Data-driven methodology for uncertainty quantification of aircraft trajectory predictions

Andrés Muñoz Hernández*, Manuel Polaina Morales*, Alejandro Güemes†, Manuel Soler‡, Daniel González-Arribas‡, Jordi Pons§, Xavier Prats§, Arinc Tutku Altun¶, Emre Koyuncu¶, Alexander Kuenz‖, Raimund Zopp**, Daniel Delahaye††

*Digital Aviation Services and Autonomy, Boeing Deutschland GmbH, Germany
†MCA Engineering, Spain
‡Aerospace Engineering Department, Universidad Carlos III de Madrid, Spain
§Department of Physics - Aerospace Division, Universitat Politécnica de Catalunya, Spain
¶Aerospace Research Center (ITU ARC), Istanbul Technical University, Turkey
‖Institute of Flight Guidance (DLR), Germany
**Flightkeys, Austria
††OPTIM GROUP, ENAC Laboratory, Ecole Nationale de l'Aviation Civile, France

***Abstract*—This work present a framework based on data-driven techniques for quantifying and chaos theory for propagating the uncertainty present in the aircraft trajectory prediction process when computing the expected trajectory from a given flight plan. The developed framework employs data assimilation models to capture real-time information from the air traffic system and introduces a novel methodology in order to account for the uncertainty of the weather conditions. The comparison of the resulting set of probabilistic trajectories and the actually flown ones proves how the former could be a key enabler to support envisioned trajectory-based operation concepts and modern airline operations planning.**

***Index Terms*—uncertainty quantification, probabilistic aircraft trajectory, ATM systems, flight planning**

I. Introduction

Within the full Air Traffic Management (ATM) system, there are multiple influencing elements that have the potential of deviating an aircraft trajectory from its nominal or planned development and introduce uncertainty in its evolution, both due to factors inherent to the trajectory (e.g., the type of aircraft executing the operation or the selected route) and due to factors exogenous to it, such as the congestion in the air traffic system at a particular point in time or spontaneous Air Traffic Control (ATC) interventions. These factors, of unknown *a priori* magnitude, play a very significant role in the generation of the observed deviations between the aircraft trajectory planned in a strategical or pre-tactical phase through a trajectory prediction process and the one actually being executed on the tactical phase.

This project has received funding from the SESAR Joint Undertaking (JU) under grant agreement No 893204. The JU receives support from the European Union's Horizon 2020 research and innovation programme and the SESAR JU members other than the Union.
The communication activity related to the action will indicate that it reflects only the author's view and that the SJU is not responsible for any use that may be made of the information it contains.

Quantifying and modelling the effect of these uncertain factors when computing the prediction for any given aircraft trajectory in any potential air traffic scenario is a key step in the development of the capabilities for implementing robust and resilient airline operations. Within this work, the focus has been placed on characterizing the micro- or trajectory-level uncertainty, paying attention to the modelling of those uncertain factors unique and inherent to the aircraft trajectory prediction process, disregarding for now those potential influences coming from the air traffic network or system, such as conflicts with other airborne aircraft or deviations coming from demand-capacity imbalances.

The characterization and propagation of the uncertainty present at the aircraft trajectory prediction level is a complex problem that is by no means a new topic, as it has been widely covered in the literature [1] [2]. Generally, most approaches are based on the identification and consideration of the potential uncertainty sources that may impact a flight prior and during its development. This results in the computation of distributions of the potential values for the uncertain variables, and that, jointly propagated, would lead to the consequent probabilistic distributions of the evolution in time of the state variables defining the aircraft trajectory.

For the earlier phase of characterizing the considered uncertainty sources, the aim is to propose a data-driven approach that relies solely on data inputs from historical instances to infer the probability distributions of the expected values for the identified uncertain factors. By doing so, the work deviates from previous approaches that take assumptions on the potential values to be expected based on models or standard distributions [3] and allows for retrieving any value distribution specific to the problem being studied [4].

Regarding the propagation of the characterized uncertain factors downstream through the aircraft trajectory prediction process, the classical approach observed in the literature relies on the application of statistical studies on numerous evaluations of the variability of the output of the dynamical system being studied (i.e., an aircraft trajectory) as a function of the variability of the inputs (i.e., uncertain factors for the aircraft trajectory prediction tool). This is mostly supported by the implementation of Monte Carlo simulations, and introducing different ways of modelling the variability of the inputs, such as with neural networks [5] or worst-case prediction algorithms [6]. However, the high computational demand associated to Monte Carlo simulations has been pointed out as well in the literature when working with aircraft trajectory predictions in large-scale air traffic situations [7].

With the purpose of finding a more efficient framework for the uncertainty propagation in aircraft trajectory prediction, a method based on Polynomial Chaos Expansions is proposed, following approaches found to be successful in similar problems [8]. This framework will allow for establishing a efficient methodology for assessing the probability distribution of the output trajectories that could be expected when accounting for possible deviations from a given flight plan based on the variability of the uncertain input factors of the aircraft trajectory prediction process.

Within this framework, a novel methodology will be embedded in order to account for the potential inaccuracies in the forecasts of the weather conditions affecting the trajectory in its development. As part of the data-driven approach proposed, this allows to avoid approaches that assume a given *a priori* distribution of the potential variations of the weather variables [9], and provides a way to employ Ensemble Prediction System (EPS) forecasts that account for perturbations in the weather conditions.

This paper is organized as follows: Section II describes the process followed for the identification and quantification of the sources of uncertainty to be considered; Section III introduces the theoretical background necessary for applying arbitrary Polynomial Chaos Expansions for the propagation of the characterized uncertainties within the prediction of the aircraft trajectories; Section IV establishes the framework that is employed for the obtainment of the probability distributions for the variables of interest defining the aircraft trajectory by integrating the quantification and propagation of the identified uncertainties; Section V proposes a case study showing the capabilities of the developed framework when applied to a relevant air traffic scenario.

II. Trajectory-level uncertainty characterization

The task of determining the relevant sources of uncertainty in the aircraft trajectory prediction activity can be implemented with different levels of rigor and sophistication, mainly depending on the specific factors that can be modelled as inputs in the implementation of the trajectory prediction tool or engine to be employed. The literature has widely covered the identification of the sources of uncertainty that should be considered in any first approach [10], classifying them in four basic groups:

- Aircraft intent uncertainty: The way in which the aircraft is actually operated (i.e., operational definition of the aircraft intent in the tactical phase) can differ from the way in which the trajectory was initially declared. In addition to this, further deviations could be incurred due to potential errors in the navigation of the intended path, although modern flight management systems cause these deviations to be negligible.
- Aircraft performance uncertainty: This group encompasses all those deviations that may arise from the differences between the aircraft performance model used to integrate the dynamic equations of motion of the aircraft in the trajectory prediction task and the real aircraft performance characteristics of the aircraft operating the actual flight.
- Initial conditions uncertainty: There may be differences between the initial flight conditions considered for the prediction of the aircraft trajectory, propagated then forward in the calculation process, and the ones that are incurred in reality due to typical modifications (e.g., flight departing later than scheduled or using a runway configuration for departure different than the one planned).
- Weather conditions uncertainty: The trajectory prediction process normally employs a given forecast of the weather conditions to be expected at the time of flight in order to model the trajectory accounting for this relevant factor. However, these forecasts do normally present deviations with respect to the incurred conditions in reality, as it is normal for a highly stochastic process such as the weather evolution, affecting then the flight in the tactical phase.

Within this work, the ambition is to deploy a methodology capable of assessing the actual impact of most of those factors characterizing each source of uncertainties. For each classification, a different approach needs to be taken according to its details and the data available, as there is too much variation between them to establish a common methodology. However, the main driver transverse to all of them is to apply a data-driven approach for the quantification of the present uncertainty, avoiding any *a priori* assumption not supported by actual data observations.

A. Aircraft intent

The uncertainty related to the aircraft intent stems from the deviations incurred from the flight plan, which is considered to be the basis information for the planned intent, with respect to the actual trajectory flown. Therefore, the characterization of these differences through a planned vs flown comparison analysis between the intent features of the flight plan and of the actual trajectory allows to retrieve probability distributions of the values of the observed deviations.

This analysis is enabled by the employment of two data pieces: Filed Traffic Flight Model (FTFM) flight plan definitions, retrieved from EUROCONTROL's Demand Data Repository (DDR2), and actual trajectory profiles from

FlightRadar24 Automatic Dependent Surveillance Broadcast (ADSB) data reconstructed following the INTRAC process [11] [12], illustrated schematically in Figure 1. The comparison between the former and the latter will then provide some insight on the statistical distribution defining the differences to be expected in the trajectory execution.

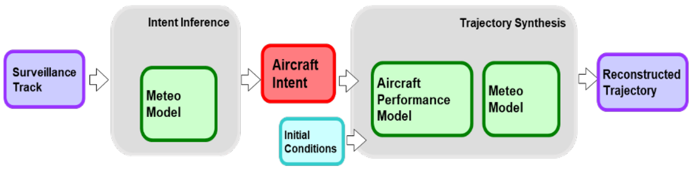

Fig. 1. Schema of the process followed to obtain reconstructed flight profiles from ADSB surveillance data.

The parametrization of these differences depend on the methodology employed to compare trajectories, which is a complex task. As a straightforward and effective simplification of the trajectory comparison activity, the differences to be studied will be focused on relevant parameters that define on a high-level the trajectory. The number and type of parameters to be considered can vary. As was introduced before and will be further explained in Section III, the output trajectory will be computed as a function of the value of the variable inputs for the trajectory computation. Therefore, the selected aircraft intent parameters need to be part of the input required by the selected computation engine for the trajectory prediction process.

For this work and attending to this constraint, the following parameters describing the trajectory were selected: constant calibrated airspeed (*CAS*) and Mach number (*M*) set during the climb and descent phases, the pressure altitude (*Hp*) attained at the Top of Climb (TOC) and Top of Descent (TOD) and the pressure altitude and Mach number set during the relevant cruise segment.

B. Aircraft performance

The study of the deviations in the planned and actual trajectory stemming from the differences between the modelled and the actual performance of the aircraft operating are quite relevant. Significant differences between the real performance characteristics and the performance model used when planning the operations can have a major impact in the tactical phase and can lead to unrealistic planned trajectories that could not be possibly flown with the real aircraft.

In order to study these differences, the data-driven approach to be followed is similar to the one proposed for the aircraft intent uncertainty, with a comparison-based analysis between the model used for planning and the actual performance observed in reality. For the former, the aircraft performance model of reference for the planning of trajectories in Europe is Base of Aircraft Data (BADA) 4 [13], which is easily accessible. Accessing real performance data of actual aircraft is complicated, since airlines tend to keep these data confidential.

Due to the lack of availability of actual performance data from any on-board Quick Access Recorder (QAR), this comparison-based analysis can not be implemented, and therefore no data-driven approach can be adopted. Therefore, for this work, deviations coming from differences between the modelled and actual aircraft performance will be neglected.

C. Initial conditions

In the trajectory prediction process, the initial conditions from which the state variables will be propagated forward have a very significant influence in the later values of the aircraft trajectory variables. Any potential variation arising from differences with respect to the planned ones must be studied in order to execute an accurate assessment of the aircraft trajectory. The typical parameters to be studied are the initial 4D position of the aircraft and the initial mass.

With respect to the latter, the lack of QAR data impedes the comparison-based analysis with assumed values in the planning phase. There are alternatives to estimate this initial mass, as for example the reconstruction process proposed by INTRAC [14], but all of them employ model-based methodologies that do not comply with a data-driven approach, so the effect of deviations in the initial mass values will be neglected within this work.

Regarding the initial position and time of the aircraft, having an accurate estimation of the first point to be propagated is essential to obtain an accurate propagation of the trajectory. For this purpose, and due to the lack of a valid way to retrieve planned (flight plans may not explicitly define the runway configuration to be used or the take-off time) and flown (surveillance data on ground tends not to be as consistent as airborne data) observations for these data pieces to perform the comparison-based uncertainty characterization, an alternative and accurate way of building estimates needs to be established.

The proposed method within this work is to develop data assimilation models that are able to capture, assimilate and process current air traffic data for the system of interest in order to issue an estimate for the expected take-off time and runway configuration to be followed by the operation of interest. These data assimilation models would then be fed historical instances of the incurred conditions of past operations. Then, using statistical and machine learning techniques, a normal approach for prediction of various aircraft trajectory variables [15], use them to estimate the operational initial conditions for the flights in the proposed scenario.

When looking specifically to the initial time, a probability distribution can then be built for the potential range of values to be adopted by this initial condition based on the estimates issued. Assuming that this initial take-off time can be approximated as the sum of the time spent in turnaround and taxi phases (i.e., assuming the parking time for the prior operation served by the tail of interest is known), probability distributions can be built for the former based on the joint probability distributions of the estimated turnaround and taxi times. Examples of these estimations are introduced below for taxi times in Figure 2 and for turnaround times in Table I.

For the taxi time estimation, both a machine learning and a deep learning model were trained on one year worth of historical taxi times for departure operations on LEMD and EGLL airports. The error incurred by the built models when estimating the taxi times on both airports during January 2018 is shown in Figure 2. This example serves to illustrate how the probability distribution characterizing the taxi time, and consequently the initial time, would then be dependent on the forecasting error.

Similarly, an statistical analysis of the historical surveillance data instances can provide the distributions characterizing the turnaround times to be observed in any airport of interest. An example of these calculated turnaround times using DDR2 data is introduced in Table I for LEMD airport during January 2018, which establishes different distributions depending on the runway configuration employed by the consecutive arrival-departure operations.

This is then complemented with the estimation of the runway configuration to be used by any operation during its departure and arrival phases. As commented before, this prediction can be used as an updated estimate of the initial point of the trajectory based on current information relevant to the air traffic scenario of interest, and as introduced in Table I, will also participate in the estimate for the turnaround time, and consequently, for the initial time. The issuance of the runway configuration estimations can also be executed by applying a data-driven approach of gathering historical surveillance data instances and tailor them to feed different machine learning models. An example of the accuracies obtained in a relevant scenario is shown in Table II, predicting the runway configuration for both departure and arrival phases on LEMD during 2018 using different classification algorithms (Decision Tree Classifier (DTC), K-Nearest Neighbor (KNN) and Naive-Bayes Classifier (NBC)).

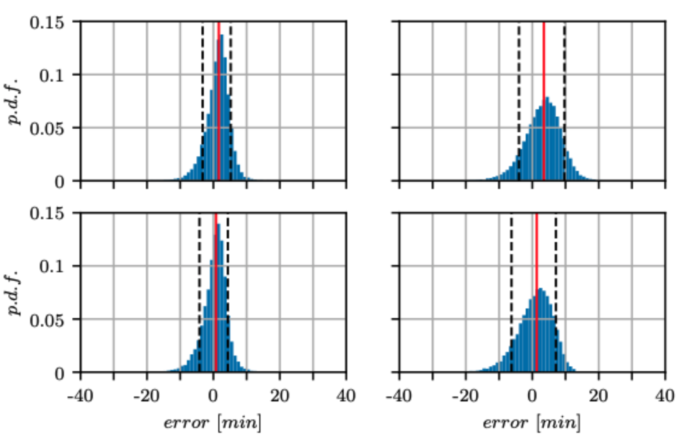

Fig. 2. Distribution of the estimation error incurred on the taxi time prediction for all departure operations during January 2018 from (left column) LEMD and (right column) EGLL when using a (top row) machine learning and a (bottom row) deep learning model. Red vertical line indicates the median of the error, black dashed lines indicate the 10th and 90th percentile.

TABLE I

PROBABILITY DISTRIBUTION PARAMETERS FOR THE ESTIMATED TURNAROUND TIMES OF DEPARTURE OPERATIONS IN LEMD DURING JANUARY 2018 FOR DIFFERENT RUNWAY CONFIGURATIONS (IN MINUTES)

Landing	Take-off	Normal Distribution	Gamma distribution
14	14	$N(91.10, 22.70)$	$\Gamma(17.07, 0.187)$
14	36	$N(86.73, 23.63)$	$\Gamma(14.32, 0.165)$
36	14	$N(87.50, 21.60)$	$\Gamma(17.69, 0.202)$
36	36	$N(87.83, 20.44)$	$\Gamma(20.27, 0.231)$

TABLE II

TRAINING AND TESTING ACCURACIES FOR THE MACHINE LEARNING MODELS BUILT FOR THE RUNWAY CONFIGURATION ESTIMATION FOR ALL OPERATIONS IN LEMD DURING JANUARY 2018.

Phase	Classifier	Training accuracy	Testing accuracy
Departure	DTC	0.73	0.67
	KNN	0.71	0.66
	NBC	0.62	0.62
Arrival	DTC	0.90	0.86
	KNN	0.89	0.87
	NBC	0.84	0.95

D. Weather conditions

Weather conditions are a very relevant factor affecting the evolution of any given aircraft trajectory. In any given aircraft trajectory prediction process, weather forecasts are used to provide an estimation of the weather conditions in which the flight will have to operate, and it is a key step in any planning process [16]. Consequently, the deviations between the weather situation indicated by the employed forecasts and the conditions that are actually incurred are an uncertainty factor that needs to be taken into account.

However, this is a very complicated task as any weather forecast is a high dimensional element whose uncertainty can not be easily quantified, as it entails two main issues. The first one is that a comparison-based analysis between the planned or expected weather conditions and the actual ones is not possible. Although the expected one could be considered as that weather forecast that was used for the aircraft trajectory planning activities, there is not a single weather element reflecting the actual conditions. Reanalysis products, which are reconstructions of past weather instances blending meteorological models and multiple real observations are the best approximation (e.g., European Control for Medium-range Weather Forecasts (ECMWF) ERA5 datasets employing Integrated Forecast System (IFS) models), but still they do not rely solely on actual data.

A second issue with the quantification of uncertainty in the weather conditions is that deviations in such a large-scale stochastic element can not be boiled down to joint probability distributions. Weather is time and location dependent, so a probability distribution of the potential deviation should be characterized for every point in which a weather estimate is issued, making the problem impossible to handle.

Solutions to the problem of including weather uncertainty in the aircraft trajectory prediction process observed in the literature entail simplifications and/or largely increased computational demand. When employing Monte Carlo simulations, the straightforward approach of evaluating the scenario of interest with different potential weather conditions [17] consequently introduces a significant computational charge. Other approaches which employ correctly the polynomial chaos theory for the aircraft trajectory prediction modelling use simplified distributions of the weather conditions [18] disregarding temporal or spatial dependence.

As an alternative, this work proposes to directly embed the weather forecast as a set of dependent variables in the polynomial chaos expansion through a dimensionality reduction process. With this method, there is no need to simplify the potential variations of the weather as provided in ensemble forecasts, and it does not entail a significant increase in the computational demand. The specifics of this process will be detailed in the next section.

III. Uncertainty propagation using arbitrary Polynomial Chaos Expansions

Following the characterization of the sources of uncertainty affecting the aircraft trajectory prediction process, it will be paramount to develop the capabilities to assess the impact of the variability of the identified stochastic factors on the operations of interest. This is an essential step to evaluate how input variability affects the output predicted trajectories to be obtained, as the spectrum of possible predictions will depend on the value distributions for each of the identified uncertain input variables.

Therefore, now the focus is on determining how to propagate these characterized uncertainties along the aircraft trajectory prediction process, so that an assessment can be obtained on the effect of the inputs' variability on the full trajectory predicted for a nominal flight plan.

The selected framework for the uncertainty propagation task is based on the Polynomial Chaos (PC) theory. This is proposed as an alternative to the classic Monte Carlo approach that still will provide the capabilities to retrieve the probabilistic trajectories that can be output for an initial flight plan as a function of the spectrum of values to be adopted by the characterized uncertain inputs for the aircraft trajectory prediction process.

A. Theoretical framework

Norbert Wiener introduced the concept of chaos theory in 1938 [20], which states that any function or model z dependent of a stochastic variable ξ can be posed as a linear combination of coefficients $a_i(t)$, independent of the stochastic variable, times a set of one-dimensional polynomials $\gamma_i(\xi)$, which form a basis orthogonal to the probabilistic distribution of the stochastic variable ξ, such that:

$$z(t,\xi) = \sum_{i=1}^{\infty} a_i(t)\gamma_i(\xi) \approx \sum_{i=1}^{d} a_i(t)\gamma_i(\xi), \tag{1}$$

where the subindex i refers to polynomial degree. Whilst the approximation error disappears when i tends to infinity, it is common practice to truncate the model at certain polynomial degree d. Nonetheless, any realistic model representing a physical mechanism depends on several stochastic parameters such that $\xi = \xi_1, \xi_2, ..., \xi_N$. Henceforth, the total number of stochastic input parameters will be referred as N. Consequently, (1) needs to be reformulated as a multidimensional polynomial expansion as follows:

$$z(t,\xi_1,\xi_2,...,\xi_N) = \sum_{i=1}^{\infty} b_i(t)\Gamma_i(\xi_1,\xi_2,...,\xi_N), \tag{2}$$

where $b_i(t)$ still quantifies the model's dependence on the polynomial expansion, while $\gamma_i(\xi_1,\xi_2,...,\xi_N)$ contains a multidimensional orthogonal polynomial basis for the stochastic variables ξ. Assuming that the stochastic input variables ξ are independent of each other, the multidimensional basis can be constructed as a simple product of the one-dimensional polynomials, such that:

$$\Gamma_i(\xi_1,\xi_2,...,\xi_N) = \prod_{j=1}^{N} \gamma_j^{\alpha_j^i}(\xi_j), \tag{3}$$

where the index α_j^i is used to indicate the combinatory information between the different independent variables and the polynomial degrees, such that:

$$\sum_{j=1}^{N} \alpha_j^i \leq M, \tag{4}$$

where index i ranges from 1 to M. The number of possible combinations between polynomials degrees and the different stochastic variables is defined as:

$$M = \frac{(N+d)!}{N!d!}, \tag{5}$$

which allows to define α_j^i as $M \times N$ matrix containing the corresponding degree of the stochastic variables for each combination.

B. Polynomial expansions definition

While there are several methods to compute these polynomials based on predefined probability density functions such as normal, gamma, beta or uniform, this project tackles these polynomials construction from a data-driven point of view, for which the arbitrary Polynomial Chaos Expansion (aPCE) [21] method is used. This method uses the statistical moments for each stochastic variable, calculated as:

$$\mu_k = \int \xi_i^m dP(\xi_i), \tag{6}$$

where each polynomial γ_i^k is defined with a set of polynomial coefficients $c_m^{(k)}$ multiplied by their corresponding power of the stochastic variable ξ_i, such that:

$$\gamma_i^k = \sum_{m=0}^{k} c_m^{(k)} \xi_i^m, \tag{7}$$

and

$$c_{m=k}^{(k)} = 1. \tag{8}$$

For each stochastic variable, the polynomial coefficients can be computed by ensuring the orthogonality between two polynomials of order k and l such that:

$$\int \gamma^k \cdot \gamma^l \cdot dP(\xi_i) = \delta_{k,l}, \tag{9}$$

where $\delta_{k,l}$ is the Kronecker delta and is equal to 0 unless k is equal to l. Subsequently, a set of equations can be defined for each polynomial degree such that:

$$\int c_0^{(0)} \cdot \left[\sum_{m=0}^{l} c_m^{(k)} \xi_i^m \right] \cdot dP(\xi_i) = 0 \tag{10}$$

$$\int \left[\sum_{m=0}^{1} c_m^{(1)} \xi_i^m \right] \cdot \left[\sum_{m=0}^{l} c_m^{(k)} \xi_i^m \right] \cdot dP(\xi_i) = 0 \tag{11}$$

$$\vdots$$

$$\int \left[\sum_{m=0}^{k-1} c_m^{(k-1)} \xi_i^m \right] \cdot \left[\sum_{m=0}^{l} c_m^{(k)} \xi_i^m \right] \cdot dP(\xi_i) = 0 \tag{12}$$

$$\int \left[\sum_{m=0}^{k} c_m^{(k)} \xi_i^m \right] \cdot \left[\sum_{m=0}^{l} c_m^{(k)} \xi_i^m \right] \cdot dP(\xi_i) = 1. \tag{13}$$

Substituting (6) into $(10) - (13)$ leads to a system of linear equations that can be written in a matrix form.

$$\begin{bmatrix} \mu_0 & \mu_1 & \cdots & \mu_k \\ \mu_1 & \mu_2 & \cdots & \mu_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{k-1} & \mu_k & \cdots & \mu_{2k-1} \\ 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} c_0^{(k)} \\ c_1^{(k)} \\ \vdots \\ c_{k-1}^{(k)} \\ c_k^{(k)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}. \tag{14}$$

C. Coefficients calculation

To compute the aPCE coefficients $b_i(t)$, the aPCE polynomials $\gamma_i^k(\xi)$ need to be evaluated at certain points of the stochastic-variable parametric space in order to solve the system of equations defined in (14), where the number of unknowns is defined by (5). Several methods can be found in the literature for this task, such as Galerkin projection [22] or collocation [23] methods, of which we will use the latter. The collocation method evaluates each polynomial expansion at certain values, known as collocation points, which are extracted from the roots of the next higher-order polynomial for each stochastic parameter [24]. This implies that, for a polynomial of order d, the number of available collocation points is $(d+1)^N$, which is always larger than the number of unknowns in the system of equations. This overdetermined system is solved by selecting the optimal M combination of collocation points based on the probability of each combination of collocation points. This probability is computed from the sum of the polynomial degree of each stochastic variable in every combination, assuming than in a standard Gaussian random variable with zero mean and unit variance the higher the root degree the lower the probability to occur.

Each of the M combinations of collocation points require to compute a solution of $z(t, \xi_1, \xi_2, ..., \xi_N)$. Following [25], the coefficients $b_i(t)$ can be computer either using Galerkin projection:

$$\left\langle \sum_{k=0}^{d} b_k \Gamma_k(\xi^{(i)}), \Gamma_l(\xi^{(i)}) \right\rangle = 0 \tag{15}$$

or least-square approximation:

$$\hat{b}_i(t) = argmin_b \sum_{i=1}^{M} \left(z^{(i)} - \sum_{k=0}^{d} b_k \Gamma_k(\xi^{(i)}) \right). \tag{16}$$

D. Weather uncertainty consideration

The computational cost of arbitrary polynomial chaos expansion increases factorially with the increase of the polynomial degree and the number of variables, as shown in (5). This poses a problem when dealing when systems affected by a large number of stochastic input variables, such as weather information that would require $\mathcal{O}(10^{12})$ collocation points, which is far beyond the scope of computational cost reduction intended for aPCE and the computational capabilities available in the consortium. To overcome this issue, a dimensionality reduction of the weather information can be carried out, allowing the aPCE to deal with a reduced version of weather. Later, the collocation points provided by aPCE for this reduced-state weather information can be translated back into the original weather dimension, allowing any aircraft trajectory predictor tool to ingest this artificial information. For that purpose, proper orthogonal decomposition (POD), also known as principal component analysis or Karhunen–Loève decomposition, has been proposed. This decomposition states that any signal dependent of time t and space x can be decomposed in a time-averaged value plus a fluctuating component defined as a linear combination of a spatial basis, composed of spatially orthonormal functions $\phi_i(x)$, times a temporal orthonormal basis made of orthonormal function $\psi_i(t)$:

$$s(x,t) \approx \bar{s}(x) + \sum_{i=1}^{N_m} \psi_i(t) \sigma_i \phi(x), \tag{17}$$

where N_m is the number of orthonormal modes used for the reconstruction and σ_i is a weighting factor. The practical implementation of POD follows the method of snapshots proposed in [26]. Each weather sample is reshaped in a vector of size N_p, where this value refers to the total number of variables in each sample. The total number of samples N_t in vector form are rearranged in a matrix:

$$S = \begin{bmatrix} s(x_1, t_1) & \cdots & s(x_{N_p}, t_1) \\ \vdots & \ddots & \vdots \\ s(x_1, t_{N_t}) & \cdots & s(x_{N_p}, t_{N_t}) \end{bmatrix}, \tag{18}$$

with size $N_t \times N_p$, where each row refers to a sample and each column to a variable. The matrix Ψ containing the POD temporal modes can be obtained solving the eigenvalue problem of the temporal correlation matrix S as follows:

$$C = SS^T = \Psi \Lambda \Psi^T, \tag{19}$$

where λ is a diagonal matrix with elements $\lambda_i = \sigma_i^2$ representing the variance content of each mode. The σ_i coefficient can be rearranged in a diagonal matrix Σ. Finally, the matrix Φ containing the spatial modes can be obtained by projecting the weather matrix of the temporal basis as:

$$\Phi = \Sigma^{-1} \Psi^T S = \Sigma^{-1} \Psi^T \Psi \Sigma \Phi. \tag{20}$$

Assuming statistical convergence of the weather dataset used to compute the POD modes, any weather sample (inside and outside of the training dataset) can be quite accurately described as a linear combination of the modes contained in the $\Sigma\Phi$ matrix by the corresponding time coefficients. Thus, a reduced version of any weather sample can be obtained by truncating the number of POD modes used in the reconstruction and requiring only the temporal coefficients to be embedded on the aPCE as inputs.

IV. Proposed Methodology

The developed framework for the implementation of the polynomial chaos expansion is a two-phase system depending on several inputs coming from different data sources, as well as on multiple standalone modules that have their own inherent complexity.

The first phase is dedicated to the obtainment of the full definition of the polynomials, posed following the PC theory, that will describe the evolution of the variables characterizing an aircraft trajectory as a function of the time elapsed and of the values adopted by the identified uncertain input variables. It is then focused on implementing a data-driven approach for retrieving the set of polynomial coefficients that are best suited to describe the trajectories of interest.

Once the full definition of the polynomials is obtained, the second phase of the system is applied to use them to retrieve the set of probabilistic trajectories that can be associated to a nominal flight plan. This second phase is then dedicated to perturbing an initial flight plan with the characterized value distributions for the uncertain variables that determine the trajectory evolution as per the PC formulation.

A. *Offline polynomial fitting*

This first phase of the developed framework aims to obtain the polynomials that describe the evolution of the variables defining the aircraft trajectory. To that end, different modules have to be coordinated based on a thorough and complex aircraft trajectory prediction process. The input datasets to this first phase are the network demand data for the defined past instances, composed of flight plan, surveillance and performance data, as well as the weather information for the established timeframe. The proposed structure is illustrated schematically on Figure 3.

In this schema, it has to be mentioned that the main focus is to build the obtainment of the boxes coloured in red, which constitute the modules that will be required in the second phase. These three modules are: the polynomials that will allow to describe the aircraft trajectory variables as a function of the possible values of the defined uncertain input parameters, as described in Section III-C; the module for integrating the retrieved weather information in order to consider the weather conditions as a potential source for uncertainty affecting the trajectory, as explained in Section III-D, and the uncertainty quantification process for the uncertain trajectory variables, as detailed in Section II.

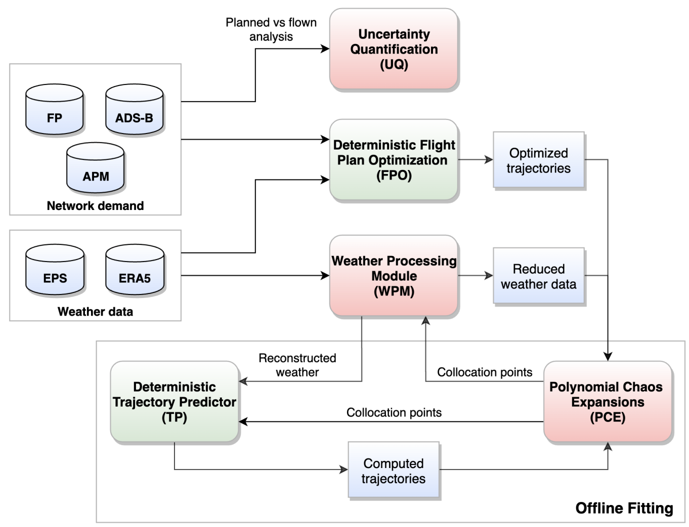

Fig. 3. Schema of the offline framework proposed for the fitting of the polynomials

Additional modules of this phase, depicted in Figure 3 with green boxes, are the tools employed for deterministic Flight Plan Optimization (FPO) and deterministic Trajectory Prediction (TP). The former will allow for the optimization of the trajectories to be followed starting from an initial flight plan, so that the dataset feeding the process for calculating the collocation points required to fit the aPCE polynomials is constituted of optimized trajectories. The latter allows for the calculation of the trajectories that are needed in order to fit the coefficients of the aPCE polynomials. Both activities are executed using a tailored implementation of the DYNAMO software [19].

B. *Probabilistic trajectory generation*

The built framework for the second phase is to be used in the tactical phase, because the modules deployed within it are prepared to consume network demand information as it is issued. Upon receiving this updated information, it will be able to compute the set of probabilistic trajectories that can be incurred when following the filed flight plans, taking

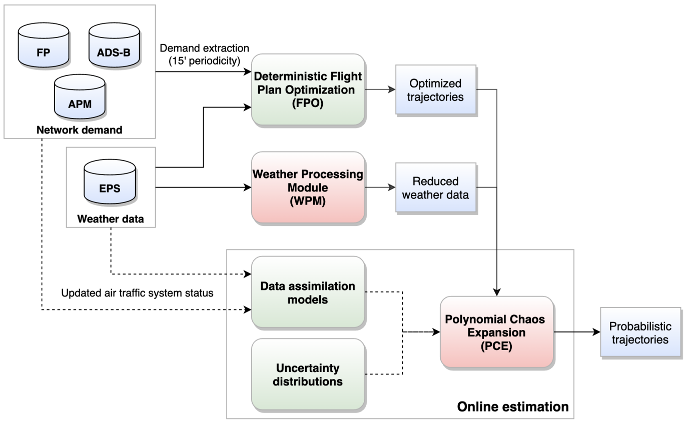

Fig. 4. Schema of the online framework proposed for the generation of probabilistic trajectories

into consideration the potential perturbations that it may suffer along its execution.

The schema for the proposed methodology to be followed is represented in Figure 4. It can be observed that both the weather processing module and the polynomials, retrieved within the first phase of the framework, constitute an integral part of the process. Regarding the input data, the information is to be consumed through a live feed, so that current data relevant for the short-term future is expected regarding the scenario of interest.

Three additional modules, represented in green in Figure 4, are required. The first one is the module for executing the deterministic optimization of the flight plans that are gathered in the tactical phase, built as well using a DYNAMO implementation. Also, the probability distributions for the uncertain input parameters will be integrated in order to retrieve the necessary perturbation values on which the polynomials depend. This is the key step that allows the deployed methodology to account for the potential deviations that the flight may suffer with respect to the declared plan, as observed in similar flights in the past. Finally, the data assimilation models proposed for the initial conditions, as described in Section II-C, are integrated in order to receive updated information about the current air traffic status that may affect the flight in its execution.

Following this process, the flight plan will be taken as initial point, consumed in its filed version from a live feed and then optimized. As a result of applying the first phase of the framework, the weather processing module and the uncertainty distributions can be employed to calculate the initial values for the input variables as calculated from the optimized flight plan. This constitutes different input cases to be fitted into the polynomials. Once fitted, the system of equations can be solved for each of the input cases to retrieve the values for the variables defining the aircraft trajectory.

Consequently, each of the input cases will lead to a different description of the final trajectory that could be described by an aircraft executing the declared flight plan. When considering all these cases together, a probability distribution of the possible values to be adopted by the aircraft trajectory variables will be obtained, and thus the required probabilistic trajectory set will be defined for the declared flight plan.

V. Study case

The objective of this study case is to exemplify the framework proposed in order to obtain a probability distribution of values for any given aircraft trajectory variable as a result of the consideration of uncertainty in the identified set of input variables affecting the trajectory prediction process. The use case developed during this section will therefore comment on all the steps to be followed for both phases of the framework.

The flight time will be the trajectory variable under evaluation, specifically at the end point of the trajectory. It is selected because it is the trajectory variable that allows for a straightforward comparison to similar values obtained by using the trajectory prediction tool or by analyzing the flown trajectories as per the surveillance data.

The data sources sources to be employed are specific to each data piece. Flight plans will be gathered from DDR2 in the ALLFT+ format, in their filed version. Surveillance data are consumed in historical instances and live feed from Flightradar24. Aircraft performance models employed are EUROCONTROL's BADA4. Regarding weather information, the ERA5 reanalysis datasets from ECMWF will be employed.

The proposed study case will consider all flights covering the route between Adolfo-Suárez Madrid Barajas airport (LEMD) and Franz Josef Strauss Munich International Airport (EDDM) during June 2018, for all present airlines and aircraft types.

In order to obtain the probability distribution of possible flight times for the trajectories that follow the considered declared flight plans, the process described by schemas in Figures 3 and 4 will be employed.

The process starts by gathering flight plan information for a relevant timeframe regarding the scenario of interest. For this purpose, 2731 flight plans declared for the flights between LEMD and EDDM were collected and optimized for the period between June 2017 and May 2018. These constitute the baseline for the calculation of the collocation points for all the considered uncertain aircraft intent variables.

The calculated collocation points for the proposed polynomial chaos expansions of order 3 are shown in Table III. Considering (5), and taking into account that 8 different aircraft intent variables are taken into account and that the selected degree of the polynomial d is 3, a total of 165 combinations of values are obtained. Then, the trajectory prediction tool will compute the trajectories following each of the input variables' combinations in order to have the required elements to perform the polynomials fitting. Once these are obtained, the coefficients of the polynomials can be calculated following (16).

Then, it is required to execute the uncertainty quantification associated to the considered aircraft intent variables to define the perturbations to be introduced. As explained in Section II,

TABLE III

COLLOCATION POINTS FOR INPUT AIRCRAFT INTENT VARIABLES FOR ALL LEMD-EDDM TRAJECTORIES BETWEEN JUNE 2017 AND MAY 2018.

$CAS_{climb}[kts]$	141.15	137.01	132.59	129.21
M_{climb}	0.764	0.756	0.745	0.733
$H_{TOC}[ft]$	41,006	40,765	38,990	37,001
M_{cruise}	0.769	0.762	0.753	0.742
$H_{cruise}[ft]$	41,521	41,006	38,990	37,001
$H_{TOD}[ft]$	41,005	39,801	38,991	37,001
M_{desc}	0.759	0.751	0.739	0.728
$CAS_{desc}[kts]$	137.03	133.93	131.04	128.63

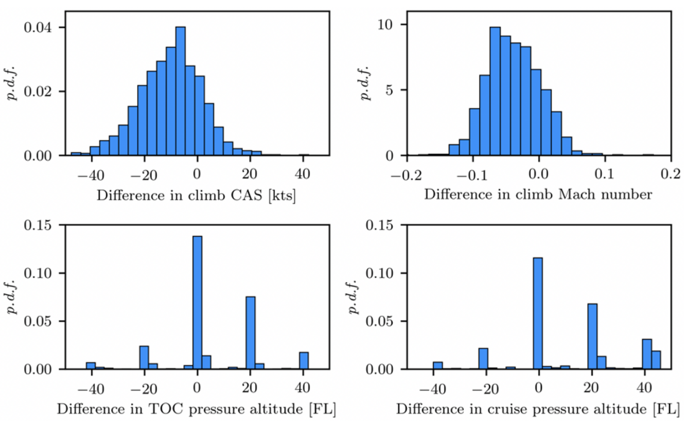

Fig. 5. Probability distributions of the differences between real and planned values for the first block of aircraft intent variables

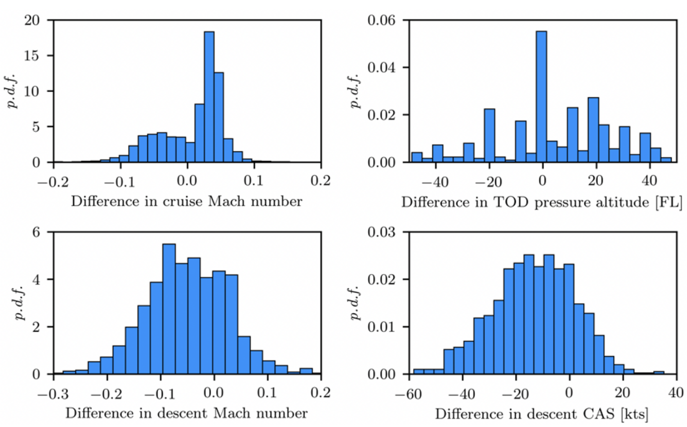

Fig. 6. Probability distributions of the differences between real and planned values for the second block of aircraft intent variables

TABLE IV

SELECTED DEVIATIONS FOR INPUT AIRCRAFT INTENT VARIABLES

$CAS_{climb}[kts]$	± 19.22	$H_{cruise}[ft]$	$\pm 2,000$
M_{climb}	± 0.051	$H_{TOD}[ft]$	$\pm 2,000$
$H_{TOC}[ft]$	$\pm 2,000$	M_{desc}	± 0.307
M_{cruise}	± 0.056	$CAS_{desc}[kts]$	± 62.13

a comparison-based analysis is implemented finding the differences between the planned and flown trajectories. The planned values are retrieved from the filed flight plan information, while the actual values are retrieved from the surveillance data for all flights between LEMD and EDDM during the considered period (June 2017 to May 2018). Executing this analysis, it is possible to build to distributions of the deviations between planned and flown trajectories, and consequently provide a certain insight on the dimension of the uncertainty to be expected on the selected variables. These distributions are presented for the selected eight uncertain aircraft intent variables on Figures 5 and 6. Knowing the range of values that can be observed, the deviations to be expected for the nominal values of the aircraft intent variables can be defined. The selected perturbations for this study case are introduced in Table III based on the observed deviations.

To include weather information into the aPCE polynomials as a few input variables, two different compressing methodologies have been evaluated: POD and CNN-AE. Both approaches have been implemented with the goal of encoding three-dimensional fields of temperature and wind velocity components into 3 latent variables. Figure 7 shows the low-order reconstruction provided by POD approach. A visual inspection makes it clear that the large-scale structures populating atmospheric flows are well captured by the reconstructions. However, the predictions show attenuated values of these temperatures close to the peak. This result is expected since the truncation of POD modes is removing energy from the reconstruction.

A dataset containing the first three POD coefficients for the flights between May 2017 and May 2018 has been included to the original dataset of aircraft intent variables, thus increasing the number of dependent variables in the aPCE polynomials up to 11. In this study case, however, the perturbations for the weather conditions were not considered using the EPS datasets due to the simpler scope proposed. Instead, the reference reconstruction of weather conditions coming from ECMWF's ERA5 datasets will be used.

In order to assess the suitability of the issued flight time estimations, the estimated flight times for the 202 flights considered during June 2018 need to be evaluated against the incurred values. Therefore, a comparison of the flight times issued against the real flight times is necessary to understand how well the developed methodology replicates the real conditions. Figure 8 illustrates this comparison for the estimations issued with the built framework, showing the error incurred for each estimated flight time of each operation. As it can be observed, results are quite satisfactory, showing a normal distribution of the error centred around zero, so a large proportion of the estimations issued are accurate with respect to reality. The inaccurate estimations may be traced to cases in which the perturbations included for some of the uncertain aircraft intent variables or for the weather conditions deviate too much the trajectory from the normal course, and therefore lead to a final estimation of the flight time that is off with respect to the incurred one.

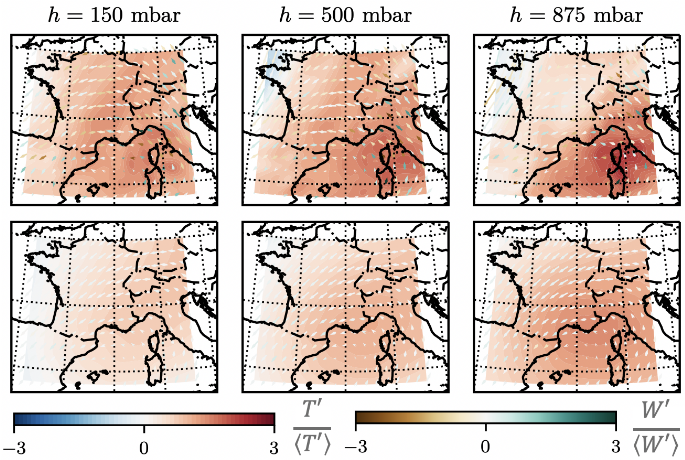

Fig. 7. Results for low-order reconstruction of weather data at three different pressure levels for the POD method. Top row refers to reference data, while bottom one refers to POD low-order reconstructions. Contour plot refers to temperature fluctuations with respect to the mean temperature at each level, while arrows heading and length refer to the direction and intensity of the wind component in the Earth-surface-parallel directions. Arrow colour denotes the magnitude of the wind in the Earth-surface-normal direction. All quantities are scaled with their corresponding standard deviation.

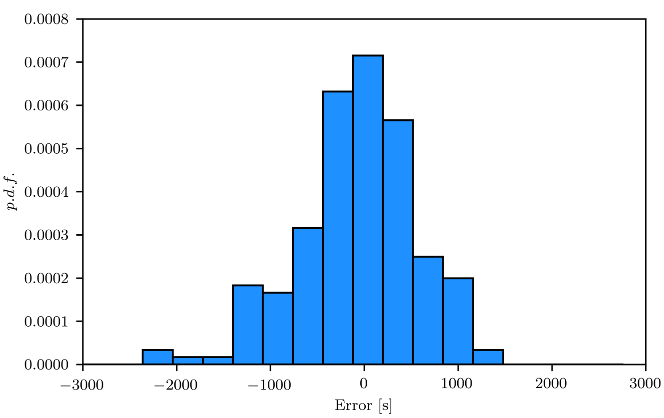

Fig. 8. Probability density function of the error incurred in the flight time estimation with the aircraft intent set of input variables together with weather variables

Thus, the proposed study case shows how, when applying the framework to a relevant scenario within the European air traffic, the results obtained for estimating the probability distribution of the flight times resembles the actual values observed in reality. This framework can be extended to estimate the associated values for other trajectory variables such as the 3D position or the speed.

References

[1] S. Sankararaman, M. Daigle, "Uncertainty Quantification in Trajectory Prediction for Aircraft Operations" AIAA 2017-1724, Session ATM and Operations, 2017.

[2] E. Casado, "Trajectory prediction uncertainty modelling for Air Traffic Management" PhD Thesis, University of Glasgow, 2016.

[3] E. Casado, M. La Civita, M. Vilaplana, M. Uzun, E. Koyuncu, G. Inalhan, "Uncertainty Quantification in Trajectory Prediction for Aircraft Operations" in IFAC, 51(9), 2018, pp. 162–167.

[4] S. Sankararaman, S. Mahadevan, "Likelihood-based representation of epistemic uncertainty due to sparse point data and/or interval data" in Reliability Engineering and System Safety, 96(7), 2011, pp. 814–824.

[5] T. Zeh, J. Rosenow, R. Alligier, H. Fricke, "Prediction of the propagation of trajectory uncertainty for climbing aircraft" in Proceedings of the 39th AIAA/IEEE Digital Avionics Systems Conference, 2020.

[6] E. Crisostomi, A. Lecchini-Visintini, J. Maciejowski, "Combining Monte Carlo and worst-case methods for trajectory prediction in air traffic control: a case study" in Journal of Guidance, Control and Dynamics, submitted, 2008.

[7] L. Lymperopoulos, J. Lygeros, "Sequential Monte Carlo methods for multi-aircraft trajectory prediction in air traffic management" in Journal of Adaptive Control and Signal Processing, 24(10), 2010, pp. 830–849.

[8] E. Casado, M. La Civita, M. Vilaplana, E.W. McGookin, "Quantification of aircraft trajectory prediction uncertainty using polynomial chaos expansions" in Proceedings of the 36th IEEE/AIAA Digital Avionics Systems Conference (DASC), 2017.

[9] SESAR Exploratory Research, "Final Project Results Report. D5.1 COPTRA" H2020-SESAR-2015-1, 2018.

[10] A. Warren, "Trajectory Prediction Concepts for Next Generation Air Traffic Management" 3rd USA/Europe Air Traffic Management R&D Seminar, 2000.

[11] M. Vilaplana, E. Gallo, F. Navarro, "Towards a Formal Language for the Common Description of Aircraft Intent" IEEE, 2005.

[12] L. Lopez Leones, M. Polaina Morales, P. Sánchez Escalonilla, D. Ferrer Herrer, M. Sanz Bravo, F. Celorrio Camara, "ADS-B based Air Traffic Performance Assessment: Develop New Metrics for Measuring ANSPs and AUs Flight Efficiency". 7th EASN Conference, 2017.

[13] EUROCONTROL, "User Manual for the Base of Aircraft Data (BADA) Revision 4.20" EEC Technical/Scientific Report No. 12/11/22-58, 2016.

[14] J. Lopez Leones, M. Polaina Morales, P. Sánchez Escalonilla, D. Ferrer Herrer, M. Sanz Bravo, F. Celorrio Camara, A. Martínez Mateo, B. Mac Namee, S. Wang, A. Grover, "Advanced Flight Efficiency Key Performance Indicators to support Air Traffic Analytics: Assessment of European flight efficiency using ADS-B data" in Proceedings 37th AIAA/IEEE Digital Avionics Systems Conference (DASC), 2018.

[15] A. M. Hernandez, E. Casado, A. Gracia, "Data-driven aircraft trajectory predictions using ensemble meta-estimators" in Proceedings 37th AIAA/IEEE Digital Avionics Systems Conference (DASC), 2018, pp. 1–10.

[16] D. González-Arribas, M. Soler, M. Sanjurjo-Rivo, "Robust aircraft trajectory planning under wind uncertainty using optimal control" in Journal of Guidance, Control, and Dynamics, 41(3), 2018, pp. 673–688.

[17] J. Rudnyk, J. Ellerbroek, J. Hooekstra, "Trajectory Prediction Sensitivity Analysis Using Monte Carlo Simulations Based on Inputs' Distributions" AIAA Journal of Air Transportation, 2019.

[18] E. Casado, M. La Civita, M. Uzun, E. Koyuncu, G. Inalhan, "Estimated Time of Arrival Sensitivity to Aircraft Intent Uncertainty" in 15th IFAC Symposium on Control in Transportation Systems, 51(9), 2018, pp. 162–167.

[19] R. Dalmau, M. Melgosa, S. Vilardaga, X. Prats, "A fast and flexible aircraft trajectory predictor and optimiser for ATM research application" in Proceedings 8th International Congress on Research in Air Transportation (ICRAT), 2018.

[20] N. Wiener, "The homogeneous chaos" American Journal of Mathematics, 60(4), 1938, pp. 897-936.

[21] S. Oladyshkin, W. Nowak, "Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion" in Reliability Engineering & System Safety, 106, 2012, pp. 179–190.

[22] R. Ghanem, P. D. Spanos, "A stochastic Galerkin expansion for nonlinear random vibration analysis" in Probabilistic Engineering Mechanics, 8(3-4), 1993, pp. 255–264.

[23] H. Li, D. Zhang, "Probabilistic collocation method for flow in porous media: Comparisons with other stochastic methods" in Water Resources Research, 43(9), 2007.

[24] M. D. Webster, M.A. Tatang, G.J. McRae, "Application of the probabilistic collocation method for an uncertainty analysis of a simple ocean model", 1996.

[25] T. Crestaux, O. Le Maıtre, J.M. Martinez, "Polynomial chaos expansion for sensitivity analysis" in Reliability Engineering & System Safety, 94(7), 2009, pp. 1161–1172.

[26] L. Sirovich, "Turbulence and the dynamics of coherent structures" in Quarterly of applied mathematics, 45(3), 1987, pp. 561–571.