Real-time forecast service for geomagnetically induced currents
WP 400 Technical Note
Forecasting dH/dt and GIC from solar wind data

P. Wintoft

October 5, 2004

Document status sheet

ISSUE	REVISION	DATE	REASON FOR CHANGE
0	0	April 26, 2004	First draft.
0	1	October 5, 2004	Most sections updated.

Summary

This document describes the models that will produce forecasts of the rate-of-change (ΔH) of the horizontal components of the local geomagnetic field in South Sweden based on ACE real time solar wind data.

It is clear that predicting ΔH with one minute resolution is with current knowledge impossible. Therefore, we motivate the use of temporal root-mean-square (RMS) ΔH formed over 10 minute intervals. A resolution of 10 minutes has been found to be a good trade off between high resolution and accurate forecasts. The optimal forecast lead time is 60 minutes and the correlation between model output and observed geomagnetic data is 0.80 ± 0.05 in $\log \mathrm{RMS}\Delta H$ and ? in $\mathrm{RMS}\Delta H$.

[Text to be added on "Computed ΔH in a dense grid over South Sweden.]

Finally, we provide a linear model relating RMS ΔH at Brorfelde, Denmark, and Uppsala, Sweden, to RMS GIC at a single location. There is also a close linear relation between RMS GIC and MAX GIC, where the latter is the maximum GIC in a 10 minute interval. This is useful as an estimate of the maximum GIC that will occur.

Acronyms

ACE	Advanced Composition Explorer
ANN	Artificial Neural Network
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
ETM	Exponential Trace Memory
GIC	Geomagnetically Induced Current
MODWT	Maximum Overlap Discrete Wavelet Transform
TN	Technical Note
URD	User Requirements Document
WP	Work Package

URD matrix

The table below lists the requirements from the URD [2] together with the relevant sections in this document.

URD req.	Section
CAP.1.	TBD
CAP.2.	TBD
CAP.3.	TBD
CAP.4.	TBD
CAP.5.	TBD
CAP.6.	TBD
CAP.7.	TBD
CAP.8.	TBD
CAP.9.	TBD
CAP.10.	TBD
CAP.11.	TBD
CAP.12.	TBD
CAP.13.	TBD
CAP.14.	TBD
CAP.15.	TBD
CAP.16.	TBD
CON.1.	TBD
CON.2.	TBD
CON.3.	TBD
CON.4.	TBD
CON.5.	TBD

Contents

1 Introduction

This document describes a module for direct solar wind – dH/dt that shall be a part of the GIC Pilot Project [1]. The requirements have been identified in the *User Requirements Document* [2]. 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, magnetosphere, 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 a public part.

Our approach to develop a forecast model for GIC follows two alternative routes. In the first approach, described in this document, a model is develop that predicts the time derivative of the ground geomagnetic field. Using the predicted ground magnetic field the GIC can then be computed [3]. In the second approach a model is developed that predicts the GIC directly from solar wind data [4]. It is then possible to explore the difference between the two techniques and to identify the weakest link.

In the following sections we describe the data used, the analysis of the data, and the development of the model.

2 Averages and prediction lead time

Having a physical time dependent parameter $x(t)$ that is collected with a sampling interval Δt results in the time series x_i. The corresponding time stamp t_i marks the beginning of the interval so that x_i is the average of $x(t)$ over the interval $t \in [t_i, t_{i+1}]$ where $t_{i+1} = t_i + \Delta t$. Similarly, we may have another variable $y(t)$ sampled to y_i. If we now wish to develop a model that predicts y from x with some lead time τ we have $\hat{y}(t+\tau) = f(x(t))$, where $\hat{y}$ is the prediction of y. This leads to the discrete model

$$\hat{y}_{i+n} = f(x_i) \tag{1}$$

where $\tau = n\Delta t$.

To understand the true forecast time assume that the current time is t_0. The latest input is x_{-1} that has been collected over the time interval $[t_{-1}, t_0]$. With a forecast time of $\tau = n\Delta t$ we will thus be forecasting y_{n-1} resulting in a true forecast time of $\tau' = \tau - \Delta t$. In order for the model to perform actual forecasts we must have $\Delta t \leq \tau$.

In the case of solar wind – magnetosphere coupling, part of the lead time is associated with the solar wind travel time from L1 to Earth. In Figure 1 the travel time is shown for velocities in the range $[300, 1000]$ km/s. If we only consider velocities up to 830 km/s then it is possible to make forecasts of 30 minutes. The ACE 2-minute average velocity exceeds 830 km/s in 27 events for the period 1998 to current, and the maximum velocity is 980 km/s corresponding to a travel time of 25 minutes. Thus, using a 30 minutes forecast lead time will capture most of the events, and for higher velocities the lead time will be shifted by mostly 5 minutes.

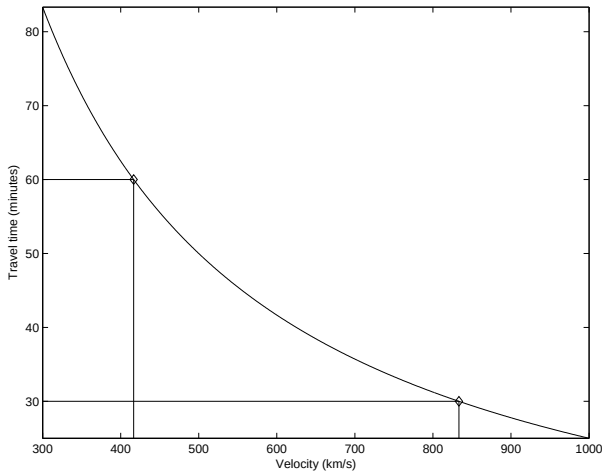

Figure 1: The figure shows the prediction lead time as a function of velocity as measured at L1. The straight lines marks the velocities at lead times of 30 and 60 minutes, respectively.

The ACE spacecraft is not located exactly on the Sun-Earth line but is on an orbit around L1. Therefore, the spacecraft does not measure the solar wind directly upstream from the Earth which will introduce uncertainties on the time of arrival and evolution of solar wind structures. Temporal averaging will reduce the uncertainties and a resolution of $\Delta t = 10$ minutes is a good trade-off

leading to a true forecast time of $30 - 10 = 20$ minutes.

3 The data

A database has been set up as described in the TN of WP 200. Here we make a short summary of the data that are used in this TN.

3.1 Solar wind data

The solar wind plasma and magnetic field data comes in two different temporal resolutions [5]: 16 second sampling of the magnetic field, and 64 second sampling of the plasma. Both the GIC and the geomagnetic data have 60 second resolution. The real-time solar wind data at SEC [6] are also given with 60 second resolution, where the plasma data have been resampled from 64 seconds to 60 seconds. However, resampling the data introduces artificial frequencies that result in differences between longer temporal averages formed from the 64- and 60-second time series. E.g., forming 5-minute averages from the 64- and 60-second data may result in occasional differences of more than 8 cm^{-3} in the density and more than 20 km/s in the velocity. Thus, resampling the data from 64 seconds to 60 seconds should be avoided for the analysis and model development.

3.2 Local geomagnetic field

We select observed geomagnetic from Brorfelde (11.67°,55.63°) and Uppsala (17.35°,59.90°) as these sites are closest to the region of Sweden we currently shall study. As stated in WP 300 the prime quantity to use when calculating GIC is the time derivatives of the horizontal magnetic field components dX/dt and dY/dt. The derivatives are approximated using the forward difference

$$\Delta X(t) = X(t+1) - X(t), \tag{2}$$

where t is in minutes and X is the north-south horizontal magnetic field component. Similarly we have

$$\Delta Y(t) = Y(t+1) - Y(t), \tag{3}$$

east-west component.

3.3 Computed geomagnetic field in a dense grid

[Text to be added.]

3.4 GIC data

As described in the database TN [5] the recorded GIC data covers three periods: 1998-09-17 – 10-28, 1999-08-15 – 11-14, and 2000-01-22 – 08-13. The first period has very few data gaps, while for the last two periods it has been stated that data has only been collected when GIC$>$ 1 Ampere.

4 Analysis

4.1 A case study – September 1998

[Text to be added.]

4.2 Auto-correlation

The auto-correlation for $T = 2$-minute average solar wind and ΔX show very different characteristics. The velocity V has an auto-correlation close to 1 for time lags ranging from 0 to 40 minutes. Thus, two measurements of V separated by 40 minutes are most of the time close to equal. Only for occasional shocks there might be a big difference in $V(t)$ and $V(t+2)$. Similarly, the density n is also highly correlated with a correlation of almost 0.9 at 40 minutes. The magnetic field component B_z has an auto-correlation that drops of quicker, but it is still above 0.6 at 40 minutes lag. Finally, the auto-correlation of ΔX is close to 0 for all time lags, thus it is not possible to predict $\Delta X(t+\tau)$ from $\Delta X(t)$ with a linear model, for any $\tau \geq 2$ minutes.

4.3 Wavelet variance – estimating the spectral density of ΔX and ΔY

As already mentioned, any temporal averaging of ΔX is not meaningful because of the weak auto-correlation. Therefore we will instead study the level of disturbance in ΔX and its relation to the solar wind. In the paper by [7] models where developed for the coupling from the solar wind to 30 minute averages of the absolute value $|\Delta X|$. Using averages of $|\Delta X|$ will capture the average disturbance level but the spectral information is lost.

Using a discrete wavelet transform (DWT) the ΔX can be decomposed into signals, called details and smooth (or approximation), that are associated with different scales, where the scale corresponds to a wavelength band. The decomposed signals can thus be thought of being a band pass filtered versions of ΔX. The DWT preserves the power in the signal but it is not time invariant, i.e. the DWT of a time shifted ΔX is not equal to the time shifted DWT of ΔX. To ensure time invariance we use a modified DWT, called the Maximum Overlap DWT (MODWT) [8]. But with the MODWT the sum of the power of the smooth and details are not equal to the power in ΔX. However, the power is preserved in the wavelet coefficients.

We apply the MODWT using the Daubechies wavelet of order 4 on one-minute ΔX for all data in 1998 resulting in the wavelet coefficients $W_{j,t}$ (details) and V_t (smooth), where the level is $j \in [1, 7]$ and time is $t \in [0, 525599]$. Level j is associated with scale

$$\tau_j = 2^{j-1}. \tag{4}$$

As the time resolution is one minute the scale is also in minutes. The variance, or power, at level j is

$$\nu_j^2 = \sum_t W_{j,t}^2 \tag{5}$$

and the power conservation means that

$$\sum_t \Delta X_t^2 = \sum_j \nu_j^2 + \sum_t V_t^2. \tag{6}$$

The signal at level j is associated with frequencies in the range

$$f_j \in \left[\frac{1}{2^{j+1}}, \frac{1}{2^j}\right] = \left[\frac{1}{4\tau_j}, \frac{1}{2\tau_j}\right]. \tag{7}$$

Thus, if we compute the power spectrum $S(f)$ of ΔX with the Fourier transform then the wavelet variance

$$\nu_j^2 \approx 2\int_{1/2^{j+1}}^{1/2^j} S(f)df \tag{8}$$

where the factor 2 in front of the integral comes from the fact the $S(f)$ exist for $f \in [-1/2, 1/2]$ and is symmetrical around $f = 0$. In Figure 2 the estimated power spectrum from the MODWT and the DFT are shown (top panel). We see that there is a close agreement between the two estimates. The power in ΔX is concentrated to small scales (high frequencies), which becomes more clear in the two lower panels. The relative power (bottom left panel) is for the first four scales: 32%, 25%, 18%, and 13%. The cumulative relative power (bottom right panel) is 88% using only the wavelet coefficients up to level 4. This is in agreement with our previous conclusion that we need the one-minute data to capture the variance in ΔX. We also see that almost 90% of the signal is found at scales of $\tau_4 = 2^{4-1} = 8$ minutes corresponding to frequencies higher than $1/32$ min^{-1}.

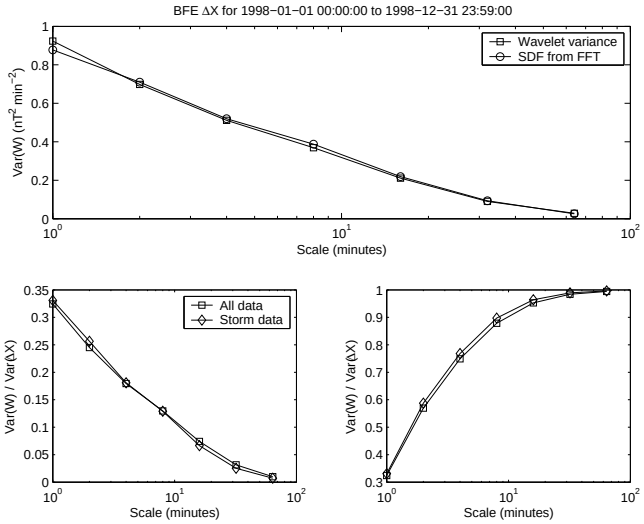

Figure 2: The figure shows the spectral density, or power spectrum, for ΔX normalised with the variance of ΔX. The labels have the following meaning: "All data", all data in 1998; "Storm data", use only one-day periods that contain events where $\Delta X > 20$ nT/min; "SDF from FFT", estimate using a Fourier transform.

We repeat the same analysis for ΔY at Brorfelde, and the result is shown in Figure 3. The spectrum is more flat up to level 4 after which the power decreases rapidly (top panel). The storm time spectrum has a shape similar to the general spectrum (bottom left) and the relative power for

the first 4 levels are: 26%, 20%, 19%, and 18%. The four levels together capture 83% of the power and at level 5 it goes over 90% (bottom right).

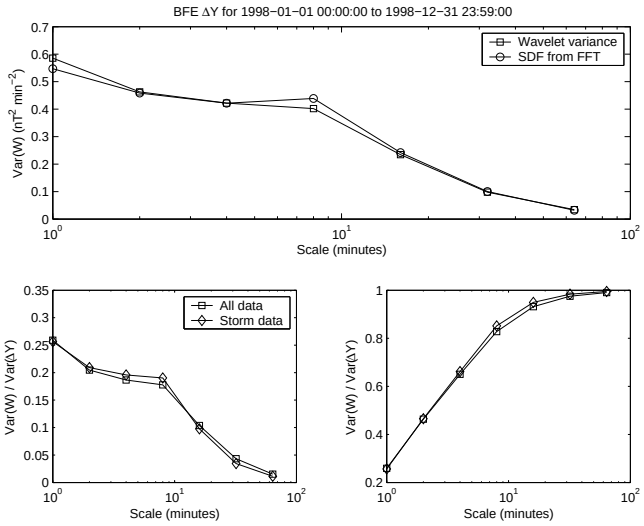

Figure 3: The figure shows the spectral density, or power spectrum, for ΔY normalised with the variance of ΔY. The labels have the following meaning: "All data", all data in 1998; "Storm data", use only one-day periods that contain events where $\Delta Y > 20$ nT/min; "SDF from FFT", estimate using a Fourier transform.

4.4 Time series of the wavelet variance

Equation 6 states that the power in ΔX equals the sum of the power of the wavelet coefficients taken over all times t. But our goal is to develop a model that predicts the variance of ΔX as a function of time and generally

$$\Delta X_t^2 \neq \sum_j W_{j,t}^2 + V_t^2. \tag{9}$$

However, the correlation C of ΔX^2 and $\overline{\Delta X}^2 = \sum_j W_{j,t}^2 + V_t^2$ is reasonable. For the one-minute values we have $C(\Delta X^2, \overline{\Delta X}^2) = 0.93$ and $C(\sqrt{\Delta X^2}, \sqrt{\overline{\Delta X^2}}) = 0.86$. Summing over 10 minute intervals, $\sum_{t=t_a}^{t_a+10} \Delta X_t$, the correlation increases to $C(\Delta X^2, \overline{\Delta X}^2) = 0.996$ and $C(\sqrt{\Delta X^2}, \sqrt{\overline{\Delta X}^2}) = 0.988$. For ΔY the correlation for one-minute data is $C(\Delta Y^2, \overline{\Delta Y}^2) = 0.89$ and $C(\sqrt{\Delta Y^2}, \sqrt{\overline{\Delta Y^2}}) = 0.85$. Using 10 minute data we have $C(\Delta Y^2, \overline{\Delta Y}^2) = 0.989$ and $C(\sqrt{\Delta Y^2}, \sqrt{\overline{\Delta Y^2}}) = 0.982$. So although Equation 9 is true the 10 minute variances of the wavelet coefficients are still good approximations to the variances in ΔX and ΔY.

From Figures 2 and 3 we get the general shape of the distribution of power at different scales. Assuming that the power distribution is constant over time we may estimate the power at different levels (or frequencies) using the 10 minute RMS of ΔX . In the next section we develop models that predicts the 10 minute RMS values.

However, the assumption of constant power distribution is not completely valid; there may be considerable differences between the variances at different scales at different times. For future work one should consider developing models that directly predict the 10 minute variance at different scales. Let

$$\nu_{j,t}^2 = \sum_{t'=t}^{t+9} W_{j,t'}^2 \tag{10}$$

be the 10-minute variance of the wavelet coefficients at level j, and let VAR(ΔX) be the 10 minute variance of ΔX. We then compute the correlation $C(\text{VAR}(\Delta X), \nu_j^2)$ and the result is shown in Figure 4. It is clear that VAR(ΔX) is a good approximation for the power at the two lowest scales, where also most of the power is. But with increasing scale the correlation drops.

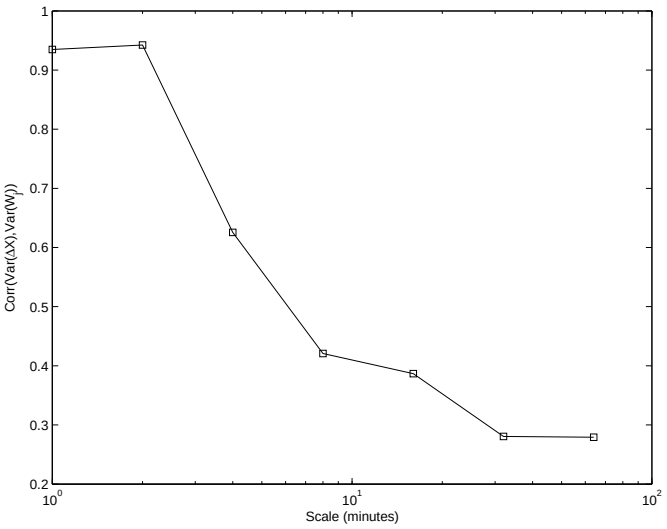

Figure 4: The figure shows the linear correlation between VAR(ΔX) and ν_j^2.

4.5 Geomagnetic field and GIC

The GIC data consist of measurements of currents flowing through a transformer neutral. The GIC may result from two different sources: space weather induced and non-space weather induced. To explore this we compute the linear correlation between the rate-of-change of the East-West component of the geomagnetic field dX/dt and the GIC. A time lag is introduced between dX/dt and GIC, and the both unfiltered and filtered GIC are used. For the filtering the DB1 wavelet is used. The wavelet approximation A and detail D are related to the original signal GIC as

$$\text{GIC} = A_1 + D_1 = A_2 + D_2 + D_1 = A_l + \sum_{n=1}^{l} D_n, \tag{11}$$

where l is the level. The filtered GIC at level l is computed as

$$\text{GIC}_l = \begin{cases} \text{GIC} & l = 0 \\ \sum_{n=1}^{l} D_n & l > 0 \end{cases} \tag{12}$$

where D_n is the detail at level l. The signal A_l contains only periods longer than 2^l minutes and thus GIC_l is the high frequency component with periods shorter than 2^l minutes. In Figure 5 we show the correlation between the Brorfelde (BFE) data and the filtered GIC data for the period 1998-09-17 to 1998-10-28. The maximum correlation is reached at a time lag of 2 minutes and filtering level 5. This means that there is a slowly varying component in GIC (period$> 2^5 = 32$ minutes) that is not related to dX/dt. The same analysis is repeated for Upsala (UPS) and the result is shown in Figure **??**. The maximum correlation is now found at a time lag of 3 minutes and filtering level 7, corresponding to $2^7 = 128$ minutes.

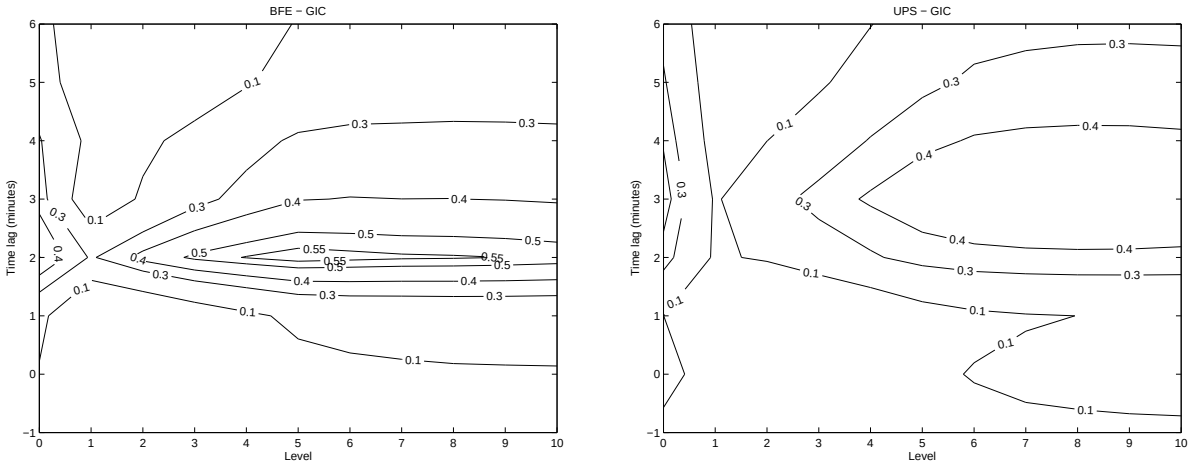

Figure 5: Correlation between dX/dt from Brorfelde (BFE), and Uppsala (UPS), and the GIC data for different levels and time lags. The level is the wavelet filtered GIC data at the corresponding level, where level 0 means unfiltered data.

We now set the filtering level to 6, which corresponds to 64 minutes, and look at the filtered GIC data. Two example periods are shown in Figures 6 and 7. In the top panels are shown a 24-hour interval around 1998-08-25 and 1998-10-18, respectively. In the bottom panels are shown a close-up covering 3 hours. The two examples have quite different characteristics. The first example (Figure 6) contains strong GIC reaching above 50 A. It is difficult to see any difference between the raw GIC and the filtered GIC in the top panel. In the bottom panel the differences becomes visible. The second example (Figure 7) is a much calmer period with maximum GIC of 12 A. However, there is a clear bias of 5-6 A in the raw GIC that is removed in the filtered GIC. For the whole period, 1998-09-17 to 1998-10-28, the raw GIC has a mean value of 1.22 A and a standard deviation of 2.2 A, while the filtered GIC has a mean of 0 and a standard deviation of 1.8 A.

The shift of 2 to 3 minutes between the geomagnetic data and the GIC data is at the moment unexplained. It could be due to different time stamping in the different sets. However, it is not crucial in this work as we will use 10 minute data.

The occasional bias term in the GIC is probably related to changes in the power grid configuration affecting the GIC measurements. Again, as we will use 10 minute data, and especially variances of the data, any bias term will be removed in the process. This will be discussed in the next section.

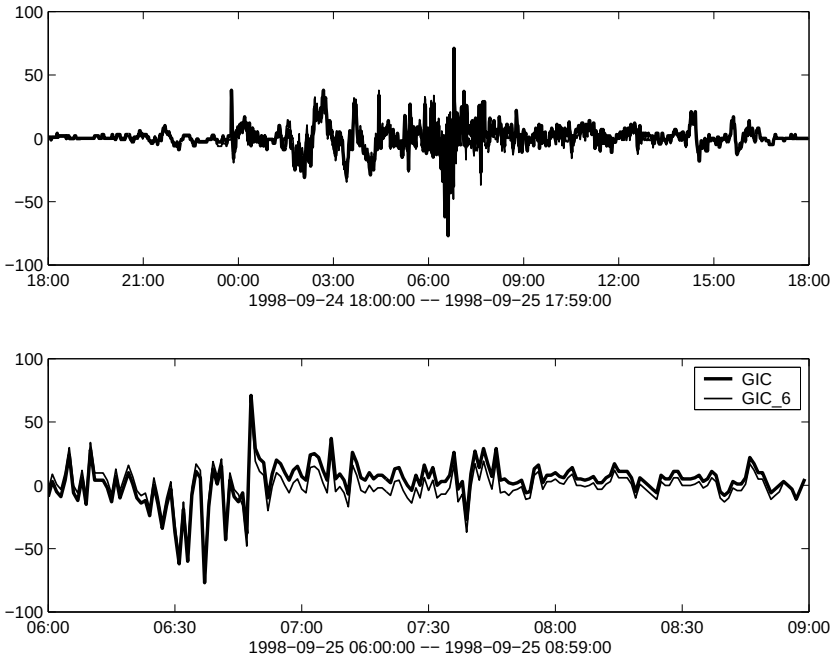

Figure 6: Correlation between dX/dt from Upsala (UPS) and the GIC data for different levels and time lags. The level is the wavelet filtered GIC data at the corresponding level, where level 0 means unfiltered data.

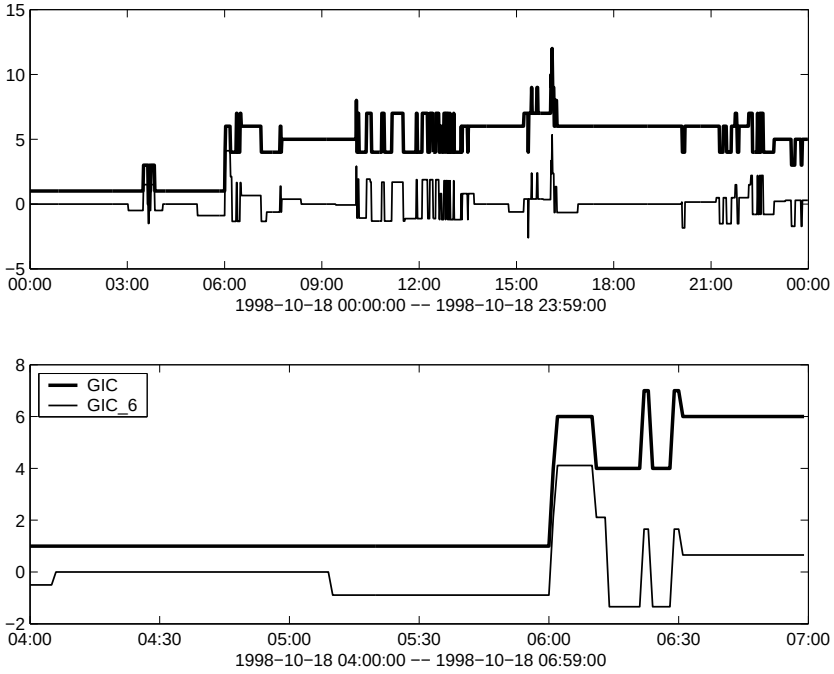

Figure 7: Correlation between dX/dt from Upsala (UPS) and the GIC data for different levels and time lags. The level is the wavelet filtered GIC data at the corresponding level, where level 0 means unfiltered data.

5 Neural network model

Here we describe the neural network model for the prediction of the 10-minute RMS ΔX and ΔY at Brorfelde and Uppsala.

5.1 10-minute resolution data

As described in Section 2 we will use data with 10 minute resolution in the forecast model. First we have the 10-minute average as

$$\mu(s) = \frac{1}{10} \sum_{t=10s}^{10s+9} x(t). \tag{13}$$

The average captures quite well the dynamics in the solar wind. Another quantity that is interesting is the standard deviation

$$\sigma(s) = \sqrt{\frac{1}{9} \sum_{t=10s}^{10s+9} (x(t) - \mu(s))^2} \tag{14}$$

as this is related to turbulence and strong gradients that are not seen in the average. Finally, we also have the root-mean-square (RMS) value

$$r(s) = \sqrt{\frac{1}{10} \sum_{t=10s}^{10s+9} x^2(t)} \tag{15}$$

which is related to the power in the signal. The solar wind data are resampled using the average and standard deviation where $x(t)$ is replaced by B_z, n, and V. The rate-of-change of the local geomagnetic field is resampled using the RMS where $x(t)$ is replaced with ΔX and ΔY at Uppsala and Brorfelde.

5.2 Architecture

The neural network takes solar wind data and time as input and predicts the log RMS ΔH, where ΔH is ΔX or ΔY at Brorfelde or Uppsala. The input is 10 minute averages and standard deviations. The inputs are collected into the input vector

$$\mathbf{X} = [d_1, d_2, l_1, l_2, \mu_{Bz}, \sigma_{Bz}, \mu_n, \sigma_n, \mu_V, \sigma_V] = [X_1, \ldots, X_{10}], \tag{16}$$

where

$$[d_1, d_2] = \left[\sin \frac{2\pi \mathrm{DD}}{365}, \cos \frac{2\pi \mathrm{DD}}{365}\right] \tag{17}$$

are the sine and cosine of the decimal day (DD),

$$[l_1, l_2] = \left[\sin \frac{2\pi \mathrm{LT}}{24}, \cos \frac{2\pi \mathrm{LT}}{24}\right] \tag{18}$$

are the sine and cosine of the local time (LT), and the $\mu_\bullet$ and $\sigma_\bullet$ are the mean and standard deviation of the solar wind data. The inputs are then normalised according to

$$x_i = \frac{X_i - a_i}{b_i} \tag{19}$$

where

$$\mathbf{a} = [0, 0, 0, 0, 0, 0.92, 8.0, 0.61, 490, 4.6] \tag{20}$$

$$\mathbf{b} = [2.1, 2.1, 2.1, 2.1, 18, 3.6, 24, 3.1, 320, 15]. \tag{21}$$

The normalisation constants have been chosen so that the mean of x_i is approximately 0, and the standard deviation is approximately 0.3. The neural network can now be written as

$$\hat{y}(t + \tau) = f(\mathbf{x}(t), N_H) \tag{22}$$

where τ is the prediction time, N_H the number of hidden neurons, and $\hat{y}$ is the network output. To capture the dynamics in the system we use internal feed-back units. The weights in network f are then adjusted so that the error between the desired output y and the network output $\hat{y}$ is minimised. The desired output is the normalised log square root of RMS dB according to

$$y = \frac{\log \mathrm{RMS}\Delta H - \alpha}{\beta}. \tag{23}$$

Four different networks are developed where ΔH is replaced with BFE ΔX, BFE ΔY, UPS ΔX, and UPS ΔY. The normalising constants are found in the table below.

Table 1: The normalising constants for the four models.

Model	α	β
BFE ΔX	0.064	1.2
BFE ΔY	-0.0079	1.2
UPS ΔX	0.020	1.4
UPS ΔY	-0.062	1.3

5.3 Training and optimisation

[Text to be added.]

5.4 Pruning

[Text to be added.]

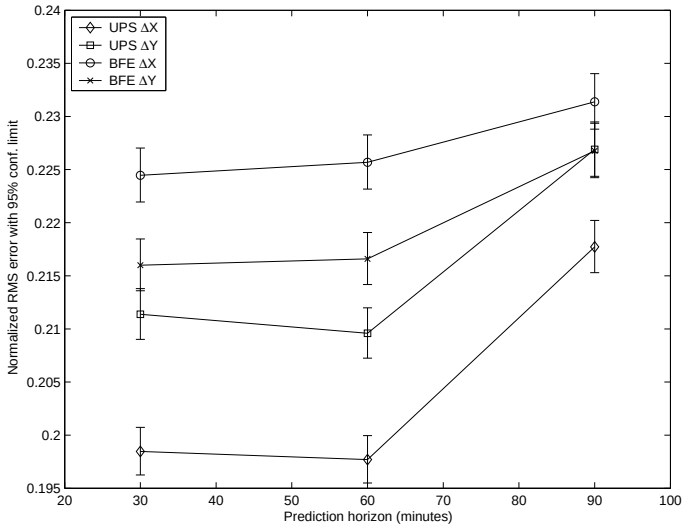

Figure 8:

5.5 Increasing the prediction horizon

[Text to be added.]

5.6 Models for ΔX and ΔY for interpolated data

[Text to be added.]

6 Forecasting of GIC

6.1 Empirical linear model from RMS ΔX and ΔY

[Text to be added.]

6.2 Theoretical model from RMS ΔX and ΔY

[Text to be added.]

7 Example predictions

7.1 ΔH

[Text to be added.]

7.2 GIC

[Text to be added.]

References

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

[2] 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.

[3] A. Pulkkinen, A. Viljanen, and R. Pirjola, "Real-time forecast service for geomagnetically induced currents technical note 2: Computation of gic from the geomagnetic field," tech. rep., Finnish Meteorological Institute, 2004.

[4] P. Wintoft, "Real-time forecast service for geomagnetically induced currents technical note 4: Direct forecasting of gic from solar wind data," tech. rep., Swedish Institute of Space Physics, 2004.

[5] P. Wintoft, A. Pulkkinen, M. Wik, A. Viljanen, R. Pirjola, H. Lundstedt, and L. Eliasson, "Real-time forecast service for geomagnetically induced currents technical note 1: Database," tech. rep., Swedish Institute of Space Physics and Finnish Meteorological Institute, 2004.

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

[7] R. S. Weigel, D. Vassiliadis, and A. J. Klimas, "Coupling of the solar wind to temporal fluctuations in ground magnetic fields," vol. 29, 2002.

[8] D. B. Percival and A. T. Walden, *Wavelet methods for time series analysis*. Cambridge University Press, 2002.