Dear Authors,

Following your thorough replies to the comments provided by both Reviewers, I would like to invite you to prepare a revised version of your study. Amongst other edits, I would encourage you to provide a thorough contextualisation of your work along the lines of your replies to Reviewer #1, and a more intuitive illustration of your methodology along the lines of your replies to Reviewer #2. As ESD is a journal with a broad readership from different fields of Earth Science, it is important that your study does not only speak to specialists in a specific sub-field. Make a careful evaluation of which edits should be included in the main text and which in the Appendix or SI, as the main paper should ideally be accessible to the ESD readership without needing extensive reference to appendices and supplements.

Best Regards,
Gabriele Messori

Dear editor,

Thanks a lot for your managing of our manuscript. We have implemented changes to address the reviewers' comments. Below, we provide a response to both reviewers with short comments on the way in which we addressed the comments. For the longer replies, please refer to the initial responses.
We are aware that there still are a significant number of references to the supplementary information, and the amount of material in supplementary information is considerable. Nonetheless, we are confident that this all is indeed strictly supplementary and that the manuscript in itself is stand-alone, since none of the main conclusions depend on any of the additionally provided analyses or considerations.

Best,
Iris de Vries, on behalf of all authors

Comments of Referee #1

Short responses, please refer to the initial response to reviewer 1 for extensive replies.

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.

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.

We would like to thank the reviewer for the important remarks and suggestions. Below we outline in brief comments how we addressed the comments. For the extensive reply to the comments we refer to the initial response to reviewer 1.

A condensed version of the reply in the initial response to reviewer 1 has been added to the methods section – ranging from **L86 to L115** – to explain our method in more detail, and explicitly relate our method to existing and upcoming methods, and hopefully make the differences more clear. In addition, a flow chart has been added – see also our response to referee #2's comments – which should further clarify the ridge regression part of the method.

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.

There is no need for the predictors of Rx1d trends to be normally distributed in our method, as explained in the first response to reviewer #1, and refer the reviewer to the online reply for a more extensive justification.

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

We have removed the acronyms for forced response (FR) and forced response estimate (FRE), which leaves Rx1d, PRCPTOT, ridge regression (RR), signal-to-noise ratio (SNR) and empirical orthogonal function (EOF). Given that these are commonly used acronyms and/or concern the essence of our study (RR), we think this is a manageable collection of acronyms.

Comments of Referee #2

Short responses, please refer to the initial response to reviewer 2 for extensive replies.

Overall comments:

This study conducts a signal detection analysis for global changes in mean and extreme precipitation using three observational datasets and CMIP6 multi-model outputs. The authors apply a ridge regression (RR) method to construct fingerprints, which helps increase a signal-to-noise ratio of precipitation change patterns. Results show a robust detection of anthropogenic signals in all observations for both mean and extreme precipitation even when removing global mean trends, further supporting the human-induced intensification of global hydrological cycle. I find this paper very well written with sufficient details provided about methods as well as various sensitivity tests and therefore suggest publication after addressing some minor issues.

Thanks again for your comments and efforts. We list the changes made below, and refer to the initial response to reviewer 2 for extensive replies.

Major comments:

1. Although method details are provided, it would be useful to explain more clearly what are benefits of the attribution approaches employed, including ridge regression, EOF-based metric for target variable, and GMST-based signal estimation. All of these procedures seem to contribute to increase signal-to-noise ratio but how they do and what step is more important. The authors provide some associated results from sensitivity tests but an overall explanation of their method possibly with a schematic would be helpful for readers to understand the contribution of each step to the final signal detection.

We added additional explanation of the method (**L86-L115**) to the methods section, as well as a flowchart **(Figure 1)** explaining the ridge regression steps. A flowchart addressing the EOF-based targets is added to the supplementary information (SI Figure S1).

In addition, we added a section to the supplementary information, **SI Section S2.4**, to provide additional information on the effects of design choices on the signal-to-noise ratio/time of emergence. The crucial points of this SI section, namely the effect of regularisation, were already discussed in the original manuscript, section 3.4.

2. An important motivation of considering different periods and datasets is opposing conclusions by previous studies about model overestimation or underestimation of the observed trends. I am wondering if the authors can go further and compare their results with some previous studies. For instance, if studies based on the latter half of 20th century trends find model underestimation, the authors can assess their model trends for the same/similar periods. Another point here is that the present study uses absolute units of precipitation while most of previous studies considered relative changes or aggregated values. It would be good to discuss possible influences of this difference.

We addressed the comparison of different precipitation metrics in supplementary **section S2.3**, and refer to here in the manuscript **L377** and **L483**.

We addressed the comparison to previous studies by adding the table below to the supplementary info. This systematic comparison to other studies has also led to the addition of a few references throughout the manuscript.

Previous studies report results ranging from model under- to model overestimation of observed trends for all trend periods and units (absolute vs. normalised). There is no systematic explanation for why opposing results are found. Therefore, we do not go further than stating this, in **L482-492**.

Paper	Model archive	Obs dataset	Spatial region	Variable	Method	Trend periods	Models w.r.t. observations?	Remarks
PRCPTOT								
Noake et al. (2012)	CMIP3	GHCN, CRU, VASClimO	Global land, separated into 5deg latitude bands. Scaling factors determined for spatiotemporal aggregate, not per latitude band.	Seasonal PRCPTOT percentage change per latitude band	Optimal fingerprinting	1952-1990, 1960–1999, 1951–1990, 1975–1999	-	"-" applies to scaling factor best estimate for seasons and observational datasets in which significant change is detected (confidence interval does not include 0), and holds for all trend periods
Wu et al. (2013)	CMIP5	GHCN	Northern hemisphere land	PRCPTOT percentage change	Optimal fingerprinting	1952-2011	-	
Polson et al. (2013)	CMIP5	GHCN, CRU, VASClimO, GPCC	Global land, separated into 5deg latitude bands. Scaling factors determined for spatiotemporal aggregate, not per latitude band.	Seasonal PRCPTOT percentage change per latitude band	Optimal fingerprinting	1951–2005 (2000 for VASClimO)	-	Applied Noake's method to CMIP5. "-" applies to scaling factor best estimate for seasons and observational datasets in which significant change is detected (confidence interval does not include 0). GPCC never shows a detectable climate signal.
Fischer & Knutti (2014)	CMIP5 + CESM initial condition ensemble	HadEX2, GHCNDEX	Global	Spatial distibution of gridpoint trends in PRCPTOT, expressed in terms of local sigma (based on 1986-2005 interannual variability)	Spatial probability distribution comparison	1960-2010	+	Models estimate more regions with positive trends in PRCPTOT, but not enough negative trends --> too much wetting
Knutson & Zeng (2018)	CMIP5	GPCC	Global, per gridpoint	Linear trend in PRCPTOT	Linear trend fitting to grid point timeseries	1901-2010, 1951-2010, 1981-2010	-	Models cannot produce the magnitude of positive nor negative trends in obs. Discrepancy gets stronger in later trend periods
Rx1d								
Min et al. (2011)	CMIP3	HadEX	NH land, separated into (overlapping) regions: mid-latitudes, tropics	Rx1d and Rx5d Probability index: 0-1 quantile per value, based on fit GEV per gridpoint	Optimal fingerprinting	1951-1999	-	"-" applies to scaling factor best estimate for regions where there is detection
Zhang et al. (2013)	CMIP5	HadEX2 + russian station data	NH land, separated into (overlapping) regions: Western Eurasia, Eastern Eurasia, North America, mid-latitudes, tropics	Rx1d and Rx5d Probability index: 0-1 quantile per value, based on fit GEV per gridpoint/station and then interpolated	Optimal fingerprinting	1951-2005	0/+	Scaling factor estimates include 1, but best estimates are still below 1
Fischer & Knutti (2014)	CMIP5 + CESM initial condition ensemble	HadEX2, GHCNDEX	Global	Spatial distibution of gridpoint trends in Rx5d, expressed in terms of local sigma (based on 1986-2005 interannual variability)	Spatial probability distribution comparison	1960-2010	-	Models don't show a large enough land fraction exhibition positive trends, and do not reproduce the magnitude of the largest trends seen in observations
Fischer & Knutti (2016)	CMIP5 and EURO-CORDEX	E-OBS/Ensembles	Europe	Changing occurrence of historical >90 percentile values of daily precipitation	Probability distribution comparison	1951-1980 and 1981-2013 distributions	-	Models show smaller increase in intensity of >90th percentile daily precipitation amounts
Borodina et al. (2017)	CMIP5 + CESM initial condition ensemble	GHCNDEX, HadEX2	Global land, selected wet regions only (wettest 40%, agreed across models)	Rx1d percentage change per gridpoint as a function of GMST [%/K], averages over wet regions, as well as land area fraction experiencing positive Rx1d trends	Trend comparison	1951-2005	-	Models show smaller trends than both observational datasets, but HadEX2 shows smaller trends than GHCNDEX
Paik et al. (2020)	CMIP5	HadEX2	Global land, separated into (overlapping) regions: Western Eurasia, Eastern Eurasia, North America, mid-latitudes, tropics, wet and dry regions.	Rx1d and Rx5d Probability index: 0-1 quantile per value, based on fit GEV per gridcell/station and then interpolated. Scaling factors	Optimal fingerprinting	1950-2020	0/+	"0" applies to EU and dry regions, where models and observations agree. For all other regions with detection, models overestimate the change ("+")
Paik et al. (2020)	CMIP5	HadEX2	Global land, separated into (overlapping) regions: Western Eurasia, Eastern Eurasia, North America, mid-latitudes, tropics, wet and dry regions.	Rx1d and Rx5d Probability index: 0-1 quantile per value, based on fit GEV per gridcell/station and then interpolated. Spatially averaged trends, normalised by GMST	Trends in %/K	1950-2020	-	Note: same study as above. In all regions where forced change is detected, models underestimate observations when trends in %/K are assessed. In these same regions, scaling factors suggest that models overestimate change.
Sun et al. (2022)	CMIP6 and CanESM2 LE	HadEX2 stations + russian and chinese station data	Global, continental, regional	Rx1d and Rx5d, Non stationary spatiotemporal varying GEV-based optimal fingerprinting, no normalisation: absolute units of precipitation (log)	Non-optimal variant of optimal fingerprinting; scaling factor determination but no internal variability covariance corrections	1950-2014	+	"+" applies to all continents/regions, and also global level, but Northwestern Europe (Scandinavia/UK) where scaling factors are around 1 ("0")

Table 1: Previous D&A studies on PRCPTOT and Rx1d, including their main findings on whether modelled forced changes are smaller (-), similar (0) or larger (+) than observed forced changes

3. The lower detectability in GHCNDEX observations are suggested to be due to the poorer spatial coverage. Regarding this issue, I would suggest using Rx5d. As I understand, Rx5d has larger spatial coverage than Rx1d and comparison with Rx1d-based results may provide a way to support the authors' interpretation. Another way would be to compare detection results from using a selected model run but with different spatial coverages applied.

We looked into this, but found that Rx5d does not provide higher coverage. Because we want to keep the difference between mean and extreme precipitation measures as large as possible, (see initial response to reviewer 2) we maintain Rx1d as the metric for extreme precipitation.

Minor comments:

L8: Indicating analysis period or trend period with signal detection would be useful here.

Changed to "[...] to assess the degree of forced change detectable in the real-world climate *in the period 1951-2020*." in **L8.**

L17-19, L58-64: Better comparisons can be made by applying the same periods as those used in previous studies. See my major comment above.

See reply to major comment above: both previous studies as well as we assess multiple trend periods. Disagreements across studies and observational datasets remain.

L20-21: Is this confirmed by repeating detection analysis using NH-extratropics only?

Yes, see supplementary information, SI section S3.

L34: "discrepancies with respect to observations". Its meaning is unclear.

Changed the sentence to "There can also be discrepancies between model representations of the water cycle and observations." in **L33-34**.

L69-71: Need to explain what the previous studies have found additionally using these "data-science methods". Also, what's the novelty of this study compared with them? Is it detection based on spatial pattern information alone?

This has now been added to the method section, from **L87** we compare to other D&A methods in general, and from **L111** we specifically refer to other data-science methods.

L108-109: "Trend biases due to this structural difference … negligible". But the cited reference considered south-east Australia only?

In **L138** we have added the reference from the previous sentence, which makes this statement in a general context.

L201: How to define S when global means are removed?

We do not think additional explanations in the paper are required here, since the forced trend is still the predicted variable in this case, which is the basis for S.

L212: "CMIP6 ssp245" should be "CMIP6 historical"?

This has been changed, **L249.**

L227: "virtually identical". adding spatial correlation would help with this.

This has been added to **L264-265.**

L314-316: This suggests possible dependence of Rx1d FRE on temperature, resembling global warming slowdown due to PDO influence?

Potentially, although we do not have enough evidence to claim that the levelling off of the trends is not simply due to shorter trend length and internal variability. Attributing changes in trend slope to lange scale modes of variability is outside the scope of this study. Hence we have not added statements on this to the manuscript.

L331-332: "results … hold when the global mean is used as FR target". Then what are benefits of using EOF-based metric for target variable?

See the discussion in the initial response to reviewer 2. We have added to the method section to further justify this choice of target in **L169-172**.

L382-383: "accuracy of the CMIP6 climate models in simulating the processes …". It's unclear how the authors get this conclusion. Observation-model agreement in residual variability? More explanation would be useful.

From **L428** we changed this to "Taken together, the above shows, first, detection of forced change in mean and extreme precipitation beyond a global mean trend, and second, the power of RR for signal extraction from high-dimensional noisy data. Finally, the fact that the relationship between relative spatial precipitation patterns and the forced precipitation trend derived from climate model simulations (the ridge model) holds in observations, suggests accuracy of the CMIP6 climate models in simulating the processes relevant to the spatial pattern of forced change in mean and extreme precipitation."

L394-395: "(not shown)". This looks important and I suggest showing them in the supplement.

Added in SI section **S2.4**

L428: "value of RR-based fingerprint construction". What happens in detection or SNR without applying RR? See my major comment above.

See the additions to the supplementary information and responses to other comments. In the manuscript, **L209-213**, unchanged text, the effect of regularisation was already described, as well as in section 3.4, from **L457** onwards.

References in table

Noake, K., Polson, D., Hegerl, G., and Zhang, X.: Changes in seasonal land precipitation during the latter twentieth-century, Geophysical Research Letters, 39, https://doi.org/10.1029/2011GL050405, 2012.

Wu, P., Christidis, N., and Stott, P.: Anthropogenic impact on Earth's hydrological cycle, Nature Climate Change, 3, 807–810, https://doi.org/10.1038/nclimate1932, 2013.

Polson, D., Hegerl, G. C., Zhang, X., and Osborn, T. J.: Causes of Robust Seasonal Land Precipitation Changes, Journal of Climate, 26, 6679 – 6697, https://doi.org/10.1175/JCLI-D-12-00474.1, 2013.

Knutson, T. R. and Zeng, F.: Model Assessment of Observed Precipitation Trends over Land Regions: Detectable Human Influences and Possible Low Bias in Model Trends, Journal of Climate, 31, 4617 – 4637, https://doi.org/10.1175/JCLI-D-17-0672.1, 2018

Min, S.-K., Zhang, X., Zwiers, F. W., and Hegerl, G. C.: Human contribution to more-intense precipitation extremes, Nature, 470, 378–381, https://doi.org/10.1038/nature09763, 2011.

Zhang, X., Wan, H., Zwiers, F. W., Hegerl, G. C., and Min, S.-K.: Attributing intensification of precipitation extremes to human influence, Geophysical Research Letters, 40, 5252–5257, https://doi.org/10.1002/grl.51010, 2013

Borodina, A., Fischer, E. M., and Knutti, R.: Models are likely to underestimate increase in heavy rainfall in the extratropical regions with high rainfall intensity, Geophysical Research Letters, 44, 7401–7409, https://doi.org/10.1002/2017GL074530, 2017.

Paik, S., Min, S.-K., Zhang, X., Donat, M. G., King, A. D., and Sun, Q.: Determining the Anthropogenic Greenhouse Gas 575 Contribution to the Observed Intensification of Extreme Precipitation, Geophysical Research Letters, 47, e2019GL086 875, https://doi.org/10.1029/2019GL086875, 2020.

Sun, Q., Zwiers, F., Zhang, X., and Yan, J.: Quantifying the Human Influence on the Intensity of Extreme 1- and 5-Day Precipitation Amounts at Global, Continental, and Regional Scales, Journal of Climate, 35, 195 – 210, https://doi.org/10.1175/JCLI-D-21-0028.1, 2022.