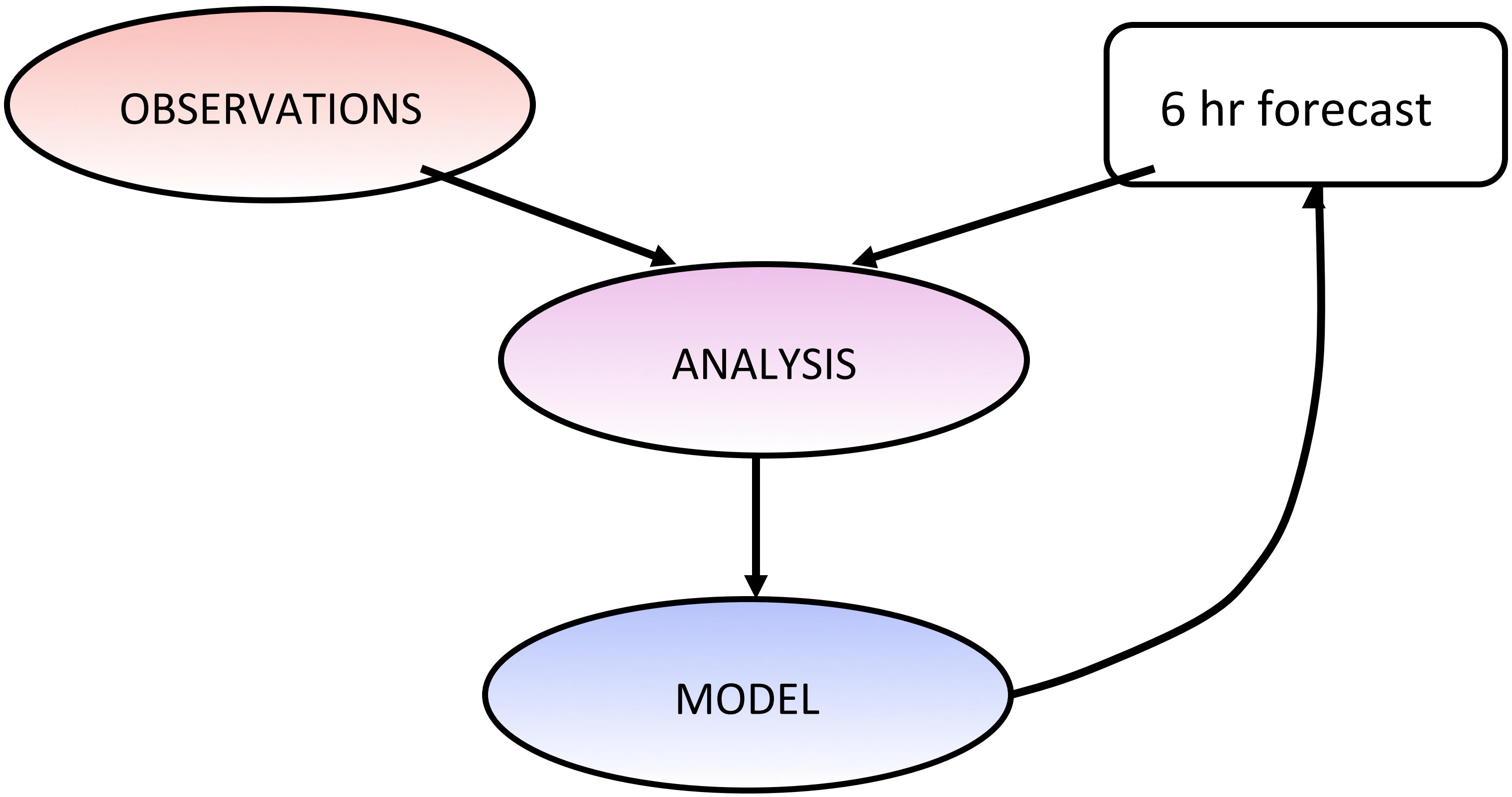
Future directions of ensemble-based data assimilation (?)

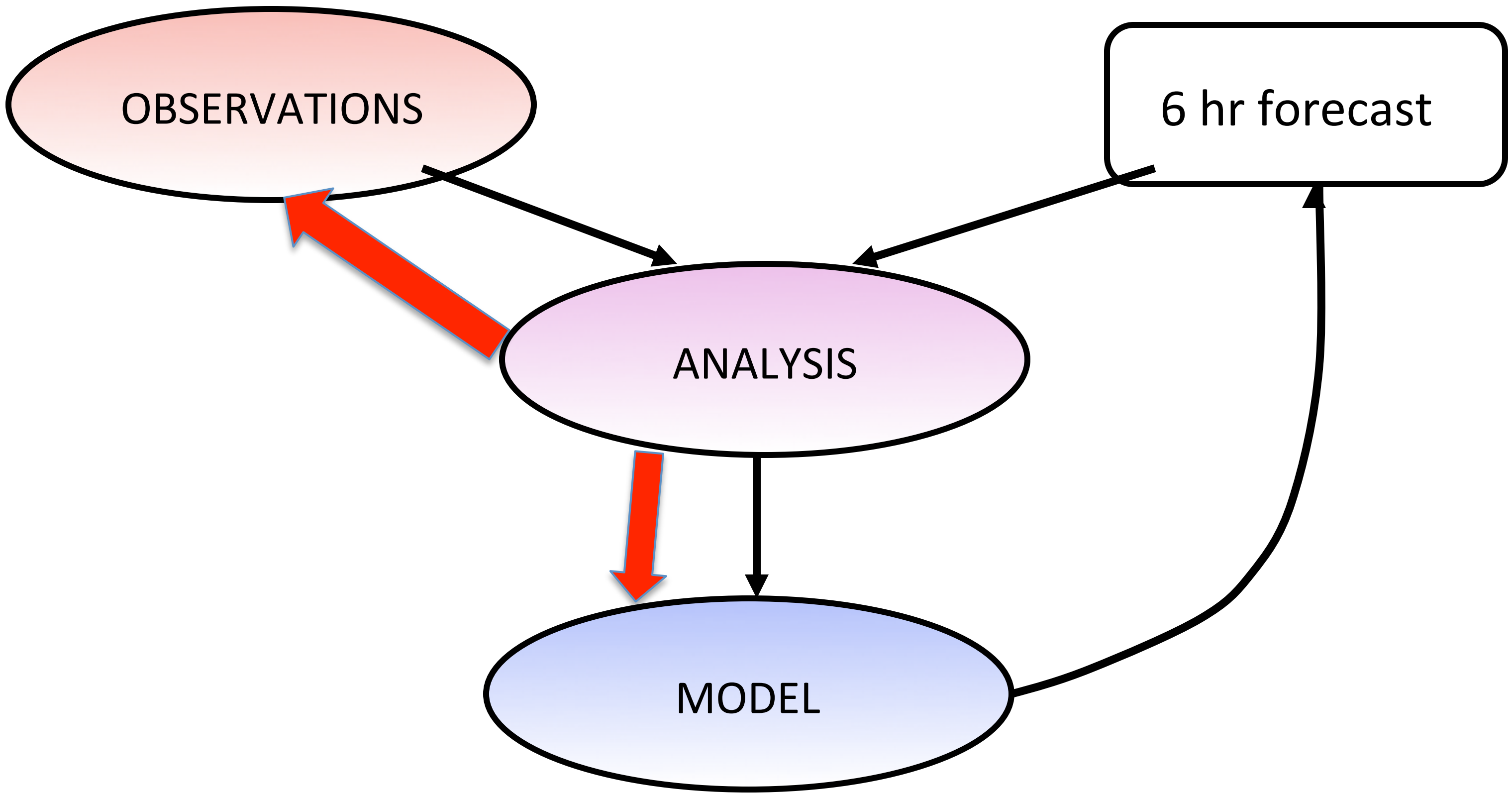
Eugenia Kalnay
with many thanks to friends and colleagues
at the University of Maryland

6^{th} EnKF Workshop, 19-22 May 2014

Data Assimilation: We need to improve observations, analysis scheme and model



Data Assimilation: We can also use it to improve observations and model



The simplicity and power of EnKF should encourage the use of DA for improvements beyond its main goal, namely

1) Combine optimally observations and model forecasts (done) ☺

- We should also use DA to:

2) Improve the observations

3) Improve the model

- Also, do more truly coupled DA:

4) Example: The ocean and the atmosphere are coupled: obviously the best DA should be coupled

- Currently the Earth System models used by IPCC for climate change do not predict population, they obtain it from UN projections.

5) We should do DA of the coupled Earth System-Human System

The simplicity and power of EnKF should encourage the use of DA for improvements beyond its main goal, namely

1) Combine optimally observations and model forecasts

- We should also use DA to:

2) Improve the observations

3) Improve the model

- Also, do more truly coupled DA:

4) Example: The ocean and the atmosphere are coupled: obviously the best DA should be coupled

- Earth system models used by IPCC have many submodels, but they don't include the Human system, which totally dominates the Earth system.

5) We should do DA of the coupled Earth System-Human System

The simplicity and power of EnKF should encourage the use of DA for improvements beyond its main goal, namely

1) Combine optimally observations and model forecasts

- We should also use DA to:

2) Improve the observations

3) Improve the model

- Also, do more truly coupled DA:

4) Example: The ocean and the atmosphere are coupled: obviously the best DA should be coupled

- Earth system models used by IPCC have many submodels, but they don't include the Human system, which totally dominates the Earth system.

5) We should do DA of the coupled Earth System-Human System

The simplicity and power of EnKF should encourage the use of DA for improvements beyond its main goal, namely

1) Combine optimally observations and model forecasts

- We should also use DA to:

2) Improve the observations

3) Improve the model

- Also, do more truly coupled DA:

4) Example: The ocean and the atmosphere are coupled: obviously the best DA should be coupled

- Earth system models used by IPCC have many submodels, but they don't include the Human system, which totally dominates the Earth system.

5) We should do DA of the coupled Earth System-Human System

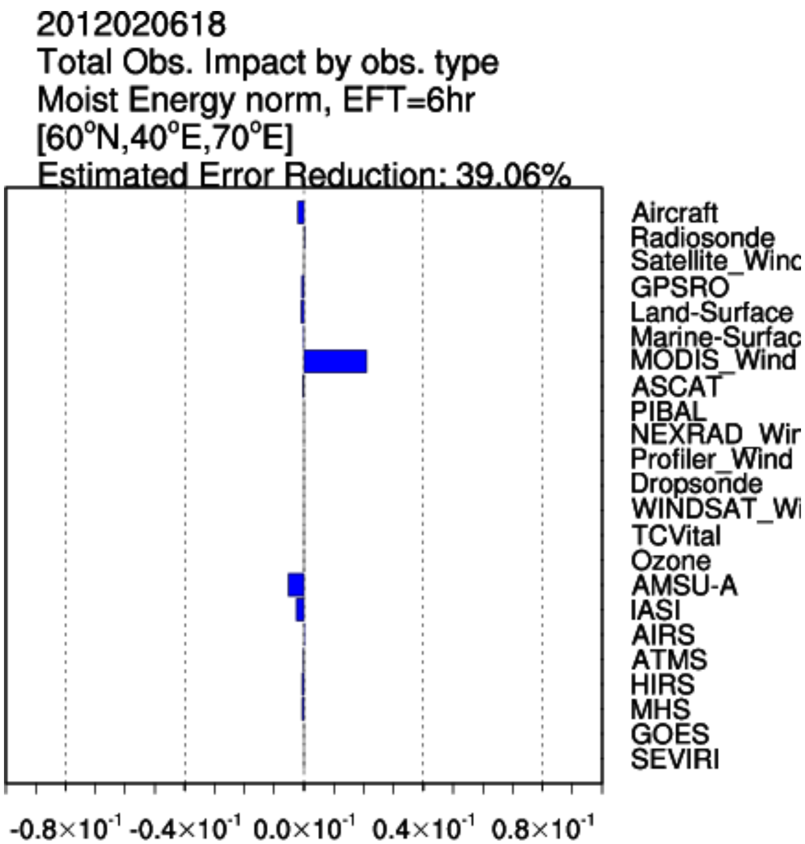
2) Improve the observations: Ensemble Forecast Sensitivity to Observations and Proactive QC

- Kalnay et al. (2012) derived EFSO
- Ota et al. (2013) tested 24hr forecasts and showed EFSO could be used to identify bad obs.
- Hotta (2014) showed that EFSO could be used after only 6 hours, so that the bad obs can be withdrawn and collected with useful metadata so they can be improved.
- We call this **Proactive QC**, much stronger than QC.
- Hotta also showed EFSO can be used to tune **R**
- Lien (2014) tested EFSO to identify useful observations of precipitation, with good results.

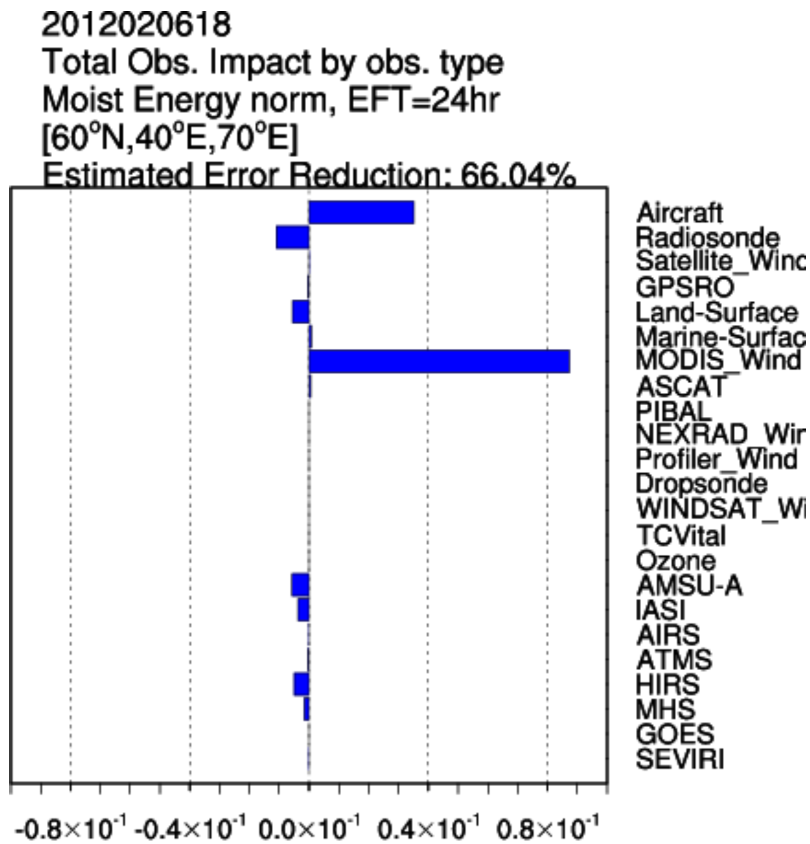
Hotta (2014)

Feb. 18 06UTC, near the North Pole (Ota et al. 2013 case). Suspect: MODIS

EFT=06 hr.

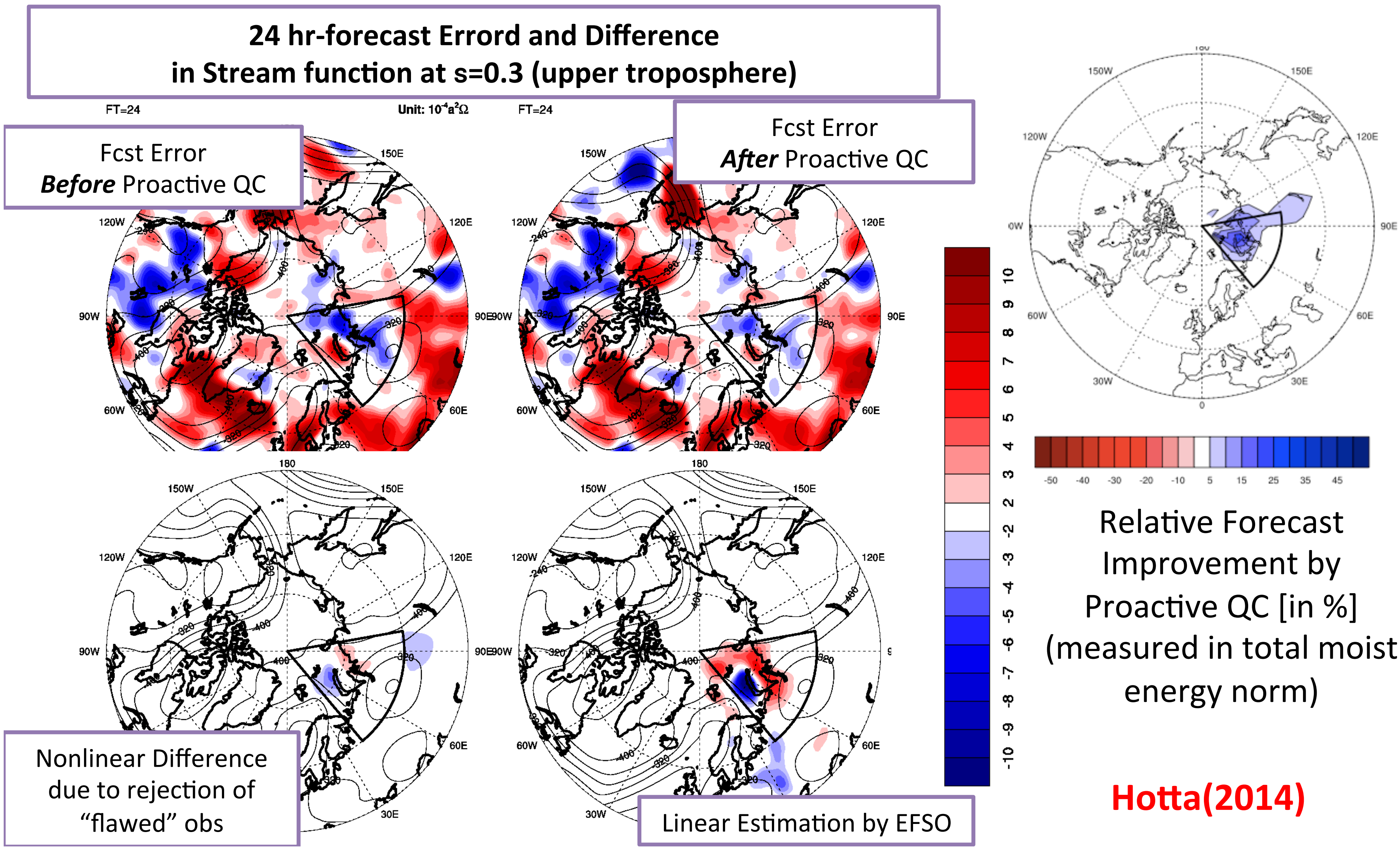


EFT=24 hr.



Can identify the bad observations after only 6 hours!

- On Feb 18, 06 UTC 2012, MODIS Winds were identified as “flawed” observation
- Rejection of the detected “flawed” observations in fact improved the forecast!
- EFSO estimated much stronger “correction” (right panel) than the actual impact (middle panel)



Ensemble Forecast Sensitivity to **Error Covariances**

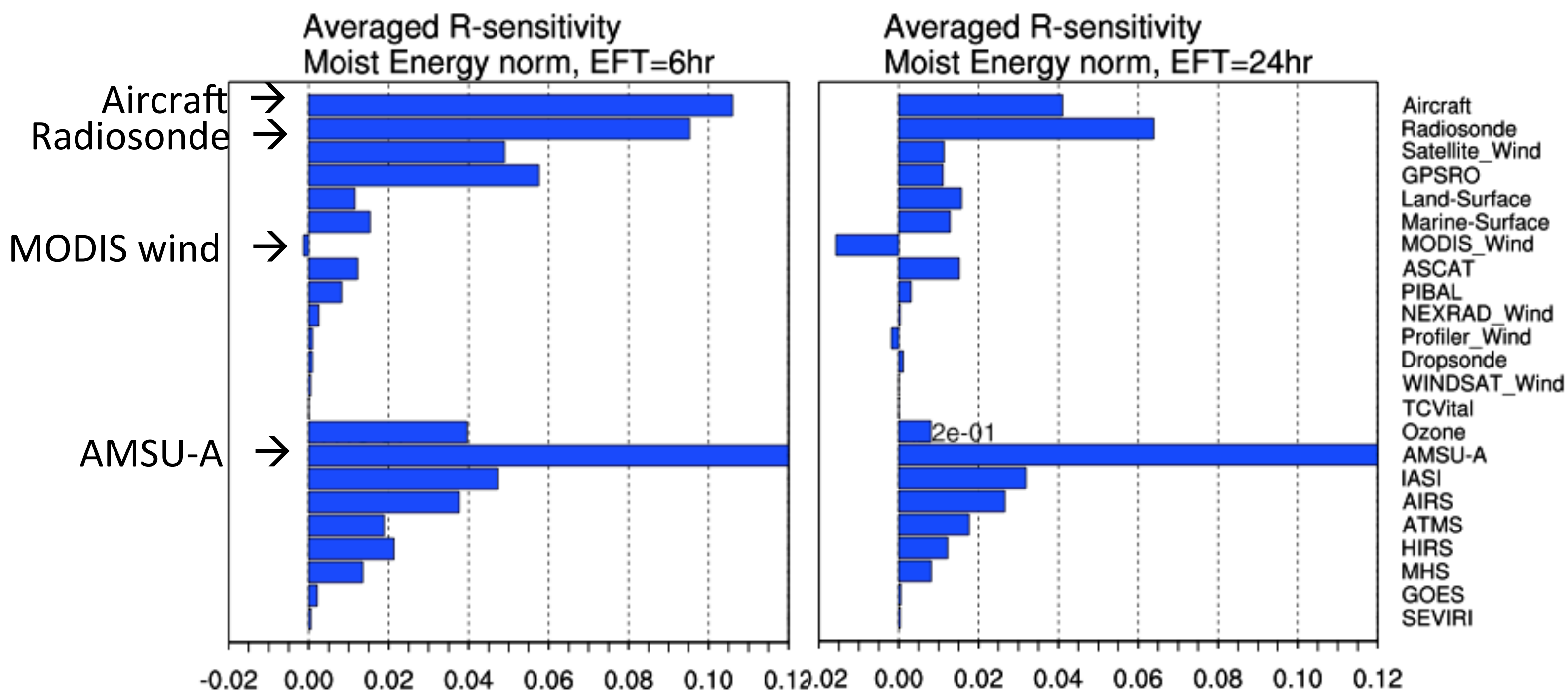
Hotta (2014)

- Daescu and Langland (2013, *QJRMS*) proposed an adjoint-based formulation of forecast sensitivity to **B** and **R** matrix.
- Daisuke Hotta formulated its ensemble equivalent for **R** using **EFSO** by Kalnay et al. (2012) :

$$\left[\frac{\partial e}{\partial \mathbf{R}}\right]_{ij} \approx \frac{\partial e}{\partial y_i} z_j \approx -\frac{1}{K-1}\left[\mathbf{R}^{-1}\mathbf{Y_0^a}\mathbf{X_{t|0}^{fT}}\mathbf{C}\left(\mathbf{e_{t|0}}+\mathbf{e_{t|-6}}\right)\right]_i \left[\mathbf{R}^{-1}\delta y^{oa}\right]_j$$

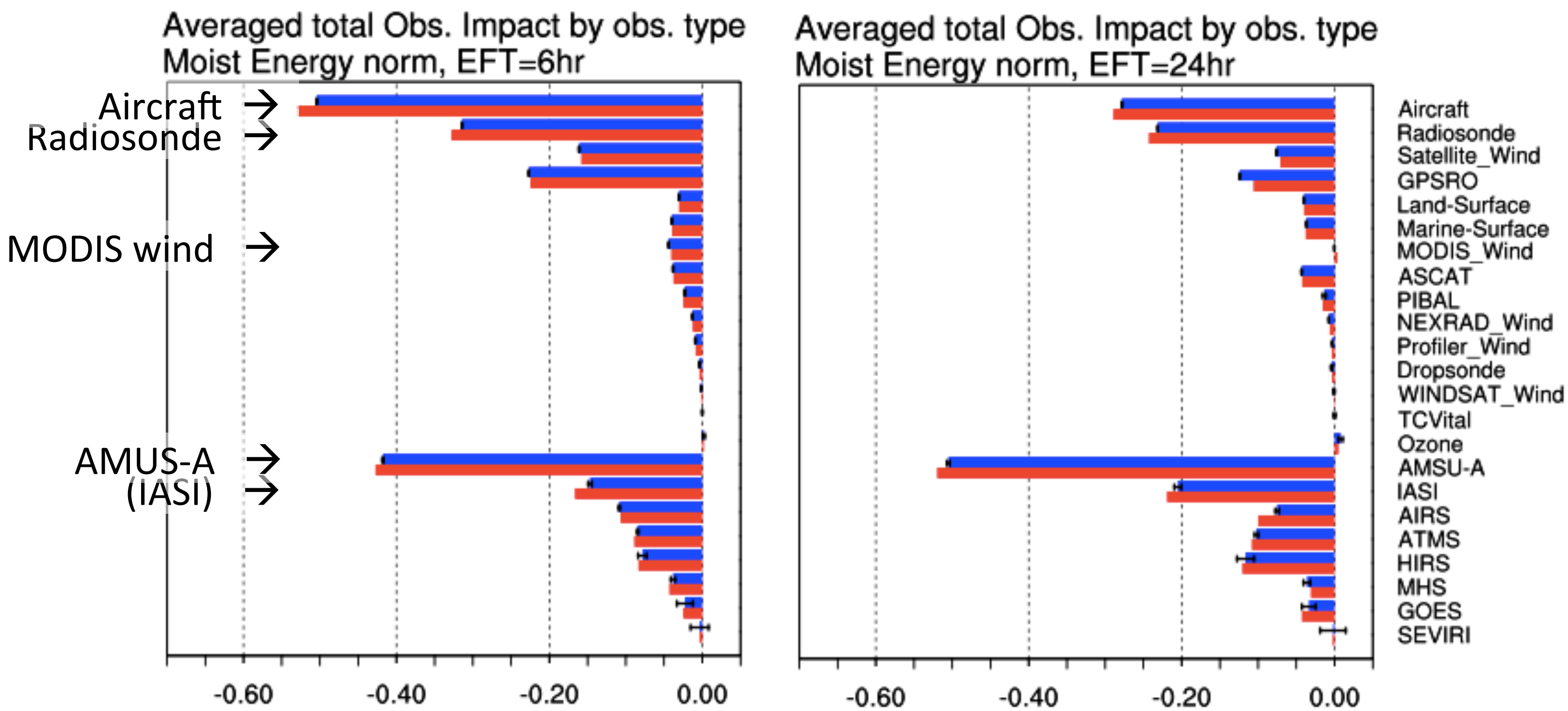
where $\mathbf{z}$ is an "intermediate analysis increment" in observation space

Result from GFS / GSI-LETKF hybrid



- Positive value: error increases as s_o^2 increases → should decrease s_o^2
- Aircraft, Radiosonde and AMSU-A: large positive sensitivity
- MODIS wind : negative sensitivity
- → **Tuning experiment**:
    - Aircraft, Radiosonde and AMSU-A: scale s_o^2 by 0.9
    - MODIS wind: scale s_o^2 by 1.1

Tuning Experiment: Result
EFSO before/after tuning of R



- Aircraft, Radiosonde and AMSU-A: significant improvement of EFSO-impact
- MODIS wind : insignificant difference in EFSO
- IASI: Significant improvement in EFSO although its error covariance is untouched!

2) Effective Assimilation of Real Precipitation

(Guo-Yuan Lien, E. Kalnay and T Miyoshi)

- Assimilation of precipitation has been done by changing the moisture *Q* in order to make the model "rain as observed".
- Successful during the assimilation: e.g., the North American Regional Reanalysis had perfect precipitation!
- However the model **forgets** about the changes soon after the assimilation stops!
- The model **will remember** potential vorticity (PV).
- EnKF should modify PV efficiently, since the analysis weights will be larger for an ensemble member that is raining more correctly, because it has a better PV.

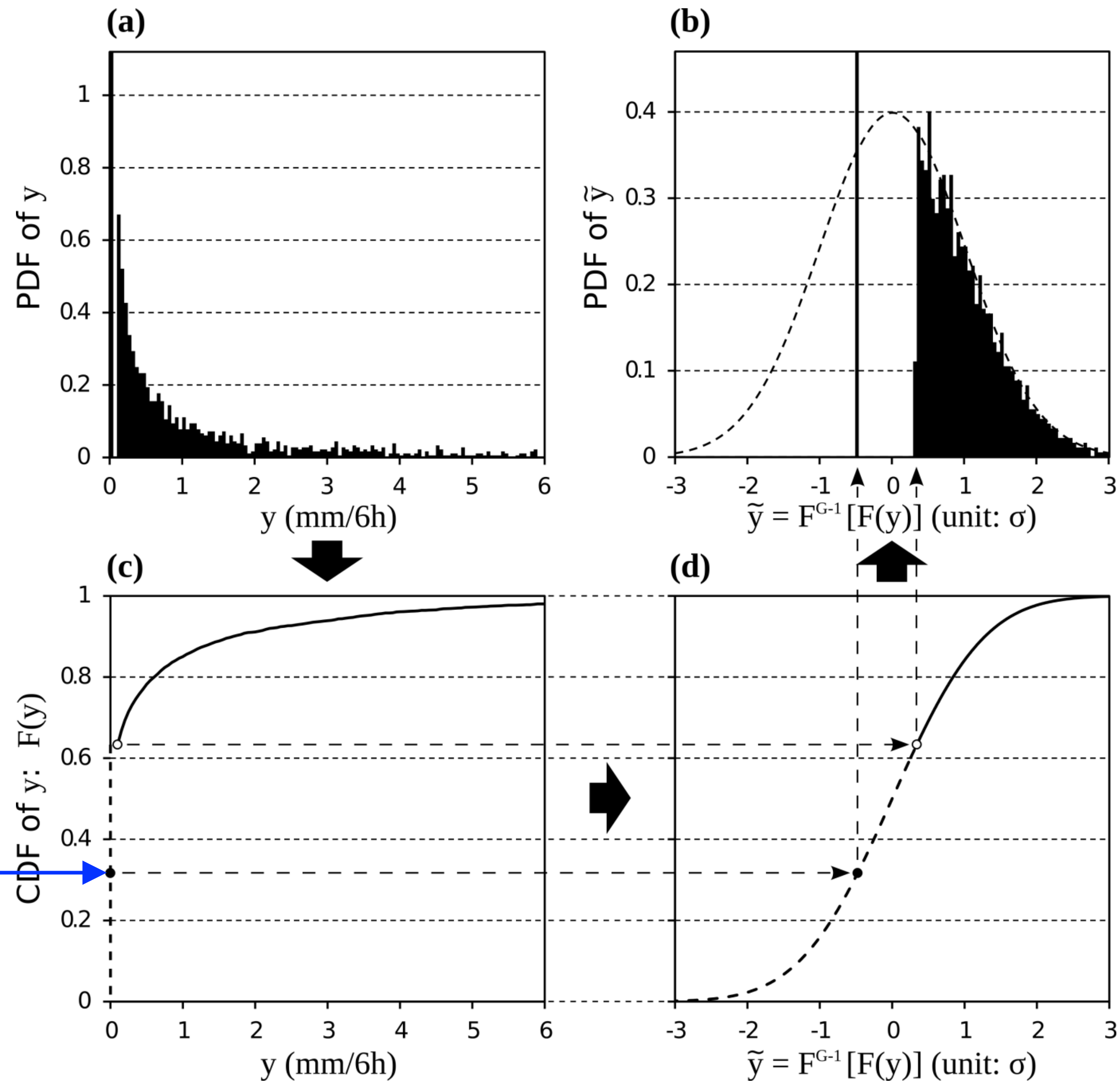
- However, ~7 years ago, we had tried assimilating precipitation observations in a LETKF-SPEEDY model simulation but the results were POOR!
- Big problem: precipitation is not Gaussian.
- We tried a Gaussian transformation of precipitation and **it worked**!

Example of Gaussian precipitation transformation

PDF

CDF

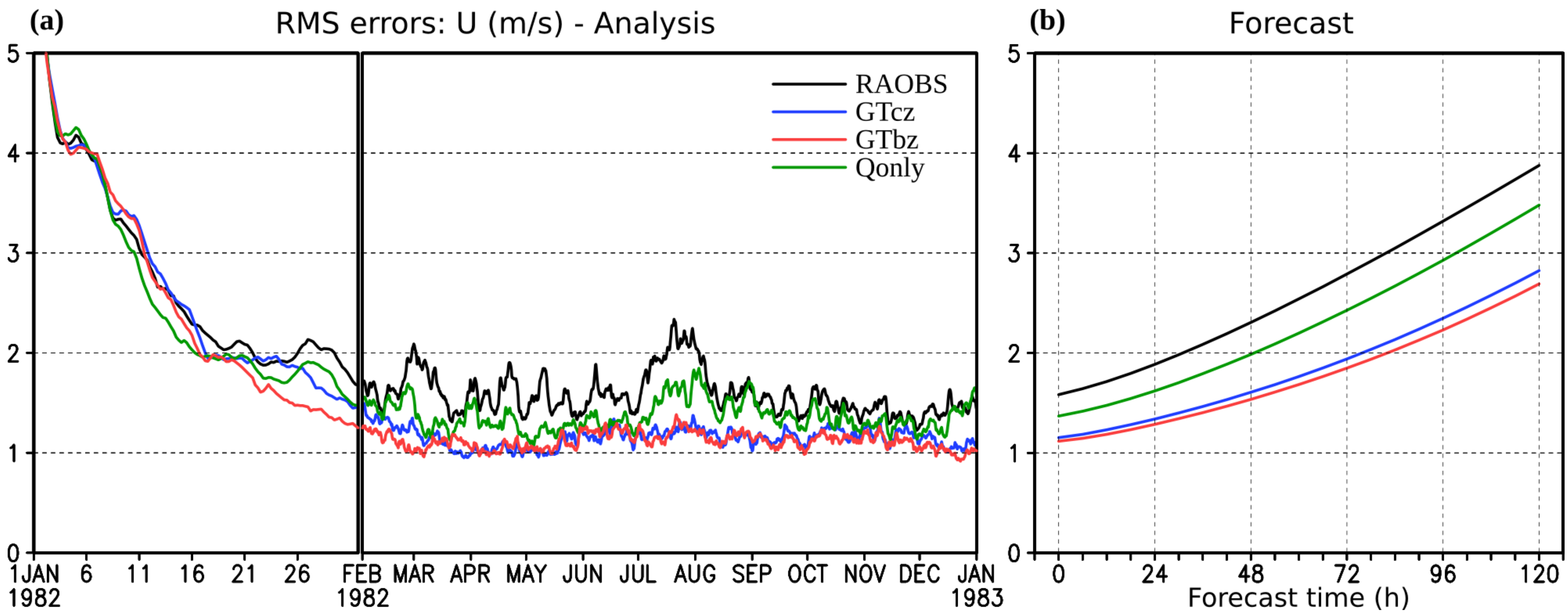
The climatological median of zero (GTcz) method



Original variable

Transformed variable

Average analysis and forecast errors (OSSE)



(Spin-up) (After the spin-up) (11-month average after the spin-up period)

RAOBS: Assimilate rawinsonde observations

GTcz: Assimilate rawinsondes + uniformly distributed global precipitation using **GTcz**

GTbz: Assimilate rawinsondes + uniformly distributed global precipitation using **GTbz**

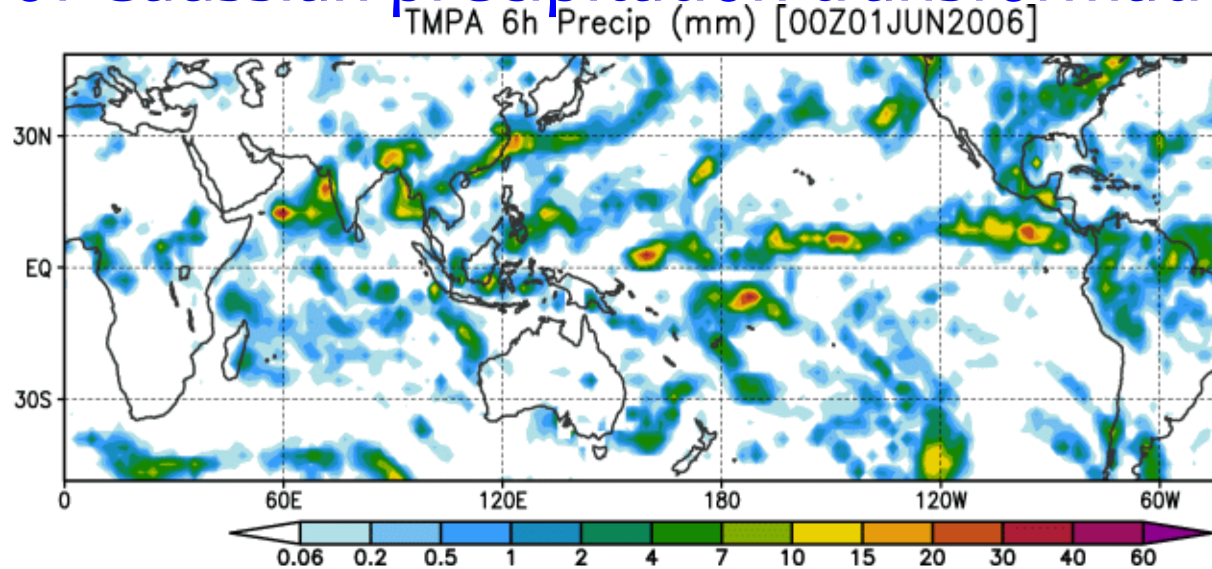
Qonly: Same as **GTcz**, but only update moisture field by precipitation assimilation

(Other variables show similar results)

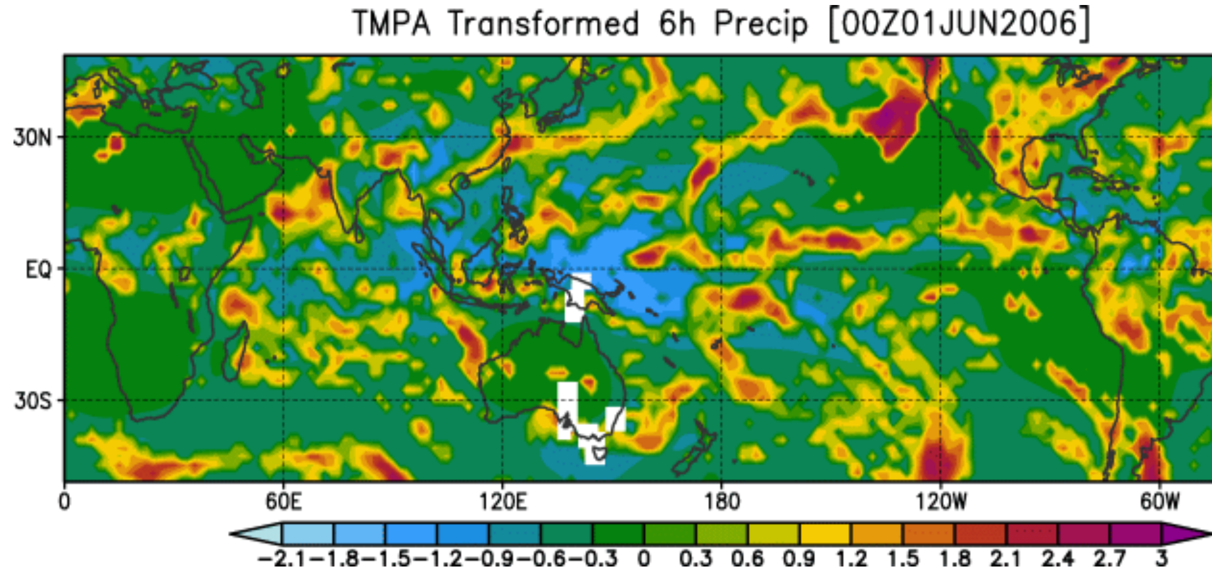
REAL OBSERVATIONS (TMPA)

Example of Gaussian precipitation transformation

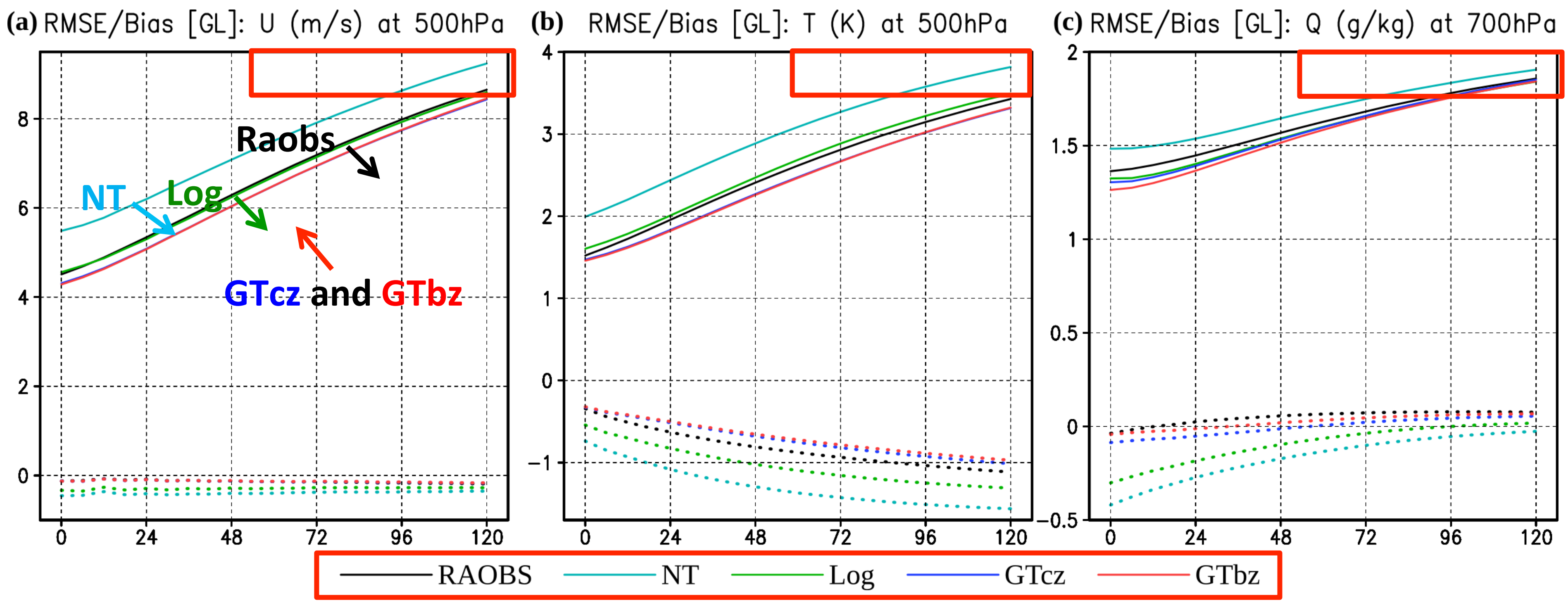
Original variable



Transformed variable



RESULTS: Average RMSE/bias vs. forecast time

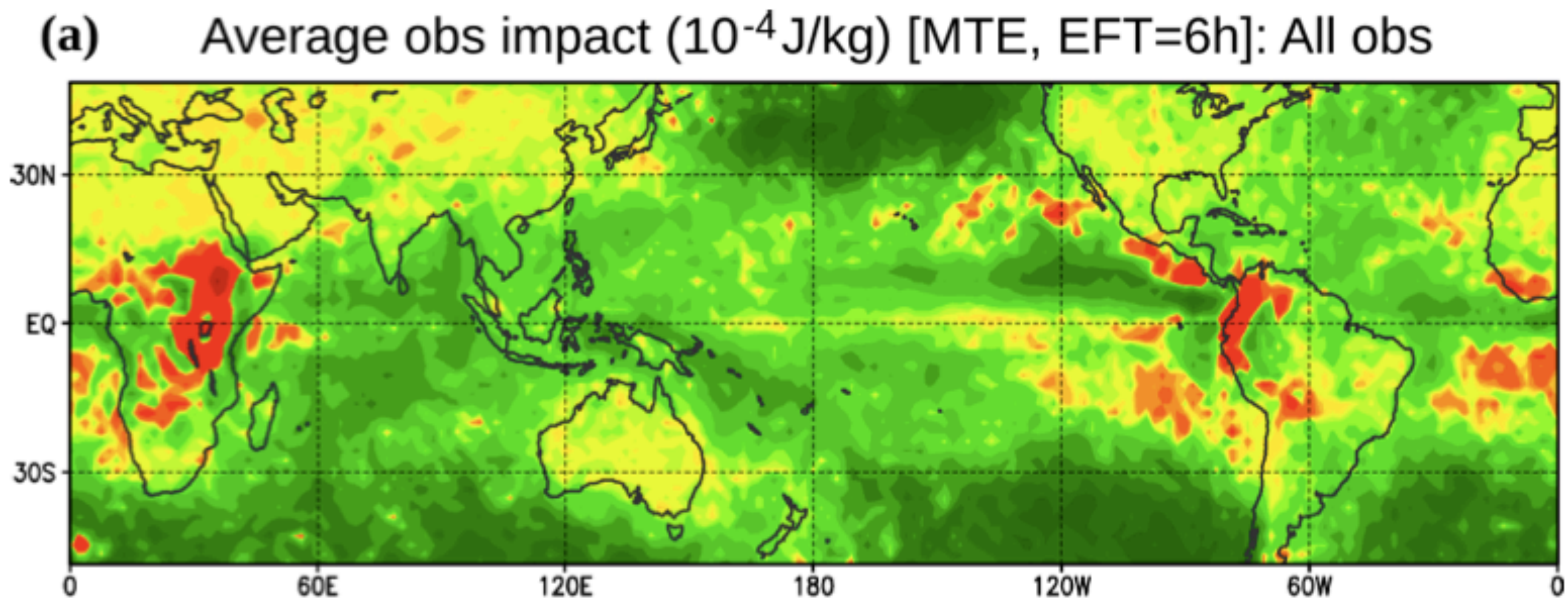


Global results Solid lines: RMS errors Dashed lines: Biases

- **RAOBS: Control**
- **No transform (NT)** gives very bad results.
- **The standard Log** transformation: marginal results.
  - Good for moisture, but bad for temperature.
- **GTcz** and **GTbz** are almost the same, both leading to clear positive impacts.
- These required several conditions to be successful: many observations don't help!

Guo-Yuan Lien (2014)

EFSO average impact of rain observations

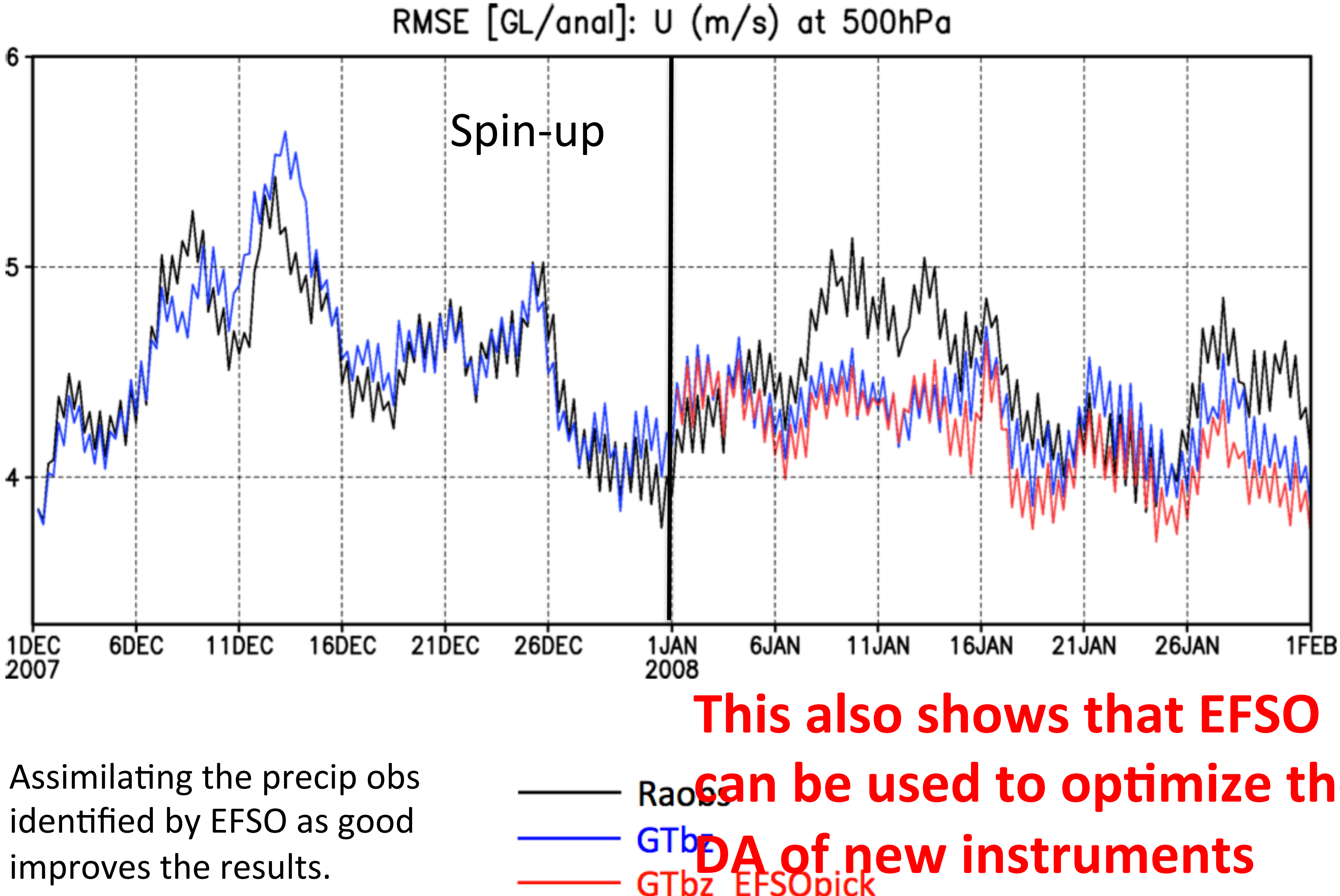


Assimilating the precip obs identified by EFSO as good improves the results.

This also shows that EFSO can be used to optimize the DA of new instruments efficiently!

One-month time series: Analysis U (m/s) at 500 hPa

Guo-Yuan Lien (2014)



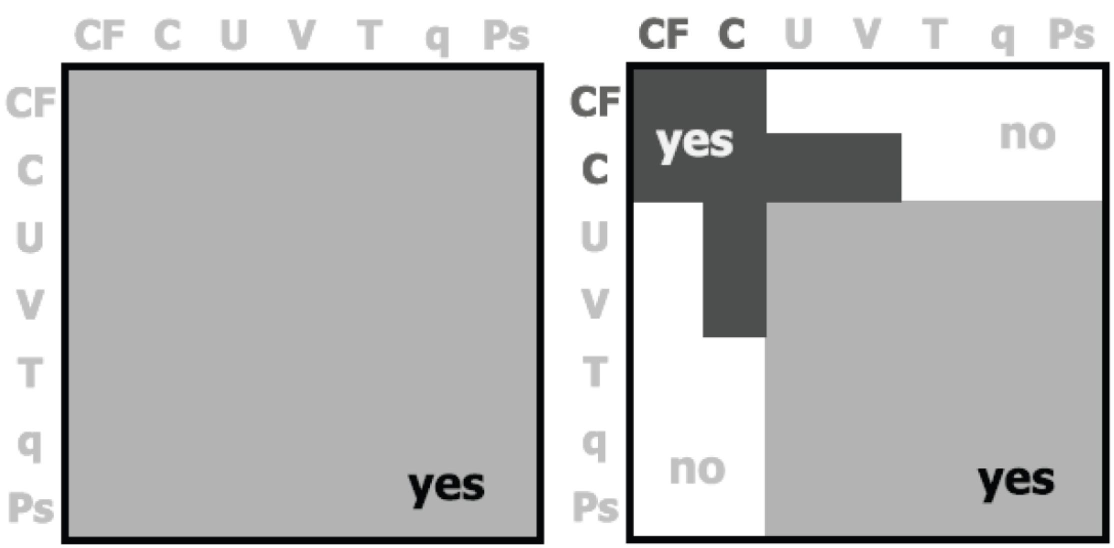
Assimilating the precip obs identified by EFSO as good improves the results.

This also shows that EFSO can be used to optimize the DA of new instruments efficiently!

3) Improve the models: Parameter estimation and estimation of bias using DA

- Model tuning on long time scales should be done with EnKF parameter estimation.
- Kang et al., JGR, 2011, 2012 showed that evolving surface carbon fluxes can be estimated accurately at the model grid resolution from simulated atmospheric CO2 observations (OCO-2) as evolving parameters.
- Another approach is the use of analysis increments to estimate model bias (Greybush et al., 2012, Mars) and even state-dependent model bias (e.g., El Niño bias), as in Danforth et al. 2007.

Surface carbon fluxes **from** atmospheric assimilation of meteorological variables and CO2 obtained as evolving parameters (OSSE). Kang et al., JGR, 2012

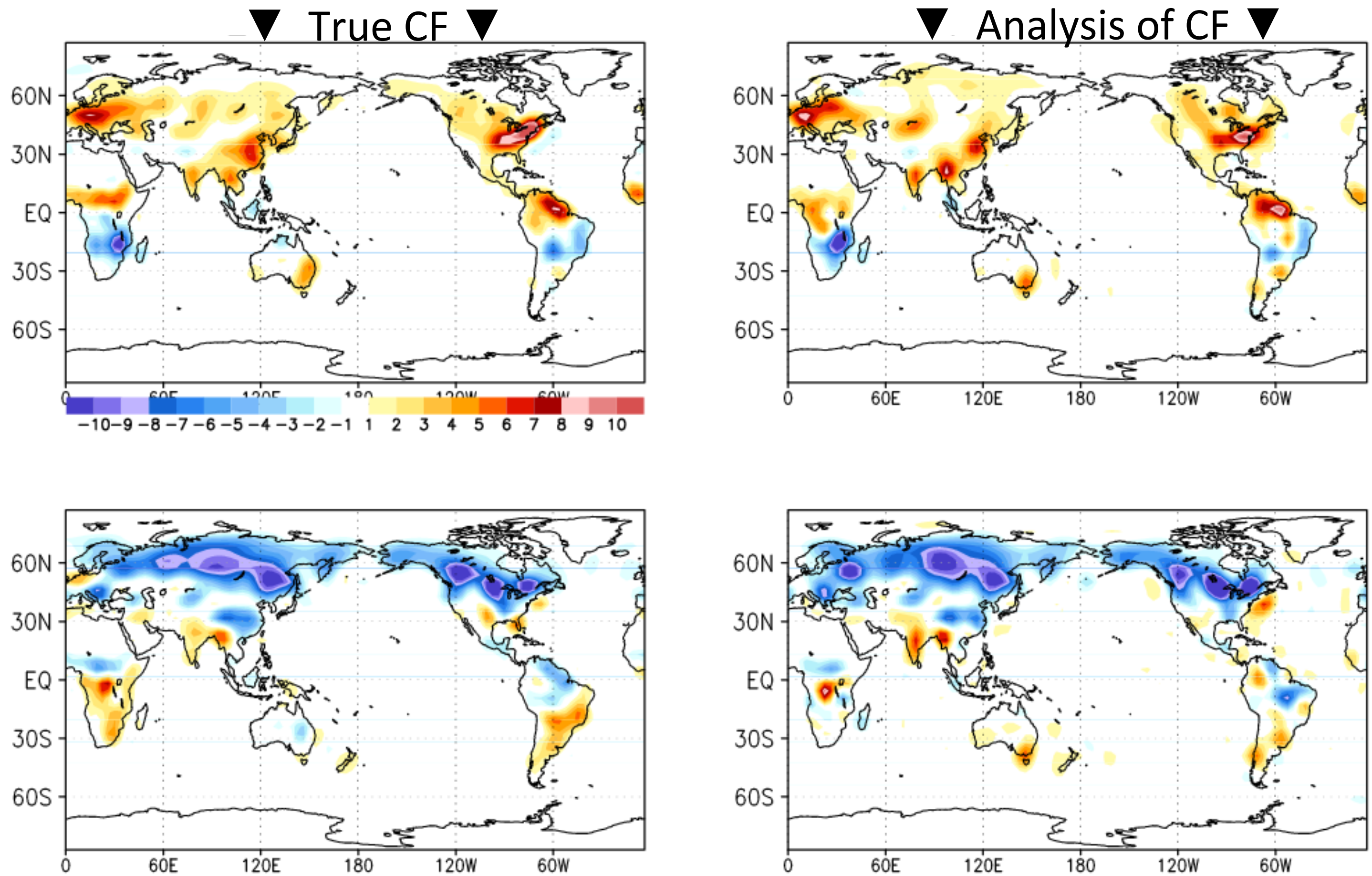


"Variable Localization"

OSSE Results

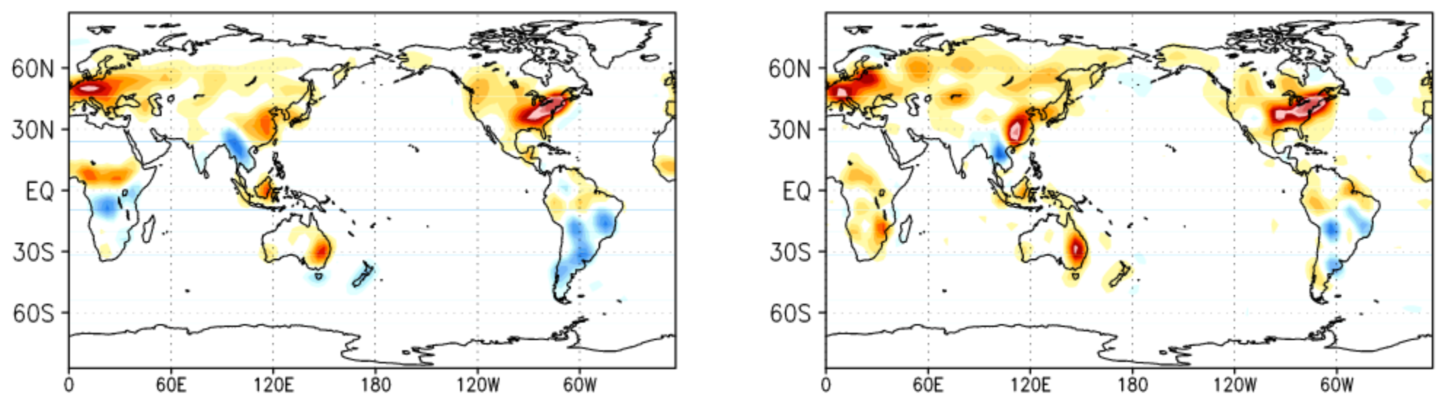
00Z01APR ▶
After three months of DA

00Z01AUG ▶
After seven months of DA



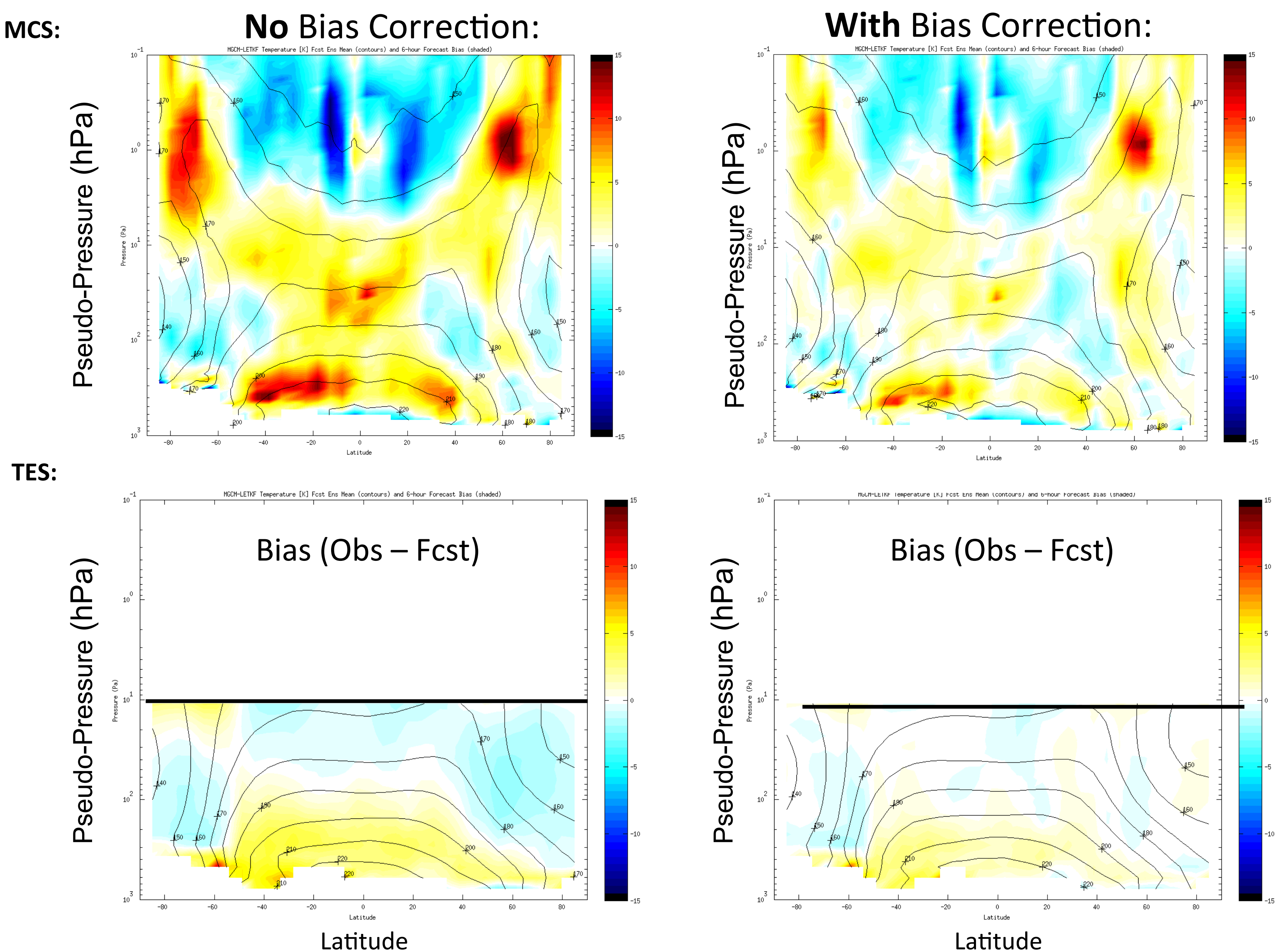
We succeed in estimating time-evolving CF at model-grid scale

00Z01JAN ▶
After one year of DA

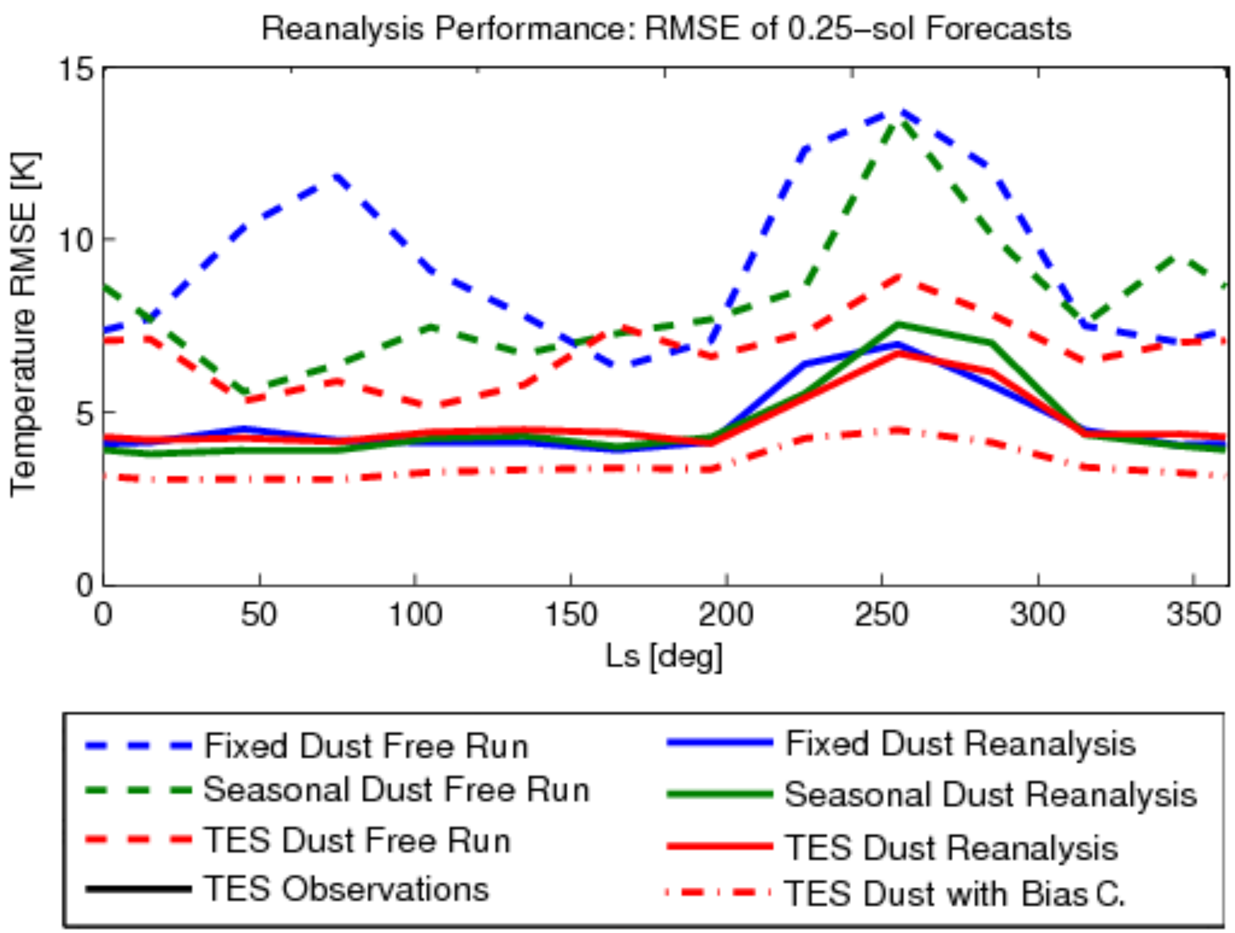


Example: Mars Bias Correction

Steve Greybush (2012)


MCS:
No Bias Correction:
With Bias Correction:
MGCM-LETKF Temperature [K] Fcst Ens Mean (contours) and 6-hour Forecast Bias (shaded)
TES:
Bias (Obs – Fcst)
Bias (Obs – Fcst)
Pseudo-Pressure (hPa)
Pressure (Pa)
Latitude

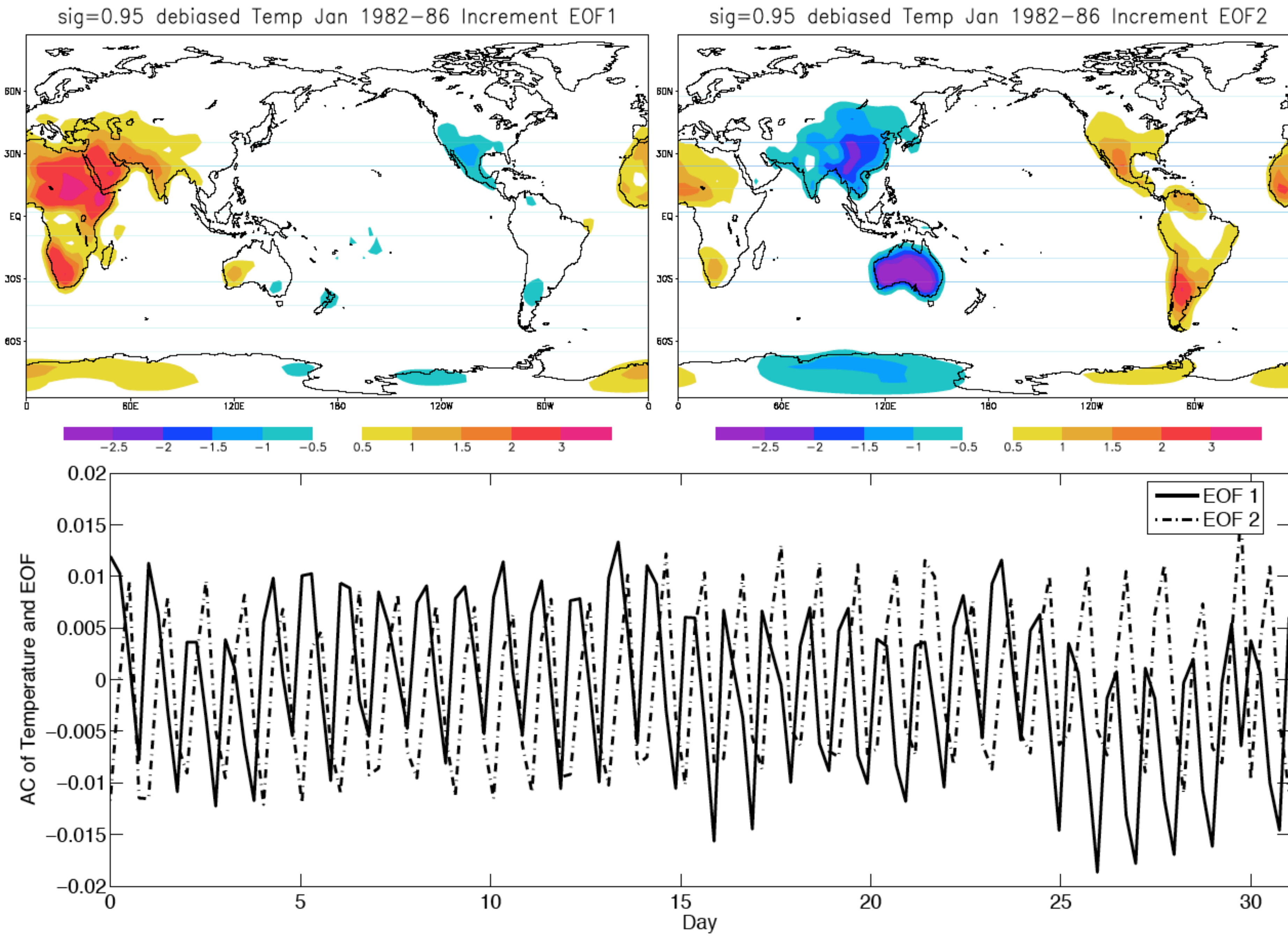
Mars TES/LETKF Performance



Example: How to define the diurnal model errors using EOFs from a Reanalysis (Danforth et al., 2007)

Estimated the average SPEEDY model error (bias) by averaging over several years the 6 hour forecast (started from reanalysis) minus the reanalysis.

Then they computed the EOFs of the anomaly in the model error, and found two dominant EOFs representing the model error in representing the diurnal cycle:

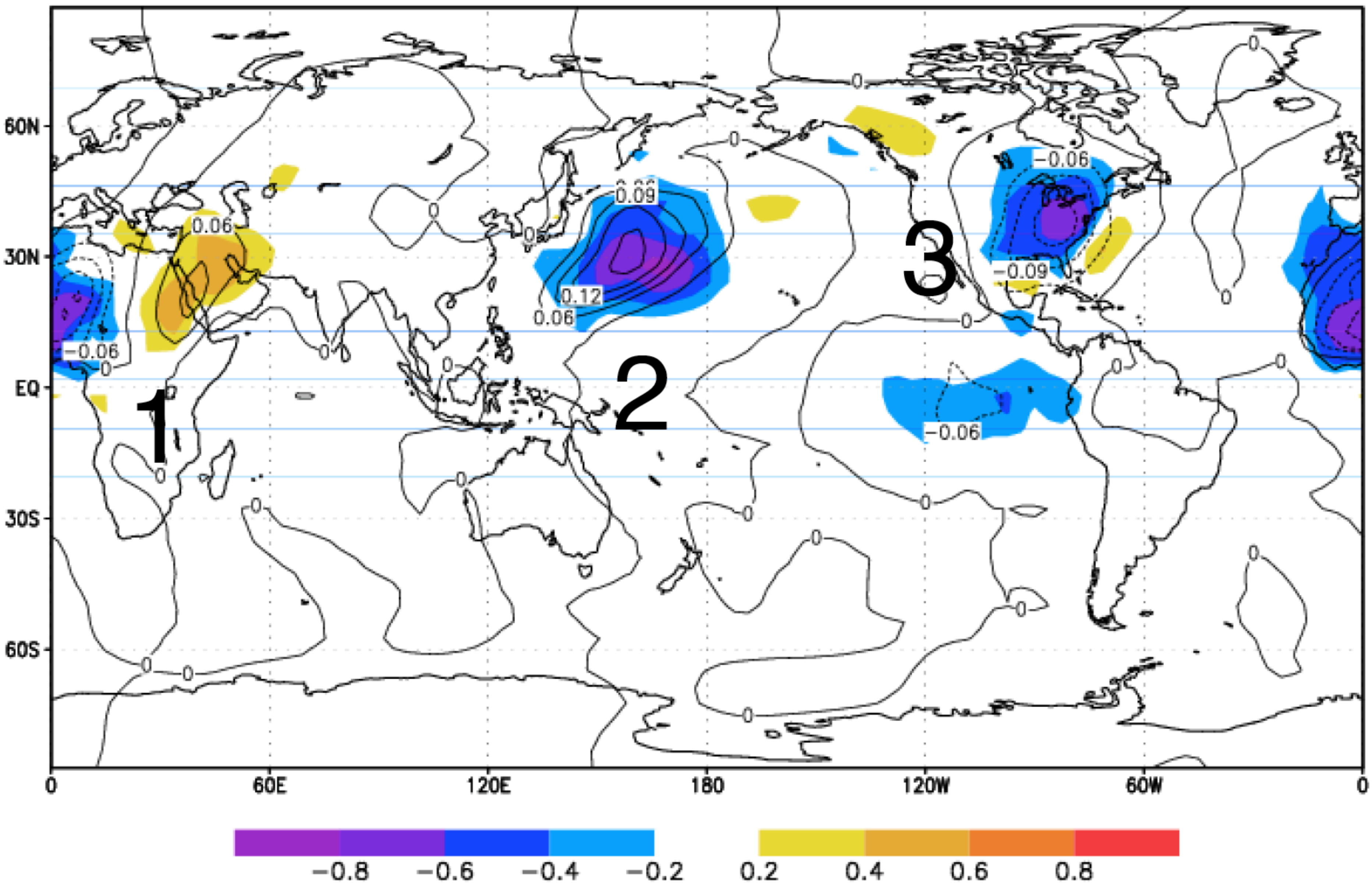


sig=0.95 debiased Temp Jan 1982−86 Increment EOF1
sig=0.95 debiased Temp Jan 1982−86 Increment EOF2
60N
30N
EQ
30S
60S
0
60E
120E
180
120W
60W
−2.5
−2
−1.5
−1
−0.5
0.5
1
1.5
2
3
EOF 1
EOF 2
AC of Temperature and EOF
Day
0.02
0.015
0.01
0.005
0
−0.005
−0.01
−0.015
−0.02
5
10
15
20
25
30

Example: How to find the state dependent errors using coupled SVD's (from Danforth et al., 2007)

Three leading coupled SVD's of the covariance of 6 hr forecast errors and corresponding model state anomaly for T at sigma=0.95. Contours: state anomaly, colors: heterogeneous correlation with forecast errors. Note that over land, the corrections suggest the anomalous temperatures are too strong, and over ocean too weak and too far to the west.

This can be extended to improving forecasts using coupled SVD's sig=0.95 Temp Jan 1982-86 Correlation Maps



4) Truly coupled data assimilation: the ocean and the atmosphere DA should be coupled

- We used to have atmospheric and oceanic models that were coupled one-way: the atmosphere could see the ocean SST, but could not change it; the ocean could see the atmospheric fluxes, but could not change it.
- Until we had the first coupled ocean-atmosphere model, we could not predict coupled phenomena, like El Niño!

DA should also be fully coupled!

Tamara Singleton’s thesis

Data Assimilation Experiments with a Simple Coupled Ocean-Atmosphere Model

Questions she addressed:

-- Which is more accurate: 4D-Var or EnKF?

-- Is it better to do an ocean reanalysis separately, or as a single coupled system?

-- ECCO is a version of 4D-Var where both the initial state and the surface fluxes are control variables. This allows ECCO to have very long windows (decades) and estimate the surface fluxes that give the best analysis.

Is ECCO the best approach for ocean reanalysis?

Answers to the Research Questions

Questions:

-- Which is more accurate: 4D-Var or EnKF?

Fully coupled EnKF (with short windows) and 4D-Var (with long windows) have about the same accuracy.

Answers to the Research Questions

Questions:

-- Which is more accurate: 4D-Var or EnKF?

Fully coupled EnKF (with short windows) and 4D-Var (with long windows) have about the same accuracy.

-- Is it better to do the ocean reanalysis separately, or as a single coupled system?

Both EnKF and 4D-Var are similar and most accurate when coupled, but uncoupled (ocean only) reanalyses are fairly good.

Answers to the Research Questions

Questions:

-- Which is more accurate: 4D-Var or EnKF?
Fully coupled EnKF (with short windows) and 4D-Var (with long windows) have about the same accuracy.

-- Is it better to do the ocean reanalysis separately, or as a single coupled system?
Both EnKF and 4D-Var are similar and most accurate when coupled, but uncoupled (ocean only) reanalyses are quite good.

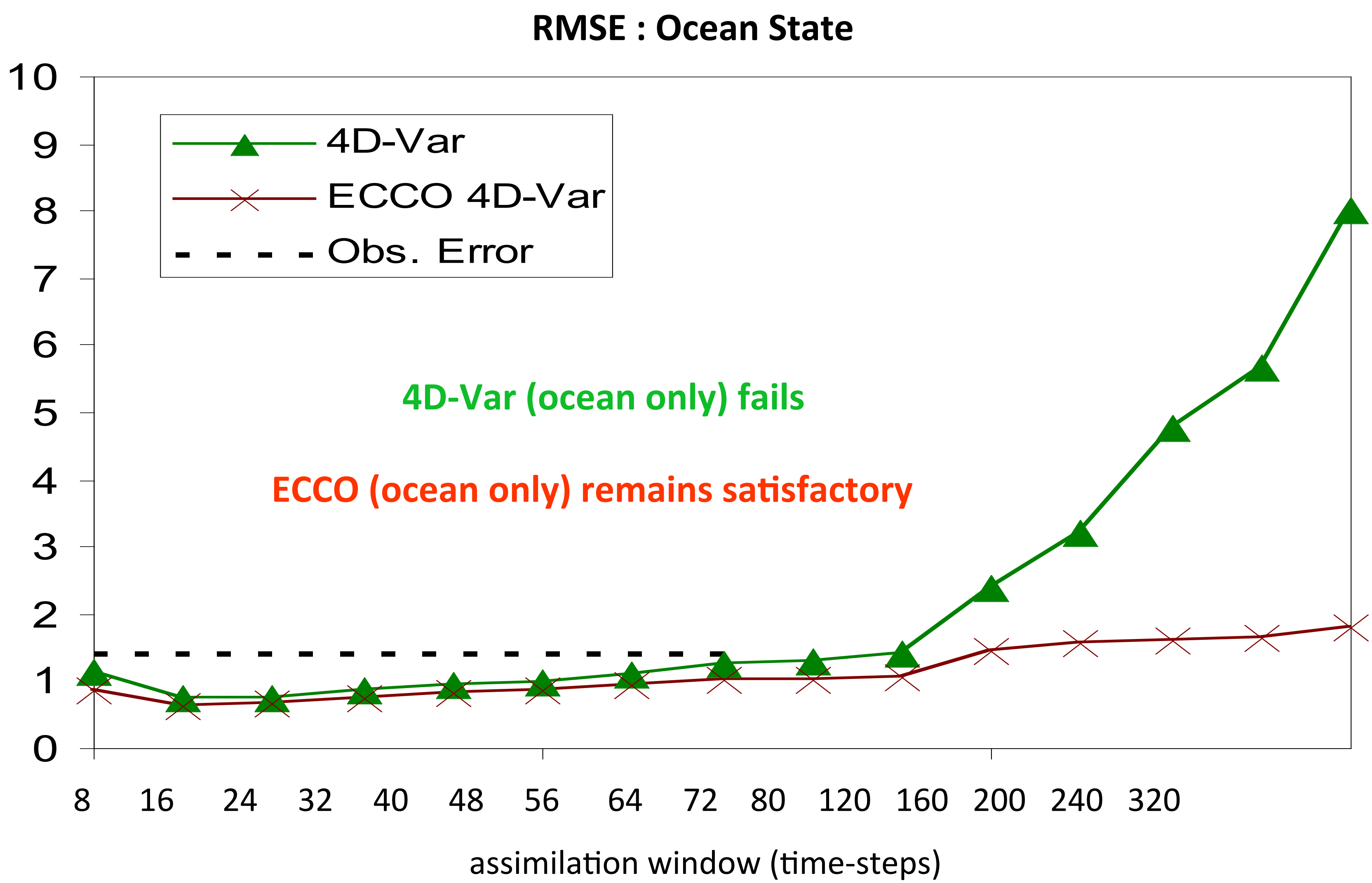
-- Is ECCO 4D-Var with both the initial state and the surface fluxes as control variables the best approach?
In our simple ocean model 4D-Var cannot remain accurate with very long windows. Our ECCO reanalysis remained satisfactory with very long windows but at the expense of less accurate fluxes.

Comparison of ECCO-like & Ocean 4D-Var

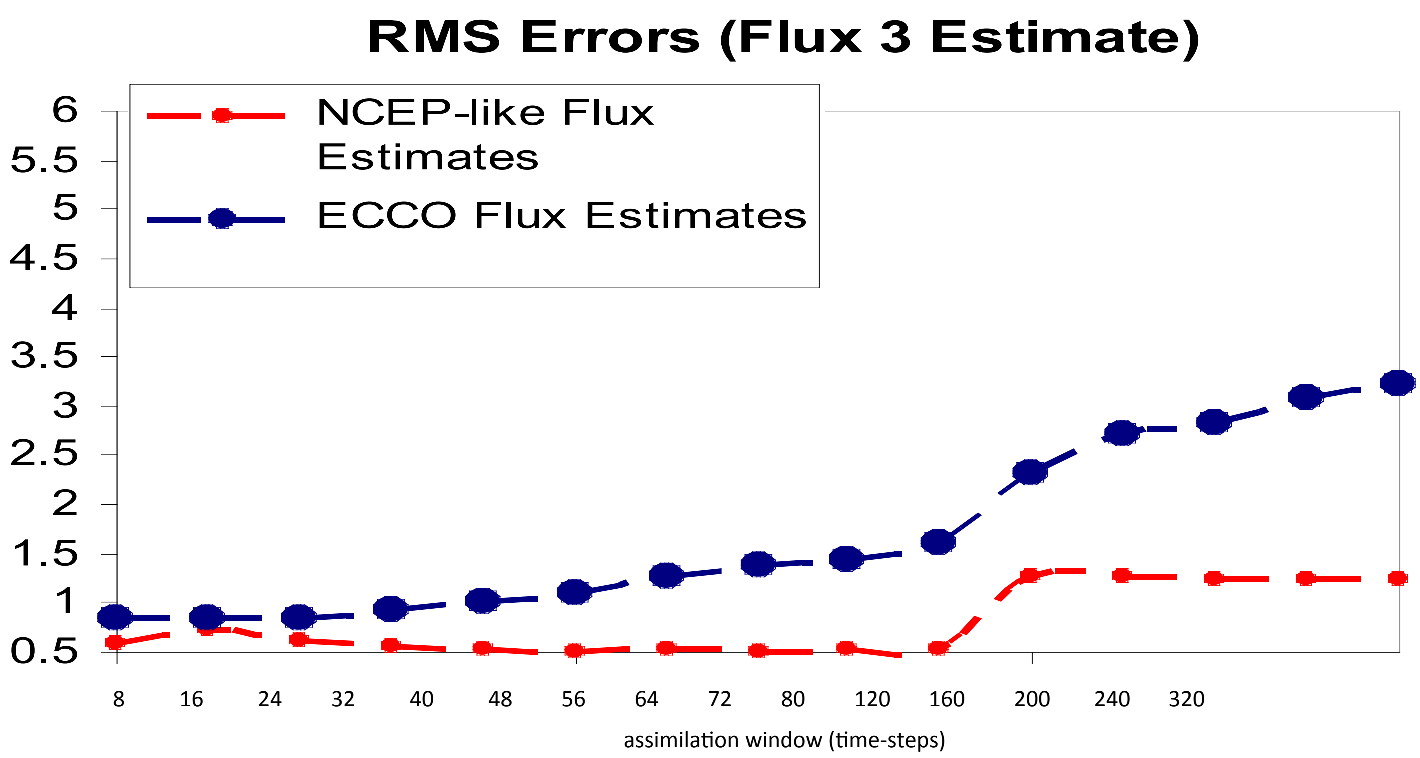
QVA APPLIED

OCEAN ONLY

Obs. s.d error = 1.41 for ocean

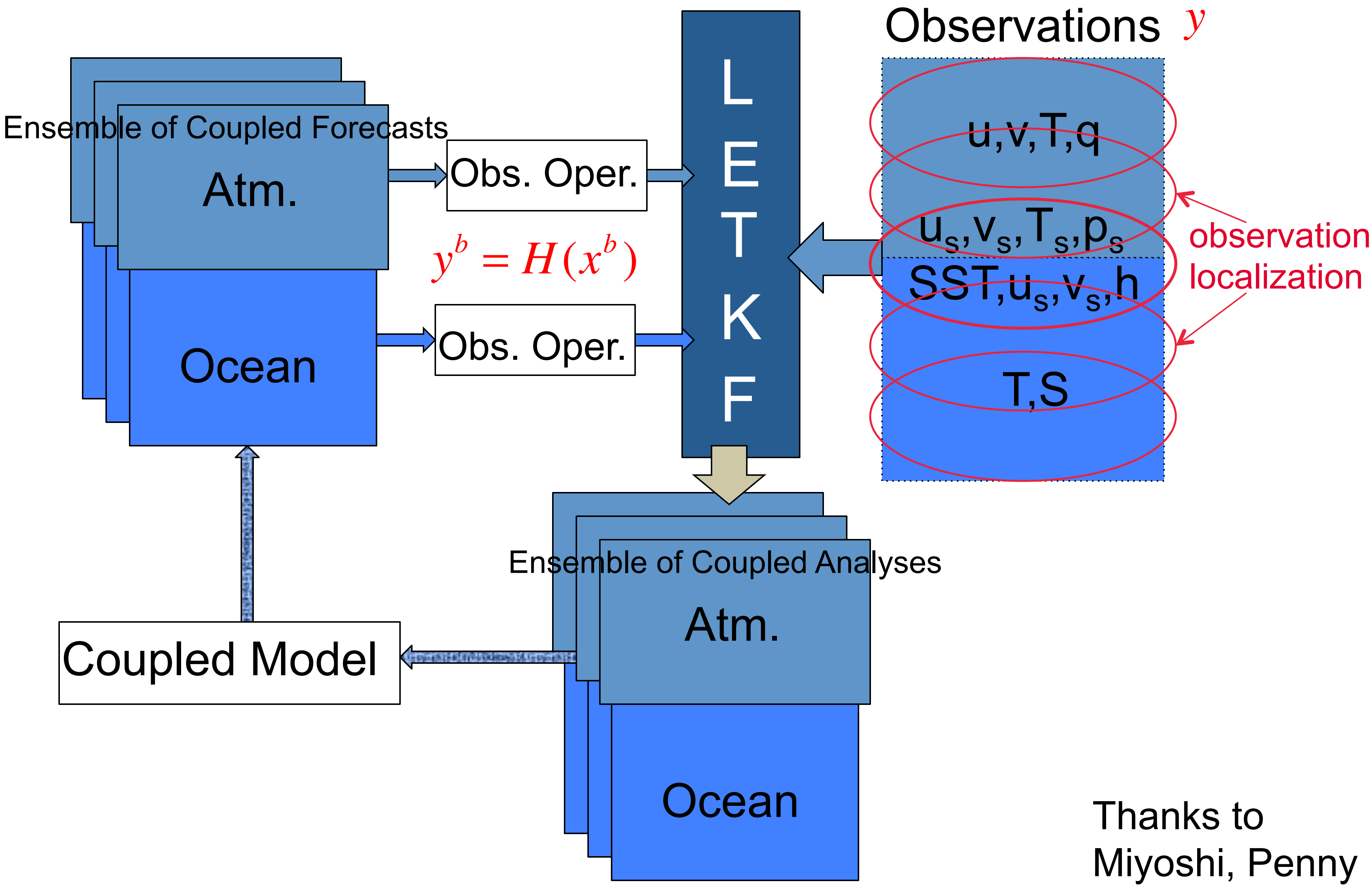


Are the ECCO fluxes more accurate?

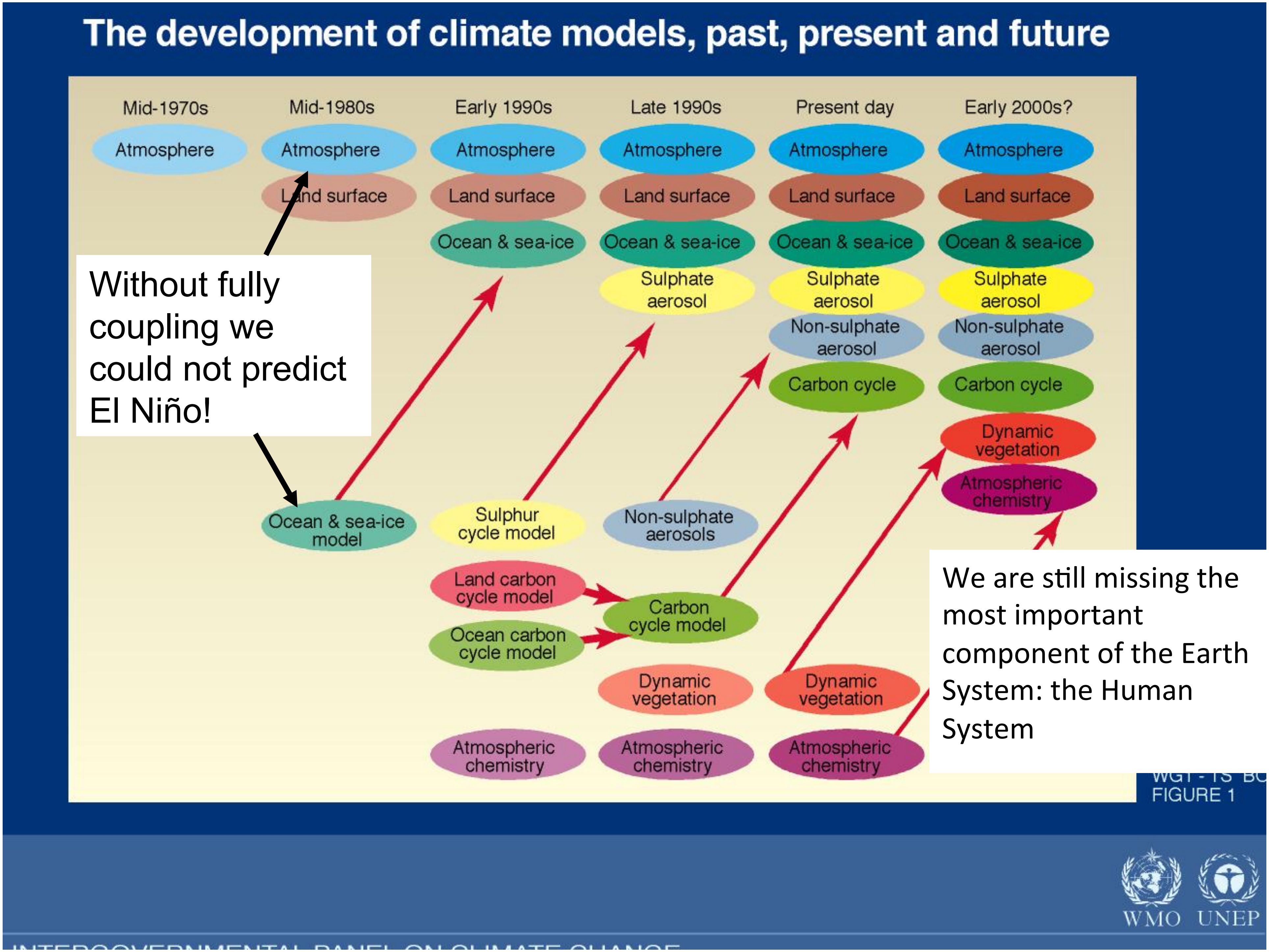


ECCO does not improve the flux estimates

Basic idea for our coupled LETKF assimilation



Thanks to
Miyoshi, Penny

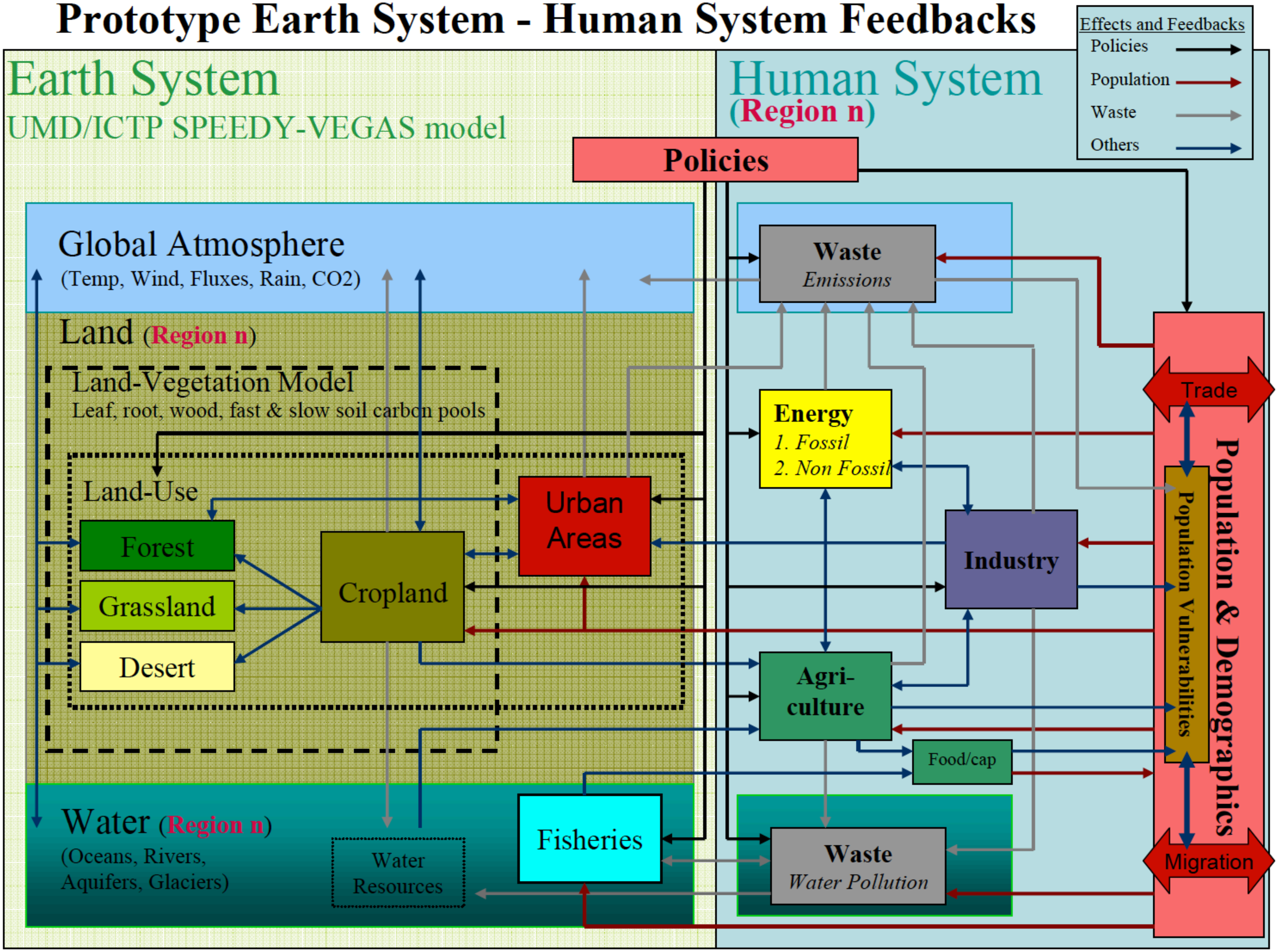


The development of climate models, past, present and future
Mid-1970s
Mid-1980s
Early 1990s
Late 1990s
Present day
Early 2000s?
Atmosphere
Atmosphere
Land surface
Ocean & sea-ice model
Atmosphere
Land surface
Ocean & sea-ice
Sulphur cycle model
Land carbon cycle model
Ocean carbon cycle model
Atmospheric chemistry
Atmosphere
Land surface
Ocean & sea-ice
Sulphate aerosol
Non-sulphate aerosols
Carbon cycle model
Dynamic vegetation
Atmospheric chemistry
Atmosphere
Land surface
Ocean & sea-ice
Sulphate aerosol
Non-sulphate aerosol
Carbon cycle
Dynamic vegetation
Atmospheric chemistry
Atmosphere
Land surface
Ocean & sea-ice
Sulphate aerosol
Non-sulphate aerosol
Carbon cycle
Dynamic vegetation
Atmospheric chemistry
Without fully coupling we could not predict El Niño!
We are still missing the most important component of the Earth System: the Human System
FIGURE 1
WMO
UNEP

5) Earth and Human System

- The Earth System is completely dominated by the Human System.
- In order to understand their interactions we need to couple them bidirectionally, i.e., with feedbacks.
- Currently, IPCC models and even Integrated Assessment models don't include population: it is exogenously obtained from UN estimates.

Prototype Earth System - Human System Feedbacks



Human **and N**ature **Dy**namical model (HANDY) with Rich and Poor: for thought experiments

Just 4 equations!

Total population: Elite + Commoners $x = x_E + x_C$

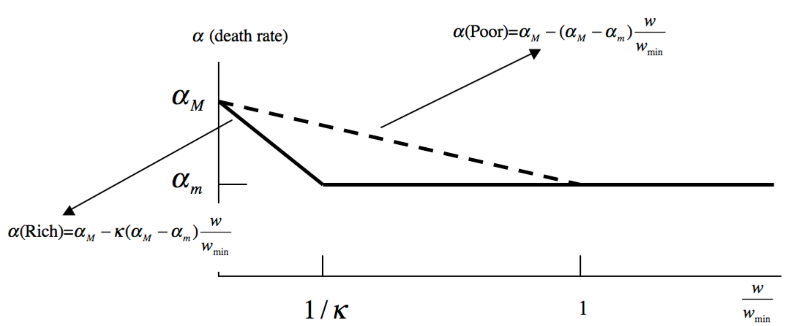
Nature equation: (only the Commoners produce)

$$\dot{y} = \text{Regeneration}\,\gamma\, y(\lambda - y) - \text{Production}\,\delta\, x_C y$$

The Wealth is managed by the Elites: Inequality factor $\kappa \sim 100$

$$\dot{W} = \text{Production-Commoner consumption-Elite consumption} = \delta x_C y - s x_C - \kappa s x_E$$

Population equations: death rate α depends on whether there is enough food:

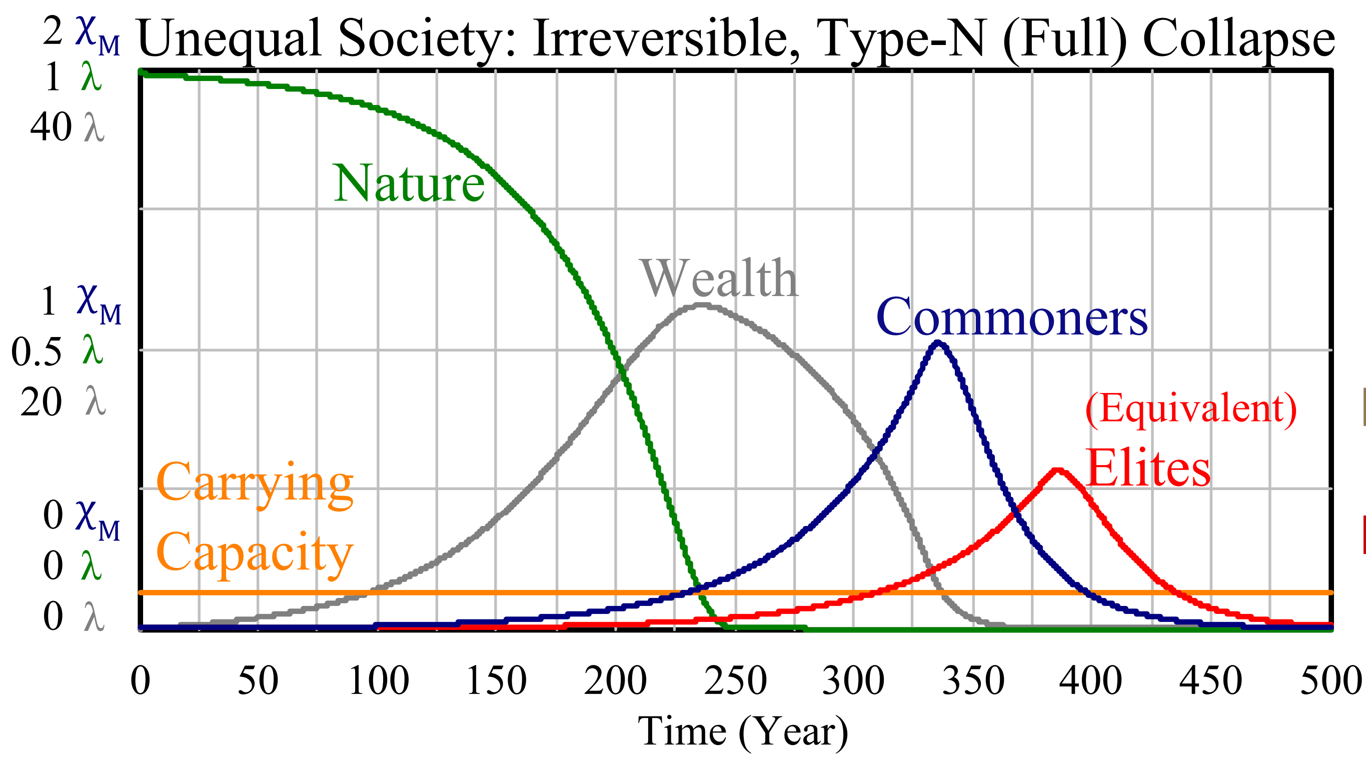


$$\dot{x}_C = -\alpha_C x_C + \beta_C x_C$$

$$\dot{x}_E = -\alpha_E x_E + \beta_E x_E$$

The **rich Elite** accumulates wealth from the work of everyone else (here referred to as the **Commoners**). When there is a crisis (e.g., famine) the elite can spend the accumulated wealth to buy food.

Human **a**nd **N**ature **Dy**namical model (HANDY) with Rich and Poor: a thought experiment



High Inequality

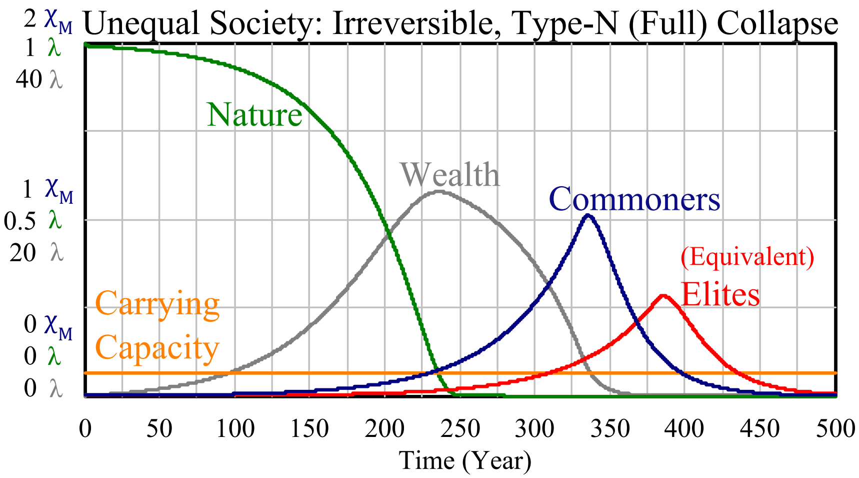
$\kappa = 100$

High Depletion

lead to collapse

The accumulated wealth starts decreasing at the time the total equivalent population crosses the Carrying Capacity. This “economic crisis” provides a very obvious indication that the population has grown beyond the sustainable level for the ecological system. If the overshoot is small, it oscillates towards equilibrium. If it is large, it leads to collapse.

Human **a**nd **N**ature **Dy**namical model (HANDY) with Rich and Poor: a thought experiment

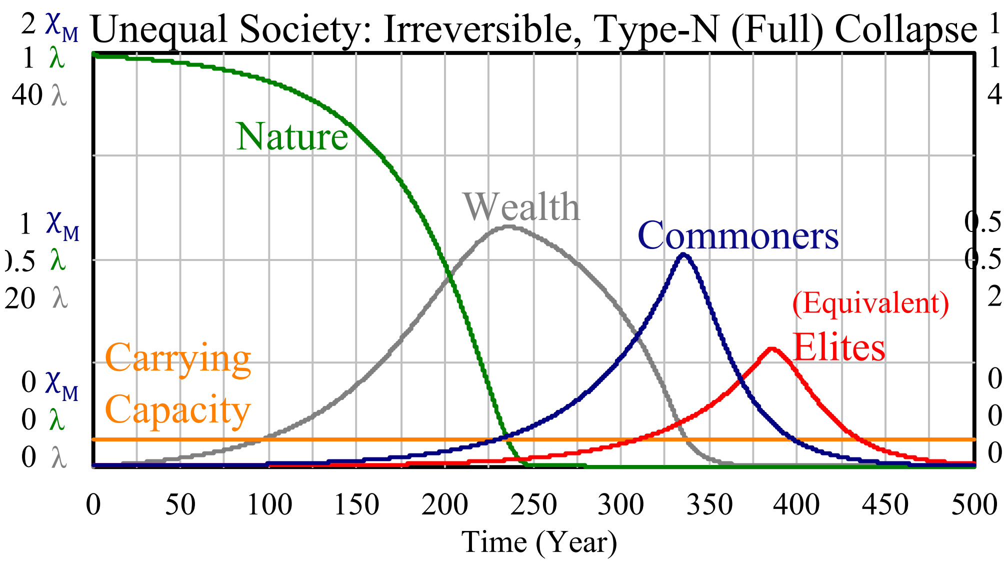


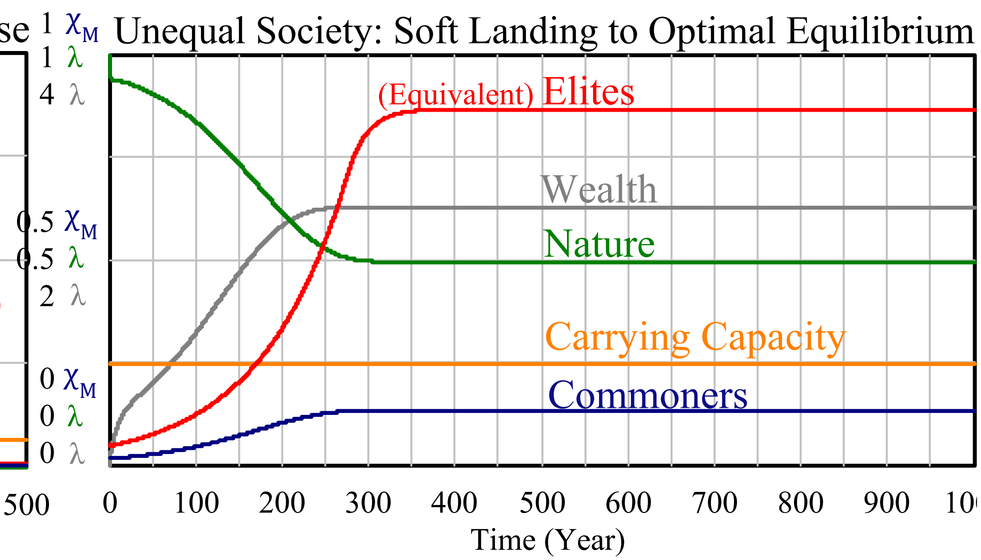
- Nature declines with population growth
- Using their wealth, the Rich can shield themselves from environmental degradation, which first affects the Poor
- Eventually it reaches the upper classes as well, when it is too late to take preventive measures

After ~250 years, having surpassed the sustainable **Carrying Capacity** of the planet, the population is drawing down the accumulated capital to survive

This thought experiment shows how a crisis can happen rapidly, even though it appears that population is rising steadily without any problems, and that the wealthy would not feel the effects of the collapse until it is too late for the poor (and then it is too late for the rich as well!).

If we reduce the *depletion per capita* to its optimal value and the *inequality* ($\kappa = 10$) it is possible to reach a steady state and survive well





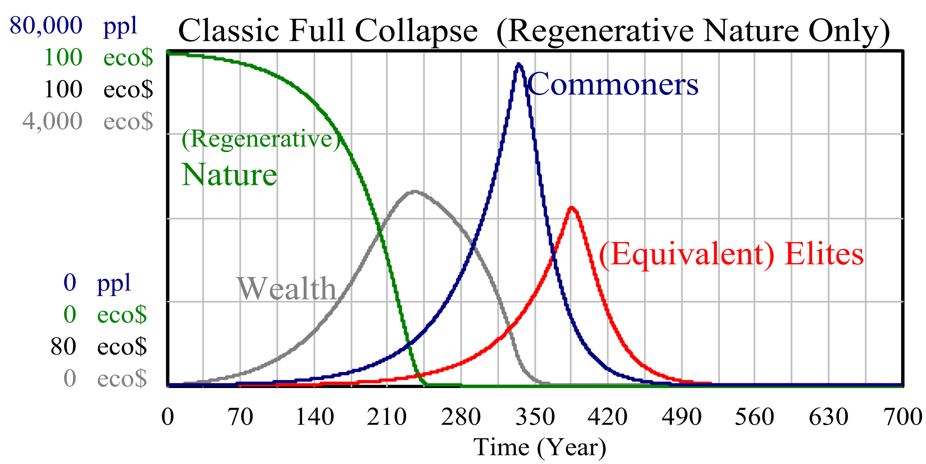
Reaching this equilibrium required **changes in policies:**

- Reduce depletion per capita
- Reduce inequality ($\kappa = 10$)
- Reduce birth rates

http://www.sciencedirect.com/science/article/pii/S0921800914000615

Journal of Ecological Economics

Consider the impact of adding fossil fuels, i.e., nonrenewable energy to Nature



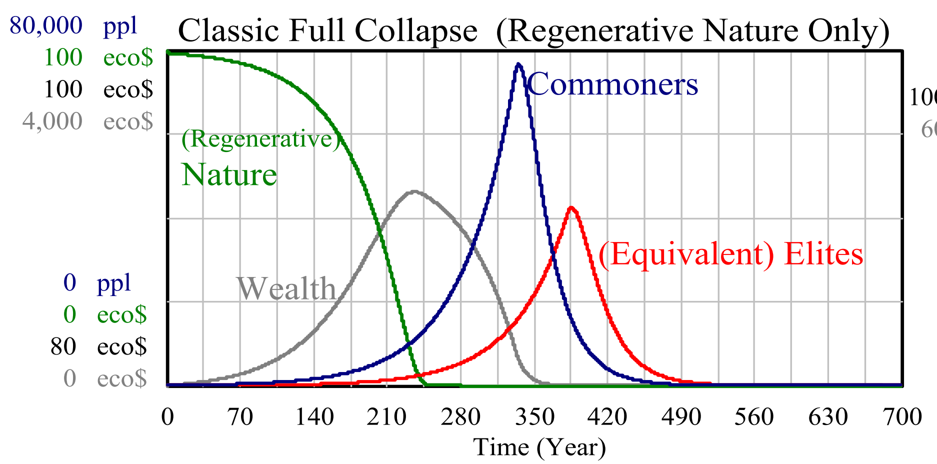
What happens when we add fossil fuels?

This is the classic HANDY1 full collapse scenario, with only regenerating Nature

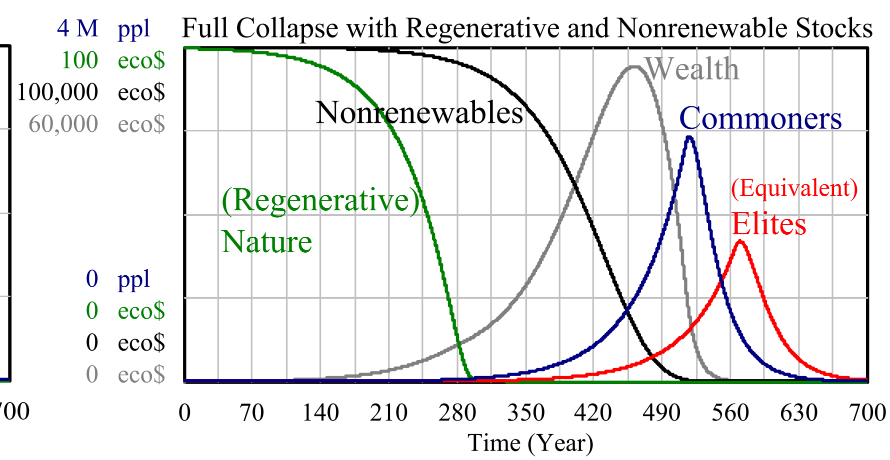
We then add to the regenerating Nature a **nonrenewable Nature**

Impact of adding fossil fuel (nonrenewable) energy to Nature

80K



4Million

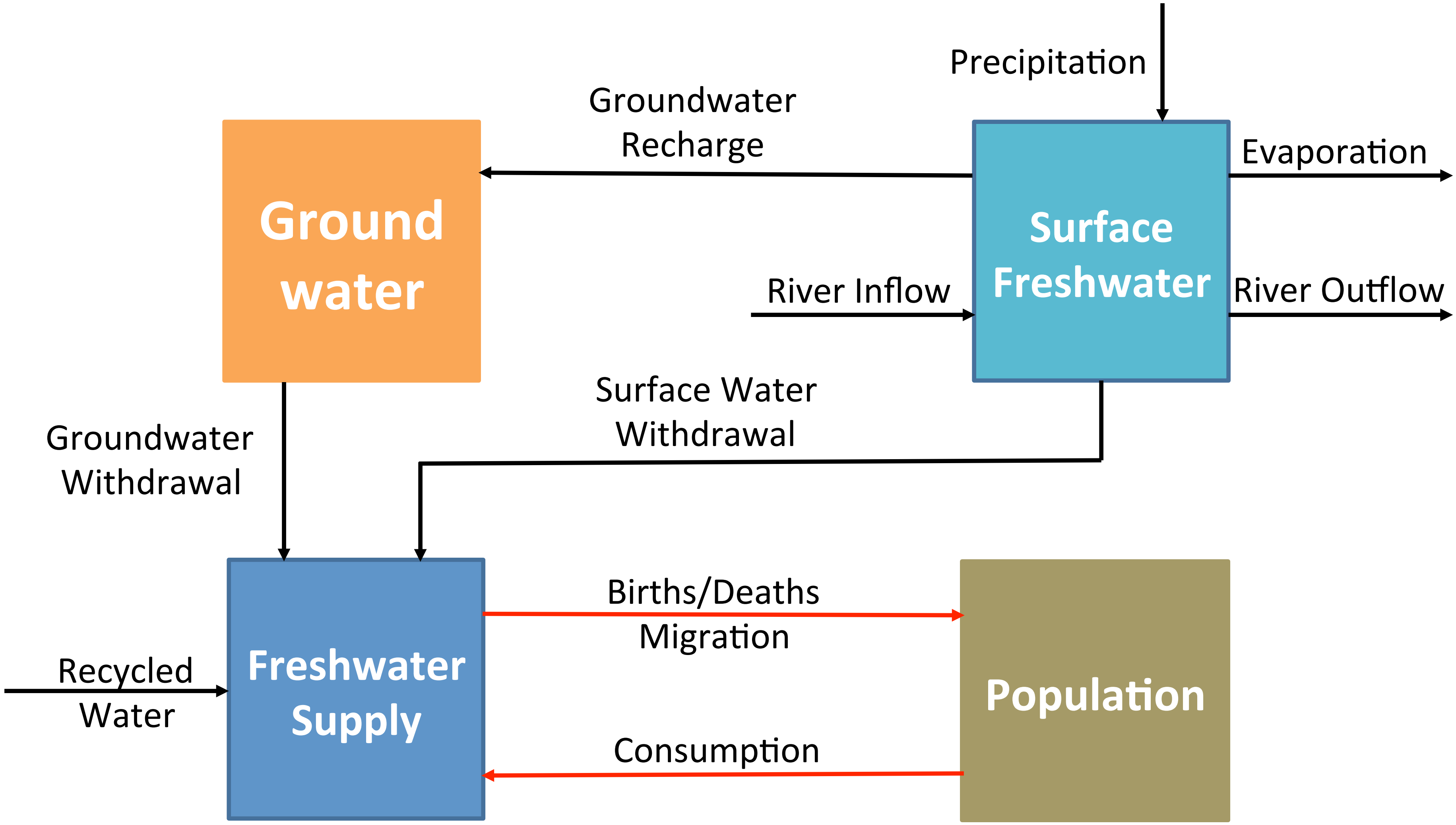


This is the classic HANDY full collapse scenario, with only regenerating Nature

We added to the regenerating Nature a **nonrenewable Nature**

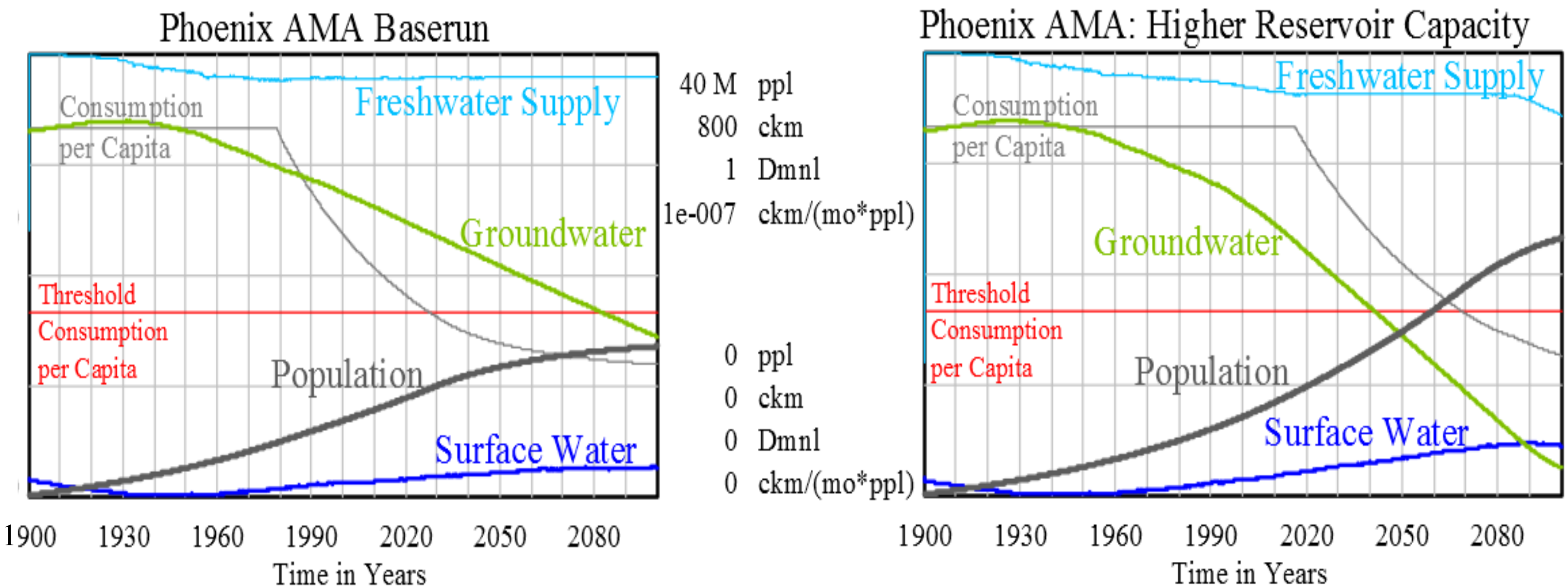
The collapse is postponed by ~200 years and the population increased by a factor of ~20!

Variables of COWA (Coupled Water model)



Applied to the Phoenix, Arizona watershed

Surprise: Double Reservoir Capacity …



- By doubling the reservoir capacity to $z_M = 10$, population can grow to a maximum of 23M, compared to 14M.
- However, groundwater ends up at a very low level of 50 ckm. Stricter groundwater withdrawal policies can prevent this.
- Without coupling the population we would not get this result.
- We should fit the observations and obtain parameters with EnKF.

SUMMARY

- Future applications of EnKF
  - 1) Combine model forecast and observations to create the best initial conditions ✓
  - 2) Improve observations
  - 3) Improve models
  - 4) Do more truly coupled data assimilation
  - 5) Do coupled Earth and Human modeling and DA.