The Fractions Skill Score for Ensemble Forecast Verification

Tobias Necker, Ludwig Wolfgruber, Lukas Kugler, Martin Weissmann, Manfred Dorninger, Stefano Serafin

Universität Wien, Institut für Meteorologie und Geophysik,
Josef-Holaubek-Platz 2, 1090 Vienna,
Austria.
Email:Manfred.dorninger@univie.ac.at

Introduction

Since weather phenomena like rainfall are difficult to predict precisely on small scales, forecasts can be unfairly judged (double penalty problem). The Fractions Skill Score (FSS) has emerged as a fair metric for the verification of deterministic rainfall forecasts. The goal of this study is to find the most suitable method for computing the (FSS) for probabilistic ensemble forecasts. It is a summary of the recently published article by Necker, et al. (2024) which presents an in-depth analysis of the FSS and its application in verifying ensemble forecasts.

Methodology

The study explores four different approaches to computing the FSS for ensemble forecasts:

- Probabilistic FSS (pFSS): This method calculates probabilities both within the ensemble and for neighborhood regions surrounding forecast points, providing a spatial and probabilistic comparison.
- Member-Averaged FSS (avFSS): FSS is computed for each individual ensemble member, and then the average is taken across all members.
- Member-Aggregated FSS (agFSS): The FSS depends on the ratio between the fractions Brier score (FBS) and the worst possible fractions Brier score (WFBS). This method computes FBS and WFBS by aggregation over the ensemble members, before calculating the FSS.
- Ensemble Mean FSS (emFSS): Instead of considering each ensemble member separately, this method uses the ensemble mean forecast to compare with the observed data.

To conduct the analysis, the authors used high-resolution precipitation forecasts from a 1000-member ensemble dataset over Germany during a severe weather event in 2016. The large ensemble allowed for an extensive evaluation of how ensemble size, neighborhood size, and the frequency of forecasted events affect the forecast skill.

Results

The study’s primary findings are based on a comparative analysis of the four FSS approaches, with the pFSS emerging as the most reliable and well-behaved method for verifying ensemble forecasts.

a. Dependence on Ensemble Size

The pFSS increases steadily with ensemble size, but it reaches a saturation point around 200 ensemble members (Figure 1). This means that larger ensembles (beyond 200 members) do not substantially improve the forecast skill judged by the pFSS. The study also derives a formula that describes the ensemble size dependence of the pFSS, which allow predicting its convergence with ensemble size. This finding is crucial for optimizing computational resources in weather forecasting models, as increasing ensemble size beyond a certain point provides diminishing returns in terms of improved accuracy. The other computation methods (avFSS, agFSS,

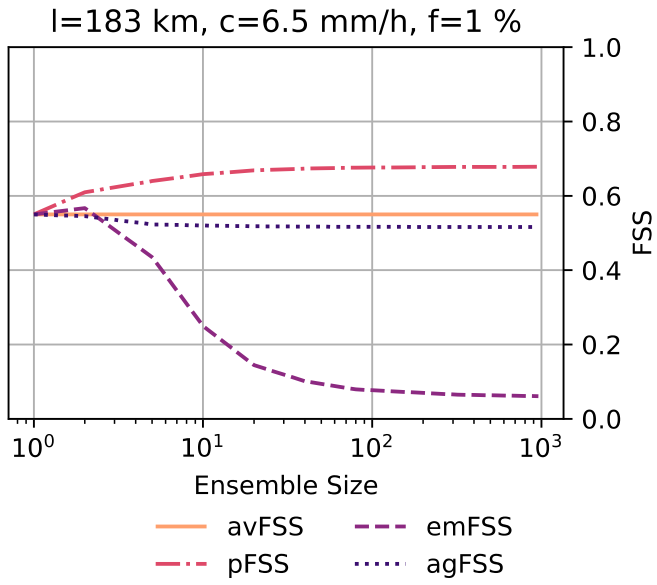

Figure 1: Comparison of four different approaches for computing a fractions skill score (FSS) for an ensemble. The FSS as a function of the ensemble size for neighborhoods of 183 km and a frequency of occurrence of 1%. When reducing the ensemble size, all approaches converge to the deterministic FSS. Note that the *x*-axis is a logarithmic scale.
agFSS: member-aggregated FSS;
avFSS: member-averaged FSS;
emFSS: ensemble mean FSS;
pFSS: probabilistic FSS

emFSS) return an ensemble FSS that is either invariant with ensemble size or worse for a larger ensemble. Thus, they do not make proper use of the information content of the ensemble forecast.

b. Influence of Neighborhood Size

Just like the FSS, the pFSS improves as the neighborhood size increases, because larger neighborhoods make the forecast less sensitive to small shifts in the location of predicted events. This mitigates the double-penalty effect. However, the trade-off is that larger neighborhoods reduce the spatial precision of the forecast.

c. Equivalence of pFSS and Brier Skill Score BSS

The pFSS revealed a distinct and predictable behaviour with changing ensemble size, similar to the BSS This behaviour can be expected as the pFSS and BSS exhibit similarities. The ensemble-size dependence of pFSS is described with an equation that depends on a single empirical parameter, α. The theoretically predicted ensemble-size dependence of the BSS corresponds to $\alpha = 1$. Further, the BSS corresponds to the pFSS with a neighbourhood size $g = 1$. The probability sample size of the BSS equals the number of ensemble members, whereas that of the FSS equals the number of grid points in the neighbourhood.

Conclusion

The study recommends using the **pFSS** as the standard method for verifying ensemble forecasts, especially in probabilistic settings where uncertainty needs to be accounted for. Unlike other ensemble FSS scores, the pFSS properly weighs the information on forecast uncertainty brought by additional ensemble members. Thus, it increases monotonically with ensemble size, as one would expect. By applying the pFSS, meteorologists and weather agencies can better understand the limitations and strengths of ensemble-based forecasts, particularly when dealing with high-impact weather events like heavy rainfall. This method offers a clear pathway to improving forecast verification techniques, ultimately leading to more accurate weather predictions.

Reference

Necker, T., Wolfgruber, L., Kugler, L., Weissmann, M., Dorninger, M. & Serafin, S. (2024) The fractions skill score for ensemble forecast verification. *Quarterly Journal of the Royal Meteorological Society*, 1–21. Available from: https://doi.org/10.1002/qj.4824