Dropout and Ensemble Networks for Thermospheric Density Uncertainty Estimation

Stefano Bonasera
University of Colorado, Boulder
stefano.bonasera@colorado.edu

Giacomo Acciarini
Frontier Development Lab
giacomo.acciarini@gmail.com

Jorge A. Pérez-Hernández
Mexico National Autonomous University
jperez@icf.unam.mx

Bernard Benson
The University of Alabama in Huntsville
bb0008@uah.edu

Edward Brown
University of Cambridge
ejeb4@cam.ac.uk

Eric Sutton
University of Colorado, Boulder
eric.sutton@colorado.edu

Moriba K. Jah
University of Texas, Austin
moriba@utexas.edu

Christopher Bridges
University of Surrey
c.p.bridges@surrey.ac.uk

Atılım Güneş Baydin
University of Oxford
gunes@robots.ox.ac.uk

Abstract

Accurately estimating spacecraft location is of crucial importance for a variety of safety-critical tasks in low-Earth orbit (LEO), including satellite collision avoidance and re-entry. The solar activity largely impacts the physical characteristics of the thermosphere, consequently affecting the trajectories of spacecraft in LEO. State-of-the-art models for estimating thermospheric density are either computationally expensive or under-perform during extreme solar activity. Moreover, these models provide single-point solutions, neglecting critical information on the associated uncertainty. In this work we use and compare two methods, Monte Carlo dropout and deep ensembles, to estimate thermospheric total mass density and associated uncertainty. The networks are trained using ground-truth density data from five well-calibrated satellites, using orbital data information, solar and geomagnetic indices as input. The trained models improve for a subset of satellites upon operational solutions, also providing a measure of uncertainty in the density estimation.

1 Introduction

Precise estimation of thermospheric total mass density is essential for accurate prediction of satellite trajectories in low-Earth orbit (LEO) [2]. Accurate knowledge of satellite location with associated uncertainties is crucial for safety and accessibility to space, representing a key aspect for a variety of space-related problems, comprising collision avoidance, re-entry, and lifetime analysis [16, 24]. State-of-the-art methods for density estimation are usually separated into empirical and physics-based models [13, 29]. Empirical models are constructed via parameter fitting utilizing satellite observations, and they represent rapid solutions for density estimation for LEO satellites. However, these models present biases, often underperform for periods of very high or low solar activity, and lack any direct estimate of the uncertainty in the estimation. Conversely, physics-based models are constructed to solve fluid equations over discretized volumes to obtain global trends of atmospheric characteristics,

Bayesian Deep Learning workshop, NeurIPS 2021.

offering greater forecasting ability than empirical models, although necessitating relatively higher computational resources [19].

The physical characteristics of the thermosphere are heavily impacted by solar activity [15]. To express the influence of the Sun on the thermospheric density, empirical and physics-based models often leverage approximated space weather indices as input. These indices are usually measured with on-ground instrumentation and are representative of the solar activity. Examples of space weather indices comprise F10.7, F30 proxies and the ap and Kp geomagnetic indices [32]. Solar information approximated by space weather indices can be used in data-driven approaches to training models for thermospheric density estimation. Once trained, these models can be employed to predict the thermospheric density encountered by a LEO satellite. A variety of studies have focused on training models for point-wise estimation and forecasting of thermospheric density. In particular, Pérez et al. [26] train feed-forward neural networks to forecast thermospheric density estimates for a well-calibrated LEO satellite, the Challenging Minisatellite Payload (CHAMP), using measured density and solar proxies as input. In a follow-up study, Pérez and Bevilacqua [25] employ as input the density estimates of three different empirically-based models to obtain an improved forecast prediction of thermospheric density for CHAMP. In a more recent study, George [11] uses long short-term memory networks to forecast thermospheric density prediction for CHAMP and the Gravity Recovery and Climate Experiment (GRACE) missions, using thermospheric density measurements and solar proxies.

In this work, we introduce dropout and ensemble neural networks (NNs) for the estimation of thermospheric density for LEO satellites. The networks are trained using real orbital data, as well as historical solar proxies and geomagnetic indices, and allow to quantify the uncertainty associated with the prediction. We demonstrate the approach in the framework of a variety of LEO satellites and well-calibrated missions, including CHAMP (GFZ), GRACE (NASA), the Global Ocean Circulation Explorer (GOCE), and the Swarm constellation (ESA) [27, 31, 30, 6]. Our work demonstrates the feasibility of incorporating data from multiple space missions to estimate thermospheric density and the associated uncertainty, towards the final effort of an accurate, reliable, and computationally lightweight global thermospheric model that can be deployed in a variety of scenarios.

2 Model uncertainty via dropout and ensemble networks

Prediction of thermospheric density with NNs is usually performed by single-point or time-series predictions [26, 25, 11]. However, it is crucial to obtain a measure of the uncertainty of the output, allowing the spacecraft operator to trust the network by knowing the associated confidence, essential for a variety of safety-critical tasks such as collision avoidance, re-entry, and lifetime analysis [4, 2].

Bayesian probability theory provides a mathematical framework that can be used to generate and train NNs with predicted uncertainty. Indeed, Bayesian deep learning methods output distributions over predictions, instead of single-point data. In Bayesian NNs, the distribution over the predictions is obtained by placing a prior on the parameters of the network $p(\omega)$, and finding the posterior that is most likely to have produced the training dataset, $p(\omega|X, Y)$, where X, Y are the training input and output data. Given a new input x^*, the predictive distribution is $p(y^*|x^*, X, Y) = \int p(y^*|x^*, \omega)p(\omega|X, Y)d\omega$. When the distribution $p(\omega|X, Y)$ cannot be analytically evaluated, variational inference can be used by introducing an approximate distribution $q(\omega)$. Therefore, the problem becomes minimizing the Kullback–Leibler (KL) divergence $\mathrm{KL}(q(\omega)||p(\omega|X, Y))$ to retrieve the approximate distribution, resulting in $q(y^*|x^*) = \int p(y^*|x^*, \omega)q(\omega)d\omega$. Given the complexity of training Bayesian NNs, a variety of approaches have been developed to obtain approximate measures of model uncertainty leveraging non-Bayesian approaches.

To cope with the computationally expensive framework of Bayesian NNs, Gal and Ghahramani [8, 9, 10] develop a theoretical framework, demonstrating that dropout training in non-Bayesian NNs can be interpreted as an approximation of Bayesian inference in deep Gaussian processes. Using dropout to approximate model uncertainty requires lower computational effort than training Bayesian NNs, with minor modifications to an NN architecture. To obtain a measure of uncertainty on a new data point at test time, the dropout NN is run multiple times stochastically with the same input [8, 9, 10]. An alternative non-Bayesian paradigm for measuring uncertainty in the model is deep ensembles [18]. Instead of training a single model, ensembles use the output of a set of different models to generate a more accurate final model with quantification on the uncertainty [18, 5].

Lakshminarayanan et al. [18] demonstrate that an ensemble of five different models sufficiently improves uncertainty estimation on a variety of datasets. Moreover, ensembles are generally more robust towards extreme events, as well as with out-of-distribution data, coping with the overfitting of the single models [28]. For space weather forecasting, ensemble techniques have become increasingly popular in recent years [21, 12, 20, 7, 28].

In this work, Monte Carlo dropout and deep ensemble are used to provide a measure of uncertainty in the thermospheric density estimation. We train a single model generating a point-wise thermospheric density estimation with Monte Carlo dropout, and an ensemble of models, each generating a mean and a standard deviation. The mean and standard deviations of the models used for the deep ensemble are then combined in a Gaussian mixture to construct a probability density function for the thermospheric density prediction, providing the density estimation and the associated uncertainty. Other methodologies for estimating model uncertainty for thermospheric density, for example using iterates of stochastic gradient descent, are the scope of future research [34, 14]

3 Thermospheric density data

In the realm of LEO satellites, just a few have onboard accelerometers for accurate measurement of the exerting forces. These spacecraft allow to accurately estimate the thermospheric mass density as a byproduct of the precise orbit determination (POD) process [6]. In particular, in this work we use POD products of five satellites: CHAMP, GRACE, GOCE, Swarm A, and B [6, 30, 27]. These POD products provide accurate estimates of thermospheric density over a combined time span of approximately 21 years, almost encompassing 2 full long solar cycles and different thermospheric altitudes, as reported in Fig. 1. Ground-truth estimates of thermospheric density are obtained from these datasets, together with orbital data. Flags are included in the dataset to exclude incorrect data. Orbital and thermospheric data is on average available with a time resolution of 10 seconds for GOCE, and 30 seconds for the other three missions. To avoid cross-correlation, the datasets are divided to have validation subsets for December of each year and three hold-out test sets, corresponding to the months of Jan 2006, Jun 2010, and Aug 2017, as depicted by the red vertical bands in Fig. 1. Moreover, each hold-out test set is selected to contain data from at least two different satellites and to be associated with different solar phases, therefore providing a large variety of input.

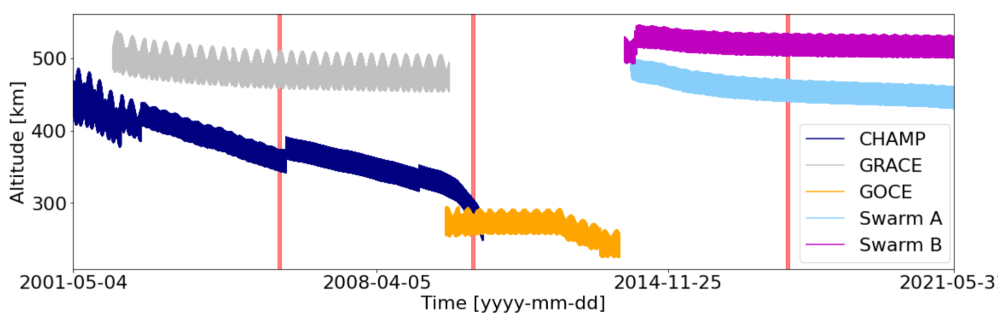

Figure 1: Trend of altitudes over mission time span. Red lines indicate time span of the test set.

To describe the total mass density of the thermospheric layer, density models are used in operational frameworks. These models are generally divided into empirical and physics-based ones. Thanks to their rapid accessibility, the NRLMSISE-00 and the Jacchia–Bowman 2008 (JB08) empirical models are leveraged in this paper as a prediction baseline representing the state of the art [23, 3]. These models require as input satellite orbital information, as well as indices for solar and geomagnetic activity. For example, NRLMSISE-00 requires as input: day of the year doy, year, seconds of the day, altitude h, latitude lat, longitude lon, local solar time lst, 81-day average of observed F10.7, the daily observed averaged F10.7 at the previous day, the daily Ap and the seven 3-hour ap indices centered at the closest date [23, 32, 33]. Comparison of the trained models with other empirical and state-of-the-art physics-based models is an avenue for future studies.

4 Experiments

The orbital information of five well-calibrated satellites and space weather indices information are leveraged as input to generate models for thermospheric density estimation with uncertainty. In particular, a Monte Carlo dropout NN with 0.2 dropout probability and an ensemble of five models are trained to predict thermospheric density and associated uncertainties. The dropout NN is constructed to output a point-wise solution, drawing 100 samples per prediction at test time. Each network in the ensemble outputs both mean and standard deviation and it is initialized with a different random number seed for 5 different instances. Each trained model contains three hidden layers with 256, 256, and 128 units, respectively. A batch size of 64, 30 epochs, a learning rate of 0.0002 and ReLu [1] activation layers are used for training each model. Adam optimizer, as implemented in PyTorch, is used to optimize the network weights [17, 22]. The experiments are run in Google Cloud, using 60 C2 vCPUs.

In the experiments, the orbital information from each satellite in Fig. 1 is used to train the models. The problem is formulated as a regression task, where the same inputs required by NRLMSISE-00 are used as input for training the networks, while the output is the thermospheric density. However, the spherical coordinates $[lat, lon, h]$ are converted into the Cartesian frame to avoid discontinuities, corresponding to a cyclic feature encoding in three dimensions [33]. After training, the networks are evaluated on the hold-out test sets. An example of the performances is shown in Fig. 2 for the CHAMP mission, for a subset of the testing set on Jan 15, 2006 spanning from 5:00:00 until 13:00:00 UTC. Figure 2a) visualizes the thermospheric density estimation of the dropout model, in black, and the ensemble model, in red, with the ground truth, in blue. In the same figure, the 1σ uncertainties associated with the dropout and the ensemble are reported in semi-transparent gray and red regions, respectively, while the output of the NRLMSISE-00 and JB08 baselines are depicted in orange and magenta, respectively. Both dropout and ensemble models outperform the two empirical models for the investigated time span. The dropout model provides larger uncertainties compared to the ensemble, although it bounds on average the ground truth within 1σ. Figure 2b) reports, for the same time span, the root mean square errors (RMSEs) for the two trained models and the two empirical models. The same coloring scheme is applied to distinguish among the models. The RMSEs reported in Fig. 2b) further confirm the improved result of the two trained models compared to operational ones on the analyzed time span and for the investigated satellite. Eventually, the average RMSEs of each satellite are reported in Table 1, for both the validation and test sets. The trained dropout and ensemble models are compared with the output of NRLMSISE-00 and JB08. Overall, the trained models outperform the two empirical ones for the thermospheric density associated with the CHAMP and GOCE missions, and partly for Swarm A, leaving room for future model improvement.

Table 1: Average RMSE for the different models over the test and validation set.

		CHAMP	GRACE	GOCE	Swarm A	Swarm B
Test	Ensemble	5.25×10^{-13}	5.54×10^{-13}	1.15×10^{-12}	4.31×10^{-14}	1.04×10^{-13}
	Dropout	6.55×10^{-13}	8.60×10^{-14}	1.29×10^{-12}	1.15×10^{-13}	7.44×10^{-14}
	NRLMSISE	1.82×10^{-12}	3.60×10^{-14}	3.17×10^{-12}	1.22×10^{-13}	4.08×10^{-14}
	JB08	1.69×10^{-12}	3.92×10^{-14}	2.69×10^{-12}	5.49×10^{-14}	2.39×10^{-14}
Vali	Ensemble	3.74×10^{-13}	6.17×10^{-13}	1.73×10^{-12}	1.11×10^{-13}	1.03×10^{-13}
	Dropout	4.10×10^{-13}	1.20×10^{-13}	1.83×10^{-12}	1.35×10^{-13}	1.08×10^{-13}
	NRLMSISE	1.02×10^{-12}	5.52×10^{-14}	5.65×10^{-12}	1.33×10^{-13}	5.95×10^{-14}
	JB08	5.80×10^{-12}	4.04×10^{-14}	5.43×10^{-12}	8.38×10^{-14}	4.67×10^{-14}

5 Conclusions

This paper presents a comparison of two data-driven models for estimating thermospheric density and associated uncertainties: dropout and deep ensemble models. Post-processed orbital data from five well-calibrated satellites and space weather indices are leveraged as input for the models. The trained models are visually tested on a 7-hour time span for a single satellite, highlighting improved performances with respect to the two operational baselines. The dropout network is observed to produce larger uncertainties for the investigated time span with respect to the deep ensemble, although

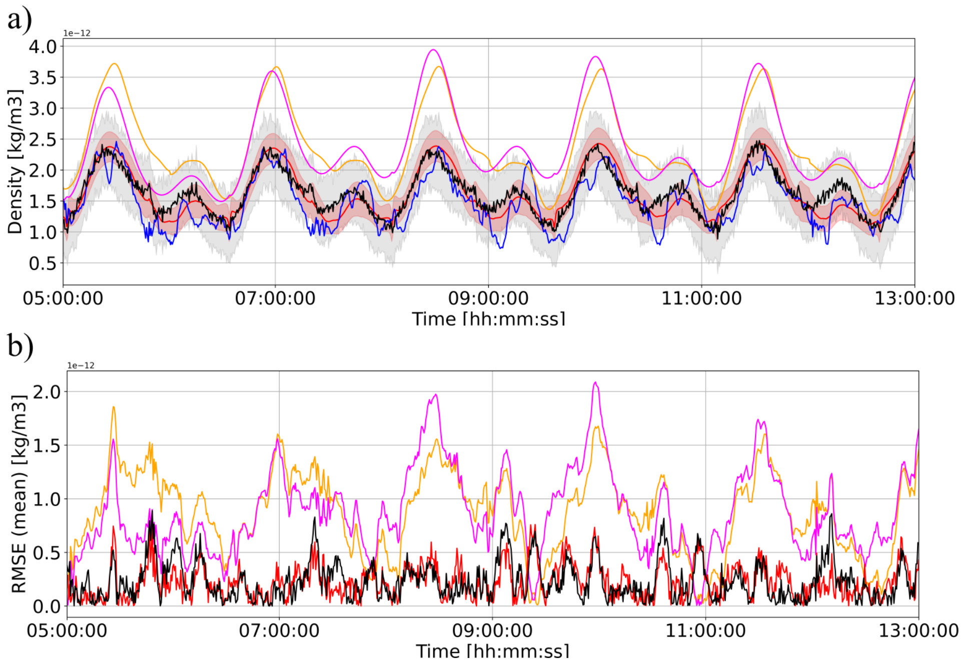

Figure 2: Trend of a) estimated thermospheric density and b) RMSEs over the time span Jan 15, 2006 from 5:00:00 until 13:00:00 UTC for CHAMP satellite. Ground truth in blue, output of empirical models NRLMSISE-00 and JB08 in orange and magenta, respectively. Dropout and ensemble models in black and red, respectively, with associated uncertainties as semi-transparent regions.

often bounding the ground truth. Average RMSE of the trained and operational models are then computed for each satellite on the test and validation sets, highlighting superior performances of the trained models for a subset of the included satellites. The presented approach provides the space operator with two methods for estimating thermospheric density and interpreting the confidence in the prediction. Future studies are expected to target improved performances of the generated models by including orbital data from other space missions. This work will ultimately lead towards a faster and more accurate global model for thermospheric density estimation, improving spacecraft operational routines, and giving the operators confidence in leveraging black-box-like models for estimating thermospheric density.

6 Acknowledgments

This work has been enabled by the Frontier Development Lab (FDL.ai). FDL is a co-operative agreement between NASA, the SETI Institute (seti.org), and Trillium Technologies Inc, in partnership with Google Cloud and Intel. We would like to thank Drs. L. Guhathakurta, M. Jin, J. van den Ijssel, E. Doornbos, A. Muñoz-Jaramillo, A. Vourlidas, I. Telezhinsky, and T.S. Kelso for sharing their technical expertise and James Parr, Jodie Hughes, and Belina Raffy for their support.

References

[1] A. F. Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018.

[2] T. Berger, M. Holzinger, E. Sutton, and J. Thayer. Flying Through Uncertainty. *Space Weather*, 18(1), 2020. doi: 10.1029/2019SW002373.

[3] B. Bowman, W. Tobiska, F. Marcos, C. Huang, C. Lin, and W. Burke. A New Empirical Thermospheric Density Model JB2008 Using New Solar and Geomagnetic Indices. *AIAA/AAS Astrodynamics Specialist Conference and Exhibit, Honolulu, HI*, 2008. doi: 10.2514/6.2008-6438.

[4] C. D. Bussy-Virat, A. J. Ridley, and J. W. Getchius. Effects of uncertainties in the atmospheric density on the probability of collision between space objects. *Space Weather*, 16(5):519–537, 2018.

[5] T. Dietterich. Ensemble methods in machine learning, 2000. In: Multiple classifier systems.

[6] E. Doornbos. Thermosphere density and wind determination from satellite dynamics, 2012. Dissertation, Springer Nature, Switzerland. ISBN 978-3-642-25129-0.

[7] S. Elvidge, H. Godinez, and M. J. Angling. Improved forecasting of thermospheric densities using multi-model ensembles. *Geosci. Model Dev.*, 9:2279–2292, 2016. doi: 10.5194/gmd-9-2279-2016.

[8] Y. Gal and Z. Ghahramani. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning, 2016. arXiv:1506.02142.

[9] Y. Gal and Z. Ghahramani. Bayesian Convolutional Neural Networks with Bernoulli Approximate Variational Inference, 2016. arXiv:1506.02158.

[10] Y. Gal and Z. Ghahramani. A Theoretically Grounded Application of Dropout in Recurrent Neural Networks, 2016. arXiv:1512.05287v5.

[11] T. George. The Use of Long Short-Term Memory Artificial Neural Networks for the Global Prediction of Atmospheric Density, 2020. MS Thesis, Unviersity of Kansas, Lawrence, KS.

[12] J. Guerra, S. Murray, and E. Doornbos. The Use of Ensembles in Space Weather Forecasting. *Space Weather*, 18(2), 2020. doi: 10.1029/2020SW002443.

[13] C. He, Y. Yanga, B. Carter, E. Kerr, S. Wua, F. Deleflie, H. Cai, K. Zhanga, L. Sagniéres, and R. Norman. Review and comparison of empirical thermospheric mass density models. *Progress in Aerospace Sciences journal*, 103:31–51, 2018. doi: 10.1016/j.paerosci.2018.10.003.

[14] L. V. Jospin, H. Laga, F. Boussaid, W. Buntine, and M. Bennamoun. Hands-on Bayesian Neural Networks – a Tutorial for Deep Learning Users. *arXiv preprint arXiv:2007.06823*, 2020.

[15] J.T. Emmert. Thermospheric mass density: A review. *Advances in Space Research*, (56(5)): 773–824, 2015.

[16] D. J. Kessler, N. L. Johnson, J. C. Liou, and M. Matney. The Kessler syndrome: Implications to future space operations. *Advances in the Astronautical Sciences*, 137(8), 2010.

[17] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization, 2017.

[18] B. Lakshminarayanan, A. Pritzel, and C. Blundell. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles, 2017. arXiv:1612.01474v3.

[19] P. Mehta and R. Linares. A methodology for reduced order modeling and calibration of the upper atmosphere. *Space Weather*, 15(10):1270–1287, 2017. doi: 10.1002/2017SW001642.

[20] S. Morley, D. Welling, and J. Woodroff. Perturbed Input Ensemble Modeling With the SpaceWeather Modeling Framework. *Space Weather*, 16:1330–1347, 2018. doi: 10.1029/2018SW002000.

[21] S. Murray. The Importance of Ensemble Techniques for Operational Space Weather Forecasting. *Space Weather*, 16:777–783, 2018. doi: 10.1029/2018SW001861.

[22] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. URL `http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf`.

[23] J. Picone, A. Hedin, D. P. Drob, and A. Aikin. NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues. *Journal ofGeophysical Research*, 107(A12), 2008. doi: 10.1117/12.2321712.

[24] F. Pinto, G. Acciarini, S. Metz, S. Boufelja, S. Kaczmarek, K. Merz, J. A. Martinez-Heras, F. Letizia, C. Bridges, and A. G. Baydin. Towards automated satellite conjunction management with Bayesian deep learning. In *AI for Earth Sciences Workshop at NeurIPS 2020, Vancouver, Canada*, 2020.

[25] D. Pérez and R. Bevilacqua. Neural network based calibration of atmospheric density models. *Acta Astronautica*, 110:58–76, 2015. doi: 10.1016/j.actaastro.2014.12.018.

[26] D. Pérez, B. Wohlberg, T. Lovell, and R. Bevilacqua. Orbit-centered atmospheric density prediction using artificial neural networks. *Acta Astronautica*, 98(1):9–23, May 2014. doi: 10.1016/j.actaastro.2014.01.007.

[27] C. Siemes, J. de Teixeira da Encarnação, E. Doornbos, J. van den IJssel, J. Kraus, R. Pereštý, L. Grunwaldt, G. Apelbaum, J. Flury, and P. Holmdahl Olsen. Swarm accelerometer data processing from raw accelerations to thermospheric neutral densities. *Earth, Planets and Space*, 68, 2016.

[28] E. Stevenson, V. Rodriguez-Fernandez, E. Minisci, and D. Camacho. A deep learning approach to solar radio flux forecasting. *Acta Astronautica journal*, 2021. doi: 10.1016/j.actaastro.2021.08.004.

[29] D. Vallado and D. Finkleman. A critical assessment of satellite drag and atmospheric density modeling. *Acta Astronautica*, 95:141–165, 2014. doi: 10.1016/j.actaastro.2013.10.005.

[30] J. van den IJssel, E. Doornbos, E. Iorfida, G. March, C. Siemes, and O. Montenbruck. Thermosphere densities derived from Swarm GPS observations. *Advances in Space Research*, 65(7): 1758–1771, 2020. doi: 10.1016/j.asr.2020.01.004.

[31] P. Visser and J. van den IJssel. Orbit determination and estimation of non-gravitational accelerations for the GOCE reentry phase. *Advances in Space Research*, 58(9):1840–1853, 2016.

[32] A. Vourlidas and S. Bruinsma. EUV Irradiance Inputs to Thermospheric Density Models: Open Issues and Path Forward. *Space Weather*, 16(1):5–15, 2018. doi: 10.1002/2017SW001725.

[33] K. Wakker. *Fundamentals of Astrodynamics*. Institutional Repository TU Delft, 2015.

[34] A. G. Wilson and P. Izmailov. Bayesian Deep Learning and a Probabilistic Perspective of Generalization. *arXiv preprint arXiv:2002.08791*, 2020.