Assessing Thermospheric Neutral Density Models using GEODYN's Precision Orbit Determination

Zachary C. Waldron[1], Katherine Garcia-Sage[2], Jeffrey P. Thayer[1], Eric K Sutton[3], Vishal Ray[4], David D Rowlands[2], Frank G Lemoine[2], Scott B Luthcke[2], Maria M. Kuznetsova[5], Rebecca Ringuette[2], Lutz Rastaetter[6], and Grant Berland[1]

[1]University of Colorado Boulder
[2]NASA Goddard Space Flight Center
[3]University of Colorado at Boulder
[4]Kayhan Space
[5]NASA Goddard SFC
[6]NASA/GSFC

June 25, 2023

Abstract

This study focuses on utilizing the increasing availability of satellite trajectory data from global navigation satellite system-enabled low-Earth orbiting satellites and their precision orbit determination (POD) solutions to expand and refine thermospheric model validation capabilities. The research introduces an updated interface for the GEODYN-II POD software, leveraging high-precision space geodetic POD to investigate satellite drag and assess density models. This work presents a case study to examine five models (NRLMSIS2.0, DTM2020, JB2008, TIEGCM, and CTIPe) using precise science orbit (PSO) solutions of the Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2). The PSO is used as tracking measurements to construct orbit fits, enabling an evaluation according to each model's ability to redetermine the orbit. Relative in-track deviations, quantified by in-track residuals and root-mean-square errors (RMSe), are treated as proxies for model densities that differ from an unknown true density. The study investigates assumptions related to the treatment of the drag coefficient and leverages them to eliminate bias and effectively scale model density. Assessment results and interpretations are dictated by the timescale at which the scaling occurs. JB2008 requires the least scaling (~-23%) to achieve orbit fits closely matching the PSO within an in-track RMSe of 9 m when scaled over two weeks and 4 m when scaled daily. The remaining models require substantial scaling of the mean density offset (~30-75%) to construct orbit fits that meet the aforementioned RMSe criteria. All models exhibit slight over or under sensitivity to geomagnetic activity according to trends in their 24-hour scaling factors.

Assessing Thermospheric Neutral Density Models using GEODYN's Precision Orbit Determination

Z.C. Waldron[1], K. Garcia-Sage[2], J.P. Thayer[1,5], E.K. Sutton[1], V. Ray[1,3], D.D. Rowlands[4], F.G. Lemoine[4], S.B Luthcke[4], M. Kuznetsova[2], R. Ringuette[6], L. Rastaetter[2], and G.D. Berland[5]

[1]CU Boulder Space Weather Technology, Research, and Education Center
[2]NASA Goddard, Space Weather Laboratory, Community Coordinated Modeling Center
[3]Kayhan Space Corp.
[4]NASA Goddard, Geodesy and Geophysics Laboratory
[5]CU Boulder, Department of Aerospace Engineering Sciences
[6]NASA Goddard, Center for HelioAnalytics

Key Points:

- Precision orbit determination solutions are expanded to study satellite drag and assess upper atmospheric density models using GEODYN.
- A proof-of-concept case study assessment of density models is presented using ICESat-2 precise science orbits and orbit fits
- Assessment results are provided for empirical (MSIS2, DTM2020, JB2008) and physics-based (TIEGCM, CTIPe) models for 14-days in November 2018

Corresponding author: Z.C. Waldron, `zachary.waldron@colorado.edu`

Abstract

This study focuses on utilizing the increasing availability of satellite trajectory data from global navigation satellite system-enabled low-Earth orbiting satellites and their precision orbit determination (POD) solutions to expand and refine thermospheric model validation capabilities. The research introduces an updated interface for the GEODYN-II POD software, leveraging high-precision space geodetic POD to investigate satellite drag and assess density models. This work presents a case study to examine five models (NRLMSIS2.0, DTM2020, JB2008, TIEGCM, and CTIPe) using precise science orbit (PSO) solutions of the Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2). The PSO is used as tracking measurements to construct orbit fits, enabling an evaluation according to each model's ability to redetermine the orbit. Relative in-track deviations, quantified by in-track residuals and root-mean-square errors (RMSe), are treated as proxies for model densities that differ from an unknown true density. The study investigates assumptions related to the treatment of the drag coefficient and leverages them to eliminate bias and effectively scale model density. Assessment results and interpretations are dictated by the timescale at which the scaling occurs. JB2008 requires the least scaling ($\sim -23\%$) to achieve orbit fits closely matching the PSO within an in-track RMSe of 9 m when scaled over two weeks and 4 m when scaled daily. The remaining models require substantial scaling of the mean density offset ($\sim 30{-}75\%$) to construct orbit fits that meet the aforementioned RMSe criteria. All models exhibit slight over or under sensitivity to geomagnetic activity according to trends in their 24-hour scaling factors.

1 Plain Language Summary

This study utilizes the increasing availability of satellite trajectory data from low-Earth orbiting satellites and their precision orbit determination (POD) solutions to expand thermospheric model validation capabilities. We introduce an updated interface for the GEODYN-II POD software to investigate satellite drag and assess density models. This work presents a case study assessment of five models (NRLMSIS2.0, DTM2020, JB2008, TIEGCM, and CTIPe) using precise science orbit (PSO) solutions of the Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2). GEODYN is used to construct orbit fits to the PSO via the five models, enabling an evaluation according to each model's ability to redetermine the orbit. Relative deviations from the PSO, quantified by in-track residuals and root-mean-square errors (RMSe), serve as proxies for model densities that differ from an unknown true density. We investigate and leverage drag coefficient assumptions to eliminate bias and scale model densities. JB2008 requires the least scaling ($\sim -23\%$) to achieve orbit fits closely matching the PSO within an in-track RMSe of 9 m when scaled over two weeks and 4 m when scaled daily. The remaining models require substantial scaling ($\sim 30 - 75\%$) to meet the aforementioned RMSe criteria.

2 Introduction

With the drastic increase in commercial satellite launches, the need to address the challenges posed by satellite drag have come to the forefront of the scientific, operational, and commercial space communities (Muelhaupt et al., 2019; Berger et al., 2020; Thayer et al., 2021; Hejduk & Snow, 2018; Bussy-Virat et al., 2018). Shortly following the 36th launch of SpaceX's Starlink constellation on 3 February 2022, 38 out of 49 satellites were lost due to the impacts of a modest geomagnetic storm that reached G1 intensity earlier that day (Berger et al., 2023; Fang et al., 2022; Hapgood et al., 2022). The satellites were placed into an initial orbit of 210 km after which they were intended to maneuver to an operational altitude of 500 km. While this low altitude plan lent itself to a quick de-orbit in the face of catastrophe, it exposed the satellites to the larger variations and uncertainties in neutral density associated with relatively meager space weather conditions. While this event happened at altitudes well below Starlink's operational orbit, it has served as a potent example to the commercial space community of the need to better model and predict atmospheric drag, which represents the most significant hurdle preventing more accurate determination and prediction of trajectories in LEO.

Precision orbit determination (POD) programs are employed in both operational and research capacities to provide high-fidelity orbit trajectories of LEO satellites. The quality of such trajectories is directly dependent on the ability of a POD's force model to realistically capture the conservative and non-conservative forces impacting a satellite's orbit. Due to advancements in conservative force modeling, the largest source of error preventing more accurate orbit trajectories is now associated with non-conservative forces (Tapley et al., 2005; Reigber et al., 2006; Velicogna & Wahr, 2005). Of these, atmospheric drag is the most variable and uncertain as a consequence of its reliance on modeling the thermospheric neutral mass density (ρ) variations and the satellite drag-coefficient (C_D) (Hejduk & Snow, 2018). The largest source of uncertainty is ρ, but for satellites with complex shapes, C_D can contribute to this uncertainty. Mehta et al. (2022) describes this issue as the interconnectedness of uncertain parameters, an extremely challenging problem to solve for the satellite drag community and one that has significant impact on the assumptions made in this work. The burden for achieving more precise and reliable LEO nowcasting and forecasting largely relies on the ability of thermospheric density models to accurately capture the behavior of neutral density and reliably predict it into the future. Adding to the problem, assessing the performance of density models presents a massive challenge due to the scarcity of data from satellite measurements, and the lack of absolute truth due to the complexity of interconnected uncertainties. This necessitates the community to seek alternative methods to add to the validation method repertoire. The growing prevalence of global navigation satellite system (GNSS)-enabled low-Earth orbiting satellites and their POD solutions represents one such potential data source, and providing methods to take advantage of these datasets will help the community expand and refine model validation capabilities.

POD programs such as the NASA Goddard Space Flight Center's (GSFC) GEODYN II software (henceforth referred to as GEODYN) have been developed within the geodesy scientific community with the above challenges in mind—implementing techniques such as reduced-dynamics paired with extremely high quality tracking measurements from GNSS to mitigate the need for highly accurate non-conservative force models when performing non-predictive orbit determination. Through these means, centimeter level radial accuracy has been demonstrated to produce precise science orbit (PSO) solutions for missions such as the Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2), which orbits at approximately 500 km (Thomas et al., 2021). These techniques—combined with GEODYN's legacy of precise conservative force and measurement modeling, meticulous time systems, and accurate coordinate reference frames—have made the program a top-tier POD tool that is well-positioned to study thermospheric neutral density models and their distinct impacts on the estimation of satellite drag (Luthcke et al., 2003; Zelensky et al., 2010; Lemoine et al., 2016; Loomis et al., 2019). This work aims to provide a method to improve the specification of satellite drag physics and the assessment of neutral density model performance to help the Ionosphere-Thermosphere (IT) community advance model predictions, and consequently improve the accuracy of POD solutions.

This paper presents the development of a modernized Python interface for the GEODYN software, leveraging the high-precision nature of space geodetic POD, but refashioned to study satellite drag and to enable density model assessment. We make use of the well-specified, low-error ICESat-2 PSO to perform a case study assessment of five thermospheric density models, three of which are empirical while the other two are physics-based. The ICESat-2 PSO serves as tracking measurements to POD-based orbit fits in which the drag effects from density models are assessed according to each model's ability to redetermine the orbit. Implications regarding the treatment of the drag coefficient are investigated and discussed. This work reports an initial result using a fixed drag coefficient of $C_D = 2.5$, followed by two methods for debiasing the assessment results using a drag acceleration scaling factor over both a two-week and a daily time interval. Each model's orbit fit contains relative in-track deviations, quantified by in-track residuals and root-mean-square errors from the ICESat-2 PSO, which are treated as proxies for model densities that differ from a true, unspecified density. By developing these methods, we aim to provide the community with the means to take advantage of emerging GNSS-tracked satellite datasets and POD solutions to objectively quantify density model performance. In addition, we hope to address deficiencies in non-conservative force modeling that may currently impede higher quality predictions of LEO trajectories. The presented model assessment results will be parsed into the Community Coordinated Modeling Center's Comprehensive Assessment of Models and Events using Library Tools (CAMEL) framework, for community use.

Section 3 gives the necessary science background needed to understand our methodology. Section 4 details the GEODYN software, provides information regarding the ICESat-2 POD solutions, and offers an overview description of the upper atmospheric density models that are assessed in this work. Section 5 details the methodology, the setup procedure for conducting the model assessment, and the methods for debiasing the assessment results using drag acceleration scaling factors. Section 6 provides the results and discussion of the assessment using ICESat-2 PSO as a case study.

3 Background

The precision of a POD solution relies on the fidelity of the tracking measurement models, the quality of the tracking data, and the ability of the POD force model to capture realistic accelerations acting on the satellite. In general, the force model defines the overall motion of a spacecraft by calculating the sum of all impacting forces, themselves being subdivided into conservative forces which are potential in nature, and non-conservative forces which act to dissipate the satellite's orbital energy. Conservative forces captured by the GEODYN force model include the Earth's static gravity field (geopotential), solid Earth and ocean tides, the effects of dynamic polar motion, the acceleration from time variable gravity, Third-body perturbations (primarily from the Sun and Moon), and contributions from general relativity. Recent improvements in conservative force modeling as well as advances in the internal measurement models have shifted the primary source of error in POD solutions to the non-conservative forces (Luthcke et al., 2006; Loomis et al., 2019; Reigber et al., 2006). The non-conservative forces modeled in GEODYN are atmospheric drag, solar radiation pressure (SRP), and Earth radiation pressure (ERP). As altitude decreases in the LEO regime, atmospheric drag increasingly becomes the largest non-gravitational force acting on satellites. In addition, the drag force's dependence on the upper atmospheric neutral mass density makes it the most error-bound perturbing force (Hejduk & Snow, 2018). While force model errors can be circumvented via reduced-dynamics and high-quality tracking measurements, this technique is limited in its application for the eventual goal of orbit prediction, which requires an improved, more-realistic force model (Tapley et al., 2004; Luthcke et al., 2019).

The drag force acting on a satellite of mass m_{sat} is proportional to the atmospheric neutral mass density ρ, the drag coefficient C_D, the projected area perpendicular to the flow direction A_{sat}, and the velocity of the satellite relative to the atmosphere $\vec{V}_{rel}$. The drag acceleration $\vec{a}_D$ due to the drag force per unit mass acting on a satellite is given in Equation 1 as

$$\vec{a}_D = -\frac{1}{2}\,\rho\, C_D\, \frac{A_{sat}}{m_{sat}}\, V_{rel}^2 \frac{\vec{V}_{rel}}{V_{rel}} \tag{1}$$

Physically, the total drag force acting on a satellite surface is given by the force due to incident atmospheric particles impacting the surface combined with the force from scattered particles departing from the surface. These effects are represented by the drag coefficient C_D, which depends on a satellite's geometry and orientation, the material and surface temperature of the spacecraft, the local atmospheric composition, and gas-surface interactions and other effects (Bernstein & Pilinski, 2022). In the context of spacecraft dynamics, the C_D is generally characterized as either fixed, fitted, or physical. Fixed C_D uses a predetermined value that does not change. Fitted C_D is derived using some form of a fitting or filtering process and is typically updated over time (every few hours or orbits). Physical C_D is computed by modeling the momentum and energy exchange between the flow-field particles and the satellite (see Mehta et al. (2022) for more details). If not physically calculated, C_D's presence in Equation 1 may be thought of as a scaling factor that effectively serves to average out errors in the atmospheric density model and gas-surface interactions. In its base state, GEODYN can use either a fitted or fixed C_D. In the fitted case C_D is an adjustable parameter that accounts for mismodeled physics and for uncertainties in ρ associated with the upper atmospheric density model.

Earth's upper atmosphere is driven by a broad range of external energy inputs, leading to complex thermal, electromagnetic, and chemical processes that result in a thermospheric neutral mass density ρ that is highly dynamic and whose variability is difficult to specify (Emmert, 2015). Upper-atmospheric density models are employed within POD force models to represent the complex behavior of ρ when calculating the force of satellite drag acting on a spacecraft, directly or indirectly through C_D. The three types of density models most commonly used by upper atmospheric communities are semi-empirical, physics-based, and data assimilative models. The simple yet effective semi-empirical models are most commonly employed in POD force models since they offer excellent climatological pictures of upper atmospheric variability and are computationally inexpensive. Physics-based models are more complex, taking the form of general circulation models which solve the first-principle equations that govern the coupled thermosphere-ionosphere system. They are not typically used in POD geodetic settings due to the computational expense. A data assimilative technique can be used to calibrate modeled density and has given rise to data assimilative (also referred to as dynamically calibrated) models. These combine analyses from a multitude of space objects to produce corrections to empirical (and occasionally physics-based) thermospheric models. The most prominent example of assimilative thermospheric density models is the United States Space Force, High Accuracy Satellite Drag Model (HASDM) (Storz et al., 2005). It is a common practice in the IT modeling community to compare model performances against HASDM outputs since it performs real-time calibration using ~ 75 space objects.

Different models, and even model types, have varying degrees of performance under specified conditions. Individual model performances are known to depend greatly on the solar flux and geomagnetic conditions that drive them, and their respective strengths make some models better qualified for some scenarios than others. Semi-empirical models are often computationally fast and accurate for climatological uses, but their ability to accurately project into the future is closely tied to the fidelity of their drivers. Physics models offer great potential for forecasting, but lack the accuracy of semi-empirical models in near real-time scenarios (Shim et al., 2014; Sutton, 2018). The vast range in model performances makes the evaluation of models a critical goal for upper atmospheric science and satellite drag communities. The scarcity, and coupled uncertainty (via C_D uncertainty) of thermospheric density measurements makes this a significant challenge. The most common method for objectively quantifying a density model's performance is to compare the sampled model outputs against satellite measurements, e.g. see Walterscheid et al. (2023)—usually in the form of accelerometer-derived densities from the Challenging Minisatellite Payload (CHAMP), Gravity Recovery and Climate Experiment (GRACE) or Gravity Field and Steady-State Ocean Circulation (GOCE) missions (Bruinsma et al., 2004; Sutton et al., 2005; Doornbos et al., 2010; Mehta et al., 2017).

In a series of papers motivated to provide community organization for conducting model comparison and evaluation, Bruinsma et al. (2017, 2018, 2021) provide commonalities for inter-model scoring. They report on chosen observed density datasets, time periods of interest, and provide a scoring metric in the form of the mean, standard deviation and root mean square error (RMSe) of the observation-to-model density ratios. He et al. (2018) similarly presents an assessment of several semi-empirical thermosphere models, focusing on their ability to reproduce spatial variations and capture complex features in thermosphere mass density. Shim et al. (2014) provides a systematic evaluation of thermospheric and ionospheric models, quantifying model performance using four skill scores calculated as functions of geomagnetic activity and geographic latitude: RMS error, prediction efficiency, ratio of maximum-to-minimum, and ratio of maximum amplitude. Thayer et al. (2023) investigates the use of the day-to-night density ratio as a metric for representing the atmosphere's response to large scale perturbations (i.e. the transition from solar maximum to solar minimum), providing inter-model and model-to-observation comparisons, and unearthing discrepancies that are not observed between models and observations when viewed using more common metrics. Each of these reports makes use of the accelerometer-derived density data sets to objectively quantify model performance.

Through this work, we aim to contribute an additional method to the community in which accurately developed and well-honed POD tools can be leveraged for assessing density model performance. For the purposes of this paper, we make a distinction when referring to the different stages of model assessment. We use the term "assessment" to refer more generally to methods and results that offer insight into model performance. "Verification" refers to using other well-specified methods and datasets to confirm the fidelity of our methods and results. "Validation" refers to the act of objectively quantifying modeled densities against observed/derived values. This paper offers a verification of our method and results by comparing against the HASDM densities, and provides an example performance assessment using two-weeks of the ICESat-2 PSO as a case study. A more formal validation scheme is the eventual goal of this work, however, this requires additional considerations and is a source of continuing effort.

4 Program and Data Descriptions

4.1 GEODYN and the Pygeodyn Wrapper

The GEODYN-II program is a precision orbit determination and parameter estimation tool that has been used on every NASA geodetic Earth and planetary altimeter mission since 1985. The program is used extensively for orbit determination, geodetic parameter estimation, tracking instrument calibration, satellite orbit prediction, as well as for many other applied research studies in satellite geodesy (Pavlis et al., 2019; Luthcke et al., 2019). GEODYN is capable of ingesting essentially all types of tracking measurements, the most common of which include observations from global navigation satellite systems (GNSS) and satellite laser ranging (SLR), as well as post-processed orbits in the form of orbit trajectories or precisely converted elements (PCE) (Pavlis et al., 2019; Lyon et al., 2004). GEODYN performs orbit propagation using Cowell's method of numerical integration, and performs data-reduction utilizing a Bayesian least-squares batch estimation process to optimally estimate parameters by minimizing the residuals between tracking data and orbit propagations (see Vallado (2013) for more information). GEODYN's long history in geodetic applications has ensured the development of very precise conservative force and measurement models, as well as accurate time systems and coordinate reference frames, making the program a top-tier POD tool. With this understanding, the errors found in the observed residuals between tracking data and determined orbit are more related to uncertainties in the satellite specific non-conservative force models, rather than being related to the quality of measurement modeling or orbit determination methods and tools. In the lower register of LEO, where atmospheric drag dominates, the observation residuals can provide valuable information on the drag model errors.

Pygeodyn is an internally-developed Python-based wrapper meant to offer improved user access to the FORTRAN-based GEODYN software. Pygeodyn offers users a streamlined and simplified tool to navigate the complex steps for modifying, controlling, running, and reading the various data sets and files that compose the GEODYN program. The main portion of GEODYN II is composed of two sequenced programs: GEODYN-IIS, a scheduling program and GEODYN-IIE, an execution program stage. The scheduling program reads and organizes input data, ancillary data files, and the user's setup options. The execution program then integrates the satellite trajectory and applies the selected models, performs orbit determination to provide computed observables, and uses the least squares scheme, along with any measured observables, to provide solutions for updated orbits as well as any requested geophysical parameters. The two stages communicate via a series of binary files which are output from the scheduling program and fed into the execution program. Historically, adding atmospheric density models to GEODYN required modification to IIS as well as subsequent data tracking and modification to IIE, a series of complications that have been circumvented with our Pygeodyn tool. Pygeodyn gives the ability to switch between different atmospheric density models that have been connected to GEODYN-IIE without the need to modify GEODYN-IIS, simplifying the user experience for adding and selecting the models. Programming in Python has also afforded Pygeodyn the ability to interface with the NASA Goddard Community Coordinated Modeling Center's (CCMC) Kamodo API (Ringuette et al., 2023), granting access to their sophisticated model readers and allowing Pygeodyn to connect physics-based density model outputs to the POD scheme.

4.2 ICESat-2 PSO Solutions as Tracking Data

ICESat-2 flies in a near-circular, near-polar, low-Earth orbit at ∼496 km altitude and an orbital period of 94.22 min. Details of the orbital parameters are reported in Luthcke et al. (2019). The ICESat-2 PSO (i.e., the science quality POD solutions), and their corresponding setup files, are provided by the Geodesy and Geophysics Laboratory within NASA/GSFC, who maintain the GEODYN program and provide science quality POD for many NASA missions. ICESat-2 is an excellent platform for orbital drag-based model assessment because of its science requirements to have such high quality orbit solutions, as well as stable attitude specifications. The ICESat-2 PSO is reported by Thomas et al. (2021) as having a radial orbit accuracy of just below 1.5 cm over a 24-hour orbit solution—performing better than the mission requirement of 3 cm. These orbit solutions are generated through the reduction of GNSS double-difference carrier phase observable residuals, and independently assessed using SLR measurement residual analysis. Technical details regarding the construction and analysis plan for the ICESat-2 PSO can be found in Luthcke et al. (2019).

The centimeter-level orbit of the PSO data is achieved using the previously mentioned reduced-dynamics technique in which GEODYN solves for empirical acceleration parameters that describe the difference between the actual positions, i.e. those derived from the GPS tracking measurements, and the positions that are calculated by the program's physical force models and satellite propagator. The PSO data includes estimations of along-track and cross-track empirical accelerations every quarter of an orbit, applying a nearest neighbor covariance constraint. With the use of reduced-dynamic empirical accelerations, it is possible to compensate for errors associated with using the MSIS86 model to calculate the effects of atmospheric drag. Luthcke et al. (2019) notes that even though a reduced-dynamic approach is commonly employed by the geodesy community to overcome any inadequacies in a force model, the technique relies on an orbit solution that has already attained sufficient radial accuracy through the use of a high-quality physical force model. Dense tracking measurements and the reduced-dynamic technique do not obviate the use of accurate orbit modeling, and improvements in the orbit fit will be realized when the force models are improved. We also note that while MSIS86 was used to estimate the drag accelerations in the ICESat-2 PSO, due to them being combined with additional empirical accelerations in the along-track and cross-track directions no related bias is found that favors the MSIS series of models in the assessments reported in this work.

This work uses the ICESat-2 PSO as tracking measurement input to a data-reduction run of GEODYN—the goal being to assess the ability of each selected density model to re-determine the orbit of ICESat-2. A data-reduction run in GEODYN is one in which orbit parameters (i.e., initial conditions) and optionally geophysical parameters (such as gravitational coefficients or the drag coefficient) are adjusted to minimize residuals and provide an improved solution. This data reduction is computed over an orbital arc, a set time period for which continuous tracking data is available. The term "orbit fit" refers to the outputs of GEODYN runs in which the ICESat-2 PSO is the tracking data type and respective density models are used to iteratively re-determine the orbit.

The following capabilities for density model assessment are enabled by using GEODYN to construct orbit fits from ICESat-2 orbit solutions:

1. Leverage GEODYN's high fidelity physical force models which have been honed by the program's long legacy in space geodesy.
2. Perform data-reduction runs in which we compare the relative ability of each atmospheric density model to re-determine the orbit of ICESat-2 given the isolated satellite drag effects.
3. Control the POD and force model parameters such that for each respective run, the only relative variable impacting the overall fit of the orbit solution for a given arc is the atmospheric density model used to estimate the drag term.
4. Control for relative errors between runs associated with an unknown drag coefficient by using a realistic fixed value of $C_D = 2.5$. This value is determined by physically calculating C_D using the Diffuse Reflection with Incomplete Accommodation (DRIA) method along the orbit of ICESat-2, as is described in Section 5.

4.3 Model Descriptions

This section provides a brief overview of the atmospheric density models that are used for verification and assessment. Table 1 lists the models, providing the Model ID used for referencing in this paper, the full name and version number, the run conditions based on the drivers, and the models' spatial and temporal resolutions. The authors acknowledge that while there are a number of ways to improve a density model's outputs at runtime (see Sutton (2018); Shim et al. (2014)), the outputs used in this work are intended to reflect typical community use, with each model being run according to the developer's operational instructions. Additional information regarding each model can be found in the references provided in the second column of Table1.

We provide a verification using SET-HASDM, a data-assimilative model, and assessment results for MSIS2, JB2008, DTM2020, TIEGCM, and CTIPe. The semi-empirical models (MSIS2, JB2008, DTM2020) are interfaced directly into GEODYN's FORTRAN-based source code. The physics-based models (TIEGCM and CTIPe) are interfaced to GEODYN via the CCMC's Kamodo program which reads and interpolates the model output files. These interpolated outputs are connected to GEODYN through the Pygeodyn wrapper using an orbit cloud interpolation technique which is detailed in Appendix C. In addition, physics-based models whose maximum altitude is below the orbit altitude of ICESat-2 include a diffusive equilibrium extrapolation of the neutral densities (see Chapter 10 of Schunk and Nagy (2009)). The use of Kamodo makes the analysis techniques in this paper easily extensible to additional models. Any thermospheric model that is supported by Kamodo, with the appropriate diffusive equilibrium extrapolation, can be added to this and similar analyses in the future.

Models Assessed in this Report

Model ID	Full name/version	Drivers, (solar $\mid$ geomagnetic)	Resolution, (spatial $\mid$ time)
Semi-Empirical			
MSIS2	Naval Research Laboratory Mass Spectrometer and Incoherent Scatter (NRLMSIS 2.0), (Emmert et al., 2021)	$F_{10.7} \mid Ap$	
JB2008	Jacchia-Bowman 2008, (Bowman et al., 2008)	$F_{10.7}$, $S_{10.7}$, $M_{10.7}$, $Y_{10.7} \mid Ap$, Dst	
DTM2020	Drag Temperature Model 2020, operational mode (Bruinsma & Boniface, 2021)	$F_{10.7}$ and K_p	
Physics-Based			
TIEGCM	National Center for Atmospheric Research (NCAR) Thermosphere-Ionosphere-Electrodynamics General Circulation Model (TIEGCM; version 2.0), (Richmond et al., 1992; Qian et al., 2014; Sutton et al., 2015)	$F_{10.7}$, EUVAC proxy model (Solomon & Qian, 2005) $\mid$ Kp via the Heelis model (Heelis et al., 1982),	$5° \times 5°$, vertically specified by logarithmic pressure surfaces in half-scale height increments from ∼97 km to ∼800 km $\mid$ 1 minute time step, hourly output
CTIPe	Coupled Thermosphere Ionosphere Plasmasphere Electrodynamics (CTIPe; version 4.1) (G. Millward et al., 1996; G. H. Millward et al., 2001)	$F_{10.7}$, solar wind inputs via magnetic field, velocity, and density measurements from Advanced Composition Explorer (ACE) $\mid$ Kp, hemispheric power index from NOAA Polar Orbiting Environmental Satellite (POES)	2° lat. $\times$ 18° lon., vertically specified by 15 logarithmic pressure surfaces from ∼80 km to ∼500 km $\mid$ 1 minute time step
Data-Assimilative			
SET-HASDM	Space Environment Technologies (SET) HASDM Database (Tobiska et al., 2021), which is derived from the US Space Force operational archive (Storz et al., 2005)		10° lat. $\times$ 15° lon., 25 km altitude steps from 175 to 825 km. $\mid$ 3 hour time step

Table 1. A summary of the thermospheric density models used in this assessment.

5 Methodology

5.1 Setup for ICESat-2 Case Study

This method uses a satellite's PSO as tracking measurements to construct dynamic POD-based orbit fits from different density models. The dynamic POD technique uses a batch least-squares approach to iteratively reduce errors between the propagating orbit fit and the ingested PSO—GEODYN refers to this as data-reduction mode. The initial conditions, and any other adjustable parameters, are iteratively estimated and updated to refine the orbit fit until it consistently reaches a convergence threshold. The remaining errors that persist between the PSO and a given model's orbit fit are understood to be primarily due to atmospheric drag effects from the respective density model. This understanding is leveraged to investigate density model performance through the assessments that are presented in this paper. Figure 1 provides a visualization showing connections between the high-level datasets and processes. While the true density along the ICESat-2 orbit remains unknown, each model's orbit fit contains in-track deviations from the ICESat-2 PSO, which are treated as proxies for model density deviations from the true density.

The GEODYN run setup that is used to construct the POD-based orbit fits is kept as similar as possible to the setup used by the team at NASA-GGL to produce the ICESat-2 PSO—meaning we modify only what is necessary to use PSO as the tracking measurement type, and to control the procedure such that drag is the only independent variable in each model's run. An extended overview of GEODYN's setup and force model parameters for the model assessment runs is provided in the appendix in Table B1, with only the most impactful considerations being discussed here. In addition to each orbit fit using the same background force models, the ICESat-2 external attitude information is also utilized to properly orient the spacecraft body and the solar array. The orbit fits are split into 24-hour, consecutive daily arcs. The arc length can theoretically be much shorter; however, orbit errors related to force model perturbations (i.e., drag) require propagation time to accumulate. An arc length on the order of 1-2 orbital periods may not depict substantial trajectory deviations in the residuals, making 24-hour arcs a balanced choice to demonstrate this assessment method. In theory, reducing the arc lengths would provide more RMSe values over shorter times and would offer higher temporal resolution towards understanding model performance, but at the cost of having accumulated less orbital error from the density model in the shorter propagation time. The choice of arc length and its ramifications on assessment results continues to be an area of study related to this work.

Other non-conservative forces that must be considered in addition to atmospheric drag are SRP and ERP which are both calculated by GEODYN according to the descriptions shown in Table B1, and the references therein. Using GEODYN and its high-fidelity force model ensures that the estimated SRP and ERP accelerations are more precise than what would be modeled by a standard satellite flythrough scheme. For each density model's orbit fit the acceleration due to drag will vary according to the error in the respective model, while the contributions from SRP and ERP will remain consistent for each arc across each model run. For the orbit fits presented in this work, the variations of SRP and ERP were found to be small relative to the variable effects of drag. This being said, the magnitude of the SRP acceleration is often on par with that of the drag acceleration at the ICESat-2's altitude. Errors related to mismodeling non-drag forces can potentially be transferred into the residuals, and as a result this method is presented as a relative assessment between controlled model runs rather than an absolute validation of performance.

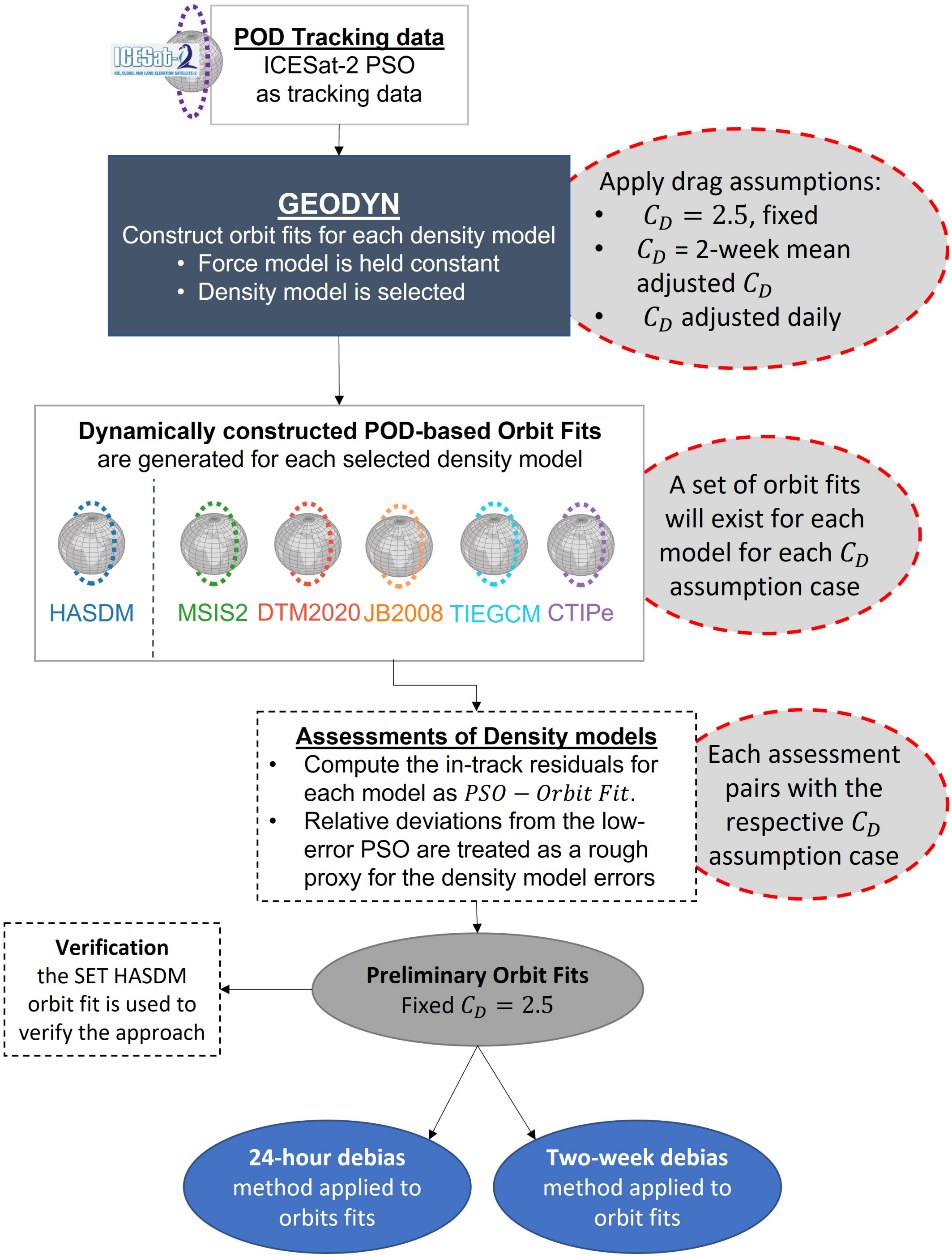


Figure 1. A flowchart visualizing the assessment process and how the datasets and POD methods fit together for model assessment.

5.2 Assessment Procedures

The interconnected, uncertain nature of C_D and ρ makes the absolute determination and assessment of either value a very complex problem that is still an active field of research within this community. In this case study, the assessment procedure is split into three subsequent orbit fit methods, each based on assumptions made to characterize C_D when calculating the drag acceleration for each model during the orbit fit procedure. In a preliminary orbit fit method, explicit biases are identified via orbit fits constructed using a fixed C_D that is held constant across all models. The constant fixed C_D is chosen to be a physically realistic value of $C_D = 2.5$ based on the average result calculated from the Diffuse Reflection with Incomplete Accommodation (DRIA) model along the ICESat-2 orbit. This demonstrates that the method is sufficiently sensitive to recognize differences in the drag effects between the models and provides an understanding of each model's approximate mean density offset relative to the ICESat-2 PSO In an alternative orbit fitting method, GEODYN's parameter estimation procedure is used to adjust the C_D for every 24-hour arc for the two-week period from the a priori estimate of $C_D = 2.5$. The mean-adjusted C_D over the two-weeks is then used as a fixed, model specific value that is constant for the time-period. This provides a fixed, unique C_D for each density model that effectively scales the density over the two weeks to account for each model's mean density offset and examine the model response to solar and geophysical dynamics. In a final orbit fitting method, the daily C_D adjustments are used without a two-week averaging. The residuals using these daily, model-specific C_D adjustments provide an assessment of model performance on time periods less than a day.

The preliminary orbit fits use a fixed C_D of 2.5 that is constant with respect to each model. This enables direct model comparison, but subjects an assessment of the density models to explicit biases depending on each model's density offset relative to this C_D value. Each model's sampled densities along the ICESat-2 orbit have an overall mean-density offset relative to each other. Fixing the C_D to a specific value will cause a particular offset amount to be favored. For instance, $C_D = 3.5$ will produce favorable orbit fits for models that trend a lower density, whereas $C_D = 1$ would favor models that trend towards higher densities. Due to these circumstances, the DRIA model is independently used to calculate a physically realistic value of the ICESat-2's C_D along its orbit. DRIA is a relatively simple, computationally fast model for capturing the gas-surface interactions between the upper atmosphere and a spacecraft. In the DRIA model, particles are always reflected with a diffuse angular distribution, but their energy exchange with the surface varies depending on the value of the energy accommodation coefficient α. This work uses the Sentman's closed-form solutions for the DRIA model as depicted in Equation 12 of Walker et al. (2014). The energy accommodation is assumed to be fixed at $\alpha = 0.89$—a tenuous assumption based on the limited empirical data for α near 500 km during solar minimum (Pilinski et al. (2010); Pilinski (2008)). This α value is likely higher than is realistic for this altitude and solar flux, therefore providing a lower limit for what a physically realistic drag coefficient might be; however, given the complex changes in atmospheric structure that occur in this altitudinal regime, this empirical value is still the most representative until further observations can be made. There are other physical C_D models that could be used instead of DRIA, but choosing and assessing the C_D models quickly expands beyond the scope of this work. For the sake of being able to conduct a model assessment as a proof-of-concept in this case study, this assumption is made with the intention to improve the treatment of C_D in future efforts. In future work, we aim to address this issue by implementing a physical satellite gas-surface interaction method to calculate the time-dependent drag coefficient, but even this will have associated assumptions and caveats. Constructing orbit fits with a fixed, common C_D of 2.5 for each model represents the type of method that is possible without being able to model the physical drag coefficient or without GEODYN's capability to adjust the parameter. This adjustment procedure was performed for different a priori C_D values and found that the final adjusted C_D for each model was consistently the same.

The 24-hour debiasing method uses GEODYN's parameter estimation capabilities to determine a daily fitted value of the C_D that accounts for accumulated errors from the force model over the 24-hour arc—the most prominent of which being due to density uncertainty. In the field of space geodetic POD, C_D is often adjusted in conjunction with reduced-dynamic empirical accelerations to account for disagreement between the observed accelerations from tracking measurements, and calculated accelerations from uncertainties in the drag force model. This technique is used to get very low error, precise orbit solutions, but limits the ability to distinguish errors that are specific to drag or the density models. By allowing only the C_D to adjust and match the orbit fit's modeled accelerations with the PSO observation, density errors over the 24-hour period are incorporated into the adjustment. A density model that is found to be over-/under-estimating the density, will have a C_D that is adjusted to be smaller/larger in a non-physical way—effectively using C_D as a scaling term between the PSO observation and uncertainty in the density model orbit fits. In practice, the C_D also absorbs any errors from mismodeled forces, but these are held constant in the model-to-model comparison. Each model is given an a priori estimate of $C_D = 2.5$ at the start of the 24-hour arc, which is allowed to adjust within a standard deviation of 10. The drag coefficient fitting occurs concurrently with the iterative orbit fit routine. Due to this non-physical use of C_D to effectively debias the density, the term "drag acceleration scaling factor" is adopted. The 24-hour scaling factor for each model (m) can be calculated for each arc (i) as,

$$f_{24,m,i} = C_{D,adj,m,i}/2.5 \tag{2}$$

The two-week debiasing method acts as a combination of the previous two methods. The C_D adjustments for each model are averaged over the two-week period to provide a mean adjusted C_D. Each model's orbit fit is then re-determined using the mean adjusted C_D for each respective model as the fixed value for the two-week period. This assessment permits a scaling of the density models over an extended period of time, highlighting errors in the orbit fits that are due to variations that take place on a longer time scale than 24-hours. This method is also motivated by the need to provide a scoring metric for each density model that can be parsed into the CCMC's CAMEL model validation infrastructure. While the $C_D = 2.5$ case is dominated by the model biases and the 24-hour debiased case demonstrates a method to debias daily densities, the two-week debiased case quantifies the ability of the models to capture dynamics caused by geomagnetic and solar activity over a more prolonged time period. This is a method that could be used in the future to assess model performance during individual stormtime periods.

5.3 Assessment Metrics

Using a PSO as tracking data makes use of GEODYN's data-reduction mode combined with a dynamic technique for estimating the orbit of a satellite. This technique uses the trajectory input to estimate updates to the initial conditions which define the motion of the satellite, thus refining the orbit. The orbit residuals obtained in this setup are the absolute differences between the PSO and each density model's orbit fit. Since other force model parameters are held constant between each density model's run, the inter-comparison of the residuals contains information primarily corresponding to relative errors in each density model's ability to replicate the drag effects seen in the ICESat-2 PSO.

To best observe satellite drag effects, all output orbits are transformed from the J2000, geocentric inertial reference system to the NTW, orbit-aligned satellite coordinate system (Vallado, 2013). This system is composed of an in-track component $\hat{T}$ that is parallel to the velocity vector, a normal component $\hat{N}$ that is perpendicular to the velocity and nominally in the radial direction, and a cross-track component $\hat{W}$ that is normal to the orbit plane and completes the right-hand coordinate frame. The in-track component $\hat{T}$ is parallel to the velocity vector direction and contains any indication that the spacecraft's trajectory has changed since orbital energy dissipations from drag will impact in the velocity direction. Information regarding this transformation as well as supporting coordinate frame details can be found in Appendix A.

For any given arc, y_o is defined to be a component of the orbit from the PSO dataset in the NTW frame, and y_m to be the orbit fit for each density model m. The residuals for each component of the orbit and orbit fit are then calculated (in terms of the in-track component) as,

$$R_{m,T} \quad = \quad y_{o,T} - y_{m,T} \tag{3}$$

The root-mean-square error (RMSe) of the residuals represents the square root of the variance of the absolute difference in the two orbits, indicating how well the density model's orbit fit matches the PSO for that arc. For the in-track component, this is computed for every i^{th} time step of an arc with n time steps as,

$$\mathrm{RMSe}_{m,T} = \sqrt{\frac{1}{n}\sum_{i}^{n}\left(y_{o,T,i} - y_{m,T,i}\right)^2} \tag{4}$$

Figure 2 provides an example of the observation residuals for sample orbits over four,

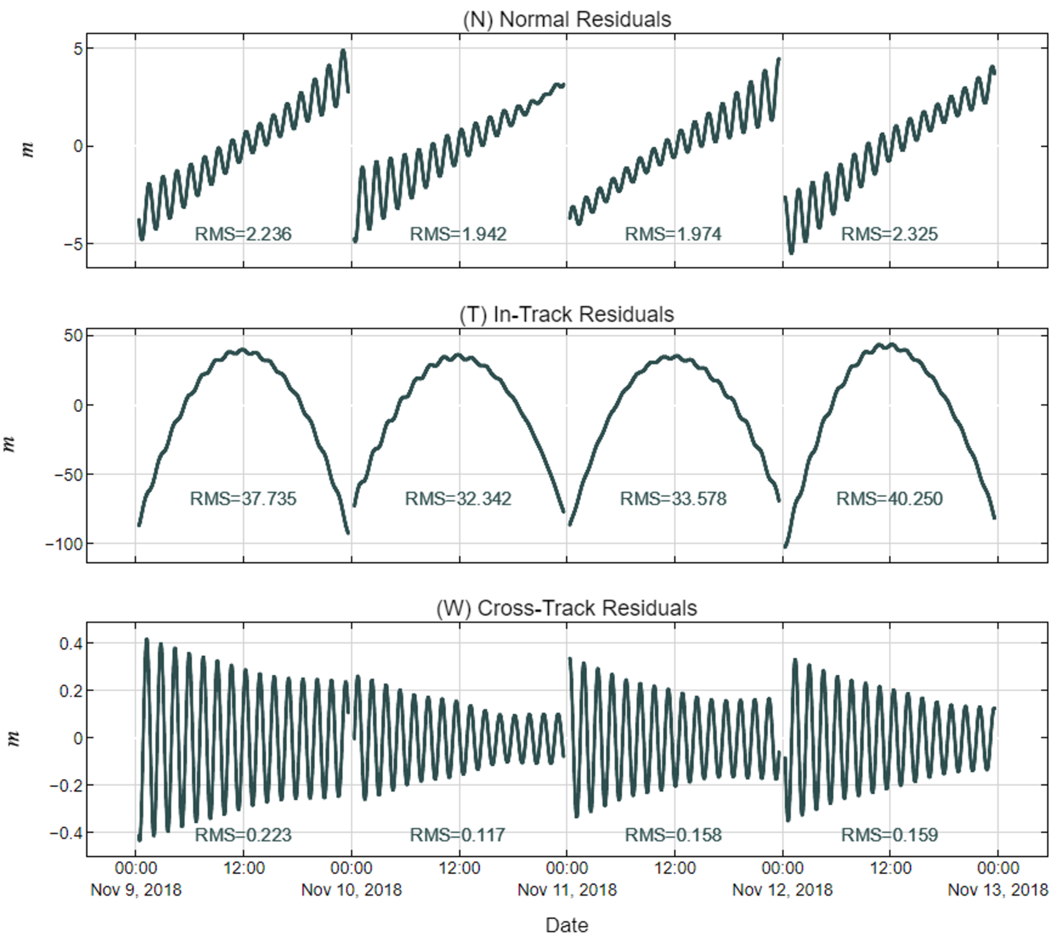


Figure 2. Depicted here are example observational residuals for each component of the NTW system across four 24-hour arcs using the MSIS2 model. The RMSe for each arc/component is given under each curve, showing that the majority of the residual variance (i.e. the orbit error due to drag) is contained within the in-track direction. Note that the vertical scale is different for each plot.

24-hour arcs. The normal $\hat{N}$ (top panel) and cross-track $\hat{W}$ (bottom panel) residuals are included in this figure only to demonstrate that the in-track component contains the majority of the variance associated with residuals in this reference frame. The RMSe values for each component and arc are included as an overlay.

During the POD process, GEODYN iteratively minimizes the discrepancies between the observed orbit (i.e., PSO) and computed orbit (i.e., orbit fits from a density model) across the entire 24-hour arc by adjusting the initial conditions to converge towards a computed trajectory. Since the minimization occurs across the entire arc, the resulting residuals take on the non-linear shapes shown in Figure 2. On a given arc and when comparing the resulting orbit fits from each density model, the only variable that has been permitted to impact each orbit fit's performance relative to the PSO is the drag effects from the selected density model. Therefore, we reason that the relative differences in the residuals for each orbit fit is indicative of density model performance. Other potential errors from mismodeled physics may persist in the residuals, but they are held constant between each model run and will impact the orbit fits consistently. If we run the same arc using the same force model and conditions, but only change the density model used to calculate satellite drag accelerations, the residuals will contain the errors related to the program attempting to reconcile errors in the density model. The RMSe for each orbit fit represents a single value for how well the program can reconcile each model's errors in density over the entire 24-hour arc.

As described in Section 4.2, the ICESat-2 PSO has been shown to have a radial orbit accuracy of below 1.5 cm, generated through the reduction of GNSS double-difference carrier phase observable residuals and independently assessed using SLR measurement residual analysis (Thomas et al., 2021). The precision of the orbit solutions were also verified in all three components using orbit overlap analysis. Given this, relative deviations from the low-error PSO are treated as a rough proxy for the density model errors relative to some unknown true density. The true density value is obscured by the various interconnected unknowns of C_D, SRP, and ERP and therefore remains unspecified. Over the course of a single arc, drag forces from each density model dissipate the satellite's orbital energy at distinct rates, resulting in drag accelerations that are either greater or less than what is represented by the in-track position of the PSO. A strongly negative in-track residual indicates a modeled density that is larger than truth, while a strongly positive in-track residual indicates a modeled density that is smaller than truth. Additional details regarding the shape of the in-track residuals and the relationship between in-track position of the PSO and orbit fits and the density can be found in Figure A2 of Appendix A.

The RMSe is the standard deviation of the residuals and serves as a measure of the difference between a respective orbit fit and the PSO over a single whole arc. Theoretically, an in-track RMSe of zero would mean no difference between an orbit fit and the PSO, indicating near-perfect agreement on average between the modeled density and the POD-based true density across the 24-hour arc. In this setup, perfect agreement for any model is unlikely since the residuals may additionally contain errors related to mismodeled forces, as well as bias/offsets related to fixing the C_D to a common value for all models. A further limitation of the metric is that the RMSe lacks information regarding timescales less than the arc length, and is unsigned, meaning it does not indicate if the modeled density is above or below the truth for a given arc. For these reasons, the in-track residuals and their respective RMSe values are assessed in conjunction with each other.

6 Results and Discussion

This section is organized as follows: (1) the preliminary method for orbit fit construction using a fixed $C_D = 2.5$ is presented, and the orbit fit method is verified using the SET-HASDM density database to determine baseline understanding; (2) an assessment of the semi-empirical and physics-based models is presented via orbit fit results that are debiased using a mean drag acceleration scaling factor over the full two week period; (3) an assessment of the semi-empirical and physics-based models is presented via orbit fit results that are debiased using a 24-hour drag acceleration scaling factor.

The specific conditions for producing density values for each model are detailed in Section 4.3. The authors acknowledge that while there are a number of ways to improve a density model's outputs at runtime (see Sutton (2018); Shim et al. (2014)), the outputs used in this work are intended to reflect typical community use, with each model being run according to the developer's operational instructions. Results are presented by focusing on a two week time period from 9 November 2018 - 23 November 2018, providing 14 adjacent daily arcs with no maneuver-based data gaps. The assessment conditions are for the altitude regime near ~ 490 km, in an atmosphere with very low solar flux, and low-to-minor geomagnetic activity. Note that this is a notoriously difficult altitude regime and activity condition for empirical models due to the minimal access to satellite density data at this altitude during times of prolonged solar minimum—increasing the potential value of this style of assessment for these models especially. Figure 3 shows low solar activity for the time period, both in terms of the magnitude and variation of solar EUV and FUV, as approximated by measurements of the 10.7 cm solar radio flux ($F_{10.7}$, top panel). The Kp geomagnetic index (bottom panel) depicts low-to-minor geomagnetic activity during the time of interest, with the two-week period being bookended by minor geomagnetic disturbances which reach no higher than $Kp = 4.3$. Only one minor-to-moderate disturbance occurs on 5 November 2018, four days before the period of interest, reaching a peak of $Kp = 5.7$. This event is mentioned only because of the possibility that its impact could be seen represented in the models as a density enhancement due to delayed heating and cooling effects.

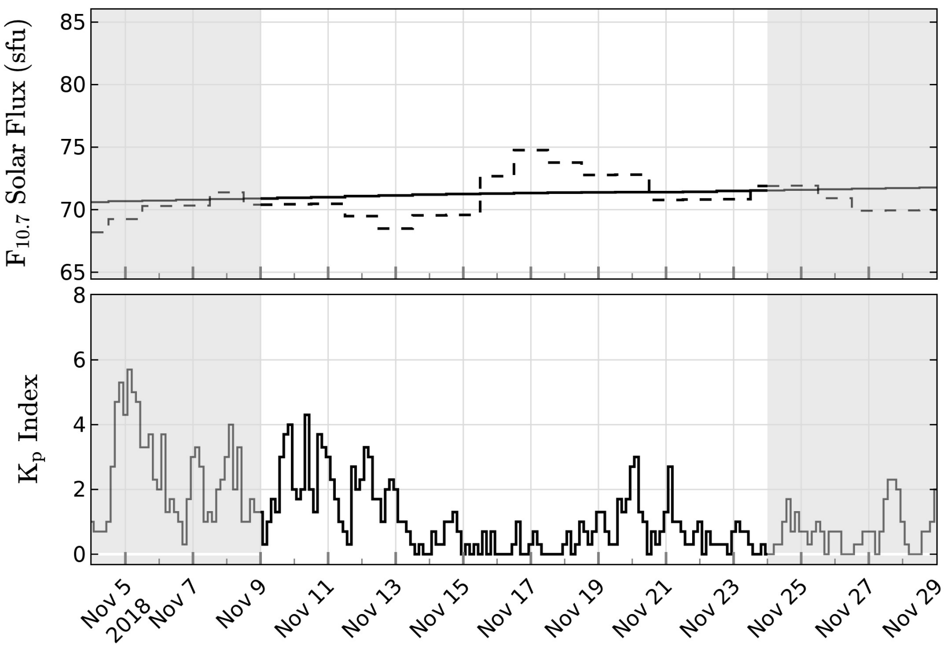


Figure 3. Top: observed solar $F_{10.7}$ radio flux. The dashed curve is the daily measured value from the Ottawa observatory normalized to 1 AU sun-earth distance; the solid curve is an 81-day (3 solar rotation) centered average. Bottom: the 3-hourly planetary magnetic index, Kp. Both panels depict the period of interest from 9 November 2018 - 23 November 2018. A few days before and after the period of interest are depicted in the shaded portions.

6.1 Preliminary Orbit Fits using a Fixed C_D of 2.5

The most straightforward way to construct the orbit fits is by calculating the acceleration due to drag from different density models using a fixed drag coefficient value for all arcs and models. This permits bias depending on the relationship between each model's mean density and the chosen C_D, but importantly demonstrates that the method is sufficiently sensitive for recognizing differences in the drag effects between the models. The in-track residual errors in this method should not be interpreted as indication of performance, but rather indicators of each model's mean density offset relative to the true-unknown density. Figure 4, shows the fixed C_D assessment results for the semi-empirical and physics-based models. The top panel shows each model's orbit averaged density along the orbit of ICESat-2, the middle panel shows the in-track residuals for each model and arc, and the bottom panel shows the in-track RMSe values.

The negative parabolic shape of the in-track residuals of MSIS2, TIEGCM, and JB2008 indicate that these modeled densities are too high—i.e., these orbit fits experience more drag acceleration and their fits tend to lag behind the PSO. The positive parabolic shape of the in-track residuals of DTM2020 and CTIPe indicate modeled densities that are too low—i.e., the drag acceleration is lower and the orbit fits tend to be in front of the PSO. In reality the PSO-to-orbit fit relationship is slightly more complex over an arc, with the above being a generalization of the overall trend. A more detailed understanding of the orbit fit movement relative to the PSO can be found in Appendix A.

The orbit fits from SET-HASDM are separated for use as verification since it uses similar assumptions of a fixed drag coefficient, and a satellite drag data assimilation technique in its internal workings. The SET-HASDM density database affords the opportunity to access historical records of HASDM densities that have been corrected through the real-time data-assimilative calibration to $\sim$80 low earth orbiters. The HASDM model is the operational standard used by the 18th Space Defense Squadron which is tasked with executing command and control over United States' space assets and all resident space objects for sake of space situational awareness. Verification with SET-HASDM provides a baseline understanding of the fidelity of the orbit fit results. Since the HASDM density values have already been effectively debiased in its data-assimilation scheme, we do not go through the steps of debiasing using the methods presented in Sections 6.2 and 6.3. Referring to Figure 5, the HASDM model consistently depicts in-track RMSe values that are on the order of 8.18 meters over the two-week period. The in-track residuals have a negative shape, indicating that the densities from SET-HASDM are slightly larger than what would be expected from the PSO. The results in Figure 5 are intended to serve as an approximate consistency check that our overall methodology, and more specifically our debiasing method, provide orbit fits with in-track errors in a reasonable range.

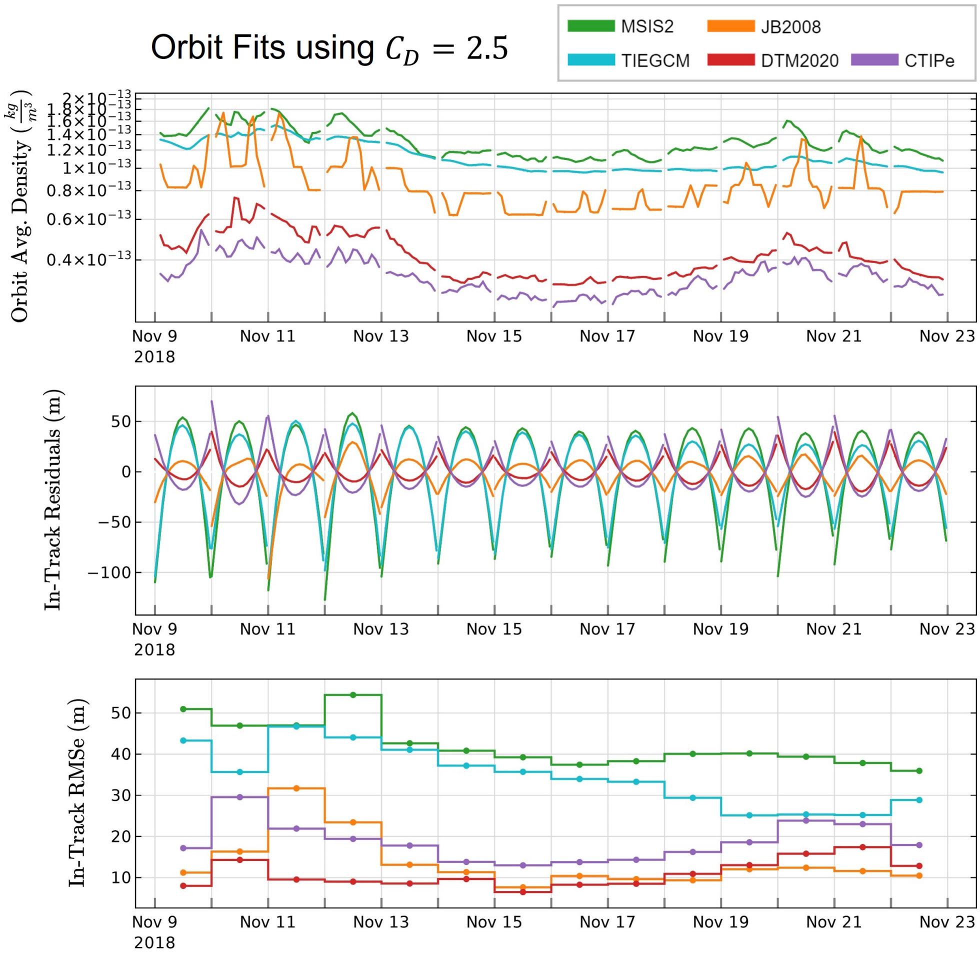


Figure 4. Assessment results given a fixed, common C_D=2.5 for MSIS2 (green), DTM2020 (red), JB2008 (orange), TIEGCM (cyan), and CTIPe (violet) during the two week time period containing 14, 24-hour arcs. Top: Orbit average neutral densities along the ICESat-2 orbit for each model. Middle: In-track orbit residuals for each arc. Bottom: In-track RMSe for each arc's in-track residuals.

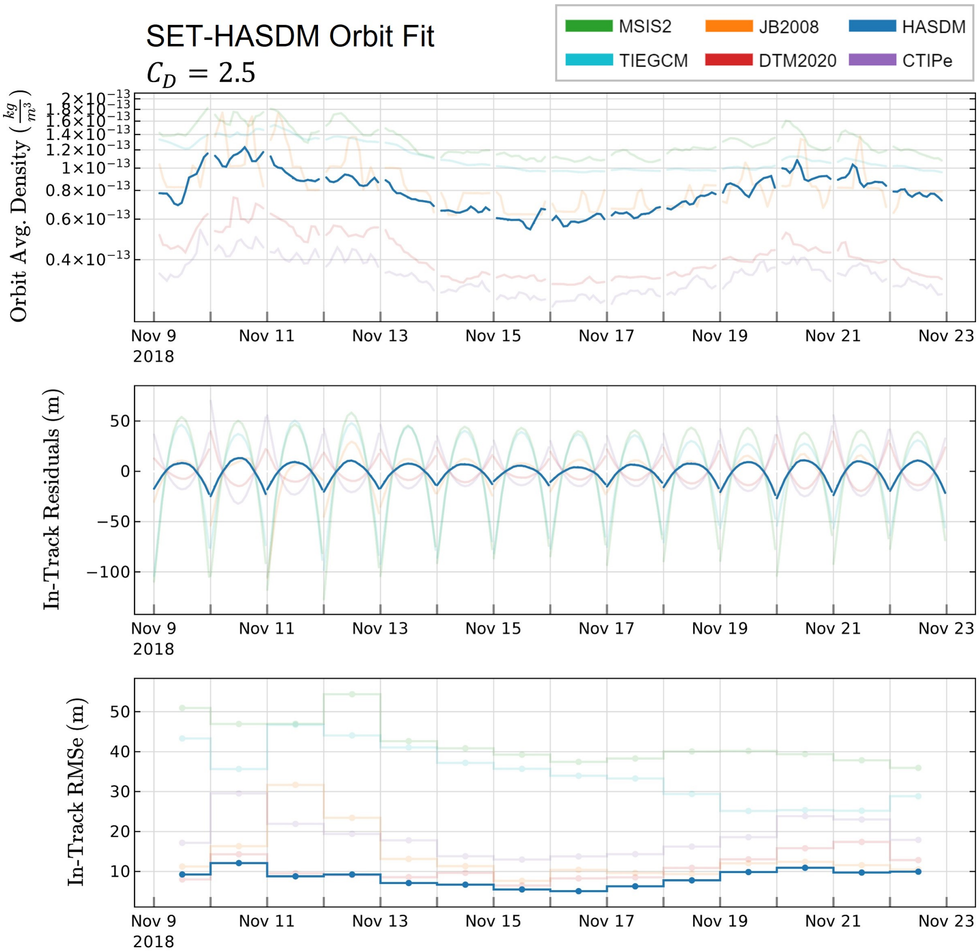


Figure 5. Verification results using SET-HASDM across 14 adjacent, 24-hour arcs from 9 November 2018 - 23 November 2018. Top: The solid blue curve depicts the neutral densities along the ICESat-2 orbit as an orbit average. Middle: In-track orbit residuals for each of the 14 adjacent, 24-hour arcs. Bottom: In-track RMSe values for each arc's in-track residuals. The range of the y-axes are chosen to facilitate comparison with Figure 4.

6.2 Debias using Two-Week Scaling Factor

The second orbit fit method debiases the density models using a mean-adjusted C_D over the two-week period (average values of the 24-hour adjusted C_Ds shown later in Figure 8). This provides a fixed C_D that is unique for each density model, adjusted to account for biases due to each model's mean density offset over the time period (see Figure 4). The in-track residuals and RMSe values for this assessment, shown in Figure 6, quantify the error due to density variation over the two-week time period. The average adjusted C_D used to construct the two-week debiased orbit fits for each model is reported in Table 2.

As mentioned in Section 5.2, the DRIA calculations indicate that $C_D = 2.5$ is a realistic value if one assumes that a fixed energy accommodation of $\alpha = 0.89$ is reasonable—an assumption limited by lack of empirical observation. According to the DRIA model, 2.5 is a realistic lower limit for the drag coefficient. Looking to Table 2, we can see that the mean adjusted C_D for MSIS2, TIEGCM, and JB2008 are all well below this lower limit, offering further evidence that these models are, on average, over-estimating the density. The upper bound is slightly more difficult to estimate in this setup, but CTIPe's adjusted C_D of 4.3 is likely too high. This will need to be investigated further in the future.

After removing the bias from the models, the relative effects of the minor geomagnetic activity become more stark in the in-track residuals. Here the changing shape of the residual curves indicate whether the model is over or underestimating the effects of geomagnetic activity on the modeled density. For example, several of the models display downward-pointing curves during geomagnetic activity, indicating densities that are too low as compared to their quiet time densities. DTM2020's residuals show an anti-correlation to the geomagnetic activity, beginning with densities that are too high and ending the two weeks with densities that are too low, which may be reflective of an overly sensitive response to geomagnetic activity and the overall downward trend in Kp. The effects of a model poorly capturing density variations during geomagnetically active times are now better quantified by the in-track RMSe after two-week debiasing is applied.

Two-week Debiasing Method: C_D = Mean Adjusted C_D

Model ID	Fixed C_D	Scaling Factor as % change
MSIS2	1.237	-49.861
TIEGCM	1.373	-44.899
JB2008	1.909	-22.967
DTM2020	3.351	34.903
CTIPe	4.368	75.663

Table 2. Summary of the assessment procedure assuming a fixed C_D that is equal to the average adjusted value for each model, assuming an a priori of $C_D = 2.5$. The second column reports mean-adjusted C_D used in each model's orbit fit construction. The third column reports the two-week scaling factor as a percent change for each model.

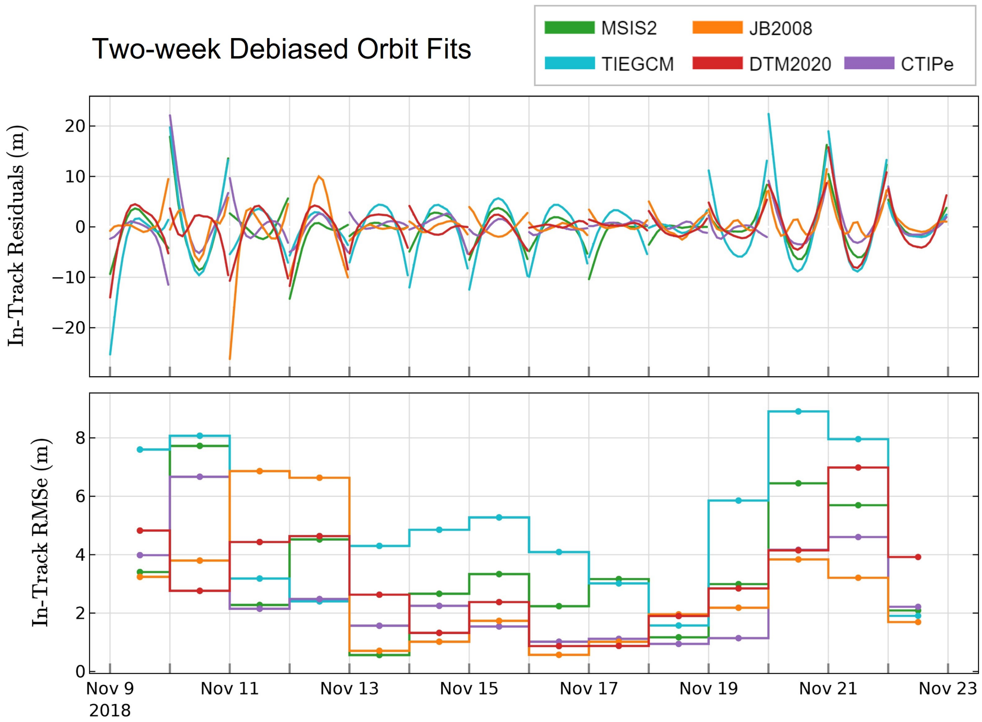


Figure 6. Assessment results for orbit fits that are debiased using two-week drag acceleration scaling factors. Top: Debiased in-track orbit residuals for each arc. Bottom: Debiased in-track RMSe for each arc's in-track residuals.

6.3 Debias using 24-hour Scaling Factor

The 24-hour debiasing procedure described in Section 5.2, is used to scale the orbit fits and their residuals to a daily cadence. Figure 7 presents the resulting in-track residuals (top panel) and in-track RMSe values (bottom panel) for each 24-hour arc for each model. The 24-hour drag acceleration scaling factor is derived by adjusting the C_D from the a priori of 2.5 over each daily arc, absorbing the average density offset for that day. The debiasing effect is seen in the overall reduction in residual error from Figure 4 to Figure 7. The calculated 24-hour scaling factors are presented in the top panel of Figure 8 as a percent change from the fixed value of 2.5. The bottom and right panels show the Kp index and Pearson's correlation coefficient between each model's scaling factors and the Kp, respectively.

The 24-hour scaling accounts for both the overall model bias and uncertainties in the density on timescales that are on the order of, or greater than, the chosen arc length of 24-hours (i.e. combination of mean density offset and daily geomagnetic variation). The remaining error depicted by the in-track residuals of Figure 7 are likely due to higher frequency variations in density that are not captured by the 24-hour debiasing (e.g. day-night variations in the neutral density). The correlation between Kp index and the scaling factors demonstrates how this metric can be used to determine how well a model accounts for geomagnetic activity. As shown by the scaling factors in Figure 8, MSIS2 ($R = 0.35$), TIEGCM ($R = 0.29$), and CTIPe ($R = 0.19$) all exhibit a subtle positive correlation, indicating a slight underestimation of density enhancements from geomagnetic activity and resulting in the scaling factors being used to compensate for these errors. Contrarily, the inverse relationships shown by DTM2020 ($R = -0.36$) and JB2008 ($R =$

695 −0.3) indicate an overestimation of these models' densities during days of increased ac-
696 tivity.

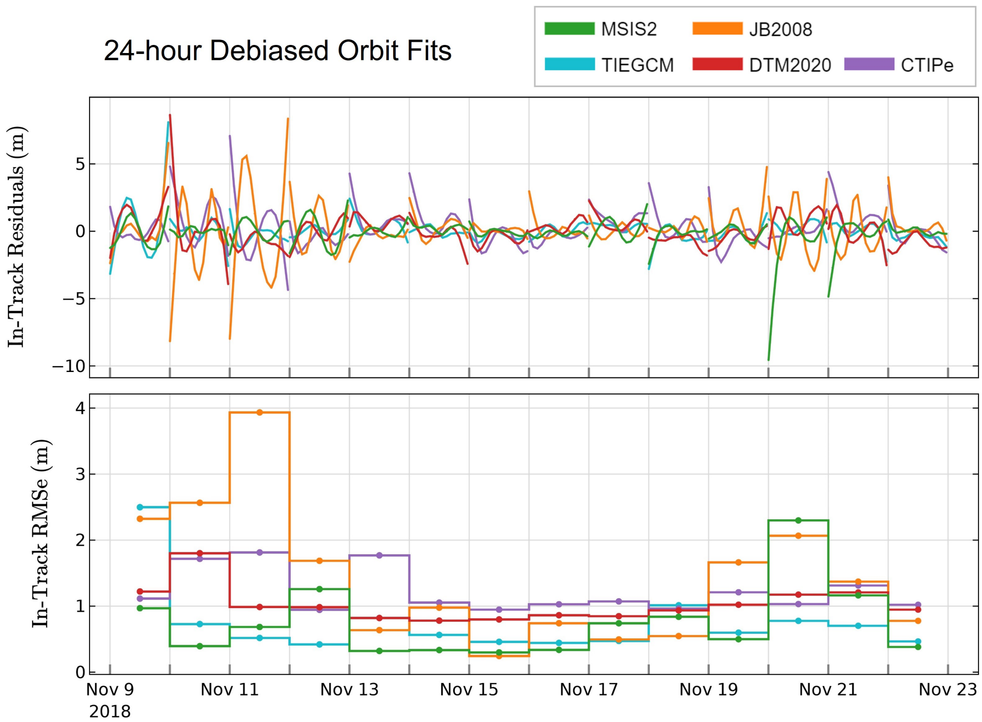


Figure 7. Assessment results for orbit fits using 24-hour drag acceleration scaling factors. Scaling factors are extracted in the least squares orbit fitting procedure by allowing the C_D to adjust once-per-arc to absorb observed errors between the PSO and the converging orbit fit. Top: In-track orbit residuals for each arc. Bottom: In-track RMSe for each arc's in-track residuals.

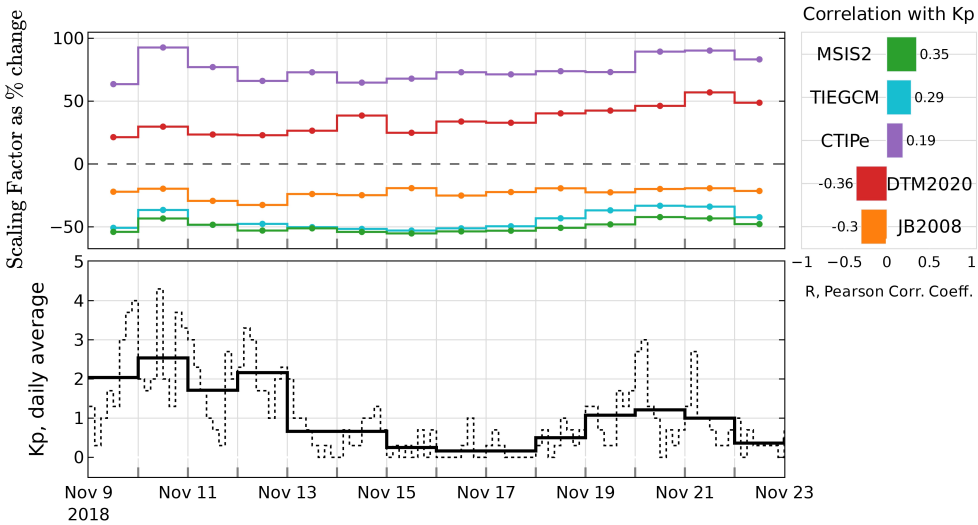


Figure 8. Top: Drag acceleration scaling factors extracted from the orbit fits shown in Figure 7, presented as a percent change from the fixed $C_D = 2.5$. Bottom left: The Kp index for the time period. Bottom right: Pearson's correlation coefficient between the scaling factors and Kp.

6.4 Discussion

While the in-track residuals of the ICESat-2 orbit fits offer an effective means of assessing the density models, most methods that use drag acceleration to study density are going to be limited by the complex interconnected uncertainties in the C_D and the density. For these reasons, we split our overall assessment into the three methods presented, each of which help to further illuminate the performance of the models. For sake of brevity, only JB2008 and TIEGCM are given in-depth discussions that synthesize an understanding of model performance from their results. The reader will then be able to apply these discussions to the remaining models, which here are discussed in more general terms.

Considering the cumulative results for JB2008 (plotted in orange for all relevant figures), Figure 4 shows that with a fixed C_D of 2.5, the in-track residuals exhibit a negative parabola shape, indicating that the JB2008 density provides larger drag accelerations than what the PSO experiences. Figure 6 provides the assessment method in which the JB2008 densities are effectively scaled by $\sim\!-22\%$ for the full two-week period. In this case the in-track residuals better highlight the arcs in which JB2008 performs poorly relative to the other models, specifically on November 11th and 12th when the geomagnetic activity fluctuates around $Kp = 3$ after having been moderately elevated for several days. In the 24-hour debiased case, JB2008's daily scaling factors (shown in orange in Figure 8) effectively compensate for density variations on the order of or greater than 24-hours. The mean density offset adjustment is $\sim\!-22\%$, and is seen to be inversely correlated ($R = -0.3$) with the geomagnetic activity—indicating that JB2008 tends to overestimate densities during active times. This effect can be clearly seen in JB2008's orbit averaged densities, represented by the orange line in the top panel of Figure 4, where the model provides much sharper density peaks than the other models during the times of slightly elevated geomagnetic activity. This effect is also clearly represented in the 24-hour debiased in-track RMSe values (bottom panel of Figure 7), which shows significant variance during active times between the PSO and the JB2008 orbit fit, even after the 24-hour scaling. The higher RMSe values shown in Figure 7 on November 9th-13th indicate that JB2008's overestimation of density during active times is not sufficiently compensated by the 24-hour scaling factor, meaning that the variation likely occurs at a higher frequency. JB2008's estimation of quiet time density is among the best in this report, needing $\leq\!-22\%$ adjustment from the fixed C_D case to provide orbit fits that match the PSO to within 1 meter (during quiet times).

Considering the cumulative results for TIEGCM (plotted in cyan for all relevant figures), Figure 4 shows that with a fixed C_D of 2.5, the in-track residuals exhibit a negative parabola shape, indicating that the TIEGCM density provides larger drag accelerations than what the PSO experiences (i.e. the TIEGCM densities are too high). TIEGCM offers interesting results from Figure 6 of the two-week scaled case. The TIEGCM densities are effectively scaled down by $\sim\!-45\%$ over the two-week period to compensate for mean density offset. Since this value is the two week average from the scaling factors shown in Figure 8, it is skewed to only partially compensate for lack of geomagnetic sensitivity (i.e., densities are scaled to be too low in active times, increase in RMSe on the bookends of the period) and partially compensate for mean density offset in quiet times (densities are scaled to be too high in quiet times, increase in RMSe from Nov. 13th to 17th). The two-week scaled case is able to clearly show in the in-track residuals that TIEGCM struggles more than the other models to properly capture variation during the period of this study. TIEGCM's daily scaling factors (cyan line in Figure 8) effectively compensate for error in density variations on the order of or greater than 24-hours. The daily variation of the scaling factors is found to slightly correlate with the Kp ($R = 0.29$)—indicating that TIEGCM tends to subtly underestimate densities during active times. This is seen in TIEGCM's orbit averaged densities (cyan line in the top panel of Figure 4) where the model provides significantly less sensitivity to the times of slightly elevated geomagnetic activity, and less variation overall. The 24-hour debiased in-track RMSe values (bottom panel of Figure 7), interestingly show very low variance between the PSO and the TIEGCM orbit fit after the 24-hour scaling. This is most likely explained by the overestimation of density during active times being sufficiently compensated for by the 24-hour scaling factors despite its orbit average densities seeming to lack much variation at all. This adds further suspicion to the higher frequency variations seen in models such as JB2008. It is possible that by moving to shorter arc lengths, such as 3-hours, the time series of scaling factors could better capture these variations, and this is a future goal of this work.

In general, MSIS2 and TIEGCM overestimate the density for all arcs, requiring $\sim\!-50\%$ and $\sim\!-45\%$ scaling factors, respectively, to bring the in-track residuals to within two meters. DTM2020 and CTIPe both underestimate the density for all arcs, each requiring a $\sim$35% and $\sim$76% increase, respectively. JB2008 requires the least overall scaling, requiring only $\sim\!-23\%$ to bring the in-track residuals to within two meters during quiet times. All models capture the geomagnetic activity relatively well as demonstrated by their scaling factors not being very highly correlated to Kp. DTM2020 ($R = -0.36$) and JB2008's ($R = -0.3$) scaling factors are inversely correlated to Kp, indicating a slight over-sensitivity to geomagnetic activity during this time period, while TIEGCM ($R = 0.29$), MSIS2 ($R = 0.35$), CTIPe ($R = 0.19$) all indicate an under-sensitivity. The scaling undergone for each model produces RMSe values that are comparable to that of the SET-HASDM orbit fit, which was separated out to serve as an approximate consistency check of our debiasing method due to its data-assimilative technique.

7 Conclusions and Future Work

This work presents the development of a modernized interface for the GEODYN-II POD software. The approach leverages the high-precision nature of space geodetic POD and an upgraded utility of the neutral density models to focus POD methods toward studying satellite drag and conducting density model assessment. The assessment method uses high-fidelity PSO as observed tracking measurements that are input into POD-based orbit fits. The drag effects from each density model are assessed according to each model's ability to redetermine the satellite's orbit. Each density model's orbit fit contains relative in-track deviations from the PSO which are treated as a proxy for model densities that differ from a true, unknown, density. These deviations are quantified with the in-track residuals and their RMS errors. We demonstrate the capabilities of this tool via a case study assessment of five thermospheric density models (MSIS2, DTM2020, JB2008, TIEGCM, and CTIPe, and a verification using SET-HASDM) using the ICESat-2 mission PSO as the observed measurements. Preliminary orbit fits are constructed after determining a mean C_D from a physics-based solution. A fixed C_D of 2.5 is applied for all models before being debiased by adjusting the C_D to account for density errors in the drag acceleration. The debiasing is performed at two different cadences, 24-hours and two-weeks, with each method highlighting different temporal aspects of the model density errors. The scaling factors extracted from the 24-hour and two-week debiasing methods are well-equipped for use in improving forecasting and modeling methods. The 24-hour scaling factors provide a more accurate representation of the true density variations for each model, while the two-week scaling factors are computationally simpler and indicate more baseline density effects. In addition, the two-week extended time period scaling factors are compatible for parsing into the CCMC's CAMEL database to move in the direction of community-oriented model validation.

We continue our efforts on this project as we move in the direction of offering a more robust thermospheric model validation scheme. Possible improvements include improving the non-conservative force modeling in GEODYN for ICESat-2 using a more realistic 3-D model of the satellite shape that would account for self-shadowing and variations in cross-sectional area with incidence angle e.g, as in March et al. (2019). The orbit determination for the primary science orbits, and the subsequent analyses described in this paper would have to use these improved geometry models. A further improvement would be to incorporate SLR measurements of ICESat-2 into the evaluations of the density models. One could include the SLR data along with the PSO trajectory data in the evaluation. See Thomas et al. (2021) for a description of these data for ICESat-2. Planned future work involves addressing the key constraints highlighted in the methodology, the foremost of which is the need to evaluate the drag coefficient more frequently along the ICESat-2 orbit. Future work will also involve expanding the study to encompass the entire ICESat-2 mission time period. Additional expansion includes incorporating additional satellites and constellations that may illuminate model performance within atmospheric regimes that lack observations of neutral density. We aim to make our expanded results available through the CCMC's CAMEL framework as well as through future publications.

The assumptions made in this paper are limited by the current status of unknowns between gas surface interaction research and thermospheric variability research. At this time, the true drag coefficient is not known for any satellite, and modeling the C_D will always introduce some inherent bias into the results. We aim to address this issue by implementing several of the satellite gas-surface interaction models currently used in the satellite drag community to calculate the time and compositionally-dependent drag coefficient. Isolating the effects of the C_D will aid to better identify the various non-density related errors that may be present in the in-track residuals. Being able to distinguish these errors and accurately quantify the amount of deviation introduced by a given density model will provide significant insight regarding model performance to the earth-space environment modeling community. As the ability to model C_D improves, the results provided by this method will similarly become more valid. The Geospace Dynamics Constellation (GDC) is an upcoming NASA mission that is intended to help fill in the gaps of understanding gas-surface interactions by providing a stable platform with full measurements of neutral composition, density, and temperature along with a high fidelity POD in which cross validation of density model assessment is possible. As a result, in addition to providing its own neutral density observations that can be used for research, operations, and model validation, these advances expected from the GDC mission will improve the accuracy and usability of density proxies derived from POD solutions like those used here. These advances will effectively multiply our density observations to be able to use any satellite with sufficiently accurate GNSS positioning and knowledge of spacecraft parameters as a density observing platform.

This work provides a step in the direction of being able to use high-fidelity GNSS-enabled LEO satellite POD solutions to objectively quantify and validate thermospheric model performance. The strength of assessment using this method is its ability to identify relative accuracy of the models in a way that is directly tied to operational use for orbit propagation. There are a multitude of uses for the tools and methods presented in this work, such as for density retrievals along the orbit of a satellite, which is a planned future effort; however, this report focuses specifically on model assessment. As work continues to refine these methods and address the caveats presented in this paper, the results of model assessments using this technique will continue to become better suited to aid satellite operators when choosing a model that will perform best under specified conditions. Having a multitude of methods for assessing upper atmospheric models under various conditions helps model developers refine the models themselves, making them better suited for orbit prediction.

8 Open Research

The ICESat-2 POD solutions, their corresponding setup files, and the GEODYN II software are provided by the Geodesy and Geophysics Laboratory within NASA-GSFC. Simulation results for the CTIPe model have been provided by the Community Coordinated Modeling Center at Goddard Space Flight Center through their publicly available simulation services (https://ccmc.gsfc.nasa.gov). Orbit fly-throughs of the TIEGCM simulation results and relevant codes used to produce the results in this paper are available at Zenodo via https://doi.org/10.5281/zenodo.8015368 (Waldron et al., 2023). The SET HASDM density data are provided for scientific use by Space Environment Technologies.

Acknowledgments
The authors acknowledge the generous support of the Community Coordinated Modeling Center under grant number 80NSSC21K1747 and support from the University of Colorado at Boulder Grand Challenge Grant: Space Weather Technology, Research, and Education Center (SWx TREC). This material is partially based upon work supported by the National Aeronautics and Space Administration under Grant 80NSSC21K1554 issued through the Heliophysics Division Space Weather Science Application initiative and the Geospace Dynamics Constellation Contract No. 80GSFC22CA012. The authors acknowledge the CCMC-Kamodo Team who designed the Kamodo API, permitting the ability to interface physics-based model outputs to the POD scheme.

Appendix A Coordinate System to Study Drag

GEODYN's input and output trajectories make use of the J2000 inertial reference system. The $\hat{X}$, $\hat{Y}$, and $\hat{Z}$ components of the inertial coordinate system offer limited information on how a satellite's orbit is impacted by atmospheric drag, leading us to convert to the more suitable Satellite Coordinate System. Two coordinate frames suited for this assessment are the NTW and RSW frames, with differences between the two being highlighted in Figure A1. We make use of the NTW system, which aligns with the orbit plane and is composed of an in-track component $\hat{T}$ that is parallel to the velocity vector $\vec{v}$, a normal component $\hat{N}$ that is perpendicular to the velocity and nominally in the radial direction, and a cross-track component $\hat{W}$ that is normal to the orbit plane and completes the right-hand coordinate frame. Being parallel to the velocity vector means that the in-track component $\hat{T}$ will contain any indication that the spacecraft's trajectory has changed since orbital energy dissipations from drag will impact in the velocity direction.

Figure A2 contains additional visualization related to the shape of the in-track residuals and how it relates to the movement of the PSO relative to the orbit-fit satellite. The overall shape of the in-track residuals is a result of the batch-least squares fitting routine as it attempts to minimize the distance between the PSO and the orbit fit across the whole arc.

Variations in the in-track component $\hat{T}$ are not the same as variations in the along-track component $\hat{S}$ of the RSW system. In-track variations act in the direction of the velocity vector, whereas along-track variations are merely along, but not necessarily parallel, to the direction of the velocity vector. We make the distinction to use the NTW system rather than the RSW system whose radial component is often used to assess orbit accuracy in geodetic POD studies. The NTW coordinate system is described in (Vallado, 2013) to have the following unit vectors and transformation:

$$\hat{T} = \frac{\mathbf{v}}{|\,\mathbf{v}\,|} \tag{A1}$$

$$\hat{W} = \frac{\mathbf{r}\times\mathbf{v}}{|\,\mathbf{r}\times\mathbf{v}\,|} \tag{A2}$$

$$\hat{N} = \hat{T}\times\hat{W} \tag{A3}$$

$$\mathbf{r}_{XYZ} = \left[\hat{N} \vdots \hat{T} \vdots \hat{W}\right]\mathbf{r}_{NTW} \tag{A4}$$

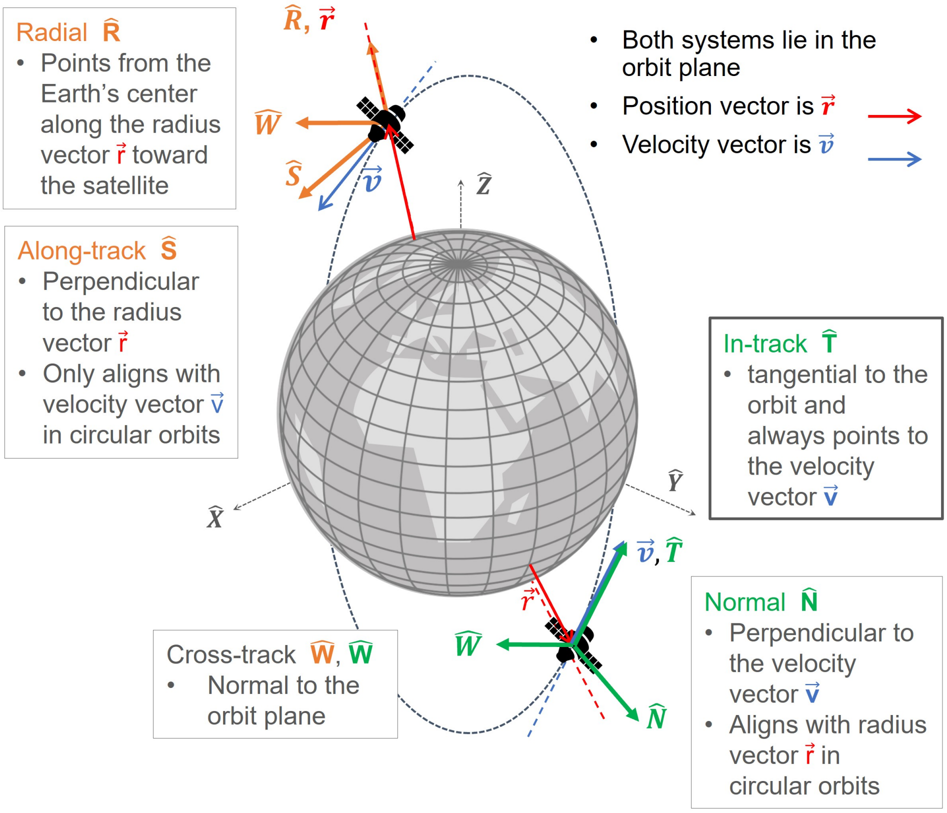


Figure A1. The above is a schematic showing the NTW and RSW satellite coordinate systems and details regarding their components. The NTW system's in-track component is parallel to the velocity vector, making it an effective tool for assessing relative effects due to atmospheric drag.

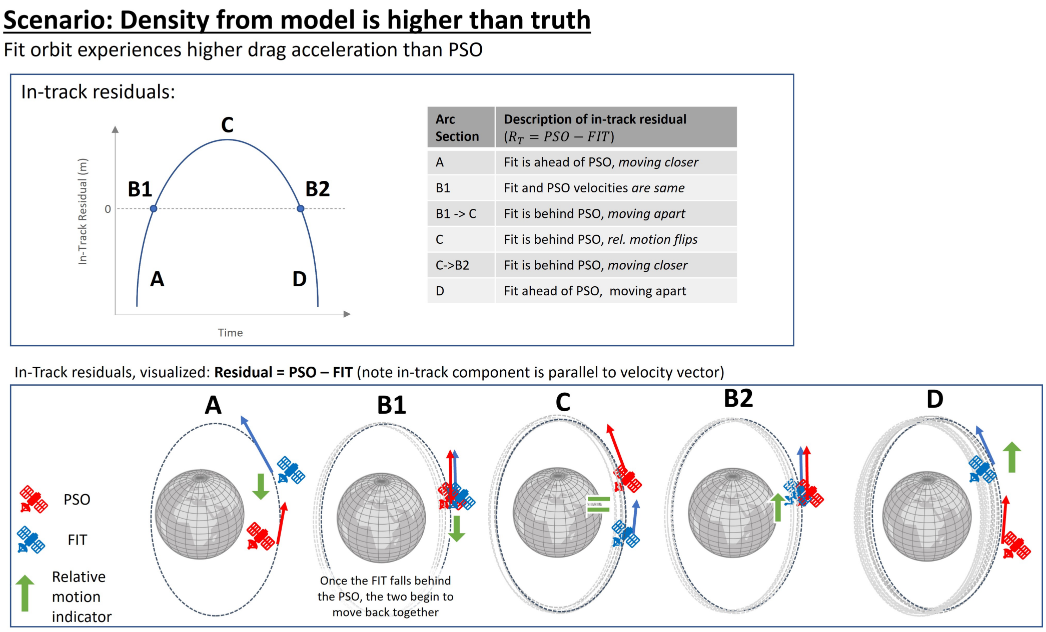


Figure A2. Above is a diagram which explores the circumstances that result in the parabolic shape of the in-track residuals. Representative in-track residuals for a single arc is given as a negative quadratic curve in the top panel. The bottom panel provides corresponding frames of schematics for each marked point to depict how the orbit fit and PSO are positioned relative to each other.

Appendix B GEODYN Run Setup for Orbit Fits

This section provides a summarized overview, in the form of Table B1, of the run setup used in GEODYN to produce the orbit fits. A similar overview is given in Table 1 of Thomas et al. (2021), which details the most relevant and important constants, models, and standards used to produce the ICESat-2 PSO.

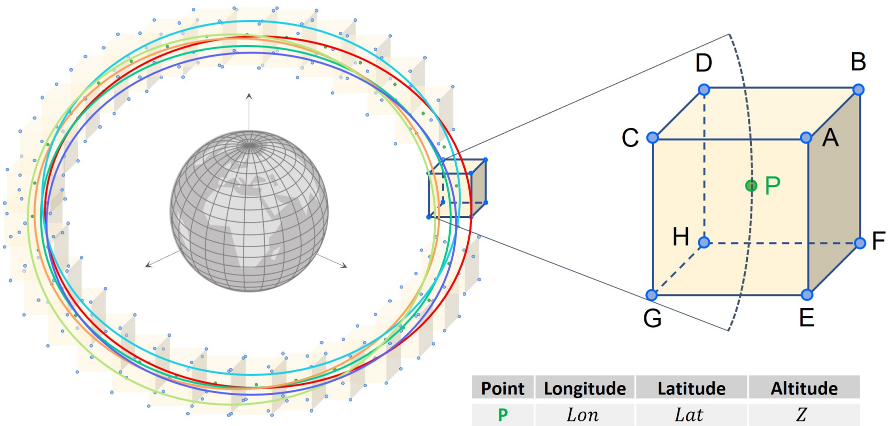
GEODYN's Setup to Produce POD-based Orbit Fits

Satellite geometry	Panel model based on pre-launch geometry and satellite surface optical properties. 14 panels make up a Box Wing model to calculate time varying area.
ICESat-2 Attitude	Telemetered spacecraft body-fixed reference frame to inertial reference frame quaternions. Telemetered solar array drive angles (for force modeling)
Non-Conservative Forces	
Atmospheric Density Models	Modified for comparison; see Table 1 for the list of assessed density models
Earth Radiation	Knocke 2nd degree zonal spherical harmonic of Earth's albedo and emissivity (Knocke et al., 1988)
Solar Radiation Pressure	Solar radiation incident on plate model (Luthcke et al., 2019; Marshall & Luthcke, 1994)
Conservative Forces	
Geopotential gravity	EIGEN6C, tide-free (Foerste et al., 2014)
Time variable gravity	Contribution from atmosphere, non-tidal oceans, hydrology, and ice; Developed from GRACE models
Earth, Pole, and Ocean tides	IERS2010 Conventions (Petit & Luzum, 2010)
Planetary ephemerides (N-Body)	JPL DE430 (Folkner et al., 2014)
Relativistic corrections	IERS2010
General Reference Frame and Constants	
Conventional inertial system	J2000 geocentric; mean equator and equinox of 2000 JAN 01 12:00:00; IERS2010
Precession - Nutation	IAU 2000A precession-nutation model
Earth Orientation Parameters	IERS 08 C04 (Bizouard & Gambis, 2011), IERS2010 conventions for diurnal, semidiurnal, and long period tidal effects on polar motion and UT1
Numerical integration	Cowell predictor-corrector; fixed and variable step; equations of motion and variational equations.
Estimation method	Partitioned Bayesian least squares.
GEODYN Controlled Setup Information	
Tracking data type	PCE (orbit trajectory) using ICESat-2 PSO
POD technique	Dynamic data reduction (no empirical accelerations)
Arc Length	24 hours
Adjusted Parameters	Initial conditions only: $X, Y, Z, \dot{X}, \dot{Y}, \dot{Z}$
Force model parameters	C_D = fixed; $C_R = 1$ (not adjusted)
Integration/orbit step	10 seconds

Table B1. A summarized overview of the GEODYN run setup for using the program to conduct density model assessment. Many of the above parameters are summarized from (Thomas et al., 2021) and (Luthcke et al., 2019).

Appendix C Orbit Uncertainty Interpolation Technique and the Kamodo Interface

Kamodo is a CCMC tool for access, interpolation, and visualization of space weather models and data in Python (Ringuette et al., 2023). Kamodo allows model developers to represent simulation results as mathematical functions which may be manipulated directly by end users. Kamodo handles unit conversion transparently and supports interactive science discovery through Jupyter notebooks with minimal coding in Python. Kamodo is chosen for this project due to its ability to offer model agnostic methods for reading data output from different model sources. Kamodo is called using its Satellite Flythrough capabilities, in which a user is able to sample the models with satellite ephemeris and return requested values from the chosen model. The orbit is pre-initialized in GEODYN using MSIS2 to get an a priori estimate for the orbit coordinates. Then using the a priori orbit, extend out the uncertainty of the coordinates to create a cube of possible values centered on the orbit. This approach accounts for possible model output differences as the orbit iteratively converges towards a solution. Finally, we plug the orbit and its uncertainty cubes into Kamodo to interpolate the model densities at all requested points. By doing this, the orbit density values from the physics model can be quickly ingested into the POD program. Figure C1 visualizes this procedure.



Point	Longitude	Latitude	Altitude
P	Lon	Lat	Z
A	$Lon + \delta_{deg}$	$Lat + \delta_{deg}$	$Z + \delta_m$
B	$Lon + \delta_{deg}$	$Lat - \delta_{deg}$	$Z + \delta_m$
C	$Lon - \delta_{deg}$	$Lat + \delta_{deg}$	$Z + \delta_m$
D	$Lon - \delta_{deg}$	$Lat - \delta_{deg}$	$Z + \delta_m$
E	$Lon + \delta_{deg}$	$Lat + \delta_{deg}$	$Z - \delta_m$
F	$Lon + \delta_{deg}$	$Lat - \delta_{deg}$	$Z - \delta_m$
G	$Lon - \delta_{deg}$	$Lat + \delta_{deg}$	$Z - \delta_m$
H	$Lon - \delta_{deg}$	$Lat - \delta_{deg}$	$Z - \delta_m$

Figure C1. A representative schematic showing the constructed “cube of uncertainty” that surrounds a given coordinate along the orbit of a satellite. Each point that makes up this cube will contain modeled neutral density values between which we can interpolate in GEODYN as the orbit drifts from the a priori orbit. This figure also demonstrates how perturbations due to different density models, represented here by different colored orbits, may necessitate a range of uncertainty for the satellite’s indexed location.

References

Berger, T. E., Dominique, M., Lucas, G., Pilinski, M., Ray, V., Sewell, R., … Thiemann, E. (2023). The thermosphere is a drag: The 2022 starlink incident and the threat of geomagnetic storms to low earth orbit space operations. *Space Weather*, *21*(3). doi: https://doi.org/10.1029/2022SW003330

Berger, T. E., Holzinger, M. J., Sutton, E. K., & Thayer, J. P. (2020). Flying Through Uncertainty. *Space Weather*, *18*(1). doi: https://doi.org/10.1029/2019SW002373

Bernstein, V., & Pilinski, M. (2022). Drag coefficient constraints for space weather observations in the upper thermosphere. *Space Weather*, *20*, e2021SW002977. doi: https://doi.org/10.1029/2021SW002977

Bizouard, C., & Gambis, D. (2011). *The combined solution C04 for Earth Orientation Parameters consistent with International Terrestrial Reference Frame 2008* (Tech. Rep.). Observatoire de Paris, Paris, France. Retrieved from `https://hpiers.obspm.fr/iers/eop/eopc04_08/C04.guide.pdf`

Bowman, B. R., Tobiska, W. K., Marcos, F. A., Huang, C. Y., Lin, C. S., & Burke, W. J. (2008). A New Empirical Thermospheric Density Model JB2008 Using New Solar and Geomagnetic Indices. *AIAA/AAS Astrodynamics Specialist Conference and Exhibit*. doi: https://doi.org/10.2514/6.2008-6438

Bruinsma, S., Arnold, D., Jäggi, A., & Sánchez-Ortiz, N. (2017). Semi-empirical thermosphere model evaluation at low altitude with GOCE densities. *Journal of Space Weather and Space Climate*, *7*(A4). doi: https://doi.org/10.1051/swsc/2017003

Bruinsma, S., & Boniface, C. (2021). The operational and research DTM-2020 thermosphere models. *Journal of Space Weather and Space Climate*, *11*(3). doi: https://doi.org/10.1051/swsc/2021032

Bruinsma, S., Boniface, C., Sutton, E. K., & Fedrizzi, M. (2021). Thermosphere modeling capabilities assessment: Geomagnetic storms. *Journal of Space Weather and Space Climate*, *11*(2018). doi: https://doi.org/10.1051/swsc/2021002

Bruinsma, S., Sutton, E., Solomon, S. C., Fuller-Rowell, T., & Fedrizzi, M. (2018). Space Weather Modeling Capabilities Assessment: Neutral Density for Orbit Determination at low Earth orbit. *Space Weather*, *16*(11), 1806–1816. doi: https://doi.org/10.1029/2018SW002027

Bruinsma, S., Tamagnan, D., & Biancale, R. (2004). Atmospheric densities derived from CHAMP/STAR accelerometer observations. *Planetary and Space Science*, *52*, 297-312. doi: https://doi.org/10.1016/j.pss.2003.11.004

Bussy-Virat, C. D., Ridley, A. J., & Getchius, J. W. (2018). Effects of Uncertainties in the Atmospheric Density on the Probability of Collision Between Space Objects. *Space Weather*, *16*(5), 519–537. doi: https://doi.org/10.1029/2017SW001705

Doornbos, E., van den IJssel, J., Luhr, H., Forster, M., & Koppenwallner, G. (2010). Neutral Density and Crosswind Determination from Arbitrarily Oriented Multiaxis Accelerometers on Satellites. *Journal of Spacecraft and Rockets*, *47*(4), 580-589. doi: https://doi.org/10.2514/1.48114

Emmert, J. T. (2015). Thermospheric mass density: A review. *Advances in Space Research*, *56*(5), 773–824. doi: https://doi.org/10.1016/j.asr.2015.05.038

Emmert, J. T., Drob, D. P., Picone, J. M., Siskind, D. E., Jones, M., Mlynczak, M. G., … Yuan, T. (2021). NRLMSIS 2.0: A Whole-Atmosphere Empirical Model of Temperature and Neutral Species Densities. *Earth and Space Science*, *8*(3), e2020EA001321. doi: https://doi.org/10.1029/2020EA001321

Fang, T. W., Kubaryk, A., Goldstein, D., Li, Z., Fuller-Rowell, T., Millward, G., … Babcock, E. (2022). Space Weather Environment During the SpaceX Starlink Satellite Loss in February 2022. *Space Weather*, *20*. doi: https://doi.org/10.1029/2022SW003193

Foerste, C., Bruinsma, S., Abrikosov, S., & et al. (2014). EIGEN-6C4 The latest combined global gravity field model including GOCE data up to degree and order 2190 of GFZ Potsdam and GRGS Toulouse. *GFZ Data Services*. doi: https://doi.org/10.5880/icgem.2015.1

Folkner, W., Williams, J., Boggs, D., Park, R., & Kuchynka, P. (2014). *The Planetary and Lunar Ephemerides DE430 and DE431* (Tech. Rep.). IPN Progress Report, 42-196, Jet Propulsion Laboratory, Pasadena, California. Retrieved from `https://ipnpr.jpl.nasa.gov/progress_report/42-196/196C.pdf`

Hapgood, M., Liu, H., & Lugaz, N. (2022). SpaceX – sailing close to the space weather? *Space Weather*. doi: https://doi.org/10.1029/2022SW003074

He, C., Yang, Y., Carter, B., Kerr, E., Wu, S., Deleflie, F., … Norman, R. (2018). Review and comparison of empirical thermospheric mass density models. *Progress in Aerospace Sciences*, *103*, 31–51. doi: https://doi.org/10.1016/j.paerosci.2018.10.003

Heelis, R. A., Lowell, J. K., & Spiro, R. W. (1982). A model of the high-latitude ionospheric convection pattern. *Journal of Geophysical Research: Space Physics*, *87*(A8), 6339–6345. doi: https://doi.org/10.1029/JA087iA08p06339

Hejduk, M. D., & Snow, D. E. (2018). The Effect of Neutral Density Estimation Errors on Satellite Conjunction Serious Event Rates. *Space Weather*, *16*(7), 849–869. doi: https://doi.org/10.1029/2017SW001720

Knocke, P. C., Ries, J. C., & Tapley, B. D. (1988). Earth radiation pressure effects on satellites. *Astrodynamics Conference, 1988*, 577–587. Retrieved from `https://arc.aiaa.org/doi/abs/10.2514/6.1988-4292` doi: 10.2514/6.1988-4292

Lemoine, F. G., Chinn, D. S., Zelensky, N. P., Beall, J. W., & Bail, K. L. (2016). The development of the GSFC DORIS contribution to ITRF2014. *Advances in Space Research*, *58(12)*, 2520-2542. doi: https://doi.org/10.1016/j.asr.2015.12.043

Loomis, B. D., Luthcke, S. B., & Sabaka, T. J. (2019). Regularization and error characterization of GRACE mascons. *Journal of Geodesy*, *93*(9), 1381–1398. doi: https://doi.org/10.1007/s00190-019-01252-y

Luthcke, S. B., Pennington, T., Loomis, B. D., Rebold, T., & Thomas, T. (2019). *ICESat-2 ATBD for Precise Orbit Determination, Orbit Design, and Geolocation Parameter Calibration* (Tech. Rep. No. Release 002). NASA Goddard Space Flight Center. Retrieved from `https://icesat-2.gsfc.nasa.gov/sites/default/files/page_files/ICESat2_POD_ATBD_r002.pdf`

Luthcke, S. B., Rowlands, D. D., Lemoine, F. G., Klosko, S. M., Chinn, D., & McCarthy, J. J. (2006). Monthly spherical harmonic gravity field solutions determined from GRACE inter-satellite range-rate data alone. *Geophysical Research Letters*, *33*(2). doi: https://doi.org/10.1029/2005GL024846

Luthcke, S. B., Zelensky, N. P., Rowlands, D. D., Lemoine, F. G., & Williams, T. A. (2003). The 1-Centimeter Orbit: Jason-1 Precision Orbit Determination Using GPS, SLR, DORIS, and Altimeter Data. *Marine Geodesy*, *26(3-4)*, 399-421. doi: 10.1080/714044529

Lyon, R. H., Engineering, A., Cefola, P. J., & Greitzer, E. M. (2004). Geosynchronous orbit determination using space surveillance network observations and improved radiative force modeling. Retrieved from `https://dspace.mit.edu/handle/1721.1/17779`

March, G., Doornbos, E. N., & Visser, P. N. (2019). High-fidelity geometry models for improving the consistency of CHAMP, GRACE, GOCE and Swarm thermospheric density data sets. *Advances in Space Research*, *63(1)*, 213-238. doi: https://doi.org/10.1016/j.asr.2018.07.009

Marshall, J. A., & Luthcke, S. B. (1994). Modeling radiation forces acting on Topex/Poseidon for precision orbit determination. *Journal of Spacecraft and Rockets*, *31*(1), 99-105. doi: https://doi.org/10.2514/3.26408

Mehta, P. M., Paul, S. N., Crisp, N. H., Sheridan, P. L., Siemes, C., March, G., & Bruinsma, S. (2022). Satellite drag coefficient modeling for thermosphere science and mission operations. *Advances in Space Research.* doi: 10.1016/J.ASR.2022.05.064

Mehta, P. M., Walker, A. C., Sutton, E. K., & Godinez, H. C. (2017). New density estimates derived using accelerometers on board the CHAMP and GRACE satellites. *Space Weather*, *15*(4), 558–576. doi: 10.1002/2016SW001562

Millward, G., Moffett, R., Quegan, S., & Fuller-Rowell. (1996). A Coupled Thermosphere-Ionosphere-Plasmasphere Model (CTIP). *STEP Handbook on Ionospheric Models (ed. R.W. Schunk).*

Millward, G. H., Müller-Wodarg, I. C., Aylward, A. D., Fuller-Rowell, T. J., Richmond, A. D., & Moffett, R. J. (2001). An investigation into the influence of tidal forcing on F region equatorial vertical ion drift using a global ionosphere-thermosphere model with coupled electrodynamics. *Journal of Geophysical Research: Space Physics*, *106*(A11), 24733–24744. doi: 10.1029/2000ja000342

Muelhaupt, T. J., Sorge, M. E., Morin, J., & Wilson, R. S. (2019). Space traffic management in the new space era. *Journal of Space Safety Engineering*, *6*(2), 80–87. doi: 10.1016/J.JSSE.2019.05.007

Pavlis, D., Mccarthy, J., Rowton, S., Moore, D., Luthcke, S., Tsaoussi, L., & Beall, J. (2019). GEODYN Systems Description: Volumes 1, 3, and 5 [Computer software manual]. Retrieved from `https://earth.gsfc.nasa.gov/sites/default/files/neptune/files/geodyn_vol1.pdf`

Petit, G., & Luzum, B. (2010). *IERS Conventions (2010)* (Tech. Rep. No. 36). Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main, Germany.

Pilinski, M. D. (2008). Analysis of a novel approach for determining atmospheric density from satellite drag. *ProQuest Dissertations and Theses*, 150. Retrieved from `https://colorado.idm.oclc.org/login?url=https://www.proquest.com/dissertations-theses/analysis-novel-approach-determining-atmospheric/docview/230692345/se-2`

Pilinski, M. D., Argrow, B. M., & Palo, S. E. (2010). Semiempirical Model for Satellite Energy-Accommodation Coefficients. *Journal of Spacecraft and Rockets*, *47*(6), 951-956. doi: https://doi.org/10.2514/1.49330

Qian, L., Burns, A. G., Emery, B. A., Foster, B., Lu, G., Maute, A., … Wang, W. (2014). The NCAR TIE-GCM. In *Modeling the ionosphere–thermosphere system* (p. 73-83). American Geophysical Union (AGU). doi: https://doi.org/10.1002/9781118704417.ch7

Reigber, C., Schwintzer, P., Stubenvoll, R., Schmidt, R., Flechtner, F., Meyer, U., … Raimondo, J. C. (2006). *A High Resolution Global Gravity Field Model Combining CHAMP and GRACE Satellite Mission and Surface Data: EIGEN-CG01C* (Tech. Rep. No. Scientific Technical Report STR ; 06/07). Potsdam : Deutsches GeoForschungsZentrum GFZ, 12 S. p. doi: https://doi.org/10.2312/GFZ.b103-06075

Richmond, A. D., Ridley, E. C., & Roble, R. G. (1992). A thermosphere/ionosphere general circulation model with coupled electrodynamics. *Geophysical Research Letters*, *19*(6), 601–604. doi: https://doi.org/10.1029/92GL00401

Ringuette, R., Zeeuw, D. D., Rastaetter, L., & Pembroke, A. (2023). Kamodo's Model-Agnostic Satellite Flythrough: Lowering the Utilization Barrier for Heliophysics Model Outputs. *Accepted in Frontiers in Astronomy and Space Sciences.* doi: 10.1016/J.ASR.2023.03.033

Schunk, R., & Nagy, A. (2009). *Ionospheres: Physics, Plasma Physics, and Chemistry* (2nd ed.). Cambridge University Press. doi: https://doi.org/10.1017/CBO9780511635342

Shim, J. S., Kuznetsova, M., Rastätter, L., Bilitza, D., Butala, M., Codrescu, M., … Sutton, E. (2014). Systematic Evaluation of Ionosphere/Thermosphere (IT)

Models: CEDAR Electrodynamics Thermosphere Ionosphere (ETI) Challenge (2009-2010). *Modeling the Ionosphere-Thermosphere System*, *9780875904917*, 145–160. doi: 10.1002/9781118704417.ch13

Solomon, S. C., & Qian, L. (2005). Solar extreme-ultraviolet irradiance for general circulation models. *Journal of Geophysical Research: Space Physics*, *110*(A10). doi: 10.1029/2005JA011160

Storz, M. F., Bowman, B. R., Branson, J. I., Casali, S. J., & Tobiska, W. K. (2005). High accuracy satellite drag model (HASDM). *Advances in Space Research*, *36*(12), 2497–2505. doi: 10.1016/j.asr.2004.02.020

Sutton, E. K. (2018). A New Method of Physics-Based Data Assimilation for the Quiet and Disturbed Thermosphere. *Space Weather*, *16*, 736–753. doi: https://doi.org/10.1002/2017SW001785

Sutton, E. K., Forbes, J. M., & Nerem, R. S. (2005). Global thermospheric neutral density and wind response to the severe 2003 geomagnetic storms from CHAMP accelerometer data. *Journal of Geophysical Research: Space Physics*, *110*, 9-40. doi: 10.1029/2004JA010985

Sutton, E. K., Thayer, J. P., Wang, W., Solomon, S. C., Liu, X., & Foster, B. T. (2015). A self-consistent model of helium in the thermosphere. *Journal of Geophysical Research A: Space Physics*, *120*(8), 6884–6900. doi: 10.1002/2015JA021223

Tapley, B., Ries, J., Bettadpur, S., Chambers, D., Cheng, M., Condi, F., … Wang, F. (2005). GGM02 - An improved Earth Gravity Field Model from GRACE. *Journal of Geodesy*, *79*(8), 467–478. doi: 10.1007/s00190-005-0480-z

Tapley, B., Schutz, B., & Born, G. H. (2004). *Statistical Orbit Determination.* Elsevier Academic Press. doi: https://doi.org/10.1016/B978-0-12-683630-1.X5019-X

Thayer, J. P., Tobiska, W. K., Pilinski, M. D., & Sutton, E. K. (2021). Remaining Issues in Upper Atmosphere Satellite Drag. *Space Physics and Aeronomy Collection*, *5*, 111–140. doi: 10.1002/9781119815570.ch5

Thayer, J. P., Waldron, Z. C., & Sutton, E. K. (2023). Solar flux dependence of upper thermosphere diurnal variations: Observed and modeled. *Journal of Geophysical Research: Space Physics*, *128*. doi: 10.1029/2022JA031146

Thomas, T., Luthcke, S., Pennington, T., Nicholas, J., & Rowlands, D. (2021). ICESat-2 Precision Orbit Determination. *Earth and Space Science.* doi: 10.1029/2020ea001496

Tobiska, W. K., Bowman, B. R., Bouwer, S. D., Cruz, A., Wahl, K., Pilinski, M. D., … Licata, R. J. (2021). The SET HASDM Density Database. *Space Weather*, *19*(4), e2020SW002682. doi: 10.1029/2020SW002682

Vallado, D. (2013). *Fundamentals of Astrodynamics and Applications* (Fourth Edi ed.). Hawthorne, CA: Microcosm Press and Springer. doi: 10.2514/2.4291

Velicogna, I., & Wahr, J. (2005). Greenland mass balance from GRACE. *Geophysical Research Letters*, *32*(18), 1–4. doi: 10.1029/2005GL023955

Waldron, Z., Garcia-Sage, K., Thayer, J., Sutton, E., Ray, V., Rowlands, D., … Berland, G. (2023, June). *Data archive accompanying "Assessing Thermospheric Neutral Density Models using GEODYN's Precision Orbit Determination".* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.8015369` doi: 10.5281/zenodo.8015369

Walker, A., Mehta, P., & Koller, J. (2014). Drag coefficient model using the Cercignani-Lampis-Lord gas-surface interaction model. *Journal of Spacecraft and Rockets*, *51*(5), 1544–1563. doi: https://doi.org/10.2514/1.A32677

Walterscheid, R. L., Chen, M. W., Chao, C. C., Gegenheimer, S., Cabrera-Guzman, J., & McVey, J. (2023). Comparative Accuracies of Models for Drag Prediction During Geomagnetically Disturbed Periods: A First Principles Model Versus Empirical Models. *Space Weather*, *21*, e2022SW003332. doi: 10.1029/2022SW003332

Zelensky, N. P., Lemoine, F. G., Ziebart, M., Sibthorpe, A., Willis, P., Beckley, B. D., ... Luceri, V. (2010). DORIS/SLR POD modeling improvements for Jason-1 and Jason-2. *Advances in Space Research*, *46(12)*, 1541-1558. doi: 10.1016/J.ASR.2010.05.008

Assessing Thermospheric Neutral Density Models using GEODYN's Precision Orbit Determination

Z.C. Waldron[1], K. Garcia-Sage[2], J.P. Thayer[1,5], E.K. Sutton[1], V. Ray[1,3], D.D. Rowlands[4], F.G. Lemoine[4], S.B Luthcke[4], M. Kuznetsova[2], R. Ringuette[6], L. Rastaetter[2], and G.D. Berland[5]

[1]CU Boulder Space Weather Technology, Research, and Education Center
[2]NASA Goddard, Space Weather Laboratory, Community Coordinated Modeling Center
[3]Kayhan Space Corp.
[4]NASA Goddard, Geodesy and Geophysics Laboratory
[5]CU Boulder, Department of Aerospace Engineering Sciences
[6]NASA Goddard, Center for HelioAnalytics

Key Points:

- Precision orbit determination solutions are expanded to study satellite drag and assess upper atmospheric density models using GEODYN.
- A proof-of-concept case study assessment of density models is presented using ICESat-2 precise science orbits and orbit fits
- Assessment results are provided for empirical (MSIS2, DTM2020, JB2008) and physics-based (TIEGCM, CTIPe) models for 14-days in November 2018

Corresponding author: Z.C. Waldron, zachary.waldron@colorado.edu

Abstract

This study focuses on utilizing the increasing availability of satellite trajectory data from global navigation satellite system-enabled low-Earth orbiting satellites and their precision orbit determination (POD) solutions to expand and refine thermospheric model validation capabilities. The research introduces an updated interface for the GEODYN-II POD software, leveraging high-precision space geodetic POD to investigate satellite drag and assess density models. This work presents a case study to examine five models (NRLMSIS2.0, DTM2020, JB2008, TIEGCM, and CTIPe) using precise science orbit (PSO) solutions of the Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2). The PSO is used as tracking measurements to construct orbit fits, enabling an evaluation according to each model's ability to redetermine the orbit. Relative in-track deviations, quantified by in-track residuals and root-mean-square errors (RMSe), are treated as proxies for model densities that differ from an unknown true density. The study investigates assumptions related to the treatment of the drag coefficient and leverages them to eliminate bias and effectively scale model density. Assessment results and interpretations are dictated by the timescale at which the scaling occurs. JB2008 requires the least scaling ($\sim -23\%$) to achieve orbit fits closely matching the PSO within an in-track RMSe of 9 m when scaled over two weeks and 4 m when scaled daily. The remaining models require substantial scaling of the mean density offset ($\sim 30{-}75\%$) to construct orbit fits that meet the aforementioned RMSe criteria. All models exhibit slight over or under sensitivity to geomagnetic activity according to trends in their 24-hour scaling factors.

1 Plain Language Summary

This study utilizes the increasing availability of satellite trajectory data from low-Earth orbiting satellites and their precision orbit determination (POD) solutions to expand thermospheric model validation capabilities. We introduce an updated interface for the GEODYN-II POD software to investigate satellite drag and assess density models. This work presents a case study assessment of five models (NRLMSIS2.0, DTM2020, JB2008, TIEGCM, and CTIPe) using precise science orbit (PSO) solutions of the Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2). GEODYN is used to construct orbit fits to the PSO via the five models, enabling an evaluation according to each model's ability to redetermine the orbit. Relative deviations from the PSO, quantified by in-track residuals and root-mean-square errors (RMSe), serve as proxies for model densities that differ from an unknown true density. We investigate and leverage drag coefficient assumptions to eliminate bias and scale model densities. JB2008 requires the least scaling ($\sim -23\%$) to achieve orbit fits closely matching the PSO within an in-track RMSe of 9 m when scaled over two weeks and 4 m when scaled daily. The remaining models require substantial scaling ($\sim 30 - 75\%$) to meet the aforementioned RMSe criteria.

2 Introduction

With the drastic increase in commercial satellite launches, the need to address the challenges posed by satellite drag have come to the forefront of the scientific, operational, and commercial space communities (Muelhaupt et al., 2019; Berger et al., 2020; Thayer et al., 2021; Hejduk & Snow, 2018; Bussy-Virat et al., 2018). Shortly following the 36th launch of SpaceX's Starlink constellation on 3 February 2022, 38 out of 49 satellites were lost due to the impacts of a modest geomagnetic storm that reached G1 intensity earlier that day (Berger et al., 2023; Fang et al., 2022; Hapgood et al., 2022). The satellites were placed into an initial orbit of 210 km after which they were intended to maneuver to an operational altitude of 500 km. While this low altitude plan lent itself to a quick de-orbit in the face of catastrophe, it exposed the satellites to the larger variations and uncertainties in neutral density associated with relatively meager space weather conditions. While this event happened at altitudes well below Starlink's operational orbit, it has served as a potent example to the commercial space community of the need to better model and predict atmospheric drag, which represents the most significant hurdle preventing more accurate determination and prediction of trajectories in LEO.

Precision orbit determination (POD) programs are employed in both operational and research capacities to provide high-fidelity orbit trajectories of LEO satellites. The quality of such trajectories is directly dependent on the ability of a POD's force model to realistically capture the conservative and non-conservative forces impacting a satellite's orbit. Due to advancements in conservative force modeling, the largest source of error preventing more accurate orbit trajectories is now associated with non-conservative forces (Tapley et al., 2005; Reigber et al., 2006; Velicogna & Wahr, 2005). Of these, atmospheric drag is the most variable and uncertain as a consequence of its reliance on modeling the thermospheric neutral mass density (ρ) variations and the satellite drag-coefficient (C_D) (Hejduk & Snow, 2018). The largest source of uncertainty is ρ, but for satellites with complex shapes, C_D can contribute to this uncertainty. Mehta et al. (2022) describes this issue as the interconnectedness of uncertain parameters, an extremely challenging problem to solve for the satellite drag community and one that has significant impact on the assumptions made in this work. The burden for achieving more precise and reliable LEO nowcasting and forecasting largely relies on the ability of thermospheric density models to accurately capture the behavior of neutral density and reliably predict it into the future. Adding to the problem, assessing the performance of density models presents a massive challenge due to the scarcity of data from satellite measurements, and the lack of absolute truth due to the complexity of interconnected uncertainties. This necessitates the community to seek alternative methods to add to the validation method repertoire. The growing prevalence of global navigation satellite system (GNSS)-enabled low-Earth orbiting satellites and their POD solutions represents one such potential data source, and providing methods to take advantage of these datasets will help the community expand and refine model validation capabilities.

POD programs such as the NASA Goddard Space Flight Center's (GSFC) GEODYN II software (henceforth referred to as GEODYN) have been developed within the geodesy scientific community with the above challenges in mind—implementing techniques such as reduced-dynamics paired with extremely high quality tracking measurements from GNSS to mitigate the need for highly accurate non-conservative force models when performing non-predictive orbit determination. Through these means, centimeter level radial accuracy has been demonstrated to produce precise science orbit (PSO) solutions for missions such as the Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2), which orbits at approximately 500 km (Thomas et al., 2021). These techniques—combined with GEODYN's legacy of precise conservative force and measurement modeling, meticulous time systems, and accurate coordinate reference frames—have made the program a top-tier POD tool that is well-positioned to study thermospheric neutral density models and their distinct impacts on the estimation of satellite drag (Luthcke et al., 2003; Zelensky et al., 2010; Lemoine et al., 2016; Loomis et al., 2019). This work aims to provide a method to improve the specification of satellite drag physics and the assessment of neutral density model performance to help the Ionosphere-Thermosphere (IT) community advance model predictions, and consequently improve the accuracy of POD solutions.

This paper presents the development of a modernized Python interface for the GEODYN software, leveraging the high-precision nature of space geodetic POD, but refashioned to study satellite drag and to enable density model assessment. We make use of the well-specified, low-error ICESat-2 PSO to perform a case study assessment of five thermospheric density models, three of which are empirical while the other two are physics-based. The ICESat-2 PSO serves as tracking measurements to POD-based orbit fits in which the drag effects from density models are assessed according to each model's ability to redetermine the orbit. Implications regarding the treatment of the drag coefficient are investigated and discussed. This work reports an initial result using a fixed drag coefficient of $C_D = 2.5$, followed by two methods for debiasing the assessment results using a drag acceleration scaling factor over both a two-week and a daily time interval. Each model's orbit fit contains relative in-track deviations, quantified by in-track residuals and root-mean-square errors from the ICESat-2 PSO, which are treated as proxies for model densities that differ from a true, unspecified density. By developing these methods, we aim to provide the community with the means to take advantage of emerging GNSS-tracked satellite datasets and POD solutions to objectively quantify density model performance. In addition, we hope to address deficiencies in non-conservative force modeling that may currently impede higher quality predictions of LEO trajectories. The presented model assessment results will be parsed into the Community Coordinated Modeling Center's Comprehensive Assessment of Models and Events using Library Tools (CAMEL) framework, for community use.

Section 3 gives the necessary science background needed to understand our methodology. Section 4 details the GEODYN software, provides information regarding the ICESat-2 POD solutions, and offers an overview description of the upper atmospheric density models that are assessed in this work. Section 5 details the methodology, the setup procedure for conducting the model assessment, and the methods for debiasing the assessment results using drag acceleration scaling factors. Section 6 provides the results and discussion of the assessment using ICESat-2 PSO as a case study.

3 Background

The precision of a POD solution relies on the fidelity of the tracking measurement models, the quality of the tracking data, and the ability of the POD force model to capture realistic accelerations acting on the satellite. In general, the force model defines the overall motion of a spacecraft by calculating the sum of all impacting forces, themselves being subdivided into conservative forces which are potential in nature, and non-conservative forces which act to dissipate the satellite's orbital energy. Conservative forces captured by the GEODYN force model include the Earth's static gravity field (geopotential), solid Earth and ocean tides, the effects of dynamic polar motion, the acceleration from time variable gravity, Third-body perturbations (primarily from the Sun and Moon), and contributions from general relativity. Recent improvements in conservative force modeling as well as advances in the internal measurement models have shifted the primary source of error in POD solutions to the non-conservative forces (Luthcke et al., 2006; Loomis et al., 2019; Reigber et al., 2006). The non-conservative forces modeled in GEODYN are atmospheric drag, solar radiation pressure (SRP), and Earth radiation pressure (ERP). As altitude decreases in the LEO regime, atmospheric drag increasingly becomes the largest non-gravitational force acting on satellites. In addition, the drag force's dependence on the upper atmospheric neutral mass density makes it the most error-bound perturbing force (Hejduk & Snow, 2018). While force model errors can be circumvented via reduced-dynamics and high-quality tracking measurements, this technique is limited in its application for the eventual goal of orbit prediction, which requires an improved, more-realistic force model (Tapley et al., 2004; Luthcke et al., 2019).

The drag force acting on a satellite of mass m_{sat} is proportional to the atmospheric neutral mass density ρ, the drag coefficient C_D, the projected area perpendicular to the flow direction A_{sat}, and the velocity of the satellite relative to the atmosphere $\vec{V}_{rel}$. The drag acceleration $\vec{a}_D$ due to the drag force per unit mass acting on a satellite is given in Equation 1 as

$$\vec{a}_D = -\frac{1}{2}\,\rho\, C_D\, \frac{A_{sat}}{m_{sat}}\, V_{rel}^2 \frac{\vec{V}_{rel}}{V_{rel}} \tag{1}$$

Physically, the total drag force acting on a satellite surface is given by the force due to incident atmospheric particles impacting the surface combined with the force from scattered particles departing from the surface. These effects are represented by the drag coefficient C_D, which depends on a satellite's geometry and orientation, the material and surface temperature of the spacecraft, the local atmospheric composition, and gas-surface interactions and other effects (Bernstein & Pilinski, 2022). In the context of spacecraft dynamics, the C_D is generally characterized as either fixed, fitted, or physical. Fixed C_D uses a predetermined value that does not change. Fitted C_D is derived using some form of a fitting or filtering process and is typically updated over time (every few hours or orbits). Physical C_D is computed by modeling the momentum and energy exchange between the flow-field particles and the satellite (see Mehta et al. (2022) for more details). If not physically calculated, C_D's presence in Equation 1 may be thought of as a scaling factor that effectively serves to average out errors in the atmospheric density model and gas-surface interactions. In its base state, GEODYN can use either a fitted or fixed C_D. In the fitted case C_D is an adjustable parameter that accounts for mismodeled physics and for uncertainties in ρ associated with the upper atmospheric density model.

Earth's upper atmosphere is driven by a broad range of external energy inputs, leading to complex thermal, electromagnetic, and chemical processes that result in a thermospheric neutral mass density ρ that is highly dynamic and whose variability is difficult to specify (Emmert, 2015). Upper-atmospheric density models are employed within POD force models to represent the complex behavior of ρ when calculating the force of satellite drag acting on a spacecraft, directly or indirectly through C_D. The three types of density models most commonly used by upper atmospheric communities are semi-empirical, physics-based, and data assimilative models. The simple yet effective semi-empirical models are most commonly employed in POD force models since they offer excellent climatological pictures of upper atmospheric variability and are computationally inexpensive. Physics-based models are more complex, taking the form of general circulation models which solve the first-principle equations that govern the coupled thermosphere-ionosphere system. They are not typically used in POD geodetic settings due to the computational expense. A data assimilative technique can be used to calibrate modeled density and has given rise to data assimilative (also referred to as dynamically calibrated) models. These combine analyses from a multitude of space objects to produce corrections to empirical (and occasionally physics-based) thermospheric models. The most prominent example of assimilative thermospheric density models is the United States Space Force, High Accuracy Satellite Drag Model (HASDM) (Storz et al., 2005). It is a common practice in the IT modeling community to compare model performances against HASDM outputs since it performs real-time calibration using ~ 75 space objects.

Different models, and even model types, have varying degrees of performance under specified conditions. Individual model performances are known to depend greatly on the solar flux and geomagnetic conditions that drive them, and their respective strengths make some models better qualified for some scenarios than others. Semi-empirical models are often computationally fast and accurate for climatological uses, but their ability to accurately project into the future is closely tied to the fidelity of their drivers. Physics models offer great potential for forecasting, but lack the accuracy of semi-empirical models in near real-time scenarios (Shim et al., 2014; Sutton, 2018). The vast range in model performances makes the evaluation of models a critical goal for upper atmospheric science and satellite drag communities. The scarcity, and coupled uncertainty (via C_D uncertainty) of thermospheric density measurements makes this a significant challenge. The most common method for objectively quantifying a density model's performance is to compare the sampled model outputs against satellite measurements, e.g. see Walterscheid et al. (2023)—usually in the form of accelerometer-derived densities from the Challenging Minisatellite Payload (CHAMP), Gravity Recovery and Climate Experiment (GRACE) or Gravity Field and Steady-State Ocean Circulation (GOCE) missions (Bruinsma et al., 2004; Sutton et al., 2005; Doornbos et al., 2010; Mehta et al., 2017).

In a series of papers motivated to provide community organization for conducting model comparison and evaluation, Bruinsma et al. (2017, 2018, 2021) provide commonalities for inter-model scoring. They report on chosen observed density datasets, time periods of interest, and provide a scoring metric in the form of the mean, standard deviation and root mean square error (RMSe) of the observation-to-model density ratios. He et al. (2018) similarly presents an assessment of several semi-empirical thermosphere models, focusing on their ability to reproduce spatial variations and capture complex features in thermosphere mass density. Shim et al. (2014) provides a systematic evaluation of thermospheric and ionospheric models, quantifying model performance using four skill scores calculated as functions of geomagnetic activity and geographic latitude: RMS error, prediction efficiency, ratio of maximum-to-minimum, and ratio of maximum amplitude. Thayer et al. (2023) investigates the use of the day-to-night density ratio as a metric for representing the atmosphere's response to large scale perturbations (i.e. the transition from solar maximum to solar minimum), providing inter-model and model-to-observation comparisons, and unearthing discrepancies that are not observed between models and observations when viewed using more common metrics. Each of these reports makes use of the accelerometer-derived density data sets to objectively quantify model performance.

Through this work, we aim to contribute an additional method to the community in which accurately developed and well-honed POD tools can be leveraged for assessing density model performance. For the purposes of this paper, we make a distinction when referring to the different stages of model assessment. We use the term "assessment" to refer more generally to methods and results that offer insight into model performance. "Verification" refers to using other well-specified methods and datasets to confirm the fidelity of our methods and results. "Validation" refers to the act of objectively quantifying modeled densities against observed/derived values. This paper offers a verification of our method and results by comparing against the HASDM densities, and provides an example performance assessment using two-weeks of the ICESat-2 PSO as a case study. A more formal validation scheme is the eventual goal of this work, however, this requires additional considerations and is a source of continuing effort.

4 Program and Data Descriptions

4.1 GEODYN and the Pygeodyn Wrapper

The GEODYN-II program is a precision orbit determination and parameter estimation tool that has been used on every NASA geodetic Earth and planetary altimeter mission since 1985. The program is used extensively for orbit determination, geodetic parameter estimation, tracking instrument calibration, satellite orbit prediction, as well as for many other applied research studies in satellite geodesy (Pavlis et al., 2019; Luthcke et al., 2019). GEODYN is capable of ingesting essentially all types of tracking measurements, the most common of which include observations from global navigation satellite systems (GNSS) and satellite laser ranging (SLR), as well as post-processed orbits in the form of orbit trajectories or precisely converted elements (PCE) (Pavlis et al., 2019; Lyon et al., 2004). GEODYN performs orbit propagation using Cowell's method of numerical integration, and performs data-reduction utilizing a Bayesian least-squares batch estimation process to optimally estimate parameters by minimizing the residuals between tracking data and orbit propagations (see Vallado (2013) for more information). GEODYN's long history in geodetic applications has ensured the development of very precise conservative force and measurement models, as well as accurate time systems and coordinate reference frames, making the program a top-tier POD tool. With this understanding, the errors found in the observed residuals between tracking data and determined orbit are more related to uncertainties in the satellite specific non-conservative force models, rather than being related to the quality of measurement modeling or orbit determination methods and tools. In the lower register of LEO, where atmospheric drag dominates, the observation residuals can provide valuable information on the drag model errors.

Pygeodyn is an internally-developed Python-based wrapper meant to offer improved user access to the FORTRAN-based GEODYN software. Pygeodyn offers users a streamlined and simplified tool to navigate the complex steps for modifying, controlling, running, and reading the various data sets and files that compose the GEODYN program. The main portion of GEODYN II is composed of two sequenced programs: GEODYN-IIS, a scheduling program and GEODYN-IIE, an execution program stage. The scheduling program reads and organizes input data, ancillary data files, and the user's setup options. The execution program then integrates the satellite trajectory and applies the selected models, performs orbit determination to provide computed observables, and uses the least squares scheme, along with any measured observables, to provide solutions for updated orbits as well as any requested geophysical parameters. The two stages communicate via a series of binary files which are output from the scheduling program and fed into the execution program. Historically, adding atmospheric density models to GEODYN required modification to IIS as well as subsequent data tracking and modification to IIE, a series of complications that have been circumvented with our Pygeodyn tool. Pygeodyn gives the ability to switch between different atmospheric density models that have been connected to GEODYN-IIE without the need to modify GEODYN-IIS, simplifying the user experience for adding and selecting the models. Programming in Python has also afforded Pygeodyn the ability to interface with the NASA Goddard Community Coordinated Modeling Center's (CCMC) Kamodo API (Ringuette et al., 2023), granting access to their sophisticated model readers and allowing Pygeodyn to connect physics-based density model outputs to the POD scheme.

4.2 ICESat-2 PSO Solutions as Tracking Data

ICESat-2 flies in a near-circular, near-polar, low-Earth orbit at $\sim$496 km altitude and an orbital period of 94.22 min. Details of the orbital parameters are reported in Luthcke et al. (2019). The ICESat-2 PSO (i.e., the science quality POD solutions), and their corresponding setup files, are provided by the Geodesy and Geophysics Laboratory within NASA/GSFC, who maintain the GEODYN program and provide science quality POD for many NASA missions. ICESat-2 is an excellent platform for orbital drag-based model assessment because of its science requirements to have such high quality orbit solutions, as well as stable attitude specifications. The ICESat-2 PSO is reported by Thomas et al. (2021) as having a radial orbit accuracy of just below 1.5 cm over a 24-hour orbit solution—performing better than the mission requirement of 3 cm. These orbit solutions are generated through the reduction of GNSS double-difference carrier phase observable residuals, and independently assessed using SLR measurement residual analysis. Technical details regarding the construction and analysis plan for the ICESat-2 PSO can be found in Luthcke et al. (2019).

The centimeter-level orbit of the PSO data is achieved using the previously mentioned reduced-dynamics technique in which GEODYN solves for empirical acceleration parameters that describe the difference between the actual positions, i.e. those derived from the GPS tracking measurements, and the positions that are calculated by the program's physical force models and satellite propagator. The PSO data includes estimations of along-track and cross-track empirical accelerations every quarter of an orbit, applying a nearest neighbor covariance constraint. With the use of reduced-dynamic empirical accelerations, it is possible to compensate for errors associated with using the MSIS86 model to calculate the effects of atmospheric drag. Luthcke et al. (2019) notes that even though a reduced-dynamic approach is commonly employed by the geodesy community to overcome any inadequacies in a force model, the technique relies on an orbit solution that has already attained sufficient radial accuracy through the use of a high-quality physical force model. Dense tracking measurements and the reduced-dynamic technique do not obviate the use of accurate orbit modeling, and improvements in the orbit fit will be realized when the force models are improved. We also note that while MSIS86 was used to estimate the drag accelerations in the ICESat-2 PSO, due to them being combined with additional empirical accelerations in the along-track and cross-track directions no related bias is found that favors the MSIS series of models in the assessments reported in this work.

This work uses the ICESat-2 PSO as tracking measurement input to a data-reduction run of GEODYN—the goal being to assess the ability of each selected density model to re-determine the orbit of ICESat-2. A data-reduction run in GEODYN is one in which orbit parameters (i.e., initial conditions) and optionally geophysical parameters (such as gravitational coefficients or the drag coefficient) are adjusted to minimize residuals and provide an improved solution. This data reduction is computed over an orbital arc, a set time period for which continuous tracking data is available. The term "orbit fit" refers to the outputs of GEODYN runs in which the ICESat-2 PSO is the tracking data type and respective density models are used to iteratively re-determine the orbit.

The following capabilities for density model assessment are enabled by using GEODYN to construct orbit fits from ICESat-2 orbit solutions:

1. Leverage GEODYN's high fidelity physical force models which have been honed by the program's long legacy in space geodesy.
2. Perform data-reduction runs in which we compare the relative ability of each atmospheric density model to re-determine the orbit of ICESat-2 given the isolated satellite drag effects.
3. Control the POD and force model parameters such that for each respective run, the only relative variable impacting the overall fit of the orbit solution for a given arc is the atmospheric density model used to estimate the drag term.
4. Control for relative errors between runs associated with an unknown drag coefficient by using a realistic fixed value of $C_D = 2.5$. This value is determined by physically calculating C_D using the Diffuse Reflection with Incomplete Accommodation (DRIA) method along the orbit of ICESat-2, as is described in Section 5.

4.3 Model Descriptions

This section provides a brief overview of the atmospheric density models that are used for verification and assessment. Table 1 lists the models, providing the Model ID used for referencing in this paper, the full name and version number, the run conditions based on the drivers, and the models' spatial and temporal resolutions. The authors acknowledge that while there are a number of ways to improve a density model's outputs at runtime (see Sutton (2018); Shim et al. (2014)), the outputs used in this work are intended to reflect typical community use, with each model being run according to the developer's operational instructions. Additional information regarding each model can be found in the references provided in the second column of Table1.

We provide a verification using SET-HASDM, a data-assimilative model, and assessment results for MSIS2, JB2008, DTM2020, TIEGCM, and CTIPe. The semi-empirical models (MSIS2, JB2008, DTM2020) are interfaced directly into GEODYN's FORTRAN-based source code. The physics-based models (TIEGCM and CTIPe) are interfaced to GEODYN via the CCMC's Kamodo program which reads and interpolates the model output files. These interpolated outputs are connected to GEODYN through the Pygeodyn wrapper using an orbit cloud interpolation technique which is detailed in Appendix C. In addition, physics-based models whose maximum altitude is below the orbit altitude of ICESat-2 include a diffusive equilibrium extrapolation of the neutral densities (see Chapter 10 of Schunk and Nagy (2009)). The use of Kamodo makes the analysis techniques in this paper easily extensible to additional models. Any thermospheric model that is supported by Kamodo, with the appropriate diffusive equilibrium extrapolation, can be added to this and similar analyses in the future.

Models Assessed in this Report

Model ID	Full name/version	Drivers, (solar $\mid$ geomagnetic)	Resolution, (spatial $\mid$ time)
Semi-Empirical			
MSIS2	Naval Research Laboratory Mass Spectrometer and Incoherent Scatter (NRLMSIS 2.0), (Emmert et al., 2021)	$F_{10.7} \mid Ap$	
JB2008	Jacchia-Bowman 2008, (Bowman et al., 2008)	$F_{10.7}$, $S_{10.7}$, $M_{10.7}$, $Y_{10.7} \mid Ap$, Dst	
DTM2020	Drag Temperature Model 2020, operational mode (Bruinsma & Boniface, 2021)	$F_{10.7}$ and K_p	
Physics-Based			
TIEGCM	National Center for Atmospheric Research (NCAR) Thermosphere-Ionosphere-Electrodynamics General Circulation Model (TIEGCM; version 2.0), (Richmond et al., 1992; Qian et al., 2014; Sutton et al., 2015)	$F_{10.7}$, EUVAC proxy model (Solomon & Qian, 2005) $\mid$ Kp via the Heelis model (Heelis et al., 1982),	$5° \times 5°$, vertically specified by logarithmic pressure surfaces in half-scale height increments from ∼97 km to ∼800 km $\mid$ 1 minute time step, hourly output
CTIPe	Coupled Thermosphere Ionosphere Plasmasphere Electrodynamics (CTIPe; version 4.1) (G. Millward et al., 1996; G. H. Millward et al., 2001)	$F_{10.7}$, solar wind inputs via magnetic field, velocity, and density measurements from Advanced Composition Explorer (ACE) $\mid$ Kp, hemispheric power index from NOAA Polar Orbiting Environmental Satellite (POES)	2° lat. $\times$ 18° lon., vertically specified by 15 logarithmic pressure surfaces from ∼80 km to ∼500 km $\mid$ 1 minute time step
Data-Assimilative			
SET-HASDM	Space Environment Technologies (SET) HASDM Database (Tobiska et al., 2021), which is derived from the US Space Force operational archive (Storz et al., 2005)		10° lat. $\times$ 15° lon., 25 km altitude steps from 175 to 825 km. $\mid$ 3 hour time step

Table 1. A summary of the thermospheric density models used in this assessment.

5 Methodology

5.1 Setup for ICESat-2 Case Study

This method uses a satellite's PSO as tracking measurements to construct dynamic POD-based orbit fits from different density models. The dynamic POD technique uses a batch least-squares approach to iteratively reduce errors between the propagating orbit fit and the ingested PSO—GEODYN refers to this as data-reduction mode. The initial conditions, and any other adjustable parameters, are iteratively estimated and updated to refine the orbit fit until it consistently reaches a convergence threshold. The remaining errors that persist between the PSO and a given model's orbit fit are understood to be primarily due to atmospheric drag effects from the respective density model. This understanding is leveraged to investigate density model performance through the assessments that are presented in this paper. Figure 1 provides a visualization showing connections between the high-level datasets and processes. While the true density along the ICESat-2 orbit remains unknown, each model's orbit fit contains in-track deviations from the ICESat-2 PSO, which are treated as proxies for model density deviations from the true density.

The GEODYN run setup that is used to construct the POD-based orbit fits is kept as similar as possible to the setup used by the team at NASA-GGL to produce the ICESat-2 PSO—meaning we modify only what is necessary to use PSO as the tracking measurement type, and to control the procedure such that drag is the only independent variable in each model's run. An extended overview of GEODYN's setup and force model parameters for the model assessment runs is provided in the appendix in Table B1, with only the most impactful considerations being discussed here. In addition to each orbit fit using the same background force models, the ICESat-2 external attitude information is also utilized to properly orient the spacecraft body and the solar array. The orbit fits are split into 24-hour, consecutive daily arcs. The arc length can theoretically be much shorter; however, orbit errors related to force model perturbations (i.e., drag) require propagation time to accumulate. An arc length on the order of 1-2 orbital periods may not depict substantial trajectory deviations in the residuals, making 24-hour arcs a balanced choice to demonstrate this assessment method. In theory, reducing the arc lengths would provide more RMSe values over shorter times and would offer higher temporal resolution towards understanding model performance, but at the cost of having accumulated less orbital error from the density model in the shorter propagation time. The choice of arc length and its ramifications on assessment results continues to be an area of study related to this work.

Other non-conservative forces that must be considered in addition to atmospheric drag are SRP and ERP which are both calculated by GEODYN according to the descriptions shown in Table B1, and the references therein. Using GEODYN and its high-fidelity force model ensures that the estimated SRP and ERP accelerations are more precise than what would be modeled by a standard satellite flythrough scheme. For each density model's orbit fit the acceleration due to drag will vary according to the error in the respective model, while the contributions from SRP and ERP will remain consistent for each arc across each model run. For the orbit fits presented in this work, the variations of SRP and ERP were found to be small relative to the variable effects of drag. This being said, the magnitude of the SRP acceleration is often on par with that of the drag acceleration at the ICESat-2's altitude. Errors related to mismodeling non-drag forces can potentially be transferred into the residuals, and as a result this method is presented as a relative assessment between controlled model runs rather than an absolute validation of performance.

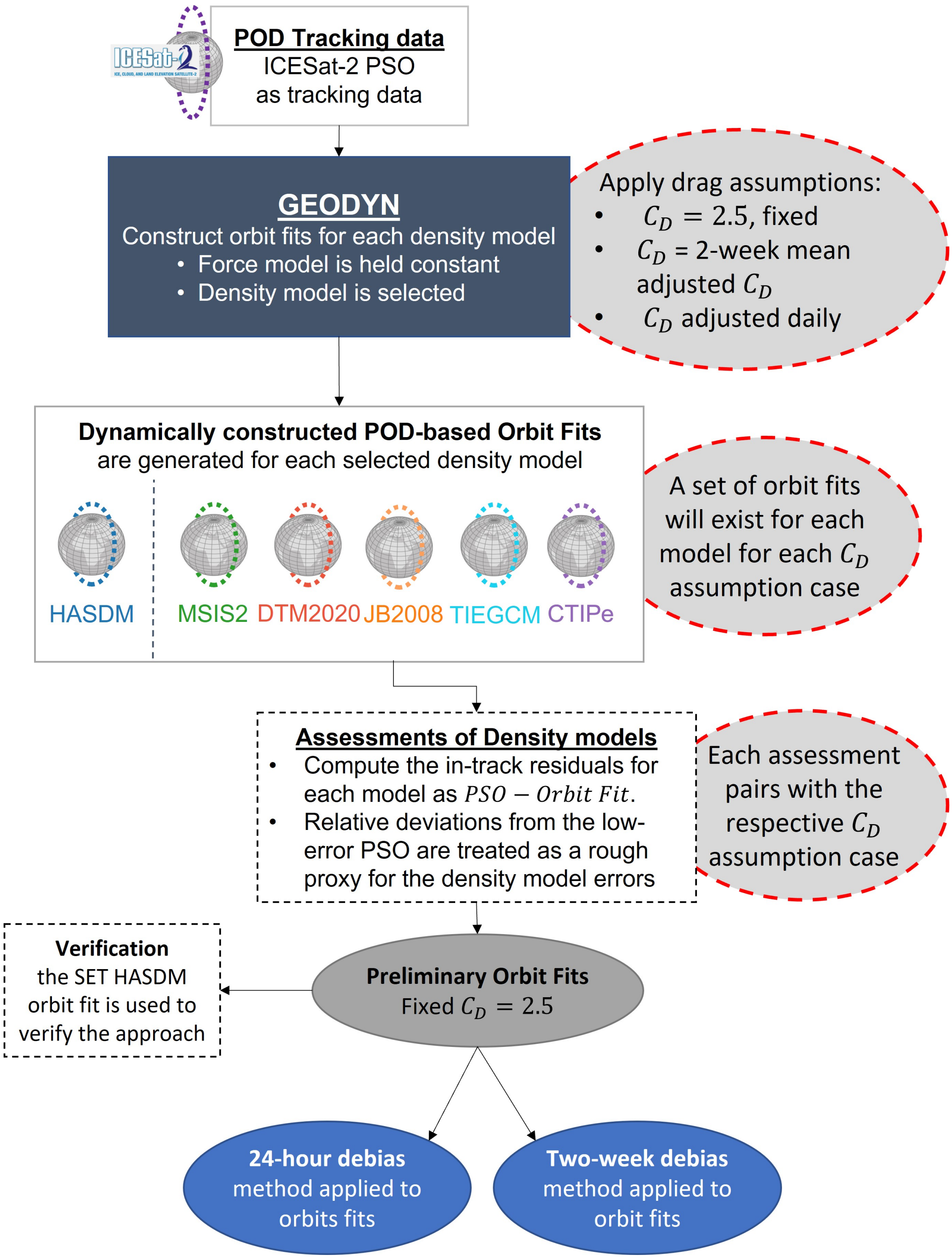


Figure 1. A flowchart visualizing the assessment process and how the datasets and POD methods fit together for model assessment.

5.2 Assessment Procedures

The interconnected, uncertain nature of C_D and ρ makes the absolute determination and assessment of either value a very complex problem that is still an active field of research within this community. In this case study, the assessment procedure is split into three subsequent orbit fit methods, each based on assumptions made to characterize C_D when calculating the drag acceleration for each model during the orbit fit procedure. In a preliminary orbit fit method, explicit biases are identified via orbit fits constructed using a fixed C_D that is held constant across all models. The constant fixed C_D is chosen to be a physically realistic value of $C_D = 2.5$ based on the average result calculated from the Diffuse Reflection with Incomplete Accommodation (DRIA) model along the ICESat-2 orbit. This demonstrates that the method is sufficiently sensitive to recognize differences in the drag effects between the models and provides an understanding of each model's approximate mean density offset relative to the ICESat-2 PSO In an alternative orbit fitting method, GEODYN's parameter estimation procedure is used to adjust the C_D for every 24-hour arc for the two-week period from the a priori estimate of $C_D = 2.5$. The mean-adjusted C_D over the two-weeks is then used as a fixed, model specific value that is constant for the time-period. This provides a fixed, unique C_D for each density model that effectively scales the density over the two weeks to account for each model's mean density offset and examine the model response to solar and geophysical dynamics. In a final orbit fitting method, the daily C_D adjustments are used without a two-week averaging. The residuals using these daily, model-specific C_D adjustments provide an assessment of model performance on time periods less than a day.

The preliminary orbit fits use a fixed C_D of 2.5 that is constant with respect to each model. This enables direct model comparison, but subjects an assessment of the density models to explicit biases depending on each model's density offset relative to this C_D value. Each model's sampled densities along the ICESat-2 orbit have an overall mean-density offset relative to each other. Fixing the C_D to a specific value will cause a particular offset amount to be favored. For instance, $C_D = 3.5$ will produce favorable orbit fits for models that trend a lower density, whereas $C_D = 1$ would favor models that trend towards higher densities. Due to these circumstances, the DRIA model is independently used to calculate a physically realistic value of the ICESat-2's C_D along its orbit. DRIA is a relatively simple, computationally fast model for capturing the gas-surface interactions between the upper atmosphere and a spacecraft. In the DRIA model, particles are always reflected with a diffuse angular distribution, but their energy exchange with the surface varies depending on the value of the energy accommodation coefficient α. This work uses the Sentman's closed-form solutions for the DRIA model as depicted in Equation 12 of Walker et al. (2014). The energy accommodation is assumed to be fixed at $\alpha = 0.89$—a tenuous assumption based on the limited empirical data for α near 500 km during solar minimum (Pilinski et al. (2010); Pilinski (2008)). This α value is likely higher than is realistic for this altitude and solar flux, therefore providing a lower limit for what a physically realistic drag coefficient might be; however, given the complex changes in atmospheric structure that occur in this altitudinal regime, this empirical value is still the most representative until further observations can be made. There are other physical C_D models that could be used instead of DRIA, but choosing and assessing the C_D models quickly expands beyond the scope of this work. For the sake of being able to conduct a model assessment as a proof-of-concept in this case study, this assumption is made with the intention to improve the treatment of C_D in future efforts. In future work, we aim to address this issue by implementing a physical satellite gas-surface interaction method to calculate the time-dependent drag coefficient, but even this will have associated assumptions and caveats. Constructing orbit fits with a fixed, common C_D of 2.5 for each model represents the type of method that is possible without being able to model the physical drag coefficient or without GEODYN's capability to adjust the parameter. This adjustment procedure was performed for different a priori C_D values and found that the final adjusted C_D for each model was consistently the same.

The 24-hour debiasing method uses GEODYN's parameter estimation capabilities to determine a daily fitted value of the C_D that accounts for accumulated errors from the force model over the 24-hour arc—the most prominent of which being due to density uncertainty. In the field of space geodetic POD, C_D is often adjusted in conjunction with reduced-dynamic empirical accelerations to account for disagreement between the observed accelerations from tracking measurements, and calculated accelerations from uncertainties in the drag force model. This technique is used to get very low error, precise orbit solutions, but limits the ability to distinguish errors that are specific to drag or the density models. By allowing only the C_D to adjust and match the orbit fit's modeled accelerations with the PSO observation, density errors over the 24-hour period are incorporated into the adjustment. A density model that is found to be over-/under-estimating the density, will have a C_D that is adjusted to be smaller/larger in a non-physical way—effectively using C_D as a scaling term between the PSO observation and uncertainty in the density model orbit fits. In practice, the C_D also absorbs any errors from mismodeled forces, but these are held constant in the model-to-model comparison. Each model is given an a priori estimate of $C_D = 2.5$ at the start of the 24-hour arc, which is allowed to adjust within a standard deviation of 10. The drag coefficient fitting occurs concurrently with the iterative orbit fit routine. Due to this non-physical use of C_D to effectively debias the density, the term "drag acceleration scaling factor" is adopted. The 24-hour scaling factor for each model (m) can be calculated for each arc (i) as,

$$f_{24,m,i} = C_{D,adj,m,i}/2.5 \tag{2}$$

The two-week debiasing method acts as a combination of the previous two methods. The C_D adjustments for each model are averaged over the two-week period to provide a mean adjusted C_D. Each model's orbit fit is then re-determined using the mean adjusted C_D for each respective model as the fixed value for the two-week period. This assessment permits a scaling of the density models over an extended period of time, highlighting errors in the orbit fits that are due to variations that take place on a longer time scale than 24-hours. This method is also motivated by the need to provide a scoring metric for each density model that can be parsed into the CCMC's CAMEL model validation infrastructure. While the $C_D = 2.5$ case is dominated by the model biases and the 24-hour debiased case demonstrates a method to debias daily densities, the two-week debiased case quantifies the ability of the models to capture dynamics caused by geomagnetic and solar activity over a more prolonged time period. This is a method that could be used in the future to assess model performance during individual stormtime periods.

5.3 Assessment Metrics

Using a PSO as tracking data makes use of GEODYN's data-reduction mode combined with a dynamic technique for estimating the orbit of a satellite. This technique uses the trajectory input to estimate updates to the initial conditions which define the motion of the satellite, thus refining the orbit. The orbit residuals obtained in this setup are the absolute differences between the PSO and each density model's orbit fit. Since other force model parameters are held constant between each density model's run, the inter-comparison of the residuals contains information primarily corresponding to relative errors in each density model's ability to replicate the drag effects seen in the ICESat-2 PSO.

To best observe satellite drag effects, all output orbits are transformed from the J2000, geocentric inertial reference system to the NTW, orbit-aligned satellite coordinate system (Vallado, 2013). This system is composed of an in-track component $\hat{T}$ that is parallel to the velocity vector, a normal component $\hat{N}$ that is perpendicular to the velocity and nominally in the radial direction, and a cross-track component $\hat{W}$ that is normal to the orbit plane and completes the right-hand coordinate frame. The in-track component $\hat{T}$ is parallel to the velocity vector direction and contains any indication that the spacecraft's trajectory has changed since orbital energy dissipations from drag will impact in the velocity direction. Information regarding this transformation as well as supporting coordinate frame details can be found in Appendix A.

For any given arc, y_o is defined to be a component of the orbit from the PSO dataset in the NTW frame, and y_m to be the orbit fit for each density model m. The residuals for each component of the orbit and orbit fit are then calculated (in terms of the in-track component) as,

$$R_{m,T} \quad = \quad y_{o,T} - y_{m,T} \tag{3}$$

The root-mean-square error (RMSe) of the residuals represents the square root of the variance of the absolute difference in the two orbits, indicating how well the density model's orbit fit matches the PSO for that arc. For the in-track component, this is computed for every i^{th} time step of an arc with n time steps as,

$$\text{RMSe}_{m,T} = \sqrt{\frac{1}{n}\sum_{i}^{n}\left(y_{o,T,i} - y_{m,T,i}\right)^2} \tag{4}$$

Figure 2 provides an example of the observation residuals for sample orbits over four,

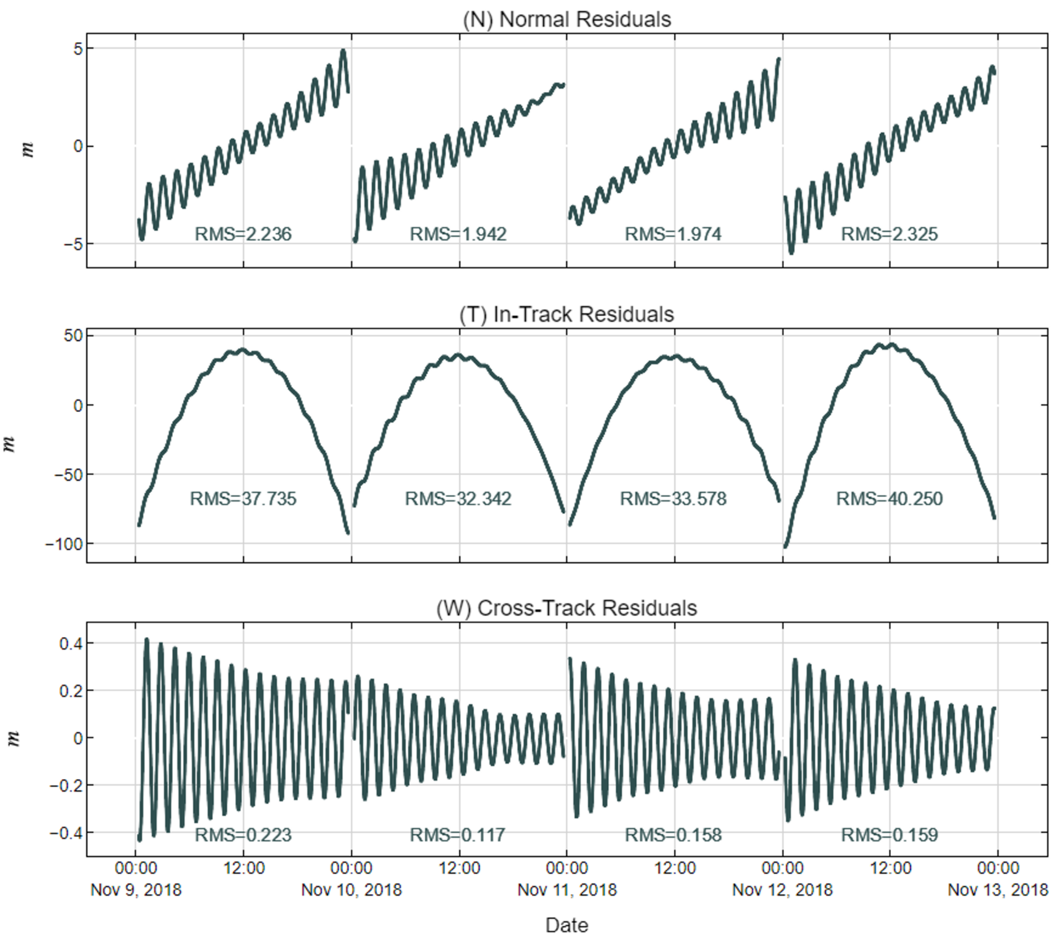


Figure 2. Depicted here are example observational residuals for each component of the NTW system across four 24-hour arcs using the MSIS2 model. The RMSe for each arc/component is given under each curve, showing that the majority of the residual variance (i.e. the orbit error due to drag) is contained within the in-track direction. Note that the vertical scale is different for each plot.

24-hour arcs. The normal $\hat{N}$ (top panel) and cross-track $\hat{W}$ (bottom panel) residuals are included in this figure only to demonstrate that the in-track component contains the majority of the variance associated with residuals in this reference frame. The RMSe values for each component and arc are included as an overlay.

During the POD process, GEODYN iteratively minimizes the discrepancies between the observed orbit (i.e., PSO) and computed orbit (i.e., orbit fits from a density model) across the entire 24-hour arc by adjusting the initial conditions to converge towards a computed trajectory. Since the minimization occurs across the entire arc, the resulting residuals take on the non-linear shapes shown in Figure 2. On a given arc and when comparing the resulting orbit fits from each density model, the only variable that has been permitted to impact each orbit fit's performance relative to the PSO is the drag effects from the selected density model. Therefore, we reason that the relative differences in the residuals for each orbit fit is indicative of density model performance. Other potential errors from mismodeled physics may persist in the residuals, but they are held constant between each model run and will impact the orbit fits consistently. If we run the same arc using the same force model and conditions, but only change the density model used to calculate satellite drag accelerations, the residuals will contain the errors related to the program attempting to reconcile errors in the density model. The RMSe for each orbit fit represents a single value for how well the program can reconcile each model's errors in density over the entire 24-hour arc.

As described in Section 4.2, the ICESat-2 PSO has been shown to have a radial orbit accuracy of below 1.5 cm, generated through the reduction of GNSS double-difference carrier phase observable residuals and independently assessed using SLR measurement residual analysis (Thomas et al., 2021). The precision of the orbit solutions were also verified in all three components using orbit overlap analysis. Given this, relative deviations from the low-error PSO are treated as a rough proxy for the density model errors relative to some unknown true density. The true density value is obscured by the various interconnected unknowns of C_D, SRP, and ERP and therefore remains unspecified. Over the course of a single arc, drag forces from each density model dissipate the satellite's orbital energy at distinct rates, resulting in drag accelerations that are either greater or less than what is represented by the in-track position of the PSO. A strongly negative in-track residual indicates a modeled density that is larger than truth, while a strongly positive in-track residual indicates a modeled density that is smaller than truth. Additional details regarding the shape of the in-track residuals and the relationship between in-track position of the PSO and orbit fits and the density can be found in Figure A2 of Appendix A.

The RMSe is the standard deviation of the residuals and serves as a measure of the difference between a respective orbit fit and the PSO over a single whole arc. Theoretically, an in-track RMSe of zero would mean no difference between an orbit fit and the PSO, indicating near-perfect agreement on average between the modeled density and the POD-based true density across the 24-hour arc. In this setup, perfect agreement for any model is unlikely since the residuals may additionally contain errors related to mismodeled forces, as well as bias/offsets related to fixing the C_D to a common value for all models. A further limitation of the metric is that the RMSe lacks information regarding timescales less than the arc length, and is unsigned, meaning it does not indicate if the modeled density is above or below the truth for a given arc. For these reasons, the in-track residuals and their respective RMSe values are assessed in conjunction with each other.

6 Results and Discussion

This section is organized as follows: (1) the preliminary method for orbit fit construction using a fixed $C_D = 2.5$ is presented, and the orbit fit method is verified using the SET-HASDM density database to determine baseline understanding; (2) an assessment of the semi-empirical and physics-based models is presented via orbit fit results that are debiased using a mean drag acceleration scaling factor over the full two week period; (3) an assessment of the semi-empirical and physics-based models is presented via orbit fit results that are debiased using a 24-hour drag acceleration scaling factor.

The specific conditions for producing density values for each model are detailed in Section 4.3. The authors acknowledge that while there are a number of ways to improve a density model's outputs at runtime (see Sutton (2018); Shim et al. (2014)), the outputs used in this work are intended to reflect typical community use, with each model being run according to the developer's operational instructions. Results are presented by focusing on a two week time period from 9 November 2018 - 23 November 2018, providing 14 adjacent daily arcs with no maneuver-based data gaps. The assessment conditions are for the altitude regime near ~ 490 km, in an atmosphere with very low solar flux, and low-to-minor geomagnetic activity. Note that this is a notoriously difficult altitude regime and activity condition for empirical models due to the minimal access to satellite density data at this altitude during times of prolonged solar minimum—increasing the potential value of this style of assessment for these models especially. Figure 3 shows low solar activity for the time period, both in terms of the magnitude and variation of solar EUV and FUV, as approximated by measurements of the 10.7 cm solar radio flux ($F_{10.7}$, top panel). The Kp geomagnetic index (bottom panel) depicts low-to-minor geomagnetic activity during the time of interest, with the two-week period being bookended by minor geomagnetic disturbances which reach no higher than $Kp = 4.3$. Only one minor-to-moderate disturbance occurs on 5 November 2018, four days before the period of interest, reaching a peak of $Kp = 5.7$. This event is mentioned only because of the possibility that its impact could be seen represented in the models as a density enhancement due to delayed heating and cooling effects.

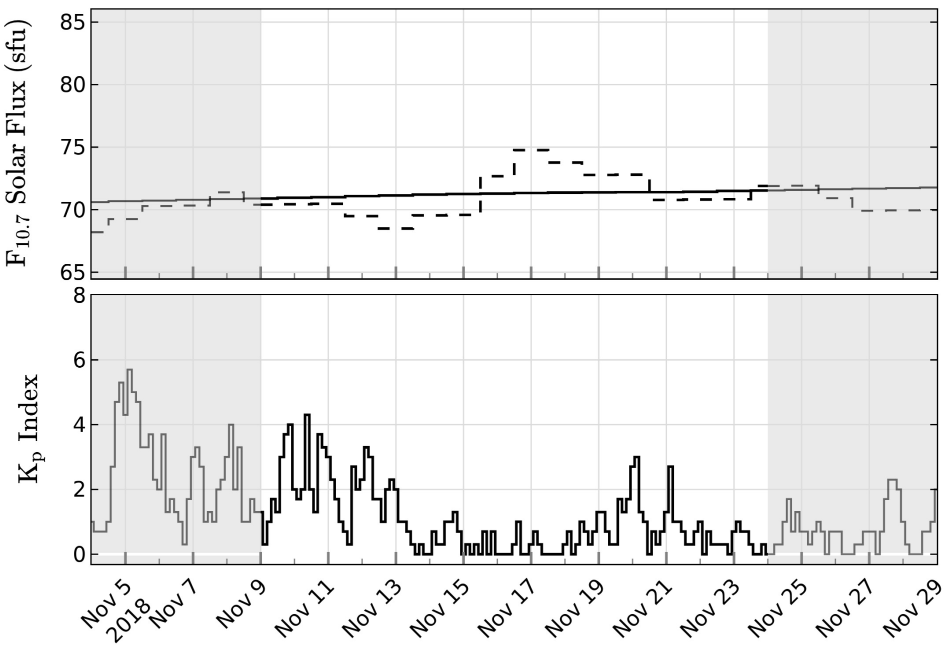


Figure 3. Top: observed solar $F_{10.7}$ radio flux. The dashed curve is the daily measured value from the Ottawa observatory normalized to 1 AU sun-earth distance; the solid curve is an 81-day (3 solar rotation) centered average. Bottom: the 3-hourly planetary magnetic index, Kp. Both panels depict the period of interest from 9 November 2018 - 23 November 2018. A few days before and after the period of interest are depicted in the shaded portions.

6.1 Preliminary Orbit Fits using a Fixed C_D of 2.5

The most straightforward way to construct the orbit fits is by calculating the acceleration due to drag from different density models using a fixed drag coefficient value for all arcs and models. This permits bias depending on the relationship between each model's mean density and the chosen C_D, but importantly demonstrates that the method is sufficiently sensitive for recognizing differences in the drag effects between the models. The in-track residual errors in this method should not be interpreted as indication of performance, but rather indicators of each model's mean density offset relative to the true-unknown density. Figure 4, shows the fixed C_D assessment results for the semi-empirical and physics-based models. The top panel shows each model's orbit averaged density along the orbit of ICESat-2, the middle panel shows the in-track residuals for each model and arc, and the bottom panel shows the in-track RMSe values.

The negative parabolic shape of the in-track residuals of MSIS2, TIEGCM, and JB2008 indicate that these modeled densities are too high—i.e., these orbit fits experience more drag acceleration and their fits tend to lag behind the PSO. The positive parabolic shape of the in-track residuals of DTM2020 and CTIPe indicate modeled densities that are too low—i.e., the drag acceleration is lower and the orbit fits tend to be in front of the PSO. In reality the PSO-to-orbit fit relationship is slightly more complex over an arc, with the above being a generalization of the overall trend. A more detailed understanding of the orbit fit movement relative to the PSO can be found in Appendix A.

The orbit fits from SET-HASDM are separated for use as verification since it uses similar assumptions of a fixed drag coefficient, and a satellite drag data assimilation technique in its internal workings. The SET-HASDM density database affords the opportunity to access historical records of HASDM densities that have been corrected through the real-time data-assimilative calibration to $\sim$80 low earth orbiters. The HASDM model is the operational standard used by the 18th Space Defense Squadron which is tasked with executing command and control over United States' space assets and all resident space objects for sake of space situational awareness. Verification with SET-HASDM provides a baseline understanding of the fidelity of the orbit fit results. Since the HASDM density values have already been effectively debiased in its data-assimilation scheme, we do not go through the steps of debiasing using the methods presented in Sections 6.2 and 6.3. Referring to Figure 5, the HASDM model consistently depicts in-track RMSe values that are on the order of 8.18 meters over the two-week period. The in-track residuals have a negative shape, indicating that the densities from SET-HASDM are slightly larger than what would be expected from the PSO. The results in Figure 5 are intended to serve as an approximate consistency check that our overall methodology, and more specifically our debiasing method, provide orbit fits with in-track errors in a reasonable range.

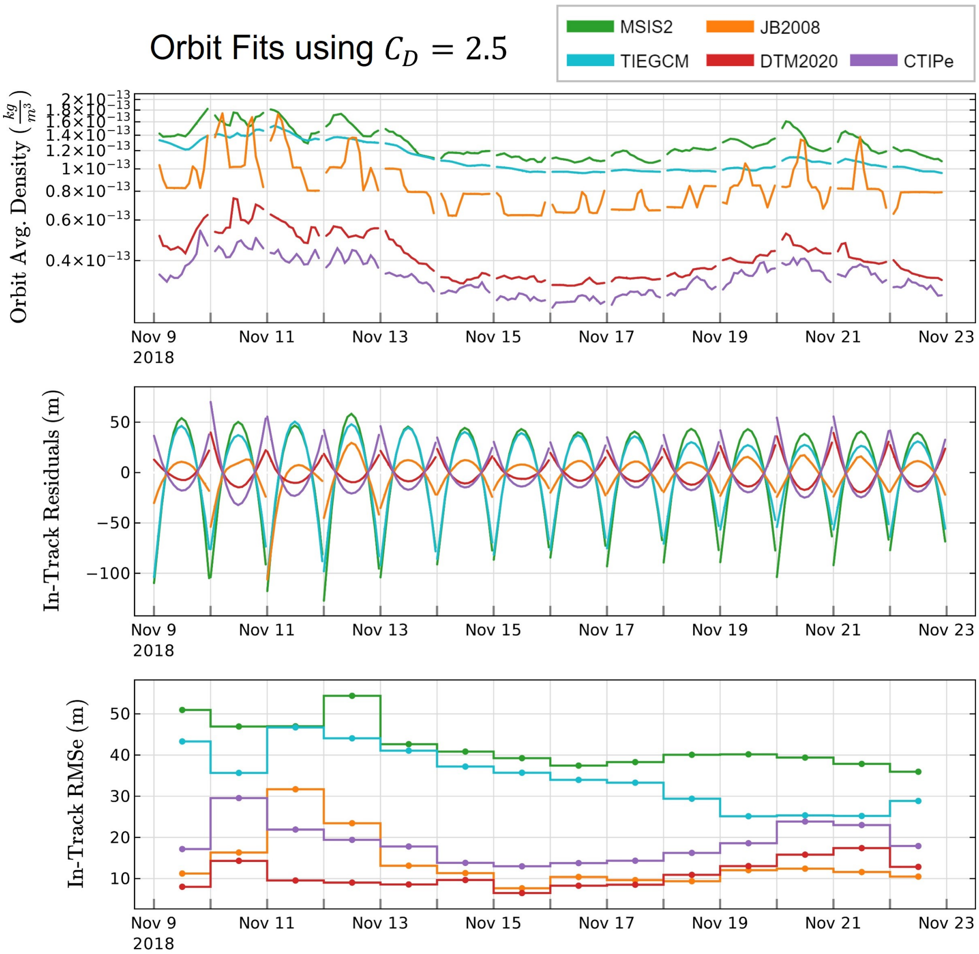


Figure 4. Assessment results given a fixed, common C_D=2.5 for MSIS2 (green), DTM2020 (red), JB2008 (orange), TIEGCM (cyan), and CTIPe (violet) during the two week time period containing 14, 24-hour arcs. Top: Orbit average neutral densities along the ICESat-2 orbit for each model. Middle: In-track orbit residuals for each arc. Bottom: In-track RMSe for each arc's in-track residuals.

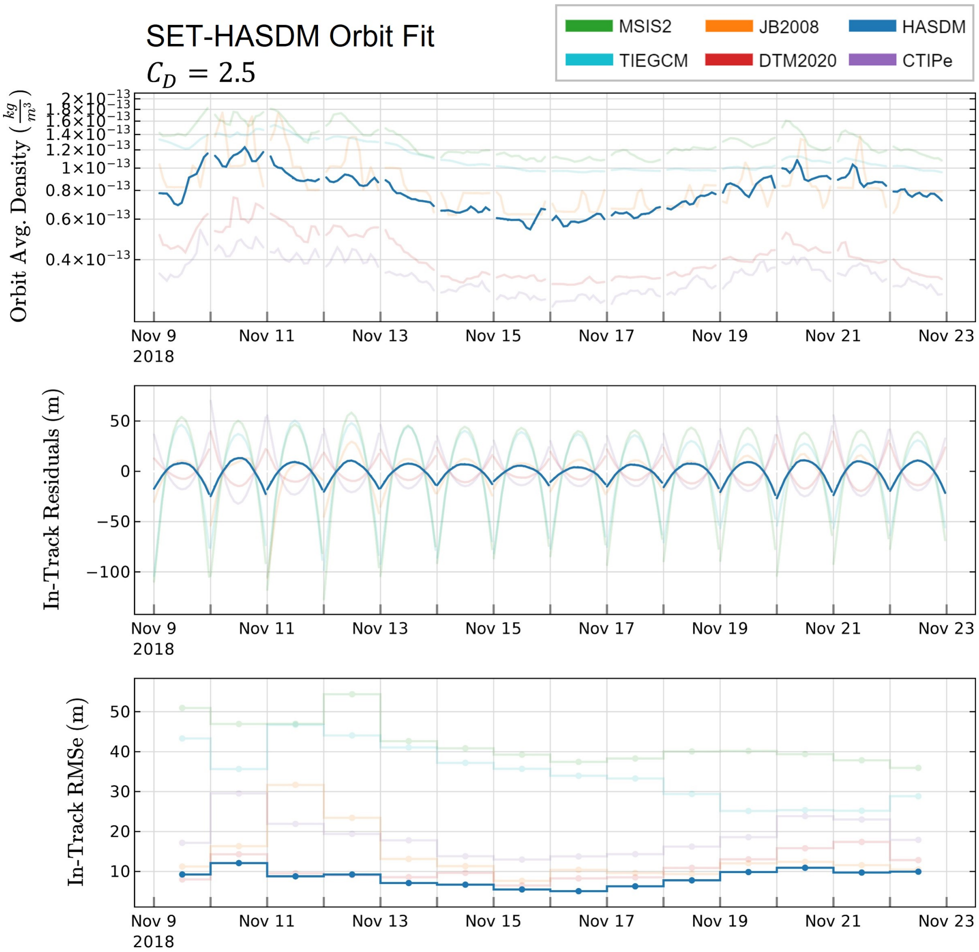


Figure 5. Verification results using SET-HASDM across 14 adjacent, 24-hour arcs from 9 November 2018 - 23 November 2018. Top: The solid blue curve depicts the neutral densities along the ICESat-2 orbit as an orbit average. Middle: In-track orbit residuals for each of the 14 adjacent, 24-hour arcs. Bottom: In-track RMSe values for each arc's in-track residuals. The range of the y-axes are chosen to facilitate comparison with Figure 4.

6.2 Debias using Two-Week Scaling Factor

The second orbit fit method debiases the density models using a mean-adjusted C_D over the two-week period (average values of the 24-hour adjusted C_Ds shown later in Figure 8). This provides a fixed C_D that is unique for each density model, adjusted to account for biases due to each model's mean density offset over the time period (see Figure 4). The in-track residuals and RMSe values for this assessment, shown in Figure 6, quantify the error due to density variation over the two-week time period. The average adjusted C_D used to construct the two-week debiased orbit fits for each model is reported in Table 2.

As mentioned in Section 5.2, the DRIA calculations indicate that $C_D = 2.5$ is a realistic value if one assumes that a fixed energy accommodation of $\alpha = 0.89$ is reasonable—an assumption limited by lack of empirical observation. According to the DRIA model, 2.5 is a realistic lower limit for the drag coefficient. Looking to Table 2, we can see that the mean adjusted C_D for MSIS2, TIEGCM, and JB2008 are all well below this lower limit, offering further evidence that these models are, on average, over-estimating the density. The upper bound is slightly more difficult to estimate in this setup, but CTIPe's adjusted C_D of 4.3 is likely too high. This will need to be investigated further in the future.

After removing the bias from the models, the relative effects of the minor geomagnetic activity become more stark in the in-track residuals. Here the changing shape of the residual curves indicate whether the model is over or underestimating the effects of geomagnetic activity on the modeled density. For example, several of the models display downward-pointing curves during geomagnetic activity, indicating densities that are too low as compared to their quiet time densities. DTM2020's residuals show an anti-correlation to the geomagnetic activity, beginning with densities that are too high and ending the two weeks with densities that are too low, which may be reflective of an overly sensitive response to geomagnetic activity and the overall downward trend in Kp. The effects of a model poorly capturing density variations during geomagnetically active times are now better quantified by the in-track RMSe after two-week debiasing is applied.

Two-week Debiasing Method: C_D = Mean Adjusted C_D

Model ID	Fixed C_D	Scaling Factor as % change
MSIS2	1.237	-49.861
TIEGCM	1.373	-44.899
JB2008	1.909	-22.967
DTM2020	3.351	34.903
CTIPe	4.368	75.663

Table 2. Summary of the assessment procedure assuming a fixed C_D that is equal to the average adjusted value for each model, assuming an a priori of $C_D = 2.5$. The second column reports mean-adjusted C_D used in each model's orbit fit construction. The third column reports the two-week scaling factor as a percent change for each model.

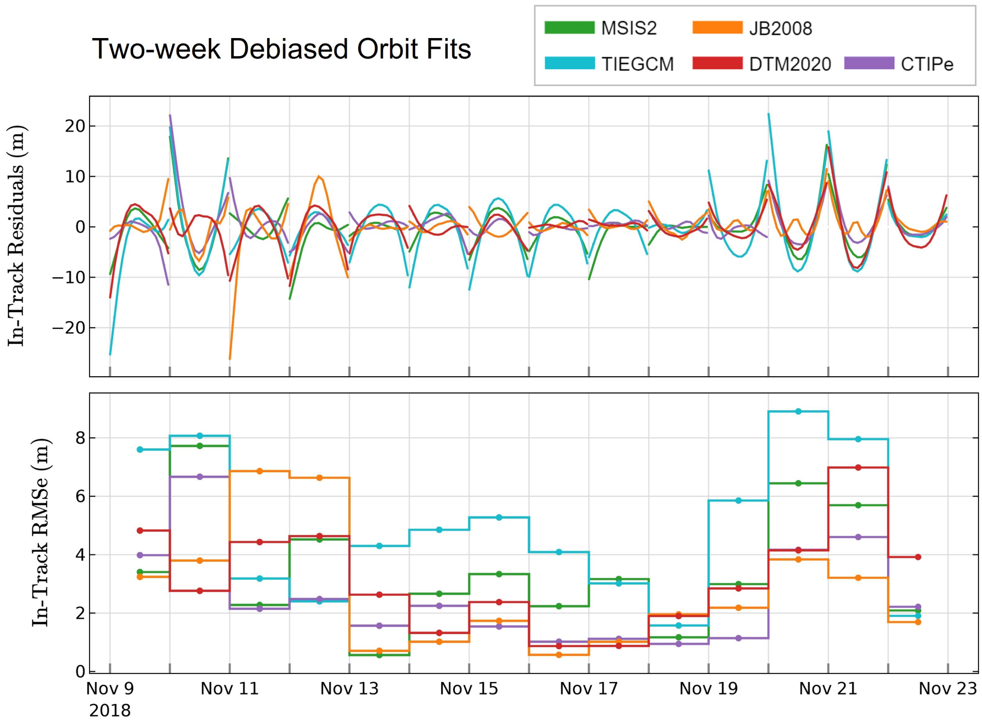


Figure 6. Assessment results for orbit fits that are debiased using two-week drag acceleration scaling factors. Top: Debiased in-track orbit residuals for each arc. Bottom: Debiased in-track RMSe for each arc's in-track residuals.

6.3 Debias using 24-hour Scaling Factor

The 24-hour debiasing procedure described in Section 5.2, is used to scale the orbit fits and their residuals to a daily cadence. Figure 7 presents the resulting in-track residuals (top panel) and in-track RMSe values (bottom panel) for each 24-hour arc for each model. The 24-hour drag acceleration scaling factor is derived by adjusting the C_D from the a priori of 2.5 over each daily arc, absorbing the average density offset for that day. The debiasing effect is seen in the overall reduction in residual error from Figure 4 to Figure 7. The calculated 24-hour scaling factors are presented in the top panel of Figure 8 as a percent change from the fixed value of 2.5. The bottom and right panels show the Kp index and Pearson's correlation coefficient between each model's scaling factors and the Kp, respectively.

The 24-hour scaling accounts for both the overall model bias and uncertainties in the density on timescales that are on the order of, or greater than, the chosen arc length of 24-hours (i.e. combination of mean density offset and daily geomagnetic variation). The remaining error depicted by the in-track residuals of Figure 7 are likely due to higher frequency variations in density that are not captured by the 24-hour debiasing (e.g. day-night variations in the neutral density). The correlation between Kp index and the scaling factors demonstrates how this metric can be used to determine how well a model accounts for geomagnetic activity. As shown by the scaling factors in Figure 8, MSIS2 ($R = 0.35$), TIEGCM ($R = 0.29$), and CTIPe ($R = 0.19$) all exhibit a subtle positive correlation, indicating a slight underestimation of density enhancements from geomagnetic activity and resulting in the scaling factors being used to compensate for these errors. Contrarily, the inverse relationships shown by DTM2020 ($R = -0.36$) and JB2008 ($R =$

−0.3) indicate an overestimation of these models' densities during days of increased activity.

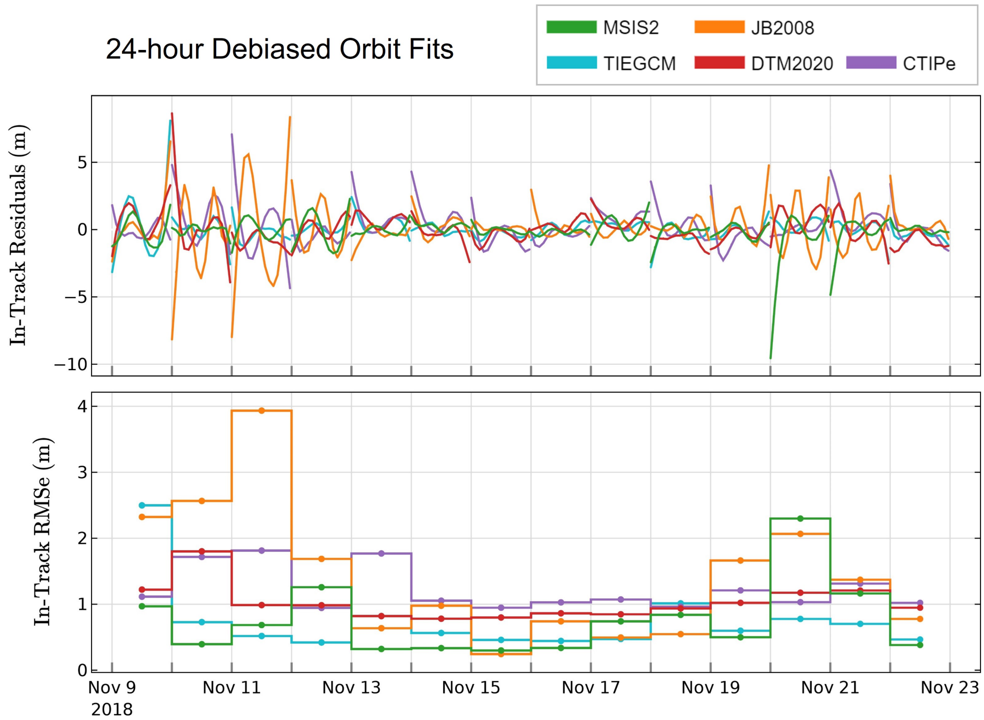


Figure 7. Assessment results for orbit fits using 24-hour drag acceleration scaling factors. Scaling factors are extracted in the least squares orbit fitting procedure by allowing the C_D to adjust once-per-arc to absorb observed errors between the PSO and the converging orbit fit. Top: In-track orbit residuals for each arc. Bottom: In-track RMSe for each arc's in-track residuals.

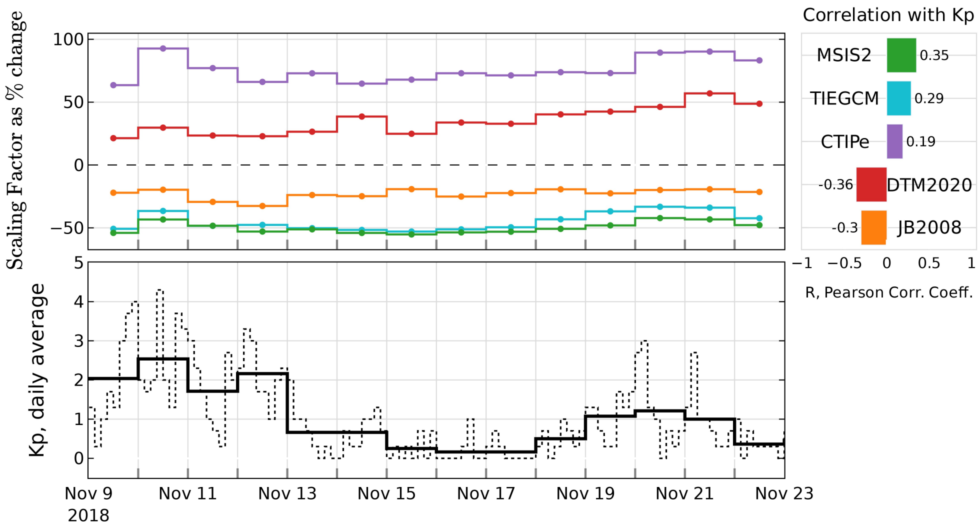


Figure 8. Top: Drag acceleration scaling factors extracted from the orbit fits shown in Figure 7, presented as a percent change from the fixed $C_D = 2.5$. Bottom left: The Kp index for the time period. Bottom right: Pearson's correlation coefficient between the scaling factors and Kp.

6.4 Discussion

While the in-track residuals of the ICESat-2 orbit fits offer an effective means of assessing the density models, most methods that use drag acceleration to study density are going to be limited by the complex interconnected uncertainties in the C_D and the density. For these reasons, we split our overall assessment into the three methods presented, each of which help to further illuminate the performance of the models. For sake of brevity, only JB2008 and TIEGCM are given in-depth discussions that synthesize an understanding of model performance from their results. The reader will then be able to apply these discussions to the remaining models, which here are discussed in more general terms.

Considering the cumulative results for JB2008 (plotted in orange for all relevant figures), Figure 4 shows that with a fixed C_D of 2.5, the in-track residuals exhibit a negative parabola shape, indicating that the JB2008 density provides larger drag accelerations than what the PSO experiences. Figure 6 provides the assessment method in which the JB2008 densities are effectively scaled by $\sim\!-22\%$ for the full two-week period. In this case the in-track residuals better highlight the arcs in which JB2008 performs poorly relative to the other models, specifically on November 11th and 12th when the geomagnetic activity fluctuates around $Kp = 3$ after having been moderately elevated for several days. In the 24-hour debiased case, JB2008's daily scaling factors (shown in orange in Figure 8) effectively compensate for density variations on the order of or greater than 24-hours. The mean density offset adjustment is $\sim\!-22\%$, and is seen to be inversely correlated ($R = -0.3$) with the geomagnetic activity—indicating that JB2008 tends to overestimate densities during active times. This effect can be clearly seen in JB2008's orbit averaged densities, represented by the orange line in the top panel of Figure 4, where the model provides much sharper density peaks than the other models during the times of slightly elevated geomagnetic activity. This effect is also clearly represented in the 24-hour debiased in-track RMSe values (bottom panel of Figure 7), which shows significant variance during active times between the PSO and the JB2008 orbit fit, even after the 24-hour scaling. The higher RMSe values shown in Figure 7 on November 9th-13th indicate that JB2008's overestimation of density during active times is not sufficiently compensated by the 24-hour scaling factor, meaning that the variation likely occurs at a higher frequency. JB2008's estimation of quiet time density is among the best in this report, needing $\leq\!-22\%$ adjustment from the fixed C_D case to provide orbit fits that match the PSO to within 1 meter (during quiet times).

Considering the cumulative results for TIEGCM (plotted in cyan for all relevant figures), Figure 4 shows that with a fixed C_D of 2.5, the in-track residuals exhibit a negative parabola shape, indicating that the TIEGCM density provides larger drag accelerations than what the PSO experiences (i.e. the TIEGCM densities are too high). TIEGCM offers interesting results from Figure 6 of the two-week scaled case. The TIEGCM densities are effectively scaled down by $\sim\!-45\%$ over the two-week period to compensate for mean density offset. Since this value is the two week average from the scaling factors shown in Figure 8, it is skewed to only partially compensate for lack of geomagnetic sensitivity (i.e., densities are scaled to be too low in active times, increase in RMSe on the bookends of the period) and partially compensate for mean density offset in quiet times (densities are scaled to be too high in quiet times, increase in RMSe from Nov. 13th to 17th). The two-week scaled case is able to clearly show in the in-track residuals that TIEGCM struggles more than the other models to properly capture variation during the period of this study. TIEGCM's daily scaling factors (cyan line in Figure 8) effectively compensate for error in density variations on the order of or greater than 24-hours. The daily variation of the scaling factors is found to slightly correlate with the Kp ($R = 0.29$)—indicating that TIEGCM tends to subtly underestimate densities during active times. This is seen in TIEGCM's orbit averaged densities (cyan line in the top panel of Figure 4) where the model provides significantly less sensitivity to the times of slightly elevated geomagnetic activity, and less variation overall. The 24-hour debiased in-track RMSe values (bottom panel of Figure 7), interestingly show very low variance between the PSO and the TIEGCM orbit fit after the 24-hour scaling. This is most likely explained by the overestimation of density during active times being sufficiently compensated for by the 24-hour scaling factors despite its orbit average densities seeming to lack much variation at all. This adds further suspicion to the higher frequency variations seen in models such as JB2008. It is possible that by moving to shorter arc lengths, such as 3-hours, the time series of scaling factors could better capture these variations, and this is a future goal of this work.

In general, MSIS2 and TIEGCM overestimate the density for all arcs, requiring $\sim -50\%$ and $\sim -45\%$ scaling factors, respectively, to bring the in-track residuals to within two meters. DTM2020 and CTIPe both underestimate the density for all arcs, each requiring a $\sim$35% and $\sim$76% increase, respectively. JB2008 requires the least overall scaling, requiring only $\sim -23\%$ to bring the in-track residuals to within two meters during quiet times. All models capture the geomagnetic activity relatively well as demonstrated by their scaling factors not being very highly correlated to Kp. DTM2020 ($R = -0.36$) and JB2008's ($R = -0.3$) scaling factors are inversely correlated to Kp, indicating a slight over-sensitivity to geomagnetic activity during this time period, while TIEGCM ($R = 0.29$), MSIS2 ($R = 0.35$), CTIPe ($R = 0.19$) all indicate an under-sensitivity. The scaling undergone for each model produces RMSe values that are comparable to that of the SET-HASDM orbit fit, which was separated out to serve as an approximate consistency check of our debiasing method due to its data-assimilative technique.

7 Conclusions and Future Work

This work presents the development of a modernized interface for the GEODYN-II POD software. The approach leverages the high-precision nature of space geodetic POD and an upgraded utility of the neutral density models to focus POD methods toward studying satellite drag and conducting density model assessment. The assessment method uses high-fidelity PSO as observed tracking measurements that are input into POD-based orbit fits. The drag effects from each density model are assessed according to each model's ability to redetermine the satellite's orbit. Each density model's orbit fit contains relative in-track deviations from the PSO which are treated as a proxy for model densities that differ from a true, unknown, density. These deviations are quantified with the in-track residuals and their RMS errors. We demonstrate the capabilities of this tool via a case study assessment of five thermospheric density models (MSIS2, DTM2020, JB2008, TIEGCM, and CTIPe, and a verification using SET-HASDM) using the ICESat-2 mission PSO as the observed measurements. Preliminary orbit fits are constructed after determining a mean C_D from a physics-based solution. A fixed C_D of 2.5 is applied for all models before being debiased by adjusting the C_D to account for density errors in the drag acceleration. The debiasing is performed at two different cadences, 24-hours and two-weeks, with each method highlighting different temporal aspects of the model density errors. The scaling factors extracted from the 24-hour and two-week debiasing methods are well-equipped for use in improving forecasting and modeling methods. The 24-hour scaling factors provide a more accurate representation of the true density variations for each model, while the two-week scaling factors are computationally simpler and indicate more baseline density effects. In addition, the two-week extended time period scaling factors are compatible for parsing into the CCMC's CAMEL database to move in the direction of community-oriented model validation.

We continue our efforts on this project as we move in the direction of offering a more robust thermospheric model validation scheme. Possible improvements include improving the non-conservative force modeling in GEODYN for ICESat-2 using a more realistic 3-D model of the satellite shape that would account for self-shadowing and variations in cross-sectional area with incidence angle e.g, as in March et al. (2019). The orbit determination for the primary science orbits, and the subsequent analyses described in this paper would have to use these improved geometry models. A further improvement would be to incorporate SLR measurements of ICESat-2 into the evaluations of the density models. One could include the SLR data along with the PSO trajectory data in the evaluation. See Thomas et al. (2021) for a description of these data for ICESat-2. Planned future work involves addressing the key constraints highlighted in the methodology, the foremost of which is the need to evaluate the drag coefficient more frequently along the ICESat-2 orbit. Future work will also involve expanding the study to encompass the entire ICESat-2 mission time period. Additional expansion includes incorporating additional satellites and constellations that may illuminate model performance within atmospheric regimes that lack observations of neutral density. We aim to make our expanded results available through the CCMC's CAMEL framework as well as through future publications.

The assumptions made in this paper are limited by the current status of unknowns between gas surface interaction research and thermospheric variability research. At this time, the true drag coefficient is not known for any satellite, and modeling the C_D will always introduce some inherent bias into the results. We aim to address this issue by implementing several of the satellite gas-surface interaction models currently used in the satellite drag community to calculate the time and compositionally-dependent drag coefficient. Isolating the effects of the C_D will aid to better identify the various non-density related errors that may be present in the in-track residuals. Being able to distinguish these errors and accurately quantify the amount of deviation introduced by a given density model will provide significant insight regarding model performance to the earth-space environment modeling community. As the ability to model C_D improves, the results provided by this method will similarly become more valid. The Geospace Dynamics Constellation (GDC) is an upcoming NASA mission that is intended to help fill in the gaps of understanding gas-surface interactions by providing a stable platform with full measurements of neutral composition, density, and temperature along with a high fidelity POD in which cross validation of density model assessment is possible. As a result, in addition to providing its own neutral density observations that can be used for research, operations, and model validation, these advances expected from the GDC mission will improve the accuracy and usability of density proxies derived from POD solutions like those used here. These advances will effectively multiply our density observations to be able to use any satellite with sufficiently accurate GNSS positioning and knowledge of spacecraft parameters as a density observing platform.

This work provides a step in the direction of being able to use high-fidelity GNSS-enabled LEO satellite POD solutions to objectively quantify and validate thermospheric model performance. The strength of assessment using this method is its ability to identify relative accuracy of the models in a way that is directly tied to operational use for orbit propagation. There are a multitude of uses for the tools and methods presented in this work, such as for density retrievals along the orbit of a satellite, which is a planned future effort; however, this report focuses specifically on model assessment. As work continues to refine these methods and address the caveats presented in this paper, the results of model assessments using this technique will continue to become better suited to aid satellite operators when choosing a model that will perform best under specified conditions. Having a multitude of methods for assessing upper atmospheric models under various conditions helps model developers refine the models themselves, making them better suited for orbit prediction.

8 Open Research

The ICESat-2 POD solutions, their corresponding setup files, and the GEODYN II software are provided by the Geodesy and Geophysics Laboratory within NASA-GSFC. Simulation results for the CTIPe model have been provided by the Community Coordinated Modeling Center at Goddard Space Flight Center through their publicly available simulation services (https://ccmc.gsfc.nasa.gov). Orbit fly-throughs of the TIEGCM simulation results and relevant codes used to produce the results in this paper are available at Zenodo via https://doi.org/10.5281/zenodo.8015368 (Waldron et al., 2023). The SET HASDM density data are provided for scientific use by Space Environment Technologies.

Acknowledgments

The authors acknowledge the generous support of the Community Coordinated Modeling Center under grant number 80NSSC21K1747 and support from the University of Colorado at Boulder Grand Challenge Grant: Space Weather Technology, Research, and Education Center (SWx TREC). This material is partially based upon work supported by the National Aeronautics and Space Administration under Grant 80NSSC21K1554 issued through the Heliophysics Division Space Weather Science Application initiative and the Geospace Dynamics Constellation Contract No. 80GSFC22CA012. The authors acknowledge the CCMC-Kamodo Team who designed the Kamodo API, permitting the ability to interface physics-based model outputs to the POD scheme.

Appendix A Coordinate System to Study Drag

GEODYN's input and output trajectories make use of the J2000 inertial reference system. The $\hat{X}$, $\hat{Y}$, and $\hat{Z}$ components of the inertial coordinate system offer limited information on how a satellite's orbit is impacted by atmospheric drag, leading us to convert to the more suitable Satellite Coordinate System. Two coordinate frames suited for this assessment are the NTW and RSW frames, with differences between the two being highlighted in Figure A1. We make use of the NTW system, which aligns with the orbit plane and is composed of an in-track component $\hat{T}$ that is parallel to the velocity vector $\vec{v}$, a normal component $\hat{N}$ that is perpendicular to the velocity and nominally in the radial direction, and a cross-track component $\hat{W}$ that is normal to the orbit plane and completes the right-hand coordinate frame. Being parallel to the velocity vector means that the in-track component $\hat{T}$ will contain any indication that the spacecraft's trajectory has changed since orbital energy dissipations from drag will impact in the velocity direction.

Figure A2 contains additional visualization related to the shape of the in-track residuals and how it relates to the movement of the PSO relative to the orbit-fit satellite. The overall shape of the in-track residuals is a result of the batch-least squares fitting routine as it attempts to minimize the distance between the PSO and the orbit fit across the whole arc.

Variations in the in-track component $\hat{T}$ are not the same as variations in the along-track component $\hat{S}$ of the RSW system. In-track variations act in the direction of the velocity vector, whereas along-track variations are merely along, but not necessarily parallel, to the direction of the velocity vector. We make the distinction to use the NTW system rather than the RSW system whose radial component is often used to assess orbit accuracy in geodetic POD studies. The NTW coordinate system is described in (Vallado, 2013) to have the following unit vectors and transformation:

$$\hat{T} = \frac{\mathbf{v}}{|\,\mathbf{v}\,|} \tag{A1}$$

$$\hat{W} = \frac{\mathbf{r}\times\mathbf{v}}{|\,\mathbf{r}\times\mathbf{v}\,|} \tag{A2}$$

$$\hat{N} = \hat{T}\times\hat{W} \tag{A3}$$

$$\mathbf{r}_{XYZ} = \left[\hat{N}\vdots\hat{T}\vdots\hat{W}\right]\mathbf{r}_{NTW} \tag{A4}$$

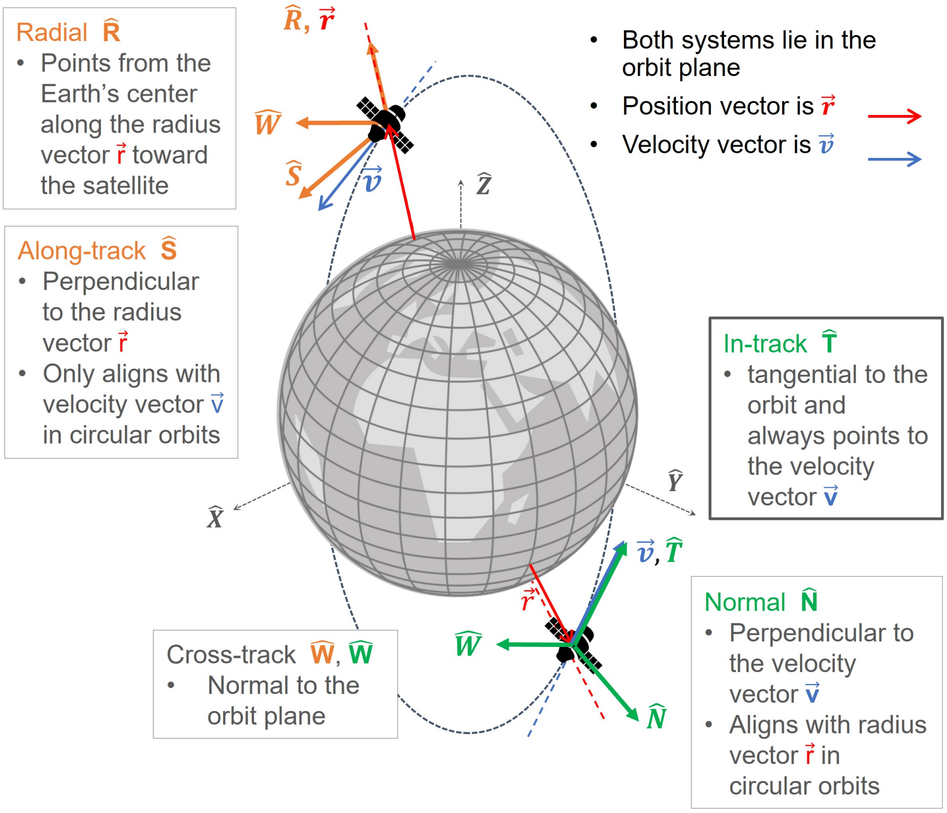


Figure A1. The above is a schematic showing the NTW and RSW satellite coordinate systems and details regarding their components. The NTW system's in-track component is parallel to the velocity vector, making it an effective tool for assessing relative effects due to atmospheric drag.

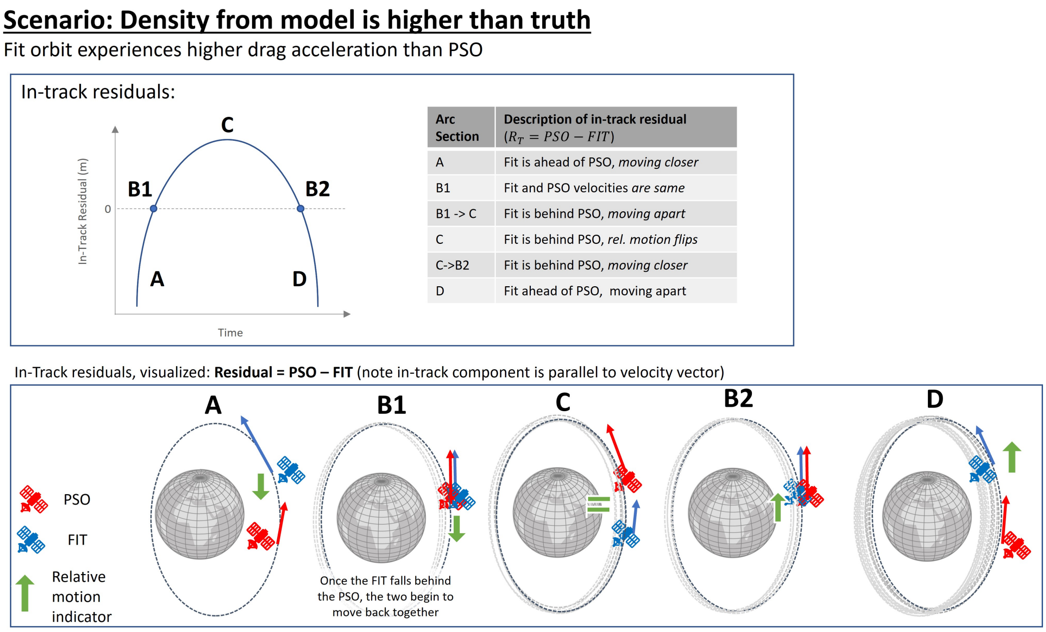


Figure A2. Above is a diagram which explores the circumstances that result in the parabolic shape of the in-track residuals. Representative in-track residuals for a single arc is given as a negative quadratic curve in the top panel. The bottom panel provides corresponding frames of schematics for each marked point to depict how the orbit fit and PSO are positioned relative to each other.

Appendix B GEODYN Run Setup for Orbit Fits

This section provides a summarized overview, in the form of Table B1, of the run setup used in GEODYN to produce the orbit fits. A similar overview is given in Table 1 of Thomas et al. (2021), which details the most relevant and important constants, models, and standards used to produce the ICESat-2 PSO.

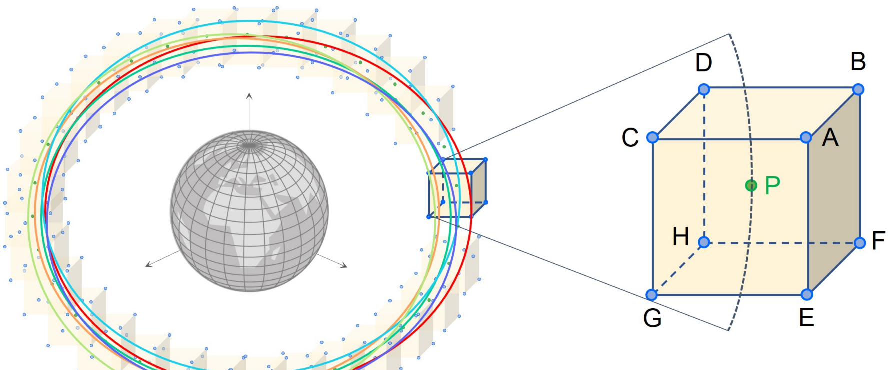
GEODYN's Setup to Produce POD-based Orbit Fits

Satellite geometry	Panel model based on pre-launch geometry and satellite surface optical properties. 14 panels make up a Box Wing model to calculate time varying area.
ICESat-2 Attitude	Telemetered spacecraft body-fixed reference frame to inertial reference frame quaternions. Telemetered solar array drive angles (for force modeling)
Non-Conservative Forces	
Atmospheric Density Models	Modified for comparison; see Table 1 for the list of assessed density models
Earth Radiation	Knocke 2nd degree zonal spherical harmonic of Earth's albedo and emissivity (Knocke et al., 1988)
Solar Radiation Pressure	Solar radiation incident on plate model (Luthcke et al., 2019; Marshall & Luthcke, 1994)
Conservative Forces	
Geopotential gravity	EIGEN6C, tide-free (Foerste et al., 2014)
Time variable gravity	Contribution from atmosphere, non-tidal oceans, hydrology, and ice; Developed from GRACE models
Earth, Pole, and Ocean tides	IERS2010 Conventions (Petit & Luzum, 2010)
Planetary ephemerides (N-Body)	JPL DE430 (Folkner et al., 2014)
Relativistic corrections	IERS2010
General Reference Frame and Constants	
Conventional inertial system	J2000 geocentric; mean equator and equinox of 2000 JAN 01 12:00:00; IERS2010
Precession - Nutation	IAU 2000A precession-nutation model
Earth Orientation Parameters	IERS 08 C04 (Bizouard & Gambis, 2011), IERS2010 conventions for diurnal, semidiurnal, and long period tidal effects on polar motion and UT1
Numerical integration	Cowell predictor-corrector; fixed and variable step; equations of motion and variational equations.
Estimation method	Partitioned Bayesian least squares.
GEODYN Controlled Setup Information	
Tracking data type	PCE (orbit trajectory) using ICESat-2 PSO
POD technique	Dynamic data reduction (no empirical accelerations)
Arc Length	24 hours
Adjusted Parameters	Initial conditions only: $X, Y, Z, \dot{X}, \dot{Y}, \dot{Z}$
Force model parameters	C_D = fixed; $C_R = 1$ (not adjusted)
Integration/orbit step	10 seconds

Table B1. A summarized overview of the GEODYN run setup for using the program to conduct density model assessment. Many of the above parameters are summarized from (Thomas et al., 2021) and (Luthcke et al., 2019).

Appendix C Orbit Uncertainty Interpolation Technique and the Kamodo Interface

Kamodo is a CCMC tool for access, interpolation, and visualization of space weather models and data in Python (Ringuette et al., 2023). Kamodo allows model developers to represent simulation results as mathematical functions which may be manipulated directly by end users. Kamodo handles unit conversion transparently and supports interactive science discovery through Jupyter notebooks with minimal coding in Python. Kamodo is chosen for this project due to its ability to offer model agnostic methods for reading data output from different model sources. Kamodo is called using its Satellite Flythrough capabilities, in which a user is able to sample the models with satellite ephemeris and return requested values from the chosen model. The orbit is pre-initialized in GEODYN using MSIS2 to get an a priori estimate for the orbit coordinates. Then using the a priori orbit, extend out the uncertainty of the coordinates to create a cube of possible values centered on the orbit. This approach accounts for possible model output differences as the orbit iteratively converges towards a solution. Finally, we plug the orbit and its uncertainty cubes into Kamodo to interpolate the model densities at all requested points. By doing this, the orbit density values from the physics model can be quickly ingested into the POD program. Figure C1 visualizes this procedure.



Point	Longitude	Latitude	Altitude
P	Lon	Lat	Z
A	$Lon + \delta_{deg}$	$Lat + \delta_{deg}$	$Z + \delta_m$
B	$Lon + \delta_{deg}$	$Lat - \delta_{deg}$	$Z + \delta_m$
C	$Lon - \delta_{deg}$	$Lat + \delta_{deg}$	$Z + \delta_m$
D	$Lon - \delta_{deg}$	$Lat - \delta_{deg}$	$Z + \delta_m$
E	$Lon + \delta_{deg}$	$Lat + \delta_{deg}$	$Z - \delta_m$
F	$Lon + \delta_{deg}$	$Lat - \delta_{deg}$	$Z - \delta_m$
G	$Lon - \delta_{deg}$	$Lat + \delta_{deg}$	$Z - \delta_m$
H	$Lon - \delta_{deg}$	$Lat - \delta_{deg}$	$Z - \delta_m$

Figure C1. A representative schematic showing the constructed "cube of uncertainty" that surrounds a given coordinate along the orbit of a satellite. Each point that makes up this cube will contain modeled neutral density values between which we can interpolate in GEODYN as the orbit drifts from the a priori orbit. This figure also demonstrates how perturbations due to different density models, represented here by different colored orbits, may necessitate a range of uncertainty for the satellite's indexed location.

References

Berger, T. E., Dominique, M., Lucas, G., Pilinski, M., Ray, V., Sewell, R., … Thiemann, E. (2023). The thermosphere is a drag: The 2022 starlink incident and the threat of geomagnetic storms to low earth orbit space operations. *Space Weather*, *21*(3). doi: https://doi.org/10.1029/2022SW003330

Berger, T. E., Holzinger, M. J., Sutton, E. K., & Thayer, J. P. (2020). Flying Through Uncertainty. *Space Weather*, *18*(1). doi: https://doi.org/10.1029/2019SW002373

Bernstein, V., & Pilinski, M. (2022). Drag coefficient constraints for space weather observations in the upper thermosphere. *Space Weather*, *20*, e2021SW002977. doi: https://doi.org/10.1029/2021SW002977

Bizouard, C., & Gambis, D. (2011). *The combined solution C04 for Earth Orientation Parameters consistent with International Terrestrial Reference Frame 2008* (Tech. Rep.). Observatoire de Paris, Paris, France. Retrieved from `https://hpiers.obspm.fr/iers/eop/eopc04_08/C04.guide.pdf`

Bowman, B. R., Tobiska, W. K., Marcos, F. A., Huang, C. Y., Lin, C. S., & Burke, W. J. (2008). A New Empirical Thermospheric Density Model JB2008 Using New Solar and Geomagnetic Indices. *AIAA/AAS Astrodynamics Specialist Conference and Exhibit.* doi: https://doi.org/10.2514/6.2008-6438

Bruinsma, S., Arnold, D., Jäggi, A., & Sánchez-Ortiz, N. (2017). Semi-empirical thermosphere model evaluation at low altitude with GOCE densities. *Journal of Space Weather and Space Climate*, *7*(A4). doi: https://doi.org/10.1051/swsc/2017003

Bruinsma, S., & Boniface, C. (2021). The operational and research DTM-2020 thermosphere models. *Journal of Space Weather and Space Climate*, *11*(3). doi: https://doi.org/10.1051/swsc/2021032

Bruinsma, S., Boniface, C., Sutton, E. K., & Fedrizzi, M. (2021). Thermosphere modeling capabilities assessment: Geomagnetic storms. *Journal of Space Weather and Space Climate*, *11*(2018). doi: https://doi.org/10.1051/swsc/2021002

Bruinsma, S., Sutton, E., Solomon, S. C., Fuller-Rowell, T., & Fedrizzi, M. (2018). Space Weather Modeling Capabilities Assessment: Neutral Density for Orbit Determination at low Earth orbit. *Space Weather*, *16*(11), 1806–1816. doi: https://doi.org/10.1029/2018SW002027

Bruinsma, S., Tamagnan, D., & Biancale, R. (2004). Atmospheric densities derived from CHAMP/STAR accelerometer observations. *Planetary and Space Science*, *52*, 297-312. doi: https://doi.org/10.1016/j.pss.2003.11.004

Bussy-Virat, C. D., Ridley, A. J., & Getchius, J. W. (2018). Effects of Uncertainties in the Atmospheric Density on the Probability of Collision Between Space Objects. *Space Weather*, *16*(5), 519–537. doi: https://doi.org/10.1029/2017SW001705

Doornbos, E., van den IJssel, J., Luhr, H., Forster, M., & Koppenwallner, G. (2010). Neutral Density and Crosswind Determination from Arbitrarily Oriented Multiaxis Accelerometers on Satellites. *Journal of Spacecraft and Rockets*, *47*(4), 580-589. doi: https://doi.org/10.2514/1.48114

Emmert, J. T. (2015). Thermospheric mass density: A review. *Advances in Space Research*, *56*(5), 773–824. doi: https://doi.org/10.1016/j.asr.2015.05.038

Emmert, J. T., Drob, D. P., Picone, J. M., Siskind, D. E., Jones, M., Mlynczak, M. G., … Yuan, T. (2021). NRLMSIS 2.0: A Whole-Atmosphere Empirical Model of Temperature and Neutral Species Densities. *Earth and Space Science*, *8*(3), e2020EA001321. doi: https://doi.org/10.1029/2020EA001321

Fang, T. W., Kubaryk, A., Goldstein, D., Li, Z., Fuller-Rowell, T., Millward, G., … Babcock, E. (2022). Space Weather Environment During the SpaceX Starlink Satellite Loss in February 2022. *Space Weather*, *20*. doi: https://doi.org/10.1029/2022SW003193

Foerste, C., Bruinsma, S., Abrikosov, S., & et al. (2014). EIGEN-6C4 The latest combined global gravity field model including GOCE data up to degree and order 2190 of GFZ Potsdam and GRGS Toulouse. *GFZ Data Services*. doi: https://doi.org/10.5880/icgem.2015.1

Folkner, W., Williams, J., Boggs, D., Park, R., & Kuchynka, P. (2014). *The Planetary and Lunar Ephemerides DE430 and DE431* (Tech. Rep.). IPN Progress Report, 42-196, Jet Propulsion Laboratory, Pasadena, California. Retrieved from `https://ipnpr.jpl.nasa.gov/progress_report/42-196/196C.pdf`

Hapgood, M., Liu, H., & Lugaz, N. (2022). SpaceX – sailing close to the space weather? *Space Weather*. doi: https://doi.org/10.1029/2022SW003074

He, C., Yang, Y., Carter, B., Kerr, E., Wu, S., Deleflie, F., … Norman, R. (2018). Review and comparison of empirical thermospheric mass density models. *Progress in Aerospace Sciences*, *103*, 31–51. doi: https://doi.org/10.1016/j.paerosci.2018.10.003

Heelis, R. A., Lowell, J. K., & Spiro, R. W. (1982). A model of the high-latitude ionospheric convection pattern. *Journal of Geophysical Research: Space Physics*, *87*(A8), 6339–6345. doi: https://doi.org/10.1029/JA087iA08p06339

Hejduk, M. D., & Snow, D. E. (2018). The Effect of Neutral Density Estimation Errors on Satellite Conjunction Serious Event Rates. *Space Weather*, *16*(7), 849–869. doi: https://doi.org/10.1029/2017SW001720

Knocke, P. C., Ries, J. C., & Tapley, B. D. (1988). Earth radiation pressure effects on satellites. *Astrodynamics Conference, 1988*, 577–587. Retrieved from `https://arc.aiaa.org/doi/abs/10.2514/6.1988-4292` doi: 10.2514/6.1988-4292

Lemoine, F. G., Chinn, D. S., Zelensky, N. P., Beall, J. W., & Bail, K. L. (2016). The development of the GSFC DORIS contribution to ITRF2014. *Advances in Space Research*, *58(12)*, 2520-2542. doi: https://doi.org/10.1016/j.asr.2015.12.043

Loomis, B. D., Luthcke, S. B., & Sabaka, T. J. (2019). Regularization and error characterization of GRACE mascons. *Journal of Geodesy*, *93*(9), 1381–1398. doi: https://doi.org/10.1007/s00190-019-01252-y

Luthcke, S. B., Pennington, T., Loomis, B. D., Rebold, T., & Thomas, T. (2019). *ICESat-2 ATBD for Precise Orbit Determination, Orbit Design, and Geolocation Parameter Calibration* (Tech. Rep. No. Release 002). NASA Goddard Space Flight Center. Retrieved from `https://icesat-2.gsfc.nasa.gov/sites/default/files/page_files/ICESat2_POD_ATBD_r002.pdf`

Luthcke, S. B., Rowlands, D. D., Lemoine, F. G., Klosko, S. M., Chinn, D., & McCarthy, J. J. (2006). Monthly spherical harmonic gravity field solutions determined from GRACE inter-satellite range-rate data alone. *Geophysical Research Letters*, *33*(2). doi: https://doi.org/10.1029/2005GL024846

Luthcke, S. B., Zelensky, N. P., Rowlands, D. D., Lemoine, F. G., & Williams, T. A. (2003). The 1-Centimeter Orbit: Jason-1 Precision Orbit Determination Using GPS, SLR, DORIS, and Altimeter Data. *Marine Geodesy*, *26(3-4)*, 399-421. doi: 10.1080/714044529

Lyon, R. H., Engineering, A., Cefola, P. J., & Greitzer, E. M. (2004). Geosynchronous orbit determination using space surveillance network observations and improved radiative force modeling. Retrieved from `https://dspace.mit.edu/handle/1721.1/17779`

March, G., Doornbos, E. N., & Visser, P. N. (2019). High-fidelity geometry models for improving the consistency of CHAMP, GRACE, GOCE and Swarm thermospheric density data sets. *Advances in Space Research*, *63(1)*, 213-238. doi: https://doi.org/10.1016/j.asr.2018.07.009

Marshall, J. A., & Luthcke, S. B. (1994). Modeling radiation forces acting on Topex/Poseidon for precision orbit determination. *Journal of Spacecraft and Rockets*, *31*(1), 99-105. doi: https://doi.org/10.2514/3.26408

Mehta, P. M., Paul, S. N., Crisp, N. H., Sheridan, P. L., Siemes, C., March, G., & Bruinsma, S. (2022). Satellite drag coefficient modeling for thermosphere science and mission operations. *Advances in Space Research.* doi: 10.1016/J.ASR.2022.05.064

Mehta, P. M., Walker, A. C., Sutton, E. K., & Godinez, H. C. (2017). New density estimates derived using accelerometers on board the CHAMP and GRACE satellites. *Space Weather*, *15*(4), 558–576. doi: 10.1002/2016SW001562

Millward, G., Moffett, R., Quegan, S., & Fuller-Rowell. (1996). A Coupled Thermosphere-Ionosphere-Plasmasphere Model (CTIP). *STEP Handbook on Ionospheric Models (ed. R.W. Schunk).*

Millward, G. H., Müller-Wodarg, I. C., Aylward, A. D., Fuller-Rowell, T. J., Richmond, A. D., & Moffett, R. J. (2001). An investigation into the influence of tidal forcing on F region equatorial vertical ion drift using a global ionosphere-thermosphere model with coupled electrodynamics. *Journal of Geophysical Research: Space Physics*, *106*(A11), 24733–24744. doi: 10.1029/2000ja000342

Muelhaupt, T. J., Sorge, M. E., Morin, J., & Wilson, R. S. (2019). Space traffic management in the new space era. *Journal of Space Safety Engineering*, *6*(2), 80–87. doi: 10.1016/J.JSSE.2019.05.007

Pavlis, D., Mccarthy, J., Rowton, S., Moore, D., Luthcke, S., Tsaoussi, L., & Beall, J. (2019). GEODYN Systems Description: Volumes 1, 3, and 5 [Computer software manual]. Retrieved from `https://earth.gsfc.nasa.gov/sites/default/files/neptune/files/geodyn_vol1.pdf`

Petit, G., & Luzum, B. (2010). *IERS Conventions (2010)* (Tech. Rep. No. 36). Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main, Germany.

Pilinski, M. D. (2008). Analysis of a novel approach for determining atmospheric density from satellite drag. *ProQuest Dissertations and Theses*, 150. Retrieved from `https://colorado.idm.oclc.org/login?url=https://www.proquest.com/dissertations-theses/analysis-novel-approach-determining-atmospheric/docview/230692345/se-2`

Pilinski, M. D., Argrow, B. M., & Palo, S. E. (2010). Semiempirical Model for Satellite Energy-Accommodation Coefficients. *Journal of Spacecraft and Rockets*, *47*(6), 951-956. doi: https://doi.org/10.2514/1.49330

Qian, L., Burns, A. G., Emery, B. A., Foster, B., Lu, G., Maute, A., … Wang, W. (2014). The NCAR TIE-GCM. In *Modeling the ionosphere–thermosphere system* (p. 73-83). American Geophysical Union (AGU). doi: https://doi.org/10.1002/9781118704417.ch7

Reigber, C., Schwintzer, P., Stubenvoll, R., Schmidt, R., Flechtner, F., Meyer, U., … Raimondo, J. C. (2006). *A High Resolution Global Gravity Field Model Combining CHAMP and GRACE Satellite Mission and Surface Data: EIGEN-CG01C* (Tech. Rep. No. Scientific Technical Report STR ; 06/07). Potsdam : Deutsches GeoForschungsZentrum GFZ, 12 S. p. doi: https://doi.org/10.2312/GFZ.b103-06075

Richmond, A. D., Ridley, E. C., & Roble, R. G. (1992). A thermosphere/ionosphere general circulation model with coupled electrodynamics. *Geophysical Research Letters*, *19*(6), 601–604. doi: https://doi.org/10.1029/92GL00401

Ringuette, R., Zeeuw, D. D., Rastaetter, L., & Pembroke, A. (2023). Kamodo's Model-Agnostic Satellite Flythrough: Lowering the Utilization Barrier for Heliophysics Model Outputs. *Accepted in Frontiers in Astronomy and Space Sciences.* doi: 10.1016/J.ASR.2023.03.033

Schunk, R., & Nagy, A. (2009). *Ionospheres: Physics, Plasma Physics, and Chemistry* (2nd ed.). Cambridge University Press. doi: https://doi.org/10.1017/CBO9780511635342

Shim, J. S., Kuznetsova, M., Rastätter, L., Bilitza, D., Butala, M., Codrescu, M., … Sutton, E. (2014). Systematic Evaluation of Ionosphere/Thermosphere (IT)

Models: CEDAR Electrodynamics Thermosphere Ionosphere (ETI) Challenge (2009-2010). *Modeling the Ionosphere-Thermosphere System*, *9780875904917*, 145–160. doi: 10.1002/9781118704417.ch13

Solomon, S. C., & Qian, L. (2005). Solar extreme-ultraviolet irradiance for general circulation models. *Journal of Geophysical Research: Space Physics*, *110*(A10). doi: 10.1029/2005JA011160

Storz, M. F., Bowman, B. R., Branson, J. I., Casali, S. J., & Tobiska, W. K. (2005). High accuracy satellite drag model (HASDM). *Advances in Space Research*, *36*(12), 2497–2505. doi: 10.1016/j.asr.2004.02.020

Sutton, E. K. (2018). A New Method of Physics-Based Data Assimilation for the Quiet and Disturbed Thermosphere. *Space Weather*, *16*, 736–753. doi: https://doi.org/10.1002/2017SW001785

Sutton, E. K., Forbes, J. M., & Nerem, R. S. (2005). Global thermospheric neutral density and wind response to the severe 2003 geomagnetic storms from CHAMP accelerometer data. *Journal of Geophysical Research: Space Physics*, *110*, 9-40. doi: 10.1029/2004JA010985

Sutton, E. K., Thayer, J. P., Wang, W., Solomon, S. C., Liu, X., & Foster, B. T. (2015). A self-consistent model of helium in the thermosphere. *Journal of Geophysical Research A: Space Physics*, *120*(8), 6884–6900. doi: 10.1002/2015JA021223

Tapley, B., Ries, J., Bettadpur, S., Chambers, D., Cheng, M., Condi, F., … Wang, F. (2005). GGM02 - An improved Earth Gravity Field Model from GRACE. *Journal of Geodesy*, *79*(8), 467–478. doi: 10.1007/s00190-005-0480-z

Tapley, B., Schutz, B., & Born, G. H. (2004). *Statistical Orbit Determination.* Elsevier Academic Press. doi: https://doi.org/10.1016/B978-0-12-683630-1.X5019-X

Thayer, J. P., Tobiska, W. K., Pilinski, M. D., & Sutton, E. K. (2021). Remaining Issues in Upper Atmosphere Satellite Drag. *Space Physics and Aeronomy Collection*, *5*, 111–140. doi: 10.1002/9781119815570.ch5

Thayer, J. P., Waldron, Z. C., & Sutton, E. K. (2023). Solar flux dependence of upper thermosphere diurnal variations: Observed and modeled. *Journal of Geophysical Research: Space Physics*, *128*. doi: 10.1029/2022JA031146

Thomas, T., Luthcke, S., Pennington, T., Nicholas, J., & Rowlands, D. (2021). ICESat-2 Precision Orbit Determination. *Earth and Space Science.* doi: 10.1029/2020ea001496

Tobiska, W. K., Bowman, B. R., Bouwer, S. D., Cruz, A., Wahl, K., Pilinski, M. D., … Licata, R. J. (2021). The SET HASDM Density Database. *Space Weather*, *19*(4), e2020SW002682. doi: 10.1029/2020SW002682

Vallado, D. (2013). *Fundamentals of Astrodynamics and Applications* (Fourth Edi ed.). Hawthorne, CA: Microcosm Press and Springer. doi: 10.2514/2.4291

Velicogna, I., & Wahr, J. (2005). Greenland mass balance from GRACE. *Geophysical Research Letters*, *32*(18), 1–4. doi: 10.1029/2005GL023955

Waldron, Z., Garcia-Sage, K., Thayer, J., Sutton, E., Ray, V., Rowlands, D., … Berland, G. (2023, June). *Data archive accompanying "Assessing Thermospheric Neutral Density Models using GEODYN's Precision Orbit Determination".* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.8015369` doi: 10.5281/zenodo.8015369

Walker, A., Mehta, P., & Koller, J. (2014). Drag coefficient model using the Cercignani-Lampis-Lord gas-surface interaction model. *Journal of Spacecraft and Rockets*, *51*(5), 1544–1563. doi: https://doi.org/10.2514/1.A32677

Walterscheid, R. L., Chen, M. W., Chao, C. C., Gegenheimer, S., Cabrera-Guzman, J., & McVey, J. (2023). Comparative Accuracies of Models for Drag Prediction During Geomagnetically Disturbed Periods: A First Principles Model Versus Empirical Models. *Space Weather*, *21*, e2022SW003332. doi: 10.1029/2022SW003332

Zelensky, N. P., Lemoine, F. G., Ziebart, M., Sibthorpe, A., Willis, P., Beckley, B. D., ... Luceri, V. (2010). DORIS/SLR POD modeling improvements for Jason-1 and Jason-2. *Advances in Space Research*, *46(12)*, 1541-1558. doi: 10.1016/J.ASR.2010.05.008