EGUsphere, referee comment RC1
https://doi.org/10.5194/egusphere-2022-568-RC1, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on egusphere-2022-568

Anonymous Referee #1

Referee comment on "Robust global detection of forced changes in mean and extreme precipitation despite observational disagreement on the magnitude of change" by Iris Elisabeth de Vries et al., EGUsphere, https://doi.org/10.5194/egusphere-2022-568-RC1, 2022

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 multi-colinear. 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 introducimg 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.

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.

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