A Bayesian Approach to Space Weather Prediction

Enrico Camporeale
Center for Mathematics and Computer Science (CWI)
Amsterdam, The Netherlands

Collaborators:
M. Chandorkar (CWI), A. Care' (U. Brescia), J. Borovsky (SSI), TAO team (INRIA), Y. Shprits (GFZ)

Bayes' formula

- Bayes' formula follows from the two fundamental rules of probability
- Product rule: $$p(X,Y) = p(Y|X)p(X)$$
- Sum rule: $$p(X) = \sum_Y p(X,Y)$$

Bayes' formula

- Bayes' formula follows from the two fundamental rules of probability
- Product rule:

$$p(X,Y) = p(Y|X)p(X)$$

- Sum rule:

$$p(X) = \sum_Y p(X,Y)$$

$$p(Y|X)p(X) = p(X|Y)p(Y)$$

Bayes' formula

- Bayes' formula follows from the two fundamental rules of probability
- Product rule: $$p(X,Y) = p(Y|X)p(X)$$
- Sum rule: $$p(X) = \sum_Y p(X,Y)$$

$$p(Y|X)p(X) = p(X|Y)p(Y) \rightarrow p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

Bayes' formula

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)} = \frac{p(X|Y)p(Y)}{\sum_Y p(X|Y)p(Y)}$$

- Why is so important?

Bayes' formula

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)} = \frac{p(X|Y)p(Y)}{\sum_Y p(X|Y)p(Y)}$$

- Why is so important?
  - Because it allows to write p(Y|X) in terms of p(X|Y), i.e. it “flips” the conditional probability

Bayes' formula

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)} = \frac{p(X|Y)p(Y)}{\sum_Y p(X|Y)p(Y)}$$

- Why is so important?
  - Because it allows to write p(Y|X) in terms of p(X|Y), i.e. it "flips" the conditional probability
  - Bayes formula is hardwired in human decision making*:

    It is our natural way of updating the probability we assign to an event, at the light of new information

    * except most of the time we get the math wrong!

Bayes formula at work
(example from Nate Silver's *The Signal and the Noise*)

- The underwear problem:

Bayes formula at work

(example from Nate Silver's *The Signal and the Noise*)

- The underwear problem:
  - What is the probability that your partner is cheating on you, conditioned on you finding an unknown pair of underwear in your drawer?

Bayes formula at work

(example from Nate Silver's *The Signal and the Noise*)

- The underwear problem:

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)} = \frac{p(X|Y)p(Y)}{\sum_Y p(X|Y)p(Y)}$$

P(cheating | underwear appearing) =

P(underwear | cheating) * P(cheating) /

[P(underwear | cheating) * P(cheating) +

P(underwear | no cheating) * P(no cheating)]

Bayes formula at work
(example from Nate Silver's *The Signal and the Noise*)

- The underwear problem:

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)} = \frac{p(X|Y)p(Y)}{\sum_Y p(X|Y)p(Y)}$$

P(partner is cheating) = 4% PRIOR

Bayes formula at work

(example from Nate Silver's *The Signal and the Noise*)

- The underwear problem:

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)} = \frac{p(X|Y)p(Y)}{\sum_Y p(X|Y)p(Y)}$$

P(partner is cheating) = 4% PRIOR

- P(underwear appearing | cheating) = 50% LIKELIHOOD
- P(underwear appearing | not cheating) = 5%

Bayes formula at work

(example from Nate Silver's *The Signal and the Noise*)

- <u>The underwear problem:</u>

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)} = \frac{p(X|Y)p(Y)}{\sum_Y p(X|Y)p(Y)}$$

P(partner is cheating) = 4% PRIOR

- P(underwear appearing | cheating) = 50% LIKELIHOOD
- P(underwear appearing | not cheating) = 5%

- P(cheating | underwear appearing) = 29 % POSTERIOR

Bayes formula at work

(example from Nate Silver's *The Signal and the Noise*)

- The 9/11 problem:

  P(terrorists would crash planes into twin towers) = 0.005%
  PRIOR

- P(plane hitting | terrorists are attacking twin towers) = 100%
  LIKELIHOOD

- P(plane hitting | terrorists are NOT attacking twin towers) = 0.008%

- P(terrorist attack | first plane hits) = 38 % POSTERIOR

Bayes formula at work

(example from Nate Silver's *The Signal and the Noise*)

- <u>The 9/11 problem:</u>

  P(terrorists would crash planes into twin towers) = 0.005%
  PRIOR

- P(plane hitting | terrorists are attacking twin towers) = 100%
  LIKELIHOOD

- P(plane hitting | terrorists are NOT attacking twin towers) = 0.008%

- P(terrorist attack | first plane hits) = 38 % ~~POSTERIOR~~ PRIOR

- P(terrorist attack | second plane hits) = 99.99 %

Bayesian inference

Let's assume to have a <u>model</u> that depends on unknown parameters **M** that we want to use to describe <u>observations</u> (data) **d**

POSTERIOR ~ LIKELIHOOD * PRIOR

$$P(M|d) \sim P(d|M) * P(M)$$

Bayesian inference

Let's assume to have a <u>model</u> that depends on unknown parameters **M** that we want to use to describe <u>observations</u> (data) **d**

POSTERIOR ~ LIKELIHOOD * PRIOR

$P(M|d) \sim P(d|M) * P(M)$

Probability of parameters BEFORE seeing the data

Bayesian inference

Let's assume to have a model that depends on unknown parameters **M** that we want to use to describe observations (data) **d**

POSTERIOR ~ LIKELIHOOD * PRIOR

$$P(M|d) \sim P(d|M) * P(M)$$

Probability that the data is generated by a given set of parameters

Probability of parameters BEFORE seeing the data

Bayesian inference

Let's assume to have a <u>model</u> that depends on unknown parameters **M** that we want to use to describe <u>observations</u> (data) **d**

POSTERIOR ~ LIKELIHOOD * PRIOR

$$P(M|d) \sim P(d|M) * P(M)$$

- $P(M|d)$: Probability of parameters AFTER seeing the data (conditioned on data)
- $P(d|M)$: Probability that the data is generated by a given set of parameters
- $P(M)$: Probability of parameters BEFORE seeing the data

Machine Learning for Space Weather

Home Team Projects Publications Resources Jobs Contact Us

Machine Learning for Space Weather

Coupling physics-based simulations with Artificial Intelligence

Goal

In this project we aim at enhancing the current state-of-the-art simulations for Space Weather, by using prior knowledge gathered from historical satellite data. Several Machine Learning techniques will be used for data-mining, classification, and regression. The long-term objective of the project is the creation of a portfolio of data-enhanced reduced models, along with automated rules for model selection. Depending on the real-time conditions observed by satellites, the resulting 'grey-box' model should choose the relative importance between physical and empirical estimations.

CWI/INRIA consortium

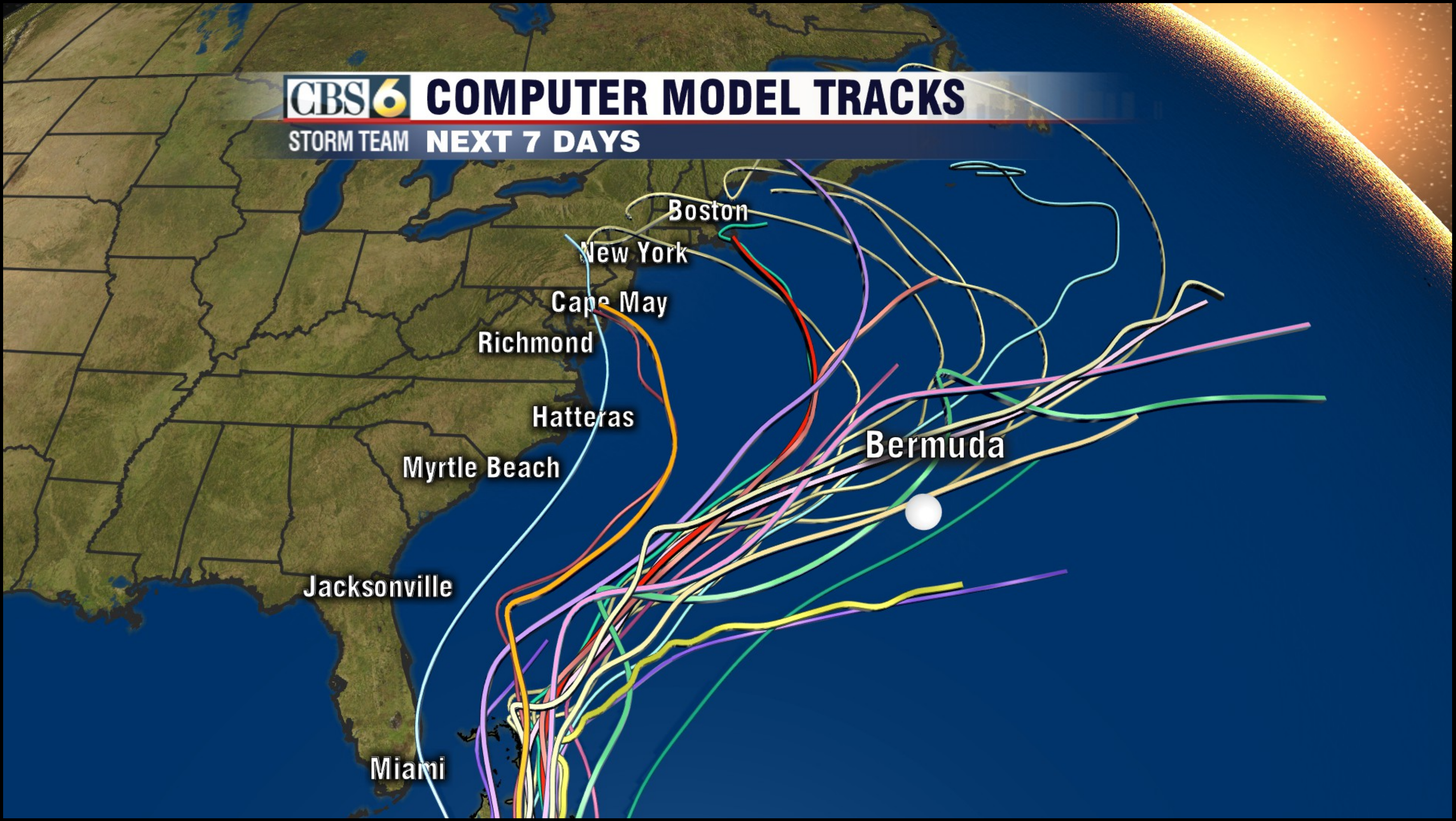
This project has started as funded by a CWI/INRIA collaboration. However, several other parties have joined in this activity, either as external collaborators, or more actively involved. Visit the team page for a list of all the people involved.

CWI Centrum Wiskunde & Informatica

Inria INVENTEURS DU MONDE NUMÉRIQUE

CWI is the Dutch National Center for Mathematics and Computer Science. INRIA is the French Institute for Research in Computer Science and Automation.

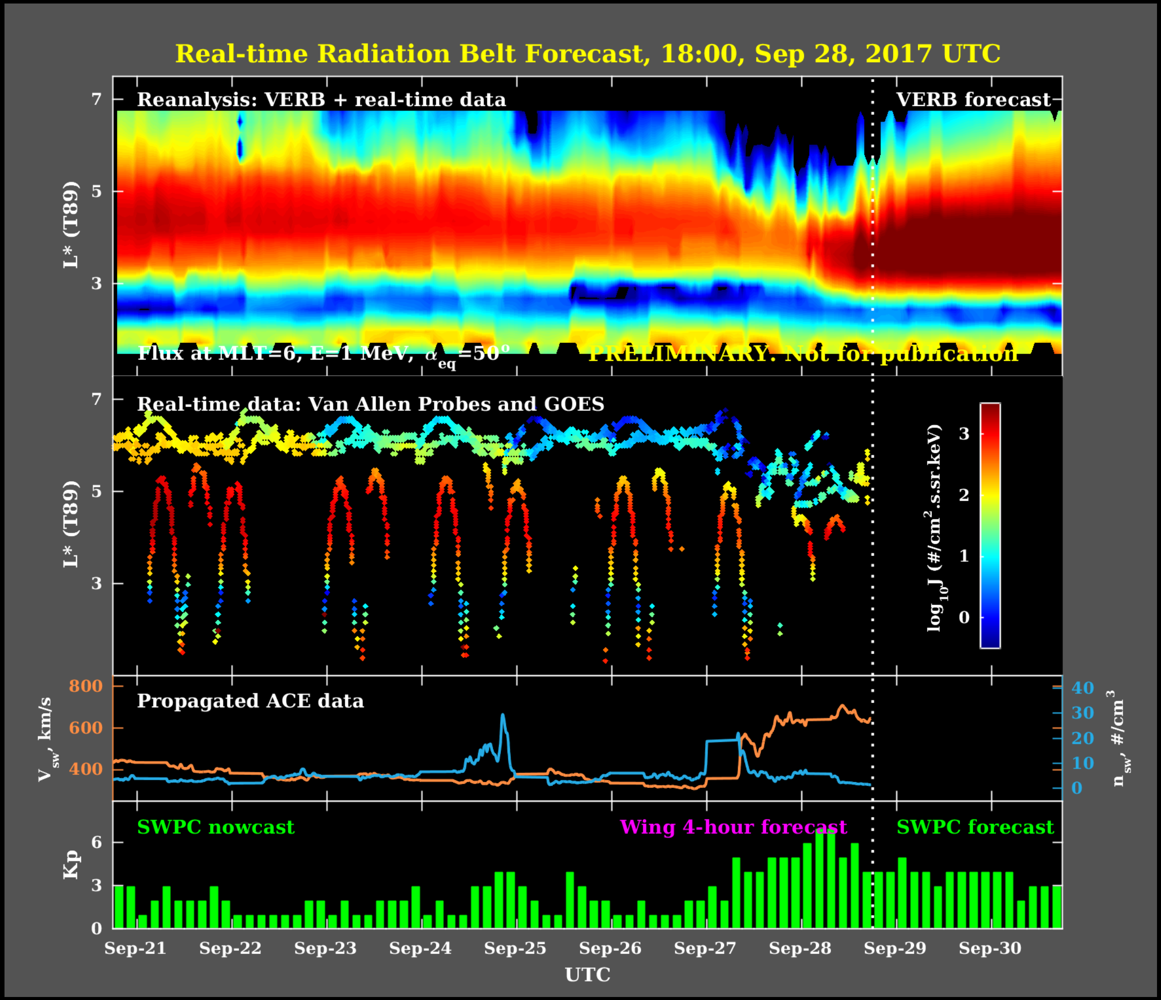
Ensemble simulations in hazard prediction



… this is not yet a routine in space weather forecasting

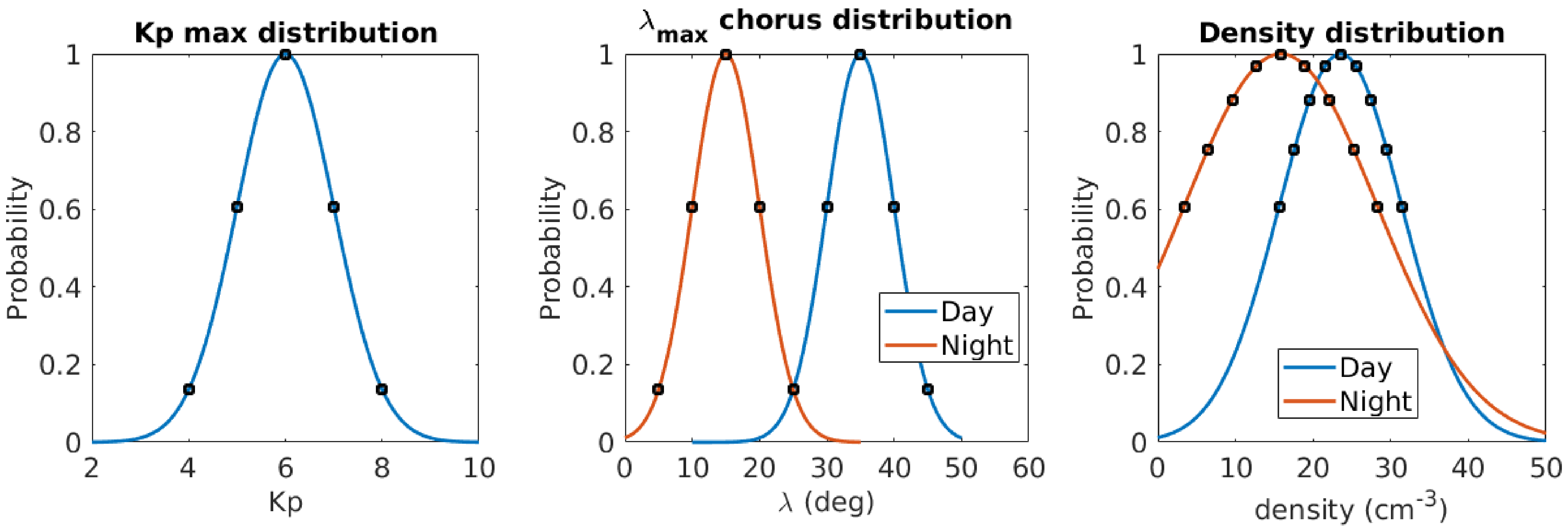
Physics-based forecast of radiation belt electron flux

- Run in real-time 24/7-ish
- One deterministic simulation!
- This is the ground truth



http://rbm.epss.ucla.edu/realtime-forecast/

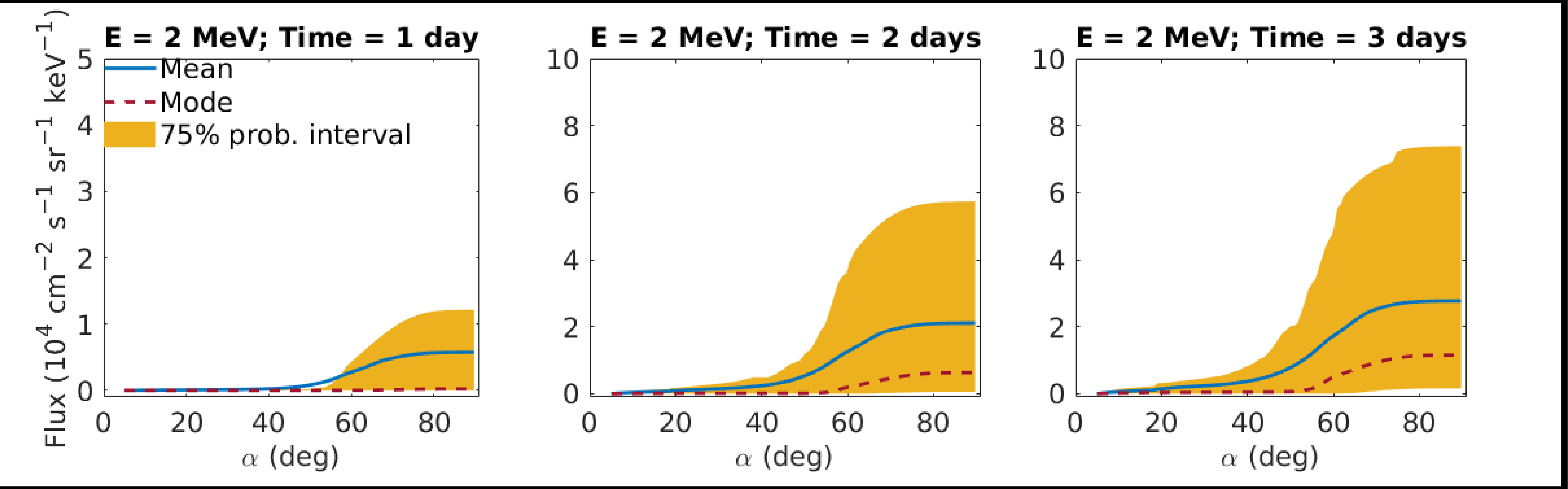
Ensemble simulations of RB electron flux via importance sampling



Three input parameters assumed independent and normally distributed.

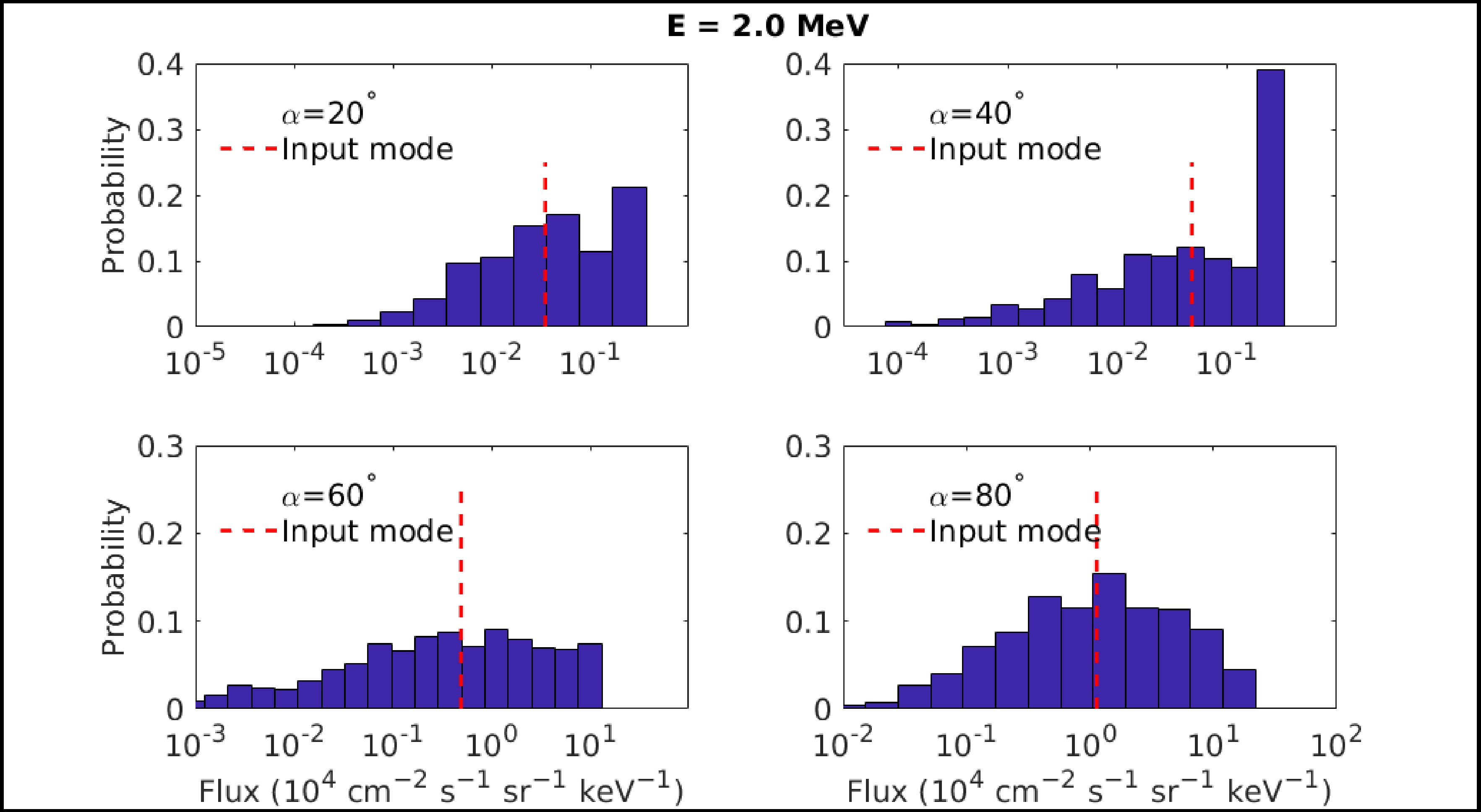
- 225 simulations in total

Ensemble simulations of RB electron flux via importance sampling



Ref: Camporeale et al. (2016) On the propagation of uncertainties in radiation belt simulations, *Space Weather, 14, 982*

Ensemble simulations of RB electron flux via importance sampling



Two orders of magnitude uncertainty!

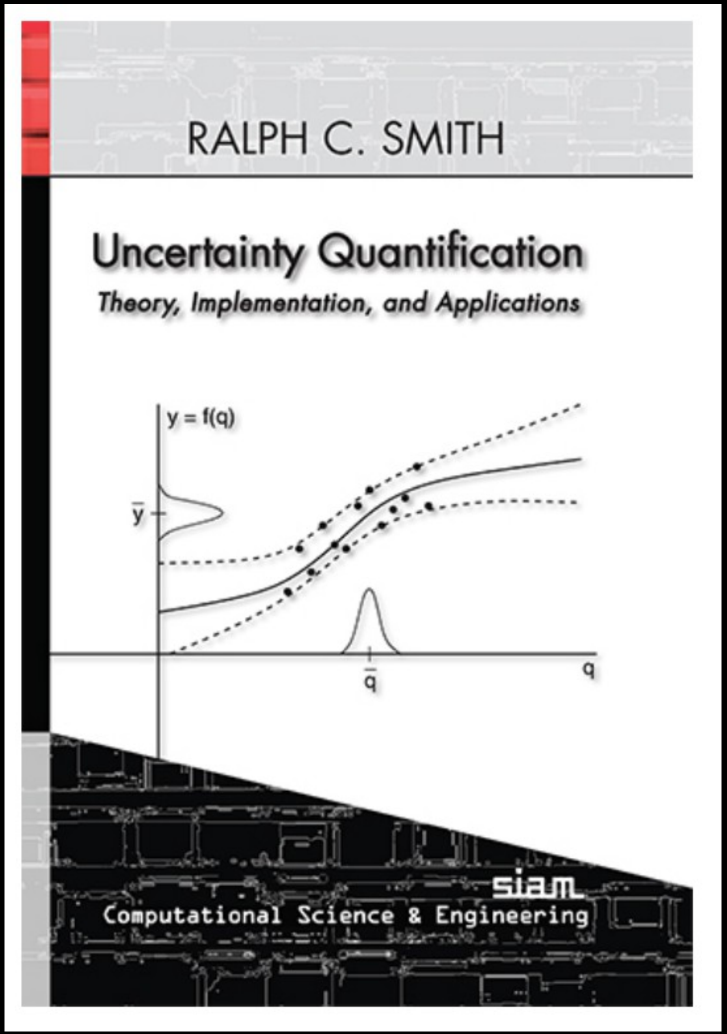
Sampling problem in UQ

- How to sample ensemble members in highly-dimensional parameter space?
- How to beat MC?

Suggested strategy: adaptive, mesh-free approach (RBF)

Proof-of-principle in

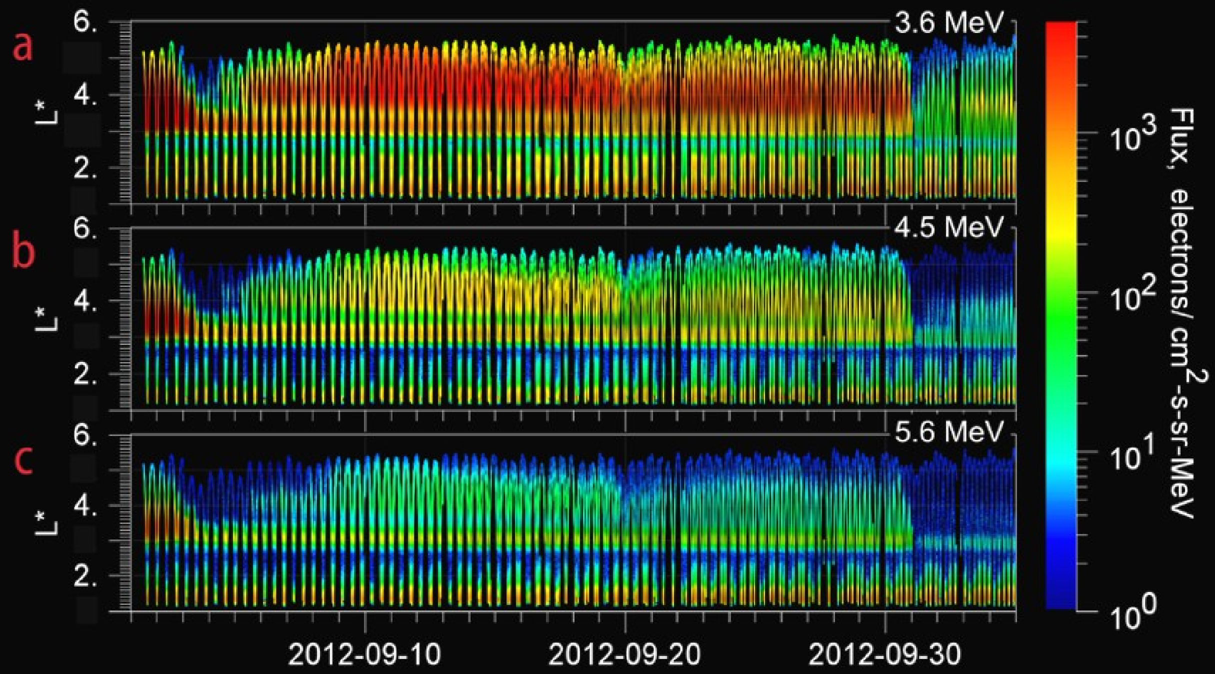
Camporeale et al. (2017) Adaptive selection of sampling points for uncertainty quantification, *Int. J. Uncertainty Quant.*



Bayesian parameter estimation

- Radial diffusion (quasi-linear theory)

$$\frac{\partial f}{\partial t} = \ell^2 \frac{\partial}{\partial \ell}\left(\frac{\kappa(\ell,t)}{\ell^2}\frac{\partial f}{\partial \ell}\right)_{\mathcal{M},J} - \lambda(\ell,t) f$$



Bayesian parameter estimation

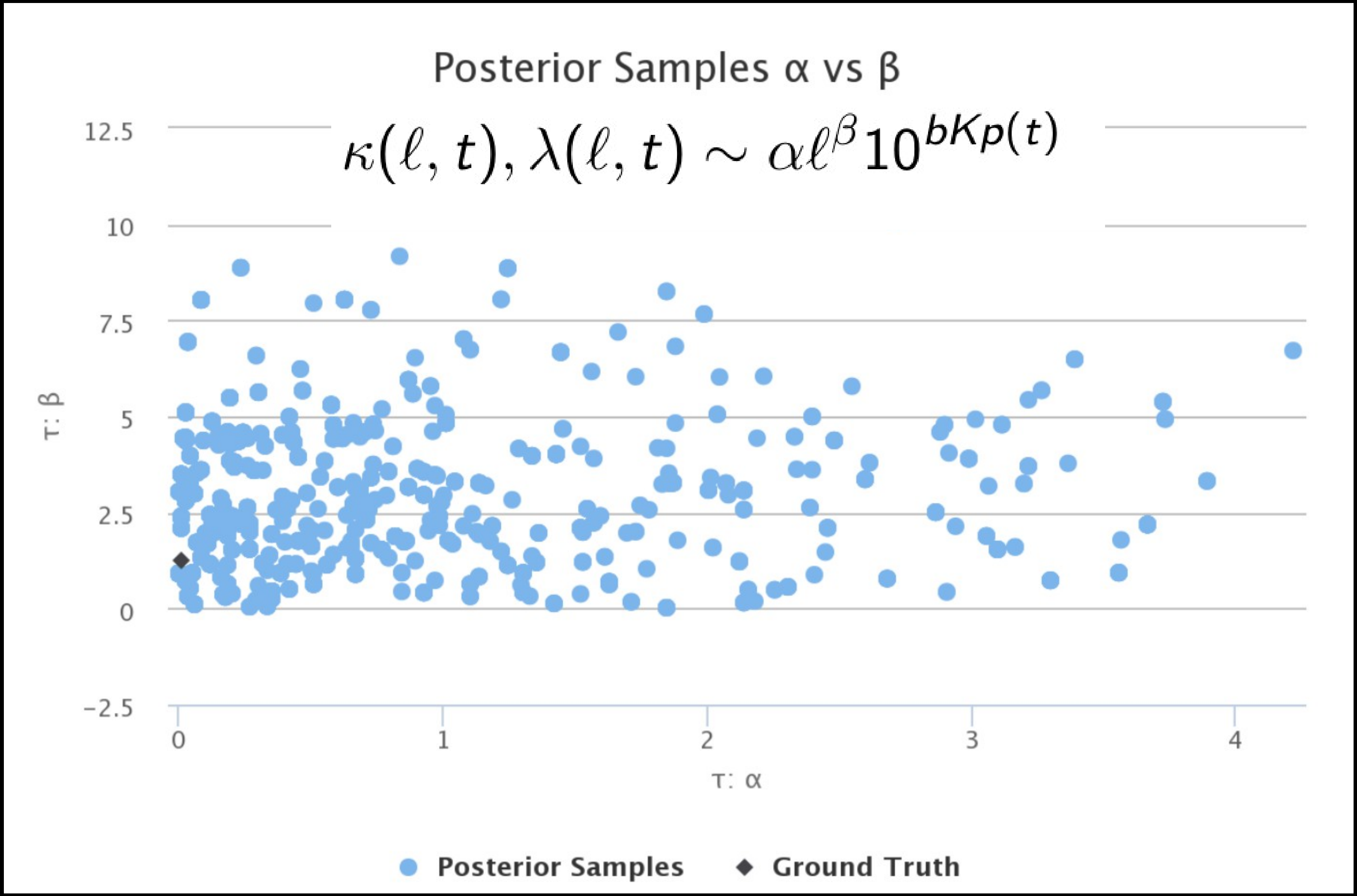
- Parametrize Diff coeff. and loss term

$$\kappa(\ell, t), \lambda(\ell, t) \sim \alpha \ell^{\beta} 10^{bKp(t)}$$

- Quantify likelihood of observed data conditioned on parameters
- Perform MCMC sampling to construct posterior distribution

Bayesian parameter estimation

- Tested on synthetic data (work in progress...)



Gaussian Process

Gaussian Process (GP) models approximate an unknown function $f(\mathbf{x})$ on a finite set of points $\mathbf{x}_1, \mathbf{x}_2, \cdots \mathbf{x}_N$ by a multivariate Gaussian distribution (equation 2).

$$y = f(\mathbf{x}) + \varepsilon \text{ with } \varepsilon \sim \mathcal{N}\left(0, \sigma^2\right) \tag{1}$$

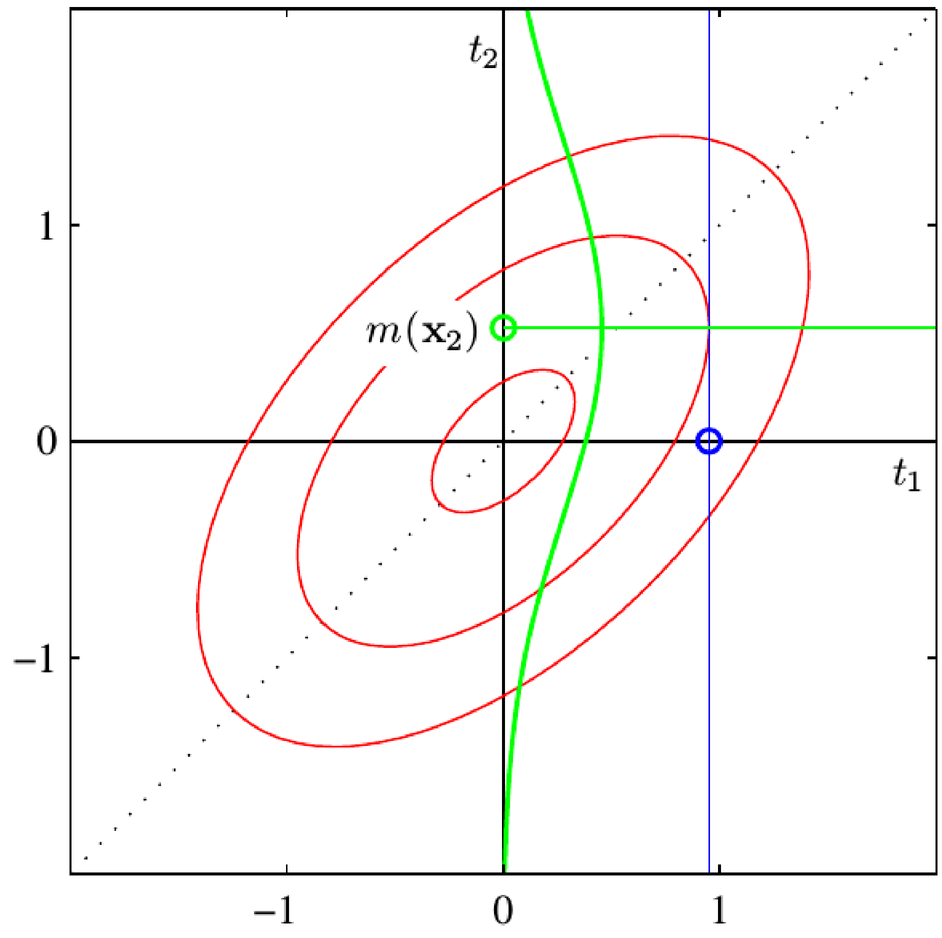
$$\left(f(\mathbf{x}_1), f(\mathbf{x}_2), \cdots f(\mathbf{x}_\mathbf{N})\right)^T \sim \mathcal{N}\left(\mu, \Sigma\right) \tag{2}$$

$$\mu_i = m(\mathbf{x}_i) \text{ (Mean)} \tag{3}$$

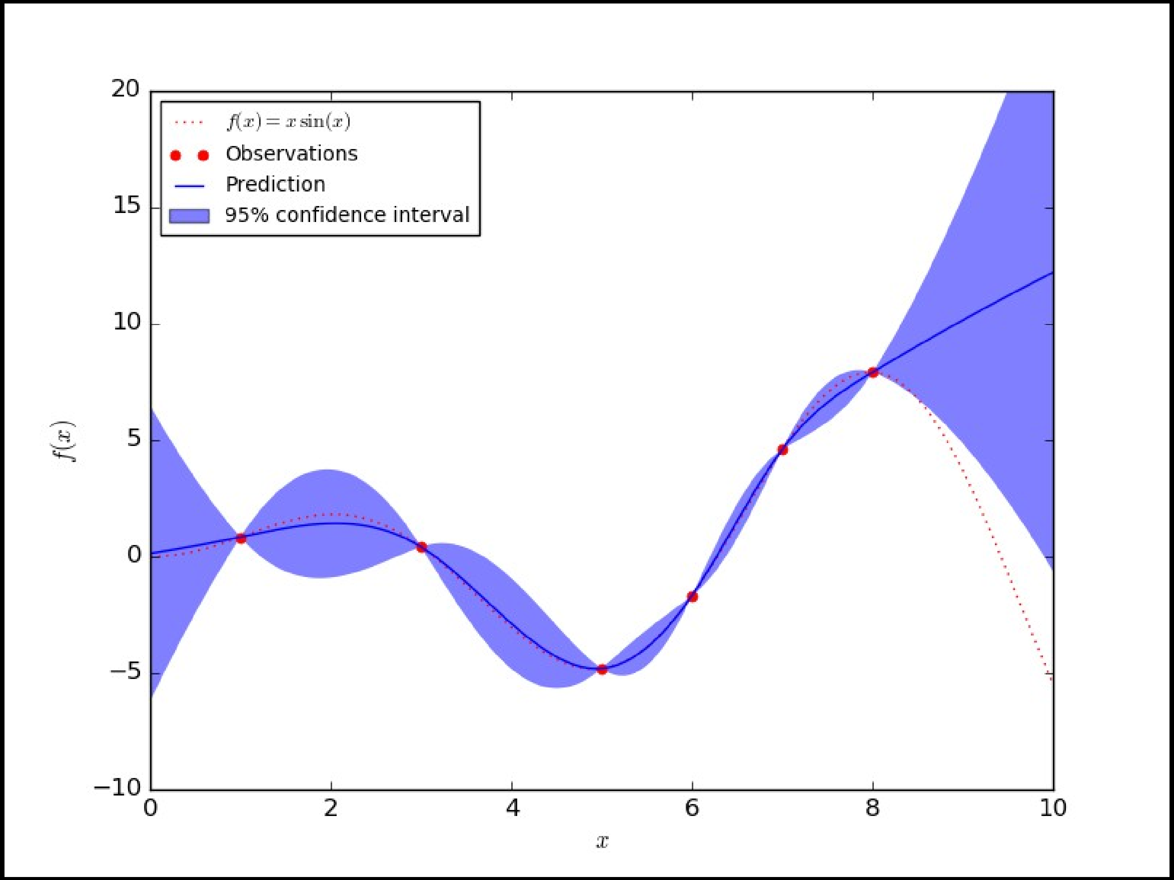
$$\Sigma_{ij} = K(\mathbf{x}_i, \mathbf{x}_j) \text{ (Covariance matrix)} \tag{4}$$

Gaussian Process Regression

Illustration of the mechanism of Gaussian process regression for the case of one training point and one test point, in which the red ellipses show contours of the joint distribution $p(t_1, t_2)$. Here t_1 is the training data point, and conditioning on the value of t_1, corresponding to the vertical blue line, we obtain $p(t_2|t_1)$ shown as a function of t_2 by the green curve.



Gaussian Process Regression



The output of GP regression is a Gaussian distribution (i.e. a mean value and a standard deviation) → probabilistic forecast

GP regression for one-hour ahead DST prediction

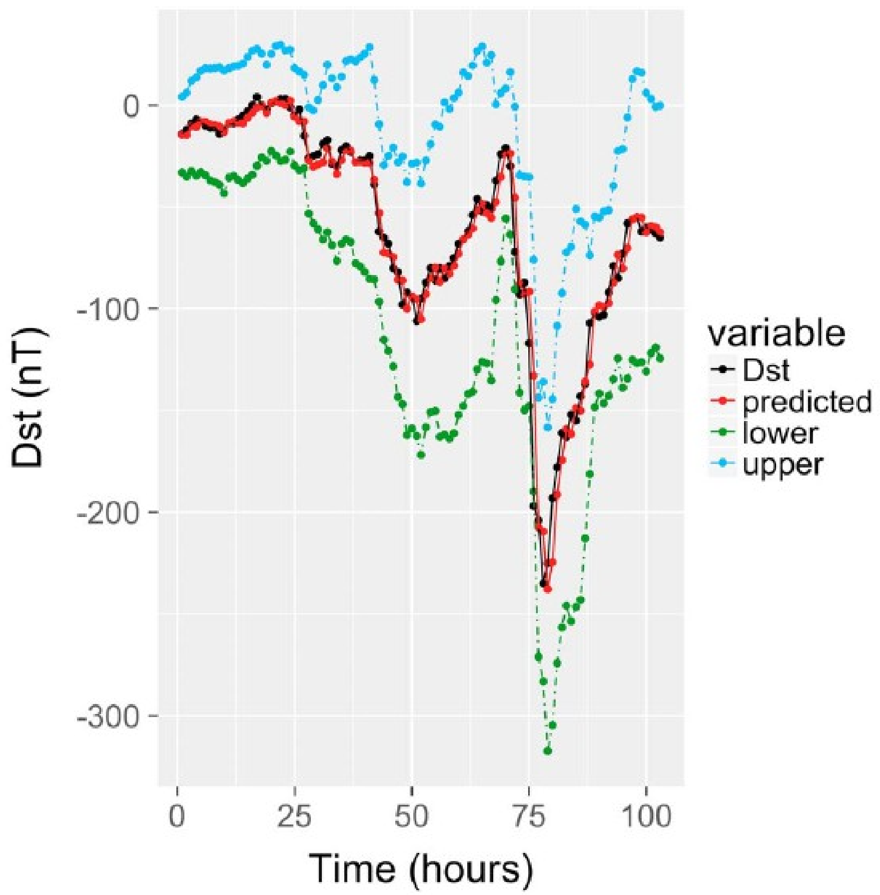
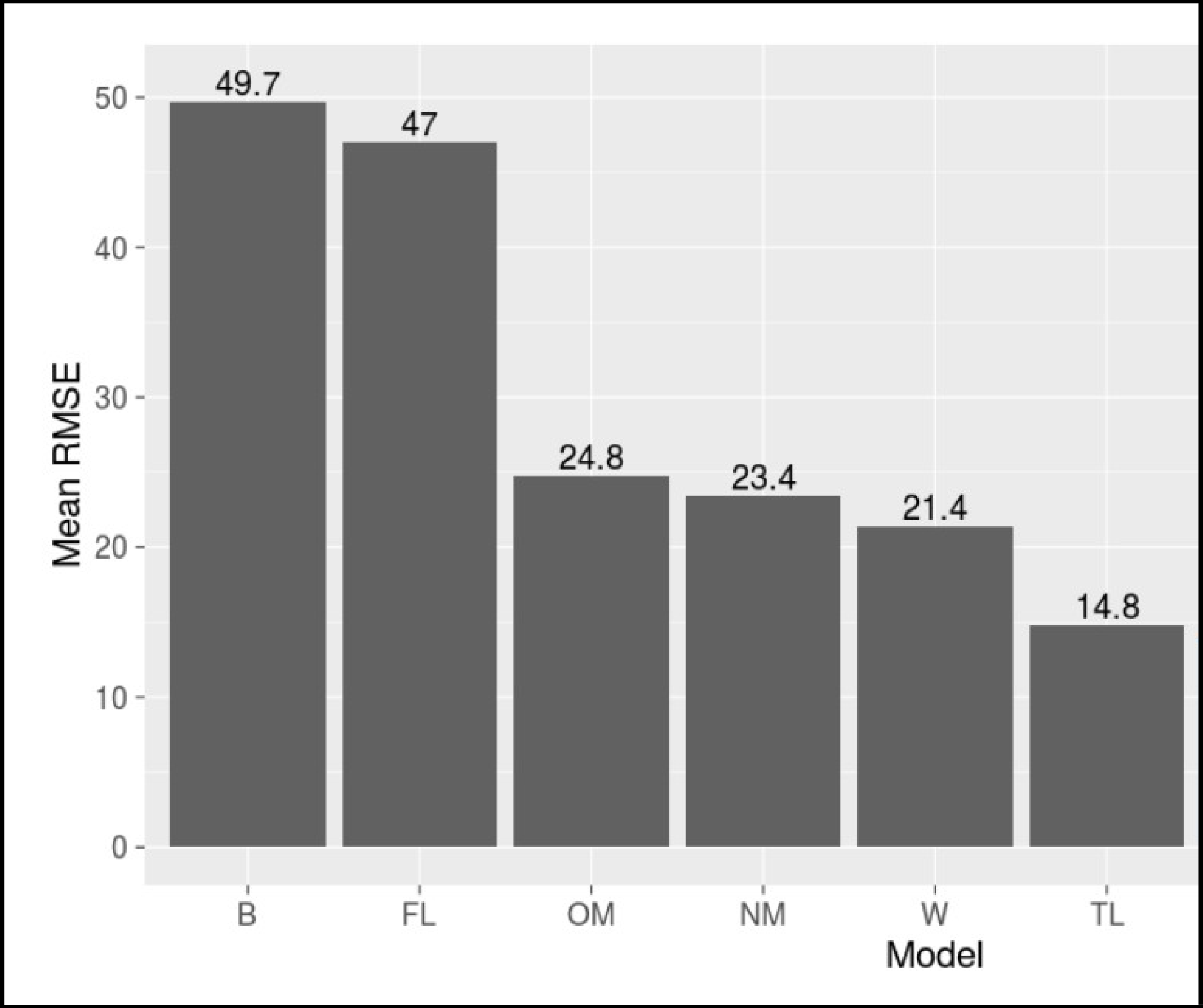


Figure 6. OSA predictions with $\pm\sigma$ error bars for event: 8 March to 10 March 2012.

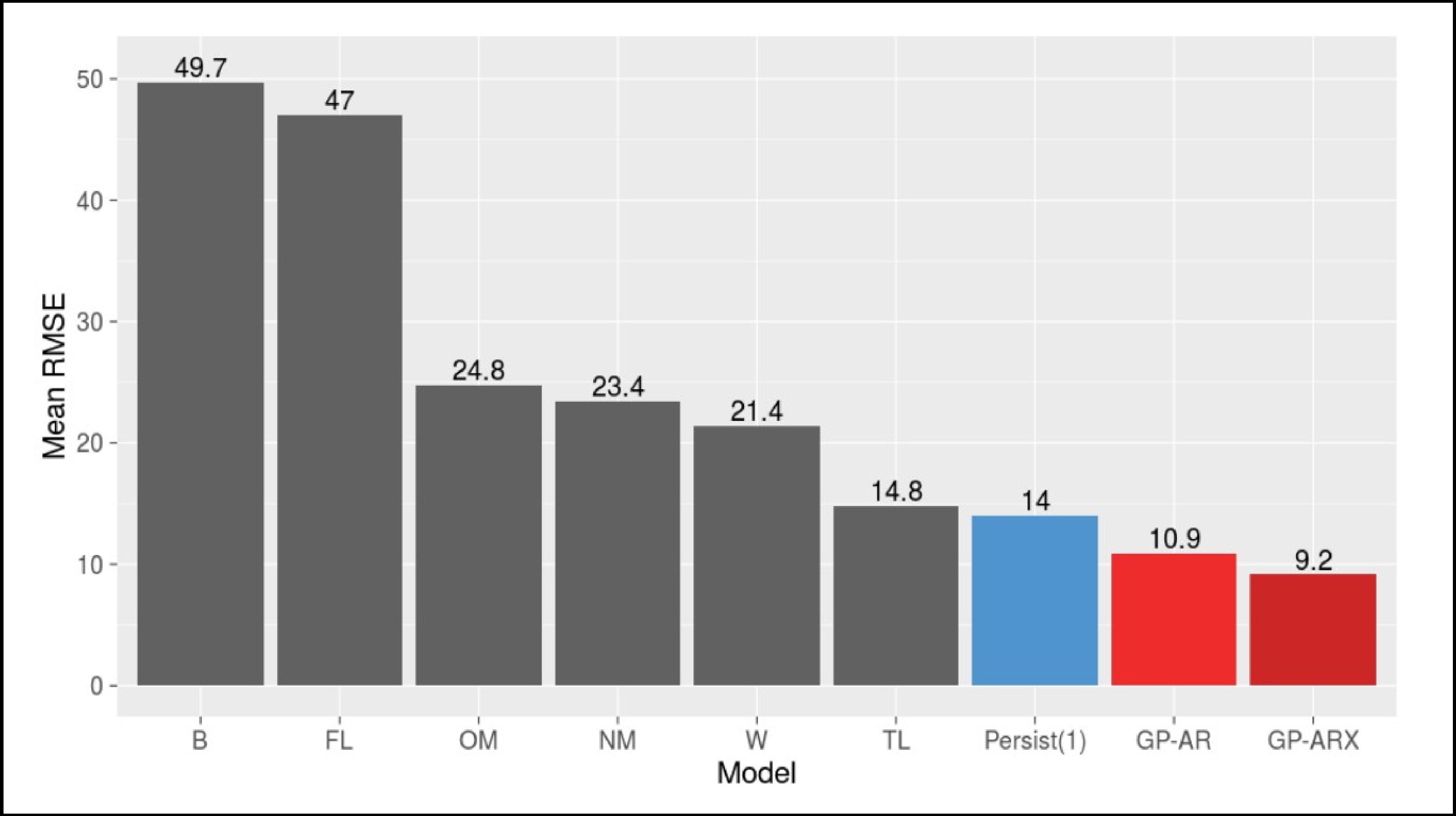
Ref: Chandorkar et al. *Space Weather* (2017)

GP regression for one-hour ahead DST prediction



Ref: Chandorkar et al. *Space Weather* (2017)
63 storms from Ji et al. *JGR* (2012)

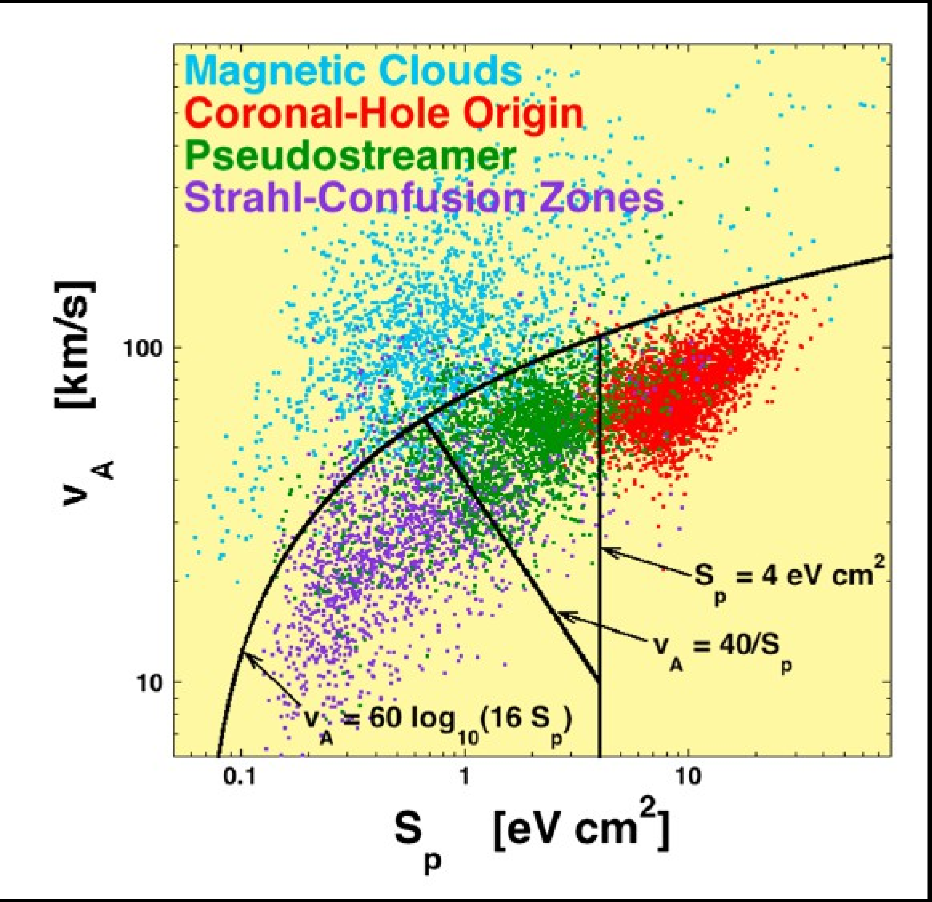
GP regression for one-hour ahead DST prediction



Ref: Chandorkar et al. *Space Weather* (2017)
63 storms from Ji et al. *JGR* (2012)

GP for solar wind classification

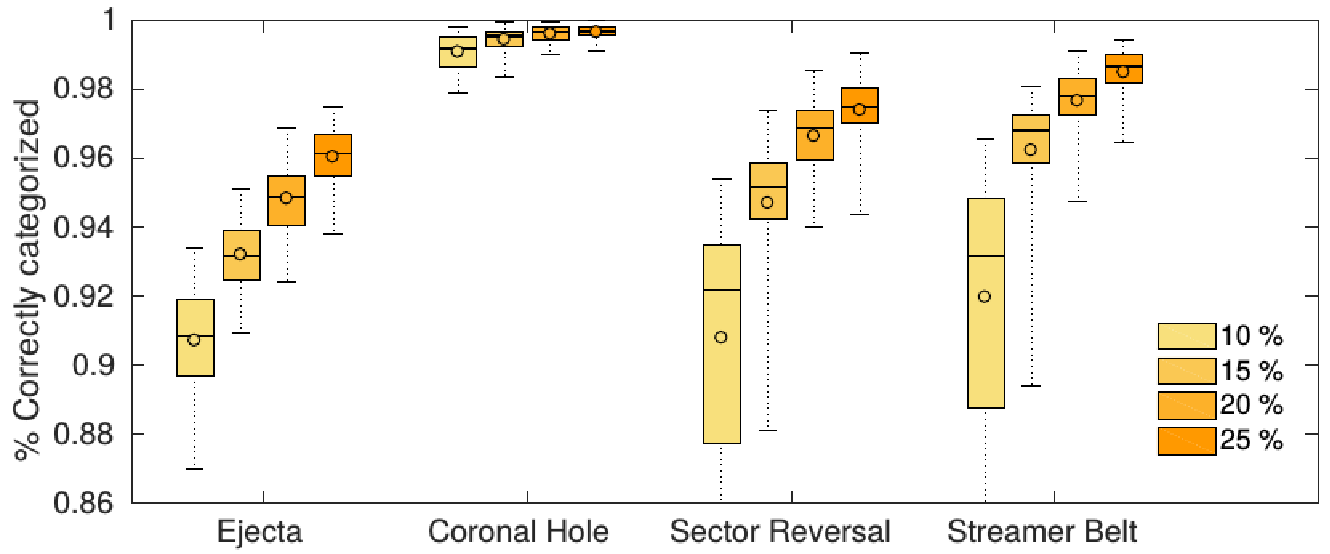
Xu & Borovsky, *JGR* (2014)



GP for solar wind classification

Table 1. List of attributes

Attribute	Symbol
Solar wind speed	V_{sw}
Proton temperature standard deviation	σ_T
Sunspot number	R
Solar radio flux (10.7 cm)	$f_{10.7}$
Alfven speed	v_A
Proton specific entropy	S_p
Temperature ratio	T_{exp}/T_p



GP for solar wind classification

Table 4. Confusion matrix for the case of 20% training set, when only probabilities larger than 50% are considered. Probabilities are conditioned on the observed category.

	Observed category			
Prediction	Ejecta	Coronal hole	Sector reversal	Streamer belt
Ejecta	97.9	0	0.8	0.6
Coronal hole origin	0.2	100	0.1	0.1
Sector reversal origin	1.0	0.0	98.5	0.3
Streamer belt origin	1.0	0.0	0.5	99.0

GP for solar wind classification

- More statistics available: ROC curves, Reliability diagram, transition probability, etc.
- 200k+ hours of OMNI database (1965-2017 classified)
- Database available on mlspaceweather.org (not yet...)
- Paper under review...

Conclusions

- Need for probabilistic forecasts
- Bayesian framework can help for:
  - Parameter estimation
  - Probabilistic regression
  - Probabilistic classification
- Data-driven (ML informed) ensemble simulations
  - Open problem for Uncertainty Quantification/Propagation