Random Forest Model of ULF wave power

S.N. Bentley, *J.R. Stout, T.E. Bloch, C.E.J. Watt, 2020, Earth and Space Science, doi.org/10.1029/2020EA001274*

Summary: Goals

- Create a predictive model of magnetospheric wave power to aid in space weather forecasting
- Investigate how several solar wind properties affect ULF (ultra-low frequency, 1-10mHz) wave occurrence

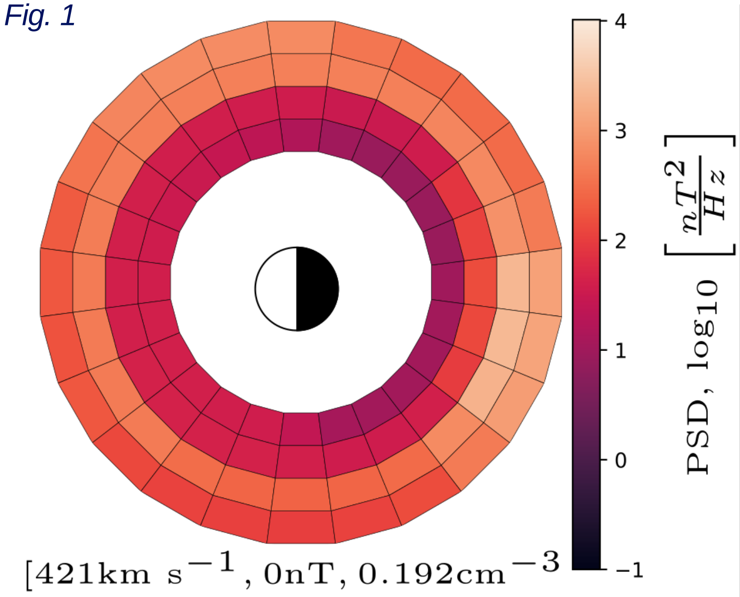
Summary: Conclusions

- The random forest model outperforms the previous "binned" empirical model
- Remaining uncertainty indicates that to improve predictions, we must include the conditions of near-Earth space and not just solar wind drivers and substorms (explosive plasma processes downstream)
- Extra power on observed on one side of the Earth is due to magnetospause perturbations, moderated by plasma density near the Earth

Summary: Model Outline

Predicts ultra-low frequency (ULF, 1-15mHz) wave power spectral density at ground magnetometer stations from solar wind observations.

Fig. 1



Model example: statistical map of wave power at 2.5 mHz, one station, for median solar wind driving.

Background: Earth’s Magnetosphere and Space Weather

EARTH’S MAGNETOSPHERE
Region of near-Earth space dominated by Earth’s magnetic field; most of the solar wind flows around this cavity as the magnetic fields must remain distinct

SUBSTORMS
Reconnection of magnetic fields, releasing lots of energy

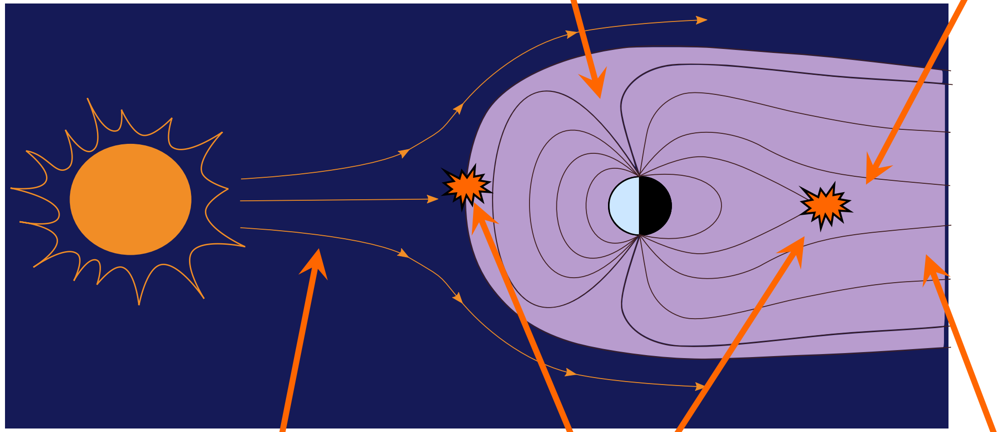


Fig. 2: The Sun-Earth System

SOLAR WIND
Constant flow of plasma from the Sun, varies in speed, density, composition, angle and magnetic field

CYCLE OF MAGNETIC FIELD MOTION
Under compressed conditions, magnetic fields can connect. This results in magnetic fields cycling around the Earth, from the nose to the tail and back again.

THE MAGNETOTAIL
Part of Earth’s magnetosphere stretches far downstream

Oscillations travel along the magnetic field and can be observed at the ground

Large perturbations in the drive waves which propagate throughout the magnetosphere

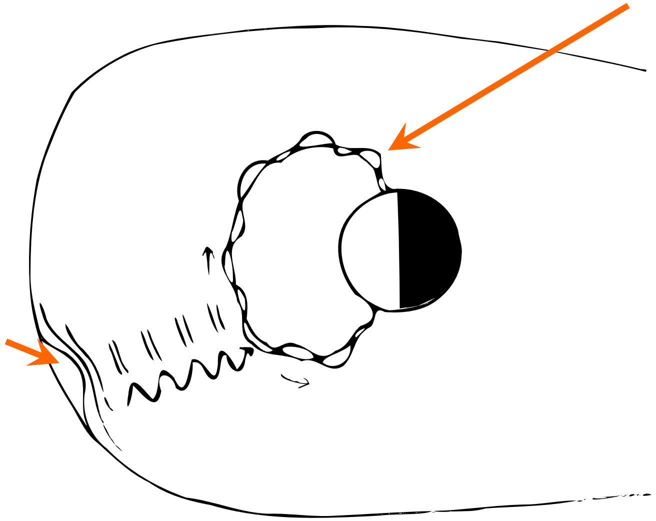


Fig. 3: Ultra-low frequency waves (1-10 mHz, periods of minutes to hours) driven by magnetopause perturbations

ULF waves: Magnetosphere-scale plasma waves which can energise radiation belt electrons and damage spacecraft
[Lam+, 2012]

Background: Datasets

1990-2005, solar wind and ground magnetic field

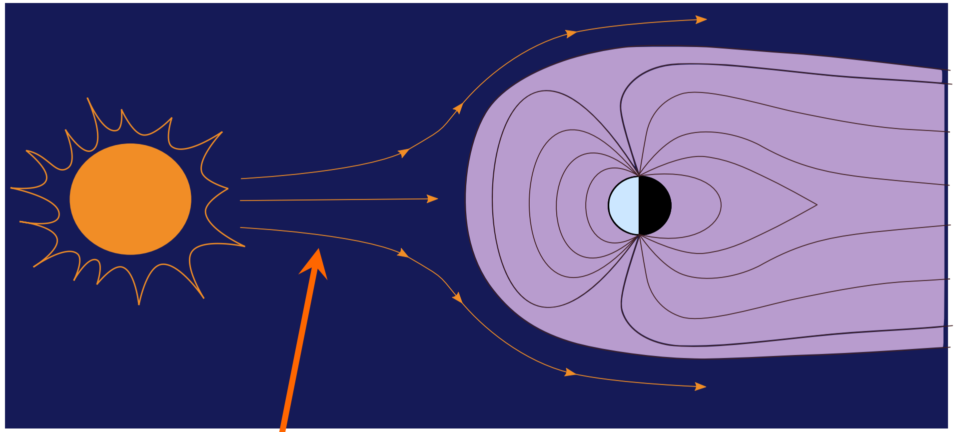


Fig. 2: The Sun-Earth System

NASA OMNIWeb data
Observations of solar wind properties, available ~45minutes before the solar wind reaches the magnetosphere

https://omniweb.gsfc.nasa.gov/
[King+, 2005]

CANOPUS/CARISMA ground magnetometer data
Wave power spectral density (PSD) from Canadian magnetometer network

https://www.carisma.ca/
[Mann+,2005]

Ground observations can be mapped to magnetospheric waves, and used in radiation belt modelling, [Ozeke+ 2009, 2014]

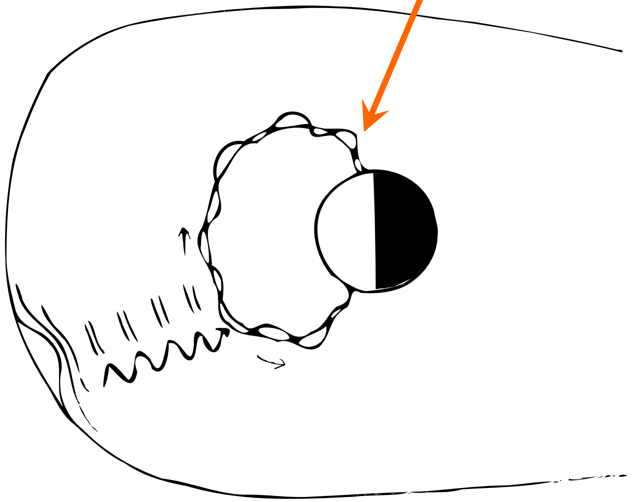


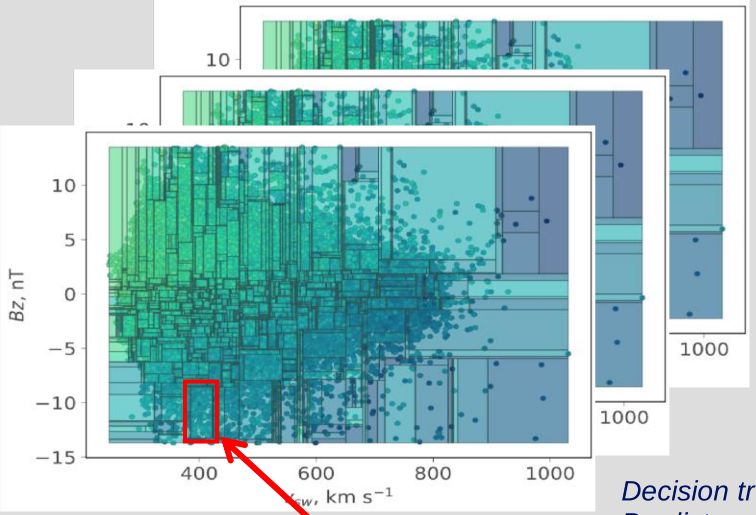
Fig. 3: Ultra-low frequency waves (1-10 mHz, periods of minutes to hours) driven by magnetopause perturbations

A Random Forest Model

Decision trees: iteratively partition parameter space to reduce variance in output.

Each tree could overfit → use a **random forest.** Final predicted value is averaged over an ensemble of 256 trees.

Fig. 4



Random forest: Average over many trees.

Decision tree: Predicts wave power as average of all data points in that bin

Solar wind: speed 400 km /s, magnetic field Bz -10 nT → Find the bin → Decision tree

Model settings:
- *Max depth 11*
- *Minimum samples per leaf 4*

Chosen using 5-fold cross validation.

Our Model: inputs and outputs

$$[\underline{MLT, vsw, Bz, var(Np)}, \underline{freq, latitude, component}] \to PSD$$

Inputs / features to train on:

- ***MLT***, magnetic local time (azimuthal angle around the Earth)
- ***v_{sw},*** solar wind speed
- ***Bz,*** solar wind north-south component of magnetic field (aligned with Earth's magnetic field)
- ***var(Np),*** variance in solar wind number density

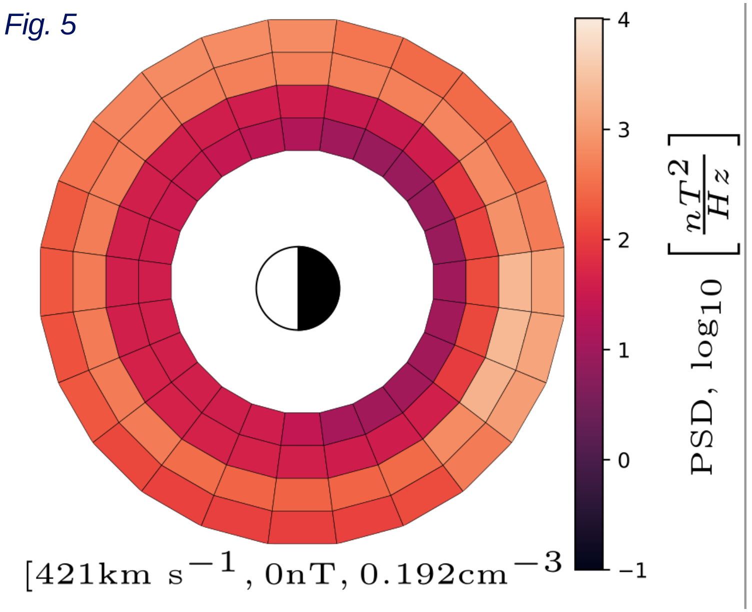
(chosen from previous work, [Bentley+ 2018])

Train one random forest for discrete values of:

- Frequency
- Latitude (CARISMA ground station, L~4.21 to 7.94)
- Horizontal component (magnetic N-S or E-W)

(for different physical reasons we keep these separate, to aid our investigation)

Fig. 5



Model example: statistical map of wave power at 2.5 mHz, one station, for median solar wind driving.

Our Model: Skill

- **Mean square error** on data subsets excluded from training: 0.13 to 0.68 log10(PSD) respectively

  (using 5-fold cross validation across all random forests)

- **Forecasting skill**: better performance than previous model or time-lagged wave power

Model tested	Skill
Random forest	81.2
Previous model	78.0
24 h lag	37.4
1 h lag	73.9

$$Skill = 100\left(1 - \frac{MSE_{model}}{MSE_{ref}}\right)$$

Positive skill scores indicate that the tested model is better than a random reference model. GILL station, 3.33 mHz.

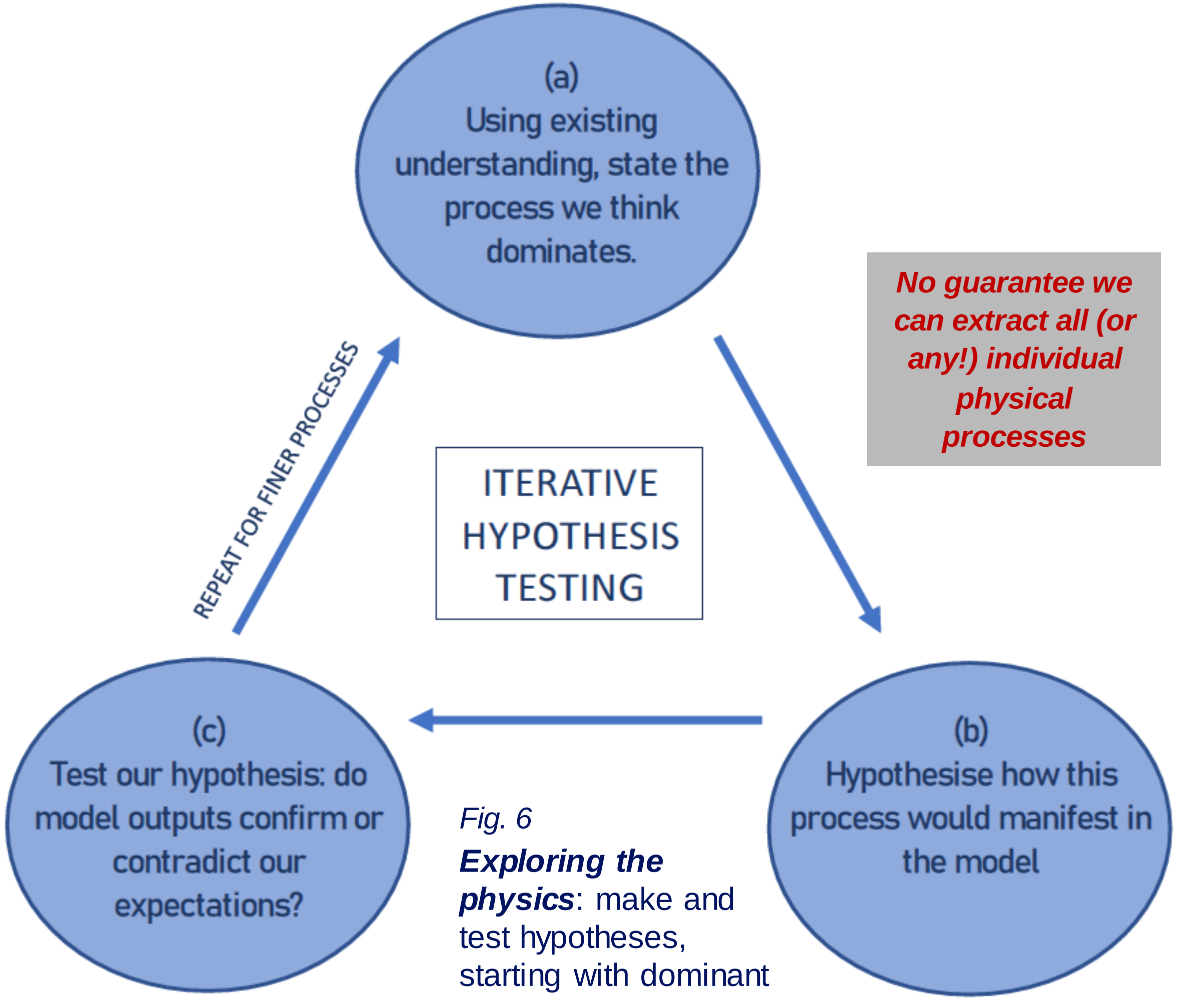
Random Forest Model of ULF wave power *Bentley et al., 2020, Earth and Space Science, doi:10.1029/2020EA001274*

Investigating wave drivers

- Both input features and the physical driving processes are highly interdependent
- Therefore driving processes don’t add linearly to final power
- Parameters were chosen because they are linked to ULF wave driving but interpretability is still not guaranteed

Some hypotheses to investigate:

1. *Asymmetries in wave power either side of the Earth (“dawn-dusk” asymmetry)*
2. *Role of wave drivers and conditions inside and outside the magnetosphere*
3. *Source of uncertainty*



(a)
Using existing understanding, state the process we think dominates.

(b)
Hypothesise how this process would manifest in the model

(c)
Test our hypothesis: do model outputs confirm or contradict our expectations?

ITERATIVE HYPOTHESIS TESTING

REPEAT FOR FINER PROCESSES

No guarantee we can extract all (or any!) individual physical processes

Fig. 6
Exploring the physics: make and test hypotheses, starting with dominant processes. Continue until no more can be distinguished.

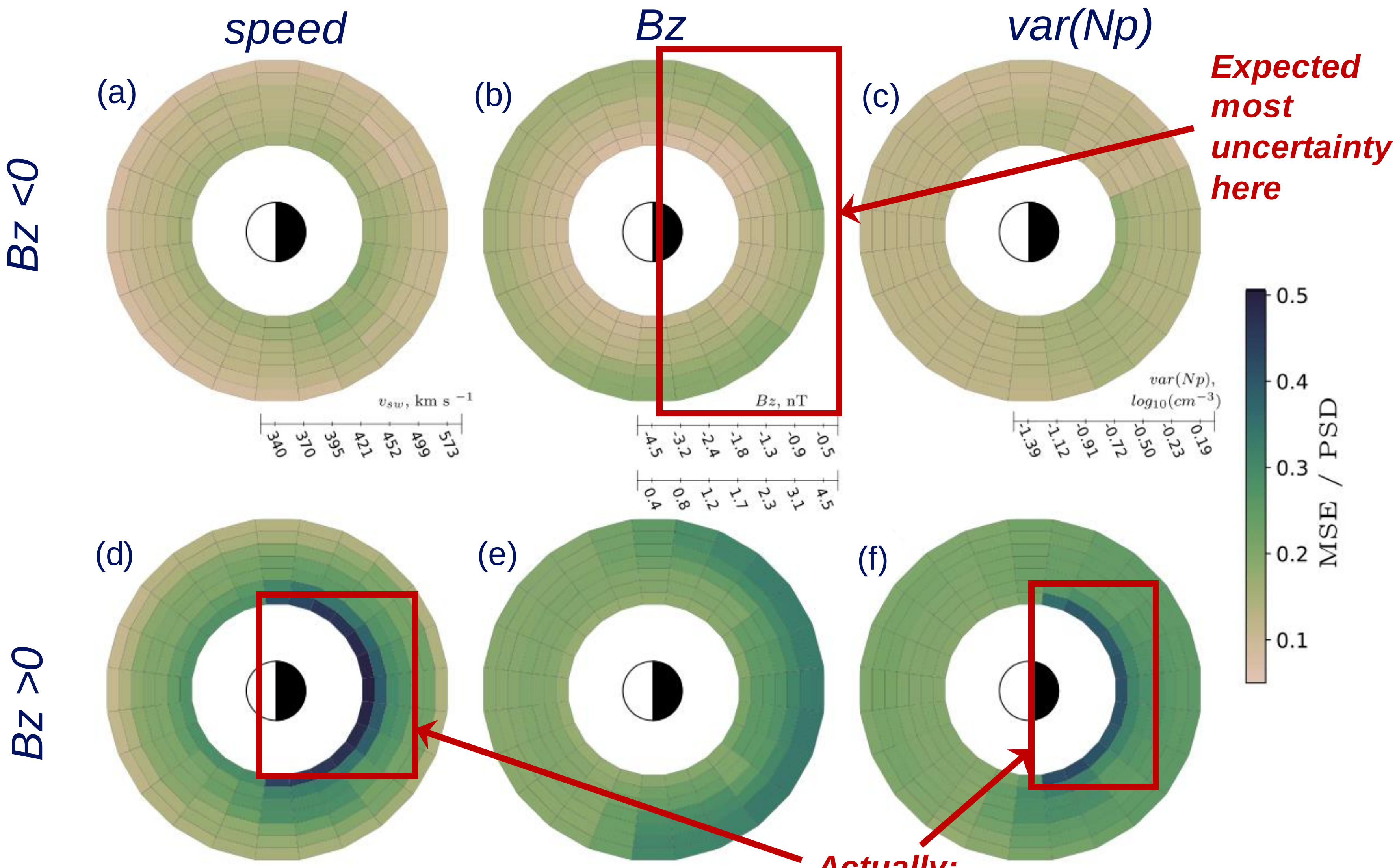
Bentley et al., 2020, Earth and Space Science, doi:10.1029/2020EA001274

Example Hypothesis:

- We expect most uncertainty where we do not represent driving processes well.
- We expect this to be substorms
  - ⟶ so expect most uncertainty for *Bz<0*.

However:

- Greatest remaining uncertainty for *Bz >0* rather than *Bz<0,*
- especially for low speed and *var(Np)*.



Expected most uncertainty here

Actually: most uncertainty where there is the least solar wind driving

Fig 7: Uncertainty remaining in model bins at given points in the parameter space.

Random Forest Model of ULF wave power

Bentley et al., 2020, Earth and Space Science, doi:10.1029/2020EA001274

Physics results summary

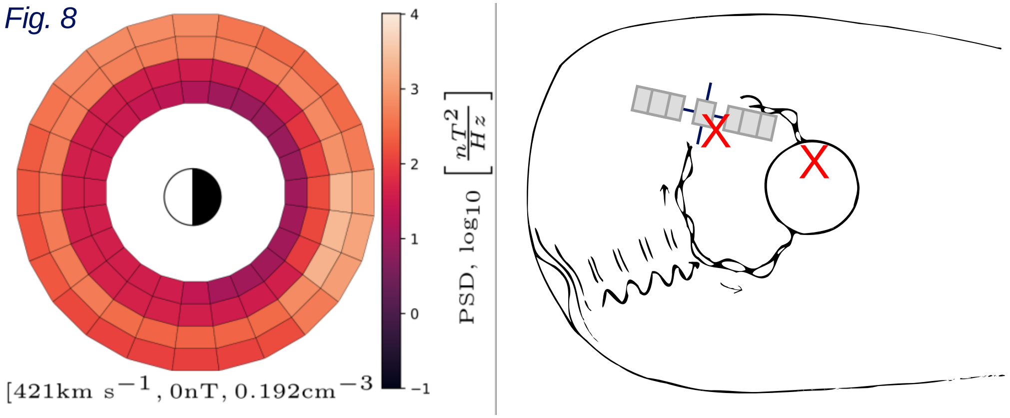
1. The dawn-dusk wave power asymmetry is due to internal moderation of magnetopause perturbations
    i. Speed-related wave power corresponds to magnetopause perturbations
    ii. *Bz*-related power does not; likely substorm related
    iii. *var(Np)*-related power does not; no clear alternative using this model as we cannot extract compression via magnetopause location

2. Internal magnetospheric processes contribute significant remaining uncertainty and should be considered next

Future potential directions

- Improve model:
    - magnetospheric driving parameters, time history, more stations
- Extend to magnetospheric wave power
    - Instead of simply assuming ground-space correspondence
- Calculate radial diffusion coefficients using this “magnetospheric map” and include in radiation belt models

Fig. 8



Goal: *Construct statistical map of in-situ wave power to use in radiation belt radial diffusion modelling.*

References

Bentley, S. N., Stout, J., Bloch, T., & Watt, C. E. J. (2020). Random Forest Model of Ultralow-Frequency Magnetospheric Wave Power, Earth and Space Science, doi:10.1029/2020EA001274

Bentley, S. N., Watt, C. E. J., Owens, M. J., & Rae, I. J. (2018). ULF wave activity in the magnetosphere: Resolving solar wind interdependencies to identify driving mechanisms. Journal of Geophysical Research: Space Physics, https://doi.org/10.1002/2017JA024740

Bentley, S. N., Watt, C. E. J., Rae, I. J., Owens, M. J., Murphy, K., Lockwood, M., & Sandhu, J. K. (2019). Capturing uncertainty in magnetospheric ultralow frequency wave models. Space Weather, https://doi.org/10.1029/2018SW002102

King, J. H., & Papitashvili, N. E. (2005). Solar wind spatial scales in and comparisons of hourly Wind and ACE plasma and magnetic field data. Journal of Geophysical Research, https://doi.org/10.1029/2004JA010649

Lam, H.-L., Boteler, D. H., Burlton, B., and Evans, J. (2012), Anik-E1 and E2 satellite failures of January 1994 revisited, Space Weather, doi:10.1029/2012SW000811.

Mann, I. R., Milling, D. K., Rae, I. J., Ozeke, L. G., Kale, A., Kale, Z. C., et al. (2008). The Upgraded CARISMA Magnetometer Array in the THEMIS Era. Space Science Reviews, https://dx.doi.org/10.1007/s11214-008-9457-6

Ozeke, L. G., Mann, I. R., Murphy, K. R., Jonathan Rae, I., & Milling, D. K. (2014). Analytic expressions for ULF wave radiation belt radial diffusion coefficients. Journal of Geophysical Research: Space Physics, https://doi.org/10.1002/2013JA019204

Ozeke, L. G., Mann, I. R., & Rae, I. J. (2009). Mapping guided Alfvén wave magnetic field amplitudes observed on the ground to equatorial electric field amplitudes in space. Journal of Geophysical Research,. https://doi.org/10.1029/2008JA013041