



# BiMA: Bias-Corrected LightGBM Model for Atmospheric Density Forecasting



Team Bimasakti

Date: May 24th, 2025

**Rasyid Ridha**

Independent Researcher  
Jakarta, Indonesia  
rasyidstat@gmail.com

## Abstract

This report presents BiMA, a Bias-Corrected LightGBM Model for Atmospheric Density Forecasting. BiMA refines the atmospheric density outputs of the recent empirical model, NRLMSIS 2.1, by training an ensemble of LightGBM models on two distinct target representations: ratio and log-ratio between observed and empirical density. This approach combines the generalization capabilities of physically-informed empirical models while mitigating their limitations in capturing complex atmospheric dynamics, particularly during geomagnetic storms. BiMA achieves significant performance improvements compared to the benchmark, with an error reduction of 44% on unseen test dataset, highlighting its strong generalization and accuracy.

## Model Report

### Model Overview

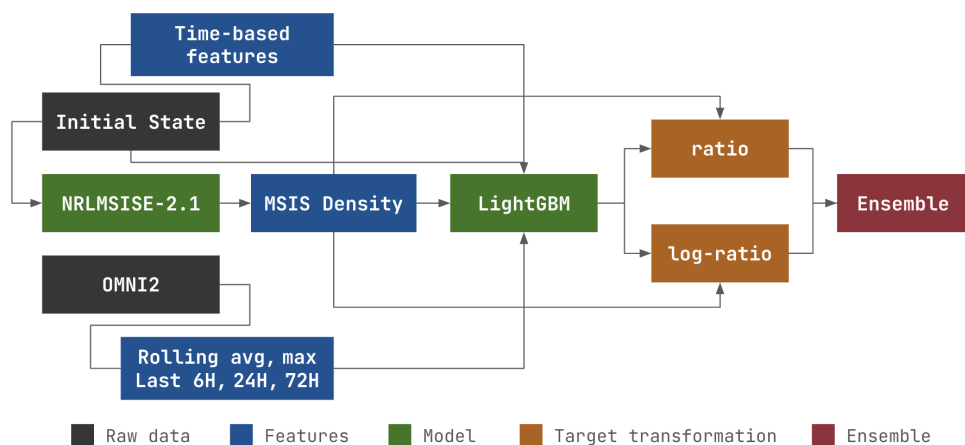


Figure 1: BiMA Architecture Diagram

BiMA uses a hybrid approach, combining the strengths of both physics-informed models and data-driven methods to achieve robust and accurate atmospheric density forecasts. Physics-informed models are grounded in established physical principles but often struggle to capture complex atmospheric dynamics, particularly during geomagnetic storms. On the other hand, purely data-driven models, while good at learning from training data, may lack generalization in unseen scenarios, especially if the dataset size is limited.

The core principle of BiMA is to learn the systematic bias in the atmospheric density output from the recent empirical model, NRLMSIS 2.1 [1], an extension of NRLMSIS 2.0 [2]. Instead of directly predicting atmospheric density, BiMA refines the outputs of the empirical model by training data-driven models on top of them. This approach, similar to that implemented in [3], effectively "calibrates" or "bias-corrects" the MSIS output.

Two models with distinct target representations are used for training: the ratio between observed density and empirical MSIS density ( $\rho_{\text{obs}}/\rho_{\text{msis}}$ ), and the log-transformed version,  $\ln(\rho_{\text{obs}}/\rho_{\text{msis}} + 1)$ , which helps normalize the distribution. LightGBM [4] was selected as the regressor for these models, due to its ability to capture nonlinear relationships, speed, flexibility, and reproducibility. During inference, predictions from these models are then transformed back, and an average ensemble of both is used to make the result more robust across varied conditions.

A novel aspect of BiMA is its ability to act as a data-driven propagator, refining the initial state density output for each forecast horizon (432 time steps, spanning 3 days at 10-minute intervals). This is necessary as empirical models like MSIS lack a built-in propagator. Instead of using computationally intensive physical orbit propagators, BiMA provides a faster and more resource-efficient solution.

## Data and Preprocessing

BiMA was trained using the STORM-AI Phase 1 Public Training Dataset [5], which includes satellite initial state, OMNI2 space weather data [6], and observed atmospheric density from five satellites: CHAMP, GRACE1, GRACE2, GRACE-FO1 and SWARM-A [7].

Empirical atmospheric density estimates were computed with `pymssis` library [8]. Specifically, the estimates are based on the latest empirical model, NRLMSIS 2.1 [1], using data inputs from CelesTrak [9], [10], and satellite position at the initial state. These estimates were used both as input and reference values for transforming the target variable (observed density), using two representations mentioned in [Model Overview](#).

A total of 139 input variables were prepared for model training, grouped into the following categories:

- **Satellite Initial State:** This includes raw orbital elements and positional data (latitude, longitude, altitude). Any missing position values were imputed using forward-fill based on known previous state.
- **Time-based Features:** Hour and day-of-year, in both UTC and local solar time, were encoded using sine and cosine transformations to capture cyclical patterns.
- **OMNI2 Space Weather Features:** Rolling aggregations (mean and maximum for last 6, 24, and 72 hours window) of solar activity and geomagnetic indices capture short and medium-term effects. Any missing values remaining after the aggregations were kept as null and handled by LightGBM's internal capability.
- **NRLMSIS 2.1 Atmospheric Density:** The MSIS density estimates at the initial state (computed as described above) were explicitly included as an input feature.
- **Forecast Horizon Encoding:** A numeric indicator of the forecast step (ranging from 1 to 432 for 72 hours at 10-minute intervals) allows the model to act as a propagator which learns forecast horizon specific-biases.

## Training and Validation

The key to building a strong model lies in choosing an appropriate data split. Initially, a standard k-fold approach was tested to evaluate model performance. However, this setup can lead to overestimated performance, as the model is prone to overfitting from time leakage within the same satellite. As a result, the model may memorize values and time instead of learning meaningful relationships with the input variables. To address this, cross-validation with satellite split was chosen to ensure the model is capable of forecasting density for entirely different satellites, locations, and altitude ranges, ensuring model generalization and reducing overfitting.

In addition, the latest year of data for each satellite-fold was held out as a supplemental evaluation to assess performance on future set, simulating real-world deployment. Both GRACE1 and GRACE2 were included in the same fold due to GRACE1 sample size limitation and the identical nature of both missions. This results into four cross-validated model training runs evaluated on CHAMP, GRACE1/GRACE2, GRACE-FO1, and SWARM-A.

The primary metric to assess model performance is OD-RMSE, aligning with the public leaderboard calculation. However, the baseline reference model used is NRLMSIS 2.1 without propagator. Evaluating the model per satellite and then averaging OD-RMSE ensures the performance metric is scale-invariant. It treats different satellites equally regardless of their absolute density magnitudes, allowing for a better holistic understanding of the forecast performance gain.

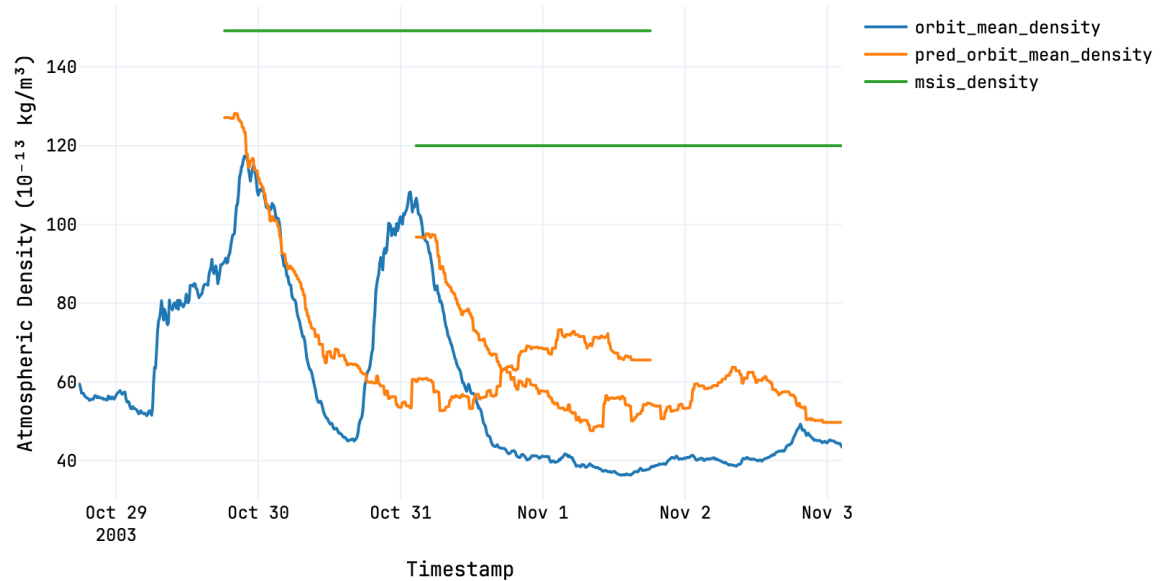
For model training, two models, as described in [Model Overview](#), were developed using an identical setup with the objective of minimizing MSE loss. Hyperparameter tuning and feature engineering choices were based on cross-validation performance, emphasizing stability and consistency across satellite performance scores. Finally, an ensemble of models trained on the full dataset was used for inference.

## Model Performance

**Table 1: Model Performance Summary**  
CH = CHAMP, G1 = GRACE-1, G2 = GRACE-2, GF1 = GRACE-FO1, SWA = SWARM-A

Model	Satellite-Fold CV						Leaderboard		
	CH	G1	G2	GF1	SWA	Avg	Public	Private	Phase I
ratio	0.361	<b>0.597</b>	<b>0.497</b>	0.111	0.332	0.380	0.743	0.408	0.693
log-ratio	<b>0.459</b>	0.569	0.460	<b>0.182</b>	<b>0.408</b>	<b>0.415</b>	0.744	0.439	0.699
ensemble	0.420	0.593	0.490	0.152	0.379	0.407	<b>0.750</b>	<b>0.444</b>	<b>0.704</b>

BiMA consistently outperforms the MSIS benchmark, achieving OD-RMSE of 0.75 on the public test and SSAOD-RMSE of 0.44 on the private test. GRACE-FO1 shows the least improvement, likely because its higher altitude and lower density are more challenging to model, given that the training data is mostly from lower altitudes. Overall, the log-ratio model performs best, particularly for CHAMP, which have highly skewed density values. While the ensemble only adds a small improvement on the leaderboard, it provides more balanced and consistent performance across satellites.



**Figure 2: Model Forecasts for CHAMP satellite during 2003 Halloween Solar Storms**

Figure 2 shows the model’s capability to forecast atmospheric density under extreme conditions even though CHAMP satellite data was not part of the training data within the cross-validation split. This highlights the model’s generalization strength, which is valuable under limited training data points and extreme space weather conditions.

## Limitations and Future Work


While BiMA achieves significant performance improvements in atmospheric density forecasting, several limitations and future research areas exist:

- Deep learning models, such as LSTMs or Transformers [11] could potentially improve forecast accuracy, especially at longer horizons. Their inherent ability to capture temporal dependencies and complex patterns in atmospheric dynamics, including responses to geomagnetic storm events, might be better suited than tree-based models.
- BiMA currently does not integrate any physics-based propagator, which might lead to physically implausible forecasts, especially at longer forecast horizons. Future work could explore physics-informed machine learning to ensure that the forecast adheres to physical principles.
- Including more satellite in-situ data points, especially from varied altitudes, or using synthetic data, could further improve generalization.
- Incorporating uncertainty estimation [3] and forecast explainability would provide valuable guidance for operational use.

## Conclusion

This work highlights BiMA as a fast, robust and accurate model for forecasting atmospheric density, combining the strengths of physics-informed models with machine learning correction. The model performs well across a wide range of conditions, including during extreme space weather, and generalizes effectively to satellites and altitudes not seen during training. BiMA also generates forecasts more quickly as it does not rely on physical orbit propagators, making it suitable for real-time forecasting applications. These qualities support its use in operational settings where timely updates are essential, helping reduce risks such as satellite drag and orbital collisions.

## Code Availability

All code used to generate the results in this report, including data preprocessing, feature engineering, model training, and evaluation, will be released upon competition on  [BiMA GitHub repository](#).

## References

- 
- [1] J. T. Emmert et al., “NRLMSIS 2.1: An Empirical Model of Nitric Oxide Incorporated Into MSIS,” *Journal of Geophysical Research: Space Physics*, 2022. DOI: [10.1029/2022JA030896](https://doi.org/10.1029/2022JA030896).
  - [2] J. T. Emmert et al., “NRLMSIS 2.0: A whole-atmosphere empirical model of temperature and neutral species densities,” *Earth and Space Science*, 2020. DOI: [10.1029/2020EA001321](https://doi.org/10.1029/2020EA001321).
  - [3] R. J. Licata, P. M. Mehta, D. R. Weimer, W. K. Tobiska, and J. Yoshii, “MSIS-UQ: Calibrated and Enhanced NRLMSIS 2.0 Model With Uncertainty Quantification,” *Space Weather*, 2022. DOI: [10.1029/2022SW003267](https://doi.org/10.1029/2022SW003267).
  - [4] G. Ke et al., “LightGBM: A highly efficient gradient boosting decision tree,” *Advances in neural information processing systems*, vol. 30, pp. 3146–3154, 2017. DOI: [10.5555/3294996.3295074](https://doi.org/10.5555/3294996.3295074).
  - [5] *STORM-AI Dataset*, <https://2025-ai-challenge.readthedocs.io/en/latest/dataset.html>.
  - [6] *OMNI2 space weather dataset (NASA Goddard Space Flight Center)*, <https://omniweb.gsfc.nasa.gov/form/dx1.html>.
  - [7] *SWARM dataset (ESA)*, <https://earth.esa.int/eogateway/missions/swarm>.
  - [8] G. Lucas, *pymsis*, Computer software, 2022. DOI: [10.5281/zenodo.5348502](https://doi.org/10.5281/zenodo.5348502).
  - [9] *Celestrak*, <https://celestrak.org/SpaceData/>.
  - [10] J. Matzka, C. Stolle, Y. Yamazaki, O. Bronkalla, and A. Morschhauser, “The geomagnetic kp index and derived indices of geomagnetic activity,” *Space Weather*, 2021. DOI: [10.1029/2020SW002641](https://doi.org/10.1029/2020SW002641).
  - [11] J. Briden, P. M. Siew, V. Rodriguez-Fernandez, and R. Linares, “Transformer-based atmospheric density forecasting,” *Advanced Maui Optical and Space Surveillance (AMOS) Technologies Conference*, 2023. [Online]. Available: <https://arxiv.org/abs/2310.16912>.