Optimal ensemble size for S2S prediction system

Sang-Wook Kim and Seok-Woo Son

School of Earth and Environmental Sciences, Seoul National University, Republic of Korea

spica557@snu.ac.kr

1. INTRODUCTION

- More extensive ensemble size increases the prediction skill of forecast model in an ideal (unbiased and fully reliable) model condition (Murphy, 1988); however, since the limit of computational resource, the optimal number of the ensembles should be considered for the efficient forecast system.
- Moreover, sub-seasonal to seasonal (S2S) prediction system is hard to be assumed as an ideal model, since it contains the considerable model bias, so that enlarging ensemble size does not always ensure increasing prediction skill.
- To determine optimal ensemble size for S2S prediction system, the effect of increasing ensemble size would be explored in a practical model with ECMWF real-time forecast that consists of 51 ensemble member. A diagnostic methodology for whether the increase in ensemble size can improve prediction skill is also examined.

3. DATA

- ECMWF real-time forecast in S2S database (Vitart et al., 2017)

Variable: Geopotential height

Level: 500-hPa (troposphere) / 50-hPa (stratosphere) pressure level

Period: 2016-2018 DJF 00UTC instantaneous forecast

Compared with ERA-Interim data (Climatology period: 1981-2010)

Table 1. The information about the ECMWF model (on the fly) configuration.

Update Date	Model Version	Time Range	Ens. Size	Atm. Resolution	Ocean Resolution	Active Sea Ice
14/05/2015	CY41R1	d 0-46	51	32 km (~10d) 64 km (10d~)	1 degree	NO
08/03/2016	CY41R2	d 0-46	51	16 km (~15d) 31 km (15d~)	1 degree	NO
22/11/2016	CY43R1	d 0-46	51	16 km (~15d) 31 km (15d~)	1/4 degree	YES
11/07/2017	CY43R3	d 0-46	51	16 km (~15d) 31 km (15d~)	1/4 degree	YES

2. THEORETICAL BASIS

- MSSS (Mean Square Skill Score)

$$MSE_f(\tau,e)=\frac{1}{N_f}\sum_{i=1}^{N_f}\left(\sum_{j=1}^{N_g}\left(\hat{f}_{ij}(\tau,e)-o_{ij}(\tau)\right)^2 w_j\right)\quad where\ \hat{f}_{ij}(\tau,e)=\frac{1}{e}\sum_{k=1}^{e}f_{ij}(\tau,k)\ and\ w_j=\frac{\cos\theta_j}{\sum_{j=1}^{N_g}\cos\theta_j}$$

$$MSE_c(\tau)=\frac{1}{N_f}\sum_{i=1}^{N_f}\left(\sum_{j=1}^{N_g}\left(c_{ij}(\tau)-o_{ij}(\tau)\right)^2 w_j\right)$$

$$MSSS(\tau,e)=\frac{MSE_f-MSE_{ref}}{MSE_{perfect}-MSE_{ref}}=\frac{MSE_c-MSE_f}{MSE_c}=1-\frac{MSE_f}{MSE_c}$$

f: forecast; o: observation
c: climatology; θ: latitude
τ: lead time
e: ensemble size ($e=1,2,\cdots,N_e$)
i: initialization date ($i=1,2,\cdots,N_f$)
j: grid point ($j=1,2,\cdots,N_g$)

- Estimation of Prediction Skill

$$MSE_f(\tau,e)=\left\langle\left(\hat{f}_{ij}(\tau,e)-o_{ij}(\tau)\right)^2\right\rangle=\left\langle\left(\frac{1}{e}\sum_{k=1}^{e}\left(f_{ij}(\tau,k)-\bar{f}_{ij}\right)\right)^2+\left(o_{ij}-\bar{f}_{ij}\right)^2-2\left(\hat{f}_{ij}-\bar{f}_{ij}\right)\left(o_{ij}-\bar{f}_{ij}\right)\right\rangle$$

[assuming fully reliable ensemble spread]

$$=\left\langle\frac{1}{e}\left(\frac{1}{e}\sum_{k=1}^{e}\left(f_{ij}(\tau,k)-\bar{f}_{ij}\right)^2\right)+\left(o_{ij}-\bar{f}_{ij}\right)^2\right\rangle=\frac{\langle D\rangle}{e}+\left\langle\left(o_{ij}-\bar{f}_{ij}\right)^2\right\rangle\quad where\ D\ is\ ens.spread$$

[assuming ensemble spread = natural variability]

$$MSE_f(\tau,e)=\frac{e+1}{e}\langle D\rangle,\qquad MSE_f(\tau,1)=2\langle D\rangle$$

$$MSE_f(\tau,e)=\frac{e+1}{2e}\overline{MSE_f(\tau,1)}$$
$$MSE_f(\tau,\infty)=\frac{1}{2}\overline{MSE_f(\tau,1)}$$

$$MSSS(\tau,e)=1-\frac{MSE_f(\tau,e)}{MSE_c(\tau)}=1-\frac{e+1}{2e}\frac{\overline{MSE_f(\tau,1)}}{MSE_c(\tau)}$$

$$MSSS(\tau,e)=1-\frac{e+1}{2e}\left(1-\overline{MSSS(\tau,1)}\right)$$
$$MSSS(\tau,\infty)=\frac{1}{2}\left(1+\overline{MSSS(\tau,1)}\right)$$

Figure 1. The theoretical MSSS in an ideal prediction system with a certain number of ensemble members. The colors indicate the averaged skill score of individual ensemble members.

4. RESULTS

- Prediction Skill of ECMWF model versus Lead Time

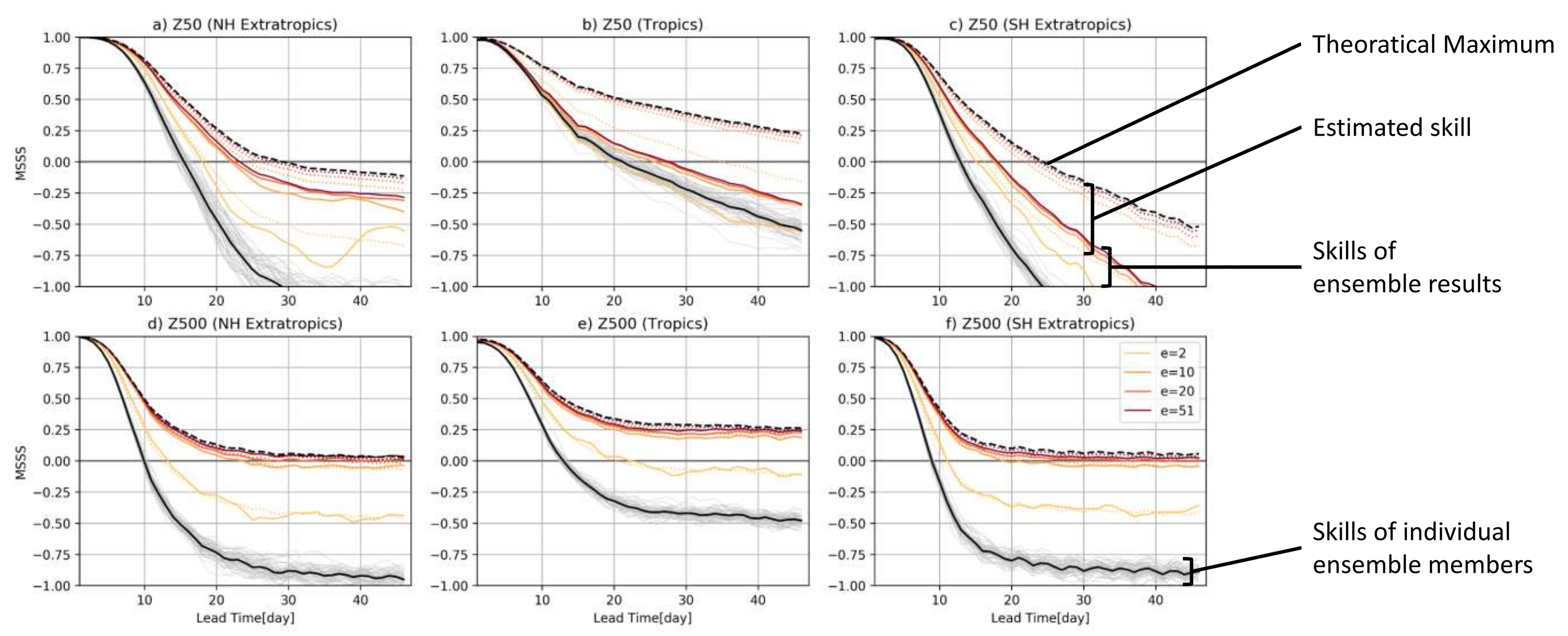

Figure 2. The MSSS of individual ensemble members (gray lines), and ensemble results (colored solid lines) and the estimated MSSS at a certain number of ensemble members (colored dotted lines) of 50-hPa (top) and 500-hPa (bottom) pressure level geopotential height forecasts initialized in DJF. The averaged skill of individual ensemble members and the theoretical skill limit are represented by a black solid line and a dashed line, respectively.

- Difference of Statistical Characteristics between OBS and FCST

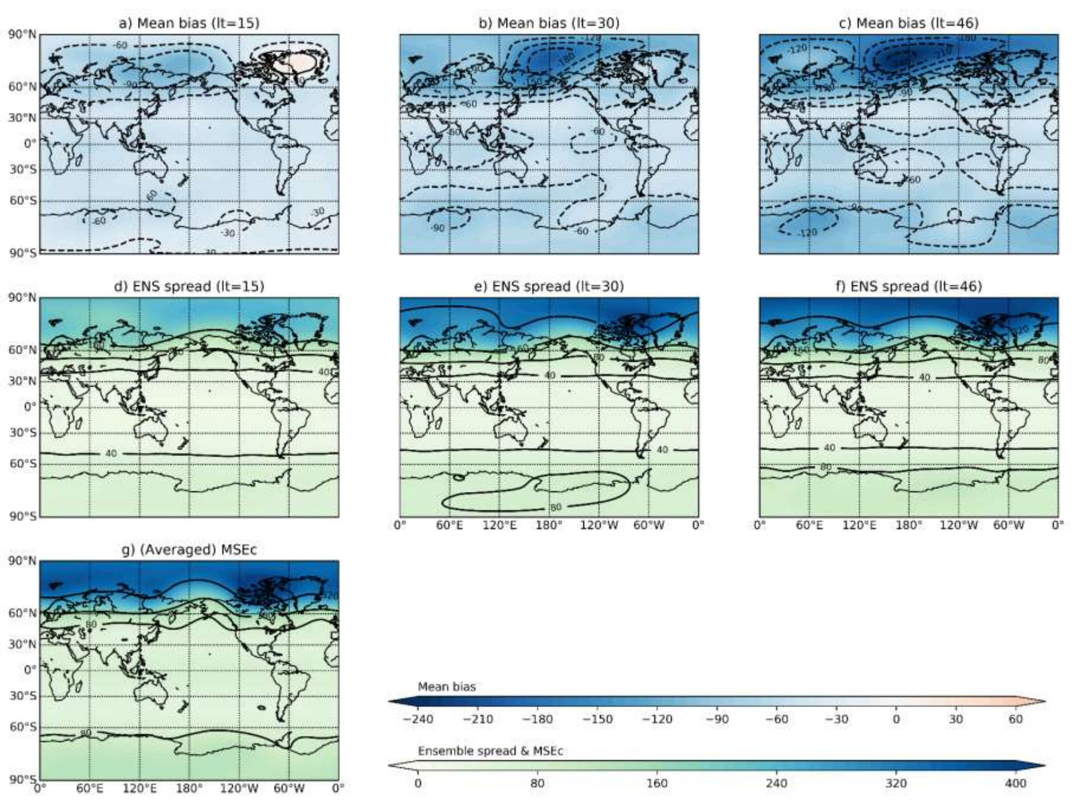

Figure 3. Mean bias (top) and square root ensemble spread (middle) distribution of 50-hPa pressure level geopotential height forecasts at the lead time of 15 (left), 30 (center), and 46 (right) days initialized in DJF. The mean distribution of square root MSE_c is represented on the bottom line.

In Z50 tropics and SH extratropics, small bias, ensemble spread, and MSE_c

In Z50 NH extratropics, huge bias, ensemble spread, and MSE_c

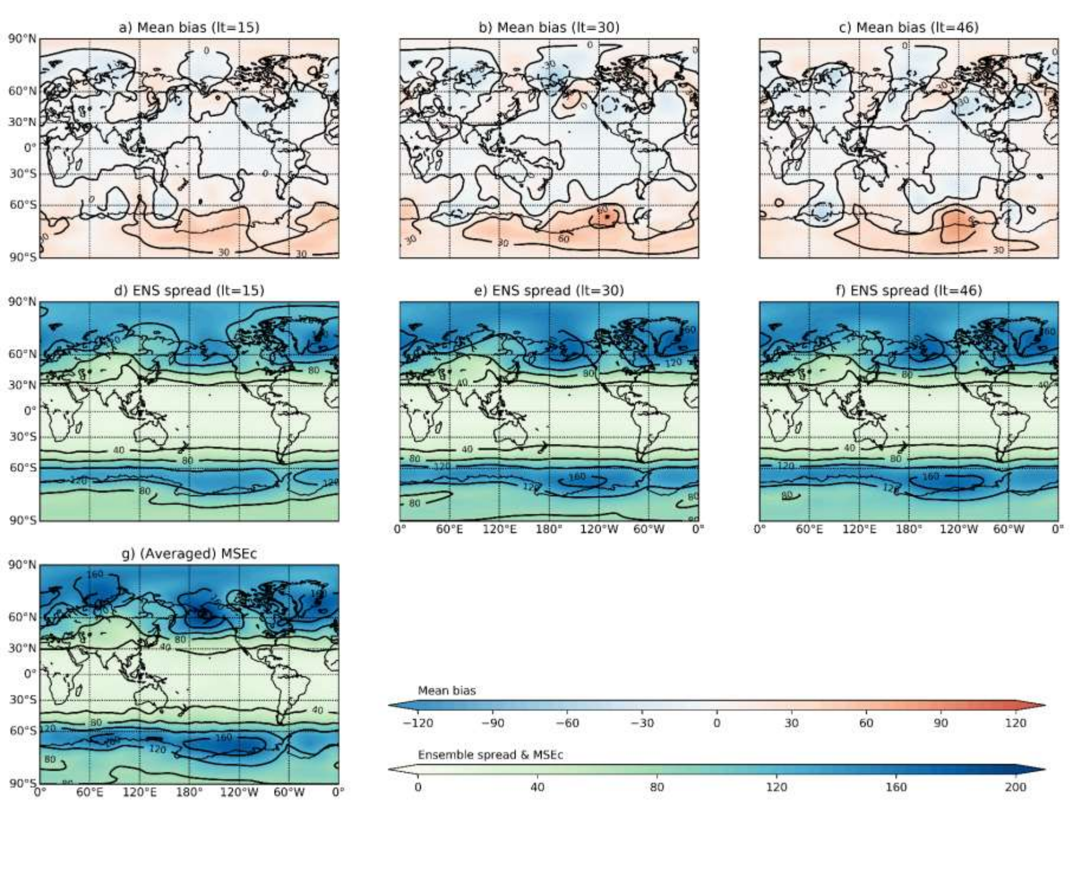

Figure 4. Same as Figure 3, but for 500-hPa pressure level geopotential height forecasts.

Mean bias

$$MB_j(\tau)=\frac{1}{N_f}\sum_{i=1}^{N_f}\left(\bar{f}_{ij}(\tau)-o_{ij}(\tau)\right)$$

$$\text{where } \bar{f}_{ij}(\tau)=\frac{1}{N_e}\sum_{e=1}^{N_e}f_{ij}(\tau,e)$$

Ensemble spread

$$Esprd_j(\tau)=\frac{1}{N_f N_e}\sum_{i=1}^{N_f}\left(\sum_{e=1}^{N_e}\left(f_{ij}(\tau,e)-\bar{f}_{ij}(\tau)\right)^2\right)$$

Natural variablilty

$$MSE_{c_j}(\tau)=\frac{1}{N_f}\sum_{i=1}^{N_f}\left(\left(c_{ij}(\tau)-o_{ij}(\tau)\right)^2\right)$$

- Diagnosis of the Coincidence between two PDFs

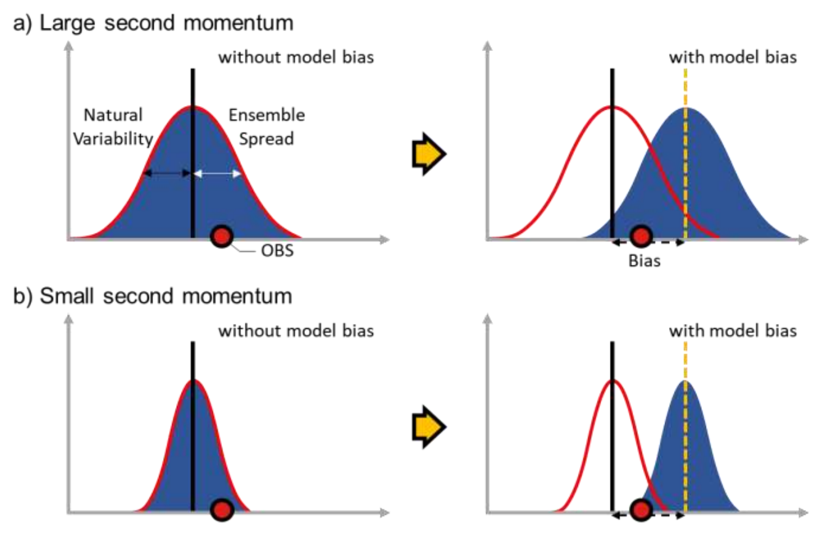

Figure 5. Schematic diagram of the probability density function (PDF) of observed state (red line) and model ensembles (blue shaded) when PDFs are having large (a) and small (b) second momentum. Without the model bias, the model ensemble PDF can capture the observation regardless of the size of the second momentum (left); however, when bias occurs, the PDF with the smaller second momentum is relatively hard to capture the observation (right).

$$H_0: MB_j=0,\qquad H_1: MB_j\neq 0\qquad |t_j(\tau)|=\frac{\left|\frac{1}{N_f}\sum_{i=1}^{N_f}\left(\bar{f}_{ij}-o_{ij}\right)\right|}{\sqrt{\frac{(N_f\times N_e-1)s_f^2+(N_f-1)s_o^2}{N_f\times N_e+N_f-2}}}\approx\frac{\left|MB_j(\tau)\right|}{\sqrt{\frac{(N_f\times N_e-1)Esprd_j(\tau)+(N_f-1)MSE_{c_j}(\tau)}{N_f\times N_e+N_f-2}}}$$

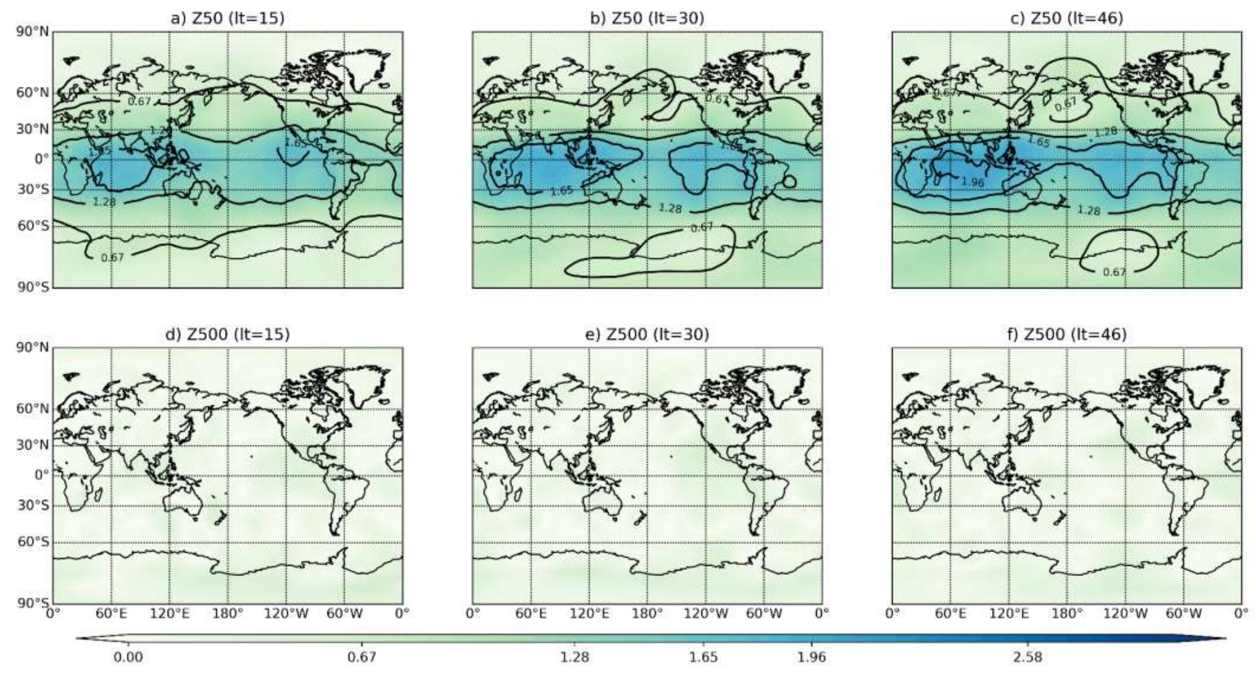

Figure 6. Distribution t-value of 50-hPa (top) and 500-hPa (bottom) pressure level geopotential height forecasts at the lead time of 15 (left), 30 (center), and 46 (right) days initialized in DJF.

$t>0.67$	H_0 is rejected with 50% confidence level
>1.28	80% confidence level
>1.65	90%
>1.96	95%
>2.58	99%

- Prediction Skill of ECMWF model versus Lead Time with Bias Correction

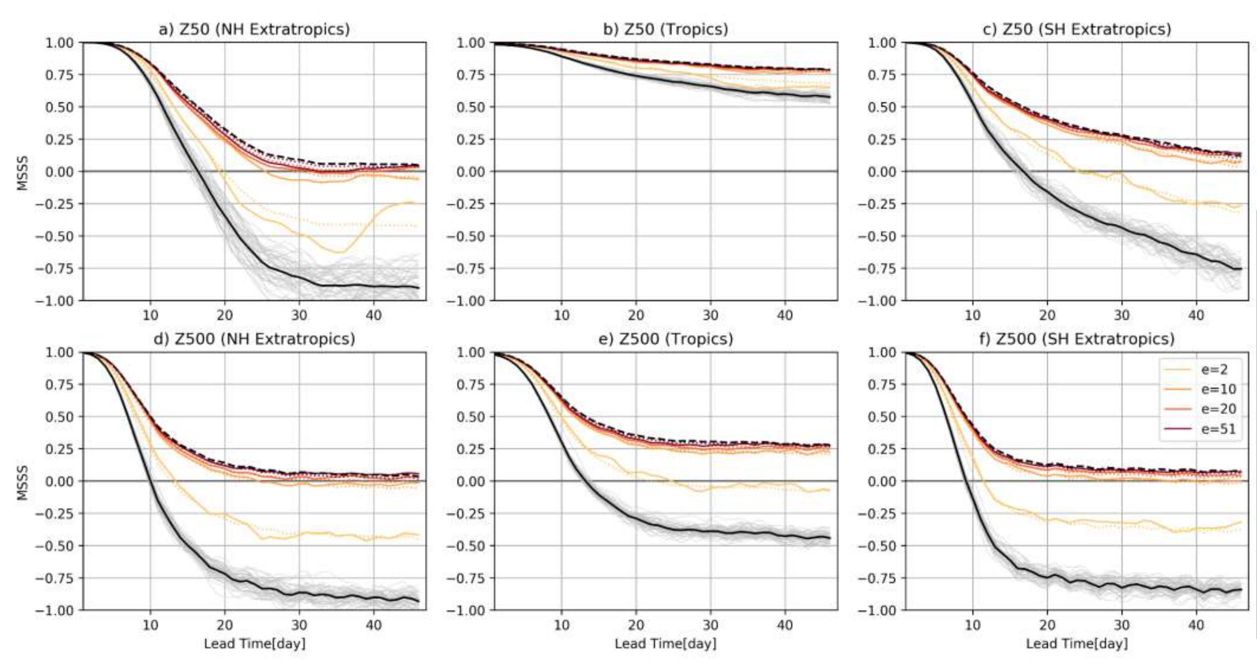

Figure 7. Same as Figure 2, but for the result with bias correction.

Quantitative skill improvement at the stratosphere for both individual members and ensemble results

Qualitative enhancement of the effect of enlarging ensemble size

Little changes in the troposphere implying that the error related to mean bias is small at this level

5. SUMMARY

- The tropospheric prediction skill, verified for 500-hPa geopotential height field, is in a good agreement with the theoretical estimation; however, the stratospheric skill, evaluated at 50 hPa, is substantially lower than the estimation.
- The disagreement between a practical skill and the estimation stems from differences of statistical characteristics, such as mean bias, ensemble spread, and natural variability.
- Through the diagnosis based on two sample t-test with equal variance, it is possible to figure out the coincidence of two PDFs (the forecast state of model ensembles and the observed state) and to guarantee the improvement of prediction skill with increasing ensemble size.
- Bias correction not only improves the prediction skill, but also it helps to ensure the fact that the extension of ensemble size increases the prediction skill, through compensating the difference between the two PDFs.