# High Confidence Intervals Applied to Aircraft Altitude Prediction

Mohammad Ghasemi Hamed, Richard Alligier, David Gianazza

**HAL Id: hal-01284915**

**https://hal-enac.archives-ouvertes.fr/hal-01284915**

Submitted on 8 Mar 2016

# High Confidence Intervals Applied to Aircraft Altitude Prediction

Mohammad Ghasemi Hamed[a], Richard Alligier[b,c], David Gianazza[b,c]

[a]*Lab-STICC UMR CNRS 6285, ENSTA Bretagne, F-29806 Brest, France*
[b]*ENAC, MAIAA, F-31055 Toulouse, France*
[c]*Univ. de Toulouse, IRIT/APO, F-31400 Toulouse, France*

*Abstract*—This paper describes the application of high confidence interval prediction methods to the aircraft trajectory prediction problem, more specifically to the altitude prediction during climb. We are interested in methods for finding two-sided intervals that contain, with a specified confidence, at least a desired proportion of the conditional distribution of the response variable. This paper introduces Two-sided Bonferroni-Quantile Confidence Intervals (TBQCI), which is a new method for obtaining high confidence two-sided intervals in quantile regression. The paper also uses the Bonferroni inequality to propose a new method for obtaining tolerance intervals in least-squares regression. This latter has the advantages of being reliable, fast and easy to calculate. We compare physical point-mass models to the introduced models on an Air Traffic Management (ATM) dataset composed of traffic at major French airports. Experimental results show that the proposed interval prediction models perform significantly better than the conventional point-mass model currently used in most trajectory predictors. When comparing with a recent state-of-the-art point-mass model with adaptive mass estimation, the proposed methods give altitude intervals that are slightly wider but more reliable.

*Index Terms*—High Confidence Interval Prediction, Tolerance Intervals, Quantile Regression, Bonferroni inequality, Aircraft Trajectory Prediction, Point-mass Model, BADA

## I. Introduction

Predicting aircraft trajectories with high confidence is fundamental to most operational concepts ([1], [2]) and is necessary to the automated tools that are expected to improve Air Traffic Management (ATM) in the near future. The automated detection and resolution of air traffic conflicts is one of the key applications of aircraft trajectory prediction. It requires a reliable prediction of the future aircraft positions in order to detect trajectory conflicts, and propose maneuvers to avoid these conflicts while minimizing trajectory deviations.

In this context, we are interested in providing a more accurate and realistic prediction of the volumes containing the future aircraft positions, based on actual radar data. As a first step, we focus on the vertical dimension, and more specifically on the aircraft climb where the uncertainties are known to be huge. Most trajectory predictors use simplified physical equations of the forces applied to the center of gravity of the aircraft. Such point-mass models require knowledge of the aircraft state (position, speed, mass, *etc.*), atmospheric conditions (wind, temperature), and aircraft intent (target speed or climb rate, for example). However, much of this information is not available to ground-based systems, and the information that is available is noisy, thus making the predictions inaccurate.

The literature on ground-based aircraft trajectory prediction is fairly large, and one may refer to [3] for a literature survey on the subject or [4] for the statistical analysis and validation of trajectory predictors. Concerning the prediction model itself, most research studies focus on point-mass models of the aircraft dynamics. Several articles propose adaptive methods adjusting some of the model parameters, such as the modeled thrust ([5]), or the modeled mass ([6], [7], [8], [9], [10], [11]). A few works use data-driven approaches or Machine Learning methods (e.g. [12], [13], [14], [15]). Some recent works ([16]) mix the data-driven and physics-based approaches by learning from historical data some of the missing inputs of the point-mass model, such as the thrust law. All these works try to improve the accuracy of the predicted nominal trajectory, and do not consider the prediction of reliable altitude intervals as their primary goal.

The subject of the current paper is to predict altitude intervals that are guaranteed to contain the future altitude with a chosen level of confidence. Considering all the unknowns and uncertainties in the parameters of the physical model, we propose to try a data-driven approach, where the prediction model is tuned on historical data. In a previous publication ([17]), we applied regression methods to the prediction of altitude intervals for the climbing aircraft. These preliminary experiments showed that regression models (linear, local linear and neural network) are much more efficient than one of the most popular point-mass models, the Eurocontrol BADA (Base of Aircraft Data) model [18] that uses default values for the parameters missing in ground systems (speed intent, thrust law, and aircraft mass).

In this paper, we introduce two new methods based on the Bonferroni inequality – one in the context of least-squares regression, and the other in the context of quantile regression – to obtain intervals that contain at least a desired proportion $\beta$ of the distribution of the conditional response variable, with a specified confidence level $\gamma$. Such $\beta$-content $\gamma$-coverage intervals are referred to as *tolerance intervals* in least-squares regression. In quantile regression, one can also define *two-sided confidence intervals* having similar content and coverage properties. The one we introduce will be denoted *Two-sided Bonferroni-Quantile Confidence Interval* (TBQCI) in the rest of this document.

We then apply the two proposed methods to the prediction of altitude intervals for climbing aircraft, using linear regression models. We compare our methods with the baseline

BADA total energy model, to which we have added a term modeling the influence of the wind on the aircraft dynamics (see section V-A, or [16], [9], [19], [20] for the equations). We also compare our approach to a state-of-the-art adaptive method recently introduced in [8], [9]. This method estimates the aircraft mass from the past trajectory points.

These comparisons are made using Paris Orly and Paris Charles-de-Gaulle Air Traffic radar reports over two months (July 2006 and January 2007). As a first step, the results in the current study are presented for a look-ahead time of 10 minutes and with an initial time $t = 0$ corresponding to a chosen altitude (FL180[1]). This allows us to compare the different methods in a context that is representative of the intended application (conflict detection and resolution in the *en-route* airspace). The proposed approach could easily be extended to various initial altitudes and look-ahead times (see II-D).

The remainder of this paper is organized as follows: Section II describes the context and intended application of the proposed altitude interval prediction methods. Section III provides background on least-squares and quantile regression. Section IV describes how tolerance intervals and confidence intervals on regression quantiles can be applied to the problem of high confidence interval prediction. Both TBQCI and the new Bonferroni tolerance interval for least-squares regression are also introduced in the same section. Section V explains the widely used point-mass model in BADA. The dataset of example trajectories and the patterns extracted from this dataset and used to train the regression models are described in Section VI. The experimental results are shown in Section VII. A conclusion is provided in the last section.

## II. Context and intended application

### A. Trajectory prediction for conflict detection and resolution

In [21], Kuchar *et al.* classify the multitude of prediction models that can be found in the literature on conflict detection and resolution (CD&R) in three categories : nominal, worst-case, or probabilistic. The nominal method uses the best estimate of where the aircraft will be in the future. Its use in CD&R applications is restricted to situations where trajectories are highly predictable (e.g. with a look-ahead time of a few seconds). The worst-case approach detects a conflict if any of a range of maneuvers induces a possibility of conflict with another aircraft. The probabilistic approach models the uncertainties involved in the trajectory prediction process in order to derive a probability of conflict. This is usually done either by adding a position error to a nominal trajectory, or by evaluating the probability of occurence of all possible future trajectories.

Many research efforts have shown the feasibility and interest of the probabilistic approach ([22], [23], [24], [25], [26]), using a variety of methods such as the analytical derivation of the conflict probability, Monte-Carlo simulations, or Bayesian methods. However, many conflict resolution algorithms rely on a modeling approach where the future position of the aircraft is assumed to be in a volume of airspace with a given degree

of certainty (close to 100%, ideally). Such algorithms compute a set of maneuvers that ensure separation of the uncertainty volumes associated to the positions of conflicting aircraft at any time in the near future. The future time is usually bounded by a maximum look-ahead time typically chosen between 10 and 20 minutes. Figure 1 shows an example of such uncertainty volumes, taken from [27], where Durand *et al.* apply genetic algorithms to solve conflicts involving multiple aircraft.

The top view shows an horizontal projection of the volumes at various look-ahead times. In this specific example, the polygons enclosing the future 2D-position are computed by adding an uncertainty to the times at which lateral maneuvers are executed. The bottom view shows a vertical profile with a constant rate of climb. Here, the height of each polygonal volume is simply obtained by adding an uncertainty in the rate of climb.
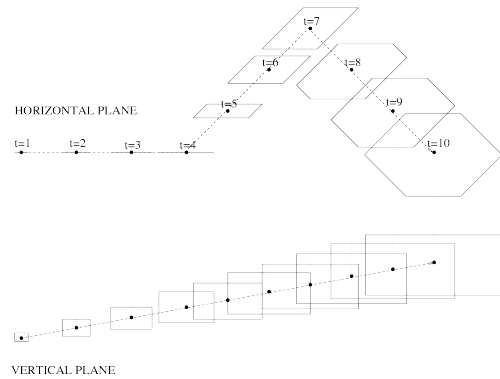


Figure 1. Uncertainty volumes in the horizontal plane (top) and vertical dimension (bottom), at various look-ahead times

Algorithms such as the one proposed in [27] or other population-based metaheuristics ([28], [29]) have proven quite effective in solving large scale problems in a realistic, albeit simulated context. The algorithms of Durand *et al.* ([30], [27], [31]) have been applied to the strategic deconfliction through speed adjustments developed in the European ERASMUS project ([32]). A more recent application is the SESAR 4.7.2 (*Separation Task in En Route Trajectory-based Environment*) project, where lateral and vertical maneuvers are also used. These algorithms do address the combinatorial aspect of the conflict resolution problem. However, their implementations have been tested in simulated environments only. They currently rely on fairly basic assumptions concerning the trajectory prediction itself.

### B. Why not use the trajectory computed on board ?

One could think that downloading the trajectory computed on board the aircraft would suffice to provide us with an accurate prediction. This might be the case for conflict detection, but this is not sufficient for many ground-based applications such as conflict resolution. The metaheuristics cited before, as well as other iterative optimization methods (e.g. the quasi-Newton method used in [33]), require to compute hundreds or

---

[1]Flight level FL180 corresponds to 18000 feet above isobar 1013.25 hPa.

even thousands of candidate trajectories[2]. Such computationally intensive methods would require to constantly downlink a large number of alternative trajectories from each aircraft, which is impractical at the present time. Consequently, we still do need a ground-based trajectory prediction that should be as fast, accurate and reliable as possible.

### C. Data-driven vs. physics-based approach

In this paper, we propose a data-driven approach using regression techniques to build a model predicting altitude intervals for climbing aircraft. Obviously, data-driven methods can only be as good as the data used to build the prediction model. The dataset used to train the model must be as representative as possible of the conditions that will be encountered when making new predictions on fresh inputs with the tuned model. For example, if one uses a dataset of examples containing only climb segments at constant CAS (Calibrated Airspeed) or Mach speed, it would not be a good idea to apply the tuned model to an aircraft operated differently – at a constant rate of climb, for example. One would need a different model for each mode of operation, tuned on specific data.

The physical model of the aircraft does not suffer from such drawbacks : it allows the parameter values (e.g. speed intent, or rate of climb) to be set and possibly changed according to the chosen mode of operation, still using the same model. However, the physical model can only be as good as its inputs (aircraft mass, thrust law, speed intent), which in the case of a ground-based prediction are full of unknowns and uncertainties.

In our case, the intended final application is to detect conflicts in the *en-route* airspace and solve them by assigning simple maneuvers to the aircraft, by modifying the aircraft heading or cruise flight level, or by adjusting the cruise speed (see [27], [31]). Currently, the set of possible maneuvers do not comprise constraints on the climb profile, such as a target speed or thrust law, or a minimum or maximum climb rate. In this context, it is justified to use a data-driven approach on a dataset of recorded traffic containing a variety of climb profiles in order to extract a reliable altitude interval prediction model. Further refinements would be to extract several datasets corresponding to different modes of operation of the aircraft and learn a separate model from each dataset. This is typically what we do when building a dataset for each aircraft type. A more sophisticated approach would be to use mixture models where several models are weighted in order to predict the output.

The data-driven approach has the advantage that it can take account of parameters external to the physical model of the aircraft (e.g. the airline operator, or the distance to go), that can significantly improve the prediction.

### D. Application of regression models to altitude prediction

Regression is a statistical technique which can be used to find the relationship between a numeric variable (response

variable) and some other numeric or non-numeric variables called independent (or explanatory) variables. In order to predict the aircraft altitude, we want to find a function which takes the past aircraft trajectory as argument (and maybe some other interesting inputs) and returns the future altitude (prediction) of the aircraft at a given look-ahead time, as illustrated in Figure 2.
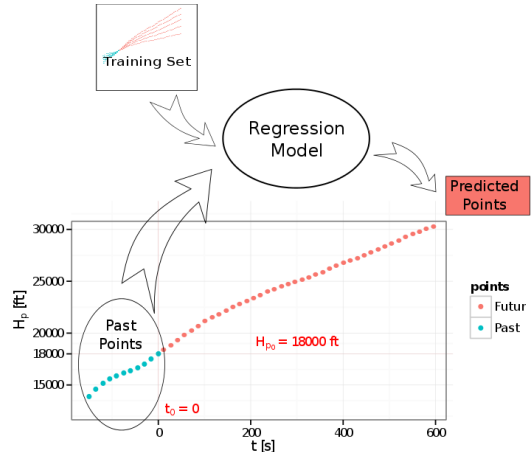


Figure 2. Application of regression models to aircraft trajectory prediction.

Before using it for prediction, the model first needs be tuned on a dataset containing examples of the desired response (here the altitude) associated with some example inputs. This is done by minimizing the overall error between the computed outputs and the desired responses. The generalization performance of the tuned model– i.e. its ability to make correct predictions for fresh inputs – must be assessed on a separate dataset, different from the *training set* used to train the model. In this paper, we use a cross-validation schema (described in Section III-D) to validate our regression based interval prediction models.

In this study, we restrict ourselves to a single initial altitude (flight level FL180) corresponding to the initial time $t = 0$ and a single look-ahead time (10 minutes ahead). The regression approach can be extended to predict a vector of altitudes corresponding to different look-ahead times, for any initial altitude. This could be done by introducing the initial altitude and look-ahead time in the input parameters of the model, and in the dataset of examples. An alternative is to train a specific model for each look-ahead time (or time interval). The latter approach might be the most consistent with the intended application (CD&R), as we might want different levels of confidence for different look-ahead times. This extension of the current study is left for future work.

### E. High-confidence intervals for altitude prediction

The construction of high-level confidence intervals starts with a pointwise prediction of the nominal value (here, the altitude), followed by an estimation of the interval bounds, based on the model properties and the observed model errors. Least-squares regression predicts the conditional mean of the response variable (altitude) for a given input $x$. Quantile regression predicts the chosen quantile (e.g. 25% or 75%) of the conditional distribution. Least-squares and quantile regression

---

[2]As an example, a conflict situation involving 10 aircraft solved by a population-based metaheuristic with 100 elements evolving over 50 generations would require to compute 50000 trajectories (assuming the population is completely renewed at each generation)

also provide a variance for their estimations. Specifying a level of confidence allows us to find, in both cases, an upper bound for the upper quantile and a lower bound for the lower quantile, thus obtaining a high-confidence interval.

Such altitude intervals could give us reliable lower and upper bounds for the polygonal volumes modeling the uncertainties on the future aircraft positions, for the purpose of detecting and solving air traffic conflicts with a maximum look-ahead time of 10 to 20 minutes.

## III. BACKGROUND ON REGRESSION

We choose a fixed regression design where the dataset $\mathcal{S} = (x_1, y_1), \ldots, (x_n, y_n)$ is a random sample. We assume that $x_i$ $(i \in \{1, \ldots, n\})$ is a deterministic vector of observations and that $y_i$ is an observation drawn from the distribution of a random variable $Y_i = Y(x_i)$. In this work, we suppose that there is a true mean regression function $f(\cdot)$ with a zero mean error and an unknown variance $\sigma^2$. $\mathcal{S}$ is a finite random sample, so the estimated regression model is the pair $(\hat{f}, \hat{\sigma})$. Note that if the errors are not normally distributed, the pair $(\hat{f}, \hat{\sigma})$ does not correctly specify the model. Thus in quantile regression, we use the symbol $P_{\mathcal{S}}$ instead of $P_{\hat{f}, \hat{\sigma}}$ to refer to the estimated regression model on the random sample $\mathcal{S}$. We use the following notations:

- $\mathcal{S} = (x_1, y_1), \ldots, (x_n, y_n)$: the random sample of regression;
- $f(x)$: the conditional mean of the response variable for a specified combination of predictors;
- $\hat{f}(\cdot)$: the estimated regression function;
- $\hat{f}(x)$: the estimated regression function at point $x$;
- $\varepsilon$: the error variable;
- $\sigma^2$: the true and unknown variance of the error variable;
- $\hat{\sigma}^2$: the estimated variance of the error variable;
- $Y(x)$: the conditional response variable for a given combination of the predictors, $Y(x) = f(x) + \varepsilon$;
- $\chi^2_{p,n}$: the $p$-quantile of a chi-square distribution with $n$ degrees of freedom;
- $Z_p$: the $p$-quantile of a standard normal distribution;
- $t_{p,n}$: the $p$-quantile of a Student distribution with $n$ degrees of freedom.

### A. Least-Squares Regression

As mentioned in fixed-design regression, the random variable $Y_i$ or $Y(x_i)$ follows a mean function $f(x_i)$ with a random error term $\varepsilon_i$ defined as:

$$Y_i = f(x_i) + \varepsilon_i, \text{ where } E(\varepsilon_i) = 0. \tag{1}$$

The model supposes that the $\varepsilon_i$ are independent and identically distributed (*iid*) random variables. The objective is to estimate the mean function $f(\cdot)$ by $\hat{f}(\cdot)$, being as close as possible to the unknown function $f(\cdot)$. The usual assumption is to suppose that the variance of the error is the same everywhere (homoscedasticity). Least-squares regression takes an estimator $\hat{f}(\cdot)$ that minimizes the Mean of Squared Errors (MSE):

$$MSE(\hat{f}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{f}(x_i))^2 \tag{2}$$

### B. Quantile Regression (QR)

Koenker and Bassett (1978) [34] introduced quantile regression, in which one finds an estimation of conditional quantiles of the response variable. Least squares regression estimates the conditional *mean* of the response variable based on given values of the independent variables, whereas quantile regression extends the regression model to the conditional quantiles of the response variable. We focus on finding $50^{th}$, $75^{th}$ or $95^{th}$ percentile of the conditional distribution of $Y$, where each random variable $Y_i$ comes from

$$Y_i = f(x_i) + \varepsilon_i \tag{3}$$

where $\varepsilon_i$ is a random error. It is also important to note that quantile regression is much more flexible than least squares regression when dealing with heterogeneous conditional distributions, because it makes no distributional assumption about the error term in model (3) and simply provides a conditional distribution of the prediction given the predictor values [35].

Let us denote $\tau$ an arbitrary quantile $(0 < \tau < 1)$. Linear quantile regression assumes that the $\tau$th conditional quantile of $Y$, denoted below by $Q_\tau(Y|x)$, is a linear function:

$$Q_\tau(Y|x) = x^T \theta_\tau.$$

Having the observations $(x_i, y_i)$ $(i = 1, \ldots, n)$, we can estimate $\theta_\tau$ by $\hat{\theta}_\tau$ by solving the following optimization problem:

$$\hat{\theta}_\tau = \arg\min_{\theta_\tau} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^T \theta_\tau),$$

where

$$\rho_\tau(u) = (\tau - I(u < 0))u,$$

and $I(\cdot)$ is the indicator function. If the error terms are *iid*, then $\hat{\theta}_\tau$ is asymptotically normal with mean $\theta_\tau$ and its asymptotic covariance is given in [35].

### C. Bonferroni Inequality for Joint Confidence Statements

The Bonferroni inequality for two events $A$ and $B$ is:

$$P(A \cup B) \leq P(A) + P(B)$$

This inequality can be rewritten as:

$$1 - P(A \cup B) \geq 1 - (P(A) + P(B)). \tag{4}$$

Let $I_{X_1}^{1-\alpha}$ and $I_{X_2}^{1-\alpha}$ be respectively two $(1-\alpha)$-confidence intervals for two arbitrary random variables $X_1$ and $X_2$. Let $A$ be the event $X_1 \notin I_{X_1}^{1-\alpha}$, and $B$ the event $X_2 \notin I_{X_2}^{1-\alpha}$. $P(A)$ and $P(B)$ respectively denote the probability of type I error for confidence intervals $I_{X_1}^{1-\alpha}$ and $I_{X_2}^{1-\alpha}$. The rewritten Bonferroni inequality (4) now becomes:

$$1 - P\left((X_1 \notin I_{X_1}^{1-\alpha}) \cup (X_2 \notin I_{X_2}^{1-\alpha})\right) \geq 1 - 2\alpha$$

This allows us to obtain the following joint confidence statement:

$$P\left((X_1 \in I_{X_1}^{1-\alpha}) \cap (X_2 \in I_{X_2}^{1-\alpha})\right) \geq 1 - 2\alpha$$

Therefore, we can always use the Bonferroni inequality on different $(1-\frac{\alpha}{n})$-confidence intervals to obtain one confidence statement over $n$ confidence intervals which has a probability greater than or equal to $1 - \alpha$.

### D. Cross-validation

Cross-validation [36] is a common schema used to assess the generalization performance of a model[3], i.e. its ability to generalize correctly on fresh inputs. Let us denote $\hat{f}_{\mathcal{S}}^{\mathcal{M}}$ the estimated prediction model found by a method $\mathcal{M}$ on a random sample $\mathcal{S}$. In a $k$-fold cross-validation schema, the random sample $\mathcal{S}$ is partitioned into $k$ folds $(\mathcal{S}_i)_{1\leqslant i\leqslant k}$, where we have $\mathcal{S}_{-i} = \mathcal{S}\backslash\mathcal{S}_i$. Here $\mathcal{S}_{-i}$ and $\mathcal{S}_i$ are respectively the training set and the validation set in the $i^{th}$ fold.

In this schema, a different prediction model $\hat{f}_{\mathcal{S}_{-i}}^{\mathcal{M}}$ is estimated on each random sample $\mathcal{S}_{-i}$ (training set in the $i^{th}$ fold). Then the prediction error of the trained model $\hat{f}_{\mathcal{S}_{-i}}^{\mathcal{M}}$ is measured on the validation set $\mathcal{S}_i$. At the end, the overall model performance measure is estimated from the prediction errors over the $k$ disjoint validation sets.

## IV. HIGH CONFIDENCE INTERVAL PREDICTION

Regression models are always built with finite sample size, thus the predicted mean or quantile is an estimate of the true unknown conditional mean or quantile of the random variable $Y(x) = f(x) + \varepsilon$. Therefore, while dealing with datasets of finite size, we need to estimate a confidence interval. In this paper, we are interested in finding intervals in regression models which contain, with a specified confidence level $\gamma$, at least a desired proportion $\beta$ of the conditional response variable.

Commonly used prediction intervals (for example, linear prediction intervals [37]) do not consider all aspects of uncertainties caused by the limited number of observations. Krishnamoorthy and Mathew [38], Hahn and Meeker [39] and Kocherginsky et al. [40] provided an in-depth justification of why tolerance intervals and confidence intervals on regression quantiles are appropriate methods for the problem of high confidence interval prediction. For an recent work on high confidence intervals in non-linear models, see [41]. The rest of this section describes tolerance intervals for least-squares regression and confidence intervals for regression quantiles, which are two classical approaches for obtaining such intervals. We introduce a new method – denoted LR Tolerance in our results – for obtaining tolerance intervals in Section IV-B, in the context of least squares regression. We also propose TBQCI (Two-sided Bonferroni-Quantile Confidence Intervals) in Section IV-E, as a new method for building two-sided high-confidence intervals in quantile regression.

### A. Tolerance Intervals for Least-Squares Regression

Tolerance intervals for Ordinary Least Squares (OLS) have been introduced by Wallis [42] in order to obtain intervals that are guaranteed, with a confidence level $\gamma$, to contain at least a proportion $\beta$ of the conditional distribution of the response variable $Y(x)$. We refer to such intervals as $\beta$-content $\gamma$-coverage regression tolerance intervals and they are denoted by

---

[3]In this work the performance is assessed through the Mean Interval Size (MIS) and the Mean Inclusion Percentage (MIP), which are described later in Section VII.

---

$I_{\gamma,\beta}^T(x)$. A two-sided tolerance interval $I_{\gamma,\beta}^T(x)$ for $Y(x)$ takes the form $\hat{f}(x) \pm \rho(x)\hat{\sigma}$, where $\rho(x)$ is a tolerance factor to be determined subject to the content $\beta$ and the desired confidence level $\gamma$. It is stated formally as:

$$P_{\hat{f},\hat{\sigma}}\left( P_{Y(x)}\left( Y(x) \in I_{\gamma,\beta}^T(x) \right) \geq \beta \right) = \gamma, \qquad (5)$$

where

$$I_{\gamma,\beta}^T(x) = [L_{\beta,\gamma}^T(x), U_{\beta,\gamma}^T(x)] = [\hat{f}(x) - \rho(x)\hat{\sigma}, \hat{f}(x) + \rho(x)\hat{\sigma}].$$

Hahn and Meeker [39], and Krishnamoorthy and Mathew [38] provided a comprehensive comparison between tolerance intervals and other statistical intervals.

### B. Bonferroni Tolerance Intervals for Least-Squares Regression

In this part, we introduce Bonferroni regression tolerance intervals for least-squares regression and we apply them to OLS. Compared to the state-of-the-art on tolerance intervals for OLS (see [38] for an in-depth reference on the subject), this method has three important advantages: *it is reliable, fast and easy to implement*. Lieberman and Miller [43] used the Bonferroni inequality to combine two confidence statements and obtain *simultaneous tolerance intervals for linear regression*. In this work, we apply the Bonferroni inequality to obtain *tolerance intervals for linear regression*. It first consists in finding a $(1 - \frac{\tau}{2})$ pointwise confidence interval for the conditional mean $f(x) = x^T\theta$ described by Equation (6) [37]. Then we obtain a $(1 - \frac{\tau}{2})$-level upper confidence bound on the error's standard deviation. In the case of OLS, it is obtained by (7) where $N$, $k$ and $\hat{\theta}$ denote respectively the number of observations, the number of predictors and the estimator of $\theta$ in the OLS model [43], [37].

$$P_{\hat{\theta}}\left( x^T\theta \in I_{1-\frac{\tau}{2}}^{pw}(x) \right) = 1 - \frac{\tau}{2}, \qquad (6)$$

$$P_{\hat{\sigma}}(\sigma \leq \mathbf{c}\hat{\sigma}) = 1 - \frac{\tau}{2}, \qquad (7)$$

where

$$L_{1-\frac{\tau}{2}}^{pw}(x) = x^T\hat{\theta} - t_{(\frac{\tau}{4}, N-k-1)}\hat{\sigma},$$

$$U_{1-\frac{\tau}{2}}^{pw}(x) = x^T\hat{\theta} + t_{(\frac{\tau}{4}, N-k-1)}\hat{\sigma},$$

$$I_{1-\frac{\tau}{2}}^{pw}(x) = [L_{1-\frac{\tau}{2}}^{pw}(x), U_{1-\frac{\tau}{2}}^{pw}(x)],$$

$$\mathbf{c} = \left( \frac{N-k-1}{\chi^2_{\frac{\tau}{2}, N-k-1}} \right)^{\frac{1}{2}}.$$

Now by injecting Equations (6) and (7) in the Bonferroni inequality, we obtain a $\gamma$-coverage $\beta$-content tolerance interval which is formulated as follows:

$$P_{\hat{\theta},\hat{\sigma}}\left( P_{Y(x)}\left( L_{1-\frac{\tau}{2}}^{pw}(x) + Z_{\frac{1-\beta}{2}}\mathbf{c}\hat{\sigma} \leq Y(x) \leq \right. \right. \qquad (8)$$

$$\left. \left. U_{1-\frac{\tau}{2}}^{pw}(x) + Z_{\frac{1+\beta}{2}}\mathbf{c}\hat{\sigma} \right) \geq \beta \right) \geq \gamma,$$

where $\gamma = 1 - \tau$. As one can notice, Equation (8) could be obtained by tables of chi-squares, Student, and normal distributions, which reduces the computational complexity

of each interval computation to $O(1)$ and is also easy and straightforward to implement. We exploit the Bonferroni inequality to combine two different confidence intervals and obtain tolerance intervals for least-squares. The application of the Bonferroni inequality does not depend on the estimated function and can be applied to linear as well as non-linear or non-parametric regression models with normally distributed errors.

### C. Interval Prediction with Quantile Regression Models

While least-squares models estimate the conditional mean function $f(\cdot)$, a quantile regression obtains estimates of conditional quantiles which, on average, are equal to the true quantile function as $n$ goes to infinity (consistent estimator). A quantile regression model gives one-sided intervals:

$$IQ_{1-\alpha}(x) = (-\infty, Q_{1-\alpha}(x)], \qquad (9)$$

where $Q_{1-\alpha}(x)$ estimates the true $(1-\alpha)$-quantile of $Y(x)$, given a particular combination of predictors. Note that, as we saw in Section III-B, $Q_{1-\alpha}(x)$ is an estimator of the unknown conditional quantile function. This means that interval $IQ_{1-\alpha}(x)$ will, on average, contain a proportion $1-\alpha$ of $Y(x)$.

### D. Confidence Interval on Regression Quantiles

Once we have estimated our conditional quantile with a quantile regression model, we need to obtain confidence intervals on the conditional quantile. Kocherginsky et al. [40] compared several methods for obtaining confidence intervals on regression quantiles and they proposed different approaches depending on the size of the dataset and the number of variables. In this work, we used the method denoted MCMB-A[4] introduced by Kocherginsky et al. [40]. This is a time-saving re-sampling method based on a modification of the Markov Chain Marginal Bootstrap (MCMB) which has a computational complexity of $O(np)$, where $p$ is the number of predictors (explanatory variables) and $n$ the number of examples. However these intervals are one-sided intervals. In the following subsection, we introduce TBQCI which obtains two-sided intervals having similar properties to tolerance intervals.

### E. Two-sided Bonferroni-Quantile Confidence Intervals (TBQCI)

In order to obtain two-sided $(1-\alpha)$-content conditional intervals ($\beta = 1 - \alpha$), one must build two distinct quantile regression models: a lower $\frac{\alpha}{2}$-quantile regression model and an upper $(1-\frac{\alpha}{2})$-quantile regression model. However each model needs itself a confidence interval as explained in IV-D.

Suppose that we have built the upper and lower quantile regression models and let $\gamma = 1 - \tau$ be the confidence level of the desired two-sided interval (which contains at least a proportion $1 - \alpha$ of $Y(x)$). We begin by obtaining a lower (one-sided) $(1-\frac{\tau}{2})$-confidence interval on the lower $\frac{\alpha}{2}$-quantile regression model and an upper (one-sided) $(1-\frac{\tau}{2})$-confidence

---

[4]This is the only implemented method and it is available in the `quantReg` package of `R`.

interval on the upper $(1-\frac{\alpha}{2})$-quantile regression model. These lower and upper $(1-\frac{\tau}{2})$-confidence intervals are respectively denoted by $IL_{\frac{\alpha}{2}}^{1-\frac{\tau}{2}}(x)$ and $IU_{1-\frac{\alpha}{2}}^{1-\frac{\tau}{2}}(x)$ in Equations (10) and (11).

$$P_{\mathcal{S}}\left( P_{Y(x)}\left( Y(x) \in IL_{\frac{\alpha}{2}}^{1-\frac{\tau}{2}}(x)|\mathcal{S}\right) \leq \frac{\alpha}{2}\right) = 1 - \frac{\tau}{2}, \quad (10)$$

where

$$IL_{\frac{\alpha}{2}}^{1-\frac{\tau}{2}}(x) = (-\infty, L_{\frac{\alpha}{2}}^{1-\frac{\tau}{2}}(x)].$$

$$P_{\mathcal{S}}\left( P_{Y(x)}\left( Y(x) \in IU_{1-\frac{\alpha}{2}}^{1-\frac{\tau}{2}}(x)|\mathcal{S}\right) \geq 1-\frac{\alpha}{2}\right) = 1-\frac{\tau}{2}, \quad (11)$$

where

$$IU_{1-\frac{\alpha}{2}}^{1-\frac{\tau}{2}}(x) = (-\infty, U_{1-\frac{\alpha}{2}}^{1-\frac{\tau}{2}}(x)].$$

In Equation (10), $L_{\frac{\alpha}{2}}^{1-\frac{\tau}{2}}(x)$ denotes a lower confidence bound on the $\frac{\alpha}{2}$-regression quantile at point $x$. Its corresponding confidence interval contains, with a confidence $1 - \frac{\tau}{2}$, at least the $\frac{\alpha}{2}$ quantile of $Y(x)$. In Equation (11), $U_{1-\frac{\alpha}{2}}^{1-\frac{\tau}{2}}(x)$ denotes an upper confidence bound on the regression quantile at point $x$ and $IU_{1-\frac{\alpha}{2}}^{1-\frac{\tau}{2}}(x)$ contains, with confidence $1 - \frac{\tau}{2}$, at least the $1 - \frac{\alpha}{2}$ quantile of $Y(x)$.

By applying the Bonferroni inequality, we merge two probability statements of $(1-\frac{\tau}{2})$ confidence and we obtain a joint confidence statement with a probability greater than or equal to $1 - \tau$:

$$P_{\mathcal{S}}\left( P_{Y(x)}\left( Y(x) \in IQ_{1-\alpha}^{1-\tau}(x)\right) \geq 1 - \alpha\right) \geq \gamma, \quad (12)$$

where

$$IQ_{1-\alpha}^{1-\tau}(x) = [L_{\frac{\alpha}{2}}^{1-\frac{\tau}{2}}(x), U_{1-\frac{\alpha}{2}}^{1-\frac{\tau}{2}}(x)], \gamma = 1 - \tau.$$

To the authors' knowledge, this method addressing the problem of obtaining two-sided high confidence interval, using one-sided confidence intervals on regression quantiles, is new to the literature. We name such intervals Two-sided Bonferroni Quantile Confidence Intervals (TBQCI).

Equation (12) states that TBQCI provides two-sided intervals that contain, with a specified confidence $\gamma$, at least a desired proportion $\beta$ of the distribution of $Y(x)$. Such intervals are similar to two-sided $\gamma$-coverage $(1-\alpha)$-content tolerance intervals defined by Equation (5). While TBQCI is for quantile regression, tolerance intervals are for least-squares regression.

## V. TRAJECTORY PREDICTION USING A POINT-MASS MODEL

Before showing the results of our regression methods on the task of predicting aircraft altitude intervals, let us describe the point-mass model that serves as baseline method in our experiments. This section describes the trajectory prediction with point-mass models, and a state-of-the-art variant where the aircraft mass is estimated from past trajectory points. We also describe how $\beta$-content altitude intervals can be built when using such models, and explain the aircraft operation during climb.
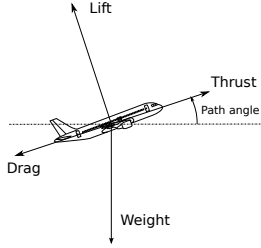
Figure 3. Simplified point-mass model.

### A. The Point-Mass Model

The point-mass model, illustrated in Figure 3, describes the forces applying to the center of gravity of the aircraft and their influence on the aircraft's acceleration, making several simplifying assumptions. It is assumed that the thrust and drag vectors are collinear to the airspeed vector $\overrightarrow{V_a}$, and that the lift is perpendicular to these vectors. We denote $Thr$ the total thrust, $D$ the aerodynamic drag of the airframe, $m$ the aircraft mass, $g$ the gravitational acceleration and $W_{UP}$ the vertical component of the wind. Introducing the geodetic altitude $h$, projecting Newton's second law along the air speed $\overrightarrow{V_a}$ gives us the following equation :

$$\frac{Thr - D}{m}V_a = V_a\dot{V_a} + g\dot{h} + \overrightarrow{W} \cdot \overrightarrow{V_a} - gW_{UP} \qquad (13)$$

Several equivalent forms of this equation can be used (see [44]), depending on what unknown variable is being calculated from the other known variables. Actually using Equation (13) to predict a trajectory requires a model of the aerodynamic drag for any airframe flying at a given speed through the air. In addition, we may need the maximum climb thrust, which depends on the engines with which the aircraft is equipped. In the experiments presented here, the Eurocontrol BADA model was used for that purpose.

However, one cannot use Equation (13) without prior knowledge of the initial state (mass, position, speed, *etc.*) of the aircraft, and of the pilot's intents as to how the aircraft will be operated in the future (thrust law, speed law, or rate of climb). The position and speed are known using radar measurements, but none of the other aforementioned parameters are available to ground-based systems.

### B. Estimating the mass

Two mass estimation methods ([8], [9], [45]) have been recently derived from Equation 13. In order to use these methods, one has to assume a thrust setting. In this work, we assume a max climb thrust. We have tested the interval prediction using the BADA point-mass model, either with the default mass given in the BADA tables, or with the mass estimated by the adaptive method ([8], [9]), which is simple to implement.

### C. Interval Prediction for BADA Models

In Section IV, we have seen how to deal with the prediction uncertainty in regression models. In the case of point-mass

models (BADA and BADA Adaptive), the uncertainty on the altitude prediction is assessed as follows:

- First, we assume that the point-mass prediction error has a zero mean, homoscedastic (constant variance) normal distribution.
- Second, we estimate $\sigma^2_{train}$, the variance of the 10-minute-ahead pointwise prediction error on the training set.
- Finally we take the $\beta$-inter-quantile of a normal distribution with mean and variance respectively equal to $f^\Phi(x)$, the point-mass prediction, and $\sigma^2_{train}$ the standard deviation of the point-mass model errors. The interval, denoted $I^\Phi_\beta(x)$, is then obtained as follows:

$$I^\Phi_\beta(x) = f^\Phi(x) \pm \sigma_{train}Z_{\frac{1-\beta}{2}} \qquad (14)$$

The term $\Phi$ stated in the above expressions can be replaced by **BADA** or **BADA Adaptive**, depending on the chosen model (BADA with default parameters, or BADA with the mass estimated by the adaptive method).

### D. Aircraft Operation During Climb

Generally, when no external constraint applies during the climb, the aircraft is operated at constant CAS[5] and variable Mach number, until a specified Mach number is reached. Above this CAS/Mach crossover altitude, the aircraft is operated at a constant Mach number, and variable CAS. External constraints may apply, however. After take-off, the aircraft cannot exceed a specified maximum CAS until Flight Level 100[6] is reached. This first climb segment is followed by an acceleration at FL100, and then a second climb segment at a higher calibrated airspeed, until the CAS/Mach crossover altitude is reached.

In this paper, we consider only this second climb segment at constant CAS, followed by the constant Mach climb, as we are mostly interested in predicting the aircraft trajectory in the en-route airspace. The actual CAS and Mach values are chosen by the airline operators, according to a cost index specific to each airline. The cost index and the chosen CAS and Mach values are not known by air traffic control systems, although some studies show the improvements that such knowledge would provide in trajectory prediction ([46], [47]). In this work, the trajectories are predicted using the standard CAS and Mach values of the BADA climb procedures file.

## VI. DATASET AND PATTERNS FOR TRAINING

This section describes how the raw data has been processed in order to obtain the climbing trajectories and the associated patterns for training. We focused on one aircraft type and we selected the Airbus A320 because it is the most common aircraft in Europe and in our trajectory set.

---

[5]CAS: the Calibrated Air Speed is the speed indicated on the pilot's instruments, disregarding the intrument errors.

[6]FL100 = 10000 feet above isobar 1013.25 hPa.

## A. The Recorded Data

The trajectories used in this study were obtained from recorded radar tracks from the Paris Air Traffic Control Center. We selected flights over two months (July 2006 and January 2007) departing from Paris Orly or Paris Charles-de-Gaulle. The raw data is made of one position report every 1 to 3 seconds. The raw Mode-C altitude[7] has a granularity of 100 feet.

The weather data grid is provided by Meteo France. The wind and temperature are available at various isobar altitudes over the same two months. The wind and temperature at every trajectory point were interpolated from the weather data grid.

## B. Data Preprocessing

The recorded aircraft trajectories were smoothed using cubic splines, in order to obtain: the aircraft position ($X$, $Y$ in a projection plan, or latitude and longitude in WGS84), the ground velocity vector ($V_x$, $V_y$), the smoothed pressure altitude $H_p$, and the rate of climb or descent (ROCD).

The climb, cruise and descent segments were identified, using triggers on the rate of climb or descent to detect the transitions between two segments. The trajectories were then filtered so as to keep only the climb segments. An additional 40 seconds were trimmed from the beginning and end of each segment, so as to remove climb/cruise or cruise/climb transitions. The trajectories were then sampled every 15 seconds, with time and distance origins at the point $P_0$ where the climb segment crosses flight level FL180[8]. The trajectory segments were sampled so as to obtain 10 points preceding $P_0$ and 40 points after $P_0$ because of the 10 minutes look-ahead time. In the end, we obtained 4939 sampled trajectories.

Using the position, velocity and wind data, we computed the true air speed, the distance flown in the air, the drift angle and the along-track and cross-track winds. The successive velocity vectors allowed us to compute the trajectory curvature at each point. The actual aircraft bank angle was then derived from true airspeed and the curvature of the air trajectory. Finally, the BADA model equations were used to compute additional data, such as: calibrated airspeed, Mach number, energy share factor[9], as well as the derivatives of these quantities with respect to time.

## C. Building Patterns for Regression

Regression models are built using a set of patterns $(x, y)$. From each sampled climbing segment, one pattern $(x, y)$ is extracted. In our application, $y$ is the pressure altitude $H_p$ at 10 minutes after $P_0$ and $x$ is built using the variables known at $t_0$. This vector includes state variables measured at $t_0$ and before. It also contains variables from the flight plan like the Requested Flight Level, the requested speed and the distance

---

[7]This altitude is directly derived from the air pressure measured by the aircraft. It is the height in feet above isobar 1013.25 hPa.

[8]FL180= 18000 feet above isobar 1013.25 hPa.

[9]The energy share factor measures how much of the energy is devoted to climb or to longitudinal acceleration. Using the observed speed, acceleration and rate of climb, it can be computed using the following formula: $\frac{g\dot{h}}{V_a\dot{V}_a + g\dot{h}}$.

between the departure and arrival airports. In addition, the temperatures and winds given by the weather grid at different altitudes and the mass estimated with the adaptive method are used. In the end, there are 338 explanatory variables and one response variable. This set contains 4939 patterns.

A principal component analysis was performed on the explanatory variables $x$, so as to reduce the dimensionality and avoid redundant input variables in the trajectory prediction. Principal components are linear combinations of the initial variables that we can use instead of the initial explanatory variables in the regression method. Once these principal components are computed, only the $p$ principal components with the largest variances are used instead of the initial 338 explanatory variables. In this case, we reduced the dimensionality of the dataset from 338 explanatory variables to two different datasets containing 35 and 65 significant Principal Components (PC) which respectively explain 95% and 99% of the variance of the feature space. Figure 4 shows the standard deviations of the first 80 principal components: the first 35 and 65 principal components are respectively located above the solid orange line and the dotted blue line and they explain respectively 95% and 99% of the variance of the 338 explanatory variables.
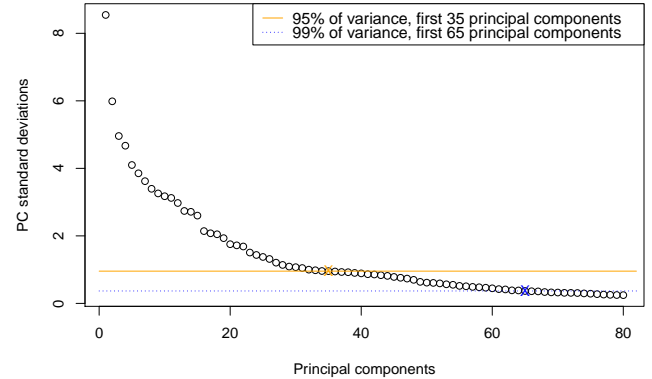


Figure 4. Principal components standard deviations for the 338 explanatory variables of the ATM described in Section VI. The first 35 and 65 principal components are respectively located above the solid orange line and the dotted blue line and they explain respectively 95% and 99% of the variance of the 338 explanatory variables.

## VII. EXPERIMENTS

Let us now compare the methods introduced in Sections IV-B and IV-E with the two flavors of point-mass model described in Section V, using the dataset described in the previous section. Here, we compare the intervals obtained by the mentioned methods regardless of any variable selection or outliers detection preprocessing.

## A. Experimental setup

Our interval prediction models are scored using 100 runs of a 10-fold cross-validation on the dataset described in Section VI. With these 100 runs, we obtain 1,000 models. Each model is tested on a validation set. From one validation set we can compute the average fraction of actual altitudes

falling inside their predicted interval and in what follows, this value is called the Mean Inclusion Percentage (MIP).

The MIP is an estimation of $P(Y(x) \in I(x))$ because it is found by the average fraction of actual altitudes falling inside their predicted interval. This value can change with different aircraft trajectory data and different sample size; it is a random variable.

As explained before, our context requires a reliable prediction of the future aircraft positions, so we are interested in interval prediction models being very likely to give a model with their $P(Y(x) \in I(x))$ greater than or equal to the desired content $\beta$. For this purpose, we will use the probability $P(P(Y(x) \in I(x)) \geq \beta)$ which can be estimated by computing the average fraction of MIP greater or equal to $\beta$. This is done using the 1,000 estimates of MIP associated to the 1,000 validation sets. It is expected to be greater than or equal to the desired coverage $\gamma$.

Please note the difference between choosing a model which has on average $P(Y(x) \in I(x)) = \beta$ and ensuring with a high probability $\gamma$ that $P(Y(x) \in I(x)) \geq \beta$.

For example, for $\beta = 0.9$ and $\gamma = 0.95$, we expect that at least 950 models among the 1,000 tuned models will have a MIP greater than or equal to 0.9. Of course, this objective is easier to achieve for very wide intervals, which are not desirable: our aim is to find the tightest interval that contains the desired proportion of altitude predictions. Consequently, the model quality is also assessed by looking at the size of the predicted interval.

The model performance criteria are the MIP and the Mean Interval Size (MIS). As for the MIP, one MIS value is computed from one validation set. The MIS is the mean size of $I(x)$ for $x$ in the validation set. In the following, when assessing the performance of the proposed methods, the term *reliability* refers to the fact that the observed content MIP is greater than or equal to the desired content $\beta$, whereas the term *precision* refers to the MIS: the tighter the interval, the more precise the method.

In our experiments, we take $\gamma = 0.95$. This choice is arbitrary and corresponds to the level of confidence that we require on the altitude interval bounds. This means that we choose a $0.05$ probability[10] of having a model which does not provide intervals that contain at least a desired proportion of the conditional distribution of the response variable (here, the altitude). From an operational point of view, we assume that such relatively rare occurences can be managed by the air traffic control system, for example through the use of existing *safety net* tools and short-term conflict alert and collision avoidance tools.

In the rest of this section, we compare the following prediction methods, using the MIP and MIS criteria:

- **BADA**: BADA point-mass model, with the BADA default values for the aircraft mass, target speeds (CAS/Mach), and assuming a maximum climb thrust. This model takes the temperature and the effect of the wind into account. In order to predict intervals one has to replace $f^\Phi(x)$

[10]This is a common choice in statistical inference.

| Method | MIP | MIS [ft] | $P(\text{MIP} \geq \beta)$ |
|---|---|---|---|
| BADA | 0.947 (0.0099) | 6237 (17) | 0.359 |
| BADA Adaptative | 0.948 (0.01) | 4057 (16) | 0.42 |
| 0.95-LR Tolerance p=65 | 0.967 (0.0078) | 4283 (27) | 0.982 |
| 0.95-LR Tolerance p=35 | 0.967 (0.0079) | 4453 (21) | 0.98 |
| 0.95-TBQCI p=65 | 0.97 (0.0078) | 4454 (97) | 0.989 |
| 0.95-TBQCI p=35 | 0.966 (0.0083) | 4451 (52) | 0.965 |

Table I
MIP AND MIS FOR DIFFERENT INTERVAL PREDICTION MODELS WITH A THEORETICAL $\beta$-CONTENT OF 0.95 AND A 10 MINUTES PREDICTION HORIZON. THE STANDARD DEVIATION OVER THE 1,000 VALUES IS SHOWN IN PARENTHESES.

and $\sigma_{train}$ in Equation (14) by the corresponding values obtained by this method.

- **BADA Adaptive**: Same as **BADA**, except that the mass is estimated using the adaptive method, introduced in [8], on eleven past trajectory points.
- **0.95-TBQCI p=65**: two-sided Bonferroni 0.95-level confidence $\beta$-content interval obtained by two different quantile regression models, as explained in Section IV-E, built on the dataset with the first 65 principal components. We used the $rq$ and $rq.predict$ function in R's `quantReg` package in order to obtain the intervals $IL_{\alpha/2}^{1-\tau/2}(x)$ and $IU_{1-\alpha/2}^{1-\tau/2}(x)$ mentioned in Equations 10 and 11.
- **0.95-TBQCI p=35**: Same as 0.95-TBQCI p=65 but built on the dataset with the first 35 principal components.
- **0.95-LR Tolerance p=65**: two-sided 0.95-coverage $\beta$-content Bonferroni least-squares regression tolerance intervals (explained in IV-B) built on the dataset with the first 65 principal components.
- **0.95-LR Tolerance p=35**: Same as 0.95-LR Tolerance p=65 but built on the dataset with the first 35 principal components.

*B. Results for $\beta = 0.95$*

Let us first show a few results for a desired content $\beta = 0.95$. This $\beta$-content should not be confused with the specified confidence level $\gamma$ used in our methods, which is also set to 0.95.

The interval prediction results for the tested methods are shown in Table I. The second column shows the mean and standard deviation of the 1,000 MIPs associated to the 1,000 validation sets. This column is to be compared with the desired content $\beta$. The third column shows the mean and standard deviation of the 1,000 values of MIS associated to the 1,000 validation sets. This column is related to the *precision* of the method. The last column shows the average fraction of MIP values greater than or equal to $\beta$. This column is to be compared with the desired coverage $\gamma$.

The tested BADA reference model has the smallest mean MIP and gives much wider altitude intervals than all other methods. This poor performance can certainly be explained by the choice of default parameters for the inputs of the physical model described in Section V-A. In the BADA method, we

| Method | CPU time in seconds |
|---|---|
| 0.95-LR Tolerance p=65 | 0.12 |
| 0.95-LR Tolerance p=35 | 0.07 |
| | |
| 0.95-TBQCI p=65 | 761 |
| 0.95-TBQCI p=35 | 230 |

Table II
CPU TIME OF LR TOLERANCE AND TBQCI MODELS FOR THE
0.95-CONTENT 10-MINUTE AHEAD INTERVAL PREDICTION TASK ON THE
ATM DATASET DESCRIBED IN SECTION VI.

have the same mass, speed intent, and thrust law for all aircraft of the same type.

The BADA Adaptive method works very well, the obtained MIP is almost the same as the desired one and it gives the tightest interval: estimating the mass highly improves the precision of the interval prediction. However, the probability that the built model contains the $\beta$ content, estimated by $P(\text{MIP} \geq \beta)$, is low compared to the $\gamma = 0.95$ coverage desired.

Unlike the BADA methods, all versions of our methods (0.95-LR Tolerance and 0.95-TBQCI) have a $P(\text{MIP} \geq \beta)$ greater than the desired coverage $\gamma$. But this higher reliability is obtained at the expense of slightly larger interval sizes, when compared with the BADA Adaptive MIS.

In Table II, we compare the CPU times of our methods, on the ATM dataset. Note that we measure the time required to compute prediction only, and not the time required to build the models. We see that the CPU times are in line with the computational complexity of the LR Tolerance and TBQCI models as seen in Section IV.

One can observe that the TBQCI model with 65 principal components is much more time consuming than the one with 35 principal components, and that both methods are much slower than the least squares tolerance methods. This is due to the computational complexity $O(np)$ of the approach adopted by TBQCI as explained in Section IV-D.

However, these CPU times are for the computation of 4939 interval predictions. The average CPU times per trajectory for the two TBQCI methods (with 35 or 65 principal components) are 0.05 and 0.15 seconds respectively. This is acceptable for an operational use if we compute only a few trajectories per aircraft. This might be too long for the CD&R application evoked in Section II. There are still some computational issues to be addressed for the TBQCI method. For want of anything better, we could parallelize the computation of the trajectories.

*C. Results for different values of $\beta$*

Figure 5 plots the MIP as a function of the desired content $\beta$. For each $\beta$ and each method, we have 1,000 values of MIP. The distribution of these 1,000 MIP values are visualized using a boxplot. The boxplot shows the 5 %, 25 %, 50 %, 75 % and 95 % quantiles. Thus, if one boxplot is over the black line $y = x$, then 95 % of the MIPs are greater than or equal to the desired content $\beta$. If this is the case we can conclude that this method have the desired $\gamma$ coverage. With Figure 5, we can see that the BADA methods have a median MIP close to the

desired $\beta$. Only the methods introduced in this paper satisfy the $\gamma$ coverage constraint.

Figure 6 plots $P(\text{MIP} \geq \beta)$ as a function of $\beta$. With this figure, we can conclude that our methods have a coverage greater than $\gamma = 0.95$ for all values of $\beta$ less than 0.95. For $\beta = 0.97$ and $\beta = 0.99$, the coverage is much less than expected for the 0.95-LR Tolerance p=65 method. Concerning the 0.95-TBQCI p=35 method, it remains high but slightly inferior to the desired coverage.

Indeed, obtaining statistical intervals in multiple regression having their true coverage equal to their nominal coverage is a hard task and this difficulty grows when increasing $\beta$ and/or the number of independent variables [38], [39]. Moreover, we know that the true data generation function is not an OLS model. The OLS (the 0.95-LR Tolerance p=65 method) is the statistical least-squares regression model which better matches our data. It is not the true function. The robustness of the 0.95-TBQCI p=35 method is simply explained by the robustness of quantile regression. As explained before, quantile regression has fewer model assumptions and is more robust to outliers than OLS.

Figure 7 plots the MIS showing the precision of the interval prediction methods for different values of the desired content $\beta$. Considering Figures 5 and 7 simultaneously, we see that the BADA Adaptive model is more precise, in terms of interval size, than our methods based on the Bonferroni inequality, but less reliable in terms of $\gamma$ coverage. All these methods perform much better than the baseline BADA model that is currently used in many trajectory predictors.
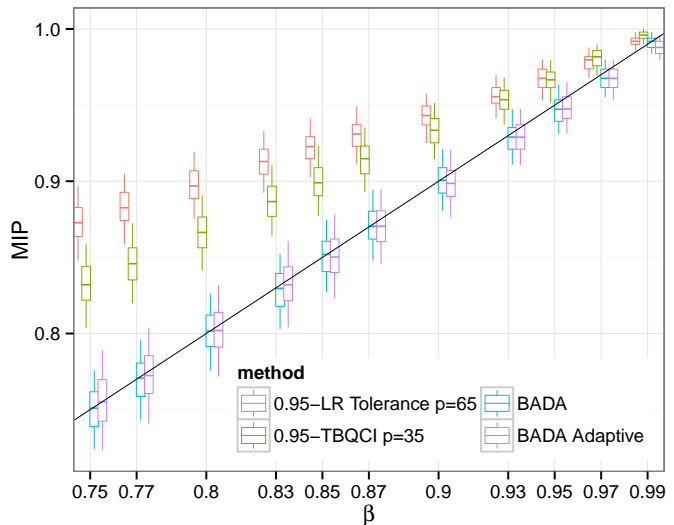


Figure 5. Mean Inclusion Percentage (MIP) for different values of $\beta$, and for the models described in Section VII-A. For each $\beta$ and each method, we have 1,000 values of MIP. The distribution of these 1,000 MIP values are visualized using a boxplot. The boxplot shows the 5 %, 25 %, 50 %, 75 % and 95 % quantiles.

VIII. CONCLUSION

This paper dealt with the application of high-confidence interval prediction techniques to the altitude prediction of
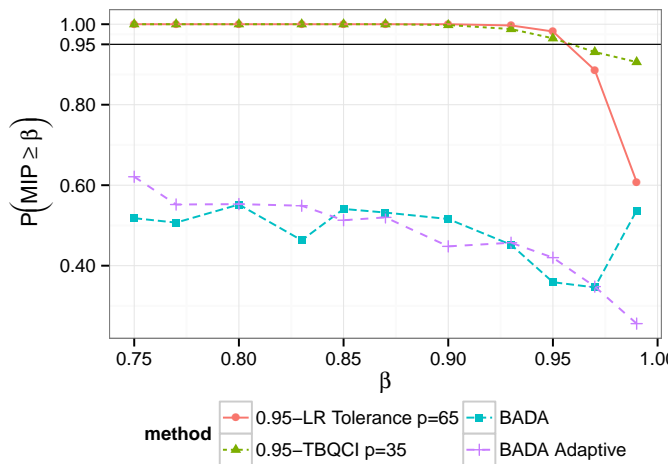
Figure 6. For each $\beta$ and each method, we have 1,000 values of MIP. This figure plots the proportion of MIP greater or equal to $\beta$ as a function of $\beta$.
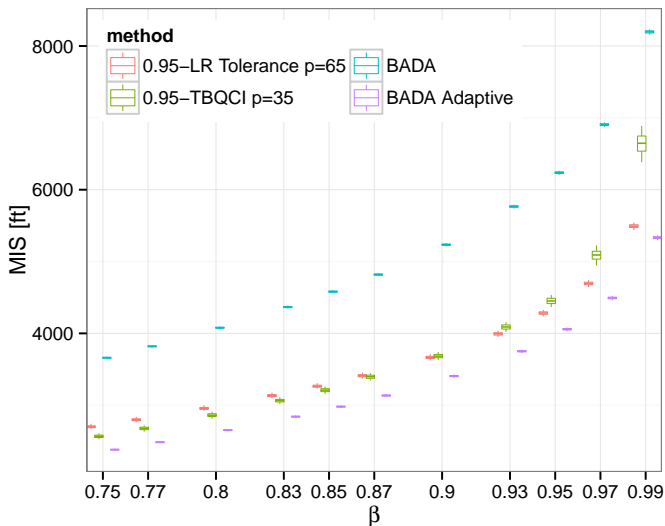


Figure 7. Mean Interval Size (MIS) for different values of $\beta$, and for the models described in Section VII-A. For each $\beta$ and each method, we have 1,000 values of MIS. The distribution of these 1,000 MIS values are visualized using a boxplot. The boxplot shows the 5 %, 25 %, 50 %, 75 % and 95 % quantiles.

climbing aircraft. We were interested in finding two-sided prediction intervals in regression models which contain, with a specified confidence level $\gamma$, at least a desired proportion $\beta$ of the conditional distribution of the predicted altitude. This paper introduced two new methods based on the Bonferroni inequality to compute such intervals. The first one (LR Tolerance) computes tolerance intervals for least squares regression. The second one – denoted Two-sided Bonferroni-Quantile Confidence Intervals (TBQCI) – obtains high confidence two-sided intervals in quantile regression.

The proposed methods have been compared with two state-of-the-art point-mass models, on the task of predicting the altitude of climbing aircraft 10 minutes ahead, starting from an initial point at flight level FL180. The results show that the proposed methods perform much better than the baseline BADA model that uses a default value for the aircraft mass. When comparing with the BADA Adaptive model where the aircraft mass is estimated using past trajectory points, our methods prove more reliable (in terms of confidence level $\gamma$), although less precise (in terms of interval size), than BADA Adaptive.

This was to be expected, as our methods explicitly build intervals so as to guarantee the desired $\beta$-content, with a probability greater than a specified confidence level $\gamma$. This is not the case for the interval prediction methods based on the point-mass models.

From an operational point of view, the proposed methods could be applied to the detection and resolution of potential conflicts between trajectories. In such a context, one might want to find the proper trade-off between reliability and precision. This trade-off may differ with the condidered look-ahead time: predictions at short look-ahead times should definitely be as reliable as possible: not detecting a conflict could lead to catastrophic consequences. When predicting far ahead, reliability is less crucial but precision is required in order to limit the number of false alarms and unnecessary maneuvers.

Future works might extend the proposed approach to various look-ahead times, considering the above discussed trade-off between reliability and precision, and also to different initial altitudes at $t_0$. The regression approach could also be extended to several aircraft types, possibly considering hierarchical models combining specialized sub-models in order to improve the prediction.

## REFERENCES

[1] H. Swenson, R. Barhydt, and M. Landis, "Next Generation Air Transportation System (NGATS) Air Traffic Management (ATM)-Airspace Project," National Aeronautics and Space Administration, Tech. Rep., 2006.

[2] S. Consortium, "Milestone Deliverable D3: The ATM Target Concept," EUROCONTROL, Tech. Rep., 2007.

[3] B. Musialek, C. F. Munafo, H. Ryan, and M. Paglione, "Literature survey of trajectory predictor technology," Federal Aviation Administration William J. Hughes Technical Center, Atlantic City, Technical report DOT/FAA/TC-TN11/1, November 2010.

[4] R. A. Vivona, M. M. Paglione, K. T. Cate, and G. Enea, "Definition and demonstration of a methodology for validating aircraft trajectory predictors," in *AIAA Guidance, Navigation, and Control Conference*, 2010.

[5] G. L. Slater, "Adaptive improvement of aircraft climb performance for air traffic control applications," in *Proceedings of the 2002 IEEE International Symposium on Intelligent Control*. IEEE conference publications, October 2002.

[6] A. Warren and Y. Ebrahimi, "Vertical path trajectory prediction for next generation atm," in *Digital Avionics Systems Conference, 1998. Proceedings., 17th DASC. The AIAA/IEEE/SAE*, vol. 2, oct-7 nov 1998, pp. F11/1 –F11/8 vol.2.

[7] A. Warren, "Trajectory prediction concepts for next generation air traffic management," in *3rd USA/Europe ATM R&D Seminar*, June 2000.

[8] C. Schultz, D. Thipphavong, and H. Erzberger, "Adaptive trajectory prediction algorithm for climbing flights," in *AIAA Guidance, Navigation, and Control (GNC) Conference*, August 2012.

[9] D. P. Thipphavong, C. A. Schultz, A. G. Lee, and S. H. Chan, "Adaptive algorithm to improve trajectory prediction accuracy of climbing aircraft," *Journal of Guidance, Control, and Dynamics*, vol. 36, no. 1, pp. 15–24, 2012.

[10] R. Alligier, D. Gianazza, M. Ghasemi Hamed, and N. Durand, "Comparison of Two Ground-based Mass Estimation Methods on Real Data," in *International Conference on Research in Air Transportation (ICRAT), Istambul, 26/05/2014-30/05/2014*. http://www.icrat.org: ICRAT, mai 2014, pp. 1–8. [Online]. Available: www.icrat.org

[11] R. Alligier, D. Gianazza, and N. Durand, "Ground-based estimation of aircraft mass, adaptive vs. least squares method," in *10th USA/Europe Air Traffic Management Research and Developpment Seminar*, 2013.

[12] A. Eckstein, "Data driven modeling for the simulation of converging runway operations," in *Proceedings of the 4th International Conference on Research in Air Transportation*. ICRAT, June 2010. [Online]. Available: http://www.icrat.org/icrat/pdf/Data%20Driven%20Modeling.pdf

[13] Y. L. Fablec and J. Alliot, "Using neural networks to predict aircraft trajectories," in *Proceedings of the 1999 International Conference on Artificial Intelligence, IC-AI*, 1999, pp. 524–529.

[14] M. Ghasemi Hamed, M. Serrurier, and D. Durand, "Possibilistic KNN regression using tolerance intervals." in *IPMU (3)*, ser. Communications in Computer and Information Science, S. Greco, B. Bouchon-Meunier, G. Coletti, M. Fedrizzi, B. Matarazzo, and R. R. Yager, Eds., vol. 299. Springer, 2012, pp. 410–419. [Online]. Available: http://dblp.uni-trier.de/db/conf/ipmu/ipmu2012-3.html#HamedSD12

[15] A. Hadjaz, , G. Marceau, P. Savéant, and M. Schoenauer, "Online learning for ground trajectory prediction," in *Proceedings of the SESAR Innovation Days (2012)*. EUROCONTROL, 2012.

[16] 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, no. 0, pp. 45 – 60, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X13001708

[17] M. Ghasemi Hamed, D. Gianazza, M. Serrurier, and N. Durand, "Statistical prediction of aircraft trajectory: regression methods vs point-mass model," in *ATM 2013, Chicago, USA, June*, 2013.

[18] "Base of aircraft data (BADA) aircraft performance modelling report," EUROCONTROL, Tech. Rep., 2009.

[19] S. G. Park and J.-P. Clarke, "Vertical trajectory optimization for continuous descent arrival procedure," in *AIAA Guidance, Navigation, and Control Conference, Minneapolis (Minnesota-USA), Paper*, vol. 4757, 2012, p. 2012.

[20] R. Slattery and Y. Zhao, "Trajectory synthesis for air traffic automation," *Journal of Guidance, Control and Dynamics*, vol. 20, no. 2, pp. 232–238, 1997.

[21] J. K. Kuchar and L. C. Yang, "A review of conflict detection and resolution modeling methods," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 1, no. 4, pp. 179–189, 2000.

[22] H. Erzberger, R. A. Paielli, D. R. Isaacson, and M. M. Eshow, "Conflict detection and resolution in the presence of prediction error," in *1st USA/Europe Air Traffic Management R&D Seminar, Saclay, France*. Citeseer, 1997, pp. 17–20.

[23] D. R. Isaacson and H. Erzberger, "Design of a conflict detection algorithm for the center/tracon automation system," in *Digital Avionics Systems Conference, 1997. 16th DASC., AIAA/IEEE*, vol. 2. IEEE, 1997, pp. 9–3.

[24] M. Prandini, J. Hu, J. Lygeros, and S. Sastry, "A probabilistic approach to aircraft conflict detection," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 1, no. 4, pp. 199–220, 2000.

[25] G. Bakker, H. Kremer, and H. Blom, "Geometric and probabilistic approaches towards conflict prediction," in *Proceedings of 3rd USA/Europe Air Traffic Management R&D Seminar, Italy*, 2000.

[26] A. Bashllari, N. Kaciroti, D. Nace, and A. Fundo, "Conflict probability estimations based on geometrical and bayesian approaches," in *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*. IEEE, 2007, pp. 479–484.

[27] N. Durand, J.-M. Alliot, and O. Chansou, "Optimal resolution of en route conflicts," in *Procceedings of the 1st USA/Europe Seminar*, 1997.

[28] N. Durand and J.-M. Alliot, "Ant colony optimization for air traffic conflict resolution," in *Proceedings of the Eighth USA/Europe Air Traffic Management Research and Development Seminar. Eurocontrol/FAA*, 2009.

[29] C. Vanaret, D. Gianazza, N. Durand, and J. B. Gotteland, "Benchmarking conflict resolution algorithms," in *5th International Conference on Research in Air Transportation, ICRAT, Berkeley, USA*, 2012.

[30] N. Durand, J. Alliot, and J. Noailles, "Automatic aircraft conflict resolution using genetic algorithms," in *Proceedings of the Symposium on Applied Computing, Philadelphia*. ACM, 1996.

[31] G. Granger, N. Durand, and J.-M. Alliot, "Optimal resolution of en route conflicts," in *AGARD MSP 1997, Workshop on Air traffic management*, 1997.

[32] F. Drogoul, P. Averty, and R. Weber, "Erasmus strategic deconfliction to benefit sesar," in *Proceedings of the 8th USA/Europe Air Traffic Management R&D Seminar*, June-July 2009. [Online]. Available: http://atm2003.eurocontrol.fr/8th-seminar-united-states-june-2009

[33] X. Prats, V. Puig, J. Quevedo, and F. Nejjari, "Multi-objective optimisation for aircraft departure trajectories minimising noise annoyance," *Transportation Research Part C*, vol. 18, no. 6, pp. 975–989, 2010. [Online]. Available: www.elsevier.com/locate/trc

[34] R. Koenker and G. Bassett, "Regression quantiles," *Econometrica*, vol. 46, no. 1, pp. 33–50, 1978. [Online]. Available: http://www.jstor.org/stable/1913643

[35] R. Koenker, *Quantile Regression*, ser. Econometric Society Monographs. Cambridge University Press, 2005.

[36] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *J. Roy. Stat. Soc.*, vol. 36, pp. 111–147, 1974.

[37] C. R. Rao and H. Toutenburg, *Linear Models: Least Squares and Alternatives (Springer Series in Statistics)*. Springer, Jul. 1999. [Online]. Available: http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0387988483

[38] K. Krishnamoorthy and T. Mathew, *Statistical Tolerance Regions: Theory, Applications, and Computation*, ser. Wiley Series in Probability and Statistics. Wiley, 2009.

[39] G. J. Hahn and W. Q. Meeker, *Statistical Intervals: A Guide for Practitioners*. John Wiley and Sons, 1991.

[40] M. Kocherginsky, X. He, and Y. Mu, "Practical confidence intervals for regression quantiles," *Journal of Computational and Graphical Statistics*, vol. 14, no. 1, pp. 41–55, 2005.

[41] M. Ghasemi Hamed, M. Serrurier, and D. Durand, "Simultaneous interval regression for k-nearest neighbor," in *Australasian Conference on Artificial Intelligence*, 2012, pp. 602–613.

[42] W. A. Wallis, "Tolerance intervals for linear regression," in *Proceedings Second Berkeley Symposium on Mathematical Statistics and Probability*. Univ. of Calif. Press, 1951, pp. 43–51.

[43] G. J. Lieberman and J. Miller, R. G., "Simultaneous tolerance intervals in regression," *Biometrika*, vol. 50, no. 1/2, pp. 155–168, 1963. [Online]. Available: http://www.jstor.org/stable/2333756

[44] A. Nuic, "User manual for Base of Aircraft DATA (BADA) rev.3.7." EUROCONTROL, Tech. Rep., 2009.

[45] R. Alligier, D. Gianazza, N. Durand *et al.*, "Energy rate prediction using an equivalent thrust setting profile," in *ICRAT 2012, 5th International Conference on Research in Air Transportation*, 2012.

[46] R. A. Coppenbarger, "Climb trajectory prediction enhancement using airline flight-planning information," in *AIAA Guidance, Navigation, and Control Conference*, 1999.

[47] E. E. Center, "Study of the acquisition of data from aircraft operators to aid trajectory prediction calculation," EUROCONTROL Experimental Center, Tech. Rep., 1998.

**Mohammad Ghasemi Hamed** received his Ph.D. (2014) degree in Computer Science from the "Institut National Polytechnique de Toulouse" (INPT) in Toulouse, France and his M.Sc. (2010) in Artificial Intelligence from the University of Toulouse III. He has a graduate degree (2010) in Computer Engineering from ENSEEIHT in Toulouse and an undergraduate degree in Computer Engineering from the University of Shahid Beheshti in Tehran, Iran. He is currently a postdoctoral fellow in Lab-STICC at ENSTA Beratgne in Brest, France.

**Richard Alligier** received his Ph.D. (2014) degree in Computer Science from the "Institut National Polytechnique de Toulouse" (INPT), his engineering degree (IEEAC, 2010) from the French university of civil aviation (ENAC) and his M.Sc. (2010) in computer science from the University of Toulouse. He is currently assistant professor at the ENAC in Toulouse, France.
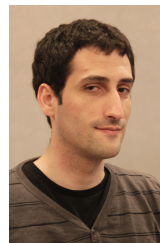
**David Gianazza** graduated from the French University of Civil Aviation (ENAC) in 1986. He received his M.Sc., Ph.D., and HDR (post-doctoral degree similar to tenure) degrees in computer science from "Institut National Polytechnique de Toulouse" in 1996, 2004 and 2016, respectively. He has held various positions in the French Civil Aviation Administration, successively as an engineer in ATC operations, technical manager, and researcher. He is currently an Associate Professor at ENAC.