Application of Spatial-Temporal Fractions Skill Score to high-resolution ensemble forecast verification
Le Duc (JAMSTEC), Kazuo Saito and Hiromu Seko (MRI)

1. Introduction

With the realism of very-high resolution models with grid spacing down to 1 km, the traditional precipitation verification methods seem not to be adequate. Figure 1a. compares hourly precipitation forecasts for two ensemble systems with the resolutions of 10 km (MF10km) and 2 km (MF2km) respectively in one month using the Brier skill score as a measurement. And also the more detail forecast behavior with respect to heavy rainfalls (using threshold 10mm/hour) is shown in Fig. 1b. The objective verification results did not support the outperformance of higher resolution forecasts, especially for heavy rainfalls, as expected, although subjective evaluations would easily identify this (not shown here). This fall of traditional scores with high resolution precipitation forecasts is attributed to the short predictability of small-scale processes, which cause a phenomenon called the double penalty (the increase of both false alarm and miss areas when forecasts are shifted from observations) on these scores.

![Fig. 1. a) BSS values of hourly precipitation forecasts from two ensemble systems MF10km and MF2km (the shaded area is the 95% confidence interval) and b) reliability diagram with respect to 10 mm/hour threshold and 5 km verification grid size.](image)

Double penalty problem occurs not only by spatial mismatch but also by timing lag. To alleviate this, current QPF verifications at JMA are mostly conducted for 3-hour precipitation. However, intense rain in short time (e.g., one hour) is sometimes more hazardous in the view point of urban-type disaster prevention.

To account for the spatial accuracy of high-resolution precipitation forecasts, a new score named Fractions Skill Score (FSS) has been proposed by Roberts and Lean (2008). This score relaxes the exact match point-point requirement between observations and forecasts by taking into account neighbor areas. By extending this concept of spatial neighborhood to temporal one, in this study, we will verify the value of cloud-resolving ensemble forecast in predicting 1-hour precipitation.

2. Fractions Skill Score

The main idea of FSS is based on forecast and observation fractions inside a spatial neighborhood or window, assumed to be a square or circle area centered at each verification point. With a given precipitation threshold and for each neighborhood, the forecast and observation fraction $p_f$, $p_o$ are computed as the ratio between the occurrence of the event and the number of grid points in this window. By moving the window over all grid points $(i,j)$ in the verification domain, a score named Fractions Brier Score FBS can be defined the same as the traditional Brier Score:

$$FBS = \frac{1}{N} \sum_{i,j}^N [(p_f(i,j) - p_o(i,j))^2],$$

where $N$ is the number of points of the verification grid. By normalizing FBS, we will have FSS. The detail mathematical treatment can be seen in Roberts and Lean (2008).
The concept of spatial fractions above can extend seamlessly to temporal domain. Instead of a neighbor area in space, a neighborhood should be rather understood as a spatial-temporal window. The mathematical treatment will be identical to the case of spatial neighborhoods taking into account an additional dimension, that is the time dimension.

FSS as defined by Roberts and Lean (2008) is applied for deterministic precipitation forecasts. If each member from an ensemble forecast is considered as a possible realization of the true state, the FSS idea can be applied for ensemble forecasts by including all realizations into each neighborhood. This is somehow similar to adding a new dimension to spatial-temporal neighborhoods.

3. Results

The forecast data in the introduction will be used here. Detail description about the ensemble systems and forecast results is shown in Seko et al. (2010). All results were aggregated for whole periods from 12-hour to 22-hour forecast ranges by MF10km which is correspondent to 6-16 forecast ranges by MF2km nested inside MF10km with time lag of 6 hours. Figure 2. represents the ensemble FSSs in comparison with the deterministic FSSs from all members. The ensemble FSSs tend to have larger values than these of other members, including the control forecasts.

Intercomparison between MF10km and MF2km using ensemble FSS is illustrated in Fig. 3. The figures clearly show that MF2km is better than MF10km in heavy rainfall forecasts significantly. In contrast, lower-resolution forecasts MF10km beats high-resolution forecasts MF2km for light rains, suggesting that the horizontal resolution of 2 km is not necessarily fine enough to completely remove the convective parameterization. For moderate rains, two systems exhibit similar performances when the FSS differences are not statistically significant. This new insight into the model performances is considerably one of advantages of FSS in comparison with the traditional scores like BSS as in Fig.1.

![Fig. 2.](image-url) a) Variation of FSSs of hourly precipitation forecasts with spatial scales in the 3-hour window from a) MF10km and b) MF2km forecasts. Besides ensemble FSSs, FSSs from all members and ensemble means are also represented.

![Fig. 3.](image-url) a) Ensemble FSSs of hourly precipitation forecasts against spatial scales from MF10km and MF2km in the 3-hour window. The shaded areas represent the 95% confidence intervals of FSS differences, which is shifted to the FSSs of MF2km. b) Corresponding FSS differences in intensity-scale diagram. A number in each box indicates the confidence level saying that the difference value is significant larger than 0.