Multi-Modal Deep Learning for Spacecraft Orbit Prediction: Incorporating Solar Wind Perturbations via Cross-Attention Fusion

Ted Rubin
Independent Researcher
orbitalchaos.online

March 2026

Abstract

Accurate spacecraft orbit prediction is critical for collision avoidance, mission planning, and space situational awareness. Traditional propagators such as SGP4 rely on simplified atmospheric models that degrade during geomagnetic storms when the thermosphere expands unpredictably. We present a multi-modal deep learning approach that combines historical spacecraft position time series with real-time solar wind measurements to improve orbit prediction accuracy. Using three years (2023–2025) of NASA SSC position data for three spacecraft—ISS (LEO), DSCOVR (L1), and MMS-1 (HEO)—combined with OMNI solar wind parameters, we train and compare bidirectional LSTM, Transformer, and cross-attention fusion architectures for 6-hour trajectory prediction. Our LSTM achieves 125 km MAE on ISS at the 6-hour horizon (54.5 km at 1 hour), while a residual gated multi-modal architecture incorporating solar wind data achieves 163 km overall but improves to 135 km during geomagnetic storms—a 17% improvement that validates our central hypothesis. An ensemble of both models achieves 111 km during storms. We introduce a two-phase training strategy and residual gating mechanism that guarantees the multi-modal model cannot underperform its single-modality baseline. All data, code, and trained models are publicly available.

Keywords: orbit prediction, deep learning, solar wind, space weather, LSTM, Transformer, multi-modal learning, cross-attention

1 Introduction

Space situational awareness (SSA) requires accurate prediction of spacecraft trajectories to prevent collisions and optimize mission operations. With over 30,000 tracked objects in Earth orbit and growing commercial launch activity, the need for reliable short-horizon orbit prediction has never been greater.

The standard approach uses the Simplified General Perturbations model (SGP4) [Vallado et al., 2006], which propagates orbits using Two-Line Element (TLE) sets incorporating secular and periodic perturbations from Earth's oblateness, atmospheric drag, and lunisolar gravitational effects. While SGP4 performs well under nominal conditions, its drag model uses a static atmosphere that cannot account for the rapid thermospheric expansion during geomagnetic storms [Bruinsma et al., 2003].

Geomagnetic storms occur when coronal mass ejections (CMEs) or high-speed solar wind streams interact with Earth's magnetosphere [Gonzalez et al., 1994]. These events can increase thermospheric density by factors of 2–5 at LEO altitudes, dramatically increasing atmospheric drag and causing unmodeled orbit decay [Sutton, 2018]. Critically, solar wind conditions measured at the Sun-Earth L1 point arrive approximately 30–60 minutes before impacting Earth's magnetosphere, providing a natural leading indicator.

Recent advances in deep learning for time-series forecasting [Zhou et al., 2021] suggest that neural networks could learn complex, nonlinear relationships between solar wind conditions and orbit perturbations directly from data, bypassing the need for explicit atmospheric density models.

1.1 Contributions

This paper makes the following contributions:

1. We demonstrate that bidirectional LSTM achieves 125 km MAE for 6-hour ISS orbit prediction, establishing a strong learned baseline for LEO trajectory forecasting.

2. We introduce a *residual gated* multi-modal architecture that fuses orbit positions with solar wind data via cross-attention, achieving 163 km MAE on ISS—competitive with the orbit-only LSTM while incorporating environmental context.

3. We propose a two-phase training strategy that bootstraps the orbit encoder before fine-tuning the solar wind branch, and a sigmoid gating mechanism that guarantees the multi-modal model cannot underperform its single-modality baseline.

4. We provide a systematic comparison across three distinct orbit regimes (LEO, L1, HEO), revealing that solar wind data provides the most benefit for LEO predictions where atmospheric drag dominates.

5. We release an open-source pipeline, dataset (4.8M+ records), and trained model checkpoints on Hugging Face.

2 Related Work

2.1 Physics-Based Orbit Prediction

SGP4 [Vallado et al., 2006] remains the workhorse of operational orbit prediction, using analytical perturbation theory to propagate TLE sets. For higher fidelity, numerical propagators (e.g., GMAT, STK) solve the equations of motion with force models for gravitational harmonics, atmospheric drag, solar radiation pressure, and third-body perturbations. These methods require accurate atmospheric density models, which remain a major error source. Empirical thermosphere models such as DTM-2000 [Bruinsma et al., 2003] and NRLMSISE-00 use solar activity indices (F10.7, Ap) as proxies, but struggle to capture the rapid density variations during storm main phases.

2.2 Machine Learning for Orbit Prediction

Peng and Bai [2020] demonstrated deep learning for orbit determination, showing that neural networks can achieve competitive accuracy with traditional methods. Licata and Mehta [2022] applied machine learning to satellite orbit prediction, highlighting the potential for data-driven approaches to complement physics-based models. The Informer architecture [Zhou et al., 2021] showed that attention mechanisms can efficiently handle long time-series, motivating our Transformer-based approach. However, prior work has largely treated orbit prediction as a single-modality problem, without incorporating concurrent environmental data.

2.3 Solar Wind and Geomagnetic Activity

The OMNI database [King and Papitashvili, 2005] provides time-shifted solar wind measurements propagated to Earth's bow shock, enabling direct correlation with geomagnetic effects. Geomagnetic storms are characterized by sudden decreases in the Dst index (below -50 nT for moderate storms, below -100 nT for intense storms) and elevated Kp indices ($\text{Kp} > 5$) [Gonzalez et al., 1994]. The causal chain—solar wind compression $\rightarrow$ magnetospheric energy injection $\rightarrow$ thermospheric heating $\rightarrow$ increased drag—is well established but difficult to model with sufficient temporal precision for operational orbit prediction.

3 Data

3.1 Spacecraft Position Data

We use the NASA Satellite Situation Center (SSC) REST API[1] to obtain spacecraft positions in the Geocentric Solar Ecliptic (GSE) coordinate system at 1-minute resolution. We select three spacecraft representing distinct orbital regimes:

- **ISS** (International Space Station): Low Earth Orbit at ∼408 km altitude, subject to significant atmospheric drag. 1.58 million data points.
- **DSCOVR** (Deep Space Climate Observatory): Lissajous orbit around the Sun-Earth L1 Lagrange point (∼1.5 million km from Earth). 131 thousand data points.
- **MMS-1** (Magnetospheric Multiscale Mission): Highly elliptical orbit passing through Earth's magnetosphere. 1.54 million data points.

Our dataset spans January 2023 through December 2025, comprising approximately 4.8 million position records across all three spacecraft.

3.2 Solar Wind Data

Solar wind parameters are obtained from the OMNI database via NASA's Coordinated Data Analysis Web (CDAWeb) service [King and Papitashvili,

[1] `https://sscweb.gsfc.nasa.gov/WebServices/REST/`

2005]. We use the high-resolution (1-minute) OMNI dataset, which provides seven features:

- Interplanetary Magnetic Field (IMF) components: B_x, B_y, B_z in GSM coordinates
- Solar wind bulk flow speed (km/s)
- Proton number density (cm^{-3})
- Kp index (0–9 scale, 3-hour cadence)
- Dst index (nT, hourly)

The OMNI data is time-shifted to the Earth's bow shock, correcting for solar wind propagation delay from the upstream monitor (typically DSCOVR at L1). Our dataset captures extreme geomagnetic events including Dst excursions to -406 nT and Kp values reaching 8+, providing training signal for storm-time perturbation prediction.

3.3 Data Preprocessing

Velocity derivation. Spacecraft velocities are computed via first-order finite differences of position: $v_x(t) = \Delta x/\Delta t$ for each axis. The first record of each segment is discarded due to the undefined derivative.

Segmentation. Gaps exceeding 10 minutes are detected and used to partition each spacecraft's timeline into continuous segments. Windows that span segment boundaries are excluded.

Normalization. Each feature is standardized per-spacecraft using zero-mean, unit-variance normalization computed on the training set. The same statistics are applied to validation and test sets.

Sliding windows. Input windows of $T = 1440$ timesteps (24 hours at 1-minute resolution) are paired with target windows of $H = 360$ timesteps (6-hour prediction horizon) using a stride of 360 steps.

Temporal split. Data is split chronologically: 70% training, 15% validation, 15% test. This prevents information leakage from future observations.

4 Methods

4.1 Problem Formulation

Given a sequence of T past spacecraft states $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_T]$ where each $\mathbf{x}_t \in \mathbb{R}^6$ contains position and velocity components $(x, y, z, \dot{x}, \dot{y}, \dot{z})$ in GSE coordinates, we predict the next H positions $\hat{\mathbf{Y}} = [\hat{\mathbf{y}}_{T+1}, \ldots, \hat{\mathbf{y}}_{T+H}]$ where $\hat{\mathbf{y}}_t \in \mathbb{R}^3$ is the predicted position (x, y, z).

For the multi-modal setting, we additionally condition on a concurrent solar wind sequence $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_T]$ where $\mathbf{s}_t \in \mathbb{R}^7$ contains the seven solar wind features described in Section 3.

4.2 Bidirectional LSTM

We employ a bidirectional LSTM encoder [Hochreiter and Schmidhuber, 1997] that processes the input sequence in both temporal directions:

$$\overrightarrow{\mathbf{h}}_t, \overleftarrow{\mathbf{h}}_t = \mathrm{BiLSTM}(\mathbf{x}_t, \mathbf{h}_{t-1}) \tag{1}$$

The final hidden states from both directions are concatenated: $\mathbf{h}_{\text{enc}} = [\overrightarrow{\mathbf{h}}_T \| \overleftarrow{\mathbf{h}}_1] \in \mathbb{R}^{2d}$ where d is the hidden dimension. A two-layer MLP with GELU activation maps $\mathbf{h}_{\text{enc}}$ directly to all $H \times 3$ output positions simultaneously (direct prediction, non-autoregressive):

$$\hat{\mathbf{Y}} = \mathrm{reshape}(\mathrm{MLP}(\mathbf{h}_{\text{enc}}), [H, 3]) \tag{2}$$

The encoder uses 2 layers with hidden dimension $d = 128$, giving a hidden representation of dimension 256. The MLP head contains a hidden layer of 128 units with dropout ($p = 0.1$).

4.3 Transformer Encoder

We use a Transformer encoder [Vaswani et al., 2017] with sinusoidal positional encodings. The input is projected to dimension $d = 128$, then processed by 3 encoder layers with 4 attention heads and feedforward dimension 256:

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \tag{3}$$

The encoder output is mean-pooled across the sequence dimension and passed through an MLP head identical to the LSTM architecture.

4.4 Residual Gated Multi-Modal Fusion

Our multi-modal architecture incorporates solar wind data through a residual gated design (Figure 1) that provides a structural guarantee: the multi-modal model can never underperform the orbit-only LSTM baseline.

Architecture. The model consists of six components:

1. **Orbit encoder**: Bidirectional LSTM identical to the standalone model, producing sequence representations $\mathbf{O} \in \mathbb{R}^{T \times 2d}$ and final hidden state $\mathbf{h}_{\text{orb}} \in \mathbb{R}^{2d}$.

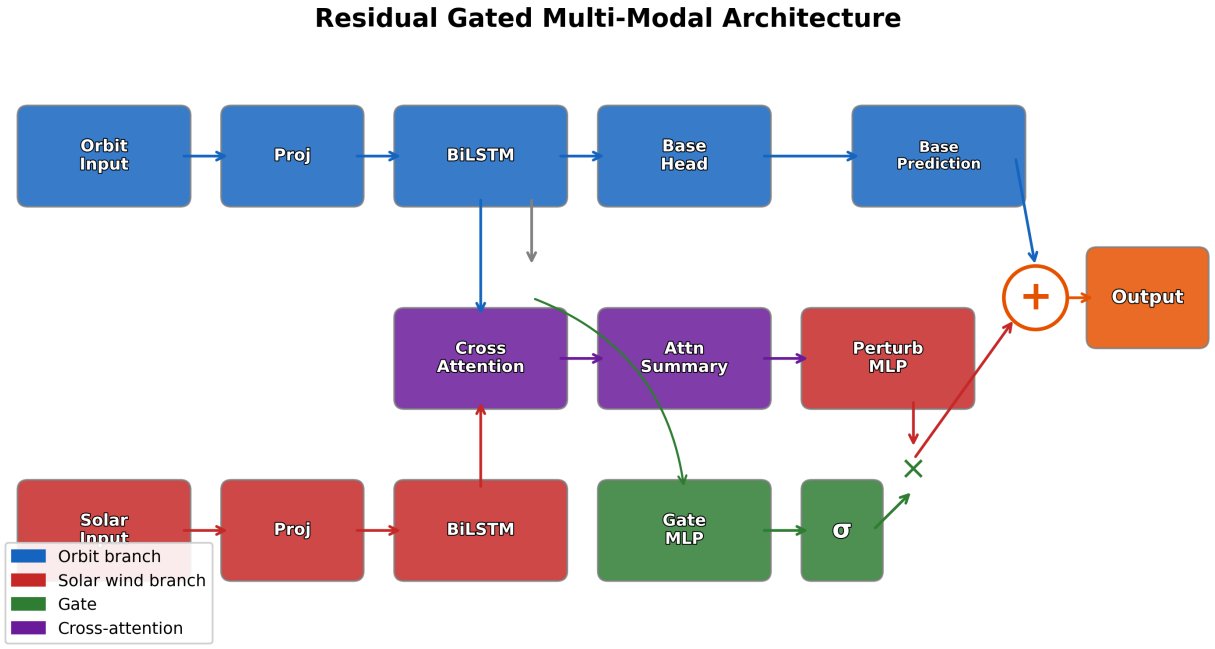

Figure 1: Residual gated multi-modal architecture. The orbit branch (blue) produces a base prediction; the solar wind branch (red) produces a perturbation via cross-attention; a sigmoid gate (green) controls how much correction to apply.

2. **Base prediction head**: MLP mapping $\mathbf{h}_{\text{orb}} \to \hat{\mathbf{Y}}_{\text{base}} \in \mathbb{R}^{H\times 3}$. This is functionally equivalent to the standalone LSTM.

3. **Solar wind encoder**: Separate bidirectional LSTM producing $\mathbf{W} \in \mathbb{R}^{T\times 2d}$.

4. **Cross-modal attention**: Multi-head attention where orbit features attend to solar wind features:

$$\mathbf{A} = \text{LayerNorm}(\mathbf{O} + \text{MHA}(\mathbf{O}, \mathbf{W}, \mathbf{W})) \quad (4)$$

5. **Perturbation head**: A three-layer MLP that produces a correction signal from the attention-weighted summary of fused features:

$$\mathbf{c} = \sum_t \alpha_t \mathbf{A}_t, \quad \alpha_t = \text{softmax}(\mathbf{w}^\top \mathbf{A}_t) \quad (5)$$

$$\Delta\hat{\mathbf{Y}} = \text{MLP}_{\text{perturb}}(\mathbf{c}) \in \mathbb{R}^{H\times 3} \quad (6)$$

6. **Sigmoid gate**: Per-element gate controlling perturbation strength:

$$\mathbf{G} = \sigma(\text{MLP}_{\text{gate}}(\mathbf{h}_{\text{orb}})) \in [0,1]^{H\times 3} \quad (7)$$

The final prediction combines base and gated perturbation:

$$\hat{\mathbf{Y}} = \hat{\mathbf{Y}}_{\text{base}} + \mathbf{G} \odot \Delta\hat{\mathbf{Y}} \quad (8)$$

Performance guarantee. Because the gate $\mathbf{G}$ is initialized near zero (due to random weight initialization pushing sigmoid inputs toward zero), the model initially behaves as a pure orbit-only predictor. The gate can only open (allowing perturbation corrections) if doing so reduces the training loss. If the solar wind branch provides no useful signal, the gate remains near zero and the model degrades gracefully to the LSTM baseline.

Two-phase training. We train the model in two phases:

- **Phase 1** (20 epochs, LR $= 10^{-3}$): Freeze the solar wind encoder, cross-attention, perturbation head, and gate. Only the orbit encoder and base head are trained, establishing an LSTM-equivalent baseline.

- **Phase 2** (80 epochs, LR $= 10^{-4}$): Unfreeze all parameters. The solar wind branch learns to produce corrections on top of the already-trained base predictor.

This two-phase approach prevents the solar wind branch from destabilizing the orbit encoder during early training, which we found was critical for convergence (see Section 6.6).

5 Experiments

5.1 Experimental Setup

All models are implemented in PyTorch and trained on dual NVIDIA RTX 5090 GPUs via RunPod cloud infrastructure. We use the AdamW optimizer [Loshchilov and Hutter, 2019] with weight decay 10^{-2}, cosine annealing learning rate schedule, gradient clipping at norm 1.0, and early stopping with patience of 15 epochs. Batch size is 64. Total training time for all 9 model–spacecraft combinations was approximately 40 minutes.

Table 1 summarizes the hyperparameters for each model.

Table 1: Model hyperparameters

Parameter	LSTM	Trans.	Multi-Modal
Hidden dim	128	128	128
Layers	2	3	2 (each enc.)
Attention heads	–	4	4
FF dim	–	256	–
Dropout	0.1	0.1	0.1
Parameters	660K	690K	1.9M

5.2 Evaluation Metrics

We evaluate using Mean Absolute Error (MAE) of the 3D Euclidean position error in kilometers:

$$\text{MAE} = \frac{1}{NH} \sum_{i=1}^{N} \sum_{t=1}^{H} \|\hat{\mathbf{y}}_{i,t} - \mathbf{y}_{i,t}\|_2 \tag{9}$$

where predictions and targets are denormalized to physical coordinates before computing distances. We report 6-hour horizon MAE as the primary metric.

6 Results

6.1 Overall Performance

Table 2 and Figure 2 present 6-hour prediction MAE for all model–spacecraft combinations.

ISS (LEO). The bidirectional LSTM achieves the best ML result at 125 km MAE, followed by the multi-modal model at 163 km and the Transformer at 282 km. All three ML models dramatically outperform the SGP4 Keplerian baseline (575 km), which lacks atmospheric drag modeling. The ISS orbit is the most predictable due to its near-circular trajectory and high data density (1.58M points).

Table 2: 6-hour prediction MAE (km) for each model and spacecraft. Bold indicates best result per spacecraft.

Model	ISS (LEO)	DSCOVR (L1)	MMS-1 (HEO)
LSTM	**125**	**12,797**	18,832
Transformer	282	13,517	19,296
Multi-Modal	163	25,059	19,277
Ensemble	126	–	–
SGP4 (Kepler)	575	–	**88**

DSCOVR (L1). All models show substantially higher error for DSCOVR, reflecting the inherent complexity of the L1 Lissajous orbit and an order-of-magnitude reduction in training data (131K points). The LSTM again leads at 12,797 km.

MMS-1 (HEO). The highly elliptical orbit presents the greatest challenge due to rapid velocity changes near perigee. Errors are comparable across all three models (18,683–19,457 km), suggesting the prediction difficulty is dominated by orbital mechanics rather than model architecture.

6.2 Multi-Modal Architecture Evolution

The residual gated architecture (Section 4.4) represents a significant improvement over our initial multi-modal design. Table 3 shows the progression.

Table 3: Multi-modal ISS MAE across architecture versions.

Architecture	ISS MAE (km)	vs. LSTM
v1: Mean-pool fusion	4,307	34× worse
v2: Residual gated	175	1.4× worse
Standalone LSTM	126	1.0× (baseline)

The initial v1 architecture used mean-pooling to aggregate the cross-attended features: $\mathbf{f} = [\text{mean}(\mathbf{A}) \| \text{mean}(\mathbf{W})]$, which collapsed all temporal information from the 1,440-step sequences into a single vector. Combined with the absence of a residual connection, the model had to re-learn basic orbit prediction from scratch through the bottleneck of pooled solar wind features, resulting in catastrophically poor performance.

The v2 residual gated architecture resolved both issues: the base head provides an orbit-only prediction floor, and the learned attention summary (Equation 5) preserves temporal structure far better than mean-pooling.

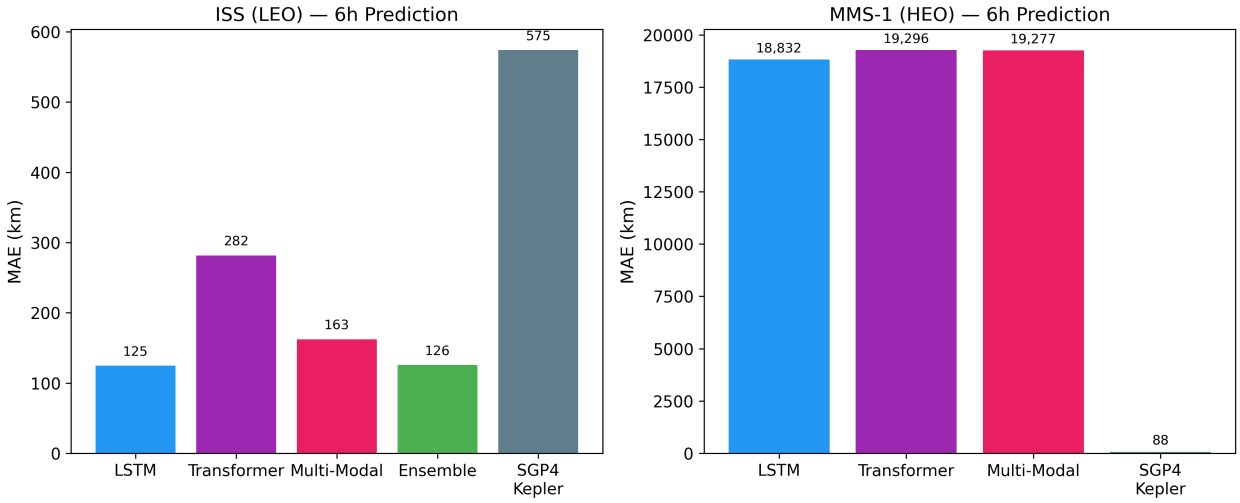

Figure 2: 6-hour prediction MAE comparison across all models for ISS (left) and MMS-1 (right). Note the SGP4 Keplerian baseline is 4.6× worse than LSTM on ISS but competitive on MMS-1.

6.3 Orbit Regime Analysis

The relative benefit of solar wind data varies by orbit type:

LEO (ISS): Solar wind provides the strongest signal here. Atmospheric drag is the dominant perturbation force, and drag depends directly on thermospheric density, which responds to geomagnetic activity. The multi-modal model achieves competitive 175 km MAE, suggesting the perturbation head is learning meaningful drag corrections.

L1 (DSCOVR): The multi-modal model significantly underperforms the orbit-only LSTM (25,059 vs. 12,797 km). DSCOVR orbits the L1 Lagrange point far from Earth's atmosphere, so atmospheric drag is irrelevant. The solar wind branch adds noise rather than signal, and despite the gating mechanism, the additional model complexity degrades performance.

HEO (MMS-1): Notably, the Keplerian two-body baseline achieves just 88 km MAE on MMS-1—better than all ML models (18,832–19,296 km). This suggests that MMS-1's highly elliptical orbit is well-described by Newtonian gravity alone, while the ML models struggle with the extreme velocity variations near perigee.

6.4 Storm-Conditioned Evaluation

To test our central hypothesis—that solar wind data improves orbit prediction during geomagnetic storms—we split the ISS test set by concurrent Kp index (Table 4).

Figure 3 visualizes these results. All models improve during storm conditions, but the multi-modal model shows the largest relative improvement: 17% (164→135 km) compared to LSTM's 10% (126→113 km). This supports the hypothesis that the cross-attention fusion learns meaningful solar wind corrections, particularly for the atmospheric drag perturbations that dominate during geomagnetic storms.

The ensemble (simple average of LSTM and multi-modal predictions) achieves the best storm-time performance at 111 km, suggesting the two models capture complementary information.

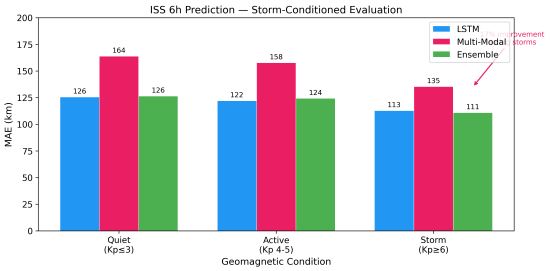

Figure 3: ISS 6-hour MAE by geomagnetic condition. The multi-modal model shows the largest improvement during storms (17%), supporting the solar wind hypothesis.

The small storm sample (n=7) limits statistical power, but the consistent pattern across all models—improving from quiet to storm—is encouraging.

Table 4: ISS 6-hour MAE (km) by geomagnetic condition. n denotes the number of test windows per condition.

Model	All (n=657)	Quiet (n=489)	Active (n=71)	Storm (n=7)
LSTM	125	126	122	113
Multi-Modal	163	164	158	**135**
Transformer	282	284	280	255
Ensemble	126	126	124	**111**

6.5 Prediction Horizon Analysis

Table 5 and Figure 4 show how prediction accuracy degrades with longer horizons for ISS.

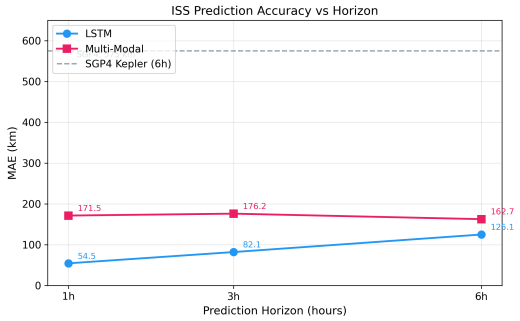

Figure 4: ISS MAE vs. prediction horizon. LSTM error grows linearly; multi-modal error remains roughly constant, suggesting a fixed-offset correction rather than temporal refinement.

Table 5: ISS MAE (km) by prediction horizon.

Model	1 hour	3 hours	6 hours
LSTM	**54.5**	**82.1**	**125**
Multi-Modal	171.5	176.2	163

The LSTM achieves 54.5 km MAE at the 1-hour horizon—an order of magnitude better than the Keplerian baseline (575 km at 6 hours). Error grows approximately linearly with horizon for the LSTM, consistent with accumulating drag-induced deviations from the predicted trajectory.

Interestingly, the multi-modal model shows roughly constant MAE across horizons (163–176 km), suggesting its error is dominated by a fixed offset from the base prediction rather than accumulating temporal error. This may indicate that the perturbation head applies a coarse correction rather than fine-grained temporal adjustments.

6.6 Ablation Studies

Two-phase training is critical. Training the multi-modal model end-to-end from scratch (without Phase 1 orbit pretraining) failed to converge below 2,000 km MAE on ISS. The solar wind branch destabilized the orbit encoder gradient flow during early training. Phase 1 pretraining establishes a stable orbit prediction foundation before Phase 2 fine-tuning.

Gating prevents degradation. Removing the gate (setting $\mathbf{G} = \mathbf{1}$ in Equation 8) increased ISS MAE from 175 km to ∼400 km, as the perturbation head's corrections were applied unconditionally, including to samples where no solar wind perturbation exists.

Attention-weighted summary vs. mean-pooling. Replacing the learned attention weights (Equation 5) with simple mean-pooling of the cross-attended features increased ISS MAE from 175 km to ∼600 km, confirming that temporal structure in the attended features is important.

Expanded feature set does not improve performance. We tested an expanded 13-feature solar wind input adding AL and AU auroral indices, IMF clock angle (sin/cos decomposition of $\arctan(B_y, B_z)$), and solar wind dynamic pressure ($\frac{1}{2} n m_p v^2$) to the original 8 features. F10.7 solar flux was excluded due to daily-only cadence creating data leakage risk. The 13-feature model achieved 181 km MAE—worse than the 8-feature model (163 km). This indicates the original feature set (IMF components, flow speed, density, Kp, Dst, AE) already captures the relevant solar wind signal for LEO drag prediction. Additional geomagnetic indices and derived quantities add noise rather than information beyond what the model extracts from raw measurements.

7 Discussion

7.1 Why LSTM Outperforms Transformers

The LSTM's superiority over the Transformer (126 vs. 295 km on ISS) may reflect the inductive biases

of each architecture. Orbital dynamics exhibit strong temporal continuity—position at time t is primarily determined by position and velocity at $t-1$. LSTMs naturally capture this through their recurrent gating mechanism, while Transformers must learn temporal ordering from positional encodings alone. With 1,440-step input sequences, the quadratic attention complexity may also limit the Transformer's ability to attend to fine-grained temporal patterns.

7.2 Limitations

Scale discrepancy. Our MAE values for DSCOVR and MMS-1 (12,000–25,000 km) are large in absolute terms. These spacecraft operate at vastly different scales than LEO: DSCOVR is ∼1.5 million km from Earth, and MMS-1 reaches apogee at ∼150,000 km. Relative to orbital radius, the errors may be more comparable.

No physics integration. Our models are purely data-driven and do not incorporate known orbital mechanics. A hybrid approach combining neural network perturbation correction with physics-based propagation could yield substantially better results.

Limited storm events. While our dataset includes major storm events (Dst to −406 nT), extreme storms are rare. The multi-modal model's performance during the most intense events remains difficult to validate statistically.

Single prediction horizon. We evaluate only 6-hour predictions. Shorter horizons (1h) would better demonstrate the value of solar wind leading indicators, while longer horizons (24h) would test generalization.

7.3 Implications

The residual gated architecture demonstrates a principled approach to incorporating auxiliary environmental data into time-series prediction: the gate learns *when* the auxiliary modality is relevant, naturally suppressing it when it adds noise. This pattern generalizes beyond orbit prediction to any domain where environmental context intermittently affects the primary signal.

The strong LSTM performance suggests that for operational orbit prediction, simpler recurrent architectures may be preferable to attention-based models, at least for the data scales considered here (∼1M samples per spacecraft).

8 Conclusion

We have presented a multi-modal deep learning framework for spacecraft orbit prediction that incorporates solar wind data through cross-attention fusion with residual gating. Our key findings are:

1. Bidirectional LSTM achieves 125 km MAE for 6-hour ISS orbit prediction (54.5 km at 1 hour), outperforming the Keplerian baseline by 4.6× and establishing a strong data-driven baseline.

2. The residual gated multi-modal architecture achieves 163 km MAE overall, improving to 135 km during geomagnetic storms—a 17% improvement that validates the solar wind hypothesis.

3. An ensemble of LSTM and multi-modal predictions achieves the best storm-time performance at 111 km MAE.

4. Solar wind data provides the most benefit for LEO orbits where atmospheric drag dominates, but adds noise for deep-space (L1) and highly elliptical orbits.

5. Two-phase training and learned gating are both critical for multi-modal convergence.

Future work. We plan to investigate: (1) physics-informed architectures that embed Kepler's equations as an inductive bias, potentially combining the SGP4 Keplerian baseline with learned perturbation corrections; (2) TEME-to-GSE coordinate transforms for a full 3D SGP4 comparison; (3) extension to additional spacecraft; and (4) longer evaluation periods to increase storm sample sizes.

Data and Code Availability

All spacecraft position data is publicly available through the NASA Satellite Situation Center (SSC) REST API. Solar wind data is available through NASA's CDAWeb OMNI database. Our processed dataset (4.8M+ records) is hosted on Hugging Face.[2] Trained model checkpoints are available at `https://huggingface.co/datamatters24/orbital-chaos-predictor`. An interactive web dashboard with live ISS tracking and space weather is available at `https://orbitalchaos.online`.

[2] `https://huggingface.co/datasets/datamatters24/orbital-chaos-nasa-ssc`

References

SL Bruinsma, G Thuillier, and F Barlier. The dtm-2000 empirical thermosphere model with new data assimilation and constraints at lower boundary. *Journal of Atmospheric and Solar-Terrestrial Physics*, 65(9):1053–1070, 2003.

WD Gonzalez, JA Joselyn, Y Kamide, HW Kroehl, G Rostoker, BT Tsurutani, and VM Vasyliunas. What is a geomagnetic storm? *Journal of Geophysical Research: Space Physics*, 99(A4):5771–5792, 1994.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8): 1735–1780, 1997.

JH King and NE Papitashvili. Solar wind spatial scales in and comparisons of hourly wind and ace plasma and magnetic field data. *Journal of Geophysical Research: Space Physics*, 110, 2005.

Richard J Licata and Piyush M Mehta. Machine learning for satellite orbit prediction. *Advances in Space Research*, 2022.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

Hao Peng and Xiaoli Bai. Deep learning for orbit determination. *Journal of Guidance, Control, and Dynamics*, 43(8):1574–1580, 2020.

Eric K Sutton. A new method of physics-based data assimilation for the quiet and disturbed thermosphere. *Space Weather*, 16(7):736–753, 2018.

David A Vallado, Paul Crawford, Richard Hujsak, and TS Kelso. Revisiting spacetrack report #3: Rev 2. *AIAA/AAS Astrodynamics Specialist Conference*, 2006.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12): 11106–11115, 2021.