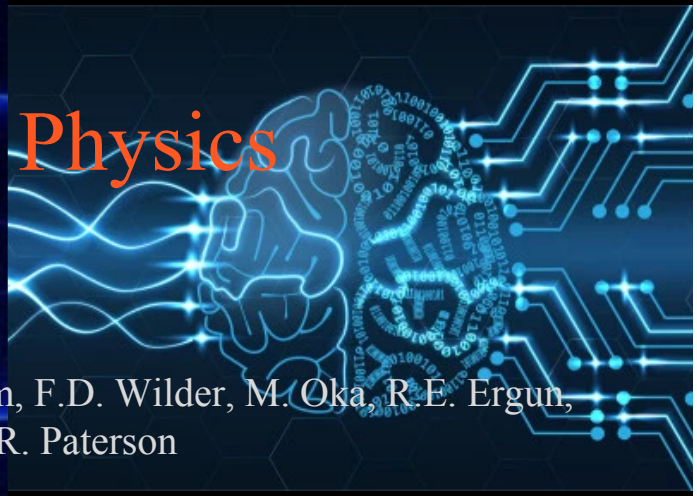
Machine Learning in Space Physics

M.R. Argall, S. Piatt, C. Small, M. Petrik, K. Larson, K. Kokkonen, J. Barnum, F.D. Wilder, M. Oka, R.E. Ergun, T.-D. Phan, B.L. Giles, R.B. Torbert, J.L. Burch, W.R. Paterson



matthew.argall@unh.edu

Matthew Argall
matthew.argall@unh.edu

Special thanks to the GEM Student Representatives

Apply for a GEM Focus Group on Machine Learning in Space Physics

Outline

Machine Learning

- What is it?
- Milestones
- Why it Matters
- What to use
- Applications in Space Physics
- Detailed Examples Using MMS Data:
  - The Problem: Selecting burst intervals in MMS data
  - Recurrent Neural Networks
  - Hierarchical Bayesian Model
    - Expectation Maximization
    - Autoregression
    - Linear regression

Machine Learning

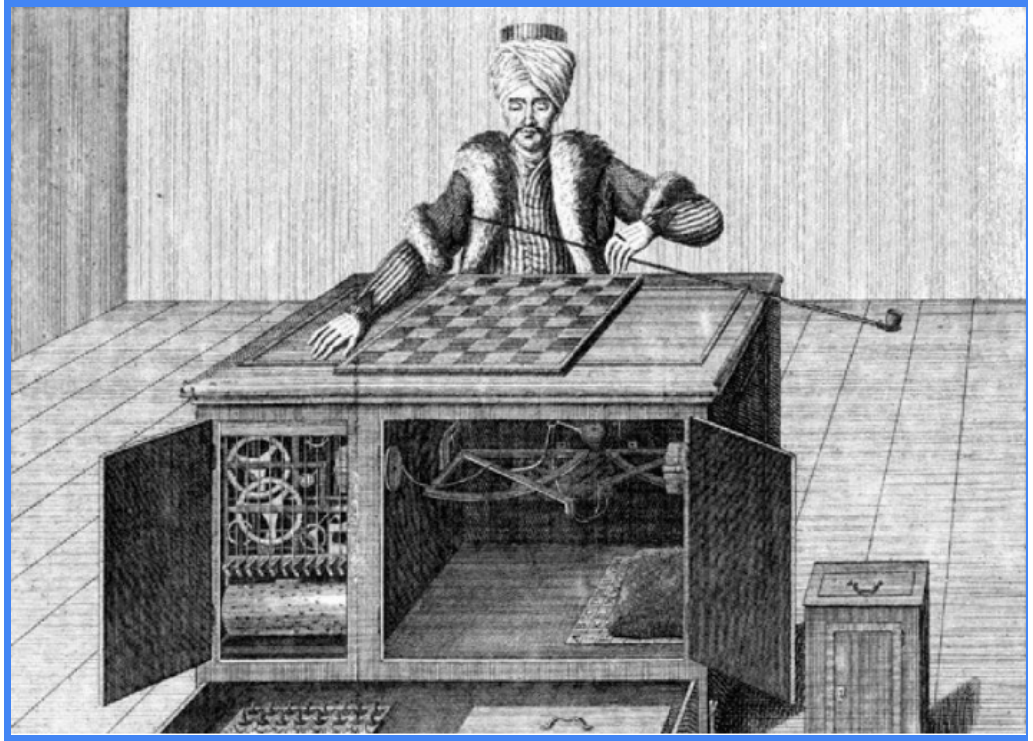
What is it?

Machine Learning is the science of getting computers to learn and act like humans do, and improve their learning over time in autonomous fashion, by feeding them data and information in the form of observations and real-world interactions.

Milestones in Machine Learning

The Turk

A chess-playing automaton built in 1770



Inspired Charles Babbage to design the Punchcard Computer

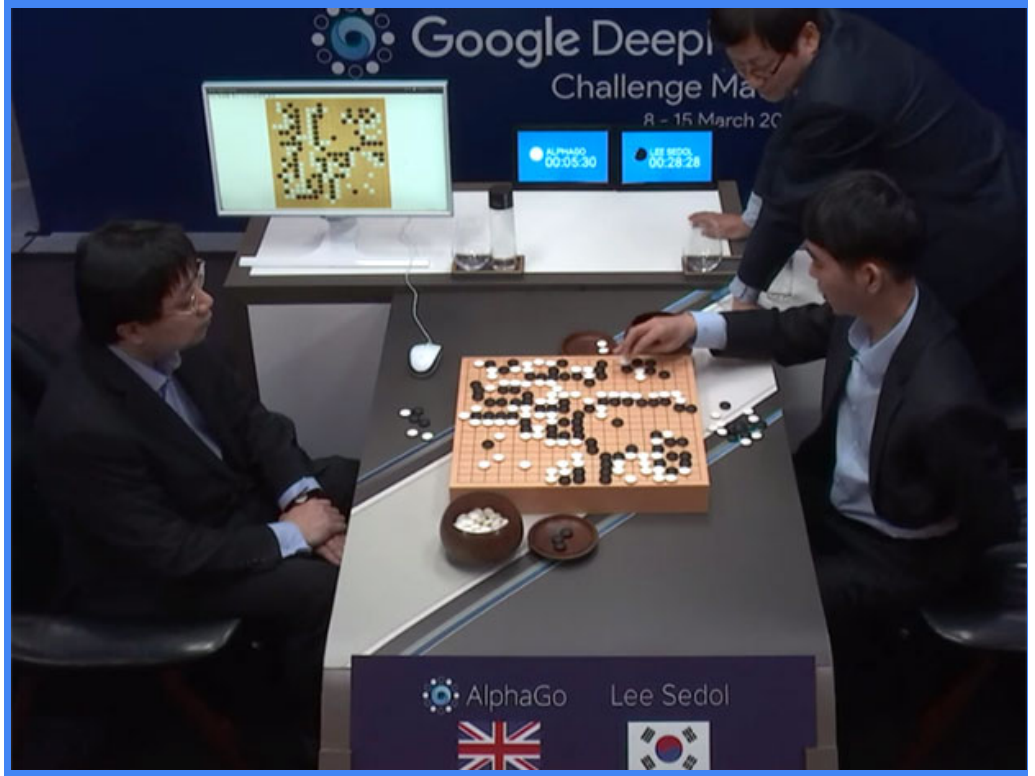
Computer Checkers

Frist application of machine learning by Arthur Samuel (IBM, 1959)

Field of study that gives computers the ability to learn without being explicitly programmed

AlphaGo

Computer beats humans at Go (2015)



A step closer to Terminator-like video game characters

Artificial Brain

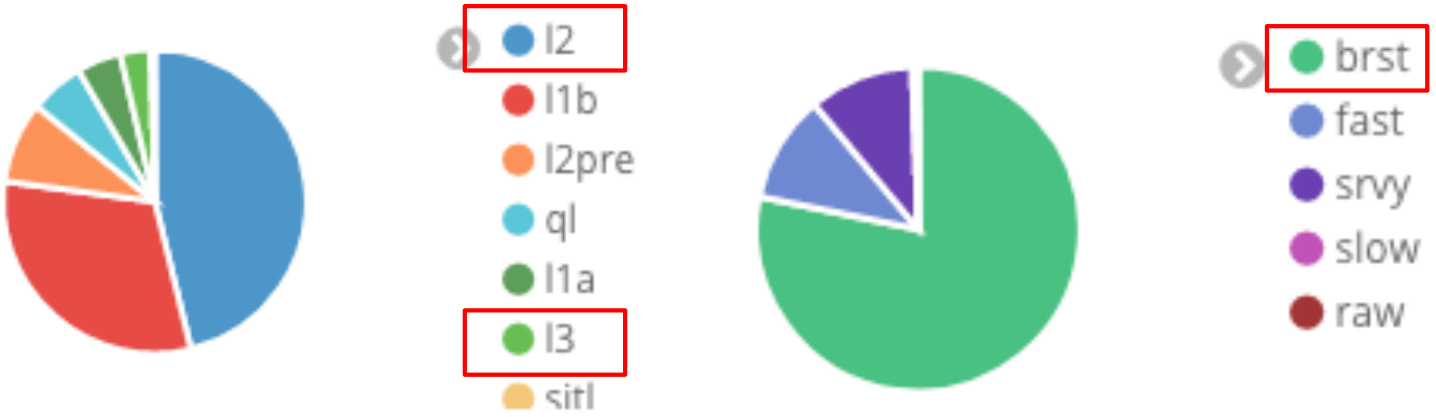
Learns to identify cats by watching YouTube videos (2012)

Computer *learns* like a human

Why it Matters

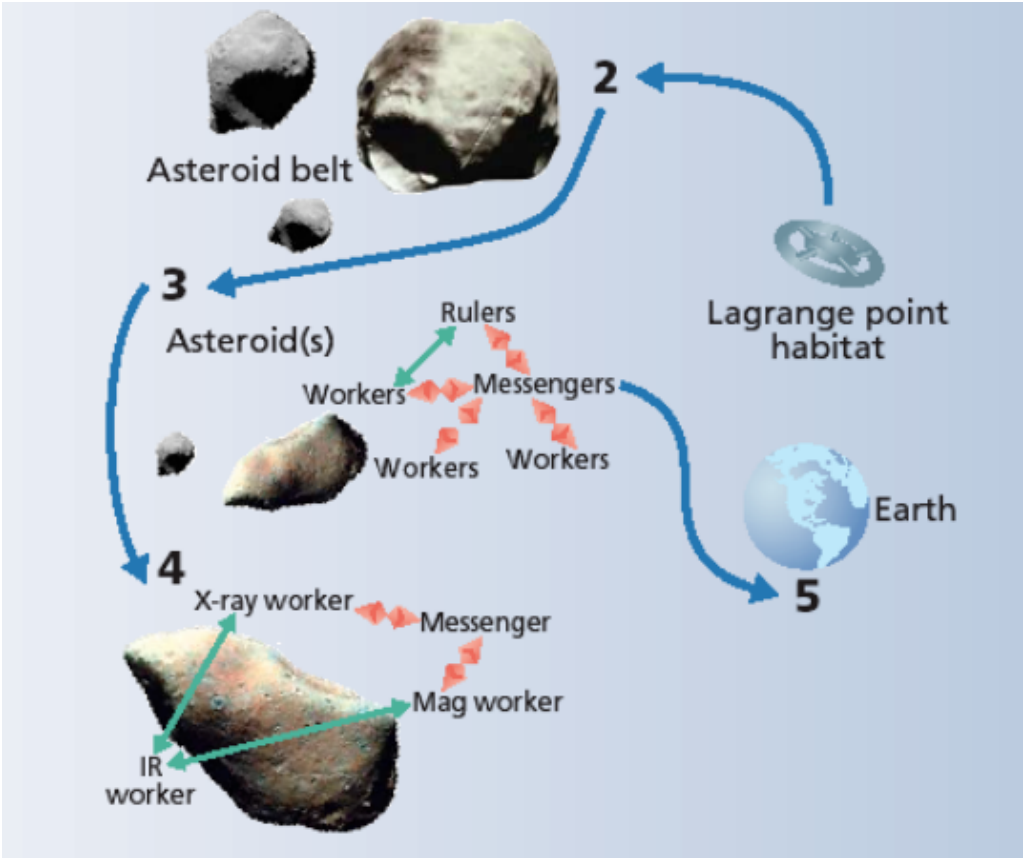
Too Much Data for One Lifetime

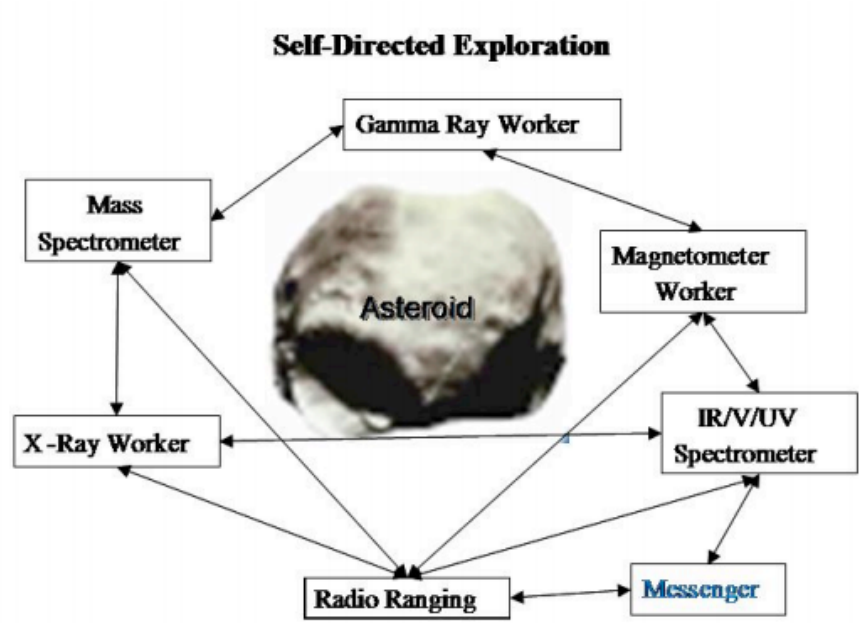
MMS: **9.4** Million Files **152.6** TB Used In ~4 years!



Machine Learning can make it tractable

Automating operations to enable more complex mission designs

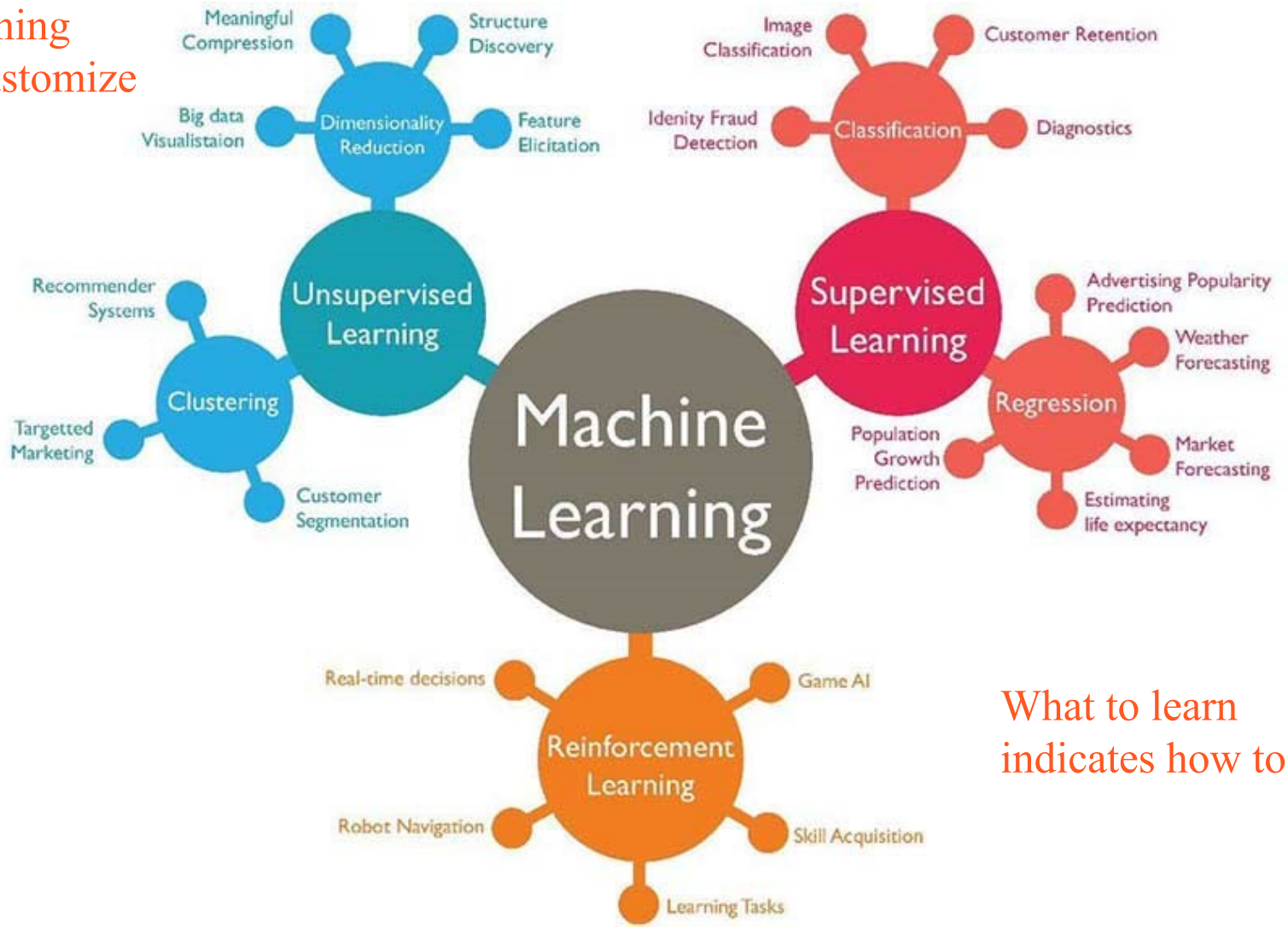




NASA's Autonomous NanoTechnology Swarm (ANTS) explores asteroids

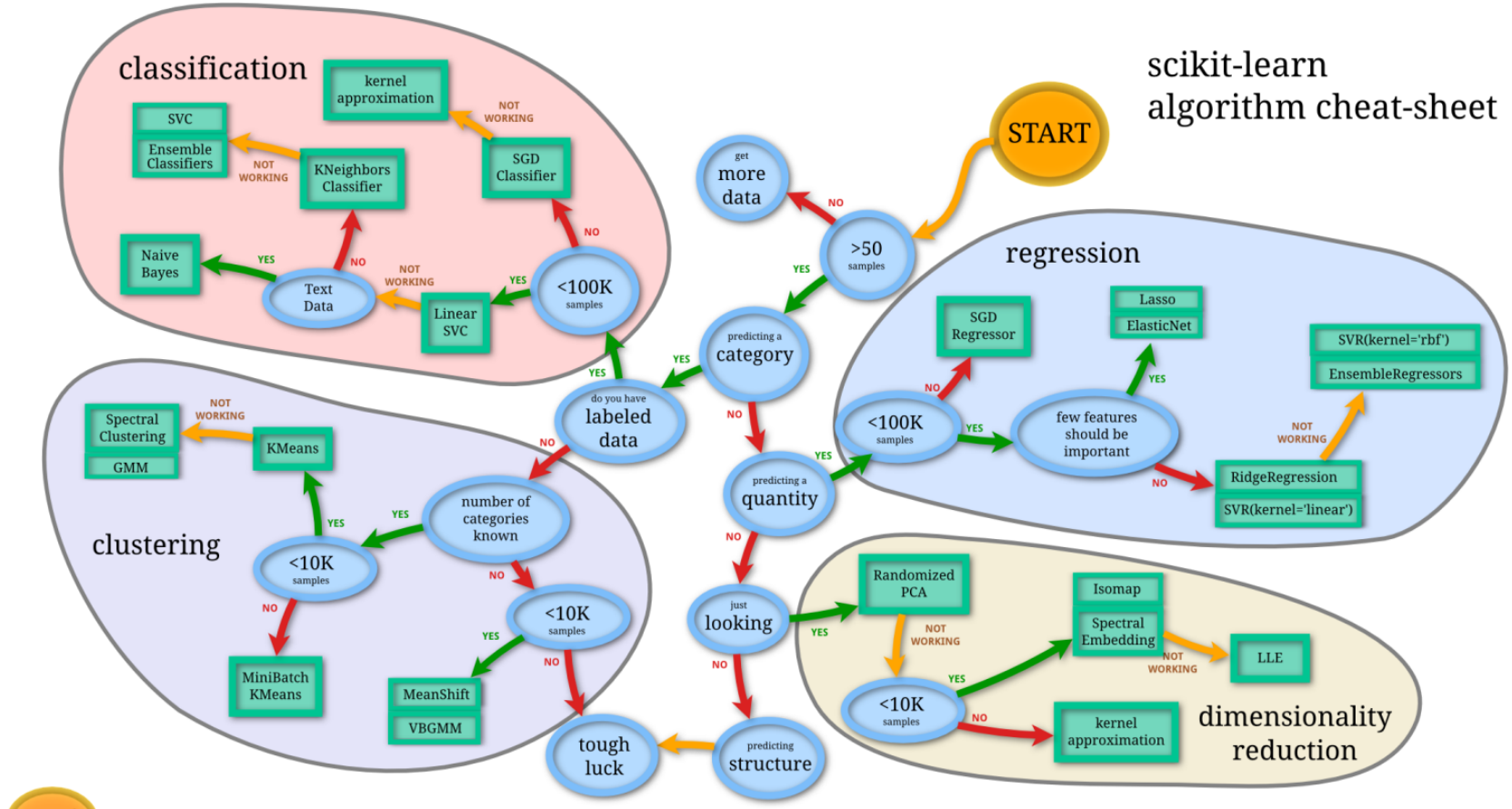
What to Use

Different learning methods to customize use



What to learn indicates how to learn.

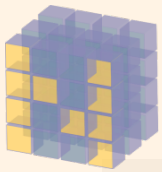
Cheat sheet for choosing initial approach



Back

Follow the workflow to find a starting point.

Packages for Python and R



NumPy

SciPy

Pandas

theano

TensorFlow

Keras

PyTorch

HelioPy

sunpy
A Community Python Library for Solar Physics

astropy

matplotlib

mlr

caret

e1071

rpart

nnet

Data Handling

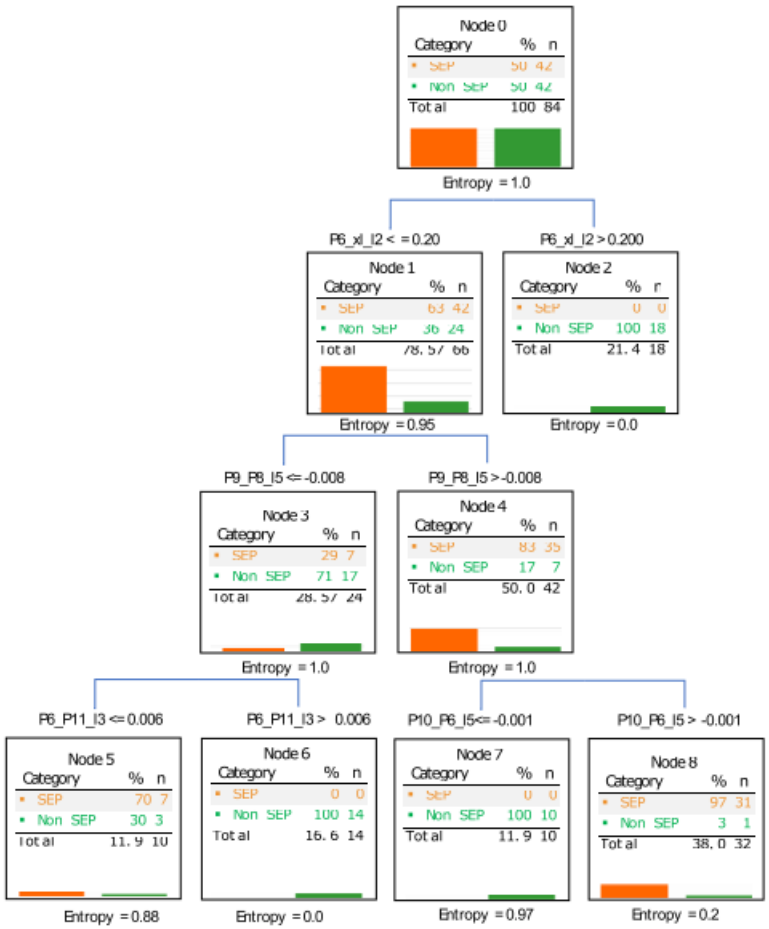
Machine Learning

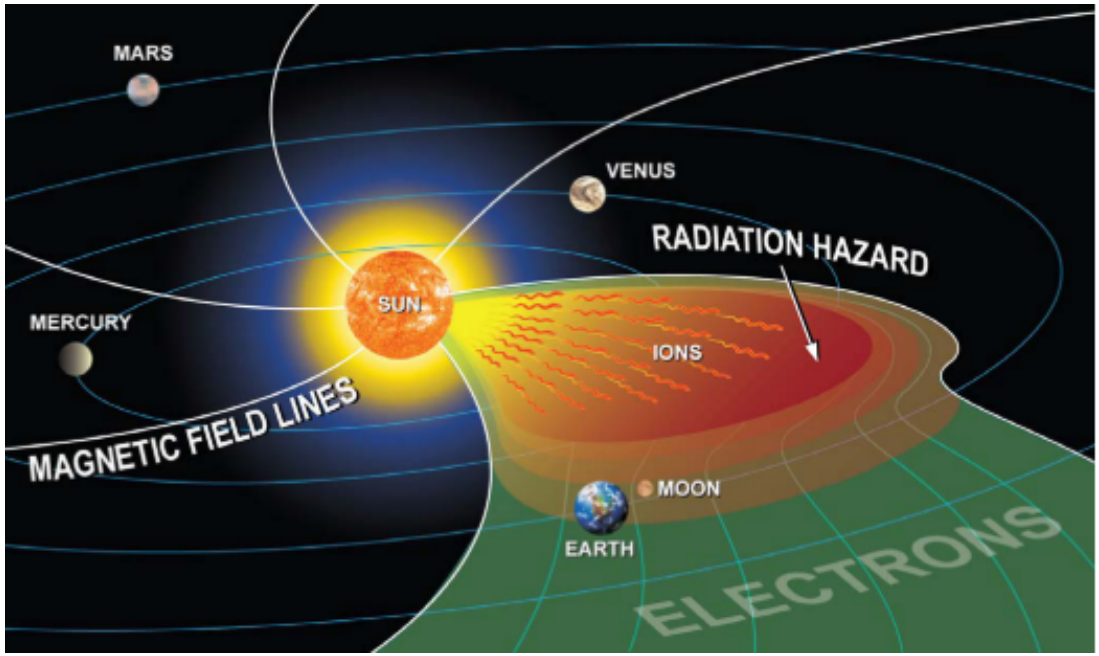
Data Visualization

Applications in Space Physics

Decision Trees are easy to understand

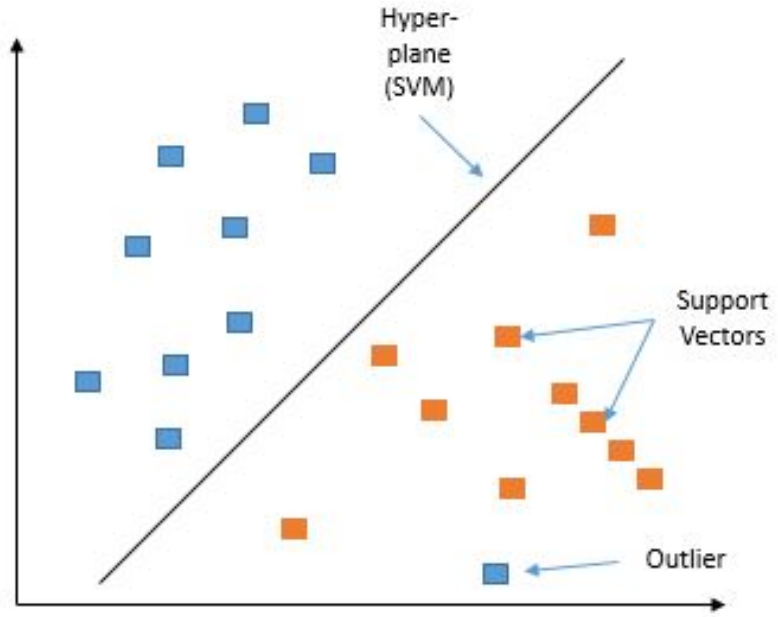
Boubrahimi, et al., 2007

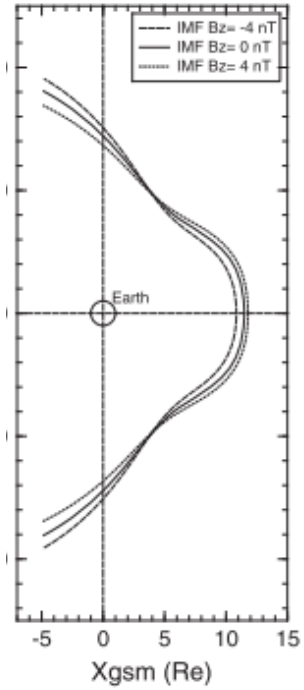


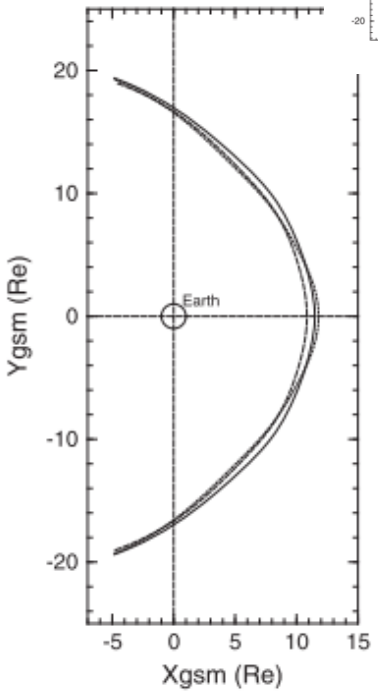


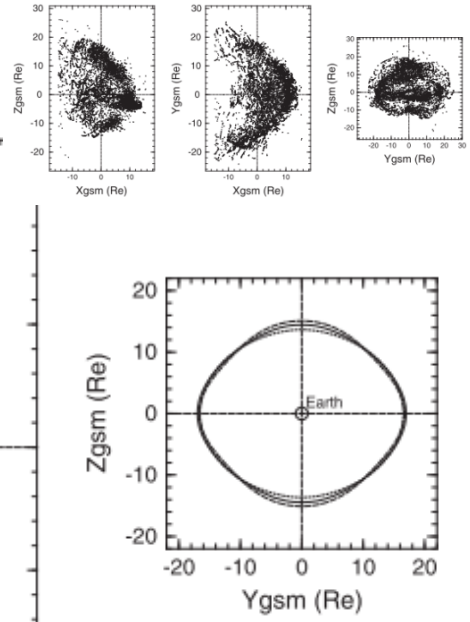
Decision Trees can be used to predict solar energetic particle events using SOHO & GOES

Support Vector Machines classify data into categories







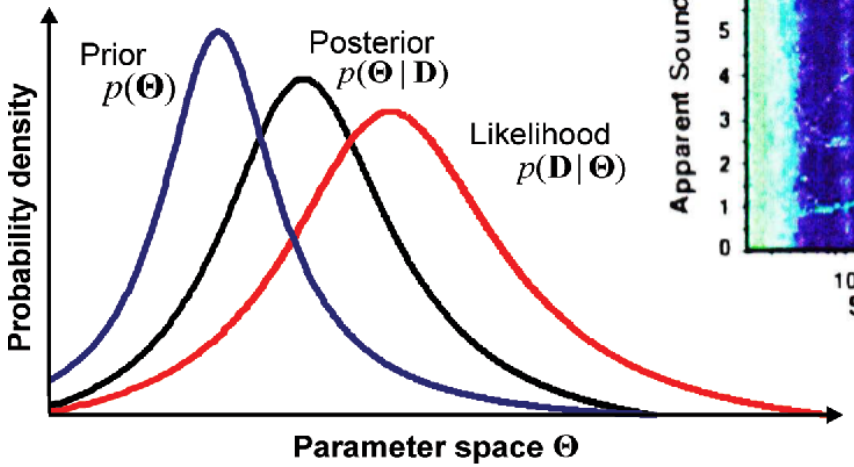


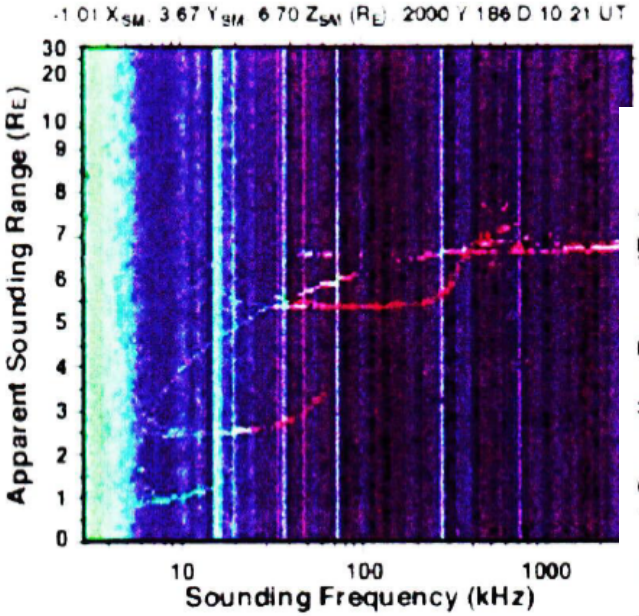
[LibSVM]

Build a 3D magnetopause model using 15,089 magnetopause crossings from 23 different satellites

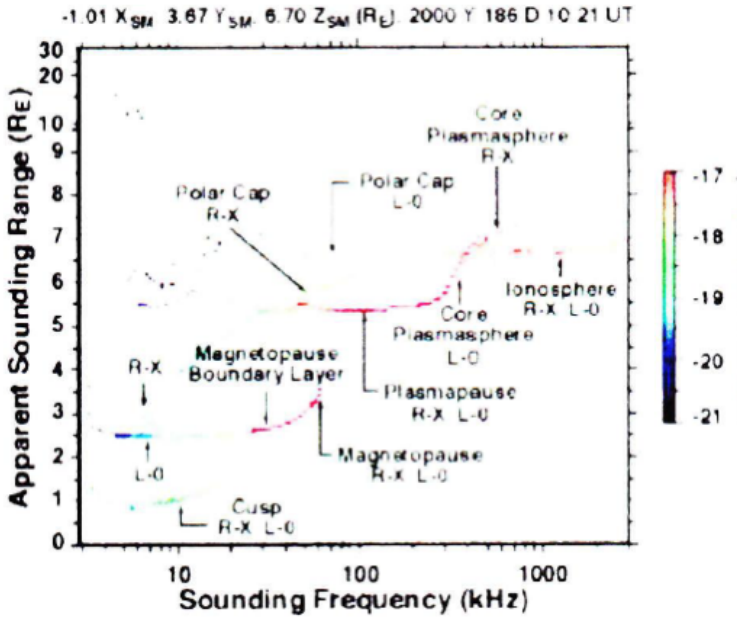
[Wang, et al., JGR, 2013]

Bayesian analysis uses probabilities to make inferences about new data





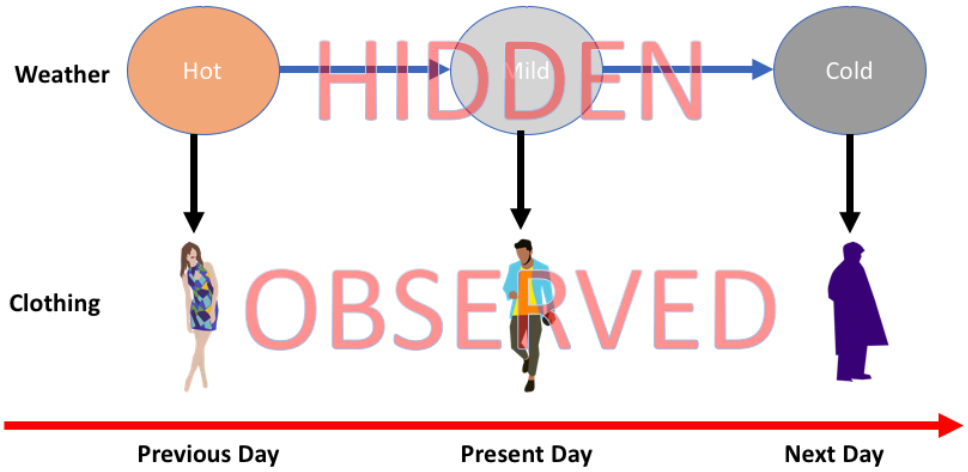
Rilee & Green, 2000

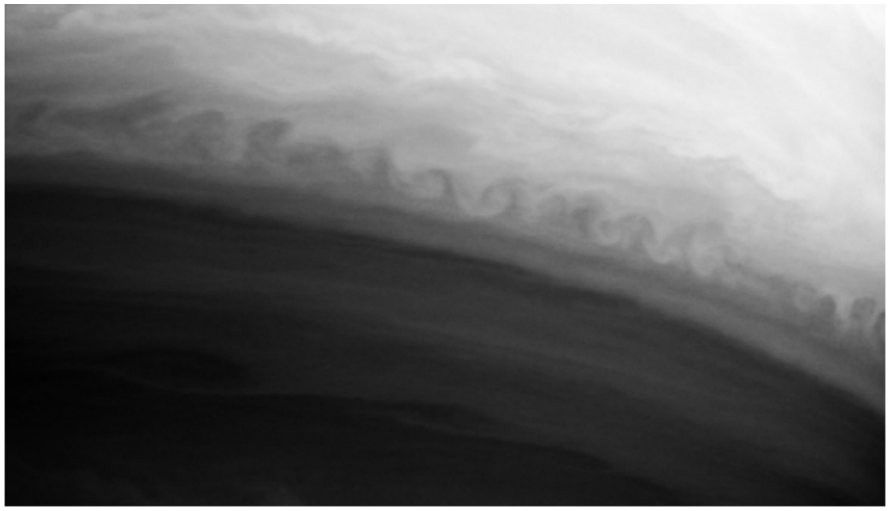


Extract signatures of the magnetopause boundaries with low SNR from IMAGE data

Hidden Markov Models predict the future using the present

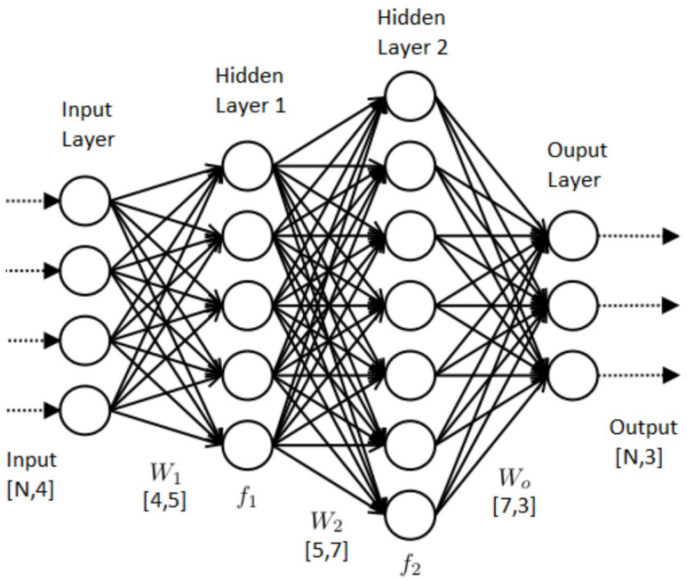
Lundgren, KTH (PhD), 2011



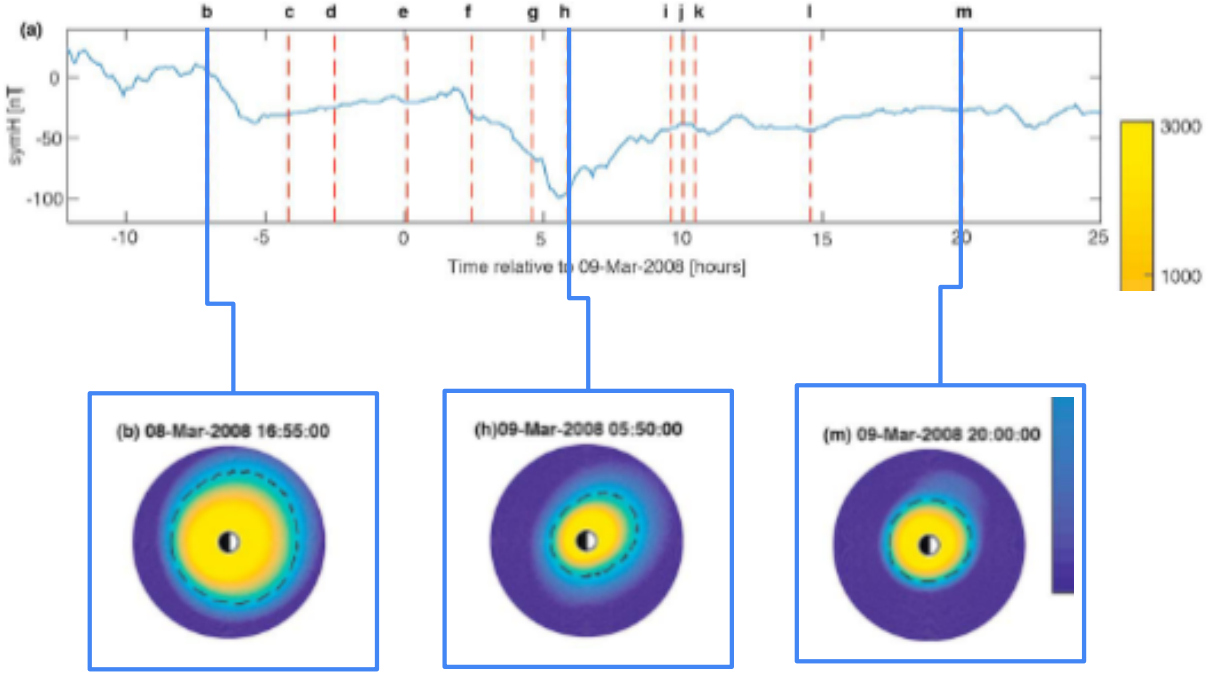


Identify Kelvin-Helmholtz waves in Geotail data

Neural networks mimic the brain's processing functionality



Bortnik, et al., JGR, 2016

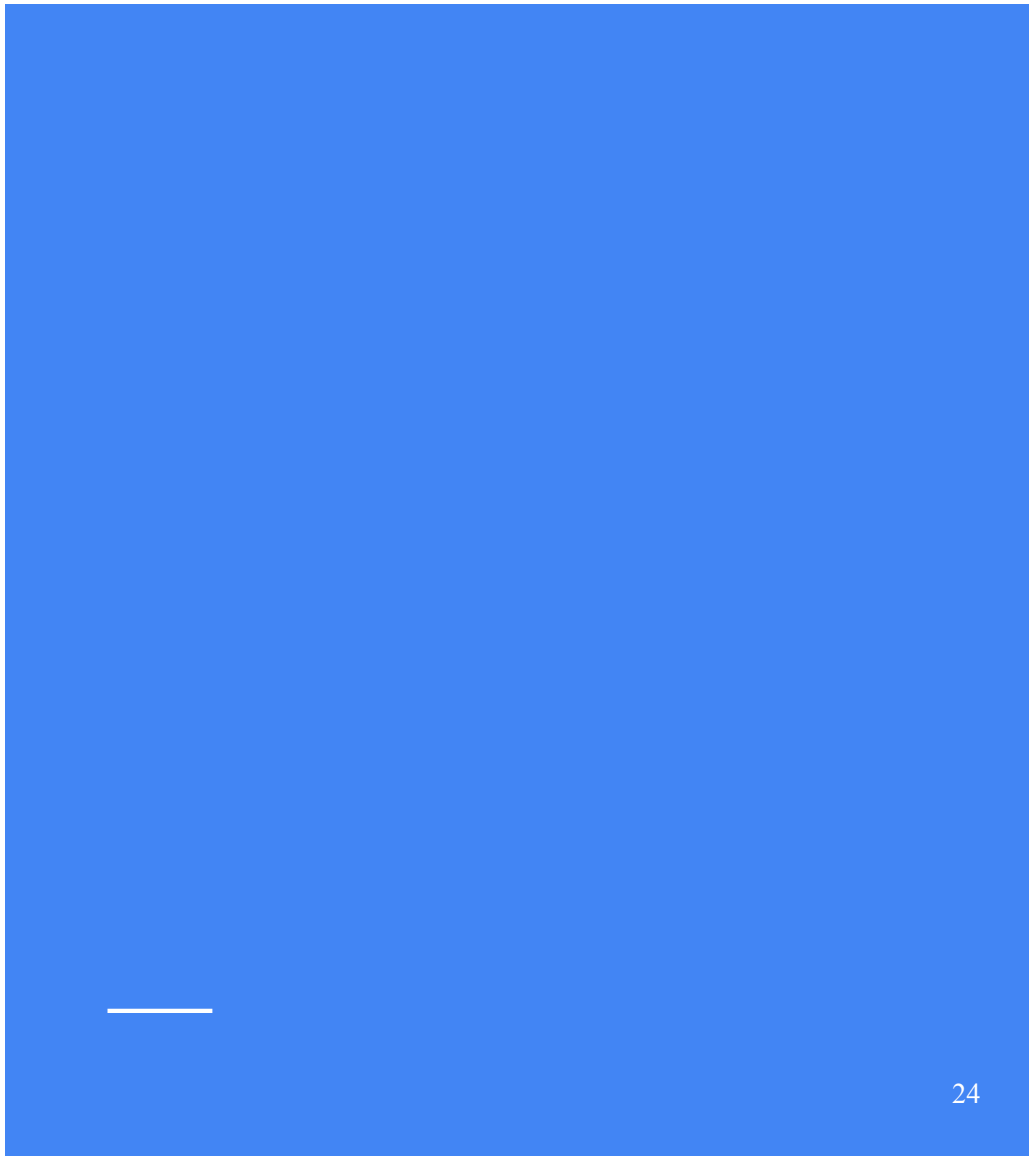


Used to predict state of inner magnetosphere using SYM-H and THEMIS data

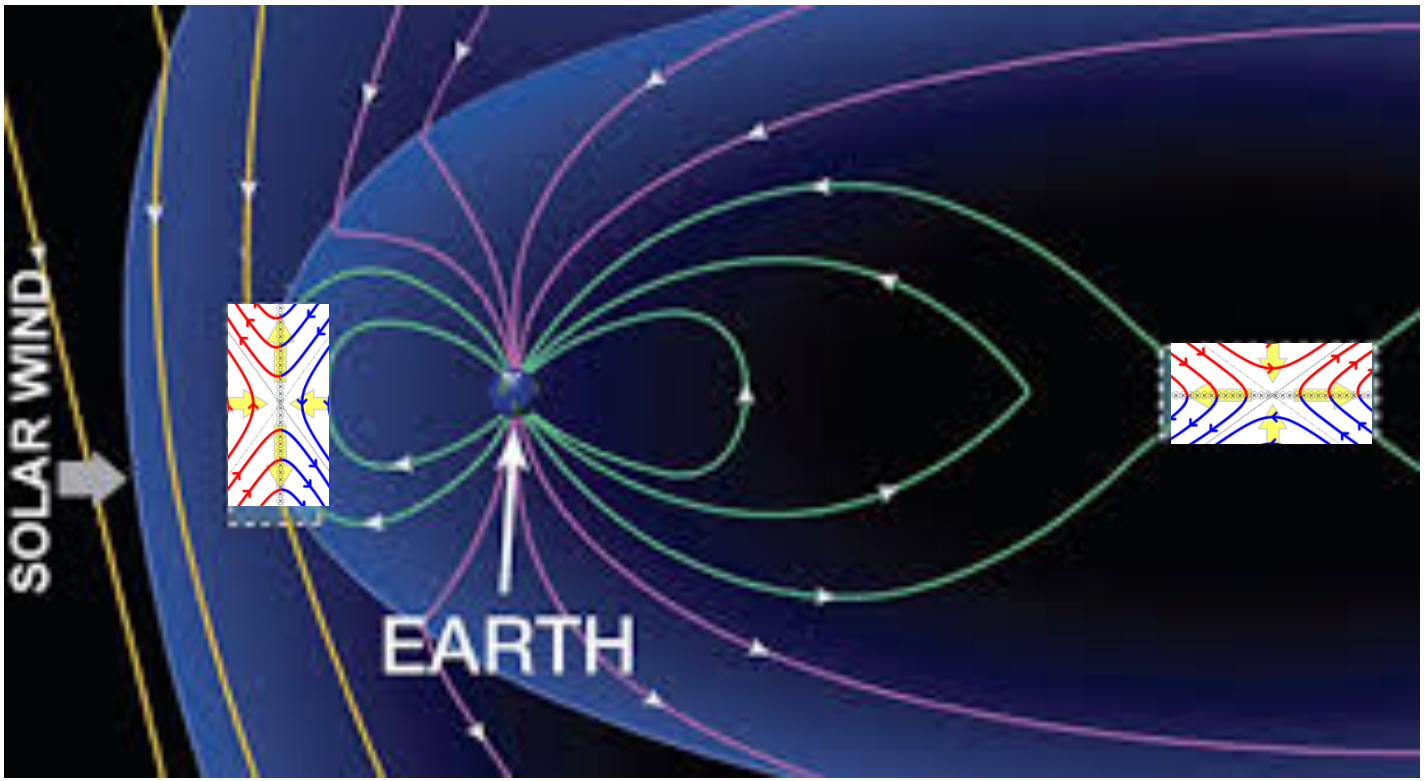
Detailed Examples Using MMS Data

The Problem

Selecting burst intervals in MMS data

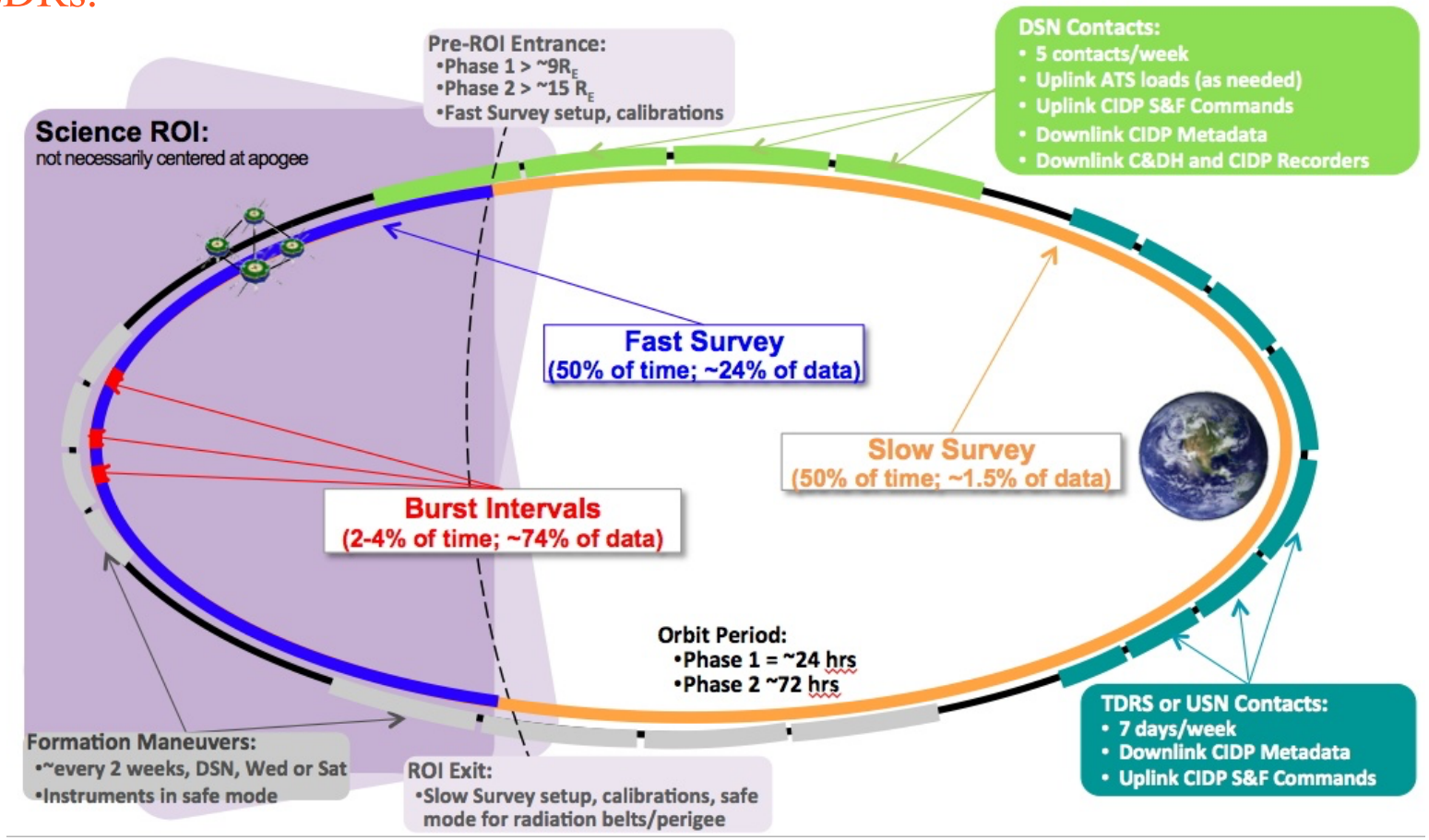


Magnetic reconnection drives the flow of energy throughout the magnetosphere



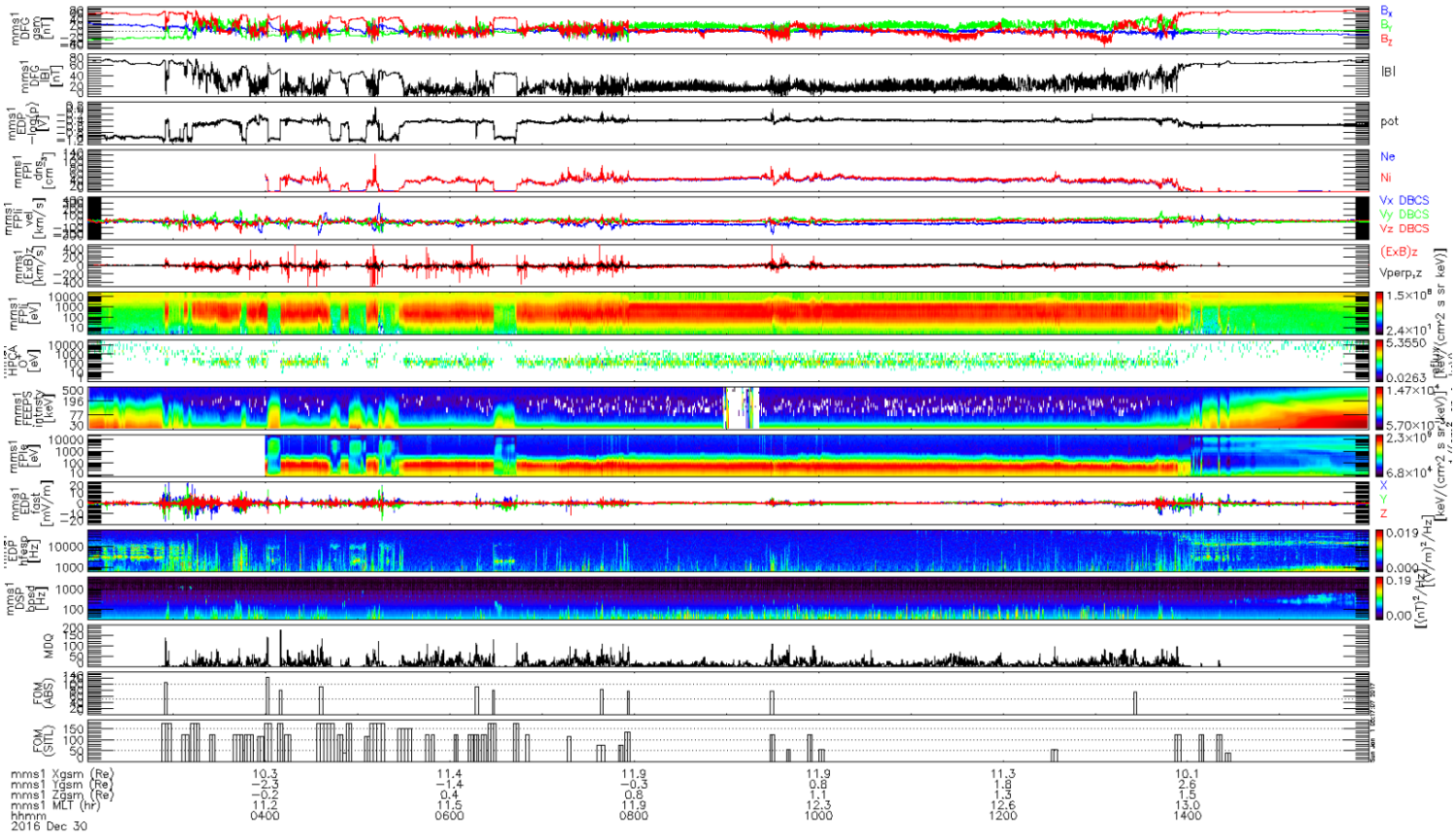
It is triggered by the microphysics of the Electron Diffusion Region (EDR).

The Region of Interest (ROI) encompasses locations likely to contain EDRs.



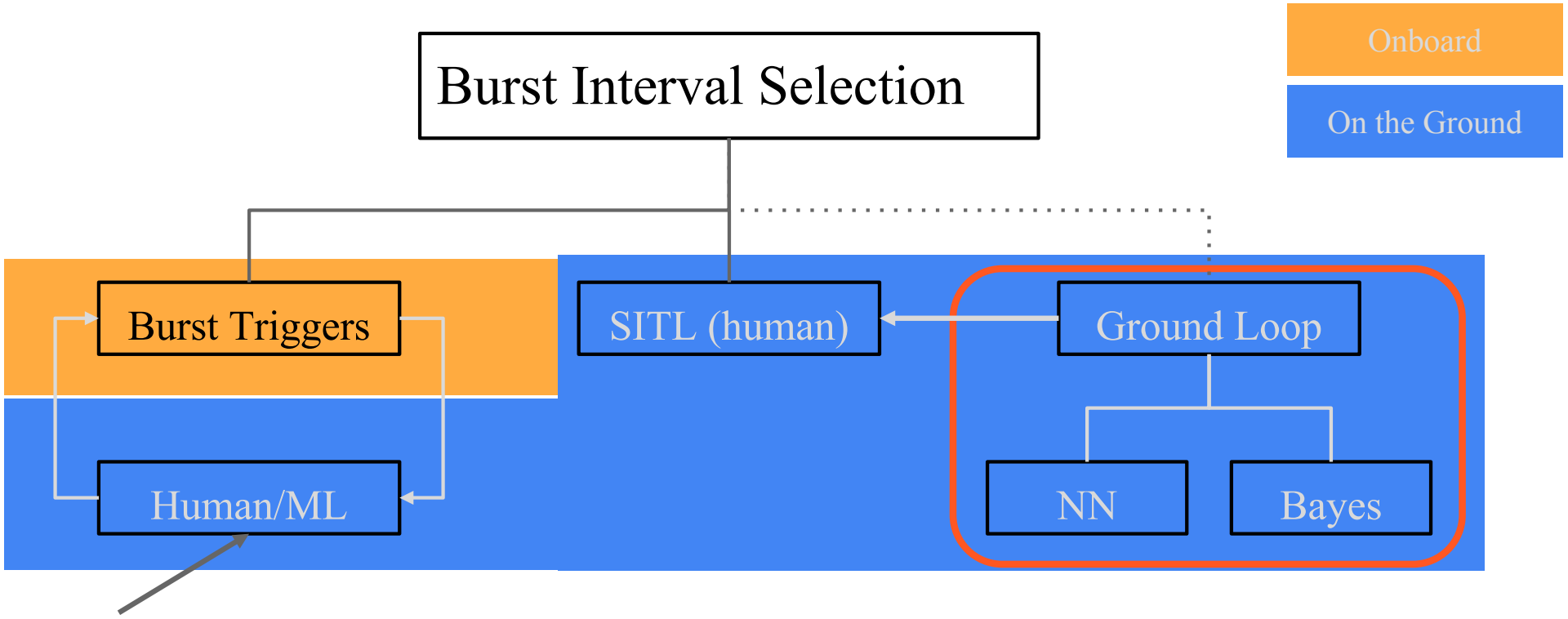
Only enough telemetry to downlink ~43 minutes of burst data per day

Burst Selections are made primarily by the Scientist in the Loop (SITL)



Need to automate the selection process

Three methods to select burst intervals



Bill Paterson

I will discuss ML and the ground loop

Small, C., In Prep

ML and MMS

Burst Triggers

Tune Burst Triggers

Equation 1. Cycle Data Quality (CDQ) Calculation

$$CDQ = \sum_{All}\left[TDN(i) + O(i)\right] * G(i) \Big/ \sum_{All} G(i)$$

Tail ABS Table

TDN	DESCRIPTION	GAIN	OFFSET
dB	|dBx|+|dBy|	1	0
dBz	|dBz|	2	0
1/|B|	B inverse	4	30
ERMS1	Power density 10 – 100 Hz	5	0
Var Ion X Flux	Variance, spinning coords	1	150
Var Ion Y Flux	Variance, spinning coords	1	150

Dayside ABS Table (MP019)

TDN	DESCRIPTION	GAIN, *G*	OFFSET, *O*
dBz	|dBz|	2	0
ERMS1	EOMNI, Power Density 10 – 100 Hz	10	0
dB	|dBx|+ |dBy|	2	0
Var Pe	Electron Pressure, Variance	1	0
Var Ni	Ion Density, Variance	1	0
Delta Ni	Ion Density, Delta of Mean	1	0

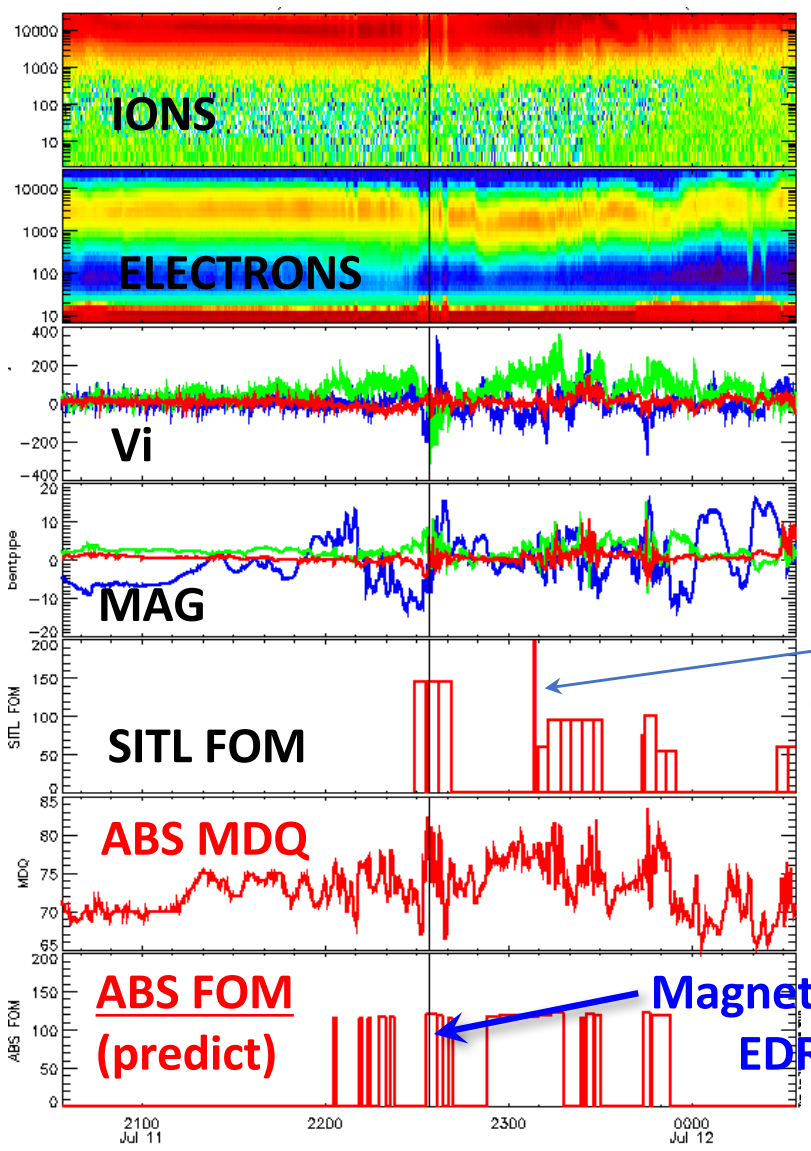
Identify EDRs

Equation 1. Cycle Data Quality (CDQ) Calculation

$$CDQ = \sum_{All}\left[TDN(i) + O(i)\right] * G(i) \Big/ \sum_{All} G(i)$$

Tail ABS Table

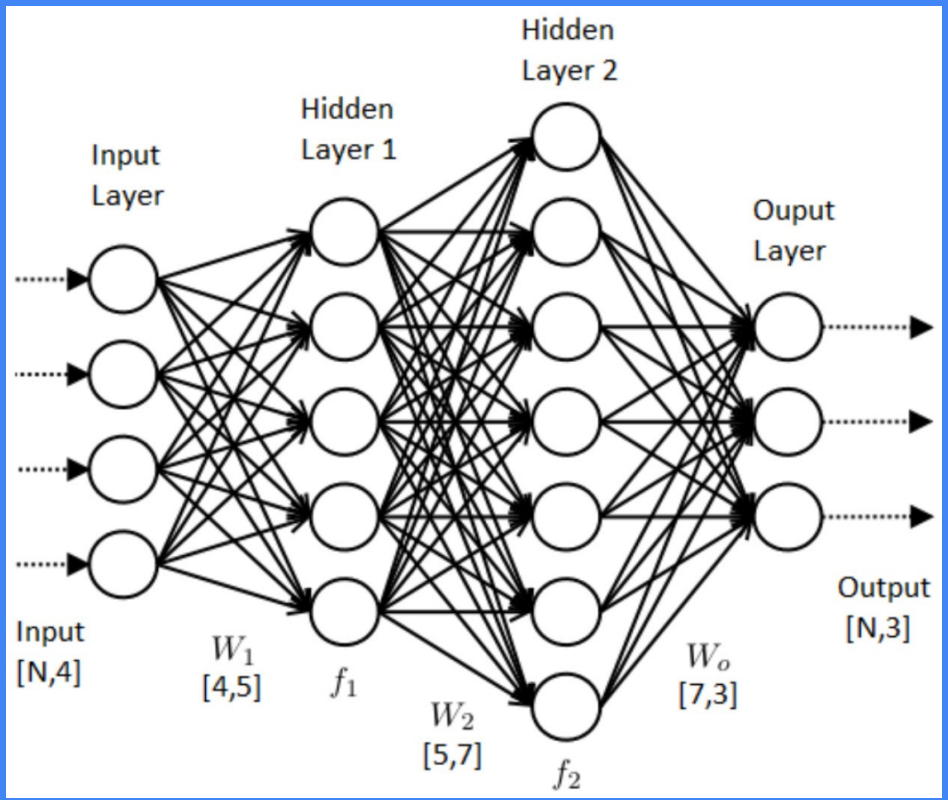
TDN	DESCRIPTION	GAIN	OFFSET
dB	|dBx|+|dBy|	1	0
dBz	|dBz|	2	0
1/|B|	B inverse	4	30
ERMS1	Power density 10 – 100 Hz	5	0
Var Ion X Flux	Variance, spinning coords	1	150
Var Ion Y Flux	Variance, spinning coords	1	150



Small, C., In Prep

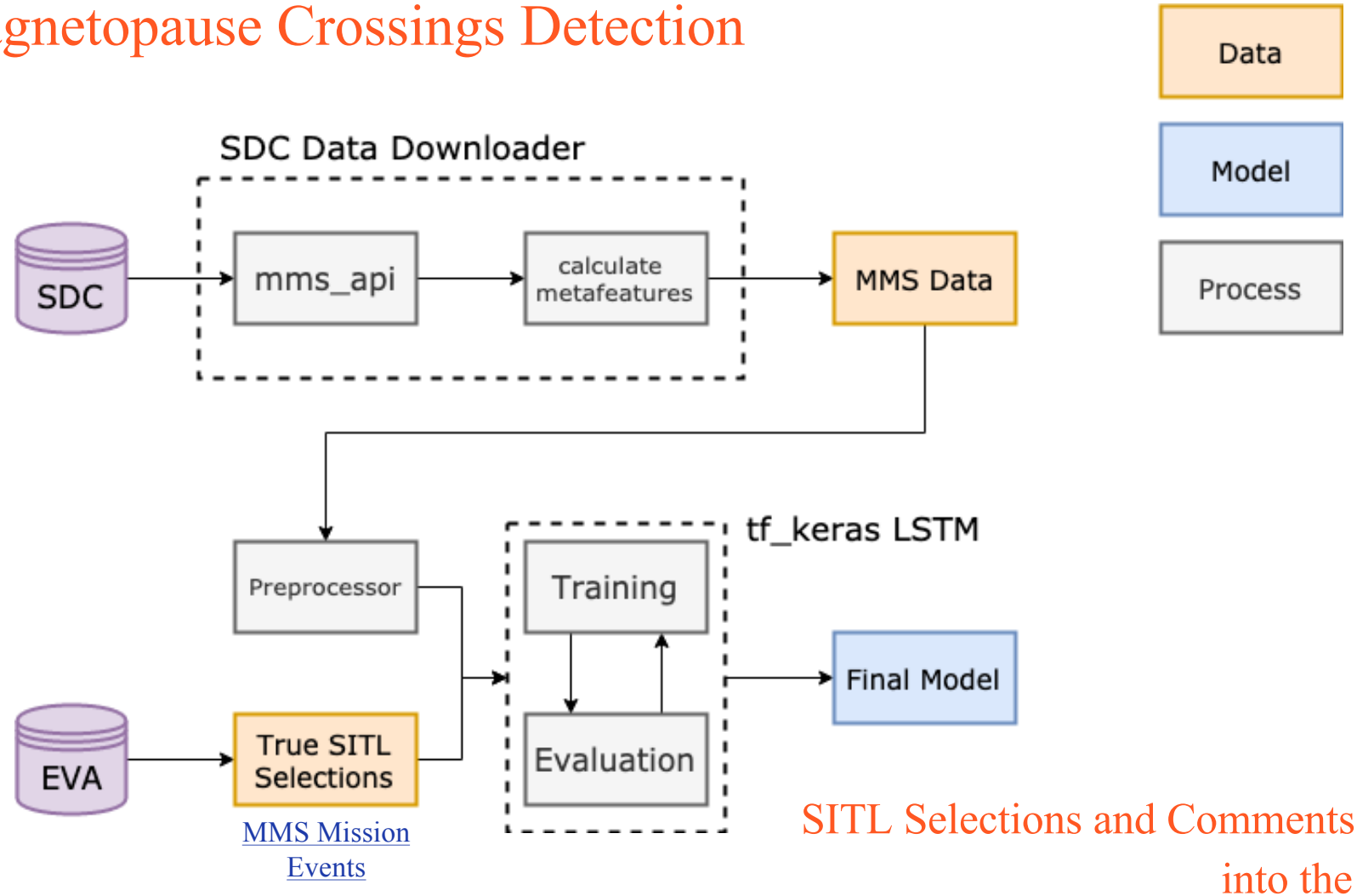
ML and MMS

Recurrent Neural Network with
Long Short-Term Memory (LSTM)



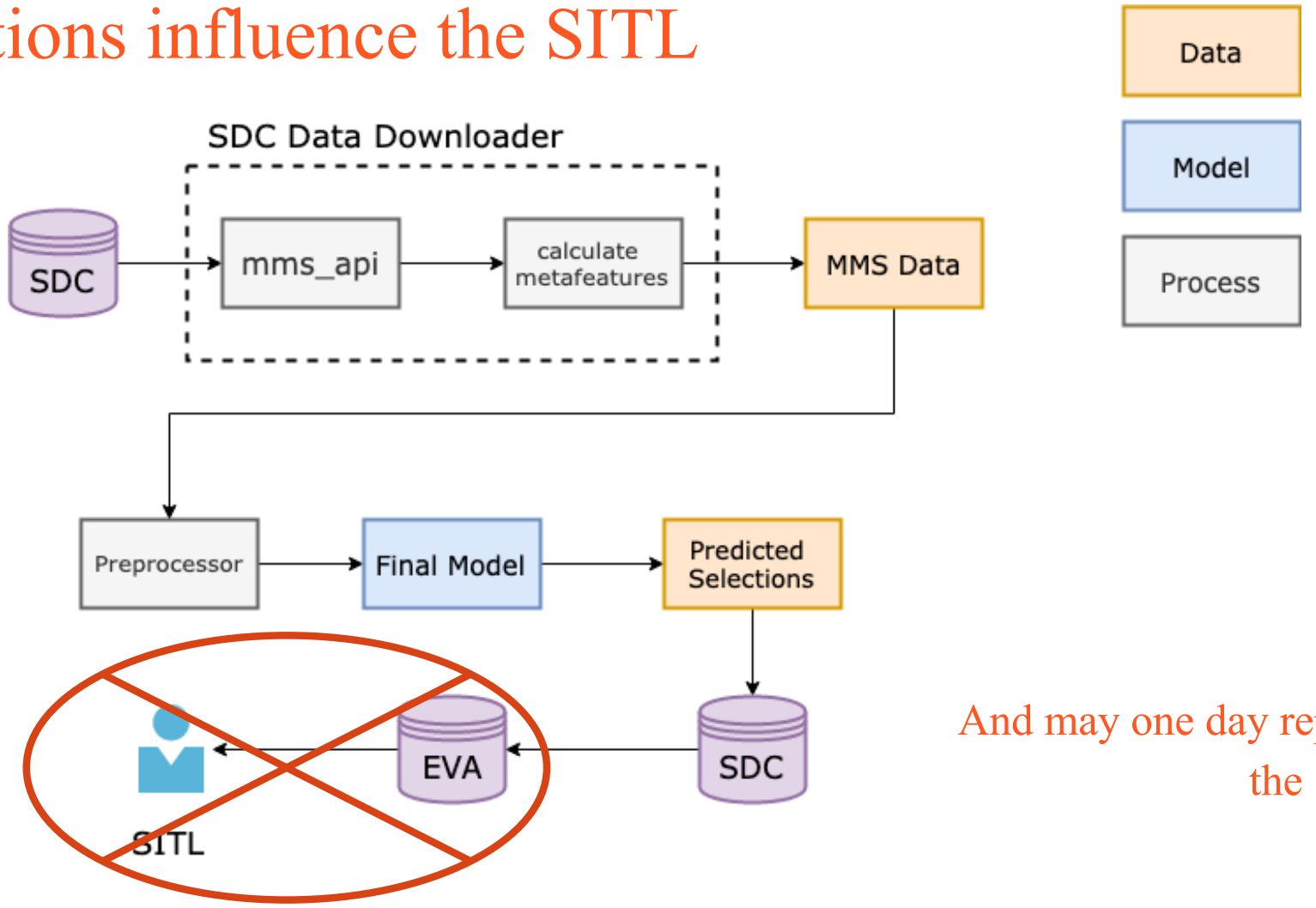
Identify magnetopause crossings where EDRs may occur

Magnetopause Crossings Detection



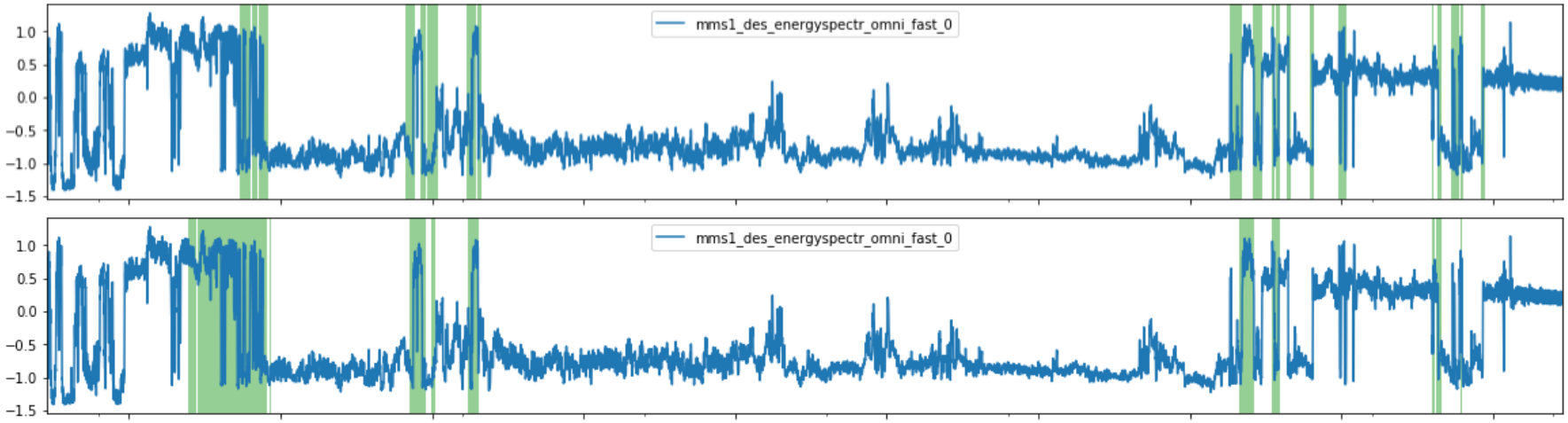
SITL Selections and Comments Factor into the Model

Selections influence the SITL



And may one day replace the SITL

Preliminary model identifies the magnetopause

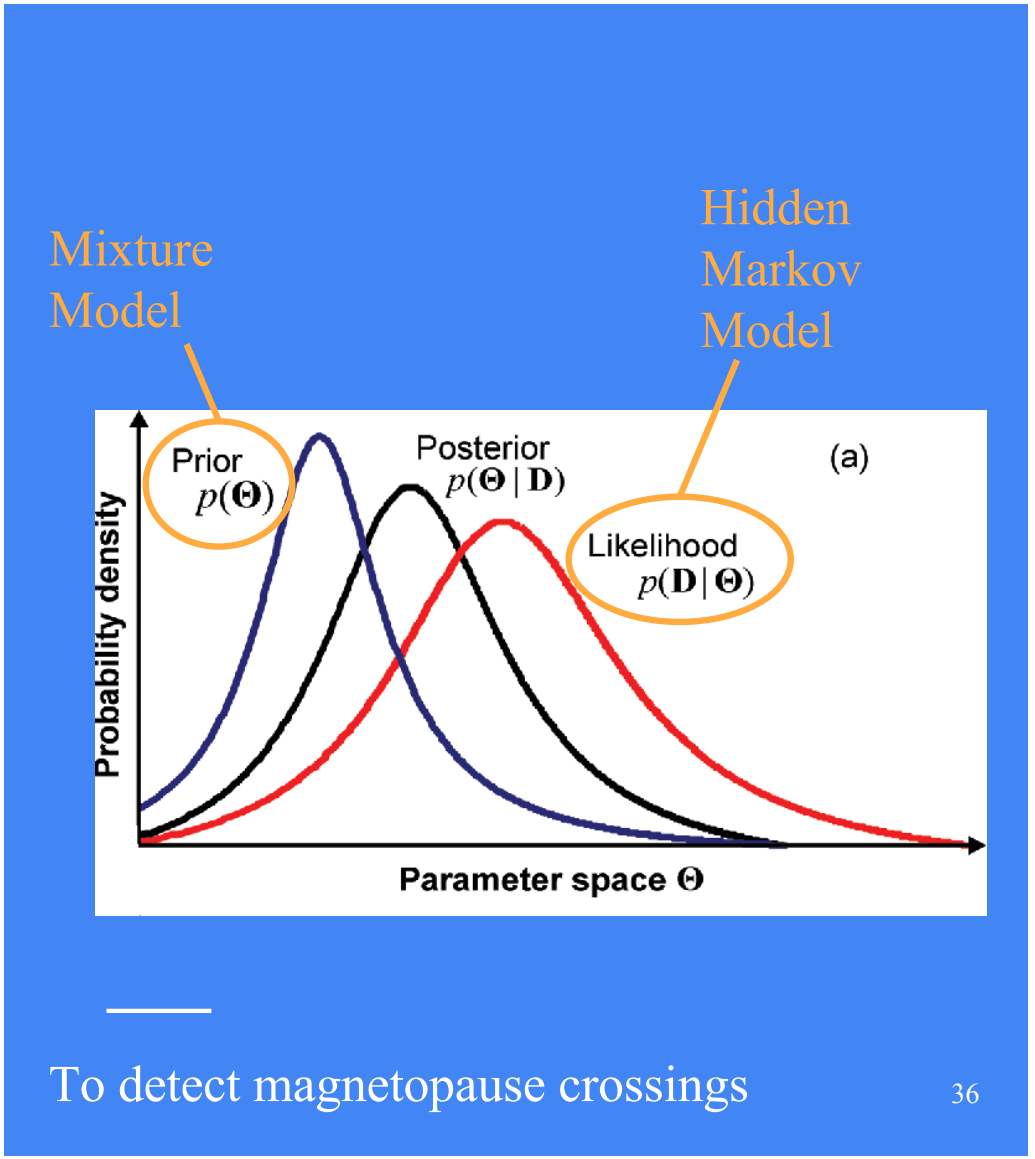


Larger training set should improve results

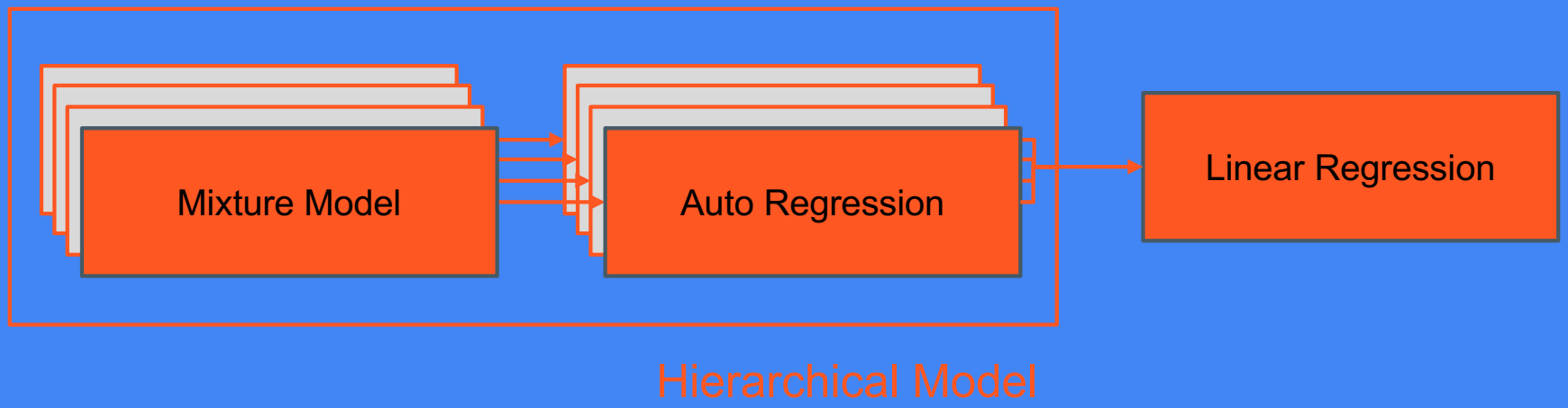
Piatt, S., UNH B.S., 2019

ML and MMS

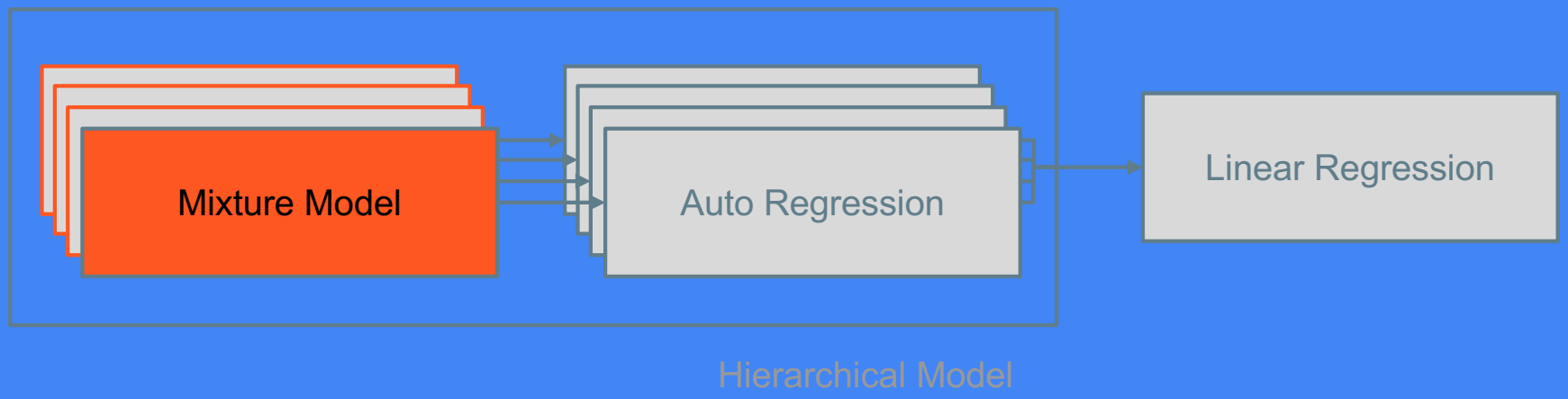
A Hierarchical Bayesian Model



To detect magnetopause crossings

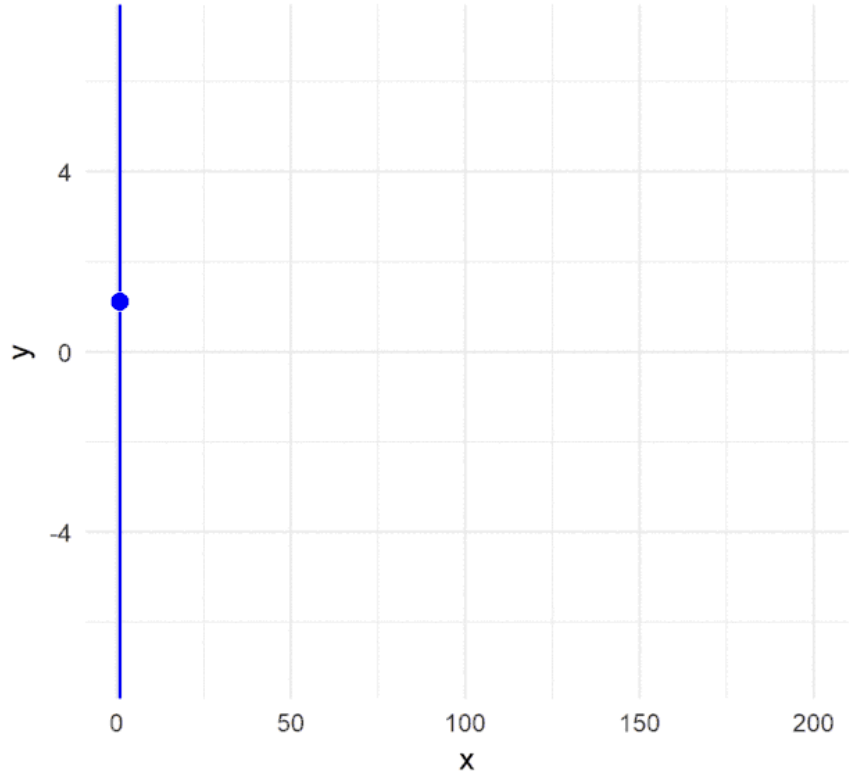


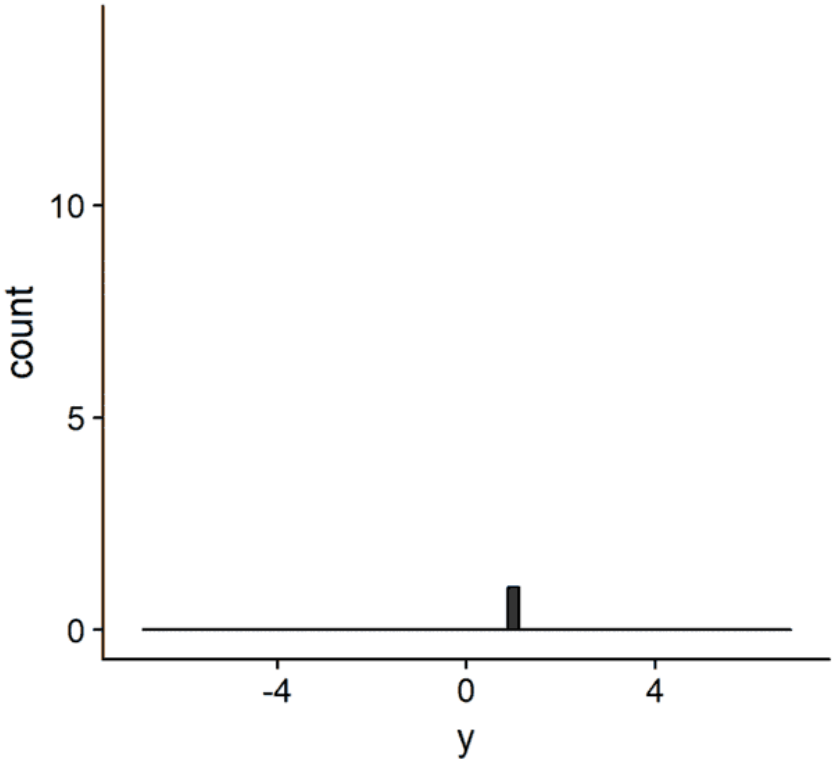
Model and Assumptions



Mixture Model

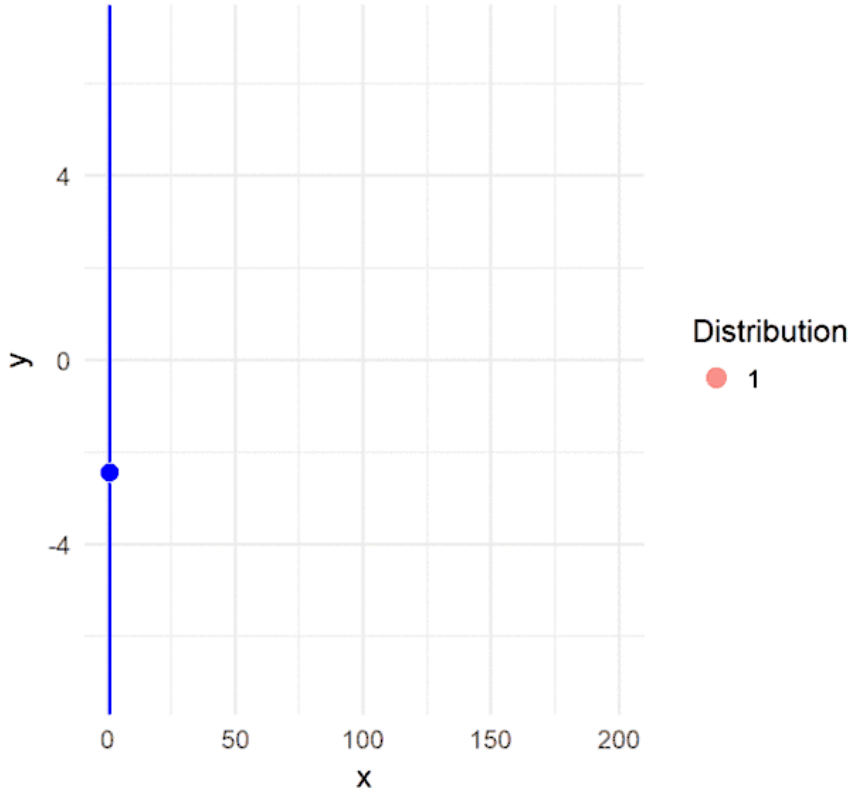
Simple Task to Identify Single Distribution

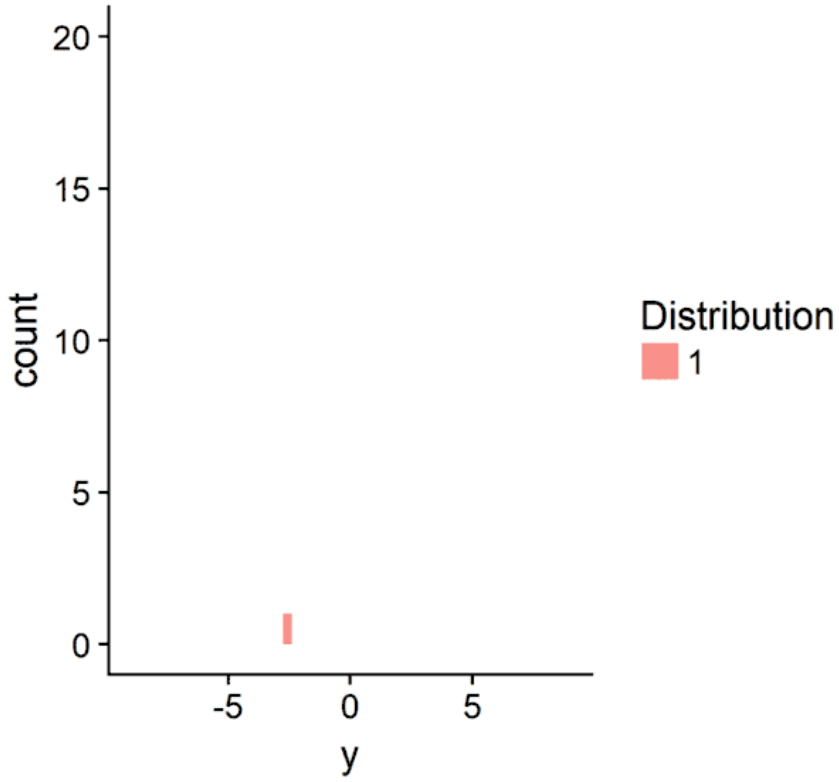




y = normal(**μ**, **σ**) = normal(0, 2)

Difficult to distinguish multi-modal data



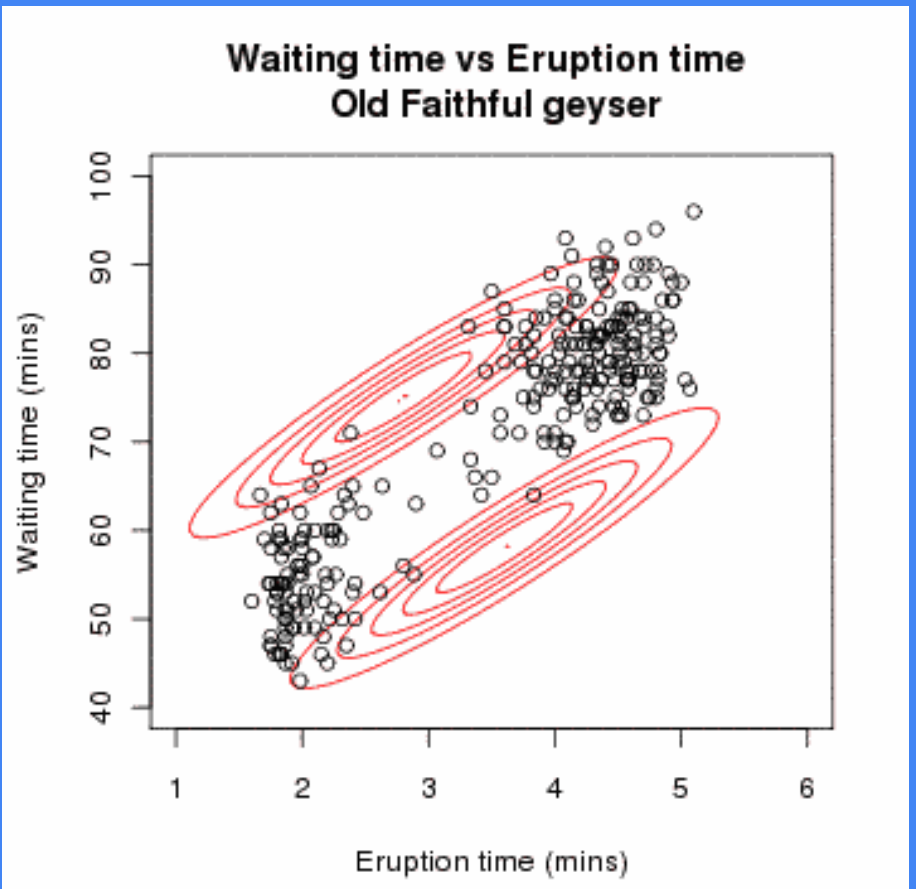


$$y = \begin{cases} 40\% & y_1 = f_N(-2, 1) \\ 60\% & y_2 = f_N(1, 2) \end{cases}$$

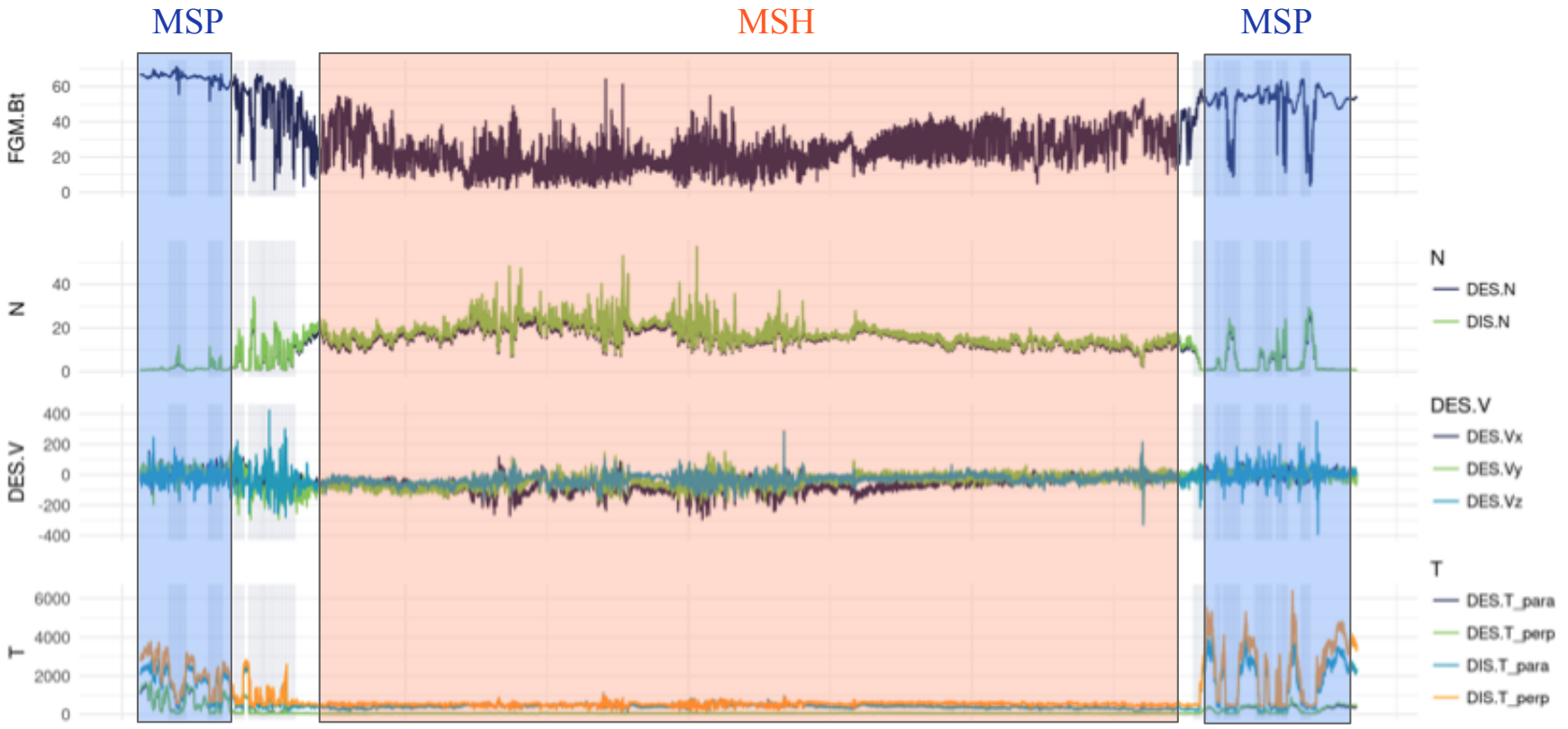
Mixture models fit mixtures of distributions to extract priors

Expectation Maximization

Expect a given data point to be within σ_i of μ_i

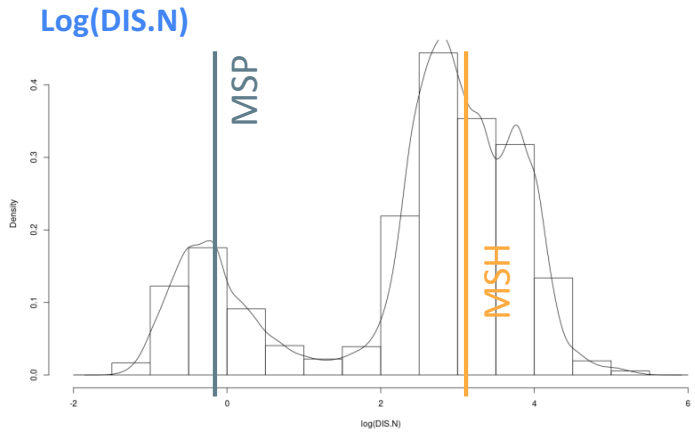


MMS's orbit takes it through different regions of space



Histogram the data to determine its distribution

Our Features are Generally Bimodal

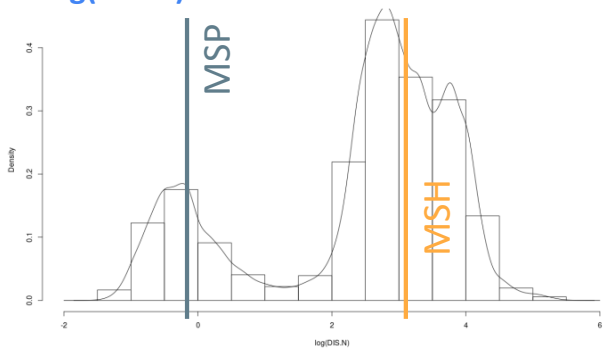


$$y_t \sim \lambda_t \cdot \mathcal{N}(\mu_{MSH}, \sigma^2_{MSH}) + (1 - \lambda_t) \cdot \mathcal{N}(\mu_{MSP}, \sigma^2_{MSP})$$

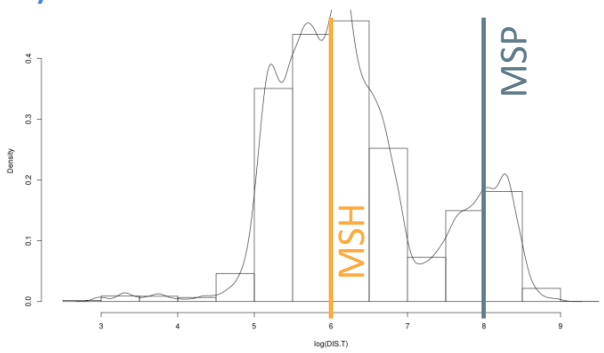
And can be represented as a mixture of two distributions

Our Data's Distribution

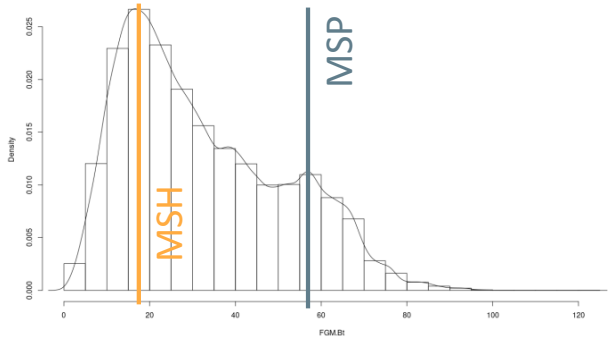
Log(DIS.N)



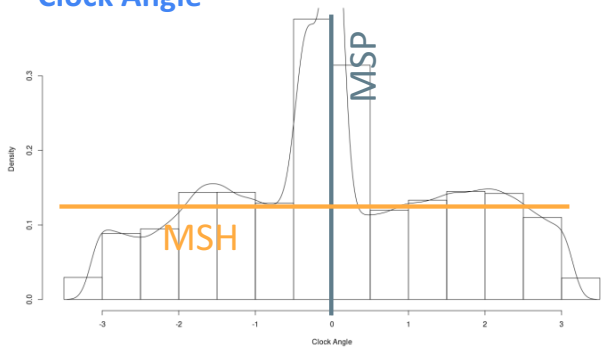
Log(DIS.T)



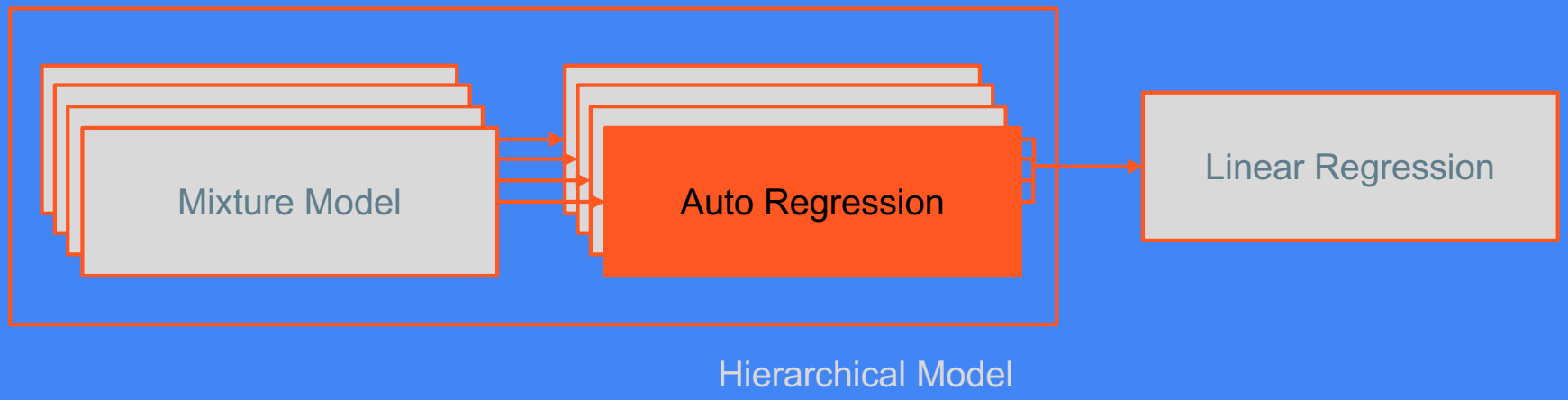
FGM.Bt



Clock Angle



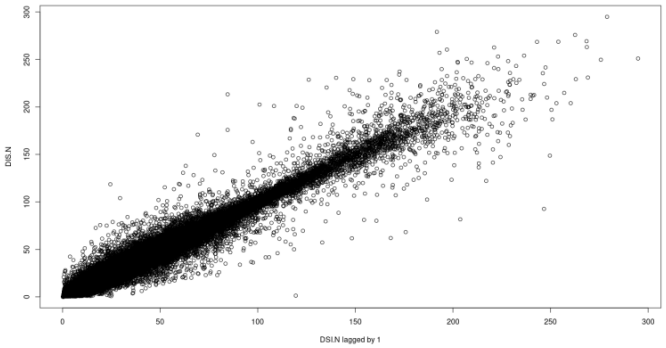
Expectataion Maximization provides our prior distributions



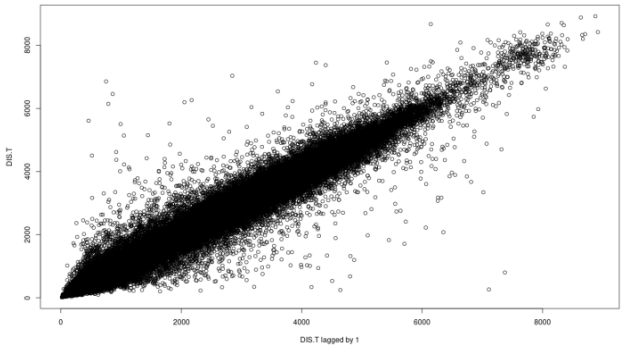
Auto Regression

Independent and Identically Distributed Assumption

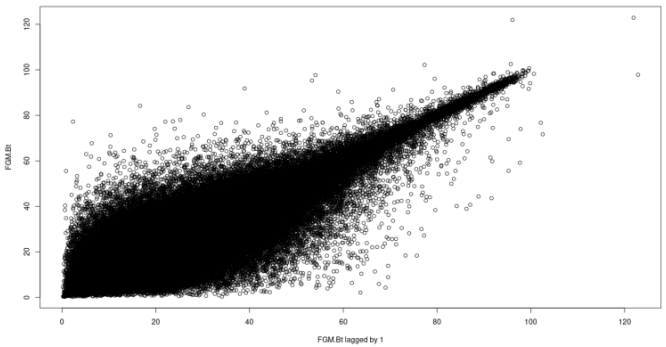
Log(DIS.T)



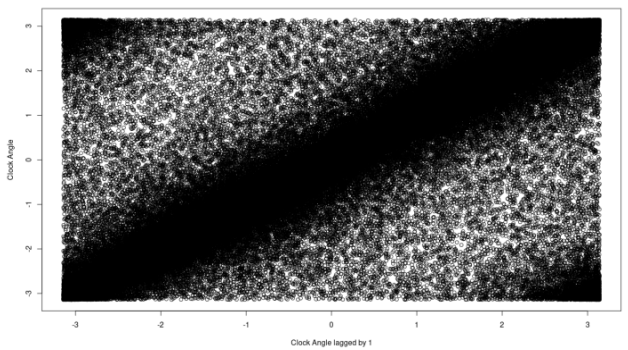
Log(DIS.N)



FGM.Bt



Clock Angle



Pearson's Correlation

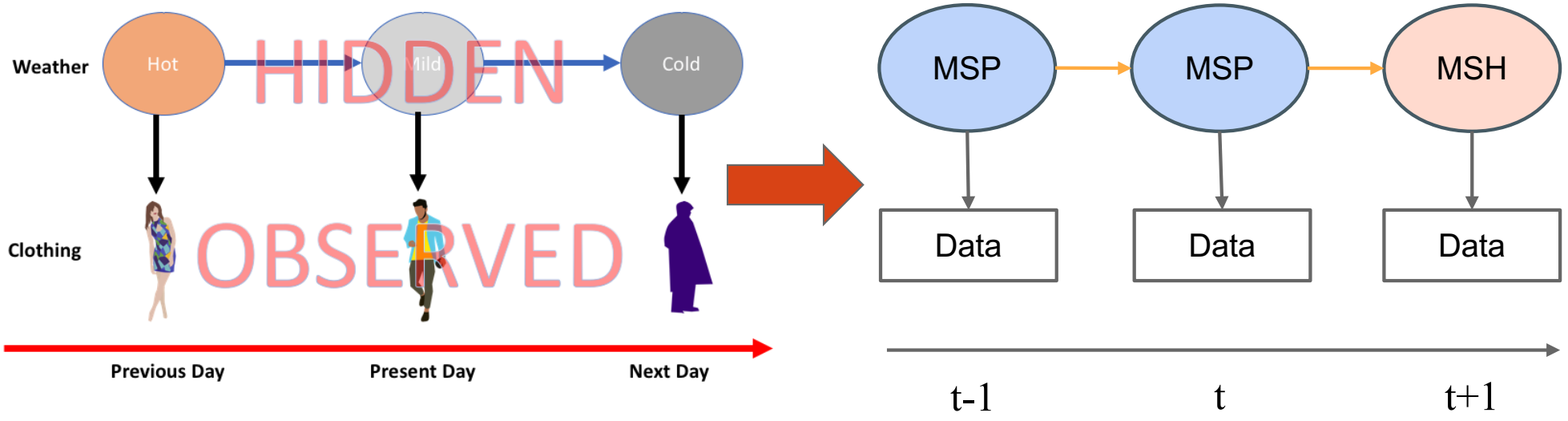
Feature	Lag-1 Correlation
DIS.T	0.9920784
DIS.N	0.9903076
FGM.Bt	0.9457811
Clock Angle	0.7755246

BASED ON 3 MONTHS OF DATA.

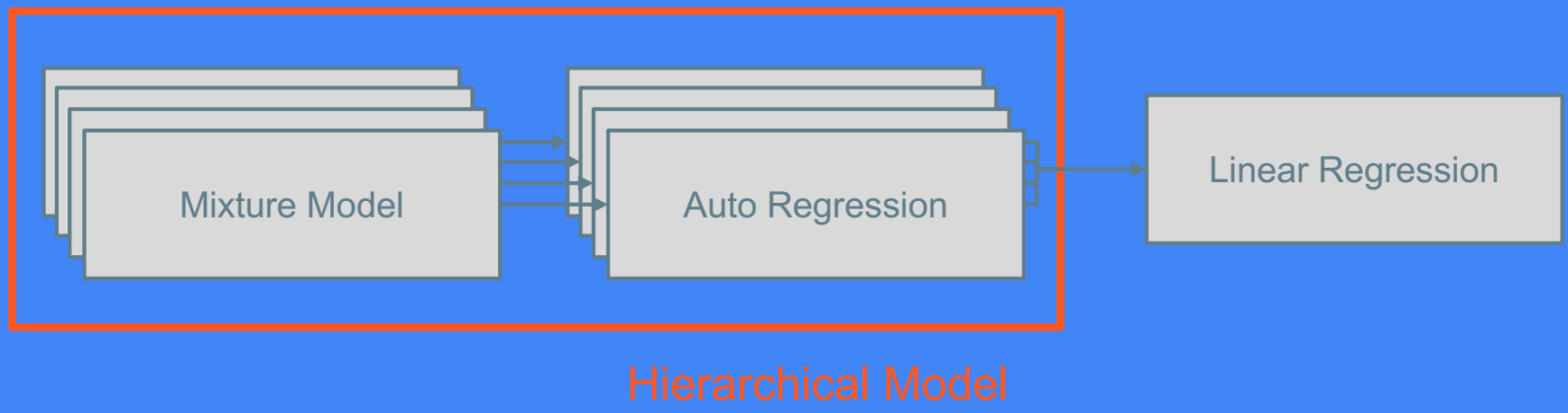
Our time series parameters are not random variables.

Temporal relationship can build in a randomized component

$$\lambda_t \sim \mathcal{N}(\lambda_{t-1}, \sigma_\lambda^2)$$



HMM provides our likelihood function



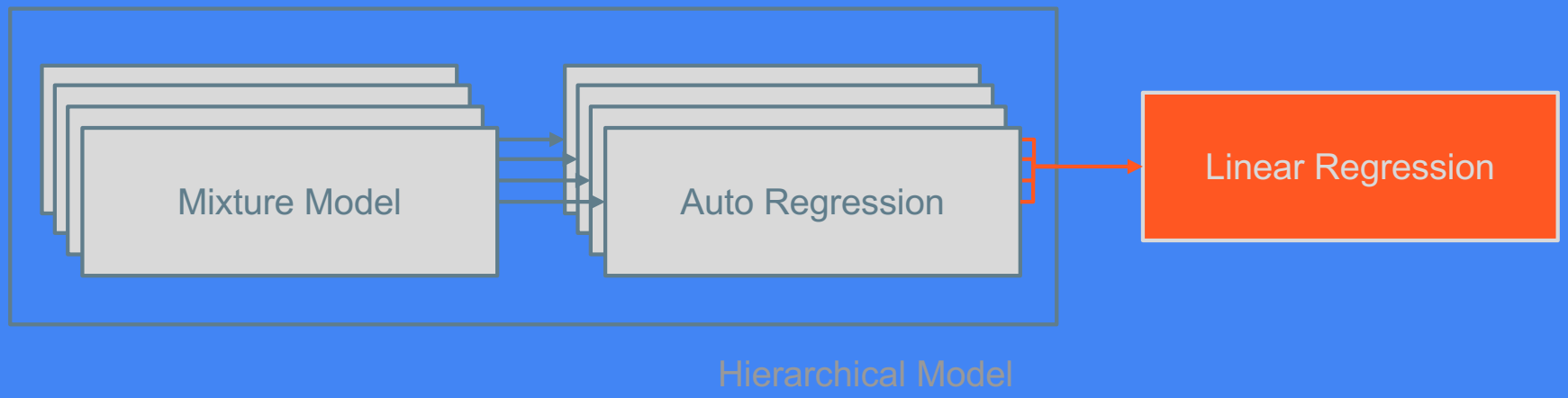
Hierarchical Bayesian Modelling

The Hierarchical Model Ranks New Data by Their Likely Location

$$\lambda_t \sim \mathcal{N}(\lambda_{t-1}, \sigma^2_\lambda)$$

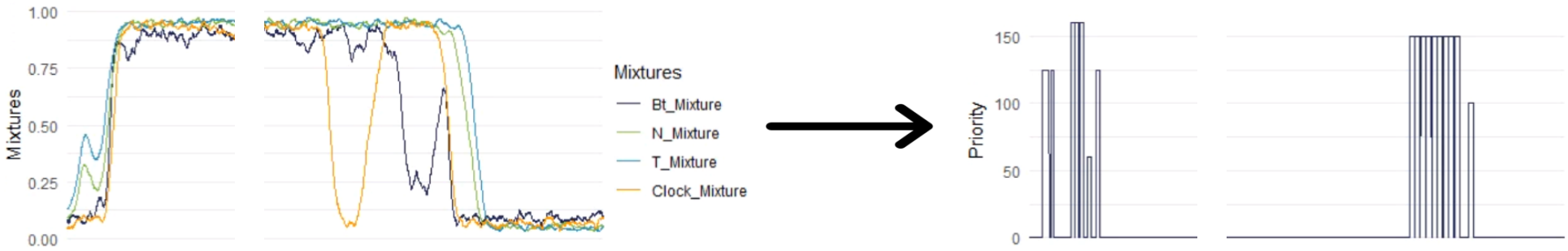
$$y_t \sim \lambda_t \cdot \mathcal{N}(\mu_{MSH}, \sigma^2_{MSH}) + (1 - \lambda_t) \cdot \mathcal{N}(\mu_{MSP}, \sigma^2_{MSP})$$

$\lambda_1 \rightarrow$ MSH; $\lambda_0 \rightarrow$ MSP; $\lambda_{0.5} \rightarrow$ MP



Linear Regression

Individual Parameter Probabilities are Combined into Single Priority Value



$$\text{Priority} \sim \alpha + \beta_{Bt}\lambda_{Bt} + \beta_N\lambda_N + \beta_T\lambda_T + \beta_\theta\lambda_\theta$$

Linear regression tells us which parameter is the best indicator of the MP

Model & Performance

This Model is Programmed in STAN

STAN:

- A probabilistic programming language written in c++.
- Uses a Gradient Descent based Markov Chain Monte Carlo algorithm for Bayesian inference.

Linear Regression Example in STAN:

```
data {
  int N;
  vector[N] x;
  vector[N] y;
}
parameters {
  real alpha;
  real beta;
  real sigma;
}
model {
  y ~ normal(alpha + beta * x, sigma);
}
```

A simple linear regression example is only a few lines long

STAN/Python/R Do All The Work For You

Training Model

```
model{
  // Mixture model
  for (n in 1:numsteps){
    target += log_mix(Bt_Mixture[n],
                      normal_lpdf(Bt[n] | Bt_mix[1], Bt_mix[3]),
                      normal_lpdf(Bt[n] | Bt_mix[2], Bt_mix[4]));

    target += log_mix(N_Mixture[n],
                      normal_lpdf(N_log[n] | N_mix[1], N_mix[3]),
                      normal_lpdf(N_log[n] | N_mix[2], N_mix[4]));

    target += log_mix(T_Mixture[n],
                      normal_lpdf(T_log[n] | T_mix[1], T_mix[3]),
                      normal_lpdf(T_log[n] | T_mix[2], T_mix[4]));

    target += log_mix(Clock_Mixture[n],
                      uniform_lpdf(Clock_Angle[n] | Clock_mix[3], C
                      normal_lpdf(Clock_Angle[n] | Clock_mix[1], Clo
  }

  // Auro-Regression
  for (n in 2:numsteps){
    Bt_Mixture[n] ~ normal(Bt_Mixture[n-1], Bt_mix_sigma);
    N_Mixture[n] ~ normal(N_Mixture[n-1], N_mix_sigma);
    T_Mixture[n] ~ normal(T_Mixture[n-1], T_mix_sigma);
    Clock_Mixture[n] ~ normal(Clock_Mixture[n-1], Clock_mix_sigma);
  }

  // Linear Regression
  Priority ~ normal(mixture_alpha + Bt_beta * Bt_Mixture +
                    N_beta * N_Mixture + T_beta * T_Mixture +
                    Clock_beta * Clock_Mixture, mixture_sigma);
}
```

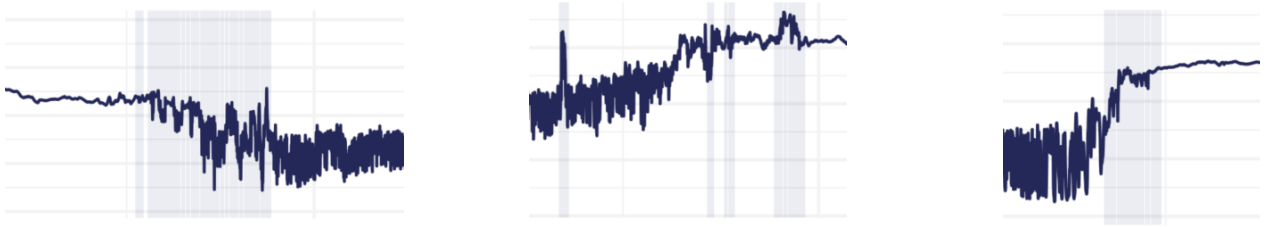
Test Model

```
model{
  // Mixture model
  for (n in 1:numsteps){
    target += log_mix(Bt_Mixture[n],
                      normal_lpdf(Bt[n] | Bt_mix[1], Bt_mix[3]),
                      normal_lpdf(Bt[n] | Bt_mix[2], Bt_mix[4]));

    target += log_mix(N_Mixture[n],
                      normal_lpdf(N_log[n] | N_mix[1], N_mix[3]),
                      normal_lpdf(N_log[n] | N_mix[2], N_mix[4]));

    target += log_mix(T_Mixture[n],
                      normal_lpdf(T_log[n] | T_mix[1], T_mix[3]),
                      normal_lpdf(T_log[n] | T_mix[2], T_mix[4]));

    target += log_mix(Clock_Mixture[n],
                      uniform_lpdf(Clock_Angle[n] | Clock_mix[3], Clock_mix[4]),
                      normal_lpdf(Clock_Angle[n] | Clock_mix[1], Clock_sigma));
  }

  // Auro-Regression
  for (n in 2:numsteps){
    Bt_Mixture[n] ~ normal(Bt_Mixture[n-1], Bt_mix_sigma);
    N_Mixture[n] ~ normal(N_Mixture[n-1], N_mix_sigma);
    T_Mixture[n] ~ normal(T_Mixture[n-1], T_mix_sigma);
    Clock_Mixture[n] ~ normal(Clock_Mixture[n-1], Clock_mix_sigma);
  }
}
generated quantities {
  vector[numsteps] Priority;

  // Linear Regression
  for (i in 1:numsteps){
    Priority[i] = normal_rng(mixture_alpha + Bt_beta * Bt_Mixture[i] +
                             N_beta * N_Mixture[i] + T_beta * T_Mixture[i] +
                             Clock_beta * Clock_Mixture[i], mixture_sigma);
  }
}
```

The entire model is only 21 lines of code

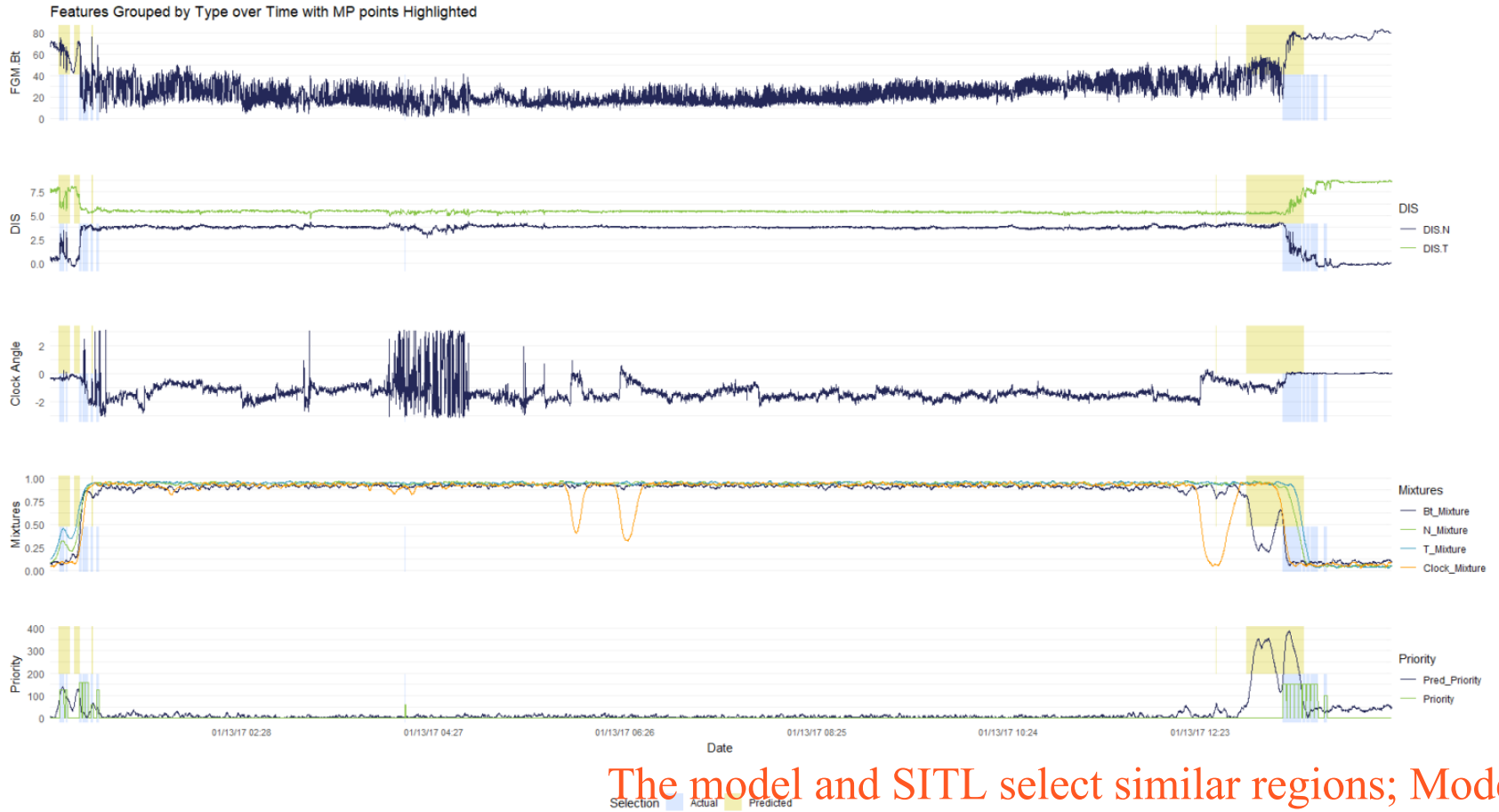
The SITL and Model are Sightly Different

- Selections tend to vary between scientists.
- Limited transmission bandwidth necessitates a priority system.
- Selections can sometimes miss parts of the magnetopause.

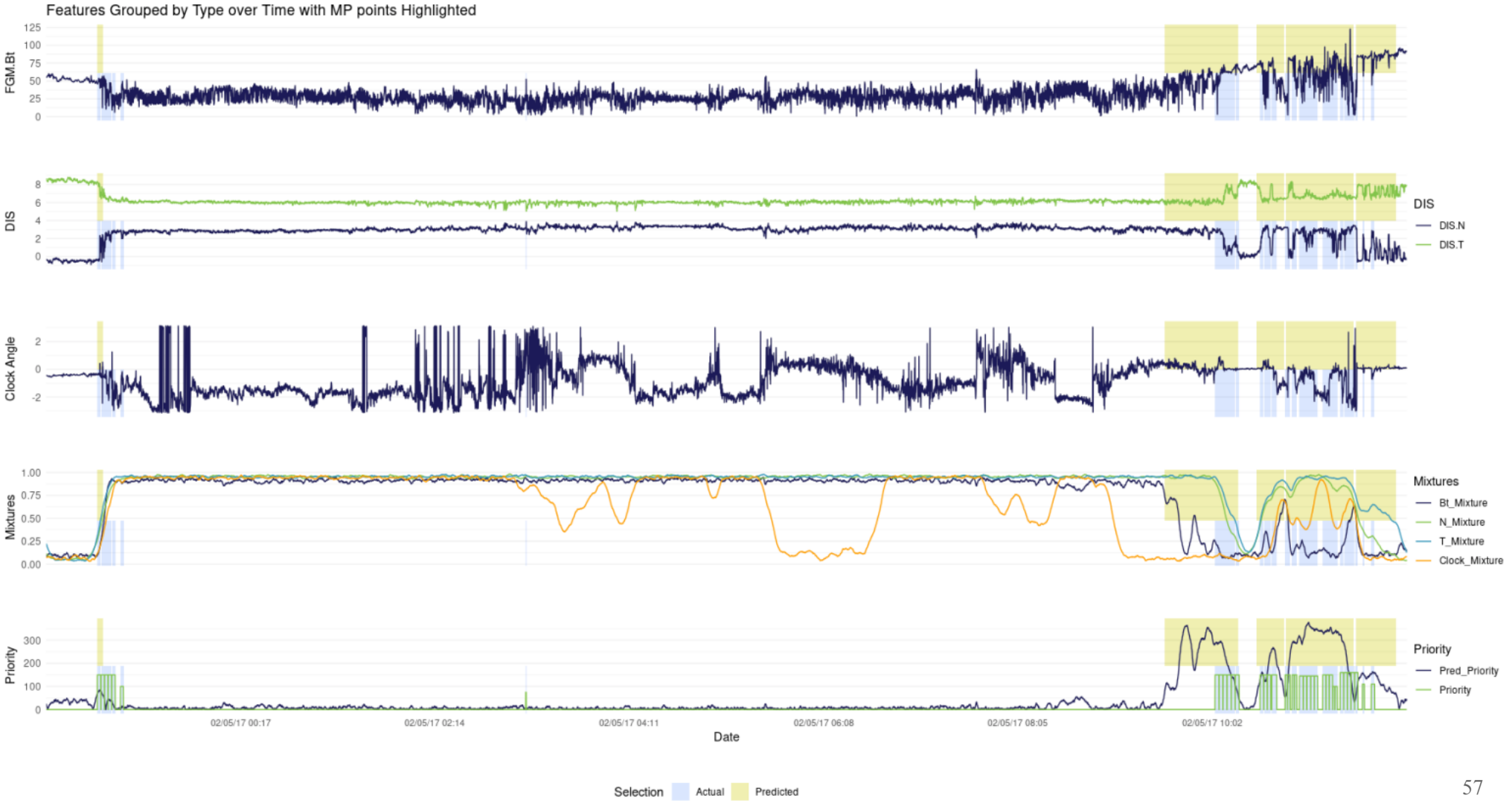


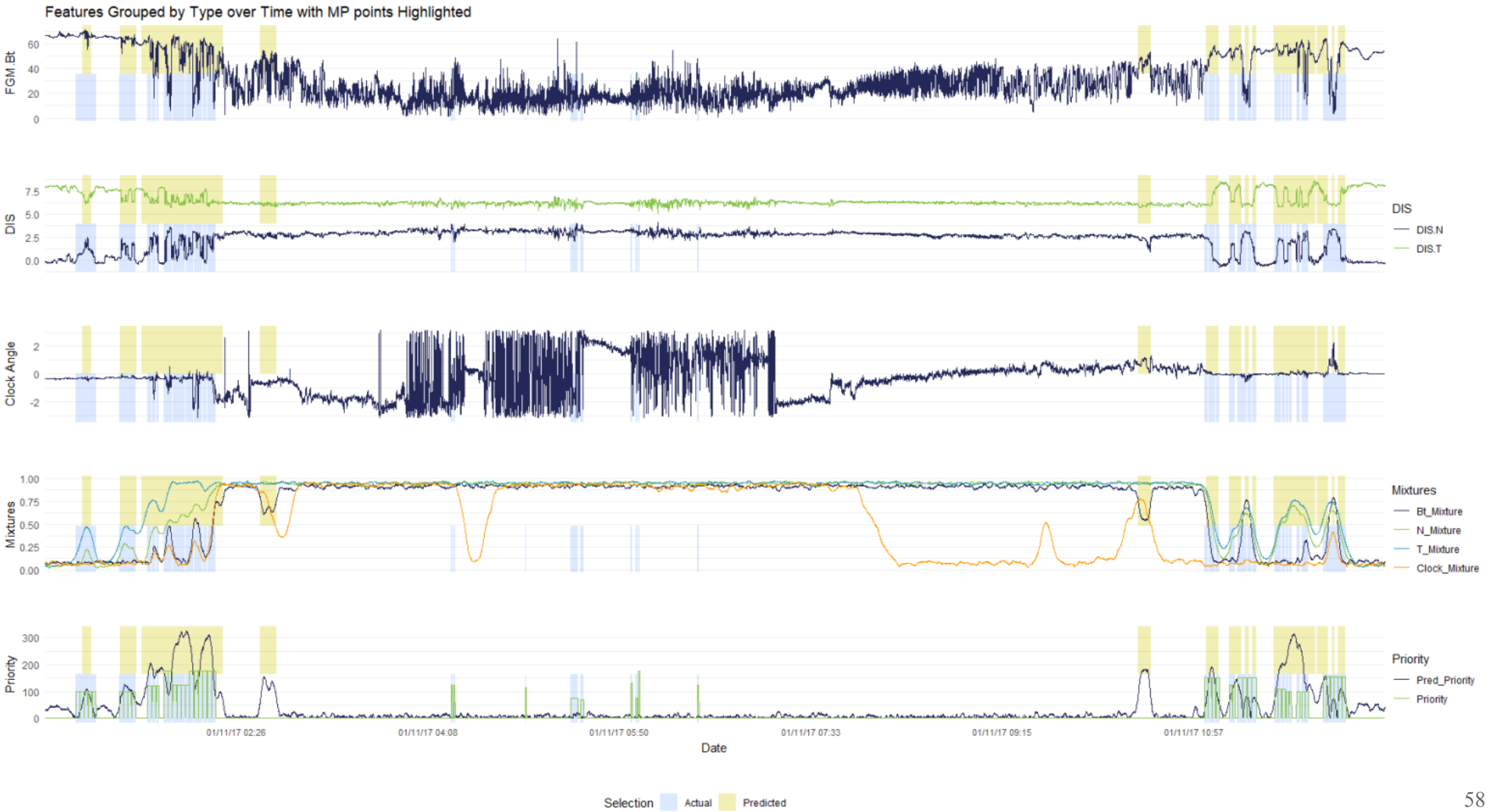
The model selects the entire magnetopause for all crossings

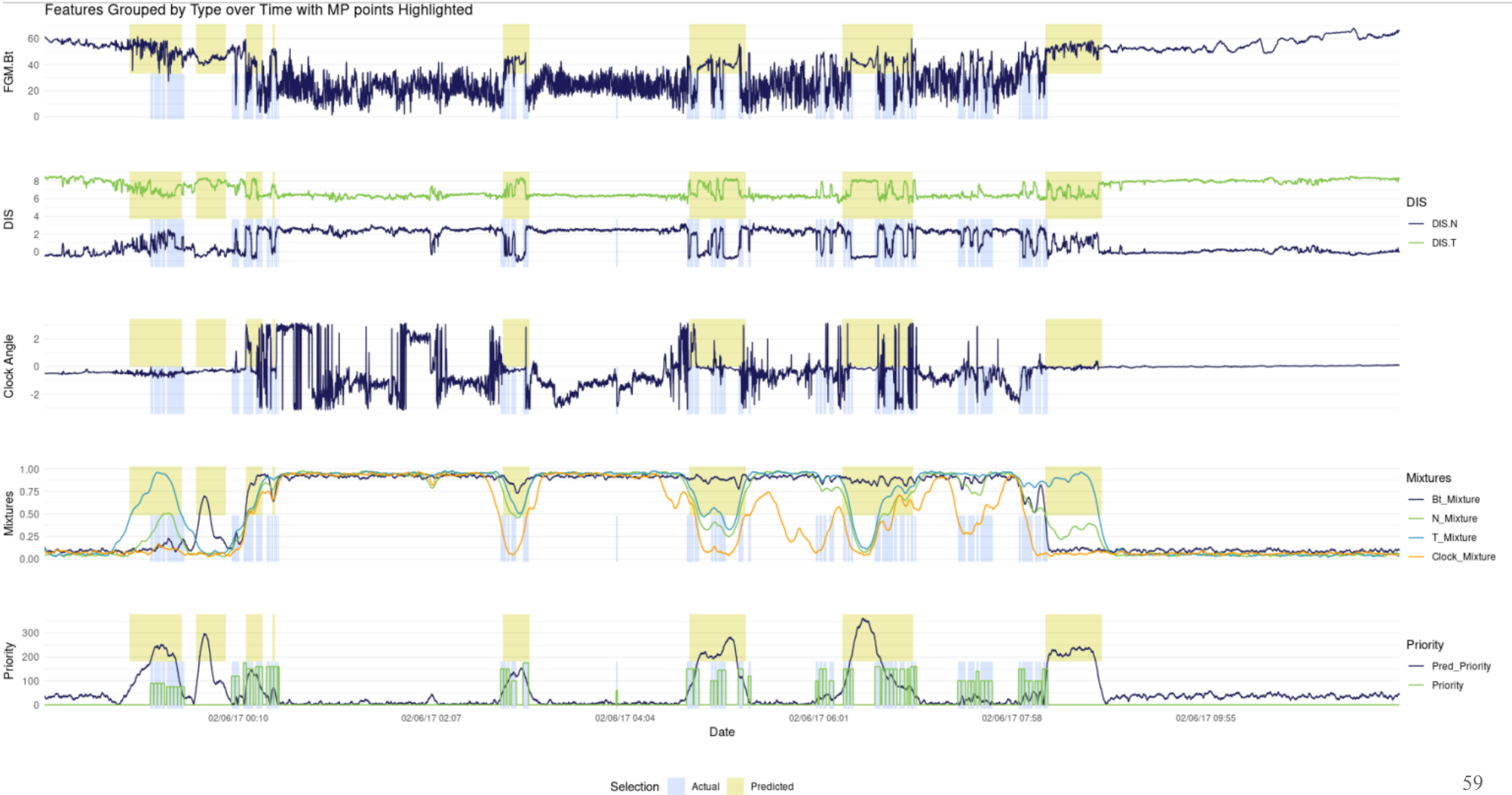
Visual Evaluation of One ROI



The model and SITL select similar regions; Model gives high weights to |B|


Features Grouped by Type over Time with MP points Highlighted
FGM.Bt
125
100
75
50
25
0
DIS
8
6
4
2
0
DIS
DIS.N
DIS.T
Clock Angle
2
0
-2
Mixtures
1.00
0.75
0.50
0.25
0.00
Mixtures
Bt_Mixture
N_Mixture
T_Mixture
Clock_Mixture
Priority
300
200
100
0
Priority
Pred_Priority
Priority
02/05/17 00:17
02/05/17 02:14
02/05/17 04:11
02/05/17 06:08
02/05/17 08:05
02/05/17 10:02
Date
Selection
Actual
Predicted


Features Grouped by Type over Time with MP points Highlighted
FGM.Bt
60
40
20
0
DIS
7.5
5.0
2.5
0.0
DIS
DIS.N
DIS.T
Clock Angle
2
0
-2
Mixtures
1.00
0.75
0.50
0.25
0.00
Mixtures
Bt_Mixture
N_Mixture
T_Mixture
Clock_Mixture
Priority
300
200
100
0
Priority
Pred_Priority
Priority
01/11/17 02:26
01/11/17 04:08
01/11/17 05:50
01/11/17 07:33
01/11/17 09:15
01/11/17 10:57
Date
Selection
Actual
Predicted


Features Grouped by Type over Time with MP points Highlighted
FGM.Bt
60
40
20
0
DIS
8
6
4
2
0
DIS
DIS.N
DIS.T
Clock Angle
2
0
-2
Mixtures
1.00
0.75
0.50
0.25
0.00
Mixtures
Bt_Mixture
N_Mixture
T_Mixture
Clock_Mixture
Priority
300
200
100
0
Priority
Pred_Priority
Priority
02/06/17 00:10
02/06/17 02:07
02/06/17 04:04
02/06/17 06:01
02/06/17 07:58
02/06/17 09:55
Date
Selection
Actual
Predicted

Four More Orbits

Point-by-Point Comparison Shows Agreement Between Model & SITL

- Average of 10,829 data points per day, 86,639 in total.
- 50% test, 50% train split.

Misclassification	13%
True Positive Rate	66%
True Negative Rate	89%
P-Value	46%

Evaluation would be more fair and favorable on a selection-by-selection basis

Mixture Model → $y_t \sim \lambda_t \cdot \mathcal{N}(\mu_{MSH}, \sigma^2_{MSH}) + (1 - \lambda_t) \cdot \mathcal{N}(\mu_{MSP}, \sigma^2_{MSP})$

Auto Regression → $\lambda_t \sim \mathcal{N}(\lambda_{t-1}, \sigma^2_\lambda)$

Linear Regression → $\text{Priority} \sim \alpha + \beta_{Bt}\lambda_{Bt} + \beta_N\lambda_N + \beta_T\lambda_T + \beta_\theta\lambda_\theta$

Model Summary

Summary

- Machine Learning is being used in more and more fields
  - Outperform humans at evermore complex tasks
- Impact of ML in Heliophysics is growing
  - Has helped analyze data in all regions of the heliosphere
- ML enables more complex mission design
  - Makes large volumes of data tractable
  - Simplifies complex mission operations schemes
- ML is helping MMS select its mission critical data
  - Ground loops can be tailored to specific campaigns or extended mission objectives
  - Neural networks and Bayesian models automate the SITL magnetopause selection process
- GEM Focus Group on Machine Learning in Space Physics
  - Email me at matthew.argall@unh.edu
  - Share your level of interest, your work, interesting papers