OUTCOME-BASED FRAMEWORK FOR ONLINE MODEL VALIDATION AND RISK AWARENESS

Juan Jurado
Air Force Test Center
jdjurado@gmail.com

Clark McGehee
Air Force Test Center
clark.mcgehee@gmail.com

1. ABSTRACT

Current Flight Test Continuation Criteria (FTCC) rely on comparing differences between expected and observable aircraft states for safety-of-flight decisions. Even state-of-the-art methods, which can make such comparisons using real-time simulations, are based on subjective thresholds derived from organizational practices. Using Wickert's Risk Awareness framework [1], where system knowledge is the "control parameter opposing drift into a mishap state," this paper proposes a novel method for real-time modeling and analysis of aircraft data to inform whether the aircraft flown is an acceptable version of the simulated aircraft. In this method, the bounds of acceptability (i.e., the Knowledge Envelope) are derived from flight control robustness ground testing. These bounds are directly traceable to the probability of experiencing an unacceptable outcome, such as departure. The proposed method is mathematically formulated, compared to existing state-of-the-art approaches, and applied to specific scenarios from recent envelope expansion programs in the Air Force Test Center.

2. INTRODUCTION

In recent years, computational advancements in aircraft design such as digital engineering have increased the use of high-fidelity models for safety-critical pre-flight predictions before test events. This paper leverages the same computational advances to improve real-time and post-test data analysis tools. Specifically, this paper aims to employ the same means of high-fidelity modeling used for safety-critical pre-flight predictions to increase test efficiency and foster real-time risk awareness. This research seeks to answer questions such as "how do we get better answers from our models?", "Are we asking the right questions from our models?" and "How do we attain rigorous efficiency?"

Arguably, any experimental test endeavor aims to safely explore an unknown domain without harming the system under test or the aircrew. Inevitably, the balance between safety and efficiency comes down to proper management of the safety, cost, and schedule pressures on the test team. In other words, though the safest thing to do would be never to fly an experimental flight test, programmatic goals inherently pressure test teams to accept risk to achieve the overarching program objectives.

In this context, the execution of a flight test program can be described as a continuous balancing of two opposing forces, where programmatic realities drive teams to take on necessary risks

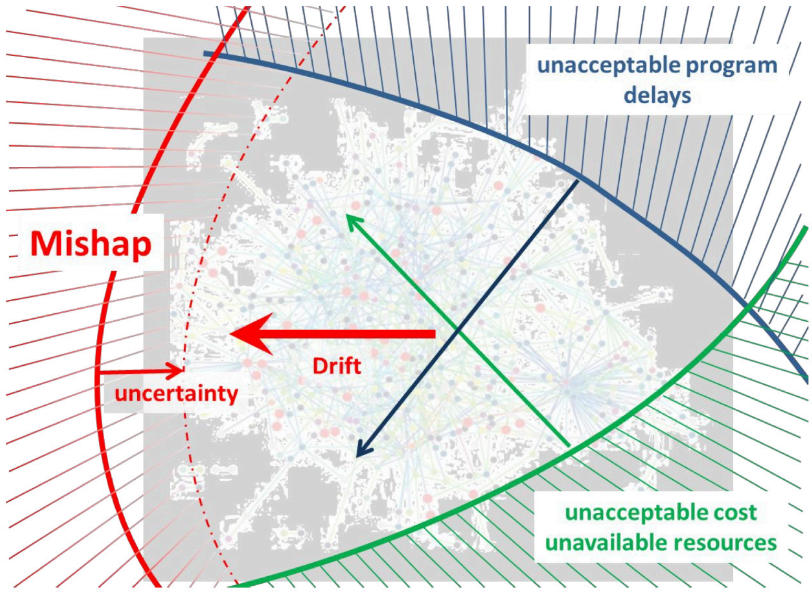

Figure 1: Wickert's uncertainty-based risk awareness model [1].

in the face of uncertainty, and the team's ability to resist drift into a mishap state is dependent on knowledge and risk awareness. As illustrated in Figure 1, Wickert [1] proposes an uncertainty-based risk awareness framework to help teams recognize and resist drift. In this framework, knowledge acts as a control parameter that prevents the phase shift from a no-mishap state into a mishap state, much like temperature as a control parameter preventing ice from phase shifting into water.

In this paper, we wish to operationalize Wickert's risk awareness framework by improving the type and quality of knowledge we draw from existing models. As models used in aircraft development trend towards higher fidelity, our systems under test tend to rely more on their use for safety and robustness. Our flight test tools should guard test teams against the potential misapplication of these increasingly complex models. The methodology proposed herein does precisely that. Though we believe this methodology can be applied to any predict-test-validate cycle, we will focus on envelope expansion examples as they typically represent the most challenging programs in terms of risk awareness and potential mishaps.

Section 3 summarizes current methods for envelope expansion, including state-of-the-art real-time techniques, and identifies their potential drawbacks. Section 4 motivates and derives a proposed improved method that leverages existing high-fidelity models and robustness ground testing. Section 5 highlights results from three notional envelope expansion scenarios with varying levels of model mismatch. Finally, Section 6 offers lessons learned and operational suggestions for applying the proposed framework to future test programs.

3. CURRENT METHODS

In general, envelope expansion test programs rely on the concept of FTCC when exploring new portions of an envelope. An FTCC is an agreed-upon quantitative measure of agreement between predictions and observed flight test results. It informs the risk assumed in continuing

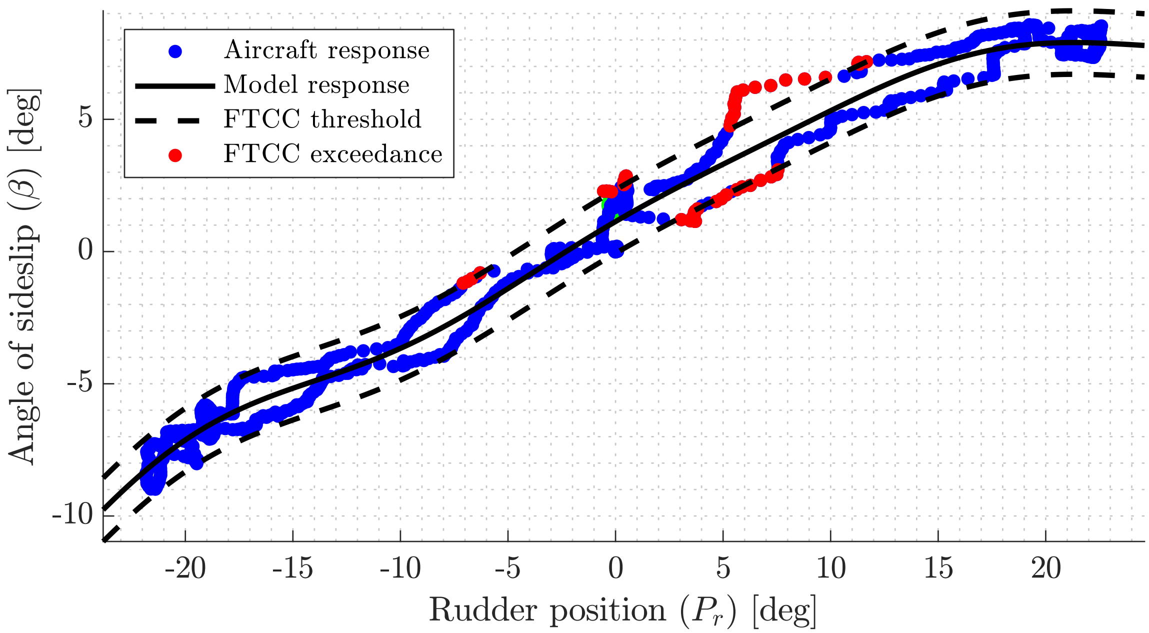

Figure 2: Notional threshold-based FTCC using real-time simulation comparisons.

onto a subsequent test condition. State-of-the-art envelope expansion techniques have recently evolved from static FTCC defined at the test planning stage to real-time aircraft-to-model comparisons in the control room at actual test conditions.

Test teams employing these methods can compare expected (simulator/model) and actual aircraft rates and states (e.g., angle of attack, angle of sideslip, load factor, and roll, pitch, and raw rates) in real-time to decide if the team should move onto the next test point as they clear each portion of the envelope. The FTCC defines the acceptable difference between expected and observed rates and states in these cases. Figure 2 illustrates a notional steady-heading sideslip test point on an instrumented T-38C with observations of Angle of Sideslip (AOS) (**blue**) compared to their expected value (**black**) in real-time. Observations falling outside of a $\pm$ 2-degree FTCC are shown in (**red**) and would potentially result in a required post-flight data review before continuing onto the next test point condition.

Both traditional and state-of-the-art FTCC methods for envelope expansion can be generalized into the information block diagram illustrated in Figure 3. As shown, whether performed in real-time or post-flight, using static or dynamic thresholds, our current criteria for deciding if the team should continue onto the next condition (and implicitly deciding if the risk in continuing is acceptable) rely on a numerical comparison between expected rates and states and observed rates and states. In essence, we are answering two fundamental questions:

1. Is the difference between the expected and observed parameter acceptable?

2. Does the aircraft match (or fly like) the simulator?

These questions are a logical result of the fact that we base flight control laws, test-point matrices, and safety plans around the nominal “design-to” case where the aircraft matches the model or simulator exactly. Though these questions have undoubtedly allowed countless teams to safely and effectively execute envelope expansion programs, we should consider a paradigm

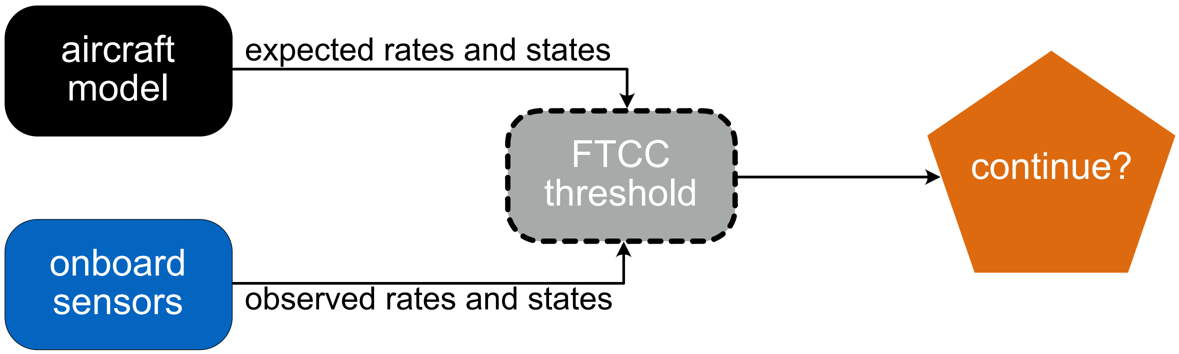

Figure 3: Information block diagram for threshold-based FTCC approaches.

shift in the type of predictions we draw from models so that we may propagate knowledge more quickly and prevent mishaps.

4. PROPOSED METHOD

As alluded to in Section 2, complex aircraft design work has become increasingly dependent on exquisite modeling and simulation to provide performance and flying qualities predictions and, more importantly, to ensure flight control robustness. Following this trend, the flight test community has also become increasingly dependent on model predictions for FTCC. However, even state-of-the-art real-time techniques like the one illustrated in Figure 2 rely on simplistic comparisons (i.e., subtractions) between expected and actual aircraft behavior. Here, we propose a fundamental change to the types of questions we ask of our models so that we may draw more insightful conclusions and improve risk awareness by accelerating the propagation of knowledge throughout the test team.

4.1 Model Transformations

In the context of model matching, let the X-62A VISTA shown in Figure 4 represent the expected or nominal aircraft model. Similarly, let the transformed image of the X-62A illustrated in Figure 5 represent the actual or observed aircraft (i.e., the as-built aircraft). As shown in Figure 3, our current FTCC assesses the immediate difference between expected and observed rates and states. Given that the probability of actually building an aircraft whose observables match the model exactly is low, we are often left to devise and accept arbitrary magnitudes for such differences as surrogates for risk awareness.

To develop a better set of criteria for model matching, let us first consider what is typically meant by a statement such as “the aircraft flies like the sim.” In general, the perception that two aircraft fly “like” each other is controlled by the underlying stability derivative curves, which control the aircraft’s response to specific deviations from trim conditions (via pilot inputs or external forces) [2]. If an aircraft flies like the simulator (i.e., the model), it is likely because the expected and actual stability derivative curves are similarly shaped.

Using this notion, flight control developers typically validate robustness and safety by stressing control laws with varying stability derivative values [3][4]. These varied models are usually generated via rigid transformations [5] of the nominal stability derivative curves. In other words,

Figure 4: Picture of X-62A VISTA aircraft representing "expected" aircraft behavior.

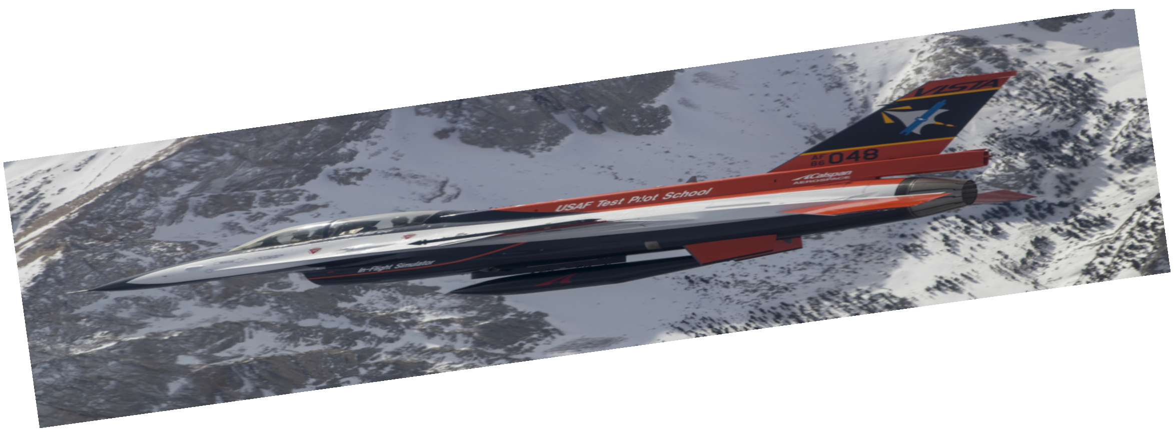

Figure 5: Transformed image of X-62A VISTA aircraft representing "actual" aircraft behavior.

they are biasing and rotating the various stability derivative curves or aeromodel tables by some percentage. In this context, differences between expected and observed aircraft rates and states are viewed as symptoms of an underlying transformation between the simulator or model and the as-built aircraft. Coincidentally, a graphical affine transformation was applied to every pixel in Figure 4 to generate Figure 5.

During flight control robustness verification, aircraft designers typically use (proprietary) Monte-Carlo simulations where simulated Flight Test Techniques (FTTs) are executed using transformations of the nominal stability derivative curves, and robustness outcomes (e.g., aircraft departed or did not depart) are cataloged against the particular variations in the model for each trial. Figure 6 illustrates the notional robustness verification process using a two-dimensional rigid transformation (bias and rotation) of the nominal C_{N_β} curve. As shown, transformations of the C_{N_β} curve leading to a non-departure outcome are shown in **green**, while transformations leading to a departure outcome are shown in **red**. Each transformed curve on the left panel of Figure 6 is represented by a transformation point on the right panel. In other words, the bias-and-rotation

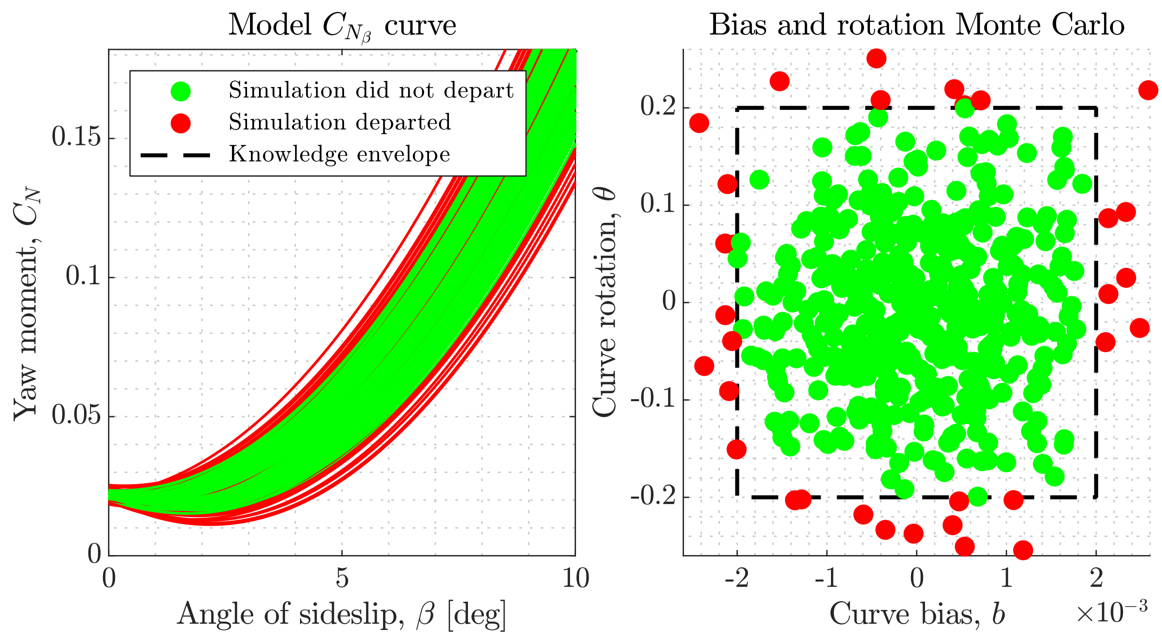

Figure 6: Notional definition of a knowledge envelope from robustness ground test.

tuple applied to generate a curve on the left is plotted as a corresponding point on the right.

As shown in Figure 7, applying a transformation to one or more stability derivative curves constitutes a “new” aircraft model. Given a set of transformed aircraft models and their outcomes in the simulated FTTs, we can draw a boundary around the transformations that generated acceptable outcomes and those that did not. Figures 6 and 7 illustrate such a boundary, henceforth defined as the Knowledge Envelope. Given a Knowledge Envelope containing model transformations with desirable outcomes, our FTCC now switches from comparing expected and observed rates and states to estimating whether the aircraft we are testing is an *acceptable version* of the nominal aircraft.

4.2 Transformation Identification

Rather than computing differences between expected and observed rates and states as shown in Figure 3, we now seek to develop a method that estimates the transformation required to match observed stability derivatives with the nominal aircraft model and determines if the estimated transformation and its uncertainty fall inside the pre-defined Knowledge Envelope, where we expect favorable outcomes. Figure 8 illustrates the modified flow of information enabling such an outcome-based FTCC. As shown, observed rates and states are still used in the analysis. However, instead of directly comparing them to the nominal rates and states, we first generate online estimates of stability derivatives (i.e., we conduct parameter identification). We then compare the estimated parameters to their nominal values to estimate the transformation required to align the two. If the transformation required (and its statistical estimation uncertainty) falls inside the Knowledge Envelope, we expect to achieve a desirable outcome and should continue testing.

Online estimation of stability derivatives (i.e., parameter identification) is a well-studied problem with proven solutions [6]. Still, it is seldom implemented in mission control rooms during flight

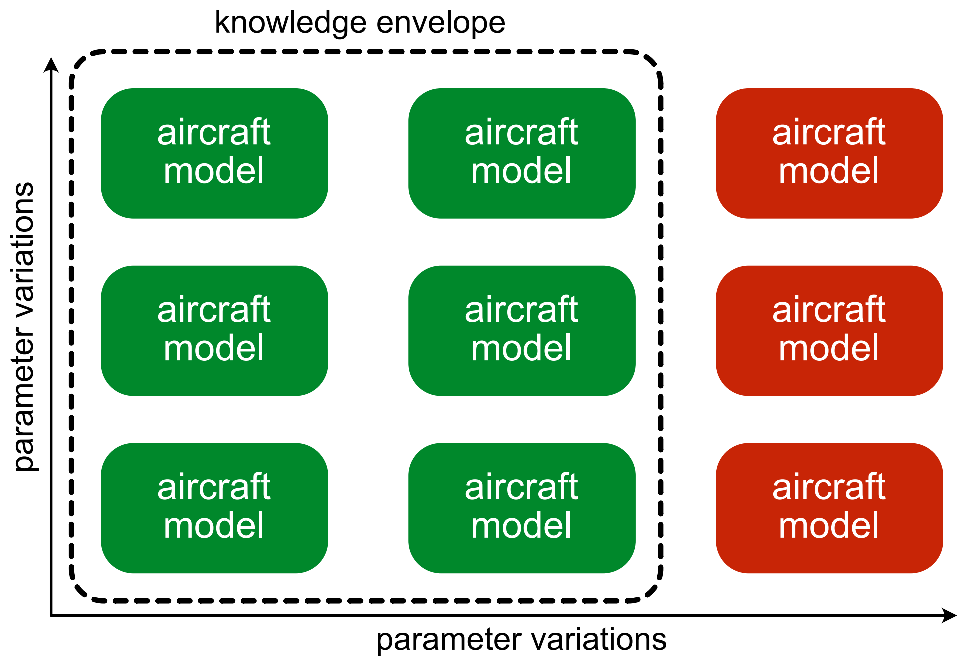

Figure 7: Illustration of knowledge envelope development using robustness ground test.

tests. Assuming this step is accomplished, we propose the use of nonlinear model regression [7][8] to estimate the transformation tuple required to align the nominal model to the observed parameter identification model. Once the transformation tuple is estimated (along with its uncertainty ellipsoid), we verify whether the transformation falls inside the Knowledge Envelope. Using this approach, we can now answer the following questions:

1. Have we seen this aircraft in simulation?

2. How likely is it that it is one of the bad ones?

4.3 Nonlinear Regression

The alignment of nominal and observed stability derivative curves can be described as a simple 2-D rigid transformation problem [9]. Letting $\{n_i\}$ and $\{t_i\}$, $i = 1 \ldots N$ represent corresponding point sets on nominal and transformed stability derivative curves, respectively, we can describe their relationship using

$$\mathbf{t}_i = \mathbf{R}_\theta \mathbf{n}_i + \mathbf{b} + \boldsymbol{\epsilon} \tag{1}$$

where $\mathbf{R}_\theta$ is a standard 2×2 rotation matrix with unknown rotation angle θ, $\mathbf{b}$ is an unknown bias vector, and $\boldsymbol{\epsilon}$ is zero-mean White Gaussian Noise (WGN). As the parameter identification algorithm [6] generates estimates of the observed stability derivative, $\mathbf{t}_i$, we can use the corresponding nominal values for the same derivative, $\mathbf{n}_i$, to estimate the unknown variables: θ and $\mathbf{b}$, using nonlinear model regression [8]. In general, the nonlinear regression problem is set up to minimize the sum of squared error, Σ^2, given by

$$\Sigma^2 = \sum_{i=1}^{N} \|\mathbf{t}_i - \mathbf{R}_{\hat{\theta}} \hat{\mathbf{n}}_i - \hat{\mathbf{b}}\| \tag{2}$$

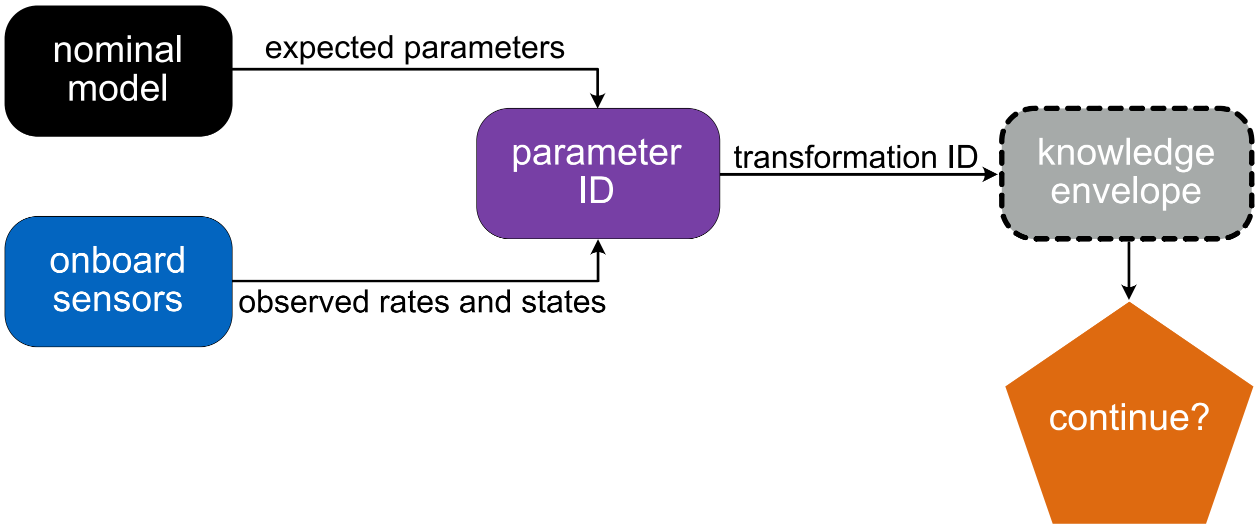

Figure 8: Information block diagram for outcome-based FTCC approach.

where $\hat{\theta}$ and $\hat{\mathrm{b}}$ represent the estimated rotation angle and biasing vector, respectively, that best align the nominal stability derivative shape with the observed shape. Given an estimated model from (2) and its Mean Squared Error (MSE), we can use [7] to compute the coefficient covariance matrix for $\hat{\theta}$ and $\hat{\mathrm{b}}$, and draw an appropriate $(1-\alpha)\%$ uncertainty ellipsoid around the estimated transformation values to account for the variance of ϵ in the parameter identification step.

5. SIMULATION RESULTS

The proposed transformation identification algorithm for FTCC was validated against three envelope expansion scenarios. In all scenarios, the test objective was to expand the directional stability envelope of a notional fighter-sized aircraft. Each scenario is designed such that the observed AOS resulting from a steady-heading sideslip maneuver would have fallen outside the legacy FTCC of ± 2 degrees.

Scenario 1 is illustrated in Figure 9. The parameter identification estimates of C_{N_β} are shown in **purple**. The least-squares model fit line shown in **green** required a transformation from the nominal model shown in **black** of approximately -0.1 degrees and a bias of approximately -1 units. As shown on the right panel, the transformation estimation tuple, along with its 99.9% uncertainty bound, fell well inside the Knowledge Envelope developed during ground robustness testing, indicating there is at least a 99.9% probability that the aircraft flown was seen during ground robustness simulation and exhibited desirable behavior (i.e., did not depart).

Scenario 2 is illustrated in Figure 10. Again, the parameter identification estimates of C_{N_β} are shown in **purple**. The least-squares model fit line shown in **red** required a transformation from the nominal model shown in **black** of approximately 0.03 degrees and a bias of approximately -2 units. As shown on the right panel, the transformation estimation tuple fell inside the Knowledge Envelope, but its 99.9% uncertainty bound fell outside. This indicates at least a 0.01% probability that the aircraft flown exhibited undesirable behavior (i.e., departed) during the ground robustness simulations.

Scenario 3 is illustrated in Figure 11. Again, the parameter identification estimates of C_{N_β} are shown in **purple**. The least-squares model fit line shown in **red** required a transformation from the nominal model shown in **black** of approximately 0.12 degrees and a bias of approximately 1.5 units. As shown on the right panel, the transformation estimation tuple fell inside the Knowledge Envelope, but its 99.9% uncertainty bound fell outside. Here, we also observe a general mismatch in the shape of the parameter estimates compared to the nominal model and the "best" transformed model fit. This behavior is typically referred to as a statistical lack of fit that can be detected by either a growth in the uncertainty ellipse throughout the maneuver (illustrated in the video animation) or a Chi-Squared lack of fit test [7]. This result is especially significant because instead of indicating the aircraft flown may exhibit undesirable behaviors, it implies the aircraft flown was never tested during ground robustness simulations. This aircraft has no predictions, and we find ourselves expanding the Knowledge Envelope in flight.

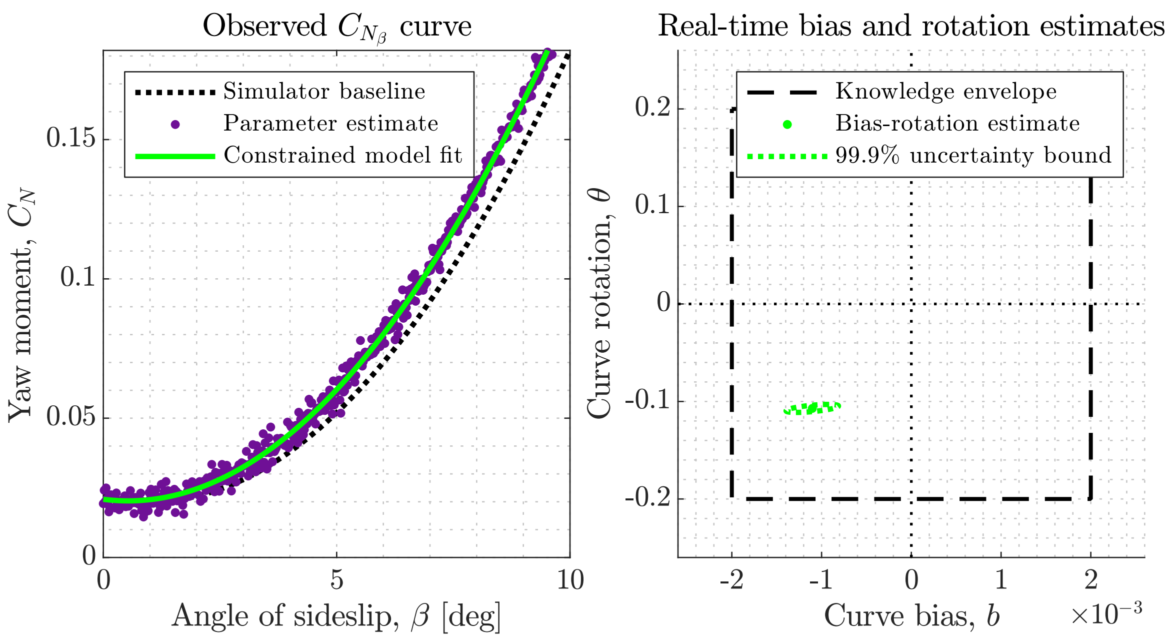

Figure 9: Simulation 1, actual C_{N_β} shows an acceptable transformation from nominal aircraft model.

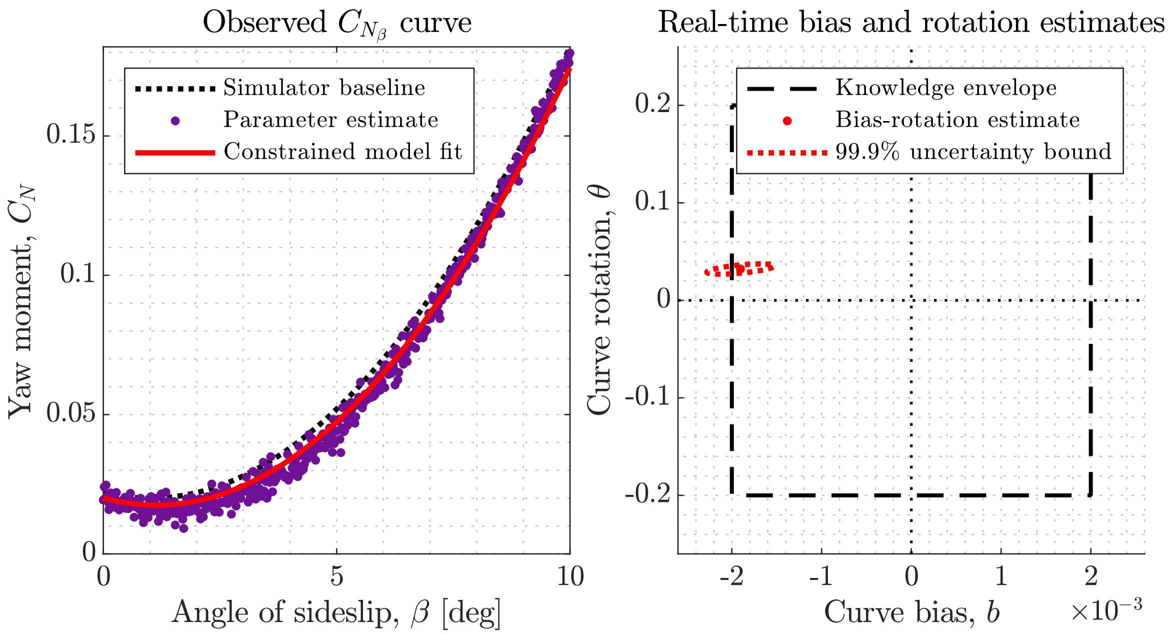

Figure 10: Simulation 2, actual C_{N_β} requires an unacceptable transformation from nominal aircraft model.

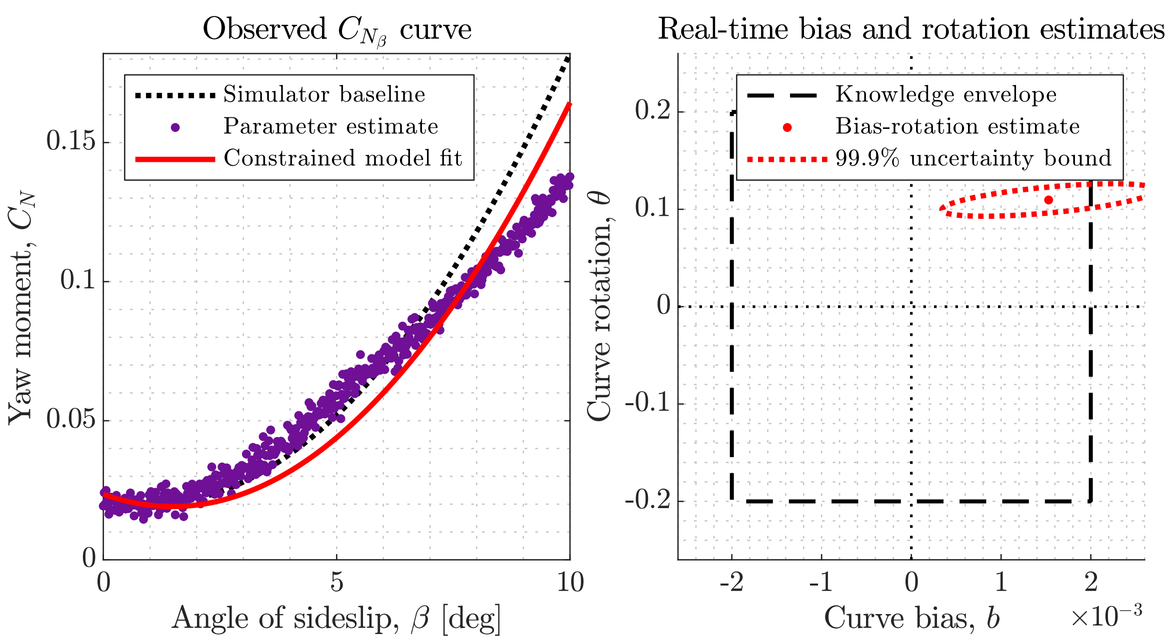

Figure 11: Simulation 3, there is a "lack-of-fit" transformation between actual and nominal C_{N_β}.

6. CONCLUSIONS

This paper proposes an improved method for defining FTCC that leverages existing high-fidelity models available in aircraft design. The proposed method shifts away from direct comparisons of expected-versus-observed rates and states (e.g., angle of attack, angle of sideslip, load factor, and roll, pitch, and raw rates) and toward parameter identification of the underlying stability derivative curves. More importantly, it estimates the transformation required to match the observed aircraft's behavior with the nominal aircraft design. Using this approach, the implicit questions answered by an FTCC evolved from "Does the aircraft match the model?" to "Have we seen this version of the aircraft in simulation?"

The proposed method defines a Knowledge Envelope using results from the ground robustness simulations typically conducted by flight control designers. Outside this envelope, "versions" of the nominal aircraft are known to have unacceptable behavior, such as instability or poor handling qualities. Using a combination of parameter identification and model transformation estimation, the stability derivatives of the as-built aircraft were compared to the Knowledge Envelope, producing a traceable probability of experiencing an undesirable outcome or hazard.

Three simulated envelope expansion scenarios were used to validate the proposed method. In all scenarios, the legacy "rates and states" FTCC was exceeded, while the stability derivative Knowledge Envelope methodology yielded more nuanced results with improved risk awareness. In the last scenario, the proposed method was used to identify that the as-built aircraft's stability derivatives significantly differed from any transformations attempted during ground robustness simulations. This highlights the ability of this methodology to alert test teams to the lack of valid predictions for the system as a whole, regardless of the differences between expected and observed rates and states.

7. REFERENCES

[1] D. Wickert, "Risk awareness: A new framework for risk management in flight test," in *Proceedings to the SETP 62nd Annual Symposium*, 2018.

[2] R. F. Stengel, *Flight Dynamics*. Princeton University Press, 2022.

[3] J. Roskam, *Airplane flight dynamics and automatic flight controls*. DARcorporation, 1995.

[4] R. Bernard, J.-J. Aubert, P. Bieber, C. Merlini, and S. Metge, "Experiments in model based safety analysis: Flight controls," *IFAC Proceedings Volumes*, vol. 40, no. 6, pp. 43–48, 2007.

[5] M. Berger, *Geometry I*. Springer Science & Business Media, 2009.

[6] G. Chowdhary and R. Jategaonkar, "Aerodynamic parameter estimation from flight data applying extended and unscented kalman filter," *Aerospace science and technology*, vol. 14, no. 2, pp. 106–117, 2010.

[7] M. H. Kutner, C. Nachtsheim, and J. Neter, *Applied Linear Regression Models*, vol. 4. McGraw-Hill/Irwin, 2004.

[8] D. M. Bates and D. G. Watts, *Nonlinear regression analysis and its applications*. Wiley series in probability and mathematical statistics. Applied probability and statistics, New York, Chichester: J. Wiley, 1988. Includes indexes.

[9] D. W. Eggert, A. Lorusso, and R. B. Fisher, “Estimating 3-d rigid body transformations: a comparison of four major algorithms,” *Machine vision and applications*, vol. 9, no. 5, pp. 272–290, 1997.

8. BIOGRAPHIES

Juan Jurado is a Flight Test Squadron Commander for the U.S Air Force Test Center. He was previously the Director of Education at the U.S. Air Force Test Pilot School. He holds a B.S.EE from Texas A&M University, an M.S.FTE from the Air Force Test Pilot School, and an M.S.EE and Ph.D. from the Air Force Institute of Technology. His research interests include real-time modeling and model validation, online sensor calibration, image processing, visual-inertial navigation, and statistical sensor management for multi-sensor navigation problems.

Clark McGehee is a student at the United States Army Command and General Staff College. He was previously the Director of Operations for a flight test squadron in the Air Force Test Center. He holds a B.S.E. in Mechanical Engineering from Duke University, an M.S.FTE from the Air Force Test Pilot School, and M.S. and Ph.D. in Mechanical Engineering from Duke University. His research interests include real-time modeling and model validation, air data, and advanced flight control systems.