Comments of Referee #1
Response to Referee #1

This paper proposes a ridge regression approach to the detection and attribution of externally forced changes in mean and extreme precipitation. This is an interesting idea that certainly merits exploration, but before devoting a lot of time to understanding the details of the paper and the results that are obtained, I think it is necessary for the authors to better explain their method and to situate it within the pantheon of methods that are already available for detection and attribution.

We would like to thank the reviewer for this important remark. We fully agree that additions to the already densely populated D&A field require justification. In the following more specific comments, we hope to address this, and in a revised paper we will make sure that the method is properly explained along with its positioning and advantages/disadvantages within the broader field of D&A methods. This method adopts several aspects of traditional detection methods (e.g. Santer et al., 2013, 2018; Marvel et al., 2013, 2020; Bonfils et al. 2021; connection explained below in detail), but is indeed not directly following from or equivalent to "optimal fingerprinting" - the method referred to by the referee below. However, we also note that our paper does not intend to propose a new method; different applications of the method we use have been described and used in several papers in recent years, e.g. Sippel et al. (2020).

Ridge regression is a technique that "regularizes" regression problems, such as that described in equation (1) of the paper, in which the predictor variables contained in matrix X are multicollinear. In the generalized least squares formulation of the regression used in detection and attribution this matrix is composed of model simulated estimates of the responses to external forcing in the form of space-time patterns of change. Depending on variable, period considered, domain of interest and how data are processed, the expected space-time patterns of responses to different forcing factors (often called fingerprints) can be strongly correlated, which results in a regression "design matrix" X that may be ill conditioned. Ridge regression is a technique that can be used to overcome this problem, although I imagine at the cost of introducing some bias into the estimated signal scaling coefficients β. Note that referring to these coefficients as "fingerprints" seems unusual to me.

The concept of regularization, however, also arises in a second way in the detection and attribution problem. Considering again equation (1), the generalized least squares approach (and also its total least squares extension) requires knowledge of the variance covariance matrix of the residuals ε, which are regarded as resulting from natural internal climate variability. Thus, the variance-covariance matrix is generally estimated from unforced control simulations, using as many climate-model simulated realisations of ε as possible. Even though many climate-model simulated realizations of ε are now generally available, the estimated variance-covariance matrix may not be of full rank or may remain uncertain. Thus, it is also often regularized, using an approach similar to the regularization used in ridge regression, but applied to the noise term rather than the signal term of equation (1). See Ribes et al (2013a, doi:10.1007/s00382-013-1735-7, and 2013b, doi:10.1007/s00382-013-1736-6). Presumably one would want to regularize both aspects of the problem, and also take signal uncertainty into account as is done in the total least squares approach to the regression problem (see again Ribes et al., 2013a and 2013b, and also Allen and Stott, 2003, doi:10.1007/s00382-003-0313-9).

How the combined model represented by equations (1-3) relates to existing techniques, and now the noise that results from internal variability comes into play and is accounted for in their subsequent application in the paper is not made clear, and I think should be clarified before results can be considered.

Thank you for your comment, and we agree that it is crucial to clarify the relationship to other methods in the D&A spectrum. Thanks also for the very nice description of the motivation behind the optimal fingerprinting method that addresses the multicollinearity problem of spatiotemporal data used in D&A, and the estimation of the variance-covariance matrix of internal variability.

Please note, however, that there is one important difference between the method we employ in our paper, and optimal fingerprinting (e.g., Allen & Stott, 2003; Ribes et al., 2013a, 2013b): In optimal fingerprinting, the observations (in space/time) are regressed onto the model's (space/time) simulated response patterns, using an estimate of the variance-covariance matrix of internal variability; and inference is made via the signal scaling factors of the regression. Unlike optimal fingerprinting, however, our method does not regress observations onto the model's simulated space-time response patterns. Instead, the goal of our regression step (Eq. 1) is to derive a projection of the model's output (i.e. space/time data of PRCPTOT and Rx1d) onto a one-dimensional detection space. We achieve this via (ridge) regression of each model's (one-dimensional) global forced response estimate Y (obtained by standard techniques) of length *n* (given by the number of CMIP6 model simulations and years considered) onto the models' simulated Rx1d or PRCPTOT patterns X (including internal variability) at *p* grid cells (where *p* is given by observational coverage). This means we obtain a set of regression coefficients β of length *p*, which represents a spatial pattern that best maps each individual modeled Rx1d/PRCPTOT pattern onto its forced response estimate $\hat{Y}$. We call this β the fingerprint. In a second step we use the fingerprint β to map observations (and piControl simulations) into the one-dimensional space in which detection is assessed. In our paper, we compare the trends in simulated and observed forced response estimates, but we do not formally determine any analogues to scaling factors in optimal fingerprinting. As correctly pointed out by the reviewer, our β coefficients (fingerprint) are thus fundamentally different to the signal scaling factor central to optimal fingerprinting, which is sometimes also called β.

We would also like to clarify the role of regularisation by juxtaposing the use of regularisation in our method with the use of regularisation in optimal fingerprinting studies (e.g. Ribes et al., 2013a, 2013b). In our application, regularisation aims to reduce the effects of internal variability on the forced response prediction, i.e., regularisation is the way we account for the noise resulting from internal variability - one of the concerns raised by the referee. To make this more concrete: if we would use ordinary least squares, the regression step (Eq. 1) would result in a highly overfit map of coefficients β that e.g. employs high weights on low-amplitude anticorrelated grid cells to explain a minor part of the variance, due to overdetermination of the problem. Therefore, we apply ridge regularisation, which penalises the squared magnitude of coefficients (L2-norm) in a way that leads to lower, spatially more uniform, but non-zero coefficients. In this way, variance in the forced response estimate due to internal variability is reduced, i.e. the robustness to internal variability is increased and the β fingerprint is more generalisable across models and between models and observations. As outlined by the referee, optimal fingerprinting employs a similar regularisation technique to make the variance-covariance matrix of internal variability more generalisable. Hence, our (ridge) regularisation step has a similar general purpose as in optimal fingerprinting (generalisability), however, is applied to a different aspect of the data (the signal in our case, the noise in optimal fingerprinting). Signal and noise are not explicitly separated in our method to estimate the β fingerprint, but noise due to internal variability is accounted for implicitly by the regularisation step. Our method is also explained in detail in Sippel et al. (2020).

In addition to the similarities and differences w.r.t. optimal fingerprinting described above, we provide more context below, to situate our method in the landscape of well-known as well as less well-known, recently developed D&A methods.

The purpose of D&A - isolating the forced climate change signal in observations and comparing this to the forced response in climate models - requires separation of the signal and the noise. Several ways to do this have been developed, which, one could argue, all derive from Hasselman (1979). From here, optimal fingerprinting (e.g. Allen & Stott, 2003; Ribes et al., 2013a, 2013b) and detection methods based on pattern similarity (e.g. Santer et al., 2013, 2018; Marvel & Bonfils, 2013, 2020; Bonfils et al., 2021), sometimes called “non-optimal detection”, have evolved. Our ridge regression based detection method is most closely related to pattern similarity methods as used in e.g. Santer et al. (2013). In Santer et al.’s work, the detection metric is defined as the projection of the variable of interest onto the leading empirical orthogonal function (EOF), derived from aggregated forced model simulations (i.e., an estimate of the signal pattern). In this method, the EOF-based signal pattern is referred to as the fingerprint, and the projection of observations and piControl simulations onto this pattern yields a one-dimensional detection metric, where detection is assessed as the deviation of the observations (e.g. trends of L years) from the distribution of unforced control simulations.

The ridge regression method described above and in our paper, builds on these pattern similarity detection methods in a straightforward way by adding a step in between signal pattern determination and projection. In our method, we project observations not onto the signal pattern directly, but onto a regression coefficient pattern that “optimally” (linearly, optimised by regularisation) maps Rx1d or PRCPTOT patterns onto a forced response metric based on the signal pattern. Hence, the ridge regression method can be seen as one step towards “optimising” the signal pattern (fingerprint) by increasing the signal-to-noise ratio (SNR).

We believe that the advantages of our method lie in (1) its relative simplicity and close links to the pattern similarity based D&A methods, while going beyond comparisons to the signal pattern (e.g. Marvel & Bonfils., 2013) or spatial aggregation techniques for Rx1d or PRCPTOT (Fischer et al., 2014; Donat et al., 2016), (2) the interpretable and relatively intuitive fingerprint map β that reflects regions exhibiting high SNR climate change signals, (3) the actual estimate of the observed forced response time series resulting from application of this map to observations, so trends in this metric can be analysed, and (4) the possibility to straightforwardly introduce additional constraints to, for instance, guard against specific climate uncertainties, such as uncertainty of whether climate models represent the correct magnitude of decadal-scale internal variability (Sippel et al., 2021). This method fits in recent developments in D&A that move towards mapping multidimensional data onto a one-dimensional detection space. Studies based on neural networks and deep learning for detection and attribution, e.g. Barnes et al. (2019, 2020), Labe & Barnes (2021), Madakumbura et al. (2021), Ham et al. (2022, preprint), employ non-linear methods - as opposed to our linear ridge regression method - but use a very similar framework with similar goals. We do not argue that ridge regression is fundamentally better than any of these older or newer methods, but we are convinced that the intuitive, physical outputs combined with high SNR can be valuable for trend detection and attribution.

We hope that these considerations address the reviewer's concerns. Based on this, we suggest to address issues raised in a revised paper by extending the Methods section with one paragraph to address in more detail the workings of our method, and relate it to existing D&A methods such as optimal fingerprinting and pattern similarity methods as explained above (in shortened and streamlined form). In addition, we will also make sure to eliminate redundant use of terminology commonly used in an optimal fingerprinting context, to avoid confusion about terms that have different meanings in our context. Nonetheless, we would prefer to continue to use the term "fingerprint" since this has been used in multiple contexts (optimal and pattern similarity fingerprinting) to describe the signature of a forcing in the climate, which is also how we use it (see the comparison with pattern similarity methods above). However, in order to prevent confusion we will make sure to add the side note that our method to create fingerprints differs from the optimal or pattern similarity methods.

Also, I think it is necessary for the authors to discuss whether the proposed methods, which basically use linear statistical models that therefore implicitly assume Gaussian, or near Gaussian errors, are suitable for the data to which they are applied. Indicators of extreme precipitation, such as Rx1day at individual grid boxes, are certainly not Gaussian.

Thanks. We fully agree that local distributions of Rx1d are not Gaussian and that methods assuming Gaussianness can not be applied to such distributions. However, we only use individual grid box time series of Rx1d as predictors in our linear regression model. Predictors in linear regression models need not be Gaussian as long as no parametric confidence intervals of regression coefficients are derived (which we don't do).

Below, we added Q-Q plots of the forced response estimates obtained by applying our detection fingerprint to piControl data (450 years per model). As can be seen, these unforced estimates are normally distributed, also for the Rx1d case. This assures the remnant natural variability in the forced response estimates is normally distributed.

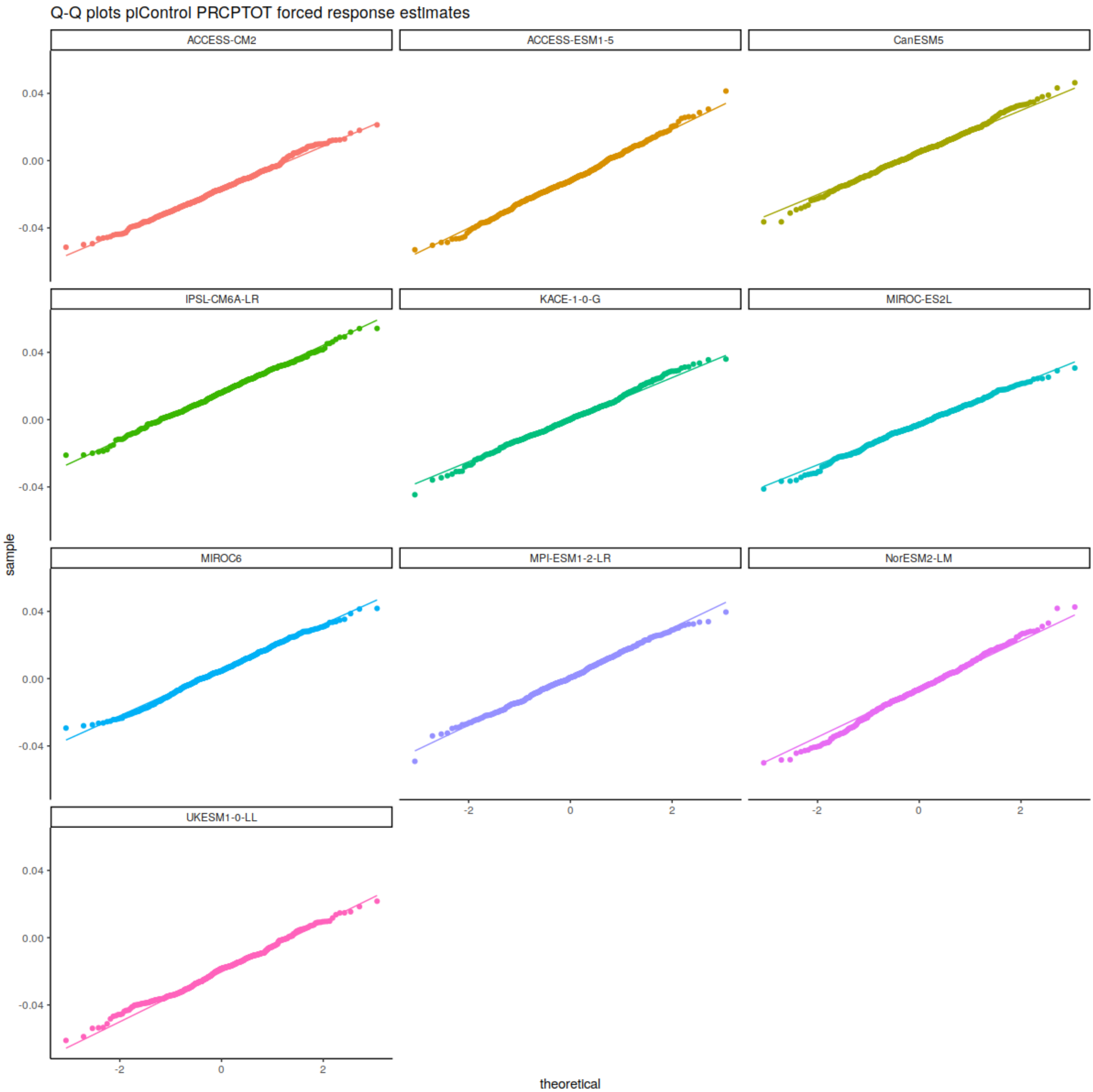

Figure 1: Q-Q plots of PRCPTOT piControl forced response estimates obtained by applying the fingerprint to piControl simulations of 10 CMIP6 models, 450 years per model. The Q-Q plots are separated by model since different models may have different spread. When points lie on the diagonal slope-1 linear line, the distribution is normal.

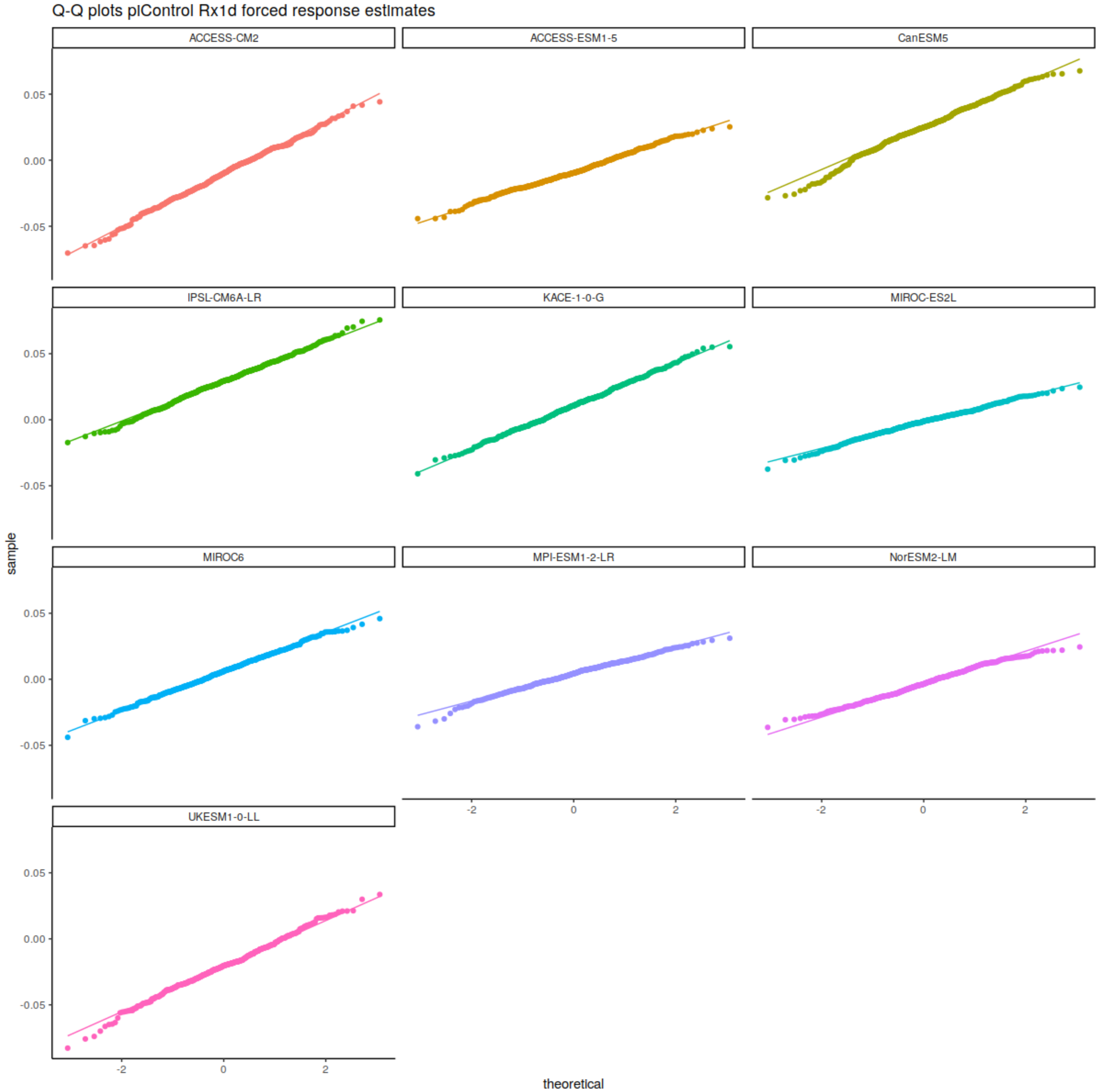

Figure 2: As figure 1 but for Rx1d.

For the estimation of time of emergence, we implicitly assume the residuals of the linear fit of forced response estimates to GMST are normally distributed. In this application the local temporal distribution of Rx1d values is no longer of influence. We do not show the ToE residuals to be normally distributed, but we can add this to the supplementary information if so required.

A final general comment is that the relatively heavy of use of acronyms in this paper is not very reader friendly.

Point taken. We'll reduce acronym use where possible.

References

Allen, M.R., Stott, P.A. (2003). Estimating signal amplitudes in optimal fingerprinting, part I: theory. *Climate Dynamics* 21, 477–491 . https://doi.org/10.1007/s00382-003-0313-9

Ribes A., Planton, S., Terray, L. (2013). Application of regularised optimal fingerprinting to attribution. Part I : method, properties, and idealised analysis, *Climate Dynamics*, 41(11-12), 2817-2836, 10.1007/s00382-013-1735-7.

Ribes A., Terray, L. (2013). Application of regularised optimal fingerprinting to attribution. Part II : application to global near-surface temperature based on CMIP5 simulations, *Climate Dynamics*, 41(11-12), 2837-2853, 10.1007/s00382-013-1736-6.

Hasselmann, K. (1979). On the signal-to-noise problem in atmospheric response studies. In D. B. Shaw (Ed.), *Meteorology over the tropical oceans* (pp. 251-259). Bracknell: Royal Meteorological Society.

Santer, B. D., Painter, J. F., Mears, C. A., Doutriaux, C., Caldwell, P., Arblaster, J. M., ... & Zou, C. Z. (2013). Identifying human influences on atmospheric temperature. *Proceedings of the National Academy of Sciences*, *110*(1), 26-33.

Santer, B. D., Po-Chedley, S., Zelinka, M. D., Cvijanovic, I., Bonfils, C., Durack, P. J., ... & Zou, C. Z. (2018). Human influence on the seasonal cycle of tropospheric temperature. *Science*, *361*(6399), eaas8806.

Marvel, K., & Bonfils, C. (2013). Identifying external influences on global precipitation. *Proceedings of the National Academy of Sciences*, *110*(48), 19301-19306.

Marvel, K., Biasutti, M., & Bonfils, C. (2020). Fingerprints of external forcings on Sahel rainfall: aerosols, greenhouse gases, and model-observation discrepancies. *Environmental Research Letters*, *15*(8), 084023.

Bonfils, C. J., Santer, B. D., Fyfe, J. C., Marvel, K., Phillips, T. J., & Zimmerman, S. R. (2020). Human influence on joint changes in temperature, rainfall and continental aridity. *Nature Climate Change*, *10*(8), 726-731.

Sippel, S., Meinshausen, N., Fischer, E. M., Székely, E., & Knutti, R. (2020). Climate change now detectable from any single day of weather at global scale. *Nature climate change*, *10*(1), 35-41.

Fischer, E. M., Beyerle, U., & Knutti, R. (2013). Robust spatially aggregated projections of climate extremes. *Nature Climate Change*, *3*(12), 1033-1038.

Donat, M. G., Lowry, A. L., Alexander, L. V., O'Gorman, P. A., & Maher, N. (2016). More extreme precipitation in the world's dry and wet regions. *Nature Climate Change*, *6*(5), 508-513.

Sippel, S., Meinshausen, N., Székely, E., Fischer, E., Pendergrass, A. G., Lehner, F., & Knutti, R. (2021). Robust detection of forced warming in the presence of potentially large climate variability. *Science advances*, *7*(43), eabh4429.

Barnes, E. A., Hurrell, J. W., Ebert-Uphoff, I., Anderson, C., & Anderson, D. (2019). Viewing forced climate patterns through an AI lens. *Geophysical Research Letters*, 46(22), 13389– 13398. https://doi.org/10.1029/2019GL084944

Barnes, E. A., Toms, B., Hurrell, J. W., Ebert-Uphoff, I., Anderson, C., & Anderson, D. (2020). Indicator patterns of forced change learned by an artificial neural network. *Journal of Advances in Modeling Earth Systems*, *12*(9), e2020MS002195.

Labe, Z. M., & Barnes, E. A. (2021). Detecting climate signals using explainable AI with single-forcing large ensembles. *Journal of Advances in Modeling Earth Systems*, 13, e2021MS002464. https://doi.org/10.1029/2021MS002464

Madakumbura, G.D., Thackeray, C.W., Norris, J., Goldenson, N., Hall, A. (2021). Anthropogenic influence on extreme precipitation over global land areas seen in multiple observational datasets. *Nat Commun* 12, 3944. https://doi.org/10.1038/s41467-021-24262-x

Ham, Y.-G. Kim, J.-H., Min, S.-K. Kim, D., Li, T., Timmermann, A., Stuecker, M. Anthropogenic fingerprints in daily precipitation revealed by deep learning, 03 August 2022, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-1860132/v1]