ESTEC/Contract No. 12854/98/NL/NB
Space Environment Database (SEDAT)
WP301. Solar protons model.

Report on the solar protons events model

RAL-SED-RP-0301
Issue 1.3, 22 December 2004

Authors: Mike Hapgood, CLRC Rutherford Appleton Laboratory

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

Table of contents

1 Preface

1.1 Document change record

Issue	Date	Notes/remarks
0.5	10 Oct 2002	Partial draft released to report progress to and seek advice from ESA
0.6	21 Nov 2002	Expand material on ingestion of Feynman catalogue (section 5.2) Add material on merging of event datasets (section 5.3) Update work on JPL-91 model (section 6.1) and on statistical analysis (section 7) to restrict analysis to solar maximum only. Update figures in these sections. Update figures in section 8 to show quantiles based on solar maximum years.
0.9	06 Jan 2003	Update ingestion of Feynman catalogue to include start and end times. Add survey of solar proton event datasets using the log-normal analysis
0.95	07 Jan 2003	Update quantile analysis to allow over-sampling if the sampling period is long (>2 years)
1.0	10 Feb 2003	Added acknowldgements in section 1 Note that IMP solar proton data is that from OMNI dataset Added more detail on event recognition in section 4.3 Added discussion of cross-calibration results at end of section 5.3 Updated Table 5 (now Table 6) to show mean fluence derivedif values of $\Phi n^{-1}(P)$ are calculated using a standard normal distribution. Added annex to summarise the SEDAT tools used in this workpackage
1.1	17 Feb 2003	Added conclusions section (9). Includes caveats and ideas for future work. Added compliance matrix (Annex B) between this document and the demo plan
1.2	11 Aug 2004	Add user names to names of SEDAT objects. Minor editorial changes. Update Annex A to give a better listing of the top-level tools, queries and their parameters.
1.3	22 Dec 2004	Complete Annex A by comparison to active server at RAL Add reference to WP302 tools, mah!get_fluences and mah!fluence_levels_max_only, used to make Figure 15.

1.2 Purpose of the document

This document is the technical note reporting the results of the application of SEDAT to the solar protons task (WP301).

1.3 Definitions, acronyms and abbreviations

ASCII	American Standard Code for Information Interchange – the standard for text-based computer files
CDF	Common Data Format – data format developed by NASA for space physics data
CDF Epoch	Time format developed for use in CDF. Time is represented as milliseconds since 0AD and stored as an 8-byte floating point number. IDL provides good support for CDF Epoch.
ESA	European Space Agency
ESTEC	European Space Technology Centre
GOES	Geosynchronous Orbiting Environment Satellite
IACG	Inter-Agency Co-ordination Group
IDL	Interactive Data Language. Commercial product with good mathematical and graphics functionality used as the scripting language in SEDAT.
IMP	Interplanetary Monitoring Platform
ISTP	International Solar Terrestrial Physics (programme)
JPL	Jet Propulsion Laboratory
MeV	Mega electron-volt
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
NSSDC	National Space Science Data Center
RAL	Rutherford Appleton Laboratory
SEDAT	Space Environment Database
SI	Système International, the international system of units.
TBD	To be done

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 test 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 Feynman, J., Spitale,G., Wang, J. And Gabriel, S. (1993) Interplanetary proton fluence model: JPL 1991, *J.Geophys.Res*. **98**, 13281-13294.

R2 Feynman, J., Armstrong, T.P., Dao-Gibner, L. and Silverman, S. (1990) New interplanetary proton fluence model, *J. Spacecraft and Rockets* **27**, 403-410.

R3 IDL exercises: image processing
http://www.astro.virginia.edu/class/oconnell/astr511/IDLexercises/IDL-Exercises-III.html

R4 Probability plotting.
http://www.public.iastate.edu/~wqmeeker/stat533stuff/psnups/chapter06_psnup.pdf
This is stated to be a summary of Chapter 6 of [R5] below.

R5 W.Q. Meeker & L. Escobar, Statistical Methods for Reliability Data John Wiley & Sons Inc; ISBN: 0471143286, 1998.

R6 Reference Document for CSDS CDF Implementation, DS-QMW-TN-0003
http://www.space-plasma.qmw.ac.uk/DOC/DS-QMW-TN-0003.ps

R7 ISTP/IACG Guidelines for CDF files
http://spdf.gsfc.nasa.gov/istp_guide/istp_guide.html

R8 ISTP/IACG Global Attributes, http://spdf.gsfc.nasa.gov/istp_guide/gattributes.htm

R9 ISTP/IACG Variable Attributes, http://spdf.gsfc.nasa.gov/istp_guide/vattributes.htm

R10 http://spdf.gsfc.nasa.gov/istp_guide/variables.htm#Epoch

R11 World Data Center for the Sunspot Index
http://sidc.oma.be/

R12 OMNIweb, Near-Earth Heliospheric Data
http://nssdc.gsfc.nasa.gov/omniweb/ow.html

R13 Demonstration of solar proton events model, RAL-SED-TN-0301

R14 www.lowell.edu/users/buie/idl/pro/slidefil.html

R15 SEDAT report on radiation environment analysis for the cruise phase of an interplanetary mission, RAL-SED-RP-0302

1.5 Acknowledgements

The funding of this work by ESA (contract 12854/98/NL/NB) is gratefully acknowledged as are the helpful comments from staff at ESTEC in particular Hugh Evans and Alain Hilgers. The supply of IMP data via the World Data Centre at RAL and of GOES data by NOAA are also gratefully acknowledged.

2 Introduction and structure

This document presents the results of SEDAT work package 301, which is a demonstration of the application of SEDAT to analyse solar proton events. There are several aspects of this application and these are reflected in the subsequent sections of the document.

Section 3 (Solar proton datasets) describes the solar proton flux datasets used in this work and outlines the techniques used to convert these to fluences, e.g. interpolation over data gaps.

Section 4 (Solar proton event detection) describes the work that was done to develop a tool to detect solar proton datasets. It first describes some ideas that were explored but eventually rejected as unsuitable (e.g. use of a sliding filter, use of digital filters). The reasons for those rejections are discussed. It then describes the two smoothing techniques that were adapted from image processing techniques and compares their ability to process solar proton data, which leads to the selection of median filtering as the preferred method because of its robust rejection of noise spikes. Finally this section describes event detection by application of a threshold to the smoothed proton data.

Section 5 (Manipulation of event files) describes the event files produced by the event detector. These have a well-defined record format and good metadata in order to support their subsequent manipulation and processing. The internal SEDAT data structure for these events records is also described. This section then discusses some of the tools developed to examine, edit and merge event files prior to processing by other tools. Finally we discuss the ingestion into SEDAT of historical proton event data (1956 to 1985) as supplied by ESTEC.

Section 6 (Using and updating the JPL-91 model) first discusses a key aspect of the JPL-91 model - namely the fitting of event fluence distributions to a log-normal form and the derivation of the mean and standard deviation of that form. The special display format used to linearise the log-normal distribution is discussed in detail and an example(s) of its application within SEDAT is presented. This second then discusses the implementation of the JPL-91 model within SEDAT. This follows the approach given in R1 and R2 (it proved useful to read both papers). The detailed implementation is discussed in some detail including the use of built-in IDL functions to generate normally-distributed pseudo-random numbers and the parallel calculation of probabilities for different fluence thresholds (following a suggestion from ESTEC). The result of the SEDAT implementation are compared with those in R1 and shown to be in good agreement.

Section 7 (Statistical sampling of the fluence distribution) discusses the direct generation of percentiles of the fluence distribution for comparison with the results of the JPL-91 model. In this section we discuss the pseudo-random sampling of a time series of accumulating fluence to derive a set of the different fluences that can accumulated over a given duration. This set can then be analysed to derive the percentiles of the fluence distribution. By plotting percentages against the equivalent percentiles we obtain a direct plot of probability versus fluence. Several examples of these plots are presented.

Section 8 (Comparison of percentile plots with the JPL-91 model) shows how we can compare the direct plot of probability versus fluence obtained in section 7 with the equivalent results of the JPL-91 model.

Section 9 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 [R13].

3 Solar proton datasets

3.1 GOES data

One of the key datasets used has been the series of solar proton measurements taken on board NOAA's GOES spacecraft. This typically yields two separate values at any time because there are usually two GOES spacecraft operational at any one time.

GOES data comes in a variety of formats. For analysis of solar proton events, the I format has been used since this provides integral fluxes above a range of thresholds, which is ideal for the required purpose.

3.2 IMP data

Data from the solar proton monitor on IMP-8 and distributed as part of NSSDC's OMNI dataset [R12] has also been used extensively. This has proved the best dataset for supporting development of the WP301 tools:

- It is a consistent dataset of long duration Jan 1973 to the present. This long duration facilitates testing by providing a long time series of fluxes and large numbers of solar proton events. Thus it is straightforward to carry out realistic load tests.
- It has a time resolution of an hour (compared with 5 minutes for GOES) so it is easier to process long periods of time than when using GOES data..
- The data format is simpler than that for GOES protons, which further improves execution time.

The bottom line is that it is much quicker to process the IMP data than the GOES data.

3.3 Construction of fluence data

Several WP301 tools require the derivation of fluence values by integration of flux values over time, i.e. $F(t) = \int_0^t f(\tau)\, d\tau$. where F(t) is the fluence at time t and f(τ) is the flux at time τ. This calculation requires interpolation over gaps of in the time series of flux data. This is needed to obtain a good estimate of the fluence. It also has the advantage that it produces a homogeneous time series that is well conditioned for further analysis such as filtering and event recognition.

We first considered the use of simple linear interpolation. However, it was quickly found that this gave spuriously high fluences if a large data gap overlapped part of a proton event. This is nicely illustrated in Figure 1, where a large data gap starts just after the peak of the proton event on day 4. Linear interpolation gives an event profile that decays too slowly and over-estimates the fluence. Thus we also considered a scheme for exponential interpolation in which the fluxes are assumed to decay or rise exponentially at the end of the gap with higher fluxes (i.e. decay if higher at the start of the gap, rise if higher at the end). An e-folding time of 2 hours was used. Figure 2 shows the effect of exponential interpolation on the previous data. The time profile of interpolated flux is better but, most importantly, the integrated fluence for this whole time series is reduced by a factor 2 or more.

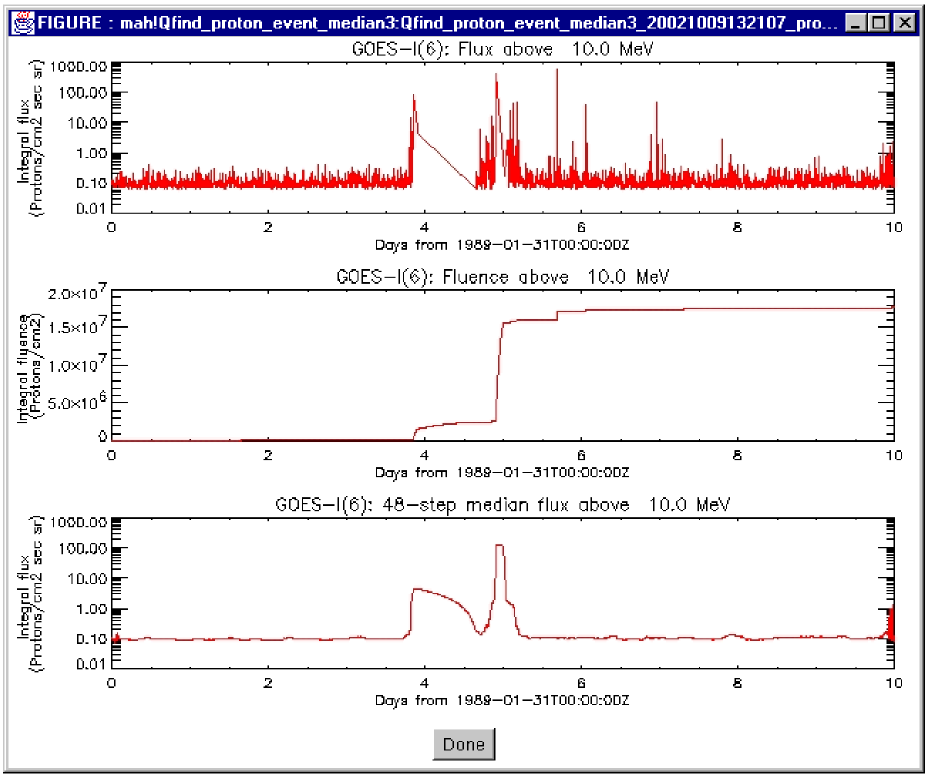

Figure 1. Linear interpolation of solar proton data.

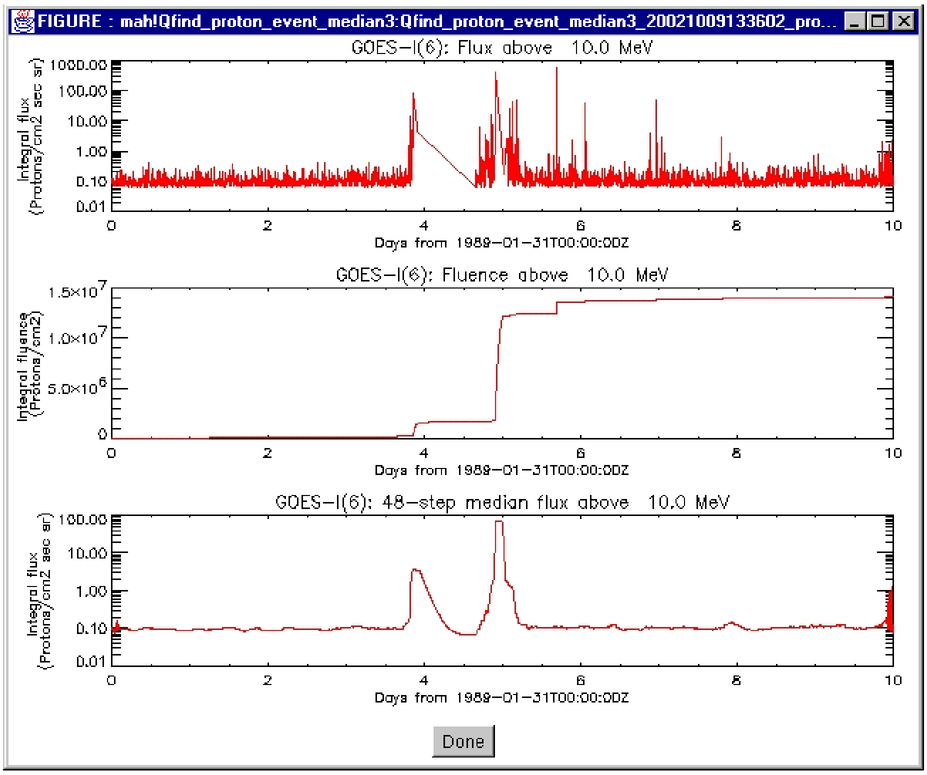

Figure 2. Exponential interpolation of solar proton data.

4 Solar proton event detection

4.1 Use of an astronomy technique

The first search tool to identify solar proton events was adapted from an astronomy IDL tool (slidefil = sliding spatial filter [R14]) designed to scan a data stream looking for short duration, non-random excursions. The basis of the tool was the use of two co-aligned time windows with markedly different sizes that can be scanned through the dataset. The larger window was used to estimate the average background signal at any time while the smaller window was used to estimate the local signal. It was then possible to apply standard statistical techniques to check if the local signal was significantly different from the background.

This tool was successfully used to search small segments of GOES proton data. For example, one year's worth of GOES proton data at 5 minutes resolution required about 9 hours processing, which could be done conveniently by overnight run. However, it was found that the execution time scaled as the square of the number of data points to be examined. This became a major inhibition in progressing work on testing and further development of the code. Some efforts were made to identify the causes of this scaling and to adapt the code so that execution time would scale linearly with the amount of data, e.g. by better exploitation of the time order of the data. This improved execution speed but was not able to eliminate the scaling problem. Thus it was decided to abandon this approach as unproductive and to develop a new tool based in better algorithms..

4.2 Use of a low-pass filter

One of the fundamental issues in building the event detection tool is the ability to distinguish real events from the spurious spikes that are sometimes found in solar proton data. One option to deal with this, as discussed earlier in the SEDAT work, is the use of low-pass filter to suppress the spikes. We then explored this using IDL's digital_filter routine but quickly came the conclusion that filtering is not suitable as a way of suppressing spikes. It inevitably leads to negative values in the filtered fluxes, which cause major problems in subsequent analysis – in particular, visualisation as this needs to use a logarithmic scale to display the range of measured fluxes. The negative fluxes reflect the fact that filters always includes strong negative terms in order to select negative phase of oscillations within the filter pass band. Thus we conclude that filtering is inappropriate as a means of processing particle fluxes. It causes more problems than it solves.

4.3 Use of image processing techniques

We then considered what techniques exist for manipulating detailed information in intrinsically positive quantities. This led us to look at image processing techniques as image data has many similarities to solar proton data – both are intrinsically positive with large dynamic range. Two image processing techniques were explored:

1. First it was noted that integration of flux time series reveals a clear pattern of fluence accumulation with sharp steps at solar proton events. Thus the issue was how to detect those steps. So we used a simple edge detection technique [R3] which is to take a stepped difference of the data with itself, i.e. we calculate $\delta F(t)=F(t+\delta t)-F(t)$, where $F(t)$ is the fluence at time t and δt is a fixed time difference. This worked well as a means of detecting proton events and significantly reduced the noise level compared with the original flux measurements.
2. Second, we used median smoothing which is a good way to identify systematic enhancements in images while suppressing noise spikes [R3]. We calculate the running median flux within a time window of width δt, i.e. $f_M(t) = \text{Median}(\{f(t); t-\delta t/2 \le t \le t+\delta t/2\})$, where $\{f(t); t-\delta t/2 \le t \le t+\delta t/2\}$ is the set of flux values within a window of width δt centred at time t. This technique also reduces the noise level compared with the original flux measurements but is much more robust against occasional extreme outliers, such as the spikes often found in GOES data.

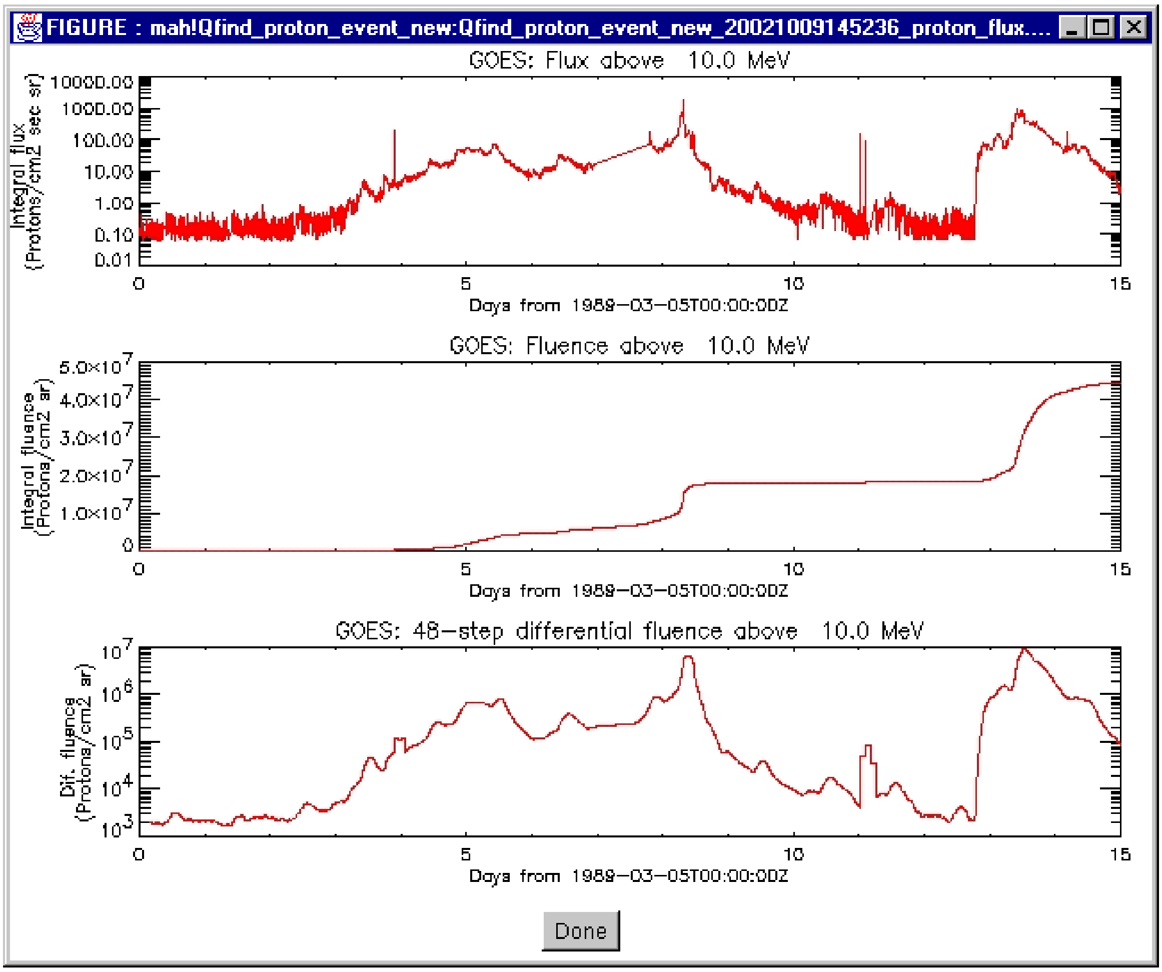

Figure 3. Smoothing of solar proton data by stepped differencing.

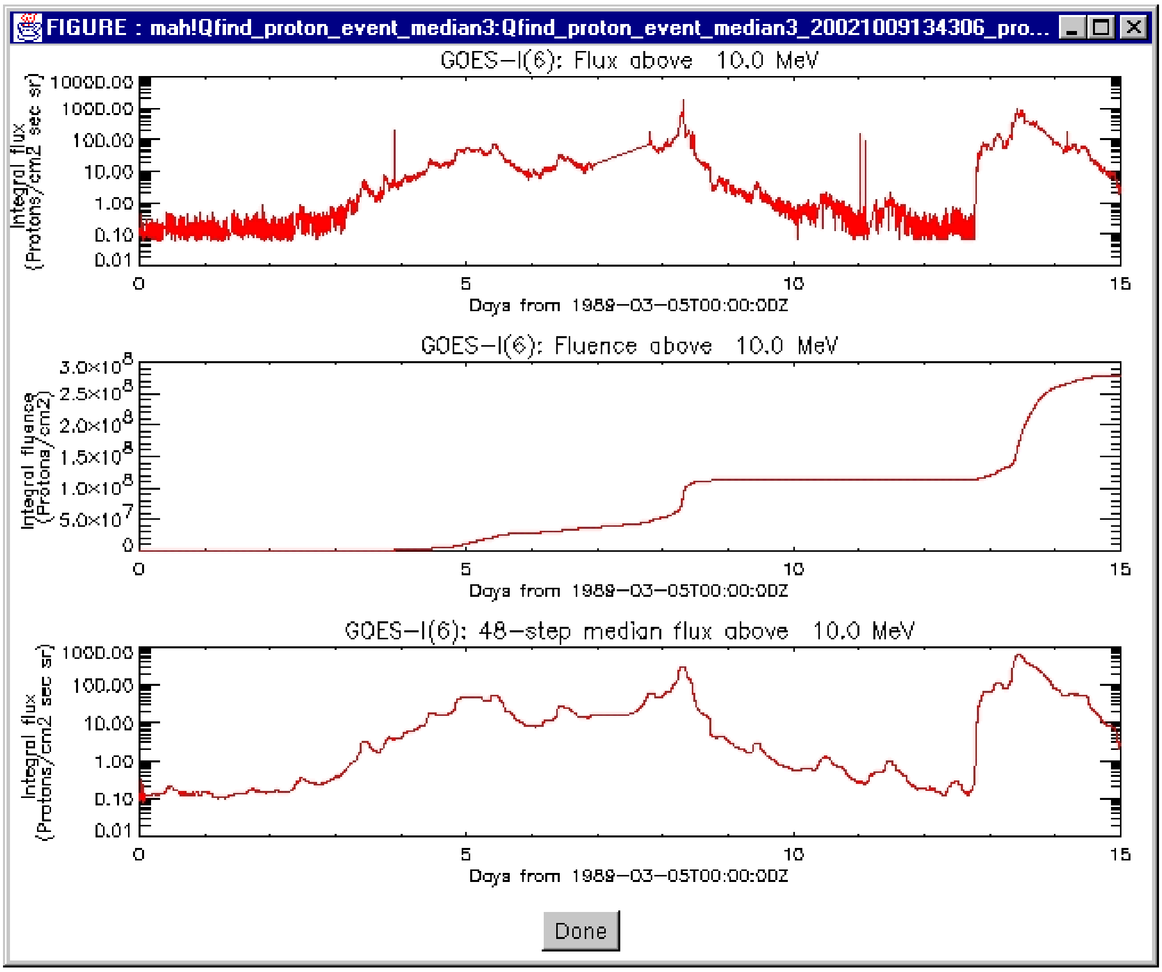

Figure 4. Smoothing of solar proton data using a running median.

The two techniques are illustrated in Figure 3 and Figure 4, which show their application to a short segment (15 days) of proton data taken by GOES 6 in March 1989. This segment was chosen because it contains a couple of real events and also two sets of spikes that we would like to suppress prior to event detection. Both techniques have good noise suppression – a noise modulation is very visible in the unprocessed data at low flux values (top panels) but is suppressed in the both sets of processed data (bottom panels). However, the median smoothing is clearly much better at suppressing spikes. The bottom panels show that the noise spikes at 4 and 11 days yield a signal if the differencing technique is used but not if median smoothing is used. Thus median smoothing has been adopted as the technique used to smooth the data and eliminate noise spikes.

To detect events in the smoothed data we proceed as follows:

- We search the smoothed data to find fluxes above the user-defined threshold
- We then retrieve the time-tags of these fluxes
- We calculate the time differences between these time-tags
- We then search for time differences that exceed a suitable limit. A limit of 4 hours is used here.
- These large differences are taken as the gaps between solar proton events and thus define the end of one event and the start of the next
- The start of the first event is the first time-tag with a flux above the threshold
- The end of the last event is the last time-tag with a flux above the threshold

This approach handles closely-spaced events in a robust manner. Close events are treated as one unless there is a significant time gap between events.

Given the start and end time of each event the tool then derives the duration, peak flux and fluence of each event. These are written to the log file and to a user dataset through which they can be made available to other tools. Each record in the user dataset is also tagged with information on data provenance:

- a data type that unambiguously identifies the source data format (e.g. GOES-I or IMP),
- a data instance code that identifies the particular source of that data type, e.g. to distinguish the same data type from different GOES spacecraft
- the energy threshold of the source dataset, e.g. if using the >10 MeV proton channel this would be 10 MeV.
- the flux threshold used to detect events within the tool

These fields are important metadata that is used by other tools, e.g. when using data derived from several sources we can check that we are processing a consistent dataset with the same energy and flux thresholds. We can also display these metadata as annotation on plots so that the provenance of the data is made known to the people viewing the plot.

The tool requires input of integral proton fluence data. At present it can read data in either the GOES-I format (as used for GOES integral fluence data) or in the IMP proton data format. The code has been structured to facilitate addition of new formats by confining format specific detail to two sections:

1. A section executed after the dataset has been opened. This handles dataset-specific details such as field names and metadata.
2. A section executed after each data record has been read. This handles dataset-specific details such as data encoding and null values.

The result of this work is a tool (mah!get_proton_data) which runs quite efficiently. It takes about 1 hour to process 1 year's worth of GOES data at 5 minutes time resolution. Thus it is straightforward to process a complete dataset from one GOES spacecraft by an overnight run. Processing of IMP proton is even better. It takes about 1 hour to process the complete dataset with results as shown in Figure 5.

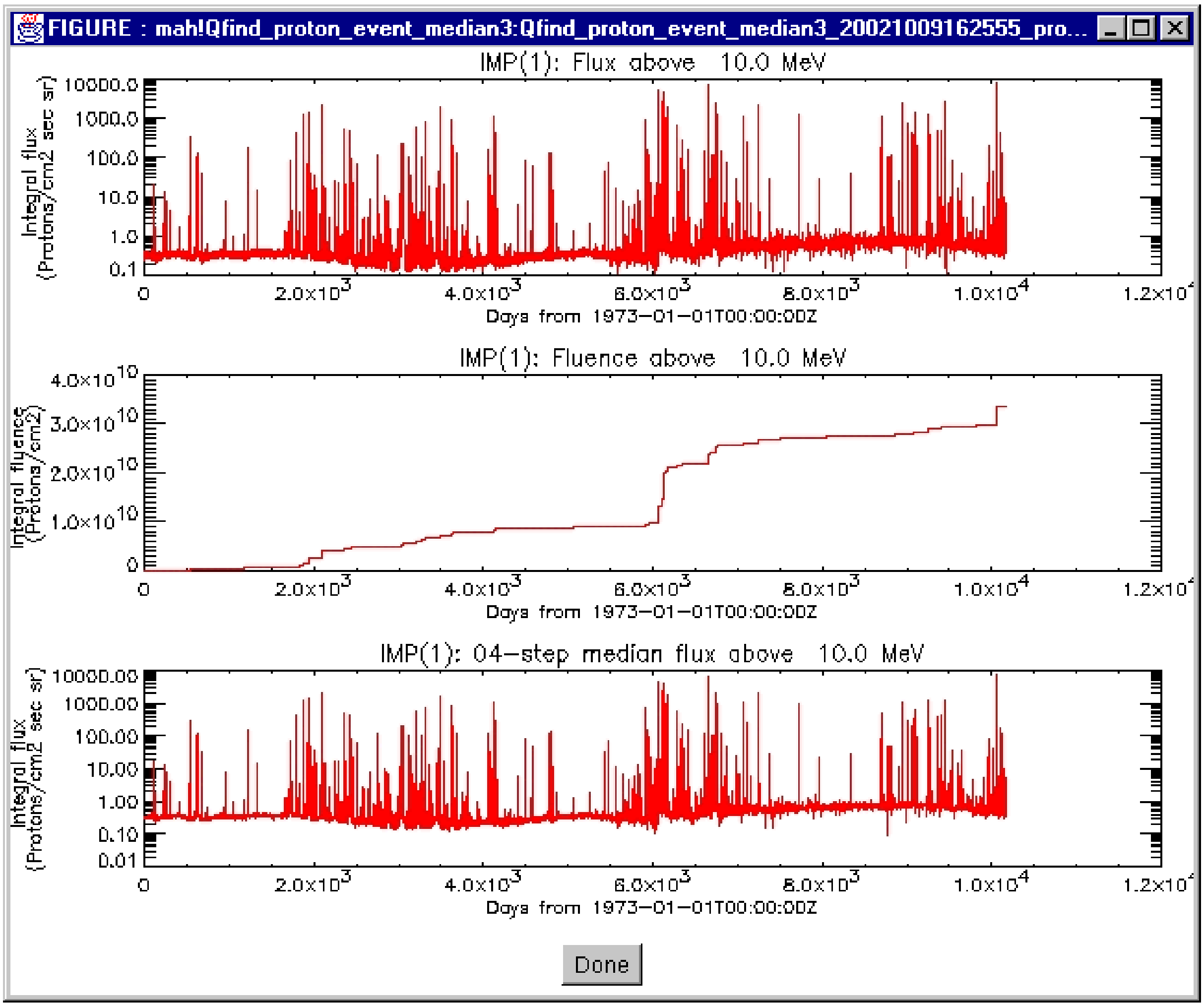

Figure 5. Fluence and solar proton events for the IMP dataset.

5 Manipulation of event files

5.1 Event data format

The event detector tool generates a set of events from a single proton flux dataset or a subset of a dataset. These are written out to event files (SEDAT user datasets) using the record format described in Table 1, together with a comprehensive set of metadata including:

- units and SI conversion strings for time, flux, fluence and energy fields,
- axis labels for all fields.

The SI conversion string is a concept originally developed by the Cluster Science Data System [R6] and now adopted more broadly, e.g. by the Inter-Agency Co-ordination Group [R7]. It is a string that can be parsed to derive the numeric relationship between the current units and the underlying SI units. See [R6] for a detailed description. It is widely used within SEDAT tools to ensure flexible handling of units.

Table 1. Record structure of events files

Field name	*IDL data type*	*Notes*
Start_time	Double	Coded in CDF Epoch format [R10]
end_time	Double	Coded in CDF Epoch format
Fluence	Double	Fluence is derived by integration of flux so use of type Double is required to ensure accuracy when integrating many thousands of values.
Peak_flux	Float	
Peak_time	Double	Coded in CDF Epoch format
Data_source	Long	Integer code as in Table 2.
Energy_level	Float	Lower bound of the energy channel used
Threshold_flux	Float	Threshold used to detect events

NB Float = 4 byte floating point number, Double = 8-byte floating point, Long = 4 byte integer

Table 2. Data sources

Data_source code	*Description of data source*
1	IMP proton fluxes
2	Feynman catalogue 1965-1985 [R2]
5	GOES-5 integral proton fluxes
6	GOES-6 integral proton fluxes
7	GOES-7 integral proton fluxes

The data format described in Table 1 is complemented by a standard data structure that is used to hold these data within SEDAT tools. This is illustrated in Table 3 in terms of the IDL code to build the data structure. A simple relational structure is used to describe each event record. This is then replicated into an array of record structures with sufficient elements to hold all the events and then inserted into a master event structure together with other relevant information such as the number of events.

Table 3. Event data structure

```
; build the event data structure
event_rec={start_time:   0.0D0, end_time:       0.0D0, $
           fluence:      0.0D0, peak_flux:      0.0,   $
           peak_time:    0.0D0, data_source:    0L,    $
           energy_level: 0.0,     threshold_flux: 0.0  }
events1 = {number: n_events, $
           events: replicate(event_rec,n_events) }
```

5.2 External data sources

The >10 MeV and the > 30 MeV solar proton events from the catalogue of Feynman et al [R2] have been loaded into the SEDAT database. The event fluences were taken from ASCII data files supplied by Alain Hilgers and the start and end dates of each event were taken directly from [R2]. These data were assembled in an Excel spreadsheet, then stored as a comma-separated value (CSV) file and finally converted to CDF format for ease of ingestion into SEDAT. The start and end dates were given in day-of-year format by Feynman but converted into year-month-day format in the Excel spreadsheet. The start time was taken to be 00^h00^m on the start date and the end time to be 23^h59^m on the end date.

The conversion to CDF was done using a standalone IDL program (TBD) run outside SEDAT. This program first created the empty CDF file with appropriate global and variable attributes [R7, R8] as shown in Table 4 below. The program then read each record of the CSV file, converted the start and end times (year, month, day, hour and minute) to CDF Epoch format and then wrote these times and the >10 MeV and >30 MeV fluences to the CDF file.

Table 4. CDF attributes for the early event 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 user dataset (mah!feynman65_85). These files contain only the start and end times and the fluences of each event (thus accurately reflecting their origin) but for wider use with WP301 tools they have been further processed within SEDAT to generate files having the event record structure described in Table 1. There is a separate file for each energy channel. The fields in the event record structure are derived as follows

- The start,and end times are taken from the ingested data file.
- The peak time is assumed to be one-third of the way through the event.

- The fluence is taken from the ingested data file.
- The peak flux is set to null.
- The data source is set to 2 (to represent the Feynman catalogues as the source of these data)
- The threshold flux is set 6.3 cm^{-2} s^{-1} (equivalent to the 1 cm^{-2} s^{-1} sr^{-1} given in [R1])
- The energy channel is 10 or 30 MeV as appropriate

5.3 Event data manipulation tools

To support further analysis of event data, we have developed a number of tools that allow the user to manipulate solar proton event datasets. These include:

- TBD- generating an ASCII listing of a dataset, e.g. to allow manual examination of the data or export for ingestion into other tools such as Excel.
- TBD- editing a dataset, e.g. to correct errors or remove dubious events.
- mah!merge_event_recs - merging two datasets into one; this includes the identification and elimination of duplicate events and support for cross-calibration.

The merging tool (mah!merge_event_recs) is important as this allows us to construct datasets covering periods longer than can be derived from any single source. However, merging requires more than a simple concatenation of datasets; it must also attempt to eliminate duplicate data and to cross-calibrate the fluences from the two sources.

To find duplicates the tool searches the second dataset. An event in those data is considered to be a duplicate if its peak time lies between the start and end times of an event in the first dataset. The tool excludes these duplicates from the data copied into the merged dataset.

The list of duplicates is also used to create the cross-calibration. For each duplicate we can plot the event fluence measured in dataset 1 against than measured in dataset 2. Figure 6 shows an example based on events derived from GOES-6 and IMP data. The event fluences cover four orders of magnitude and are therefore shown on a log-log plot. To find a simple relationship that can be used for cross-calibration we fit a straight line between the logarithms of the fluences, i.e. $\log F2 = a + b \log F1$, where F1 and F2 are the fluences on the x and the y axes. A least absolute deviation technique is used (the IDL LADFIT tool) as there is significant scatter around the general trend of the plot. This yields the blue line shown in Figure 6, which is a good measure of the common trend between the two sets of measurements.

To align the two datasets, we adjust the dataset with the lower values so that its values are increased to match the other dataset. This approach has been adopted as it identifies the maximum plausible risk. To determine which dataset has higher values we compare the maximum value of the dataset plotted on the x-axis with the y-axis value of the fitted line at that x-axis value. The comparison is made a high values as it is these fluences that most influence the risk arising from solar protons. If the adjustment is to be made to the x-axis dataset we transform values as $\log F1a = a + b \log F1$ and if the adjustment is to be made to y-axis dataset we transform values as $\log F2a = -(a/b) + (1/b) \log F2$.

To verify the adjustment we replot the fluences from the common events with the appropriate adjustment. Figure 7 shows an example, and it can be seen that there is now a good alignment between the datasets. Note however, that these common event data are only included once in the merged dataset.

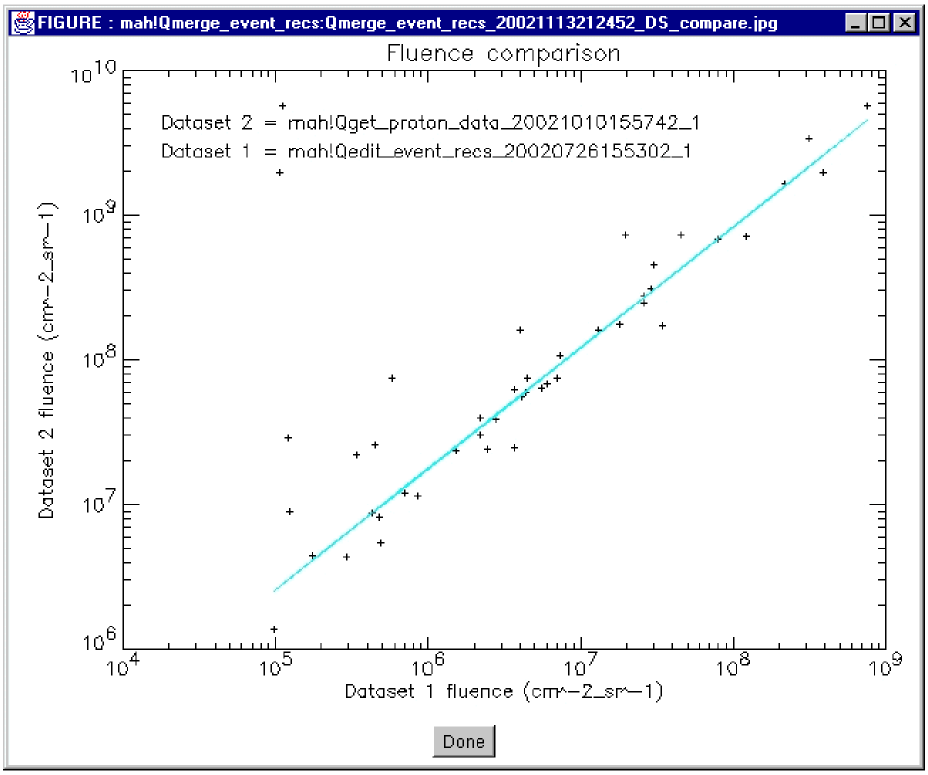

Figure 6. Comparison of fluences of common events derived from GOES6 (x-axis) and IMP (y-axis) data.

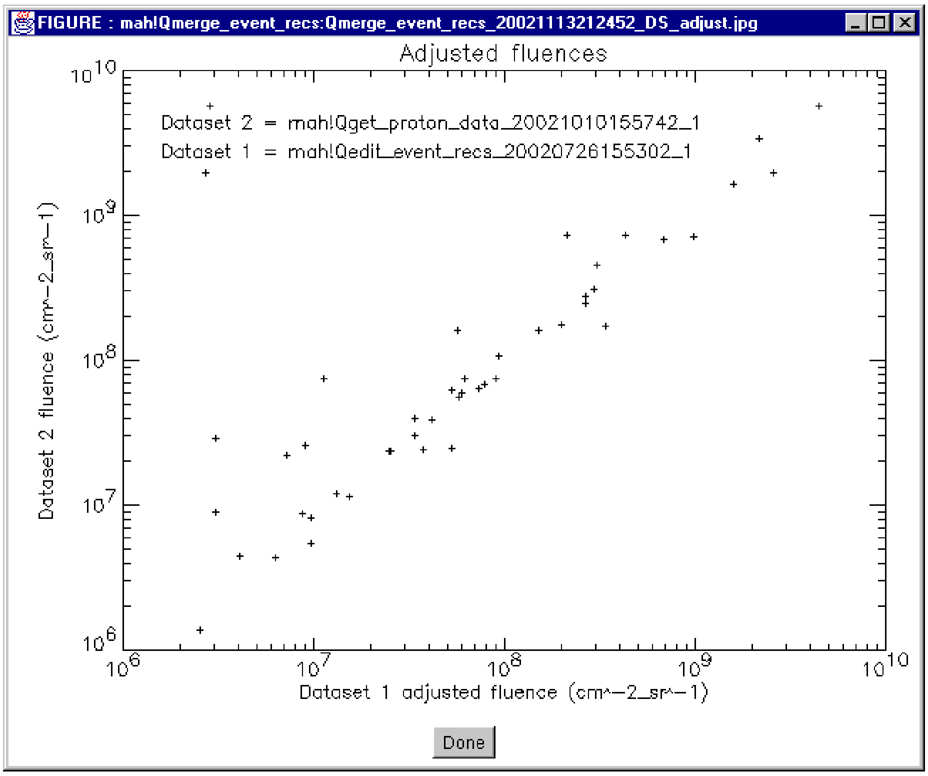

Figure 7. Alignment of the fluences shown in Figure 6. The GOES-6 fluences have been adjusted to match the higher measurements from IMP.

Table 5 below shows results of merging solar proton event datasets from a variety of data sources. On the left of the table we have the names of the two datasets, the numbers of events in each and the energy channel for which the events were derived (which must be the same for both datasets in order to have a valid merger). On the right of the table we have the number of events that the tool finds to be common to the two datasets and the intercept and slope of the fitted straight line (as shown in Figure 6).

Table 5. Cross-calibration results for solar proton event datasets

	Dataset 1		Dataset 2			Fit	
Energy	Name	Count	Name	Count	N common	Intercept	Slope
> 10 Mev	Feynman	133	IMP	166	17	-0.91	1.07
> 10 Mev	IMP	166	GOES-6	189	41	-1.03	1.10
> 10 Mev	IMP	166	GOES-7	113	51	0.09	1.01
> 10 Mev	GOES-6	189	GOES-7	113	56	2.10	0.79
> 30 MeV	Feynman	133	IMP	93	5	2.10	0.69
> 30 MeV	IMP	93	GOES-6	129	20	-1.14	1.17
> 30 MeV	IMP	93	GOES-7	66	21	-0.22	1.03
> 30 MeV	GOES-6	129	GOES-7	66	33	2.27	0.77

Examination of the table suggests that the value of the slope is a good measure of the quality of the cross calibration. A slope near unity indicates that the two datasets are well-correlated whereas a slope markedly different of one indicates a problem. There are three examples of the latter in the table:

- When merging the Feynman catalogue with events from IMP data for the >30 MeV channel. The problem here is that the number of common events is low (5) and thus the fit is poor.
- When merging the GOES-6 and GOES-7 events for both the >10 MeV and >30 MeV channels. In this case the problem is the presence of too many outlier ev ents with high fluence measured on GOES-7 and low fluence measured on GOES-6. The cause of this disparity is beyond the scope of this demonstration but may be related to the difficulty of distinguishing events close in time.

The other cases in the table appear to be good cross-calibrations with slopes near unity. In these cases the intercept is a logarithmic measure of the difference in calibration (since the straight line fit was made between the logarithms of the fluences). An intercept near zero (e.g. IMP versus GOES-7) indicates close agreement between calibrations, whereas an intercept near –1 (e.g. Feynman versus IMP; IMP versus GOES 6) indicates that the first dataset yields fluences about 10 times lower than those in the second dataset.

6 Using and updating the JPL-91 model

6.1 Special display format

A key element in the JPL-91 model [R1, R2] is the use of a special display format in which a log-normal distribution of cumulative probability follows a straight line. This is an example of *probability plotting*, which is a set of techniques for linearising various cumulative probability distributions [R4]. It is then possible to test for these distributions by plotting the data in the special display format and verifying that the data follow a straight line. The parameters of the distribution can then be derived using standard fitting techniques to derive the gradient and intercept of the straight line.

For a log-normal distribution the cumulative probability P of having an event with an attribute A greater than some value F is $P = \Phi n((\log(F) - \mu) / \sigma)$, where Φn is the normal distribution and μ & σ are the mean and standard deviation of the log-normal distribution. Thus the value of F at the P percentile of the log-normal distribution is given by $\log(F) = \mu + \sigma\, \Phi n^{-1}(P)$, where $\Phi n^{-1}(P)$ is the P percentile of the normal distribution. Thus to test for a log-normal distribution we:

- calculate P the cumulative probability of occurrence of each event, this is $P=n/(N+1)$ where n is the rank of each event in terms of fluence (1= lowest fluence) and N is the total number of events.
- plot the fluence F of each event (on a logarithmic scale) against the inverse normal distribution corresponding to cumulative probability P (on a linear scale).
- fit a straight line to the plotted data. The intercept of this line gives μ and the gradient gives σ.

This approaches was first prototyped in Excel and then implemented in SEDAT (and thus IDL) as the mah!plot_log_norm_max_only tool. The inverse normal distribution, $\Phi n^{-1}(P)$, is readily calculated in both programming environment using built-in functions:

- in Excel it is implemented using the function *NORMINV(P,μ,σ)*
- in SEDAT/IDL it is implemented using the IDL function μ *- gauss_cvf(P)**σ

For the standard normal distribution we have mean μ=0 and standard deviation σ=1.

Figure 8 shows an example of the special display format using events identified using the tool described in Section 4.3 to analyse all available data from IMP. Following the procedure of Feynmann et al [R1] we consider only events with fluences greater than some user-defined threshold (in this case 10^7 protons cm^{-2}). As found by Feynmann et al the resulting plot follows a straight line for high fluences but flattens out at low fluences, so following their approach we obtain the straight line by fitting to the points above the median (i.e. P > 50%). The resulting distribution describes the high fluence events - and, since those events dominate the total fluence, it provides a good model of the total fluence.

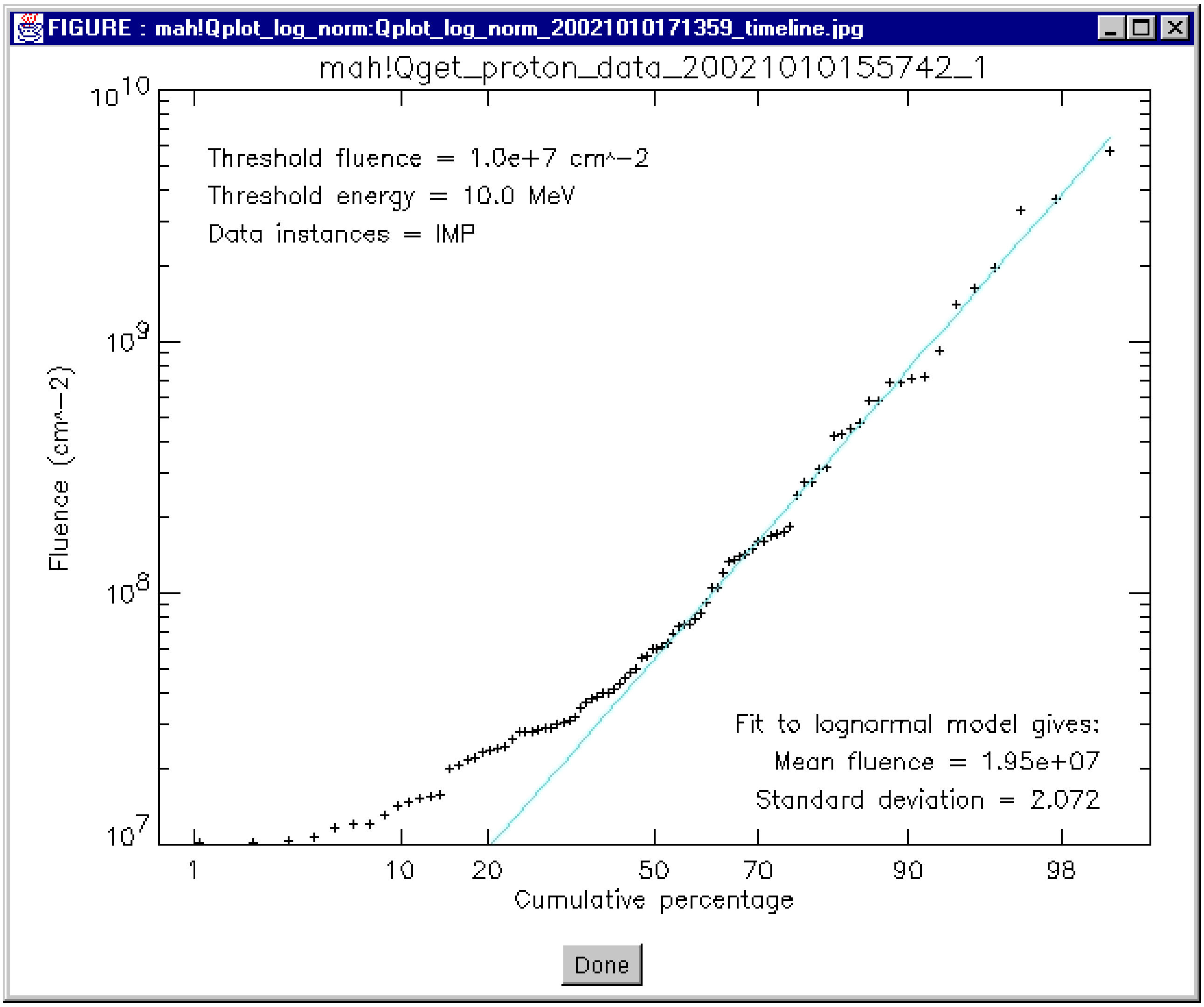

Figure 8. The fluence values of solar proton events plotted against the percentage ranking. The percentage ranking is plotted in proportion to the inverse normal distribution of that ranking so that a log-normal distribution is linearised as shown above. The plot shows the log-normal parameters deduced from a linear fit for P > 50%.

The work of Feynman et al [R1] also showed that solar proton events predominantly occur in a period around solar maximum. For this reason the JPL-91 model assumes that the fluence from solar proton events is confined to a 7-year period about solar maximum. This period is taken to start 2.5 years before solar maximum and to end 4.5 years after solar maximum, with the time of solar maximum being that derived from the twelve-month running mean value of the international sunspot number [R11]. The coefficients of the JPL-91 model were therefore derived using only events in these maximum years. Thus, when fitting the log-normal distribution using the method described above, we must consider only event in maximum years. This was implemented in two steps:

1. Writing a SEDAT function (TBD) that determines if a given date lies in the solar maximum years. If so, the function returns the cycle number of the maximum and if not it returns a null value of –1. For convenience the cycle numbers and the months of their maxima are held in a data structure that is initialised before examining individual events. The initialisation is done by another function in which contains hard-coded values of the cycle numbers and the months of their maxima. Given the small number of the solar cycles to be considered, this is more efficient that putting values in a data file.
2. On reading the events into the log-normal analysis tool, the tool checks if the event falls in a solar maximum period. If so, the event is added to the set for analysis as described above. If not, it excluded from the analysis.

Figure 9 and Figure 10 show an example of the analysis restricted to solar maximum years. These two figures are based on event fluences from the Feynman catalogue (for years up to 1985) and on SEDAT analysis of IMP proton data (for 1985 to 2000).

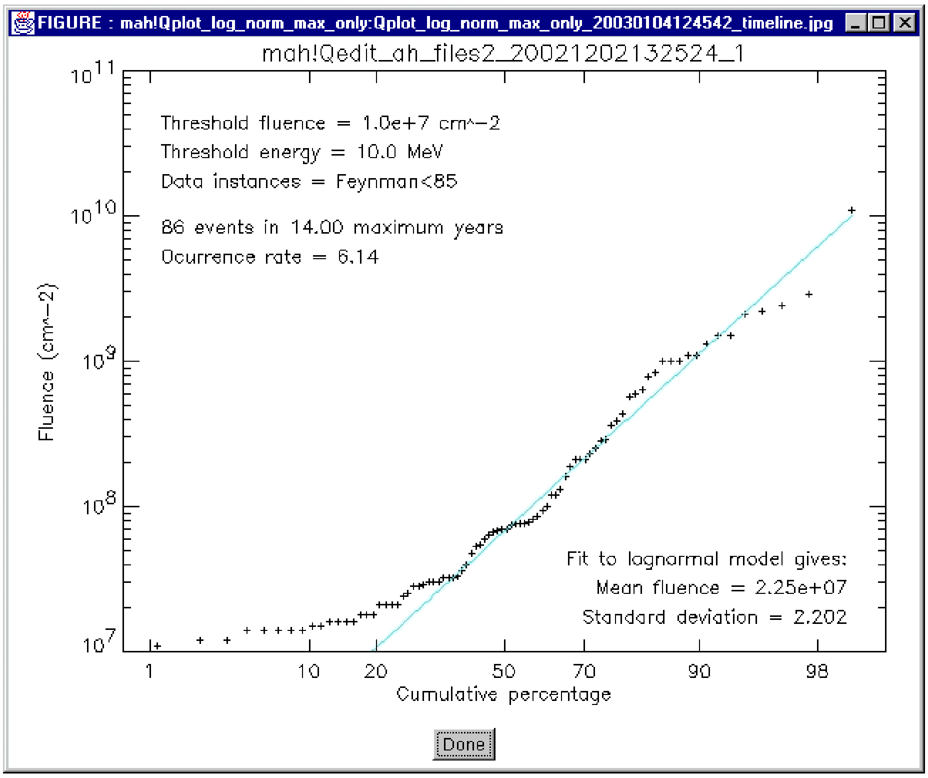

Figure 9. >10 MeV solar proton event fluences from the Feynman catalogue (1965 to 1985).

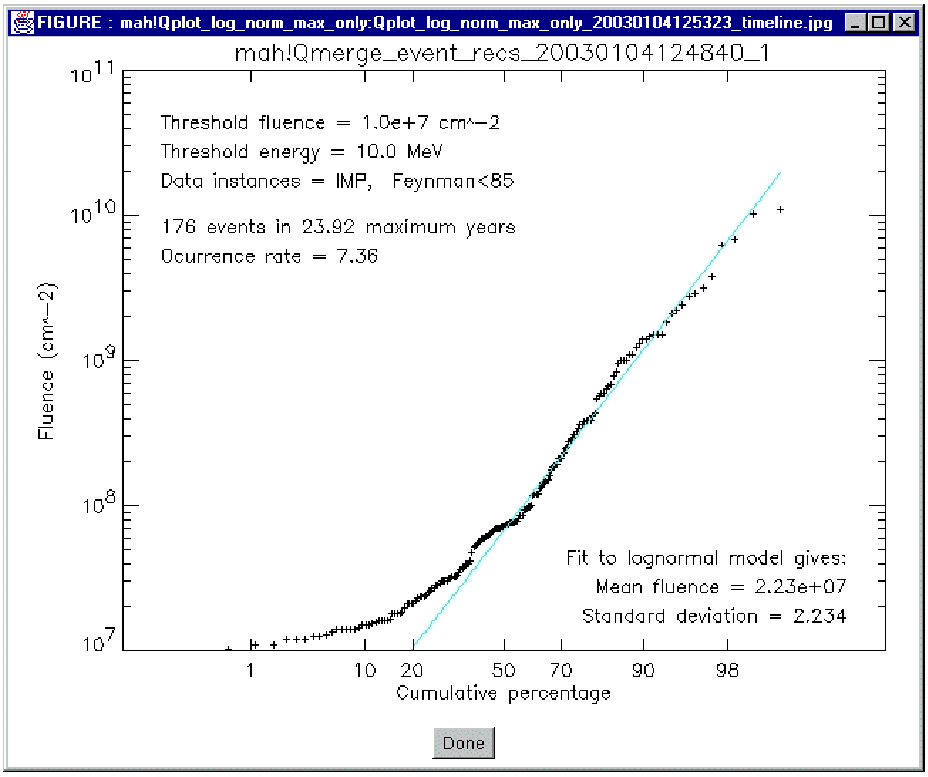

Figure 10. >10 MeV solar proton event fluences from Feynman + IMP (1965 to 2000).

6.2 Some survey results

Table 6 below shows the results of applying the solar-maximum log-normal analysis to the various solar proton event datasets including the Feynman catalogue, the event datasets derived from analysis of GOES and IMP data and combinations of these generated using the merging tool described in section 5.3.

Table 6. Results from log-normal analyses performed by SEDAT.

Dataset	Energy	Mean μ	e^{μ}	Standard deviation σ	Occurrence rate	Maximum years
Feynman	>10Mev	18.03	6.77E+07	2.202	6.14	14.00
Feynman + IMP	>10MeV	18.04	6.81E+07	2.234	7.36	23.92
GOES6	>10Mev	18.14	7.59E+07	2.356	4.04	6.94
GOES7	>10Mev	18.68	1.30E+08	2.319	4.90	6.94
IMP	>10Mev	17.99	6.49E+07	2.063	4.77	16.97
Feynman	>30Mev	17.58	4.33E+07	1.789	3.42	14.00
GOES6	>30Mev	18.30	8.88E+07	2.029	2.01	6.94
GOES7	>30Mev	18.70	1.32E+08	1.700	2.45	6.94
IMP	>30Mev	17.55	4.20E+07	1.926	2.38	16.80
GOES6	>60Mev	17.71	4.93E+07	1.740	1.89	6.35
GOES7	>60Mev	18.01	6.61E+07	1.488	1.74	6.89
IMP	>60Mev	17.21	2.97E+07	1.801	1.97	16.75

6.3 Implementing the model

The JPL-91 model has implemented as a SEDAT tool using the approach described by Feynmann et al [R1, R2]. The probability that the fluence level due to solar proton events will exceed F in time t is given by $P(F,t) = \Sigma_{n=1}^{\infty}\ p(n,F,t)$, where p(n,F,t) is the probability that the fluence level will exceed F due to n solar proton events in time t. In practice this summation must be made finite; to achieve this, we monitor the relative change in the value of P(F,t) as n is increased and stop the summation when that relative change is sufficiently small; less than one part in 10^4 is used in the SEDAT tool. For durations of a few years, this requires 20 to 40 steps in the summation.

The individual terms p(n,F,t) are calculated using the equation p(n,F,t) = A(n,t) Q(F,n), where A(n,t) is the probability that n events occur in time t, while Q(F,n) is the probability that n events generate a fluence exceeding F.

1. The first term is just a Poisson probability. It may be conveniently calculated using the iterative formula $A(n,t) = A(n-1,t) * \omega t/n$, where ω is the occurrence rate of solar proton events. The first step (n=1) in the iteration takes $A(0,t) = \exp(-\omega t)$.
2. The second term is calculated by a large set of numerical simulations as discussed in R1: For each simulation step (a) we calculate x a normally-distributed pseudo-random number using the IDL function RANDOMN, (b) we convert this into a log-normally distributed fluence using $10.0^{(x*\sigma+\mu)}$, where μ and σ are the mean and standard deviation of the log-normal distribution (as derived in Section 6.1), and (c) we repeat the process n times to accumulate the fluence from n events. We repeat the simulation 100000 times and determine M, the number of times, that the accumulated fluence exceeds F. Then Q(F,n)=M/100000.

To improve execution speed of calculation, we use an idea suggested by Alain Hilgers (email 31 March 2000). The tool can process a complete set of fluence levels $\{F_1, F_2, Fi, \ldots\}$ in parallel. For each simulation step we separately determine if the simulated fluence exceeds each value F_i in that set. This yields a related set of counts M_i of the number of times that the accumulated fluence exceeds each value F_i. Thus the probability of exceeding each value is $Q(F_i,n)=M_i/100000$.

The results compare well with those reported in the Feynmann et al papers. For example Figure 11 shows the results of running the SEDAT JPL-91 tool using the parameters for > 10 MeV protons as reported in [R1]. This figure can be directly compared with Figure 4c of R1 and shows good general agreement. Similarly, Figure 12 shows similar results from SEDAT for > 30 MeV protons and can be directly compared with Figure 4d of R1. Again there is a good general agreement.

The JPL-91 model is implemented in SEDAT as a function (mah!jpl_91_array), so that it can easily be called from other tools. The results shown in this section were derived using a top-level tool (mah!test_jpl_model) dedicated to running the model.

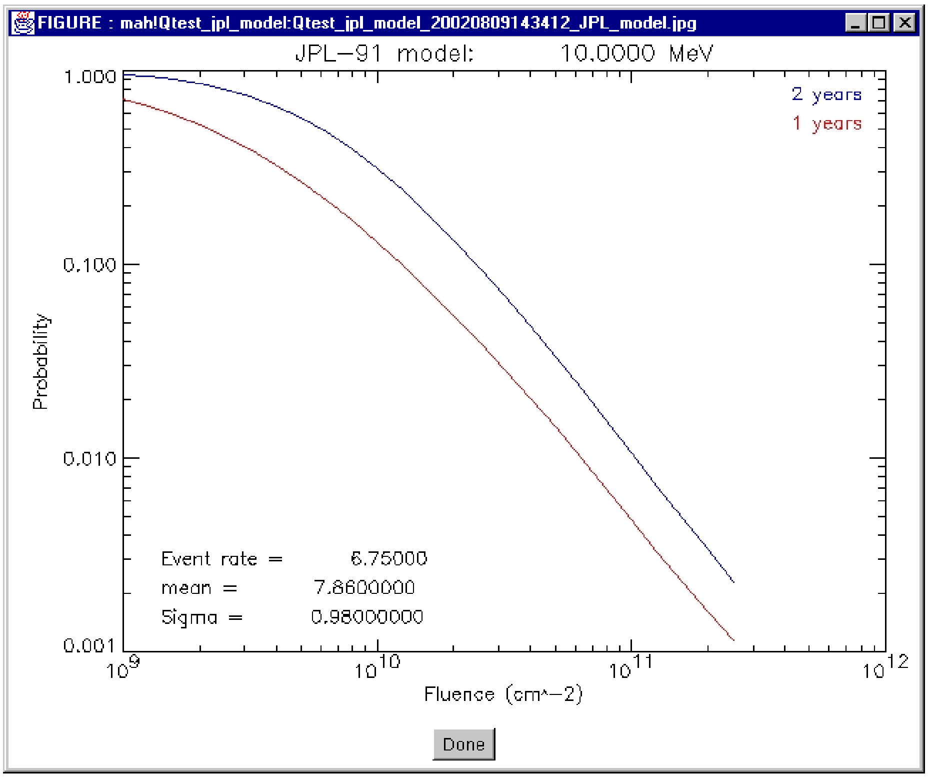

Figure 11. Probability vs fluence for 1 and 2 years duration using the JPL-91 model at 10 MeV.

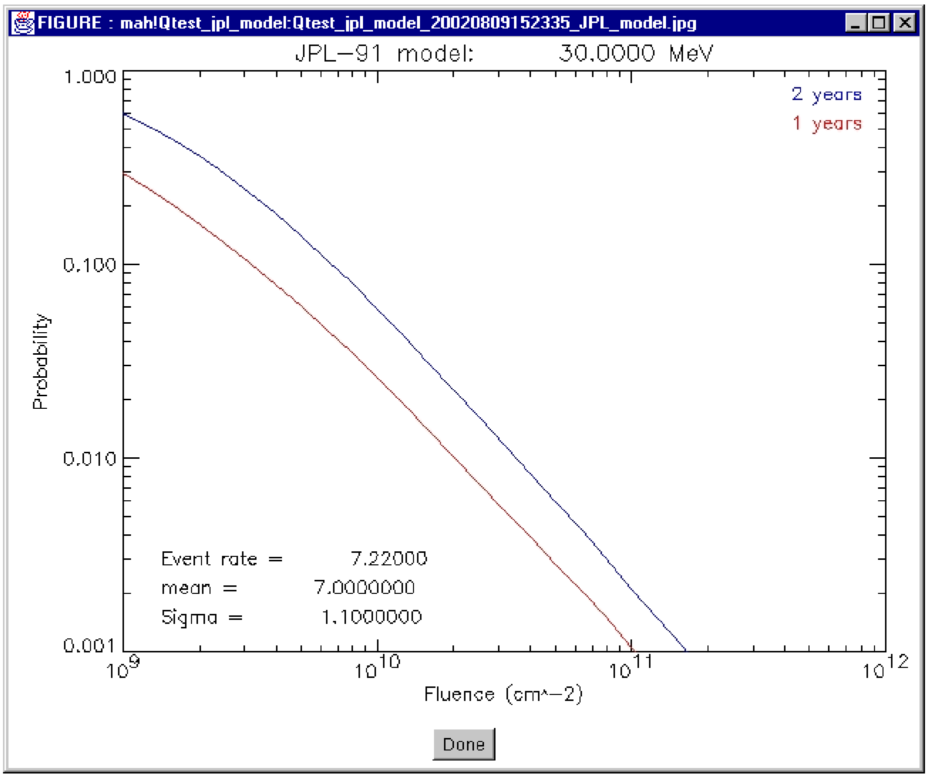

Figure 12. Probability vs fluence for 1 and 2 years duration using the JPL-91 model at 30 MeV.

7 Statistical sampling of the fluence distribution

As an alternative, and a comparison, with the JPL-91 model, a SEDAT tool (mah!fluence_stats_max_only) has been written to derive probability versus fluence curves for various durations (as in Figure 11) by direct statistical analysis of the solar proton data as follows:

- we take a time series of accumulated fluence as shown in the middle panel of Figure 5
- we select a set of pseudo-random samples from the time series, each with the required duration, and derive the fluence accumulated in each sample, i.e. the difference between the accumulated fluences at the start and end of the sample.
- for consistency with the JPL-91 model we consider only samples that come from solar maximum years.
- this yields a distribution of fluences for the required duration, this can then be analysed to derive the percentiles of the distribution, An example distribution is shown in Figure 13, where it is clear that these distributions are highly skew-symmetric - with a high density at low fluences but a long tail extending to high fluences. This type of distribution mandates the use of percentiles rather than mean and standard deviation.
- we then simply plot each percentage against the corresponding percentile of the distribution as shown in Figure 14.

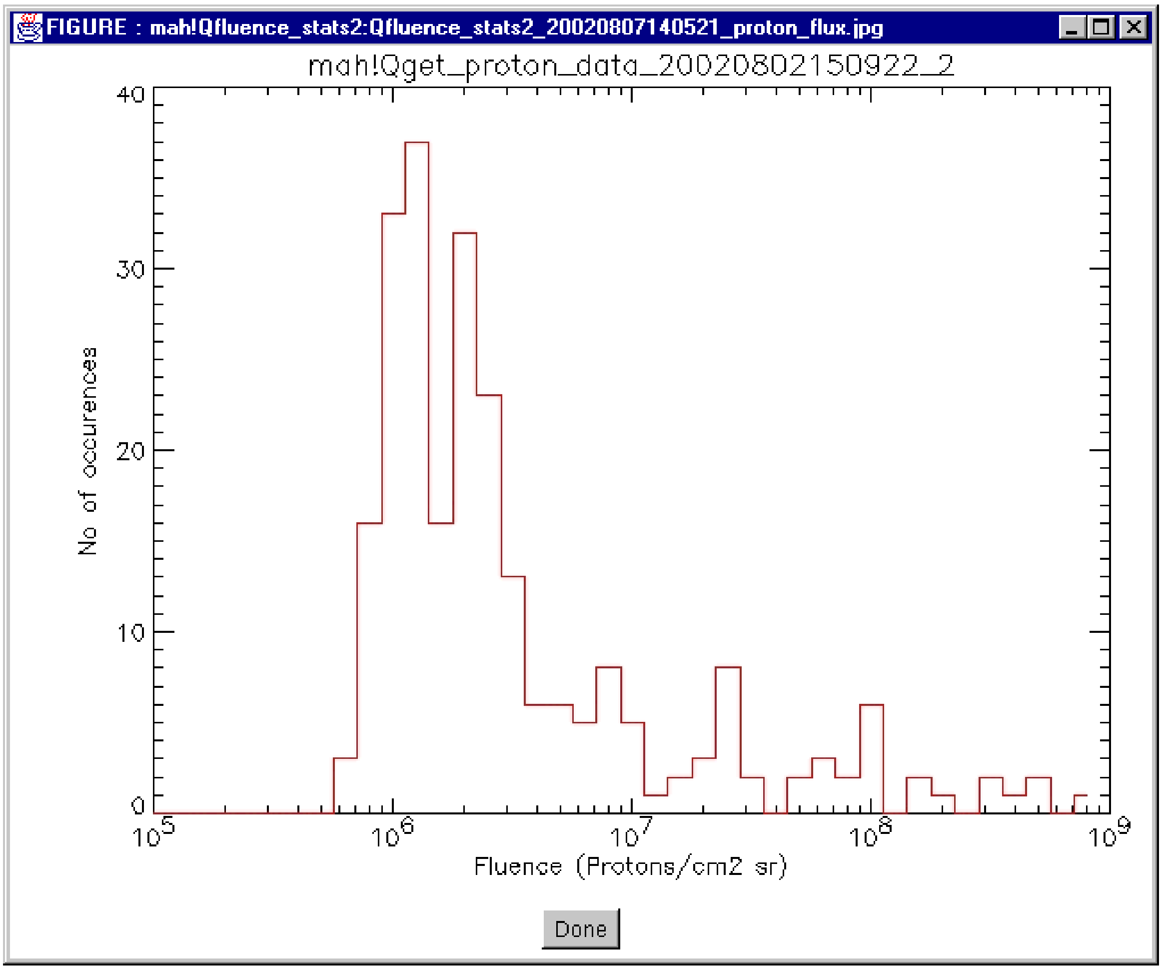

Figure 13. Distribution of fluence values in a set of pseudo-random 1000-hour bins taken from the 10 MeV channel of the whole IMP dataset. …

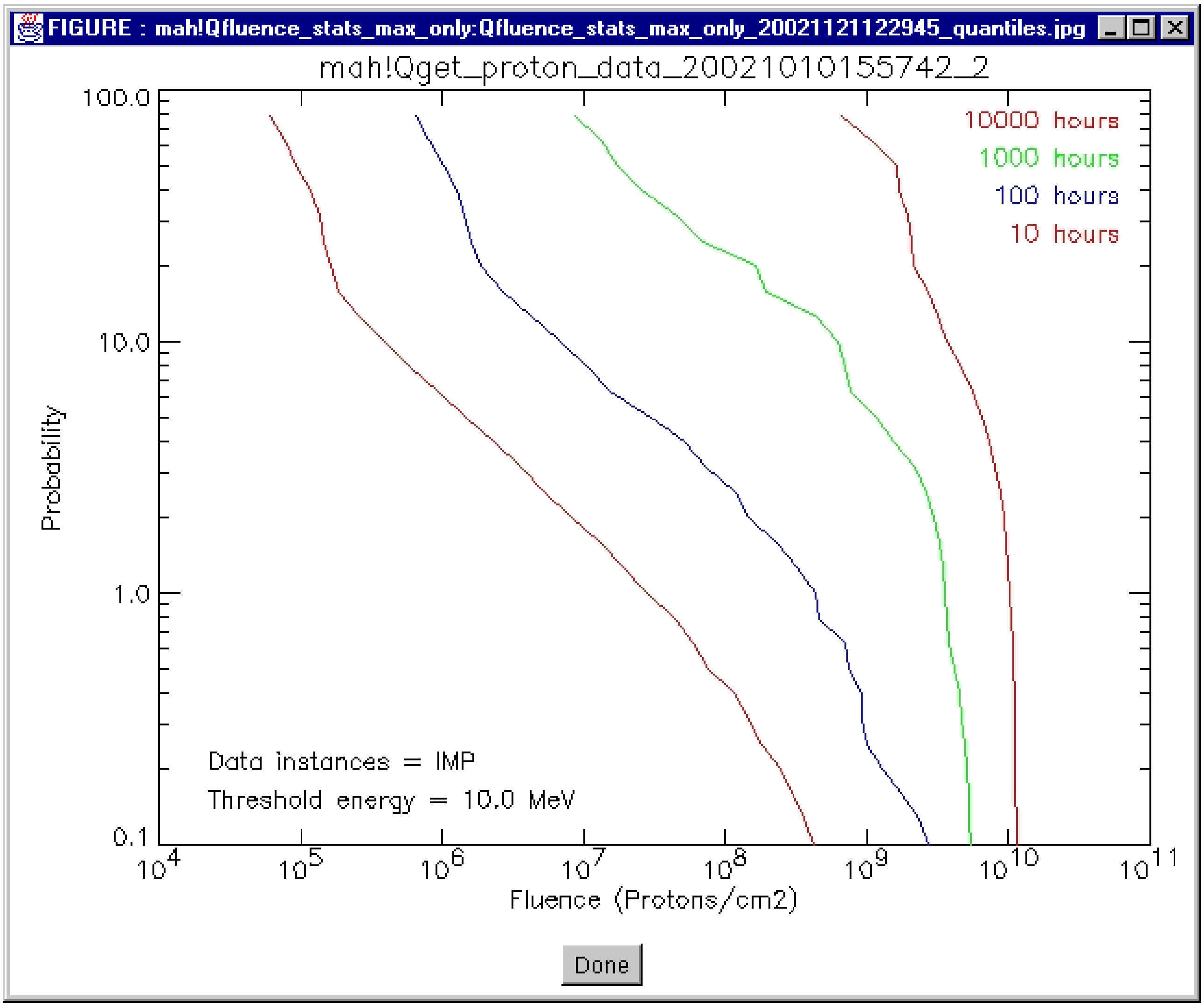

Figure 14. Percentiles of the fluence distribution for 10 MeV IMP proton data for durations between 10 and 10000 hours

The pseudo-random sampling of the fluence dataset is performed by choosing a series of start times which step through the dataset with time increasing but where the step size is a pseudo-random values between 0.5 and 1.5 times the sample duration. Thus the average step size matches the sample duration and so we expect that the number of samples will be close to the number of independent samples that can be taken from the fluence dataset.

For each sample, we determine whether or the start time and end time fall in solar maxima (using the cycle check function described in section 6.1). When, and only when, both start and end occur in the same maximum, we add the sample to the fluence distribution.

The mah!fluence_stats_max_only tool allows the user to apply a scale factor so that we can under- or over-sample the dataset. This has only proved necessary when the sample duration becomes large (> 2 years), such that only very few independent samples can be extracted (especially as there is a maximum sample size of 7 years set by the requirement to take data from solar maximum years only). In this case it is necessary to over sample the data in order to get sufficient samples to allow the software to function.

We can also generate plots that simultaneously display percentiles in all available energy channels in a dataset. This uses tools developed in WP302 and described in the report on that workpackage [R15], namely,

- mah!get_fluences, which generates fluence data files similar to those produced by mah!get_proton_data except that the energy and fluence are vectors rather than scalars. Thus information on all available energy channels is available.
- mah!fluence_levels_max_only, which is adapted from mah!fluence_stats_max_only. It reads the vector fluence data files, generates percentiles in all available energy channels and display them as shown in Figure 15.

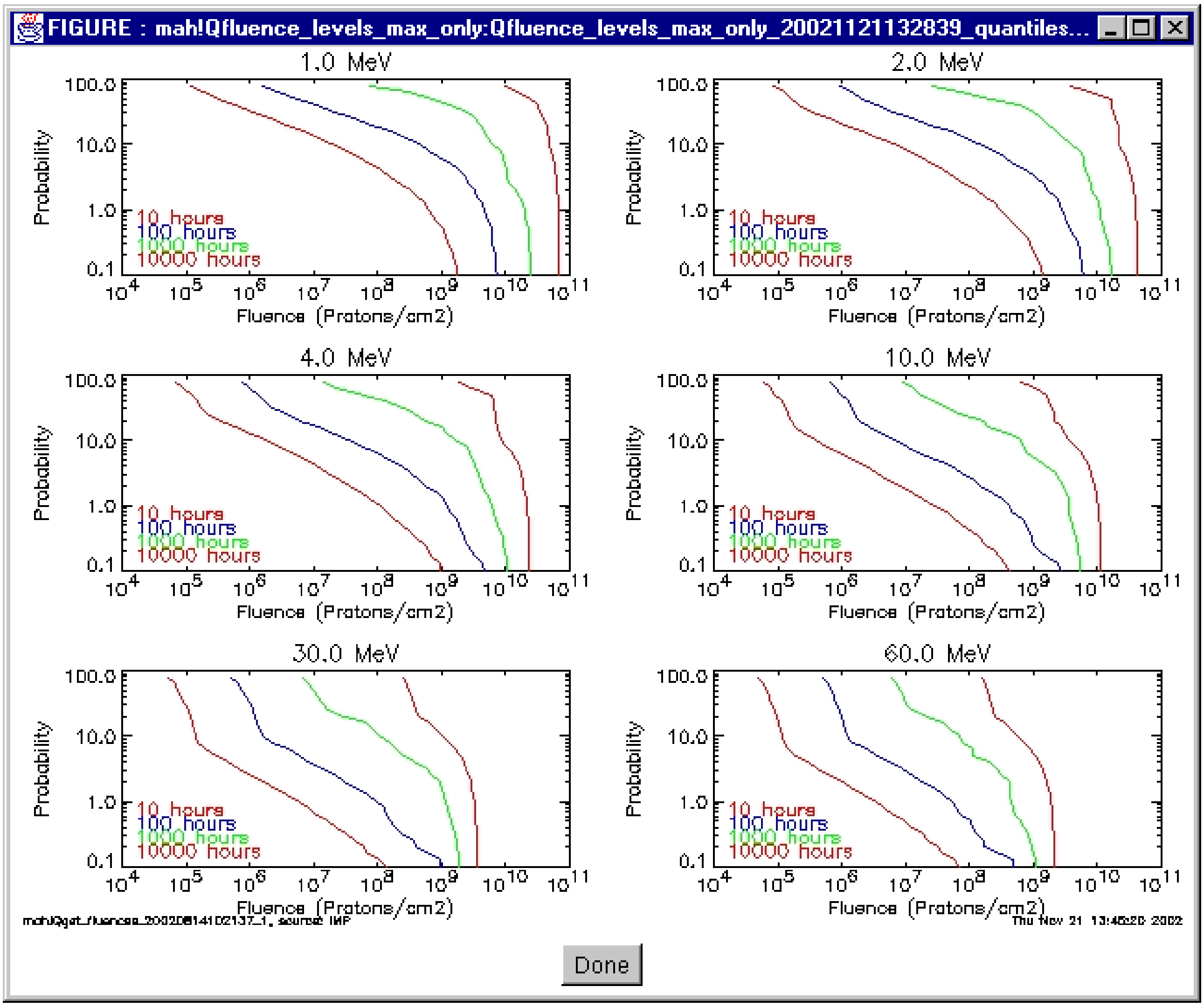

Figure 15. Percentiles of the fluence distribution for all IMP proton data channels for durations between 10 and 10000 hours

8 Comparison of percentile plots with the JPL-91 model

The percentile plots shown in the previous section can easily be compared with the results of the JPL-91 model by displaying both sets of results on the same plot. Because the JPL-91 model is implemented in SEDAT as a function, it is straightforward to invoke this function from a modified version of the mah!fluence_stats_max_only tool described in Section 7 and thus display both results on a common plot. This is illustrated in Figure 16 below.

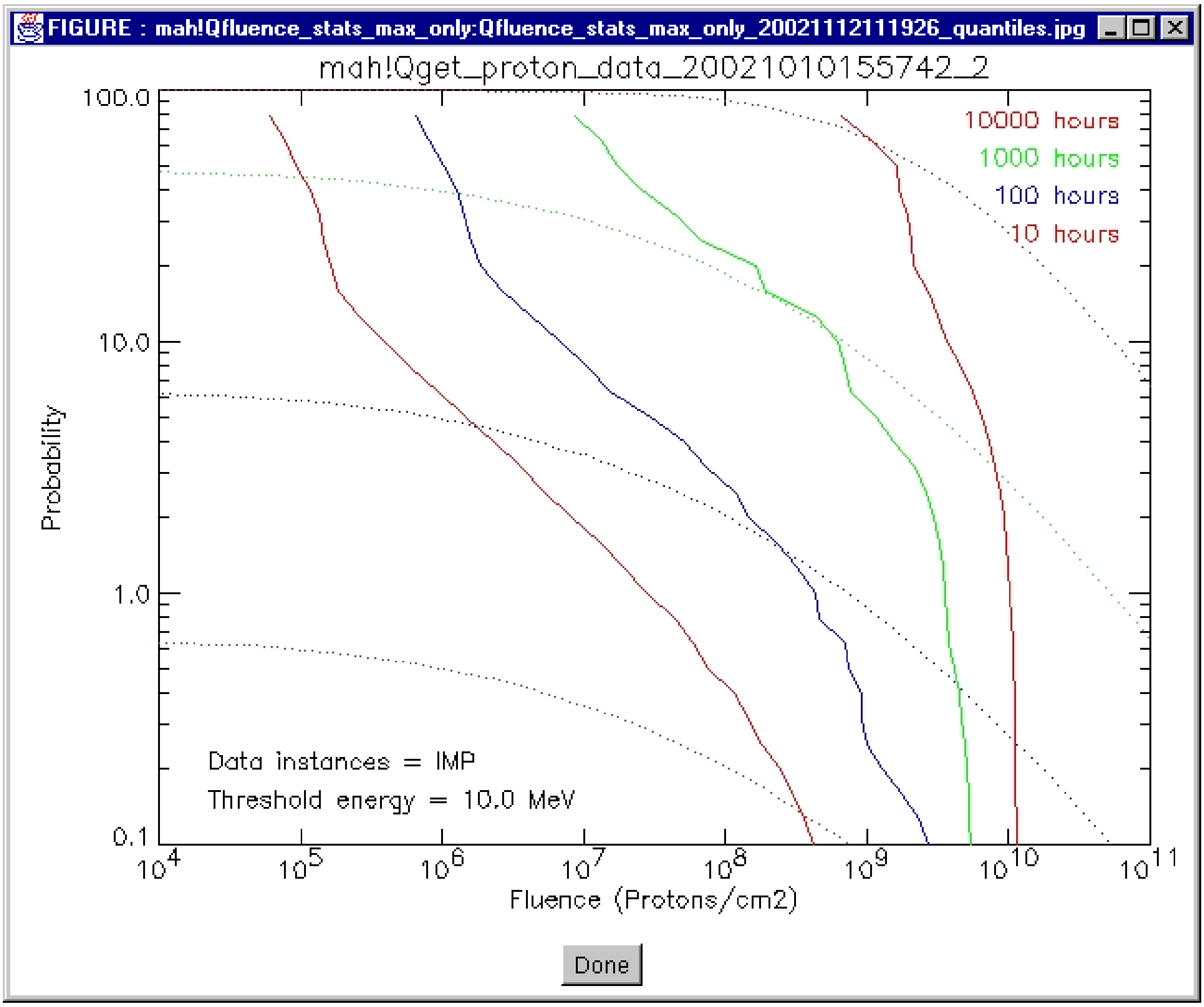

Figure 16. Percentiles of the IMP proton fluence data (solid lines) overplotted with the predictions of the JPL-91 model (dashed lines).

Figure 16 shows that there is a marked difference between the percentiles derived via the JPL-91 model and those derived by random sampling of the fluence timeline. The probabilities derived by sampling the timeline are higher than the model values at low fluences and lower than the model at high fluences. This difference requires some explanation. To do this we must look at the strengths and weaknesses of the two approaches.

First let us look at the sampling approach. Table 7 shows the number of fluence samples contributing to each of the durations for which solid curves are shown in Figure 16. You can see that the short duration curves are based on large numbers of samples but only small numbers of samples contribute to the longest duration curves. Thus the short duration curves are clearly more reliable than the longest duration curves.

Table 7. Numbers of samples for IMP proton data

Sample duration	Number of samples
10	16068
100	1537
1000	147
10000	11

In addition, we note that the longest duration curves show a near-vertical decline. This shows that the fluence values are saturating at the highest fluence available in the dataset under study. The sampling approach cannot extrapolate beyond the range of data available.

The modelling approach has very different strengths and weaknesses. It is based on a model of fluence occurrence and thus can easily extrapolate beyond the range of data available. Thus the model-derived curves of probability versus fluence do not saturate and have slopes such that higher fluences are possible but at monotonically decreasing probabilities. The weaknesses of the modelling approach come at low fluences. These are poorly represented because the low fluence tail of the event fluence distribution (see Figure 9 and Figure 10) is a poor fit to the log-normal distribution. Indeed it is clear from those two figures that low fluences occur more frequently than would be predicted by the log-normal distribution (the straight-line in Figure 9 and Figure 10). This is fully consistent with the result in Figure 16 where the sampling approach shows higher probabilities of low fluences. This result is consistent with the papers of Feynman et al [R1, R2]; they state clearly that their aim in developing the model was to address high fluence events.

Thus it is clear that the strength of the sampling approach is as a means of estimating the fluence probabilities for short periods and especially at low fluence levels. In this regime it is clearly superior to the modelling approach. However, sampling is a weak approach when estimating fluences over long periods; in this regime the modelling approach is superior (as might be expected from the objectives of Feynman et al in developing the model). Both approaches can be applied to particular applications. The modelling approach is ideal for estimating long duration risks from solar proton events, e.g. proton impacts on spacecraft sub-systems over mission durations of years. The sampling approach is suited to estimating short-term risks from solar proton events, e.g. extra-vehicular activity.

9 Conclusions

9.1 Solar proton events

This report describes the development of SEDAT tools for identifying solar proton events and for exploiting the event lists to model solar proton fluences. The key issue in this development was the establishment of robust and efficient algorithms for conditioning the raw data (integral proton fluxes) which is subject to spurious spikes and to data gaps. Given the large dynamic range of proton fluxes and their intrinsic positive nature, it was realised that techniques from image processing would be more appropriate for spike suppression than classical signal processing techniques. After some experimentation a running median was adopted as a robust way to remove spurious spikes from the time series of proton flux. Data gaps were filled by interpolation. However, it was quickly apparent that a simple linear interpolation scheme was over-estimating fluxes when gaps started or ended during solar proton events – and that this lead to a significant over-estimate of total fluence. To reduce this problem an exponential interpolation scheme was adopted. This provides a minimum realistic estimate of the missing fluence.

Once this conditioning was done it was straightforward to identify solar proton events by selecting times when the proton flux exceeds some user-defined threshold. One issue that arises here is how to distinguish events close together in time. The event detection tool uses a simple criterion - that the flux must fall below threshold for longer than some minimum duration. A value of four hours was used.

The main output of the event detection tool is an event list including attributes such as start time, end time and total fluence. Thus the report specifies a standard content and format for event lists so that they can easily be manipulated. Most importantly this manipulation includes merging lists so that duplicates are identified and used to establish a cross-calibration. The merged list has the duplicates eliminated and the event fluences adjusted to a common calibration. This assumes the worst case condition, i.e. the lower fluence dataset adjusted upwards to match the higher fluence dataset.

The standard content and format for event lists also allows us to create user datasets that hold copies of externally-sourced solar proton event catalogues, e.g. the 1965 to 1985 catalogue published by Feynman and her co-workers [R2].

The event lists can be analysed to obtain the event occurrence rate and to fit their fluence distribution to a log-normal distribution (following the approach of Feynman et al [R1, R2]). This has been applied to a number of datasets and yields results similar to those reported by Feynman et al. The probabilities of exceeding various solar proton fluence levels are then predicted using these results and a tool that implements the JPL-91 model algorithms of Feynman et al [R1, R2]. For efficiency of the computation this tool uses some simplifications suggested by A. Hilgers (private communication).

In summary, these aspects of the report show that SEDAT can be used to develop tools that implement, and perhaps extent, standard techniques for solar proton event studies.

9.1.1 Caveats

There are two main caveats on this work on solar proton events:

1. **The effect of interpolation.** The solar proton flux data contain data gaps which we have filled by interpolation. That interpolation has a significant effect on the fluences if it overlaps with part of a solar proton event. This is discussed in detail in section 3.3, where we show that simple linear interpolation is very likely to overestimate fluence and that an exponential interpolation scheme gives a much more realistic result (see Figure 1 and Figure 2). The exponential interpolation scheme assumes that the flux decays throughout the data gap and thus would underestimate fluence if the flux were to rise during the gap. However, the gap constitutes a lack of evidence about the flux variations and so exponential interpolation provides a minimum realistic estimate of the fluence.

2. **Peak times in Feynman catalogue.** The event merging tool identifies duplicates between two solar proton event lists by finding cases where peak times of events in one list fall between the start and end times of events in the other list. This is a simple robust criterion but one that cannot be directly applied to the Feynman catalogue as the published event list [R2] does not contain times of peak flux. To work round this constraint, we have assumed that the peak time is one-third of the way through the event, i.e. peak_time = (2*start_time + end_time)/3. This assumption is based on the observation that solar proton fluxes generally rise more quickly than they decline. Thus if we estimate a peak time it should be nearer the start time than the end time. However, note that our only use of the peak time is in the event matching algorithm above. This is relatively insensitive to the accuracy of the peak time unless it is very close to the start time. Thus a simple estimate of peak time will suffice – and so the one-third criterion was chosen as the simplest rational division of the event duration consistent with the rise being faster than the decline.

9.2 Statistical sampling of the fluence distribution

The event detection tool also produces a well-conditioned time series of the accumulating solar proton fluence. This is the basis of an alternative approach to predicting solar proton fluences. A tool has been written that can take random samples from the time series but with fixed sample duration. This yields an estimate of the statistical distribution of the fluence that accumulate in that fixed duration and thus we can derive the probability of exceeding a particular fluence level in that duation. This is the same quantity as that calculated using the JPL-91 model.

We have compared the results from this sampling approach with those obtained from the model approach. There is a marked difference between the two. Detailed examination of these differences indicates that the modelling approach is superior for estimating fluence probabilities for long durations and high fluence levels. But the sampling approach is superior for estimating fluence probabilities for short durations. This difference is not surprising as the modelling approach was designed to address the problem of long durations and high fluence levels.

In summary, this aspect of the report shows that SEDAT can be used to develop novel tools for analysis of solar proton data.

9.3 Future work

In this section we identify some options for future work:

1. **Time constant for interpolation.** The data gaps in the solar proton fluxes have been filled by exponential interpolation as discussed above. However, one open issue with this technique is the choice of the e-folding time. A value of two hours was used in the results presented in this document. This is a fairly arbitrary choice, though it does give a plausible interpolation as shown in Figure 2. One issue for future work might be to study this e-folding time in more detail:
   - to assess the sensitivity of the total fluence to its value,
   - to compare the flux profiles of interpolated events with those of events for which full data coverage is available.

2. **Prediction of solar proton fluences.** The results in this paper suggest that a modelling approach [R1, R2] is the better technique for predicting fluences over long periods (years) but that a sampling approach, as developed here, is better for short periods (hours, days). This should be explored further to confirm the results of this paper. In addition, further work is needed to gain a better understanding of the transition between the two regimes. In Figure 16 the fluences derived by the sampling approach cross those from modelling in the region where the model curves steepen towards higher fluences. It would be interesting to determine if this is just a co-incidence or if it is physically meaningful.

10 Annex A – summary of tools used in this demonstration

Figure 17 below shows the top-level tools used in this workpackage (blue boxes) and the SEDAT datasets flowing between those tools (green boxes). The full set of top-level tools and queries is listed in Table 8 together with a summary of the functionality that they provide. The parameters for these queries are listed in Table 9.

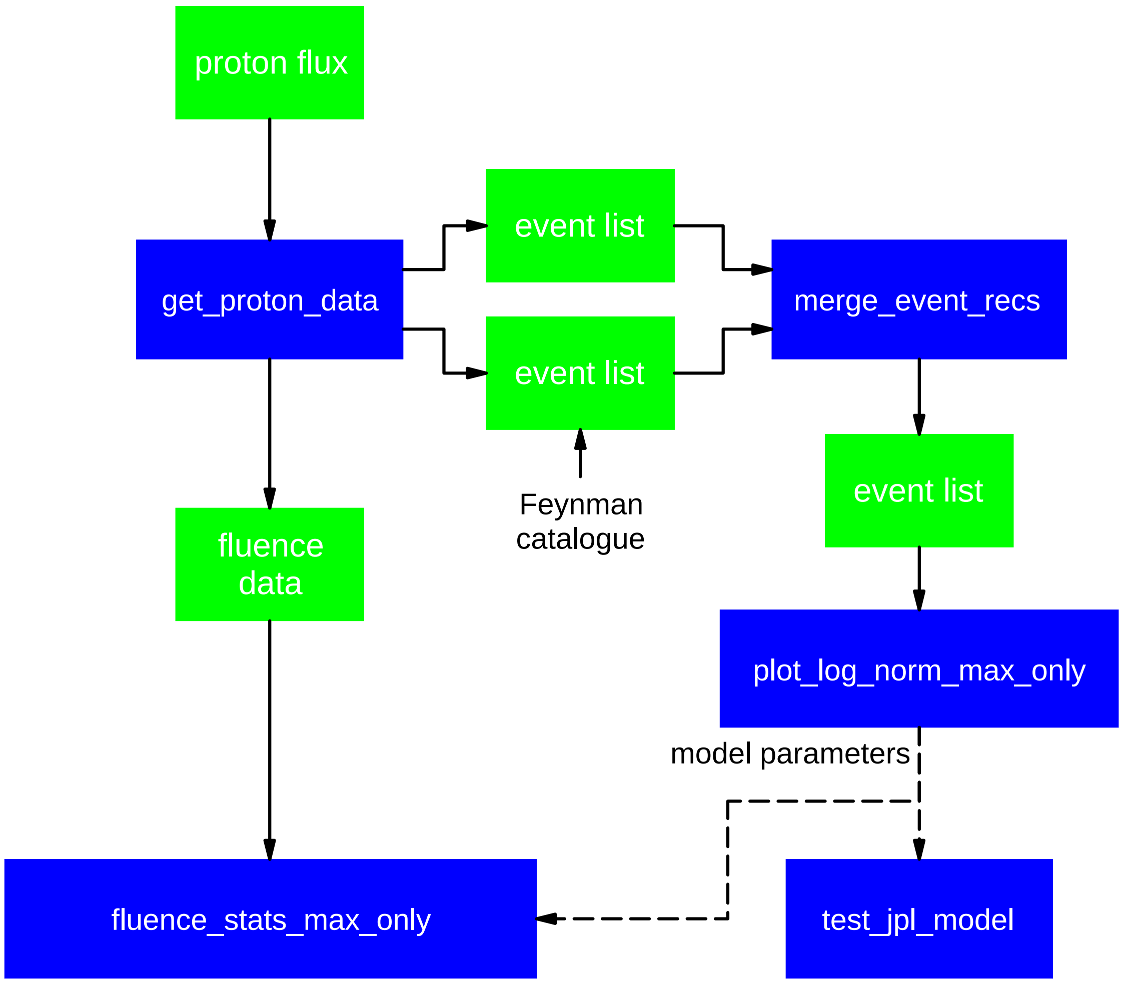

Figure 17. Data flow for WP301.

The primary data input is a solar proton flux dataset (top left of figure). This is then processed by the ma!get_proton_data tool to derive two products: (a) a list of solar proton events and their attributes (start time, total fluence, etc) and (b) a well-conditioned time series of the proton fluence accumulated from the start of the dataset. The list of proton events has a standard format so that it is easy to combine lists from different data sources.

The merging tool (mah!merge_event_recs) will ingest two lists, identify any common events and compare the fluence values for those common events. This comparison allows cross-calibration of the fluences in the two files. To exploit this, we have created a standard format copy of the Feynman catalogue of solar proton events for 1965 to 1985. Thus we can compare and merge event lists derived by SEDAT with the list from that catalogue. The output of the merging tool is a single event list with duplicates removed and fluences adjusted to a common scale.

The event lists can be analysed using the mah!plot_log_norm_max_only tool which implements the log-normal distribution test developed by Feynman et al. This provides model parameters in the same style as the JPL-91 model, i.e. an occurrence rate plus the mean and standard deviation of the log-normal distribution of events. These parameters are reported to the log file for use in other tools.

A function (mah!jpl_model_array) has been written to implement the methodology of the JPL-91 model. This estimates the probability of particular fluence values accumulating in a set of given times. It is controlled by a set of model parameters as described above and can be run in standalone mode via the top-level tool mah!test_jpl_model. These probabilities can also be estimated directly using the mah!fluence_stats_max_only tool; this samples the time series of proton fluence data generated by the first tool above. This last tool also overplots the equivalent probabilities from the JPL-91-like model (derived by calling the mah!jpl_model_array function).

Table 8. Top-level tools for WP301 plus their queries

Tool	**Query**	**Function**
mah!get_proton_data	mah!Qget_proton_data	Identify solar proton events
mah!merge_event_recs	mah!Qmerge_event_recs	Merge solar proton event list
mah!plot_log_norm_max_only	mah!Qplot_log_norm_max_only	Analyse a solar proton event list in terms of the log-normal distribution
mah!test_jpl_model	mah!Qtest_jpl_model2	Run a JPL-91-like model of solar proton fluences
mah!fluence_stats_max_only	mah!Qfluence_stats_max_only	Generate probability-fluence curves by direct statistical analysis of solar proton fluences

Table 9. Parameters used by WP301 queries

Parameter	*Description*	*Recommended values*
mah!Qget_proton_data		
start	Start time of data search	Start time as CCSDS ASCII code A, i.e. yyyy-mm-ddThh:mm:ssZ
stop	Stop time of data search	Stop time as CCSDS ASCII code A
data_source	String uniquely specifying the data format in use.	String describing the data format: a. GOES-I for GOES data in I format; b. IMP for IMP-J data
data_instance	Integer code identifying the spacecraft on which data were taken.	Spacecraft code as in Table 2.
threshold_flux	flux threshold for event detection in units of cm^-2_s^-1_sr^-1	10

Parameter	*Description*	*Recommended values*
energy_channel	number of energy channel to analyse (1 to 7)	3 for > 10 MeV channel in GOES-I data format; 4 for > 10 MeV channel in IMP data
Proton_dataset	Dataset that contains the data to be searched for events. This should logically comprise a series of time-ordered records, each containing a time (epoch) field and the proton flux field(s).	Select “Dataset” from dropdown menu, then use second dropdown menu to select the name of the proton dataset to analyse
mah!Qmerge_event_recs		
n/a	Name of dataset 1	Select “Dataset” from dropdown menu, then use second dropdown menu to select the name of the first dataset to merge
n/a	Name of dataset 2	As for dataset 1
mah!Qplot_log_norm_max_only		
min_fluence	Minimum fluence level to consider when analysing distribution of solar proton event fluences. Units of cm^{-2}.	1.0e+7
event_file	Name of data file containing list of solar proton events and their attributes	Select “Dataset” from dropdown menu, then use second dropdown menu to select the name of the event dataset to analyse
mah!Qtest_jpl_model2		
energy	Energy level (in MeV) for which the model is to be run	10
event_rate	Rate of events per year	Take value from JPL-model: for 10 MeV rate=6.75
sigma	Standard deviation of log-normal distribution	Take value from JPL-model: for 10 MeV sigma=0.97
mean	Mean of log-normal distribution	Take value from JPL-model: for 10 MeV mu= 7.3e+7
mah!Qfluence_stats_max_only		
time_window	Time windows for which we calculate probability of exceeding fluence levels. Can enter multiple values separated by commas.	10, 100, 1000, 10000 (Four values of the window at steps of 10 in window size from 10 upwards)
window_si_rel	Units of time window as an SI_conversion string	3.6e+3>s for units of hours
scale_factor	Scale factor for sampling. We aim to sample the fluence time series with an average period = time window * scale factor	1.0

Parameter	*Description*	*Recommended values*
fluence_data		Select “Dataset” from dropdown menu, then use second dropdown menu to select the name of the fluence dataset to analyse

Table 10. Other important tools used in WP301

Name	Type	Description
mah!check_cycle_phase	Function	Determines if a given date lies in the solar maximum years.
mah!edit_event_recs, mah!Qedit_event_recs	Tool, Query	editing a dataset, e.g. to correct errors or remove dubious events
mah!events_file_meta	Function	Build metadata for SEDAT solar proton event file format
mah!extract_si_con	Function	process an SI conversion string supplied in the format
mah!fluence_file_meta	Function	Build metadata for SEDAT fluence file format
mah!get_events	Function	Analyse thresholded flux data to identify solar proton events and derive their attributes
mah!get_quantile	Function	Get quantiles of a distribution
mah!jpl_model_array	Function	Implements JPL-91 model to calculate probability of exceeding each value in a set of fluence levels $\{F_1, F_2, Fi, \ldots\}$.
mah!proton_data_instances	Function	Returns string with the name of a data instance given its code as in Table 2
mah!read_event_recs, mah!Qread_event_recs	Tool, Query	Generating an ASCII listing of a dataset, e.g. to allow manual examination of the data or export for ingestion into other tools such as Excel.
mah!time_lib	Functions	Library of tools for manipulating time values: • CDFepoch_CCSDS - function to convert CDF epoch to CCSDS A format • CCSDS_CDFepoch - function to convert CCSDS A format to CDF epoch • CDFepoch_MJD - function to convert CDF epoch to MJD • MJD_CDFepoch - function to convert MJD to CDF epoch
mah!valid_num	Function	Check if string is a valid IDL number – routine from SOHO library

11 Annex B - Compliance matrix

The table below shows the compliance between this report and the demonstration plan [R13]. The item numbers relate to items in the demonstration procedure section (4.2) of that plan and compliance section numbers relate to the sections of this document.

Item from Plan	Description	Compliance in section
Section 4.2.1 Model update		
1-4	Solar proton event selection	3.3 and 4
5	Event list files	5.1
6	Ingestion of the Feynman catalogue	5.2
7-10	Merging of event lists	5.3
11-14	Analysis of event lists against a log-normal model	6.1 and 6.2
15-18	Fluence prediction by a JPL-like model	6.3
Section 4.3.2 Use of empirical data		
1-5	Fluence prediction by sampling	7 and 8