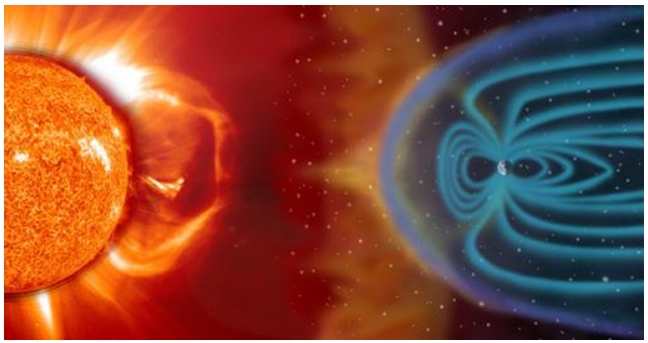
Prediction of Soft Proton Contamination in XMM-Newton: a Machine Learning Approach

E. Kronberg[1], F. Gastaldello[2], S. Haaland[3], A. Smirnov[4], M. Berrendorf[1], S. Ghizzardi[2], K. Kuntz[5], N. Sivadas[6], R. Allen[5], A. Tiengo[2], R. Ilie[7], Y. Huang[7] and L. Kistler[8]

[1]LMU, Germany; [2]INAF-IASF, Italy; [3]University of Bergen, Norway; [4]GFZ, Germany; [5]JHU, USA; [6]BU, USA; [7]University of Illinois, USA and [8]UNH, USA

INTERNATIONAL SPACE SCIENCE INSTITUTE

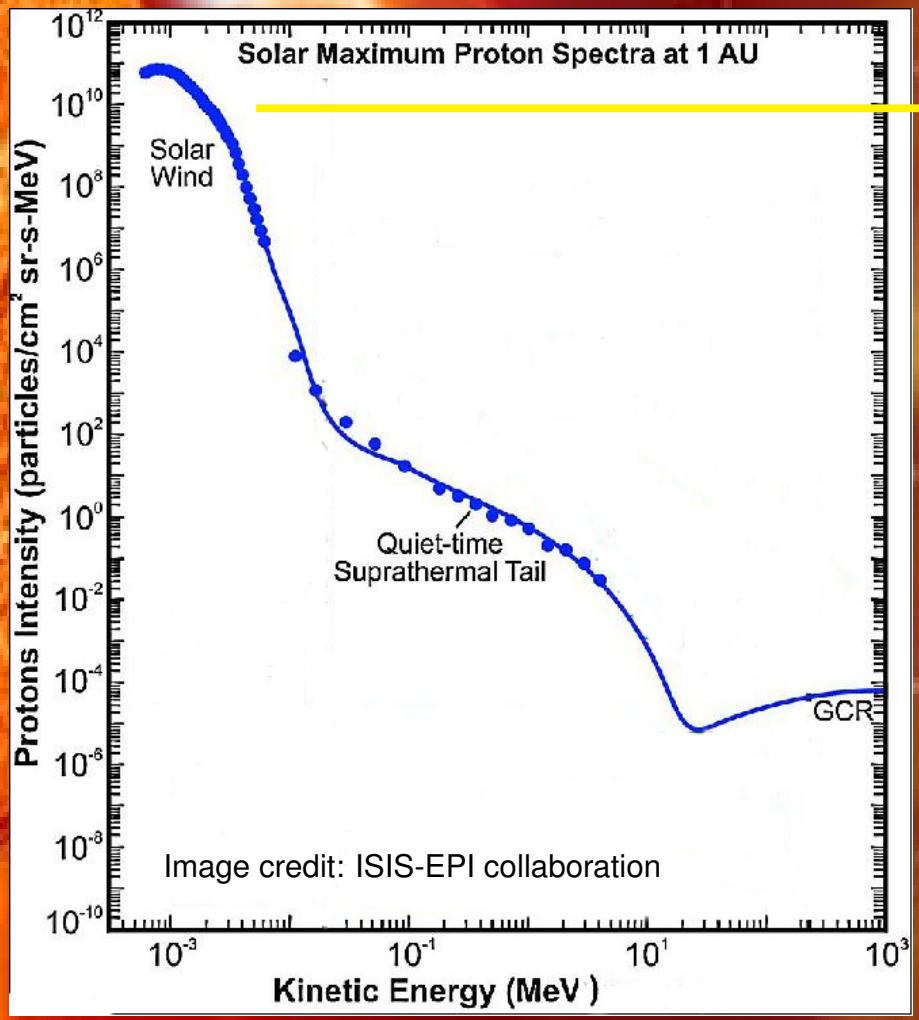


DFG Deutsche Forschungsgemeinschaft German Research Foundation

VolkswagenStiftung

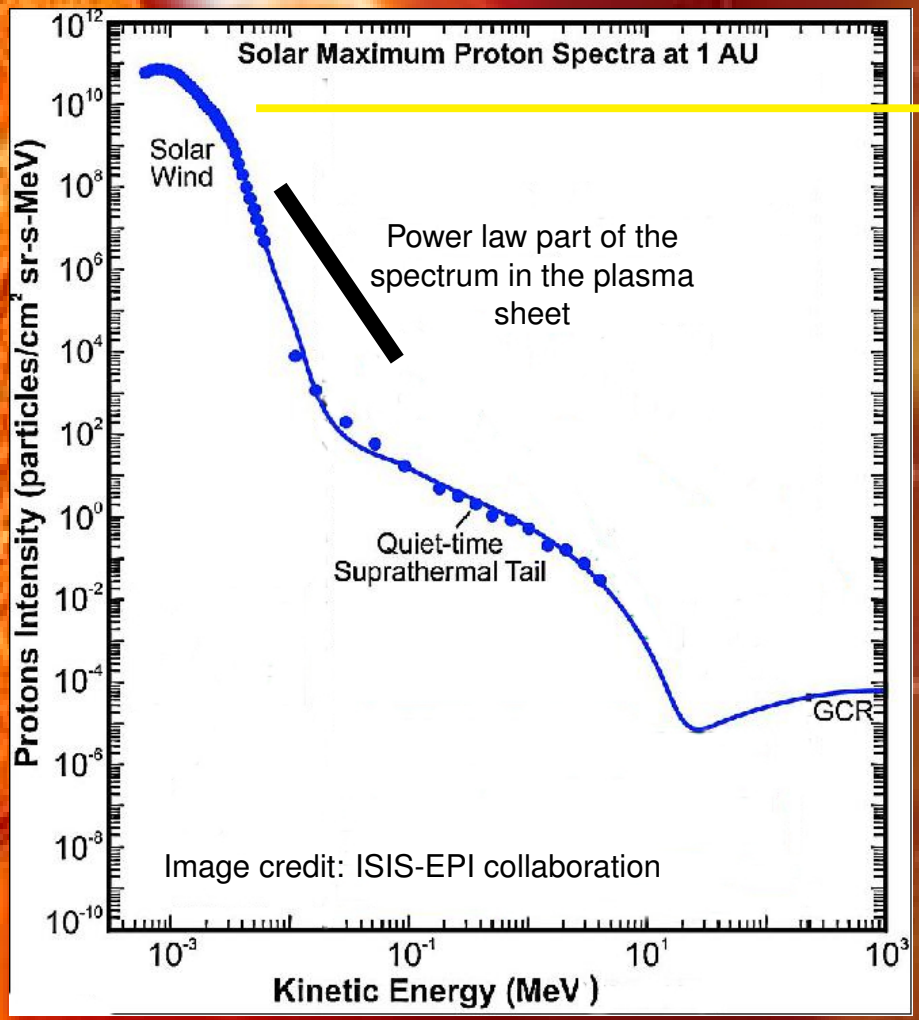
The magnetosphere is an effective proton accelerator

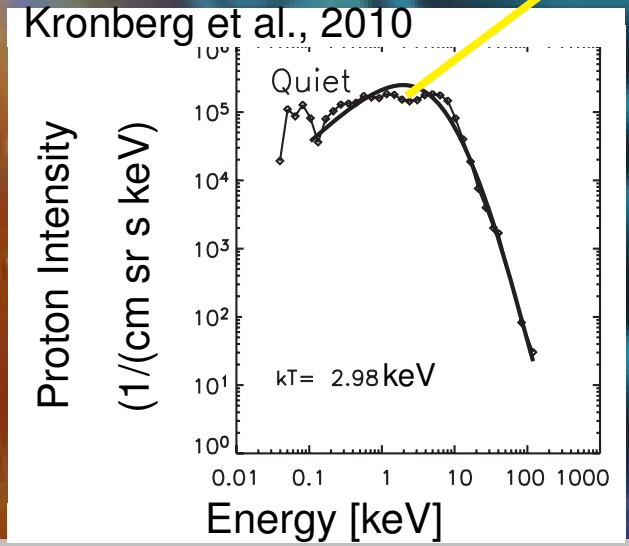
- Proton spectrum in the solar wind: $E = 1\,\mathrm{keV}$



The magnetosphere is an effective proton accelerator

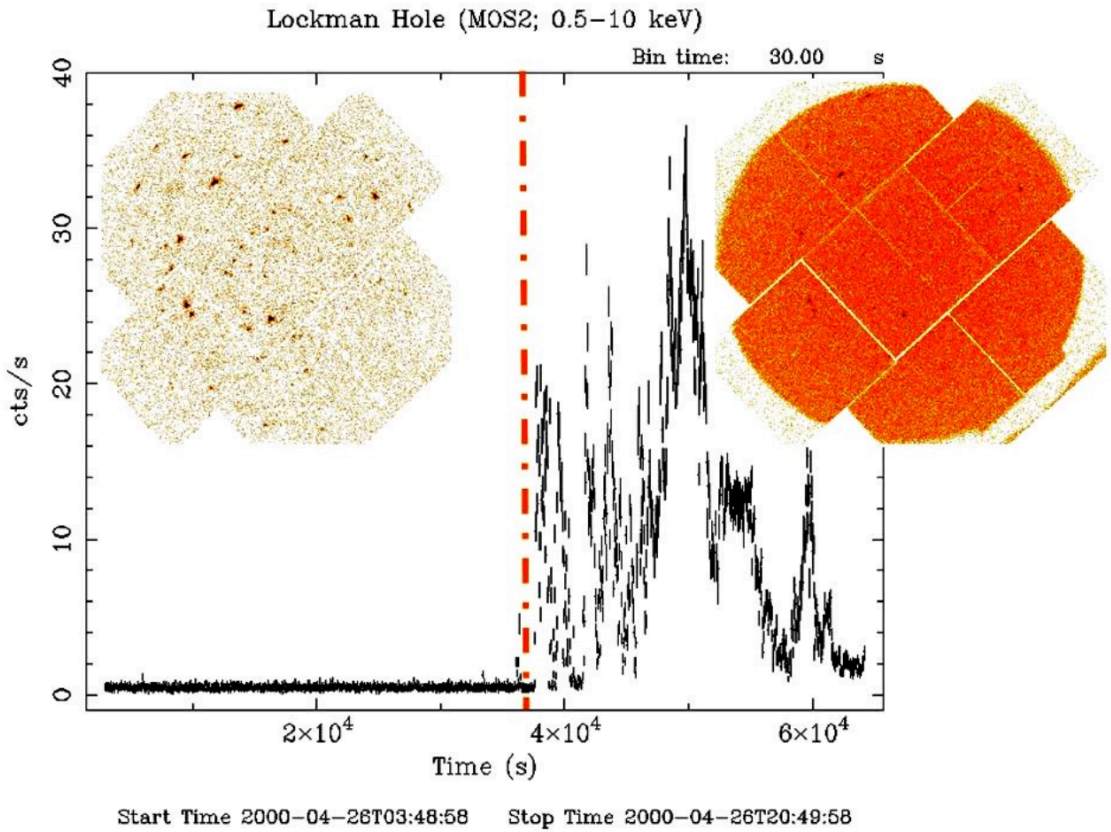
- Proton spectrum in the solar wind: $E = 1\,\mathrm{keV}$
- Proton spectrum in the plasma sheet: $E \sim 3\,\mathrm{keV}$ + hard energetic tail





Motivation: why do we care about energetic protons?

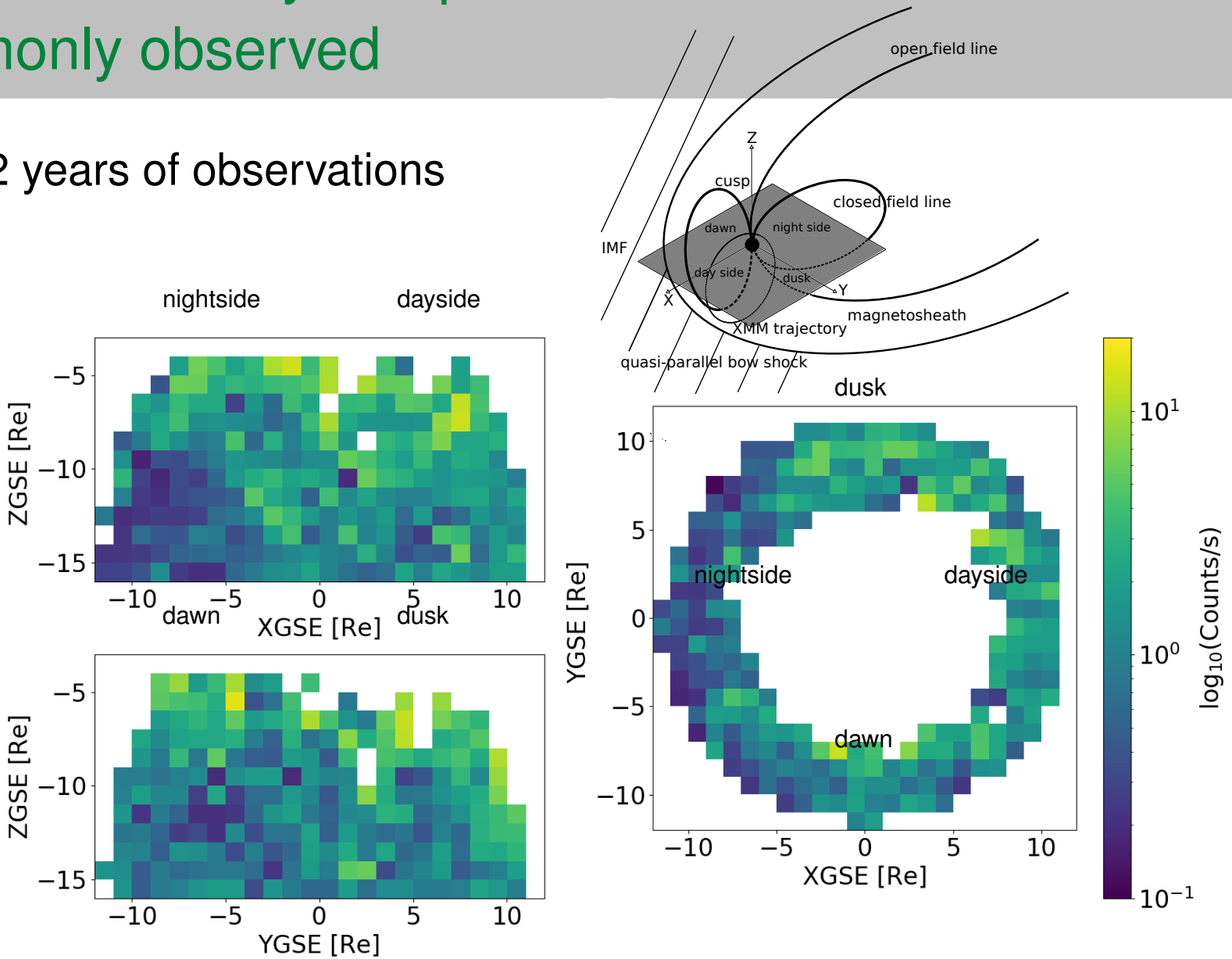
- Example: observations by XMM Newton (X ray telescope) are affected by $\sim$100 keV protons
- Astronomers need to understand the sources of this contamination and to predict it



Kronberg et al., ApJ, 2020

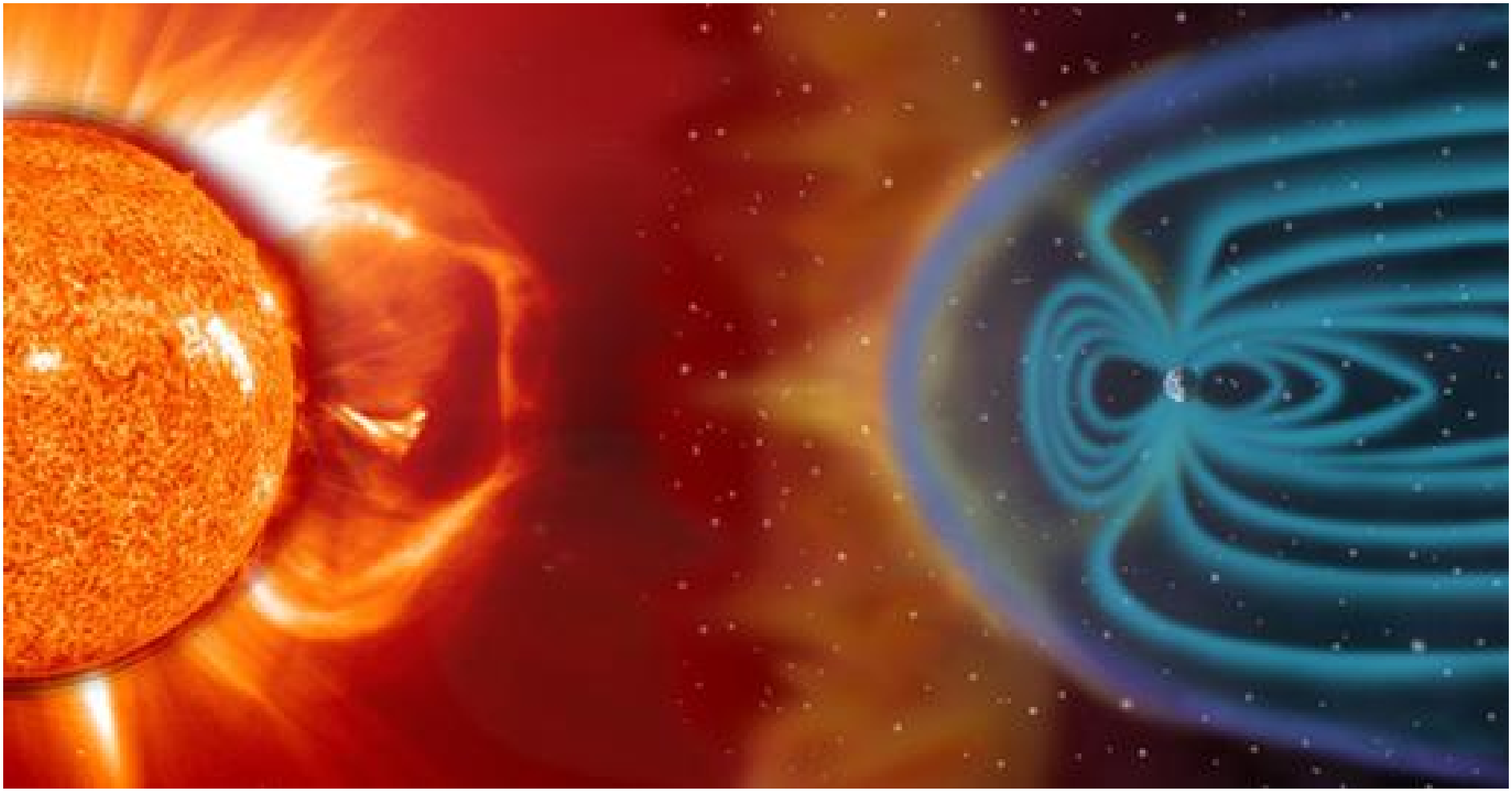
Contamination by soft protons is commonly observed

- 12 years of observations

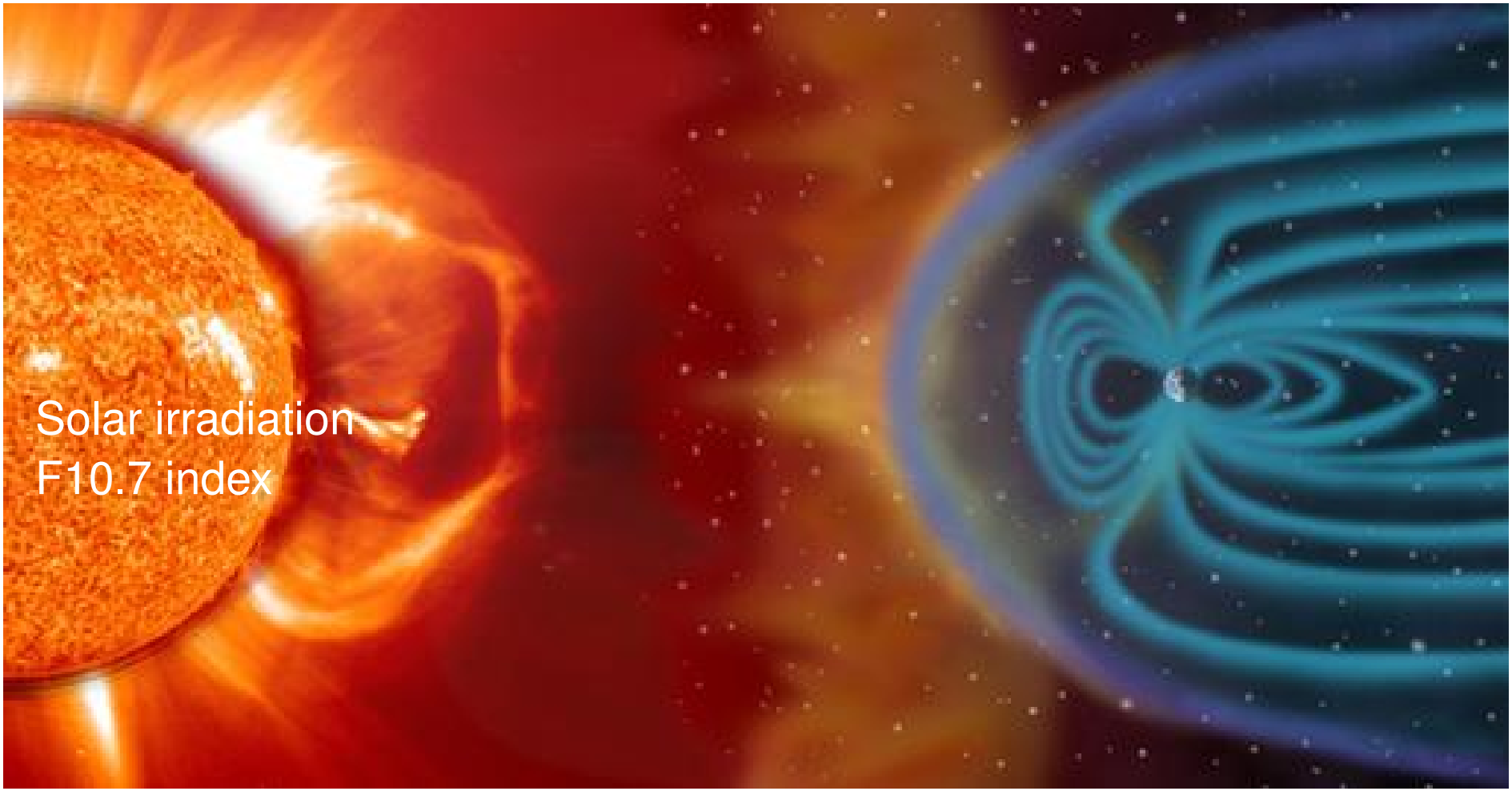


Kronberg et al., ApJ, 2020

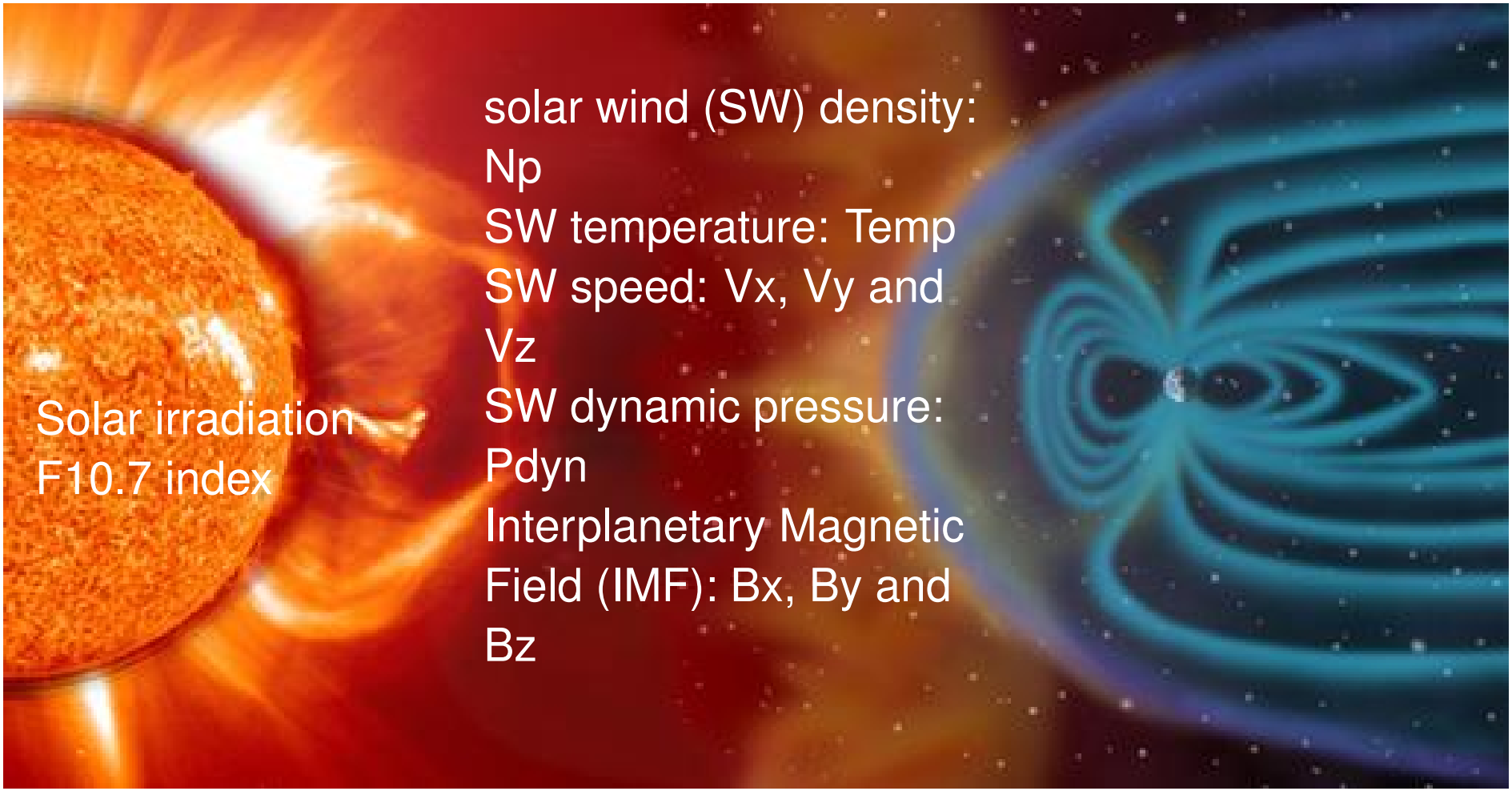
Which parameters affect the contamination?



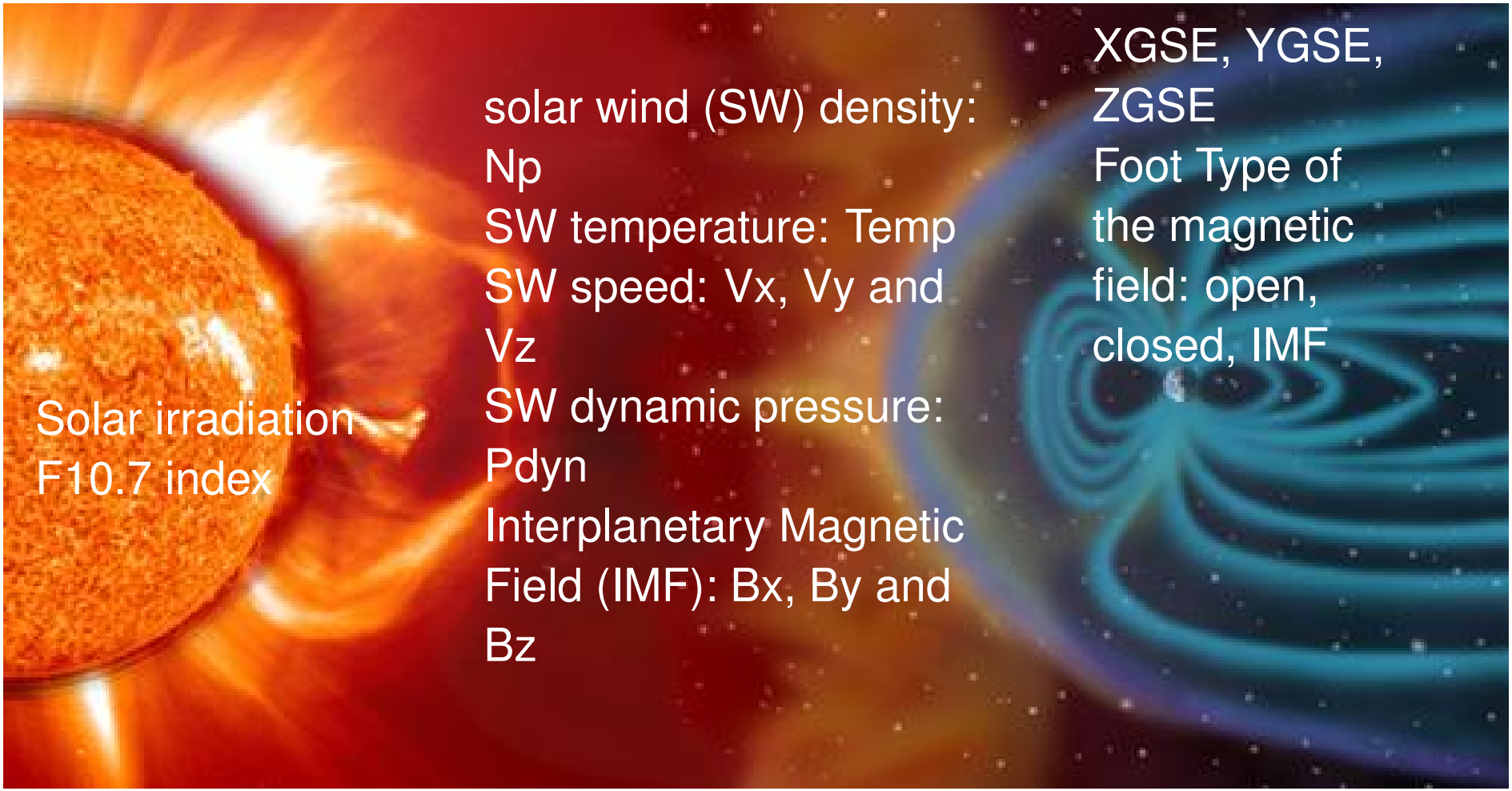
Which parameters affect the contamination?



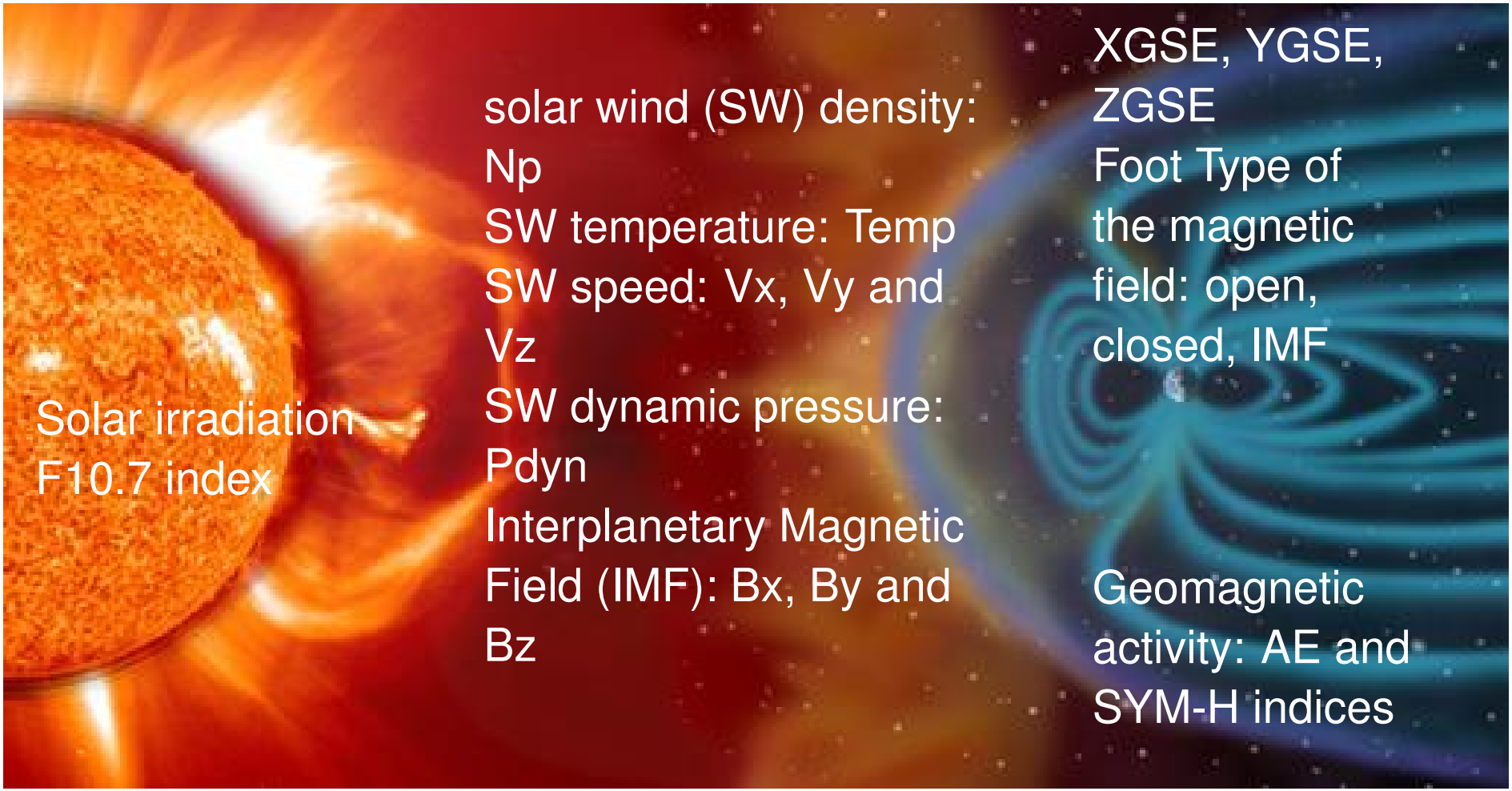
Which parameters affect the contamination?



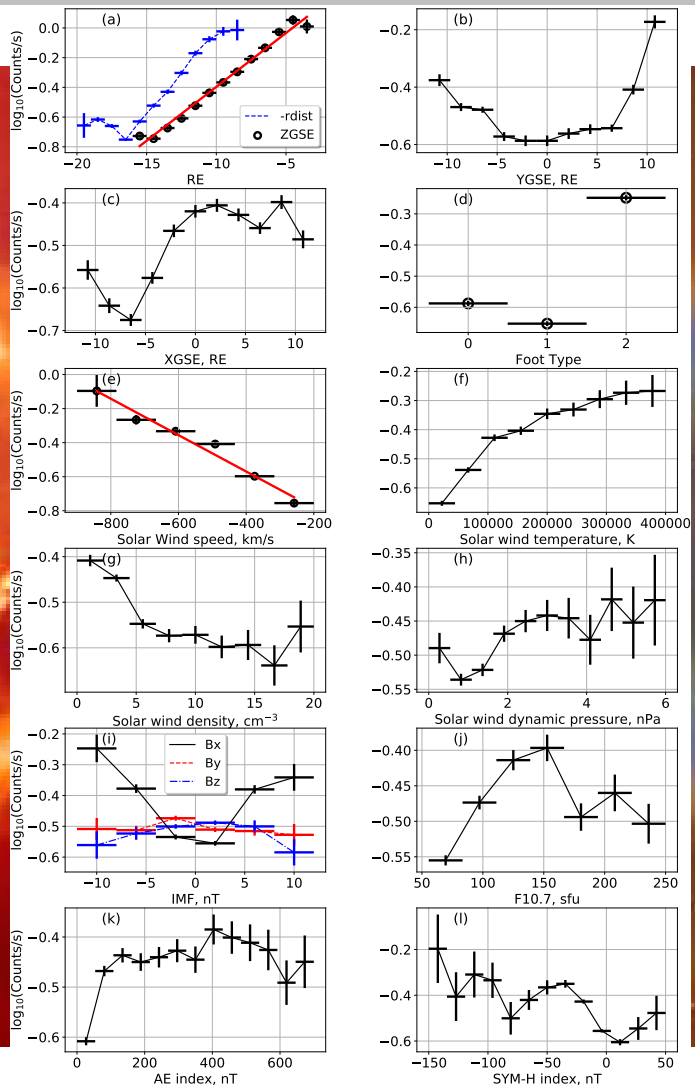
Which parameters affect the contamination?

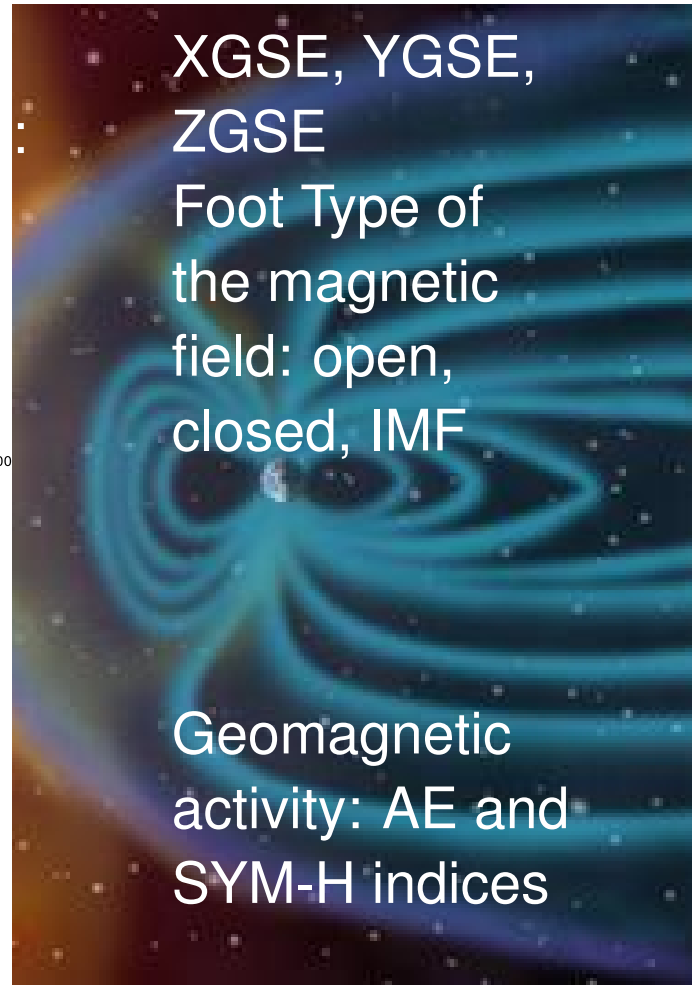


Which parameters affect the contamination?



Some parameters may have complex non-linear relationships with the contamination





We forecast the contamination using location, solar and geomagnetic parameters by regression.

We forecast the contamination using location, solar and geomagnetic parameters by regression.

linear regression

Gradient Boosting

Support Vector Machines

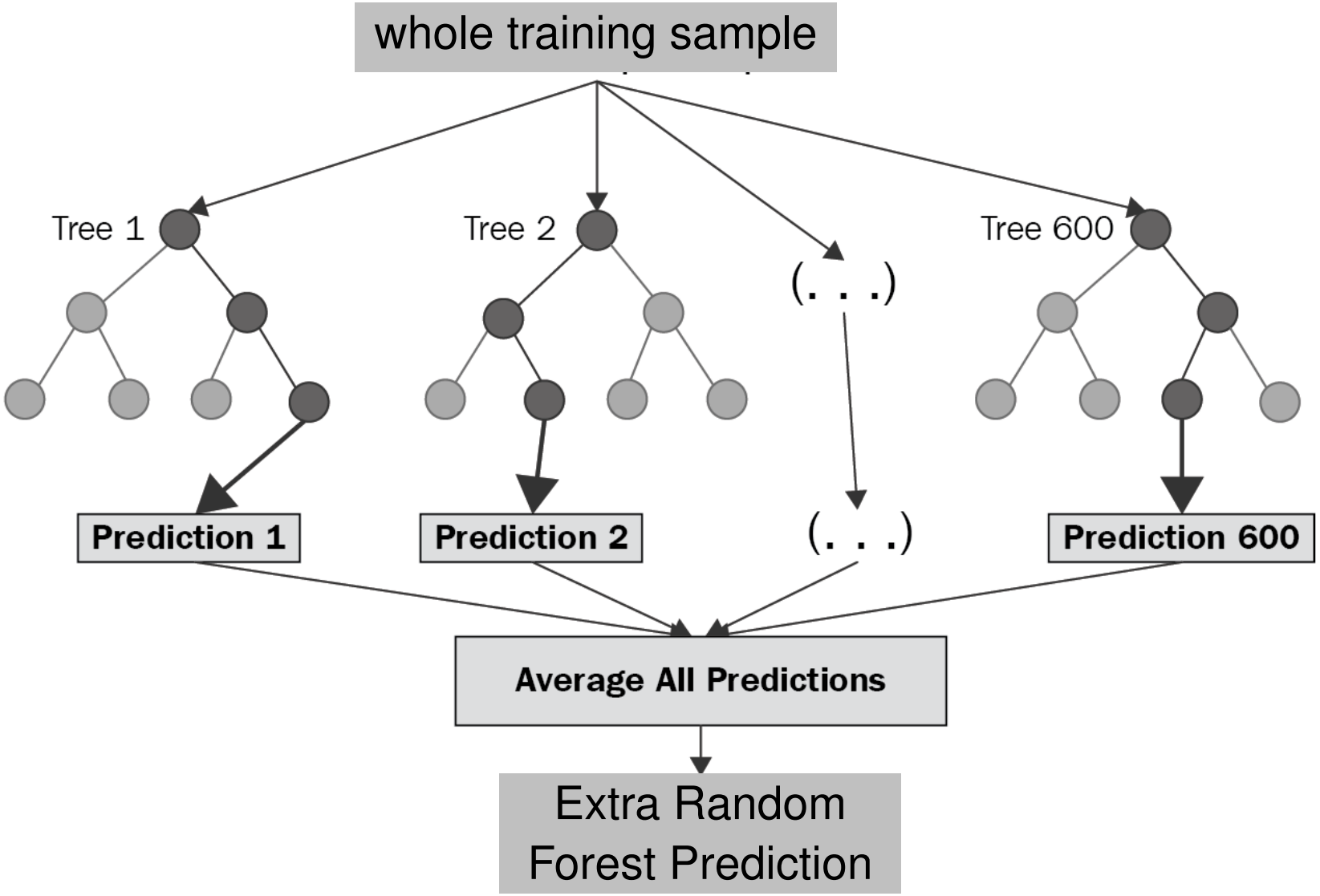
Gradient Descent

Decision Trees and Random Forests

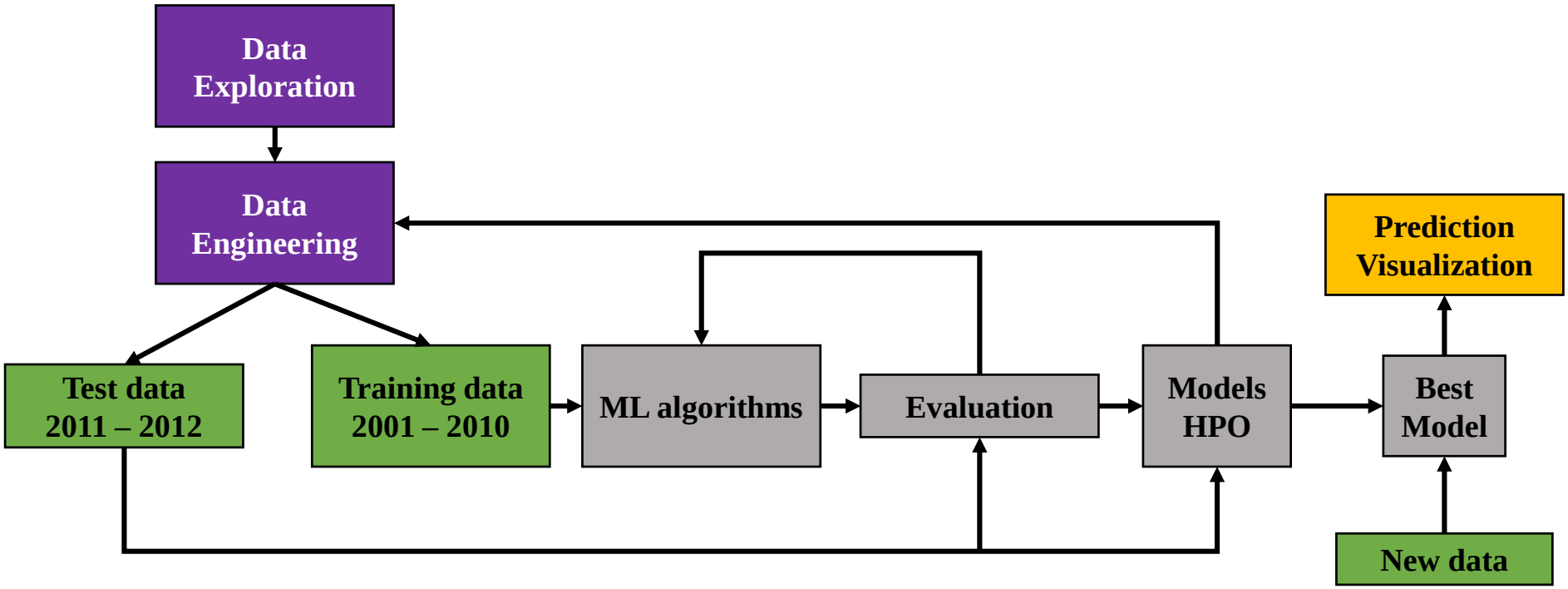
Neural networks

We forecast the contamination using location, solar and geomagnetic parameters by regression.

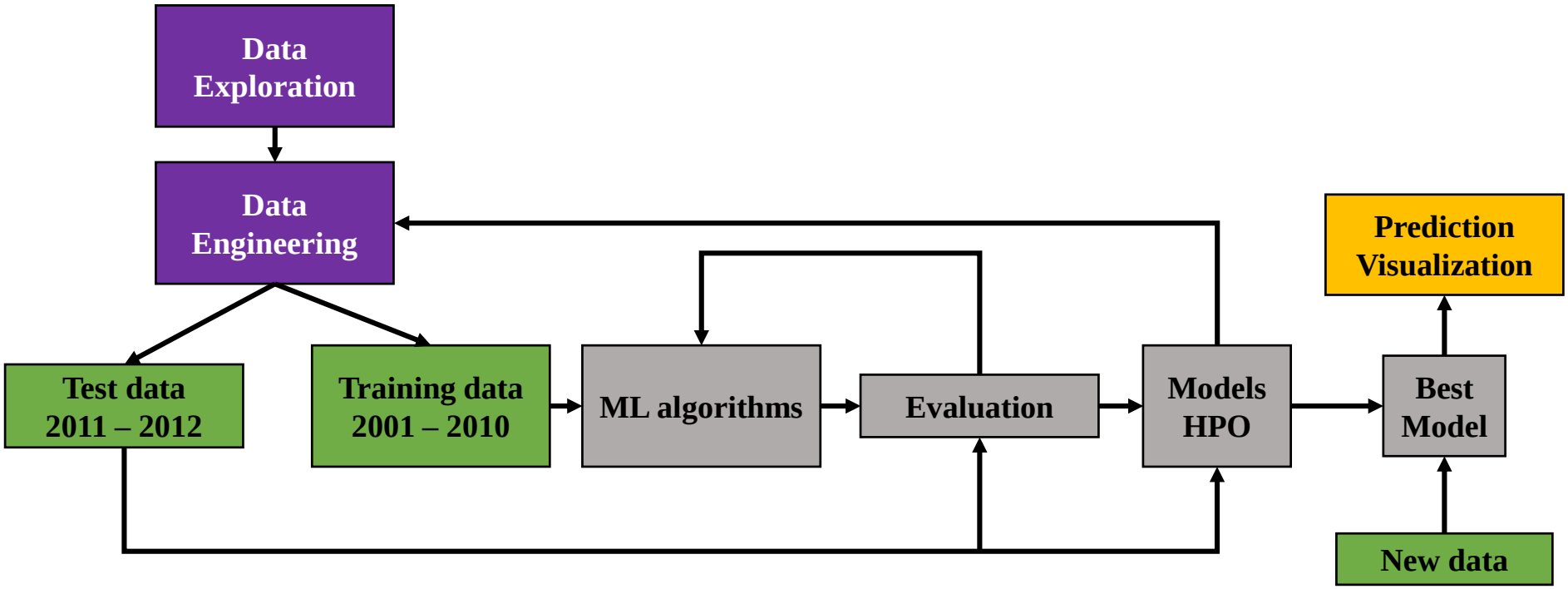
Extra Random Forest: each tree is trained using a randomly chosen set of predictors and top-down splitting

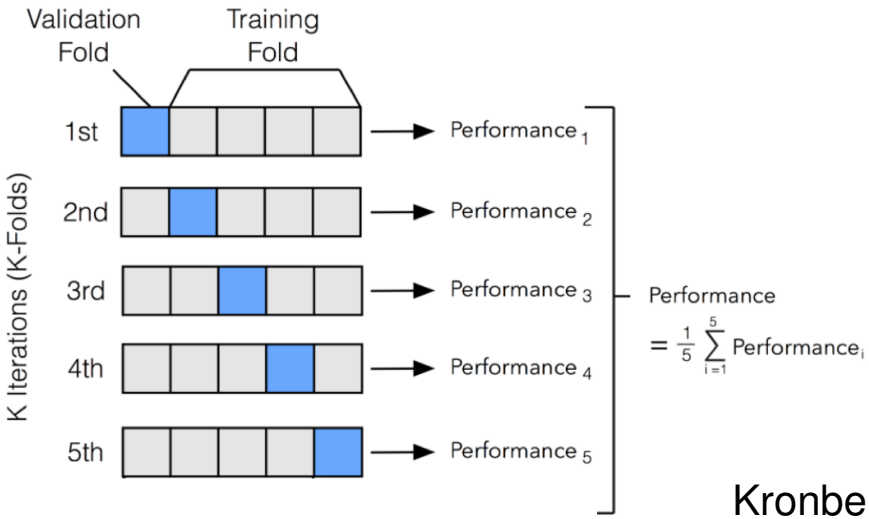


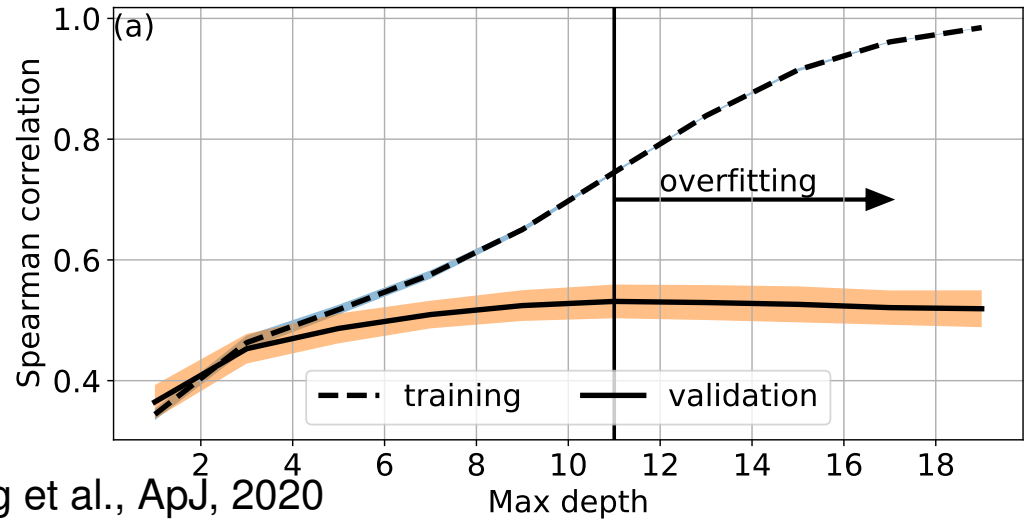
Machine Learning pipeline



Machine Learning pipeline

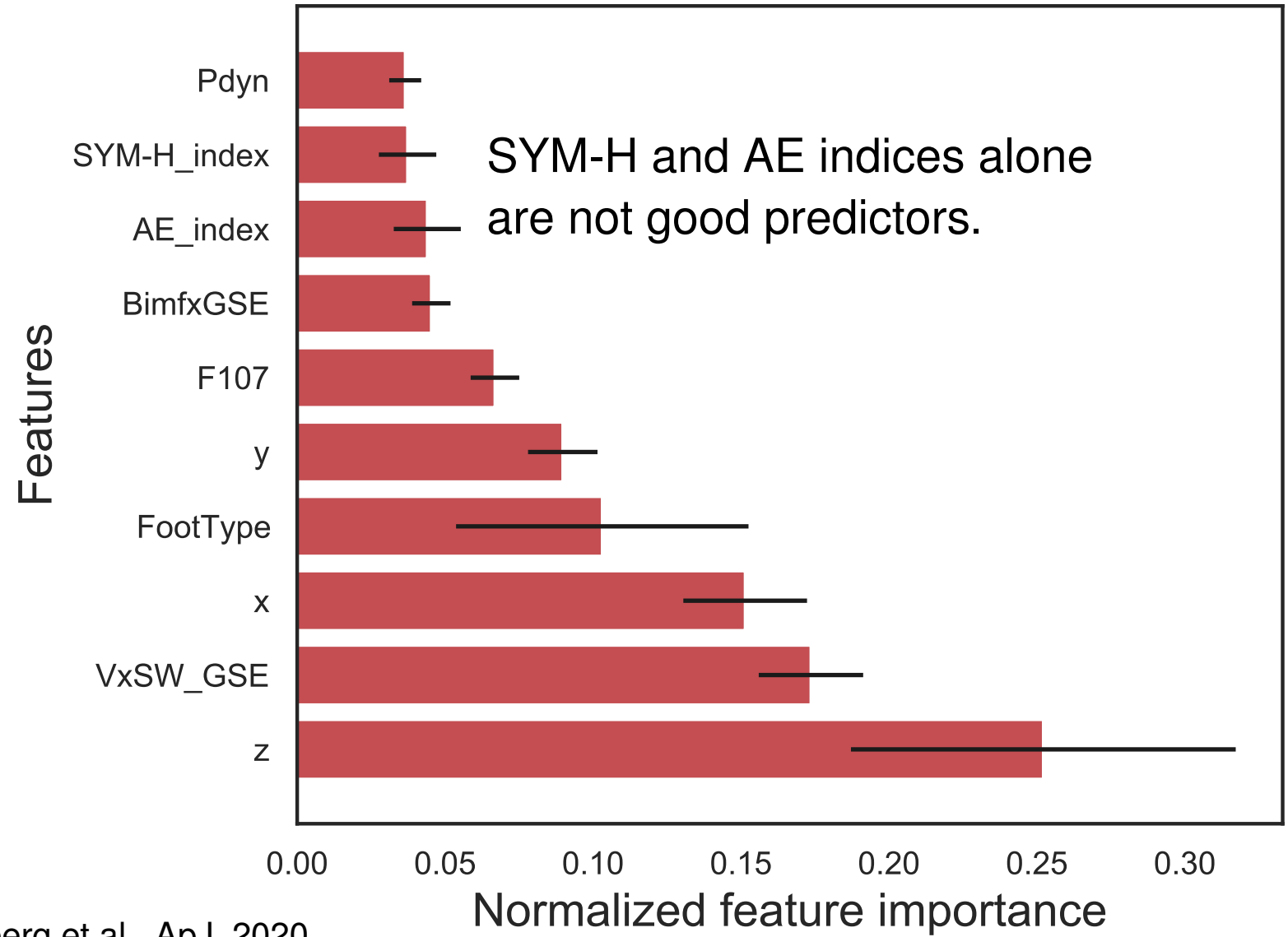






Kronberg et al., ApJ, 2020

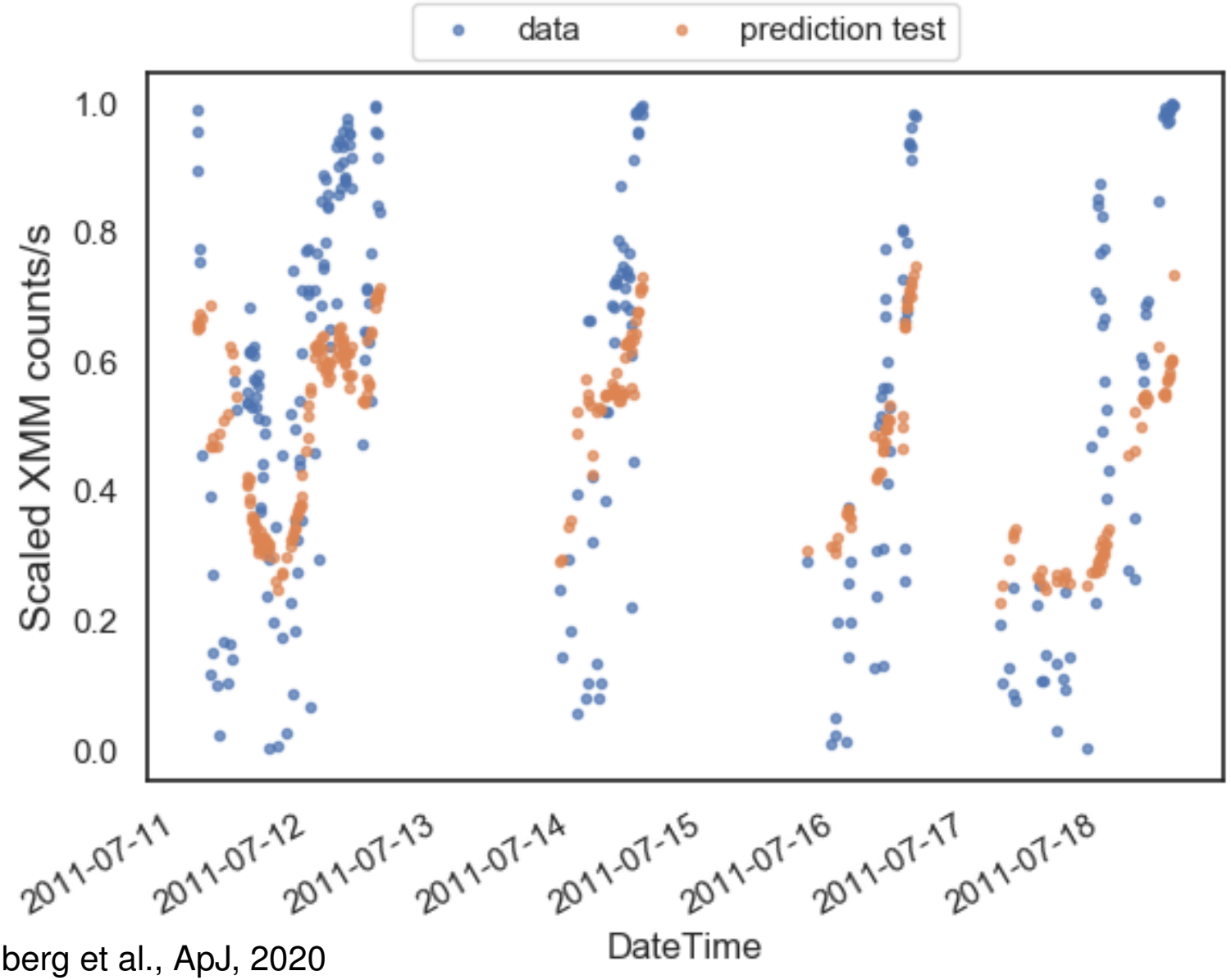
Results: the most important predictors are location and VxSW



Kronberg et al., ApJ, 2020

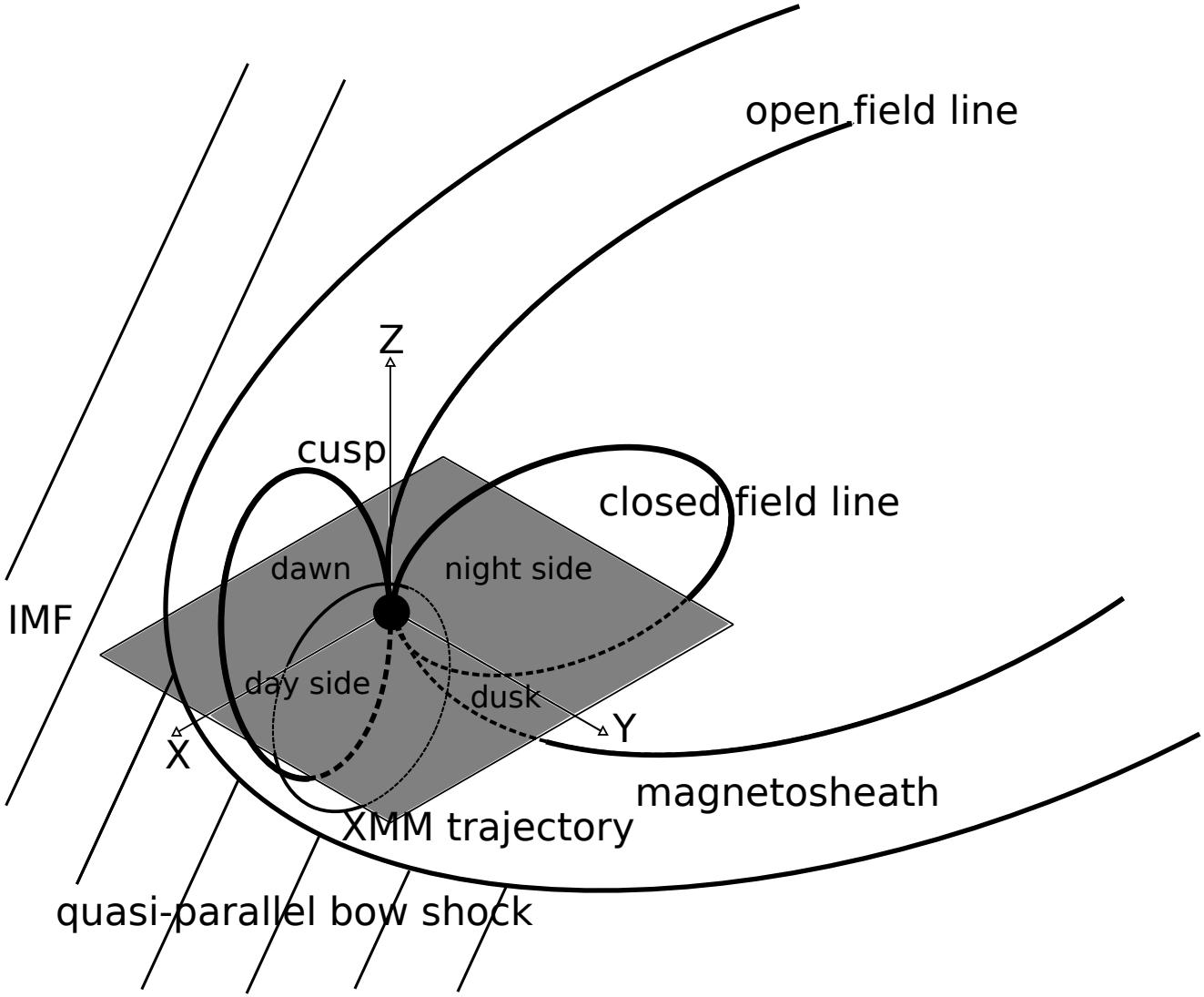
Results: the model is significantly better than the baseline

Spearman CC: Extra Random Forest $\simeq$0.48 vs. baseline $\simeq$0.35



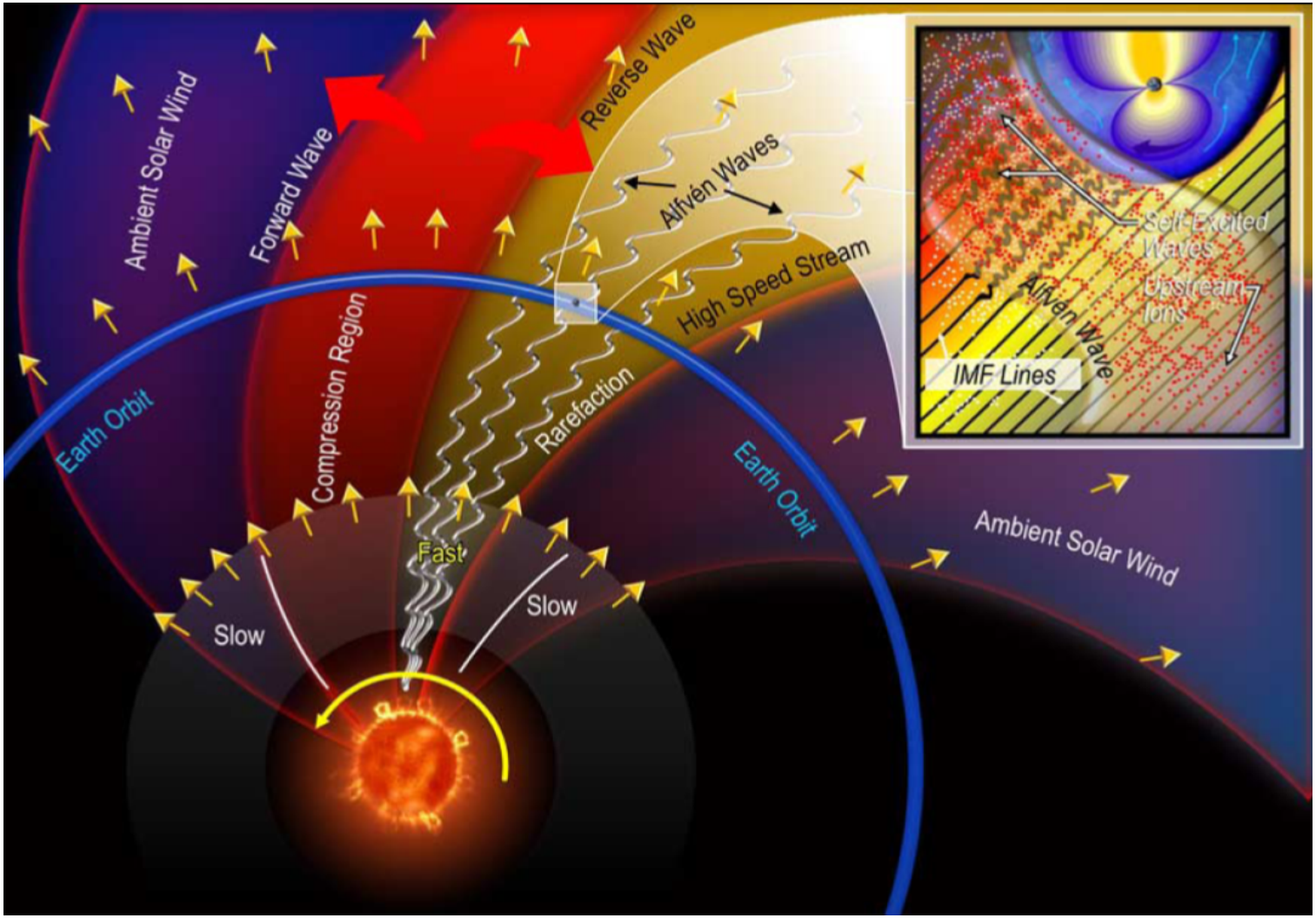
Kronberg et al., ApJ, 2020

Conclusions: future missions should avoid closed magnetic field lines, especially at the dusk



Kronberg et al., ApJ, 2020

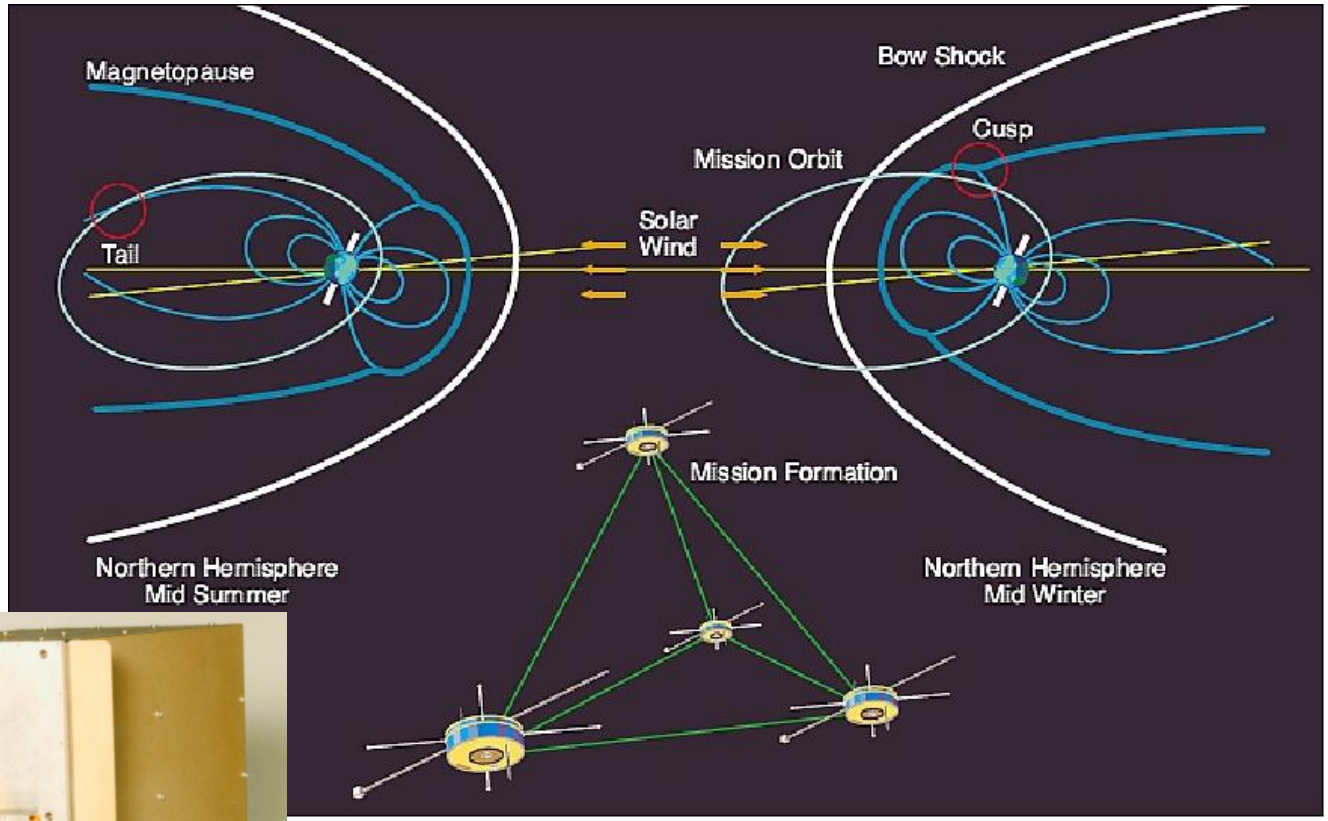
Conclusions: future missions should minimize observations associated with high solar wind speed



Desai et al., 2007

Outlook: we develop a model based on energetic proton observations by Cluster to predict the contamination

Contact: kronberg@geophysik.uni-muenchen.de



RAPID energetic particle detector