Author response about "Building long-term and high spatio-temporal resolution precipitation and air temperature reanalyses by mixing local observations and global atmospheric reanalyses: the ANATEM method".

A. Kuentz, T. Mathevet, B. Hingray, J. Gailhard

May 10, 2015

Contents

Response to the reviews with detail of the related changes in the manuscript

The point-by-point response to the reviewer's comments is given in the individual replies to each comment and copied below. All the lines, pages and figures numbers refers to the HESSD version. The answers have been written during the revision process and some parts of the paper have been changed various times, making the first suggested change out of date. The mathematical notations have also been changed after the responses to reviewers 1 and 2. All final changes are presented in the marked-up manuscript at the end of this author's response.

Response to reviewer #1

The authors would like to thank Referee 1 for his positive evaluation and for his constructive comments and suggestions. The detailed answers to the specific comments are presented below.

P313, L7: Please rephrase: "time-series of different regions and climates."

We propose the following sentence: *"When considering climate or hydrology, numerous studies aim at characterising variability, trends or breaks using observed time-series representing different regions or climate of the world."*

P313, L10: Please rephrase: "...time-series that suffer from ..."

We propose the following sentence: *"However, in hydrology, these studies are usually limited to reduced temporal scale (mainly few decades, seldomly a century) because they are dependant on observed time-series which have a limited spatio-temporal density."*

P313, L12: the correct term is "climatic information" (without s)

This will be corrected in the revised version

P314, L7: The "related uncertainties" refer to uncertainties related to multi-decadal variations? If so, please indicate the type of uncertainty.

Yes, that is right. This will be clarified as:

"In a non-stationary climate, multi-decadal variations can remain high above the long-term trend. In climate projections for the coming decades, they often represent a major source of uncertainty (e.g. Hawkins and Sutton, 2009; Deser et al., 2012). For precipitation or hydrometeorological variables such as streamflow, uncertainties related to multi-decadal variations can be as large as or even larger than uncertainties due to climate models (e.g. Terray and Boé, 2013; Lafaysse et al., 2014)."

P315, L1: "longer than 100 years" (plural)
P315, L 19: "streamflow variations" instead of "streamflows variations"?
P317, L3: Maybe "bounded" might be more appropriate in this context than "limited".
P318, L13: The correct longitude should be 8°W.

These points will be corrected in the revised version

P318, L18-19: This statement is not clear: The methodologies you are discussing here are based on the reconstruction site only?

Yes, this is what we meant; we propose this rephrasing: *"Different methods are classically used to reconstruct climatic observations. Some of them are only based on the series at the reconstruction site itself (long-term average or regime, temporal interpolation techniques...), others are based on external data (proxy data) used to calibrate and run a reconstruction model."*

P319, L19: $\hat{x}$ instead of x ?

Yes, the correct phrase should be: *"The estimate $X_{LM,d}$ of the target variable X obtained with LM for a given day d has the classical following expression :"*

P321, L10: What do you mean exactly by this? If I understand that correctly, only the ANATEM approach incorporates uncertainty in terms of Eq. 4, whereas the local model itself is parameterized through neglecting ε_d.

Yes the part mentioning "another way of considering uncertainty" refers to the ANATEM approach only. To improve understanding we suggest removing the last part. The new sentence reads: "As explained previously for the air temperature reconstruction, a simple version of this model with a residual term considered equal to zero is used in this paper."

P322, L2: I expected k being the index variable, whereas n indicates the total number of days used for the similarity analyses. If so, I recommend replacing k by n since k is used for specific days later on in the manuscript.

Yes; it should be n here, this will be corrected.

P322, L6: Please indicate which archive is used here (SPAZM?).
SPAZM, this will be added.

P323, L1-7: It remains unclear how you have derived the ensembles using ANA and ANATEM. This is my point of criticism as described in the general comments section. It is clear that we can

select among n days for which the spatial geo-potential height distribution is similar to that observed for the day of interest. Have the ensembles been achieved through drawing random numbers using the distributions (e.g., box plot in Fig. 2) derived for each day? Please provide some more details with respect to the ensembles.

For each day where an analog reconstruction is made, the 50 nearest analogs days are selected (the analogy being defined with the TW criterion from the atmospheric features described earlier in the paper). Then, the distribution used in ANA and ANATEM reconstructions is the empirical distribution constituted from the 50 values of air temperature (or precipitation) observed respectively for these 50 nearest analogs days. There is therefore no random process in the elaboration of the distribution. Note that in some other papers, authors use an analog method where they calibrate a gamma distribution on the empirical distribution and then they randomly draw in this distribution (e.g. Marty et al., 2008). This allows to better represents extreme values. Since, we were more interested in the generality of the ANATEM method, we didn't add this modelling and further generation process in our ANA method.

We propose this rephrasing of the paragraph:

"The reconstruction is deterministic when only one analog is used (classically the nearest analog). The analog day can be also selected among the n nearest analogs. An ensemble of reconstructions can be produced when all n nearest analogs 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 analogs respectively. As a result, an ensemble of reconstructions can be produced. This allows characterizing the uncertainty in the reconstruction. The ensemble of reconstructions obtained with ANA model for the variable X and day d will be written in the following $\left[X_{ANA_d^k}\right]_{k=1\ldots n}$ *where* $k = 1 \ldots n$ *refers to the n nearest analogs selected for the day d. In the present case, the selection is done among the 50 nearest analogs* ($n = 50$)*."*

P325, L12: Why does the local model yield a value of 9.0 °C? From the figure, I would expect 9.8 °C.

Yes, this is a mistake that will be corrected

P326, Eq. 9: It remains unclear to me, why you have chosen this type of equation. Could you please provide some more information with respect to the theoretical back- ground (e.g., appropriate shape for typical values of xd and the parameters).

We acknowledge that the reference, which is an EDF internal report, is not published. We made nevertheless this choice because we already had an experience with this formulation in the field of data assimilation for operational streamflow forecasts. This formulation is used to post-process streamflow forecasts based on the analysis of Rainfall-Runoff model past residuals. Despite its rather empirical nature, the formulation proved to give satisfactory results for this post-processing application. Depending on the current hydrological processes, we may prefer to make a "multiplicative" post-processing of the forecast (typically during drought events) or an "additive" post-processing of the forecast (typically during floods). Due to these two basic properties, we decided to use this formulation for ANATEM, suitable with the problems encountered with rainfall. Another formulation could be obviously tested (as suggested by one of the examiner of Anna Kuentz PhD). Note however that this would not change the principle of the ANATEM combination. We also expect it would not drastically change the conclusions of our work.

a_d^k and b_d^k coefficients are deduced from two conditions proposed by Dufour and Garçon (1997) :

- The slope of the tangent to the curve in $x = 0$ should be $\left(\frac{P_{\mathrm{ANA}_d^k}}{P_{\mathrm{LM,ANA}_d^k}}\right)^2$
- When $P_{\mathrm{ANA}_d^k} = P_{\mathrm{LM,ANA}_d^k}$, the following should be obtained : $\widehat{P}_d^k = P_{\mathrm{LM},d}$

The first condition has been imposed empirically and selected because it gave satisfactory results, while the second condition is logically deduced from the idea of the correction model.

The first condition gives the equality :

$\frac{a_d^k}{b_d^k} = \left(\frac{P_{\mathrm{ANA}_d^k}}{P_{\mathrm{LM,ANA}_d^k}}\right)^2$

The second condition gives the equivalence relation :

$P_{\mathrm{ANA}_d^k} = P_{\mathrm{LM,ANA}_d^k} \Leftrightarrow x_d = \frac{x_d^2 + a_d^k \cdot x_d}{x_d + b_d^k} \Leftrightarrow a_d^k = b_d^k$

From these two relations the coefficients can be defined as :

$$a_d^k = P_{\mathrm{ANA}_d^k} \quad \text{and} \quad b_d^k = \frac{\left(P_{\mathrm{LM,ANA}_d^k}\right)^2}{P_{\mathrm{ANA}_d^k}}$$

Note that in the paper there are some notation mistakes that will be corrected: the value of the local model for day d is sometimes noted LM_d instead of $P_{\mathrm{LM},d}$.

P326, Eq. 11: The first approximation for very small values of x_d is clear to me. However, I do not understand why $x_d \cdot \left(1 + \frac{a_d^k}{x_d}\right) \cdot \left(1 + \frac{b_d^k}{x_d}\right)^{-1}$ yields $x_d + (a_d^k - b_d^k)$ for $x_d \to +\infty$.

Even though it becomes evident from Fig. 4 that this approach represents an additive transformation for high precipitation intensities, I would like to ask you to explain this approximation more in detail.

It comes from Taylor series expansion, see the detail below :

Using the usual first order Taylor expansion $(1+y)^{-1} = 1 + y + o(y)$ when y is close to 0 for the variable $y = \frac{b_d^k}{x_d}$:

$$x_d \cdot \left(1 + \frac{a_d^k}{x_d}\right) \cdot \left(1 + \frac{b_d^k}{x_d}\right)^{-1} \quad \sim \quad x_d \cdot \left(1 + \frac{a_d^k}{x_d}\right) \cdot \left(1 - \frac{b_d^k}{x_d}\right) \text{ when } x_d \to +\infty$$

After expansion,

$$x_d \cdot \left(1 + \frac{a_d^k}{x_d}\right) \cdot \left(1 - \frac{b_d^k}{x_d}\right) = x_d + a_d^k - b_d^k + \frac{a_d^k \cdot b_d^k}{x_d}$$

The last term tends to 0 when x tends to infinity.

P328, L11: What does SD mean? Is it the standard deviation of the time series? Please explain this abbreviation.

Yes it is the standard deviation; *"The ratio between the SD of the reconstructed and of the observed values..."* will be replaced by *"The ratio between the standard deviations of the reconstructed and of the observed time-series..."*

P328, Eq. 14: This equation is incomplete, as is it returns zero for an ideal model while the ideal value of the KGE criterion is 1 (as it is obvious from your results). The correct equation for the KGE criterion is (Gupta et al., 2009):

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

That is right, this mistake will be corrected.

P329, L9: "The ANATEM model does not capture..." instead of "do"

This will be corrected in the revised version

P330, L15-18: By definition, the local model has no mean bias. Please check the other values as well. When regarding the figure, the mentioned values are not clear to me.

That is right, there were some mistakes with the values; here is the proposed corrected paragraph:

"The distributions of criteria at the annual time-step (Fig. 8, right part) confirm these statements:

- *ANA has a moderate correlation (mean r close to 0.5), LM and ANATEM have a rather good correlation (mean r greater than 0.8);*

- *LM has no mean bias (by construction), ANA and ANATEM have a moderate mean bias (less than 0.05);*
- *ANA has a noticeable variability bias (up to 0.15), TEM and ANATEM have a limited variability bias (around 0.03).*

The hierarchy between the three models is comparable at daily and monthly time-steps, with KGE values ranging from 0.35 to 0.7 for ANA, ranging from 0.78 to 0.88 for LM and ranging from 0.73 to 0.85 for ANATEM (Fig. 8). ANA is clearly poor at a daily timestep, 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 of the models to reproduce the low and high frequency variability while the intra-annual variability is well-captured."

P330, L25: intra-annual?

Yes, this will be corrected.

P332, L6: Do you mean α instead of β ?

Line 6 is correct. Line 7 should be *"(mean r between 0.94 and 0.99)"* instead of *"(mean β between 0.95 and 0.99)"*

P332, L15-16: Please check these values carefully as they seem to differ from the values in the figure.

Yes there is again a problem with the values, the corrected sentence is: *"This is also expressed by mean KGE values, ranging from 0.25 to 0.87 for ANA, from 0.88 to 0.99 for LM and from 0.92 to 0.97 for ANATEM respectively."*

P333, L14-15: Do you mean "spatial robustness"?

This comment is unclear to us. If it means that the "spatial robustness" is not well explained, we propose to complete the text P331 L22 as follows: *"At different time-steps and for different criteria, ANA also exhibits a rather good spatial robustness of performances (i. e., homogeneity of the results at a regional scale, which could be expressed by a rather limited spread of the distribution, as shown by the distance between quantile 0.1 and 0.9)"*

P334, L11-16: Please add a brief description how to relate your statements in the text to the findings achieved through evaluating the figure (e.g., ANATEM-ANA is suitable to investigate the contribution of LM,...). This might improve the comprehensibility of the model inter-comparison.

We propose the new following formulation: *"The contribution of LM model to the performance of ANATEM is highlighted by the difference of* *performance between ANATEM and ANA models, showed in Fig. 12d. This contribution decreases from south-west to north-east, ranging from 0.06 to 0.04. Conversely, the contribution of ANA model to ANATEM performance (showed in Fig. 12e presenting the difference of performance between ANATEM and LM models), slightly increases from south-west to north-east, ranging from 0.0 to 0.02. The contribution of large scale information (through ANA model) is stronger when LM model (local information) is less efficient, that is, when the location at reconstruction is far from the reference temperature station. "*

P334, L. 23: 0.69 to 0.89

OK

P335, L5: This statement is somewhat confusing, as I would expect the spatial distribution to be dependent on the distance to the Gap meteorological station

Here is the proposed new sentence : *"Conversely, the contribution of ANA to the performance of ANATEM is close to zero for the stations closest to Gap and slightly increases (up to 0.07) with the distance to Gap (Fig. 13e)."*

P336, L9-12: Please define "annual precipitation multiplicative anomaly" plotted in Fig. 15 (0.5 = 150% precipitation depth with respect to the mean value?).

We will slightly modify the paragraph (p. 336 L9-12): *"Figure 15 presents the 1883–2010 annual multiplicative anomaly time-series of precipitation 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 is also given."*

We will also add the following sentence in the caption of Fig. 15 : *"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."*

P350, Fig. 4: In my opinion, the term "observed precipitation" is confusing as these values represent the analogue days (which have been derived from observations).

We propose the following sentence: *"Left panel: observed precipitation at the target site for each of the analogue days as a function of the precipitation estimate from LM for these same days."*

P358, Fig. 12, P359, Fig. 13: These figures are difficult to read. The numbers on the map are too small in my opinion. I would suggest rearranging the panels of both figures and adjust their

size. Would it make sense to create a new figure that includes the panels d and e of Fig. 12 and 13, respectively? You could increase the size of each panel, which would greatly improve readability

It is true that the figure can't be correctly read in the current format, but they have been produced in a portrait layout thinking on the final format of the page(e;g. in the format of the HESS journal). It does therefore not really fit with the current format which is that of the HESS Discussion publications. Don't you think that they would be readable in the final format?

We nevertheless retain you suggestion and will see what the possibilities to improve readability are.

Response to reviewer #2

General comments

The authors would first like to thank Referee #2 for his positive evaluation and for his interesting questions and suggestions. The answers to the general comments are detailed below

It might be important to tell why those 2 geopotential heights were chosen.

For our domain, these geopotential fields were found to be the most informative predictors by Bontron (2004). We can add this sentence p322, line 18.

According to the spatial grid upon which the ANA model is based, it would be worth precising that large-scale information refers here to meso-scale circulation (rather than large synoptic scale).

That is right, we will correct this in the revised version.

p.320 : The authors mention a general, stochastic form of the local model and then state they would only consider uncertainty using the ANA model turned probabilistic by taking 50 analog days instead of the only nearest one. But what would have been the value-added of using a stochastic LM instead of a pure deterministic model ?

We decided not to present the use of a stochastic LM mainly because of the statistical distribution of precipitation residuals which is not straightforward to model. This would have also introduced some additional complexity level in the ANATEM formulation, which we wanted to avoid. We however agree that introducing a stochastic LM is obviously one of the points that would be worth to explore as a perspective of our work. This would however potentially require some other approach for combining ANA and LM estimates. A major advantage of using different combination methods as suggested in the conclusion (p338 l.21-27) would be actually the possibility to combine two probabilistic models. Another advantage of a stochastic LM would be also to that it would allow an extended comparison of the three LM, ANA and ANATEM reconstruction with probabilistic scores (This probabilistic evaluation was for instance carried out for both the ANA and ANATEM approaches in Kuentz 2013).

Upon which criteria was the spatial domain chosen in order to implement the analog model ?

The predictor spatial domain was optimized by maximizing the mean performance of the prediction for a number of precipitation stations over south-eastern France. The performance was estimated from the mean over the simulation period 1953-1993 of the Ranked Probability Score (RPS) (Epstein, 1969; Murphy, 1971). The spatial domain optimization results from the exploration of different growing rectangular analogy domain, as explained by Obled et al. (2002).

I would recommend presenting the ANATEM model for precip of section 3.3.2 another way : as is, it is not clear what the rationale was that eventually lead to such a formu- lation, although the results and mathematical formulation show the model is definitely appropriate for dealing with both low and high values issues. For instance, the Dufour and Garçon (1997) reference is very difficult to obtain whereas it is needed to under- stand how parameters a(k,d) and b(k,d) were defined. I would suggest adding a short description of it.

As explained in our response to referee 1, we acknowledge that the reference, which is an EDF internal report, is not published. We made nevertheless this choice because we already had an experience with this formulation in the field of data assimilation for operational streamflow forecasts. This formulation is used to post-process streamflow forecasts based on the analysis of Rainfall-Runoff model past residuals. Despite its rather empirical nature, the formulation proved to give satisfactory results for this post-processing application. Depending on the current hydrological processes, we may prefer to make a "multiplicative" post-processing of the forecast (typically during drought events) or an "additive" post-processing of the forecast (typically during floods). Due to these two basic properties, we decided to use this formulation for ANATEM, suitable with the problems encountered with rainfall. Another formulation could be obviously tested (as suggested by one of the examiner of Anna Kuentz PhD). Note however that this would not change the principle of the ANATEM combination. We also expect it would not drastically change the conclusions of our work.

a_d^k and b_d^k coefficients are deduced from two conditions proposed by Dufour and Garçon (1997) :

- The slope of the tangent to the curve in $x = 0$ should be $\left(\frac{P_{\mathrm{ANA}_d^k}}{P_{\mathrm{LM,ANA}_d^k}}\right)^2$
- When $P_{\mathrm{ANA}_d^k} = P_{\mathrm{LM,ANA}_d^k}$, the following should be obtained : $\widehat{P}_d^k = P_{\mathrm{LM},d}$

The first condition has been imposed empirically and selected because it gave satisfactory results, while the second condition is logically deduced from the idea of the correction model.

The first condition gives the equality :

$$\frac{a_d^k}{b_d^k} = \left(\frac{P_{\mathrm{ANA}_d^k}}{P_{\mathrm{LM,ANA}_d^k}}\right)^2$$

The second condition gives the equivalence relation :

$$P_{\mathrm{ANA}_d^k} = P_{\mathrm{LM,ANA}_d^k} \Leftrightarrow x_d = \frac{x_d^2 + a_d^k \cdot x_d}{x_d + b_d^k} \Leftrightarrow a_d^k = b_d^k$$

From these two relations the coefficients can be defined as :

$$a_d^k = P_{\mathrm{ANA}_d^k} \quad \text{and} \quad b_d^k = \frac{\left(P_{\mathrm{LM,ANA}_d^k}\right)^2}{P_{\mathrm{ANA}_d^k}}$$

Note that in the paper there are some notation mistakes that will be corrected: the value of the local model for day d is sometimes noted LM_d instead of $P_{\mathrm{LM},d}$.

Do you think it would be worth adding the description of these two conditions in the paper ?

Also, explain eq.11 (just mention it comes from Taylor expansion)

We will add the mention of Taylor expansion in the revised version. You can find the details of the calculations below.

Using the usual first order Taylor expansion $(1+y)^{-1} = 1 + y + o(y)$ when y is close to 0 for the variable $y = \frac{b_d^k}{x_d}$:

$$x_d \cdot \left(1 + \frac{a_d^k}{x_d}\right) \cdot \left(1 + \frac{b_d^k}{x_d}\right)^{-1} \quad \sim \quad x_d \cdot \left(1 + \frac{a_d^k}{x_d}\right) \cdot \left(1 - \frac{b_d^k}{x_d}\right) \text{ when } x_d \to +\infty$$

After expansion,

$$x_d \cdot \left(1 + \frac{a_d^k}{x_d}\right) \cdot \left(1 - \frac{b_d^k}{x_d}\right) = x_d + a_d^k - b_d^k + \frac{a_d^k \cdot b_d^k}{x_d}$$

The last term tends to 0 when x tends to infinity.

Minor comments and suggestions

Thank you for these detailed suggestions and corrections that will be integrated in the revised version.

Figure 10 should have been (and will be) deleted, since it is the same as Figure 9 with a larger scale (but ANA results are not seen in Fig. 9 because they are very poor). This sentence will be added in the caption of figure 9 : "For the annual time step, ANA results are smaller than 0.6; they therefore do not appear on the figure."

Response to reviewer #3

The authors are grateful to Referee #3 for his positive evaluation, his interesting comments and detailed corrections. The answers to the general and specific comments, as well as for a selection of technical comments are detailed below.

General comments

I am writing with some advantage, as two earlier reviews have already been made. Disagreeing with them on one point, I did not find the manuscript particularly well written. Some effort should be placed in order to make it so. Future contributions would benefit enormously from the review of a native speaker.

The revised version of the manuscript will be reviewed by a native speaker.

The presented results are encouraging. Nonetheless, in many examples there is relatively little improvement from the application of a simple linear regression and ANATEM. I believe the biggest advantage of applying ANATEM is not the improved accuracy of mean estimates, but rather the representation of uncertainty it produces. This is not sufficiently emphasized in the text.

We agree with reviewer #3 that there might be a limited improvement of ANATEM, compared to linear regression. However, in our opinion, the other improvement is to give a better spatial robustness of the results. We also agree that another interest of ANATEM is to produce a representation of uncertainties, which is rather difficult concerning precipitation with a linear model (Wu et al., 2011). Another improvement of the modelling of precipitation uncertainty is that it is conditioned by atmospheric circulation patterns.

The following paragraph (p. 323 l.13-15) :
"The originality and the strength of the ANATEM method introduced here is to combine the two previous models and to consequently take advantage of both local and large scale information."

has been modified as :
"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."

This sentence has also been added to the conclusion (after p338 l.2):
"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."

The introduction is interesting and provides a nice overview of the scientific relevance of the work and the challenges associated with it. I would have benefited, however, from a deeper overview of mathematical models and approaches employed to similar ends. ANATEM is solely compared with linear regression (a very simple model) and the analog method (developed in 1969). Ideally, it should be compared with more recent and potentially more performing alternatives (one that instantly comes to mind is non-linear regression). Understanding that this could require a large amount of work, I believe the authors should introduce at least a list of "competing" models.

We agree with reviewer #3 and ANATEM has only been compared to rather simple and classical models (analog method and linear regression).

This paragraph has been added in the introduction (after p.315 l.27): *"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)."*

Although straightforward in hindsight, I found the goals of the work hard to precise at the first stages of the reading. I recommend that a graphical scheme is added to the manuscript in order to facilitate its reading. Also, I believe a simple scheme covering what periods and stations are used in order to calibrate the models, as well as what periods and stations are used in their evaluation would be worthwhile.

As suggested, a graphical scheme (Fig. 1) representing the three methods compared in the study has been added.

Finally, from the introduction and conclusion sections, one is inclined to think the manuscript is focused on the analysis of "long-term" climate records. From the remaining sections it appears the focus is on the proposal and evaluation of the novel ANATEM model that aims at reconstructing (not analyzing) long-term series. I consider the introduction of ANATEM a worthy objective and find the introduction and conclusion sections a bit misleading.

Perhaps they could be adapted in order to increase the value of the paper.

Some changes have been made in the introduction, we hope that the focus of the paper is now easier to understand from it.

Specific comments

As mentioned before, the manuscript could probably benefit from a number of writing corrections. One prevalent issue is the use of the word "we", which I believe should be avoided. In the technical corrections, below, the authors will find some suggestions.

As suggested, the revised version of the manuscript has been reviewed by a native speaker.

There is a fair amount of text which, in essence, is explaining how a simple linear regression works and how it is applied to the problem at hand. This occurs in §3.1 and, again, in §3.3.1. I hope readers will be mostly familiar with such concepts. If the sections could be made shorter, particularly, 3.3.1, it would add to the clarity and flow of the text. (and technical comment "4.13 Page 324 l. 10-25. + page 325 l.1-5 : Review this section. This is too long in order to explain something as well established as a simple linear regression. The equations are also a bit redundant in my opinion.")

We have reduced both paragraphs 3.1 and 3.3.1.

The paragraph 3.1 (p319 l.9-26) has been replaced with :

"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 Tg) from a local neighbour series (referred to as Ne) using a classical linear regression model."

The paragraph 3.3.1 (p324 l.2 to p325 l.8) has been replaced with :

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

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

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

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

The methods are applied using only one neighbouring station. Why not to use more? If more were available a range of interpolation techniques would become available (e.g. Kriging with covariates: KED, co-Kriging, universal Kriging).

We agree with reviewer #3 that more sophisticated interpolation techniques could be interesting when multiple stations were available. In this case study on the Durance watershed, only one station for precipitation and one station for temperature have been found with data available on the whole reconstruction period (1883-2010). This is the reason why the model has been developed with only one neighbouring station. As mentioned in the conclusion (p339 l.2-4), it would be worth to use more if more were available. This was however not the case for our study. The following sentences of the conclusion were probably misleading with this respect.

Extract from the discussion version (p339 l.1-4):

"A thorough sensitivity analysis to the selection 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 to make use of all historical stations available in the close or farer neighbourhood of the region under construction."

We thus modified it. It now reads:

"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 in-

troduction).”

Notation could be revised. Estimates are denoted with a circumflex accent in some parts of the manuscript, yet not throughout. They should be. An example is §3.3.1, line 11 “... $T_{\mathrm{LM},d}$ is the air temperature estimate ...”

According to this suggestion and to the editor comment, mathematical notations of the whole paper have been revised in the final version.

Figures 12 and 13 are too small and hard to read. Also, there is some spelling in French (besides the location names, evidently).

These figures have been modified in the revised version, the two figures have been split into three figures (see new figures attached to this comment).

In the analysis sections, the authors refer the stations by their name, but this can become confusing. To some extent they could make use of the numbers put forward in Figure 1.

That is true, the numbers used in Fig. 1 and Tab. 1 have been added in the text when a reference is made to a station name.

Personally, I find the claims based on the correlation metric are taken too far. For instants, why does a high correlation show that a model captures well both short and long-term variability?

The correlation coefficients have been computed comparing the observed and reconstructed time-series at different aggregation time-steps. The correlation estimated on the daily (resp. annual) series gives information on the capacity of the model to reproduce the high-frequency (resp. low-frequency) variability.

For more clarity, the terms “short-term variability” and “long-term variability” have been replaced by “high-frequency variability” and “low-frequency variability” respectively (e.g. p. 332 l.6).

More importantly, I have doubts regarding the interpretation of figures 12 and 13. Low differences in terms of correlation should hint that the models are very similar. For instants, a low correlation difference between ANATEM and LM, should mean that most of the information contained in ANATEM comes from LM. I am not sure the analyses of the figure’s results – particularly regarding spatial patterns – go in that sense.

The reviewer is right. This is the first result to highlight. We wanted also nevertheless to highlight the noticeable spatial pattern of these contributions, which are highly dependent on the distance to the reference station. We have modified the text accordingly :

Extract from the discussion version (p. 334 l.11-16) :

"The contribution of LM model (Fig. 12d) to the performance of ANATEM, decreases from south-west to north-east, ranging from 0.06 to 0.04. Conversely, the contribution of ANA model (Fig. 12e) to the performance of ANATEM, slightly increases from southwest to north-east, ranging from 0.0 to 0.02. The contribution of large scale information 15 (through ANA model) is stronger when LM model (local information) is less efficient, that is, when the location at reconstruction is far from the reference temperature station."

Modified paragraph :

"The contribution of the LM (resp. ANA) model to the performance of the ANATEM model is presented in Fig. 11d (resp. 11e). 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."

A similar clarification has been added in the paragraph 4.4.2 about precipitation reconstructions.

Extract from the discussion version (p.335 l.1-5) :

"ANATEM increases the global reconstruction performance but it also notably smooths local contrasts. The contribution of LM to the performance of ANATEM decreases as the distance to Gap increases, ranging from 0.22 to 0.02 (Fig. 13d). Conversely, the contribution of ANA to the performance of ANATEM slightly increases from south-west to 5 north-east, ranging from 0.0 to 0.07 (Fig. 13e)."

Modified paragraph:

"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. 12d). 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. 12e)."

Also, the references to plots d) and e) might be switched.

The references are not switched but we agree that the legend was perhaps confusing. It has been changed to:

"Spatial patterns of (d) the contribution of the LM model to ANATEM perfor-

mance (estimated by the difference between the performance of the ANATEM and ANA models) and of (e) the contribution of the ANA model to ANATEM performance (estimated by the difference between the performance of the ANATEM and LM models)"

The conclusion ends with the mention of an application of the ANATEM results to the reconstruction of hydrological long-term series. I can imagine why the authors – having conducted the work – felt inclined to add this to the manuscript. I also believe that, however interesting the topic is, it requires a number of additional considerations which have not, nor should be, addressed in the present contribution. The paper is already valuable due to the introduction of ANATEM – particularly its uncertainty estimation feature. I see no need to close it with a 15-line long reference to another work.

We have drastically reduced this 15-lines reference to this natural extension of the work (see new version in the technical comments). We expect actually to publish it in a fully dedicated publication.

Technical corrections

Most of the technical corrections suggested by referee #3 have been made in the revised version. We answer below to a selection of them that doesn't imply only phrasing or spelling but could also impact understanding.

4.11 Page 322. 24. Please clarify what a moving seasonal filter is. What was the window size, etc.

The clarification is given in the following sentence. The point has been replaced by a column to evidence the link between the two sentences. We also replaced "a moving seasonal filter" by "a moving calendar filter".

4.12 Page 323. 16-25. An overall confusing paragraph. Should be rephrased with an emphasis on clarity.

Paragraph from the discussion version:

"The principle of ANATEM is the following: the local variable reconstructed for the target day d is the local variable estimate obtained by the local model, corrected by the errors of the Local Model identified when it is applied for the prediction of the local variable on the n analogs days. In other words, for any target day, the Analog Model ANA allows the identification of n analog days in terms of atmospheric circulation (see Sect. 3.2). The n prediction errors respectively obtained when the Local Model LM is used for predicting the local observed value for each of these n days are used to define the error distribution associated to the prediction obtained with the Local Model for the

target day d. The prediction obtained with ANATEM for the target day is therefore probabilistic."

New modified paragraph:

"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 predictions 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."

4.20 Page 331. 1-12. Please rephrase the section.

Paragraph from the discussion version:

"For the sake of conciseness, we consider here for the evaluation only one reference time-series for each model. For the Local Model, this is simply the reconstruction obtained with the model. For the probabilistic reconstruction models ANA and ANATEM, this is the mean time-series derived from the ensemble of 50 time-series reconstructions (the daily value for a given day is the mean of the probabilistic reconstruction for this day). As it will be noticed later, these mean time series obviously present a much lower variability than each time series of the reconstruction ensemble. For the sake of simplicity, these mean time series will be also referred to as reconstructed time series. In the following, the performance of a given model will be presented with the distributions of r, α, β and KGE criteria obtained for the 22 watersheds at the daily, monthly and annual time-steps."

New modified paragraph:

"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, α, β and KGE criteria obtained for the 22 watersheds at the daily, monthly and annual time steps."

4.23 Page 334. 24-25 Rephrase.

Paragraph from the discussion version:

"However, the similarity in terms of large scale forcing influences probably influences the performance. Hence, two watersheds at the same distance to the Gap station have rather different performances (i.e. Buech watershed have a very good correlation of 0.88 and the Durance at Briançon a moderate correlation of 0.77)."

New proposed paragraph:

"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. 1 – have a very good correlation of 0.88 and the reconstructions for the Durance at Briançon watershed – #3 in Fig. 1 – 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."

4.24 Page 335. 18-20. What does the sentence inform the reader about really? The variability is even larger from year to year... By the other hand, the periods referred to are not related to long-term trends...

Sentence in the discussion version:

"This series (red curve in Fig. 14) highlights a relatively strong variability: mean air temperature can vary of nearly 1°C in less than 10 years (e.g.: 1890–1900, 1940–1950)."

This sentence has been removed.

4.25 Page 336. 18-20. Please clarify what is meant, give values, and check phrasing.

Paragraph in the discussion version:

"ANATEM series present a very homogeneous temporal behaviour when compared to the high dispersion observed between the five HISTALP series. This may be partly explained by the fact that ANATEM series are reconstructed for all watersheds based on a same reference series (Gap). The main reason is however probably the high spatial variability of precipitation and the fact that HISTALP series cover a much wider spatial domain than ANATEM series. The low dispersion between the reconstructed series is otherwise coherent with the limited dispersion obtained between time-series observed for the same 22 watersheds on the observation period (not shown here)."

New modified paragraph:

"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."

4.28 Page 339. 7-10. Could be improved. Please rephrase.

Paragraph in the discussion version:

"The region we have considered covers a rather narrow domain. However, we can expect that the interest of the reference station is much lower if we would do reconstructions for much more distant locations. We can expect conversely that the relative interest of the large scale information would be much larger for distant sites."

New proposed paragraph:

"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."

20-end. Too long and, in my opinion, off topic and not needed to value the paper.

Paragraph in the discussion version:

"A major application of such reconstructions will be obviously the possibility to reconstruct long term variations for a number of climate driven variables. As an illustration, the long-term climatic time-series produced in the present work have been used for reconstructing long-term hydrological time-series at multiple hydrometric stations of the Durance basin (Kuentz, 2013; Mathevet et al., 2013).
(line 20) An outstanding result of this reconstruction is that the time series obtained for the whole 20th century present a very high correlation level with historical discharges time series obtained from rescued hydrometric archives for the catchment. In our case, the availability of historical streamflow time series allowed us to demonstrate the overall quality of the meteorological reconstruction. This independent hydrological validation is not expected to be feasible everywhere but it gives high confidence in this hydrometeorological reconstruction approach. Even when such an independent validation cannot be carried out, the reconstructed time series definitively produce a high-value information for researchers or water resources managers. Further works for other hydroclimatic contexts are therefore also worth to better identify the potential of the method and the possibility for improving it."

The end from (line 20) has been replaced by:

"Thanks to the availability of long observed discharge series, this study provided an independent hydrological validation of the climatic reconstructions

over the entire 20th century."

Bibliography

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.

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, URL http://www.sciencedirect.com/science/article/pii/S0022169414006350, 2014.

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, URL http://www.sciencedirect.com/science/article/pii/S0022169407002557, 2007.

Deser, C., Phillips, A., Bourdette, V., and Teng, H.: Uncertainty in climate change projections: the role of internal variability, Climate Dynamics, 38, 527–546, doi:10.1007/s00382-010-0977-x, URL http://link.springer.com/article/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.

Epstein, E. S.: A Scoring System for Probability Forecasts of Ranked Categories, Journal of Applied Meteorology, 8, 985–987, doi:10.1175/1520-0450(1969)008<0985:ASSFPF>2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0450(1969)008%3C0985:ASSFPF%3E2.0.CO%3B2, 1969.

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, Journal of Hydrology, 377, 80–91, doi:10.1016/j.jhydrol.2009.08.003, URL http://www.sciencedirect.com/science/article/pii/S0022169409004843, 2009.

Hawkins, E. and Sutton, R.: The Potential to Narrow Uncertainty in Regional Climate Predictions, Bulletin of the American Meteorological Society, 90, 1095–1107, doi:10.1175/2009BAMS2607.1, URL http://journals.ametsoc.org/doi/abs/10.1175/2009BAMS2607.1, 2009.

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.

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 Resources Research, 50, 3317–3341, doi:10.1002/2013WR014897, URL http://onlinelibrary.wiley.com/doi/10.1002/2013WR014897/abstract, 2014.

Marty, R., Zin, I., and Obled, C.: On adapting PQPFs to fit hydrological needs: the case of flash flood forecasting, Atmospheric Science Letters, 9, 73–79, doi:10.1002/asl.176, URL http://onlinelibrary.wiley.com/doi/10.1002/asl.176/abstract, 2008.

Murphy, A. H.: A Note on the Ranked Probability Score, Journal of Applied Meteorology, 10, 155–156, doi:10.1175/1520-0450(1971)010<0155:ANOTRP>2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0450(1971)010%3C0155:ANOTRP%3E2.0.CO%3B2, 1971.

Obled, C., Bontron, G., and Garçon, R.: Quantitative precipitation forecasts: a statistical adaptation of model outputs through an analogues sorting approach, Atmospheric Research, 63, 303–324, doi:10.1016/S0169-8095(02)00038-8, URL http://www.sciencedirect.com/science/article/pii/S0169809502000388, 2002.

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, URL http://dx.doi.org/10.1080/02626667.2012.665994, 2012.

Terray, L. and Boé, J.: Quantifying 21st-century France climate change and related uncertainties, Comptes Rendus Geoscience, 345, 136–149, doi:10.1016/j.crte.2013.02.003, 2013.

Wu, L., Seo, D.-J., Demargne, J., Brown, J. D., Cong, S., and Schaake, J.: Generation of ensemble precipitation forecast from single-valued quantitative precipitation forecast for hydrologic ensemble prediction, Journal of Hydrology, 399, 281–298, doi:10.1016/j.jhydrol.2011.01.013, URL http://www.sciencedirect.com/science/article/pii/S0022169411000382, 2011.

Manuscript version marked-up with all changes

Manuscript prepared for Hydrol. Earth Syst. Sci. Discuss.
with version 2014/09/16 7.15 Copernicus papers of the LaTeX class copernicus.cls.
Date: 10 May 2015

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

A. Kuentz[1,2,*], **T. Mathevet**[1], **B. Hingray**[3,4], **and J. Gailhard**[1]

[1]Électricité de France, DTG, BP 41, 38040 Grenoble CEDEX 09, France
[2]Irstea, Hydrosystems and Bioprocesses Research Unit (HBAN), 1 rue Pierre-Gilles de Gennes, CS 10030, 92761 Antony CEDEX, France
[3]CNRS, LTHE UMR 5564, Grenoble, F-38000, France
[4]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

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

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

1 Introduction

~~As highlighted by the even larger number of publications in the recent decades, estimating the hydrological impacts of climate change is a key societal requirement for relevant planning and adaptation. It is however difficult because of the numerous sources of uncertainty associated to climate projections. They are related to emission scenarios, models and also to the internal variability of the climate system,~~ Multi-decadal variations of climate variables, intrinsically arising from ~~its~~ the chaotic and non-linear nature ~~. Internal variability, leading to multi-scale variations – from multi-year to multi-decadal scales, has~~ the climate system, have long been observed for a number of large ~~but also~~ as well as local scale climate features (Madden, 1976).

In a non-stationary climate, multi-decadal variations can remain ~~high above~~ substantially above or below the long-term trend. In climate projections for the coming decades, they often ~~represent a major source of~~ lead to large uncertainties (e.g. Hawkins and Sutton, 2009; Deser et al., 2012). For precipitation or hydrometeorological variables such as streamflow, ~~related uncertainty can be as large or even larger than~~ 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 ~~do not account for this uncertainty source. As an illustration~~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 ~~200runoff time-series~~ 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 ~~with the~~ on the basis of such a study may be questionable. ~~To~~ In our opinion, they ~~at least~~ suffer from a certain lack of large historical perspective, which would at least require characterising the multi-scale variability of climate variables.

~~Characterizing~~ Today, characterising the multi-scale variability of climate variables ~~appears today to be necessary~~ would appear to be important (if not mandatory) ~~to put into perspective~~ in order to put future climate projections into perspective. Numerous studies worldwide have investigated past variability of climate and related variables. ~~For~~ 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), ~~west North-America (Rood et al., 2005) ,~~ 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. ~~As~~ However, given that the density of observations ~~is~~ was significantly lower before 1960 (Hannah et al., 2011), most ~~time-series however~~ time series usually cover a few decades only, which is obviously not sufficient for a relevant analysis of ~~multidecadal~~ multi-decadal variations (Mathevet and Garçon, 2010; Hannaford et al., 2013). Long-term historical ~~time-series~~ time series (covering a period longer than 100 years) are of course the ideal material for ~~this~~ such an analysis. Such historical series ~~were for instance~~ have been used for the Loire river in France (Renard, 2006), the ~~Colombia~~ Columbia and Missouri rivers in the USA (Rood et al., 2005), the Murray-Darling basin in Australia (CSIRO, 2010) ~~or more recently~~and, more recently, for a larger ~~panel~~ set of French stations by Boé and Habets (2013). Long-term streamflow ~~time-series are obviously rather~~ time series are rare, with a ~~ ~~ typically very low spatial density. Some could still be ~~rescued~~ recovered from various national and regional ~~archives~~ archive sources but the ~~rescue~~ recovery process is long and requires demanding ~~digitizing~~ 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.)~~preventing~~ , hindering the use of ~~series for the~~ some series for variability analysis.

~~Characterizing the long-term~~ Characterising the low-frequency variability of climate and related variables from observations is therefore ~~usually not~~ 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 ~~can~~

use environmental markers ~~like for instance~~ such as tree rings (Frank and Esper, 2005) ~~,~~ or lake sediments (Wilhelm et al., 2013, 2012), narrative ~~evidences~~ evidence of droughts (Pfister et al., 2006) ~~,~~ or geochemical tracers in ice ~~core~~ cores from glaciers (Jouzel et al., 2007). ~~For the reconstruction of past streamflows variations,~~ Simulations provide an efficient way ~~is simulation, where simulated discharges~~ to reconstruct past flow variations. Simulated discharge times series are obtained ~~with~~ using a hydrological model ~~from~~ forced with past variations of meteorological variables available for the region. ~~In some particular cases,~~ When meteorological observations required for such ~~an analysis may cover a much longer time period than the period for which hydrological data are available. They obviously suffer, in most cases, from the same scarcity and length limitations than hydrological data. In such a case, meteorological data~~ 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~~ 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 ~~for the~~ covering the entire 20th century (from 1871 ~~to present year~~ for the NOAA 20CR, Compo et al., 2011 ~~,~~ and from 1900 for the ECMWF ERA-20C, Poli et al., 2013) ~~present~~ provides a great opportunity for such a reconstruction. ~~As~~ Unfortunately, their spatial and temporal resolutions ~~do unfortunately~~ rarely fit the resolution ~~standards~~ needs (typically sub-daily time step, up to 1000 ~~km^2) required for hydrological applications, the reconstruction of~~ km2) of hydrological applications. In such a case, the required local meteorological data ~~is~~ can be obtained through downscaling.

This study compares 3 different statistical approaches for the reconstruction of ~~high resolution~~ high-resolution precipitation and temperature data. Reconstructions are respectively obtained from observations available at a neighbouring station, from ~~large scale~~

large-scale (meso-scale) atmospheric variables extracted from the 20CR ~~reanalyses, and from both data at a time. If the two first approaches have been already applied for~~ reanalysis and from a combination of both. Although the first two approaches have already been applied in similar studies, the last ~~one is original, as it~~ is an original approach in that makes use of both local observations and ~~large scale atmospheric informationsimultaneously.~~ 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 ~~scale~~ step for the 22 subcatchments of the Upper Durance River basin, a mesoscale catchment located in the ~~South-Eastern~~ south-eastern Alps. They ~~were produced for further hydrological reconstructions for~~ 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 ~~. They are~~ and evaluated and compared in Sect. 4. Section 5 ~~shortly discusses the long-term~~ briefly discusses the low-frequency climatic variability reconstructed over the 1870–2010 period~~with the models~~. Finally, conclusions and perspectives ~~of~~ 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~~ south-eastern France (Fig. 2). The main characteristics of the watersheds are detailed in Table 1.

~~Limited in the North~~ Bounded in the north by the Écrins ~~Alpine massif~~ mountain range of the Alps and in the ~~South~~ south by the Mediterranean Sea, the various ~~sub-watersheds~~

~~highlight very different climate~~subcatchments display very different climates. Upstream hydrological regimes are ~~snow dominated~~ 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 ~~this~~ that period.

~~Daily~~ For each watershed, daily mean areal air temperature and precipitation ~~have been estimated for each watershed~~ 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 "~~observed~~ observation period" and the SPAZM series will be referred to as "observations" ~~, although it is not direct observations but~~ 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 ~~Local~~ Long local reference ~~long~~ series

To reconstruct the mean areal air temperatures and precipitation of the 22 watersheds, ~~we searched~~ 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 ~~four~~ 4 air temperature and ~~forty~~ 40 precipitation measurement stations in the watershed and its ~~neighborhood~~neighbourhood. Unfortunately, most ~~data~~ of the data from these stations have been lost ~~. Very~~ and only very few and incomplete series ~~are still available nowadays~~remain available today. For precipitation, ~~we were able~~ it was possible to rebuild a 1883–2010 series for the Gap location by merging two sources of data, provided respectively by Electricité de France (EDF) and Météo-France. For air temperature, the nearest daily series found, provided by Météo-France, was ~~found in~~ for Marseille and covers the period 1868–2010.

For a qualitative assessment of the reconstructed series, ~~we also considered~~ 5 monthly time series from the HISTALP project database (monthly series, Auer et al., 2007) ~~. The series also started in~~ were also used. They also go back to the 1870's. For air temperature, the ~~corresponding~~ selected stations are located around the ~~South-Eastern~~ south-eastern part of the Alps, ~~in~~ at Genova University, Milano-Brera, Montpellier, Nice airport and Nîmes airport. For precipitation, they are ~~quite~~ 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~~ Large-scale climatic data

~~Large scale atmospheric data are~~ 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 ~~by~~ the Climate Program Office of the National Oceanic and Atmospheric Administration (NOAA). This reanalysis ~~has been~~ was produced by assimilating only sea level pressure data, ~~which allows it starting as soon as~~ making it possible to go back to the end of the 19th century. The reanalysis covers the 1871–2010 period.

In the present work, ~~large scale~~ the large-scale variables used for the reconstruction are the fields of ~~geopotential heights at~~ 700 and 1000 hPa geopotential heights in the rectangular spatial domain situated between ~~the points (~~longitudes 8° ~~O, 38~~W and 12° ~~N) and (12~~E and latitudes 38° ~~E,~~ N and 50° N~~).~~ .

3 Methodology: ~~mixing~~ combining two sources of information

In climatology ~~or~~ and hydrology, the reconstruction of past climatic data is usually necessary ~~, either~~ to estimate missing values, assess data quality or build ~~long term~~ long-term climatic reanalyses. Different methods are classically used to reconstruct climatic observations. Some of them are ~~only~~ solely based on the series ~~at reconstruction itself~~ 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 ~~either be local or regional scale observations and either have the same or a different nature of the reconstructed series~~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, ~~we present~~ the three methods used for the reconstruction are presented. The first one uses local neighbour ~~observation~~ observations of a similar proxy (respectively, air temperature or precipitation observation). The second is basically a downscaling approach using regional ~~large scale~~ large-scale information of a different proxy (geopotential fields). The third ~~one~~ approach uses both proxies~~at a time~~.

As in most reconstruction works, ~~the~~ these methods rely on a period ~~on which both proxy data and data~~ over which both data at the reconstruction point ~~are available. This concomitant~~ and proxy information are available (see Fig. 1). This period will be referred to as the observation period. The reconstruction period is the period ~~on~~ over which the reconstruction model is applied~~: it corresponds~~ , corresponding to the period where proxy ~~data are available while~~ 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~~ regression-like models, where predictors ~~are local observed data~~ should be well correlated with the data ~~at reconstruction~~to be reconstructed. This model is calibrated against observations ~~on their common period of availability (observation period) and the series at reconstruction is either filled or extended by the modelled series.~~ during the observation period.

In the following, the ~~series to be reconstructed will be called target series (Tg), and the local series used as reference will be called neighbour series (Ne) , although the neighbourhood of these two series is variable in space or time and concretely depends on the availability of other local observed series.~~

~~In this paper,~~ principle of the local model (LM) is to reconstruct the target series ~~, a local observed neighbour series will be used through a simple~~ (referred to as Tg) from a local neighbour series (referred to as Ne) using a classical linear regression model~~, called local model (LM). The estimate $\widehat{X}_{\mathrm{LM},d}$ of the target variable X obtained with LM for a given day d~~

~~has the classical following expression:~~

$$\text{LM}: X_{\text{LM},d} = \alpha \cdot X_{\text{Ne},d} + \beta + \varepsilon_d$$

~~where $X_{\text{Ne},d}$ is the value of the neighbour series for the day d, α is a multiplicative correction factor, β is an additive correction factor and ε_d is a residual assumed to have zero mean. Depending on the nature of the reconstructed series (air temperature or precipitation), the correction factor is either only multiplicative (i.e. $\beta = 0$) or additive (i.e. $\alpha = 1$).~~ .

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 ~~difference of altitude~~ altitude difference between the target and neighbour series. However, even when the target and neighbour series are very well correlated, residuals of such ~~models usually exhibits~~ a model usually exhibit a strong seasonal pattern. ~~Then,~~ In this case, the LM model can be slightly improved ~~assuming that the additive correction~~ 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~~ local model for air temperature reconstruction can thus be written as:

$$\text{LM}: \cancel{T_{\text{LM},d}}\widehat{T}_{\text{LM}}(d) = T_{\cancel{\text{Ne},d}\text{Ne}}(d) + \beta_{\cancel{d}}(d) + \varepsilon_{\cancel{d}}(d) \tag{1}$$

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

In ~~this paper, we chose to use this model~~ 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 ~~more adequate for precipitation and~~ compatible with the asymmetrical distribution of precipitation values (~~no negativevalues~~never negative). The correction factor is taken ~~constant all over~~ to be constant throughout the year. The improvement ~~brought by~~ obtained by using a variable correction has ~~nevertheless~~ been assessed and shown ~~as~~ to be negligible (Kuentz, 2013). The constant multiplicative correction factor is calibrated over the ~~common data availability~~ observation period as:

$$\alpha = \frac{\cancel{\overline{P_{\mathrm{Tg}}}}\,\overline{P}_{\mathrm{Tg}}}{\cancel{\overline{P_{\mathrm{Ne}}}}\,\overline{P}_{\mathrm{Ne}}} \tag{2}$$

where $\cancel{\overline{P_{\mathrm{Ne}}}}\ \overline{P}_{\mathrm{Ne}}$ is the mean value of the neighbour series and $\cancel{\overline{P_{\mathrm{Tg}}}}\ \overline{P}_{\mathrm{Tg}}$ is the mean value of the target series, both calculated ~~on the common data availability~~ over the observation period.

The ~~Local Model~~ local model used for precipitation reconstruction reads:

$$\mathrm{LM}_{\cancel{d}} : \cancel{P_{\mathrm{LM},d}}\,\widehat{P}_{\mathrm{LM}}(d) = \alpha \cdot\ P_{\cancel{\mathrm{Ne},d}\,\mathrm{Ne}}(d) + \varepsilon_{\cancel{d}}(d) \tag{3}$$

where $\cancel{P_{\mathrm{LM},d}}\ \widehat{P}_{\mathrm{LM}}(d)$ is the estimate of the target precipitation for ~~the~~ day d, $\cancel{P_{\mathrm{Ne},d}}\ P_{\mathrm{Ne}}(d)$ is the value of the neighbour series precipitation for this same day and $\cancel{\varepsilon_d}\ \varepsilon(d)$ is a residual with zero mean. ~~As explained previously for the air temperature reconstruction~~Once again, a simple version of this model with a residual term ~~considered~~ equal to zero is used in ~~this paper, as uncertainty will be taken into account in another way~~the present work.

3.2 ~~Large scale~~ Large-scale climatic information: the ~~Analog~~ analogue method

The second reconstruction ~~method is the analog~~ model is based on the analogue method introduced by Lorenz (1969). ~~The method is currently~~ 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 ~~over the past as made~~ as done by Timbal et al. (2006). Nevertheless, the release of the long atmospheric reanalyses for the 20th century opens ~~doors for more development of~~ the door to more such uses, allowing ~~for~~ the reconstruction of long climatic series covering the entire 20th century.

The ~~analog~~ analogue method is based on the fact that local meteorological variables are strongly influenced by the state of the atmosphere and its ~~circulation at the synoptic scale. As long as a long enough~~ meso-scale circulation. Provided that a sufficiently long archive with concomitant local and ~~large scale~~ large-scale observations is available, it is ~~then~~ therefore possible to produce local meteorological scenarios for any other day for which the required ~~large scale~~ large-scale atmospheric predictors are available. For this, the ~~k~~ 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 ~~analog days are next used as~~ analogue days are then used as the weather scenario for the target day.

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

The ~~Analog Model~~ analogue method has some parameters to be set such as the type ~~or the~~ and level of predictors, the number of ~~analog~~ analogue days selected for the prediction, the spatial domain used to compute the similarity criterion or the similarity criterion itself. Numerous variations of the ~~analog~~ analogue method have been developed. In the present work, ~~we use the Analog Model~~ the analogue model (ANA) presented by Obled et al. (2002) and further explored by Bontron and Obled (2005), Ben Daoud et al. (2010)~~and~~

~~Horton et al. (2012)~~ , Horton et al. (2012) and Chardon et al. (2014) is used. Its main features are presented below.

- *The predictors* are the ~~geopotential fields at~~ 700 and 1000 hPa ~~for the times 0~~ geopotential height fields at times 00:00 and 24~~h~~: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 ~~have~~ 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 ~~the~~ longitudes 8° W and 12° E and ~~the~~ latitudes 38 and 50° N, with a step of 2°.

- *A moving ~~seasonal~~ calendar filter* is used for the determination of candidate ~~analog days. For~~ analogue days: for each target day, candidate ~~analog days are days which calendar day is~~ analogue days are the days included in a ~~60days~~ 60-day interval around the calendar day of the target.

The reconstruction is deterministic when only one ~~analog~~ analogue is used (classically the nearest ~~analog~~analogue). The ~~analog~~ analogue day can be also selected among the n nearest ~~analogs. As a result, an~~ analogues. An ensemble of reconstructions can be produced ~~. This allows characterizing~~ 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 ~~in the following $[X_{\mathrm{ANA}_d^k}]_{k=1\ldots n}$ where $k = 1\ldots n$~~ as $[X(d_k)]_{k=1..n}$ where $[d_k]_{k=1..n}$ refers to the n nearest ~~analogs selected for the~~ analogue days selected for day d. In the present case, the ~~selection is done among~~ ensemble of reconstructions is obtained from the 50 nearest ~~analogs~~ analogues ($n = 50$).

3.3 Mixing formulation: the ANATEM ~~method~~model

Both local and ~~large scale~~ large-scale predictors are available for the 1870–2010 period. The ~~Local Model~~ local model (LM~~and the Analog Model~~) and the analogue model (ANA~~can be therefore~~) 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~~ large-scale atmospheric information (~~synoptic~~ meso-scale variables). The originality and ~~the~~ strength of the ANATEM ~~method introduced here is to combine~~ model introduced here lies in an approach that combines the two previous models~~and to therefore~~. In this way, it can take advantage of both local and ~~synoptic information~~ large scale information and produce an original representation of uncertainties, conditioned by atmospheric circulation patterns.

The principle of ANATEM is the following: ~~the local variable reconstructed for the target day d is the local variable estimate obtained by the local model, corrected by the errors of the Local Model identified when it is applied for the prediction of the local variable on the n analogs days. In other words, for~~ for any target day, the ~~Analog Model ANA~~ analogue model allows the identification of n ~~analog~~ analogue days in terms of atmospheric circulation (see ~~Sect.~~Section 3.2). The ~~n prediction errors respectively obtained when the Local Model LM is used for predicting the local observed value~~ 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~~are~~, allowing the calculation of n predictions errors. These n error values are finally used to define the error distribution associated ~~to~~ with the prediction obtained with the ~~Local Model~~ 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

~~Let ANA$_d^k$ be the $k$thanalog day selected~~ The probabilistic air temperature prediction from the ANATEM model for day d has the following expression:

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

where $\left[\widehat{T}^k_{\text{ANATEM}}(d)\right]_{k=1...n}$ is the ensemble of reconstructed values for the target day ~~–~~ d ~~. Let's write $T_{\text{ANA}_d^k}$ the observed air temperature for this day, and $T_{\text{LM,ANA}_d^k}$ the~~ (ANATEM stands for "Combined Model" and refers to the ANATEM model), $\widehat{T}_{\text{LM}}(d)$ the air temperature estimate ~~that would have been obtained with the local model~~ obtained with LM ~~when applied for the prediction of temperature for this same~~ for target day d, d_k the k~~thanalog day .~~

~~The ANATEM method assumes that the error made by the local model LM for the~~ th analogue day selected for target day d~~could be the same as the error made by this same model for the day ANA$_d^k$. In such a case, the estimate of the local temperature $\widehat{T}_d^k$ obtained for day d through ANATEM using the~~ , $T(d_k)$ the observed air temperature for this k~~thanalog day would read:~~

$$\cancel{\widehat{T}_d^k = T_{\text{LM},d} + T_{\text{ANA}_d^k} - T_{\text{LM,ANA}_d^k}}$$

~~where $T_{\text{LM},d}$ is~~ th analogue day and $\widehat{T}_{\text{LM}}(d_k)$ the air temperature estimate ~~that would have been obtained with~~ obtained with the local model (LM~~for the target day~~) for the same day d_k.

~~According to Eq. (5), the reconstructed temperature $\widehat{T}_d^k$~~ 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~~thanalog day can be written as a linear function of the value of $T_{\text{LM},d}$ as follows:~~

$$\cancel{\widehat{T}_d^k = f(x_d) = x_d + a_d^k}$$

~~with $x_d = T_{\mathsf{LM},d}$ and $a_d^k = [T_{\mathsf{ANA}_d^k} - T_{\mathsf{LM,ANA}_d^k}]$~~$^{\text{th}}$ analogue day d_k.

~~Thus, the value of $\widehat{T}_d^k$ can be read in the~~ The statistical dressing of the LM prediction for the target day d can be simply represented on a graph in a (T_{LM},T) space~~as the ordinate of the point with abscissa $T_{\mathsf{LM},d}$ in the line which slope is 1 and intercept is $T_{\mathsf{ANA}_d^k} - T_{\mathsf{LM,ANA}_d^k}$. Such representation is~~ , as shown in Fig. ~~3 (left): each analog day selected for the day d is represented by a black point in the (T_{LM},T) space, and the line described above is plotted in red. The vertical distance between this line and the first bisector plotted in green represents the error a_d^k made by the local model for the analog day ANA_d^k, which is transposed to the target day to obtain the ANATEM estimation.~~

~~By repeating this operation for each analog day, a set of parallel lines is obtained as shown in Fig. 3 (right). The ensemble of reconstructed temperatures for day d is the local model estimate $T_{\mathsf{LM},d}$ corrected by the ensemble of errors made by LM over the~~ 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 ~~analog days. It can be represented by the ensemble of points with abscissa $T_{\mathsf{LM},d}$ in each of the n parallel linesplotted in Fig. 3 (right). The distribution of these reconstructed values $[\widehat{T}_d^k]_{k=1...n}$ is represented by a boxplot (10~~analogue days. Each cross is simply the intercept of two lines: the vertical line at the $\widehat{T}_{\mathsf{LM}}(d)$ value on the x axis and the 1:1 line passing through the point ($\widehat{T}_{\mathsf{LM}}(d_k)$, ~~25, 50, 75 and 90quantiles)on the right part of the figure, and the mean value is shown by a blue point~~$T(d_k)$). This is illustrated for a given analogue day in Fig. 3 (left).

~~Finally, the probabilistic prediction from the ANATEM method for day d has the following expression:~~

$$\left[\widehat{T}_d^k\right]_{k=1...n} = T_{\mathsf{LM},d} + \left[T_{\mathsf{ANA}_d^k} - T_{\mathsf{LM,ANA}_d^k}\right]_{k=1...n}.$$

For the example shown in Fig. 3, over the n ~~analog days having a similar synoptic situation than~~ analogue days with meso-scale situations similar to that of day d, the local model estimate T_{LM} was ~~in~~ on the average higher than the observed temperature at the target point. ~~This lead to negatively correct the model for the~~ 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 $-$~~9.0~~9.8°C, the ~~fifty~~ 50 air temperature values produced by the ANATEM ~~method~~ model have a mean of -11.2°C and their 10 and 90% quantiles are respectively -13.1 and -9.3°C.

3.3.2 Precipitation reconstruction

~~The ANATEM method~~ Although the ANATEM model uses the same basic principle for precipitation reconstruction~~. Another formulation was however proposed , due to~~ , a somewhat different formulation is proposed to account for the specific features of ~~precipitations~~ precipitation (asymmetric distribution ~~, lot of~~ and many zero values).

The additive correction formulation used for the probabilistic reconstruction of temperature (Eq. 5) is not suitable ~~for precipitation~~here. It can actually produce negative values as illustrated in Fig. 4 (left), elaborated ~~on~~ following the same principle ~~than~~ as explained in Fig. 3 (right).

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

$$\left[\widehat{P}^k{}_{\cancel{d}\,\mathsf{MULT}}(d)\right]_{k=1\ldots n} = \cancel{P_{\mathsf{LM},d}}\,\widehat{P}_{\mathsf{LM}}(d)\cdot\left[\frac{P_{\mathsf{ANA}_d^k}}{\cancel{P_{\mathsf{LM,ANA}_d^k}}}\,\frac{P(d_k)}{\widehat{P}_{LM}(d_k)}\right]_{k=1\ldots n}. \tag{5}$$

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 ~~seem~~ appear to be reasonable for common values, but ~~the reconstruction can produce unreasonable high values of precipitation~~can be unreasonably high in certain cases.

In the following, ~~we have therefore chosen to build~~ the probabilistic reconstruction of precipitation has therefore been built with a correction model ~~that has~~ intended to have a multiplicative behaviour for low values of ~~LM_d~~ $\widehat{P}_{\text{LM}}(d)$ and an additive behaviour for high values of ~~LM_d~~ $\widehat{P}_{\text{LM}}(d)$. Its analytical formulation and its asymptotic ~~behaviours when x_d~~ behaviour when x tends to zero or infinity (through a Taylor expansion) have the following expressions:

$$\widehat{P}^k{}_{\cancel{d}\text{ANATEM}}(d) = f(x_{\cancel{d}}(d)) = \cancel{\frac{x_d^2 + a_d^k \cdot x_d}{x_d + b_d^k}} \frac{x(d)^2 + a(d_k)\cdot x(d)}{x(d) + b(d_k)} \tag{6}$$

where ~~$x_d = \text{LM}_d$ and where a_d^k and b_d^k~~ $x(d) = \widehat{P}_{\text{LM}}(d)$ and $a(d_k)$ and $b(d_k)$ are parameters to be expressed ~~in function of $P_{\text{ANA}_d^k}$ and $\text{LM}_{\text{ANA}_d^k}$.~~

$$\cancel{\widehat{P}_d^k = \frac{x_d^2 + a_d^k \cdot x_d}{x_d + b_d^k} = x_d \cdot \frac{x_d + a_d^k}{x_d + b_d^k} \underset{x_d \to 0}{\sim} x_d \cdot \frac{a_d^k}{b_d^k}}$$

$$\cancel{\widehat{P}_d^k = \frac{x_d^2 + a_d^k \cdot x_d}{x_d + b_d^k} = x_d \cdot \left(1 + \frac{a_d^k}{x_d}\right) \cdot \left(1 + \frac{b_d^k}{x_d}\right)^{-1} \underset{x_d \to +}{\sim}}$$

~~We define the~~ 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 $x(d)$.

$$\widehat{P}^k_{\text{ANATEM}}(d) = \frac{x^2 + a\cdot x}{x + b} = x \cdot \frac{x + a}{x + b} \underset{x \to 0}{\sim} x \cdot \frac{a}{b} \tag{7}$$

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

The two model parameters ~~a_d^k and b_d^k following~~ 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 ~~expressed~~ defined as a function of ~~$P_{\text{ANA}_d^k}$ and $P_{\text{LM,ANA}_d^k}$~~ $P(d_k)$ 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 $\widehat{P}^k_{\mathrm{ANATEM}}(d) = \widehat{P}_{\mathrm{LM}}(d)$ must be obtained.

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

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

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:

$$\left[\widehat{P}^k{}_{d\,\mathrm{ANATEM}}(d)\right]_{k=1\ldots n} = \left[\frac{{P_{\mathrm{LM},d}}^2 + P_{\mathrm{ANA}^k_d}\cdot P_{\mathrm{LM},d}}{P_{\mathrm{LM},d} + \frac{{P_{\mathrm{LM,ANA}^k_d}}^2}{P_{\mathrm{ANA}^k_d}}} \frac{\widehat{P}_{\mathrm{LM}}(d)^2 + P(d_k)\cdot \widehat{P}_{\mathrm{LM}}(d)}{\widehat{P}_{\mathrm{LM}}(d) + \frac{\widehat{P}_{LM}(d_k)^2}{P(d_k)}}\right]_{k=1\ldots n}. \tag{10}$$

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

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

– if ~~$P_{\mathrm{ANA}_d^k} > 10 \cdot P_{\mathrm{LM,ANA}_d^k}$~~ $P(d_k) > 10 \cdot \widehat{P}_{LM}(d_k)$ then the value of ~~$P_{\mathrm{ANA}_d^k}$~~ $P(d_k)$ = ~~$10 \cdot P_{\mathrm{LM,ANA}_d^k}$~~ $10 \cdot \widehat{P}_{LM}(d_k)$, and

– if ~~$P_{\mathrm{ANA}_d^k} < \frac{1}{10} \cdot P_{\mathrm{LM,ANA}_d^k}$~~ $P(d_k) < \frac{1}{10} \cdot \widehat{P}_{LM}(d_k)$ then the value of ~~$P_{\mathrm{ANA}_d^k}$~~ $P(d_k)$ = ~~$\frac{1}{10} \cdot P_{\mathrm{LM,ANA}_d^k}$~~ $\frac{1}{10} \cdot \widehat{P}_{LM}(d_k)$.

The filtering threshold (10) has been chosen ~~arbitrary~~arbitrarily. A sensitivity analysis with different values from 2 to 100 showed that this threshold has ~~fairly no impact on the reconstruction performances~~ little impact on reconstruction performance. This is true because very few ~~analog~~ analogue days are generally affected by this filtering operation. The filters are represented by blue zones ~~on~~ 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 ~~performances~~performance

4.1 Evaluation process

The data presented in Sect. 2 ~~allow reconstructing~~ 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 ~~reconstruction~~ reconstructed time series are generated~~from the probabilistic reconstructions obtained each day of the period~~. In the present section, the three reconstruction models are evaluated based on their reconstruction skill for the 1948–2010 ~~observed period~~observation period.

The evaluation is based on three criteria. The ratio β between the mean estimated value and ~~the~~ mean observed value of the variable evaluates the bias of the reconstruction. The ratio between the ~~SD~~ standard deviations of the reconstructed and ~~of the observed~~

~~values~~ observed time series (α) 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 ~~additionally summarized within~~ summarised by the Kling–Gupta Efficiency criterion (KGE; Gupta et al., 2009) defined as ~~following:~~ follows:

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

The ability of the reconstruction to reproduce the variability and variations of observations ~~is carried out~~ was evaluated for multiple temporal resolutions: daily (~~high frequency~~ 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, ~~we first present~~ the performance of the three models for an illustrative watershed (~~L'Ubaye~~ Ubaye River at Barcelonette) 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, α, β and KGE when estimated for the daily, monthly and annual ~~time-step~~ time step from the 50 ensembles. ~~We next present~~ Then, the results obtained for the 22 watersheds of the Durance ~~Basin~~basin are presented.

4.2 Performance for ~~L'~~the Ubaye River at Barcelonnette watershed

4.2.1 Air temperature reconstruction

Figure 6 presents the mean annual ~~time-series~~ time series of mean air temperature ~~of~~ for the watershed. These figures ~~firstly~~ first show that the observed temperature has increased ~~during~~ 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 ~~do~~ does not capture the temporal evolution and the variability of air temperature, ~~contrary to~~ as opposed to the LM and ANATEM ~~. We also notice that ANA ensembles are much wider than~~ models. Note also that the spread of the ANA ensembles is much higher than that of the ANATEM ensembles. ~~This is highlighted~~ These observations are consistent with the distributions obtained for the different criteria at the annual ~~time-step~~ time step (Fig. 7, right~~part~~):

- ANA ~~has~~ shows a limited mean bias (β close to 1), but ~~has rather bad~~ a rather poor temporal correlation and significant bias of variability, which is exhibited by ~~rather~~ relatively low mean values of r, α ~~,~~ and KGE (between 0.2 and 0.6, not visible ~~on~~ in the figure);

- LM and ANATEM ~~present~~ show very good temporal correlations (mean r greater than 0.9) and limited mean and variability bias. The two models ~~have~~ show slightly different skills: LM ~~has~~ shows no mean bias (by construction) but a significant variability bias (mean α less than 0.9) whereas ANATEM ~~has~~ shows a limited mean and variability bias (mean β and mean α greater than 0.95).

Figure 7 ~~presents also~~ also shows the distributions of the criteria for daily and monthly ~~time-steps~~time steps. The hierarchy ~~between the three model is comparable at these time-steps~~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, ~~ranging from~~ 0.95 to 0.98 for LM and ~~ranging from~~ 0.93 to 0.97 for ANATEM. Moreover, the mean criteria are higher at a monthly ~~time-step and then~~ time step than at a daily ~~time-step, compared to the annual time-step finally~~time step, while the annual time step performance is the lowest. This means that ~~models have more difficulties to reproduce annual~~ these models have greater difficult in reproducing inter-annual or daily variability than intra-annual variability (this is partly due to the seasonality of air temperature). ~~LM model is performing~~ The LM model performed slightly better than ANATEM at daily and monthly ~~time-steps~~time steps. Conversely, ANATEM ~~is performing~~ performs better than LM at annual ~~time-steps~~time steps.

4.2.2 Precipitation reconstruction

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

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

- ANA ~~has~~ shows a moderate correlation (mean r close to 0.5), while LM and ANATEM ~~have~~ show a rather good correlation (mean r greater than 0.8);
- ~~ANA and LM have a limited~~ LM shows no mean bias (~~less than 0.02), ANATEM has~~ by construction), while ANA and ANATEM show a moderate mean bias (less than 0.05);
- ANA ~~and ANATEM have a limited~~ shows a noticeable variability bias (~~less than 0.02), LM has a moderate~~ up to 0.15), while LM and ANATEM show a limited variability bias (~~less than 0.05~~around 0.03).

The hierarchy ~~between the~~ of the performance of the three models is ~~comparable at~~ the same for both daily and monthly ~~time-steps~~time steps, with KGE values ranging from ~~0.4~~ 0.35 to 0.7 for ANA, ~~ranging from 0.8 to 0.9~~ 0.78 to 0.88 for LM and ~~ranging from 0.78 to 0.88~~ 0.73 to 0.85 for ANATEM (Fig. 9). ANA performance is clearly poor at a daily ~~time-step~~time step, with a very limited correlation (r less than 0.4). The mean criteria are higher at a monthly ~~time-step~~ time step and similar at daily and annual ~~time-steps~~time steps. As for air temperature, this highlights the difficulty ~~of the models~~ the models have to reproduce the ~~low and high frequency~~ low- and high-frequency variability while the ~~infra-annual~~ intra-annual variability is well-captured.

4.3 Performance for all 22 watersheds

For the sake of ~~conciseness, we consider here for the evaluation only one reference time-series~~ readability, only one time series is considered for each model. ~~For the Local Model, this is simply the reconstruction obtained with the model. For the probabilistic reconstruction models~~ ANA and ANATEM ~~, this is the mean time-series~~ probabilistic reconstructions are represented by the mean time series derived from the ensemble ~~of 50 time-series reconstructions~~ (the daily reconstructed value for a ~~ ~~given day is the mean of the ~~probabilistic reconstruction~~ 50 probabilistic reconstructions for this day). ~~As it will be noticed later, these~~ 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 ~~obviously present a much lower~~ logically present a much lower temporal variability than each ~~time series~~ individual component of the reconstruction ensemble. ~~For the sake of simplicity, these mean time series will be also referred to as reconstructed time series.~~ In the following, the performance of a given model will be presented with the distributions of r, α, β and KGE criteria obtained for the 22 watersheds at the daily, monthly and annual ~~time-steps~~time steps.

4.3.1 Air temperature reconstruction

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

– ~~at a daily and a monthly time-steps, ANA~~ At daily and monthly time steps, the ANA model suffers from a limited positive mean bias (mean β around 1.03) and a significant negative variability bias (mean α from 0.85 to 0.88). Correlation with observations is very good (mean r greater than 0.93). At the annual ~~time-step, ANA is not able~~ time step, ANA fails to capture the ~~long-term~~ low-frequency variability and trend, with a very low correlation (mean r close to 0.53, not shown ~~on~~ in the figure) and a very strong negative variability bias (mean α close to 0.42, not shown ~~on~~ in the figure). At different ~~time-steps~~ time steps and for different criteria, ANA ~~also~~ exhibits a rather good spatial robustness of ~~performances (~~performance (i.e. homogeneity of the results at a regional scale, which could be related to a rather limited spread of the distribution ~~: distance between quantile 0.1 and 0.9),~~ compared to LM and ANATEM models, as shown by the distance between quantiles 0.1 and 0.9).

- ~~at the different time-steps, LM model presents~~ At all the different time steps, the LM model provides very satisfactory results. It ~~has~~ shows no mean bias (by construction) and a moderate to limited variability bias (mean α between 0.91 and 0.99). The ~~high to low frequency~~ high- to low-frequency variability is very well captured (mean r between 0.92 and 0.99). LM ~~has~~ shows moderate spatial robustness for correlation and variability bias, for daily and annual ~~time-steps~~time steps.

- ~~at the different time-steps, ANATEM model has~~ At all the different time steps, the ANATEM model provides very satisfactory results. It ~~has~~ shows a moderate mean negative bias (mean β close to 0.97) and a limited to moderate variability bias (mean α between 0.95 and 0.98). The ~~short-term to long-term~~ high-frequency to low-frequency variability is very well captured (mean ~~β between 0.95~~ r between 0.94 and 0.99). ANATEM ~~has~~ exhibits moderate robustness concerning mean bias ~~, but also~~ and concerning correlation and variability bias, for daily and annual ~~time-steps~~time steps.

The LM and ANATEM models ~~thus clearly outperforms~~ 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 ~~performances are~~ performance is better and more robust at a monthly ~~time-step~~time step, compared to daily and annual ~~time-step. ANATEM has~~ time steps. ANATEM exhibits a slightly better spatial robustness of ~~performances~~ performance than LM. This is also expressed by mean KGE values, ranging from ~~0.52 to 0.97~~ 0.25 to 0.87 for ANA, ~~from 0.92~~ 0.88 to 0.99 for LM and ~~from 0.95 to 0.99~~ 0.92 to 0.97 for ANATEM respectively.

4.3.2 Precipitation reconstruction

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

- ~~at~~ At a daily ~~time-step~~ time step, ANA suffers from a very moderate mean negative mean bias (mean β close to 0.95) and a strong variability bias (mean α around 0.55). It also ~~has~~ shows a limited correlation (mean r close to 0.65). At monthly and annual ~~time-step, ANA has~~ time steps, ANA shows a moderate to limited mean bias (mean β close to 0.95), a significant variability bias (mean α around 0.8) and an acceptable level of correlation (mean r between ~~between~~ 0.7 to 0.8).

- ~~at the different time-steps, LM model has~~ At all the different time steps, the LM model shows very satisfactory results. It ~~has a~~ shows no mean bias (by construction) ~~,~~ and a limited variability bias (mean α from 0.97 to 1.05). The ~~short-term to long-term~~ high-frequency to low-frequency variability is well captured (mean r between 0.77 and 0.84).

- ~~at the different time-steps, ANATEM model has~~ At all the different time steps, the ANATEM model shows very satisfactory results. It ~~has~~ shows a limited negative mean bias (mean β around 0.96) ~~,~~ and a limited variability bias (mean α from 0.94 to 1.02). The ~~short-term to long-term~~ high-frequency to low-frequency variability is well captured (mean r between 0.75 and 0.87).

The LM and ANATEM models ~~perform thus better than~~ thus perform better than the ANA model, particularly ~~concerning the~~ 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 ~~performances are~~ performance is better and more robust at a monthly ~~time-step,~~ time step compared to daily and annual ~~time-step~~time steps. The spatial robustness of ~~performances~~ performance is slightly lower for the variability ~~criteria~~ criterion than for the ~~others criterion. LM has the lowest~~ 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 ~~models~~ model performance

In the present section, ~~we discuss the spatial pattern of performances~~ the spatial patterns of performance (in terms of correlation, at a daily ~~time-step~~time step) of the three models ~~. We also present the spatial pattern~~ and the spatial patterns of the gain in performance obtained with ANATEM reconstructions when either compared to ANA or LM ~~reconstructions~~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 ~~is fairly not influenced by~~ appears to be mostly independent of the location of the watershed, with a mean correlation ranging from to 0.92 to 0.94 (Fig. 12a). For LM (Fig. 12b), the location of the watershed ~~has~~ 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 ~~of model performances~~ in model performance driven by the distance ~~to~~ 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 to~~ distance from Marseille) also influences the performance of the reconstruction, with a mean correlation ranging from 0.97 to 0.99. However, ANATEM ~~has slightly better performances~~ shows slightly better performance than LM and ~~then~~ ANA and the range of correlation values is slightly ~~thinner than the range observed~~ smaller than for LM.

The contribution of ~~LM model (Fig. 12d)~~ the LM (resp. ANA) model to the performance of ~~ANATEM, decreases from south-west to north-east, ranging from 0.06 to 0.04. Conversely, the contribution of ANA model (~~the ANATEM model is presented in Fig. ~~12e)to~~

~~the performance of ANATEM, slightly increases from~~ 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 ~~, ranging from~~ 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~~ large-scale information (through the ANA model) is stronger when the LM model (local information) is less ~~efficient~~effective, that is, when the location ~~at reconstruction~~ 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 ~~really influences~~ appear to really influence the performance, with a mean correlation ranging from ~~to~~ 0.62 to 0.68 (Fig. 13a). ~~For~~ On the other hand, for LM (Fig. 13b) and ANATEM (Fig. 13c), watersheds close to the local reference station ~~highlight conversely better performances~~ (Gap) show better performance than the others (the correlation ranges from 0.62 to 0.88 for LM and from ~~69~~ 0.69 to 0.89 for ANATEM). However, the ~~similarity in terms of large scale forcing influences probably influences the performance. Hence,~~ distance from the local reference station is probably not the only factor influencing performance, as two watersheds at the same distance ~~to~~ from the Gap station ~~have rather different performances~~ displayed somewhat different performance (i.e. ~~Buech watershed~~ 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 ~~increases the global~~ slightly increases the overall reconstruction performance but ~~it also notably smooths~~ 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 ~~to~~ from Gap increases, ranging from 0.22 to 0.02 (Fig. ~~13d). Conversely~~14c). On the other hand, the contribution of ANA to the performance of ANATEM ~~slightly increases from south-west to north-east, ranging from 0.0 to~~ is close to 0 for the stations closest to Gap and slightly increases (up to 0.07) with the distance from Gap (Fig. ~~13e~~14d). As observed for air temperature reconstruction and here in a more pronounced way, the contribution of ~~large scale~~ large-scale information (through the ANA model) is stronger when the LM model (local information) is less ~~efficient~~effective, as a result of ~~an increasing distance to~~ 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 ~~anomaly time-series~~ time series of air temperature anomalies reconstructed by the ANATEM ~~method~~ model (mean model) for the 22 watersheds of the Durance river. Anomalies have been computed as ~~the differences to~~ 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 ~~nowadays~~the present. In order to better ~~characterize~~ characterise low-frequency variability, ~~a smoothed mean of the~~ the mean of all 22 ~~series reconstructed for the 22 watersheds respectively has been computed by LOESS (Cleveland, 1979) . This series (red curve in Fig. 15) highlights a relatively strong variability: mean air temperature can vary of nearly 1C in less than 10(e.g.: 1890–1900, 1940–1950).~~ 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 ~~homogeneised~~ 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 ~~of~~ between ANATEM and HISTALP annual series is close to 0.8). However, the warming trend in the HISTALP series is ~~larger~~ 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~~ time series are obviously sensitive to the reference ~~time-series~~ 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 ~~precipitation reconstructions obtained~~ annual multiplicative time series of precipitation anomalies reconstructed with ANATEM for the 22 ~~watershed~~ 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 ~~is~~ are also given.

~~ANATEM series present a very homogeneous temporal behaviour when compared to the high dispersion~~ 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 ~~the high spatial variability of precipitation and the fact that~~ that the HISTALP series cover a much wider spatial domain ~~than ANATEM series. The low dispersion between the reconstructed series is otherwise coherent with the limited dispersion obtained between time-series observed for the same 22 watersheds on the observation period (not shown~~

~~here).~~ with a high spatial variability of atmospheric influences and thus precipitation regimes and times series.

~~Besides, the~~ The smoothed time series from ANATEM reconstructions is highly correlated ~~to~~ with the smoothed time series from HISTALP data. The ANATEM reconstruction is therefore able to reproduce the ~~low frequency~~ 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 ~~translate~~ correspond to different precipitation indexes, as long as they are estimated from different stations. As ~~noticed~~ 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 ~~would be worth~~ should be considered to explore the importance of ~~these~~ such issues.

6 Conclusions

Reconstructing local scale meteorological variables over long periods is ~~challenging and necessary for better understanding the low frequency~~ 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 ~~kind~~ types of data for the reconstruction: the ~~regression based Local Model~~ regression-based local model (LM) uses local observations of the variable from neighbouring stations as ~~predictor, the Analog Model~~ a predictor; the analogue model (ANA), a so-called downscaling model, uses ~~large scale information of atmospheric circulation ,~~ large-scale information concerning atmospheric circulation and the ANATEM model uses a mix of both local and ~~large scale atmospheric information combining therefore both the Local Model and the Analog Model~~large-scale atmospheric information by combining the local and analogue models.

The three models have been developed and applied ~~for~~ to the reconstruction of mean air temperature and precipitation ~~time-series of~~ time series for a sample of 22 watersheds situated in the Durance ~~watershed,~~ 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 ~~observation~~ observations used for the reconstruction are respectively Marseille air temperature, Gap precipitation historical ~~time-series and geopotential~~ time series and geopotential height fields from the 20CR reanalysis.

The ~~multicriteria and multiscale~~ multi-criteria and multi-scale performance assessment highlights that the best reconstructions are obtained when local information is used. ~~ANA~~ The ANA model is clearly less efficient than the two ~~others methods~~other models, particularly concerning ~~long-term~~ low-frequency (annual) air temperature variability or ~~short-term~~ high-frequency (daily) precipitation variability. ~~The regression based~~ On the other hand, the regression-based model and the ANATEM model ~~have conversely~~ provide very satisfactory results for all criteria. ANATEM ~~has~~ offers a slight advantage and the spatial patterns of the reconstruction skills show that it ~~takes benefit~~ benefits from the qualities of both underlying models. Hence, the ANATEM ~~method allow to reconstruct very satisfactory~~ model can be used to reconstruct adequate air temperature and precipitation ~~reanalyses~~ series at a high temporal resolution (daily) and different spatial scales (from 4 to 3500 km^2), while improving the spatial robustness of ~~performances~~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~~ time series highlight the large inter-annual variability of annual precipitation for the Durance region. Long-term climatic reanalyses ~~exhibits~~ exhibit some particular periods with rather strong rainfall anomalies, such as the wet periods at the beginning of the 1910's and ~~mid 1930~~mid-1930's (known for ~~floods~~ flood events), or ~~rather~~ relatively dry periods such as the 1940's and 1970's (known for ~~droughts~~ drought events).

~~This work has a number of interesting perspectives and raises a variety of challenging questions.~~

The potential for improving the method is ~~not negligible~~considerable. The ANA method ~~we used here has been firstly~~ 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~~ large-scale predictors could potentially ~~allow for~~ provide a better reconstruction of air temperature variations. This also applies ~~for~~ to the precipitation reconstruction. The predictors used here ~~do only inform~~ provide information only on the atmosphere dynamics. The inclusion of thermodynamic predictors and ~~of~~ humidity predictors for the identification of ~~analog~~ analogue days has been proved to improve ~~for the studied region the~~ the performance of the method for the studied region (Marty et al., 2011; Chardon et al., 2014).

Another possibility ~~of progress~~ for improvement concerns the combined formulation used for the ANATEM ~~method~~model. The formulation presented in this paper has been applied ~~straightforwardly and has not been modified on the basis of these results~~in a straightforward fashion. However, ~~we~~ the authors are convinced that ~~fancy statistical developments on the way both~~ 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) ~~would allow to~~ could improve temporal correlation, ~~or~~ reduce mean and variability bias and ~~consider~~ allow probabilistic calibration (not ~~shown~~ addressed in this paper).

The ~~issue~~ choice of the reference series used for the ~~Local Model is also challenging. We have~~ local model also presents a challenging issue. It has been shown that the good performance of the ~~methods is achieved thanks to~~ models largely depends on this local information. A ~~thorough sensitivity analysis to the selection~~ thorough analysis of the sensitivity to the choice of the reference ~~time-series~~ time series should be carried out. Considering the importance of local information, an extension of the method should also consider the possibility ~~to make use of all historical stationsavailable in the close or farer~~ of making use of other historical stations, if available, in the neighbourhood of the region ~~under construction.~~ 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. ~~Our results also highlight~~ The results also show that the reconstruction skill decreases ~~when the distance to~~ as the distance from the reference station increases. The region ~~we have considered covers a rather narrow domain. However, we can expect that the interest~~ considered in the present study is relatively small. The importance of the reference station ~~is much lower if we would do reconstructions for much more distant locations. We can expect conversely that the relative interest of the large scale~~ 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 ~~much larger for distant sites~~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 ~~characterization of long term variability, we~~ characterisation of long-term variability, the authors are confident that all of these questions will be tackled by the scientific community in the ~~next~~ coming years. A major application of such reconstructions will ~~be obviously the possibility to reconstruct long term~~ obviously be the reconstruction of long-term variations for a number of ~~climate driven variables. As an illustration, we used~~ climate-driven variables. For example, the long-term climatic ~~time-series~~ time series produced in the present work ~~for reconstructing~~ have been used to reconstruct long-term hydrological ~~time-series~~ time series at multiple hydrometric stations of the Durance basin (Kuentz, 2013; Mathevet et al., 2013). ~~An outstanding result of this reconstruction is that the time series obtained for the whole~~ 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 present a very high correlation level with historical discharges time series obtained from rescued hydrometric archives for the catchment. In our case, the availability of historical streamflow time seriesallowed us to demonstrate the overall quality of the meteorological reconstruction. This independent hydrological validation is not expected to be feasible everywhere but it gives high confidence in this hydrometeorological~~

~~reconstruction approach. Even when such an independent validation cannot be carried out, the reconstructed time series definitively produce a high-value information for researchers or water resources managers. Further works for other hydroclimatic contexts are therefore also worth to better identify the potential of the method and the possibility for improving it~~century.

Acknowledgements. The authors are ~~grateful~~ 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 ~~the NOAA which provided~~ 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 autors are very grateful to Bettina Schaefli 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 downscalling 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 # ~~correspondance with~~ column correspond to those indicated in Fig. 2, P_A is the annual mean ~~precipitations~~ precipitation (~~on~~ over the period 1948–2010) ~~,~~ and T the mean air temperature (~~on~~ over the period 1948–2010)

#	Name	Altitude m	Area km^2	$\bar{P}_A$ mm yr^{-1}	$\bar{T}$ °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

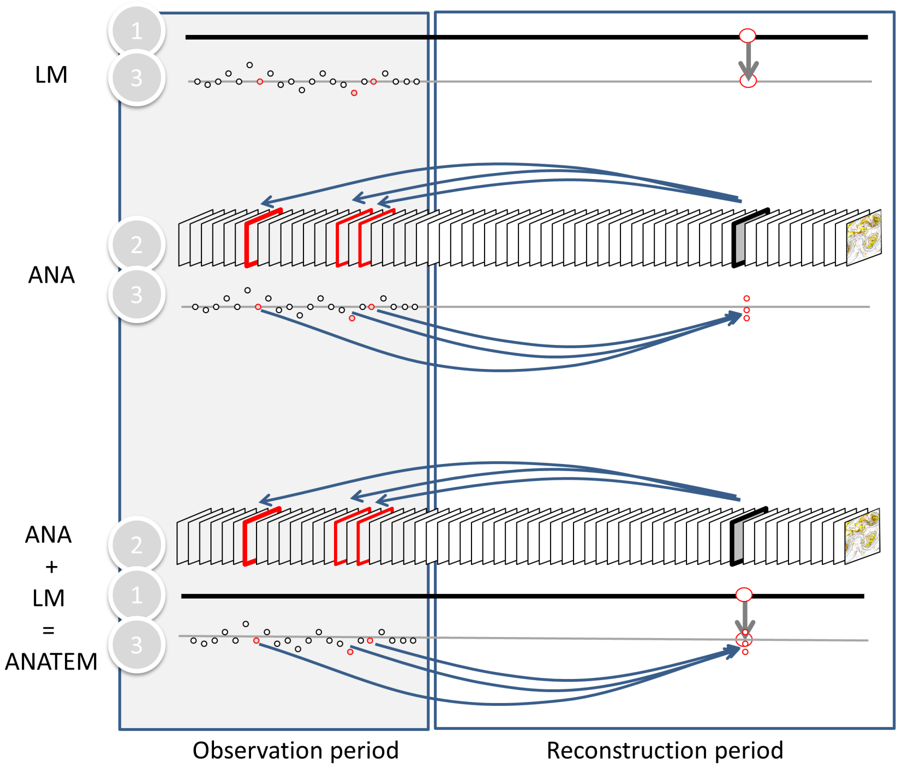

1
LM
3
2
ANA
3
2
ANA
+
1
LM
=
3
ANATEM
Observation period
Reconstruction period

Figure 1. ~~Map~~ Schemes of the ~~study area with~~ 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 ~~22 watersheds selected~~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.

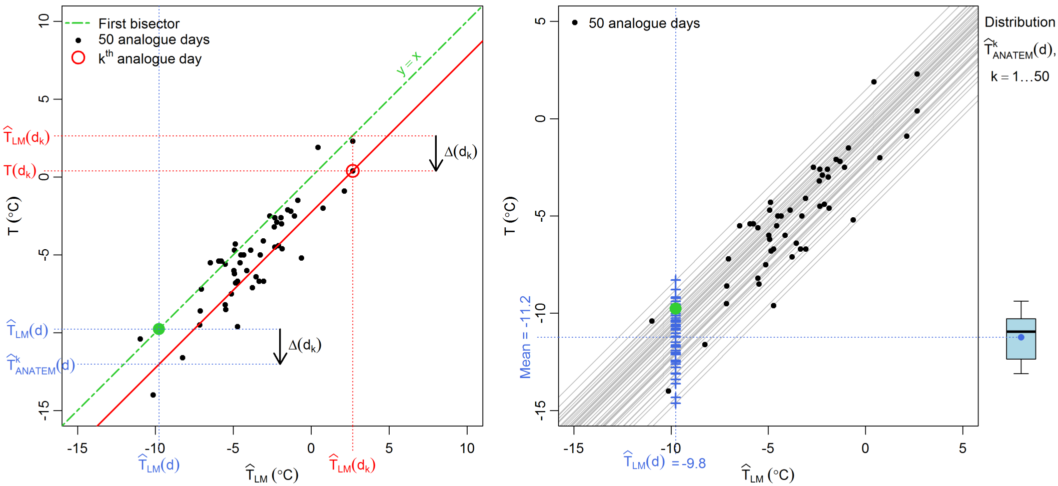


Figure 3. Representation of the ANATEM formulation for air temperature reconstruction of a given day d. Left panel: observed temperature for ~~the target~~ any analog day as a function of the temperature estimate ~~from LM~~with LM for this day. ~~For~~ The green point corresponds to the target day d.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

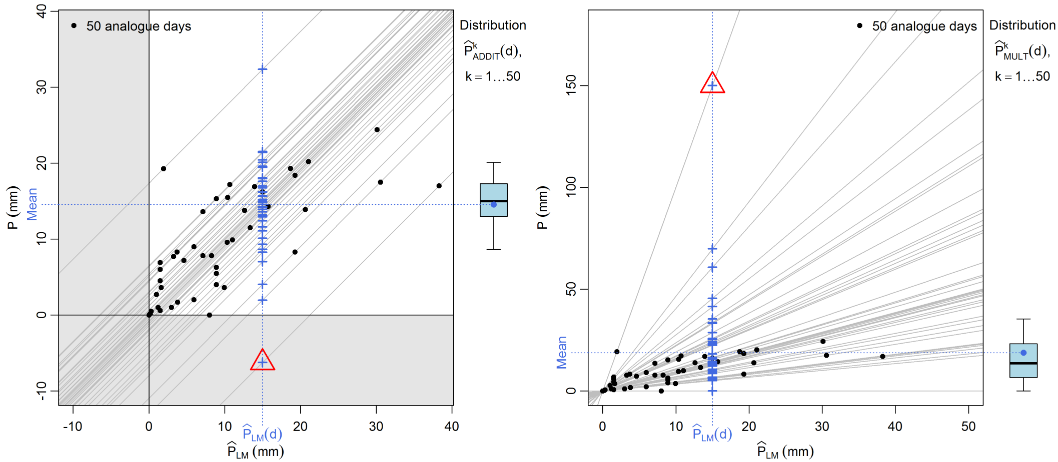


Figure 4. Representation of the additive and multiplicative formulation for precipitation reconstruction from a local model and 50 ~~analog~~ analogue days ~~of~~ for a given day d. Left panel: ~~case of an~~ additive formulation for the correction. Right panel: ~~case of a~~ multiplicative formulation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

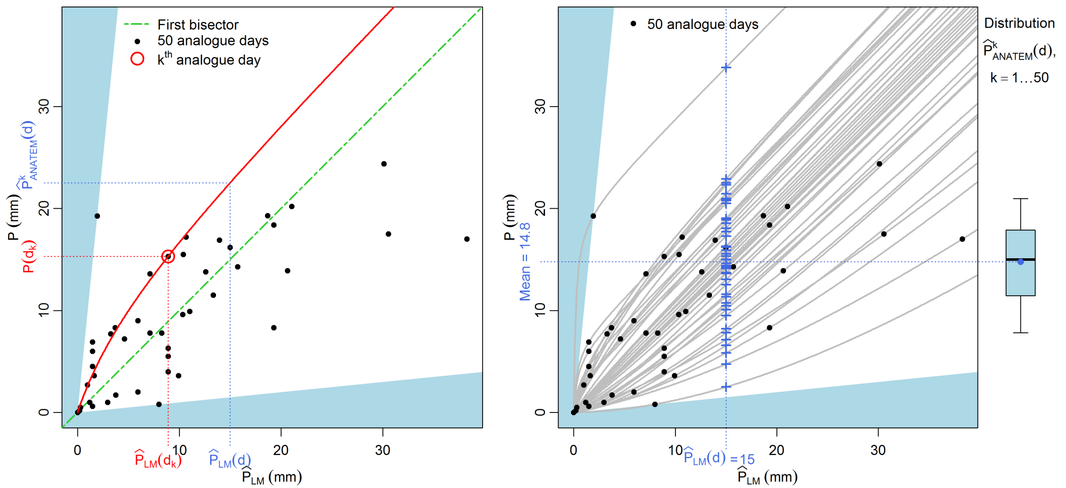


Figure 5. Representation of the ANATEM formulation for ~~precipitations~~ precipitation reconstruction ~~of~~ for a given day d. Left panel: observed precipitation ~~for~~ at the target site for each of the analogue days as a function of the precipitation estimate from LM for the same days. For a given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

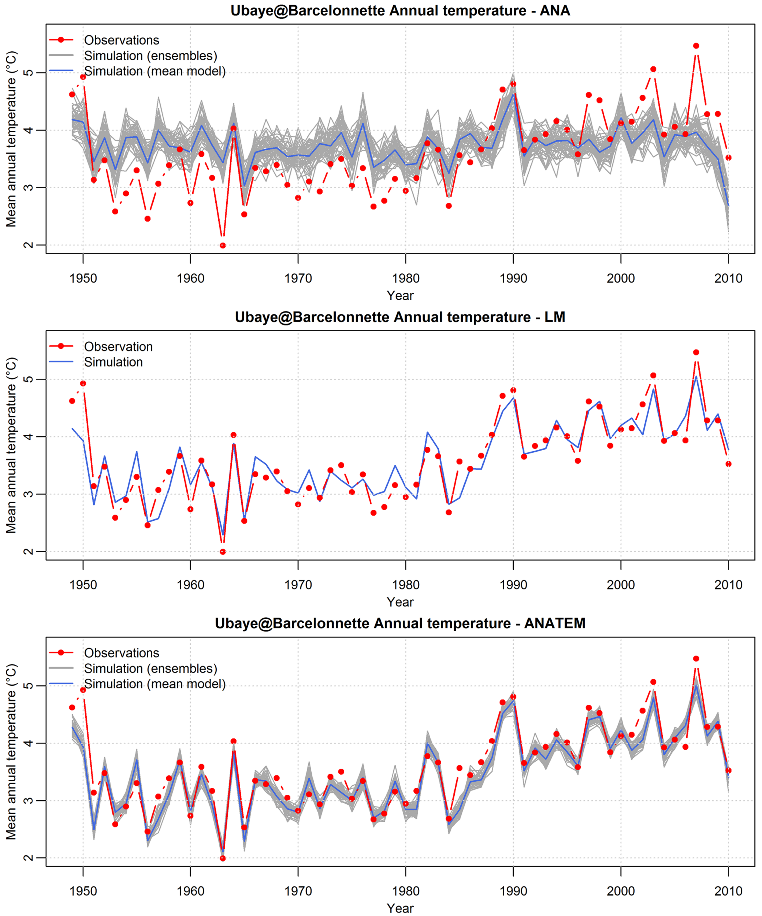


Figure 6. Annual ~~time-series~~ time series of air temperature reconstructions for the Ubaye River at Barcelonnette watershed by ~~analog method~~ the analogue (ANA), local ~~model~~ (LM) and ANATEM ~~method~~models.

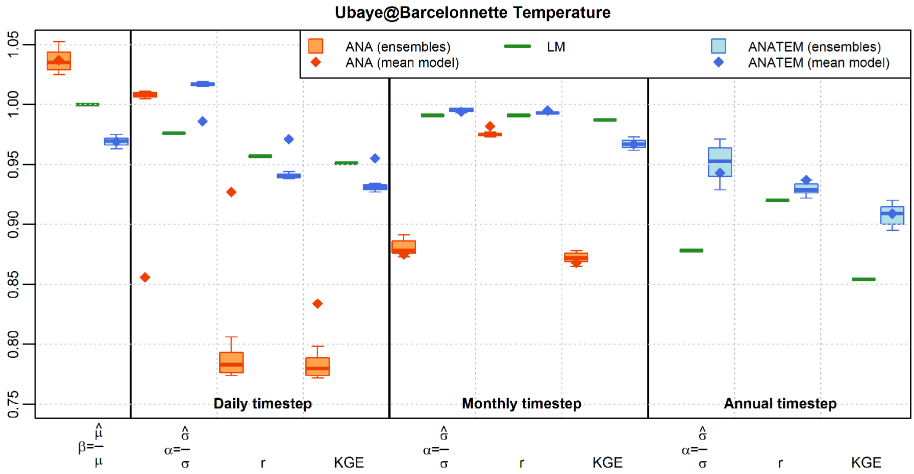


Figure 7. Daily, monthly and annual performance criteria of air temperature reconstructions for the Ubaye River at Barcelonnette watershed by ~~analog method~~ the analogue (ANA), local ~~model~~ (LM) and ANATEM ~~method~~models. For the annual time step, ANA results are smaller than 0.75; they therefore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

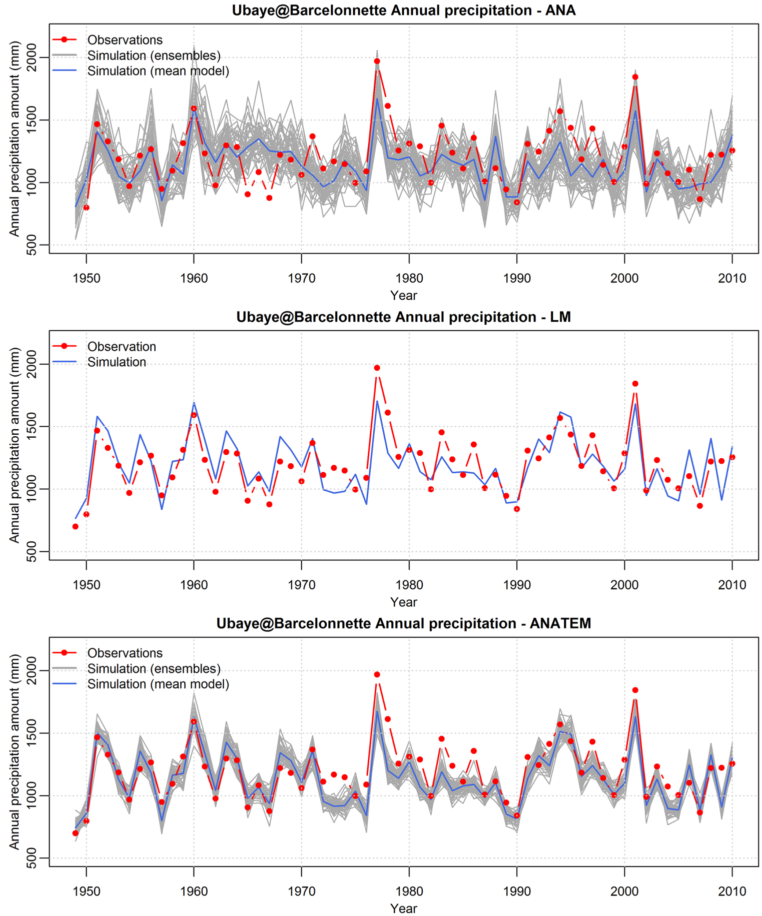


Figure 8. Annual ~~time-series~~ time series of precipitation reconstructions for the Ubaye River at Barcelonnette watershed by ~~analog method~~ the analogue (ANA), local ~~model~~ (LM) and ANATEM ~~method~~models.

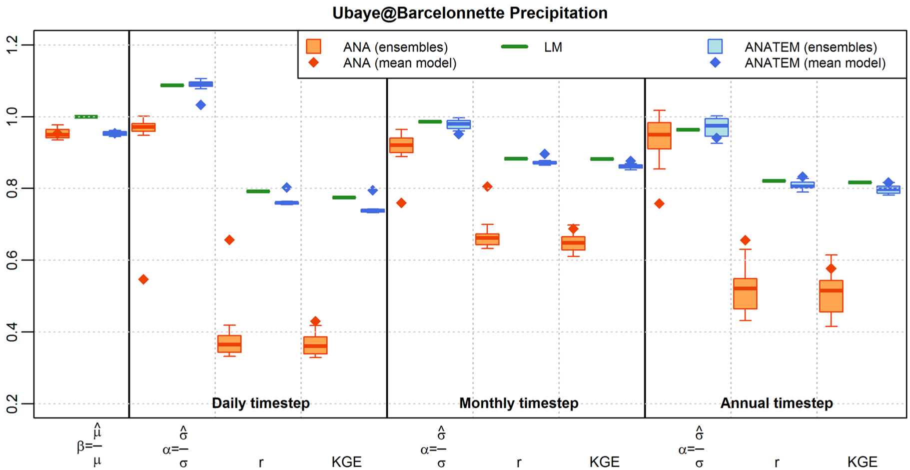


Figure 9. Daily, monthly and annual performance criteria of precipitation reconstructions for the Ubaye River at Barcelonnette watershed by ~~analog method~~ the analogue (ANA), local ~~model~~ (LM) and ANATEM ~~method~~models.

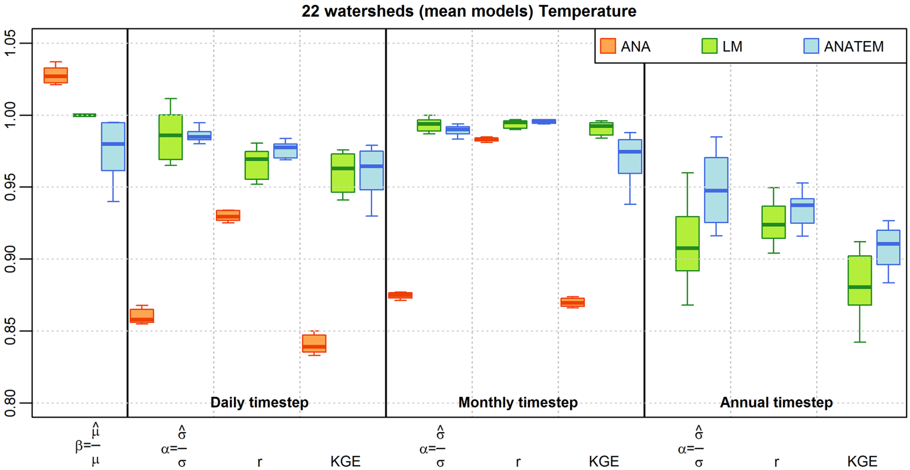


Figure 10. Daily, monthly and annual performance criteria of air temperature mean reconstructions for 22 watersheds by ~~analog method~~ the analogue (ANA), local ~~model~~ (LM) and ANATEM ~~method~~models. ~~Annual performance criteria of air temperature mean reconstructions for 22 watersheds by analog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

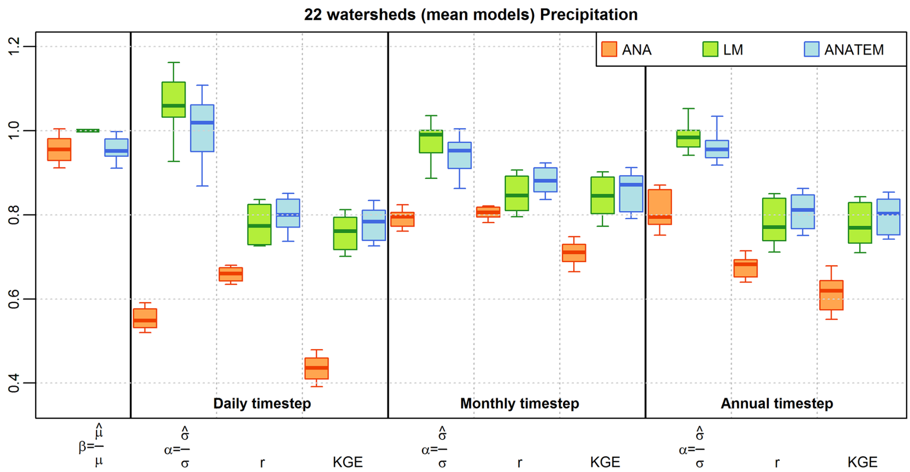
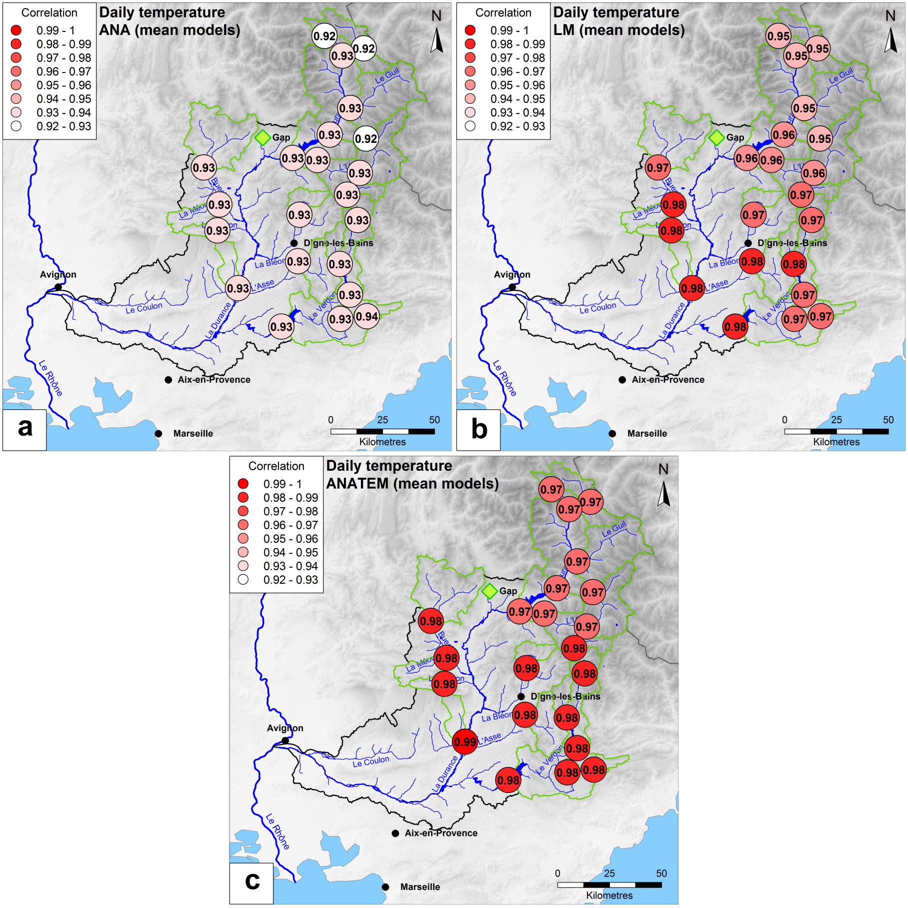


Figure 11. Daily, monthly and annual performance criteria of precipitation mean reconstructions for 22 watersheds by ~~analog method~~ the analogue (ANA), local ~~model~~ (LM) and ANATEM ~~method~~models.



Correlation
0.99 - 1
0.98 - 0.99
0.97 - 0.98
0.96 - 0.97
0.95 - 0.96
0.94 - 0.95
0.93 - 0.94
0.92 - 0.93
Daily temperature
ANA (mean models)
Daily temperature
LM (mean models)
Daily temperature
ANATEM (mean models)
N
Gap
Digne-les-Bains
Avignon
Aix-en-Provence
Marseille
Le Guil
La Bléone
L'Asse
Le Coulon
La Durance
Le Verdon
Le Rhône
0 25 50
Kilometres
a
b
c

Figure 12. Regional correlation patterns of air temperature mean reconstructions by the **(a)** ~~analog method~~ analogue model (ANA), **(b)** local model (LM) and **(c)** ANATEM ~~method~~model.~~Spatial pattern of the gain in correlation obtained with ANATEM reconstructions compared to ANA **(d)** or LM **(e)** reconstructions.~~

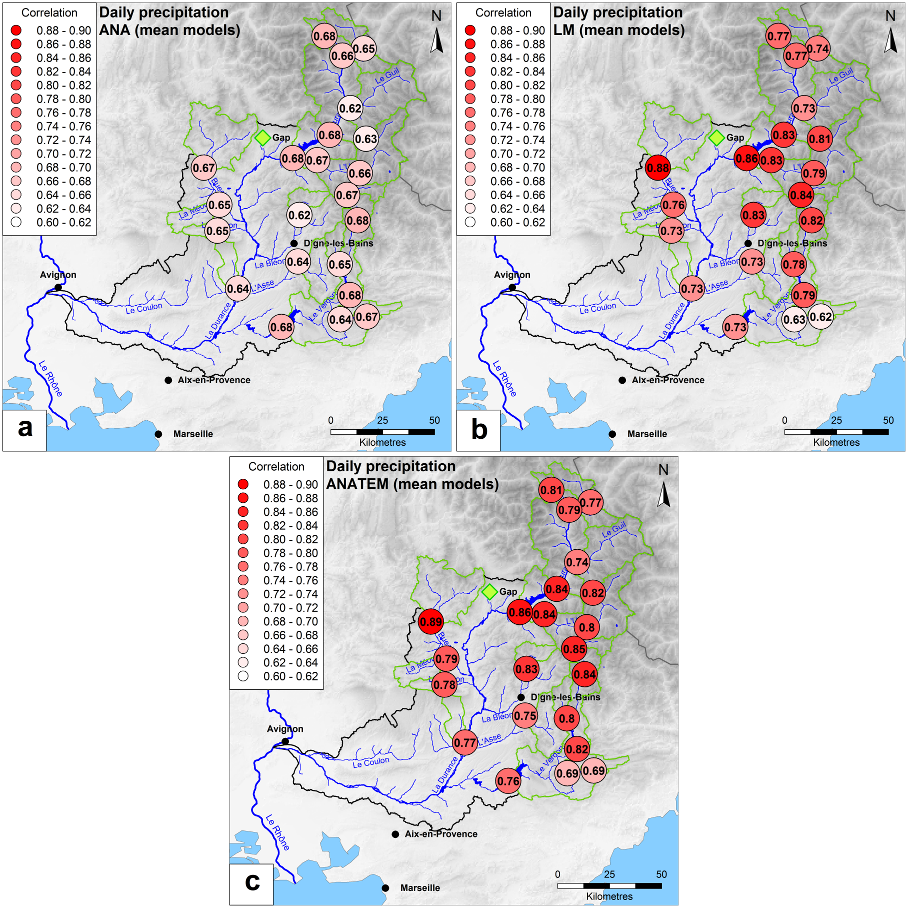


Figure 13. Regional correlation patterns of precipitation mean reconstructions by the **(a)** ~~analog method~~ analogue model (ANA), **(b)** local model (LM) and **(c)** ANATEM ~~method~~model.

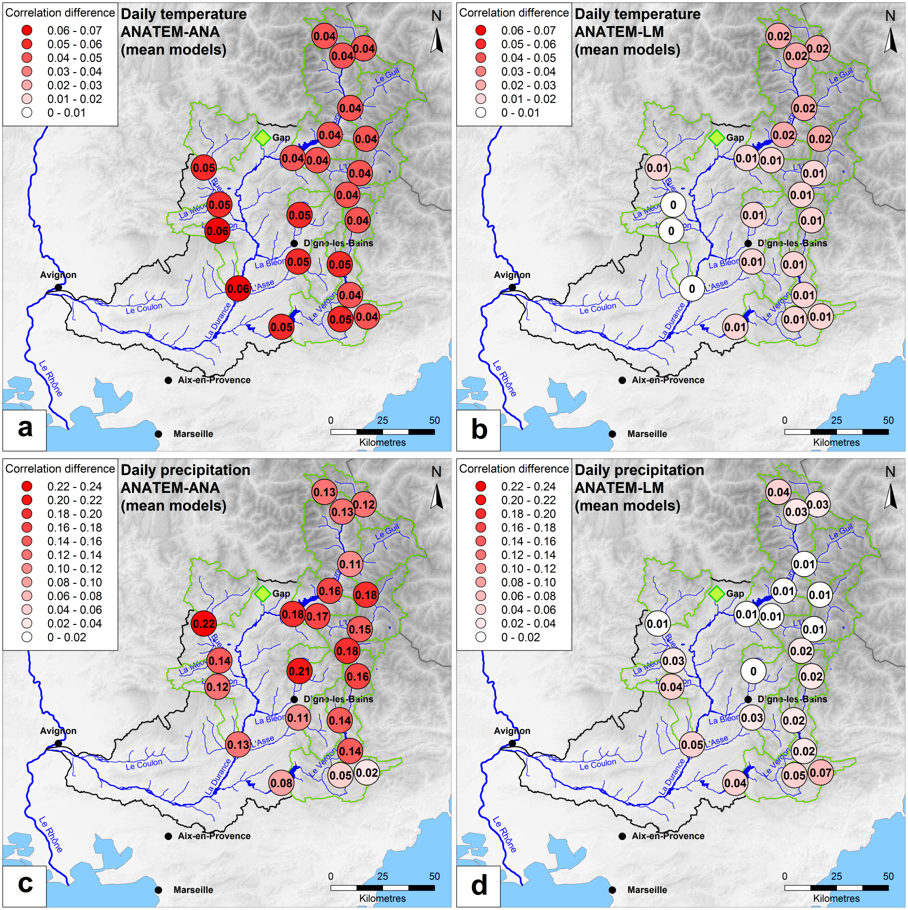

Correlation difference
Daily temperature
ANATEM-ANA
(mean models)
Daily temperature
ANATEM-LM
(mean models)
Daily precipitation
ANATEM-ANA
(mean models)
Daily precipitation
ANATEM-LM
(mean models)
0.06 - 0.07
0.05 - 0.06
0.04 - 0.05
0.03 - 0.04
0.02 - 0.03
0.01 - 0.02
0 - 0.01
0.22 - 0.24
0.20 - 0.22
0.18 - 0.20
0.16 - 0.18
0.14 - 0.16
0.12 - 0.14
0.10 - 0.12
0.08 - 0.10
0.06 - 0.08
0.04 - 0.06
0.02 - 0.04
0 - 0.02
Gap
Digne-les-Bains
Avignon
Aix-en-Provence
Marseille
Le Guil
La Bléone
L'Asse
Le Coulon
La Durance
Le Verdon
Le Rhône
N
0
25
50
Kilometres
a
b
c
d

Figure 14. Spatial ~~pattern~~ patterns of **(a,c)** the ~~gain in correlation obtained with~~ 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~~compared to~~ .

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

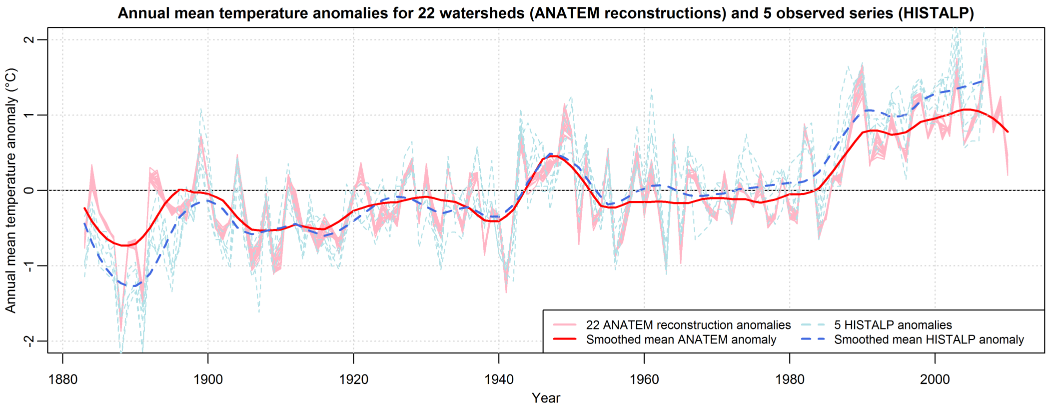


Figure 15. Mean annual air temperature additive anomaly for the 22 watersheds (ANATEM) and 5 stations (HISTALP). The additive anomaly for a given year has been computed as the difference between the annual temperature for this year and the 1883–2010 mean.

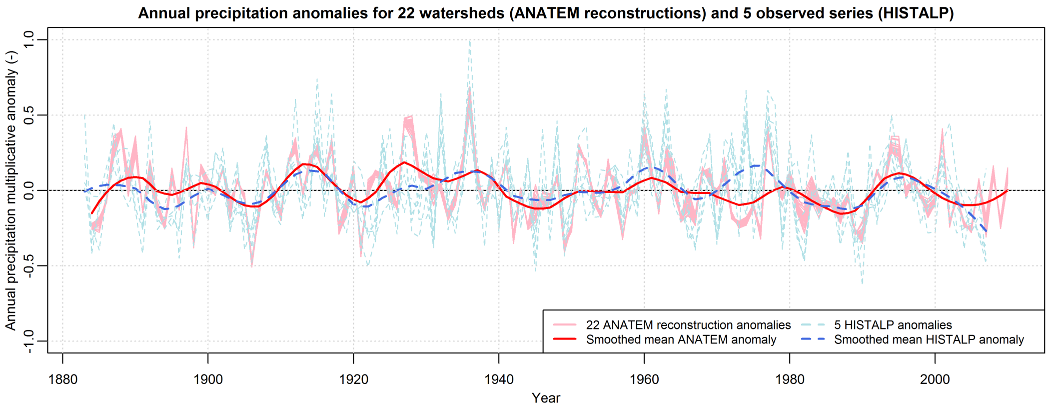


Figure 16. Mean annual precipitation multiplicative anomaly for the 22 watersheds (ANATEM) and 5 stations (HISTALP). The multiplicative anomaly for a given year has been computed as the ratio between the annual precipitation for this year and the 1883-2010 mean.