

1 Verification scores

1.1 Ensemble mean

The ensemble mean is obtained by averaging all ensemble forecasts:

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n \quad (1)$$

where \mathbf{x}_n is n th ensemble member, N is the number of ensemble members.

1.2 Ensemble spread

The ensemble spread measures the differences between the members in the ensemble forecast. The ensemble spread is the rms-difference between the ensemble members and ensemble mean defined as:

$$spread = \sqrt{\frac{1}{N} \sum_{n=1}^N (\mathbf{x}_n - \bar{\mathbf{x}})^2} \quad (2)$$

where N is the number of ensemble members, \mathbf{x}_n is n th ensemble member. A small (large) spread indicates low (high) forecast uncertainty. However, a small (large) spread does not necessarily indicate high (low) skill, although it could give an indication of high (low) predictability.

1.2.1 Root Mean Square Error (RMSE)

The RMSE is defined by the following equation:

$$RMSE = \sqrt{\frac{1}{D} \sum_{d=1}^D (x_f^d - x_a^d)^2} \quad (3)$$

where x_f^d and x_a^d indicate the forecast and analysis values at the grid-point d , respectively. D is the number of grid points in the spatio-temporal for seasonal score or temporal domains for daily score, namely which indicates all grid points over the NH in the verification period. **Each control run at the initial time is regarded as each analysis. Note that each single-center ensemble is verified against its own analysis.** The RMSE indicates a forecast error, and the RMSE score of zero (0.0) demonstrates a perfect skill. The RMSE is expected to be comparable with the ensemble spread at the same verification time.