Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2017-62-AC1, 2017

Hydrology and Earth System Sciences Discussions

Open Access

Interactive comment on "An adaptive two-stage analog/regression model for probabilistic prediction of local precipitation in France" *by* Jérémy Chardon et al.

Jérémy Chardon et al.

benoit.hingray@univ-grenoble-alpes.fr

Received and published: 29 June 2017

[hess, manuscript]copernicus

We thank the referee for this thorough review and for the numerous constructive suggestions that we will consider for incorporation in the modified manuscript. We give here the detailed responses to all his comments and questions.

Printer-friendly version

Discussion paper

1 Topic and general comments

1.1 Topic

The paper presents a new two-stage hybrid perfect prognosis SDM called SCAMP. SCAMP was applied to a large number of grid points in France and was proven to be adaptive to different weather types and seasons which is illustrated nicely with visually appealing figures. The method seems very interesting given the issues encountered with some other very popular downscaling- or bias correction methods (e.g. lack of variance for pure transfer functions or physical inconsistency that easily occurs with quantile mapping and related techniques). There are a couple of issues though that I think should be addressed before publication. Some of them might be just a matter of clarification, but some might be more fundamental depending on the intended use of the method. These issues are outlined in the following.

1.2 What is the intended use of the method?

In the introduction you mention regional climate studies of present, past and future climate as well as numerical weather prediction (NWP) but without being very clear for which of these cases SCAMP is actually made for. Given that you downscale from 1.125 degree resolution to a 8km grid I suppose that SCAMP is not designed to do NWP, given that the ECMWF global deterministic model runs at 9km resolution and most national weather services in Europe operationally run limited area models at 1-2km resolution and limited area ensembles at 2-10km resolution. If however that is the intended use, please explain in which context and for which users you think it could be useful. What made me doubting that SCAMP is intended for regional climate studies, is the use of the word “prediction” throughout the paper. If the intended use are regional climate studies, I would recommend to either use “simulation” rather than “prediction”

Printer-friendly version

Discussion paper

or to precisely define what "prediction" means in this context. The same applies to section 3.4.

A widely used argument for the development of statistical downscaling models (SDMs) is that they allow producing local scale weather scenarios. We obviously agree that high resolution ensembles are operationally available from most national weather services. SCAMP would not be of any interest in respect to this point.

As mentioned in the manuscript, another important argument for the development/use of SDMs is that the outputs of GCM and/or NWP models are generally 1) biased and, from a statistical point of view, 2) not reliable (the ensembles are often underdispersive – (see for instance, Leutbecher and Palmer, 2008). In a number of cases however, impact studies require unbiased and reliable meteorological scenarios. This is for instance a critical requirement for hydrological impact studies as a result of the strong linearities in the hydrological response of river basins to meteorological forcings.

In the present work, we did not select a given context for the application of SCAMP. SCAMP could be used for either forecasting, reconstruction or simulation. We will precise this in the new manuscript version. Some specific requirements would apply for each context. For instance, the temporal transferability of the model in a modified climate context would be required for the development of climate projections. The quality of large scale predictors would have to be checked for reconstructions over the XXth century or for climate prediction (as often reported, thermodynamic predictors are of lower quality than dynamic ones – see questions + responses to 'specific comments' 5 and 6 below).

We will precise what the word "prediction" means in this context. We find this word more suited than "simulation" because this latter suggests that times series of precipitation are produced. This is here not the case (although some postprocessing generation process could be used for this but this is out of the scope of the present *work) as we issue for each day the statistical distribution of precipitation amount (thus a probabilistic prediction).*

1.3 Manuscript organization and conciseness

1. The introduction is to my mind rather long and could be written more concisely. In addition it should contain some more precise statement on the intended use of SCAMP (see section 1.2).

We will adapt the introduction as suggested and clarify the intended use of SCAMP.

2. I don't understand why the description of the analog stage (stage 1, section 3.2 and 3.3) comes after the description of the GLM stage (stage 2, section 3.1). In my view this should be reversed. The first part of section 3 (page 5) should contain a concise outline of SCAMP. There is a start at page 5 line 10-12 that should be completed with one or two sentences on the backup model.

We thank the referee for these suggestions. We will complete the outline of SCAMP and we will describe the analog stage before the GLM one as suggested.

3. The last paragraph of section 3.2 could go in a tightened section 3.3 as well (The AM as benchmark and backup model). Its last two sentences are already a very concise summary of section 3.3.

The Analog Model can be indeed presented as a benchmark and backup model. We will merge the last paragraph of section 3.2 with section 3.3 as suggested.

4. I wonder about sections 2 and 4.1 as well: I found it somewhat difficult to figure out which potential predictors were actually used during the first read. There are a few things said in section 2, during section 3 things are quite vague (concerning predictors) and only in section 4.1 things became more clear. If you consider 4.1 to be a central result of the study the information in this subsection should be split into a “methods part” right after or included in section 2 and a “results part” remaining in section 4. If this is not the case I’d suggest to entirely include section 4.1 after or into section 2, but rewritten (together with section 2 from the fourth paragraph on, page 4 line 17 et seq.) in a much more concise manner. For example saying first what you used in the end and then concisely explain why. I think this would allow to be more specific and to use more precise wording in section 3. With a more clear structure lengthy transitions, such as the page 5 last sentence or page 11 lines 4-6, might not be necessary any more.

We thank the referee for these different suggestions. As suggested we will include section 4.1 into section 2 and modify the text and transitions consequently. This will make indeed the paper more clear.

1.4 Language issues.

Please check your paper thoroughly for language/grammar issues during the revision, especially

1. tenses stick to simple past for things you did avoid future tense for things you finally did, otherwise it induces unnecessary doubt. 2. reduce the use of modal verbs (may, could etc.) where possible in order to be more precise and quantitative.

3. prepositions
4. word order in the context of adjectives and adverbs
5. remove superfluous adverbs for more clarity
6. add missing definite articles
7. mind French to English translation pitfalls

See the technical correction section for examples.

We will carefully check for these different issues and a native English person will read the paper. Thank you for these recommendations.

2 Specific comments

1. Is SCAMP an abbreviation for something? (I'm just curious)

In the previous work of Raynaud et al. (2016), we first worked on a multivariate Analog version, for multivariate prediction (precipitation, temperature and radiation). SCAMP is the abbreviation defined in this previous work and stands for Sequential Constructive atmospheric Analogs for Multivariate weather Prediction. We kept this abbreviation for the present work even if we are in a monovariate configuration. This will be clarified.

2. In the introduction (first paragraph) SDM and post-processing are used synonymously. Are they? And if yes, in which context?

SDM and post-processing are sometimes rather synonymous for instance when SDM are used to produce local weather scenarios from GCM output data (their ability to do some bias correction is an important feature here as mentioned previously). We agree that this is not always the case. Other applications of SDM are possible as those mentioned in the second paragraph of the introduction (weather generation, climate change attribution...). We will reformulate the text to avoid the confusion.

3. some references seem slightly out of context. For example:

(a) Page 2 line 12-13: Maraun et al 2010a review paper already cited at Page 1 line 25

This reference has been indeed already mentioned in a previous paragraph and will be removed here.

(b) Page 2 line 28: Citation Maraun et. al 2010b: please cite something more specific in this context.

As mentioned by referee 1, we agree that a more specific reference can be given in this context. Moreover these lines were somehow confusing. Generalized Linear Models (GLM) are regression models specifically introduced by statisticians to model non-gaussian data (see Nelder and Wedderburn, 1972). They were first used by Stern and Coe (1984) for the generation of precipitation. The vector generalized linear models (VGLM, Yee and Wild 1996), closely related to the class of GLMs, are the most general class of linear regression models available. The work of Maraun et al. (2010b) is "just" one recent application of VGLMs for the case of precipitation. We will simplify this section and remove the mention to VLGMs, which is not necessary here.

(c) Page 7 line 2: Radanovics et al., 2013: isn't this one more on predictor domains?

Radanovics et al., 2013 is indeed one work focusing on France where the issue of the predictor domains has been explored. The iterative process followed to identify the best predictor domain is however slightly different from that used in the present work. It is repeated 5 times with different initial conditions whereas in the present work, we do it only one. This is the reason why we did not mention it for this domain optimization issue.

4. Page 3 line 24-26: The last sentence of the paragraph is unclear. Please rewrite.

Our point is that the type of model used in the work of Ibarra-Berastegi et al. (2011) is not really optimal. A linear regression model is indeed not suited to the non-gaussian nature of precipitation amounts. The approach of Ibarra-Berastegi et al. (2011) would thus benefit from using a model suited to precipitation. We will reformulate the sentence to make it clear.

5. Do you think that the selected predictors may depend on the data set used, or its resolution? Please comment.

Different studies have shown that the predictors depend on the predictand. For precipitation, predictors can differ from one location to the other (e.g. Cavazos and Hewitson, 2005; Timbal et al., 2009; Chardon et al., 2014). They are also not necessary the same for precipitation, radiation or other surface weather variables (e.g. Raynaud et al. 2016). We could also expect that the predictors depend on the dataset used, for the atmospheric reanalyses especially. To our knowledge this analysis has not been carried out yet. Some dependence to the resolution is also probably to expect. A higher *resolution would definitively allow for a better description of the shapes of geopotential fields. It would also allow for a more relevant simulation of thermodynamic processes. It would likely lead in turn to have higher quality variables for some quantity such as air instability (as mentioned in the following question). The quality of simulation does however not only depend on the resolution of reanalyses but also on the quality of the model and of the observed data available for assimilation. We could thus expect that data with higher resolution do not necessary always lead to better quality predictions.*

These issues are obviously very interesting and would be worth specific analyses in the future. A comment will be introduced in the perspectives of the modified manuscript.

6. Page 4 line 20: How meaningful are quantities describing instability at 1.125 degrees resolution? and related, if the aim is to do downscaling of climate model outputs or reconstructions how well are the instability and humidity variables simulated by these models, and could the quality of this simulations be an issue for SCAMP? Please comment.

We agree that atmospheric variables describing instability do not give a very good picture of instability when available atmospheric variables are at 1.125° resolution. To our opinion however, they can have some predictive power as a "proxy" of the instability.

The quality of such predictors in climate model outputs or reconstruction is obviously an issue. When applied in a reconstruction context for the whole XXth century, we indeed found that the added value of such predictors was much smaller than when applied with the recent reanalyses available for the last decades. We will add a comment in the discussion on this issue.

7. Page 4 line 20: Be more specific on the predictors used. For example by referring

to table 1 here.

Thank you for the suggestion. We will precise the text as suggested.

8. Figure 1: The caption text is unclear. What is highlighted in black? Is there a reason to use “quantity” and not “amount”? (Same for figures 2 and 9, page 11 line 32, page 13 lines 3, 21,25 and 26, page 15 line 2, page 22 line 6)

We have used “quantity” to specifically refer to the positive values of precipitation. “amount” is often used to describe all precipitation values, including zeroes. Using “quantity” allows us to make a rather clear link between the “quantity model” and the non-zero amounts it is used for. For us, a “precipitation amount” model would have been more confusing as it could have referred to both the zero amounts and the non-zero amounts.

In figure 1, we use equation 1 to decompose $F_{AM}(y)$ *into two parts. We highlight in black the contribution of the empirical cdf of the non-zero precipitation amount to the overall cdf. As expressed in equation 1, this last pdf is “weighted” by the probability of occurrence of probability. We will reformulate the caption of Figure 1 to make it clearer.*

9. Page 5: I’d suggest to add “SCAMP” to the section title of section 3.

This will be done.

10. Page 7 line 6: What does “+12h and +24h UTC” refer to? are this lead times? but then UTC is strange, because time differences don’t have a time zone. Or does it refer to the time of the day? But if so, for which hour is the simulation?

For any given day, SAFRAN precipitation amounts correspond to the cumulated precipitation from 6h00 UTC of the day to 6h00 UTC of the next day, or in other words they correspond to precipitation from 6h00 UTC to 6h00 UTC +24h (where UTC stands for Coordinated Universal Time). We acknowledge that the times used for the geopotential were confusing. We should not have used the "+" symbol. The atmospheric circulation of the current day is described with the geopotential fields at two different UTC times, namely 12h UTC and 24h UTC of that day. The two fields retained to compare days and identify circulation analogs are thus centred over the temporal window used to determine daily precipitation. We will correct the text accordingly

11. Section 3.2: What is the archive length used for the analog model?

The archive length is 1982-2001 which corresponds to the period considered for the predictions with SCAMP and / or the backup AM25 model. We will precise it in the revised version

12. Section 3.2: Which period was used for the optimization of the predictor domains? Is it the same as for the simulation in this work? What are the implications?

The period used for the optimization of the predictor domains is 1982-2001. The period is thus the same as the period of the predictions. The prediction skill of SCAMP presented in our manuscript may therefore be slightly overestimated.

To assess the influence of the optimization period, we could have followed a leave-one-out approach, where for instance, the best analogy domain would have been identified from all years except that of the current prediction day. This would have required *much larger computing resource than those already used for the work presented in the manuscript. The process used for the optimization of the analogy domain is indeed rather long (as mentioned in the manuscript, it is first iterative where different spatial domains of increasing size and considered in turn. The identification of the best analogs days for a given analogy domain is also rather time consuming as a result of the similarity criterion used to compare days). It had also to be applied successively for the 8,981 grid cells of France. This optimization plus the re-estimation for each prediction day of the different regression models considered in the regression stage of SCAMP already required the use of the Grenoble University High Performance Computing centre CIMENT (https://ciment.ujf-grenoble.fr/wiki-pub/index.php/). A leave-one-out approach would have required too many computing resource and was thus not applied here.*

We agree that the optimal domain may depend on the period used for the optimization of the method. We however expect that the domains would be rather similar when obtained from different periods and that their influence on the main results we present in our work would be limited. In a recent work carried out with an Analog Model similar to AM25, we have actually shown that slightly different domains may lead to identify – for a given prediction day, rather different sets of analog dates. We have however also shown that this does not lead to a significant difference in the prediction skill (Chardon et al., 2014). For the context of the present manuscript, an interesting work would be to explore if analogs from different but similar analogy domains would influence the choice of the predictors in the regression stage and/or also if the coefficients of the regression would change. This could contribute to assess the robustness of the approach. We will mention this perspective work in the discussion.

13. Section 3.3 first line: Please specify briefly what the significance conditions are.

We used the 5% significance level for each predictor. This information will be added in the new version of the paper.

14. Section 3.1: are there discrete values drawn from the Gamma distribution for the final prediction? And if so, how?

We aim to model the distribution of precipitation amount, its day dependency, and to further use this distribution as probabilistic prediction. For any given prediction day, we do thus not draw some realization from the distribution.

15. Page 9: I think it is a good thing to look at the skill with respect to climatology as you do, especially for comparison with other studies or methods, but you could have used the AM25 benchmark as P as well, right? Would that be equivalent to your BSS or CRPSS? If not, what is the difference and which one should be preferred under which circumstances?

Yes, we could have chosen the AM25 model as reference for the evaluation of the combined model. We have preferred to use the climatological reference as this allows for normalized scores which can be compared, as mentioned by the referee, with those obtained in other studies.

We compared the combined approach and the AM25 model with gains in skill scores estimated for both approach with respect to the climatology. Such gains, given in terms of BSS or CRPSS percentage points, are also widely used to compare different prediction models. They thus present the advantage to be rather easy to interpret.

16. In section 4.1 you describe several steps of restrictions applied in terms of the candidate predictors for the sake of robustness and clarity of the article. I appreciate these goals, but at present the description is a bit confusing and it remains unclear

Printer-friendly version

Discussion paper

which of these restrictions are a feature of SCAMP and would be kept for a general application of SCAMP and which ones aren't and what would be the potential impact on robustness and skill.

The issue of predictor restrictions is an interesting point which was not easy to tackle. The main goal of those restrictions was indeed to improve the clarity of the article. The manuscript does thus not present a definitive configuration of SCAMP but more a proof of concept for an adaptive model which could use a much larger set of potential predictors, when relevant.

The impact of fewer restrictions on the robustness of the method is potentially an important issue and would be worth a detailed analysis. This will be suggested as a perspective of the work.

17. Page 13 line 8: The phrase is very unclear. Please rewrite.

Thanks for this remark. We will rewrite the phrase.

18. Page 13 second paragraph: What exactly causes the GLM to "fail" in the southeast for the occurrence? Are there not enough wet analogues to estimate the occurrence probability or does it fail the significance test for the parameters? please comment.

We agree that the text of the second paragraph of p. 13 is somehow confusing. As mentioned in the paragraph, the GLM which models the occurrence does actually almost never fail, even in the southeast. From the sum of the frequencies obtained for the two cases "case 2" and "case 4" of Figure 5, we can see that, whatever the region, the GLM which models the occurrence is indeed activated most of the time (more than

97% of the days) (remember that case 2 corresponds to (Success of GLM modeling the occurrence + Failure of GLM modeling the quantity); case 4 corresponds to (Success of GLM modeling the occurrence + Success of GLM modeling the quantity). Consequently, and whatever the region, the situation where AM25 is used as backup for the prediction of the occurrence probability is very rare (see the sum of the frequencies obtained for case 1 and case 3).

Nevertheless, Figure 5 indeed highlights a very specific behavior in the southeast when compared to the remaining of France. Case 2 is activated much more often in this region (increase of 30% percentage point) than elsewhere and, in a symmetric way, Case 4 is activated much less often than elsewhere (decrease of 30% percentage points).

The reason underlying this result is to be related to the much higher proportion of dry days in southeast as illustrated in figure R1a below. For a number of predictions days, the number of analog days that are wet is indeed to be small in the southeast. This is obviously not a difficulty for the estimation of the GLM modeling the occurrence. This is conversely likely one for the estimation of the GLM modeling the quantity. For days for which the number of wet analogs is small, the size of the dataset available to fit the GLM modeling the quantity can be too small to allow for a fit with significant parameters. This very likely explains the spatial disparities in both graphs "Case 2" and "Case 4" of Figure 5.

In our work, a GLM (GLM modeling the occurrence or GLM modeling the quantity) was said to fail for a given prediction day when the significance test failed for the parameters. The link of these failures with the number of wet/dry analog days is not as direct as we could have presented it in the text. It is however strongly suggested from Figure R1a and from what is described in the different graphs of Figure 6. We will reformulate this section for clarification.

Figure R1a. Probability of a dry days over the 1982-2001 period (percentage of days which are dry) (see supplementary material)

19. Page 15 line 2-3: The predictor set optimized for the whole of France? I thought they were optimized for each grid cell and time step. Is this only for this experiment or in general? This is confusing and will hopefully get more clear with a restructured version of sections 2 and 4.1.

We agree that the text is somehow confusing. For the sake of clarity, a single set of potential predictors was used for all grid points of France (see section 4.1). The most interesting set may be however rather different from one region to the other. This is a possible reason for which we obtain no gain in southern France when we activate the quantity model. This should become clearer with the restructured version of the manuscript.

20. Page 16 line 20: Please quantify which proportion of days you would consider as “reasonable”.

We had actually not in mind to suggest a “reasonable” proportion of days, which could be used to retain a reduced number of regression sets. Considering a reduced number of regression sets would obviously allow for reducing the computational time required for the model identification/evaluation.

As mentioned in the next sentence of the manuscript, this may however limits the possibility to achieve a better prediction for some (rare) events which would activate very unusal predictors. This is what is highlighted with some of the graphs in the discussion section. We will clarify this point in the revised manuscript version.

21. Looking at figure 9, I wonder if the high frequency of the AM25 model in the south-east might be related to the Gamma distribution being a suboptimal approximation of the precipitation amount distribution in this region. Did you test this?

Thank you very much for this input. The distribution of non-zero precipitation amounts is indeed rather different in the southeast region than that of the other regions of France. The main reason is the existence of much more frequent and intense heavy precipitations in southeastern France (this is the reason why a lot of works have focused in the last decades on precipitation extremes in southeastern France). This is suggested with the much higher Coefficient of Variation of daily precipitation in this region (cf. figure R1b below).

Figure R1b. Variation coefficient of daily precipitation (see supplementary material)

The gamma distribution is obviously flexible and widely used in the hydro-meteorology literature to model strictly positive precipitation. It may be however not optimal in this specific context and a distribution with a heavy tail would be probably more appropriate (e.g. the extended GPD distribution introduced by **?***). This may improve the prediction but this may also lead to estimation difficulties as a more complex distribution would require more data for a robust fit. Here, the number of data is voluntarily limited to the number of analog dates considered for the fit. A more flexible model (with more parameters) would require considering more analog dates. This may be detrimental for the prediction skill as poorer analogs would be integrated in the set of dates used for the estimation of the regression. Another possibility would be to fit a more flexible model with the same number of analog dates than in the present case. Due to the estimation problems mentioned above, this would however likely lead to a much more frequent use of the backup analog prediction model in our case. Such an analysis would be worth to do and this issue will be mentioned in the perspective section.*

22. Why is figure 10 a line chart? There is no order in the WPs, is there? I'd recommend to transform this in a series of bar charts (one for each WP). This would further avoid all the colors and line types and thus solve the issue with the invisible (probably yellow) dotted line for R700+H+Occ-1 in a) and R700+T700+W700 in b).

We agree that a series of bar charts would have been more relevant and we will change the figure for this representation.

23. Depending on the intended use of SCAMP, the temporal structure of the simulated precipitation might be relevant. I suppose that a detailed analysis of the representation of the annual cycle, the autocorrelation and the interannual variability in both SCAMP and AM25 is beyond the scope of the present paper, especially since this is not straight forward for probabilistic simulations, and you might have a look at this in future work, but could you make a statement on the overall variance of the SCAMP simulations as compared to the benchmark and the observations? Typically analog models reproduce the observed variance quite well while deterministic regression models suffer from reduced variance. Since SCAMP is a hybrid model it would be interesting to know which characteristics it "inherits".

In a context where time series have to be simulated, additional criteria would be actually relevant to evaluate / compare the different modelling approaches; they should especially include as mentioned by the referee, the ability to reproduce the observed variance of precipitation from daily to interannual time scales.

We will add a comment on this in the discussion section.

Deterministic regression models are indeed known to underestimate the variance of

Printer-friendly version

Discussion paper

daily precipitation. However, scenarios obtained with deterministic regression models disregard the variance of the residuals in the regression. However, regression models can be also used in a stochastic simulation framework, where a random variable is drawn from the statistical distribution associated to the regression. In **?** *for instance, such a generation process was used, to first identify if the prediction day was wet or dry (based on a random variable compared to the occurrence probability obtained from a first occurrence model) and to next generate some precipitation amount (based on a random realization within the gamma distribution used to model the distribution of precipitation amount in case of a wet day). The observed variance of daily precipitation was well reproduced.*

If time series would have to be simulated with SCAMP, a similar stochastic generation process would be followed for the regression stage. We thus would expect that the variance of observed daily precipitation would be rather well reproduced as in **?***. The regression stage is also not expected to increase the variance that would be obtained for the benchmark analog (in a configuration where one of the k-nearest analogs is randomly sorted each day and used as weather scenario for the day). This is one of the preliminary results we obtained for a similar work we currently develop in western Switzerland.*

24. page 22 lines 16-20: This part is not clear, please rewrite.

We will reformulate this paragraph.

25. page 22 lines 28-32: I don't understand what "classically" means in this part. please use some more precise wording.

The word "often" is indeed more suited there.

26. page 22 line 35: This sentence is not clear to me. In what sense is the set of days homogeneous?

The days are homogeneous with respect to their large-scale atmospheric circulation configuration. We will clarify this.

27. page 23 line 2: The sentence is not clear to me. Which context? and who leaves room for improvement?

The context we wanted to refer is that of very specific meteorological configurations that may be observed from time to time and for which the usual predictors are sub-optimal. The analog/regression approach presented here is expected to allow for the identification of better suited predictors and thus for an improved prediction skill for the prediction day under consideration. We will rephrase this sentence in order to make it clearer.

28. Is SCAMP transferable to other regions or countries? To what extent? Under which circumstances would it be necessary/unnecessary to redo the predictor selection? Please comment.

SCAMP is indeed transferable to other regions. We will add a comment on this in the discussion. The development of SCAMP for another context would obviously first require the identification of the best predictor set. In a number of previous studies, the most relevant predictors to be used for statistical downscaling are indeed found to be region dependent (e.g. Cavazos and Hewitson, 2005; Timbal et al., 2009; Chardon *et al., 2014). Similar conclusions have likely to be expected for SCAMP also. Part of these conclusions may also result from the occurrence frequency of weather regimes observed for the different regions. They can actually vary a lot from one region to the other. In such a case, the best predictors identified for a given region could partly be a result of the fact that the weather situations for which they are more relevant are more frequently observed for that region. This issue would be worth future investigations.*

29. It would be helpful to mark or highlight the predictors that were preselected for the occurrence and amount models respectively in table 1.

Thank you for the suggestion. This will be marked in table 1.

3 Technical corrections

We thank the reviewer for the very careful reading of the paper and for all the technical corrections pointed out here. The text will be corrected accordingly.

Cavazos T, Hewitson BC. 2005. Performance of NCEP-NCAR reanalysis variables in statistical downscaling of daily precipitation. Climate Research 28: 95–107.

Chardon J, Hingray B, Favre AC, Autin P, Gailhard J, Zin I, Obled C. 2014. Spatial similarity and transferability of analog dates for precipitation downscaling over France. J. Clim. 27: 5056–5074, doi: 10.1175/JCLI-D-13-00464.1.

Leutbecher, M.; Palmer, T. N. 2008. Ensemble forecasting. JOURNAL OF COMPUTATIONAL PHYSICS. 227(7). 3515-3539.

Maraun, D., et al. 2010a: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Rev. Geophys., 48,RG3003, 20 doi:201010.1029/2009RG000314, 2010.

Maraun, D., Rust, H. W., and Osborn, T. J.: 2010b. Synoptic airflow and UK daily precipitation extremes Development and validation of a vector generalised linear model, Extremes, 13, 133–153, doi:10.1007/s10687-010-0102-x.

Mezghani, A. and Hingray, B., 2009. A combined downscaling-disaggregation weather generator for stochastic generation of multisite hourly weather variables in complex terrain. Development and multi-scale validation for the Upper Rhone River Basin. J. Hydrology. 377 (3-4) : 245-260.

Naveau, Philippe; Huser, Raphael; Ribereau, Pierre; et al. 2016. Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. WATER RESOURCES RESEARCH. 52(4) 2753-2769

Radanovics, S., Vidal, J.-P., Sauquet, E., Ben Daoud, A., and Bontron, G.: Optimising predictor domains for spatially coherent precipitation 5 downscaling, Hydrol. Earth Syst. Sci., 17, 4189–4208, doi:10.5194/hess-17-4189-2013, 2013.

Raynaud, D., Hingray, B., Zin, I., Anquetin, S., Debionne, S., Vautard, R. 2016. Atmospheric analogs for physically consistent scenarii of surface weather in Europe and Maghreb. Int. J. Climatology. doi :10.1002/joc.4844.

Timbal B, Fernandez E, Li Z. 2009. Generalization of a statistical downscaling model to provide local climate change projections for Australia. Environ. Model. Software 24(3): 341–358.

Yee, TW; Wild, CJ. 1996. Vector generalized additive models. JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES B-METHODOLOGICAL. 58(3) 481-493