Hydrology and
Earth System
Sciences

Open Access

Discussions

EGU

HESSD

Interactive
comment

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

Anonymous Referee #1

Received and published: 19 May 2017

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

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).

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.

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.

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.

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.

2 Specific comments

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

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

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

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

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

C4

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

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

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

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.

7. Page 4 line 20: Be more specific on the predictors used. For example by referring to table 1 here.

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)

9. Page 5: I'd suggest to add "SCAMP" to the section title of section 3.

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?

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

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?

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

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

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 AM_{25} benchmark as P_ϕ as well, right? Would that be equivalent to your ΔBSS or $\Delta CRPSS$? If not, what is the difference and which one should be preferred under which circumstances?

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 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.

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

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.

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.

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

21. Looking at figure 9, I wonder if the high frequency of the AM_{25} 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?

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).

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 AM_{25} 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".

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

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

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

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

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.

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

3 Technical corrections

1. page 1 line 4: ...and the regression link are likely to also vary in time. → ... and the regression link are likely to vary in time too.

2. page 1: does SDM stand for statistical downscaling **methods** (abstract) or statistical downscaling **models** (Introduction, e.g., line 24)? Please unify.

3. Page 2 line 1: Perfect Prog → Perfect Prognosis

4. Page 2 line 8-9: Another possible reference in the reconstruction context: Caillouet et al. (2016)

5. Page 2 line 9: weather numerical models → numerical weather prediction models

6. Page 3 line 6: ...a set of parameters is classically estimated for each weather-type... → ...a set of parameters is estimated for each weather-type...

7. Page 3 line 12: could consist → consists

8. Page 4 line 1: The paper structures... → The paper is structured...

C8

9. Page 4 line 1: You forgot section 3 in the list.

10. Page 4 line 5: metropolitan French territory → mainland of France (also Page 10 line 11. "metropolitan territory" could be misunderstood as areas occupied by large cities)

11. Page 4 line 10: 6-hour → 6-hourly

12. Page 4 line 12: predictors are 1000 and 500 HGT geopotential fields... → predictors are the 1000hPa and 500hPa geopotential height fields...

13. Page 4 line 19: ...the water atmosphere content... → the atmospheric water content (or total column water content?)

14. Page 4 line 20: we also consider the occurrence of precipitation for the previous day. → we also considered the occurrence of precipitation on the previous day.

15. Page 4 line 29: it's "dependent on" but "independent of"

16. page 7 line 5: the similarity is assessed by the TWS → the similarity is assessed using the TWS

17. Page 7 line 6: applied to the geopotential 1000 hPa and 500 hPa respectively at +12h and +24h UTC → applied to the geopotential height at 1000 hPa and 500 hPa at +12h and +24h UTC respectively.

18. Page 7 line 7: low and high pressures → low and high pressure systems

19. Page 7 line 12: → A different analog model was thus considered for each of the 8981 SAFRAN grid cells.

20. Page 7 line 15: we use the 100 → we used the 100

21. Page 7 line 15: identified with AM → identified with the AM

22. Page 7 line 17: we will also consider → we also consider

23. Page 7 line 20: This analog model ... will give a benchmark → This analog model ... is used as a benchmark

24. Page 7 line 20-21: They will be additionally used → In addition they were used

25. Page 8 line 12: and possibly benchmark → and benchmark

26. Page 6 line 3: pi is modeled → pi was modeled

27. Page 6 line 6: we use a GLM → we used a GLM

28. Page 6 line 7: is here expressed as → is therefore expressed as

29. Page 6 line 32: F_Y is finally obtained → F_Y is obtained

30. Page 8 line 19: to evaluate some Ensemble Prediction System → to evaluate Ensemble Prediction Systems

31. Page 9 line 3: y is here practically described → y is here described

32. Some more in section 4.1

33. Page 11 line 21: Figure 4a similarly presents → Figure 4a shows

34. page 11 line 26: Similarly to what was obtained with the BSS gain → similarly to the BSS gain

35. Page 11 line 30: dependency to regional features → dependency on regional features

36. Page 12 line 1: no apostrophe in SCAMPs behavior. (Same page 13 line 27)

37. Page 12 line 2: I don't understand "to updating" in this context.

38. Page 13 line 6: predict occurrence probability → predict the occurrence probability

39. page 13 line 11: it is met → it occurs

40. page 13 line 19: the proportion ... is next too low → the proportion ... is therefore too low

41. page 15 line 6: not necessary → not necessarily

42. page 16 line 2: we assess the frequency each structure has been selected → we assess how often each structure has been selected

43. page 16 line 3: This allows us to give some insight → this allows for some insight

44. page 16 line 3: information really used for the regression stage → information used in the regression stage

45. page 16 line 13: region depend → region dependent

46. page 16 line 20: This would be however not necessary relevant or desirable." Strange sentence, that could be deleted starting the next sentence with "However" for example.

47. page 19 line 23: For some few frequent WPs → for some less frequent WPs

48. page 19 line 24: is much more selected → is more frequently selected for WP7

49. page 19 line 26: for this really few frequent WP → for this rare WP7

50. page 19 line 29: less frequent WPs as WP7 → less frequent WPs such as WP7

51. page 19 line 29: this suggests the spatial robustness → this suggests spatial robustness or this suggests some spatial robustness

52. page 20 line 2: The interest of a hybrid model → the relevance? of a hybrid model

53. page 20 line 3: analogs of the prediction day are identified to estimate a two-part regression model → analogs of the prediction day were identified to estimate the parameters of a two-part regression model

54. page 20 line 5: this hybrid approach actually updates → this hybrid approach updates

55. page 22 line 1: that would have been achieved directly with a → from a

56. page 22 line 4: compared to the prediction that could be directly obtained from the atmospheric analogs of the analog stage → compared to the reference analog model

57. page 22 line 5: CRPSS skill scores → CRPSS

58. page 22 line 6: thanks to → due to

59. page 22 line 8-9: seems also to be decisive → seems to be decisive as well

60. page 22 line 9: twice higher then → either it is "twice as high as" or "two times higher than" which is by the way equivalent to "three times as high as"

61. table 1: baroclinicity → baroclinity

62. Make sure that especially the figures 2, 7, 8, 9 and 11 are rendered large enough in the final version, that is at least not smaller than now.

References

Caillouet, L., Vidal, J.-P., Sauquet, E., and Graff, B. (2016). Probabilistic precipitation and temperature downscaling of the twentieth century reanalysis over france. *Climate of the Past*, 12(3):635–662.

HESSD

Interactive comment

Printer-friendly version

Discussion paper