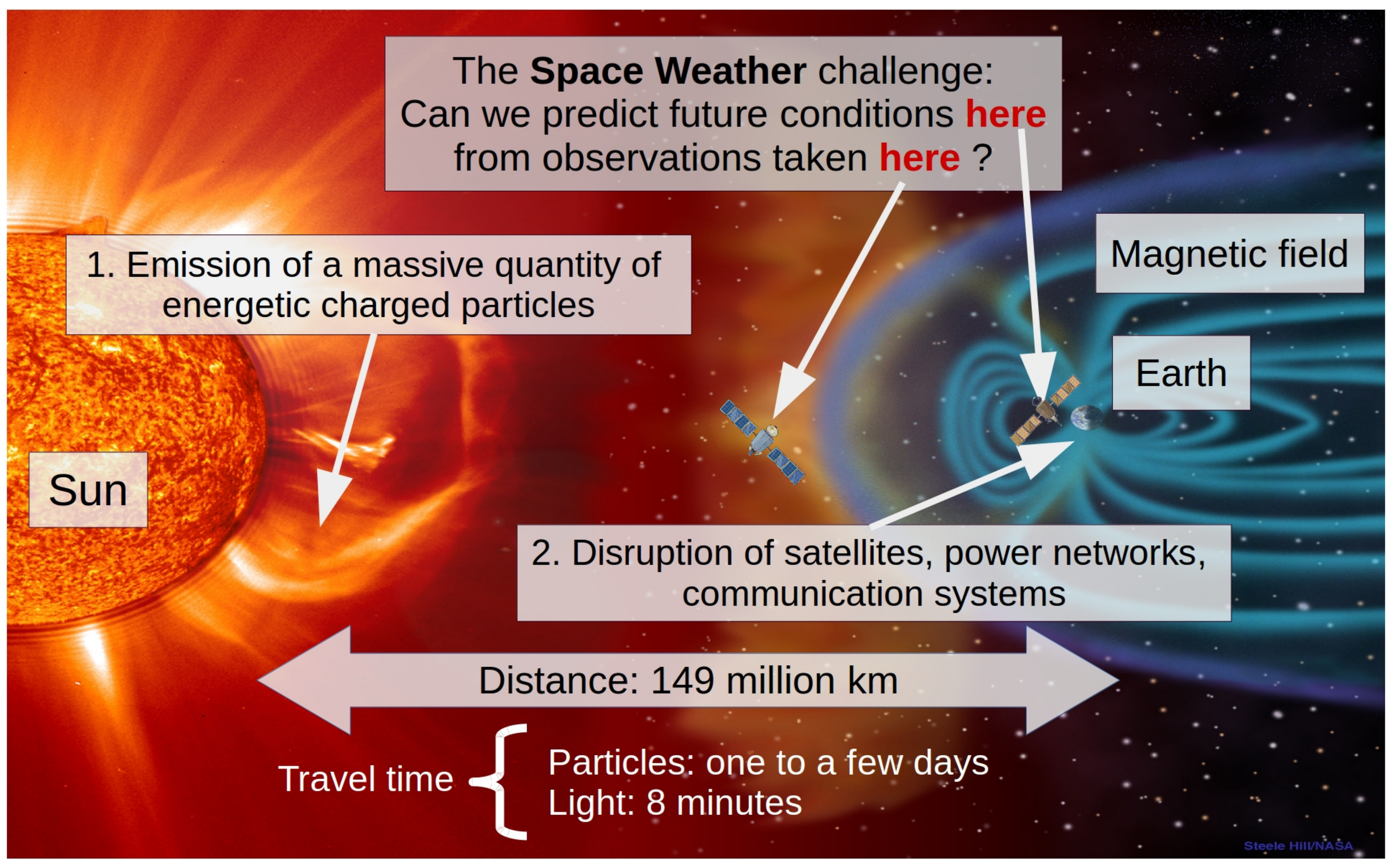
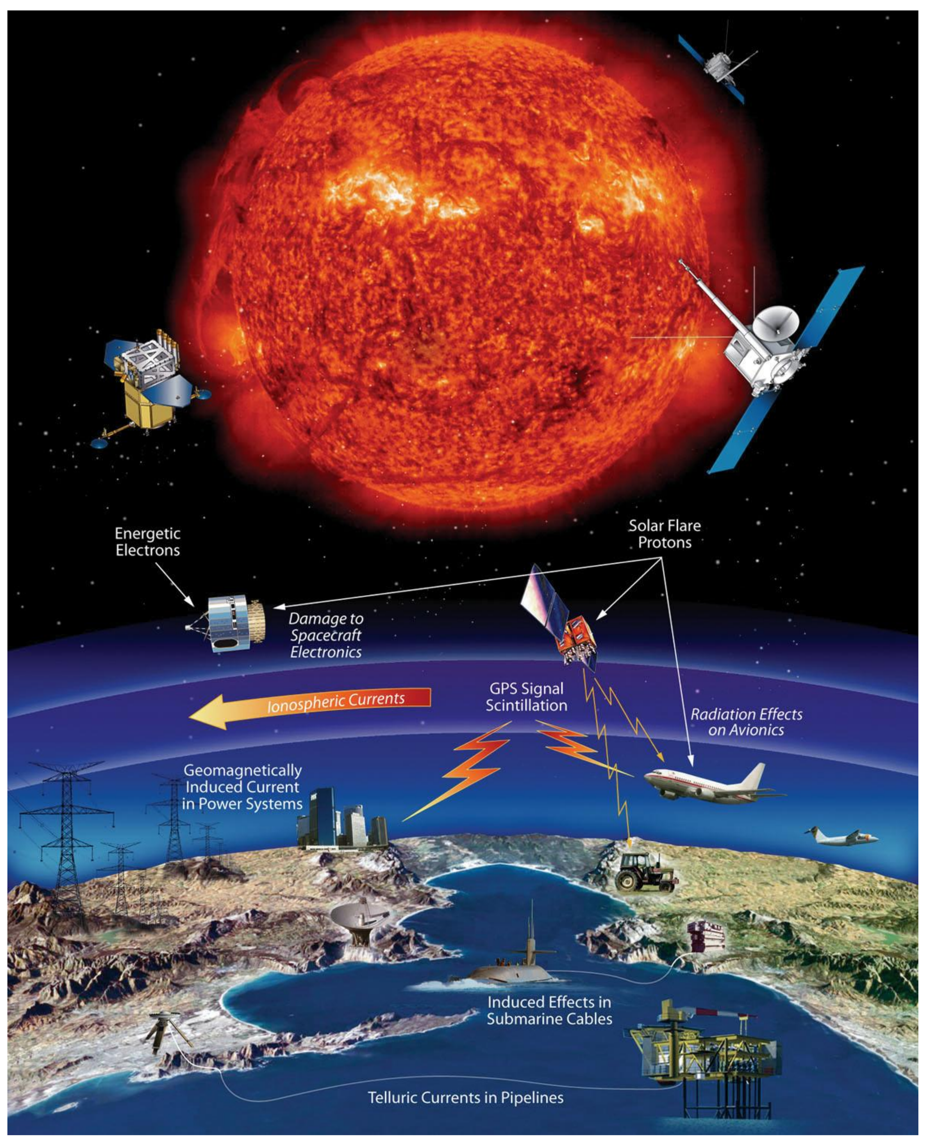
Machine Learning for Space Weather

Mandar Chandorkar, Enrico Camporeale
Multiscale Dynamics, CWI, Amsterdam
http://mlspaceweather.org/

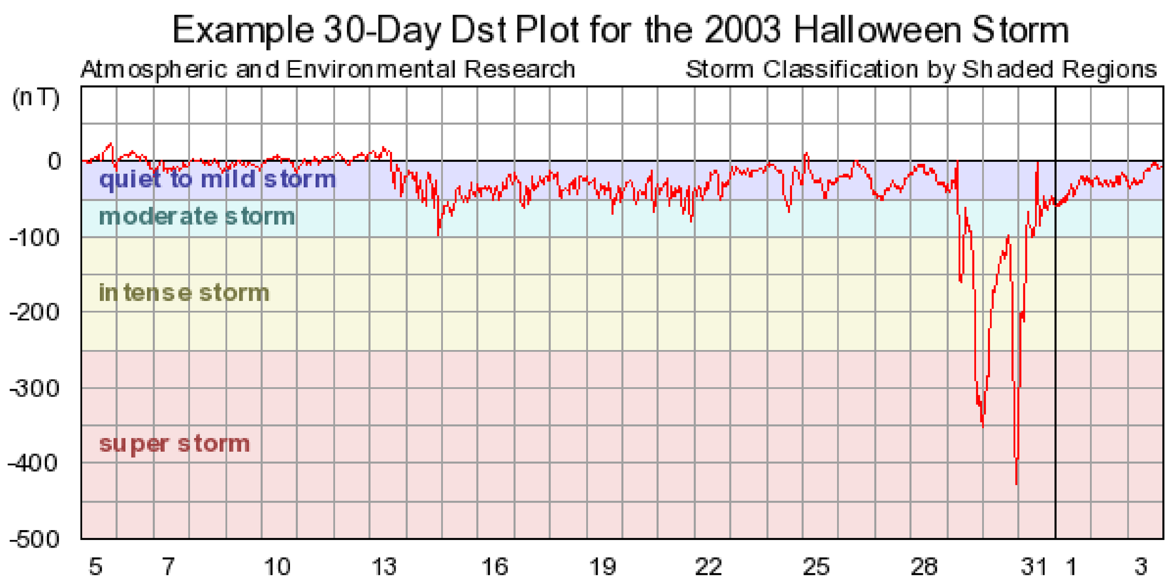
Space weather is the branch of space physics concerned with the prediction of the conditions close to Earth (the magnetosphere) driven by the variability of the Sun.

- Coronal Mass Ejections (CMEs) can cause Geomagnetic Storms at Earth and induce extra currents in the ground that can affect power grid operations.
- Geomagnetic storms can also modify the signal from radio navigation systems (GPS and GNSS) causing degraded accuracy and produce auroras.



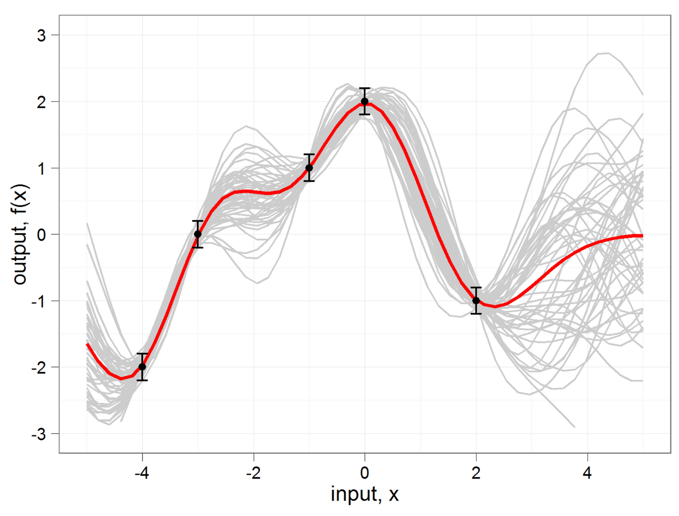


Geomagnetic Activity and Indexes



Due to the complex nature of geomagnetic response to the solar wind, it is useful to use activity indexes to record and predict the magnetosphere's state. The Dst index is an index of magnetic activity derived from a network of near-equatorial geomagnetic observatories. Hourly records of Dst are available since 1957.

Gaussian Process Regression



Gaussian Process (GP) models specify statistical distributions over functions. In GP models, the finite dimensional distribution of the output data is a multivariate Gaussian specified by equation 3.

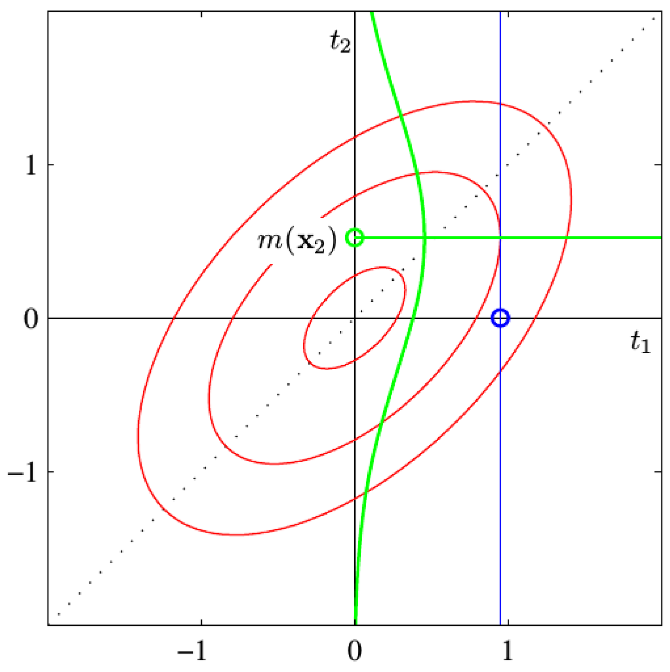
$$y = f(x) + \epsilon \tag{1}$$

$$f \sim \mathcal{GP}(m(x), C(x, x')) \tag{2}$$

$$(\mathbf{y} \;\; \mathbf{f}_*)^T \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) + \sigma^2 I & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right) \tag{3}$$

In order to make predictions using GP models, one must calculate the posterior predictive distribution $\mathbf{f}_*|X, \mathbf{y}, X_*$ which is also a multi-variate Gaussian.

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 Dst prediction

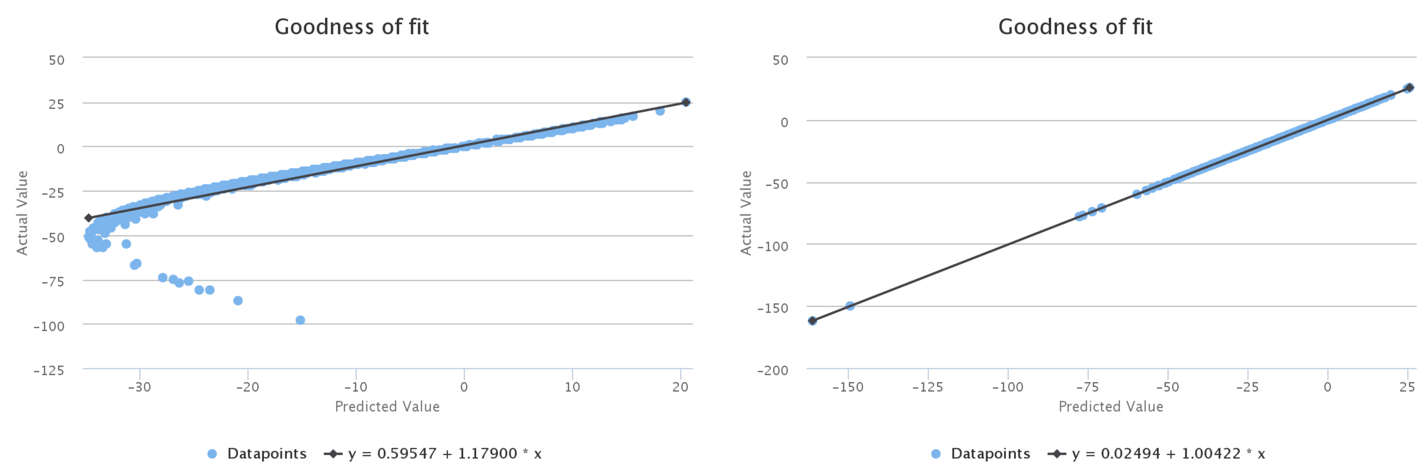
$$Dst(v) \sim \mathcal{GP}(m(v), C(u, v)) \tag{4}$$

$$C_{rbf}(u, v) = \mathbb{E}[Dst(u) \times Dst(v)] = e^{-\frac{1}{2}|u-v|^2/\sigma^2} \tag{5}$$

$$C_{fbm}(u, v) = \mathbb{E}[Dst(u) \times Dst(v)] = |u|^{2H} + |v|^{2H} - |u-v|^{2H} \tag{6}$$

Using the solar wind speed as a predictive variable, we train a Gaussian Process model to predict the Dst index 4. We compare the performance of the *Radial Basis Function* (RBF) kernel versus the *Fractional Brownian Motion* (FBM) kernel (eq. 5,6).

The figures below show the goodness of fit of the two GP models trained and tested for the years 2007 and 2006 respectively (Left: RBF kernel. Right: FBM kernel)



Kernel	Data: Train, Test	MAE	RMSE	R^2
RBF	300,1000	1.5044	6.9752	0.7925
FBM	300,1000	0.0312	0.0461	0.9999

Although the RBF based model gives a satisfactory fit, it performs poorly in predicting outlying events like geomagnetic storms ($Dst \leq -100nT$). the FBM based GP model provides more robust Dst predictions in greatly varying solar wind conditions.