

Forecasting space weather: Predicting interplanetary shocks using neural networks

Jon Vandegriff *, Kiri Wagstaff, George Ho, Janice Plauger

The Johns Hopkins University, Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, MD 20723, USA

Received 1 May 2003; received in revised form 25 August 2003; accepted 2 September 2004

Abstract

We are developing a system to predict the arrival of interplanetary (IP) shocks at the Earth. These events are routinely detected by the Electron, Proton, and Alpha Monitor (EPAM) instrument aboard NASA's ACE spacecraft, which is positioned at Lagrange Point 1 (L1). In this work, we use historical EPAM data to train an IP shock forecasting algorithm. Our approach centers on the observation that these shocks are often preceded by identifiable signatures in the energetic particle intensity data. Using EPAM data, we trained an artificial neural network to predict the time remaining until the shock arrival. After training this algorithm on 37 events, it was able to forecast the arrival time for 19 previously unseen events. The average uncertainty in the prediction 24 h in advance was 8.9 h, while the uncertainty improved to 4.6 h when the event was 12 h away. This system is accessible online, where it provides predictions of shock arrival times using real-time EPAM data.

© 2005 Published by Elsevier Ltd on behalf of COSPAR.

Keywords: Space weather; Predicting interplanetary shocks; Neural networks

1. Introduction

Interplanetary (IP) shocks are often launched by coronal mass ejections (CMEs). The shocked region of the solar wind can be a source of locally accelerated particles whose energies typically reach into the MeV range (Kahler et al., 1984), but can extend into the GeV range for strong shocks. Because the shock is created near the Sun, the initial particle signature as seen by ACE at L1 shows strong velocity dispersion (i.e., the higher energy particles reach the Electron, Proton, and Alpha Monitor (EPAM) detector sooner than low-energy particles originating from the same event). Forewarning of this initial particle burst from a newly launched CME would require more detailed knowledge and monitoring of the Sun's active regions than is currently available.

Intensity evolution following the velocity-dispersed onset typically shows increasing numbers of particles, especially for shocks which stay magnetically connected to L1 as they propagate through the solar wind. As the shock gets closer to the spacecraft, particle intensities rise dramatically in a non-dispersive pattern (i.e., all energies of particle rise simultaneously) since the source of acceleration is now close to the detector. We describe here an empirical technique which uses the evolution of the particle intensity after the initial, velocity-dispersed onset of the shock to predict the arrival time of the subsequent peak in intensity associated with the shock passage.

The enhanced particle intensities associated with IP shock passage are commonly referred to as energetic storm particle (ESP) events due to their close association with geomagnetic storms. During an ESP event, the number of incident ions with energies above 10 MeV can increase by several orders of magnitude, and some

* Corresponding author. Tel.: +1 240 228 8558; fax: +1 240 228 0727.
E-mail address: jon.vandegriff@jhuapl.edu (J. Vandegriff).

events show significant increases in fluence for ions above 30 MeV. Because of the radiation hazards to astronauts and space-embedded technology posed by energetic particles, there is considerable interest in forecasting large IP shock-driven particle events. Furthermore, because large ESP events are infrequent, they are difficult to study. However, the same acceleration mechanism operates on high energy (above 30 MeV) and low energy (below 10 MeV) particles, and thus we gain significant insight into ESP characteristics and evolution by studying the lower energy, lower intensity events, which are common.

2. ACE/EPAM data description

NASA's Advanced Composition Explorer (ACE) spacecraft is stationed in a halo orbit around Lagrange point L1, which is about 1.5 million km from the Earth (235 Earth radii, R_E), or about 1% of the distance from the Earth to the Sun. The EPAM instrument on ACE measures low energy charged particles at L1 (Gold et al., 1998). Data from EPAM, as well as solar wind data from other instruments on ACE, are telemetered back and published on the web by NOAA in real-time (Zwickl et al., 1998), providing a continuous upstream monitor of the particle activity and solar wind behavior. Simple monitoring of real-time intensity values observed at L1 can be used to provide a short-term prediction of what conditions may soon reach the Earth. For example, given a shock traveling with the solar wind at 500 km/s, EPAM real-time data gives roughly a 45 min lead time on possible future particle behavior at the Earth. Our goal is to increase this lead time by using EPAM data ahead of the shock to predict the arrival time of the intensity burst associated with shock passage.

EPAM measurements include ion intensities for several energy channels. Almost all of the ions measured are protons, but other less abundant species (mostly He^{2+} , at the few percent level) are also present. Neither this ion contamination nor the occasional electron contamination are removed from our input data, since it is unlikely that this could be easily and reliably done in a real-time system.

Based on a list of about 160 shocks from August 1997 to December 2001 (Ho et al., 2003), we selected 56 events showing both velocity dispersion in the shock onset and a peak intensity greater than 10^5 particles/(s cm^2 ster keV) for the 47–65 keV proton channel. These selection criteria excluded 75 small shocks (i.e., peak intensity below 10^5) and 24 shocks which, although strong enough, did not exhibit the dispersive precursor. The start time for each event was established using a simple trigger designed to detect velocity dispersion; this trigger can also be used in

Table 1
Inputs to the neural network

Input name	Description
P1	Proton intensity, energy 47–65 keV
P3	Proton intensity, energy 112–187 keV
P5	Proton intensity, energy 310–580 keV
P6	Proton intensity, energy 760–1220 keV
P7	Proton intensity, energy 1060–1910 keV
ξ	Anisotropy coefficient
SS	Spectral slope
IMP	Intensity at midpoint
SS'	Spectral slope derivative
IMP'	Intensity at midpoint derivative

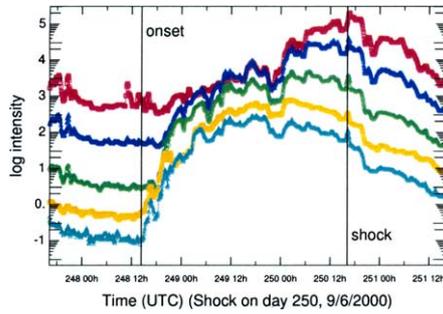
real-time. The trigger examines the spectral slope, the average height of the energy spectrum, and the time derivatives of these quantities. It reliably detect onsets with a very low false alarm rate. We divided these shocks into *training* and *test* data sets. Each sequence (from onset to shock arrival) included an average of 530 data points (spanning 44.2 h).

As inputs to our empirical model, we use the five ion channels provided by the NOAA real-time system, which are listed in Table 1. These intensities are 5-min averages of proton intensity in non-contiguous energy bands ranging from 47 keV to 1.9 MeV. We also use an anisotropy coefficient, ξ , which characterizes angular intensity variations that indicate the direction of shock propagation (Zwickl and Webber, 1976). The remaining four inputs listed in Table 1 are derived quantities based on the energy spectrum of the five proton channels. The spectral slope (SS) is the steepness of a line (in log intensity – log energy space) fit to the intensity points. The intensity midpoint (IMP) is the intensity level at the middle of the spectrum. The time derivatives of these quantities (SS' and IMP') are also included. All inputs are smoothed based on preceding values in the data stream.

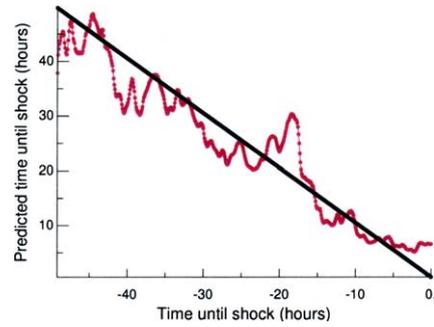
3. IP shock prediction method

In many cases, precursor signatures exist in the particle data which give clues about the impending shock arrival. The data set shown in Fig. 1(a) provides a clear example of this velocity-dispersed precursor, which appears as a contraction or squeezing of the proton intensity values beginning at hour 14 of day 248. However, as the shock gets closer to the Earth, the particle intensity values for various energy channels exhibit more complex behavior. Therefore, instead of attempting to heuristically model the shock evolution, we have devised an empirical approach that predicts the arrival time of the shock.

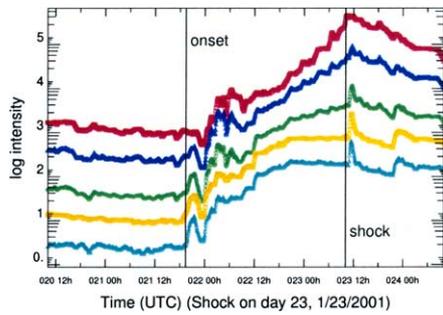
The intensity time series data plus other derived quantities listed in Table 1 serve as input data to an



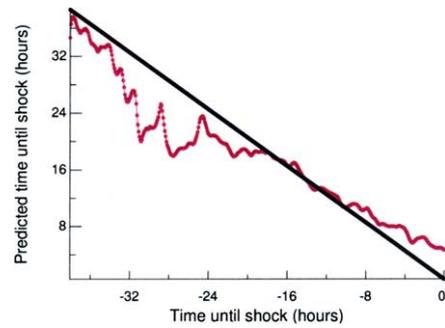
(a) Particle intensities for test sequence 9.



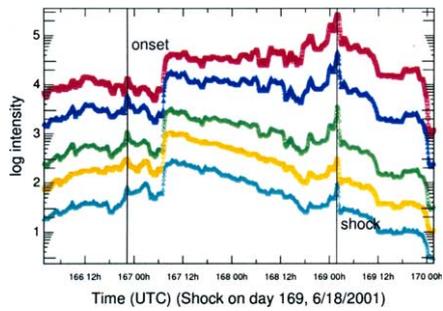
(b) Predicted time until shock arrival.



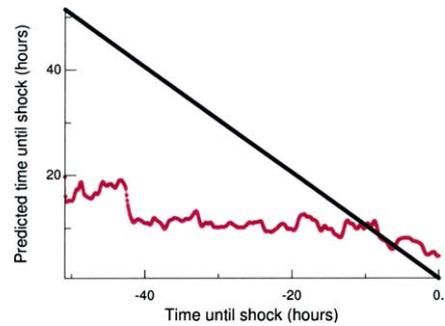
(c) Particle intensities for test sequence 12.



(d) Predicted time until shock arrival.



(e) Particle intensities for test sequence 15.



(f) Predicted time until shock arrival.

Fig. 1. Each pair of panels shows an intensity evolution profile on the left and a forecast evaluation on the right. The intensity plots show the four proton channels used as inputs; the lowest-energy channel is the topmost line. The start and stop times for the data fed into the forecasting algorithm are indicated with vertical lines labeled according to the onset and shock arrival times for the event. In the forecast evaluations, the neural network is attempting to provide a countdown to the shock arrival, similar to the straight line shown as the ideal forecast.

artificial neural network which is then trained to predict the time remaining until the shock arrives. We use a relatively new form of recurrent neural net (Lo, 1994, 1999), which contains time-delayed links connecting all the nodes in any hidden layer with all the other nodes in that same hidden layer. Such time-delayed connections: (1) represent an explicit treatment of the time domain of the problem, (2) give the network a “memory” with a decay time proportional to the time step, and (3) are crucial in allowing the networks to be able to capture the behavior of arbitrary dynamical systems (Lo, 1998). We use a fixed step size and momentum constant, both of which are manually tuned.

The network we used consisted of 10 input nodes (one for each of the inputs in Table 1), one output node (to predict the remaining time), and two layers of four hidden nodes each. Input values to the network were scaled and shifted to ensure that all values fall within the sensitivity range of the hyperbolic tangent used as an activation function. Logarithmic transformations were used on intensity values and linear transformations on all other values. There is an inherent assumption, when using this network, that there is in fact an impending event in the data stream. If not, the predictions will be unreliable (as the network was only trained on true events), usually outputting a single very high value.

4. Results and analysis

At each time point in the time-series input data for a given shock, the network provides its guess for the time remaining until the shock arrival. During training, differences between this guess and the actual time to arrival are used to adjust the internal weights of the network. During testing, the weights are fixed, and differences between the network estimate and the actual remaining time are used to assess the performance of the network on input sequences it has not seen before.

Training was performed using 37 of the 56 selected IP shocks. After presenting each training sequence to the network 100 times, the performance of the network was assessed by having it forecast arrival time countdowns for the 19 shocks not used during training.

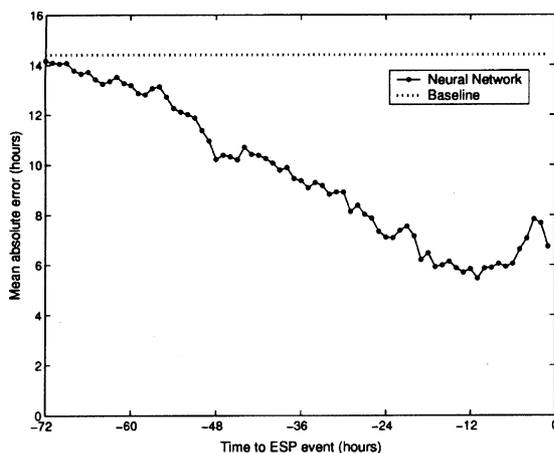
Fig. 1 shows input sequences and forecasts for 3 of the 19 shocks used in testing. The left panels (a, c, and e) show the time evolution of intensity values for each particle channel. The topmost trace in each intensity plot corresponds to the lowest energy channel P1 and each successively lower intensity trace corresponds to the next higher energy channel, from P3 to P7. Labeled vertical lines indicate the onset and shock times. Only the data between these times were used as input data, although it is helpful to see data outside this region to visually understand the event. Although also used as inputs, the time evolution of the anisotropy coefficient and derived spectral properties are not shown.

The panels on the right (b, d, and f) show a comparison of the forecasted and actual arrival times. Since the true output is always a countdown until the shock arrival, the ideal output is always a straight line with a slope of -1 . This plot provides a simple way of determining

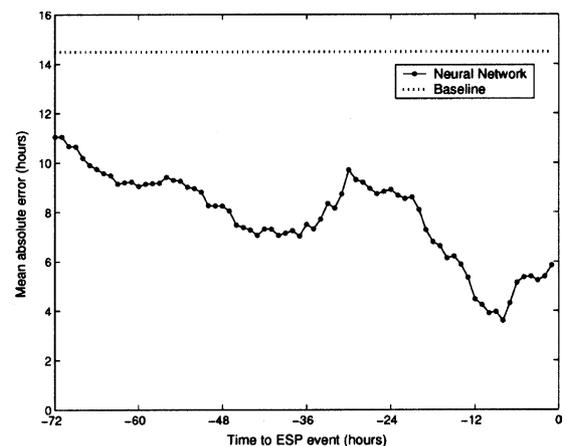
how the forecast arrival time compares with the true countdown.

Test sequence 9 (Fig. 1(a) and (b)) exhibits very good performance. The neural network briefly under-estimates the shock arrival time by about 10 h when it is still 44 h away, and it over-estimates the arrival time by about 10 h when it is 18 h away, but overall its predictions align well with the true arrival time. Five hours before the shock arrival, the predicted countdown flattens out, predicting longer arrival times than those in the actual countdown. This is most likely caused by the fact that there is a smaller peak just before the main peak, and the decline in flux following the smaller peak causes the forecasting algorithm to delay its countdown. In test sequence 12 (Fig. 1(c) and (d)), predictions made more than 24 h in advance of the shock under-estimate the arrival time by up to 8 h. From 20 to 10 h before the event, the prediction is very accurate, although in the final 10 h before the shock, the forecast arrival time exceeds actual shock arrival time. Test sequence 15 (Fig. 1(c) and (d)) indicates a situation where the algorithm essentially failed to detect the oncoming burst. Probably this is due to the fact that, except for a non-dispersive rise at hour 6 of day 167, the intensity traces are relatively flat for most of this event. In most events, it is the rising intensity values that drive the countdown. Note that this event has only a weakly dispersive signature at the onset.

Results over the entire training and test sets are summarized in Fig. 2. At time t , we calculate two values: P_t is the predicted time until arrival, and A_t is the actual time until arrival. The absolute error in the prediction time is calculated as $|P_t - A_t|$. Fig. 2(a) shows the absolute errors averaged over all training datasets as a function of countdown time. Because each event used to construct the average extends back a different amount of time before the



(a) Mean absolute error on 37 training data sets.



(b) Mean absolute error on 19 test data sets.

Fig. 2. The bottom axis is time until shock arrival and the left axis is mean absolute error in the forecasted arrival time. Mean absolute error is obtained at each time point by averaging the absolute error made for each sequence that had a forecast at that time. The horizontal “Baseline” error corresponds to the error made by a simplistic, fixed event length forecasting model, which is based on the average event length in the training data.

shock, the absolute error values are averages of a variable number of points. For example, at 60 h before the shock, there are only a few datasets contributing to the mean absolute error at this time. However, at 12 h before the shock, all of the datasets have contributing arrival estimates. Fig. 2(b) plots the mean absolute error for the 19 testing data sets. Both training and testing datasets indicate forecasting errors which are decreasing until about 12 h before the shock arrival, after which they rise slightly. This is likely due to the fact that particle behavior in the few hours before the shock arrives is very different from longer-term trends. A single network cannot adequately predict both kinds of behavior. We are investigating the use of an additional network, specifically trained for short-term predictions.

Both (a) and (b) in Fig. 2 also show the error for a baseline prediction scheme that uses the average event length in the training events to estimate the arrival time. The average time from onset to shock arrival for the 37 training events is 1.84 days (or 44.2 h). As a forecast, this fixed event duration gives a constant prediction error for each event. An average of the absolute error using this approach is about 14.5 for both training and test data sets. The network consistently beats this simple forecast.

Other models exist for computing arrival times of IP shocks, including the shock time of arrival (STOA) model (Dryer and Smart, 1984) and the Interplanetary Shock Propagation Model (ISPM) (Smith and Dryer, 1990). These models rely on radio burst information (start time, location, drift speed, and burst duration) as well as H-alpha, GOES, and X-ray data, but no particle data. A recent study (Kadinsky-Cade et al., 1998) of shocks in the period 1991–1997 using these models along with IMP-8 and WIND data (for evaluation) found that the average deviations between the predicted shock arrival time and the nearest actual arrival time were ± 39 h for the ISPM model and ± 36 h for the STOA model. The neural network described here, provides an improved indicator of shock events mainly through its direct use of particle data.

5. Conclusions

We have demonstrated that an artificial neural network can be trained to predict the shock arrival with better accuracy than existing methods. We have several

ideas for expanding this work. First, we plan to include other sources of information about the interplanetary medium, including the observed magnetic field and plasma characteristics. This information is continuously collected by other instruments on ACE. Second, we are also developing methods to provide an estimate for the error in our predictions. Finally, we are in the process of establishing this system for real-time access, at <http://sd-www.jhuapl.edu/ACE/EPAM/RISP/>. We view this work as a promising first step in providing predictive information about impending shocks. These advance warnings are useful as space weather monitors, especially to satellite and communications operators, as well as to the human space flight community.

References

- Dryer, M., Smart, D.F. Dynamical methods of coronal transients and interplanetary disturbances. *Adv. Space Res.* 4, 291–301, 1984.
- Gold, R.E., Krimigis, S.M., Hawkins, S.E., Haggerty, D.K., Lohr, D.A., Fiore, E., Armstrong, T.P., Holland, G., Lanzerotti, L.J. Electron, proton and alpha monitor on the advanced composition explorer spacecraft. *Space Sci. Rev.* 86, 541–562, 1998.
- Ho, G.C., Lario, D., Decker, R.B., Roelof, E.C., Desai, M.I., Smith, C.W. Energetic electrons associated with transient interplanetary shocks: evidence for weak interaction, in: *Proceedings of the 28th International Cosmic Ray Conference*, pp. 3689–3692, 2003.
- Kadinsky-Cade, K., Quigley, S., Ginet, G. Validation of interplanetary shock propagation models. *Trans. Am. Geophys. Un.* 79, F712, 1998.
- Kahler, S.W., Sheeley Jr., N.R., Howard, R.A., Koomen, M.J., Michels, D.J., McGuire, R.E., von Roseninge, T.T., Reams, D.V. Associations between coronal mass ejections and solar energetic proton events. *J. Geophys. Res.* 89, 9683–9693, 1984.
- Lo, J.T. Synthetic approach to optimal filtering. *IEEE Trans. Neural Networ.* 5 (5), 803–811, 1994.
- Lo, J. Universal neuroapproximation of dynamic systems for robust identification, in: *Proceedings of the 1998 International Joint Conference on Neural Networks*, Anchorage, Alaska, pp. 2429–2434, 1998.
- Lo, J., Recursive neural filters. US Patent 5, 963, 929, 1999.
- Smith, Z., Dryer, M. MHD study of temporal and spatial evolution of simulated interplanetary shocks in the ecliptic plane within 1 AU. *Sol. Phys.* 129, 387–405, 1990.
- Zwickl, R.D., Dogget, K.A., Sahm, S., Barrett, W.P., Grubb, R.N., Detman, T.R., Raben, V.J., Smith, C.W., Riley, P., Gold, R.E., Mewaldt, R.A., Maruyama, T. The NOAA real-time solar-wind (RTSW) system using ACE data. *Space Sci. Rev.* 86, 633–648, 1998.
- Zwickl, R.D., Webber, W.R. Limitations of the COS approximation as applied to the cosmic-ray anisotropy. *Nucl. Instrum. Methods* 138 (1), 191–199, 1976.