*31st EPS Conference on Plasma Phys. London, 28 June - 2 July 2004 ECA Vol.**28G**, P-1.055 (2004)*

Applications of information theory in the near-Earth plasma environment

T. K. March,[1] S. C. Chapman[1] and R. O. Dendy[2,1]

[1]*Department of Physics, University of Warwick, Coventry CV4 7AL, UK*
[2]*UKAEA Culham Division, Culham Science Centre, Abingdon, Oxfordshire OX14 3DB, UK*

1. Introduction

We consider the problem of estimating solar wind conditions at Earth given only single-point measurements from a spacecraft, approximately 200 Earth radii closer to the Sun in halo orbit around the point L1. The solar wind velocity is variable in magnitude (400 to 900 kms^{-1}) and in direction, so that the exact time for the solar wind to reach Earth from L1 can only be estimated; its average value is approximately 45 minutes. Here we take single-point solar wind measurements from the WIND spacecraft [1, 2], and use the AE geomagnetic indices [3] as a proxy for solar wind arrival at Earth. Figure 1 displays examples of these timeseries, and shows that a delay is necessary to obtain optimal correlation. In this work we investigate which method of time delay estimation is optimal. In doing so, we also quantify the correlation between the solar wind and the geomagnetic effects, in terms of their mutual information content.

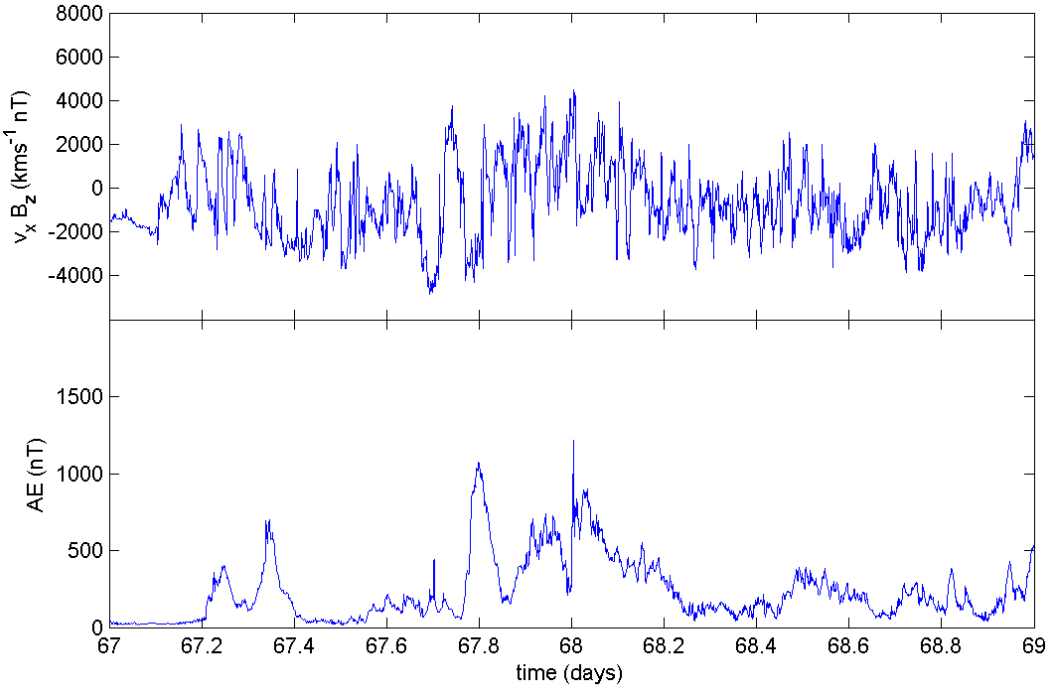

Figure 1: Timeseries of (upper) solar wind v_xB_z from WIND and (lower) AE geomagnetic index, for days 67-69 of 1995.

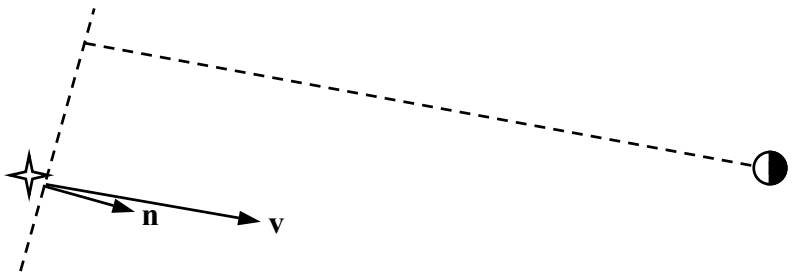

Figure 2: Geometry of planar propagation from the spacecraft (left) to the Earth. Plane normal **n** and solar wind velocity **v** shown.

2. Method

Typically, one assumes that a single-point measurement from the spacecraft specifies the solar wind conditions over a plane with normal **n**. Let **v** denote the solar wind velocity, and $\mathbf{P}_{sc}$ and $\mathbf{P}_E$ the spacecraft and Earth positions, see Fig. 2. The time for the solar wind to reach Earth is then given by [4]

$$\Delta t = \frac{(\mathbf{P}_{sc} - \mathbf{P}_E).\mathbf{n}}{\mathbf{v}.\mathbf{n}} \tag{1}$$

In this work we use the GSE (Geocentric Solar Ecliptic) coordinate system: *x*-axis from the Earth to the Sun, *y*-axis in ecliptic plane in dusk direction. We use the time shift defined by Eq.(1) to generate projected solar wind timeseries at Earth from the WIND data at L1, using the following eight hypotheses:

1. control set: raw data alone, with $\Delta t = 0$
2. **n** is oriented along the *x*-axis (Earth-Sun line)
3. **n** is parallel to the solar wind velocity **v**
4. **n** is perpendicular to **B** and lies in the *x*-*y* plane (ecliptic)
5. **n** is perpendicular to **B** and to a line joining the spacecraft to Earth
6. **n** is parallel to the vector (B_yB_z, B_xB_z, B_xB_y), see [5]
7. **n** is at 45º to the *x*-axis (the Parker spiral angle), in the ecliptic plane
8. **n** is given by minimum variance analysis [4] of **B**

We evaluate the success of different methods for estimating **n** by calculating the mutual information between WIND satellite data and ground-based AE index measurements. For two timeseries **a** and **b** the mutual information between them is defined by [6]

$$I = -\sum_{i=1}^{m} p_i^a \log_2 p_i^a - \sum_{j=1}^{m} p_j^b \log_2 p_j^b + \sum_{i=1}^{m}\sum_{j=1}^{m} p_{i,j}^{a,b} \log_2 p_{i,j}^{a,b} \tag{2}$$

where the timeseries are discretised into *m* bins, p_i^a is the measured probability of observing timeseries **a** in bin *i*, and $p_{i,j}^{a,b}$ is the joint probability of observing timeseries **a** in bin *i* and timeseries **b** in bin *j*. Here logarithms are taken to base two so that the mutual information is measured in bits.

Method	Maximum mutual information (bits)	Additional time lag to maximize (mins)
(7) Parker Spiral	0.2727	45.8
(8) Minimum variance analysis	0.2708	49.4
(4) **n** ⊥ **B** and in ecliptic plane	0.2705	46.4
(5) **n** ⊥ **B** and spacecraft position	0.2698	48.8
(2) **n** parallel to *x*-axis	0.2696	47.6
(3) **n** parallel to wind velocity	0.2694	49.4
(6) **n** = (B_yB_z, B_xB_z, B_xB_y)	0.2674	50.0
(1) Raw data	0.2585	96.8

Figure 3: Mutual information between the solar wind v_xB_z and geomagnetic AE timeseries for the eight different time delay estimation methods. The second column gives the maximum mutual information obtained for each method while the third column gives the additional time-lag necessary to obtain this maximum.

3. Results

Figure 3 shows the results of the comparison using a four bin mutual information calculation (the results are qualitatively unchanged for 2-64 bins). The mutual information is evaluated with respect to an additional time lag added to the projected timeseries. This allows for any finite response time of the magnetospheric system. We consider the maximum mutual information obtained for each method and conclude that the Parker spiral angle method (7) in this case yields the best results, giving over a quarter of a bit of mutual information. However, this is only slightly better than methods (2-6) whose results are almost indistinguishable. Method (8), where **n** is given by minimum variance analysis [4] of **B**, gives the second highest mutual information. This technique is still under development by the authors of [4]. An additional fixed time lag of around 45-50 minutes maximises the mutual information found using methods (2-8). This is presumably an internal propagation time for information through the upper regions of the magnetosphere to the ground. A consistency check is provided by the difference between the maximizing time lag for the raw data (96.8 mins), and the mean time lag (47.5 mins) for the data projected using the models. Using the average *x*-distance between WIND and Earth, this implies a mean transmission speed for information in the solar wind of about 350kms^{-1}, which is a typical slow solar wind

speed. Several of the time delay estimations considered here give similar results. This is perhaps due to an intrinsic limitation imposed by the geometry of the problem. Since the WIND spacecraft lies close to the Earth-Sun line, and the solar wind velocity is to zeroth-order radially outward, one expects the greatest contribution to the time delay to come from the *x*-distance. Improving on the *x*-axis delay method (2) is thus difficult.

4. Conclusions

We have shown that it is possible, using mutual information, to quantitatively distinguish hypotheses concerning time delay estimation. This measure quantifies the degree to which the far upstream solar wind can be used to predict AE, and can also be used to estimate the response time of the magnetospheric system to solar wind input. Irrespective of the specific model used to link the solar wind and AE time series, two robust quantitative results emerge. First, their mutual information is 0.26 ± 0.01 bits when using a four bin partition. Second, there is an additional time lag of 47.5 ± 2.5 minutes reflecting transmission and propagation within the magnetosphere to ground.

Acknowledgements

This work was supported in part by the United Kingdom Engineering and Physical Sciences, and Particle Physics and Astronomy Research Councils. Geomagnetic AE index data was from the Kyoto World Data Centre [7], WIND data was from the NASA Coordinated Data Analysis Web [8].

References

[1] K. W. Ogilvie et. al., *Space Science Reviews*, **71**, 55-77, 1995
[2] R. P. Lepping et. al., *Space Sicence Reviews*, **71**, 207-229, 1995
[3] Davis, T. N., and M. Sugiura, *Journal of Geophysical Research*, **71**, 785-, 1966
[4] D. R. Weimer et. al., *Journal of Geophysical Research*, **108**(A1), 2003
[5] A. J. Ridley, *Journal of Atmospheric and Solar-Terrestrial Physics*, **62**:757-771, 2000
[6] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, 1949
[7] *http://swdcwww.kugi.kyoto-u.ac.jp*
[8] *http://cdaweb.gsfc.nasa.gov*