

ESTEC/Contract No. 12854/98/NL/NB
Space Environment Database (SEDAT)
WP303. Spacecraft anomalies analysis

Report on demonstration of correlation of spacecraft anomalies with environment data

RAL-SED-RP-0303
Issue 1.3, 22 January 2004

Authors: Mike Hapgood, CLRC Rutherford Appleton Laboratory

ESA technical Officer: A Hilgers D/TOS, TOS-EMA



SEDAT	Doc. No: Issue: 1.3	RAL-SED-RP-0303 Date: 22/01/2004
Report on spacecraft anomalies analysis		Page 2

© CCLRC 2004

This document is approved for wider release by ESA under the terms of ESA Contract 12854/98/NL/NB

SEDAT	Doc. No: Issue: 1.3	RAL-SED-RP-0303 Date: 22/01/2004
Report on spacecraft anomalies analysis		Page 3

Table of contents

1	PREFACE.....	4
1.1	DOCUMENT CHANGE RECORD.....	4
1.2	PURPOSE OF THE DOCUMENT.....	4
1.3	DEFINITIONS, ACRONYMS AND ABBREVIATIONS.....	4
1.4	IMPORTANT DOCUMENTS.....	5
1.4.1	<i>Applicable documents.....</i>	5
1.4.2	<i>Reference documents.....</i>	5
1.5	ACKNOWLEDGEMENTS.....	5
2	INTRODUCTION AND STRUCTURE.....	6
3	DATASETS USED IN THE DEMONSTRATION	7
3.1	ANOMALY DATA.....	7
3.2	METEOSAT POSITION.....	7
3.3	METEOSAT DATA.....	8
3.4	GOES DATA.....	8
4	MLT ANALYSIS	9
5	SUPERPOSED EPOCH ANALYSIS	10
5.1	INTRODUCTION.....	10
5.2	TIME-TAG TOOL.....	11
5.3	AVERAGING TOOL.....	12
5.4	PLOTTING TOOL.....	13
5.5	CONTROL TEST - RANDOM DATA SELECTION.....	16
6	CONCLUSIONS	18
6.1	CAVEATS.....	18
6.2	FUTURE WORK.....	19
7	ANNEX A – SUMMARY OF TOOLS USED IN THIS DEMONSTRATION.....	20
8	ANNEX B - COMPLIANCE MATRIX	24

SEDAT	Doc. No: Issue: 1.3	RAL-SED-RP-0303 Date: 22/01/2004
Report on spacecraft anomalies analysis		Page 4

1 Preface

1.1 Document change record

Issue	Date	Notes/remarks
0.5	21 Nov 2002	Intermediate draft released for review by ESA
1.0	09 Feb 2003	Added annex to summarise the SEDAT tools used in this workpackage
1.1	18 Feb 2003	Added conclusions section (6). Includes caveats and ideas for future work. Added compliance matrix (Annex B) between this document and the demo plan
1.2	08 Dec 2003	Add user names to names of SEDAT objects Minor editorial changes
1.3	22 Jan 2003	Update Annex A to give a better listing of the top-level tools, queries and their parameters.

1.2 Purpose of the document

This document is the technical note reporting the results of the application of SEDAT to the demonstration of correlation of spacecraft anomalies with environment data task (WP303).

1.3 Definitions, acronyms and abbreviations

ASCII	American standard code for information interchange
CDF	Common Data Format
CSV	Comma separated variables
EPS	energetic particle monitor
ESA	European Space Agency
GOES	Geosynchronous Orbiting Environment Satellite
IDL	Interactive Data Language. Commercial product with good mathematical and graphics functionality used as the scripting language in SEDAT.
keV	kilo electron-volt
MeV	Mega electron-volt
MLT	Magnetic local time
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
RAL	Rutherford Appleton Laboratory
SEDAT	Space Environment Database
SEM	Space Environment Monitor
SI	Système International, the international system of units.
SM	Solar magnetic (co-ordinate system)
TBC	To be confirmed
TBD	To be done

SEDAT	Doc. No: Issue: 1.3	RAL-SED-RP-0303 Date: 22/01/2004
Report on spacecraft anomalies analysis		Page 5

1.4 Important Documents

We list here the various documents used as source material for this report. These include both hardcopy and web sources. Documents may be referenced in the text and this is indicated by a sequential code of the form Xn, where n is an integer and X = A or R (for applicable and reference documents respectively). The series of integers are separate for applicable and reference documents.

1.4.1 Applicable documents

- A1 SEDAT Statement of Work. Appendix 1 to AO/1-3306/97/NL/NB
- A2 Space Environment Database and Analysis Tools. Proposal in response to ESA ITT AO/1-3306/97/NL/NB. RAL/RRS/201/97. January 1998.

1.4.2 Reference documents

- R1 Orbital Equations
http://liftoff.msfc.nasa.gov/academy/rocket_sci/orbmech/formulas.html
- R2 An introduction to space physics coordinate systems,
http://sspg1.bnsc.rl.ac.uk/Share/Coordinates/ct_home.htm
- R3 M.A. Hapgood, T.G. Dimbylow, D.C. Sutcliffe, P.A. Chaizy, P.S. Ferron, P.M. Hill and X.Y. Tiratay, (1997) "The Joint Science Operations Centre", *Space Sci. Rev.*, **79**, 487-525.
- R4 ISTEP/IACG Global Attributes, http://spdf.gsfc.nasa.gov/istp_guide/gattributes.htm
- R5 ISTEP/IACG Variable Attributes, http://spdf.gsfc.nasa.gov/istp_guide/vattributes.htm
- R6 http://spdf.gsfc.nasa.gov/istp_guide/variables.htm#Epoch
- R7 Reference Document for CSDS CDF Implementation, DS-QMW-TN-0003
<http://www.space-plasma.qmw.ac.uk/DOC/DS-QMW-TN-0003.ps>
- R8 Correlation of spacecraft anomalies with environment data, RAL-SED-TN-0303

1.5 Acknowledgements

The funding of this work by ESA (contract 12854/98/NL/NB) is gratefully acknowledged as are the helpful comments and METEOSAT data from staff at ESTEC in particular Hugh Evans and Alain Hilgers. The METEOSAT data are used with the kind permission of Andrew Coates at MSSL and were supplied to us by ESTEC. The supply of GOES data by NOAA is also gratefully acknowledged.

SEDAT	Doc. No: Issue: 1.3	RAL-SED-RP-0303 Date: 22/01/2004
Report on spacecraft anomalies analysis		Page 6

2 Introduction and structure

This document presents the results of SEDAT work package 303 which is a demonstration of the application of SEDAT to perform a demonstration of correlation of spacecraft anomalies with environment data. There are several aspects of this application and these are reflected in the subsequent sections of the document.

Section 3 describes the datasets used in this analysis. These include the SEDAT system datasets that provide data on those aspects of the space environment that may be linked with spacecraft anomalies, e.g. energetic electron fluxes. But they also include the data on anomalies and the positions of the spacecraft on which the anomalies occurred. The ingestion of these latter datasets is an activity within WP303, so their ingestion into SEDAT is discussed.

Section 4 describes the correlation of anomalies with magnetic local time. This reveals the well-known effect that anomalies are much more frequent in the morning sector.

Section 5 describes the superposed epoch analysis in which we search for consistent patterns in space environment data taken around the time of anomalies (but focusing mainly on the period before an anomaly as required to look for casual links). The method of superposed epoch analysis is outlined and then its implementation within SEDAT is described. The analysis is applied both to data taken on the spacecraft on which the anomalies occurred and to data from other distant spacecraft. The results reveal both a broad general increase in electron fluxes (seen on all spacecraft) and a local increase just before the anomaly (seen only on the anomaly spacecraft). Finally to verify the analysis method we perform a superposed analysis using pseudo-random sampling rather than sampling around the time (or epoch) of anomalies. We show that the variations in the results of this control test are consistent with random noise.

Section 6 presents overall conclusions to the report including some caveats and ideas for follow-up work. This is followed by two annexes: Annex A gives a summary of the SEDAT tools used in this WP is given in an annex at the end of the report, while Annex B presents a table showing how the sections of this report relate to the demonstration plan [R8].

SEDAT	Doc. No: Issue: 1.3	RAL-SED-RP-0303 Date: 22/01/2004
Report on spacecraft anomalies analysis		Page 7

3 Datasets used in the demonstration

3.1 Anomaly data

The anomaly data used in this demonstration are those from the Meteosat 3 spacecraft in geosynchronous orbit. They consist of a time (accurate to 1 second) and a short anomaly description. They were made available by ESA in ASCII format and were converted to CDF format for ease of ingestion into SEDAT.

The conversion to CDF was done in two stages.

1. The data were imported into an Excel spreadsheet. The Excel functionality was then used to reformat the time to give separate fields for year, month, day, hour, minutes and seconds. These fields together with the event id were then saved in comma-separated values (CSV) file
2. The CSV file was converted to CDF format using a standalone IDL program. This program first created the empty CDF file with appropriate global and variable attributes [R4, R5] as shown in Table 1 below. The program then read each record of the CSV file, converted the time fields to a date in CDF_EPOCH format [R6] and then wrote the event id and the time to the CDF file.

Table 1. CDF attributes for the anomaly data

Global attributes	Variable attributes
Acknowledgement	Fieldnam
Data_version	Fillval
Generated_by	Lablaxis
Generation_date	SI_conversion
Title	Units

The CDF file was then ingested into SEDAT as a system dataset (SYSTEM!METEOSAT_ANOMALIES).

3.2 Meteosat position

The position of Meteosat-3 during its lifetime was provided by ESA in the form of a table as shown in Table 2 below. To use these data within the demonstration (e.g. to calculate magnetic local time at each anomaly) it is necessary to interpolate these data. This was done by writing a SEDAT tool (mah!meteosat_position) that could be called with any arbitrary date as its argument and returns the position of the spacecraft as X, Y and Z in geocentric co-ordinates.

The data in the table are hardcoded into the tool (an efficient solution as the table is small). Given the date of interest, the tool identifies which is the appropriate record in the table and takes the geocentric longitude from the position field in that record. The X, Y and Z components are then calculated assuming that the geocentric latitude is zero and the geocentric distance is 42164.0 km (as appropriate to the geosynchronous orbit of Meteosat) [R1]. We ignore small deviations from this ideal, e.g. the inclined orbit noted in the last record of Table 2. If the date lies outside the range of the data, the function indicates there is no valid position data by returning a null value of -9.9e+6 km in each of X, Y and Z.

SEDAT	Doc. No: Issue: 1.3	RAL-SED-RP-0303 Date: 22/01/2004
Report on spacecraft anomalies analysis		Page 8

Table 2. Meteosat-3 position data. The transition between different positions is assumed to be rapid and ignored in this analysis.

From	To	Position
Launch 15 June 1988	Jun-89	0°E
Jun-89	Jan-90	50°W
Jan-90	Apr-90	0°E
Apr-90	Nov-90	5°W
Jul-91		50°W
late 1992		75°W
Apr-93		72.8°W
Feb-95	Nov-95	70°W, inclined

The mah!meteosat_position tool returns position data using the standard position structure that has been developed for SEDAT applications. This has the form: position = {sys: ", form: ", null: 0.0, si_conversion: ", dat: fltarr(3) } where the tags as specified in Table 3 below.

Table 3. SEDAT position data structure

Sys	String indicating the co-ordinate system of the position data, e.g. 'GEOC' for geocentric	GEOC
Form	String indicating whether the position data are expressed in cartesian ('CA') or polar ('PO') form.	CA
Null	Value that represents a null value in the dat field	-9.9e+6
Si_conversion	SI conversion string [R7]	1.0e+3>m
dat	Three values that specify a position	

3.3 Meteosat data

The primary dataset used in this demonstration is the medium energy electron measurements made by the SEM instrument on Meteosat itself. This has proved to be an excellent dataset because it gives a record of the total environment experienced by Meteosat. Thus we can distinguish large-scale effects (which may also be seen by other spacecraft) from effects local to Meteosat (which will not be seen by other spacecraft).

The Meteosat data provide differential electron fluxes at 5 energies between 40 and 200 keV. We carry out superposed epoch analyses on all five channels.

3.4 GOES data

Another dataset used in this demonstration are the high energy (>2 MeV) electron fluxes measured by the EPS instrument on the GOES spacecraft, which are also in geosynchronous orbit (though at different longitudes to Meteosat). Thus these data provide a different view of a similar particle population to that observed by Meteosat.

4 MLT analysis

The first step in the anomaly analysis is to examine the distribution of anomalies with respect to magnetic local time. This has been implemented as a single tool (`mah!plot_meteosat_anomalies`) that reads the anomaly dataset and for each case:

- calls `mah!meteaset_position` tool to obtain the geocentric position of the spacecraft
- converts this position to solar magnetic coordinates [R2] using the `SYSTEM!convcoord` system tool
- calculates the magnetic local time in hours using $MLT = 12.0 + \phi * 12.0 / 3.14159$ where ϕ is the spacecraft longitude in SM coordinates [R3] (the value of MLT must be reduced modulo 24 to obtain a value between 0 and 24).

The resulting set of magnetic local times is then analysed to obtain the required distribution which is plotted as a histogram as shown in Figure 1. This shows the well-known tendency for anomalies to occur in the early morning sector.

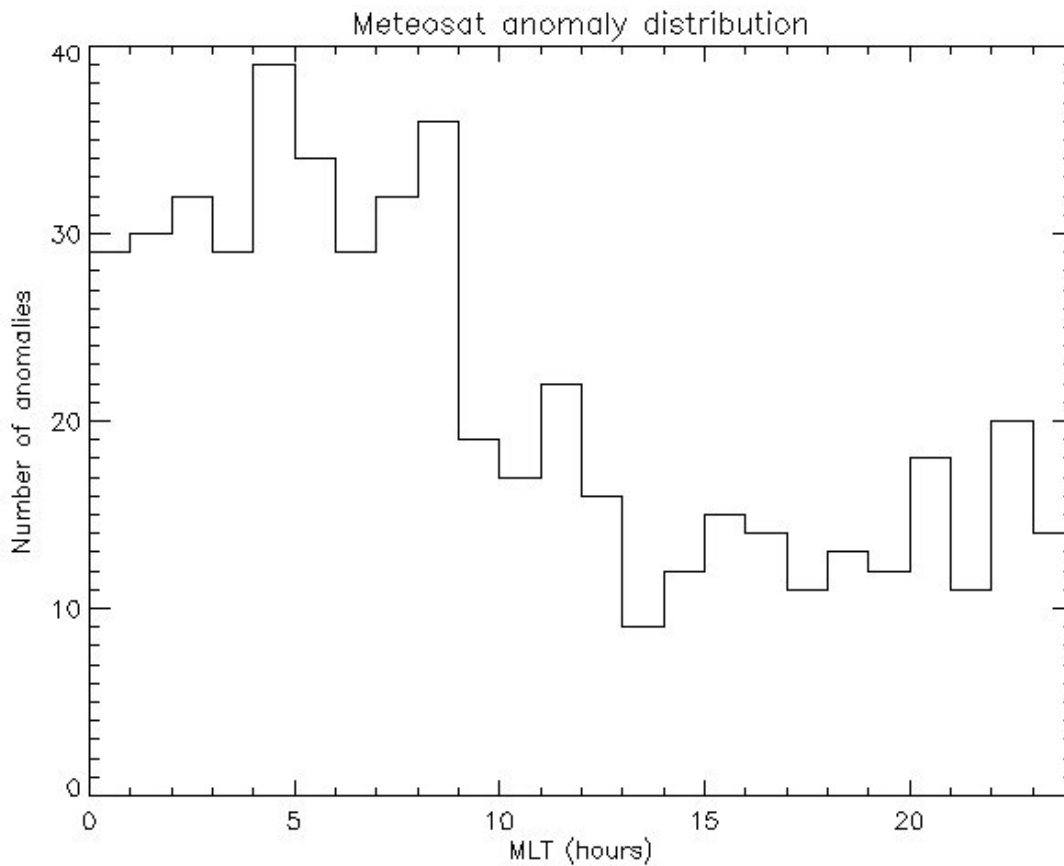


Figure 1. Meteosat anomaly distribution by magnetic local time

5 Superposed epoch analysis

5.1 Introduction

Superposed epoch analysis is a technique to search data for consistent patterns around events and thus establish a correlation between that pattern and the event. Such correlations can provide a evidence for a casual relationship between the events and the parameters in which patterns are found.

The technique is illustrated in Figure 2 below. We plot the parameter of interest as a function of time from each event. The time series for individual events (as shown by the black and light blue curves) does not show any marked event. But if we superpose the time series for a large number of events and calculate average values in a set of time bins we obtain a clear pattern (as shown by the red curve). This example was constructed using the function $0.2 \cdot \exp(-(t+5)^2/10) + r$ where r is a psuedo-random number between 0 and 1. The randomness dominates in the individual curves but the signal is clearly visible in the red curve, which is the mean of 100 individual curves.

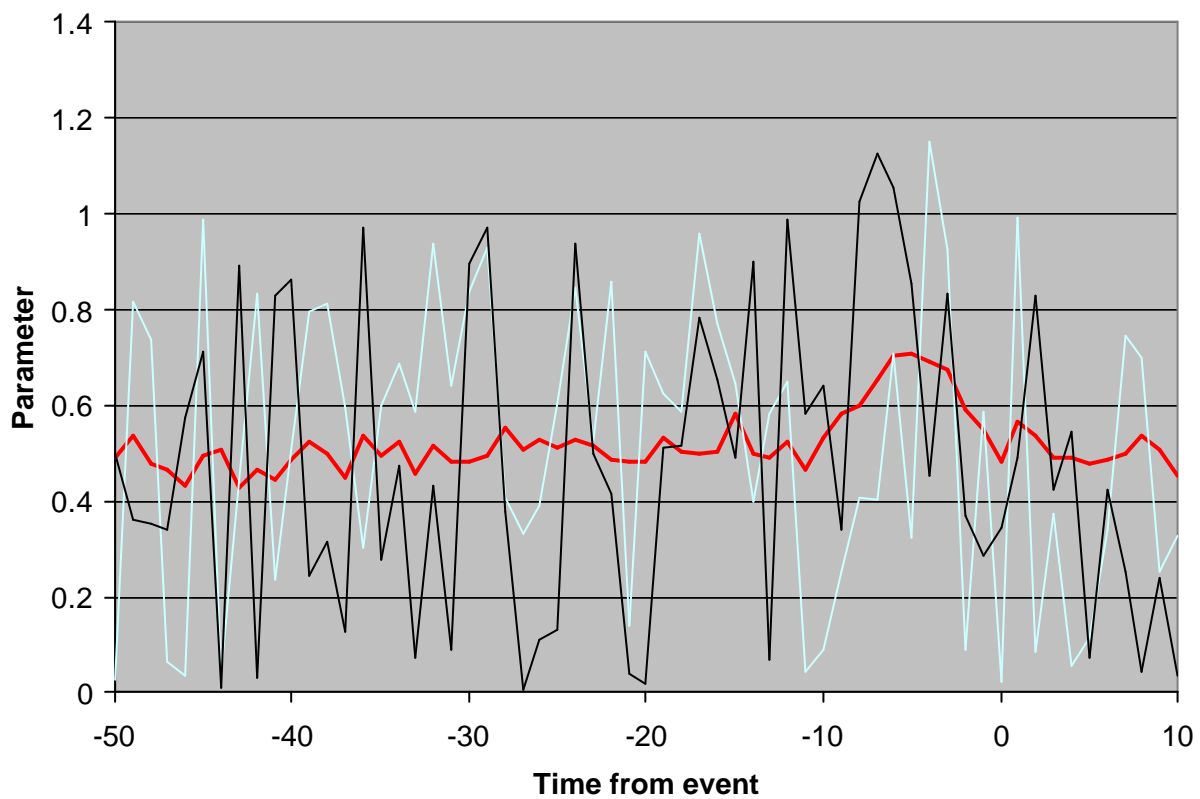


Figure 2. Principle of superposed epoch analysis

In the present case we will apply superposed epoch analysis to look for a correlation between spacecraft anomalies and enhanced energetic electron fluxes. Thus the event times (sometimes called the key times) will be the times of the anomalies. The parameters to be superposed will be measurements of electron fluxes in various energy channels; these must be retrieved from their source databases and time-tagged relative to events as discussed in the next section.

SEDAT	Doc. No: Issue: 1.3	RAL-SED-RP-0303 Date: 22/01/2004
Report on spacecraft anomalies analysis		Page 11

5.2 Time-tag tool

As the first step in the superposed epoch analysis we process the source dataset to extract the source data records close to each event time and write those records to an intermediate data file but tagged with the offset time from the relevant event. This file can then be used to perform a superposed epoch analysis on any field in the dataset. This approach has the advantage of reducing overall execution time; the data selection, which is time-consuming, is done only once and not repeated every time we analyse a different field.

The tool to carry out this task (`mah!evt_win_sel_drl`) requires that the user specify the time window, around each event; for which data records shall be selected and written to the intermediate data file. The window boundaries are specified separately for times before the event and times after the event. Thus it is possible to direct the superposed epoch analysis to focus on fluctuations that occur before events – as is essential if looking for a casual relationship.

The tool operates in a two-stage process

1. It reads the times of each event and then performs a binary search of the source dataset times to determine the indices of the source dataset records that cover the required time window around each event. These indices are stored in an internal data structure for use in the second stage; this also records the event time and reference number (a sequential integer). If no records are found in the required window, the event is skipped and no entry is made in the internal data structure.
2. It retrieves the source dataset records in the time window around each event (using the indices previously determined), prepends each record with the event reference number and time and with the source dataset identifier, and writes the resulting record to an intermediate data file.

Note that this tool does not require any detailed knowledge of the source dataset, or its metadata, other than the name of the time-tag field (and even this would not be required if the usual field name of epoch is available). SEDAT retrieves the source dataset records and metadata as IDL data structures, so the three addition fields can be prepended to data records by extending the data structure thus:

```
new_structure = { {event_number: 0L, offset_time: 0.0, data_source: 0}, old_structure }
```

A similar approach applies to metadata: the metadata structures for the three additional fields should be prepended to the var structure from the source dataset metadata and the number of fields value should be incremented by three.

SEDAT	Doc. No: Issue: 1.3	RAL-SED-RP-0303 Date: 22/01/2004
Report on spacecraft anomalies analysis		Page 12

5.3 Averaging tool

The next step, and the core of the superposed epoch analysis, is to process the intermediate file to derive average values of the parameter of interest. These averages are calculated in a series of bins covering the time window around each event. The bin size is selected by the user and is an input parameter for the tool. This binning approach allows the tool (mah!spe_analysis2) to handle any arbitrary distribution of offset times and thus it can be applied to any combination of data sampling and event times.

Some parameters of interest are vectors, e.g. particle fluxes at several different energies. To facilitate analysis of such data, the tool can perform superposed epoch analysis in parallel on some or all element of a vector. To support this, the tool includes several important functions

- To check if the field name supplied by the user is a scalar or a vector parameter (by examination of the metadata associated with the intermediate file)
- To allow the user to restrict analysis to a subset of the elements of a vector – and check that this is valid subset (again by examination of metadata)
- To manipulate values of the selected parameter as a vector or a scalar as appropriate

To perform this averaging tool:

- reads each record in the intermediate file
- extracts the offset time and assigns this to a time bin
- extracts the parameter of interest and updates the running values of the count (n), mean (M) and variance (V) for that bin. The running mean and variance are derived as follows, where subscript n indicates the values after n counts and X is the current, (n+1)th, value of the parameter:

$$M_0=0 \qquad V_0=0$$

$$M_{n+1}=(n*M_n+X)/(n+1) \qquad V_{n+1}=n*V_{n-1}/(n+1) + n*((X- M_n)/(n+1))^2$$
- if the parameter is a vector, the running values are updated separately for each element of the vector
- repeats the process for each record

After processing all records in the intermediate file, the tool writes a new result file in which each record corresponds to a specific offset time. The main fields are the offset time plus the final values of the count, mean and variance for the parameter (these are vectors if the parameter is a vector). In addition each record is tagged with a data source code and a value that characterises the channel in which the parameter is measured, e.g. the energy at which particle measurements are made. If the parameter is a vector, the channel value is a matching vector.

This new result file also contains as full a set as possible of the metadata associated with the parameter and the channel value. These metadata are vital for later processing, e.g. plotting the result file.

5.4 Plotting tool

This tool (mah!Qspe_plot_vector) allows us to plot the results of the superposed epoch analysis. We plot the mean value of the selected parameter as a function of offset time from the events; we also show the standard error on each mean. The plot is annotated to show various useful items of metadata: the source dataset, the energy channel plotted and names and units of the plotted parameters. It also shows the name of the intermediate data file that was input to the plot tool and the date and time at which the plot was produced. Finally we show a typical value for the number of independent measurements that contribute to each mean value; the median is used as most bins contain a similar number of measurements but a few bins contain very different numbers of measurements.

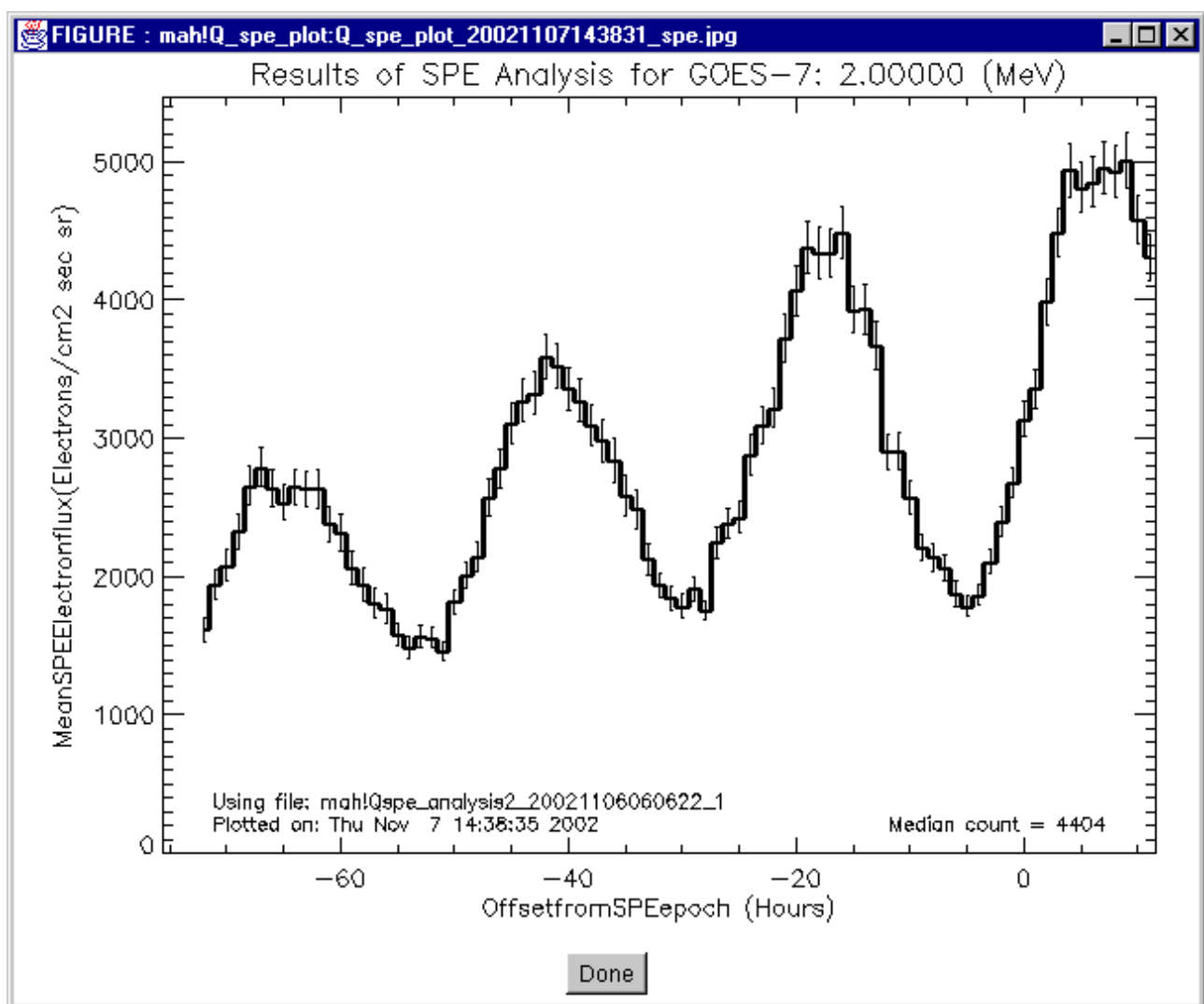


Figure 3. Superposed epoch analysis of GOES electron fluxes

An example is shown in Figure 3 above. The thick line in this figure represents the mean 2 MeV electron fluxes measured by GOES-7, while the thin error bars indicate the standard error on each mean value. The thick line is plotted in histogram style in order to make clear the size of the bins used in the superposed epoch analysis. This plot shows a clear diurnal pattern in the fluxes together with a general rising trend in the 72 hours prior to the event time. The phase of the peaks in the

SEDAT	Doc. No:	RAL-SED-RP-0303
	Issue: 1.3	Date: 22/01/2004
Report on spacecraft anomalies analysis		Page 14

diurnal pattern lags the event time by a few hours. This probably reflects the difference in longitude between Meteosat-3 and GOES-7.

Figure 3 shows a single channel so this plot style is appropriate if we wish to examine the superposed epoch analysis for a scalar parameter or for a single channel of a vector parameter. But in many cases we want to examine (and compare) the different channels of a vector parameter (e.g. from measurements of a particle energy spectrum). To handle this case we have developed a modified plotting tool that produces a multi-panel plot with a panel for each channel of the vector parameter. Each panel then shows the results of the superposed epoch analysis for its channel. The same metadata is displayed but those items that apply to the whole plot (source dataset, name of the intermediate data file, date and time at which the plot was produced, typical value for the number of independent measurements behind each mean) are shown once at the bottom of the plot. Those that apply to particular channels (energy channel) are shown on the relevant panel. We also show the names and units of the plotted parameters on each panel.

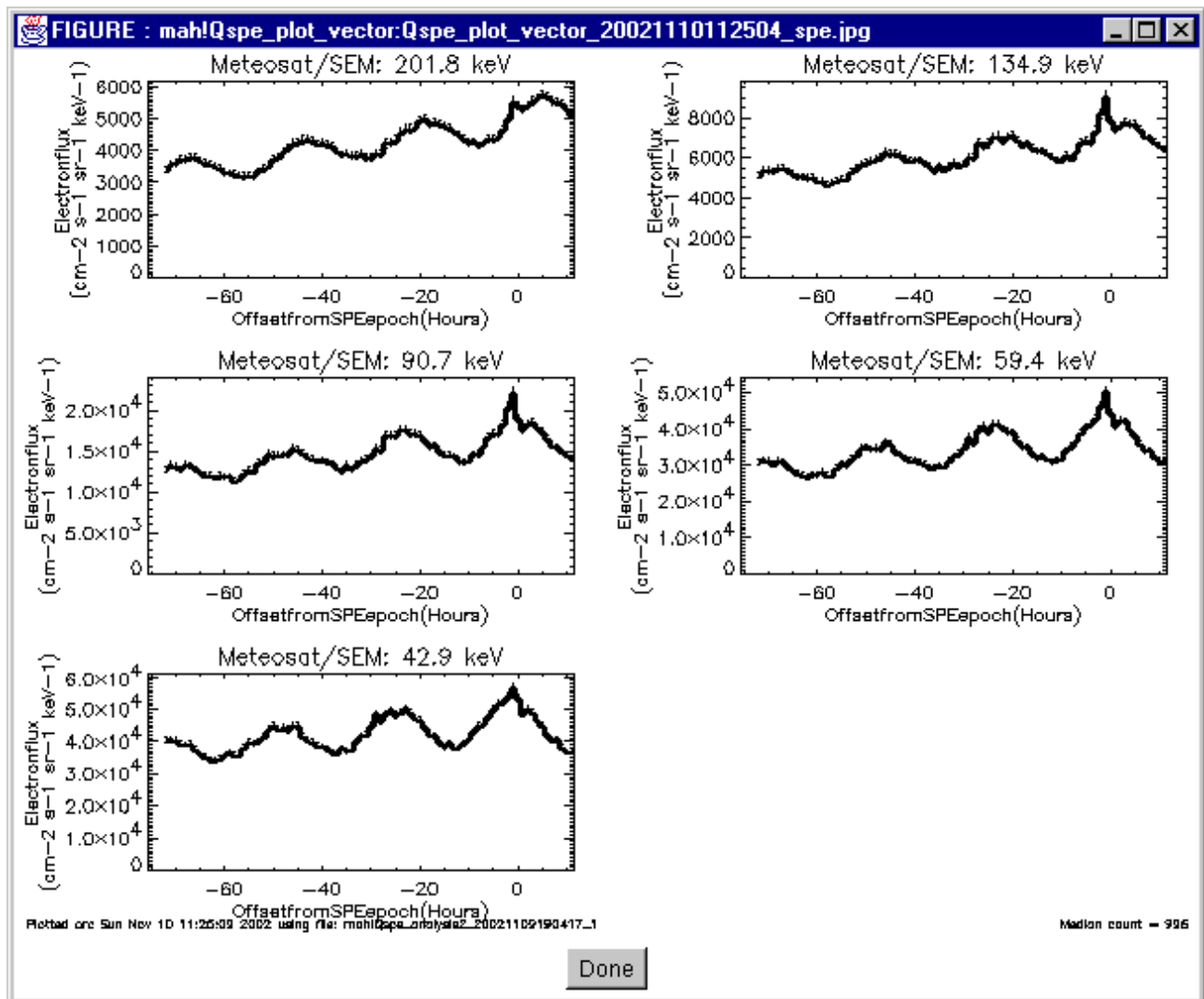


Figure 4. Superposed epoch analysis of Meteosat electron fluxes

An example is shown in Figure 4 below. As before the thick line represents the mean electron fluxes, while the thin error bars indicate the standard error on each mean value. Note that the energy

SEDAT	Doc. No: Issue: 1.3	RAL-SED-RP-0303 Date: 22/01/2004
Report on spacecraft anomalies analysis		Page 15

channels are plotted in descending order of energy. This reflects their order in the underlying dataset.

This plot also shows the clear diurnal pattern in the fluxes together with a general rising trend in the 72 hours prior to the event time. But in this case the phase of the peaks in the diurnal pattern closely matches the event time – except perhaps in the highest energy channel. This is better alignment of the peak with the events is to be expected if the electron fluxes are responsible for the anomaly events. However, these plots also show a sharp peak in the fluxes one or two hours before the event time; this is visible in all energy channels but most clearly in the three intermediate energy channels. This suggests that a localised increase in electron fluxes may also play a role in triggering anomalies.

SEDAT	Doc. No:	RAL-SED-RP-0303
	Issue: 1.3	Date: 22/01/2004
Report on spacecraft anomalies analysis		Page 16

5.5 Control test - random data selection

To provide a control test of the superposed epoch analysis, we have also written a modified version of the time-tag tool (mah!random_select) that simulates events at pseudo-random intervals, extracts the source data records close to each random event and write those records to an intermediate data file in the same form as the time-tag tool. The output can then be processed by the averaging and plotting tools in the same manner as data from real events. The resulting plot should be close to a horizontal line.

The number of simulated events N is controlled by setting the average separation of events to be $\langle \Delta t \rangle = T/(N+1)$ where T is the duration of the dataset. The separation between any pair of simulated events is $\Delta t = \langle \Delta t \rangle (r+0.5)$ where r is a pseudo-random number between 0 and 1. The first event is offset from the start of the dataset by a similar interval.

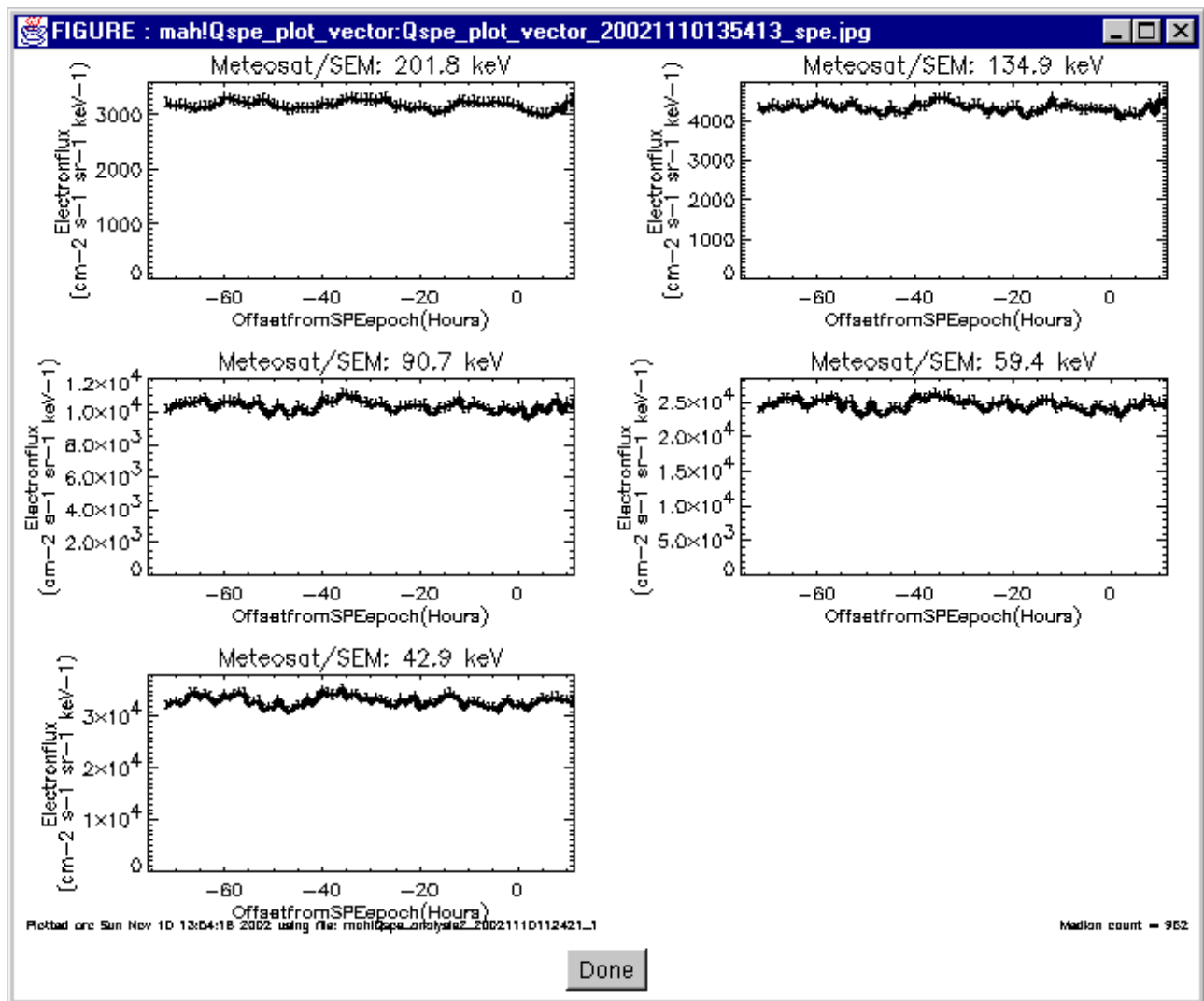


Figure 5. Random superposition of Meteosat data

An example of this control analysis is shown in Figure 5 above. This shows a random superposition of Meteosat data and should be compared with the superposed epoch analysis of the same data as

SEDAT	Doc. No:	RAL-SED-RP-0303
	Issue: 1.3	Date: 22/01/2004
Report on spacecraft anomalies analysis		Page 17

shown in Figure 4. To test that the small deviations from flatness in Figure 5 are not significant we have calculated the mean (M) and standard deviation (SD) of the curves shown in this figure. We also calculate SE, the mean value of the standard errors plotted on these curves. The results are given in Table 4 below. The values of SD and SE for each channel are very similar. Thus we can conclude that the small deviations shown in Figure 5 are due to random sampling and have no physical significance.

Table 4. Statistical analysis of data shown in Figure 5

Energy	201.8	134.9	90.7	59.4	42.9
M	3193.925	4353.059	10426.36	24744.25	33205.12
SD	75.21326	110.9456	307.5036	766.4125	936.459
SE	86.79038	112.8732	276.8418	647.0764	768.366

Note that small deviations in Figure 5 appear to be correlated between the different energy channels. This correlation is real but is simply an artifact of the analysis method. Each energy channel is sampled pseudo-randomly – but the same pseudo-random pattern is used for each channel. Thus the correlation is simply a consequence of the correlation between the different energy channels in the underlying data.

Note that the common pseudo-random sampling of all channels occurs because the sampling tools work at the level of complete data records rather than individual fields. This facilitates the analysis of vector fields but leads the correlation reported above. The separate sampling of individual vector elements would greatly increase the complexity of tool and its operation. It would also degrade execution speed.

SEDAT	Doc. No: Issue: 1.3	RAL-SED-RP-0303 Date: 22/01/2004
Report on spacecraft anomalies analysis		Page 18

6 Conclusions

This report describes the development of SEDAT tools for analysing spacecraft anomaly data supplied by ESA.

The first step was to write a simple tool to determine the distribution of anomalies in magnetic local time. This showed a strong bias towards the morning sector as expected for anomalies on a geosynchronous spacecraft.

The main part of this demonstration was the development of a set of tools to perform superposed epoch analysis with respect to anomaly times. The target dataset for this analysis would typically be an energetic electron dataset, as these electrons are suspected to be a prime cause of the effects leading to anomalies. The tool set comprises three tools:

- One to scan through a target dataset and extract data that falls in a user-defined time window around each anomaly event. The output is an intermediate dataset that contains the whole of each extracted record prefixed by the offset time from the anomaly.
- The second tool scans through the intermediate dataset, bins each record according to the offset time and a user-defined binning scheme, selects a field specified by the user and then updates the running count, mean and variance of that field for the appropriate bin. The output is a results dataset that contains the count, mean and variance of the selected field at a series of offset times, i.e. a classic superposed epoch result. One special feature of the tool is that the field selected by the user may be a vector or scalar. If it is a vector, the results dataset contains a separate count, mean and variance for each element of the vector.
- The last tool plots the results dataset. It plots the mean value as a function of offset time; an error bar on each mean is also shown and is the standard error derived in the usual manner from the variance and count. If the field selected by the user is a vector it generates a separate plot from each element of the vector.

One important point to note is the separation of the data selection stage from averaging stage. This separation means that the data selection is only done once and then averaging can be applied separately to each field of interest. This is efficient because the data selection stage requires a lot of time to execute. However, it means that the whole of each selected input record must be placed in the intermediate dataset.

This demonstration also includes a control test. To do this we developed a second selection tool similar to that described above, but with the key times provided at random rather than by the anomaly dataset. This yields an intermediate dataset with identical structure to that above so that it can be processed by the averaging tool above and the results displayed by the same plotting tool. Thus we generate a result equivalent to the superposed epoch analysis by with data selected randomly. This is an excellent control mechanism to test the reality of signals found in the superposed epoch analysis – and as clearly demonstrated in this report.

6.1 Caveats

The superposed epoch analysis has one major limitation; it is a time-consuming process – both in terms of processor usage and data input/output. This is true for both the data selection and the averaging tools (the plotting tool runs quickly). This limitation has been partially mitigated by optimising the code for execution speed and running the tool on an adequately powerful computer.

SEDAT	Doc. No: Issue: 1.3	RAL-SED-RP-0303 Date: 22/01/2004
Report on spacecraft anomalies analysis		Page 19

However, it is also important to develop a suitable discipline when running these tools – namely, to run them in parallel with other activities. Those may be work activities (if the tool is run during the working day) or private activities (if run overnight or over the weekend). It is also important to monitor progress, e.g.

- To verify that start-up has executed correctly by monitoring the query closely during the first minute of execution. There is nothing worse than to discover several hours later that the query failed at start-up (e.g. due to a minor user error).
- To monitor the progress of the query from time to time. To do this it is important that the underlying tool should issue periodic reports, e.g. listing every 10000th input record. This will also pick up query problems or failures caused by resource limitations on multi-user systems. This may be facilitated if it is possible to login remotely to check query progress.

6.2 Future work

It would be very interesting to extend this work to fine time resolution datasets in order to better resolve the peak near zero offset time as shown in Figure 4. However, that will require very careful preparation of the input datasets to enable the analysis to run in a reasonable time and without hitting resource limits.

7 Annex A – summary of tools used in this demonstration

Figure 6 below shows the main top-level tools used in this workpackage (blue boxes) and the SEDAT datasets flowing between those tools (green boxes). This figure excludes the tool to generate the MLT distribution of anomalies (mah!plot_meteosat_anomalies) as this is a standalone item outside the main data flow. The full set of top-level tools and queries is listed in Table 6 together with a summary of the functionality that they provide. The parameters for these queries are listed in Table 6.

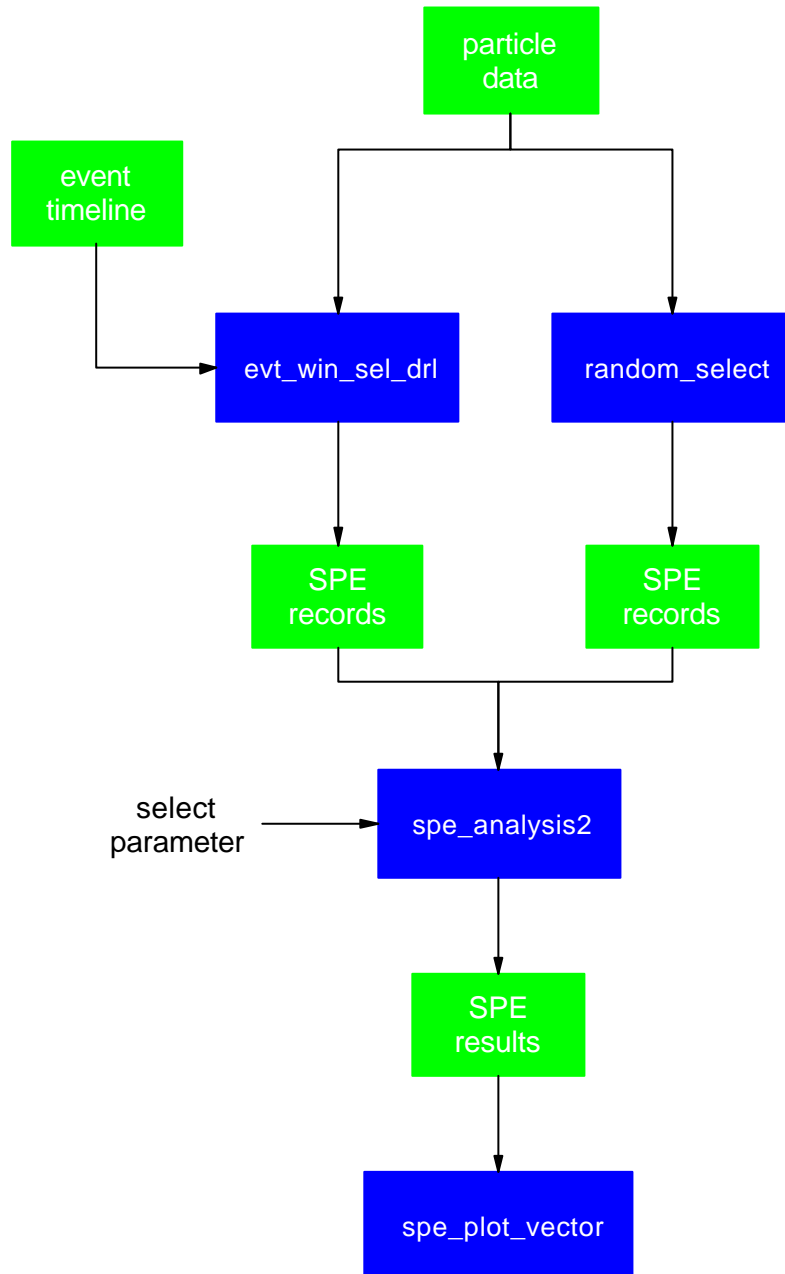


Figure 6. Main data flow for WP303.

SEDAT	Doc. No: Issue: 1.3	RAL-SED-RP-0303 Date: 22/01/2004
Report on spacecraft anomalies analysis		Page 21

The primary data inputs are the event data (top left of figure) that provide the key times for the superposed epoch analysis and the particle dataset (top centre) which will be the target of that analysis. This is then processed to derive a set of “superposed epoch” or SPE records. These are copies of those records in the original dataset that lie with a user-specified time window around each event. The SPE records are prepended with the number of the event in whose window they lie and the offset time from that event. In the event that an original record lies in the time windows around several events, it will be copied into a separate SPE record for each event.

The mah!spe_analysis2 tool is then used to bin the SPE records according to their offset time from the key epoch and a user-specified width of the time bins. In each bin it calculates the running mean and variance of a parameter specified by the user. This parameter may be a scalar, a vector or a subset of a vector. If the parameter has more than one element, the mean and variance are calculated separately for each element. On completion, the tool writes a dataset with one record for each bin in the analysis. Each record contains the offset time at the centre of the bin, the number of values placed in the bin plus the mean and variance of those values. If the parameter selected by the user had more than one element, the number of values, mean and variance are written out as vectors with length equal to the number of elements.

This output file can then be read and displayed by the spe_plot_vector tool. If the number of values, mean and variance are vectors, it produced a separate plot for each element. The plots show the mean value as a function of offset time. A standard error on that mean is also displayed and is derived from the variance and number of values as $\text{SQRT}(\text{variance}/\text{number-of-values})$.

To verify the superposed epoch results, we can also superpose randomly selected data. This done using the random_select tool. This generates an SPE records file with the same format as the file generated by the key period selection tool, but with its content randomly selected at random times throughout the particle dataset. We can then apply the spe_analysis2 and spe_plot_vector tools to generate a plot showing the effects of random superposition. This shows the level of variability arising from random effects in the data and thus allows us to verify whether or not any patterns in the superposed epoch results are statistically significant.

Table 5. Top-level tools for WP303 plus their queries

Tool	Query	Function
mah!plot_meteosat_anomalies	mah!Qplot_meteosat_anomalies	Calculate MLT distribution of anomalies
mah!evt_win_sel_drl	mah!Qevt_win_select	Select data around key times
mah!spe_analysis2	mah!Qspe_analysis2	Perform superposed analysis on selected data
mah!spe_plot_vector	mah!Qspe_plot_vector	Plot superposed analysis results
mah!random_select	mah!Qrandom_select3	Select data around random times

SEDAT	Doc. No: Issue: 1.3	RAL-SED-RP-0303 Date: 22/01/2004
Report on spacecraft anomalies analysis		Page 22

Table 6. Parameters used by WP303 queries

<i>Parameter</i>	<i>Description</i>	<i>Recommended values</i>
mah!Qplot_meteosat_anomalies		
event_data	List of anomaly times	SYSTEM! METEOSAT_ANOMALIES
mah!Qevt_win_select		
target_dataset:	Dataset to be subject to superposed epoch analysis	Any of METEOSAT_HR, METEOSAT_LR or GOES G-format datasets
event_dataset:	List of anomaly times to use as key times	SYSTEM! METEOSAT_ANOMALIES
lower_window	Offset from key time to start of selection window in units set by win_si_con	72
upper_window	Offset from key time to end of selection window in units set by win_si_con	12
win_si_con	SI conversion string to specify time units. Recommend string = 1 hour.	3.6e+3>s
data_instance	Code to identify data type	7 for GOES-7 data or 21 for Meteosat data
mah!Qspe_analysis2		
windowed_data	Windowed data records for superposed epoch analysis	Output dataset from mah!Qevt_win_select or mah!Qrandom_select3
spe_lower_window	Offset from key time to start of selection window in units set by spe_win_si_con	72
spe_upper_window	Offset from key time to end of selection window in units set by spe_win_si_con	12
spe_bin_size	Time bin for superposed epoch analysis in units set by spe_win_si_con	1
spe_win_si_con	SI conversion string to specify time units. Recommend string = 1 hour	3.6e+3>s
spe_target	Name of field to be subject to superposed epoch analysis	flux

SEDAT	Doc. No: Issue: 1.3	RAL-SED-RP-0303 Date: 22/01/2004
Report on spacecraft anomalies analysis		Page 23

<i>Parameter</i>	<i>Description</i>	<i>Recommended values</i>
mah!Qspe_plot_vector		
spe_results	Results of superposed epoch analysis	Output dataset from mah!Qspe_analysis2
mah!Qrandom_select3		
particle_data	Dataset to be subject to superposed epoch analysis	Any of METEOSAT_HR, METEOSAT_LR or GOES G-format datasets
lower_window	Offset from key time to start of selection window in units set by win_si_con	72
upper_window	Offset from key time to end of selection window in units set by win_si_con	12
win_si_con	SI conversion string to specify time units. Recommend string = 1 hour.	3.6e+3>s
n_windows	Number of random selections to be made from particle_data	500
data_instance	Code to identify data type	7 for GOES-7 data or 21 for Meteosat data

SEDAT	Doc. No: Issue: 1.3	RAL-SED-RP-0303 Date: 22/01/2004
Report on spacecraft anomalies analysis		Page 24

8 Annex B - Compliance matrix

The table below shows the compliance between this report and the demonstration plan [R8]. The item numbers relate to items in the procedure section (4.3) of that plan and compliance section numbers relate to the sections of this document.

Item from Plan	Description	Compliance in section
Section 4.3.1 Distribution in magnetic local time		
1	Anomaly dataset	3.1 and 3.2
2-4	MLT distribution	4
Section 4.3.2 Superposed epoch analysis		
1-4	Data selection around key times	5.2
5-7	Superposed epoch analysis	5.3
8-10	Plotting results	5.4
Section 4.3.3 Control results		
1-4	Data selection at random	5.5
5	Analysis and plotting	5.3 and 5.4