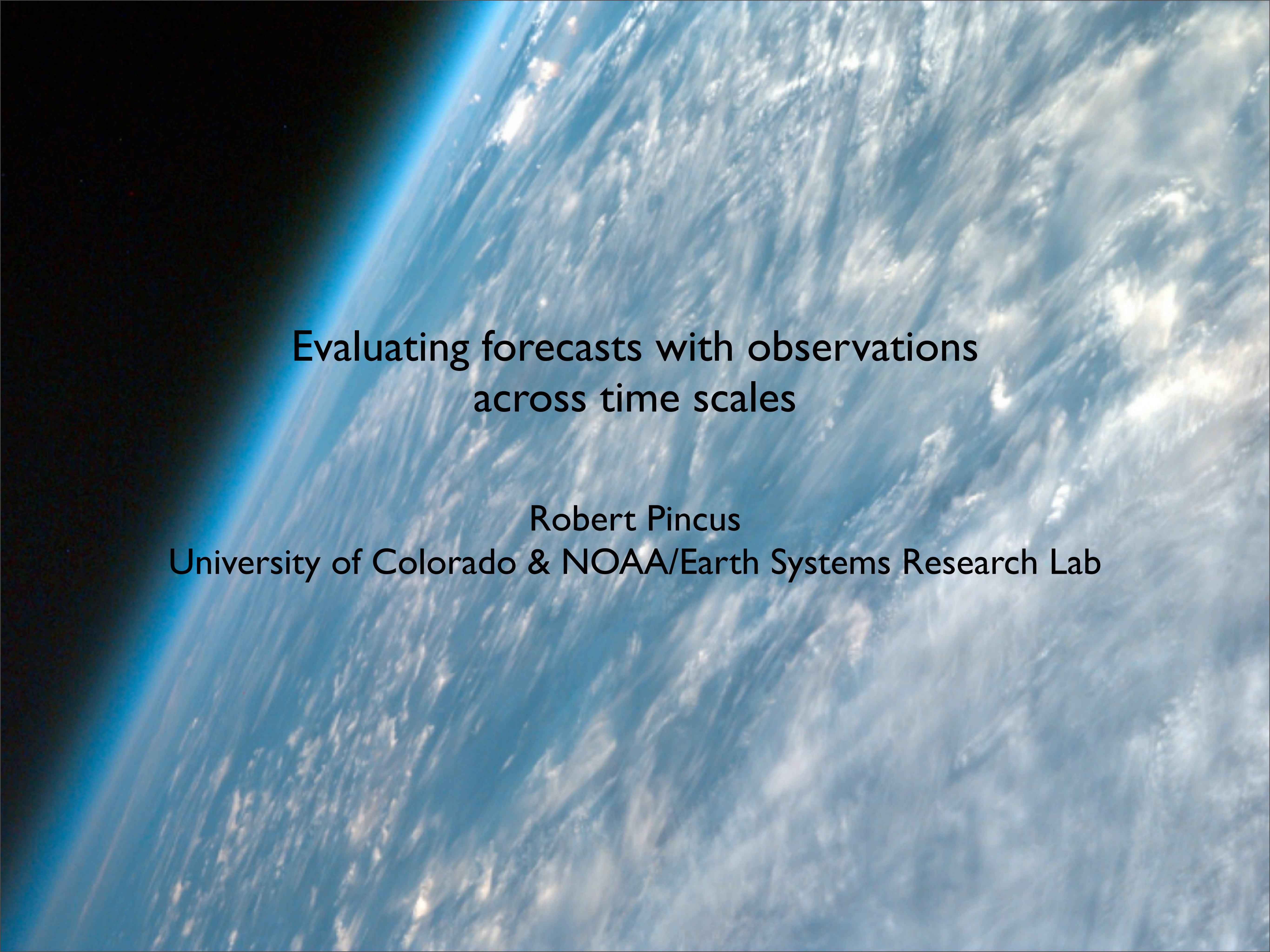
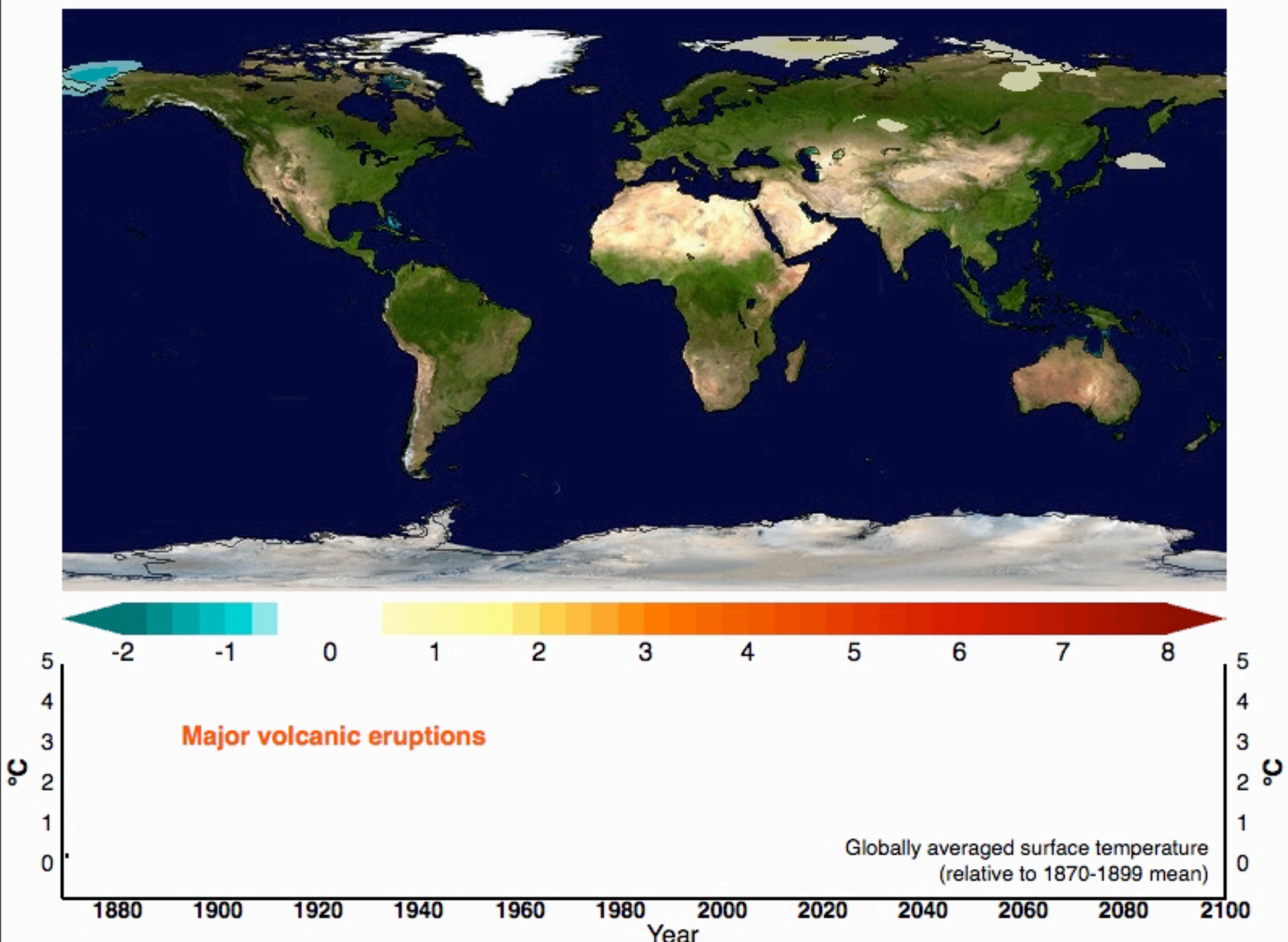
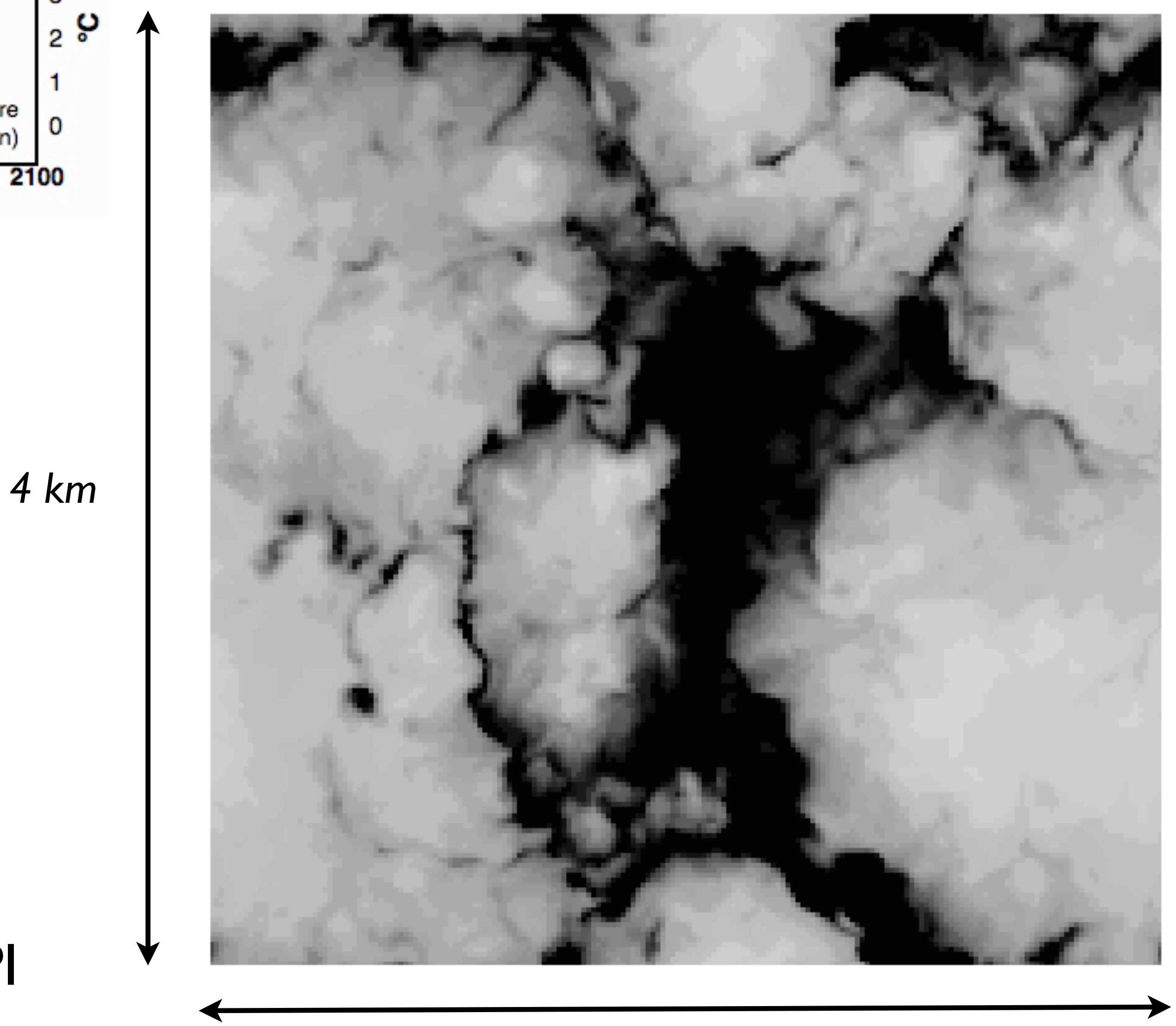


Evaluating forecasts with observations across time scales

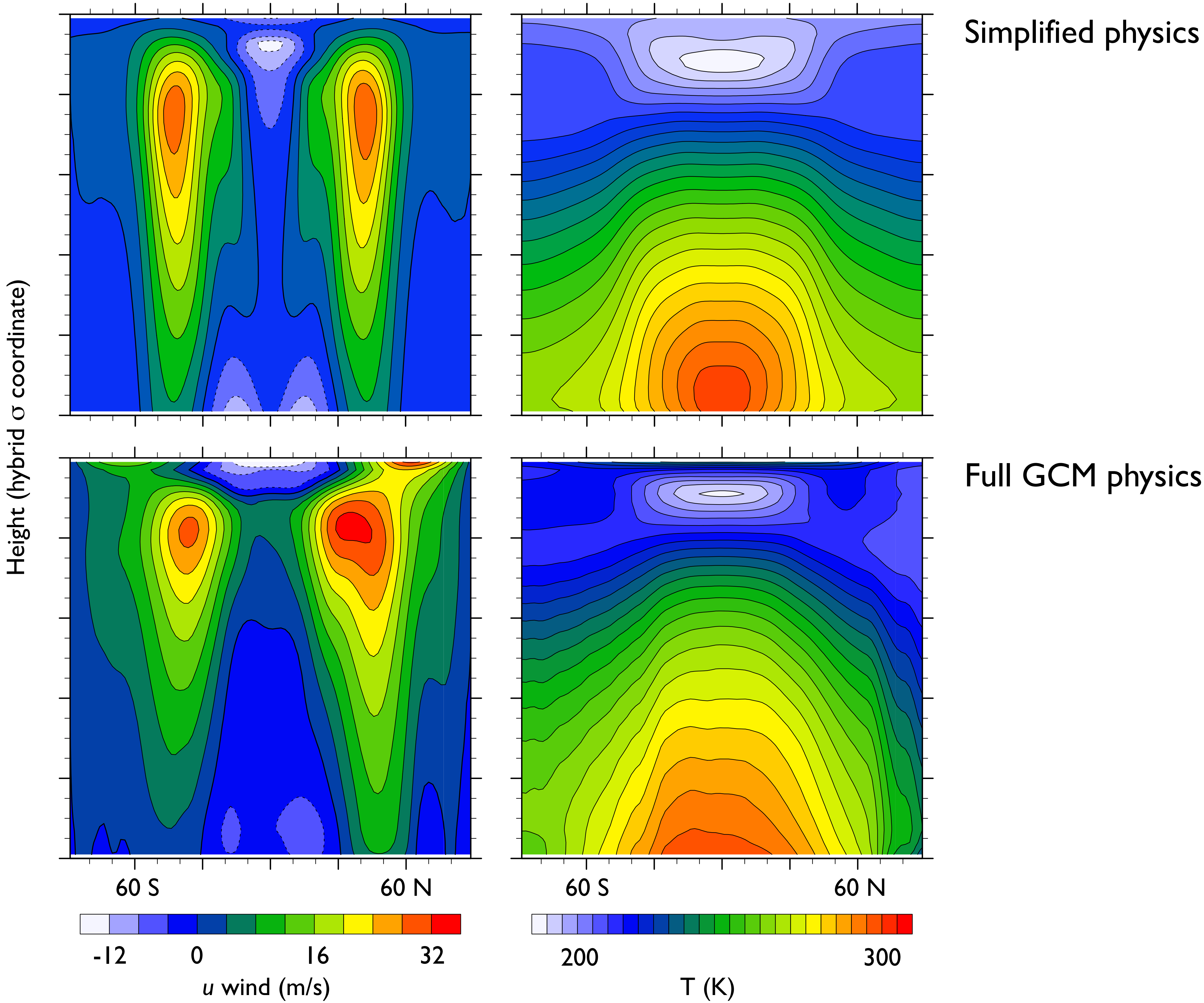
Robert Pincus
University of Colorado & NOAA/Earth Systems Research Lab

Gary Strand, NCAR





Bjorn Stevens, MPI



Hui Wan, Max-Planck-Institute

Observing the system *in situ*

To evaluate the models we need to have relevant measurements

State variables are momentum (winds), energy (temperature), and tracer masses

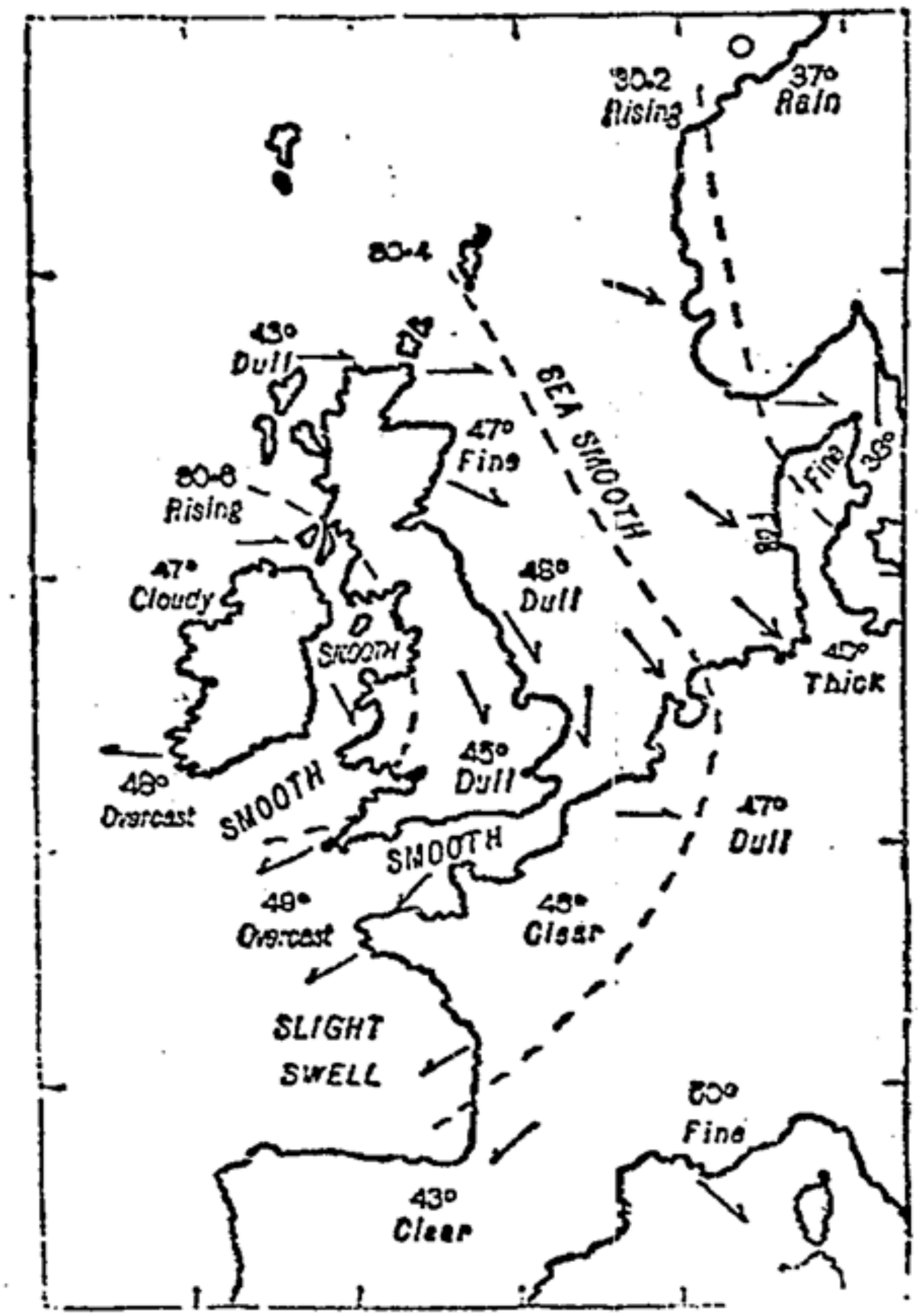
The most important tracer is water, divided among various states including two(+) flavors of condensate

Climate-relevant observations are long term, geographically diverse, with well-understood sampling

Observing the system *in situ*

Go outside.

WEATHER CHART, MARCH 31, 1875.

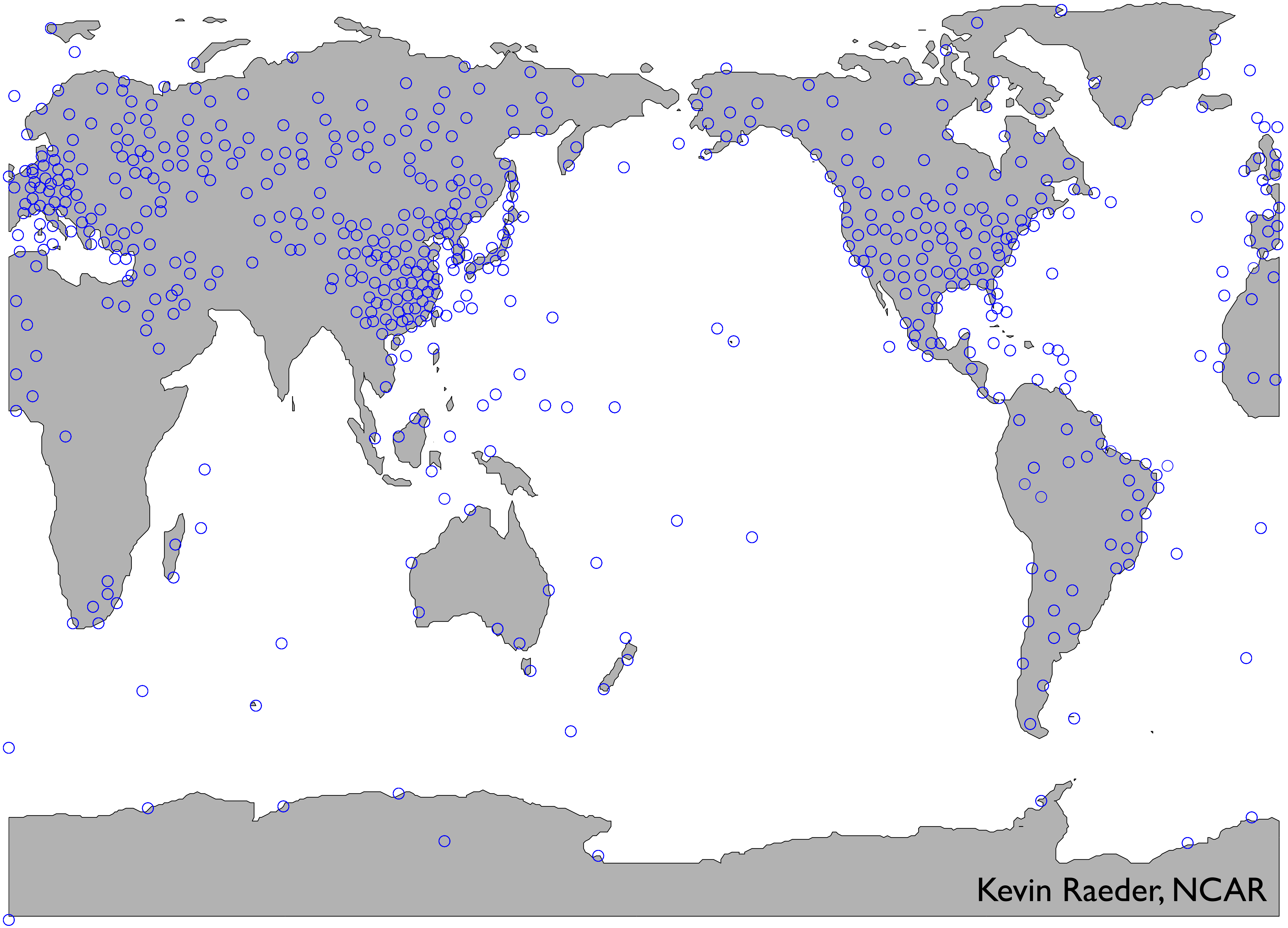


The dotted lines indicate the gradations of barometric pressure The variations of the temperature are marked by figures, the state of the sea and sky by descriptive words, and the direction of the wind by arrows—barbed and feathered according to its force. ⊙ denotes calm.

Image: http://galton.org

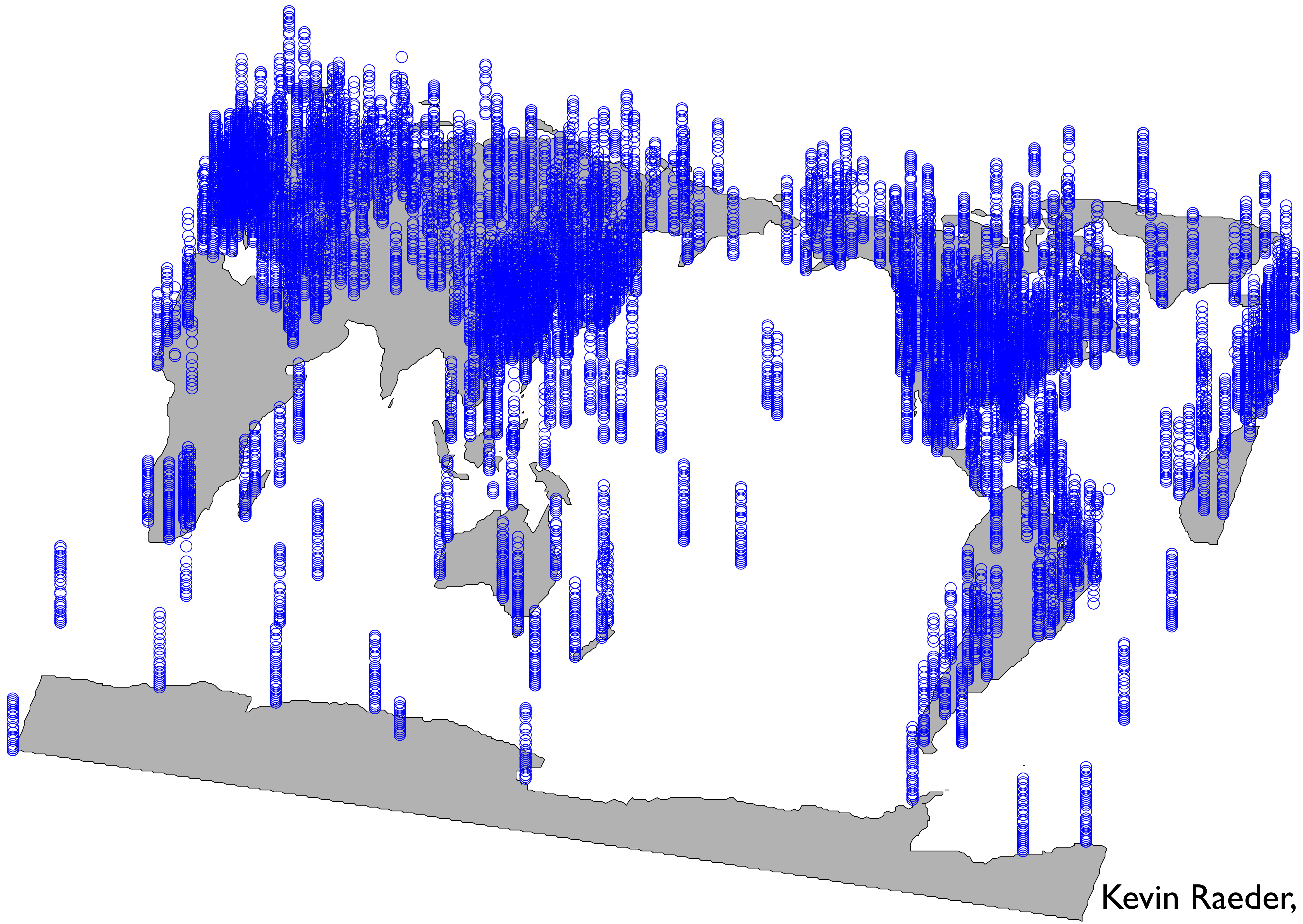

NOAA Photo Library

Radiosonde locations, 1 Dec 2006, 6 hr window around 12Z



Kevin Raeder, NCAR

Radiosonde locations, 1 Dec 2006, 6 hr window around 12Z



Kevin Raeder, NCAR

Observing the system remotely

If you want to measure more of the atmosphere you look at it - i.e. sense it remotely using electromagnetic radiation

$$\begin{aligned}\boldsymbol{\Omega} \cdot \nabla I(\mathbf{x}, \boldsymbol{\Omega}) = \quad & - \sigma(\mathbf{x}) I(\mathbf{x}, \boldsymbol{\Omega}) \\ & + \sigma(\mathbf{x}) \frac{\omega_0(\mathbf{x})}{4\pi} \int_{4\pi} I(\mathbf{x}, \boldsymbol{\Omega}) P(\mathbf{x}, \boldsymbol{\Omega}' \to \boldsymbol{\Omega}) d\boldsymbol{\Omega}' \\ & + \sigma(\mathbf{x})(1 - \omega_0(\mathbf{x})) B(T(\mathbf{x}))\end{aligned}$$

Observing the system remotely

If you want to measure more of the atmosphere you look at it - i.e. sense it remotely using electromagnetic radiation

$$\begin{aligned}\mathbf{\Omega} \cdot \nabla I(\mathbf{x}, \mathbf{\Omega}) = & -\sigma(\mathbf{x}) I(\mathbf{x}, \mathbf{\Omega}) \\ & + \sigma(\mathbf{x}) \frac{\omega_0(\mathbf{x})}{4\pi} \int_{4\pi} I(\mathbf{x}, \mathbf{\Omega}) P(\mathbf{x}, \mathbf{\Omega}' \to \mathbf{\Omega}) d\mathbf{\Omega}' \\ & + \sigma(\mathbf{x})(1 - \omega_0(\mathbf{x})) B(T(\mathbf{x}))\end{aligned}$$

$$\begin{aligned}\mu \tfrac{dI_\lambda(\tau,\mu,\phi)}{d\tau} = & -I_\lambda(\tau, \mu, \phi) \\ & + \frac{\omega_0}{4\pi} \int_0^{2\pi} \int_{-1}^{1} P_\lambda(\mu', \phi' \to \mu, \phi) I_\lambda(\tau, \mu', \phi') d\phi' d\mu' \\ & + (1 - \omega_0) B_\lambda(T(\tau))\end{aligned}$$

Observing the system remotely

Three main characteristics

- Passive vs. active remote sensing
- Emission vs. scattering regimes
- View from above or below

Observing the system remotely (emission/absorption)

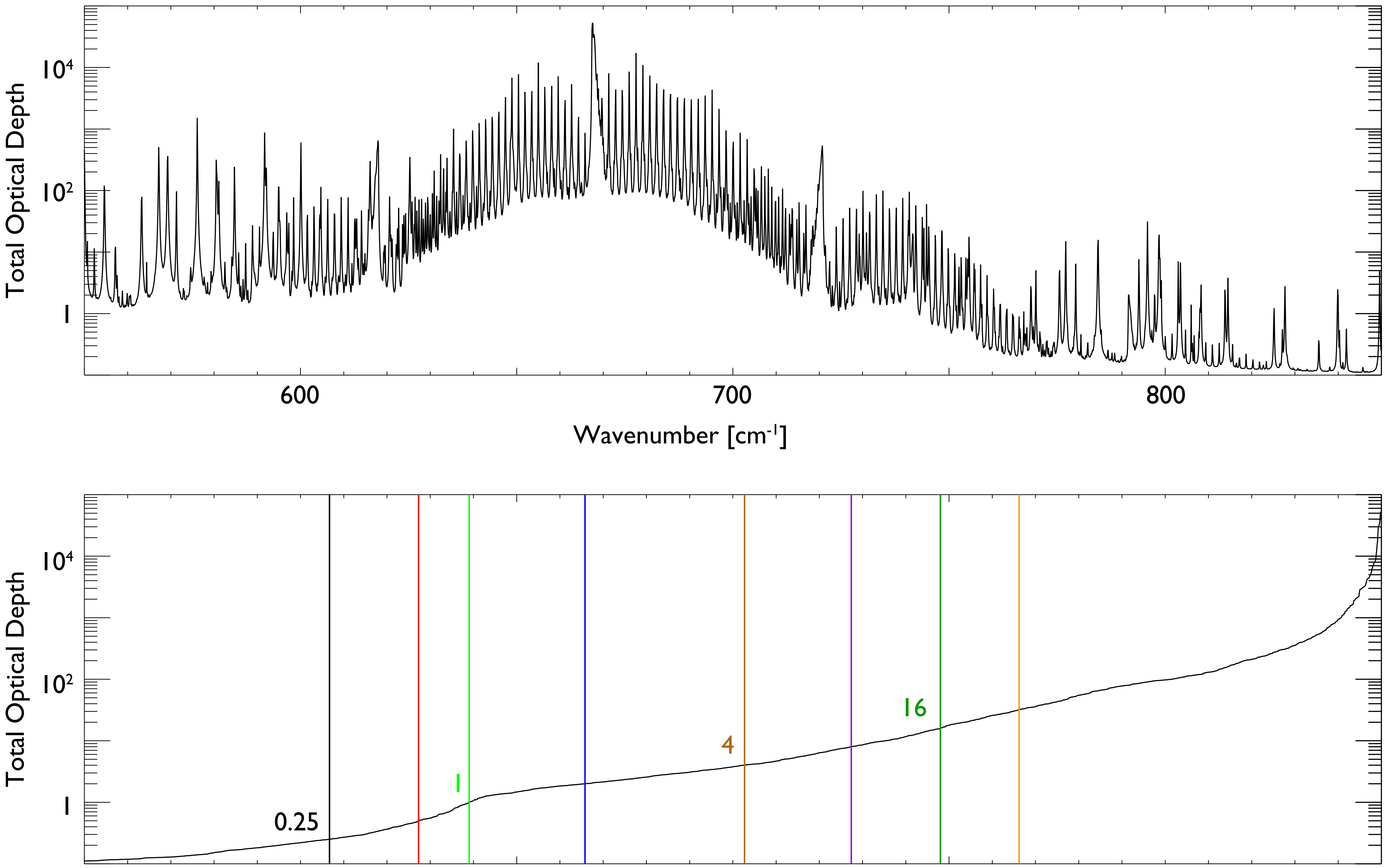
In the infrared and microwave scattering is small

At nadir (looking straight down)

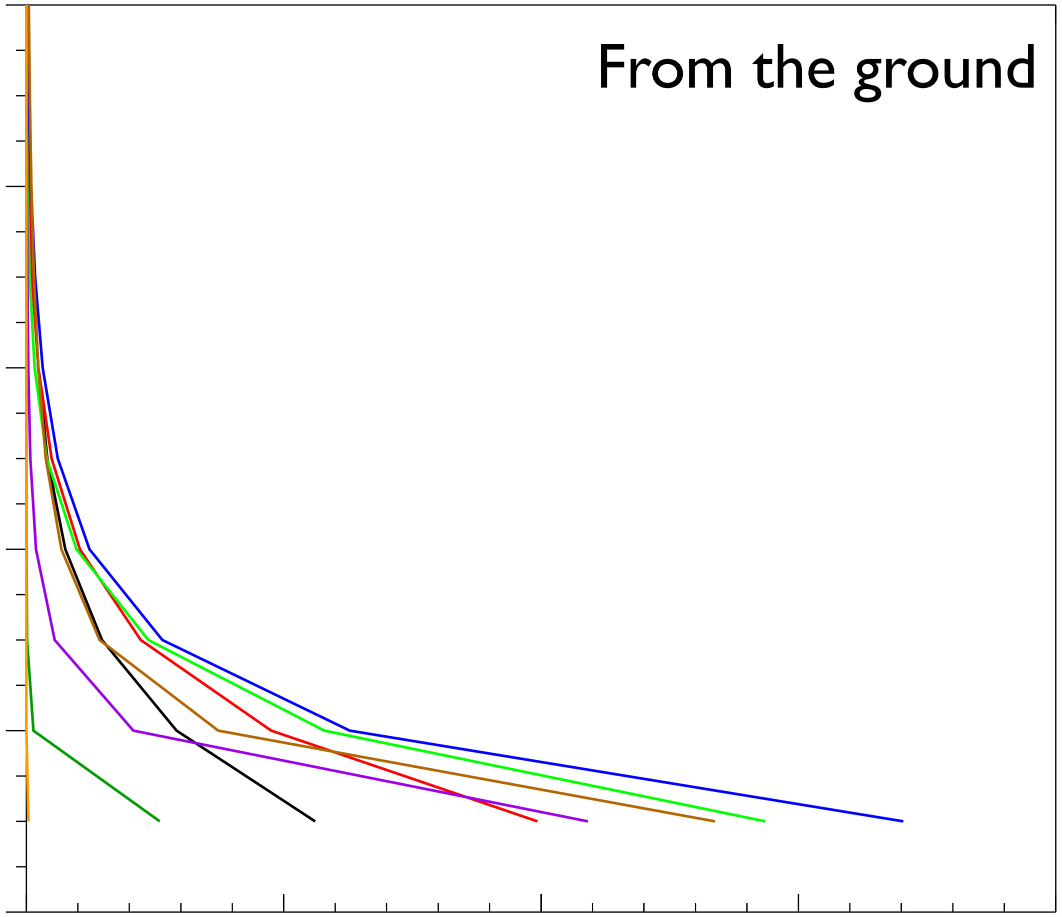
$$\mu \frac{dI_\lambda(\tau,\mu,\phi)}{d\tau} = -I_\lambda(\tau,\mu,\phi) \ + \ (1-\omega_0)B_\lambda(T(\tau))$$

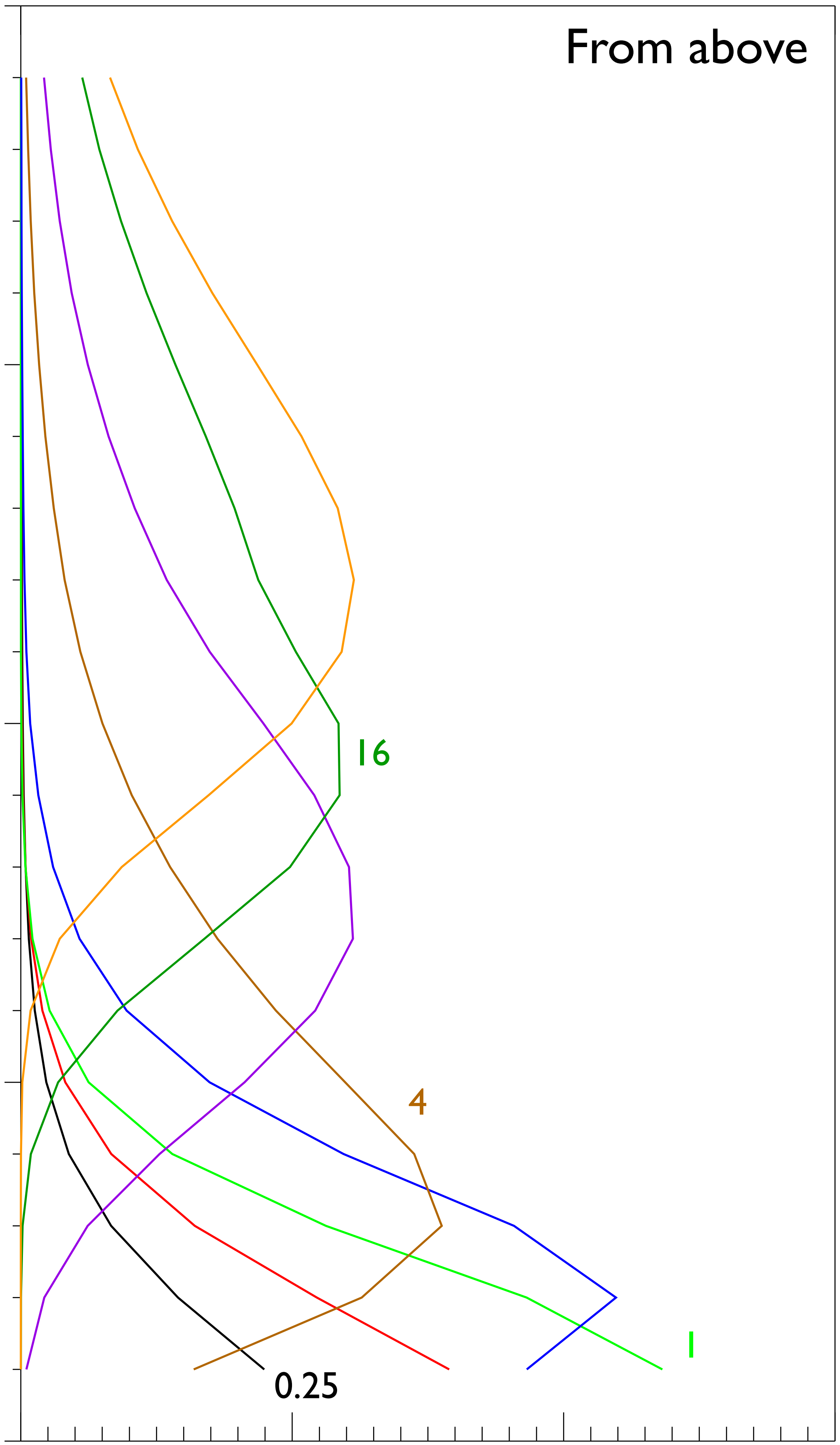
In an absorbing atmosphere of total thickness τ^*

$$I_\lambda(0) = I_\lambda(\tau^*)e^{-\tau^*} + \int_{\tau^*}^{0} e^{-\tau'} B_\lambda(\tau')d\tau'$$



Dave Turner, U. Wisconsin





Dave Turner, U. Wisconsin

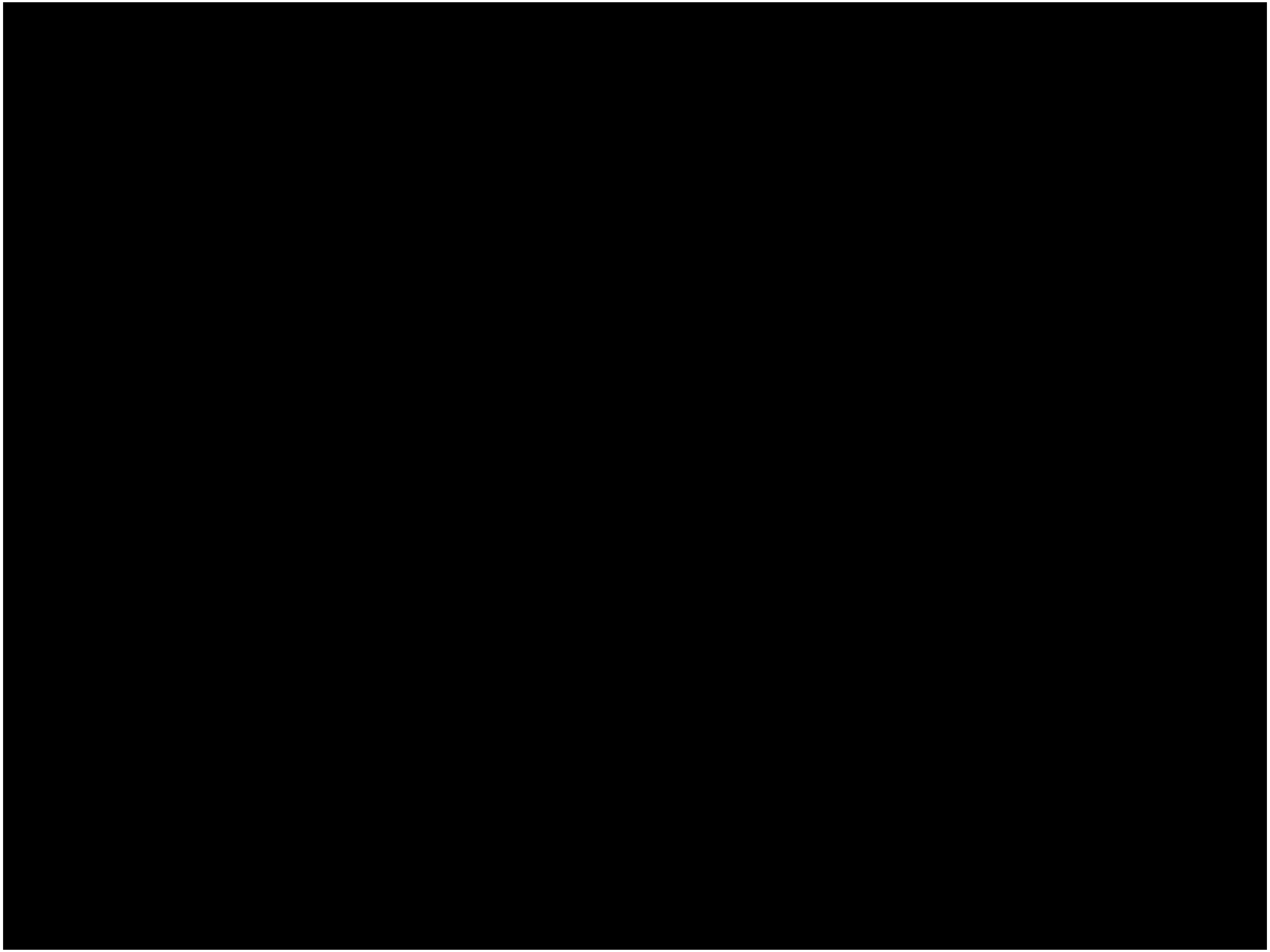
Observing the system remotely (scattering)

Scattering dominates in the visible and near-infrared in clouds (emission is negligible)

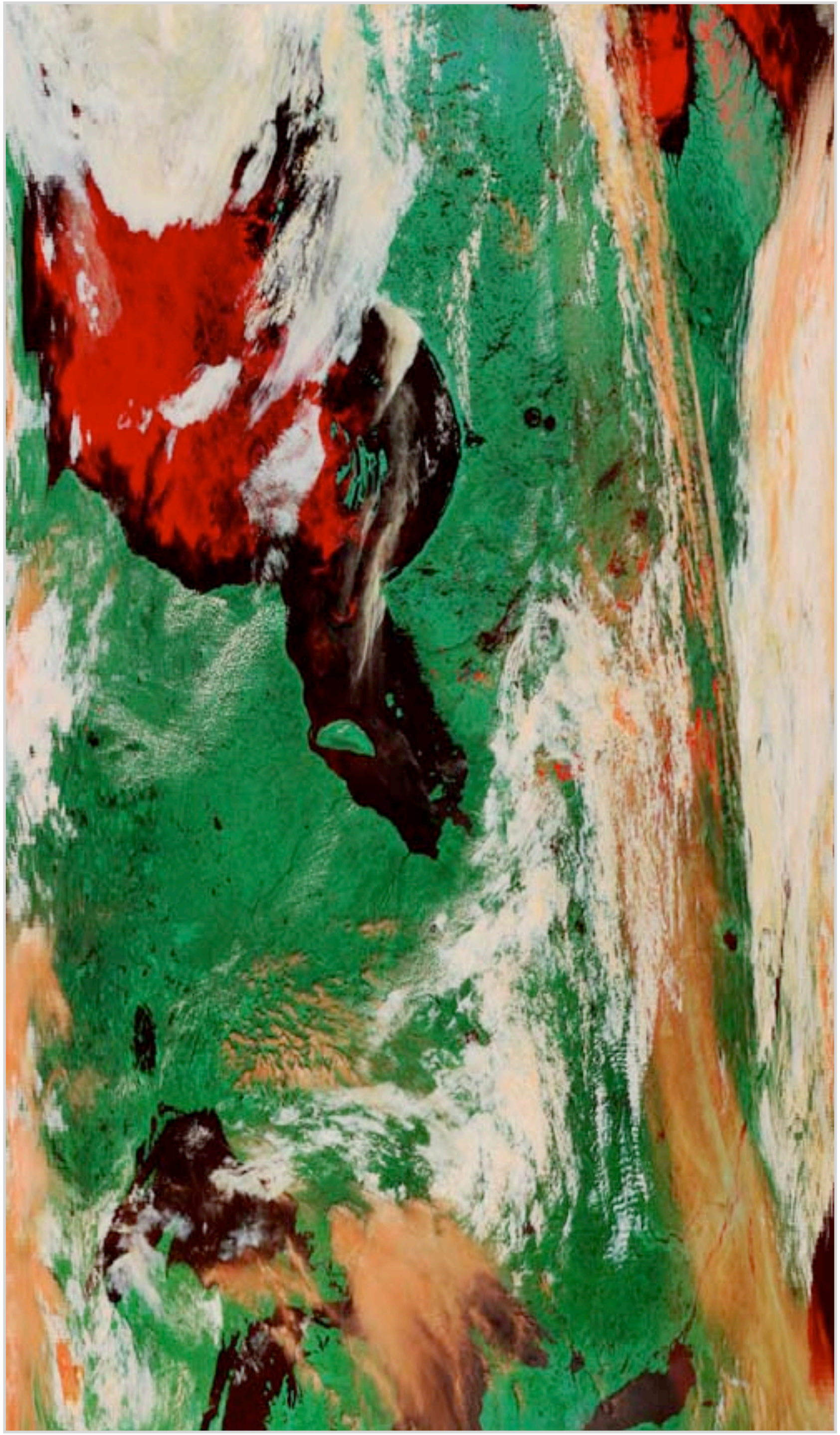
Solutions are costly to compute

Parameters vary significantly with wavelength, particle size, shape, ...

Observing the system remotely



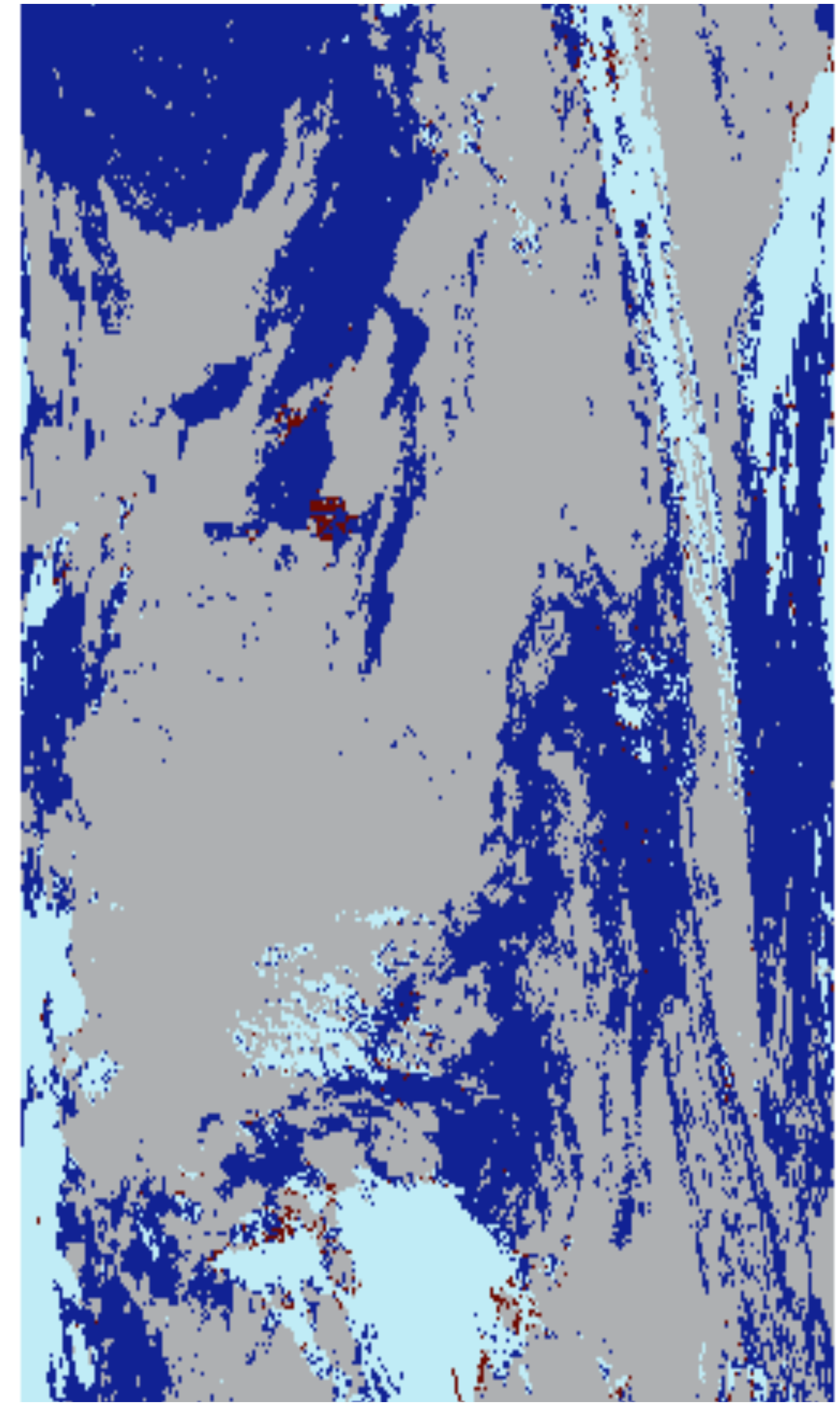
SWIR composite



Cloud Mask overall confidence

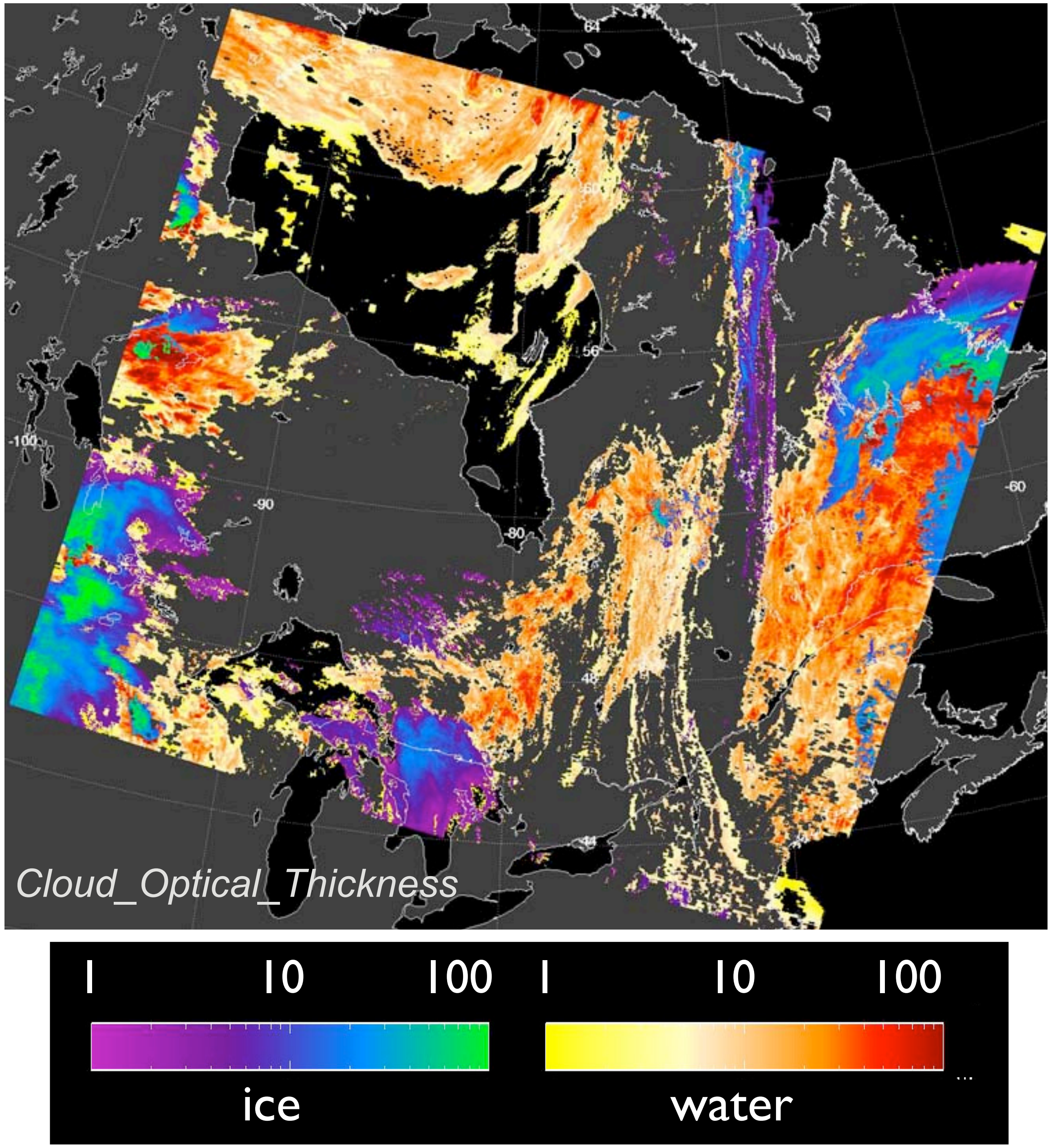
probably clear
cloudy
clear
probably cloudy

Thermodynamic phase
Cloud_Phase_Optical_Properties

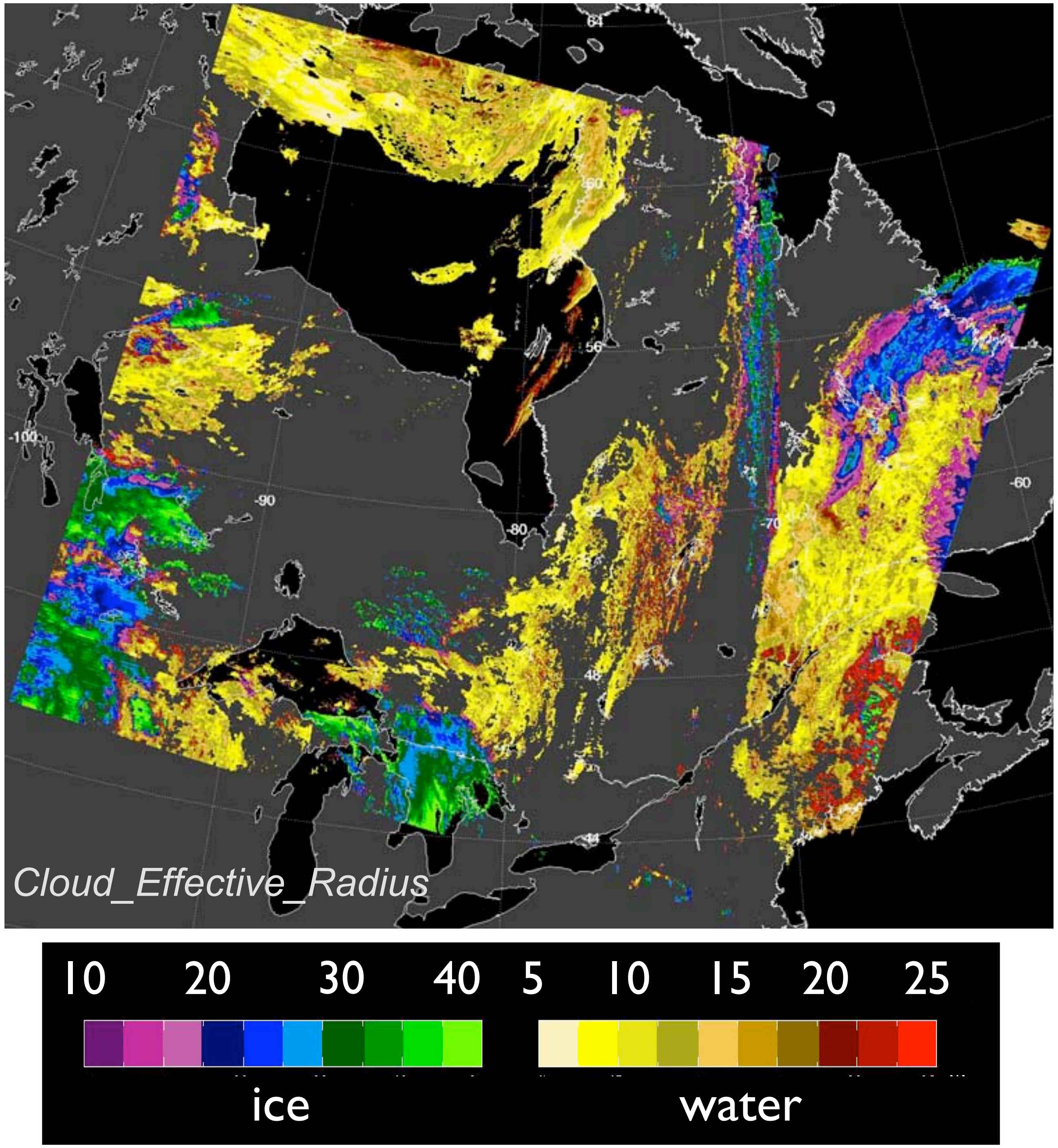


liquid water
ice
undetermined

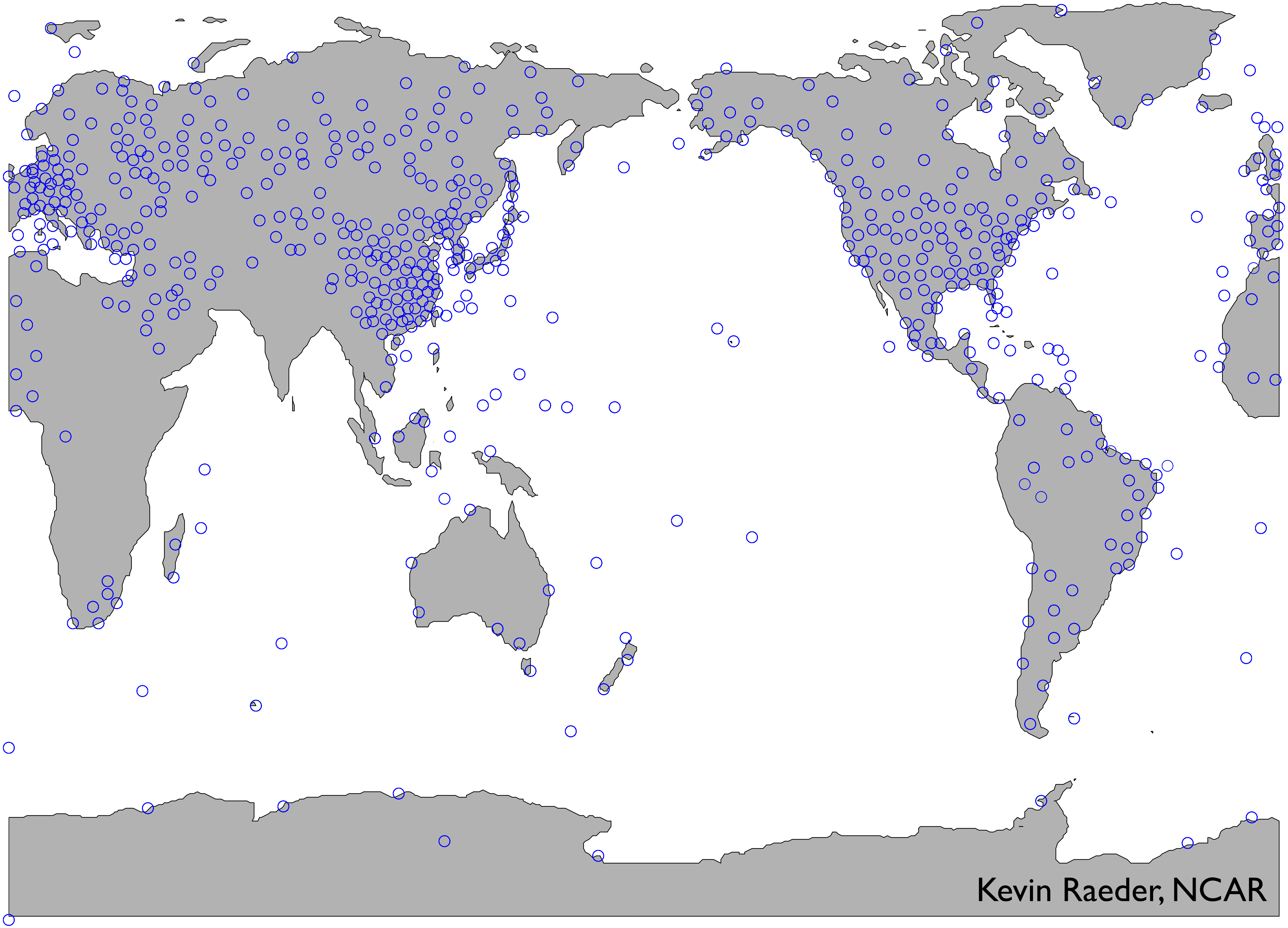
Optical thickness



Particle size (μm)



Radiosonde locations, 1 Dec 2006, 6 hr window around 12Z



Kevin Raeder, NCAR

What's the likelihood of a given state given our forecast and observations?

$$p(\mathbf{x}|\mathbf{y},\mathbf{x_b}) = \frac{p(\mathbf{y},\mathbf{x_b}|\mathbf{x})\,p(\mathbf{x})}{p(\mathbf{y},\mathbf{x_b})}$$

Introduce a cost function

$$J(\mathbf{x}) = -\log(p(\mathbf{x}|\mathbf{y},\mathbf{x_b})) + c$$

Assume no correlation between forecast and observation errors

$$p(\mathbf{y},\mathbf{x_b}|\mathbf{x}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x_b}|\mathbf{x})$$

$$J(\mathbf{x}) = -\log(p(\mathbf{x}|\mathbf{x_b})) - \log(p(\mathbf{x}|\mathbf{y})) + c$$

Assume Gaussian distributions of forecast and observations errors

$$
\begin{aligned}
J(\mathbf{x}) &= \frac{1}{2}(\mathbf{x_b}-\mathbf{x})^T\mathbf{P_b}^{-1}(\mathbf{x_b}-\mathbf{x}) \\
&+ \frac{1}{2}(\mathbf{y}-\mathcal{H}(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-\mathcal{H}(\mathbf{x}))
\end{aligned}
$$

Our cost function minimum is at

$$\nabla J(\mathbf{x}) = \mathbf{P_b}^{-1}(\mathbf{x_b}-\mathbf{x}) + \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y}-\mathcal{H}(\mathbf{x})) = 0$$

Assuming linear observation operator

$$\mathcal{H}(\mathbf{x}) = \mathcal{H}(\mathbf{x_b}) + \mathbf{H}(\mathbf{x}-\mathbf{x_b})$$

So the most likely solution is given by

$$\mathbf{P_b}^{-1}(\mathbf{x}-\mathbf{x_b})+\mathbf{H}^T\mathbf{R}^{-1}(\mathcal{H}(\mathbf{x_b})+\mathbf{H}(\mathbf{x}-\mathbf{x_b})-\mathbf{y})=0$$

Rearranging gives something more familiar

$$\mathbf{x_a}=\mathbf{x_b}+\mathbf{K}(\mathbf{y}-\mathcal{H}(\mathbf{x_b}))$$

with

$$\mathbf{K}=[\mathbf{P_b}^{-1}+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}]^{-1}\mathbf{H}^T\mathbf{R}^{-1}$$

This is the Kalman filter equation

Data assimilation for weather forecasting

There are two main classes of implementation of the Kalman filter in the atmospheric sciences

Variational methods minimize the cost function iteratively

"3D-Var" minimizes cost function at a single time; "4D-Var" minimizes cost function over some extended time window

Variational methods need tangent-linear and adjoint models

Ensemble methods use Monte Carlo integration of Bayes's formula; take sample covariance matrices from a finite ensemble

Ensemble data assimilation recipe (for Brian Mapes)

For each observation in turn

- compute the prior/background observation PDF
  - fit the ensemble of expected observations with a Gaussian
  - find the product of the background and observation PDFs
  - find the increment needed for each ensemble member
- update "each" element in the state vector in each member
  - regress the expected observation against the element using the ensemble
  - apply (observation increment) * (correlation coefficient) * (localization)

Data assimilation is model evaluation

- Analysis quality is measured by quality of fit to observations
    - "Observation biases" make this complicated
- Forecast skill is measured against verifying analysis
    - Note: Forecast skill includes errors in forecast model and initial conditions (including errors in assimilation)

Standard deterministic verification metrics are boring

Geographic regions:

Tropics (20S - 20N), extratropics (20-90 S, N)

Grid: 2.5° lat-lon

Quantities:

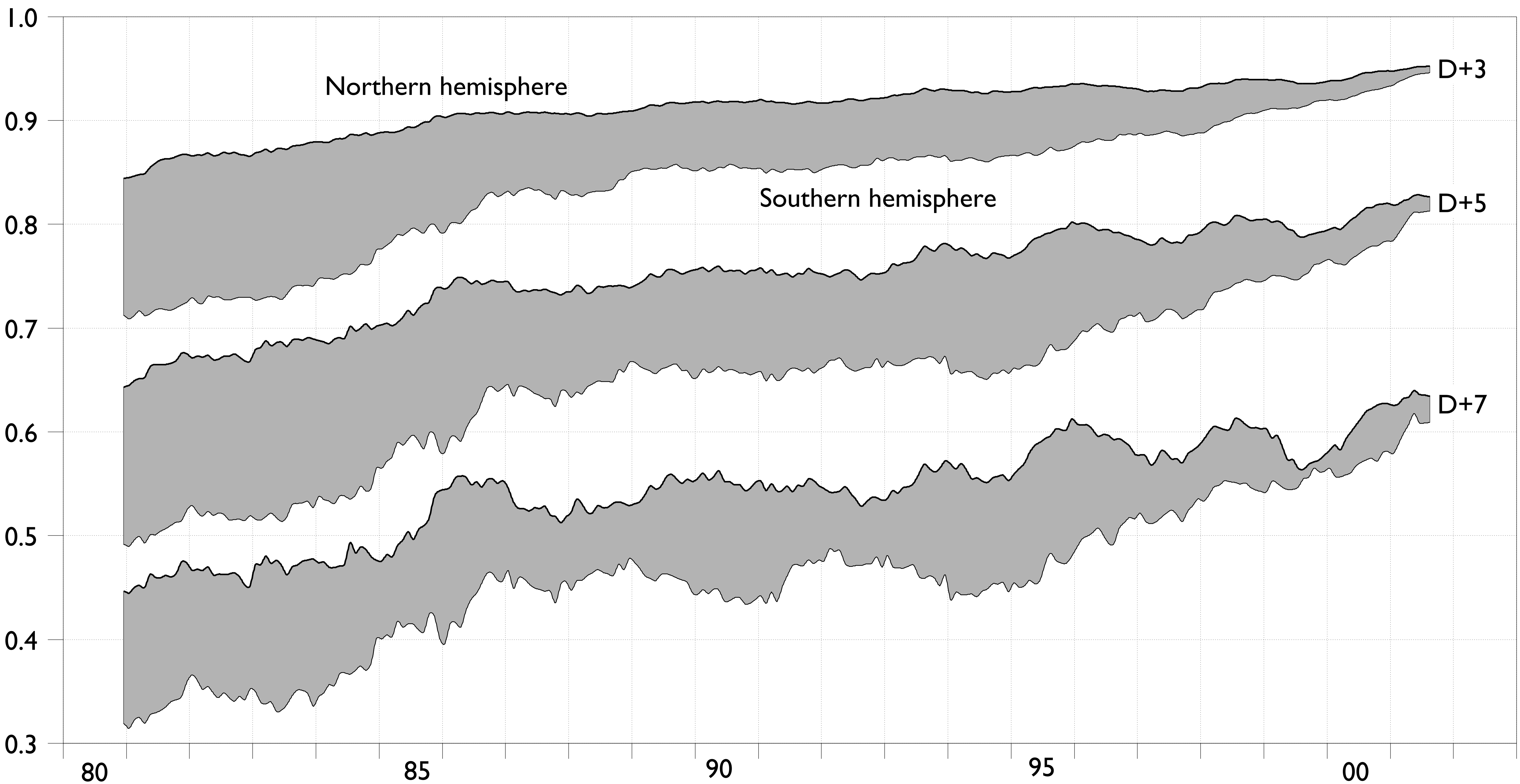
Sea-level pressure, geopotential height, winds, temperature

Heights: 850 hPa, 250 hPa (+ sea level in extratropics)

Statistics:

bias, root-mean-square error, anomaly correlation, S1 skill (measures magnitude of gradients)

Anomaly correlation of ECMWF 500hPa height forecasts



Simmons and Hollingsworth, 2001

Deterministic forecast evaluation

A lot of effort goes into analyses - best estimates of the state of the atmosphere

Verification against those analyses is straightforward

- Tests large-scale circulation, which is not so sensitive to model details
- But evaluation is routine - 2-4 times each day, year after year

Probabilistic evaluation is more fun

Ensembles of forecasts

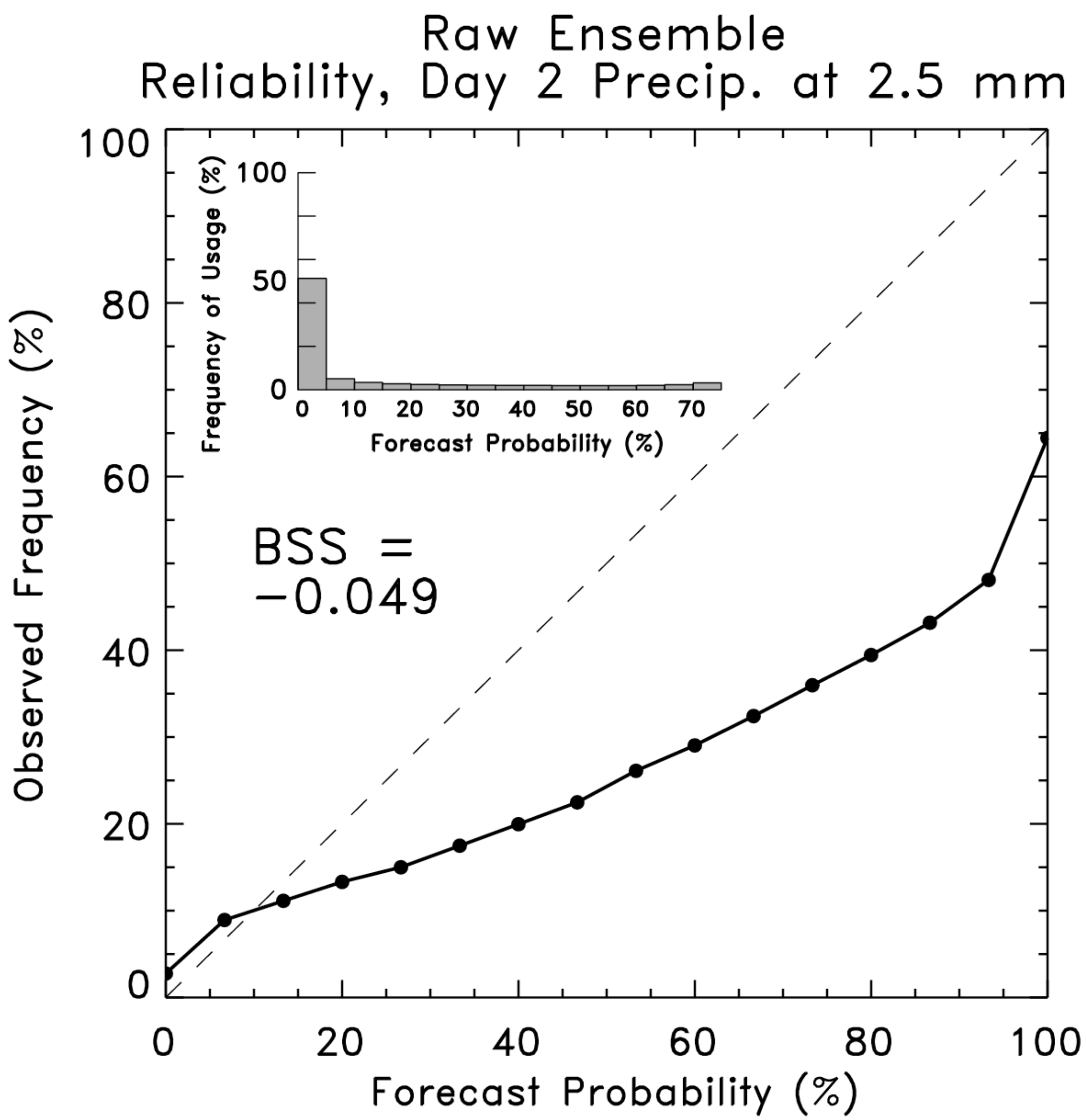
produce sample distributions

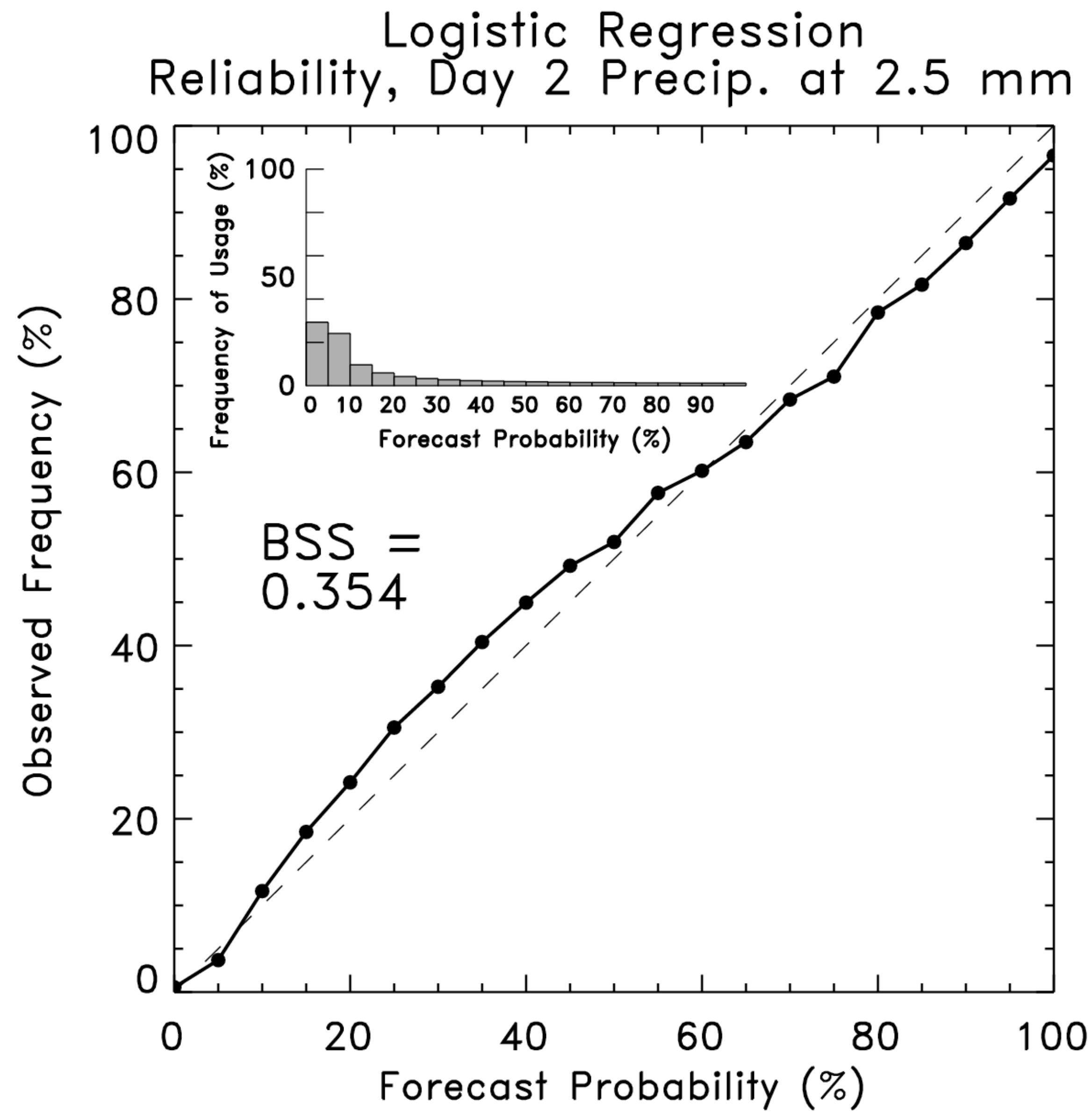
are expressed as the probability of some discrete event (e.g. "chance of rain > 1 mm")

Verification is binary

Brier score is discrete analog to mean-squared error

Includes
sharpness (differentiation of forecasts) and
reliability (forecast frequencies matching observations)





Tom Hamill, notes for
Short Course on Ensemble Forecasting

From weather forecasting to climate projections

Weather forecasting and climate modeling communities (and codes) are mostly distinct

Few climate models are able to make weather forecasts

Retrospective simulations are loosely coupled to historical state of the atmosphere

The most sought-after forecasts by climate models (e.g. "climate sensitivity") can't be verified

Observing the climate system: data reduction

How are all those observations summarized?

Univariate averaging on regular grids, for the most part

Maybe histograms, joint histograms, linear regressions

Life without time correspondence (evaluating "climate")

Choose a variable ϕ

Build a composite season cycle

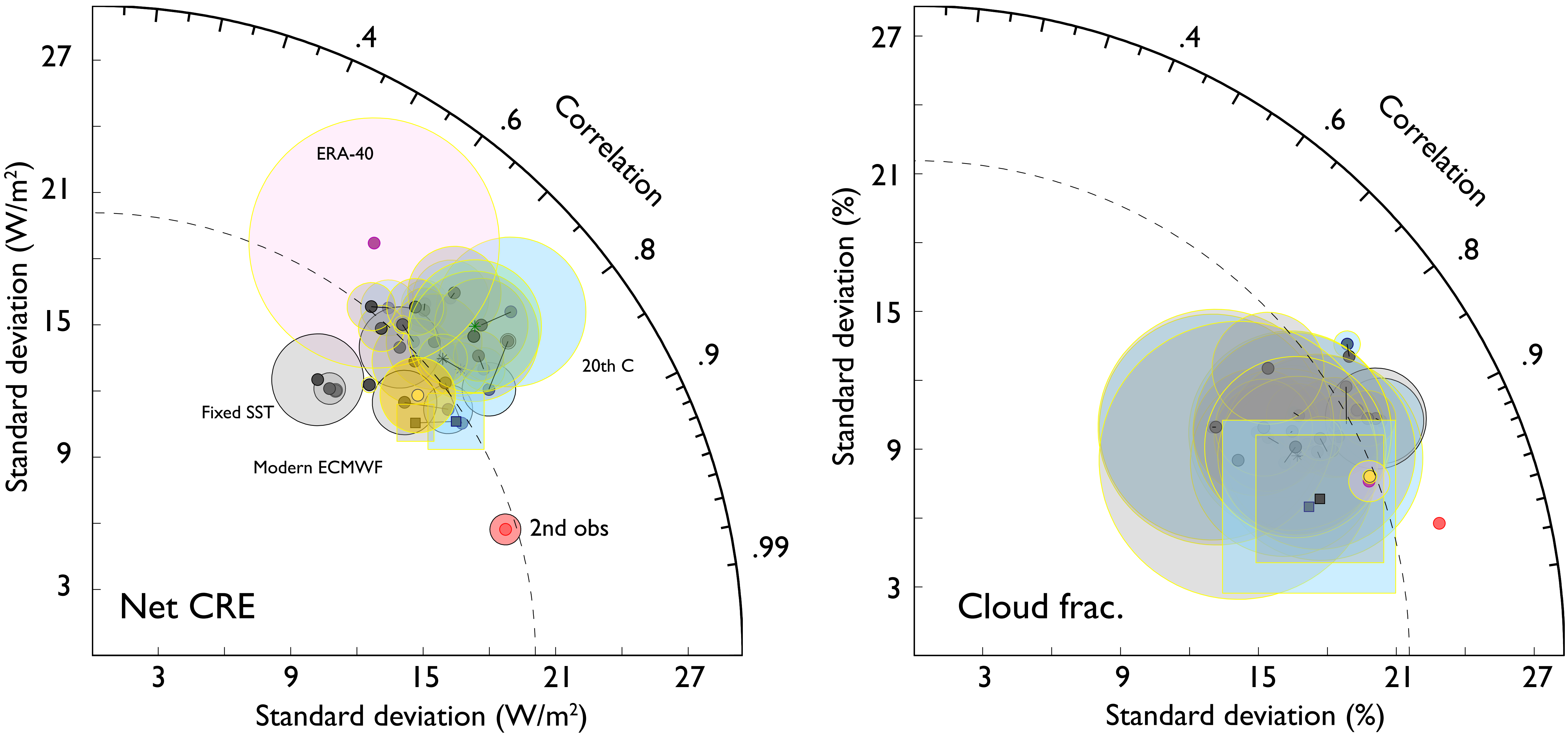
$$\hat{\phi}(x, y; m) = \frac{1}{Y} \sum_{y=1}^{Y} \phi(x, y; m, y)$$

Compute the mean squared error

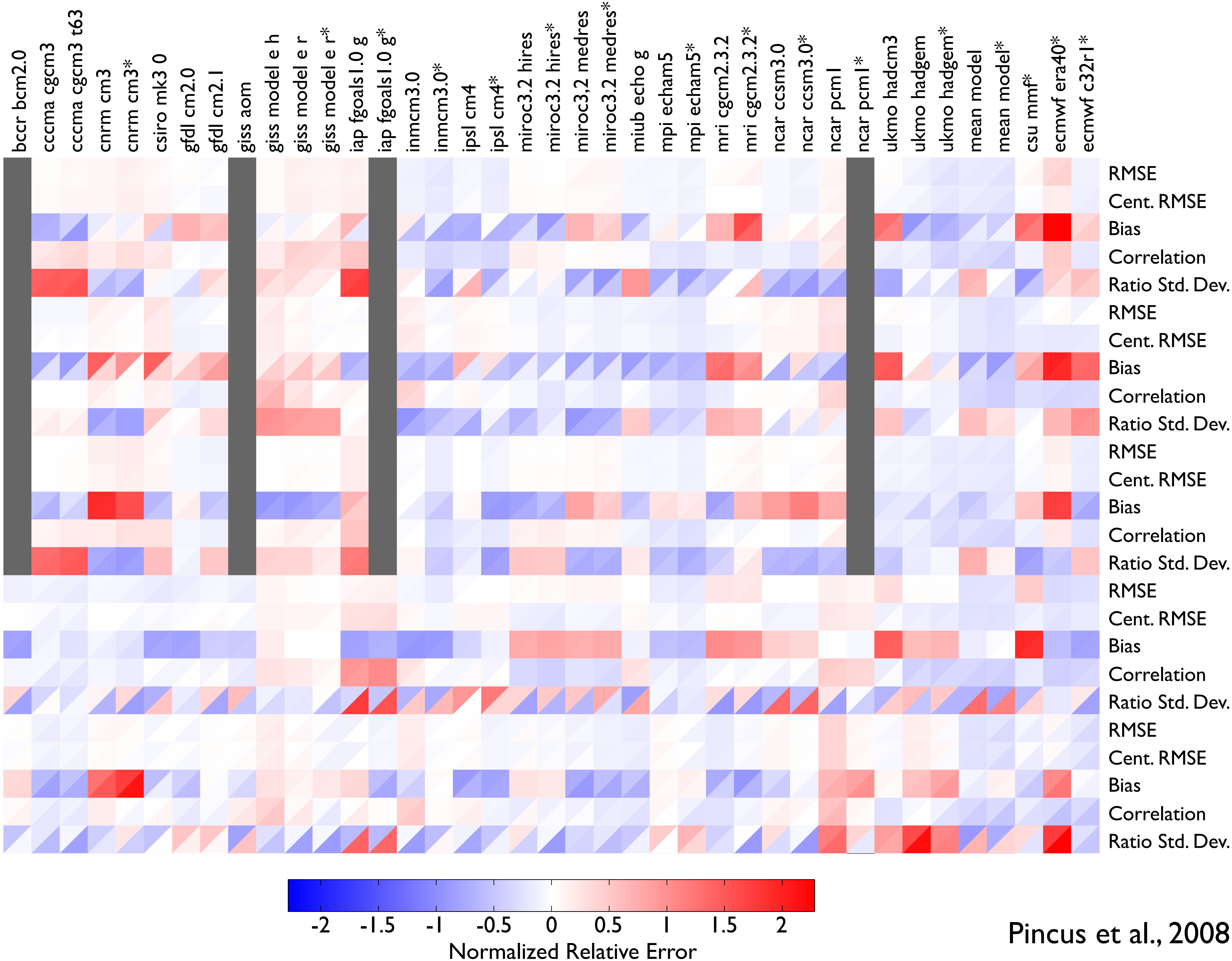
$$e^2 = \frac{1}{T} \int_T \frac{1}{A} \int_A (\hat{\phi}_m - \hat{\phi}_o)^2 \; da \; dt$$

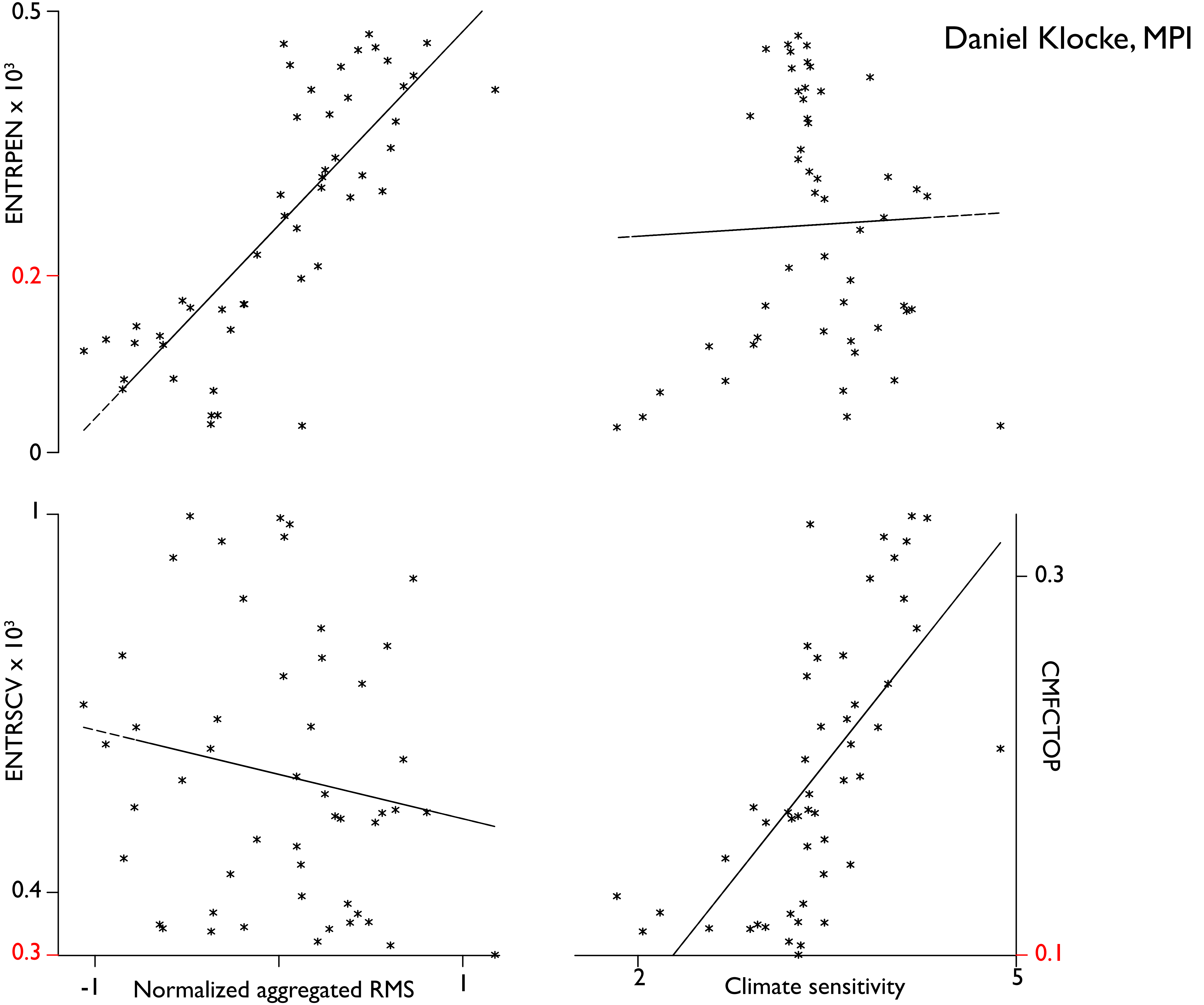
Invoke

$$\begin{aligned} e^2 &= \overline{e}^2 + e'^2 \\ &= \overline{e}^2 + \sigma_o^2 (1 + s^2 - 2sr) \end{aligned}$$



K.E. Taylor, 2001
(c.f. Pincus et al., 2008)


bccr bcm2.0
cccma cgcm3
cccma cgcm3 t63
cnrm cm3
cnrm cm3*
csiro mk3 0
gfdl cm2.0
gfdl cm2.1
giss aom
giss model e h
giss model e r
giss model e r*
iap fgoals1.0 g
iap fgoals1.0 g*
inmcm3.0
inmcm3.0*
ipsl cm4
ipsl cm4*
miroc3.2 hires
miroc3.2 hires*
miroc3,2 medres
miroc3.2 medres*
miub echo g
mpi echam5
mpi echam5*
mri cgcm2.3.2
mri cgcm2.3.2*
ncar ccsm3.0
ncar ccsm3.0*
ncar pcm1
ncar pcm1*
ukmo hadcm3
ukmo hadgem
ukmo hadgem*
mean model
mean model*
csu mmf*
ecmwf era40*
ecmwf c32r1*
RMSE
Cent. RMSE
Bias
Correlation
Ratio Std. Dev.
Net CRE
LW CRE
SW CRE
CF
PR
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
Normalized Relative Error
Pincus et al., 2008



Daniel Klocke, MPI
ENTRPEN x 10³
0.5
0.2
0
ENTRSCV x 10³
1
0.4
0.3
-1
1
Normalized aggregated RMS
2
5
Climate sensitivity
CMFCTOP
0.3
0.1

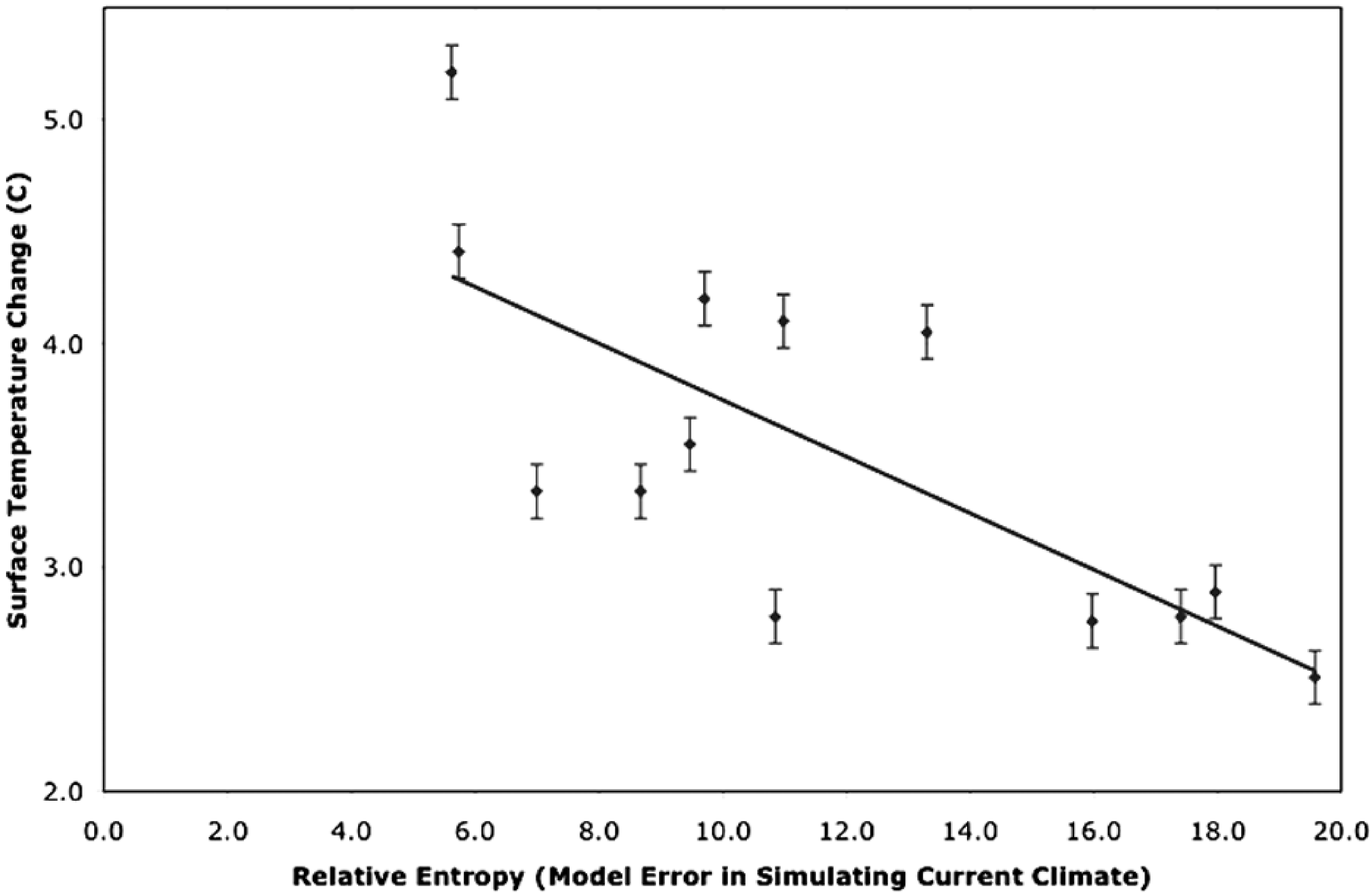
There are three more broad classes of evaluation based on

relationships among variables
(e.g. Bony et al., 2003; Bennhold and Sherwood, 2008;
Clement et al., 2009)

characteristics of emergent behavior
(e.g. Williams and Tselioudis, 2008; see also
phenomena-based metrics)

information theoretic
(e.g. Shukla et al., 2005, Majda and Gershgorin 2010)

There may be an infinity of ways in which the models are wrong



Shukla el al., 2005

The climate model evaluation conundrum

Models are evaluated during development

- Each group, each individual has their own set of metrics
- Centers trade-off skill in mean vs. skill in variability
- Community metrics are being developed

But there is no basis on which to select or prefer a given metric

- Weak-to-no observational basis linking historical skill with climate change response
- Relationships in existing ensembles may be fortuitous and/or have simple causes