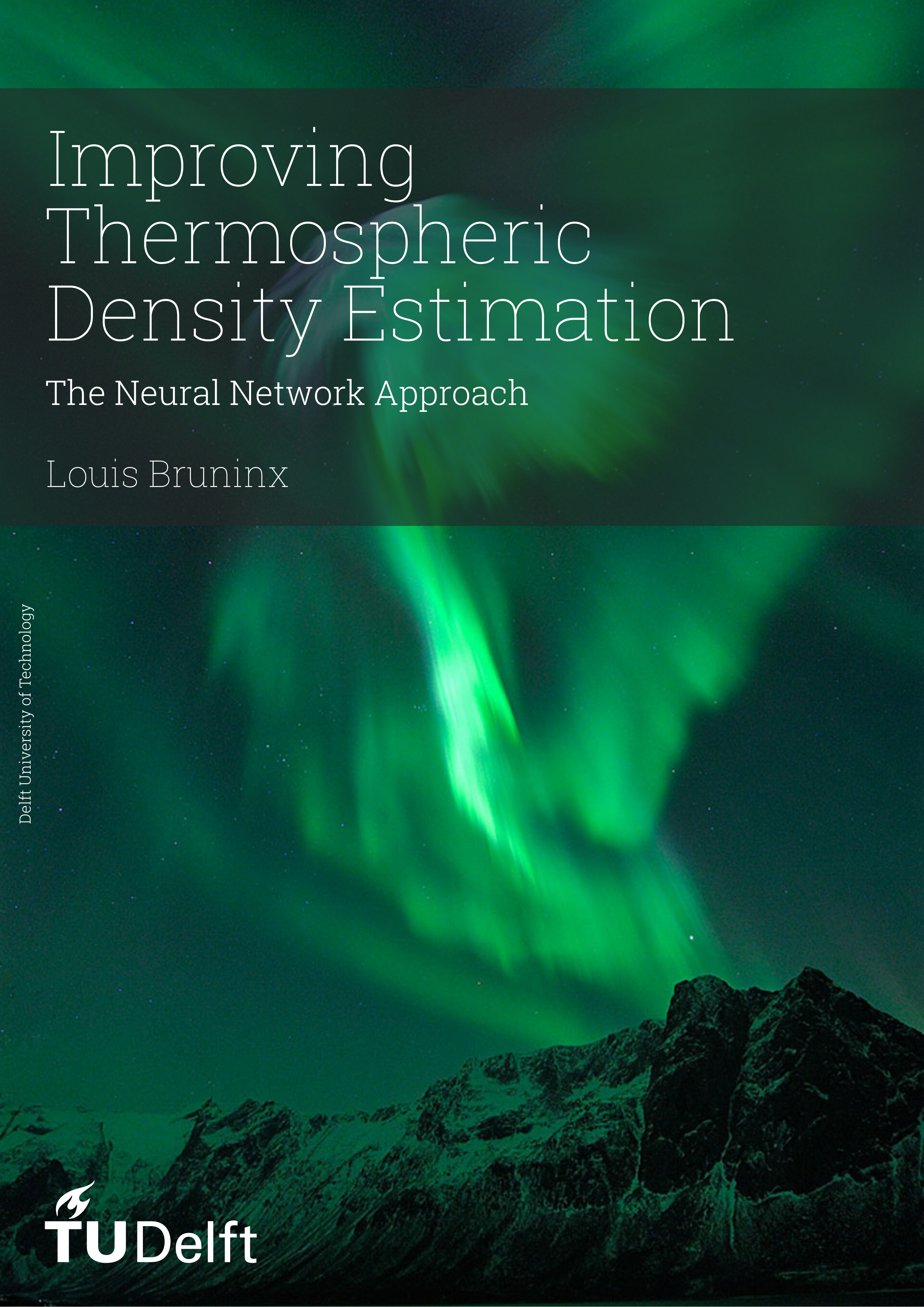


Improving Thermospheric Density Estimation

The Neural Network Approach

Louis Bruninx

Delft University of Technology

TU Delft

Improving Thermospheric Density Estimation

The Neural Network Approach

by

Louis Bruninx

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Friday May 1 2026 at 14:00.

Supervisor TU Delft:	Steve Gehly
Supervisor Ecosmic:	Francesca Andreoli
Chair:	Jeannette Heiligers
Independent Examiner:	Onur Çelik
Project Duration:	May, 2025 - May, 2026
Faculty:	Faculty of Aerospace Engineering, Delft

Cover: A spooky skyscape to celebrate Halloween by B. Jørgensen and
ESA under ESA Standard Licence

Preface

This thesis was carried out as the final part of the MSc programme in spaceflight dynamics at TU Delft, in collaboration with Ecosmic. The work focuses on thermospheric density modelling for low Earth orbit applications, with particular relevance to orbit propagation and space traffic management. The subject of this thesis attracted me because it combines astrodynamics, data analysis, space weather and neural networks within a single research project, allowing me to gain experience in multiple fields at the same time.

I would like to sincerely thank Steve Gehly and Francesca Andreoli for their guidance, feedback and support throughout this thesis. I am also grateful to Benedetta Cattani, Gaia Roncalli and Imane Marouf for allowing me to become a part of the Ecosmic team throughout my internship and thesis. Additionally, I would also like to thank Jeannette Heiligers and Onur Çelik for agreeing to serve on the thesis committee.

Finally, I would like to thank my family and friends for their support and for always encouraging me to be the best version of myself.

Louis Bruninx
Delft, April 2026

Abstract

Atmospheric drag remains one of the main sources of uncertainty in low Earth orbit propagation because the thermospheric density depends strongly on solar and geomagnetic activity and can vary substantially over time and space. This uncertainty directly affects orbit determination, orbit propagation and conjunction assessment and is therefore highly relevant for operational applications such as space traffic management. At the same time, the density model used in such applications must not only be sufficiently accurate, but also computationally efficient and compatible with near-real-time data availability constraints. This creates a practical trade-off between physical fidelity, operational robustness and computational cost.

This thesis investigates whether a neural network-based correction to the JB2008 empirical thermospheric density model can improve atmospheric density estimation and, more importantly, whether such improvements translate into better orbit propagation performance. Rather than replacing the empirical model entirely, a hybrid correction framework is developed in which JB2008 provides the baseline density estimate and a lightweight feed-forward neural network learns a correction using baseline density, solar, geomagnetic, spatial and temporally embedded input features. This design is motivated by the desire to preserve the physical and operational utility of the empirical baseline while still allowing the model to capture systematic, time-varying deviations linked to changing space weather conditions.

To support operational relevance, the methodology is explicitly designed with near-real-time use in mind. The input feature set includes solar and geomagnetic drivers, physically motivated temporal embeddings that represent thermospheric memory effects and trust score features that quantify the reliability of outage-prone high-cadence space weather inputs. The model is trained on accelerometer-derived density data from the CHAMP, GOCE and GRACE missions. A key methodological choice is the use of a simple feed-forward architecture rather than a recurrent sequence model. By embedding the relevant temporal history directly into the input features, the approach remains compatible with fragmented datasets and avoids dependence on continuous sequences, while still allowing delayed thermospheric responses to be represented.

Model performance is evaluated in two complementary stages. In stage A, the neural network is validated directly against accelerometer-derived density estimates. The results show that the correction model reduces the mean and spread of the log-residual error distribution relative to JB2008 and yields reconstructed density time series that better match the reference densities in both phase and amplitude. These findings indicate that the model learns more than a static offset correction and captures part of the time-varying thermospheric variability that is not fully represented by the empirical baseline. In stage B, the corrected density model is assessed in orbit propagation against Swarm A precise orbit determination data over seven validation arcs spanning quiet, moderate and severe geomagnetic conditions. The neural network correction model achieves the lowest overall 24-hour along-track RMS error among the tested density configurations and performs best during disturbed periods, when empirical density models tend to degrade most strongly. This suggests that the learned correction is not only beneficial when reconstructing density, but can also improve orbit propagation performance.

Overall, the results of this thesis indicate that a lightweight neural-network correction to JB2008 can provide a promising and operationally relevant improvement to empirical thermospheric density modelling. In particular, the findings suggest that a comparatively simple feed-forward architecture, when combined with physically informed temporal feature design and explicit data-reliability indicators, is already capable of learning useful corrections to an empirical baseline. At the same time, the final orbit propagation validation is limited to a single satellite and a restricted set of validation arcs. The results should therefore be interpreted as a promising proof of concept rather than as definitive evidence of broad operational superiority across all low Earth orbit regimes. Broader POD-based validation across additional satellites, altitude regimes and operational scenarios is required before more general conclusions can be drawn.

Nomenclature

List of Abbreviations

Term	Description
CAD	Computer-Aided Design
CHAMP	Challenging Minisatellite Payload
CTIPe	Coupled Thermosphere Ionosphere Plasmasphere Electrodynamics model
DSMC	Direct Simulation Monte Carlo
DTM	Drag Temperature Model
ECDF	Empirical Cumulative Distribution Function
EUV	Extreme Ultraviolet
EWMA	Exponentially Weighted Moving Average
FFILL	forward-filling
GCM	General Circulation Model
GFZ	German Research Centre for Geosciences
GITM	Global Ionosphere-Thermosphere Model
GOCE	Gravity Field and Steady-State Ocean Circulation Explorer
GPS	Global Positioning System
GRACE	Gravity Recovery and Climate Experiment
GRS80	Geodetic Reference System 1980
GSM	Geocentric Solar Magnetospheric
HASDM	High Accuracy Satellite Drag Model
IDEA	Iterative Driver Estimation and Assimilation
IMF	Interplanetary Magnetic Field
ISGI	International Service of Geomagnetic Indices
JB2008	Jacchia-Bowman 2008 model
LEO	Low Earth Orbit
LN	Layer Normalisation
LST	Local Solar Time
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MLP	Multilayer Perceptron
MSE	Mean Squared Error
NCAR	National Center for Atmospheric Research
NN	Neural Network
NRLMSISE-00	Naval Research Laboratory Mass Spectrometer and Incoherent Scatter Radar Exosphere model

Term	Description
POD	Precise Orbit Determination
RKF78	Runge–Kutta–Fehlberg 7(8)
RMS	Root Mean Square
RMSE	Root Mean Square Error
SET	Space Environment Technologies
SiLU	Sigmoid Linear Unit
SMA	Simple Moving Average
SRP	Solar Radiation Pressure
SSN	Sunspot Number
STM	Space Traffic Management
TDB	Barycentric Dynamical Time
TIE-GCM	Thermosphere-Ionosphere-Electrodynamics General Circulation Model
TPMC	Test Particle Monte Carlo
UKF	Unscented Kalman Filter
UTC	Coordinated Universal Time
UV	Ultraviolet
WDC	World Data Center

List of Symbols

Term	Description
AE	Auroral electrojet index
AL	Lower auroral electrojet index
α	Decay coefficient of the exponentially weighted moving average
a_p	Linearised planetary geomagnetic activity index
A_{ref}	Reference area used in Stage B propagations
$ASY\text{-}D$	Asymmetric declination-component storm-time disturbance index
$ASY\text{-}H$	Asymmetric horizontal-component storm-time disturbance index
AU	Upper auroral electrojet index
BC	Ballistic coefficient
$B_{\mathrm{epoch\_count}}$	Number of mini-batches per epoch
B_{size}	Mini-batch size
B_x	Sun–Earth aligned component of the interplanetary magnetic field
B_y	East–west component of the interplanetary magnetic field
B_z	North–south component of the interplanetary magnetic field
C_d	Drag coefficient
$C_{d,k}$	Drag-coefficient estimate at secant iteration k
$\mathrm{corr}_{\hat{r},r}$	Pearson correlation between predicted and target log-residuals
C_{var}	Trust score for an outage-prone variable category
D	Atmospheric drag force
$\Delta\Phi$	Angular displacement of a periodic variable over an averaging interval
$\Delta\mathbf{r}(t)$	Position deviation between propagated and POD states
Δt	Sampling interval

Term	Description
d_{emb}	Satellite embedding dimension
d_{input}	Input dimension
$d\Phi_{MP}/dt$	Newell coupling function; rate of magnetic flux opening at the magnetopause
D_{st}	Disturbance storm time index
d_{trunk}	Hidden width of the shared trunk
$e_{\text{AT}}(t)$	Along-track position error
ϵ	Akasofu coupling parameter
e_r	Log-residual prediction error
$e(s)$	Learned embedding associated with satellite identifier s
η	Learning rate
$F_{10.7}$	Solar radio flux proxy used to represent EUV-related solar activity
F_{30}	Solar radio flux proxy at 30 cm wavelength
$f(C_d)$	Mean along-track error objective used in C_d tuning
γ	Weight of the correlation term in the main loss
h	Geodetic altitude used in the atmospheric density callback
H	EWMA half-life
$h(t)$	Learned hidden representation of the input
$h_0(t)$	Initial hidden representation after the input projection
I_N	Discrete time integral over a finite window
K	Number of residual blocks in the trunk
K_p	Planetary geomagnetic activity index
λ	Geodetic longitude used in the atmospheric density callback
λ_{sat}	Penalty coefficient for the satellite refinement output
λ_{wd}	Weight-decay coefficient
$\mathcal{L}_{\text{corr}}$	Correlation-based loss term
$\mathcal{L}_{\text{main}}$	Main training objective combining regression and correlation loss
$\mathcal{L}_{\text{main,val}}$	Validation value of the main training objective
l_{out}	Outage length in minutes
$\mathcal{L}_{\text{reg}}$	Regression loss term
$\mathcal{L}_{\text{sat}}$	Penalty applied to the satellite refinement branch
$\mathcal{L}$	Total loss function
m	Satellite mass
M_{10}	Solar proxy used in JB2008 to represent delayed solar emission effects
MAPE_ρ	Mean absolute percentage error in density space
$\overline{\text{MSE}}_{\text{AT}}^{(\text{arc})}$	Arc-level mean of the window-wise along-track MSE values
$\text{MSE}_{\text{AT}}^{(w)}$	Mean squared along-track error for window w
μ_s	Satellite-specific mean of the raw log-residual target
μ_{var}	Trust-score offset parameter controlling the midpoint of the decay
μ_x	Mean of feature x in the training set
N	Number of samples in an evaluation set
N_{int}	Window length used for discrete time integration
N_{SMA}	Window length of the simple moving average
N_{train}	Number of training samples
ν	True anomaly
N_w	Number of valid comparison points in a 24-hour propagation window

Term	Description
ω	Argument of periapsis
PCN	Polar Cap North index
PCS	Polar Cap South index
p_{drop}	Dropout probability applied in hidden layers
p_{feat}	Feature-dropout probability
ϕ	Geodetic latitude used in the atmospheric density callback
P	Early-stopping patience
P	Period of a periodic variable
p_{sat}	Satellite-ID dropout probability
ψ	Weight of the regression term in the main loss
p_{sw}	Solar wind dynamic pressure
$r(t)$	Logarithmic density residual relative to the JB2008 baseline
$\hat{r}(t)$	Predicted log-residual correction
ρ	Atmospheric mass density
ρ_{JB2008}	Baseline JB2008 thermospheric density
$\hat{\rho}_{\mathrm{NN}}(t)$	Neural-network-corrected thermospheric density
ρ_{obs}	Observed accelerometer-derived thermospheric density
ρ_{sw}	Solar wind proton density
$\mathrm{RMS}_{\mathrm{AT}}^{(w)}$	Root mean square along-track error for window w
$\mathrm{RMSE}_{\hat{r},r}$	Root mean square error in log-residual space
$\mathbf{r}_{\mathrm{POD}}(t)$	POD reference position vector
R_{sat}	Satellite-head usage ratio, defined as the mean absolute satellite-refinement magnitude divided by the mean absolute global-branch magnitude
$\mathbf{r}_{\mathrm{sim}}(t)$	Propagated position vector
S	Reference cross-sectional area
s	Satellite identifier
S_{10}	Solar proxy used in JB2008 to represent additional solar spectral variability
σ_r	Global standard deviation of the raw log-residual target
σ_x	Standard deviation of feature x in the training set
$\sigma(z)$	Sigmoid activation function
$\mathcal{S}_{\mathrm{val}}$	Set of satellites used in validation
$SYM\text{-}D$	Symmetric declination-component storm-time disturbance index
$SYM\text{-}H$	Symmetric horizontal-component storm-time disturbance index
τ	Lag amount used in time-shifted features
τ_{var}	Trust-score scale parameter controlling the steepness of the decay
T_{sw}	Solar wind proton temperature
t_{TDB}	Simulation epoch expressed in TDB seconds since J2000
t_{yr}	Decimal year used in the empirical spacecraft-mass model
u	Argument of latitude
V	Relative velocity between satellite and atmosphere
vB_s	Rectified solar wind electric field proxy
$\hat{\mathbf{v}}_{\mathrm{POD}}(t)$	Unit vector in the POD along-track direction
$\mathbf{v}_{\mathrm{POD}}(t)$	POD reference velocity vector
v_{sw}	Solar wind speed
W_{arc}	Number of evaluated 24-hour windows within an arc

Term	**Description**
$x_{\cos}$	Cosine embedding of a periodic variable
$\vec{x}_{\text{final}}$	Renormalised periodic-feature vector used as final input
$x_{\log}$	Log-transformed version of feature x
$x_{\max}$	Maximum training-set value used for min–max scaling
$\vec{x}_{\text{mean}}$	Averaged periodic-feature vector before renormalisation
$x_{\min}$	Minimum training-set value used for min–max scaling
$x_{\sin}$	Sine embedding of a periodic variable
$x(t)$	Input feature vector at time t
x	Generic input variable or feature
$y(t)$	Normalised bias-corrected training target
Y_{10}	Solar proxy used in JB2008 to represent longer-lagged solar forcing effects
$\hat{y}_{\text{global}}(t)$	Global shared prediction branch
$\hat{y}(t)$	Predicted normalised log-residual target
$\hat{y}_{\text{sat}}(t)$	Satellite-conditioned refinement term
z	Generic scalar input to an activation function

Contents

1

Introduction

The low Earth orbit (LEO) environment is becoming increasingly congested due to the rapid growth in operational satellites, large constellations and debris objects. In this setting, accurate orbit determination and propagation is essential for maintaining safe and sustainable space operations. A key application is space traffic management (STM), where reliable orbit forecasts are required for conjunction assessment, collision risk analysis and operational decision making. For objects in LEO, however, one of the main factors limiting orbit determination and propagation accuracy is atmospheric drag.

Atmospheric drag depends strongly on the neutral mass density of the thermosphere. Unlike many other perturbing forces, thermospheric density is highly dynamic and varies in response to solar and geomagnetic forcing. Changes in solar ultraviolet and extreme ultraviolet radiation alter the background thermal structure of the upper atmosphere, while geomagnetic disturbances can inject additional energy into the high-latitude thermosphere and produce substantial storm-time density increases. As a result, the density encountered by a satellite at a given altitude is not fixed, but changes continuously with solar activity, geomagnetic conditions, latitude, local solar time and the recent history of atmospheric forcing. This makes thermospheric density one of the most difficult quantities to model accurately for operational orbit propagation.

A wide range of thermospheric density modelling approaches has been developed to address this problem. Empirical and semi-empirical models, such as JB2008 and NRLMSISE-00, remain widely used in orbit propagation because they are computationally efficient, robust and easy to evaluate. However, their performance often degrades during strongly disturbed geomagnetic conditions, where rapid and localised density variations become more difficult to capture. Physics-based and data-assimilative models can provide improved physical realism and better responsiveness to changing conditions, but they are typically more computationally demanding and more difficult to integrate into lightweight operational workflows. This creates a practical trade-off between physical fidelity and operational usability.

Within this context, machine learning (ML) offers a promising intermediate strategy. Rather than replacing an empirical model entirely, a neural network can be used to learn a correction to an existing baseline density estimate. Such a hybrid approach retains the speed and operational convenience of the underlying empirical model while allowing for data-driven corrections. This is particularly attractive for STM-related applications, where models must not only be accurate, but also fast and stable.

This thesis investigates such a hybrid correction framework in collaboration with the Italian space start-up Ecosmic [1]. The central idea is to use a neural network to provide a correction to the JB2008 empirical density model. Instead of predicting absolute density directly, the neural network learns the residual difference between accelerometer-derived density observations and the JB2008 baseline, using solar, geomagnetic, spatial and temporal inputs. In this way, the work aims to improve thermospheric density estimation while preserving the operational advantages of a widely used empirical model.

---

[1] `https://www.ecosmic.space/`

The research has two main objectives. The first objective is to determine whether a neural-network-based correction to JB2008 can improve density reconstruction relative to accelerometer-derived thermospheric density observations across multiple satellites and varying space weather conditions. The second objective is to determine whether such improvements in density reconstruction also translate into improved orbit propagation performance. This second objective is essential, since better agreement with density reference data does not automatically guarantee reduced orbit prediction error in an operational propagation setting.

To address these objectives, a two-stage evaluation strategy is adopted. First, the correction model is evaluated directly in density space against accelerometer-derived density references. This provides a development and validation environment in which different model configurations can be compared efficiently. Second, the final model is tested in an independent orbit propagation setting against precise orbit determination (POD) data, where the corrected density model is assessed on its ability to reduce along-track propagation errors relative to standard empirical baselines. This two-stage design makes it possible to distinguish between improvement in reconstructed density and improvement in the operational quantity of interest, namely orbit propagation accuracy.

The remainder of this thesis is structured as follows. Chapter 2 reviews the theoretical background of atmospheric drag uncertainty, the space weather processes that drive thermospheric density variability and the main model classes used for thermospheric density estimation, before motivating the selected hybrid correction approach. Chapter 3 describes the acquisition, preprocessing and feature engineering of the satellite, solar and geomagnetic datasets used in this work. Chapter 4 presents the neural network correction framework, including the target formulation, model architecture, training procedure and two-stage evaluation design. Chapter 5 then presents the validation and orbit propagation results. Finally, the thesis closes with the main conclusions in Chapter 6 and recommendations for future work are presented in Chapter 7. For reference, Appendix A provides a complete overview of the final input features and their associated normalisation strategies, while Appendix B presents a high-level planning overview of the thesis.

2

Theoretical Foundation and Literature Review

Atmospheric drag is one of the key non-gravitational forces affecting the motion of satellites in LEO. Accurate drag modelling is essential for orbit determination, orbit propagation and conjunction assessment. However, drag remains a major source of uncertainty in orbit prediction due to the complex and highly dynamic nature of the thermosphere. This uncertainty propagates directly into conjunction assessment and collision-risk analysis, both of which rely on accurate orbital state information. As a result, atmospheric drag modelling remains a central challenge in maintaining the safety and sustainability of the low Earth orbit environment.

This chapter reviews the main sources of uncertainty in atmospheric drag modelling and examines the principal approaches used to model thermospheric density. It first discusses the main contributors to drag uncertainty in Section 2.1. It then introduces the physical space weather processes that drive thermospheric density variability in Section 2.2. Next, the main classes of thermospheric density models are reviewed in Section 2.3, including empirical, physics-based and data-assimilative approaches together with their operational trade-offs. The chapter then considers drag coefficient estimation as an additional source of uncertainty in Section 2.4 and motivates the use of a neural network correction approach in Section 2.5. Finally, the research motivation, goals and hypotheses of this thesis are presented in Section 2.6.

2.1. Sources of Uncertainty in Atmospheric Drag Modelling

The atmospheric drag force depends on four components: the atmospheric density, the drag coefficient (C_d), the relative velocity between the satellite and the surrounding atmosphere and the reference area exposed to the flow. Accurate orbit propagation therefore requires that each of these quantities be estimated as reliably as possible. In practice, several of these terms are affected by factors that vary strongly in space and time, including solar and geomagnetic activity, atmospheric composition and spacecraft attitude. Consequently, drag modelling is inherently uncertain and requires accurate representation of multiple coupled variables.

The drag force D acting on a satellite is commonly expressed as

$$D = \frac{1}{2}\rho C_d V^2 S \tag{2.1}$$

where ρ is the atmospheric density, C_d is the drag coefficient, V is the relative velocity between the satellite and the atmosphere and S is the reference cross-sectional area perpendicular to the flow.

According to Equation (2.1) and Mostaza Prieto, Graziano, and Roberts (2014), the main contributors to drag-force uncertainty are the atmospheric density, the drag coefficient and the relative velocity term, which may itself be affected by atmospheric winds. The reference area S can also contribute to drag uncertainty, particularly for satellites and debris with changing attitude and complex or unknown geometry. In practice, the effects of drag coefficient and reference area are often combined into a single parameter through the ballistic coefficient (BC), defined as

$$BC = \frac{m}{C_d S} \tag{2.2}$$

where m is the mass of the satellite. The ballistic coefficient describes the extent to which an object resists deceleration due to atmospheric drag, with a higher ballistic coefficient corresponding to a lower sensitivity to drag.

Among these uncertainty sources, atmospheric density is generally the most difficult to characterise accurately because it varies strongly with solar forcing, geomagnetic activity, altitude, latitude and local time. For this reason, thermospheric density modelling has become a central topic in satellite drag research.

2.2. Space Weather and Thermospheric Response

Thermospheric density is not solely determined by altitude and local time, but varies continuously in response to external forcing from the Sun and the coupled Sun-Earth system. At the most general level, this variability is driven by two main mechanisms. First, solar ultraviolet (UV) and extreme ultraviolet (EUV) radiation directly heat the upper atmosphere and set much of its background state. Second, solar wind-magnetosphere coupling injects additional energy into the high-latitude upper atmosphere during disturbed geomagnetic conditions, leading to increased heating and storm-time density increases. As a result, the thermosphere density experiences substantial fluctuations across timescales ranging from minutes to the solar cycle, making density one of the most difficult quantities to predict accurately for orbit propagation and drag modelling.

For the present thesis, the most relevant point is that the observed density in the thermosphere is the combined outcome of multiple linked processes, including solar radiative forcing, solar wind driving, magnetospheric energy transfer, high-latitude ionospheric heating and the delayed cooling of the thermosphere. The following subsections therefore introduce the main physical processes through which space weather affects the thermosphere, with particular attention to the drivers and indices used later in the data processing and methodology chapters.

2.2.1. Solar Radiative Forcing of the Thermosphere

The Sun is the primary energy source of the thermosphere under both quiet and disturbed conditions. In particular, solar UV and EUV radiation are absorbed in the upper atmosphere, where they drive photoionisation and direct heating. This radiative input largely determines the background thermal structure of the thermosphere and is therefore the primary neutral density driver. When solar EUV irradiance increases, the thermosphere warms, its scale height increases and the upper atmosphere expands. As a consequence, the neutral density at a fixed orbital altitude can increase significantly even in the absence of strong geomagnetic activity.

This solar radiative forcing varies over multiple timescales. On long timescales, the approximately 11-year solar cycle controls the background level of EUV output, producing large changes in mean thermospheric density between solar minimum and solar maximum. On shorter timescales, solar rotation and active-region evolution introduce day-to-day and 27-day variability, while events such as solar flares can produce rapid but typically shorter-lived irradiance enhancements (Kutiev et al. 2013).

Because direct continuous EUV observations are not always available over the full historical record, thermospheric modelling has long relied on solar activity proxies. The most widely used of these is the $F_{10.7}$ index, which measures solar radio flux at a wavelength of 10.7 cm and is strongly correlated with solar EUV variability relevant to thermospheric heating (Dudok De Wit and Bruinsma 2017). In operational density models, $F_{10.7}$ is therefore commonly used as a practical proxy for the solar radiative state of the upper atmosphere. More advanced empirical models such as JB2008 extend this idea by using additional proxy indices, including S_{10}, M_{10} and Y_{10}, to represent solar spectral bands and delayed solar influences that are not fully captured by $F_{10.7}$ alone (Bowman et al. 2008).

The main point is that solar radiative forcing provides the baseline state upon which geomagnetic disturbances are superimposed. Variables such as $F_{10.7}$, S_{10}, M_{10} and Y_{10} are compact representations of the evolving solar energy input that drives the long-term thermospheric density variability relevant for atmospheric drag.

2.2.2. Solar Wind-Magnetosphere Coupling and Geomagnetic Driving

In addition to direct radiative forcing, the Sun influences the thermosphere indirectly through the solar wind and its embedded interplanetary magnetic field (IMF). The solar wind is a continuous outflow of magnetised plasma from the solar corona, carrying mass, momentum and magnetic field into interplanetary space. When this flow reaches Earth, it interacts with Earth's magnetosphere, which forms the main interface through which solar wind energy can be transferred into the Earth system. The efficiency of this transfer depends not only on the solar wind conditions, but also strongly on the orientation of the IMF relative to the geomagnetic field.

A particularly important quantity is the north-south component of the IMF, denoted B_z. When B_z is directed southward, opposite to the northward geomagnetic field at the dayside magnetopause, magnetic reconnection becomes more efficient as illustrated in Figure 2.1. This process opens previously closed geomagnetic field lines on the dayside and allows solar wind energy, momentum and magnetic flux to enter the magnetosphere more readily. The opened field lines are then convected over the polar caps into the magnetotail, where they can reconnect again on the nightside. This large-scale circulation enhances magnetospheric convection and drives stronger geomagnetic activity (Lockwood, McWilliams, et al. 2020; Lockwood 2022). By contrast, when B_z is directed northward, large-scale dayside reconnection is generally less effective, so the coupling between the solar wind and the magnetosphere is weaker, as illustrated in Figure 2.2.

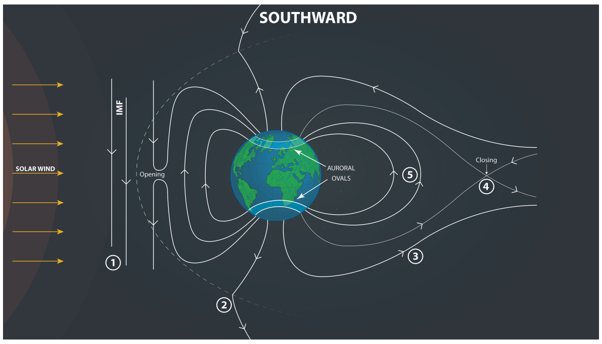


Figure 2.1: Schematic illustration of solar wind-magnetosphere interaction during southward IMF. Southward B_z promotes more efficient dayside magnetic reconnection, opening field lines and enhancing energy transfer into the magnetosphere. Credit: Dylan Leivers (Royal Belgian Institute for Space Aeronomy 2020)

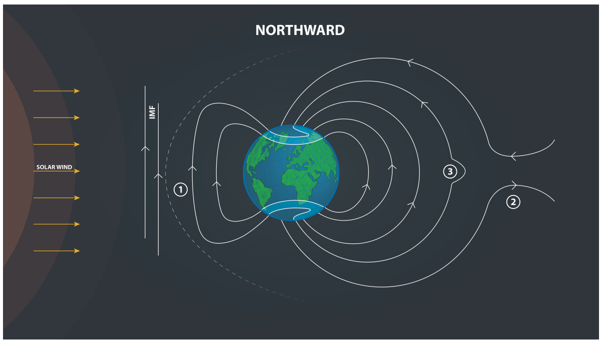


Figure 2.2: Schematic illustration of solar wind-magnetosphere interaction during northward IMF. Under northward B_z, large-scale dayside reconnection is generally less effective, leading to weaker solar wind–magnetosphere coupling. Credit: Dylan Leivers (Royal Belgian Institute for Space Aeronomy 2020)

Although the north-south IMF component B_z is usually the dominant factor controlling the overall strength of solar wind-magnetosphere coupling, the other IMF components can also influence the system response. In particular, the east-west component B_y can modify the pattern and asymmetry of high-latitude ionospheric convection and auroral activity and can therefore affect how and where energy is deposited into the upper atmosphere. By contrast, the role of B_x, the IMF component approximately aligned with the Sun–Earth direction, is generally more subtle and is often treated as secondary in comparison with B_z, although it may still influence coupling geometry and geomagnetic activity under some conditions (Lockwood, Owens, et al. 2016; Cai et al. 2022; Laundal et al. 2018).

The effects of enhanced solar wind-magnetosphere coupling are strongest near the poles. When coupling becomes more efficient, more energy is transferred into the high-latitude upper atmosphere. This leads to stronger electrical currents in the auroral regions and more particle precipitation into the atmosphere (Verkhoglyadova et al. 2016). Together, these processes heat the polar thermosphere and cause it to expand, which increases the local neutral density at LEO altitudes. During stronger geomagnetic disturbances, this density enhancement does not remain confined to the polar regions, but can spread more broadly through the thermosphere.

The IMF orientation is not the only factor that determines how strongly the solar wind affects the Earth system. The speed and strength of the solar wind also matter. Faster solar wind streams generally carry more energy and stronger magnetic fields can lead to more effective coupling when their orientation is favourable. In addition, solar wind density and dynamic pressure influence how strongly the magnetosphere is compressed during disturbed periods (Tenfjord and Østgaard 2013). For this reason, no single solar wind parameter fully describes the level of upstream driving and multiple variables are needed to characterise the forcing acting on the magnetosphere-ionosphere-thermosphere system.

For this reason, many studies employ solar wind coupling functions, which combine multiple upstream variables into a single quantity intended to approximate the efficiency of solar wind-magnetosphere energy transfer. Common examples include the Akasofu ϵ parameter (Akasofu 1981; Koskinen and Tanskanen 2002), the Newell coupling function (Newell et al. 2007) and simpler quantities such as vB_s. These are not direct measurements of thermospheric heating, but compact proxies for the upstream drivers that can ultimately lead to enhanced geomagnetic activity, high-latitude energy deposition and storm-time density increases. They can provide a physically motivated way to supplement the raw solar wind and IMF variables with features that better represent the effectiveness of the transfer of energy from the solar wind into the Earth system.

2.2.3. High-Latitude Energy Deposition and Storm-Time Thermospheric Response

During geomagnetically disturbed conditions, a substantial fraction of the energy transferred from the solar wind into the Earth system is deposited at high latitudes. This occurs primarily through two mechanisms: Joule heating associated with enhanced ionospheric currents and electric fields and particle precipitation from the magnetosphere into the upper atmosphere (Verkhoglyadova et al. 2016; Wilson et al. 2006). Together, these processes provide the main path through which geomagnetic forcing influences the thermosphere during storms.

Joule heating arises when ionospheric plasma is driven across the geomagnetic field by magnetospheric convection. Through collisions between ions and neutral atmosphere, part of this electrodynamic energy is converted into heat in the high-latitude thermosphere. At the same time, precipitating particles deposit energy directly into the upper atmosphere, especially in the auroral regions. The combined effect is a rapid increase in thermospheric temperature, followed by expansion of the neutral atmosphere and an increase in density at a given altitude. These density enhancements are often strongest first at high latitudes, where the energy input is largest (Wilson et al. 2006).

The storm-time response is not purely local. Once the polar thermosphere is heated, pressure gradients intensify and mass and energy are distributed towards lower latitudes. In addition, geomagnetic storms can launch travelling atmospheric disturbances that propagate away from the auroral source regions and contribute to density increases over a much broader spatial domain as explained by Gardner and Schunk (2010). As a result, geomagnetic forcing can influence thermospheric density far beyond the immediate polar regions, which is one of the reasons why atmospheric drag can increase strongly during storm periods.

Besides changing temperature and density, geomagnetic storms can also alter thermospheric composition. Storm-time upwelling at high latitudes can change the relative abundance of major neutral particles and these composition disturbances may carry on into the recovery phase (Gan et al. 2024). This is important because thermospheric density is shaped not only by direct heating, but also by the evolving state of the atmosphere during and after geomagnetic forcing.

The key takeaway is that storm-time thermospheric density is the result of both the instantaneous high-latitude energy input and the subsequent reaction of the global thermosphere. This explains why geomagnetic indices and coupling functions are relevant variables for density modelling.

2.2.4. Geomagnetic Indices and Their Physical Interpretation

To characterise geomagnetic activity, a range of geomagnetic indices have been developed, each summarising a different aspect of the coupled magnetosphere-ionosphere system. In the context of thermospheric density modelling, these indices are useful because they provide compact descriptions of how strongly the Earth system is disturbed and where the associated energy input is most likely to occur.

A first important group consists of the auroral electrojet indices AE, AL and AU, which describe geomagnetic activity in the auroral zones. These indices are derived from ground-based magnetometer measurements at high latitudes and are commonly used as proxies for enhanced auroral electrodynamic activity. The AU and AL indices represent the upper and lower envelope of the magnetic perturbations, while $AE = AU - AL$ provides a measure of the overall auroral electrojet activity (Toyohisa, Masahisa, and Tohru 1998). Because they respond rapidly to high-latitude current fluctuations, these indices are particularly useful for characterising short-term geomagnetic forcing linked to auroral heating of the upper atmosphere.

A second group describes the more global storm-time disturbance field. According to T. et al. (2010), the D_{st} index is derived from near-equatorial magnetometer stations and is widely used as a measure of geomagnetic storm intensity, being closely associated with the strength of the storm-time ring current. The related SYM and ASY indices separate the disturbance field into a symmetric and an asymmetric component. In particular, $SYM\text{-}H$ can be interpreted as a 1-minute version of the hourly D_{st} index, while $ASY\text{-}H$ provides information on asymmetries in the disturbance field that are not captured by D_{st} or $SYM\text{-}H$ alone. These indices are therefore useful for representing both the overall severity and the structure of storm-time geomagnetic activity.

A third group consists of the planetary geomagnetic indices K_p and a_p, which provide a broader measure of global geomagnetic activity. The K_p index is a quasi-logarithmic index derived over successive 3-hour intervals from a global network of ground observatories, while a_p is its linearised equivalent (Matzka et al. 2021). Although these indices have lower temporal resolution than the 1-minute auroral and storm-time indices, they remain widely used in empirical thermospheric modelling because they provide a robust and operationally established summary of geomagnetic disturbance on global scales.

Taken together, these indices describe complementary aspects of geomagnetic forcing. These aspects are auroral electrodynamic activity, global storm-time disturbance intensity, asymmetry in the disturbance field and global geomagnetic variability. For this reason, they provide physically meaningful inputs for modelling the thermospheric density response to geomagnetic activity. Additional high latitude indices, such as polar cap activity indices (e.g. PCN, PCS), can further complement this picture by describing geomagnetic forcing in the polar regions.

2.2.5. Thermospheric Response Timescales and Memory Effects

The thermosphere does not respond instantaneously to changes in solar and geomagnetic forcing. Instead, its state at any given moment reflects both the current external conditions and the recent history of energy input (Kutiev et al. 2013). This delayed behaviour arises because the thermosphere has thermal inertia and because storm-time forcing triggers dynamical adjustments, such as changing winds, circulation patterns and composition, that continue to evolve after the initial energy deposition (Sheng et al. 2017).

As a result, changes in thermospheric density often lag behind the forcing that caused them. The timing of the response depends on the type and strength of the forcing, the altitude and latitude considered and the initial condition of the atmosphere. For example, studies have shown that the thermospheric density response to geomagnetic activity can be delayed by several hours and that stronger storms are often associated with longer response times and more prolonged recovery phases (Wang, Miao, et al. 2022; Zesta and Oliveira 2019). Additionally, the response to periodic solar EUV forcing is not perfectly immediate, but can exhibit a measurable phase delay relative to the driving solar variability (Ren et al. 2021).

The recovery of the thermosphere after geomagnetic disturbances is also gradual rather than instantaneous. Even after the strongest external driving has weakened, elevated temperature, altered circulation and storm-time composition changes can persist for many hours or longer, allowing density anomalies to remain present into the recovery phase (Sheng et al. 2017; Gan et al. 2024). This means that thermospheric density cannot be represented adequately using only instantaneous input values, since part of the observed state is inherited from prior forcing.

The delayed and cumulative nature of the thermospheric response implies that neutral density cannot be represented accurately using only the instantaneous values of solar and geomagnetic inputs. At any given moment, the observed density is shaped not only by the current external forcing, but also by the recent history of heating, circulation changes and composition disturbances. As a result, any modelling approach that aims to describe thermospheric density realistically must account for this temporal dependence and for the fact that part of the thermospheric state is a result of prior forcing.

For convenience, Table 2.1 summarises the main solar, interplanetary and geomagnetic indices discussed in this section together with their physical interpretation and relevance for thermospheric density modelling.

Table 2.1: Summary of the main solar, interplanetary and geomagnetic indices discussed in this section and their physical interpretation for thermospheric density modelling.

Index / Parameter	Category	Physical meaning / relevance
$F_{10.7}$	Solar flux proxy	Solar radio flux proxy widely used to represent EUV-driven background thermospheric heating.
S_{10}	Solar flux proxy	Proxy representing additional solar spectral variability relevant to thermospheric heating in JB2008.
M_{10}	Solar flux proxy	Proxy representing solar emission effects with delayed thermospheric influence in JB2008.
Y_{10}	Solar flux proxy	Proxy representing longer-lagged solar forcing effects used in JB2008.
B_x	IMF component	IMF component approximately aligned with the Sun-Earth direction. Usually secondary, but can affect coupling geometry under some conditions.
B_y	IMF component	East-west IMF component. Influences asymmetry of high-latitude convection and auroral activity.
B_z	IMF component	North-south IMF component. Dominant control on dayside reconnection efficiency and therefore solar wind–magnetosphere coupling strength.
v_{sw}	Solar wind plasma	Solar wind speed. Faster streams generally carry more energy.
ρ_{sw}	Solar wind plasma	Solar wind proton density. Contributes to dynamic pressure and magnetospheric compression.
p_{sw}	Solar wind plasma	Solar wind dynamic pressure. Affects magnetospheric compression and disturbed-time response.
T_{sw}	Solar wind plasma	Solar wind proton temperature. Provides additional context on upstream solar wind conditions.
AE	Auroral index	Overall auroral electrojet activity. Proxy for enhanced high-latitude electrodynamic activity.
AL	Auroral index	Lower envelope of auroral-zone magnetic perturbations. Commonly associated with the westward electrojet.
AU	Auroral index	Upper envelope of auroral-zone magnetic perturbations. Commonly associated with the eastward electrojet.
D_{st}	Storm-time index	Global storm-time disturbance index closely associated with ring-current intensity.
$SYM\text{-}H$	Storm-time index	1-minute analogue of D_{st}. Represents the symmetric part of the storm-time disturbance field.
$ASY\text{-}H$	Storm-time index	Represents the asymmetric part of the storm-time disturbance field and local-time asymmetries in current systems.
$SYM\text{-}D$	Storm-time index	Declination-component symmetric disturbance index.
$ASY\text{-}D$	Storm-time index	Declination-component asymmetric disturbance index.
K_p	Planetary index	Quasi-logarithmic 3-hour planetary geomagnetic activity index summarising global disturbance level.
a_p	Planetary index	Linearised equivalent of K_p, widely used in empirical thermospheric models.
PCN / PCS	Polar cap index	Polar-cap geomagnetic activity indices describing forcing in the polar cap region.
vB_s	Coupling function	Rectified solar wind electric field proxy. Captures the southward B_z driving.
ϵ	Coupling function	Akasofu coupling parameter. Proxy for solar wind energy input into the magnetosphere.
$d\Phi_{MP}/dt$	Coupling function	Newell coupling function. Proxy for the rate of magnetic flux opening and overall coupling efficiency.

2.3. Thermospheric Density Modelling Approaches

The atmosphere is in constant motion, expanding, shrinking and turning. Therefore, the density ρ constantly varies per location and over time. As discussed previously, one of the main driving factors behind this is the Sun and its activity. There is the impact of the 11-year-long solar cycle (solar wind), combined with the shorter-term solar activity, such as coronal mass ejections, solar flares and solar energetic particle emissions.

During periods of elevated solar activity, increased EUV flux heats the thermosphere, causing it to expand and thereby increasing density at a given altitude (Emmert et al. 2021). Conversely, during solar minima, the atmosphere contracts, resulting in lower densities. Additionally, geomagnetic storms can cause density enhancements of up to 100% or more at altitudes between 300–500 km (Krauss et al. 2015; Sutton, Forbes, and Nerem 2005).

Given these dynamics, accurate modelling of atmospheric density is critical for operational orbit determination, re-entry prediction and space traffic management applications. The following subsections provide an overview of empirical, physics-based and data-assimilative atmospheric density models, their methodologies and their respective advantages and limitations.

2.3.1. Empirical and Semi-Empirical Models

Empirical atmospheric density models are constructed by fitting large datasets of historical satellite and ground-based measurements to statistical formulations. In practice, many widely used models also incorporate simplified physical relationships and are therefore more accurately described as semi-empirical models. These models are popular in operational orbit determination because they are computationally efficient, robust and they provide reasonable global density estimates. However, their performance often degrades during strong geomagnetic events, where rapid local density variations are more difficult to capture.

The most commonly used semi-empirical thermospheric models include:

- **NRLMSISE-00**: Developed by the Naval Research Laboratory, this model extends the MSIS series by incorporating additional incoherent scatter radar and mass spectrometer data to estimate temperature and composition profiles from the surface up to the exosphere (Picone et al. 2002).
- **JB2008**: The Jacchia-Bowman 2008 model enhances the Jacchia series by using density data from CHAMP and GRACE, incorporating additional solar proxies such as S_{10}, M_{10} and Y_{10} to improve density specification accuracy, particularly during periods of high solar activity (Bowman et al. 2008).
- **DTM series**: The Drag Temperature Model series, including DTM2013 and DTM2020, uses satellite accelerometer data to derive thermospheric density and temperature variations. DTM2013 showed comparable or slightly superior performance to JB2008 under most solar conditions (He et al. 2018). The latest DTM2020 model incorporated GOCE mission data, which provided continuous density measurements from 260 km down to 180 km during its deorbiting phase, enhancing model calibration at lower thermospheric altitudes (Bruinsma and Boniface 2021).

Recent studies have shown that DTM2013 and JB2008 outperform legacy models such as NRLMSISE-00 during both solar minimum and maximum conditions, with only minor differences observed between DTM2013 and JB2008 themselves (He et al. 2018). However, the introduction of DTM2020 further improved model accuracy, particularly by replacing the $F_{10.7}$ solar flux proxy with the F30 index and implementing an hourly Ap geomagnetic index along with using GOCE data at very low altitudes, leading to improved modelling of short-term density variations and of lower thermosphere density.

Despite these improvements, empirical models inherently rely on historical datasets and thus have limited capability to predict density variations under unseen geomagnetic or solar conditions. This motivates the development of physics-based, data assimilation and hybrid approaches, which are discussed in the following section.

2.3.2. Physics-based Models

Physics-based atmospheric density models simulate the thermosphere by solving the fundamental conservation equations of mass, momentum and energy as explained by He et al. (2018). Unlike empirical or semi-empirical models, which rely primarily on statistical fits to observational data, physics-based models derive the atmospheric state from the physical relationships. This allows them to represent the thermospheric response to changing solar and geomagnetic influences in a more physically consistent way.

The most widely used physics-based model for thermospheric density estimation is the Thermosphere-Ionosphere-Electrodynamics General Circulation Model (TIE-GCM), developed by the National Center for Atmospheric Research (NCAR). TIE-GCM is a three-dimensional, time-dependent model of the coupled thermosphere-ionosphere system, covering a global domain from approximately 97 km up to around 600 km altitude, depending on solar activity (Qian et al. 2014). Chiu and Masha (2009) explain that TIE-GCM uses external drivers such as $F_{10.7}$ and geomagnetic activity parameters to simulate temperature, neutral composition, winds and electrodynamic quantities.

A major strength of TIE-GCM is that it explicitly represents the coupling between solar forcing, geomagnetic activity and thermospheric state changes (Qian et al. 2014). For this reason, physics-based models are better suited than empirical models to representing complex spatial structures and time-dependent storm responses, although their practical performance still depends on model assumptions, external drivers and observational constraints (He et al. 2018).

Despite its comprehensive physical representation, TIE-GCM has limitations in operational contexts. Its high computational cost and dependence on accurate real-time geophysical drivers have restricted its use to research applications or retrospective studies. Additionally, biases in input proxies and missing lower atmospheric coupling (e.g. planetary waves, tides) can limit forecast accuracy without observational corrections (Chiu and Masha 2009).

Other physics-based general circulation models (GCMs), such as the global ionosphere-thermosphere model (GITM) (Ridley, Deng, and Tóth 2006) and coupled thermosphere ionosphere plasmasphere electrodynamics model (CTIPe) (Codrescu et al. 2012), exist and are used for similar research purposes. However, TIE-GCM remains the most mature and widely used model in the space weather and thermosphere modelling community.

Due to these limitations, recent developments have focused on combining physics-based models with data assimilation techniques to produce near real-time density specifications with improved accuracy, as discussed in the next section.

2.3.3. Data-Assimilative Approaches

Data assimilation combines observational data with numerical models to produce a more accurate representation of the atmospheric density than either could provide independently. In thermospheric density modelling, data assimilation techniques adjust model outputs based on satellite observations to reduce biases and improve real-time accuracy, particularly during geomagnetically disturbed conditions.

Although empirical and semi-empirical models are also derived from observations, they differ fundamentally from data-assimilative approaches. Empirical models are calibrated offline using historical datasets, whereas data assimilation updates a model or density estimate during operation using newly available measurements. As a result, data-assimilative methods are able to improve responsiveness to changing thermospheric conditions beyond what can be achieved with historical statistical relationships alone.

A well-known example is the High Accuracy Satellite Drag Model (HASDM), the operational density model used by the United States Air Force Space Command for orbit determination and conjunction assessment. HASDM assimilates radar-tracking observations from approximately 70-90 calibration satellites to update a global density field every three hours (Storz, Bowman, and Branson 2002). While this provides an accurate global density field of the thermosphere, HASDM has limited ability to capture realistic spatial structures or to forecast density ahead of geomagnetic storms. Furthermore, the real-time density outputs generated by HASDM are not publicly available, limiting its applicability for operational use without direct access granted by the United States Air Force.

A more physics-based data-assimilative approach is the Iterative Driver Estimation and Assimilation (IDEA) framework as explained by Sutton, Thayer, et al. (2021). IDEA is a data assimilation technique developed to combine physics-based models such as TIE-GCM with satellite-derived density observations. It employs an iterative least-squares filter similar to an unscented Kalman filter (UKF), adjusting external driving parameters (e.g. solar flux, geomagnetic indices) to minimise differences between model outputs and observations. IDEA accounts for the lagged thermospheric response to changes in external drivers, enabling more accurate nowcasting and short-term forecasting of neutral densities, even during geomagnetic storms. However, IDEA remains computationally expensive and has not yet been implemented operationally.

More recent work has explored the use of alternative observational sources for assimilation. Mutschler et al. (2019) propose a method, called SoleiTool, that uses routine debris tracking data to infer drag accelerations and estimate thermospheric density through a two-stage filtering framework. In a follow-up study, Mutschler (2022) extends this concept using GPS precise orbit determination data from CubeSats to estimate density scaling factors and infer geomagnetic driving conditions for TIE-GCM-based density specification. Together, these studies suggest that large constellations of tracked objects may provide a scalable observational basis for improving real-time thermospheric density estimation and that observation-driven density correction can provide practical orbit prediction benefits.

Despite their strong potential, data-assimilative approaches encounter several practical challenges. They often rely on computationally expensive underlying models, require continuous access to high-quality measurements and must deal with sparse or uneven observational coverage. Consequently, although data assimilation represents an important step toward more accurate and responsive thermospheric density modelling, its integration into operational workflows remains challenging.

2.3.4. Machine Learning and Hybrid Approaches

In recent years, ML and neural network-based methods have become increasingly common in thermospheric density research. They have been used for several different purposes, including the study of long-term density variability, direct density prediction, the calibration of empirical models and the development of more operationally oriented frameworks that aim to remain reliable under changing space weather conditions.

One application of ML in this field is the study of long-term thermospheric density variability. For example, Weng et al. (2020) used an artificial neural network to derive long-term trends in thermospheric density and showed that the ML approach could better capture density variations than earlier empirical approaches, particularly during the unusually deep solar minimum of 2007-2009. This demonstrates that ML methods can be useful not only for short-term forecasting but also for describing variability and trend behaviour in the thermosphere.

A second line of research uses ML for direct density prediction from solar, geomagnetic and orbital inputs. Wang, Chen, et al. (2022) developed a long short-term memory (LSTM) ensemble learning model for storm-time thermospheric density prediction and validated the model's generalisation capability by training on GOCE data and testing on independent Swarm C data. Their results showed that the model remained robust under different geomagnetic activity levels and outperformed NRLMSISE-00 during storm periods. Similarly, Acciarini et al. (2024) trained black-box ML models on thermospheric density data obtained using precise orbit determination (POD) from multiple satellites, including CHAMP, GOCE, GRACE and Swarm A/B and benchmarked them directly against NRLMSISE-00 and JB2008 using the same solar and geomagnetic inputs. They reported that the ML models reduced the mean absolute percentage error (MAPE) from roughly 40-60% to about 20%, corresponding to improvements of 61% over NRLMSISE-00 and 39% over JB2008. These direct density prediction studies show that ML models can outperform standard empirical models, but they do not represent the only way in which ML has been applied to thermospheric density modelling.

Another approach uses ML not as a full replacement for existing models, but as a means of calibrating or enhancing empirical density estimates. Pérez et al. (2015) propose a neural-network-based calibration framework that takes density estimates from DTM-2013, NRLMSISE-00 and JB2008 as inputs and uses accelerometer-derived density from CHAMP and GRACE as target data. Their results showed that the neural network could reduce the density estimation error relative to the original empirical models. More recently, Guo et al. (2025) proposed a two-step enhancement method for an empirical thermospheric density model, in which a spherical harmonic correction and ML-based enhancement coefficients are combined. Their study reported consistent performance improvements across different satellites, altitudes and space weather conditions. These works demonstrate that hybrid approaches, in which ML is used to enhance an existing empirical model rather than replace it entirely, have become an important direction in the literature.

ML has also begun to be used in more operationally oriented density modelling frameworks. Licata et al. (2022) developed an ML thermospheric mass density model based on the HASDM database and incorporated uncertainty quantification into the modelling process. Their results showed that the model could provide robust density predictions together with meaningful uncertainty estimates during diverse space weather conditions. This is particularly interesting from an operational perspective, as it suggests that ML density models may contribute not only to improving density estimates, but also to providing information on the reliability of those estimates.

Overall, the literature shows that ML methods have already been applied in several meaningful ways within thermospheric density modelling. These approaches range from long-term trend estimation and direct density prediction to empirical-model calibration and uncertainty-aware operational modelling. Collectively, these studies show that ML provides a framework for improving thermospheric density estimation across different modelling objectives, ranging from long-term trend analysis to direct prediction and the enhancement of empirical models.

2.3.5. Comparison and Operational Trade-Offs

The reviewed approaches address the thermospheric density estimation problem from a different perspective and exhibit different strengths and weaknesses. Empirical and semi-empirical models remain attractive for operational use because they are computationally efficient, robust and easy to evaluate at arbitrary points along an orbit. However, their dependence on historical statistical relationships and their limited access to solar and geomagnetic inputs limit their ability to capture rapid density fluctuations, particularly during geomagnetic disturbances. As a result, their simplicity and speed are often achieved at the cost of reduced responsiveness.

Physics-based models provide a more explicit representation of the coupled thermosphere-ionosphere system by solving the governing conservation equations. This gives them a stronger physical basis and makes them better suited to representing complex spatial structure and storm-time dynamics. Their operational use, however, is limited by their computational cost and their dependence on accurate external drivers that are not always available in real time. In addition, modelling assumptions can still introduce systematic errors. Data-assimilative approaches address part of this issue by incorporating newly available observations to improve the model output or external inputs. These methods can significantly improve density estimation accuracy, but they also require continuous high-quality measurements and often remain computationally demanding.

Machine-learning-based approaches introduce a different type of trade-off. Their main strength lies in their flexibility as they can learn complex non-linear relationships directly from data and have been shown to improve density estimates in several settings, including long-term variability analysis, direct density prediction and the enhancement of empirical models. At the same time, their performance depends strongly on the quality of the training data. Additionally, it must be ensured that the model is able to generalise outside of the training window. Recent work has therefore focused on hybrid and operationally oriented ML frameworks that aim to combine improved predictive skill with more robustness and, in some cases, uncertainty quantification.

Overall, no single model class fully satisfies all requirements for accurate, responsive and operationally practical thermospheric density estimation. Empirical models offer speed and simplicity, physics-based models offer physical interpretability, data-assimilative approaches offer improved responsiveness to current conditions over the physics-based models and ML-based methods offer flexibility and the ability to learn corrections from data. The literature therefore suggests that the main challenge is not simply to select one class over another, but to find a modelling strategy that balances accuracy, responsiveness and operational feasibility.

2.4. Drag Coefficient Estimation as an Additional Drag Uncertainty Source

In addition to atmospheric density uncertainty, the drag coefficient (C_d) represents an important source of uncertainty in atmospheric drag modelling. The drag coefficient is a dimensionless parameter that characterises how a spacecraft's shape, surface properties and orientation relative to the incoming flow influence the aerodynamic drag it experiences. Historically, a constant value of $C_d = 2.2$ has often been assumed in orbit propagation. This approximation used to be considered acceptable because uncertainties in atmospheric density models dominated the overall drag force error, making high-fidelity C_d estimation less impactful compared to improving the density estimation (Mostaza Prieto, Graziano, and Roberts 2014).

As shown in Equation (2.1), the drag force depends directly on both the drag coefficient C_d and the reference cross-sectional area S. In practice, these quantities are often combined into the ballistic coefficient, introduced in Equation (2.2), which combines mass, drag coefficient and reference area into a single parameter representing an object's sensitivity to atmospheric drag. As atmospheric density modelling has improved and satellite operators have strived towards increasingly precise orbit prediction, the accurate estimation of C_d and the ballistic coefficient has become more important. As a result, a range of techniques have been developed to estimate drag coefficients for objects in low Earth orbit.

One approach to estimating the ballistic coefficient is to use tracking data. In this case, the ballistic coefficient is treated as a tuneable parameter during the orbit determination and orbit propagation fitting procedure. By fitting the modelled trajectory to observations, an effective ballistic coefficient can be estimated as shown by Ergene (2024). If the spacecraft mass and reference area are known, this estimate can then be transformed into the corresponding drag coefficient using a modified version of Equation (2.2). Although this method is operationally useful, the estimated parameter might also include remaining errors from the atmospheric density model.

Another approach involves using analytical, semi-empirical or geometric approximations. Analytical approaches based on free molecular flow theory provide rapid estimates of C_d for idealised shapes such as spheres, flat plates and cylinders. These models typically assume simplified gas and surface interaction behaviour, such as fully diffuse or fully specular reflection. They may be complemented using semi-empirical corrections to account for accommodation effects or surface temperature differences (Moe and Moe 2005). Panel methods provide a more geometrically realistic alternative by discretising the object into flat surfaces with assigned material and reflection properties. Thereafter, the overall drag coefficient is obtained by summing the contribution of each panel relative to the flow direction (Doornbos 2011). These approaches offer a useful balance between computational efficiency and geometric realism, but their fidelity remains limited for highly complex spacecraft shapes and surface interactions.

Higher-fidelity physical modelling approaches aim to simulate the underlying gas-surface interactions more directly. Test Particle Monte Carlo (TPMC) methods statistically model the interaction of atmospheric particles with the object's surface, accounting for effects such as accommodation, diffuse and specular reflection and thermal re-emission. Direct Simulation Monte Carlo (DSMC) extends this framework by additionally modelling molecular collisions, allowing the representation of non-equilibrium gas dynamics around complex geometries (Mostaza Prieto, Graziano, and Roberts 2014). These methods can provide significantly more accurate drag coefficients than analytical approximations, but they are also computationally expensive and therefore not always practical for operational orbit determination.

In some cases, high-fidelity C_d estimation can be improved further when satellite operators provide detailed CAD models, material data and spacecraft attitude data. Such information enables more realistic aerodynamic modelling under varying attitudes and environmental conditions. However, these data are often proprietary and are typically unavailable for debris objects and many operational satellites. Experimental approaches, such as rarefied-flow or hypersonic wind tunnel testing, can also be used to validate drag models or determine drag coefficients for components and simplified spacecraft geometries. In practice, however, accurately reproducing the free molecular flow regime encountered in low Earth orbit remains challenging.

More recently, machine learning and surrogate modelling approaches have been proposed to approximate high-fidelity aerodynamic computations at much lower computational cost. In these approaches, neural networks or response-surface models are trained on precomputed TPMC or DSMC results to rapidly estimate C_d across different flow conditions and spacecraft orientations. An example is the Satellite Drag Coefficient Response Surface Modelling toolkit, which uses precomputed TPMC outputs to provide nearly instantaneous drag coefficient estimates for orbit determination and re-entry analysis (Sheridan et al. 2022). These methods aim to combine the efficiency required for operational applications with the accuracy of physics-based aerodynamic modelling.

Despite the wide range of available techniques, accurate drag coefficient estimation remains challenging because aerodynamic uncertainty is strongly coupled to uncertainty in the atmospheric density itself. In practice, errors in density and errors in C_d can compensate each other, making them difficult to retrieve from orbit data alone. This motivates more research into combined estimation strategies in which density and drag-related parameters are solved jointly. Nevertheless, the present thesis focuses specifically on improving the density side of the drag problem, while acknowledging that the total drag uncertainty also depends on the treatment of C_d.

2.5. Hybrid Neural Network Correction Approach Motivation

Among the different sources of drag uncertainty, atmospheric density is the most natural primary target for improvement in the present work. As discussed in the previous sections, density exhibits strong spatial and temporal variability and remains one of the most difficult quantities to model accurately under changing solar and geomagnetic conditions. It is also the drag-related quantity for which the broadest range of modelling approaches already exists, including empirical, physics-based, data-assimilative and ML-based methods. This makes density modelling an interesting focus for improvement. In addition, because atmospheric density directly affects drag acceleration along the full orbit, even small improvements in density estimation can translate into meaningful improvements in orbit prediction.

A second motivation follows from the operational trade-offs identified in the literature. Empirical and semi-empirical models such as JB2008 remain attractive because they are fast, robust and straightforward to evaluate in an operational orbit propagation workflow. However, their reliance on historical statistical relationships and limited solar and geomagnetic input limits their responsiveness during rapidly changing conditions. Physics-based and data-assimilative approaches can provide improved physical realism and responsiveness, but they are generally more computationally demanding. This suggests that a useful intermediate strategy is not necessarily to replace an empirical model entirely, but rather to keep its speed and baseline physical structure while learning a correction to reduce the remaining error.

From this perspective, a neural-network correction of an empirical baseline provides several advantages. First, the baseline model already provides a physically meaningful first-order density estimate, so the learning objective is reduced from reconstructing the full thermospheric density model to modelling the residual error that remains after the empirical estimation. Second, the corrected model inherits the computational efficiency and broad applicability of the underlying empirical model, which is valuable for operational use. Third, the non-linear learning capabilities of neural networks allow the model to capture the highly non-linear and complex atmospheric responses during extreme space weather events. Finally, such a hybrid formulation reduces the risk of complete black-box behaviour, since the neural network acts on top of an established density model rather than replacing it.

Although the drag coefficient is also an important contributor to drag uncertainty, it is not modelled explicitly in this thesis. This choice was made with full awareness that density and drag-coefficient related errors are intertwined in practice and cannot be perfectly separated in this analysis. In many cases, errors in atmospheric density and errors in C_d can partially compensate for one another, making a strict attribution of the drag model error difficult. Nevertheless, the primary focus of this thesis is placed on improving the density component, since atmospheric density remains a dominant and dynamically variable source of drag uncertainty under changing solar and geomagnetic influences. In addition, density correction offers a more broadly applicable modelling use across multiple satellites, whereas the estimation of the C_d often requires detailed information on spacecraft geometry, surface properties and attitude that is not consistently available for most satellites. The present work therefore does not attempt to isolate density from C_d completely, but instead investigates whether improving the density side of the drag problem can still provide a meaningful benefit for orbit determination and propagation.

The literature reviewed in this chapter therefore motivates a hybrid strategy in which the speed and operational convenience of an empirical model are retained, while a neural network is used to learn a correction, based on solar, spatial and temporal inputs, to the original empirical density estimate. The following sections build on this motivation by defining the research goals and research questions of this thesis.

2.6. Research Motivation and Goals

The literature reviewed in this chapter shows that accurate thermospheric density modelling remains difficult, particularly when both operational practicality and responsiveness to changing space weather conditions are required. Building on the hybrid motivation outlined in the previous section, this thesis investigates whether a neural-network-based correction strategy can improve density estimation while remaining suitable for orbit determination and propagation applications.

Within this global problem, this thesis focuses specifically on improving the atmospheric density component of the drag model. The first goal is to investigate whether a neural-network-based correction, when applied to an empirical density model, can reduce the error relative to accelerometer-derived thermospheric density observations across multiple satellites and varying space weather conditions. In this context, the accelerometer-derived density is treated as the high-fidelity reference against which the model performance is measured. Rather than replacing the empirical model entirely, the thesis uses a hybrid formulation in which the empirical model provides the baseline density estimate and the neural network learns a correction using solar, geomagnetic, temporal and spatial inputs.

A second goal of the thesis is to determine whether this correction method translates into practical orbit prediction improvements over the empirical model. This is a necessary step, since better agreement with accelerometer-derived density targets does not automatically imply better orbit propagation performance. For this reason, the thesis adopts a two-stage evaluation strategy. First, by assessing whether the correction improves density reconstruction and second, by evaluating whether the corrected model reduces orbit propagation error in an independent validation setting.

Based on the refined thesis scope and the research gaps identified in this literature study, the following main research question is formulated:

> **Research Question 1 (RQ1)**
> *Can a neural network-based correction to a baseline empirical atmospheric density model, using the baseline density together with solar, geomagnetic, temporal and spatial inputs, provide more accurate reconstructions of accelerometer-derived thermospheric density than the baseline alone across multiple satellites and varying space weather conditions?*

Hypothesis 1 (H1): A neural network trained on accelerometer-derived density residuals relative to an empirical model and informed by solar, geomagnetic and orbital context, can learn a meaningful correction that reduces density errors and better captures thermospheric variability than the empirical model alone.

Because improved density reconstruction does not automatically imply improved orbit prediction accuracy, a second research question is posed:

> **Research Question 2 (RQ2)**
> *Does the neural-network-corrected density model reduce 24-hour along-track orbit propagation error relative to the baseline when evaluated against POD truth data over multiple validation arcs spanning quiet and geomagnetically disturbed conditions?*

Hypothesis 2 (H2): The neural-network-corrected density model will yield lower 24-hour along-track root mean square propagation errors than the baseline when evaluated against Swarm A POD truth data, with the largest improvements occurring during moderate and severe geomagnetic disturbances.

2.7. Conclusion

This chapter has shown that atmospheric drag modelling in low Earth orbit is affected by multiple coupled sources of uncertainty, of which thermospheric density remains the most difficult to represent accurately. The review has highlighted that thermospheric density is controlled by both solar radiative forcing and geomagnetic driving and that its response is delayed, history-dependent and strongly variable across space and time. As a result, accurate drag modelling requires not only a suitable density model, but also a meaningful representation of the solar and geomagnetic drivers that shape the thermosphere.

The literature further shows that no single modelling approach fully satisfies all operational requirements. Empirical models remain attractive because of their speed and robustness, but their responsiveness is limited during disturbed conditions. Physics-based and data-assimilative models provide greater physical realism and improved responsiveness, but at increased computational and operational cost. Machine learning-based approaches, particularly hybrid correction frameworks, offer a promising intermediate strategy by combining the efficiency of empirical models with the flexibility of data-driven correction.

On this basis, this thesis adopts a neural network correction approach focused on improving the density component of the drag problem while retaining the operational advantages of empirical models. This motivation leads directly to the two central research questions of the thesis: first, whether such a correction can improve accelerometer-derived density reconstruction and second, whether these improvements translate into more accurate orbit propagation. The next chapter therefore turns to the data processing strategy and the construction of the solar, geomagnetic, spatial and temporal inputs used to address these questions.

3

Data Processing

This chapter describes how the raw satellite, solar and geomagnetic data are transformed into the final dataset used for training and validation of the neural network. It first introduces the underlying data sources and their temporal and orbital coverage in Section 3.1. It then explains the preprocessing pipeline used to align the data streams, handle periodic variables and preserve operational realism in Section 3.2, followed by the construction of the final solar, geomagnetic, temporal and trust-based inputs in Section 3.3. The feature scaling and distribution-aware normalisation strategy are presented in Section 3.4, after which the target variable formulation and train-validation split strategy are defined in Sections 3.5 and 3.6. Together, these steps provide the final dataset on which the methodology of Chapter 4 is built.

3.1. Data Acquisition and Sources

This thesis utilises a dataset spanning January 2000 to December 2024, capturing the latter half of Solar Cycle 23, the entirety of Cycle 24 and the ascending phase of Cycle 25. This temporal range ensures the model is exposed to a wide range of solar activity, from extreme geomagnetic storms to prolonged solar minima. This dataset is composed by collecting data from three main domains: satellite-based accelerometer data, solar irradiance indices and geomagnetic space weather parameters.

3.1.1. Accelerometer-Derived Density Data

The accelerometer-derived density data are obtained from the TU Delft thermosphere repository (Siemes, De Teixeira Da Encarnação, et al. 2016; Siemes, Borries, et al. 2023), providing high-fidelity thermospheric density observations. The accelerometer-derived density data is available for the CHAMP, GOCE, GRACE, Swarm and GRACE-FO missions. The primary features extracted from these files include the timestamp (UTC/GPS), the satellite's position (GRS80) in the form of geodetic altitude, longitude and latitude and the local solar time. Additionally, the argument of latitude is included to provide the model with orbital phase context. The accelerometer-derived density (often denoted as 'dens_x' in the source files) serves as the basis for the target variable. To ensure the integrity of the data, the quality flags, when provided, are used to filter out anomalous data. These flags indicate whether a data point is anomalous.

Table 3.1 provides an overview of the satellite features, including the respective units and a short description of each parameter.

Table 3.1: Satellite Feature Summary

Parameter	Units	Description
Timestamp	YYYY-MM-DD HH:MM:SS	Standardised to UTC across all missions.
Altitude	meters (m)	Geodetic altitude above the GRS80 ellipsoid.
Latitude	degrees (°)	Geodetic latitude ($-90°$ to $+90°$).
Longitude	degrees (°)	Geodetic longitude ($0°$ to $360°$).
Local Solar Time	hours (h)	Position of the satellite relative to the Sun (0 to 24).
Argument of Latitude	degrees (°)	Angle along the orbit between the satellite and the ascending node.
Measured Density	kg/m³	High-cadence density derived from accelerometer measurements or GPS accelerations.
Data Flags	Boolean/Int	Status indicators used to filter non-nominal or anomalous measurements.

To provide insight into these measurements, it is useful to visualise the orbital regimes occupied by the satellites in relation to the solar regime. Figure 3.1 presents the altitude evolution of the selected missions alongside the solar cycle progression. This visualisation not only highlights the spatial and temporal coverage of the dataset, but also illustrates the correlation between high solar activity (indicated by the Sunspot Number) and increased atmospheric drag, which accelerates orbital decay.

Complementing the visual overview, Table 3.2 provides a quantitative overview of the dataset. The table details the precise operational windows and mean geodetic altitudes for each satellite, alongside the total volume of raw samples collected. Additionally, the table includes a ”Missing Data” metric quantifying the extent of instrument outages and quality-flagged exclusions. Note that while the GRACE-FO D satellite is listed for completeness, it is explicitly marked as excluded. The specific instrumental degradation warranting this exclusion is detailed in the subsequent discussion on mission-specific limitations.

Table 3.2: Satellite mission data availability and mean orbital regimes

Mission	Sat.	Start Date	End Date	Mean Alt. (GRS80, km)	Raw Data Count (10^6)	Missing Data (%)
CHAMP	-	2000-07-29	2010-09-04	378	29.4	7.7
GOCE	-	2009-11-01	2013-10-20	266	11.0	12.1
GRACE	A	2002-04-04	2017-10-31	460	43.2	12.2
	B	2002-04-04	2017-05-22	466	39.5	17.4
Swarm	A	2014-01-01	2025-08-31	464	12.2	0.9
	B	2014-01-01	2025-08-31	519	12.2	0.6
	C	2014-01-01	2025-08-31	464	12.2	0.7
GRACE-FO	C	2018-05-29	2025-04-30	502	21.7	0.7
	D	*Excluded due to accelerometer degradation (see text)*				

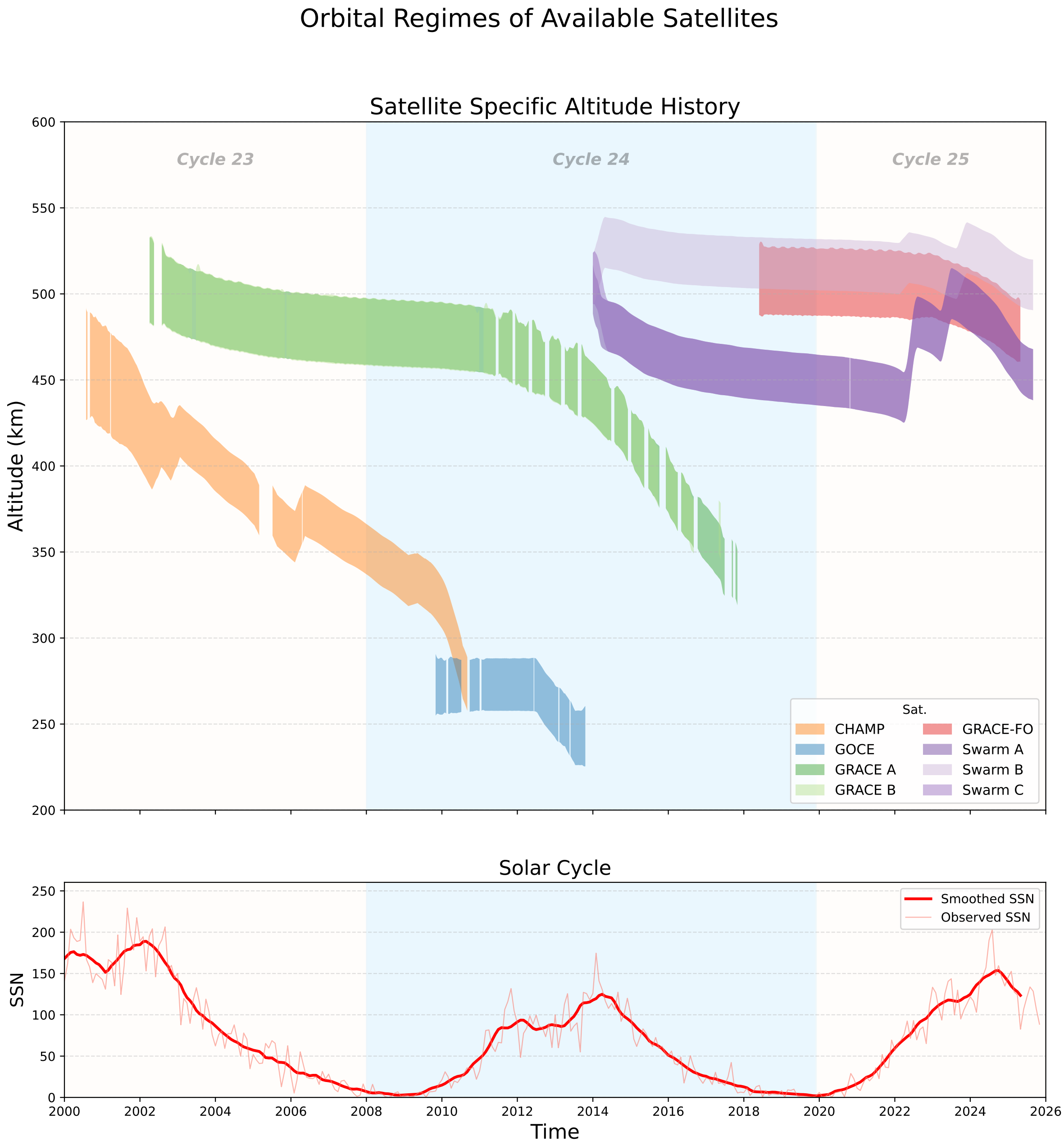


Figure 3.1: Orbital regimes of relevant satellites including solar activity. The top panel shows the altitude history of each satellite considered in this study. The bottom panel shows the smoothed and raw Sunspot Number (SSN), illustrating the correlation between high solar activity and increased orbital decay rate. For visual clarity, the altitude time series in this figure are daily-binned (one value per day), so it may visually differ slightly from the higher-cadence representation in Figure 3.2.

As indicated by the missing data percentages in Table 3.2, the dataset is not continuous. Significant gaps exist due to instrument outages and orbital manoeuvres. Figure 3.2 presents a timeline of this availability for each satellite, explicitly marking outages that exceed one hour. These discontinuities are a defining characteristic of the dataset and require a modelling approach that treats time-steps independently, rather than relying on continuous sequence models that would be sensitive to interruptions.

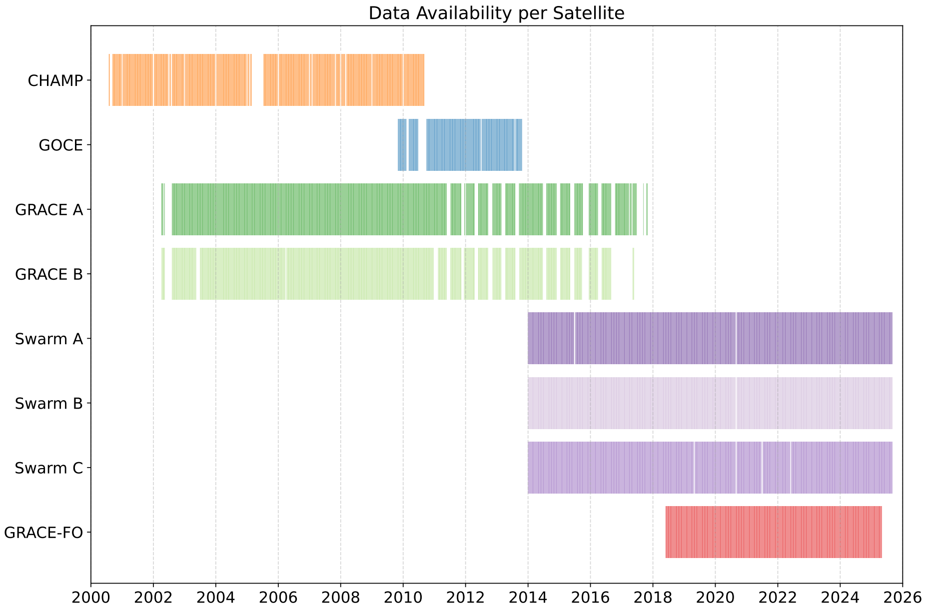


Figure 3.2: Temporal availability of accelerometer data across all missions. The coloured bars indicate periods of nominal data availability, while white gaps represent instrument outages or excluded data segments greater than 1 hour.

Although accelerometer-derived densities are widely regarded as the most reliable in-situ measurements of thermospheric mass density, they are not free from limitations. The conversion from measured non-gravitational acceleration to density requires assumptions regarding the satellite's drag coefficient, cross-sectional area and instrument calibration that can introduce systematic offsets between missions (Doornbos 2011). Consequently, while accelerometer-derived density data provides the best available proxy for true density, it must be treated as a high-fidelity estimate rather than absolute ground truth.

Given these limitations, not all missions are included in the neural network training, validation and test dataset. However, they remain part of the general data inventory described in this chapter to illustrate the full extent of the available (albeit degraded) thermospheric measurements. Previous studies have reported calibration challenges and increased uncertainty in the Swarm and GRACE-FO accelerometer datasets. For the Swarm mission, Siemes, De Teixeira Da Encarnação, et al. (2016) identified extreme thermal sensitivity, up to 100 times higher than anticipated, along with frequent sudden bias steps that required complex, manual correction strategies. Similarly, regarding GRACE-FO, Behzadpour, Mayer-Gürr, and Krauss (2021) report that accelerometer measurements from both satellites are contaminated by instrumental noise. Additionally, the systematic degradation of the GRACE-FO D accelerometer shortly after launch was documented. This accelerometer failure required the substitution of observations with synthetic data derived from GRACE-FO C. Given these issues related to Swarm and GRACE-FO accelerometer-derived data and to ensure a consistent and reliable target variable during training, validation and testing, these missions are excluded from the dataset. However, Swarm-A Precise Orbit Determination (POD) data has been retained for independent validation purposes. The POD-based validation strategy, described in Chapter 4, enables an assessment of the model's performance during orbit propagation without relying directly on potentially erroneous accelerometer-derived densities.

This exclusion of Swarm and GRACE-FO has implications for the temporal and orbital coverage of the training dataset. By omitting these missions, the effective dataset period is limited to missions operating before 2018. Consequently, the model is trained predominantly on data spanning Solar Cycles 23 and 24, without direct exposure to the early phase of Solar Cycle 25. In addition, the altitude regime represented in the dataset is constrained by the orbital characteristics of the CHAMP, GRACE and GOCE missions, thereby limiting the altitude range available for density correction learning to $240 - 530$ km.

3.1.2. Solar and Geomagnetic Drivers

To complement the satellite measurements, solar and geomagnetic activity data are integrated from multiple space weather services. An overview of these features, including their respective categories, data sources and temporal resolutions, is provided in Table 3.3. These parameters characterise the solar radiative forcing, upstream solar wind and geomagnetic response that together influence thermospheric density. This includes an extensive OMNI dataset, geomagnetic indices, Polar Cap (PC) indices and solar irradiance indices. Each parameter is explained in detail in Section 2.2.

High-cadence OMNI data provides a 1-minute resolution view into the solar wind and Interplanetary Magnetic Field (IMF). The OMNI dataset provides a 1-minute resolution record of upstream solar wind and IMF conditions, together with several high-cadence geomagnetic indices.

- **Interplanetary Magnetic Field (IMF):** The field magnitude and vector components (B_x, B_y, B_z) in Geocentric Solar Magnetospheric (GSM) coordinates.
- **Solar Wind Plasma:** flow speed (v_{sw}), proton density (ρ_{sw}), flow pressure (p_{sw}) and temperature (T_{sw})
- **High-Cadence Geomagnetic Indices:** The auroral electrojet indices AE, AL and AU, together with the SYM and ASY indices in their horizontal (H) and declination (D) components, are included to capture rapid high-latitude and storm-time geomagnetic variability at 1-minute cadence as explained in Section 2.2.4. Compared with the hourly D_{st} index, these indices provide finer temporal resolution for resolving the onset and short-timescale evolution of geomagnetic disturbances.

Global and polar geomagnetic indices provide insight into the magnetosphere's overall state. The K_p and a_p indices (3-hourly resolution) represent the global geomagnetic climate. The Disturbance Storm Time (Dst) index is obtained separately to provide an hourly measure of the ring current intensity (Campbell 1996), which is critical for capturing the storm-time response. Furthermore, the 1-minute resolution Polar Cap North (PCN) index provides a localised measurement of geomagnetic activity over the magnetic north pole. While Polar Cap South (PCS) data exists, it is excluded as it is not consistently available in near-real-time.

Solar irradiance indices show the long-term solar heating trends and the Extreme Ultraviolet (EUV) variability. These are obtained from the SOLFSMY.TXT dataset, provided by Space Environment Technologies (SET). This includes the commonly used $F_{10.7}$ index alongside additional indices S_{10}, M_{10}, Y_{10}.

The specific characteristics and sources of these integrated features are detailed in Table 3.3.

Table 3.3: Solar feature summary

Category	**Features / Parameters**	**Source**	**Resolution**
Interplanetary	B_x, B_y, B_z (GSM, nT), v_{sw} (km/s), ρ_{sw} (n/cc), p_{sw} (nPa), T_{sw} (K)	NASA OMNI [1]	1 minute
Auroral	AE, AL, AU (nT)	NASA OMNI [1]	1 minute
Ring Current	$SYM/D, SYM/H, ASY/D, ASY/H$ (nT)	NASA OMNI [1]	1 minute
Storm Context	D_{st} (nT)	WDC Kyoto [2]	1 hour
Geomagnetic	K_p, a_p (-)	GFZ Potsdam [3]	3 hours
Polar Activity	PCN, PCS* (mV/m)	ISGI [4]	1 minute
Solar Flux	$F_{10.7}$, S_{10}, M_{10}, Y_{10} (sfu)	SET (SOLFSMY) [5]	1 day

*Note: PCS is retrieved for completeness but excluded from the final set of features due to real-time availability concerns.

---

[1] Accessible at `https://cdaweb.gsfc.nasa.gov/cgi-bin/eval1.cgi`

[2] Accessible at `https://wdc.kugi.kyoto-u.ac.jp/wdc/Sec3.html`

[3] Accessible at `https://kp.gfz.de/en/data`

[4] Accessible at `https://isgi.unistra.fr/data_download.php`

[5] Accessible at `https://spacewx.com/jb2008/`

3.2. Preprocessing and Temporal Alignment

The features described in Section 3.1 exhibit a wide range of native temporal resolutions, from 10-second accelerometer observations to daily solar flux updates. To maintain high information density while avoiding redundant data, a preprocessing pipeline is implemented to synchronise these data streams into a uniform 1-minute temporal grid. This resolution serves as the middle ground, aligning with the high-cadence OMNI drivers while effectively downsampling the satellite measurements to reduce high-frequency noise.

3.2.1. Temporal Synchronisation Strategy

The synchronisation of various data cadences required different strategies for the satellite measurements and the solar drivers. For the high-cadence satellite data (10 or 30s intervals), both linear interpolation and 1-minute averaging are considered. Since high-frequency accelerometer data is often subject to noise, linear interpolation would preserve these high-frequency noise components, propagating these artefacts into the model's target variable. As a solution, 1-minute averaging is selected as it suppresses high-frequency noise and provides a smoothed density value that is more representative of the mean atmospheric state within the 60-second interval. Additionally, given that the raw dataset exceeds 100 million measurements, linear interpolation is deemed suboptimal due to the additional computational burden without increasing the physical information content.

In addition to the accelerometer-derived density, the JB2008 model density is evaluated at the native satellite measurement cadence (10 or 30s) using the corresponding satellite position and solar-geomagnetic inputs required. This model density is subsequently subjected to the same 1-minute averaging procedure applied to the measured density. This approach ensures temporal consistency between the model baseline and the observational target. By applying identical temporal handling to both variables, the comparison between JB2008 and the accelerometer-derived density reflects differences in physical modelling rather than discrepancies introduced by using different sampling strategies.

Regarding the lower-cadence solar features, such as the 3-hourly K_p or hourly D_{st}, a forward-filling (FFILL) solution is adopted. In detail, this strategy sets the value of an index constant for every 1-minute timestamp within the index's native reporting interval (e.g. the K_p value for the 03:00–06:00 UTC window is assigned to every minute between 03:00 and 06:00 UTC). While linear/cubic interpolation might provide a smoother transition between entries, it introduces a critical operational flaw called data leakage. To perform linear interpolation at a point between times t and $t+1$, the value at $t+1$ must be known. In a real-time operational scenario, the next K_p or $F_{10.7}$ is not yet available, making interpolation impossible. Using the FFILL method, it is ensured that at any timestamp t, the model only has access to the most recent observation, thus preventing leakage.

It must be noted that the FFILL method described above assumes that, for example, the K_p value for the 03:00-06:00 UTC window is already known at 03:00 UTC. However, in real-time operations, this value is not determined until after 06:00 UTC. This is called the operational lag and must be accounted for during the feature engineering process. The method used to handle operational lags, ensuring that the model only utilises data that would be realistically available in real-time operations, is described in Section 3.3.

3.2.2. Periodic Feature Engineering

Certain spatial and temporal features, specifically local solar time (LST), geodetic longitude and the argument of latitude, possess a periodic topology that presents a significant feature engineering challenge. On a standard linear scale (e.g. 0 to 24 hours for LST), the values 23.99 and 0.01 are physically adjacent, yet numerically represent opposite ends of the feature space. A neural network would interpret 0.01 as a near-minimum state and 23.99 as a near-maximum state, creating a large, non-physical step-change at the boundary.

This discontinuity introduces a secondary issue during the 1-minute averaging process performed during the satellite data downsampling. If two data points within a single minute have LST values of 23.99h and 0.01h, simply averaging these would result in 12.00h. This is a physically incorrect state positioned nearly 12 hours away from the actual LST. As a result, it is crucial to resolve this discontinuity before downsampling.

The first approach, explicit rollover handling (or angle unwrapping), involves detecting crossings of the $0/24$ h (or $0/360°$) boundary and shifting one side of the boundary by one full period prior to computing the mean. While this can produce a valid linearised average, it treats the variable as a purely linear scalar and potentially loses the geometric context of the traversal.

The second approach, the one adopted in this thesis, is the mapping of periodic features onto a unit circle using their sine and cosine components. This transformation ensures continuity throughout the entirety of the feature's range and provides insight into the rate at which the periodic variable changes within the averaging interval. When the satellite traverses a large fraction of the unit circle within a one-minute bin (e.g. during a polar-orbit pass rapid LST variations occur), the averaged sine/cosine vector exhibits a reduced norm due to phase cancellation. This norm attenuation may serve as an additional informative signal that could be exploited in future extensions of the proposed method. For a variable x with a known period P, the mapping is defined as:

$$x_{sin} = \sin\left(\frac{2\pi x}{P}\right) \tag{3.1}$$

$$x_{cos} = \cos\left(\frac{2\pi x}{P}\right) \tag{3.2}$$

Where P represents the period (e.g., 24 for LST, 360 for longitude and argument of latitude). By converting the raw parameters to sine and cosine components, the discontinuity is eliminated.

3.2.3. Vector Norm Renormalisation

While transforming periodic features into sine and cosine components resolves the discontinuity, it introduces a geometric side effect. The 1-minute averaging strategy causes the vector norm for periodic features to decrease. Theoretically, any coordinate mapped to the unit circle using the sine and cosine transformations defined in the previous section should maintain a norm of exactly 1. However, because the 1-minute average of points situated along a circular arc results in a coordinate that lies along the chord connecting those points, the resulting mean vectors always fall inside the unit circle. As seen in Figure 3.3, the LST of the downsampled dataset shows distinct concentric rings located at radii significantly less than 1.

This reduction in norm is not uniform across the dataset. It is a direct function of the satellite's rate of change across the periodic variable ($\Delta\text{LST}/\Delta t$). As shown in Figure 3.4, the norm deflation is mission-specific and highly correlated with orbital inclination.

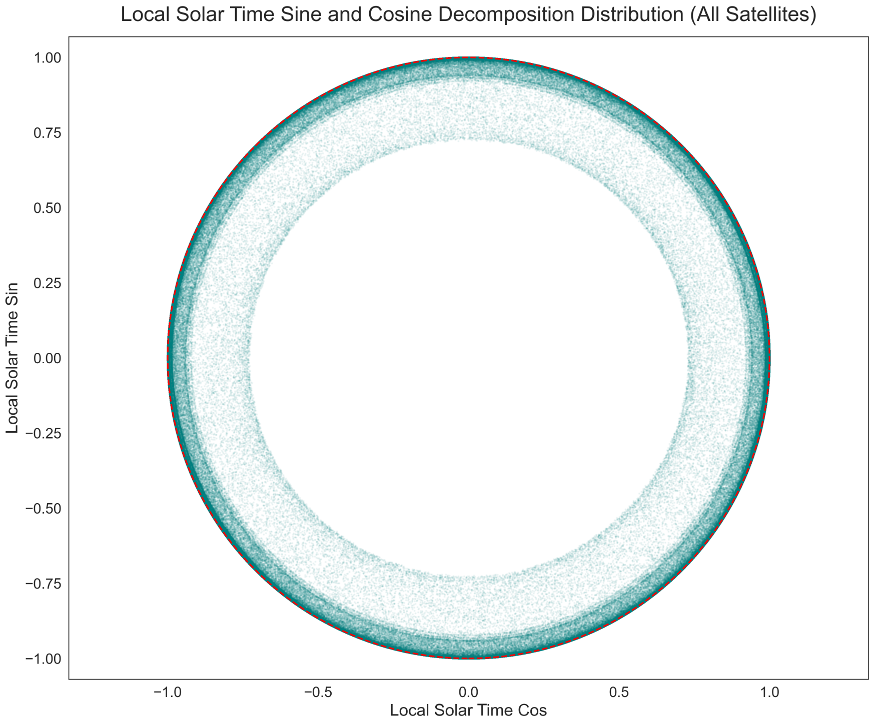


Figure 3.3: Distribution of the Local Solar Time (LST) vector after 1-min averaging. The deviation from the unit circle manifests as concentric rings, indicating signal attenuation caused by averaging over rapid changes in periodic features.

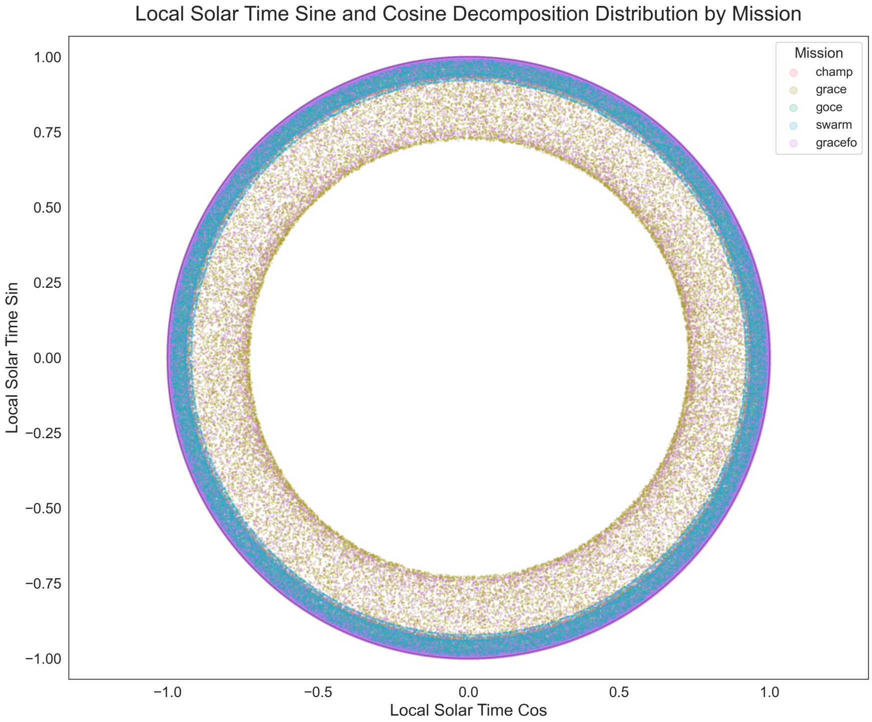


Figure 3.4: LST vector distribution separated by satellite mission. The magnitude of the norm decrease is highly dependent on the mission. Near-polar missions (GRACE, GRACE-FO) exhibit the most severe reduction, while sun-synchronous missions (GOCE) maintain a norm closer to unity.

Missions with an inclination near $90°$, such as GRACE and GRACE-FO ($i \approx 90°$), traverse LST extremely rapidly as they pass over the Earth's poles. When averaging the 10 or 30-second raw samples into a 1-minute sample during these polar crossings, the satellite covers a significant arc length. The resulting downsampled sine and cosine components lie well within the unit circle.

In contrast, missions like GOCE ($i \approx 96.7°$) or those in specific sun-synchronous orbits experience a slower LST change over time. This results in a smaller arc and, consequently, a vector norm significantly closer to the unit circle during a polar crossing. Table 3.4 provides mission-specific insights into the inclination and the related LST vector behaviour.

Table 3.4: Impact of orbital inclination on LST vector norm stability

Mission	Inclination i	Min. Norm	Mean Norm
CHAMP	$87.30°$	0.9230	0.9981
GOCE	$96.70°$	0.9850	0.9991
GRACE	$89.00°$	0.7229	0.9957
Swarm	$87.35° - 87.75°$	0.9201	0.9984
GRACE-FO	$89.00°$	0.7345	0.9959

To investigate the mechanism driving these decreases in norm values for near-polar satellites, an analysis is conducted on a single polar crossing event. As indicated in Table 3.4, the GRACE mission exhibits the most severe attenuation with a minimum norm of 0.7229. A 1-minute window from the GRACE A dataset containing a polar pass (2005-02-27 04:02:00 UTC) is isolated to examine the raw 10-second accelerometer behaviour compared to the downsampled average. Table 3.5 presents the local solar time vectors for this interval, explicitly demonstrating how the averaging of valid unit-norm vectors can result in a significant loss of magnitude.

Table 3.5: Raw and averaged LST vectors during polar crossing example (2005-02-27 04:02:00 UTC). While each individual raw sample maintains a unit norm of 1.0, the rapid LST change (spanning from 5.49h to 13.34h within 60 seconds) results in a significant attenuation of the LST norm.

#	Time (+s)	LST (h)	LST_{sin} (-)	LST_{cos} (-)	Norm LST (-)
1	+02s	5.49	0.9911	0.1329	1.0000
2	+12s	6.42	0.9941	-0.1087	1.0000
3	+22s	8.16	0.8442	-0.5360	1.0000
4	+32s	10.60	0.3574	-0.9340	1.0000
5	+42s	12.39	-0.1022	-0.9948	1.0000
6	+52s	13.34	-0.3439	-0.9390	1.0000
1-Min Mean		**9.40**	**0.4568**	**-0.5633**	**0.7252**

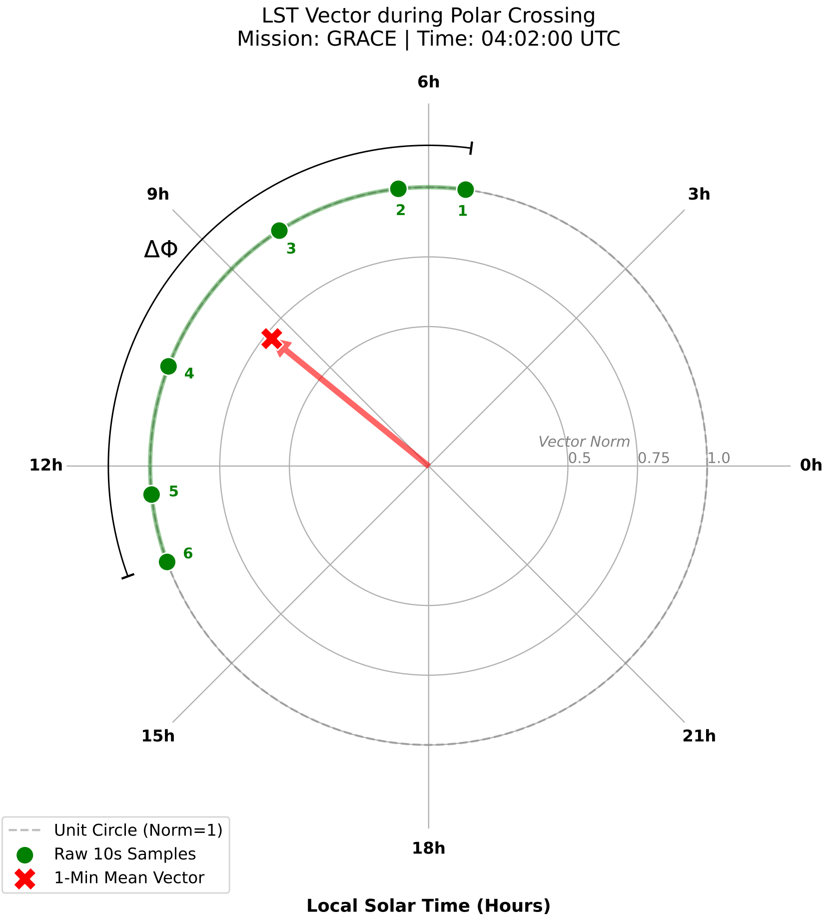


Figure 3.5: Attenuation of the LST vector for a near-polar mission (GRACE, $i \approx 89°$) during 2005-02-27 04:02:00 UTC. The six raw 10-second samples (green, indexed 1–6) span a wide angular range $\Delta\Phi \approx 135°$ due to the rapid traversal of time zones at the pole. Consequently, the 1-minute mean vector (red X) represents the geometric centroid of these samples, resulting in a significantly reduced norm of 0.7252. However, the mean vector's angle remains a valid estimator of the mean LST, justifying the renormalisation step.

The attenuation mechanism is illustrated in Figure 3.5. The six raw observations lie on the unit circle and span an angular range defined as $\Delta\Phi$, representing the total angular displacement in local solar time during the 1-minute interval. For periodic variables encoded as $(\sin, \cos)$ pairs, this angular span directly determines the magnitude of the averaged vector. During near-polar crossings, the satellite experiences a rapid change in local solar time, causing $\Delta\Phi$ to increase substantially. In the example shown, the LST varies from 5.49h to 13.34h within 60 seconds, corresponding to an angular sweep of approximately $\Delta\Phi \approx 135°$. The 1-minute mean vector therefore lies at the geometric centroid of the six raw observations, resulting in a reduced norm of 0.7252

A direct comparison between Figure 3.5 and Figure 3.6 highlights the dependence of attenuation on angular span. In the GRACE example, the large $\Delta\Phi$ during a polar crossing produces a substantial norm reduction. In contrast, the GOCE example exhibits a much smaller angular displacement within the same 1-minute window, resulting in a minimal deviation from the unit circle. This contrast confirms that the observed LST norm deflation is closely tied to the angular displacement in LST during the 1-minute window.

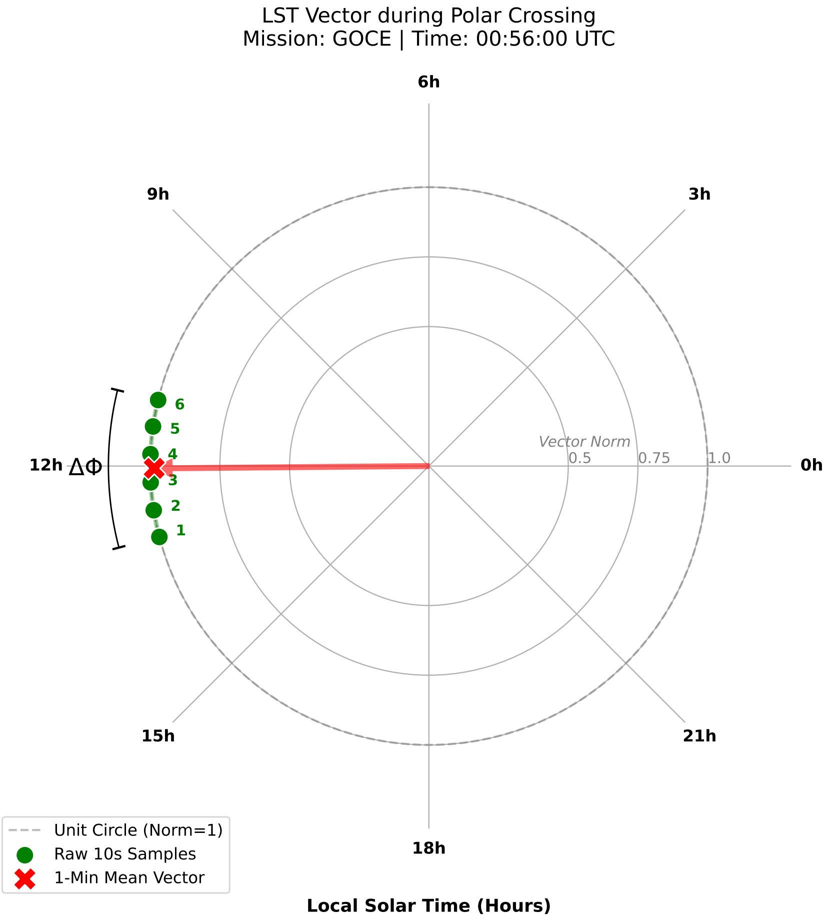


Figure 3.6: Stability of the LST vector for a sun-synchronous mission (GOCE, $i \approx 96.7^{\circ}$) during 2010-01-26 00:56:00 UTC. In contrast to the polar case, the angular range $\Delta\Phi$ covered during a 1-minute window is significantly smaller, causing the raw samples to cluster tightly. The resulting mean vector remains approximately on the unit circle with a norm of > 0.98.

Despite the decrease in magnitude of the averaged vector given an increase of $\Delta\Phi$, the angular direction of the mean vector remains a valid approximation of the mean local solar time over the 1-minute interval. The attenuation affects only the radial component in the embedded $(\sin, \cos)$ space, while the angular orientation, which encodes the physical information, is preserved. Consequently, renormalising the averaged vector to unit length restores the expected geometric structure without altering the underlying angular information.

This renormalisation step is particularly important because the sine and cosine components are subsequently used as neural network input features. Without correction, variations in vector norm would introduce unintended amplitude differences that are unrelated to the physical quantity being represented. A norm of, for example, 0.75 does not represent a weaker or less meaningful local solar time state, but is solely the geometric consequence of temporal averaging. If left uncorrected, this norm variability could introduce an additional, unintended feature encoding the rate of change of the periodic variable rather than its phase. In other words, the magnitude of the $(\sin, \cos)$ pair would begin to reflect orbital geometry and sampling effects instead of solely representing local solar time. By restoring the vector's norm to 1, the representation remains purely angular, ensuring that only physically meaningful phase information is presented to the model. The correction is applied element-wise to the averaged sine and cosine components of variable x as follows:

$$\vec{x}_{final} = \frac{\vec{x}_{mean}}{||\vec{x}_{mean}||} \tag{3.3}$$

This transformation is applied to both local solar time and geodetic longitude, as both variables exhibit significant variations in angular velocity during near-polar crossings. In contrast, the argument of latitude progresses at a nearly constant rate throughout the orbit regardless of inclination or latitude. Therefore, its angular sweep $\Delta\Phi$ remains relatively consistent across all 1-minute averaging windows, resulting in a stable vector norm that does not require renormalisation.

3.3. Feature Engineering and Coupling Functions

This section describes how the raw solar and geomagnetic inputs are transformed into a physically more informative feature set for the neural network. The objective is not merely to increase the number of inputs, but to embed the key processes that influence thermospheric density variability in a form that the model can use effectively. To this end, three main feature engineering steps are applied. First, physically motivated solar wind coupling functions are added to better represent the efficiency of energy transfer from the solar wind into the magnetosphere-ionosphere-thermosphere system. Second, temporal feature embeddings are constructed to capture the delayed and history-dependent response of the thermosphere to external forcing. Third, dedicated trust metrics are introduced to account for outages in several high-cadence space weather inputs while preserving a continuous feature stream. Together, these steps are designed to improve physical interpretability and provide the neural network with an enhanced description of the drivers relevant to thermospheric density variation.

3.3.1. Solar Wind Coupling Functions

According to Newell et al. (2007), the raw solar parameters (e.g. B_z, v_{sw}, ρ) do not linearly map to the energy transferred into the magnetosphere-thermosphere system. The interaction is highly non-linear, governed by the efficiency of magnetic reconnection at the dayside magnetopause. To provide the neural network with additional features that better represent the physical energy input driving thermospheric heating, specific solar wind coupling functions are computed and added to the set of neural network inputs.

In the comprehensive comparative analysis performed by Newell et al. (2007) the performance of 20 candidate coupling functions is evaluated against ten magnetospheric state variables, ranging from global indices (K_p, D_{st}) to localised auroral power (e.g. AE, AU). To avoid redundancy and to keep the set of inputs lean, two coupling functions are selected based on statistical performance and physical interpretability. The first is the Newell coupling function ($d\Phi_{MP}/dt$), which has been identified as the statistically optimal coupling function. The second is the rectified electric field (vB_s), selected for its physical significance as a direct proxy for the convection electric field. This combination ensures that the model benefits from both the most statistically predictive coupling function and the classical physical driver used in semi-empirical density models.

As discussed in section 2.2, solar wind-magnetosphere coupling is strongest when the IMF is oriented southward or antiparallel to the Earth's dayside magnetic field, because this favours magnetic reconnection at the magnetopause. However, the coupling strength does not change in a purely binary way between fully active and inactive conditions. Instead, it varies continuously with both the solar wind strength and the IMF orientation. Coupling functions are designed to summarise this combined upstream driving in a compact form, so that the model receives features that are more directly related to the efficiency of energy transfer into the magnetosphere-ionosphere-thermosphere system than the raw solar wind variables alone.

The Newell Coupling Function $d\Phi_{MP}/dt$

The primary coupling function selected is the Newell coupling function, expressed as the rate of magnetic flux opening at the magnetopause, denoted by $d\Phi_{MP}/dt$. This directly relates to the physical explanation in section 2.2. When dayside magnetic reconnection becomes more efficient, closed geomagnetic field lines are opened more rapidly, allowing solar wind energy and momentum to enter the magnetosphere more effectively. In the comparative study of Newell et al. (2007), the Newell coupling formulation showed the strongest overall statistical relationship with a broad set of magnetospheric activity indicators, including Dst, K_p and AE. Its main advantage is that it does not assume that coupling depends only on whether the IMF is strictly southward. Instead, it allows the coupling strength to vary smoothly with the IMF clock angle, while also accounting for the magnitude of the transverse magnetic field and the solar wind speed.

This is physically relevant because magnetic reconnection at the dayside magnetopause is strongest when the IMF is southward, but weaker coupling can still occur at intermediate IMF orientations. The Newell coupling function therefore generalises the southward IMF picture introduced in section 2.2. Rather than treating the IMF effects as either on or off, it represents coupling as a continuous process whose strength depends on how favourably the field is oriented for reconnection. The function is defined as:

$$\frac{d\Phi_{MP}}{dt} = v_{sw}^{4/3} B_T^{2/3} \sin^{8/3}\left(\frac{\theta_c}{2}\right) \tag{3.4}$$

where v_{sw} is the solar wind velocity, $B_T = \sqrt{B_y^2 + B_z^2}$ is the magnitude of the transverse IMF component and $\theta_c = \operatorname{atan2}(B_y, B_z)$ is the IMF clock angle. The clock angle describes the IMF orientation in the Y-Z plane and therefore indicates how favourably the field is aligned for magnetic reconnection at the dayside magnetopause.

The $\sin^{8/3}(\theta_c/2)$ term is particularly important because it acts as a smooth orientation-dependent weighting. When the IMF is strongly southward, this term becomes large and the coupling strength increases. When the IMF is northward, the term becomes small, reflecting weaker but not necessarily zero coupling. In this sense, the Newell function acts as a soft rectifier since it strongly favours southward IMF, while still allowing intermediate clock angles to contribute to the coupling strength. This makes it better suited to representing the gradual variation in solar wind-magnetosphere coupling efficiency than a purely binary on/off version.

The Rectified Electric Field vB_s

In addition to the Newell coupling function, the rectified electric field vB_s is included as a simpler and more classical coupling proxy. Physically, this quantity represents the part of the solar wind electric field that is most relevant for dayside magnetic reconnection and magnetospheric convection. In practical terms, it combines solar wind speed with the southward component of the IMF, thereby providing a compact measure of how strongly the solar wind is influencing the Earth system. It is defined as

$$vB_s = v_{sw} \cdot \max(0, -B_z). \tag{3.5}$$

Here, the $\max(0, -B_z)$ term acts as a hard rectifier. Only southward IMF contributes to the coupling term, while northward IMF is set to zero. This reflects the classical picture that southward B_z is the most effective condition for dayside magnetic reconnection. Compared with the Newell coupling function, vB_s is therefore a simpler approximation as it does not account for intermediate clock-angle dependence, but it retains the most important effect of IMF orientation.

The importance of vB_s lies in its physical interpretability and its long history of use in empirical and semi-empirical space weather and density modelling. Although Newell et al. (2007) showed that more advanced formulations such as $d\Phi_{MP}/dt$ often achieve stronger statistical performance across a wide range of geomagnetic indices, vB_s remains a useful baseline proxy. Including both $d\Phi_{MP}/dt$ and vB_s therefore allows the model to use both a high-performing general coupling function and a simple, physically interpretable proxy of southward IMF energy input.

3.3.2. Temporal Feature Embedding

Earth's thermosphere exhibits significant thermal and dynamic inertia. Therefore, it does not respond instantaneously to energy injected by the solar wind or geomagnetic storms. Rather, it takes time to heat up, expand and subsequently cool down. Consequently, the density at any given moment is an integrated result of both the instantaneous conditions and the history of recent solar activity. As a result, providing a neural network solely with instantaneous measurements is insufficient. To accurately model the physical system, the model needs to be equipped with knowledge of prior solar and geomagnetic activity.

To capture this historical context, several temporal feature variants are engineered from the base input parameters. The considered temporal features include simple time-shifted lags, simple moving averages (SMA), exponentially weighted moving averages (EWMA) and discrete time integrations over varying windows. EWMAs are particularly advantageous in this context as they apply a decay factor that assigns greater weight to recent events while also retaining a diminishing memory of older disturbances. This formulation approximates the gradual decay behaviour characteristic of thermospheric recovery following geomagnetic disturbances (Fuller-Rowell et al. 1994).

The mathematical formulations of these transformations are provided below. Let $x(t)$ denote a base input parameter sampled at discrete time steps t with sampling interval Δt. The following temporal transformations are constructed.

Time-Shifted Lags

For a given lag τ, the lagged feature is defined as

$$x_{lag}(t, \tau) = x(t - \tau) \tag{3.6}$$

Where τ is an integer multiple of Δt. Multiple τ values are considered to expose the model to past feature states at various points in time.

Simple Moving Average (SMA)

For a window of length N_{SMA}, the SMA is defined as

$$\text{SMA}(t, N_{SMA}) = \frac{1}{N_{SMA}} \sum_{i=0}^{N_{SMA}-1} x(t - i\Delta t) \tag{3.7}$$

This represents a uniformly weighted average over the past N_{SMA} samples.

Exponentially Weighted Moving Average (EWMA)

The EWMA is defined as

$$EWMA(t, \alpha) = \alpha x(t) + (1 - \alpha) EWMA(t - \Delta t) \tag{3.8}$$

where $\alpha \in (0, 1)$ determines the rate of decay. To ensure the decay remains physically interpretable regardless of the sampling cadence, α is calculated as a function of a specified half-life H and the sampling interval Δt:

$$\alpha = 1 - 2^{-\Delta t / H} \tag{3.9}$$

In this study, where the temporal grid is synchronized to $\Delta t = 1$ minute, the formulation ensures that the contribution of a previous state is reduced by 50% after H minutes. The half-life therefore represents a physically interpretable decay timescale, consistent with the gradual thermospheric recovery behaviour following geomagnetic disturbances.

Discrete Time Integral

To approximate cumulative energy input, discrete integrations are computed over finite windows:

$$I_N(t, N_{int}) = \sum_{i=0}^{N_{int}-1} x(t - i\Delta t)\Delta t \tag{3.10}$$

which corresponds to a Riemann-sum approximation of the time integral over the previous N_{int} samples.

These complementary transformations allow the model to capture both discrete past states (lags), uniform finite-memory behaviour (SMA), exponentially decaying memory (EWMA)and cumulative energy deposition into the system (integration).

When constructing these temporal features, it is critical to account for the operational lags for each parameter. Several indices are only reported once the reporting window is closed. For example, as highlighted in Section 3.2, the 3-hourly Kp value is only reported once the 3-hour reporting window closed. Therefore, the current Kp value is not yet known. Rather, the most recent known value is that of the last 3-hour reporting window. Hence, the operational lags of each parameter are identified and applied to each parameter. This ensures that the neural network only relies on data that would be available in a real-time operational setting. This strictly prevents any leakage of future information into the feature set. A summary of each parameter and its operational lag is provided in Table 3.6.

Table 3.6: Operational reporting lags applied to solar and geomagnetic input parameters to ensure strict near-real-time compliance.

Data Category	**Indices**	**Operational Lag (hours)**
Solar Irradiance	$F_{10.7}$, S_{10}, M_{10}, Y_{10}	24
Global Geomagnetic	K_p, a_p	3
Ring Current	D_{st}	1
High-Cadence Geomagnetic	SYM (H/D), ASY (H/D), AE, AL, AU	0*
Polar Cap Activity	PCN	0*
Solar Wind	v_{sw}, ρ_{sw}, p_{sw}, T_{sw}	0*
IMF	B_x, B_y, B_z	0*

*Assumed instantaneous. Operational transmission latency (≤ 30 min) falls within acceptable near-real-time limits.

A novel contribution of the methodology developed in this thesis is the shift from sequence-dependent modelling to a history augmented feature architecture. By embedding the temporal history directly into the feature space via the engineered temporal features described above, the model's reliance on continuous, unbroken data streams is eliminated.

Unlike conventional recurrent architectures, such as Long Short-Term Memory (LSTM) networks or Gated Recurrent Units (GRUs), which require uninterrupted sequences to maintain internal hidden states, this approach renders each individual minute of data entirely self-contained. This design choice is one of the most critical aspects of the proposed setup, as it directly addresses the problem of having gaps in the accelerometer-derived datasets. Since each sample carries its own historical context, the presence of gaps in the accelerometer-derived density data does not pose an issue.

Furthermore, this continuous data independence permits the use of random shuffling of the data during training and validation. This helps mitigate the tendency of sequential training to over-emphasise the most recently seen periods and reduces the risk of forgetting patterns learnt from earlier years, thereby providing a more temporally consistent and generalisable correction model.

As will be demonstrated in the results presented in Chapter 5, this architecture allows the model to achieve significant performance gains over the JB2008 and NRLMSISE-00 empirical baselines. By resolving the constraints of data continuity, this setup provides a robust framework for high-fidelity thermospheric density reconstruction and orbit propagation, even when utilising fragmented density datasets.

In conclusion, the temporal feature embedding process enriches the raw solar and geomagnetic inputs with physically meaningful memory, while preserving operational realism through the implementation of operational lags. The resulting feature set enables the model to represent thermospheric response timescales without requiring recurrent sequence processing and provides a robust foundation for training on long and partially discontinuous satellite measurements.

3.3.3. Data Availability and Trust Metrics

Data Availability of High-Cadence Space Weather Inputs

A practical challenge in constructing a solar feature dataset is the outages of several high-cadence space-weather feeds. In particular, the IMF and solar wind parameters exhibit recurring outages in the historical records. The PCN index also contains occasional gaps. These interruptions occur frequently enough to not be considered rare edge cases. They should be treated as core data engineering and modelling considerations rather than minor preprocessing artefacts. An inspection of the original minute-level inputs showed that the largest data unavailability occurs in the solar wind and IMF channels. As seen in Table 3.7, around 21.6% of solar wind data is missing. About 7.24% of the IMF data is missing. The PCN index is considerably more reliable, with less than 1% missing data, but still requires explicit outage handling to preserve operational robustness.

Table 3.7: Summary of data availability for outage-prone high-cadence space-weather inputs in the original minute-level record. Reported percentages are aggregated over the full dataset.

Data Category	**Indices**	**Missing Data (%)**
Solar Wind (SW)	v_{sw}, ρ_{sw}, T_{sw}, p_{sw}	21.68
Interplanetary Magnetic Field (IMF)	B_x, B_y, B_z	7.24
Polar Cap Index (PCN)	PCN	< 1.00

Outage Handling Strategy

A naive treatment of these gaps would be to interpolate the missing values. However, interpolation would artificially synthesise intermediate space weather data, particularly during disturbed periods where sharp transitions in IMF and solar-wind conditions are physically meaningful. Additionally, interpolation would inject leakage into the dataset as discussed in Section 3.2. However, simply discarding rows containing missing IMF, SW or PCN values would result in substantial data loss and would degrade the continuity of the engineered temporal features. Since the objective is to construct a continuous minute-level solar feature stream, the outage handling approach must preserve temporal continuity while remaining physically sensible.

To satisfy these requirements, missing IMF, SW and PCN values are forward-filled (FFILL) rather than interpolated. This preserves the last known measurement and avoids introducing fabricated intermediate values. However, forward-filling alone is insufficient since a forward-filled value becomes progressively less representative as the outage duration increases. To account for this, a trust score feature is introduced for each outage-prone data category (IMF, SW and PCN). These features provide the neural network with insight into the current outage duration, allowing the neural network to distinguish between fresh observations and stale inputs.

Trust Score Formulation

Forward-filling preserves continuity, but it does not distinguish between a value that is one minute old and a value that has been carried forward for several hours. Treating both cases as equally reliable would be physically unrealistic. At the same time, a purely binary reliability flag is too coarse. Assigning the same label to a value forward-filled by one minute and a value forward-filled by 1000 minutes discards useful information about the severity of the outage.

To address this, a continuous trust score is introduced for each outage-prone data category. The trust score is designed to satisfy three requirements:

1. It should remain close to 1 for short outages (limited penalty for brief interruptions),
2. It should drop rapidly once the outage becomes operationally significant,
3. It should approach zero for prolonged outages (severely stale data).

These requirements are well matched by a shifted and scaled hyperbolic tangent function. Let $l_{\text{out}}(t)$ denote the outage length in minutes, i.e., the number of consecutive minutes since the last valid observation for a given data category. The trust score is then defined as

$$C_{\text{var}}(l_{\text{out}}) = 1 - \frac{1}{2}\left[\tanh\left(\frac{l_{\text{out}} - \mu_{\text{var}}}{\tau_{\text{var}}}\right) + 1\right] \tag{3.11}$$

where μ_{var} is an offset parameter and τ_{var} is a scaling parameter, both defined separately for each outage-prone data category.

This formulation maps outage length to a trust value C_{var} bound between 0 and 1 where μ_{var} controls the location of the transition and corresponds to the outage duration at which the trust score crosses approximately 0.5. The parameter τ_{var} controls the steepness of the transition and therefore determines how gradually or abruptly trust decays around μ_{var}.

As a result, short outages receive only a mild penalty, medium-duration outages are penalised increasingly strongly and prolonged outages are effectively marked as unreliable. This provides the neural network with a physically interpretable and smoothly varying indication of data freshness, while avoiding hard thresholds.

Selection of trust score Parameters

The trust score parameters μ_{var} and τ_{var} are chosen separately for IMF, SW and PCN, since these data categories differ in both physical variability and outage characteristics. A single shared trust-decay curve would therefore be suboptimal since some data streams become stale very quickly, while others remain informative for longer outage durations. These values are chosen based on the qualitative reasoning presented below. However, a more thorough sensitivity analysis could be considered for future work.

IMF data can change rapidly, especially during disturbed geomagnetic conditions, so stale IMF measurements become unrepresentative on relatively short timescales. At the same time, IMF outages are not negligible (approximately 7% of the record), which means the trust decay must not be so aggressive that IMF information is discarded too quickly during common short interruptions. To balance these considerations, the IMF trust score is parametrised with $\mu_{imf} = 10$ and $\tau_{imf} = 2.5$. This places the rapid trust reduction around a 10-minute outage and ensures a relatively sharp transition once the outage becomes non-trivial.

Solar-wind variables are generally smoother than IMF components and their predictive utility can remain meaningful over somewhat longer outages. However, solar-wind outages are the most frequent in the dataset (approximately 21.6% missing), so the trust score formulation must be sufficiently tolerant of short-to-moderate interruptions. For this reason, a slower and broader decay is used with $\mu_{sw} = 20$ and $\tau_{sw} = 7.5$. This yields a gradual reduction in trust, with the 50% trust level occurring at an outage duration of approximately 20 minutes.

PCN is considerably more complete ($< 1\%$ missing) than the other two categories, but it is a particularly valuable near-real-time proxy for high-latitude geomagnetic activity and thermospheric forcing. Because outages are rare, the model can afford to be more conservative when PCN is unavailable. A sharper and earlier decay is therefore selected with $\mu_{pcn} = 5$ and $\tau_{pcn} = 1$. This causes trust to drop rapidly after only a few minutes of outage, effectively signalling to the neural network that one of the most informative real-time drivers is temporarily unavailable.

The selected μ and τ values for each outage-prone data category are summarised in Table 3.8, together with the corresponding design rationale. This table provides a compact overview of how the trust score decay behaviour has been adapted to the outage characteristics and physical behaviour of IMF, SW and PCN inputs.

Table 3.8: Category-specific trust score parameters used in Equation (3.11). The offset parameter μ controls the outage duration at which trust drops to approximately 50%, while τ controls the steepness of the decay.

Data Category	μ_{var} [min]	τ_{var} [min]	**Design rationale**
IMF	10	2.5	Rapid variability. Trust should decay quickly once outages exceed short interruptions.
SW	20	7.5	Smoother dynamics and frequent outages. Trust decay should be more gradual.
PCN	5	1.0	High-value and rarely missing. Trust should drop sharply when unavailable.

The resulting category-specific trust score decay curves are shown in Figure 3.7.

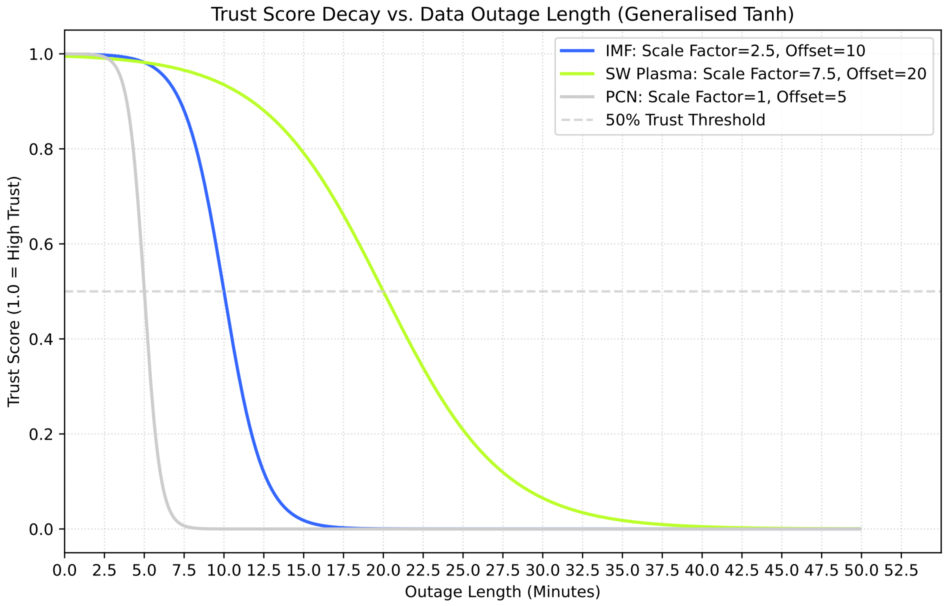


Figure 3.7: Category-specific trust score decay curves used for outage-prone input streams. The shifted and scaled hyperbolic tangent function allows short outages to retain high trust while rapidly down-weighting stale values once the outage duration exceeds a category-specific threshold.

Together with forward-filling, these trust scores enable a continuous feature stream while explicitly encoding the reliability of outage-affected inputs. Rather than treating all measurements as equally reliable, the network can learn to down-weight stale carried-forward values during prolonged outages and rely more strongly on other predictors when confidence in a given data category is low.

In addition to the trust scores attached to the instantaneous forward-filled IMF, SW and PCN variables, corresponding trust indicators are also constructed for the engineered temporal features. In the feature-generation pipeline, each variable containing outages is first mapped to its category (IMF, SW, or PCN), while coupling functions such as vB_s and the Newell coupling function are assigned a combined trust defined by the minimum of the relevant IMF and SW trust scores. For each feature, the operationally lagged version of the trust score is first generated, after which trust metrics are propagated through the temporal operators in parallel with the physical feature itself. Specifically, for simple moving-average features, a rolling mean of the trust score is computed, representing the average reliability of the samples contributing to that window. For exponentially weighted moving-average features, an EWMA of the trust score is computed so that recent reliability contributed more strongly than older reliability. For integration features, a rolling minimum trust score is used as a conservative indicator, flagging whether any part of the integration window is affected by low-confidence values. In this way, the model receives not only the engineered temporal features themselves, but also an estimate of how trustworthy the information content of each temporal feature is.

3.3.4. Final Feature Selection

To capture the thermal and dynamic inertia of the thermosphere, the base solar and geomagnetic parameters are augmented with various temporal embeddings. Table 3.9 provides a summary of the transformations applied to each variable category.

The final feature selection is primarily driven by physical interpretability, standard space weather practices and the desire to reduce the amount of highly correlated features. It is acknowledged that this selection relies on personal judgment and may not represent the absolute optimal feature set. A comprehensive feature ablation study is recommended for future work to provide a more justified feature selection.

To capture the thermal and dynamic inertia of the thermosphere, the base solar and geomagnetic parameters are augmented with various temporal variants. However, several raw parameters are entirely excluded from the final input vector to minimise redundancy and noise. The K_p index is dropped due to its high correlation with the linear a_p index. The solar wind proton temperature (T_{sw}) is removed as it is highly dependent on flow speed, density and pressure. The AU index is excluded because the auroral electrojet activity is already represented by the AE and AL indices since $AE = AU - AL$. The B_x component of the Interplanetary Magnetic Field (IMF) is discarded, as Trattner, Petrinec, and Fuselier (2021) indicate that the influence of the B_x component on magnetic reconnection is uncertain. Furthermore, the D-component disturbance indices (ASY/D and SYM/D) are discarded, since the H-components provide a cleaner ring-current representation (T. et al. 2010). Finally, the PCS index is excluded due to near real-time data availability concerns, as detailed in Section 3.1.

The selection of specific temporal windows for the remaining features is designed to balance short-term atmospheric reactions with long-term recovery dynamics. For solar irradiance proxies ($F_{10.7}$, S_{10}, M_{10}, Y_{10}) an 81-day average is incorporated. This is a standard space weather convention, often used in empirical models such as JB2008 (Bowman et al. 2008), which captures the base level of solar extreme ultraviolet (EUV) heating. Crucially, a trailing 81-day simple moving average is used rather than the traditional centred average to prevent future data leakage. The specific discrete lags of 24h, 48h and 120h are adopted from the standard operational reporting conventions found in the SOLFSMY dataset. Note that the SOLFSMY conventions specify a 1 day lag for $F10$ and $S10$, an additional 2 day lag is added in this version to allow more ability to capture recent solar irradiance behaviour. Additionally, a 1 day lag is added to $M10$ to complement the conventional 2 day lag.

For geomagnetic and solar wind drivers, memory windows are tailored to the physical decay rates of the phenomena they represent. Ring current indices (e.g., D_{st}, SYM/H) track the trapping of energetic particles, which decay slowly over several days following a geomagnetic storm. Consequently, these features are assigned longer memory variants, such as an EWMA with a 24h half-life. In contrast, auroral indices (e.g., AE, AL) and solar wind drivers more closely reflect rapid, localised energy injections via Joule heating. To capture the quicker responses, these features are assigned shorter memory windows, typically 1-hour SMAs or 3-hour integrals (Pulkkinen 2007).

To properly represent the orientation of the interplanetary magnetic field without discontinuities, the IMF clock angle $\theta_c = \text{atan2}\left(\frac{B_y}{B_z}\right)$ is decomposed into its sine, cosine and half-sine components. This decomposition resolves the periodic boundary discontinuity. Specifically, the half-sine term enables the model to emulate the $\sin^{8/3}\left(\frac{\theta_c}{2}\right)$ merging behaviour observed in the Newell coupling function Equation (3.4), recognising that energy transfer peaks during southward IMF (Borovsky 2023).

Finally, to ensure the model evaluates data reliability, temporal trust features are added that represent the temporal windows of the physical features they monitor. As combined trust serves as a reliability metric for derived coupling functions (which require inputs from both IMF and solar wind data sources), its temporal variants directly mirror those chosen for those coupling functions.

Table 3.9 summarises the complete, finalised selection of temporal variants per base feature used to construct the neural network's input vector. In this summary, the numerical values within each column specify the temporal window or decay constant applied to that transformation, while a '–' indicates that the specific variant is omitted for that feature to avoid redundancy.

Table 3.9: Final solar feature selection showing the temporal variants per base feature. The applied temporal transformations are defined by the lag amount τ, the simple moving average window N_{SMA}, the exponentially weighted moving average half-life α and the integration window length N_{int}. The entries filled with – indicate that for that specific feature, the respective temporal feature is not included in the dataset.

Base Feature	**Lag (τ)**	**SMA (N_{SMA})**	**EWMA (α)**	**Integral (N_{int})**
Interplanetary				
bx_gsm_nT	–	–	–	–
by_gsm_nT	0m	1h	3h	–
bz_gsm_nT	0m	1h	3h	–
sw_speed_km_s	0m	3h, 6h	–	–
sw_proton_density_n_cc	0m	–	–	–
sw_proton_temperature_K	–	–	–	–
sw_flow_pressure_nPa	0m	1h	–	–
Auroral				
ae_index_nT	0m	1h	–	–
al_index_nT	0m	–	–	3h, 12h
au_index_nT	–	–	–	–
Ring Current				
sym_d_nT	–	–	–	–
sym_h_nT	0m	–	3h, 24h	–
asy_d_nT	–	–	–	–
asy_h_nT	0m	1h	–	–
Storm Context				
dst_nT_1hr	1h	–	24h	–
Geomagnetic				
Kp_3hr	–	–	–	–
ap_3hr	3h, 6h	24h	3h	–
Polar Activity				
pcn_index	0m	–	6h	3h
pcs_index	–	–	–	–
Solar Flux				
F10_daily	24h, 48h	81d	–	–
S10_daily	24h, 48h	81d	–	–
M10_daily	24h, 48h	81d	–	–
Y10_daily	24h, 120h	81d	–	–
Coupling Functions				
vBs	0m	–	–	1h
Newell_Coupling_dPhi_dt	0m	–	–	1h, 3h
Additional				
IMF_theta_sin_half	0m	–	–	–
IMF_theta_sin	0m	–	–	–
IMF_theta_cos	0m	–	–	–
Trust Metrics				
IMF_Trust_Score	0m	1h	3h	–
SW_Trust_Score	0m	1h, 3h, 6h	–	–
PCN_Trust_Score	0m	–	6h	3h
Combined_Trust	0m	–	–	1h, 3h

In addition to the temporally expanded solar and geomagnetic drivers listed in Table 3.9, the final input vector also includes a set of spatial, geometric and identifier features that are used directly without additional temporal variants. These features are summarised in Table 3.10.

Table 3.10: Spatial, geometric and identifier features included directly in the final input vector without additional temporal expansion.

Feature	Description
alt_km	Geodetic altitude in kilometres (above GRS80 ellipsoid)
lat_deg	Geodetic latitude in degrees
arg_lat_cos	Cosine of the argument of latitude
arg_lat_sin	Sine of the argument of latitude
lon_cos	Cosine of geodetic longitude
lon_sin	Sine of geodetic longitude
lst_cos	Cosine of local solar time
lst_sin	Sine of local solar time
sat_id_int	Integer satellite identifier used for the satellite embedding

3.4. Data Normalisation Strategy

Neural networks generally train more reliably when input features are placed on comparable numerical scales. If raw features with very different magnitudes and distributions are used directly, optimisation becomes less stable, gradients can become poorly conditioned and the model may over-emphasise variables with a higher magnitude. For this reason, a dedicated normalisation strategy is applied to all input features prior to training.

Rather than applying a single normalisation method to every feature, a strategy based on the feature distribution is adopted. The feature set contains a mixture of bounded geometric variables, approximately Gaussian variables, strongly right-skewed heavy-tailed variables, cyclic variables represented as sine–cosine pairs and trust score features that are already bounded and physically interpretable. Applying the same transformation to all of these would be suboptimal.

The normalisation strategy for each feature is selected based on inspection of its empirical distribution. For each candidate variable, the raw distribution and the log-transformed distribution have to be inspected. Features with approximately symmetric, bell-shaped distributions are standardised directly using z-scaling, while strongly right-skewed variables are first transformed with a logarithm and then z-scaled. Bounded variables are scaled using min–max normalisation and cyclic sine–cosine features are treated separately through vector renormalisation (discussed previously in Section 3.2). The final decisions for each feature are documented in a normalisation decision table and implemented in the training pipeline.

3.4.1. Bounded Distributions

Bounded variables are scaled using min–max normalisation. This category primarily includes geometric features such as latitude and altitude, which are naturally constrained to finite ranges. For a bounded variable x, min–max scaling is applied as

$$x_{\text{norm}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}, \tag{3.12}$$

where x_{min} and x_{max} are estimated from the training set only. This maps the training distribution to the interval $[0, 1]$, while preserving the relative ordering of values. Min–max scaling is preferred for these variables because their physical ranges are bounded and their distributions are not necessarily well described by a Gaussian assumption. In addition, preserving the bounded nature of these features improves interpretability and avoids unnecessarily amplifying edge values.

3.4.2. Approximately Gaussian Distributions

Features with approximately symmetric, near-Gaussian distributions are standardised using z-scaling. For a variable x, the transformation is

$$x_{\text{norm}} = \frac{x - \mu_x}{\sigma_x}, \tag{3.13}$$

where μ_x and σ_x are the training-set mean and standard deviation, respectively. This transformation centres each feature around zero and rescales it to unit variance, which is beneficial for gradient-based optimisation. Many solar and geomagnetic features (and several of their engineered lag/int/SMA/EWMA variants) sufficiently resembled a Gaussian distribution to be normalised in this way without any additional transformation.

3.4.3. Heavy-Tailed Distributions

Several variables in the feature set exhibit strong right skew and heavy tails, particularly among coupling functions, auroral indices and solar-wind-derived quantities. For these variables, direct z-scaling would leave a highly asymmetric distribution and would cause the bulk of samples to be compressed near the mean while a small number of extreme values dominate the variance.

To mitigate this, a logarithmic transform is applied prior to z-scaling. In the final implementation, a $\log_{10}$ transform is used with a small positive floor ϵ to ensure numerical stability:

$$x_{\log} = \log_{10}(\max(x, \epsilon)) \,, \tag{3.14}$$

followed by standard z-scaling of $x_{\log}$ using

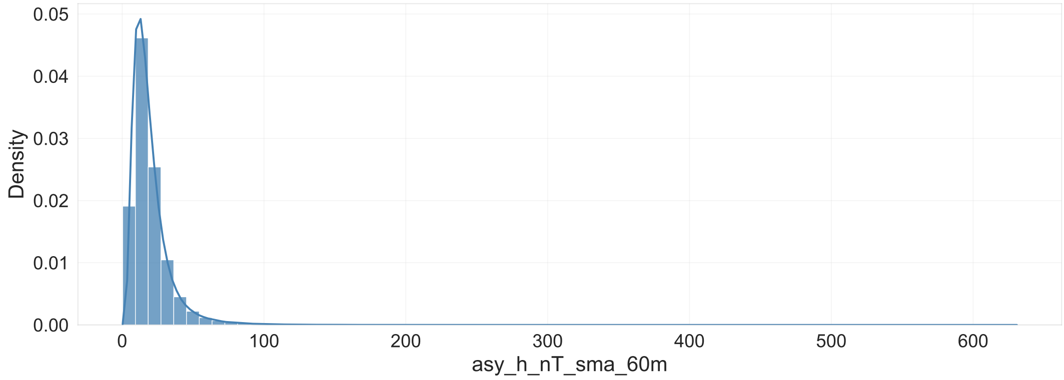
$$x_{\mathrm{norm}} = \frac{x_{\log} - \mu_{x,\log}}{\sigma_{x,\log}}. \tag{3.15}$$

This $\log_{10} +$ z-scaling strategy converts strongly skewed needle-like distributions into shapes that are much closer to a Gaussian distribution. Figure 3.8 illustrates the distribution-based decision process. The feature `asy_h_nT_sma_60m` exhibits a strongly skewed, needle-like raw distribution, while a logarithmic transformation produces a substantially more Gaussian-like shape. In contrast, `bz_gsm_nT_shift_0m` is already approximately symmetric in raw form and is therefore standardised directly using z-scaling.

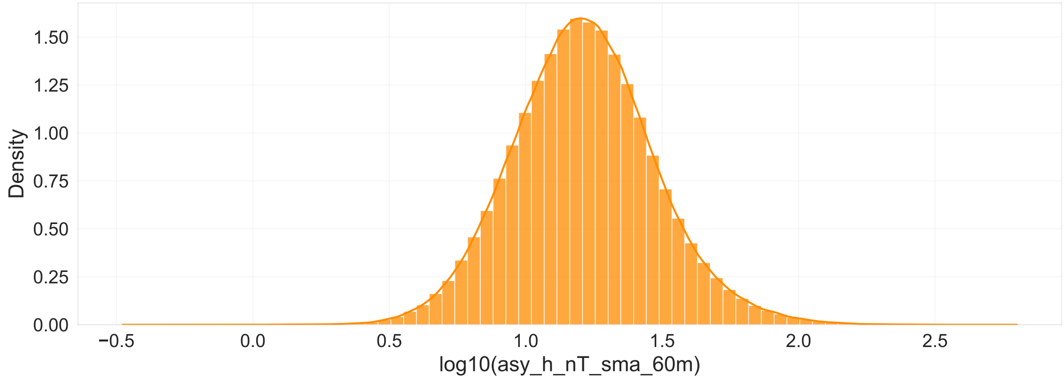
The same $\log_{10}$ +z-scaling logic is also applied to the density target and the baseline JB2008 density feature, which span multiple orders of magnitude. The processing of the target variable is discussed in more detail in Section 3.5.

Distribution-Aware Normalisation Examples

(a) asy_h_nT_sma_60m — Raw distribution



(b) asy_h_nT_sma_60m — Log-transformed distribution



(c) bz_gsm_nT_shift_0m — Raw distribution

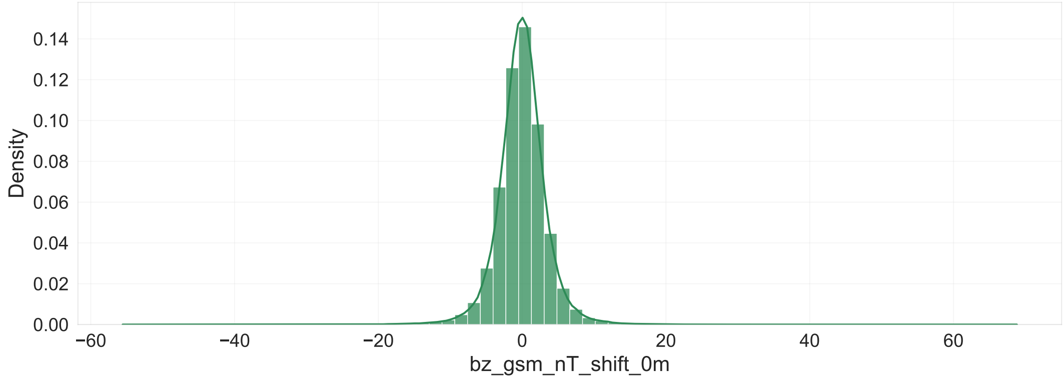


Figure 3.8: Illustrative examples of the distribution-aware normalisation strategy. (a) Raw distribution of `asy_h_nT_sma_60m`, showing a strongly skewed, needle-like shape. (b) Log-transformed view of `asy_h_nT_sma_60m` demonstrating a substantially more Gaussian-like shape. (c) Raw distribution of `bz_gsm_nT_shift_0m`, which is already approximately Gaussian and is therefore suitable for direct z-scaling. The y-axis shows probability density, so values above 1 are possible for concentrated distributions.

3.4.4. Cyclic Vector Renormalisation

Cyclic variables (e.g., local solar time, longitude and argument of latitude) are represented as sine–cosine pairs. As discussed earlier in Section 3.2.3, downsampling can cause the sine–cosine encoding vector to deviate from unit norm. To preserve the geometric meaning of the angular encoding, these pairs are renormalised prior to use. In the preprocessing pipeline, these cyclic variables are therefore assigned a dedicated `renorm` strategy and handled separately from standard column-wise scaling, ensuring that the angular encoding remains geometrically consistent.

3.4.5. Pass-Through Features

Not all features required explicit numerical scaling. In particular, the trust score features introduced in Section 3.3.3 (e.g., IMF, SW and PCN trust metrics and their derived temporal variants) are already bounded and physically interpretable. Applying an additional normalisation step to these variables would reduce their interpretability without providing a clear optimisation benefit. These features are therefore left unchanged.

3.4.6. Normalisation Summary

The resulting preprocessing pipeline applies a feature-specific normalisation strategy tailored to the statistical structure of each input variable. Instead of enforcing a single transformation across all features, each feature is assigned the scaling method that best preserves its physical meaning while placing it on a numerically suitable scale for neural-network training. Table 3.11 provides a high-level summary of the normalisation strategies used in this work.

Table 3.11: Summary of the feature-specific, distribution-aware normalisation strategy applied to the input dataset.

Normalisation Strategy	Applied To
Min–max scaling	Bounded geometric variables
Z-scaling	Approximately Gaussian variables
Log_{10} + z-scaling	Positive heavy-tailed variables
Vector renormalisation	Cyclic sine–cosine pairs
Pass-through	Trust scores

All normalisation statistics (mean, standard deviation, minimum and maximum) are computed from the training split only and then reused for validation and testing. This ensures that the preprocessing pipeline prevents leakage of information from the validation or test sets into training, thereby ensuring that validation and test performance remain unbiased and reflect a truly out-of-sample evaluation. The resulting normalisation statistics are stored and reused during inference, ensuring consistent preprocessing across training, validation, testing and operational deployment.

A complete feature-level overview of the selected normalisation strategy is provided in Table A.1 located in the appendix. Collectively, these transformations place the inputs on numerically comparable scales and substantially reduce optimisation issues caused by varying orders of magnitude and strongly skewed distributions. However, it should be noted that the transformed features do not become identical in distribution. In particular, variables that are standardised directly using z-scaling retain different shape characteristics from variables that are first log-transformed and then z-scaled. This may influence how easily the neural network can extract signal from different feature groups and therefore may affect the relative sensitivity of the model to specific inputs. This trade-off is intentional since preserving physically meaningful structure while improving numerical conditioning was considered more important than forcing all features into an identical distribution.

As indicated in Table A.1, the feature `sat_id_int` is treated differently from the other inputs. Since it encodes a categorical mission identifier rather than a continuous physical variable, it is excluded from numeric normalisation. Section 4.3 describes how the satellite ID is incorporated into the model architecture as a learned categorical embedding.

3.5. Target Variable Formulation

The neural network is not trained to predict absolute thermospheric density directly. Instead, it is trained to predict a correction relative to the baseline JB2008 density model. The rationale for selecting JB2008 as the reference model is discussed in Section 4.1. By using an existing model as a baseline, numerical stability and learning efficiency improve, since the network can focus on modelling the residual structure that JB2008 does not capture, rather than relearning the entire density behaviour from scratch.

Let $\rho_{\mathrm{obs}}(t)$ denote the accelerometer-derived density target and $\rho_{\mathrm{JB2008}}(t)$ the corresponding baseline model density at the same epoch. A residual target is formulated in logarithmic space as

$$r(t) = \log_{10}\rho_{\mathrm{obs}}(t) - \log_{10}\rho_{\mathrm{JB2008}}(t). \tag{3.16}$$

Thermospheric density spans multiple orders of magnitude. Expressing the target in logarithmic form compresses the range and turns multiplicative density errors into additive residuals, which are generally easier to learn for a neural network.

A further refinement is introduced to reduce the impact of mission-dependent offsets. Although the residual time series from different satellites exhibit similar physical variability, their mean offsets can differ due to factors such as drag coefficient tuning. If these offsets are not taken into account in the target, the neural network may overemphasise learning trivial constant corrections (e.g. a near-constant scale factor per mission) instead of learning the more challenging and physically relevant time-varying fluctuations.

To mitigate this, a satellite-specific bias term is removed from the raw log residual. Let s denote the satellite identity and let μ_s denote the mean value of $r(t)$ for satellite s, estimated from the training set only. A global standard deviation σ_r is then computed over the raw log-residuals across the training set and used to scale the bias-corrected residual target. The final training target is defined as

$$y(t) = \frac{r(t) - \mu_s}{\sigma_r}, \tag{3.17}$$

where $y(t)$ is the normalised bias-corrected residual target used for training.

This formulation encourages the model to focus primarily on reducing the residual variability (i.e., the standard-deviation component) rather than merely correcting the mean offset. In other words, the network is discouraged from converging to a trivial mission-specific multiplicative correction and is instead forced to learn the response of density to solar and geomagnetic influences.

It should be noted that using JB2008 as the baseline target reference is both a strength and a limitation. It is a strength because it provides a physically meaningful first-order estimate and stabilises the learning problem. However, it also implies that the neural network is trained to correct JB2008 rather than to model density independently and therefore the learned correction may inherit some limitations of the baseline representation. This trade-off is accepted in the present work because the objective is to improve near-real-time operational density estimation relative to an established empirical baseline.

3.6. Dataset Split Strategy

A dedicated dataset split strategy is designed to ensure a physically meaningful and leakage-resistant separation between training, validation and final testing. Rather than using a random row-wise split, the accelerometer-derived density dataset is split into several training and validation periods.

This choice is necessary for two reasons. First, the feature set contains temporal lags, moving averages, exponentially weighted moving averages and integration windows, which introduce strong temporal dependence between neighbouring samples. A random split would therefore lead to substantial leakage between training and validation samples. Second, multiple satellites are active during overlapping time periods. If one satellite is used for training while another satellite at the same epoch is used for validation, the model could indirectly benefit from exposure to the same underlying space weather conditions, weakening the independence of the validation set.

For these reasons, validation is defined using explicit time blocks that are applied consistently across all satellites active during those periods. In other words, if a given time interval is assigned to validation, all eligible accelerometer data from all active satellites within that interval are assigned to validation. This ensures that validation measures the model's ability to generalise to unseen time periods and physical regimes, rather than merely unseen rows.

3.6.1. Train-Validation Split on Accelerometer Data

The train–validation split is constructed from the accelerometer-derived density data of CHAMP, GRACE-A, GRACE-B and GOCE. Swarm and GRACE-FO are excluded from this splitting procedure, since their accelerometer data is not used for model training, as mentioned in Section 3.1.

The validation intervals are selected based on the solar cycle phases and mission-specific occurrences. In particular, the chosen blocks are designed to span both solar minimum and more active periods, so that validation performance reflects the model's behaviour across the different regimes rather than a single background state. The selected intervals also include mission transition phases, which are operationally important and often associated with changes in orbital altitude, drag environment, or data characteristics.

In addition, the validation set is constructed to include the final GOCE re-entry regime, which provides a particularly relevant high-drag orbital scenario. Care was also taken to ensure that each training mission contributes at least two validation periods, so that validation is not tied to a single mission phase but instead samples multiple orbital regimes for each satellite. Together, these choices make the validation set more representative of the range of conditions the model is expected to handle.

A further benefit of this split design is that the validation set includes a phase, from 2016 onward, in which GRACE validation data remains available while no corresponding training data exists around that period. This is intentional and beneficial, as it makes this part of the validation set more challenging. It provides a truly out-of-sample addition to the validation set since the validation set tests not only interpolation across familiar conditions but also generalisation to later unseen periods.

An important additional design rule is that validation periods are universal across all relevant satellites active at the time. This prevents cross-satellite leakage, i.e. the situation where the model is trained on one satellite at a given timestamp and validated on another satellite at the same timestamp under nearly identical solar conditions.

3.6.2. Buffer Zones to Prevent Temporal Leakage

To further reduce leakage at the train–validation boundaries, a buffer zone of 7 days is applied around each validation block. All samples within this buffer are excluded from training. This buffer serves two purposes. First, it prevents the model from validating on samples that occur immediately after or before nearby training samples, which would otherwise remain strongly autocorrelated in time. Second, it mitigates leakage from engineered temporal features, many of which use recent history through lags, moving averages and integrations. Although the longest solar-index averages span much longer windows, the 7 day exclusion buffer is chosen as a practical and conservative break that substantially reduces boundary contamination for memory features.

3.6.3. Final Test Strategy and Separation from Train--Validation

No conventional test split is created from the accelerometer training dataset. Instead, final testing is reserved for an orbit-propagation evaluation using Swarm A precise orbit determination (POD) data. This final test is intentionally separated from the train–validation accelerometer pipeline and is discussed in detail in Section 4.8.

This split design reflects the intended use case of the model. The train-validation split is used to train, optimise and compare candidate neural network configurations on accelerometer-derived density targets, while the final assessment is performed through orbit propagation and benchmarked against empirical density models (including NRLMSISE-00 and JB2008) on an independent Swarm-A POD dataset.

3.6.4. Summary of Split Design

The dataset split strategy adopted in this work is summarised in Table 3.12. The table lists the mission groups used in the accelerometer-based train and validation pipeline, their assigned role, the realised train and validation proportions per satellite and the validation time periods. Because the split is defined using explicit time periods rather than random sampling, the final train-validation percentages differ between satellites depending on mission lifetime and overlap with the selected validation blocks.

Table 3.12: Data split configuration used for the accelerometer-derived density dataset. Validation intervals are applied consistently across all active training satellites at the same epoch. A 7 day exclusion buffer is applied around each validation block to reduce temporal leakage. The realised train-validation percentages differ by satellite because the split is period-based rather than randomly sampled.

Mission	Sat.	Role	Actual Train – Val Split	Validation Blocks
CHAMP	–	Train & validation set	94.7% / 5.3%	2008-06-01 – 2008-09-01 2009-11-01 – 2010-02-01
GRACE	A	Train & validation set	84.1% / 15.9%	2008-06-01 – 2008-09-01 2009-11-01 – 2010-02-01 2013-06-01 – 2013-11-01 2016-01-01 – 2017-10-31
	B	Train & validation set	88.6% / 11.4%	2008-06-01 – 2008-09-01 2009-11-01 – 2010-02-01 2013-06-01 – 2013-11-01 2016-01-01 – 2017-05-22
GOCE	–	Train & validation set	81.9% / 18.1%	2009-11-01 – 2010-02-01 2013-06-01 – 2013-11-01
Swarm	A/B/C	Excluded from train & validation set	–	–
GRACE-FO	C	Excluded from train & validation set	–	–

A visual overview of the resulting split is provided in Figure 3.9. The figure shows the altitude-time coverage of all missions together with the solar-cycle context, similar to Figure 3.1. The figure confirms that the validation intervals are applied consistently across all active training satellites. It also illustrates that the selected validation blocks span multiple orbital and solar-activity regimes, including the late GRACE-only validation phase, while Swarm and GRACE-FO remain excluded from the train and validation dataset and are shown only for contextual reference.

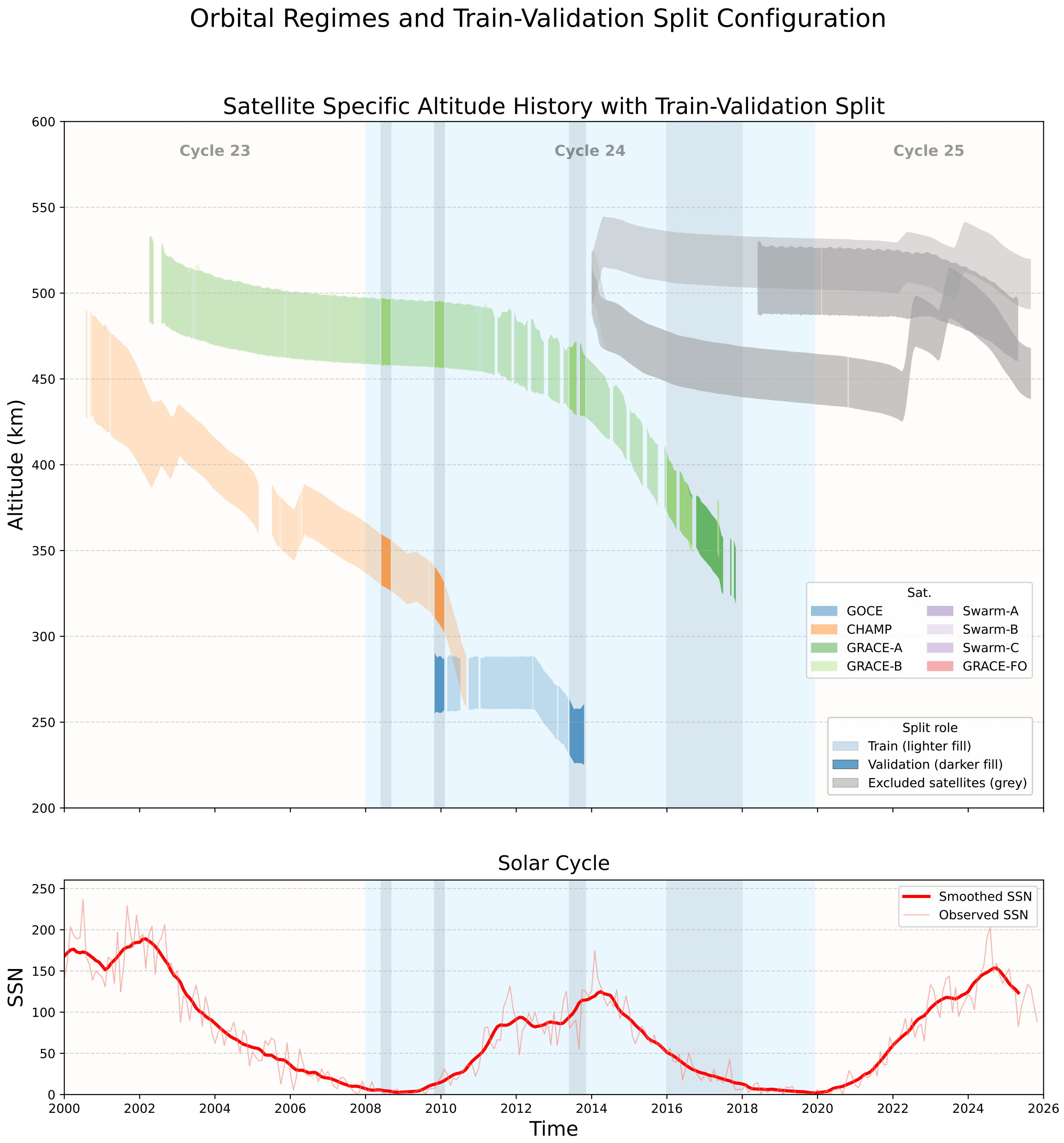


Figure 3.9: Overview of the satellite-specific train-validation split configuration over altitude and time, shown together with the solar cycle. The top panel shows the altitude history of all considered satellites as daily altitude envelopes. For the selected training missions (CHAMP, GRACE-A/B and GOCE), colour indicates the satellite identity, while opacity indicates the split role. Lighter segments correspond to training data and darker segments correspond to validation blocks. Missions excluded from training and validation (Swarm and GRACE-FO) are shown in grey for contextual reference only. The bottom panel shows the observed and smoothed sunspot number (SSN), providing solar-cycle context for the chosen validation intervals.

4

Methodology

This chapter describes the methodological framework used to develop, train and evaluate the proposed neural network-based correction to the JB2008 density model. Building on the physical motivation and literature review presented in Chapter 2 and the data processing strategy described in Chapter 3, the chapter first introduces the overall problem formulation and method overview in Section 4.1. It then defines the correction target and its inverse mapping in Section 4.2, presents the neural network architecture in Section 4.3 and describes the training procedure and hyperparameter selection strategy in Sections 4.4 and 4.5. Finally, the two-stage evaluation pipeline is formalised in Section 4.6, with the accelerometer-derived density validation stage defined in Section 4.7 and the independent POD-based orbit-propagation validation stage defined in Section 4.8. Together, these elements provide the methodological basis for assessing whether a lightweight correction to an empirical density model can improve both density reconstruction and orbit-propagation performance.

4.1. Problem Definition and Method Overview

This thesis has two main objectives. First, to improve atmospheric density reconstruction compared to accelerometer-derived density data and second, to assess whether improved density reconstruction leads to more accurate orbit propagation. In this context, the focus is not on replacing existing empirical thermospheric models outright, but on developing a practical correction framework that can be integrated into an operational workflow and subsequently evaluated in orbit propagation. The proposed approach therefore aims to improve atmospheric density estimates while preserving the computational efficiency provided by commonly used empirical models.

Rather than directly predicting atmospheric density, the chosen approach relies on a neural network trained to provide a correction to the JB2008 estimate. JB2008 is adopted as the baseline atmospheric model for this work due to its operational relevance, computational efficiency and strong performance in disturbed conditions. This correction-based design preserves the physical and operational utility of the baseline model while allowing the neural network to learn systematic, time-varying deviations linked to solar and geomagnetic influences.

A key requirement of the proposed framework is operational relevance under near-real-time conditions. In practice, this means the model should rely on input features that can be obtained with limited latency and should remain lightweight enough for repeated inference in an orbit propagation context. Although a strict latency threshold is not imposed in this work, the methodology used is explicitly designed for near-real-time use, which is reflected by the evaluation strategy.

To assess whether the proposed correction framework provides real value, a two-stage evaluation philosophy is adopted. Stage A is a development validation stage based on accelerometer-derived density estimates. This stage provides a relatively fast and controlled way to compare model configurations and training epochs. Hence, it is used for model selection during development. However, because it remains tied to accelerometer-derived density targets, it does not by itself demonstrate improved orbit propagation performance.

For that reason, stage B is defined as the final independent validation stage. In this stage, the corrected density model is evaluated in orbit propagation against precise orbit determination (POD) data and compared against standard empirical models. This downstream test is designed to determine whether improvements observed in the accelerometer-derived density validation translate into reduced along-track orbit propagation error, which is a key operational performance metric in density modeling.

The remainder of this chapter formalises the target framework used by the neural network correction model, describes the model architecture and training procedure. Additionally, it defines the two-stage evaluation protocol used to assess performance.

4.2. Model Target Framework

This section explains the formulation of the target variable used by the neural network correction model. The processed feature set and the initial target representation were introduced in Chapter 3. Building on that foundation, this section defines the final target used during training, motivates the adopted normalisation strategy for the target and describes how the neural network output is mapped back to a physically meaningful corrected density estimate.

4.2.1. JB2008-Centred Correction Framework

Although the final goal of this thesis is to improve operational density estimation, the model is not formulated as a direct predictor of absolute thermospheric density. Instead, the neural network is trained to estimate a correction to the JB2008 density. This choice is made for both practical and methodological reasons.

First, JB2008 already provides a physically meaningful and computationally efficient baseline estimate and is widely used in orbital applications. Framing the task as a correction problem allows the model to preserve the baseline behaviour where JB2008 already performs well, while learning systematic deviations in regimes where additional information from solar and geomagnetic drivers is advantageous. Furthermore, the correction-based formulation is more naturally aligned with operational deployment, since the neural network can be interpreted as a lightweight model layer on top of an established empirical density model rather than as a stand-alone replacement, given that the correct operational data pipeline is implemented.

4.2.2. Log-Residual Target Definition

The log-residual target is defined in full in Equation (4.1). Here, it is only recalled briefly because it forms the direct prediction target of the neural network. The target is written as

$$r(t) = \log_{10}\left(\frac{\rho_{\mathrm{obs}}(t)}{\rho_{\mathrm{JB2008}}(t)}\right), \tag{4.1}$$

so that the network predicts a correction relative to the JB2008 baseline rather than the full density directly.

The motivation for this formulation is discussed in Equation (4.1). In brief, it provides a physically interpretable multiplicative correction to JB2008 and yields a more numerically well-behaved learning target than density in linear space.

4.2.3. Target Normalisation and Bias Handling

As discussed in Equation (4.1), inspection of the accelerometer-derived density data and JB2008 outputs shows a systematic mean offset in the raw log-residuals that differs between satellites. A substantial part of this offset is likely not caused by time-varying thermospheric physics, but is instead a consequence of satellite-specific data handling (e.g. bias introduced by choosing a specific C_d during density estimate derivation from accelerometer data). If a standard loss is applied directly to the raw residual target in **??**, the neural network can reduce the loss substantially by learning this quasi-static offset, even if it does not improve the modelling of time-varying density variability.

Since the primary objective of this thesis is to improve drag-relevant density variability for operational orbit propagation, the target is normalised such that the neural network is encouraged to model the dynamic component of the residual rather than the satellite-specific mean offset. To achieve this, the raw log-residual is transformed as

$$y(t) = \frac{r(t) - \mu_s}{\sigma_r}, \tag{4.2}$$

where μ_s is the satellite-specific mean of $r(t)$and σ_r is the global standard deviation of $r(t)$ across the training set.

In this formulation, μ_s captures the satellite-specific offset, while σ_r scales the target by the overall variability magnitude. The resulting target therefore emphasises the time-varying component of the residual rather than the static mean offset.

To avoid information leakage, both μ_s and σ_r are computed using training data only. For satellites not present in the final training dataset, a neutral fallback of $\mu_s = 0$ is used during inference. In the final POD-based propagation validation, any remaining mean offset is further controlled through the model-specific C_d tuning procedure, such that the comparison remains focused on the relevant model behaviour instead of satellite-specific offsets.

4.2.4. Inverse Transformation and Density Reconstruction

The neural network outputs a prediction $\hat{y}(t)$ in the normalised target space defined by Equation (4.2). During inference, this prediction is first mapped back to the log-residual space:

$$\hat{r}(t) = \sigma_r \hat{y}(t) + \mu_s, \tag{4.3}$$

after which the corrected density estimate is reconstructed as

$$\hat{\rho}_{\mathrm{NN}}(t) = \rho_{\mathrm{JB2008}}(t)\, 10^{\hat{r}(t)}. \tag{4.4}$$

This inverse mapping is used during validation and orbit-propagation inference. As such, it is ensured that the neural network output is consistently and correctly mapped to the intended JB2008 correction.

4.3. Model Architecture

This section describes the neural network architecture used to predict the normalised log-residual target defined in Section 4.2. The model is implemented as a residual multilayer perceptron (MLP) with a shared global prediction and a small satellite-specific refinement. This design allows for a lightweight and fast inference model suitable for operational deployment and it allows minor mission-dependent differences to be captured without encouraging the network to learn satellite-specific bias as its primary signal.

4.3.1. Input Representation

The network input is a single feature vector $x(t) \in \mathbb{R}^d$, constructed from the processed dataset described in Chapter 3. For clarity, the inputs can be grouped into four categories:

- **Baseline density:** density values derived from the baseline model (JB2008) that provide the reference state for correction.
- **Solar and geomagnetic drivers:** indices and coupling functions, augmented with temporal embeddings (lags, moving averages and integration windows) to encode thermospheric memory effects.
- **Temporal and spatial information:** variables denoting the temporal and spatial state of the satellite at each specific time (e.g. latitude, local solar time).
- **Trust scores:** additional inputs describing the reliability of the IMF, solar wind and PCN data.

The exact feature list and the feature-specific normalisation procedure is defined in Table A.1. The reasoning behind the table is presented in Chapter 3.

In addition to $x(t)$, the model receives an integer satellite identifier $s \in \{0, 1, \dots\}$. Here, $s = 0$ is reserved for an unknown satellite identity, allowing the model to represent satellite-independent behaviour when the satellite identity is withheld during training using the dropout method. This method is explained in more detail in Section 4.4.

4.3.2. Architecture Overview

The network maps the input features $x(t)$ to a scalar prediction $\hat{y}(t)$, representing the normalised log-residual target defined in Section 4.2. Conceptually, the architecture consists of three steps. First, a feed-forward network shared by all satellites (the trunk, i.e. the main sequence of layers) transforms the input vector into a hidden representation $h(t)$ (an internal, learned feature vector that summarises the input information relevant for prediction). Second, a global output head (the final module that maps $h(t)$ to an output) maps $h(t)$ to a baseline prediction $\hat{y}_{\text{global}}(t)$, which is shared across all satellites. Third, a small satellite-conditioned refinement head produces an additional term $\hat{y}_{\text{sat}}(t)$, allowing minor mission-dependent deviations to be modelled when beneficial.

The final prediction is obtained through

$$\hat{y}(t) = \hat{y}_{\text{global}}(t) + \hat{y}_{\text{sat}}(t), \tag{4.5}$$

where $\hat{y}_{\text{sat}}(t)$ is explicitly constrained during training to remain a refinement rather than the dominant contribution as explained in Section 4.4.

Because thermospheric memory effects are encoded explicitly through the temporal feature embeddings introduced in Chapter 3 (lags, moving averages and integration windows), the model does not require a recurrent or sequence-based architecture to access historical context. A feed-forward trunk can therefore operate on each time sample independently, which is advantageous in the presence of data gaps. This design choice supports both operational robustness and fast inference, while retaining sensitivity to delayed thermospheric response through the engineered temporal features.

The following subsections describe the trunk, the satellite-specific refinement method and the regularisation mechanisms used in the forward pass.

Feed-Forward Network

The core of the model is a residual MLP trunk that maps the input feature vector $x(t)$ to a learned internal representation $h(t)$. Let d_{input} denote the input dimension (number of features), d_{trunk} the trunk dimension (hidden width) and K the number of residual blocks. The role of the trunk is to transform the raw engineered inputs into a compact set of internal features that are useful for predicting the target $\hat{y}(t)$.

An MLP can be viewed as a sequence of simple building blocks where each block applies a linear transformation (a weighted sum of the inputs) followed by a non-linear activation function. The linear layers allow the network to combine the input features, while the non-linear activations are what enable the model to represent non-additive relationships. This is important in the present setting because thermospheric density responds to solar and geomagnetic influences in a complex manner.

In the first step, the input is projected from $\mathbb{R}^{d_{input}}$ to the trunk dimension $\mathbb{R}^{d_{trunk}}$ using a linear layer followed by the SiLU activation function. The Sigmoid Linear Unit (SiLU) is a smooth non-linear activation defined as

$$\mathrm{SiLU}(z) = z \cdot \sigma(z) = \frac{z}{1 + e^{-z}}, \tag{4.6}$$

where $\sigma(z)$ denotes the sigmoid function. The initial hidden representation is then given by

$$h_0(t) = \mathrm{SiLU}(W_{\text{in}} x(t) + b_{\text{in}}), \tag{4.7}$$

where W_{in} and b_{in} are learned weights and biases. This projection creates the initial hidden representation $h_0(t)$, which is then refined through K residual blocks. Each residual block updates the hidden representation using the structure

$$h_{k+1}(t) = h_k(t) + f_k(\mathrm{LN}(h_k(t))), \tag{4.8}$$

where $f_k(\cdot)$ is a two-layer feed-forward mapping.

Here, $\mathrm{LN}(\cdot)$ denotes Layer Normalisation, which rescales the components of the hidden representation to have a more stable mean and variance within each sample before applying the block transformation.

In simplified form, for a hidden vector h, Layer Normalisation can be written as

$$\mathrm{LN}(h) = \gamma \odot \frac{h - \mu(h)}{\sqrt{\sigma^2(h) + \epsilon}} + \beta, \tag{4.9}$$

where $\mu(h)$ and $\sigma^2(h)$ are the mean and variance of the components of h, ϵ is a small constant for numerical stability and γ and β are learned scale and shift parameters.

The mapping $f_k(\cdot)$ uses SiLU activation and dropout. Dropout is a regularisation technique in which a random subset of intermediate activations is temporarily set to zero during training, which helps reduce overfitting and encourages the network to rely on distributed patterns rather than any single activation or feature. Here, the dropout probability within the trunk blocks is denoted by p_{drop} and corresponds to standard activation dropout applied to intermediate hidden activations during training (different from feature dropout p_{feat} applied to the input vector and satellite-ID dropout p_{sat} applied to the satellite identifier, described in Section 4.4.4).

In detail, each block first normalises the current representation, then applies a small MLP transformation and finally adds the result back to the original representation providing an updated representation in what is called a skip connection. It allows the block to learn a refinement to the current representation rather than having to replace it entirely.

The residual (skip) connection is included for optimisation stability. Rather than forcing each block to learn a completely new representation from scratch, the block learns a refinement to the current representation. This typically improves gradient flow during training and makes deeper networks easier to train without degradation in performance. Layer Normalisation further stabilises training by keeping the scale of intermediate activations well behaved.

After the final block, the trunk output is mapped to a scalar global prediction through a final LayerNorm and a linear layer:

$$\hat{y}_{\mathrm{global}}(t) = W_{\mathrm{out}}\, \mathrm{LN}(h_K(t)) + b_{\mathrm{out}}. \tag{4.10}$$

This produces a global normalised prediction, which is refined per satellite by the embedding layer described in the next subsection.

Satellite-Conditioned Refinement Pathway

To allow the model to capture small, mission-dependent deviations, a satellite-specific refinement is added. The satellite identifier s is mapped to a low-dimensional embedding vector $\vec{e_s}$, with dimension d_{emb}, via a learnable embedding table. The refinement head then operates on the concatenation of the trunk representation and the embedding:

$$\hat{y}_{\mathrm{sat}}(t) = g([\, h_K(t),\, e(s)\,])\,, \tag{4.11}$$

where $g(\cdot)$ is a small MLP (also using dropout with probability p_{drop}) that outputs a scalar satellite-specific correction term.

The final model prediction is the sum of the global and satellite-conditioned outputs as shown in Equation (4.5). This additive formulation is intentionally chosen so that $\hat{y}_{\mathrm{sat}}(t)$ acts only as a refinement term. During training, an explicit penalty is applied to discourage the satellite branch from dominating the prediction as explained in more detail in Section 4.4.

To encourage generalisation to satellites not seen during training, satellite identity is randomly withheld during training via satellite-ID dropout. This ensures with a fixed probability that the true satellite identifier s is set to $s = 0$, corresponding to the unknown satellite embedding. This exposes the model to both satellite-aware and global training conditions, supporting use for both satellites in the training set, as well as new satellites.

During training, feature and satellite-ID dropout are applied within the forward pass to improve robustness. Their role in training is explained in Section 4.4. The numerical values of these dropout probabilities, as well as trunk width, number of residual blocks, embedding dimension and refinement head size, are treated as hyperparameters and summarised in Section 4.5.

4.4. Training Procedure

This section describes the procedure used to train the neural network defined in Section 4.3. Details of data sources, preprocessing, feature construction and the train-validation split strategy are provided in Chapter 3. This section focuses on the training configuration, the loss function, the sampling strategy and the model selection procedure. The final values of the hyperparameters introduced in this section will be provided in Section 4.5.

4.4.1. Datasets and Normalisation Statistics

Training and validation are performed on pre-processed datasets constructed according to the split strategy defined in Chapter 3. The target used for training is the normalised log-residual $y(t)$ defined in Equation (4.2). The target normalisation parameters (satellite-specific means μ_s and global standard deviation σ_r) and the feature normalisation parameters are computed on the training data only and stored as fixed statistics for consistent application during training and validation.

To ensure the accelerometer-derived density data is consistent, a final check is added to remove any potential outliers. A floor value of 10^{-25} kg/m^3 is used for the density data as it was observed that some samples had a density values of 10^{-30} kg/m^3 and even negative.

4.4.2. Mini-Batches and Balanced Satellite Sampling

For clarity, the training data hierarchy used in this work is as follows. A sample corresponds to one training example at one time step, i.e. one row of the processed training dataset containing the input feature vector and its associated target value. A mini-batch consists of B_{size} samples processed together in one optimisation step. An epoch is defined here not as a full pass through all unique training samples, but as B_{epoch_count} mini-batches generated by the balanced satellite sampler. A full training run then consists of multiple epochs, with validation performed after each epoch to monitor progress and determine checkpoint selection and stopping.

Mini-batch training is performed using batches of size B_{size}. To avoid the training process being dominated by satellites with substantially larger sample counts, batches are constructed using a balanced satellite sampling strategy. Concretely, each batch element is formed by first selecting a satellite ID uniformly from the set of satellites present in the training dataset, followed by sampling a random row index from that satellite's pool of samples. This procedure is repeated B_{size} times per batch. This yields approximately uniform satellite exposure during training while retaining randomness at the sample level.

Because indices are sampled independently, this procedure implies sampling with replacement (each sample is returned to the population before the next selection), resulting in individual observations being drawn multiple times within a single epoch, while other observations may not be drawn in that epoch. As a consequence, an epoch is not defined as a strict single pass through all training samples. Instead, in this work an epoch is defined as a fixed number of mini-batches with approximately uniform satellite exposure. This provides a consistent unit for validation monitoring and checkpointing. The number of mini-batches per epoch B_{epoch_count} is selected such that each epoch provides sufficient training progress while keeping validation evaluations computationally manageable. The specific value of B_{epoch_count} is treated as a hyperparameter and reported in Section 4.5.

The total number of training epochs is determined by validation-based stopping, subject to a predefined maximum epoch count. In this way, training can terminate early once validation performance no longer improves, while still enforcing an upper bound on computational cost.

In the satellite ID convention used throughout this work, $s = 0$ is reserved for an unknown satellite identity and does not correspond to a physical satellite. Therefore, it does not appear as a label in the training dataset and is not sampled by the balanced batch sampler. Instead, $s = 0$ is only introduced during training through satellite ID dropout (see Section 4.4.4), which replaces the satellite ID of a fraction of samples by the unknown satellite $s = 0$.

4.4.3. Optimisation and Training Objective

Model parameters (the weights and biases of the neural network layers) are learned by minimising a scalar loss function over the training data. Training is performed with the AdamW optimiser, which is a gradient-based optimisation method that updates the model parameters in the direction that reduces the loss. The learning rate η controls the update step size, while the weight decay term λ_{wd} acts as an L_2-type regularisation that discourages excessively large weights and can improve generalisation.

The network is trained to predict the normalised target $\hat{y}(t)$ defined in Section 4.2. The training objective is constructed to guide the model to predict the correct residual magnitude and to ensure the model reproduces the correct temporal variability (i.e. the model correctly predicts the residual shape irrespective of small offsets). To obtain these requirements, a hybrid loss is used that combines a robust regression term and a correlation-based term.

Let $\mathcal{L}_{\mathrm{reg}}$ denote a SmoothL1 loss between $\hat{y}$ and y. SmoothL1 behaves similarly to mean squared error for small errors, but becomes closer to an absolute error loss for larger errors, reducing sensitivity to occasional outliers. In addition, let $\mathcal{L}_{\mathrm{corr}}$ denote a correlation loss that encourages $\hat{y}$ to follow the temporal variability (i.e. the shape) present in y by maximising the linear correlation between predictions and targets. The combined objective is written as

$$\mathcal{L}_{\mathrm{main}} = \psi\, \mathcal{L}_{\mathrm{reg}}(\hat{y}, y) + \gamma\, \mathcal{L}_{\mathrm{corr}}(\hat{y}, y), \tag{4.12}$$

where ψ and γ are scalar weights that control the relative emphasis on matching the pointwise residual magnitude versus matching the temporal variability.

In addition, an explicit penalty is applied to the satellite-specific refinement output $\hat{y}_{\mathrm{sat}}(t)$. As described in Section 4.3, the satellite refinement is intended to provide only a small correction on top of the global prediction. To discourage the refinement factor from dominating the total output, the following quadratic penalty is added:

$$\mathcal{L}_{\mathrm{sat}} = \lambda_{\mathrm{sat}}\, \mathbb{E}\left[\hat{y}_{\mathrm{sat}}(t)^2\right], \tag{4.13}$$

where λ_{sat} controls the strength of this constraint. The final loss optimised during training is therefore

$$\mathcal{L} = \mathcal{L}_{\mathrm{main}} + \mathcal{L}_{\mathrm{sat}}. \tag{4.14}$$

4.4.4. Training Regularisation

To improve generalisation and reduce large dependence on one specific feature, several regularisation mechanisms are applied during training. In this context, regularisation refers to techniques that deliberately make the learning problem slightly harder during training, so that the resulting model is less likely to rely on accidental patterns in the training data and is more robust when applied to new time periods or satellites.

First, feature dropout is applied to the input vector $x(t)$. With probability p_{feat}, individual input features are temporarily set to zero during a forward pass. The model is therefore forced to learn how to perform when various features are masked, rather than overfitting to a single strongly correlated feature. This prevents the model from relying heavily on one specific feature.

It is noted that this element-wise dropout is not intended to explicitly simulate realistic data outages. For example, an instantaneous driver like B_z may be masked while a derived lag or integration feature remains non-zero. Additionally, the associated trust score inputs are not modified during dropout. In this work, feature dropout is therefore treated as a generic regularisation mechanism that aims to improve robustness to feature degradation instead of an outage simulation. This choice is justified since the training dataset already contains periods with real solar and geomagnetic data outages. These gaps are incorporated in the data through the forward-filling procedure and trust score calculations described in Chapter 3, which exposes the model to operationally realistic data availability conditions during training.

Second, satellite ID dropout is applied to the satellite identifier s. With probability p_{sat}, the true satellite ID is replaced by $s = 0$, corresponding to an unknown satellite. This routes the sample through a global embedding and encourages the model to learn a shared mapping that generalises across different satellites. At the same time, when the satellite ID is not dropped, the model can still learn a small mission-dependent refinement as described in Section 4.3.2. This mechanism is therefore intended to balance satellite-specific improvements with generalisation to satellites not seen during training.

In addition to the dropout mechanisms described above, regularisation is also applied at the parameter level through weight decay in the AdamW optimiser. Weight decay adds a penalty that discourages excessively large network weights, which reduces the odds of overfitting. In practice, this acts as a constraint on model complexity and typically improves training stability and generalisation.

4.4.5. Checkpoint Selection During Training

Checkpointing is performed using the validation value of the main training objective, denoted $\mathcal{L}_{\mathrm{main,val}}$. This metric is an exact copy of the $\mathcal{L}_{\mathrm{main}}$ metric as described in Equation (4.12) and reflects the same two requirements used to guide optimisation during training being accurate prediction of the residual magnitude through the regression term and accurate reconstruction of the target shape through the correlation-based term. Checkpoint selection based on $\mathcal{L}_{\mathrm{main,val}}$ therefore ensures the model state that best balances these two objectives on unseen validation data is retained.

The model state is saved whenever $\mathcal{L}_{\mathrm{main,val}}$ improves, i.e. decreases. Training is terminated early if this validation objective does not improve for P consecutive epochs, where P is treated as a hyperparameter.

Although $\mathcal{L}_{\mathrm{main,val}}$ is used for checkpoint selection, complementary validation metrics such as per-satellite correlation, log-residual RMSE, density-space MAPE and tail statistics are also monitored and reported in the Stage A evaluation in Section 4.7.

4.5. Hyperparameter Selection

This section summarises the hyperparameter selection procedure used for the neural network architecture and training configuration described in sections 4.3–4.4. In this work, hyperparameters refer to user-defined settings that control the model capacity and the training dynamics (e.g. network width, number of residual blocks, dropout probabilities, learning rate, loss weights and the number of update steps per epoch). These values are not learned from data directly, but must be selected during model development.

4.5.1. Selection Approach

Hyperparameter selection is performed through manual, iterative tuning. This approach has been adopted primarily for computational practicality as an extensive sweep over architecture, loss-weighting and regularisation settings would have been computationally expensive given the dataset size. Instead, hyperparameters were adjusted manually based on observed training and validation behaviour and on the model's ability to reproduce physically plausible residual variability.

During development, each change has been evaluated on the validation set using the checkpoint score defined in Section 4.4.5, complemented by visual inspection of reconstructed density time series compared to accelerometer-derived density time series as explained in Section 4.5.2. This iterative process is sufficient to obtain a stable and performant configuration for the subsequent stage A and stage B evaluations.

4.5.2. Selection Criteria

The primary quantitative criterion used during tuning is the mean linear correlation between predicted and target log residuals over the validation satellites $\mathcal{S}_{\mathrm{val}}$, consistent with the checkpoint score defined in Section 4.4.5. In addition to the mean linear correlation, several metrics were added to gain insight into the model's performance. The full set of metrics used is explained in detail in Section 4.7. In addition to this scalar score, qualitative inspection was performed by visually comparing reconstructed density time series $\hat{\rho}_{\mathrm{NN}}(t)$ against the accelerometer-derived density reference $\rho_{\mathrm{obs}}(t)$. This visual inspection has been particularly informative for identifying configurations that produced predictions that were unrealistically smooth (insufficient variability capture) or excessively spiky (overly sensitive to noise or unstable training).

4.5.3. Reported Hyperparameters

Table 4.1 summarises the final set of hyperparameters used in this thesis. For clarity, they are grouped into architecture hyperparameters, optimisation, loss and objective parameters, regularisation, sampling, checkpointing and data hyperparameters. The batches per epoch B_{epoch_count} is denoted as $\lceil N_{\text{train}}/B_{size}\rceil$ where N_{train} is the total number of training samples.

Table 4.1: Final hyperparameters used for training the JB2008 correction model. Symbols follow the notation introduced throughout Chapter 4. The reported settings correspond to the final model and training configuration used in this work.

Group	Hyperparameter (symbol / code name)	Setting
Architecture		
	Input dimension d_{input}	73
	Number of satellites S	5 (IDs $0..4$, 0 unknown)
	Trunk width d_{trunk}	256
	Residual blocks K	4
	Satellite embedding dim d_{emb}	8
	Sat-head hidden width $d_{\text{sat-head}}$	64
	Block/head dropout p_{drop}	0.1
Optimisation		
	Optimiser	AdamW
	Learning rate η	1×10^{-4}
	Weight decay λ_{wd}	5×10^{-4}
	Number of epochs	30
Loss and objective (V2)		
	SmoothL1 parameter β_{SL1}	0.5
	Regression weight ψ	0.1
	Correlation weight γ	1.0
	Sat-head penalty λ_{sat}	3×10^{-3}
Regularisation (training-time)		
	Feature dropout p_{feat}	0.05
	Satellite-ID dropout p_{sat}	0.2
Sampling		
	Sampler strategy	Balanced satellite sampling
	Batch size B_{size}	1024
	Batches per epoch B_{epoch_count}	$\lceil N_{\text{train}}/B_{size} \rceil$ (computed at runtime)
Early stopping / checkpointing		
	Patience P	30
	Checkpoint metric	Validation objective $\mathcal{L}_{\mathrm{val}}$
Data / filtering and implementation details		
	Physics floor	1×10^{-25}
	Satellites	(CHAMP, GOCE, GRACE A/B)

4.6. Model Evaluation Pipeline

The performance of the proposed JB2008 correction model is assessed using a two-stage evaluation process. The motivation for this design is to separate model development and selection from the final independent operational validation. A purely accelerometer-derived density validation is relatively fast and enables efficient comparison of training epochs and model configurations, but it remains exposed to potential errors in the accelerometer-derived density reference data and does not directly quantify orbit propagation improvement. Therefore, a full orbit propagation validation against Precise Orbit Determination (POD) data, which is computationally expensive, provides an operationally meaningful final test.

Accordingly, stage A is defined as an accelerometer-derived density validation on the satellites used during training (CHAMP, GOCE, GRACE A/B). Stage A is used for checkpoint selection and hyperparameter tuning and provides a rapid assessment of whether the model captures the desired logarithmic residual behaviour. Stage B is defined as a POD-based orbit propagation validation on an out-of-sample satellite (Swarm-A). Stage B is not used for any training, checkpoint selection or hyperparameter tuning. Instead, it serves as a final, independent test to determine whether the learned density correction translates into reduced orbit propagation error in an operational setting.

In stage A, the neural network correction model is evaluated relative to the baseline JB2008 model. In Stage B, the corrected JB2008 density model is compared against both JB2008 and an additional empirical baseline, namely NRLMSISE-00, in order to provide a broader reference for operational orbit propagation performance.

4.7. Stage A - Accelerometer-Derived Density Validation

Stage A evaluates the neural network correction model using accelerometer-derived density references. The primary purpose of this stage is to provide a computationally efficient evaluation for comparing training epochs and hyperparameter configurations, allowing for fast and practical model development. The neural network prediction $\hat{y}(t)$ is mapped back to the log residual $\hat{r}(t)$ and reconstructed density $\hat{\rho}_{\mathrm{NN}}(t)$ using Equation (4.3)–Equation (4.4). These variables are then used to evaluate performance relative to the JB2008 baseline.

To assess both the ability of the model to track the variability of the correction signal and the magnitude of remaining errors, Stage A reports metrics in both log-residual space and density space. Let $r(t) = \log_{10}(\rho_{\mathrm{obs}}(t)/\rho_{\mathrm{JB2008}}(t))$ denote the target log residual and let $\hat{r}(t)$ denote its prediction by the neural network. For a set of N validation samples $\{r_i, \hat{r}_i\}_{i=1}^{N}$ (computed per validation satellite), the log residual correlation is evaluated using the Pearson correlation coefficient

$$\mathrm{corr}_{\hat{r},r} = \frac{\sum_{i=1}^{N}\left(\hat{r}_i - \overline{\hat{r}}\right)\left(r_i - \overline{r}\right)}{\sqrt{\sum_{i=1}^{N}\left(\hat{r}_i - \overline{\hat{r}}\right)^2}\sqrt{\sum_{i=1}^{N}\left(r_i - \overline{r}\right)^2}}, \tag{4.15}$$

where $\overline{\hat{r}} = \frac{1}{N}\sum_{i=1}^{N}\hat{r}_i$ and $\overline{r} = \frac{1}{N}\sum_{i=1}^{N} r_i$ denote the sample means of $\hat{r}$ and r, respectively. In addition to correlation, the root mean square error is evaluated using

$$\mathrm{RMSE}_{\hat{r},r} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{r}_i - r_i\right)^2}, \tag{4.16}$$

The correlation metric primarily captures whether the model reproduces the temporal variability of the observed correction, while the RMSE quantifies the typical correction magnitude mismatch and penalises large deviations. In addition, an interpretable density-space metric is reported using the mean absolute percentage error (MAPE) between reconstructed density $\hat{\rho}_{\mathrm{NN}}(t)$ and the accelerometer-derived density reference $\rho_{\mathrm{obs}}(t)$

$$\mathrm{MAPE}_{\rho} = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{\hat{\rho}_{\mathrm{NN},i} - \rho_{\mathrm{obs},i}}{\rho_{\mathrm{obs},i}}\right|. \tag{4.17}$$

Finally, the distribution of log-residual errors is analysed to characterise typical versus worst-case behaviour. Defining the log residual error at point i as $e_i = \hat{r}_i - r_i$, stage A reports the standard deviation $\mathrm{std}(e_r)$ and tail statistics such as the 95th and 99th percentiles of $|e_r|$. These distribution statistics provide insight into whether the model reduces extreme error events in addition to improving average performance.

While Stage A is informative for model development, it does not, by itself, prove that the correction model provides an improvement during orbit propagation over current industry standards. The accelerometer-derived density reference depends on satellite-specific calibration and aerodynamic modelling choices and improved accelerometer-derived density validation metrics do not by themselves guarantee improved orbit propagation performance. The ultimate objective of this thesis is not to reproduce the accelerometer-derived densities, but to reduce orbit propagation errors in an operational context. For this reason, stage B provides the final independent validation using POD orbit truth.

4.8. Stage B - Orbit Propagation Validation Against POD Data

Stage B provides the final operational validation by assessing the performance of the learned density correction in orbit propagation against POD orbit truth for Swarm A. This test is independent of the accelerometer-derived density training targets and directly evaluates whether the corrected density model reduces the in-track error during propagation. All orbit propagations in stage B are performed using the tudatpy astrodynamics framework. Tudat provides a consistent numerical environment to construct the force models, propagate the orbits and compute state deviations relative to the POD truth data under different atmospheric density configurations. Using a single propagation framework ensures that differences in performance between JB2008, JB2008+NN and NRLMSISE-00 arise from the atmospheric density modelling choices, rather than from inconsistencies in the implementation.

A limitation of this validation stage is that, although the density correction model is trained using data spanning a broader range of orbital altitudes, the orbit propagation validation presented here is restricted to Swarm A only and therefore to its specific orbital regime during the periods of interest. Stage B should therefore be interpreted as a proof of concept rather than as a complete validation across all altitude regimes represented in the training dataset. In particular, additional orbit propagation tests using lower altitude satellites would be required to assess whether the observed performance is representative across the full orbital regime covered during training.

4.8.1. Compared Density Models

Orbit propagation is performed under three atmospheric density configurations. First, JB2008 is used as the baseline empirical model. Second, JB2008+NN applies the trained neural network correction to the JB2008 density according to Equation (4.4). Third, NRLMSISE-00 is included as an additional empirical baseline for completeness (implemented through Tudat's NRLMSISE-00 atmosphere model).

4.8.2. Propagation Configuration

All propagations use the same force model and numerical integration settings for every density configuration to ensure a fair comparison. The adopted propagation configuration is summarised in Table 4.2. The aerodynamic drag coefficient C_D is treated separately and tuned per arc and per density configuration as described in Section 4.8.3.

Table 4.2: Propagation force model and numerical integration configuration used for all Stage B orbit propagations.

Component	**Configuration**
Spacecraft mass	Set per arc using an estimated Swarm-A mass.
Reference area	Constant reference area $A_{\mathrm{ref}} = 0.70\ m^2$ for all propagations.
Earth gravity	Spherical harmonics, degree/order $(90, 90)$.
Third-body gravity	Sun and Moon modelled as point-mass perturbations.
Solar radiation pressure	Cannonball SRP model with Earth occultation.
Atmospheric drag	Drag acceleration using selected density model.
Integrator	Variable-step Runge–Kutta–Fehlberg 7(8) (RKF78).
Initial step size	30 s.
Step-size bounds	1–100 s.
Error control	Scalar absolute and relative tolerance 10^{-11}.

In addition to the force model settings in Table 4.2, the spacecraft properties used in the drag model are kept consistent across all density configurations. The Swarm A mass is set per arc using a simple empirical mass model evaluated at the arc start epoch and then held fixed over the full tuning day and subsequent evaluation days within that arc. Specifically, the mass is modelled as a piecewise-linear function of time since launch, initialised at $473\ \mathrm{kg}$ at launch and decreasing over the mission to reflect propellant consumption, with a lower bound of $367\ \mathrm{kg}$.

In the implementation, the mass model is evaluated using $t_{\mathrm{yr}} =$ year $+$ month$/12$ and the following piecewise-linear schedule:

$$m(t_{\mathrm{yr}}) = \begin{cases} 473 - 14.3\,(t_{\mathrm{yr}} - 2013.9), & t_{\mathrm{yr}} < 2016.0, \\ 430 - 5.0\,(t_{\mathrm{yr}} - 2016.0), & 2016.0 \leq t_{\mathrm{yr}} < 2020.0, \\ \max(410 - 6.0\,(t_{\mathrm{yr}} - 2020.0),\ 367)\,, & t_{\mathrm{yr}} \geq 2020.0, \end{cases} \tag{4.18}$$

with m in kilograms. This provides a faster depletion of propellant during the initial stage of the mission, since several manoeuvres were performed. The stages after are assumed to consume less propellant.

The aerodynamic reference area A_{ref} was treated as constant value of $0.7\ \mathrm{m}^2$ for all propagations and was obtained from the published Swarm spacecraft ram surface specification (ESA, CNES, and CSA 2024). Holding m and A_{ref} fixed within each arc ensures that the model-specific C_d tuning procedure described in Section 4.8.3 remains effective over the entire arc. Any mismatches in the ballistic coefficient $m/(C_d A_{\mathrm{ref}})$ caused by offsets in the reference area or mass are absorbed by the tuned C_d, allowing the stage B comparison to focus on atmospheric model performance comparison.

Within tudatpy, NRLMSISE-00 can be used directly through the standard atmospheric environment interface. However, no equivalent built-in environment setup is available for JB2008. The neural network density correction model (JB2008+NN) is inherently user-defined. Therefore, for both JB2008 and JB2008+NN, Earth's atmosphere is implemented using tudatpy's custom four dimensional constant interface (`custom_four_dimensional_constant_temperature`), which allows specification of a custom density function $\rho = \rho(h, \lambda, \phi, t)$ as a callback evaluated during propagation. Concretely, tudat requests the density at the current spacecraft state through a function with signature $\rho(h, \lambda, \phi, t_{\mathrm{TDB}})$, where h is the geodetic altitude above the GRS80 reference ellipsoid (in meters), λ is the geodetic longitude and ϕ is the geodetic latitude (in radians). Finally, t_{TDB} is the current simulation epoch expressed as TDB seconds since J2000. In this work, the callback evaluates and returns either the baseline JB2008 density $\rho_{\mathrm{JB2008}}(h, \lambda, \phi, t)$ or the neural-network corrected density $\hat{\rho}_{\mathrm{NN}}(h, \lambda, \phi, t)$, depending on the selected propagation configuration.

For the JB2008 configuration, the callback computes ρ_{JB2008} using the pyatmos JB2008 implementation driven by the JB2008 space weather input files. For the JB2008+NN configuration, the same JB2008 density ρ_{JB2008} is first computed and then corrected by evaluating the trained neural network at the current epoch. The network input vector is assembled from the instantaneous spatial and temporal quantities (e.g. latitude, longitude), the instantaneous baseline density ρ_{JB2008} and the remaining precomputed solar wind, geomagnetic and trust score features matched to the current epoch. In addition, the argument of latitude embedding used by the neural network is computed at runtime from the current propagated spacecraft Cartesian state by conversion to Keplerian elements and evaluation of $u = \omega + \nu$, after which $\sin(u)$ and $\cos(u)$ are provided as inputs. This ensures that all geometry-dependent network inputs are derived from the propagated state and not from precomputed trajectory data. The resulting feature vector is normalised using the same processing logic as in training and stage A validation and then passed through the trained neural network to obtain a predicted normalised log residual correction. The corrected density is reconstructed as described in Section 4.2.4.

The `custom_four_dimensional_constant_temperature` interface additionally requires a constant atmospheric temperature, a specific gas constant and the ratio of specific heats. Since this study is only concerned with the drag acceleration magnitude given a specific aerodynamic model, these additional factors are assumed to have no influence and are therefore set at typical values. The temperature, gas constant and ratio of specific heats are therefore provided as fixed placeholder values ($T = 1000\,\mathrm{K}$, $R = 287\,\mathrm{J\,kg^{-1}\,K^{-1}}$ and $\gamma = 1.4$) to satisfy the environment interface inputs.

4.8.3. Arc Design and C_d Tuning Protocol

Seven POD arcs are selected to represent a range of operationally relevant solar regimes, including quiet conditions, storm-time responses and recovery phases under both solar minimum and solar maximum activity. For each arc, the day immediately preceding the arc start is included as a dedicated tuning day. This tuning day is used exclusively to tune the aerodynamic drag coefficient C_d for each of the three density models and each arc, after which the tuned C_d value is held fixed during the subsequent arc propagation evaluation for the respective density model. The selected arcs, their tuning days and the motivation for each arc are summarised in Table 4.3.

Table 4.3: Selected Swarm-A POD validation arcs for Stage B orbit-propagation testing. For each arc, the day immediately preceding the arc start is used as a model-specific C_d tuning interval, after which the tuned C_d is held fixed during the propagation evaluation. The duration column excludes the tuning day and is computed inclusively over the arc start and end dates.

Arc	Arc timespan	Duration (days)	C_d tuning day	Rationale
1	22/12/2018 – 03/01/2019	13	21/12/2018	Solar minimum regime with extended quiet periods and a storm on 28/12, allowing evaluation across quiet conditions, storm-time response and post-storm cooling.
2	07/05/2024 – 15/05/2024	9	06/05/2024	Solar maximum regime containing an exceptionally strong geomagnetic storm, with the subsequent recovery phase captured within the arc window.
3	14/03/2020 – 22/03/2020	9	13/03/2020	Solar minimum regime with a smaller storm near the vernal equinox.
4	19/06/2021 – 29/06/2021	11	18/06/2021	Rising activity as $F_{10.7}$ begins increasing, spanning the summer-solstice period.
5	31/01/2022 – 08/02/2022	9	30/01/2022	Moderate storm period associated with the widely reported Starlink re-entry event, including the recovery phase.
6	13/09/2022 – 22/09/2022	10	12/09/2022	Solar maximum regime with a storm and recovery, occurring near the autumnal equinox.
7	20/04/2023 – 30/04/2023	11	19/04/2023	Solar maximum regime with a severe storm and subsequent cooling, providing another disturbed test case.

The C_d tuning is formulated as a one-dimensional search that aims to remove mean along-track bias over the tuning day. Concretely, for a candidate C_d, a 24-hour propagation is performed over the tuning interval and the resulting along-track error time series is computed against the POD truth. The tuning objective is defined as the mean along-track error over the tuning interval,

$$f(C_d) = \frac{1}{N} \sum_{i=1}^{N} e_{\mathrm{AT}}(t_i, C_d), \tag{4.19}$$

and the tuned coefficient is chosen such that $f(C_d) \approx 0$, i.e. the propagation has near-zero mean along-track bias on the tuning day. This strategy is consistent with the intent of separating static offset effects (absorbed by C_d) from variability-driven error behaviour, which is evaluated in the subsequent validation propagations.

To efficiently solve for the root $f(C_d) = 0$, the tuning process employs the Secant method, a numerical root-finding algorithm. Given two initial estimates, the algorithm iteratively refines the drag coefficient using the recurrence relation:

$$C_{d,k+1} = C_{d,k} - f(C_{d,k}) \frac{C_{d,k} - C_{d,k-1}}{f(C_{d,k}) - f(C_{d,k-1})}, \tag{4.20}$$

where k denotes the iteration index. For this analysis, the algorithm is initialised with physically representative baseline guesses of $C_{d,0} = 2.2$ and $C_{d,1} = 2.5$. To ensure numerical stability within the orbit propagator, subsequent iteration updates are constrained to a range of $C_d \in [0.5, 10.0]$. The iterative procedure evaluates the signed mean along-track error of the 24-hour tuning propagation and terminates when the absolute mean error converges below a strict tolerance of 10 m.

The tuned C_d value should not be interpreted as a physically meaningful aerodynamic coefficient. In the present drag formulation, tuning C_d primarily compensates for scaling errors in the ballistic coefficient, including uncertainty in the reference area A_{ref}, uncertainty in the spacecraft mass m and any bias in the density of the selected atmospheric model over the tuning interval. As a result, the tuned C_d acts as an arc- and model-dependent calibration factor that removes the dominant mean along-track bias, thereby shifting the subsequent evaluation emphasis toward the time-varying (variability-driven) behaviour of the density model rather than its mean level.

A further limitation is that a single 24-hour tuning day is used to set C_d for the remaining evaluation sub-arcs within the arc. This is a simplification as the effective C_d required to match POD truth can drift over time due to changes in spacecraft operating conditions (e.g. attitude). Consequently, the tuned C_d is expected to become progressively less optimal as the propagation windows move farther from the tuning day. Despite this, the approach is retained because it reflects a practical operational workflow in which a single recent calibration interval is used to reduce mean bias before short-term predictions, while still allowing a fair comparison between atmospheric density configurations under identical force model and numerical settings.

Following tuning, each arc is evaluated through sequential 24-hour propagations (sub-arcs). Each 24-hour window is initialised from the POD truth state at the start of the window, propagated forward with the tuned C_D and compared against the POD truth over the same window. This daily re-initialisation ensures that the evaluation focuses on 24-hour prediction performance and prevents error accumulation over multi-day intervals from obscuring day-to-day model behaviour.

Along-track errors are computed by aligning the propagated states with the POD truth states and projecting the position error onto the along-track direction. Specifically, the POD truth states are first sorted by epoch and interpolated to the simulation epochs using a component-wise cubic interpolator. This interpolation is performed to obtain $\mathbf{r}_{\mathrm{POD}}(t)$ and $\mathbf{v}_{\mathrm{POD}}(t)$ at each simulation time t. The position deviation is then defined as $\Delta\mathbf{r}(t) = \mathbf{r}_{\mathrm{sim}}(t) - \mathbf{r}_{\mathrm{POD}}(t)$. The along-track unit vector is constructed from the POD velocity direction,

$$\hat{\mathbf{v}}_{\mathrm{POD}}(t) = \frac{\mathbf{v}_{\mathrm{POD}}(t)}{\|\mathbf{v}_{\mathrm{POD}}(t)\|}, \tag{4.21}$$

and the along-track error is obtained by the scalar projection

$$e_{\mathrm{AT}}(t) = \Delta\mathbf{r}(t) \cdot \hat{\mathbf{v}}_{\mathrm{POD}}(t). \tag{4.22}$$

This definition is used because, over short prediction horizons, errors in atmospheric drag are usually observed most clearly in the in-track (along-track) direction. Density errors mainly affect the orbital energy making the along-track component a sensitive indicator of performance of the drag model. In many cases, if the density is overestimated, the propagated satellite experiences too much drag, loses too much orbital energy, causing a negative radial error which in turn causes the propagated state to be ahead of the POD state in terms of in-track position. If the density is underestimated, the opposite occurs. The propagated satellite experiences too little drag and the propagated state will be behind the POD state in terms of in-track position. The sign of the resulting along-track error depends on the chosen sign convention and on how this phase difference develops relative to the POD trajectory. By contrast, radial and cross-track errors are more strongly affected by other modelling sources, such as gravity-field errors, solar radiation pressure, or attitude-related effects. For this reason, the along-track component is taken as the primary metric over 24-hour propagation windows.

4.8.4. Evaluation Metric

Stage B quantifies orbit-propagation performance using the along-track error time series $e_{\mathrm{AT}}(t)$ defined in Section 4.8.3. For each 24-hour propagation window, the simulator produces a discrete set of epochs $\{t_i\}_{i=1}^{N_w}$ at which the propagated state is compared against the POD truth where N_w denotes the number of valid comparison points within that window after POD interpolation and NaN filtering. Since a variable-step integrator is used, N_w can differ slightly between windows, but the metric definition remains identical.

The primary Stage B performance metric is the mean squared error (MSE) of the along-track propagation error over a given 24-hour window w,

$$\mathrm{MSE}_{\mathrm{AT}}^{(w)} = \frac{1}{N_w} \sum_{i=1}^{N_w} e_{\mathrm{AT}}(t_i)^2, \tag{4.23}$$

which directly reflects propagation-relevant drag modelling performance. For interpretability, results are also reported as the corresponding root-mean-square (RMS) error,

$$\mathrm{RMS}_{\mathrm{AT}}^{(w)} = \sqrt{\mathrm{MSE}_{\mathrm{AT}}^{(w)}}. \tag{4.24}$$

For each arc and density configuration, $\mathrm{MSE}_{\mathrm{AT}}^{(w)}$ (or equivalently $\mathrm{RMS}_{\mathrm{AT}}^{(w)}$) is computed for every 24-hour window w within that arc. The arc-level score is then obtained by averaging the window-level metrics across all windows in the arc,

$$\overline{\mathrm{MSE}}_{\mathrm{AT}}^{(\mathrm{arc})} = \frac{1}{W_{\mathrm{arc}}} \sum_{w=1}^{W_{\mathrm{arc}}} \mathrm{MSE}_{\mathrm{AT}}^{(w)}, \tag{4.25}$$

where W_{arc} is the number of 24-hour windows evaluated in that arc.

4.9. Methodology Summary

This chapter presented a correction-based neural network aimed at improving operational thermospheric density estimation for low-Earth orbit propagation. Rather than replacing established empirical models, the approach uses the JB2008 baseline with a neural-network correction layer that learns systematic, time-varying residual structure linked to solar and geomagnetic forcing. This design preserves the physical interpretability and computational efficiency of JB2008 while aiming to provide better model performance due to solar influences.

The correction model is trained in log-space using a JB2008 residual target. The raw target is defined as a base-10 log residual between accelerometer-derived density estimates and JB2008 density, $r(t) = \log_{10}(\rho_{\mathrm{obs}}/\rho_{\mathrm{JB2008}})$, which represents the multiplicative correction required to scale JB2008 to the observed density. To reduce sensitivity to satellite-dependent bias, the target is normalised by subtracting a satellite-specific mean and scaling by a global standard deviation (shared across all satellites) computed on training data only. During inference, the neural-network output is mapped back to log-residual space and converted to a physically meaningful corrected density via $\hat{\rho}_{\mathrm{NN}} = \rho_{\mathrm{JB2008}} 10^{\hat{r}}$.

The predictive model is implemented as a residual multilayer perceptron with a shared global trunk and a small satellite-conditioned refinement. The trunk transforms an engineered feature vector into a compact internal representation, which is mapped to a global prediction and thereafter refined through a low-dimensional satellite embedding. Satellite-ID dropout exposes the model to both satellite-aware and global conditions, supporting generalisation to satellites outside the training set. Feature dropout and weight decay provide additional regularisation. Training uses a balanced satellite sampling strategy to avoid dominance by satellites with large sample counts and the optimisation objective combines a regression term with a correlation-based term to prioritise reproduction of the residual shape while remaining stable to outliers.

Model performance is assessed through a two-stage evaluation pipeline designed to separate development validation from independent operational verification. Stage A provides a rapid validation using accelerometer-derived density references on the training/validation satellites. Metrics are reported in both log-residual space (correlation and RMSE) and density space (MAPE), enabling checkpoint selection and hyperparameter tuning.

Stage B provides the final operational validation by testing whether the learned density correction improves orbit propagation against Swarm A precise orbit determination (POD) truth data. Propagations are performed in a consistent tudatpy environment under three density configurations: JB2008, JB2008+NN and NRLMSISE-00. A common force model and numerical integration configuration is used across all cases to ensure fair comparison, while spacecraft properties are held constant within each arc. The Swarm A mass is estimated from a simple piecewise linear mission timeline model and fixed per arc and the reference area is set to $A_{\mathrm{ref}} = 0.70\ \mathrm{m}^2$ based on published Swarm ram surface specifications. For JB2008 and JB2008+NN, custom tudatpy atmosphere callbacks are used to provide tudatpy with the required density as a function of altitude, longitude, latitude and epoch. For JB2008+NN the neural-network inputs are obtained from the geometric quantities, JB2008 density and the pre-computed solar and geomagnetic features matched to the current epoch.

Stage B uses seven POD arcs spanning quiet periods and geomagnetically disturbed conditions under both solar minimum and solar maximum regimes. For each arc, the day preceding the arc start is used as a dedicated tuning day to calibrate C_d per density configuration. The tuning objective targets near-zero mean along-track bias over the tuning day. It should be noted that the tuned C_d acts primarily as an effective calibration factor that absorbs scale offsets in the ballistic coefficient and density rather than a physically meaningful aerodynamic coefficient. The tuned C_d is then held fixed for the subsequent daily 24-hour evaluation windows within the arc, reflecting a practical operational workflow while acknowledging that the optimal effective C_d may drift with time.

Propagation error is quantified using the along-track error time series computed by interpolating POD truth states to the simulation epochs and projecting the position deviation onto the POD velocity direction. The primary stage B metric is the mean squared along-track error over each 24-hour window, with RMS reported for interpretability. Window-level metrics are converted to arc-level scores by averaging across daily windows within each arc. Together, stages A and B provide a complete evaluation framework that supports efficient model development using accelerometer-derived densities and independently validates operational performance through reduced orbit-propagation error against POD truth.

5

Results and Analysis

This chapter presents the results of the proposed thermospheric density neural network correction model and evaluates whether the learned correction provides a meaningful improvement over the empirical baselines. The analysis is divided into two complementary stages. Stage A, presented in Section 5.1, assesses the ability of the neural network to reconstruct accelerometer-derived density more accurately than JB2008 in density space. Stage B, presented in Section 5.2, examines whether these improvements translate into improved orbit propagation accuracy. The chapter concludes with an overall validation summary in Section 5.3.

5.1. Stage A - Accelerometer-Derived Density Validation Results

Stage A provides the first level of validation for the proposed correction framework by evaluating the neural network directly against accelerometer-derived density references. The purpose of this stage is not yet to assess full operational propagation performance, but rather to determine whether the model has learned a physically meaningful correction to the JB2008 baseline in density space. This makes stage A particularly suitable for analysing training and validation behaviour and identifying whether the learned correction captures relevant thermospheric variability before proceeding to the more computationally expensive and operationally focused propagation validation of stage B.

5.1.1. Evaluation Recap

As defined in Section 4.7, stage A serves as the development validation stage of this thesis. It is used for checkpoint selection, model comparison and hyperparameter tuning against accelerometer-derived density references, but it does not provide the final operational validation.

The neural network is trained to predict the normalised log-residual target, which represents the correction required to scale JB2008 to the accelerometer-derived density. During evaluation, the network output is mapped back to the log-residual space and subsequently reconstructed into a corrected density estimate. Performance is then assessed relative to both the target log-residual and the reconstructed density.

To capture both absolute error magnitude and the density shape resemblance, stage A reports metrics in two complementary spaces. In the log residual space, the Pearson correlation coefficient is used to evaluate how well the model reproduces the shape of the required correction, while the log-residual RMSE quantifies the typical error in correction magnitude. In density space, the reconstructed density is compared against the accelerometer-derived reference using the mean absolute percentage error (MAPE), which provides a more directly interpretable measure of physical density error. In addition, the distribution of the log residual errors is analysed through statistics such as the standard deviation and percentiles in order to assess whether the model reduces not only average error but also extreme deviations.

Together, these metrics provide a compact but informative view of stage A performance. Correlation indicates whether the model captures the shape of the necessary density correction, whereas RMSE, MAPE and the residual-distribution statistics quantify how accurately that correction is reproduced.

5.1.2. Training Behaviour and Model Selection

The training stability and generalisation capability of the neural network were continuously monitored using the validation set described in Section 4.7. Figure 5.1 illustrates the progression of the training loss, described in Equation (4.14), against the validation loss over the training epochs. The model demonstrates a rapid decrease in training loss during the early epochs. It then stabilises towards epoch 30. The validation loss is consistently higher than the training loss, which is expected when evaluating the model on unseen data. Crucially, the validation curve remains flat towards the latter half of the epochs, indicating that the model has successfully learned how to convert the input features into a useful correction factor without overfitting to the training set (this would be the case if the validation loss started increasing towards the latter half of the epochs).

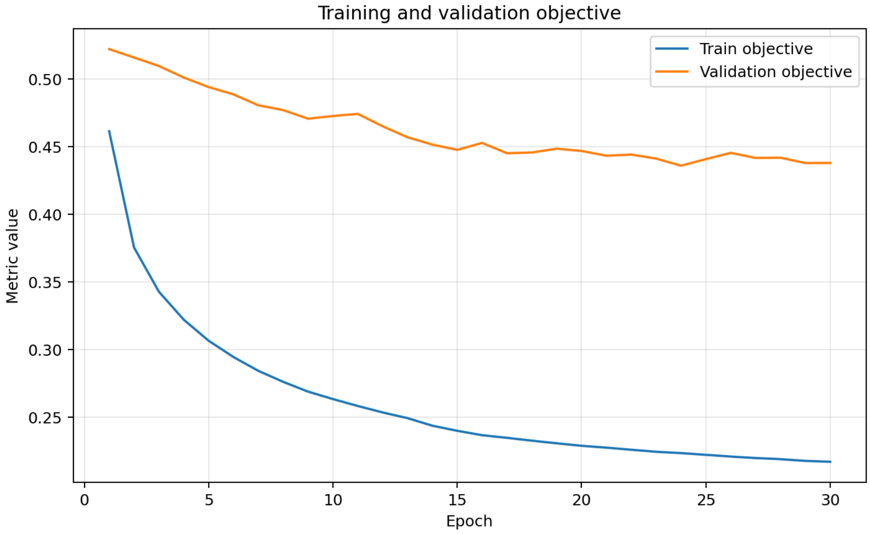


Figure 5.1: Progression of the training and validation objectives over 30 epochs. The training objective decreases rapidly, while the validation objective shows a smaller but consistent reduction before flattening. This behaviour is expected, since the model is optimised directly on the training set, whereas improvement on unseen validation data is typically more gradual. The absence of a strong increase towards the latter half of epochs in the validation curve indicates that the model continues to generalise without clear evidence of severe overfitting.

A critical component of the model architecture is the inclusion of both a global shared trunk and a satellite-specific correction. To verify that the model learned generalised physical drivers rather than memorising individual satellite behaviour, the relative magnitude of the satellite head's output was tracked during training. During the forward pass, the model isolates the outputs of the global physics trunk ($\hat{y}_{\mathrm{global}}$) and the satellite-specific embedding head ($\hat{y}_{\mathrm{sat}}$) as described in Equation (4.5). For each epoch, the satellite-head usage ratio is calculated by dividing the mean absolute magnitude of the satellite-specific corrections by the mean absolute magnitude of the global physics predictions as defined in Equation (5.1).

$$R_{\mathrm{sat}} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left|\hat{y}_{\mathrm{sat}}^{(i)}\right|}{\frac{1}{N}\sum_{i=1}^{N}\left|\hat{y}_{\mathrm{global}}^{(i)}\right|} \tag{5.1}$$

where N represents the total number of training samples evaluated during the epoch.

As shown in Figure 5.2, the network initially shows a heavier reliance on the satellite-specific corrections to minimise the loss, starting with a ratio of approximately 0.37. However, as training progresses, this ratio rapidly declines by nearly 40%, ultimately stabilising around 0.23 by epoch 30. To express this ratio relative to the combined prediction magnitude, the satellite-head share is computed as $R_{\mathrm{sat}}/(1+R_{\mathrm{sat}})$, since R_{sat} is defined relative to the global branch alone rather than the full output. In terms of total predictive share, this means the satellite head's contribution to the final predicted correction decreases from roughly 27% in early epochs down to less than 19% by the end of training. This shift suggests that the network transitions from using the satellite IDs as a crucial predictive factor to relying predominantly on the generalised physics trunk.

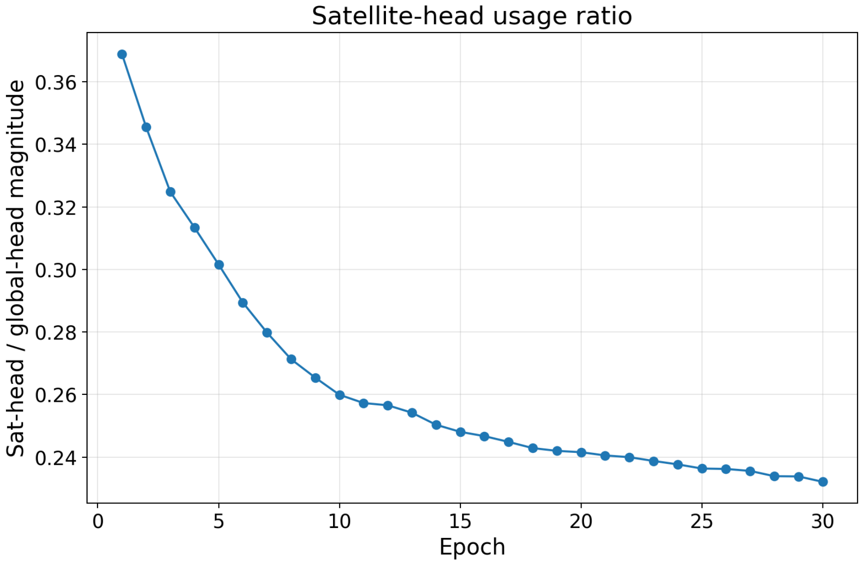


Figure 5.2: The ratio of the satellite-specific head magnitude to the global physics trunk magnitude over the training epochs. The decline and stabilisation of this ratio support the interpretation that the network learns to prioritise shared thermospheric physics over localised satellite memorisation.

5.1.3. Performance Breakdown by Satellite

The primary objective of Stage A validation is to quantify the model's performance improvement over the baseline JB2008 model across diverse orbital regimes. Figure 5.3 and Figure 5.4 present the mean absolute percentage error (MAPE) and root mean square error (RMSE), respectively, for each evaluated satellite.

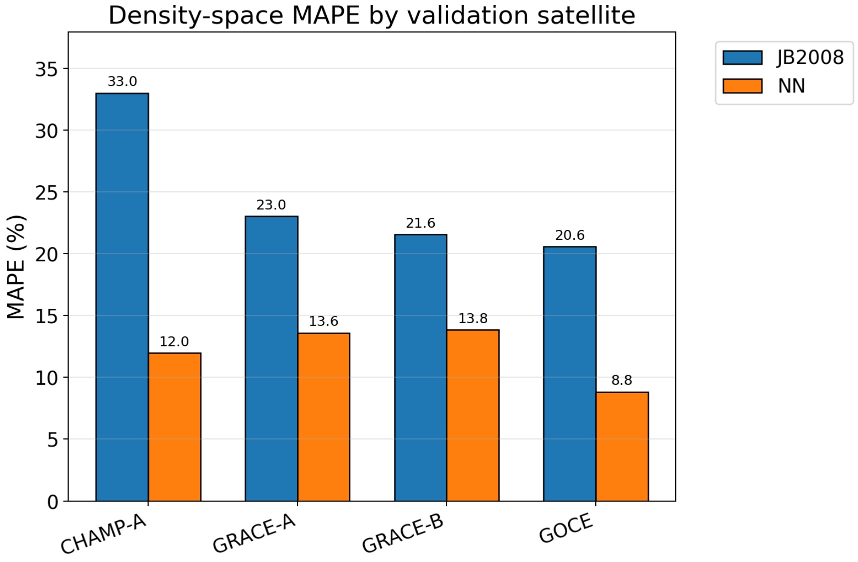


Figure 5.3: Density-space mean absolute percentage error (MAPE) for the stage A validation satellites. For all four satellites, the neural network correction model reduces the error relative to the baseline JB2008 model. The largest relative improvements are observed for CHAMP A and GOCE, while substantial reductions are also obtained for GRACE A and GRACE B.

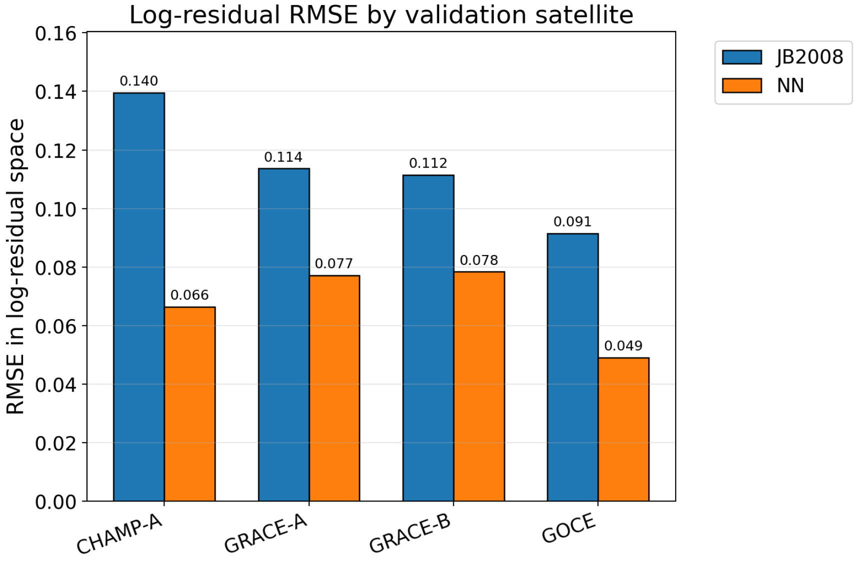


Figure 5.4: Log-residual-space root mean square error (RMSE) for the stage A validation satellites. The neural network correction model achieves lower RMSE than the baseline JB2008 model for every validation satellite, indicating that the learned correction consistently improves reconstruction accuracy across the evaluated orbital regimes.

When evaluating these absolute error metrics, it is important to address the difference between the JB2008 data and the neural network outputs. The baseline JB2008 MAPE and RMSE are computed using raw, uncalibrated empirical predictions while the neural network's outputs are unnormalised during inference using each satellite's specific historical mean and the global standard deviation. Because this mean-shift inherently removes static offsets, it could cause the neural network outputs to more closely align with the accelerometer-derived densities. Therefore, reductions in RMSE and MAPE alone do not definitively prove that the model has learned any meaningful corrections based on the underlying physics.

To demonstrate a true physical performance increase, the model can be evaluated using the Pearson correlation in the log-residual space. Because correlation is invariant to static mean shifts, it provides an explicit way of evaluating the model's ability to learn density effects caused by solar activity that JB2008 misses. The target variable of the neural network is the log residual, i.e. the exact correction required atop of JB2008 to match the accelerometer-derived density. Because the baseline JB2008 model cannot correct itself, its predicted residual correction is a flat line, yielding a correlation of exactly zero.

Instead, the neural network achieves a log-residual correlation of 0.65 to 0.70 as seen in Figure 5.5. This strong structural alignment suggests that the network is not simply applying a bias correction. This result suggests that the model is able to capture the dynamic density fluctuations that JB2008 misses. As seen in Figure 5.3 and Figure 5.4, this combination of static bias correction and improved correlation results in substantial error reductions across all altitudes, from GOCE's extremely low-Earth orbit to the higher altitudes of the GRACE and CHAMP missions.

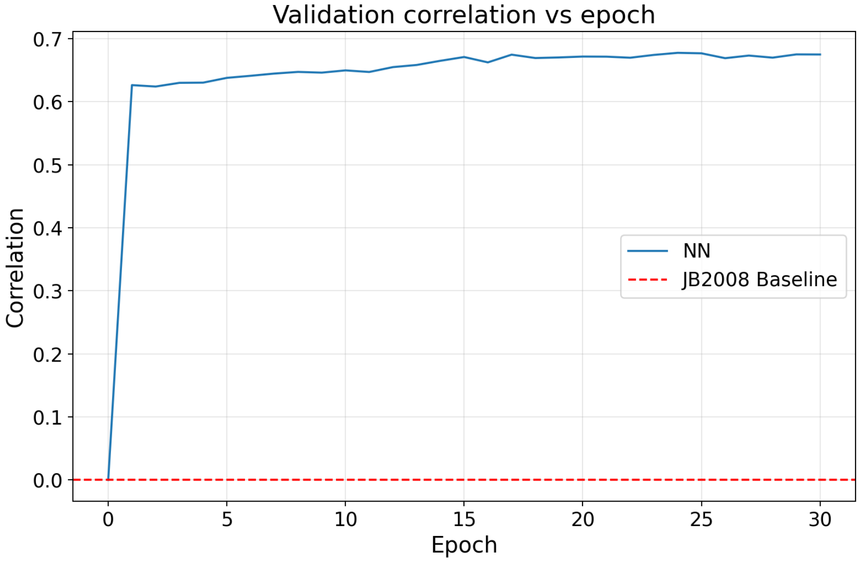


Figure 5.5: Validation correlation in log-residual space as a function of training epoch. The neural network rapidly reaches a correlation above 0.6 and gradually improves toward approximately 0.67, while the JB2008 baseline remains at zero because it provides no learned residual correction and therefore corresponds to a flat prediction in residual space. The stable plateau at later epochs indicates that the model captures the temporal structure of the residuals without evidence of strong degradation in validation performance.

To further visualise the performance, Figure 5.6 illustrates the distribution of the log residual errors across the validation set. The empirical JB2008 baseline exhibits a wider variance and a slightly positive mean. In contrast, the neural network's residual distribution is centred closer to zero while also showing a reduction in variance causing a more needle-like distribution.

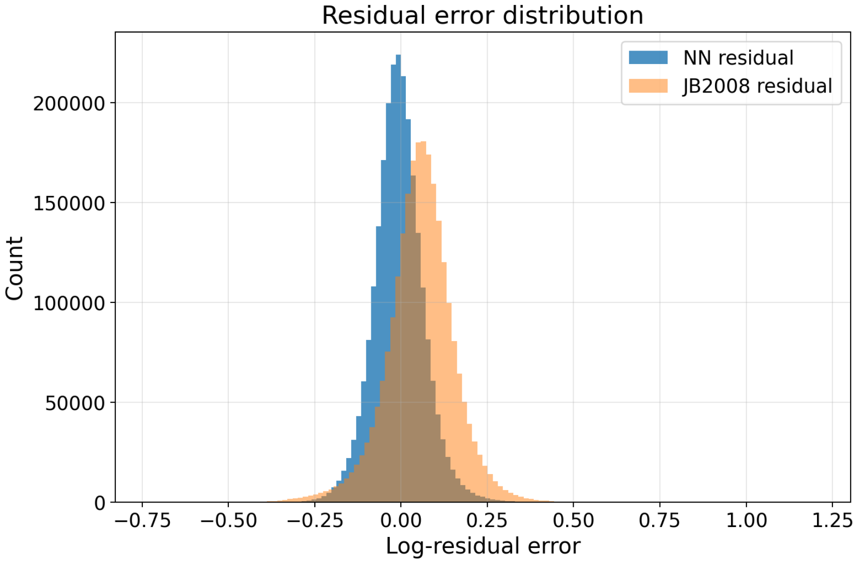


Figure 5.6: Distribution of log-residual errors on the validation set for the baseline JB2008 model and the neural network correction model. Relative to JB2008, the neural network error distribution is more tightly concentrated, shifted closer to zero and exhibits a reduced tail width. This indicates that the learned correction reduces both the systematic offset and the spread of the residual errors, including a reduction in the frequency of larger errors.

To quantify the visual improvements observed in the error distributions, Table 5.1 summarises the statistical properties of the log residual errors. The reduction in mean bias is shown by the mean shift from 0.05618 to a near-zero value of -0.00647, while the overall standard deviation decreases by approximately 25% in the logarithmic space.

For operational orbit propagation, large residual errors are most likely to occur during strongly disturbed conditions where empirical models tend to degrade. The 95th and 99th absolute percentiles therefore provide a useful summary of the performance regarding these heavily disturbed times. As shown in Table 5.1, the neural network reduces both percentiles, with the 99th percentile decreasing from 0.31472 to 0.21425. This indicates that the learned correction not only reduces the average spread of the residuals, but also improves behaviour in the higher-error portion of the validation distribution.

Table 5.1: Statistical evaluation of log residual errors for the validation set. The neural network demonstrates a near-zero mean bias and a significant reduction in extreme prediction errors compared to the JB2008 baseline.

Metric	JB2008 Baseline	Neural Network
Mean (Bias)	0.05618	-0.00647
Standard Deviation (Spread)	0.09848	0.07298
95th Percentile (Absolute)	0.22301	0.14696
99th Percentile (Absolute)	0.31472	0.21425

The impact of these statistical improvements is also visible when examining the reconstructed density time series. Figure 5.7 presents a six-hour example for the GRACE-A satellite and helps illustrate why the neural network achieves better global validation statistics than the JB2008 baseline. The baseline JB2008 solution exhibits both phase errors and amplitude mismatches relative to the target accelerometer-derived density. For example, near the 01:00 and 04:00 timestamps, JB2008 overestimates the amplitude of the density peaks and does not align as well with the timing of the observed variations. By contrast, the neural network correction shifts the reconstruction closer to the target in both phase and magnitude. It also follows smaller secondary structures more closely, while remaining smoother than the raw target signal, showing that the model captures relevant physical variability without simply reproducing short-timescale noise.

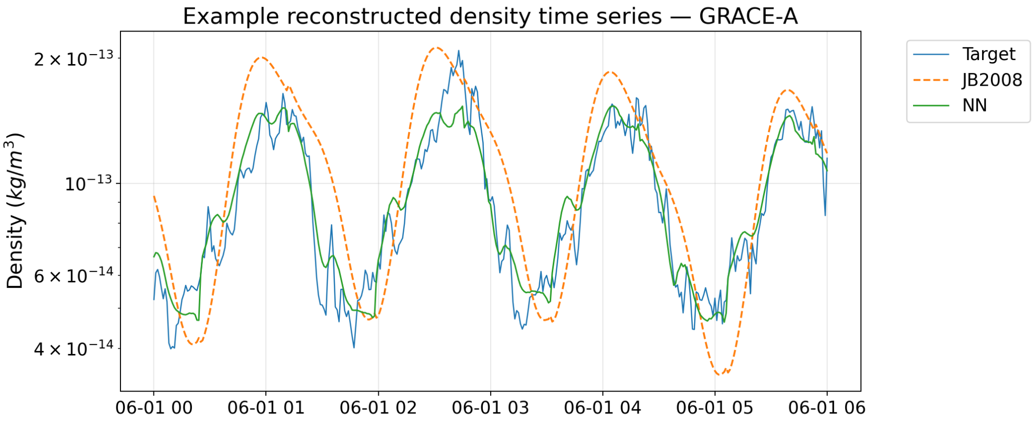


Figure 5.7: A six-hour reconstructed density time series starting on 2008-06-01 00:00:00 UTC for the GRACE-A satellite. The neural network (green) provides corrections to the JB2008 baseline (orange) showcasing its ability to better capture the structural intricacies present in the target accelerometer-derived density (blue) when compared with the baseline.

The validation behaviour is also consistent with the leakage-resistant data design described in Section 3.6. Temporal leakage is mitigated by the use of explicit train-validation time blocks applied consistently across satellites, together with the handling of temporal lags, moving averages and other temporal features using only past information. In this setting, the validation results are therefore evaluated on genuinely unseen time periods. The observed performance is consistent with this design. The network improves substantially over JB2008, but still exhibits a clear gap between training and validation performance and does not reproduce the target signal with unrealistic precision. This supports the interpretation that the model has learned a meaningful JB2008 correction from the input features, rather than benefiting from look-ahead information.

In summary, the stage A validation results indicate that the neural network provides a meaningful improvement over the JB2008 baseline when compared directly against accelerometer-derived densities. It is able to significantly eliminate systemic biases, significantly reduces extreme errors during volatile periods and dynamically corrects complex sub-orbital phase and amplitude inaccuracies. However, achieving higher accuracy when comparing to accelerometer-derived densities is not the end goal. The primary operational objective of this model is to translate these refined density predictions into meaningful improvements for satellite orbit propagation. To evaluate this end goal, the model must be tested in a propagation environment that mimics real-time orbit propagation. This is the focus of the stage B orbit propagation evaluation detailed in the following section.

5.2. Stage B - Orbit Propagation Validation Results

While Stage A demonstrated the neural network's ability to more accurately reconstruct accelerometer-derived thermospheric density compared to JB2008, the ultimate test of an atmospheric model is its performance during orbit propagation. Stage B evaluates the along-track error of 24-hour propagations across multiple distinct arcs, comparing the neural network against the empirical NRLMSISE-00 and JB2008 models.

5.2.1. Overall Propagation Accuracy

To establish the overall performance of the models, the 24-hour root mean square (RMS) along-track errors were aggregated across all evaluation arcs. As shown in Figure 5.8, the neural network successfully reduced the overall mean along-track error to 87 m. This represents a nearly 19% reduction compared to the primary operational baseline, JB2008 (107 m) and a vast improvement over NRLMSISE-00 (196 m).

Furthermore, Figure 5.9 illustrates the distribution of these daily RMS errors. While NRLMSISE-00 exhibits a massive error variance, the neural network and JB2008 display similar relative error spreads. However, the neural network's median and the interquartile bounds are shifted downward. Crucially, the neural network also reduces the ceiling of the extreme outliers, demonstrating a higher degree of predictive consistency and a safer worst-case operational bound.

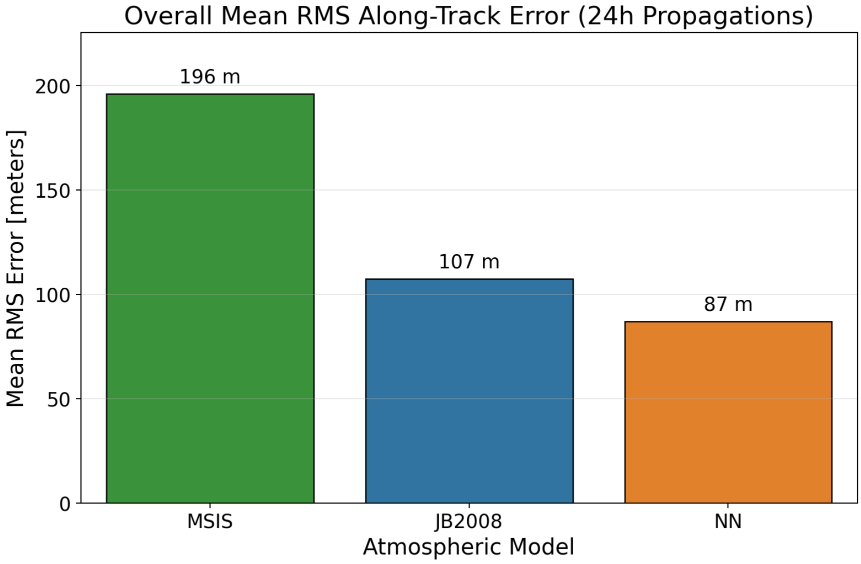


Figure 5.8: Overall mean RMS along-track error across all 24-hour propagation windows.

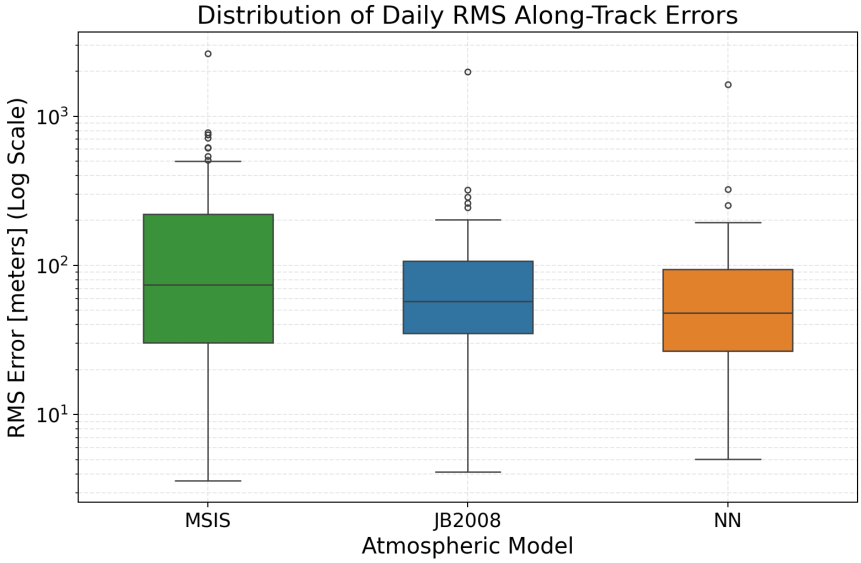


Figure 5.9: Distribution of daily RMS along-track errors (log scale), demonstrating the neural network's tighter interquartile range and reduced outlier ceiling.

5.2.2. Temporal Consistency and Storm Resilience

A main goal for the neural network density model is the ability to improve accuracy during severe geomagnetic storms. Figure 5.10 breaks down the mean RMS error by evaluation arc. The neural network exhibits high temporal stability, remaining consistently on the lower end across all arcs, whereas JB2008 and NRLMSISE-00 experience severe fluctuations during highly active solar periods (arcs 2, 5, 6 and 7).

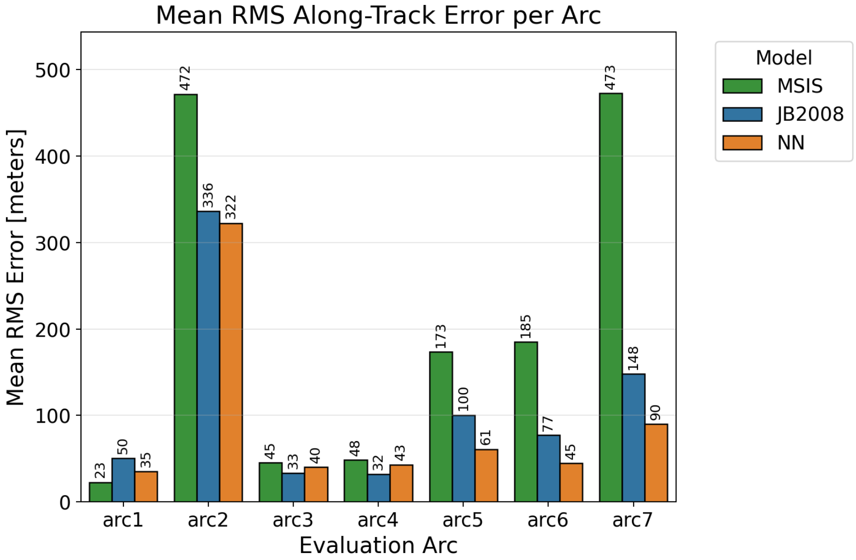


Figure 5.10: Mean RMS along-track error per evaluation arc, highlighting the temporal stability of the Neural Network across varying space weather conditions.

When categorising the evaluation arcs by their respective space weather conditions, an interesting pattern can be observed. During geomagnetically quiet periods, represented by arcs 1, 3 and 4 according to Table 4.3, the empirical models perform adequately. In these stable environments, the predictive accuracy of all three models is relatively comparable, as the baseline thermospheric dynamics are captured by traditional empirical formulations. However, the neural network demonstrates its operational value during periods of elevated solar activity. Arc 2 captures an unusually severe geomagnetic storm. Under these extreme conditions, the neural network provides only a minor accuracy advantage over the significantly degraded baselines. The correction model doesn't provide a large improvement over the empirical models since the neural network uses the severely degraded JB2008 estimates as the baseline. However, during arcs 5, 6 and 7, which represent moderate to severe geomagnetic disturbances, the neural network yields a substantial predictive improvement. In these highly active scenarios, the network suppresses the severe error fluctuations that compromise the reliability of both JB2008 and NRLMSISE-00.

This resilience is further illustrated by examining the daily error evolution during specific periods of increased solar activity. During the severe geomagnetic event captured in arc 2 (Figure 5.11), the empirical baselines experience a severe degradation in predictive accuracy. At the peak of this storm on day 4, the NRLMSISE-00 and JB2008 along-track RMS errors reach approximately 2600 m and 2000 m, respectively. The neural network, however, reduces the maximum error to approximately 1600 m, indicating a slight improved capacity to model extreme thermospheric disturbances. This stability is not limited to isolated extremes as it is also present during moderate storm events, such as arc 5 (Figure 5.12). Under these actively disturbed conditions, the neural network consistently provides improved predictions compared to the baseline empirical models.

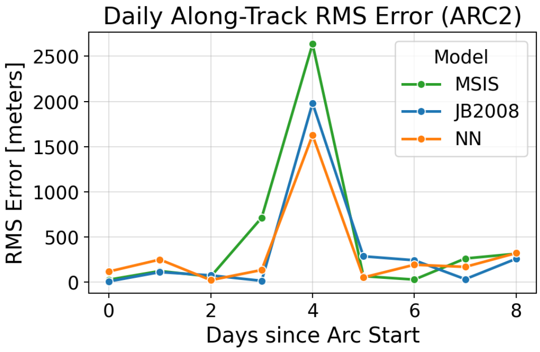


Figure 5.11: Daily along-track RMS error evolution during arc 2, starting on 07/05/2024 and covering the exceptionally severe geomagnetic storm of May 2024. All three models show a strong degradation on the most disturbed day, but the neural-network-corrected model retains the lowest RMS error among the compared configurations.

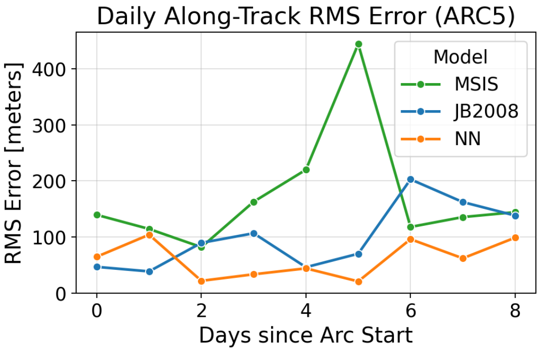


Figure 5.12: Daily along-track RMS error evolution during arc 5, starting on 31/01/2022 and covering the moderate storm period associated with the February 2022 Starlink event. The neural network correction model generally maintains lower daily RMS error than the empirical baselines throughout the arc.

5.2.3. Error Characterisation

While the RMS error provides a useful overall performance metric, it does not distinguish between systematic offset and day-to-day variability. For example, a high RMS error may arise because a model produces a persistent along-track bias, because it fails to capture the time-varying response to changing space weather conditions, or because both effects are present simultaneously. To better characterise the source of the propagation error, the daily along-track results were decomposed into their mean value (bias) and standard deviation (spread), as shown in Figure 5.13.

The figure shows that the empirical models, particularly NRLMSISE-00, exhibit a broader distribution in both daily bias and daily spread. JB2008 performs more favourably, but still shows a wider spread than the neural-network-corrected model. By contrast, the JB2008 correction model results are more tightly clustered around zero bias and generally achieve lower daily spread. This indicates that the neural network not only reduces systematic along-track offset, but also improves the consistency with which short-term propagation errors are controlled across different daily conditions. A few large outliers remain for all models, particularly during strongly disturbed periods, but the neural network correction model shows the most compact overall error distribution.

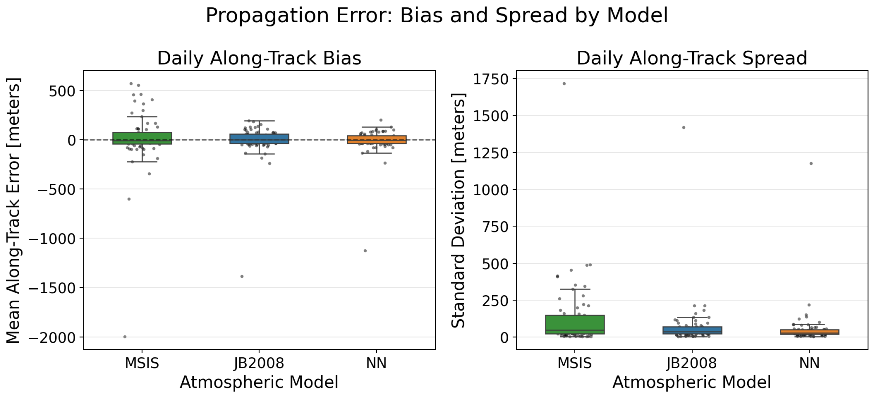


Figure 5.13: Distribution of daily along-track bias and spread for the three atmospheric density models across all evaluated 24-hour propagation windows. The left panel shows the daily mean along-track error, while the right panel shows the daily standard deviation of the along-track error within each daily window. Under the sign convention of Equation (4.22), a negative along-track bias indicates that the propagated state is behind the POD reference in the along-track direction, while a positive bias indicates that it is ahead. In this convention, storm-time density underestimation is generally expected to produce a negative along-track bias because the propagated satellite experiences too little drag and therefore tends to lag the POD trajectory, whereas density overestimation tends to produce the opposite behaviour. The empirical models exhibit broader bias and spread distributions, especially during disturbed conditions, while the JB2008 correction model remains more tightly clustered near zero bias and lower spread.

Validity of Ballistic Coefficients

In orbit determination, the ballistic coefficient (C_d) is often estimated as a tuneable parameter to absorb unmodeled aerodynamic offsets and structural biases in the density models. However, while it acts as a mathematical sink for these errors, it must still remain within a physically justifiable regime. If a density model generates highly erroneous or unphysical outputs, the optimiser could force C_d to extreme values (e.g. < 1.0 or > 10.0).

For a satellite in low Earth orbit with the geometry of Swarm A, realistic tuned C_d values are expected to roughly reside between 1 and 4. As shown in Figure 5.14, the C_d values obtained during tuning for the neural network are constrained within this regime. They align closely with the empirical baselines, confirming that the network's orbit prediction improvements are the result of structurally sound improvements, rather than the optimiser exploiting the drag coefficient tuning to compensate for poor density estimations.

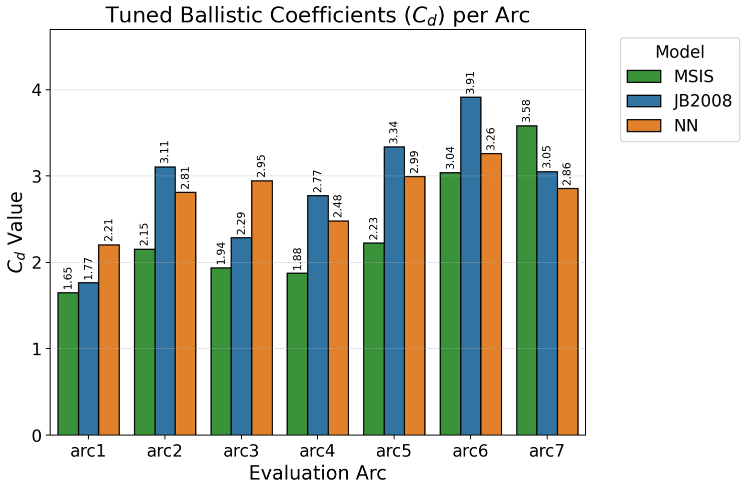


Figure 5.14: Tuned ballistic coefficients (C_d) per arc. The neural network maintains strictly physical values consistent with established empirical models, validating the integrity of its density predictions.

5.2.4. Operational Reliability

To translate the error metrics into actionable data for satellite operators, an empirical cumulative distribution function (ECDF) of the RMS errors is presented in Figure 5.15. This plot defines the probability that a model will maintain its 24-hour prediction RMS error below a given safety threshold.

The neural network consistently improves over the empirical baselines across all thresholds. For example, assuming a strict operational RMS error tolerance of 200 meters, the neural network satisfies this requirement on approximately 95% of the evaluated days. In comparison, JB2008 meets this threshold on 90% of days and NRLMSISE-00 on only 75%. The convergence of the neural network's curve toward 1.0 (100%) at lower error thresholds suggests it is a safer and more stable operational model.

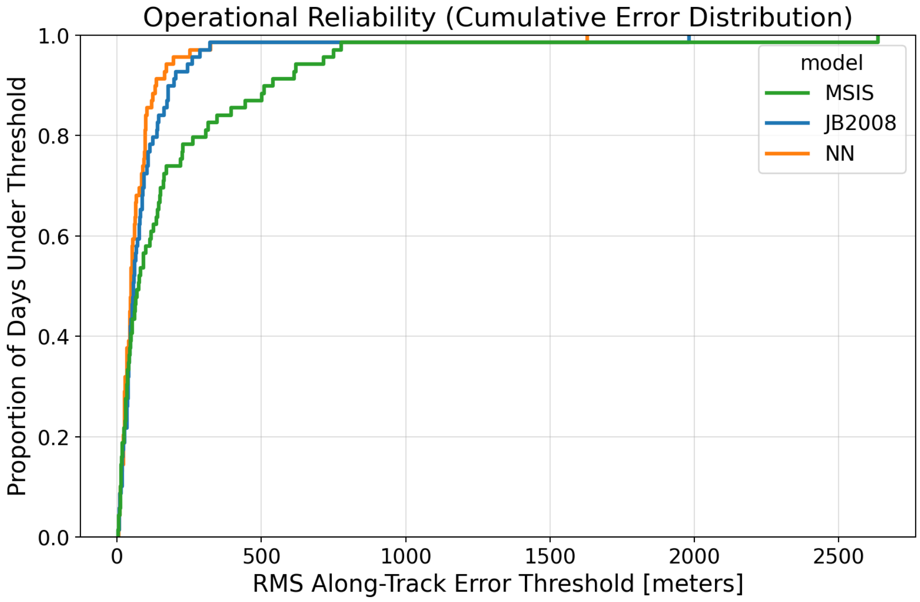


Figure 5.15: Empirical Cumulative Distribution Function (ECDF) of the RMS errors, illustrating the superior operational reliability of the Neural Network across all safety thresholds.

5.3. Validation Overview

The results of this chapter suggest that the proposed neural network correction model provides a consistent and operationally meaningful improvement over the empirical baselines. In stage A, the model demonstrated the ability to reduce the error of the reconstructed density relative to accelerometer-derived observations across all validation satellites. The improvements were not limited to a reduction in absolute error metrics such as MAPE and RMSE, but were also reflected in the correlation measures, where the neural network achieved a strong correlation with the target log residuals. Additionally, the neural network reduced the variance of the log residual distribution and suppressed extreme outliers. These findings indicate that the model learned more than a simple static offset correction and was able to capture part of the dynamic thermospheric variability that is not represented by JB2008 alone.

More importantly, stage B showed that the stage A improvements translated into improved orbit propagation performance, which is the primary objective of this thesis. Across the full set of 24-hour propagation arcs, the neural network achieved the lowest overall mean RMS along-track error and delivered the most reliable behaviour across a wide range of operating conditions. The largest gains were observed during moderate and severe geomagnetic disturbances, where the empirical baselines exhibited substantial degradation and the neural network was able to maintain lower errors. At the same time, the tuned ballistic coefficients remained within a physically plausible range, indicating that the performance gain was not simply the result of an unrealistic compensation through the drag parameter tuning process.

Overall, the combined stage A and stage B results suggest that this neural network density correction model, trained on accelerometer-derived density residuals and injected with solar wind, geomagnetic and orbital context, can improve upon traditional empirical thermospheric density models in both physical reconstruction and operational orbit prediction.

6

Conclusions and Limitations

This thesis investigated whether a neural network-based correction to the JB2008 empirical density model can improve atmospheric density estimation and, more importantly, whether such improvements translate into better orbit propagation performance. The work was motivated by the need for thermospheric density models that are not only more accurate under varying space weather conditions, but also sufficiently lightweight and operationally practical for repeated use in low Earth orbit propagation workflows.

To address this problem, a hybrid correction framework was developed in which JB2008 provides the baseline density estimate and a neural network learns a correction using solar, geomagnetic, spatial and temporal input features. Rather than relying on a recurrent sequence model, the proposed method uses a comparatively simple feed-forward neural network architecture combined with physically motivated temporal feature engineering. In addition, trust score features were introduced to account for outages in the high-cadence space weather inputs. Together, these design choices aimed to preserve operational simplicity while still allowing the model to represent delayed and non-linear thermospheric responses to solar and geomagnetic forcing.

6.1. Conclusions

The results of stage A indicate that this correction-based approach provides a meaningful improvement over JB2008 when evaluated directly against accelerometer-derived density data. In particular, the neural network reduced the mean and spread of the log-residual error distribution and produced a reconstructed density signal that more closely followed the target density than the baseline model. The example reconstructed time series further showed that the neural network improves both phase alignment and amplitude representation relative to JB2008, rather than solely applying a static shift. These results suggest that the model learned part of the time-varying density variability that is not fully captured by the empirical baseline.

The results of stage B provide further evidence that these improvements are not limited to density reconstruction alone. In the POD-based orbit propagation validation, the neural network correction model achieved the best overall along-track propagation performance among the tested density configurations. The most convincing improvements were observed during disturbed periods, particularly in arcs 5, 6 and 7, where the neural network consistently outperformed the empirical baselines. This indicates that the learned correction may be especially beneficial under geomagnetically active conditions, which are also the conditions under which empirical models tend to degrade most strongly. In this sense, the proposed framework shows clear promise as a practical improvement layer on top of an existing empirical density model.

Overall, the results of this thesis indicate that a lightweight neural network correction to JB2008 can improve both thermospheric density reconstruction and short-term orbit propagation performance, with the largest improvements occurring during disturbed space weather conditions. They further suggest that a simple feed-forward architecture, when combined with physically motivated temporal embeddings and trust score features, is already capable of learning useful corrections to an empirical baseline.

6.2. Answers to the Research Questions

The findings discussed above can now be related directly to the research questions formulated in Chapter 2. Since the thesis was structured around two central questions, one concerning density reconstruction and one concerning orbit propagation performance, it is useful to state their answers explicitly. The following summary therefore links the main results of stage A and stage B back to the original research questions.

Answer to Research Question 1 (RQ1)
The results of stage A indicate that the answer to RQ1 is positive. The neural network-based correction model provides more accurate reconstructions of accelerometer-derived thermospheric density than the JB2008 baseline across all four validation satellites. The improvements are reflected not only in reduced absolute error metrics, but also in the reduced mean and spread of the log-residual error distribution and in the improved reconstruction of the density time series. These findings indicate that the model learns a meaningful part of the thermospheric variability that is not captured by JB2008 alone.

Answer to Research Question 2 (RQ2)
The results of stage B indicate that the answer to RQ2 is also positive, but it should be interpreted more carefully. The neural network correction model achieved the best overall 24-hour along-track propagation performance among the tested density configurations and showed its clearest advantages during disturbed periods, particularly in arcs 5, 6 and 7. However, since the POD-based validation was restricted to Swarm A and a limited set of validation arcs, this result should be interpreted as promising proof of concept rather than as definitive evidence of general operational superiority across all low Earth orbit regimes.

6.3. Interpretation and Limitations

At the same time, the implications of these findings must be interpreted carefully. The final orbit propagation validation was performed only on Swarm A and over a limited number of validation arcs. Although these arcs were specifically selected to include quiet, moderate and severe geomagnetic conditions, they do not provide a comprehensive validation across the full range of orbital regimes represented in the training dataset. In particular, the stage B analysis does not prove that the observed performance improvements will necessarily generalise to lower altitude satellites. The results should therefore be interpreted as a promising proof of concept rather than as definitive evidence of broad operational superiority across all relevant LEO regimes.

Taken together, the results of this thesis provide encouraging evidence that a neural network-based correction to an empirical density model can improve thermospheric density modelling and short-term orbit propagation performance. At the same time, the limited scope of the POD-based validation means that these findings should be interpreted as a promising proof of concept rather than as definitive evidence of broad operational superiority across all low Earth orbit regimes. Broader validation across additional satellites, altitude regimes and operational scenarios is therefore still required before the method can be regarded as an improvement over the current baseline density models. Within those limits, the proposed hybrid framework represents a promising direction for future thermospheric density modelling research and development. The next chapter builds on these conclusions and limitations by outlining recommendations for future work.

7

Recommendations

Based on the findings and limitations of this thesis, several recommendations can be formulated for future work. These recommendations aim to strengthen the evidence for the proposed correction framework, improve the quality of the training data and further explore the design choices made in this work. The recommendations are grouped into three themes: validation and operational testing, data quality and robustness and model development and benchmarking. Within each theme, the recommendations are ordered by practical importance.

7.1. Validation and Operational Testing

The most immediate priority for future work is to strengthen the operational evidence for the proposed correction framework. Although the current stage B results are promising, the validation remains limited in both orbital coverage and total evaluation duration. The following recommendations therefore focus on expanding the orbit propagation validation.

7.1.1. Extend the POD-based Orbit Propagation Validation

The most important recommendation is to extend the stage B validation beyond Swarm A. While the current results provide promising evidence that the neural network correction model improves short-term orbit propagation performance, the final validation is restricted to a single satellite and therefore to a limited orbital regime. This means that the present work demonstrates proof of concept, but does not yet establish that the method generalises across the full altitude range represented in the training data.

A particularly valuable next step would be to perform additional POD-based validation on satellites flying at lower altitudes than Swarm A. This would allow the correction framework to be tested in denser drag regimes, where thermospheric density errors generally have a stronger impact on orbit propagation. Such an analysis would provide a much stronger basis for assessing whether the observed improvement extends across the full training orbital regime rather than only to the Swarm A conditions considered here.

7.1.2. Expand the Stage B Validation Procedure

The current orbit propagation validation is based on seven selected arcs, each spanning approximately 7 to 10 evaluation days. Although these arcs were chosen to cover quiet, moderate and severe geomagnetic conditions, the total amount of stage B validation data remains limited. A more extensive validation would therefore strengthen the conclusions of this thesis considerably.

In practice, this means extending the number of POD arcs, increasing the temporal coverage across different solar-cycle phases and including a larger number of disturbed and recovery periods. A broader validation set would make it possible to assess whether the observed improvements are robust across a wider range of operational conditions and would reduce the chance that the present conclusions depend too strongly on a limited set of hand-selected cases.

7.1.3. Adopt a More Operationally Realistic C_d Tuning Strategy

In the present stage B setup, the drag coefficient C_d is tuned once per model on the day preceding each validation arc and is then held fixed for the full duration of that arc. This provides a controlled comparison between density models, but it is still a simplified representation of how a practical operational workflow would likely behave.

A useful follow-up study would therefore be to implement a more realistic C_d update strategy, for example by recalibrating C_d daily. In addition, the tuning itself could be made more extensive, for example by using longer calibration intervals or more operationally representative assimilation-style updates. This would allow the stage B framework to move closer to a realistic operational propagation setting.

7.2. Data Quality and Robustness

A second important theme concerns the quality and operational robustness of the input data used by the model. Since the proposed framework depends on accelerometer-derived density targets and on outage-prone space weather inputs, further work on data quality and feature reliability could directly improve both model performance and operational resilience. The following recommendations therefore focus on improving the target data and better understanding the effect of degraded feature availability.

7.2.1. Rebuild the Training Dataset Using More Aggressively Cleaned Accelerometer-Derived Density Data

A major recommendation for future work is to repeat the analysis using an accelerometer-derived density dataset that has been cleaned more aggressively. The present work already includes standard filtering and quality handling, but additional cleaning through more advanced outlier detection, smoothing or mission-specific noise reduction could further improve the quality of the training target.

Such a procedure would likely reduce the total amount of usable data. However, given the large size of the full dataset, this trade-off is likely acceptable. A cleaner target dataset could allow the neural network to focus more strongly on physically meaningful thermospheric variability rather than on mission-specific artefacts or measurement noise. This may improve both the learned correction in stage A and the downstream propagation performance in stage B.

7.2.2. Analyse the Effect of Feature Outages on Model Performance

The trust score framework developed in this thesis is one of the more novel practical elements of the methodology. It provides the model with explicit information on the freshness of IMF, solar wind and PCN inputs during outages. However, due to time constraints, the present work does not yet quantify in detail how model performance changes when these trust scores indicate degraded feature availability.

A particularly interesting extension would therefore be to analyse performance as a function of trust score level. This would give direct insight into how strongly the model depends on PCN, solar wind and IMF inputs and how much their outages degrade the model performance. Such a study would be highly relevant from an operational perspective, because it would provide a more explicit understanding of the resilience of the framework under realistic data availability limitations.

7.3. Model Development and Benchmarking

A third set of recommendations concerns the modelling choices made in this thesis itself. The present framework was intentionally designed to remain lightweight, physically motivated and computationally feasible, but several opportunities remain to test how strongly the results depend on the chosen baseline model, feature set and neural network configuration. The following recommendations therefore focus on benchmarking and on refining the modelling framework more systematically.

7.3.1. Benchmark the Correction Framework Against Additional Baseline Density Models

The current work adopts JB2008 as the baseline because of its strong operational relevance and comparatively good performance during disturbed conditions. However, it would be highly informative to repeat the full methodology using NRLMSISE-00 as the underlying baseline model and compare the resulting corrected model directly against the JB2008-based version.

This would allow the assessment of whether the quality of the corrected model depends strongly on the quality of the empirical baseline to which the neural network is attached. Given that NRLMSISE-00 appears to degrade more strongly than JB2008 during active periods, it is plausible that a correction built on top of NRLMSISE-00 would perform worse. Testing this explicitly would provide useful insight into the role of the baseline model in hybrid density correction frameworks.

In addition, future benchmarking should ideally compare the final correction model not only against JB2008 and NRLMSISE-00, but also against other important reference models such as DTM2020 and, where possible, the HASDM model. This would indicate the performance of the proposed method more clearly compared to the current options regarding operational thermospheric density modelling.

7.3.2. Perform a Systematic Feature Ablation and Feature Selection Study

The present feature set and temporal embedding strategy were designed using physical reasoning, standard space weather practice and practical judgement. While this produces a well-motivated final feature set, it does not establish whether the selected combination is optimal.

A valuable extension would therefore be to perform a systematic feature ablation study. One approach would be to begin with a much larger feature pool containing more or less all candidate parameters and temporal variants and then remove selected features or groups of features in a structured way. The resulting performance differences would reveal which drivers, temporal embeddings and trust features are most important to the final model. Such a study would also provide a stronger empirical basis for future feature selection decisions.

It must be noted, however, that this type of analysis would be computationally expensive due to the dataset size and the large number of possible feature combinations. For that reason, it was not performed in this work and is considered a follow-up study.

7.3.3. Perform a More Extensive Architecture and Hyperparameter Study

The neural network used in this thesis is intentionally relatively simple. The selected architecture and associated hyperparameters were chosen through manual, iterative tuning based primarily on the observed stage A validation performance. This was a practical and appropriate choice within the scope of the present work, as it allowed a stable and operationally lightweight correction framework to be developed without the computational burden of a large tuning process. However, it also means that the final configuration should not be interpreted as the optimal architecture for this problem.

A natural next step would therefore be to perform a more extensive study of both architecture choice and hyperparameter selection. This could include a broader exploration of trunk depth and width, embedding dimension, refinement-head design, dropout levels, loss weighting and batch construction, as well as comparisons with somewhat more advanced architectures. Such an analysis would clarify whether the current performance is already close to the achievable limit of this correction framework, or whether additional gains remain available through more systematic model selection. In this way, future work could determine more clearly to what extent the observed improvement arises from the physically motivated feature design itself, from the hybrid correction concept, or from the specific model configuration used in this thesis.

Overall, the recommendations of this chapter indicate that the most important next step is not to replace the current framework entirely, but to strengthen and extend it in a systematic way. In particular, broader POD-based validation across additional satellites and orbital regimes is required to establish whether the promising stage B results generalise beyond the limited proof-of-concept considered in this thesis. At the same time, improvements in target data quality, outage robustness, benchmarking and model selection offer clear opportunities to further refine the method. Within this perspective, the proposed hybrid correction framework provides a useful foundation for future research into more accurate and operationally practical thermospheric density modelling.

References

Acciarini, Giacomo et al. (Feb. 2024). “Improving Thermospheric Density Predictions in Low-Earth Orbit With Machine Learning”. In: *Space Weather* 22.2, e2023SW003652. ISSN: 15427390. DOI: 10.1029/2023SW003652. URL: /doi/pdf/10.1029/2023SW003652%20https://onlinelibrary.wiley.com/doi/abs/10.1029/2023SW003652%20https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023SW003652.

Akasofu, S. I. (June 1981). “Energy coupling between the solar wind and the magnetosphere”. In: *Space Science Reviews* 28.2, pp. 121–190. ISSN: 00386308. DOI: 10.1007/BF00218810/METRICS. URL: https://link.springer.com/article/10.1007/BF00218810.

Behzadpour, Saniya, Torsten Mayer-Gürr, and Sandro Krauss (May 2021). “GRACE Follow-On Accelerometer Data Recovery”. In: *Journal of Geophysical Research: Solid Earth* 126.5. ISSN: 21699356. DOI: 10.1029/2020JB021297.

Borovsky, Joseph E. (Jan. 2023). “Further investigation of the effect of upstream solar-wind fluctuations on solar-wind/magnetosphere coupling: Is the effect real?” In: *Frontiers in Astronomy and Space Sciences* 9, p. 975135. ISSN: 2296987X. DOI: 10.3389/fspas.2022.975135.

Bowman, Bruce R. et al. (2008). “A new empirical thermospheric density model JB2008 using new solar and geomagnetic indices”. In: *AIAA/AAS Astrodynamics Specialist Conference and Exhibit*. DOI: 10.2514/6.2008-6438;WEBSITE:WEBSITE:AIAA-SITE;WGROUP:STRING:AIAA. URL: /doi/pdf/10.2514/6.2008-6438.

Bruinsma, Sean and Claude Boniface (2021). “The operational and research DTM-2020 thermosphere models”. In: *Journal of Space Weather and Space Climate* 11.3, p. 47. ISSN: 2115-7251. DOI: 10.1051/SWSC/2021032. URL: https://www.swsc-journal.org/articles/swsc/full_html/2021/01/swsc210039/swsc210039.html%20https://www.swsc-journal.org/articles/swsc/abs/2021/01/swsc210039/swsc210039.html.

Cai, Xuguang et al. (May 2022). “The Effects of IMF By on the Middle Thermosphere During a Geomagnetically “Quiet” Period at Solar Minimum”. In: *Journal of Geophysical Research: Space Physics* 127.5, e2021JA029816. ISSN: 21699402. DOI: 10.1029/2021JA029816;ISSUE:ISSUE:DOI. URL: /doi/pdf/10.1029/2021JA029816%20https://onlinelibrary.wiley.com/doi/abs/10.1029/2021JA029816%20https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JA029816.

Campbell, Wallace H. (July 1996). “Geomagnetic storms, the Dst ring-current myth and lognormal distributions”. In: *Journal of Atmospheric and Terrestrial Physics* 58.10, pp. 1171–1187. ISSN: 00219169. DOI: 10.1016/0021-9169(95)00103-4. URL: https://doi.org/10.5636/jgg.36.257.

Chiu, Wiegand and Kuznetsova Masha (2009). *CMR*. URL: https://kauai.ccmc.gsfc.nasa.gov/CMR/view/model/SimulationModel?resourceID=spase%3A%2F%2FCCMC%2FSimulationModel%2FTIE-GCM%2F2.0.

Codrescu, Mihail V. et al. (Feb. 2012). “A real-time run of the Coupled Thermosphere Ionosphere Plasmasphere Electrodynamics (CTIPe) model”. In: *Space Weather* 10.1. ISSN: 15427390. DOI: 10.1029/2011SW000736;PAGE:STRING:ARTICLE/CHAPTER. URL: /doi/pdf/10.1029/2011SW000736%20https://onlinelibrary.wiley.com/doi/abs/10.1029/2011SW000736%20https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2011SW000736.

Doornbos, Eelco (2011). *Thermospheric density and wind determination from satellite dynamics*. URL: https://repository.tudelft.nl/record/uuid:33002be1-1498-4bec-a440-4c90ec149aea.

Dudok De Wit, Thierry and Sean Bruinsma (2017). “The 30 cm radio flux as a solar proxy for thermosphere density modelling”. In: *Journal of Space Weather and Space Climate* 7, A9. ISSN: 2115-7251. DOI: 10.1051/SWSC/2017008. URL: https://www.swsc-journal.org/articles/swsc/full_html/2017/01/swsc160042/swsc160042.html%20https://www.swsc-journal.org/articles/swsc/abs/2017/01/swsc160042/swsc160042.html.

Emmert, J. T. et al. (Mar. 2021). “NRLMSIS 2.0: A Whole-Atmosphere Empirical Model of Temperature and Neutral Species Densities”. In: *Earth and Space Science* 8.3. ISSN: 23335084. DOI: 10.1029/2020EA001321.

Ergene, Emre (2024). “Ballistic coefficient estimation methods using two-line elements”. In: URL: https://www.politesi.polimi.it/handle/10589/231177.

ESA, CNES, and CSA (2024). *Swarm (Geomagnetic LEO Constellation) - eoPortal*. URL: https://directory.eoportal.org/satellite-missions/swarm#space-segment-concept.

Fuller-Rowell, T. J. et al. (Mar. 1994). “Response of the thermosphere and ionosphere to geomagnetic storms”. In: *Journal of Geophysical Research: Space Physics* 99.3, pp. 3893–3914. ISSN: 21699402. DOI: 10.1029/93JA02015. URL: /doi/pdf/10.1029/93JA02015%20https://onlinelibrary.wiley.com/doi/abs/10.1029/93JA02015%20https://agupubs.onlinelibrary.wiley.com/doi/10.1029/93JA02015.

Gan, Quan et al. (Jan. 2024). “Thermospheric Responses to the 3 and 4 November 2021 Geomagnetic Storm During the Main and Recovery Phases as Observed by NASA’s GOLD and ICON Missions”. In: *Geophysical Research Letters* 51.1, e2023GL106529. ISSN: 19448007. DOI: 10.1029/2023GL106529;REQUESTEDJOURNAL:JOURNAL:19448007;WGROUP:STRING:PUBLICATION. URL: /doi/pdf/10.1029/2023GL106529%20https://onlinelibrary.wiley.com/doi/abs/10.1029/2023GL106529%20https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023GL106529.

Gardner, L. C. and R. W. Schunk (Aug. 2010). “Generation of traveling atmospheric disturbances during pulsating geomagnetic storms”. In: *Journal of Geophysical Research: Space Physics* 115.A8. ISSN: 2156-2202. DOI: 10.1029/2009JA015129. URL: /doi/pdf/10.1029/2009JA015129%20https://onlinelibrary.wiley.com/doi/abs/10.1029/2009JA015129%20https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2009JA015129.

Guo, Yu et al. (May 2025). “A Two-Step Method for Enhancing Thermospheric Mass Density Empirical Model”. In: *Space Weather* 23.5, e2024SW004317. ISSN: 15427390. DOI: 10.1029/2024SW004317. URL: /doi/pdf/10.1029/2024SW004317%20https://onlinelibrary.wiley.com/doi/abs/10.1029/2024SW004317%20https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024SW004317.

He, Changyong et al. (Nov. 2018). “Review and comparison of empirical thermospheric mass density models”. In: *Progress in Aerospace Sciences* 103, pp. 31–51. ISSN: 03760421. DOI: 10.1016/J.PAEROSCI.2018.10.003. URL: https://www.researchgate.net/publication/328857274_Review_and_comparison_of_empirical_thermospheric_mass_density_models.

Koskinen, Hannu E.J. and Eija I. Tanskanen (Nov. 2002). “Magnetospheric energy budget and the epsilon parameter”. In: *Journal of Geophysical Research: Space Physics* 107.A11, pp. 42–1. ISSN: 21699402. DOI: 10.1029/2002JA009283;SUBPAGE:STRING:FULL. URL: /doi/pdf/10.1029/2002JA009283%20https://onlinelibrary.wiley.com/doi/abs/10.1029/2002JA009283%20https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2002JA009283.

Krauss, S. et al. (Oct. 2015). “Thermospheric and geomagnetic responses to interplanetary coronal mass ejections observed by ACE and GRACE: Statistical results”. In: *Journal of Geophysical Research: Space Physics* 120.10, pp. 8848–8860. ISSN: 21699402. DOI: 10.1002/2015JA021702;WGROUP:STRING:PUBLICATION. URL: /doi/pdf/10.1002/2015JA021702%20https://onlinelibrary.wiley.com/doi/abs/10.1002/2015JA021702%20https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2015JA021702.

Kutiev, Ivan et al. (2013). “Solar activity impact on the Earth’s upper atmosphere”. In: *Journal of Space Weather and Space Climate* 3, A06. ISSN: 2115-7251. DOI: 10.1051/SWSC/2013028. URL: https://www.swsc-journal.org/articles/swsc/full_html/2013/01/swsc120043/swsc120043.html%20https://www.swsc-journal.org/articles/swsc/abs/2013/01/swsc120043/swsc120043.html.

Laundal, K. M. et al. (May 2018). “Interplanetary Magnetic Field Bx Component Influence on Horizontal and Field-Aligned Currents in the Ionosphere”. In: *Journal of Geophysical Research: Space Physics* 123.5, pp. 3360–3379. ISSN: 21699402. DOI: 10.1002/2017JA024864;WEBSITE:WEBSITE:AGUPUBS;ISSUE:ISSUE:DOI. URL: /doi/pdf/10.1002/2017JA024864%20https://onlinelibrary.wiley.com/doi/abs/10.1002/2017JA024864%20https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2017JA024864.

Licata, Richard J. et al. (Apr. 2022). “Machine-Learned HASDM Thermospheric Mass Density Model With Uncertainty Quantification”. In: *Space Weather* 20.4, e2021SW002915. ISSN: 15427390. DOI: 10.1029/2021SW002915. URL: /doi/pdf/10.1029/2021SW002915%20https://onlinelibrary.wiley.com/doi/abs/10.1029/2021SW002915%20https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021SW002915.

Lockwood, Mike (Feb. 2022). “Solar Wind—Magnetosphere Coupling Functions: Pitfalls, Limitations, and Applications”. In: *Space Weather* 20.2, e2021SW002989. ISSN: 15427390. DOI: 10.1029/2021SW002989;REQUESTEDJOURNAL:JOURNAL:15427390;WEBSITE:WEBSITE:AGUPUBS;WGROUP:STRING:PUBLICATION. URL: /doi/pdf/10.1029/2021SW002989%20https://onlinelibrary.wiley.com/doi/abs/10.1029/2021SW002989%20https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021SW002989.

Lockwood, Mike, Kathryn A. McWilliams, et al. (2020). “Semi-annual, annual and Universal Time variations in the magnetosphere and in geomagnetic activity: 2. Response to solar wind power input and relationships with solar wind dynamic pressure and magnetospheric flux transport”. In: *Journal of Space Weather and Space Climate* 10, p. 30. ISSN: 2115-7251. DOI: 10.1051/SWSC/2020033. URL: https://www.swsc-journal.org/articles/swsc/full_html/2020/01/swsc190052/swsc190052.html%20https://www.swsc-journal.org/articles/swsc/abs/2020/01/swsc190052/swsc190052.html.

Lockwood, Mike, Mathew J. Owens, et al. (June 2016). “On the origins and timescales of geoeffective IMF”. In: *Space Weather* 14.6, pp. 406–432. ISSN: 15427390. DOI: 10.1002/2016SW001375;PAGE:STRING:ARTICLE/CHAPTER. URL: /doi/pdf/10.1002/2016SW001375%20https://onlinelibrary.wiley.com/doi/abs/10.1002/2016SW001375%20https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2016SW001375.

Matzka, J. et al. (May 2021). “The Geomagnetic Kp Index and Derived Indices of Geomagnetic Activity”. In: *Space Weather* 19.5. ISSN: 15427390. DOI: 10.1029/2020SW002641.

Moe, Kenneth and Mildred M. Moe (July 2005). “Gas–surface interactions and satellite drag coefficients”. In: *Planetary and Space Science* 53.8, pp. 793–801. ISSN: 0032-0633. DOI: 10.1016/J.PSS.2005.03.005. URL: https://www.sciencedirect.com/science/article/abs/pii/S0032063305000486.

Mostaza Prieto, David, Benjamin P. Graziano, and Peter C.E. Roberts (2014). “Spacecraft drag modelling”. In: *Progress in Aerospace Sciences* 64, pp. 56–65. ISSN: 03760421. DOI: 10.1016/J.PAEROSCI.2013.09.001.

Mutschler, Shaylah et al. (2019). “Physics-based Approach to Density Estimation and Prediction using Orbital Debris Tracking Data”. In: URL: www.amostech.com.

Mutschler, Shaylah Maria (2022). *Global Thermospheric Density Estimation using CubeSat GPS Data and a Physics-based Space Environment Model*. Tech. rep.

Newell, P. T. et al. (Jan. 2007). “A nearly universal solar wind-magnetosphere coupling function inferred from 10 magnetospheric state variables”. In: *Journal of Geophysical Research: Space Physics* 112.1. ISSN: 21699402. DOI: 10.1029/2006JA012015.

Pérez, David et al. (May 2015). “Neural Network based calibration of atmospheric density models”. In: *Acta Astronautica* 110.1, pp. 58–76. ISSN: 00945765. DOI: 10.1016/j.actaastro.2014.01.007. URL: https://doi.org/10.1016/j.actaastro.2014.01.007.

Picone, J. M. et al. (Dec. 2002). “NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues”. In: *Journal of Geophysical Research: Space Physics* 107.A12, pp. 15–1. ISSN: 21699402. DOI: 10.1029/2002JA009430,. URL: /doi/pdf/10.1029/2002JA009430%20https://onlinelibrary.wiley.com/doi/abs/10.1029/2002JA009430%20https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2002JA009430.

Pulkkinen, Tuija (May 2007). “Space weather: Terrestrial perspective”. In: *Living Reviews in Solar Physics* 4.1, pp. 1–60. ISSN: 16144961. DOI: 10.12942/lrsp-2007-1. URL: https://link.springer.com/article/10.12942/lrsp-2007-1.

Qian, Liying et al. (May 2014). “The NCAR TIE-GCM: A Community Model of the Coupled Thermosphere/Ionosphere System”. In: *Wiley Blackwell 6* 9780875904917, pp. 73–83. DOI: 10.1002/9781118704417.CH7;PAGE:STRING:ARTICLE/CHAPTER. URL: /doi/pdf/10.1002/9781118704417.ch7%20https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118704417.ch7%20https://agupubs.onlinelibrary.wiley.com/doi/10.1002/9781118704417.ch7.

Ren, Dexin et al. (Feb. 2021). “Observations and Simulations of the Peak Response Time of Thermospheric Mass Density to the 27-Day Solar EUV Flux Variation”. In: *Journal of Geophysical Research: Space Physics* 126.2, e2020JA028756. ISSN: 21699402. DOI: 10.1029/2020JA028756;WGROUP:STRING:PUBLICATION. URL: /doi/pdf/10.1029/2020JA028756%20https://onlinelibrary.wiley.com/doi/abs/10.1029/2020JA028756%20https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020JA028756.

Ridley, A. J., Y. Deng, and G. Tóth (May 2006). “The global ionosphere–thermosphere model”. In: *Journal of Atmospheric and Solar-Terrestrial Physics* 68.8, pp. 839–864. ISSN: 1364-6826. DOI: `10.1016/J.JASTP.2006.01.008`. URL: `https://www.sciencedirect.com/science/article/abs/pii/S1364682606000071?via%3Dihub`.

Royal Belgian Institute for Space Aeronomy (Oct. 2020). *The Cluster Epic: unravelling the consequences of the electromagnetic forces at the planetary scale*. URL: `https://www.aeronomie.be/en/news/2020/cluster-epic-unravelling-consequences-electromagnetic-forces-planetary-scale`.

Sheng, Cheng et al. (Apr. 2017). “Thermospheric recovery during the 5 April 2010 geomagnetic storm”. In: *Journal of Geophysical Research: Space Physics* 122.4, pp. 4588–4599. ISSN: 2169-9402. DOI: `10.1002/2016JA023520`. URL: `/doi/pdf/10.1002/2016JA023520%20https://onlinelibrary.wiley.com/doi/abs/10.1002/2016JA023520%20https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2016JA023520`.

Sheridan, Phillip Logan et al. (May 2022). “Updates and improvements to the satellite drag coefficient Response Surface Modeling toolkit”. In: *Advances in Space Research* 69.10, pp. 3828–3846. ISSN: 0273-1177. DOI: `10.1016/J.ASR.2022.02.044`. URL: `https://www.sciencedirect.com/science/article/pii/S0273117722001569#b0275`.

Siemes, Christian, Claudia Borries, et al. (2023). “New thermosphere neutral mass density and crosswind datasets from CHAMP, GRACE, and GRACE-FO”. In: *Journal of Space Weather and Space Climate* 13, p. 16. ISSN: 2115-7251. DOI: `10.1051/SWSC/2023014`. URL: `https://www.swsc-journal.org/articles/swsc/full_html/2023/01/swsc230004/swsc230004.html%20https://www.swsc-journal.org/articles/swsc/abs/2023/01/swsc230004/swsc230004.html`.

Siemes, Christian, João De Teixeira Da Encarnação, et al. (Dec. 2016). “Swarm accelerometer data processing from raw accelerations to thermospheric neutral densities 2. Aeronomy Swarm Science Results after two years in Space”. In: *Earth, Planets and Space* 68.1. ISSN: 18805981. DOI: `10.1186/s40623-016-0474-5`.

Storz, Mark F., Bruce R. Bowman, and Major James I. Branson (2002). “High accuracy satellite drag model (HASDM)”. In: *AIAA/AAS Astrodynamics Specialist Conference and Exhibit*. DOI: `10.2514/6.2002-4886;CTYPE:STRING:BOOK`. URL: `/doi/pdf/10.2514/6.2002-4886`.

Sutton, E. K., J. M. Forbes, and R. S. Nerem (Sept. 2005). “Global thermospheric neutral density and wind response to the severe 2003 geomagnetic storms from CHAMP accelerometer data”. In: *Journal of Geophysical Research: Space Physics* 110.A9, pp. 9–40. ISSN: 2156-2202. DOI: `10.1029/2004JA010985`. URL: `/doi/pdf/10.1029/2004JA010985%20https://onlinelibrary.wiley.com/doi/abs/10.1029/2004JA010985%20https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2004JA010985`.

Sutton, Eric K., Jeffrey P. Thayer, et al. (June 2021). “Toward Accurate Physics-Based Specifications of Neutral Density Using GNSS-Enabled Small Satellites”. In: *Space Weather* 19.6. ISSN: 15427390. DOI: `10.1029/2021SW002736`.

T., Iyemori et al. (Feb. 2010). “Mid-latitude Geomagnetic Indices ”ASY” and ”SYM” for 200 (Provisional)”. In.

Tenfjord, P. and N. Østgaard (Sept. 2013). “Energy transfer and flow in the solar wind-magnetosphere-ionosphere system: A new coupling function”. In: *Journal of Geophysical Research: Space Physics* 118.9, pp. 5659–5672. ISSN: 21699402. DOI: `10.1002/JGRA.50545;PAGEGROUP:STRING:PUBLICATION`. URL: `/doi/pdf/10.1002/jgra.50545%20https://onlinelibrary.wiley.com/doi/abs/10.1002/jgra.50545%20https://agupubs.onlinelibrary.wiley.com/doi/10.1002/jgra.50545`.

Toyohisa, Kamei, Sugiura Masahisa, and Araki Tohru (1998). *Auroral Electrojet (AE) Indices*. URL: `https://wdc.kugi.kyoto-u.ac.jp/aedir/ae2/onAEindex.html`.

Trattner, K. J., S. M. Petrinec, and S. A. Fuselier (Apr. 2021). “The Location of Magnetic Reconnection at Earth’s Magnetopause”. In: *Space Science Reviews* 217.3, p. 41. ISSN: 15729672. DOI: `10.1007/s11214-021-00817-8`. URL: `https://pmc.ncbi.nlm.nih.gov/articles/PMC8550343/`.

Verkhoglyadova, Olga et al. (2016). “Estimation of energy budget of ionosphere-thermosphere system during two CIR-HSS events: observations and modeling”. In: *Journal of Space Weather and Space Climate* 6, A20. ISSN: 2115-7251. DOI: `10.1051/SWSC/2016013`. URL: `https://www.swsc-journal.org/articles/swsc/full_html/2016/01/swsc150046/swsc150046.html%20https://www.swsc-journal.org/articles/swsc/abs/2016/01/swsc150046/swsc150046.html`.

Wang, Peian, Zhou Chen, et al. (Mar. 2022). “The Prediction of Storm-Time Thermospheric Mass Density by LSTM-Based Ensemble Learning”. In: *Space Weather* 20.3, e2021SW002950. ISSN: 15427390. DOI: `10.1029/2021SW002950;ISSUE:ISSUE:DOI`. URL: `/doi/pdf/10.1029/2021SW002950%20https://onlinelibrary.wiley.com/doi/abs/10.1029/2021SW002950%20https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021SW002950`.

Wang, Xin, Juan Miao, et al. (Apr. 2022). “Using Temporal Relationship of Thermospheric Density With Geomagnetic Activity Indices and Joule Heating as Calibration for NRLMSISE-00 During Geomagnetic Storms”. In: *Space Weather* 20.4, e2021SW003017. ISSN: 15427390. DOI: `10.1029/2021SW003017;PAGE:STRING:ARTICLE/CHAPTER`. URL: `/doi/pdf/10.1029/2021SW003017%20https://onlinelibrary.wiley.com/doi/abs/10.1029/2021SW003017%20https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021SW003017`.

Weng, Libin et al. (Mar. 2020). “A Machine-Learning Approach to Derive Long-Term Trends of Thermospheric Density”. In: *Geophysical Research Letters* 47.6, e2020GL087140. ISSN: 19448007. DOI: `10.1029/2020GL087140`. URL: `/doi/pdf/10.1029/2020GL087140%20https://onlinelibrary.wiley.com/doi/abs/10.1029/2020GL087140%20https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020GL087140`.

Wilson, G. R. et al. (Oct. 2006). “Response of the thermosphere to Joule heating and particle precipitation”. In: *Journal of Geophysical Research: Space Physics* 111.10. ISSN: 21699402. DOI: `10.1029/2005JA011274;PAGEGROUP:STRING:PUBLICATION`. URL: `/doi/pdf/10.1029/2005JA011274%20https://onlinelibrary.wiley.com/doi/abs/10.1029/2005JA011274%20https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2005JA011274`.

Zesta, Eftyhia and Denny M. Oliveira (Nov. 2019). “Thermospheric Heating and Cooling Times During Geomagnetic Storms, Including Extreme Events”. In: *Geophysical Research Letters* 46.22, pp. 12739–12746. ISSN: 19448007. DOI: `10.1029/2019GL085120;REQUESTEDJOURNAL:JOURNAL:19448007;WGROUP:STRING:PUBLICATION`. URL: `/doi/pdf/10.1029/2019GL085120%20https://onlinelibrary.wiley.com/doi/abs/10.1029/2019GL085120%20https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019GL085120`.

A

Full Feature Normalisation Decision Table

Table A.1: Feature-level normalisation decisions used in the final preprocessing pipeline.

Feature	Normalisation Strategy
Min-max scaling	
alt_km	Min-max scaling
lat_deg	Min-max scaling
Z-scaling	
F10_daily_shift_1440m	Z-scaling
F10_daily_shift_2880m	Z-scaling
F10_daily_sma_116640m	Z-scaling
IMF_theta_cos_shift_0m	Z-scaling
IMF_theta_sin_half_shift_0m	Z-scaling
IMF_theta_sin_shift_0m	Z-scaling
M10_daily_shift_1440m	Z-scaling
M10_daily_shift_2880m	Z-scaling
M10_daily_sma_116640m	Z-scaling
S10_daily_shift_1440m	Z-scaling
S10_daily_shift_2880m	Z-scaling
S10_daily_sma_116640m	Z-scaling
Y10_daily_shift_1440m	Z-scaling
Y10_daily_shift_2880m	Z-scaling
Y10_daily_sma_116640m	Z-scaling
al_index_nT_int_180m	Z-scaling
al_index_nT_int_720m	Z-scaling
al_index_nT_shift_0m	Z-scaling
asy_h_nT_shift_0m	Z-scaling
asy_h_nT_sma_60m	Z-scaling
bx_gsm_nT_shift_0m	Z-scaling
by_gsm_nT_ewma_180m	Z-scaling
by_gsm_nT_shift_0m	Z-scaling
by_gsm_nT_sma_60m	Z-scaling
bz_gsm_nT_ewma_180m	Z-scaling
bz_gsm_nT_shift_0m	Z-scaling
bz_gsm_nT_sma_60m	Z-scaling
dst_nT_1hr_ewma_1440m	Z-scaling

Feature	Normalisation Strategy
dst_nT_1hr_shift_60m	Z-scaling
sw_speed_km_s_shift_0m	Z-scaling
sw_speed_km_s_sma_180m	Z-scaling
sw_speed_km_s_sma_360m	Z-scaling
sym_h_nT_ewma_1440m	Z-scaling
sym_h_nT_ewma_180m	Z-scaling
sym_h_nT_shift_0m	Z-scaling
Log10 + Z-scaling	
Newell_Coupling_dPhi_dt_int_180m	Log10 + Z-scaling
Newell_Coupling_dPhi_dt_int_60m	Log10 + Z-scaling
Newell_Coupling_dPhi_dt_shift_0m	Log10 + Z-scaling
ae_index_nT_shift_0m	Log10 + Z-scaling
ae_index_nT_sma_60m	Log10 + Z-scaling
ap_3hr_ewma_180m	Log10 + Z-scaling
ap_3hr_shift_180m	Log10 + Z-scaling
ap_3hr_sma_1440m	Log10 + Z-scaling
density_jb2008_kg_m3	Log10 + Z-scaling
density_kg_m3	Log10 + Z-scaling
pcn_index_ewma_360m	Log10 + Z-scaling
pcn_index_int_180m	Log10 + Z-scaling
pcn_index_shift_0m	Log10 + Z-scaling
sw_flow_pressure_nPa_shift_0m	Log10 + Z-scaling
sw_proton_density_n_cc_shift_0m	Log10 + Z-scaling
vBs_int_60m	Log10 + Z-scaling
vBs_shift_0m	Log10 + Z-scaling
sw_flow_pressure_nPa_sma_60m	Log10 + Z-scaling
Vector renormalisation	
arg_lat_cos	Vector renormalisation
arg_lat_sin	Vector renormalisation
lon_cos	Vector renormalisation
lon_sin	Vector renormalisation
lst_cos	Vector renormalisation
lst_sin	Vector renormalisation
Pass-through	
Any_Outage_Flag_shift_0m	Pass-through
Combined_Trust_int_min_180m	Pass-through
Combined_Trust_int_min_60m	Pass-through
Combined_Trust_shift_0m	Pass-through
IMF_Trust_Score_ewma_180m	Pass-through
IMF_Trust_Score_shift_0m	Pass-through
IMF_Trust_Score_sma_60m	Pass-through
PCN_Trust_Score_ewma_180m	Pass-through
PCN_Trust_Score_shift_0m	Pass-through
PCN_Trust_Score_sma_60m	Pass-through
SW_Trust_Score_ewma_180m	Pass-through
SW_Trust_Score_shift_0m	Pass-through
SW_Trust_Score_sma_60m	Pass-through
Embedding	
sat_id_int	Embedding

B

High-Level Thesis Planning

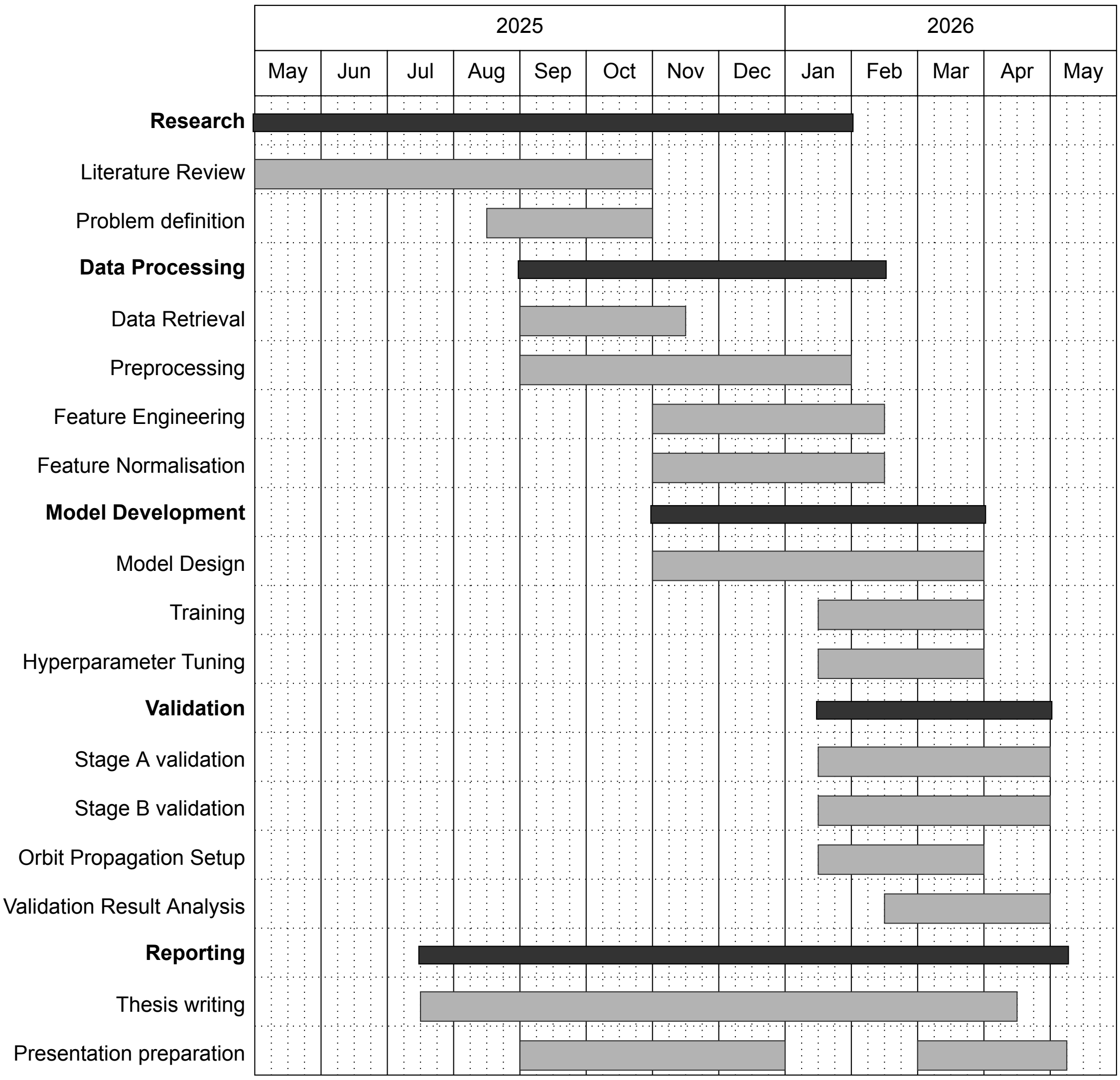


Figure B.1: High-level thesis planning overview spanning May 2025 to May 2026.