On the Importance of Data Assimilation for the Thermosphere Ionosphere System

Mihail V. Codrescu
Space Weather Prediction Center/ NOAA, Boulder, Colorado, USA

Modern technological systems like GNSS positioning, HF communications, radar ranging, satellite communications, and power distribution are all affected by space weather and can become unreliable during disturbed conditions. For large space weather events the thermosphere and ionosphere, driven by strong external forcing and under the influence of feed-back loops, exhibit large deviations from climatology. Such extreme space weather conditions can have high impacts on systems and are notoriously difficult to reproduce by models. Successful specification and forecasting during such events requires physics based ionosphere thermosphere models and Data Assimilation (DA) schemes. DA in the thermosphere ionosphere system is required because of the impossibility to measure the forcing of the system with the necessary spatial and temporal resolution. The Coupled Thermosphere Ionosphere Plasmasphere and Electrodynamics (CTIPe) model is currently evaluated to determine the possibility of its use as the background model for the development of a modular data assimilation system. In parallel, the uncertainty associated with the external forcing of the system, i.e. high-latitude convection and precipitation patterns, solar UV and EUV fluxes, and the waves propagating from below, and the uncertainties associated with them are being evaluated to establish requirements for the DA scheme.

Outline

Why Data Assimilation?
What kind of Data Assimilation?

System behaviors

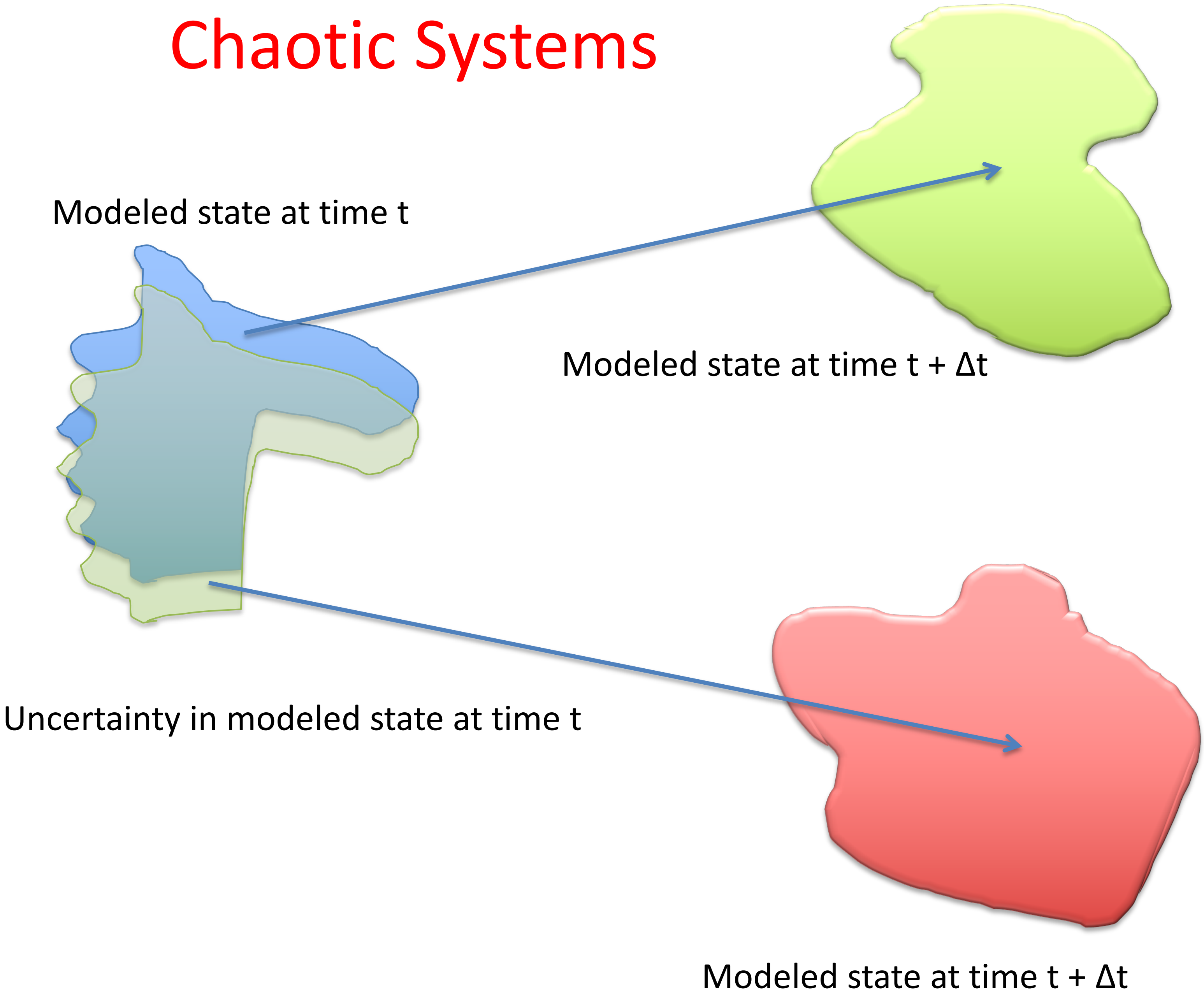
Models and data for Data Assimilation

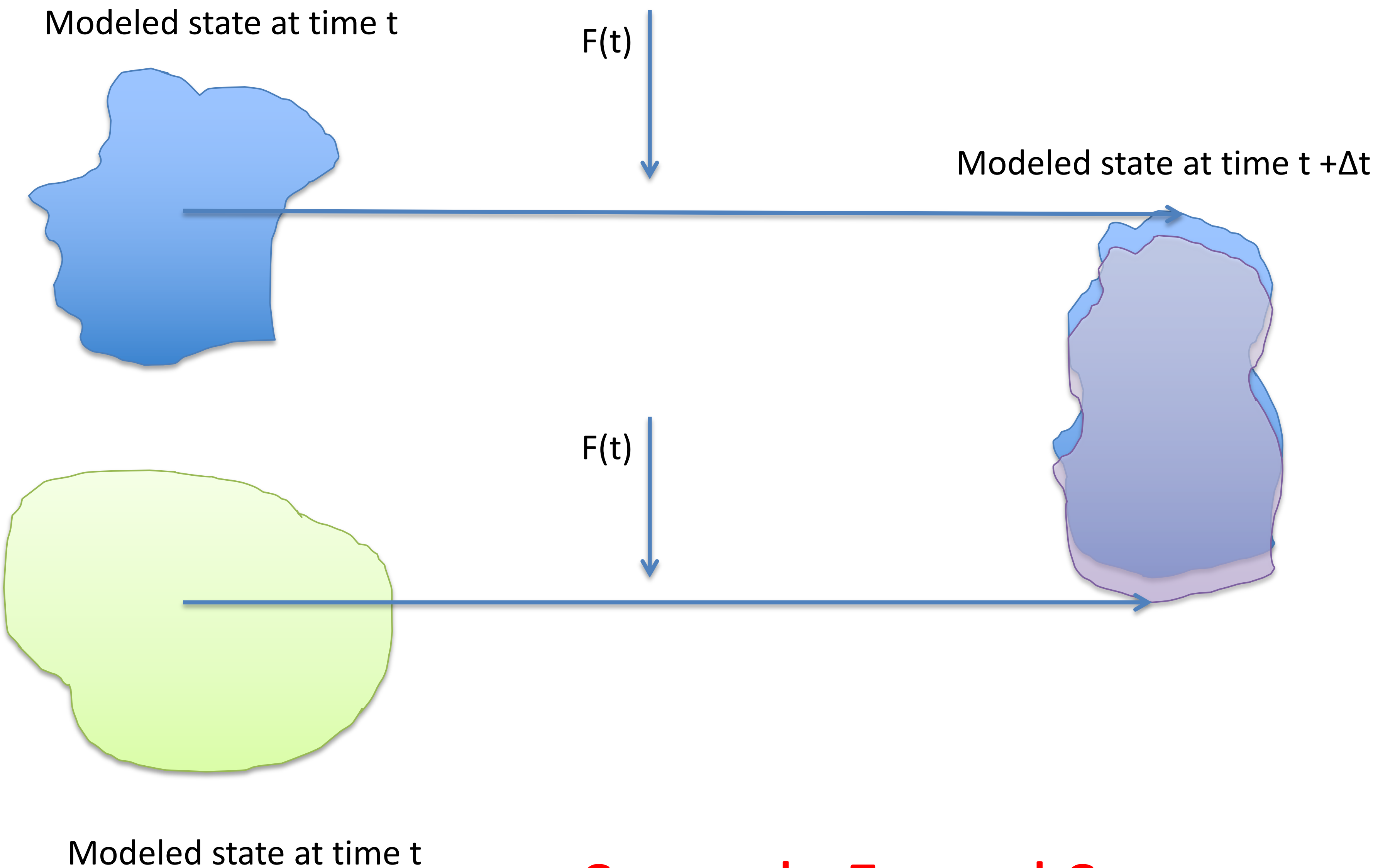
Data Assimilation for the thermosphere-ionosphere system

Data assimilation problems

The Future (COSMIC II, GOLD, ICON)

Chaotic Systems





Strongly Forced Systems

The Need for Data Assimilation

In addition to compensating for missing physics in the models, data assimilation schemes have system dependent requirements:

Chaotic systems need DA because the initial conditions cannot be determined with the necessary accuracy

Strongly forced systems need DA because the forcing cannot be accurately determined

The different needs result in different optimal assimilation schemes

More on Systems

No system is purely chaotic or purely forced

There is a mixture of behaviors in all systems

The dominant behavior dictates the optimal Data Assimilation scheme

Chaotic behavior dominates the global terrestrial weather system (the external forcing is diurnally reproducible)

However, on a regional scale (hurricane) external forcing can become important

Chaotic systems require better initial conditions

Strongly forced systems require better forcing as a function of time

A hybrid system requires both

Data Assimilation Issues/Questions

What Assimilation Scheme to use (enKF)?

What model (physics based, resolution, missing physics)?

What should be in the Kalman state?

How many measurements are necessary?

How good is the Error Covariance matrix (enough members in the enKF?)

State and Forcing not self-consistent; Consequences?

State elements not self-consistent; Consequences?

Running the DA scheme

Changing only the model results

Updating state not forcing

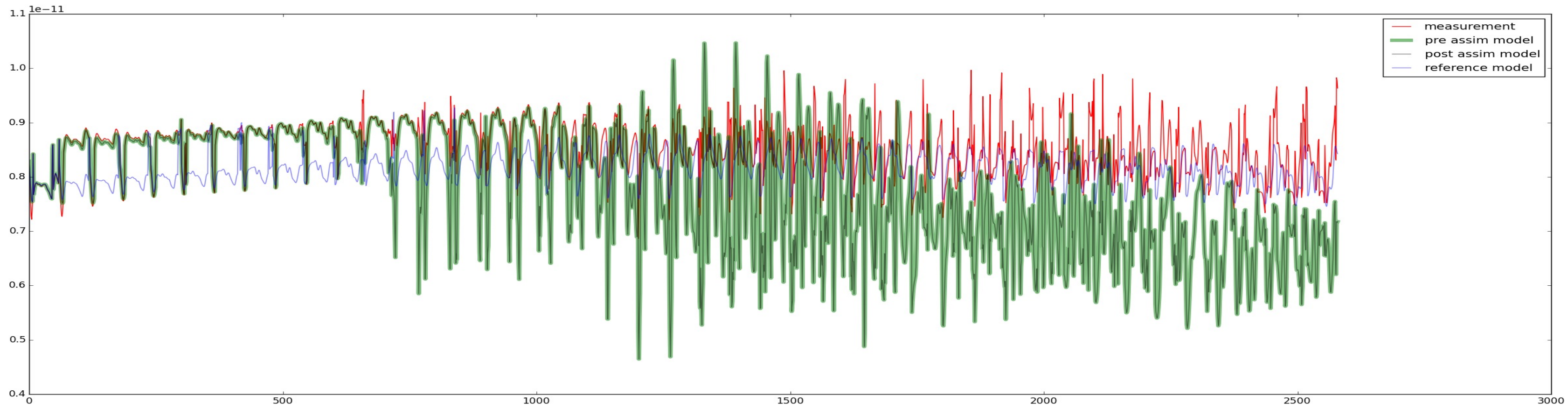
Updating forcing not state

Updating both forcing and state

Fitting structure in the model & data

Results

Experiment 41



Large Kalman state (500000 elements)
60 measurements every 10 minutes
20 members in the ensemble
$\Rightarrow$ Not enough data!!
$\Rightarrow$ Not enough members in the ensemble

The Future

- Data Assimilation
- GNSS ground and RO TEC
- Digisondes, dynasondes and radars
- COSMIC II, GOLD and ICON
- SWARM, ...

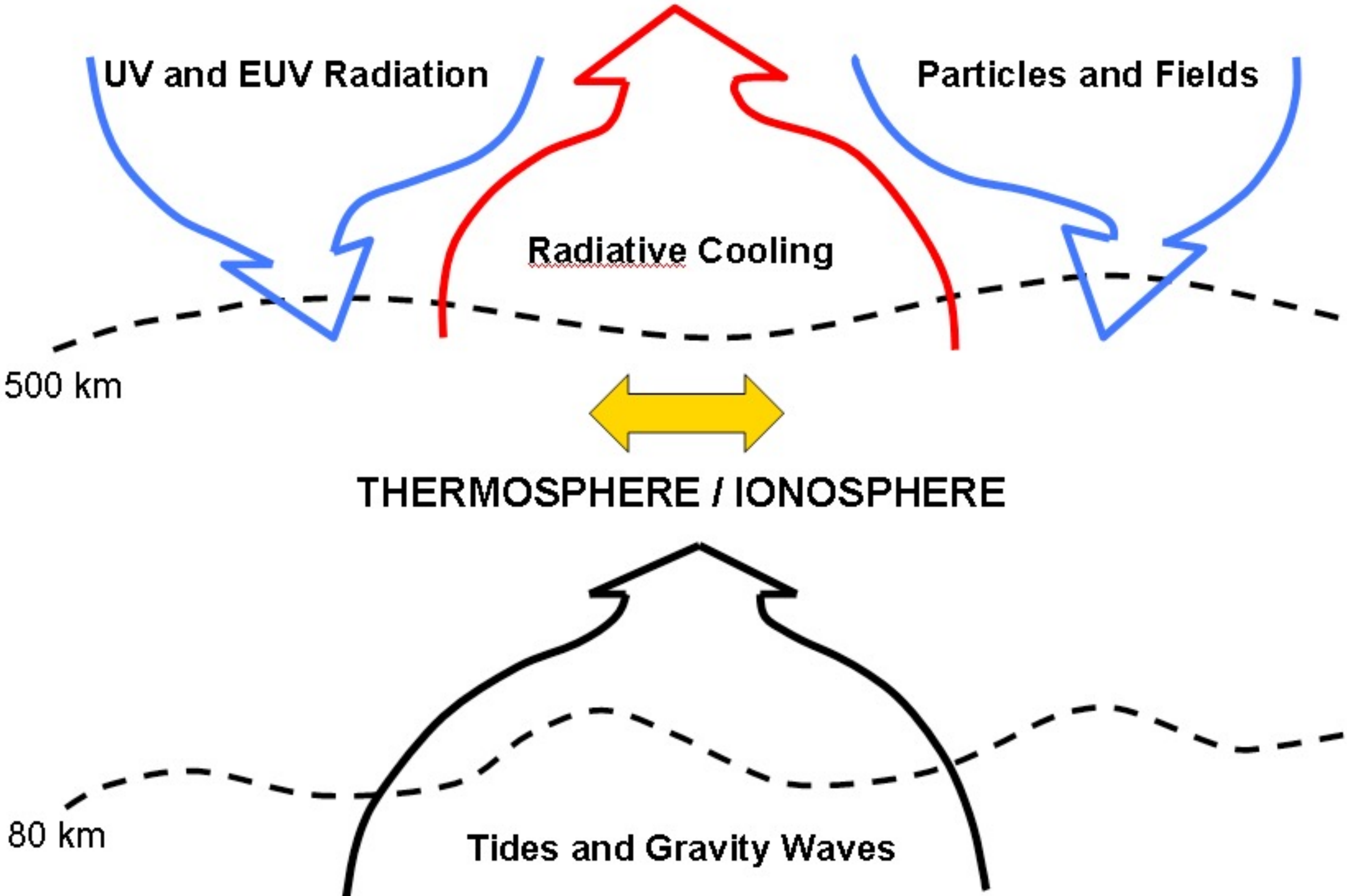
Not just for operations but for Research also!

Reanalysis!

Thank You

Questions?

THE SYSTEM



Strongly forced

On Models and Data for Data Assimilation

Good models require fewer measurements for specification and forecast

Good, abundant measurements can work with less sophisticated models

A perfect model requires no measurements

Perfect measurements require no model

This is true for both chaotic and strongly forced systems

The characteristics of the system, together with the quality of model and data dictate the choice of assimilation scheme!

Solar EUV Heating

We use a proxy (F10.7) for the solar flux.

F10.7 correlates well with the solar flux over long time scales but not so well over short time scales.

The use of a proxy combined with uncertainties in heating efficiencies combine to produce uncertainties of at least 50% in the thermosphere heating.

Heat transport complicates the picture even more.

Important for global dynamics and electrodynamics

Particles and Fields

E-field patterns are statistical and consequently smooth

Conductivity calculations are based on statistical precipitation patterns that are extrapolated from one orbit, in-situ, measurements or ACE data

=> Joule heating calculations based on statistical patterns have large uncertainties:

50% globally

factor of ten locally

One cannot remove the uncertainty entirely

Important for global circulation, neutral composition, and electrodynamical processes

Tides from below

Only propagating tides included in most models

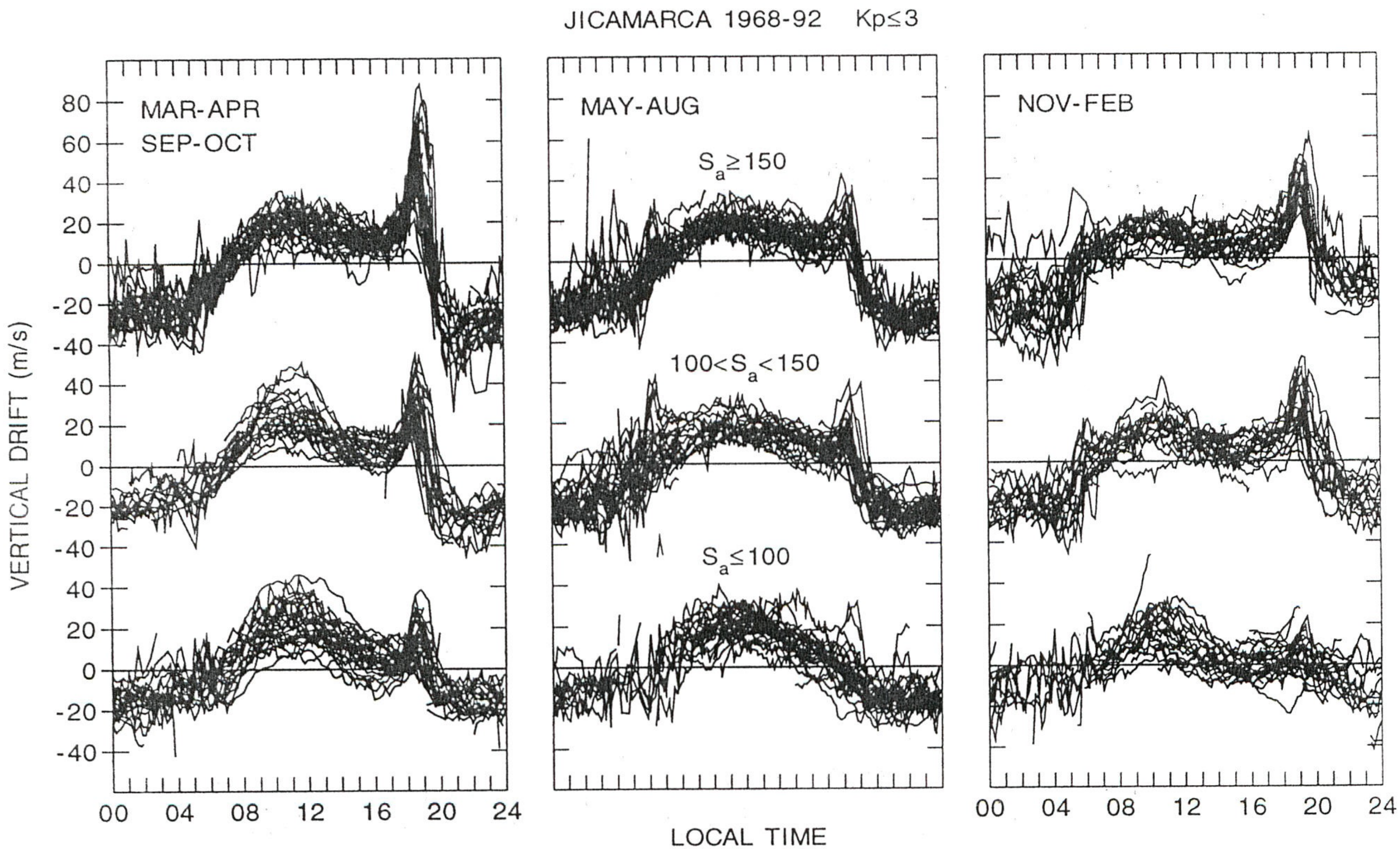
No gravity waves included

Amplitudes and phases are uncertain by at least 50%

Very important for the D- and E-regions (80-150 km)

Important for the F-region (300-500 km) variability

Large Variability



One or a few measurements are not enough to pinpoint the state of the system.
Need to interpret self-consistently many different measurements.

Fejer and Scherliess, 2001

A Modeling Problem

- Uncertainty in high-latitude convection and particle precipitation produce large uncertainty in Joule heating

- Joule Heating affects the global neutral temperature structure, circulation, and chemical composition (neutral system state)

- Neutral changes affect production, loss and transport of ionization and have dramatic effects on global electron density and TEC structures (ionosphere state)

- Global Joule heating cannot be satisfactory modeled at this time

  => We can model generic storms but not specific ones

Improve Joule Heating through data assimilation

Data Assimilation

Combine model and data based on their statistical errors

Challenges

Find the best model representation for state evolution in time

Obtain accurate statistical error estimations for model and data

Availability of quality data

- latency
- spatial coverage
- statistical errors

Data Assimilation Schemes

Ideally, any measurement should improve the estimate of all state elements

The covariance matrix can captures these relationships

In some systems the covariance matrix is stationary, i.e. it does not depend on the state of the system and can be determined a-priori

In strongly nonlinear, forced systems the covariance matrix is not stationary and must be determined at each assimilation step

In Space Weather the systems are large and non-linear and calculating the covariance matrix at each assimilation step is not practical.

Monte-Carlo methods have been successfully used in other fields to estimate the covariance

A DA Example: Global TEC from GPS

System: Ionosphere(-Thermosphere) is Nonlinear and Strongly Forced

Background Models: Empirical

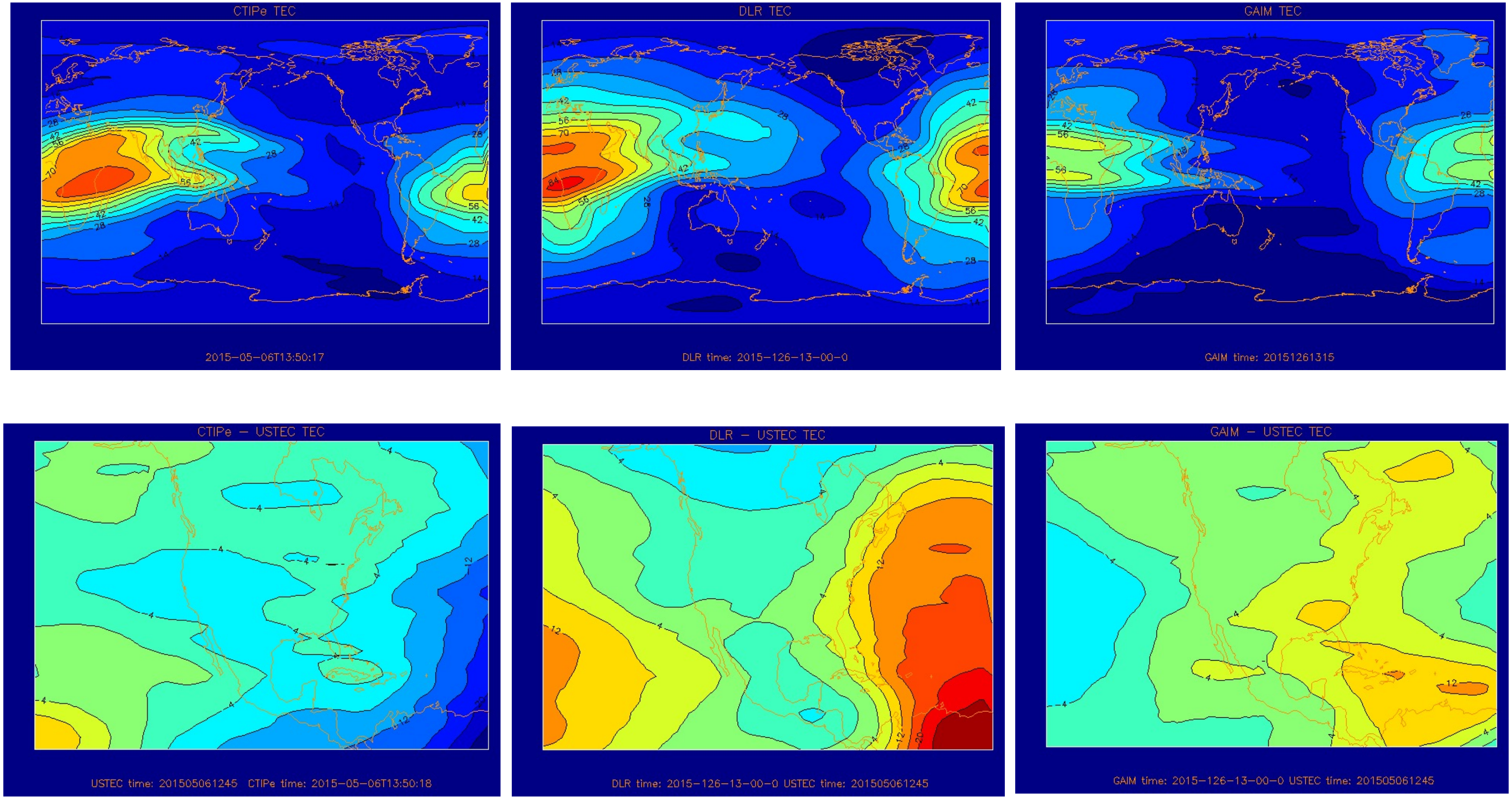
Data: Abundant in places, scarce in others

Assimilation scheme: Gauss Markov Kalman Filter
Localized covariance
No forcing estimates

DLR http://swaciweb.dlr.de/data-and-products/public/tec/tec-global/?L=1
GAIM Schunk et al. 2004 doi:10.1029/2002RS002794

Physics based model without DA
CTIPe http://helios.swpc.noaa.gov/ctipe/index.html

TEC Comparison

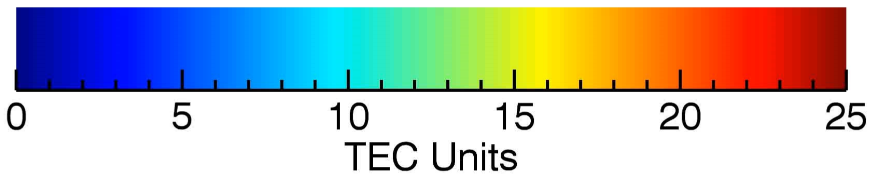


CTIPe DLR GAIM

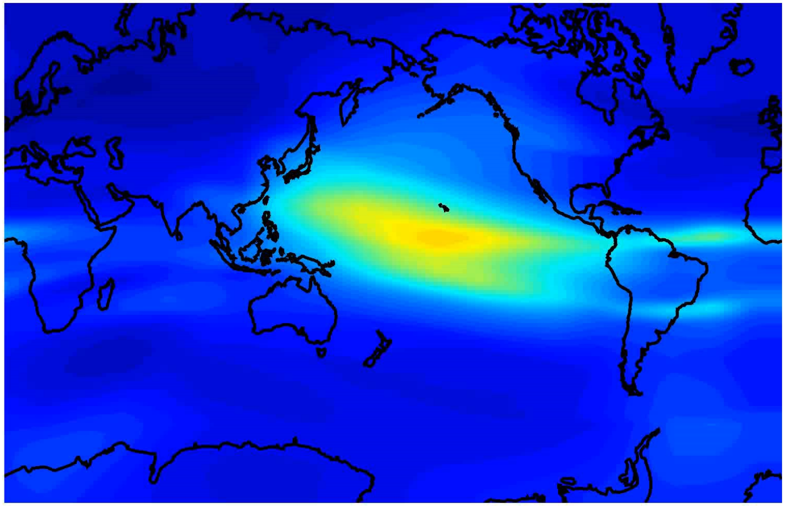
http://helios.swpc.noaa.gov/ctipe/teccomp.html

Results from Bill Frey

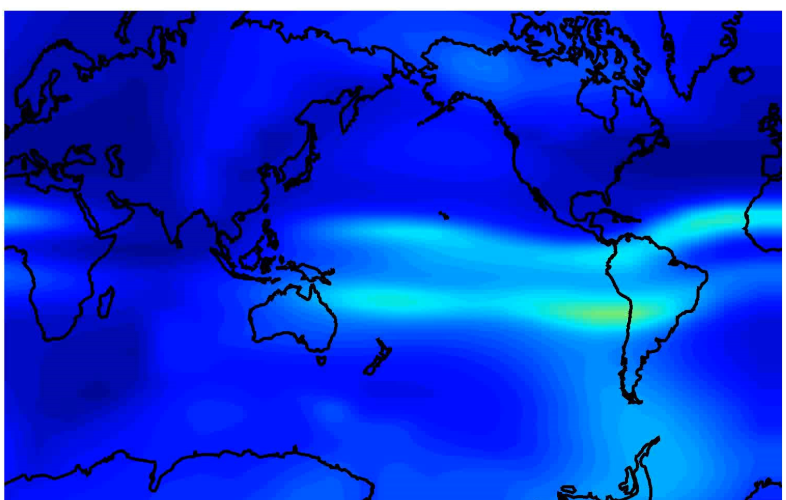
TEC Standard Deviation at 0Z
61 Model Runs Between
8-28-2014 and 11-17-2014



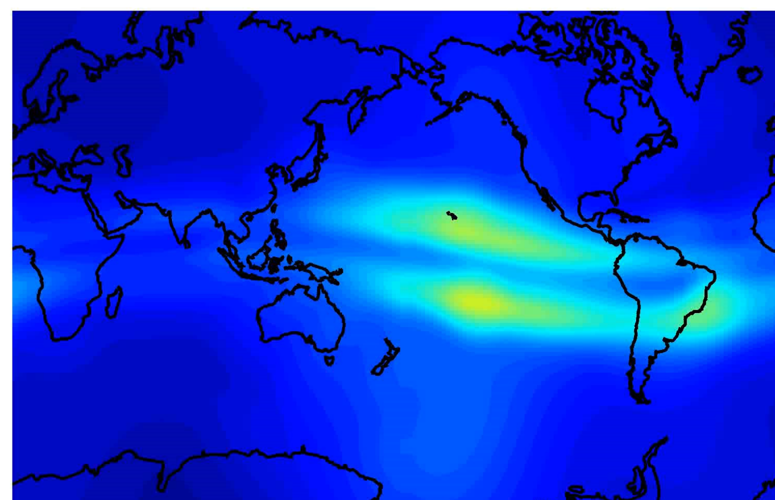
CTIPe



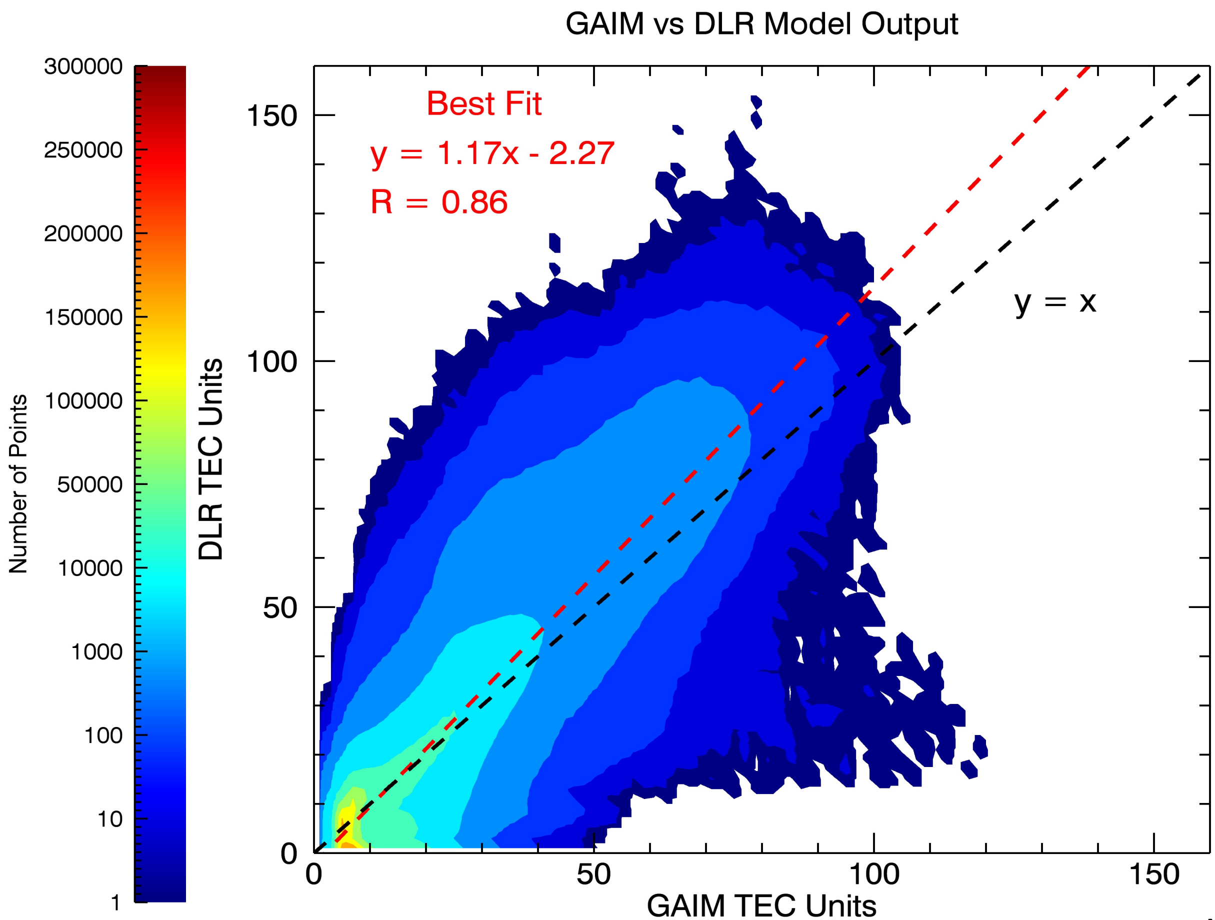
GAIM



DLR



Data Assimilation today



Results from Bill Frey

To do better we need to estimate the forcing based on all available measurements

An Ensemble Kalman Filter for Neutral Composition

System: Thermosphere-(Ionosphere) is Nonlinear and Strongly Forced

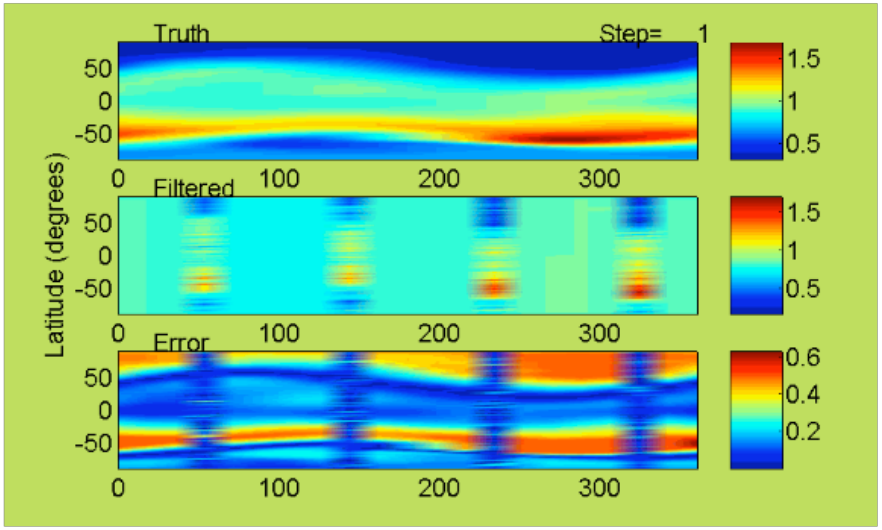
Truth state generated by a more sophisticated numerical model

Background Model: Physics based

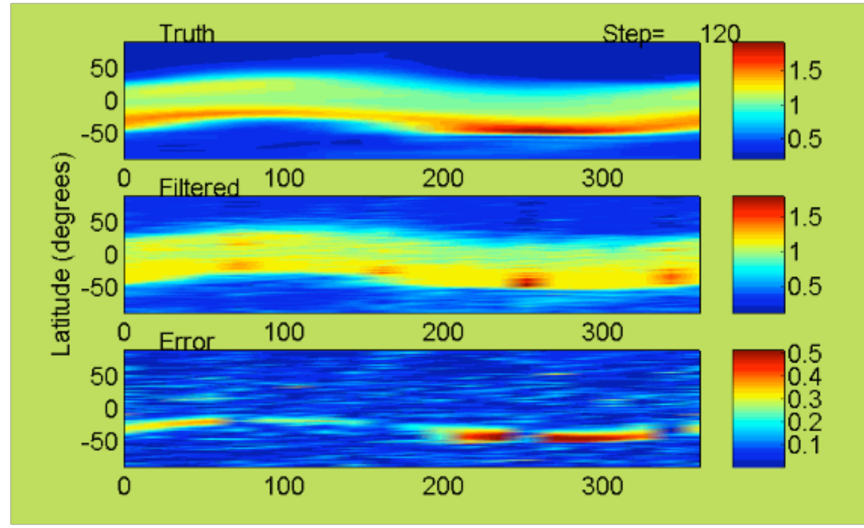
Data: Measurements from two sun-synchronous satellites (20%)
    10% error on the measurements

Assimilation scheme: Ensemble Kalman Filter
    No forcing included in the state
    Forcing inferred from the assimilated state

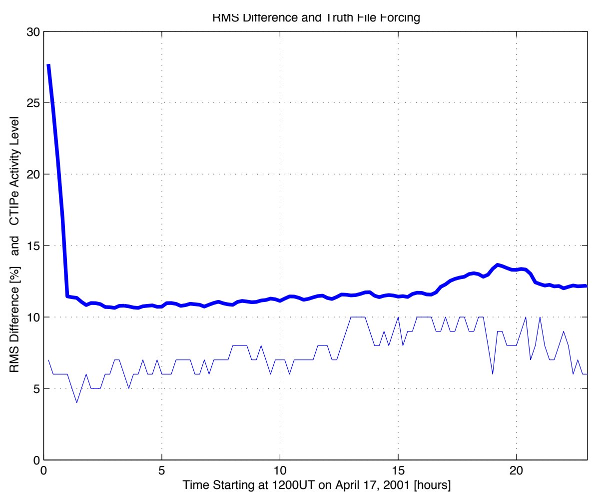
Ensemble Kalman Filter example for O/N2

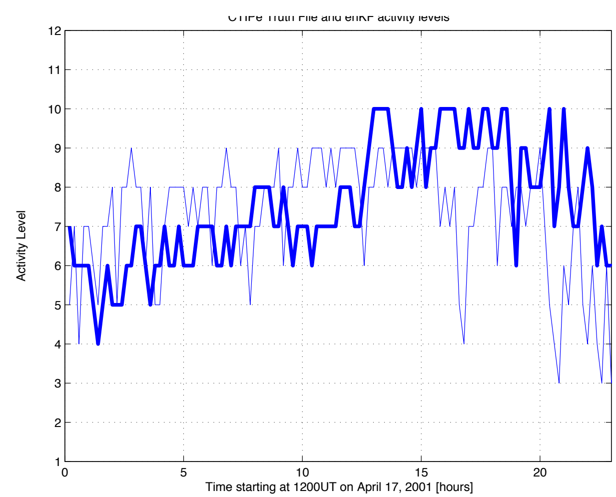


State after first assimilation step



State after 120 assimilation steps





Codrescu et al. 2004; SPACE WEATHER, doi:10.1029/2004SW000088

Data Assimilation and the future

Ensemble Kalman filters offer the best hope for data assimilation for large strongly forced dynamical systems with non-stationary covariance

One can do a better modeling job without perfect knowledge of the forcing

Data assimilation in multiple model fields can result in much more accurate forcing patterns than direct measurements can offer

One could infer forcing patterns using EOFs. Successive orders of EOFs can be determined from different assimilated fields using better data and physical understanding.

New missions SWARM, COSMIC II, GOLD, ICON

In nonlinear strongly forced systems Data Assimilation should be primarily about correcting the forcing and not about correcting the model results

Challenges

Run GCMs in semi-operational environment + data assimilation

Specify input variability and uncertainty

Coupling with the lower atmosphere

UV and EUV specification and effects

Convection E-fields and particle precipitation during extreme events

High-latitude forcing specification now from ACE

- Solar wind structure influence on high-latitude forcing
- Solar wind-magnetosphere-ionosphere coupling + data assimilation

See you at the poster!

Developing An Ensemble Kalman Filter for Data Assimilation in CTIPE

Stefan Codrescu
CU/CIRES Boulder
NOAA/NCEI Boulder
stefan.codrescu@noaa.gov

Mihail Codrescu
NOAA/SWPC Boulder

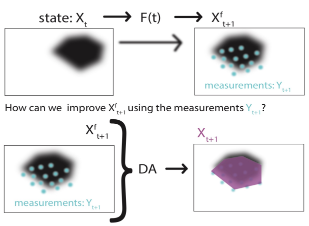
Mariangel Fedrizzi
CU/CIRES Boulder
SWPC/NOAA Boulder

Abstract

We are developing an ensemble kalman filter to assimilate data into the Coupled Thermosphere Ionosphere Plasmasphere and Electrodynamics (CTIPE) model. The Ensemble Kalman Filter (EnKF) approach is useful for approximating the non stationary covariance of the state, especially in high dimensional state spaces like that of CTIPE. Challenges include creating a representative distribution of the state uncertainty in the ensemble, limiting the computational complexity of the scheme, and managing the effects of measurement biases. Finally, we present preliminary results in assimilating simulated measurements.

Introduction

Data assimilation is a process in which measurements are incorporated into the state of a model to improve specification and forecasting. Our scheme implements an Ensemble Kalman Filter (EnKF).



Conceptually, a data assimilation (DA) scheme combines the model forecast X^f_{t+1} with measurements Y_{t+1} to estimate the state X_{t+1} with a smaller uncertainty.

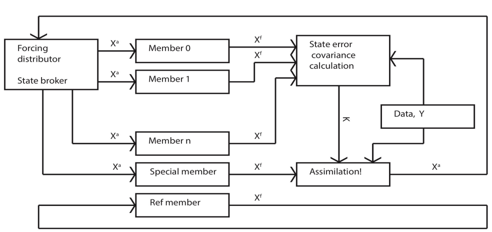
Here, we represent uncertainty roughly with blur.

Many techniques for data assimilation exist. Some techniques such as the extended Kalman Filter or 4DVAR linearize the dynamics which may be more epensive and underestimate the correlations compared to the Ensemble Kalman Filter.

Architecture

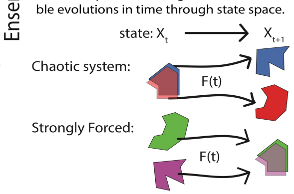
CTIPE is a non-linear, coupled thermosphere-ionosphere-plasmasphere physically based numerical code that includes a self-consistent electrodynamics scheme for the computation of dynamo electric fields. The model consists of four distinct components which run concurrently and are fully coupled.

The assimilation scheme is written from scratch in C++11. The program manages the ensemble, gathering measurements, and assimilation calculations. The Eigen library is used for matrix computations. Additionally, the GNU Multi-precision Library is needed.



Ensemble

Performance of the filter depends critically on the covariance estimate obtained from the ensemble. The goal is to distribute the ensemble over the state space according to the error of the current state and the possible evolutions in time through state space.



Note the difference in behavior between a strongly forced system and a chaotic system due to some forcing F(t).

Chaotic systems are characterized by a strong sensitivity to initial condition, thus it makes sense to perturb the initial condition across the ensemble to distribute members over the state space.

CTIPE in contrast, has less dependency on the initial condition and more dependency on the forcing applied. We call this a strongly forced system. To distribute the ensemble over the state space, we perturb the forcing.

The forcing in the ensemble is perturbed according to independent draws from normal distributions described to the right.

These are values that seemed to have acceptable performance. Investigating optimal distribution parameters is future work.

$$B_n \sim \mathcal{N}(0, 10)$$
$$B_\theta \sim \mathcal{N}(0, 15)$$
$$V_{sw} \sim \mathcal{N}(0, 50)$$
$$\rho_{sw} \sim \mathcal{N}(0, 5)$$
$$F_{10.7} \sim \mathcal{N}(0, 50)$$

Challenges

- The filter can cause the state to become inconsistent and cause model instance to crash.
- Density data from CHAMP and GRACE are more sparse than our simulated data and appears to have bias with respect to each other. Assimilating real density makes the filter diverge.
- How many members required to accurately approximate the covariance?
- How do we scale to more measurements? The most expensive calculation is the pseudo-inverse required for the Kalman gain. Can we avoid this by assimilating one measurement at a time?

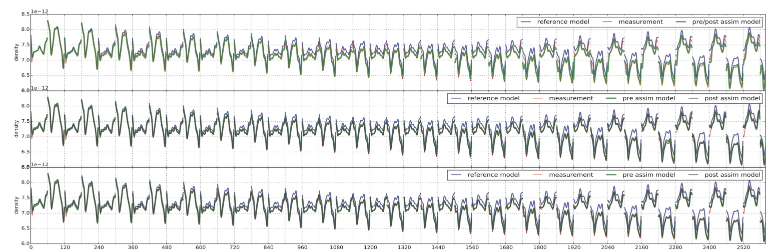
Results

We show results from assimilating simulated neutral density measurements from March 20, 2007 (quiet day near solar minimum). Neutral density measurements are simulated by sampling a run of CTIPE forced with an artificial F10.7 value. Although the test is weak, it indicates that we are on the right track.

The combination of Panel 1 highlight the strongly forced nature of the system, the filter is able to reproduce the simulated measurements only by modifying the forcing Panel 2 illustrates that fixing the state under a certain forcing will maintain the trajectory. Panel 3 shows the behvavior when both forcing and state are modified by the assimilation.

Three modes

- Panel 1: only allowing the filter to modify the forcing.
- Panel 2: the filter is only allowed to modify the Kalman state elements.
- Panel 3: both state and forcing are modified by the filter.



Codrescu, M. V., T. J. Fuller-Rowell, and C. F. Minter. "An ensemble-type Kalman filter for neutral thermospheric composition during geomagnetic storms." Space Weather 2.11 (2004).
Evensen, G. (1994), Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99, 10,143--10,162.
Gillijns, Steven, et al. "What is the ensemble Kalman filter and how well does it work?." American Control Conference, 2006. IEEE, 2006.
Houtekamer, P. L., and H. L. Mitchell (1998), Data assimilation using an ensemble Kalman filter technique, Mon. Weather Rev., 126, 796-- 811.
Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, Trans. ASME J. Basic Eng., 82, 35--45.