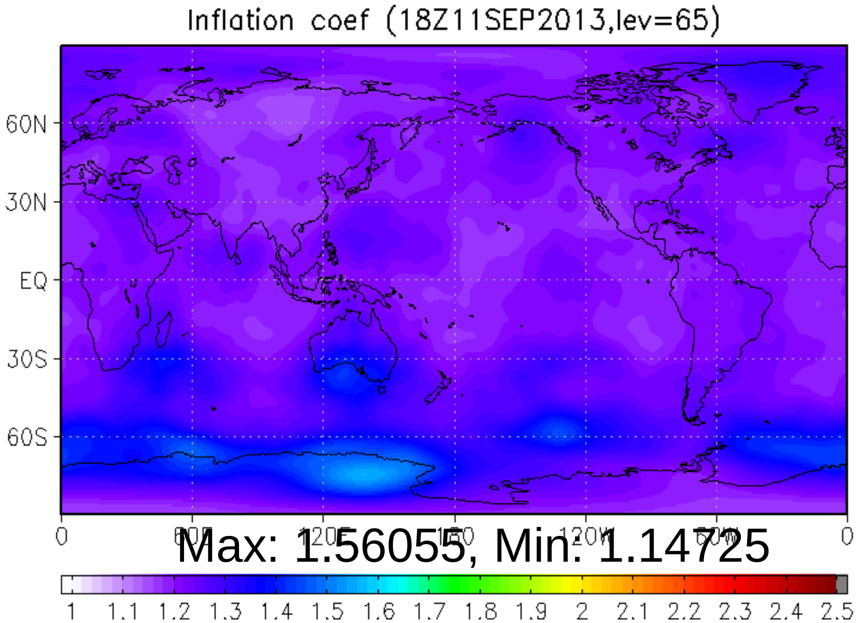
The 7th EnKF data assimilation workshop

Adaptive inflation scheme using the relaxation to prior spread method

Yoichiro Ota (Japan Meteorological Agency)

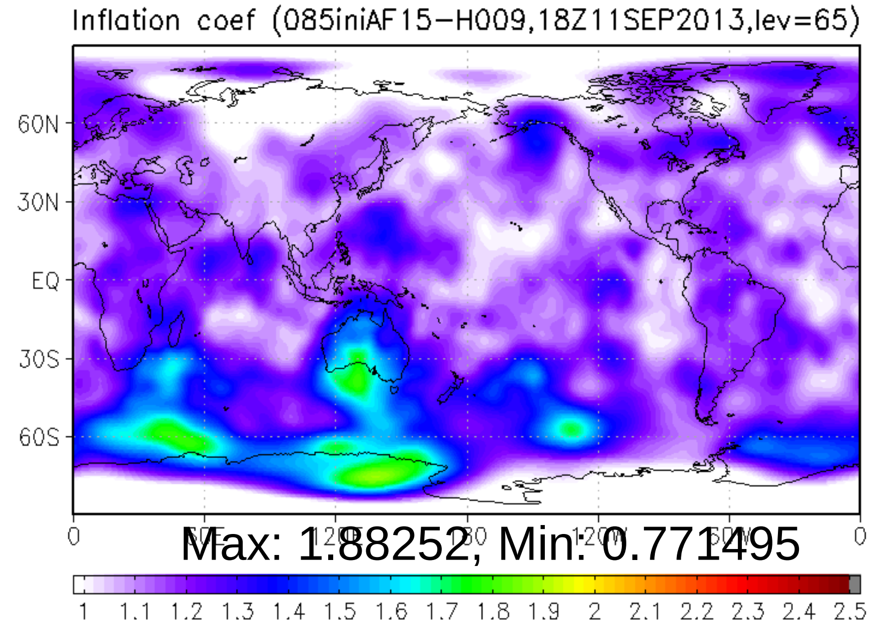
Special thanks: Shunji Kotsuki and Takemasa Miyoshi (RIKEN/AICS)

Motivations



Max: 1.56055, Min: 1.14725

Inflation without AIRS and IASI



Max: 1.88252, Min: 0.771495

Inflation with AIRS and IASI

Estimated inflation coefficients in the hybrid 4DVAR-LETKF on lower stratosphere.

- Current adaptive inflation scheme does not work well with observations that are assimilated with wrong R (observation error covariance)
  - In this case, AIRS and IASI are assimilated with larger than optimal R

Seek more robust methods to estimate adaptive covariance inflation.

Current methods

(1) Adaptive inflation in JMA

Based on $\left\langle d_{a-b} d_{o-b}^{\mathbf{T}} \right\rangle = \mathbf{HBH}^{\mathbf{T}}$

da-b is approximated with do-b, HδXf, and transformation matrix on grid i.

$$\alpha_i = \frac{tr\left[\left\langle d_{a-b} d_{o-b}^{\mathbf{T}} \right\rangle \left(\tilde{\rho}_i^{-1} \circ R_i\right)^{-1}\right]}{tr\left[\tilde{\rho}_i \circ \left(\frac{H_i \delta X_f \left(H_i \delta X_f\right)^{\mathbf{T}}}{m-1}\right)\left(\tilde{\rho}_i^{-1} \circ R_i\right)^{-1}\right]}$$

ρ is localization function (inflation to R)

R is used as the weight to each observation

(2) Formulation using R (Miyoshi 2011)

Based on $\left\langle d_{o-b} d_{o-b}^{\mathbf{T}} \right\rangle = \mathbf{HBH}^{\mathbf{T}} + \mathbf{R}$

$$\alpha_i = \frac{tr\left[\left\langle d_{o-b} d_{o-b}{}^{\mathbf{T}} \right\rangle \left(\tilde{\rho}_i^{-1} \circ R_i\right)^{-1}\right] - tr\left(\tilde{\rho}_i\right)}{tr\left[\left(\frac{H_i \delta X_f \left(H_i \delta X_f\right)^{\mathbf{T}}}{m-1}\right)\left(\tilde{\rho}_i^{-1} \circ R_i\right)^{-1}\right]}$$

R is directly used (more sensitive to the setting of R)

Inflation coefficients are estimated on each analysis grid.
Inflation is applied to the forecast ensembles.

Adaptive RTPS (Based on Ying and Zhang 2015)

Relaxation to prior spread (RTPS):

$$\mathbf{X}_{i,inf}^{a} = \left(\frac{\alpha\sigma_i^b + (1-\alpha)\sigma_i^a}{\sigma_i^a} \right) \mathbf{X}_i^a$$

σ^b: forecast spread
σ^a: analysis spread

Based on $\left\langle d_{a-b} d_{o-a}^{\mathbf{T}} \right\rangle = \mathbf{HAH}^{\mathbf{T}}$

$$\beta = \sqrt{\frac{(d_{a-b})^{\mathbf{T}} (d_{o-a}) \circ \mathbf{R}^{-1}}{tr[\mathbf{HAH}^{\mathbf{T}} \circ \mathbf{R}^{-1}]}}$$

weights to each obs.

Mean inflation λ on observation space

$$\lambda = \frac{1}{p} \sum_{i}^{p} \frac{\alpha\left(\mathbf{HBH}^{\mathbf{T}}\right)_i^{1/2} + (1-\alpha)\left(\mathbf{HAH}^{\mathbf{T}}\right)_i^{1/2}}{\left(\mathbf{HAH}^{\mathbf{T}}\right)_i^{1/2}}$$

p: number of obs.

$$\alpha = \frac{\beta - 1}{\frac{1}{p} \sum_{i=1}^{p} \frac{\left(\mathbf{HBH}^{\mathbf{T}}\right)_i^{1/2} - \left(\mathbf{HAH}^{\mathbf{T}}\right)_i^{1/2}}{\left(\mathbf{HAH}^{\mathbf{T}}\right)_i^{1/2}}}$$

Blue: modifications to original Ying and Zhang (2015)

Inflation pattern is determined with RTPS method.
Global coefficient α is estimated from all assimilated observations.
Inflation is applied to the analysis ensembles.

For real applications, time smoothing is applied to the estimation of α.

Experiment with SPEEDY model

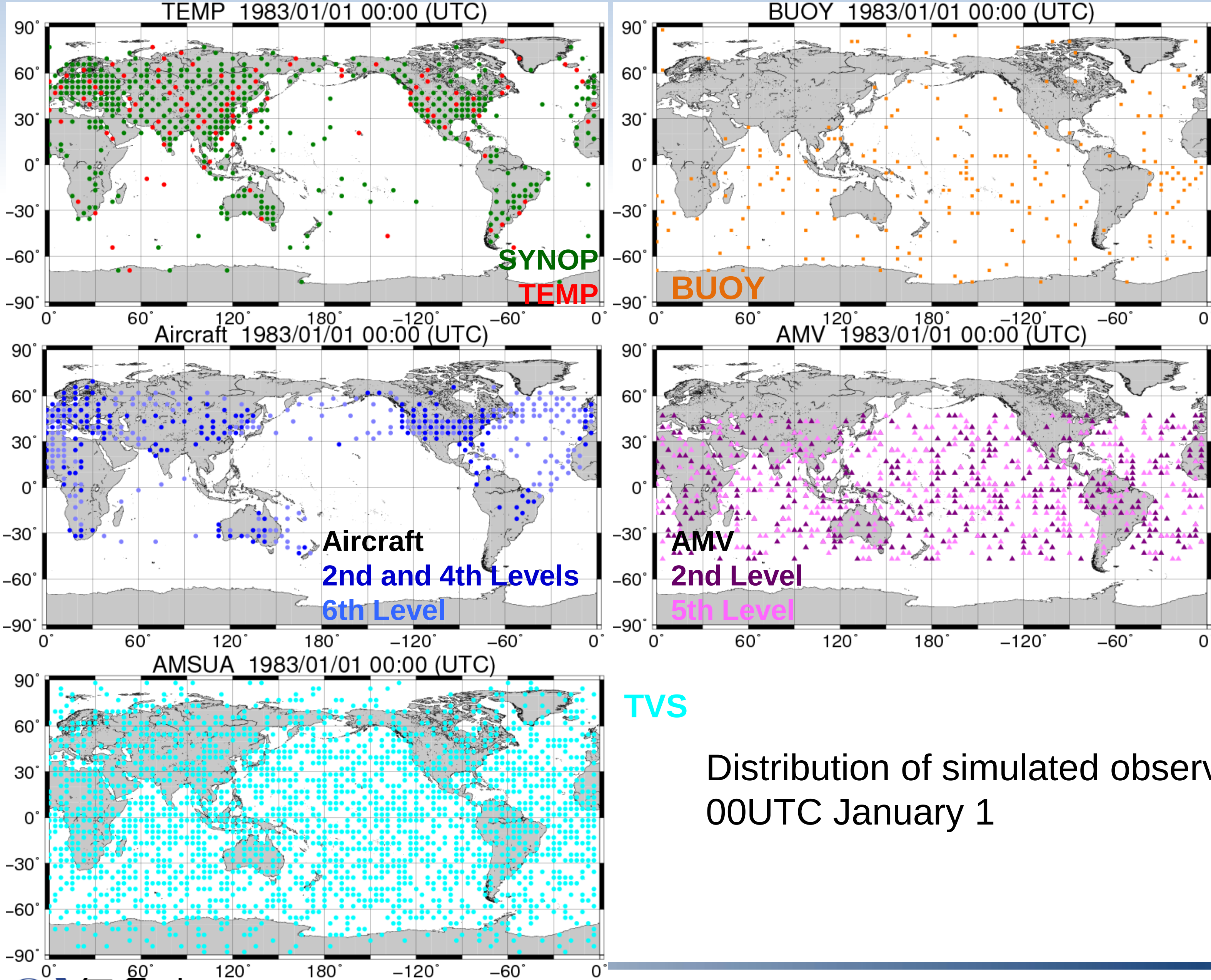
- Forecast model: SPEEDY model (T30L7)
- Identical twin experiments
  - Perfect model setting, correct R is known, well tuned localization scale
- Observations are assimilated every 6 hours (setting is shown in next slide)
- 32 members, localization scale 800km in horizontal, 0.4 scale height in vertical
  - Optimal RTPS: manually tuned (with 0.05 intervals) coefficient (α=0.50)
  - Adaptive RTPS
    - α starts from 0.30, 0.50 and 0.70
    - temporal smoothing parameter: 0.005 and 0.025
- O-A and HAH are derived from LETKF update
- Period: 1 year (later half year is used for verification)

Observation settings

Type	Elements	Numbers	Observed level	Errors
SYNOP	Ps	550	Surface	1hPa
TEMP	U,V,T,RH	104(00,12UTC)	All levels	1m/s,1K,5%
BUOY	Ps	200 (Sea)	Surface	1.5hPa
AIRCRAFT	U,V,T	180 (lower) 220 (upper)	2nd, 4th levels 6th level	2m/s,1.5K
AMV	U,V	400 (Low latitudes)	2nd and 5th levels	3m/s
TVS	Vertically accumulated T	2000	2nd, 4th, and 6th levels	0.5K

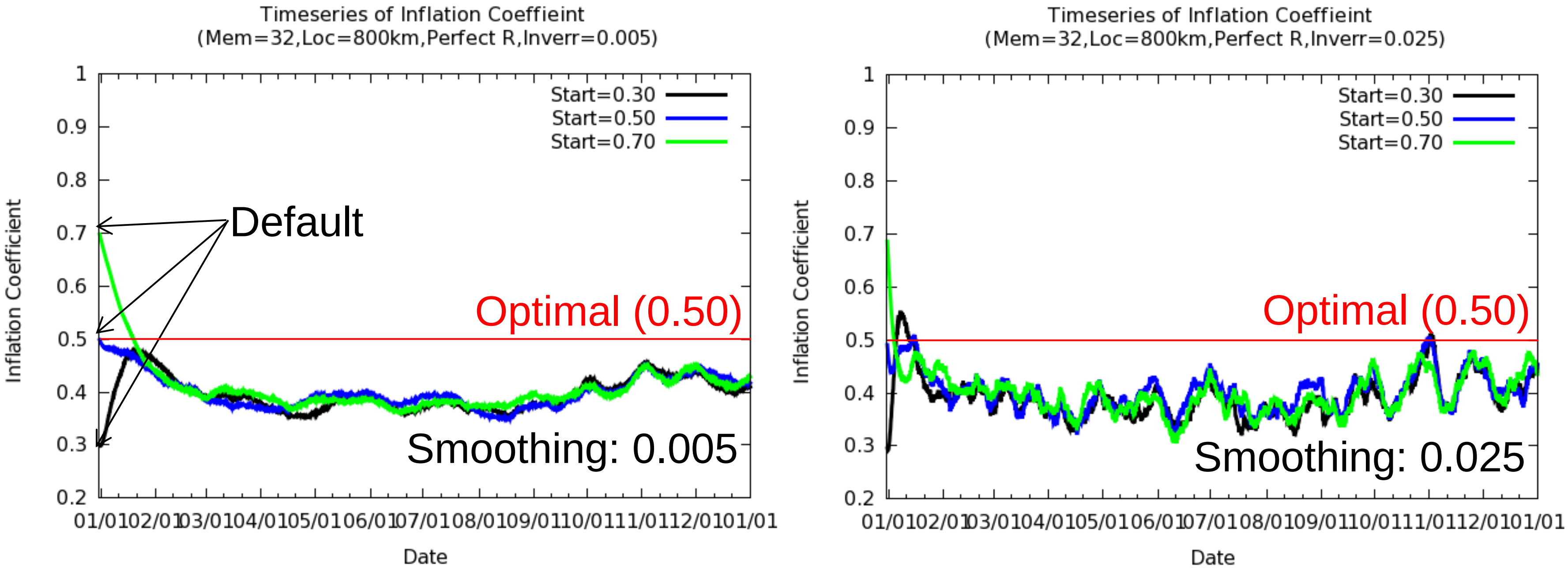
TVS on kth model level is defined as $\frac{1}{4}\left(T_{k-1} + 2T_k + T_{k+1}\right)$

Observation distribution



Distribution of simulated observations on 00UTC January 1

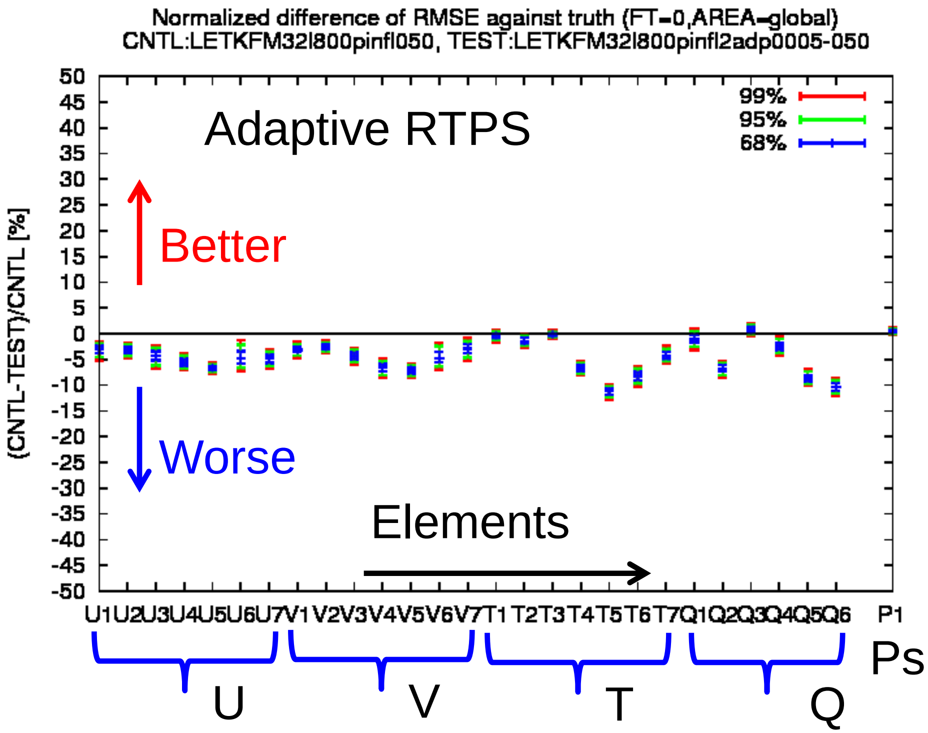
Estimated coefficients of RTPS

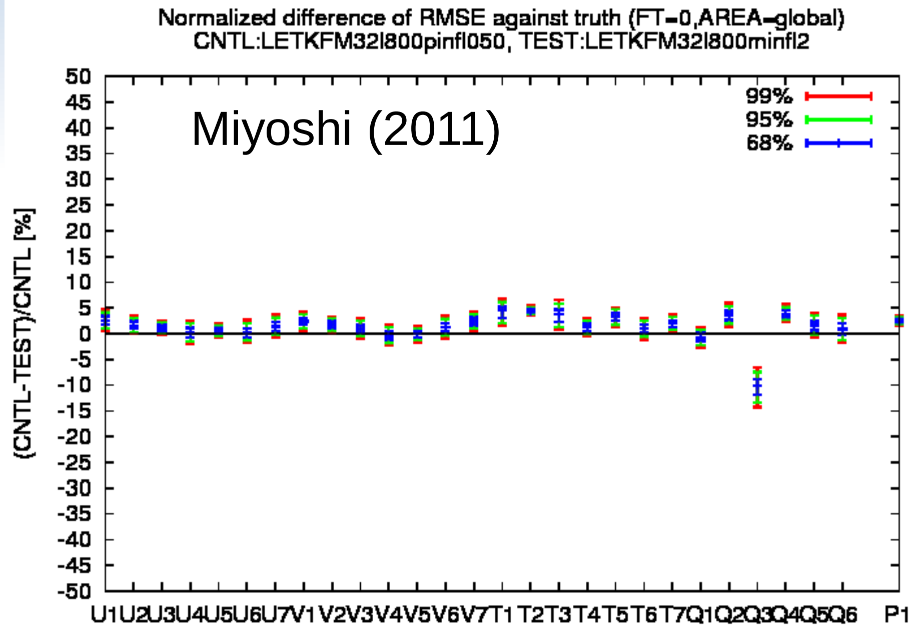


- Estimated coefficients are converged to the similar value for all experiments.
- It is slightly smaller than the optimal value (0.50 in this case).
- 0.005 is good enough for smoothing parameter.

Only the results with default value 0.5 and time smoothing 0.005 are shown in the following.

Analysis accuracy





Normalized difference of analysis RMSE:
CNTL is Optimal RTPS experiment

$$\frac{RMSE(CNTL) - RMSE(TEST)}{RMSE(CNTL)} \times 100\%$$

- Analysis RMSE of Adaptive RTPS is slightly worse than Optimal RTPS.
- Adaptive multiplicative inflation of Miyoshi (2011) works well in the ideal settings.
- Current adaptive inflation used in JMA is not good as Miyoshi (2011) (not shown).

Possible reasons of underestimation

1. Problems on non-linear (RH in this case) or non-local (TVS in this case) observation operators: Not likely (as shown in next slide)

$\mathbf{HAH^T}$ can be derived by 2 ways:

- Directly compute HXa from analysis ensembles
- Update HXf using the LETKF ← Use this for the experiment

Two methods are equivalent when observation operator is linear and local.
It is not trivial when observation operator is non-linear or non-local.

2. Model error term Q is not considered on B

In the adaptive RTPS, inflation coefficient is estimated based on

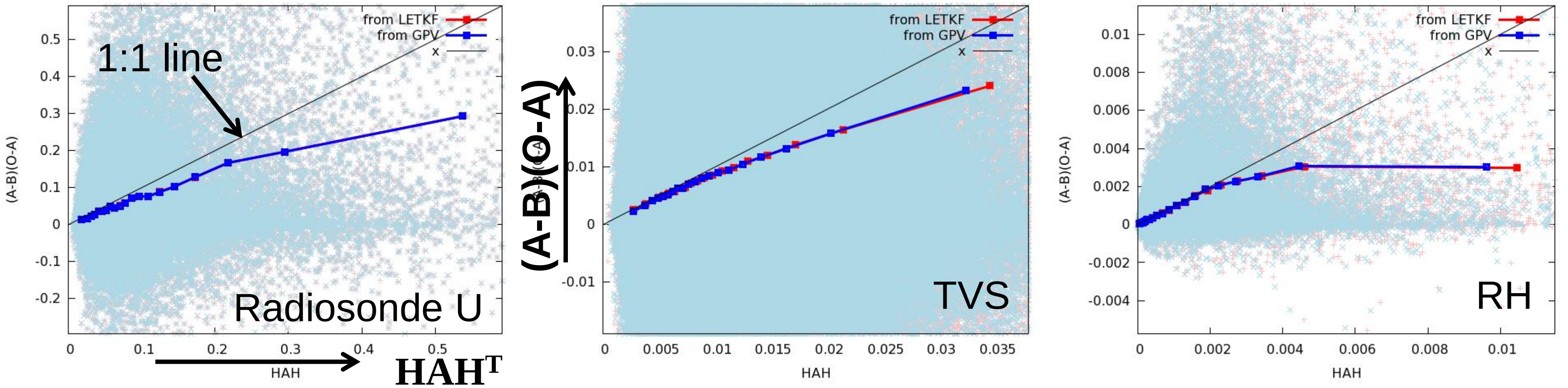
$$\left\langle d_{a-b} d_{o-a}^{\mathbf{T}} \right\rangle = \mathbf{HAH^T}$$

Using nonlinear model, model error term Q can not be neglected even if the model is perfect.

$$\mathbf{B} = \mathbf{MAM^T} + \mathbf{Q}$$

B may not satisfy optimal relation even when A is optimal.

Effect of non-linear or non-local H



Scatter plot and binned average of $\mathbf{HAH^T}$ (x-axis) and (A-B)(O-A) (y-axis) over the last 1 month of the Optimal RTPS experiment.

Blue: $\mathbf{HAH^T}$ is computed directly from analysis ensemble, Red: updated with LETKF

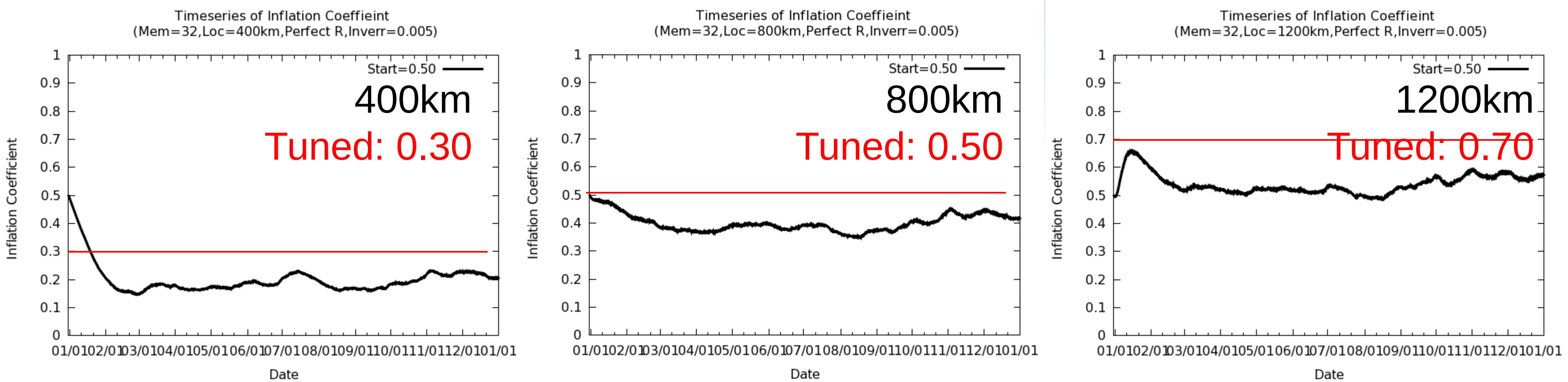
- $\mathbf{HAH^T}$ tends to be larger than (A-B)(O-A) especially for large $\mathbf{HAH^T}$ even for observations with linear and local H.
- For linear and local H (left), two methods are equivalent.
- For non-local (middle) or non-linear (right) H, $\mathbf{HAH^T}$ updated with LETKF is slightly larger than that is directly computed from analysis ensembles.

Effect of non-local or non-linear observation operators seems not be a critical reason of underestimation.

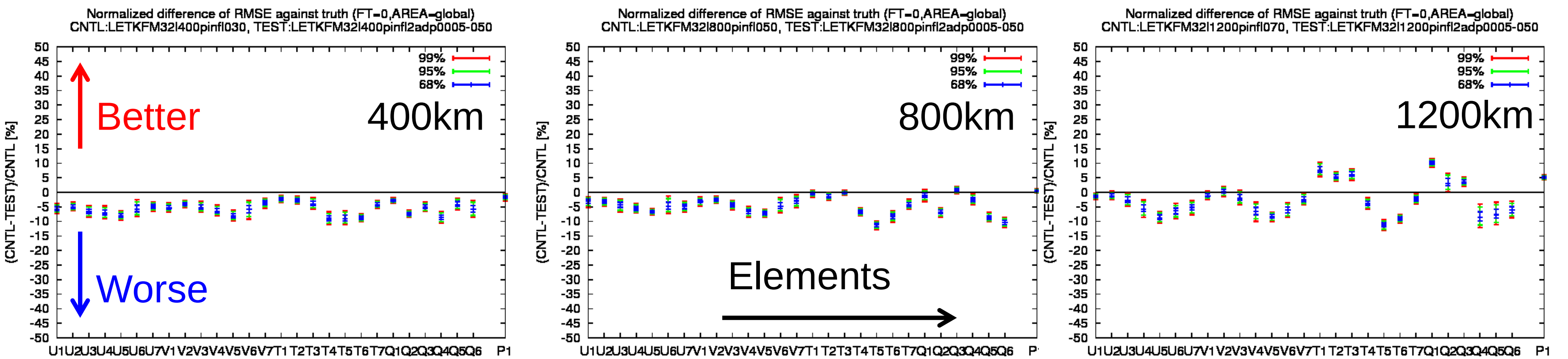
Adaptive RTPS with imperfect DA settings

- Suboptimal localization length scale
  - Smaller than optimal: 400 km in horizontal
  - Larger than optimal: 1200 km in horizontal
- Larger observation error for TVS
  - 2 times and 4 times of actual observation error standard deviation
- Imperfect model
  - Some parameters are changed from the model that produced the truth run

(1) Suboptimal localization scale



Estimated inflation coefficients (black lines) are slightly smaller than manually tuned coefficients (red lines) for all experiments.



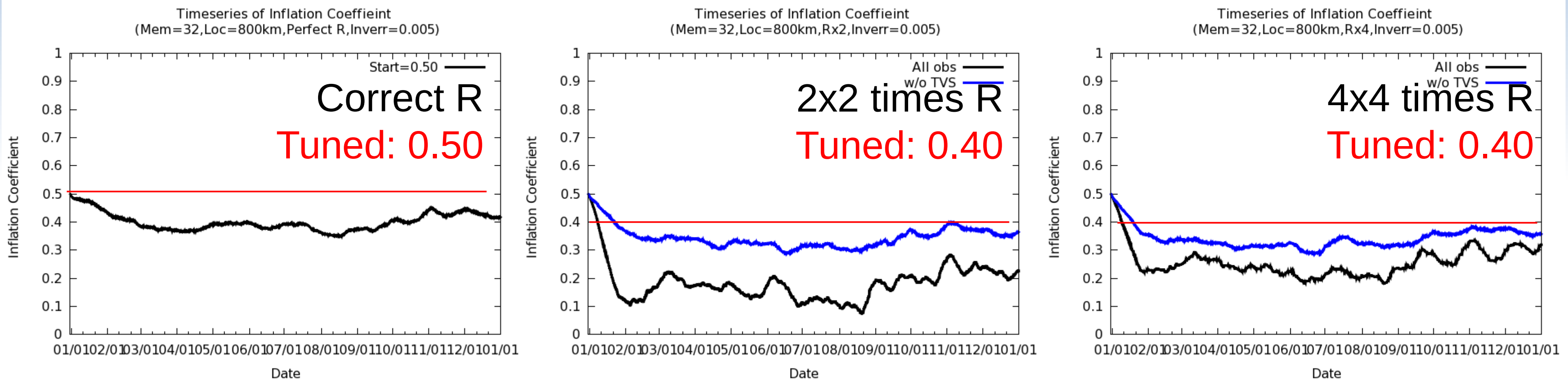
Normalized analysis RMSE difference from that of optimal setting show slight degradations

Observation settings

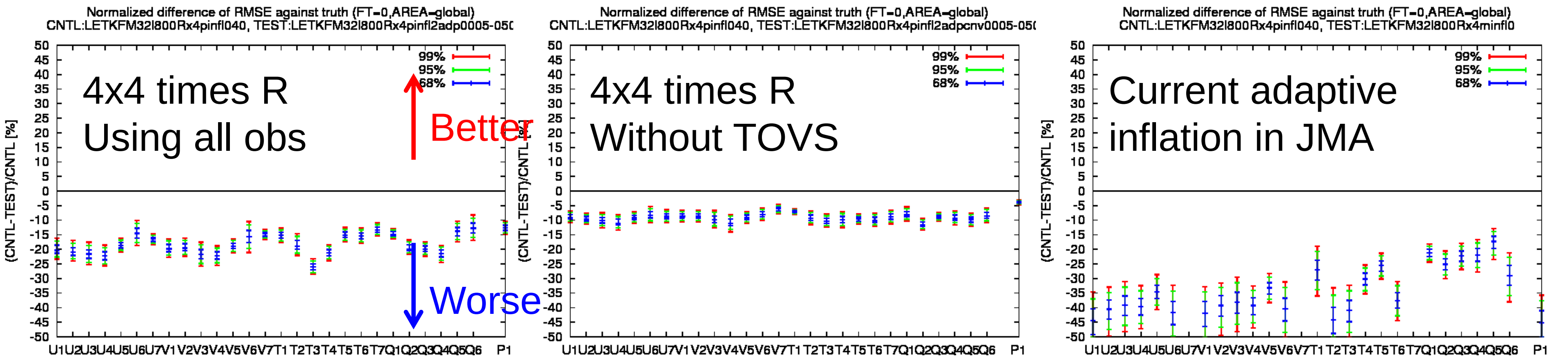
Type	Elements	Numbers	Observed level	Errors
SYNOP	Ps	550	Surface	1hPa
TEMP	U,V,T,RH	104(00,12UTC)	All levels	1m/s,1K,5%
BUOY	Ps	200 (Sea)	Surface	1.5hPa
AIRCRAFT	U,V,T	180 (lower) 220 (upper)	2nd, 4th levels 6th level	2m/s,1.5K
AMV	U,V	400 (Low latitudes)	2nd and 5th levels	3m/s
TVS	Vertically accumulated T	2000	2nd, 4th, and 6th levels	**0.5K**

Make observations with this setting, but assimilate them with 2 times or 4 times of actual observation error standard deviations.

(2) Observation error larger than actual

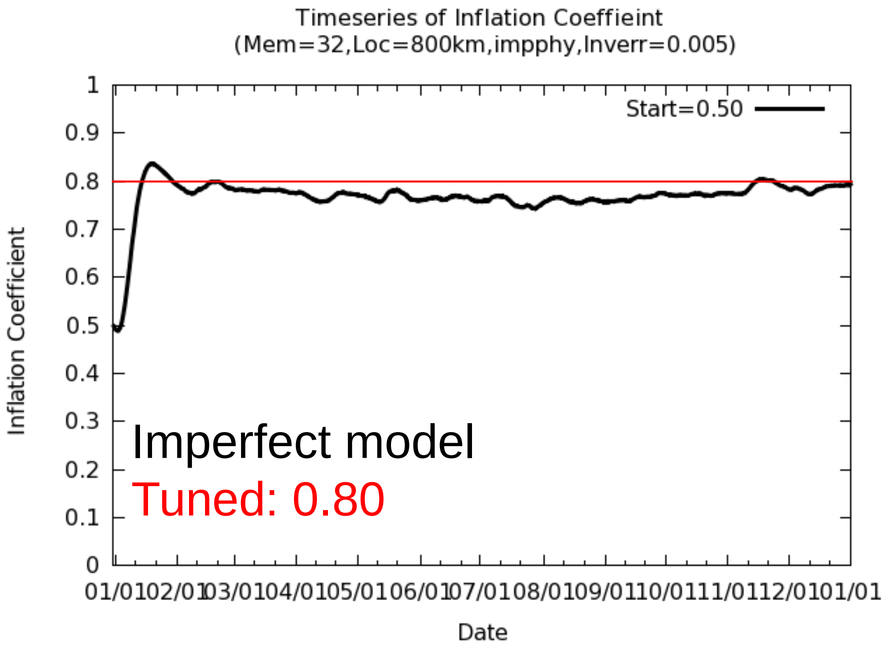


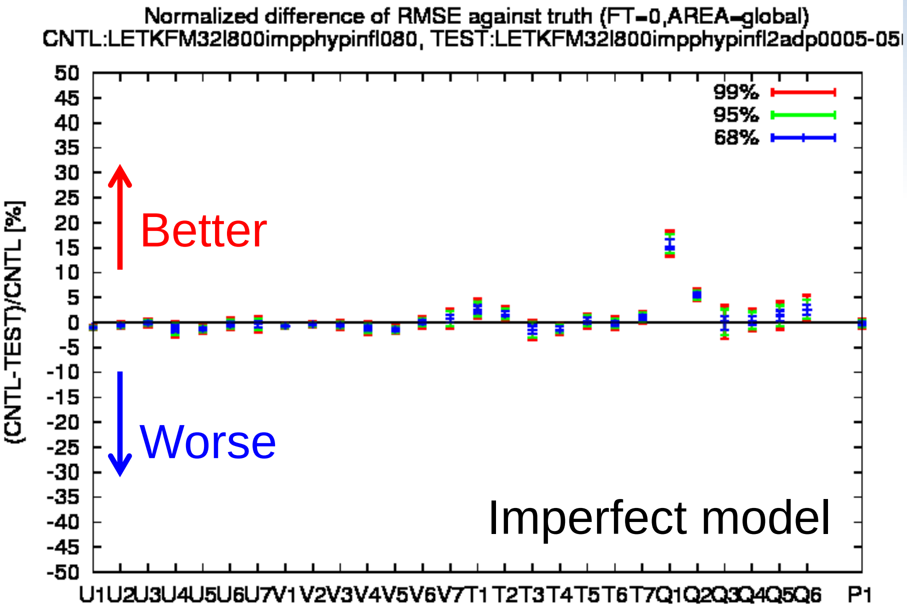
Significant underestimations using R larger than actual errors.
Better estimation if removing observations with wrong R from estimation (blue lines).



- Adaptive RTPS is better than current adaptive multiplicative inflation (note that Miyoshi 2011 does not work for this setting)
- Better analysis if removing observations with wrong R from estimation.

(3) Imperfect model





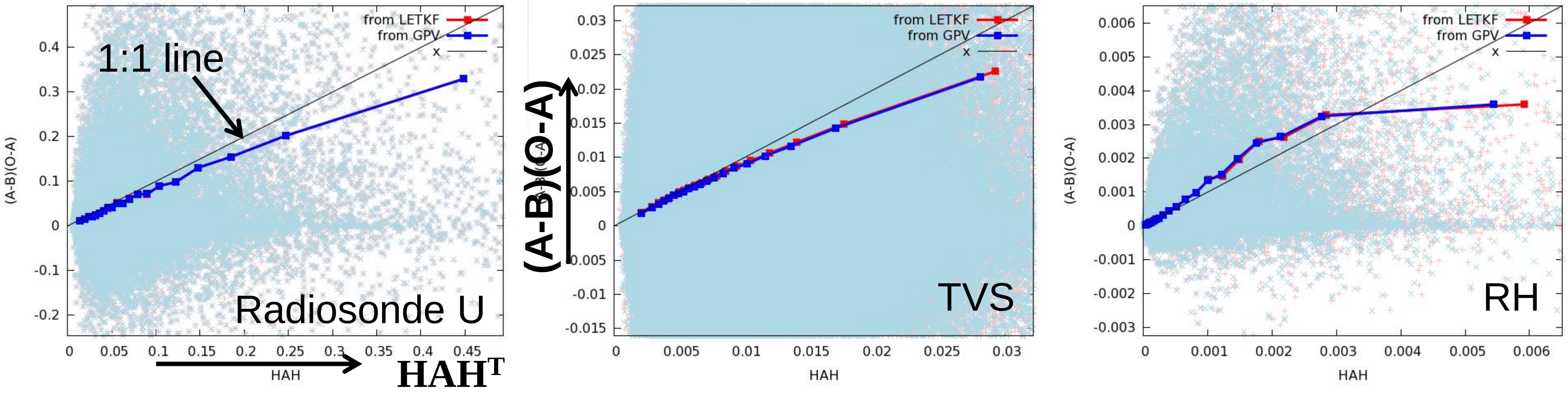
- Estimated coefficient is nearly optimal for imperfect model setting.
- Current adaptive multiplicative inflation does not work well with this setting (too large spread for Q: not shown)

Summary and conclusions

- Adaptive RTPS (Ying and Zhang 2015) is applied to the SPEEDY-LETKF.
  - Estimated coefficient is slightly smaller than optimal, suggesting optimal A will not produce optimal B
  - Adaptive RTPS is more robust in R settings and imperfect models than current adaptive multiplicative inflation
  - Excluding observations with improper R from the estimation would be beneficial.
- Future plan
  - Apply adaptive RTPS to hybrid 4DVAR-LETKF using actual observations and operational model.

Backup

Scatter plots for Adaptive RTPS

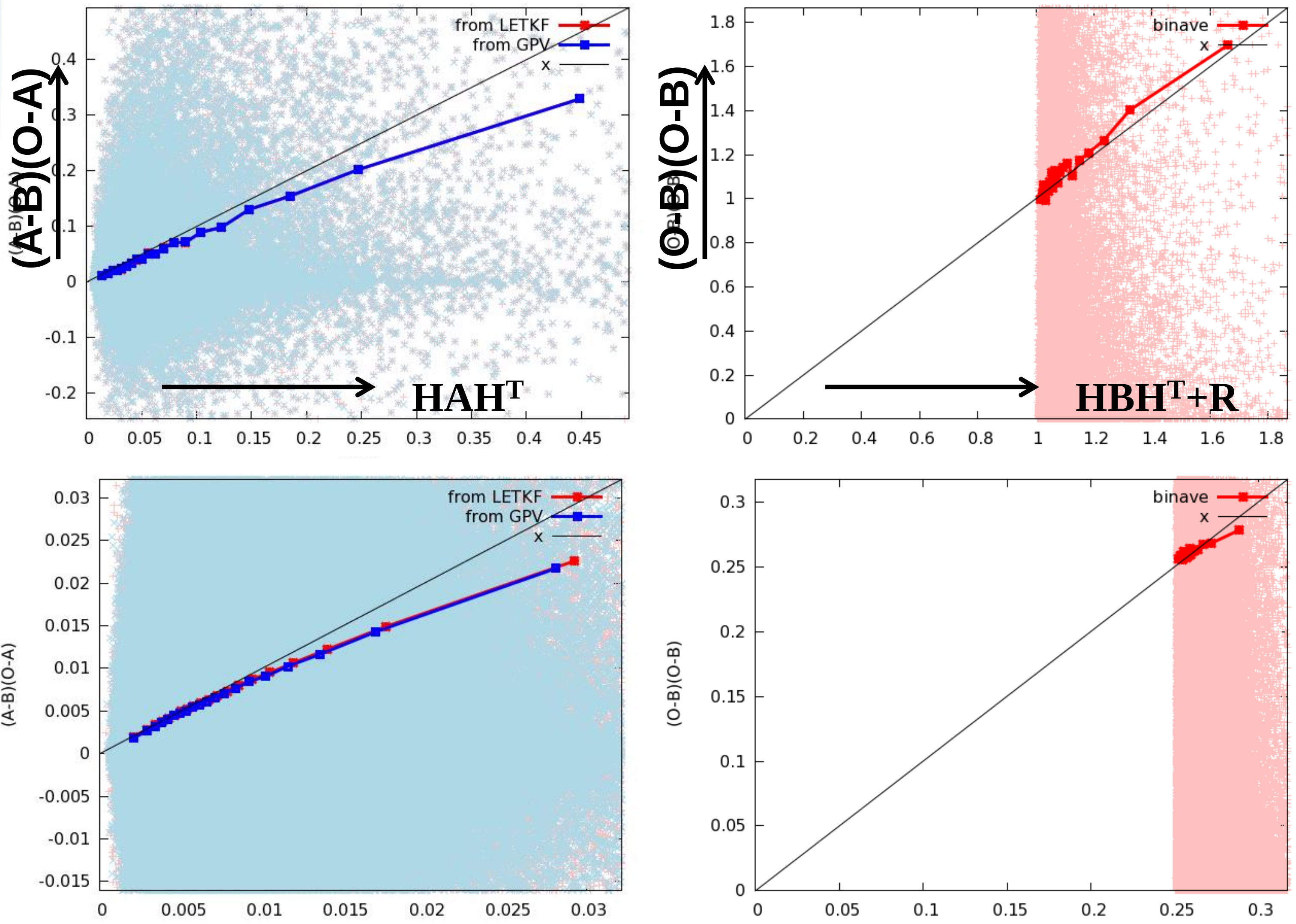


Scatter plot and binned average of $\mathbf{HAH^T}$ (x-axis) and (A-B)(O-A) (y-axis) over the last 1 month of the Adaptive RTPS experiment.

Blue: $\mathbf{HAH^T}$ is computed directly from analysis ensemble, Red: updated with LETKF

- For Radiosonde U and TVS, $\mathbf{HAH^T}$ are slightly larger than (A-B)(O-A). For RH, it is opposite.

Analysis spread and forecast spread



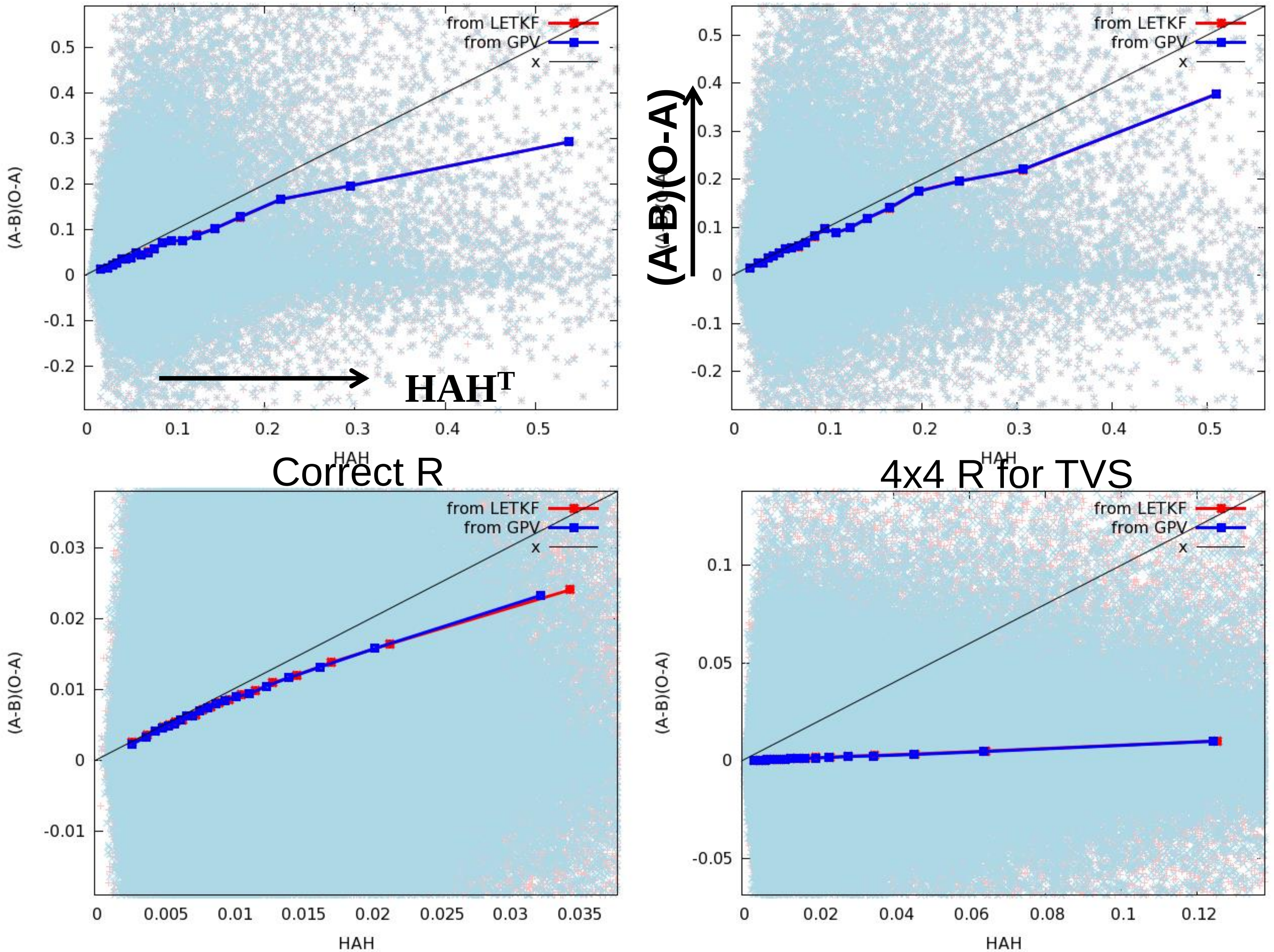
Radiosonde U

Adaptive RTPS experiment

TVS

HAH is overdispersive but HBH is not necessarily overdispersive and underdispersive for some observations.

Scatter plots for imperfect observation error

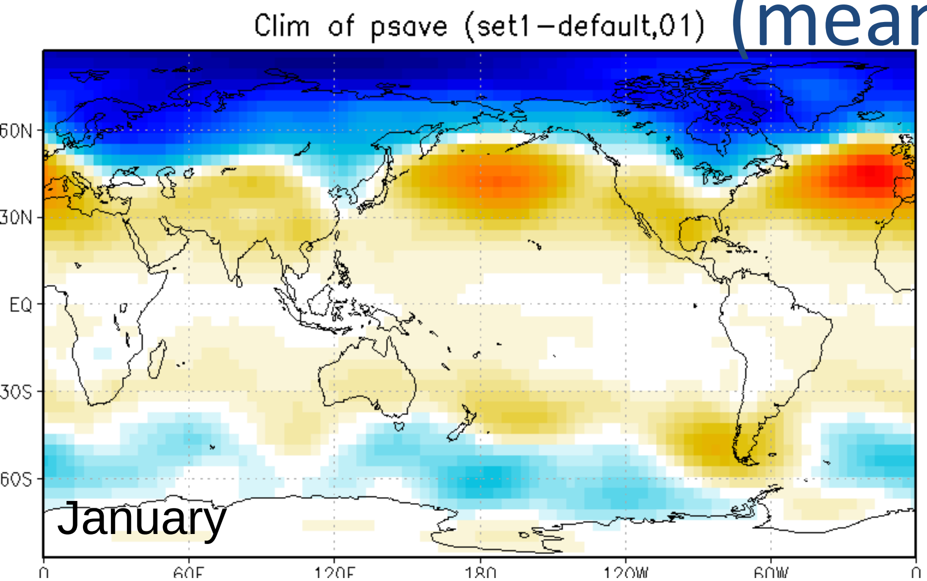


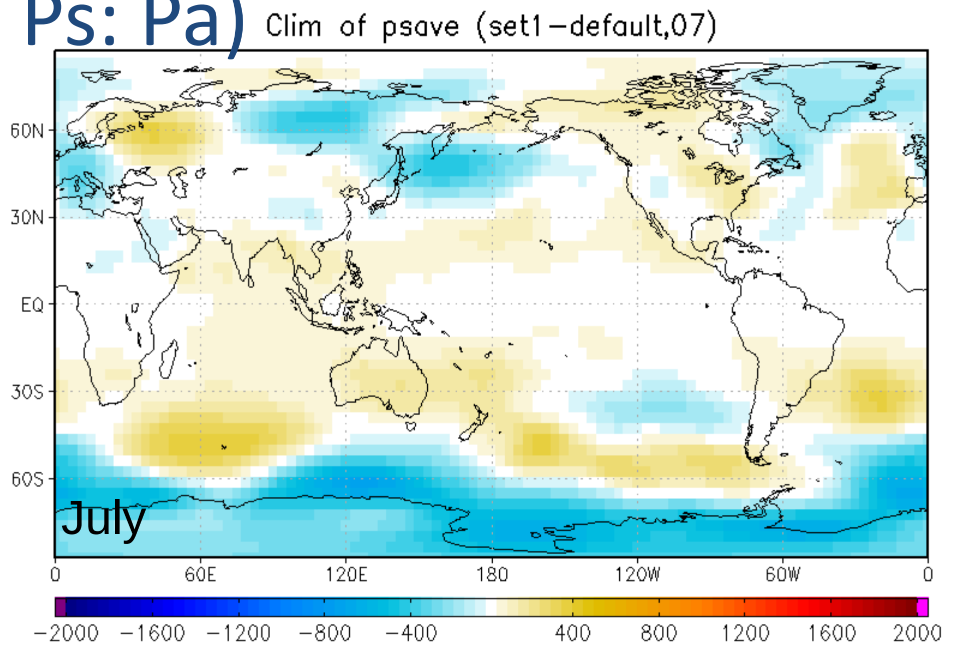
Estimation with observations with inflated R leads to underestimation.

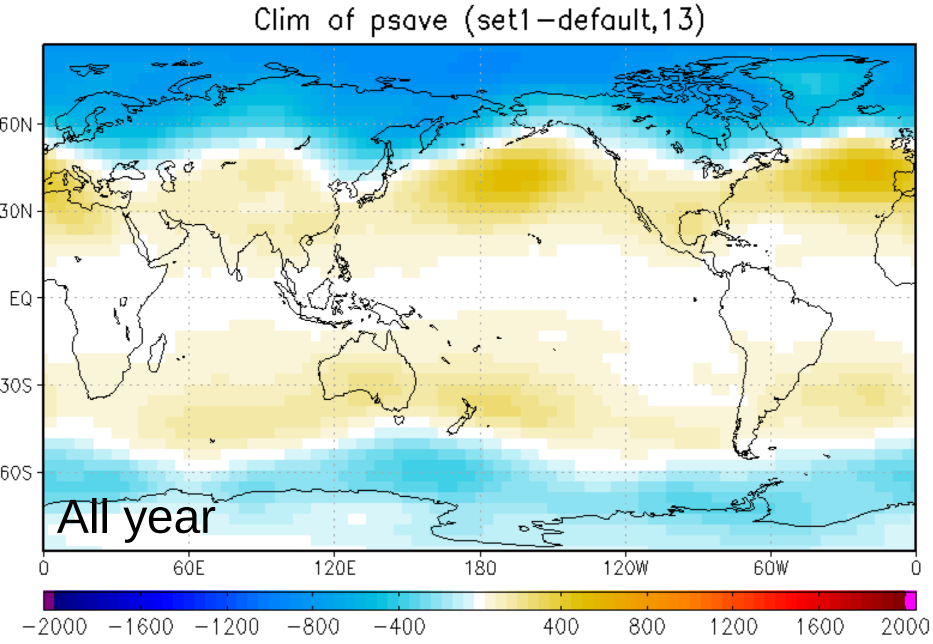
Imperfect model settings

- Modify following parameters from the model used in nature-run
  - Time scale of 6th order hyper-diffusion on T and Vorticity: 18 hours→9 hours
    - Time scale for divergence is 9 hours
  - Albedo on sea-ice: 0.6→0.9
  - Albedo on snow: 0.6→0.9
  - Relaxation time for convection to the reference state: 6 hours→9 hours
  - Drug coefficient for momentum over land: 2.4×10^{-3}→1.2×10^{-3}
    - 0.8×10^{-3} over sea
  - Relaxation time for specific humidity in large scale condensation : 4 hours→6 hours

Difference of model climatology from free-run (mean Ps: Pa)

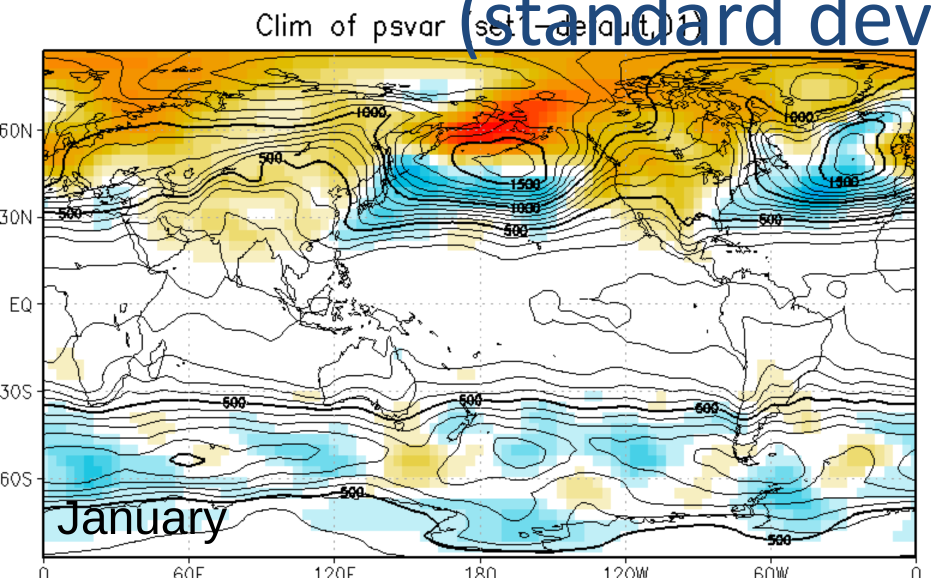


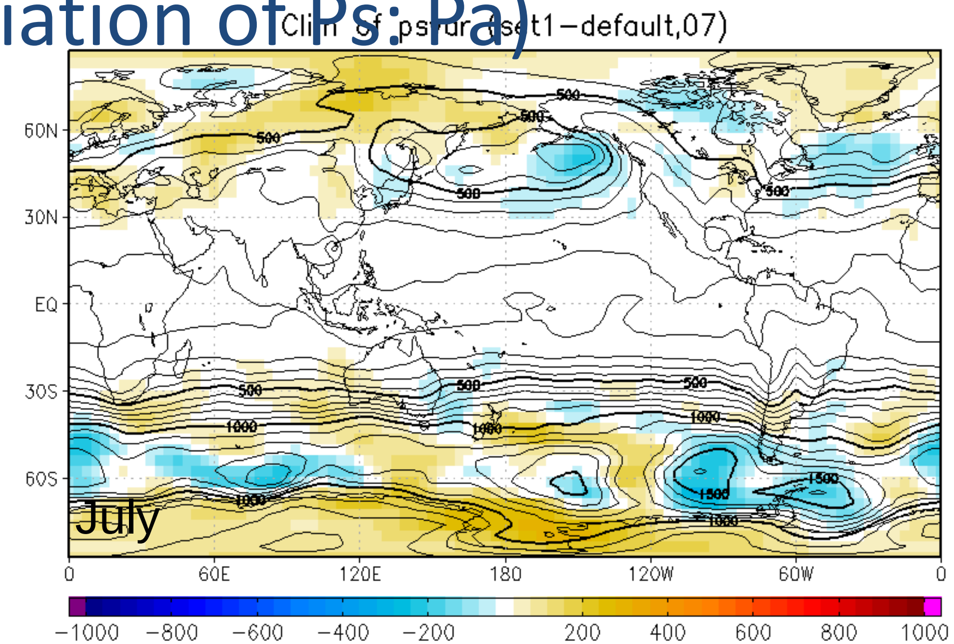


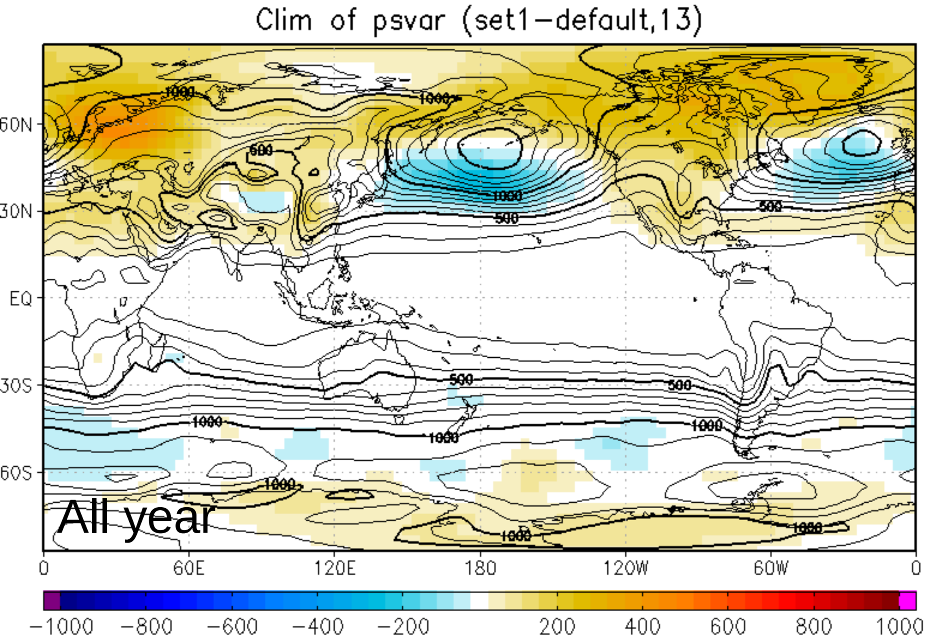


Difference of average surface pressure (Pa) from 10 years free model run (Imperfect model – Perfect model)

Difference of model climatology from free-run (standard deviation of Ps: Pa)

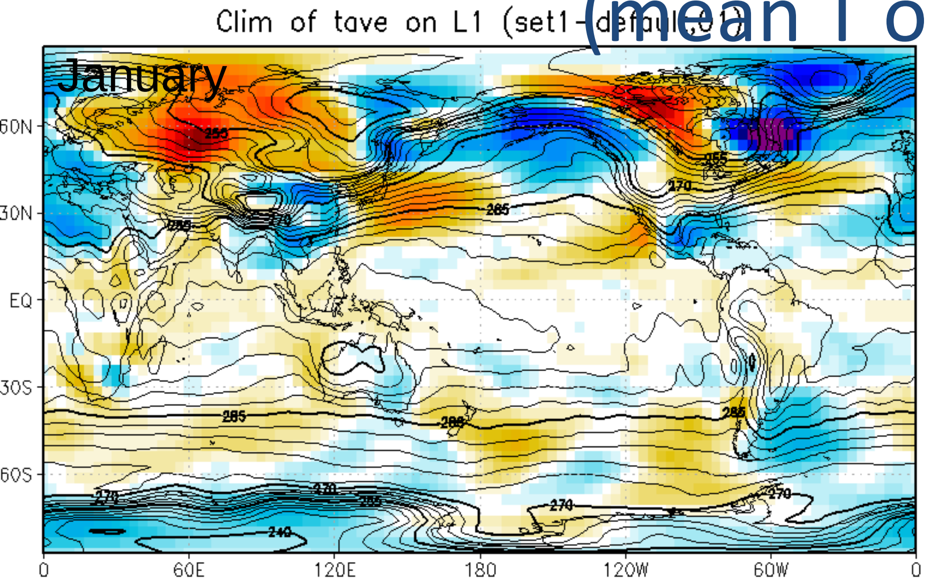


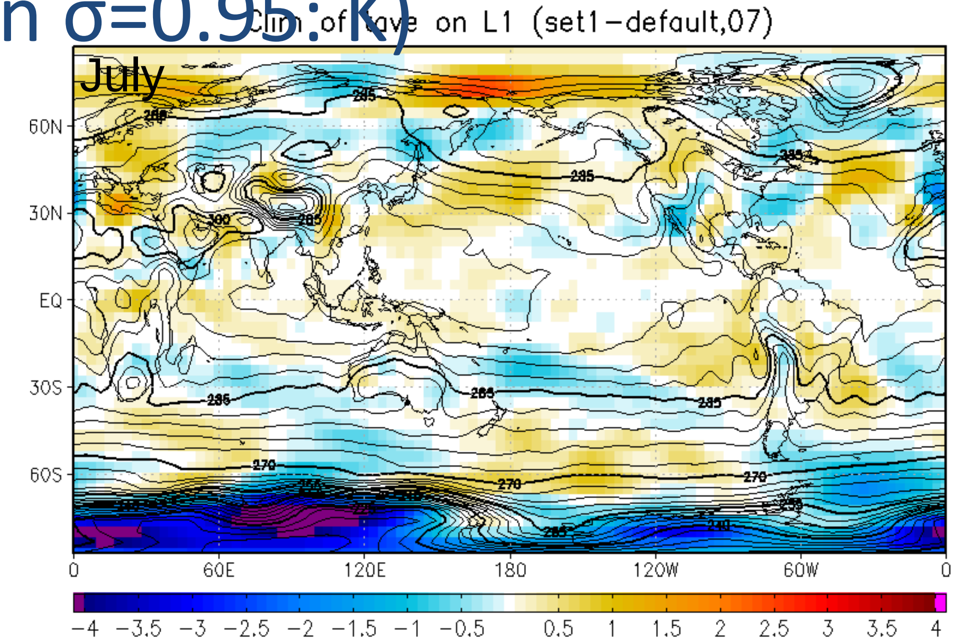


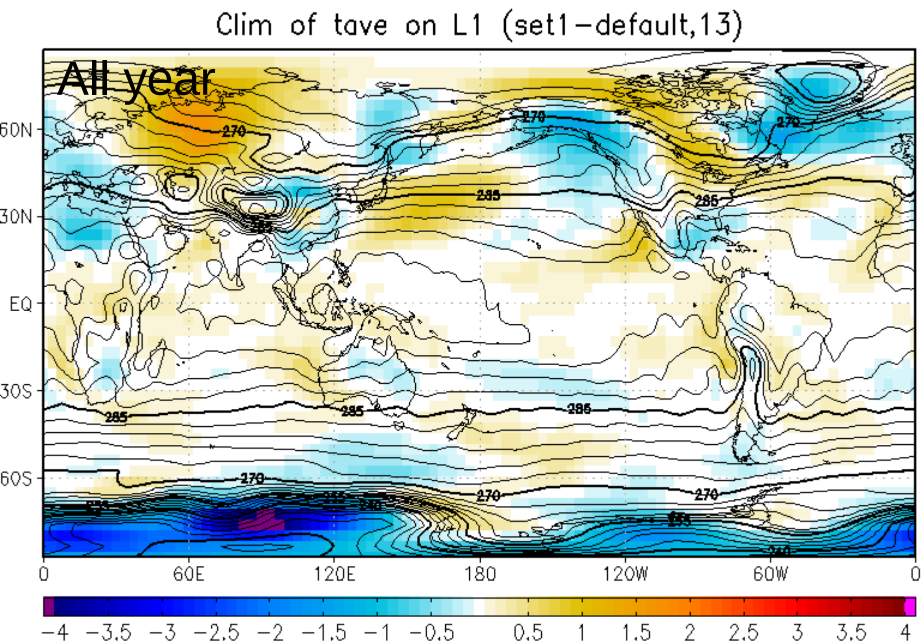


Difference of standard deviation of surface pressure (Pa) from 10 years free model run
(Imperfect model – Perfect model)

Difference of model climatology from free-run (mean T on σ=0.95: K)

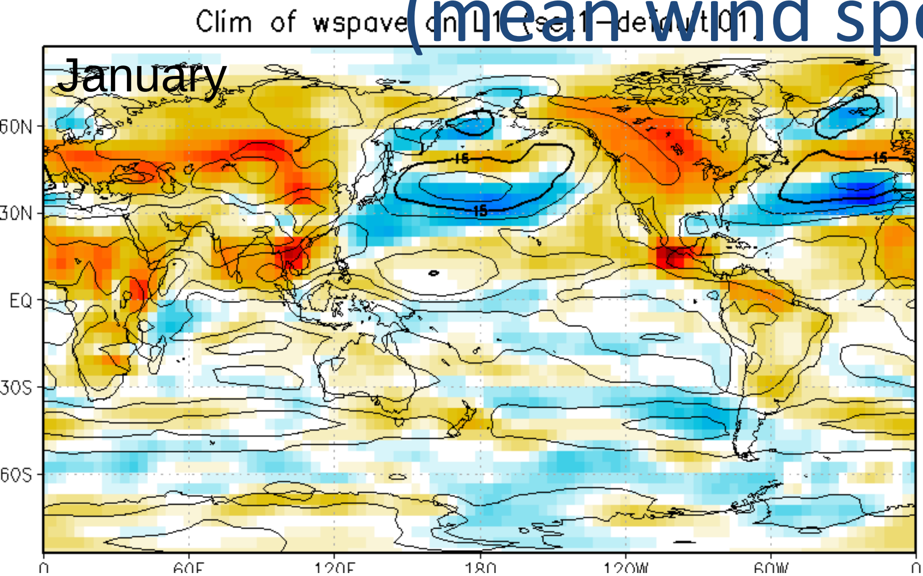


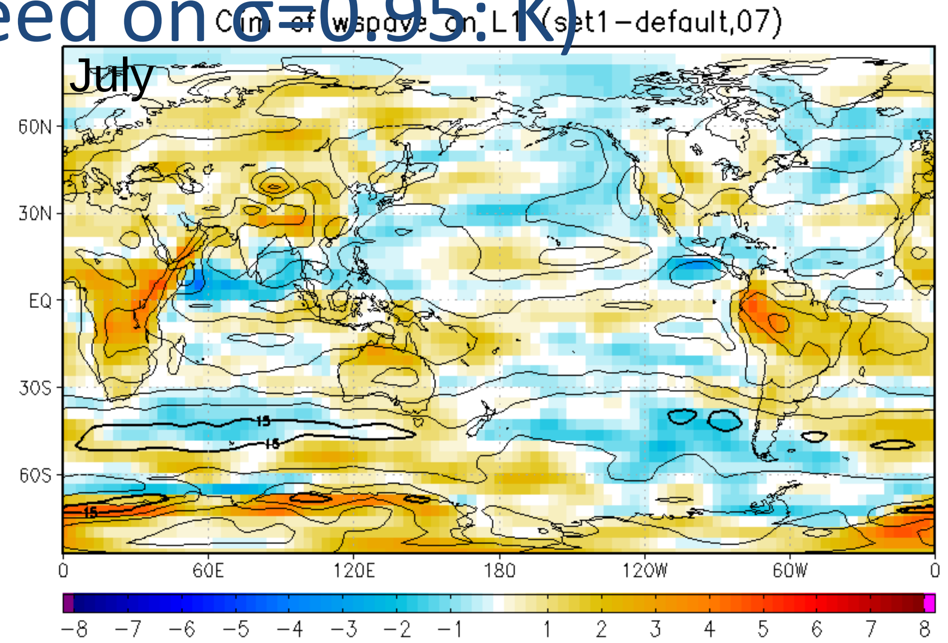


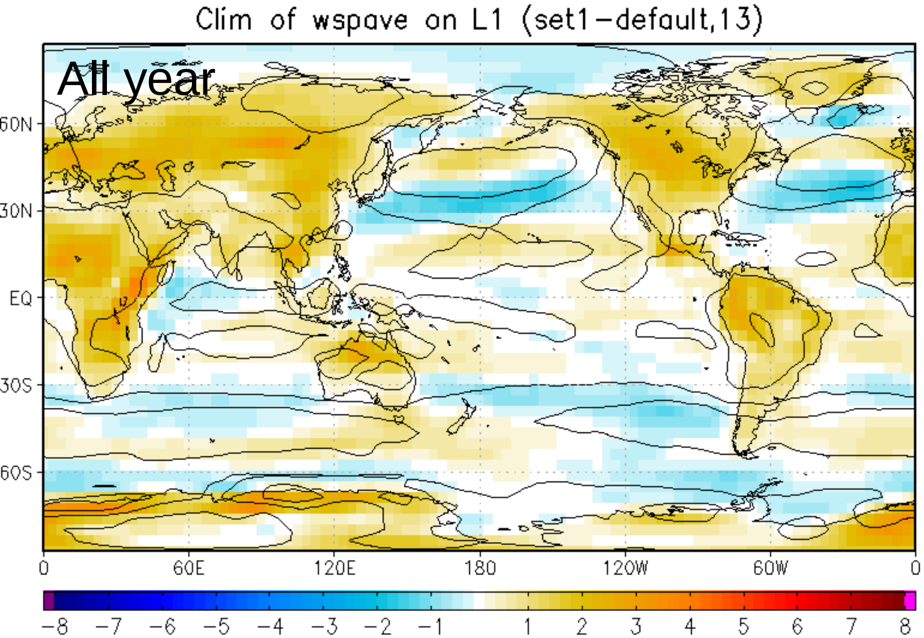


Difference of average temperature on σ=0.95 (K) from 10 years free model run (Imperfect model – Perfect model)

Difference of model climatology from free-run (mean wind speed on σ=0.95: K)

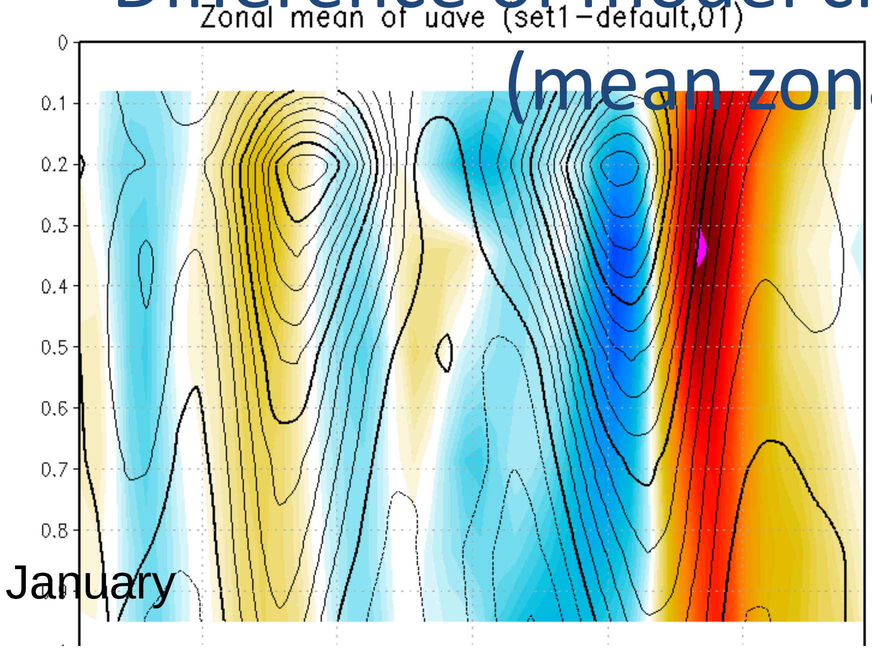


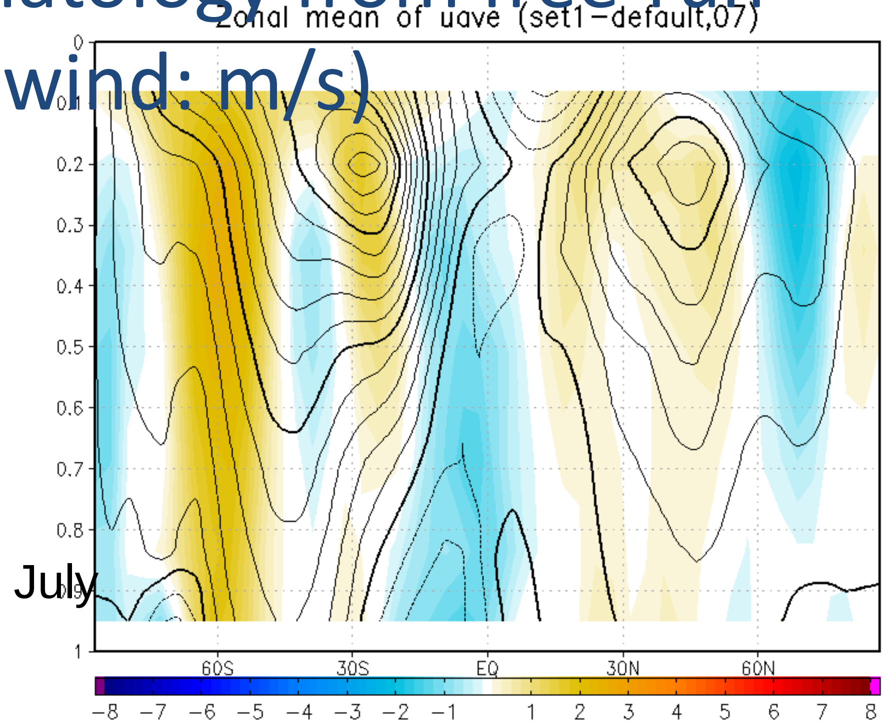


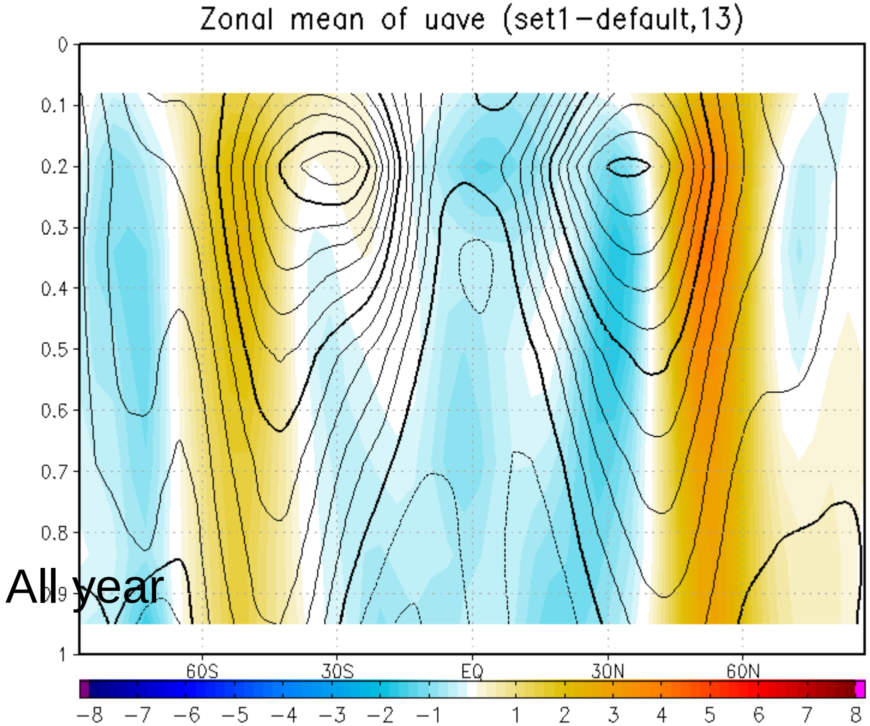


Difference of average wind speed on σ=0.95 (K) from 10 years free model run (Imperfect model – Perfect model)

Difference of model climatology from free-run (mean zonal wind: m/s)

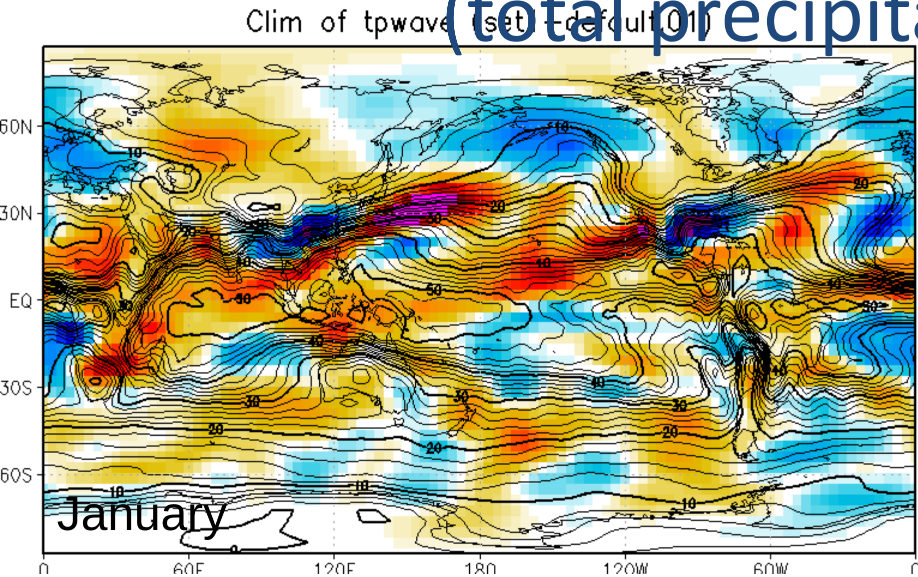


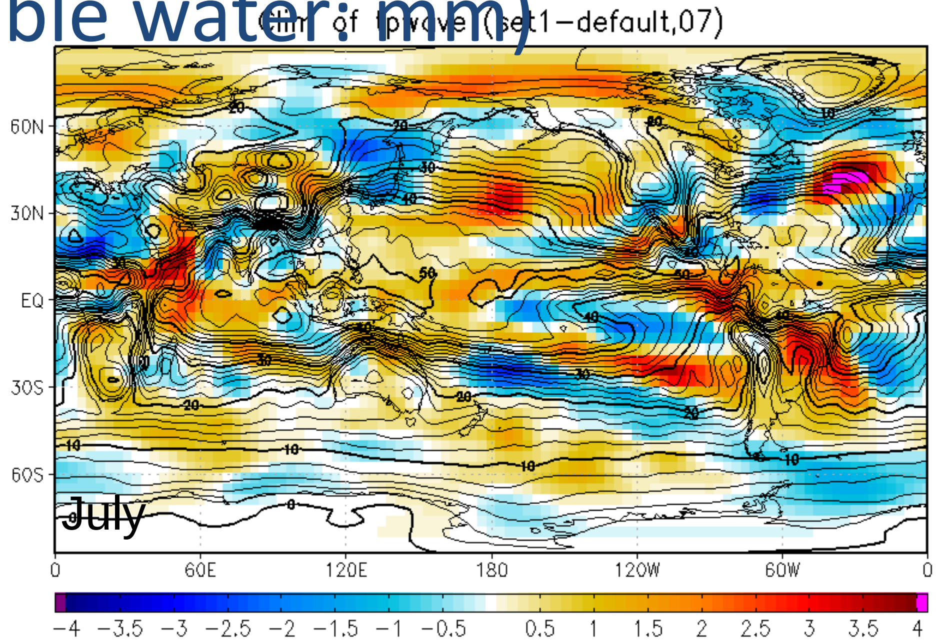


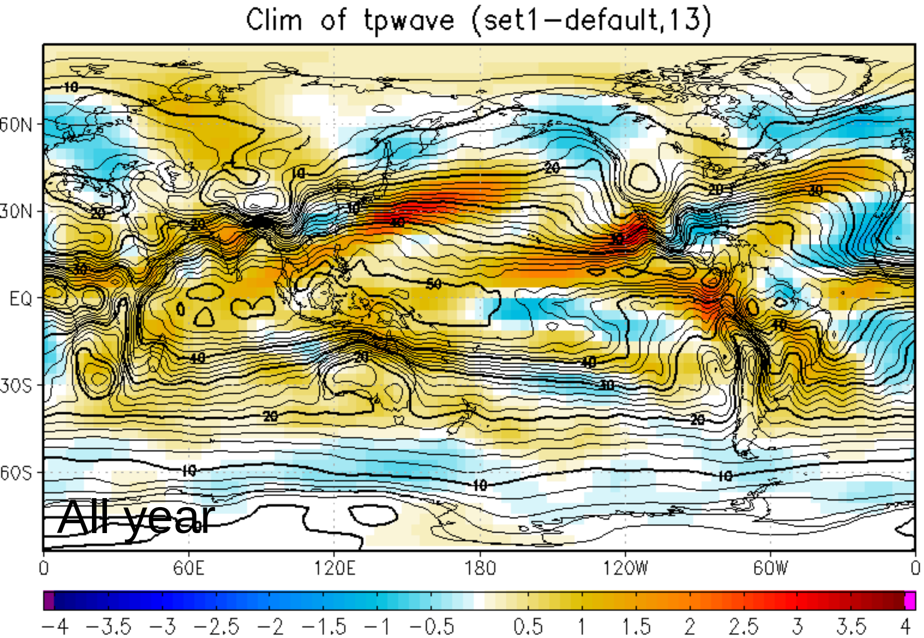


Difference of average zonal wind (m/s) from 10 years free model run (Imperfect model – Perfect model)

Difference of model climatology from free-run (total precipitable water: mm)

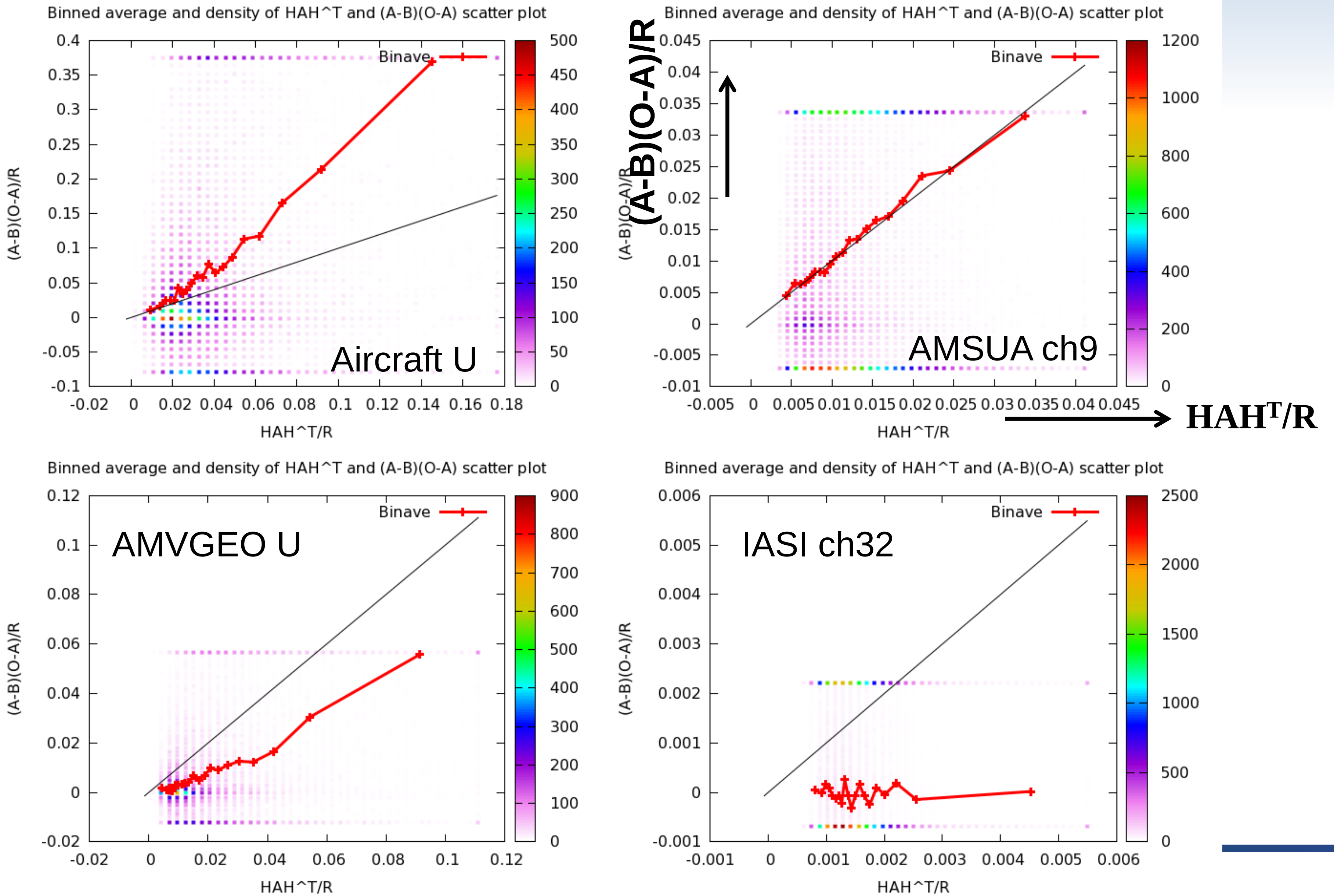






Difference of average total precipitable water (mm) from 10 years free model run (Imperfect model – Perfect model)

Scatter plots for pseudo hybrid DA with real observations



Scatter plots for pseudo hybrid DA with real observations

- Experimental settings
  - 50 member TL319L100 LETKF analysis re-centered around TL959L100 4DVar analysis on every analysis time
  - Localization scale: 400km in horizontal, 0.4 scale height in vertical
  - Period: from 00UTC Dec. 1 2014 to 18UTC Feb. 28 2015
  - Fixed inflation coefficient estimated on 18UTC Nov. 29 2014 with adaptive multiplicative inflation without assimilating AIRS and IASI
- Results
  - Underdispersive for conventional direct observations
  - Adequate spread for AMSUA, CSR and GNSS
  - Overdispersive for AMV, MHS, TMI and SSMIS
  - No clear correlation between (A-B)(O-A) and HAH for hyperspectral IR sounders
  - Suggesting hyperspectral IR sounders should not be used for inflation estimation.
  - Indicating the specified observation errors for some observations are not optimal.