Observation and Analysis Opportunities Collaborating with the ICON and GOLD Missions
Sept 27-28, 2016 • NCAR HAO, Boulder CO

The Roles of Data Assimilation in a Science Investigation

Anthony J Mannucci
Xiaoqing Pi
Attila Komjathy
Olga Verkhoglyadova
Xing Meng
Bruce T Tsurutani

Jet Propulsion Laboratory, California Institute of Technology

Chunming Wang and Gary Rosen

University of Southern California

September 28, 2016

Copyright 2016. All Rights Reserved.
Government sponsorship acknowledged: NASA/NSF Partnership for Collaborative Space Weather Modeling

Overview

- **Origins of data assimilation**
- **Addressing the ICON and GOLD science questions?**
- **Data assimilation in the ionosphere-thermosphere**
- **Addressing ICON and GOLD science questions**
- **Concluding remarks**

ICON and GOLD Science Goals

- What are the sources of strong ionospheric variability
- How do large-scale atmospheric waves control the ionosphere at low latitudes?
- How do ion-neutral coupling processes respond to increases in solar forcing and geomagnetic activity?
- How do geomagnetic storms alter the temperature and composition structure of the thermosphere?
- What is the global-scale response of the thermosphere to solar extreme-ultraviolet variability?
- How does the nighttime equatorial ionosphere influence the formation and evolution of equatorial plasma density irregularities?
- How significant are the effects of atmospheric waves and tides propagating from below on thermospheric temperature structure?

Conundrum

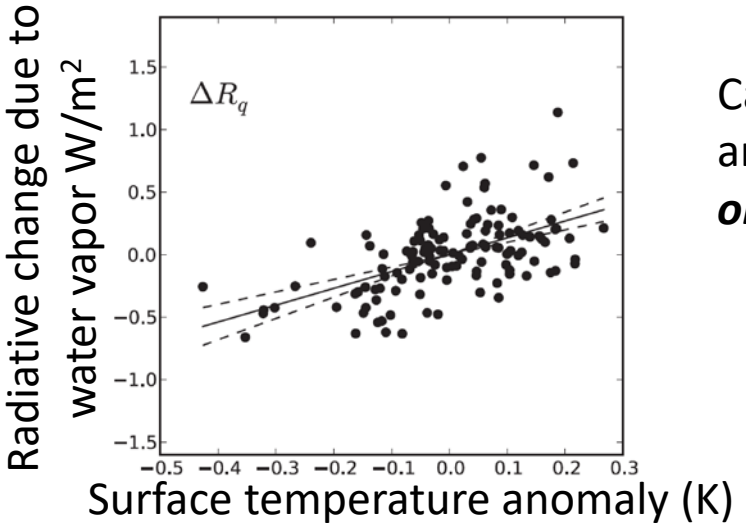
- **Numerical simulations with widely used coupled ionosphere-thermosphere models can address some of these science goals**
  - **Resolution limits notwithstanding**
- **How do we effectively capture the new science using ICON and GOLD?**

Origins of Data Assimilation

- **"Data assimilation was developed to improve forecasts"**
- **DA, through a process called "3DVAR", determines the initial state of a forecast model that is run forward in time**
  - **3DVAR is a spatial interpolation procedure**
  - **Uncertainty assignment and quality control are important to obtain good forecasts**
- **DA is not about discovery? It is about using models that embody "sufficient physics" to predict the future**

Or *Is* Data Assimilation About Discovery?

- **Atmospheric scientists use "reanalyses" (consistently processed data assimilative model fields) as a form of observation, to make discoveries about Earth's atmosphere**
- **Uses a dynamical time step (short-range forecast)**
- **Earth science missions provide information that is missing or poorly represented in the reanalyses**



Caption: Scatterplot of the temperature…flux anomalies vs surface temperature anomaly in ***the observations (using the ERA-Interim reanalysis).***

Dessler, A. E., M. R. Schoeberl, T. Wang, S. M. Davis, and K. H. Rosenlof (2013), Stratospheric water vapor feedback, *Proceedings of the National Academy of Sciences*, *110*(45), 18087–18091, doi:10.1073/pnas.1310344110.

Data Assimilation in the I-T

Model	Approach	Driver Estimation	Name	Comment
GloTEC	Tomography w/ climatological background	N	Yue et al.	
IDA4D	Kalman w/ climatological background	N	Bust et al.	No dynamical time step
Kalman w/ IRI-Background	Kalman w/ climatological background	N	Lin et al.	
GAIM/USU Gauss Markov	Kalman w/ climatological (physics) background	N	Schunk et al.	
GAIM/JPL Band Limited Kalman	Kalman w/ physics	N	Pi et al.	
EMPIRE	Regional w/ Driver Estimation	Y	Bust, Datta-Barua et al.	
GAIM/JPL 4DVAR	Adjoint-based variational w/ physics	Y	Pi et al.	
GAIM/USU Full Physics	Ensemble w/ physics	Y	Schunk et al.	
GAIM/USU Mid-Low Electro	Ensemble w/ physics	Y	Eccles et al.	Dyn time step
GAIM/USU High Lat	Ensemble w/ physics	Y	Eccles et al.	
USU Thermosphere	Ensemble w/ physics	N	Scherliess et al.	
TIEGCM based Ensemble	Ensemble Kalman w/ Coupled T-I physics	Y	Matsuo et al.	
Multi-Model Ensemble USU/JPL	Multi-model	Y	Schunk et al.	

Apologies for omissions (e.g. RT IRI). Names are “representative.”

Assimilation Approaches

Kalman – "time-sequential least squares fitting"

- New state = weighted_sum_of (prior state, new observations–projected onto state)
- Weighting based on: covariance of prior state and observation covariance
- Time update may involve dynamics or something simpler (e.g. nothing)
- Covariance is updated also (expensive calculation)

4DVAR – "optimization procedure"

- Minimize difference between prior state and new observations (projected), w/ a penalty for deviating too far from prior state
- Use adjoint model to efficiently compute gradient (wrt state) of the difference
- State includes electron density, thermospheric and other drivers

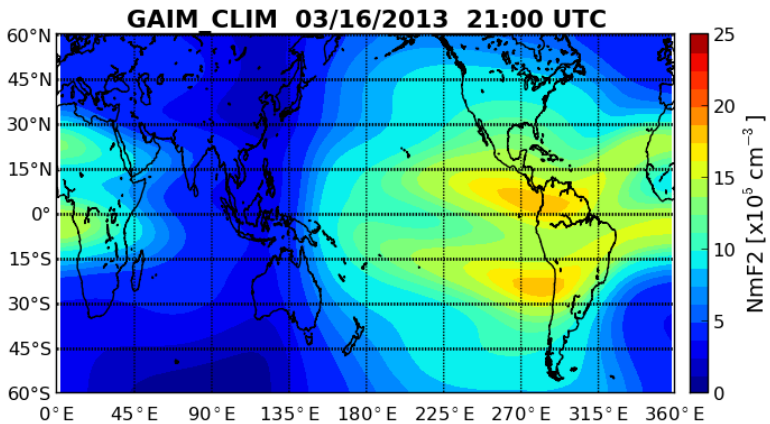
Ensemble Kalman – Kalman with a computational simplification

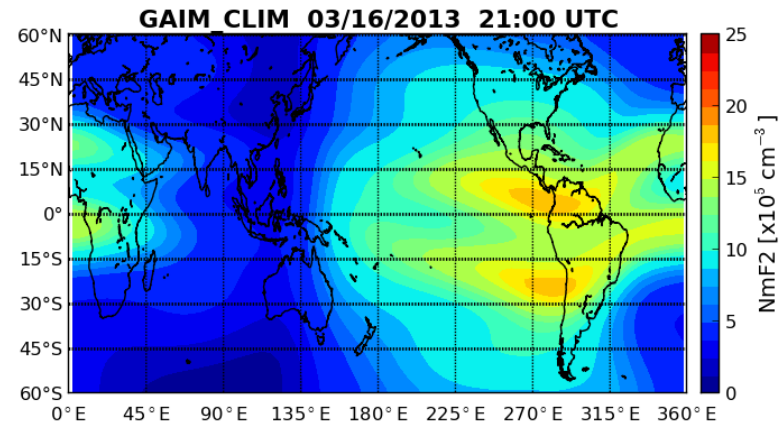
- Monte Carlo approach to estimating covariance update
- State often includes drivers

- Driver estimation depends on how the "state" is defined: electron or neutral densities alone, or drivers also
- Active research area

Comparison between First-Principle and Assimilative Modeling

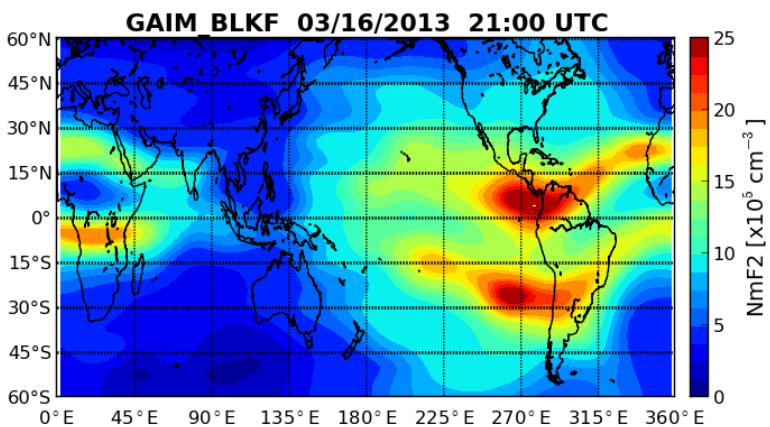
CLIM

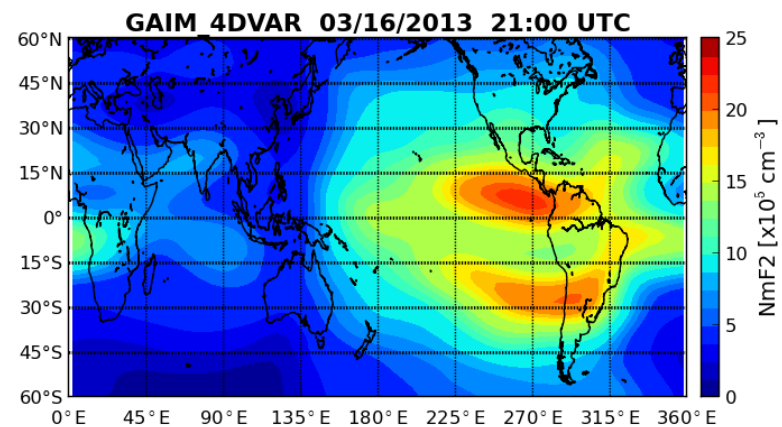




CLIM

BLKF

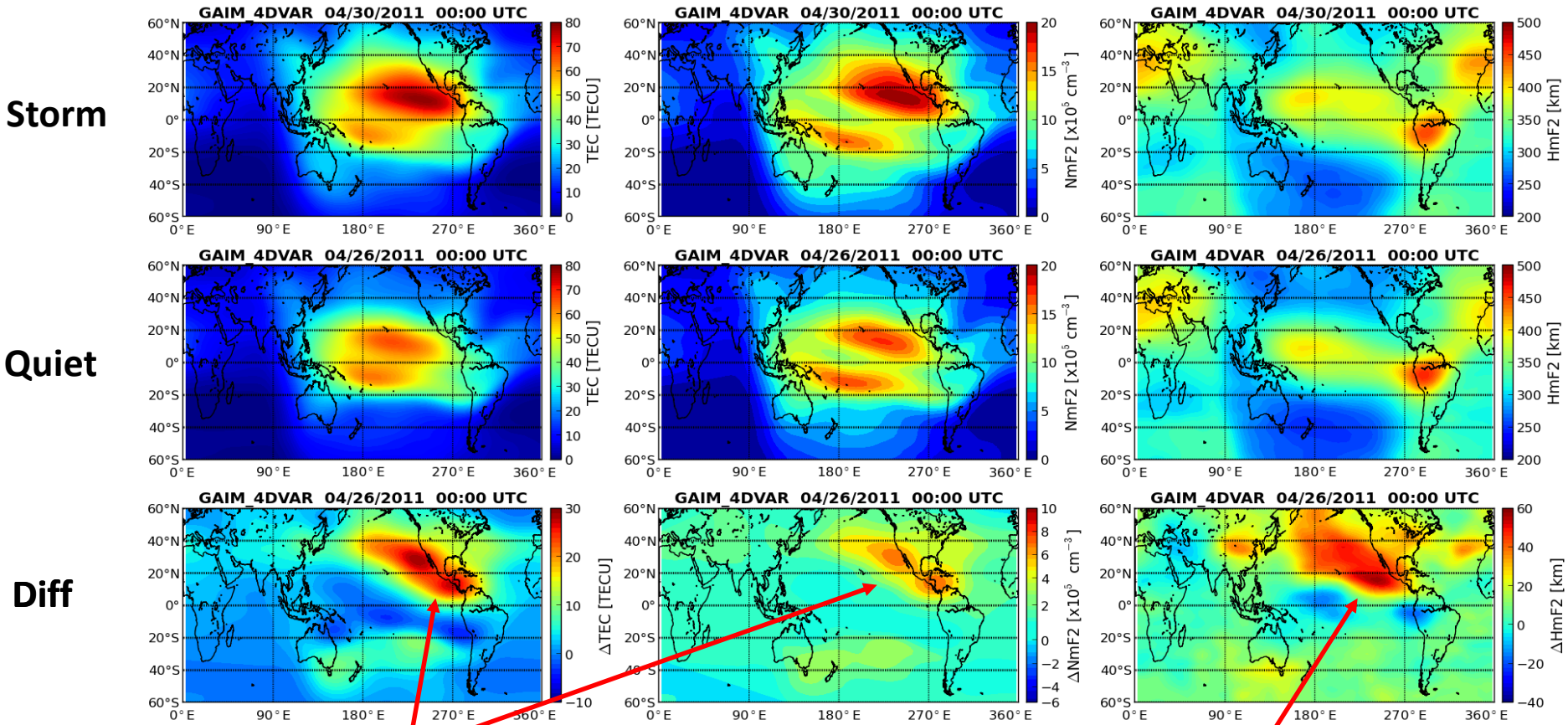




4DVAR

- **NmF2** under **quiet** conditions
- Significantly larger peak density in “weather” is shown in data assimilation results

Features of Ionospheric Storm: GAIM-4DVAR (2011-04-30, HSS)

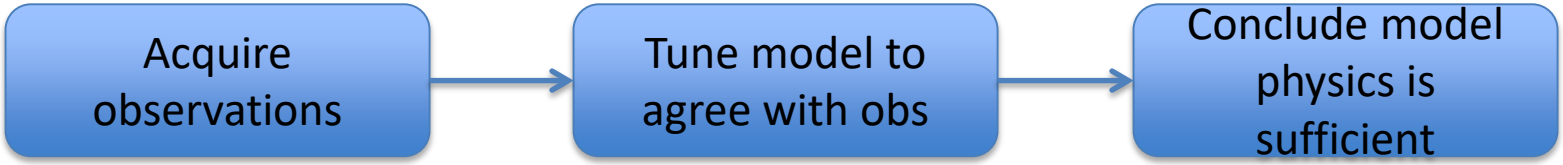


Hemispherically asymmetric TEC and peak density ↑ at mid lat in the afternoon during the storm main phase.

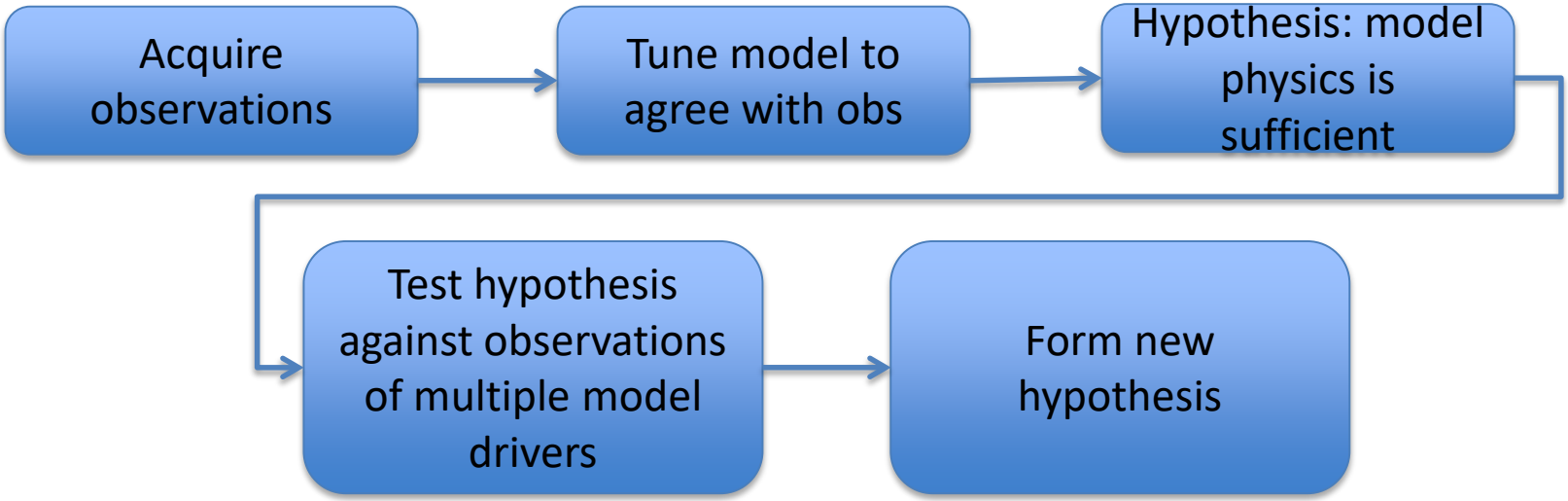
Height change (↑) in the region of density increase.

Resolving the Conundrum "Effectively Capture the New Science"

What we tend to do:



What we should do:

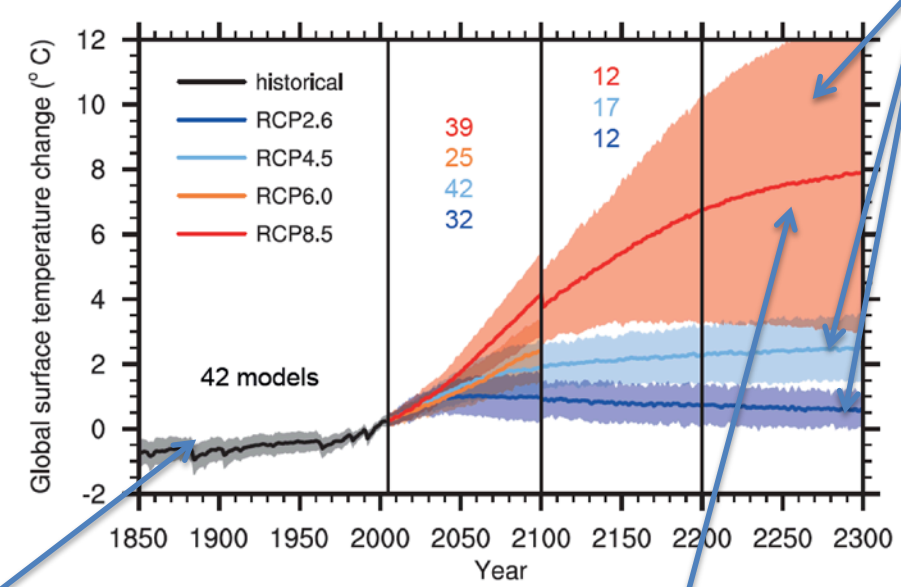


See Mannucci et al., "On scientific inference in geophysics and the use of numerical simulations for scientific investigations" *Earth and Space Science*, 2015

The Forecasts of 42 Climate Models for the IPCC

42 different climate models run for the IPCC assessment agree well with past data, but disagree about future climate change

Different scenarios for future CO_2 amount



Large inter-model spread

Model results for past data

Average of all the models

- **Interpretation: all GCMs are tuned to match data. Tuning creates the possibility that agreement with observations can be achieved with diverse physics (i.e. "confusion")**

Conclusion

- **The value of data assimiliation for science is in the estimates of state or drivers that are accurate enough and complete enough to be useful for hypothesis testing**
- **ICON and GOLD will provide the critical "driver" information for understanding cause and effect relationships within the I-T**
- **Use models to form hypotheses**
- **Move hypotheses forward by comparing multiple model outputs (drivers) to observations of those drivers**
- **ICON and GOLD will surprise us!**

BACKUP

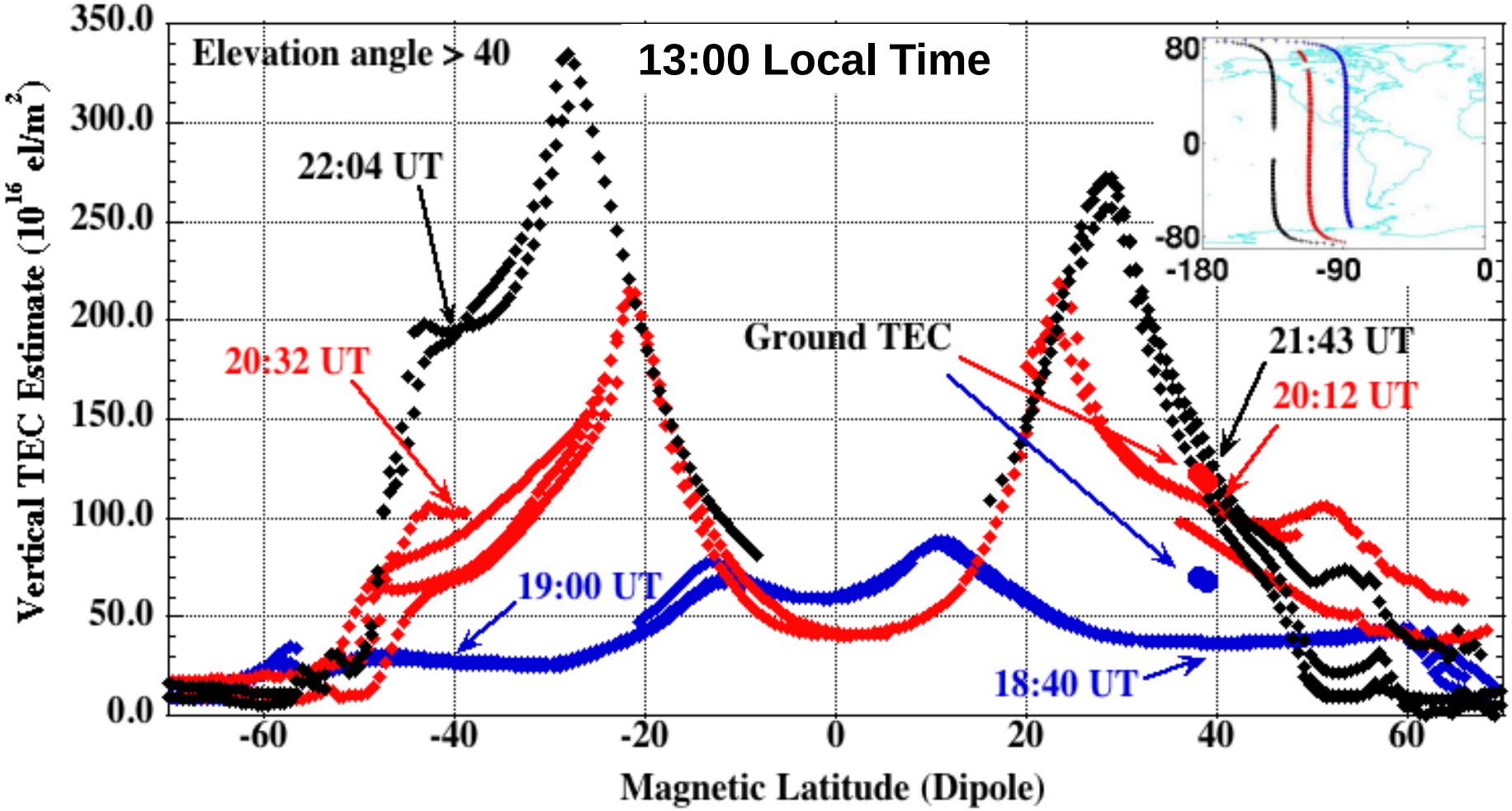
Why Forecasts Fail

- **Models with differing underlying physical parametrizations can be tuned to agree with past observations**
- **No such tuning is possible when the models are run for forecast purposes**
- **Standard "model validation" approaches leave us with limited confidence that we can use these models to accurately forecast**

- **A famous example of this problem in the climate domain is shown in the next slide**

Example of an Extreme Storm: October 30, 2003

Electron content above 400 km altitude



Mannucci et al., *GRL* 2005

"Global Ionospheric Storm"