Final author response

Submission “Latent Linear Adjustment Autoencoders v1.0: A novel method for estimating and emulating dynamic precipitation at high resolution”

We thank both referees for their insightful comments, helpful feedback and their positive evaluation. We address all points in detail below where we show the referees’ comments in italics for ease of exposition.

Review 2

Major comments:

1. The training (1955-2070) and testing (2071-2100) periods are consecutive, which I do not think is the best choice, as the training data is likely to contain a forced precipitation trend. Separating the training and testing datasets (e.g. training 1955-2020) would provide a more rigorous test of whether the dynamical adjustment method can separate internal variability from a forced signal, without much of the forced signal being present in the training dataset. The authors should test at least some of their results for sensitivity to the choice of training period.

Thank you, we provide two additional analyses: (1) to identify the **sensitivity to the training period**, and (2) to identify the **sensitivity to the trend removal procedure**. In short, reducing the training period to 1955-2020 does not have a noticeable effect on the results. The SLP trend removal procedure used in the paper is shown to yield residuals that match the ensemble mean precipitation trend very well. However, there is a known sensitivity of dynamical adjustment to different detrending choices, and we show this for the case when SLP would not be detrended but precipitation would be detrended (in a simplistic manner). We discuss the implication of this limitation of dynamical adjustment (generally) in the revised manuscript. Please see our response to Major Comment 1 from the first referee for all details on both additional sensitivity tests.

2. I am not convinced that the forced signal that is extracted using the dynamical adjustment method is a purely thermodynamic signal of precipitation change, for two reasons. Firstly, the residual trend will include not only Clausius-Clapeyron-related increases in moisture, but also any other change in the relationship between SLP and precipitation under climate change. This could include, for example, changes in land-atmosphere interactions or weather system dynamics.

Secondly, there may be changes in the pattern of the individual SLP EOFs under climate change. Even small changes could have large consequences for regional precipitation. The authors have tried to address this point by detrending the SLP time-series, based on trends in EOF1, but I was slightly confused by the description of this detrending, and am not convinced that it would account for any (possibly subtle) changes in the shape of EOFs.

The dynamical adjustment method will remove any signal caused by temporal variation in the frequency of the SLP EOFs that were identified during the training period. The removed component will likely be due mainly to internal variability, though it could also include some forced signal if forcing were to drive any systematic change in the relative frequency of SLP EOFs. The residual will likely reflect a forced signal but calling it a thermodynamic precipitation change is too much of an oversimplification to be useful. Other factors could also be important.

The reviewer raises very important points, and we agree. Referring to the residual as a pure “thermodynamic signal” is clearly an oversimplification. What we meant to say, and this is is agreement with the use of terminology in many dynamical adjustment papers (e.g. Deser et al 2016, Lehner et al 2017, Lehner et al. 2018), is that we expect the residual time series to *contain* the imprint of thermodynamical signals, in particular thermodynamical changes (for example, as pointed out by the reviewer, increases in temperature that induce higher water-holding capacity of the atmosphere via the C-C relation). We will rephrase the revised manuscript such that, (1) it becomes very clear that we are not claiming that the residual is a purely thermodynamical signal, but that it may contain effects of feedbacks, remaining internal variability, circulation components not directly captured by SLP, etc. In addition, long-term dynamical changes may even be part of the residuals; (2) we will also clarify that the choice of detrending (i.e., which variables to detrend, etc.), remains a key uncertainty in dynamical adjustment. Hence, clearly, more work is needed to fully understand the different implications of trend removal, but we believe this work is beyond the present study as the goal of this study was to illustrate Latent Linear Adjustment Autoencoders as a versatile tool to simulate daily precipitation variability based on a coarse SLP field.

3. Dynamical adjustment appears to have the potential to significantly reduce the size of ensembles needed to reliably extract forced trends. However, a certain number of model years are needed to train the algorithm, so it is not clear exactly what the computational cost saving would be overall. Could the authors provide an estimate of the overall fractional saving in computational cost, taking algorithm training into account?

Please see our reply to Point 2 made by Reviewer 1. In short, we agree, making an exact calculation of “number of ensemble members saved” is hard to make because of the training. But this is also not our main point: Our main point was to provide a proof-of-concept for applying dynamical adjustment to high-resolution regional precipitation fields (which is novel), and we anticipate eventually our model to be trained on large ensembles, but to be applied to, e.g. models where only runs are available, or even reanalysis.

4. Is there an alternative type of machine learning algorithm that could be used to link SLP EOFs as input directly to the 2D precipitation fields as output (e.g. some form of neural network)? What are the benefits of using the intermediate stage of the autoencoder? I am not suggesting any extra analysis here, only for the authors to justify their choice of method a bit more.

Linking SLP EOFs as input with the 2D precipitation fields as output without having the intermediate stage of the autoencoder seems very challenging from an estimation point of view. Here, the autoencoder helps to estimate the decoder. We are not aware of alternative ML algorithms for this input/output combination and our methodology is novel in this regard.

More generally, one could extend the method of Sippel et al. (2019) by using a neural network instead of regularized linear regression. In that case, however, one would have a separate fit for each grid point (i.e. not the 2D precipitation field as output). This would be computationally demanding and it is also questionable whether the resulting predicted spatial field would be as coherent. We will discuss this point in the Summary/Conclusion of the revised manuscript.

Minor points

1. It is not clear from the objectives in section 1 that the dynamical adjustment will be used to separate forced precipitation trends from internal variability. It would be useful to the reader for this objective to be spelled out here.

Thank you. We have clarified this in the revised manuscript.

2. Fig. 3: How were these examples chosen? Are they representative of the data as a whole? It might be more useful to show high, medium and low skill cases rather than a random selection.

Thanks for the suggestion. The examples were chosen randomly but we have now updated the figure to show examples for different quantiles of the loss (high, medium and low skill cases).

3. Figure colour scales. It is quite difficult to get much information out of the current single shading colour scales. I appreciate this is not a simple problem, but perhaps these could be improved to show the spatial features more clearly.

Thank you for the suggestion. We have experimented with different color maps and different numbers of color levels but have not yet found alternative settings that yielded better figures.

4. Fig. 4 & 5: Why only use a single holdout ensemble member for this? Why not use all of them? Also, relative error might be more informative for Fig. 4, rather than absolute error which mainly picks out the regions of high precipitation.

Fig. 4 and 5 look fairly similar for the other holdout members.
Regarding the relative error, we show the R^2 values in Fig. 3: Note that the R^2 values are computed as 1 - relative error where relative error = mean(residual sum of squares)/mean(total sum of squares).

Please also note that we have updated Fig. 8 to include all holdout ensemble members.

References

Deser, C., Terray, L. and Phillips, A.S., 2016. Forced and internal components of winter air temperature trends over North America during the past 50 years: Mechanisms and implications. *Journal of Climate*, *29*(6), pp.2237-2258.

Lehner, F., Deser, C. and Terray, L., 2017. Toward a new estimate of “time of emergence” of anthropogenic warming: Insights from dynamical adjustment and a large initial-condition model ensemble. *Journal of Climate*, *30*(19), pp.7739-7756.

Lehner, F., Deser, C., Simpson, I.R. and Terray, L., 2018. Attributing the US Southwest's recent shift into drier conditions. *Geophysical Research Letters*, *45*(12), pp.6251-6261.

Sippel, S., Meinshausen, N., Merrifield, A., Lehner, F., Pendergrass, A. G., Fischer, E. M., and Knutti, R. (2019) Uncovering the forced climate response from a single ensemble member using statistical learning, Journal of Climate, 32, 5677-5699. doi:10.1175/JCLI-D-18-0882.1