for any other day for which the required large-scale atmospheric predictors are available. For this, the  $n$  days that are the most similar to the target day in terms of atmospheric circulation are first identified in the archive. The surface meteorological variables observed for one of those analogue days are then used as the weather scenario for the target day.

In the present case, the archive is the SPAZM meteorological analysis (Gottardi et al., 2012) covering the 1948–2010 observation period. As large-scale atmospheric predictors are available for each day of the 1883–2010 period covered by the 20CR atmospheric reanalysis, the method allows the reconstruction of a 127-year time series of daily and local meteorological variables.

The analogue method has some parameters to be set such as the type and level of predictors, the number of analogue days selected for the prediction, the spatial domain used to compute the similarity criterion or the similarity criterion itself. Numerous variations of the analogue method have been developed. In the present work, the analogue model (ANA) presented by Obled et al. (2002) and further explored by Bontron and Obled (2005), Ben Daoud et al. (2010), Horton et al. (2012) and Chardon et al. (2014) is used. Its main features are presented below.

- *The predictors* are the 700 and 1000 hPa geopotential height fields at times 00:00 and 24:00. For the spatial domain of the present study, these geopotential fields were found to be the most informative predictors by Bontron (2004).

- *The similarity criterion* is proposed by Teweles and Wobus (1954). This score is based on the shape of the geopotential fields and has been shown to perform better than a classical Euclidean distance for this type of use (e.g. Wetterhall et al., 2005).
- *The spatial domain* used to estimate the similarity includes all grid points between longitudes 8° W and 12° E and latitudes 38 and 50° N, with a step of 2°.
- *A moving calendar filter* is used for the determination of candidate analogue days: for each target day, candidate analogue days are the days included in a 60-day interval around the calendar day of the target.

The reconstruction is deterministic when only one analogue is used (classically the nearest analogue). The analogue day can be also selected among the  $n$  nearest analogues. An ensemble of reconstructions can be produced when all  $n$  nearest analogues are successively used for the reconstruction. In the following, the ensemble is simply defined with the empirical distribution of the  $n$  observations from the  $n$  nearest analogues respectively. This ensemble of reconstructions makes it possible to evaluate the uncertainty in the reconstruction. The ensemble of reconstructions obtained with the ANA model for the variable  $X$  and day  $d$  will be written as  $[X(d_k)]_{k=1\dots n}$  where  $[d_k]_{k=1\dots n}$  refers to the  $n$  nearest analogue days selected for day  $d$ . In the present case, the ensemble of reconstructions is obtained from the 50 nearest analogues ( $n = 50$ ).

### 3.3 Mixing formulation: the ANATEM model

Both local and large-scale predictors are available for the 1870–2010 period. The local model (LM) and the analogue model (ANA) can therefore be used to produce two different reconstructions of precipitation or air temperature for this period, one based on local observed data (another station with available data), the other from large-scale atmospheric information (meso-scale variables). The originality and strength of the ANATEM model introduced here lies in an approach that combines the two previous models. In this way, it can take advantage of both local and large scale information and produce an original representation of uncertainties, conditioned by atmospheric circulation patterns.

The principle of ANATEM is the following: for any target day, the analogue model allows the identification of  $n$  analogue days in terms of atmospheric circulation (see Section 3.2). The local model is then used to obtain an estimate of the variable to be reconstructed (precipitation or air temperature at the target site) for each of the selected analogue days. These  $n$  estimates are respectively compared with the corresponding observed values for these  $n$  days, allowing the calculation of  $n$  prediction errors. These  $n$  error values are finally used to define the error distribution associated with the prediction obtained with the local model for the target day  $d$ . The prediction obtained with ANATEM for the target day is therefore probabilistic.

### 3.3.1 Air temperature reconstruction

The probabilistic air temperature prediction from the ANATEM model for day  $d$  has the following expression:

$$\left[ \widehat{T}_{\text{ANATEM}}^k(d) \right]_{k=1\dots n} = \widehat{T}_{\text{LM}}(d) + \left[ T(d_k) - \widehat{T}_{\text{LM}}(d_k) \right]_{k=1\dots n} \quad (4)$$

where  $\left[ \widehat{T}_{\text{ANATEM}}^k(d) \right]_{k=1\dots n}$  is the ensemble of reconstructed values for the target day  $d$ ,  $\widehat{T}_{\text{LM}}(d)$  the air temperature estimate obtained with LM for target day  $d$ ,  $d_k$  the  $k^{\text{th}}$  analogue day selected for target day  $d$ ,  $T(d_k)$  the observed air temperature for this  $k^{\text{th}}$  analogue day and  $\widehat{T}_{\text{LM}}(d_k)$  the air temperature estimate obtained with the local model (LM) for the same day  $d_k$ .

In this expression,  $\left[ T(d_k) - \widehat{T}_{\text{LM}}(d_k) \right]$  is the error obtained with the LM model when it is applied to estimate the temperature of the  $k^{\text{th}}$  analogue day  $d_k$ .

The statistical dressing of the LM prediction for the target day  $d$  can be simply represented on a graph in a  $(T_{\text{LM}}, T)$  space, as shown in Fig. 3 (right). In this figure, the green point is the value obtained for the target day with the LM model. The different blue crosses in the  $y$  direction around this estimate define the distribution of the  $n$  errors obtained with the LM model respectively applied to the  $n$  analogue days. Each cross is simply the intercept of

two lines: the vertical line at the  $\widehat{T}_{LM}(d)$  value on the x axis and the 1:1 line passing through the point  $(\widehat{T}_{LM}(d_k), T(d_k))$ . This is illustrated for a given analogue day in Fig. 3 (left).

For the example shown in Fig. 3, over the  $n$  analogue days with meso-scale situations similar to that of day  $d$ , the local model estimate  $T_{LM}$  was on the average higher than the observed temperature at the target point. Applying this error distribution to the reconstructed day  $d$  leads to a negative correction on most of the ensemble. While the value of the local model was  $-9.8^{\circ}\text{C}$ , the 50 air temperature values produced by the ANATEM model have a mean of  $-11.2^{\circ}\text{C}$  and their 10 and 90% quantiles are respectively  $-13.1$  and  $-9.3^{\circ}\text{C}$ .

### 3.3.2 Precipitation reconstruction

Although the ANATEM model uses the same basic principle for precipitation reconstruction, a somewhat different formulation is proposed to account for the specific features of precipitation (asymmetric distribution and many zero values).

The additive correction formulation used for the probabilistic reconstruction of temperature (Eq. 4) is not suitable here. It can actually produce negative values as illustrated in Fig. 4 (left), elaborated following the same principle as explained in Fig. 3 (right).

An alternative formulation uses a multiplicative correction for each analogue date. The probabilistic reconstruction is here defined by the following expression:

$$\left[ \widehat{P}_{MULT}^k(d) \right]_{k=1\dots n} = \widehat{P}_{LM}(d) \cdot \left[ \frac{P(d_k)}{\widehat{P}_{LM}(d_k)} \right]_{k=1\dots n}. \quad (5)$$

The multiplicative formulation obviously avoids the estimation of negative precipitation values. A graphical representation of this reconstruction strategy is given in Fig. 4 (right). As illustrated, the reconstructed values appear to be reasonable for common values, but can be unreasonably high in certain cases.

In the following, the probabilistic reconstruction of precipitation has therefore been built with a correction model intended to have a multiplicative behaviour for low values of  $\widehat{P}_{LM}(d)$  and an additive behaviour for high values of  $\widehat{P}_{LM}(d)$ . Its analytical formulation and its asymp-

otic behaviour when  $x$  tends to zero or infinity (through a Taylor expansion) have the following expressions:

$$\widehat{P}_{\text{ANATEM}}^k(d) = f(x(d)) = \frac{x(d)^2 + a(d_k) \cdot x(d)}{x(d) + b(d_k)} \quad (6)$$

where  $x(d) = \widehat{P}_{LM}(d)$  and  $a(d_k)$  and  $b(d_k)$  are parameters to be expressed as a function of  $P(d_k)$  and  $\widehat{P}_{LM}(d_k)$ . In what follows, for the sake of simplicity, the day indexes will be omitted from  $a(d_k)$ ,  $b(d_k)$  and  $x(d)$ .

$$\widehat{P}_{\text{ANATEM}}^k(d) = \frac{x^2 + a \cdot x}{x + b} = x \cdot \frac{x + a}{x + b} \underset{x \rightarrow 0}{\sim} x \cdot \frac{a}{b} \quad (7)$$

$$\widehat{P}_{\text{ANATEM}}^k(d) = \frac{x^2 + a \cdot x}{x + b} = x \cdot \left(1 + \frac{a}{x}\right) \cdot \left(1 + \frac{b}{x}\right)^{-1} \underset{x \rightarrow +\infty}{\sim} x + (a - b) \quad (8)$$

The two model parameters  $a$  and  $b$  are defined as in the work of Dufour and Garçon (1997) for the assimilation of streamflow data in a hydrological model. The parameters are defined as a function of  $P(d_k)$  and  $\widehat{P}_{LM}(d_k)$  in order to reach a compromise between a good multiplicative behaviour for low values and a good additive behaviour for high values.

Two conditions have been set to define the parameters:

- The slope of the tangent to the curve at  $x = 0$  must be  $\left(\frac{P(d_k)}{\widehat{P}_{LM}(d_k)}\right)^2$  ;
- When  $P(d_k) = \widehat{P}_{LM}(d_k)$ , the equality  $\widehat{P}_{\text{ANATEM}}^k(d) = \widehat{P}_{LM}(d)$  must be obtained.

These two conditions lead to the following expressions for the parameters:

$$a = P(d_k) \text{ and } b = \frac{\left(\widehat{P}_{LM}(d_k)\right)^2}{P(d_k)}. \quad (9)$$

More detailed calculations of the asymptotic behaviour when  $x$  tends to zero or infinity and the use of the two conditions are provided in the supplementary material published along with this paper.

The probabilistic reconstruction obtained with ANATEM for precipitation finally reads:

$$\left[\widehat{P}_{\text{ANATEM}}^k(d)\right]_{k=1\dots n} = \left[\frac{\widehat{P}_{LM}(d)^2 + P(d_k) \cdot \widehat{P}_{LM}(d)}{\widehat{P}_{LM}(d) + \frac{\widehat{P}_{LM}(d_k)^2}{P(d_k)}}\right]_{k=1\dots n}. \quad (10)$$

The graphical representation of this formulation is shown in Fig. 5. The graph on the left side shows the curve corresponding to Eq. (10) applied for one analogue day  $d_k$  and the graph on the right side shows the ensemble of curves associated respectively with the  $n$  analogue days. The distribution of the reconstructed values  $[\widehat{P}_{\text{ANATEM}}^k(d)]_{k=1\dots n}$  is represented by the boxplot.

In the case of very different values of  $P(d_k)$  and  $\widehat{P}_{LM}(d_k)$ , Eq. (10) can potentially produce unreasonably high values of corrected precipitation  $\widehat{P}_{\text{ANATEM}}^k(d)$ . In order to avoid such values the following filters have been applied:

- if  $P(d_k) > 10 \cdot \widehat{P}_{LM}(d_k)$  then the value of  $P(d_k) = 10 \cdot \widehat{P}_{LM}(d_k)$ , and
- if  $P(d_k) < \frac{1}{10} \cdot \widehat{P}_{LM}(d_k)$  then the value of  $P(d_k) = \frac{1}{10} \cdot \widehat{P}_{LM}(d_k)$ .

The filtering threshold (10) has been chosen arbitrarily. A sensitivity analysis with different values from 2 to 100 showed that this threshold has little impact on reconstruction performance. This is true because very few analogue days are generally affected by this filtering operation. The filters are represented by blue zones in Fig. 5.

For the example day shown in Fig. 5, the local model LM gives a reconstructed value of 15.0 mm. The mean and the 10 and 90% percentiles of the probabilistic reconstruction obtained with ANATEM are respectively 14.8, 7.8 and 21.0 mm.

## 4 Analysis of ANA, LM and ANATEM performance

### 4.1 Evaluation process

The data presented in Sect. 2 can be used to reconstruct the daily air temperature and precipitation series for the 22 selected watersheds over the period 1883–2010. The reconstruction is deterministic for the LM model. For ANATEM and ANA, 50 reconstructed time series are generated. In the present section, the three reconstruction models are evaluated based on their reconstruction skill for the 1948–2010 observation period.

The evaluation is based on three criteria. The ratio  $\beta$  between the mean estimated value and mean observed value of the variable evaluates the bias of the reconstruction. The ratio between the standard deviations of the reconstructed and observed time series ( $\alpha$ ) evaluates the ability of the reconstruction to reproduce the observed variability of the variable. The coefficient of correlation  $r$  between the observed and reconstructed series additionally measures the ability of the reconstruction to reproduce the observed temporal variations (e.g. alternating dry/wet or warm/cold periods). The overall performance obtained for these three criteria is summarised by the Kling–Gupta Efficiency criterion (KGE; Gupta et al., 2009) defined as follows:

$$\text{KGE} = 1 - \sqrt{(1 - r)^2 + (1 - \alpha)^2 + (1 - \beta)^2}. \quad (11)$$

The ability of the reconstruction to reproduce the variability and variations of observations was evaluated for multiple temporal resolutions: daily (high-frequency variability), monthly (accounting thus for the infra-annual variability) and annual (low-frequency variability) resolutions. For the annual resolution, the series are aggregated by hydrological year, i.e. from 1 October to 30 September.

In the following sections, the performance of the three models for an illustrative watershed (Ubaye River at Barcelonnette) is first presented. The evaluation relies (1) on the graphical comparison of the observed and reconstructed annual series for the 1948–2010 period and (2) on the distributions obtained for  $r$ ,  $\alpha$ ,  $\beta$  and KGE when estimated for the daily,



monthly and annual time step from the 50 ensembles. Then, the results obtained for the 22 watersheds of the Durance basin are presented.

## 4.2 Performance for the Ubaye River at Barcelonnette watershed

### 4.2.1 Air temperature reconstruction

Figure 6 presents the mean annual time series of mean air temperature for the watershed. These figures first show that the observed temperature has increased over the last 60 years, with a mean value of around 3°C in the 50's and a mean value of around 4°C nowadays. The ANA model does not capture the temporal evolution and the variability of air temperature, as opposed to the LM and ANATEM models. Note also that the spread of the ANA ensembles is much higher than that of the ANATEM ensembles. These observations are consistent with the distributions obtained for the different criteria at the annual time step (Fig. 7, right):

- ANA shows a limited mean bias ( $\beta$  close to 1), but a rather poor temporal correlation and significant bias of variability, which is exhibited by relatively low mean values of  $r$ ,  $\alpha$  and KGE (between 0.2 and 0.6, not visible in the figure);
- LM and ANATEM show very good temporal correlations (mean  $r$  greater than 0.9) and limited mean and variability bias. The two models show slightly different skills: LM shows no mean bias (by construction) but a significant variability bias (mean  $\alpha$  less than 0.9) whereas ANATEM shows a limited mean and variability bias (mean  $\beta$  and mean  $\alpha$  greater than 0.95).

Figure 7 also shows the distributions of the criteria for daily and monthly time steps. The hierarchy of the performance of the three models is the same for both time steps, with KGE values ranging from 0.77 to 0.87 for ANA, 0.95 to 0.98 for LM and 0.93 to 0.97 for ANATEM. Moreover, the mean criteria are higher at a monthly time step than at a daily time step, while the annual time step performance is the lowest. This means that these models have greater difficulty in reproducing inter-annual or daily variability than intra-annual variability (this is partly due to the seasonality of air temperature). The LM model performed slightly

better than ANATEM at daily and monthly time steps. Conversely, ANATEM performs better than LM at annual time steps.

#### 4.2.2 Precipitation reconstruction

Figure 8 presents the observed and reconstructed annual time series of mean precipitation on the watershed. As shown, observed annual precipitation presents a high variability, ranging from  $1000 \text{ mm yr}^{-1}$  over certain periods to  $1500$  to  $2000 \text{ mm yr}^{-1}$  for some exceptional years. All three models capture relatively well this variability and are able to reproduce wet and dry periods. The spread of the ANA ensembles is much greater than that of the ANATEM ensembles.

The distributions of criteria at the annual time step (Fig. 9, right part) confirm these statements:

- ANA shows a moderate correlation (mean  $r$  close to 0.5), while LM and ANATEM show a rather good correlation (mean  $r$  greater than 0.8);
- LM shows no mean bias (by construction), while ANA and ANATEM show a moderate mean bias (less than 0.05);
- ANA shows a noticeable variability bias (up to 0.15), while LM and ANATEM show a limited variability bias (around 0.03).

The hierarchy of the performance of the three models is the same for both daily and monthly time steps, with KGE values ranging from 0.35 to 0.7 for ANA, 0.78 to 0.88 for LM and 0.73 to 0.85 for ANATEM (Fig. 9). ANA performance is clearly poor at a daily time step, with a very limited correlation ( $r$  less than 0.4). The mean criteria are higher at a monthly time step and similar at daily and annual time steps. As for air temperature, this highlights the difficulty the models have to reproduce the low- and high-frequency variability while the intra-annual variability is well-captured.

### 4.3 Performance for all 22 watersheds

For the sake of readability, only one time series is considered for each model. ANA and ANATEM probabilistic reconstructions are represented by the mean time series derived from the ensemble (the daily reconstructed value for a given day is the mean of the 50 probabilistic reconstructions for this day). For the sake of simplicity, these mean time series will be referred to as the reconstructed time series in the following. As will be illustrated later, note that these ensemble mean time series logically present a much lower temporal variability than each individual component of the reconstruction ensemble. In the following, the performance of a given model will be presented with the distributions of  $r$ ,  $\alpha$ ,  $\beta$  and KGE criteria obtained for the 22 watersheds at the daily, monthly and annual time steps.

#### 4.3.1 Air temperature reconstruction

The main results obtained for air temperature reconstruction are (Fig. 10):

- At daily and monthly time steps, the ANA model suffers from a limited positive mean bias (mean  $\beta$  around 1.03) and a significant negative variability bias (mean  $\alpha$  from 0.85 to 0.88). Correlation with observations is very good (mean  $r$  greater than 0.93). At the annual time step, ANA fails to capture the low-frequency variability and trend, with a very low correlation (mean  $r$  close to 0.53, not shown in the figure) and a very strong negative variability bias (mean  $\alpha$  close to 0.42, not shown in the figure). At different time steps and for different criteria, ANA exhibits a rather good spatial robustness of performance (i.e. homogeneity of the results at a regional scale, which could be related to a rather limited spread of the distribution compared to LM and ANATEM models, as shown by the distance between quantiles 0.1 and 0.9).
- At all the different time steps, the LM model provides very satisfactory results. It shows no mean bias (by construction) and a moderate to limited variability bias (mean  $\alpha$  between 0.91 and 0.99). The high- to low-frequency variability is very well captured

(mean  $r$  between 0.92 and 0.99). LM shows moderate spatial robustness for correlation and variability bias, for daily and annual time steps.

- At all the different time steps, the ANATEM model provides very satisfactory results. It shows a moderate mean negative bias (mean  $\beta$  close to 0.97) and a limited to moderate variability bias (mean  $\alpha$  between 0.95 and 0.98). The high-frequency to low-frequency variability is very well captured (mean  $r$  between 0.94 and 0.99). ANATEM exhibits moderate robustness concerning mean bias and concerning correlation and variability bias, for daily and annual time steps.

The LM and ANATEM models clearly outperform the ANA model. LM is characterised by a very good correlation and no mean bias, but a moderate variability bias. ANATEM is characterised by a very good correlation, and limited mean and variability bias. Model performance is better and more robust at a monthly time step, compared to daily and annual time steps. ANATEM exhibits a slightly better spatial robustness of performance than LM. This is also expressed by mean KGE values, ranging from 0.25 to 0.87 for ANA, 0.88 to 0.99 for LM and 0.92 to 0.97 for ANATEM respectively.

### 4.3.2 Precipitation reconstruction

The three models present slightly different results for precipitation reconstruction (Fig. 11):

- At a daily time step, ANA suffers from a very moderate mean negative mean bias (mean  $\beta$  close to 0.95) and a strong variability bias (mean  $\alpha$  around 0.55). It also shows a limited correlation (mean  $r$  close to 0.65). At monthly and annual time steps, ANA shows a moderate to limited mean bias (mean  $\beta$  close to 0.95), a significant variability bias (mean  $\alpha$  around 0.8) and an acceptable level of correlation (mean  $r$  between 0.7 to 0.8).
- At all the different time steps, the LM model shows very satisfactory results. It shows no mean bias (by construction) and a limited variability bias (mean  $\alpha$  from 0.97 to

1.05). The high-frequency to low-frequency variability is well captured (mean  $r$  between 0.77 and 0.84).

- At all the different time steps, the ANATEM model shows very satisfactory results. It shows a limited negative mean bias (mean  $\beta$  around 0.96) and a limited variability bias (mean  $\alpha$  from 0.94 to 1.02). The high-frequency to low-frequency variability is well captured (mean  $r$  between 0.75 and 0.87).

The LM and ANATEM models thus perform better than the ANA model, particularly in terms of correlation. LM is characterised by a good correlation, no mean bias and a limited variability bias. ANATEM is also characterised by a good correlation, limited mean and variability bias. Model performance is better and more robust at a monthly time step compared to daily and annual time steps. The spatial robustness of performance is slightly lower for the variability criterion than for the other criteria. LM shows the lowest spatial robustness, then ANATEM and finally ANA. This is again illustrated by the mean KGE values ranging from 0.43 to 0.71 for ANA, from 0.75 to 0.83 for LM and from 0.76 to 0.84 for ANATEM.

## 4.4 Spatial patterns of model performance

In the present section, the spatial patterns of performance (in terms of correlation, at a daily time step) of the three models and the spatial patterns of the gain in performance obtained with ANATEM reconstructions when either compared to ANA or LM alternatives will be discussed.

### 4.4.1 Air temperature reconstruction

For temperature reconstructions, the spatial patterns of model performance are presented in Fig. 12. For ANA, the performance of the reconstruction appears to be mostly independent of the location of the watershed, with a mean correlation ranging from 0.92 to 0.94 (Fig. 12a). For LM (Fig. 12b), the location of the watershed had a slightly higher influence on the performance, with a mean correlation ranging from 0.95 to 0.98. This spatial pattern

has a clear south-west to north-east structure, with a decrease in model performance driven by the distance from the local reference time series (located in Marseille, south-west of the watersheds). Finally, for the ANATEM model (Fig. 12c), the location of the watershed (i.e. distance from Marseille) also influences the performance of the reconstruction, with a mean correlation ranging from 0.97 to 0.99. However, ANATEM shows slightly better performance than LM and ANA and the range of correlation values is slightly smaller than for LM.

The contribution of the LM (resp. ANA) model to the performance of the ANATEM model is presented in Fig. 14a (resp. 14b). It is estimated by the difference between the performance of the ANATEM and ANA (resp. LM) models. The contribution of the LM model is much higher than that of ANA, whatever the location, meaning that most of the information provided by ANATEM comes from LM. Note however that for both the LM and ANA models, the contribution of the model presents a clear south-west to north-east gradient, which decreases for LM (from 0.06 to 0.04) and increases for ANA (from 0.0 to 0.02). The contribution of large-scale information (through the ANA model) is stronger when the LM model (local information) is less effective, that is, when the location to be reconstructed is far from the reference temperature station.

#### 4.4.2 Precipitation reconstruction

The spatial patterns of model performance obtained for precipitation are slightly different than those obtained for temperature (Fig. 13). For ANA, the location of the watershed does not appear to really influence the performance, with a mean correlation ranging from 0.62 to 0.68 (Fig. 13a). On the other hand, for LM (Fig. 13b) and ANATEM (Fig. 13c), watersheds close to the local reference station (Gap) show better performance than the others (the correlation ranges from 0.62 to 0.88 for LM and from 0.69 to 0.89 for ANATEM). However, the distance from the local reference station is probably not the only factor influencing performance, as two watersheds at the same distance from the Gap station displayed somewhat different performance (i.e. the reconstructions for the Buech watershed – #10 in Fig. 2 – have a very good correlation of 0.88 and the reconstructions for the Durance at Briançon watershed – #3 in Fig. 2 – a moderate correlation of 0.77). This could be due to large-scale

climatic influences that give some watersheds a higher proximity to Gap in terms of the precipitation pattern.

ANATEM slightly increases the overall reconstruction performance but at the same time notably smoothes local contrasts. The contribution of LM to the performance of ANATEM is generally higher than that of ANA, but decreases as the distance from Gap increases, ranging from 0.22 to 0.02 (Fig. 14c). On the other hand, the contribution of ANA to the performance of ANATEM is close to 0 for the stations closest to Gap and slightly increases (up to 0.07) with the distance from Gap (Fig. 14d). As observed for air temperature reconstruction and here in a more pronounced way, the contribution of large-scale information (through the ANA model) is stronger when the LM model (local information) is less effective, as a result of the increasing distance from the reference station.

## 5 Climatic variability assessment

### 5.1 1883–2010 reconstructions of air temperature

Figure 15 presents the 1883–2010 annual time series of air temperature anomalies reconstructed by the ANATEM model (mean model) for the 22 watersheds of the Durance river. Anomalies have been computed as differences with respect the 1883–2010 mean. This figure exhibits a pseudo-stationary period from 1880 to 1940, then a slight temperature increase between 1940 and 1980 and a stronger increase from 1980 until the present. In order to better characterise low-frequency variability, the mean of all 22 reconstructed series was computed and smoothed using a LOESS low-pass filter (Cleveland, 1979, smoothing parameter value used: 0.15).

The ANATEM reconstructions have been qualitatively compared to five series of air temperature anomalies obtained from homogenised series of the HISTALP project (Genova University, Milano-Brera, Montpellier, Nice airport and Nîmes airport).

The ANATEM model reproduces fairly well the annual and low-frequency variability of air temperature anomalies from the HISTALP stations (the mean correlation between ANATEM

and HISTALP annual series is close to 0.8). However, the warming trend in the HISTALP series is stronger than in the ANATEM reconstructions, HISTALP temperatures being significantly lower than ANATEM temperatures before 1900 and significantly higher after 1980. ANATEM reconstructions and HISTALP time series are obviously sensitive to the reference time series (i.e. Marseille for ANATEM) and the homogenisation process applied to the observations (for both Marseille and HISTALP stations). Further research is required to explore the sensitivity of the ANATEM reconstructions to these key features (partly tested in Kuentz, 2013).

## 5.2 1883–2010 reconstructions of precipitation

Figure 16 presents the 1883–2010 annual multiplicative time series of precipitation anomalies reconstructed with ANATEM for the 22 watersheds along with five precipitation HISTALP series (Aix-en-Provence, Nice (Cap-Ferrat), Orange, Saint-Paul-les-Durance and Toulon). For both the reconstructions and the HISTALP series, the mean smoothed series are also given.

The dispersion between the 22 ANATEM reconstructed time series is relatively low. It is actually similar to the dispersion obtained between the time series of observations available for the same 22 watersheds over the 1960–2010 period (not shown here). The dispersion observed between the five HISTALP series is comparatively much higher. This may be partly explained by the fact that the ANATEM series are reconstructed for all watersheds based on a same reference series (Gap). The main reason is however probably that the HISTALP series cover a much wider spatial domain with a high spatial variability of atmospheric influences and thus precipitation regimes and times series.

The smoothed time series from ANATEM reconstructions is highly correlated with the smoothed time series from HISTALP data. The ANATEM reconstruction is therefore able to reproduce the low-frequency variability of precipitation resulting from climate variability. Some differences can be observed, for example between 1920 and 1930 or between 1970 and 1980. They may be due again to the large spatial variability of precipitation which would also correspond to different precipitation indexes, as long as they are estimated from differ-



ent stations. As already noted for air temperature reconstructions, these differences could also be due to the reference series used in ANATEM and to the homogenisation process for the HISTALP series. Additional work should be considered to explore the importance of such issues.

## 6 Conclusions

Reconstructing local scale meteorological variables over long periods is a challenging but necessary task in order to obtain a better understanding of the low-frequency variability of regional climate and climate driven variables. Three models are compared in the present work, using different types of data for the reconstruction: the regression-based local model (LM) uses local observations of the variable from neighbouring stations as a predictor; the analogue model (ANA), a so-called downscaling model, uses large-scale information concerning atmospheric circulation and the ANATEM model uses a mix of both local and large-scale atmospheric information by combining the local and analogue models.

The three models have been developed and applied to the reconstruction of mean air temperature and precipitation time series for a sample of 22 watersheds situated in the Durance basin, in the south-east of France. This sample of watersheds represents a wide range of climatic conditions, from highly mountainous to Mediterranean. The local observations used for the reconstruction are respectively Marseille air temperature, Gap precipitation historical time series and geopotential height fields from the 20CR reanalysis.

The multi-criteria and multi-scale performance assessment highlights that the best reconstructions are obtained when local information is used. The ANA model is clearly less efficient than the two other models, particularly concerning low-frequency (annual) air temperature variability or high-frequency (daily) precipitation variability. On the other hand, the regression-based model and the ANATEM model provide very satisfactory results for all criteria. ANATEM offers a slight advantage and the spatial patterns of the reconstruction skills show that it benefits from the qualities of both underlying models. Hence, the ANATEM model can be used to reconstruct adequate air temperature and precipitation series at

a high temporal resolution (daily) and different spatial scales (from 4 to 3500 km<sup>2</sup>), while improving the spatial robustness of performance. Besides these results in terms of performances, the ANATEM model provides an original representation of uncertainties, which are conditioned by atmospheric circulation patterns through the use of an ensemble of analogue days.

Time series of air temperatures reconstructed for the 1883–2010 period exhibit the well-known warming experienced since the middle of last century, with a higher rate since the 1980's. Reconstructed precipitation time series highlight the large inter-annual variability of annual precipitation for the Durance region. Long-term climatic reanalyses exhibit some particular periods with rather strong rainfall anomalies, such as the wet periods at the beginning of the 1910's and mid-1930's (known for flood events), or relatively dry periods such as the 1940's and 1970's (known for drought events).

The potential for improving the method is considerable. The ANA method used here was first developed for precipitation forecasts (e.g. Obled et al., 2002). The poor reconstruction skill obtained for temperature was therefore not a surprise and other large-scale predictors could potentially provide a better reconstruction of air temperature variations. This also applies to the precipitation reconstruction. The predictors used here provide information only on the atmosphere dynamics. The inclusion of thermodynamic predictors and humidity predictors for the identification of analogue days has been proved to improve the performance of the method for the studied region (Marty et al., 2011; Chardon et al., 2014).

Another possibility for improvement concerns the combined formulation used for the ANATEM model. The formulation presented in this paper has been applied in a straightforward fashion. However, the authors are convinced that further statistical developments concerning the way the two models are combined (e.g. forecast combination methods as in Winkler and Makridakis, 1983 or Hoeting et al., 1999) could improve temporal correlation, reduce mean and variability bias and allow probabilistic calibration (not addressed in this paper).

The choice of the reference series used for the local model also presents a challenging issue. It has been shown that the good performance of the models largely depends on this

local information. A thorough analysis of the sensitivity to the choice of the reference time series should be carried out. Considering the importance of local information, an extension of the method should also consider the possibility of making use of other historical stations, if available, in the neighbourhood of the region of reconstruction. Cases with multiple historical stations available would open the door to other alternative reconstruction approaches (as stated in the introduction). Of course, historical local scale data covering long historical periods are very scarce and sparse. The results also show that the reconstruction skill decreases as the distance from the reference station increases. The region considered in the present study is relatively small. The importance of the reference station would be expected to decrease for reconstructions concerning larger regions. At the same time, in such cases, the contribution of the large-scale information would be expected to be higher. Additional work is definitively required to assess the relative interest of both components of the ANATEM model in this context.

Because of the numerous scientific and operational stakes associated with the characterisation of long-term variability, the authors are confident that all of these questions will be tackled by the scientific community in the coming years. A major application of such reconstructions will obviously be the reconstruction of long-term variations for a number of climate-driven variables. For example, the long-term climatic time series produced in the present work have been used to reconstruct long-term hydrological time series at multiple hydrometric stations of the Durance basin (Kuentz, 2013; Mathevet et al., 2013). Thanks to the availability of long observed discharge series, this latter study provided an independent hydrological validation of the climatic reconstructions over the entire 20th century.

*Acknowledgements.* The authors are thankful to Meteo-France and the HISTALP project (Historical Instrumental Climatological Surface Time Series Of The Greater Alpine Region) for providing long-term climatic data, as well as to NOAA for the Twentieth Century Reanalysis Project dataset. The authors also wish to thank warmly Vazken Andréassian, Anne-Catherine Favre and Cristian Perret for their support and their helpful comments during different phases of this work. Last but not least, the authors are very grateful to Bettina Schaeffli and the three anonymous reviewers for their positive, encouraging and demanding comments that significantly contributed to the improvement of this paper.

## References

- Auer, I., Böhm, R., Jurkovic, A., Lipa, W., Orlik, A., Potzmann, R., Schöner, W., Ungersböck, M., Matulla, C., Briffa, K., Jones, P., Efthymiadis, D., Brunetti, M., Nanni, T., Maugeri, M., Mercalli, L., Mestre, O., Moisselin, J.-M., Begert, M., Müller-Westermeier, G., Kveton, V., Bochnicek, O., Stastny, P., Lapin, M., Szalai, S., Szentimrey, T., Cegnar, T., Dolinar, M., Gajic-Capka, M., Zaninovic, K., Majstorovic, Z., and Nieplova, E.: HISTALP historical instrumental climatological surface time series of the Greater Alpine Region, *Int. J. Climatol.*, 27, 17–46, doi:10.1002/joc.1377, 2007.
- Bárdossy, A. and Pegram, G.: Infilling missing precipitation records – A comparison of a new copula-based method with other techniques, *Journal of Hydrology*, 519, Part A, 1162–1170, doi:10.1016/j.jhydrol.2014.08.025, 2014.
- Ben Daoud, A., Sauquet, E., Lang, M., Obled, C., and Bontron, G.: La prévision des précipitations par recherche d'analogues: état de l'art et perspectives, *Houille Blanche*, 2009-6, 60–65, doi:10.1051/lhb/2009079, 2010.
- Boé, J. and Habets, F.: Multi-decadal river flow variations in France, *Hydrol. Earth Syst. Sci.*, 18, 691–708, doi:10.5194/hess-18-691-2014, 2014.
- Bontron, G.: Prevision quantitative des precipitations: adaptation probabiliste par recherche d'analogues ; utilisation des reanalyses NCEP-NCAR et application aux precipitations du sud-est de la France, Thèse doctorat, Institut national polytechnique, Grenoble, France, 2004.
- Bontron, G. and Obled, C.: A probabilistic adaptation of meteorological model outputs to hydrological forecasting, *Houille Blanche*, 2005-1, 23–28, 2005.
- Bourqui, M., Mathevet, T., Gailhard, J., and Hendrickx, F.: Hydrological validation of statistical downscaling methods applied to model projections, *IAHS-AISH Publ.*, 344, 32–38, 2011.
- Chardon, J., Hingray, B., Favre, A.-C., Autin, P., Gailhard, J., Zin, I., and Obled, C.: Spatial similarity and transferability of analog dates for precipitation downscaling over France, *J. Climate*, 27, 5056–5074, doi:10.1175/JCLI-D-13-00464.1, 2014.
- Cleveland, W. S.: Robust locally weighted regression and smoothing scatterplots, *J. Am. Stat. Assoc.*, 74, 829–836, doi:10.1080/01621459.1979.10481038, 1979.
- Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.:

- The twentieth century reanalysis project, *Q. J. Roy. Meteorol. Soc.*, 137, 1–28, doi:10.1002/qj.776, 2011.
- Coulibaly, P. and Evora, N. D.: Comparison of neural network methods for infilling missing daily weather records, *Journal of Hydrology*, 341, 27–41, doi:10.1016/j.jhydrol.2007.04.020, 2007.
- CSIRO: Climate variability and change in south-eastern Australia: a synthesis of findings from Phase 1 of the South Eastern Australian Climate Initiative (SEACI), Tech. rep., South Eastern Australian Climate initiative, Australia, 2010.
- Deser, C., Phillips, A., Bourdette, V., and Teng, H.: Uncertainty in climate change projections: the role of internal variability, *Clim. Dynam.*, 38, 527–546, doi:10.1007/s00382-010-0977-x, 2012.
- Dufour, C., and Garçon, R.: Méthode statistique de recalage du modèle de prévision au pas journalier MORDOR dans le cadre du projet Vienne, Rapport technique, Electricité de France, Grenoble, France, 1997.
- Frank, D. and Esper, J.: Temperature reconstructions and comparisons with instrumental data from a tree-ring network for the European Alps, *Int. J. Climatol.*, 25, 1437–1454, doi:10.1002/joc.1210, 2005.
- Gottardi, F., Obled, C., Gailhard, J., and Paquet, E.: Statistical reanalysis of precipitation fields based on ground network data and weather patterns: Application over French mountains, *J. Hydrol.*, 432–433, 154–167, doi:10.1016/j.jhydrol.2012.02.014, 2012.
- Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling, *J. Hydrol.*, 377, 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.
- Hänggi, P., and Weingartner, R.: Inter-annual variability of runoff and climate within the Upper Rhine River basin, 1808–2007, *Hydrolog. Sci. J.*, 56, 34–50, doi:10.1080/02626667.2010.536549, 2011.
- Hannaford, J., Buys, G., Stahl, K., and Tallaksen, L. M.: The influence of decadal-scale variability on trends in long European streamflow records, *Hydrol. Earth Syst. Sci.*, 17, 2717–2733, doi:10.5194/hess-17-2717-2013, 2013.
- Hannah, D. M., Demuth, S., van Lanen, H. A. J., Looser, U., Prudhomme, C., Rees, G., Stahl, K., and Tallaksen, L. M.: Large-scale river flow archives: importance, current status and future needs, *Hydrol. Process.*, 25, 1191–1200, doi:10.1002/hyp.7794, 2011.
- Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in regional climate predictions, *B. Am. Meteorol. Soc.*, 90, 1095–1107, doi:10.1175/2009BAMS2607.1, 2009.
- Hingray, B., Hendrickx, F., Bourqui, M., Creutin, J.-D., François, B., Gailhard, J., Lafaysse, M., Lemoine, N., Mathevet, T., Mezghani, A., and Monteil, C.: RIWER 2030 – climats régionaux et

- incertitudes, ressource en eau et gestion associée de 1960 à 2030, Rapport final projet ANR VMCS 2009-2012, ANR, CNRS/LTHE, EDF/LNHE, EDF/DTG, Grenoble, France, 2013.
- Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T.: Bayesian model averaging: a tutorial, *Stat. Sci.*, 14, 382–401, 1999.
- Horton, P., Jaboyedoff, M., Metzger, R., Obled, C., and Marty, R.: Spatial relationship between the atmospheric circulation and the precipitation measured in the western Swiss Alps by means of the analogue method, *Nat. Hazards Earth Syst. Sci.*, 12, 777–784, doi:10.5194/nhess-12-777-2012, 2012.
- Imbeaux, E.: La Durance: Régime, Crues et Inondations, Vve Ch. Dunod, Paris, France, 1892.
- Jouzel, J., Masson-Delmotte, V., Cattani, O., Dreyfus, G., Falourd, S., Hoffmann, G., Minster, B., Nouet, J., Barnola, J. M., Chappellaz, J., Fischer, H., Gallet, J. C., Johnsen, S., Leuenberger, M., Loulergue, L., Luethi, D., Oerter, H., Parrenin, F., Raisbeck, G., Raynaud, D., Schilt, A., Schwander, J., Selmo, E., Souchez, R., Spahni, R., Stauffer, B., Steffensen, J. P., Stenni, B., Stocker, T. F., Tison, J. L., Werner, M., and Wolff, E. W.: Orbital and millennial antarctic climate variability over the past 800,000 years, *Science*, 317, 793–796, doi:10.1126/science.1141038, 2007.
- Kuentz, A.: Un siècle de variabilité hydro-climatique sur le bassin de la Durance: recherches historiques et reconstitutions, Thèse de doctorat, AgroParisTech, Paris, France, 2013.
- Kuentz, A., Mathevet, T., Perret, C., Gailhard, J., and Andréassian, V.: Uncertainty estimation and reconstruction of historical streamflow records, *Geophys. Res. Abstr.*, 14, EGU2012-5632-1, 2012.
- Kuentz, A., Mathevet, T., Coeur, D., Perret, C., Gailhard, J., Guérin, L., Gash, Y., and Andréassian, V.: Hydrométrie et hydrologie historiques du bassin de la Durance, *Houille Blanche*, 2014-4, 57–63, doi:10.1051/lhb/2014039, 2014.
- Lafaysse, M., Hingray, B., Mezghani, A., Gailhard, J., and Terray, L.: Internal variability and model uncertainty components in future hydrometeorological projections: The Alpine Durance basin, *Water Resour. Res.*, 50, 3317–3341, doi:10.1002/2013WR014897, 2014.
- Lorenz, E. N.: Atmospheric predictability as revealed by naturally occurring analogues, *J. Atmos. Sci.*, 26, 636–646, doi:10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2, 1969.
- Lorenzo-Lacruz, J., Vicente-Serrano, S., López-Moreno, J., Morán-Tejeda, E., and Zabalza, J.: Recent trends in Iberian streamflows (1945–2005), *J. Hydrol.*, 414–415, 463–475, doi:10.1016/j.jhydrol.2011.11.023, 2012.
- Madden, R. A.: Estimates of the natural variability of time-averaged sea-level pressure, *Mon. Weather Rev.*, 104, 942–952, doi:10.1175/1520-0493(1976)104<0942:EOTNVO>2.0.CO;2, 1976.

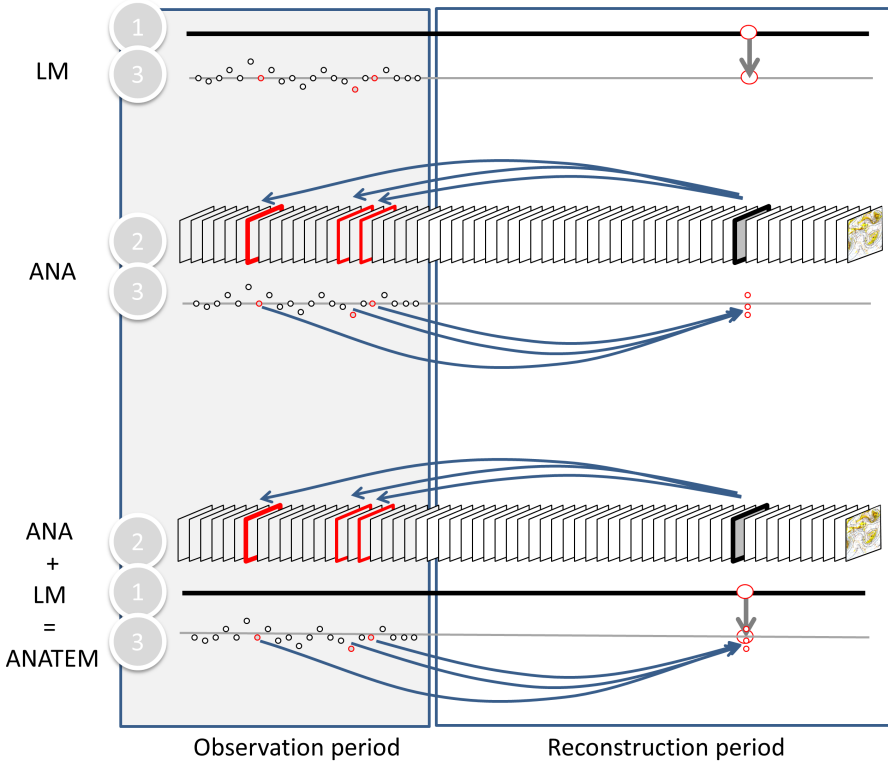
- Marty, R., Zin, I., Obled, C., Bontron, G., and Djerboua, A.: Toward real-time daily PQPF by an analog sorting approach: application to flash-flood catchments, *J. Appl. Meteorol. Clim.*, 51, 505–520, doi:10.1175/JAMC-D-11-011.1, 2011.
- Mathevet, T., and Garçon, R.: Tall tales from the hydrological crypt: are models monsters?, *Hydrolog. Sci. J.*, 55, 857–871, doi:10.1080/02626667.2010.503934, 2010.
- Mathevet, T., Kuentz, A., Gailhard, J., and Andreassian, V.: Characterizing a century of climate and hydrological variability of a mediterranean and mountainous watersheds: the Durance River Case-Study, in: *AGU Fall Meeting Abstracts, #H31B-1156*, San Francisco, California, USA, 2013.
- Obled, C., Bontron, G., and Garçon, R.: Quantitative precipitation forecasts: a statistical adaptation of model outputs through an analogues sorting approach, *Atmos. Res.*, 63, 303–324, doi:10.1016/S0169-8095(02)00038-8, 2002.
- Pfister, C., Weingartner, R., and Luterbacher, J.: Hydrological winter droughts over the last 450 years in the Upper Rhine basin: a methodological approach, *Hydrolog. Sci. J.*, 51, 966–985, doi:10.1623/hysj.51.5.966, 2006.
- Poli, P., Hersbach, H., Tan, D., Dee, D., Thépaut, J. N., Simmons, A., Peubey, C., Laloyaux, P., Komori, T., and Berrisford, P.: The data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C), *ERA Report Series 14*, ECMWF, UK, 2013.
- Renard, B.: *Détection et prise en compte d'éventuels impacts du changement climatique sur les extrêmes hydrologiques en France*, Thèse de doctorat, Institut National Polytechnique de Grenoble, Grenoble, 2006.
- Renner, M. and Bernhofer, C.: Long term variability of the annual hydrological regime and sensitivity to temperature phase shifts in Saxony/Germany, *Hydrol. Earth Syst. Sci.*, 15, 1819–1833, doi:10.5194/hess-15-1819-2011, 2011.
- Rood, S. B., Samuelson, G., Weber, J., and Wywrot, K.: Twentieth-century decline in streamflows from the hydrographic apex of North America, *J. Hydrol.*, 306, 215–233, doi:10.1016/j.jhydrol.2004.09.010, 2005.
- Stahl, K., Hisdal, H., Hannaford, J., Tallaksen, L. M., van Lanen, H. A. J., Sauquet, E., Demuth, S., Fendekova, M., and Jódar, J.: Streamflow trends in Europe: evidence from a dataset of near-natural catchments, *Hydrol. Earth Syst. Sci.*, 14, 2367–2382, doi:10.5194/hess-14-2367-2010, 2010.

- Teegavarapu, R. S.: Spatial interpolation using nonlinear mathematical programming models for estimation of missing precipitation records, *Hydrological Sciences Journal*, 57, 383–406, doi:10.1080/02626667.2012.665994, 2012.
- Teng, J., Chiew, F., Timbal, B., Wang, Y., Vaze, J., and Wang, B.: Assessment of an analogue downscaling method for modelling climate change impacts on runoff, *J. Hydrol.*, 472–473, 111–125, doi:10.1016/j.jhydrol.2012.09.024, 2012.
- Terray, L., and Boé, J.: Quantifying 21st-century France climate change and related uncertainties, *CR Geosci.*, 345, 136–149, doi:10.1016/j.crte.2013.02.003, 2013.
- Teweles, S. and Wobus, H.: Verification of prognostic charts, *B. Am. Meteorol. Soc.*, 35, 455–63, 1954.
- Timbal, B., Arblaster, J. M., and Power, S.: Attribution of the late-twentieth-century rainfall decline in Southwest Australia, *J. Climate*, 19, 2046–2062, doi:10.1175/JCLI3817.1, 2006.
- Van Den Dool, H. M.: A new look at weather forecasting through analogues, *Mon. Weather Rev.*, 117, 2230–2247, doi:10.1175/1520-0493(1989)117<2230:ANLAWF>2.0.CO;2, 1989.
- Wetterhall, F., Halldin, S., and Xu, C.-Y.: Statistical precipitation downscaling in central Sweden with the analogue method, *J. Hydrol.*, 306, 174–190, doi:10.1016/j.jhydrol.2004.09.008, 2005.
- Wilhelm, B., Arnaud, F., Sabatier, P., Crouzet, C., Brisset, E., Chaumillon, E., Disnar, J.-R., Guiter, F., Malet, E., Reyss, J.-L., Tachikawa, K., Bard, E., and Delannoy, J.-J.: 1400 years of extreme precipitation patterns over the Mediterranean French Alps and possible forcing mechanisms, *Quaternary Res.*, 78, 1–12, doi:10.1016/j.yqres.2012.03.003, 2012.
- Wilhelm, B., Arnaud, F., Sabatier, P., Magand, O., Chapron, E., Courp, T., Tachikawa, K., Fanget, B., Malet, E., Pignol, C., Bard, E., and Delannoy, J. J.: Palaeoflood activity and climate change over the last 1400 years recorded by lake sediments in the north-west European Alps, *J. Quaternary Sci.*, 28, 189–199, doi:10.1002/jqs.2609, 2013.
- Winkler, R. L. and Makridakis, S.: The combination of forecasts, *J. Roy. Stat. Soc. A-G.*, 146, 150–157, doi:10.2307/2982011, 1983.
- Zhang, X., Harvey, K. D., Hogg, W. D., and Yuzyk, T. R.: Trends in Canadian streamflow, *Water Resour. Res.*, 37, 987–998, doi:10.1029/2000WR900357, 2001.

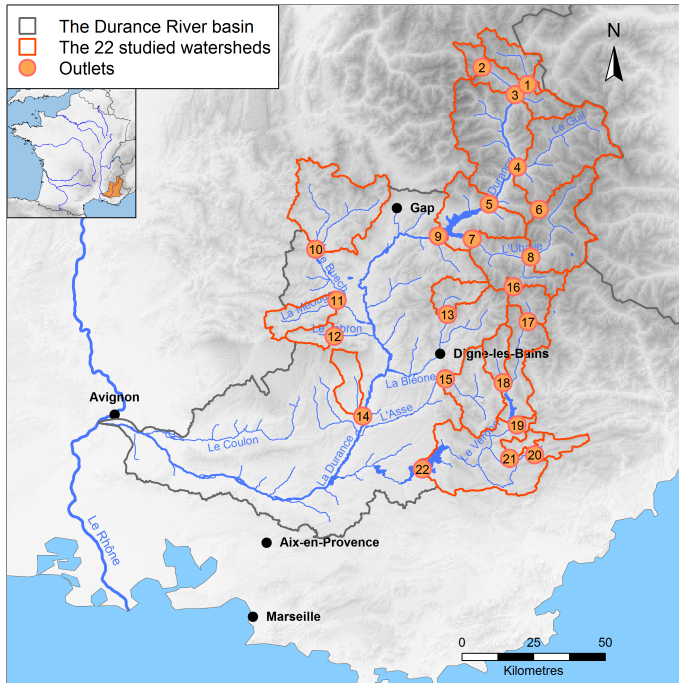


**Table 1.** Main characteristics of the 22 selected watersheds. The number in the # column correspond to those indicated in Fig. 2,  $P_A$  is the annual mean precipitation (over the period 1948–2010) and  $T$  the mean air temperature (over the period 1948–2010)

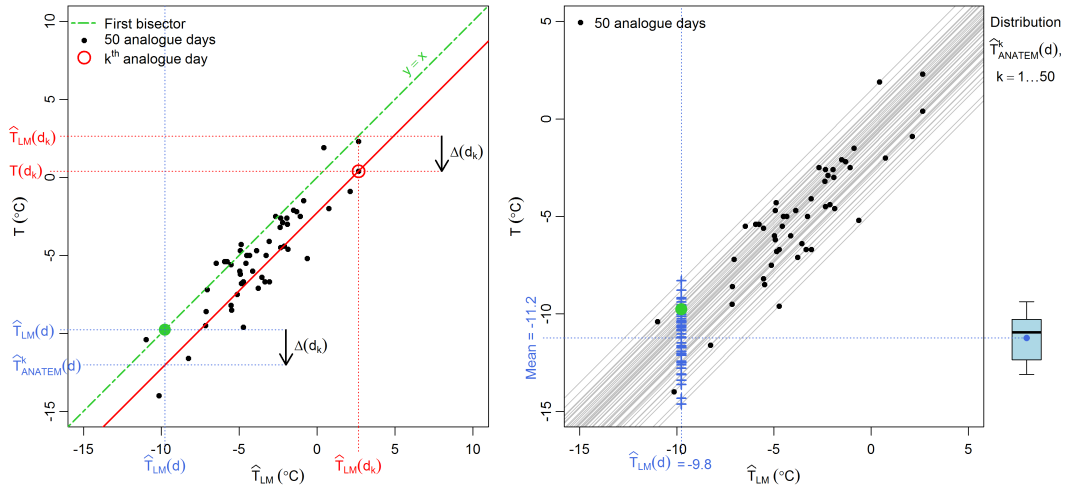
#	Name	Altitude m	Area km <sup>2</sup>	$\bar{P}_A$ mm yr <sup>-1</sup>	$\bar{T}$ °C
1	The Durance river at Val-des-Près	1360	203	1322	2.5
2	The Guisane river at Monêtier-les-Bains	1510	78	1627	2.3
3	The Durance river at Briançon	1187	548	1381	2.9
4	The Guil river at Montdauphin	895	725	1087	3.3
5	The Durance river at La Clapière	787	2170	1352	3.5
6	The Riou de Crachet river at Saint-Paul	2020	4	1532	1.6
7	The Ubaye river at Roche-Rousse	790	946	1235	4.1
8	The Ubaye river at Barcelonnette	1132	549	1201	3.6
9	The Durance river at Serre-Ponçon	652	3582	1301	4.0
10	The Buech river at Les Chambons	662	723	1259	7.4
11	The Méouge river at Méouge	545	221	1094	8.9
12	The Jabron river at Piedguichard	593	89	1206	9.1
13	The Bes river at La Javie	805	165	1085	6.6
14	The Lauzon river at Villeneuve	341	124	1097	10.4
15	The Asse river at La Clue de Chabrières	605	375	1077	8.6
16	The Verdon river at Allos	1780	10	1592	2.7
17	The Verdon river at Colmars	1230	158	1453	4.3
18	The Issole river at Saint-André-les-Alpes	931	137	1229	6.8
19	The Verdon river at Castillon	790	657	1319	6.2
20	The Artuby river at La Bastide	1008	91	1272	8.4
21	The Jabron river at Comps-sur-Artuby	782	66	1116	9.0
22	The Verdon river at Sainte-Croix	400	1625	1176	8.2



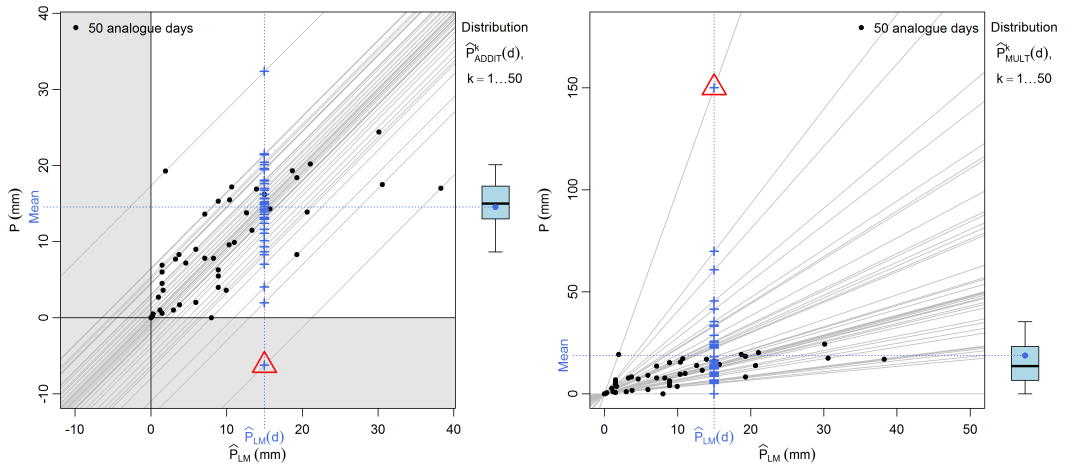
**Figure 1.** Schemes of the 3 reconstruction models : local model (LM – top scheme), analogue model (ANA – middle scheme), combined local+analogue model (ANATEM – bottom scheme). Predictors are either (1) local scale meteorological predictors (LM model), (2) mesoscale atmospheric predictors (ANA model) or both (1)+(2) (ANATEM model). Local scale predictors are daily observations of the variable at one (possibly several) neighbouring precipitation or temperature station (for the present work, Gap rain gauge, Marseille temperature station for precipitation / temperature reconstruction respectively). Mesoscale predictors are fields of atmospheric variables (700 and 1000hPa geopotential heights over a mesoscale European domain). Local and mesoscale predictors cover the whole period (observation + reconstruction). The three reconstruction models are first developed and evaluated based on their reconstruction skill for the observation period where concomitant observations of the target variable are available (dots of series 3 in the scheme, period 1948-2010 in the present work). Models are next applied for the reconstruction of each day of the reconstruction period (period 1883-2010 in the present work). Note : The reconstruction period can also include the observation period (this is the case in the present work).



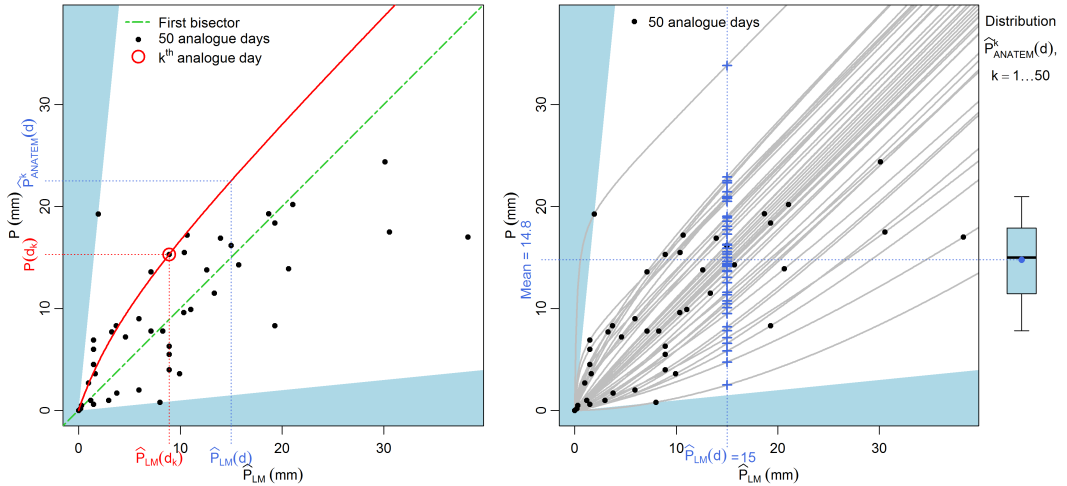
**Figure 2.** Map of the study area with the 22 selected watersheds.



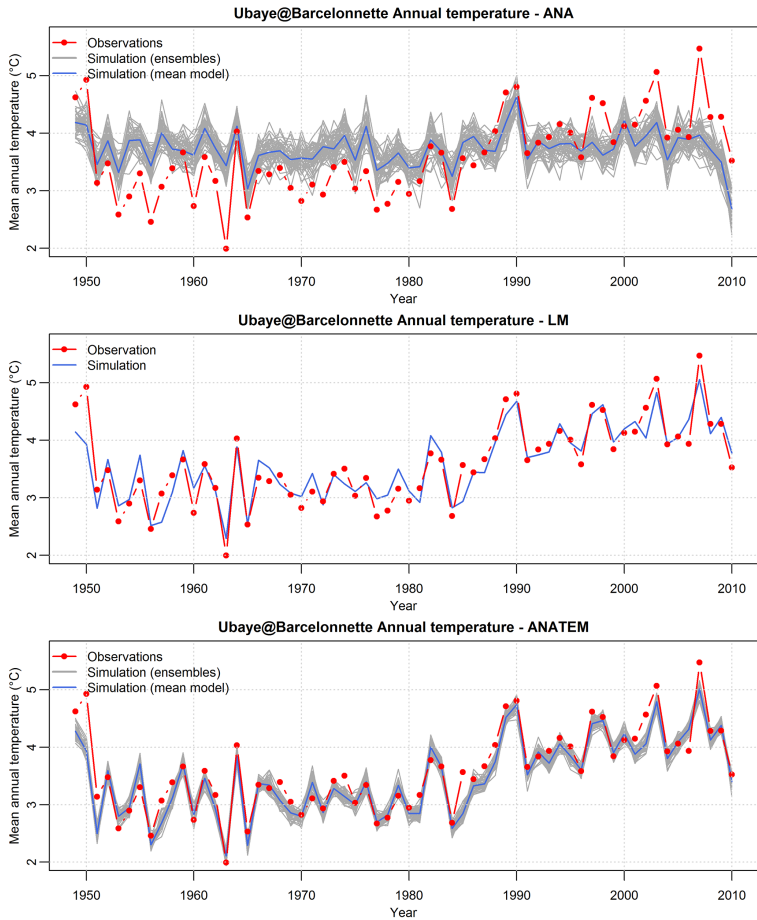
**Figure 3.** Representation of the ANATEM formulation for air temperature reconstruction of a given day  $d$ . Left panel: observed temperature for any analog day as a function of the temperature estimate with LM for this day. The green point corresponds to the target day  $d$ . The red-circled point corresponds to a particular analog day  $d_k$ . For this analog day, the corrected estimate for day  $d$  is  $\hat{T}_{ANATEM}^k(d)$  which is obtained as  $\hat{T}_{LM}(d) + \Delta(d_k)$  where  $\Delta(d_k) = T(d_k) - \hat{T}_{LM}(d_k)$ . Right panel: probabilistic prediction obtained for  $d$  from the 50 analogues. The corresponding boxplot (10, 25, 50, 75 and 90% quantiles) is given to the right of the figure (the blue point indicates the mean value).



**Figure 4.** Representation of the additive and multiplicative formulation for precipitation reconstruction from a local model and 50 analogue days for a given day  $d$ . Left panel: additive formulation for the correction. Right panel: multiplicative formulation for the correction. The triangles highlight anomalous or potentially anomalous corrected predictions.

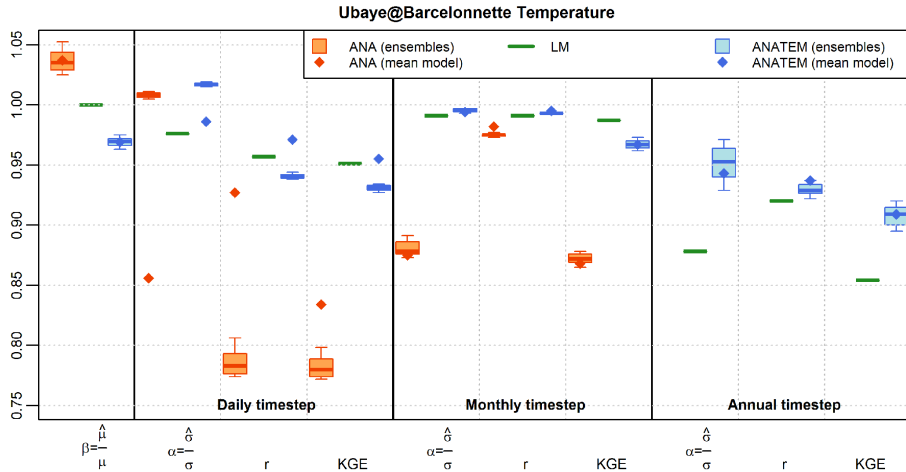


**Figure 5.** Representation of the ANATEM formulation for precipitation reconstruction for a given day  $d$ . Left panel: observed precipitation at the target site for each of the analogue days as a function of the precipitation estimate from LM for the same days. For a given analogue day  $k$ , the corrected estimate for day  $d$  is the ordinate of the point on the red curve (defined by Eq. (10)) crossed by abscissa  $\hat{P}_{LM}(d)$ . Right panel: probabilistic prediction obtained for  $d$  from the 50 analogues. The corresponding boxplot (10, 25, 50, 75 and 90% quantiles) is given to the right of the figure (the blue point indicates the mean value).

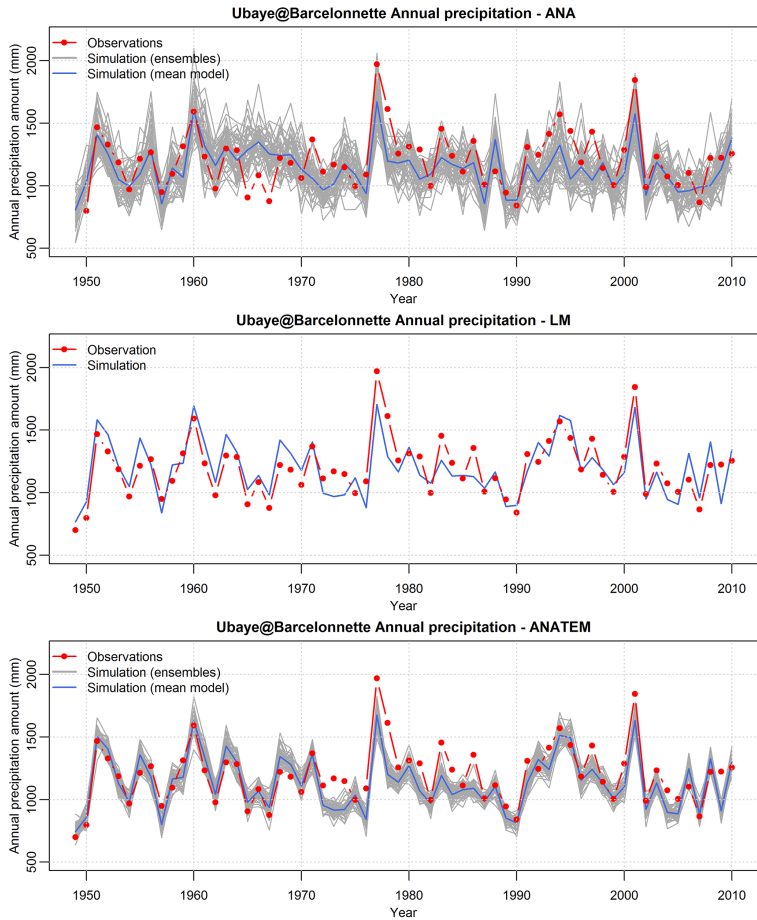


**Figure 6.** Annual time series of air temperature reconstructions for the Ubaye River at Barcelonnette watershed by the analogue (ANA), local (LM) and ANATEM models.

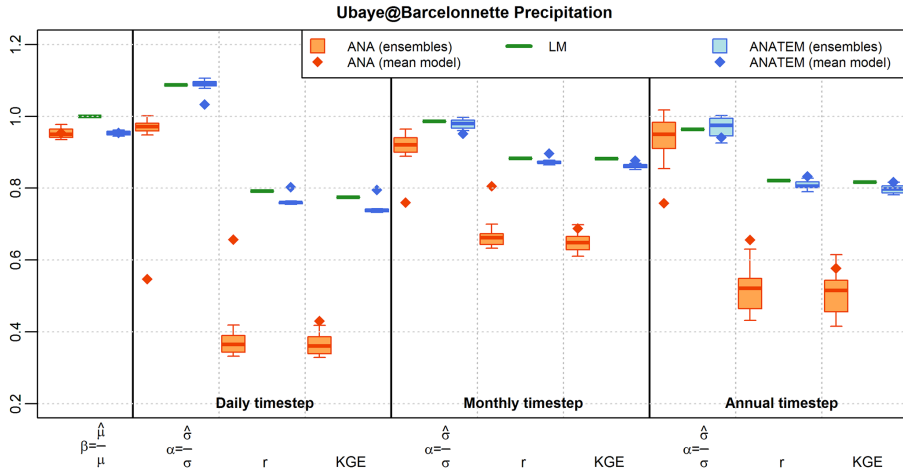




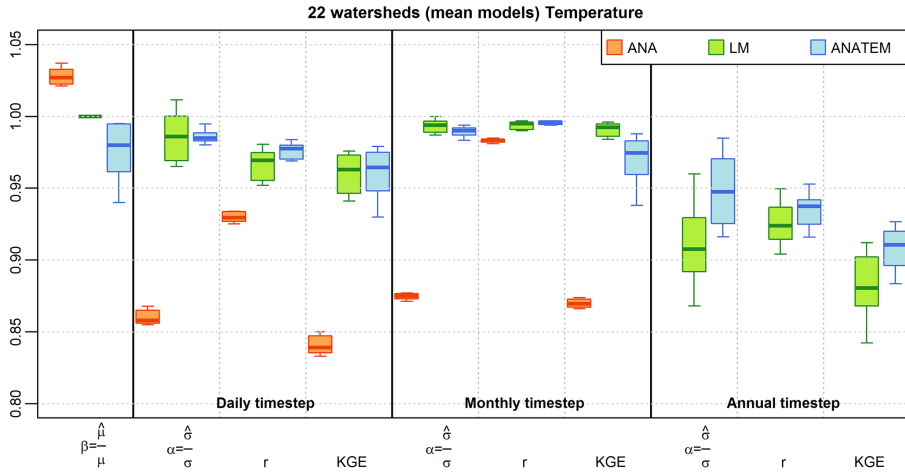
**Figure 7.** Daily, monthly and annual performance criteria of air temperature reconstructions for the Ubaye River at Barcelonnette watershed by the analogue (ANA), local (LM) and ANATEM models. For the annual time step, ANA results are smaller than 0.75; they therefore do not appear in the figure.



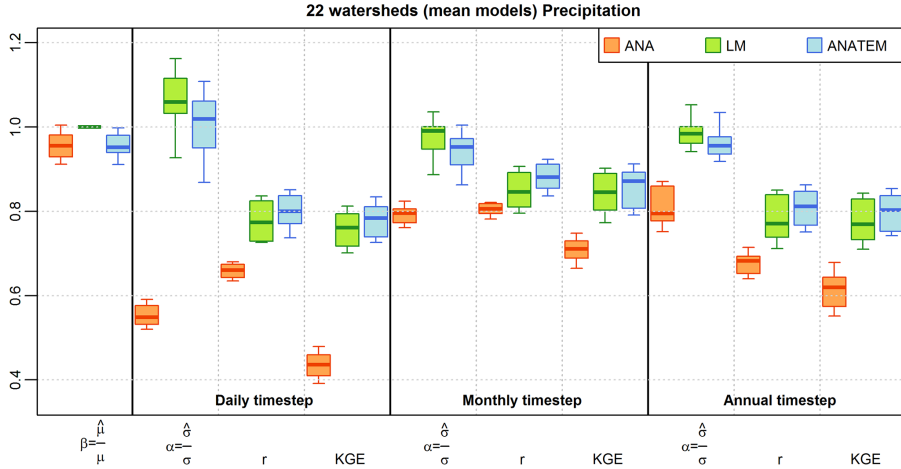
**Figure 8.** Annual time series of precipitation reconstructions for the Ubaye River at Barcelonnette watershed by the analogue (ANA), local (LM) and ANATEM models.



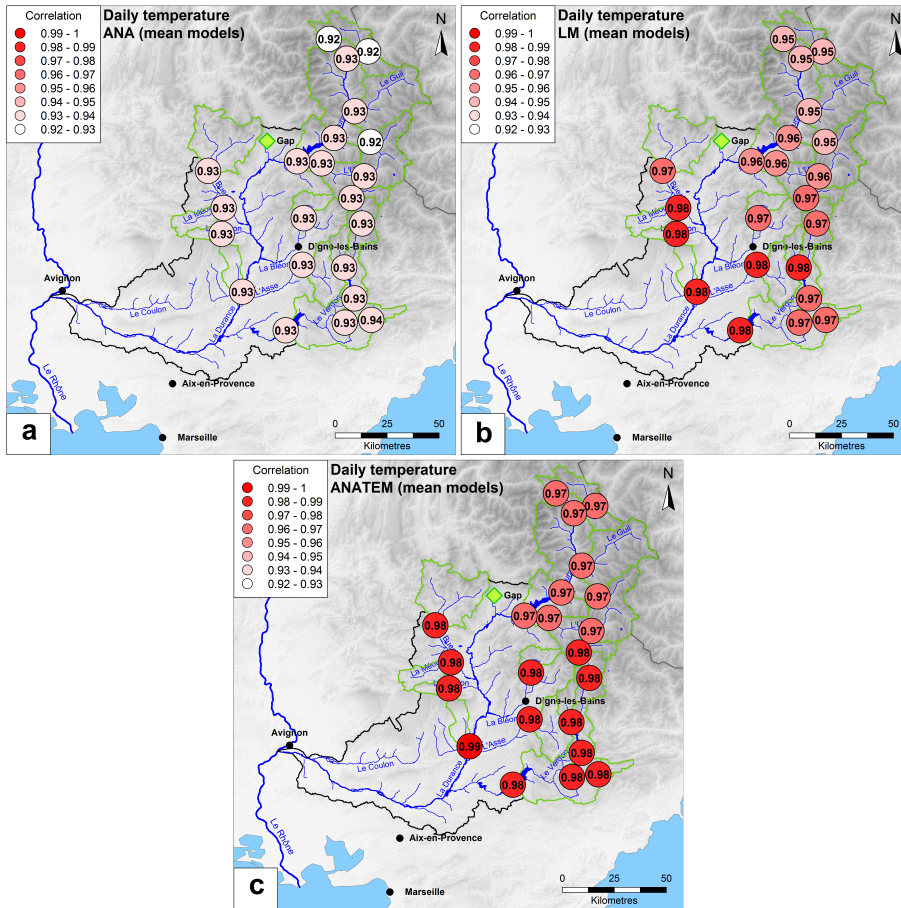
**Figure 9.** Daily, monthly and annual performance criteria of precipitation reconstructions for the Ubaye River at Barcelonnette watershed by the analogue (ANA), local (LM) and ANATEM models.



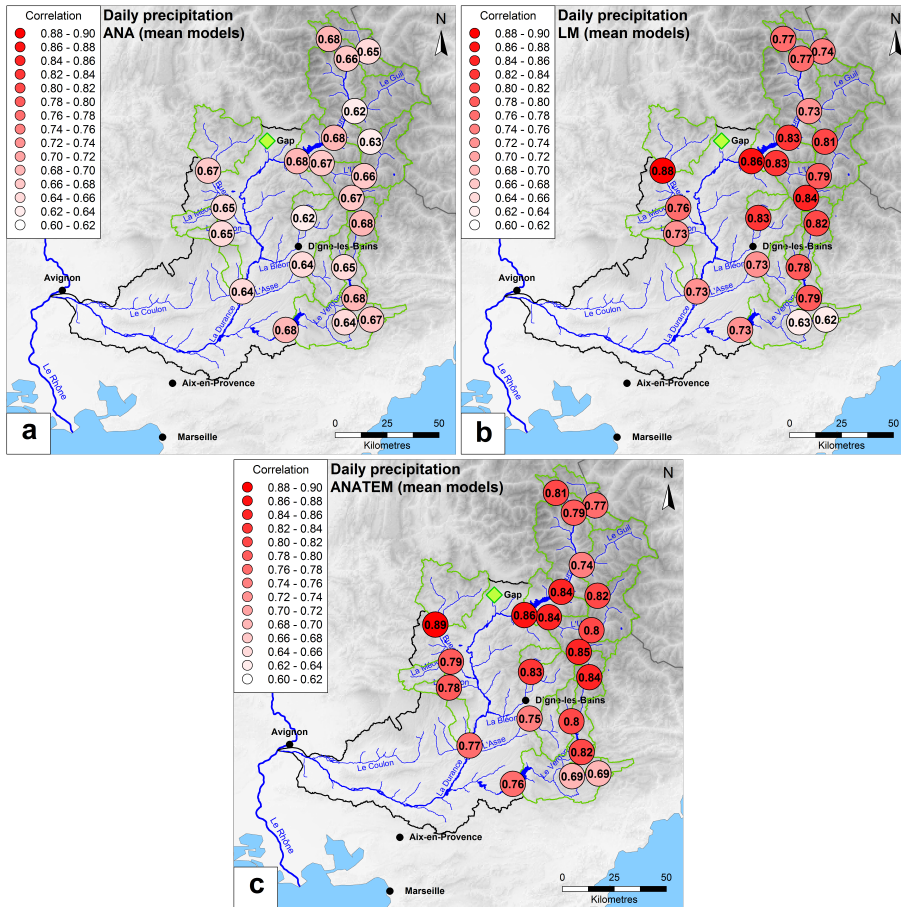
**Figure 10.** Daily, monthly and annual performance criteria of air temperature mean reconstructions for 22 watersheds by the analogue (ANA), local (LM) and ANATEM models. For the annual time step, ANA results are smaller than 0.6; they therefore do not appear in the figure.



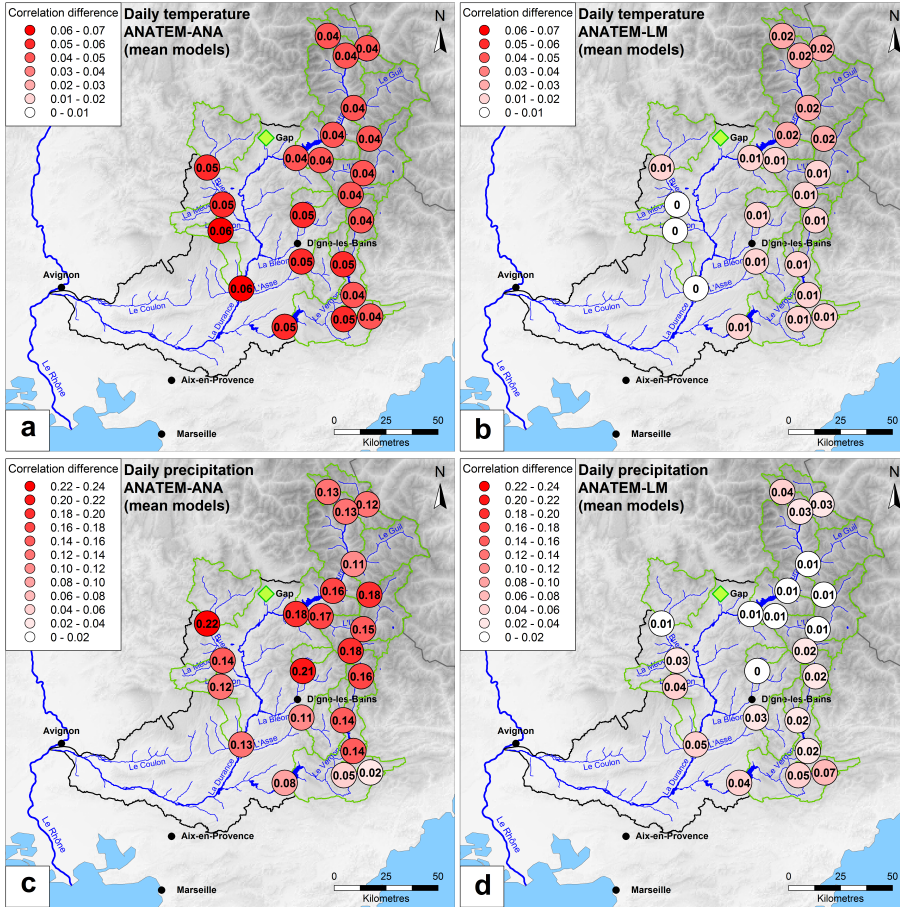
**Figure 11.** Daily, monthly and annual performance criteria of precipitation mean reconstructions for 22 watersheds by the analogue (ANA), local (LM) and ANATEM models.



**Figure 12.** Regional correlation patterns of air temperature mean reconstructions by the (a) analogue model (ANA), (b) local model (LM) and (c) ANATEM model.



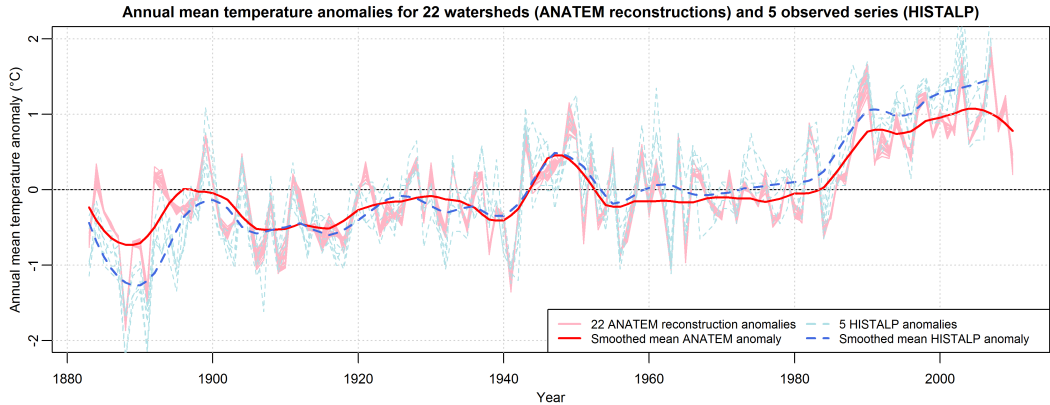
**Figure 13.** Regional correlation patterns of precipitation mean reconstructions by the (a) analogue model (ANA), (b) local model (LM) and (c) ANATEM model.



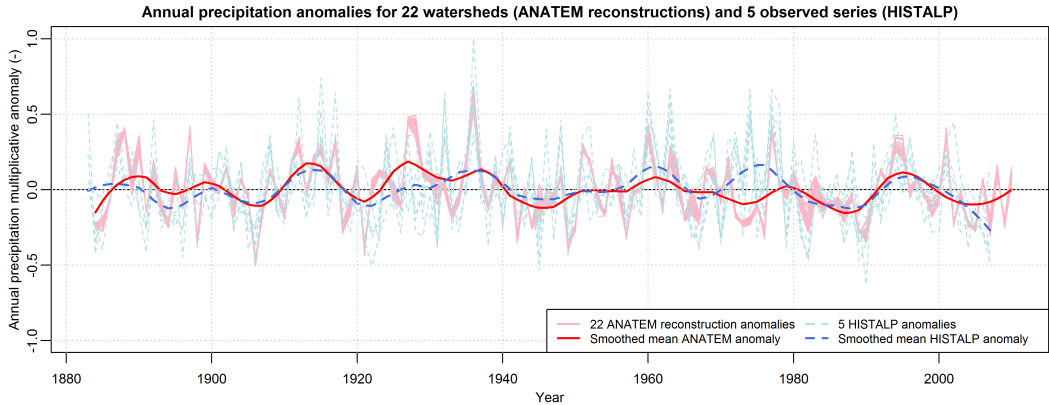


**Figure 14.** Spatial patterns of **(a,c)** the contribution of the LM model to ANATEM performance (estimated by the difference between the performance of the ANATEM and ANA models) for air temperature (a) and precipitation (c) reconstructions.

Spatial patterns and of **(b,d)** the contribution of the ANA model to ANATEM performance (estimated by the difference between the performance of the ANATEM and LM models) for air temperature (b) and precipitation (d) reconstructions. Note the different color scales for precipitation and temperature reconstructions.



**Figure 15.** Mean annual air temperature additive anomaly for the 22 watersheds (ANATEM) and 5 stations (HISTALP). The additive anomaly for a given year has been computed as the difference between the annual temperature for this year and the 1883–2010 mean.



**Figure 16.** Mean annual precipitation multiplicative anomaly for the 22 watersheds (ANATEM) and 5 stations (HISTALP). The multiplicative anomaly for a given year has been computed as the ratio between the annual precipitation for this year and the 1883-2010 mean.