Challenges in Inverse Modeling and Data Assimilation of Atmospheric Constituents

Ave Arellano
Atmospheric Chemistry Division
National Center for Atmospheric Research
NCAR is sponsored by the National Science Foundation

with input from Monika Kopacz
Woodrow Wilson School of International and Public Affairs
Princeton University

the chemical data assimilation problem

Our goal is to combine observations and models to improve our estimate of the state (constituents) of a system (atmosphere) .

from a Bayesian framework:

(state given observations) α (observations given state) (state)

$[\,\mathbf{x} \mid \mathbf{Y}\,] \quad \alpha \quad [\,\mathbf{Y} \mid \mathbf{x}\,] \quad [\,\mathbf{x}\,]$

(posterior) α (likelihood) (prior)

the inverse problem

Our goal is to combine observations and models to improve our estimate of the sources (of constituents) in the system (atmosphere) .

from a Bayesian framework:

(source given observations of the state)

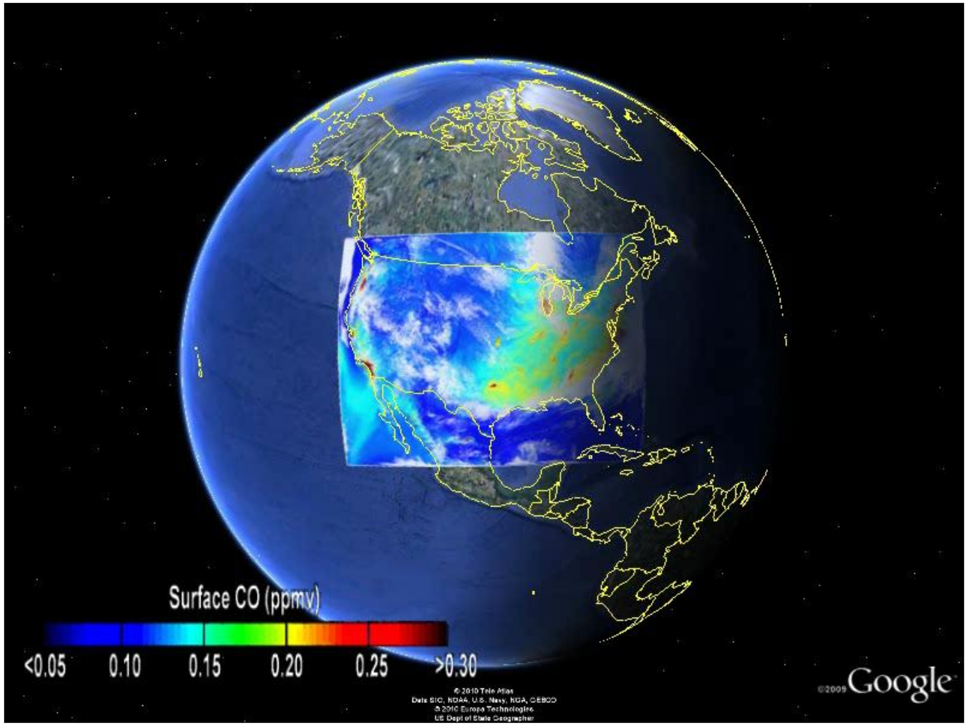
α (observations given the source) (source)

[**x** | **Y**]	α	[Y | **x**]	[**x**]
(posterior)	α	(likelihood)	(prior)

our model of the system

estimate = M(state) + error

$$\mathbf{x}_{t+1} = \mathbf{M}(\mathbf{x}_t) + \eta_t \text{ with } \eta_t \sim N(0,Q)$$



from a 4km WRF-Chem
North American Monsoon (NAM) Simulation
(~July/Aug 2006)
P.I.s Mary Barth/Alma Hodzic
http://acd.ucar.edu/~barthm/namcase.html

In the case of [CO]:

$-\vec{v} \cdot \nabla[CO]$

natural /anthropogenic combustion-related processes

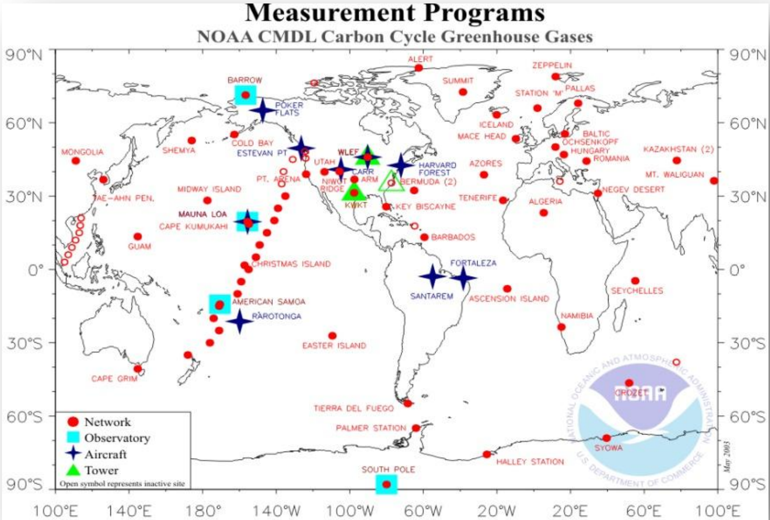
$$\frac{d[CO]}{dt} = \left(\frac{\partial[CO]}{dt}\right)_{transport} + \left(\frac{\partial[CO]}{dt}\right)_{emissions} + \left(\frac{\partial[CO]}{dt}\right)_{chemistry} + \left(\frac{\partial[CO]}{dt}\right)_{deposition}$$

$-k_{CO-OH}[CO][OH]$

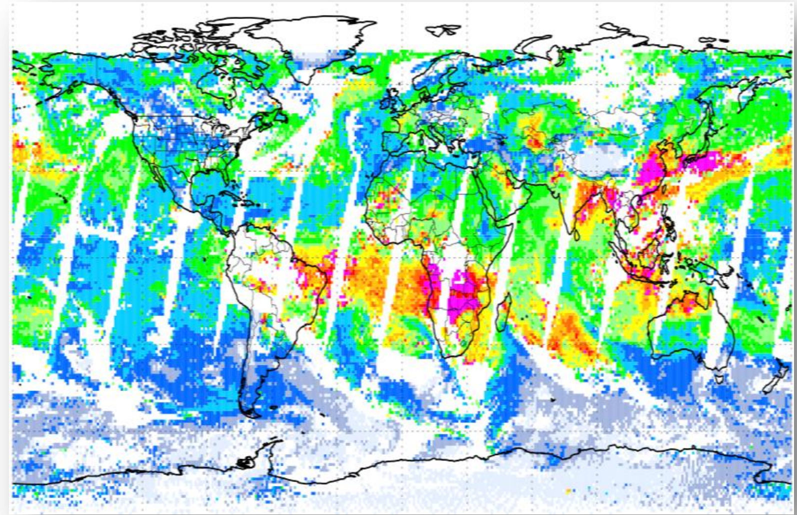
$k_{CH_4-OH}[CH_4][OH]$

observations of the system

In-Situ Measurements

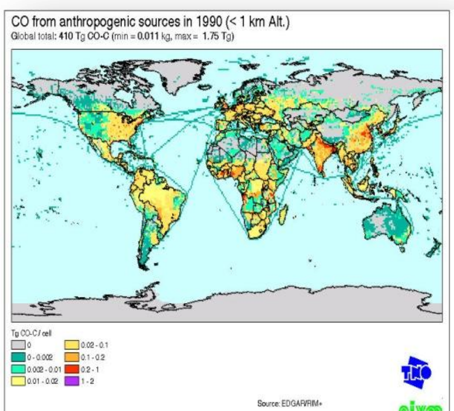


Remote-sensed Measurements



e.g. Terra MOPITT, Aura TES, Envisat SCIAMACHY, Aqua AIRS, Aura MLS, ACE FTS, Metop IASI

Ancillary Data



Source Inventories

Fire Data

Aircraft Measurements

DA problem

observation = h(state) + error

$$\mathbf{Y} = \mathrm{h}\,(\mathbf{x}) + \boldsymbol{\varepsilon} \quad \text{with } \boldsymbol{\varepsilon} \sim N\,(\mathbf{0}, \mathbf{R})$$

$$\mathbf{Y} = \mathbf{Hx} + \boldsymbol{\varepsilon} \quad \text{linear case}$$

In the case of direct observations of CO, $\mathbf{H}$ is simply a linear interpolation of the model CO state to the observation

However, in the case of remote-sensed measurements of CO, $\mathbf{H}$ can be more complicated (may not be non-linear too!). For example,

$$\mathbf{y}_{\mathrm{radiance}} = \mathrm{h}(\mathbf{x}) + \mathbf{e}_{\mathrm{y}}$$ where $\mathrm{h}\,(\,)$ can be a radiative transfer model (RTM)

$$\mathbf{y}_{\mathrm{radiance}} = \frac{\delta\mathbf{y}}{\delta\mathbf{x}}\mathbf{x} + \mathbf{e}_{\mathrm{y}}$$ linear case where $\frac{\delta y}{\delta x}$ is the measurement sensitivity

$$\mathbf{y}_{\mathrm{retrieval}} = \hat{\mathbf{x}} = \mathbf{x}_{\mathrm{a}} + \frac{\delta\hat{\mathbf{x}}}{\delta\mathbf{x}}(\mathbf{x} - \mathrm{x}_{\mathrm{a}}) + \mathbf{e}_{\mathrm{x}}$$ where $\mathbf{A} = \frac{\delta\hat{\mathbf{x}}}{\delta\mathbf{x}}$ is the averaging kernel

observation and model

Inverse problem

observation = h(source) + error

$$\mathbf{Y} = \mathrm{h}\,(\mathbf{x}) + \boldsymbol{\varepsilon} \qquad \text{with } \boldsymbol{\varepsilon} \sim N\,(\mathbf{0}, \mathbf{R})$$

$$\mathbf{Y} = \mathbf{H}\mathbf{x} + \boldsymbol{\varepsilon} \qquad \text{linear case}$$

In the case of direct observations of CO, $\mathbf{H}$ is matrix of response functions (mapping the source to the state) + a linear interpolation of the model CO state to the observation

However, in the case of remote-sensed measurements of CO, $\mathbf{H}$ can be more complicated (may not be non-linear too!). For example,

$$\mathbf{y}_{\mathrm{radiance}} = \mathrm{h}(\mathbf{x}) + \mathbf{e}_{\mathrm{y}}$$ where $\mathrm{h}\,(\,)$ can be a radiative transfer model (RTM)

$$\mathbf{y}_{\mathrm{radiance}} = \frac{\delta \mathbf{y}}{\delta \mathbf{x}}\mathbf{x} + \mathbf{e}_{\mathrm{y}}$$ linear case where $\frac{\delta y}{\delta x}$ is the measurement sensitivity

$$\mathbf{y}_{\mathrm{retrieval}} = \hat{\mathbf{x}} = \mathbf{x}_{\mathrm{a}} + \frac{\delta \hat{\mathbf{x}}}{\delta \mathbf{x}}(\mathbf{x} - \mathrm{x}_{\mathrm{a}}) + \mathbf{e}_{\mathrm{x}}$$ where $\mathbf{A} = \frac{\delta \hat{\mathbf{x}}}{\delta \mathbf{x}}$ is the averaging kernel

the inverse problem

$[\,\mathbf{x} \mid \mathbf{Y}\,]$ (posterior) $[\,\mathbf{Y} \mid \mathbf{x}\,]$ (likelihood) $[\,\mathbf{x}\,]$ (prior)

is $N(\mathbf{x}^f, \mathbf{P}^f)$ and $\mathbf{Y} = \mathbf{Hx} + \varepsilon$ with $\varepsilon \sim N(0,\mathbf{R})$, the posterior is $N(x^a,P^a)$

$$J_x = \left(y - Hx^f\right)^t R^{-1}\left(y - Hx^f\right) + \left(x - x^f\right)^t {P^f}^{-1}\left(x - x^f\right)$$

An estimate $\mathbf{x}^a$ of the state can be expressed as:

$$x^a = \left(H^t R^{-1} H + {P^f}^{-1}\right)^{-1}\left(H^t R^{-1} y + {P^f}^{-1} x^f\right), \quad P^a = \left(H^t R^{-1} H + {P^f}^{-1}\right)^{-1}$$

Kalman Filter

$$x^a = x^f + K\left(y - Hx^f\right), \quad P^a = (I - KH)P^f$$

$$x^a = x^f + P^f H^t\left(HP^f H^t + R\right)^{-1}\left(y - Hx^f\right)$$

Can be recast as:

$$x^a = x^f + \underbrace{P^f H^t\left(HP^f H^t\right)^{-1}}_{\text{least-squares}} \quad \underbrace{\left(HP^f H^t\right)\left(HP^f H^t + R\right)^{-1}}_{\text{shrink to } \mathbf{Hx}^f}\left(y - Hx^f\right)$$

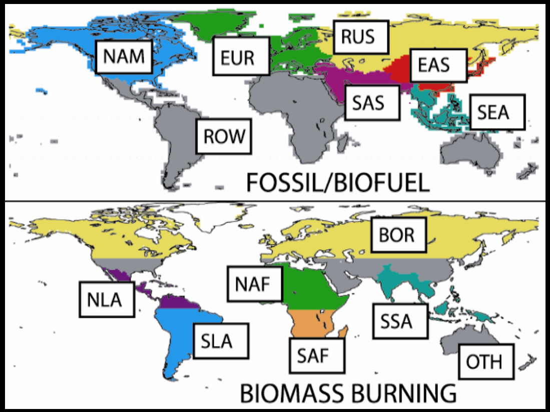
inverse modeling of CO sources

e.g. [emission | MOPITT] ~ [MOPITT | emission] [emission]

$N(x^a, P^a)$ $N(Hx, \mathbf{R})$ $N(\mathbf{x}^f, \mathbf{P}^f)$

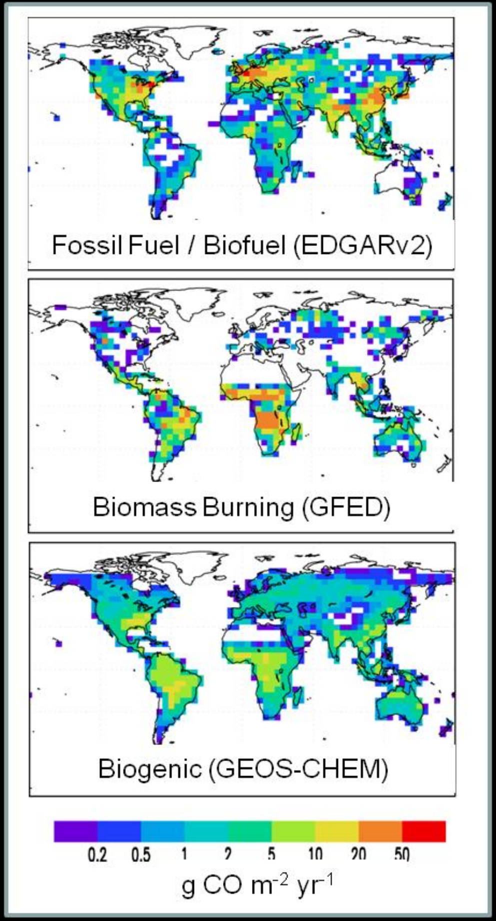
We solve for regional source scaling factors

e.g.



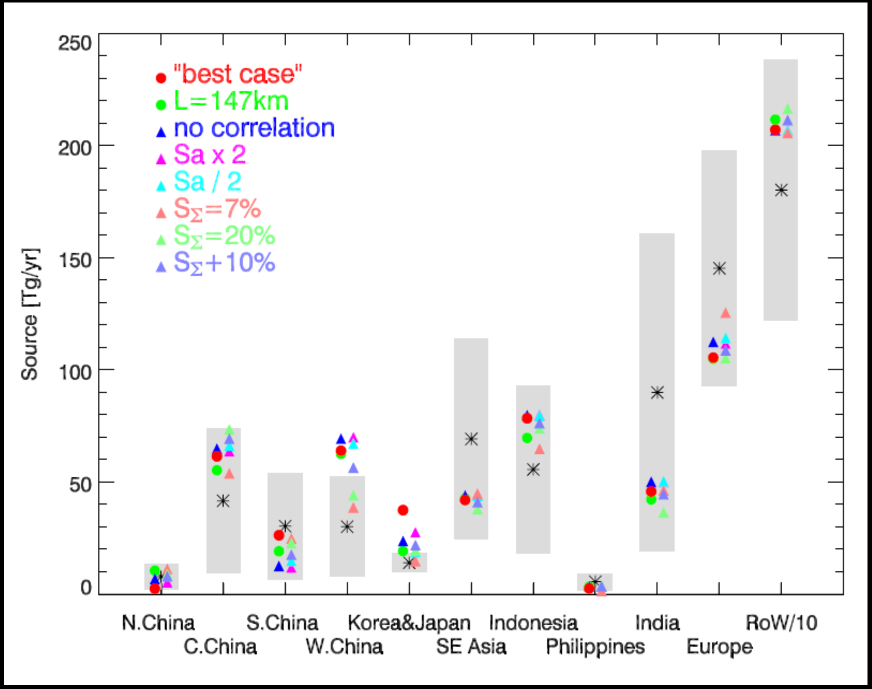
Note that the posterior mean is sensitive to data, prior estimates, error specification and model leading to persistent discrepancies in source estimates

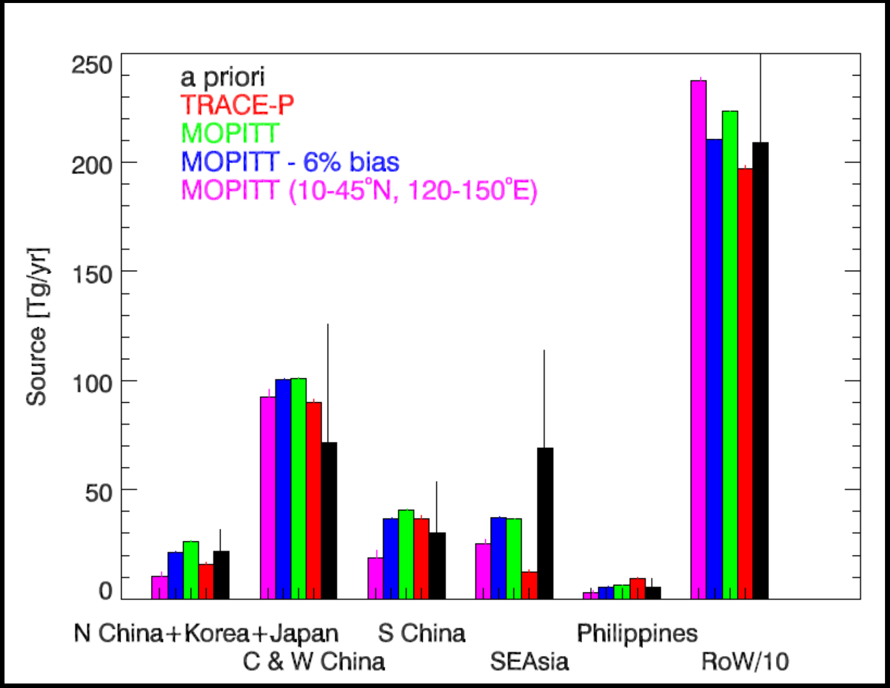
(No TransCom framework for CO)



ensemble of inversions

Sensitivity Tests of Error Assumptions and Obs Data Choice

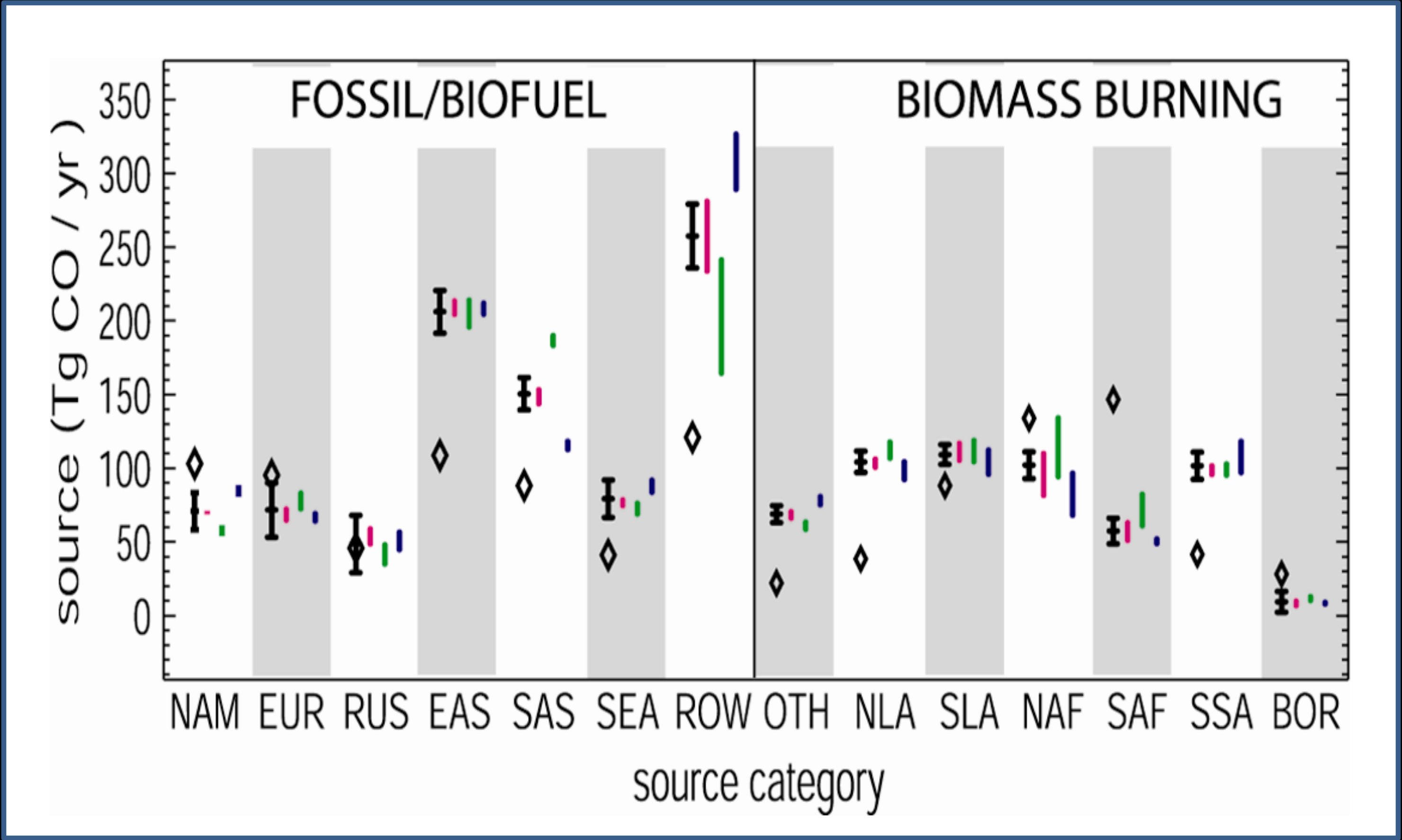




Heald et al. (2004)

ensemble of inversions

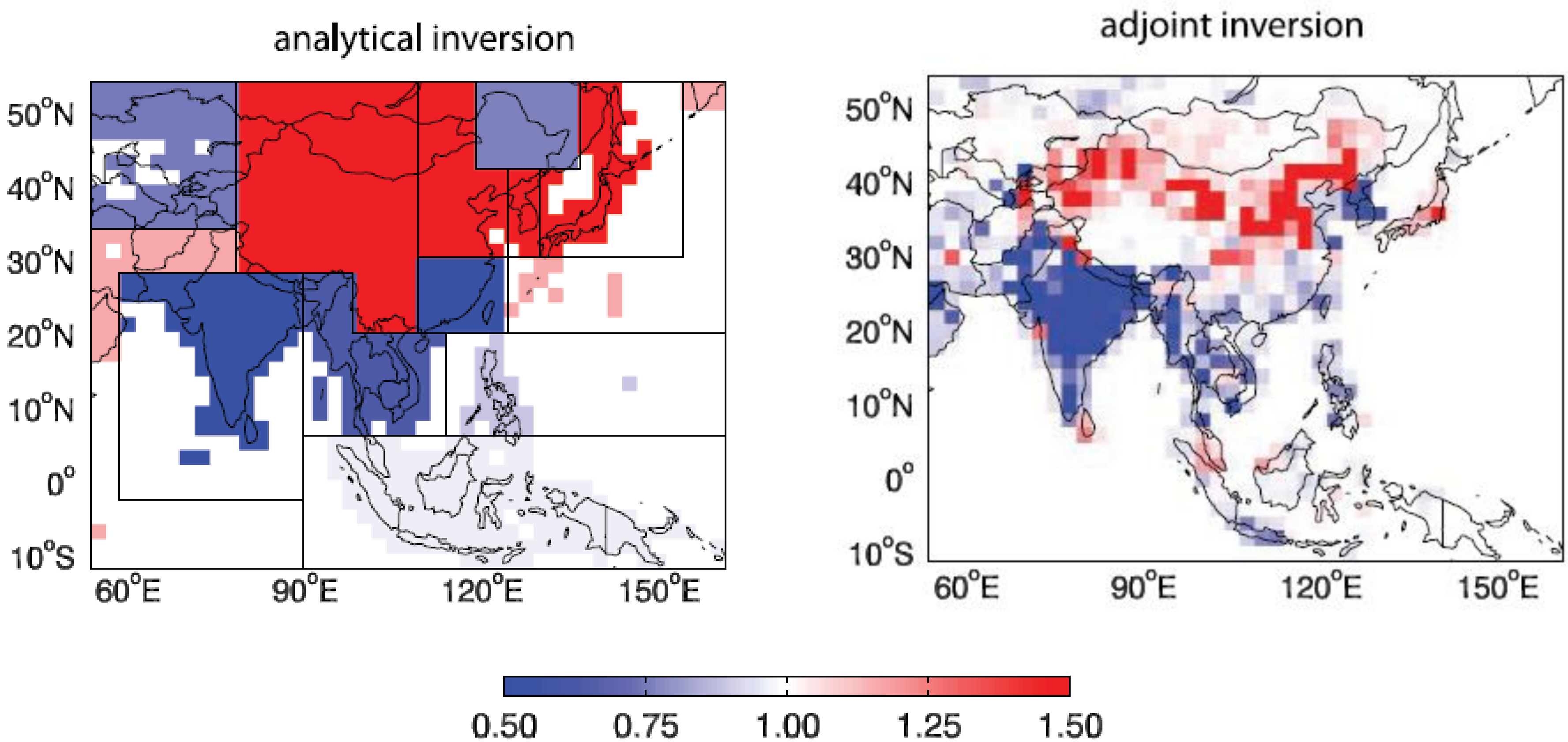
Sensitivity Tests of Error Assumptions and Obs Data Choice



Arellano et al. (2004)

sensitivity to methodology

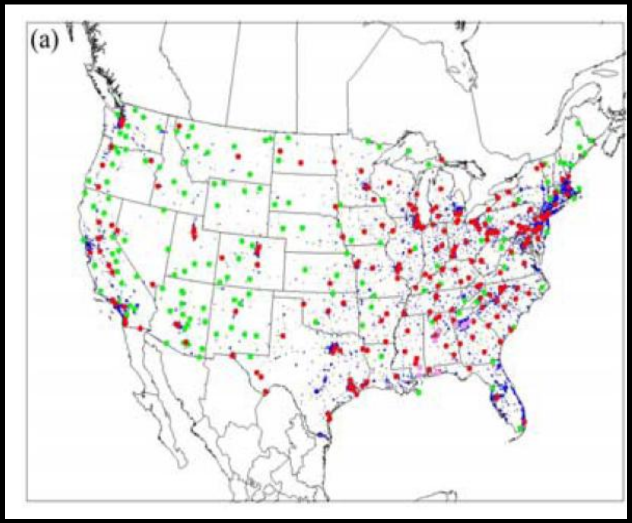
Comparison between Adjoint and Analytical Solution

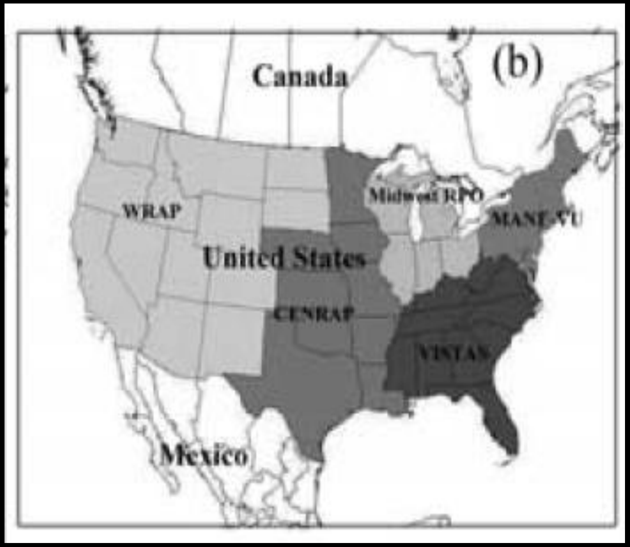


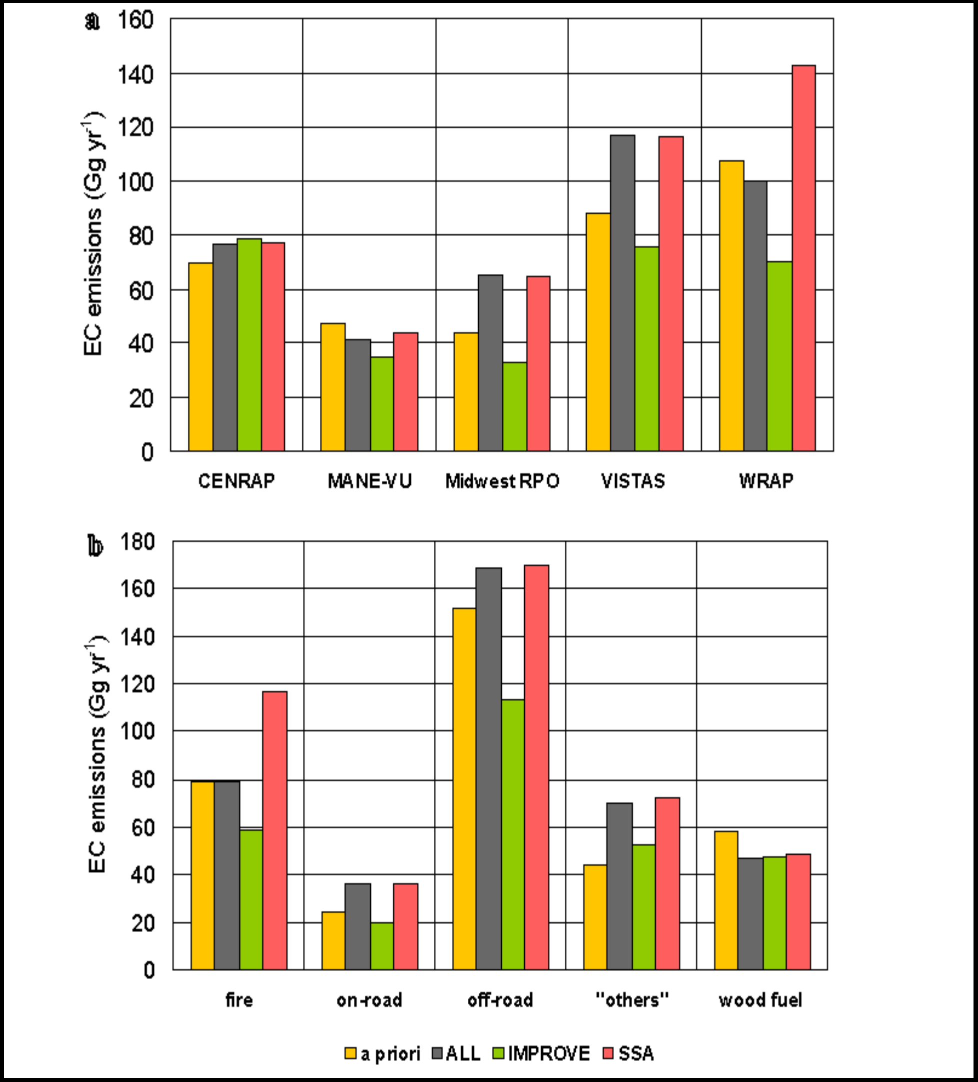
Kopacz et al. (2009)

sensitivity to observing network

Top-down Estimates of Black Carbon



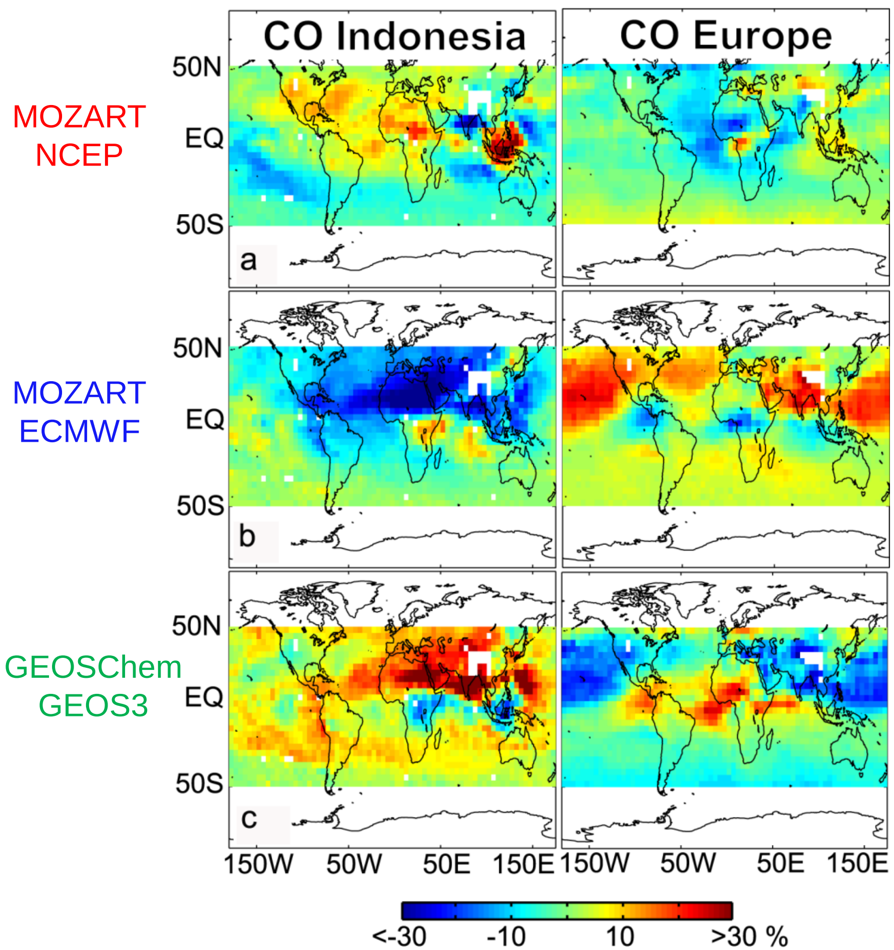




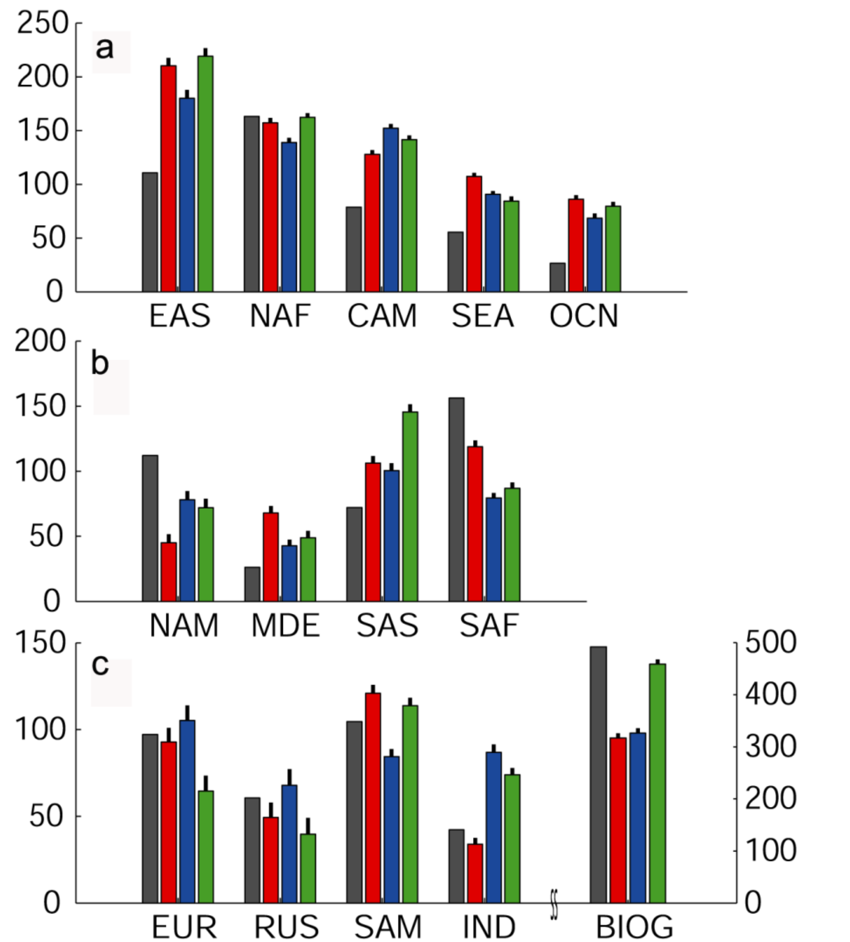
Hu et al. (2009)

sensitivity to GCTM transport

differences in response functions



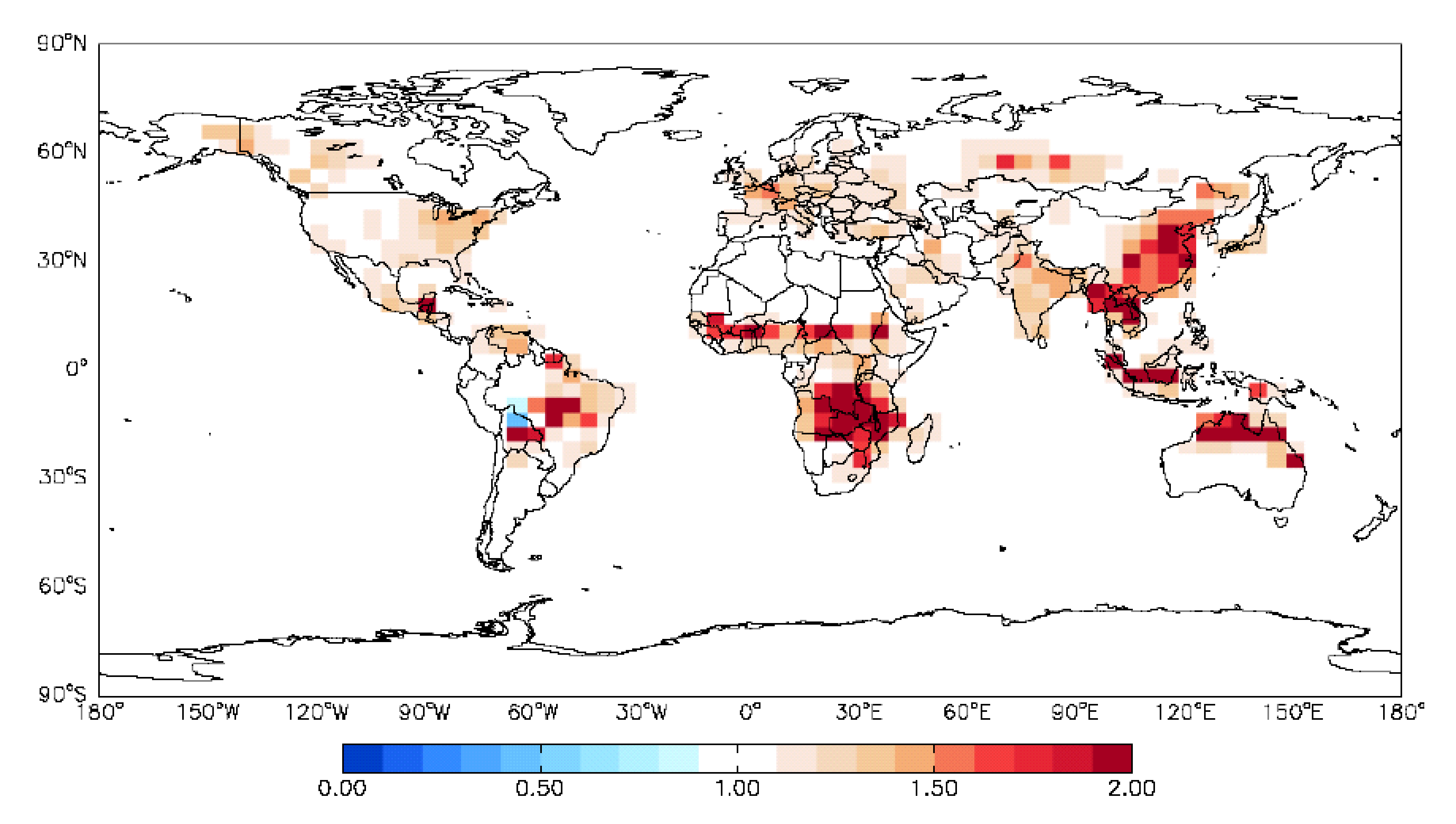
differences in inverse results



Arellano and Hess, (2006)

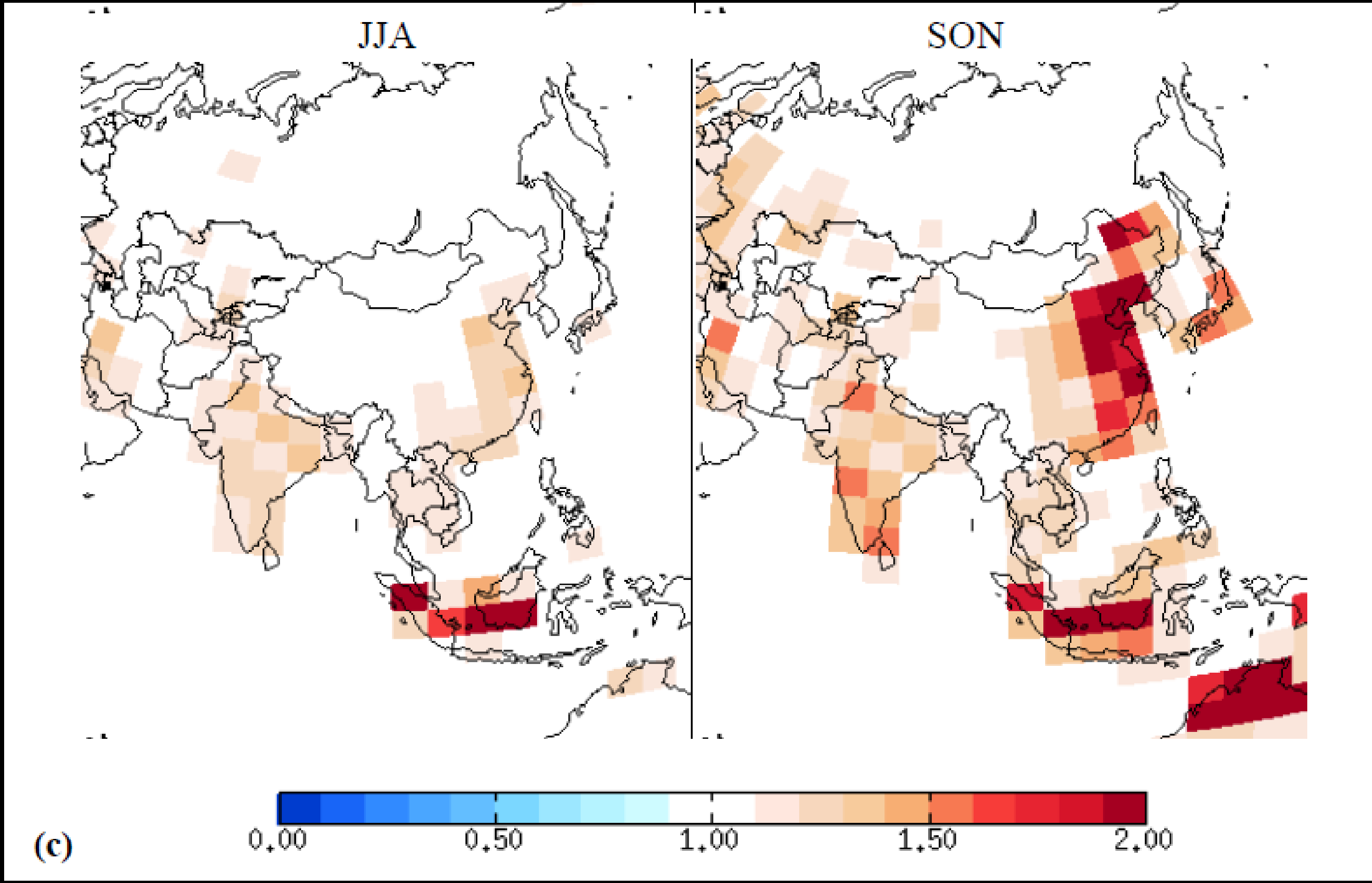
recent top-down estimates of CO sources

annual estimates of scaling factors



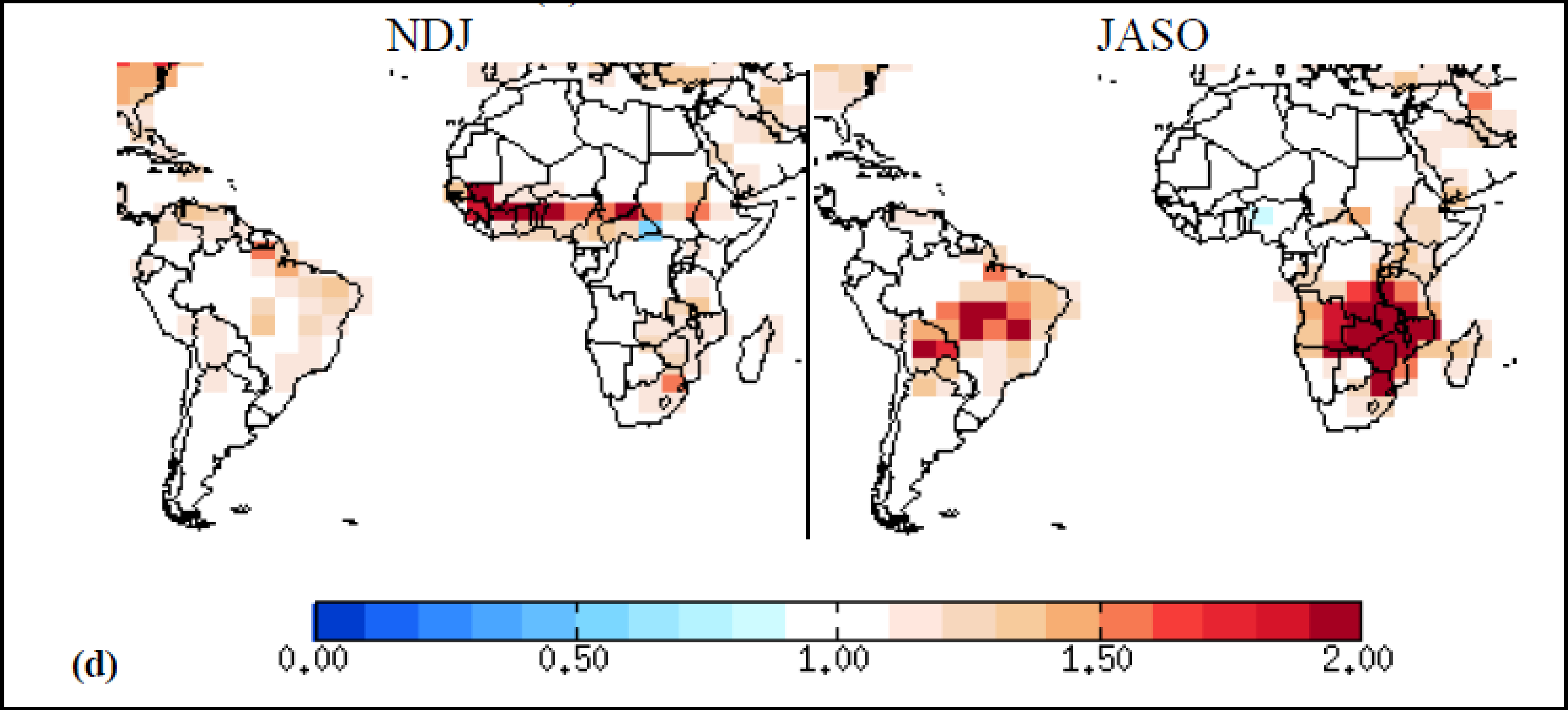
Kopacz et al., (2010)

recent top-down estimates of CO sources



Kopacz et al., (2010)

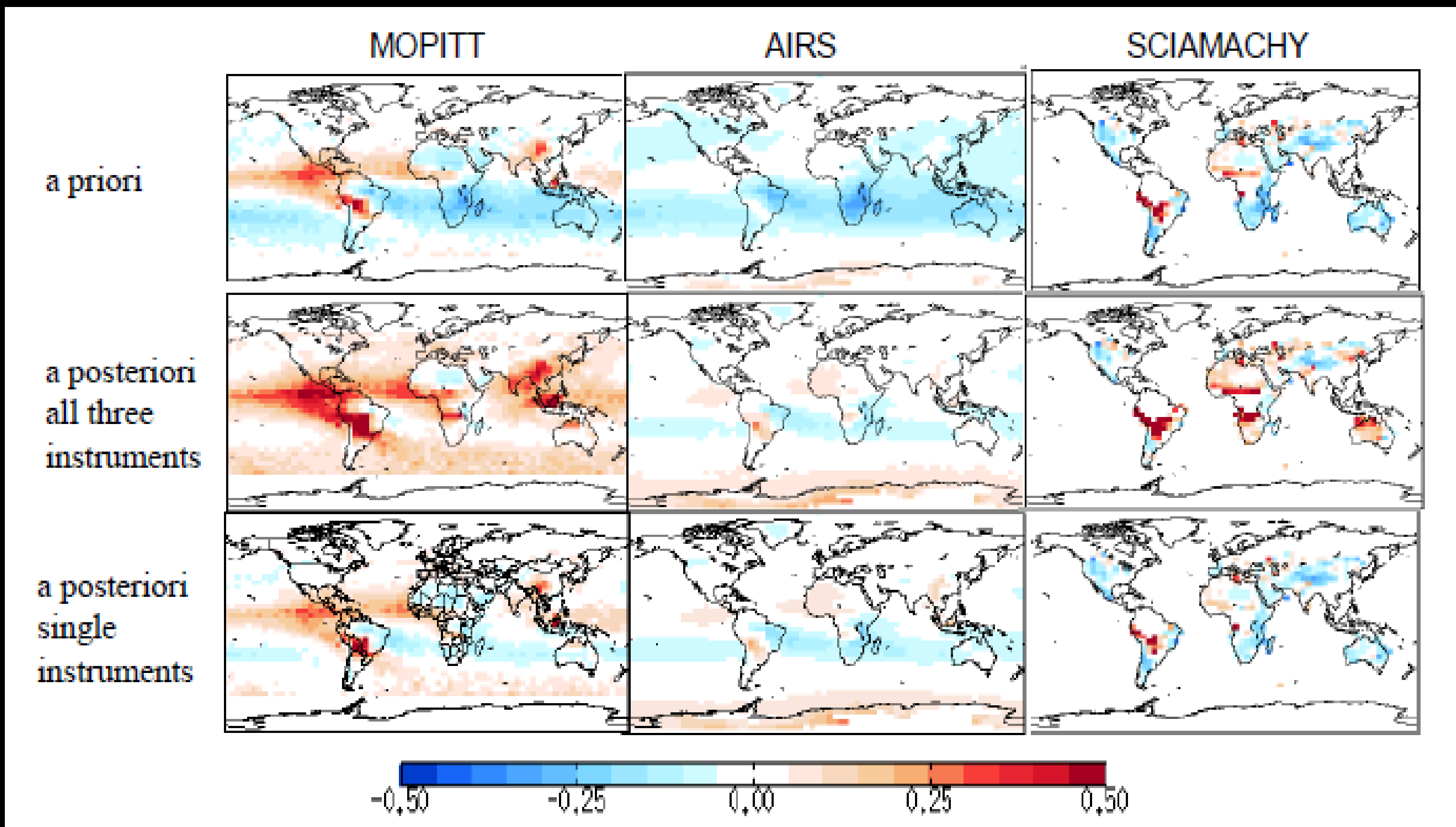
recent top-down estimates of CO sources



Kopacz et al., (2010)

problems?

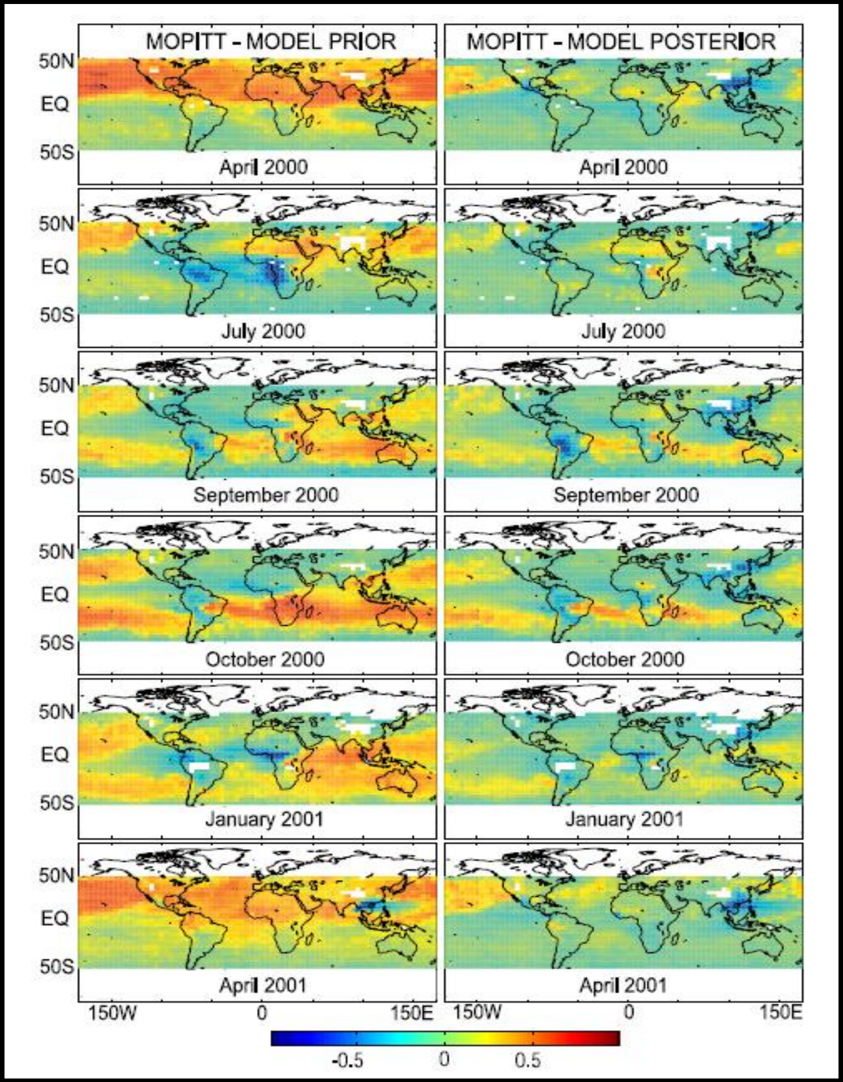
posterior CO still exhibit large biases



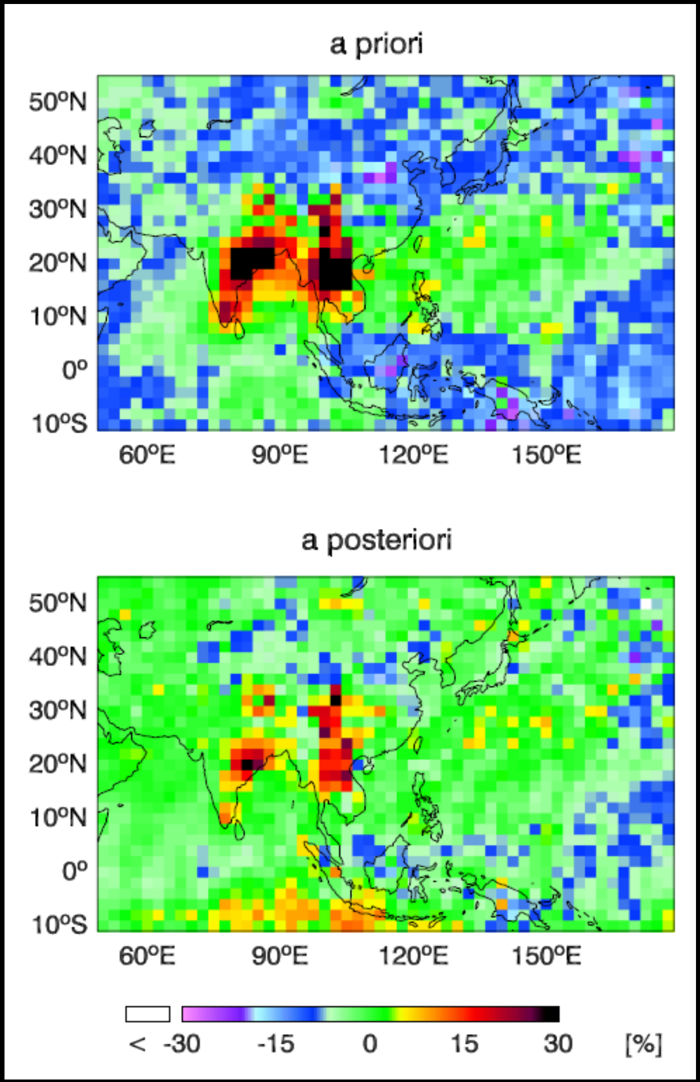
Kopacz et al., (2010)

problems?

posterior CO still exhibit large biases



Arellano et al. (2006)



Heald et al. (2004)

recommendations

1) Inversions with a regional focus

- reconciling differences in inverse results
- reducing aggregation errors

2) TransCom-like CO experiments

- understanding errors in GCTM
- elucidating differences in methodology (perfect model experiments)

3) Towards a more integrated/synergistic approach

- multi-species/multi-platform inversions
- state/source estimation

4) Better error characterization

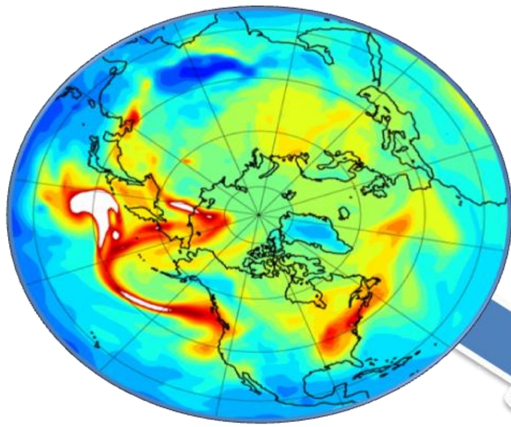
- observational errors (biases/inconsistencies)
- model errors
- emission errors

chemical data assimilation system @ NCAR

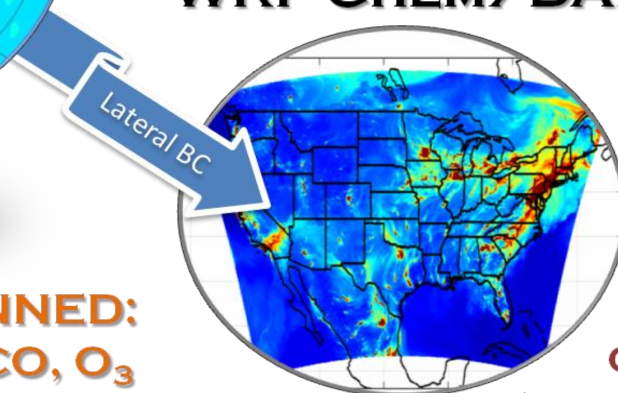
Research-Based Regional to Global NWP with Chemistry

- leveraging on state-of-the-art community models (CAM-Chem, WRF-Chem) and Community DA facility (DART)

CAM-Chem/DART



WRF-Chem/DART



Observations

Current:	Planned:
Rawinsondes	TES CO, O_3
ACARS	IASI CO, O_3
Aircraft	OMI NO_2, O_3
Satwinds	GOME O_3
MOPITT CO	
MODIS AOD	

Applications

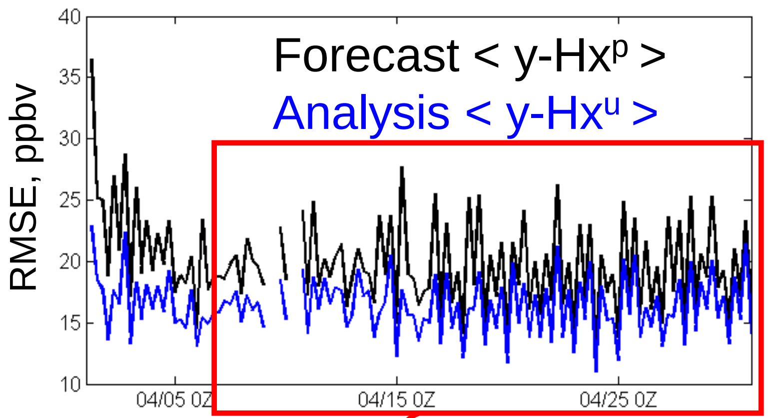
Chemical OSSEs
Chemical Weather
Field campaign support

In collaboration with NCAR/IMAGe & NCAR/MMM

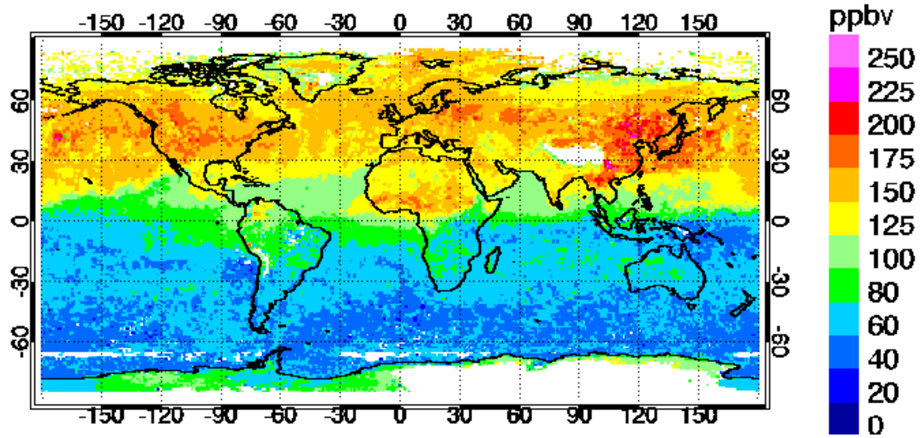
utility of a chemical DA system

capability to evaluate the forecast model

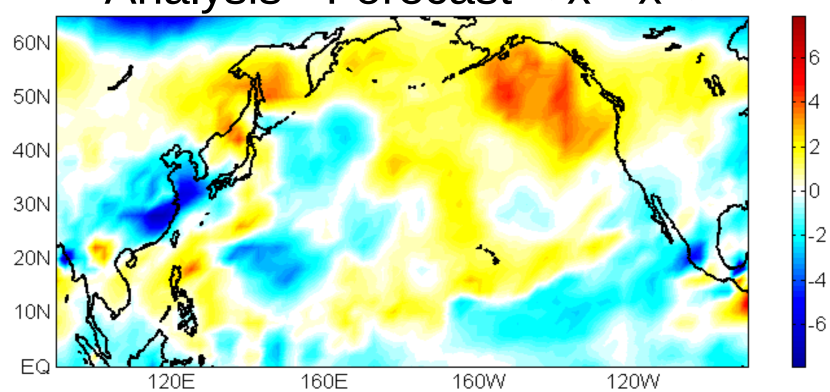
RMSE relative to MOPITT



MOPITT CO @ 700 hPa (April 2006)



Analysis - Forecast $< x^u - x^p >$

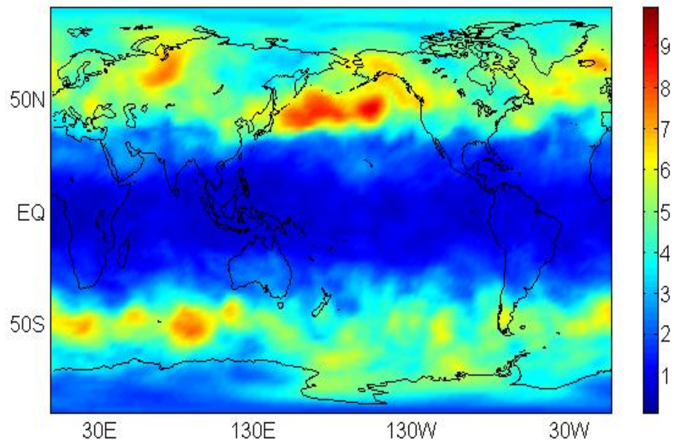


system diagnostic for short-term (6-hourly) forecast errors

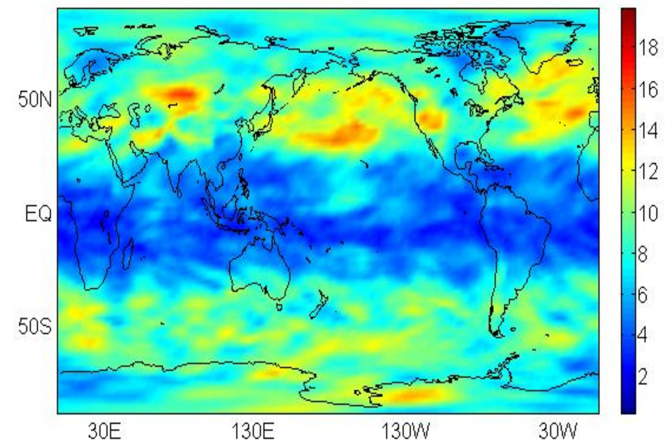
model overpredicts in source region while underpredicts in downwind region

error characterization using ensembles

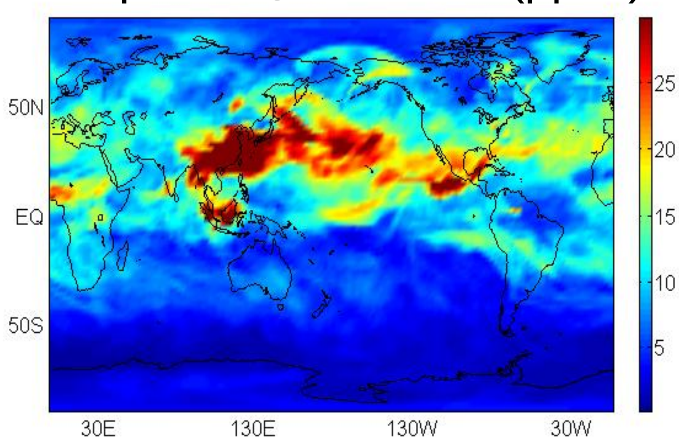
Temperature Spread @ 500 hPa



U Wind Spread @ 500 hPa (m/s)



CO Spread @ 500 hPa (ppbv)



CO Emissions Spread (%)