Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2019-63-RC2, 2020

Nonlinear Processes in Geophysics
Discussions

Interactive comment on "Beyond Univariate Calibration: Verifying Spatial Structure in Ensembles of Forecast Fields" *by* Joshuah Jacobson et al.

Anonymous Referee #2

Received and published: 17 June 2020

This manuscript discusses multivariate verification of ensemble forecasts with a focus on spatial structures. A new approach is proposed, illustrated and discussed using synthetic and precipitation datasets.

The proposed approach consists in transforming a multivariate quantity (a spatial field) into a univariate quantity. This is performed by thresholding and counting the fraction of grid points exceeding a threshold over a domain. The popular rank histogram tool is applied to the resulting ensemble and observed fractions of threshold exceedance (FTE). The method is easy to implement and is appealing because of the simplicity of both the interpretation of the derived univariate product and of the derived rank

histograms.

In order to better understand FTE histograms, a toy model is used to assess the sensitivity of the method to predefined discrepancies in a controlled environment. In a real-life example context, FTE histograms are derived for precipitation forecasts and limitations of the underlying forecasting system diagnosed. The paper is well-written, well-structured, and pleasant to read. The description of the method and datasets is clear, the discussion of the results convincing.

However, the manuscript would benefit from some clarifications. First, the authors could clarify the link between the FTE histogram approach and the fraction skill score (FSS, Roberts and Lean 2008), a popular verification metric applied to deterministic forecast of precipitation fields. The first step is similar for the two methods (transformation of a real-value field into a binary field and then a focus on the fraction of events over a domain). It would be beneficial for the verification community to clearly state the link between FTE histogram and FSS.

Secondly (and more importantly), the author could discuss the impact of miscalibration of the univariate distribution on the interpretation of the FTE. It is cleared mentioned in the text that this aspect should not be disregarded. In line 86, it is stated that the FTE histogram can be used "once the marginal distribution has been checked", and in Section 4.2 that "the possible non-uniformity of FTE histograms [...] could be due to univariate miscalibration". Perfect calibration of univariate distributions cannot be expected in reality so a discussion about the permeability of FTE histograms to univariate miscalibration would be more than useful for the interpretation of real case FTE histograms.

In addition, the authors could clarify why the parameters derived from the beta distribution fitting differ from the parameters defined in the Keller and Hense 2011 paper. The list of references could also be checked and the duplicated doi information removed.

Reference: Roberts NM, Lean HW. 2008. Scale-selective verification of rainfall ac-

cumulations from high-resolution forecasts of convective events. Mon. Weather Rev. 136: 78– 97.

Interactive comment on Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2019-63, 2020.