Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2020-352-AC1, 2021

Interactive comment on "Improving snowfall representation in climate simulations via statistical models informed by air temperature and total precipitation" *by* Flavio Maria Emanuele Pons and Davide Faranda

Flavio Maria Emanuele Pons and Davide Faranda

flavio.pons@lsce.ipsl.fr

Received and published: 14 March 2021

General Comments: In this manuscript, the authors test existing methods for the estimation of the apportionments of rain/mixed/snowfall and present a more robust way of selecting threshold temperatures to be used in single/multiple threshold models. The tests are performed based on the reanalysis ERA5 dataset and the IPSL_WRF climate projection model for the period 1979-2005, and at 0.25-degree spatial resolution in Europe, including the Scandinavian peninsula. Daily temperatures are used for the

models. In general, I find that the authors did a good job explaining the different methods and performance analyses, and as a data exercise, the procedures are somehow clear in the manuscript. However, I find that there is a disconnect between the objectives and how realistic it would be to apply these models over such domains and spatial resolution.

A: We are aware that these statistical models have been initially formulated for, and applied to, the estimation of snowfall from station data, and that it would be ideal to work at very small scales to catch all modes of spatio-temporal variability of any given phenomenon (not only in atmospheric sciences).

This is true at all levels, in fact any simulation model for meteorology, climate, oceanography, land-atmosphere interaction and so on is integrated at a finite time step that necessarily cuts off high temporal frequencies, and over finite grids that require sometimes coarse approximations of the sub-grid scale processes. It is not ideal, but if scientists in the climatology field chose to wait until it will be possible to integrate climate models at the molecular dissipation scales, present climate studies would not be able to address much more than global averages. As climatologists, we must deal with these approximations, and do our best to mitigate their effects.

Q: I also find that the use of the reanalysis data as an "observation" (e.g.,Figure 5) to test all the models against can easily be challenged given the uncertainties in such datasets, especially to determine snowfall in complex terrain. I would like to point out that the type of models that the authors are applying are generally derived from meteorological data at in-situ stations, while the article refers to modeling grid scales of several km (0.25 degrees, which at 70 deg lat isâĹij10 km inE and 28 km in N coordinates, https://www.opendem.info/arc2meters.html).

A: We agree that the term "observation" is incorrect in this context, and will be replaced by "reanalysis" or "reference dataset" in the future version of the paper.

Q: How can the same type of models be appropriate for both scales? A gridcell of such dimensions would easily cover a mountain valley and surrounding mountains, with a very wide elevational range. Throughout my reading, I just kept wondering how such models could be applied in this spatial context. One would expect that over the domain of a single gridcell, one would potentially encounter a portion of the area to receive rainfall, while another portion would receive snowfall and a middle portion would be in the transitional zone if an event occurs around the freezing point. I find this to be very problematic in the context of the manuscript. Another relevant issue has to do with the use of daily temperatures for the models,given that sub-daily temperature fluctuations would have marked effects on precipitation phase apportionments.

A: The reviewer has concerns about the idea that these models are applied to gridded datasets with a relatively coarse resolution and daily frequency. However, as also mentioned in the literature review in Section 1, this type of model is already applied in the recent climate literature focusing on snowfall. For example, Bai et al., 2019 consider observations projected on a 0.25° grid, while Chen et al., 2020 use a sigmoid function to estimate snowfall data from GCM simulations, at a 1.5° resolution. A comparison of the performance of the simple single threshold method to estimate snowfall over Europe, compared to E-OBS, can be found in Faranda 2020. All of these studies consider daily data, as well as others dealing with station data, such as Liu et al., 2018.

One may rather wonder if, given the level of approximation present in climate simulations, the use of S-shaped functions is inappropriate, and simpler models such as the binary threshold should not be applied instead. This is indeed the research question we addressed in the paper: the results in terms of reconstruction of the snow in the reference period suggest that models admitting a nonlinear function provides better performances than naive binary apportionment or simple linear regressions, and we think that this point is clearly proven by the results discussed in Section 4 of our paper.

This happens, despite all the concerns about neglected complexity due to sub-grid scales, because the relationship between snow fraction and temperature can be seen

as a transition between two fixed boundaries (i.e. a snow fraction equal to 0 and 1), regardless of the coarse spatio-temporal resolution. Most smooth transitions of this type result in S-shaped relationships, regardless of the scales involved. For example, analogous S-shaped transitions can be observed when studying turbulence in a nocturnal stable boundary layer (Van de Wiel et al., 2017), involving different processes and much smaller scales than in snow hydrology.

We do not find this fact particularly surprising. Statistical models are always wrong with respect to the reality of the phenomenon they try to catch: their efficacy must be measured by how useful they are in terms of the goal, in this case the reconstruction of an unobserved variable from two observed covariates (technically, a prediction task). Since our goal is purely predictive, the effectiveness of the method should be judged based on their prediction performance, and not comparing it to the way it is used in different contexts.

The fact that the same statistical model can be used at very different scales (or even in completely different fields, or branches of a field) is not an exception in quantitative research. A time series model featuring intermittency can catch features of the generating process in small-scale turbulence as well as at the climate scale; econometric models can be applied to study the micro-performance of a single agent as well as of entire countries; exponential or gamma laws describe the waiting times between two quantistic as well as macroscopic events; Lotka-Volterra equations can be used to model the relationships both between populations of large prey-predators, and between small pathogens such as phages and bacteria, et cetera.

The nature of the phenomenon suggests that an S-shaped relationship could perform better than anything less smooth such as a binary threshold, and we believe that our results show that this is indeed a realistic conclusion.

Q: Also, what would the near surface temperature be representative of in the manuscript? A mean elevation? If so, would this be realistic? I argue it is not, especially to illustrate how the proposed methods can enhance the estimation of the model's parameters.

A: As standard in gridded climate simulations and reanalysis, the near-surface temperature is representative for the mean elevation in the grid cell. Again, we understand how these scales may seem disproportionately large when compared to the domain of a single weather station, but this logic would basically exclude using climate simulation models for anything but computing yearly/global statistics.

Q: As snow hydrologists, in our group's modeling efforts we use a similar two-temperature threshold model to estimate the precipitation apportionments at grid scales between 10-100 m in mountainous terrain, with a linear model between the two thresholds. Even at such scales, we understand that there are drawbacks to such model, but the uncertainties in precipitation amounts and temperatures are primary, and the model for the precipitation apportionments takes a secondary role. However, even at such scales, determining the performance of the model and parameters is very challenging because of the difficulties in determining accurate precipitation amounts, particularly snowfall.

A: We find that this consideration indeed marks one of the greatest differences between a climate study and a small scale study. Of course there is a lot of uncertainty also in the performance of reanalysis models, in addition to the coarse graining of the process. However, when conducting an analysis on climate simulation model outputs, the reference period is chosen to be a previously validated gridded observation or reanalysis dataset. Given the validation, the dataset is taken as the reality, and used to bias correct the climate models. It is even possible to conduct a so called "perfect model" experiment, where the output of a climate model is used as the reality to check the performance of the bias correction on another model.

While we remain aware of the limitations of reanalysis datasets, it is once again a level of approximation that we must accept, as the alternative would be to drop climate

studies altogether.

In such a framework, after validation against another reliable dataset, we neglect uncertainty in the reanalysis and we use it for the bias correction of the climate models. At this point, we assume that the bias corrected variables in the model are unbiased. This removes the issues of observational error in each model, and the use of model ensembles (at least tries to) mitigate the effects of the model-specific approximations.

Clearly, the resulting outputs cannot be considered representative of sub-grid scales. However, we do consider the bias corrected variables representative for the resolved scales. The issue with snowfall, as stressed in the paper, is that it poses more challenges than other variables in terms of direct bias correction, while the raw data are not realistic, and they also become incompatible with temperature and total precipitation once these two variables are bias corrected. Our objective is not to use models that represent correctly subgrid scales, but simply to leverage on variables that are relatively easy to adjust and provide a better reconstruction of the snowfall compared to raw data.

We remark again that our goal is purely predictive and the effectiveness of the method should be judged based on results in these terms, and not comparing it to the way it is used in different contexts. Not only results discussed in Section 4.2 - 4.3 clearly point to an improvement of the representation of snowfall in these terms, but such improvement is particularly dramatic over areas characterized by complex orography, as shown by the case studies focusing on the Alps and Norway, showing that indeed these models work well even at a coarser level.

Q: Threshold values also seem unrealistic in some locations, as low as -15 or as high as 5+ C. There are examples of snowfall events at high temperatures (e.g., late June 2019 summer events in Colorado), but as a generalized modeling threshold it would seem unrealistic to have such high and/or low values.

A: We agree, in many areas the estimated thresholds are not directly interpretable in meteorological terms. However, our objective is not to obtain results that reproduce realistic microphysics, but to estimate models that provide the best predictive power. The choice of the knots for spline regression is completely arbitrary: for example, one common choice is to use the deciles of the covariate distribution. We choose to use breakpoint analysis with up to 2 thresholds because we expect relatively smooth and monotonic relationships, but we do not imply that the recovered thresholds are necessarily realistic for every grid point.

We also stress that we presented results from two different techniques, showing that actually Eq. 11 provides thresholds that are more physically realistic, while the unadjusted breakpoint analysis provides less interpretable values. Please also notice that, since we work with standardized anomalies (due to the different scales of the variables involved in the regressions) as specified in the article, these values are not meant to be interpreted as absolute temperature, but as deviations from the long-term climatology.

Q: This would also highlight the issues regarding the data and spatial scales of the analysis. Because of these issues, I am recommending that the manuscript be rejected. I ultimately consider that the results in the manuscript do not accomplish demonstrating how the proposed methods deliver improvements in model accuracy.

A: We are frankly puzzled by this decision, and especially by the way it is justified. The only comment actually referring to the results concerns the threshold temperatures, an instrumental value obtained at the beginning of the modelling procedure, and not one of the primary objectives of our analysis. This last comment leaves us with the impression that the reviewer formed their opinion based on their previous knowledge about these models (applied in a different context) and did not take into proper account the results that should indicate whether or not we answered the initial research question. We underline once again that we propose an improvement of techniques that are already used for similar or analogous tasks, so rejecting the present paper on this basis means to also challenge part of the existing literature.

Q: Specific Comments:

A: Thank you, we will make sure to make all the corrections listed below.

ll. 35-36: Odd sentence.l.

52: typo in “observationa”.

ll. 74 & 79: I suggest using past tense to refer to findings or proposed models, as theones in these lines (Pipes and Quick (1977) and L’hôte et al. (2005)).

l. 116: typo “Section ??”

l. 122: I suggest adding a comma before “which”, but the sentence would need changes.

l. 125: what are you calling “large scales” here? Suggest clarifying.

l. 141: “worth mentioning”.l.

291: suggest changing “It is easy to prove” to “It can be shown”.

l. 259: Revise “The latter resulted not enough numerically stable”.

References

Bai, L., Shi, C., Shi, Q., Li, L., Wu, J., Yang, Y., ... & Meng, J. (2019). Change in the spatiotemporal pattern of snowfall during the cold season under climate change in a snowâĂŘdominated region of China. International Journal of Climatology, 39(15), 5702-5719.

Chen, H., Sun, J., & Lin, W. (2020). Anthropogenic influence would increase intense snowfall events over parts of the Northern Hemisphere in the future. Environmental Research Letters, 15(11), 114022

Liu, S., Yan, D., Qin, T., Weng, B., Lu, Y., Dong, G., & Gong, B. (2018). Precipitation phase separation schemes in the Naqu River basin, eastern Tibetan plateau. Theoretical and applied climatology, 131(1), 399-411.

Van de Wiel, B. J., Vignon, E., Baas, P., van Hooijdonk, I. G., van der Linden, S. J., Antoon van Hooft, J., ... & Genthon, C. (2017). Regime transitions in near-surface temperature inversions: A conceptual model. Journal of the Atmospheric Sciences, 74(4), 1057-1073.

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2020-352, 2020.