New Applications of Advanced Data Assimilation to Improve Models and Observations

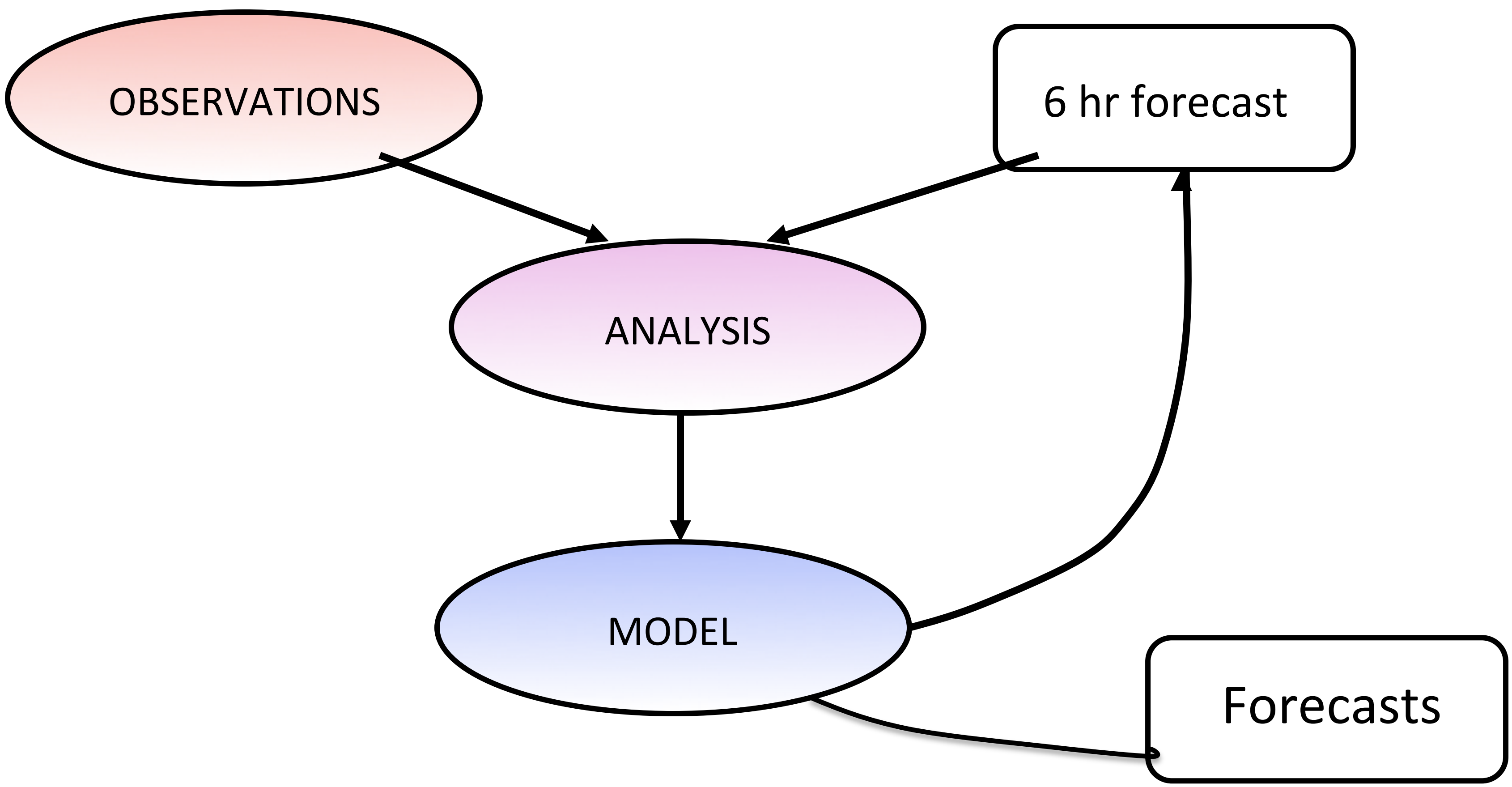
E. Kalnay[1], T. C. Chen[1], D. Hotta[2], Y. Ota[2], T. Miyoshi[3,1], Kriti Bhargava[1], J. Carton[1], F. Yang[4]

[1] UMD, [2]JMA, [3]RIKEN, [4]NCEP

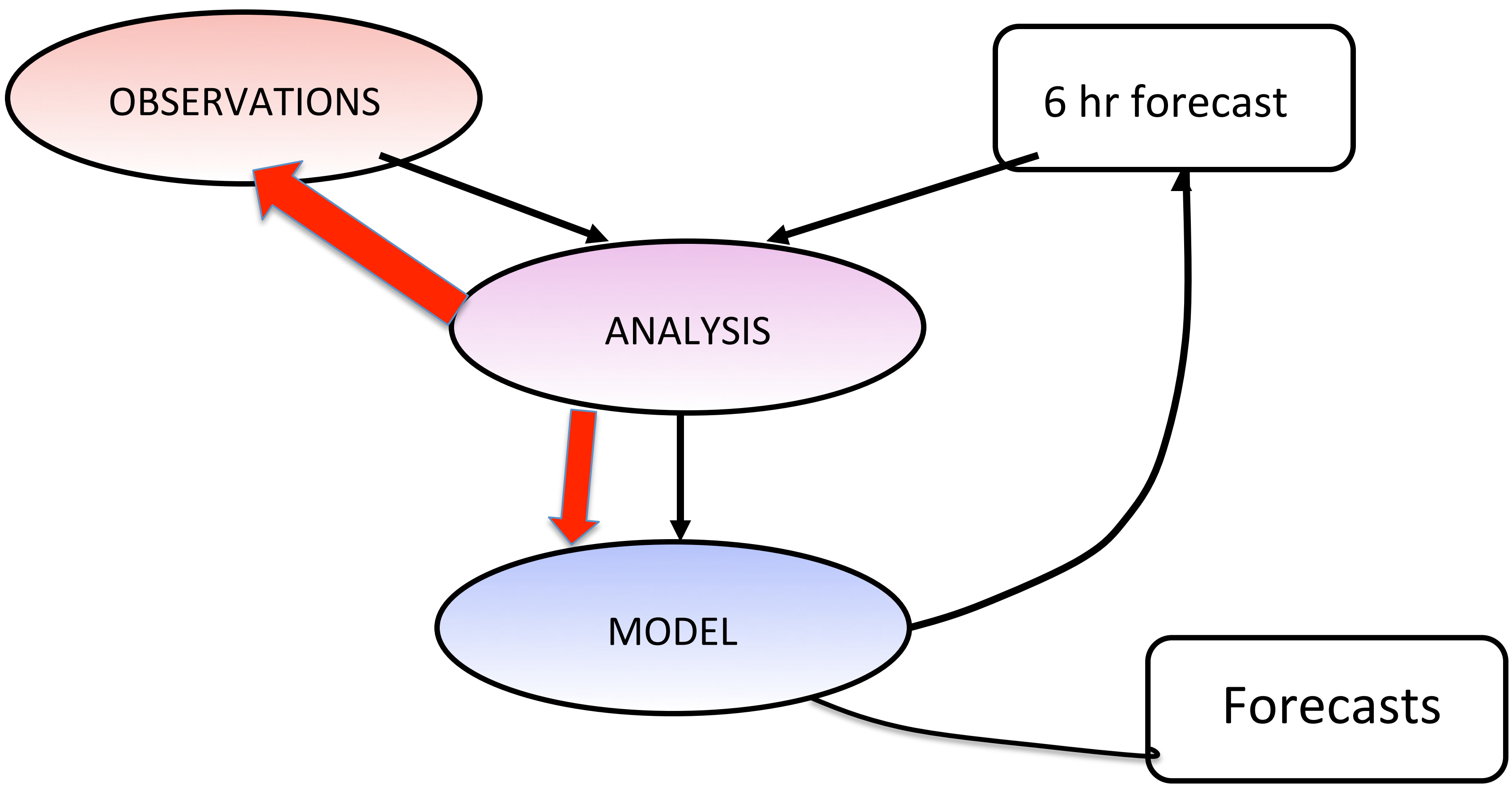
With many thanks to students, friends and colleagues
from the University of Maryland

Big Data and Environment Workshop
10-13 November 2015 – Buenos Aires

Classic Data Assimilation: For NWP we need to improve observations, analysis scheme and model



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



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

Combine optimally observations and model forecasts (mostly done! ☺)

- We should also use DA to:

Improve the observations

Improve the model

- Improve the models by parameter estimation

Example: Estimate the surface carbon fluxes as evolving parameters.

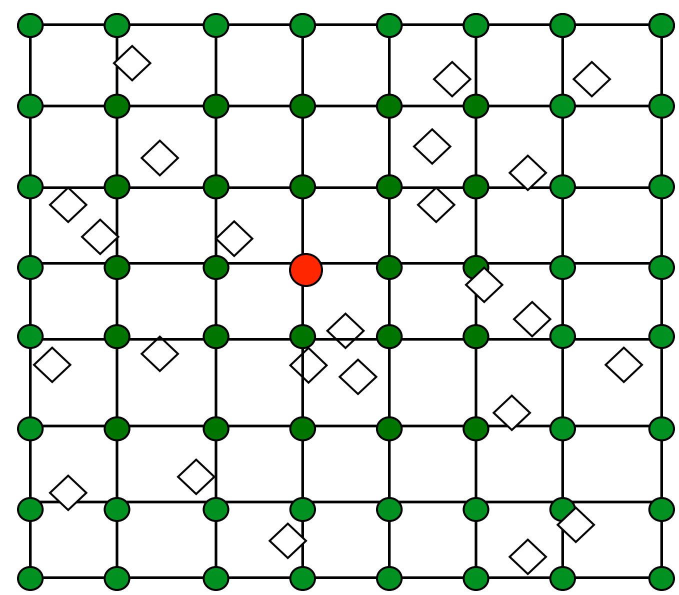
- Earth system models used by IPCC have many sub-models, but they don't include the Human System, which totally dominates the Earth system.

We should do DA of the <u>two-way coupled</u> Earth System-Human System, and use DA for parameter tuning

LETKF: Localization based on observations

Perform data assimilation in a local volume, choosing observations

The state estimate is updated at the central grid red dot

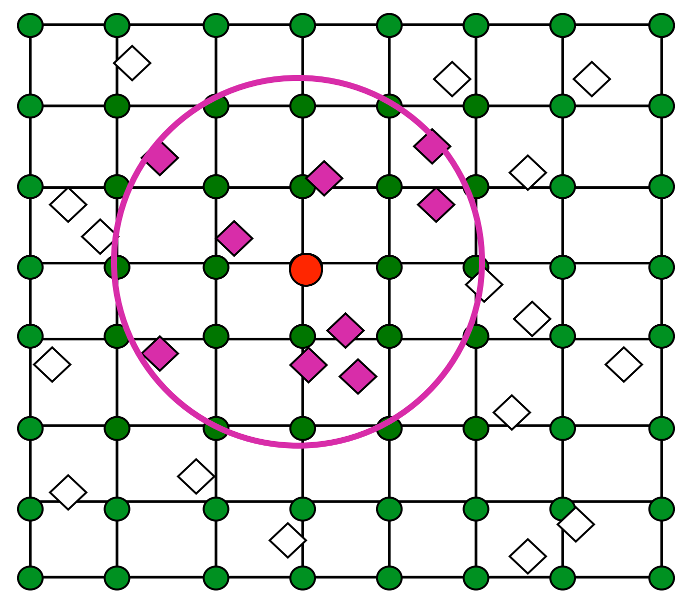


LETKF: Localization based on observations

Perform data assimilation in a local volume, choosing observations

The state estimate is updated at the central grid red dot

All observations (purple diamonds) within the local region are assimilated



The LETKF algorithm can be described in a single slide!

Local Ensemble Transform Kalman Filter (Hunt et al, 2007)

Globally:

Forecast step: $\mathbf{x}^b_{n,k} = M_n\left(\mathbf{x}^a_{n-1,k}\right)$

Analysis step: construct $\mathbf{X}^b = \left[\mathbf{x}^b_1 - \overline{\mathbf{x}}^b \mid \ldots \mid \mathbf{x}^b_K - \overline{\mathbf{x}}^b\right];$

$$\mathbf{y}^b_i = H(\mathbf{x}^b_i); \mathbf{Y}^b_n = \left[\mathbf{y}^b_1 - \overline{\mathbf{y}}^b \mid \ldots \mid \mathbf{y}^b_K - \overline{\mathbf{y}}^b\right]$$

Locally: Choose for each grid point the observations to be used, and compute the local analysis error covariance and perturbations in ensemble space:

$$\tilde{\mathbf{P}}^a = \left[(K-1)\mathbf{I} + \mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y}\right]^{-1}; \mathbf{W}^a = [(K-1)\tilde{\mathbf{P}}^a]^{1/2}$$

Analysis mean in ensemble space: $\overline{\mathbf{w}}^a = \tilde{\mathbf{P}}^a\mathbf{Y}^{bT}\mathbf{R}^{-1}(\mathbf{y}^o - \overline{\mathbf{y}}^b)$

and add to $\mathbf{W}^a$ **to get the analysis ensemble in ensemble space.**

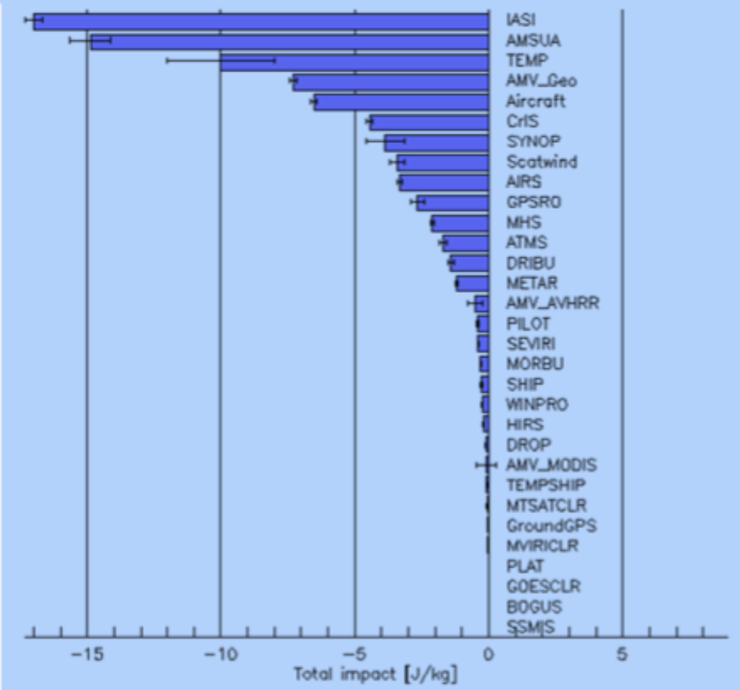
The new ensemble analyses in model space are the columns of $\mathbf{X}^a_n = \mathbf{X}^b_n\mathbf{W}^a + \overline{\mathbf{x}}^b$**. Gathering the grid point analyses forms the new global analyses. Note that the the output of the LETKF are analysis weights** $\overline{\mathbf{w}}^a$ **and perturbation analysis matrices of weights** $\mathbf{W}^a$**. These weights multiply the ensemble forecasts.**

Forecast Sensitivity to Observations (Langland and Baker, 2004)

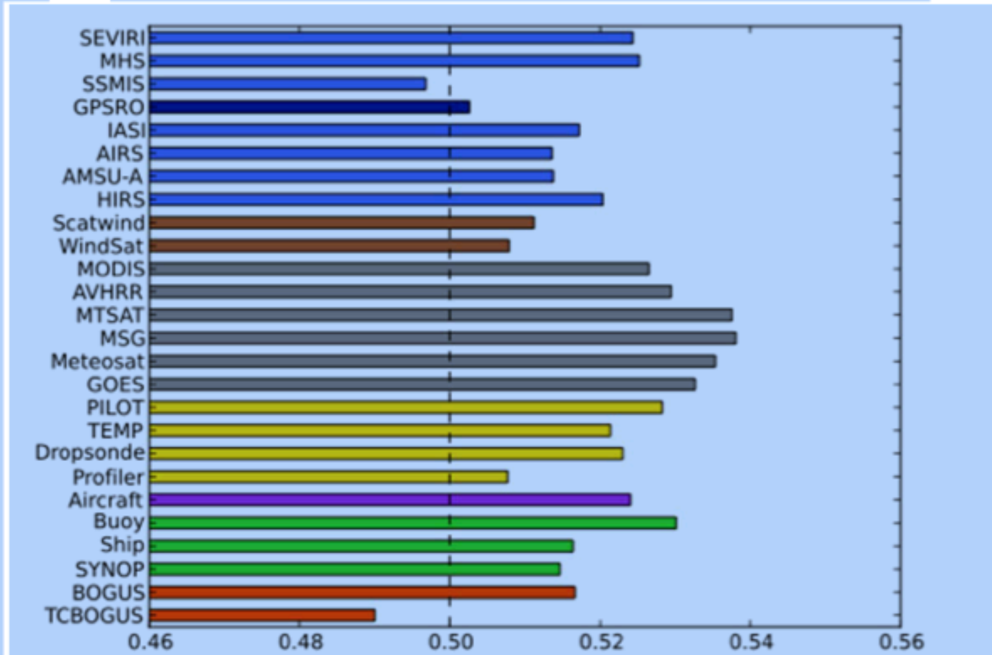
Met Office

FSOI in Global NWP

Total Observation Impact (Aug 2014)



Fraction Obs that Improve Forecast



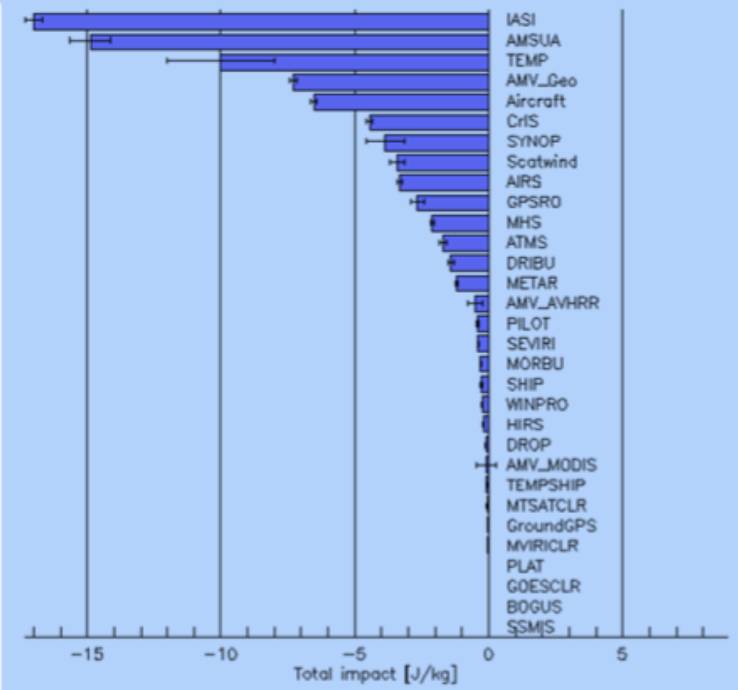
- Infra-Red (IASI) and microwave (AMSUA) radiances now biggest impact.
- Note only ~50% of observations reduce forecast error(!).
- Estimate: need 6 months time series to assess impact for single observing site.
- EFSO methodology now being considered when no adjoint available

Can we identify bad observations?

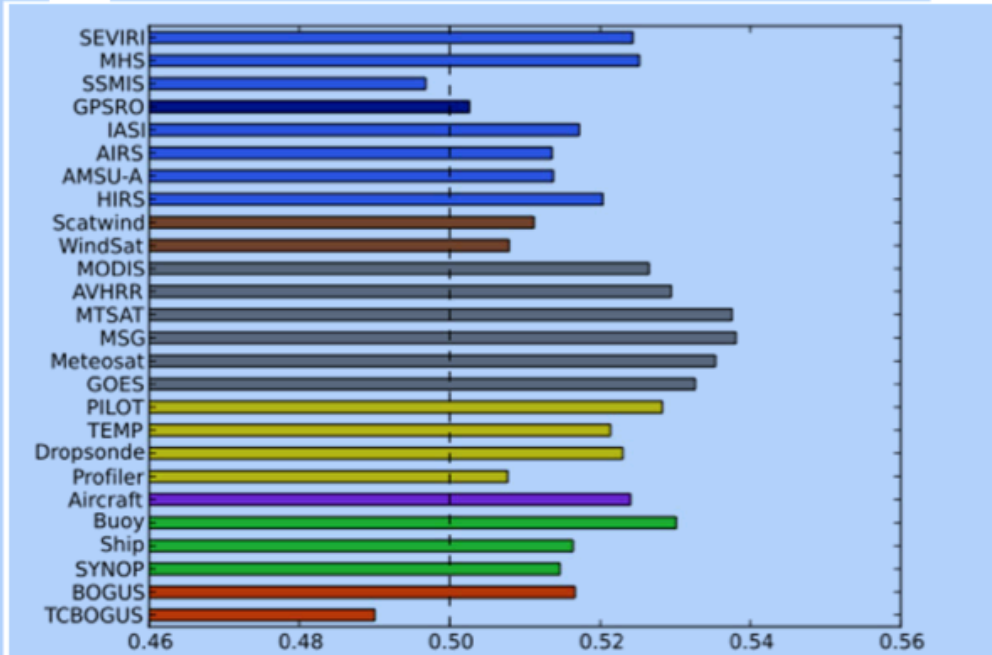
Met Office

FSOI in Global NWP

Total Observation Impact (Aug 2014)



Fraction Obs that Improve Forecast



- Infra-Red (IASI) and microwave (AMSUA) radiances now biggest impact.
- Note only ~50% of observations reduce forecast error(!).
- Estimate: need 6 months time series to assess impact for single observing site.
- EFSO methodology now being considered when no adjoint available

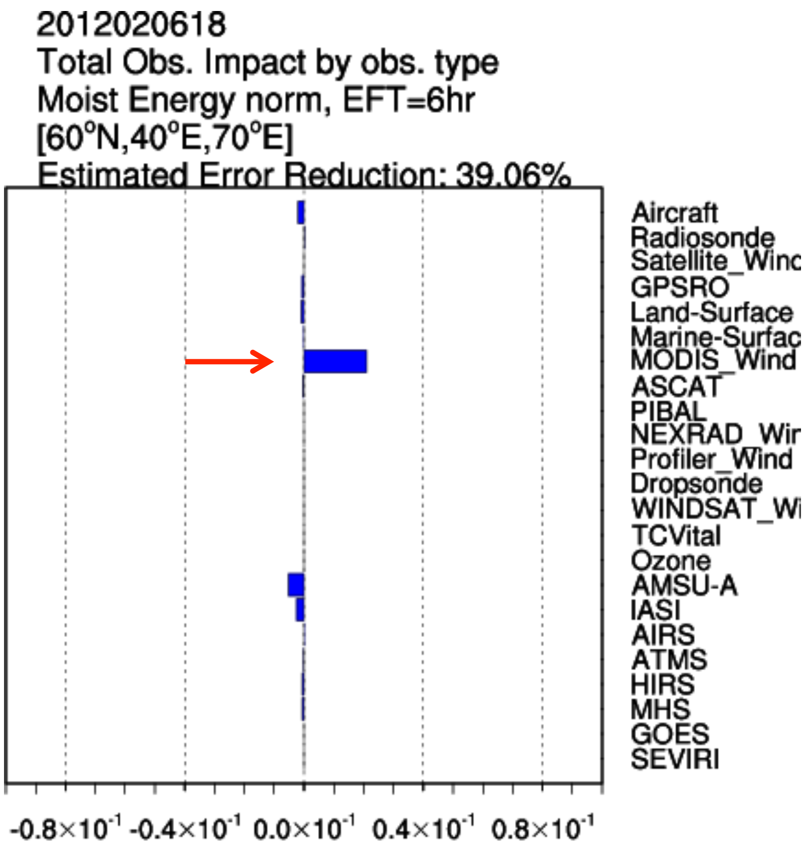
1) Improve the observations: Ensemble Forecast Sensitivity to Observations and Proactive QC

- Kalnay et al. (2012) derived EFSO.
- Ota et al. (2013) tested 24hr GFS forecasts and showed EFSO could be used to identify bad obs.
- D. Hotta (2014): EFSO can be used after only 6 hours, so that the bad obs. can be collected and withdrawn, with useful metadata, so they can be improved. The analysis is corrected with EFSO.
- We call this **Proactive QC**, much stronger than QC.
- Hotta also showed EFSO can be used to tune **R**
- Tse-Chun Chen (2015) tested impact of EFSO/PQC over 5 day forecasts: **PROMISING RESULTS**

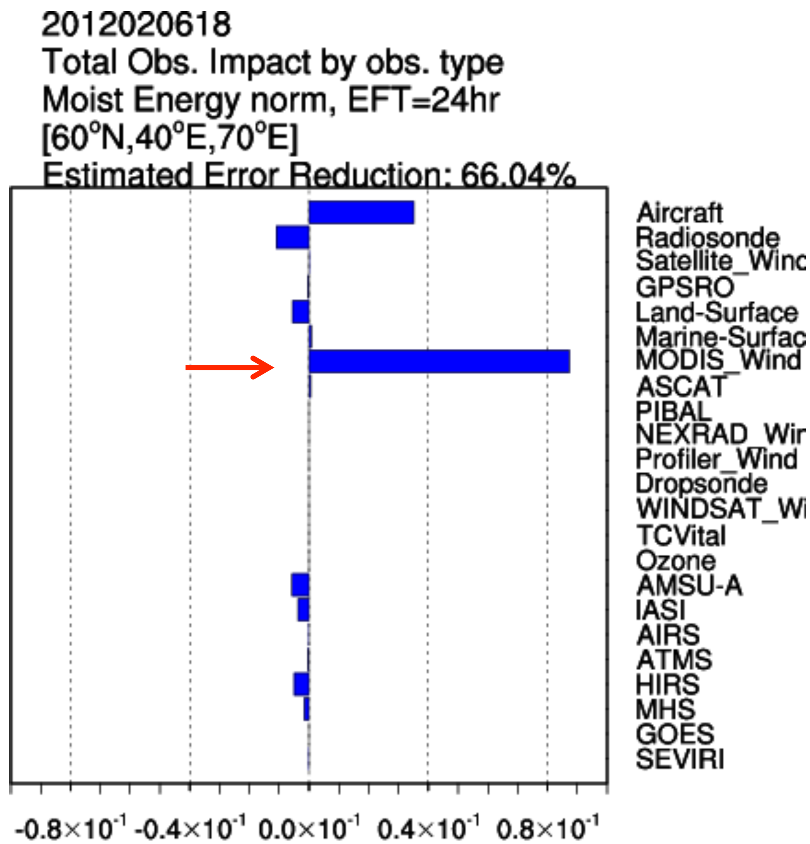
Hotta (2014)

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

FT=06 hr.



FT=24 hr.

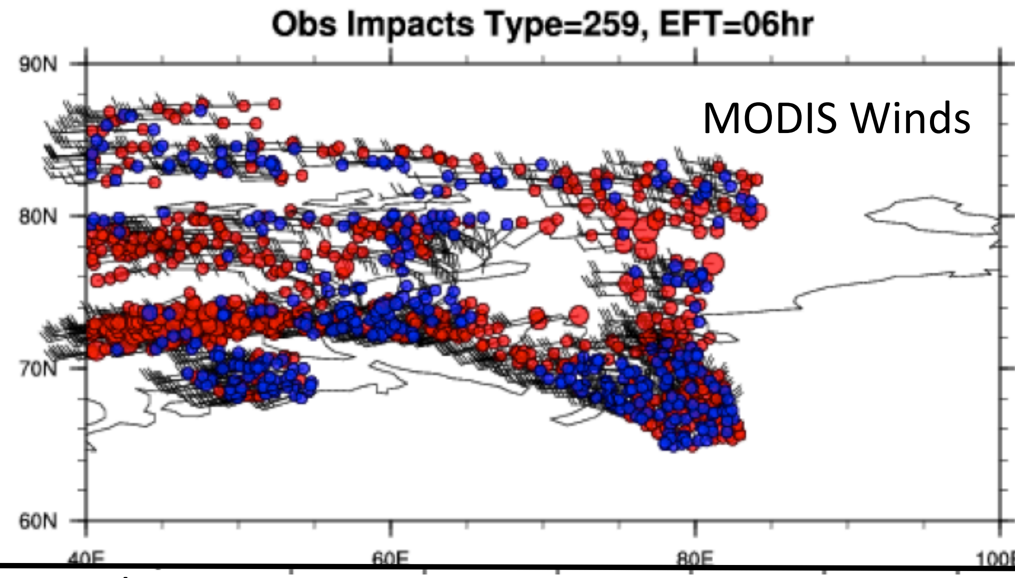


Can identify the bad observations after only 6 hours!

Improve observations:

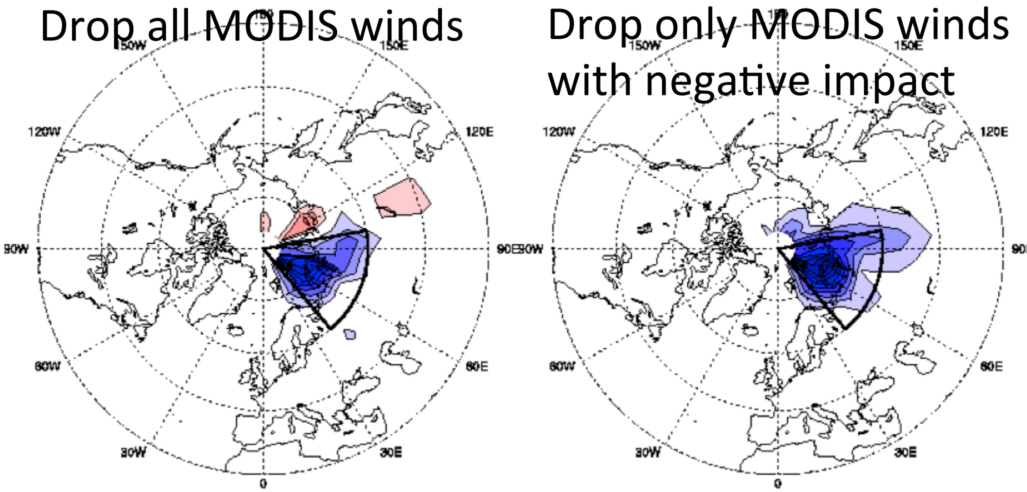
Proactive QC: Find and delete the obs that make the 6hr forecast worse using EFSO

Dr. Daisuke Hotta (2014):
EFSO is able to find whether <u>each</u> observation improves (blue) or makes the 6hr forecast worse (red)



Drop all MODIS winds

Drop only MODIS winds with negative impact



Impact of 6hr PQC on 24hr fcst

PQC with metadata can be used to improve the algorithm!

It should accelerate optimal assimilation of new instruments!

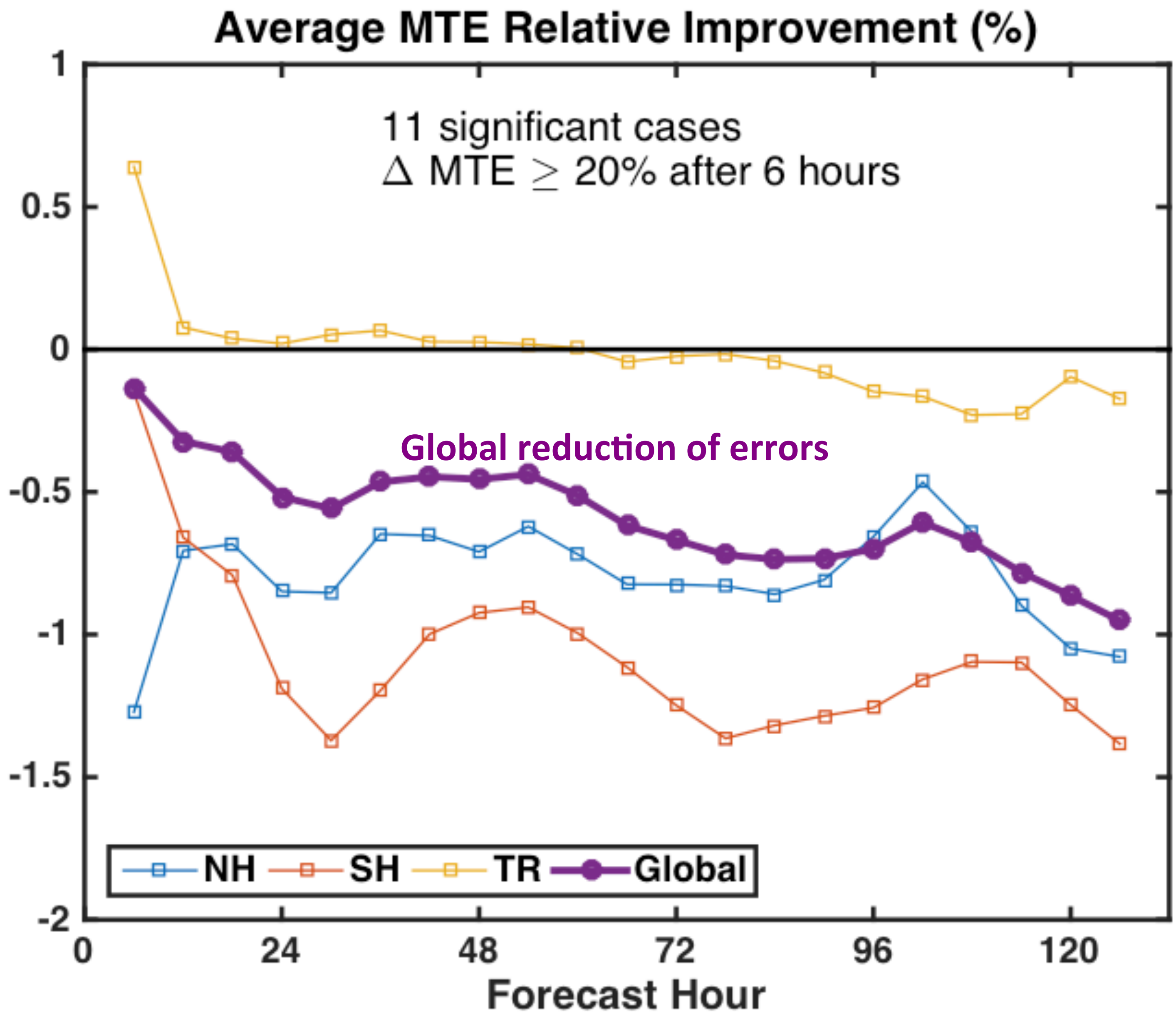
In one month: 20 cases of skill dropout due to flawed observations that passed the operational QC

Tse-Chun Chen: classify the 20 cases into

- 11 SIGNIFICANT cases, where EFSO estimates that withdrawing the flawed observations reduces the 6hour forecast error by more than 20% (in Total Moist Energy)
  - **ΔTME>20%**
- 9 NON-SIGNIFICANT CASES, with **ΔTME<20%**

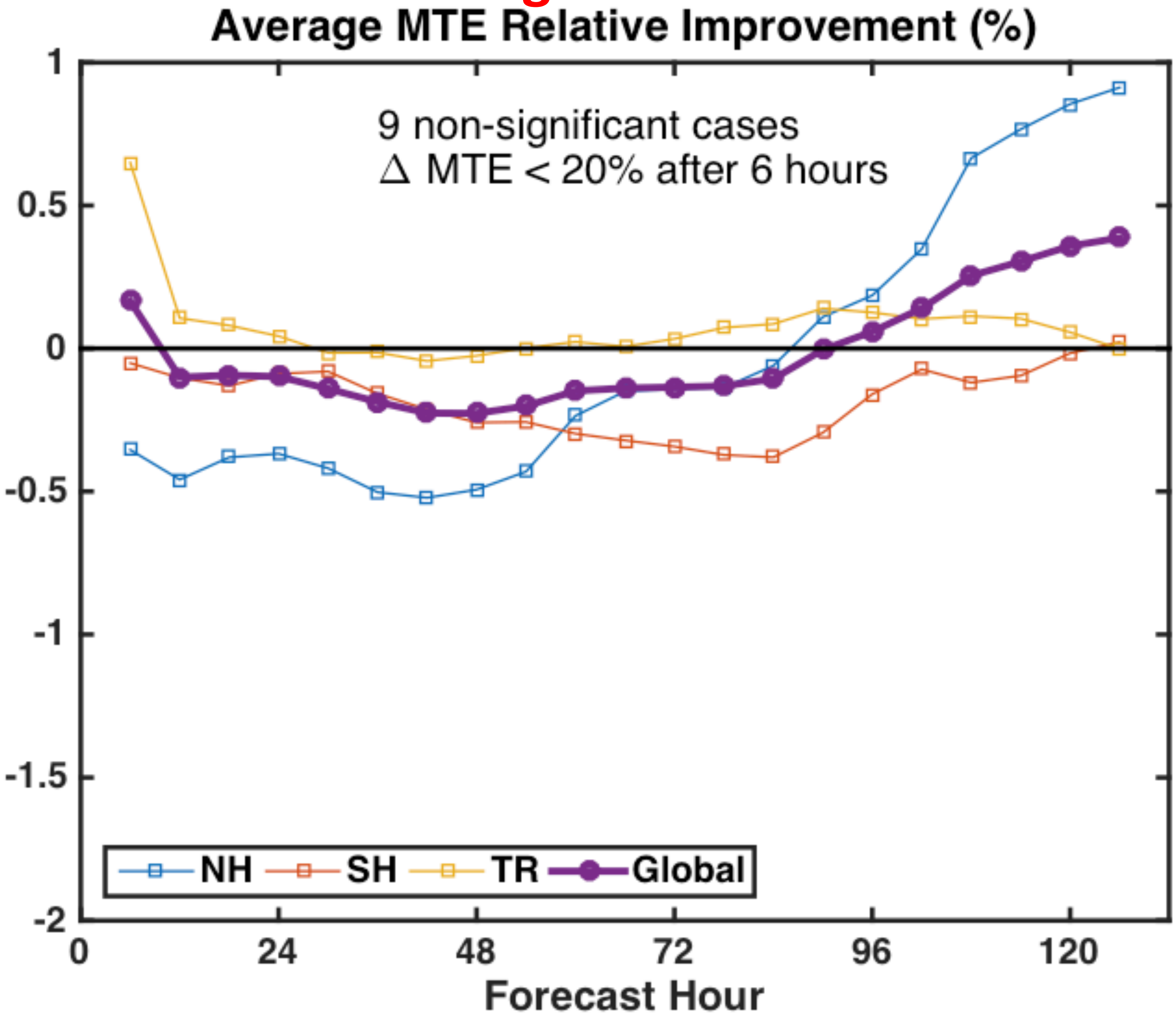
5-day reduction of total moist energy of the forecast error

11 significant cases



5-day reduction of total moist energy of the forecast error

9 non-significant cases



2) Ensemble Forecast Sensitivity to Error Covariances

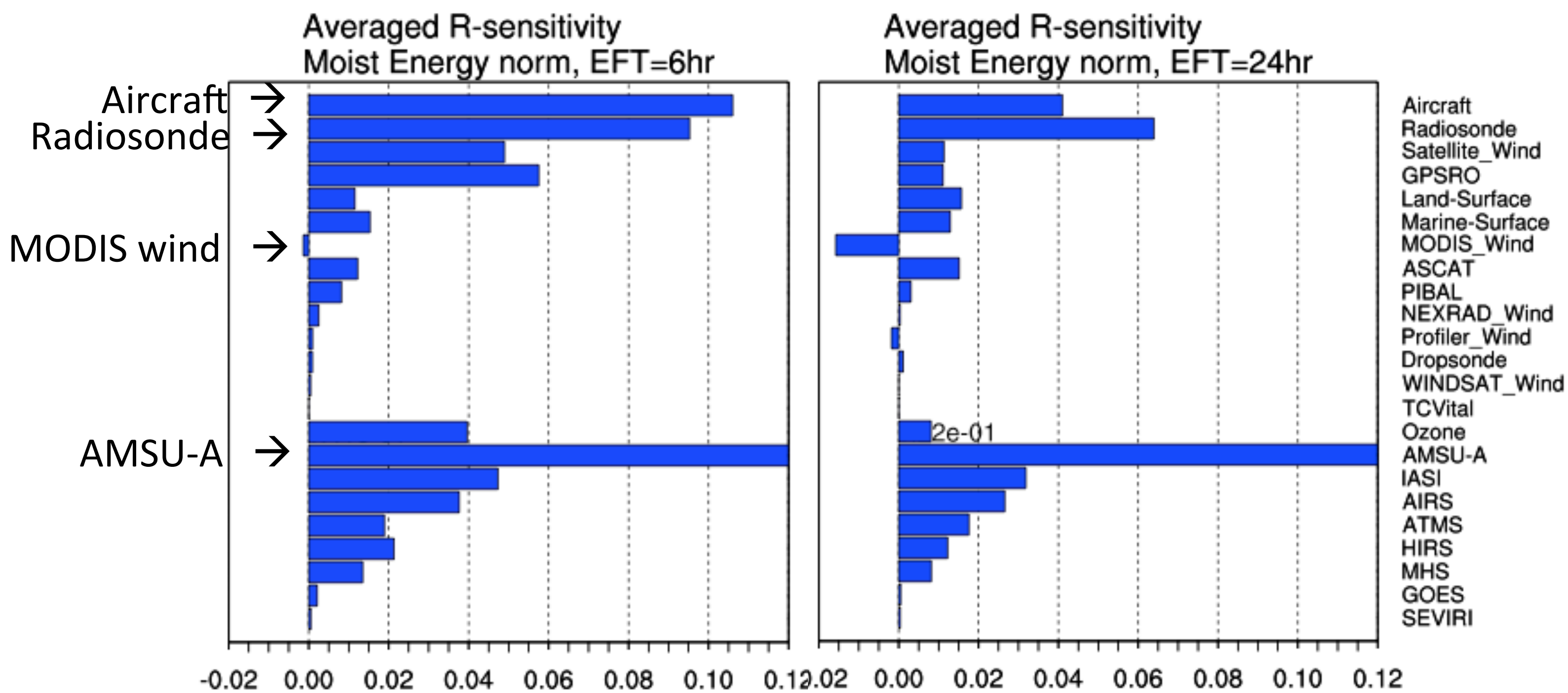
Hotta (2014)

- 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) :

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

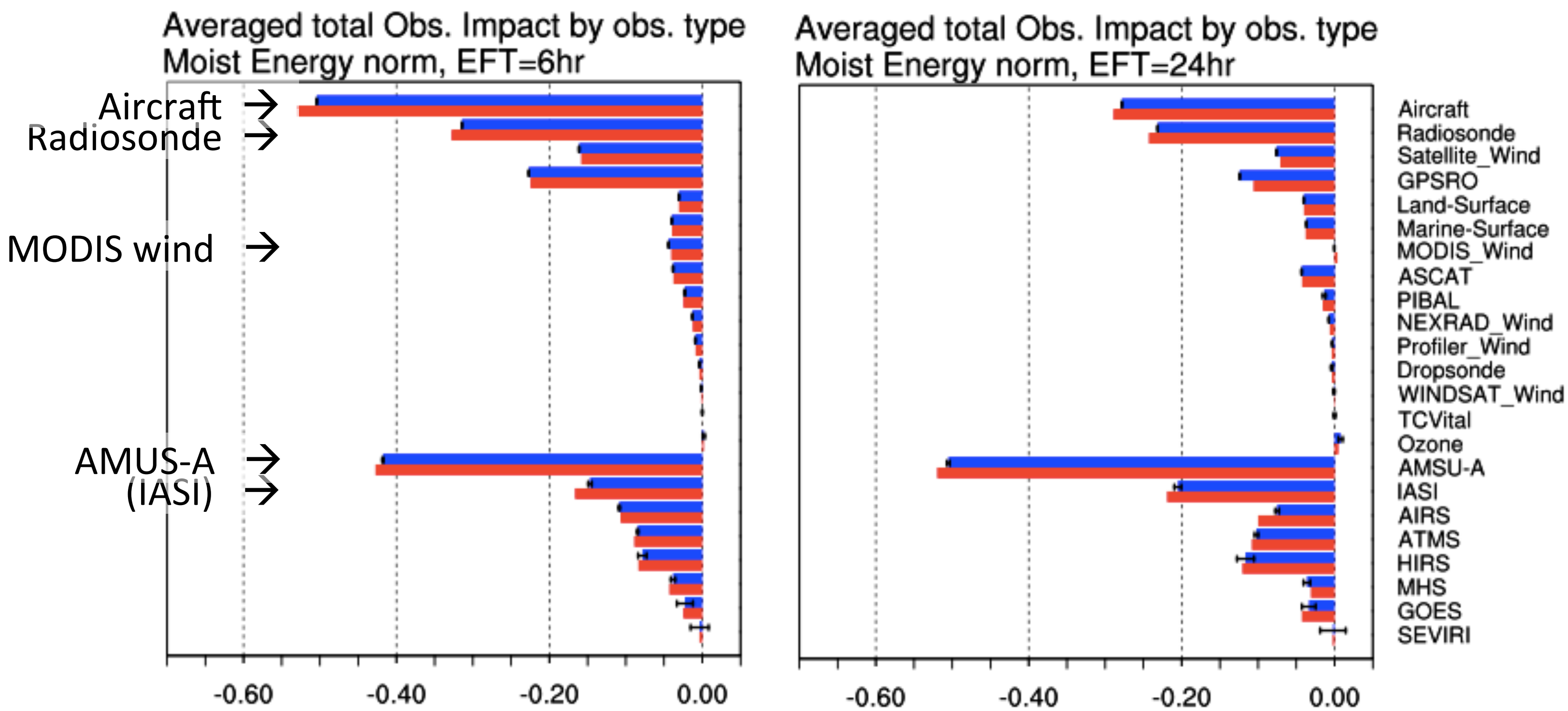
where $\mathbf{z}$ is an "intermediate analysis increment" in observation space

R-sensitivity results 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

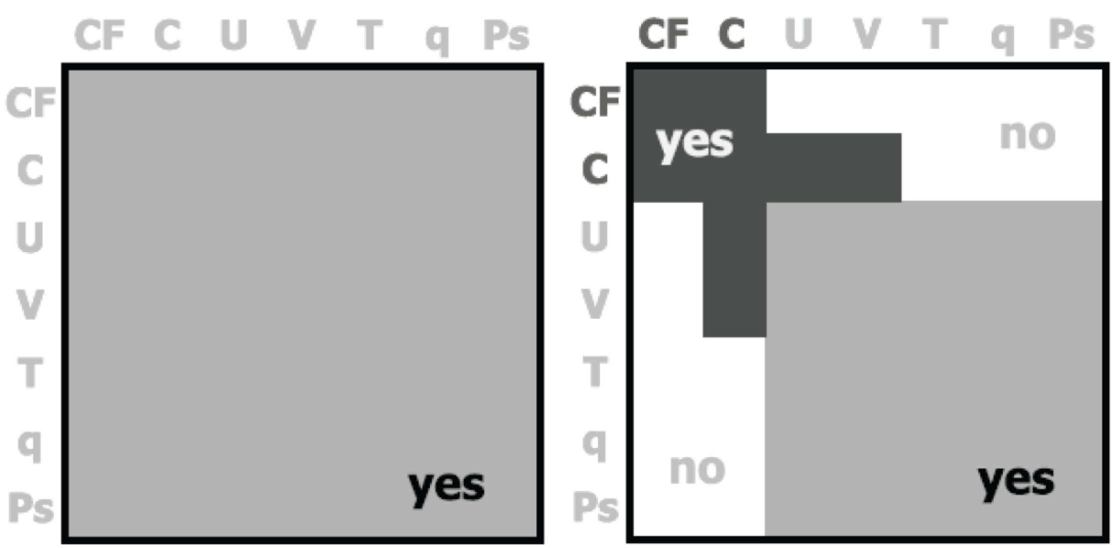


- Aircraft, Radiosonde and AMSU-A: significant improvement of EFSO-impact
- IASI: Significant improvement in EFSO although its error covariance is untouched!
- Very promising results for quick testing of new observing systems!

4) 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.

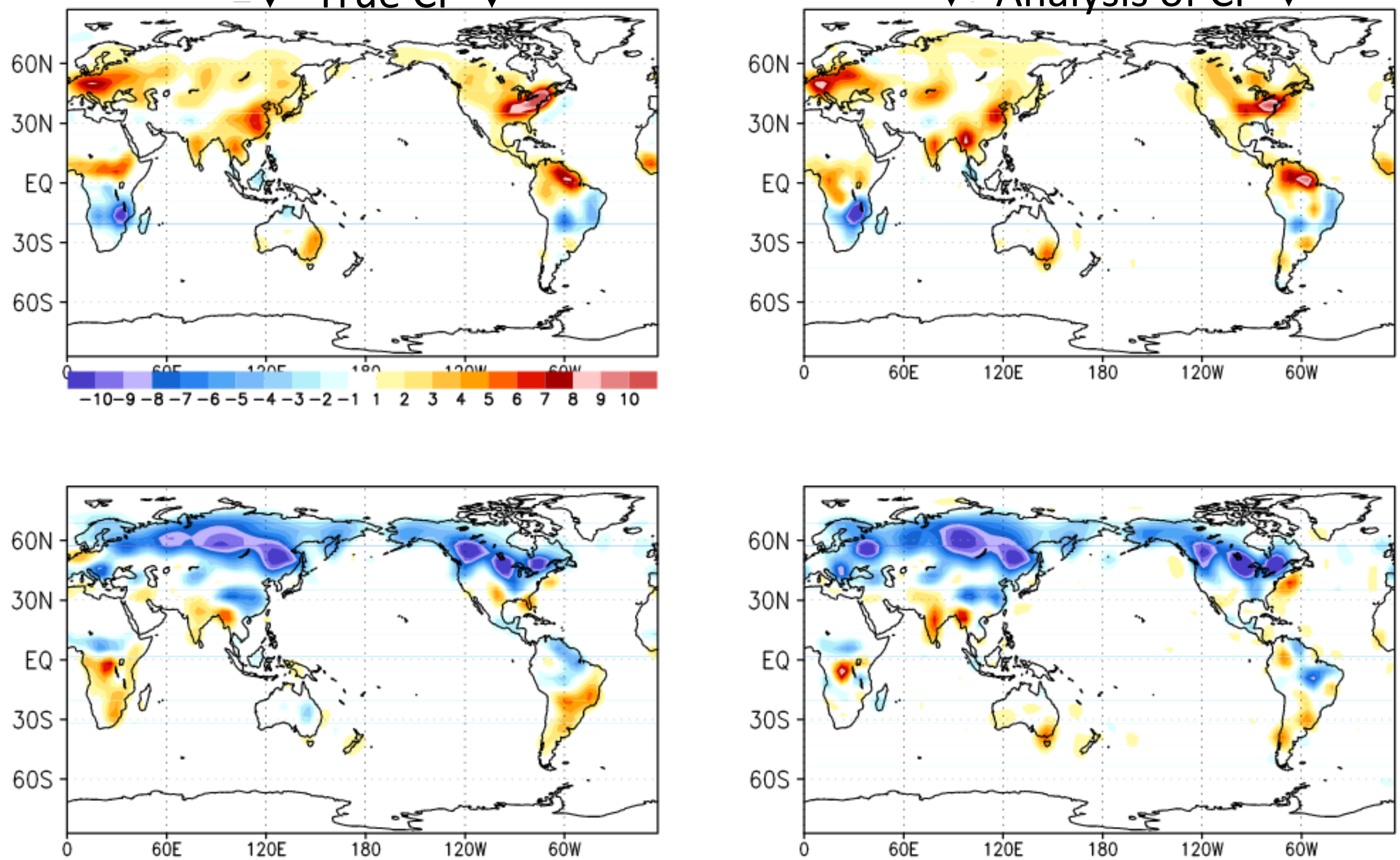
Surface carbon fluxes CF from atmospheric assimilation of meteorological variables and CO_2 obtained as evolving parameters (OSSE). Kang et al., JGR, 2011, 2012



"Variable Localization" in the B matrix

OSSE Results

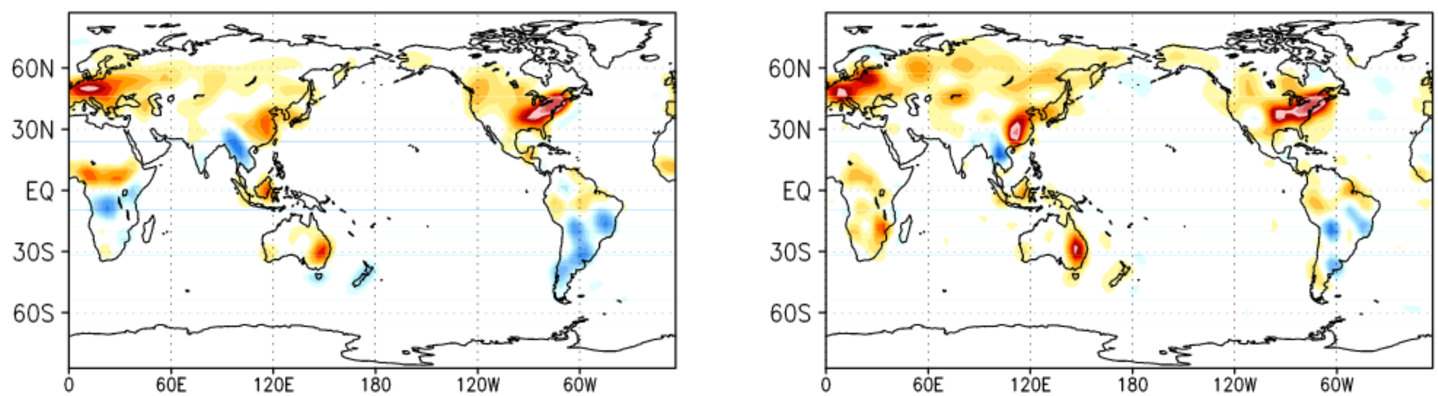
▼ True CF ▼ ▼ Analysis of CF ▼



00Z01APR ▶
After three months of DA

00Z01AUG ▶
After seven months of DA

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



00Z01JAN ▶
After one year of DA

5) How can we estimate and correct Big Model bias?

- The best current estimate of nature is the Analysis.
- The First Guess (6hr forecast) contains the initial forecast errors (before they grow nonlinearly).
- Analysis - First Guess = Analysis Increments (**AI**) =
  - Initial (linear) model errors.
- **The time average of AI is the best estimate of the error growth due to model bias in 6 hr.**
- Danforth, Kalnay and Miyoshi (DKM-2007) estimated the 6hr errors of the SPEEDY model.
- Estimated the average SPEEDY model error (bias) by averaging over several years the 6 hour forecast (started from reanalysis R1) minus the reanalysis.

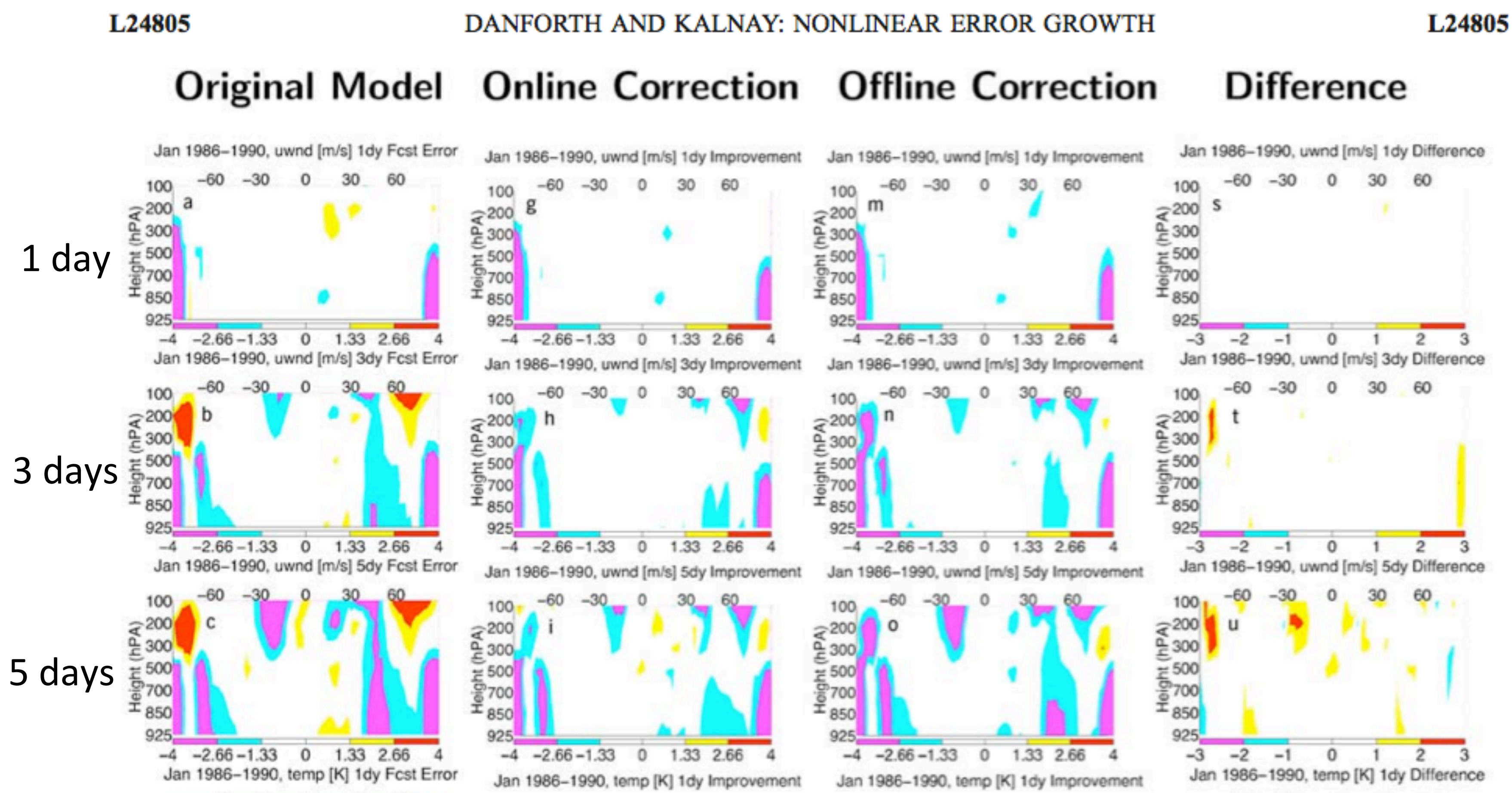
DKM-2007 results

- Estimated the monthly mean 6hr forecast bias
- Corrected the model by adding (–bias/6hr) to each variable time derivative, at each grid point.

Results

- The bias correction after 3 or 5 days was the same as the best *a posteriori bias* correction.
- But the random errors were **smaller**.
- The dominant EOFs of the 6hr debiased forecast errors were the errors in the diurnal cycle.
- It was possible to estimate the systematic errors for anomalies (e.g., ENSO, lows over land or over ocean)

The model corrected online did same or better than the model statistically corrected off-line



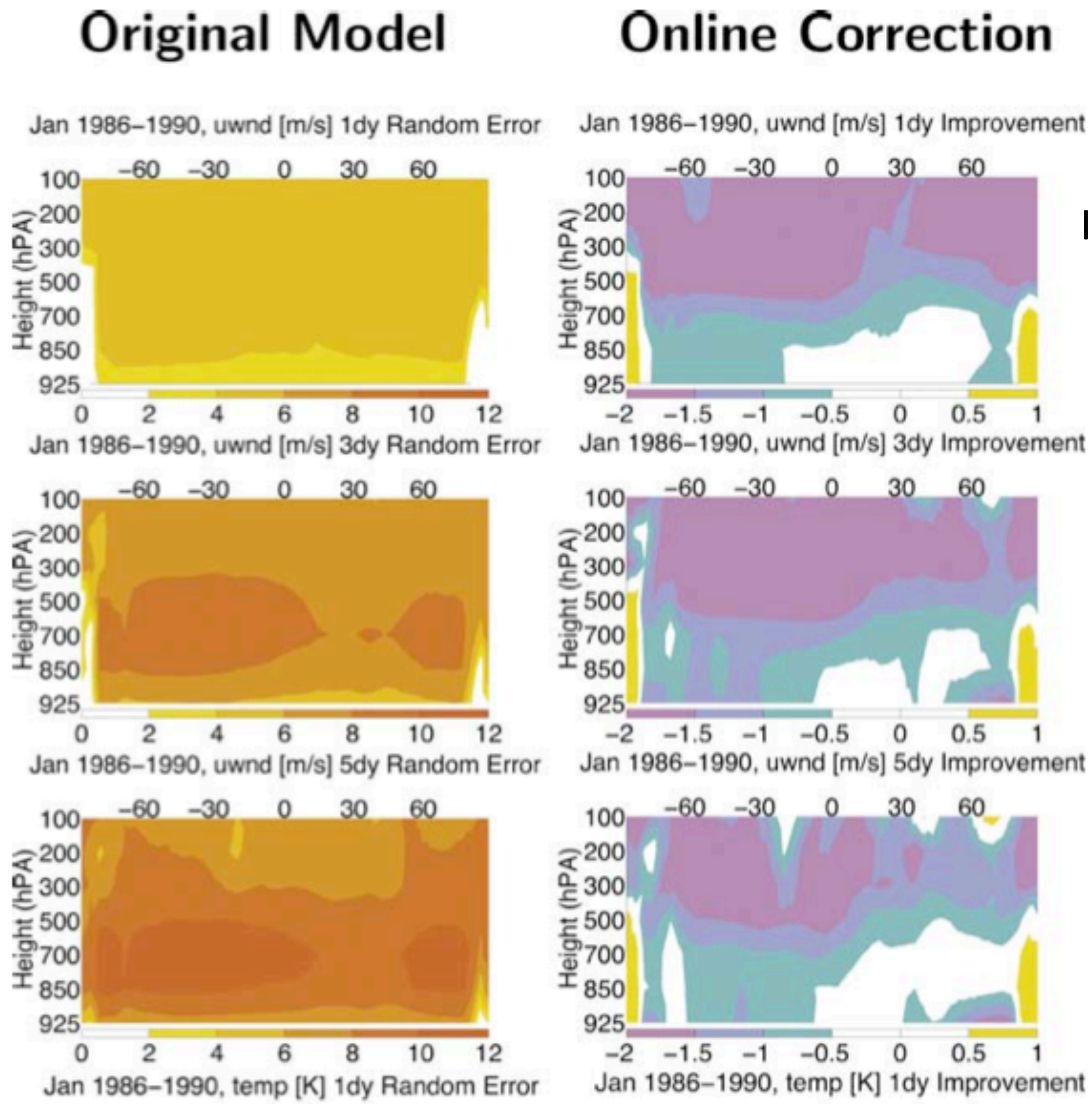
And the random errors were significantly smaller!

1 day

3 days

5 days

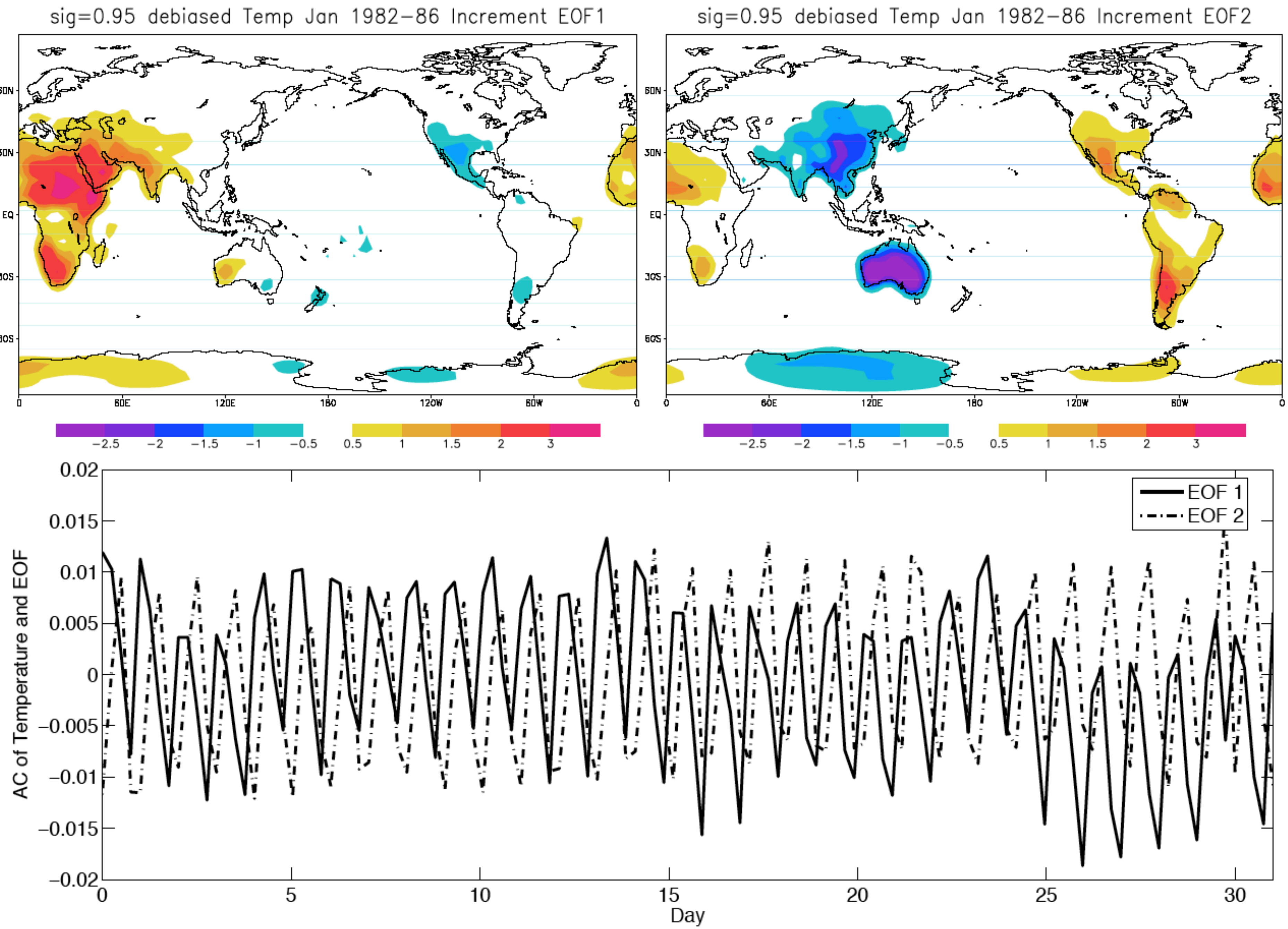
Improvements



How to find the diurnal cycle 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.

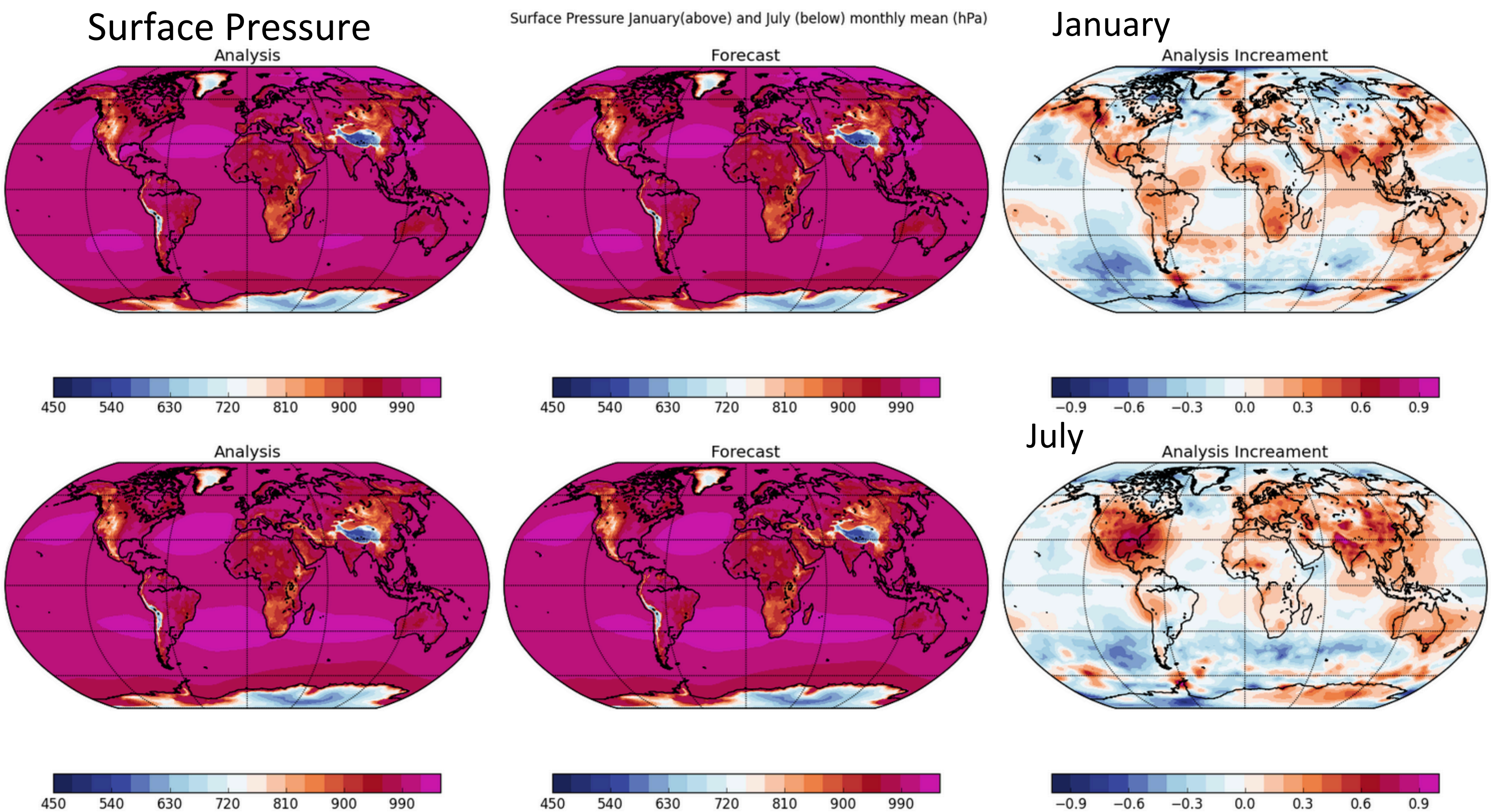
Then they computed the EOFs of the anomaly in the model error, and found two dominant EOFs representing the model error in representing the diurnal cycle:


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
0
5
10
15
20
25
30

Implications for improving the model bias

- The DKM2007 method gave very good results with the SPEEDY model, using R1 as an approximation of the true atmosphere.
- The -bias/6hr was added to the SPEEDY time derivatives (u,v,T,p_s).
- This corrected the bias, getting similar or better results than an *a posteriori* bias correction! In addition, random forecast errors were also reduced.
- It was also used to improve the diurnal cycle and to find the state dependent systematic errors (e.g., during an El Niño).
- **It can be tried on the GFS (or the CFS!) taking advantage of the *Analysis Increments*, i.e., the difference between the Analysis and the Forecast.**
- Dr. Fanglin Yang (NCEP) very kindly provided us with 2012, 2013, and 2014 Analyses and Forecasts.

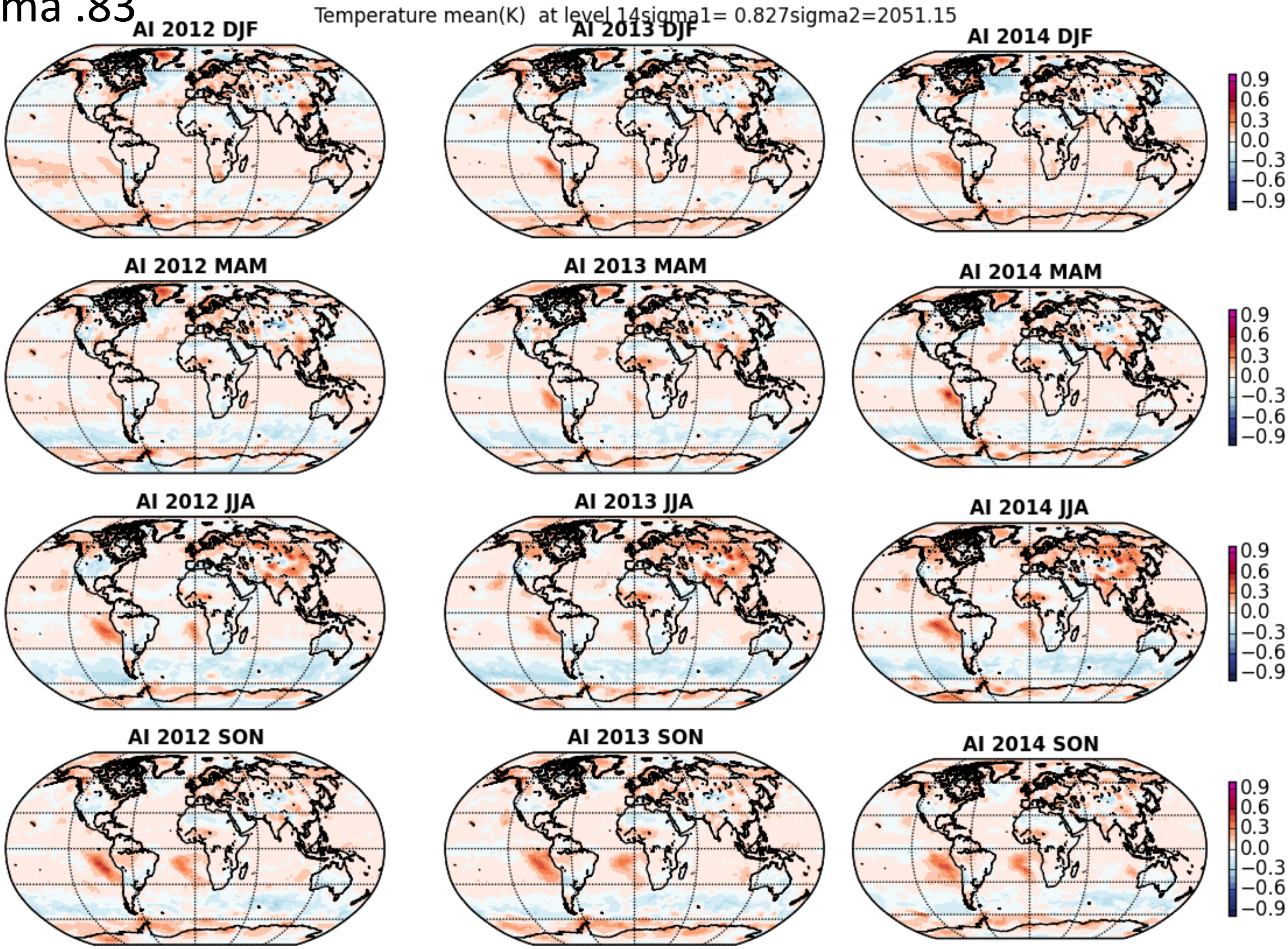
First results: 2014 Analyses, Forecasts and AIs



P_s is too low over continents, too high over oceans in both winter and summer.

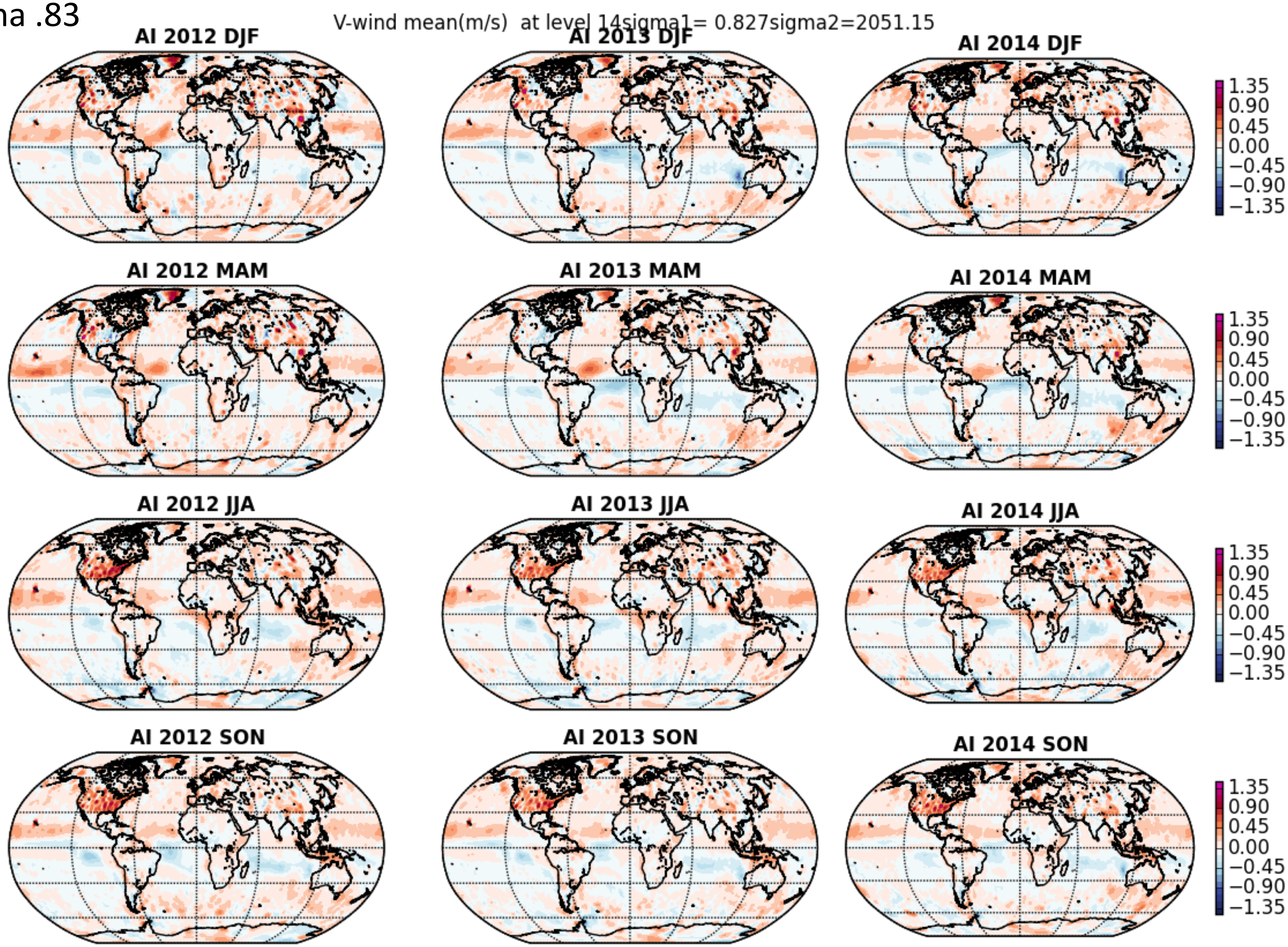
Seasonal Analysis Increment 2012-2013-2014

T sigma .83



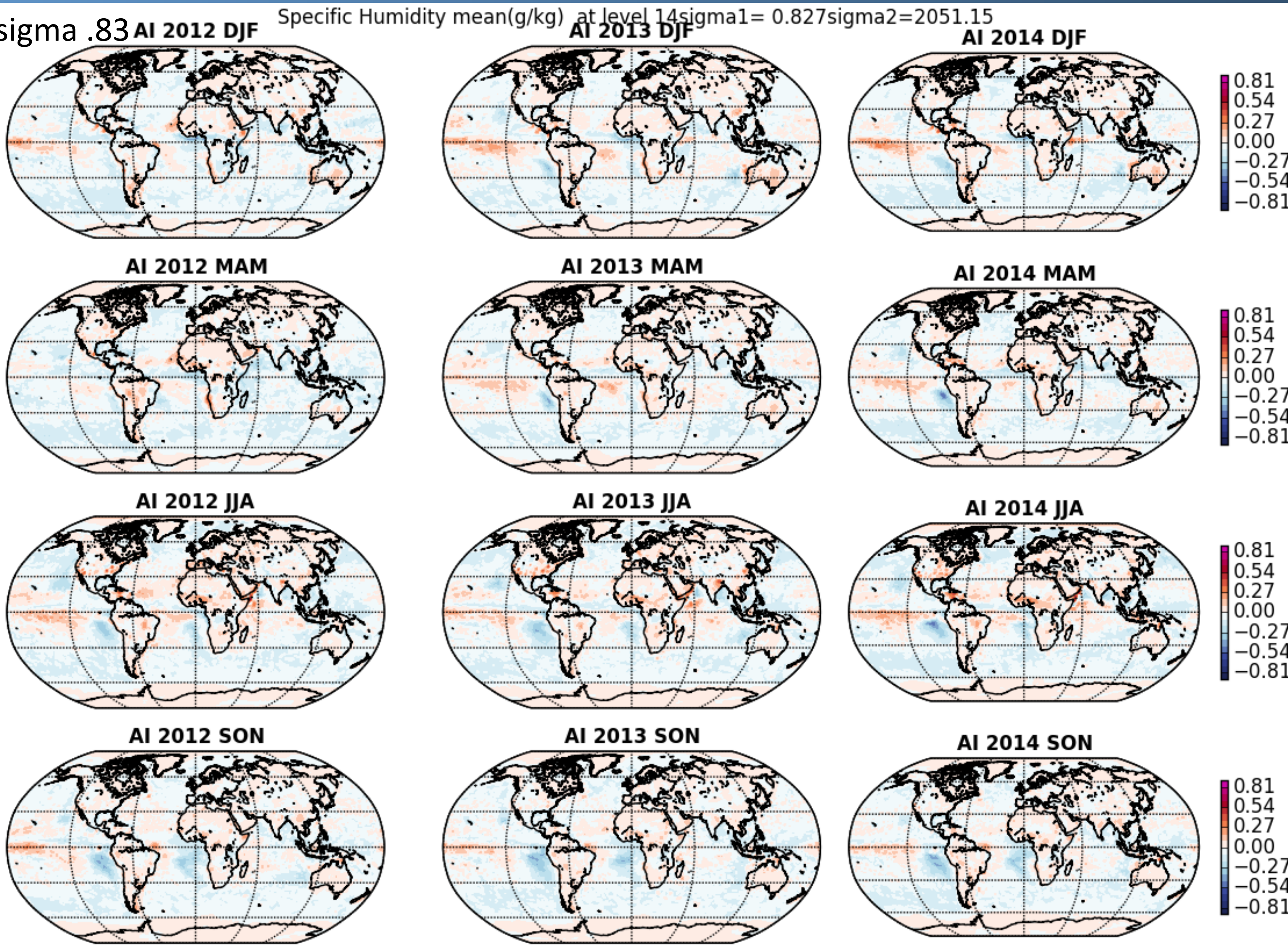
Seasonal Analysis Increment 2012-2013-2014

V sigma .83



Seasonal Analysis Increment 2012-2013-2014

Q sigma .83



How do we plan to reduce model bias?

- Check the robustness of the monthly or seasonal averaged AI (2014 vs. 2013 vs. 2012) ✓
- Perform exploratory low resolution (T254) experiments correcting the perceived model bias by adding AI/6hr to each variable time derivative.
- Test the impact on the forecast skill.
- Explore the diurnal cycle of the AI. Test if the diurnal cycle errors can be reduced.
- **If successful, the AI bias correction will also guide the development of the physical parameterizations.**

SUMMARY 1

- Applications of EnKF-based data assimilation can go beyond providing the best initial conditions:
- They can improve both observations and models:
- Improve observations with EFSO and PQC.
- Improve the Obs Error Covariance R
- Improve the model parameters
- Improve the models by using the Analysis Increments to correct the models' bias

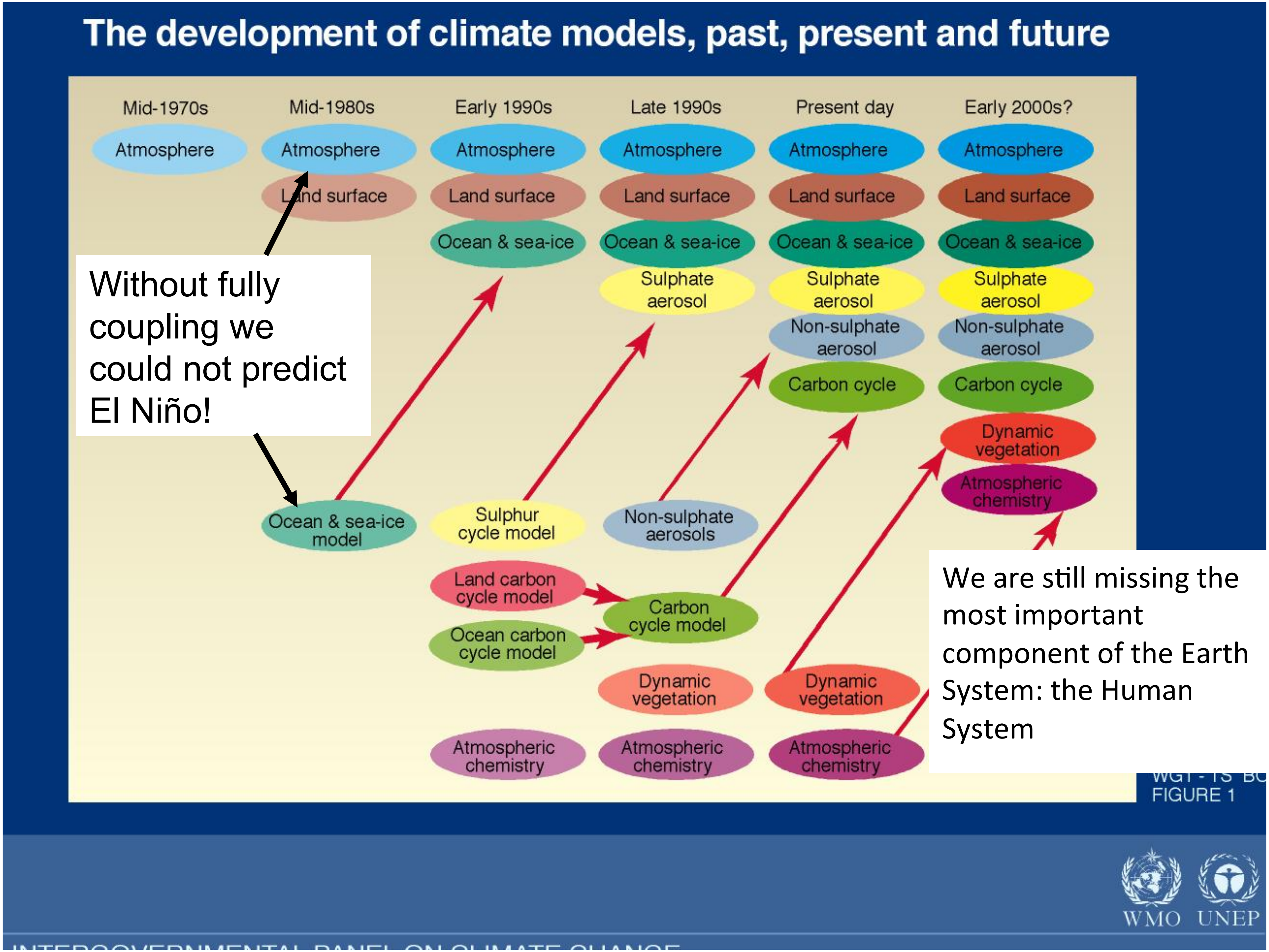
The Human and Earth System models should be fully coupled, and could be tuned with Big Data DA

Jorge Rivas[2], Safa Motesharrei[1,3]
and Eugenia Kalnay[1]
[1]University of Maryland; [2]IGES/U of Minnesota;
[3]National Socio-Environmental Synthesis Center (SESYNC)

Big Data and Environment Symposium
University of Buenos Aires, 10-13 November 2015

Earth and Human Systems

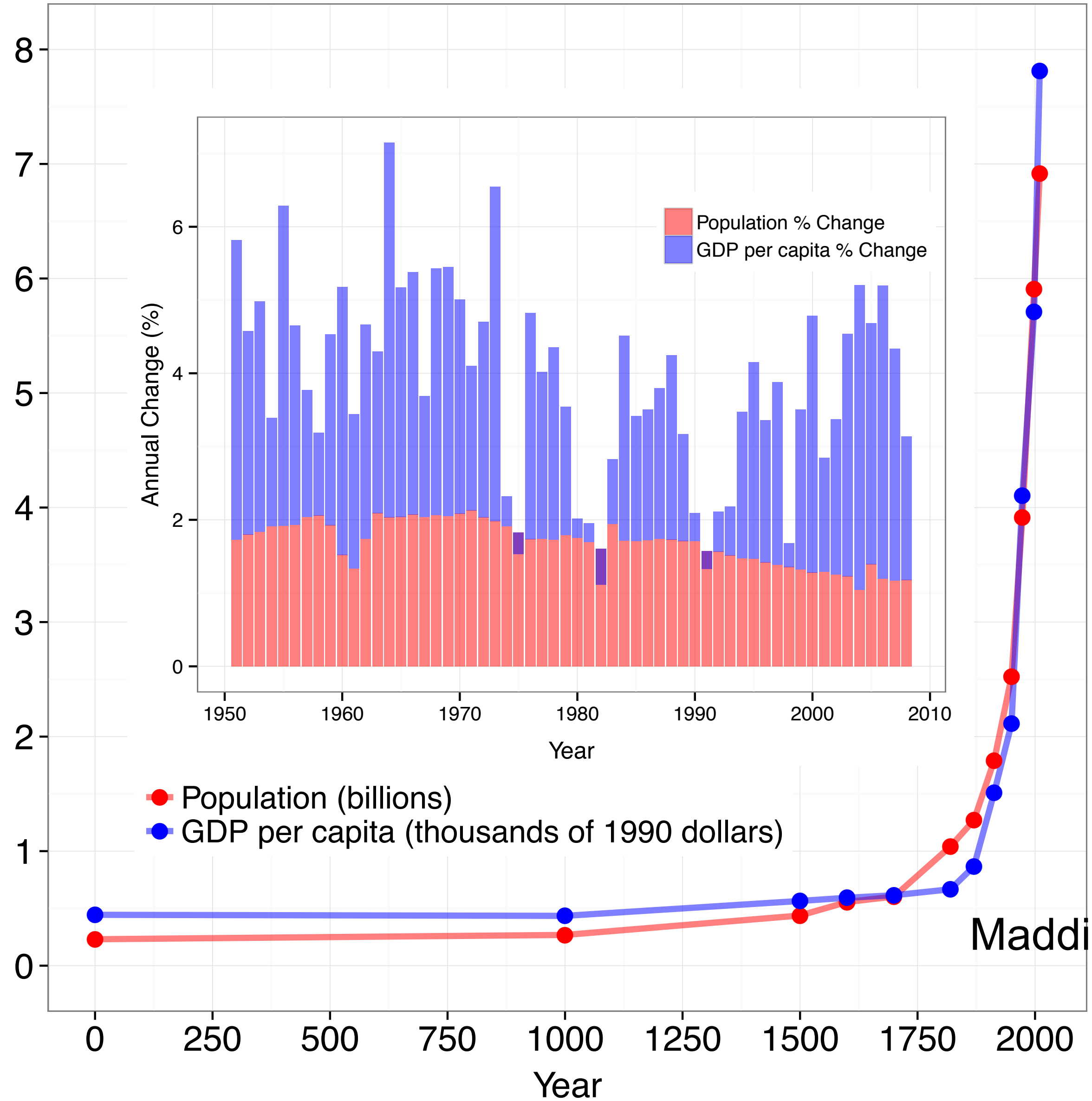
- The Earth System is completely dominated by the Human System.
- In order to understand their interactions we need to couple them bidirectionally, i.e., with feedbacks.
- Currently, IPCC models and even Integrated Assessment Models **do not couple population**. Instead, population is exogenously obtained from UN projections.


The development of climate models, past, present and future
Mid-1970s
Mid-1980s
Early 1990s
Late 1990s
Present day
Early 2000s?
Atmosphere
Land surface
Ocean & sea-ice
Sulphate aerosol
Non-sulphate aerosol
Carbon cycle
Dynamic vegetation
Atmospheric chemistry
Ocean & sea-ice model
Sulphur cycle model
Non-sulphate aerosols
Land carbon cycle model
Ocean carbon cycle model
Carbon cycle model
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

Growth of Population and GDP/Capita: Consumption of Resources is their Product!

1AD	0.3b
1650	0.5b
1800	1.0b
1927	2.0b
1960	3.0b
1975	4.0b
1987	5.0b
1998	6.0b
2011	7.0b

2100 | 10.9b



Maddison, (2001)

Why was the population able to grow so fast since the 1950's?

Two reasons:

1) Sanitation and Antibiotics (Public Health → living longer)

2) Use of fossil fuels in agriculture starting in the 1950's:

- fertilizers, pesticides, irrigation, mechanization (Green Revolution).

1950 to 1984: production of grains increased by 250% and the population doubled

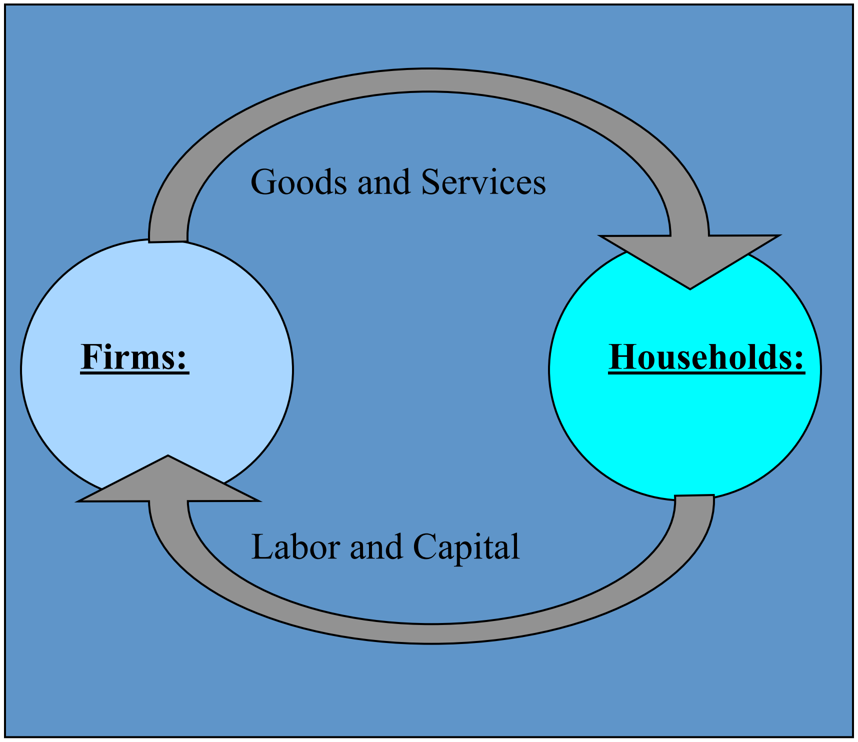
Without fossil fuels population would be much smaller!

- Growth in grain production is now flattening out
- Industrial farming is destroying forests, soil
- Urban and suburban sprawl is overrunning best farmland

This is not sustainable: "We are drawing down the stock of natural capital as if it was infinite" (Herman Daly)

Standard Neoclassical Economic Model

As Herman Daly, Robert Costanza, and other scholars in the field of Ecological Economics describe,



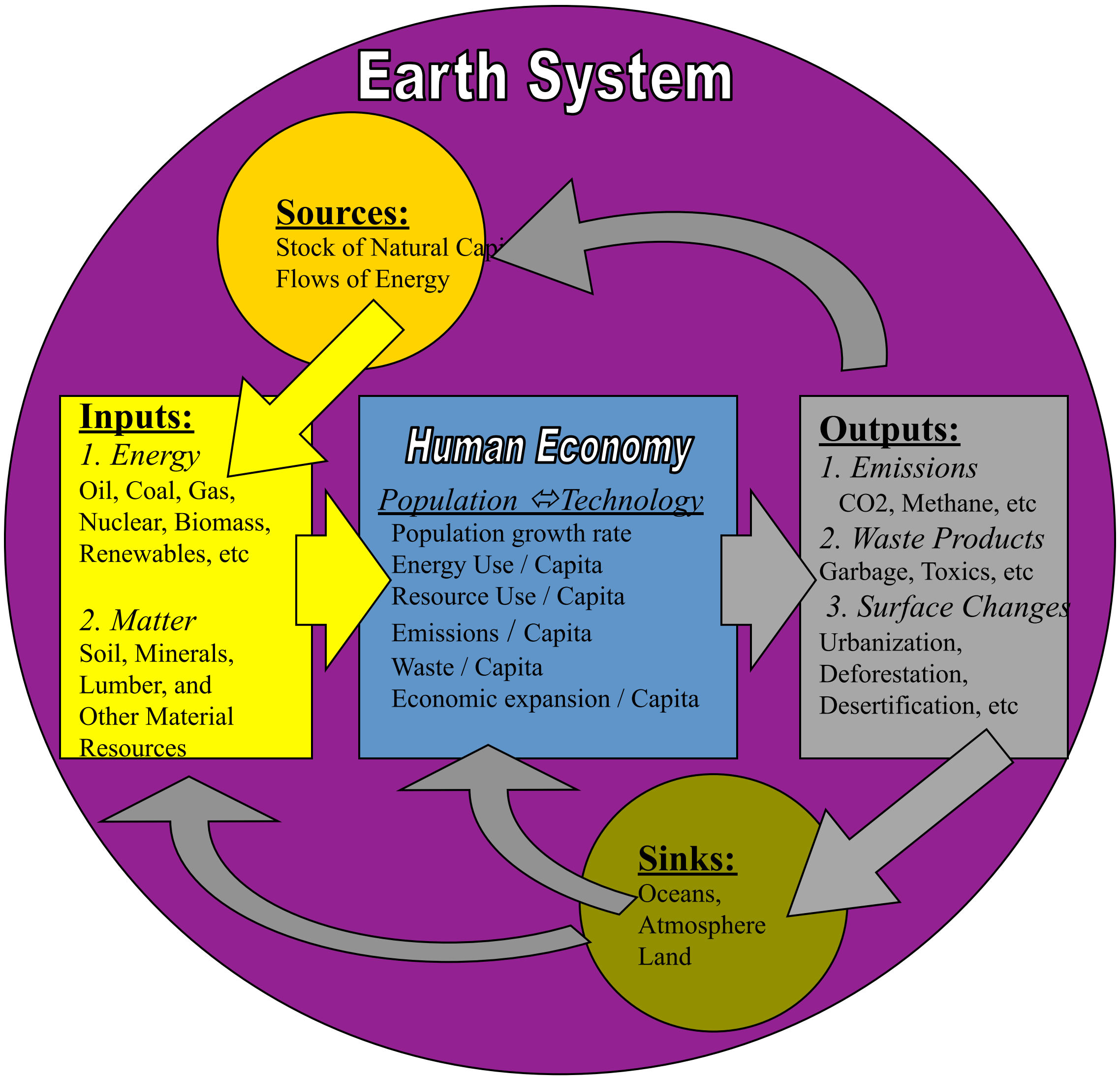
The standard Neoclassical Economic Model does not account for:

- Inputs (resources)
- Outputs (pollution)
- Stocks of Natural Capital
- Dissipation of Energy (i.e., a Perpetual Motion Machine)
- Depletion, Destruction or Transformation of Matter

Therefore, no *effects on the Earth System*, and *No Limits to Growth*.

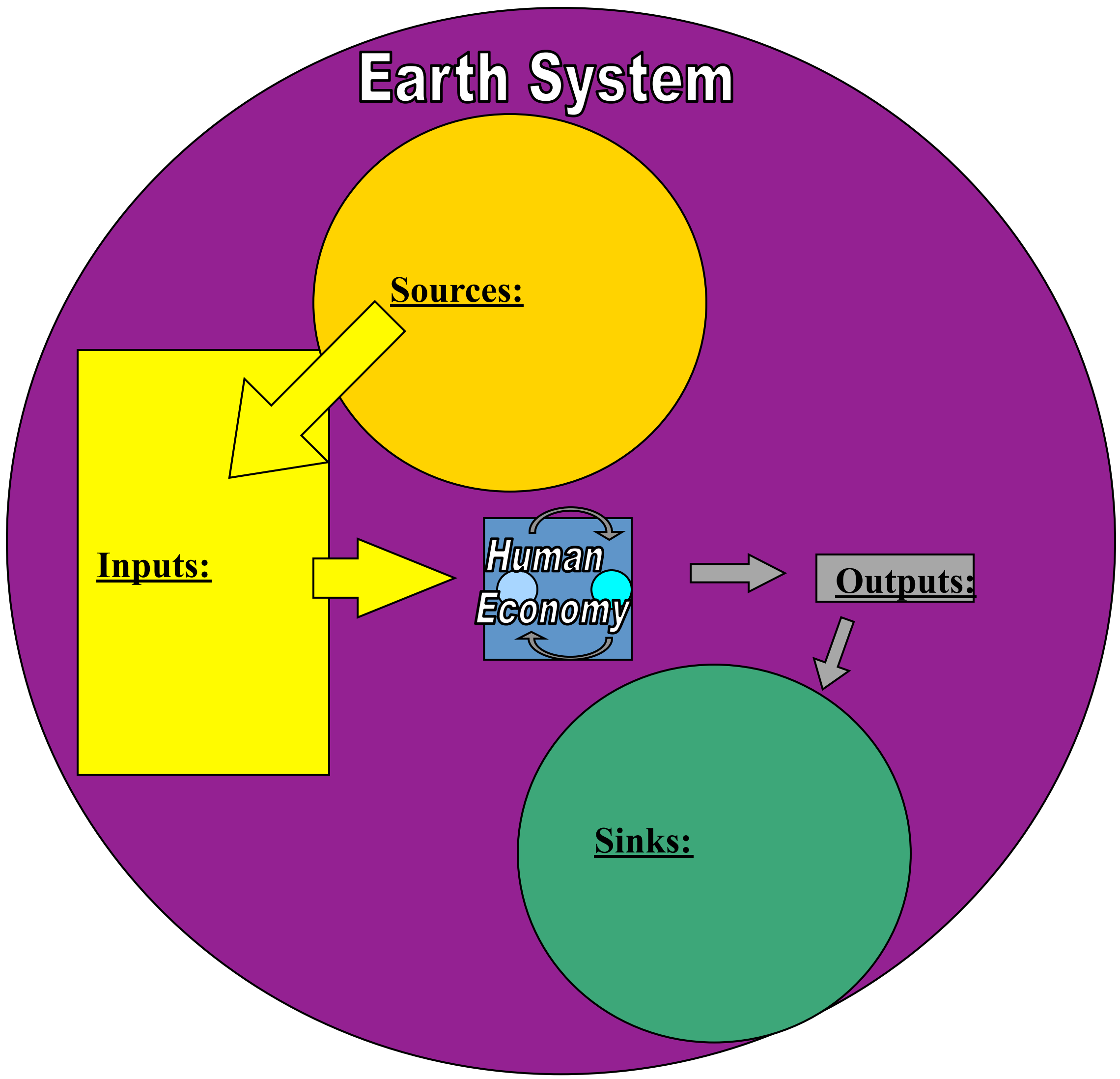
Realistic **Ecological** Economic Model (Herman Daly)

- Incorporates INPUTS, including **DEPLETION** of **SOURCES**
- Incorporates OUTPUTS, including **POLLUTION** of **SINKS**



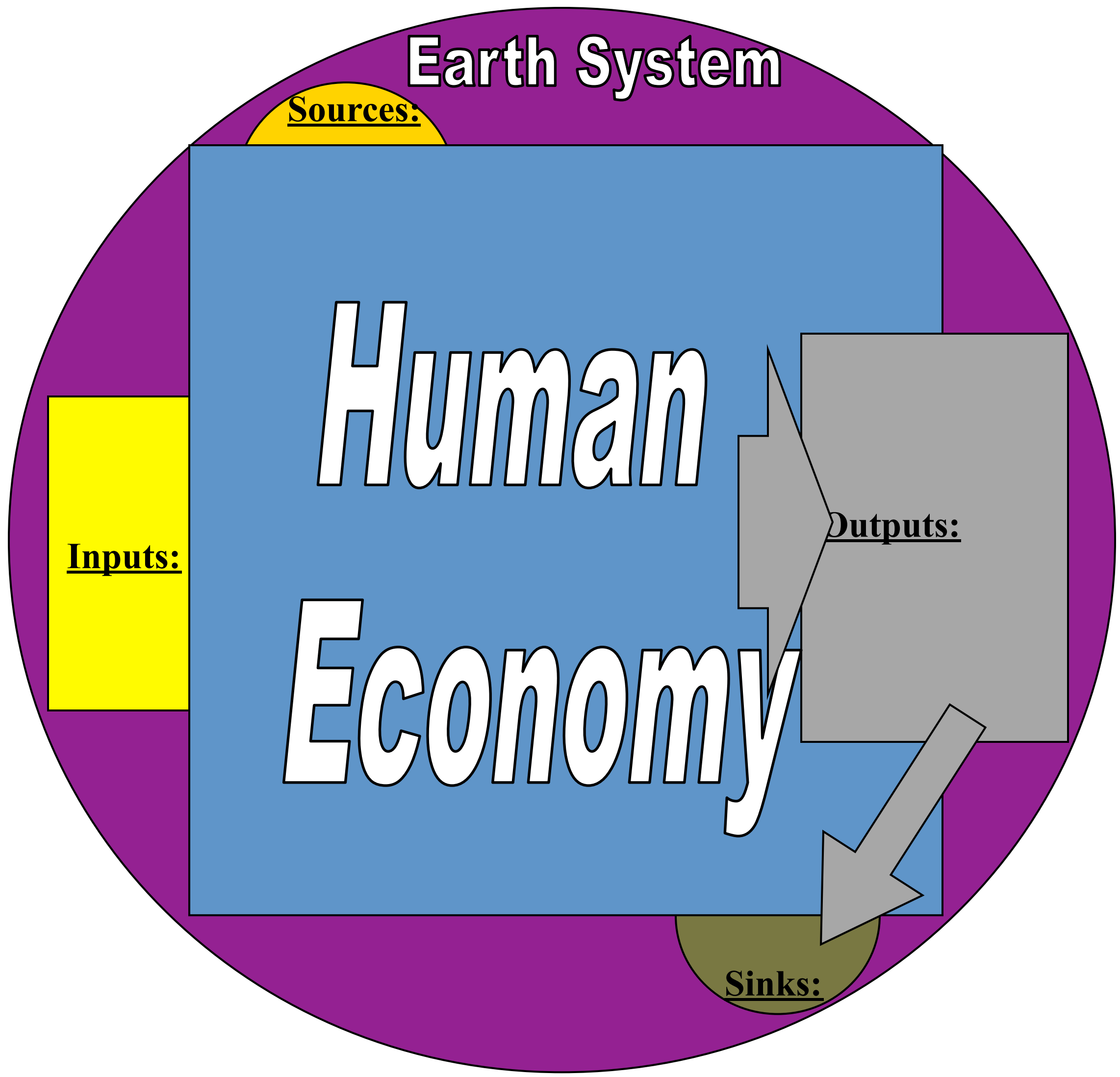
“Empty World” Model

- Throughout <u>most of human history</u>, the **Human Economy** was so small relative to the **Earth System**, that it had little impact on the **Sources** and **Sinks**.
- In this scenario, the standard isolated economic model might have made sense.

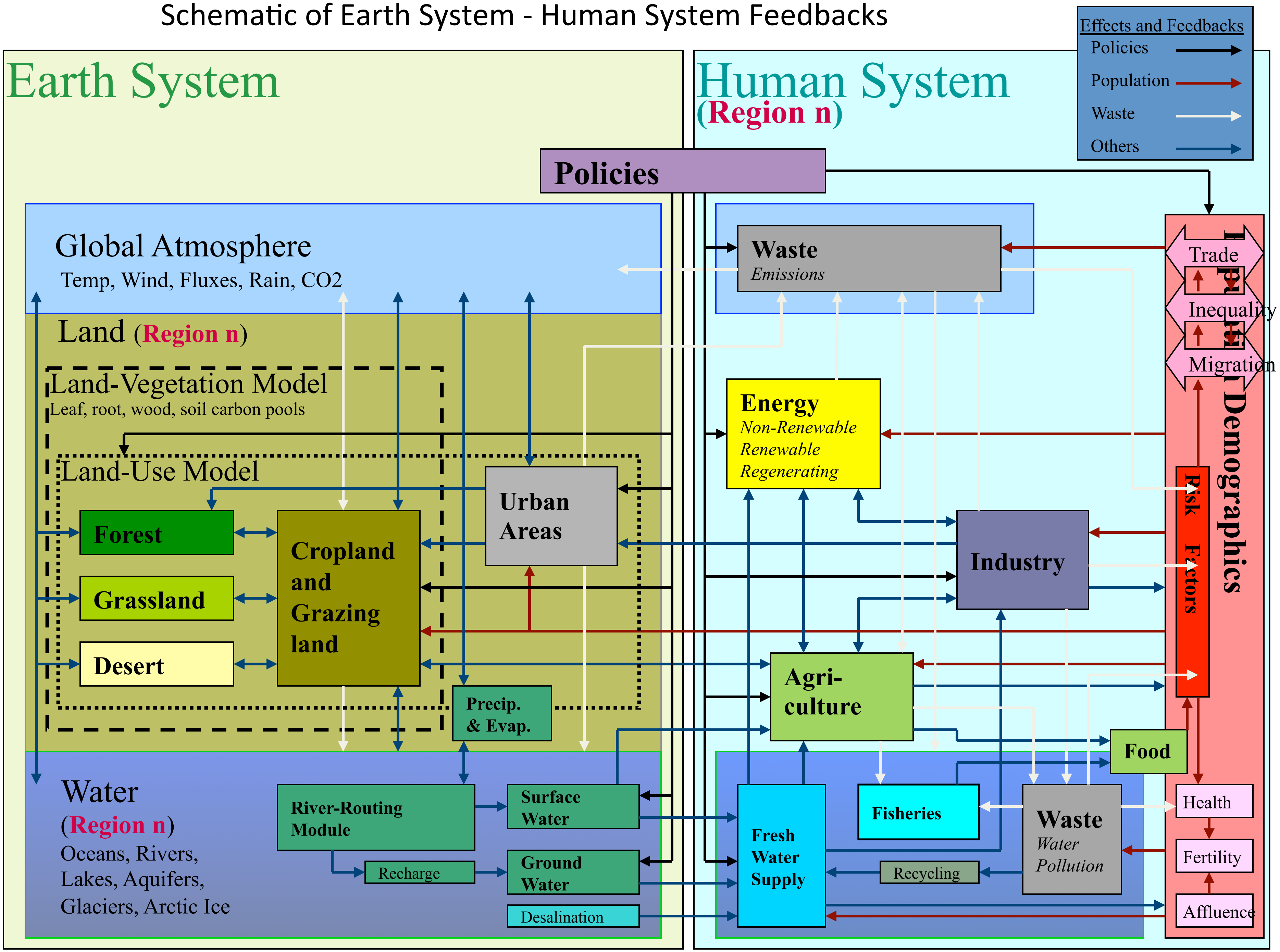


“Full World” Ecological Economic Model

- Today, the **Human Economy** has grown so large, it has very large **Effects** on the **Earth System**, ***Depleting*** the **Sources** and ***Filling*** the **Sinks**. It is clear that ***growth cannot continue forever***.



Schematic of Earth System - Human System Feedbacks



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

Just 4 equations!

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

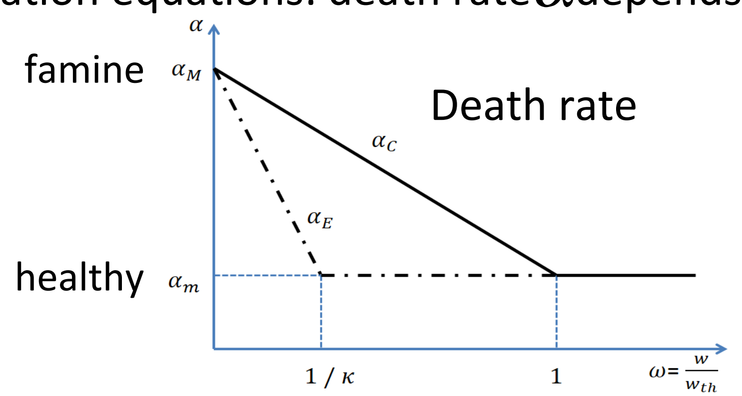
Nature equation: (only the Commoners produce)

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

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:

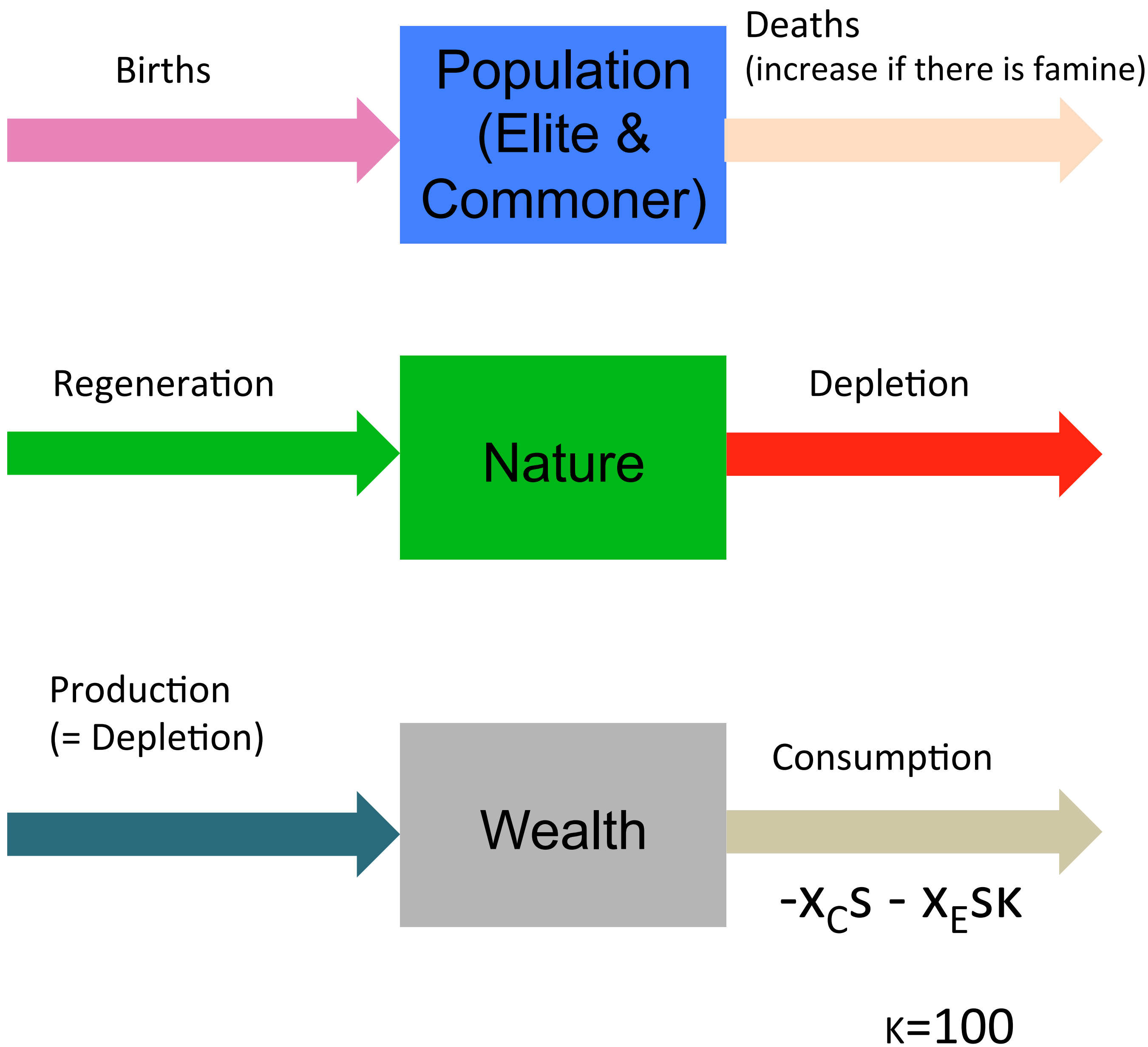


$$\dot{x}_C = -\alpha_C x_C + \beta_C x_C$$

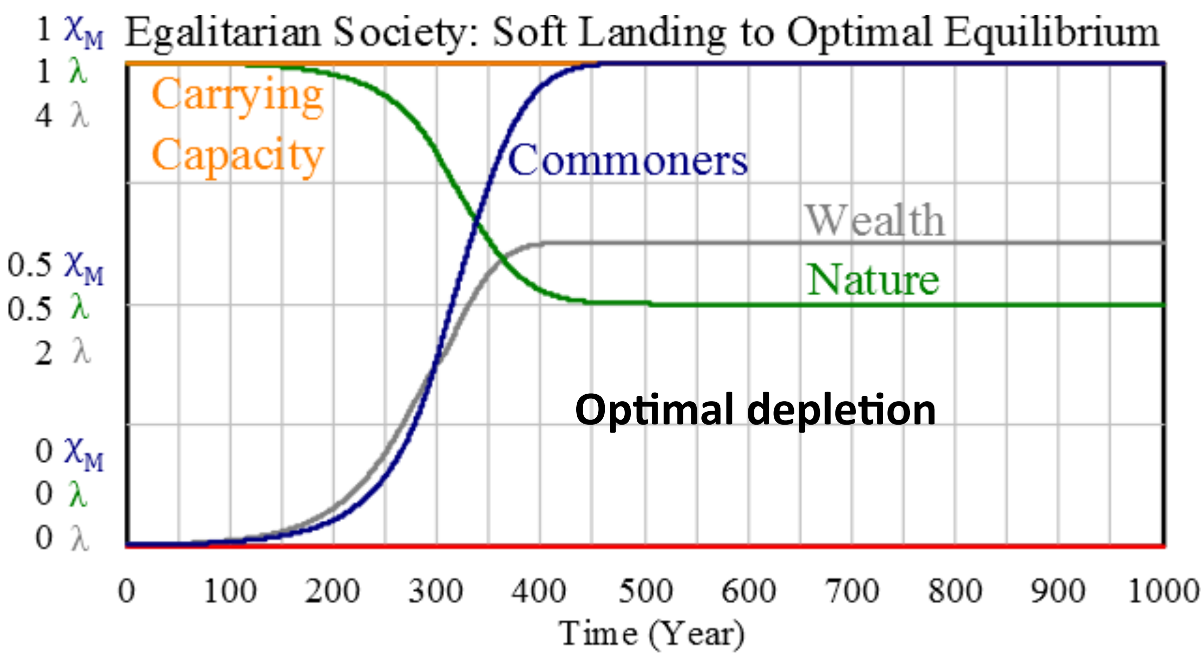
$$\dot{x}_E = -\alpha_E x_E + \beta_E x_E$$

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 and survive longer.

State Variables (Stocks) and Flows in HANDY1

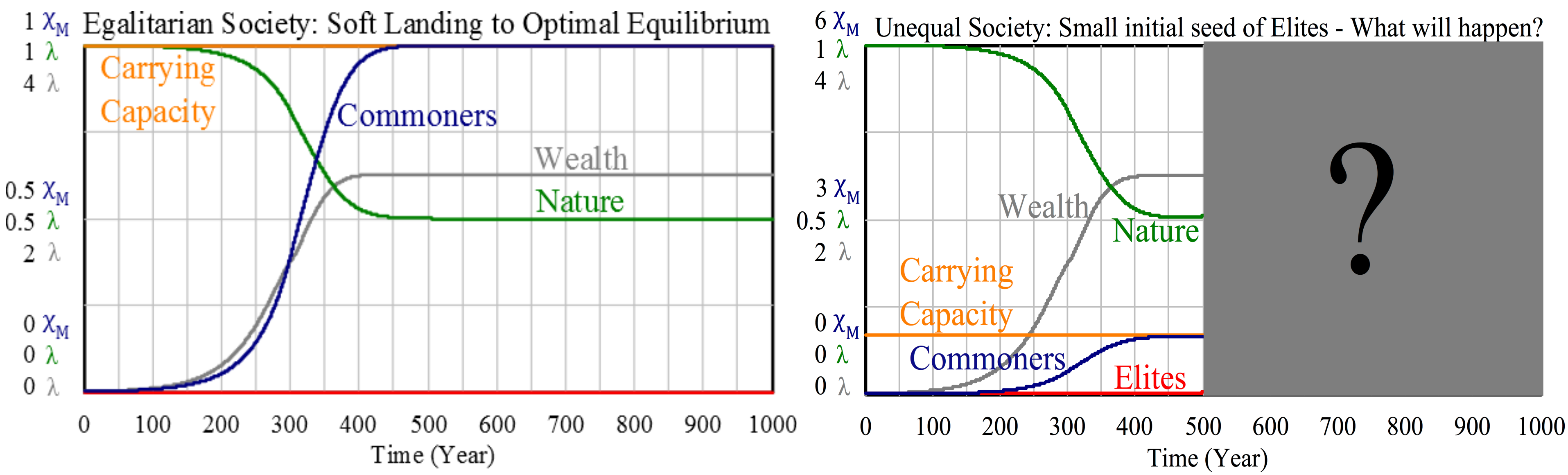


Experiments for an Egalitarian Society (K=1)



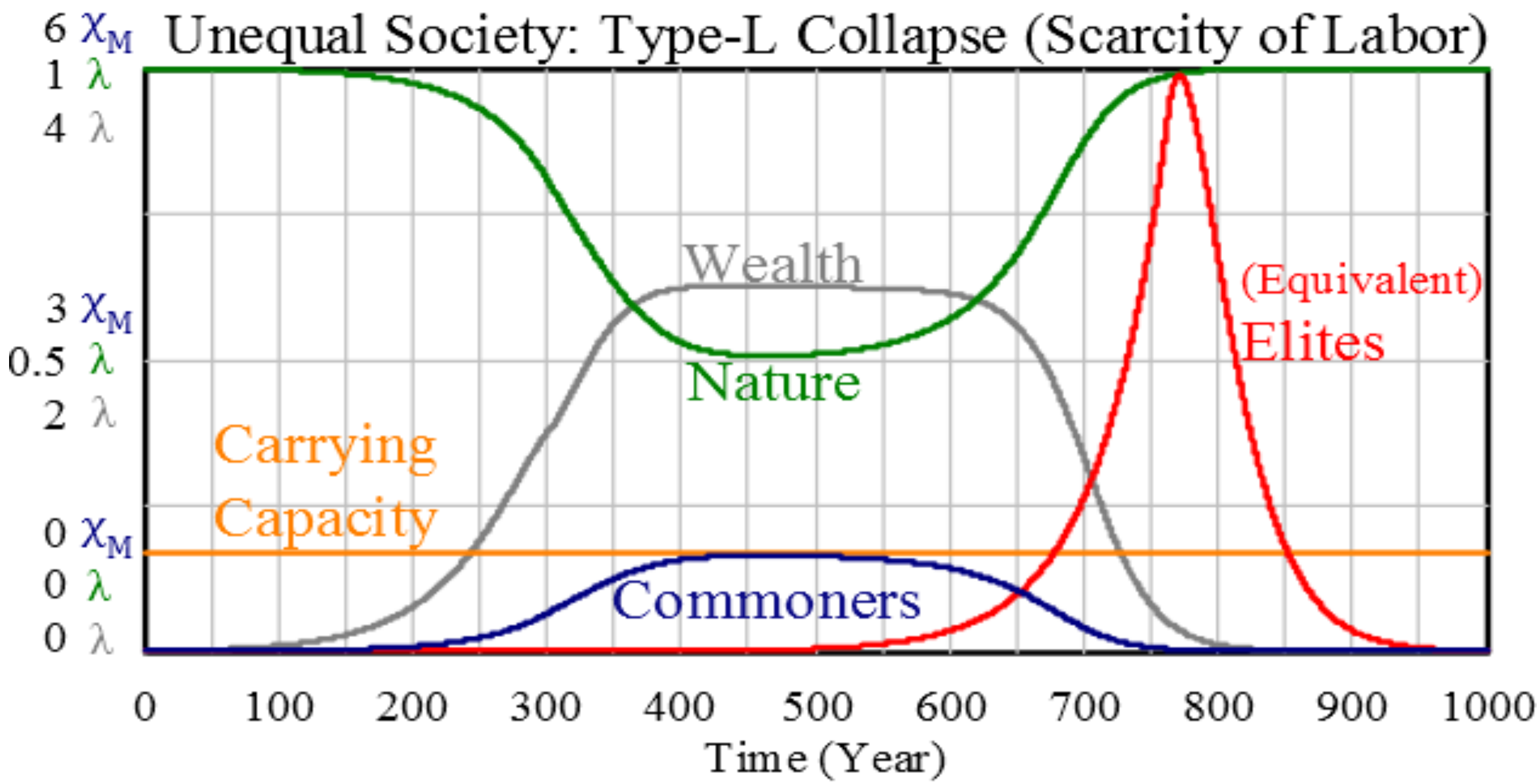
With optimal depletion an egalitarian society reaches equilibrium at the maximum Carrying Capacity

What happens if we introduce Inequality? Optimal depletion, but K=100

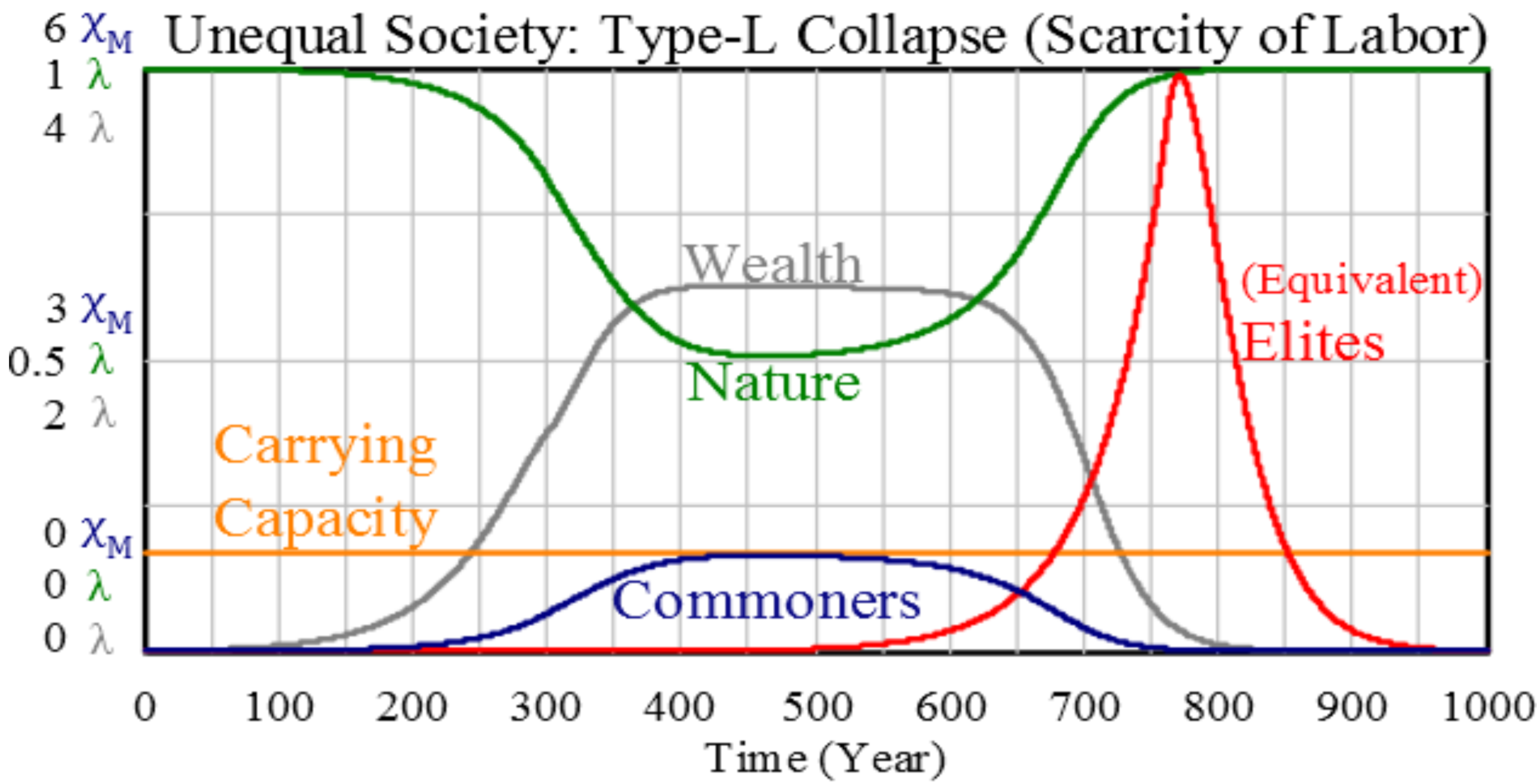


Up until t = 500 years,
both scenarios show the exact same dynamics.

An otherwise *sustainable* society will collapse if there is high inequality ($\kappa = 100$).

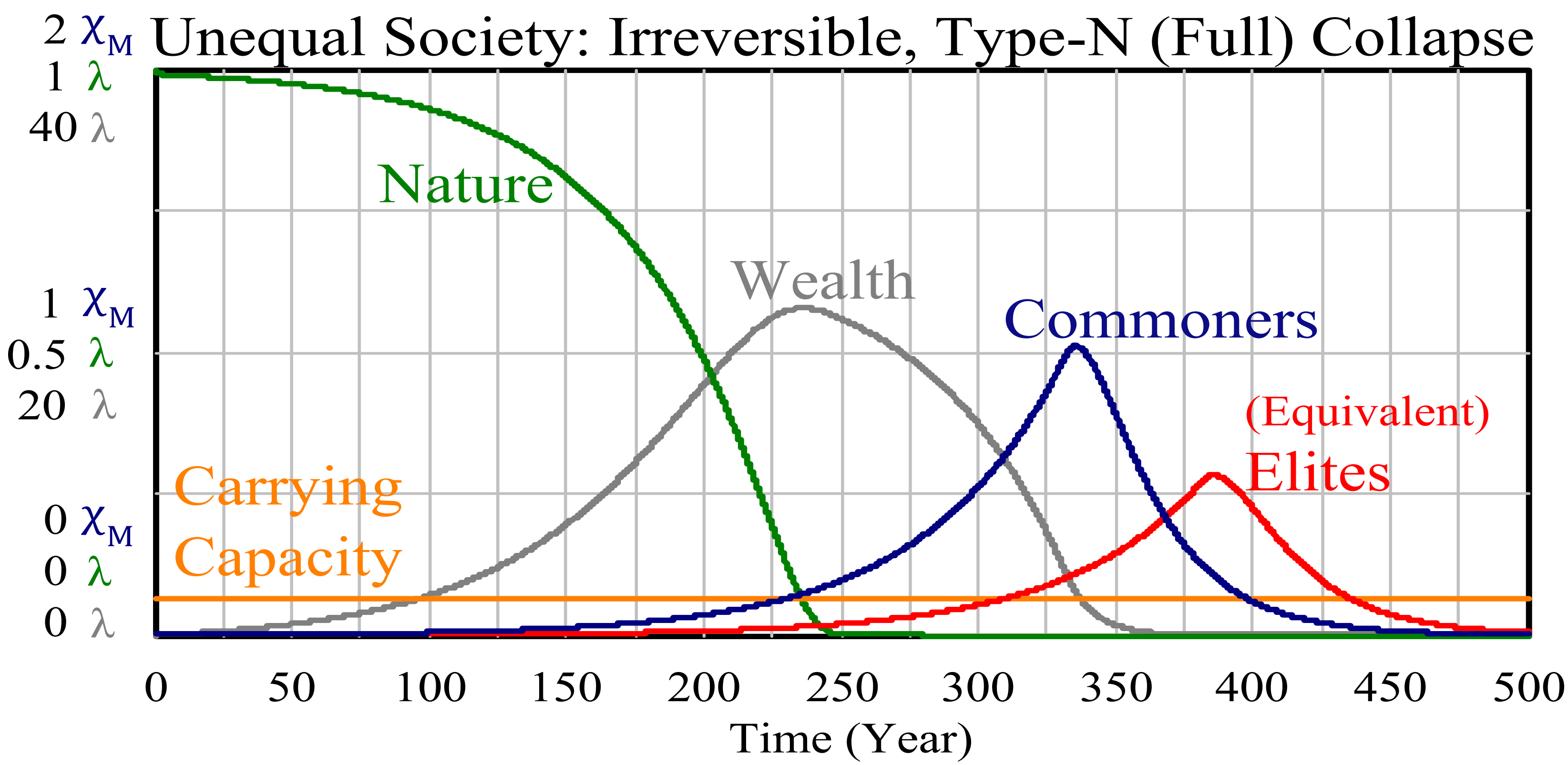


An otherwise *sustainable* society will collapse if there is high inequality ($\kappa = 100$).

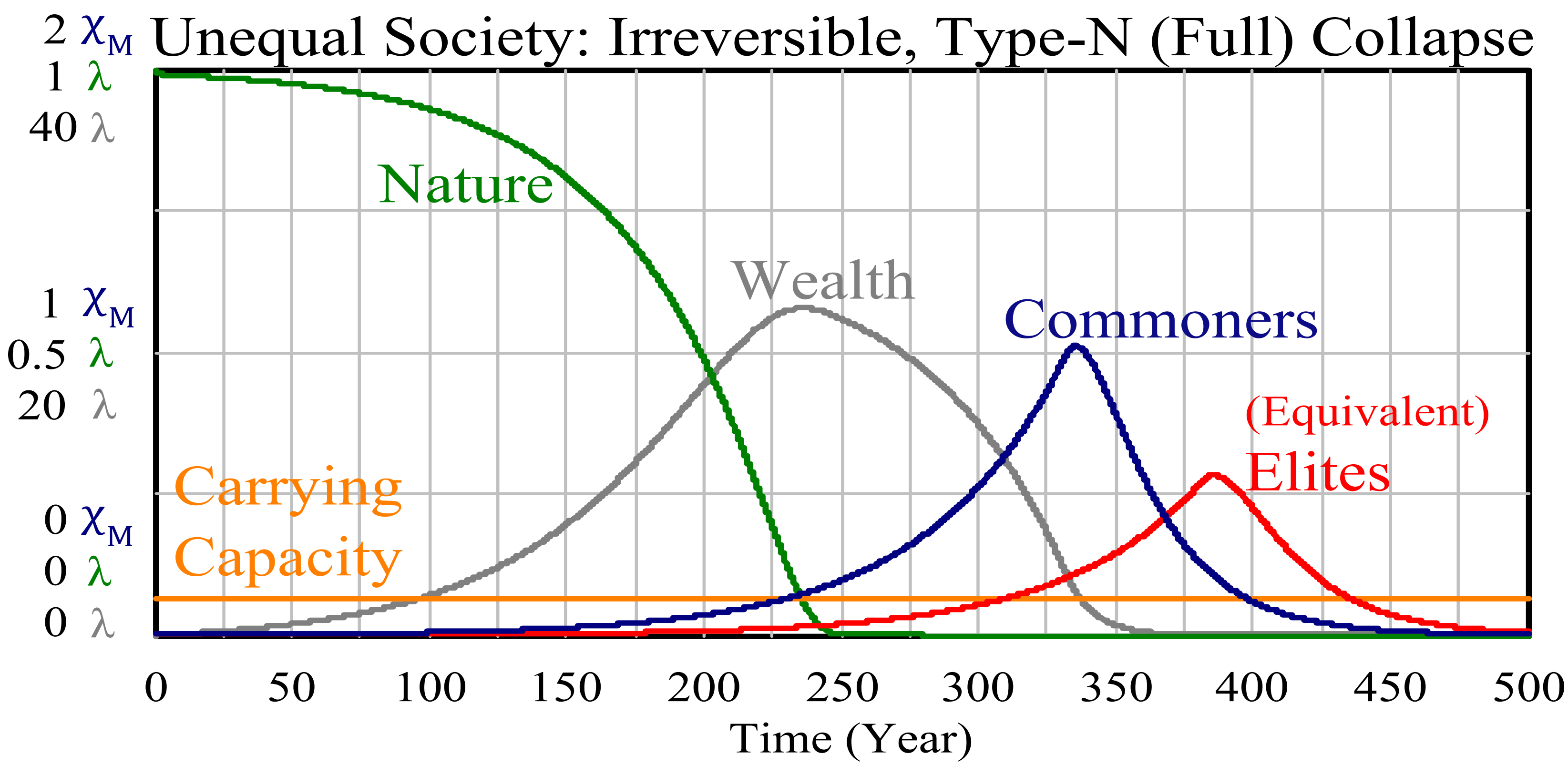


What happens if we have *both* high inequality and high depletion rate?

Typical Collapse: High Depletion Rates and High Inequality at the same time

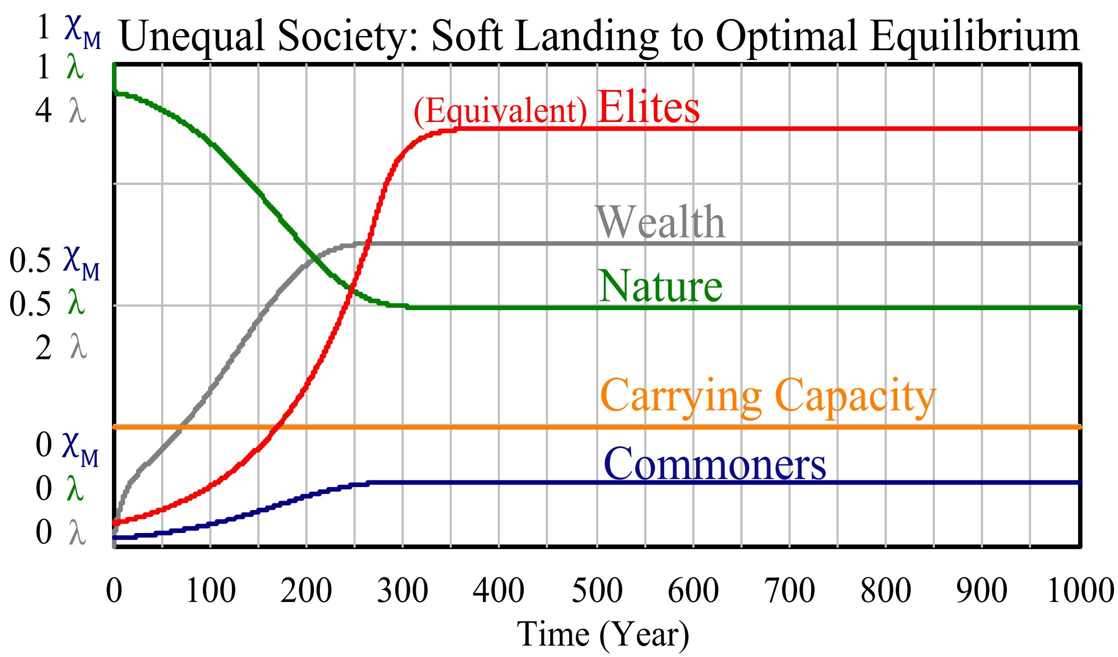


Typical Collapse: High Depletion Rates and High Inequality at the same time



Is there any hope for an unequal society to survive?

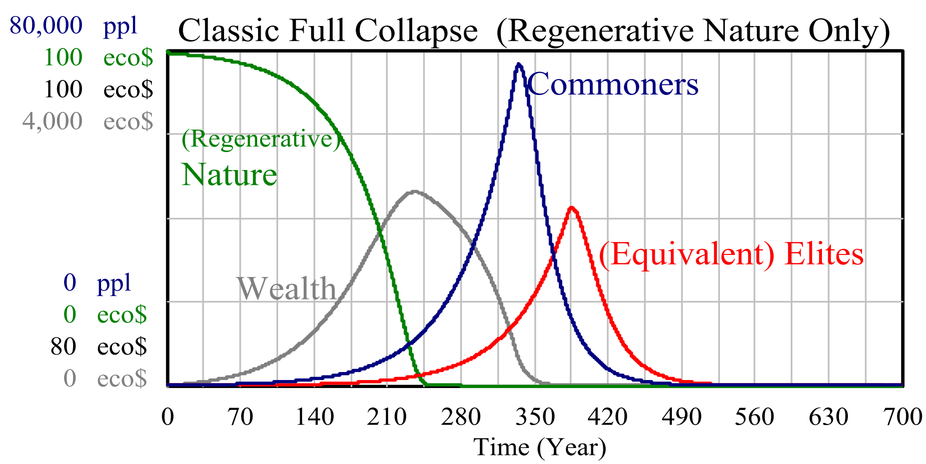
If we reduce the *depletion per capita* and *inequality*, and slow down the *population growth*, it is possible to reach a steady state and survive well.



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

- Reduce depletion per capita
- Reduce inequality ($\kappa = 10$) (as estimated by Daly)
- Reduce birth rates

Could a collapse be prevented if we have large stocks of Nonrenewable Energy?



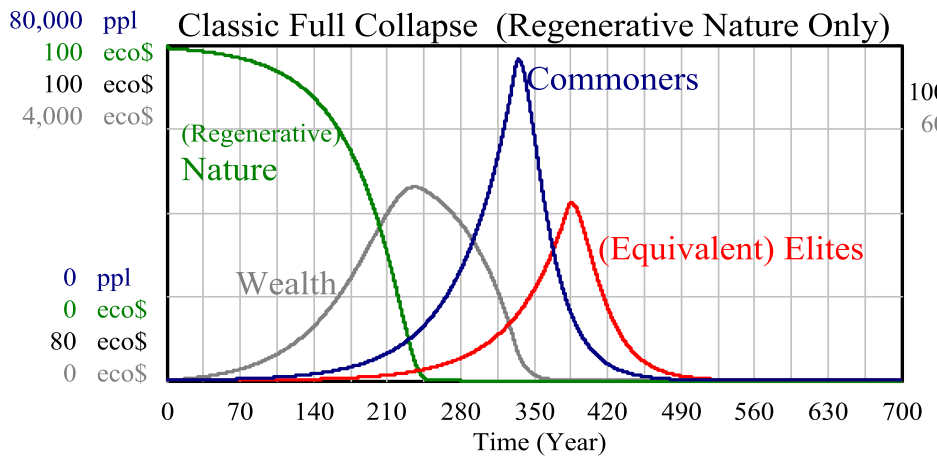
What happens when we add fossil fuels?

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

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

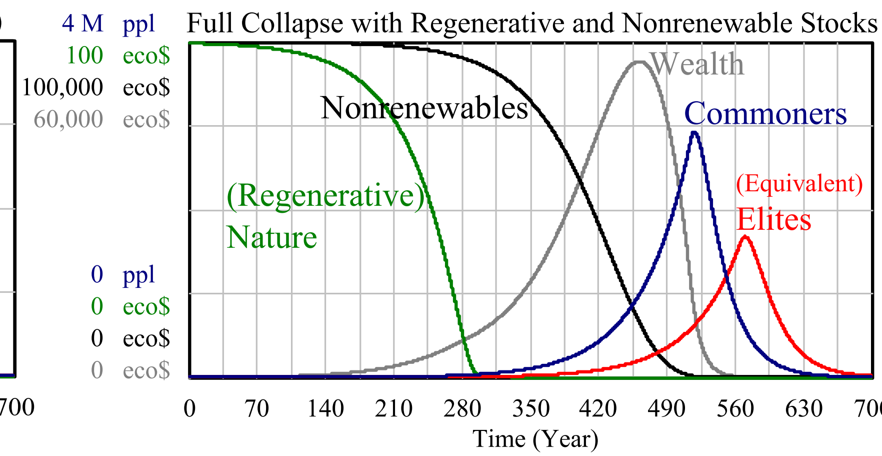
Impact of adding fossil fuels (nonrenewable energy resources)

80K



Regenerating Nature Only

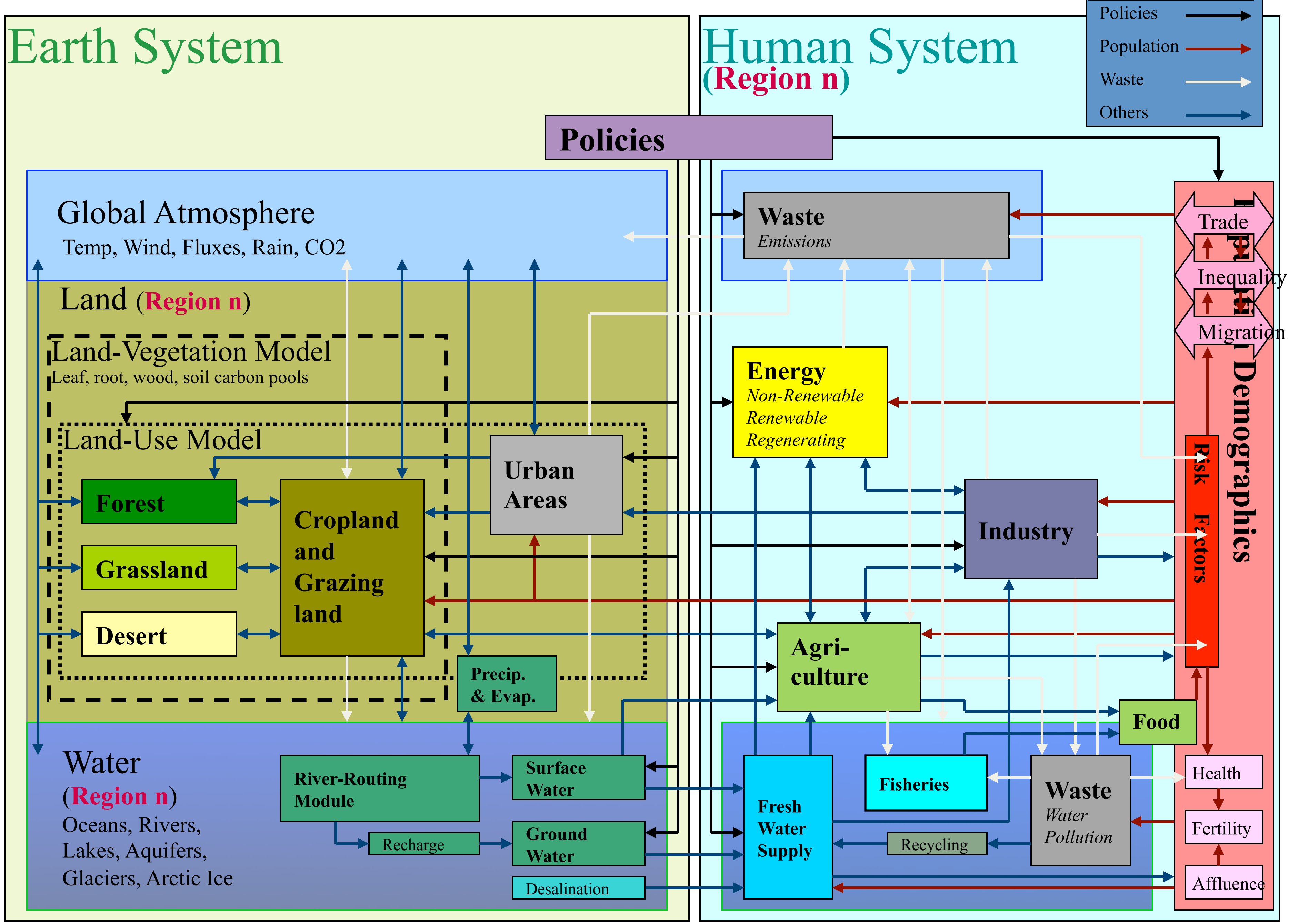
4Million



Both Regenerating and **Nonrenewable** Resources

The collapse is postponed by ~200 years and the peak population increases by a factor of ~20! Reminiscent of the Industrial Revolution!

Schematic of Earth System - Human System Feedbacks



Summary

- We are using up in 200+ years the fossil fuels that nature accumulated over millions of years
- The use of fossil fuels for agriculture increased food production and population after 1950.
- HANDY I "thought experiments" show that reducing:
  1. Social inequality
  2. Population growth
  3. Depletion per capita allow society to become sustainable.
- HANDY II: Adding non-renewables
  1. Increases maximum population by ~20 times.
  2. Postpones collapse by about 200-300 years
  3. If the transition from fossil to renewables is done early enough, it is possible to avoid the collapse.

We are NOT modeling the coupled Earth-Human System

- We need to couple them to provide feedbacks
- Data assimilation can help tune the big coupled models