Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2019-49-AC1, 2019


Nonlinear Processes in Geophysics
Discussions
Open Access

# *Interactive comment on* "Remember the past: A comparison of time-adaptive training schemes for non-homogeneous regression" *by* Moritz N. Lang et al.

**Moritz N. Lang et al.**

moritz.lang@uibk.ac.at

Received and published: 10 December 2019

*This manuscript compares the effect of different schemes to compose training data for statistical post-processing methods (here: non-homogeneous regression) on the performance of the resulting forecasts. It is well written and highly relevant to operational forecasting where availability of reforecast data may be limited and the consequences of changes in the NWP model on forecast calibration must be understood in order to decide whether forecasts from an older NWP model version can be used to fit the parameters defining the post-processing model. This last point is the only one where I feel the manuscript could benefit from a more detailed discussion.*



We want to thank you for your fruitful and constructive review. We have been carefully going trough your comments to address each point individually including your general comment on NWP model changes. According to your suggestions and the comments of reviewer 2, the most substantial changes in the manuscript are the following:

- The main goal of the article is now more clearly stated in the manuscript. The objective is to cover a wide range of methods as proposed in the literature – rather than finding the universally best method – in order to provide guidance on strengths and weaknesses of the underlying strategies. Therefore, to show a wide spectrum of possible approaches in a unified setup, we consider typical basic applications of these training schemes and refrain from more elaborate tuning or combinations. We have adjusted the introduction (Sect. 1), the conclusion (Sect. 4) and the corresponding paragraphs in the methodology (Sect. 2.2.2) accordingly.
- We have added more information on the 2016-03-08 change in the horizontal resolution of the ECMWF EPS (cycle 41r2). This specific change was chosen to construct the data sets A-C because it is likely to affect coefficient estimates more substantially. We also now clarify that, in fact, further model changes occurred in the time periods considered but that these did not affect the horizontal resolution and hence can be expected to have much smaller effects on the coefficient estimates.
- An additional comparison of the different time-adaptive training schemes has been performed on daily precipitation sums employing a left-censored Gaussian model for post-processing. All results are very similar to the analyses for the 2 m temperature forecasts presented in the manuscript and hence nicely support the conclusions given in the paper. Therefore, we feel that it is not necessary to report these additional results in the main manuscript but we do include them in an online supplement.

Below, you can find a detailed point-to-point reply to your comments and suggestions.

*Specifically:*

*The CRPS skill scores in Fig. 5 h) suggest that the regularization scheme struggles with the adjustment to the NWP model upgrade and to the annual cycle, but also the SW plus and the smooth model have an overall neutral effect on skill even though these schemes increase the training sample size significantly. It would be interesting to better understand the causes of this result.*

Figure 5 h shows the CRPS skill scores for alpine sites for data set B. Thus, the *smooth model* is trained on the 'old EPS version' while the predictions are for the 'new EPS version' which explains its relative performance loss. This is also true for the *sliding-window plus* which is, in large parts, based on data from the 'old EPS version'. The classical *sliding-window* approach and the *regularized sliding-window* approach both adjust to the 'new EPS version' more rapidly at the same pace and, hence, show a similar predictive performance difference as for data set A (Fig. 5 g).

*Figure 3 gives some good idea about the problems with the regularization scheme (parameters adjust very slowly to changes) but it is not ideal to illustrate problems with 'SW plus' and the 'Smooth model' since no NWP model upgrade happens in data set A. Wouldn't it be better to use data set B for this figure, where we can expect some adjustment during the first days/weeks of the validation period? Also, is Innsbruck the best location to illustrate the effect of a NWP model upgrade? As 'best' alpine location in this context I would consider the one that is most strongly impacted by the horizontal resolution change (this could be studied by considering changes in biases in the raw ensemble forecasts) in the ECMWF model and therefore presents a worst case scenario in terms of adjustment to a NWP model upgrade.*

Figure 1 shows estimated coefficient paths for the weather station Altdorf for the first calendar year of the validation period of data set B. Altdorf is the station which is most strongly affected by the model change with a height difference in the model topography



of 47.4 m. The first 40 days within the validation, which correspond to the period directly after the EPS change, is highlighted in pink. The variability of the coefficients within the first 40 days compared to the rest of the year are in the same order and hence no clear adjustment can be detected.

Despite the well reasoned comment, the analyses for data set B provide no further insights to the adjustment phases of the *sliding-window* and *regularized sliding-window* approaches. Hence, to restrict the presented analysis to the error sources (v) and (vi) described in Sect. 2.1 of the manuscript, we suggest to keep showing the coefficient paths for data set A in the paper. Consequently, as for data set A no EPS change has to be considered, we have kept the results for Innsbruck in the manuscript.

*I would encourage the authors to provide some more discussion along these lines, since NWP model upgrades have been the main argument to justify the need for re-forecasts, and I am not aware of any previous study that looks at the effect of NWP model upgrades on the performance of post-processed ensemble forecasts in a quantitative way.*

We agree that analyzing the effects of EPS changes would be a very interesting research question on its own. We have tried to account for this issue by our study design (data set A, B, C), however, for a comprehensive perspective on this topic one would need to perform an extensive analysis on more than 15 stations. This is beyond the objective of this paper which mainly aims at presenting how time-adaptive post-processing scheme are related to each other and how these perform under specific restrictions, such as the EPS change which affects the horizontal resolution investigated in our current study. Other studies focus more on investigating the effect of model changes themselves such as, e.g., Demaeyer and Vannitsem (2019) by studying the impact of changes in a quasi-geostrophic model on post-processing.

To address your initial comment on when it is beneficial to use data from a previous NWP model for forecast calibration: As the results in our study show, the time-adaptive training schemes using multiple years of data are superior to the ones using the most recent days only, even in case of the EPS change investigated. However, this might look different for a different NWP model and/or future model changes and must be evaluated individually in each case. This is now explicitly stated in the conclusion of the manuscript (Sect. 4).

*It is interesting to see in Figs. 3 and 4 that this type of regularization seems to introduce too much inertia, i.e. the parameters only adapt with a certain delay or sometimes not at all. Have the authors tested alternative stopping criteria? A simple and obvious variant would be to perform 2 or 3 iterations on each (except the first) new day.*

Figure 2 shows the temporal evolution of regression coefficients exemplary for Innsbruck using three instead of only one iteration. In comparison to the *regularized sliding-window* model version presented in Fig. 3 of the manuscript, the temporal evolution of the coefficients for the modified *regularized sliding-window* approach is much more comparable to the evolution of the classical *sliding-window* approach. This increased similarity is also visible in the aggregated CRPS skill scores shown in Fig. 3. In comparison to original *regularized sliding-window* approach discussed in the manuscript, the modified *regularized sliding-window* approach has both less profound performance losses for alpine stations, and less visible performance gains for stations in the plain and foreland with the classical *sliding-window* approach as a reference.

In summary, three iterations used in the estimation process for each new day is not generally superior to a single iteration for the employed data set. To show a wide spectrum of commonly-used training schemes in a unified setup, we kept the different approaches as close as possible to the originally proposed version to show their advantages and possible disadvantages Thus, we refrain from introducing further modifications such as e.g., additional hyperparameter tuning as now stated in the conclusion (Sect. 4). In addition, we now explicitly emphasize in Sect. 2.2.3 of the paper that an increased number of iterations might be more appropriate for the *regularized sliding-window* approach depending on the employed data.



*Minor comments:*

*244-245: While it's possible (even likely) that a larger slope coefficient is due to higher skill of the EPS temperature forecasts, one cannot be sure if at least to some extent the larger slope coefficient is due to an amplitude bias of the raw ensemble forecasts, i.e. the ensemble underpredicts high temperatures and overpredicts low temperatures, and increasing the slope coefficient compensates for that.*

A very good remark! We have corrected the statement in Sect. 3.1 according to your comment.

Demaeyer J, Vannitsem S (2019). Correcting for Model Changes in Statistical Post-Processing – An approach based on Response Theory. *Nonlinear Processes in Geophysics Discussions*, **2019**, 1–27. 10.5194/npg-2019-57.

Interactive comment on Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2019-49, 2019.

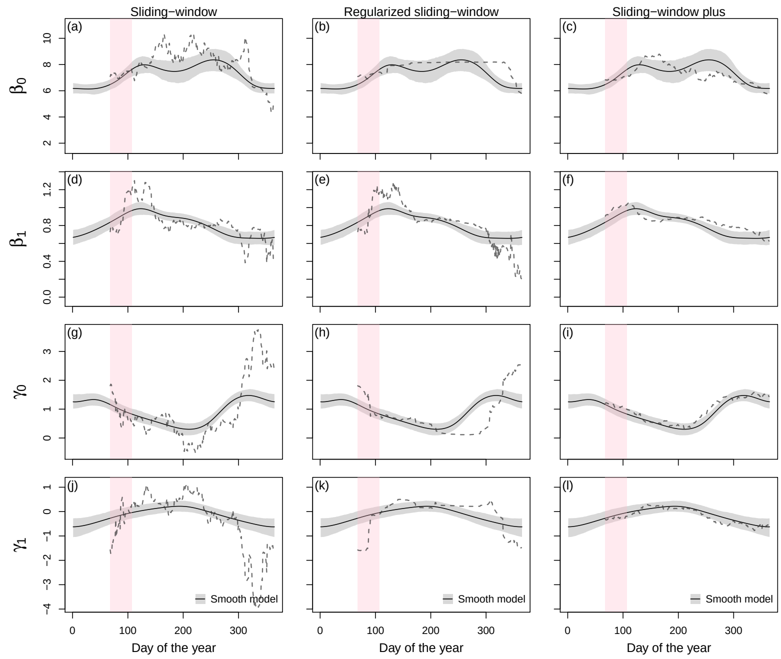


**Fig. 1.** As Fig. 3 and 4 in the paper, but for station Altdorf for the first calendar year during the validation period in data set B. The first 40 days within the validation period, which correspond to the period directly after the change in the horizontal resolution of the ECMWF EPS on March 8, 2016, is highlighted in pink.



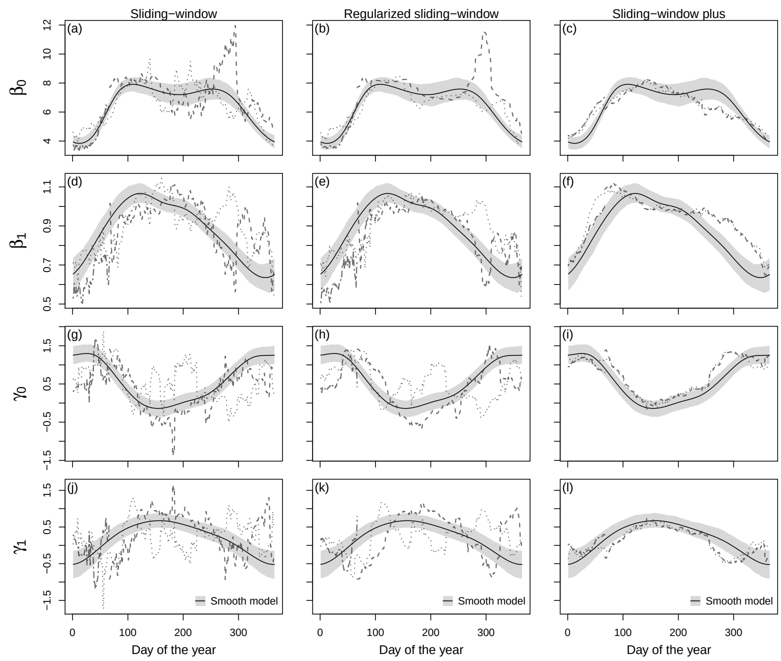


**Fig. 2.** As Fig. 3 in the paper, the temporal evolution of regression coefficients is shown for the validation period in data set A for Innsbruck at forecast step +36h (valid at 1200 UTC). Contrary to Fig. 3 in the paper, the modified regularized sliding-window approach uses three iterations in the estimation process before the optimizer is stopped. The coefficient paths are plotted for the consecutive calendar years 2014, 2015, and 2016 as dashed, dotted, and two-dashed line, respectively.

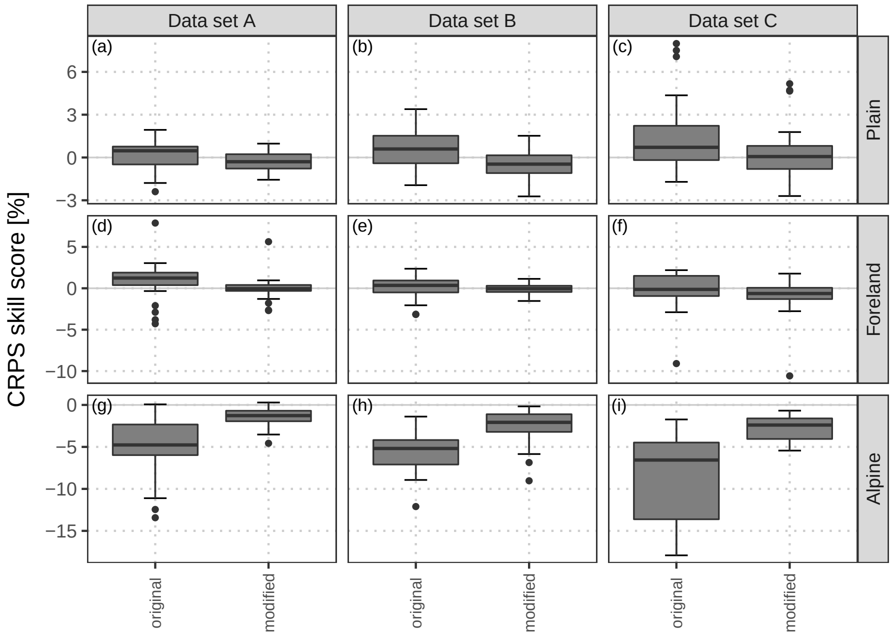


**Fig. 3.** Similar to Fig. 5 in the paper, but only for the regularized sliding-window approaches with the classical sliding-window approach as a reference. The original version as presented in the manuscript uses a single iteration, whereas the modified version uses three iterations in the estimation process before the optimizer is stopped.