PRediction Of Geospace Radiation Environment and Solar wind parameterS

Work Package 3
Forecast of the evolution of geomagnetic indices

Deliverable 3.2
Identify and collect relevant data

P. Wintoft and S. Walker
June 26, 2015

This project has received funding from the *European Unions Horizon 2020 research and innovation programme* under grant agreement No 637302.

Document Change Record

Issue	Date	Author	Details
v1	April 10, 2015	P. Wintoft	Initial draft
v2	May 6, 2015	P. Wintoft	Updated draft
v3	June 25, 2015	P. Wintoft	Final draft
v4	June 26, 2015	P. Wintoft and S. Walker	Final version

Contents

Summary

The overall aim of WP 3 concerns improvement and new development of models based on data driven modelling, such as CNN and NARMAX. Existing models for *Dst* and *Kp* will be analysed and verified with the aim of finding weaknesses and to suggest improvements. Solar wind and geomagnetic indices shall also be analysed in order to develop models for the identification of features, such as (but not limited to) shocks, sudden commencements, and substorms. Such categorisation will aid the model development and verification, and can also serve as alternative approach to models providing numerical input-output mapping. In addition to the development of *Dst* and *Kp* models new models will be developed to forecast *AE*. The models will be implemented for real-time operation at IRF and data and plots will be provided on a web server.

This deliverable is targeted at collecting historic real time ACE data, Science Level 2 ACE data, Kp, Dst, and AE. An SQL database shall be set up where the data are collected. The data sets shall be analysed with respect to quality and coverage. Also include the coming DSCOVR spacecraft in the study.

Acronyms

ACE	Advanced Composition Explorer
DSCOVR	Deep Space Climate Observatory
GFZ	GeoForschungsZentrum
GSFC	Goddard Space Flight Center
NASA	National Aeronautic and Space Administration
NCEI	National Centers for Environmental Information
NOAA	National Oceanographic and Atmospheric Administration
SWPC	Space Weather Prediction Center
WDC	World Data Center

1 Introduction

The development and verification of models are crucially dependent on high quality data that captures the processes under study and with good temporal coverage. As outlined in the WP description the solar wind ACE and *Kp*, *Dst*, and *AE* indices should be addressed.

2 Datasets – archived data

The relevant archived data are maintained at WDC-Kyoto[1] for the *Dst* and *AE* indices, GFZ Potsdam[2] for the *Kp* index, and the ACE science center[3] for the Level 2 ACE solar wind data. The National Center for Environmental Information (NCEI, former NGDC)[4]

[1] `http://wdc.kugi.kyoto-u.ac.jp/`
[2] `http://www.gfz-potsdam.de/`
[3] `http://www.srl.caltech.edu/ACE/ASC/`
[4] `http://www.ngdc.noaa.gov/stp/geomag/indices.html`

also holds the indices from WDC-Kyoto.

At NASA/GSFC compiled datasets containing solar wind data and indices are provided through OMNIWeb[5]. The OMNI datasets come with three different temporal resolutions: hourly, 5 minute, and 1 minute. The solar wind data come from different spacecraft, such as ACE, IMP-8 and WIND. The solar wind data have been propagated from the spacecraft location to a point close to the Earth bow shock nose, see Section 4 for more details. The OMNI datasets also contain the geomagnetic indices Kp, Dst, AE, AU, and AL. The indices have been obtained from NCEI, which means that they are the same as provided by the official sources. Thus, the OMNI datasets can be used for the propagated solar wind data and for the indices.

The ACE Level 2 magnetic field and plasma data have temporal resolution of 16 and 64 seconds, respectively. At the ACE science center the magnetic field and plasma data have also been merged into one dataset with 64 second resolution, which is the dataset used in this study. As several other datasets have one minute cadence a dataset with one minute resolution ACE data is also created. The values at one minute cadence are linearly interpolated from the 64 second data.

Missing data is always troublesome for time series analysis and modelling. For the ACE data period, around 1998 and onwards, there are no missing data for the Kp, Dst, and AE index series. However, the solar wind data series have datagaps and an analysis is made in Section 4.

A summary of the datasets are given in Section 6.

3 Datasets – real time data

The ACE real time data are provided by SWPC[6]. The ACE magnetic field data, in GSM coordinates, are provided with a cadence of one minute. The plasma data also have one minute cadence with proton density, speed, and ion temperature. Note that only speed is given, not velocity vector. The data typically lags real time by a couple of minutes.

In the future solar wind real time data will be provided by SWPC from the DSCOVR spacecraft. DSCOVR entered the L1 orbit on June 7, 2015, and it is estimated instrument calibration will take about 30 days. We expect that data will be provided from SWPC in a similar way as for the ACE spacecraft, and as described by NOAA[7]:

> The DSCOVR mission is a partnership between NOAA, NASA, and the U.S. Air Force. The Air Force provided the Space X Falcon 9 launch vehicle for the mission. NOAA will operate DSCOVR from its NOAA Satellite Operations Facility in Suitland, Maryland, and process the space weather data at NOAAs Space Weather Prediction Center (SWPC) in Boulder, Colorado, one of NOAAs nine National Centers for Environmental Prediction. SWPC will distribute these data to users within the United States and around the world. The data will be archived at NOAAs National Center for Environmental Information.

[5] `http://omniweb.gsfc.nasa.gov/hw.html`

[6] `http://www.swpc.noaa.gov/`

[7] `http://www.nesdis.noaa.gov/news_archives/DSCOVR_L1_orbit.html`

The estimated *Kp* data are also provided by SWPC. The data are provided in near real time and typically lag by a couple of hours.

The real time *Dst* data are provided by WDC-Kyoto. Data are listed for the current month and typically lags real time by a couple of hours. Note that *Dst* is listed with time stamps 1–24 UT, which means that time stamps mark the end of each one hour interval. In the PROGRESS database they are instead stored with time stamps 0–23 UT marking the beginning of each interval.

Real time *AE*, *AL*, and *AU* are also provided by WDC-Kyoto. However, currently only plots are available.

4 Solar wind ACE and OMNI sets

4.1 A summary of OMNI propagation results

The OMNI datasets consist of solar wind data from several different spacecraft time shifted to a location close to the Earth's bow shock nose King & Papitashvili (2005). The OMNIweb contains a detailed description[8] of different techniques used to shift the data from the spacecraft location to the location of Earth's bow shock nose. The advantage of using the time shifted data is that the L1–Earth propagation delay need not to be handled for modelling and analysis of solar wind–magnetosphere interaction. Here follows a summary of the discussions from the web page.

The time shifting of the data builds on the assumption that the solar wind is organised in phase fronts that convect from the spacecraft location to the Earth bow shock nose. The bow shock position is determined from the model by Farris & Russell (1994). To determine the phase fronts the phase front normals (PFN) are found using a minimum variance analysis (MVA) (Sonnerup et al. 2004). Five different techniques to find the PFNs are discussed: 1) "Modified" MVA (Weimer et al. 2003, Bargatze et al. 2005); 2) MVAB-0 (Haaland et al. 2006); 3) Cross Product (based on Knetter et al. (2004)); 4) a combination of 2) and 3) where 3) is used if it gives an acceptable PFN, else 2) is used; 5) both 2) and 3) must give the same PFN. The final OMNI datasets have been created using technique 4.

The time shifting technique does not change the values of the magnetic field or plasma data, it only changes the associated time tags. However, on occasions with large time derivatives in velocity the time shifted plasma elements may overtake earlier elements. The approach to handle this is to take the average of all magnetic or plasma data that belong to the same time shifted one minute interval.

The techniques were tested by shifting ACE data to the location of WIND, for the years 1998–2000, and comparing the shifted ACE data to the unshifted WIND data. For each parameter a cross correlation function was applied in bins as function of the impact parameter (IP). The IP is a measure of the transverse distance between the ACE and WIND spacecraft. The result is shown in Table 1. Thus, for IP$<$ 30 Re the correlation is high, but then falls off with increasing IP.

[8] `http://omniweb.gsfc.nasa.gov/html/HROdocum.html`

Table 1: Cross correlation between propagated ACE to WIND and WIND observations of Bz binned as function of impact parameter (IP).

IP (Re)	0 – 15	15 – 30	30 – 60	60 – 90	90 – 120	120 – 150	> 150
C	0.87	0.85	0.75	0.63	0.54	0.34	0.35

The cross correlation was also studied as function of the variability in Bz using the standard deviation σ_{Bz}. Only data with IP< 60 Re were used. The cross correlation increased as function of σ_{Bz} as seen in Table 2.

Table 2: Cross correlation between propagated ACE to WIND and WIND observations of Bz binned as function of σ_{Bz} for IP< 60 Re.

σ_{Bz} (nT)	0 – 1	1 – 2	2 – 3	3 – 4	> 4
C	0.66	0.76	0.82	0.85	0.91

The cross correlation as function of the distance in the X direction between ACE and WIND was computed and is shown in Table 3. There is not much variability as function of distance.

Table 3: Cross correlation between propagated ACE to WIND and WIND observations of Bz binned as function of distance in the X direction for IP< 60 Re.

Distance (Re)	< 50	50 – 125	125 – 200	> 200
C	0.78	0.77	0.81	0.74

Finally, the cross correlation as function of speed was studied for IP< 60 and $\sigma_{Bz} > 1$ nT, and the result is shown in Table 4. There is a slight decrease in correlation for increasing speed.

4.2 Discussion of the OMNI propagation

The previous section shows that the impact parameter (IP) has the largest effect on the Bz correlation between ACE and WIND, a result that we also expect to hold for any spacecraft around L1 and Earth. Accepting a correlation of 0.7, which is rather low, indicates that IP should be less than 60 Re for the data to be useful.

When the variability in Bz is small ($\sigma_{Bz} < 1$ nT) the correlation is low, which is expected and is not of any real concern.

The upstream distance between ACE and WIND does not have any significant effect on the correlation and it is expected the same will be true for L1–Earth.

Table 4: Cross correlation between propagated ACE to WIND and WIND observations of Bz binned as function of speed for IP< 60 Re and $\sigma_{Bz} > 1$ nT.

Speed (km/s)	< 350	$350 - 450$	$450 - 550$	> 550
C	0.84	0.83	0.79	0.72

For increasing speed the correlation consistently drops, although Table 4 indicates a correlation above 0.7. However, the last bin contains all speeds above 550 km/s, which is a rather low limit. Most likely the correlation will continue to drop for higher speeds as the effect of overtaking plasma elements will increase.

4.3 Missing magnetic field data

The number of missing 1-minute values in the OMNI dataset is much larger that the number of missing values in the ACE dataset. For the period 1997-09-02 to 2015-05-11 there are 23 759 missing magnetic field values in the ACE set, while there are 666 913 missing magnetic field values in the OMNI set. In total there are 9 302 400 minutes in the ACE set, thus the fraction of missing values are 0.26% for ACE and 7.2% for OMNI. The number of missing values per year for the two datasets are shown in Figure 1.

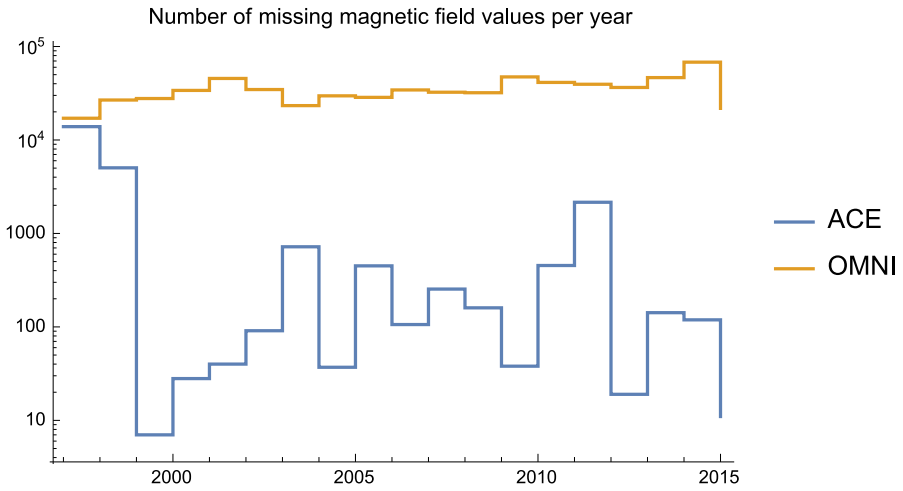

Figure 1: Number of missing 1-minute values in magnetic field data per year for the ACE and OMNI sets.

Next, we define a datagap as a contiguous sequence of missing values, thus a datagap is associated with a start time and an end time, and extends over a period of time. If only one 1-minute value is missing then the datagap will have a length of one minute. Figure 2 shows the lengths of datagaps longer than 10 minutes for the ACE set. It is seen that the coverage is very good, e.g. datagaps longer than a day (1440 minutes) are only seen before 1999. The longest datagap occur for the period 1997-12-08 – 1997-12-14.

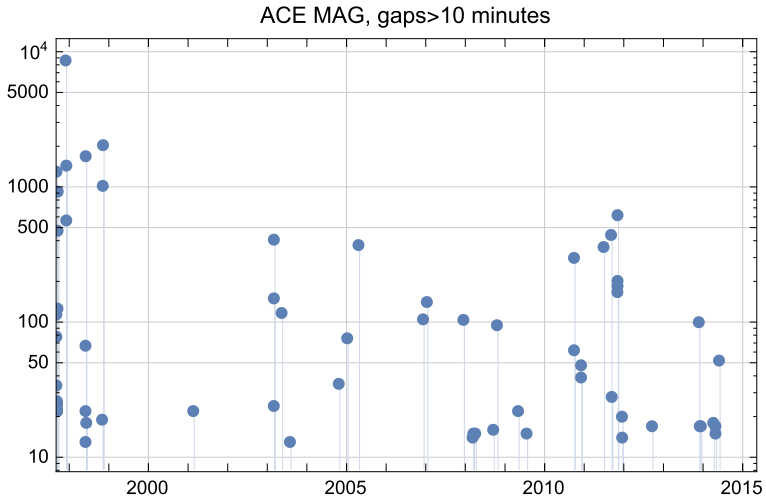

Figure 2: Length of datagaps, in minutes, for ACE magnetic filed data. Only gaps longer than 10 minutes are shown.

The OMNI magnetic field set contains a large number of datagaps and gaps longer than 180 minutes are shown in Figure 3. There are many gaps with lengths larger than a day.

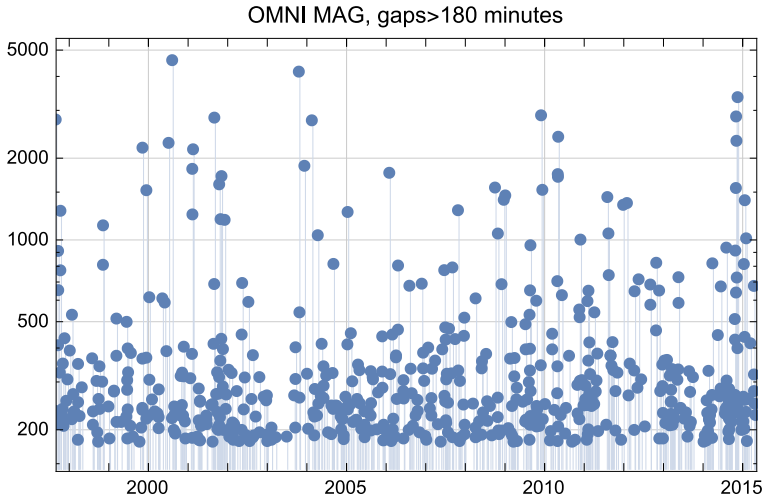

Figure 3: Length of datagaps, in minutes, for OMNI magnetic filed data. Only gaps longer than 180 minutes are shown.

Periods with disturbances caused by CMEs and the related geomagnetic storms are the most interesting. However, depending on storm level, these events are rare. E.g., the fraction of Kp values larger than 5 is only 4%. The fraction of 3-hour intervals with more than 80% magnetic field data in the solar wind as function of Kp level is shown in Figure 4. For the ACE set all 3-hour intervals have better than 80% coverage for all $Kp \geq 5$. At the the $Kp = 9$ level there are 23 cases. The OMNI set does not have full

coverage, and especially important for large Kp the fraction drops.

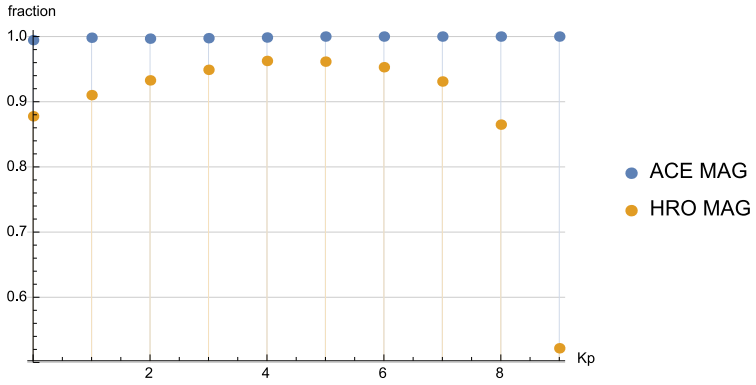

Figure 4: Fraction of magnetic field data with more than 80% data coverage for each 3-hour Kp interval as function of Kp.

4.4 Missing plasma data

For the plasma data the OMNI set has better coverage than the ACE set before 1999 as other spacecraft data have been used. But after that the ACE set has better coverage (Figure 5).

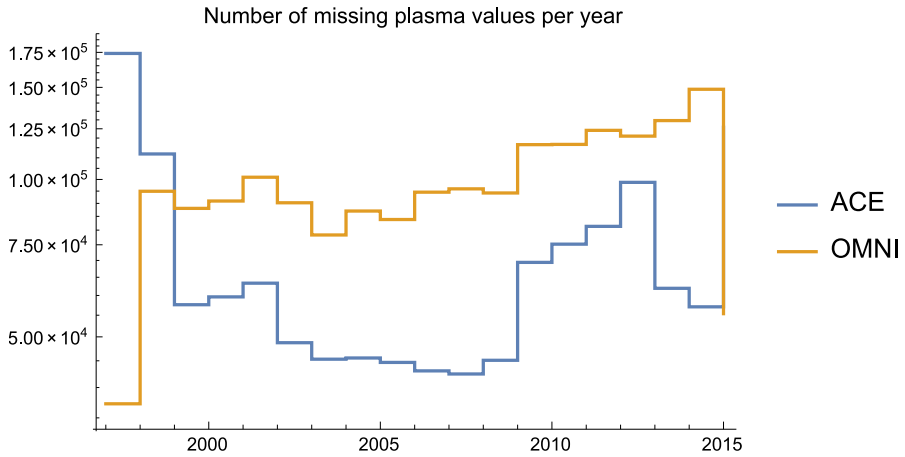

Figure 5: Number of missing 1-minute values in plasma data per year for the ACE and OMNI sets.

The distribution of ACE plasma datagap lengths longer than 60 minutes is shown in Figure 6. Clearly there are more gaps in the ACE plasma set than in the magnetic field set. The two longest gaps occur at 1997-09-02 – 1998-02-05 and 2015-02-18 – 2015-05-11. Not counting the two longest gaps, there are 25 gaps that are longer than a day.

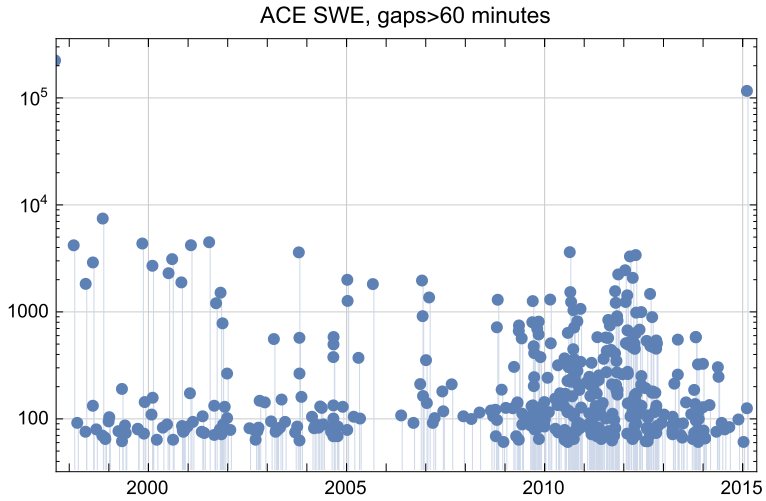

Figure 6: Length of datagaps, in minutes, for ACE magnetic filed data. Only gaps longer than 60 minutes are shown.

The OMNI plasma gap lengths are shown in Figure 7 for gaps larger than 180 minutes. There are more gaps in OMNI compared to ACE, but the number of gaps longer than a day is similar and equal to 26.

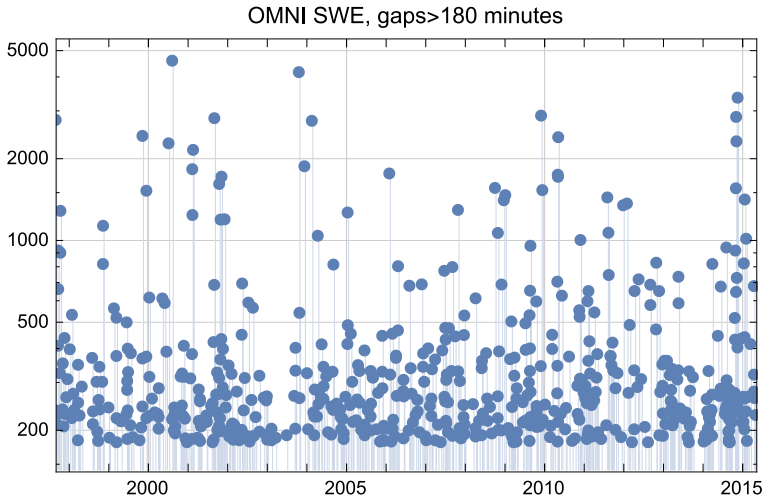

Figure 7: Length of datagaps, in minutes, for OMNI magnetic filed data. Only gaps longer than 180 minutes are shown.

The fraction of plasma data with more than 80% coverage for each 3-hour interval as function of Kp is shown in Figure 8. For high levels, $Kp \geq 8$, the fraction is similar for ACE and OMNI, while at lower levels the ACE coverage is better.

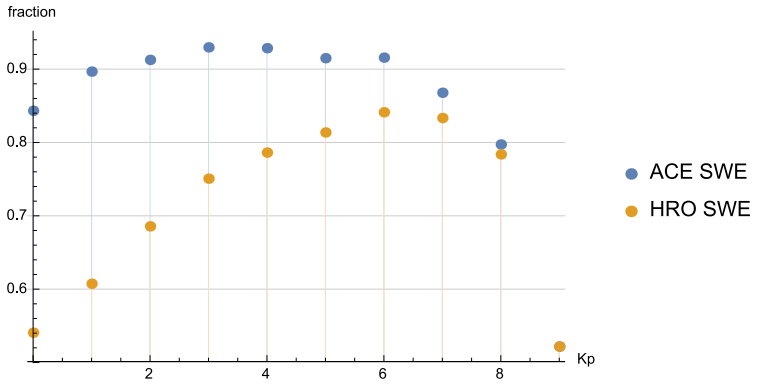

Figure 8: Fraction of plasma data with more than 80% data coverage for each 3-hour Kp interval as function of Kp.

5 Database

All data are stored in a PostgreSQL[9] database. Each dataset is stored in a table with names following the syntax given in Section 6, e.g. the ACE Level 2, 64 second, data are stored in a table with name `ace_l2_64s`. Data can then be efficiently retrieved either directly or over a REST service. Details to access the data will be given at `http://corona.lund.irf.se/progress/internal/`.

6 Summary of datasets

Here a summary of the datasets that will be used in WP3 is given. The source location and format of data of each set is provided. Each dataset is then downloaded and stored in the SQL database.

6.1 OMNI – hourly resolution

ID	omni-1h
Resolution	1 hour
Range	1963–2015
Format	Text
`ftp://spdf.gsfc.nasa.gov/pub/data/omni/low_res_omni/`	
Hourly averaged solar wind data from multiple spacecraft advected to Earth bow shock nose position. Dataset also includes other data such as indices Kp, Dst AE, AL, and AU for each hour. The main web site is at `http://omniweb.gsfc.nasa.gov/ow.html`.	

[9] `http://www.postgresql.org`

6.2 High resolution OMNI – one minute resolution

ID	omni-1m
Resolution	1 minute
Range	1981–2015
Format	Text
`ftp://spdf.gsfc.nasa.gov/pub/data/omni/high_res_omni/`	
One minute averaged solar wind data from multiple spacecraft advected to Earth bow shock nose position. Dataset also includes other data such as indices AE, AL, and AU. The main web site is at `http://omniweb.gsfc.nasa.gov/ow_min.html`.	

6.3 High resolution OMNI – five minute resolution

ID	omni-5m
Resolution	5 minutes
Range	1981–2015
Format	Text
`ftp://spdf.gsfc.nasa.gov/pub/data/omni/high_res_omni/`	
Five minute averaged solar wind data from multiple spacecraft advected to Earth bow shock nose position. Dataset also includes other data such as indices AE, AL, and AU. The main web site is at `http://omniweb.gsfc.nasa.gov/ow_min.html`.	

6.4 ACE Level 2 – 64 s merged MAG and SWE

ID	ace-l2-64s
Resolution	64 seconds
Range	1997–2015
Format	HDF
`ftp://mussel.srl.caltech.edu/pub/ace/level2/magswe/`	
`http://www.srl.caltech.edu/ACE/ASC/level2/lvl2DATA_MAG-SWEPAM.html`.	

6.5 ACE Level 2 – one minute resampled

ID	ace-l2-1m
Resolution	1 minute
Range	1997–2015
Format	SQL
See ace-l2-64s	
ACE 64 second data resampled to one minute resolution using linear interpolation.	

6.6 ACE real time plasma data – one minute resolution

ID	ace-swe-rt-1m
Resolution	1 minute
Range	Past two hours
Format	Text
`http://services.swpc.noaa.gov/text/ace-swepam.txt`	
ACE real time proton density, speed, and ion temperature.	

6.7 ACE real time magnetic field data – one minute resolution

ID	ace-mag-rt-1m
Resolution	1 minute
Range	Past two hours
Format	Text
`http://services.swpc.noaa.gov/text/ace-magnetometer.txt`	
ACE real time magnetic field in GSM coordinates.	

6.8 Estimated *Kp* real time data – 3 hour resolution

ID	kp-rt-3h
Resolution	3 hours
Range	Past month
Format	Text
`http://services.swpc.noaa.gov/text/daily-geomagnetic-indices.txt`	
Estimated *Kp* index.	

6.9 *Dst* real time data – 1 hour resolution

ID	dst-rt-1h
Resolution	1 hour
Range	Current month
Format	Text
`http://wdc.kugi.kyoto-u.ac.jp/dst_realtime/presentmonth/index.html`	
Real time *Dst* index.	

References

Bargatze, L. F., McPherron, R. L., Minamora, J. & Weimer, D. (2005), 'A new interpretation of weimer et al.'s solar wind propagation delay technique', *Journal of Geophysical Research* **110**, A07105.

Farris, M. H. & Russell, C. T. (1994), 'Determining the standoff distance of the bow shock: Mach number dependence and use of models', *Journal of Geophysical Research* **99**(A9), 17,681–17,689.

Haaland, S., Paschmann, G. & Sonnerup, B. U. Ö. (2006), 'Comment on "a new interpretation of weimer et al.'s solar wind propagation delay technique" by bargatze et al.', *Journal of Geophysical Research* **111**, A06102.

King, J. H. & Papitashvili, N. E. (2005), 'Solar wind spatial scales in and comparisons of hourly wind and ace plasma and magnetic field data', *Journal of Geophysical Research* **110**, A02104.

Knetter, T., Neubauer, F. M., Horbury, T. & Balogh, A. (2004), 'Four point discontinuity observations using cluster magnetic field data: A statistical survey', *Journal of Geophysical Research* **109**, A06102.

Sonnerup, B. U. Ö., Haaland, S., Paschmann, G., Lavraud, B., Dunlop, M. W., Réme, H. & Balogh, A. (2004), 'Orientation and motion of a discontinuity from single-spacecraft measurements of plasma velocity and density: Minimum mas flux residue', *Journal of Geophysical Research* **109**, A03221.

Weimer, D. R., Ober, D. M., Maynard, N. C., Collier, M. R., McComas, D. J., Ness, N. F., Smith, C. W. & Watermann, J. (2003), 'Predicting interplanetary magnetic field (imf) propagation delay times using the minimum variance delay technique', *Journal of Geophysical Research* **108**(A1), 1026.