Statistical modeling of plasmaspheric hiss amplitude using solar wind measurements and geomagnetic indices

D. I. Golden,[1] M. Spasojevic,[1] W. Li,[2] and Y. Nishimura[2]

Received 31 January 2012; revised 22 February 2012; accepted 24 February 2012; published 23 March 2012.

[1] Plasmaspheric hiss plays a major role in the scattering and loss of electrons from the Earth's radiation belts, thereby contributing to the maintenance of the slot region between the inner and outer electron belt. As such, models of radiation belt dynamics require accurate estimates of hiss amplitudes. We present a straightforward empirical model, based on THEMIS data from June 2008 through December 2010, which uses multiple regression to predict observed equatorial hiss amplitudes on the three magnetospheric THEMIS probes as a function of *L* and MLT. The model utilizes solar wind measurements (e.g., dynamic pressure and IMF) and geomagnetic indices (e.g., *AE* and *SYM-H*) as model inputs. This model explains on average 26% of the variance of hiss amplitude observed on THEMIS. We present example output of the model for a minor geomagnetic storm that reveals, for the first time, the evolution of plasmaspheric hiss amplitude as a function of storm phase. **Citation:** Golden, D. I., M. Spasojevic, W. Li, and Y. Nishimura (2012), Statistical modeling of plasmaspheric hiss amplitude using solar wind measurements and geomagnetic indices, *Geophys. Res. Lett.*, *39*, L06103, doi:10.1029/2012GL051185.

1. Introduction

[2] Plasmaspheric hiss is a naturally occurring extremely low frequency (ELF) electromagnetic emission that is often observed within the plasmasphere in the near-Earth magnetosphere. Pitch-angle scattering of energetic electrons in the Earth's radiation belts due to cyclotron-resonant interactions with plasmaspheric hiss is a driver of large-scale electron loss and is partially responsible for the formation of the slot region that separates the inner and outer electron belts [e.g., *Lyons et al.*, 1972; *Lyons and Thorne*, 1973; *Meredith et al.*, 2007, 2009]. Global estimates of loss time scales for radiation belt electrons are typically calculated via quasi-linear diffusion modeling [e.g., *Khazanov et al.*, 2003; *Glauert and Horne*, 2005; *Albert*, 2005]. These models require accurate global estimates of plasmaspheric hiss amplitude to obtain plausible results.

[3] Previously, models of radiation belt dynamics have relied on simplistic assumptions of hiss amplitude, such as linear models of average hiss amplitude as a function of *AE** (the maximum *AE* in the preceding three hours) [e.g., *Meredith et al.*, 2007; *Shprits et al.*, 2008] based on statistics from CRRES measurements [*Meredith et al.*, 2004]. However, although average hiss amplitude increases with increasing *AE**, this correlation is weak. We show here that a linear model of hiss amplitude using a diverse assortment of solar wind measurements and geomagnetic indices is far more effective at capturing the measured hiss variance than one which uses only *AE** as an input. We present the first global statistical model that captures the variability of plasmaspheric hiss amplitude as a function of solar wind and geomagnetic drivers; this model may be used in its present form as an input to global diffusion models.

2. Instrumentation and Modeling Methodology

2.1. THEMIS Data

[4] The THEMIS mission is composed of five separate probes in near-equatorial orbits with apogees above 10 R_E and perigees below 2 R_E [*Angelopoulos*, 2008], generally within 20° magnetic latitude of the dipole equator. The Digital Fields Board (DFB) [*Cully et al.*, 2008] calculates the mean amplitude of the magnetic field in 6 logarithmically-spaced frequency bands from 0.1 Hz to 4 kHz using data from the Search Coil Magnetometer (SCM) [*Roux et al.*, 2008; *Le Contel et al.*, 2008]. The filter bank wave amplitude is available in survey mode, covering most orbits with a cadence of four seconds [*Cully et al.*, 2008]. This study uses survey mode SCM data from June 2008 through December 2010 for the three inner magnetospheric THEMIS probes, A, D and E. Only THEMIS data recorded within the plasmasphere, as determined via the approach described in section 2 of *Li et al.* [2010], is used in this study.

[5] To determine observed hiss amplitude, we use the highest three recorded filter bank (FBK) levels of the first SCM channel (SCM1), which have frequency ranges of 80–227 Hz, 316–904 Hz and 1390–4000 Hz (passbank levels 4, 2 and 0, respectively, from Table 6 of *Cully et al.* [2008]). The total hiss power is taken to be the sum of the power of these three FBK levels, where any channel power is set to zero in the sum if it is below the noise floor. If all three channels are at or below their noise floors, then the data are discarded. Also, we exclude any data collected within one minute of an instrument anomaly. Note that because we use data from only one component of the magnetic field, and this component is perpendicular to the ambient magnetic field, the measured hiss amplitude is somewhat below the true amplitude by, on average, a factor of $\sqrt{2}$. The gaps between the FBK levels introduce up to an additional factor of 1.05 reduction in measured versus true amplitude.

[6] Besides plasmaspheric hiss, the other emission of note which may appear with significant amplitude on the SCM data between 80–4000 Hz is equatorial magnetosonic waves.

[1]Electrical Engineering Department, Stanford University, Stanford, California, USA.
[2]Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, California, USA.

0094-8276/12/2012GL051185

Table 1. Model Inputs

Predictor	Number Hours History
SYM-H	0,1,3,4,5,6
$log_{10}AE$	0,2,3,5,6
$log_{10}P_{dyn}$	0,2,6
$d\Phi_{MP}/dt$	3,5,6
$P_{dyn}^{1/2}d\Phi_{MP}/dt$	1,2,5
$cos(\lambda)$	0

Magnetosonic waves are typically confined to within a few degrees of the equator [e.g., *Russell et al.*, 1970; *Gurnett*, 1976; *Němec et al.*, 2005] and may appear at frequencies up to the lower hybrid resonance, generally within the same frequency range as hiss inside the plasmasphere. Because the THEMIS survey mode data includes only one channel each of electric and magnetic fields sampled at a cadence of four seconds, it is not possible to distinguish between magnetosonic and whistler mode waves based on their respective wavenormals or polarization. Therefore, to mitigate excessive contamination from magnetosonic waves, we exclude THEMIS measurements within four degrees of the dipole equator.

2.2. The Multiple Regression Model

[7] We model the observed hiss amplitude using a linear multiple regression model of the form $Y = Xb$ where Y (the "response") is an $n \times 1$ column vector of n samples of THEMIS hiss amplitude on a log scale, X (the model inputs) is an $n \times (m + 1)$ matrix of n samples of m different solar wind measurements and geomagnetic indices plus a constant, and b (the "weights") is an $(m + 1) \times 1$ column vector of weights, one for each input feature plus one for the constant. The optimal weights, b, are determined by a traditional least squares solution based on simultaneous measurements of the inputs and response.

[8] We include two geomagnetic indices as model inputs, *SYM-H*, and the logarithm of *AE*. *AE* has been shown in the past to be correlated with hiss amplitude and occurrence [e.g., *Meredith et al.*, 2004; *Golden et al.*, 2011]. We also include three solar wind parameters as model inputs using solar wind data from *Qin et al.* [2007], including the logarithm of solar wind dynamic pressure, P_{dyn}; the rate at which magnetic flux is opened at the magnetopause, $d\Phi_{MP}/dt = v_{sw}^{4/3}B_T^{2/3}\sin^{8/3}(\theta_c/2)$ [*Newell et al.*, 2007], where v_{sw} is the solar wind velocity, B_T is the tangential IMF magnitude, and θ_c is the MF clock angle; and a version of the same that is corrected for magnetopause currents, $P_{dyn}^{1/2}d\Phi_{MP}/dt$. Time integration of $d\Phi_{MP}/dt$ and $P_{dyn}^{1/2}d\Phi_{MP}/dt$ have been shown by *Newell et al.* [2007] to be predictive of a variety of different geomagnetic indices and phenomena. Solar wind data from the *Qin et al.* [2007] database are used if they have quality 2 (full availability) or 1 (interpolated); epochs where any solar wind parameter has quality 0 (not available) are excluded. We also include as a model input the cosine of the dipole magnetic latitude, λ, of the hiss measurement.

[9] We consider the possibility of including the current value of each solar wind measurement and geomagnetic index (with respect to the time of the THEMIS hiss measurement) as well as the average value between 0 and 1 hrs ago, 1 and 2 hrs ago, etc., up to the average between 5 and 6 hrs ago. Instead of including all time histories for each measurement, we use sequential feature selection with 10-fold cross validation to choose only the time histories that are relevant for modeling hiss amplitude. This feature selection step was run only once, across the entire data set at all L and MLT.

[10] The resulting time histories which were chosen as inputs to the model are shown in Table 1 (0 hrs history represents the instantaneous measurement of the feature). The inter-hour variations of *SYM-H* and *AE* appear to be important in most of the six hours preceding the hiss measurements, while coarser time histories of P_{dyn}, $d\Phi_{MP}/dt$ and $P_{dyn}^{1/2}d\Phi_{MP}/dt$ are needed, possibly due to the slower response of hiss generation to these exo-magnetospheric measurements.

[11] We bin THEMIS hiss measurements by dipole L and MLT into bins with $\Delta L = 1$ and ΔMLT = 2 hrs. The average observed hiss amplitude on THEMIS as a function of L and MLT is shown in Figure 1a. The shown values are linearly interpolated between bin centers (which is why they appear "smoothed"). Grid points with fewer than 30 samples are white. In this and all subsequent figures, the values in a given bin are based on data obtained exclusively when THEMIS is inside the plasmasphere. The number of 4-sec samples per bin is shown in Figure 1b. Although Figure 1b shows that the number of observations of hiss is nearly constant across MLT, Figure 1a shows that hiss appears with

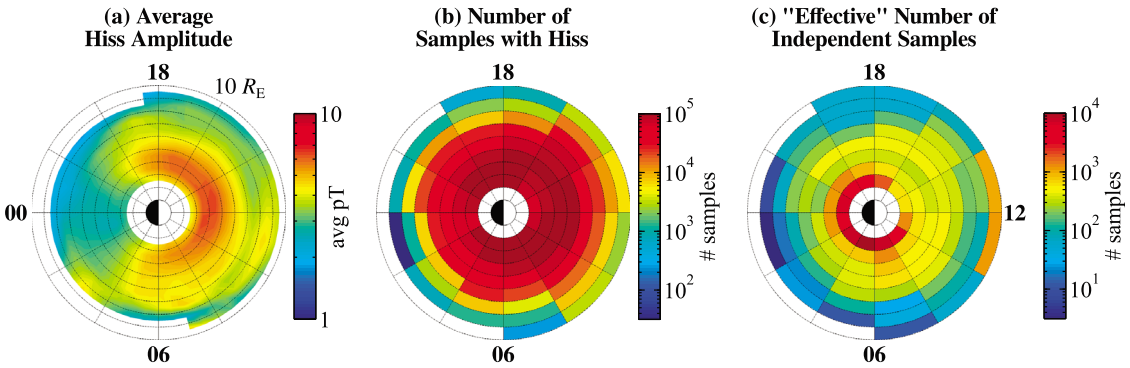

Figure 1. Hiss statistics used in this study, from 1 June 2008 to 25 November 2010. (a) Average hiss amplitude (the average is taken over the logarithm of hiss amplitude). (b) Number of 4-sec samples where hiss was detected above the instrument noise floor (bin size is $\Delta L = 1$, ΔMLT = 2 hrs). (c) "Effective" number of *independent* samples.

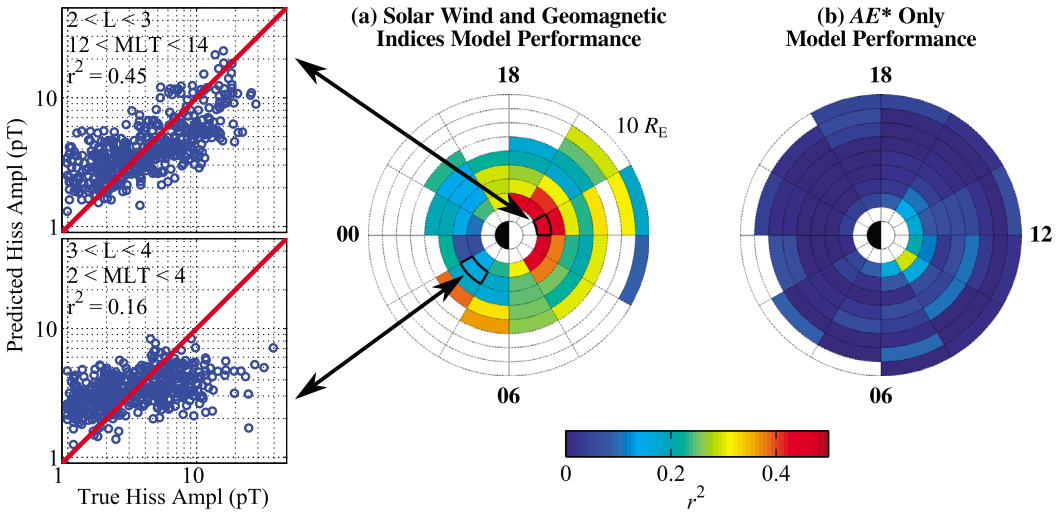

Figure 2. (a) Model performance based on coefficient of determination, r^2 (the square of the correlation coefficient), between the true hiss amplitude and the amplitude predicted by this model. Bins with too few independent samples to be modeled are colored white. Scatter plots of the predicted hiss amplitude versus the true amplitude, i.e., the constituent data used to calculate the coefficient of determination, are shown for two select bins. The mean r^2 is 0.26. (b) Performance for a model which only uses the current value of AE^* as an input. This model performs poorly, with a mean r^2 of 0.04.

minimum amplitude in the midnight sector and maximum amplitude in the noon sector at $L = 1$. We discuss the L and MLT structure of the hiss amplitude distribution with respect to its dynamic response to a geomagnetic storm in section 4.

[12] To avoid overfitting in our model of hiss amplitude, we require at least 10 times as many independent samples as model inputs in each bin ($n > 10\ m$). However, because our samples of hiss amplitude are drawn from a time series, each sample is not actually independent of the previous sample. Therefore, we calculate the "effective" number of independent samples in each bin [*Bayley and Hammersley*, 1946], which, for a time series with high inter-sample autocorrelation, may be significantly less than the total number of samples. The effective number of samples of a first order autoregressive time series is a function of the time series length and autocorrelation coefficient, and is given by equation 2.7 of *Mudelsee* [2010]. The effective number of independent samples of hiss amplitude is shown in Figure 1c. As shown in Table 1, there are a total of $m = 21$ inputs in the model; therefore, to model a given bin, we require that bin to have $n \geq 210$ effective independent samples.

[13] For each L/MLT bin with sufficient independent samples of hiss, we use least squares to calculate coefficients for the linear model containing the inputs from Table 1. These coefficients are supplied in the auxiliary material.[1]

3. Goodness of Fit

[14] After specifying a linear regression model for each L/MLT bin, we evaluate the performance in each bin by comparing our modeled hiss amplitude, $\hat{Y} = X\hat{b}$ (where $\hat{b}$ is the estimated weights that we calculate via least squares), with the true hiss amplitude, Y. We use as our goodness of fit metric the coefficient of determination, which is the square of the correlation coefficient, between Y and $\hat{Y}$. This metric may be interpreted as "the proportion of the variance of measured hiss amplitude explained by the model" [e.g., *Chatterjee and Hadi*, 2008, section 7.2].

[15] Figure 2a shows the per-bin coefficient of determination for our linear regression model for all modeled bins with $n \geq 210$. Bins which were not modeled because they contained too few independent samples are colored white. Because the constituent hiss amplitude data of each bin has a different distribution and was collected during different geomagnetic conditions, the performance varies significantly across different bins. Scatter plots of the predicted hiss amplitude versus the true hiss amplitude ($\hat{Y}$ versus Y) of some representative bins are shown. Higher performing bins, such as $2 < L < 3$, $12 < \text{MLT} < 14$ (top left), show a good correlation between Y and $\hat{Y}$, while lower performing bins, such as $3 < L < 4$, $2 < \text{MLT} < 4$ (bottom left), show a poor correlation between Y and $\hat{Y}$. Model performance is generally best at low L on the day side and worst at low L on the night side. The mean of r^2 across all modeled bins is 0.26. We could say that, on average, this model explains 26% of the variance of observed hiss amplitude.

[16] These results should be contrasted with a linear regression model which uses only the current value of AE^*, as used in recent global quasi-linear diffusion modeling [e.g., *Meredith et al.*, 2007; *Shprits et al.*, 2008]. In this case, the per-bin modeling procedure is the same, where the hiss amplitude in each bin is modeled as $\log(\text{hiss amplitude}(t)) = b(AE^*(t)) + C$, where b is a scalar weight which is estimated from the data and C is a constant. The per-bin performance of this model is shown in Figure 2b. The mean of r^2 is 0.04. On average, a model which includes only AE^* explains a mere 4% of the variance of observed hiss amplitude.

4. Example Model Output

[17] An example of our hiss model output is shown in Figure 3. This example is meant not to validate the model

[1]Auxiliary materials are available in the HTML. doi:10.1029/2012GL051185.

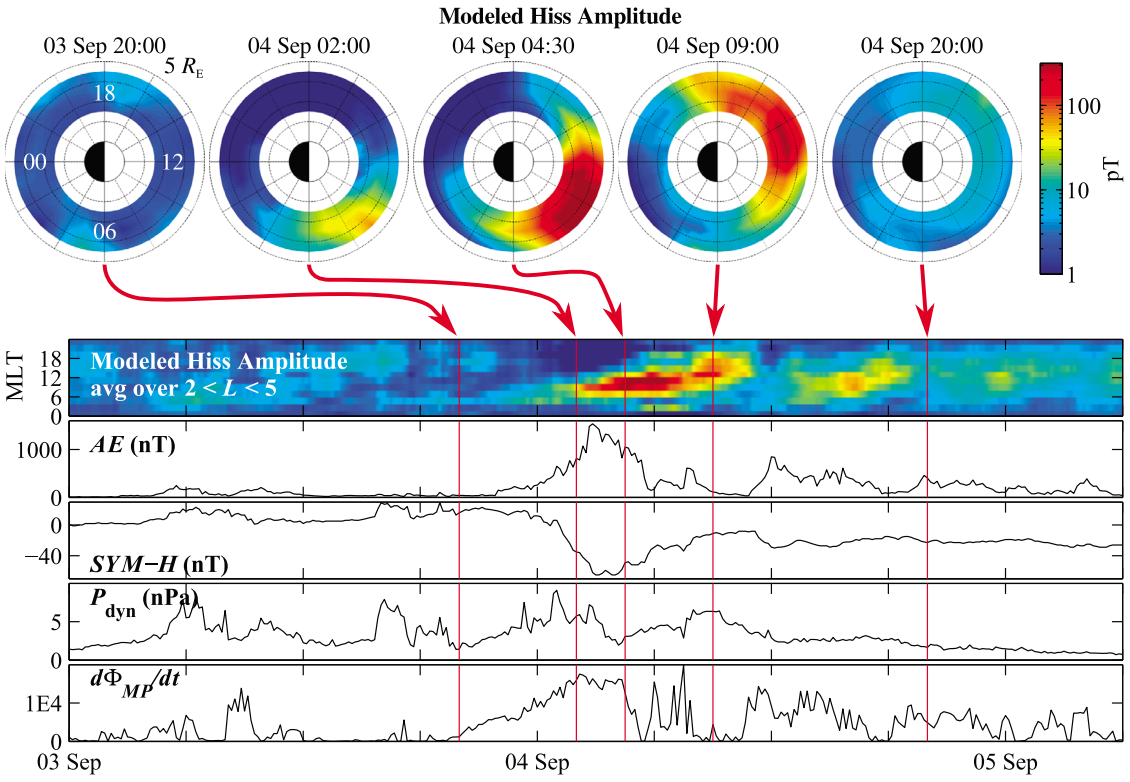

Figure 3. (top) Output of the hiss model for the 04 September 2008 storm. (bottom) Hiss amplitude as a function of MLT and time averaged over $2 < L < 5$, and *AE*, *SYM-H*, P_{dyn} and $d\Phi_{MP}/dt$ as a function of time.

(being an empirical model, the model is already "validated" via the measure of its goodness of fit in section 3), but to illustrate new insight that may be gained using the model. The interval from 3–5 September 2008 includes a moderate geomagnetic storm with a minimum *SYM-H* of about −65 nT. Prior to storm onset (max *SYM-H*) at ∼21:00 UTC on 3 September, the modeled hiss amplitude is low across all *L* and MLT. At storm onset, hiss amplitude initially increases significantly, particularly in the pre-noon sector between 04 and 12 MLT where amplitudes exceed 100 pT. As the storm progresses through the minimum *SYM-H* at 03:30 UTC on 4 September, the region of high hiss amplitude remains present and rotates Eastward to the post-noon sector. Some regions of high hiss amplitude continue until around 02:00 UTC on 5 September, well into the storm recovery phase.

[18] Although the mechanism of plasmaspheric hiss generation remains an area of active research, it is likely that cyclotron resonant amplification plays an important role in the process. The minimum resonant energy for first order cyclotron resonance with plasmaspheric hiss in the frequency range of 100–1000 Hz is ∼30–300 keV at $L = 3.5$. The initial amplification of waves in the pre-noon quadrant near 00:00 UTC on 4 September is likely the result of increased magnetospheric convection, as indicated by the increase in $d\Phi_{MP}/dt$, leading to enhanced transport of these medium energy electrons from the plasmasheet to the dawn side inner magnetosphere. Further, the intensification of the waves that occurs after 04:30 UT on 4 September is likely the result of the explosive injection of substorm electrons, as indicated by the peak in *AE*.

[19] In the recovery phase of the storm, the peak in the hiss local time distribution moves to the noon and post-noon sectors. However, we note that the apparent motion of the most intense hiss region is considerably slower than the azimuthal drift time of hiss resonant electrons. For example for plasmaspheric hiss at 1 kHz near $L = 3.5$, the minimum resonant energy is 35 keV corresponding to a drift time of about 8.5 hrs, with higher energies (lower resonant frequencies, closer to the peak of the hiss spectrum) drifting even faster. However, the Eastward movement of the hiss over the course the storm appears to correspond to a much longer drift time in the range of approximately 18–22 hrs. Further work is needed to understand the evolution of the local time distribution of plasmaspheric hiss as a function of storm phase.

[20] An animation showing the progression of hiss amplitude versus *L* and MLT during this interval is included in the auxiliary material.

5. Conclusion

[21] The presented model for plasmaspheric hiss amplitude, which includes as inputs time histories of various geomagnetic indices and solar wind parameters, is able to estimate hiss amplitudes more accurately than a model which uses only *AE** as an input; its output is therefore more suitable as an input to radiation belt models. The greater precision of estimated hiss amplitudes that may be determined using this model has vast potential to improve the accuracy of radiation belt models and to further our understanding of the role of plasmaspheric hiss in the maintenance of the structure of the electron radiation belts.

[22] **Acknowledgments.** The work at Stanford University was supported by the NSF under awards 1043442 and 1137411. *AE* and *SYM-H* data are from the World Data Center for Geomagnetism, Kyoto. Solar wind data are from the Virtual Radiation Belt Observatory, based on data from OMNIWeb, courtesy of R. Denton and Z. Qin, who acknowledge support from NSF award 0751007. THEMIS data are from NASA CDAWeb, and we acknowledge NASA contract NAS5-02099 and V. Angelopoulos for use of THEMIS data, J. W. Bonnell and F. S. Mozer for EFI data, C. W. Carlson and J. P. McFadden for ESA data, A. Roux and O. LeContel for SCM data and C. Cully for FBK data.

[23] The Editor thanks Jichun Zhang and Andrew Collier for their assistance in evaluating this paper.

References

Albert, J. M. (2005), Evaluation of quasi-linear diffusion coefficients for whistler mode waves in a plasma with arbitrary density ratio, *J. Geophys. Res.*, *110*, A03218, doi:10.1029/2004JA010844.

Angelopoulos, V. (2008), The THEMIS Mission, *Space Sci. Rev.*, *141*, 5–34, doi:10.1007/s11214-008-9336-1.

Bayley, G. V., and J. M. Hammersley (1946), The "effective" number of independent observations in an autocorrelated time series, *J. R. Stat. Soc.*, *8*(2), 184–197.

Chatterjee, S., and A. S. Hadi (2008), *Regression Analysis by Example*, 2nd ed., Wiley-Interscience, Hoboken, N. J.

Cully, C. M., R. E. Ergun, K. Stevens, A. Nammari, and J. Westfall (2008), The THEMIS Digital Fields Board, *Space Sci. Rev.*, *141*, 343–355, doi:10.1007/s11214-008-9417-1.

Glauert, S. A., and R. B. Horne (2005), Calculation of pitch angle and energy diffusion coefficients with the PADIE code, *J. Geophys. Res.*, *110*, A04206, doi:10.1029/2004JA010851.

Golden, D. I., M. Spasojevic, and U. S. Inan (2011), Determination of solar cycle variations of midlatitude ELF/VLF chorus and hiss via automated signal detection, *J. Geophys. Res.*, *116*, A03225, doi:10.1029/2010JA016193.

Gurnett, D. A. (1976), Plasma wave interactions with energetic ions near the magnetic equator, *J. Geophys. Res.*, *81*, 2765–2770, doi:10.1029/JA081i016p02765.

Khazanov, G. V., K. V. Gamayunov, and V. K. Jordanova (2003), Self-consistent model of magnetospheric ring current and electromagnetic ion cyclotron waves: The 2–7 May 1998 storm, *J. Geophys. Res.*, *108*(A12), 1419, doi:10.1029/2003JA009856.

Le Contel, O., et al. (2008), First results of the THEMIS search coil magnetometers, *Space Sci. Rev.*, *141*, 509–534, doi:10.1007/s11214-008-9371-y.

Li, W., R. M. Thorne, J. Bortnik, Y. Nishimura, V. Angelopoulos, L. Chen, J. P. McFadden, and J. W. Bonnell (2010), Global distributions of suprathermal electrons observed on THEMIS and potential mechanisms for access into the plasmasphere, *J. Geophys. Res.*, *115*, A00J10, doi:10.1029/2010JA015687.

Lyons, L. R., and R. M. Thorne (1973), Equilibrium structure of radiation belt electrons, *J. Geophys. Res.*, *78*, 2142–2149, doi:10.1029/JA078i013p02142.

Lyons, L. R., R. M. Thorne, and C. F. Kennel (1972), Pitch-angle diffusion of radiation belt electrons within the plasmasphere, *J. Geophys. Res.*, *77*, 3455–3474.

Meredith, N. P., R. B. Horne, R. M. Thorne, D. Summers, and R. R. Anderson (2004), Substorm dependence of plasmaspheric hiss, *J. Geophys. Res.*, *109*, A06209, doi:10.1029/2004JA010387.

Meredith, N. P., R. B. Horne, S. A. Glauert, and R. R. Anderson (2007), Slot region electron loss timescales due to plasmaspheric hiss and lightning-generated whistlers, *J. Geophys. Res.*, *112*, A08214, doi:10.1029/2007JA012413.

Meredith, N. P., R. B. Horne, S. A. Glauert, D. N. Baker, S. G. Kanekal, and J. M. Albert (2009), Relativistic electron loss timescales in the slot region, *J. Geophys. Res.*, *114*, A03222, doi:10.1029/2008JA013889.

Mudelsee, M. (2010), *Climate Time Series Analysis: Classical Statistical and Bootstrap Methods, Atmos. Oceanogr. Sci. Libr.*, vol. 42, Springer, Dordrecht, Netherlands, doi:10.1007/978-90-481-9482-7.

Němec, F., O. Santolík, K. Gereová, E. Macúšová, Y. de Conchy, and N. Cornilleau-Wehrlin (2005), Initial results of a survey of equatorial noise emissions observed by the Cluster spacecraft, *Planet. Space Sci.*, *53*, 291–298, doi:10.1016/j.pss.2004.09.055.

Newell, P. T., T. Sotirelis, K. Liou, C.-I. Meng, and F. J. Rich (2007), A nearly universal solar wind-magnetosphere coupling function inferred from 10 magnetospheric state variables, *J. Geophys. Res.*, *112*, A01206, doi:10.1029/2006JA012015.

Qin, Z., R. E. Denton, N. A. Tsyganenko, and S. Wolf (2007), Solar wind parameters for magnetospheric magnetic field modeling, *Space Weather*, *5*, S11003, doi:10.1029/2006SW000296.

Roux, A., O. Le Contel, C. Coillot, A. Bouabdellah, B. de La Porte, D. Alison, S. Ruocco, and M. C. Vassal (2008), The Search Coil Magnetometer for THEMIS, *Space Sci. Rev.*, *141*, 265–275, doi:10.1007/s11214-008-9455-8.

Russell, C. T., R. E. Holzer, and E. J. Smith (1970), OGO 3 observations of ELF noise in the magnetosphere: 2. The nature of the equatorial noise, *J. Geophys. Res.*, *75*, 755–768, doi:10.1029/JA075i004p00755.

Shprits, Y. Y., D. A. Subbotin, N. P. Meredith, and S. R. Elkington (2008), Review of modeling of losses and sources of relativistic electrons in the outer radiation belt II: Local acceleration and loss, *J. Atmos. Terr. Phys.*, *70*, 1694–1713, doi:10.1016/j.jastp.2008.06.014.

D. I. Golden and M. Spasojevic, Electrical Engineering Department, Stanford University, 350 Serra Mall, Stanford, CA 94305, USA. (dgolden1@stanford.edu)

W. Li and Y. Nishimura, Department of Atmospheric and Oceanic Sciences, University of California, 7127 Math Sciences Bldg., 405 Hilgard Ave., Los Angeles, CA 90095, USA.