新的分析误差估计方法及其在中国气象局业务全球同化预报系统中的应用

冯杰

fengjiefj@fudan.edu.cn

大气与海洋科学系/大气科学研究院，复旦大学

王金成

中国气象局地球系统数值预报中心

致谢：戴国锟，周菲凡，段晚锁，刘永柱，张林，穆穆，雷荔傈等

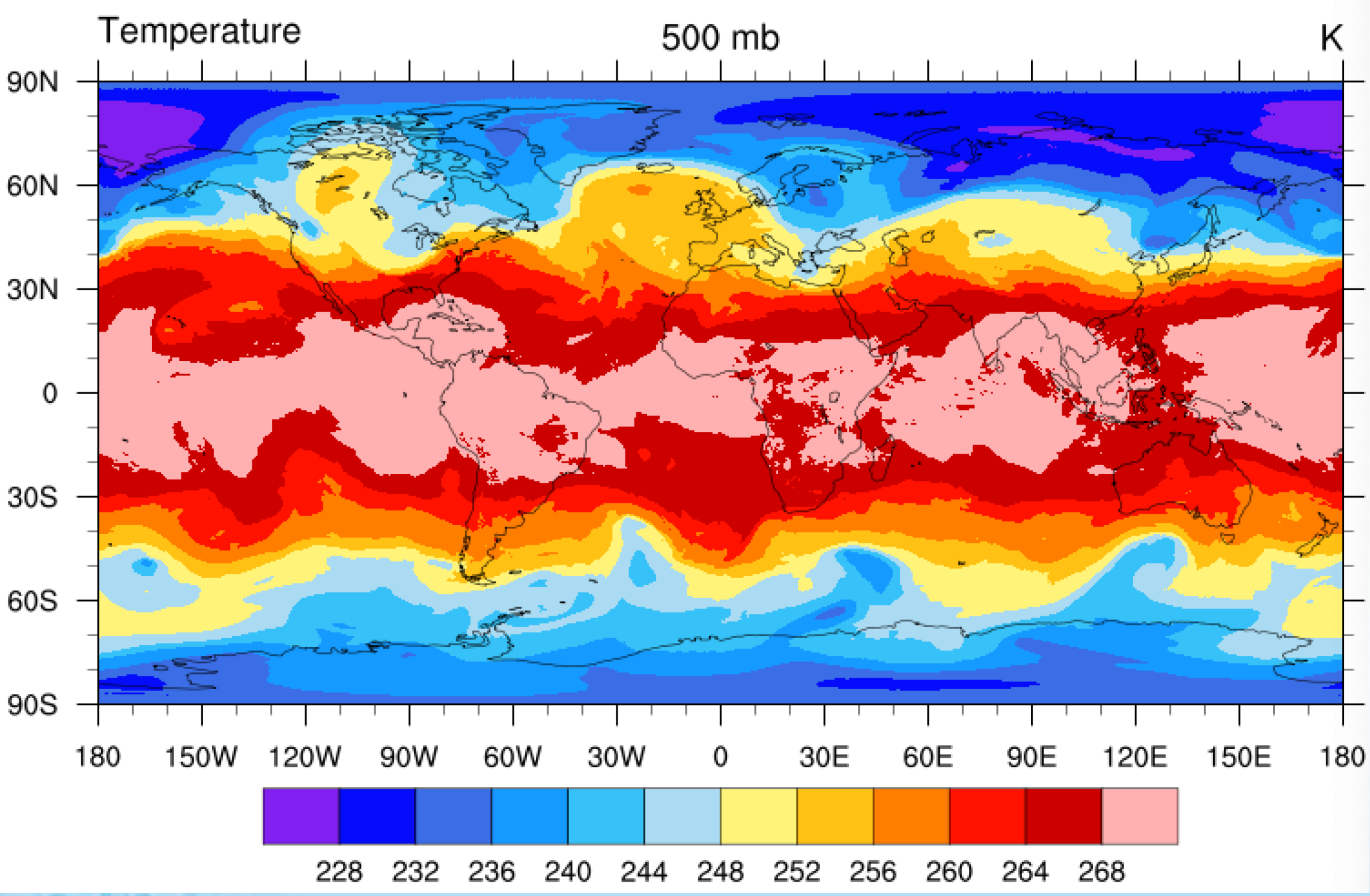
第五届全国中尺度气象学论坛，2023-8-11，银川

Outline

- **Background and motivation**
- **Methodology and validation**
- **Results of the operational GRAPES_GFS model at CMA**
- **Conclusions and discussion**

Background

A randomly selected analysis field



Analysis error (Analysis - Truth), how to measure?

Background

Accurate estimates of error variances in numerical analyses and forecasts are critical:

- Evaluation of Data Assimilation (DA) and forecast system
- Tuning of DA system
- Proper initialization of ensemble forecasts
- Understanding uncertainties in Climate reanalysis

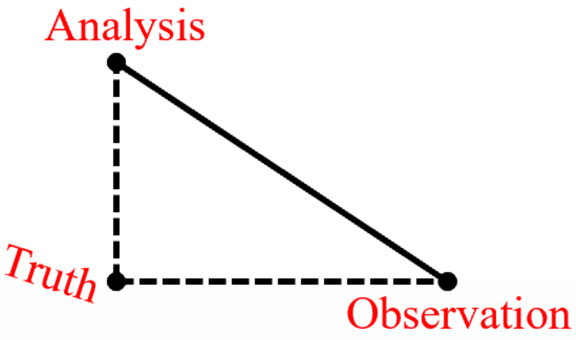
Traditional methods:

- ➢ Observations as proxy
  - Sparse observations – no gridded information
  - Fraught with observational error (including representativeness error)
- ➢ DA schemes themselves
  - Computationally expensive
  - Affected by same assumptions used in DA scheme, potentially biased/inaccurate estimates
- ➢ Analysis in other DA systems as proxy
  - Independent?
  - Errors in analysis

Background

Accurate estimates of error variances in numerical analyses and forecasts are critical:

- Evaluation of Data Assimilation (DA) and forecast system
- Tuning of DA system
- Proper initialization of ensemble forecasts
- Understanding uncertainties in Climate reanalysis



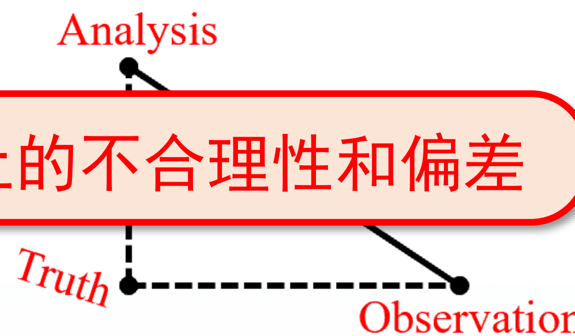
Traditional methods:

- Observations as proxy
  - Sparse observations – no gridded information
  - Fraught with observational error (including representativeness error)
- DA schemes themselves
  - Computationally expensive
  - Affected by same assumptions used in DA scheme, potentially biased/inaccurate estimates
- Analysis in other DA systems as proxy
  - Independent?
  - Errors in analysis

Background

Accurate estimates of error variances in numerical analyses and forecasts are critical:

- Evaluation of Data Assimilation (DA) and forecast system
- Tuning of DA system
- Proper initialization of ensemble forecasts
- Understanding uncertainties in Climate reanalysis

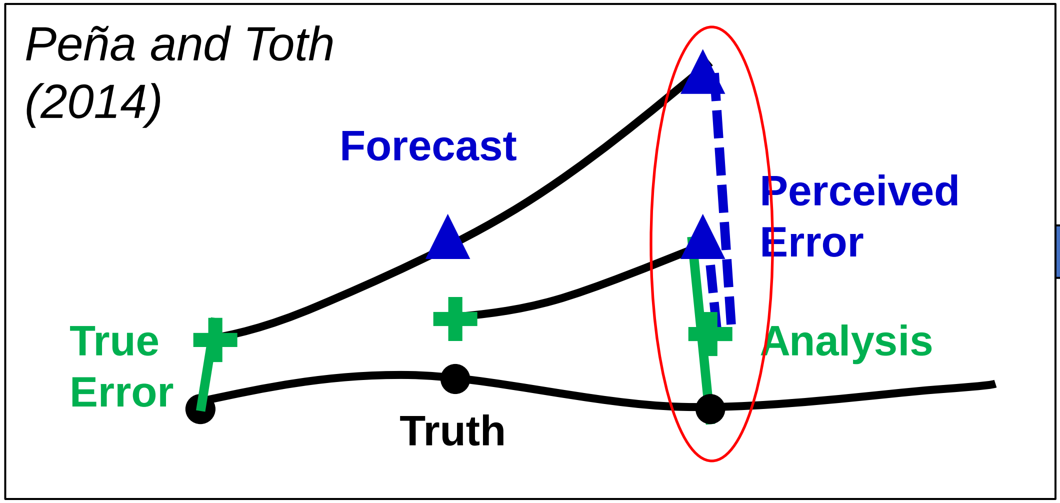


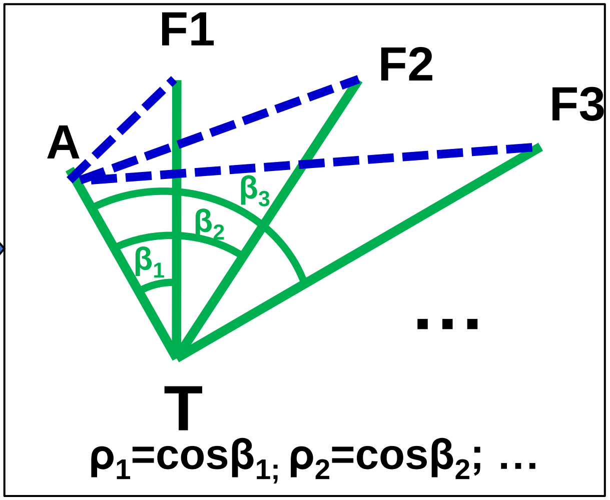
传统方法大多是经验性的，存在理论上的不合理性和偏差

Tra

- Observations as proxy
  - Sparse observations – no gridded information
  - Fraught with observational error (including representativeness error)
- DA schemes themselves
  - Computationally expensive
  - Affected by same assumptions used in DA scheme, potentially biased/inaccurate estimates
- Analysis in other DA systems as proxy
  - Independent?
  - Errors in analysis

Statistical Analysis and Forecast Error (SAFE) Estimation





Forecast State — Analysis State — Analysis Error

$$d_i^2 = (F_i - A)^2 = ((F_i - T) - (A - T))^2 \equiv (x_i - x_0)^2$$

Perceived Error — True State — Forecast Error

Measurements

$$d_i^2 = x_0^2 + x_i^2 - 2\rho_i x_0 x_i$$

Estimated quantities

Can we estimate unknown parameters with observed quantities?

Cost Function and Relevant Assumptions

$$d_i^2 = x_0^2 + x_i^2 - 2\rho_i x_0 x_i$$

Measurements **Estimated quantities**

Cost Function $$J = \max\left(\left|d_i^2 - \hat{d}_i^2\right|\right) \cdot w_i^{-1}$$

- **max : L_∞ norm**
- **Sampling standard error of the mean (SEM)**
- **Minimization: Limited-memory BFGS**

$$SEM_i = \frac{sd_i}{\sqrt{N}} \cdot f$$

$$f = \sqrt{(1+r_1)(1-r_1)}$$

$$w_i = \frac{SEM_i}{\sum_i SEM_i}$$

Peña and Toth (2014)

Connect measurements to estimates:

1.How true error grows in time; →

2.How true forecast errors get decorrelated from true analysis errors with increasing lead time.

$$\rho_i = \rho_1^{\,i}$$

Exponential

$$x_i^2 = x_0^2 e^{\alpha t_i}$$

Logistic

$$x_i^2 = \frac{S_\infty \cdot c}{e^{-\alpha t_i} + c}$$

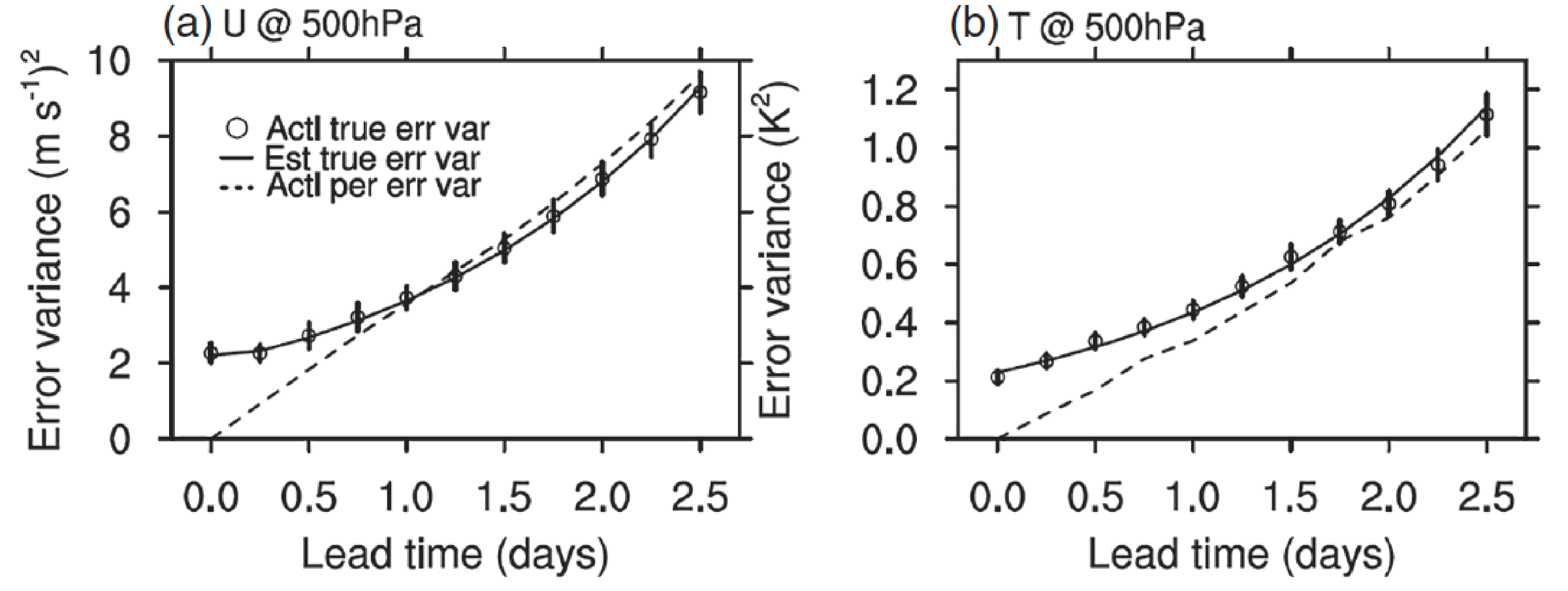
$$c = x_0^2 / (S_\infty - x_0^2)$$

α : Growth Rate

S_∞ :Saturation Value

Validate Assumption 1 using GFS OSSE data
Truth is known

Ground truth: ECMWF operational model version c31r1
Analysis & Forecast model: NCEP GFS model
Data assimilation: 3DVar
Time period of samples: 3 July to 26 August 2005, 24-hour interval



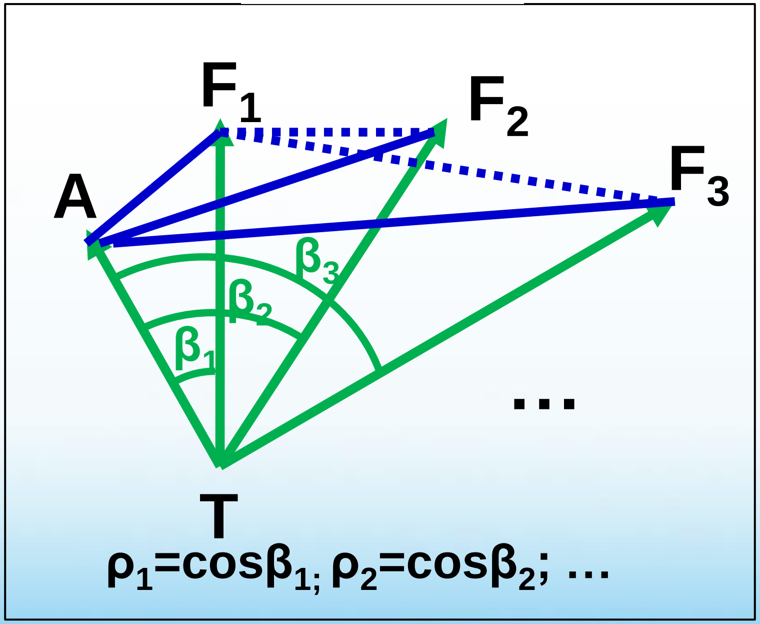
Short-range true forecast error variance almost grows exponentially!

Feng et al. (2020, QJRMS)

Analysis / Forecast Error Correlation

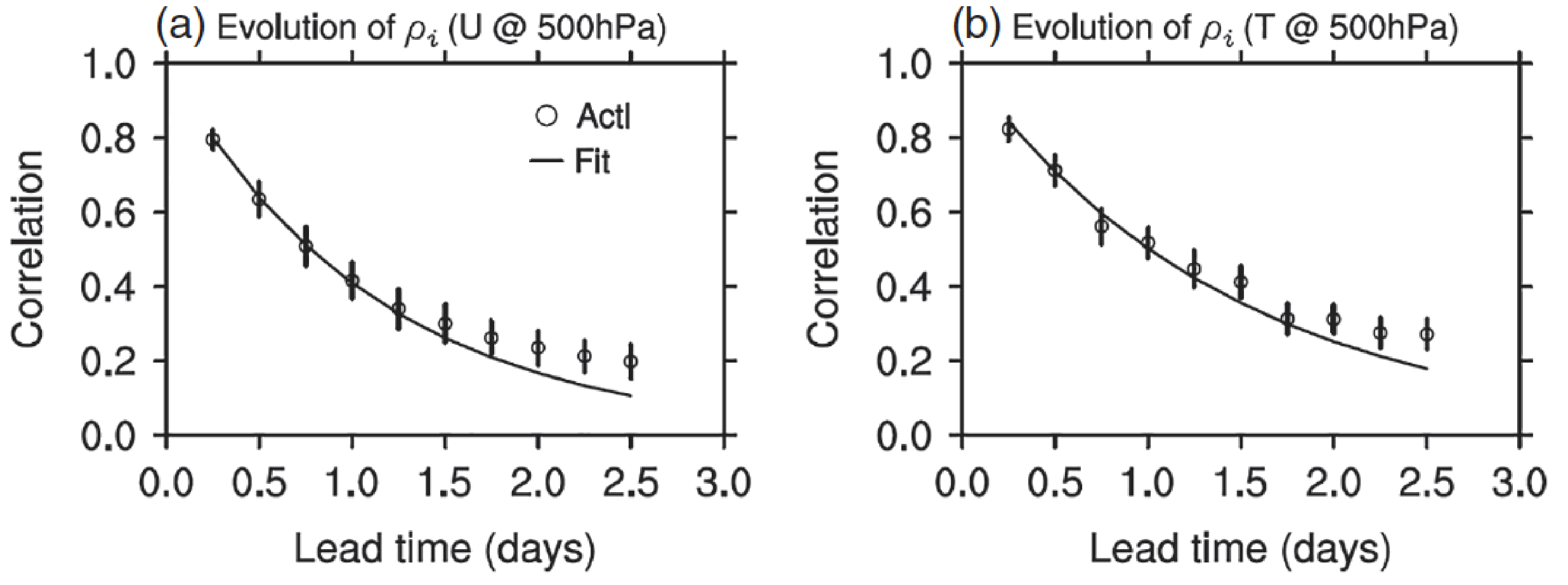
- With no DA step, analysis & forecast errors correlate at 1.0
- With one DA step, errors become de-correlated, 1 > ρ1 >0;
- The effectiveness of forecast errors in sampling analysis errors decays exponentially in time

$$\rho_i = \rho_1{}^i$$



Validate Assumption 2 using GFS OSSE data
Truth is known

Ground truth: ECMWF operational model version c31r1
Analysis & Forecast model: NCEP GFS model
Data assimilation: 3DVar
Time period of samples: 3 July to 26 August 2005, 24-hour interval



$$\rho_i = \rho_1^{\ i}$$

Feng et al. (2020, QJRMS)

SAFE方法的发展（*Feng et al. 2017,Tellus; 2020,2023 QJRMS*）

- SAFE方法的cost function是多元高阶非线性函数，最小化求解极易受采样误差的影响而不稳定，无法估计格点分析误差
- Feng et al. (2017) 在cost function中引入了额外的限制项
  - ✓ 利用滞后预报偏差的指数误差增长对真实误差增长参数(α)进行限制（假定模式误差很小），极大减小采样误差的影响

$$
\begin{aligned}
\mathrm{J} = & \max\left(\left|f_i^{2\prime} - f_i^2\right| \cdot w_i\right) \ (i = 0, 1, \ldots, n) \\
& + \max\left(\left|f_{i,j}^{2\prime} - f_{i,j}^2\right| \cdot w_{i,j}\right) \ (i = 0, 1, \ldots, m),
\end{aligned}
$$

SAFE方法的发展（*Feng et al. 2017,Tellus; 2020,2023 QJRMS*）

- SAFE方法的cost function是多元高阶非线性函数，最小化求解极易受采样误差的影响而不稳定，无法估计格点分析误差

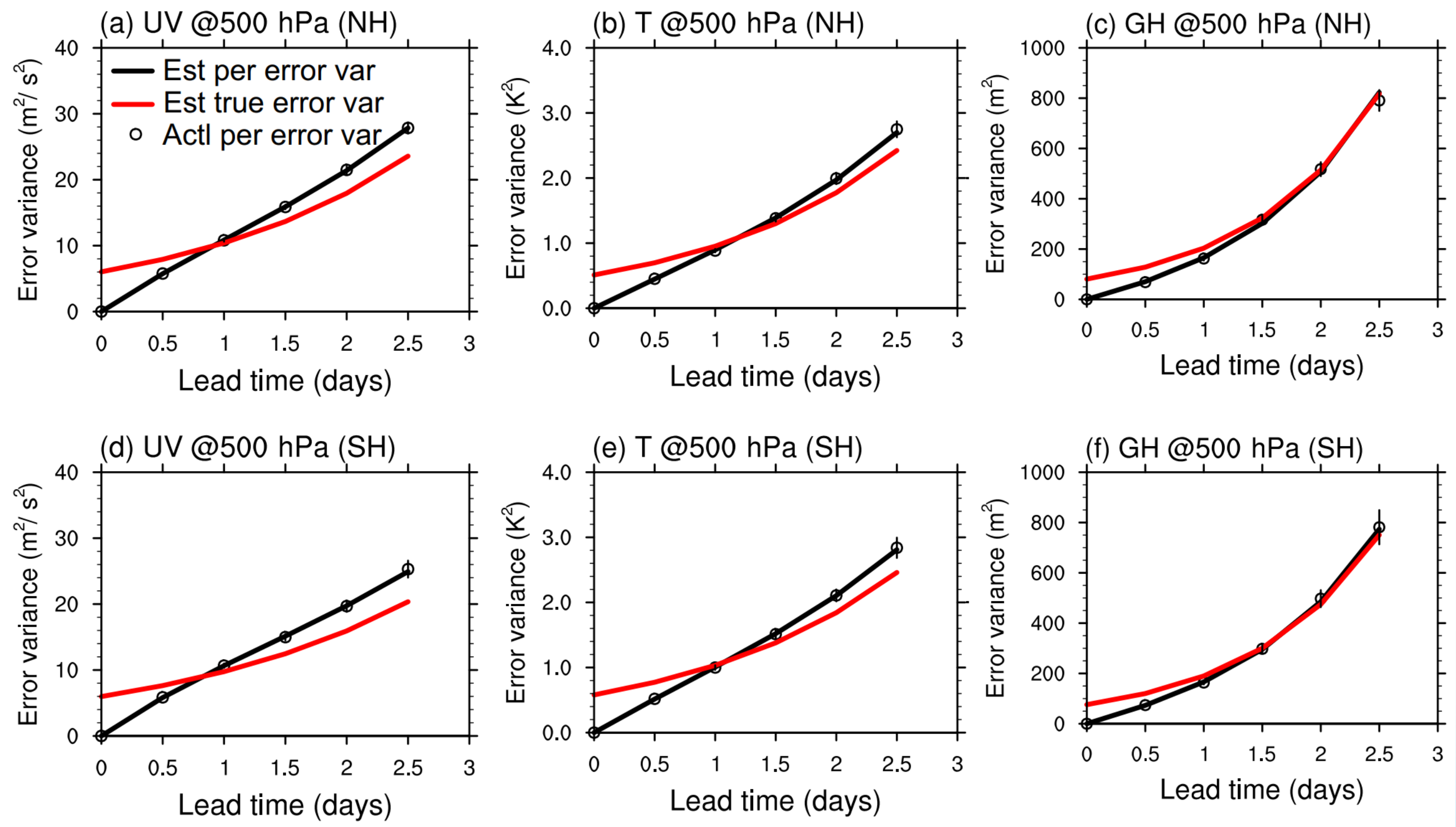
- Feng et al. (2017) 在cost function中引入了额外的限制项

  ✓ 利用滞后预报偏差的指数误差增长对真实误差增长参数(α)进行限制（假定模式误差很小），极大减小采样误差的影响

$$J = \max\left(\left|f_i^{2\prime} - f_i^2\right| \cdot w_i\right)\ (i = 0, 1, \ldots, n)$$
$$+ \max\left(\left|f_{i,j}^{2\prime} - f_{i,j}^2\right| \cdot w_{i,j}\right)\ (i = 0, 1, \ldots, m),$$

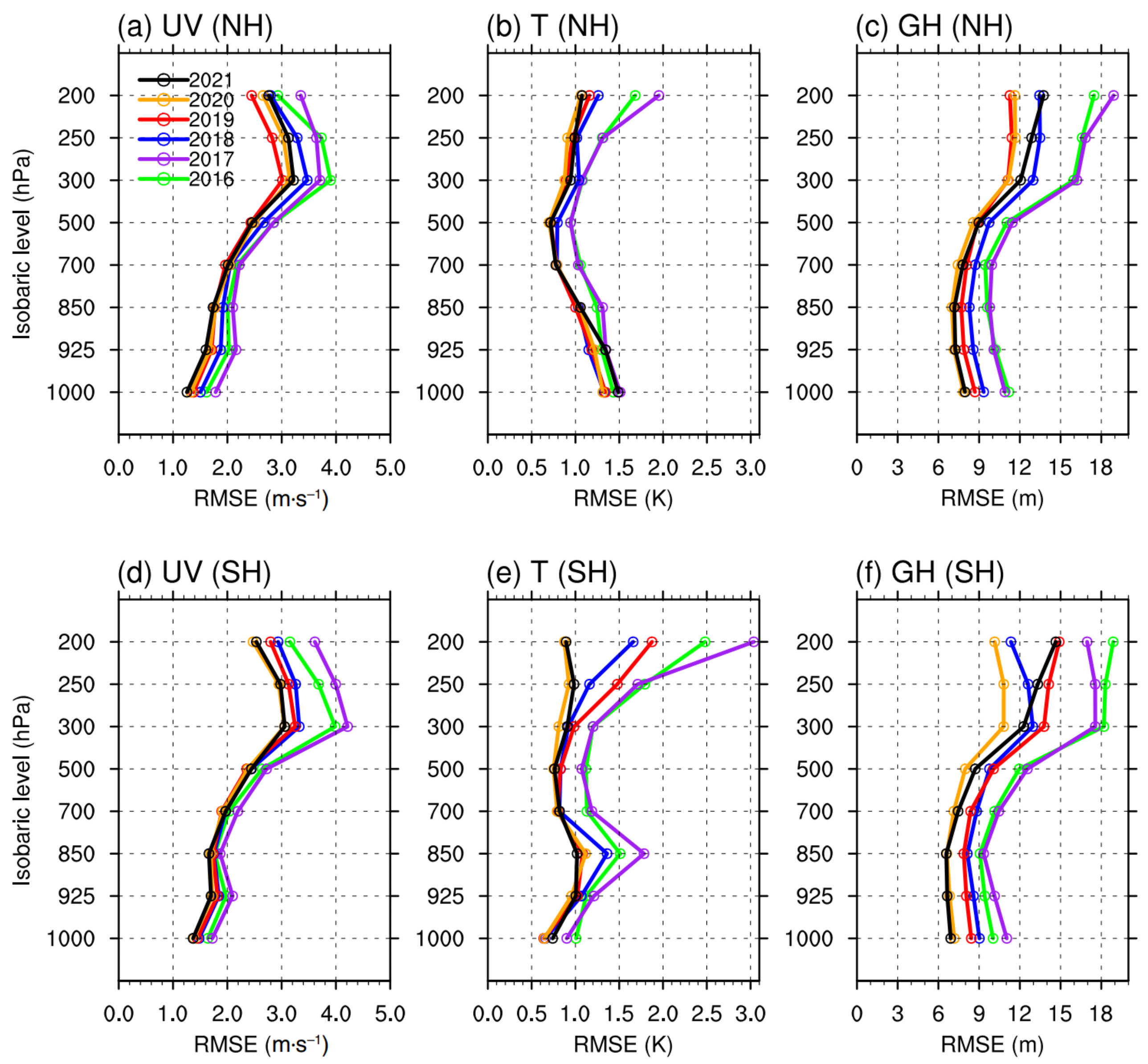
Short-range Evolution of Forecast Error Variance @Dec2021-Feb2022

GRAPES_GFS real-time DA/Prediction system
NH: N30-70 degs; SH: S30-70 degs;

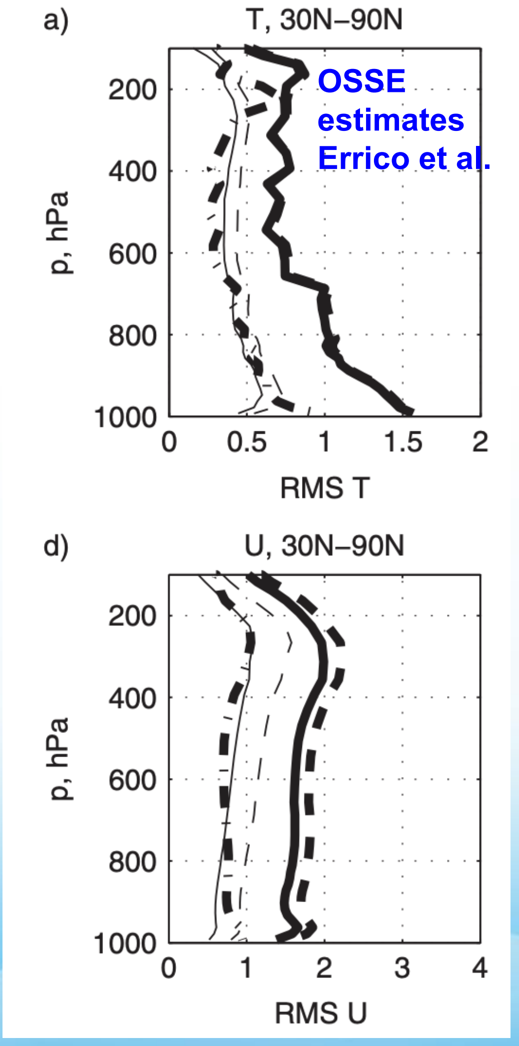


- **Good fitting of perceived error variance**
- **Analysis error variance unnegligible at early lead times**

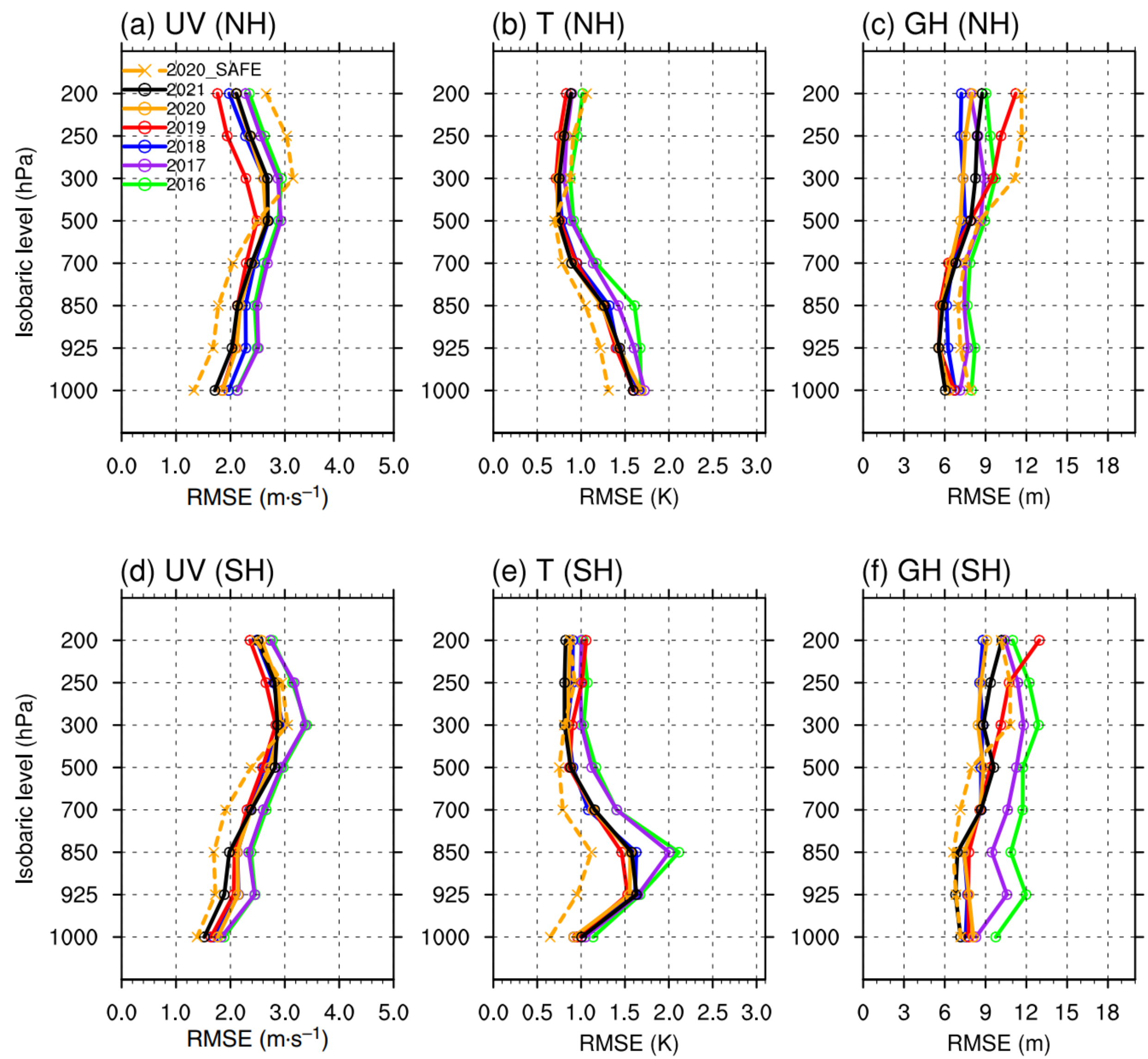
Annual Variation of Analysis Error - SAFE



- 分析误差廓线结构较为合理
- 2016,17和之后分析误差明显减小

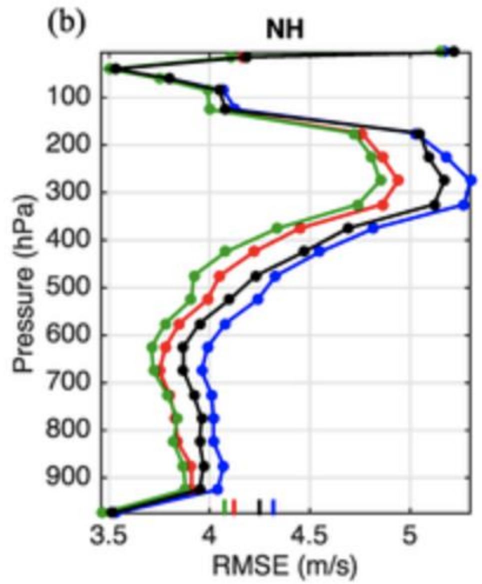


Annual Variation of Analysis Error - vs. ERA Reanalysis

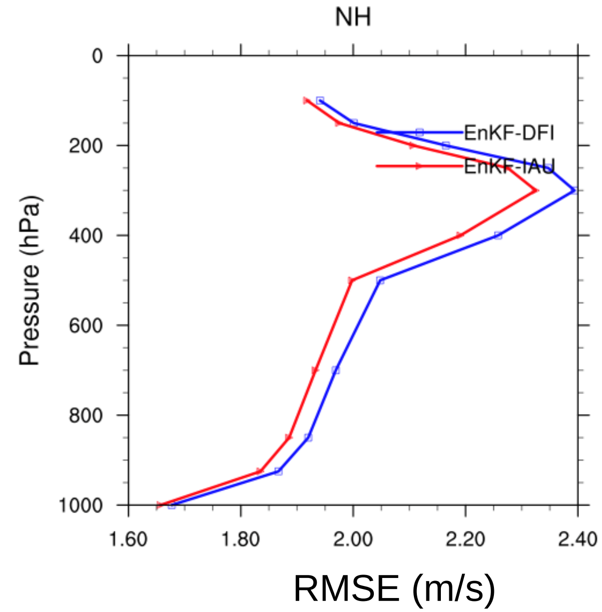


- ERA检验存在一定不合理
  - ➢ UV,GH的300和500hPa分析误差非常接近
- ERA检验相比SAFE
  - 500hPa以上低估
  - 500hPa以下高估

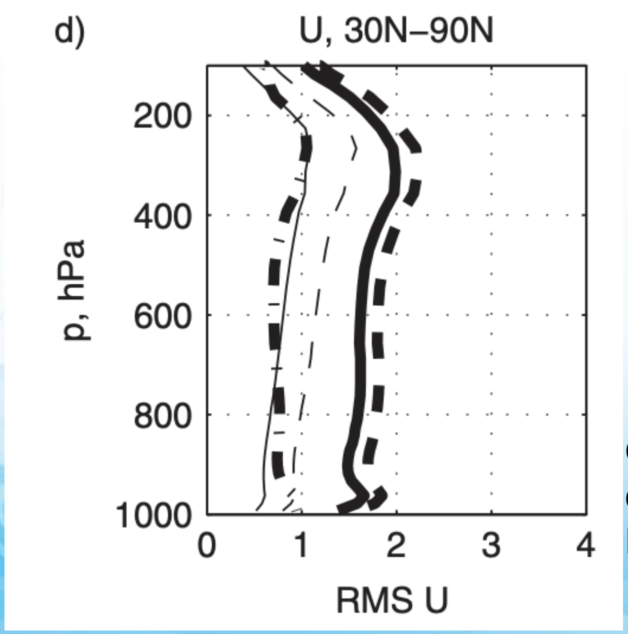
代表性误差？



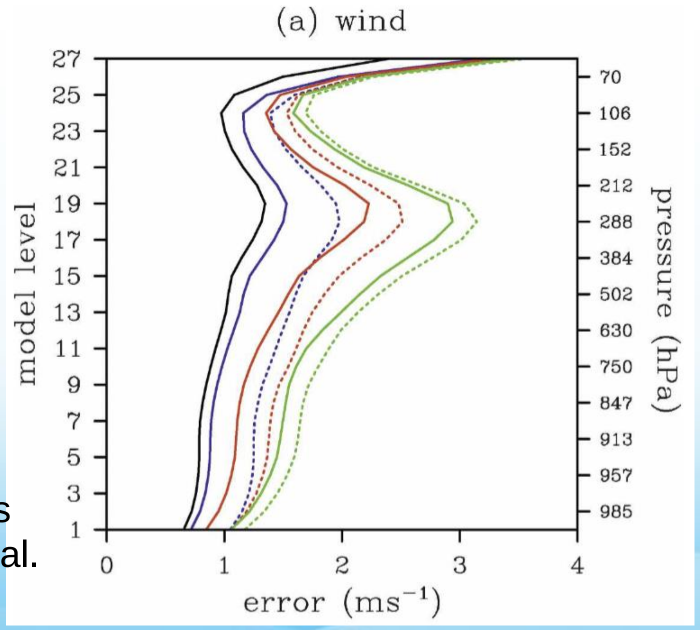
Lei et al. 2016



Lei et al.



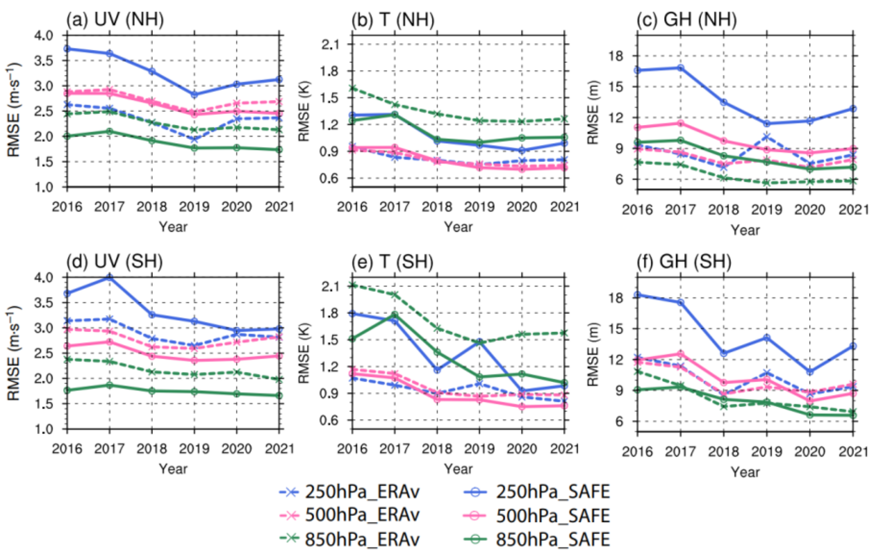
OSSE estimates Errico et al.



Wang et al. 2008

Annual Variation of Analysis Error - SAFE

分析误差随时间演变



- ERAv和SAFE反映的趋势相似，但存在系统性偏差

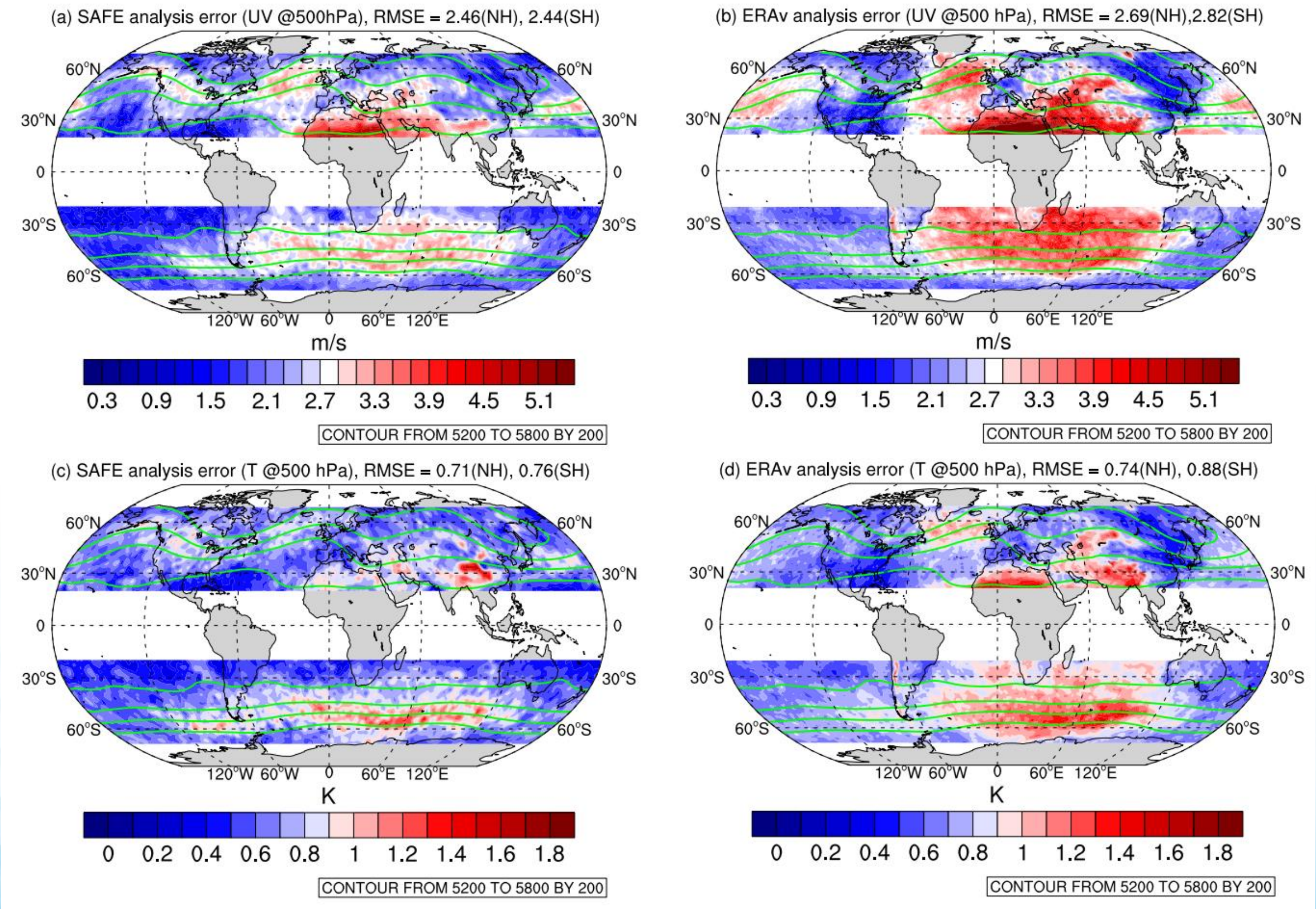
- 2017vs2016变化相对较小
  - ➢ 平均4%
- 2018vs2017显著减小
  - ➢ 平均17%，尤其在250hPa，~22%
  - ➢ UV (10%),T(23%),GH(19%)
  - ➢ **3DVAR to 4DVAR**
- 2019vs2018有一定减小
  - ➢ 尤其对NH~9%
  - ➢ 模式物理改进CU；更多卫星观测同化，e.g., FY-4A GIIRS
- 2020vs2019有一定减小
  - ➢ 尤其对T, GH in SH, ~18%
  - ➢ 更多卫星观测同化，e.g., FY-3D

分析误差相对前一年的相对变化

			2017	2018	2019	2020	2021
UV	250 hPa	NH	-2%	-10%	-14%	7%	3%
UV	500 hPa	NH	0%	-7%	-9%	3%	-2%
UV	850 hPa	NH	5%	-9%	-8%	0%	-2%
UV (6年12.5%)	250 hPa	SH	9%	-18%	-4%	-6%	1%
UV	500 hPa	SH	3%	-10%	-3%	1%	3%
UV	850 hPa	SH	6%	-6%	-1%	-2%	-2%
T	250 hPa	NH	1%	-23%	-5%	-6%	9%
T	500 hPa	NH	0%	-16%	-10%	-3%	2%
T	850 hPa	NH	5%	-21%	-3%	5%	1%
T (6年29%)	250 hPa	SH	-5%	-32%	27%	-37%	6%
T	500 hPa	SH	-4%	-23%	0%	-9%	1%
T	850 hPa	SH	18%	-24%	-20%	3%	-9%
GH	250 hPa	NH	1%	-20%	-15%	2%	10%
GH	500 hPa	NH	4%	-15%	-9%	-3%	5%
GH	850 hPa	NH	2%	-15%	-7%	-9%	3%
GH (6年24.5%)	250 hPa	SH	-4%	-28%	12%	-23%	23%
GH	500 hPa	SH	5%	-22%	3%	-21%	9%
GH	850 hPa	SH	3%	-13%	-3%	-16%	-1%

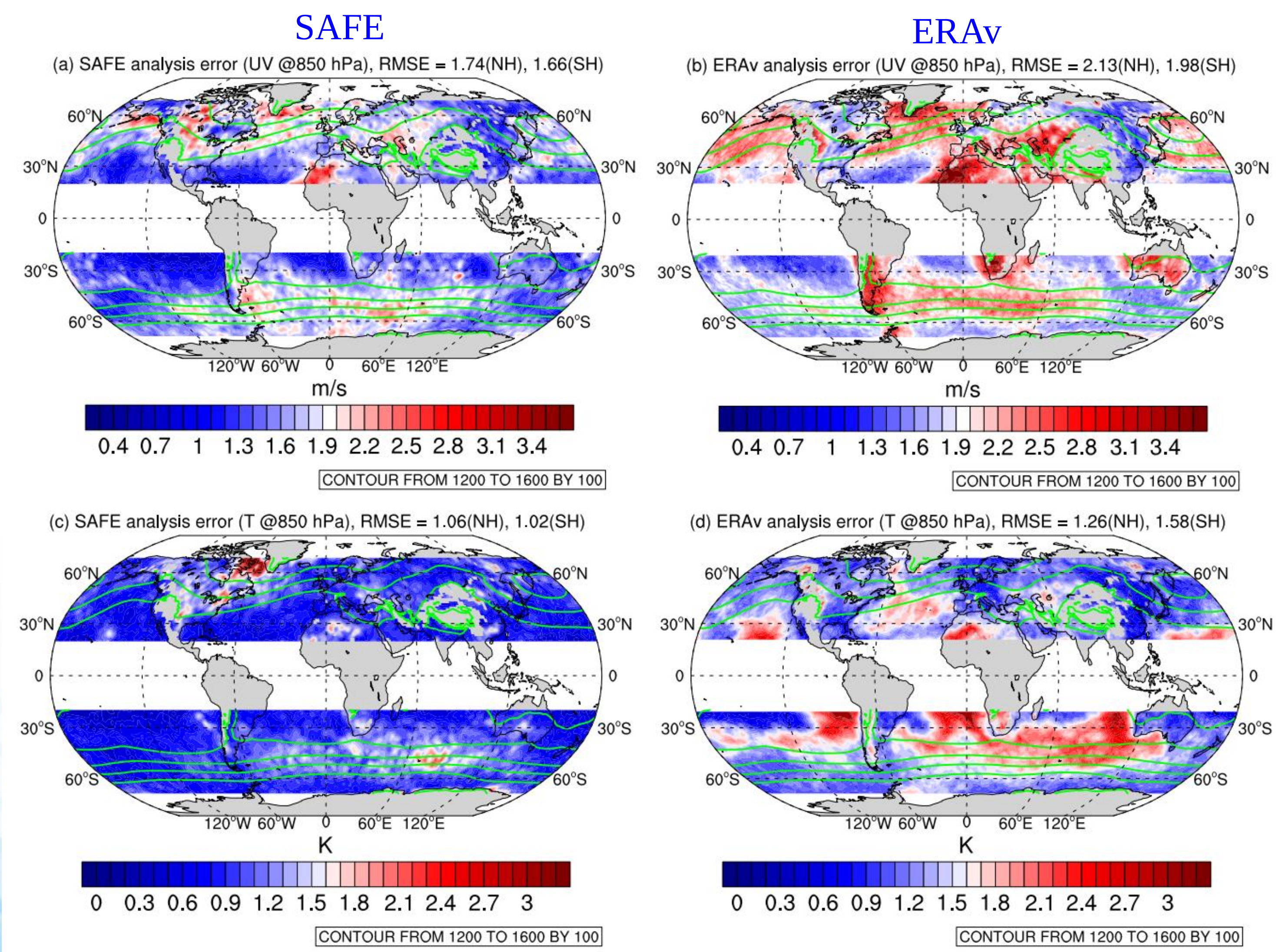
Spatial Distribution of Analysis Error@500 hPa (2021)

SAFE ERAv



SAFE和ERAv分析误差空间结构类似，但后者整体高估

Spatial Distribution of Analysis Error@850 hPa (2021)



SAFE和ERAv分析误差空间结构相似性比500hPa下降，
后者整体高估

Summary

- SAFE方法可以提供GRAPES系统同化分析场误差的可信估计
- GRAPES区域平均分析误差2016-2021改进明显

  - 2017-2018, UV (10%),T(23%),GH(19%)

  同化方案3DVar升级至4DVar

  - 2018-2020, NH~9%, T, GH in SH, ~18%

  模式物理改进，加入多种卫星观测同化
- GRAPES格点分析以及24-hr预报误差估计较合理

  - ERAv存在明显bias

Discussion

- 和其它中心分析场对比的检验方法存在问题
- 不同的误差估计方法

  - SAFE (1. 有理论基础 ; 2. 假定合理无偏)

  - Other methods (1. 经验性估计, 2. 存在偏差)
- 资料同化、集合预报中的应用，大气海洋模式中的应用......

Thanks for listening!

fengjiefj@fudan.edu.cn

- **Feng, J.***, Z. Toth, and M. Peña, **2017**: Spatial Extended Estimates of Analysis and Short-Range Forecast Error Variances. *Tellus A*, 69:1, 1325301.
- **Feng, J.***, Z. Toth, M. Pena, and J. Zhang, **2020**: Partition of Analysis and Forecast Error Variance into Growing and Decaying Components. *Quart. J. Roy. Meteor. Soc.,* 146(728), 1302-1321.
- **Feng, J**, et al. **2022**: Spatiotemporal estimation of analysis errors in the operational global data assimilation system at the China Meteorological Administration using a novel statistical method, *QJRMS*, DOI: 10.1002/qj.4507.
- Pena, M. and Toth, Z. 2014. Estimation of analysis and forecast error variances. *Tellus* 66A, 21767.