# The Tractrix Magnetopause: A Novel Physics-Based Functional Form for the Magnetopause Shape

C. J. O'Brien[1], Michael R Collier[2], Brian Walsh[1], David Gary Sibeck[3], and E. Taylor[4]

[1]Boston University
[2]NASA Goddard SFC
[3]GSFC
[4]Howard University

November 22, 2022

**Abstract**

A new model for the shape of the magnetopause is presented using a closed-form analytic function known as a tractrix. This shape is derived from several physics-based underpinnings, eliminating the need for fitting ad-hoc functional forms that, while convenient, are not physically motivated. One feature of the magnetopause predicted by this model is that the magnetotail flares outward until it reaches a constant width, a feature that has significant observational evidence but is seldom represented in functional forms of the magnetopause shape. To optimize the parameters of this model, a dataset of over 13,000 magnetopause crossings from THEMIS/ARTEMIS, Cluster, Geotail, Interball, and several other spacecraft is utilized. Using a Bayesian approach combined with a Markov Chain Monte Carlo (MCMC) method to estimate the posterior probability distribution in parameter space, the maximum likelihood parameters for the model that optimize its performance on this dataset are determined. The modelâ\euros performance is compared to that of other popular models of the magnetopause with a focus on their relative performance, and is shown to outperform models that assume the tail flares outward to infinity at far distances. The optimized model more accurately predicts magnetopause position along the tail than other popular static analytic magnetopause models, while still being easy to implement for a variety of applications.

# The Tractrix Magnetopause: A Novel Physics-Based Functional Form for the Magnetopause Shape

**C. J. O'Brien[1], M. R. Collier[2], B.M. Walsh[1], D.G. Sibeck[2], E. Taylor[3]**

[1]Center for Space Physics, Boston University, Boston, MA, USA
[2]NASA/GSFC, Greenbelt, MD, USA
[3]Howard University, Washington, DC, USA

**Key Points:**

- A new functional form for the shape of the magnetopause is derived directly from a handful of physical assumptions about the magnetopause.
- The tractrix model was optimized on a novel dataset of 901 magnetopause crossings at lunar distances observed with ARTEMIS-P1 combined with a very large dataset of near-Earth and nose magnetopause crossings.
- The tractrix model shows good performance across both regimes covered by this dataset, and is even able to outperform more complex non-axisymmetric models.

**Abstract**

A new model for the shape of the magnetopause is presented using a closed-form analytic function known as a tractrix. This shape is derived from several physics-based underpinnings, eliminating the need for fitting ad-hoc functional forms that, while convenient, are not physically motivated. One feature of the magnetopause predicted by this model is that the magnetotail flares outward until it reaches a constant width, a feature that has significant observational evidence but is seldom represented in functional forms of the magnetopause shape. To optimize the parameters of this model, a dataset of over 13,000 magnetopause crossings from THEMIS/ARTEMIS, Cluster, Geotail, Interball, and several other spacecraft is utilized. Using a Bayesian approach combined with a Markov Chain Monte Carlo (MCMC) method to estimate the posterior probability distribution in parameter space, the maximum likelihood parameters for the model that optimize its performance on this dataset are determined. The model's performance is compared to that of other popular models of the magnetopause with a focus on their relative performance, and is shown to outperform models that assume the tail flares outward to infinity at far distances. The optimized model more accurately predicts magnetopause position along the tail than other popular static analytic magnetopause models, while still being easy to implement for a variety of applications.

## Plain Language Summary

The boundary marking the edge of the extent of Earth's magnetic field into interplanetary space is known as the magnetopause. At this boundary, a great deal of physically interesting phenomena occur. Namely, it is the surface at which magnetic reconnection takes place, locally converting magnetic energy into kinetic energy and broadly opening Earth's magnetosphere to the driving solar wind which fuels geomagnetic storms. In order to predict the location of this surface, empirical models describing the location of the magnetopause have been developed in the past. However, none of these models have been more directly derived from the physics of the magnetopause. Instead, models generally assume some convenient shape, such as a conic section. In this paper, a new empirical model for the magnetopause shape that is based on the physics of the magnetopause is presented, along with some demonstrations that it has as good or better predictive power than other popular, more complex magnetopause models.

## 1 Introduction

Empirical models describing the shape of the magnetopause generally use ad-hoc functional forms that are fit to spacecraft observations of magnetopause crossing locations. This approach is about as old as spacecraft observations of the magnetopause themselves (e.g. Spreiter et al. (1966)), and as understanding of the magnetospheric system has increased, so too has the complexity of models trying to describe it. Some early models were completely static with no dependence on solar wind parameters (Howe & Binsack, 1972), whereas some models created more recently have functional forms with more than 20 tunable parameters (Lin et al., 2010).

The most popular functional forms for empirical magnetopause models have been conic sections. The first model to include a dependence on the z component of the IMF used an ellipse and hyperbola to describe an open or closed magnetopause (Fairfield, 1971). The form has been adopted in numerous models at the Earth (Holzer & Slavin, 1978; Shue et al., 1997, 1998; Chao et al., 2002; Lin et al., 2010; Lu et al., 2011) and elsewhere in the solar system (Winslow et al., 2013). These models generally control the subsolar point as well as the opening angle $\alpha$, which determines to what extent the magnetopause is open (parabola or hyperbola) or closed (elliptical). This angle is often called the "flaring angle". These two parameters are controlled by solar wind conditions, typically the IMF $B_z$ and dynamic pressure, using some tuning parameters. Most are axisymmetric about the Sun-Earth line, but there are notable exceptions that predict North-South and/or East-West asymmetries. These models are generally highly complex, with some models having 18 or even 20 tuning parameters (Lin et al., 2010; Lu et al., 2011).

One notable feature of conic section based models is that the magnetopause distance either expands indefinitely in the distant tail or closes at some distance to form an ellipse. Although some numerical modeling results find a tail width consistent with this behavior (Ogino et al., 1992; Welling & Ridley, 2010; Park et al., 2015), many observations of the distant tail magnetopause show a similar opening width for large stretches of the distant tail without any outward flaring (Maezawa, 1975; Slavin et al., 1983; Fairfield, 1992; Kivelson et al., 1993; Hasegawa, 2002). Such observations have led many to propose that the magnetotail reaches a constant asymptotic opening width, a proposal that is supported by some magnetohydrodynamic (MHD) modeling of the magnetotail (Borovsky, 2012). Conic sections mathematically cannot reproduce this behavior. This could make such models less suitable for use in the tail, which can be a problem when incorporating observations from lunar satellites, for example.

Another popular method is to use a piecewise approach, specifying a separate functional form for different areas of the magnetopause. Frequently these regions are the dayside and the nightside, generally using a conic section for the dayside and an inverse trigonometric function (Petrinec & Russell, 1993, 1996), another conic section (Kuznetsov & Suvorova, 1998), or a cylinder (Kawano et al., 1999) for the nightside. There are also models that attempt to account for North-South or dawn-dusk asymmetries by describing different magnetic latitude regimes such as the cusp with different functions (Boardsen et al., 2000).

Only recently have databases of magnetopause crossings become large enough for machine learning algorithms to be implemented for magnetopause modeling (A. Dmitriev et al., 2011; Wang et al., 2013). Machine learning algorithms have proven to be very effective tools for uncovering underlying structure in higher-dimensional data, and magnetopause models constructed using machine learning algorithms can be very accurate. However, they do have limitations, such as strong dependence on the assumptions fed into the model as well as dependence on the criterion used to evaluate the performance of the model after each training cycle (known as the loss function). Perhaps the biggest barrier is that the model has no set functional form, which means that users must have the database in hand and run the model for a desired set of driving conditions in order to get a magnetopause surface. This drastically limits the utility of these models.

One limitation shared by all current empirical magnetopause models that despite all their complexity, the functional forms they use are still ad-hoc. In this paper a new functional form that has the goal of describing the physical world through known and observed physics is presented. This functional form is a closed-form analytic function that is easy to use to fit and predict magnetopause crossings on the dayside and nightside, including the asymptotic constant-diameter magnetotail.

## 2 Derivation of the Functional Form

First, some basic physical assumptions are used to build up the overall form of the model and make the mathematics tractable. When constructing a model of any type, the decisions that most affect the representational power of the end product are the assumptions and simplifications made in order to construct it. Taking MHD models of the magnetosphere as an example, the closures for higher order moments of the Vlasov equation, the abbreviated model for ionosphere coupling, and other simplified versions of highly complex processes are deeply important. For empirical magnetopause models the assumptions made when constructing the model function in order to simplify the mathematics into an analytic closed-form solution have similar importance. If one assumes that the magnetopause is a conic section, that is necessarily a simplification of the physics involved made so the problem is tractable. Here a similar approach is taken, except the initial assumptions are not made about the functional form itself but about a simplified model of solar wind flow diverting around the magnetosphere. By moving one level deeper, so to speak, the aim is to construct a model with greater representational power than current empirical models that can still be packaged as an analytic function, a form factor that has proven to be of great utility to the magnetospheric modeling community (Howe & Binsack, 1972; Roelof & Sibeck, 1993; Petrinec & Russell, 1996; Shue et al., 1998; Chao et al., 2002; Lin et al., 2010; Lu et al., 2011). By deriving the function on this basis, subsequent modifications to the model can also be made by examining the assumptions, modifying them, and observing the resulting change, instead of simply substituting one ad-hoc functional form for another.

To that end, the problem is considered in two dimensions so that it is tractable and so that an analytic closed-form solution can be found. In practice the model can be rotated about the Sun-Earth line to generate a three dimensional model surface. The x axis is defined as the solar wind flow direction, which is considered to be parallel to itself everywhere as it strikes the magnetopause surface and is diverted. This is a simplification because upstream of the magnetopause there exists a standing bow shock. When the solar wind flow passes through this shock, it senses the downstream magnetopause and starts to divert around it, therefore the flow is not truly one dimensional when it encounters the magnetopause surface. The function that is eventually derived has two control parameters: the subsolar point distance from Earth and the asymptotic magnetotail width, both of which are expected to vary with solar wind conditions. Based on how they are expected to vary with solar wind conditions, functional forms for these control parameters are motivated. This yields the full form of the magnetopause model.

### 2.1 Derivation of the Tractrix Functional Form

To derive the functional form for the magnetopause shape, assume that the magnetopause is a constant-pressure surface defined by pressure balance between the Earth's magnetosphere and solar wind pressure sources. The assumption that the magnetopause is a constant pressure surface is of course not strictly true in reality. However, the assumption is justified for several reasons. In steady state MHD simulations pressure instabilities between regions of the magnetopause surface are equalized fairly efficiently, resulting in a magnetopause that is close enough to a constant pressure surface for this assumption to be justified. For non-steady-state solar wind conditions, it could certainly be the case that sharp gradients or shears in solar wind conditions could occur so quickly that the instability cannot be equalized fast enough for the assumption to hold. However, since the tractrix model is explicitly a static model that does not apply to situations where the time history of the solar wind flow is important, there is sufficient justification for the assumption to be made to motivate a model. To simplify, consider the solar wind ram pressure (also known as dynamic pressure) to be the only solar wind pressure source impinging on the magnetosphere. This is a good assumption, as in the entirety of the currently available one-minute-averaged solar wind conditions available in the OMNI database [1981-2019] dynamic pressure is generally two orders of magnitude larger than magnetic or proton thermal pressures (see Figure 1). Thermal pressure does become more important, however, as tail flare angle gets closer and closer to zero as one moves down the tail (Collier et al., 1998).

The geometry referred to in the subsequent derivation is shown in Figure 2. Consider a tangent line (shown in grey) to the equal pressure surface (shown in teal), and call the angle this tangent line forms with the horizontal $\theta$. If $v(x)$ is the solar wind flow velocity at some x position, the component of the solar wind flow velocity normal to the magnetopause surface at that x position is therefore $v(x)sin\theta$, a purely geometrical argument relying on the assumption that the flow is everywhere parallel to the horizon-

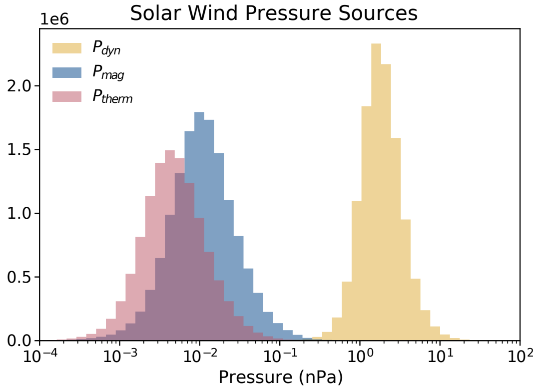


**Figure 1.** Histogram of solar wind pressure sources in the past 38 years [1981-2019] of one minute averaged solar wind data available in the OMNI database. Ram pressure (yellow), also known as dynamic pressure, is generally two orders of magnitude greater than magnetic (blue) or proton thermal (pink) pressure for the entirety of the times considered.

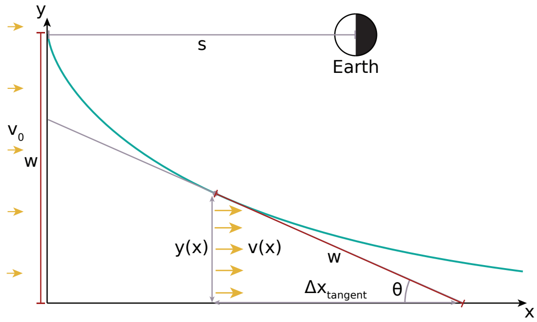


**Figure 2.** The tractrix model in two dimensions, with the x axis starting at the nose of the magnetopause. The magnetopause surface is shown in teal, and the tangent line is shown in grey. Via the relation given in Equation 6, the distance from the tangent point to the x axis is a constant value along the surface $w$ (shown in red), equal to the height the curve reaches above the x axis at 0 (shown also in red). The solar wind flow direction is shown in gold. The distance from the subsolar point to the Earth $s$ is shown at top, and is used to convert from the coordinates shown in the figure to GSE coordinates. Geometrically, the solar wind flow velocity normal to the magnetopause surface is $v(x)sin\theta$.

tal. The ram pressure on the magnetopause surface can therefore be written as $K\rho_m v(x)^2 sin^2\theta$. In this case, $\rho_m$ is the solar wind mass density, and $K$ is the momentum transfer coefficient ($K < 1$ for fluid-like momentum transfer, $K = 1$ for inelastic momentum transfer, and $K = 2$ for elastic momentum transfer (Fairfield, 1971)). Note that here $\theta$ denotes the angle the tangent line to the surface makes with the horizontal, not a polar coordinate. In reality this ram pressure is balanced through the magnetosheath to the magnetopause, and in the process is transferred into other pressure sources.

Calling the assumed constant surface pressure $P_{surf}$ allows the prior argument to be represented as the equation

$$K\rho_m v(x)^2 \sin^2\theta = P_{surf} \tag{1}$$

Then assume that the flow is incompressible, an approach which has ample precedent (Roberts et al., 1991; Goldstein & Roberts, 1999). At the subsolar point, the flow velocity is equal to the upstream flow velocity $v_0$ and $\theta = \pi/2$ (the flow is entirely perpendicular). The ram pressure applied to this point on the equal pressure surface is therefore

$$K\rho_m v_0^2 = P_{surf} \tag{2}$$

which can be combined with Equation 1 to obtain

$$K\rho_m v_0^2 = K\rho_m v(x)^2 \sin^2\theta \tag{3}$$

which can be rearranged to yield the relation

$$\sin\theta = \frac{v_0}{v(x)} \tag{4}$$

Since $\sin\theta < 1$, Equation 4 implies that the flow speed increases around the flanks of the magnetopause, which is routinely observed by spacecraft (Walsh et al., 2012; Dimmock & Nykyri, 2013) and is a feature of gas dynamic calculations (Spreiter & Alksne, 1969). It is important to note that gas dynamic models predict no solar wind flow at the subsolar point (referred to as the stagnation point) with the plasma pressure being entirely thermal at this point. Since in this analysis the fact that dynamic pressure is balanced from the solar wind into the magnetosheath and then to the magnetopause is neglected, there exists some flow at the subsolar point for this model.

The particle flux per unit length that enters from the left of Figure 2 (Sun direction) at $x = 0$ is $n * v_0 * w$ where $n$ is the number density of the solar wind plasma. Since the flow is assumed to be incompressible, $n$ is constant and particle flux conservation yields the relation (Also known as Bernoulli's Principle)

$$n * v_0 * w = n * v(x) * y(x) \Rightarrow \frac{v_0}{v(x)} = \frac{y(x)}{w} \tag{5}$$

which, combined with Equation 4, yields the relation

$$\sin\theta = \frac{y(x)}{w} \tag{6}$$

Assuming the flow is incompressible results in some tension with assumptions made previously, namely that the "diverting" component of the flow velocity is ignored which requires some compression at the magnetopause boundary. Referring again to Figure 2, Equation 6 means that the distance from the tangent point on the curve to the point where the tangent line intersects the x axis is always a constant value $w$. This is the definition of a curve known as a tractrix or "hundkurve" (Lawrence, 2014).

The differential equation describing this curve can be obtained geometrically through evaluating the reciprocal of the slope of the tangent line. Considering the triangle formed in the bottom right of Figure 2, the reciprocal of the slope can be calculated by noting that $\Delta x_{tangent} = \sqrt{w^2 - y(x)^2}$ and that $\Delta y = y(x)$:

$$\frac{\Delta x_{tangent}}{\Delta y} = -\frac{\sqrt{w^2 - y(x)^2}}{y(x)} = \frac{dx}{dy} \tag{7}$$

which has solution

$$x = w \ln \frac{w + \sqrt{w^2 - y(x)^2}}{y(x)} - \sqrt{w^2 - y(x)^2} \tag{8}$$

which is the formula for a tractrix in Cartesian coordinates.

To recast Equation 8 into geophysically useful coordinates, it can be transformed into aberrated Geocentric Solar Ecliptic (GSE) coordinates (with negative $x_{GSE}$ directed downtail and centered on Earth rather than defined by the subsolar point) with the coordinate transform

$$x = -(x_{GSE} - s) \tag{9}$$

$$y(x) = w - y_{GSE} \tag{10}$$

where $s$ is the upstream distance of the magnetopause subsolar point and the y axis is oriented so that the subsolar point is located on the x axis instead of above it. Since this model is axisymmetric, one could also use $x_{GSM}$ (which is the same as $x_{GSE}$) and $y_{GSM}$ (which is still in the $y_{GSE} - z_{GSE}$ plane). Thus the overall form of the tractrix magnetopause surface is given by a more convenient form of Equation 8:

$$x_{GSE} = s - w \ln \frac{w + \sqrt{w^2 - (w - y_{GSE})^2}}{w - y_{GSE}} + \sqrt{w^2 - (w - y_{GSE})^2} \tag{11}$$

### 2.2 Width and Standoff Functional Form

The two parameters $s$ and $w$ in Equation 11 control the model's response to solar wind conditions. $s$, the distance of the subsolar point from the Earth, is a common parameter in empirical models (Petrinec & Russell, 1996; Shue et al., 1997, 1998, e.g.). $w$, the asymptotic tail width, is the asymptotic width to which the magnetopause flares outward along the tail. Many other empirical models, such as Shue et al. (1998), predict that the tail continues to flare outward infinitely with distance from the Earth, or close for northward IMF. Some work has found that the magnetotail continues to flare in one plane but flattens out in another (Sibeck & Lin, 2014). On the basis that both parameters represent a pressure standoff between solar wind pressure sources and the Earth's magnetosphere, the same overall functional form shall be used for both in this study.

For this study, $s$ and $w$ are chosen to be controlled by solar wind dynamic pressure and the sine rectifier of the interplanetary magnetic field (IMF) only. As previously shown, the dominant pressure term in the solar wind is the dynamic pressure, therefore it is the only solar wind pressure term included here. The sine rectifier of the IMF is a measure of the magnitude of the shear between the IMF and the Earth's magnetic field that takes into account the total magnitude of the draped IMF, unlike using $B_z$ alone. Mathematically, the sine rectifier of the IMF is given by $B_S = B_{IMF} \sin^2(\theta_C/2)$ where $B_{IMF}$ is the IMF magnetic field magnitude and $\theta_C = \tan^{-1}(\frac{B_y}{B_z})$ is the solar wind clock angle. For instance, a clock angle $\theta_C = \pi$ (i.e. $B_y = 0, B_z < 0$) results in $B_S = B_{IMF}$. This is because the IMF is entirely oppositely directed to Earth's magnetic field. A clock angle $\theta_C = \pi/2$ (i.e. $B_y > 0, B_z = 0$) results in $B_S = 0.5B_{IMF}$. This is because the draped IMF still has some shear against Earth's magnetic field, despite the fact that $B_z = 0$. The sine rectifier of the IMF is frequently used in solar wind-magnetosphere coupling functions, and has been shown to be more highly correlated with many different indicators of magnetospheric activity resulting from magnetic reconnection than $B_z$ alone (Perreault & Akasofu, 1978; Kan & Lee, 1979; Vasyliunas et al., 1982; Newell et al., 2007). Magnetic reconnection controls magnetopause current systems and therefore the position and shape of the boundary (Sibeck et al., 1991; Roelof & Sibeck, 1993; Borovsky, 2013), thus $B_S$ should be included as a control parameter in the functional form for $s$ and $w$. Other parameters have been shown to have an effect on the shape of Earth's magnetopause, but will not be considered in this study for simplicity. Namely, Earth's dipole tilt angle has been shown to drive oblateness in the magnetopause shape (Liu et al., 2012;

Wang et al., 2013), but since the tractrix model is axisymmetric, it cannot reproduce oblateness and therefore dipole tilt dependence is not included in the model.

If one were to assume that Earth's magnetic field intensity falls off radially as some power law, i.e.

$$B(r) = B_0(\frac{R_E}{r})^n \tag{12}$$

for some $n > 1$ where $B_0$ is the equatorial magnetic field at the Earth's surface, balancing the solar wind dynamic pressure against the magnetic pressure of Earth's magnetic field at a standoff distance $r_s$ yields

$$P_{dyn} = \frac{B_0^2(\frac{R_E}{r_s})^{2n}}{2\mu_0} \tag{13}$$

which can be rearranged to yield the relation

$$r_s \propto P_{dyn}^{-1/2n} \tag{14}$$

for some $n > 1$. Therefore one should expect that standoff distance varies with dynamic pressure as a power law, an approach that is used in many empirical models (Shue et al., 1998; Chao et al., 2002; Lin et al., 2010).

Magnetic reconnection at the nose drives currents that erode Earth's magnetic field and allows the standoff position to move inward. Using $B_S$ as an indicator of the reconnection rate at the nose, one expects a generally smaller standoff position with increasing $B_S$. Therefore a simple linear dependence is used for $B_S$, yielding the overall forms of the standoff and width functions

$$s = (s_0 + s_1 B_S)P_{dyn}^{-1/s_2} \tag{15}$$

$$w = (w_0 + w_1 B_S)P_{dyn}^{-1/w_2} \tag{16}$$

Each of these functions has three tuning parameters given by $s_0$ through $s_2$ and $w_0$ through $w_2$, respectively. Thus the overall form of the tractrix model has six tuning parameters and accepts only two solar wind parameters ($P_{dyn}$ and $B_S$) as input, making it one of the simplest empirical models in terms of functional complexity.

It is important to note that this account of magnetic reconnection has some limitations. Magnetic reconnection-driven magnetopause erosion has been observed to saturate for extreme values of $B_S$ or shear magnetic field (Siscoe, 2002; Yang, 2003; Ridley, 2005; Shepherd, 2007). A simple linear relation as used in this study cannot reproduce this behavior. This simple linear relation also assumes that the magnetopause position varies instantaneously with IMF variations. For more northward-oriented IMF, magnetic reconnection is anticipated to still occur however at a lower rate due to reconnection poleward of the cusp. Past observations and modeling have shown that magnetic reconnection does indeed occur at these high latitudes for northward IMF, but offered inconclusive evidence for a rearrangement of the magnetopause location and shape (Avanov et al., 2001; Le et al., 2001). The relation presented in this study depends only on shear between the IMF and Earth's magnetic field on the equatorward side of the cusp, therefore it cannot reproduce any effects caused by cusp reconnection (if they exist).

## 3 Model Optimization

In order to constrain the tractrix model's tuning parameters, it is fit to a combined dataset of over 13,000 magnetopause crossings. These crossings cover a broad spatial extent and include an extensive dataset used in Wang et al. (2013). One of the major strengths of the tractrix model is its potential ability to more accurately model tail behavior compared to other popular empirical models. It follows that to best optimize this model to a dataset, it is desirable for that dataset to include magnetopause crossings at points downtail where the tail has reached or is close to reaching its asymptotic width. However, magnetopause crossings in the tail are relatively rare when compared to crossings close to the Earth and in the nose due to the limited amount of spacecraft that have flown through the distant tail. This scarcity has driven some researchers to fit empirical models to magnetohydrodynamic (MHD) simulations (Lu et al., 2011) rather than trying to fit them to small spacecraft datasets. Here a sufficient database with over 900 magnetopause crossings obtained with the ARTEMIS-P1 (formerly THEMIS-B) spacecraft in the deep magnetotail was compiled in order to model the tail. Finally, a Markov Chain Monte Carlo (MCMC) method is used to find the optimal parameters for the tractrix model for these combined datasets.

### 3.1 ARTEMIS Dataset

The Acceleration, Reconnection, Turbulence and Electrodynamics of the Moon's Interaction with the Sun (ARTEMIS) constellation is a continuation of the Time History of Events and Macroscale Interactions During Substorms (THEMIS) mission consisting of two spacecraft (formerly THEMIS-B and THEMIS-C) inserted into an elliptical lunar orbit (Angelopoulos, 2008). This orbit with radius of about $60R_E$ provides an excellent environment to measure magnetopause crossings far enough downtail to constrain the tail width function.

Magnetopause crossings were selected according to the following paradigm: Plasma and field measurements were surveyed through visual inspection for clear transitions from narrow to broad distributions in the ion spectra (magnetosheath to magnetosphere) and rotations in the magnetic field. Importantly, one spacecraft orbital pass was allowed to contain multiple magnetopause crossings, corresponding to the boundary sweeping over the spacecraft repeatedly. This is done for several reasons, the first being that in MHD simulations the tail has been observed to move rapidly due to shears in the solar wind conditions (Borovsky, 2012). Including multiple crossings due to this "flapping" can help to mitigate or "average out" these effects. This is also done to preserve continuity with the Wang et al. (2013) data set, detailed in Section 3.2. While this does weight the database slightly toward passes with multiple crossings, studies using this paradigm have found that it does not prevent the model from being a good fit for observations (Wang et al., 2013). An example magnetopause crossing is shown in Figure 3 to demonstrate the selection criteria.

Using this paradigm, 901 magnetopause crossings were identified in ARTEMIS-P1 data from August $8^{th}$, 2011 to December $5^{th}$, 2017. In order to take into account the realignment of the magnetotail in response to the direction of the solar wind, the positions of these crossings were transformed from GSE coordinates into solar wind aberrated GSE coordinates. As implemented by A. V. Dmitriev (2003), this coordinate system attempts to correct for aberrations in the central position of the magnetotail due to the motion of the Earth around the sun through the solar wind as well as the direction of the solar wind flow relative to the Earth via a series of two rotations. These rotations attempt to align the x axis of the GSE coordinate system with the solar wind flow direction instead of the Sun-Earth line. The angles $\alpha$ and $\beta$ are defined to be the rotation angles of the x axis in the y and the z flow direction, respectively. With all position variables and velocities given in GSE coordinates, the angles and combined rotation matrix are given by

$$\alpha = \tan^{-1}(\frac{v_y + 30\frac{km}{s}}{|v_x|}) \tag{17}$$

$$\beta = \tan^{-1}(\frac{v_z}{\sqrt{v_x^2 + (v_y + 30\frac{km}{s})^2}}) \tag{18}$$

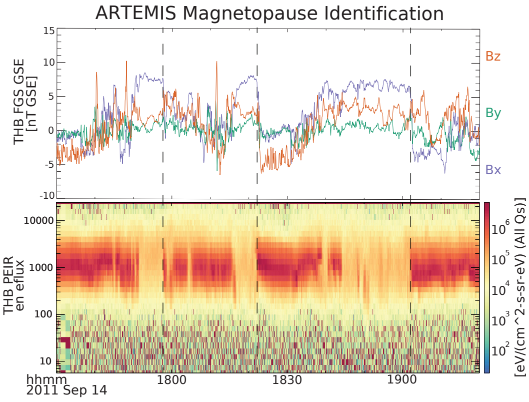


**Figure 3.** ARTEMIS-P1 magnetic field and ion energy spectrum data displaying several magnetopause crossings that occurred on September $14^{th}$ 2011 from 1750 to 1900 UTC. These data show three clear magnetic field rotations from a strongly GSE X-aligned field to a turbulent field, which occurs at the same time as transitions from a low energy plasma energy spectra to a region with highly thermalized plasma. This is consistent with a transition from the magnetosheath into the southern lobe. The positions of the crossings are chosen to be the black vertical dotted lines at 1750, 1820, and 1900 UTC.

$$\begin{bmatrix} x_{GSE,ab} \\ y_{GSE,ab} \\ z_{GSE,ab} \end{bmatrix} = \begin{bmatrix} \cos\alpha\cos\beta & -\sin\alpha\cos\beta & -\sin\beta \\ \sin\alpha & \cos\alpha & 0 \\ \cos\alpha\sin\beta & -\sin\alpha\sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} x_{GSE} \\ y_{GSE} \\ z_{GSE} \end{bmatrix} \tag{19}$$

Note the presence of $30\frac{km}{s}$ in Equations 17 and 18, which is included to re-insert the orbital velocity of the Earth through the solar wind that is subtracted off in the OMNI dataset.

Using data from a number of missions collected in the OMNI database (King, 2005) each crossing was associated with solar wind velocity and dynamic pressure, as well as IMF sine rectifier $B_S$. Any crossing missing one or more of these parameters was removed from the dataset. Finally, the GSE positions of the remaining crossings were shifted into this aberrated coordinate system. Note that the solar wind parameters were not shifted to the location of each crossing downtail. This is done mainly for the reason that the solar wind data for the tail crossings should be treated in the same way as solar wind conditions for the near-Earth crossing detailed in the subsequent section. Furthermore, the goal of this model is to generate a magnetopause surface from given instantaneous solar wind conditions, so training the model on solar wind data propagated to each crossing would not be in line with the end use of the model. In the end, cuts for incomplete OMNI data resulted in a magnetotail dataset of 649 magnetopause crossings, the spatial distribution of which is can be seen in Figure 4.

When considering Figure 4, it is important to note that the apparent elliptical structure of magnetopause observations on the right side of the figure is not due to underlying structure of the magnetopause, but due to the fact that the ARTEMIS spacecraft are in orbit around the moon and are thus constrained to a roughly $10R_E$ wide band around lunar orbit. This orbital constraint induces a sampling bias, where the farther downtail the observation is made, the smaller the magnetotail widths that can be sampled. The reverse is true closer to the Earth, where only large tail widths can be sampled. This effect is clearest in the "top-down" ($x_{GSE}$-$y_{GSE}$ plane) plot in Figure 4, where the observations are spread about a roughly circular path with radius $\approx 60R_E$, which is the average lunar orbit distance. Phrased another way, it is not the case that the magnetopause is never observed in the regions shown in the bottom left and top right of the right plot

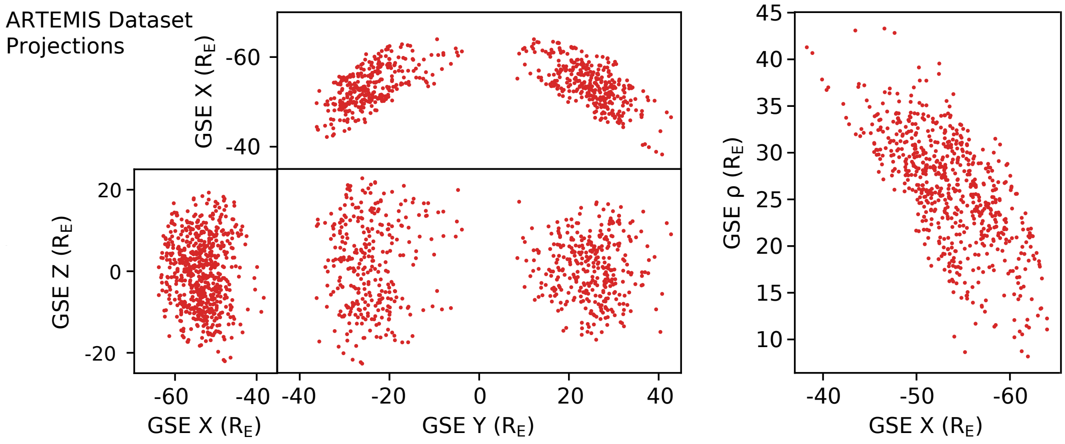


**Figure 4.** Projections of the spatial distribution of the ARTEMIS-obtained dataset. On the left, clockwise from top are GSE Y-X, Y-Z, and X-Z plots which provide a "top-down", "nose-on", and "left-facing" view of the dataset, respectively. On the right is the same dataset folded about the GSE X axis. The Y axis is labeled as "$\rho$" with reference to a cylindrical coordinate system oriented along the GSE x-axis ($\rho = \sqrt{y^2 + z^2}$). Note a strong sampling bias in GSE X induced by the lunar orbit: in the very far tail only small magnetotail widths are sampled, whereas in the "near" tail only very large magnetotail widths are sampled.

on Figure 4 because it doesn't exist in those regions, it is instead the case that there is never a spacecraft present in those regions to sample the magnetopause location. Similar effects can be observed in other magnetopause crossing datasets acquired using the ARTEMIS spacecraft (Mieth et al., 2018; Gencturk Akay et al., 2019).

### 3.2 Wang et al. Dataset

In order to constrain the tractrix model in spatial regions close to the Earth, a very large collection of magnetopause crossings previously used to train and validate a machine learning model of the near-Earth magnetopause (Wang et al., 2013) is utilized. The original dataset included 15,089 crossings assembled from 23 different satellites between November 1966 and November 2008, with associated solar wind conditions. Multiple magnetopause crossings per spacecraft pass were counted as separate crossings, as with the ARTEMIS dataset. Since the crossings in this dataset were obtained over a long time period, some of the crossings have higher quality solar wind data than others. In order to ensure that only crossings with the highest-quality solar wind data are used, all crossings obtained before 1981 are cut out, which is when minute-averaged solar wind conditions became available in the OMNI database. All crossings obtained after this year use the minute-averaged solar wind conditions. Performing this cut yields a reduced dataset of 12,522 magnetopause crossings obtained with eight spacecraft missions, the specific distribution of which is detailed in Table 1. Over half of these crossings come from the THEMIS constellation. The spatial distribution of this dataset is shown in Figure 5.

### 3.3 MCMC Fit Procedure

In order to optimize the tractrix model's performance on this combined dataset, a Bayesian framework was used to construct the posterior probability distribution of the model parameters given the combined dataset, then a Markov Chain Monte Carlo (MCMC) method was used to estimate this posterior distribution and determine which parame-

| Satellite | Date Range | No. |
|---|---|---|
| AMTE CCE | 1984 August - 1988 December | 29 |
| AMTE IRM | 1984 August - 1986 January | 36 |
| Cluster 1 | 2001 January - 2004 December | 2,556 |
| Geotail | 1992 October - 1997 June | 1,352 |
| Interball 1 | 1995 August - 1998 December | 1,771 |
| Magion 4 | 1996 March - 1997 August | 119 |
| Prognoz 8 | 1981 January - 1981 September | 71 |
| Prognoz 10 | 1985 May - 1985 November | 31 |
| THEMIS A | 2007 June - 2008 November | 1,183 |
| THEMIS B | 2007 June - 2008 November | 1,693 |
| THEMIS C | 2007 June - 2008 November | 1,984 |
| THEMIS D | 2007 June - 2008 November | 877 |
| THEMIS E | 2007 June - 2008 November | 820 |
| Total | 1981 January - 2008 November | 12,522 |

**Table 1.** Distribution of magnetopause crossings obtained with each spacecraft in the Wang et al. (2013) dataset, and the time ranges of those magnetopause crossings. Note that more than half come from the THEMIS spacecraft.

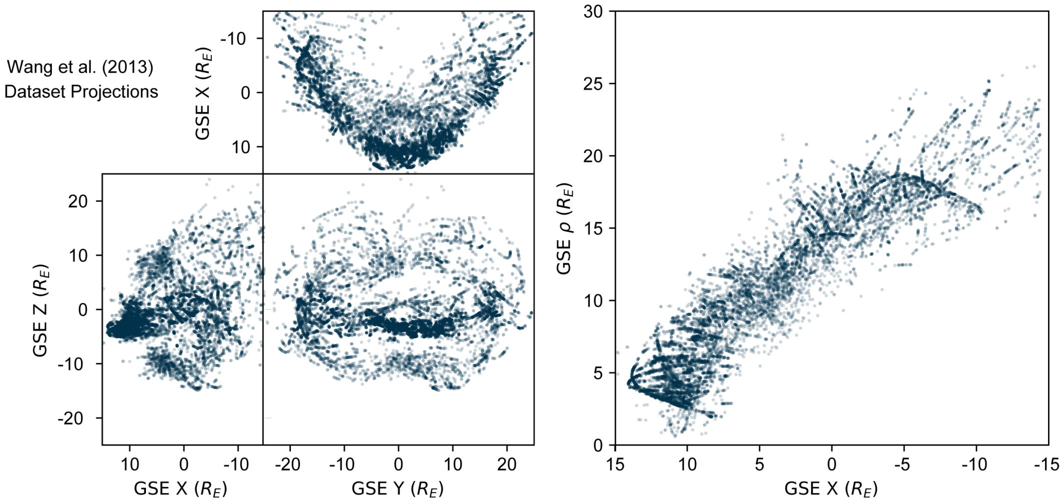


**Figure 5.** Projections of the spatial distribution of the (Wang et al., 2013) dataset. On the left, clockwise from top are GSE Y-X, Y-Z, and X-Z plots which provide a “top-down”, “nose-on”, and “left-facing” view of the dataset, respectively. On the right is the same dataset folded about the GSE X axis. The Y axis is labeled as “$\rho$” with reference to a cylindrical coordinate system oriented along the GSE x-axis ($\rho = \sqrt{y^2 + z^2}$). Note the extreme density of crossings near the nose of the magnetopause.

ters maximize it, thereby obtaining a maximum likelihood estimate (MLE) of the optimal parameters for this dataset. This framework was chosen for several reasons. First, this framework is statistically valid for almost any model function, unlike many other optimization frameworks such as nonlinear least squares fitting. It can also deal with highly correlated parameters well, as it does not involve calculation of the derivative of the covariance matrix. It also gives a sense of the correlation and uncertainty in each parameter automatically via analysis of the posterior distribution.

The "posterior distribution" in Bayesian statistics is the conditional probability of event $A$ occurring given event $B$, written as $P(A|B)$. For model optimization, consider event $A$ to be the model parameters to be in some state, and event $B$ to be the observed data the model is being fit to. The parameters for which this probability distribution is maximized are therefore the parameters most likely responsible for producing the observed data and can be thought of as "best fit" parameters. Denote the parameters of the tractrix model as a vector $\boldsymbol{\theta}$ consisting of $s_0, s_1, s_2$ and $w_0, w_1, w_2$, the $n$ observed magnetopause crossing positions $x_n$ as $\boldsymbol{x}$, and the solar wind conditions $B_{S,n}$ and $P_{dyn,n}$ associated with each crossing as vectors $\boldsymbol{B_S}$ and $\boldsymbol{P_{dyn}}$. Via Bayes's theorem, this distribution is given by

$$P(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{B_S}, \boldsymbol{P_{dyn}}) = \frac{P(\boldsymbol{\theta})P(\boldsymbol{x}, \boldsymbol{B_S}, \boldsymbol{P_{dyn}}|\boldsymbol{\theta})}{P(\boldsymbol{x}, \boldsymbol{B_S}, \boldsymbol{P_{dyn}})} \tag{20}$$

$P(\boldsymbol{\theta})$, the probability of the parameters being in a particular state $\boldsymbol{\theta}$, is known as the prior, and represents what is believed to be true about the parameters. For parameters representing physical quantities, this would encode physical constraints on the quantities involved (e.g. mass must always be positive, velocities must not exceed $c$, frequencies are expected to be observed in a Gaussian distribution around some natural frequency, etc.). The more information one has about the parameters involved, the more complicated a prior one could construct. However, the shape of the prior heavily influences the shape of the posterior, so in general it is good to be conservative with one's choice of prior. In the case of this fit procedure, for $s_2, w_2 < 0$, note that $-\frac{1}{s_2}, -\frac{1}{w_2} > 1$, and therefore that standoff position would increase for increasing dynamic pressure, which is known to be nonphysical (Spreiter & Alksne, 1969). The prior distribution is therefore given by $P(\boldsymbol{\theta}) = 1$ for $s_2 > 0$ and $w_2 > 0$, and $P(\boldsymbol{\theta}) = 0$ everywhere else. This is known as an "uninformed uniform prior", and assumes within the stated bounds all parameters are equally likely (without considering the data).

$P(\boldsymbol{x}, \boldsymbol{B_S}, \boldsymbol{P_{dyn}})$, the probability of the evidence, is the probability that the specific data was obtained. Since the dataset does not change over the course of this analysis, $P(\boldsymbol{x}, \boldsymbol{B_S}, \boldsymbol{P_{dyn}})$ is always constant. $P(\boldsymbol{x}, \boldsymbol{B_S}, \boldsymbol{P_{dyn}}|\boldsymbol{\theta})$, the likelihood, is the probability the data was obtained given a set of parameters $\boldsymbol{\theta}$. This is akin to a "goodness-of-fit" statistic in traditional fitting algorithms. Define a function $d_{trac}(\boldsymbol{x_n}, \boldsymbol{\theta}, B_{S,n}, P_{dyn,n})$ representing the distance between a given magnetopause crossing $x_n$ with solar wind conditions $B_{S,n}$ and $P_{dyn,n}$ and the tractrix surface defined by the model parameters $\boldsymbol{\theta}$. For a perfect prediction, this distance would be zero, i.e. the crossing would be exactly on the magnetopause surface. The ideal fitting state would be this quantity being zero for every crossing in the dataset, so the likelihood is constructed to be a multidimensional Gaussian distribution centered on this hypothetical point:

$$P(\boldsymbol{x}|\boldsymbol{\theta}, B_S, P_{dyn}) \propto \sigma^2 exp(-\frac{1}{2}\sum_n \frac{d_{trac}(\boldsymbol{x_n}, \boldsymbol{\theta}, B_{S,n}, P_{dyn,n})^2}{\sigma^2}) \tag{21}$$

This distribution is maximized where $d_{trac}$ is minimized in parameter space. Note also the parameter $\sigma$, which is a constant uncertainty in position for all crossings. This parameter, the width of the higher dimensional Gaussian, will also be estimated as part of the MCMC procedure. It can be used as a measure on the general uncertainty inherent to the model.

Via taking the natural logarithm of Equation 20 (using the likelihood given by Equation 21) the expression for the posterior can be found without needing to worry about the constant normalization factors such as $P(\boldsymbol{x}, \boldsymbol{B_S}, \boldsymbol{P_{dyn}})$:

$$\ln P(\boldsymbol{\theta}, B_S, P_{dyn}|\boldsymbol{x}) = -\frac{1}{2}\sum_n \frac{d_{trac}(\boldsymbol{x_n}, \boldsymbol{\theta}, B_{S,n}, P_{dyn,n})^2}{\sigma^2} + \ln(\sigma^2) + \ln(P(\boldsymbol{x})) + const. \quad (22)$$

Due to the natural logarithm, the normalization constants fall out of the expression and do not affect where the distribution is maximized. Furthermore, the term $\ln \sigma^2$ that results from taking the natural logarithm serves to "punish" the posterior if an arbitrarily large uncertainty is assumed.

Instead of directly sampling Equation 22, which would be computationally expensive for seven parameters and the very large dataset, an affine-invariant ensemble sampler for Markov chain Monte Carlo (MCMC) is used to estimate the posterior distribution (Goodman & Weare, 2010). This is a version of a Metropolis-Hastings algorithm. Qualitatively, a Metropolis-Hastings algorithm involves initializing a large number of "walkers" in parameter space that iteratively random walk through the parameter space and sample the posterior at each location in parameter space they are located at for each step in the chain. They then randomly accept or reject their last move in the chain, with weighting based on the relative values of the posterior at each location. Since the equilibrium state of this process is the true probability density of the posterior, with enough iterations an estimate of the posterior distribution is obtained.

The combined dataset is randomly split into a training set containing 80% of the crossings, and a validation dataset containing the remaining 20% of the crossings. Initial positions for the walkers are obtained by running the algorithm on training data that has been randomly downsampled to 10% of its original size. Since the computation time of each step in the chain is directly related to the size of the dataset, this reduces computation time by a factor of 10 and allows the larger chain to converge more quickly. The python package emcee was used to wrap the setup of the walkers and chain (Foreman-Mackey et al., 2013). The full chain is run for 100,000 iterations, discarding the first 398 steps in the chain and thinning by a factor of 45 to account for autocorrelation ($\tau_{max} = 199, \tau_{min} = 90$). A parameter called the "autocorrelation time" $\tau$ can be calculated for each parameter, and serves as a measure of how many iterations it takes for walkers to "forget" their previous positions in each dimension of the parameter space. $\tau_{max}$ and $\tau_{min}$ are the maximum and minimum such times across the dimensions of the parameter space. Qualitatively, autocorrelation is the tendency of walkers to take some time to move to a new location in parameter space and "forget" their old one. By discarding a large number of points ($\approx 2\tau_{max}$) at the start of the chain, the walkers are given time to both "forget" their initial position and distribute themselves evenly throughout parameter space. By discarding all but one position of the chain every $n$ iterations for some $n \approx 0.5\tau_{min}$, one can ensure that each iteration in the chain can be considered to be independent.

The maximum of this downsampled distribution is taken as initial parameters for the full-dataset chain. The larger chain was iterated 10,000 times, discarding the first 352 steps in the chain and thinning by a factor of 33 ($\tau_{max} = 176, \tau_{min} = 66$). The maximum likelihood estimate of the optimal parameters for the tractrix model on the training dataset is given in Table 2. A representation of the full multidimensional posterior distribution is given as Figure 6. On the diagonal of the figure are the one dimensional probability distributions for each parameter, and the off-diagonal elements are two dimensional histograms representing the joint probability distributions of each pair of parameters. By inspecting the off-diagonal elements of the figure, covariances and correlations between parameters can be identified. The ideal case is for all distributions to

| Parameter | Value | Uncertainty |
|---|---|---|
| $s_0$ | 14.56 | $\pm 0.06$ |
| $s_1$ | -0.0398 | $\pm 0.0200$ |
| $s_2$ | 5.70 | $\pm 0.21$ |
| $w_0$ | 32.34 | $\pm 0.16$ |
| $w_1$ | -0.247 | $\pm 0.042$ |
| $w_2$ | 12.24 | $\pm 1.43$ |
| $\sigma$ | 1.91 | $\pm 0.06$ |

**Table 2.** Optimal parameters for the tractrix model on the training dataset. Quoted uncertainties are the standard deviation of each one dimensional probability distribution.

be one dimensional Gaussian distributions on the diagonal and two dimensional Gaussian distributions off the diagonal.

The posterior distribution is locally Gaussian for all parameters except for $w_1$ and $w_2$, which have a slight double-peak structure. This means that there are two values of these parameters that locally maximize the posterior, with only one being the global maximum. What this implies physically is that there may be two solar wind condition dependencies sampled by this study, with one being dominant. This could be two "modes" that the magnetosphere operates in with one being dominant, or some difference between the dawn and dusk tail magnetopause. This behavior is difficult to identify in the full dataset due to its size, and would be missed with a simple optimization algorithm that would leave the local minimum. This bimodal tail behavior is outside the scope of this project, but is an example of potentially new physics being uncovered via a machine learning algorithm, and will be the subject of future investigation. Since the global maximum has more than twice the likelihood of the other local maximum, the global maximum will be used in subsequent analyses and is quoted in Table 2.

The optimized functions for subsolar point distance and asymptotic tail width are illustrated in the form of contour plots in Figure 7 and as slices through each contour plot in Figure 8. The optimized function for the standoff distance $s$ predicts that the subsolar distance moves inward for increasing $B_S$, which is consistent with shear magnetic field reconnecting in the nose eroding the dayside magnetopause. It is worth noting that the countours shown here appear very different from usual contours of this type due to the use of $B_S$ instead of $B_z$. Specifically, the discontinuity associated with the transition from southward to northward $B_z$ is not present, since $B_S$ is a continuous control parameter for dayside reconnection in this model. The tractrix model predicts that tail width $w$ depends more strongly on the IMF orientation and strength than the subsolar point position does, which is supported by some prior observations and modeling (Maezawa, 1975). Additionally the model incorporates that the effect of $B_S$ is smaller for high dynamic pressure and the effect of dynamic pressure is smaller for large $B_S$, a phenomenon that has been reported previously (Roelof & Sibeck, 1993).

## 4 Comparison with Existing Models

To examine the strengths and weaknesses of this model, its performance is compared against several other empirical models. The models selected are Shue et al. (1998); Petrinec and Russell (1996); Chao et al. (2002); Lin et al. (2010); and Lu et al. (2011). Shue et al. (1998) is the most widely used magnetopause model, and consists of a conic section rotated about the Sun-Earth line with eight tunable parameters. Petrinec and Russell (1996) was one of the first models with dependence on solar wind parameters to focus on the tail, and has performed well in other global studies. Chao et al. (2002) uses

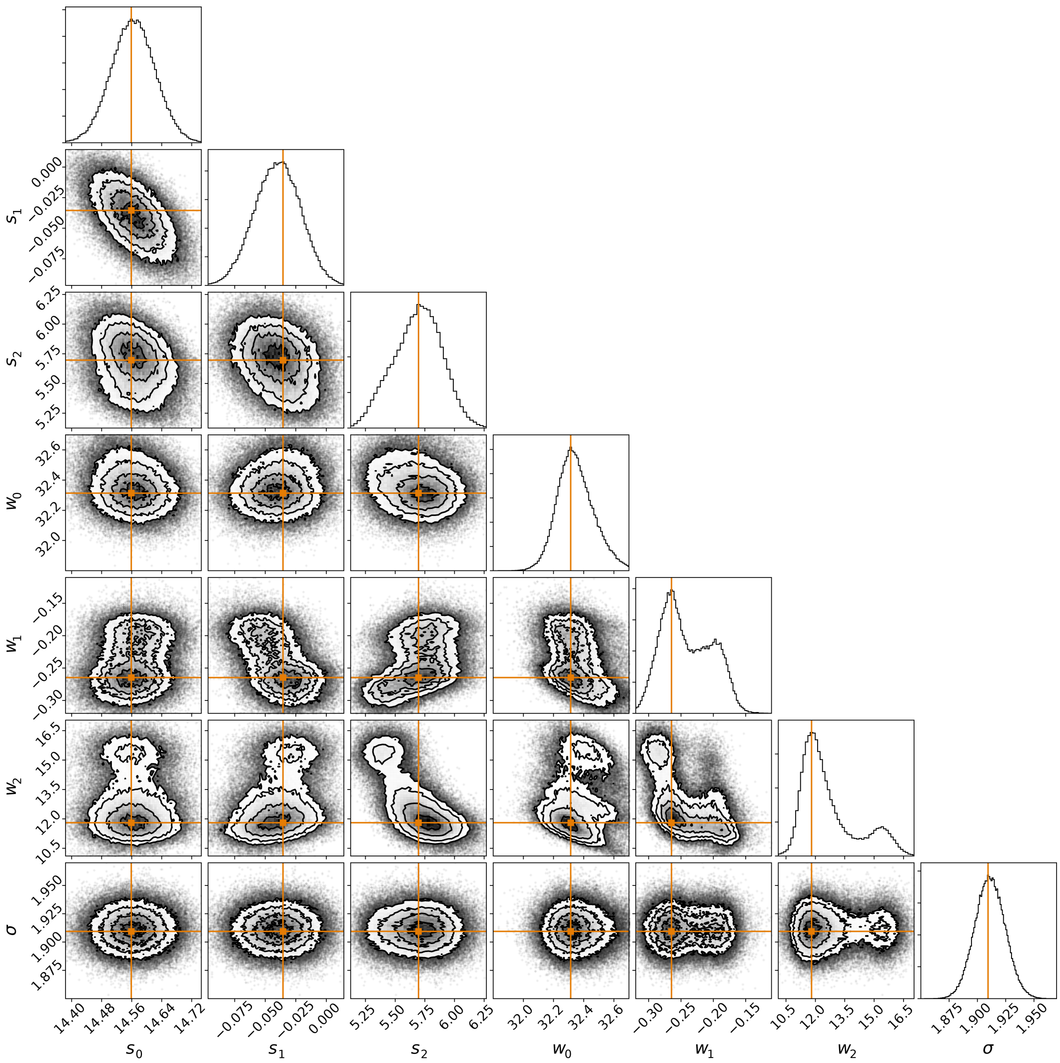


**Figure 6.** A corner plot showing the full seven dimensional probability distribution. Along the diagonal is the one dimensional probability distribution for each parameter, whereas off the diagonal are two dimensional convolved probability distributions for each pair of parameters. The locations of the maximum likelihood parameters are highlighted in orange. Close to the maximum likelihoods, contours of equal probability are estimated, with the bins of the two dimensional histograms plotted overtop. Far from the maximum likelihood, the locations of each walker in the chain at each iteration are plotted. Note that $\sigma$, the higher dimensional Gaussian width of the posterior, is also estimated as part of this procedure, the 1D distribution of which is shown in the bottom right. Created using the python package corner (Foreman-Mackey, 2016).

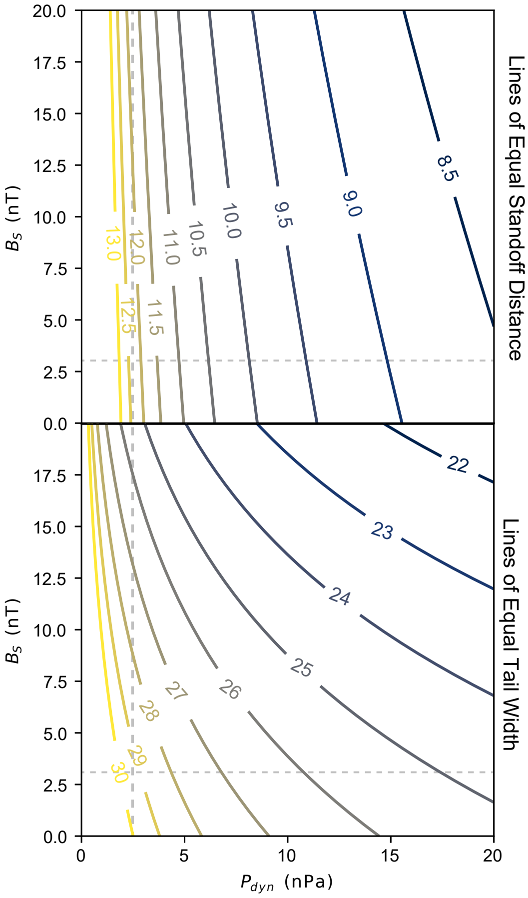


**Figure 7.** Lines of equal subsolar point distance (standoff distance) and asymptotic tail width predicted by the tractrix model with the optimized Equations 15 and 16, with dynamic pressure on the x axis and sine rectifier on the y axis. The grey dotted lines show the average value of $P_{dyn}$ and $B_S$ in the total dataset ($2.40nPa$ and $2.64nT$, respectively).

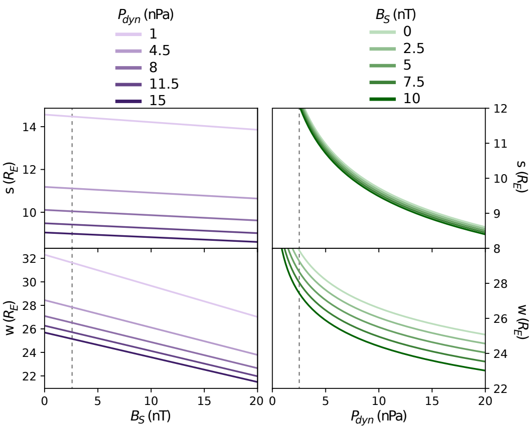


**Figure 8.** Slices through the contours shown in Figure 7 along lines of equal $P_{dyn}$ (left) and $B_S$ (right). The above plots show the subsolar standoff distance $s$, while the bottom plots show tail width $w$. The grey dotted lines show the average value of $P_{dyn}$ and $B_S$ in the total dataset ($2.40nPa$ and $2.64nT$, respectively).

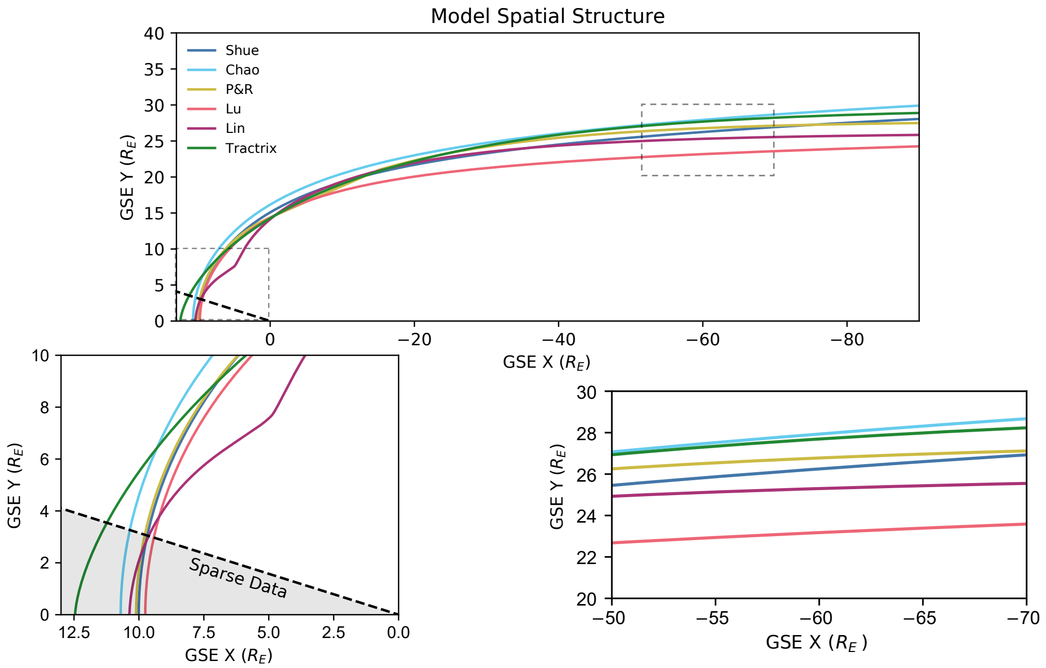


**Figure 9.** Spatial structure of the tractrix model and the five other models compared in this study on the dayside (left) and the far magnetotail (right). The black dotted line on the left is the lowest inclination of nose crossings in the dataset. The input parameters used are average input parameters for the combined dataset ($P_{dyn} = 2.34nPa$, $B_{IMF} = [0.058nT, -0.26nT, -0.049nT]$, $B_S = 2.58nT$, and a dipole tilt angle of 6.3°). For non-axisymmetric models Lu et al. (2011) and Lin et al. (2010), their structure in the meridional plane is shown. The tractrix has a larger subsolar distance than other models because it does not assume the magnetopause is locally circular at the nose.

the same overall form as Shue et al. (1998) with an updated standoff function. It was trained on higher quality solar wind data, and it is the best-performing model that uses the functional form of Shue et al. (1998). Lu et al. (2011) is an empirical model trained on an MHD simulation of the Earth's magnetosphere instead of satellite observations, which was done in part to account for the fact that magnetopause crossing data in the tail is sparse. Lin et al. (2010) is one of the most complicated empirical magnetopause models, with over twenty tunable parameters. It takes into account dawn/dusk and north/south asymmetries, as well as cusp indentations. One notable absence from this list is the machine learning model of Wang et al. (2013), whose dataset is used again in this study. While at the time of publication the authors of Wang et al. (2013) intended to provide the model as a publicly accessible utility, such plans never came to fruition. The Wang et al. (2013) machine learning model is sufficiently involved that reproducing elements of the model without the the original code and environment would not provide a fair comparison. In subsequent sections, the performance of these models and the tractrix model close to the Earth are quantitatively compared using the portion of the validation dataset from the Wang et al. (2013) dataset, and in the tail using the portion from the ARTEMIS dataset.

To qualitatively illustrate the differences between the models considered here, their shape in the meridional plane is plotted on the dayside and in the far tail in Figure 9. On the dayside, the tractrix predicts a much larger standoff position than any other model. This is due to the fact that the tractrix magnetopause is not locally circular at the nose as models based on conic sections are, but is instead "blunt" at the nose. This is reflected in gas dynamic models (Spreiter et al., 1966) and machine learning models (Wang et al., 2013). Observations of the magnetopause made very close to the nose are actually quite rare, most crossings that are called "nose crossings" actually occur some distance from the Sun-Earth line. If one considers the angle between the vector to a given crossing and the Sun-Earth line, there are very few crossings in this dataset that are observed within 17.5 degrees of the Sun-Earth line. It is difficult to locate this cone by eye without folding the dataset about the Sun-Earth line (See the right side of Figure 5). A black line with an inclination angle of 17.5° is plotted in Figure 9, and corresponds to the location of this cone in the dataset. The only model compared to in this study that shows the spatial distribution of magnetopause crossings used to optimize it, Lin et al. (2010), observed a similar lack of crossings in this region of the nose in their dataset (Figure 1d). For models that are locally circular at the nose, the magnetopause distance from the Earth at some inclination angle is essentially the magnetopause distance at the subsolar point. In the case of the tractrix, there is a significant difference between the magnetopause position at the subsolar point and the magnetopause distance at some inclination angle. Subsequent analysis shows that the tractrix has good predictive performance in this area of the magnetopause (see Figure 11), so this may expose a performance vulnerability of the other models in this study.

In the tail, the tendency of conic section-based models Shue et al. (1998), Chao et al. (2002), and Lu et al. (2011) to flare outward continuously along the tail can be clearly seen. Petrinec and Russell (1996) reaches an asymptotic tail width using an inverse trigonometric function, but generally predicts a smaller tail width than the tractrix model. While Lin et al. (2010) is a conic section-based model that does not asymptotically reach a constant tail width, its functional form suppresses this effect such that it cannot be observed in the figure. The slope of the tractrix curve is smaller on the dayside than any other model, since every other model in this study has a blunt, spherical nose. On the nightside, the tractrix has a larger slope than any other model until about $-45R_E$ GSE X, after which models that flare outward strongly ((Shue et al., 1998), (Chao et al., 2002)) overtake the tractrix's downtail expansion rate. (Lin et al., 2010) and (Petrinec & Russell, 1996) always expand more slowly than the tractrix over reasonable distances downtail. This means that the tractrix gets to its maximum opening (asymptotic width) slower than both of these models.

### 4.1 Near-Earth Performance

To quantify the performance of a given model, the model magnetopause surface is constructed for each crossing in the validation dataset using its associated solar wind conditions, then the distance from that crossing to the closest point on the model surface is calculated. This distance will subsequently be referred to as the uncertainty. A smaller uncertainty corresponds to a more accurate prediction for the associated crossing. This is akin to the $d_{trac}$ function used in the likelihood (Equation 21), but can be extended to all the models compared to in this study. By calculating the uncertainties for the portion of the validation dataset that lies between $-15R_E$ and $15R_E$ GSE X, the tractrix's performance can be investigated in a region that has been the focus of most empirical modeling efforts.

To investigate the performance of the tractrix in different spatial regions, the uncertainties are split into $1R_E$ square bins in GSE X and GSE $\rho$, then each bin is averaged ($\rho$ in this case referencing a cylindrical coordinate system oriented around the GSE X axis, see Figure 4). Thus by considering which bins have smaller uncertainties a sense can be obtained of where the tractrix model has the best performance. The binned uncertainties are presented in Figure 10, with the relative number of crossings in each bin represented as the size of each bin. The highest density of crossings occurs close to the nose roughly from $9R_E$ to $14R_E$ GSE X and $3R_E$ to $5R_E$ GSE $\rho$. In this area, the trac-

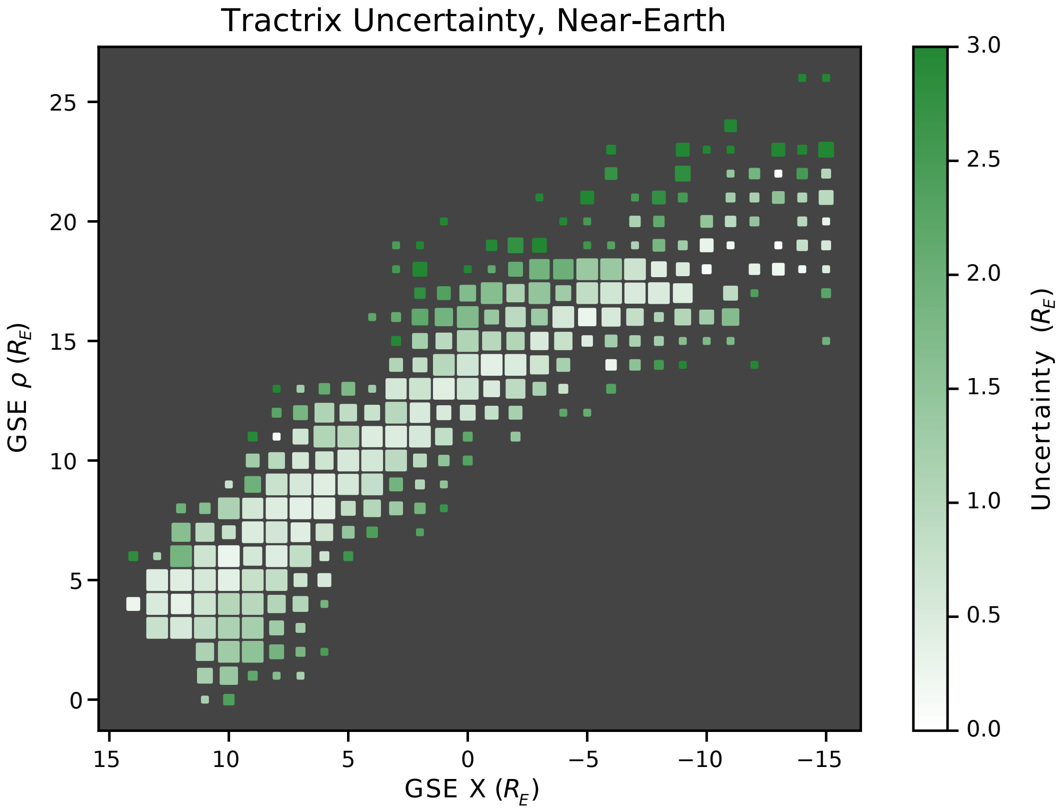


**Figure 10.** Binned and averaged uncertainties of the tractrix model on the near-Earth portion of the validation dataset, plotted as a function of GSE X and GSE $\rho$ (see Figure 4 for a definition of $\rho$). A darker bin corresponds to a higher uncertainty. The relative size of each bin corresponds to the relative number of crossings it contains (bins with more than 15 crossings are full sized). For reference, the largest bin at $x = 9R_E, \rho = 3R_E$ contains 76 crossings.

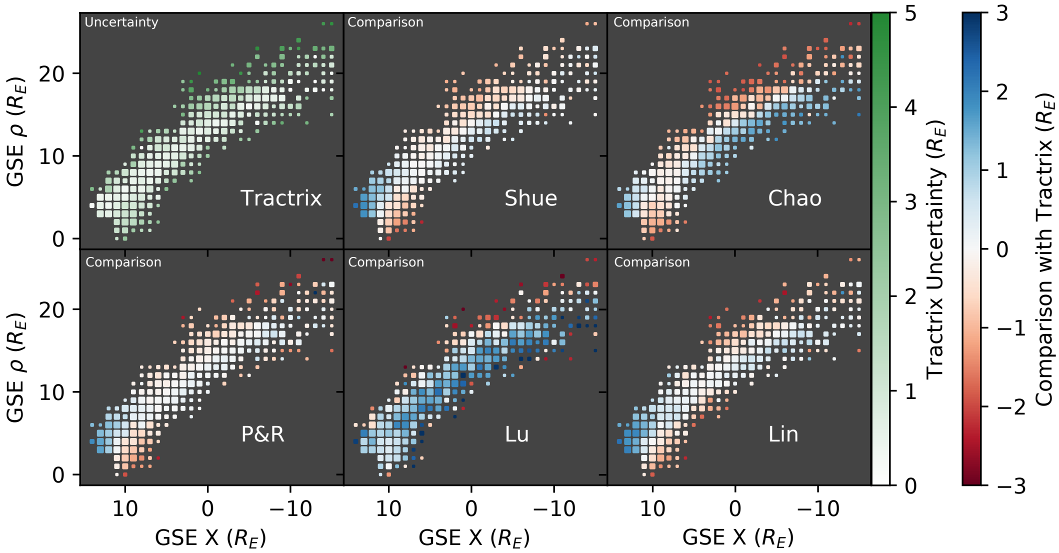


**Figure 11.** Difference between the average uncertainty for each model compared to in this study and the tractrix model in each $1R_E$ square bin in GSE X and and GSE $\rho$ (as in Figure 10), with the tractrix uncertainties also plotted for convenience. A blue bin indicates the tractrix outperforming the model being compared to, and a red bin indicate the tractrix performing worse than the model being compared to. The relative size of each bin corresponds to the relative number of crossings in each bin (bins with more than 15 crossings are full sized). The tractrix has better performance than all other models near the nose when the nose is extended (within about $3R_E$ of the largest GSE X position) and on the "closer half" of the nightside region.

trix has an average uncertainty of $0.71R_E$. Over the entirety of this portion of the validation dataset, the tractrix has an average uncertainty of $0.98R_E$.

For each of the other models, the model magnetopause surface is constructed for each crossing in the validation dataset using its associated solar wind conditions, then the shortest distance from that crossing to the model surface is calculated exactly as was done with the tractrix. These uncertainties were then binned and averaged according to the same scheme as was used to create the data presented in Figure 10. Then, the tractrix uncertainty in each bin was subtracted from the model uncertainty in order to compare the performance of the tractrix relative to each model. In Figure 11 the difference between each model's uncertainty and the tractrix's uncertainty are plotted in the same manner as Figure 10, but in this case blue bins indicate the tractrix outperforming the model being compared to, and the red bins indicate the tractrix performing worse than the model being compared to. This technique is analogous to a "skill score", a technique used in weather prediction (Benedetti, 2010). Note that for (Lin et al., 2010) and (Lu et al., 2011), magnetopause crossing positions are shifted into GSM coordinates since the models output surfaces in GSM coordinates. Since GSM and GSE share an x axis, the cylindrical binning scheme is the same in GSE and GSM coordinates.

Some systematic trends are visible in the performance, likely due to the forced shapes of the models. It is important to point out that the tractrix has a performance vulnerability for $x \leq 10R_E$, $\rho \leq 4R_E$ as compared to all models except Lu et al. (2011). There are fewer crossings in these bins than in those in the far nose, so it's possible that the magnetopause is less likely to be located in this spatial domain than in the other regions closest to the Sun-Earth line where there are more crossings. However, it is also impor-

| Model | % Tractrix Outperforms ($-15R_E < x < 15R_E$) | % Tractrix Outperforms (Nose, $\rho < 8R_E, x > 6R_E$) |
|---|---|---|
| Shue et al. (1998) | 50.6 | 56.9 |
| Chao et al. (2002) | 49.1 | 44.3 |
| Petrinec and Russell (1996) | 49.1 | 55.2 |
| Lu et al. (2011) | 70.5 | 72.2 |
| Lin et al. (2010) | 55.2 | 62.0 |

**Table 3.** Percentage of crossings more accurately predicted by the tractrix than each model, in two regimes of the near-Earth portion of the dataset. A higher percentage corresponds to better performance by the tractrix model. Note that the comparable/coin flip performance of the tractrix overall is drastically boosted at the nose, with the exception of its performance relative to Chao et al. (2002).

tant to point out that the tractrix outperforms all other models near the magnetopause nose particularly when the nose is extended (within about $3R_E$ of the largest GSE X position on Figure 11). This standoff position is commonly used for space weather application and planning for science missions, demonstrating value for the model. This also indicates that the extended nose of the tractrix model may allow it to predict the magnetopause location more accurately in this region. Additionally, the fact that the tractrix opens more slowly than the other models compared to in this study allows it to also have superior performance on the nightside for the inner half of the crossings as compared to all other models. Furthermore, the tractrix can outperform the Lu et al. (2011) model over the majority of the spatial domain considered here.

By calculating what percent of the time the tractrix has a closer prediction than a given model, that percentage can be used as a performance metric. The binning method used previously can be deceptive in that each bin does not necessarily contain the same amount of crossings; comparing the overall percentage of the time the tractrix is outperforming a given model gives a different view of the tractrix's performance that does not suffer from this effect. Table 3 contains the percentage of the crossings that were more accurately predicted by the tractrix model than each other model compared to in this study in two regimes. These regimes are the entire near-Earth dataset and a region roughly corresponding to the nose, chosen to be GSE $\rho < 8R_E, x > 6R_E$. The average uncertainty of each model in Earth radii can also be calculated for these two regimes, which is given in Table 4. It confirms what one could gather from Figure 11 qualitatively: the only model that it performs worse against near the nose than over the entire near-Earth dataset is Chao et al. (2002), which is likely due to the fact that it is trained on magnetopause crossings that occurred during extreme solar wind conditions, which boosts its performance near the nose when the magnetopause is compressed. Over the entire near-Earth dataset the tractrix has essentially comparable performance to any given model (except Lu et al. (2011), which it consistently outperforms), whereas near the nose it has drastically increased performance for most nose geometries.

### 4.2 Performance in the Magnetotail

The tractrix model's performance in the tail region can be evaluated in a similar manner as the previous section. The closest distance from each crossing in the ARTEMIS-obtained dataset to the tractrix magnetopause surface is calculated and taken as the uncertainty for that crossing. Then the uncertainties are split into $2R_E$ square bins in GSE X and and GSE $\rho$ and each bin is averaged. These binned uncertainties are plotted in Figure 12, with larger uncertainties represented by darker boxes, and the relative amount

| Model | Uncertainty ($R_E$) ($-15R_E < x < 15R_E$) | Uncertainty ($R_E$) (Nose, $\rho < 8R_E, x > 6R_E$) |
|---|---|---|
| Tractrix | 0.989 | 0.837 |
| Shue et al. (1998) | 1.02 | 1.07 |
| Chao et al. (2002) | 0.983 | 0.742 |
| Petrinec and Russell (1996) | 1.00 | 0.979 |
| Lu et al. (2011) | 1.90 | 1.74 |
| Lin et al. (2010) | 1.14 | 1.28 |

**Table 4.** Average uncertainty of each model for each of the regimes in Table 3 given in Earth radii (entire near-Earth dataset and approximate nose region).

of crossings in each bin represented as the relative sizes of each bin. It is worth reiterating that the curved shape of the magnetopause crossing distribution is the result of the crossings being obtained with ARTEMIS-P1, a spacecraft constrained to a roughly $10R_E$ diameter orbit around the moon, and not the structure of the magnetopause. For example, it is not the case that the magnetopause is never found in the regions in the top right and bottom left of Figure 12, it is simply that ARTEMIS-P1 never flies through those regions and thus cannot sample the magnetopause there.

The average width of the magnetotail in GSE is $\bar{\rho} = 25.9R_E$. Within $\Delta\rho = \pm 1R_E$ of $\bar{\rho}$, the tractrix model can predict the magnetopause position within $1.25R_E$. Within $\Delta\rho = \pm 4R_E$ of $\bar{\rho}$ ($22R_E \leq \rho \leq 30R_E$) the tractrix has an average uncertainty of $2.75R_E$. The more extreme tail widths that are sampled, the more the performance of the tractrix falls off. For the most extreme (and commensurately most rare) tail widths, the tractrix has an uncertainty of up to $20R_E$. This is likely due to the fact that the tractrix model is static, whereas the tail has been observed to be an environment for which the time history of the solar wind conditions can influence its instantaneous shape. For instance, the magnetotail has been observed in simulations to "flap around" significantly in response to gradients in solar wind conditions, which could produce these extremely small ($< 20R_E$) and extremely large ($> 30R_E$) tail widths that could not be reproduced by a static model such as the tractrix (Borovsky, 2012). For prolonged periods of low $B_S$, flux could also accumulate in the tail causing the tail to gradually increase in size in a way that the tractrix model cannot capture. It may be the case that certain solar wind conditions would produce a magnetopause that does not open to a asymptotic width in the way that the tractrix predicts.

Using the same uncertainty calculation, binning, and averaging scheme, the same comparison can be constructed for the other models. Then the performance of the models can be compared in the same way as Figure 11. From Figure 13, it can be seen that no model reproduces these extreme tail widths particularly accurately. All models in this comparison are static in the sense that they only use instantaneous solar wind conditions to predict a global magnetopause shape, so it follows that they are not able to reproduce these time-dependent effects either. Some models are able to consistently predict the location of the magnetopause better than the tractrix for large or small tail widths. Lin et al. (2010) consistently predicts small tail width crossings better than the tractrix, which is possibly due to the fact that it incorporates dawn-dusk asymmetries which can compress the magnetopause on the foreshock side.

To account for the fact that the crossings are not evenly distributed in the bins of Figure 13, what percent of the time the tractrix has a closer prediction than a given model can again be calculated crossing-by-crossing. The percentage of the time the tractrix outperforms each given model on this dataset is given in Table 5. The average uncertainty of each model in Earth radii over the regimes considered in Table 5 is given in Table 6.

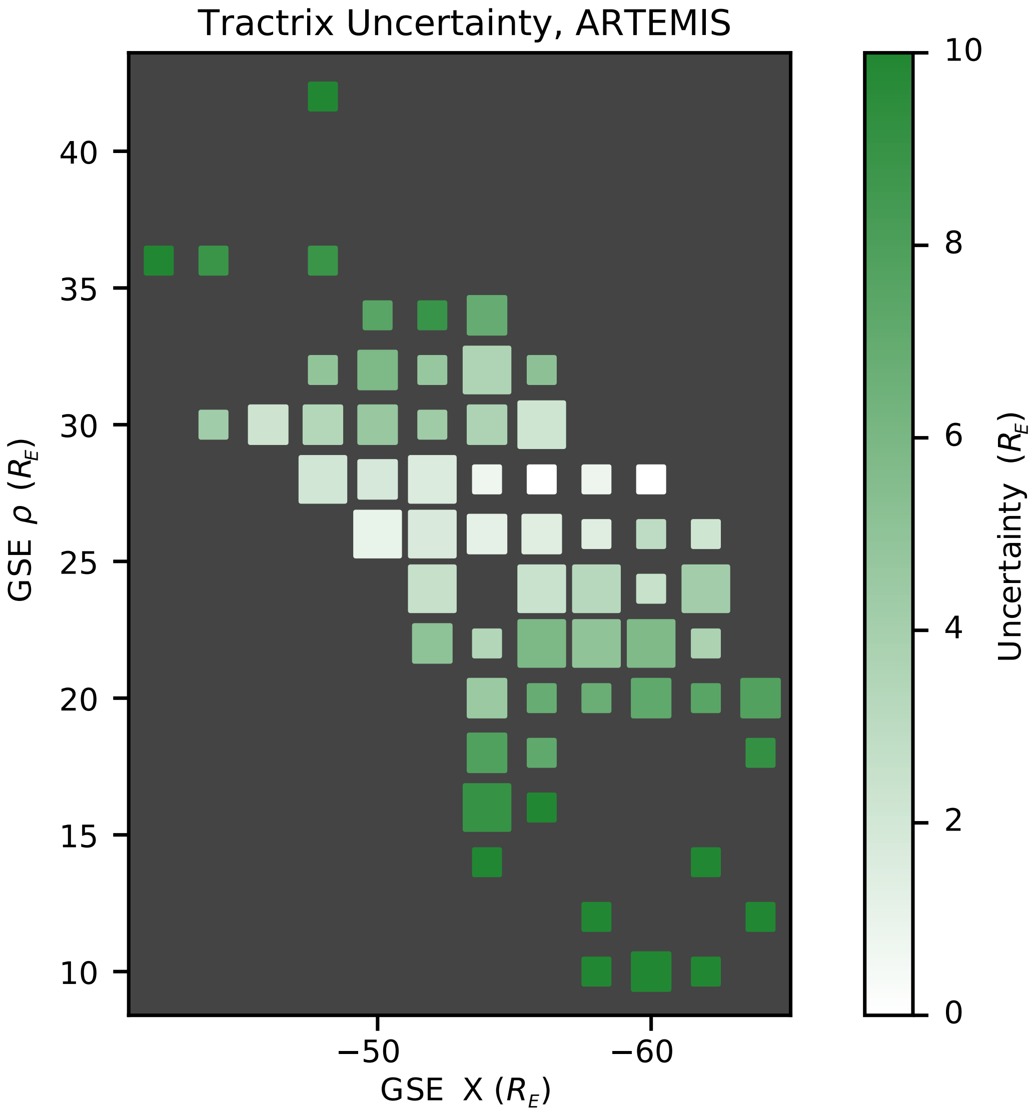


**Figure 12.** Binned and averaged uncertainties of the tractrix model on the ARTEMIS-obtained portion of the validation dataset, plotted as a function of GSE X and GSE $\rho$ (see Figure 4 for a definition of $\rho$). A darker bin corresponds to a higher uncertainty. The relative size of each bin corresponds to the relative number of crossings it contains (bins with more than 3 crossings are full sized). For reference, the largest bins at $x = -52R_E, \rho = 24R_E, 28R_E$ contain 7 crossings.

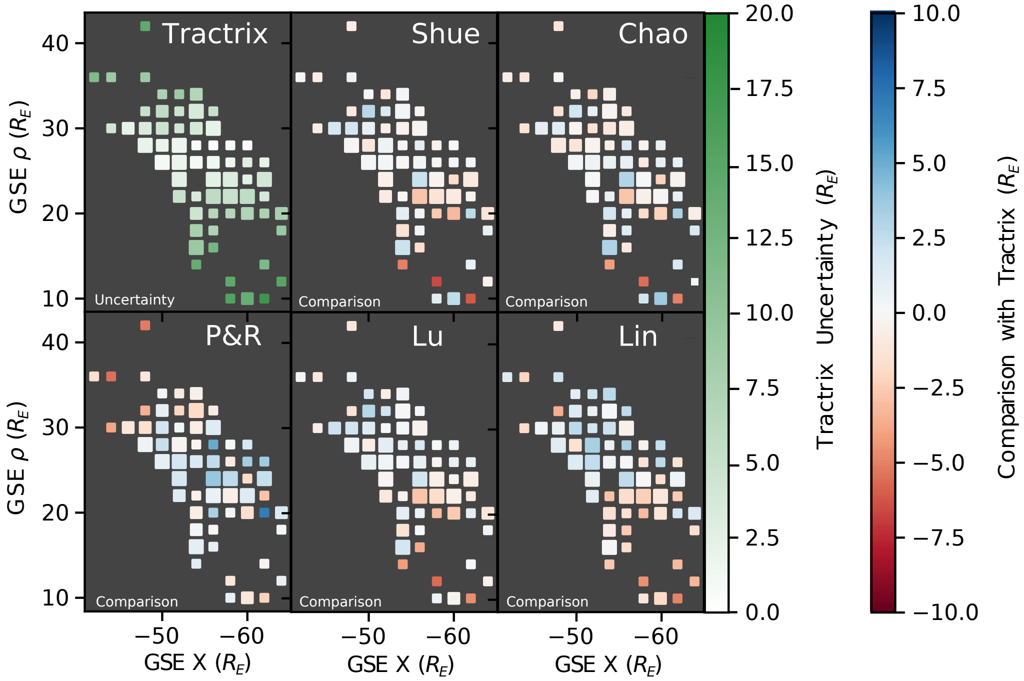


**Figure 13.** Difference between the average uncertainty for each model compared to in this study and the tractrix model in each $2R_E$ square bin in GSE X and and GSE $\rho$ (as in Figure 12), with the tractrix uncertainties also plotted for convenience. A blue bin indicates the tractrix outperforming the model being compared to, and a red bin indicate the tractrix performing worse than the model being compared to. The relative size of each bin corresponds to the relative number of crossings in each bin (bins with more than 3 crossings are full sized).

| Model | % Tractrix Outperforms (Total) | % Tractrix Outperforms ($25.9R_E \pm 5R_E$) | % Tractrix Outperforms ($B_S \geq 3.14nT$) |
|---|---|---|---|
| Shue et al. (1998) | 45.8 | 39.2 | 56.3 |
| Chao et al. (2002) | 50.0 | 50.0 | 56.3 |
| Petrinec and Russell (1996) | 53.3 | 63.5 | 58.3 |
| Lu et al. (2011) | 55.1 | 54.1 | 70.8 |
| Lin et al. (2010) | 53.4 | 60.8 | 64.6 |

**Table 5.** Percentage of crossings more accurately predicted by the tractrix than each model in the tail, across the entire tail dataset (column two), close to the average tail width (column three), and for $B_S$ greater than its average value of $3.14nT$ (column four). This is akin to "Southward" IMF $B_z$. Note that the comparable/coin flip performance of the tractix overall is improved significantly close to the average tail width, with the exception of its performance relative to Shue et al. (1998). The tractrix also has significantly improved performance for large $B_S$, as it outperforms all other models considered in this study for $B_S$ greater than its average value.

| Model | Uncertainty ($R_E$) (Total) | Uncertainty ($R_E$) ($25.9R_E \pm 5R_E$) | Uncertainty ($R_E$) ($B_S \geq 3.14nT$) |
|---|---|---|---|
| Tractrix | 4.98 | 3.20 | 5.55 |
| Shue et al. (1998) | 4.81 | 2.95 | 6.18 |
| Chao et al. (2002) | 4.94 | 3.14 | 6.33 |
| Petrinec and Russell (1996) | 5.48 | 4.45 | 6.42 |
| Lu et al. (2011) | 4.89 | 3.15 | 6.07 |
| Lin et al. (2010) | 5.12 | 3.17 | 6.43 |

**Table 6.** Average uncertainty of each model for each of the regimes in Table 5 given in Earth radii (entire ARTEMIS-obtained dataset, within $5R_E$ of the average tail width, and larger than average $B_S$).

Over the entire dataset, note that the tractrix has better performance than two of the models, but for all models except Petrinec and Russell (1996) each model has essentially coin-flip odds of being the better model. If the considered crossings are limited to ones that occurred within $5R_E$ of the average tail width, one can see that the tractrix significantly outperforms every model except Shue et al. (1998) (which also does better in total performance in the tail). Interestingly, for crossings with a large associated $B_S$, the tractrix also has significantly improved performance. For crossings associated with a greater-than-average $B_S$ (i.e. $B_S \geq 3.14nT$) the tractrix outperforms all models considered in this study. Since prolonged large values of $B_S$ correspond to the formation of substorms in the same way prolonged southward IMF $B_z$ does, this could indicate that the tractrix is better at predicting the position of the tail as its diameter changes during substorms than other empirical models, or that steady reconnection helps mitigate flux accumulation in the tail thereby achieving a more constant tail width (Maezawa, 1975).

## 5 Conclusions

The tractrix model of the magnetopause advances our understanding of the magnetospheric system because it allows us to perform magnetopause fitting to a functional form that has physical basis, unlike any previously derived model. To draw an analogy, we could in principle fit particle distributions to arbitrary functional forms (as we currently do for magnetopause crossings). Instead, we fit them to functions such as Maxwellians, power laws, and kappa functions (Collier, 1993), which have physical basis. In the case of Maxwellians, a connection to the central limit theorem and collision operators allows us to connect the parameters of the fit function to physical properties of the plasma. The tractrix model of the magnetopause is a step towards placing the characterization of the magnetopause shape onto a similar physical basis.

Another advantage of the tractrix model is its simplicity. John von Neumann famously stated "With four parameters I can fit an elephant, and with five I can make him wiggle his trunk", which has been shown to be literally true (Mayer et al., 2010). Even though the magnetosphere is not an elephant (despite it having a nose and a tail), John von Neumann's point that models with fewer free parameters have greater predictive power still applies. The tractrix model has only six tunable parameters, less than half that of some models considered in this study such as Lin et al. (2010). However, it still has comparable or better performance than the models considered in this study for the majority of regimes, and outperforms Lin et al. (2010) across both sections of the dataset.

The MCMC machine learning method utilized in this study highlights the potential of machine learning methods to uncover new physics. Instead of an optimization algorithm that seeks to simply minimize some loss parameter, estimating the posterior distribution allows one to uncover local extrema that could represent other physical states of the system that are not the dominant state captured by the maximum likelihood parameters. Even though the tractrix model is fairly simple, its scientific utility is boosted by the algorithm used to calculate its optimal state.

One limitation of current observations emphasized by this study is the fact that single-point spacecraft observations make comparing any model to the instantaneous global shape of the magnetopause impossible. This is a major challenge for the study of many processes that control the arrangement of Earth's magnetosphere, especially magnetic reconnection. Upcoming magnetopause imaging missions such as the Lunar Environment Heliospheric X-ray Imager (LEXI) and the joint ESA-CSA Solar wind Magnetosphere Ionosphere Link Explorer (SMILE) mission will provide near instantaneous (integration time in minutes or tens-of-minutes) imaging of the global magnetopause for comparison to models like the one presented here.

## Appendix A Libraries for Python and IDL

An implementation of the tractrix model is available for Python via pip or manual download from the public GitHub repository at the URL `https://github.com/connor-obrien888/tractrix-python`. An implementation of the tractrix model is available for IDL via manual download from the public GitHub repository at the URL `https://github.com/connor-obrien888/tractrix-IDL`. Versions compatible with SPEDAS and pySPEDAS are under development and will be available at the above URLs.

**Acknowledgments**

Special thanks to Pat Tamburo for guidance on the optimal use of MCMC. Codes used in this project are available on the GitHub repository at the URL `https://github.com/connor-obrien888/tractrix-paper` where links to convenient implementations of the tractrix model in popular programming languages can also be found. Both datasets are available from NASA's Space Physics Data Center in their entirety at the URL `https://spdf.gsfc.nasa.gov/pub/data/aaa_special-purpose-datasets/aaa_boundary_crossings/magnetopause_crossings`. Solar wind data used in this analysis can be obtained via the OMNI database which can be accessed at the URL `https://omniweb.gsfc.nasa.gov`.
Authors C. J. O'Brien and B. M. Walsh acknowledge support by NASA Grant 80NSSC20K1710.
David G. Sibeck and Michael R. Collier acknowledge support from the USPI program.

## References


Angelopoulos, V. (2008, December). The THEMIS Mission. *Space Science Reviews*, *141*(1-4), 5–34. Retrieved 2020-11-11, from http://link.springer.com/10.1007/s11214-008-9336-1 doi: 10.1007/s11214-008-9336-1

Avanov, L. A., Smirnov, V. N., Waite, J. H., Fuselier, S. A., & Vaisberg, O. L. (2001, December). High-latitude magnetic reconnection in sub-Alfvénic flow: Interball Tail observations on May 29, 1996. *Journal of Geophysical Research: Space Physics*, *106*(A12), 29491–29502. Retrieved 2020-11-19, from http://doi.wiley.com/10.1029/2000JA000460 doi: 10.1029/2000JA000460

Benedetti, R. (2010, January). Scoring Rules for Forecast Verification. *Monthly Weather Review*, *138*(1), 203–211. Retrieved 2020-11-04, from https://journals.ametsoc.org/mwr/article/138/1/203/70799/Scoring-Rules-for-Forecast-Verification doi: 10.1175/2009MWR2945.1

Boardsen, S. A., Eastman, T. E., Sotirelis, T., & Green, J. L. (2000, October). An empirical model of the high-latitude magnetopause. *Journal of Geophysical Research: Space Physics*, *105*(A10), 23193–23219. Retrieved 2020-05-06, from http://doi.wiley.com/10.1029/1998JA000143 doi: 10.1029/1998JA000143

Borovsky, J. E. (2012, June). The effect of sudden wind shear on the Earth's magnetosphere: Statistics of wind shear events and CCMC simulations of magnetotail disconnections: WIND SHEAR ON EARTH'S MAGNETOSPHERE. *Journal of Geophysical Research: Space Physics*, *117*(A6), n/a–n/a. Retrieved 2020-06-04, from http://doi.wiley.com/10.1029/2012JA017623 doi: 10.1029/2012JA017623

Borovsky, J. E. (2013, November). Physics-based solar wind driver functions for the magnetosphere: Combining the reconnection-coupled MHD generator with the viscous interaction: SOLAR WIND DRIVER FUNCTIONS. *Journal of Geophysical Research: Space Physics*, *118*(11), 7119–7150. Retrieved 2020-11-19, from http://doi.wiley.com/10.1002/jgra.50557 doi: 10.1002/jgra.50557

Chao, J., Wu, D., Lin, C.-H., Yang, Y.-H., Wang, X., Kessel, M., … Lepping, R. (2002). Models for the size and shape of the earth's magnetopause and bow shock. In *COSPAR Colloquia Series* (Vol. 12, pp. 127–135). Elsevier. Retrieved 2020-05-11, from https://linkinghub.elsevier.com/retrieve/pii/S0964274902802128 doi: 10.1016/S0964-2749(02)80212-8

Collier, M. R. (1993, August). On generating Kappa-like distribution functions using velocity space Lévy flights. *Geophysical Research Letters*, *20*(15), 1531–1534. Retrieved 2020-11-06, from http://doi.wiley.com/10.1029/93GL01702 doi: 10.1029/93GL01702

Collier, M. R., Slavin, J. A., Lepping, R. P., Ogilvie, K., Szabo, A., Laakso, H., & Taguchi, S. (1998, August). Multispacecraft observations of sudden impulses in the magnetotail caused by solar wind pressure discontinuities: Wind and IMP 8. *Journal of Geophysical Research: Space Physics*, *103*(A8), 17293–17305. Retrieved 2020-10-22, from http://doi.wiley.com/10.1029/97JA02870 doi: 10.1029/97JA02870

Dimmock, A. P., & Nykyri, K. (2013, August). The statistical mapping of magnetosheath plasma properties based on THEMIS measurements in the magnetosheath interplanetary medium reference frame: MAGNETOSHEATH STATISTICAL MAPPING. *Journal of Geophysical Research: Space Physics*, *118*(8), 4963–4976. Retrieved 2020-11-11, from http://doi.wiley.com/10.1002/jgra.50465 doi: 10.1002/jgra.50465

Dmitriev, A., Suvorova, A., & Chao, J.-K. (2011, May). A predictive model of geosynchronous magnetopause crossings: PREDICTIVE MODEL OF GMCS. *Journal of Geophysical Research: Space Physics*, *116*(A5). Retrieved 2020-05-06, from http://doi.wiley.com/10.1029/2010JA016208 doi: 10.1029/2010JA016208

Dmitriev, A. V. (2003). Comparative study of bow shock models using Wind and Geotail observations. *Journal of Geophysical Research*, *108*(A12), 1464. Retrieved 2020-10-28, from http://doi.wiley.com/10.1029/2003JA010027 doi: 10.1029/2003JA010027

Fairfield, D. H. (1971, October). Average and unusual locations of the Earth's magnetopause and bow shock. *Journal of Geophysical Research*, *76*(28), 6700–6716. Retrieved 2020-10-22, from http://doi.wiley.com/10.1029/JA076i028p06700 doi: 10.1029/JA076i028p06700

Fairfield, D. H. (1992). On the structure of the distant magnetotail: ISEE 3. *Journal of Geophysical Research*, *97*(A2), 1403. Retrieved 2021-01-20, from http://doi.wiley.com/10.1029/91JA02388 doi: 10.1029/91JA02388

Foreman-Mackey, D. (2016, jun). corner.py: Scatterplot matrices in python. *The Journal of Open Source Software*, *1*(2), 24. Retrieved from https://doi.org/10.21105/joss.00024 doi: 10.21105/joss.00024

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. (2013, March). emcee : The MCMC Hammer. *Publications of the Astronomical Society of the Pacific*, *125*(925), 306–312. Retrieved 2020-11-02, from http://iopscience.iop.org/article/10.1086/670067 doi: 10.1086/670067

Gencturk Akay, I., Kaymaz, Z., & Sibeck, D. G. (2019, January). Magnetotail boundary crossings at lunar distances: ARTEMIS observations. *Journal of Atmospheric and Solar-Terrestrial Physics*, *182*, 45–60. Retrieved 2020-05-11, from https://linkinghub.elsevier.com/retrieve/pii/S136468261730562X doi: 10.1016/j.jastp.2018.11.002

Goldstein, M. L., & Roberts, D. A. (1999, November). Magnetohydrodynamic turbulence in the solar wind. *Physics of Plasmas*, *6*(11), 4154–4160. Retrieved 2020-10-22, from http://aip.scitation.org/doi/10.1063/1.873680 doi: 10.1063/1.873680

Goodman, J., & Weare, J. (2010, January). Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, *5*(1), 65–80. Retrieved 2020-11-02, from http://msp.org/camcos/2010/5-1/p04.xhtml doi: 10.2140/camcos.2010.5.65

Hasegawa, H. (2002). Plasma entry across the distant tail magnetopause 1. Global properties and IMF dependence. *Journal of Geophysical Research*, *107*(A5), 1063. Retrieved 2020-04-13, from http://doi.wiley.com/10.1029/2001JA900139 doi: 10.1029/2001JA900139

Holzer, R. E., & Slavin, J. A. (1978). Magnetic flux transfer associated with expansions and contractions of the dayside magnetosphere. *Journal of Geophysical Research*, *83*(A8), 3831. Retrieved 2020-11-09, from http://doi.wiley.com/10.1029/JA083iA08p03831 doi: 10.1029/JA083iA08p03831

Howe, H. C., & Binsack, J. H. (1972, July). Explorer 33 and 35 plasma observations of magnetosheath flow. *Journal of Geophysical Research*, *77*(19), 3334–3344. Retrieved 2020-05-11, from http://doi.wiley.com/10.1029/JA077i019p03334 doi: 10.1029/JA077i019p03334

Kan, J. R., & Lee, L. C. (1979, July). Energy coupling function and solar wind-magnetosphere dynamo. *Geophysical Research Letters*, *6*(7), 577–580. Retrieved 2020-11-24, from http://doi.wiley.com/10.1029/GL006i007p00577 doi: 10.1029/GL006i007p00577

Kawano, H., Petrinec, S. M., Russell, C. T., & Higuchi, T. (1999, January). Magnetopause shape determinations from measured position and estimated flaring angle. *Journal of Geophysical Research: Space Physics*, *104*(A1), 247–261. Retrieved 2020-11-09, from http://doi.wiley.com/10.1029/98JA02479 doi: 10.1029/98JA02479

King, J. H. (2005). Solar wind spatial scales in and comparisons of hourly Wind and ACE plasma and magnetic field data. *Journal of Geophysical Research*, *110*(A2), A02104. Retrieved 2020-05-26, from http://doi.wiley.com/

10.1029/2004JA010649 doi: 10.1029/2004JA010649

Kivelson, M. G., Kennel, C. F., McPherron, R. L., Russell, C. T., Southwood, D. J., Walker, R. J., ... Hughes, T. J. (1993). The Galileo Earth encounter: Magnetometer and allied measurements. *Journal of Geophysical Research*, *98*(A7), 11299. Retrieved 2021-02-08, from http://doi.wiley.com/10.1029/92JA03001 doi: 10.1029/92JA03001

Kuznetsov, S., & Suvorova, A. (1998, January). Solar wind magnetic field and pressure during magnetopause crossings at geosynchronous orbit. *Advances in Space Research*, *22*(1), 63–66. Retrieved 2020-11-09, from https://linkinghub.elsevier.com/retrieve/pii/S0273117797011010 doi: 10.1016/S0273-1177(97)01101-0

Lawrence, J. D. (2014). *A catalog of special plane curves.* Dover Publications.

Le, G., Raeder, J., Russell, C. T., Lu, G., Petrinec, S. M., & Mozer, F. S. (2001, October). Polar cusp and vicinity under strongly northward interplanetary magnetic field on April 11, 1997: Observations and MHD simulations. *Journal of Geophysical Research: Space Physics*, *106*(A10), 21083–21093. Retrieved 2020-11-19, from http://doi.wiley.com/10.1029/2000JA900091 doi: 10.1029/2000JA900091

Lin, R. L., Zhang, X. X., Liu, S. Q., Wang, Y. L., & Gong, J. C. (2010, April). A three-dimensional asymmetric magnetopause model: THREE-DIMENSIONAL MAGNETOPAUSE MODEL. *Journal of Geophysical Research: Space Physics*, *115*(A4), n/a–n/a. Retrieved 2020-03-12, from http://doi.wiley.com/10.1029/2009JA014235 doi: 10.1029/2009JA014235

Liu, Z.-Q., Lu, J. Y., Kabin, K., Yang, Y. F., Zhao, M. X., & Cao, X. (2012, July). Dipole tilt control of the magnetopause for southward IMF from global magnetohydrodynamic simulations: DIPOLE TILT CONTROL OF THE MAGNETOPAUSE. *Journal of Geophysical Research: Space Physics*, *117*(A7), n/a–n/a. Retrieved 2020-05-06, from http://doi.wiley.com/10.1029/2011JA017441 doi: 10.1029/2011JA017441

Lu, J. Y., Liu, Z.-Q., Kabin, K., Zhao, M. X., Liu, D. D., Zhou, Q., & Xiao, Y. (2011, September). Three dimensional shape of the magnetopause: Global MHD results: THREE DIMENSIONAL MAGNETOPAUSE. *Journal of Geophysical Research: Space Physics*, *116*(A9), n/a–n/a. Retrieved 2020-05-06, from http://doi.wiley.com/10.1029/2010JA016418 doi: 10.1029/2010JA016418

Maezawa, K. (1975, September). Magnetotail boundary motion associated with geomagnetic substorms. *Journal of Geophysical Research*, *80*(25), 3543–3548. Retrieved 2020-12-14, from http://doi.wiley.com/10.1029/JA080i025p03543 doi: 10.1029/JA080i025p03543

Mayer, J., Khairy, K., & Howard, J. (2010, June). Drawing an elephant with four complex parameters. *American Journal of Physics*, *78*(6), 648–649. Retrieved 2020-11-06, from http://aapt.scitation.org/doi/10.1119/1.3254017 doi: 10.1119/1.3254017

Mieth, J. Z. D., Frühauff, D., & Glassmeier, K.-H. (2018, July). *Statistical Analysis of Magnetopause Crossings at Lunar Distances* (preprint). Magnetosphere & space plasma physics/Magnetopause, cusp, and boundary layers. Retrieved 2021-01-27, from https://angeo.copernicus.org/preprints/angeo-2018-66/angeo-2018-66.pdf doi: 10.5194/angeo-2018-66

Newell, P. T., Sotirelis, T., Liou, K., Meng, C.-I., & Rich, F. J. (2007, January). A nearly universal solar wind-magnetosphere coupling function inferred from 10 magnetospheric state variables: UNIVERSAL COUPLING FUNCTION. *Journal of Geophysical Research: Space Physics*, *112*(A1), n/a–n/a. Retrieved 2020-11-11, from http://doi.wiley.com/10.1029/2006JA012015 doi: 10.1029/2006JA012015

Ogino, T., Walker, R., & Ashour-Abdalla, M. (1992, December). A global magnetohydrodynamic simulation of the magnetosheath and magnetosphere when the interplanetary magnetic field is northward. *IEEE Transactions on Plasma Science*, *20*(6), 817–828. Retrieved 2021-02-17, from http://ieeexplore.ieee.org/document/199534/ doi: 10.1109/27.199534

Park, K. S., Lee, D., Ogino, T., & Lee, D. H. (2015, September). MHD simulations using average solar wind conditions for substorms observed under northward IMF conditions. *Journal of Geophysical Research: Space Physics*, *120*(9), 7672–7686. Retrieved 2021-02-17, from https://onlinelibrary.wiley.com/doi/abs/10.1002/2015JA021005 doi: 10.1002/2015JA021005

Perreault, P., & Akasofu, S.-I. (1978, September). A study of geomagnetic storms. *Geophysical Journal International*, *54*(3), 547–573. Retrieved 2020-11-24, from https://academic.oup.com/gji/article-lookup/doi/10.1111/j.1365-246X.1978.tb05494.x doi: 10.1111/j.1365-246X.1978.tb05494.x

Petrinec, S. M., & Russell, C. T. (1993, March). External and internal influences on the size of the dayside terrestrial magnetosphere. *Geophysical Research Letters*, *20*(5), 339–342. Retrieved 2020-11-09, from http://doi.wiley.com/10.1029/93GL00085 doi: 10.1029/93GL00085

Petrinec, S. M., & Russell, C. T. (1996, January). Near-Earth magnetotail shape and size as determined from the magnetopause flaring angle. *Journal of Geophysical Research: Space Physics*, *101*(A1), 137–152. Retrieved 2020-05-11, from http://doi.wiley.com/10.1029/95JA02834 doi: 10.1029/95JA02834

Ridley, A. J. (2005, December). A new formulation for the ionospheric cross polar cap potential including saturation effects. *Annales Geophysicae*, *23*(11), 3533–3547. Retrieved 2020-11-19, from https://angeo.copernicus.org/articles/23/3533/2005/ doi: 10.5194/angeo-23-3533-2005

Roberts, D. A., Ghosh, S., Goldstein, M. L., & Mattheaus, W. H. (1991, December). Magnetohydrodynamic simulation of the radial evolution and stream structure of solar-wind turbulence. *Physical Review Letters*, *67*(27), 3741–3744. Retrieved 2020-10-22, from https://link.aps.org/doi/10.1103/PhysRevLett.67.3741 doi: 10.1103/PhysRevLett.67.3741

Roelof, E. C., & Sibeck, D. G. (1993). Magnetopause Shape as a Bivariate Function of Interplanetary Magnetic Field $B_z$ and Solar Wind Dynamic Pressure. *Journal of Geophysical Research*, *98*(A12), 21421–21450. Retrieved 2020-07-14, from http://doi.wiley.com/10.1029/93JA02362 doi: 10.1029/93JA02362

Shepherd, S. G. (2007, March). Polar cap potential saturation: Observations, theory, and modeling. *Journal of Atmospheric and Solar-Terrestrial Physics*, *69*(3), 234–248. Retrieved 2020-11-19, from https://linkinghub.elsevier.com/retrieve/pii/S136468260600263X doi: 10.1016/j.jastp.2006.07.022

Shue, J.-H., Chao, J. K., Fu, H. C., Russell, C. T., Song, P., Khurana, K. K., & Singer, H. J. (1997, May). A new functional form to study the solar wind control of the magnetopause size and shape. *Journal of Geophysical Research: Space Physics*, *102*(A5), 9497–9511. Retrieved 2020-03-12, from http://doi.wiley.com/10.1029/97JA00196 doi: 10.1029/97JA00196

Shue, J.-H., Song, P., Russell, C. T., Steinberg, J. T., Chao, J. K., Zastenker, G., … Kawano, H. (1998, August). Magnetopause location under extreme solar wind conditions. *Journal of Geophysical Research: Space Physics*, *103*(A8), 17691–17700. Retrieved 2020-05-11, from http://doi.wiley.com/10.1029/98JA01103 doi: 10.1029/98JA01103

Sibeck, D. G., & Lin, R.-Q. (2014, February). Size and shape of the distant magnetotail: Size/Shape of the Distant Magnetotail. *Journal of Geophysical Research: Space Physics*, *119*(2), 1028–1043. Retrieved 2020-12-08, from http://doi.wiley.com/10.1002/2013JA019471 doi: 10.1002/2013JA019471

Sibeck, D. G., Lopez, R. E., & Roelof, E. C. (1991). Solar wind control of the magnetopause shape, location, and motion. *Journal of Geophysical Research*, *96*(A4), 5489. Retrieved 2020-11-19, from http://doi.wiley.com/10.1029/

90JA02464 doi: 10.1029/90JA02464
Siscoe, G. L. (2002). Hill model of transpolar potential saturation: Comparisons with MHD simulations. *Journal of Geophysical Research*, *107*(A6), 1075. Retrieved 2020-11-19, from http://doi.wiley.com/10.1029/2001JA000109 doi: 10.1029/2001JA000109
Slavin, J. A., Tsurutani, B. T., Smith, E. J., Jones, D. E., & Sibeck, D. G. (1983, October). Average configuration of the distant (<220 R $_e$ ) magnetotail: Initial ISEE-3 magnetic field results. *Geophysical Research Letters*, *10*(10), 973–976. Retrieved 2021-02-08, from http://doi.wiley.com/10.1029/GL010i010p00973 doi: 10.1029/GL010i010p00973
Spreiter, J. R., & Alksne, A. Y. (1969). Plasma flow around the magnetosphere. *Reviews of Geophysics*, *7*(1, 2), 11. Retrieved 2020-10-22, from http://doi.wiley.com/10.1029/RG007i001p00011 doi: 10.1029/RG007i001p00011
Spreiter, J. R., Summers, A. L., & Alksne, A. Y. (1966, March). Hydromagnetic flow around the magnetosphere. *Planetary and Space Science*, *14*(3), 223–253. Retrieved 2020-11-09, from https://linkinghub.elsevier.com/retrieve/pii/0032063366901243 doi: 10.1016/0032-0633(66)90124-3
Vasyliunas, V. M., Kan, J. R., Siscoe, G. L., & Akasofu, S.-I. (1982, April). Scaling relations governing magnetospheric energy transfer. *Planetary and Space Science*, *30*(4), 359–365. Retrieved 2020-11-24, from https://linkinghub.elsevier.com/retrieve/pii/0032063382900411 doi: 10.1016/0032-0633(82)90041-1
Walsh, B. M., Sibeck, D. G., Wang, Y., & Fairfield, D. H. (2012, December). Dawn-dusk asymmetries in the Earth's magnetosheath: MAGNETOSHEATH ASYMMETRIES. *Journal of Geophysical Research: Space Physics*, *117*(A12), n/a–n/a. Retrieved 2020-10-22, from http://doi.wiley.com/10.1029/2012JA018240 doi: 10.1029/2012JA018240
Wang, Y., Sibeck, D. G., Merka, J., Boardsen, S. A., Karimabadi, H., Sipes, T. B., . . . Lin, R. (2013, May). A new three-dimensional magnetopause model with a support vector regression machine and a large database of multiple spacecraft observations: 3-D MAGNETOPAUSE MODEL WITH SVRM. *Journal of Geophysical Research: Space Physics*, *118*(5), 2173–2184. Retrieved 2020-03-12, from http://doi.wiley.com/10.1002/jgra.50226 doi: 10.1002/jgra.50226
Welling, D. T., & Ridley, A. J. (2010, April). Exploring sources of magnetospheric plasma using multispecies MHD: PLASMA SOURCES IN BATSRUS. *Journal of Geophysical Research: Space Physics*, *115*(A4), n/a–n/a. Retrieved 2021-02-17, from http://doi.wiley.com/10.1029/2009JA014596 doi: 10.1029/2009JA014596
Winslow, R. M., Anderson, B. J., Johnson, C. L., Slavin, J. A., Korth, H., Purucker, M. E., . . . Solomon, S. C. (2013, May). Mercury's magnetopause and bow shock from MESSENGER Magnetometer observations: MERCURY'S MAGNETOPAUSE AND BOW SHOCK. *Journal of Geophysical Research: Space Physics*, *118*(5), 2213–2227. Retrieved 2020-11-09, from http://doi.wiley.com/10.1002/jgra.50237 doi: 10.1002/jgra.50237
Yang, Y.-H. (2003). Saturation of IMF $B_z$ influence on the position of dayside magnetopause. *Journal of Geophysical Research*, *108*(A3), 1104. Retrieved 2020-04-27, from http://doi.wiley.com/10.1029/2002JA009621 doi: 10.1029/2002JA009621

# Supporting Information for "The Tractrix Magnetopause: A Novel Physics-Based Functional Form for the Magnetopause Shape"

C. J. O'Brien[1], M. R. Collier[2], B.M. Walsh[1], D.G. Sibeck[2], E. Taylor[3]

[1]Center for Space Physics, Boston University, Boston, MA, USA

[2]NASA/GSFC, Greenbelt, MD, USA

[3]Howard University, Washington, DC, USA

## Contents of this file

1. Figures S1 to S4

## Introduction

Below are supporting figures giving additional insight into the structure of the tractrix model and how that structure compares to the other models in this study, as well as figures showing the mean and standard deviation of the dynamic pressure in the magnetopause crossing dataset.

## References

Chao, J., Wu, D., Lin, C.-H., Yang, Y.-H., Wang, X., Kessel, M., ... Lepping, R. (2002). Models for the size and shape of the earth's magnetopause and bow shock. In *COSPAR Colloquia Series* (Vol. 12, pp. 127–135). Elsevier. Retrieved 2020-05-

11, from https://linkinghub.elsevier.com/retrieve/pii/S0964274902802128 doi: 10.1016/S0964-2749(02)80212-8

Lin, R. L., Zhang, X. X., Liu, S. Q., Wang, Y. L., & Gong, J. C. (2010, April). A three-dimensional asymmetric magnetopause model: THREE-DIMENSIONAL MAGNETOPAUSE MODEL. *Journal of Geophysical Research: Space Physics*, *115*(A4), n/a–n/a. Retrieved 2020-03-12, from http://doi.wiley.com/10.1029/2009JA014235 doi: 10.1029/2009JA014235

Petrinec, S. M., & Russell, C. T. (1996, January). Near-Earth magnetotail shape and size as determined from the magnetopause flaring angle. *Journal of Geophysical Research: Space Physics*, *101*(A1), 137–152. Retrieved 2020-05-11, from http://doi.wiley.com/10.1029/95JA02834 doi: 10.1029/95JA02834

Shue, J.-H., Song, P., Russell, C. T., Steinberg, J. T., Chao, J. K., Zastenker, G., ... Kawano, H. (1998, August). Magnetopause location under extreme solar wind conditions. *Journal of Geophysical Research: Space Physics*, *103*(A8), 17691–17700. Retrieved 2020-05-11, from http://doi.wiley.com/10.1029/98JA01103 doi: 10.1029/98JA01103

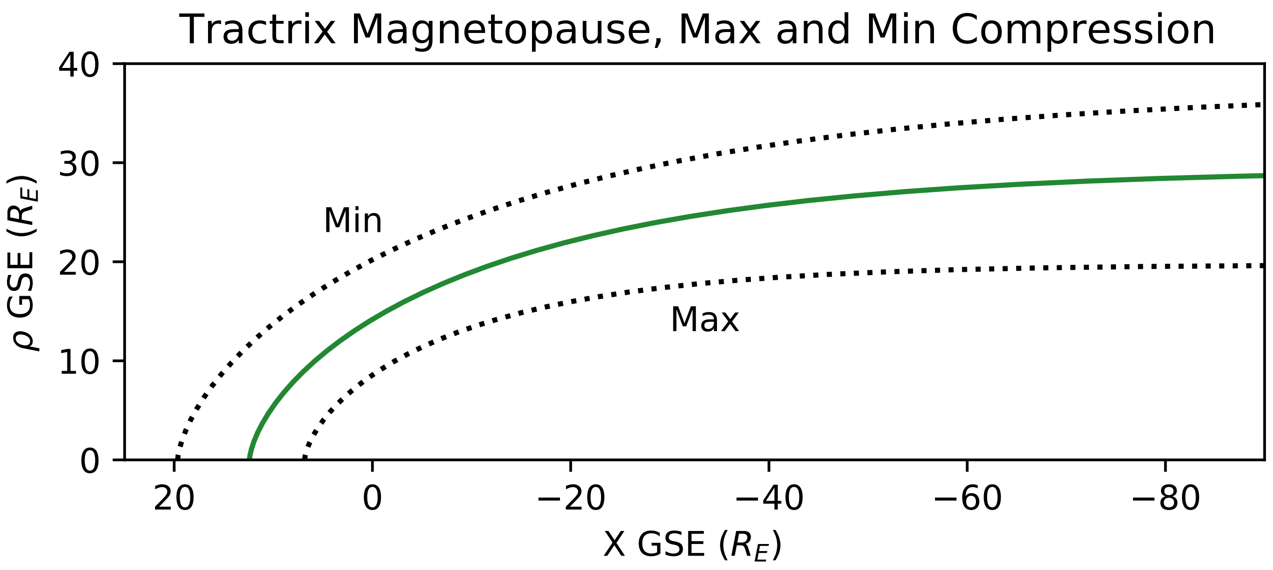


**Figure S1.** The tractrix model surface for mean values of $B_S$ and $P_{dyn}$ ($2.64nT$ and $2.40nPa$), as well as for maximum and minimum compression. Maximum compression corresponds to the maximum values of $B_S$ and $P_{dyn}$ ($17.0nT$ and $58.2nPa$). Minimum compression corresponds to the minimum values of $B_S$ and $P_{dyn}$ ($0.00nT$ and $0.182nPa$).

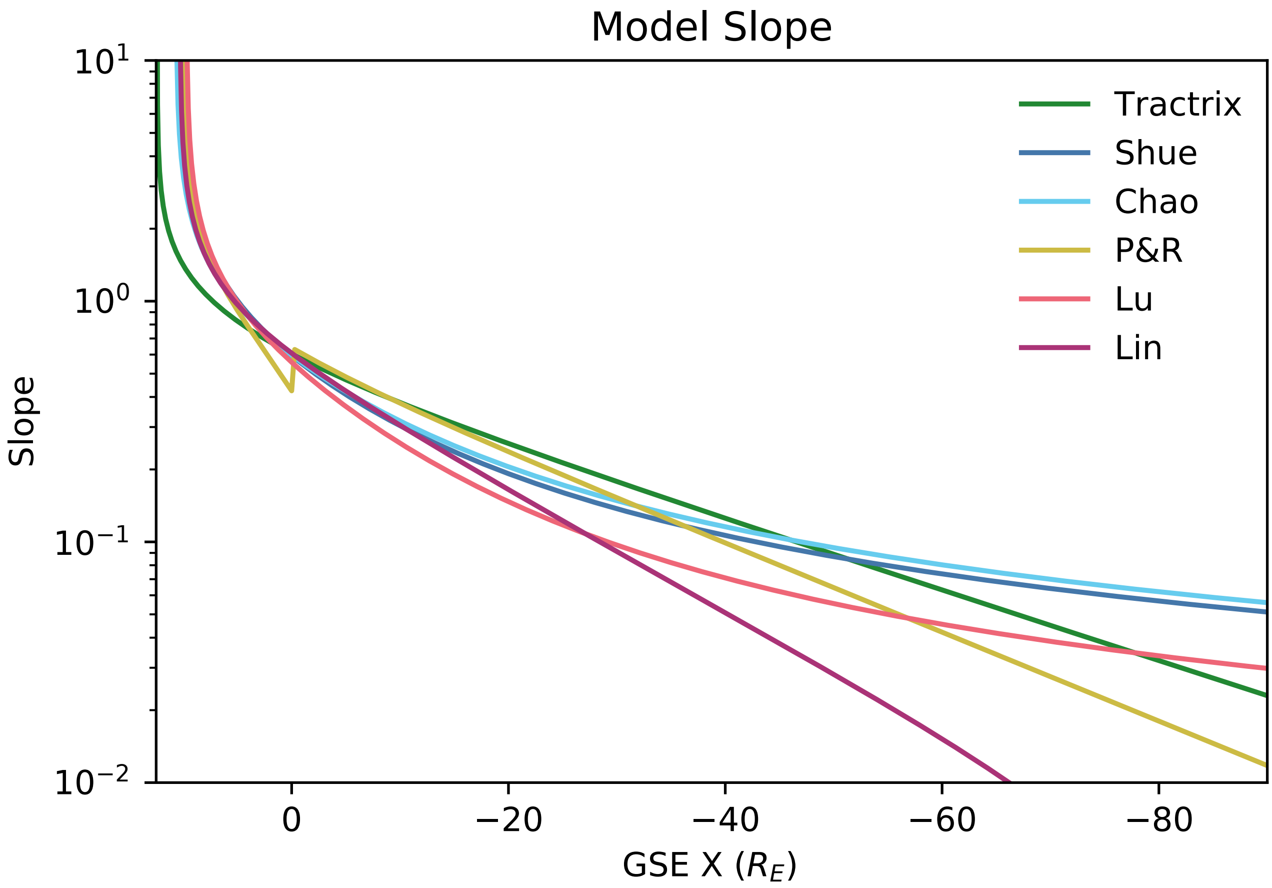


**Figure S2.** Slope of model curves for each model considered in this study. On the nightside, the tractrix has a larger slope than any other model until about $-45R_E$ GSE X, after which models that flare outward strongly (Shue et al. (1998), Chao et al. (2002)) overtake the tractrix's downtail expansion rate. Lin et al. (2010) and Petrinec and Russell (1996) always expand more slowly than the tractrix over reasonable distances downtail. Note the small discontinuity at $-45R_E$ GSE X for Petrinec and Russell (1996) where the model switches from its dayside functional form to its nightside functional form.

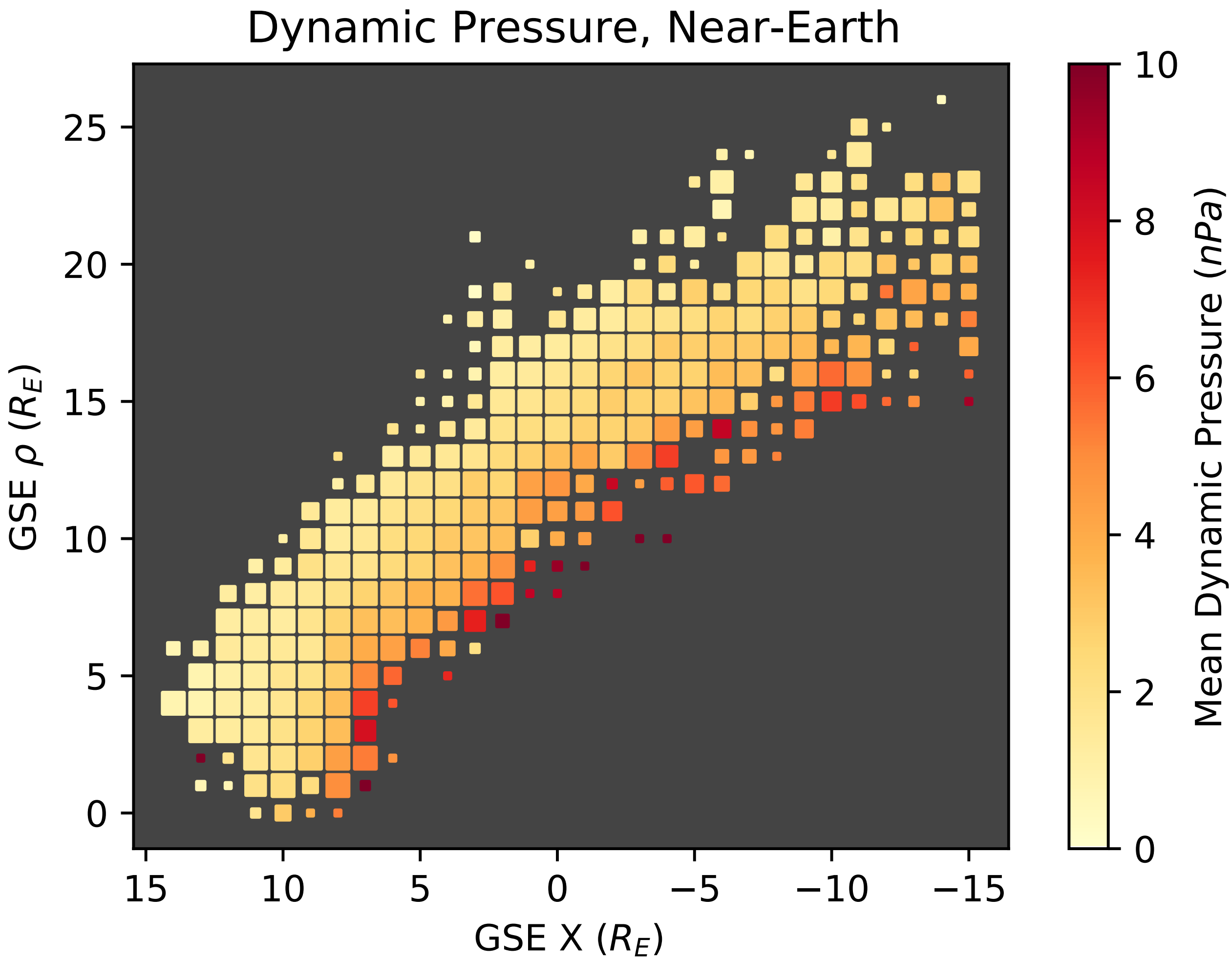


**Figure S3.** Mean solar wind dynamic pressure associated with each crossing, binned using the same scheme as Figure 10 in the main text. Note that the closest crossings to the Earth in the nose happen for extreme solar wind dynamic pressure, likely contributing to Chao et al. (2002) outperforming the tractrix in this region due to its specialized training dataset.

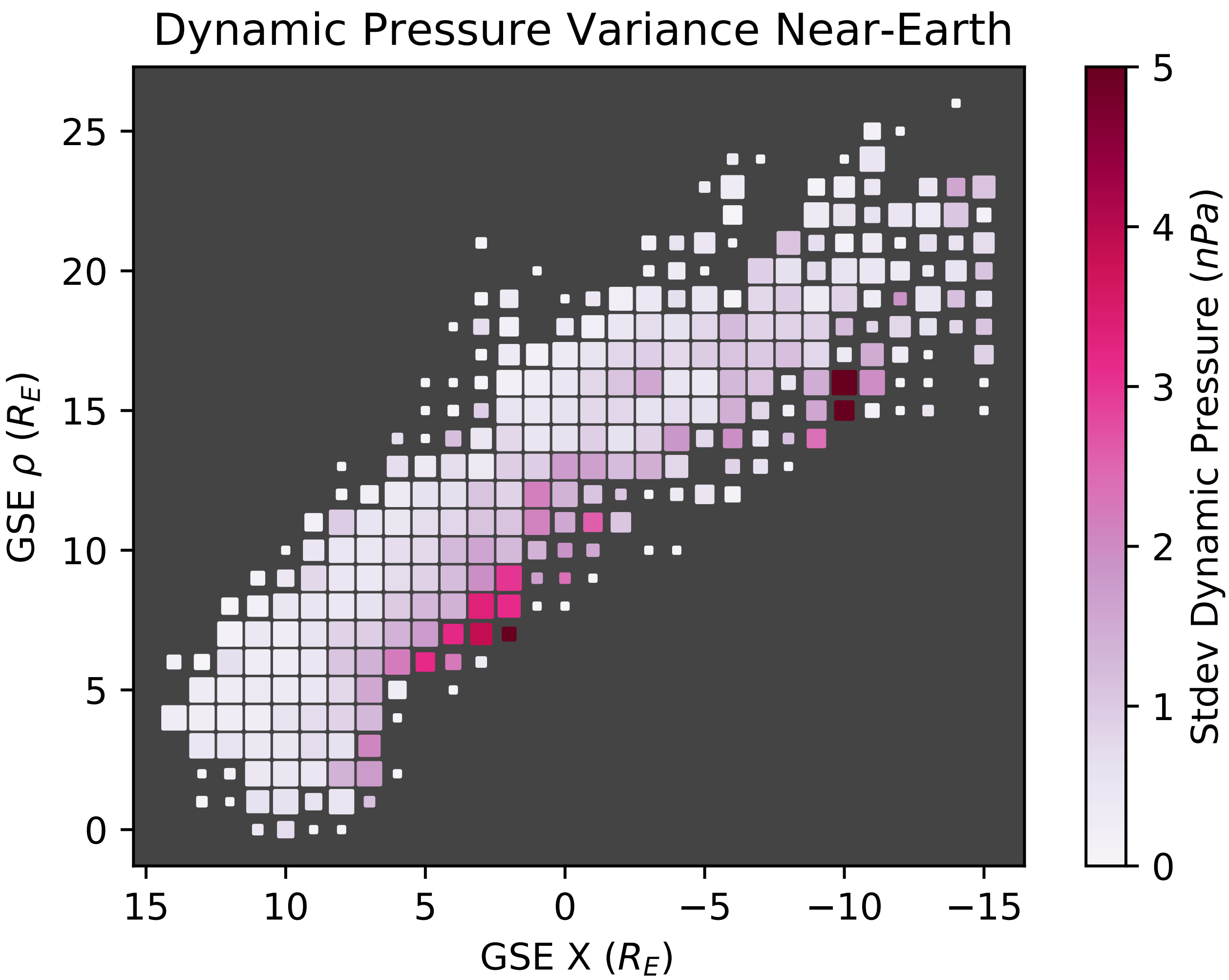


**Figure S4.** Standard deviation of solar wind dynamic pressure associated with each crossing, binned using the same scheme as Figure 10 in the main text.