Real-time ionospheric data processing and modeling: an empirical orthogonal function approach

Shunrong Zhang

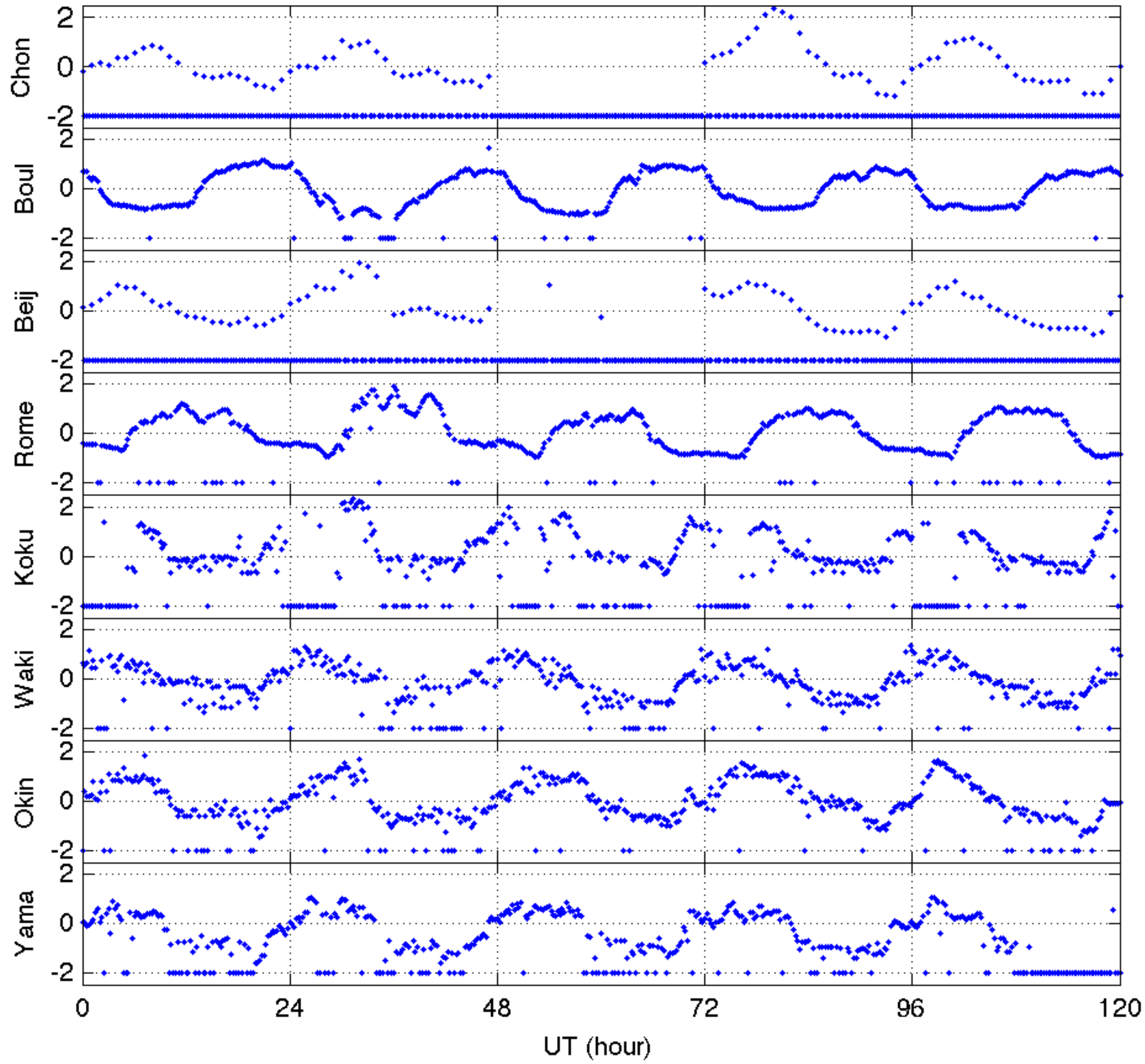
MIT Haystack Observatory

with contributions from A. Coster and Z. Chen

Outline

- Motivation
- EOF concept
- EOF-based North America GPS-TEC modeling
- EOF-based Near Real-time data processing
  - DIEOF procedure
  - Case studies
- Summary

Data gaps and non-uniform data resolution



Test data of normalized foF2 from WDC during October 2013.

Empirical Orthogonal Function (EOF)

$$T(t, n, \phi, \lambda) = \sum_{i=1}^{I} b_i(\phi, \lambda) P_i(t, n)$$

- A decomposition of a data set in terms of orthogonal basis functions which are determined from the data.
- The ith basis function is chosen to be orthogonal to the basis functions from the first through i − 1, and to minimize the residual variance. (The basis functions are chosen to be different from each other, and to account for as much variance as possible.)

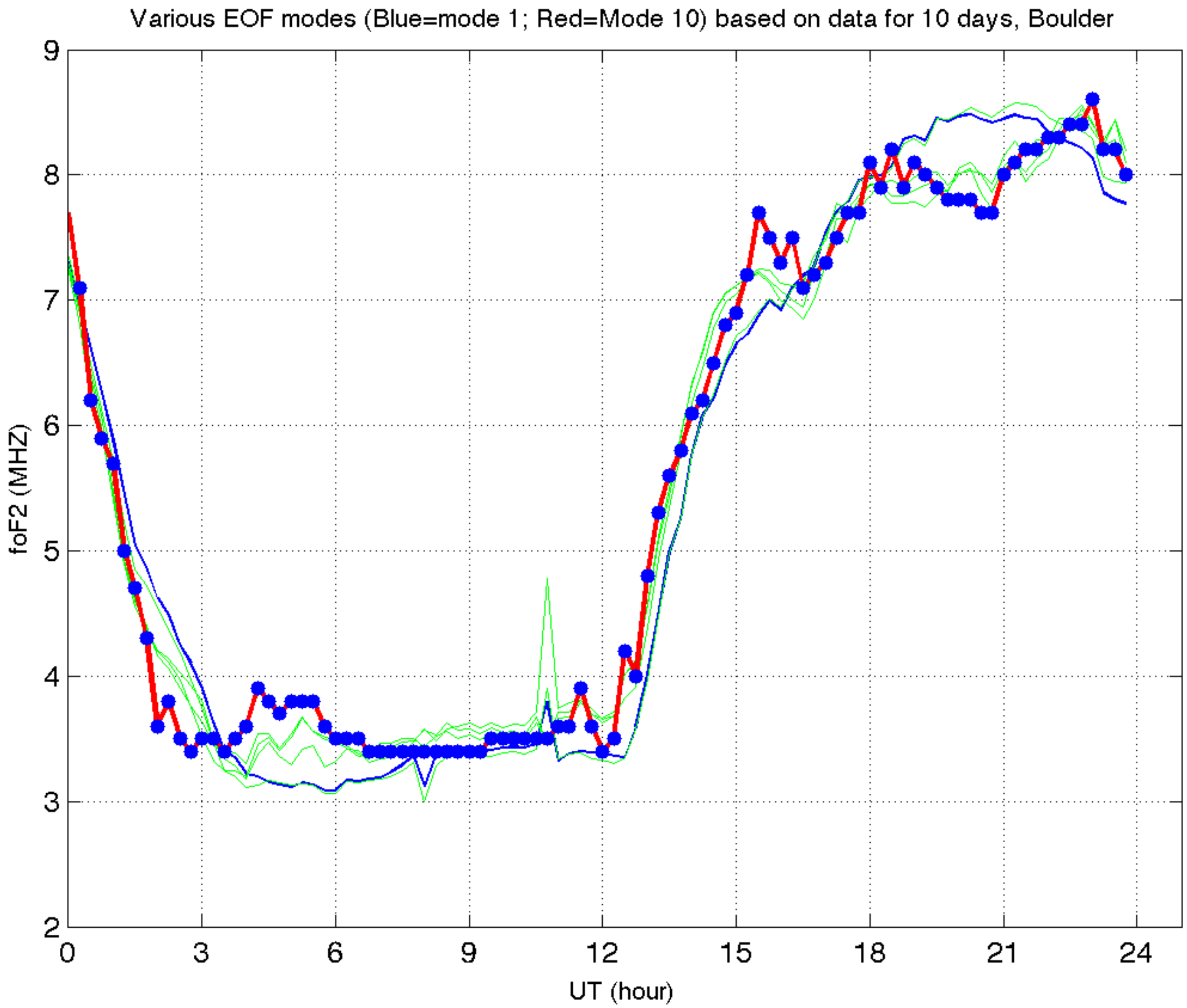
Principal Component Analysis (PCA)

$$\mathbf{C}\boldsymbol{b}_i = \omega_i \boldsymbol{b}_i$$
$$\mathbf{C} = \boldsymbol{cov}\,[\,\boldsymbol{T}\,(\boldsymbol{t}, \boldsymbol{n}, \boldsymbol{\varphi}, \boldsymbol{\lambda})\,]$$

Singular Value Decomposition (SVD)

$$T = U\,\Sigma\,W^T$$

EOF modes



A test dataset for 10 days over Boulder, CO was processed with EOF decomposition.

Faithfully reproduce the original data with a small number of modes

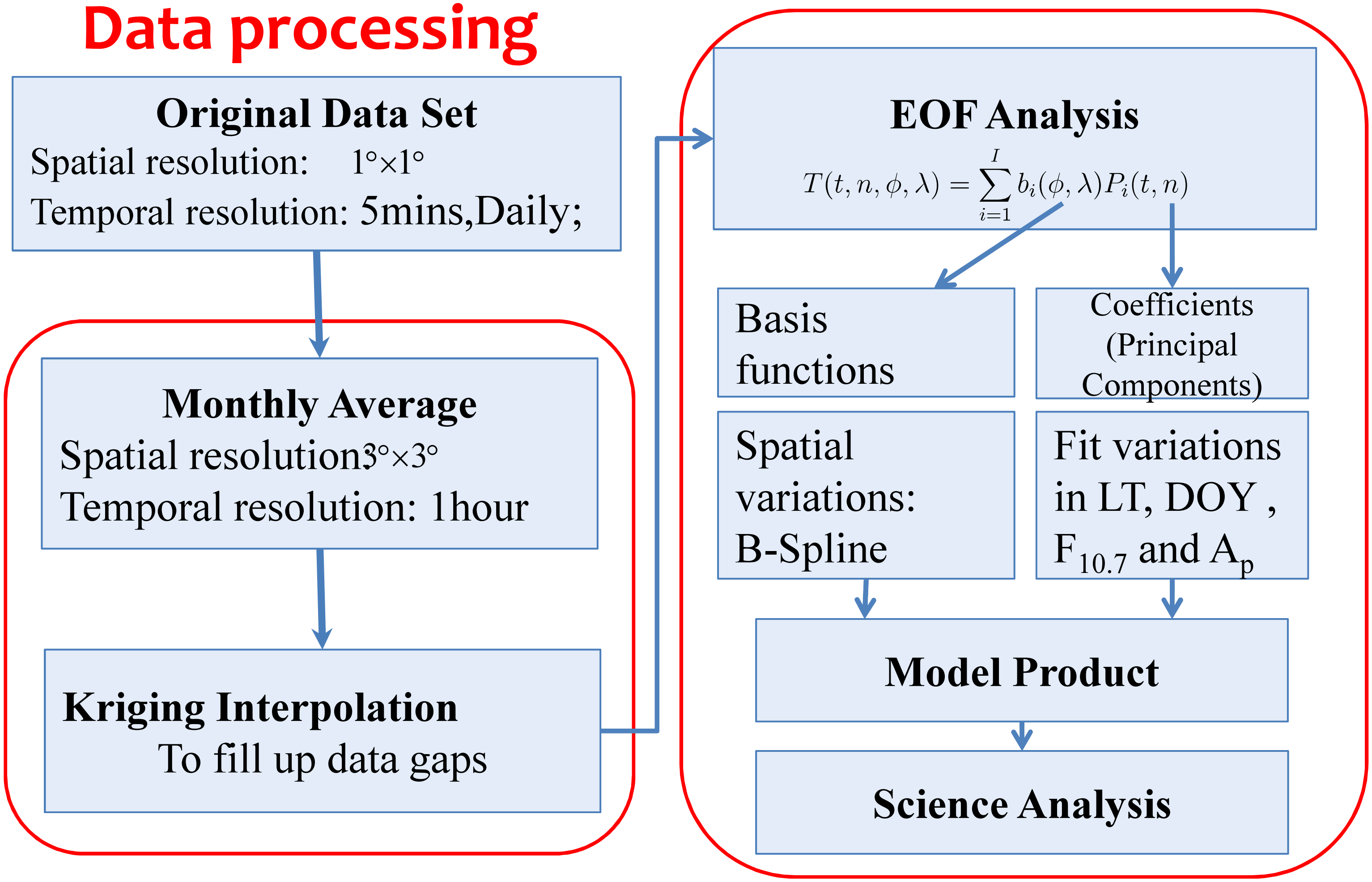
Outline

- Motivation
- EOF concept
- EOF-based North America GPS-TEC modeling
- EOF-based Near Real-time data processing
  - DIEOF procedure
  - Case studies
- Summary

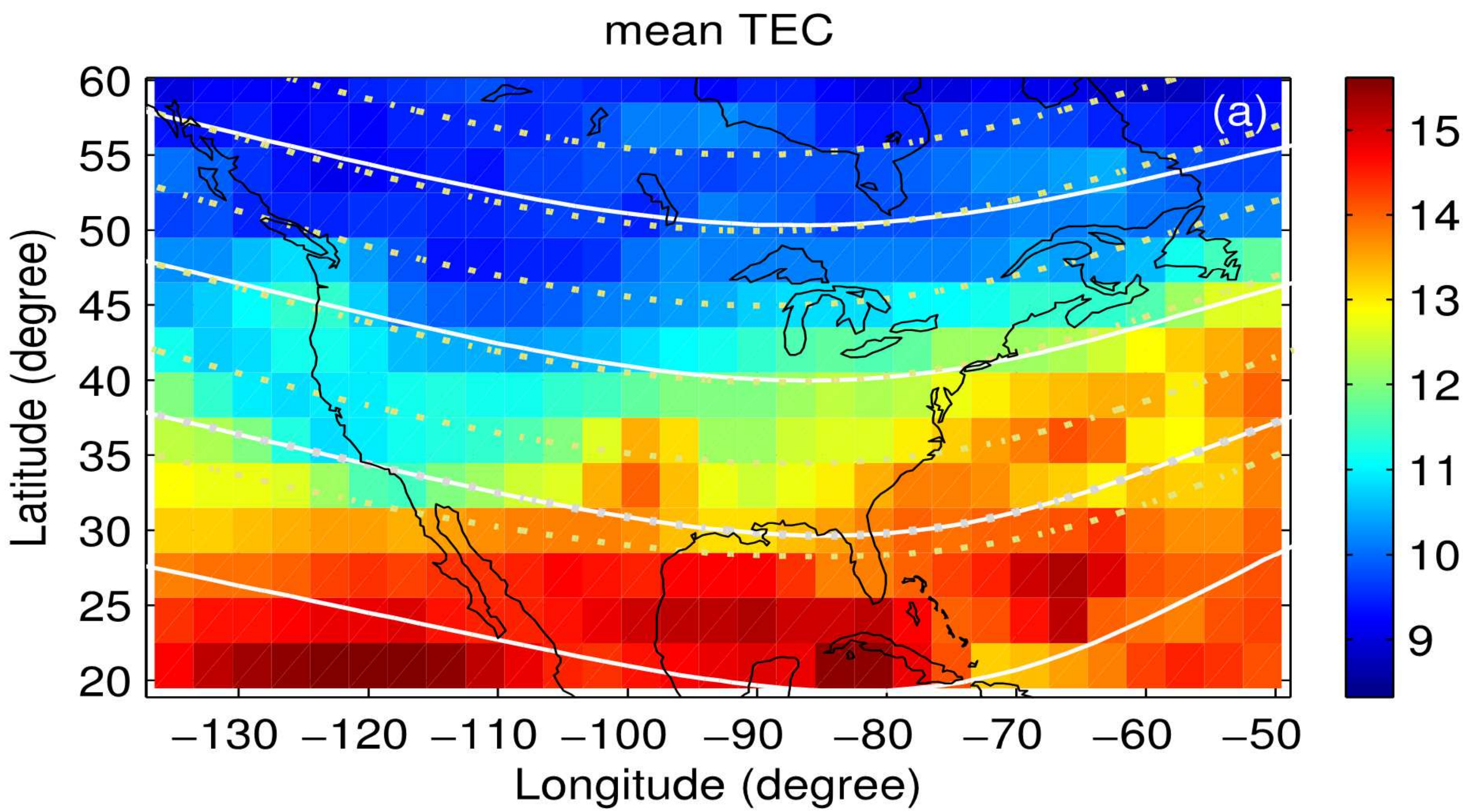
North America GPS TEC Model

based on GPS TEC data for a full solar cycle 2000 – 2012

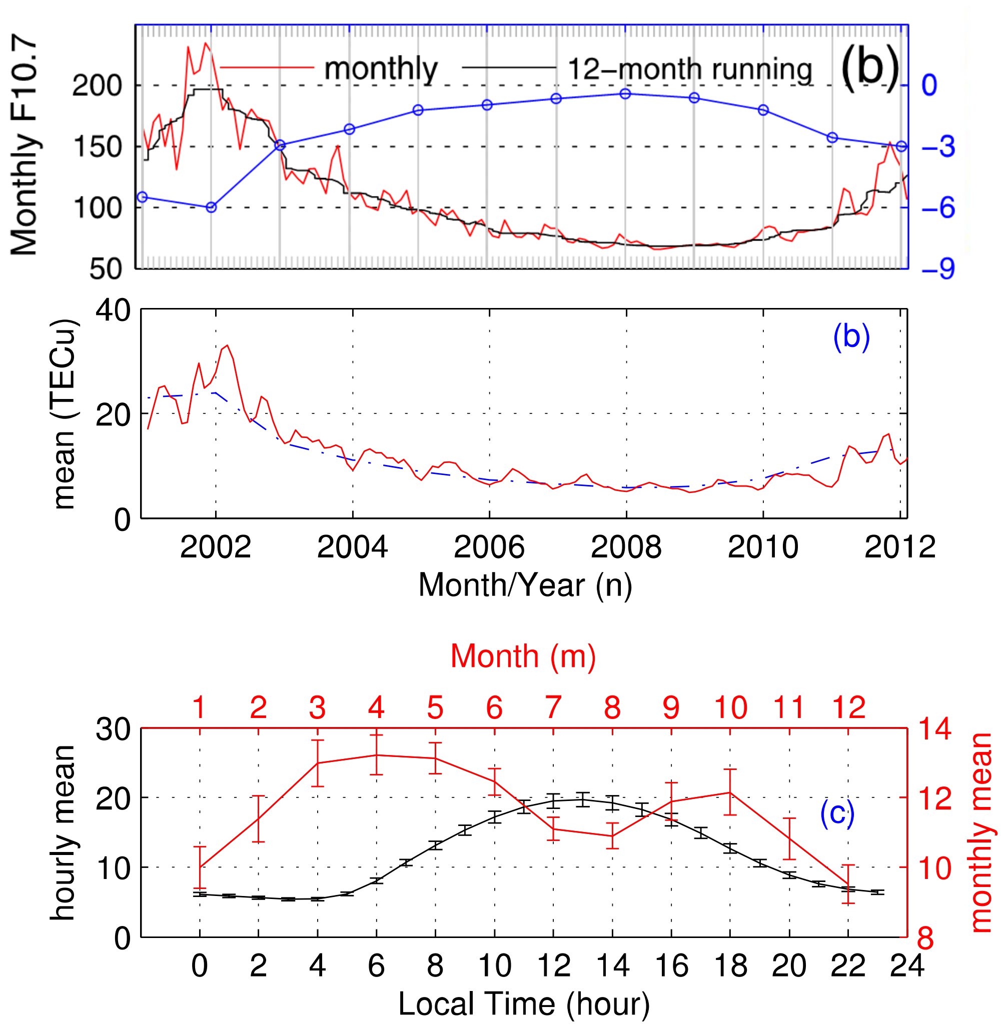
Data processing



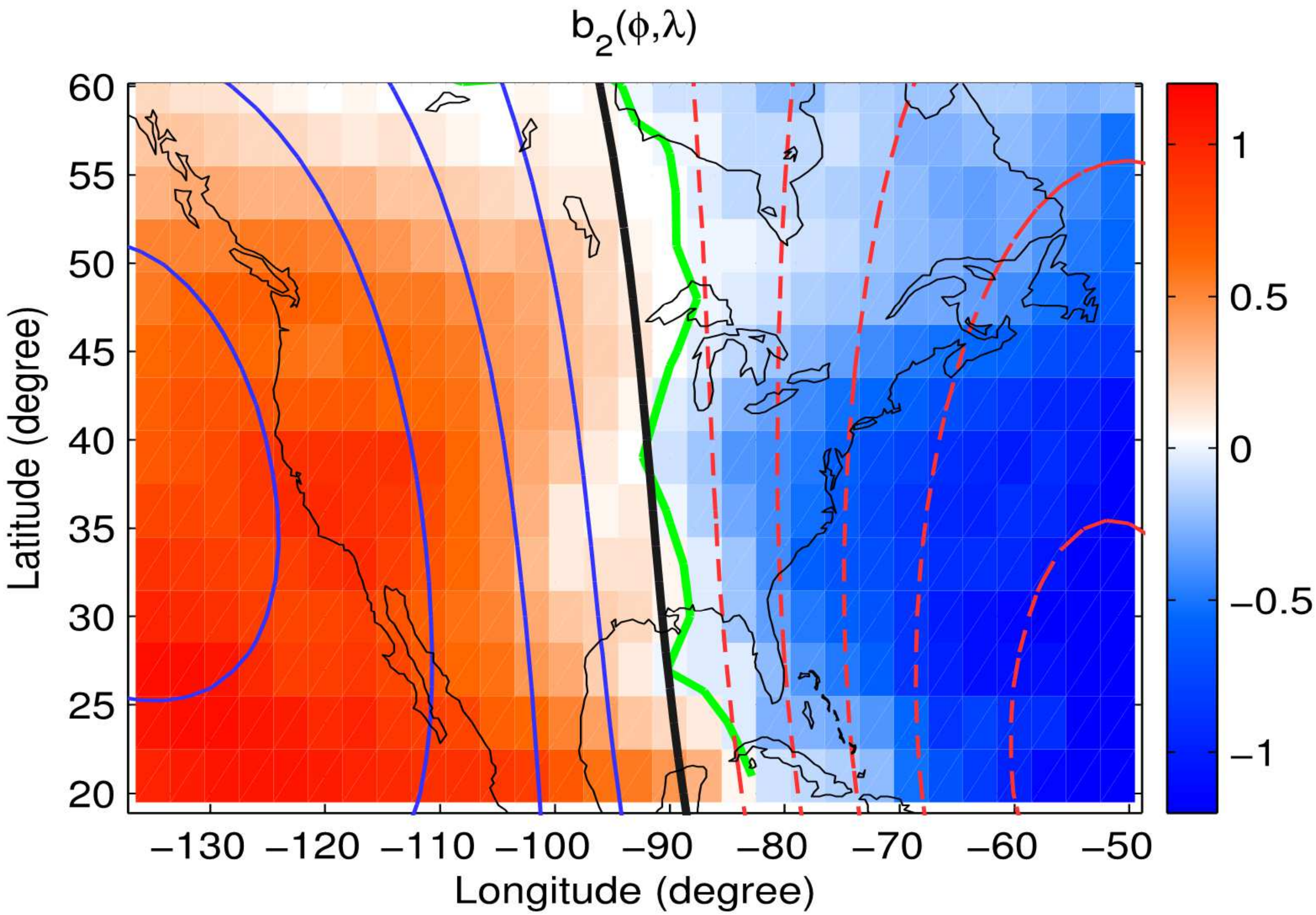
Mean TEC (spatial variations from EOF basis function mode 1)



Mean temporal variation (EOF mode 1)

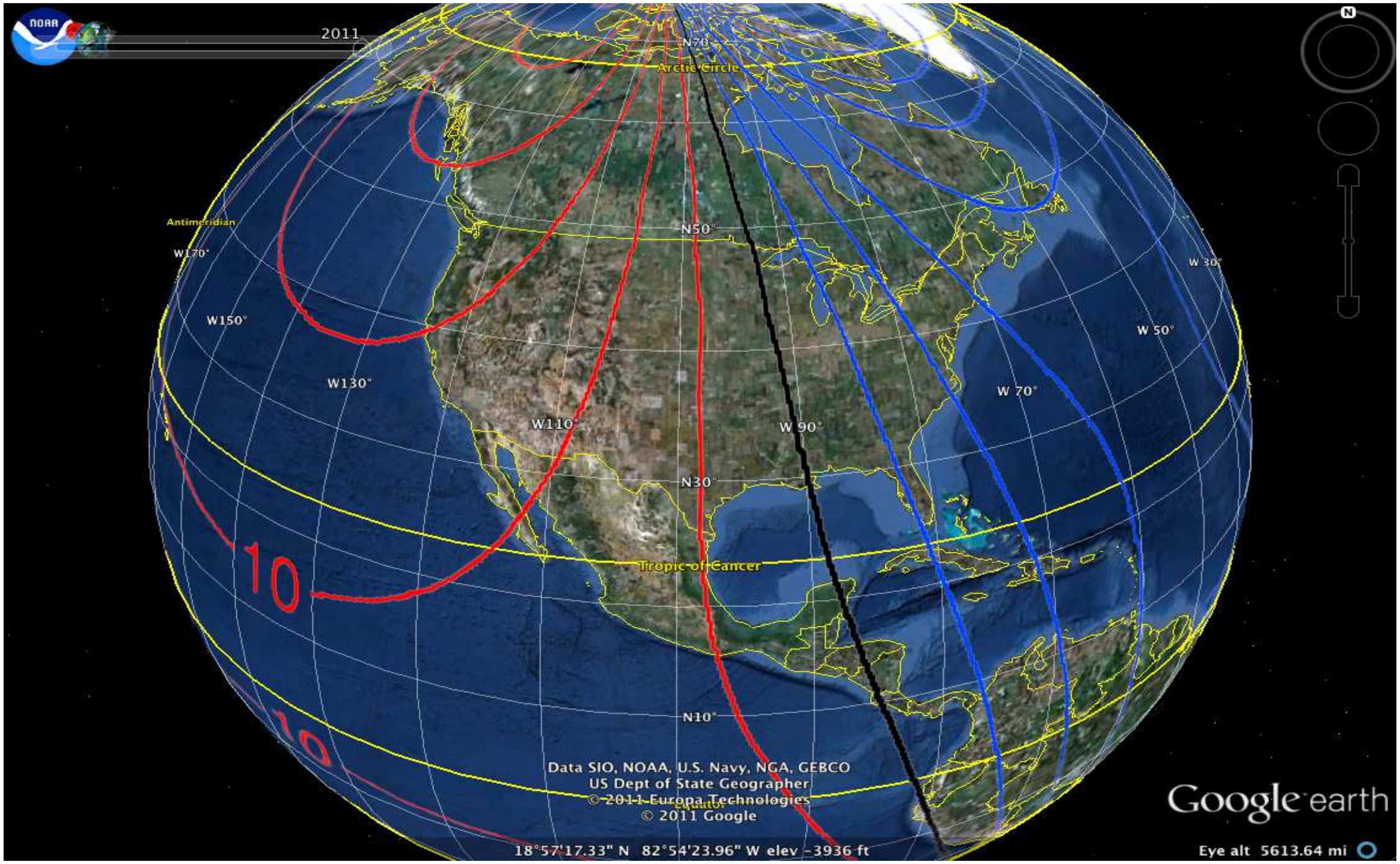


East-west Symmetry (spatial variations from EOF basis function mode 2)

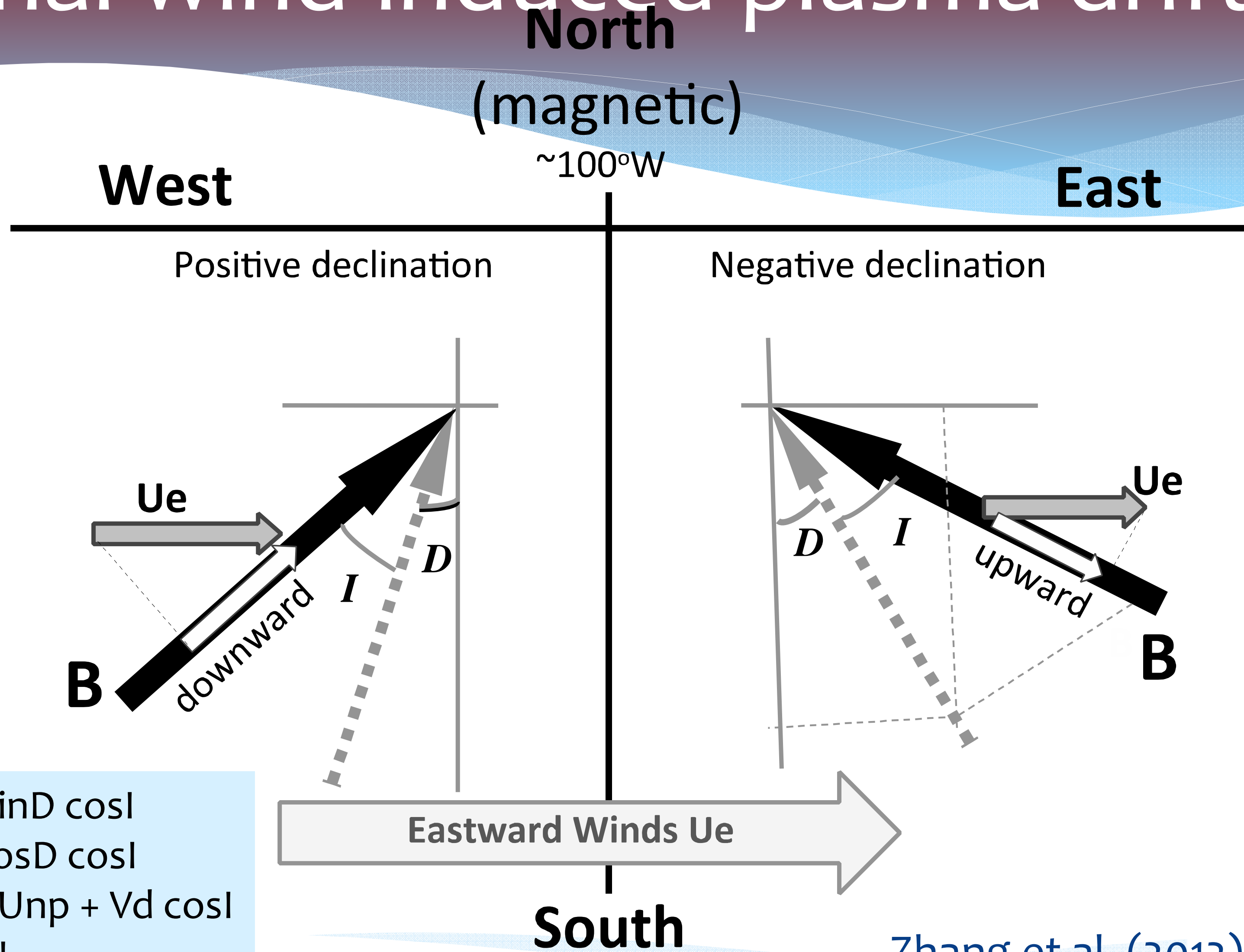


Zhang et al. (JGR, 2013; JGR, 2012; GRL; 2011)

Magnetic declination



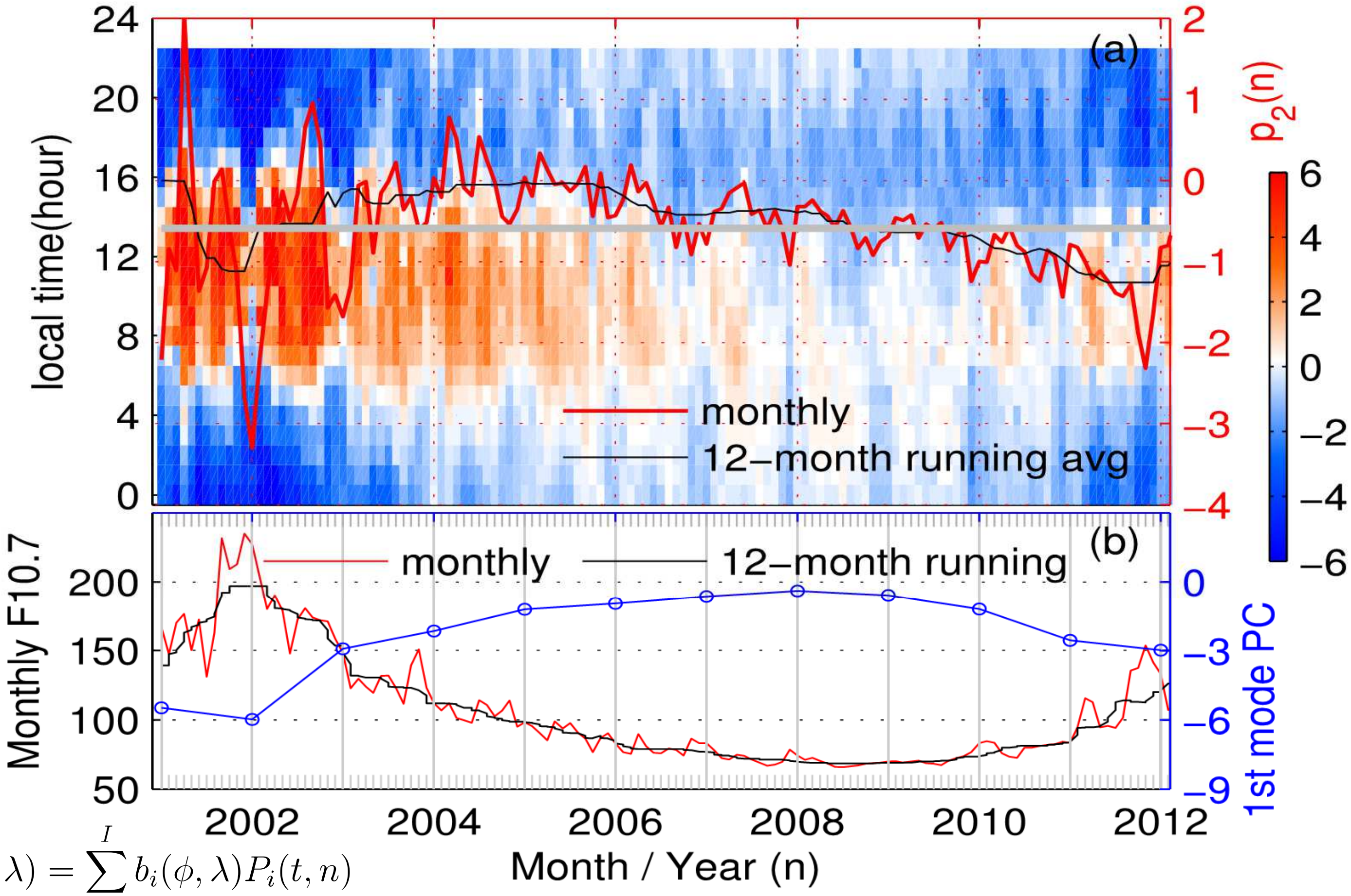
Zonal wind-induced plasma drifts



Uep = -Ue sinD cosI
Unp = Un cosD cosI
Vp = Uep + Unp + Vd cosI
Vzp = VpsinI

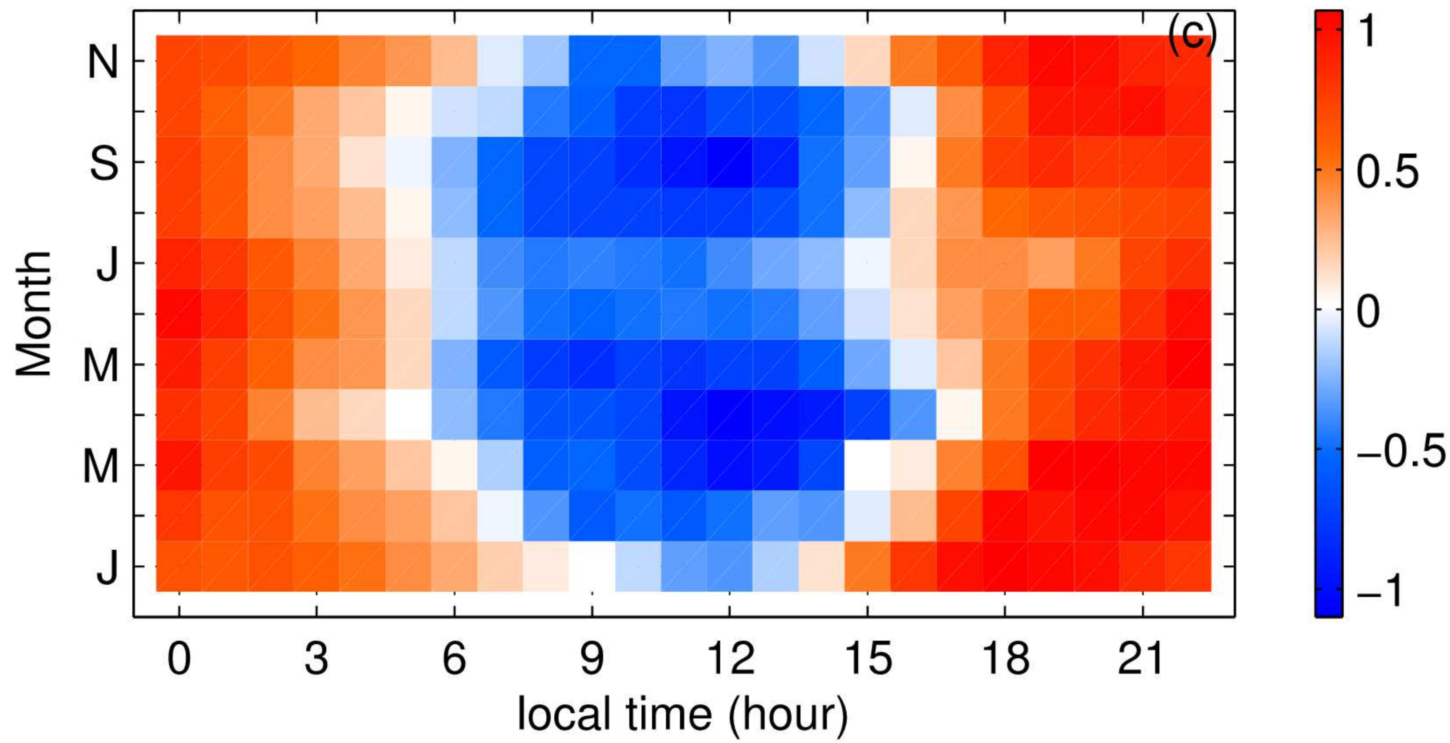
Zhang et al. (2012)

Temporal variation (EOF model 2)

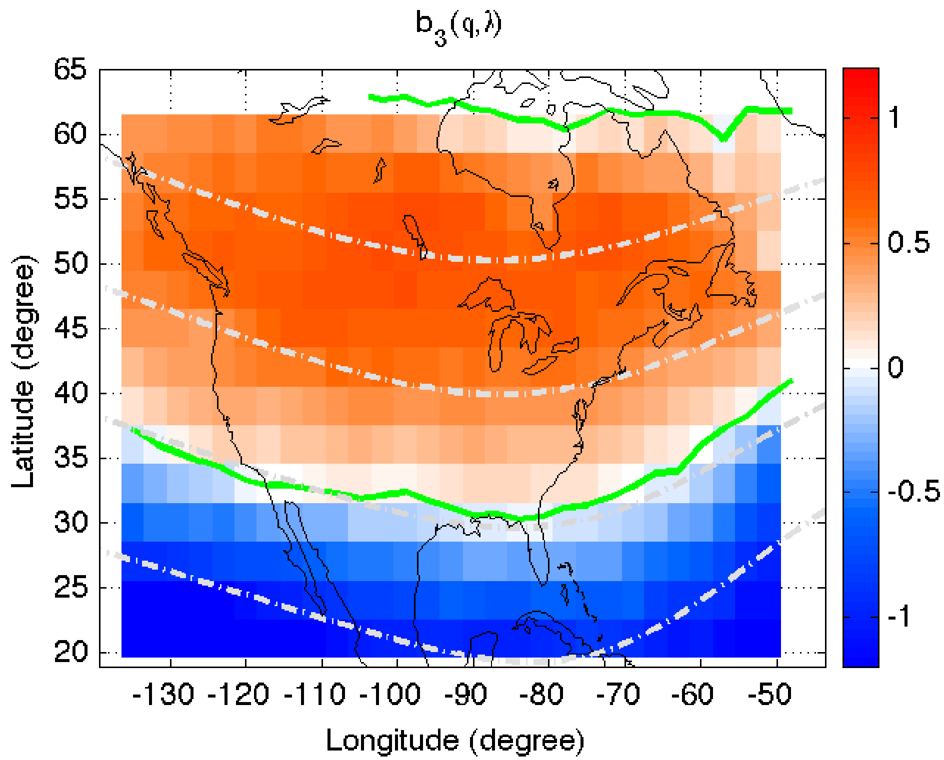


$$T(t,n,\phi,\lambda)=\sum_{i=1}^{I}b_i(\phi,\lambda)P_i(t,n)$$

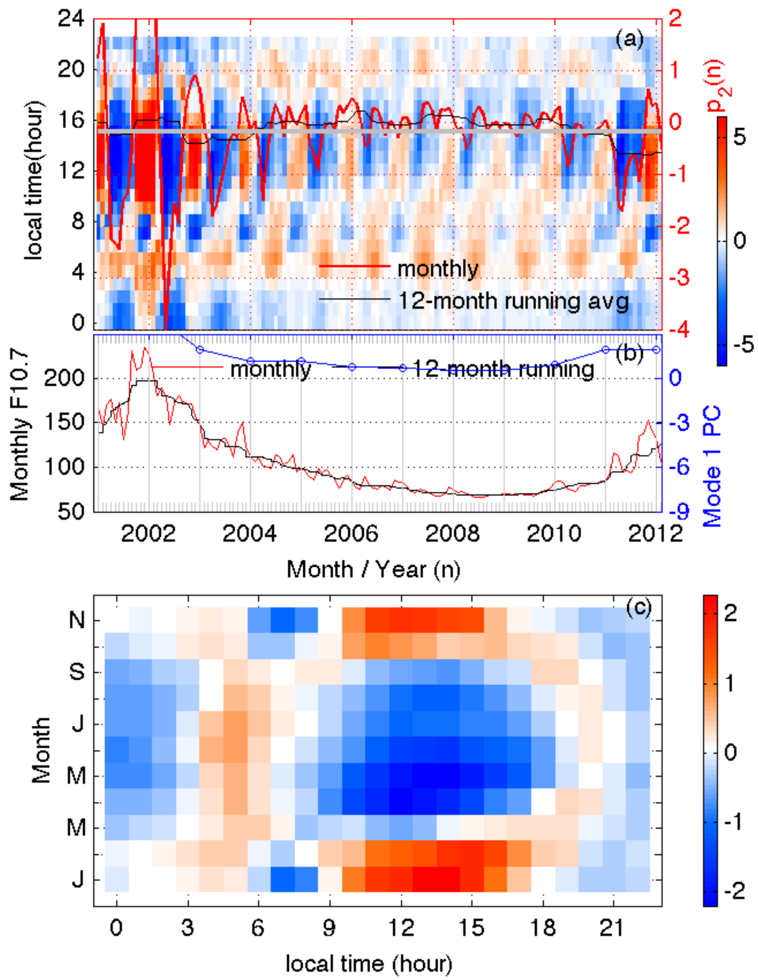
Further decomposed temporal variations (EOF mode 2)



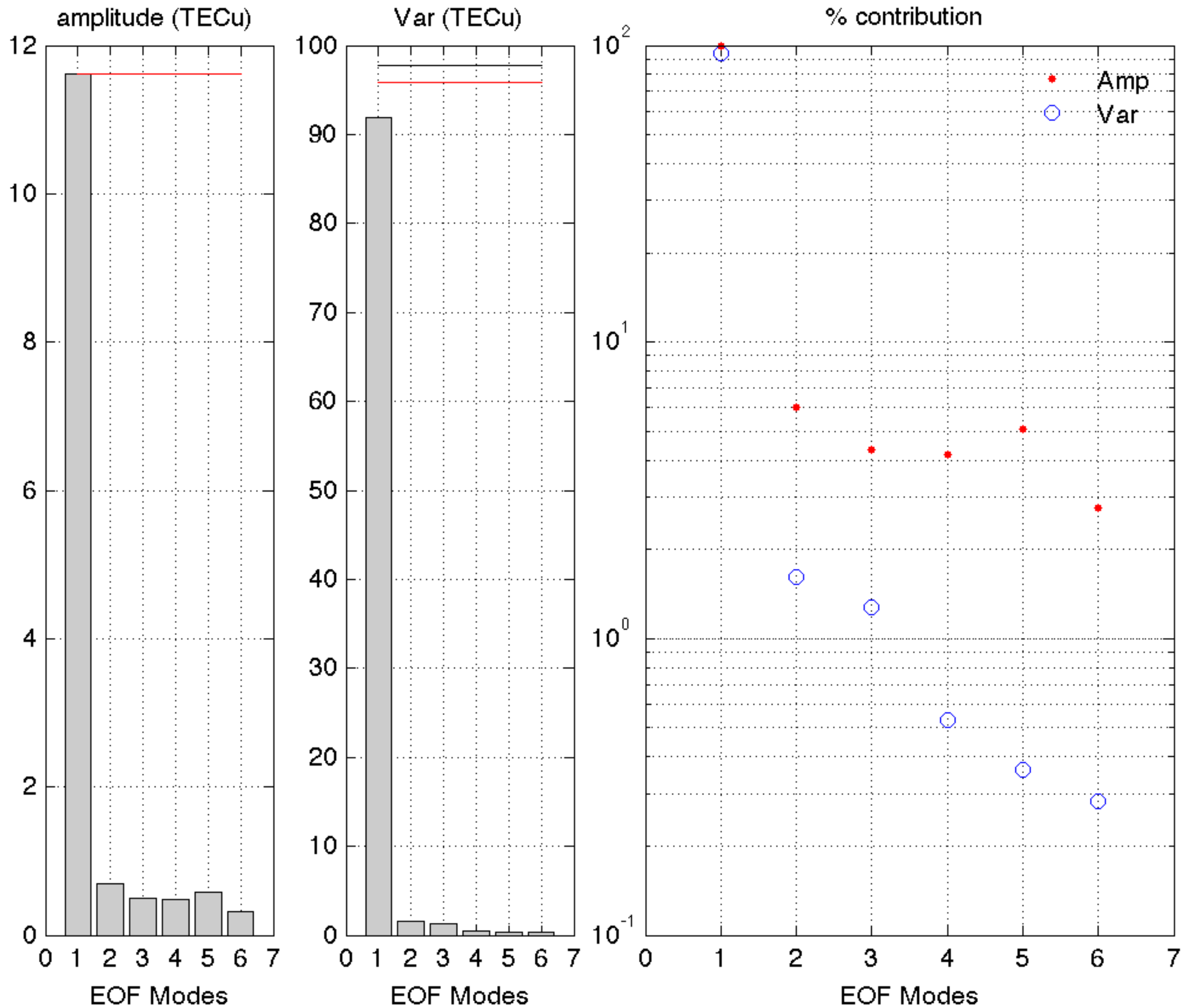
Spatial variations from EOF basis function mode 3



Temporal variations of EOF mode 3



Contributions from various EOF modes



North America Total Electron Content Model

Model NATEC: North America TEC **Output Parameter** TEC

Day Number (day)	Start 90	End 90	Interval 45	between 1-365
Local Time (hour)	Start 12	End 12	Interval 6	between 0-24
Geodetic Longitude (degree)	Start -13	End -48	Interval 3	between -135 --- -48 E
Geodetic Latitude (degree)	Start 21	End 65	Interval 3	for NATEC: 21-65
Solar Activity	F107 135	Fbar 135		F107: monthly mean; Fbar: 81-day mean between 70-300
Magnetic Activity	ap index 15			Ap index; between 0-100

RUN Reset

GPS TEC over North America

Outline

- Motivation
- EOF concept
- EOF-based North America GPS-TEC modeling
- EOF-based Near Real-time data processing
  - DIEOF procedure
  - Case studies
- Summary

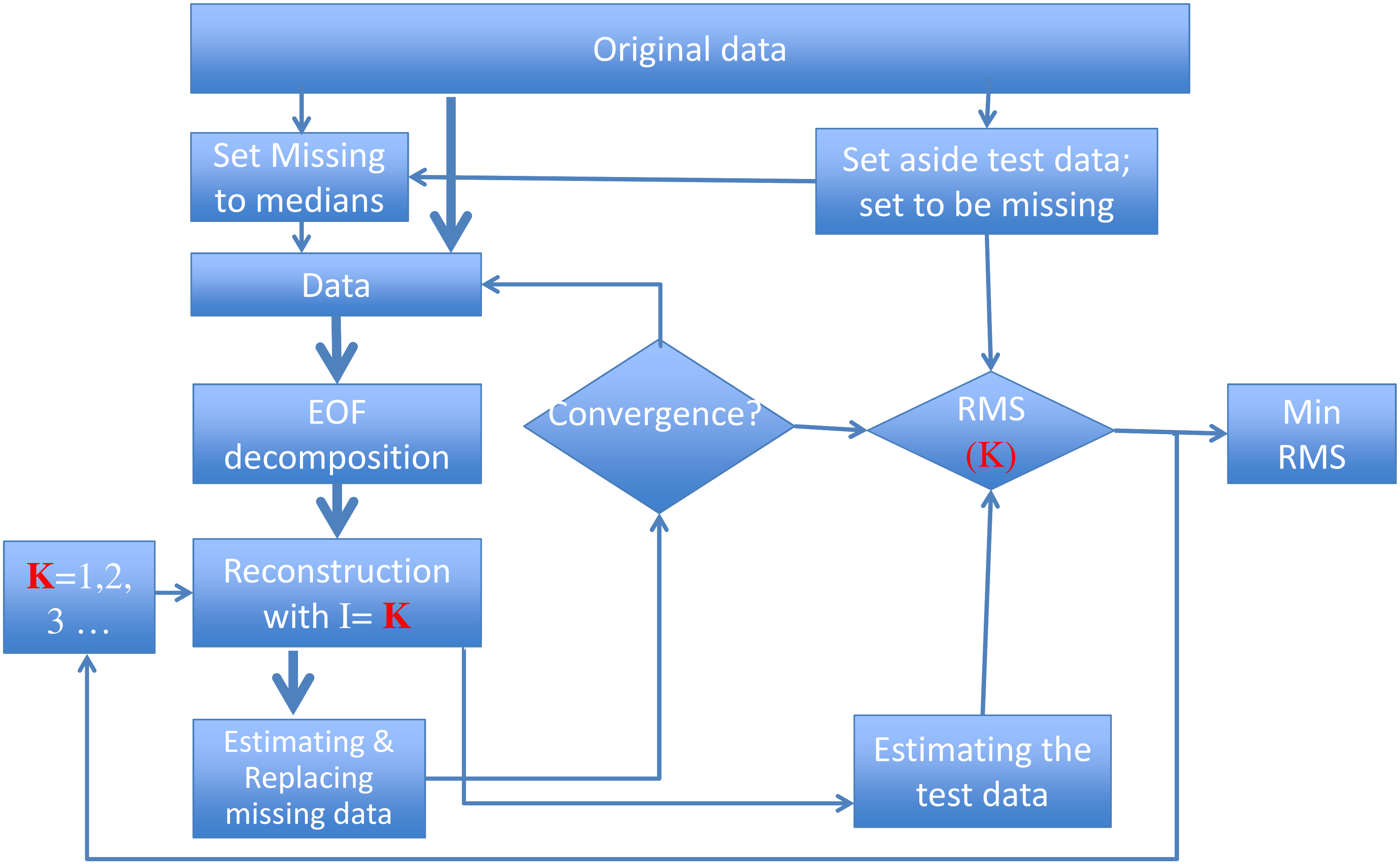
DIEOFs: Data Interpolating Empirical Orthogonal Functions

$$T(t,n,\phi,\lambda) = \sum_{i=1}^{I} b_i(\phi,\lambda) P_i(t,n)$$ $T_{i,j}$: i=1,2, ..., n; j=1:2, …, m

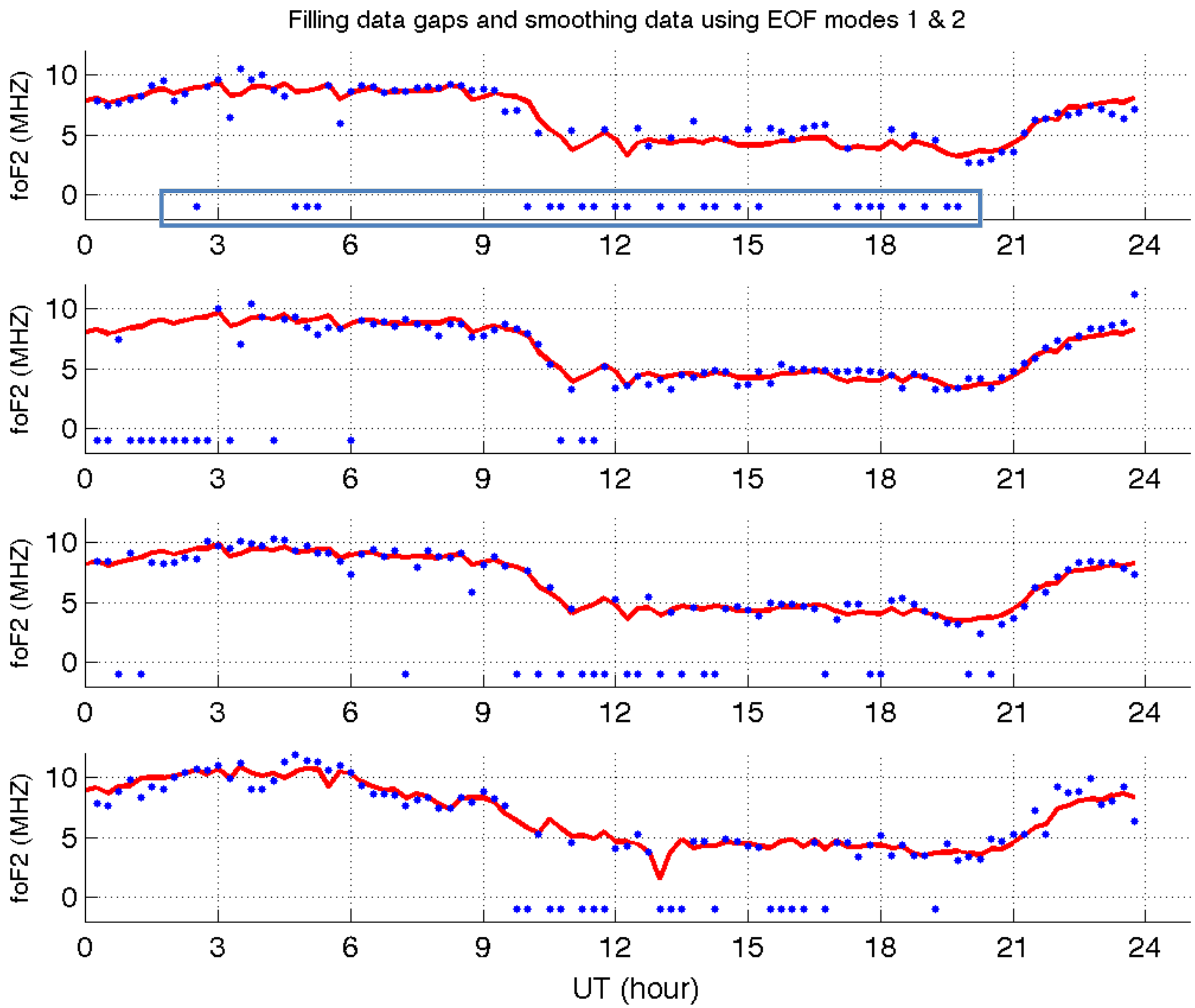
- The number of modes I equals to n or m;
- To deal with noisy data and data gaps:
  - When I is truncated to K, the EOF decomposition provides a reconstruction data, smooth and approximated to the original data.
  - If data represented by $T_{i,j}$ is incomplete, an iteration approach may be applied for deriving the best basis functions that may be used to fill in data gaps or even for prediction

Stoneback et al. (2013)

DIEOF procedure



First example

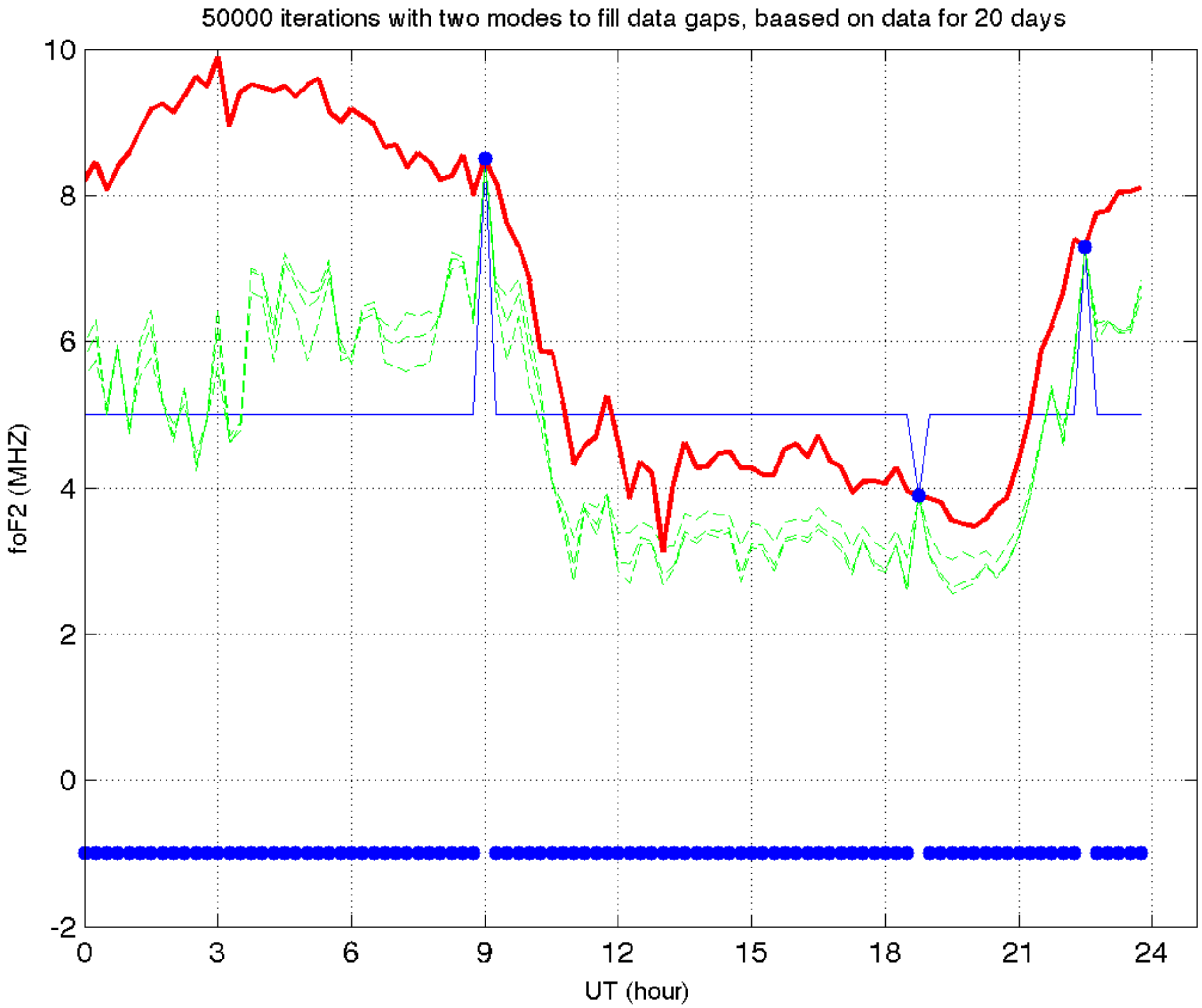


- Single site
- 10 days
- 15 min interval

- Using EOF modes 1 and 2
- After 50,000 iterations

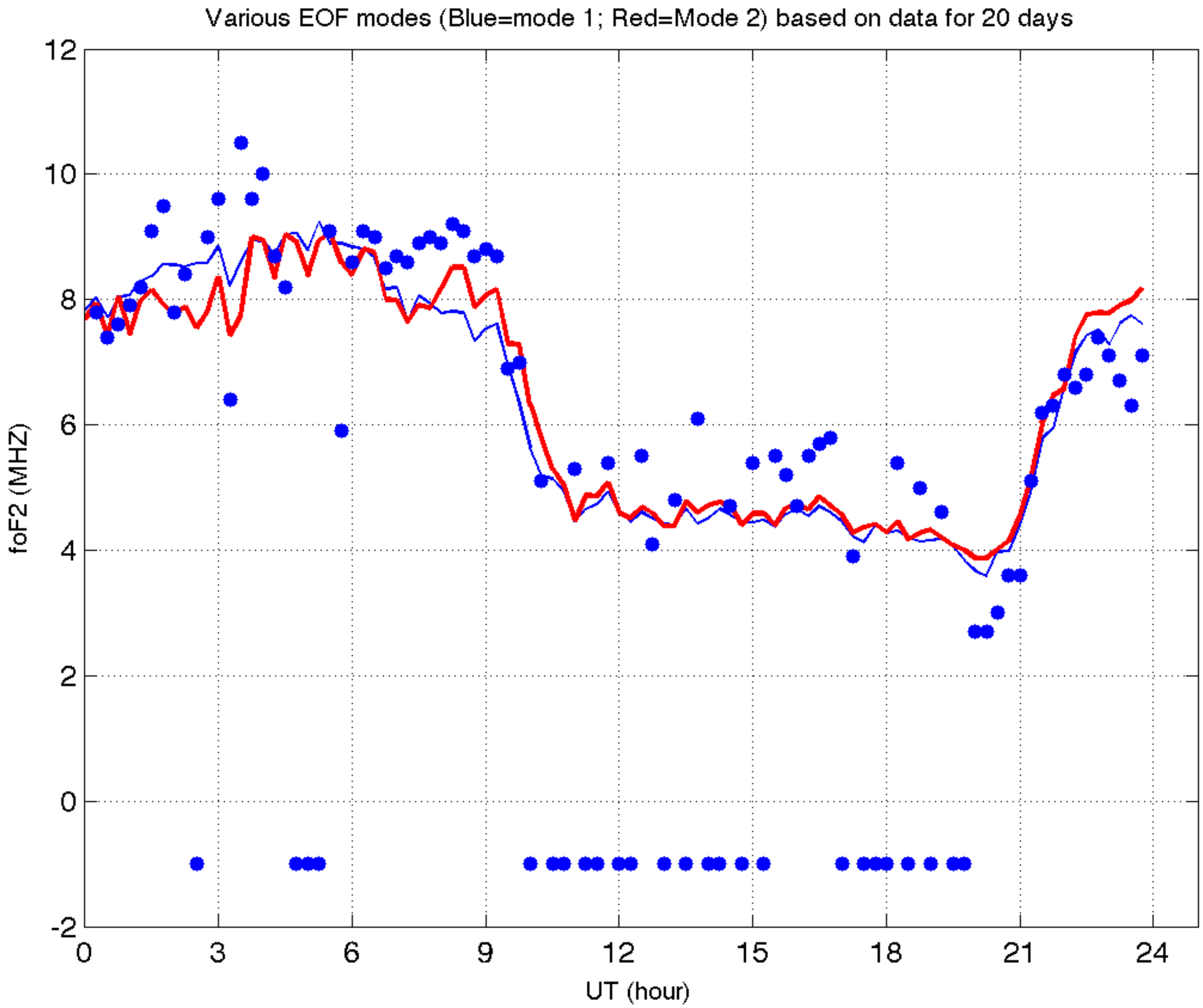
- Shown for days 1, 4,7, and 10.

Data gaps on day 6

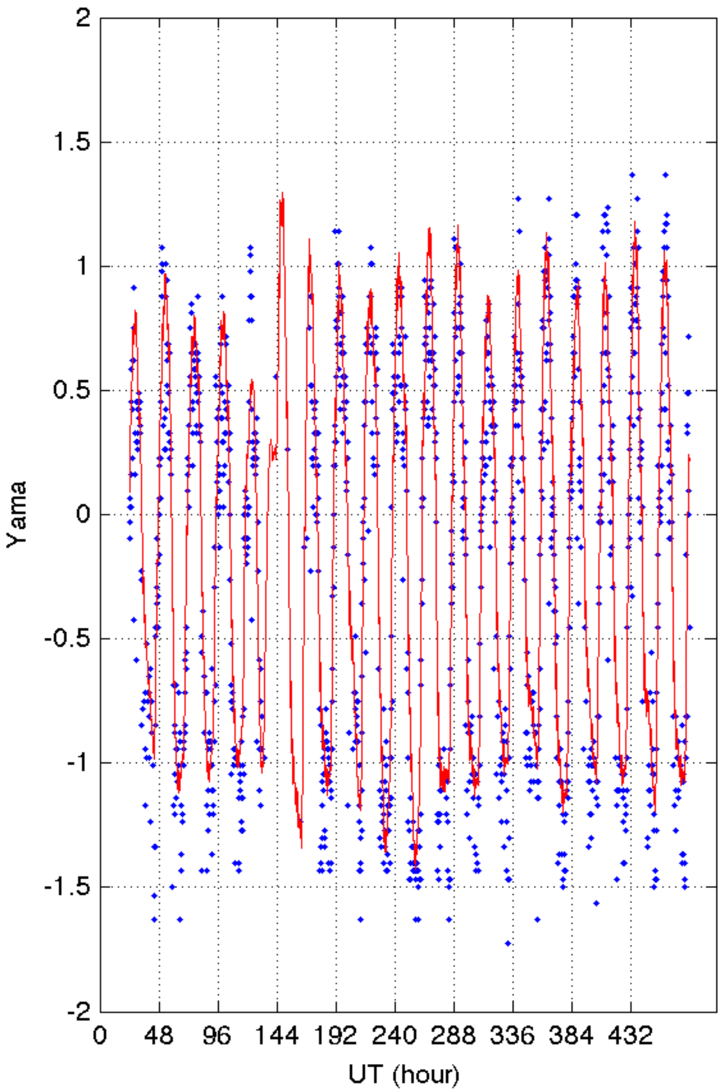


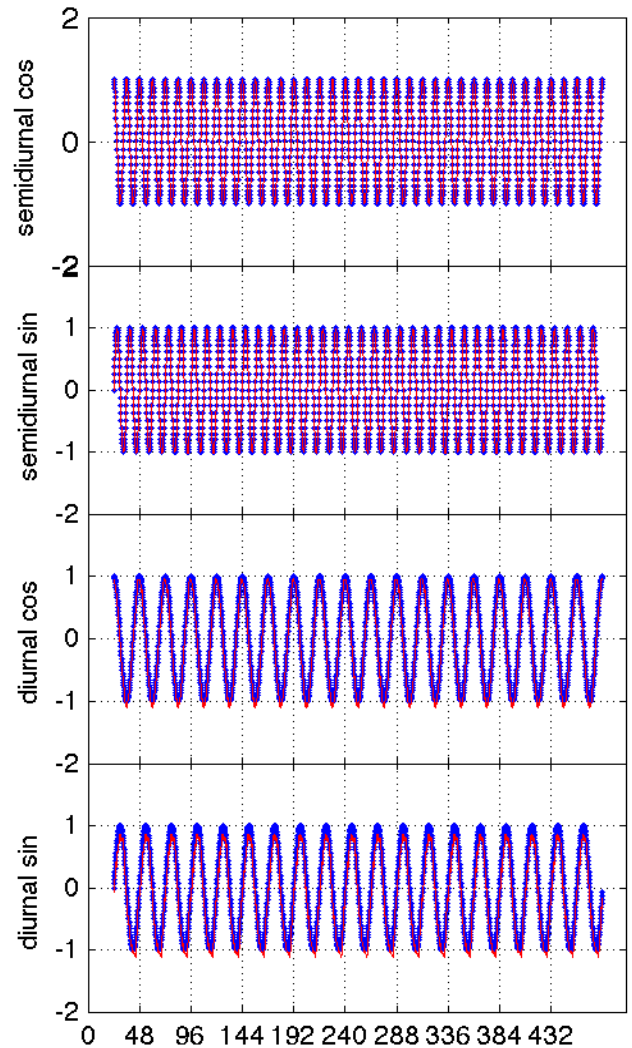
- Only 3 observational points
- Filled the gaps based on DIEOF
- 3 data points were assimilated

Modes 1 and 2


Various EOF modes (Blue=mode 1; Red=Mode 2) based on data for 20 days
12
10
8
6
4
2
0
-2
foF2 (MHZ)
0
3
6
9
12
15
18
21
24
UT (hour)

Second example





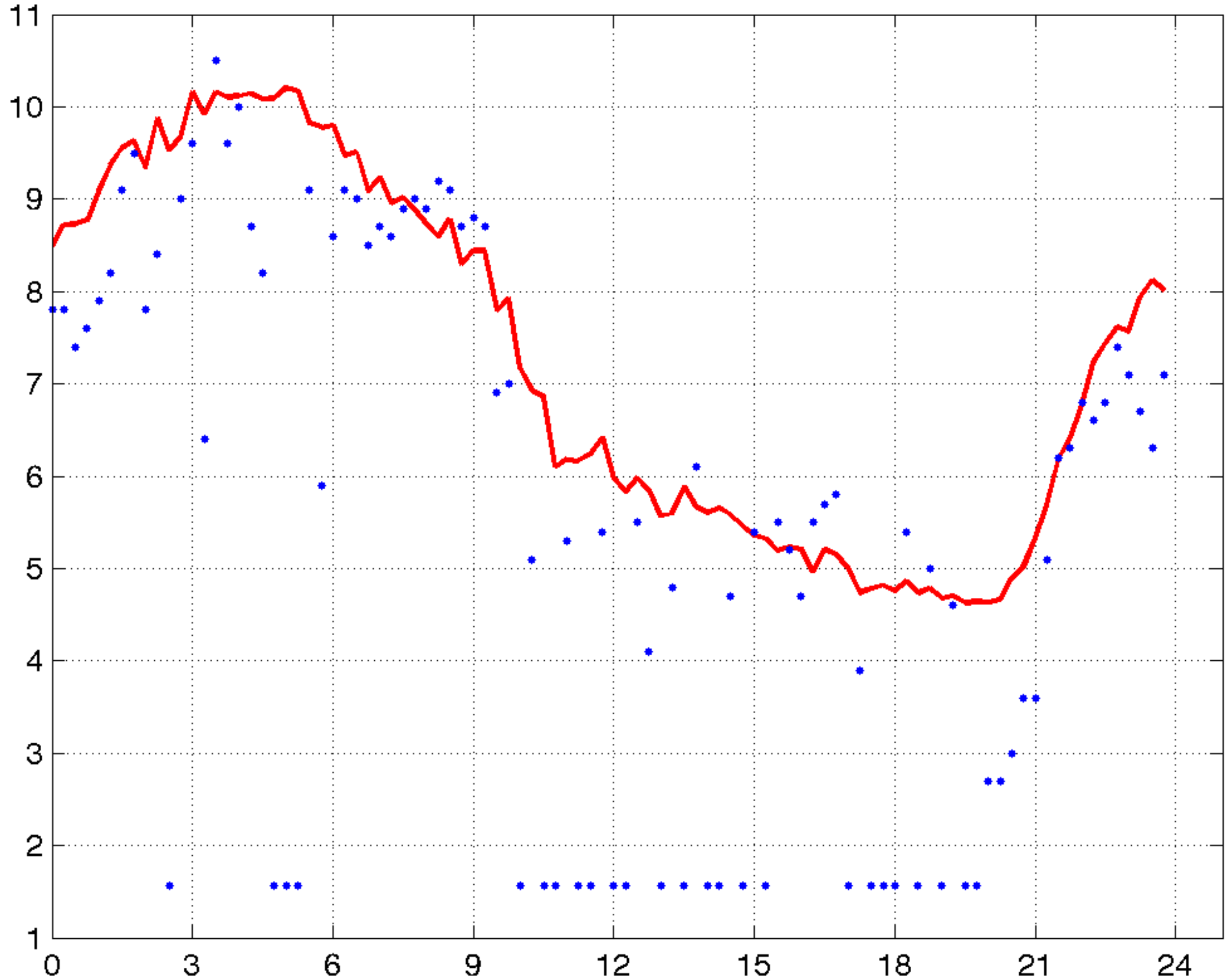
Single site

20 days

Data + 4 physical terms: (diurnal and semidiurnal component)

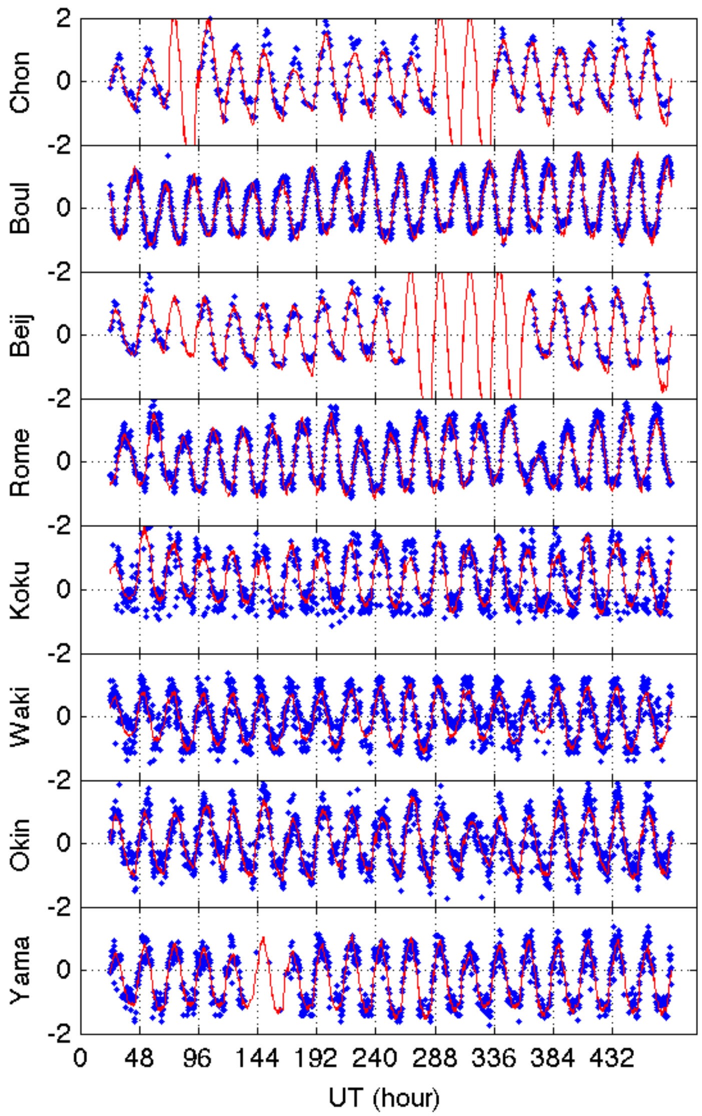
Fast convergence (but computational intensive)

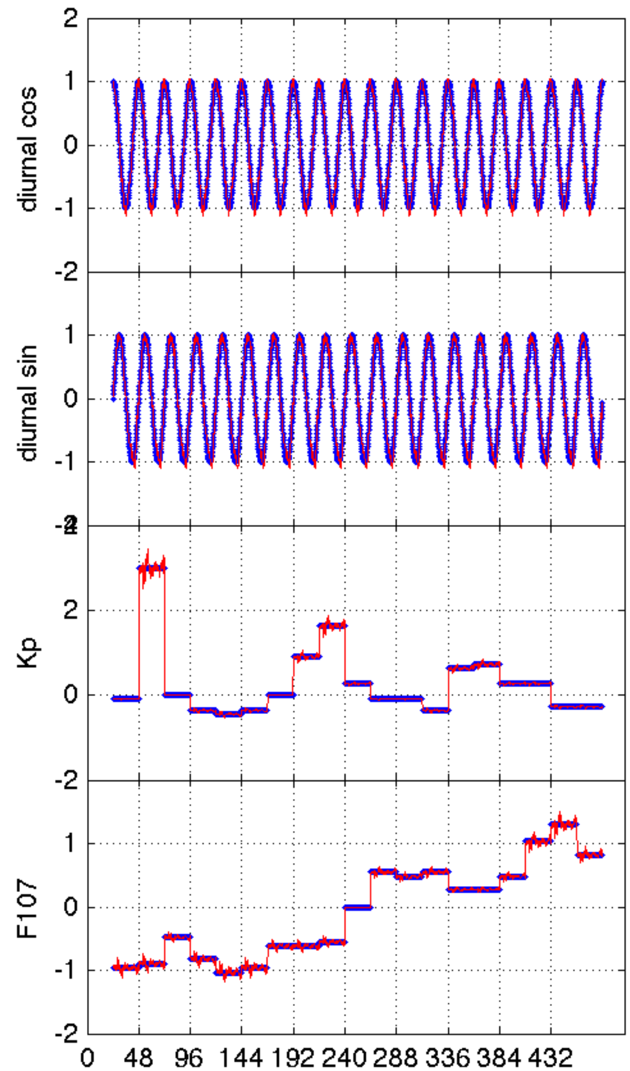
Modes are truncated to 4



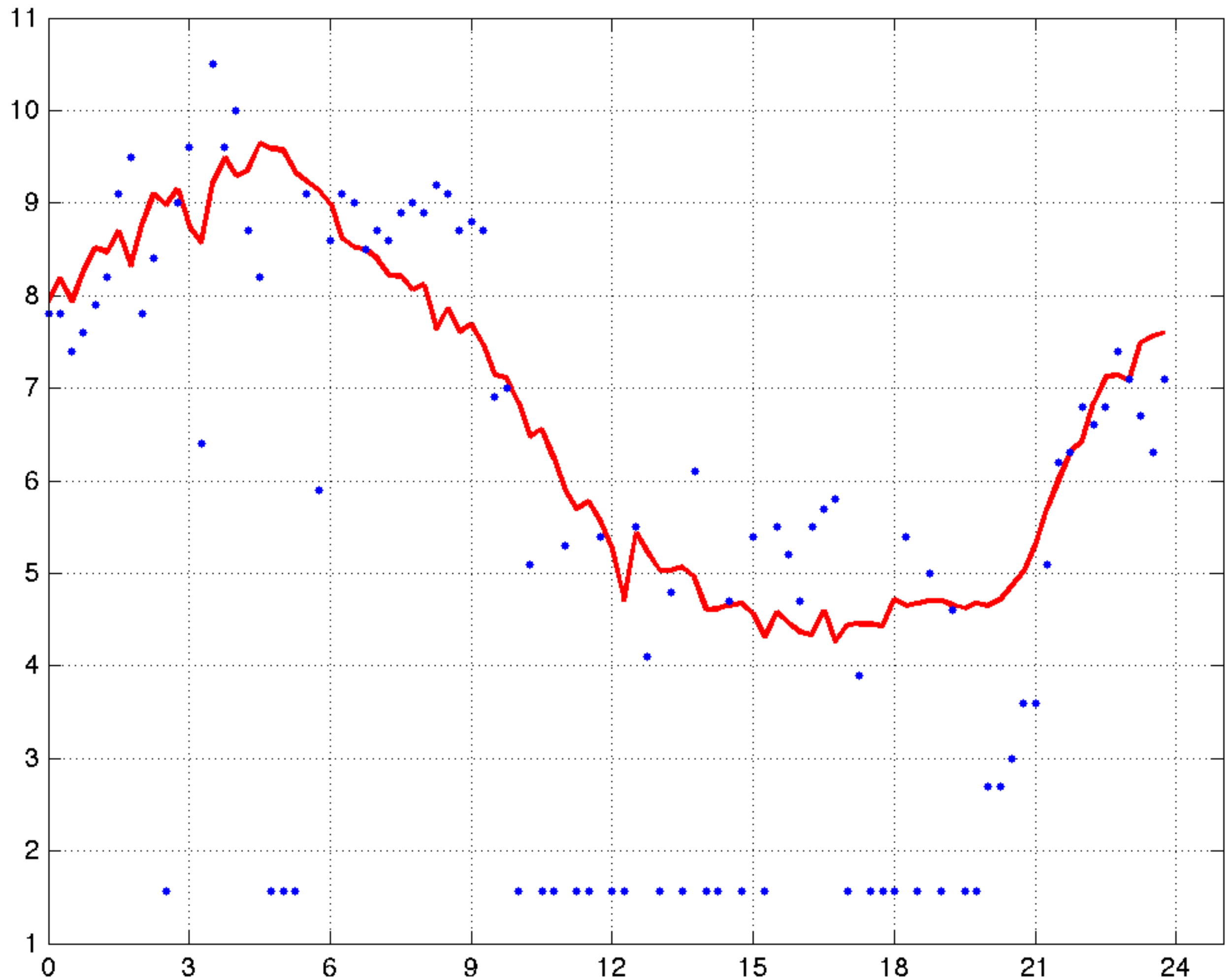
With up to 5000 iterations (if not convergent)

Third example

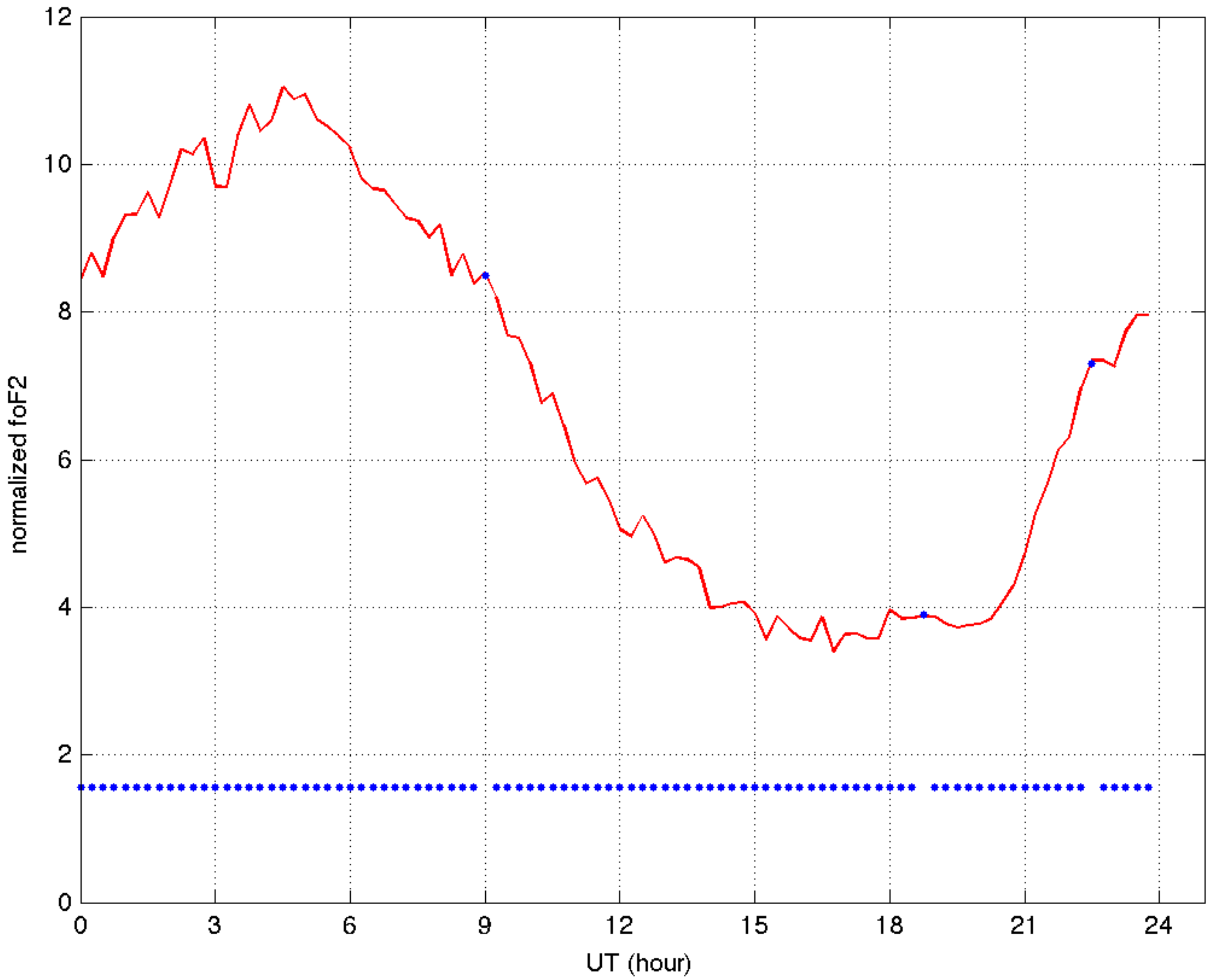




- 8 sites
- 20 days
- 4 physical terms



Third example



Third example

Error estimation, etc

- Set aside a portion of the data, and treat these test data as missing
- Go through the entire DIEOF procedure and estimate those test data, and compare these to original data to calculate RMS

Summary

- A powerful tool for modeling and real-time data processing
  - Smoothing data
  - Filling data gas
- Application for ionospheric predication
  - Using statistical correlation between each of EOF basis functions with geophysical parameters
  - Using EOF basis functions to fill “future data gaps” for predication

Real-time IRI @ MIT