EMS Annual Meeting Abstracts
Vol. 15, EMS2018-160, 2018

Understanding the characteristics of the Fractions Skill Score: The limiting case and implications for aggregation

Marion Mittermaier
Met Office, Weather Science, Exeter, United Kingdom (marion.mittermaier@metoffice.gov.uk)

The Fractions Skill Score (FSS) is arguably one of the most used spatial verification metrics in use today. The fraction of forecast and observed analysis grid points within a neighbourhood, exceeding a threshold, are used to compute the numerator of the score. By definition a perfect forecast has a FSS of 1, and a “no skill” forecast has a score of 0.

In this presentation the score characteristics are illustrated using a selection of contrived and real examples. The real example is the core case of the MesoVICT project. It is shown that the denominator defines the score’s characteristics. The limiting case here refers to cases where, using precipitation as an example, either the forecast and observed fields are completely dry (null) or when precipitation is present in both but there are no exceedances in either the forecast or observed fields for a given threshold. In both cases the score is undefined, which is mathematically problematic and intuitively wrong. Therefore, a perfectly dry forecast can only achieve a perfect score of 1 if a threshold of >=0 is used, and a forecast that has no exceedances above a threshold, the same as the analysis field, is also penalised.

Secondly, the FSS is 0 if either the forecast or the observed field does not exceed a threshold. This symmetry means it cannot differentiate between what are traditionally referred to as false alarms or misses. The score can also not differentiate between a major mismatch, where there is a lot of precipitation in the one field and not in the other, or a minor one, where there is little (but not negligible) precipitation in one and nothing in the other. Whatever the magnitude of the differences in the forecast and analysis fields, these are non-trivial zero scores, i.e. they have meaning.

Finally, the FSS is only greater than 0 if and only if there are values exceeding a given threshold in both the forecast and the observed field. The magnitude of an average score computed over many forecasts is sensitive to the aggregation method. The zero scores matter for aggregation, as excluding them implies excluding all situations associated with what could be considered "forecast busts". Omitting near-zero scores is a more credible decision, but to do this supplementary information is required. With appropriate information it could then be possible to identify and exclude small non-zero scores if and only if it can be proven that these are related to spurious artefacts in the observed field. Furthermore, to avoid ambiguity, the numerator and denominator of the FSS should be aggregated separately before an average score is computed.