



## Assessing the calibration of multivariate ensemble forecasts: E-values and the choice of pre-rank function

**Sam Allen** and Johanna Ziegel

University of Bern, Institute of Mathematical Statistics and Actuarial Science, Switzerland ([sam.allen@stat.unibe.ch](mailto:sam.allen@stat.unibe.ch))

It is often stated that the goal of probabilistic forecasting is to issue predictive distributions that are as sharp as possible, subject to being calibrated. To assess the calibration of ensemble forecasts, it is customary to employ rank histograms. Rank histograms not only assess whether or not an ensemble prediction system is calibrated, but they also reveal what (if any) systematic biases are present in the forecasts. This information can readily be relayed back to forecasters, helping to improve future predictions. Such is the utility of rank histograms, several extensions have been proposed to evaluate the calibration of probabilistic forecasts for multivariate outcomes. These extensions typically introduce a so-called pre-rank function that condenses the multivariate forecasts and observations into univariate objects, from which a standard rank histogram can be constructed. Several different approaches to construct multivariate rank histograms have been proposed, each of which differs in the choice of pre-rank function. Existing pre-rank functions typically aim to preserve as much information as possible when condensing the multivariate forecasts and observations into univariate objects. Although this is sensible when testing for multivariate calibration, the resulting rank histograms are often difficult to interpret, and are therefore rarely used in practice.

We argue that the principal utility of these histogram-based diagnostic tools is that they provide forecasters with additional information regarding the deficiencies that exist in their forecasts, in turn allowing them to address these shortcomings more readily; interpretation is therefore a key requirement. We demonstrate that there are very few restrictions on the choice of pre-rank function when constructing multivariate rank histograms, meaning forecasters need not restrict themselves to the few proposed already, but can instead choose a pre-rank function on a case-by-case basis, depending on what information they want to extract from their forecasts. We illustrate this by introducing a range of possible pre-rank functions when assessing the calibration of probabilistic spatial field forecasts. The pre-rank functions that we introduce are easy to interpret, easy to implement, and they provide complementary information. Several pre-rank functions can therefore be employed to achieve a more complete understanding of the multivariate forecast performance. Finally, having chosen suitable pre-rank functions, tests for univariate calibration based on rank histograms can readily be applied to the multivariate rank histograms. We illustrate this here using e-values, which provide a theoretically attractive way to sequentially test for the calibration of probabilistic forecasts.