Development of AI Methods in Spacecraft Anomaly Prediction

Technical Note 2

Satellite Anomaly Analysis

Version 1.0

ESA/ESTEC Contract No. 11974/96/NL/JG(SC)

P. Wintoft and L. Eliasson
Swedish Institute of Space Physics

17 September 2001

Document status sheet

Technical Note 2, SAAPS-SAAM		
Version	Date	Comment
0.1	16 December 1999	
0.2	3 April 2000	
1.0	17 September 2001	Moved the ONMI-ACE correlation study from TN 1 to this TN.

Contents

Chapter 1

Introduction

The satellite anomaly analysis module (SAAM) shall provide five different functions as described in the URD [*Wintoft,* 1999]: plotting functions, filters, statistics, guidelines, and estimate of the best prediction model. In this document it will be examined how this shall be achieved.

We first analyse part of the data in the SAAPS database by studying correlations between various parameters. We then continue to describe the tools in the analysis tool.

Chapter 2

Various analyses

2.1 Analysis of satellite anomalies

Problems are regularly experienced during the operation of satellites. These problems, or anomalies, range from change in the memory state in onboard computers to physical damage on circuitry. Lists of satellite anomalies exist in both public [*Wilkinson,* 1994] and non-public databases. The origin of the anomaly can either be the space environment or a technical problem. Several studies have shown clear links between the space environment and anomaly times [*Wrenn and Smith,* 1996] which makes it feasible to develop a system for the analysis and prediction of space environment induced anomalies.

When a satellite is exposed to electrons with energies of 1-20 keV electric charge may build up on the surface of the satellite [*Wrenn and Smith,* 1996] and cause electrostatic discharge (ESD). Electrons in this energy range at GEO are accelerated by geomagnetic substorms and are thus clustered around the midnight-morning local time sector [*Wrenn and Smith,* 1996, *Dyer and Rodgers,* 1999]. The anomalies from the S001 data set show a clear clustering around 3 hours local time. The interpretation is thus that the anomalies are due to surface charging. [*Wrenn and Smith,* 1996] also studies the probability for anomalies as a function of both local time and Kp, where Kp serves as an indicator of keV electron flux. This type of analysis can be used to identify surface ESD effects.

Internal charging, or deep dielectric charging, can occur at times of enhanced fluxes of MeV electrons. Electrons are trapped in dielectric materials and charge can build up over several hours to a few days until a discharge may occur. [*Wrenn and Smith,* 1996] analyzed some 140 anomalies from the DRAδ satellite. A key feature of the anomalies were that they where preceded by a charging time of more than 30 hours. Based on this a correlation was made between the anomalies and the daily average flux of the > 2 MeV electrons measured at GOES-7. There was a clear threshold in the electron flux below which no anomalies occurred.

[*López Honrubia and Hilgers,* 1997] studied five years of anomaly data from two consecutive Meteosat satellites, MOP-1 and MOP-2, together with the daily average electron flux for energies above 2 MeV. It was shown that there were a clear trend that the anomalies occur during days with high flux values. However, for individual anomaly events the flux values for the preceding days showed a large degree of variation with no unique pattern leading to the anomaly. Two different methods were applied to make a classification of the anomaly and non-anomaly events: a linear correlation method and a non-linear neural network.

In the work by [*Andersson et al.* 1998] and [*Wu et al.* 1998] neural network models were developed for the prediction of anomalies for specific anomaly data sets based on both local and non-local space environment data. These prediction models will be further addressed in Technical Note 3 [*Wintoft and Eliasson* 2001].

2.1.1 Correlation between different anomaly data sets

It is interesting to examine the correlation between the different anomaly sets as this will indicate how general a prediction model can be. We perform this analysis using conditional probabilities and relative mutual information as defined in Sections 3.3 and 3.4, respectively.

To proceed we select only those anomalies that are believed to be ESD events. Then we select two variables X and Y which corresponds to anomalies from data set A and B, respectively. X and Y can have the values $x = 0, 1$ and $y = 0, 1$, respectively, where a zero indicates no anomalies during a day and a one indicates one or more anomalies during the day. Table 2.1 summarizes conditional probabilities and relative mutual information calculated on the daily anomaly data. The number of days for which the data sets overlaps are also given. The conditional probability $P(X = 0|Y = 0)$ states that if a 'no-anomaly' is observed in set B then what is the probability that a 'no-anomaly' is also observed in set A. Similarly we also have $P(X = 1|Y = 1)$ for the anomaly events. We can also compute the probability that the observations are the same in both set A and B given an observed 'no-anomaIy' or 'anomaly' event in set B, i.e. $P(X = Y|Y)$. A value $P(X = Y|Y) > 0.5$ means that we are doing better than guessing. The last column gives the relative mutual information $I_r(X;Y)$. If $I_r(X;Y) = 0$ then there is no relation between the two sets. If $I_r(X;Y) = 1$ then there is a perfect correlation.

From the P_{00} column we see that if we observe a 'no-anomaly' event in one data set it is likely that we will also observe a 'no-anomaly' event in any other set. However, from the P_{11} column we see that the probability is much lower for observing simultaneous 'anomaly' events in any two sets. The P_{00} and P_{11} conditional probabilities are summarized in the P_{XY} column which gives the probability that an observed event in one set will be the same as in

any other set. The numbers indicate that using the events from one data set to predict whether there will be the same event in another set will perform slightly better than just guessing the event. The highest success will be to use the S003 data to predict the S004-S022 data which gives a probability of 63% that the prediction will be correct. The final column, $I_r(X;Y)$, is the relative mutual information which also shows the very weak correlation between any two data sets, except for the S003 and S004-S022 data.

To be more precise, we have in the above actually calculated the probabilities in going from set B to set A ($P(A|B)$). However, if we reverse the ordered so that we use A to predict B the probabilities and relative mutual information will be very similar to the numbers in Table 2.1.

To conclude, we see that if we have a model for the prediction of daily ESD anomalies for one satellite then that model will not work for the other satellites. However, it might still be possible that a specific model may be tuned, by e.g. changing some threshold value or time delays, to better predict the anomalies for another satellite.

Table 2.1: Cross analysis between the different anomaly sets. A and B are the data sets; N is the number of overlapping days; $P_{00} = P(X = 0|Y = 0)$, i.e. given that there is an observed 'no-anomaly' in set B what is the probability that a 'no-anomaly' will also be observed in set A; $P_{11} = P(X = 1|Y = 1)$; $P_{XY} = P(X = Y|Y)$, i.e. given an observed event ('anomaly' or 'no-anomaly') in set B what is the probability that the observed event in set A will be the same as in B; and $I_r(X,Y)$ is the relative mutual information.

A	B	N	P_{00}	P_{11}	P_{XY}	$I_r(X,Y)$
S001	S002	2623	0.88	0.26	0.57	0.02
S001	S003	1728	0.95	0.17	0.56	0.02
S001	S004-S022	4462	0.91	0.23	0.57	0.02
S002	S003	1601	0.95	0.16	0.56	0.02
S002	S004-S022	2070	0.97	0.11	0.54	0.02
S003	S004-S022	1048	0.97	0.29	0.63	0.10

2.2 $\sum K_p$ and its relation to anomalies

In [*Wintoft and Eliasson* 2001] we will explore models for the predictions of daily anomalies from $\sum K_p$. However, we start here by analysing the relation between $\sum K_p$ and the anomalies.

We calculate the relative mutual information from $\sum K_p$ to daily anomalies. From past experience we know that the response of anomalies usually include the variation of $\sum K_p$ over several days. However, to calculate the mutual information we must bin $\sum K_p$ and assign one class to each bin. If we thus study a time delay line over 10 days we get (number of bins)^{10} com-

binations which becomes impossible to handle even if the number of bins is small. To be able to proceed we instead form averages of $\sum K_p$ extending from one to ten days. In this process some information is lost but we can get an overall picture of the situation.

We start with selecting all data in the five anomaly sets and compare it with the average $\sum K_p$. The result is shown in Figure 2.1. We see that the relative mutual information (RMI) $(I(X;Y)/H(Y))$ is generally below 0.15 and that the different anomaly sets respond to $\sum K_p$ differently. The S001 anomalies shows the most direct response to $\sum K_p$ and S002 and S003 anomalies are best related to about four days averages of $\sum K_p$.

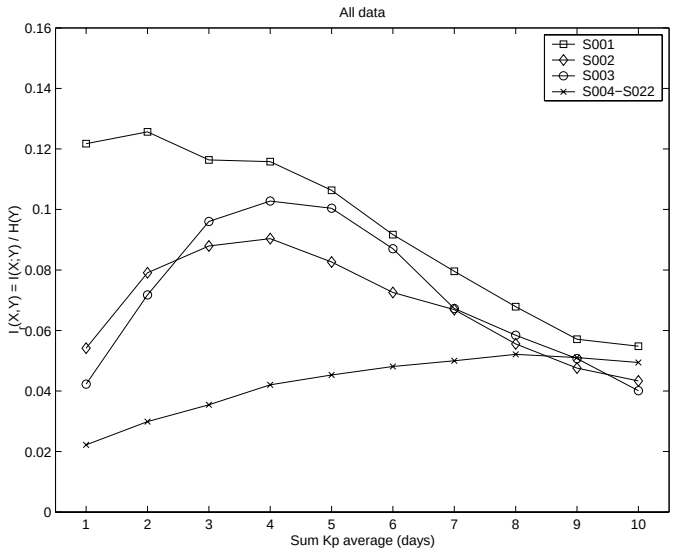

Figure 2.1: The relative mutual information $I(X;Y)/H(Y)$ for the different anomaly sets as a function of the average $\sum K_p$, where the average goes from 1 to 10 days. The anomaly sets contain all the anomaly data.

We can refine the analysis by selecting a subset of the anomalies that are believed to be more related to $\sum K_p$. We make the same selection as in Section 2.1.1 which means that we only select the ESD anomalies. Now the relation between $\sum K_p$ and the anomalies becomes stronger as seen in Figure 2.2. The S001 anomalies are best related to one to two day averages of $\sum K_p$, the S002 anomalies for two to four day averages, S003 for three to five day averages, and S004-S022 for seven to eight day averages. This is again a confirmation of the results shown in Table 2.1 that the anomalies are quite different for the different satellites.

Finally, we can examine the effect of one day $\sum K_p$ with time delays from -10 to 10 days. The result is shown in Figure 2.3. We see that the S001 anomalies are most related to $\sum K_p$ for the same day. The S002 anomalies peaks at 1 day delay and S003 at 2 days.

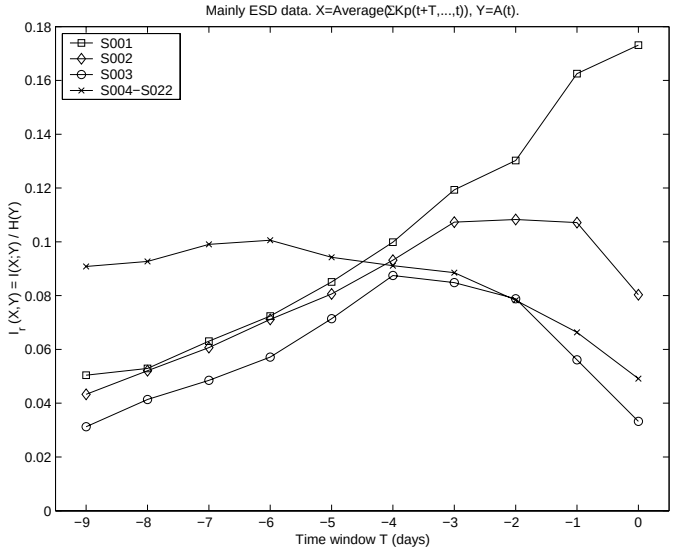

Figure 2.2: The relative mutual information $I(X;Y)/H(Y)$ for the different anomaly sets as a function of the average $\sum K_p$, where the average goes from 1 to 10 days. A subset of the anomaly data is selected that mainly related to ESD as explained in the text.

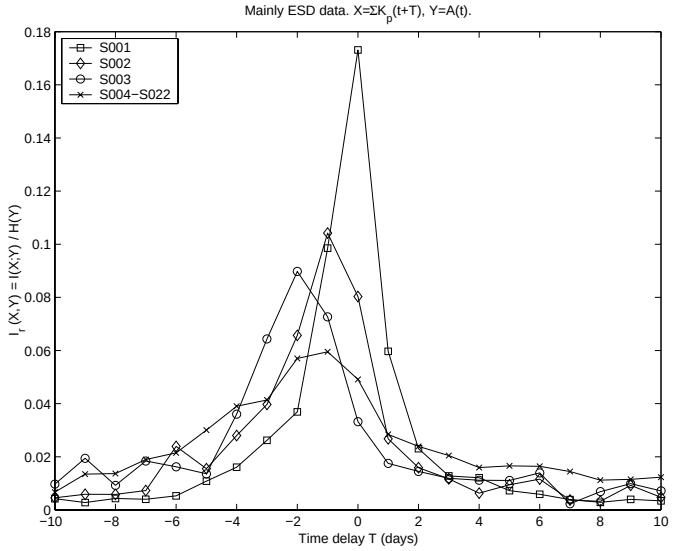

Figure 2.3: The relative mutual information $I(X;Y)/H(Y)$ for the different anomaly sets as a function of $\sum K_p$ delayed -10 to 10 days. A subset of the anomaly data is selected that mainly related to ESD as explained in the text.

2.3 Daily GOES $>$ 2 MeV electron fluence and anomalies

We repeat the same analysis used for $\sum K_p$ to the daily $>$ 2 MeV electron fluence. The result is shown in Figure 2.4. The time delay between the fluence and the anomalies is again varied from -10 to 10 days.

The most striking difference between Figure 2.3 and Figure 2.4 is that $\sum K_p$ is leading the anomalies, except for S001 which is simultaneous, while the fluence is lagging the anomalies, except for S003 which is also simultaneous. This, at first appearance, strange result will be further addressed below.

The S001 anomaly set contain mainly surface ESDs and $\sum K_p$ works as an proxy for the keV electron flux. The peak correlation is $I_r \approx 0.17$ with no time delay ($T = 0$ days). The correlation drops to a maximum $I_r \approx 0.05$ at $T = 2$ to 3 days for the $>$ 2 MeV electron fluence. Thus, both the correlation drops and the nowcasting becomes a post-casting when we change $\sum K_p$ to the $>$ 2 MeV electron fluence. It is known that increases in the keV electron flux is more directly related to the solar wind structures causing substorms, while the MeV flux is also related to the solar wind but not directly to substorms [*Baker et al.* 1997]. The MeV flux tend to lag the keV flux by up to several days. Therefore, for surface ESDs we expect this kind of behaviour.

The $>$ 2 MeV electron fluence and S003 anomaly data has a peak correlation of $I_r \approx 0.25$ at $T = 0$ (Figure 2.4). This indicates that the anomalies are dominated by internal ESDs. At the same time we see from Figure 2.3 that the correlation between $\sum K_p$ and anomalies is weaker ($I_r \approx 0.09$) but that $\sum K_p$ is leading with 2 days.

The other two anomaly sets probably contain both internal and surface ESDs in more equal numbers and thus have correlation and time delay in between the S001 and S003 sets.

2.4 The correlation between OMNI and ACE solar wind data

The solar wind data in the OMNI data set come from a number of near-Earth solar wind spacecraft. The aim of the OMNI set is to compile the data from several sources and to make the set as compatible as possible. The data comes from spacecraft that has mainly been close to the Earth, like the IMP-8 that was in an $30 \times 40 R_E$ geocentric orbit. The cross-correlation between the different parameters for different spacecraft was then examined, and when the systematic errors were larger than the random errors a cross-normalization was adopted. It was found that only the density and temperature needed to be cross-normalized, whereas the IMF parameters

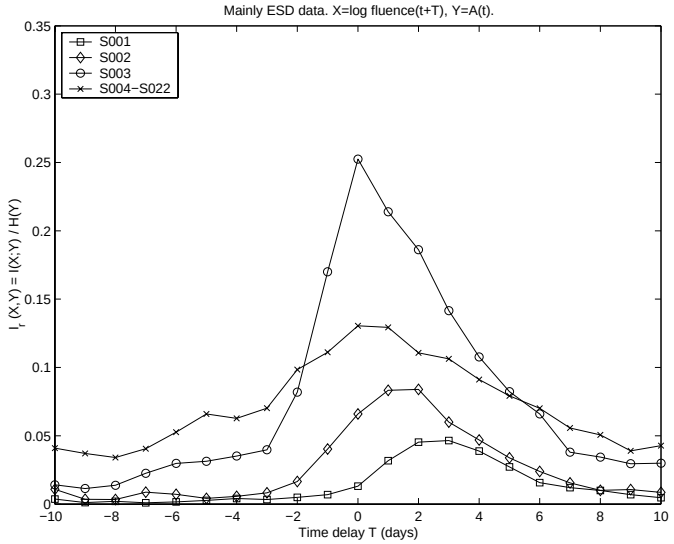

Figure 2.4: The relative mutual information $I(X;Y)/H(Y)$ for the different anomaly sets as a function of log daily > 2 MeV electron fluence delayed -10 to 10 days. A subset of the anomaly data is selected that mainly related to ESD as explained in the text.

and the flow speed always had systematic differences smaller than the random errors. As the ISEE-3 spacecraft was located at the Lagrange L1 point about $240R_E$ upstream from the Earth this data was time shifted to a near-Earth location. The time shifts are corotation

$$\tau_{\rm rot} = \frac{x}{V}\left\{\frac{1+\frac{V}{R\Omega}\frac{y}{x}}{1-\frac{V_E}{R\Omega}}\right\}, \tag{2.1}$$

and convection

$$\tau_{\rm vec} = \frac{x_2 - x_1}{V}. \tag{2.2}$$

It is now interesting to study the relation between the OMNI data and the ACE data. Figure 2.5 shows the correlation between the OMNI magnetic field data and the ACE magnetic field data for 1998. The quantity plotted is the average field magnitude

$$F = \langle |\mathbf{B}| \rangle, \tag{2.3}$$

where the angle brackets denote the time average. Generally, the two data sets show a very good agreement. The few outliers are marked with squares or diamonds. The diamonds represent times when $B_{\rm ACE} - B_{\rm OMNI} < 5$ nT and the squares when $F_{\rm ACE} - F_{\rm OMNI} < -5$ nT.

Next we consider the magnitude of the average magnetic field and the magnetic field components. Figure 2.6 shows the correlation plots for the

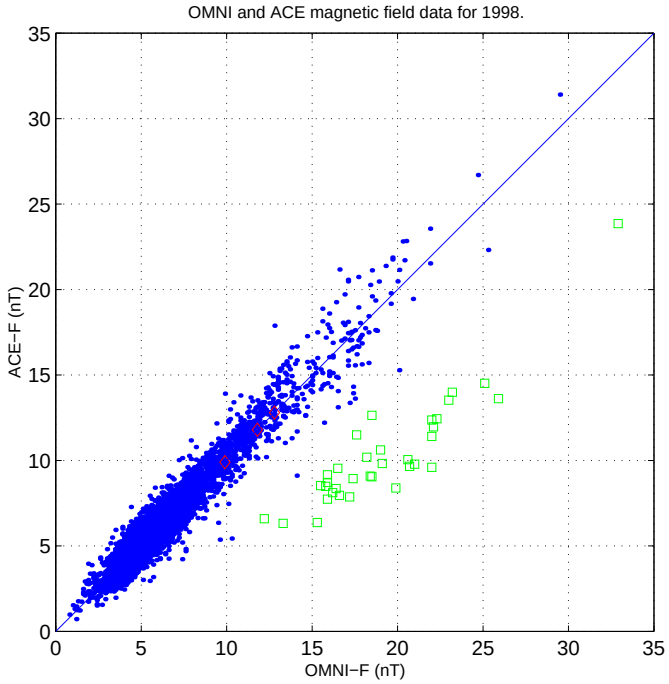

Figure 2.5: Correlation plots between the hourly OMNI magnetic field data and the ACE magnetic field data over the year 1998. The figure shows the average field magnitude (F).

same data set as in Figure 2.5. Again, there is a very good agreement between the two data sets with only a few outliers.

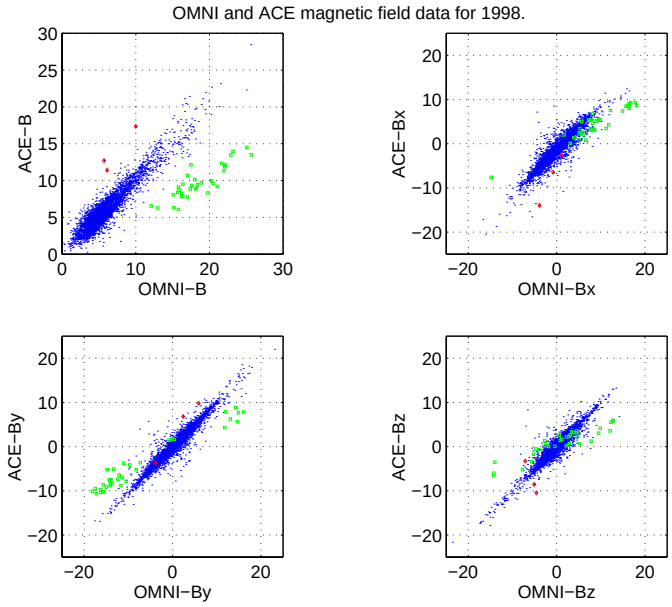

Figure 2.6: Correlation plots between the hourly OMNI magnetic field data and the ACE magnetic field data over the year 1998. The figures show the magnitude of the average magnetic field (B), and the magnetic field components (B_x,B_y,B_z).

To conclude we can say that there is a very good agreement between the hourly ACE and OMNI data sets. Models developed on data based on the OMNI set can thus also be used with the ACE data. Care has only to be taken in how the hourly averages are formed and how the L1-Earth time shift is introduced. In Figures 2.5 and 2.6 the data has been averaged differently. The ACE data are lagging averages $\langle B \rangle(t) = \int_{t'=t-1}^{t} B(t')dt'$, while the OMNI data are following averages $\langle B \rangle(t) = \int_{t'=t}^{t+1} B(t')dt'$. The two different averages will thus introduce a one hour time shift between the two data sets, and one hour is approximately the time it takes the solar wind to travel from L1 to Earth.

Chapter 3

The satellite anomaly analysis module

3.1 Basic operations

3.1.1 The output from the database

A time series of a parameter are obtained from the SAAPS database by calling the request method from the the database tool. All data, except the anomaly data, are contiguous. Any data gaps in the time series are indicated with NaN (Not a Number). The output from the database tool is a vector of objects, were each object contain the time of the observation and one or several data values depending on which parameter that has been requested. There are three arguments that must be specified when requesting data: the parameter, start time, and end time. E.g., if the requested parameter is the magnetic field data from the ACE spacecraft (ACE-MFI), for a period from t_0 to t_1, then the database tool would return

$$X = \begin{bmatrix} t_0 & B(t_0) & B_x(t_0) & B_y(t_0) & B_z(t_0) \\ t_0 + \Delta t & B(t_0 + \Delta t) & B_x(t_0 + \Delta t) & B_y(t_0 + \Delta t) & B_z(t_0 + \Delta t) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_1 & B(t_1) & B_x(t_1) & B_y(t_1) & B_z(t_1) \end{bmatrix}, \tag{3.1}$$

where Δt is the sampling interval.

3.1.2 Handling data gaps

Generally, all the parameters in the SAAPS database contain occasional data gaps. To be able to make any further mathematical analysis these data gaps have to be treated.

The safest approach is to simply to create a data set in which the times with data gaps have been removed. However, this may lead to small data

sets.

The data gaps can also be replaced with linearly interpolated data values. This is achieved by searching one column at at time for NaN's and then interpolate the value. Data gaps can be contiguous and thus extend over several time steps. The number of time steps over which it is acceptable to interpolate must be given. This number can be estimated from e.g. a cross-correlation analysis. The algorithm for removing data gaps is given below.

1. Set i_1 to the first instance of NaN in $X(i, j)$ for column j.

2. If $X(i_1 + 1, j) = \text{NaN}$ set $i_2 = i_1 + 1$.

3. Continue with step 2 until $X(i_n + 1, j) \neq \text{NaN}$.

4. If $n \leq m$, where m is the maximum number of contiguous time steps to be interpolated, then

$$\hat{X}(i_k, j) = \frac{k}{n+1}(X(i_1 - 1, j) + X(i_n + 1, j),\ 1 \leq k \leq n.$$

5. Set i_1 to the next instance of NaN after i_n.

6. Continue with step 2 for rows i and all columns j.

3.1.3 Averaging time series data

To be able to perform studies between parameters with different sample intervals an averaging must be performed to reduce the time resolution for the parameter with the highest sampling rate to the resolution of the parameter with the lowest sampling rate. Also, models my not require the highest sampling rate.

Assume that the original time series x_0, x_1, ..., x_m are equidistant sampled at times $t_0^{(x)}$, $t_1^{(x)}$, ..., $t_m^{(x)}$ where the sample interval is $\Delta t^{(x)}$. We thus have

$$x_i = x(t_i^{(x)}) = x(t_{i-1}^{(x)} + \Delta t^{(x)}). \tag{3.2}$$

We also assume that the resolution $\Delta T^{(x)}$ equals the sample interval, i.e. $\Delta T^{(x)} = \Delta t^{(x)}$. We now want to average and resample the time series x_i to a new time series y_j so that

$$y_j = y(t_j^{(y)}) = y(t_{j-1}^{(y)} + \Delta t^{(y)}), \tag{3.3}$$

where $t_j^{(y)}$ and $\Delta t^{(y)}$ are the new sample times and sample interval, respectively. The time resolution of the new series is $\Delta T^{(y)}$ and is not necessary

equal to the sample interval $\Delta t^{(y)}$. The relation between the two series becomes

$$y_j = \frac{1}{p} \sum_{i=sj}^{sj+p-1} x_i, \tag{3.4}$$

where the number of points to average is

$$p = \frac{\Delta T^{(y)}}{\Delta T^{(x)}} = \frac{\Delta T^{(y)}}{\Delta t^{(x)}} \tag{3.5}$$

and the point sample interval is

$$s = \frac{\Delta t^{(y)}}{\Delta t^{(x)}}. \tag{3.6}$$

The new sample times becomes

$$t_j^{(y)} = t_{sj}^{(x)} + \frac{\Delta T^{(y)} - \Delta t^{(x)}}{2} + t_{\text{off}}^{(y)}, \tag{3.7}$$

where we also introduce a time offset $t_{\text{off}}^{(y)}$. The time offset determines to which time an average belongs. E.g. if $t_{\text{off}}^{(y)} = 0$ then we have a central average. If $t_{\text{off}}^{(y)} = -(\Delta T^{(y)} - \Delta t^{(x)})/2$ then the time is in the beginning of the average interval (forward average), while if $t_{\text{off}}^{(y)} = (\Delta T^{(y)} - \Delta t^{(x)})/2$ it is at the end of the average interval (lagging average). From a physical point of view the central average is to prefer, while for real time operation the lagging average must be used. The hourly average data from the Space Environment Center (SEC) and the OMNI data set use forward average data.

The latest time in the original time series is $t_m^{(x)}$. When performing the average in Equation 3.4 the last position to use is $sj + p - 1$. This means that $sj + p - 1$ has a maximum value of m, thus $sj \leq m - p + 1$.

$\Delta T^{(y)}$ and $\Delta t^{(y)}$ should also be multiples of $\Delta t^{(x)}$ to ensure that p and s are integer numbers.

Any data gaps that exist in the time series that are averaged will also be present in the resulting time series. This means that if a times series with five minute resolution is averaged to one hour resolution, and if there is a NaN for one point, then that one hour interval will also be a NaN. If this is to be avoided the time series should first be interpolated to remove any NaN's.

3.1.4 Error measures

To asses the performance of a model several different error measures exist.

The mean-square-error (MSE) is defined as

$$\text{RMSE} = \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2. \tag{3.8}$$

The root-mean-square error (RMSE) is defined as

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}}. \tag{3.9}$$

The variance is closely related to the MSE and is defined as

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2. \tag{3.10}$$

Taking the square root of the variance we get the standard deviation σ. The prediction efficiency (PE) [*Detman,* 1998] is the MSE normalized with respect to the variance

$$\mathrm{PE} = 1 - \frac{\mathrm{MSE}}{\sigma^2}. \tag{3.11}$$

If the MSE equals the variance the PE becomes 0, and if MSE is zero the PE becomes equal to 1. To compare different prediction techniques skill scores [*Detman,* 1998] can be used. The skill score (SS) is defined as

$$\mathrm{SS} = 1 - \frac{\mathrm{MSE}}{\mathrm{MSE}_{\mathrm{ref}}}, \tag{3.12}$$

where the MSE of the model is related to the MSE of a reference model ($\mathrm{MSE}_{\mathrm{ref}}$). If $\mathrm{SS} > 0$ then the new model is better than the reference model, while if $\mathrm{SS} < 0$ the new model is less good than the reference model.

3.2 Linear correlation

The linear correlation coefficient is calculated as

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}}, \tag{3.13}$$

where $\bar{x}$ is the mean of the x_i's and $\bar{y}$ is the mean of the y_i's [*Press et al.,* 1992].

The correlation coefficient r will be equal to -1 if the two variable are perfectly anti-correlated, and equal to 1 if they are perfectly correlated. If $r = 0$ then there exist no linear correlation. However, even if r is small or zero there might exist a non-linear relation.

3.3 Probability and conditional probability

Assume we have the two discrete random variables X and Y that can assume the values $x = 0, 1, \ldots, m-1$ and $y = 0, 1, 2, \ldots, n-1$, respectively. Further, assume we have k observations of the two variables where the number of outcomes for each value in x and y are a_x and b_y, respectively. We see that

$\sum_x a_x = \sum_y b_y = k$. We can now calculate the probability that X will take on the value x as

$$P(X = x) = p(x) = \frac{a_x}{k}. \tag{3.14}$$

Similarly we get for Y

$$P(Y = y) = p(y) = \frac{b_y}{k}. \tag{3.15}$$

The joint probability then becomes

$$P(X = x, Y = x) = P(X = x)P(Y = y) = \frac{a_x b_y}{k^2}. \tag{3.16}$$

The conditional probability $P(X = x|Y = y)$ gives us the probability the x will be observed given that y has been observed. The conditional probabilities are calculated as

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} \tag{3.17}$$

and

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)}. \tag{3.18}$$

It is worth noting that $\sum_x P(X = x|Y = y) = 1$.

3.4 Entropy and mutual information

The linear correlation coefficient estimates how well linearly correlated two variables are. Thus, a small value of the correlation coefficient only indicates that the two variables are weakly linearly related, but there might be a non-linear correlation. The mutual information is another way to examine the correlation between two variables, and it does not assume anything about the functional dependency between the two variables [*Swingler,* 1996, *Deco and Obradovic* 1996].

As in the previous section, assume that we have a discrete random variable X that can take m different discrete values x with probabilities $p(x)$. The entropy of X then becomes

$$H(X) = -\sum_x p(x) \ln p(x). \tag{3.19}$$

As $0 \leq p(x) \leq 1$ the entropy will always be positive. The entropy will be maximum for a uniformly distributed variable; $p(x) = 1/m$ gives $H(X) = \ln m$.

Similar to the joint and conditional probabilities we also have the joint entropy

$$H(X_1, X_2) = -\sum_{x_1} \sum_{x_2} p(x_1, x_2) \ln p(x_1, x_2), \tag{3.20}$$

and the conditional entropy

$$H(X_1|X_2) = -\sum_{x_1}\sum_{x_2} p(x_1, x_2) \ln p(x_2|x_1). \tag{3.21}$$

Finally, the mutual information is defined as

$$I(X_1; X_2) = \sum_{x_1}\sum_{x_2} p(x_1, x_2) \ln \frac{p(x_1, x_2)}{p(x_1)p(x_2)}. \tag{3.22}$$

Note that the ";" in $I(X;Y)$ means that X and Y are not interchangeable, thus $I(X;Y) \neq I(Y;X)$. The mutual information lies in the interval $0 \leq I(X;Y) \leq H(Y)$. If $I(X;Y) = 0$ then X and Y are uncorrelated, whereas if $I(X;Y) = H(Y)$ they are perfectly correlated. We can normalize the mutual information with respect to the entropy to get the relative mutual information

$$I_r(X;Y) = \frac{I(X;Y)}{H(Y)} \tag{3.23}$$

which now is a number between 0 and 1.

As an example we can study the mapping $y = \sin x$ on the interval $x \in [0, 2\pi]$. We bin the continuous values into discrete values with a bin size of 0.1 and calculate the probabilities $p(x)$, $p(y)$, and $p(x, y)$. The result is shown in Figure 3.1. The computed entropies are $H(X) = 3.45$ and $H(Y) = 2.35$ and the mutual information is $I(X;Y) = 1.88$. The relative mutual information then becomes $I_r(X;Y) = 0.801$ and thus predicting y from x should be successful. On the other hand we have $I_r(Y;X) = I(X;Y)/H(X) = 0.546$ which means that making the inverse prediction $y \mapsto x$ is more difficult. By examining the sine curve or the probabilities we see that going from x to y is single valued while going from y to x is double valued.

3.5 Superposed epoch analysis

An efficient way to examine if there are trends in a parameter associated with a list of events is to use superposed epoch analysis. For this one needs a list of events and the size and location of the analysis window. Assume we have a list of anomalies at times

$$t_i\ ,\ 1 \leq i \leq m. \tag{3.24}$$

Next, we set the analysis window to start at

$$t_i^{(s)} = t_i + p\Delta t, \tag{3.25}$$

and end at

$$t_i^{(e)} = t_i + q\Delta t. \tag{3.26}$$

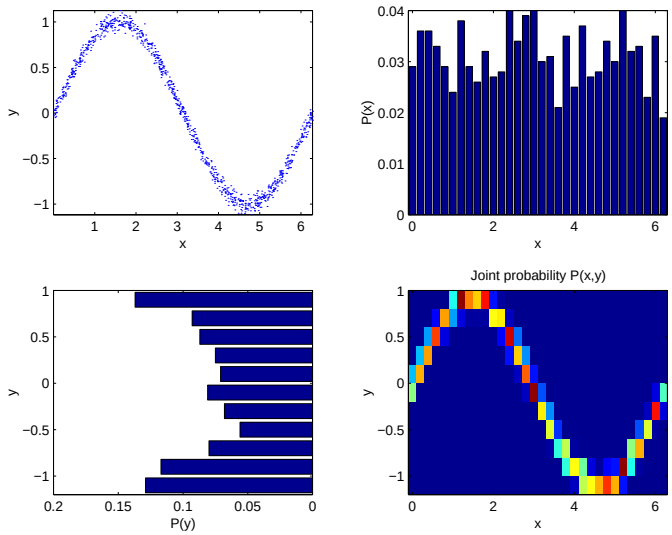

Figure 3.1: A sine curve $y = \sin x$ (top left) and the probabilities $p(x)$ (top right) and $p(y)$ (bottom left) using a bin size of 0.1, and the joint probabilities $p(x, y)$ (bottom right).

Generally $p < 0$ and $q \geq 0$ so that the window start before the event and end at or after the event. The sample interval for the parameter is Δt. A matrix is created from the parameter $x(t)$ that should be superposed as

$$X_{ij} = x(t_i^{(s)} + j\Delta t)\ ,\ 0 \leq j \leq n, \tag{3.27}$$

where $n = q - p$. In the above it is assumed that the event times t_i are positioned at the sample times of the parameter $x(t)$. If this is not the case the event times are easily moved to the sample times by rounding the event time to the closest sample time. Then, the final step is to calculate the superposed values

$$s_j = \sum_{i=1}^{m} X_{ij}, \tag{3.28}$$

or alternatively the superposed average

$$\hat{s}_j = \frac{1}{m} s_j. \tag{3.29}$$

If there are any data gaps in the original time series $x(t)$ the matrix X_{ij} will contain NaN's at the corresponding positions. The value s_j at position j will then have a NaN if one or more rows of X_{ij} contain a NaN at position j. Obviously, if there are a large number of events (m large) then the probability that s_j will only contain NaN's increases. Therefore, it is desirable to replace data gaps with interpolated values.

3.6 Determining the best prediction model

From a user submitted list of anomalies we can determine the anomaly prediction model that best matches the submitted list. The prediction models are covered in [*Wintoft and Eliasson* 2001].

Assume that we have a collection of models

$$F^i(t)\ ,\ i = 1, 2, \ldots, q \tag{3.30}$$

that return the probability for an anomaly at time t. If for a specific time $t = t^{\mathrm{na}}$ we have $F^i(t^{\mathrm{na}}) < 0.5$ then the prediction is 'no anomaly'. At another time $t = t_{\mathrm{a}}$ we have $F^i(t_{\mathrm{a}} > 0.5$ which means that we have an 'anomaly'. Based on the user submitted lists of anomaly events and no anomaly events each model can be run for all the events.

Let the user event list contain the times $t = [t_1^{\mathrm{na}}, t_2^{\mathrm{na}}, \ldots, t_m^{\mathrm{na}}, t_1^{\mathrm{a}}, t_2^{\mathrm{a}}, \ldots, t_n^{\mathrm{a}}]$. The desired output for these events is $y = [0, 0, \ldots, 0, 1, 1, \ldots, 1]$ where the first sequence is made up of m zeros (no-anomaly) and the second sequence of n ones (anomaly).

The output from model i is

$$x^i = F^i(t). \tag{3.31}$$

The conditional probability can be used to calculate the probability that the observed event is predicted by model i

$$P^i_{XY} = P(X = x^i | Y = y). \tag{3.32}$$

The probability the a no-anomaly event will be predicted is

$$P^i_{XY}(\mathrm{na}) = P(X = x^i(\mathrm{na}) | Y = y(\mathrm{na})) \tag{3.33}$$

and the probability that an anomaly will be predicted

$$P^i_{XY}(\mathrm{a}) = P(X = x^i(\mathrm{a}) | Y = y(\mathrm{a})). \tag{3.34}$$

These two probabilities can be averaged

$$Q = \frac{P^i_{XY}(\mathrm{na}) + P^i_{XY}(\mathrm{a})}{2}. \tag{3.35}$$

The best prediction model is the model that maximizes Q. Using Q we will get the model that gives the highest probabilities of predicting both anomalies and no-anomalies without getting biased by an unbalanced data set. A model is considered useful if $Q > 0.5$.

To demonstrate the average conditional probability Q we can study a few extreme cases.

Lets assume that we have an event set that contain 10% anomalies. A simple model based only on this percentage of anomalies would always

predict no-anomalies. Then we would get $P^i_{XY}(\text{na}) = 1$, $P^i_{XY}(\text{a}) = 0$, and $Q = 0.5$. This model will always be correct for the no-anomalies and always be wrong on the anomalies, i.e. not a very useful model.

As another example assume that the model correctly predict 80% of the no-anomalies but only 50% of the anomalies. Then we get $P^i_{XY}(\text{na}) = 0.8$, $P^i_{XY}(\text{a}) = 0.5$, and $Q = 0.65$.

One value of Q can correspond to range of probabilities for the no-anomaly and anomaly predictions. I.e. two different models that give $Q = 0.8$ may have in one case $P_{XY}(\text{na}) = 0.8$ and $P_{XY}(\text{a}) = 0.8$, but in another case $P_{XY}(\text{na}) = 0.9$ and $P_{XY}(\text{a}) = 0.7$. The first model is equally good in predicting no-anomalies as anomalies, while the second model is more biased to the no-anomaly predictions. The optimization procedure will not distinguish between these two cases.

Bibliography

[*Andersson et al.* 1998] Andersson, L., L. Eliasson, and O. Norberg, Spacecraft anomaly forecasting using local environment data, *Technical Note, WP 210, Study of Plasma and Energetic Electron Environment and Effects*, ESTEC/Contract No. 11974/96/NL/JG(SC), 1998.

[*Baker et al.* 1997] Baker, D.N., X. Li, N. Turner, J.H. Allen, L.F. Bargatze, J.B. Blake, R.B. Sheldon, H.E. Spence, R.D. Belian, G.D. Reeves, S.G. Kanekal, B. Klecker, R.P. Lepping, K. Ogilvie, R.A. Mewaldt, T. Onsager, H.J. Singer, and G. Rostoker, Recurrent geomagnetic storms and relativistic electron enhancements in the outer magnetosphere: ISTP coordinated measurements, *J. Geophys. Res. 102*, 14,141–14,148, 1997.

[*Deco and Obradovic* 1996] Deco, G., and D. Obradovic, *An Information-Theoretic Approach to Neural Computing*, Springer-Verlag, New York, 1996.

[*Detman,* 1998] Detman, T.R., Toward real-time operational model predictions, *Workshop on AI Applications in Solar-Terrestrial Physics*, Lund, Sweden, 29-31 July 1997, ESA, WPP-148, 135–143, 1998.

[*Dyer and Rodgers,* 1999] Dyer, C., and D. Rodgers, Effects on spacecraft and aircraft electronics, *Workshop on space weather,* ESTEC, Noordwijk, WPP-155, 17–27, 1999.

[*López Honrubia and Hilgers,* 1997] López Honrubia, F.J. and A. Hilgers, Some correlation techniques for environmentally induced anomalies analysis, *J. Spacecraft and Rockets,* 670–674, 1997.

[*Press et al.,* 1992] Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C*, Cambridge University Press, New York, 1992.

[*Swingler,* 1996] Swingler, K., *Applying neural networks: a practical guide*, Academic Press Ltd, London, 1996.

[*Wilkinson,* 1994] Wilkinson, D.C., Spacecraft Anomaly Manager, National Oceanic and Atmospheric Administration, National Geophysical Data Center, Solar-Terrestrial Physics Division, Boulder, CO, 1994.

[*Wintoft,* 1999] Wintoft, P., SAAPS WP 210 – Satellite Anomaly Analysis Module User Requirements Document, Version 1.1, 1999.

[*Wintoft,* 1999] Wintoft, P., SAAPS WP 120 – Technical Note 1, Satellite Anomaly Analysis Module, Version 0.1, 1999.

[*Wintoft and Eliasson* 2001] Wintoft, P., and L. Eliasson, Satellite Anomaly Prediction Module, *Technical Note 3, Development of AI Methods in Spacecraft Anomaly Predictions,* ESA/ESTEC Contract No. 11974/96/NL/JG(SC), 2001.

[*Wrenn and Smith,* 1996] Wrenn, G.L., and R.J.K. Smith, The ESD threat to GEO satellites: empirical models for observed effects due to both surface and internal charging, *ESA Symposium Proceedings on Environment Modelling for Space-based applications,*, ESTEC, Noordwijk, SP-392, 121–124, 1996.

[*Wu et al.* 1998] Wu, J.G., H. Lundstedt, L. Andersson, L. Eliasson, and O. Norberg, Spacecraft anomaly forecasting using non-local environment data, *Technical Note, WP 220, Study of Plasma and Energetic Electron Environment and Effects*, ESTEC/Contract No. 11974/96/NL/JG(SC), 1998.