He Fei (Orcid ID: 0000-0003-0542-2686)
Zhang Xiaoxin (Orcid ID: 0000-0002-7759-7402)
Fok Mei-Ching, H. (Orcid ID: 0000-0001-9500-866X)
Liemohn Michael, W. (Orcid ID: 0000-0002-7039-2631)
Gallagher Dennis (Orcid ID: 0000-0003-3924-3571)
Nakano Shin'ya (Orcid ID: 0000-0003-0772-4610)

A new solar wind driven global dynamic plasmapause model: 2. Model and Validation

Fei He, Key Laboratory of Earth and Planetary Physics, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China (Email: hefei@mail.iggcas.ac.cn)

Xiao-Xin Zhang, Key Laboratory of Space Weather, National Center for Space Weather, China Meteorological Administration, Beijing, China (**Corresponding author, Email: xxzhang@cma.gov.cn**)

Rui-Lin Lin, National Space Science Center, Chinese Academy of Sciences, Beijing, China (Email: linrl@nssc.ac.cn)

Mei-Ching Fok, NASA Goddard Space Flight Center, Code 673, Greenbelt, MD 20771, U.S.A. (Email: mei-ching.h.fok@nasa.gov)

Roxanne M. Katus, Department of Mathematics, Eastern Michigan University, Ypsilanti, Michigan, USA (Email: rkatus@emich.edu)

Mike W. Liemohn, Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, Michigan, USA (Email: liemohn@umich.edu)

Dennis L. Gallagher, NASA Marshall Space Flight Center, Huntsville, Alabama, U.S.A. (Email: dennis.gallagher@nasa.gov)

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/2017JA023913

Shinya Nakano, The Institute of Statistical Mathematics, Tachikawa, Tokyo, Japan (Email: shiny@ism.ac.jp)

Abstract.

A new solar wind driven global dynamic plasmapause (NSW-GDP) model has been constructed based on the largest currently-available database containing 49119 plasmapause crossing locations and 3957 plasmapause profiles (corresponding to 48899 plasmapause locations), from 18 satellites during 1977 – 2015 covering four solar cycles. This model is compiled by the Levenberg-Marquardt method for nonlinear multiparameter fitting and parameterized by V_{SW}, B_Z, *SYM-H*, and *AE*. Continuous and smooth MLT-dependence controlled mainly by the solar wind driven convection electric field E_{SW} is also embedded in this model. Compared with previous empirical models based on our database, this new model improves the forecasting accuracy and capability for the global plasmapause. The diurnal, seasonal and solar cycle variations of the plasmapause can be captured by the new model. The NSW-GDP model can potentially be used to forecast the global plasmapause shape with upstream solar wind and IMF parameters and corresponding predicted values of *SYM-H* and *AE*, and can also be used as input parameters for other inner magnetospheric coupling models, such as dynamic radiation belt and ring current models and even MHD models.

Key Points:

1. A new solar wind driven global dynamic plasmapause model based on multi-satellite observations is constructed
2. This model is parameterized by V_{SW}, IMF B_Z, SYM-H, and AE, and has continuous and smooth MLT-dependence

3. This model is potentially applicable to inner magnetospheric research studies and space weather forecasts

1. Introduction

The plasmapause is the outer boundary of the plasmasphere, where the plasma density drops dramatically by at least half an order of magnitude in a short distance of ~0.5 R_E (Earth's radii, 1 R_E=6378.0 km) [*Carpenter*, 1963; *Gringauz*, 1963]. The plasmapause configuration is one of the key parameters in the coupling interaction of the plasmasphere, the ring current, and the radiation belts in the magnetosphere and is also an important indicator for geomagnetic activity in near-Earth space [*Carpenter and Anderson*, 1992; *Fok et al.*, 1995; *Khazanov and Liemohn*, 1995; *Goldstein et al.*, 2003a; *Spasojević et al.*, 2003].

The plasmapause locations can be characterized by physics-based models or statistics-based empirical models. For the physics-based models, some are based on a fluid approach, such as the dynamic global core plasma model (DGCPM) [*Ober et al.*, 1997; *Liemohn et al.*, 2004; *He et al.*, 2013] in which the plasmapause is identified from the radial steep gradient in the density profiles, and the others are based on a kinetic approach, such as the dynamic kinetic model of the plasmasphere by *Pierrard and Stegen* [2008] in which the plasmapause is derived from zero parallel force (ZPF) surface [*Lemaire*, 1989]. For a detailed review of the physics-based models of the plasmasphere, please refer to *Pierrard et al.* [2009]. After the review of *Pierrard et al.* [2009], a physics-based reconstruction of the density in the plasmasphere was presented by *Verigin et al.* [2012] and *Kotova et al.* [2015], where the plasmapause was described as the last closed stream line but with rather flexible shape.

Many statistics-based empirical models of the plasmapause have been developed in past studies.

The functional representation of the geocentric radius of the plasmapause (L_{PP}, in R_E) was first proposed by *Carpenter and Anderson* [1992] (hereafter, this model is referred to as the CAA-1992 model). The CAA-1992 model is a *Kp*-based empirical model with $L_{PP} = 5.6 - 0.46\ Kp_{max}$, where Kp_{max} is the maximum *Kp* value in the preceding 24 hours, and where the model was developed for the 0 h – 15 h magnetic local time (MLT) sector based on plasmapause crossing events taken from International Sun-Earth Explorer (ISEE-1) data in 1977, 1982 and 1983. Following the CAA-1992 model, *Gallagher et al.* [2000] constructed the empirical global core plasma model (GCPM) to characterize the core plasma density and composition in the inner magnetosphere. This is the first model to provide smooth, continuous total density globally.

A new *Kp*-dependent model of $L_{PP} = (5.39 \pm 0.072) - (0.382 \pm 0.019)\ Kp_{max}$, where Kp_{max} is the maximum *Kp* value in the preceding 12 hours, was proposed by *Moldwin et al*. [2002] (hereafter referred to as the MOL-2002 model) based on Combined Release and Radiation Effects Satellite (CRRES) measurements in 1990-1991. Using the same database of *Moldwin et al.* [2002], *O'Brien and Moldwin* [2003] further developed an MLT-dependent L_{PP} model with the following formula (hereafter referred to as the OBM-2003 model):

$$L_{PP} = a_1 [1 + a_{mlt} \cos(\boldsymbol{\Phi} - a_{\boldsymbol{\Phi}})]\ Q + b_1 [1 + b_{mlt} \cos(\boldsymbol{\Phi} - a_{\boldsymbol{\Phi}})] \qquad (1)$$

where $Q = \max_{-36,-2} Kp$, $Q = \log_{10} \max_{-36,0} AE$, or $Q = \log_{10} |\min_{-24,0} Dst|$, $\boldsymbol{\Phi} = 2\pi(\mathrm{MLT}/24)$, a_1, a_{mlt}, $a_{\boldsymbol{\Phi}}$, b_1, b_{mlt}, and $b_{\boldsymbol{\Phi}}$ are fitted parameters. The notation $\max_{t1,t2} X$ or $\min_{t1,t2} X$ indicates the maximum or minimum of X taken from t_1 to t_2 hours before the plasmapause crossing. Recently, *Bandić et al.* [2016] revisited the same CRRES data and constructed a new model (BAN-2016) similar to

OBM-2003 but with different coefficients.

The plasmapause locations extracted from the Imager for Magnetosphere-to-Auroral Global Exploration (IMAGE) Extreme Ultraviolet Imager (EUV) images were used by *Larsen et al.* [2007] to construct the first solar wind driven plasmapause model (hereafter referred to as the LAR-2007 model) without considering MLT-dependence. The LAR-2007 model was a function of interplanetary magnetic field (IMF) B_Z component with a shift time of 155 min, and $\boldsymbol{\phi} = V_{SW}\, B \sin^2(\theta_c/2)$ with shift time of 275 min for solar wind speed V_{SW}, IMF magnitude B, and IMF clock angle θ_c:

$$L_{PP} = 0.050\, B_{Z,155} - 1.110 \times 10^{-4}\, \boldsymbol{\phi}_{275} + 4.23 \qquad (2)$$

A new fit function of L_{PP} to V_{SW}, B_Z and *AE* was recently proposed by *Cho et al.* [2015] (hereafter referred to as the CHO-2015 model) based on the plasmapause crossings from the Time History of Events and Macroscale Interactions during Substorms (THEMIS) during the ascending phase of Solar Cycle 24, achieving better prediction performance compared with previous models. *Liu et al.* [2015] used the THEMIS data to further construct an MLT-dependent L_{PP} model with input parameters of *SYM-H*, *AL*, *AU*, *AE* and *Kp*. Plasmapause crossings extracted from the Waves of High frequency and Sounder for Probing of Electron density by Relaxation (WHISPER) on Cluster were used by *Verbanac et al*. [2015] to construct the new L_{PP} formula as a function of B_Z, $V_{SW}B_Z$, $d\Phi_{mp}/dt$, *Dst*, *Ap*, and *AE* in three MLT sectors (1 h – 7 h, 7 h – 16 h, and 16 h – 1 h) (hereafter referred to as the VER-2015 model).

For the convenience of comparisons, all the above empirical models are summarized in Table 1. It is

shown that each of these models just used plasmaspheric observations from a single satellite, and the databases for these models were not large enough to fully and adequately cover the temporal-spatial-geomagnetic activity dimensions. Some of the models with MLT-dependence have just used geomagnetic indices as input parameters (e.g., LIU-2015), while other models using solar wind parameters have no MLT-dependence (e.g., LAR-2007 and CHO-2015). Since the plasmapause locations are highly MLT-dependent and also highly dependent on solar wind and geomagnetic activity, construction of a solar wind driven global dynamic plasmapause model with MLT-dependence is necessary.

In Paper 1 [*Zhang et al.*, this issue], the largest currently-available database containing 49119 plasmapause crossing locations and 3957 plasmapause profiles (corresponding to 48899 plasmapause locations in 1 h MLT intervals) was established for four solar cycles (21 – 24). Statistical characteristics of this database were discussed in detail in Paper 1, especially the diurnal, seasonal and solar cycle variations of the plasmapause. This database, with full MLT and solar activity coverage, provides a unique opportunity to construct the new solar wind driven global dynamic plasmapause (NSW-GDP) model in this investigation.

The paper will be arranged as follows. The correlations between the plasmapause locations and the solar wind and geomagnetic parameters will be investigated in section 2 to help us select the parameters important for controlling the size and the shape of the plasmapause. In section 3, we will get the relationships between the selected parameters and the plasmapause locations by the Levenberg-Marquardt method for nonlinear multiparameter fitting. In section 4, the NSW-GDP

model will be compared with previous models and relevant discussions will be given. Finally, a summary and conclusion will be presented in section 5.

2. Parameters Selection

As a core region of the inner magnetosphere, the dynamics of the plasmasphere is mainly controlled by two sources of drivers. The first is external convective driving from the solar wind and IMF [*Goldstein et al.*, 2003a, 2003b, 2005a, 2005b; *Sandel et al.*, 2003; *Darrouzet et al.*, 2009; *Katus et al.*, 2015; and references therein], which modifies large scale convection in the inner magnetosphere that drives the dynamic distribution of plasmaspheric plasma through $\mathbf{E} \times \mathbf{B}$ drifts. The second is internal driving due to the dynamics of magnetospheric energetic particles and the ionosphere [*Goldstein et al.*, 2003c, 2007; *Liemohn et al.*, 2004, 2006; *He et al.*, 2016; *Zhang et al.*, 2017], especially by auroral substorms that produce strong ion and electron precipitation in the ionosphere [*Akasofu*, 1964; *McPherron et al.*, 1973] and dipolarization in the magnetosphere [*Runov et al.*, 2009; *Ge et al.*, 2012]. In the following sections, we will investigate the correlations of the plasmapause location with geomagnetic indices and solar wind parameters to optimize the parameters that drive the NSW-GDP model.

2.1. Correlation with Geomagnetic Indices

The 3-hour averaged *Kp* index, the 1-hour averaged *Dst* index, the 5-minute averaged *SYM-H* index, and the 5-minute averaged *AE* index are used in this section to investigate correlations of the plasmapause location to geomagnetic activity. Like previous studies [e.g., *O'Brien and Moldwin*, 2003; *Cho et al.*, 2015; *Liu et al.*, 2015], a time window of two days is set to find the best delay

time of the geomagnetic indices to plasmapause variations. We correlate the plasmapause locations with $\max_{t_1,t_2} Kp$, $\log_{10}(|\min_{t_1,t_2} Dst|)$, $\log_{10}(|\min_{t_1,t_2} SYM\text{-}H|)$, and $\log_{10}(\max_{t_1,t_2} AE)$, where t_1 and t_2 both varies from -48 h to 0 h with 0 h corresponding to the plasmapause crossing time, always keeping t_1 earlier than t_2. The Spearman rank-order correlation coefficient (ROCC) [*Press et al.*, 1992] of the three quantities to the plasmapause locations at different delay times are calculated and the times t_1 and t_2 corresponding to the strongest correlation are shown in Figure 1. It is indicated that the correlation is always negative for *Kp*, *Dst*, *SYM-H*, and *AE*. The correlation is always strong in the 0 h – 12 h and 21 h – 24 h MLT sectors and weak in the 12 h – 21 h MLT sector. The MLT-averaged values of t_1 are -26 h, -18 h, -18 h, and -18 h for *Kp*, *Dst*, *SYM-H*, and *AE*, respectively, and the corresponding t_2 values are -3 h, -1 h, -1 h, and -2 h. The *Kp*, *Dst* and *SYM-H* indices are all good indicators of geomagnetic storms but with temporal resolutions of 3 hours, 1 hour, and 1 minute, respectively. In consideration of constructing a dynamic plasmapause model, it's better to use high resolution indices, though the ROCC's for these three indices are all high in Figures 1*a*-1*c*. Therefore, only the 5-minute averaged *SYM-H* and the 5-minute averaged *AE* are used in our model given that *SYM-H* is a good proxy of the geomagnetic storm and *AE* is a good indicator of the geomagnetic substorm. Inclusion of *SYM-H* and *AE* into the model can improve the adaptability of the model since sometimes a geomagnetic disturbance can cause dynamic variations of the plasmapause even when the solar wind conditions are quiet and stable [*Goldstein et al*., 2007; *Zhang et al*., 2017].

2.2. Correlation with Solar Wind

For the solar wind and IMF parameters, since the OMNI data have been time shifted to the nose of the Earth's bow shock and the delay time from the bow shock to the inner magnetosphere has been considered in matching solar wind parameters to the plasmapause crossings, the time window method in the section above will not be used here. The solar wind parameters in our database will be directly correlated with the plasmapause locations. The ROCC's between L_{PP} and IMF B_X, B_Y, B_Z, $B_T = \sqrt{B_Y^2 + B_Z^2}$ (all in the GSM coordinate system in nT), IMF clock angle θ_c =atan (B_Y, B_Z), V_{SW} (km/s), N_{SW} (cm^{-3}), solar wind driven convection electric field $E_{SW}=10^{-3}B_Z V_{SW}$ (mV/m), solar wind dynamic pressure $P_{dyn}= 2\times10^{-6} N_{SW}V_{SW}^2$ (nPa), and Akasofu's solar wind-magnetosphere coupling function $\varepsilon=V_{SW}B_T\sin^4(\theta_c/2)$ [*Perrault and Akasofu*, 1978] at different MLT sectors (in 1-h intervals) are calculated and plotted in Figure 2.

Characteristics in Figure 2 can be summarized as follows:

1. V_{SW} has the strongest correlation with L_{PP} compared with other parameters, with the ROCC being ~ -0.6 at all MLT sectors indicating a very strong negative correlation.
2. N_{SW} is positively correlated with L_{PP} though not so strong, with ROCCs around 0.2.
3. IMF B_Z and θ_c have the same correlations with L_{PP}. Although their ROCCs are symmetric about the line of ROCC=0, their physical means are the same since an increase of θ_c from 0° to 180° corresponds to a decrease of B_Z from positive to negative. Such correlations indicate that a southward turning of the IMF can cause the earthward shrinkage of L_{PP}.
4. Due to the strong correlation of V_{SW} with L_{PP}, the ε parameter, P_{dyn}, and E_{SW}, which are all related with V_{SW} according to their definitions, are all weakly correlated with L_{PP} with the

correlations becoming even weaker in the pre-dusk region around 16 h MLT. This is because correlation of B_Z with L_{PP} is the weakest and the positive correlation of N_{SW} to L_{PP} is the strongest in that region.

5. IMF B_X and B_Y are almost uncorrelated with L_{PP} at all MLT sectors.

Based on the above analysis, the two basic parameters, V_{SW} and IMF B_Z, which are the most important two parameters for solar wind-magnetosphere coupling and inner magnetospheric dynamics, are selected to be the input parameters of the NSW-GDP model. Therefore, in the following section, we will construct the NSW-GDP model driven by V_{SW}, B_Z, *AE*, and *SYM-H*.

3. Construction of the NSW-GDP Model

The basic framework of the model will be built on V_{SW} and B_Z in the first step. In the second step, the MLT-dependence of the plasmapause will be built on E_{SW}, and then *AE* and *SYM-H* will be included in the model in the final step. Note that only 80% of the database is randomly chosen for modeling, the other 20% is used for comparison, and all calculations throughout the rest of this paper are performed in the magnetic equatorial plane of solar magnetic (SM) coordinates.

3.1. Basic Framework

To get the basic framework of the NSW-GDP model, it is better to eliminate the effects of MLT-dependence of the plasmapause. According to the statistics in Paper 1, the plasmapause shape in the 0 h to 6 h MLT sector is almost circular under many different solar wind and geomagnetic activity levels. Thus, we resample all the L_{PP} values from the 0 h to 6 h MLT sector to different V_{SW} bins with intervals of 50 km/s and different B_Z bins with intervals of 1 nT, respectively, to get the

dots in Figure 3 with standard deviations represented by the vertical error bars. It should be noted that plasmapause crossings with V_{SW} greater than 800 km/s and the magnitude of IMF B_Z greater than 15.0 nT are excluded in Figure 3 since the corresponding sample numbers are too small to ensure a reasonable statistical significance level for the *t*-test [*Press et al.*, 1992]. Best fitting of the dots in Figure 3*a* shows that the plasmapause decreases exponentially with V_{SW} as represented by the function in equation (3). As the magnitude of the IMF B_Z increases, the plasmapause shrinks towards the Earth, with the reduction more dramatic for southward IMF than for northward IMF. Such variations of L_{PP} with B_Z can be well depicted by an inverse polynomial function as shown by the black curve in Figure 3b. Therefore, the two fitted functions in Figure 3 are as follows with the fitted coefficients for $L_{PP}(V_{SW})$ and $L_{PP}(B_Z)$ listed in Table 2.

$$L_{PP}(V_{SW}) = f_0 + f_1 \exp(-V_{SW} / f_2) \qquad (3)$$

$$L_{PP}(B_Z) = g_0 + \frac{g_1}{1 + g_2[(B_Z - g_3)/g_4]^2 + g_5[(B_Z - g_3)/g_4]^4 + g_6[(B_Z - g_3)/g_4]^6} \qquad (4)$$

The two functions in Figure 3 are independent. It is expected that, for different IMF conditions, the coefficients for $L_{PP}(V_{SW})$ should change, and the same for $L_{PP}(B_Z)$ under different solar wind conditions. Therefore, establishment of a two-dimensional L_{PP} function of both V_{SW} and B_Z should be the most useful. To achieve this, we first present the B_Z variations of L_{PP} for different V_{SW} in Figure 4. Curves in Figure 4 are all best-fittings of equation (4) to averaged values in each V_{SW} interval. Note that the typical standard deviation for the symbols in Figure 4 is ~0.5 R_E. The *t*-test indicates that the maximum values at IMF B_Z = +3.0 nT in Figure 4 are significantly different with

each other with significance level of 99%, except for the comparison between the star and the diamond, for which a significance level of 95.5% is shown. For the symbols at IMF B_Z = -15.0 nT or +15.0 nT, the significance levels are all above 95%. This reveals that the symbols and curves in Figure 4 are truly different from each other.

It is demonstrated that the functional form of equation (4) can well depict the L_{PP}-B_Z relationship under different V_{SW} conditions. So, we have rewritten equation (4) in the following form:

$$L_{PP}(V_{SW}, B_Z) = g_0(V_{SW}) + \frac{g_1(V_{SW})}{1 + g_2(V_{SW})u^2 + g_5(V_{SW})u^4 + g_6(V_{SW})u^6}, \quad (5)$$
$$u = [B_Z - g_3(V_{SW})] / g_4(V_{SW})$$

where the seven V_{SW} dependent g functions can control the configurations of the fitted curves and the fitting method is as follows:

1. Divide the L_{PP}-B_Z data into different subsets according to the V_{SW} bins with an interval of 50 km/s;
2. Resample the L_{PP} data into B_Z bins with an interval of 1.0 nT;
3. Fit the data points to equation (4) to get g_0, g_1, g_2, g_3, g_4, g_5, and g_6 for different V_{SW}, as shown in Figure 5.
4. Fit g_0, g_1, g_2, g_3, g_4, g_5, and g_6 to V_{SW} to get the forms of the seven functions, respectively.
5. Adjust the coefficients of the seven functions iteratively to get an optimized function of equation (5).

After best-fitting with the Levenberg-Marquardt method [*Press et al.*, 1992], the forms of the seven g functions are:

$$\begin{cases} g_0(V_{SW}) = g_{00} + g_{01}\exp(-V_{SW}/g_{02}), \\ g_1(V_{SW}) = g_{10} + g_{11}\exp(-V_{SW}/g_{12}), \\ g_2(V_{SW}) = g_{20} + g_{21}\exp(-V_{SW}/g_{22}), \\ g_3(V_{SW}) = g_{30} + g_{31}V_{SW}, \\ g_4(V_{SW}) = g_{40} + g_{41}\exp(-V_{SW}/g_{42}), \\ g_5(V_{SW}) = g_{50} + g_{51}\exp(-V_{SW}/g_{52}), \\ g_6(V_{SW}) = g_{60}\exp(-V_{SW}/g_{61}) \end{cases} \quad (6)$$

where the coefficients for these functions will be finally determined in section 3.4 with the bootstrap resampling method.

3.2. MLT-Dependence

In the previous section, we have established the functional representation of the L_{PP} by V_{SW} and IMF B_Z in the 0 h to 6 h MLT sector. The MLT variations of the plasmapause will be investigated to establish the MLT-dependence of the model. Since the MLT-dependence of the plasmapause is mainly controlled by the large-scale convection in the inner magnetosphere, the solar wind driven convection electric field E_{SW}, therefore, may be suitable to investigate the MLT-dependence of the plasmapause. Note that this statistical analysis cannot represent the often highly structured plasmapause profile that appears to result from meso-scale processes in the inner magnetosphere. While these structures can have profound radial and MLT variations, only their statistical properties are included.

Figure 6 shows the MLT-dependence of the L_{PP} under different conditions of E_{SW}. Firstly, the plasmapause locations are classified into six groups according to the values of E_{SW} as shown in the upper left corner of Figure 6. Secondly, the plasmapause locations in each group are binned into

each MLT sector with 1 h intervals. Finally, the averaged L_{PP} are obtained and plotted as colored dots in Figure 6. Each set of dots in Figure 6 is best fitted with the following function:

$$f(\mathrm{MLT}) = \frac{h_0 + h_1 \sin^2 \chi}{1 + h_3 \sin^2 \chi}, \qquad \chi = \frac{\pi(\mathrm{MLT} - h_2)}{24} \tag{7}$$

where h_0, h_1, h_2 and h_3 are functions of E_{SW}. After best fitting with the Levenberg-Marquardt method in each E_{SW} group, the forms of the four h functions are:

$$\begin{cases} h_0(E_{SW}) = h_{00} + h_{01}E_{SW} + h_{02}E_{SW}^2, \\ h_1(E_{SW}) = h_{10} + h_{11}E_{SW} + h_{12}E_{SW}^2, \\ h_2(E_{SW}) = h_{20} + h_{21}E_{SW}, \\ h_3(E_{SW}) = h_{30} + h_{31}E_{SW} + h_{32}E_{SW}^2, \end{cases} \tag{8}$$

where the coefficients for these functions will be finally determined in section 3.4.

3.3. Including Storm and Substorm Effects

As expressed in section 2, apart from the solar wind and IMF, the dynamics in the inner magnetosphere can also affect the shape and size of the plasmapause. The most significant dynamics in the inner magnetosphere are geomagnetic storms and substorms that can modify the convection patterns in the plasmaspheric region, thus change the shape and size of the plasmapause. As common proxies for geomagnetic activity, *SYM-H* and *AE* are used to denote the storm and substorm activities, respectively. Therefore, we further attempt to add functions of *SYM-H* and *AE* to the NSW-GDP model.

According to the statistics in section 2.1, the slopes of the linear fitting functions of L_{PP} to $\alpha \log_{10}$

|*SYM-H*| and alog_{10} *AE* at different MLT sectors are shown in Figure 7*a* and 7*b*, respectively. It is shown that both *SYM-H* and *AE* have different effects on the plasmapause at different MLT sectors. For *SYM-H*, the slope decreases from 0 h to 10 h MLT and then increases until 16 h MLT when the slope experiences a decrease again. For *AE*, there is only one peak at 16 h MLT, around which the slopes are both decreasing.

According to the variations of the slopes in Figure 7, piecewise linear functions are adopted to characterize the relationship between L_{PP} and *SYM-H* and that between L_{PP} and *AE*. Based on Figure 1, the MLT-averaged time window is t_1 = -18 h and t_2 = -1 h for both *SYM-H* and *AE*. The function is:

$$\Delta L_{PP}(SYM\text{-}H, AE, \text{MLT}) = k_0 \text{alog}_{10} |\min_{-18,-1} SYM\text{-}H| + k_1 \text{alog}_{10} \max_{-18,-1} AE + k_2 \quad (9)$$

where the parameters k_0, k_1 and k_2 are piecewise linear functions of MLT in the following forms:

$$k_0 = \begin{cases} k_{00} + k_{01}\text{MLT}, & 0 \le \text{MLT} < 10, \\ k_{02} + k_{03}\text{MLT}, & 10 \le \text{MLT} < 16, \\ k_{04} + k_{05}\text{MLT}, & \text{MLT} \ge 16 \end{cases} \quad (10)$$

$$k_1 = \begin{cases} k_{10} + k_{11}\text{MLT}, & 0 \le \text{MLT} < 16, \\ k_{12} + k_{13}\text{MLT}, & \text{MLT} \ge 16 \end{cases} \quad (11)$$

$$k_2 = \begin{cases} k_{20} + k_{21}\text{MLT}, & 0 \le \text{MLT} < 15, \\ k_{22} + k_{23}\text{MLT}, & \text{MLT} \ge 15 \end{cases} \quad (12)$$

The coefficients in equations (10) – (12) are to be determined in the final fitting of the model functions to the plasmapause location database.

Additionally, the contributions of storm or substorm activity to plasmaspheric evolution are different under different conditions of solar wind driven large scale convection. The plasmapause

locations are classified into four subsets based on the absolute values of E_{SW}, as listed in Table 3. The values of $dL_{PP\text{-}Substorm}$ and $dL_{PP\text{-}Storm}$ in Table 3 are obtained in the following manner. In each subset of E_{SW}, $dL_{PP\text{-}Substorm}$ equals the averaged plasmapause location during non-substorm periods with *AE* < 200 nT subtracted by the averaged plasmapause location during substorm periods with *AE* > 500 nT. For $dL_{PP\text{-}Storm}$, the subsets for non-storm and storm periods are *SYM-H* > -20 nT and *SYM-H* < -50 nT, respectively. It is demonstrated that as $|E_{SW}|$ increases, the contributions of both substorms and storms to the shrinkage of the plasmapause weaken almost linearly. Therefore, a switch function of $k_3(|E_{SW}|)$ has been added to the first two terms of equation (9), which is now rewritten as:

$$\Delta L_{PP}(SYM\text{-}H, AE, \mathrm{MLT}) = \left\{k_0 \mathrm{alog}_{10} \left| \min_{-18,-1} SYM\text{-}H \right| + k_1 \mathrm{alog}_{10} \max_{-18,-1} AE\right\} \times \left\{k_{30} + k_{31} \tanh[(|E_{SW}| - k_{32})/k_{33}]\right\} + k_2 \quad (13)$$

where the coefficients for the switch function will be determined in the following optimization procedure.

3.4. Determination of the Model Coefficients

Finally, the complete form of the NSW-GDP model has been constructed as follows:

$$L_{PP}(V_{SW}, B_Z, SYM\text{-}H, AE, \mathrm{MLT}) = \left\{g_0(V_{SW}) + \frac{g_1(V_{SW})}{1 + g_2(V_{SW})u^2 + g_5(V_{SW})u^4 + g_6(V_{SW})u^6}\right\} \times \frac{h_0(E_{SW}) + h_1(E_{SW})\sin^2\chi}{1 + h_3(E_{SW})\sin^2\chi}$$
$$+\{k_0(\mathrm{MLT})\mathrm{alog}_{10} \left| \min_{-18,-1} SYM\text{-}H \right| + k_1(\mathrm{MLT})\mathrm{alog}_{10} \max_{-18,-1} AE\} \times k_3(|E_{SW}|) + k_2(\mathrm{MLT}), \quad (14)$$
$$u = [B_Z - g_3(V_{SW})]/g_4(V_{SW}),$$
$$\chi = \frac{\pi[\mathrm{MLT} - h_2(V_{SW})]}{24},$$

where detailed definitions of the functions have been presented in previous sections. A bootstrap

resampling method [*Efron and Tibshirani*, 1993; *Huber and Ronchetti*, 2009], which has been used successfully before in space physics research to quantify the error on statistical results [e.g., *Reiff*, 1990; *Liemohn et al.*, 2012; *Katus et al.*, 2013], is conducted to determine the coefficients and test the robustness of the model. The bootstrap method is to randomly choose 80% of the database for the creation of the model, and then compare against the remaining 20%. For this investigation 1000 bootstrap resamplings are conducted to investigate the mean and spread of the coefficients in equation (14).

By fitting equation (14) to each resampling of the database in each MLT sector to maximize the ROCC and minimize the root-mean-square errors (RMSE) of the fitted plasmapause locations to the observed plasmapause locations, the coefficients for equations (14) are determined. Repeating the fitting procedure for 1000 times, the mean value and standard deviation (σ) of each coefficient in equation (14) are calculated and listed in Table 4. The standard deviations are generally within 5% of the mean values, indicating that the model fittings are robust. It is noted that, in equation (14), the upper limit values for V_{SW}, $|B_Z|$, and $|E_{SW}|$ are 800 km/s, 15.0 nT, and 10 mV/m, respectively, as analyzed in section 3.1.

4. Discussions on the NSW-GDP Model

4.1. General Performance

Comparisons between the observed plasmapause locations (L_{OBS}) and the model results (L_{MOD}) at different MLT sectors are shown in Figure 8 and the ROCC and RMSE are plotted in Figure 9. The error bars in Figure 9 indicate the standard deviations of the ROCC and RMSE calculated through

the bootstrap resampling method. Figure 8 is obtained in the following manner. In a specific MLT sector with N plasmapause locations, the L_{OBS} and corresponding L_{MOD} are resampled into two-dimensional grids with size of 0.2 R_E × 0.2 R_E, and the number of points in each grid is normalized by the total number of plasmapause locations N to get the normalized distributions in Figure 8. Generally, ~ 70% of the points are located in the range of L_{MOD} ± 0.5 R_E, except for the 16 – 19 MLT sectors, the percentage is about 50% ~ 60%. The ROCC of the NSW-GDP model is between 0.7 and 0.8 in 21 h to 12 h MLT sectors. In the afternoon to dusk region, a minimum value of ROCC = 0.56 is reached at 17 h MLT. The variation of RMSE with MLT is inversely changing with that of ROCC in Figure 9, indicating that a higher ROCC corresponds to a lower RMSE and thus a more successful modeling. The error bars in Figure 9 show that the model compilation method is robust regardless of the 80% of the database chosen to create the coefficient set.

4.2. Comparison with Previous Models

In this section, the NSW-GDP model will be compared with those models listed in Table 1. Based on the remaining 20% of the database, the RMSE of the various models listed in Table 1 are calculated and plotted in Figure 10. The most significant feature in Figure 10 is that the NSW-GDP model has the best performance in all MLT sectors compared with all other models. The maximum RMSE of NSW-GDP is 0.91 R_E at 17 h MLT sector, with a minimum of 0.57 R_E at midnight. The maximum RMSE of other models are all larger than 1.0 R_E with some models achieving 1.8 R_E. Generally, the performances of all the models become poor in the afternoon to dusk MLT sectors. This might be caused by the fact that the plasmasphere is the most dynamic in the afternoon to dusk

MLT sectors where different structures of plumes may be generated [*Moldwin et al*., 2004; *Goldstein and Sandel*, 2005; *Darrouzet et al*., 2006, 2008; *He et al*., 2011, 2013]. Besides, all the old models are based on plasmaspheric observations from single satellites in limited periods, and such models may be difficult to fit such a large database with a time span of nearly 40 years.

4.3. Modeled Plasmapause Variations

To analyze the variations of the plasmapause locations simulated by the NSW-GDP, e.g., the diurnal, seasonal, and solar cycle variations presented in Paper 1, the global plasmapause locations are calculated from 1980 to 2015 (a 36-year long interval) based on the hourly OMNI solar wind and IMF parameters and the geomagnetic indices. Then, the plasmapause locations are binned into an MLT-UT coordinate frame with 1 h intervals as shown in Figure 11*e*, an MLT-month coordinate frame with 1 month intervals as shown in Figure 11*j*, and an MLT-year coordinate frame with 1 year intervals as shown in Figure 11*o*, respectively. Figure 11 also shows are the variations of the four input parameters of the NSW-GDP model and the averaged curves are obtained using data from 1980 to 2015.

Figure 11*e* shows that the plasmapause shape experiences two weak valleys around 0 h UT and 12 h UT at all MLT sectors, basically consistent with those in Paper 1. Figures 11*a*-11*d* indicate that V_{SW}, IMF B_Z, and *SYM-H* have almost no diurnal variation; *AE* has two weak peaks between ~12 h and ~18 h UT. The differences between Figure 11*e* and Figure 11*a* in Paper 1 might be caused by the fact that the plasmapause locations are considered to be equal in the 0 h – 6 h MLT sector in the construction of the basic framework of the model. Also, the diurnal variations of the plasmapause

may be faded through the average of the 36-year data in 1 h UT intervals.

The seasonal variations of the plasmapause locations characterized by equinox valleys and solstice peaks are clearly shown in Figure 11*j*. Such variations may possibly be caused by the seasonal variations of *SYM-H* shown in Figure 11*h* since the seasonal variations of V_{SW}, IMF B_Z, and *AE* are not so obvious as shown in Figures 11*f*, 11*g* and 11*i*.

The solar cycle variations characterized by solar maximum valleys and solar minimum peaks are well captured by the NSW-GDP model, as shown in Figure 11*o*. A new feature in Figure 11*o* is that the solar cycle variations of the plasmapause locations seems to have a time delay relatively to the sunspot number as shown by the dashed curve Figure 11*o*. This might be caused by the fact that the solar wind conditions and geomagnetic activities are more disturbed during the descending phase of the solar cycle as indicated in Figure 11*k*-11*n* and also in the statistics of *Papitashvili et al.* [2000].

4.4. Potential Applications of the NSW-GDP Model

The newly compiled NSW-GDP model is a solar wind driven, dynamic and MLT-continuous plasmapause model. It could be included in dynamic radiation belt and ring current models, such as the Comprehensive Ring Current Model (CRCM) [*Fok et al.*, 2001], the Radiation Belt Environment (RBE) model [*Fok et al.*, 2011], and the Comprehensive Inner Magnetosphere-Ionosphere (CIMI) Model [*Fok et al.*, 2014], instead of the CA1992 (or any of these other) plasmapause models. The NSW-GDP model can also be embedded into the MHD models [e.g., *Raeder et al.*, 2001; *Hu et al*., 2007; *Pulkkinen et al.*, 2008; *Lu et al.*, 2013] to better characterize the inner boundary of the magnetosphere which is usually set to be at 3 R_E in these

MHD models for simplicity.

The NSW-GDP model can also be used to forecast the status of the plasmasphere. If upstream observations of solar wind and IMF parameters and prediction models of *SYM-H* [*Cai et al.*, 2010] and *AE* [*Takalo and Timonen*, 1997; *Luo et al.*, 2013] are used, the NSW-GDP model can potentially be used to forecast the plasmapause location and shape on the magnetic equatorial plane. Once the plasmapause location is determined, the plasmaspheric plasma density and composition can be predicted through extrapolation of ionospheric density to magnetosphere, such as the method used in the GCPM model [*Gallagher et al.*, 2000].

5. Conclusion

In this paper, the NSW-GDP model is developed based on the largest database to date, containing 49119 plasmapause crossing locations and 3957 plasmapause profiles (corresponding to 48899 plasmapause locations in 1 h MLT intervals), from 18 satellites during 1977 – 2015. Through investigation of the correlations of the plasmapause locations with solar wind parameters and with geomagnetic indices, the 5-min V_{SW}, 5-min B_Z, 5-min *SYM-H*, and 5-min *AE* are chosen as the free, driving parameters of the NSW-GDP model. A time shift in the solar wind parameters and geomagnetic indices relative to derived plasmapause locations is considered in the model development. The driving parameters for the plasmapause model are selected according to the correlation analysis, and the relationships between these parameters and the plasmapause location and shape are constructed by the Levenberg-Marquardt method for nonlinear multiparameter fitting. It is noted that continuous MLT-dependence is embedded in this new model. The limitations of this

model are also discussed. The main results are obtained as follows:

1. The functions of NSW-GDP model are given in equation (14) with the coefficients listed in Table 4.
2. The new model is developed and parameterized by V_{SW}, B_Z, *SYM-H*, and *AE*. The plasmapause locations are calculated on the magnetic equatorial plane in SM coordinate system.
3. Continuous and smooth MLT-dependence is embedded in this model and controlled mainly by the solar wind driven convection electric field E_{SW}.
4. The NSW-GDP model can well capture the seasonal and solar cycle variations of the plasmapause.
5. This new model improves the forecast accuracy and capability for the global plasmapause, and the best performance of agreements between model results and observations is achieved compared with selected previous models.

In this paper, we have preliminarily developed an initial plasmapause model. Assuming that the plasmasphere is aligned with magnetic field lines and using empirical geomagnetic field models (IGRF and Tsyganenko models), an empirical three-dimensional plasmapause shape can be obtained through field line tracing. The NSW-GDP can potentially be included to current magnetospheric dynamic models and be applied to forecast the dynamic evolution of the plasmasphere. More work should be done in the future to improve the global dynamic plasmapause model, especially for extreme solar wind conditions. Besides, if the time history of the plasmapause and the corotation effect can be embedded into the NSW-GDP model, the performance of the model

can be further improved in the future.

Acknowledgements

The authors sincerely thank the National Astronomical Observatories, Chinese Academy of Sciences, for provision of the CE-3 EUVC data, NASA/CDAWEB for provision of the ISEE-1 PWI data, DE-A PWI data, Polar PWI data, and IMAGE RPI data, ESA/Cluster Science Archive for provision of the Cluster WHISPER data, Iowa University for the provision of the VAP EMFISIS data available from http://emfisis.physics.uiowa.edu/data/index, the THEMIS team for provision of the THEMIS ESA and EFI data available from http://themis.ssl.berkeley.edu/index.shtml, JAXA/DARTS for provision of the Akebono PWS data, and Dr. T. Forrester of IMAGE EUV team for provision of the IMAGE EUV data and relevant processing software. The authors give thanks to NASA-CCMC for providing the code of IGRF and Tsyganenko model, NASA/GSFC OMNIWeb for providing the solar wind and IMF data, the Kyoto World Data Center for providing the geomagnetic indices (*Dst*, *SYM-H*, and *AE*) and the German Research Centre for Geosciences (GFZ) for providing the *Kp* and *Ap* indices. This work was supported by the National Natural Science Foundation of China (41274147 and 41674155) and the Youth Innovation Promotion Association (No. 2017258), Chinese Academy of Sciences. Support for the US authors was provided by NASA. The newly compiled plasmapause location model coded in Interactive Data Language (IDL) is available by request through email to the corresponding author.

References:

Akasofu, S.-I. (1964), The development of the auroral substorm, *Planet. Space Sci*., 12, 273-282, doi:10.1016/0032-0633(64)90151-5.

Bandić, M., G. Verbanac, M. B. Moldwin, V. Pierrard, and G. Piredda (2016), MLT dependence in the relationship between plasmapause, solar wind, and geomagnetic activity based on CRRES: 1990–1991, *J. Geophys. Res. Space Physics*, 121, 4397–4408, doi:10.1002/2015JA022278.

Cai, L., S. Y. Ma, and Y. L. Zhou (2010), Prediction of SYM-H index during large storms by NARX neural network from IMF and solar wind data, *Ann. Geophys.*, 28, 381-393, doi:10.5194/angeo-28-381-2010.

Carpenter, D. L. (1963), Whistler evidence of a 'Knee' in the magnetospheric ionization density profile, *J. Geophys. Res.*, 68(6), 1675-1682, doi:10.1029/ JZ068i006p01675.

Carpenter, D. L., and R. R. Anderson (1992), An ISEE/whistler model of equatorial electron density in the magnetosphere, *J. Geophys. Res.*, 97, 1097–1108, doi:10.1029/91JA01548.

Cho, J., D.-Y. Lee, J.-H. Kim, D.-K. Shin, K.-C. Kim, and D. Turner (2015), New model fit functions of the plasmapause location determined using THEMIS observations during the ascending phase of Solar Cycle 24, *J. Geophys. Res. Space Physics*, 120, 2877–2889, doi:10.1002/2015JA021030.

Darrouzet, F., J. De Keyser, P. M. E. Décréau, D. L. Gallagher, V. Pierrard, J. F. Lemaire, B. R. Sandel, I. Dandouras, H. Matsui, M. Dunlop, J. Cabrera, A. Masson, P. Canu, J. G. Trotignon, J. L. Rauch, and M. André (2006), Analysis of plasmaspheric plumes: CLUSTER and IMAGE

observations, *Ann. Geophys.*, 24, 1737–1758, doi:10.5194/angeo-24-1737-2006.

Darrouzet, F., De Keyser, J., and Pierrard, V. (Eds.) (2009), *The Earth's Plasmasphere: A Cluster and IMAGE Perspective*, Springer, New York, U.S.A., doi:10.1007/978-1-4419-1323-4.

Efron, B., and R. J. Tibshirani (1993), *An Introduction to the Bootstrap*, 436pp., Chapman and Hall, New York.

Fok, M.-C., T. E. Moore, J. U. Kozyra, G. C. Ho, and D. C. Hamilton (1995), Three-dimensional ring current decay model, *J. Geophys. Res.*, 100(A6), 9619–9632, doi:10.1029/94JA03029.

Fok, M.-C., R. A. Wolf, R. W. Spiro, and T. E. Moore (2001), Comprehensive computational model of the Earth's ring current, *J. Geophys. Res.*, 106, 8417–8424, doi:10.1029/2000JA000235.

Fok, M.-C., A. Glocer, Q. Zheng, R. B. Horne, N. P. Meredith, J. M. Albert, and T. Nagai (2011), Recent developments in the radiation belt, environment model, *J. Atmos. Sol. Terr. Phys.*, 73, 1435–1443, doi:10.1016/j.jastp.2010.09.033.

Fok, M.-C., N. Y. Buzulukova, S.-H. Chen, A. Glocer, T. Nagai, P. Valek, and J. D. Perez (2014), The Comprehensive Inner Magnetosphere-Ionosphere Model, *J. Geophys. Res. Space Physics*, 119, doi:10.1002/2014JA020239.

Gallagher, D. L., P. D. Craven, and R. H. Comfort (2000), Global core plasma model, *J. Geophys. Res.*, 105(A8), 18819-18833, doi: 10.1029/1999JA000241.

Ge, Y. S., X.-Z. Zhou, J. Liang, J. Raeder, M. L. Gilson, E. Donovan, V. Angelopoulos, and A. Runov (2012), Dipolarization fronts and associated auroral activities: 1. Conjugate observations and perspectives from global MHD simulations, *J. Geophys. Res.*, 117, A10226,

doi:10.1029/2012JA017676.

Goldstein, J., B. R. Sandel, W. T. Forrester, and P. H. Reiff (2003a), IMF driven plasmasphere erosion of 10 July 2000, *Geophys. Res. Lett.*, 30(3), 1146, doi:10.1029/2002GL016478.

Goldstein, J., R. W. Spiro, B. R. Sandel, R. A. Wolf, S.-Y. Su, and P. H. Reiff (2003b), Overshielding event of 28–29 July 2000, *Geophys. Res. Lett.*, 30(8), 1421, doi:10.1029/2002GL016644, 2003.

Goldstein, J., B. R. Sandel, M. R. Hairston, and P. H. Reiff (2003c), Control of plasmaspheric dynamics by both convection and sub-auroral polarization stream, *Geophys. Res. Lett.*, 30(24), 2243, doi:10.1029/2003GL018390.

Goldstein, J., B. R. Sandel, W. T. Forrester, M. F. Thomsen, and M. R. Hairston (2005a), Global plasmasphere evolution 22–23 April 2001, *J. Geophys. Res.*, 110, A12218, doi:10.1029/2005JA011282.

Goldstein, J., J. L. Burch, B. R. Sandel, S. B. Mende, P. C:son Brandt, and M. R. Hairston (2005b), Coupled response of the inner magnetosphere and ionosphere on 17 April 2002, *J. Geophys. Res.*, 110, A03205, doi:10.1029/2004JA010712.

Goldstein, J. and Sandel, B. R. (2005), The Global Pattern of Evolution of Plasmaspheric Drainage Plumes, in *Inner Magnetosphere Interactions: New Perspectives from Imaging* (eds J. Burch, M. Schulz and H. Spence), American Geophysical Union, Washington, D. C.. doi: 10.1029/159GM02

Goldstein, J., B. R. Sandelm H. U. Frey, and S. B. Mende (2007), Multiple plasmapause undulations observed by the IMAGE satellite on 20 March 2001, *J. Atmos. Sol.-Terr. Phys.*, 69(3), 322-333,

doi:10.1016/j.jastp.2006.08.010.

Gringauz, K. I. (1963), The structure of the ionized gas envelope of earth from direct measurements in the USSR of local charged particle concentrations, *Planet. Space Sci.*, 11(3), 281-296, doi: 10.1016/0032-0633(63)90030-8.

He, F., X. X. Zhang, B. Chen, and M. ‐ C. Fok (2011), Reconstruction of the plasmasphere from Moon-based EUV images, *J. Geophys. Res.*, 116, A11203, doi:10.1029/2010JA016364.

He, F., X.-X. Zhang, B. Chen, M.-C. Fok, and Y.-L. Zou (2013), Moon-based EUV imaging of the Earth's plasmasphere: Model simulations, *J. Geophys. Res. Space Physics*, 118, 7085–7103, doi:10.1002/2013JA018962.

He, F., X.-X. Zhang, B. Chen, M.-C. Fok, and S. Nakano (2016), Determination of the Earth's plasmapause location from the CE-3 EUVC images, *J. Geophys. Res. Space Physics*, 121, 296–304, doi:10.1002/2015JA021863.

Hu, Y. Q., X. C. Guo, and C. Wang (2007), On the ionospheric and reconnection potentials of the Earth: Results from global MHD simulations, *J. Geophys. Res.*, 112, A07215, doi:10.1029/2006JA012145.

Huber, P. J., and E. M. Ronchetti (2009), *Robust Statistics*, second edition, 363pp., John Wiley and Sons, Inc., New Jersey.

Katus, R. M., M. W. Liemohn, D. L. Gallagher, A. Ridley, and S. Zou (2013), Evidence for potential and inductive convection during intense geomagnetic events using normalized superposed epoch analysis, *J. Geophys. Res. Space Physics*, 118, 181–191, doi:10.1029/2012JA017915.

Katus, R.M., D. L. Gallagher, M.W. Liemohn, A. M. Keesee, and L. K. Sarno-Smith (2015), Statistical storm time examination of MLT-dependent plasmapause location derived from IMAGE EUV, *J. Geophys. Res. Space Physics*, 120, 5545–5559, doi:10.1002/2015JA021225.

Khazanov, G. V., and M. W. Liemohn (1995), Nonsteady state ionosphere-plasmasphere coupling of superthermal electrons, *J. Geophys. Res.*, 100(A6), 9669–9681, doi:10.1029/95JA00526.

Kotova, G.A., M. I.Verigin, and V. V. Bezrukikh (2015), Physics-based reconstruction of the 3-D density distribution in the entire quiet time plasmasphere from measurements along a single pass of an orbiter, *J. Geophys. Res. Space Physics*, 120, 7512–7521, doi:10.1002/2015JA021281.

Larsen, B. A., D. M. Klumpar, and C. Gurgiolo (2006), Correlation between plasmapause position and solar wind parameter, *J. Atmos. Sol. Terr. Phys.*, 69, 334–340, doi:10.1016/j.jastp.2006.06.017.

Lemaire, J. (1989), Plasma distribution models in a rotating magnetic dipole and refilling of plasmaspheric flux tubes, *Phys. Fluids B*, 1, 1519, doi:10.1063/1.858928.

Liemohn, M. W., A. J. Ridley, D. L. Gallagher, D. M. Ober, and J. U. Kozyra (2004), Dependence of plasmaspheric morphology on the electric field description during the recovery phase of the 17 April 2002 magnetic storm, *J. Geophys. Res.*, 109, A03209, doi:10.1029/2003JA010304.

Liemohn, M. W., A. J. Ridley, J. U. Kozyra, D. L. Gallagher, M. F. Thomsen, M. G. Henderson, M. H. Denton, P. C. Brandt, and J. Goldstein (2006), Analyzing electric field morphology through data-model comparisons of the GEM IM/S Assessment Challenge events, *J. Geophys. Res.*, 111, A11S11, doi: 10.1029/2006JA011700.

Liemohn, M. W., and R. Katus (2012), Is the storm time response of the inner magnetospheric hot

ions universally similar or driver dependent?, *J. Geophys. Res.*, 117, A00L03, doi:10.1029/2011JA017389.

Liu, X., W. Liu, J. B. Cao, H. S. Fu, J. Yu, and X. Li (2015), Dynamic plasmapause model based on THEMIS measurements, *J. Geophys. Res. Space Physics*, 120, 10,543–10,556, doi:10.1002/2015JA021801.

Lu, J. Y., H. Jing, Z. Q. Liu, K. Kabin, and Y. Jiang (2013), Energy transfer across the magnetopause for northward and southward interplanetary magnetic fields, *J. Geophys. Res. Space Physics*, 118, 2021–2033, doi:10.1002/jgra.50093.

Luo, B., X. Li, M. Temerin, and S. Liu (2013), Prediction of the AU, AL, and AE indices using solar wind parameters, *J. Geophys. Res. Space Physics*, 118, 7683–7694, doi:10.1002/2013JA019188.

McPherron, R. L., C. T. Russell, and M. P. Aubry (1973), Satellite studies of magnetospheric substorms on August 15, 1968. IX. Phenomenological model for substorms, *J. Geophys. Res., 78*, 3131-3149, doi:10.1029/JA078i016p03131.

Moldwin, M. B., L. Downward, H. K. Rassoul, R. Amin, and R. R. Anderson (2002), A new model of the location of the plasmapause: CRRES results, *J. Geophys. Res.*, 107(A11), 1339, doi:10.1029/2001JA009211.

Moldwin, M. B., J. Howard, J. Sanny, J. D. Bocchicchio, H. K. Rassoul, and R. R. Anderson (2004), Plasmaspheric plumes: CRRES observations of enhanced density beyond the plasmapause, *J. Geophys. Res.*, 109, A05202, doi:10.1029/2003JA010320.

Ober, D. M., J. L. Horwitz, and D. L. Gallagher (1997), Formation of density troughs embedded in

the outer plasmasphere by subauroral ion drift events, *J. Geophys. Res.*, 102(A7), 14595–14602, doi:10.1029/97JA01046.

O'Brien, T. P., and M. B. Moldwin (2003), Empirical plasmapause models from magnetic indices, *Geophys. Res. Lett.*, 30(4), 1152, doi:10.1029/2002GL016007.

Papitashvili, V. O., N. E. Papitashvili, and J. H. King (2000), Solar cycle effects in planetary geomagnetic activity: Analysis of 36-year long OMNI dataset, *Geophys. Res. Lett.*, 27, 2797–2800, doi:10.1029/2000GL000064.

Perreault, P., and S. I. Akasofu (1978), Study of geomagnetic storms, *Geophys. J. R. Astron. Soc.*, 54(3), 547–573, doi:10.1111/j.1365-246X.1978.tb05494.x.

Pierrard, V., J. Goldstein, N. Andre, V. K. Jordanova, G. A. Kotova, J. F. Lemaire, M. W. Liemohn, and H. Matsui (2009), Recent progress in physics-based models of the plasmasphere, *Space Sci. Rev.*, 145, 92–229, doi:10.1007/s11214-008-9480-7.

Pierrard, V., and K. Stegen (2008), A three-dimensional dynamic kinetic model of the plasmasphere, *J. Geophys. Res.*, 113, A10209, doi:10.1029/2008JA013060.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992), *Numerical Recipes*, Cambridge Univ. Press, Cambridge, U. K.

Pulkkinen, T. I., M. Palmroth, and T. Laitinen (2008), Energy as a tracer of magnetospheric processes: GUMICS-4 global MHD results and observations compared, *J. Atmos. Sol. Terr. Phys.*, 70(5), 687–707, doi:10.1016/j.jastp.2007.10.011.

Raeder, J., R. L. McPherron, L. A. Frank, S. Kokubun, G. Lu, T. Mukai, W. R. Paterson, J. B.

Sigwarth, H. J. Singer, and J. A. Slavin (2001), Global simulation of the Geospace Environment Modeling substorm challenge event, *J. Geophys. Res.*, 106(A1), 381–395, doi:10.1029/2000JA000605.

Reiff, P. H. (1990), The use and misuse of statistics in space physics, *J. Geomagn. Geoelectr.*, 42, 1145–1174, doi:10.5636/jgg.42.1145.

Runov, A., V. Angelopoulos, M. I. Sitnov, V. A. Sergeev, J. Bonnell, J. P. McFadden, D. Larson, K.-H. Glassmeier, and U. Auster (2009), THEMIS observations of an earthward propagating dipolarization front, *Geophys. Res. Lett.*, 36, L14106, doi:10.1029/2009GL038980.

Sandel, B. R., J. Goldstein, D. L. Gallagher, and M. Spasojević (2003), Extreme ultraviolet imager observations of the structure and dynamics of the plasmasphere, *Space Sci. Rev.*, 109(1), 25–46, doi:10.1023/B:SPAC.0000007511.47727.5b.

Spasojević, M., J. Goldstein, D. L. Carpenter, U. S. Inan, B. R. Sandel, M. B. Moldwin, and B. W. Reinisch (2003), Global response of the plasmasphere to a geomagnetic disturbance, *J. Geophys. Res.*, 108(A9), 1340, doi:10.1029/2003JA009987.

Takalo, J., and J. Timonen (1997), Neural network prediction of AE data, *Geophys. Res. Lett.*, 24(19), 2403-2406, doi:10.1002/97GL02457.

Verbanac, G., V. Pierrard, M. Bandic, F. Darrouzet, J.-L. Rauch, and P. Décréau (2015), Relationship between plasmapause, solar wind and geomagnetic activity between 2007 and 2011, *Ann. Geophys.*, 33, 1271–1283, doi:10.5194/angeo-33-1271-2015.

Verigin, M. I., G. A. Kotova, V. V. Bezrukikh, and O. S. Aken'tieva (2012), Restoration of the

proton density distribution in the Earth's plasmasphere from measurements along the INTERBALL-1 satellite orbit, *Geomagn. Aeron.*, 52(6), 725-729, doi: 10.1134/S0016793212060138

Tables:

Table 1. Information of Various Empirical Plasmapause Models

Model	Data Source	Year	Parameter [a]	MLT-Dependence
CAA-1992	ISEE-1	1977-1983	*Kp*	No
MOL-2002	CRRES	1990-1991	*Kp*	No
OBM-2003	CRRES	1990-1991	*Kp*, *Dst*, or *AE*	Yes
LAR-2007	IMAGE	2001	B_Z and $\boldsymbol{\phi}$	No
CHO-2015	THEMIS	2008-2012	B_Z, and *AE*	No
LIU-2015	THEMIS	2009-2013	*SYM-H*, *AL*, *AU*, *AE*, and *Kp*	Yes [b]
VER-2015	Cluster	2007-2011	B_Z, BV_{SW}, $d\Phi_{mp}/dt$ [c], *Kp*, *Dst*, or *AE*	Yes [d]
BAN-2016	CRRES	1990-1991	V_{SW}, BV_{SW}, $d\Phi_{mp}/dt$, *Ap*, *Dst*, or *AE*	Yes [e]
NSW-GDP	Multiple	1977-2015	V_{SW}, B_Z, *AE*, and *SYM-H*	Yes

[a] In this column, "or" indicates that one of the parameters is used in the model, e.g. the OBM-2003 contains three models, VER-2015 contains six models, and "and" indicates that all of the parameters are included in the model.

[b] The MLT resolution is 1 h.

[c] $d\Phi_{mp}/dt = V_{SW}^{4/3} B_T^{2/3} \sin^{8/3}(\theta_c/2)$, where $B_T = \sqrt{B_Y^2 + B_Z^2}$ is the projection of IMF on the GSM yz plane, $\theta_c = \text{atan}(B_Y, B_Z)$, and B is the IMF magnitude.

[d] Just considered three MLT sectors (1 h – 7 h, 7 h – 16 h, and 16 h – 1 h), in each of which the plasmapause location is constant.

[e] Only the *Ap*, *Dst*, or *AE* model is MLT-dependent.

Table 2. Fitted Coefficients for the Curves in Figure 3

f_0	f_1	f_2	g_0	g_1	g_2	g_3	g_4	g_5	g_6
2.7883	9.3345	268.6071	1.7569	3.3058	0.7945	3.6678	27.1041	0.1983	0.0010
±0.0753	±0.4005	±12.6382	±0.1105	±0.1498	±0.0871	±0.0311	±1.8705	±0.0925	±0.0006

Table 3. Contributions of Storm and Substorm Activities to Plasmapause Shrinkage under Different Solar Wind Conditions.

Soar Wind Condition (mV/m)	$\lvert E_{SW}\rvert \le 2.0$	$2.0 < \lvert E_{SW}\rvert \le 4.0$	$4.0 < \lvert E_{SW}\rvert \le 6.0$	$\lvert E_{SW}\rvert > 6.0$
$dL_{PP\text{-}Substorm}$ (R_E)	1.28	1.20	1.08	1.00
$dL_{PP\text{-}Storm}$ (R_E)	1.59	1.41	1.31	1.16

Table 4. Fitted Coefficients for the NSW-GDP Model and Corresponding Standard Deviations (σ)

Coefficient	Value	σ	Coefficient	Value	σ	Coefficient	Value	σ
g_{00}	2.051	0.039	g_{52}	143.584	5.011	k_{02}	-0.405	0.0132
g_{01}	28.753	1.938	g_{60}	10.815	0.768	k_{03}	0.0175	0.00093
g_{02}	71.077	3.593	g_{61}	95.789	4.668	k_{04}	-0.045	0.0022
g_{10}	1.627	0.073	h_{00}	1.139	0.0262	k_{05}	-0.005	0.00019
g_{11}	9.065	0.389	h_{01}	-0.0522	0.0050	k_{10}	-0.513	0.0134
g_{12}	259.032	11.952	h_{02}	0.0124	0.0017	k_{11}	0.0165	0.0011
g_{20}	1.317	0.027	h_{10}	0.630	0.0234	k_{12}	0.276	0.0129
g_{21}	34.702	3.286	h_{11}	-0.236	0.0045	k_{13}	-0.0329	0.0024
g_{22}	99.965	6.463	h_{12}	0.0124	0.0015	k_{20}	1.175	0.0456
g_{30}	3.722	0.038	h_{20}	16.656	0.251	k_{21}	-0.075	0.0043
g_{31}	-0.00124	0.00007	h_{21}	0.268	0.0073	k_{22}	-1.825	0.078
g_{40}	15.691	0.443	h_{30}	0.752	0.0227	k_{23}	0.125	0.014
g_{41}	130.797	7.632	h_{31}	-0.301	0.0131	k_{30}	0.75	0.05
g_{42}	179.516	8.620	h_{32}	0.238	0.0147	k_{31}	-0.25	0.01
g_{50}	0.0239	0.003	k_{00}	-0.165	0.0089	k_{32}	4.0	0.25
g_{51}	8.068	0.551	k_{01}	-0.0065	0.0003	k_{33}	2.0	0.10

Figures and Captions:

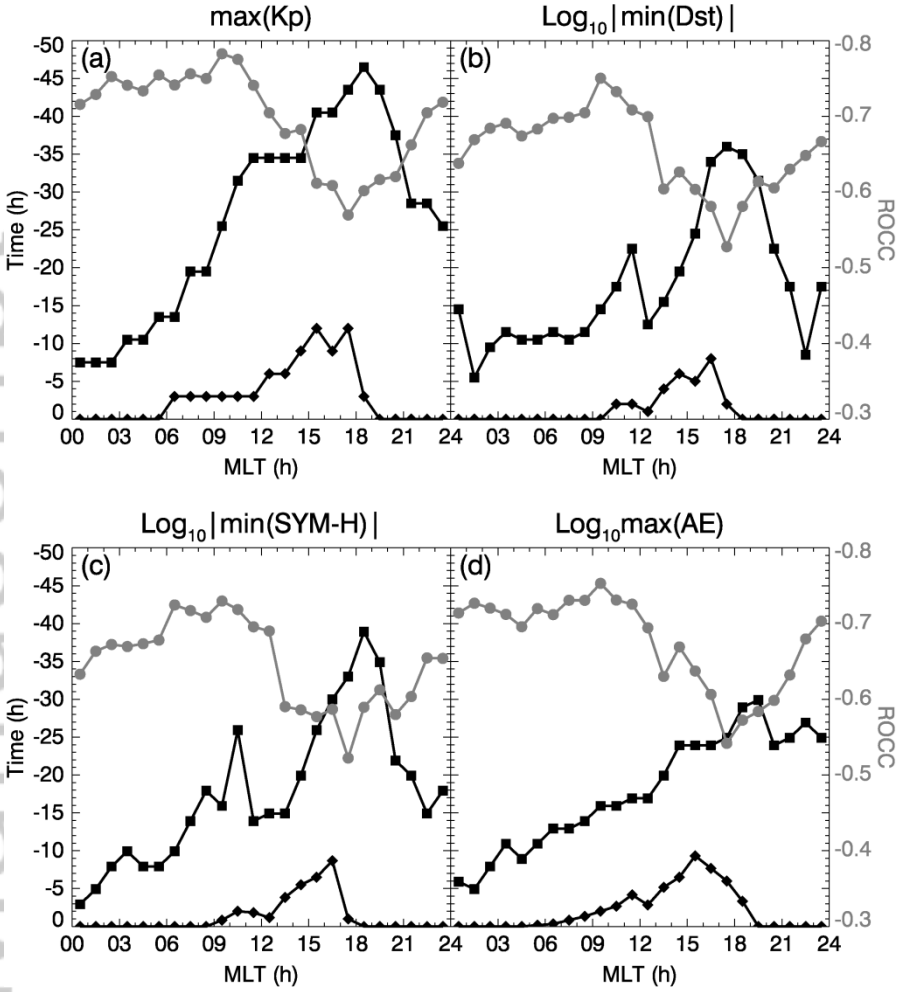


Figure 1. Variations of t_1 (*squares*) and t_2 (*diamonds*) for (a) *Kp*, (b) *Dst*, (c) *SYM-H*, and (d) *AE*, respectively, for the strongest correlation with the corresponding ROCC (*dots*) shown in black.

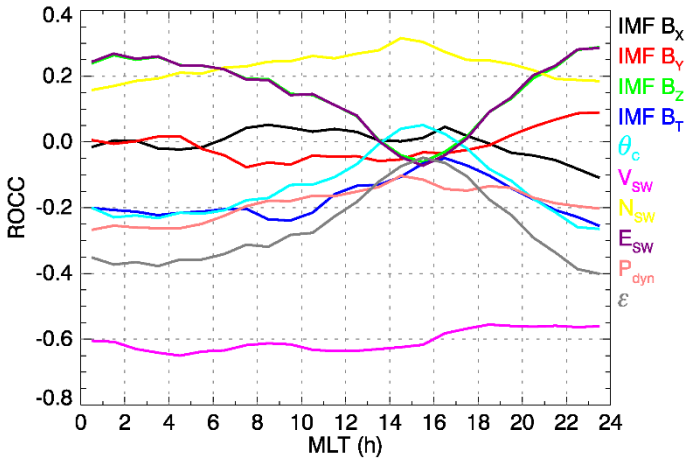


Figure 2. ROCCs between L_{PP} and various solar wind and IMF parameters. Different colors representing different parameters as shown at the right side. Note that the green IMF B_Z curve is directly beneath the purple θ_c curve.

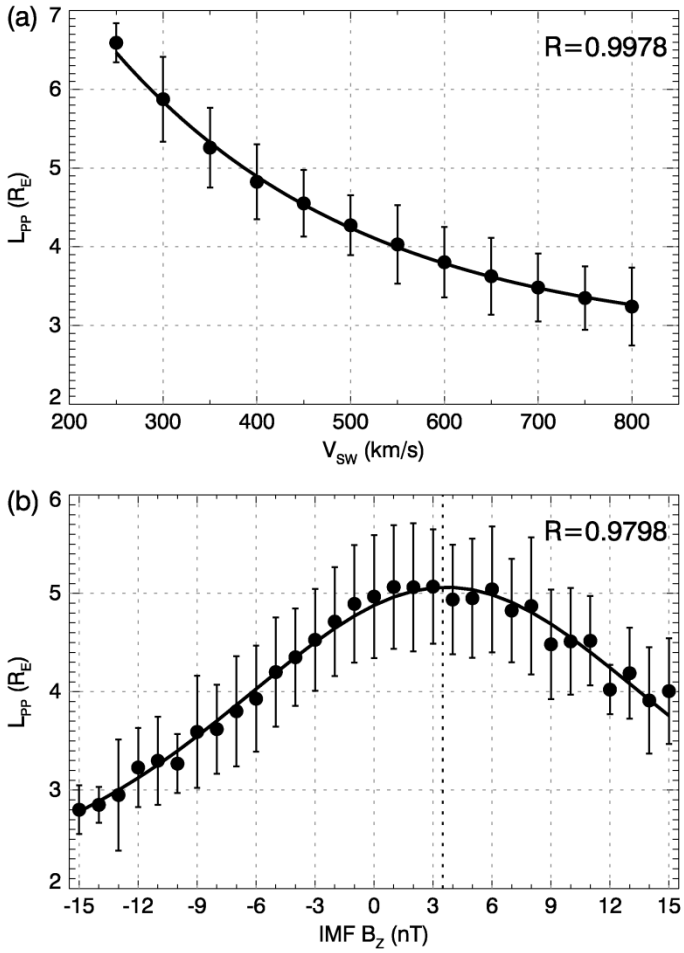


Figure 3. Averaged L_{PP} versus (a) V_{SW} and (b) IMF B_Z, respectively. The black dots represent the averaged L_{PP} in 50 km/s intervals for V_{SW} in the top panel and the averaged L_{PP} in 1 nT intervals for IMF B_Z in the bottom panel with the standard deviations shown by the vertical bars. The curve in (a) represents the best fitting of the dots to an exponentially decreasing function, and the curve in (b) represents the best fitting of the dots to an inverse polynomial function with the peak represented by the vertical dashed line. The correlation coefficients (R) are shown at the top left corners of each panel.

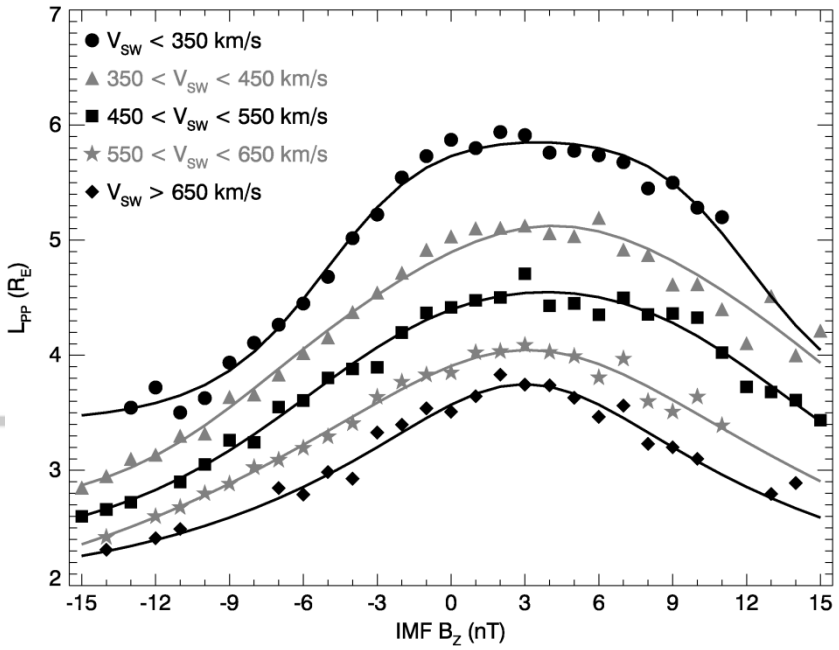


Figure 4. L_{PP}-B_Z variations for different V_{SW}. The dots represent the averaged L_{PP} in 1 nT B_Z intervals. The curves represent the best fittings to equation (4). Different symbols representing different V_{SW} values are shown at the upper left corner.

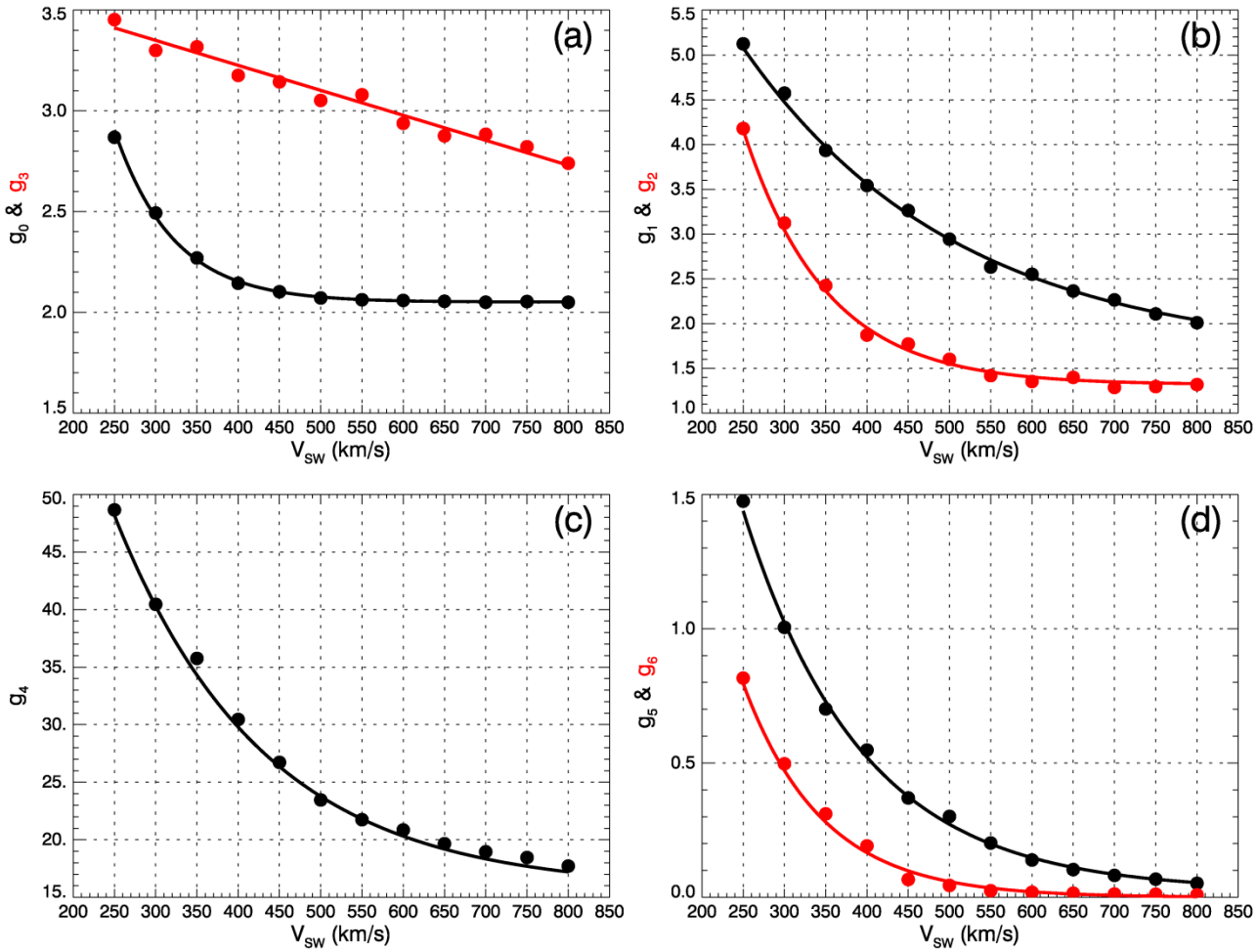


Figure 5. Example of best fittings of the functions in equation (5) for (a) g_0 (black) and g_3 (red), (b) g_1 (black) and g_2 (red), (c) g_4, and (d) g_5 (black) and g_6 (red). In each panel, the dots represent the values of the parameters fitted using equation (5), and the thick lines represent best fittings of the dots to the equation (6).

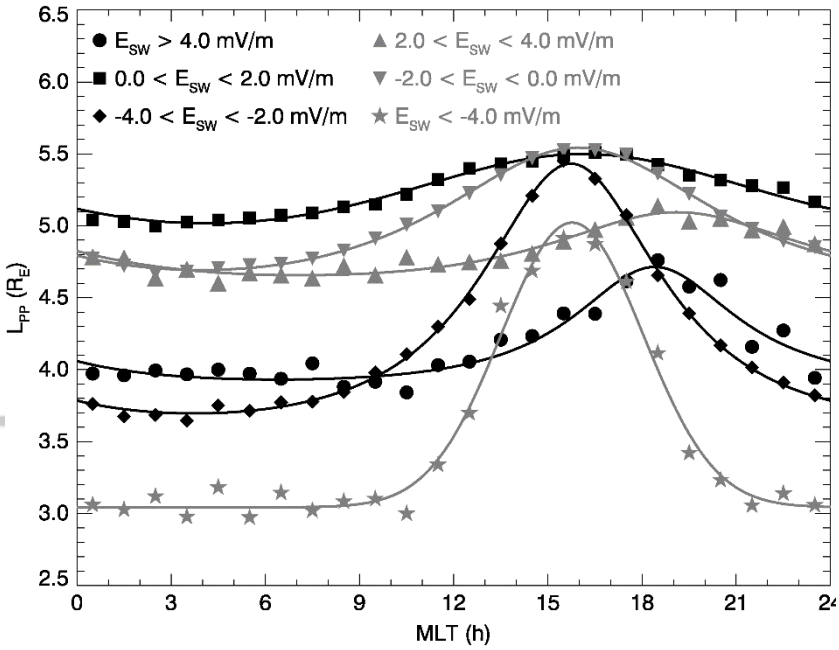


Figure 6. MLT-dependence of the plasmapause locations under different conditions of E_{SW}. The dots represent the averaged L_{PP} in 1 h MLT intervals. The curves represent the best fittings to equation (7). Different symbols representing different E_{SW} values are shown at the upper left corner.

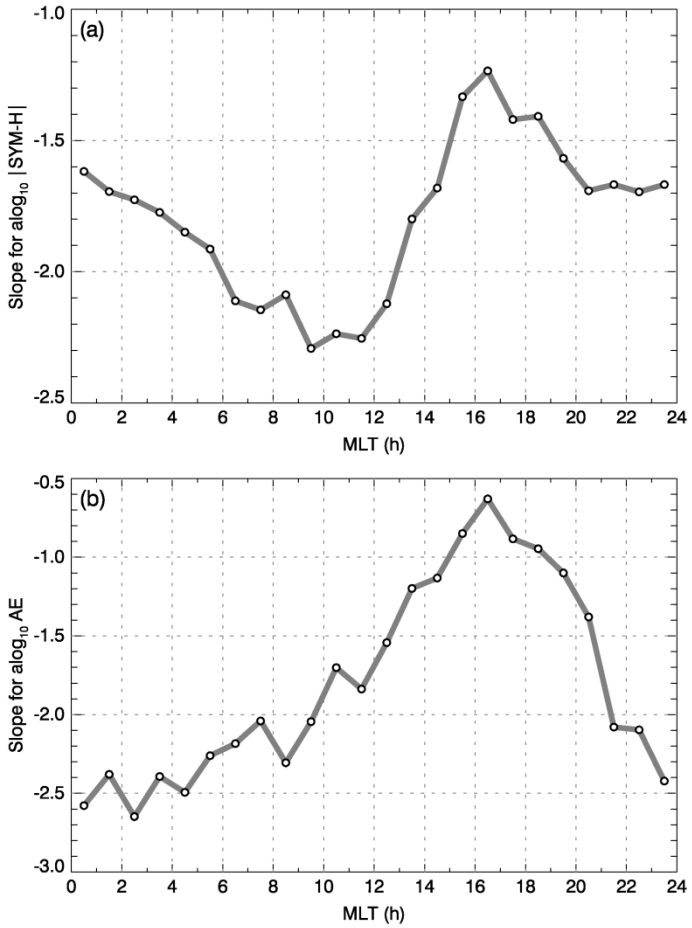


Figure 7. Slopes of the linear fitting functions of L_{PP} to (a) alog_{10} $|SYM\text{-}H|$ and (b) alog_{10} AE at different MLT sectors in 1 h intervals, respectively.

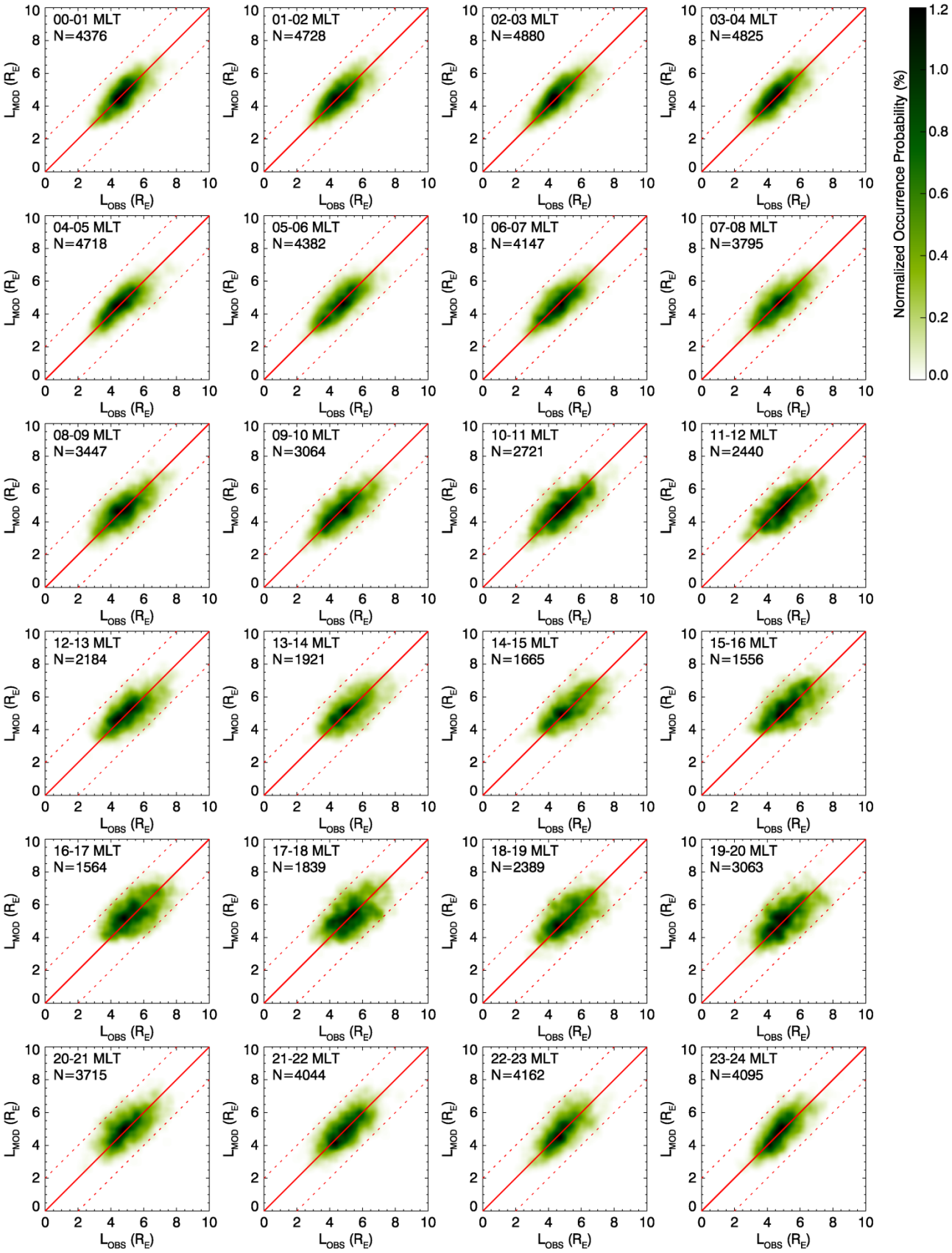


Figure 8. Normalized occurrence probabilities of the L_{MOD} relative to L_{OBS}. The black solid lines indicate $L_{MOD} = L_{OBS}$, and the dashed lines are plotted at $L_{MOD} = L_{OBS} \pm 2.0$, respectively. The linear scaled color bar denoting the occurrence probabilities are shown at the upper right corner. The numbers of plasmapause locations in each MLT sector are shown at the upper left corners of each panel.

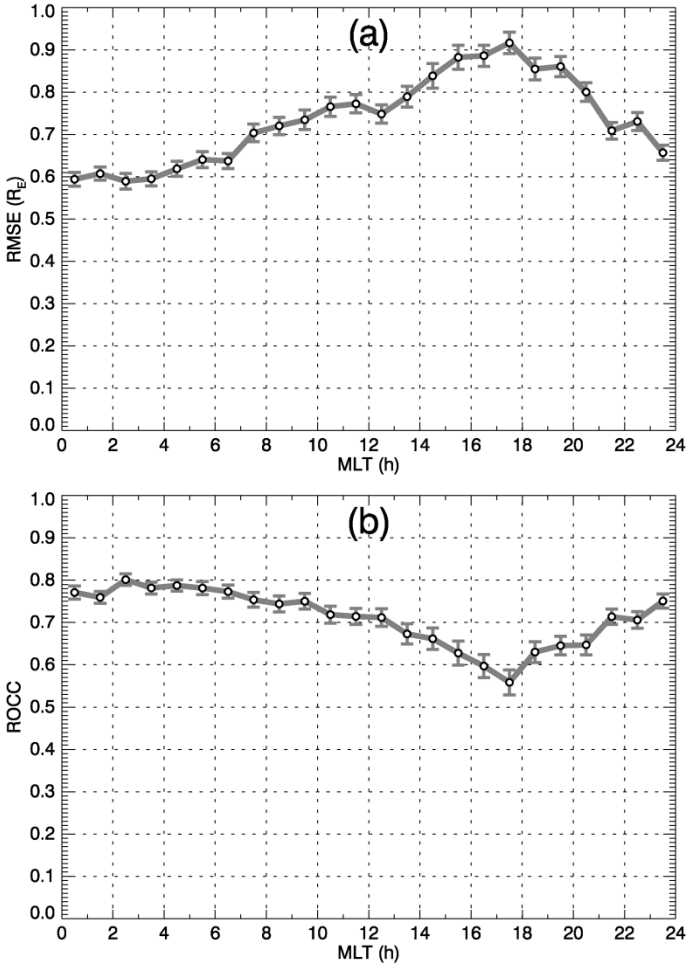


Figure 9. The RMSE and ROCC of the NSW-GDP model versus MLT. The error bars denote the 1σ confidence found using bootstrap resampling.

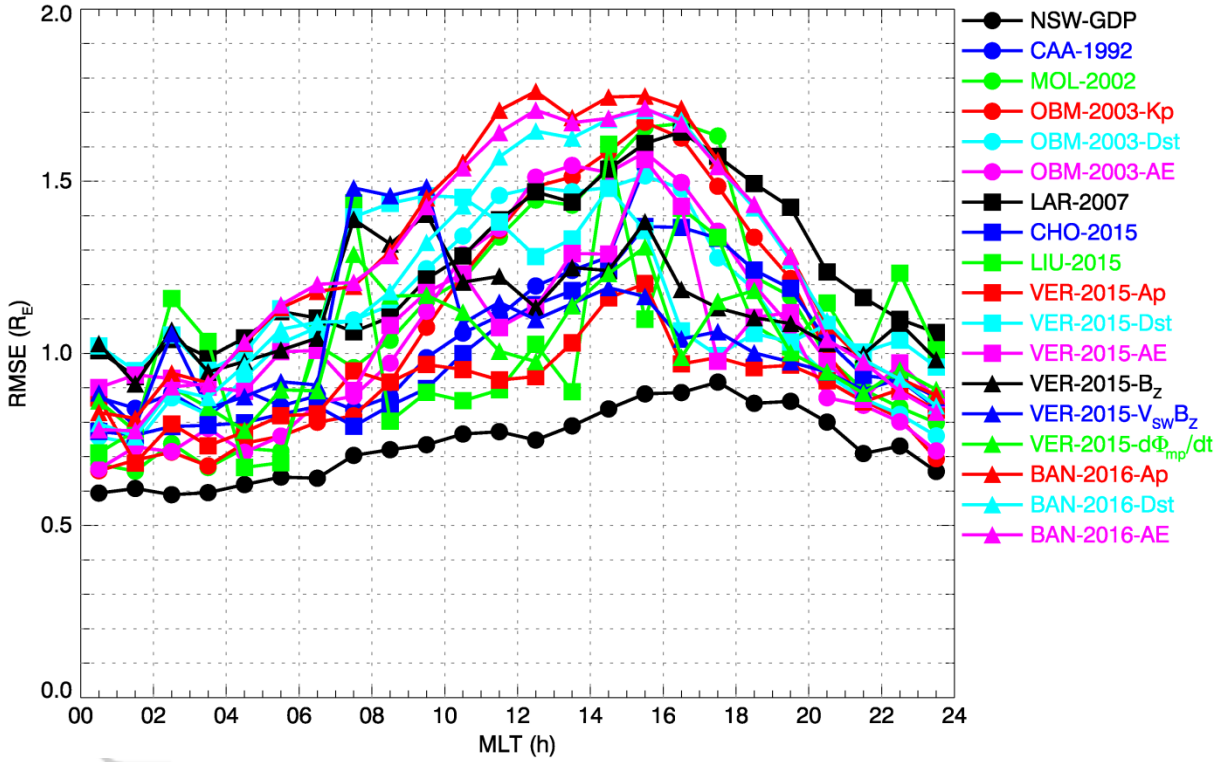


Figure 10. RMSE of various models as a function of MLT. Different colors and symbols representing different models are shown at the right.

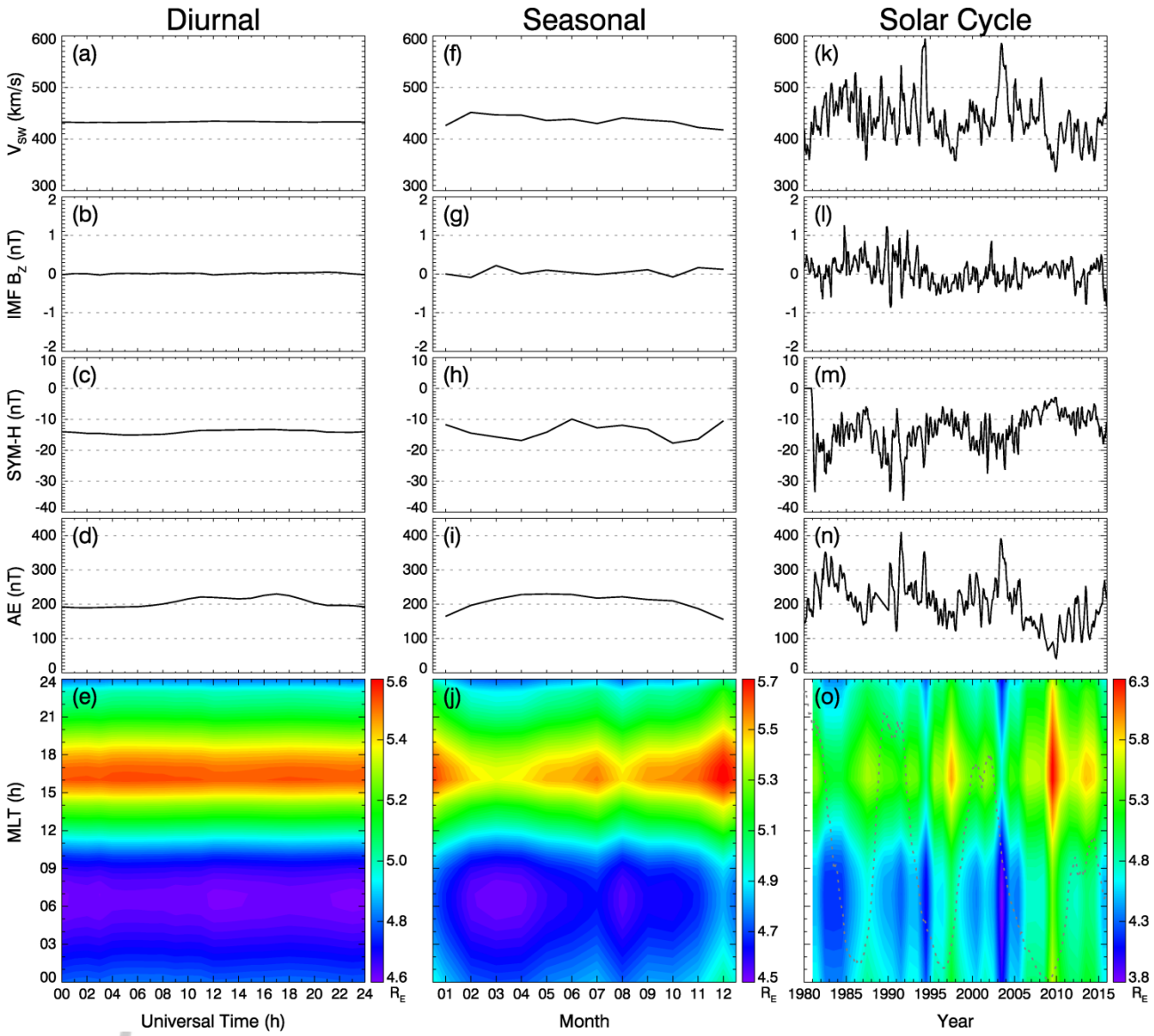
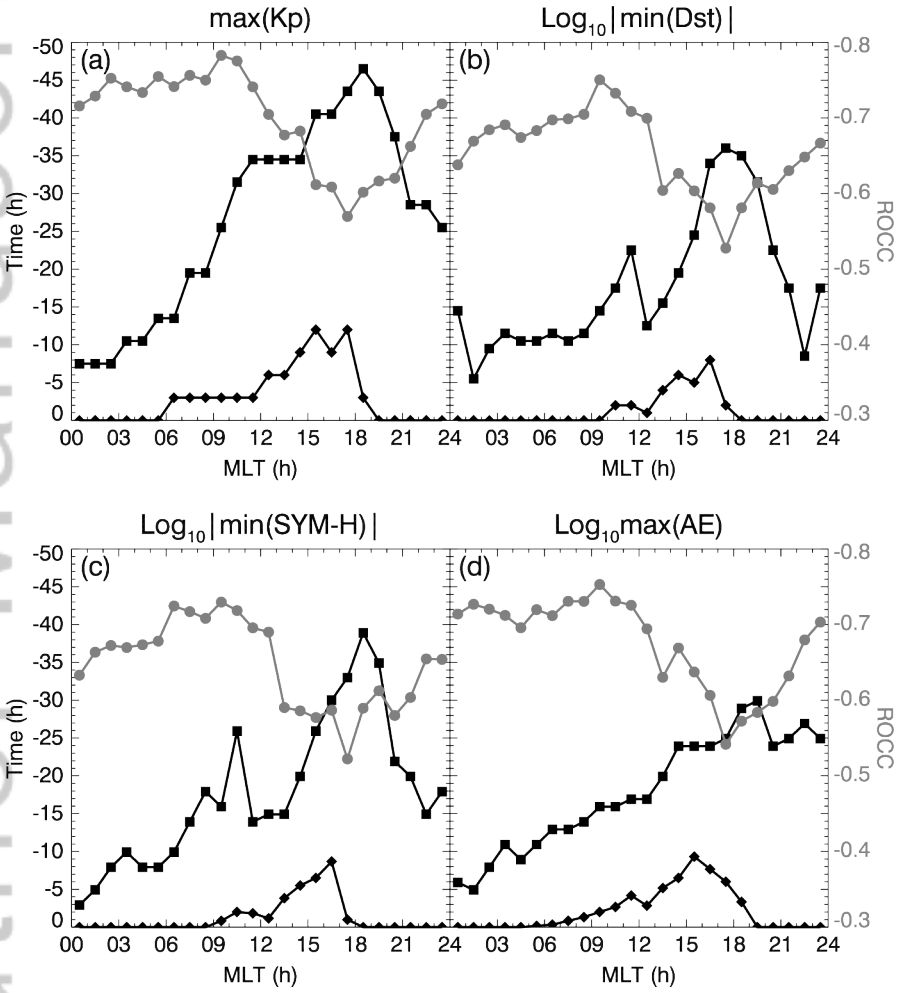
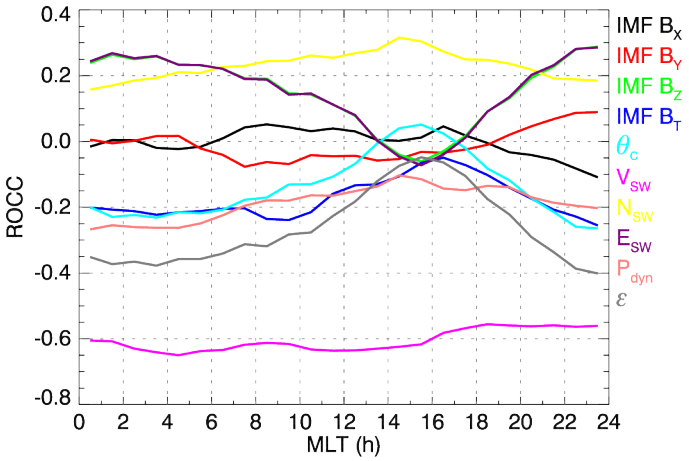


Figure 11. Diurnal (*left column*), seasonal (*middle column*), and solar cycle (*right column*) variations of the plasmapause locations simulated by the NSW-GDP model. From top to bottom in each column, shown are variations of the averaged V_{SW}, IMF B_Z, *SYM-H*, and *AE*, respectively. From left to right, the averages are calculated hourly, monthly, and monthly, respectively, except for panel (o) where the plasmapause locations are averaged yearly. The grey dashed curve in (o) is the scaled sunspot numbers indicating solar cycles. Note that panels (e), (j) and (o) have different color

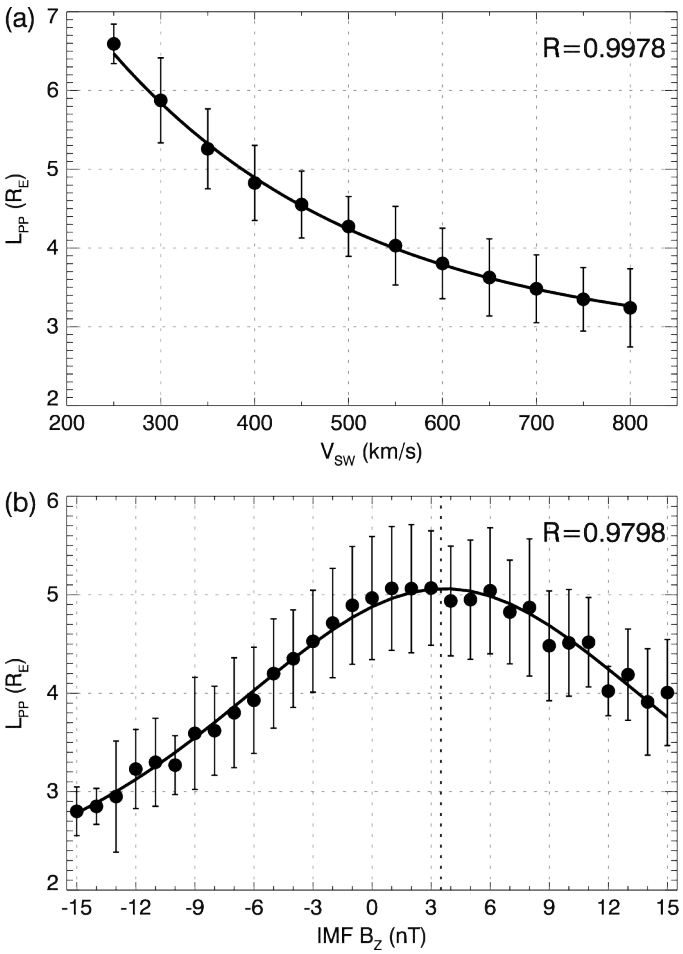
bars, which are scaled as shown to the right of each panel.



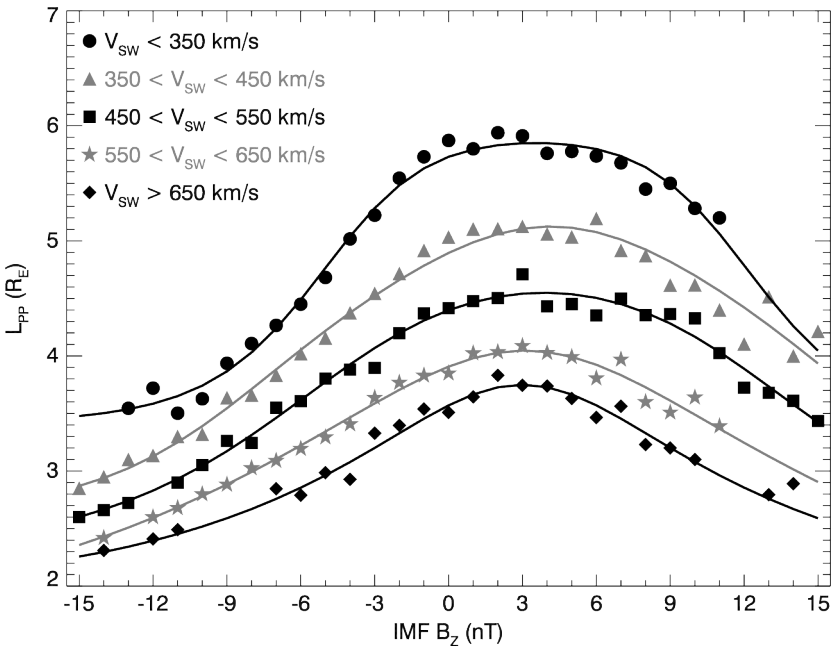
2017ja023913-f01-z-.eps



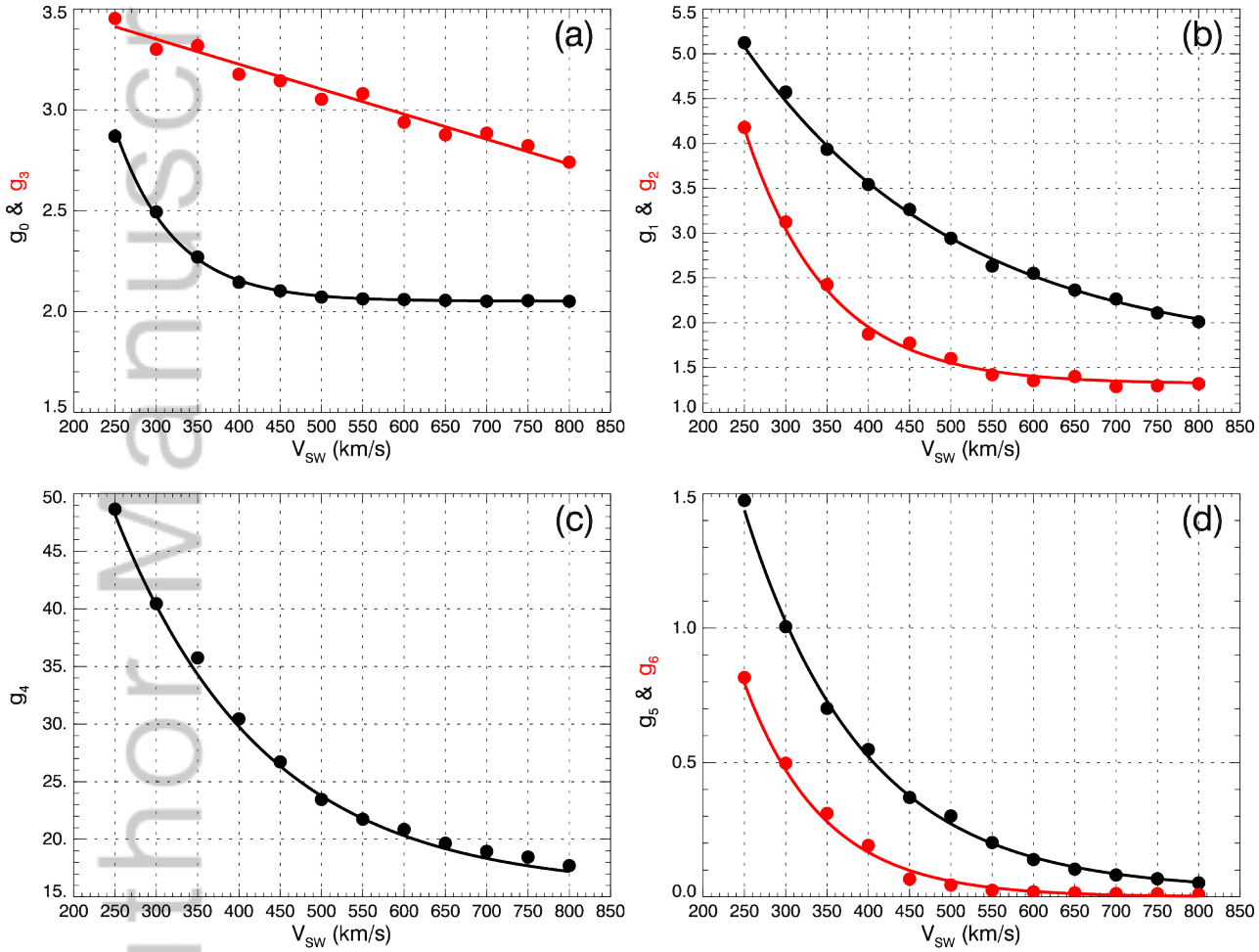
2017ja023913-f02-z-.eps



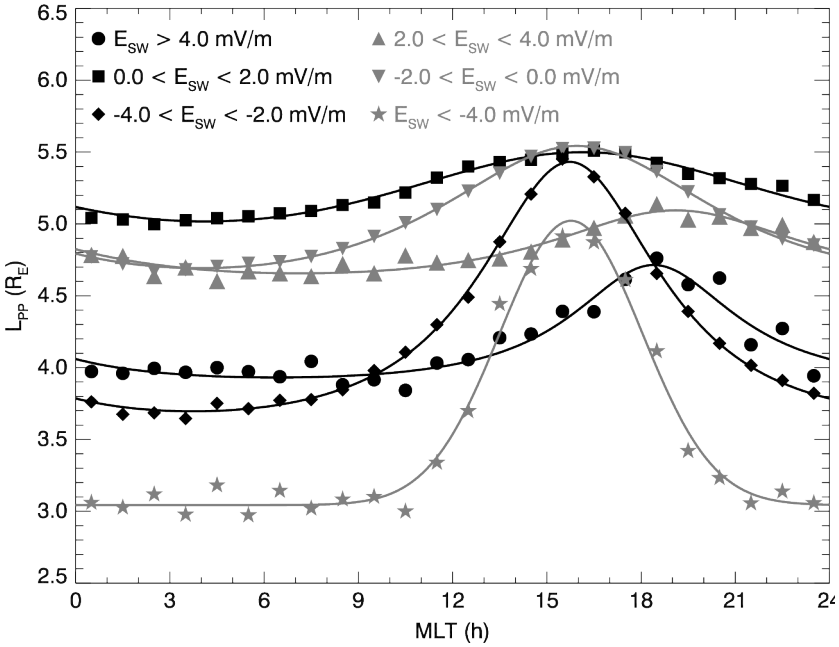
2017ja023913-f03-z-.eps



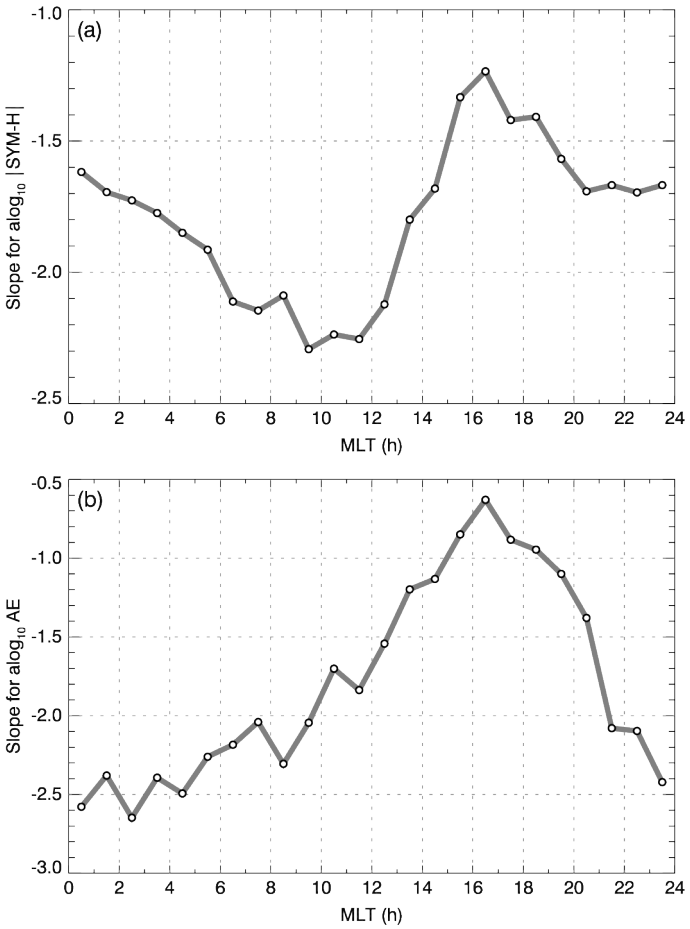
2017ja023913-f04-z-.eps



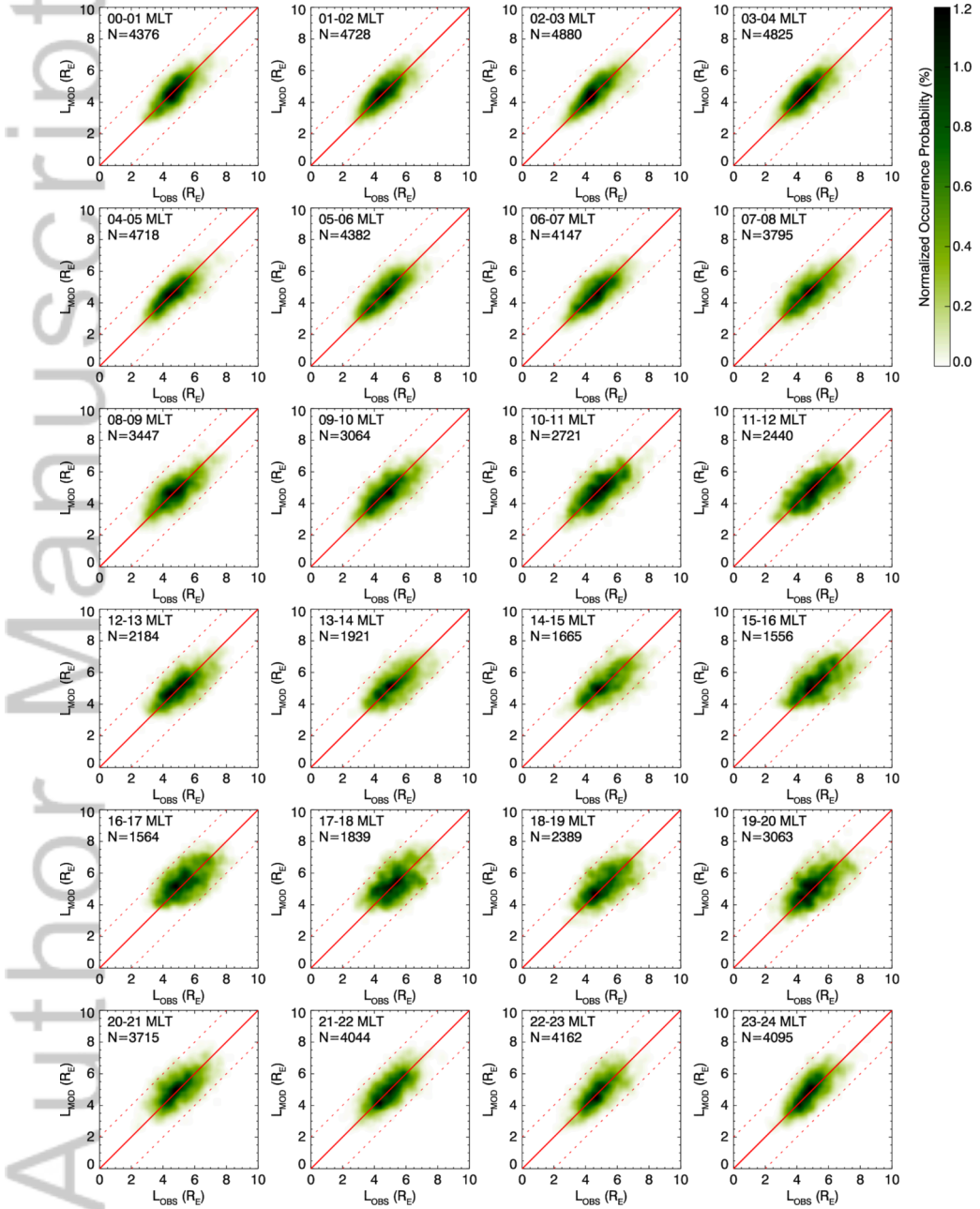
2017ja023913-f05-z-.eps



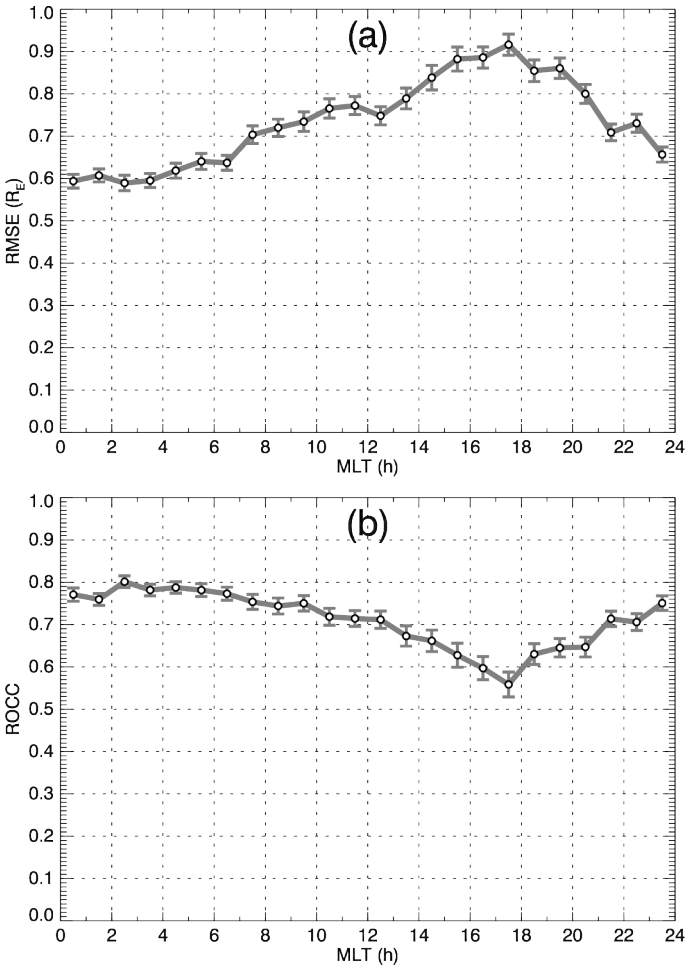
2017ja023913-f06-z-.eps



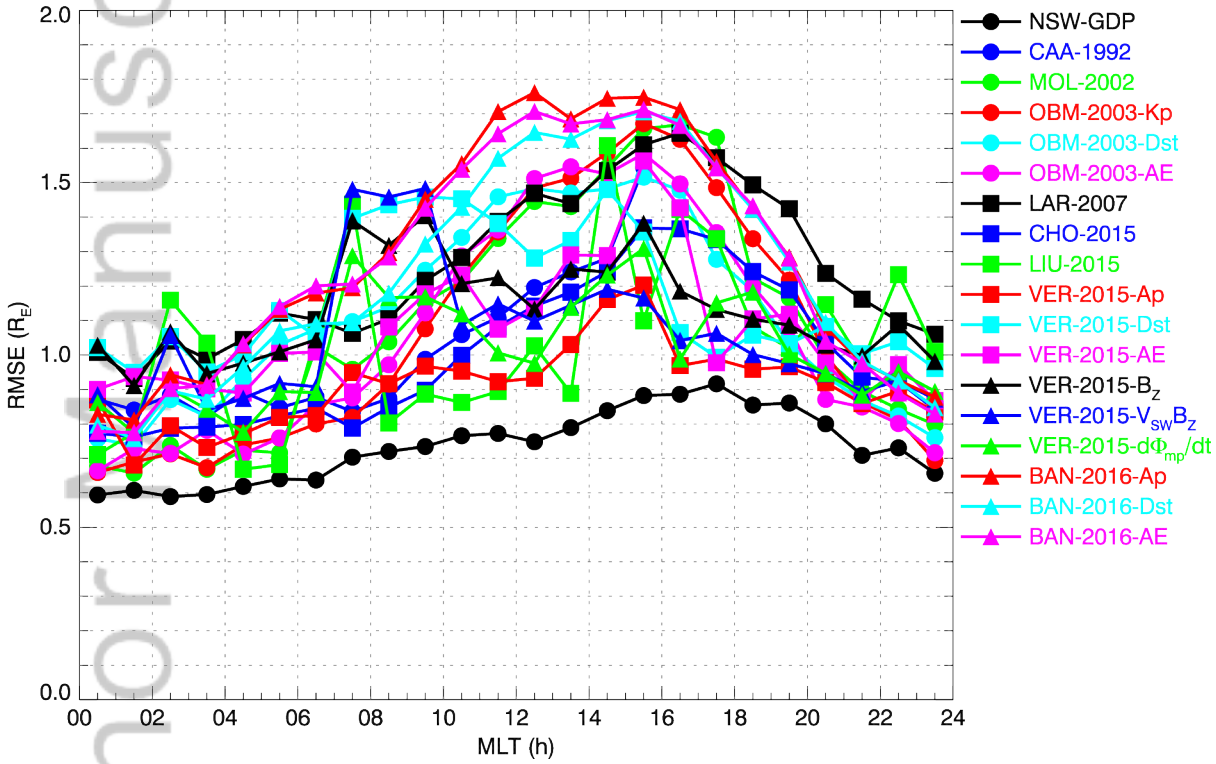
2017ja023913-f07-z-.eps



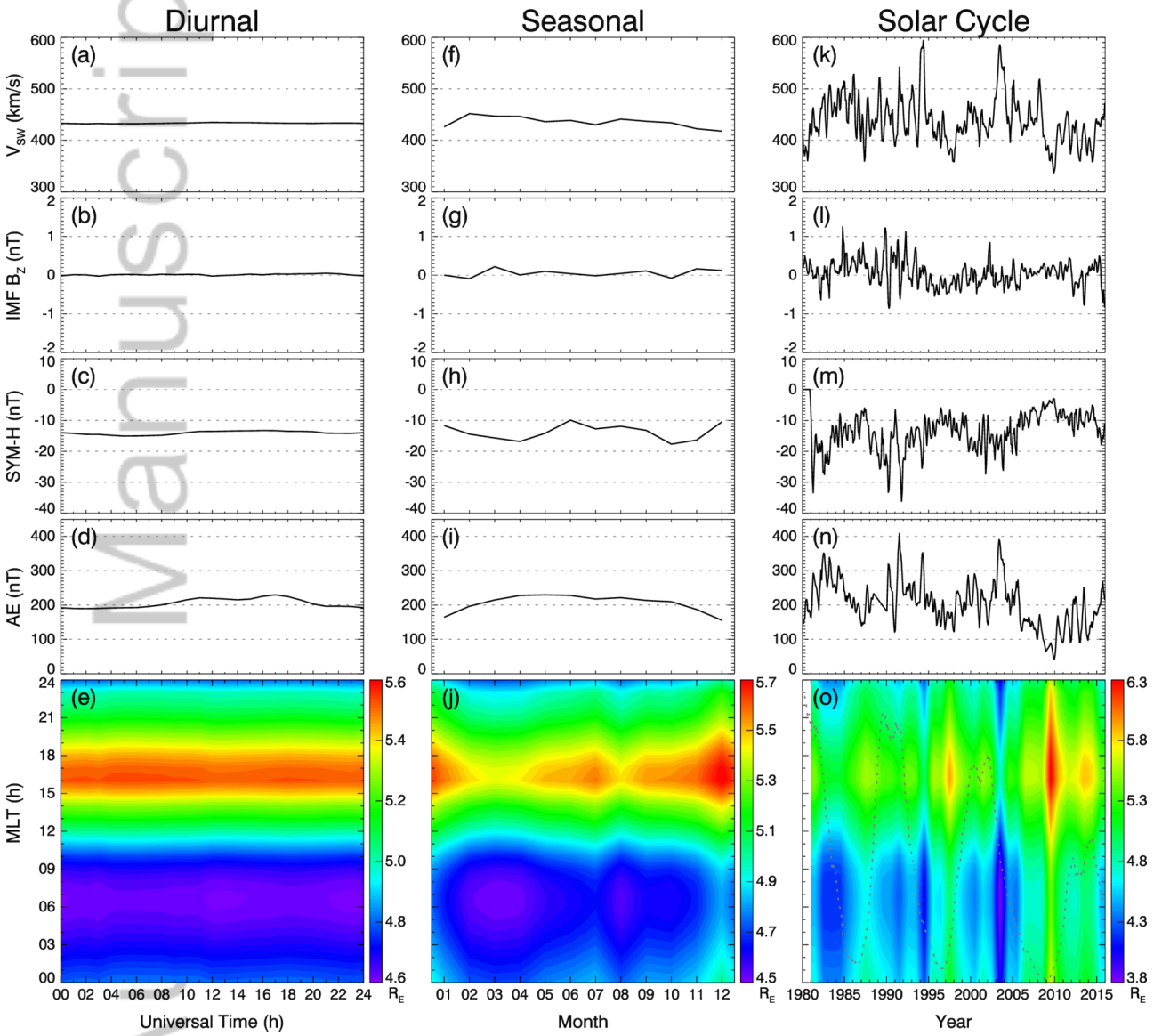
2017ja023913-f08-z-.eps



2017ja023913-f09-z-.eps



2017ja023913-f10-z-.eps



2017ja023913-f11-z-.eps