Preprints.org

Article

Building Credible VTOL Flight Models for Handling Qualities Certification by Simulation

Lorenzo Favaro [‡] , Agata Rylko [‡] , Giuseppe Quaranta [*]

Article

Building Credible VTOL Flight Models for Handling Qualities Certification by Simulation

Lorenzo Favaro ‡, Agata Rylko ‡ and Giuseppe Quaranta *

Dipartimento di Scienze e Tecnologie Aerospaziali, Politecnico di Milano

* Correspondence: giuseppe.quaranta@polimi.it; Tel.: +39-02-2399-8405 (G. Q.)

‡ These authors contributed equally to this work.

Abstract: Certifying novel VTOL aircraft handling qualities (HQ) may be challenging, relying on costly and high-risk flight testing. This paper presents a methodology to establish the credibility of flight simulation models for certification by simulation aiming to bridge the gap between the model input uncertainty and certification confidence. The core objective is to assess if an model, despite its inherent uncertainties, can reliably predict the handling qualities compliance for specific flight tasks. This is achieved by quantifying the impact of input parameter uncertainties on predicted handling qualities and, crucially, by evaluating the envelope of the resulting uncertain aircraft transfer functions — scaled by a confidence ratio —— against established maximum unnoticeable added dynamics boundaries. Applied to a lift+cruise VTOL model performing a deceleration to hover manoeuvre, the study demonstrates that while longitudinal control dynamics largely remained within MUAD limits, indicating the model credibility for those aspects, vertical axis dynamics coupled with longitudinal inputs for some uncertain configurations exceeded these limits, correlating with observed flight test performance variability. Readers will find a structured, quantitative approach to model validation for HQ certification by simulation, leveraging MUAD to determine if a nominal model is sufficiently representative for certification, thereby supporting safer and more efficient VTOL development.

Keywords: handling qualities; certification; simulation; Uncertainty Analysis; eVTOL; UAM

---

1. Introduction

Integrating vertical take-off and landing (VTOL) vehicles into urban airspace comes with enormous challenges from a design and technological standpoint, and considering the necessary certification efforts. Demonstrating compliance with the flight-related airworthiness rules set by aviation authorities is indeed one of the most critical phases of the vehicle development process: the evidence collection and compliance demonstration processes, which are primarily based on flight testing activities, are highly complex, very time demanding, and require a lot of effort to minimise safety risks while keeping the representativeness of the flight test. Certification may become even more difficult in the context of Urban Air Mobility (UAM), where the safety of eVTOL operations needs to be fully assessed and understood[1]. Therefore, appropriate means of compliance (MoC) and guidelines for the certification of this class of vehicles are still under development. In the publication of the MoC for the SC-VTOL Standard, the European Union Aviation Safety Agency (EASA) in MOC VTOL.2500(b) states that flight simulators "may also support some certification tests"[1], explicitly allowing the possibility of exploit simulators as an alternative to flight tests under specific conditions and when appropriate means are undertaken.

---

[1] See for instance the European Union Commission implementing regulation (EU) 2024/1111 of 10 April 2024 which deals with the establishment of requirements for the operation of manned aircraft with a vertical take-off and landing capability

1.1. VTOL Handling Qualities Regulatory Framework

The certification of aircraft configurations, which differs from conventional rotorcraft or fixed-wing aircraft, requires special care. EASA decided to tackle the unique characteristics of these aircraft that have lift/thrust units through the definition of a complete set of dedicated technical specifications in the form of a special condition for VTOL-capable aircraft, that culminated in the publication by EASA of *Special Condition for small-category VTOL aircraft* (SC VTOL) [2]. This standard, which descends directly from EASA's CS-23 and CS-27, is articulated in different subparts and covers a wide range of rules, ranging from stall characteristics and flight envelope definition to handling qualities (HQ) assessment. The topic of VTOLs HQ certification takes on particular importance, because adequate HQs lead to a limited pilot workload and sufficient spare capacities that provide full situational awareness and so a higher level of safety. To this aim, three specific rules of the SC-VTOL deal directly with HQs: VTOL.2135 Controllability, VTOL.2140 Control Forces, and VTOL.2145 Flying Qualities. The role of each rule and the relationship with the different aspects of the piloted flight are described in Figure 1.
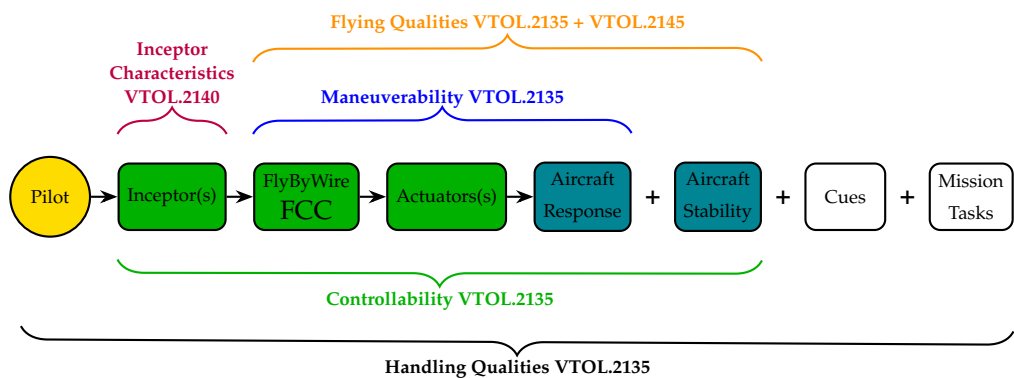

Figure 1. SC-VTOL Subpart B Controllability and Handling Qualities concerning the different aspects of piloted flight.

VTOL.2135 rule [2] explicitly focuses on controllability and manoeuvrability. Controllability links pilot control inputs to the resulting aircraft response and combines VTOL inherent properties with the mechanical characteristics of control inceptors. In quantitative terms, it measures the levels of compensation required to hold a given flight condition and the availability of control margins. On the other hand, manoeuvrability can be defined as a performance of the aircraft, representing its ability to transition from one trim condition to another. VTOL.2145 puts the accent on VTOL stability, referring to its tendency, either initial (i.e. static stability) or long-term (i.e. dynamic stability), to return to its trim condition after undergoing an external disturbance. Figure 1 shows that the aforementioned rules are not separate but have several intersection points. Stability and manoeuvrability are indeed intrinsic aircraft properties; as such, they contribute to defining the aircraft's flying qualities. On the other hand, the HQ should consider the pilot in the loop, taking into account the capability to perceive correctly the state of aircraft using the simulator cues (visual, vestibular, audible, etc.) and perform the required mission tasks without excessive workload that may hinder the performance. So, as expressed by EASA in [1], to comply with VTOL.2135 it is necessary to show that the aircraft possess a satisfactory HQ to "give the opportunity for the crew to better manage high-workload situations, and allow them to operate safely for longer periods, and to be able to deal with aircraft system failures and contingencies." To show compliance with these rules, EASA proposed as a MoC the Modified Handling Qualities Rating Method (MHQRM) [1] where the applicant must demonstrate reaching a minimum acceptable HQ rating for each flight phase and condition that has a probability of being encountered greater than 10^{-9}. The suggestion is to use the well-established Cooper-Harper Handling Qualities Rating (CHR) Scale [3], but applicants may choose other scales. The probability of a flight condition X_{FltC} is defined by a combination of individual probabilities associated with

- the **Flight Envelope** (Normal, Operational, Limit) the VTOL operates in, x_{FE};
- possible aircraft **Failure Conditions** (Nominal up to Major, Hazardous), x_{FC};
- the level of **Atmospheric Disturbances** (AD) acting on the vehicle, from total absence of AD up to the ones exciting structural limits (Light, Moderate, Severe), x_{AD}.

EASA drafted standardised tables to help the applicant define individual terms. To obtain the probability, under the assumption of independence of the three factors, these should be combined as follows

$$x_{FltC} = x_{FE} \cdot x_{FC} \cdot x_{AD}. \tag{1}$$

Should a clear correlation between the factors exist, the expression shall be modified accordingly.

Therefore, to certify an aircraft according to VTOL.2135, collected evidence based on MHQRM must cover the entire aircraft flight envelope, accounting for failure events and atmospheric disturbances. EUROCAE ED-295 [4] provides some guidance on how to collect the evidence required by the MHQRM. Inspired by the concept of Mission Task Element (MTE) of Ref. [5], ED-295 developed a set of guidelines specifically tailored for VTOLs in the form of Flight Test Manoeuvres (FTM) to be performed with the pilot-in-the-loop. For each FTM the document proposes a set of limits, in terms of desired or adequate performance, that are quantitative measures, specifically tailored to the FTM under study, and aimed at identifying performance deficiencies . Furthermore, a subjective CHR – provided by properly trained test pilots – must follow; it is hence a pilot-assigned evaluation, similarly to Ref. [5] where it is defined as *Assigned HQs*. MHQRM [1] reinterprets the CHR, presenting three possible HQ levels: Satisfactory, Adequate, and Controllable (Table 1). The ultimate purpose of FTMs is to quantify HQs according to the level of pilot workload considered appropriate for the intended task. As ED-295 highlights, FTMs work on a dual level: first of all, they stress aircraft's capabilities, looking for possible Flight Control System deficiencies that, combined with an excessive workload request, could severely challenge the pilot ("*without exceptional piloting skills*"). The performance parameters serve to validate the pilot ranking, as they should correspond to the CHR assigned by the pilot. Secondly, a set of CS-based guidelines ensures repeatability of the tasks and direct comparison between the feedback of different pilots.

Table 1. HQ Ratings definition [1].

HQ Rating	Description	FC	CHR
Satisfactory (SAT)	HQ allow achievement of **desired performance criteria without exceptional piloting skills and with no or minimal pilot compensation**	Up to Minor	1–3
Adequate (ADQ)	HQ allow achievement of **desired performance criteria** or **adequate performance criteria** without exceptional piloting skills and **with moderate to extensive pilot compensation**	Major	4–6
Controllable (CON)	HQ DO NOT allow achievement of adequate performance criteria WITHOUT exceptional piloting skills. Allows, however, **continued safe flight and landing**, **without exceptional piloting skills**, after a transient condition or reconfiguration to retain control, if necessary	Hazardous	7–9

To apply the MHQRM approach, a large number of flight tests may be required, with many tests performed in challenging (e.g., outside normal flight envelope or with failures) or difficult to repeat (e.g., atmospheric disturbance) conditions. To reduce the risk, time, and cost of flight test activities for VTOL aircraft, in some scenarios it may be possible to proceed with a demonstration using flight simulators, as suggested by EASA, following an approach called "Certification by Analysis". However, appropriate Flight Simulation Models (FSM) must be developed first, so it is necessary to provide a reliable methodology that leads to such models.

1.2. Certification by Simulation Approach

With the term Certification by Simulation (CbS), we refer to all cases where flight simulation can be used to support, augment or replace flight testing in the demonstration of certification compliance "without sacrificing the level of safety" [6]. This approach, already accepted for helicopters by certification authorities in the past on a case-by-case basis [7,8], has been made systematic and more structured for the certification of rotorcraft under the EASA CS-27 and CS-29 in Ref. [6]. The process presented in Ref. [6] is a structured, requirement-based approach that starts from the definition of the Applicable Certification Rule (ACR) (Phase 1 in Figure 2) and then moves to the subsequent simulation-related phases that allow one to build up the simulation ecosystem composed by the Flight Simulation Model (FSM), the Flight Simulator (FS) the and Flight test activities developed to collect validation data for the FSM and the FS (Phase 2 in Figure 2). In the development of FSMs it is crucial to evaluate the error in comparison with tests, i.e. validate the models, showing that they have a sufficient level of fidelity that translates into an error that is below a certain set of tolerances agreed with the certification authority. However, for certification activities, this is typically not enough, because the models can be used in conditions that cannot be compared with equivalent tests, so it is also necessary to provide evidence of the credibility of the models. The concept of credibility of the predictions made by simulations is somewhat elusive, but it expresses the level of confidence that the users of the FSM have in the fact that the results obtained with the FSM reflect the behavior of the real aircraft.
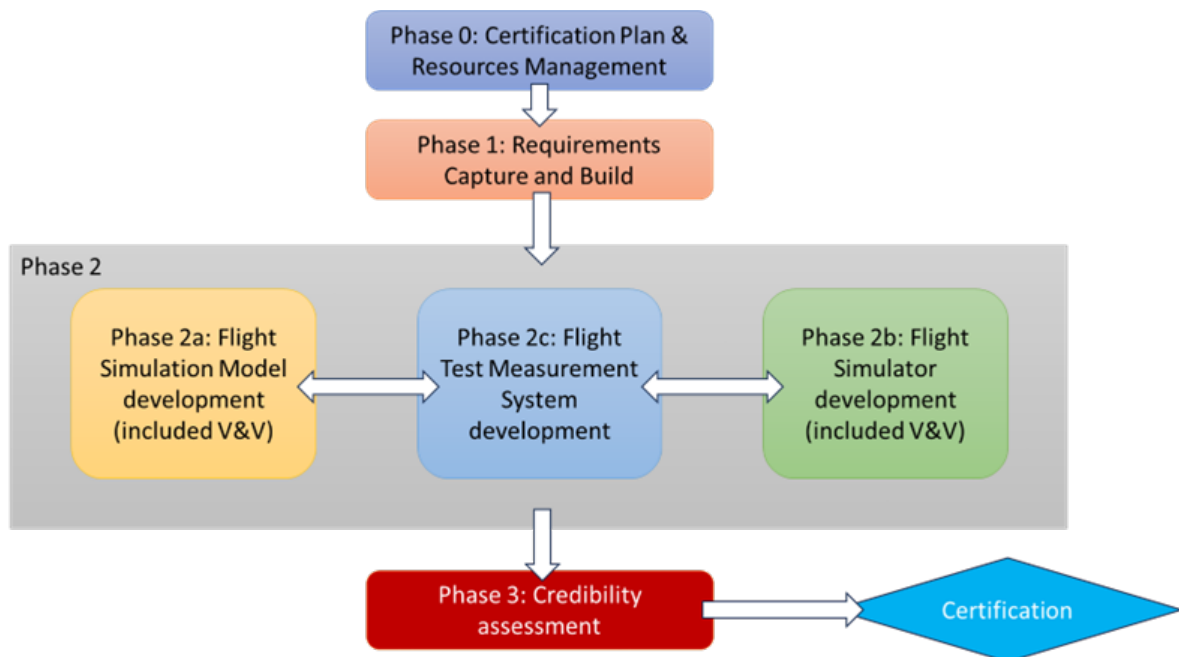
Figure 2. Overall structure of the Certification by Simulation Process, as from [6]

In Ref. [6] the credibility is managed by the definition of a *confidence ratio CR* that is the ratio between the value of the distance between the performance obtained by the FSM and the performance limit allowed for the certification, called the margin M, and the uncertainty U in the prediction of that performance:

$$CR = \frac{M}{U} \tag{2}$$

where the uncertainty U collects the experimental uncertainty, the uncertainty of input data and parameters of the model, and the numerical uncertainty associated with the method used to solve the mathematical models.

Basically, the users accept that the higher the risk of using the model for certification, defined through the combination of the influence level and the predictability level (see Ref. [6] for more details), the higher the confidence ratio CR must be (in any case greater than 1). In this way, asking for a margin

$M = CR \cdot U$ to be significantly higher than the uncertainty, it is possible to ensure that the predictions have a higher degree of credibility.

The concept of a confidence ratio allows us to keep the credibility of models at the required level; however, it applies only to situations where there is a quantified requirement. In case of ACRs that rely on subjective pilot evaluations based on discrete – and often nonlinear [9] – rating scales, such as the CHR, other criteria must be defined to assess credibility. Since this is exactly the case of HQs certification for VTOLs, this paper will try to extend the concept of CR proposed in Ref. [6] to this case.

1.3. Proposed Approach for Credibility Assessment of a Flight Simulation Model

The first step of a credibility assessment is always composed of the validation and fidelity assessment of the FSM. In principle, since the ACRs in this case require performing a pilot-assigned evaluation, i.e., assigned HQ tests, it is the fidelity of these evaluations that should be the objective of the fidelity assessment and uncertainty quantification. However, we must acknowledge that any quantity measured during these tests will be affected not only by the FSM but also by the FS characteristics and by the test pilot characteristics. As such, it will not represent a measure of the FSM quality. A different approach could be defined by making use of the concept of *Predicted HQ* defined in Ref. [5]. The predicted HQs (pHQ) are obtained through an open-loop flight test and therefore relate to the intrinsic flying qualities of the aircraft. They are chosen as a way to anticipate how an aircraft will behave in terms of HQ, and in Ref. [5] is used as support for aircraft design, as it is possible to show that helicopters that have certain pHQs are expected to reach a certain level of Handling Quality Rating (HQR) in the aHQ tests. Similar behavior may be expected for VTOLs. As a consequence, it is possible to say that a FSM of a VTOL that shows pHQs sufficiently close to those measured during flight tests is a faithful reproduction of the real aircraft. At the same time, it is possible to assess the level of uncertainty of the FSM by looking at the uncertainty that results in the pHQs.

However, once the uncertainty on the pHQ of the FSM is quantified, it is necessary to understand how it translates into an uncertainty for the aHQs and how the confidence ratio could be taken into account. The process we suggest here is as follows:

- select among the uncertain configurations those that lead to the worst pHQs.
- Compute the transfer functions for these worst configurations (wTFs) and compare them with the nominal transfer function (TF) obtained with the reference model.
- Define new transfer functions that are the nominal transfer function plus the difference between the WTFs and the nominal one, multiplied by the confidence ratio.

If the difference between the nominal TF and the wTFs multiplied by the CR stays within the range of the mismatches that proved unnoticeable to pilots, defined here [10] as MUAD (Maximum Unnoticeable Added Dynamics), it is possible to assume that the results obtained by using the nominal FSM must lead to assigned HQ ratings that have a sufficient degree of credibility. If this is not the case, i.e. the wTFs are beyond the MUAD limits, the model developer has two options:

- Improve the models to reduce the uncertainty and/or reduce the uncertainty in the flight test measures.
- Perform several flight simulator test campaigns using nominal and worst-case models to quantify the impact of uncertainties on the rating for the aHQ.

Of these two options, the former should be the preferred option whenever possible. The latter, in fact, may require a large number of simulated flight tests.

To verify whether this approach is feasible, we will start in this work investigating the relationship that exists between the uncertain models that present the worst pHQs and the corresponding assessment of the aHQ. To avoid blurring the evaluation with subjective pilot evaluations, it has been decided to apply the approach to a drone model in which pilots are replaced by a virtual system that controls the aircraft. So, no flight simulator tests are presented here. The paper is organised as follows: after presenting the aircraft model, the uncertainty quantification analysis is presented to identify the relationship between the uncertainty of the input data and the resulting pHQ. Then, a first set of

assigned HQ tests is presented, considering both the reference model and several modified models selected to represent the worst configurations, followed by an analysis of how the transfer functions of the worst cases compare with the MUAD boundaries. Finally, some conclusions are drawn.

2. VTOL Aircraft Description

VTOL aircraft present a wide variety of configurations. In this paper, it has been chosen to investigate the approach to credibility of a flight simulation model using an aircraft in the *lift+cruise* configuration. Vertical lift is provided by variable rpm propellers that operate during take-off and landing, and low-speed manoeuvres, while forward thrust is generated by separated propellers during cruise flight. EASA developed SC-VTOL for passenger transport with a maximum take-off mass of 3175 kg [2] (up to 9 passengers). However, in the context of this research, it has been decided to perform the certification exercise on a laboratory small-scale prototype to allow for a direct comparison with data from experimental tests.

The aircraft under analysis is a small-scale unmanned eVTOL, designed, manufactured and tested in flight by the Aerospace Systems and Control Laboratory (ASCL) of the Department of Aerospace Science and Technology at the Politecnico di Milano (DAER) [11–13] (depicted in Figure 3). It has two distinct sets of rigid propellers for thrust generation: eight fixed-pitch, variable-RPM three-bladed propellers, are responsible for controlling the vehicle in helicopter mode (using differential thrust), while a couple of two-bladed propellers produce forward-flight thrust in aeroplane mode. The baseline structure is of a fixed-wing UAV which features a small-sized fuselage and two rectangular wings with zero built-in twist. The main geometry and performance characteristics are detailed in Table 2.
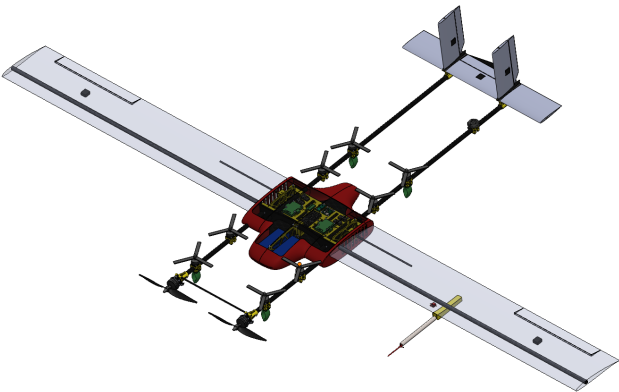
Figure 3. eVTOL CAD model overview

Table 2. eVTOL main design and performance specifications.

Parameter	Value	Unit
Cruise speed	15	m/s
Stall speed	11	m/s
Maximum speed	22	m/s
Max Gross Weight	7	kg
Wing and fuselage span	3.12	m
Wing chord	0.25	m
Wing incidence	4.5	deg
Horizontal stabilizer span	0.721	m
Horizontal stabilizer chord	0.13	m
Horizontal stabilizer arm	1.08	m
Vertical propeller diameter	0.178	m
Vertical propeller chord at 80% radius	0.0124	m

2.1. Flight Simulation Model Architecture

The FSM used is based on a Simulink-based simulation platform, which can cover the entire VTOL flight envelope. The simulator is built using a component-based *stitched quasi-Linear Parameter Varying* (qLPV) modelling approach [14,15] and is composed of an aircraft dynamics block and a control unit.

The flight dynamics block simulates VTOL's 6DOFs rigid-body dynamics by integrating the nonlinear equations of motion, taking into account aerodynamic, propulsive, and gravitational contributions. The aerodynamic forces of the airframe are computed based on numerical data [16]. After computing perturbations as the difference between the current trim states (given by aerodynamic angles α and β) and the interpolated ones, these are multiplied by the interpolated aerodynamic coefficients and stability derivatives: the resulting information can be exploited to compute the perturbed overall aerodynamic forces and moments . The FSM does not take into account stall or propeller-airframe interaction effects. For the former, it is always possible to add a post-processing check that stall angles are never reached during simulations, as suggested in Ref. [6]. The latter has been the objective of the analysis presented in [17], and its impact at low speed is very limited. Propulsive terms computation follows an analogous path: employing the airspeed and the throttle percentage command coming from the Control Unit for each rotor, the angular speed, thrust and torque are obtained from multidimensional look-up tables. After applying mixer matrix conversion, multicopter body-axes thrust and moments are available. The look-up tables are obtained as results of experimental tests in the wind tunnel [12]. The local dynamics of the electrical motors which are connected to the propellers is not modelled.

The Control Unit is structured into a cascade control arrangement made of four loops: body angular rates, attitude angles, inertial velocity and inertial position; while velocities and rates loops use full PID controllers, position and attitude rely on proportional ones. The output of the control unit is the required throttle command. By activating or deactivating the control loops, different levels of assistance can be provided to the pilot. The standard setting for the unmanned VTOL is *Off-board Mode*; by setting the position in the NED frame and the yaw angle, this mode allows automatic control from the ground station. However, to simulate the presence of a virtual pilot within the control loop for certification purposes, the eVTOL is operated here in *Position Mode*, using longitudinal and lateral control input to prescribe the ground speeds, V_N and V_E, respectively, vertical control input to prescribe the vertical velocity V_U, and a last control input to prescribe the yaw attitude.

2.2. Applicable Certification Requirement

The CbS approach presented in Ref. [6] is requirement-based. As such, any analysis of the fidelity of the model and its credibility should always begin by capturing the requirements and objectives of the analysis from the knowledge of the specific applicable certification specifications.

Given the width of the class of VTOLs currently under development, the Performance Standards and the intended manoeuvre should be consistent with the aircraft's specific features. To tailor the prescribed requirements to the specific vehicle , ED-295 [4] conceives its performance standards in a parametric manner, through a scaling technique which defines the expected performance and precision to perform the task based on the eVTOL size. The scaling process is formulated using the *D-value*, which "*represents the diameter of the smallest circle enclosing the VTOL aircraft projection on a horizontal plane [...]*" a quantity defined in the SC-VTOL [2]; in this work, given the size of the VTOL (see Table 2), a value of 3 m has been selected to scale the test course task and requirements. For this exercise, it has been decided to focus on a symmetric manoeuvre performed at low speed, so the chosen FTM is the *Deceleration to Hover*, a manoeuvre representative of the final approach to landing that a VTOL is typically expected to perform and during which good HQ should be considered essential for a safe flight. The test course described in [4] is structured as follows:

- The starting point is a stabilized hovering flight condition between the "minimum height" h_1 and a maximum of 40 ft ($\approx$ 12 m) over a ground reference point, heading towards the longitudinal

direction along which the manoeuvre is to be performed; since for the eVTOL under study h_1 is not defined, the hovering height is set to 12 m, so that the aircraft is outside any ground effect.

- A longitudinal acceleration – as aggressive as possible – must be applied to reach the "*maximum recommended speed in the low-speed envelope or a maximum of 40 kn, whatever is lower*", here set to $\Delta V_{\mathrm{lon}} = 8$ m/s.
- The planned forward speed must be maintained for 2 s before initiating a high-attack-rate deceleration to hover. The starting point for deceleration is provided to the pilot by a ground reference, closing the manoeuvre within a distance of $0.6D$ from the intended endpoint, without any allowable overshoot.
- In the end, a stabilised hovering flight must be maintained for at least 5 s.

Figure 4 shows the test course setup. A key aspect is represented by the course length: ED-295 prescribes a standard distance of $12.5D$, from which an allowable difference of $0.6D$ is subtracted to define the boundaries of the desired arrival point. However, the document clarifies that "*the course length [...] should be long enough to ensure that a sustained translation at desired forward speed can be achieved prior to the deceleration. If the suggested* $12.5D$ *is not sufficient, the course length should be increased accordingly*". Given the relatively slow response of the eVTOL to the acceleration command, the course length is here increased to 53.2 m ($\approx 17.7D$), as a trade-off between the aggressiveness of the manoeuvre (which requires a shortened length) and the necessity to maintain forward speed for a prolonged time, ultimately ensuring that the manoeuvre is sufficiently aggressive to challenge the aircraft without compromising the FTM.
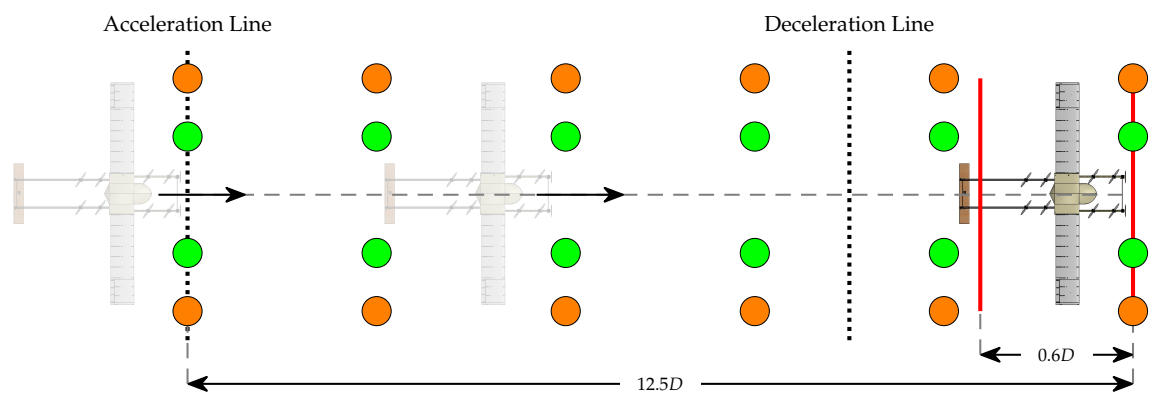
Figure 4. Deceleration to Hover Test Course (adapted from Reference [4])

According to ED-295 [4], deviations of up to 10 ft (≈ 3 m) are allowed to meet desired performance, which appears too loose in this case, considering the small size of the eVTOL. Given that this is the only requirement not affected by the scaling procedure, the altitude limits are here revised, reducing the threshold to 3 ft (≈ 1 m) for desired and 6 ft (≈ 2 m) for adequate performance. So, in summary, to evaluate the success of the FTM and the level of performance standards achieved, the requirements in Table 3 hold.

Table 3. Deceleration to Hover FTM Performance Criteria [4].

Description	Desired	Adequate
Maintain lateral track within $\pm X$ meters from ground reference line	$\pm 0.3D$	$\pm 0.6D$
Maintain the heading within $\pm X$ degrees	± 10 deg	± 15 deg
Maintain radar altitude within $\pm X$ feet	3 ft	6 ft
The aircraft must be brought to a controlled hover within X meters of the intended endpoint. Overshoots are not permitted	$0.6D$	$0.6D$

3. Uncertainty Analysis Framework

As described in Ref. [6], validation and verification activities during FM development should eliminate or reduce all known error sources to acceptable levels, that is, below the sufficiency threshold established in agreement with the certification authority. What remains will be by it's nature not known and can be quantified in terms of uncertainty. This uncertainty in turn will provide a tool to assess the credibility of the model also in the conditions where it is necessary to use the FSM to perform a prediction because no flight data are available to perform a validation. The total uncertainty, used in the CR expression eq.(2), is considered the combination of three sources (see [6]):

- uncertainty due to numerical errors u_{num};
- uncertainty due to experimental errors (present only in the validation cases where the model is compared with experiments) u_r;
- uncertainty due to input errors u_{inp}, i.e. due to an erroneous knowledge of the parameters used as input for the FSM.

In the case analysed in this paper, then numerical models should be considered verified up to a level where the numerical error has been brought to values at least an order of magnitude smaller than the other sources of error. At the same time, we will not investigate the experimental error, since we do not have experimental data to compare with, but we will concentrate on the assessment of the input errors, making the hypothesis that this part is the determinant component of the total error. At this stage, we will make the hypothesis that all the uncertainties associated with input parameters are epistemic and not aleatoric. In fact, most of them are the results of lack of knowledge, due also to limited chances to have accurate experimental methodologies to measure them, more than a true intimate aleatoric nature, that may be presented but often to a lower extent. Additionally, even for the cases where the aleatoric nature may appear as the best description, often we lack experimental measures capable of capturing the appropriate probabilistic characteristic for such parameters. In contrast, the usage of epistemic uncertainties requires only the definition of ranges of uncertainties that can be assessed in many cases using engineering judgement.

So, the uncertainty analysis presented here starts from the definition of epistemic uncertainty ranges for the most relevant parameters of the FM model and continues with the usage of efficient methodologies that may led to quantifying the impact of these combined uncertainties on set of well chosen predicted handling qualities that are considered indicative of the performance that may be expected from the eVTOL drone in performing deceleration to hover assigned handling quality test described in the previous section.

3.1. FSM-Dakota Coupling for Uncertainty Analysis

To achieve the primary HQ certification goal, the work aims at characterising the role and the implications of uncertainty on the eVTOL flight dynamics model. In this scenario, a statistical simulation tool was needed to wrap the existing FSM and allow to perform a series of Design of Experiments (DOE) analysis. The main goal of DOE is to rank the uncertain input factors and quantify to what extent each of them affects the output metrics.

Due to its flexibility and ease of use, Dakota — a state-of-the-art software for Uncertainty Quantification, model calibration, and optimisation — was the natural choice, offering a wide range of variance-based and sensitivity analysis methods for DOE analysis, along with meta-modelling capabilities for building powerful surrogate models [18]. The latter is especially well suited for integration into the eVTOL virtual simulation environment, especially for complex, time-demanding iterative analyses. Moreover, thanks to its extensible interface, Dakota can be easily coupled with external computational codes like the eVTOL FSM, communicating via script files and managing physics-based models as black boxes, where by injecting input it is possible to measure output. In principle, any analysis to propagate the uncertainty in the parameters through the model all the way to the output can be done using Monte Carlo (MC) analysis [19]. However, when the number of parameters is large, the well-known 'curse of dimensionality' may easily lead to intractable problems. Sensitivity

Analysis (SA), especially if done with approximate models that use limited resources, may be used as a tool to determine which of the input factors can be fixed anywhere in the range of variation without significantly affecting the output and which are instead primarily influential [20]. Furthermore, in cases where all input parameters are defined in terms of intervals given their epistemic nature, the SA, as long as global and not local, may be all is needed to perform a UQ analysis, since it provides the range of variability of the output [6]. Several SA techniques are available in the literature [21]. Since each method shows advantages and drawbacks a hybrid, comprehensive UQ framework is proposed and tested here [22].

3.2. Morris-One-at-a-Time Method

The first technique employed is the Morris One-at-a-time (MOAT) method [23], which allows to discern whether an input factor has negligible, linear and additive, or nonlinear interaction effect on the model output. This approach evaluates the influence of each parameter, which can take only a discrete set of values called levels, by varying them individually along predefined trajectories: a trajectory consists of $k + 1$ points, where k represents the number of model inputs. At each step along a trajectory, only one input variable is modified relative to the previous step, following a "one-at-a-time" variation approach. This process enables the computation of elementary effects (EE) as the ratio of the incremental change in the output caused by the modified input. The EE values are then averaged, yielding to two key sensitivity indices: the modified mean μ^* and the standard deviation σ of the EE distribution. A high μ^* indicates that a factor has a strong overall impact on output, while a high σ indicates the presence of either nonlinear effects in the response or significant interactions with other factors. For additional details on the meaning of μ^* and σ, the reader should refer to the comprehensive analysis presented in [20].

3.3. Variance-Based Decomposition

The MOAT method is conceptually simple and so convenient to use when the number of factors is large and the computational times to elaborate the output are significant. However,it is primarily used for preliminary SA and does not explicitly isolate potential interaction effects among the FSM inputs arising from nonlinearities. A more quantitative evaluation is provided by *Variance-based Decomposition* (VBD) methods which, as described in [20], enable ranking inputs "according to the amount of output variance that is removed when we learn the true value of a given input factor." To achieve this, it is crucial to distinguish between first-order and total-effect contributions. The first-order index serves as a measure of the importance of an individual factor and, in a purely additive model, represents the exact proportion of the total variance attributed to that factor. This is quantitatively expressed through Sobol' main-effect index S_i. However, if the total variance V of the output Y cannot be fully explained by the sum of the first-order effects alone, interactions between input variables must be considered, as follows

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>1} V_{ij} + \ldots + V_{12\ldots k} \tag{3}$$

where $V(Y)$ is the total output unconditional variance. For the factor X_i the Sobol index $S_i = \frac{V_i}{V(Y)}$. Computing high-order interaction terms $V_{12\ldots k}$ is often impractical for SA purposes. The computation of Total Effect index S_{T_i} — which aggregates both the first-order effect of a factor and all its interactive contributions — represents a valid alternative

$$S_{T_i} = S_i + \sum_{j \neq i} S_{ij} + \sum_{j \neq i} \sum_{l \neq i} S_{ijl} + \cdots \tag{4}$$

The total effect index accounts for the total contribution to the output variation due to factor i, including first-order effect and all higher-order effects due to interactions. This metric is widely recognized in the literature as a standard SA index and has shown a strong correlation with MOAT's μ^*, making it

a useful basis for comparative analysis between methods. For a detailed discussion on the different options to perform this computation, the reader is again referred to [20].

3.4. Proposed Multimodal Approach

To reduce the number of FSM executions, the problem is first tackled qualitatively, performing a preliminary analysis using MOAT, which gives a first characterization of the most influent parameters. Then, a VBD quantitative analysis follows: however, performing a full VBD directly on the FSM would require an excessively high number of MC simulations. Hence, a surrogate model is developed for each pHQ Quantity of Interest (QoI), generating an input-output dataset through Latin Hypercube Sampling (LHS), and constructing a meta-model that emulates the full behaviour of the FSM through Kriging Interpolation. By leveraging a limited set of simulations, Kriging exploits spatial correlations between sampled points to predict the value of the response function at unsampled locations. This results in an efficient computational emulator that can replace the original FSM, ultimately enabling the VBD and the computation of Sobol indices. Figure 5 presents an overview of the methodology adopted. Information about Dakota's implementation of the methods can be found in [24].
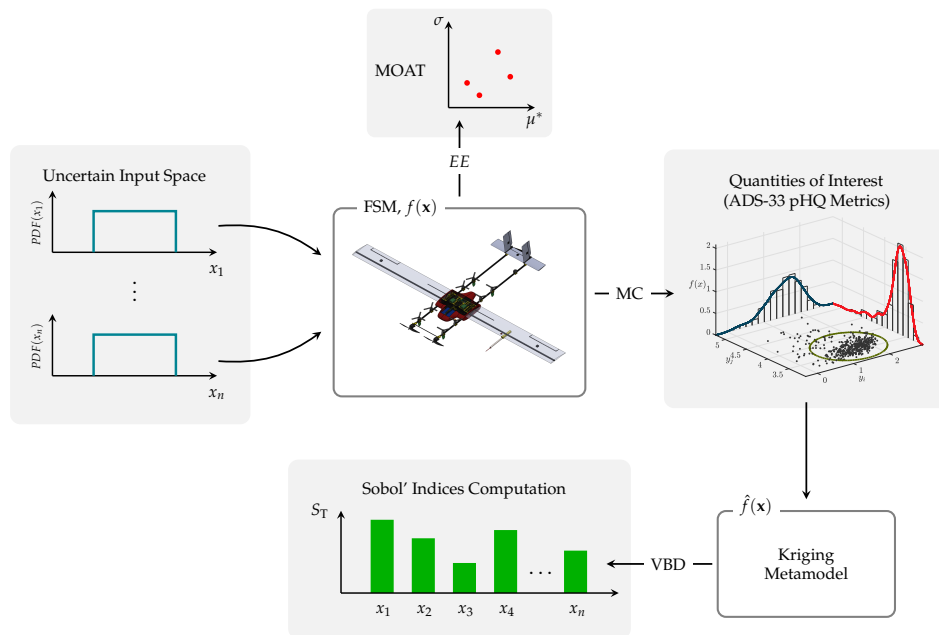
Figure 5. Uncertainty analysis workflow

4. Low-Speed Predicted HQs: Uncertainty Analysis Results

The analysis begins with the SA and UQ phases, hereafter jointly referred to as Uncertainty Analysis (UA), to investigate the relationship between the uncertainty of the FSM input data and the output u_{inp} identified through selected pHQ quantities of interest. This analysis is structured over two consecutive levels:

- by means of a preliminary SA the key parameters in driving pHQs output uncertainty are identified, providing an insight into model's dependence upon input parameters, including possible nonlinear/interactions effects arising throughout the simulations;
- relying on UQ techniques, the effect of input space uncertainties on the selected pHQs is assessed and the related output uncertainty is quantified.

Focusing on the deceleration to hover FTM, four QoI are identified in [5][2] pHQs are selected as the corresponding fidelity metrics, to characterize the longitudinal dynamics behaviour of VTOL:

---

[2]   This document is often colloquially identified as 'ADS-33' making reference to the original design standard from which it originates [25].

1.   Roll-due-to-pitch Coupling for Aggressive Agility (Section 3.3.9.2 of [5]);
2.   Moderate Amplitude Pitch Attitude Changes (Section 3.3.3 of [5]);
3.   Short-term Response to Pitch Control Inputs (Section 3.3.2.1 of [5]);
4.   Mid-term Response to Pitch Control Inputs (Section 3.3.2.3 of [5]).

In the following section, the main results of the UA are presented, highlighting the challenges encountered and the solutions proposed.

4.1. FSM Input Uncertainty Definition

The first step in the UA process is the characterization of the uncertainties related to model input parameters. Table 4 summarizes the main FSM factors to whom an uncertainty range has been assigned, chosen to ensure sufficient coverage of the different building blocks within the FSM — following its component-based nature — from design parameters to aerodynamic characteristics. Starting from roll-due-to-pitch coupling UA, no assumption on the possible decoupling between the longitudinal and lateral dynamics was made, hence generating $n = 22$ input factors, comprehensive of eVTOL's mass and inertia parameters (namely M, pitch inertia moment I_y and cross-inertia moment I_{xy}), aerodynamic coefficients and stability derivatives related to the longitudinal dynamics, and vertical rotors' characteristics, including thrust coefficient K_T [3], rotors' thrust dependence on the longitudinal speed T_u [4] and a time-delay on rotors' response to controller inputs. Since preliminary coupling analysis showed that there were no significant off-axis effects, the inter-axes factors, that are I_{xy}, the rolling moment terms $C_{\mathcal{M}_{l_i}}$ and the roll rate terms C_{j_p}, were discarded in the subsequent analyses, reducing the input space to $n = 14$ dimensions.

Table 4. Uncertain input factors for UA, pitch test.

Description	Factor	Value	Range	Unit
Mass/Inertia	M	7	± 0.5	kg
	I_y	0.7237	$\pm 10\%$	kg·m^2
Rotor Parameters	K_T	$6.96 \cdot 10^{-8}$	$\pm 15\%$	–
	T_u	1	$\pm 5\%$	–
	τ_d	0	$+0.03$	s
0-AoA Coefficients	$C_{\mathcal{L}_0}$	0.7745	$\pm 20\%$	–
	$C_{\mathcal{D}_0}$	0.0183	$\pm 20\%$	–
	$C_{\mathcal{M}_{m_0}}$	-0.0240	$\pm 20\%$	–
α-derivatives	$C_{\mathcal{L}_\alpha}$	5.1270	$\pm 20\%$	–
	$C_{\mathcal{D}_\alpha}$	0.2558	$\pm 20\%$	–
	$C_{\mathcal{M}_{m_\alpha}}$	-0.5402	$\pm 20\%$	–
q-derivatives	$C_{\mathcal{L}_q}$	7.6136	$\pm 20\%$	–
	$C_{\mathcal{D}_q}$	0.7179	$\pm 20\%$	–
	$C_{\mathcal{M}_{m_q}}$	-16.169	$\pm 20\%$	–

Uncertain ranges for each class of factors are assigned following a separate reasoning, given their intrinsic differences: aerodynamic and stability terms are derived from their nominal values, with a 20% increase/decrease applied. Similarly, I_y, K_T and T_u are defined in percentage terms, limiting their range of variability to avoid divergent simulations; rotors' delay τ_d follows the same logic. The mass range varies in both directions by 0.5 kg.

---

3   As from Ref. [11,16], K_T is defined as $K_T = C_T \rho A R^2 = T/\Omega^2$.
4   T_u is modelled as a gain on the airspeed information fed by the FSM sensor unit to the thrust-related LUT, to not influence thrust dependence on throttle command.

4.2. Moderate-Amplitude Pitch Attitude Changes

The first pHQ studies the attitude quickness: this criterion aims "to characterize the helicopter's ability to achieve rapid, precise attitude changes when performing sharp, moderate amplitude manoeuvres" [26]. Given the dynamics and control architecture of the eVTOL under study, a new interpretation of the requirement was necessary for two main reasons. The first concerns the input specification: ADS-33 suggests applying a step in the longitudinal cyclic for Attitude Command/Attitude Hold (ACAH) response types, or a pulse for Rate Response types [27]. In this case, since a positive control input ΔV_{lon} causes a pitch-down movement and the vehicle returns to its previous trim state after command release (similar to an ACAH-flown helicopter), a step input was selected as the most suitable for the test. The second challenge involves defining the output metrics. ADS-33 requires the attitude change "to be made as rapidly as possible from one steady attitude to another" [5], extracting two key quantities from the response, which are the ratio of the peak angular rate to the peak attitude change $q_{\text{pk}}/\Delta\theta_{\text{min}}$, and the minimum attitude excursion before settling at the new trim state $\Delta\theta_{\text{min}}$. However, due to the small pitch angles at which the eVTOL flies, after the step input and its corresponding $\Delta\theta_{\text{pk}}$, the vehicle reaches a new trim without any undershoot, making $\Delta\theta_{\text{min}}$ difficult to identify clearly. In this study, by examining the recorded q time history, $\Delta\theta_{\text{min}}$ is then taken as the angle reached when the angular rate approaches the second stationary point after injection of the input as shown in Figure 6.
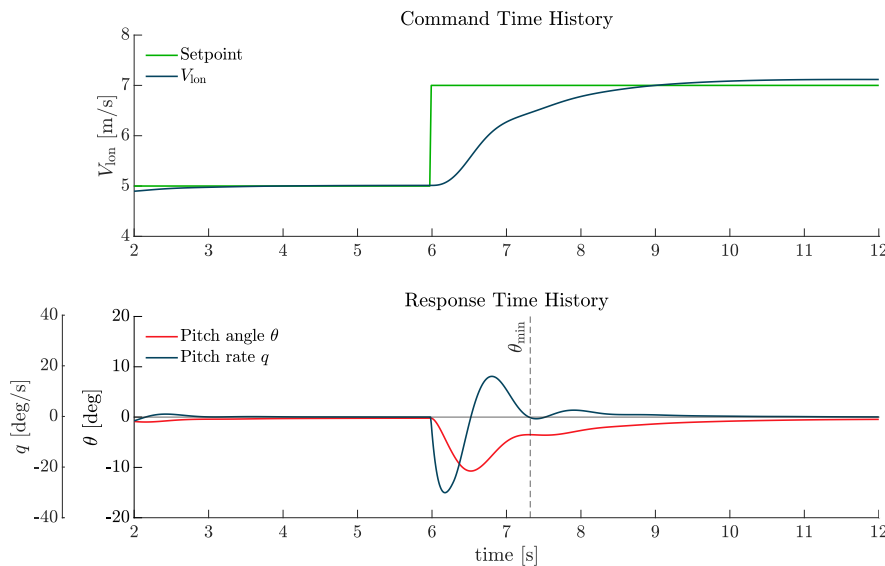

Figure 6. Pitch attitude quickness, low-speed test manoeuvre.

Figure 7 compares the results from the MOAT and VBD analyses. Focusing on MOAT results, Figure 7(**a**) and 7(**b**) display the modified mean μ^* against the standard deviation σ of the computed *EE*s as in [28]. The parameters located in the lower right-hand area behave mostly monotonically/linearly and appear to be highly significant in driving model's behaviour, but without significant interactions with the other uncertain variables. On the other hand, the variables placed in the upper left-hand region show a high nonlinear/interactional effect. To compare factors and assess their relative importance, the criterion based on μ^* is selected; alternatively, the Euclidean distance from the origin in the $\mu^* - \sigma$ plane could be used [21]. Figure 7(**a**) shows $\Delta\theta_{\text{min}}$ MOAT results: almost all factors behave monotonically, with little or no nonlinearities/interactions. Moreover, a large number of them seem to affect the output measure, from mass to aerodynamic parameters. Indeed aerodynamics plays a primary role. All lift-related coefficients are present, with a predominance of $C_{\mathcal{L}_0}$ and $C_{\mathcal{L}_\alpha}$. Pitching-moment stability derivatives appear too, together with drag-related ones; among the others, I_y and K_T prevail. Switching to the quickness metrics (Figure 7(**b**)) the behaviour changes drastically and the distinction between parameters becomes clearer: the eVTOL is mainly guided by mass and rotor-related factors, with response delay τ_d assuming a bigger role; as for aerodynamics, only $C_{\mathcal{L}_0}$ and

$C_{\mathcal{L}_\alpha}$ are visible, while the other terms are relegated to low order impacts. In Figure 7(c) Sobol indices and total effect indices for both pHQ metrics are shown. The threshold of 10^{-2} is indicated in the figure to easily discriminate the relevant factors, i.e. all those with a Sobol index above the threshold. Focusing on the minimum attitude change, good agreement with the MOAT results is visible, showing that total effects are often only relevant for low-influence factors. Both methods identify M, I_y, K_T and aerodynamic terms, confirming the importance of $C_{\mathcal{L}_0}$, $C_{\mathcal{L}_\alpha}$ and $C_{\mathcal{M}_{mq}}$.

(a) MOAT $\Delta\theta_{min}$, low-speed

(b) MOAT $q_{pk}/\Delta\theta_{min}$, low-speed

(c) Sobol indices and total effect, hover/low-speed
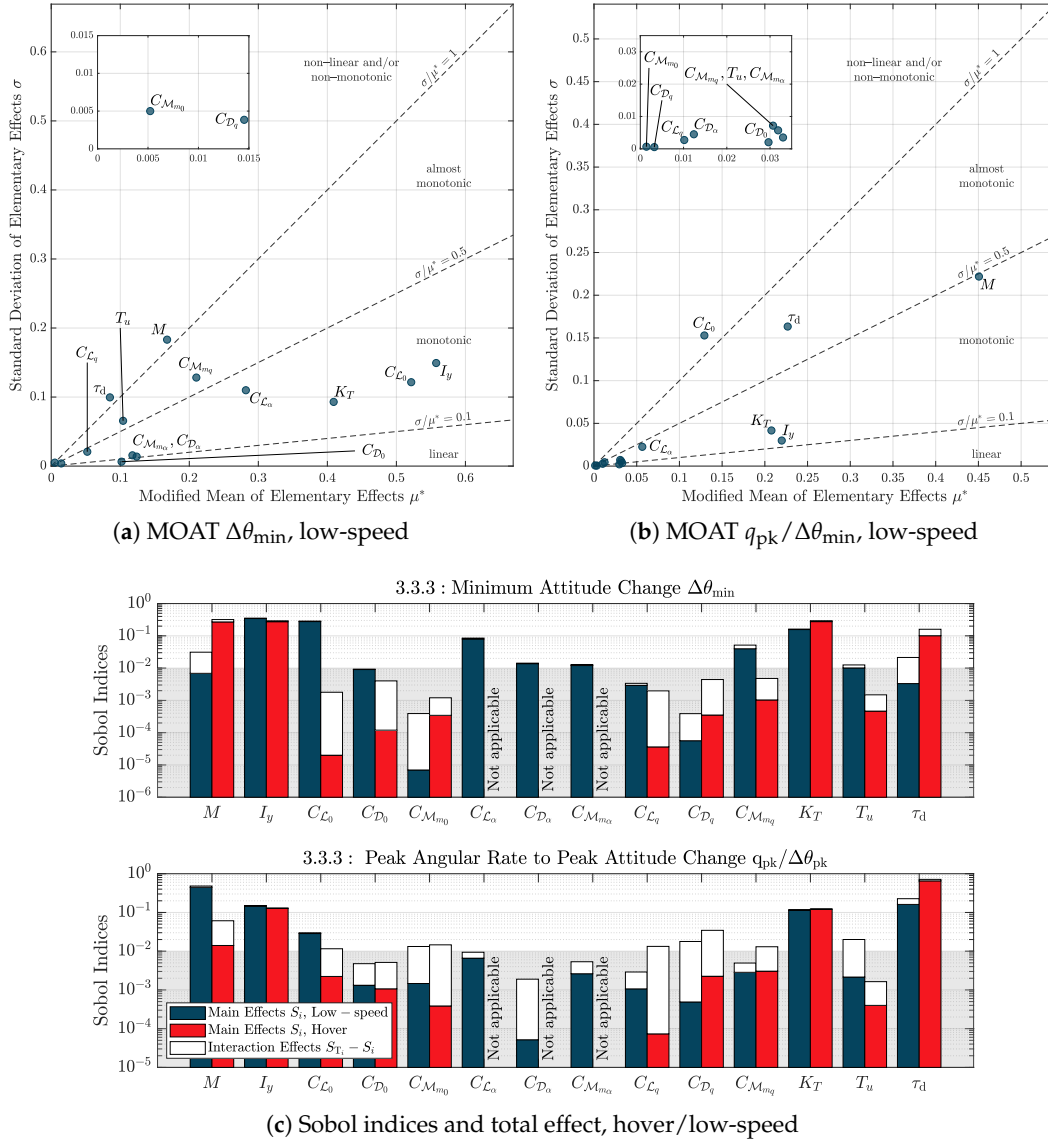
Figure 7. Pitch attitude quickness UA results.

The results also for hover are shown for completeness only for the VBD cases, showing trends similar to the low-speed case.

Figure 8 shows the results of the MC analysis in the attitude-quickness requirement plane. Most of the variability is located close to the high part of the y-axis. This is particularly evident in the hover case, which is affected by a high dispersion for the quickness metrics (from 3.5 to a maximum of 5 1/s). A similar trend for low-speed manoeuvring is shown in Figure 8(b). In this case, the cluster is limited to the 2.5-3.5 1/s range. Despite the greater output variance, the vehicle shows better manoeuvrability when operated in hover. Due to the small dimensions of the aircraft, the results significantly exceed the HQ Level 1 limits prescribed for classical helicopters. This was expected and should not be considered indicative of the expected good HQ of the drone VTOL here under analysis, because the pHQ limits

provided by [5] must be considered as design indications only for the specific vehicle for which they have been drawn, which are manned helicopters.
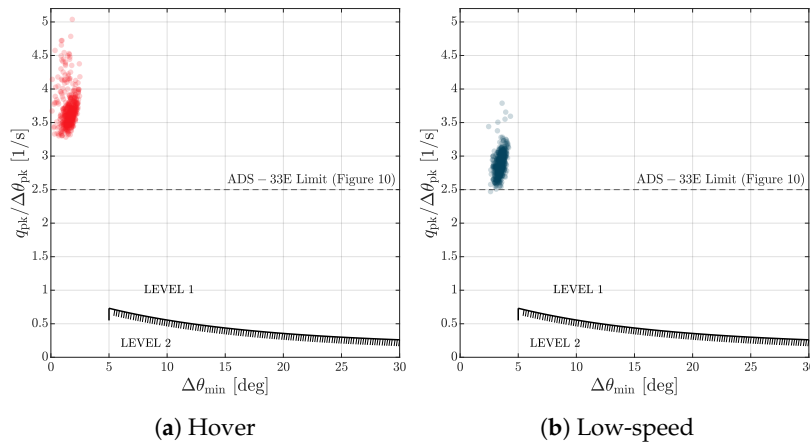

(a) Hover   **(b)** Low-speed

Figure 8. Attitude quickness ADS-33 results from MC analysis

4.3. Mid-Term Response to Pitch Control Inputs

The requirement of Section 3.3.2.3 in Ref. [5] assesses the dynamic stability of the vehicle by looking at the eigenvalues derived from the q-response to a ΔV_{lon} pulse input. Figure 9 shows typical time histories. As per Ref. [5], "the portion of the time history used to calculate the damping ratio should occur after the input is completed", thus the initial oscillation caused by the pulse is disregarded. Additionally, since the damping ratio must be measured from the free response, concerns arose regarding the onboard attitude and rate controllers quickly restoring the trim state. To minimise their influence, the damping ratio ζ was evaluated using the Transient Peak Ratio (TPR) method [5] considering only the first peak and first valley. The second positive peak was used to compute the characteristic period and the natural frequency ω_{n}.
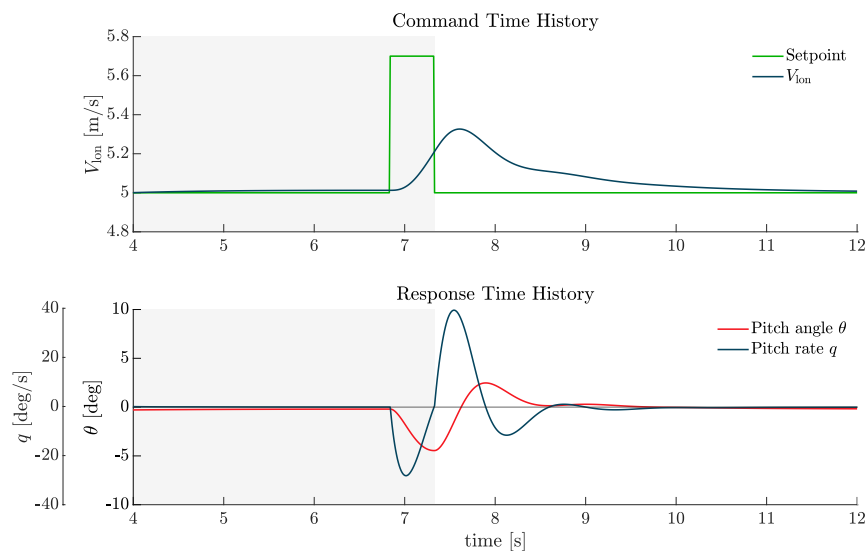
Figure 9. Pitch dynamic stability, low-speed test manoeuvre

Two main trends emerge from the damping ratio (ζ) results in Figure 10(a). Firstly, aerodynamics plays a greater role than in the quickness case, since $C_{\mathcal{L}_0}$ and $C_{\mathcal{M}_{m_q}}$ rank among the factors with the highest μ^*. Secondly, a high standard deviation is associated with M and τ_{d} at low speeds, suggesting possible nonlinearities or interactive effects between variables. Regarding the natural frequency ω_{n}, MOAT identifies three groups in the $\mu^* - \sigma$ plane: the first dominated by M, K_T and

τ_{d}, the second by the zero-lift coefficient $C_{\mathcal{L}_0}$, pitch-axis inertia I_y and $C_{\mathcal{M}_{mq}}$ — confirming damping trends for aerodynamics —, and a third combining α-derivatives, rotor speed dependency T_u and the pitch-rate derivative of the lift coefficient $C_{\mathcal{L}_q}$. Sobol indices in Figure 10(c) confirm the MOAT findings, reasserting the relevance of $C_{\mathcal{L}_0}$ and $C_{\mathcal{M}_{mq}}$, already highlighted in the quickness simulations. Also in this case hover shows similar trends.

(a) MOAT ζ, low-speed

(b) MOAT ω_{n}, low-speed

(c) Sobol indices, hover/low-speed
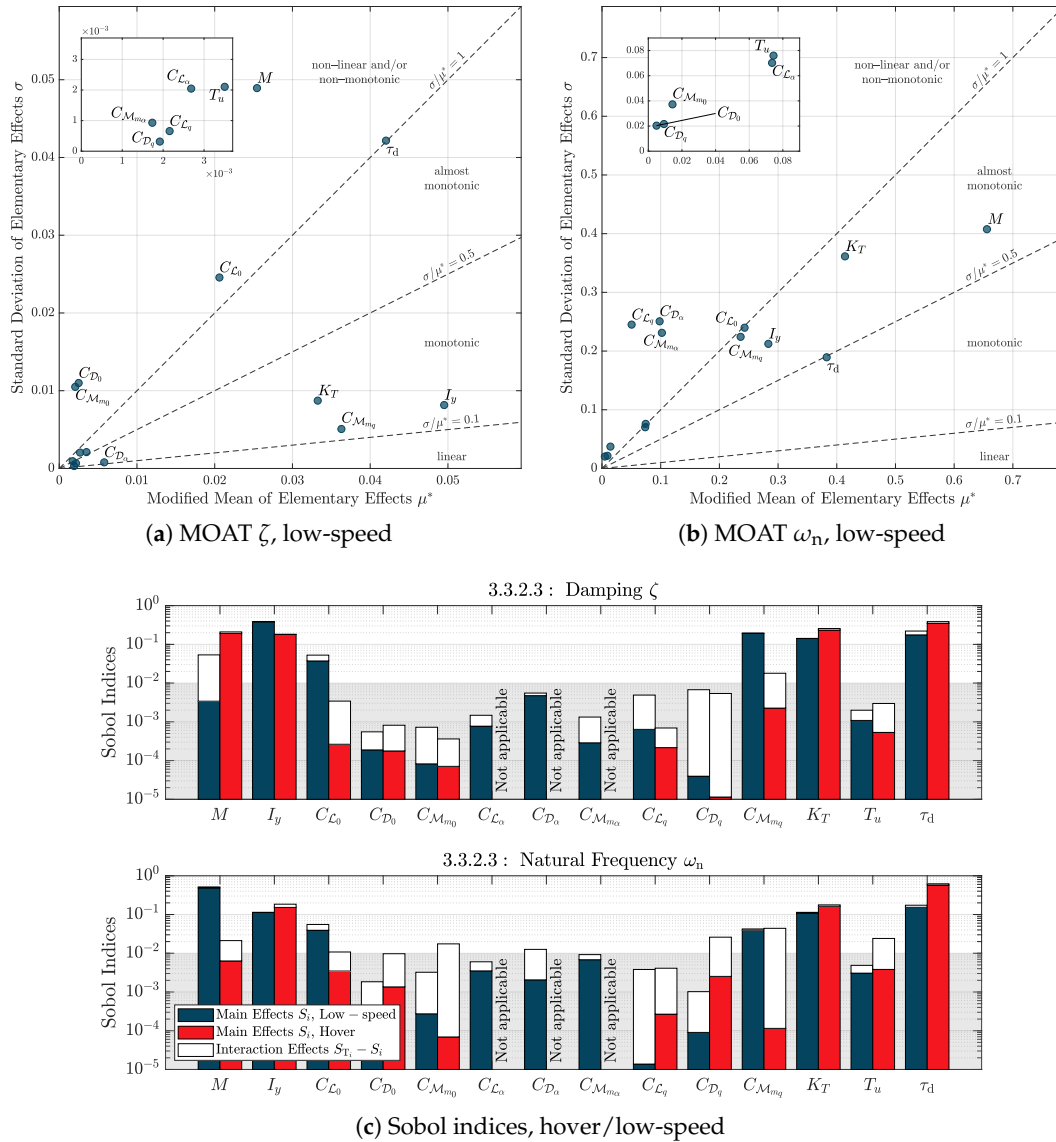
Figure 10. Pitch dynamic stability UA results

Figure 11 illustrates that the aircraft operates near the boundary between Level 1/2 established in Ref. [5] for helicopter. Although the damping fluctuates around $\zeta = 0.35$ by approximately $\pm\, 0.05$, the frequency shows much greater variability and significantly exceeds the helicopter values , with $\omega_n = 5.61$ rad/s corresponding to a natural oscillation frequency of 0.9 Hz. This represents a very high frequency and is again attributed to the small size of the vehicle compared with the size of a manned helicopter.

(**a**) Hover   (**b**) Low-speed

Figure 11. Dynamic stability ADS-33 results obtained from MC analysis

4.4. Short-Term Response to Pitch Control Inputs

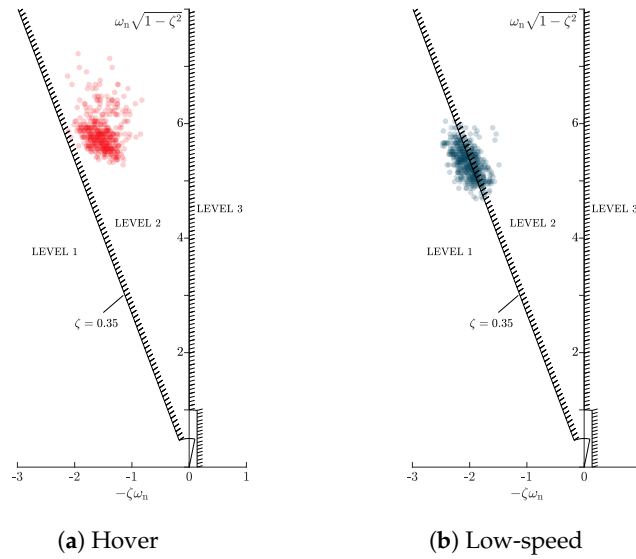The last pHQ analysed is the bandwidth ω_{BW} described in Section 3.3.2.1 of Ref. [5], which investigates the behaviour of the aircraft in the frequency domain "to determine if there is sufficient phase margin to allow the pilot to close the attitude loop [...] without threatening stability" [5]. The test used requires a longitudinal frequency sweep input and recording of the q response, processed using the Fast Fourier Transform algorithm (FFT) to obtain Bode plots and evaluate ω_{BW} and the phase delay τ_p (see [25] for the exact definition). Following [27], a 90 s automated sweep is generated, linearly increasing from 0.3 rad/s to 12 rad/s (approximately 0.05 Hz to 2 Hz). The ADS-33 test guide [27] also suggests recording 20 s before and after the sweep. Figure 12 illustrates the reference commands in the longitudinal speed and the related q-rate response.
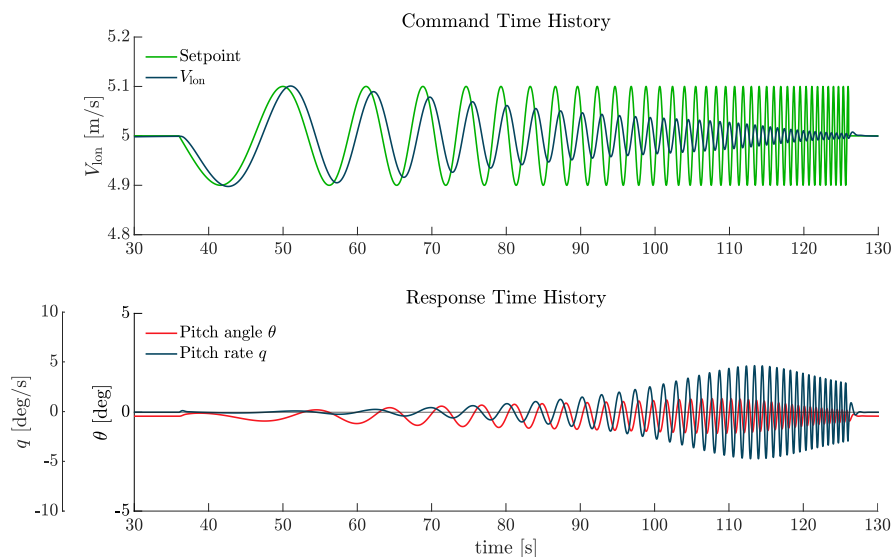
Figure 12. Pitch bandwidth, low-speed test manoeuvre

However, performing a 130 s simulation leads to inherently time-demanding computation to obtain the QoI which could make the UA prohibitively expensive. So, instead of analysing the complete nonlinear response, the criterion is applied to a linearized form of the FSM near a $V_{\text{lon}} = 5$ m/s trim point, reducing the computational costs but still capturing the main trends in bandwidth behaviour.

Figure 13 compares the Bode plots obtained by the frequency sweep with those of the linearized system. It should be noted that the agreement between the models holds as long as the sweep applied to the nonlinear model stays within reasonably small amplitudes[5]. As the applied sweep amplitude increases, differences may arise. A coherence plot is also shown for the FFT-based results. The ADS-33 Test Guide [27] indicates a minimum value of 0.6 for the coherence to check the goodness of the nonlinear system sweep results. As specified in Reference [5], since the phase is nonlinear between the crossover frequency ω_{180} and $2\omega_{180}$, the phase delay τ_p must be determined by a linear least squares fit.
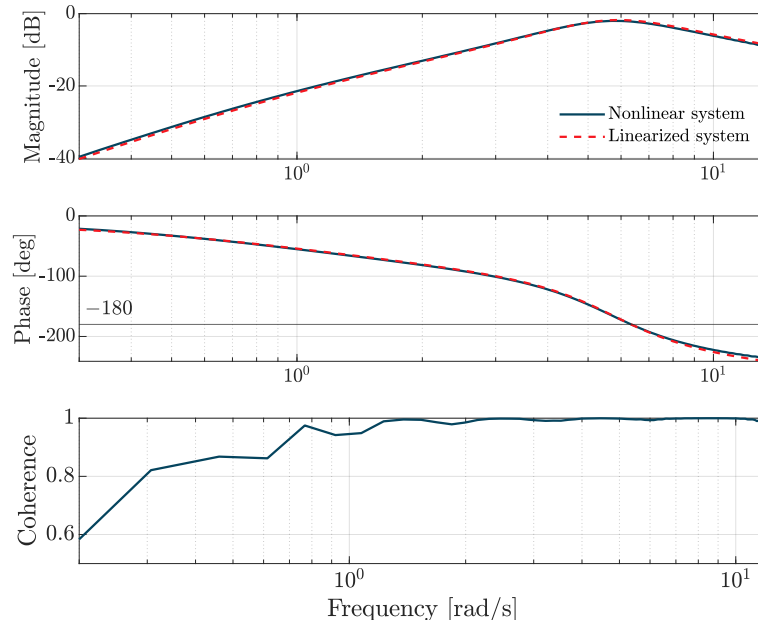

Figure 13. Bode plots for $q/\Delta V_{\text{lon}}$

Unlike the other pHQs, the bandwidth results are much less variant (see Figure 14). This is reflected in both the MOAT analyses — featuring extremely small σ-values, which make most of the uncertain inputs fall into the linear/monotonic area — and in the VBD, where Sobol indices of relevant factors are mainly composed of first-order effects. Apart from VTOL's mass — responsible for most of the output variance —, three subgroups are identified. The first is again composed of $C_{\mathcal{L}_0}$ and $C_{\mathcal{M}_{mq}}$, together with K_T and I_y. These factors contribute to $\omega_{BW_{\text{phase}}}$ with Sobol indices varying between values of 0.05 and 0.1. The second subset group is composed of α-derivatives, T_u and τ_d, and shows values less relevant but close to the threshold. The third group contains irrelevant factors. τ_p (Figure 14(**b**)) exhibits a joint participation of all factors, as seen for $\Delta\theta_{\text{min}}$ in moderate-amplitude pitch attitude changes requirement, with an increased importance played by aerodynamics. Figure 14(**c**) shows the Sobol indices, again in agreement with MOAT μ^*.

---

5   Figure 13 has been derived from a $\Delta V_{\text{lon}} = \pm 0.1$ m/s, resulting in $|\Delta\theta| < 1$ deg; up to $|\Delta\theta| \approx 5$ deg.

(a) MOAT ω_{BW}, low-speed

(b) MOAT τ_p, low-speed

(c) Sobol' Indices, hover/low-speed
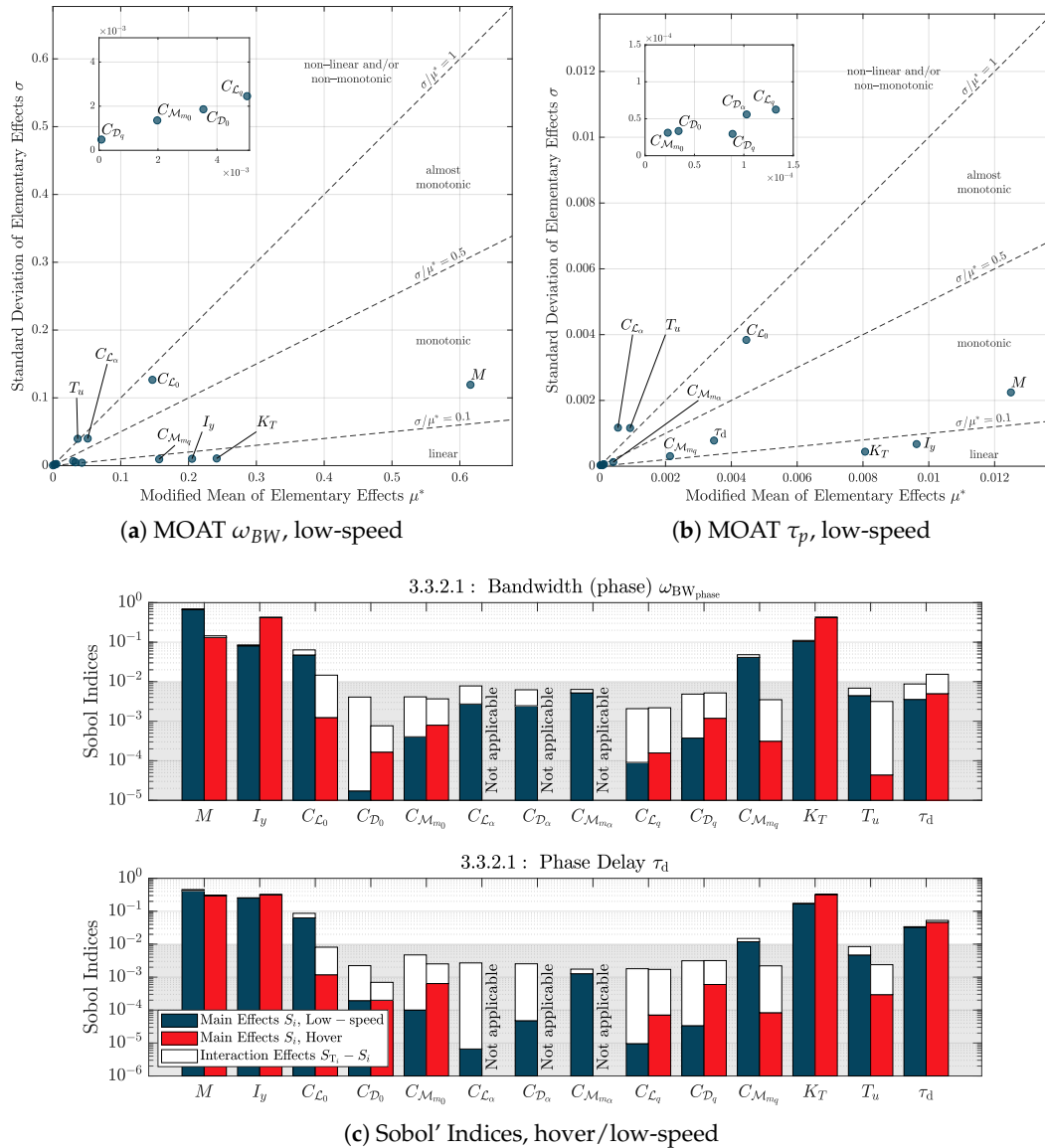
Figure 14. Pitch bandwidth UA results

The results of the bandwidth and phase delay UA are finally compared with the ADS-33 design limits defined for helicopters in Figure 15. A cluster of points exists at a significant distance from the threshold between the levels. $\omega_{BW_{\text{phase}}}$ shows a wider range of variability than τ_p. The Bode phase curve of the uncertain models showed a greater variation in the range between bandwidth and crossover frequency (ω_{180}), shifting bandwidth values over the interval of 4-5 rad/s. Since the phase curve is less sensitive to variations for $\omega > \omega_{180}$, τ_p instead remained almost unchanged after ω_{180}, resulting in a much lower variance in the phase delay results.

Figure 15. Bandwidth ADS-33 results from MC analysis

In summary, it is possible to say that the SA analysis could be used effectively to identify the most relevant factors that influence the input uncertainty in terms of the selected pHQ. The efficient MOAT analysis was very effective in identifying the relevant factors, providing results very close to those obtained by the more demanding VBD analysis. Additionally, the MOAT provided also a good iden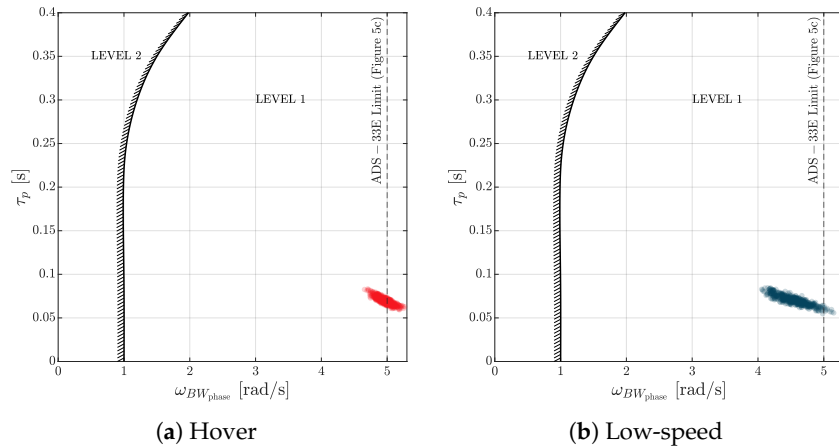tification of the factors that may present a significant nonlinearity, which in turn may lead to total effect indexes significantly different from Sobol indexes. For these cases, the VBD analysis could be used. Then MC results, mainly extracted from analysis already used to determine the VBD, could be used to assess the uncertainty of the FSM in the space of pHQ.

5. Flight Test Manoeuvres for Assigned HQs Evaluation

To verify if the uncertain models that are identified through the previous UA are also those that lead to the largest differences in terms of HQR evaluated for the assigned HQ case under investigation, it is presented here an analysis of the response of several most critical uncertain cases identified obtained performing the deceleration to hover manoeuvre. In principle, these tests must be performed considering the presence of a pilot-in-the-loop, since a real pilot will probably adapt the piloting strategy while perceiving a change in the aircraft dynamic behavior. However, since we are only interested here in the potential impact if model input uncertainties on the aHQ we will simply perform the simulation closing all control loops of the flight controller of the aircraft and checking *de facto* if the same virtual pilot, without strategy adaptation is able to still control the aircraft and keep the same desired/adequate performance of the nominal model.

5.1. Selection of Test Points

At this stage, the focus shifts to the assigned dimension of the HQs. The main objective of this section is to execute the selected FTM to collect the HQ evidence required by the MHQRM [1]. Exploiting UA results, a series of uncertain aircraft configurations is selected from the MC output dataset for the execution of the FTM. Two sets of HQ tests are presented: the first is carried out in the Normal Flight Envelope (NFE) [6], in calm air. The second, is still in the NFE but now operates with atmospheric disturbances, and is tested with the intent of verifying how much the uncertainty impacts on the evaluation with AD as required by [1].

Since the FSM do not allow for the pHQ metrics to be controlled as direct input variables, the employed configurations are sampled from pHQs MC output distributions. For each pHQ (i.e. attitude quickness, dynamic stability, bandwidth), four equally spaced points are chosen moving in the "least favourable" directions, proceeding from each index' mean value to the extreme of the distribution. These direction are identified as those who go in the direction of worsening the "level" of the aircraft

---

[6]   The NFE "*corresponds to the normal use of the VTOL aircraft according to its Concept of Operations*" [4].

HQR as indicated by the pHQ maps of Ref. [5]. Since each pHQ features two individual metrics, a total of 24 configurations have been tested, summarised case by case in Figure 16.
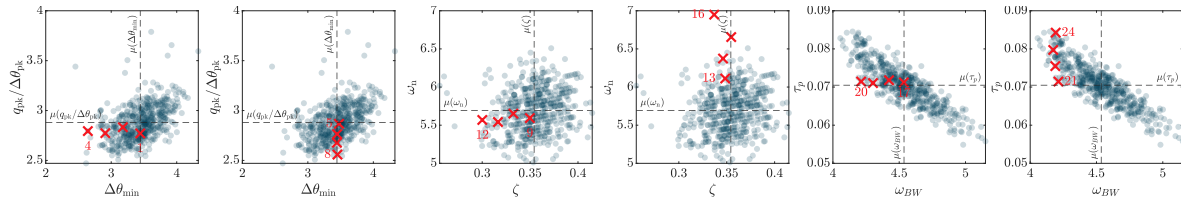

Figure 16. Selected test configurations, graphical representation

Each group of four points is equally spaced and minimize the distance from the mean value of the other index[7] pertaining to the same pHQ. In this way, we expect each configuration to show the variability trends only for the considered QoI, with limited interference from the secondary one. Table 5 provides a complete picture of the 24 defined configurations in terms of pHQ.

Table 5. Selected test configurations predicted HQs.

	00*	01	02	03	04	05	06	07	08	09	10	11	12
$\Delta\theta_{\min}$	3.50	3.46	3.20	2.93	2.66	3.49	3.47	3.47	3.48	3.43	3.19	3.07	2.82
$q_{\mathrm{pk}}/\Delta\theta_{\mathrm{pk}}$	2.78	2.77	2.84	2.77	2.80	2.86	2.76	2.68	2.55	2.61	2.65	2.61	2.62
ζ	0.37	0.37	0.33	0.32	0.31	0.35	0.33	0.35	0.34	0.35	0.33	0.32	0.30
ω_{n}	5.92	5.84	6.10	5.91	5.85	5.98	6.04	5.69	5.44	5.59	5.65	5.54	5.57
ω_{BW}	4.52	4.43	4.62	4.49	4.46	4.34	4.30	4.37	4.18	4.12	4.30	4.24	4.18
$\tau_p \cdot 10^{-2}$	6.72	7.70	7.58	6.93	6.83	7.78	7.18	7.97	7.66	8.58	8.14	7.73	8.09

* Baseline configuration

	13	14	15	16	17	18	19	20	21	22	23	24
$\Delta\theta_{\min}$	3.28	3.50	3.43	3.76	3.45	3.39	3.56	3.53	3.53	3.36	3.18	3.15
$q_{\mathrm{pk}}/\Delta\theta_{\mathrm{pk}}$	2.88	2.95	3.17	3.75	2.82	2.78	2.72	2.65	2.65	2.61	2.61	2.58
ζ	0.35	0.35	0.36	0.34	0.35	0.36	0.39	0.40	0.40	0.37	0.34	0.31
ω_{n}	6.11	6.37	6.66	6.95	6.02	5.91	5.83	5.73	5.73	5.61	5.56	5.53
ω_{BW}	4.62	4.68	5.06	3.99	4.52	4.40	4.27	4.21	4.21	4.18	4.15	4.18
$\tau_p \cdot 10^{-2}$	7.04	7.01	6.15	8.11	7.14	7.09	7.11	7.04	7.14	7.53	7.86	8.42

5.2. Normal Flight Envelope Results

Figure 17 presents the results for the HQs in the 24 configurations, for each of the performance listed in the Table 3. First, it must be noted that the reference configuration is not able to ensure Level 1 HQ due to the inability to stay within the desired limits for altitude. A more extensive look at Table 3 reveals that the altitude is the performance that is the most affected by the uncertainty of the FSM. The analysis of the results reveals that the most critical performance standard —— and the most sensitive to uncertainty in the pHQs —— is the altitude. Figure 17 shows that the eVTOL gains significant altitude during the FTM. At the start of the deceleration phase, the controller commands a sharp pitch-up manoeuvre to initiate braking. Due to the high aggressiveness of the command and the sustained forward speed at which the deceleration begins, the aircraft responds with a rapid climb. A similar effect occurs during acceleration, where the eVTOL loses altitude. However, since this phase starts from a hover, the impact is less marked.

In terms of lateral-directional performance, the vehicle comfortably achieves the desired level, with lateral track deviations and heading excursions remaining around the order of 10^{-2} throughout all simulations. This outcome directly reflects the findings of roll-due-to-pitch coupling, which confirmed the absence of coupling when the aircraft is operated in a purely longitudinal manner. The numeric results have not been reported in the Table 3 given their low relevance.

---

7   This is true except for τ_p. Due to the skewed shape of the MC distribution, to consider its worst-case values, it has been necessary to make a different choice.

Regarding the longitudinal requirement, although the aircraft can maintain a stable hover for the expected duration, configuration uncertainties can result in minor deviations from the desired region.

The analysis highlights two main outcomes:

- Firstly, aside from the altitude requirement, the FTM can generally be completed at the desired level when flying in the NFE (i.e. in the absence of external disturbances). This is primarily due to the effectiveness of the FCS, which is capable — even under high attack rates — of keeping the eVTOL within the required boundaries, especially along the lateral-directional axes.
- Secondly, despite the positive contribution of the FCS to the closed-loop response of the aircraft , the uncertainty in the pHQs still has a significant impact on the overall FTM performance. Although no consistent trend emerges between the worse pHQ simulations and worse aHQ obtained, small variations in the tested configurations can easily push the eVTOL to different levels of performance threshold. It must be noted, that it was not possible to select configurations where only a single pHQ characteristic was modified keeping all the others constant.

Figure 17. NFE task performance results. 00 represents the baseline configuration. The colour code refer to the performances: green desired, yellow adequate, red controllable, grey uncontrollable.

Overall, it is possible to say that parameter sets in the range of uncertainty that lead to worse pHQs, also show significant effect on the aHQ behaviour. So, the idea of performing the uncertainty assessment at the level of pHQ, using much simpler open-loop tests, represents a good approach for the identification of the most critical FSM. The results suggest that the level of uncertainty in the input models is large and consequently the impact on the aHQ is significant. So, it is possible to conclude that a lower uncertainty in the FSM input parameter will be necessary to employ only the nominal model in the aHQ simulations with confidence.

5.3. Atmospheric Disturbances Flight Test Results

The FTM operations are now extended to include the effect of atmospheric disturbances, entering is an area of the flight envelope where the credibility of the FSM becomes crucial to assess HQ results. To analyse the VTOL behaviour under external disturbances, a wind turbulence model is introduced.

The method adopted here follows MIL-F-8785C (Reference [29]), which states that the "*engineering model may be considered as the simplest model which is consistent with any related usage in aircraft design, but still correctly identifies the primary parameters of particular interest*". Accordingly, the Von Kármán spectral model is used. The model defines the power spectral density for the three turbulent velocity components - taking white noise as input - and the turbulent roll angular rate spectrum; pitch and yaw rates are derived subsequently from the velocity components.

In addition to random fluctuations, a mean wind component dependent only on altitude is applied. Two mean wind velocities V_{wind}[8], 0.5 and 1 m/s, are analysed, each combined with three wind directions ($\psi_{\text{wind}} = 0°, 90°, 180°$). Since the FTM operates at low altitude, the Integral Length Scales $L_{u,v,w}$ and the Turbulence Intensities $\sigma_{u,v,w}$ depend only on altitude, as per MIL-F-8785C [29].

---

[8] Evaluated at 6 m altitude and adjusted according to the FTM altitude considering wind shear.

To eliminate any bias associated with the stochastic nature of turbulence, a preliminary set of 40 runs is performed using the nominal aircraft configuration. . For each turbulent component (u, v, w, p), 10 simulations are run, varying only that component while keeping the others constant. The random seed yielding the worst performance is selected for each component. This procedure is repeated for every wind speed and direction choosing the worst-case performance for each wind condition. The selected seed vectors are then used consistently across the 24 test configurations, as defined in Table 5, leading to the results summarized in Figure 18.
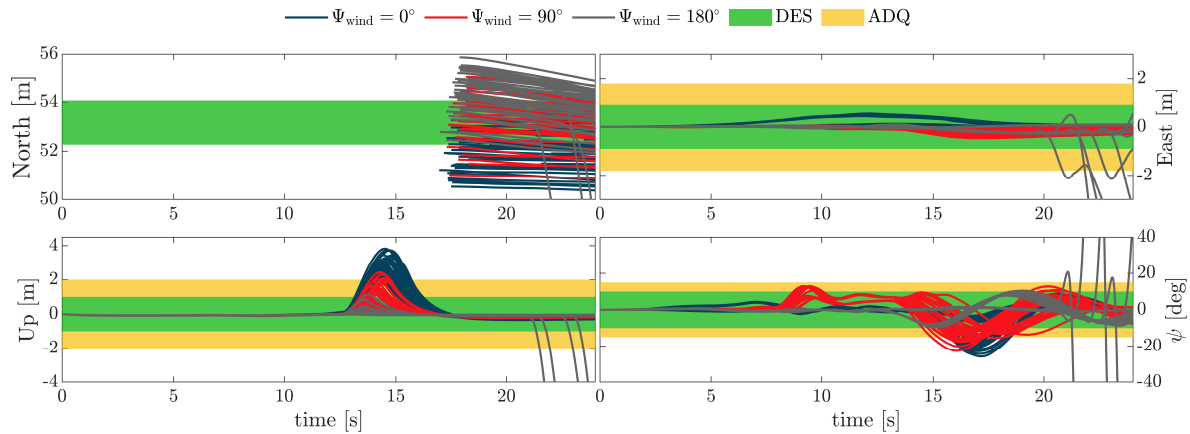

Figure 18. Recorded time histories with atmospheric disturbance

The performance parameters deviations due model input uncertainties u_{inp} are significantly amplified in the presence of disturbances. The aircraft is no longer able to consistently meet the longitudinal requirement. Furthermore, during the hover phase, substantial deviations are observed from the intended arrival point (top-left plot in Figure 18), mainly due to wind effects during hover and the use of the multicopter controller's position mode. Enhancing position control during this phase may help maintain compliance with the requirement, allowing the aircraft to hold its intended arrival point.

An interesting aspect is furthermore represented by the tailwind case, i.e. $\psi_{wind} = 0$ deg, as in general it reveals to be the only wind condition in which the FTM acceleration-deceleration phase generally satisfies the desired lateral, vertical, and heading requirements. However, 8 configurations with tailwind conditions result in divergent simulations. In these cases the eVTOL struggles to maintain stable hover in the final phase, when the tailwind excites a divergent pitch oscillation that ultimately leads to complete loss of control. This behaviour will need to be investigated further in the future.

Heading performance also suffers a marked degradation, with the eVTOL exhibiting large ψ deviations once the deceleration ΔV_{lon} is commanded. This effect is primarily driven by the p-rate turbulence component, which, combined with lateral velocity v, strongly couples with the pitch and yaw axes. As a result, the aircraft frequently exceeds adequate boundaries, especially under headwind and crosswind conditions. Figure 18 highlights how this issue becomes more prominent near the end of the deceleration phase, when the aircraft reaches its maximum pitch angle.

Similarly, altitude performance drops to the controllable level, with headwind conditions proving particularly critical (as shown in Figure 19). The longitudinal turbulence component intensifies the aggressiveness of the deceleration phase, causing the vehicle to exceed desired thresholds. In contrast, lateral performance is comparatively less affected, as the FCS still manages to track the test course within the desired standards.

Lateral Track Deviation [m]

	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
N-1	0.09	0.10	0.09	0.10	0.14	0.12	0.10	0.10	0.10	0.11	0.11	0.15	0.12	0.10	0.10	0.09	0.08	0.09	0.11	0.12	0.11	0.11	0.11	0.11	0.11
N-2	0.48	0.48	0.46	0.50	0.48	0.54	0.48	0.48	0.49	0.52	0.50	0.51	0.54	0.49	0.48	0.44	0.44	0.47	0.48	0.51	0.50	0.50	0.51	0.51	0.50
E-1	0.24	0.18	0.23	0.32	0.29	0.20	0.29	0.19	0.35	0.38	0.34	0.35	0.37	0.18	0.26	0.12	0.13	0.26	0.29	0.30	0.23	0.23	0.31	0.32	0.36
E-2	0.11	0.10	0.14	0.39	0.47	0.11	0.23	0.10	0.25	0.33	0.30	0.46	0.27	0.09	0.12	0.06	0.10	0.14	0.37	0.15	0.12	0.12	0.14	0.18	0.31
S-1	0.09	0.09	0.07	0.08	0.07	0.08	0.11	0.09	0.10	0.09	0.09	0.09	0.08	0.08	0.08	0.07	0.06	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.10
S-2	0.25	0.23	0.27	UNC	0.74	UNC	UNC	0.26	0.28	UNC	UNC	0.28	UNC	0.23	0.26	UNC	UNC	0.26	0.26	0.23	0.26	0.26	0.26	0.26	0.28

Heading [deg]

	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
N-1	5.40	5.77	4.68	3.53	6.69	5.41	3.98	6.45	6.16	5.93	5.15	6.16	5.56	5.43	5.32	4.89	4.65	5.20	3.45	4.93	6.61	6.61	5.78	5.61	5.70
N-2	19.33	14.21	24.24	20.83	20.70	13.83	25.74	17.76	21.59	22.04	22.78	15.99	22.36	14.67	20.16	19.17	16.55	19.61	18.00	13.61	17.01	17.01	19.03	18.50	22.59
E-1	12.22	10.73	15.69	12.22	12.11	10.67	12.30	10.83	12.92	13.06	13.31	10.84	13.80	10.15	12.79	11.57	10.72	12.33	10.75	10.87	11.73	11.73	12.52	12.55	12.74
E-2	12.52	12.74	15.39	20.08	22.54	12.42	14.54	12.50	13.90	16.45	17.28	16.63	14.60	12.43	12.59	13.58	13.16	12.48	17.62	12.52	12.58	12.58	12.75	12.80	16.67
S-1	1.28	1.43	1.15	1.36	1.24	1.41	1.16	1.40	1.45	1.56	1.38	1.64	1.49	1.37	1.26	1.35	1.15	1.28	1.46	1.60	1.49	1.49	1.47	1.48	1.47
S-2	9.14	8.49	10.22	UNC	10.15	UNC	UNC	8.77	9.52	UNC	UNC	10.45	UNC	8.59	9.43	UNC	UNC	9.23	9.77	8.54	8.96	8.96	9.27	9.32	9.72

Altitude [m]

	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
N-1	2.41	2.55	2.63	1.98	0.71	2.33	2.21	2.70	2.23	2.04	2.18	0.32	2.14	2.49	2.46	2.96	2.76	2.33	1.22	1.99	2.54	2.54	2.19	2.09	2.21
N-2	3.27	3.42	3.45	2.92	2.82	3.25	3.25	3.73	3.20	3.02	3.05	2.10	3.02	3.33	3.28	3.83	3.75	3.23	2.55	2.92	3.45	3.45	3.06	2.98	3.16
E-1	0.82	1.32	0.99	0.27	0.18	1.21	0.50	1.40	0.41	0.29	0.36	0.09	0.31	1.29	0.77	1.86	1.62	0.63	0.20	0.43	1.23	1.23	0.52	0.41	0.33
E-2	1.77	1.97	2.03	0.38	0.22	1.85	1.29	2.08	1.11	0.79	0.86	0.09	0.92	1.93	1.81	2.43	2.29	1.50	0.27	1.21	1.94	1.94	1.46	1.24	0.73
S-1	0.22	0.45	0.28	0.08	0.08	0.38	0.14	0.48	0.13	0.09	0.11	0.09	0.10	0.39	0.21	1.13	0.83	0.19	0.08	0.12	0.36	0.36	0.15	0.12	0.11
S-2	0.08	0.07	0.08	UNC	0.08	UNC	UNC	0.08	0.09	UNC	UNC	0.09	UNC	0.07	0.08	UNC	UNC	0.08	0.08	0.08	0.08	0.08	0.09	0.08	0.09

Longitudinal Arrival Point [m]

	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
N-1	-1.36	-0.56	-2.43	-1.15	-1.04	-0.40	-1.44	-0.90	-0.45	+0.11	-0.94	+0.89	-0.37	-0.83	-1.55	-2.54	-2.23	-1.49	+0.11	+0.48	-0.45	-0.45	-0.40	+0.07	-0.54
N-2	-1.75	-1.12	-2.59	-1.88	-2.28	-0.41	-1.67	-1.42	-1.09	-0.62	-1.73	-0.81	-1.08	-1.32	-1.92	-2.75	-2.06	-1.85	-1.49	-0.25	-0.95	-0.95	-1.11	-0.89	-1.21
E-1	+0.27	+0.79	-1.00	+0.55	-0.55	+0.96	+0.10	+0.52	+1.29	+1.52	+0.71	+1.43	+1.27	+0.50	+0.08	-1.79	-1.43	+0.24	+0.96	+1.90	+1.04	+1.04	+1.28	+1.48	+1.28
E-2	-0.54	+0.06	-1.78	+0.38	-0.52	+0.31	+0.06	-0.48	+0.66	+0.86	+0.27	+1.28	+0.78	-0.43	-0.80	-2.25	-1.39	-0.51	+0.78	+1.28	+0.18	+0.18	+0.58	+0.94	+0.64
S-1	+0.65	+1.16	-0.75	+0.79	+0.13	+1.38	+0.45	+0.98	+1.57	+1.82	+1.05	+1.53	+1.54	+0.93	+0.51	-1.11	-0.93	+0.63	+1.15	+2.18	+1.54	+1.54	+1.65	+1.75	+1.51
S-2	+1.25	+1.82	+0.27	UNC	+0.64	UNC	UNC	+1.76	+2.18	UNC	UNC	+2.06	UNC	+1.55	+1.08	UNC	UNC	+1.22	+1.68	+2.70	+2.37	+2.37	+2.31	+2.27	+2.06

| $\Delta\theta_{\min}$ [deg] | $q_{pk}/\Delta\theta_{pk}$ [s^{-1}] | ζ [−] | ω_n [rad/s] | ω_{BW} [rad/s] | τ_p [s] |

Figure 19. Task performance results with atmospheric disturbances, for varying incoming wind direction (N-E-S) and magnitude (1 → V_{wind} = 0.5 m/s and 2 → V_{wind} = 1 m/s). 00 represents the baseline configuration. The colour code refer to the performances: green desired, yellow adequate, red controllable, grey uncontrollable

In summary, these analyses highlight that the effects of wind disturbances can be assessed using these simulations. Trends with wind intensity and direction are identifiable using the nominal FSM. These trends are reproduced similarly by the modified configurations, excluding some cases. However, it must be remembered that the uncertainty levels considered have been assessed as large in the previous section.

6. Maximum Unnoticeable Added Dynamics

MUAD boundaries are defined as the alterations in aircraft dynamics that can be introduced in the aircraft without the pilot consciously perceiving them. The approach proposed in the introduction says that if the envelope of the uncertain transfer functions pertaining to the ACR under consideration multiplied by the selected *CR* falls within the limits of the MUAD, then the nominal model could be considered credible and used to perform ACR evaluations with flight simulators. In this case, we will

make the hypothesis that the minimal acceptable CR is equal to 1 (see [6] for more details on it). In contrast, if the uncertainty ranges encompassing the nominal results exceed the MUAD boundaries, credibility should not be considered appropriate. So, the applicant will need to reduce the uncertainty or assess the impact of uncertainty on the simulations.

The envelopes (summarised in Table 6) are valid across the frequency range from 10^{-2} to 10^{2} rad/s and exhibit a "hourglass" shape, narrowing around 3 rad/s, which is the frequency region most relevant to manual control [30].

Table 6. MUAD Envelopes Transfer Functions [31].

Envelope Type	Transfer Function
Upper Gain Envelope	$\dfrac{3.16s^2 + 31.61s + 22.79}{s^2 + 27.14s + 1.84}$
Lower Gain Envelope	$\dfrac{0.0955s^2 + 9.92s + 2.15}{s^2 + 11.60s + 4.95}$
Upper Phase Envelope	$\dfrac{68.89s^2 + 1100.12s - 275.22}{s^2 + 39.94s + 9.99}e^{0.0059s}$
Lower Phase Envelope	$\dfrac{475.32s^2 + 184100s - 29456.1}{s^2 + 11.66s + 0.0389}e^{-0.0072s}$

In light of this, MC generated FSM can be employed to assess the influence u_{inp} has on the aircraft transfer function (TF), computing the envelope of uncertain TFs (eTFs) and comparing them with the nominal one. In performing this assessment, it must be clearly taken into account that the MUAD boundaries were originally conceived for a class of aircraft very different from eVTOLs and may therefore present limitations in their applicability. Nevertheless, they are used in this work to demonstrate the approach which enhances FSM credibility.

In Figure 20 below, the main outcome of this assessment is shown. The nominal system $q/\Delta V_{\text{lon}}$ Bode plots are displayed together with their uncertain region (Figure 20(**a**)) — derived using the MC results of the UA phase — and the corresponding "added-dynamics interpretation" is given (Figure 20(**b**)). The difference between the 500 MC output eTFs and the reference TF is interpreted as an unmodelled dynamics, showing that the uncertainty leading to the eTFs is bounded by MUAD limits for all the relevant frequencies (0.3-12 rad/s), except for some limited excursion outside magnitude boundaries slightly above 6 rad/s. Consequently, it can be fairly assumed that the FTM results obtained using the baseline configuration of the FSM would lead to the same HQ ratings as for the worst or best configuration in the uncertainty range, at least for the performance that depend on this transfer function.: according to the proposed approach, these results are indeed consistent with the performances achieved in the NFE. The differences between the 24 tested configurations are not such as to drive the eTF out of the MUAD envelopes for this TF.

Following a similar line of reasoning, the previous altitude performance results can be interpreted in MUAD terms. The FTM simulation phase highlighted indeed an increased altitude gain in the VTOL's response, revealing a strong coupling between the longitudinal and vertical axes in the assigned dimension of HQs. Therefore, it is necessary to investigate whether this behaviour may be attributed to the shape of the added dynamics related to input uncertainties. To explore this aspect, the eTFs between the longitudinal input command and the resulting vertical velocity response $V_z/\Delta V_{\text{lon}}$[9] are defined and the corresponding added dynamics are assessed.

---

[9] $X_z/\Delta V_{\text{lon}}$ analysis yields analogous results, preserving the shape of Figure 21(**b**) added dynamics envelope.

(**a**) MC uncertain Bode plots envelope
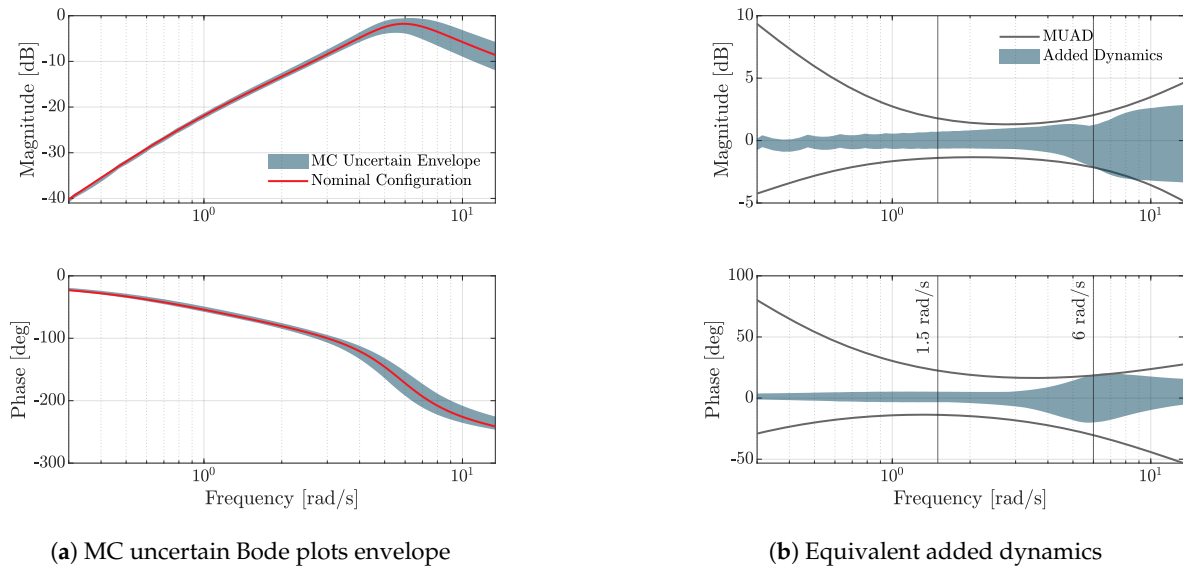
(**b**) Equivalent added dynamics

Figure 20. MUAD analysis on $q/\Delta V_{\mathrm{lon}}$ transfer function, due to MC uncertainty

Although the eTFs remains fully enclosed within the MUAD envelopes for the phase angles, the TFs of several configurations exceed the magnitude boundaries, especially in the central region of the frequency range, where the system is expected to be more sensitive to such dynamics. The increased spread of the wTFs correlates with the degradation of altitude performance, resulting in a high degree of FTM performance variability, as seen in Figure 17. What must be noted is the fact that not all the most critical configurations identified in Section 5.2 showed a TF outside the MUAD range, highlighting the importance of considering a sufficiently complete set of pHQs in the UQ phase of the investigation.

(**a**) Added dynamics

(**b**) Equivalent added dynamics
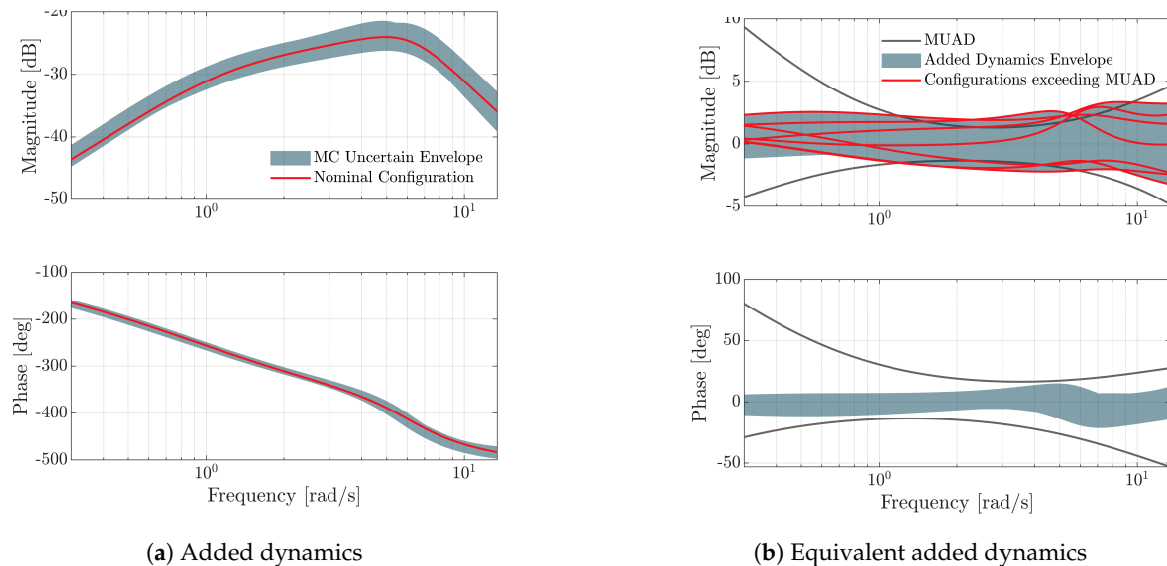
Figure 21. MUAD analysis on $V_z/\Delta V_{\mathrm{lon}}$ transfer function, for the 24 aircraft configurations considered

Once this first assessment has been carried out, the sensitivity of the FSM to added dynamics across the frequency range can be studied, selectively assessing the validity of original MUAD boundaries on its dynamics. This process, which has been widely studied in the literature (see Refs. [31–33]) requires to investigate the aircraft response to several added dynamics, specifically tailored to modify the baseline transfer function at several frequency values. This allows exploring the pilot control sensitivity and adaptation in terms of HQ rating. The simplest approach to perturb the transfer

function is to model the added dynamics H_{add} either as first-order lag-lead and lead-lag dynamics, as in Ref. [32], or through second-order in the form of dipoles [31,33]

$$H_{add}(s) = \frac{s^2 + 2\omega_{dp}\zeta_z s + \omega_{dp}^2}{s^2 + 2\omega_{dp}\zeta_p s + \omega_{dp}^2}. \tag{5}$$

This work employs the second-order form, as it allows to intuitively control the frequency position of the disturbance and gradually increase/decrease the corresponding amplitude. To preserve the shape of H_{add} within the control loop, the dipole is placed in series with the controller's velocity setpoint ΔV_{lon}. Following the approach discussed in Ref. [33], the dipole central frequency value was relocated iteratively between 1 and 6 rad/s (Figure 20(b)) while ζ_z was gradually increased from 0.25 to 0.70. The damping ratio of the denominator was always fixed at $\zeta_p = 0.2$. Figure 22 shows an example of the adopted approach, for both the added dynamics H_{add} and the resulting effect on the baseline TF.

(**a**) Added Dynamics   (**b**) Controlled Dynamics
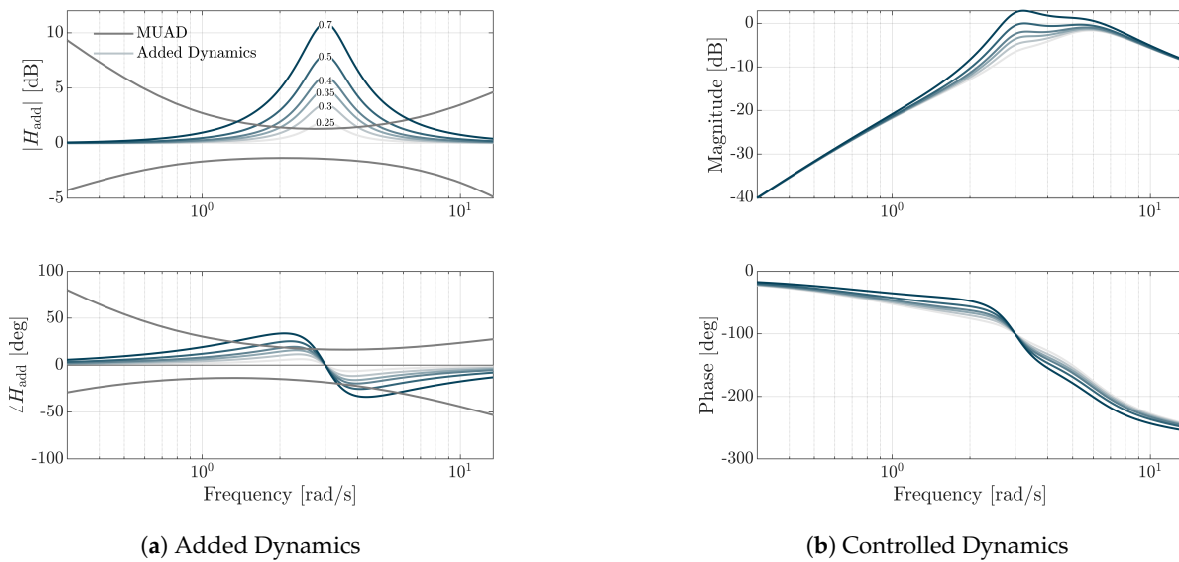
Figure 22. Bode plots of added and resulting dynamics for varying ζ_z (0.25-0.70)

To plan the FTM simulations, three configurations were analysed: the case of NFE operations ($V_{wind} = 0$ m/s) and the cases of headwind ($\psi_{wind} = 0°$), for both $V_{wind} = 0.5$ m/s and $V_{wind} = 1$ m/s. Figure 23 collects the results for altitude performance, which is the most sensitive to variations in H_{add}. For all of the tested operating conditions, a clear dependency upon the dipole's location exists. while in the high part of the frequency range ($\omega \geq 3$ rad/s) H_{add} effect is not driving the rating out of the baseline level, at lower frequencies (in the range 1.5-3 rad/s) the results show the greatest variability, eventually bringing the system outside the adequate boundary. It can also be observed that, while at high frequencies the trend of the performance decreases monotonically with increasing the frequency of ω_{dp} and increases monotonically with the intensity of the added disturbance (i.e. with ζ_z), at low frequencies this monotonicity is lost.

This behaviour can be attributed to two main factors. The first reason can be sought in the effect H_{add} exerts on ΔV_{lon}. When a high-ζ_z low-frequency dynamics is inserted within the control loop, the dynamics of the aircraft is affected to such an extent that the velocity command is inhibited. As a result, the aircraft cannot longer adequately follow the input imposed by the controller, leading to oscillations in both the input and the response that disrupt the FTM. In this context, the observed improvement, visible at 1 and 2 rad/s in Figure 23, should not be interpreted as a beneficial effect of the added dynamics, but rather as a consequence of the piloting algorithm's inability to correctly execute the manoeuvre, ultimately making the results unreliable. Specifically, during the acceleration phase, the eVTOL reaches the required maximum forward speed much more slowly than in the nominal case,

drastically reducing the level of aggressiveness needed for the manoeuvre and both the deceleration aggressiveness and the associated altitude gain. So, the improvement in altitude performance does not correspond to an overall improvement of the FTM performance.

A second interpretation can be sought in the nature of the HQ requirement assigned to the aircraft, as the required FTM input does not contain any high-frequency component and is likely to be mostly affected by dipoles located in the lower part of the frequency range. A different task (e.g. a *pitch-tracking* task for the measurement of control effort, as done in [33]) may result in a different frequency-related behaviour.

It is interesting to note that, as the H_{add} vary and, specifically, as the bandwidth parameter decreases (i.e., approaching the threshold between Level 1 and Level 2 of ADS-33), the degradation in the pHQs, now directly controlled by means of the added TF — rather than indirectly influenced via different aircraft configurations — directly translates into a deterioration in the corresponding assigned HQs, making the relationship between the two HQs dimensions clearer.
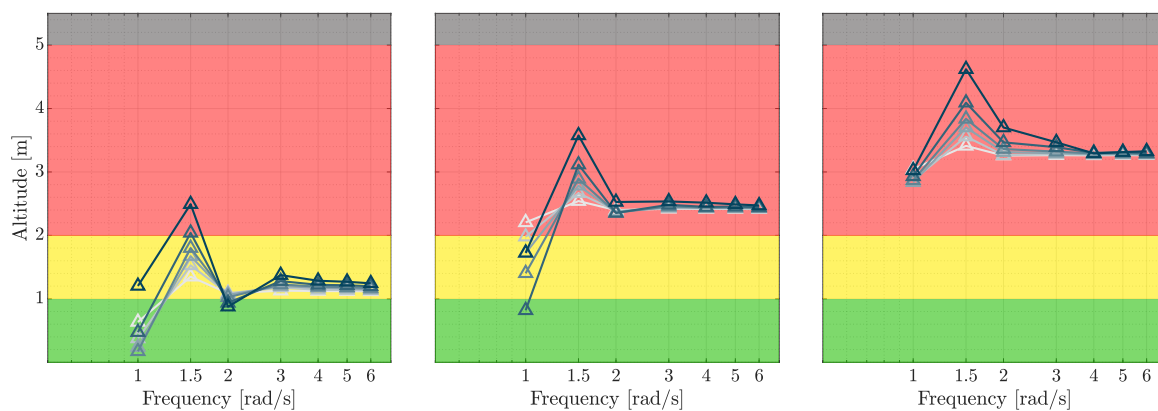
Figure 23. Results for varying H_{add}, No Wind (left), $V_{wind} = 0.5\,\mathrm{m/s}$ (centre), $V_{wind} = 1\,\mathrm{m/s}$ (right)

In summary, this analysis showed that taking the envelope of uncertain models in the frequency domain and comparing it with MUAD limits may be a good way to verify if the uncertainties are able to generate modification that can affect or not the aHQ performance. However, it is important to consider uncertainties that affect a complete set of pHQ and compare all direct and cross transfer functions that affect the FTM with the MUAD limits.

7. Conclusions

This paper successfully proposed and demonstrated a methodology for assessing the credibility of flight simulation models for VTOL handling qualities certification by simulation. The key results obtained are:

- The impact of FSM input parameter uncertainties was effectively propagated through sensitivity analysis (MOAT, VBD) to quantify variations in open-loop predicted HQs and, subsequently, in the aircraft's envelope of transfer functions.
- A novel application of the Maximum Unnoticeable Added Dynamics concept was introduced: by comparing the envelope of uncertain TFs, scaled by a confidence ratio, against MUAD boundaries, a quantitative criterion for FSM credibility for specific HQ-related tasks was established.

For the studied lift+cruise VTOL model and the deceleration to hover flight test manoeuvre the eTFs for longitudinal control dynamics largely satisfied the MUAD limits across the uncertain configurations, suggesting the nominal FSM's credibility for assessing performance aspects primarily driven by these dynamics. However, the eTFs for vertical axis dynamics coupled with longitudinal inputs for several uncertain configurations exceeded MUAD magnitude limits. This finding correlated well with the observed variability and degradation in FTM altitude-keeping performance, indicating

that FSM uncertainty in these coupled dynamics could lead to non-credible predictions for altitude-critical tasks if the nominal model alone were used.

The deliberate introduction of added dynamics in the form of dipoles further confirmed the frequency-dependent sensitivity of FTM performance, particularly altitude control, to deviations from the nominal aircraft dynamics.

These findings are highly promising for the advancement of a standardized certification by simulation approach for VTOL aircraft. The MUAD-based framework offers an objective, engineering-driven method to link FSM fidelity directly to pilot-perceptible dynamic changes. This is a significant step towards reducing reliance on extensive and potentially hazardous flight testing, especially when evaluating off-nominal conditions or the impact of model uncertainties. By demonstrating that an FSM's predictions remain within "unnoticeable" dynamic boundaries despite its own uncertainties, confidence in its use for certification can be substantially increased.

However, to fully mature this approach into a standardized practice, further investigation is crucial in several areas:

- VTOL-Specific MUAD Boundaries. The existing MUAD boundaries were originally conceived for conventional aircraft. Research is needed to tailor, refine, and validate these boundaries for the diverse range of VTOL configurations (e.g., lift+cruise, multicopter, tiltrotor) and their unique critical flight tasks and control strategies.
- Comprehensive Transfer Function Assessment, While this study focused on key longitudinal and coupled vertical dynamics, a standardized approach would require a systematic assessment of all relevant direct and cross-coupling transfer functions that significantly influence critical FTM performance against MUAD limits.

Addressing these areas will solidify the proposed CbS methodology, paving the way for more efficient, cost-effective, and robust certification pathways for the rapidly evolving VTOL aircraft sector.

References

1. Anon.. *Third Publication of Means of Compliance with Special Condition VTOL, MOC-3 SC-VTOL Issue 2.* Cologne, Germany, 2023. EASA.
2. Anon.. *SC-VTOL, Special Condition for small-category VTOL aircraft, Issue 1*, 2019. EASA.
3. Cooper, G.E.; Harper, R.P. The use of pilot rating in the evaluation of aircraft handling qualities. Technical Report TN D-5153, National Aeronautics and Space Administration, 1969.
4. Anon.. *ED-295, Guidance on VTOL Flight Control Handling Qualities Verification*, 2024. EUROCAE.

5.   Anon.. *MIL-DTL-32742(AR), Department of Defense Detail Specification Handling Qualities for Military Rotorcraft*. Redstone Arsenal, Alabama, 2023.

6.   Padfield, G.; van't Hoff, S.; Hofmeister, P.; Lu, L.; White, M.; Quaranta, G. *Rotorcraft Certification by Simulation and Analysis. An Introduction to the Principles and Practices*; Springer, 2025. https://doi.org/10.1007/978-3-031-86398-1.

7.   Bianco-Mengotti, R.; Ragazzi, A.; Del Grande, F.; Cito, G.; Brusa Zappellini, A. AW189 Engine-Off-Landing Certification by Simulation. In Proceedings of the AHS 72nd Annual Forum, West Palm Beach, FL, 2016.

8.   Ragazzi, A.; Bianco Mengotti, R.; Sabato, P.; Afruni, G.; Hyder, C. AW169 loss of tail rotor effectiveness simulation. In Proceedings of the European Rotorcraft Forum, Milano, Italy, 2017.

9.   Wilson, D.; Riley, D. Cooper–Harper pilot rating variability. In Proceedings of the AIAA 16th Atmospheric Flight Mechanics Conference, 1989, pp. 89–3358 CP. https://doi.org/10.2514/6.1989-3358.

10. Anon. Flying Qualities of Piloted Aircraft. Technical Report MIL-STD-1797A, US Department of Defense Interface Standard, 1990.

11. Martinelli, E. Modelling, control, integration and testing of an eVTOL drone. Master's thesis, Politecnico di Milano, Italy, 2022.

12. Bara, K. Aerodynamics, propulsion and testing of an eVTOL drone. Master's thesis, Politecnico di Milano, Italy, 2023.

13. Rylko, A.; Lovera, M.; Quaranta, G. Aerodynamic Model Development and Fidelity Assessment for eVTOL Predicted Handling Qualities. In Proceedings of the 50th European Rotorcraft Forum (ERF 2024), Marseille, France, 2024; pp. 1–14.

14. Tobias, E.L.; Tischler, M.B. A model stitching architecture for continuous full flight-envelope simulation of fixed-wing aircraft and rotorcraft from discrete-point linear models. Technical Report Special Report RDMR-AF-16-01, US Army AMRDEC, 2016.

15. Nabi, H.N.; Visser, C.C.d.; Pavel, M.D.; Quaranta, G. Development of a quasi-linear parameter varying model for a tiltrotor aircraft. *CEAS Aeronautical Journal* **2021**, *12*, 879–894.

16. Martello, N.C. Modelling and integration of an eVTOL UAV. Master's thesis, Master Thesis, Politecnico di Milano, Italy, 2021.

17. Rylko, A.; Quaranta, G.; M., L. Aerodynamic Model Development and Fidelity Assessment for eVTOL Predicted Handling Qualities. In Proceedings of the 50th European Rotorcraft Forum, September 2024.

18. Dalbey, K.; Eldred, M.; Geraci, G.; Jakeman, J.; Maupin, K.; Monschke, J.; Seidl, D.; Swiler, L.; Tran, A.; Menhorn, F.; et al. *Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.14 Theory Manual*. Sandia National Laboratories, Albuquerque, NM, 2021.

19. Gentle, J.E. *Random number generation and Monte Carlo methods*; Vol. 381, Springer, 2003.

20. Saltelli, A.; Ratto, M.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Saisana, M.; Tarantola, S. *Global Sensitivity Analysis*; Wiley, 2008.

21. Campolongo, F.; Saltelli, A. Sensitivity analysis of an environmental model: an application of different analysis methods. *Reliability Engineering & System Safety* **1997**, *57*, 49–69.

22. Anon.. *ASME V&V 20 Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer*; American Society of Mechanical Engineers: New York, NY, 2009.

23. Morris, M.D. Factorial sampling plans for preliminary computational experiments. *Technometrics* **1991**, *33*, 161–174.

24. Adams, B.; Bohnhoff, W.; Dalbey, K.; Ebeida, M.; Eddy, J.; Eldred, M.; Hooper, R.; Hough, P.; Hu, K.; Jakeman, J.; et al. *Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.16 User's Manual*. Sandia National Laboratories, Albuquerque, NM, 2022.

25. Anon.. *Proposed Revisions to Aeronautical Design Standard – 33E (ADS-33E-PRF) Toward ADS-33F-PRF*. U.S. Army Combat Capabilities Development Command, Redstone Arsenal, AL, 2019.

26. Pavel, M.; Padfielf, G. Progress in the development of complementary handling and loading metrics for ADS-33 manoeuvres. In Proceedings of the American Helicopter Society 59th Annual Forum Proceedings, May 6–8, 2003.

27. Anon.. *Test Guide for ADS-33E-PRF*. U.S. Army Research, Development, and Engineering Command, Moffet Field, CA, 2008.

28. Menberg, K.; Heo, Y.; Choudhary, R. Sensitivity analysis methods for building energy models: Comparing computational costs and extractable information. *Energy and Buildings* **2016**, *133*, 433–445. https://doi.org/10.1016/j.enbuild.2016.10.005.

29. Air Force Wright Aeronautical Laboratories, Air Force Systems Command Wright-Patterson Air Force Base, OH. *Background Information and User Guide for MIL-F-878SC, Military Specification - Flying Qualities of Piloted Airplanes*, 1982.

30. McRuer, D.; Jex, H. A Review of Quasi-Linear Pilot Models. *IEEE Transactions on Human Factors in Electronics* **1967**, *HFE-8*, 231–249. https://doi.org/10.1016/j.ifacol.2017.08.2328.

31. Mitchell, D.; He, C.; Strope, K. Determination of Maximum Unnoticeable Added Dynamics. In Proceedings of the AIAA Atmospheric Flight Mechanics Conference and Exhibit, August 21–24, 2006.

32. Butijn, M.; Lu, T.; Pool, D.; .; van Paassen, M. Assessment of Maximum Unnoticeable Added Lag-Lead or Lead-Lag Dynamics with a Cybernetic Approach. In Proceedings of the AIAA Scitech Forum, January 7–11, 2019.

33. Matamoros, I.; Lu, T.; van Paassen, M.; Pool, D. A Cybernetic Analysis of Maximum Unnoticeable Added Dynamics for Different Baseline Controlled Systems. *IFAC-PapersOnLine* **2017**, *50*, 15847–15852. https://doi.org/10.1016/j.ifacol.2017.08.2328.