Hydrol. Earth Syst. Sci. Discuss., 12, C425–C439, 2015
www.hydrol-earth-syst-sci-discuss.net/12/C425/2015/

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Building long-term and high spatio-temporal resolution precipitation and air temperature reanalyses by mixing local observations and global atmospheric reanalyses: the ANATEM method" *by* A. Kuentz et al.

A. Kuentz et al.

anna.kuentz@gmail.com

Received and published: 5 March 2015

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 geopotential 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 near-

est 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 $\left[X_{ANA_d^k}\right]_{k=1...n}$ where $k = 1 \dots n$ 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 $\left(\frac{P_{ANA_d^k}}{P_{LM,ANA_d^k}}\right)^2$
- When $P_{ANA_d^k} = P_{LM,ANA_d^k}$, the following should be obtained : $\widehat{P}_d^k = P_{LM,d}$

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 :

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

The second condition gives the equivalence relation :

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

From these two relations the coefficients can be defined as :

$a_d^k = P_{ANA_d^k}$ and $b_d^k = \frac{\left(P_{LM,ANA_d^k}\right)^2}{P_{ANA_d^k}}$

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 $P_{LM,d}$.

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 $y = \frac{b_d^k}{x_d}$:

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

References

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, 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.

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, http://www.sciencedirect.com/science/article/pii/S0022169409004843, 2009.

Hawkins, E. and Sutton, R.: The Potential to Narrow Uncertainty in Regional Climate Pre-

dictions, Bulletin of the American Meteorological Society, 90, 1095–1107, doi:10.1175/2009BAMS2607.1, http://journals.ametsoc.org/doi/abs/10.1175/2009BAMS2607.1, 2009.
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, 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, http://onlinelibrary.wiley.com/doi/10.1002/asl.176/abstract, 2008.
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.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 12, 311, 2015.