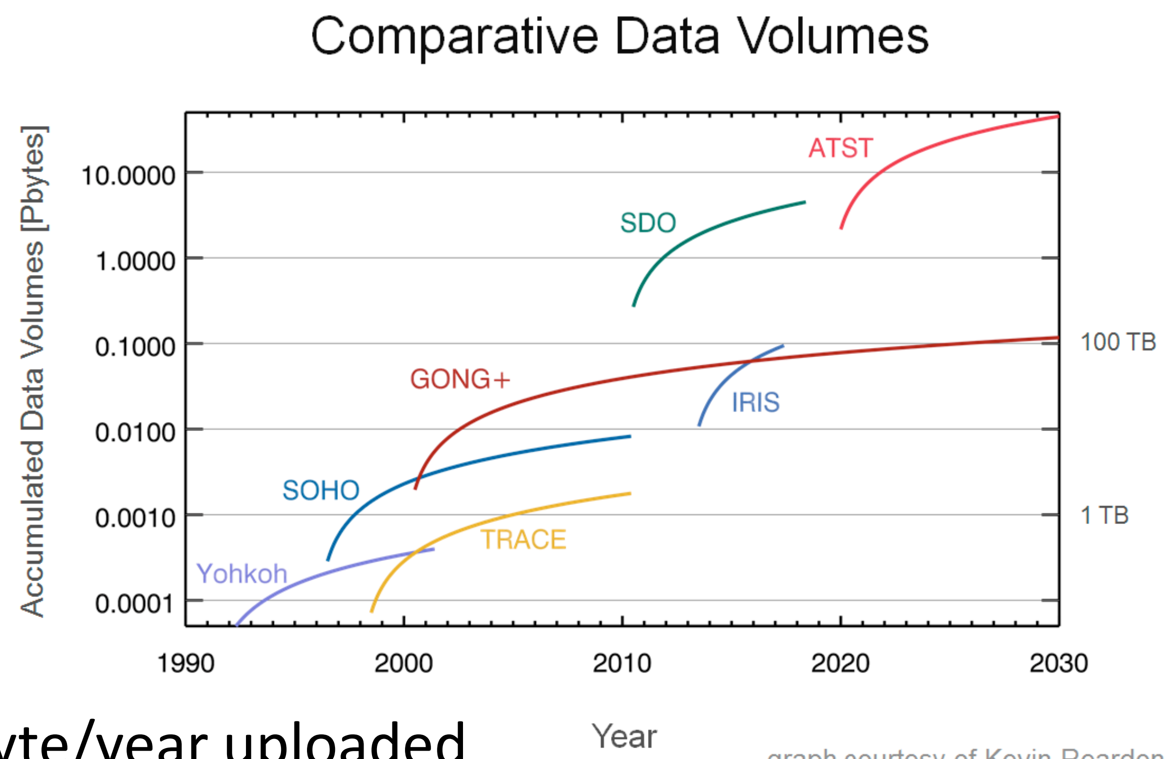
A neural-network-based reconstruction of spatio-temporal datasets with application to the near-Earth space environment

Jacob Bortnik, UCLA

Wen Li, Richard Thorne, Chao Yue, Xiangning Chu, Vassilis Angelopoulos, Lauren Blum, Qianli Ma, Craig Kletzing, Geoff Reeves, Harlan Spence Shri Kanekal, Dan Baker

Reference: Bortnik, J., W. Li, R. M. Thorne, and V. Angelopoulos (2016), A unified approach to inner magnetospheric state prediction, J. Geophys. Res., 121, 2423–2430, doi:10.1002/ 2015JA021733.

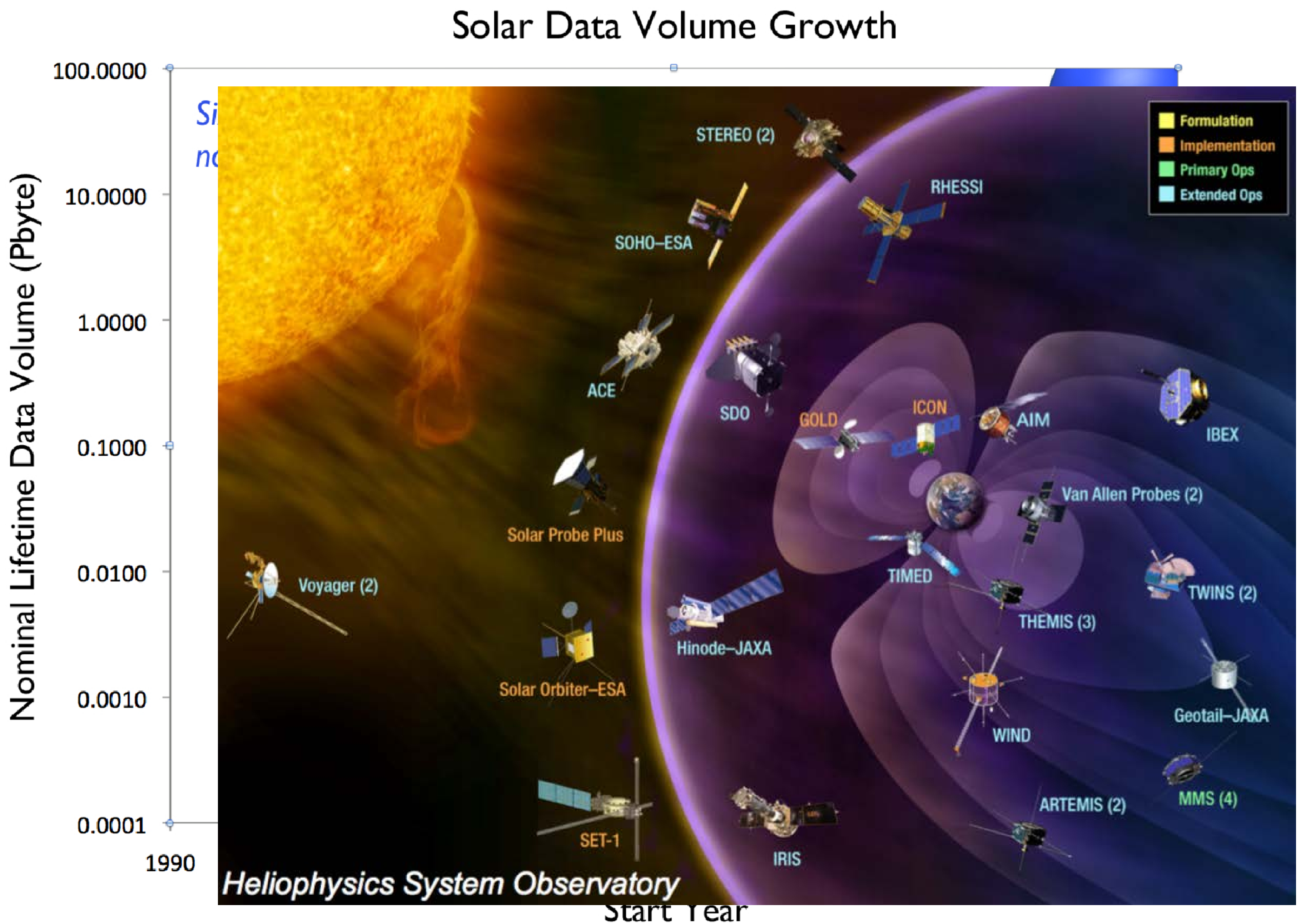
Motivation



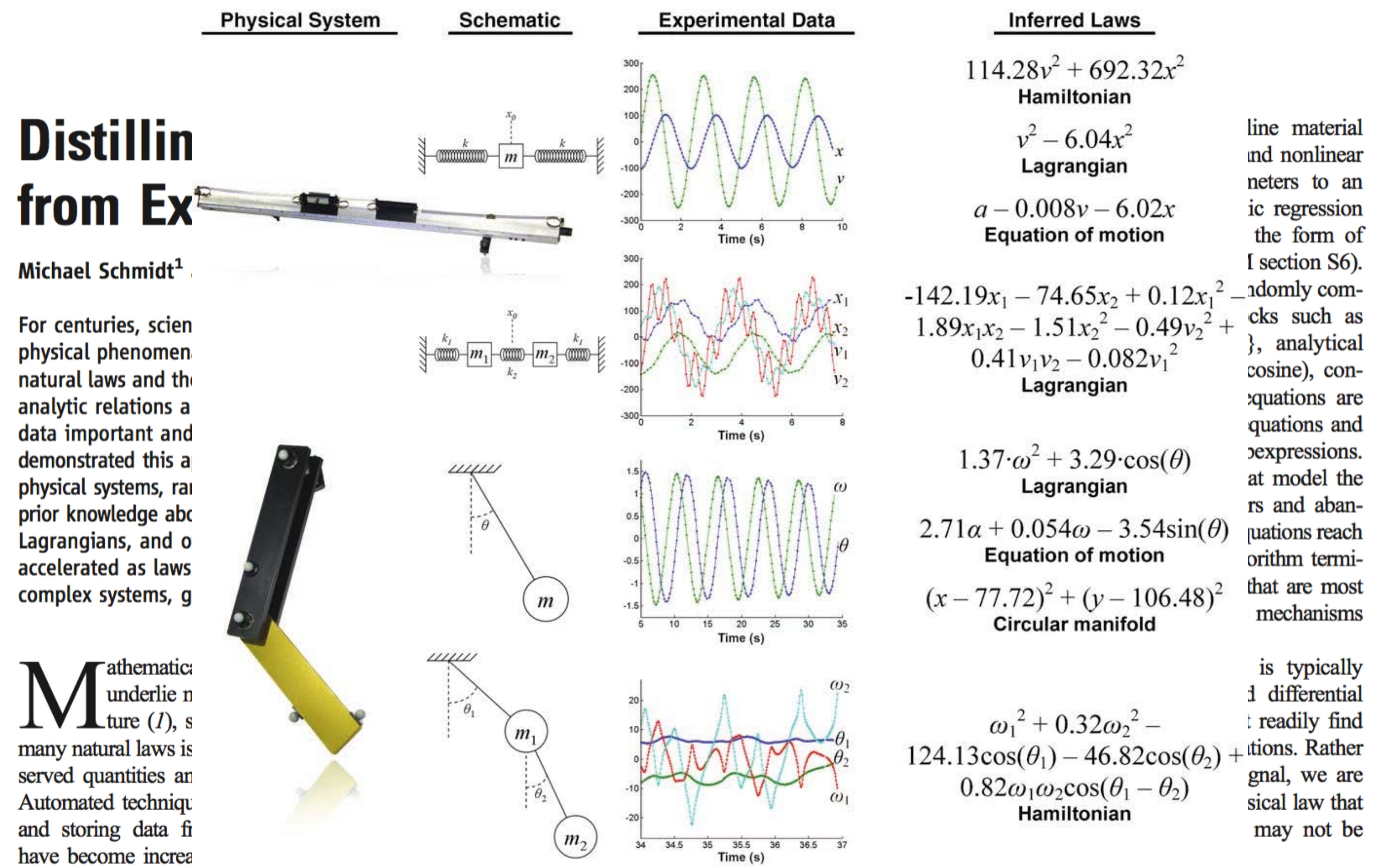
graph courtesy of Kevin Reardon

- Volumes of data are growing, e.g.,
  - Webpages: ~10 PB
  - youtube: ~2.5 petabyte/year uploaded
  - LHC: 15 petabytes/year
  - Radiology: ~70 petabytes/year
  - Large synoptic survey telescope LSST: several 100s PB/10 yrs
  - Square kilometer array: ~1 petabyte/min, ~500,000 PB/yr
- The ‘normal’ approach of downloading data to a local machine for analysis will no longer be feasible
- How do we extract “science” (specification? prediction? insight/intuition? set of equations?) from big data volumes

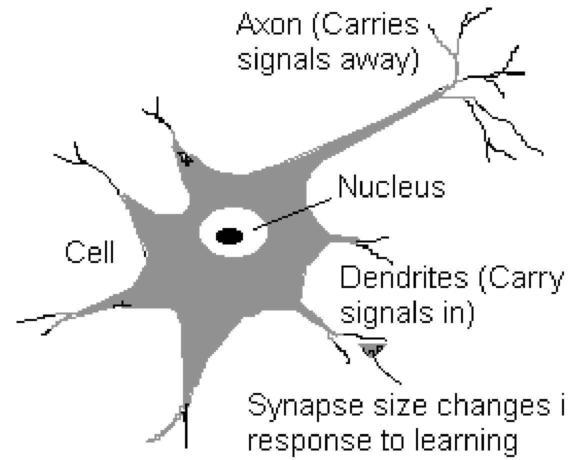
How does this impact the world of science?

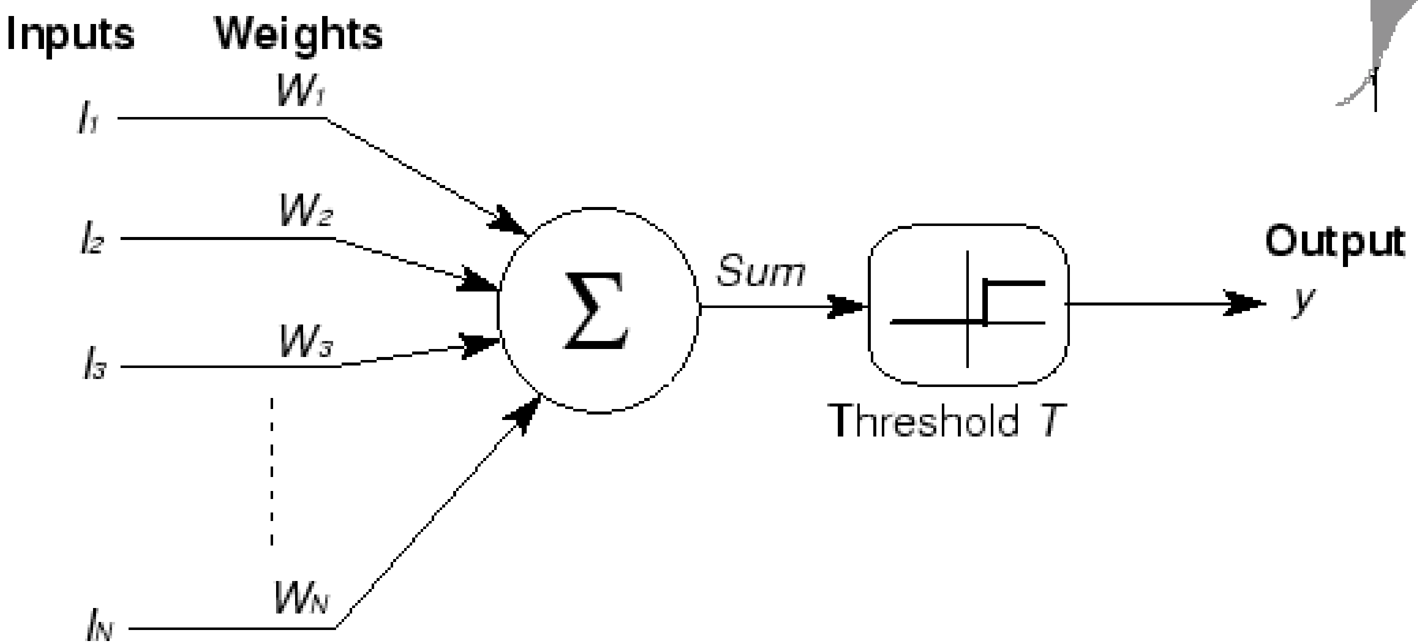


Looking for fundamental physics



Neural network background

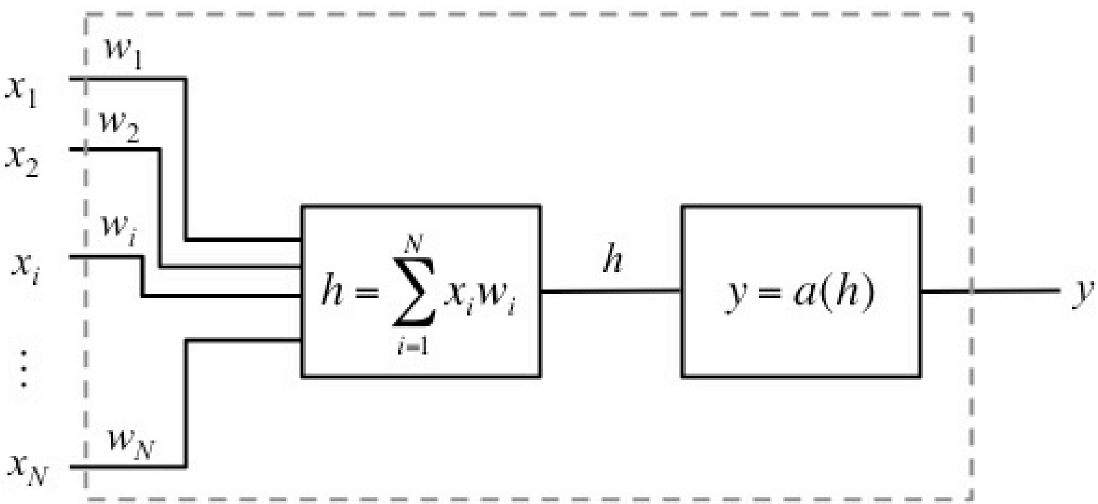




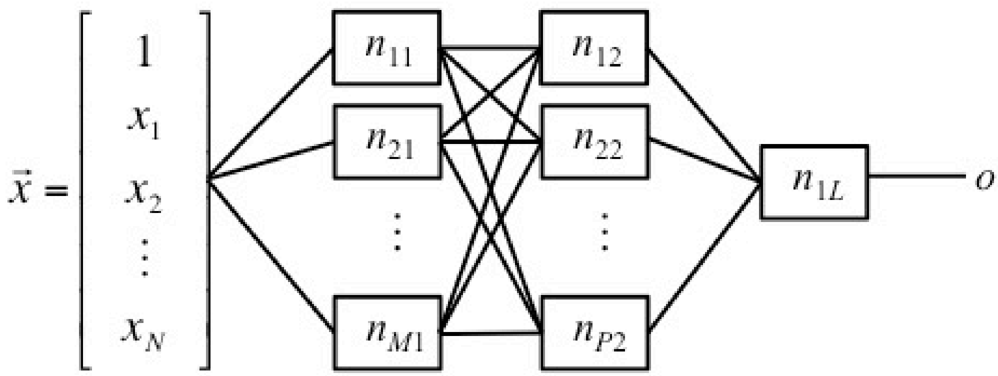
- McCullough & Pitts neuron [1943] to replicate biological neuron
- Built into “perceptron” [Rosenblatt, 1957] with learning rule
- Uses a simple step (“Heaviside”) activation function
- Problem with convergence of the learning rule (step discontinuity)

General neuron model and network

(a) Neuron



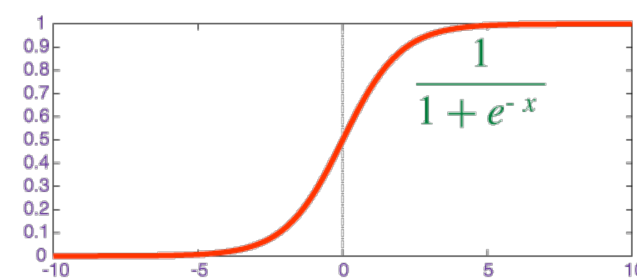
(b) Neural network



$h_k = [1, x_1 \ldots x_m][w_{k0} \ldots w_{km}]^T$

$y_k = \Phi(v_k) = 1/(1 + \exp[-v_k])$

Smooth and differentiable -> training by backpropagation algorithm ~1986



Training the neural network

- Training: finding the value of network weights (i.e., free parameters) the minimize the error/cost function to some "training set".
- Error

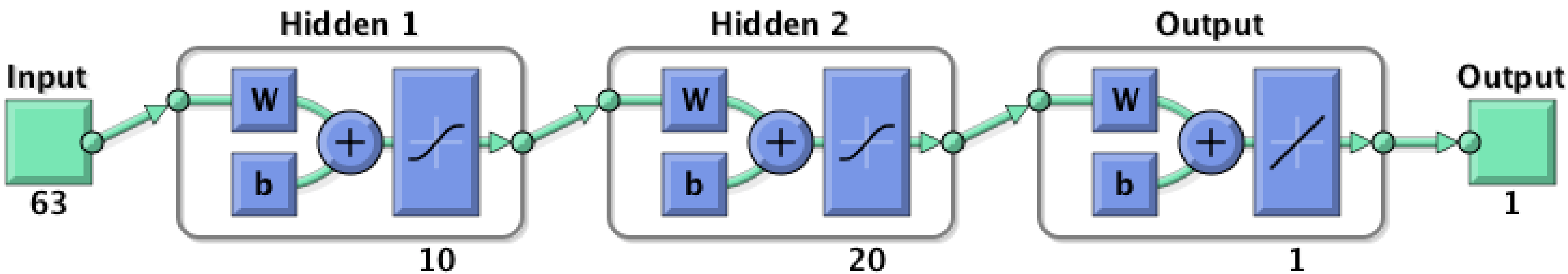
$$E = \frac{1}{2}\sum_{i=1}^{N}\left(t_i - y_i\right)^2$$

- Use gradient descent: $w_j^{n+1} = w_j^n - \eta \frac{\partial E}{\partial w_j^n}$
- Use an algorithm called "Backpropagation" (of errors), [Rumelhart, Hinton & William, *Nature* 1986]

Goal: *Given a set of sparse measurements of quantity Q, at location r and time t, reconstruct Q over all r at any t*

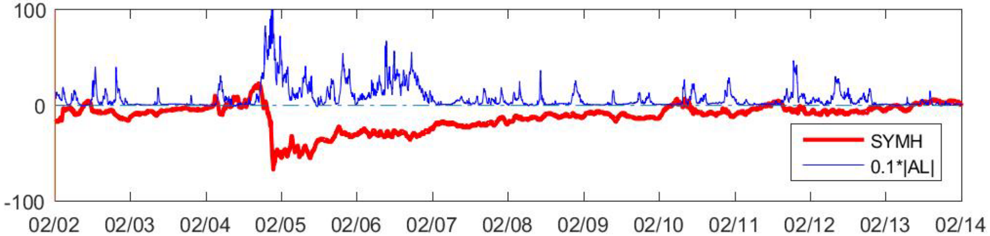
- What is Q? Any quantity that can be measured, for example on a satellite, and there are a large number of observations.
- Examples I'll show now:
  1. Electron number density: Use THEMIS density data (from S/C potential) June 2008 – Oct 2014, TH-A, D, E in 5-min cadence (~10^6 samples)
  2. Energetic electron fluxes: Use Relativistic Electron Proton Telescope (REPT) data, Oct 2012-Oct 2014, 8 energy channels: 1.8 MeV-7.7 MeV
  3. Chorus wave intensity: upper and lower band waves, measured on THEMIS and RBSP, ~372k samples, May 2010-June 2014.
  4. Hiss wave intensity: RBSP data, Oct 2012-Sep 2014, 280k samples.
- Examples I won't show:
  – MagEIS data, available for whole RBSP mission
  – EMIC waves
  – Magnetosonic (equatorial noise) waves
  – Integrated ULF wave intensity
- Regressed against a time history of a geomagnetic index at 5 minute cadence
  – usually symH, occasionally AE, time history of 5-10 hours
  – Why geomagnetic index (and not SW)? Because it is simple, readily available, unlike SW which often has gaps, and should contain all the information in the SW.
  – Also "historic" (following SAMI3 model), but we will include in later versions.

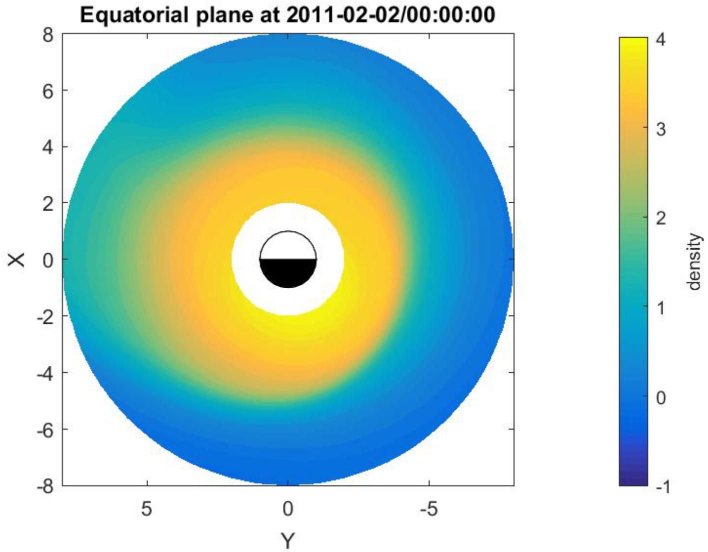
Network Architecture

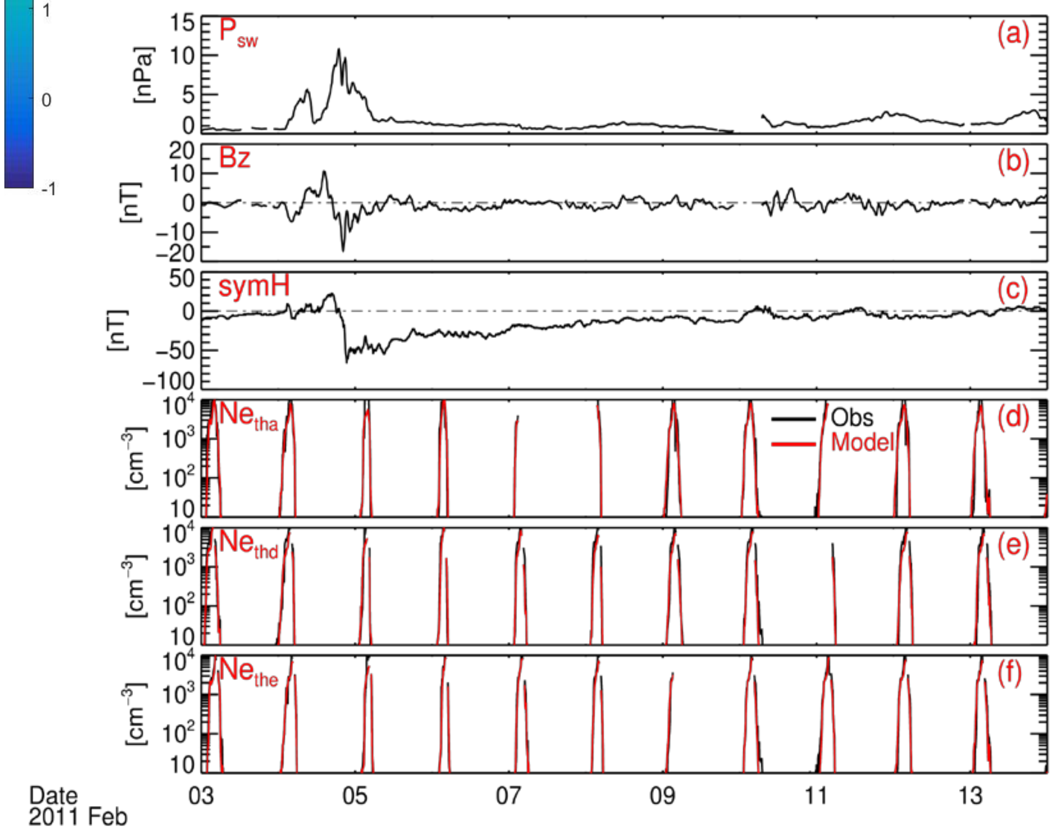


- Use a "deep" neural net architecture with 2 hidden layers. Why deep NN?
  - NN: t is a universal approximator, even with 1 layer [*Hornik et al*. 1989; *Cybenko*, 1989]
  - Deep: Don't need to know the feature set a-priori, deep architecture is more efficient and learns its own optimal feature set
  - First layer: dimensionality reduction, optimal feature construction
  - Second layer: more complex representations
  - Sigmoid activation function in hidden layers, linear in output layer
- Does it have to be a neural net? No! Just need a high variance model (SVM, HMM, etc.) and LOTS of data [*Banko & Brill*, 2001]
- Divide data into 3 parts: Training (70%), Validation (15%), and Test (15%)
- Continue "training" the neural net until error on validation set increases for 10 consecutive times, pick minimum error point. Use Scaled Conjugate Gradient or Levenberg-Marquardt method to optimize.
- Object is to pick the most generalized representation without over-fitting

Dynamic plasmaspheric density model: equatorial



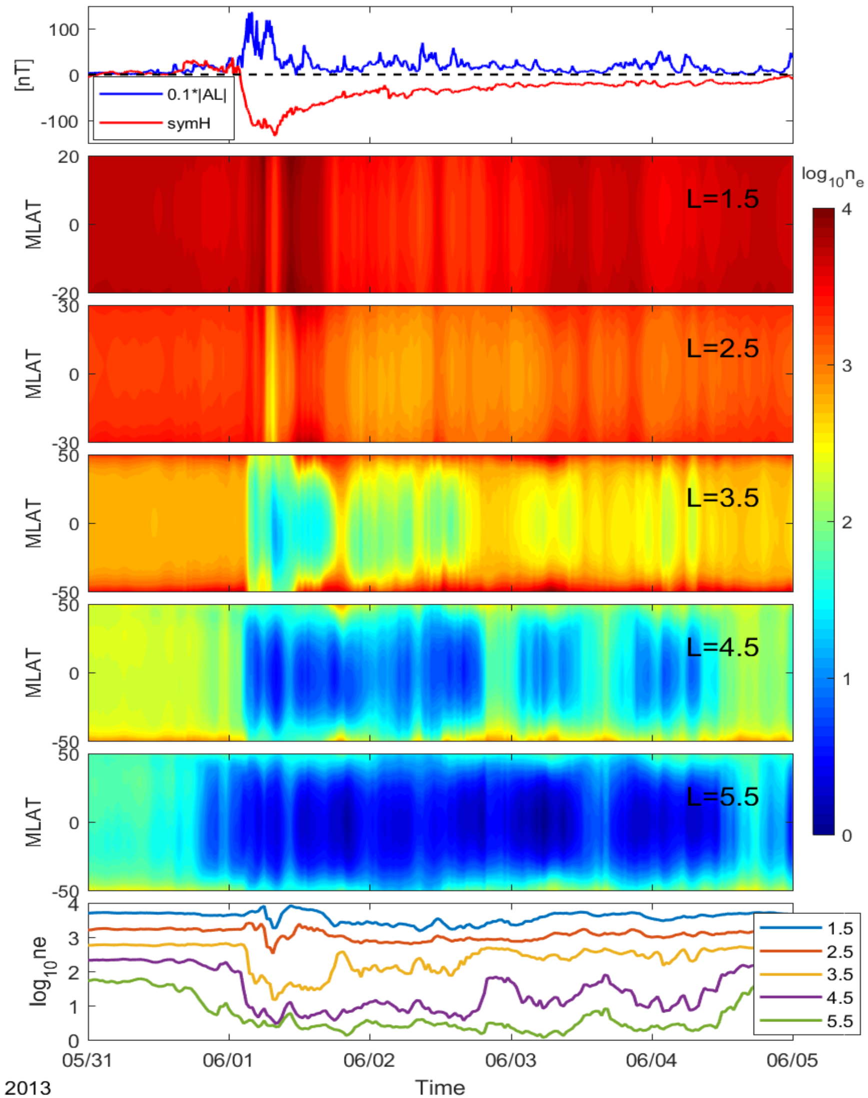


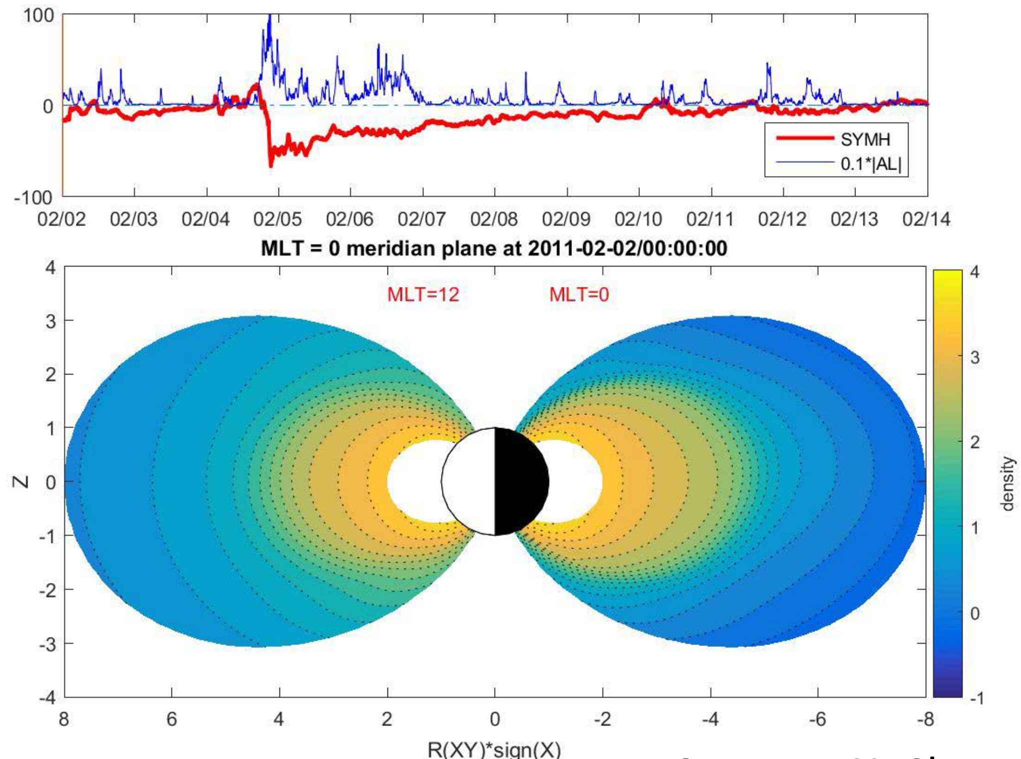


THEMIS data for training
June 2008 – Dec 2012
Courtesy Xiangning Chu

Dynamic density model: meridional

Insight discovery! Low L-shell density enhancements

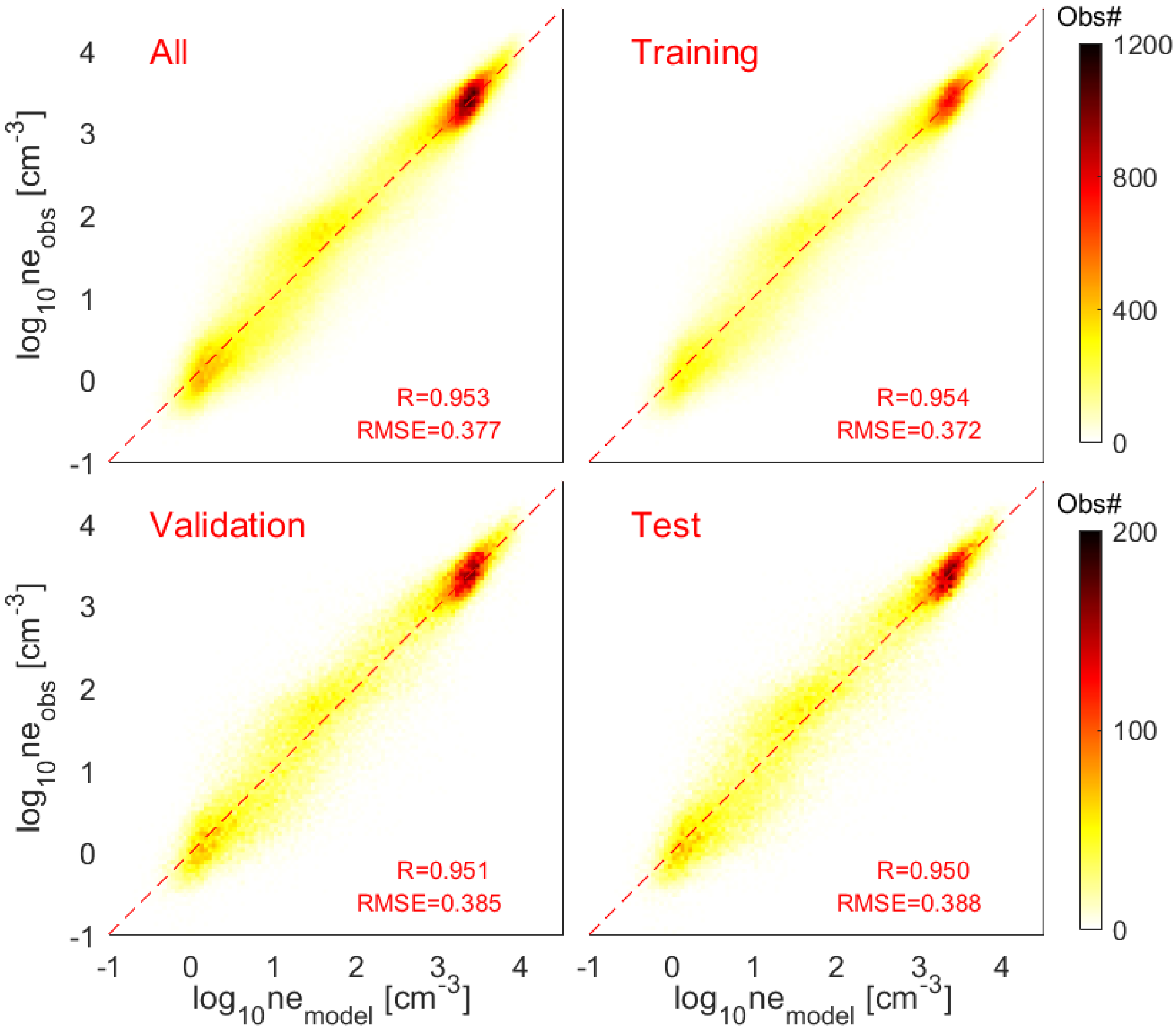




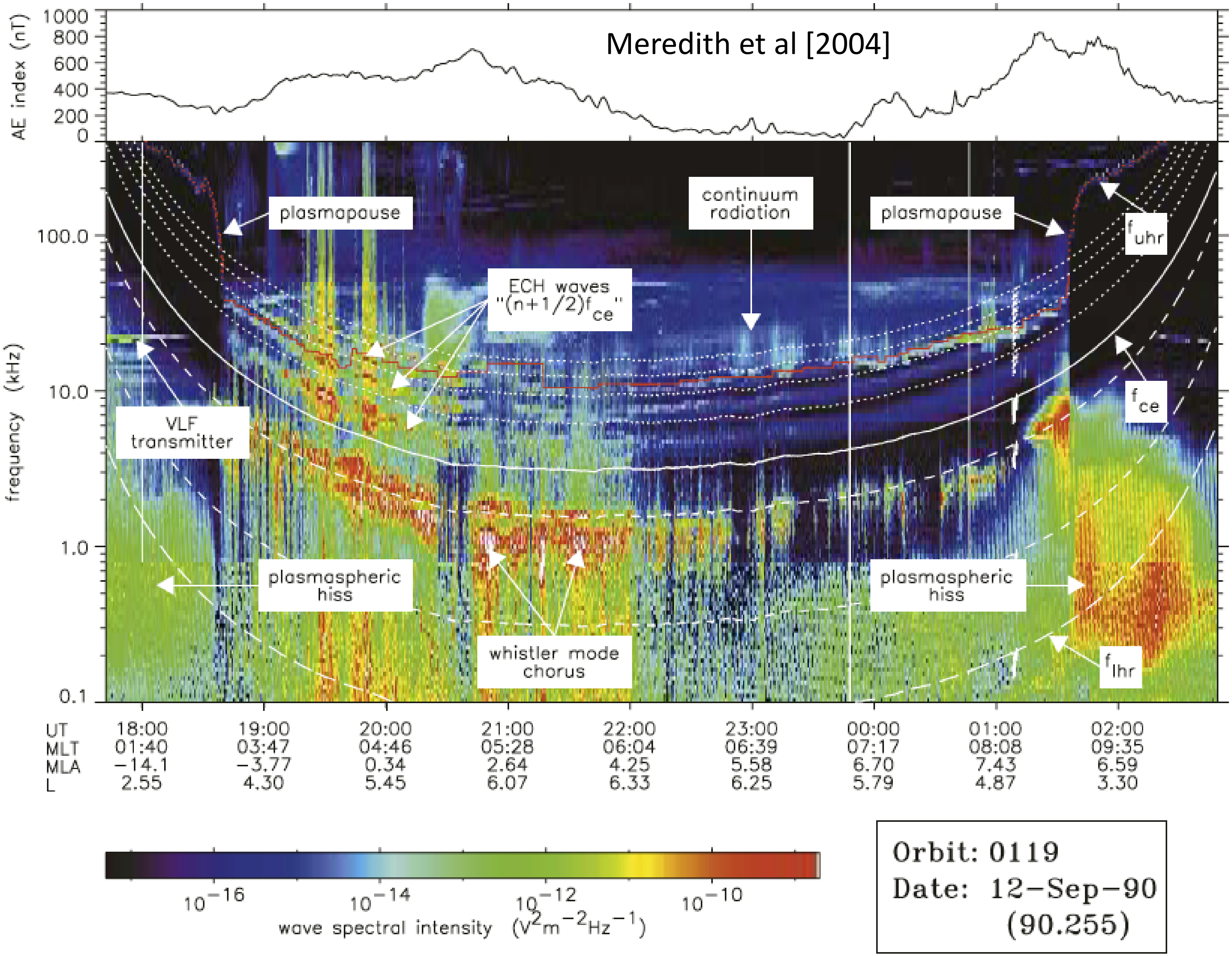
IMAGE, CRRES, POLAR, ISEE, CLUSTER,

Courtesy X. Chu

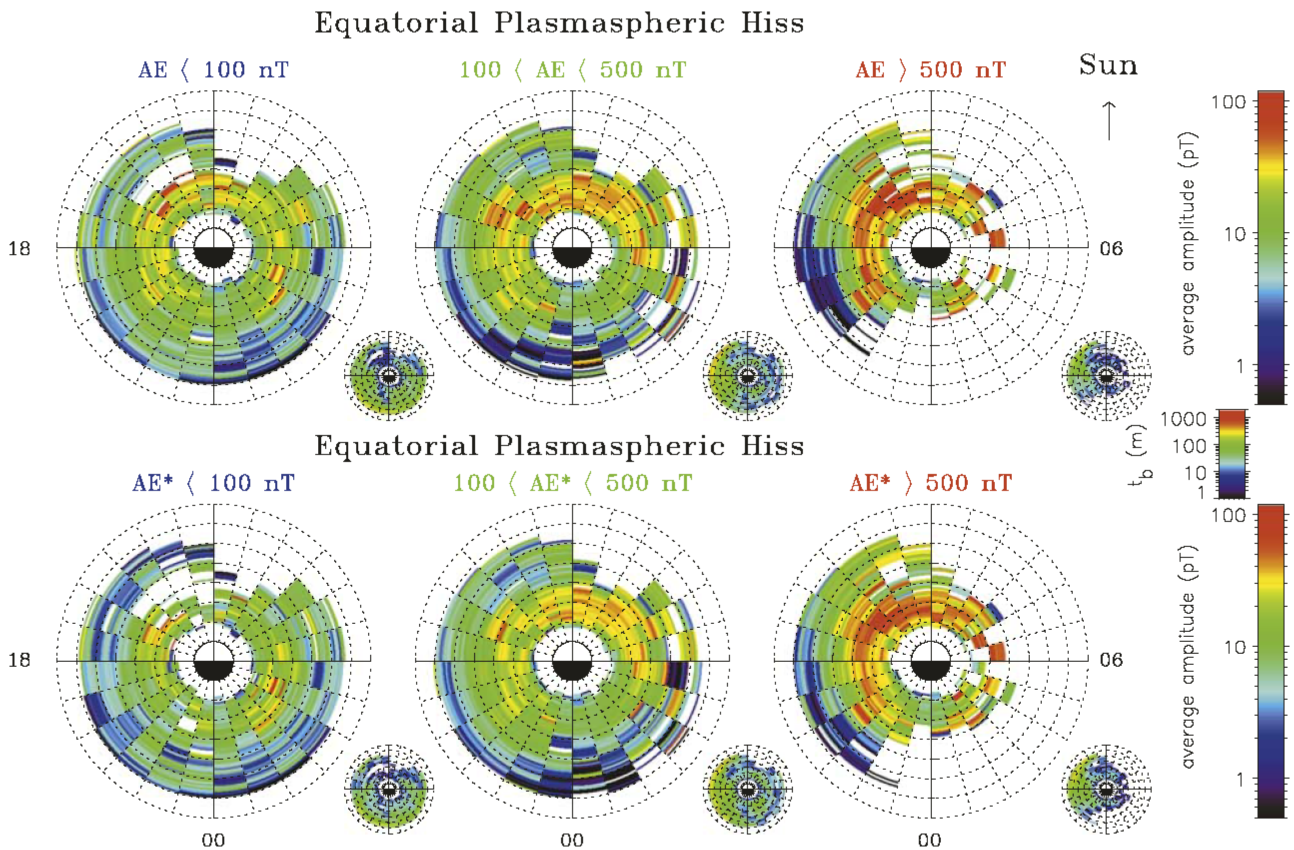
Density: Prediction-observation comparison



The wave environment in space



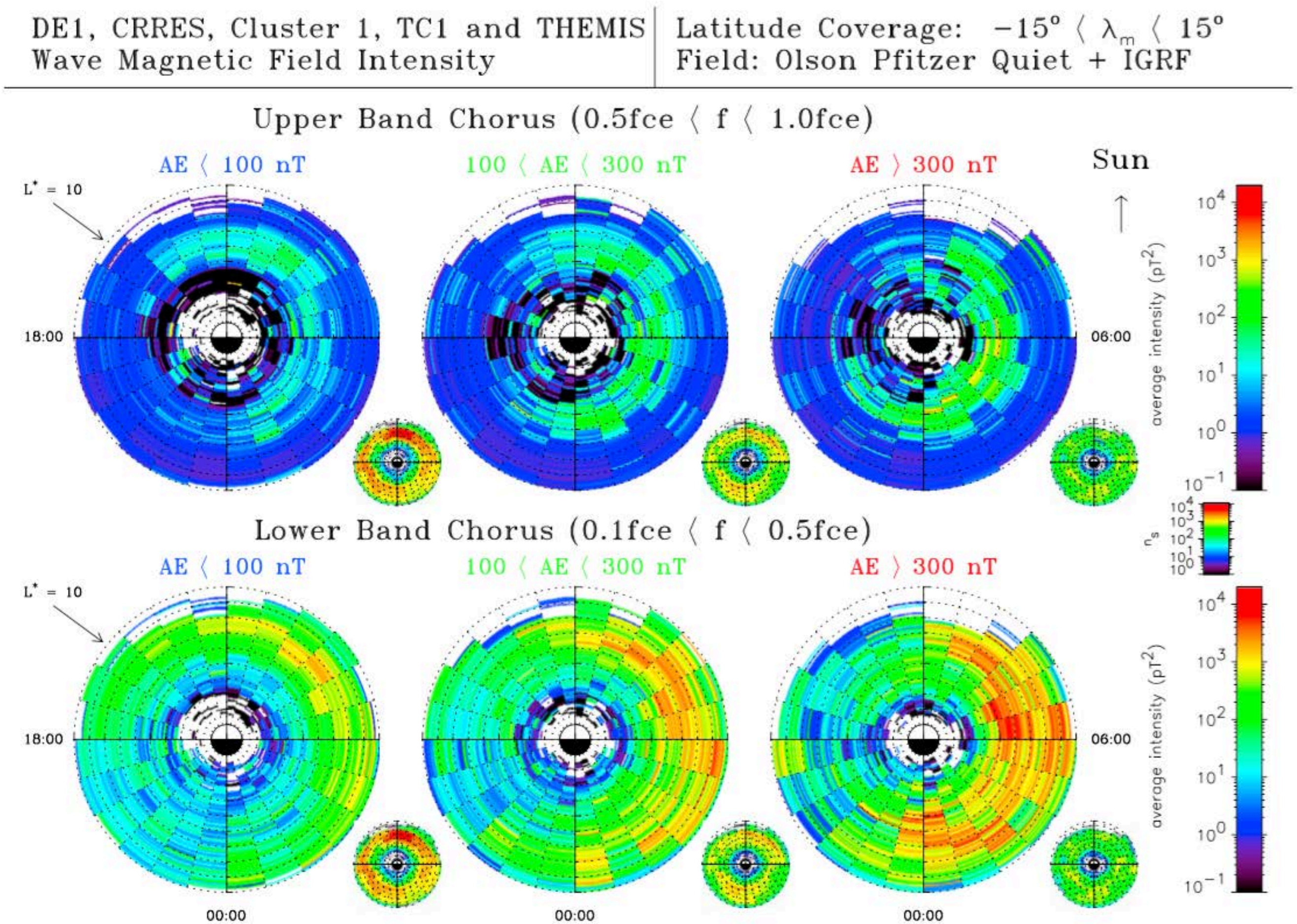
Plasmaspheric hiss statistical distribution



Meredith et al. [2004]

- Standard statistical picture of the plasmaspheric hiss distribution, showing geomagnetic control and local time asymmetry

Whistler-mode chorus statistical distribution



Meredith et al. [2012]

- Geomagnetic control and local time distribution
- More satellites (DE1, CRRES, Cluster 1, TC1, THEMIS), but same single-spacecraft approach!

Chorus wave environment (movie)

<u>Whistler-mode chorus</u>: Target: 5-min resolution log10(Bw) from THEMIS A, D, E (~6 years) and both RBSP-A, -B from 2012-10-01 to 2015-12-01.

Input : L, sin(MLT), cos(MLT), MLAT, 5-65: AL index in 5 min resolution for the previous 3 hours, 66-114: symH index in 30 min resolution for the previous 24 hours.

Weight: 4+log10(Bw) as the weights of the targets values.

Architecture: [20 10]; Perform: factor of ~2, r=0.7263.

Hiss wave environ-ment (movie)

Plasmaspheric hiss: Target:

30.

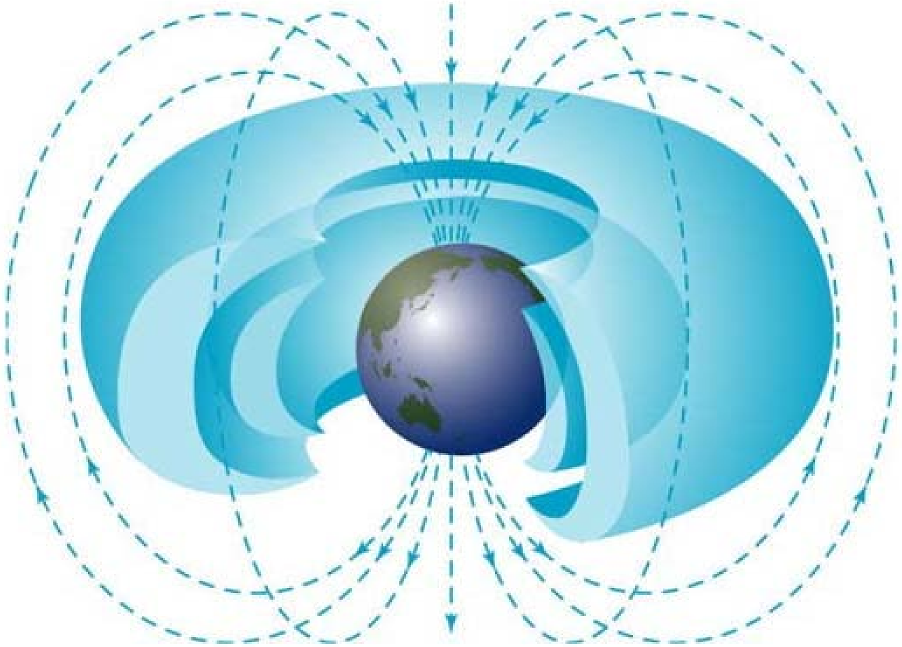
Inputs: L, sin(MLT), cos(MLT), MLAT, one present value of the AL index, 6-11: symH index in 60 min resolution for the previous 5 hours.

Architecture: [20 10]; Perform: factor of ~2, r=0.617.

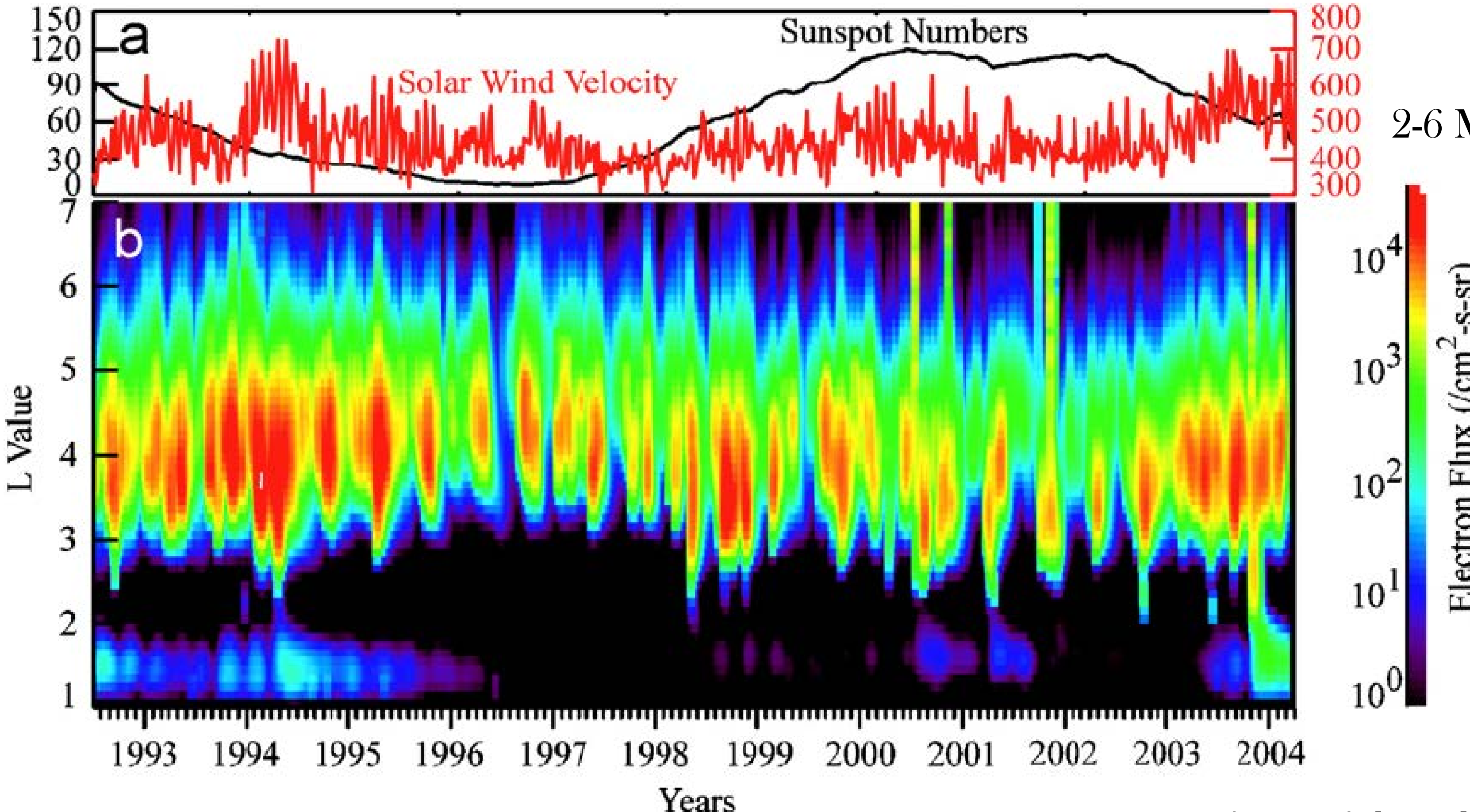
Prediction of radiation belt fluxes

Explorer 1 launch: Jan. 31st 1958

- High energy >1 MeV ("killer") electrons
- Satellite & astronaut hazard
- Occur in 2 zones



Variability of Outer belt

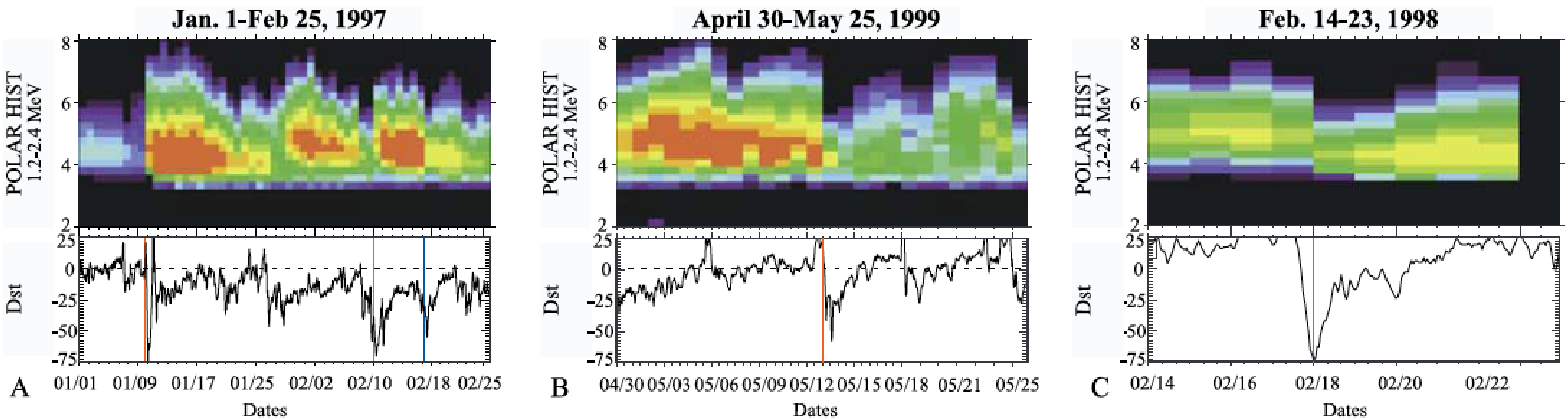


2-6 MeV

Baker et al. [2008]

Outer radiation belt exhibits variability, several orders of magnitude, timescale ~minutes.

Predictability of outer belt fluxes

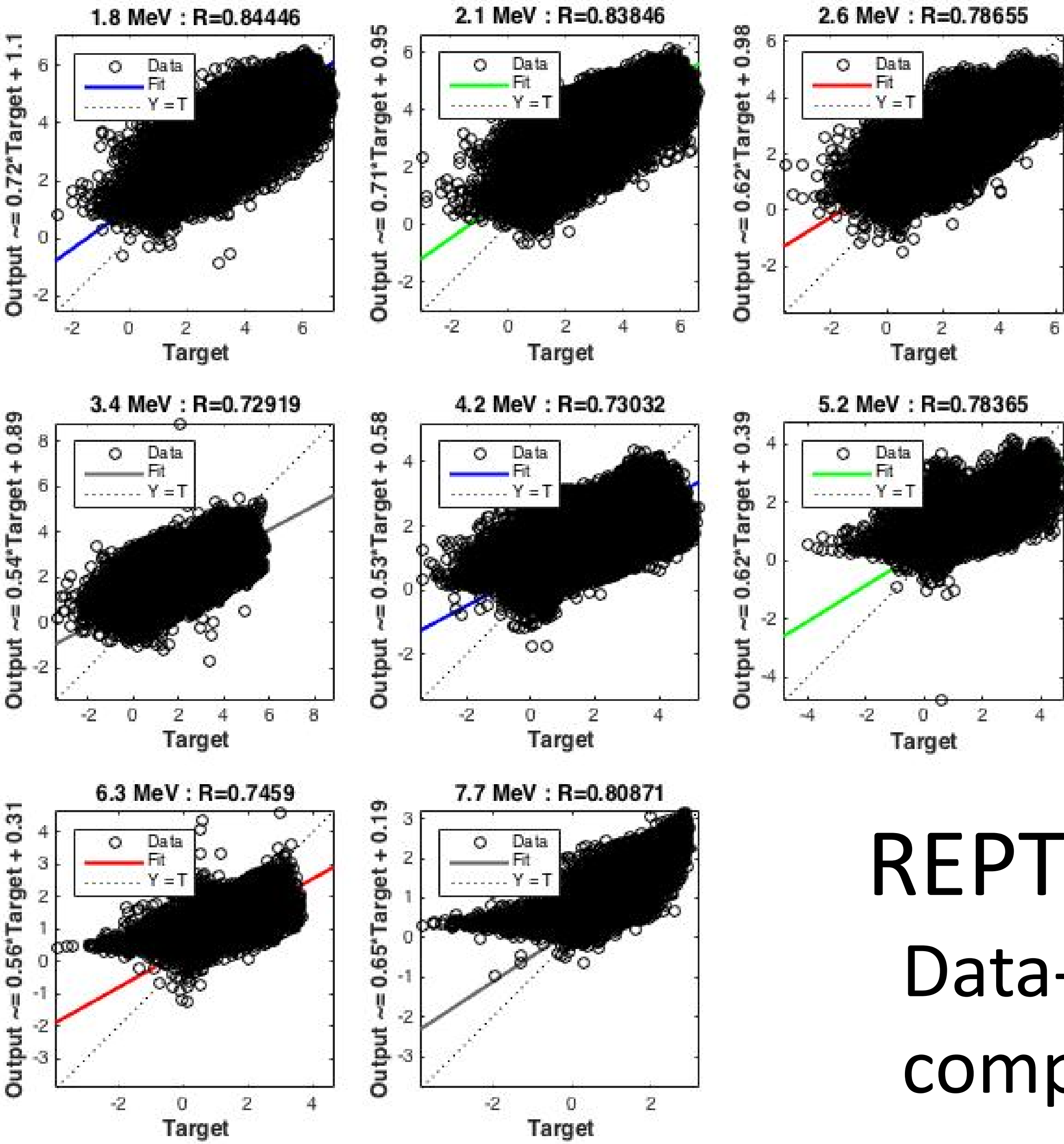


Reeves et al. [2003]

- Similar sized storms can produce net increase (53%), decrease (19%), or no change (28%). "*Equally intense post-storm fluxes can be produced out of nearly any pre-existing population*"
- Delicate balance between acceleration and loss, both enhanced during storm-time, "*like subtraction of two large numbers*".

REPT movie

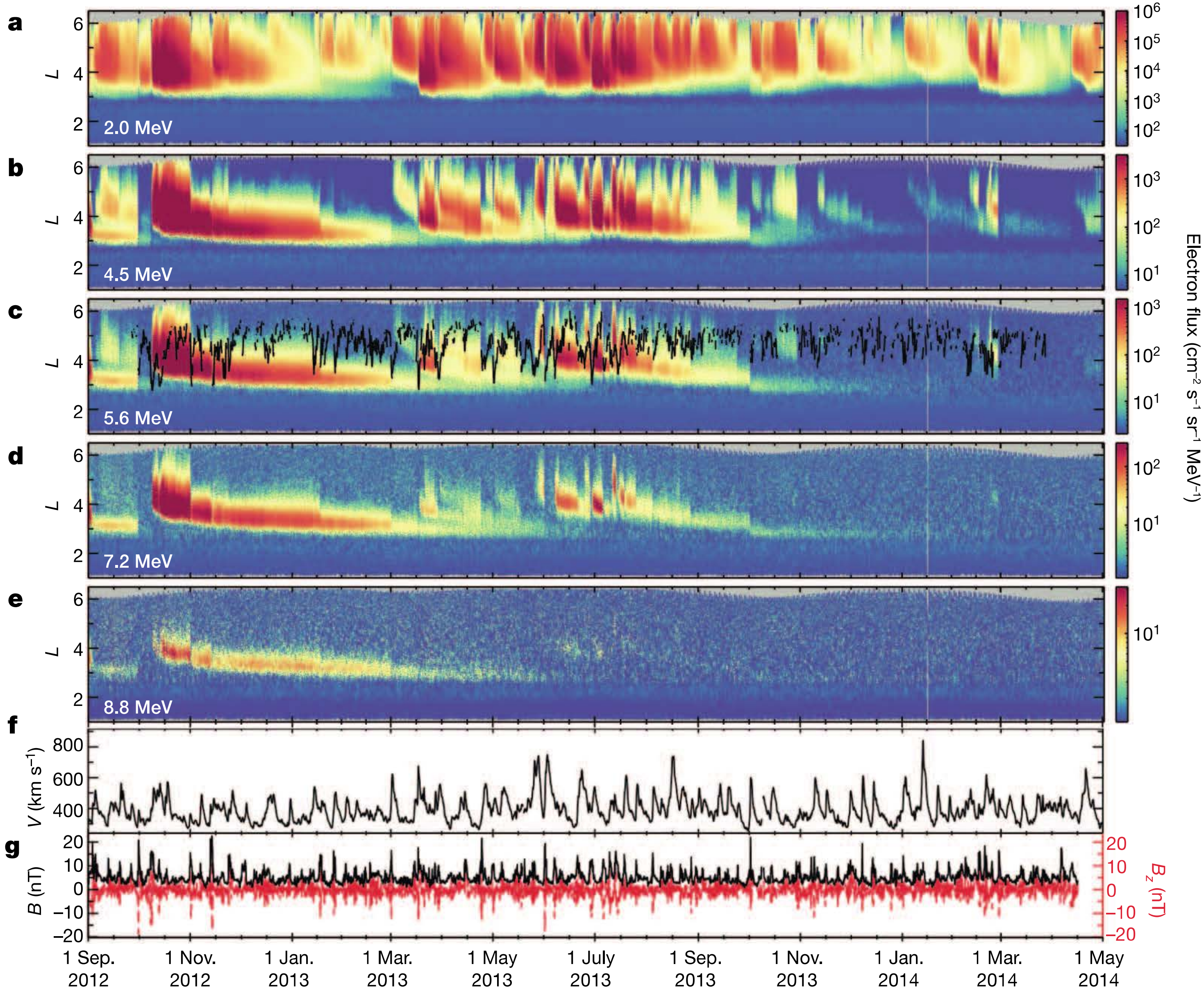
- REPT data: 8 energy channels 1.8, 2.1, 2.6, 3.4, 4.2, 5.2, 6.3, 7.7 MeV
- Regressed on 10 hours of Dst only
- Small number of samples, ~188k in total. Artefacts show up in higher energy channels since few accelerations reach those energies!



REPT fluxes

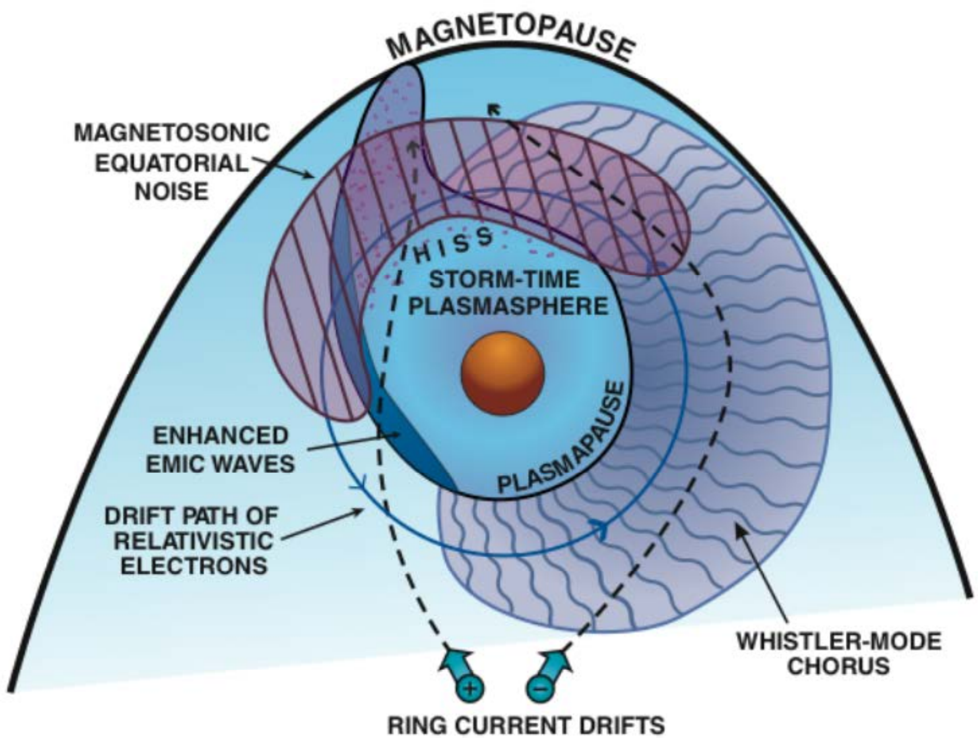
Data-model comparison

Rich vs poor data environments



Collective wave effects

- Particles drift around the earth
- Accumulate scattering effects of:
  - ULF
  - Chorus
  - Hiss (plumes)
  - Magnetosonic
- Characteristic effects of each waves are different and time dependent



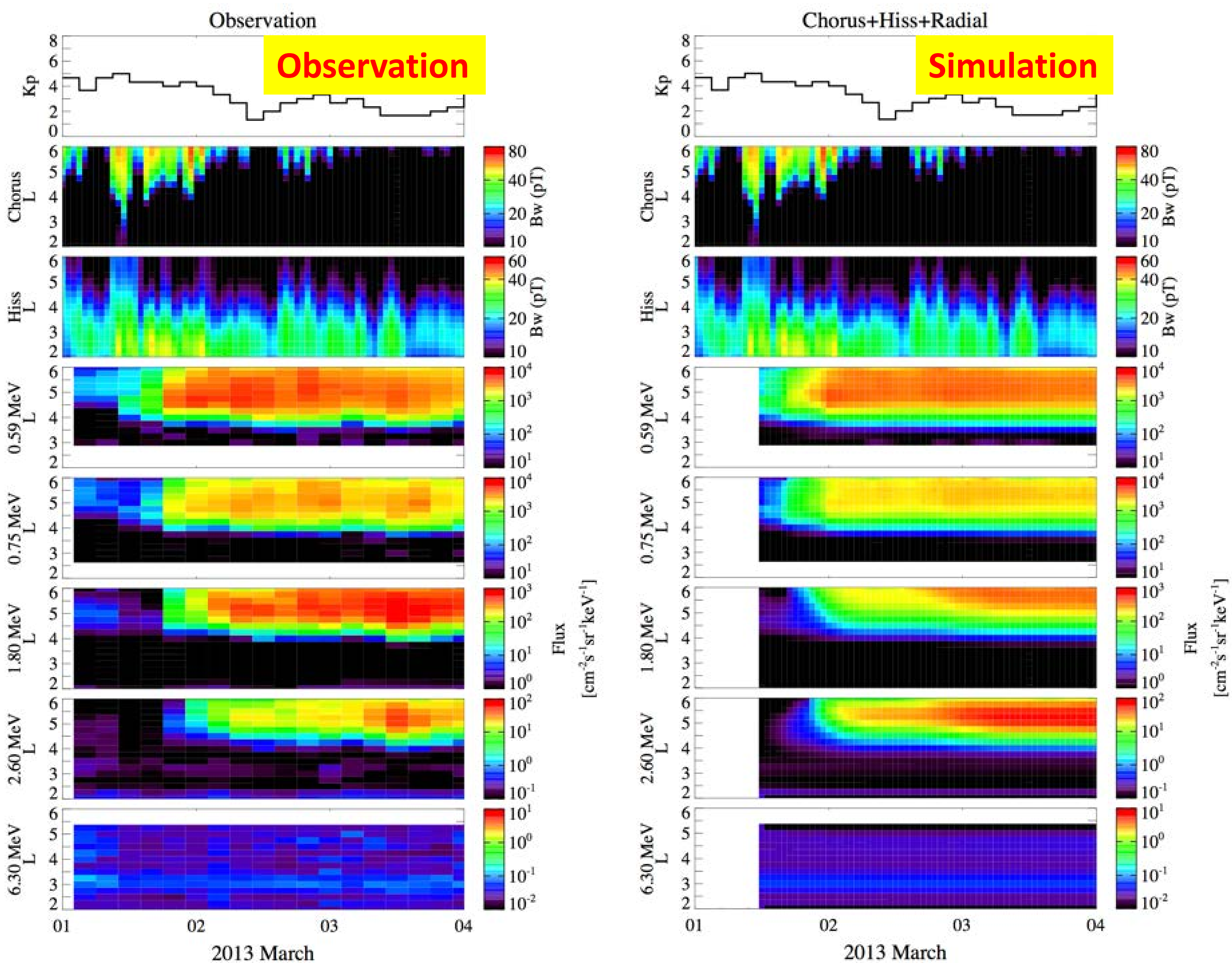
Thorne [2010] GRL “frontiers” review

Data/model fusion approach

March 1-3 2013

Moderately disturbed event, Kp~4, Dst~-80 nT

Modest acceleration to ~2.6 MeV



Summary

- Scientific data volumes are growing and a new approach is required to extract “science”.
- We presented a “unified approach to inner magnetospheric state prediction” Bortnik et al., JGR
- Take a set of observations of some quantity Q measured at (**r**, t), and reconstruct Q at all **r** as a function of t. Q can be anything, e.g., density, energetic particle fluxes, and different wave modes.
- Used a 2-layer neural network, 5-10 hour history of sym (or AE) as regressor, and 5 min cadence, we get few 100k samples from THEMIS and/or RBSP data sets.
- Preliminary results show good agreement (R~0.8-0.9) between model and data, the “physics” are baked into the model and need to be interpreted (the data deluge does NOT make the scientific method obsolete, cf Chris Anderson WIRED magazine), e.g.
  - dawnside chorus occurrence, increasing MLT with latitude,
  - hiss more intense on dayside and
  - relativistic REPT fluxes exhibit dynamic variability, drift shell splitting (more intense on nightside than dayside)
- Good potential for specification models (or part of physical models)