Real-time forecast service for geomagnetically induced currents
Technical Note 1: Database

P. Wintoft, A. Pulkkinen, M. Wik, A. Viljanen, R. Pirjola, H. Lundstedt, L. Eliasson

September 29, 2004

Document status sheet

ISSUE	REVISION	DATE	REASON FOR CHANGE
0	0	January 8, 2004	First draft.
0	1	February 27, 2004	Updated Table 1. Updated Sections 2.4.1, 2.4.3, and 2.5.
0	2	September 29, 2004	Updated according comments by Alexi Glover, April 23, 2004. Pages with changes are: iii, 3, 6, 7, 8, 10, and 14. Section 2.4.3 has been removed. Section 2.3.3 has been added.

Summary

This document describes the database that shall be a part of the GIC Pilot Project. The data has been identified in the *User Requirements Document* [1]. The database extends back to 1998 and contains data with high temporal resolution. The database consists of solar wind data (16 second, 1 minute, and 64 second resolution), ground geomagnetic data (1 minute resolution), GIC data (1 minute resolution), and power grid data. The purpose of the database is to facilitate data analysis, model development, and real-time forecast.

Acronyms

GIC Geomagnetically Induced Current

URD User Requirements Document

WP Work Package

URD matrix

The table below lists the requirements from the URD [1] together with the relevant sections in this document.

URD req.	Section
CAP.1.	2.2, 2.4
CAP.2.	2.2.2
CAP.3.	2.2.1
CAP.4.	2.4.2
CAP.5.	Not applicable
CAP.6.	Not applicable
CAP.7.	Not applicable
CAP.8.	Not applicable
CAP.9.	Not applicable
CAP.10.	Not applicable
CAP.11.	Not applicable
CAP.12.	Not applicable
CAP.13.	Not applicable
CAP.14.	Not applicable
CAP.15.	2.2.1
CAP.16.	2.2.2
CON.1.	Not applicable
CON.2.	Not applicable
CON.3.	Not applicable
CON.4.	2.5
CON.5.	2.4.1

Contents

1 Introduction

This document describes the database that shall be a part of the GIC Pilot Project [2]. The data has been identified in the *User Requirements Document* [1]. The purpose of the project is described in the URD as:

> The space weather refers to conditions on the Sun and in the solar wind, magneto-sphere, ionosphere, and thermosphere that can influence the performance and reliability of space-borne and ground-based technological systems, and can endanger human life and health. The space weather can at times induce currents in electrical power grids generally known as GIC (geomagnetically induced currents).
>
> The purpose of this project is to provide a software package that can be used for realtime forecasting of GIC in the Swedish power grid. The software shall be used by power grid operators and tested for a one-year period. During this period, the accuracy and reliability of the software shall be determined, and the usefulness of the software shall be formulated through a cost-benefit analysis. Another aspect is the need to educate the public and decision makers of the potential hazards of GIC and how forecasts can help to mitigate the effects. Thus, the software shall also have apublic part.

To support the model development and real-time operation a database need to be set up. As described in the URD the forecast models shall be driven by solar wind data. Ground geomagnetic field shall be used to compute the ground electric field, and with knowledge about the power grid configuration the GIC shall be computed. Thus, the following basic data shall be included:

- solar wind data,

- ground geomagnetic data,

- power grid data,

- and GIC data.

In addition to this, intermediate data sets may be computed such as the electric field.

To be able to compute the GIC accurately high temporal resolution is needed. Therefore, the data shall be stored with approximately one minute sampling rate, where the solar wind plasma data sets the lower limit at 64 second sampling rate.

We do not carry out any in-depth analysis of the data here, that is left to the technical notes of WP 300 and WP 400. However, we examine the quality of the data in terms of data gaps and coverage.

In the next section we describe the data sources, and then the database.

2 The data

2.1 Time series

In a time series each observed quantity is associated with a unique time at which the observation took place. The observed values in the time series are averages extending over some time interval. Thus the time data in the time series is a pointer, hereafter called time stamp, that refers to an interval. In the database the time stamps are given in Universal Time (UT) and indicates the start of an interval. Thus, a one-minute average at time 12:56 means that data on the interval 12:56≤ t <12:57 has been used.

2.2 Solar wind data

The upstream solar wind plasma and magnetic field drive electrical currents, particle transportation and convection of magnetic fields in the magnetosphere. Changes in the solar wind are manifested on the ground as variations in the magnetic field. Using historic solar wind data, extending over the past 30 years, models have been developed that forecast indices of geomagnetic field variations on time scales of 10 minutes, 1 hour, and 3 hours [3, 4, 5]. The last two models have been implemented for real time operation using solar wind data from the *Advanced Composition Explorer* (ACE) spacecraft [*ACE*]. ACE is located at the Lagrange point L1 which is about 1.5 million km upstream from the Earth. It has been identified in the URD [1] that the ground magnetic field should be predicted with the highest possible resolution, therefore the highest resolution solar wind data shall be used.

2.2.1 Historic data

The ACE spacecraft was placed at L1 in 1998 and from the ACE web page [6] high resolution data can be downloaded back to 1997-1998. In Fig. 1 the data coverage for the different instruments is shown. For our purpose the interesting data comes from the MAG instrument (magnetic fields) and the SWEPAM instrument (plasma density and bulk velocity). The data comes in two different processing levels, where level 2 is recommended for scientific purposes [7]. The highest resolution SWEPAM data are available as 64-second averages, and the MAG data are available as 16-sec averages.

2.2.2 Real-time data

The real time data are available with a one minute resolution from the *Space Environment Center* (SEC) [8]. The data are available as text files at an ftp site [9]. The latest one-minute data is usually delayed by two minutes and covers the last two hours. At the same site the data are available over the latest month organised into daily files. No data exist further back in time at that site. Through the ESA project *Development of AI methods in spacecraft anomaly predictions*
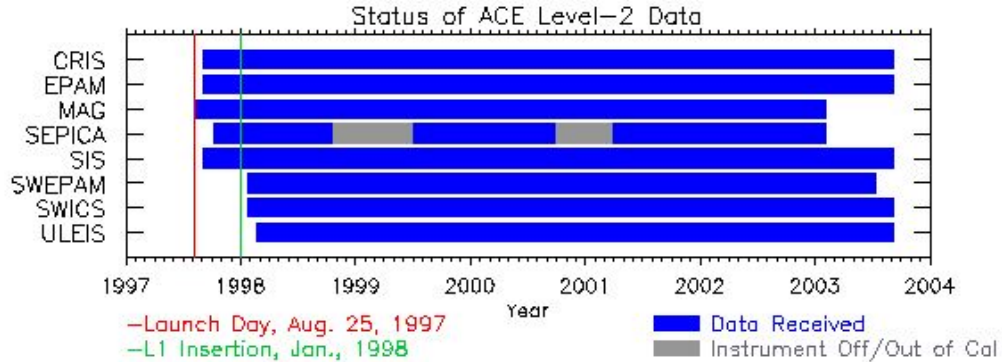
Figure 1: The data coverage for the instruments on the ACE spacecraft. The image was obtained from [6].

[10] a software package called SAAPS were developed. SAAPS contains a database with various space weather data. The ACE one-minute real-time data from SEC have been downloaded into the SAAPS database since March-April 2000, and is thus available to this project.

2.3 Geomagnetic data

2.3.1 Basic geomagnetic database

A basic geomagnetic database with one minute values was generated using recordings from stations nearby southern Sweden (see Fig. 2). Included are all IMAGE magnetometer array sites having data for selected period and seven INTERMAGNET stations BFE, ESK, HAD, HLP, LER, NGK, WNG. The stations that are included in the database are listed in Table 1. Finnish Meteorological Institute (FMI) has the main responsibility of the IMAGE and all data was accessible directly from the server at FMI. INTERMAGNET data was downloaded automatically via anonymous ftp from the INTERMAGNET ftp-server. Both data are freely available for scientific studies. Detailed information about IMAGE and INTERMAGNET can be obtained from www.geo.fmi.fi/image and www.intermagnet.org, respectively. The database was created for a period 1998-2002 for which ACE solar wind and definitive geomagnetic data exists. Written MatLab software enables an easy extension of the database if new stations or new periods with definitive data needs to be included to the geomagnetic database. All geomagnetic data are stored in ASCII format described in www.geo.fmi.fi/image/column.html (new style).

3

Table 1: The name and geographical location of the stations included in the database.

| Station	Latitude	Longitude
BFE	55.63	11.67
ESK	55.32	356.80
HAD	51.00	355.52
HLP	54.61	18.82
LER	60.13	358.82
NGK	52.07	12.68
WNG	53.74	9.07
NAL	78.92	11.95
LYR	78.20	15.82
HOR	77.00	15.60
HOP	76.51	25.01
BJN	74.50	19.20
SOR	70.54	22.22
ALT	69.86	22.96
KEV	69.76	27.01
TRO	69.66	18.94
MAS	69.46	23.70
AND	69.30	16.03
KIL	69.02	20.79
KAU	69.02	23.05
IVA	68.56	27.29
ABK	68.35	18.82
LEK	68.13	13.54
MUO	68.02	23.53
LOZ	67.97	35.08
KIR	67.84	20.42
SOD	67.37	26.63
PEL	66.90	24.08
RVK	64.94	10.98
LYC	64.61	18.75
OUJ	64.52	27.23
HAN	62.30	26.65
DOB	62.07	9.11
NUR	60.50	24.65
UPS	59.90	17.35
TAR	58.26	26.46
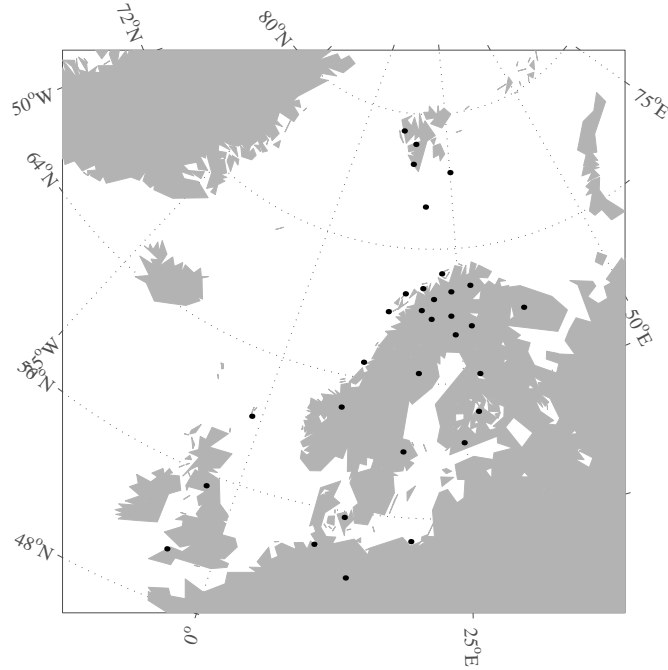
Figure 2: Location of the stations included to the basic geomagnetic database. Geographical coordinates are used.

2.3.2 Interpolated geomagnetic database

The geomagnetic data collected to the basic database was interpolated to a spatially denser grid covering the area where the GIC modeling is carried out (see Fig. 3). The list of significant GIC events for which interpolation was carried out is shown in Table 2. The interpolation procedure utilizes novel Spherical Elementary Current System (SECS) method (for validation and application in GIC studies see e.g. [11, 12, 13] and is executed essentially in two steps: 1.) geomagnetic recordings are used to derive ionospheric equivalent current system for each time step separately, 2.) from the derived equivalent current system, the magnetic field can be computed at any point on the surface of the Earth. The computed field satisfies the physical properties of the geomagnetic field. Prior to the computation of the equivalent currents, datagaps were removed using a linear interpolation and visually selected baselines were subtracted from the data. The written MatLab software utilizes earlier SECS-related tools and is very flexible in terms of fine tuning the interpolation parameters like spatial density of the grid and smoothness of the interpolated field.

2.3.3 Real time data

[Discussions are taking place on the access of real time data from Brorfelde and Uppsala. This section will be updated accordingly.]
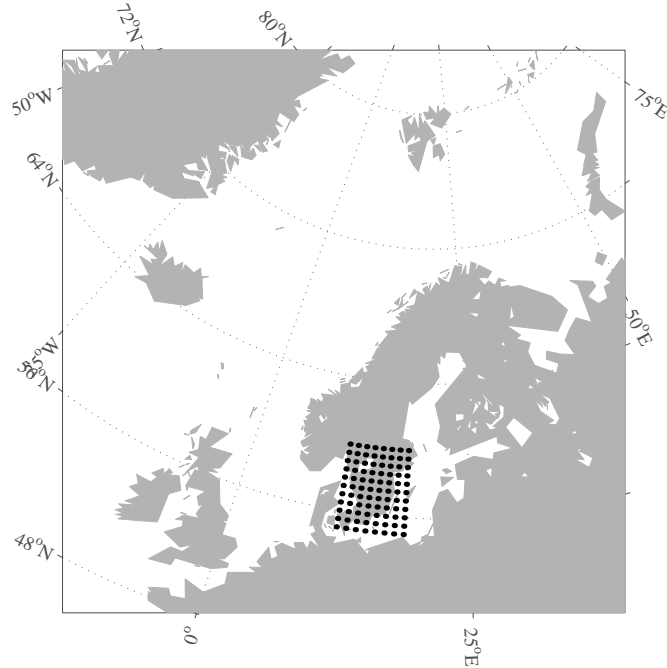
5

Figure 3: Location of the points to which geomagnetic data is interpolated. Geographical coordinates are used.

2.4 GIC data

2.4.1 Measurements at a transformer in southern Sweden

Measurements of GIC has been carried out in a transformer grounding at OKG in southern Sweden for several years. GIC data are at the moment available for the period 1998-2000 for transformer T7 (O3) in Oskarshamn. Data are also available for the period Sep-Okt 1998 for additional 7 transformers. The GIC data all have a resolution of 1 minute. Measurements are performed at all times (T7) but data is saved only when GIC> 1 A except during the campaign Sep-Okt 1998 when all the measurements was saved. The measurements at OKG was performed in the transformer grounding between the voltage levels 20kV and 400kV. Below is a summary of the amount of data available:

- Transformer O3 (Oskarshamn): 1998.09.17-1998.10.28 (the whole period) 1999.08.15-1999.11.13 (when GIC> 1 A) 2000.01.22-2000.08.13 (when GIC> 1 A)

- Transformers: Alvesta, Glan, Kimsta, Kolstad, Nybro, O1, O2 and T7: GIC data from these transformers are from the campaign Sep-Okt 1998. The measurements was performed at transformer stations in Simpevarp (connection from OKG to the 400kV voltage level) and connecting stations.

The measured GIC data from the Swedish power grid may not be published. However, plots and other summary information may be published.

2.4.2 Table of events

From a previous study [14] a list of events were identified based on observed GIC in South Sweden. The list is shown in the Table 2. An event is defined as an occasion when the difference between the maximum and minimum GIC in an one-hour interval exceeds 30A. The definition builds on engineering practice and is related to what levels GIC can be accepted in the system. Thus, the 30 A threshold should not be taken to strictly and may change with technology development. If several events takes place in one day it is still only idicated by one row in the table. The columns are: the date when the event started; the day of the year; the largest absolute GIC for the day; the average solar wind speed during the day; the expected CME arrival time; the actual CME arrival time; whether the CME hit the Earth. The table is based on GIC data that has been modified in accordance to changes that has taken place in the power grid configuration.

2.5 Power grid data

For the GIC modelling a DC-description of the network will be used. The power grid data required are as follows:

- Geographic coordinates of the nodes (latitude and longitude).

- Transformer resistances and earthing resistances.

- Transmission lines resistances.

- Type of transformer e.g. autotransformer.

The 400 kV network in southern Sweden will be used in this project. We consider GIC in the transmission lines to be the sum of the induced currents in all three phases, that is $\text{GIC} = 3 \cdot \text{GIC}_{\text{phase}}$. The transmission line resistances should therefore be used in a similar way. The power grid data is confidential and shall be accessible only to the project team. The data shall not be stored in the same database as the other data.

Table 2: GIC events and their relation to observed CMEs. A GIC disturbance that extends over several consecutive days is counted as one event in the table. In 13 cases out of 23, it seems that the GIC event is related to a CME, four cases are uncertain and in five cases no CME has been observed at the Sun. The GIC and solar wind speed values are both 4-minute average values. The * denotes that the spacecraft SOHO was not available for observations at the time. Table adopted from [14].

Start date	Day no.	Largest absolute GIC (A)	Solar wind speed (km/s)	Expected CME date	CME occured date	Hit
24 Sep 1998	267	78	800		*	*
25 Sep	268	18	600		*	*
2 Oct	275	25	700		*	*
20 Oct	293	140	400		*	*
15 Aug 1999	227	20	375	10 Aug	9 Aug partial	uncertain
20 Aug	232	25	650	17 Aug	17 Aug partial	yes
30 Aug	242	22	500	26-27 Aug	27 Aug partial	yes
22 Sep	265	80	600	19 Sep (20 Sep)	20 Sep halo	uncertain
26 Sep	269	19	550	23 Sep	-	no
10 Oct	283	24	500	6-7 Oct	-	no
12 Oct	285	25	575	9 Oct	-	no
22 Oct	295	65	675	18-19 Oct	18 Oct partial	yes
28 Oct	301	29	410	24 Oct	25 Oct partial	uncertain
13 Nov	317	36	460	9-10 Nov	-	no
22 Jan 2000	22	30	375	17-18 Nov	16 Nov partial	uncertain
11 Feb	42	22	580	8-9 Feb	8,9 Feb halo	yes
24 Feb	55	19	750	20-21 Feb	-	no
6 Apr	97	270	680	4 Apr	4 Apr halo	yes
23 May	144	30	600	20 May	20 May partial	yes
8 Jun	160	62	775	5-6 Jun	6 Jun halo	yes
10 Jul	192	23	475	6-7 Jul	7 Jul halo	yes
11 Jul	193	43	475	7-8 Jul	7 Jul halo 8 Jul partial	yes
13 Jul	195	21	700	10-11 Jul	11 Jul halo	yes
14 Jul	196	48	750	11-12 Jul	11 Jul halo	yes
15 Jul	197	222	980	13-14 Jul	14 Jul halo	yes
19 Jul	201	40	625	15-16 Jul	-	no
12 Aug	225	35	700	9-10 Aug	9 Aug halo	yes

3 The database

3.1 MySQL

SQL stands for *Structured Query Language* and MySQL is an open source database management system [15]. The history of MySQL begins in 1979 and in 1996 a binary distribution for Linux and Solaris was released. Today MySQL is implemented for all major platforms. MySQL is freely available under *GNU General Public License* (GPL), which means that it is free for in-house use.

The data in the database are organised in tables, where each table has a fixed number of columns and variable number of rows. A column corresponds to a physical parameter, e.g. solar wind speed, and a row corresponds to an observation. The time of an observation is a critical and unique parameter and is stored in the SQL DATETIME format. The format is "year-month-day hour:minute:second", e.g. "2003-12-22 12:45:04". It is also defined as a PRIMARY KEY which means that there can only be one row in a table with a specific time stamp.

In the following we describe the contents of the database and how it is organised. Administration tasks, such as creating the database and tables, or updating the database, are described in Appendix B.

3.2 Overview

The data are stored in a database called `gicpilot`. The database is selected with the command

```
mysql> use gicpilot
```

where `mysql>` is the MySQL command prompt. The database contains the tables that hold the data. The tables can be listed using the command

```
mysql> show tables;
```

which should produce an output like

```
+-------------------+
| Tables_in_gicpilot |
+-------------------+
| geomag            |
| gic               |
| mag16s            |
| magswe            |
| swepam64s         |
+-------------------+
```

9

The contents of each table are described below.

3.3 Historic solar wind data

As described in Section 2.2.2 the ACE one minute real time data from SEC have been stored in the SAAPS database since 2000. This archived real-time data differs slightly from the final ACE solar wind data. The final data extends back to 1998. For scientific purposes the final data should be used.

3.3.1 16 sec magnetic field data

The 16-second magnetic field data are stored in the table `mag16s` (see Appendix B). The are 9 fields (SQL columns) in the table and they are described in Table 3.

Table 3: Definition of the fields in the `mag16s` table.

Field	Meaning	Unit
dt	Time stamp	year-month-date hour:minute:second
B	Magnitude of the magnetic field	nT
BxGSE	Magnetic field x-component in GSE	nT
ByGSE	Magnetic field y-component in GSE	nT
BzGSE	Magnetic field z-component in GSE	nT
BxGSM	Magnetic field x-component in GSM	nT
ByGSM	Magnetic field y-component in GSM	nT
BzGSM	Magnetic field z-component in GSM	nT
Quality	A quality flag	> 0 for data gaps

The data covers the period 1998-01-01 00:00:02 to 2003-11-25 23:59:45 and will be regularly updated. There are 11 637 011 rows in the table. Nominally the data are sampled with a 16-second interval, however, there are occasions when there are 15 seconds between two samples. Thus, there is a small discrepancy between the number of 16-second intervals in the time period and the number of rows in the table. This will, however, not be a problem when using the data.

There are 26 642 data gaps (Quality> 0) which is less than 0.23% of the total number of data points. Most of the data gaps take place in 1998 (21 652) and especially during July and November. Figure 4 shows the distribution of data gaps for each month.
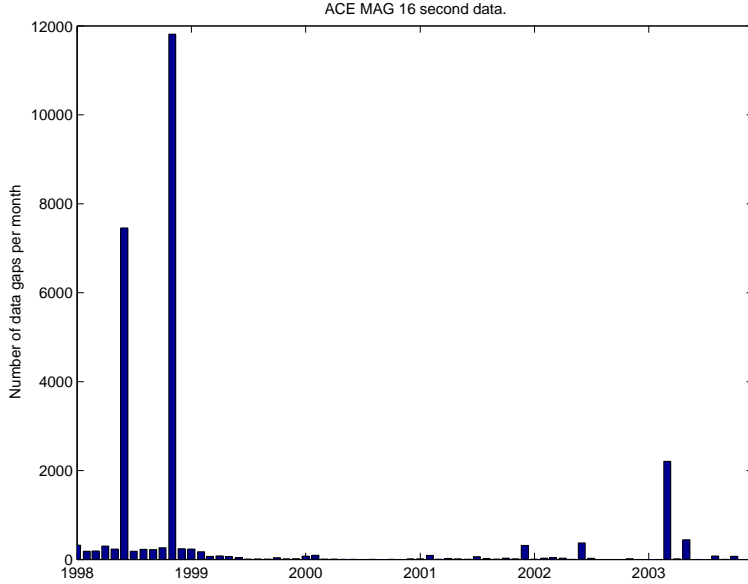
Figure 4: Distribution of data gaps for the ACE MAG 16 second data for each month.

3.3.2   64 sec plasma data

The 64-second plasma data are stored in the table `swepam64s`. There are 14 fields as described in Table 4.

The data covers the period 1998-01-23 00:00:31 to 2003-09-05 23:59:23 and contains 2 770 207 samples. There are 855 sample intervals that are not equal to 64 seconds, out of which 541 intervals are 63 seconds and 271 intervals are 65 seconds. The rest span from 1 second to 127 seconds.

There are 424 195 data gaps in the density n, which corresponds to 15% of the data. The velocity data V contains 193 547 data gaps which is 7% of the data. The distribution of data gaps is shown in Figure 5. Thus, there is quite a large number of data gaps in the plasma data.

It is interesting to examine the distribution of the lengths of continuous data gaps. In Figure 6 the distribution of the data gaps lengths is shown. It is seen that data gaps of 1-minute length, or more exactly 64-second length, completely dominates. By replacing the 1 and 2-minute data gaps with the interpolated value reduces the fraction of data gaps from 15% to 12%.

There are also sequences with many data gaps, especially during 1998. By also identifying the longest data gap sequences and removing them from the data set the fraction of data gaps decreases according to Table 5.

11

Table 4: Definition of the fields in the `mag16s` table.

Field	Meaning	Unit
dt	Time stamp	year-month-date hour:minute:second
n	Particle density	cm^{-3}
T	Temperature	K
He4p	Ratio of He over protons	
V	Bulk flow speed	km/s
VxGSE	x-component of the velocity in GSE	km/s
VyGSE	y-component of the velocity in GSE	km/s
VzGSE	z-component of the velocity in GSE	km/s
VxGSM	x-component of the velocity in GSM	km/s
VyGSM	y-component of the velocity in GSM	km/s
VzGSM	z-component of the velocity in GSM	km/s
xGSE	Location of ACE in GSE	km
yGSE	Location of ACE in GSE	km
zGSE	Location of ACE in GSE	km

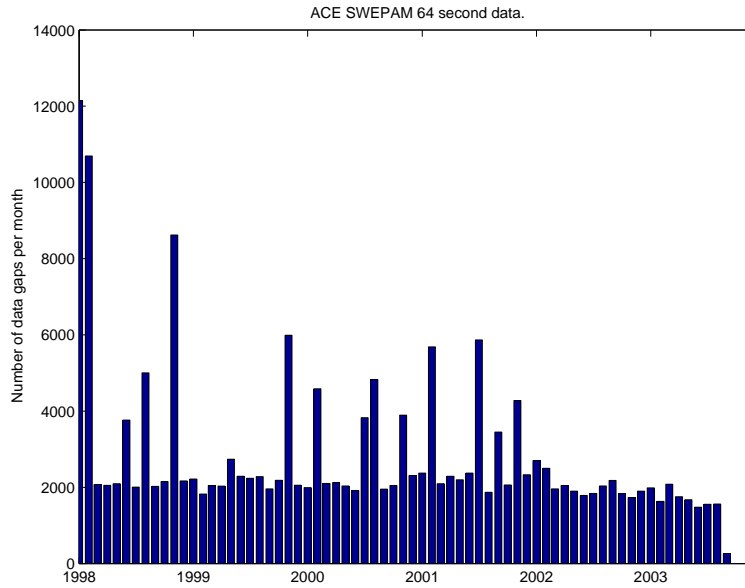

Figure 5: Distribution of data gaps for the ACE SWEPAM 64 second data for each month.

Table 5: default

| Percent data gap	Comment
15.3	All data
11.9	Interpolated 1-minute gaps
11.6	Interpolated 2-minute gaps
11.0	1998-01-23 00:00:31, 17549 data points, 13.0 days
10.7	2001-02-03 08:33:46, 7182 data points, 5.3 days
10.5	1998-11-15 12:25:33, 7006 data points, 5.2 days
10.2	1999-06-20 22:53:23, 6942 data points, 5.1 days
10.0	2000-03-14 06:46:43, 5906 data points, 4.4 days
9.8	2000-08-16 17:08:16, 5780 data points, 4.3 days
9.6	2002-09-26 08:19:22, 5426 data points, 4.0 days
9.4	2001-03-15 11:38:45, 5329 data points, 3.9 days
9.2	2000-10-07 17:30:33, 5200 data points, 3.9 days
9.1	2002-01-03 14:04:32, 4736 data points, 3.5 days
8.9	2002-08-30 18:16:47, 4482 data points, 3.3 days
8.8	1999-11-15 13:33:03, 4260 data points, 3.2 days
8.6	2001-07-23 15:34:13, 4210 data points, 3.1 days
8.5	1999-03-20 01:25:02, 4153 data points, 3.1 days
8.3	2000-08-21 09:19:59, 3977 data points, 2.9 days
8.2	1998-02-20 18:35:09, 3939 data points, 2.9 days
8.0	2000-05-09 21:21:16, 3908 data points, 2.9 days
7.9	1999-01-31 22:53:40, 3597 data points, 2.7 days
7.8	1999-07-12 04:33:36, 3503 data points, 2.6 days
7.6	2002-03-15 22:23:34, 3497 data points, 2.6 days
7.5	2003-01-07 13:45:32, 3427 data points, 2.5 days
7.4	2000-12-31 09:14:22, 3356 data points, 2.5 days
7.3	2001-01-26 10:43:55, 3309 data points, 2.5 days
7.2	2001-10-17 06:07:54, 3097 data points, 2.3 days
7.1	2001-09-20 15:47:10, 3055 data points, 2.3 days
7.0	1998-08-16 21:04:09, 2888 data points, 2.1 days
6.9	2000-04-13 21:11:43, 2858 data points, 2.1 days
6.8	2001-01-19 14:23:40, 2828 data points, 2.1 days
6.7	2002-08-24 10:21:05, 2725 data points, 2.0 days
6.6	1999-12-21 22:11:08, 2676 data points, 2.0 days
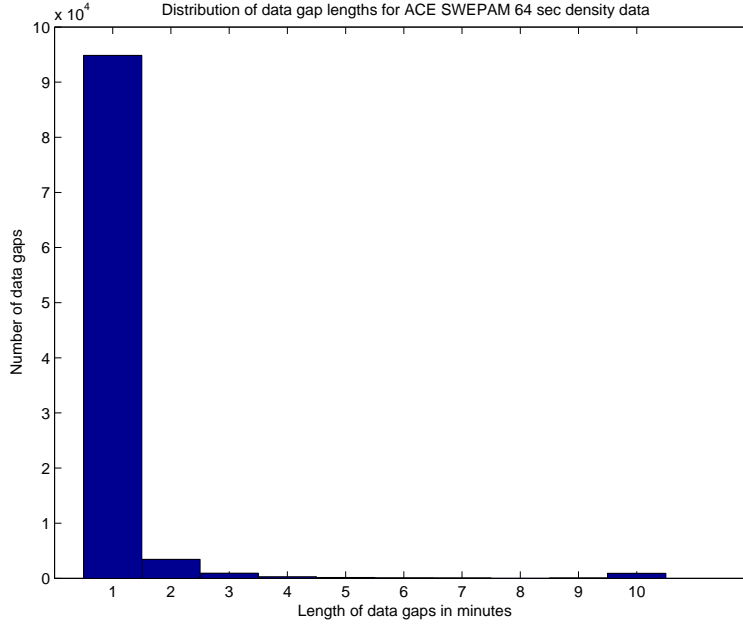
Figure 6: Distribution of data gaps lengths for the ACE SWEPAM 64 second density data.

3.3.3 Real time solar wind data

The real-time solar wind plasma and magnetic field data are merged and stored into the table `magswe` with the fields shown in Table 6. Data have been copied from the SAAPS database to the `gicpilot` database. The data covers the period 2000-05-01 to present, which at the moment of writing is 2004-01-13. There is a total of 1 947 696 data points and increases by 1440 points per day. There are 7 866 data gaps in the magnetic field data, which corresponds to 0.4%. In the plasma data there are 35 639 data gaps corresponding to 1.8% of the data.

3.4 GIC data

The GIC data are stored in the table `gic` and contains the fields shown in Table 7. The time resolution is one minute. Measurements have been made for three time periods: 1998-09-17 to 1998-10-28, 1999-08-15 to 1999-11-14, and 2000-01-22 to 2000-08-13.

For the first period (1998-09-17 02:30:00 to 1998-10-28 00:00:00) there are basically no data gaps. However, because the original data were labelled in local time there is a one-hour data gap from 1998-10-24 23:59:00 to 1998-10-25 01:00:00 due to the transition from summer time to normal time.

For the second and third periods it is stated [private communications, I. Kronfeldt] that data have only been recorded when the GIC measurement exceed 1 Ampere. However, there are 17 943 GIC values that are zero in the two data sets. To examine if there is a variation in the sampling rate over the day we plot the distributions of UT hours in the three data sets are shown in Figure 7. The

Table 6: Definition of the fields in the `magswe` table.

Field	Meaning	Unit
dt	Time stamp	year-month-date hour:minute:second
BxGSM	x-component of the magnetic field in GSM	nT
ByGSM	y-component of the magnetic field in GSM	nT
BzGSM	z-component of the magnetic field in GSM	nT
n	Particle density	cm^{-3}
V	Bulk flow speed	km/s
T	Temperature	K
lat	Latitude of magnetic field	degrees
lon	Longitude of magnetic field	degrees
MagStat	Status flag for magnetic field, > 0 indicates data gap	
SweStat	Status flag for plasma data, > 0 indicates data gap	

Table 7: Definition of the fields in the `gic` table.

Field	Meaning	Unit
dt	Time stamp	year-month-date hour:minute:second
gic	GIC value	A

top panel (first set) shows a flat distribution indicating that we have a good coverage for all hours. The third and fourth panels, second and third data sets, show a clear UT variation. To understand whether the variation is related to the fact that only data with $|GIC| \geq 1$ A are included in the second and third data sets, we also plot the distribution for the first set including only data with $|GIC| \geq 1$ A (second panel). A diurnal variation appears with peaks around midnight-morning and afternoon. Thus, the diurnal variation in the second and third data sets may be related to GIC selection effect, but it is also observed that bins with more samples coincide with office hours.

3.5   Geomagnetic data

The geomagnetic data are stored in the table `geomag` and covers the period 1998-01-01 to 2002-12-31. The stations included are listeed in Table 1. In the table we use the longitude and latitude of the station instead of the station name. The fields in the table are described in Table 8. The latitude and longitude are given in units of 1/100 of a degree so that they can be stored as integer values, but still with a high enough accuracy.

Data gaps are indicated by the value 99999.9 ($10^5 - 0.1$). Generally there are very few data gaps in the geomagnetic data set, and when they occur it is most as single one-minute gaps. It is therefore safe to interpolate data gaps to produce a continuous time series.
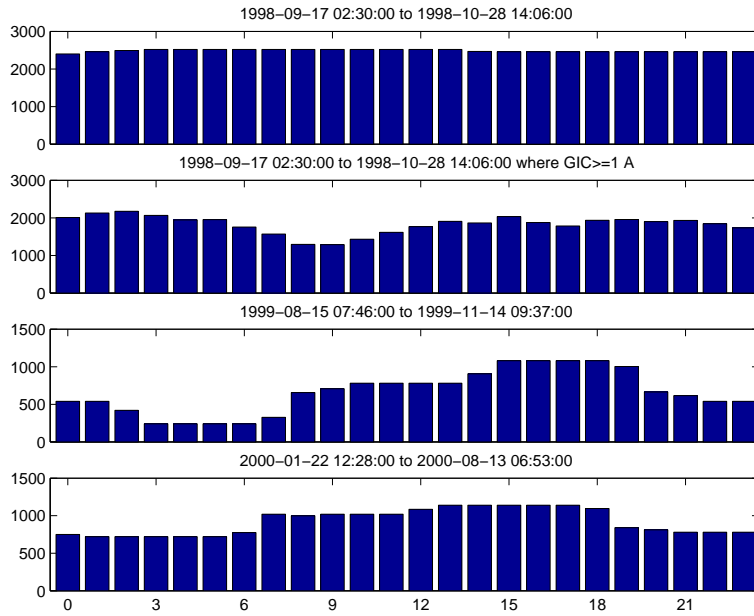
Figure 7: Distribution of UT hours for the data sets..

Table 8: Definition of the fields in the `geomag` table.

Field	Meaning	Unit
dt	Time stamp	year-month-date hour:minute:second
lon	Geographic longitude of the station	1/100 degree
lat	Geographic latitude of the station	1/100 degree
x	x-component of the magnetic field	nT
y	y-component of the magnetic field	nT
z	z-component of the magnetic field	nT

3.6 Technical issues

3.6.1 Java interface

To access the database from Java a Java Database Connector (JDBC) must be installed. The official driver for MySQL is MySQL Connector/J [16]. The driver consists of a JAR file (Java Archive) that is placed in the Java search path. The code on the following lines shows a class that can be used to connect to the database and read data.

```java
/*
 * Connector.java
 */

import java.sql.*;
import java.util.Vector;

/**
 * @author  Peter Wintoft
 */
public class Connector {

    private Connection conn = null;
    private ResultSet result;
    private String host;
    private String database;
    private String port;
    private String user;
    private String passwd;

    /** Creates a new instance of Connector */
    public Connector() {
        this.host="some_host.domain";
        this.database="db_name";
        this.port="port_number";
        this.user="a_user";
        this.passwd="a_password";
    }

    /** Register and connect to the database server.*/
    public boolean connect() throws SQLException {
        //Register the MySQL driver.
        try
        {
            Class.forName("com.mysql.jdbc.Driver").newInstance();
        }
        catch(ClassNotFoundException e) {
```

```java
            System.err.println(e);
        }
        catch(InstantiationException e) {
            System.err.println(e);
        }
        catch(IllegalAccessException e) {
            System.err.println(e);
        }
        //Connect to the database.
        String c = "jdbc:mysql://"+this.host+"/"+this.database+":"+this.port;
        this.conn = DriverManager.getConnection(c,this.user,this.passwd);
        if (conn == null)
            return false;
        else
            return true;
    }


    /** Get the data that resulted from a previous query.*/
    public Vector getData() throws SQLException {
        int c = this.getNumOfColumns();
        Vector rows = new Vector();
        while (this.result.next()) {
            Vector row = new Vector(c);
            for(int i=1;i<=c;i++)
                row.add(this.result.getString(i));
            rows.add(row);
        }
        return rows;
    }


    /**
    * Get a description, the meta-data, of the data
    * that resulted from a previous query.
    */
    public Vector getMetaData() throws SQLException {
        ResultSetMetaData rsmd = this.result.getMetaData();
        int c = rsmd.getColumnCount();
        Vector v = new Vector(c);
        for(int i=1;i<=c;i++)
            v.add(rsmd.getColumnName(i));
        return v;
    }


    /** Get the number of data columns. */
    public int getNumOfColumns() throws SQLException {
        ResultSetMetaData rsmd = this.result.getMetaData();
        return rsmd.getColumnCount();
    }
```

```
    /** Submit an SQL query. */
    public void query(String q) throws SQLException
    {
        PreparedStatement statement = this.conn.prepareStatement(q);
        this.result = statement.executeQuery();
    }


    /**
     * Run a Connector object using the argument string that
     * is given on the command line as the SQL query.
     *
     * The resulting Vector object is written to standard out.
     *
     * Example:
     *    java Connector "show tables;"
     */
    public static void main(String[] args) {
        Connector c = new Connector();
        Vector data = null;
        Vector metaData = null;
        try {
            c.connect();
            c.query(args[0]);
            data = c.getData();
            metaData = c.getMetaData();
        }
        catch(SQLException e) {
            System.err.println(e);
        }
        System.out.println(metaData);
        System.out.println(data);
    }
}
```

3.6.2   Matlab interface

To access the database from Matlab a special function must be used that is available in the search path. To connect to the database the `mysql` function is executed with the following arguments:

```
matlab>> mysql('open',host,user,password)
```

To load data the SQL query is passed as a string to the `mysql` function. The number of columns that the query results in must be matched by the same number of output arguments from the function call, e.g.:

```
matlab>> [d,bz] = mysql('select dt,bzgsm from mag16s where
dt>="2000-01-01 00:00:00" and dt<"2000-02-01 00:00:00"');
```

3.6.3 Database privileges

The database has been set up with three different levels of user privileges using the MySQL `GRANT` command. At the top level is the administrator of the database that have full access and may add and remove tables, alter tables, add and remove data from tables, add and remove users, etc. At the next level we have the database operator that may read and write data to/from the database. The operator is not allowed to remove data. At third level there is the user who may only read data from the database.

3.6.4 Performance

The tables in the database that contain time series data all have a `DATETIME` field that is defined as a `PRIMARY KEY`. This means that retrieving data for a selected time range is fast. The following example shows this:

```
mysql> select count(*) from mag16s where dt>="2000-01-01 00:00:00"
and dt<"2000-02-01 00:00:00";
+----------+
| count(*) |
+----------+
|   167400 |
+----------+
1 row in set (0.42 sec)
```

If we instead retrieve data with the search criteria on some other field the whole data set must be searched. The example below results in approximately the same number of rows but it now takes 40 times longer.

```
mysql> select count(*) from mag16s where bzgsm>0 and bzgsm<0.1;
+----------+
| count(*) |
+----------+
|   132293 |
+----------+
1 row in set (18.33 sec)
```

References

[1] P. Wintoft, H. Lundstedt, and R. Pirjola, "A space weather forecast system for geomagnetically induced currents," Tech. Rep. 1.5, Swedish Institute of Space Physics, Scheelevägen 17, 223 70 Lund, Sweden, 2003.

[2] H. Lundstedt and P. Wintoft, "Real-time forecast service for geomagnetically induced currents." http://www.lund.irf.se/gicpilot/, 2004.

[3] H. Gleisner and H. Lundstedt, "Auroral electrojet predictions with dynamic neural networks," *Journal of Geophysical Research*, vol. 106, pp. 24,541–24,550, 2001.

[4] F. Boberg, P. Wintoft, and H. Lundstedt, "Real time kp predictions from solar wind data using neural networks," *Physics and Chemistry of Earth*, vol. 25, no. 4, pp. 275–280, 2000.

[5] H. Lundstedt, H. Gleisner, and P. Wintoft, "Operational forecasts of the geomagnetic dst index," *Geophysycal Research Letters*, vol. 29, pp. 34–1–34–4, 2002.

[6] A. Davis, "Advanced composition explorer (ace)." http://www.srl.caltech.edu/ACE, 2004.

[7] A. S. Center. http://www.srl.caltech.edu/ACE/ASC/level2/index.html, 2004.

[8] L. Combs and B. Poppe, "Space environment center." http://www.sec.noaa.gov, 2004.

[9] V. Raben, "The space environment center anonymous ftp server." ftp://ftp.sec.noaa.gov/pub/lists/ace, 2004.

[10] P. Wintoft and L. Eliasson, "Saaps, satellite anomaly analysis and prediction system, tech. note 1, database and database tools," tech. rep., ESA/ESTEC Contract No. 13561/99/NL/SB, 2001.

[11] A. Pulkkinen, O. Amm, A. Viljanen, and B. W. Group, "Ionospheric equivalent current distributions determined with the method of spherical elementary current systems," *J. Geophys. Res.*, vol. 108, p. doi: 10.1029/2001JA005085, 2003.

[12] A. Pulkkinen, *Geomagnetic induction during highly disturbed space weather conditions: Studies of ground effects*. PhD thesis, Finnish Meteorological Institute, 2003.

[13] A. Viljanen, A. Pulkkinen, O. Amm, R. Pirjola, T. Korja, and B. W. Group, "Fast computation of the geoelectric field using the method of elementary current system," *Ann. Geophys.*, p. Accepted for publication, 2003.

[14] I. Kronfeldt, "Forecasting and calculating geomagnetically induced currents," tech. rep., Sydkraft AB, 2002.

[15] P. DuBois, *MySQL, The definitive guide to using, programming, and administering MySQL 4 databases*. Sams Publishing, 2003.

[16] D. Axmark, A. Larsson, and M. Widenius, "Mysql: The world's most popular open source database." http://www.mysql.com/, 2004.