2018 Summer Workshop

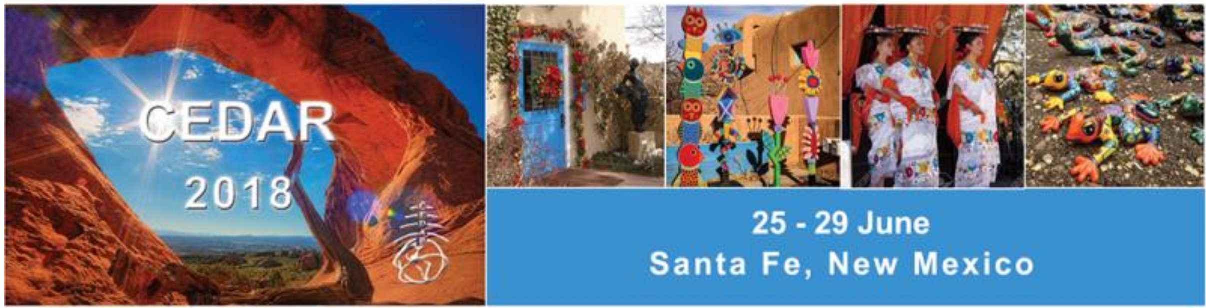
June 17-23, 2018

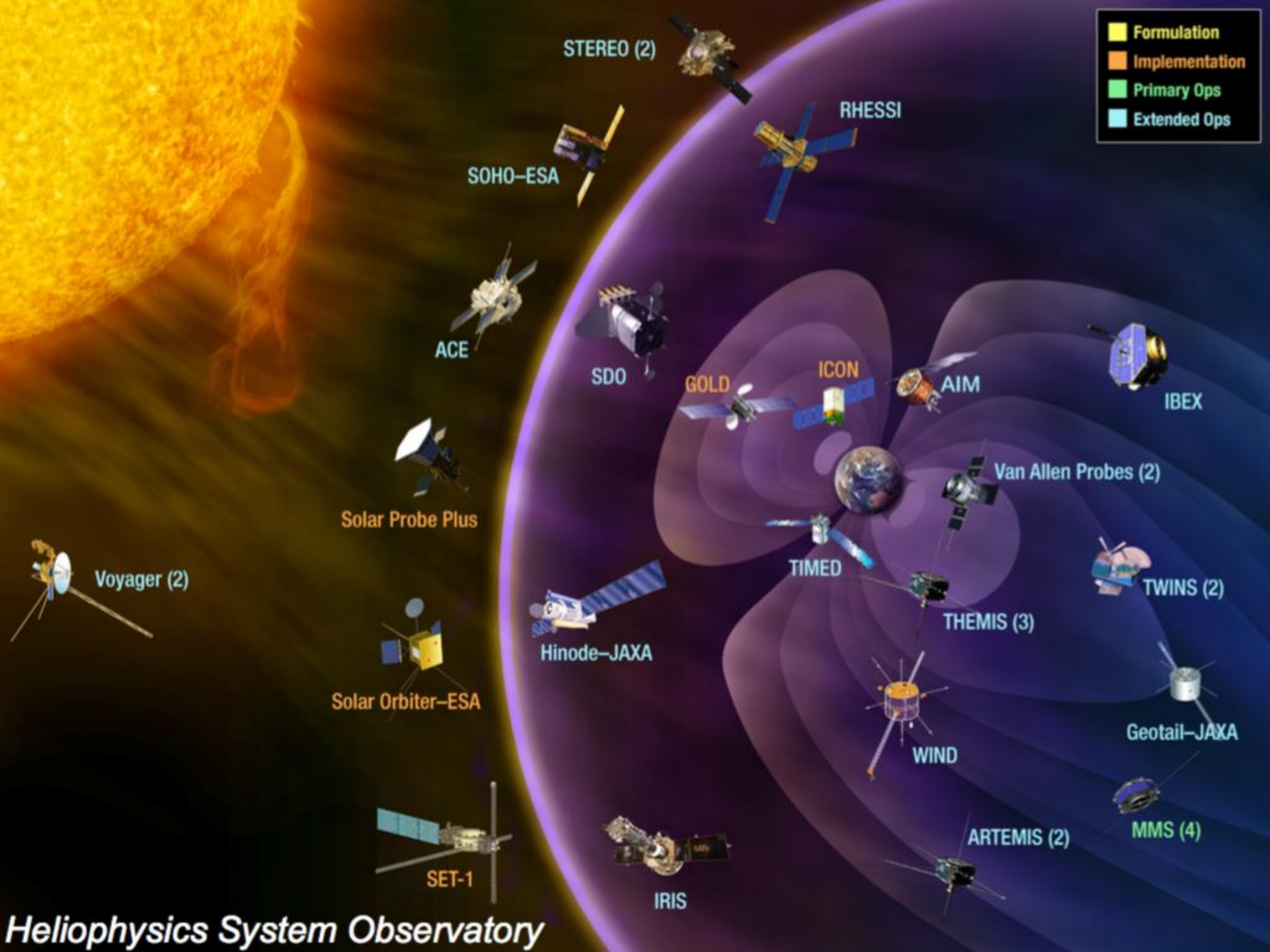
Eldorado Hotel & Spa ▪ Santa Fe, New Mexico

Magnetospheric research in the age of data science

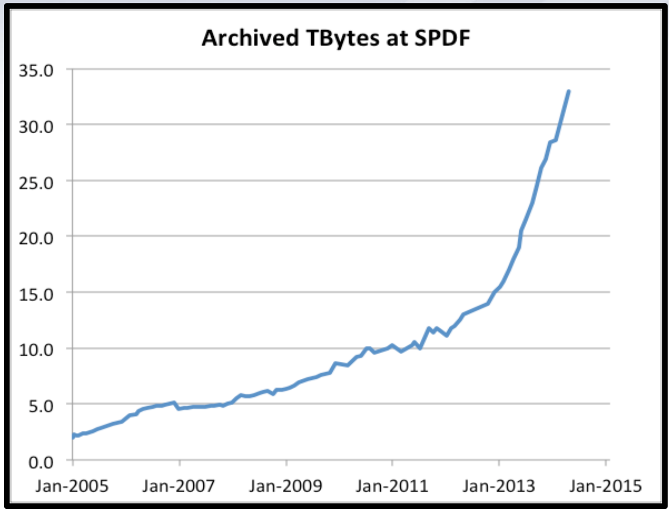
Jacob Bortnik, Xiangning Chu, Wen Li, Qianli Ma, Vassilis Angelopoulos, Richard Thorne

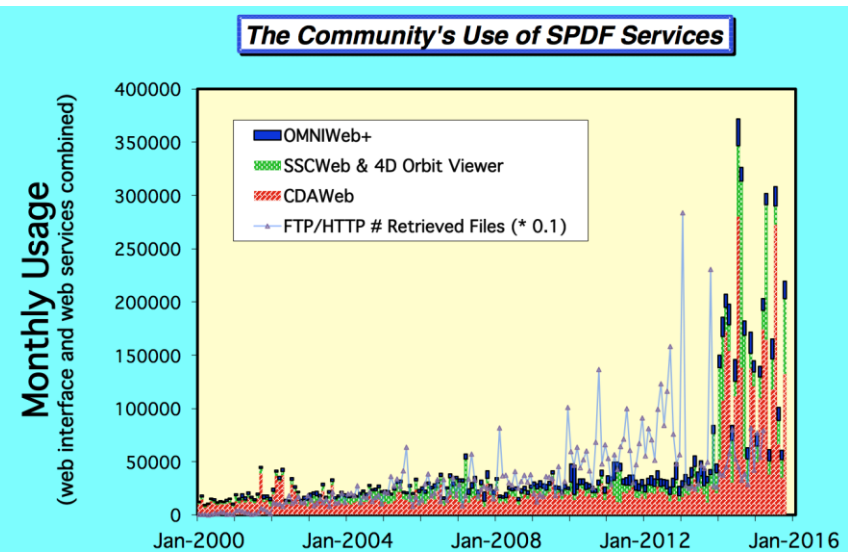
With thanks to: Craig A. Kletzing, George B. Hospodarsky, Harlan E. Spence, Geoffrey D. Reeves, Shrikanth G. Kanekal, Daniel N. Baker

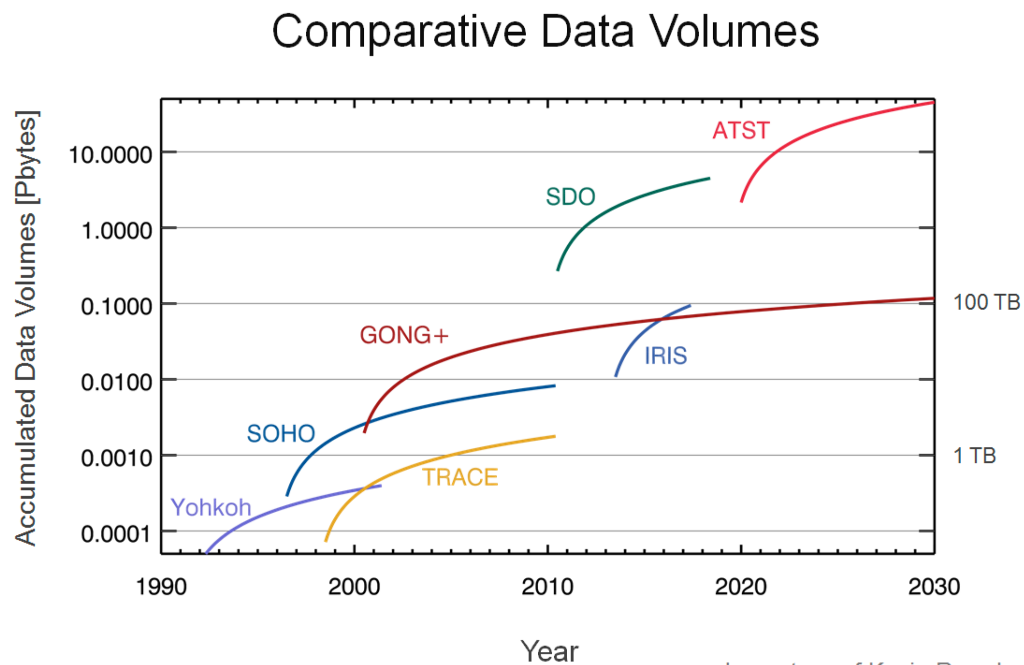



Formulation
Implementation
Primary Ops
Extended Ops
STEREO (2)
RHESSI
SOHO–ESA
ACE
SDO
GOLD
ICON
AIM
IBEX
Solar Probe Plus
Van Allen Probes (2)
TIMED
Voyager (2)
TWINS (2)
THEMIS (3)
Hinode–JAXA
Solar Orbiter–ESA
WIND
Geotail–JAXA
MMS (4)
ARTEMIS (2)
SET-1
IRIS
Heliophysics System Observatory

Growth in data volume



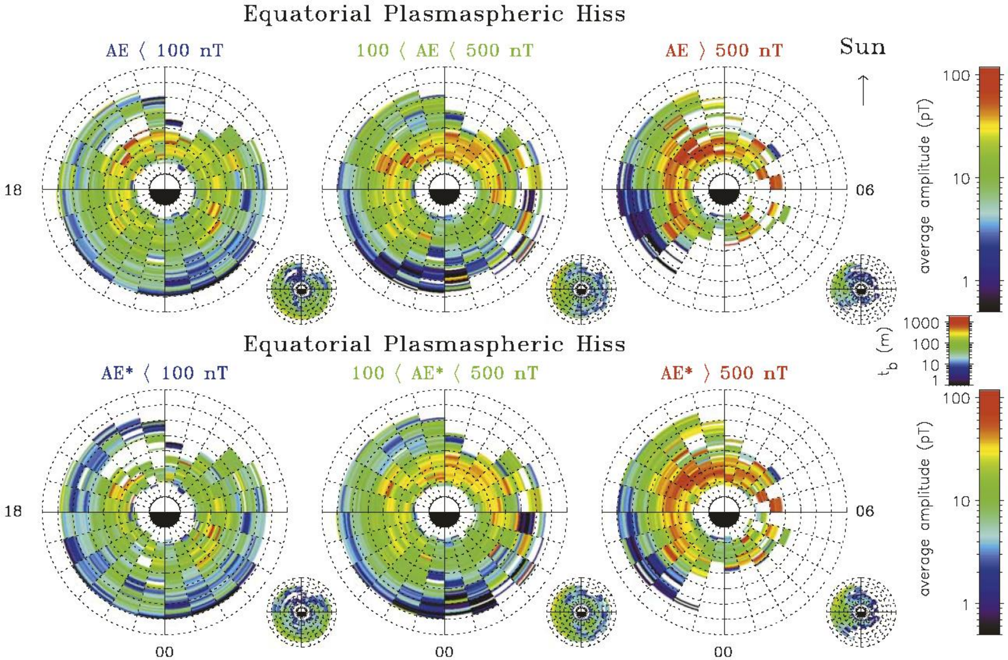




graph courtesy of Kevin Reardon

As of Feb 2016: 91TB of holdings; 23 current spacecraft: ~130 past. Monthly ~6TB are served (~730k files). To add: MMS to be made public 3/1/2016. Van Allen Probes, Cluster WBD, etc.

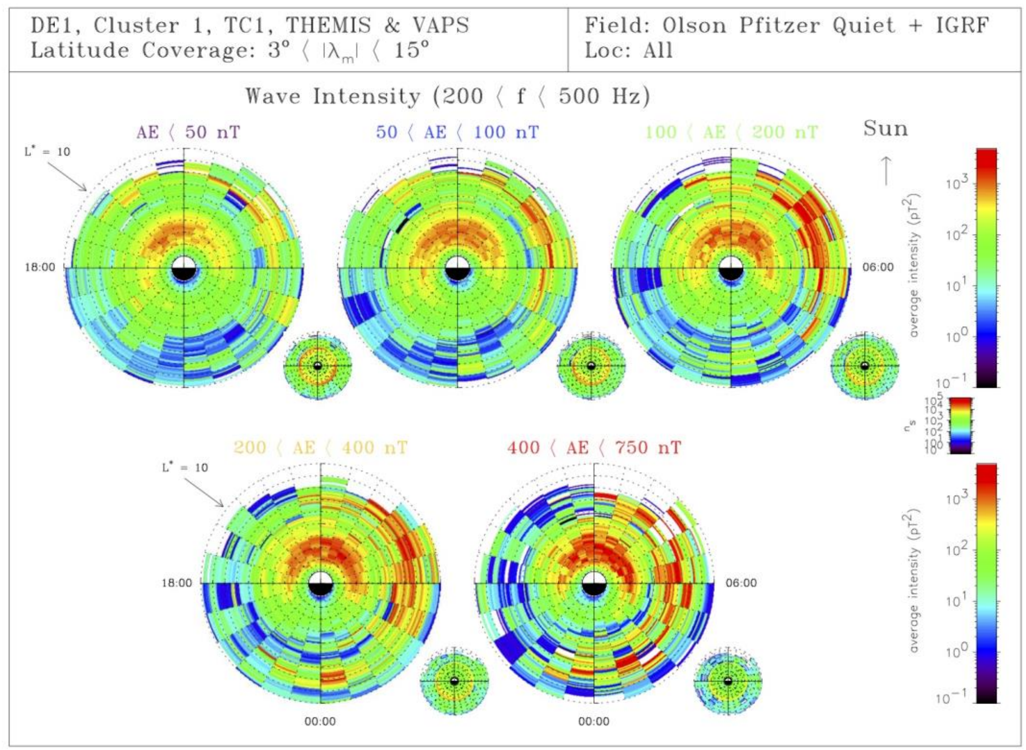
How do we extract "science" (specification? prediction? insight/ intuition? set of equations?) from big data volumes



What are we doing now?

Same Author et al. [2018]

Author et al. [2004]



Whistler-mode wave (hiss) distribution, showing geomagnetic control and local time asymmetry

How would this distribution change with 10x, 100x, etc. more data?

What should we be doing?

1. Be able to "exploit" archived data, i.e., extrapolate what past mission data would look like for a current event or case study

2. Make sense of large (and increasing) volumes of data currently being collected, e.g., petaByte data-sets cannot be analyzed in the "usual way"

3. A model that "improves" the more data is added, i.e., if data quantity increases 10x, model should be able to improve its temporal and spatial resolution and pick up more subtle patterns or trends in behavior

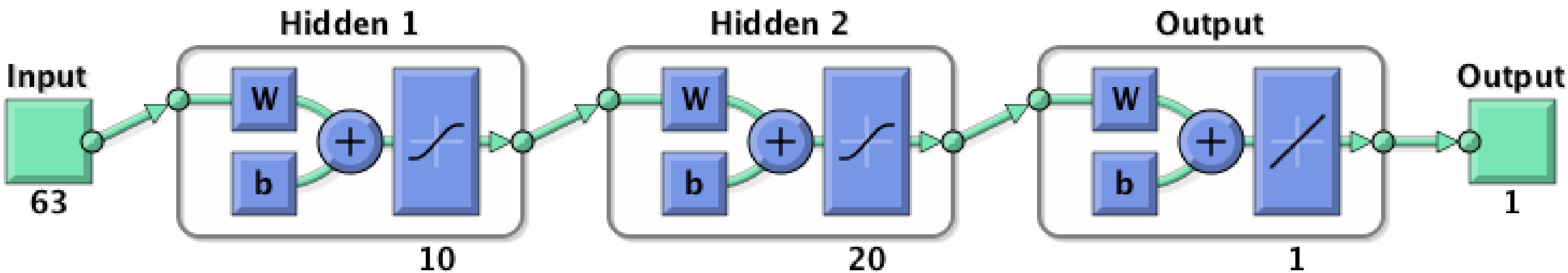
4. Take full advantage of distributed arrays of instrumentation collecting simultaneous, homogeneous, i.e., don't treat multi-spacecraft (distributed array) data as a sum of individual satellites (instruments)!

What could we be doing?

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.
- Regressed against a time history of a geomagnetic index at 5 min 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.
  - "historic" (following SAMI3 model), but we will include in later versions.

Neural network approach



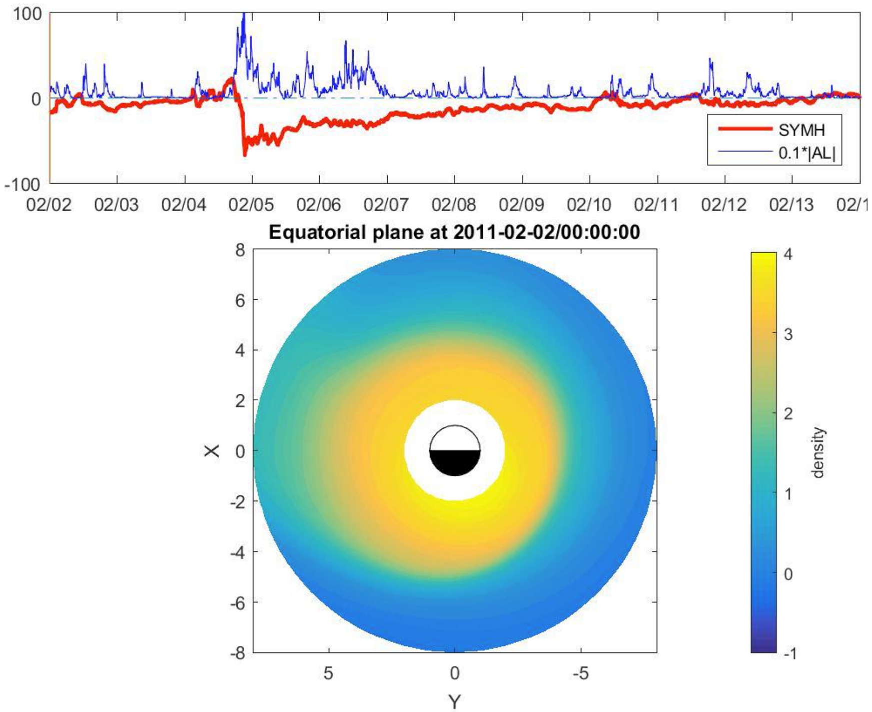
- Use a "deep" neural net architecture with 2 hidden layers. Why deep NN?
  - NN: t is a universal approximator, even with 1 layer [*Hornik et al*. 1989; *Cybenko*, 1989]
  - Deep: Don't need to know the feature set a-priori, deep architecture is more efficient and learns its own optimal feature set
  - 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**

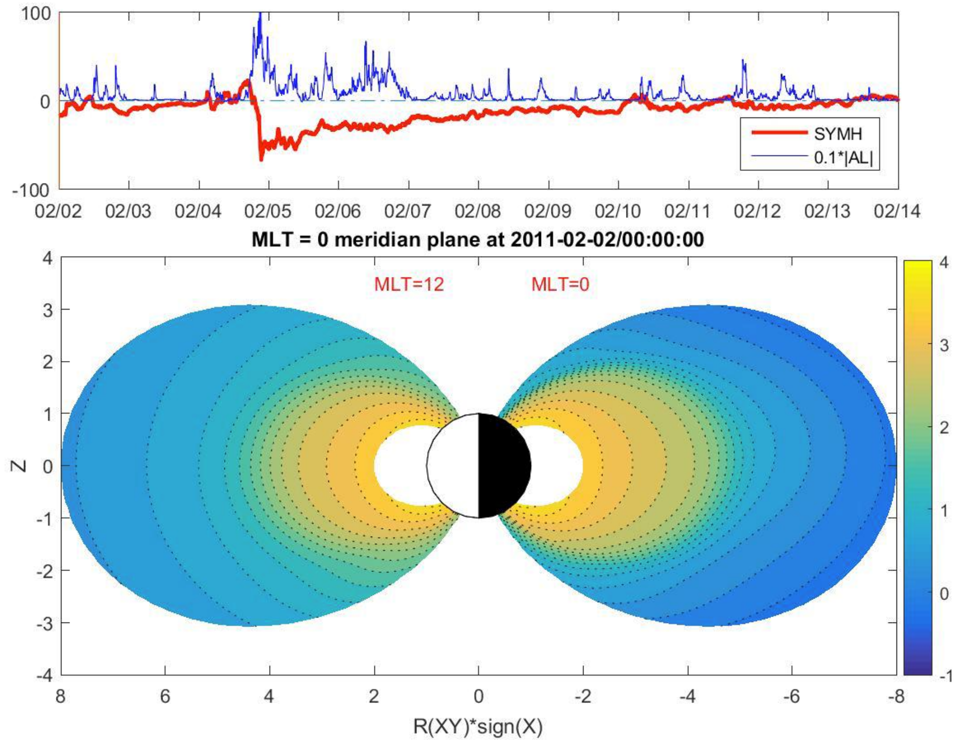
Reconstruction of 3D plasmasphere

Equatorial Electron number density: 5-min resolution electron density from THEMIS A, D, E probes, 2008-06-01 to 2014-11-30.

Inputs: 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. Architecture: [20 10]; Perform: factor of ~1.5, r – 0.957

3D Electron number density: 5-min resolution electron density from ISEE, CRRES, POLAR, and IMAGE. L~1-11, MLAT: -50 to 50 deg, all MLT, 1977-2005 resulting in 217,500 data points. R~0.954





Approach is general: can be applied to any quantity: hiss and chorus waves

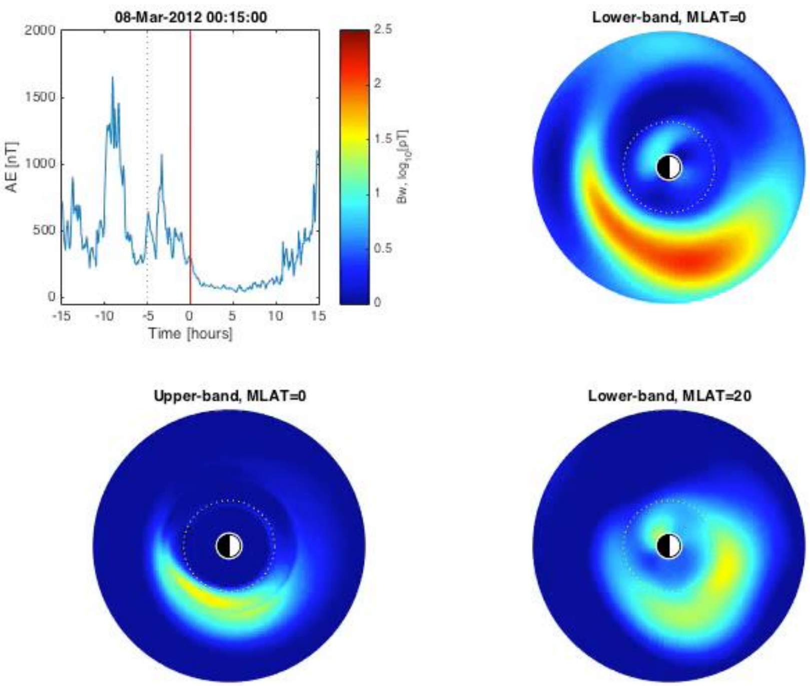
<u>Whistler-mode chorus waves</u>

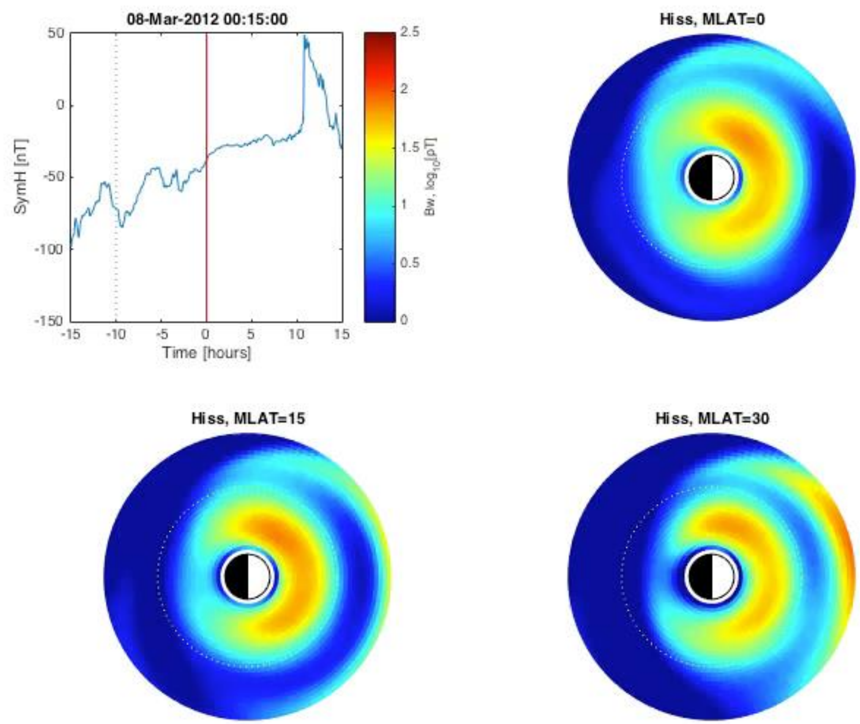
Upper band (0.5-0.8 fce) and lower band (0.1-0.5 fce) waves, measured on THEMIS and RBSP, ~372k samples, May 2010-June 2014.

<u>Plasmaspheric hiss waves</u>

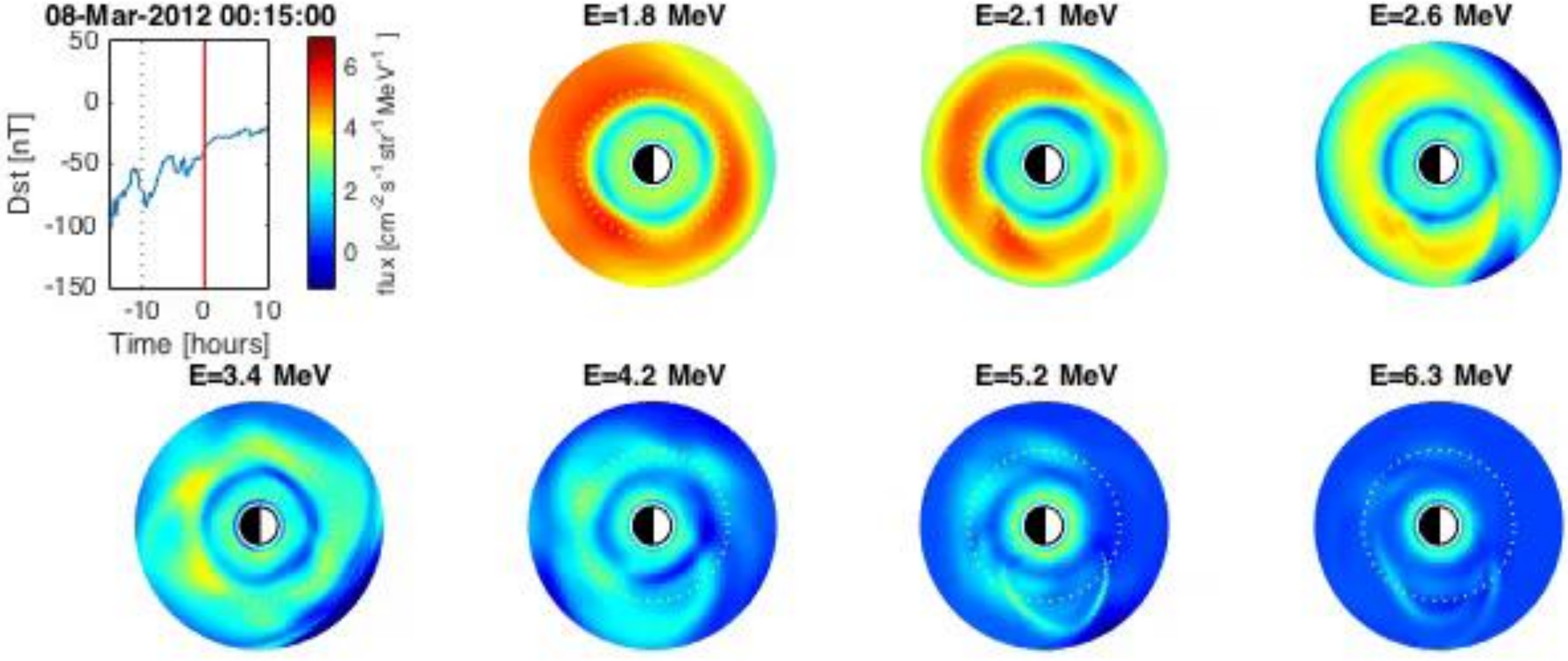
Van Allen Probes data, EMFISIS 0.1-2 kHz Bw; Oct 2012-Sep 2014, ~280k samples.

Regressed on 10-hrs of sym-H



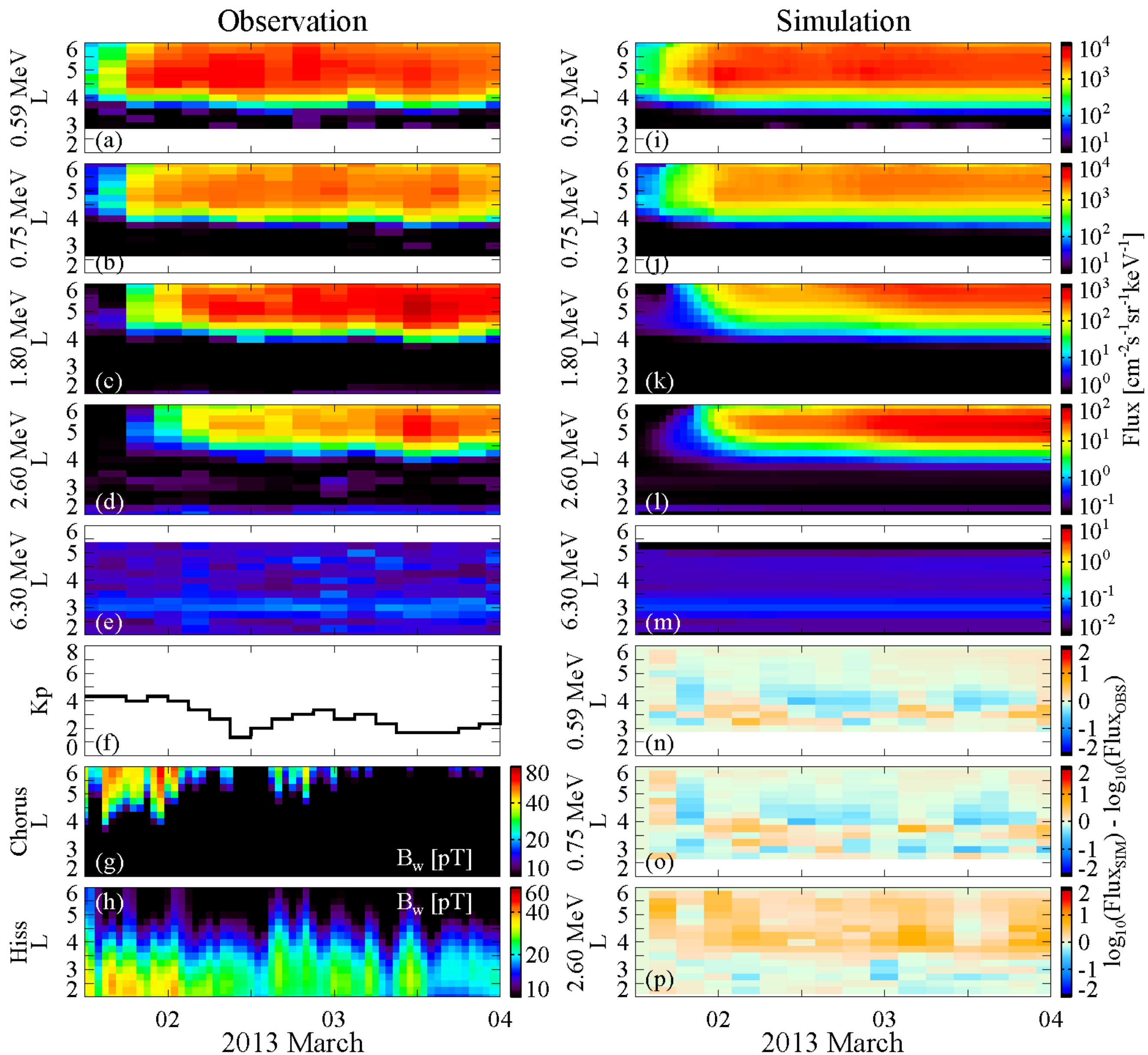


Relativistic electron (MeV) fluxes



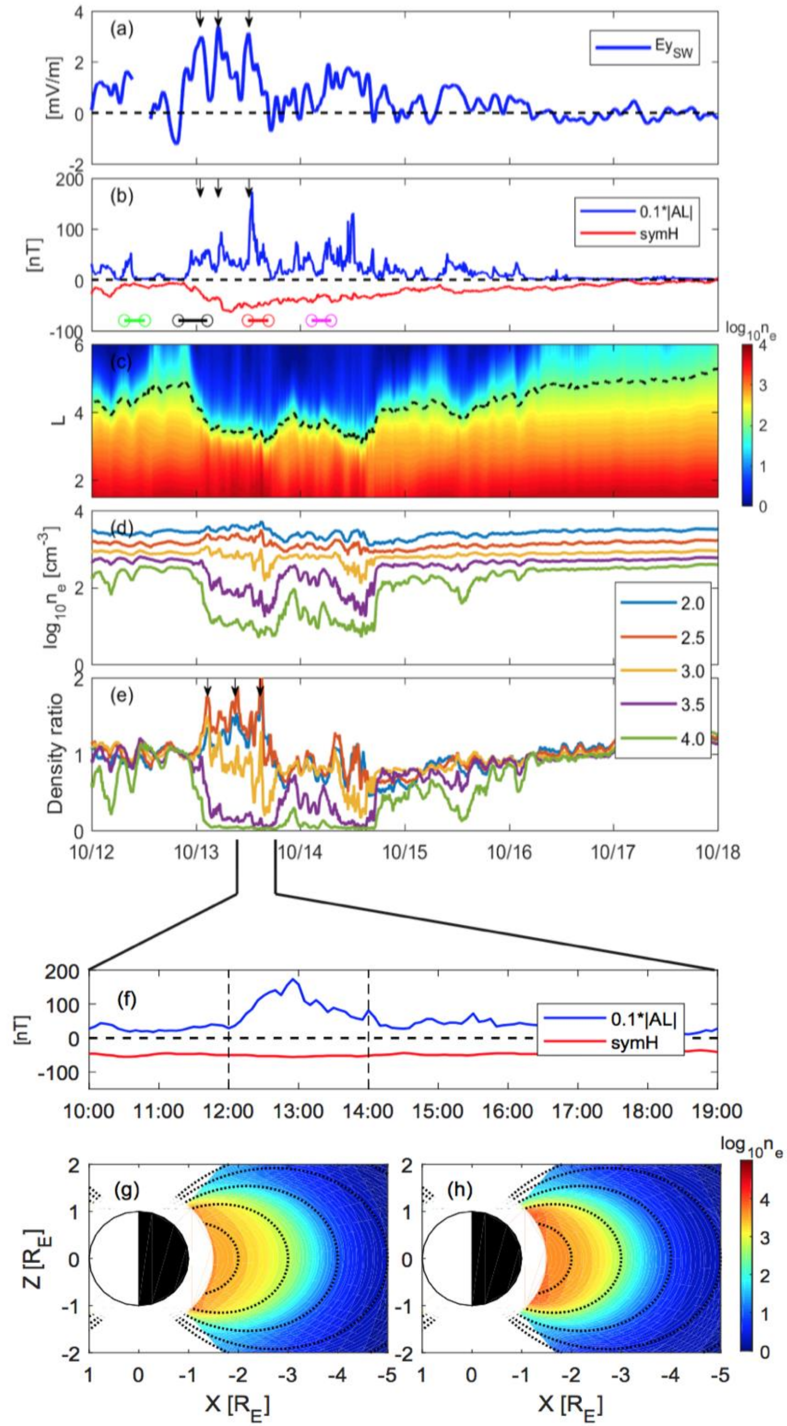
- REPT data: 8 energy channels 1.8, 2.1, 2.6, 3.4, 4.2, 5.2, 6.3, 7.7 MeV
- Regressed on 10 hours of Dst only
- Small number of samples, ~188k in total. Artifacts show up in higher energy channels since few accelerations reach those energies!

Application in modeling boundary and driving conditions



"Learned" chorus and hiss wave environments can be used as an input to (physics-based) Fokker-Planck models to predict ~MeV flux, i.e., **use data-rich environment to map into data-starved environment.**

Insight discovery: low L-shell density enhancements



- DEN3D (3D neural net density model) discovered enhancement at low L-shells in associationed with substorm injections
- Verified by independent satellite and modeling measurements

Summary

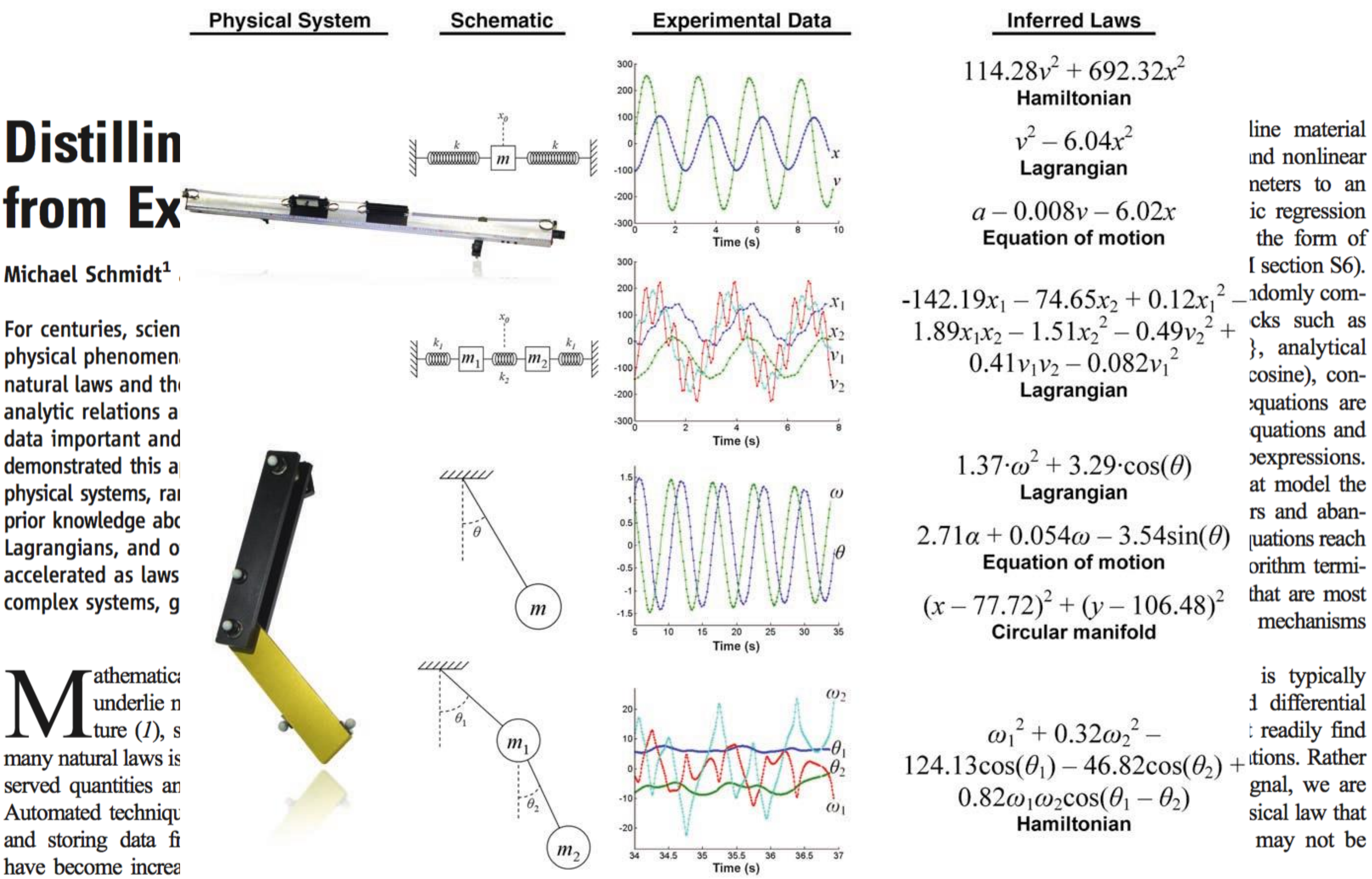
- **Scientific data growing: new approach needed to extract "science".**
- We presented a "unified approach to inner magnetospheric state prediction" *Bortnik et al.*, [2016] JGR
- Take a set of observations of some quantity Q measured at (**r**, t), and reconstruct Q at all **r** as a function of t. Q can be anything, e.g., density, energetic particle fluxes, and different wave modes.
- Preliminary results show excellent 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.
  - 3D plasmaspheric density model, evolves on 5 min cadence
  - Hiss and chorus 3D wave fields
  - REPT (~MeV) fluxes cannot be learned directly: data starved environment, need to translate from data rich (waves, keV particles) to data starved environment.
- For specification models, input to physical models, insight discovery

THE END

Summary

- Geospace in the modern era is exciting!
  - Unprecedented observational resources and data
  - Space weather is becoming an increasingly important consideration for our modern, technological society.
- We have made key contributions to our field in understanding the interconnectedness of waves, wave-particle physics, and machine learning
- Many questions remain that are of scientific importance and societal interest.
- A key area of expansion is in the exploitation of large, heterogenous datasets to extract insights, fundamental physics, and achieve specification/prediction.

Looking for fundamental physics



Big data/Machine learning

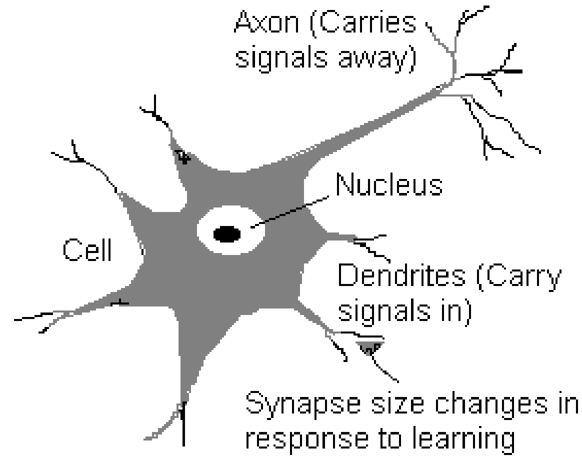
- What it is:

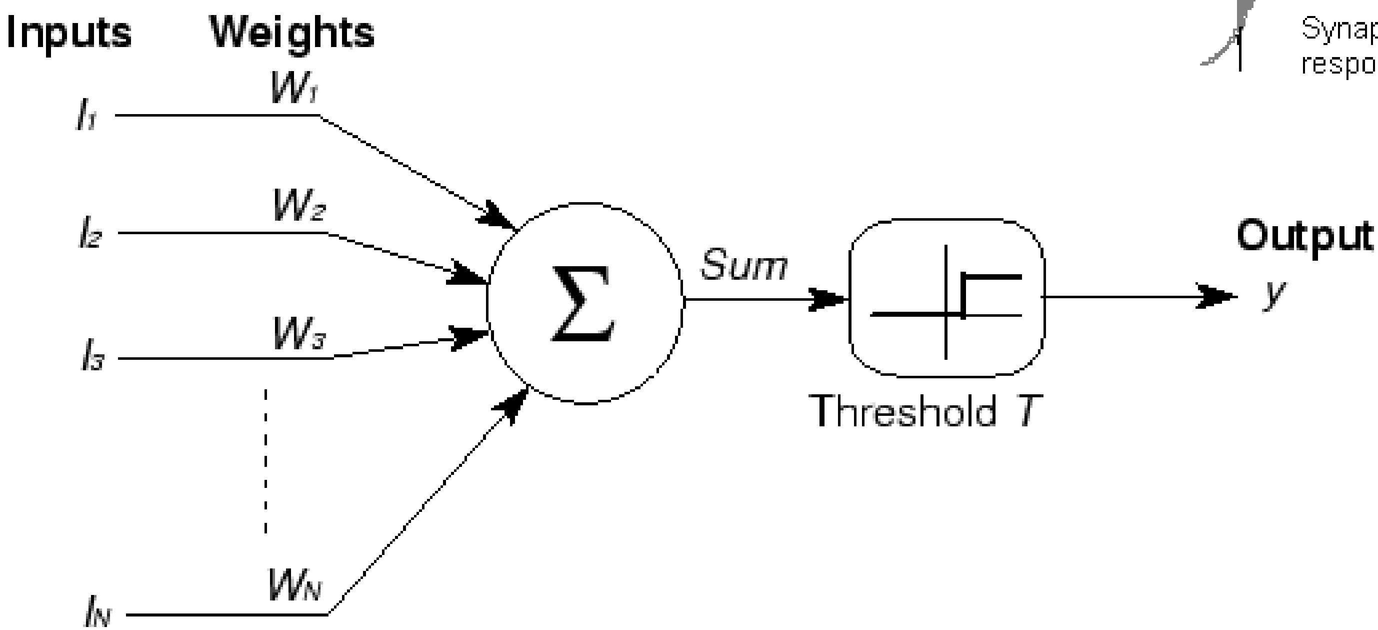
  A collection of algorithms that are able to autonomously 'learn' or infer subtle patterns in data without being explicitly programmed to do so.

- What it is not:

  It is not a substitute for measurement, modeling, or physical understanding, e.g., "*The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, by Chris Anderson [WIRED, 2008/6]". It acts as a complement, discerning patterns in large volumes of data that cannot be achieved with the naked eye.

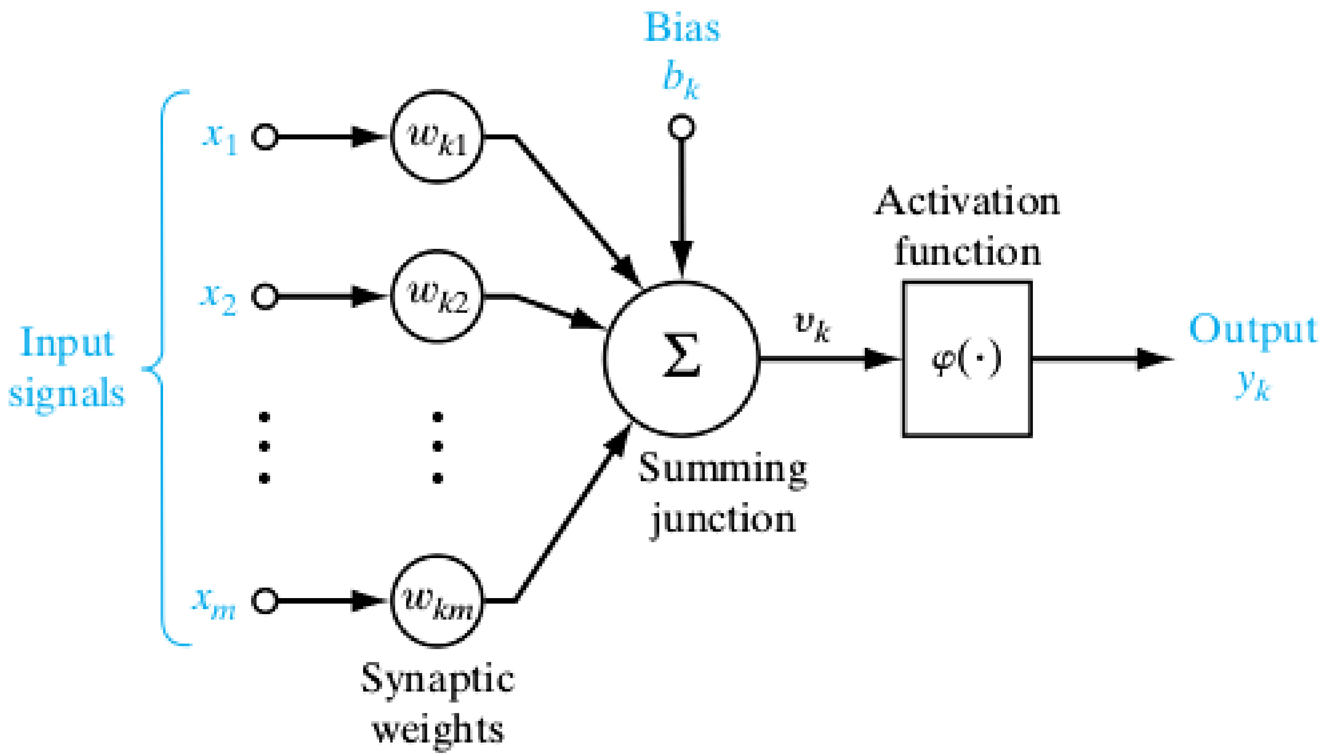
Background on neural network models



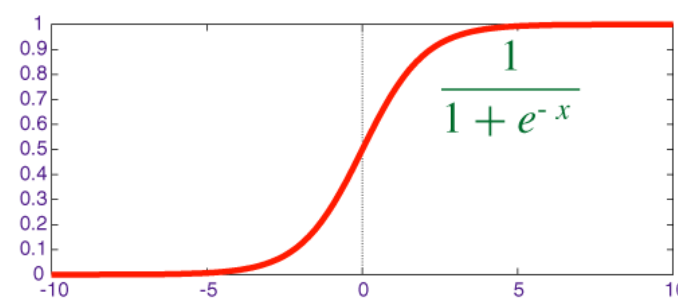


- McCullough and Pitts neuron model, 1943 “perceptron”
- Replicates biological neuron
- Uses a simple step thresholding function
- Problem with convergence of the learning rule (step discontinuity)

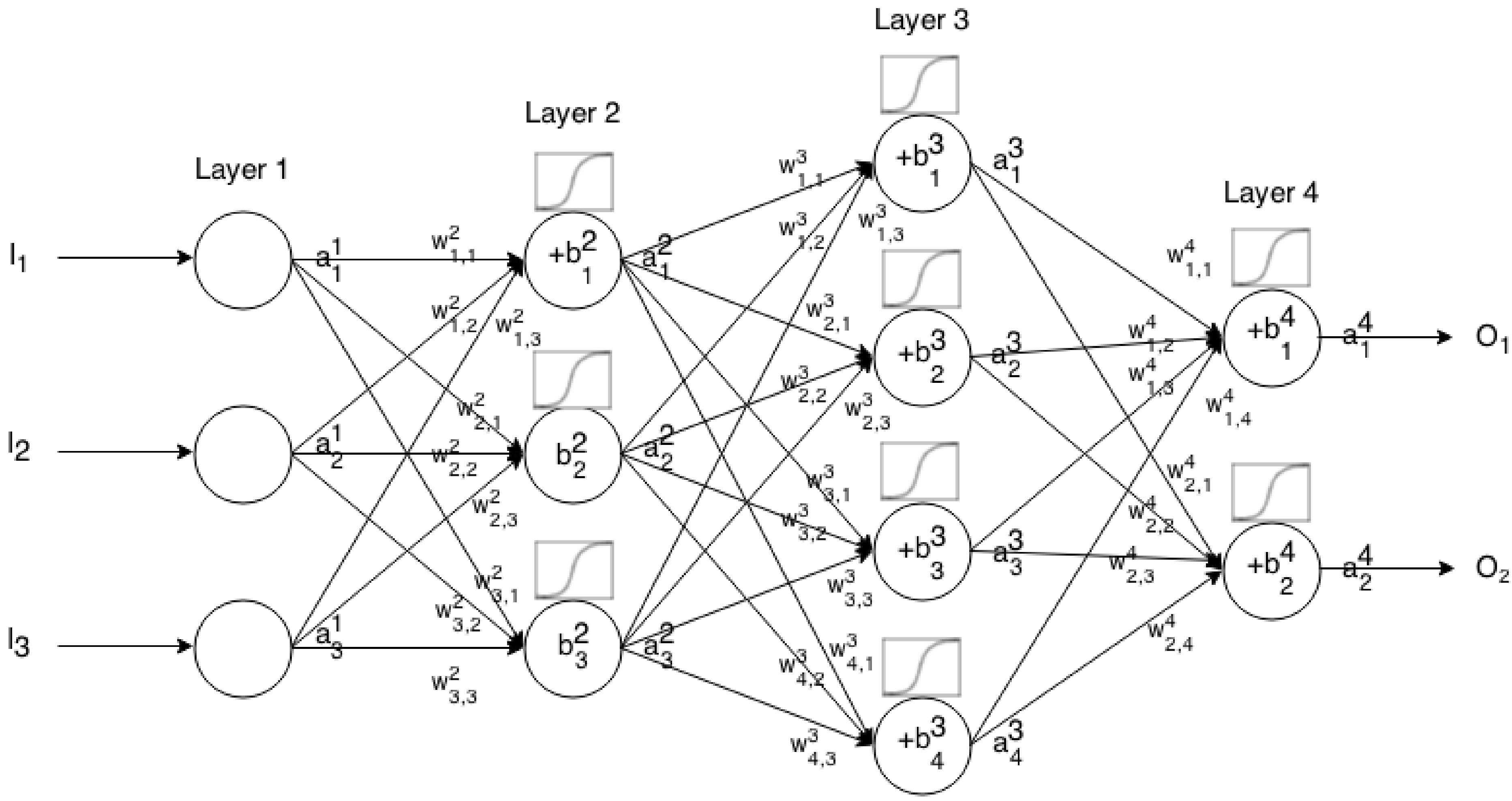
General neuron model



$v_k = [b_k=1, x_1 \ldots x_m][w_{k0} \ldots w_{km}]^T$
$y_k = \Phi(v_k) = 1/(1 + \exp[-v_k])$
Learning by gradient descent



Neuron "networks"



- Connect individual neurons together to form a network.