

# **SAAPS**

## **Satellite Anomaly Analysis and Prediction System**

Software Requirements Document

Version 0.2

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

P. Wintoft

3 February 2000

## Document status sheet

SAAPS SRD	
Version	Date
0.1	12 December 1999
0.2	3 February 2000

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
1.1	Purpose . . . . .	4
1.2	Scope . . . . .	4
1.3	Definitions, acronyms and abbreviations . . . . .	4
1.4	References . . . . .	4
1.5	Overview . . . . .	4
<b>2</b>	<b>General description</b>	<b>6</b>
2.1	Relation to current projects . . . . .	6
2.2	Relation to predecessor and successor projects . . . . .	6
2.3	Function and purpose of SAAPS . . . . .	6
2.3.1	Database and database tools . . . . .	6
2.3.2	Satellite Anomaly Analysis Module . . . . .	7
2.3.3	Satellite Anomaly Prediction Module . . . . .	7
2.4	Model description . . . . .	7
2.4.1	Database and database tools . . . . .	7
2.4.2	Satellite Anomaly Analysis Module . . . . .	8
2.4.3	Satellite Anomaly Prediction Module . . . . .	8
<b>3</b>	<b>Specific requirements</b>	<b>9</b>
3.1	Time series objects . . . . .	9
3.2	Plotting functions . . . . .	10
3.3	Database and database tools . . . . .	11
3.3.1	Data and meta-data . . . . .	11
3.3.2	Storing data . . . . .	11
3.3.3	Retrieving data . . . . .	12
3.3.4	Converting to/from ASCII . . . . .	12
3.3.5	Automatic download data . . . . .	13
3.4	Satellite anomaly analysis module . . . . .	13
3.4.1	Coordinate transformations . . . . .	13
3.4.2	Statistics . . . . .	13
3.4.3	Superposed epoch analysis . . . . .	13
3.4.4	Cluster analysis . . . . .	13

3.4.5	Pattern search . . . . .	13
3.4.6	Guidelines . . . . .	13
3.4.7	Estimate best model . . . . .	14
3.5	Satellite anomaly prediction module . . . . .	14
3.5.1	Prediction model . . . . .	14
<b>4</b>	<b>User requirements vs. software requirements traceability matrix</b>	<b>16</b>

# Chapter 1

## Introduction

### 1.1 Purpose

The Satellite Anomaly Analysis and Prediction System (SAAPS) contain 3 subsystems: the database and database tool (DB&T), the satellite anomaly analysis module (SAAM), and the satellite anomaly prediction module (SAPM). The user requirements for each subsystem are described in the URD-WP110 (DB&T) [1], URD-WP210 (SAAM) [2], and the URD-WP310 (SAPM) [3]. The purpose of this document is to specify the requirements of the software system from the developers point of view. The SRD incorporates the user requirements described in the URDs.

### 1.2 Scope

### 1.3 Definitions, acronyms and abbreviations

### 1.4 References

1. URD-WP110
2. URD-WP210
3. URD-WP310
4. ESA software engineering standards, ESA PSS-05-0 Issue 2, February 1991.
5. Guide to the user requirements definition phase, ESA PSS-05-02 Issue 1, October 1991.

### 1.5 Overview

The structure of this document follows the document template as described in the ESA software engineering standards [4] and user requirements guide

[5].

Section 2 describes, in general terms, the software requirements of SAAPS. In Section 3 the specific requirements are given.

## Chapter 2

# General description

### 2.1 Relation to current projects

Lund Space Weather Model

### 2.2 Relation to predecessor and successor projects

Lund Space Weather Model and SPEE

### 2.3 Function and purpose of SAAPS

The SAAPS user requirements are specified in three separate URDs ([1],[2],[3]), each capturing the requirements of the three subsystems: database and database tools (DB&T), satellite anomaly analysis module (SAAM), and satellite anomaly prediction module (SAPM).

The purpose of SAAPS is to provide a tool for the analysis and prediction of satellite anomalies. As the users of SAAPS will work in many different geographical regions, and at the same time they share a common database, it is decided that SAAPS shall be operated over the Internet.

#### 2.3.1 Database and database tools

The purpose of the DB&T is to maintain a database of space weather and anomaly data, and to provide this data to the SAAM and SAPM. The DB&T shall therefore have functions to build the database from existing databases. To keep the SAAPS database up to date with real time data the DB&T shall also have functions to automatically retrieve and store new data into the database. This updating ranges from once every 10 minutes to once every day.

### **2.3.2 Satellite Anomaly Analysis Module**

[WP 220]

### **2.3.3 Satellite Anomaly Prediction Module**

[WP 320]

## **2.4 Model description**

The SAAPS model is naturally divided into the three subsystems defining the function of SAAPS, namely the database and database tools (DB&T), the satellite anomaly analysis module (SAAM), and the satellite anomaly prediction module (SAPM). The DB&T can work by itself and do not depend on SAAM or SAPM.

From a users point of view the entrance point to SAAPS will be a web page where he can select whether to run the real time predictions module or the analysis module. In addition to this there shall be links to help pages. The regular user shall thus never be able to directly run the DB&T.

### **2.4.1 Database and database tools**

The access to the database is via the database tools (DBT). The DBT can be run by the system manager, automatically, and by the SAAM and SAPM.

Initially the database does not exist, but must be built. To perform this the data source must be selected, which e.g. can be a URL. Then the data format and the relevant parameters must be identified. This information is provided to the DBT and an object in the database is created. At this stage only the meta data for the database object has been created.

Data can now be stored in the database object. In the process of building the database the system manager must manually run the DBT to add new data for an existing database object.

Existing database objects can also be updated automatically by the DBT at an interval specified by a scheduling program.

The DBT shall provide interfaces to the SAAM and the SAPM. The interfaces are of the type that a request for data comes from SAAM/SAPM and the DBT then returns the data. The SAAM/SAPM can thus only retrieve data from the database and never alter the database.

The database objects can generally be of one of two types: contiguous data or non-contiguous data. As a major part of the database will contain contiguous data the model will be optimized for this. One database object might e.g. contain several years of 1 minute resolution data and it must be ensured that any given time period can be uploaded quickly.



The satellite anomaly data in the database shall have a special status due to the confidentiality issues. The system manager shall always have full access to the database. A user, who always enters via the SAAM or SAPM, shall only retrieve summary information about the anomalies in a way so that the source will never be disclosed.

#### **2.4.2 Satellite Anomaly Analysis Module**

[WP 220]

#### **2.4.3 Satellite Anomaly Prediction Module**

[WP 320]

# Chapter 3

## Specific requirements

### 3.1 Time series objects

A central part in the SAAPS are the handling of time series data. It is therefore convenient to define a class of time series objects (TSOs). It shall be possible to perform the following operations on a TSO:

- create a new TSO,
- store in database,
- retrieve from database,
- linearly interpolate to remove data gaps,
- time average.

The TSO shall always have the format described in Table 3.1. The time field contains the time of the observations and the data fields the values. The data field can e.g. be the electron flux or spacecraft position. The time field shall increase strictly monotonically, i.e.  $t_1 < t_2 < \dots < t_m$ . If the TSO is continuous then the observations are sampled at equidistant times, i.e.  $t_i = t_{i-1} + \Delta t$  where  $\Delta t$  is the sample interval. If there are any data gaps these will be indicated with NaN, Not a Number.

Table 3.1: The format of the time series object (TSO).

Time field	Data fields			
$t_1$	$x_{1,1}$	$x_{1,2}$	$\dots$	$x_{1,n}$
$t_2$	$x_{2,1}$	$x_{2,2}$	$\dots$	$x_{2,n}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$t_m$	$x_{m,1}$	$x_{m,2}$	$\dots$	$x_{m,n}$

When a TSO is created from an external source, e.g. via ftp, it shall be ensured that the above holds. Typically, the data received extends over a certain time period and contain several fields. It must thus be checked that the time field increases strictly monotonically. If this is not the case it must be sorted and possible double times be removed. This shall be done automatically but the system manager shall be notified. If the time series is continuous it shall also be checked that all times in the received time interval exist. If some times are missing these shall be created and the data fields shall be set to NaN.

The storing and retrieving of a TSO is described in Section 3.3.

To further process the TSO it may be necessary to first replace data gaps with linear interpolation. To perform this it must be stated how many contiguous positions with data gaps that shall be replaced by the linearly interpolated value [*Wintoft, 2000a*].

To average a time series three parameters must be given: window size, time offset, and sample interval. The window size determines the resolution of the averaged time series. The time offset is the time with which each average is associated with. Finally, the sample interval is the time steps of the averaged time series. To give an example we have the ACE 1-minute solar wind data that we wish to average to 1-hour resolution. Then the window size is 1 hour. If we want a central average [*Wintoft, 2000a*] then the time offset is 0 hours, whereas a lagging average would have a time offset of +0.5 hours. With a sample interval of 1 hour means that we will have one 1-hour-average per hour, whereas a sample interval of 1 minute will give sixty 1-hour-averages per hour.

In addition to these core functions that operate on the TSO the different modules can extend it with further operations. Figure 3.1 shows the steps from retrieving a TSO from the database, perform preprocessing, running a model, and presenting the result as a plot.

## 3.2 Plotting functions

The plotting functions applies to TSOs. A plot can be produced as a time series plot or as a general plot.

In a time series plot the data values are plotted as a function of time. The time field is formatted so that the meaning becomes clear on the plot. The formatting can be automatic or manual. With automatic formatting appropriate labels and tick marks are determined from the length of the time series. If e.g. the plot extends over a few days then the tick marks should be in days while the label should contain the year and month. With manual formatting the tick marks and labels must be defined. It must also be stated which data fields that should be plotted.

In a general plot one defines which field that should be on the horizontal

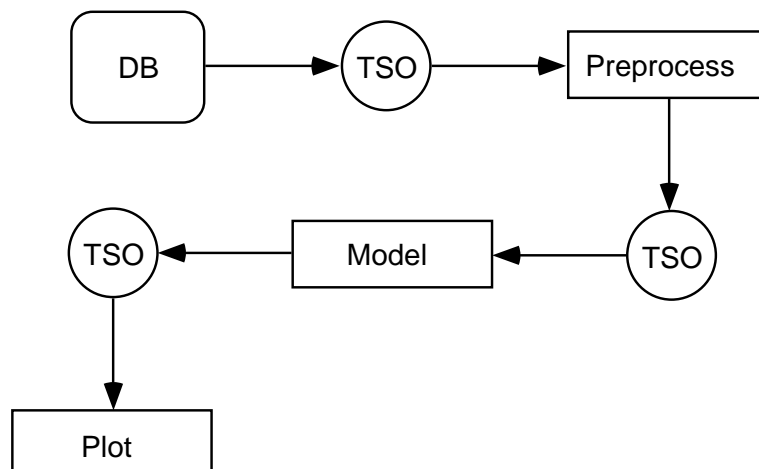


Figure 3.1: The flow of data from the database (DB) using a time series object (TSO) that is preprocessed and run in a model to produce a plot.

axis which field(s) that should be used on the vertical axis.

### 3.3 Database and database tools

#### 3.3.1 Data and meta-data

The database contain two types of objects: database objects (DBO) and meta-data of the DBOs (MDBO). When a DBO is created the first time the MDBO is also created. Then, the DBO can be updated from TSOs. When the DBO is automatically updated it will grow with time. Taking the ACE magnetic field data as an example, which contain the time field plus six data fields for every minute, it will grow with  $7 \times 8 \times 60 \times 24 = 80640$  bytes per day (79 kB/day) or 29 MB/year.

The MDBO contain a description of the DBO and shall contain the fields given in Table 3.2.

#### 3.3.2 Storing data

For a continuous TSO that shall be stored into the database a few things must be checked. If there is a time overlap between the TSO and the database object (DBO) then the times in the TSO must be the same as the times in the DBO for the overlapping period, and if this is not the case a new DBO will be created from the TSO and the system manager shall be notified.

The data values from the TSO shall be written into the existing places of the DBO for the overlapping period. This means that any NaN in the DBO may be replaced by a value from the TSO. If a value in the DBO exist and

Table 3.2: The format of the meta data.

Name	Description
DBO name	The name of the DBO
Description	A short description of the data
Start time	The time of the first data in the DBO
End time	The time of the last data in the DBO
Continuous	True if the DBO is continuous, else false
Sample interval	If continuous, the time step between samples
Field name 1	The name of the first data field
Field name 2	The name of the second data field
⋮	⋮
Field name $n$	The name of the $n$ th data field
Field unit 1	The unit of the first data field
Field unit 2	The unit of the second data field
⋮	⋮
Field unit $n$	The unit of the $n$ th data field

is overwritten by a new and different value from the TSO then the system manager shall be notified.

If there is no overlap between the TSO and the DBO then the times between the last time in the DBO and the first time in the TSO shall be created and the data fields shall be set to NaNs.

### 3.3.3 Retrieving data

A TSO can be retrieved from the database. Four parameters must be given: the name of the DBO, the start time, the end time, and the interval type (open, closed, semi-open, see [Wintoft, 2000a]).

### 3.3.4 Converting to/from ASCII

The data in the database are stored in an internal SAAPS format. It shall be possible to convert the data to/from ASCII data.

To convert a DBO to ASCII the following parameters must be given: the name of the DBO, the time interval, the fields, and the format. The fields could e.g. be the time field and all data fields. The format determines how the time field and data fields should be formatted. The format of the time field could e.g. be YYYY-MM-DD hh:mm:ss in UT. An additional parameter specifies whether the meta-data also should be included in the ASCII data header.

To convert ASCII data to a TSO the format of the ASCII data must be given.

### 3.3.5 Automatic download data

A schedule service shall run on the SAAPS server that automatically runs data downloads at regular intervals, e.g. every 10 minutes or every hour. The program shall connect to the server that provides the data, download the data, create a TSO, and store it in the SAAPS database. Various checks and preprocessing are made, as described in Section 3.1, when the TSO is created.

## 3.4 Satellite anomaly analysis module

### 3.4.1 Coordinate transformations

### 3.4.2 Statistics

### 3.4.3 Superposed epoch analysis

In superposed epoch analysis (SEA) two TSOs are compared to each other. The first TSO is the data that should be superposed and the second TSO contain the list of events that should be used for the analysis. The time interval over which the SEA should be made shall also be given. Finally, the SEA time window [Wintoft, 2000a] shall be given. Thus, to perform SEA four parameters must be given: data TSO, event TSO, time interval, and time window. The result from the SEA is a new TSO which extends over a time interval equal to the SEA window.

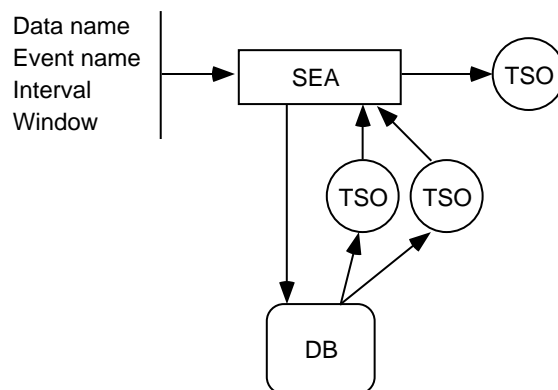


Figure 3.2: Superposed epoch analysis.

### 3.4.4 Cluster analysis

### 3.4.5 Pattern search

### 3.4.6 Guidelines

### 3.4.7 Estimate best model

## 3.5 Satellite anomaly prediction module

### 3.5.1 Prediction model

To run a prediction model two parameters are needed: the model name and the time interval over which the prediction is to be made. Based on the model name the model definition (MD) is loaded. The MD contains which data should be used and various parameters needed for the normalization, pre-processing, and post-processing. The output from the model is a TSO containing the predicted values. Figure 3.3 shows the flow of data in the model.

The first step is to merge the data from possible several TSOs. The TSOs are averaged and resampled to match each other. Then the relevant data fields are extracted and a new TSO is created.

The TSO is then normalized so that all input values lie in the correct range for the network.

The data fields in the TSO are extracted (PreNet) to create the input matrix for the network. This step also includes possible rearrangement of the matrix elements to create a time delay line.

The input matrix is run through the neural network (NN) which generates the output matrix.

Finally, the output matrix is inserted into the merged TSO (PostNet) to generate the model output TSO. The data fields available in the output TSO is thus both the input data and the predicted data. The predicted values are the time stamps that corresponds to the time of the input data plus the prediction time.

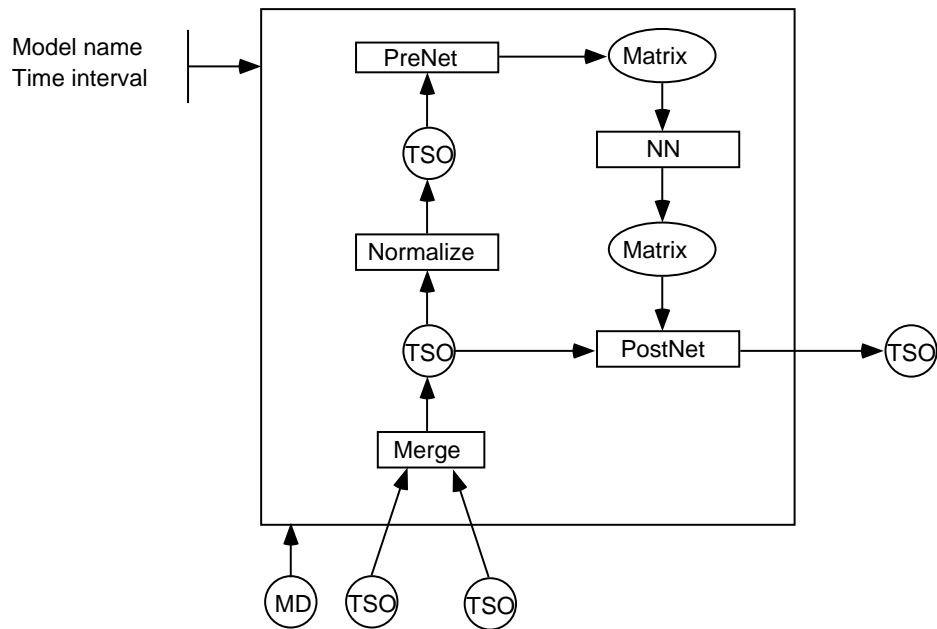


Figure 3.3: The prediction model. Inputs are model name and time interval. Then the model definition (MD) and time series objects (TSO) are loaded. The result is a new TSO.



## Chapter 4

# User requirements vs. software requirements traceability matrix

## References

- Andersson, L., L. Eliasson, and P. Wintoft, Prediction of times with increased risk of internal charging on spacecraft, *Workshop on space weather*, ESTEC, Noordwijk, WPP-155, 427–430, 1999.
- Boberg, F., P. Wintoft, and H. Lundstedt, Real time Kp predictions from solar wind data using neural networks, To be published in *Physics and Chemistry of Earth*, 1999.
- Baum, E.B., and D. Haussler, What size net gives valid generalization?, *Neural Computation* 1, 151–160, 1989.
- Cybenko, G., Approximation by superpositions of a sigmoidal function, *Math. Control Signals Systems* 2, 303–314, 1989.
- Hagan, M.T., and M. Menhaj, Training feedforward networks with the Marquardt algorithm, *IEEE Trans. Neural Networks*, 5, 989–993, 1994.
- Haykin, S., *Neural Networks: A Comprehensive Foundation*, Macmillan College Publishing Company, Inc., New York, 1994.
- Jacobs, R.A., Increased rates of convergence through learning rate adaption, *Neural Networks* 1, 295–307, 1988.
- Freeman, J.W., T.P. O'Brien, A.A. Chan, and R.A. Wolf, Energetic electrons at geostationary orbit during the November 3-4, 1993 storm: Spatial/temporal morphology, characterization by a power law spectrum and, representation by an artificial neural network, *J. Geophys. Res.*, 103, 26,251–26,260, 1998.
- Koons, H.C., and D.J. Gorney, A neural network model of the relativistic electron flux at geosynchronous orbit, *J. Geophys. Res.*, 96, 5,549–5,556, 1991.
- Mayaud, P.N., *Derivation, Meaning, and Use of Geomagnetic Indices*, AGU Geophys. Monograph 22, Washington, D.C., 1980.
- Nguyen, D., and B. Widrow, Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights, in *Proc. IJCNN*, 3, 21–26, 1990.
- Rodgers, D.J., Correlation of Meteosat-3 anomalies with data from the spacecraft environment monitor, *Internal ESTEC Working Paper No. 1620*, ESTEC, Noordwijk, 1991.
- Stringer, G.A., and R.L. McPherron, Neural networks and predictions of day-ahead relativistic electrons at geosynchronous orbit, *Proceedings of the International Workshop on Artificial Intelligence Applications in Solar-Terrestrial Physics, Lund, Sweden, 22-24 September 1993*, Edited by J.A. Joselyn, H. Lundstedt, and J. Trolinger, Boulder, Colorado, 139–143, 1993.
- Stringer, G.A., I. Heuten, C. Salazar, and B. Stokes, Artificial neural network (ANN) forecasting of energetic electrons at geosynchronous orbit, in

- 
- Radiation Belts: Models and Standards, Geophys. Monogr. Ser.*, vol 97, edited by J.F. Lemaire, D. Heynderickx, and D.N. Baker, AGU, Washington, D.C., 291–295, 1996.
- Swingler, K., *Applying neural networks: a practical guide*, Academic Press Ltd, London, 1996.
- Wintoft, P., Satellite Anomaly Analysis Module, *SAAPS, Technical Note 2*, 2000.
- Wintoft, P., Satellite Anomaly Prediction Module, *SAAPS, Technical Note 3*, 2000.
- Wrenn, G.L., and R.J.K. Smith, The ESD threat to GEO satellites: empirical models for observed effects due to both surface and internal charging, *ESA Symposium Proceedings on Environment Modelling for Space-based applications*, ESTEC, Noordwijk, SP-392, 121-124, 1996.
- Wu, J.-G., H. Lundstedt, L. Andersson, L. Eliasson, and O. Norberg, Spacecraft anomaly forecasting using non-local environment data, *Study of plasma and energetic electron environment and effects (SPEE)*, ESTEC Technical Note, WP 220, 1998.
-