Quantifying Accuracy and Uncertainty in Data-Driven Flight Trajectory Predictions with Gaussian Process Regression

Graas, R.; Sun, Junzi; Hoekstra, J.M.

Publication date
2021

Document Version
Final published version

Published in
11th SESAR Innovation Days

Quantifying Accuracy and Uncertainty in Data-Driven Flight Trajectory Predictions with Gaussian Process Regression

Rik Graas, Junzi Sun*, Jacco Hoekstra
Faculty of Aerospace Engineering, Delft University of Technology, the Netherlands
*Corresponding author: j.sun-1@tudelft.nl

***Abstract*—Several initiatives are developed to shift the current paradigm in Air Traffic Management from the tactical-based approach to more strategic-based coordination of flights. This transformation of the ATM system relies on the improvement of predictive models for the 4D flight trajectories. A variety of performance-based and data-driven approaches are developed for trajectory predictions. The accuracy of the predictions is often deterministic and can be highly impacted by uncertainties that occur in each flight. These uncertainties are commonly related to the lack of detailed information concerning the flight intent, or the inaccuracy of positional and weather-related data. To better understand prediction errors and uncertainties in data-driven predictions, this study proposes a novel two-stage Gaussian Process Regression (GPR) approach. By combining historical flight data and flown trajectory of a given flight, the predictive distributions from the GPR allow us to study both prediction errors and uncertainties. To evaluate the model, we applied the method for flights arriving at the Amsterdam Airport Schiphol. We also evaluate and quantify how flight-plan and meteorological information help to reduce prediction error and uncertainty.**

***Index Terms*—Trajectory prediction, Gaussian Process Regression, Prediction accuracy, Uncertainty quantification**

I. Introduction

Trajectory Based Operations (TBO) have been identified as a key enabler for future ATM [1]. Effective implementation of such an approach relies on the accurate prediction of the flight path. The quality of these predictions is often impacted by uncertainties, which may vary during different phases of the flight. Traditionally, TPs are modeled using deterministic techniques and models which do not explicitly capture the sources of uncertainty that affect the prediction accuracy [2]. Many of these models provide a single trajectory forecast without the ability to express prediction uncertainties.

Modeling uncertainty in predictive analytics involves the process of uncertainty quantification, which aims to describe how the uncertainty in input parameters of a predictive model affects the uncertainty of the predictions of the target variable. This is commonly addressed by Monte Carlo simulations, where performance parameters are expressed by Probability Density Functions (PDF). And subsequently, the input uncertainties are propagated through deterministic models to identify the joint effect of the stochastic factors on the predictions of the target variable [3], [4].

The increasing availability of trajectory data has given rise to the popularity of data-driven techniques that apply machine learning models to predict aircraft trajectories. Radar data and Automatic Dependent Surveillance-Broadcast (ADS-B) data are widely used sources [5]–[7]. Aircraft intent data is commonly derived from flight plans, which typically contain the type of aircraft, cruising speed, cruising level, and waypoints describing the intended route [8], [9].

The data-driven predictive models aim to exploit different data sources to extract relevant trajectory features that could be used to predict aircraft positions. A broad variety of machine learning techniques are applied in the literature. For example, [10] proposed statistical regression models that assumed the aircraft position to be a function of a set of dependent variables. The study concluded that the regression model obtains more accurate predictions compared to the performance model-based (e.g. BADA) approach. In [11], Generalized Linear Models are employed to predict arrival times of descending aircraft. This study identified the aircraft type, initial altitude, and initial ground speed as the input variables with the greatest statistical significance to predict the arrival time. Also, Neural Networks have been widely adopted in data-driven trajectory predictions globally, for example, in [12]–[14], as well as in studies conducted by EUROCONTROL. [15], [16].

Other studies have identified the main sources of uncertainty in aircraft trajectory predictions [17], [18]. These inaccuracies come from errors in position and speed measurement, unknown aircraft mass, and meteorological conditions [2]. The lack of knowledge concerning the operational strategy of the airline and deviation of common ATC practices are also major sources of uncertainty [18], [19]. Another major source of uncertainty occurs in the temporal domain, for example, caused by the inaccurate take-off time prediction [16], [20].

The majority of these studies focus on the evaluation of prediction accuracy, which could either be expressed by the spatial-temporal errors between the prediction and actual trajectory [8], [21]. In addition to quantifying the errors, quantification of prediction uncertainties remains to be a challenging topic for trajectory predictions. How to properly make use of historical flights and flown trajectory of an existing flight also needs further research.

In this paper, we propose a two-stage Gaussian Process Regression (GPR) approach for estimating trajectory prediction errors and uncertainties. One GPR model is designed to consider historical flights with similar flight patterns, while another to incorporate the partially flown trajectories of a flight. Finally, the GPR models provide predictive distributions for quantifying the uncertainties.

II. Core methodology

A. Trajectory clustering

Clustering is used in this study to extract historical flights that have similar flight tracks. We make use of the common DBSCAN method [22] that generates clusters based on density of the sampled flight trajectories. For situations where main flight patterns are known (e.g., in terminal maneuvering area), k-means methods are used in combination with DBSCAN to further refine the clusters according to Standard Instrument Departure Routes and Standard Arrival Routes. We also adopted the simple Euclidean distance metric for the trajectory clustering.

B. The general Gaussian Process regression

Gaussian Processes can be used to represent a collection of random variables that are temporally and spatially related. Any finite set of those variables is assumed to form a multivariate normal distribution. The random variables express the evaluation of a function $f(x)$ at a possibly multidimensional input location x. With training data, GPR learns the underlying distribution, which could be fitted by potentially infinitely many functions. GPR provides an elegant approach to assign a probability to each of these functions [23], where the mean of this distribution represents the most probable regression model of the data.

We can formulate the GPR as mean function $m(x)$ and covariance function $k(x_i, x_j)$.

For simplicity, the mean function $m(x)$ is often assumed to be zero. The characteristics of $f(x)$ are fully specified by a combination of covariance functions (or kernels) $K(X, X)$, which essentially specifies the correlation between different data points.

A variety of kernels are discussed in literature [23], [24]. Once the covariance function is selected and the hyperparameters are learned from training data, GPR can be applied to estimate the value of a function evaluated at any set of new inputs $X*$. The joint distribution of (possibly noisy) observations y and predicted values $y*$ is expressed as a multivariate normal distribution. The predictive distribution y^*, conditional on the training data (X, y) and the provided test data X^*, is then represented as follows:

$$\mathrm{y}^* \mid X^*, X, y \sim \mathcal{N}(\mu, \Sigma) \tag{1}$$

where μ and Σ can be calculated as:

$$\begin{aligned} \mu &= \mathrm{K}\left(\mathrm{X}^*, \mathrm{X}\right) \mathrm{K}(\mathrm{X}, \mathrm{X})^{-1} \mathrm{y} \\ \Sigma &= \mathrm{K}\left(\mathrm{X}^*, \mathrm{X}^*\right) - \mathrm{K}\left(\mathrm{X}^*, \mathrm{X}\right) \mathrm{K}(\mathrm{X}, \mathrm{X})^{-1}\ \mathrm{K}\left(\mathrm{X}, \mathrm{X}^*\right) \end{aligned} \tag{2}$$

Here, parameters X, y, and X^* are obtained from trajectory data. The choice of kernel(s) K is described in the next section.

C. A two-stage GPR approach for trajectory prediction

Three main building blocks are required to train two GPR models, which are training data, kernel functions, and hyperparameters.

Firstly, A training dataset is constructed that contained both the target variables (position and altitude) and predictor variables. We tested three different models that are trained with different predictors (Table II).

Secondly, a covariance function (kernel) is selected. This function highly influences the shape of the predicted trajectory. Several kernels are commonly available for modeling these relatively smooth functions. They include Radial Basis Function (RBF) kernel, RBF kernel, Rational Quadratic Function kernel, and the Matérn kernel. These kernels are combined with a linear kernel, which accounts for modeling straight segments.

Thirdly, the alpha parameter is specified to prevent numerical issues during fitting and could be interpreted as the additional variance on the training data.

Table I: Three different GPR models trained on different sets of predictor variables.

GPR Model	Predictor variables
Model-A	ADS-B data only
Model-B	ADS-B and Flight Plan data
Model-C	ADS-B, Flight Plan, and ERA5 data

The first stage aims to develop a *History GPR* model (G_H) based on the training data from historical trajectories for each cluster. To limit the processing time, the number of data points that form each trajectory is reduced with the Ramer-Douglas-Peucker (RDP) algorithm [25], which simplifies a curve connected by points, by representing that curve with fewer points. We make use of k-fold cross-validation to identify the best-fitted kernel and its corresponding hyperparameters. The final G_H model is a stochastic regression model that allows us to sample flight data points from a cluster based on the predictive distributions of GPR models.

Next, we develop another *Predictive GPR* model (G_P), which is to be trained on a specifically constructed set of data for each flight. Such dataset is constructed as follows:

- We extract the current flight data up to the prediction horizon T_0.
- We sample from G_H specific to the cluster a set of data with flight parameters beyond T_0, represented by historical data.
- The means of the samples are aggregated with the partially flown trajectory.

Figure 1 provides an example of the construction of training data, which shows both the historical observations ($< T_0$) as well as the G_H samples ($> T_0$). In this paper, we use the time when the trajectory reached FL250.

The final dataset (represented by the solid and empty dots in Figure 1) is used to develop G_P using the same procedure for training G_H. Finally, the variance of the samples beyond T_0 from G_H helps to identify the hyperparameters parameter on the training data.

In Figure 2, we can see the sample data from G_H model (in gray), prediction of trajectory from G_P model (in red), and actual flight trajectory (in green). We can also observe

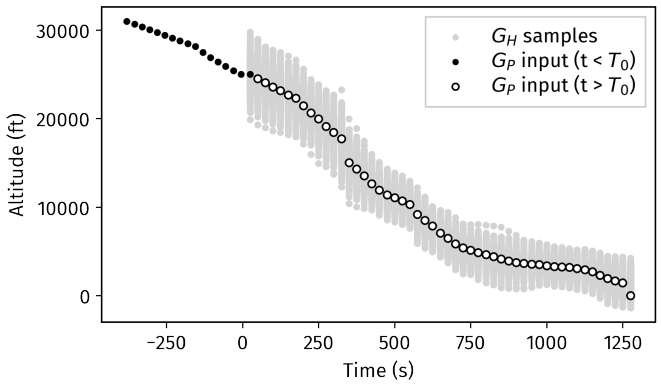

Figure 1: Data for constructing G_P, based on flown trajectory and samples from historical GPR model G_H.

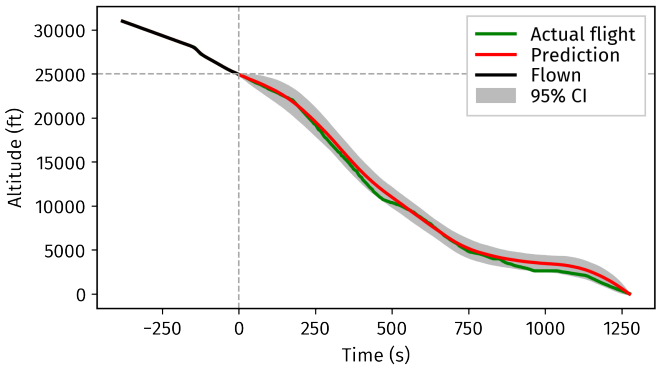

Figure 3: Obtain 95% confidence interval from predictive GPR mode G_P.

that the horizontal ground track profile has a large error when compared to the altitude profile.

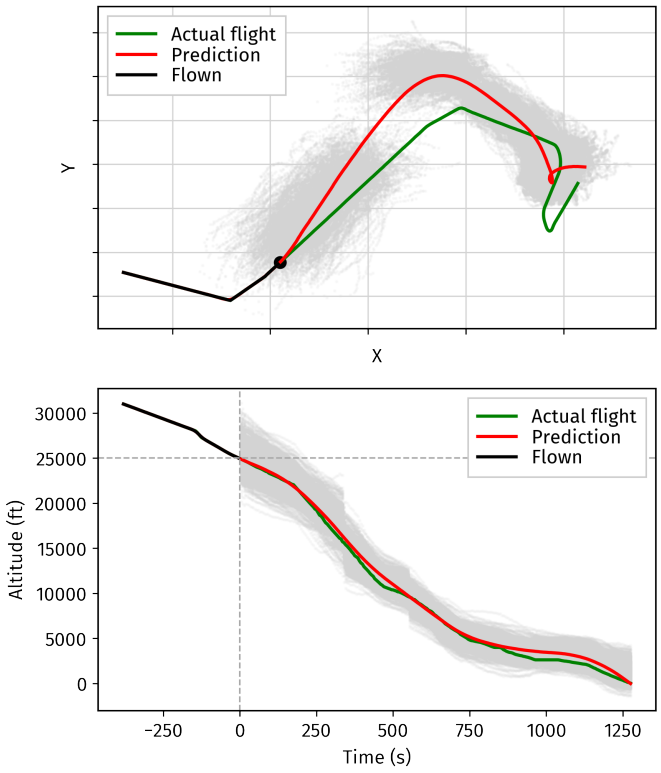

Figure 2: An Example of a predicted flight trajectory

From G_P model, we can also construct the prediction confidence interval based on the predictive distributions, as shown in Figure 3,

D. Error and uncertainty metrics

Based on posterior distributions of GRP models, both prediction uncertainty and accuracy can be studied. The mean of the predictive distribution is considered as the most probable prediction and thus used to compute the accuracy metrics. The accuracy of the predictions is considered as the spatial and temporal errors between the actual and predicted trajectory. The errors are:

- *Horizontal errors*: They rely on the flat earth approximation base on relatively small distances that are covered. Specifically, the along-track error (ATE) and the cross-track error (CTE) are used.
- Vertical error: It is the difference in altitude between the predicted and actual trajectory, with a negative error indicating that the predicted position is lower than the actual aircraft position.
- Temporal error: It is measured at FL100 (initial approach fix) and indicates the difference in time when the predicted- and actual trajectory have reached this altitude.

The uncertainty is quantified by computing the standard deviation (σ) of the sampled predictions of the 3D-position, indicating the spread of the predicted position of the aircraft.

Horizontal errors, vertical errors, and uncertainty measures are evaluated at every 1000 feet starting from the start of the prediction horizon at T_0. Besides, different look-ahead times are evaluated to analyze the progression of the predictive metrics over time.

III. Experiment

A. ADS-B and aircraft data

The ADS-B data provide the basis for the flight trajectories in the final dataset. This data is gathered by a Mode S receiver located at TU Delft, with a coverage of approximately 400 kilometers radius. Positions and velocities are decoded and aircraft are identified by their unique 24-bit Mode S transponder code. Individual flights are extracted from the decoded ADS-B data [26], and we select the ones approaching Amsterdam Airport Schiphol (EHAM) for the experiments.

In addition, we use the aircraft database from the OpenSky network, to provide additional information about the aircraft. Based on the ICAO transponder code, information like aircraft type code, registration, and operator are aggregated to each flight.

With this dataset, we also aggregate the Wake Turbulence Category (WTC), which are heavy, medium, and light. ICAO

specifies the WTC based on the Maximum Take-Off Weight (MTOW) of the aircraft.

Based on the airline information, we also identified the operation type of the flight, including cargo, business aviation, unscheduled (e.g. charter), low-cost flight, or traditional flight.

B. Flight plan and intent data

The aircraft intent is expressed using information extracted from flight plans from the Eurocontrol R&D data archive [27]. Each ADS-B observation is aggregated with its corresponding ECTRL ID using the ICAO registration and rounded timestamps of the actual flight points.

The aircraft intent is expressed as three next waypoints at each location in the ADS-B dataset. To identify these three waypoints, the distance from departure airport is computed to establish a variable that describes the progress of the flight. Based on this variable, the next three waypoints are selected from the filed flight points. Each waypoint comprises the latitude, longitude, and altitude of the aircraft together with a time component:

$$\Delta t_{\text{wp}} = t_{\text{plan}} - t_{\text{actual}} \tag{3}$$

which expresses the difference in total flight time up till the specific waypoint (t_{plan}) and the actual flight time since take-off as observed from the ADS-B record (t_{actual})

C. Meteorological data

Meteorological forecasts are extracted from the ERA5 database. This database provides estimates of a large variety of meteorological parameters on an hourly basis. The data is formatted in a grid with a spatial resolution of 30 kilometers and divides the atmosphere into 137 different pressure levels up to a height of 80 kilometers.

For this study, the wind speeds in three dimensions and the temperature are extracted from the database. Since the available ADS-B data covers a sub-region of Europe, the extraction of meteorological data is limited to this region with longitudes ranging from -10 to 30 degrees and latitudes ranging from 30 to 70 degrees.

A simple linear interpolation model is developed to express the parameters as a function of the four dimensions (latitude, longitude, altitude, and time). This function is evaluated at the given ADS-B records, such that each observation would be aggregated with the meteorological forecasts.

D. Data preparation

As observed from Figure 4, a variety of preparation steps are executed to construct the final dataset. These steps are described below.

1) We first extract only the descent trajectory using the fuzzy logic identification process proposed by [28].
2) Common data cleaning and filtering tasks are applied. The final dataset comprised partial descent trajectories whose initial data point is found above 25,000 feet (FL250), while the final data point at the arrival airport.

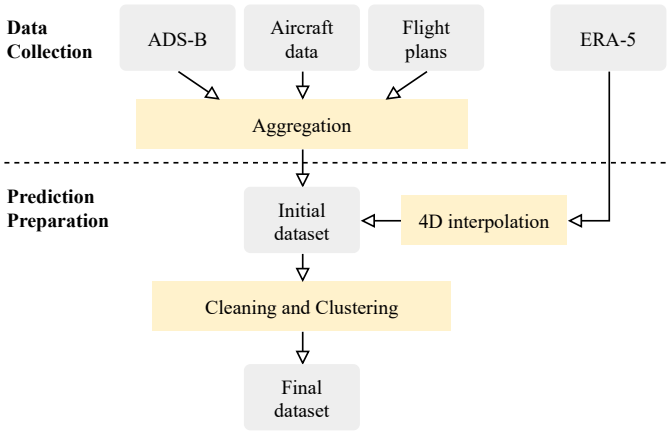

Figure 4: Process overview of the collection and preparation of data.

3) Latitude and longitude are transformed to Cartesian coordinate. All categorical features are converted to numeric variables using one-hot encoding.

An overview of the features that are used as predictors to the predictive models is found in Table II.

Table II: Features used in trajectory prediction models.

ADS-B	FP[1]	ERA5	Other
rate of climb [ft/min]	x_i [m]	wind x [m/s]	WTC [-]
ground speed [kts]	y_i [m]	wind y [m/s]	Market [-]
track [deg]	alt_i [ft]	wind z [Pa/s]	
	t_i [s]	temp. [K]	

[1] The upcoming three waypoints are included (i ∈ {1,2,3})

IV. Results

The final dataset comprises around 10 thousand arrival trajectories at EHAM Airport in June 2018. Based on the aforementioned clustering approach, we have identified seven clusters of arrival trajectories, which are shown in Figure 5. For the analysis in the rest of this section, the GPR predictive models are applied to all trajectories. Error metrics are obtained for each cluster separately.

A. GPR prediction errors and uncertainties

Figure 6 summarizes the errors and uncertainty for each clusters. The results visualize the distribution of the median value for each metric evaluated over each individual predicted trajectory from a cluster. Three different GPR models with an increasing number of predictors are evaluated (see Table I).

The mean vertical error is found to be centered around zero for all clusters and models. However, Model-A consistently shows the largest spread of vertical errors. Especially in clusters 0 till 4, Model-A provides the largest ATE, while Model-B and Model-C show comparable results. The distribution of CTE shows smaller order of magnitudes compared to the ATE, with negligible differences among the three models.

The predictive uncertainties for position predictions are visualized in Figure 7. The contours in the plot show the

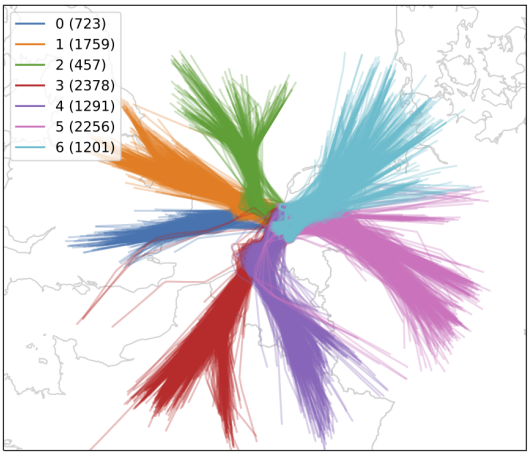

Figure 5: Clusters of arrival flights at EHAM. Labels shows the cluster name and number of flights.

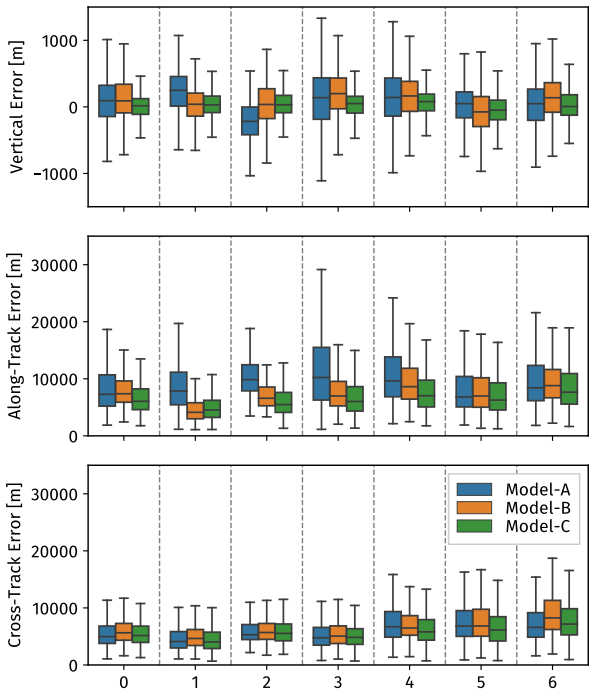

Figure 6: Predictive metrics of all clusters showing the predictive errors (top row) and the standard deviation of the predictive distributions (bottom row).

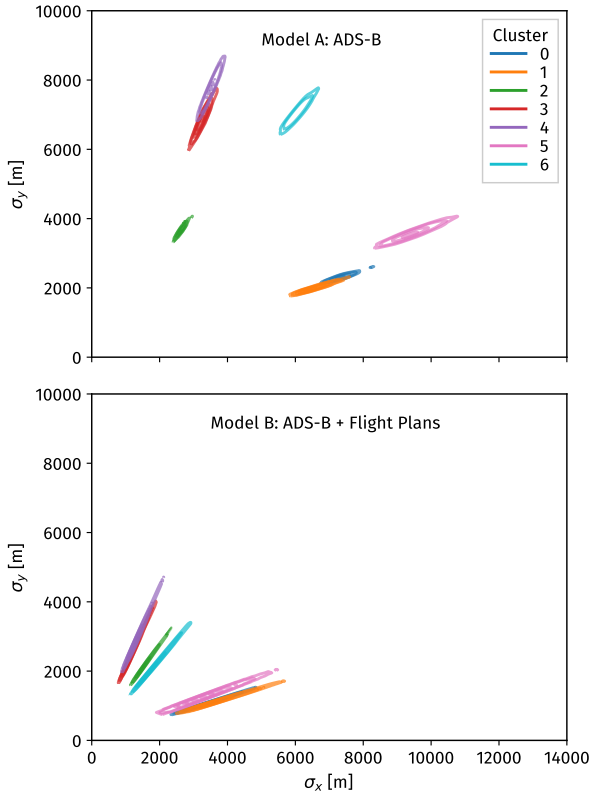

Figure 7: Contour plot showing the average standard deviations of the prediction of the horizontal position for each cluster obtained from model Model-A and Model-B. Outer line of each contours corresponding to 5% of the probability mass.

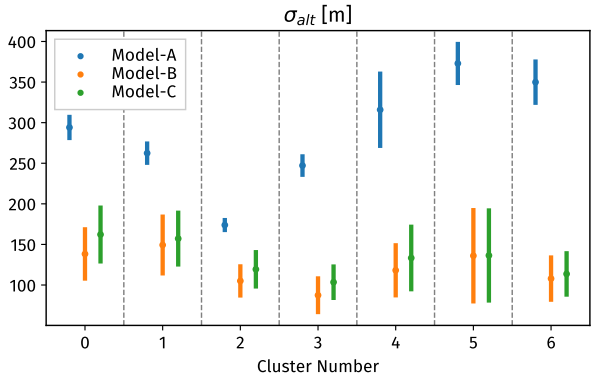

Figure 8: Vertical prediction uncertainties obtained based on 95% confidence intervals

average standard deviations in each cluster's position predictions. The uncertainties are much higher in the direction of along-track than the cross-track directions for each cluster. We can also qualify the reduction in prediction uncertainty by introducing flight intent information in the prediction (bottom plot).

Vertical uncertainties can be more easily visualized. In Figure 8, the error bars illustrate the variation in the confidence intervals that are obtained by GPR models for all trajectories, grouped by cluster number.

B. *Errors at different flight levels*

To further analyze the accuracy and uncertainty of the predictions at different flight levels, we now focus on the results for one cluster (cluster 3). The predictive metrics are evaluated at different flight levels to analyze the progression of the predictive capability over the descent profile (Figure 9).

Until FL070, the distribution of vertical errors is centered around zero with Model-C showing the smallest variance in predictive errors. During the final stage of the descent, below FL070, all models tend to overestimate the altitude of the

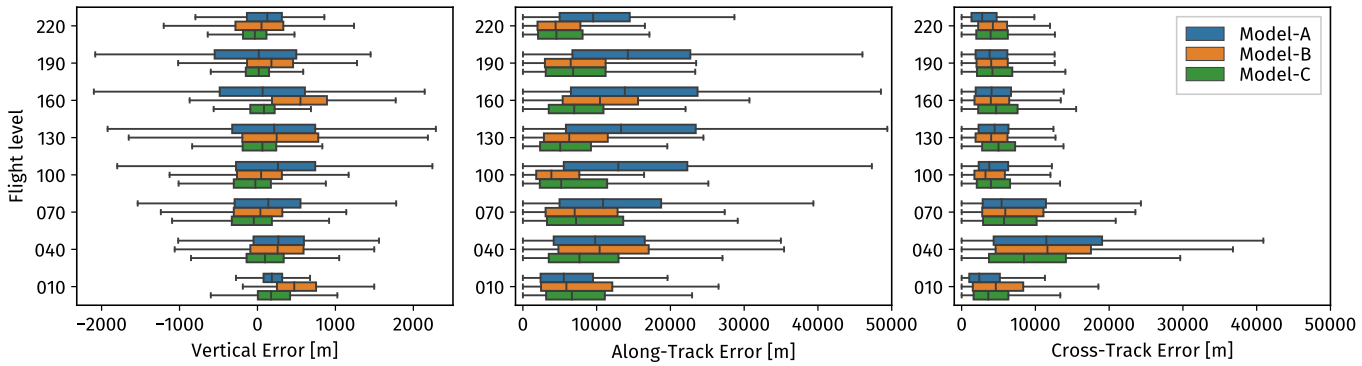

Figure 9: Predictive metrics of cluster 3 showing the predictive errors (top row) and the standard deviation of the predictive distributions (bottom row) for a variety of flight levels.

aircraft. Initially, at FL240, the ATE does not deviate in each model. Hereafter, Model-B and Model-C produce significantly smaller errors. Below FL070, the three models produce comparable results in terms of the ATE. While Model-B and Model-C improve the spatial accuracy along the flight track in the initial stage of the prediction horizon, the CTE does not differ among the different models. The CTE remains constant in the initial stage of the descent but rapidly increases once the altitude drops below FL070. Eventually, the errors decrease again once the aircraft gets closer to its final destination. It is worth noting that at FL040, the cross-track error increases drastically due to the vectoring procedures frequently executed in EHAM.

C. *Errors and uncertainties at different look-ahead time*

Figure 10 shows the accuracy metrics for all models evaluated at different look-ahead times for flights from cluster 3. It can be seen that after 5 minutes, the vertical and along-track errors remain at a similar level for all three models. However, the cross-track errors do increase with a longer look-ahead time. This is likely to the vectoring procedure, which is commonly practiced at EHAM.

Figure 11 shows the prediction uncertainties at different look ahead intervals. Similarly, we can see that the uncertainties also remain at the same level for each model after 5 minutes.

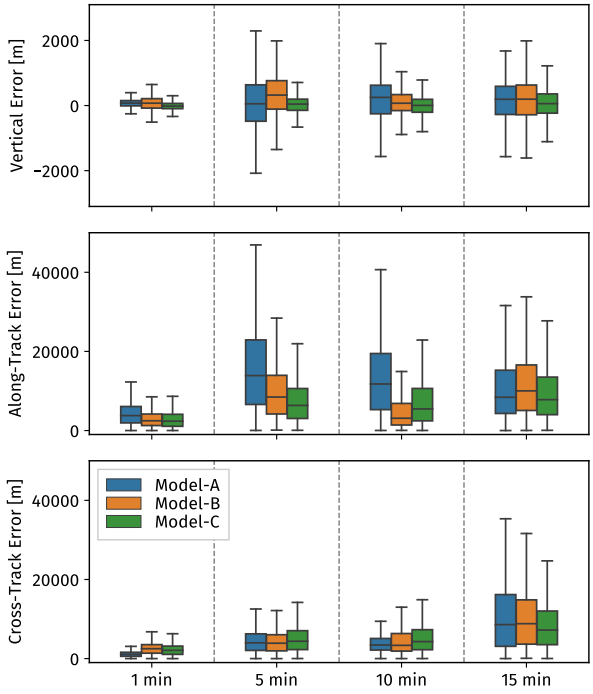

Figure 10: Comparison of predictive accuracy with the GPR models for increasing look-ahead times (cluster 3)

D. *Effect of aircraft and airline type*

In this study, we also analyze the effects of the aircraft and airline type in the prediction models. This is evaluated by extending model Model-C with training data that incorporated predictor variables describing the WTC category and the airline market segment.

Half of the flights are operated with medium WTC, while the other half is operated by heavy WTC. The vast majority of airlines operate in the traditional scheduled market segment. The indices of the predictive metrics, relative to model Model-C, are presented in Table III to show the relative difference between the models. A comparison of both models shows that the standard deviation of the predicted 3D-position increases by 33%, while the horizontal and vertical errors of the predictions are comparable (Table III). This has shown that aircraft WTC and operation information do not bring significant improvement in accuracy, while largely increasing the prediction uncertainties due to the increased dimensionalities in features.

Table III: Comparison between model Model-C and an extended model including aircraft and airline data

Model	**HTE**	**VE**	σ_x	σ_y	σ_{alt}
Model-C	100	100	100	100	100
Model-C (extended)	104	98	133	133	133

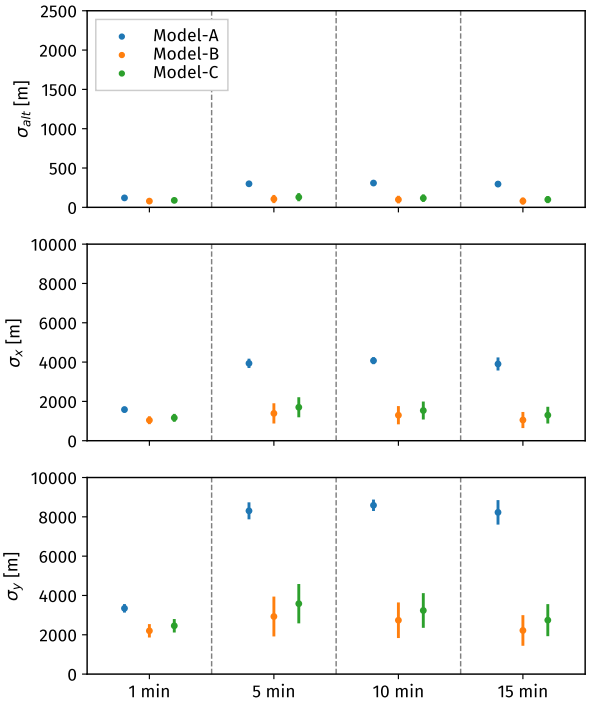

Figure 11: Comparison of the predictive uncertainty, expressed by the standard deviation of the predictions,for increasing look-ahead times (cluster 3)

V. Discussion

Three different GPR models with varying sets of predictor variables are trained based on the different clusters of trajectories according to arrival patterns. The results showed that the predictive accuracy could be improved when incorporating flight plan data and meteorological data when training the GPR models. Generally, the distribution of the results obtained from Model-C shows smaller variances, which indicates that this model obtains more consistent results compared to Model-A. The major improvements are observed in clusters 0 up till 4, which showed significant reductions in the ATE. The differences in spatial accuracy between the models in clusters 5 and 6 are less distinct. These clusters are less distinctive as both show a larger spread in trajectory shapes. The effect of including more predictor variables diminishes when the model is trained on less distinct clusters. This proves the importance of the effective clustering of trajectories before training the data-driven models.

The results also showed that the predictive uncertainty, expressed by the standard deviation of the predictions, of Model-A is considerably higher compared to Model-B and Model-C. Together with the improvement in accuracy, this proves the benefits of including aircraft intent in data-driven trajectory prediction. However, in contrast to our initial hypothesis, further extending the model with meteorological data does not result in further improvement (at least for our test case).

The standard deviation of the predicted x- and y-position is largely dependent on the direction of the flight track. Trajectories oriented in either a Northerly or Southerly direction (Cluster 2, 3, 4) showed larger deviations in the prediction y-coordinate, while the opposite effect is found for trajectories flying in either an Easterly or Westerly direction (Cluster 0, 1). To better describe the uncertainty, along-track and cross-track should also be used. However, constructing such uncertainty metrics is less obvious than error metrics.

The predictive metrics are evaluated for different flight levels to investigate the progression of the predictive capability of the models along with the descent profile. Above FL100, the effect of including aircraft intent and meteorological data results in the reduction of the ATE compared to model Model-A. However, the results showed the difficulty of predicting the final stage of the descent trajectory below FL040, where both the vertical error and the CTE increase significantly for all models. During this final stage, the flights are subjected to ATC commands that guide the aircraft to the appointed runway. This causes the flight tracks, within a single cluster, to diverge in this final stage of the flight as aircraft are assigned to different approach tracks. This complicates the training of the models and results in larger predictive errors, with large increases in the CTE. Also, the number of waypoints in a filed FP below FL100 is sparse. Generally, only two or three waypoints describe the aircraft's intent in this final stage. This causes the effect of FP data to diminish in this final stage. The initial stage of the descent, between the top of descent and FL100, is usually represented by more waypoints.

The uncertainty of the predictions, quantified by the GPR models, initially increases right after the start of the prediction horizon. Hereafter, the standard deviation gradually decreases until FL100 is reached. The Initial Approach Fix (IAF) is located at this flight level. The aircraft proceeds from the en-route segment to the IAF to start the initial segment of the instrument approach. Therefore, many routes will converge to the IAF, which is captured by the GPR models as shown by the decreasing uncertainty until FL100 is reached.

In this study, we also found that the multivariate GPR in the common GPR implementation (Scikit-Learn) does not consider the correlation among different predictors. Most GPR implementations found in the literature treated the multidimensional case by modeling each response variable individually without considering the correlation between the variables [24]. The key challenge in modeling multivariate response variables in GPR is the specification of a covariance function that both incorporates the correlation between data points as well as the correlation among target variables [29]. This could be improved in future research.

VI. Conclusion

In this paper, the Gaussian Process Regression is proposed for quantifying data-driven trajectory accuracies and uncertainties. At first, the trajectories of flights arriving at Schiphol Airport are clustered. Flights in each cluster are then treated with a two-stage GPR approach to study the predictability.

GPR is a pure data-driven prediction approach. This has the advantage that it does not require explicit knowledge of source uncertainty to study the prediction errors and uncertainties. The quantification of the predictive uncertainty can also contribute to the improvement of the prediction and management of 4D-trajectories. The two-stage GPR method is able to utilize both information from all historical flights and the past trajectory data from the flight of interest.

Based on the GPR models, we confirm that uncertainty of the descent trajectory predictions could be reduced by incorporating flight plan data when training the models. Additional meteorological data does not result in a significant reduction of prediction uncertainty but do show an improvement in predictive accuracy.

The main improvements are observed throughout the initial stage of the descent. The uncertainty of the GPR predictions decreased until FL100 is reached, which is the effect caused by the IAF that represents the position where the aircraft trajectories are merged to initiate the approach segment. In the final stage of the descent, the predictive errors increase due to the complex dynamics of arrival procedure and ATC commands that guide the aircraft to the appointed runway.

This study evaluated the predictive models on the descent segment of the flight. Additional research could be performed to apply the probabilistic predictive models to different phases of the flight, like the climb and cruise phases. Also, this study focused on the prediction of single trajectories without considering the interactions with other aircraft. A collaborative trajectory predictor that fuses the predicted trajectories of multiple aircraft has the potential to further improve the predictive accuracy.

References

[1] A. Rodríguez-Sanz, F. Gómez Comendador, R. M. Arnaldo Valdés, J. A. Pérez-Castán, P. González García, and M. N. G. Najar Godoy, "Air Traffic Management based on 4D Trajectories: A Reliability Analysis using Multi-State Systems Theory," *Transportation Research Procedia*, vol. 33, pp. 355–362, 2018. [Online]. Available: https://doi.org/10.1016/j.trpro.2018.11.001

[2] E. Casado, C. Goodchild, and M. Vilaplana, "Identification and Initial Characterization of Sources of Uncertainty Affecting the Performance of Future Trajectory Management Automation Systems," *Proceedings of the 2nd International Conference on Application and Theory of Automation in Command and Control Systems*, no. May, pp. 170–175, 2012.

[3] S. Sankararaman and M. Daigle, "Uncertainty quantification in trajectory prediction for aircraft operations," *AIAA Guidance, Navigation, and Control Conference, 2017*, pp. 1–11, 2017.

[4] J. Rudnyk, J. Ellerbroek, and J. M. Hoekstra, "Trajectory prediction sensitivity analysis using monte carlo simulations," *2018 Aviation Technology, Integration, and Operations Conference*, pp. 1–20, 2018.

[5] R. Alligier, D. Gianazza, and N. Durand, "Learning the aircraft mass and thrust to improve the ground-based trajectory prediction of climbing flights," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 45–60, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.trc.2013.08.006

[6] I. Lymperopoulos and J. Lygeros, "Improved multi-aircraft ground trajectory prediction for air traffic control," *Journal of Guidance, Control, and Dynamics*, vol. 33, no. 2, pp. 347–362, 2010.

[7] W. Kun and P. Wei, "A 4-D trajectory prediction model based on radar data," *Proceedings of the 27th Chinese Control Conference, CCC*, pp. 591–594, 2008.

[8] M. A. Konyak, "Improving Ground-Based Trajectory Prediction through," no. August, pp. 1–25, 2009.

[9] R. A. Coppenbarger, "En route climb trajectory predicition enhancement using airline flight-planning information," *1999 Guidance, Navigation, and Control Conference and Exhibit*, pp. 1077–1087, 1999.

[10] M. G. Hamed, D. Gianazza, M. Serrurier, and N. Durand, "Statistical prediction of aircraft trajectory: regression methods vs point-mass model," *10th USA/Europe ATM R&D Seminar*, vol. 2013, no. June, p. pp, 2013. [Online]. Available: https://hal-enac.archives-ouvertes.fr/hal-00911709

[11] A. M. de Leege, M. M. van Paassen, and M. Mulder, "A machine learning approach to trajectory prediction," *AIAA Guidance, Navigation, and Control (GNC) Conference*, pp. 1–14, 2013.

[12] Y. Pang, N. Xu, and Y. Liu, "Aircraft trajectory prediction using lstm neural network with embedded convolutional layer," *Proceedings of the Annual Conference of the Prognostics and Health Management Society, PHM*, vol. 11, no. 1, 2019.

[13] K. Yang, M. Bi, Y. Liu, and Y. Zhang, "LSTM-based deep learning model for civil aircraft position and attitude prediction approach," *Chinese Control Conference, CCC*, vol. 2019-July, pp. 8689–8694, 2019.

[14] Z. Wang, M. Liang, and D. Delahaye, "Short-term 4D trajectory prediction using machine learning methods," *SESAR Innovation Days*, no. November, 2017.

[15] H. Naessens, T. Philip, M. Piatek, K. Schippers, and R. Parys, "Predicting flight routes with a deep neural network in the operational air traffic flow and capacity management system," *EUROCONTROL Maastricht Upper Area Control Centre, Maastricht Airport, The Netherlands, Tech. Rep*, 2017.

[16] R. Dalmau, F. Ballerini, H. Naessens, S. Belkoura, and S. Wangnick, "Improving the predictability of take-off times with machine learning," *Proceedings of the Ninth SESAR Innovation Days, Athens, Greece*, pp. 2–5, 2019.

[17] P. Weitz, "Determination and visualization of uncertainties in 4D-trajectory prediction," *Integrated Communications, Navigation and Surveillance Conference, ICNS*, pp. 1–9, 2013.

[18] K. T. Mueller, J. A. Sorensen, and G. J. Couluris, "Strategic aircraft trajectory prediction uncertainty and statistical sector traffic load modeling," *AIAA Guidance, Navigation, and Control Conference and Exhibit*, no. August, pp. 1–11, 2002.

[19] E. Casado, M. L. Civita, M. Vilaplana, and E. W. McGookin, "Quantification of aircraft trajectory prediction uncertainty using polynomial chaos expansions," *AIAA/IEEE Digital Avionics Systems Conference - Proceedings*, vol. 2017-Septe, no. 699274, 2017.

[20] S. Badrinath, H. Balakrishnan, E. Joback, and T. G. Reynolds, "Impact of off-block time uncertainty on the control of airport surface operations," *Transportation Science*, vol. 54, no. 4, pp. 920–943, 2020.

[21] J. Bronsvoort, G. McDonald, M. Paglione, C. Garcia-Avello, I. Bayraktutar, and C. M. Young, "Impact of missing longitudinal aircraft intent on descent trajectory prediction," *AIAA/IEEE Digital Avionics Systems Conference - Proceedings*, pp. 1–14, 2011.

[22] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *kdd*, vol. 96, no. 34, 1996, pp. 226–231.

[23] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Massachusetts Institute of Technology, 2006. [Online]. Available: www.GaussianProcess.org/gpml

[24] S. A. Goli, B. H. Far, and A. O. Fapojuwo, "Vehicle Trajectory Prediction with Gaussian Process Regression in Connected Vehicle Environment," *IEEE Intelligent Vehicles Symposium, Proceedings*, vol. 2018-June, no. Iv, pp. 550–555, 2018.

[25] L. Basora, J. Morio, and C. Mailhot, "A trajectory clustering framework to analyse air traffic flows," *7th SESAR Innovation Days*, pp. 1–8, 2017.

[26] J. Sun, J. Ellerbroek, and J. Hoekstra, "Flight extraction and phase identification for large automatic dependent surveillance-broadcast datasets," *Journal of Aerospace Information Systems*, vol. 14, no. 10, pp. 566–571, 2017.

[27] "R&D Data Archive." [Online]. Available: https://www.eurocontrol.int/dashboard/rnd-data-archive

[28] J. Sun, J. Ellerbroek, and J. Hoekstra, "Flight Extraction and Phase Identification for Large Automatic Dependent Surveillance–Broadcast Datasets," *Journal of Aerospace Information Systems*, vol. 14, no. 10, pp. 566–571, 2017.

[29] B. Wang and T. Chen, "Gaussian process regression with multiple response variables," *Chemometrics and Intelligent Laboratory Systems*, vol. 142, pp. 159–165, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.chemolab.2015.01.016