FORECASTING THE EVOLUTION OF SATELLITE CONJUNCTION EVENT PARAMETERS WITH PHYSICS-INFORMED TRANSFORMERS

Vincent Chirio[(1)], Fabio Chiappina[(2)], Jeffrey Small[(3)], Alinda Mashiku[(4)],

[(1)] University of Chicago, 5801 S. Ellis Ave, Chicago, IL 60637, USA;
Email: vincentchirio@gmail.com
[(2)] a.i. solutions, NASA Goddard Space Flight Center, 8800 Greenbelt Rd, Greenbelt, MD 20771, USA;
Email: fabio.chiappina@nasa.gov
[(3)] NASA CARA, NASA Goddard Space Flight Center, 8800 Greenbelt Rd, Greenbelt, MD 20771, USA;
Email: jeffsmall381@gmail.com
[(4)] NASA CARA, NASA Goddard Space Flight Center, 8800 Greenbelt Rd, Greenbelt, MD 20771, USA;
Email: alinda.k.mashiku@nasa.gov

ARTIFICIAL INTELLIGENCE (AI) USAGE DISCLOSURE

This document was created with assistance from ChatGSFC for code generation and text revision. The content has been reviewed and edited by Fabio Chiappina. More information on the extent and nature of AI usage is included in the Acknowledgements section of the document.

ABSTRACT

In modern conjunction assessment workflows, operators make decisions based on collision risk estimates that change as new space object measurements arrive. This paper introduces a physics-informed transformer architecture that learns from patterns in historical conjunction event data to forecast how key parameters – including covariance matrices around object states and the probability of collision of those objects – will likely evolve with future measurements, potentially enabling more proactive risk assessment decisions even before new observations become available. The model is trained on data from three NASA satellites and an assortment of objects that closely approach those satellites, and the processes of data selection and curation, feature engineering, and model training are outlined. Preliminary results indicate that the physics-informed conjunction transformer predicts conjunction parameters at future timesteps more accurately than a naive baseline method, serving as promising evidence that making effective predictions on conjunction event evolution for better-informed maneuver decisions may be possible.

1. INTRODUCTION

NASA's Conjunction Assessment Risk Analysis (CARA) program receives ephemeris data from all Earth-orbiting robotic NASA spacecraft. The predicted trajectories represented by those ephemerides are compared to those of a catalog of known space objects to identify predicted close approaches, known as conjunctions, between two space objects. The properties and predicted trajectories of the primary object (i.e., the NASA asset being screened against the rest of the space object catalog) and the secondary object (i.e., any other cataloged object, including other NASA assets, commercial and international spacecrafts, rocket bodies, and debris fragments) are then used to compute probability of collision, which quantifies the risk of the conjunction event. When probability of collision of a conjunction event exceeds a threshold of 1e-4, NASA's CARA recommends mitigation (e.g., maneuvering to change trajectories and reduce collision risk) [1].

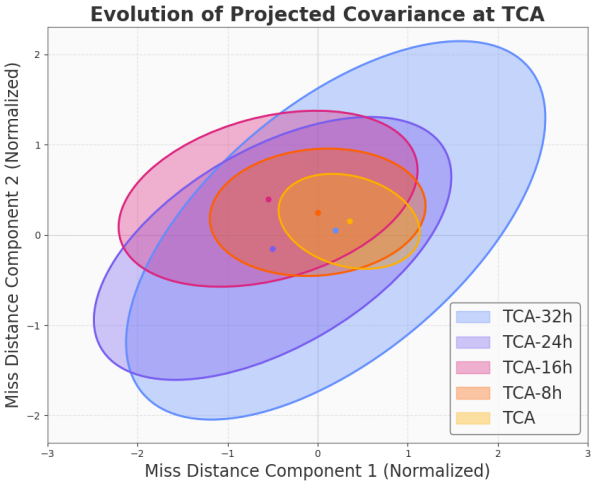


Figure 1. Illustration of the evolution of projected covariance at TCA across measurement updates. Each ellipse represents relative state uncertainty (e.g., 1-σ) in the conjunction plane at TCA as estimated from the OD solution available at the labeled screening time. Ellipse centers indicate the predicted miss distance.

An increasingly congested orbital environment demands effective frameworks for making timely risk mitigation decisions to protect operational spacecraft. While existing conjunction assessment (CA) workflows that rely on physics-based state propagation are well-established, they do not fully exploit the rich temporal structure present in historical conjunction event parameter sequences. Artificial intelligence and/or machine learning (AI/ML) techniques excel at

extracting and leveraging such temporal structure from vast datasets. This paper explores the potential to learn predictable temporal trends in conjunction event parameters from historical data via AI/ML. Physics-informed transformers trained on historical conjunction data display predictive improvement over baseline naive persistence models and provide analytical insights valuable to CA research and operations.

2. LITERATURE REVIEW

2.1. AI/ML for Orbit Determination

Recent work shows AI/ML adding value to orbit determination (OD) workflows as an augmentation or improvement upon certain components of classical methods rather than a wholesale replacement of the end-to-end process. Promising examples include the use of neural networks and differentiable programming to increase the precision of the widely used SGP4 propagator without lengthening computation time [2] and the use of physics-informed and/or traditional neural networks to perform initial OD using angles-only data in the cislunar regime [3, 4]. An adjacent domain in which AI/ML methods have recently demonstrated significant value is solar and/or atmospheric modeling, where ML techniques have enabled the construction of data-driven thermospheric density models that can reduce orbit propagation errors and/or uncertainties [5]. Overall, AI/ML appears to be a promising avenue for advancement of OD techniques, whether by introducing better coupling to physical constraints or by leveraging empirical data to improve analytical models [6]; in one study, for example, neural networks trained on historical OD data learned and corrected orbit prediction errors, suggesting that data-driven ML models can effectively augment processes governed by dynamical systems without requiring increased fidelity in the underlying physical models [7]. The use of AI/ML is thus a promising avenue for improving the quality of the OD data products in the CA pipeline upstream of the computation and modeling of the temporal evolution of conjunction event parameters.

2.2. AI/ML for Conjunction Assessment

Downstream of OD, AI/ML methods have also been investigated for a range of tasks in the conjunction assessment pipeline. NASA CARA's recent AI/ML compendium concludes that, given the scarcity of "ground-truth" collision data and the need for explainability of model results, the AI/ML approaches explored so far (fuzzy inference, clustering, LSTMs on evolving CDMs) have not yet demonstrated consistent operational advantages over established probability of collision (PC)/Monte-Carlo methods [8].

However, recent academic studies have begun to explore, with some success, the application of ML techniques to various aspects of CA. Deep-learned classifiers have been trained to detect close conjunctions with improved computational efficiency over classical filters [9]; graph neural networks have shown promise as a conjunction detection mechanism [10]; models have been trained on decisions made by human analysts to make go/no-go collision avoidance maneuver decisions [11]. Other efforts have used AI/ML to model the temporal evolution of conjunction events and their associated parameters: a competition surrounding an ESA-provided dataset of conjunction events and the creation of models predicting their collision risk showed that ML approaches are a promising avenue for improving upon the naive forecasting baseline [12]; a Bayesian approach has been applied to effectively and directly predict probability of collision using historical CDM data [13]; Bayesian deep learning methods have also been proposed to model CDM time-series with richer feature representations [14].

While these studies demonstrate that AI/ML techniques can add value to the conjunction assessment pipeline, no such technique has yet been widely adopted into operational workflows, and further exploration remains necessary. Many existing approaches rely on relatively simple models and/or narrow feature spaces and thus do not fully exploit the rich temporal structure present across successive measurement updates. An abundance of data remains unutilized, and state-of-the-art deep learning approaches for integrating this data in a way that incorporates physical constraints – on which this analysis focuses – have yet to be fully explored.

2.3. Transformer Architecture

The transformer architecture, a type of deep neural network first introduced in its modern formulation by a foundational technical paper in 2017, has recently been the subject of much excitement and research in the domain of AI and machine learning [15]. Though best known for enabling large language models (LLMs), the transformer architecture is now considered a state-of-the-art tool for sequential data modeling and time-series forecasting in general due to its ability to extract semantic content from complex sequential data, leading to its adoption by many disciplines. In spacecraft guidance, navigation, and control, transformers have been used to learn control policies and accurately predict action sequences [16]; in atmospheric modeling, transformers have demonstrated their superior performance over traditional linear propagation methods for atmospheric density forecasting [17]; and in orbit prediction, transformers have achieved low positioning error while maintaining a much lower computational cost than high-fidelity propagators [18]. The widespread success of transformers in modeling complex temporal patterns in physical systems makes compelling their use in CA, where conjunction parameters and uncertainties form structured temporal sequences across successive measurement updates.

2.4. Physics-Informed Neural Networks

Despite the widespread success of neural networks and their variants in scientific and engineering domains, challenges still emerge when these networks attempt to model physical systems. One such challenge is the failure of trained models to satisfy known properties, constraints, or dynamical equations governing physical systems, resulting in poor extrapolation despite adequate performance with inputs within the provided training data. Physics-informed neural networks (PINNs) can mitigate these issues for neural networks applied to tasks involving known physical properties [19]. PINNs incorporate known partial differential equations (PDEs) representing dynamical constraints of the modeled system into the loss function, thus penalizing the model during training for failing to uphold these constraints and improving physical understanding in the learning process. For instance, one study applied PINNs to gravitational potential modeling, improving upon the robustness of traditional methods [20].

Taking inspiration from PINNs, this paper introduces an approach to incorporating physics knowledge into this training process in which a physics-informed loss penalizes the model for producing position covariance matrices that are not positive-definite, influencing the model to produce output that is physically coherent. This technique differs slightly from the typical approach of leveraging PDEs to incorporate known relationships between inputs and outputs, but it is still derived from the same principle of augmenting the training loss function with physics knowledge to improve generalization and physical realism.

2.5. Summary and Motivation

Prior work demonstrates that AI/ML techniques have the potential to enhance certain aspects of the CA pipeline. Though existing approaches have found some success, they have yet to fully exploit the temporal relationships embedded within historical conjunction data measurement update sequences while also incorporating physical constraints into the learning process. This analysis builds on these foundations by combining transformer-based sequence modeling with physics-informed loss to forecast the evolution of conjunction event parameters. The eventual integration of such a transformer model into operational risk assessment workflows would allow decisions to be informed by the conjunction event evolution patterns extracted from a vast repository of currently underutilized historical data.

3. DATA COLLECTION AND PROCESSING

The selected dataset is limited to a subset of conjunction events from between January 2020 and June 2025 and involving one of three well-tracked and rarely maneuvered NASA satellites, henceforth referred to as NS1, NS2, and NS3. This dataset comprises 18,001 conjunction events spanning over 6,000 unique secondary objects, where each event contains at least two sequential measurement updates. Conjunction events are randomly split into training (70%), validation (15%), and test (15%) sets. Each sample represents a complete sequence of features associated with a unique conjunction event, independent of all others.

3.1. Feature Selection and Engineering

Each update in a sequence constituting a conjunction event includes the twenty-one elements that fully define the combined object position and velocity covariance matrix at time of closest approach (TCA), propagated from the OD solution available at that sample time, as well as four elements that define the combined covariance ellipse in the conjunction plane: the semi-major and semi-minor axes and the cosine and sine of the conjunction clock angle, defined as the angle between the major axis of the combined covariance ellipse and the relative position vector between the two objects [21]. For probability of collision forecasting (see section 5.4), two additional features – the computed probability of collision and the miss distance from the latest available OD solution – are included.

The transformation of the clock angle into two derived features through doubling and subsequent cosine and sine operations, rather than using the angle itself as a single feature, is useful in that the resulting representations are smooth and continuous. In a feature space considering only raw clock angle, two clock angles of 0 and 180 degrees would be far apart; in the transformed, two-feature space, these angles map to nearby points, so consecutive event updates with clock angles near 0 or 180 degrees do not yield large changes in the feature values. The latter feature space is thus a more meaningful representation of the conjunction geometry and promotes more effective learning.

3.2. Ellipsoid Orientation Clustering

To motivate the use of AI/ML techniques to make predictions on conjunction event parameters, an agglomerative hierarchical clustering procedure with a Dynamic Time Warping distance metric is applied on event sequences in search of coherent groupings revealing useful latent structure [22, 23]. Clustered feature vectors contain only two features: clock angle sine and cosine. A silhouette score is computed for each clustering result to measure how effectively the number of produced clusters captures the natural structure of the dataset [24]. By performing multiple clustering trials and selecting the result with the highest average silhouette score across all samples, the most effective number of clusters is determined.

The clustering procedure is applied to events from each of the three selected NASA missions separately; interestingly, for each mission, four clusters are discovered, each with distinct clock angle evolution patterns (see Figure 2). Since the evolution of conjunction event clock angles does appear to be organizable into distinct clusters, the clustering results serve as compelling evidence that conjunction event data does contain latent structure that can be exploited by AI/ML methods to make effective predictions – and that conjunction clock angle cosines and sines are useful features for representing that structure.

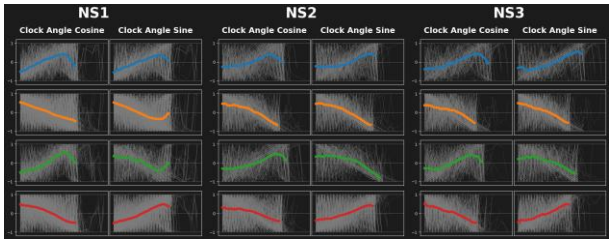


Figure 2. Conjunction clustering results for all three NASA satellites in this investigation using two features: the cosine and sine of conjunction clock angles.

4. PHYSICS-INFORMED TRANSFORMER DESIGN

The CA domain necessitates extracting meaningful information from complex data while maintaining interpretability to understand operational implications. Transformer architectures are thus selected, and architectural and training decisions aim to preserve an intuitive understanding of model predictions.

4.1. Model Architecture

The transformer model in this analysis is designed to predict feature values in future measurement updates given the values of those same features from previous measurement updates. Since sequences of conjunction event parameters encode complex temporal patterns, the transformer architecture is an optimal choice for modeling conjunction evolution. The underlying attention mechanisms effectively enable the model to "look back" at every preceding measurement before predicting the next, learning deep relationships defining how sequences of updates interact with each other.

The chosen architecture, illustrated in Figure 3, is segmented into three separate components, each associated with a distinct aspect of conjunction assessment workflows: orbit determination, projection onto the conjunction plane, and collision risk parameter estimation. Segmenting the architecture in this way preserves the traceability of model predictions through the conventional conjunction assessment workflow. The total parameter count of each component is presented in Table I below.

Table I. Number of parameters in each transformer component used throughout this analysis.

Component	Parameter Count
OD Transformer	1.1M
Conjunction Transformer	680K
Collision Transformer	630K
Total	2.41M

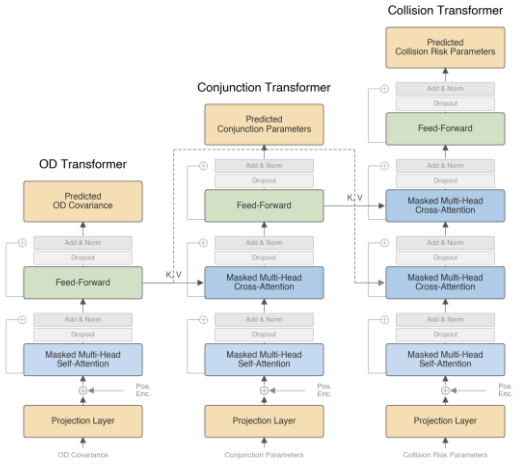


Figure 3. A high-level diagram illustrating the full model architecture, segmented into orbit determination, projection onto the conjunction plane, and collision risk parameter estimation.

4.2. Layers

Various layer types are repeated across the three transformer components. Each component begins with an input projection layer that maps raw features into a higher-dimensional latent space, followed by multi-headed attention blocks that enable the model to learn dependencies across sequential updates. Self-attention allows the data from each timestep in a sequence to attend to each earlier measurement, thus enabling the transformer to learn how conjunction event parameters evolve throughout an event's lifecycle. Layer normalization, residual connections, and dropout layers stabilize training and minimize overfitting to the training data, while feed-forward layers transform the learned representations back into predictions of conjunction event parameters.

Importantly, the OD transformer implements two stages of self-attention: first, only the projected OD features are considered, and then additional features are incorporated to encode the identity and properties of the space object under analysis; thus, any patterns in conjunction event parameter evolution that are unique to certain objects can be captured. In the conjunction and collision transformers, self-attention similarly models temporal dependencies within conjunction plane and collision risk parameters, respectively, and cross-attention layers integrate covariance predictions and conjunction information to capture how all such factors,

from object-specific qualities to patterns in covariance sequences across space objects as a whole, drive the evolution of collision risk.

4.3. Training

Each model is trained for 25 epochs using the AdamW optimizer with cosine annealing to gradually reduce the learning rate [25]. Validation dataset loss is monitored at every epoch, and early stopping is utilized such that training halts if seven consecutive epochs fail to produce an improvement in the validation loss.

4.4. OD Training

In initial experiments, trained OD models predicted high rates of position covariance matrices that were not positive-definite, which is physically impossible and represents an invalid covariance ellipsoid. To address this in the OD portion of the model, a physics-informed penalty term is incorporated into the loss function in addition to the standard regression loss to ensure physically meaningful predictions. The baseline loss is given by the Huber loss (L_{Huber}), which measures the discrepancy between the predicted covariance elements y_p and the corresponding ground-truth values y while remaining robust to outliers [26]. To penalize physically invalid predictions, for each event in a single training step, the eigenvalues of the predicted covariance matrix are computed. Since a physically valid covariance must be positive semi-definite, negative eigenvalues indicate an unphysical solution, and thus, an additional loss penalty is added proportional to the magnitude of those eigenvalue(s), discouraging (though not entirely preventing) the model from learning such solutions. The weighting of this penalty relative to the Huber loss becomes an additional hyperparameter requiring tuning. Let λ_i denote the i-th eigenvalue of a predicted covariance matrix. The parameter that governs the weighting of the physics-informed penalty with respect to the Huber loss term is α. Thus, the combined Huber and physics-informed losses are defined as follows:

$$L = L_{Huber}(y, y_p) + \alpha \sum_i \min(0, -\lambda_i)^2$$

The training pipeline for the OD model is as follows:

1. The training data is split into batches (n=32) of unique events to be processed in parallel at each step.
2. For each event, a forward pass through the model is performed, producing a vector of 21 predicted covariance elements.
3. For each timestep within the batch that has a valid label (observed next step), the absolute difference between the actual and predicted covariance values are recorded and averaged, the physics-informed penalty is applied, and Huber Loss is calculated across the batch.
4. A backward pass calculates the partial derivatives of loss with respect to model parameters.
5. All model parameters are greedily updated in the direction of decreasing batch loss according to the partial derivatives.

4.5. Conjunction and Collision Training

For the conjunction and collision transformers, the training pipeline remains the same aside from a few nuances. First, when performing a forward pass, features are extracted by passing the necessary data through preceding components. During each training step for the conjunction model, object covariances are first passed through the OD model to produce predicted orbital covariance states incorporated as additional features. Similarly, additional features are extracted from both the OD and conjunction models to be incorporated into the training of the collision model. Second, because the conjunction and collision components output fundamentally different parameters, there is no physics-informed penalty implemented – only basic Huber loss.

5. RESULTS

5.1. Naïve Persistence Baseline

An important step in evaluating the effectiveness of any model is a comparison against one or more elementary models to ensure that the complexity added by the model, which may confound interpretability and cost resources to train and/or deploy, is justified by significant performance improvements over a minimum benchmark. For time-series forecasting tasks, a commonly used baseline is often called a naive or persistence model [27, 28]. This model always predicts that the observation at the next timestep (and any subsequent timesteps) will be identical to the previous observation. For short forecast horizons, naive persistence models can perform reasonably well, but their performance diminishes as they predict further into the future. Outperforming the persistence model is a useful indicator that a model has accomplished meaningful learning. All other results in this section are discussed in comparison to the results of this model.

5.2. OD Transformer Forecasting Results

The resulting trained OD Transformer demonstrates a 45.17% decrease in Huber loss over the naive persistence baseline on test data, indicating a noteworthy degree of predictive power (see Table II). These results may be observed qualitatively in Figure 4, which displays a graphical representation of the progression of uncertainty for the primary object during a conjunction event. In this figure, the covariance values at each time step are predicted based on all preceding observations; the naive baseline propagates the previous

state forward, and the model performs a forward pass with all preceding observations to obtain the current state prediction. Both the naive and model next-measurement-update predictions track actual covariance trends closely, suggesting that model's improvement stems from detecting subtle trends and learning to anticipate and/or ignore sudden changes.

The performance gains are significantly more apparent when the capability of the model design to chain together predictions to forecast values multiple measurement updates into the future is leveraged. Figure 5 displays the chained primary covariance predictions propagated only from the initial observation. Here, the model predictions generally follow the true parameter evolutions while diverging significantly from the naive predictions, which remain stagnant at the initial values. This demonstrates the promise of the OD transformer to reliably estimate object uncertainty at TCA far in advance.

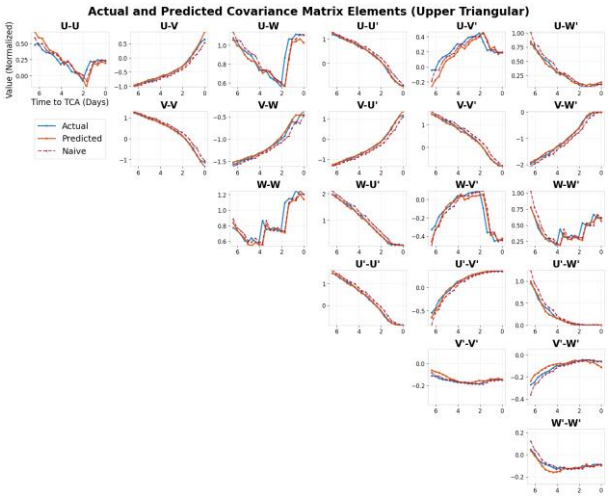


Figure 4. Predicted versus actual covariance values for the primary object at each timestep during a conjunction event. Each prediction is re-generated at each timestep using all data from before that timestep.

5.3. Conjunction Transformer Forecasting Results

Similarly, the trained conjunction transformer demonstrates a 31.91% decrease in Huber loss compared to the naive persistence baseline (see Table III). In Figure 6, this improvement comes through trending the gradual decrease of the conjunction ellipsoid semi-major axis as TCA approaches and the anticipation of shifts in ellipsoid orientation.

Figure 7 illustrates the conjunction transformer's capability to forecast conjunction ellipsoid parameters many measurement updates into the future. Not only does it roughly trend the decrease in semi-major and semi-minor axes over time, but it also adequately approximates the final ellipsoid orientation based only on the very initial state. The orientation of the conjunction ellipsoid close to TCA has significant

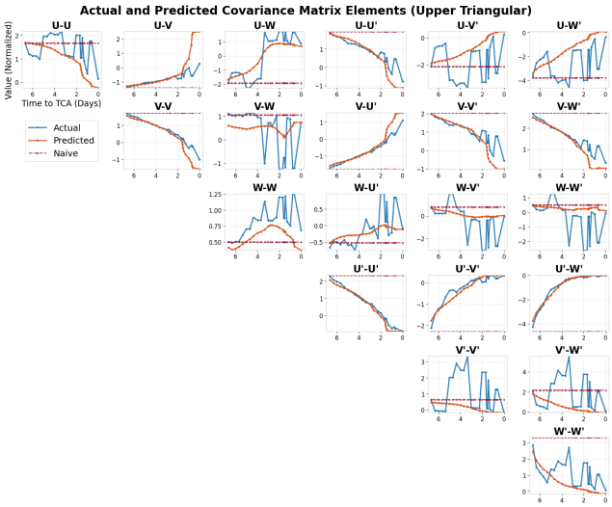


Figure 5. Propagated covariance predictions from initial state only versus actual covariance for a primary object during a conjunction event.

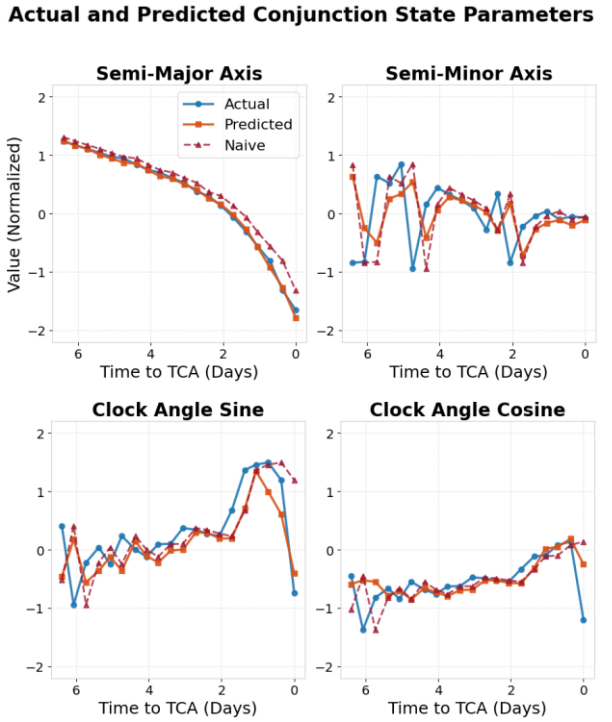


Figure 6. Predicted versus actual conjunction parameter values at each timestep during a conjunction event. Each prediction is of only the single next timestep, using observed values from all available previous timesteps.

implications for the collision risk associated with an event. For the evaluation of collision risk, both the magnitude and direction of uncertainty are critical, represented by the axes of the ellipsoid and the clock angle of the ellipsoid, respectively. Uncertainty in a direction can simultaneously be large but uncorrelated

with miss distance or small but highly correlated with miss distance, meaning there may be significant operational value in forecasting how future measurements will affect these parameters, enabling informed decision-making even before new data arrives.

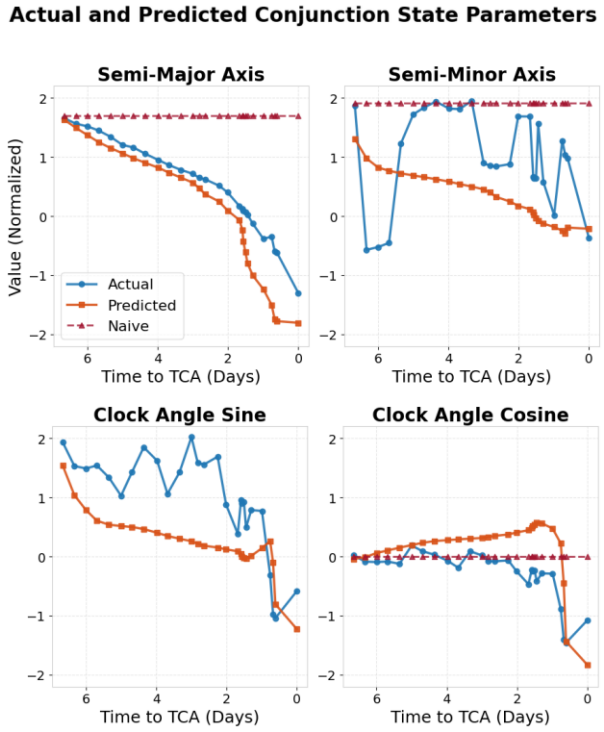


Figure 7. Predicted versus actual conjunction parameter values throughout a conjunction event, where predictions are propagated from the initial observed state only; subsequent observations are not incorporated into the predictions.

5.4. Probability of Collision Transformer Forecasting Results

Finally, the probability of collision transformer exhibits an 18.28% decrease in Huber loss (averaged across both features – probability of collision and miss distance) compared to the naive persistence baseline (see Table IV). Among the three transformers, this transformer is the worst performing, with the lowest improvement in Huber loss over the baseline. Though the transformer does significantly improve upon the baseline for certain events, it performs considerably worse than the baseline for many others, especially for longer forecast horizons. Further investigation is necessary to improve the performance of this transformer, to identify shared characteristics of the events on which it performs best, and to determine why it performs comparatively worse than the other two transformers when the underlying parameters upon which probability of collision depends (e.g., covariances) are forecasted more accurately.

Figure 8 demonstrates an example of model predictions near TCA at varying forecast windows for a particular conjunction event. When the transformer only predicts the next observation in the sequence, its performance is quite similar to that of the baseline. When the last 10 observations are cut off, requiring the model to propagate its forecasts 10 observations into the future, the model projects a gradual drop-off in both miss distance and probability of collision; while the miss distance forecast does not decrease quickly enough to keep pace with the truth data, the model is correct to predict decreases in both features. When the last 20 observations are cut off, meaning for this conjunction event that the model is given only the initial state and one subsequent observation, the model still predicts that both features will gradually decline, and does so quite accurately for probability of collision. The relative accuracy with which this decline is predicted from just the initial conditions of the conjunction event suggest that the collision transformer model may be capable of anticipating important changes in collision risk far in advance of TCA. Thus, these modeling techniques show promise for operational utility in collision risk forecasting.

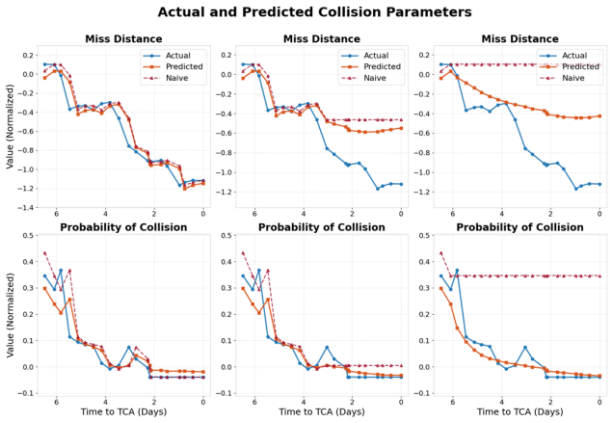


Figure 8. Propagated probability of collision predictions from varying forecast windows for a particular conjunction event. From left to right: always predicting the next state given all previous observations; cut off 10 observations before TCA; cut off 20 observations before TCA.

5.5. Quantitative Performance Comparison

Tables II-IV quantify the errors made by the naive persistence model versus the transformer model for the tasks discussed above, highlighting the improvements made by the transformer over the naive model.

Table II. Percentage decrease of the Huber loss of the OD transformer compared to the persistence model averaged across various forecast horizons.

Forecast Horizon	Avg. Loss, Persistence	Avg. Loss, OD Transformer	Loss Percent Decrease
1-29	0.2078	0.1140	45.17
1-5	0.2725	0.1804	33.81
6-10	0.1358	0.0981	27.77
11-15	0.1549	0.0984	36.49
15-20	0.0940	0.0566	39.82
21-25	0.0899	0.0434	51.79
26-29	0.5728	0.2303	59.80

Table III. Percentage decrease of the Huber loss of the conjunction transformer compared to the persistence model averaged across various forecast horizons.

Forecast Horizon	Avg. Loss, Persistence	Avg. Loss, OD Transformer	Loss Percent Decrease
1-29	0.5072	0.3453	31.91
1-5	0.5684	0.3912	31.18
6-10	0.4707	0.3378	28.23
11-15	0.5002	0.3495	30.13
15-20	0.4124	0.3416	17.17
21-25	0.4147	0.2713	34.57
26-29	0.7191	0.3893	45.86

Table IV. Percentage decrease of the Huber loss of the collision transformer compared to the persistence model averaged across various forecast horizons.

Forecast Horizon	Avg. Loss, Persistence	Avg. Loss, OD Transformer	Loss Percent Decrease
1-29	0.5072	0.3453	31.91
1-5	0.5684	0.3912	31.18
6-10	0.4707	0.3378	28.23
11-15	0.5002	0.3495	30.13
15-20	0.4124	0.3416	17.17
21-25	0.4147	0.2713	34.57
26-29	0.7191	0.3893	45.86

5.6. Limitations

A common phenomenon observed in the evolution of conjunction events is that the uncertainty and/or probability of collision of the event is high and relatively constant for the majority of the event duration but then drops off substantially near TCA as new measurements become available. The results of this investigation suggest that the transformer model captures this common behavior better than the persistence model. However, for any given conjunction where the transformer model correctly predicts this behavior, whether the model is using information specific to that conjunction to predict the decline of covariance and/or probability of collision near the time of closest approach remains unclear; the model may have simply learned that this is a commonly observed phenomenon and that consistently predicting this behavior is an effective enough heuristic to result in relatively low overall loss. Not all conjunction events evolve in this way; in some cases, the probability of collision may significantly increase near the time of closest approach. Given the characteristics of a particular conjunction, a truly effective model would be able to discern how those characteristics will evolve over time, rather than predicting a single common evolution pattern across all conjunctions. Further analysis with larger datasets including sample conjunctions the evolution of which does not adhere to this commonly observed pattern is necessary to ensure that the model predicts those evolutions equally well.

Additionally, the effectiveness of the positive definite penalty imposed upon covariance matrices is limited for velocity covariance matrices due to the common occurrence of extremely small values that are small enough to exceed the precision of standard Python and PyTorch. The penalty also does not strictly enforce positive definiteness in predicted position covariances; it merely penalizes the model for learning to predict non-positive definite covariances, reducing their likelihood but not eliminating them. More sophisticated techniques to circumvent these issues are recommended.

5.7. Interpretability

An important facet of the evaluation of a machine learning model is an analysis on the interpretability of its predictions. Understanding why a model makes the predictions that it does helps engineers build trust in those predictions and diagnose whether the model is learning meaningful relationships rather than spurious correlations. Segmenting the overall architecture into three constituent transformers assists in this interpretability effort; loss is computed separately for each transformer, helping to narrow down where in the CA process contributes most strongly to overall error. Inspecting attention layer weights for particular conjunctions helps indicate which timesteps are contributing most heavily to model predictions. An object embedding layer in the OD transformer, which learns a vector representation of each object in the dataset encoding its unique conjunction evolution patterns; dot product similarity between object embedding vectors reveals that the two objects with greatest embedding similarity are NS1 and NS2, which may be explained by their similar sensor coverage – NS2 receives the overwhelming majority of its tracks from sensors that also frequently track NS1. Further inspection of hidden layers, embedding vectors, and their correlations with object orbital geometries, sensor coverage, and other object- or conjunction-level properties is recommended to understand the factors that most contribute to successful conjunction event parameter forecasting.

6. FUTURE WORK

This analysis focused on a relatively small dataset from only three hand-selected missions. A substantial broadening of the dataset under consideration is necessary to ensure that the resulting model is

generalizable across missions and that the learned patterns are not easily reducible to simple heuristics (see section 5.6). Detailed characterization (e.g., error mean and variance) of the model's performance not only forecasting one timestep into the future, but also all subsequent timesteps until the time of closest approach, is also recommended and would allow for the construction of confidence intervals around each forecasted value.

Given the potential for effective feature selection to significantly improve model performance, a natural extension of this investigation is the expansion of the feature space to incorporate additional phenomena that may inform the evolution of conjunction event characteristics. One such additional phenomenon to consider is space weather. Stronger solar activity increases atmospheric heating, and the resulting atmospheric expansion results in higher atmospheric density and a stronger drag force exerted upon spacecraft, particularly those in lower orbits. Incorporating current or forecasted values of one or multiple solar proxies may help the model attend to how changes in solar and atmospheric activity informs the evolution of conjunction events. Orbital elements, object type and characteristics, and the locations and characteristics of sensors and their probabilities of detecting the objects may also provide valuable context that may inform and improve conjunction event parameter forecasting accuracy.

Finally, while the probability of collision transformer discussed in section 5.4 directly computes estimates of probability of collision as outputs of its transformer network, several analytical methods already exist for computing this parameter without relying on machine learning. Thus, an alternative approach would be to forecast each time-varying parameter used in its computation (e.g., covariance, etc.) and then use an established method to compute the probability of collision from those forecasted parameters. Further analysis is necessary to determine whether direct forecasting versus forecasting of time-varying parameters and subsequent analytical computations yields closer approximations to the true probability of collision at future timesteps.

7. ACKNOWLEDGEMENTS

Generative AI (i.e., a large language model) was used throughout this investigation for code generation and technical writing assistance (e.g., grammar, narrative flow). All generated code was reviewed, edited, and tested by the authors of this document, and all generated text was manually revised.

8. CONCLUSION

This investigation centers on the application of machine learning techniques to make predictions on the evolution of parameters describing conjunction events, including covariance and probability of collision. Transformer architectures are used to forecast conjunction event parameters and demonstrate promising results, capturing complex patterns and outperforming the naive persistence forecaster baseline, while discouraging the learning of non-positive-definite position covariance matrices via a physics-informed loss penalty. Forecasting performance is evaluated on a small dataset containing conjunction events from three NASA satellites. To build a more accurate forecaster with greater operational utility, a significant expansion of the dataset under consideration, both in number of conjunction events and in external features that may serve as useful context, is recommended, as is further interpretability analysis via examination of hidden layers and embeddings.

REFERENCES

1. NASA. (2023). Spacecraft Conjunction Assessment and Collision Avoidance Best Practices Handbook. Office of the Chief Engineer, NASA/SP-20230002470 (Rev 1). https://ntrs.nasa.gov/citations/20230002470
2. Acciarini, G., Baydin, A.G. & Izzo, D. (2025). Closing the gap between SGP4 and high-precision propagation via differentiable programming. Acta Astronautica 226, 694–701.
3. Scorsoglio, A., D'Ambrosio, A., Ghilardi, L., Furfaro, R. & Reddy, V. (2023). Physics-Informed Orbit Determination for Cislunar Space Applications. Advanced Maui Optical and Space Surveillance Technologies (AMOS) Conference, Maui, Hawaii, USA.
4. Ojeda Romero, J., Schlei, W., Whipps, G., LaFarge, N., Fritsch, G. & Phillips, S. (2025). An Adaptive Approach to Cislunar Initial Orbit Determination Using Machine Learning. The Journal of the Astronautical Sciences 72, 45.
5. Acciarini, G., Brown, E., Berger, T., Guhathakurta, M., Parr, J., Bridges, C. & Baydin, A.G. (2024). Improving Thermospheric Density Predictions in Low-Earth Orbit With Machine Learning. Space Weather 22(2), e2023SW003652.
6. Caldas, F. & Soares, C. (2024). Machine Learning in Orbit Estimation: a Survey. Acta Astronautica 220, 97–107.
7. Peng, H. & Bai, X. (2018). Artificial Neural Network–Based Machine Learning Approach to Improve Orbit Prediction Accuracy. Journal of Spacecraft and Rockets 55(5), 1248–1260.
8. Mashiku, A.K., Newman, L.K. & Highsmith, D.E. (2025). NASA Conjunction Assessment Risk Analysis (CARA) Compendium for

Artificial Intelligence and Machine Learning for Satellite Collision Avoidance. Advanced Maui Optical and Space Surveillance Technologies (AMOS) Conference, Maui, Hawaii, USA.

9. Stevenson, E., Rodriguez-Fernandez, V., Urrutxua, H. & Camacho, D. (2023). Benchmarking deep learning approaches for all-vs-all conjunction screening. Advances in Space Research 72(7), 2660–2675
10. Stevenson, E., Rodriguez-Fernandez, V. & Urrutxua, H. (2022). Towards graph-based machine learning for conjunction assessment. Advanced Maui Optical and Space Surveillance Technologies (AMOS) Conference, Maui, Hawaii, USA.
11. Ravi, P., Zollo, A. & Fiedler, H. (2023). AI for Satellite Collision Avoidance – Go/No Go Decision-Making. Orbital Debris Conference, Houston, Texas, USA.
12. Uriot, T., Izzo, D., Simões, L.F., Abay, R., Einecke, N., Rebhan, S., Martinez-Heras, J., Letizia, F., Siminski, J. & Merz, K. (2022). Spacecraft collision avoidance challenge: Design and results of a machine learning competition. Astrodynamics 6(2), 121–140.
13. Simões Catulo, J., Soares, C. & Guimarães, M. (2023). Predicting the Probability of Collision of a Satellite with Space Debris: A Bayesian Machine Learning Approach. arXiv preprint arXiv:2311.10633.
14. Pinto, F., Acciarini, G., Metz, S., Boufelja, S., Kaczmarek, S., Merz, K., Martinez-Heras, J.A., Letizia, F., Bridges, C. & Baydin, A.G. (2020). Towards Automated Satellite Conjunction Management with Bayesian Deep Learning. arXiv preprint arXiv:2012.12450.
15. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. & Polosukhin, I. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems 30, 5998–6009.
16. Posadas-Nava, A., Scorsoglio, A., Ghilardi, L., Furfaro, R. & Linares, R. (2025). Action Chunking with Transformers for Image-Based Spacecraft Guidance and Control. AAS/AIAA Astrodynamics Specialist Conference (ASC), Paper AAS 25-897.
17. Briden, J., Siew, P.M., Rodriguez-Fernandez, V. & Linares, R. (2023). Transformer-based Atmospheric Density Forecasting. Advanced Maui Optical and Space Surveillance Technologies (AMOS) Conference, Maui, Hawaii, USA.
18. Caldas, F. & Soares, C. (2024). Precise and Efficient Orbit Prediction in LEO with Machine Learning using Exogenous Variables. IEEE Congress on Evolutionary Computation (CEC).
19. Raissi, M., Perdikaris, P. & Karniadakis, G.E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics 378, 686–707.
20. Martin, J.R. & Schaub, H. (2023). The Physics-Informed Neural Network Gravity Model Revisited: Model Generation III. AAS/AIAA Astrodynamics Specialist Conference (ASC), Paper AAS 23-110.
21. Newman, L.K. (2008). The NASA Robotic Conjunction Assessment Process: Overview and Operational Experiences. International Astronautical Congress (IAC), Paper IAC-08-A.6.2.6.
22. Ward, J.H., Jr. (1963). Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association 58(301), 236–244.
23. Berndt, D.J. & Clifford, J. (1994). Using Dynamic Time Warping to Find Patterns in Time Series. AAAI-94 Workshop on Knowledge Discovery in Databases (KDD-94), 359–370.
24. Rousseeuw, P.J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20, 53–65.
25. Loshchilov, I. & Hutter, F. (2019). Decoupled Weight Decay Regularization. International Conference on Learning Representations (ICLR).
26. Huber, P.J. (1964). Robust Estimation of a Location Parameter. The Annals of Mathematical Statistics 35(1), 73–101.
27. Hewamalage, H., Ackermann, K. & Bergmeir, C. (2022). Forecast evaluation for data scientists: common pitfalls and best practices. Data Mining and Knowledge Discovery 37, 788–832.
28. Beck, N., Dovern, J. & Vogl, S. (2025). Mind the naive forecast! A rigorous evaluation of forecasting models for time series with low predictability. Applied Intelligence 55, 395.