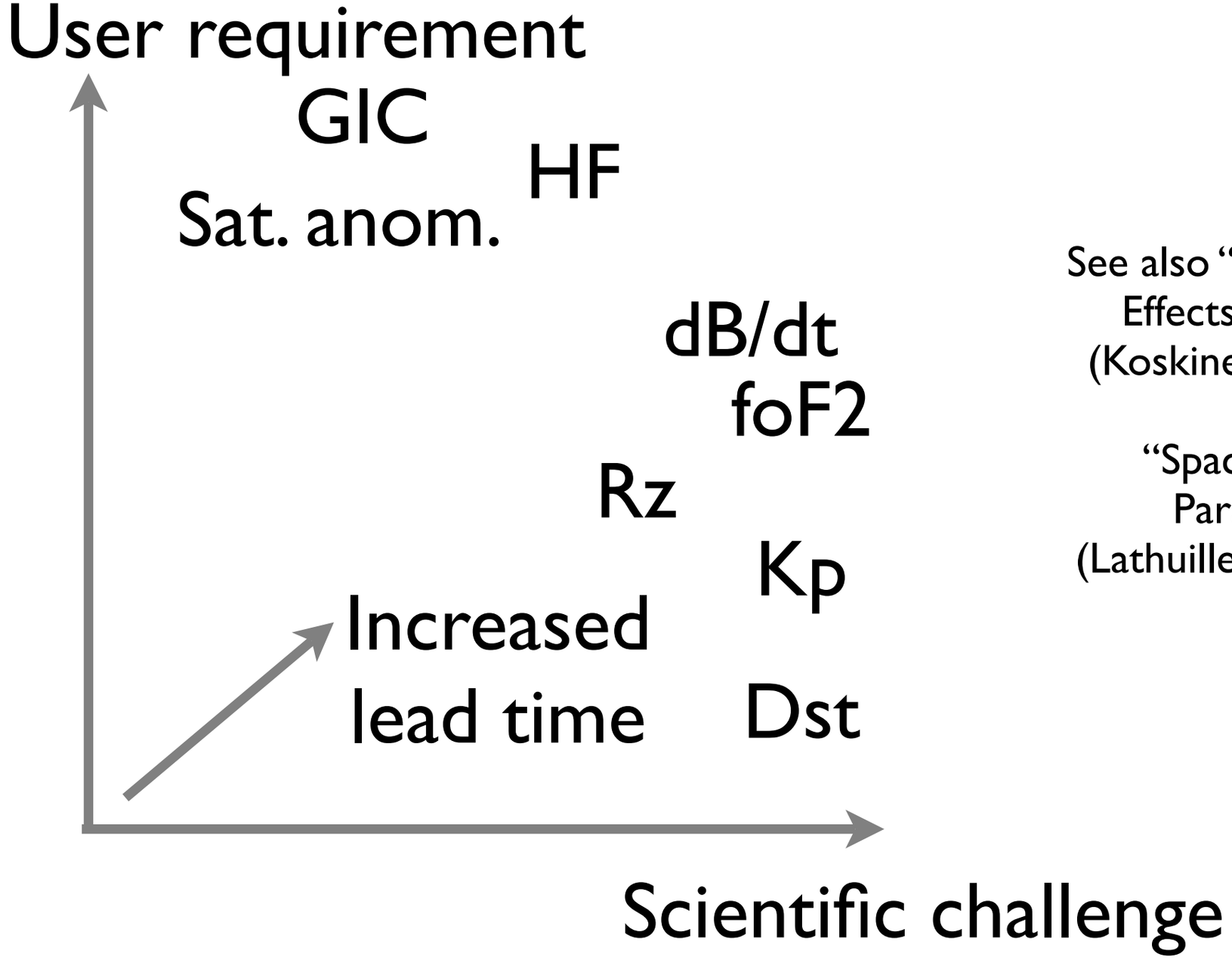
Data acquisition, development, and implementation of real time forecast models

Peter Wintoft, Henrik Lundstedt, Magnus Wik
Swedish Institute of Space Physics

Why forecast?

- Human need for planning future activities.
- Human need for responding to environmental changes.
- Ultimate test of a [physical] model.
- NOAA SWPC lists 25 users [G. Siscoe, 2007].
- ESA SWAPP list contain 19 users.

Forecast what and when?



See also "Space Weather Effects Catalogue" (Koskinen et al., 2001) and "Space Weather Parameters" (Lathuillere et al., 2001)

Prediction or forecast

- Wikipedia:
  - A prediction is a statement or claim that a particular event will occur in the future in more certain terms than a forecast.
  - Forecasting is the process of estimation in unknown situations. Prediction is a similar, but more general term, and usually refers to estimation of time series, cross-sectional or longitudinal data.
- Commercial business do forecasts (h=2.45 [5.40])
- Scientists do predictions (h=0.67 [0.12])
- Prediction (historic comparison), forecast (future) [Vassiliadis, IEEE P.S., 2000]

h=Google hit ratio of forecast/prediction using site:com or site:gov [site:co.uk or site:ac.uk].

Data

Data

- It's complex!

Data

- It's complex!
- Most scientists set up their on databases (text files, SQL, CDF, ...).

Data

- It's complex!
- Most scientists set up their on databases (text files, SQL, CDF, ...).
- Historic data:

Data

- It's complex!
- Most scientists set up their on databases (text files, SQL, CDF, ...).
- Historic data:
  - NGDC (SPIDR, OMNI), SIDC, SWENET, WDC, ...

Data

- It's complex!
- Most scientists set up their on databases (text files, SQL, CDF, ...).
- Historic data:
  - NGDC (SPIDR, OMNI), SIDC, SWENET, WDC, ...
- Real time data:

Data

- It's complex!
- Most scientists set up their on databases (text files, SQL, CDF, ...).
- Historic data:
  - NGDC (SPIDR, OMNI), SIDC, SWENET, WDC, ...
- Real time data:
  - SWPC (former SEC), SIDC, SWENET, ...

Data

- It's complex!
- Most scientists set up their on databases (text files, SQL, CDF, ...).
- Historic data:
  - NGDC (SPIDR, OMNI), SIDC, SWENET, WDC, ...
- Real time data:
  - SWPC (former SEC), SIDC, SWENET, ...
- Data gaps are annoying. Especially when not in a standardised format. (NULL, NaN, 999.9, 1e33, ...)

Data

- It's complex!
- Most scientists set up their on databases (text files, SQL, CDF, ...).
- Historic data:
  - NGDC (SPIDR, OMNI), SIDC, SWENET, WDC, ...
- Real time data:
  - SWPC (former SEC), SIDC, SWENET, ...
- Data gaps are annoying. Especially when not in a standardised format. (NULL, NaN, 999.9, 1e33, ...)
- User data: Sat. anom., GIC, MUF, ...

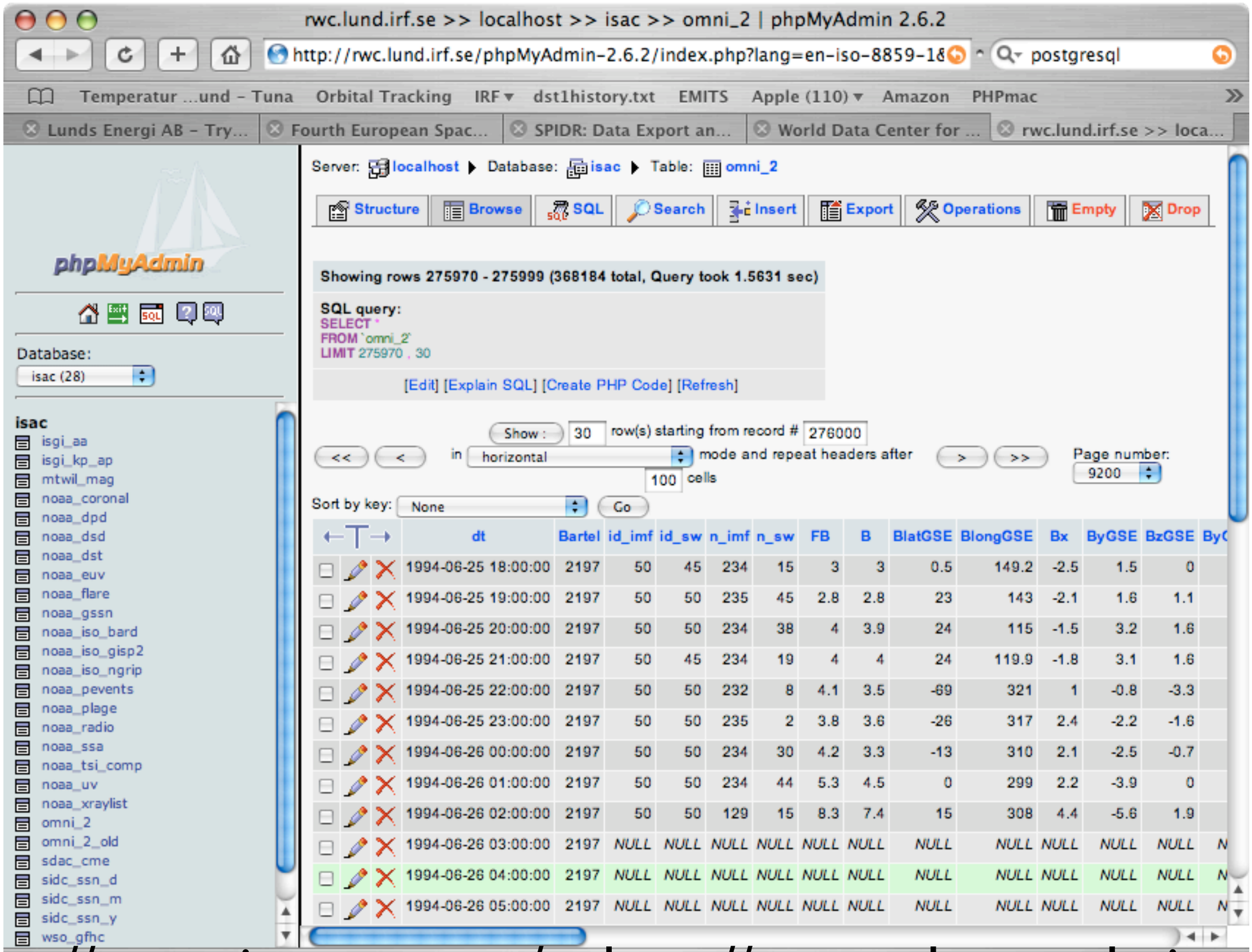
Databases (I)

- Common Data Format (CDF).
  - The CDF software package is used by hundreds of government agencies, universities, and private and commercial organizations as well as independent researchers on both national and international levels. CDF has been adopted by the International Solar-Terrestrial Physics (ISTP) project as well as the Central Data Handling Facilities (CDHF) as their format of choice for storing and distributing key parameter data.
- Hierarchical Data Format (HDF).
  - A versatile data model that can represent very complex data objects and a wide variety of metadata. A completely portable file format with no limit on the number or size of data objects in the collection. A software library that runs on a range of computational platforms, from laptops to massively parallel systems, and implements a high-level API with C, C++, Fortran 90, and Java interfaces. A rich set of integrated performance features that allow for access time and storage space optimizations. Tools and applications for managing, manipulating, viewing, and analyzing the data in the collection.

Databases (II)

- Structured Query Language (SQL).
  - PostgreSQL (SWENET)
    - PostgreSQL is a powerful, open source relational database system. It has more than 15 years of active development and a proven architecture that has earned it a strong reputation for reliability, data integrity, and correctness. It runs on all major operating systems, including Linux, UNIX (AIX, BSD, HP-UX, SGI IRIX, Mac OS X, Solaris, Tru64), and Windows.
  - MySQL (ISAC, ESA Pilot GIC)
    - The MySQL® database has become the world's most popular open source database because of its consistent fast performance, high reliability and ease of use. It's used in more than 11 million installations ranging from large corporations to specialized embedded applications on every continent in the world. (Yes, even Antarctica!)
- Large (huge) SQL user community.
- Large collection of tools (Java, C, Perl, PHP, Matlab, Mathematica, IDL, ...).
- SWPC setting up open MS SQL database [R. Zwickl, Yesterdays talk].

Example of SQL database: ESA ISAC project



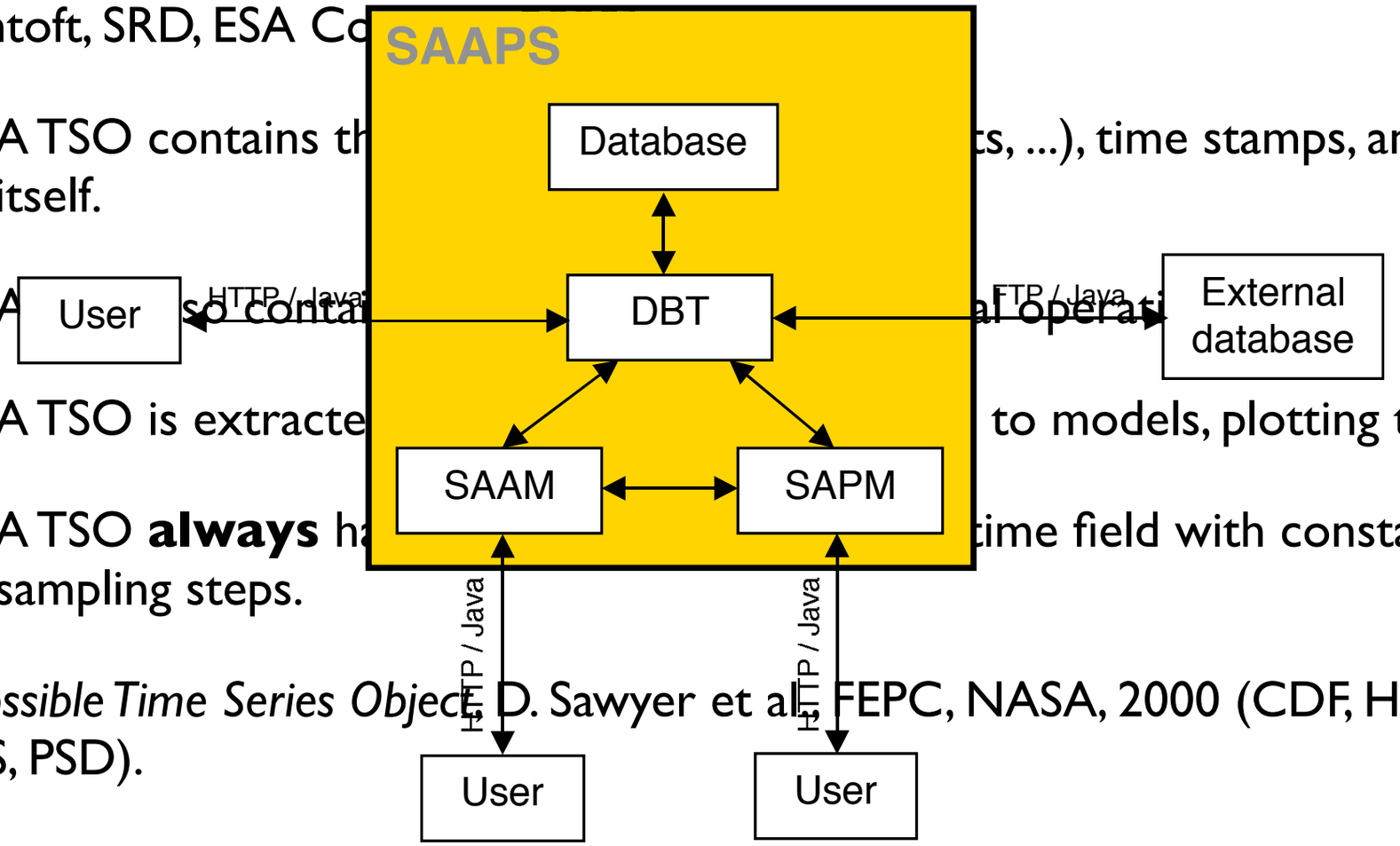
http://www.isac-esa.org/ http://www.phpmyadmin.net/

Time series object

- During the SAAPS project a notion of a time series object (TSO) emerged [P. Wintoft, SRD, ESA Contract, 2000]:
  - A TSO contains the meta data (field names, units, ...), time stamps, and the data itself.
  - A TSO also contains basic (often used) statistical operations.
  - A TSO is extracted from a database and passed to models, plotting tools, ...
  - A TSO **always** have monotonically increasing time field with constant sampling steps.
- *A Possible Time Series Object*, D. Sawyer et al., FEPC, NASA, 2000 (CDF, HDF4, IDFS, FITS, PSD).
- Other TSO, e.g. Matlab [`timeseries`], R [`ts`], ...

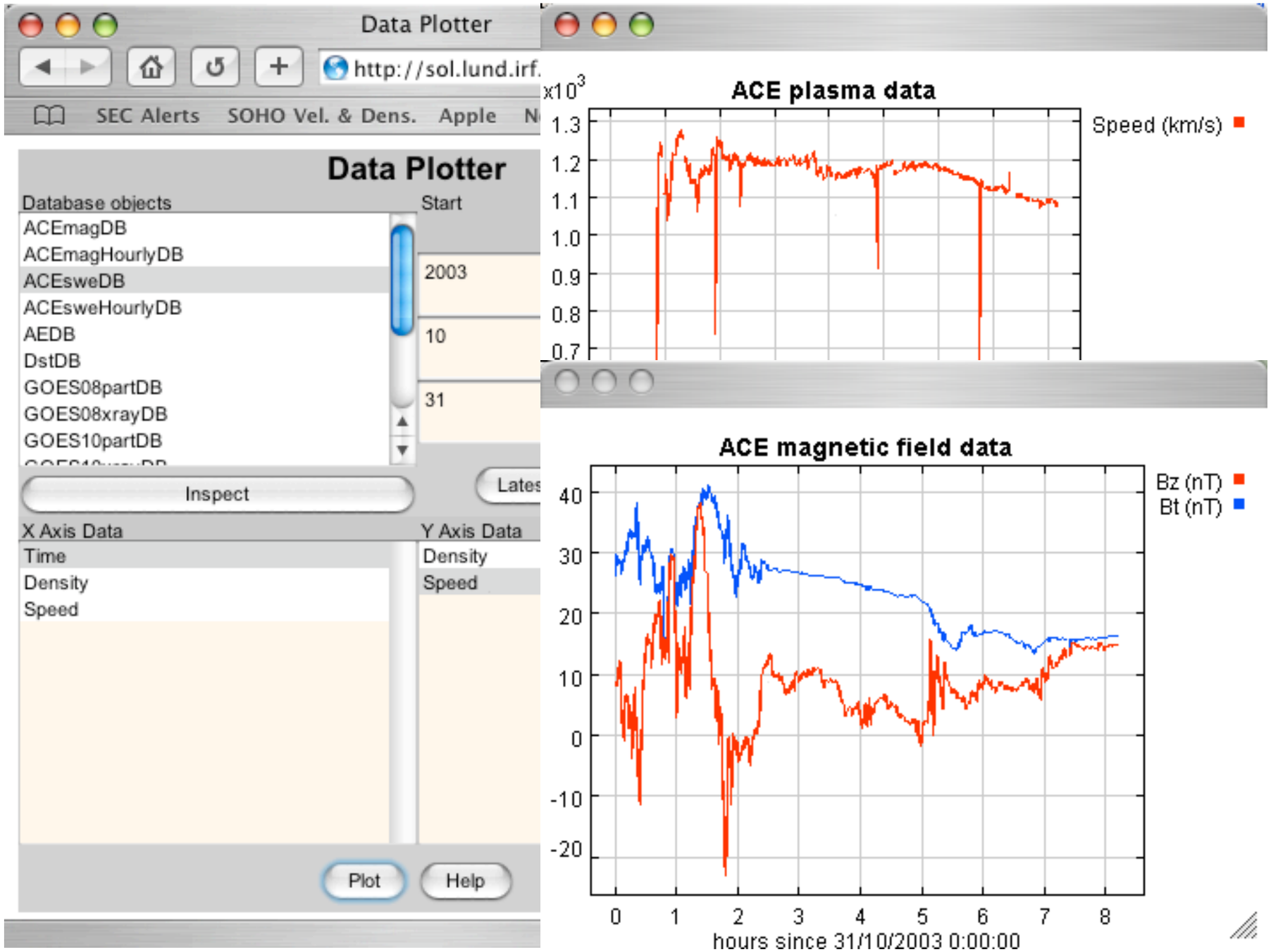
Time series object

- During the SAAPS project a notion of a time series object (TSO) emerged [P. Wintoft, SRD, ESA Co
  - A TSO contains th ts, ...), time stamps, and the data itself.
  - A so contai al operati
  - A TSO is extracte to models, plotting tools, ...
  - A TSO **always** ha time field with constant sampling steps.
- *A Possible Time Series Object*, D. Sawyer et al., FEPC, NASA, 2000 (CDF, HDF4, IDFS, FITS, PSD).
- Other TSO, e.g. Matlab [`timeseries`], R [`ts`], ...



TSO plot

Java TSO



ptplot from http://ptolemy.berkeley.edu/java/ptplot/

Models

- The prime goal of empirical models, and basically all models, is to make predictions of the future given information of the past (Farmer and Sidorowich, 1987).

Relationship between SW modelling & services

- Two-way development
  - Scientists use current knowledge implemented for real time operation.
  - Users have certain requirements on parameters, accuracy, lead time, ...
  - Education goes both ways.
- Sometimes they meet!

ESA Pilot GIC

ESA Pilot GIC

- Power company UR (2003): 95% accuracy with 1 hour lead time of GIC.

ESA Pilot GIC

- Power company UR (2003): 95% accuracy with 1 hour lead time of GIC.
- Model provides GIC forecast with 30 min. lead time.

ESA Pilot GIC

- Power company UR (2003): 95% accuracy with 1 hour lead time of GIC.
- Model provides GIC forecast with 30 min. lead time.
- "New UR": 1 hour forecast not prime interest, rather days.

http://www.lund.irf.se/gicpilot/gicforecast/

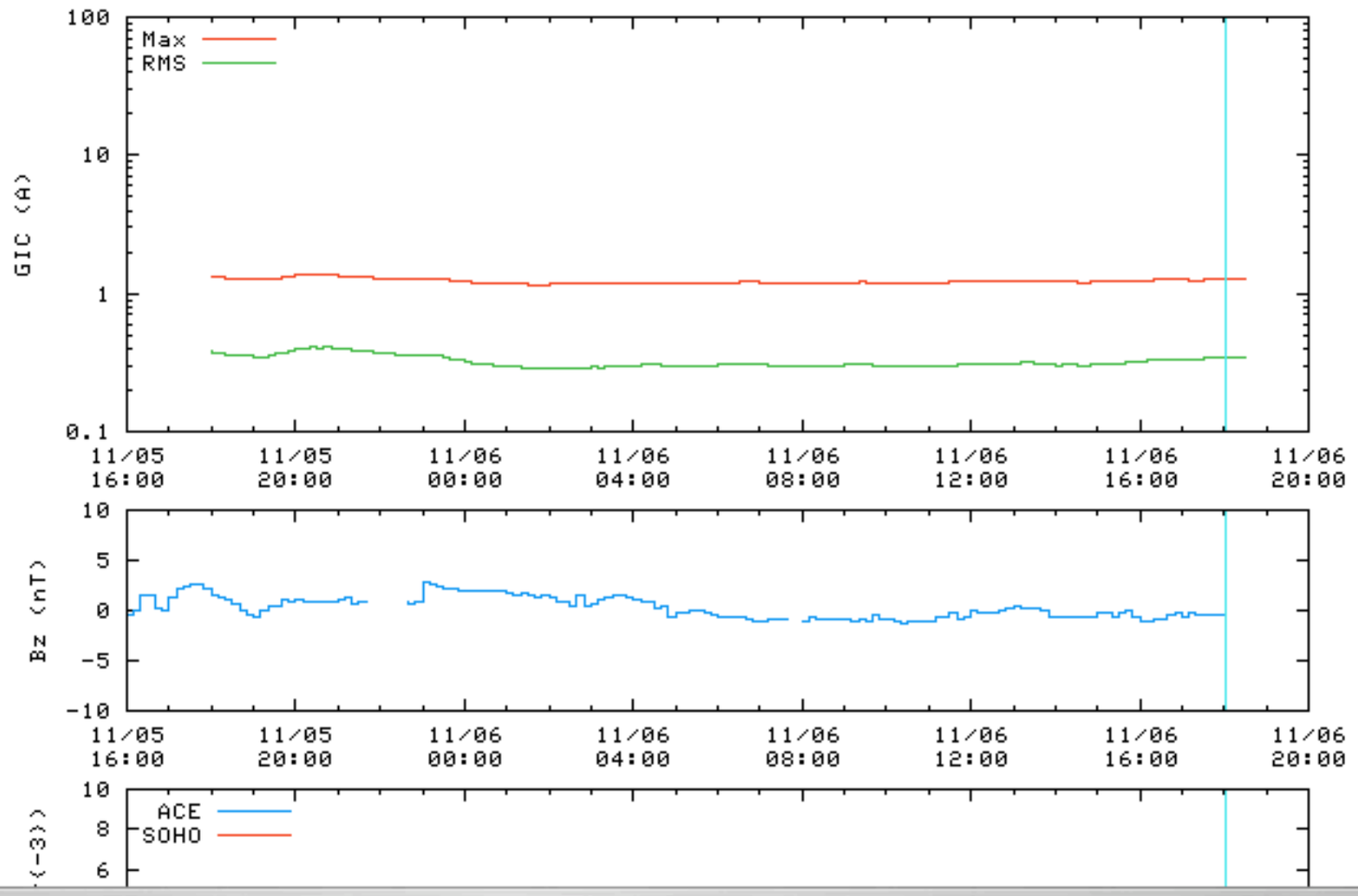
Real-time forecast service
for
Geomagnetically Induced Currents

Real time forecast of local GIC

Last 24 hours | Last 4 hours | Archive | Models | Changes

Forecast issued 2007-11-06 18:01:08 CET.

30 minute forecast of local 10 minute RMS and maximum GIC.

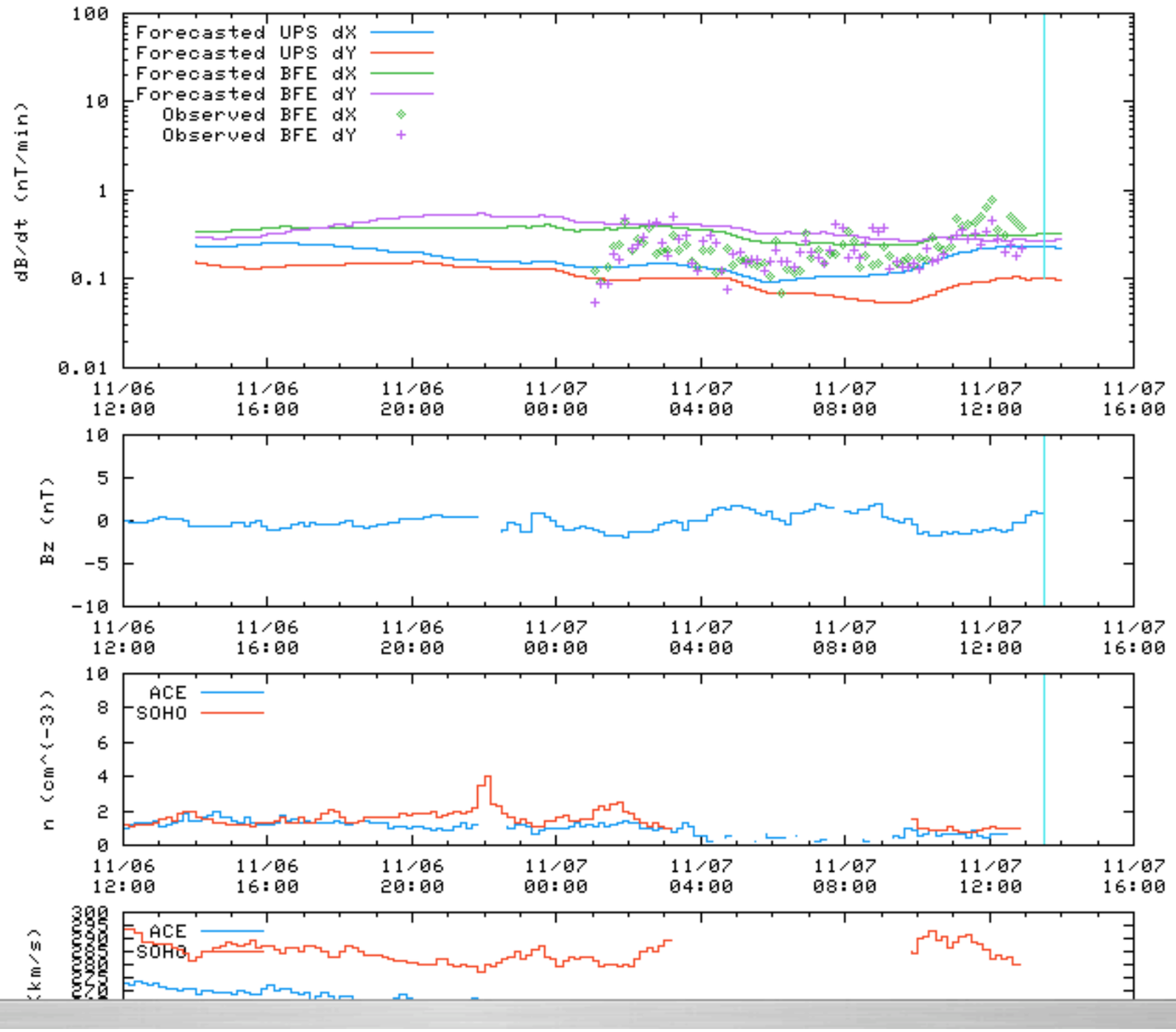


Real time RMS ΔB forecast for Uppsala and Brorfelde

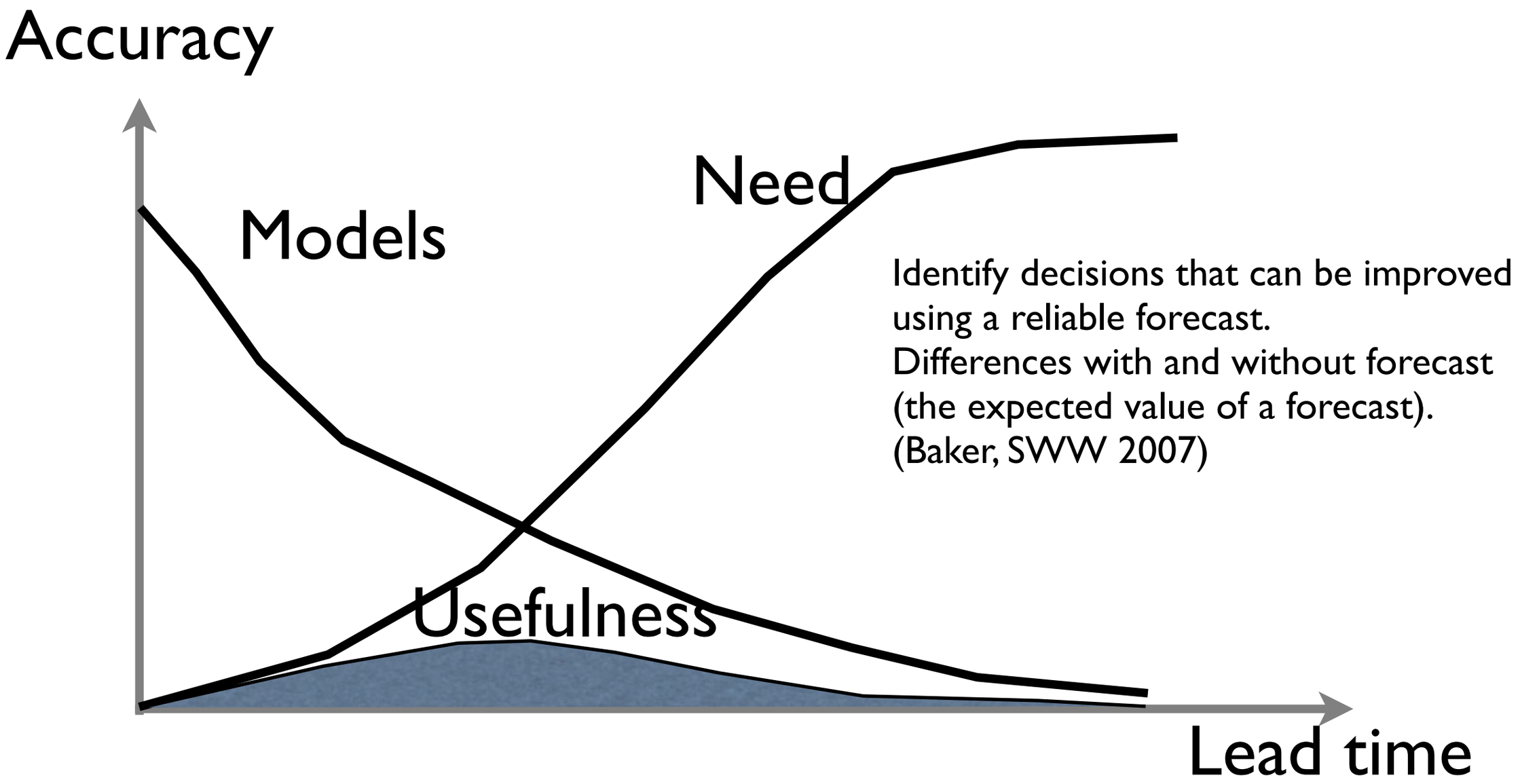
Last 24 hours | Last 4 hours | Archive | Models | Changes

Forecast issued 2007-11-07 13:31:07 CET.

30 minute forecast of local 10 minute RMS dB.

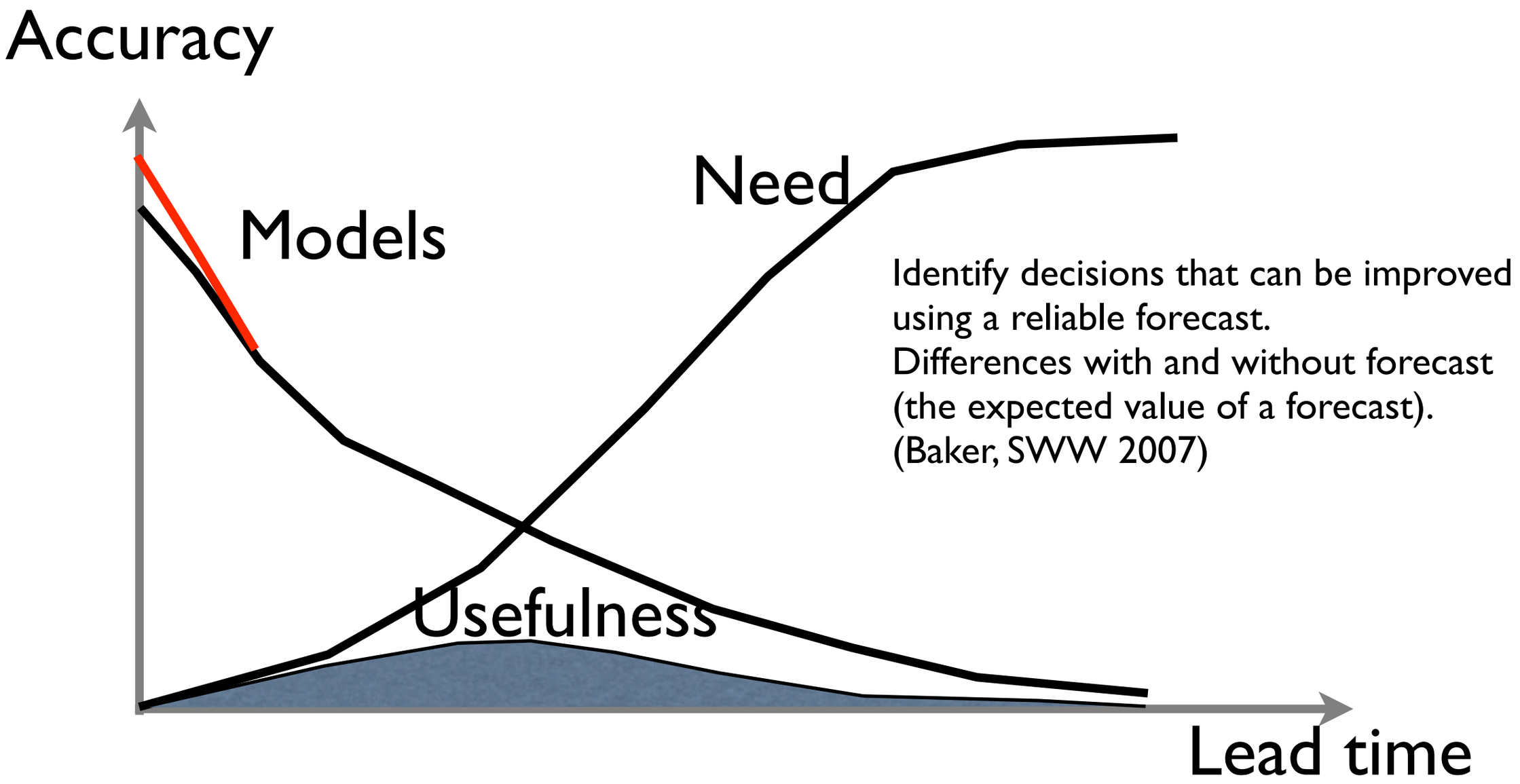


Prediction lead time



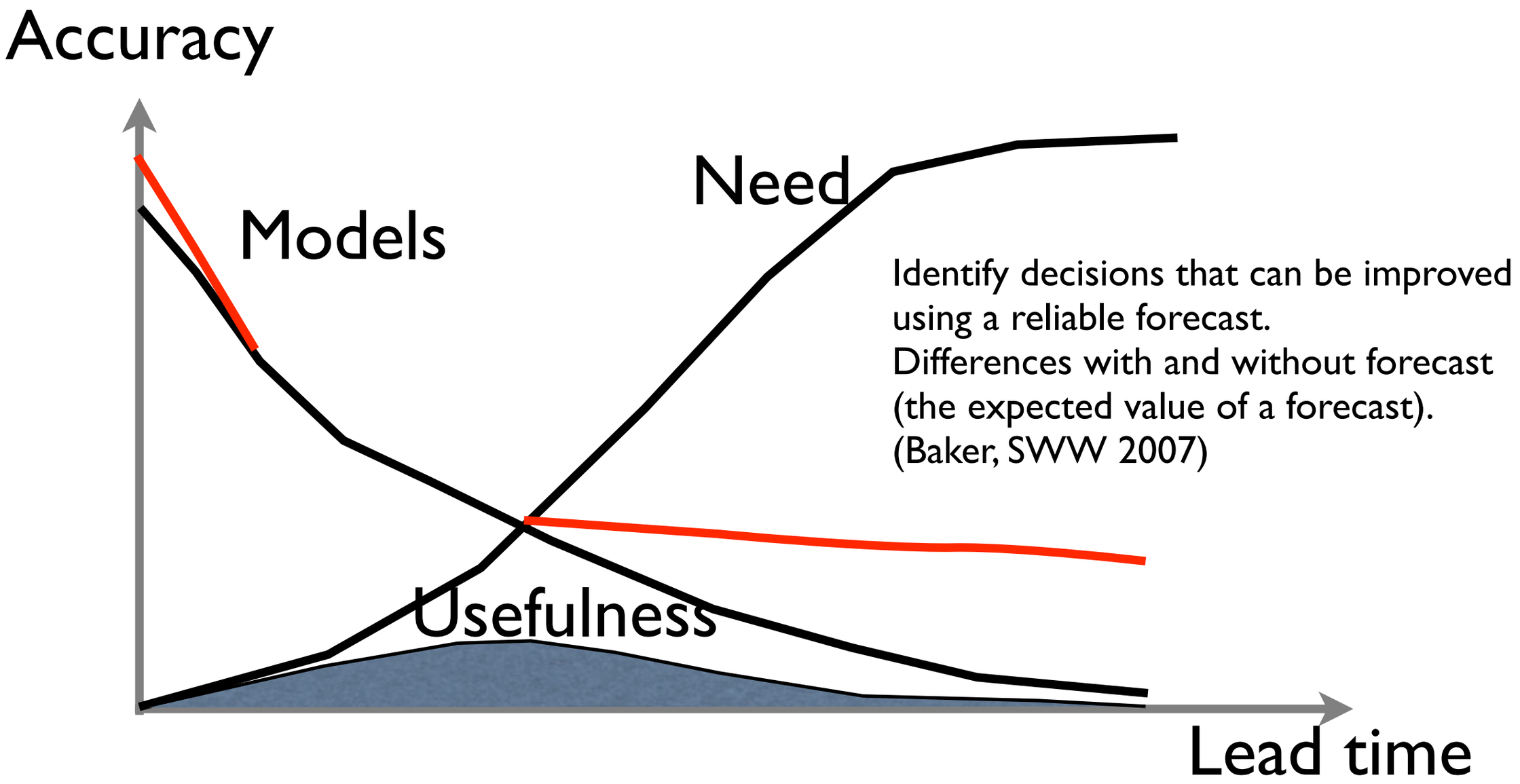
Identify decisions that can be improved using a reliable forecast.
Differences with and without forecast (the expected value of a forecast).
(Baker, SWW 2007)

Prediction lead time



Identify decisions that can be improved using a reliable forecast.
Differences with and without forecast (the expected value of a forecast).
(Baker, SWW 2007)

Prediction lead time



Identify decisions that can be improved using a reliable forecast.
Differences with and without forecast (the expected value of a forecast).
(Baker, SWW 2007)

Evaluating lead time

- Judging the true prediction lead time is quite difficult.
- Often a “statistical” lead time is given.
- However, more interesting to judge lead time from the dynamics.

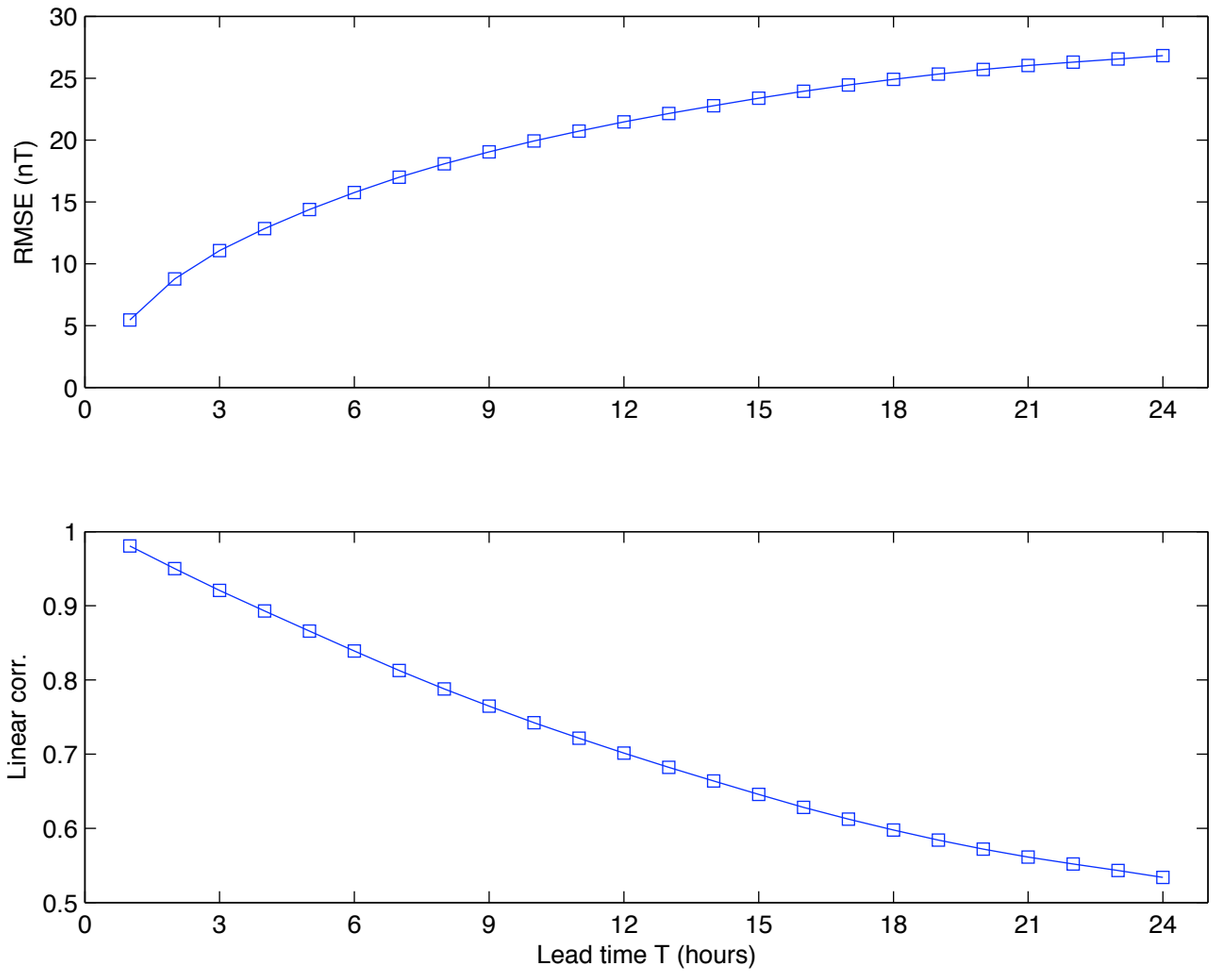
Evaluating lead time using wavelet coherence

Grinsted, et al., Application of the cross wavelet transform and wavelet coherence to geophysical time series, Nonlinear processes in geophysics, 11, 561–566, 2004.

Wintoft et al, 2007, Submitted to Space Weather Journal.

Evaluating lead time using wavelet coherence

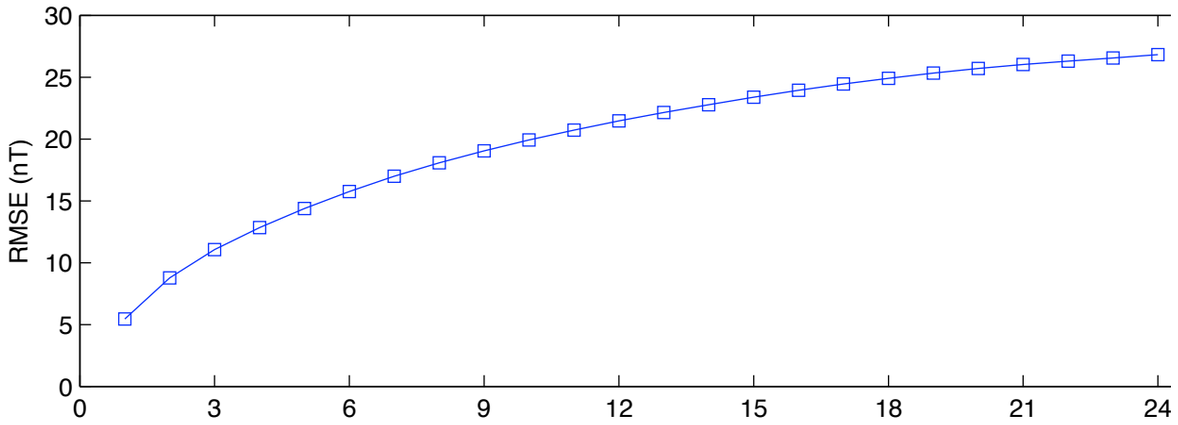
Grinsted, et al., Application of the cross wavelet transform and wavelet coherence to geophysical time series, Nonlinear processes in geophysics, 11, 561–566, 2004.

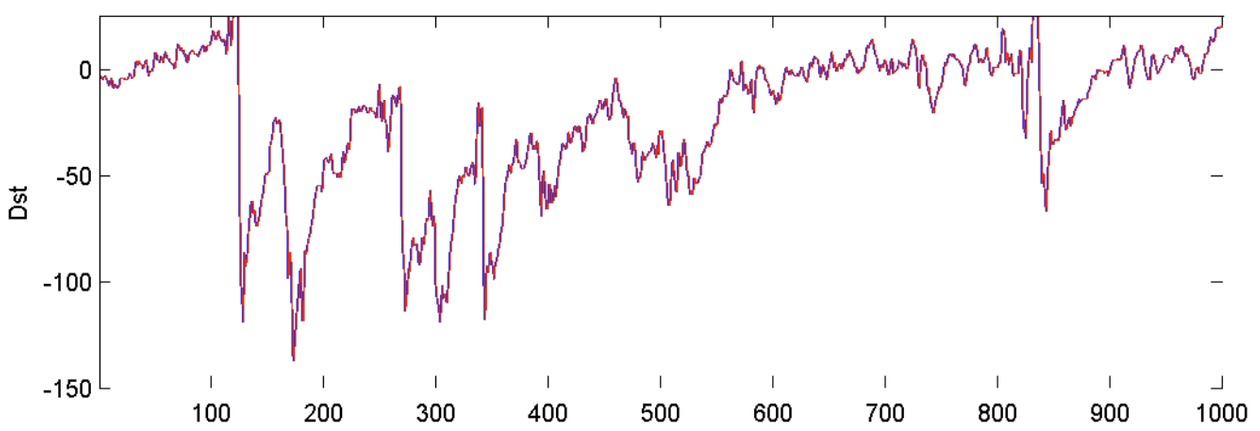


Wintoft et al, 2007, Submitted to Space Weather Journal.

Evaluating lead time using wavelet coherence

Grinsted, et al., Application of the cross wavelet transform and wavelet coherence to geophysical time series, Nonlinear processes in geophysics, 11, 561–566, 2004.

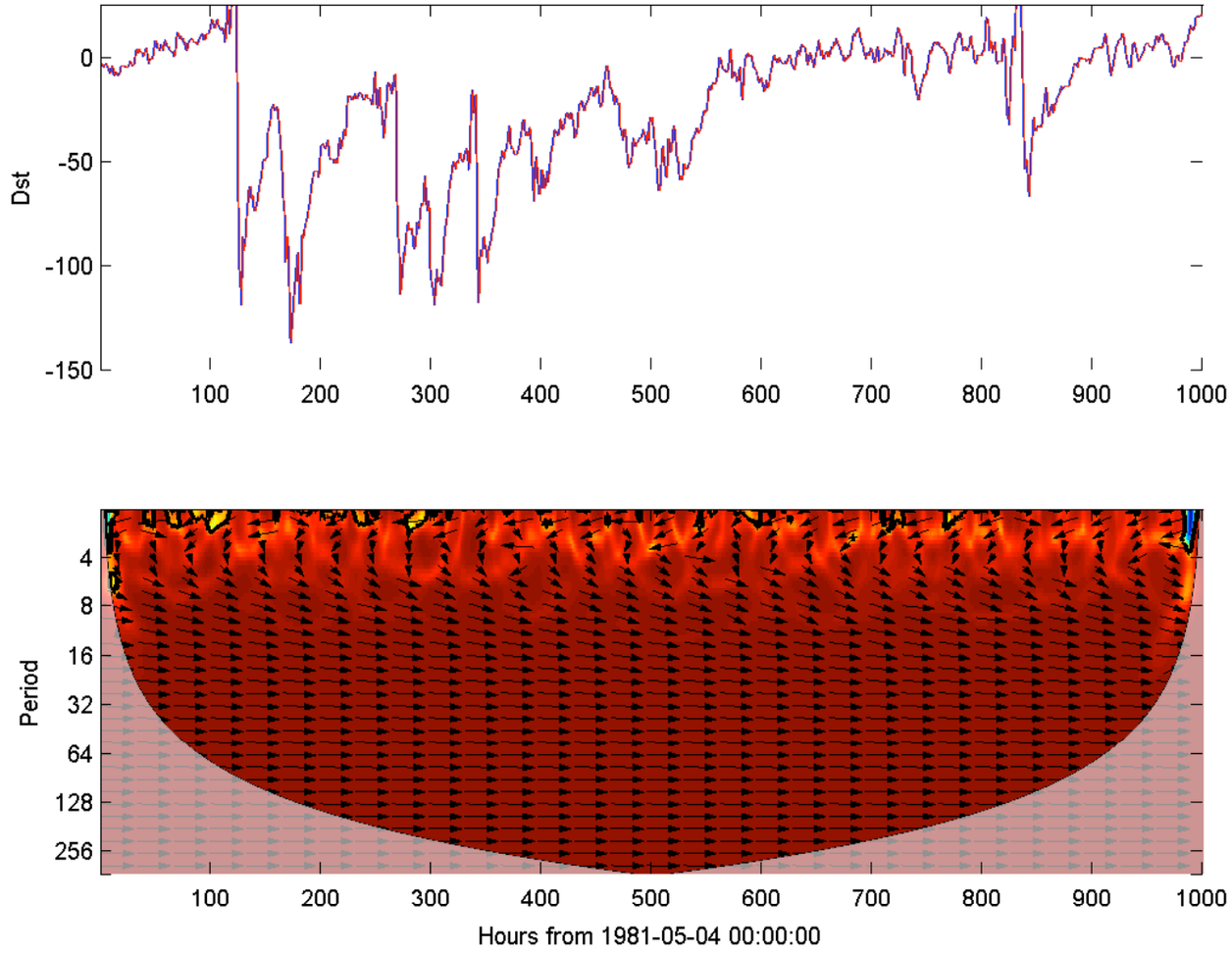




Wintoft et al, 2007, Submitted to Space Weather Journal.

Evaluating lead time using wavelet coherence

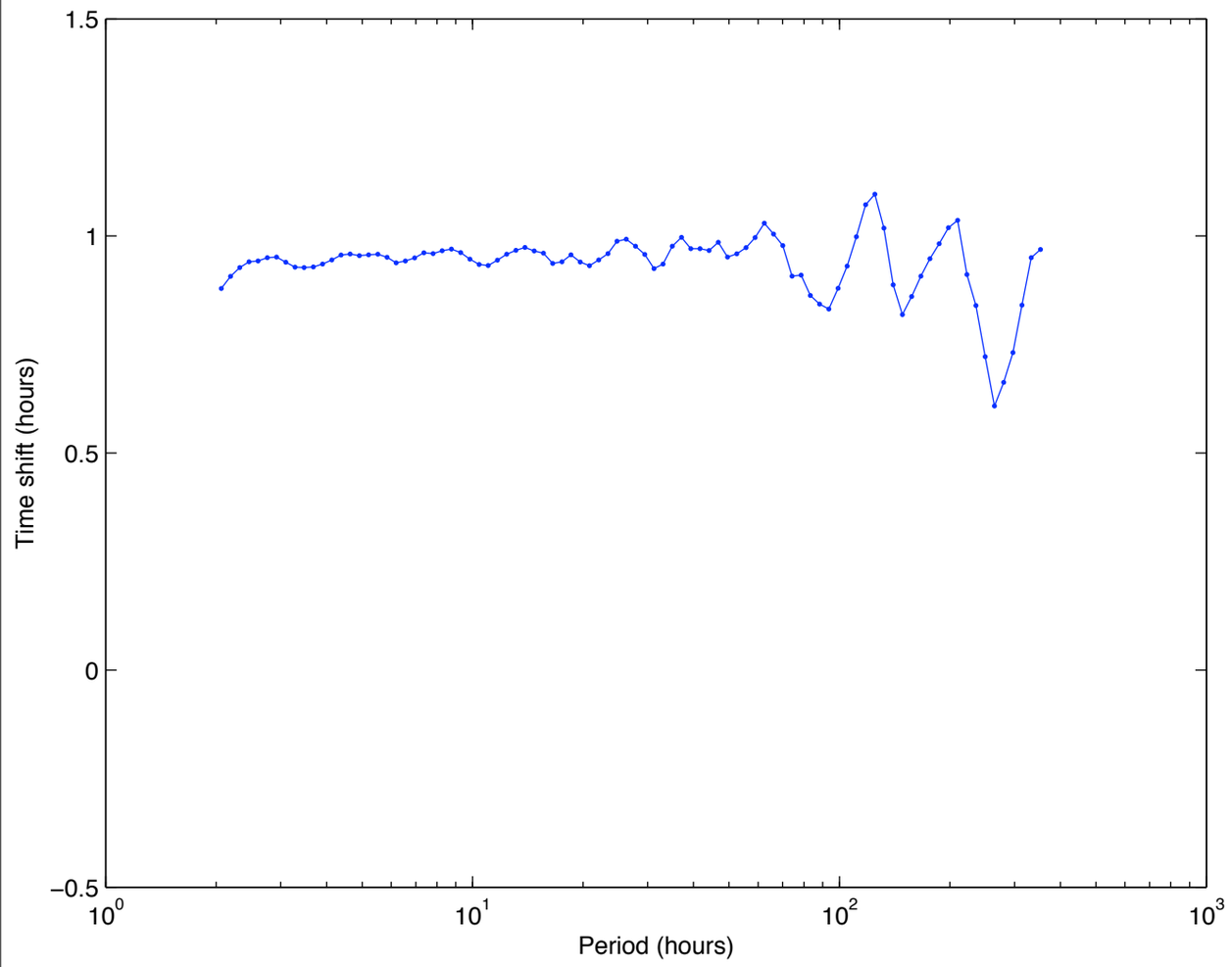
Grinsted, et al., Application of the cross wavelet transform and wavelet coherence to geophysical time series, Nonlinear processes in geophysics, 11, 561–566, 2004.

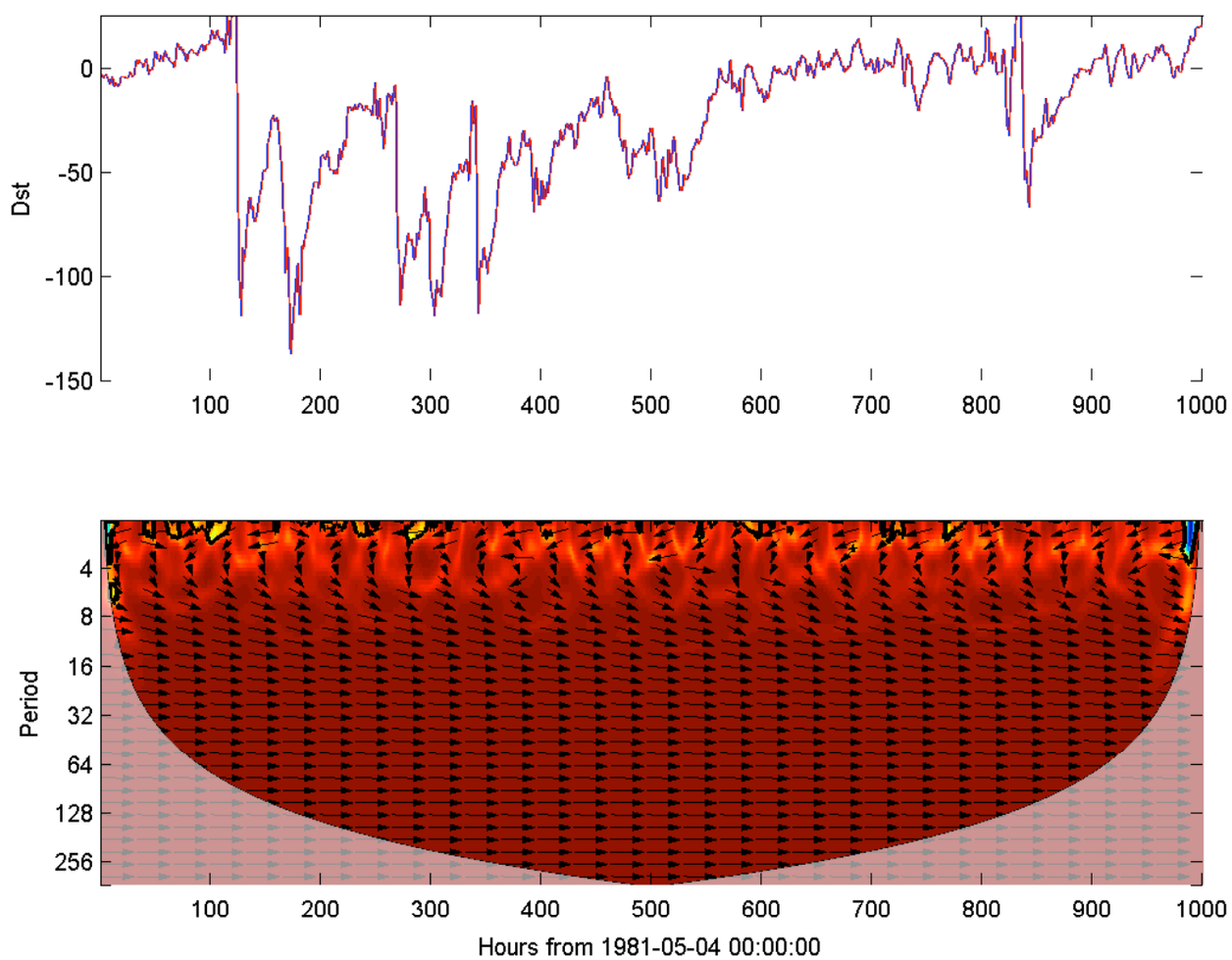


Wintoft et al, 2007, Submitted to Space Weather Journal.

Evaluating lead time using wavelet coherence

Grinsted, et al., Application of the cross wavelet transform and wavelet coherence to geophysical time series, Nonlinear processes in geophysics, 11, 561–566, 2004.

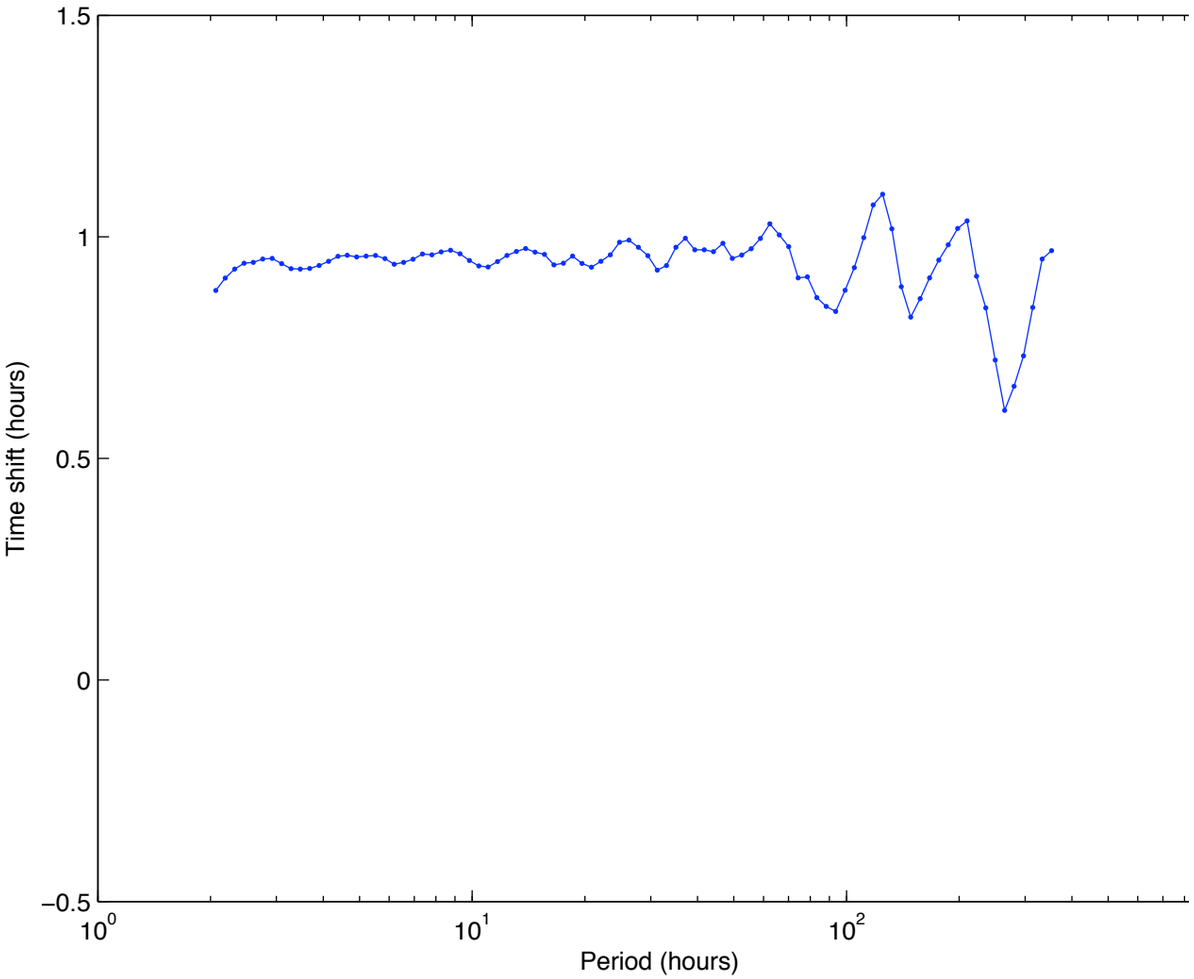


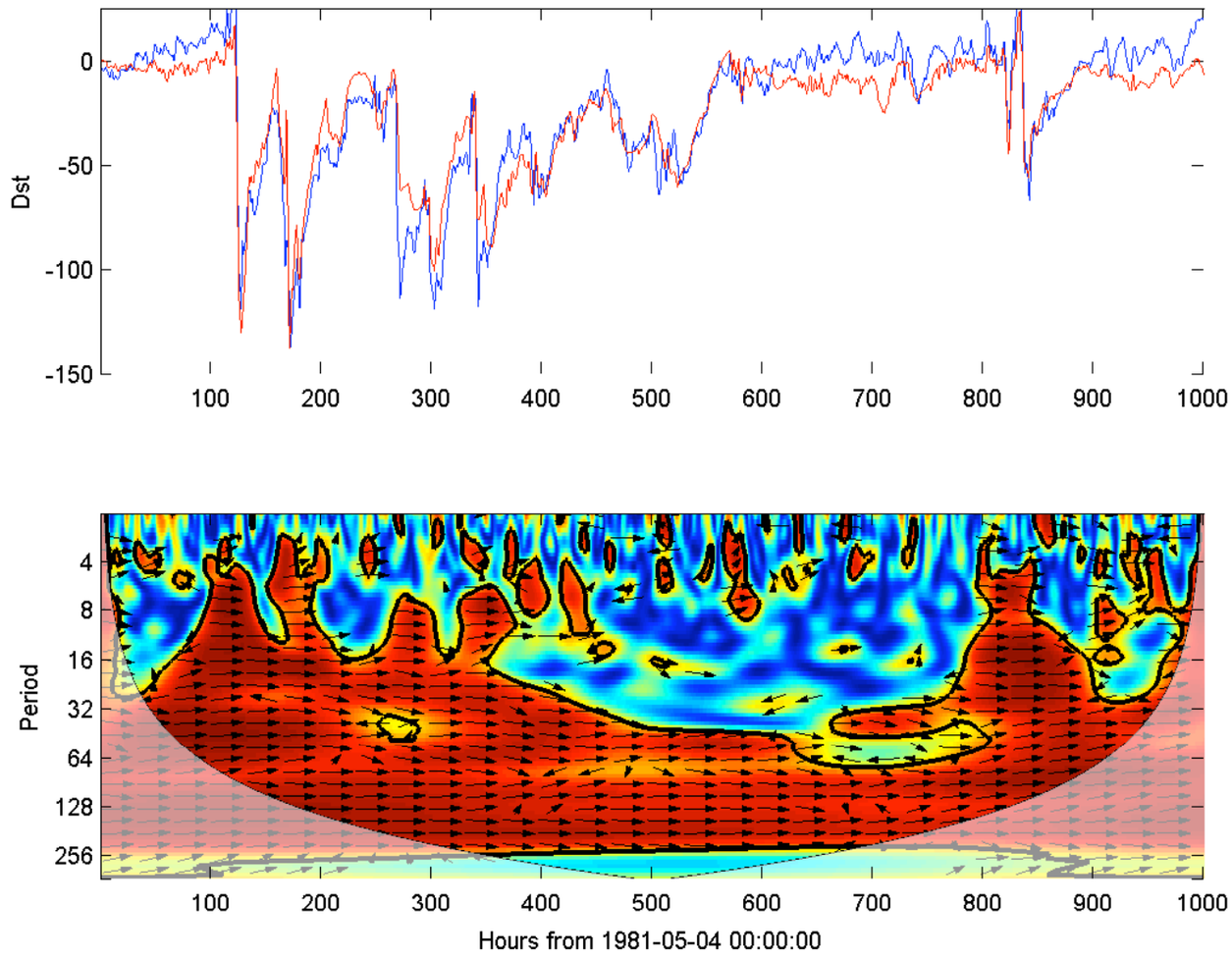


Wintoft et al, 2007, Submitted to Space Weather Journal.

Evaluating lead time using wavelet coherence

Grinsted, et al., Application of the cross wavelet transform and wavelet coherence to geophysical time series, Nonlinear processes in geophysics, 11, 561–566, 2004.

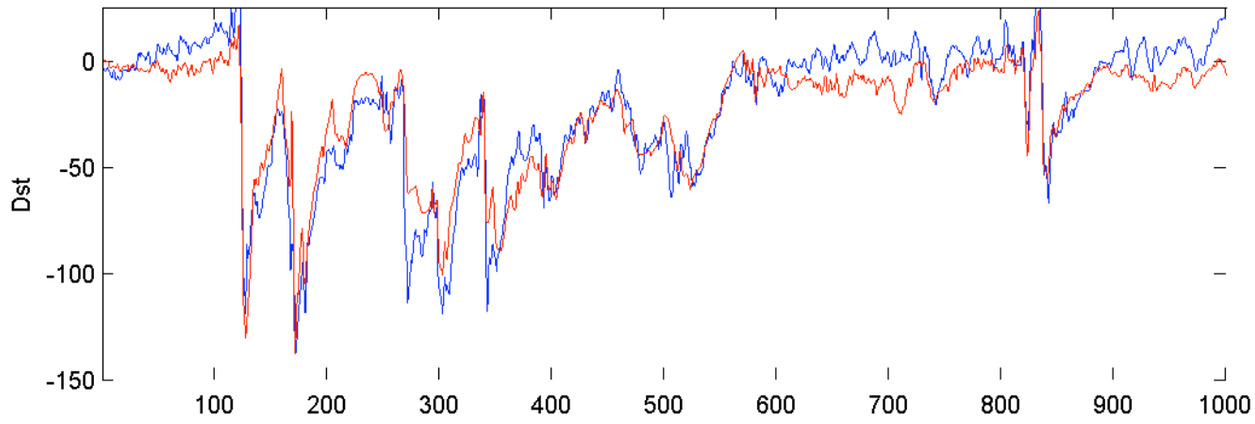


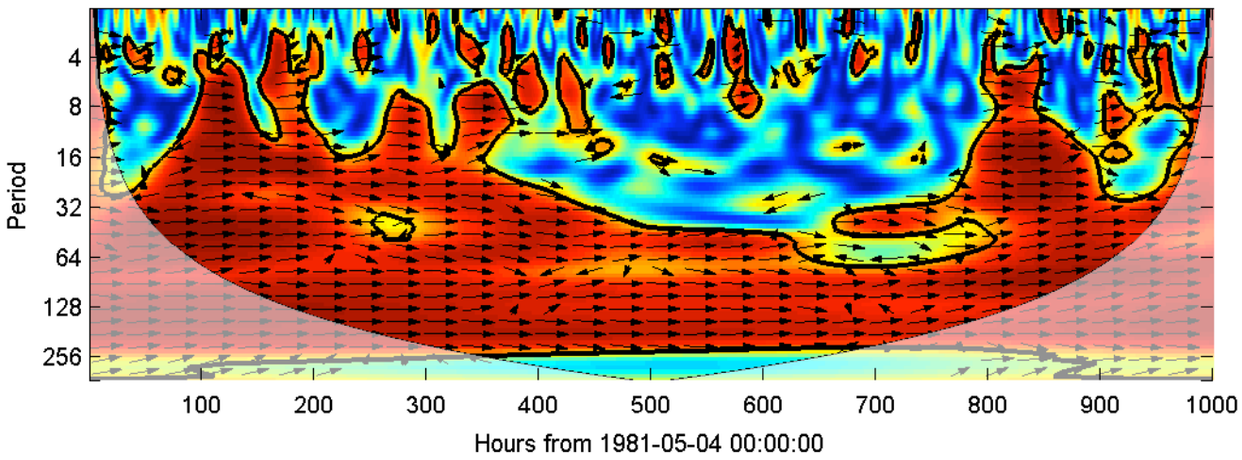


Wintoft et al, 2007, Submitted to Space Weather Journal.

Evaluating lead time using wavelet coherence

Grinsted, et al., Application of the cross wavelet transform and wavelet coherence to geophysical time series, Nonlinear processes in geophysics, 11, 561–566, 2004.

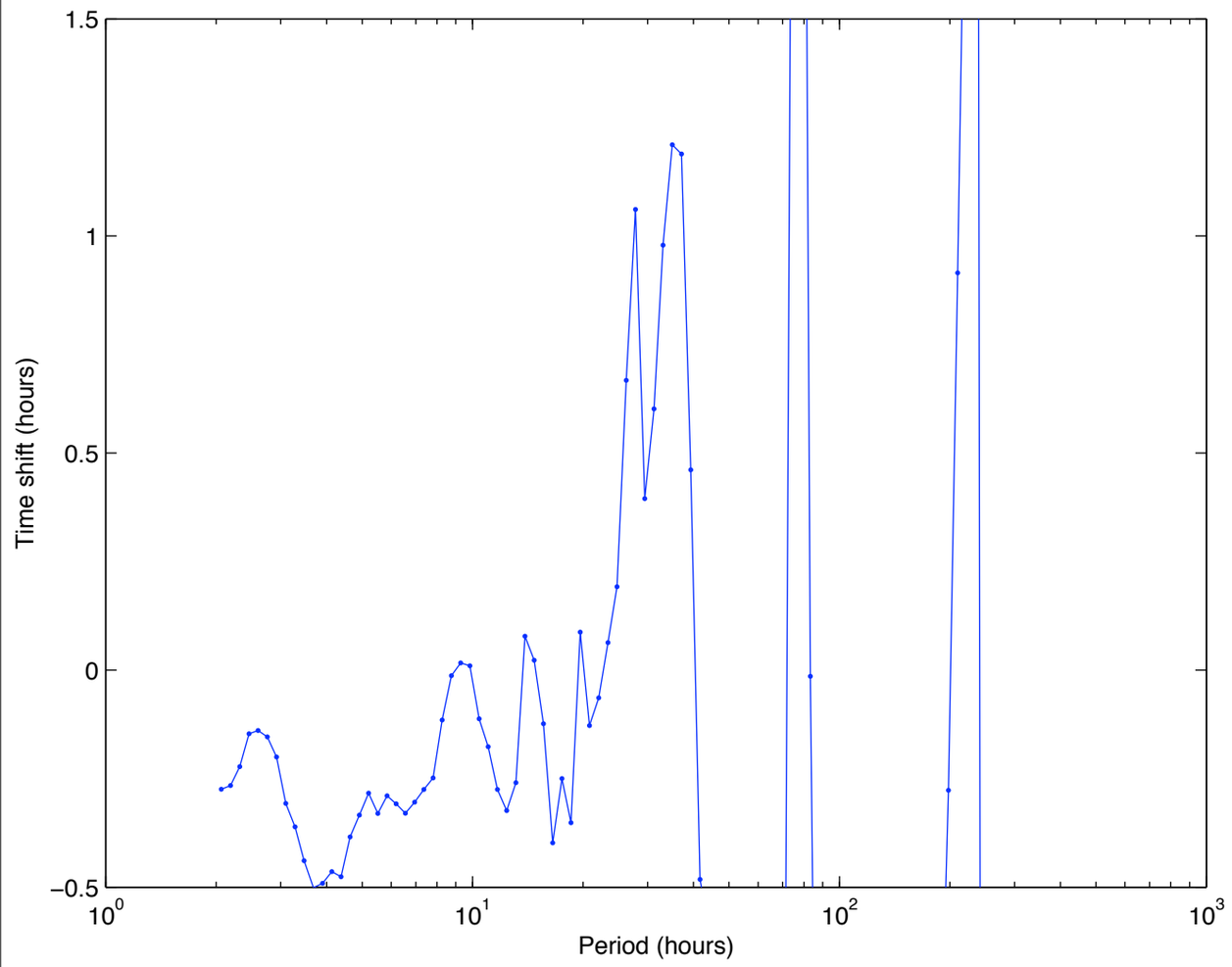


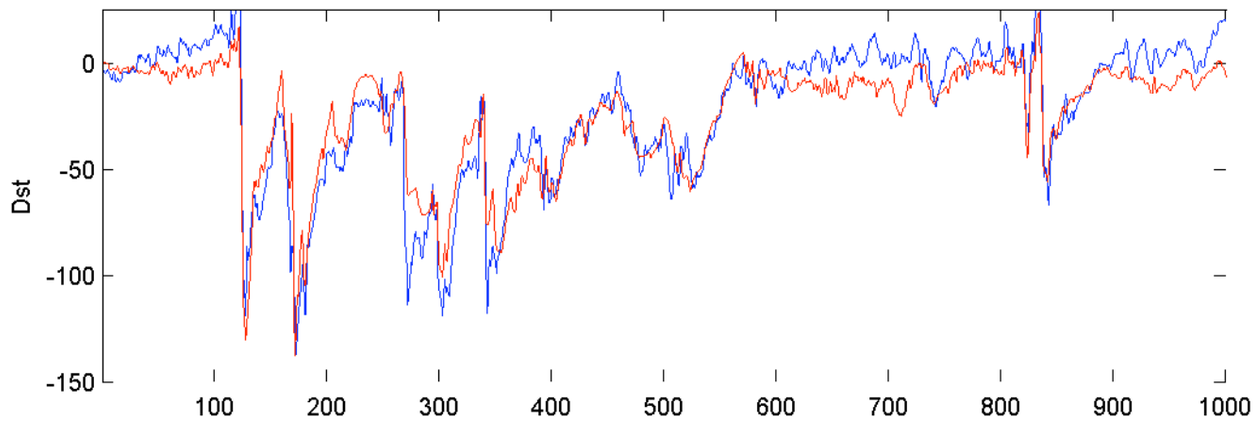


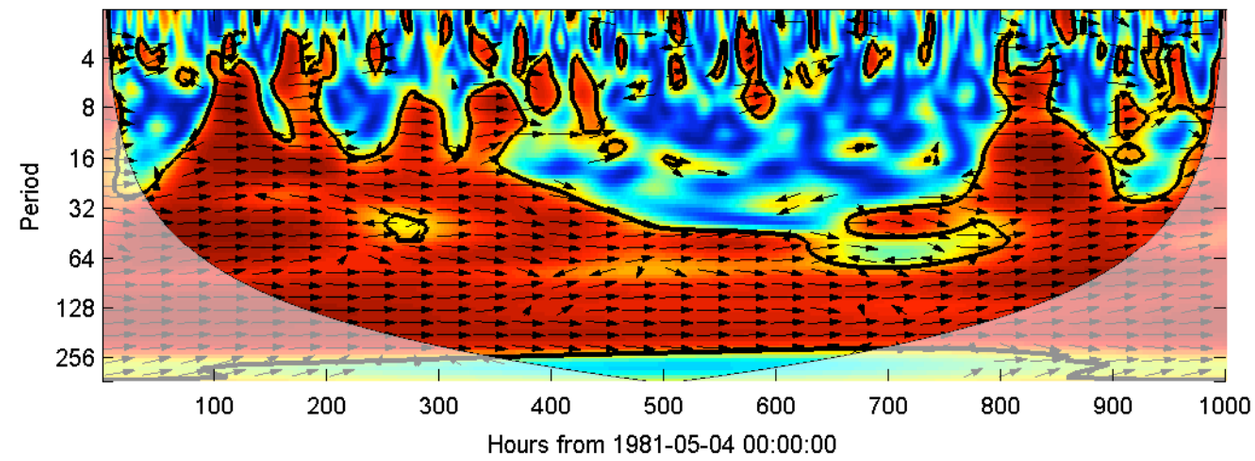
Wintoft et al, 2007, Submitted to Space Weather Journal.

Evaluating lead time using wavelet coherence

Grinsted, et al., Application of the cross wavelet transform and wavelet coherence to geophysical time series, Nonlinear processes in geophysics, 11, 561–566, 2004.







Wintoft et al, 2007, Submitted to Space Weather Journal.

Operational models (publicly available)

- Most models are empirical!
- Types of models:
  - Physical reasoning together with certain assumptions, however, still some data fitting.
  - Data driven models (statistical, polynomial, neural network).

Models predicting indices

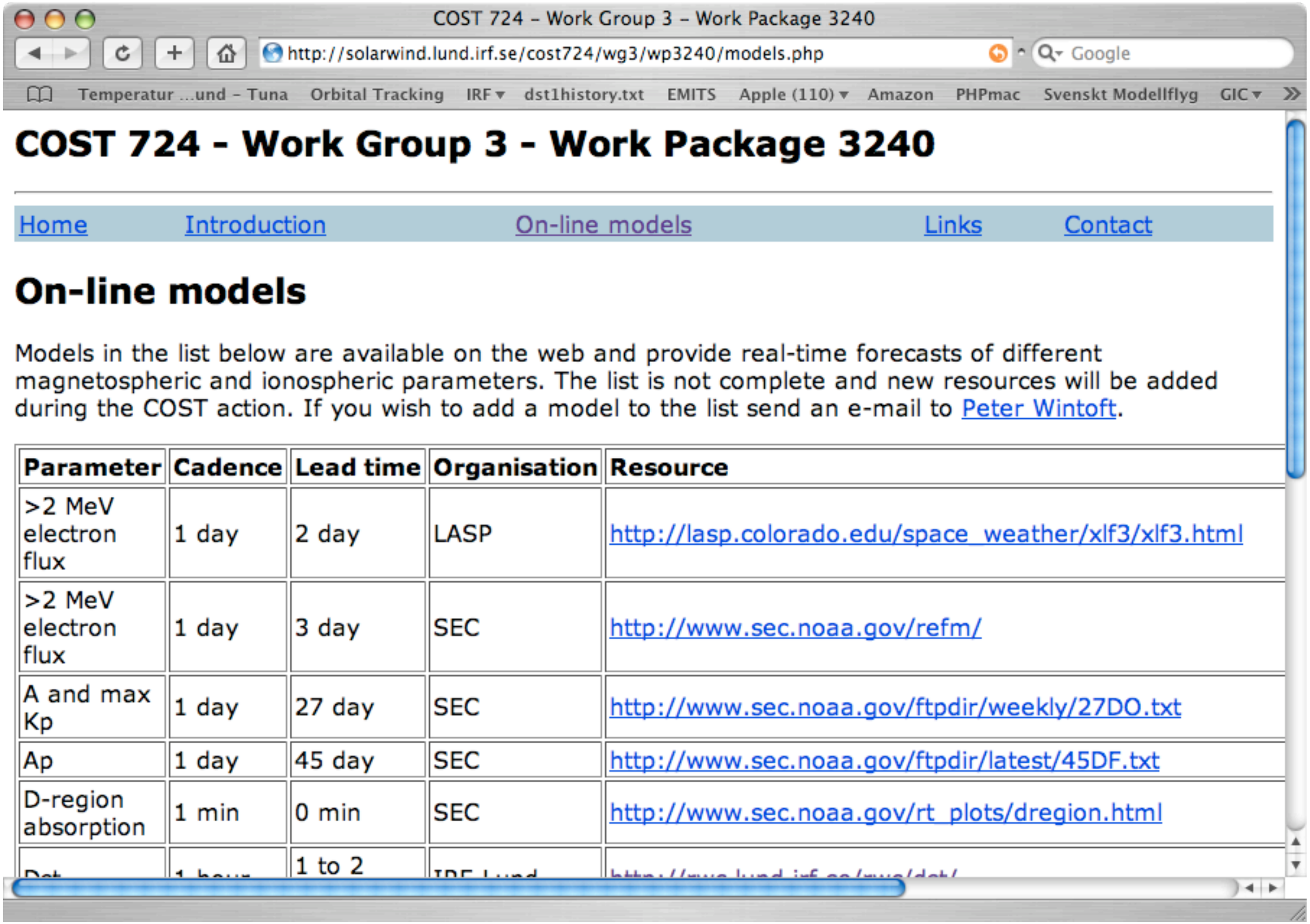
- AE (GIFINT) Pallocchia et al., 2006 [Elman neural network].
- Dst (RWC-Sweden) Lundstedt et al., 2002 [Elman neural network].
- Dst (GIFINT) Pallocchia et al., 2006 [Elman neural network].
- Dst (LASP) Temerin and Li, 2006 [Complex trial-and-error].
- Kp (SEC) Costello, 1997 [Neural network].
- Kp (RWC-Sweden) Boberg et al., 2000 [Neural network].
- Kp (UPOS) Wing et al., 2005 [Recurrent neural network].
- Kp (BAS) Kutiev, 2007 [].
- Maximum daily Kp (SEC) [Statistical].
- Pseudo K (RWC-Canada) (Lam, 2006) [Statistical].
- Ap (SEC) [Statistical].
- Ap (GIFS) Thomson et al., 1993 [ARIMA].

Models for observed data

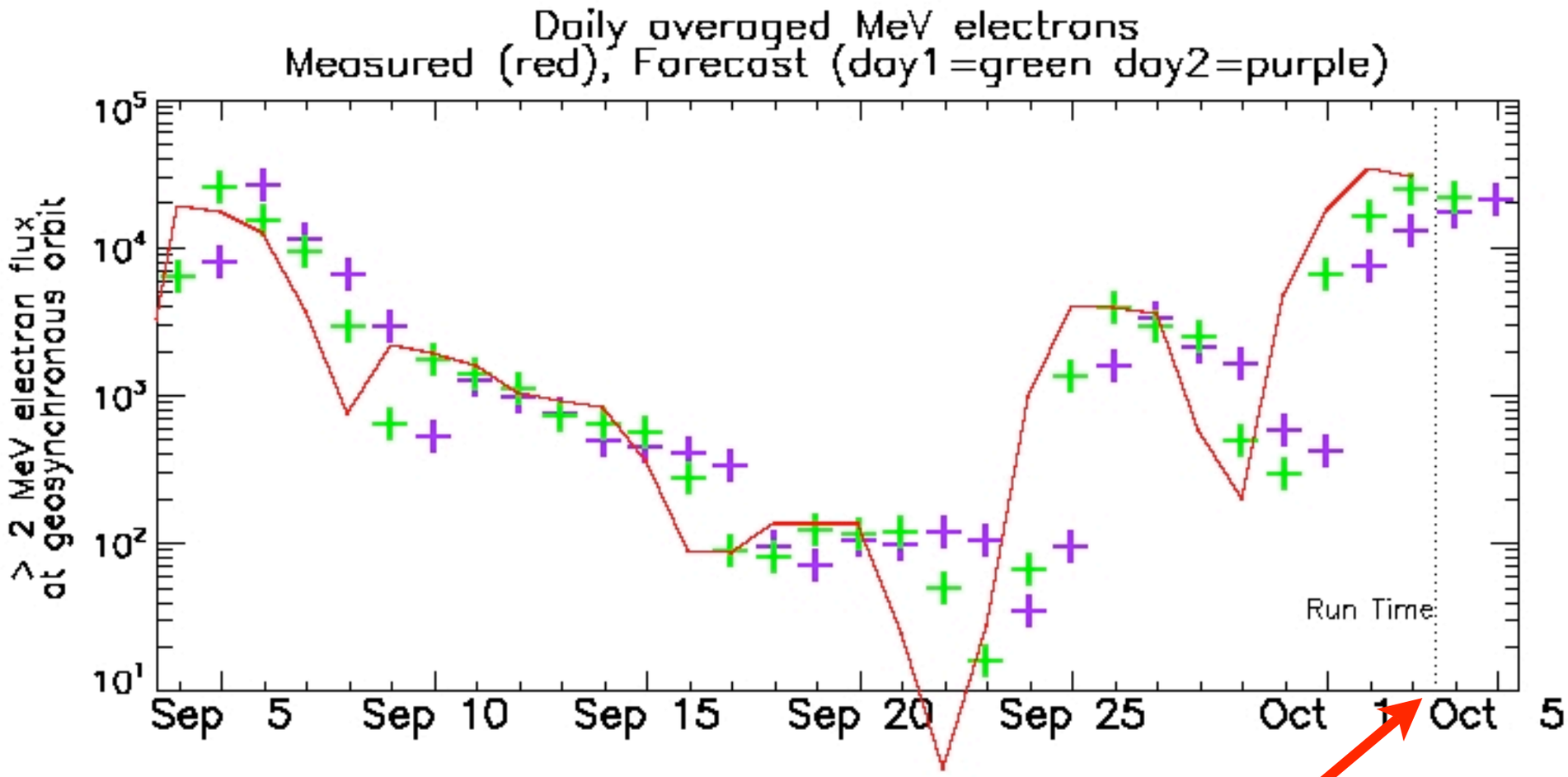
Indices have their limitations (Baumjohann, 1986; Campbell, 2004; Kamide and Rostoker, 2004; Lam, 2006) and sometimes there is no index that capture the process under study.

- 10-minute RMS |dB/dt| (RWC-Sweden), Wintoft et al., 2005 [Elman neural network].
- foF2 (SEC) Fuller-Rowell et al., 2001 [3rd order polynomial].
- foF2 (RAL) Muhtarov et al., 2002 [ARMA].
- D-region absorption (SEC) [].
- MUF (DIFS) Butcher, 2005 [Secret].
- > 2 MeV flux (LASP) Li et al., 2001 [Radial diffusion model].
- > 2 MeV flux (SEC) Baker et al., 1990 [Linear filter].

http://solarwind.lund.irf.se/cost724/wg3/wp3240

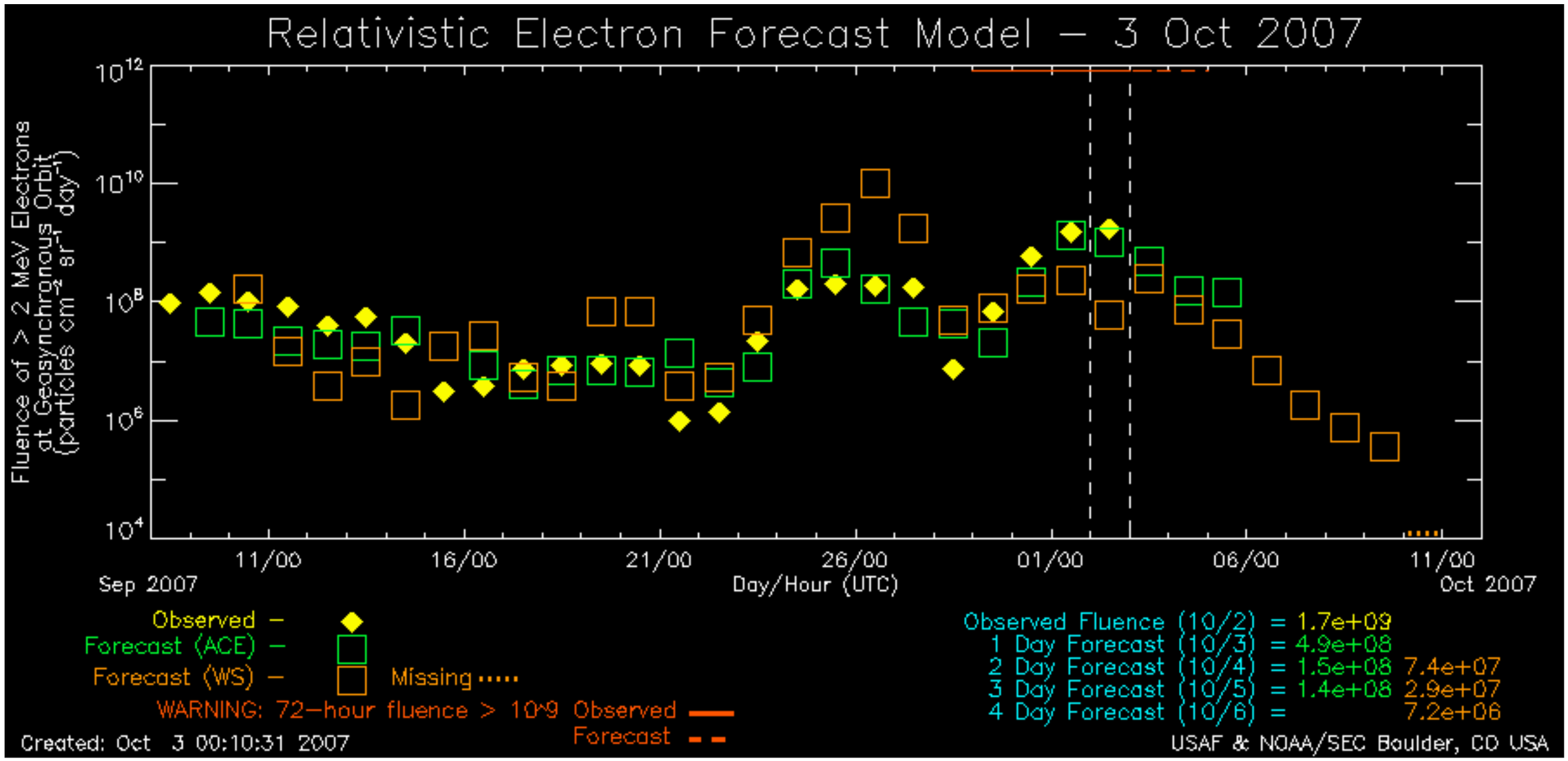


LASP >2 MeV forecast

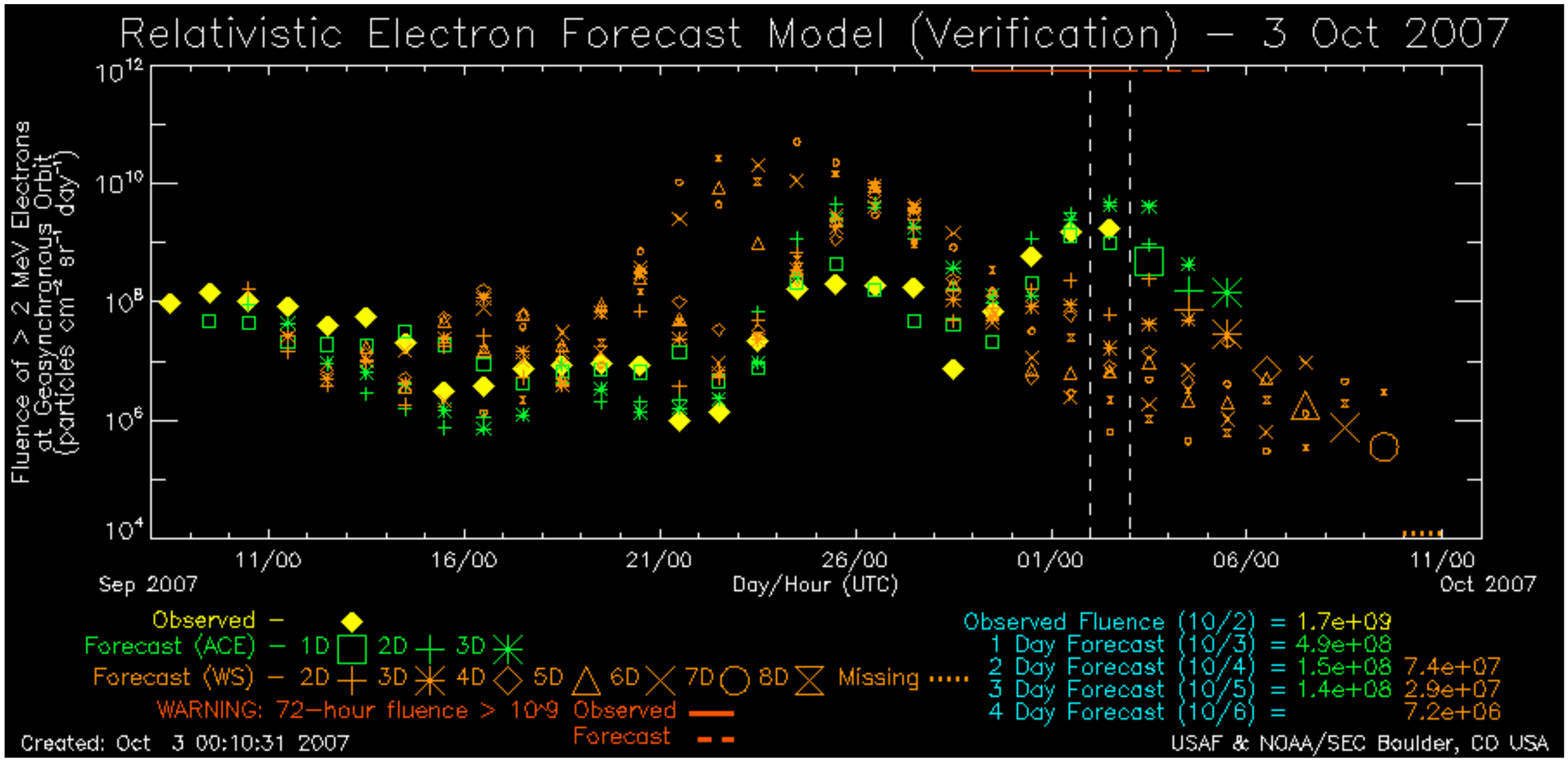


Forecast issued on 3 Oct 2007

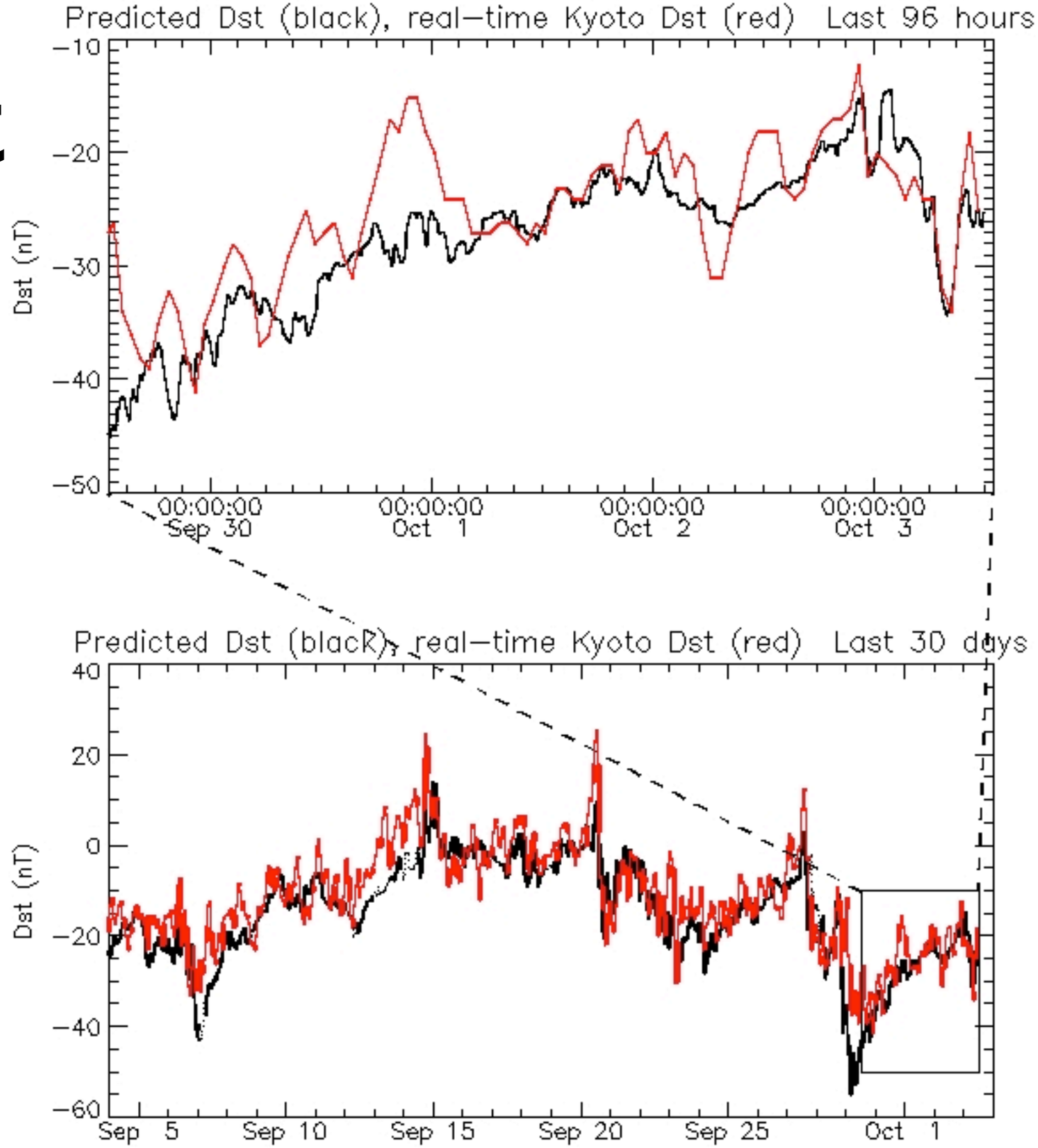
SEC >2 Mev forecast



SEC >2 Mev forecast



LASP Dst


Predicted Dst (black), real–time Kyoto Dst (red) Last 96 hours
Dst (nT)
−10
−20
−30
−40
−50
00:00:00 Sep 30
00:00:00 Oct 1
00:00:00 Oct 2
00:00:00 Oct 3
Predicted Dst (black), real–time Kyoto Dst (red) Last 30 days
Dst (nT)
40
20
0
−20
−40
−60
Sep 5
Sep 10
Sep 15
Sep 20
Sep 25
Oct 1

Geomagnetic Activity

http://www.geophys.bas.bg/kp_for/kp_mod.htm

Temperatur ...und - Tuna | Orbital Tracking | IKF | dst1history.txt | EMITS | Apple (110) | Amazon | PHPmac | Svenskt Modellflyg | GIC | eBay

GPhI
GEOPHYSICAL INSTITUTE-BAS

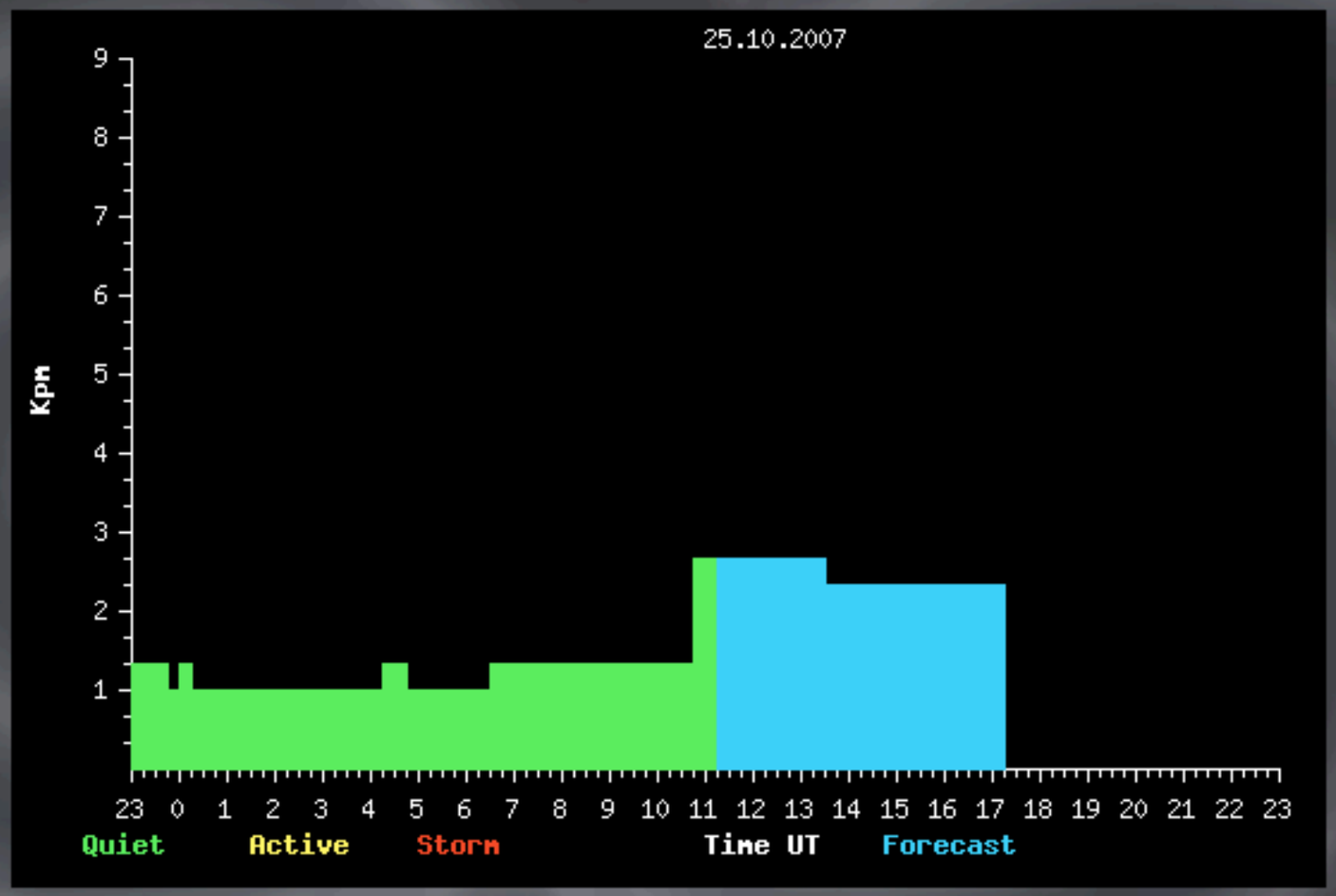
Geomagnetic Activity Today

| Kpm-index | Values | Yesterday |

Kpm more from 5
STORM

Kpm from 4 to 5
ACTIVE

Kpm less then 4
QUIET



View History

Year	Month	Day	
2007	01	01	View

Lund Dst model

Lund Dst model

Data analysis and selection (OMNI, NGDC, ACE)

Lund Dst model

Data analysis and selection (OMNI, NGDC, ACE)

Optimisation of neural network (Lundstedt et al., GRL, 2002)

Lund Dst model

Data analysis and selection (OMNI, NGDC, ACE)

Optimisation of neural network (Lundstedt et al., GRL, 2002)

Implementation

Lund Dst model

Data analysis and selection (OMNI, NGDC, ACE)

Optimisation of neural network (Lundstedt et al., GRL, 2002)

Implementation

UNIX cron executes a UNIX script every 10 minutes

Lund Dst model

Data analysis and selection (OMNI, NGDC, ACE)

Optimisation of neural network (Lundstedt et al., GRL, 2002)

Implementation

UNIX cron executes a UNIX script every 10 minutes

ACE solar wind data (MySQL, Java) → Solar wind (ASCII)

Lund Dst model

Data analysis and selection (OMNI, NGDC, ACE)

Optimisation of neural network (Lundstedt et al., GRL, 2002)

Implementation

UNIX cron executes a UNIX script every 10 minutes

ACE solar wind data (MySQL, Java) → Solar wind (ASCII)

NN model (Java) → Dst forecast (ASCII)

Lund Dst model

Data analysis and selection (OMNI, NGDC, ACE)

Optimisation of neural network (Lundstedt et al., GRL, 2002)

Implementation

UNIX cron executes a UNIX script every 10 minutes

ACE solar wind data (MySQL, Java) → Solar wind (ASCII)

NN model (Java) → Dst forecast (ASCII)

Text formatting (Perl) → 7-day, 24-hour, and real time data (ASCII)

Lund Dst model

Data analysis and selection (OMNI, NGDC, ACE)

Optimisation of neural network (Lundstedt et al., GRL, 2002)

Implementation

UNIX cron executes a UNIX script every 10 minutes

ACE solar wind data (MySQL, Java) → Solar wind (ASCII)

NN model (Java) → Dst forecast (ASCII)

Text formatting (Perl) → 7-day, 24-hour, and real time data (ASCII)

Kyoto n.r.t. data (MySQL, Perl) → Kyoto Dst (ASCII)

Lund Dst model

Data analysis and selection (OMNI, NGDC, ACE)

Optimisation of neural network (Lundstedt et al., GRL, 2002)

Implementation

UNIX cron executes a UNIX script every 10 minutes

ACE solar wind data (MySQL, Java) → Solar wind (ASCII)

NN model (Java) → Dst forecast (ASCII)

Text formatting (Perl) → 7-day, 24-hour, and real time data (ASCII)

Kyoto n.r.t. data (MySQL, Perl) → Kyoto Dst (ASCII)

24-hour plot (GNU Plot) → PNG file

Lund Dst model

Data analysis and selection (OMNI, NGDC, ACE)

Optimisation of neural network (Lundstedt et al., GRL, 2002)

Implementation

UNIX cron executes a UNIX script every 10 minutes

ACE solar wind data (MySQL, Java) → Solar wind (ASCII)

NN model (Java) → Dst forecast (ASCII)

Text formatting (Perl) → 7-day, 24-hour, and real time data (ASCII)

Kyoto n.r.t. data (MySQL, Perl) → Kyoto Dst (ASCII)

24-hour plot (GNU Plot) → PNG file

7-day plot (GNU Plot) → PNG file

Lund Dst model

Data analysis and selection (OMNI, NGDC, ACE)

Optimisation of neural network (Lundstedt et al., GRL, 2002)

Implementation

UNIX cron executes a UNIX script every 10 minutes

ACE solar wind data (MySQL, Java) → Solar wind (ASCII)

NN model (Java) → Dst forecast (ASCII)

Text formatting (Perl) → 7-day, 24-hour, and real time data (ASCII)

Kyoto n.r.t. data (MySQL, Perl) → Kyoto Dst (ASCII)

24-hour plot (GNU Plot) → PNG file

7-day plot (GNU Plot) → PNG file

PHP

Lund Dst model

Data analysis and selection (OMNI, NGDC, ACE)

Optimisation of neural network (Lundstedt et al., GRL, 2002)

Implementation

UNIX cron executes a UNIX script every 10 minutes

ACE solar wind data (MySQL, Java) → Solar wind (ASCII)

NN model (Java) → Dst forecast (ASCII)

Text formatting (Perl) → 7-day, 24-hour, and real time data (ASCII)

Kyoto n.r.t. data (MySQL, Perl) → Kyoto Dst (ASCII)

24-hour plot (GNU Plot) → PNG file

7-day plot (GNU Plot) → PNG file

PHP

Simple
(UNIX script)

Lund Dst model

Data analysis and selection (OMNI, NGDC, ACE)

Optimisation of neural network (Lundstedt et al., GRL, 2002)

Implementation

UNIX cron executes a UNIX script every 10 minutes

ACE solar wind data (MySQL, Java) → Solar wind (ASCII)

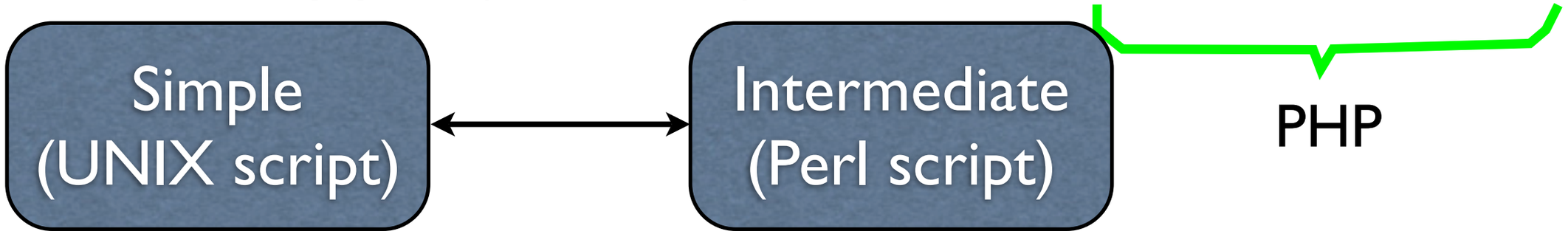
NN model (Java) → Dst forecast (ASCII)

Text formatting (Perl) → 7-day, 24-hour, and real time data (ASCII)

Kyoto n.r.t. data (MySQL, Perl) → Kyoto Dst (ASCII)

24-hour plot (GNU Plot) → PNG file

7-day plot (GNU Plot) → PNG file



Lund Dst model

Data analysis and selection (OMNI, NGDC, ACE)

Optimisation of neural network (Lundstedt et al., GRL, 2002)

Implementation

UNIX cron executes a UNIX script every 10 minutes

ACE solar wind data (MySQL, Java) → Solar wind (ASCII)

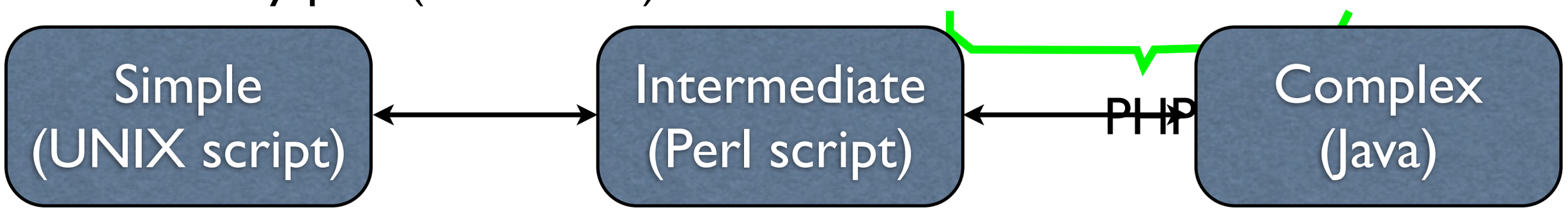
NN model (Java) → Dst forecast (ASCII)

Text formatting (Perl) → 7-day, 24-hour, and real time data (ASCII)

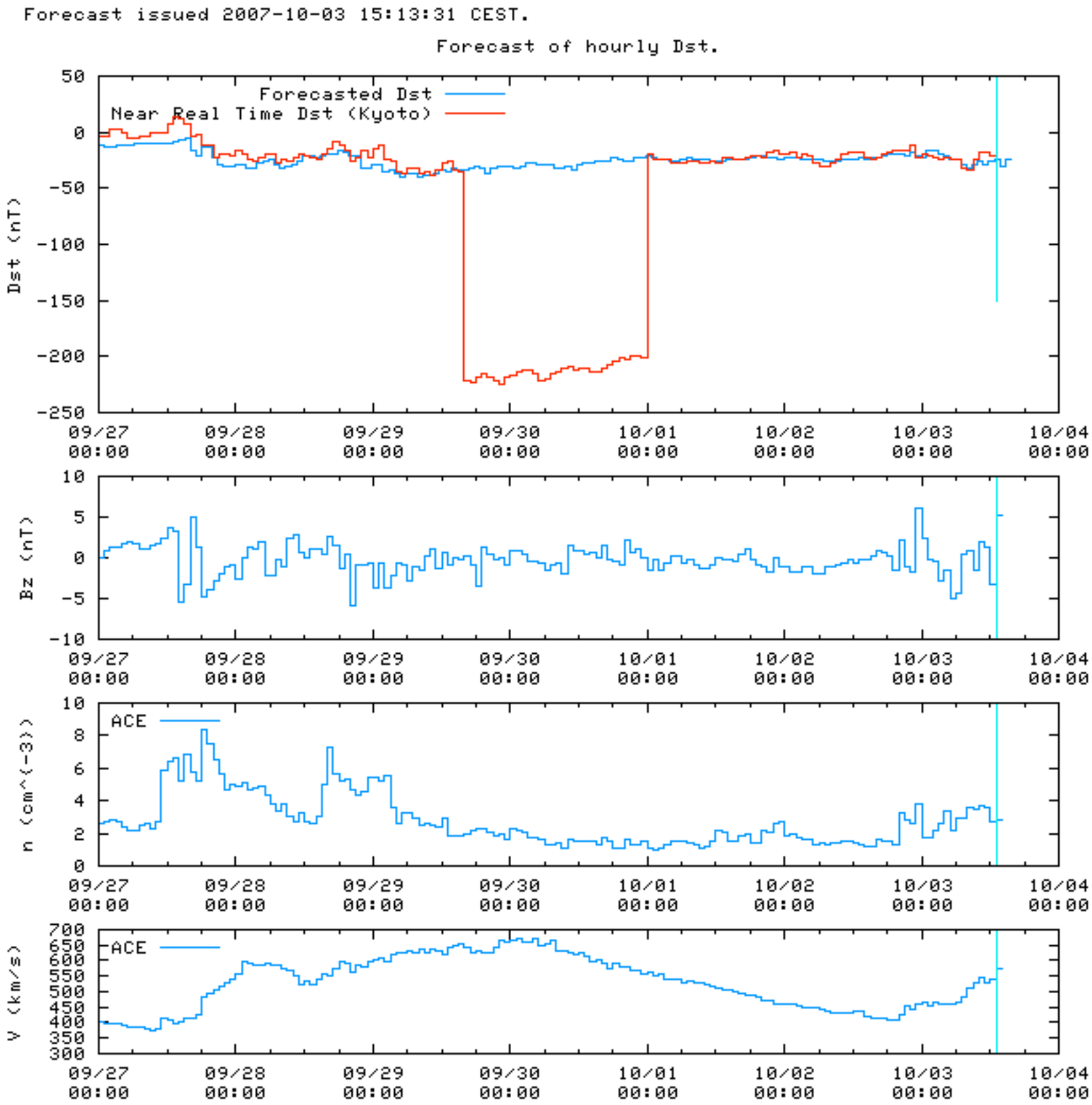
Kyoto n.r.t. data (MySQL, Perl) → Kyoto Dst (ASCII)

24-hour plot (GNU Plot) → PNG file

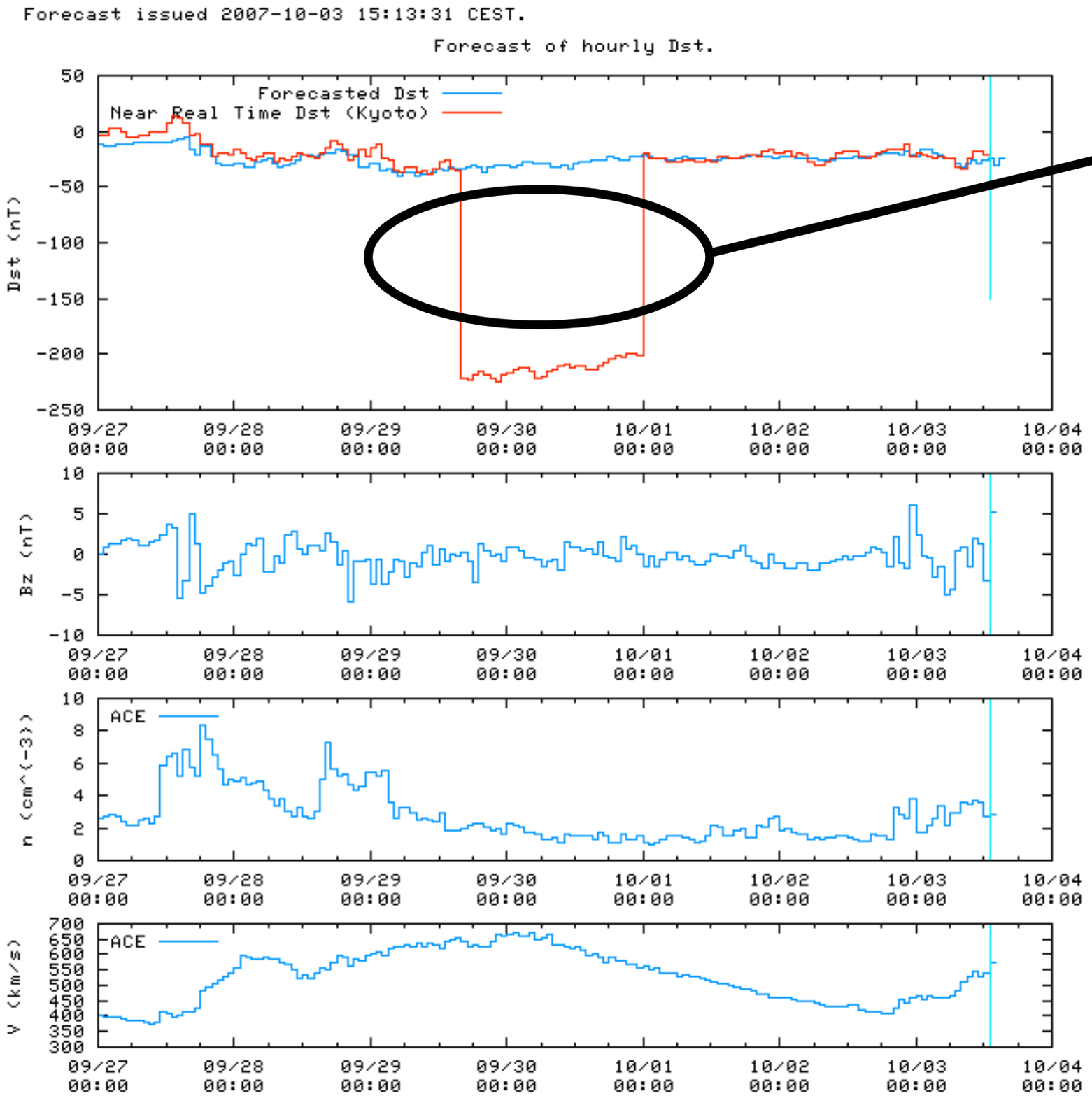
7-day plot (GNU Plot) → PNG file



Lund Dst

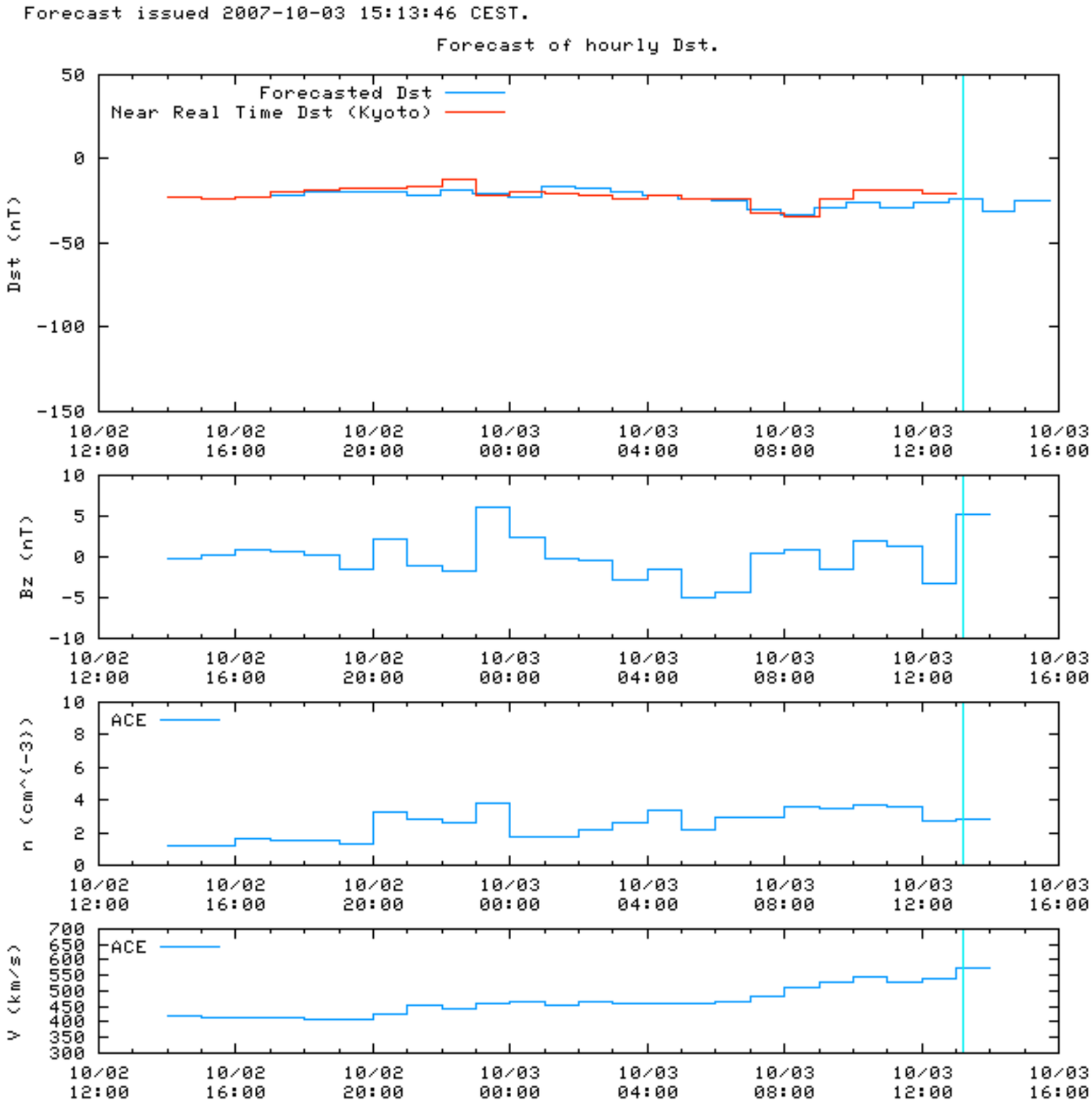

Forecast issued 2007-10-03 15:13:31 CEST.
Forecast of hourly Dst.
Forecasted Dst
Near Real Time Dst (Kyoto)
Dst (nT)
50
0
-50
-100
-150
-200
-250
Bz (nT)
10
5
0
-5
-10
n (cm^(-3))
ACE
10
8
6
4
2
0
V (km/s)
ACE
700
650
600
550
500
450
400
350
300
09/27 00:00
09/28 00:00
09/29 00:00
09/30 00:00
10/01 00:00
10/02 00:00
10/03 00:00
10/04 00:00

Lund Dst



Real time data problem during download to IRF-Lund database

Lund Dst close look



Lund Dst data

http://rwc.lund.irf.se/rwc/dst/dst1.txt

http://rwc.lund.irf.se/rwc/dst/dst1.txt

Google

Temperatur ...und - Tuna | Orbital Tracking | IRF | dst1history.txt | EMITS | Apple (110) | Amazon | PHPmac

```
2007-10-25 12:10:00 -13.271944030236398
```

Real time Dst forecast

http://rwc.lund.irf.se/rwc/dst/data.php

Google

Temperatur ...und - Tuna | Orbital Tracking | IRF | dst1history.txt | EMITS | Apple (110) | Amazon | PHPmac

Real time Dst forecast

Last 7 days | Last 24 hours | Archive | Data | Models

Solar wind input data

YYYY-MM-DD	HH:MM:SS	Bz	n	V
2007-10-19	00:00:00	2.6	3.2	588
2007-10-19	01:00:00	1.0	3.9	590
2007-10-19	02:00:00	3.4	4.2	589
2007-10-19	03:00:00	2.6	4.0	591
2007-10-19	04:00:00	2.0	4.7	593
2007-10-19	05:00:00	3.2	4.4	599
2007-10-19	06:00:00	1.6	3.3	630

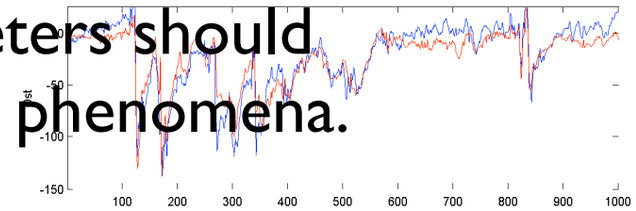
Dst forecast

YYYY-MM-DD	HH:MM:SS	Dst
2007-10-19	00:44:00	-32
2007-10-19	01:42:00	-27
2007-10-19	02:42:00	-26
2007-10-19	03:42:00	-22
2007-10-19	04:42:00	-20
2007-10-19	05:42:00	-20

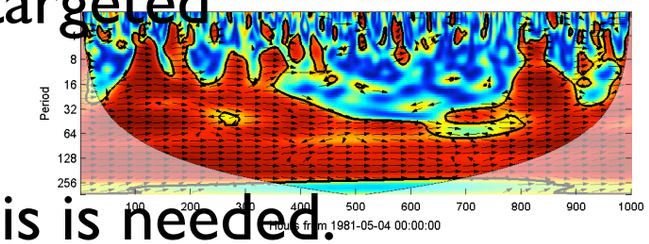
http://www.lund.irf.se/rwc/

Summary

- Collecting data and building the database for model development is time consuming.
- Often models need to be adopted to handle real time data.
- Models forecasting indices or general physical parameters should aim at lead times comparable to the time scale of the phenomena.

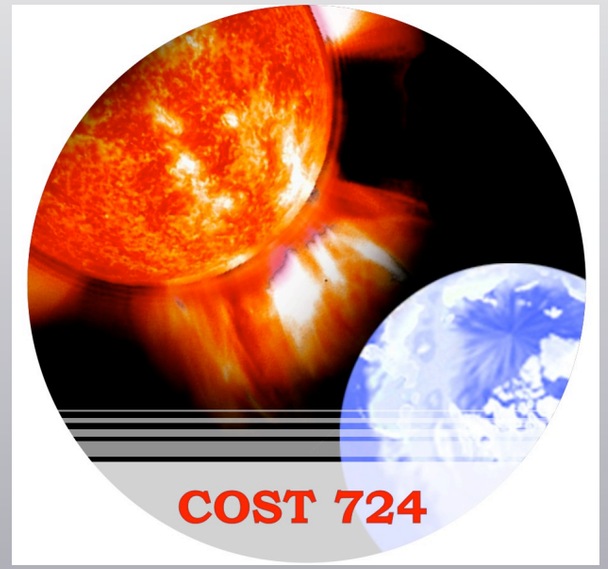


- Models forecasting with shorter lead times must be targeted towards parameters specified by the end users.

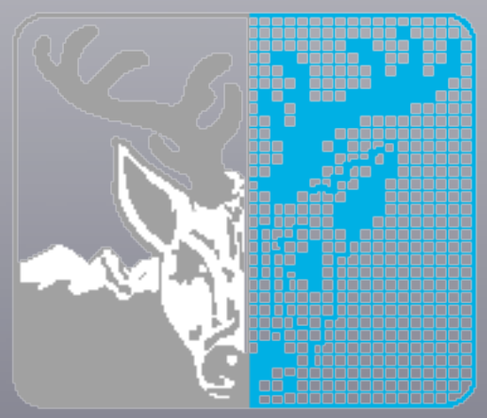


- To properly evaluate forecast lead time careful analysis is needed.
- Implementing for real time operation is a very important test of the model.
- The Time Series Object (TSO) is a useful concept for transferring time series data.
- Plots are essential to get overall view (end user) ...
- but data is sometimes more useful (service provider).

Acknowledgement



COST 724



High Performance Computing
Center North (HPC2N)

Daniel Heynderickx organising this session.

Ronald Van der Linden and the organisers of ESWW 4.