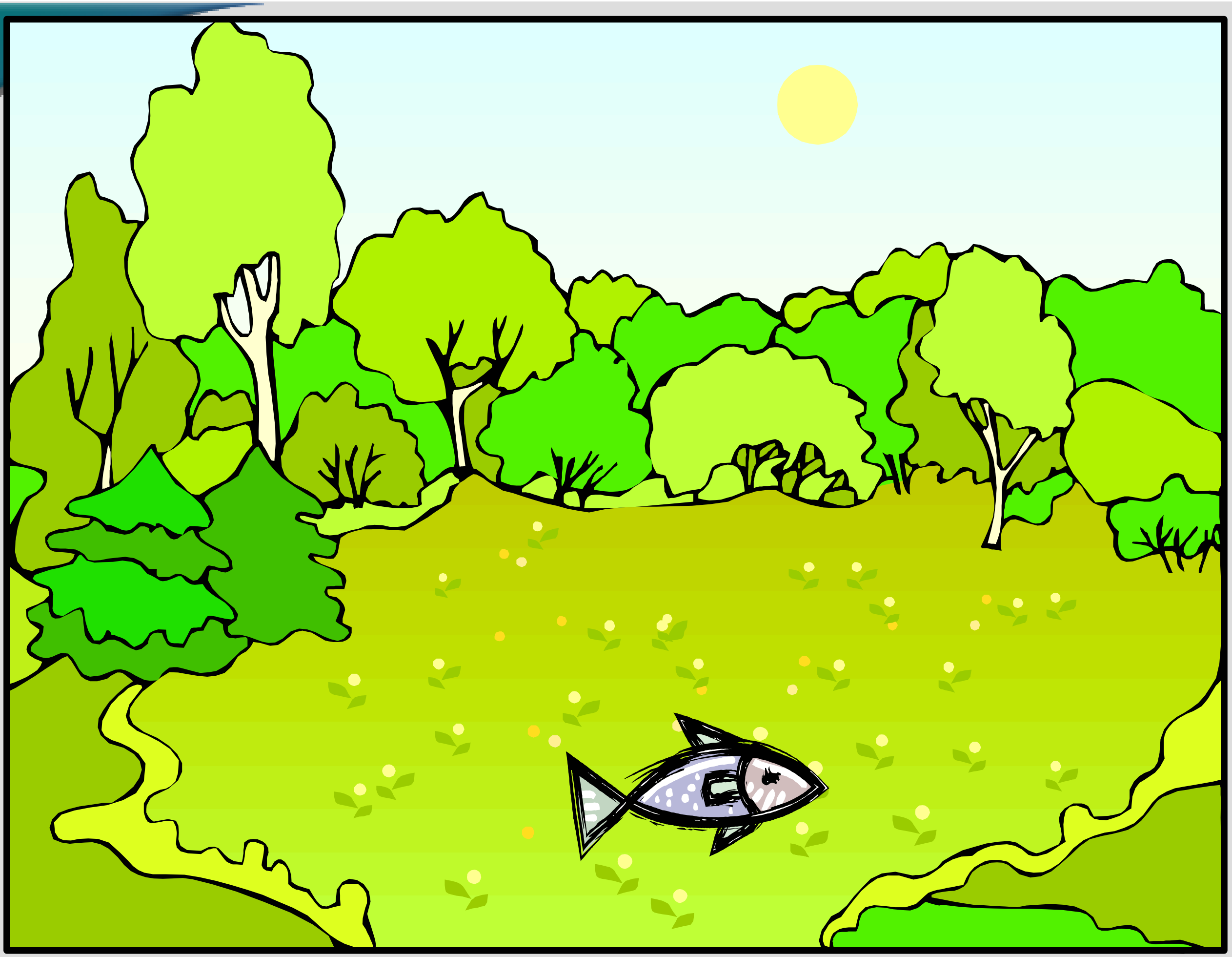
Using Validation Sites and Field Campaigns to Evaluate Observational and Model Bias *(and perhaps correct the bias)*

Tom Ackerman

Battelle Fellow, PNNL

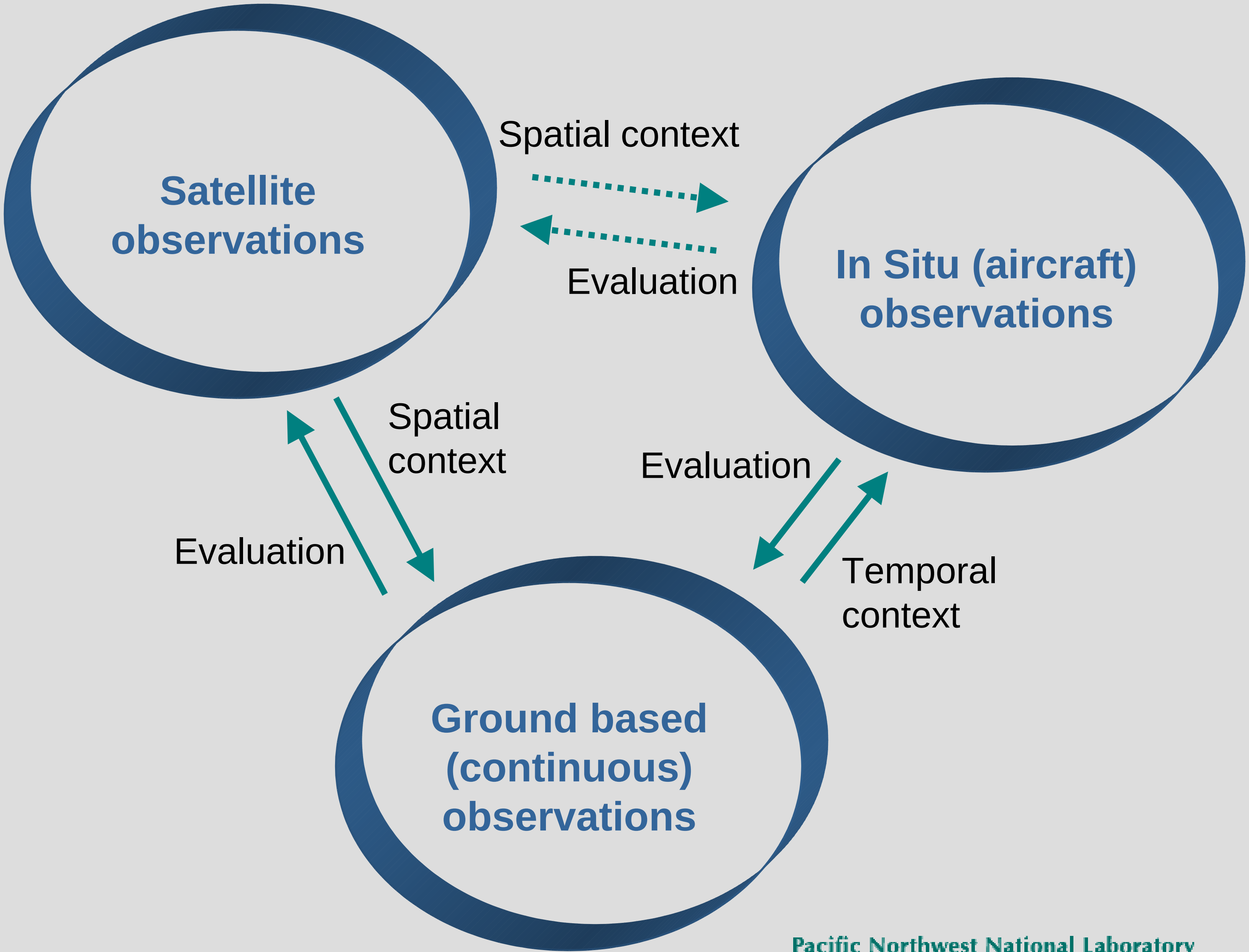
Professor, University of Washington



My primary interest is cloud properties and effects, where the bias errors in our measurements are pretty insignificant compared to model-observation differences

A brief history ….

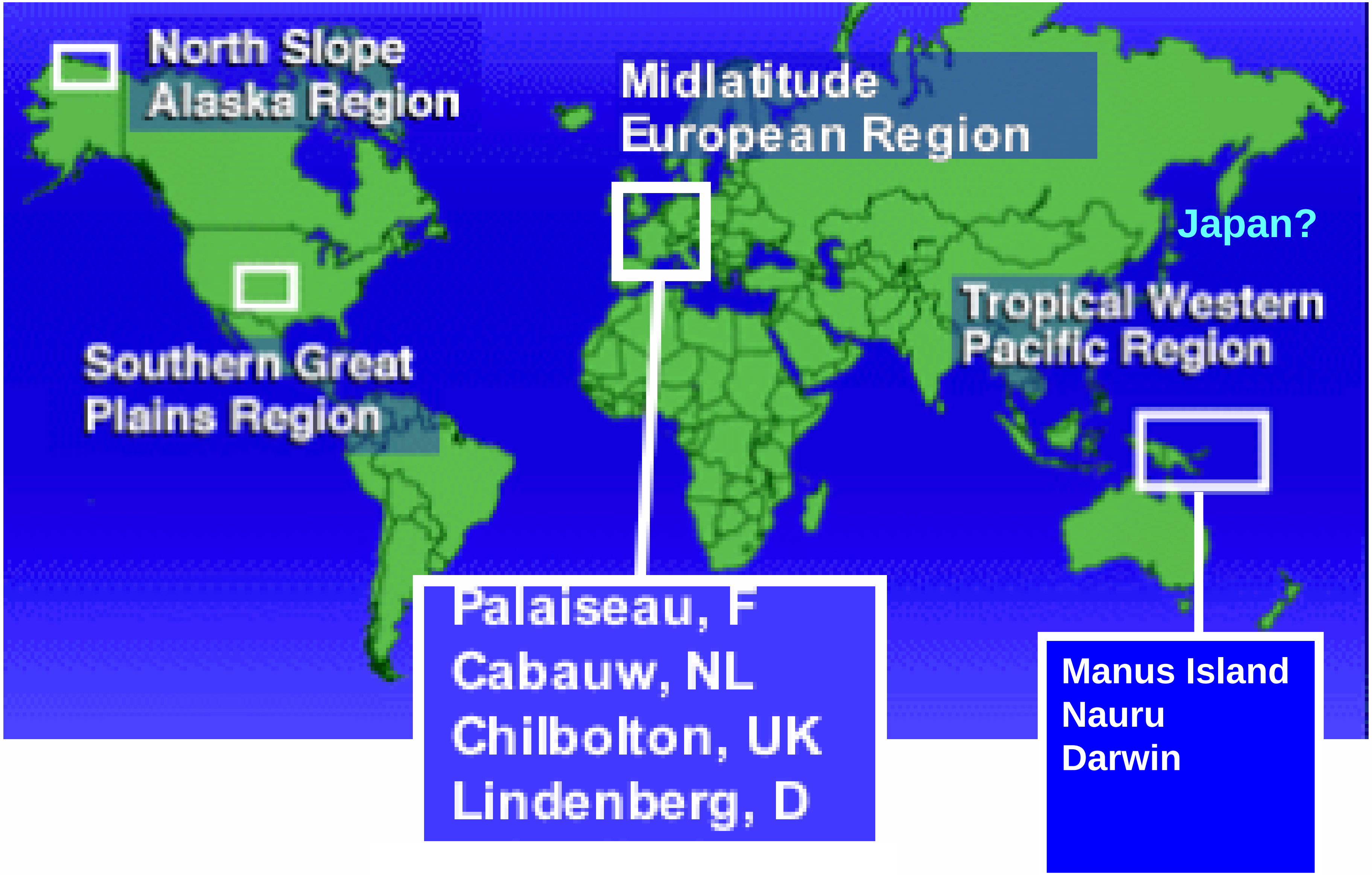
- In the beginning, there were surface measurements
  - Continuous measurements of state variables
  - Networks => climatology, initialization for forecast models
- Then there were field programs
  - Short duration, multiple platforms (aircraft and ground)
  - Process studies => elucidate the physics of the atmosphere
- Then there were satellites
  - Global, single platform, one (or a few) instrument
  - Climatology, spatial snapshots
- But, the satellites needed validation
  - More field programs
  - Multiple locations, repeated efforts
- And then there were profiling sites
  - Continuous measurements of many variables and fluxes
  - Process studies, satellite evaluation, climatology

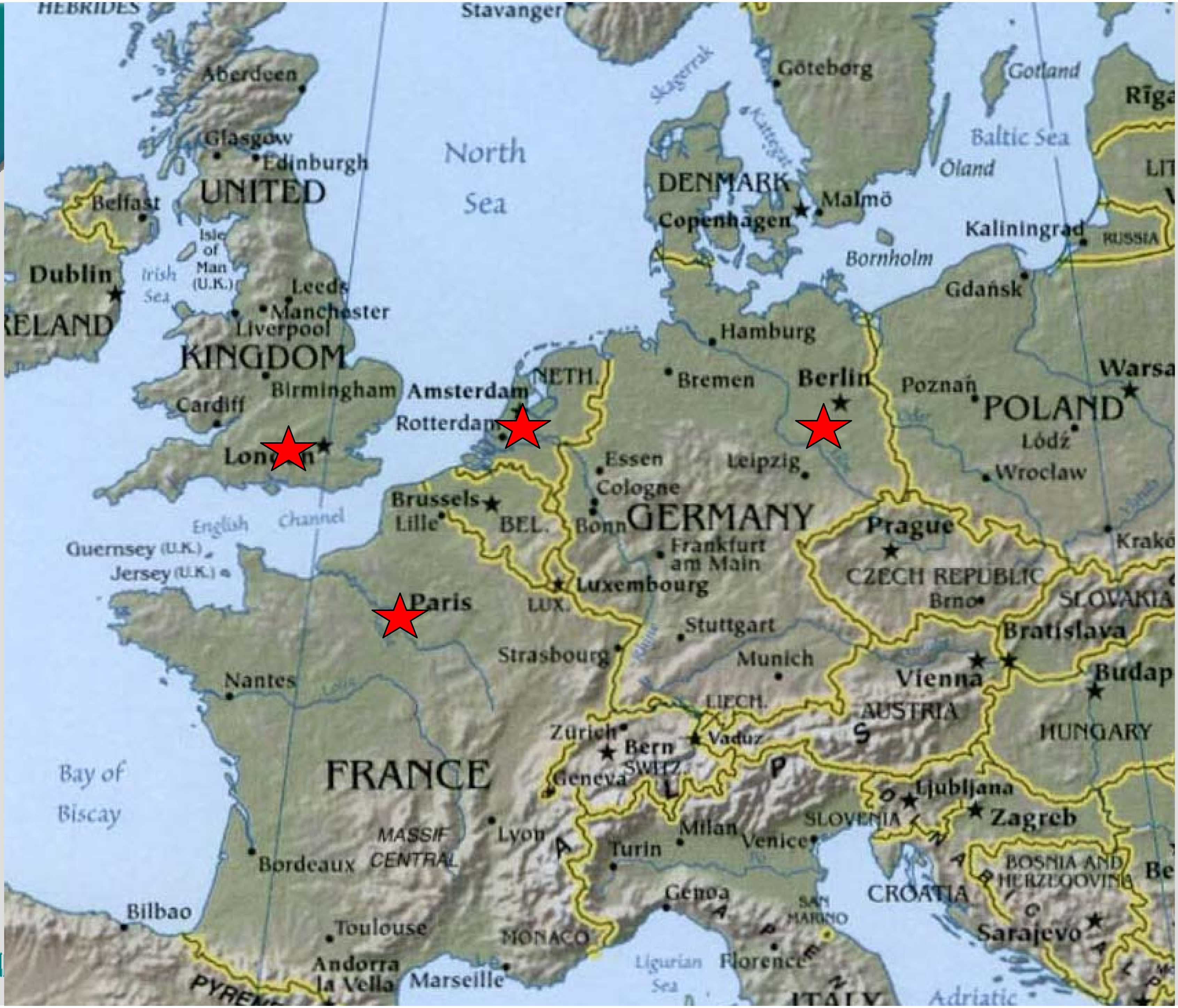

Satellite observations
Spatial context
Evaluation
In Situ (aircraft) observations
Spatial context
Evaluation
Evaluation
Temporal context
Ground based (continuous) observations

Types of Ground-based sites

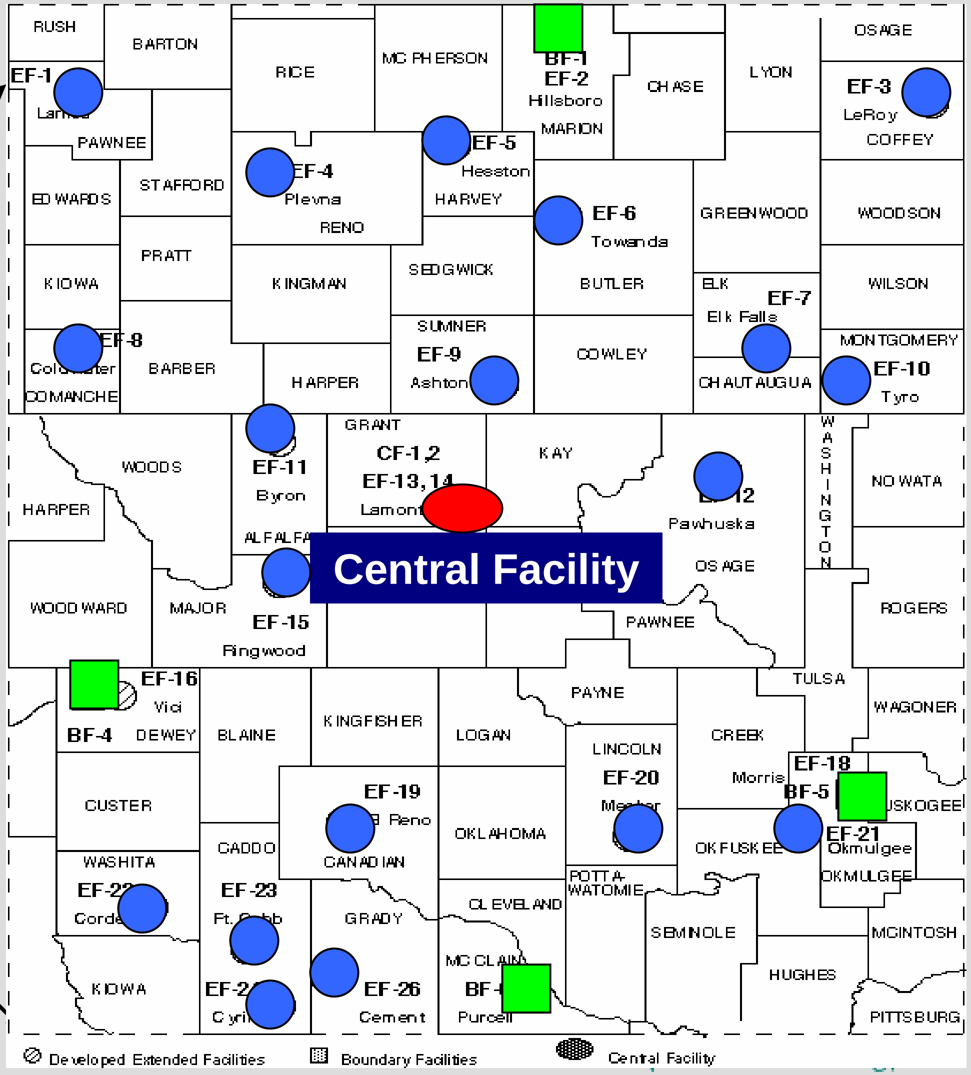
- Standard meteorology – radiosondes
- Special networks: Baseline Surface Radiation Network (BSRN) or Aerosol Robotic Network (AERONET)
- Cloud and Aerosol Profiling (CAP) Sites

GEWEX Atmospheric Profiling Network




HEBRIDES
Stavanger
Aberdeen
Göteborg
Gotland
Riga
Skagerrak
Kattegat
Baltic Sea
Glasgow
Edinburgh
North
Sea
DENMARK
Öland
Belfast
UNITED
Malmö
Copenhagen
Kaliningrad
RUSSIA
Isle of Man (U.K.)
Bornholm
Dublin
Irish Sea
Leeds
Gdańsk
Manchester
IRELAND
Liverpool
Hamburg
KINGDOM
NETH.
Bremen
Berlin
Poznań
Warsa
Birmingham
Amsterdam
POLAND
Cardiff
Rotterdam
Łódź
London
Essen
Leipzig
Wrocław
Cologne
Brussels
English
Channel
Lille
BEL.
Bonn
GERMANY
Prague
Guernsey (U.K.)
Frankfurt am Main
Jersey (U.K.)
Luxembourg
CZECH REPUBLIC
Brno
SLOVAKIA
Paris
LUX.
Stuttgart
Bratislava
Strasbourg
Munich
Vienna
Budap
Nantes
LIECH.
AUSTRIA
Zürich
Bern
Vaduz
HUNGARY
Bay of
Biscay
FRANCE
Geneva
SWITZ.
Ljubljana
SLOVENIA
Zagreb
MASSIF CENTRAL
Lyon
Milan
Venice
Bordeaux
Turin
BOSNIA AND HERZEGOVINA
CROATIA
Genoa
SAN MARINO
Bilbao
Toulouse
MONACO
Sarajevo
Andorra la Vella
Ligurian Sea
Florence
Marseille
ITALY
Adriatic

ARM Southern Great Plains



Battelle

Southern Great Plains Central Facility

Bias errors in measurements

- Most difficult errors to diagnose
  - By definition, if we know about a bias error, we remove it
- How do we find them?
  - Instrument to instrument comparison (water vapor)

    But, often have only one instrument or we cannot sort out source of inconsistency
  - Instrument to model comparison (diffuse flux, MPACE)

    But, which do we trust? (instrument, of course!)
  - Consistency among multiple measurements (single-scatter albedo, aerosol closure experiments)

    But, can we reduce solution to bias in only one instrument?

Some stories from ARM

Story #1: Water vapor

- ARM has invested more effort and money in the study of water vapor measurements than any other quantity
- Multiple instrumentation
- FIVE intensive campaigns
- Many science team research projects
- Countless hours of debate

Revercomb et al., 2003, BAMS (and a host of references)

Soden et al., 2005, JGR and references therein

Table 2. Operational water vapor instrumentation at the ARM SGP Central Facility. (Additional information about the instruments listed may be found online at www.arm.gov/docs/instruments.html.)

Instrument	Primary quantity observed and typical resolutions	References
AERI retrievals	Water vapor mixing ratio profiles: 10 min, 100-m resolution, 24 h day^{-1}	Feltz et al. (1998); Turner et al. (2000)
Cimel sun photometer (CE-318)	Total precipitable water vapor: every quarter air mass for air masses greater than 2, and every 15 min for airmasses less than 2	Holben et al. (1998); Schmid et al. (2001)
GPS at Lamont, OK	Total precipitable water vapor: 30-min resolution, 24 h day^{-1}	Wolfe and Gutman (2000); King and Bock (1996); Rotacher (1992)
In situ probes (Vaisala HMP35D*)	Water vapor mixing ratio: at surface, 25 m, and 60 m; 1-min resolution	Richardson and Tobin (1998); Richardson et al. (2000)
MFRSR	Total precipitable water vapor: 1-min resolution during daytime	Harrison et al. (1994); Schmid et al. (2001)
MWR (Radiometrics WVR-1100)	Total precipitable water vapor: 20-s resolution, 24 h day^{-1}	Liljegren and Lesht (1996); Liljegren (1999)
Radiosonde (Vaisala RS-80H)	Relative humidity profiles: 10-m resolution, eight launches per day	Turner et al. (2003); Lesht (1998)
Raman lidar (CARL)	Water vapor mixing ratio profiles: 10 min, 78-m resolution, 24 h day^{-1}	Goldsmith et al. (1998); Turner and Goldsmith (1999)
RSS	Total precipitable water vapor: 1-min resolution during daytime (installed after the 1996 WVIOP)	Harrison et al. (1999); Schmid et al. (2001)

*Installed after the 1996 WVIOP, replacing the original Qualimetrics 5120-E and 5134-E sensors.

Battelle

TABLE 3. Additional instrumentation brought to the ARM SGP central facility for the 1996 and1997 WVIOPs.

Instrument	Primary quantity observed and typical resolutions	References
AATS-6	Total precipitable water vapor: 12-s resolution during daytime (during 1997 WVIOP only)	Matsumoto et al. (1987); Schmid et al. (2001)
Chilled mirrors (Meteor AG) on kite and tethersonde	Relative humidity profiles: 2-s data during most evenings	Porch et al. (1998); Turner and Goldsmith (1999)
Chilled mirrors on tower (General Eastern D2/M4)	Dewpoint temperature: 1-min resolution, 24 h day^{-1}[a]	Richardson and Tobin (1998); Richardson et al. (2000)
GPS receiver at SGP Central Facility	Total precipitable water vapor: 30-min data, 24 h day^{-1}	Wolfe and Gutman (2000)
MPI-DIAL	Water vapor density profiles: 30 s, 75-m resolution during multiple 12-h periods (operations restricted by FAA)[b]	Wulfmeyer and Bösenberg (1998); Linné et al. (2001)
NOAA ETL 20.6/31.65-GHz microwave radiometer (ETL 1)	Atmospheric brightness temperatures and total precipitable water vapor:[c] 30-s resolution, 24 h day^{-1}	Hogg et al. (1983); Han and Westwater (2000)
NOAA ETL 23.87/31.65-GHz microwave radiometer (ETL 2)	Atmospheric brightness temperatures and total precipitable water vapor: 30-s resolution, 24 h day^{-1} (during 1997 WVIOP only)	Hogg et al. (1983); Han and Westwater (2000)
Scanning AERI in trailer	Downwelling infrared radiance: 8 min, 1-wavenumber resolution, 24 h day^{-1}	Feltz et al. (1998)
SRL	Water vapor mixing ratio profiles: 1 min, 75-m resolution primarily at night	Whiteman and Melfi (1999); Whiteman et al. (2001)

Ba

Microwave Radiometer vs. Vaisala Radiosonde

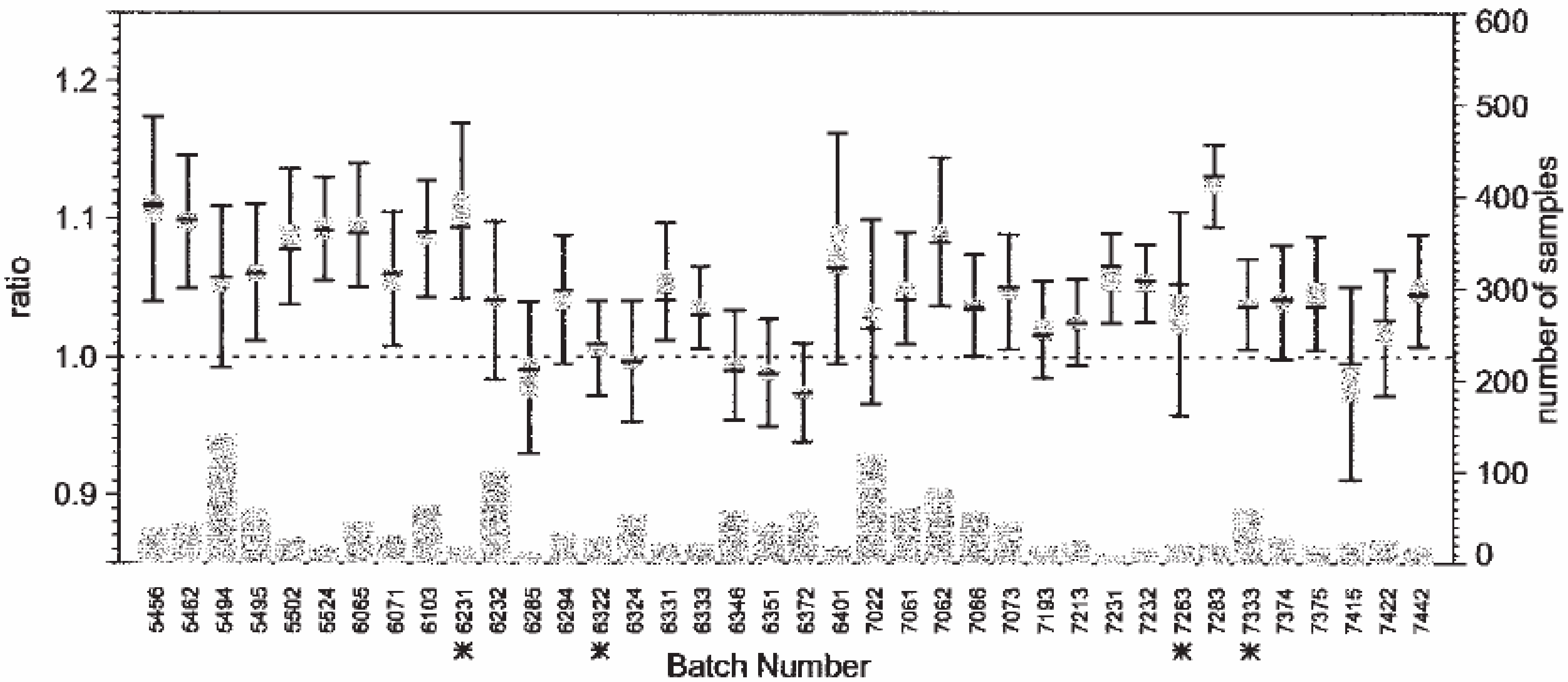


FIG. 6. Analysis of the ratio of the PWV observed by the MWR to the radiosonde as a function of radiosonde calibration batch. The quasi-box-plot indicates the spread (±1 standard deviation about the mean by the error bars and ±1 standard error of the mean by the gray box), the mean value (center of the gray box), and the median ratio (intermediate horizontal black line). The histogram at the bottom indicates the number of radiosondes analyzed in each batch. The standard deviation of the PWV ratio for the

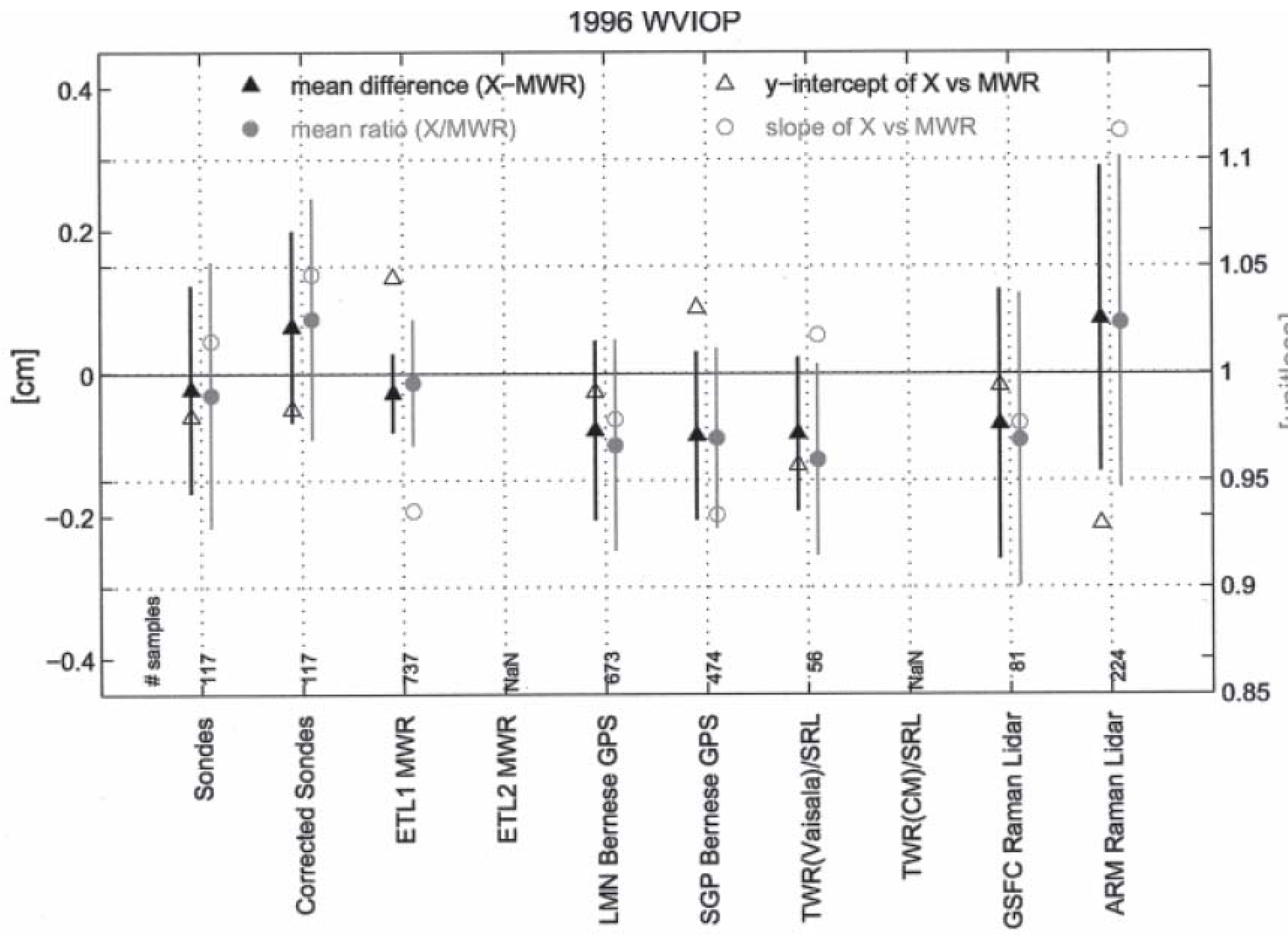


Fig. 10. Comparisons of various techniques that derive PWV with the ARM MWR for the 1996 WVIOP. Data have been averaged to 30-min values before comparison. The mean bias (solid triangles, left axis) and mean ratio (solid

Comparison of Upper Tropospheric Humidity (UTH) from GOES 6.7 m channel with radiosondes and Raman lidar

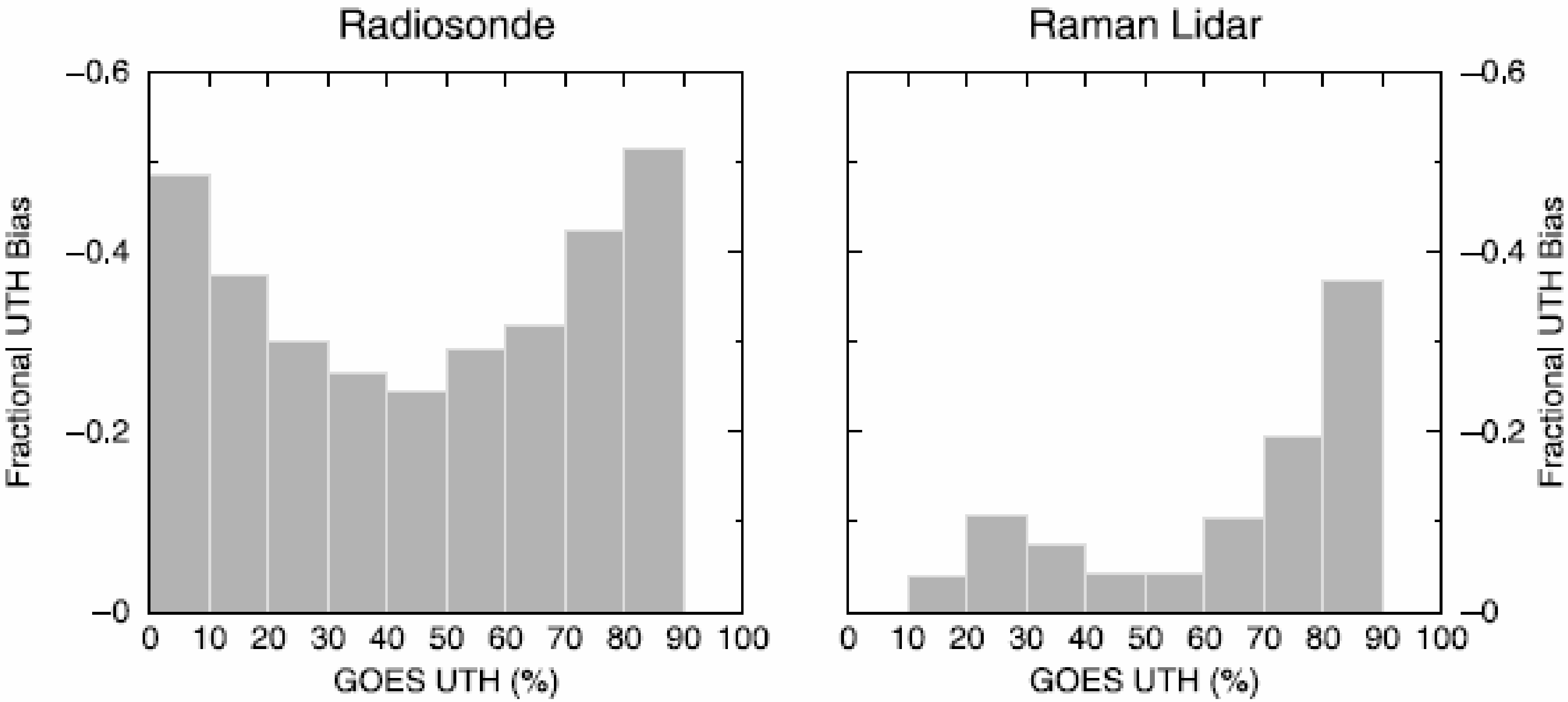


Figure 5. Fractional UTH bias relative to GOES plotted as a function of UTH for the Vaisala RS80-H radiosonde observations (left) and Raman lidar observations (right). The results are averages from all four IOPs.

Lessons learned

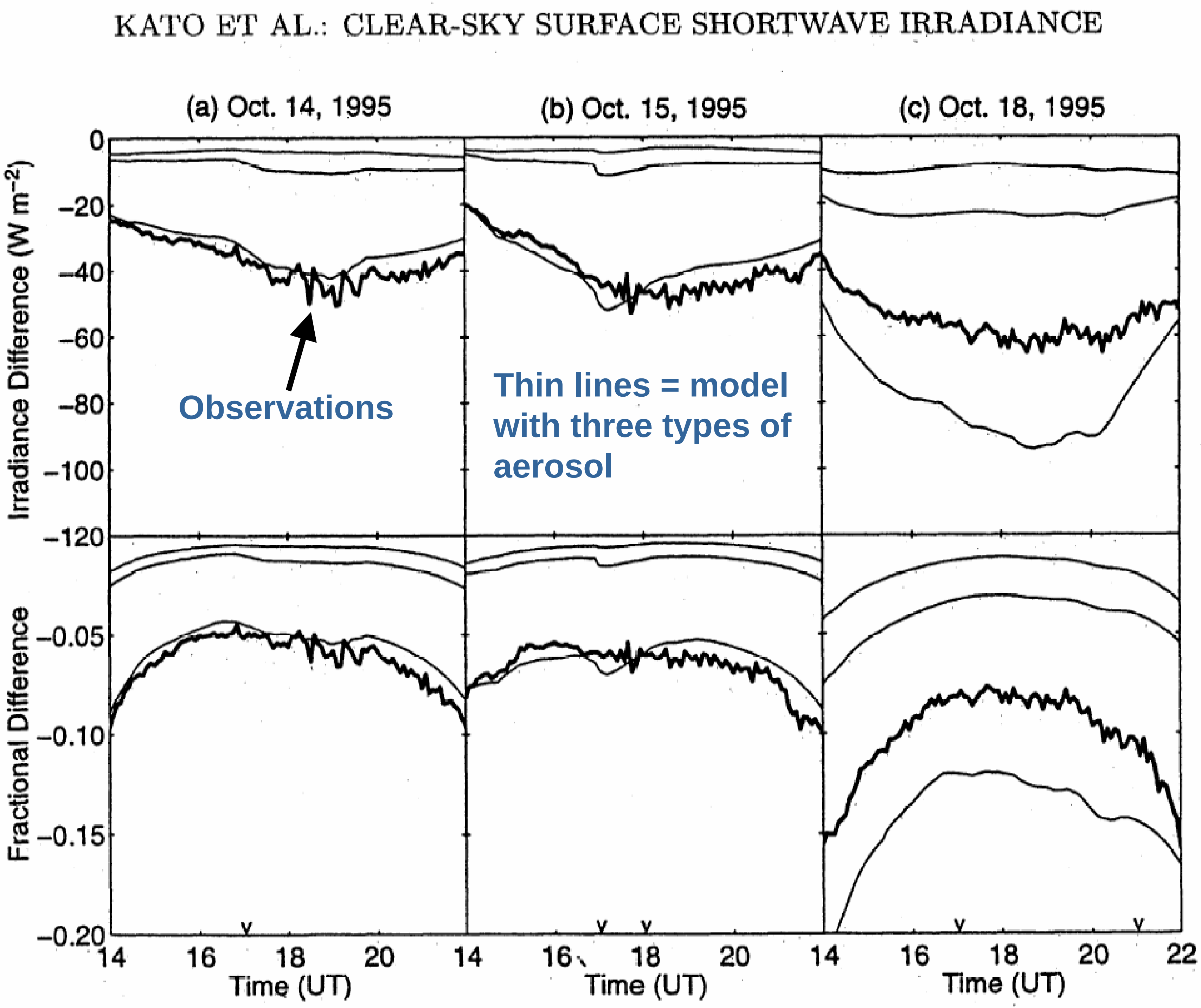
- We can measure water vapor to better than 2% in the column and better than 5% in upper troposphere
- Radiosondes have to be corrected to get close to this accuracy
- We have schemes to do that for Vaisala sondes and they seem to work pretty well (we can quantify this)
- This information does not seem to be penetrating the operational side of the field (case in point: we cannot get the US Weather Service to switch to Vaisala sondes at Barrow despite our identification of gross errors in the current sondes; Reason: climate record!)

Sidebar (courtesy of A. Tomkins)

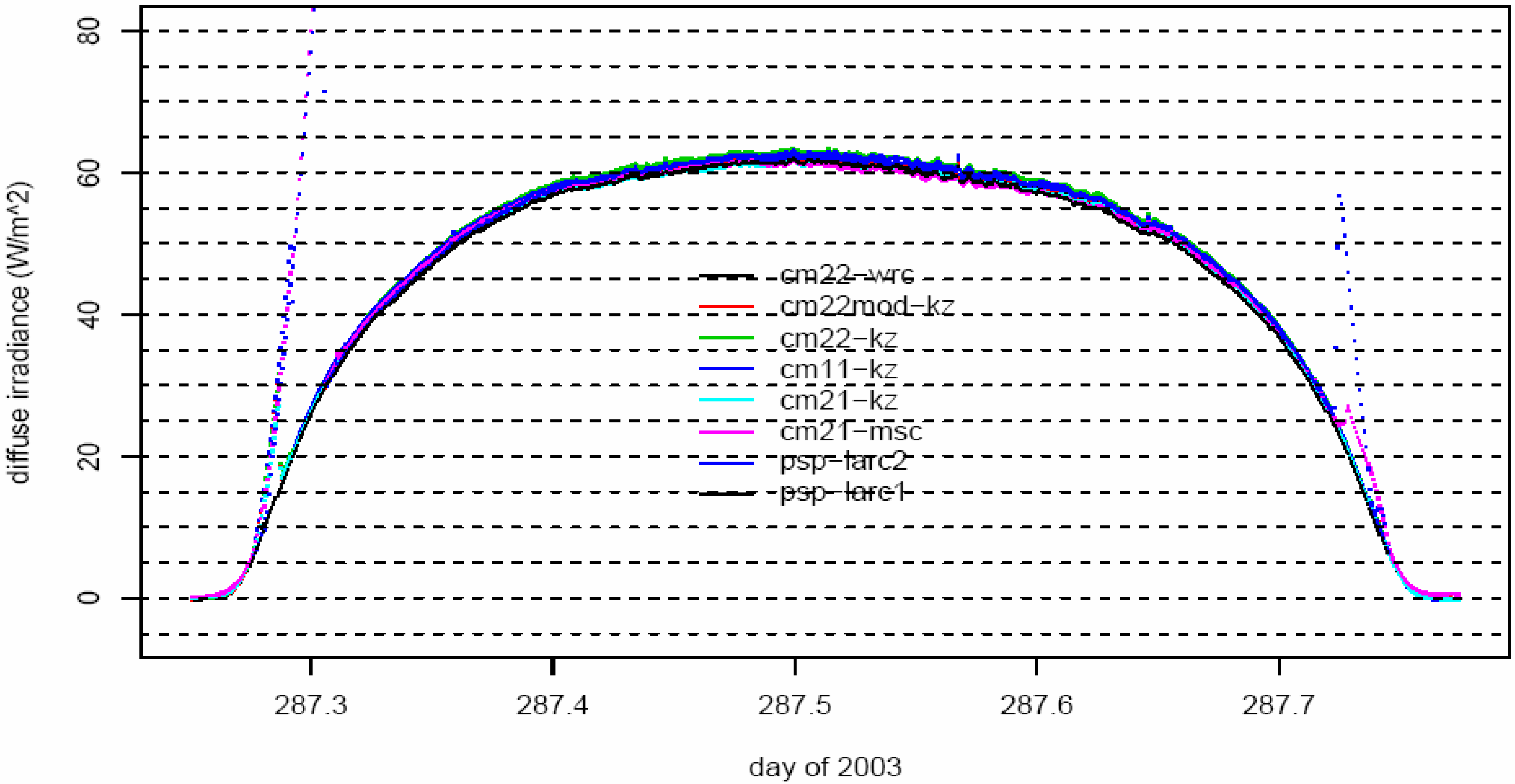
- Fact: we can measure PWV continuously (every 20 seconds) with a MWR to a column error < 2%
- MWR measurements are accurate over land (and water)
- Cost of MWR (bulk discount) is ~$100,000
- Cost of SSM/I is ? (but let's estimate $20,000,000)
- So for the cost of 1 SSM/I, I can deploy 200 operational MWRs at land sites
- If you have $20M, which would you prefer?

Story #2: Solar diffuse flux

- Started with identification of a discrepancy between measured and calculated clear-sky diffuse flux at the SGP
- Resulted in a large number of model investigations
- Spawned one aircraft IOP and two later ground-based IOPs
- Identified problem with thermal correction in solar broad-band radiometers
- Calculated a correction factor that is now standard across all thermal-pile radiometers


KATO ET AL.: CLEAR-SKY SURFACE SHORTWAVE IRRADIANCE
(a) Oct. 14, 1995
(b) Oct. 15, 1995
(c) Oct. 18, 1995
Irradiance Difference (W m^-2)
0
-20
-40
-60
-80
-100
-120
Observations
Thin lines = model with three types of aerosol
Fractional Difference
-0.05
-0.10
-0.15
-0.20
14
16
18
20
22
Time (UT)

Michalsky et al.: Diffuse Irradiance IOP in 2003 Comparison of 8 BB radiometers



Lessons learned

- Sometimes it ISN'T the model
- Continuous, well-calibrated measurements can produce new problems
- Well-designed experiments can identify the errors and correct them
- Why do you care? If you are removing the bias in your model (adding aerosol?) when it is a bias in the instrument ….

Story #3: Arctic clouds

An Assessment of ECMWF Model Analyses and Forecasts over the North Slope of Alaska Using Observations from the ARM Mixed-Phase Arctic Cloud Experiment

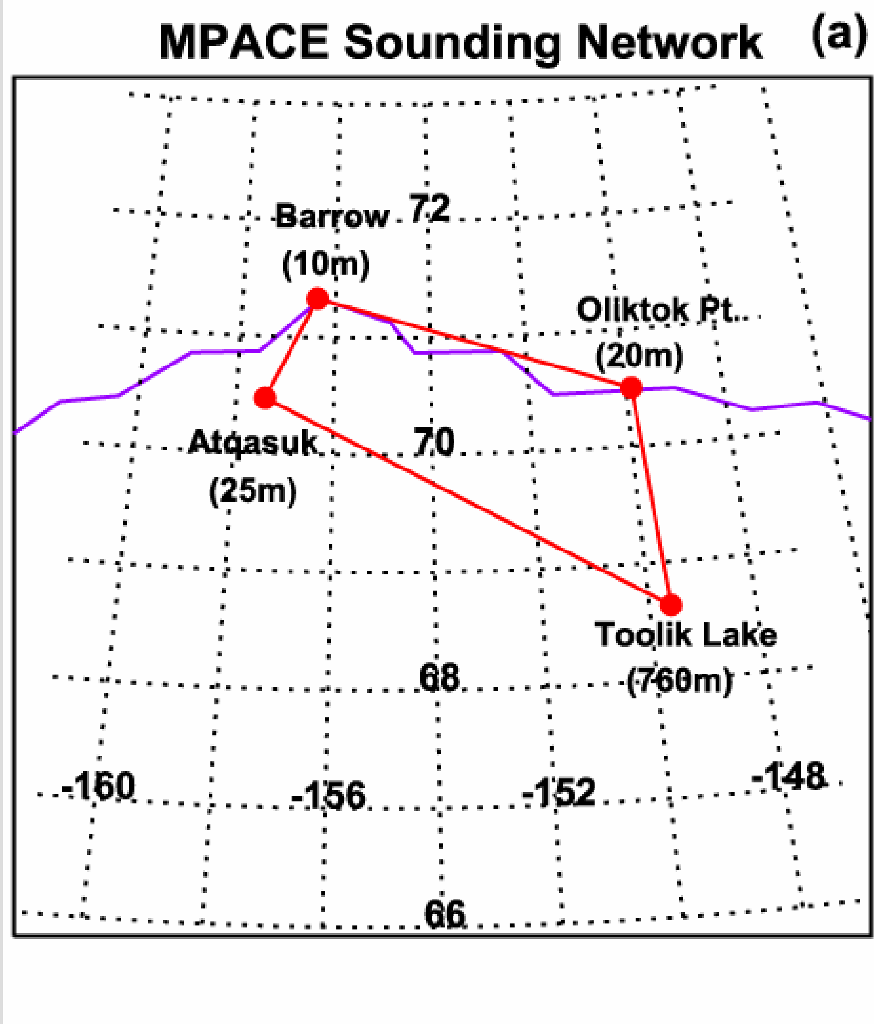
Shaocheng Xie, Stephen A. Klein, John J. Yio,

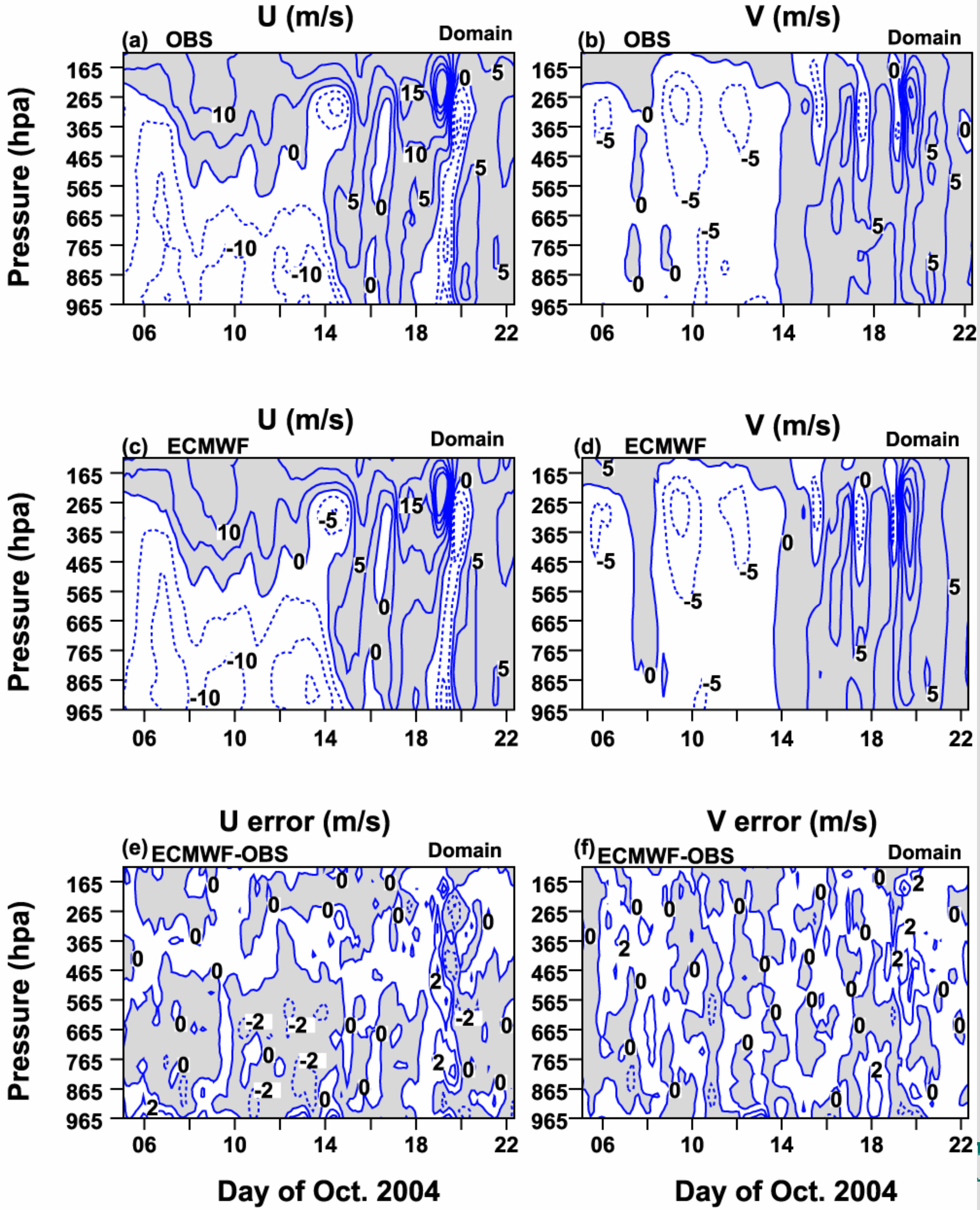
Anton C. M. Beljaars, Charles N. Long,

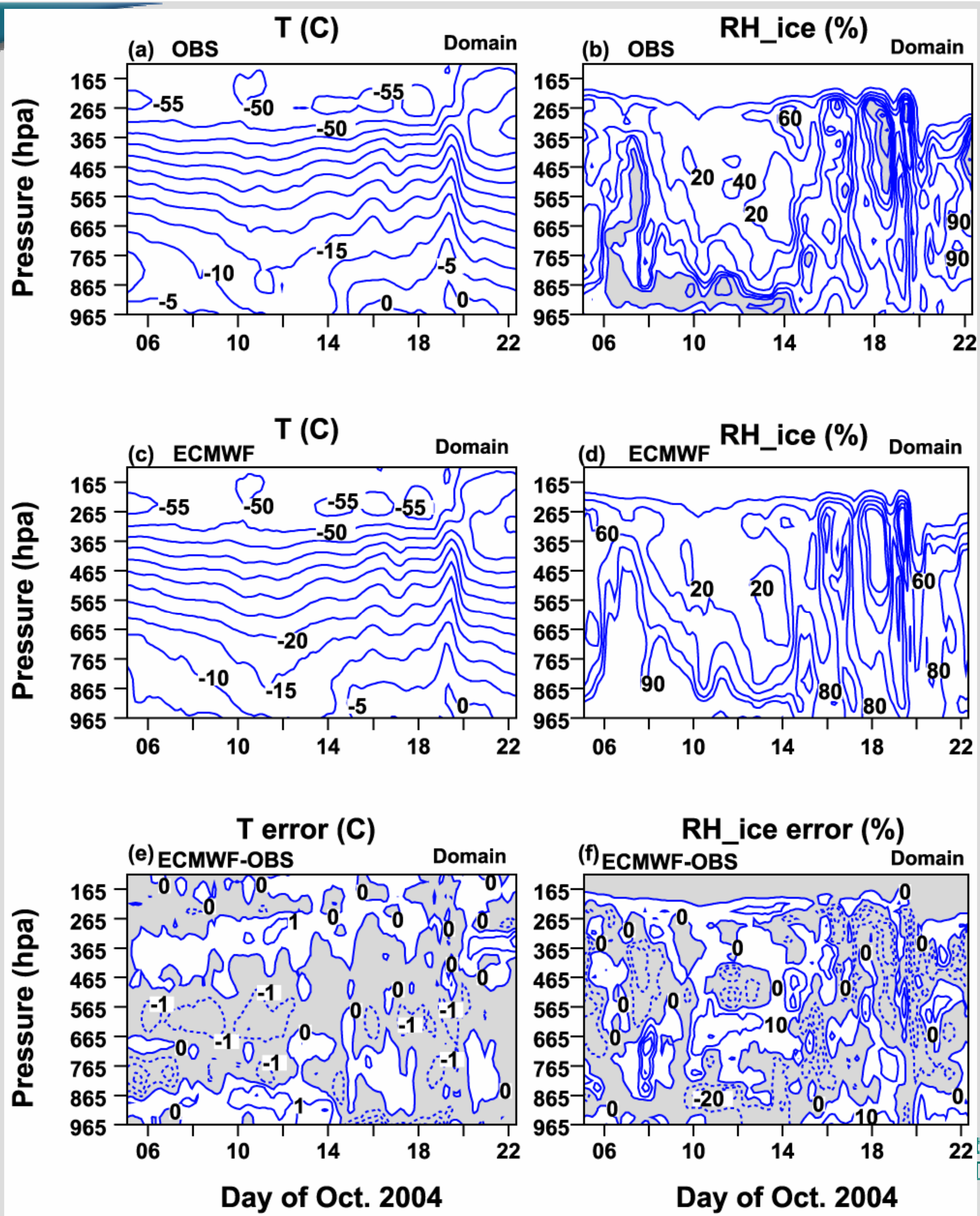
and Minghua Zhang

Mixed-Phase Arctic Cloud Experiment

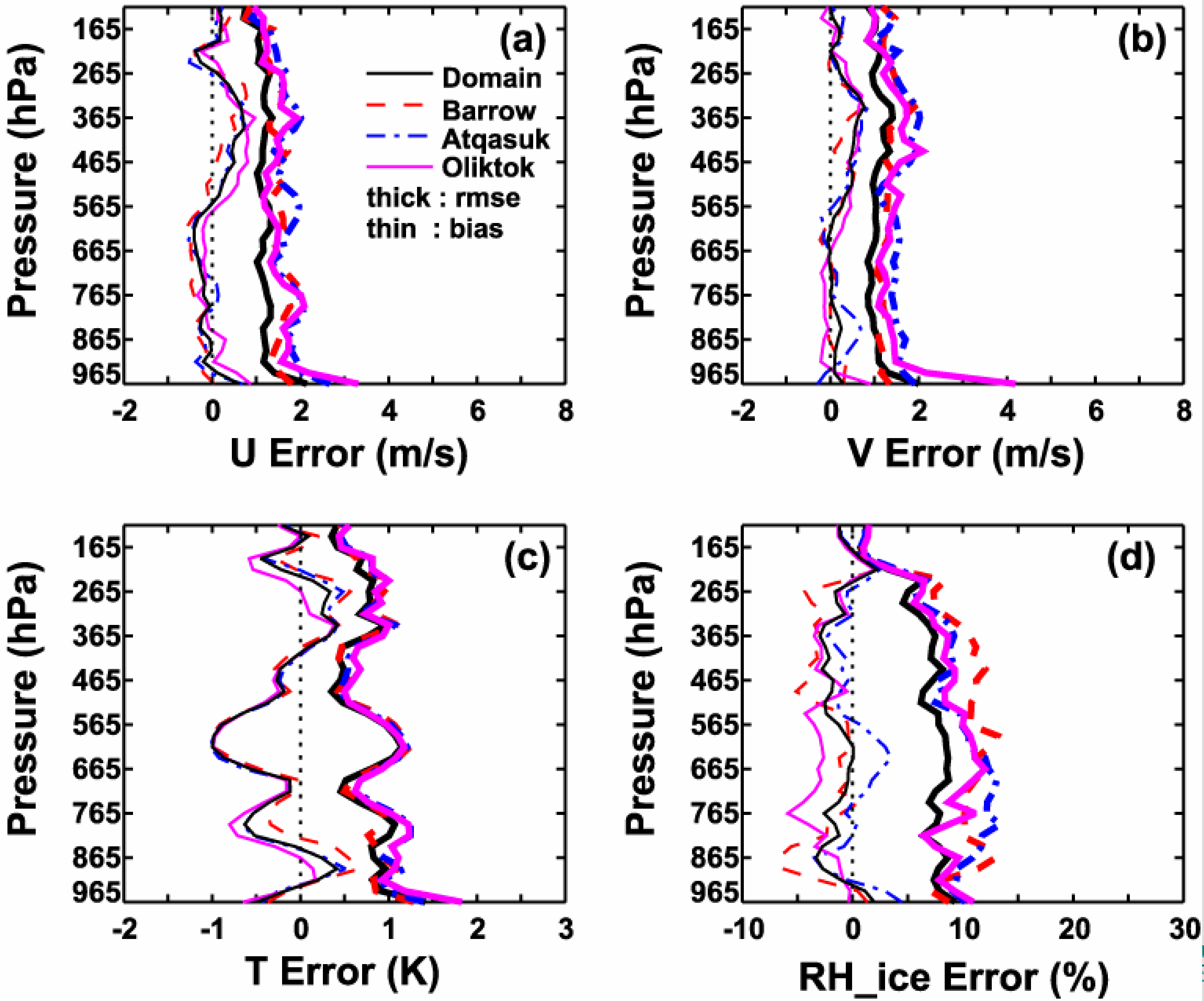
- Start with MPACE domain
- Create domain-average values
  - Variational analysis
  - Time and space averaging
  - 3-hourly values
- Compare with ECMWF analysis (6 hourly values)

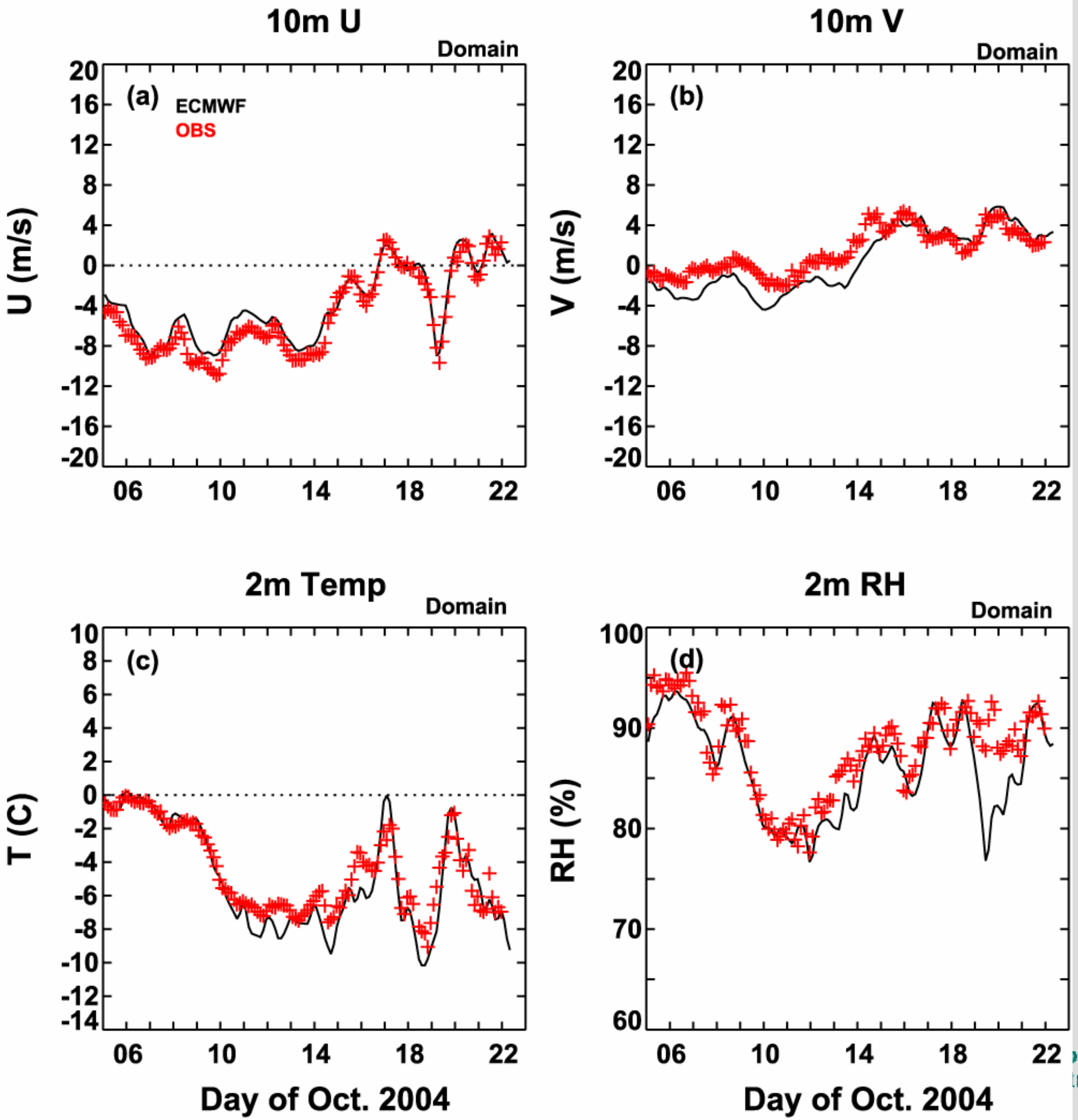



U (m/s)
(a) OBS
Domain
V (m/s)
(b) OBS
Domain
U (m/s)
(c) ECMWF
Domain
V (m/s)
(d) ECMWF
Domain
U error (m/s)
(e) ECMWF-OBS
Domain
V error (m/s)
(f) ECMWF-OBS
Domain
Pressure (hpa)
Day of Oct. 2004

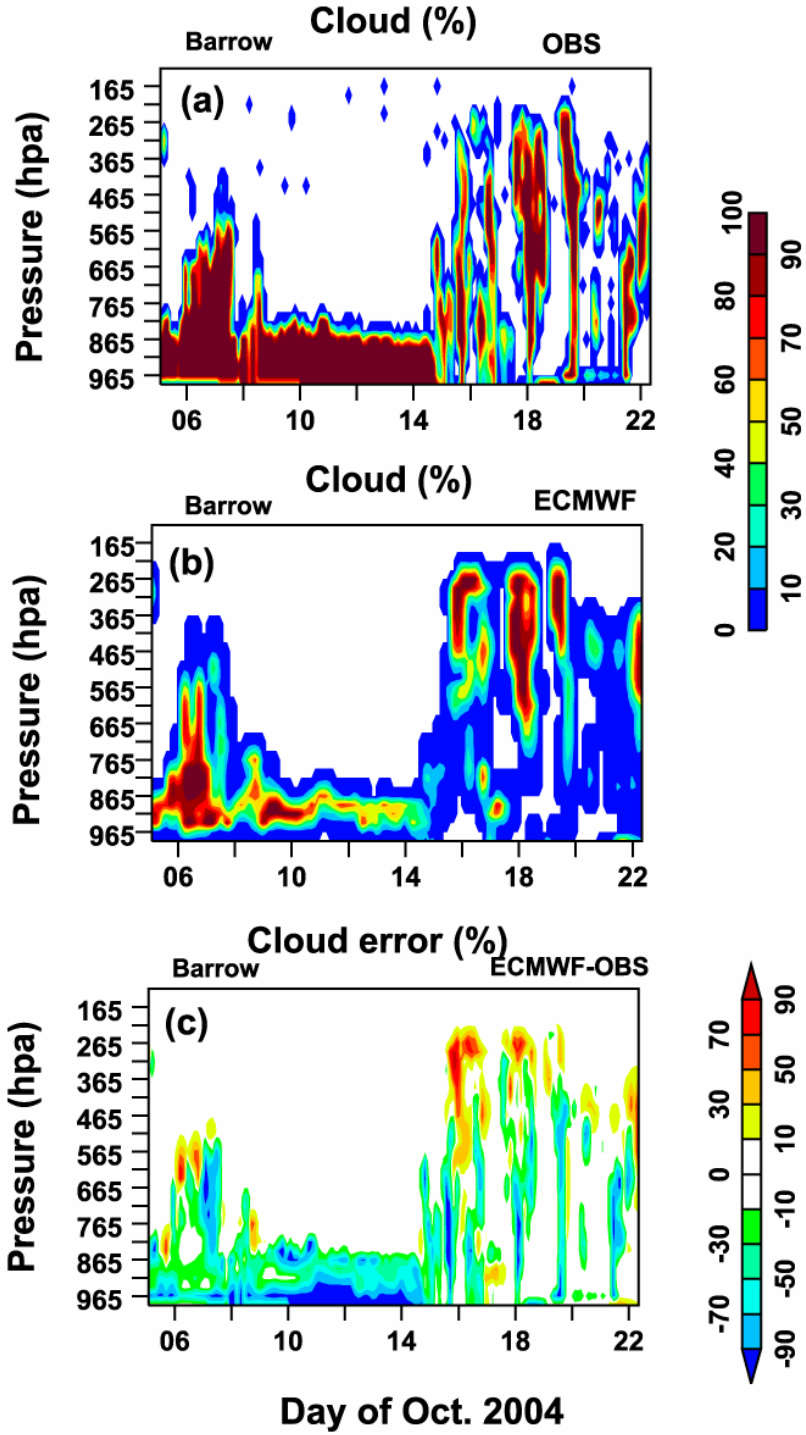

T (C)
(a) OBS
Domain
RH_ice (%)
(b) OBS
Domain
T (C)
(c) ECMWF
Domain
RH_ice (%)
(d) ECMWF
Domain
T error (C)
(e) ECMWF-OBS
Domain
RH_ice error (%)
(f) ECMWF-OBS
Domain
Pressure (hpa)
165
265
365
465
565
665
765
865
965
06
10
14
18
22
Day of Oct. 2004

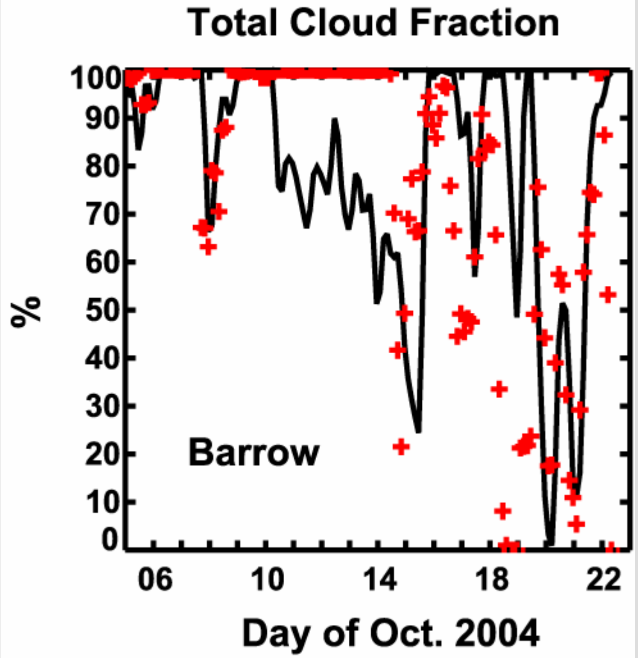
Mean Errors and RMS Errors

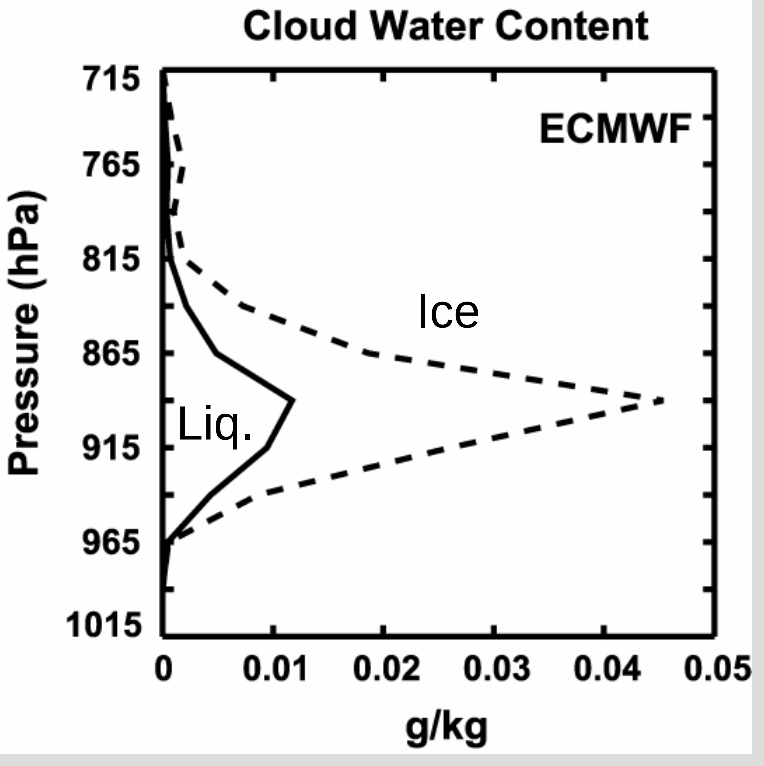


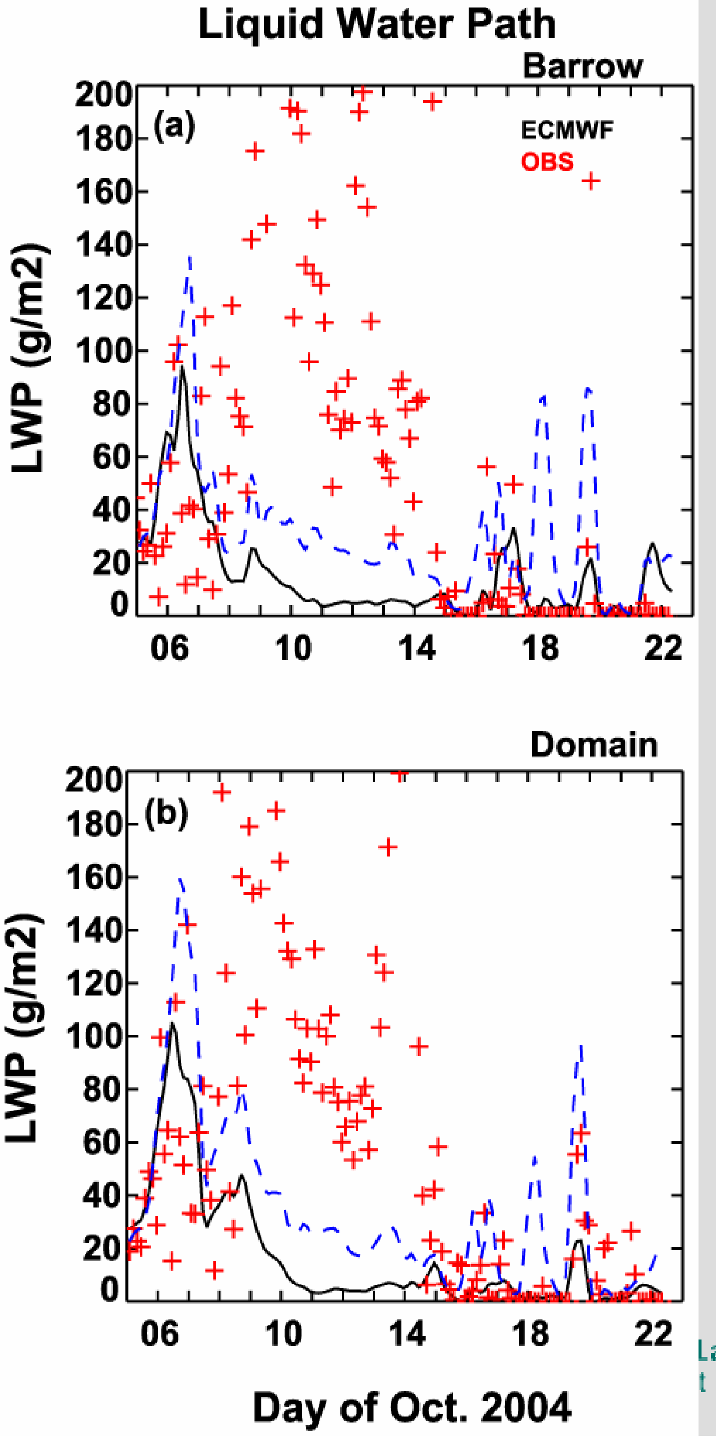


10m U
Domain
(a)
ECMWF
OBS
U (m/s)
10m V
Domain
(b)
V (m/s)
2m Temp
Domain
(c)
T (C)
2m RH
Domain
(d)
RH (%)
06
10
14
18
22
Day of Oct. 2004


Cloud (%)
Barrow
OBS
(a)
Pressure (hpa)
165
265
365
465
565
665
765
865
965
06
10
14
18
22
Cloud (%)
Barrow
ECMWF
(b)
Cloud error (%)
Barrow
ECMWF-OBS
(c)
Day of Oct. 2004
0
20
40
60
80
100
10
30
50
70
90
-90
-70
-50
-30
-10
0
10
30
50
70
90

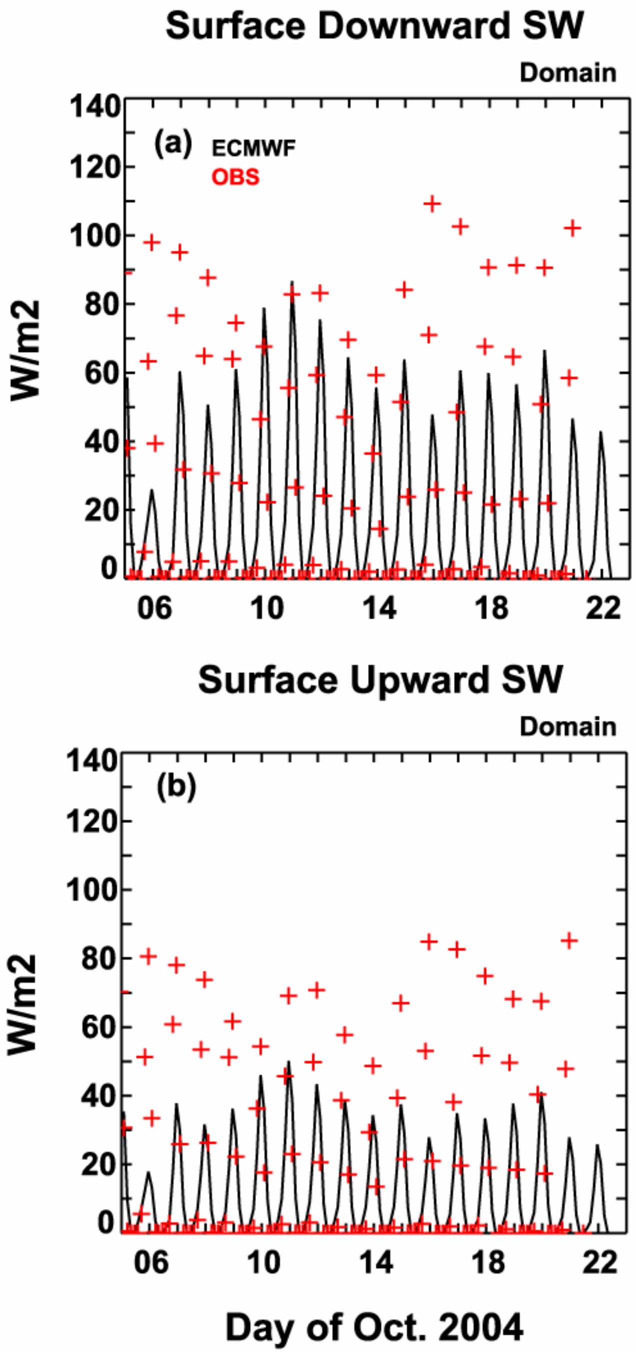

Total Cloud Fraction
%
Barrow
0
10
20
30
40
50
60
70
80
90
100
06
10
14
18
22
Day of Oct. 2004

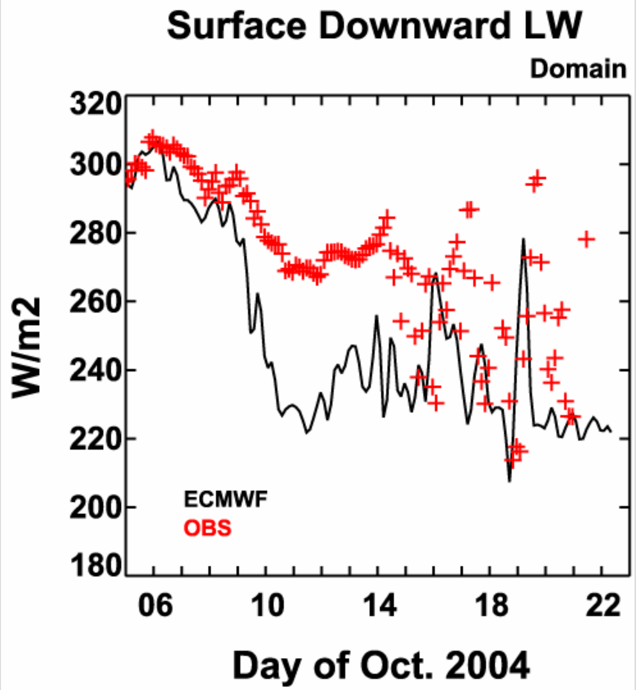




Battelle

Laboratory
t of Energy





Net Energy loss (surface to atm):

Observations	**-9.6 W/m2**
ECMWF	**-20.9 W/m2**

Lessons learned

- We can use field data to diagnose model biases
- In this case,
  - Best results are obtained over the domain => single point values may not be representative of the domain
  - Model does very well in capturing synoptic variation of large scale fields
  - Model represents cloud occurrence fairly well
  - Model clouds have too little liquid water and poor representation of ice/water vertical distribution
  - Results in a severe underestimate of downwelling LW and corresponding errors in surface radiation budget

Story #4: Operational comparison

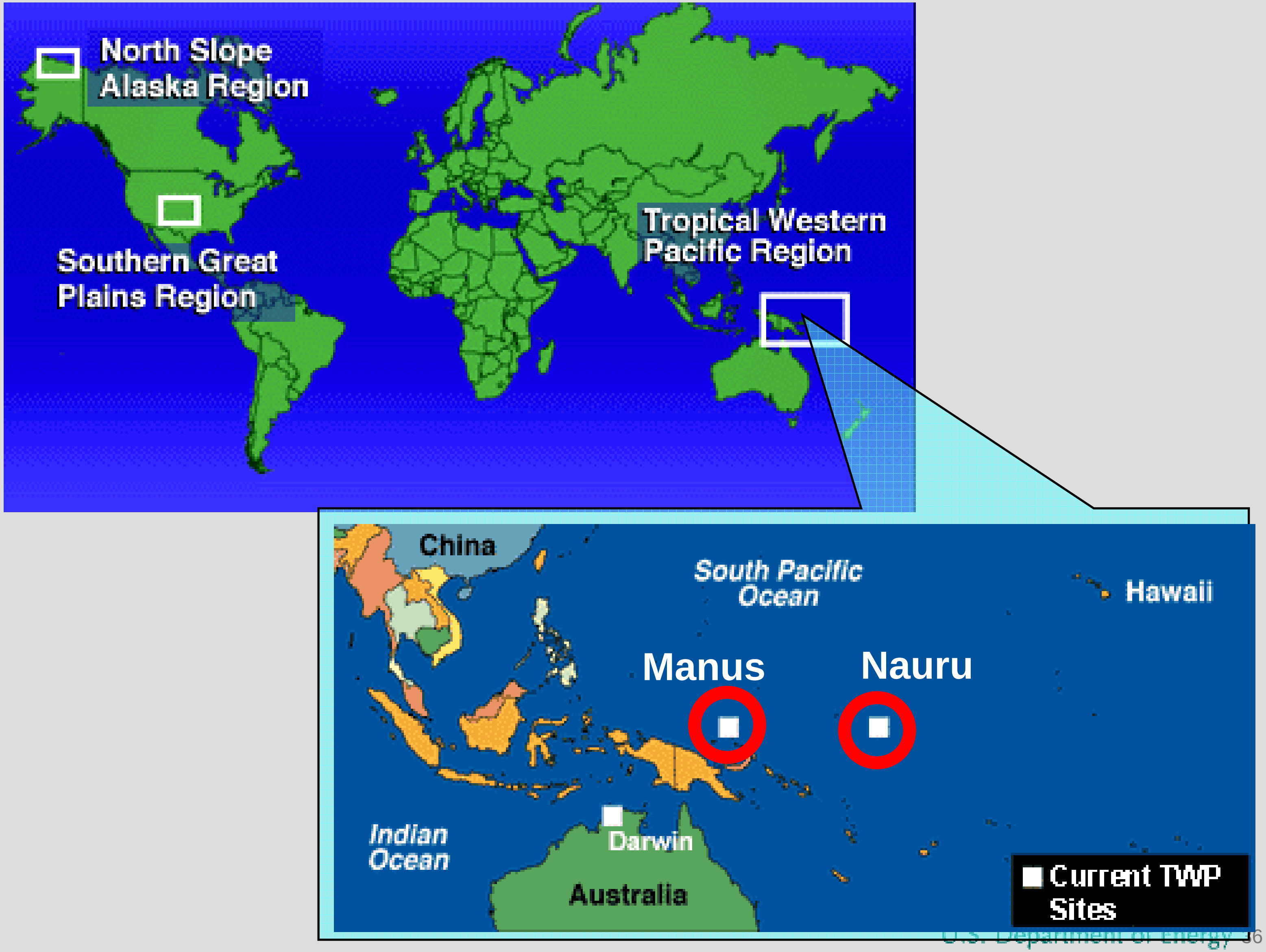
- CloudNet project
- PI: Prof. Anthony Illingworth, U. Reading
- Comparison of data from 3 European sites (Cabauw, Chilbolton, Palaiseau) with forecast model output
- Brilliant webpage
- Being extended to Lindenberg and ARM sites

Story #5: Heating Rate Profiles

Acknowledgements to:

Sally McFarlane, Jim Mather, Roger Marchand

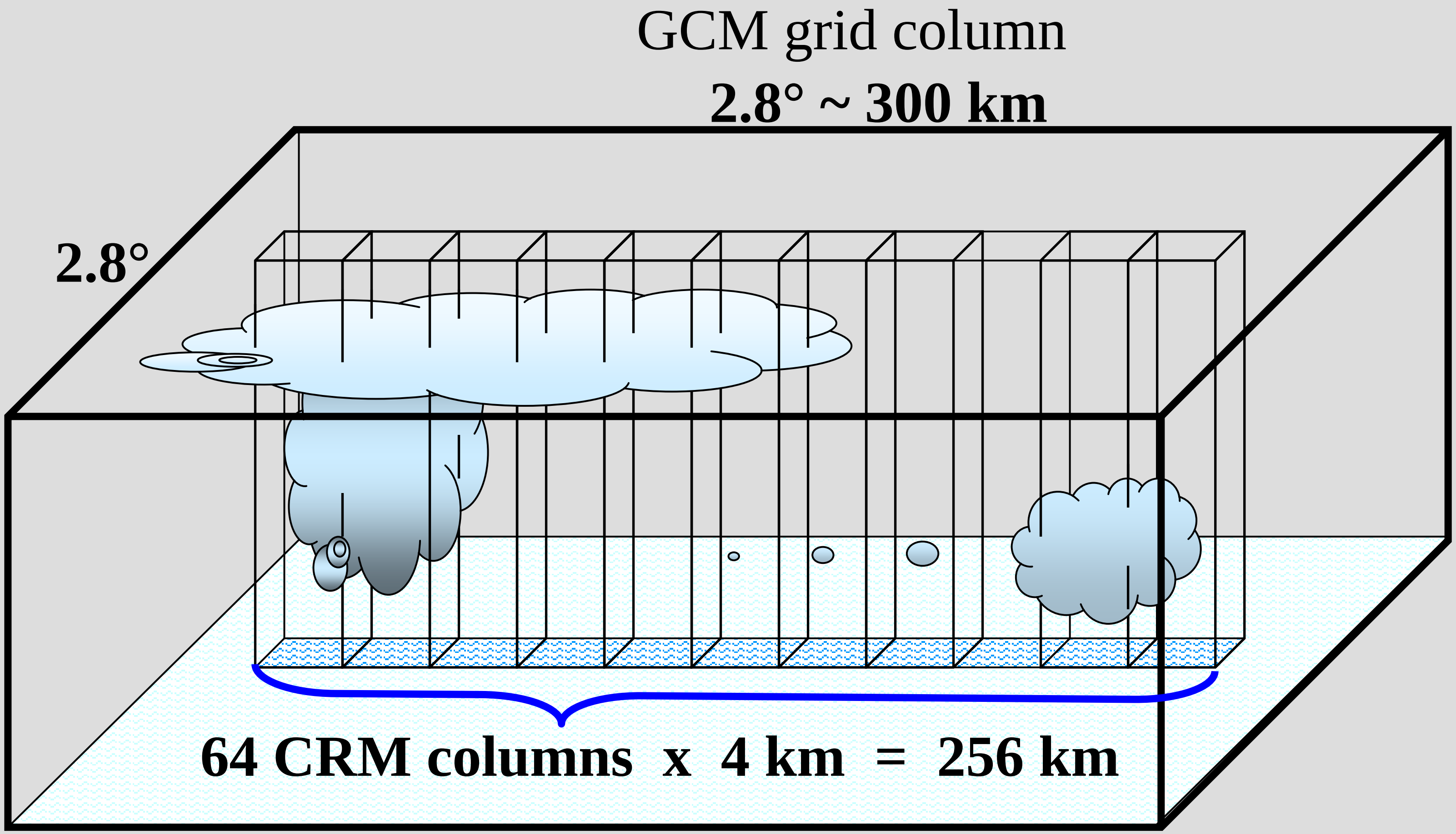
ARM Tropical Western Pacific Sites



ARM Data Processing

- Heating Rate Profiles
  - Temperature and water vapor profiles from radiosondes, scaled to microwave radiometer precipitable water and surface temperature
  - Vertical profiles of cloud microphysical properties calculated from ARM millimeter wave radar data (data has 10-second temporal and 45 m vertical resolution)
  - Sample the cloud properties every 5 minutes and perform radiative transfer only on the sampled profiles.
  - Calculate broadband fluxes and vertical profile of heating rates.

MMF Schematic

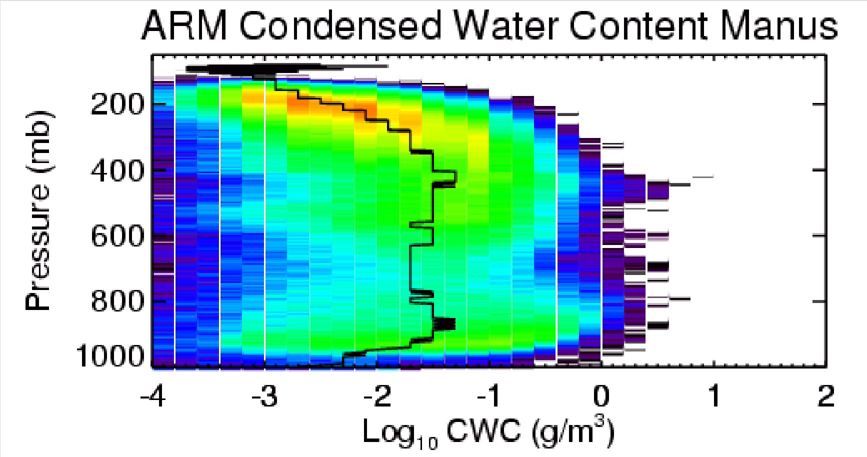


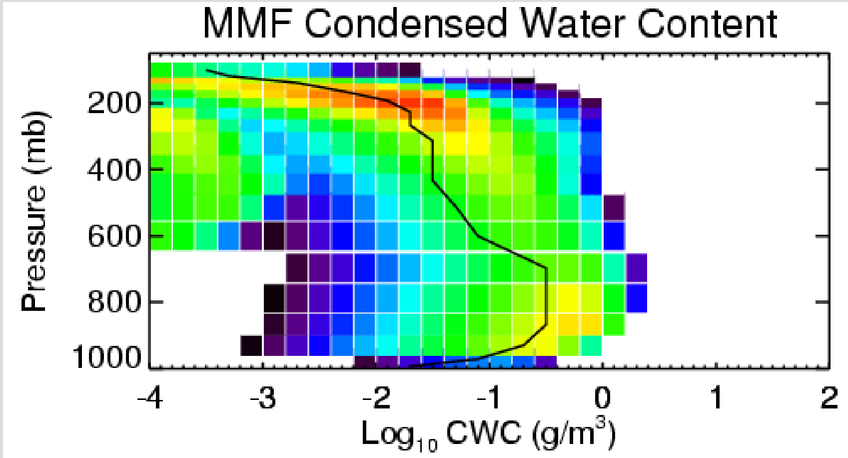
Simulations

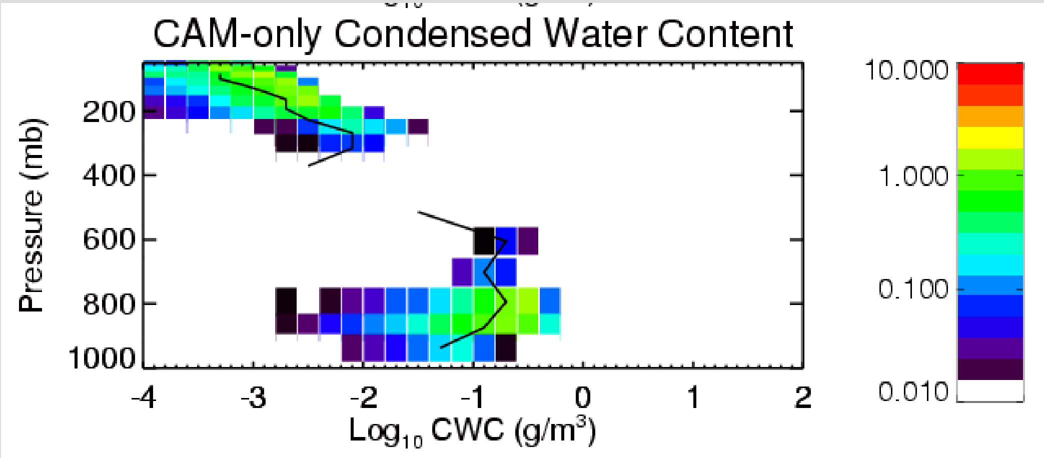
- MMF simulations with CSU model *
  - Run with observed SST values
  - Start in January 1998 and run into 2001
  - Second run for 2000 started from different initial conditions
- CAM simulations
  - Run with observed SST values for same period
- For the CAM-only runs, we examine output from the gridbox containing the ARM site
- For MMF runs, we examine the average over the 64 CRM columns within the gridbox containing the ARM site

** Model output available to any interested scientists*

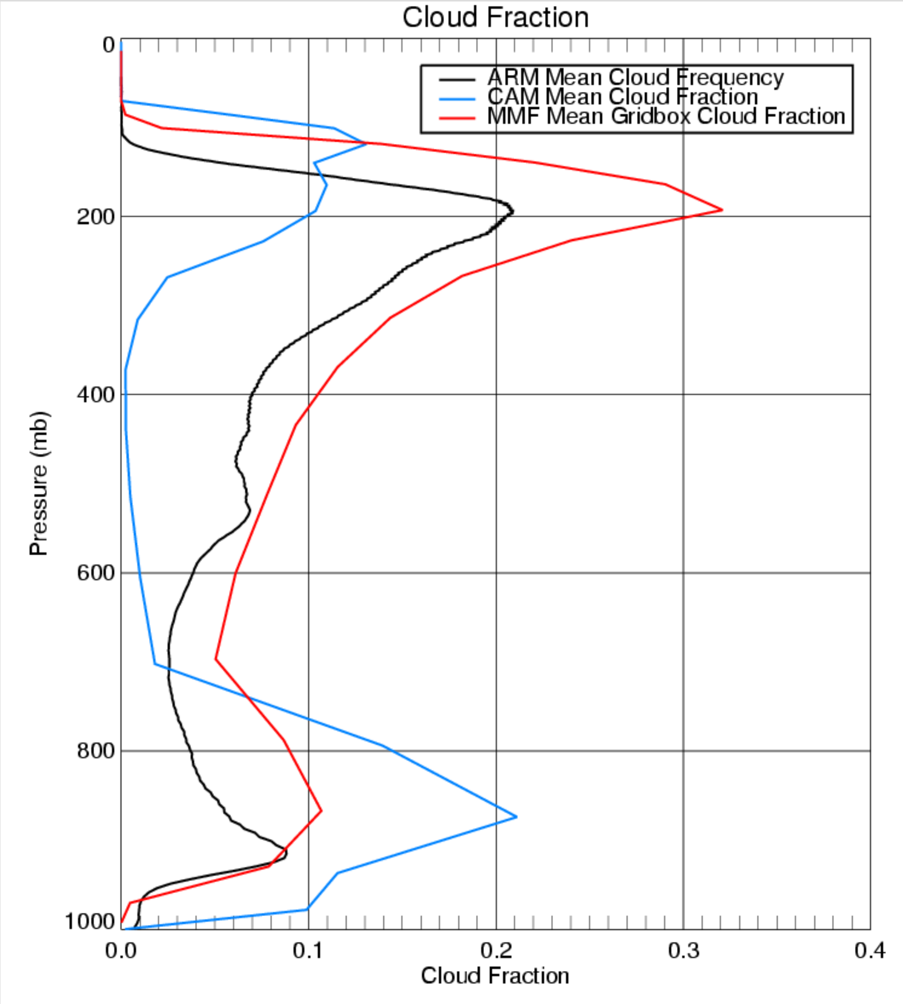
Condensed Water Frequency Distributions - Manus







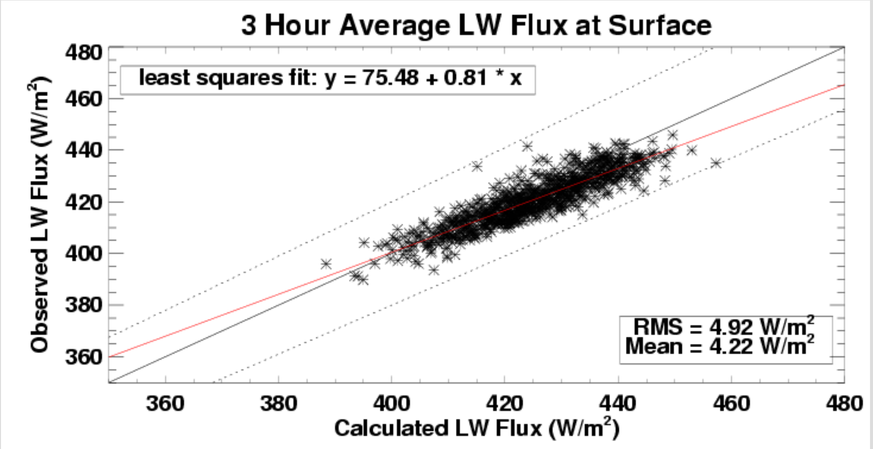
Cloud Fraction

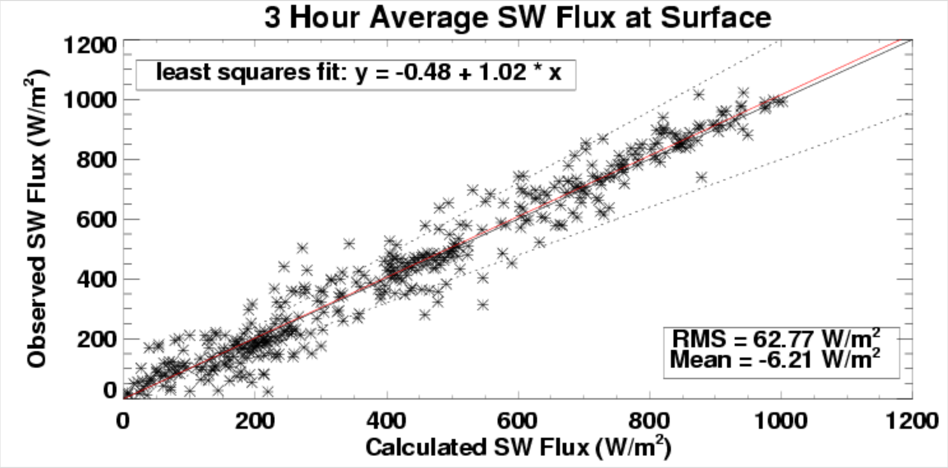


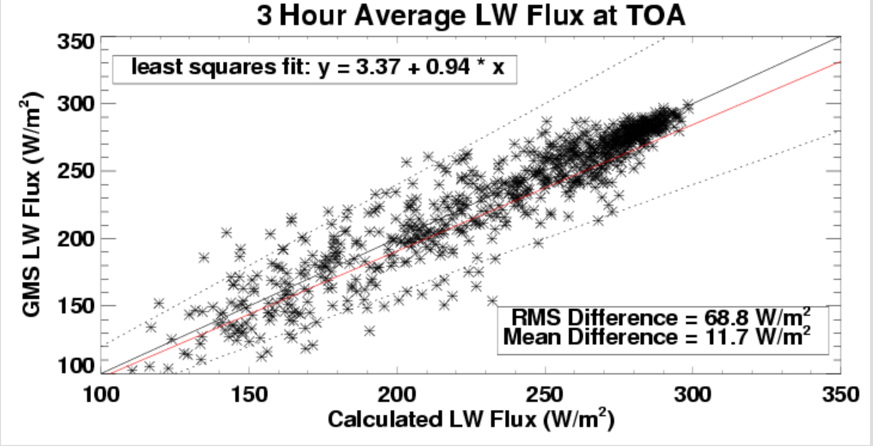
- ARM cloud frequency is percent of time reflectivity is greater than -40 dBZ at given level
- CAM cloud fraction is mean gridbox cloud fraction from cloud parameterization
- MMF cloud fraction is number of cloudy CRM columns within CAM gridbox

Surface and TOA Flux Comparisons (RT model and Data)

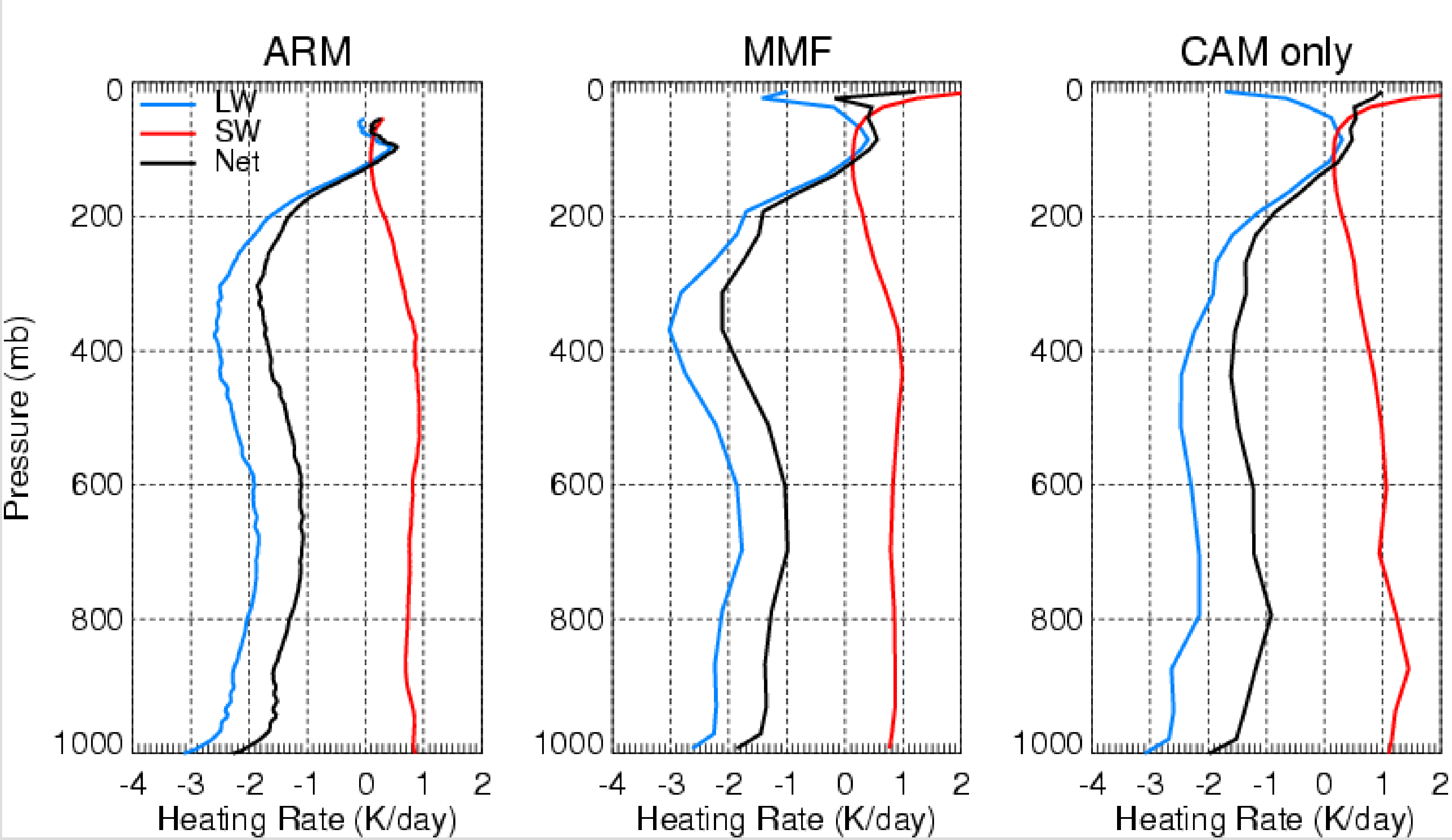
- SW surface comparisons show strong correlation, little bias
- LW surface comparisons are biased towards model overestimate of LW flux under cloudy conditions
- LW TOA comparisons show large scatter under cloudy conditions



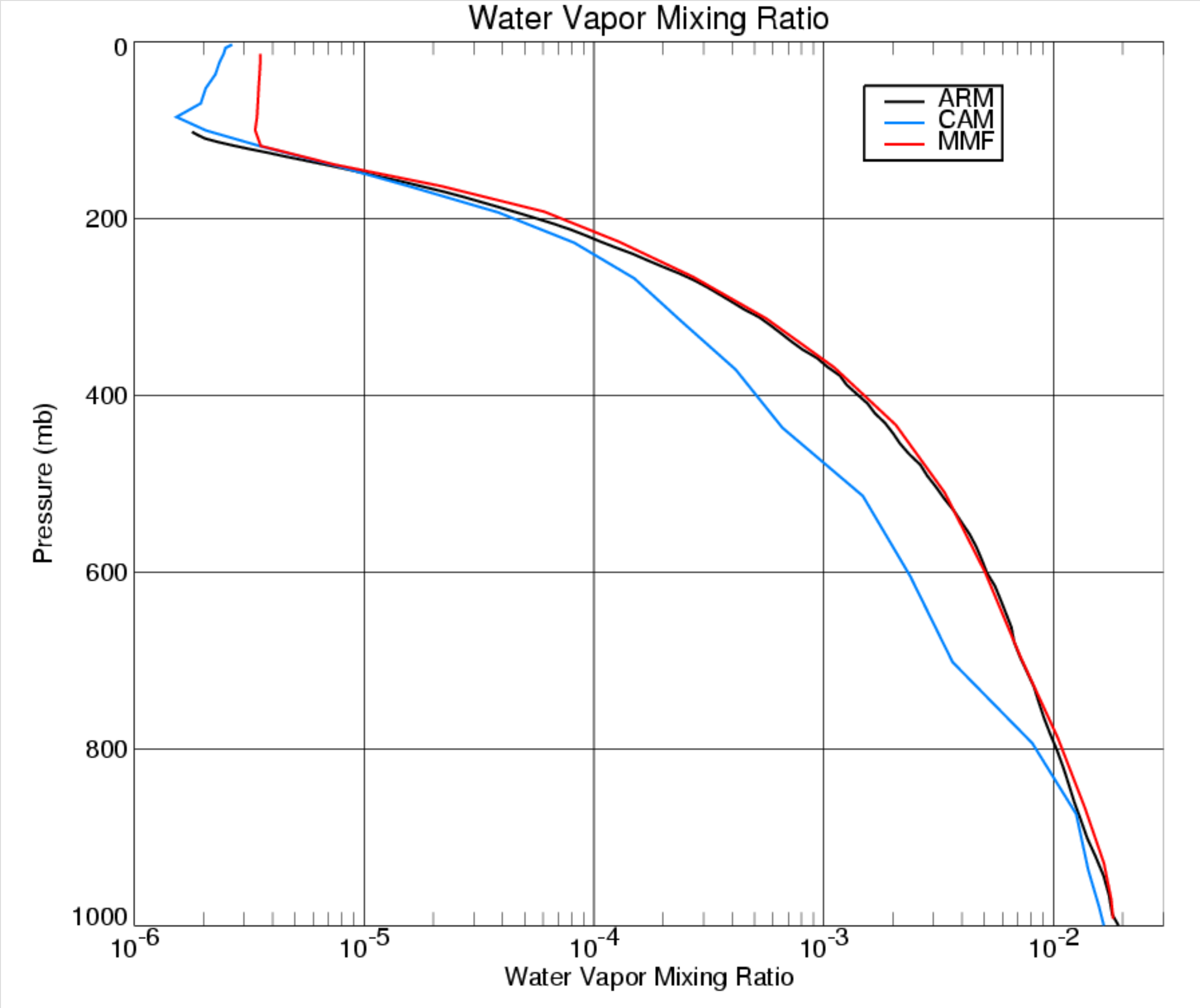




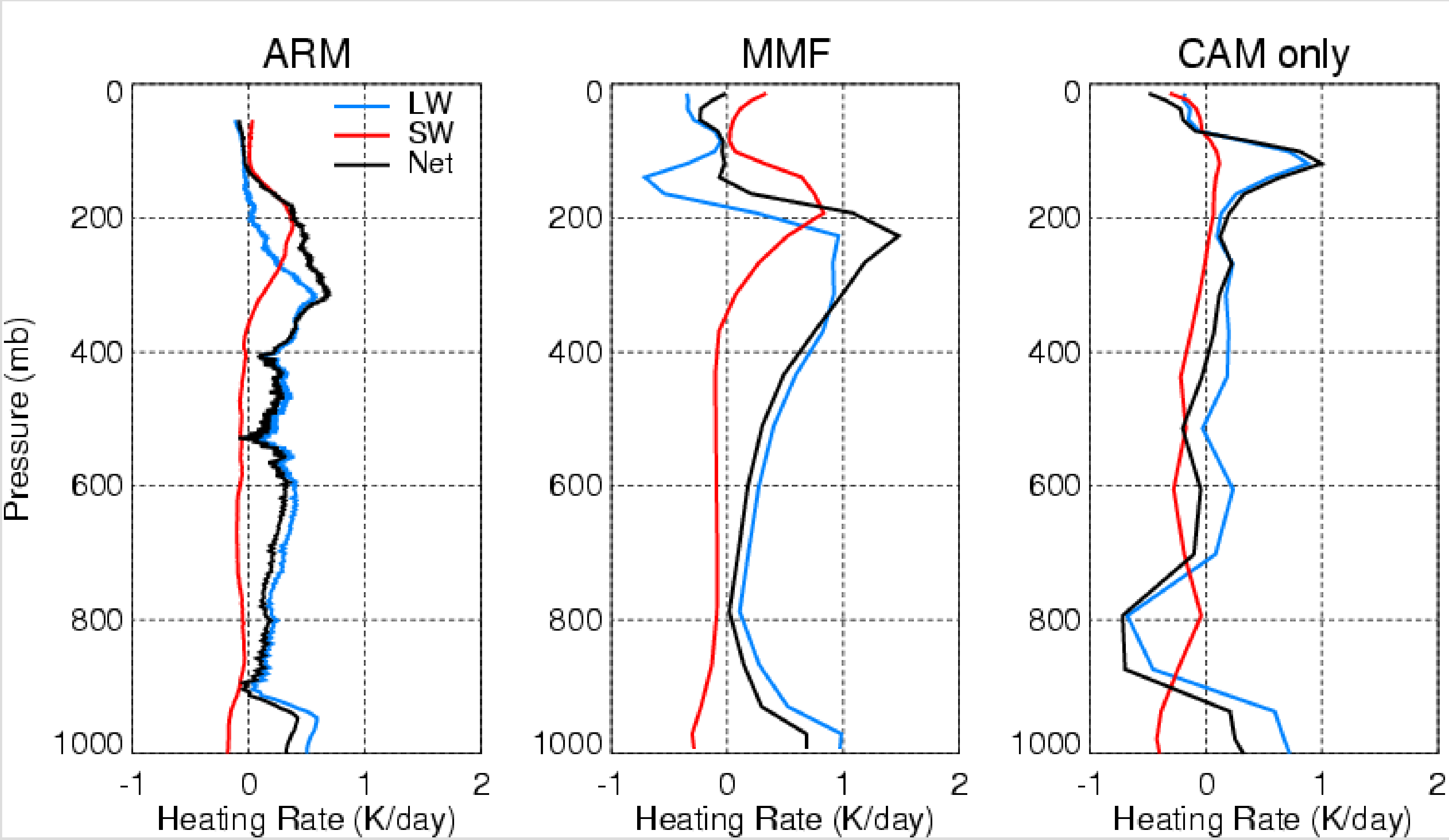
Heating Rates: Clear Sky

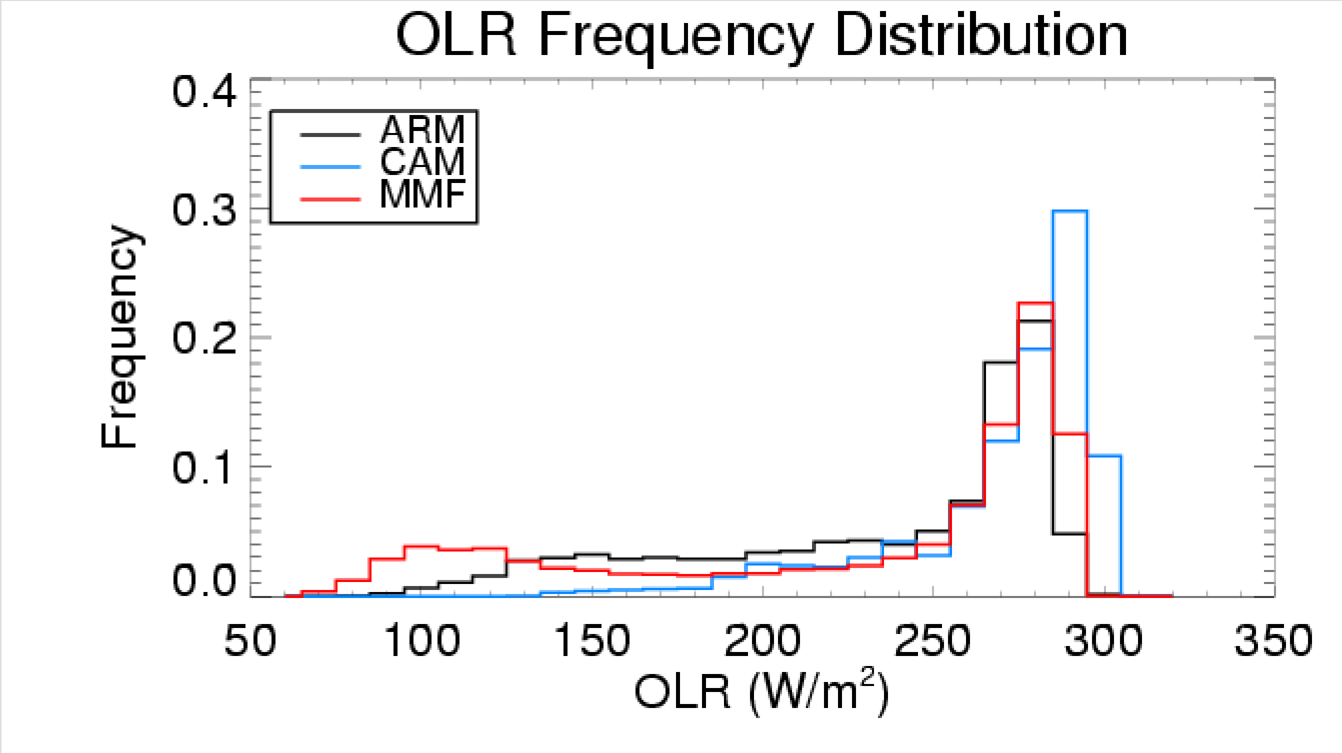


Average Water Vapor Profiles



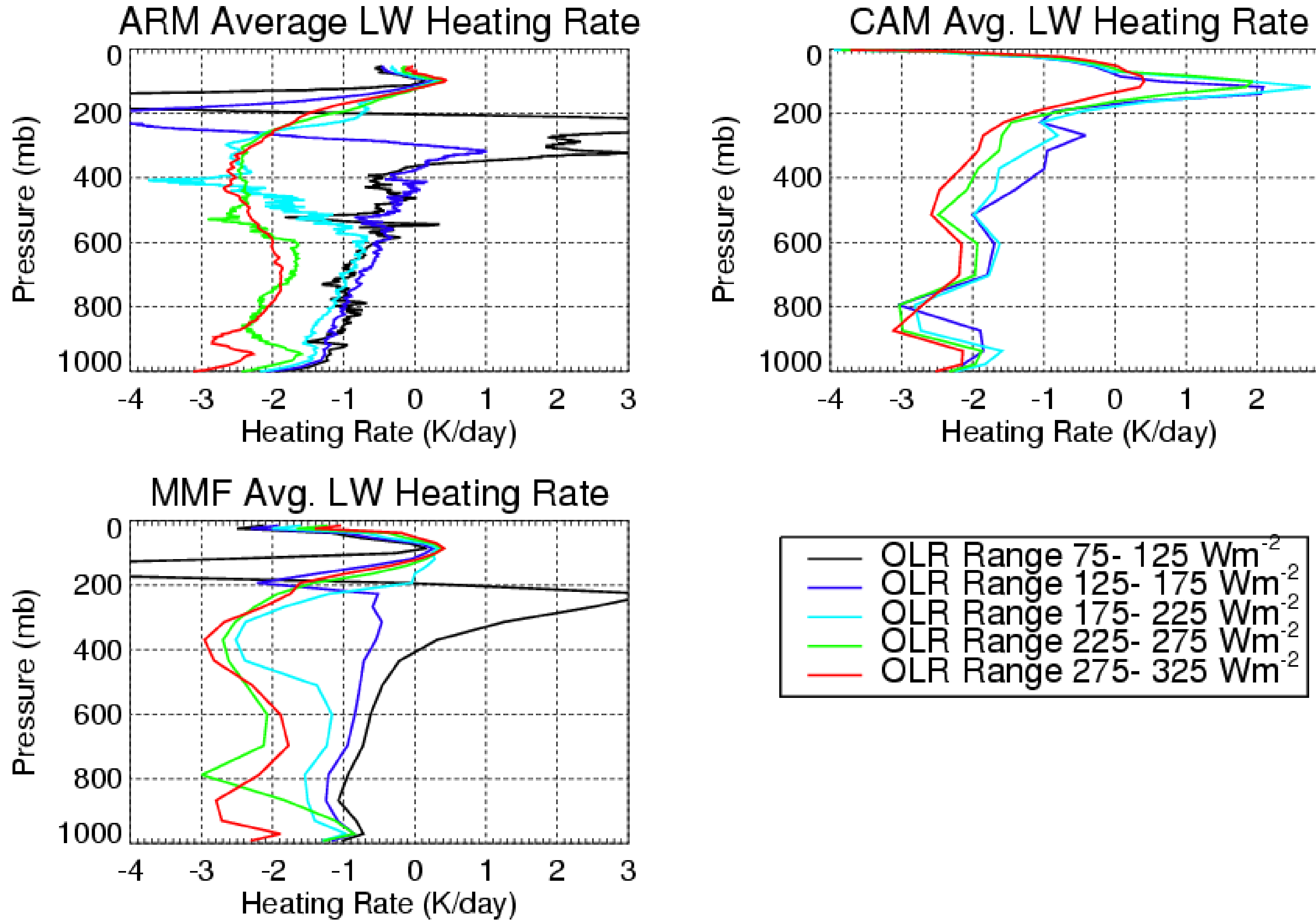
Heating Rates
All Sky – Clear Sky





- CAM has no OLR values below 175 W/m^2; larger frequency of very high OLR
- MMF/ARM frequency distributions similar; MMF has more very low OLR values

Heating Rates for Various OLR Ranges



Lessons learned

- We can use routine CAP data to carry out statistical comparisons to identify model biases
- Results from this study
  - MMF reproduces observed water vapor profile and clear sky heating rates better than CAM
  - Both models have problems with clouds, but CAM are more severe
  - Heating rates errors can be directly tied to deficiencies in cloud properties
  - Classification is a very helpful diagnostic

Summary

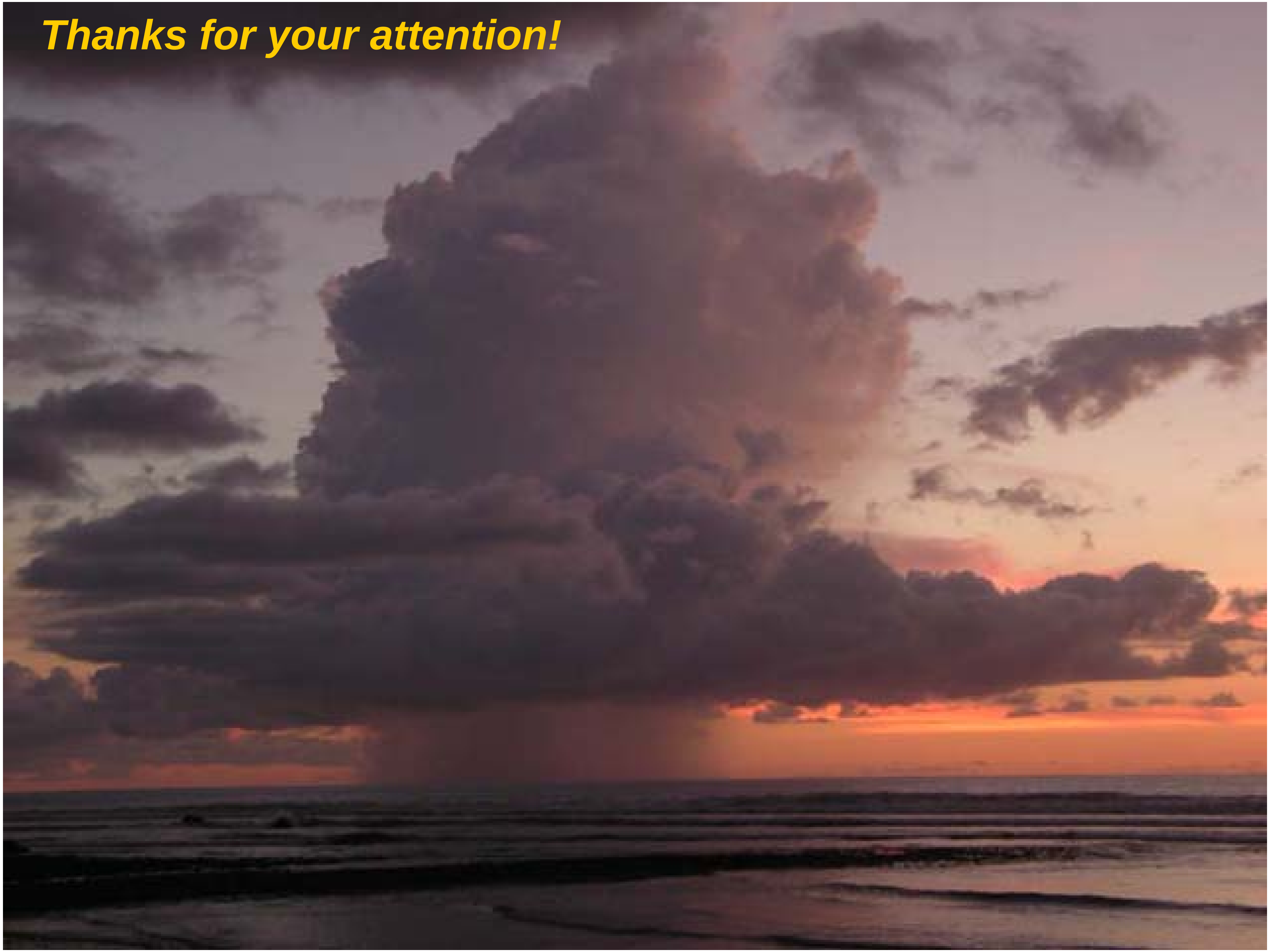
- Extensive networks (limited instrumentation)
  - Continuous well-calibrated observations of a few important variables
  - Provide constraint on model bias
- CAP sites (heavily instrumented, few in number)
  - Continuous, well-calibrated observations of many variables
  - Testing ground for fundamental physics and chemistry
  - Development framework for process models
  - Evaluation facility for satellite measurements
  - Evaluation facility for model performance

Summary

- Best way (currently) to evaluate model bias
  - Continuous comparison with CAP site data => multiple measurements of variables where possible
  - Identify discrepancies
  - Target field campaigns at one or more sites to study processes and assign cause to discrepancy
  - Correct
  - Continue

Issues

- Are we making the “right” measurements?
- Are there simple data streams that we could generate that would be useful?
- Do we know the absolute accuracy of the measurements?
- How good is the data quality assessment?
- How much measurement detail do we have to communicate to the user (NWP) community in order to make the measurements useful?


Thanks for your attention!