Observation-Based

Ensemble Spread-Error Relationship

Twenty-ninth session of
the WWRP/WCRP Working Group on Numerical Experimentation

Melbourne, Australia

10-14 March 2014

Munehiko Yamaguchi[1,2], Simon Lang[1], Martin Leutbecher[1],
Mark Rodwell [1], Gabor Radnoti[1], Niels Bormann[1] and
Jean-Noël Thépaut[1],

1: European Centre for Medium-Range Weather Forecasts
2: Meteorological Research Institute, Japan Meteorological Agency

Introduction

- Assessing the performance of numerical weather prediction (NWP) including ensemble prediction is of great importance in order to **understand the skill of the prediction**, **identify the deficiencies and limitations**, and **further improve the NWP system** (model dynamics, physics, data assimilation, etc.).

- A conventional way to do so is analysis-based evaluations. In other words, **predictions are evaluated against their own analyses**, which are the outcomes of a data assimilation used in the NWP system yielding the predictions.

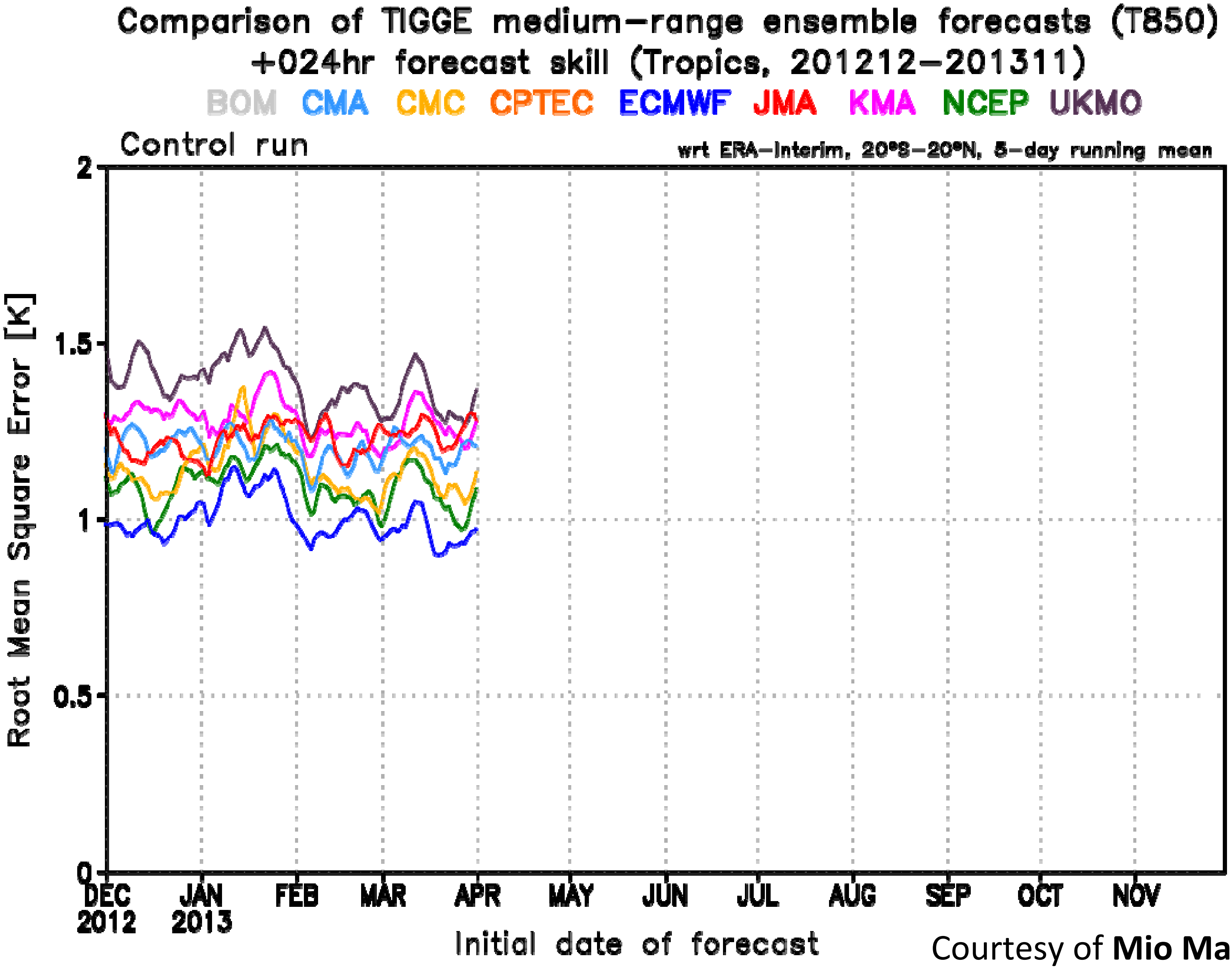
- However, **verifying against own analyses often creates a problematic situation in the interpretation of the verification results** (Brown et al. 2012 WGNE Final Report).

Example
Problematic situations in the analysis-based verification

Park et al. (2008) demonstrate using TIGGE data set **that the advantage of verifying predictions against their own analyses can be large**. Since NWP models for predictions are also used in data assimilation schemes to provide so-call first-guess fields, the analysis and forecast fields are in the same model world (as opposed to the real world). If they have the same or similar systematic error patterns, the advantage of the own analysis-based verification could be rather large.

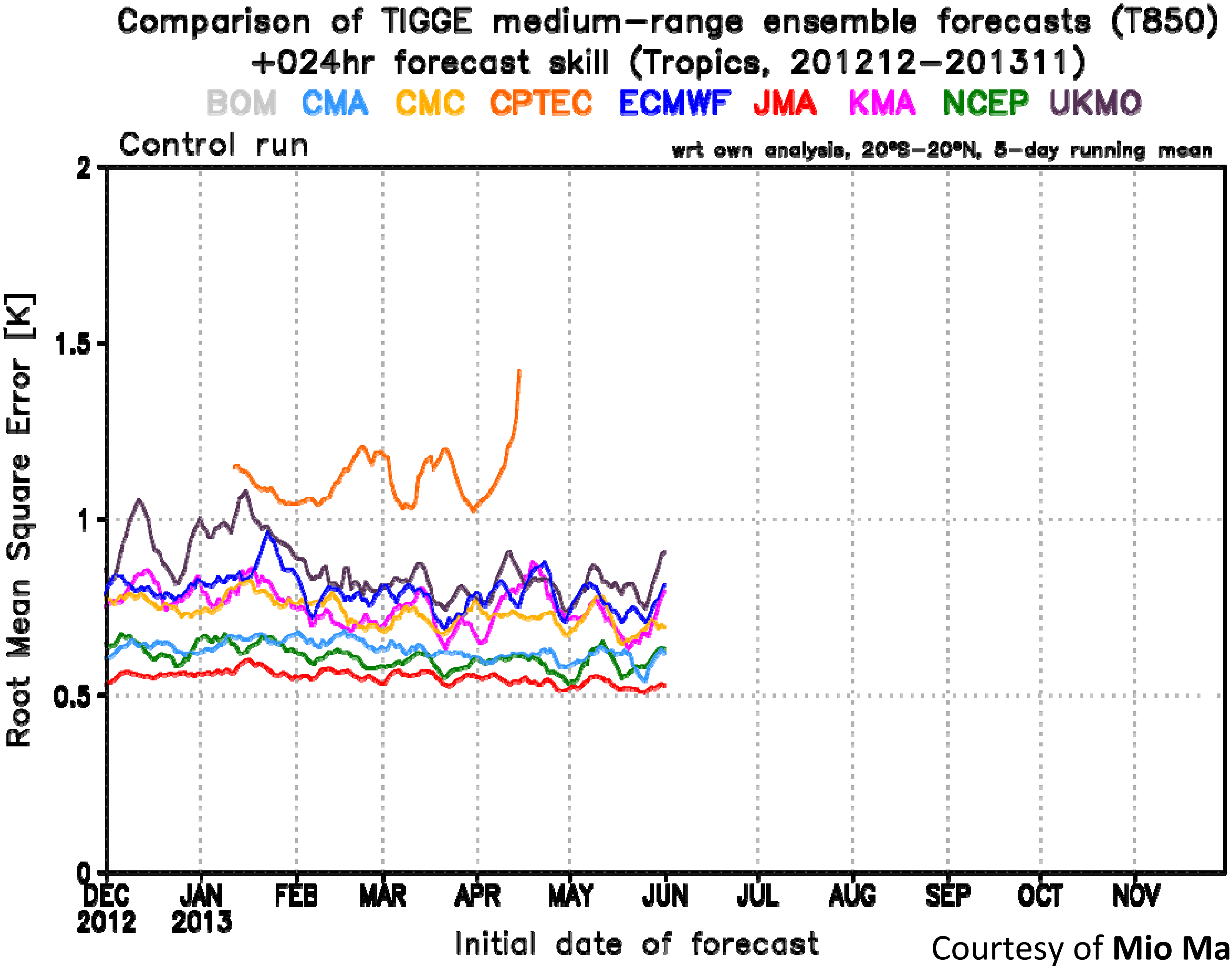
Another troublesome situation associated with the analysis-based verification is when new observations are added to a data assimilation scheme. As new observations can produce significant changes in analysis fields, especially in data-sparse areas, relative to changes in forecast fields, **the verification results may look worse when the forecasts are verified against the analyses with the new observations assimilated** (Bouttier and Kelly 2001). **Such a questionable degradation of forecast performance can be seen more distinctively at short range forecasts** (Geer et al. 2010).

T850 TR(20S-20N) FT=024 **against ERA**



Courtesy of **Mio Matsueda**

T850 TR(20S-20N) FT=024 **against own analysis**



Courtesy of **Mio Matsueda**

Motivation of this study

- Ensemble forecasts are usually verified against their own analyses and the limited number of conventional data (e.g. radiosondes, synop)

- Satellite observations are less used. In this study, **ensemble forecasts are verified against satellite observations** using a newly developed verification tool at ECMWF.

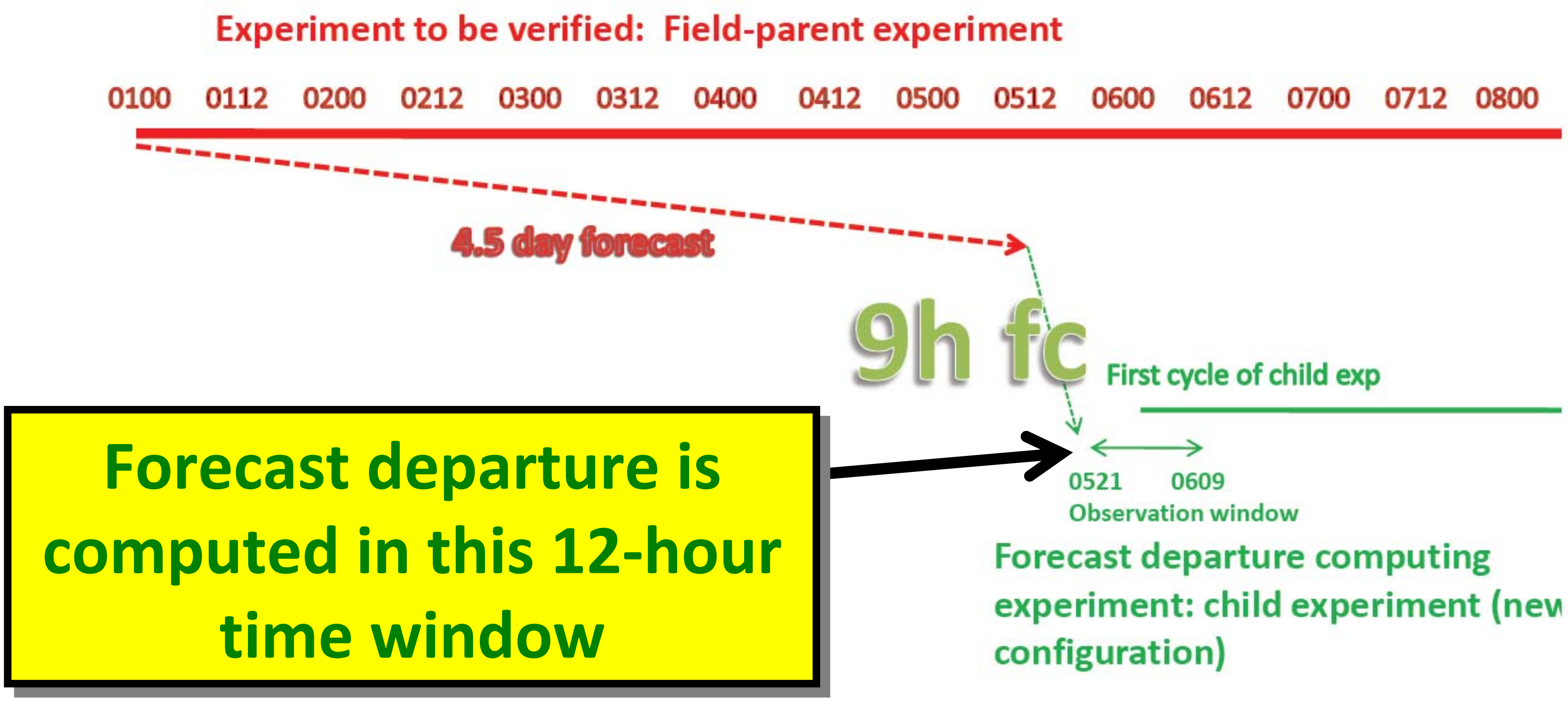
- Verifying ensemble forecasts against satellite observations will provide **additional views of understanding the forecast performance**.

- Moreover, it may lead to **future possible developments and/or amendments of the ensemble forecasting system**.

What we have done 1/2

1. We have developed a new verification tool called “**forecast obstat**”, in which (deterministic) forecasts are verified against observations including a various types of satellite observations.

Schematic of the Forecast Obstat (example of verifying 5-day forecast)



Forecast departure is computed in this 12-hour time window

What we have done 2/2

2. We modified the forecast obstat so that it could handle with ensemble forecasts (hereafter referred to as **ensemble forecast obstat**: EFO).

3. We modified the EFO so that both the parent and child experiments can be run with and without **stochastic physics scheme** (SPPT + SKEB).

4. Then we have conducted a set of experiments (next slide).

Ensemble Forecast Obstat Experiments

Initial time: 2012070100 – 2012091500 with a **4 day interval**

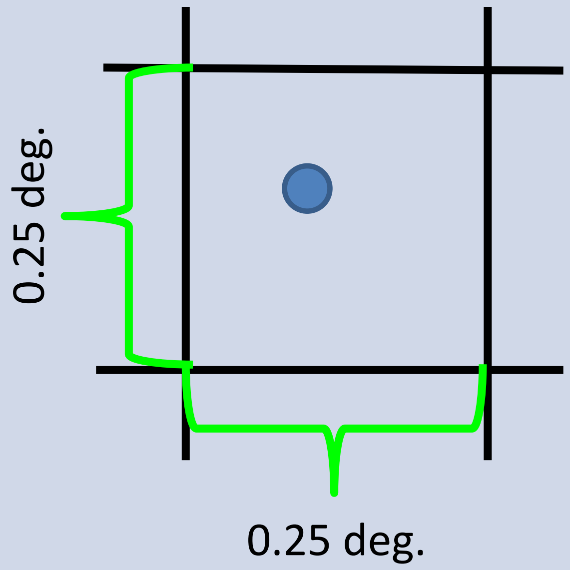
	With stochastic physics		**W/O stochastic physics**	
Resolution	**TL639L91**			
Parent exp.	**fvmd**		**fwbi**	
Child exp.	cf	pf	cf	pf
FT=24	**fvn8**	**fvn5**	**fwrg**	**fwlu**
FT=48	**fvt2**	**fvt1**	**fvt2**	**fx3h**
FT=120	**fvq3**	**fvpx**	**fvq3**	**fwre**

20 ensemble members

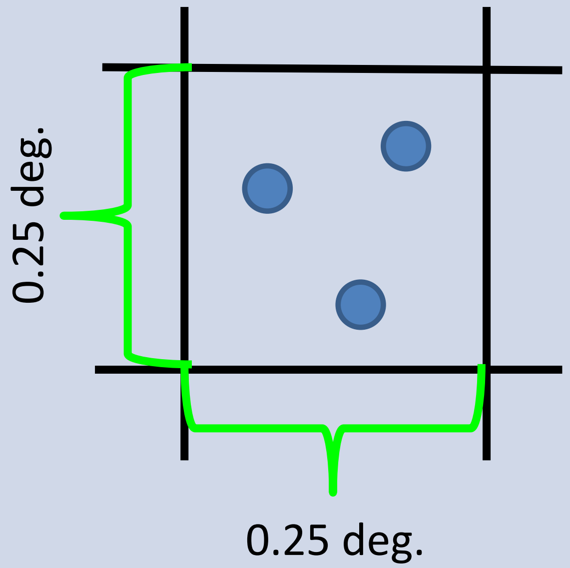
Ensemble Spread-Error Relationship

Ensemble spread-error relationship is verified. The ensemble spread and ensemble mean error is calculated in grid-point space with a horizontal resolution of 0.25 x 0.25 degrees (roughly equivalent to TL639).

When the number of observations in a grid box is one, the ensemble spread and ensemble mean error at the observation point represent the value of the grid point.



When the number of observations in a grid box is two or more, the ensemble spread and ensemble mean error at each observation point are first calculated. Then a simple average of them represents the value of the grid point.



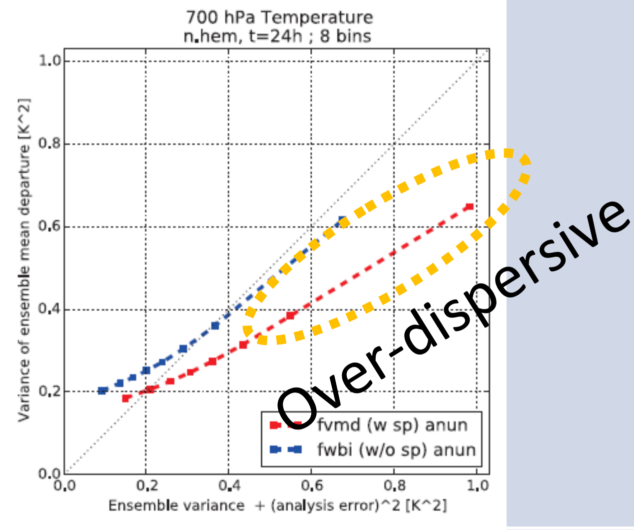
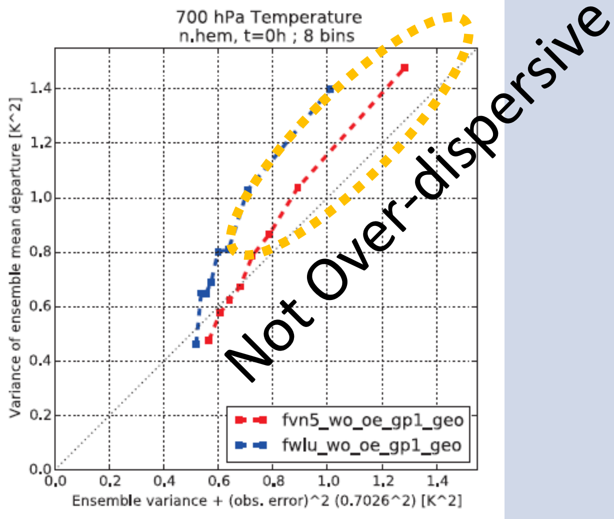
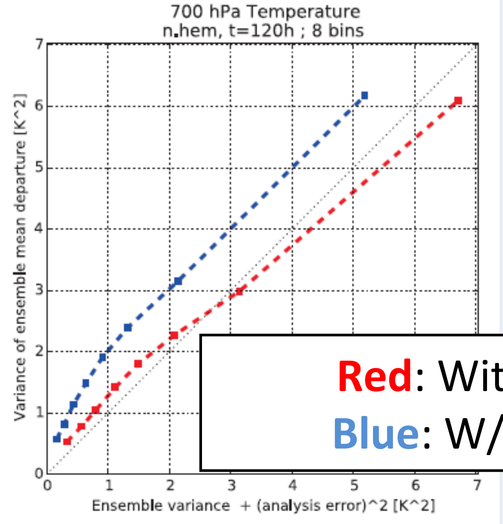
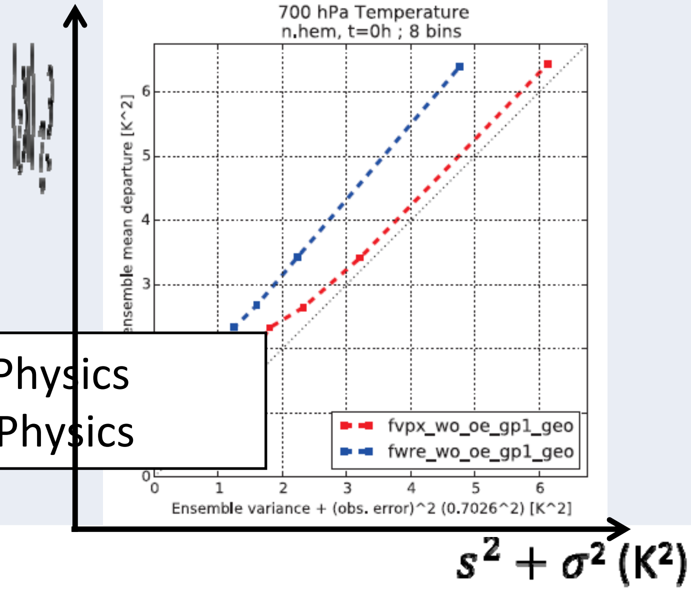
Ensemble Spread-Error Relationship -2-

For reliable ensemble,

$$\frac{1}{N}\sum_{i=1}^{N}\left(\tilde{\epsilon}_i^2 - \sigma_i^2 - \frac{M+1}{M-1}s_i^2\right) \to 0, \quad \text{for } N \to \infty. \quad (7)$$

- N: Independent cases (e.g. different dates)
- i: ith case
- $\tilde{\epsilon}_i^2$: dispersion of ensemble mean error against the observational state
- σ_i : the observation error σ_i
- s_i^2: dispersion of the ensemble
- M: Ensemble size

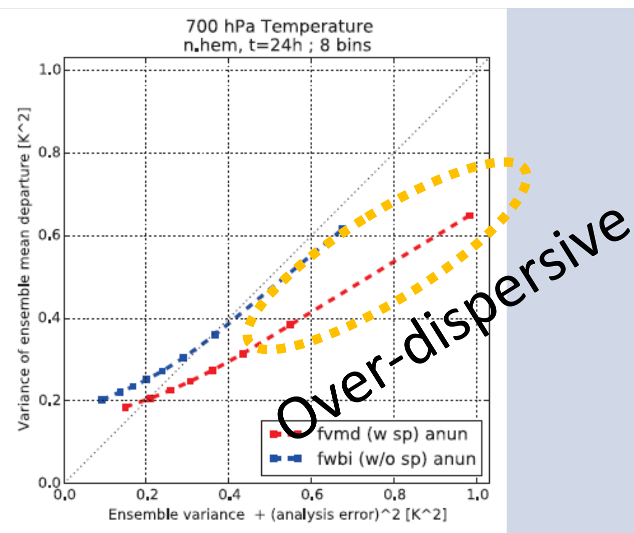
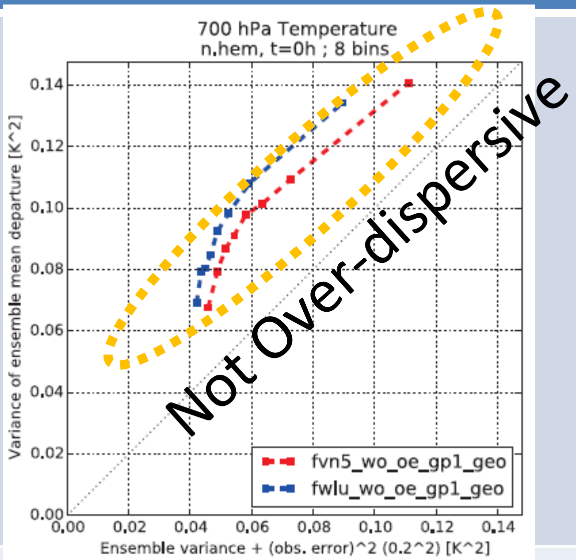
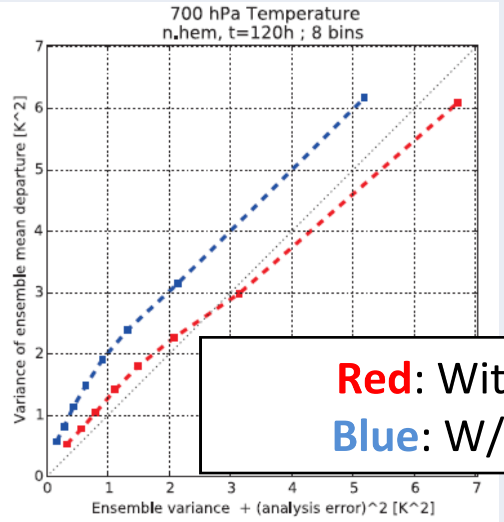
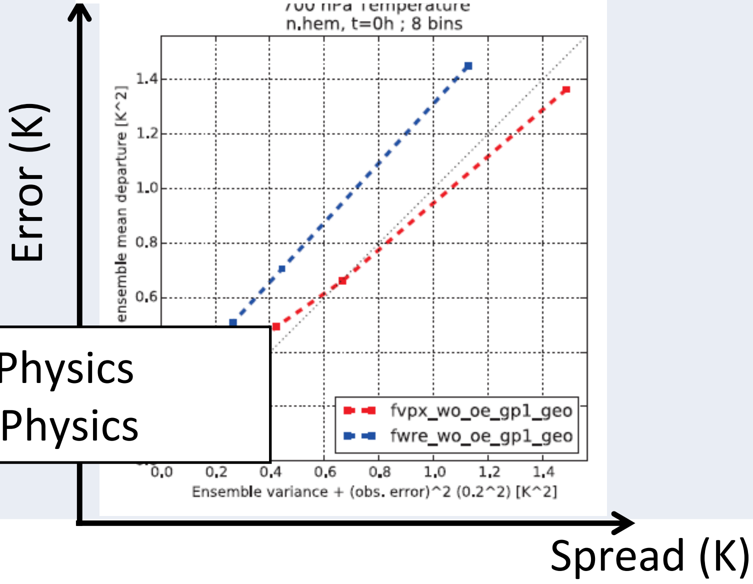
Evaluation of Spread-Error Relationship -NH T700-

FT	**Agaist analysis** (Conventional Method)	**Against observation -Radiosonde-**
24 (hrs)		
120 (hrs)		

Red: With Stochastic Physics

Blue: W/O Stochastic Physics

Evaluation of Spread-Error Relationship -NH 700-

FT	**Agaist analysis** (Conventional Method)	**Against observation -AMSUA-**
24 (hrs)		
120 (hrs)		

Red: With Stochastic Physics
Blue: W/O Stochastic Physics

Results and conclusion

The over-dispersive relationship seen in the verification against analysis is not seen in the verification against both radiosonde and AMSU-A observations at the short range forecasts.

In the 5-day verifications, the relationship looks similar between analysis- and observation-based verifications.

Similar spread/error relationship is seen in the analysis-based verification even when the verified grid boxes are limited where the AMSU-A observations exist (not shown).

Conclusion

- **Observation-based evaluations provide different views of the spread/error relationship from those of the analysis-based evaluations.**
- **Anaysis-based verification only may lead to wrong decision-making in the future improvement of the ensemble system.**