Aerospace Technology Congress
11-12 October 2016, Solna, Stockholm

Data-driven Pilot Behavior Modeling Applied to a V_{MCG} Determination Flight Test Task

F. M. S Turetta*, H.V.H. Ayala*, L.G. Trabasso** and J. Alfredson***.

* EMBRAER, Department of Systems Modeling and Simulation.
**ITA, Department of Mechanical Engineering.
***SAAB, Aeronautics Human Machine Interaction & LiU, Department of Computer and Information Science.

ABSTRACT
Human models have been studied and used in engineering analysis for over 70 years to allow predictions of the pilot-vehicle system behavior. The difficulties in pilot modeling are evident due to the complexity of the brain, lack of repeatability in behavior and the great number of variables that can affect the human performance. This complexity, associated with the fact that there are no explicit laws to allow modeling based in first principles, could indicate that data-driven modeling techniques would be the most efficient way to obtain pilot models, such as black-box system identification methods that construct dynamic models according to measured input and output data, and where the parameters have no physical meaning. With this approach, it is advantageous to seek knowledge from other fields to allow a better understanding of the pilot behavior, select adequate input/output variables and define the experimental conditions and data. Criteria for evaluating the modeling approaches include adaptability as well as feasibility. Adaptability concerns coping with dynamic and uncertain conditions and feasibility refers to the models contribution to an applied context. This paper presents the results of the application of data-driven theoretical linear dynamic models in the task of representing the behavior of the pilot trying to keep the centerline of the runway after an engine failure. Real data is used, where PID with anti-windup and Hammerstein-Wiener model structures are compared. Results show that the Hammerstein-Wiener structure seems more appropriate to represent this specific behavior.

1. INTRODUCTION
Aircraft design as all other complex systems design is on a continuous trend of increasing the use of modeling and simulation techniques. The goal of this trend is to be able to detect design issues as early as possible in the product lifecycle in order to reduce costs, shorten time to market and increase maturity on entry into service. Simulation has shown to be an efficient tool for that, allowing performing tests and validations before even the first metal cut, and in addition to that, to perform automated tests in a scale that would never be possible without the use of computers.

It is known, however, that human operators should perform a large part of the modeling validation. This does not diminishes the part of modeling and simulation in bringing test capabilities to early stages of the design as simulation with the pilot in the loop are widely used in the industry. But the need for the pilot in the loop prevents from automating tests and performing them in a large scale, consequently. In such scenario, the need for a mathematical representation of the human pilot arises. It is evident that science is very far from being able to represent all the complex human behavior with mathematical equations, but in some specific cases, it is possible to isolate a specific action that is required for an engineering test. If this is possible, it is also possible to automate tests that require that specific human behavior, by implementing a mathematical abstraction of the pilot's actions.

Mathematical models of the human pilot are not new to science; in fact, they have been used for over 70 years to allow predictions of the pilot-vehicle system behavior (McRuer, 1967) with linear and quasi-linear models. Those models evolved to optimal control models in the 1970s by Wierenga (1970) and Kleinman (1969), and then to nonlinear models such as Hess´s proposal of 'pulsive control behaviour' (1979). Other disciplines in addition to control, from biomechanical and vibration analysis to sensorial and perception have also been using pilot models. Lone and Cooke (2014) did an extensive review of those models.

Although all this development has been made, their use in the well established Validation & Verification process from the Systems Engineering perspective appears to be quite new. Regarding the difficulty of model validation, Lone and Cooke (2014) already mentioned this: regardless of the abstraction done to obtain the model, it could never be validated against the pilots mind, only with a black box approach. Still, the abstractions of the pilots' minds are useful, because they are necessary - at least - to identify inputs and outputs. Yet, the mathematical structure is flexible and validation metrics should be taken into account to determine them.

Human-in-the-loop simulator based design has been used by the aviation industry since the 1970s for cost effective, safer design and creating applied knowledge early in the design process (Alm, 2007). Pilots in simulators can be used both in the process of generating a model for a specific scenario and for validating the model. This paper proposes to explore this approach to a specific test that is V_{MCG} (Minimum Control Speed on the Ground) determination of an aircraft. In other words, the behavior modeled is the rudder pedal input by the pilot in reaction to an engine failure, in order to maintain the runway centerline. The scope is to apply data-driven techniques, more specifically the ones based on the formalism of system identification, to identify a pilot model based on data. Real human pilots on an aircraft were used to generate datasets; then models were generated and validated in a simulation environment, reproducing the maneuver. The contribution of this work is evaluating a new form of obtaining pilot models with little or no previous use, and allowing to use those models to automate systems development tests that otherwise would require a human pilot in the loop and an expensive infrastructure.

This paper is organized as follows: Section 2 contains the behavioral model development rationale, Section 3 presents the results and discussion using the identified models; the conclusions and further work are the contents of Section 4.

2. BEHAVIORAL MODEL DEVELOPMENT

This section consists of four parts. First, there is a discussion on the cognitive architecture and process performed by the human pilot during the event of an engine failure during takeoff. This is required in order to identify the relevant parameters that should be considered as inputs/outputs and the structure of the model. The discussion on the design of the experiment to obtain data for the identification process is the content of part two. Data already available from other experiments and flight tests are also considered. Third, the data obtained from the experiment execution is qualitatively analyzed against the real data and data usage decisions are made. Fourthly, the model identification possibilities and adaptations to its structure and representation are discussed. The model developed, its validation and results are presented in Section 3.

2.1 Cognitive architecture and its influence on the model structure and representation

As stated earlier, the model to be identified is supposed to represent the human pilot behavior only in the specific scenario to be analyzed. This scenario is a takeoff run followed by an engine failure with corrective pilot inputs to minimize the runway centerline deviation. Figure 1 shows the physics involved in this situation for a dual engine aircraft, that is, how the thrust asymmetry generated by failure of one of the engines creates a yawing moment due to the distance of the working engine to the aircraft centerline. The 30 ft deviation shown in Figure 1 is a reference to the certification requirement that states the maximum deviation from the runway centerline following and engine failure should be 30 ft (FAR, 2002). This value is used to define the aircraft V_{MCG} that, in turn, impacts on minimum V_1 which has an influence on aircraft performance. In other words, this shows how important it is for the design to guarantee that the pilot is able to maintain this maximum deviation and illustrates an automatic test that could be used in early design phases, where aircraft tail volume and other characteristics are feasible to change.

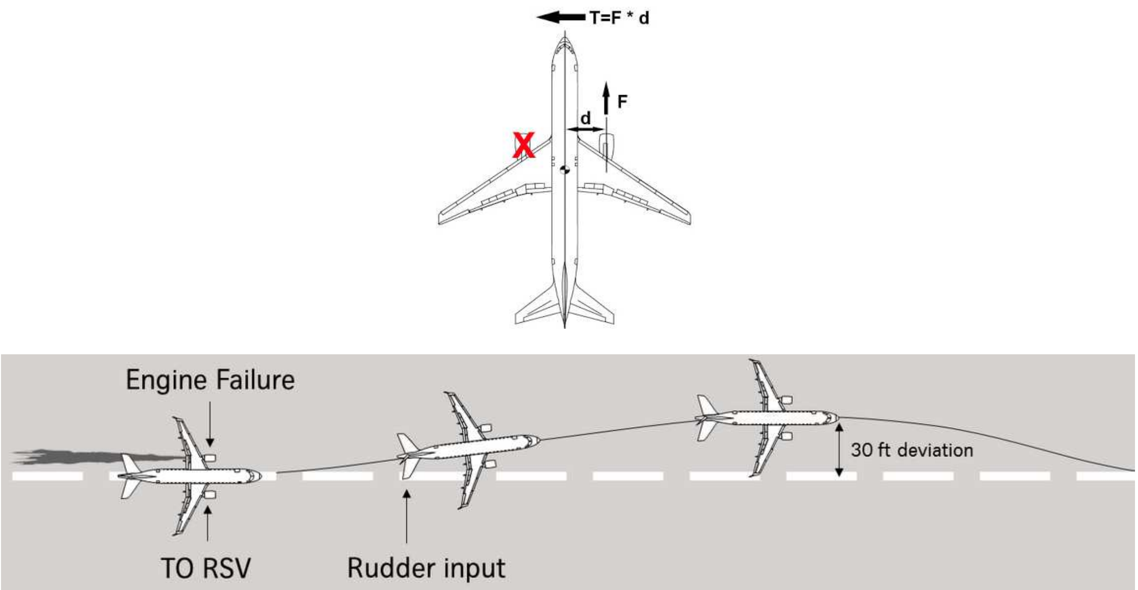

Figure 1: Centerline deviation due to engine failure

Within the described scenario, the questions relevant to the cognitive architecture (and consequently to the pilot model identification) are: which are the variables the pilot is monitoring? How does s/he uses each one of them to process his/her reaction? What is the difference considering dual and single pilot operations? Which equipment such as head up display affects his/her performance and how? Initially, a single variable is selected and the model is based on it. Later complexity may be added to the model in order to make it more accurate and representative. Thus, the model proposed is a single input single output model as depicted in Figure 2.

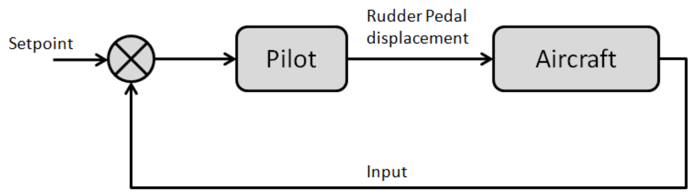

Figure 2: Single Input Single Output Pilot Model

The model output is the rudder pedal displacement, but the input the pilot uses to process his/her reaction is not trivial to enumerate. On the one hand, one could think of using as input the centerline deviation, but it was discarded because although the pilot tries to minimize it, s/he is not really controlling it to a defined set point, as long as the aircraft is aligned with the runway. On the other hand, the angle between aircraft heading and the runway centerline is controlled to remain zero by the pilot. It seemed like a good selection, but since it requires the use of the runway heading which is "outside" of the aircraft, it would inconvenient to implement from the modeling standpoint. Therefore, the alternative was to use the yaw rate that does not depend on variables on other referential systems. Intuitively, the pilot reaction appears to be linked primarily to the yaw rate, rather than to the angle between the centerline and the aircraft nose. The drawback of this choice is that the yaw rate set to zero does not imply that the aircraft is aligned with the runway. Conveniently, this is solved simply by using the integral of the yaw rate, which is null whenever the simulation begins with the aircraft aligned with the runway and retains the feature of apparent proportionally to the pilot's reaction.

Another important consideration is the mathematical representation to be used. For instance, is the pilot response a linear with the selected input? As already mentioned, PID models have been used to represent pilot behavior implying that the response is linear, but could it be represented by another linear representation such as an ARX (Auto Regressive with Exogenous Inputs) model? In addition, it seems intuitive that a human operator is able to adapt to rapidly changing conditions switching the "gains" or even the "structure" of the model.

Additionally although most of the dynamic behavior may be linear, there might be inherent non-linearity to the man-machine interface. For instance, in the model proposed herein there is at least one static nonlinearity which is a saturation of the rudder pedal output imposed by its mechanical stopper. It could be argued that this circumstance originates from the aircraft, not the pilot. It could be considered that the pilot intuitively increases the force against the pedal stopper hoping that the "rudder would deflect more" (this behavior can be seen in daily life events, like when using a remote control with low battery, where we press the button strongly hoping the TV will respond!). As the pilot is readily able to revert his/her command when the desired response is achieved, s/he does not appear to suffer integration windup effects, thus when using PID-like structures the model would need to incorporate at least some kind of anti-windup feature.

As an initial approach, after data collection two different mathematical representations were tried: a non-linear Hammerstein-Wiener model and a PID with anti windup. All data processing was performed off-line using a commercial software package (MATLAB™ 2015b, 2015).The identification algorithms were obtained in the System Identification MATLAB™ toolbox.

2.2 Design of the experiment

Having defined the input as the [Yaw Rate Integral] and the output as the [Rudder Pedal Displacement], the difficulties in designing and experiment for pilot identification are still numerous. From the strict system identification point of view, the input signal should be persistently exciting, so that all the frequency and amplitudes responses of interest are exited (Aguirre, 2015). Since there is a human pilot in the loop of the experiment, it is unpractical and too expensive to perform a large number of test points. Also using complex and random input signals would be very far from the real operational scenario of interest, which is, basically, a large 'step-like' moment on the aircraft due to an engine failure or crosswind gust generating a deviation from the centerline. In scenarios like this, engineering judgment and prior knowledge is used to design the input signal. Those considerations are limitations to be taken into account during the use of the model (Billings, 2013).

Other variables potentially affecting the experiment results should be controlled and the ones that cannot be controlled should be randomized (Montgomery, 2001). Based upon previous practical experiments, it has been decided to keep fixed the following variables: Aircraft type; Wind conditions = zero and Pilot.

The variation from pilot to pilot is indeed a variable of interest, but at this stage of the research, the goal is to confirm if an acceptably representative model for a given pilot can be identified. In addition, there are evidences in Turetta (2013) that well trained pilots do not have such a large variation in performance.

The speed at which the engine failure occurs is also of great interest as it has a very complex relationship with the pilot and the aircraft response. At greater speeds, the pilot will probably be more vigilant, but the rudder efficiency will be greater and so will the yaw rate at the moment of the engine failure. This prompts the following questions: will a model identified in a range of speeds work in different speed ranges? Should the model identification be carried out at different speeds? If so, how to accommodate this requirement with the system identification techniques?

Based upon the considerations above, the designed experiment consists of a simulated takeoff with engine failure, with a human pilot in the loop, as detailed in Table 1. This design is advantageous as it is a normal procedure trained in flight simulators by pilots as well as it is representative of a real operational scenario.

Table 1. Experiment data sample.

Test Parameters		Test Results	
Pilot	Engine Failure speed	Rudder Input	Yaw Rate
Pilot A	70 [kts]	Time series X	Time series X´
	80 [kts]	Time series Y	Time series Y´
	90 [kts]	Time series Z	Time series Z´

2.3 Experiment execution and real flight test data

Real data from flight tests was used in the model identification. Two suitable sets of rejected takeoffs were selected for the present study. They were all performed by the same pilot (an experienced flight test pilot with approximately 6000 flight hours), on the same day, in the same aircraft and same environmental conditions.

2.4 Model Identification

With the yielded data, the first approach to identification was to use the Least Squares method (Luenberger, 1996) to identify a polynomial model, but this was proved ineffective and the two approached bellow were used.

Nonlinear Hammerstein-Wiener Model

The proposed system has a static non linearity at the output, which is the saturation due to the rudder pedal mechanical stopper. This can be represented by using a Hammerstein-Wiener model which composed of a linear dynamic system with two static nonlinearities at its input and output. Figure 3 represents the model structure. In the present case, we set solely the input nonlinearity as a saturation.

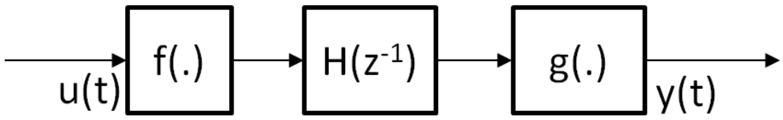

Figure 3: Hammerstein-Wiener model structure. Note that the dynamic part is linear. The nonlinearities are defined by both static functions f(.) and g(.)

As for the order of the linear part of the model, Tustin (1947) proposed that a PID controller could emulate the human pilot behavior, and that information was used as a first estimative for the model order (equation 3).

PID anti windup model

It is possible to represent a PID as a discrete difference equation. The discrete PID equation using backward Euler (Ogata, 1995) is:

$$C(z) = K_p + \frac{K_i T_s z}{z-1} + \frac{K_d(z-1)}{T_s z} \quad (1)$$

which can be manipulated into:

$$C(z) = \frac{U(z)}{E(z)} = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{a_0 + a_1 z^{-1} + a_2 z^{-2}} \quad (2)$$

where:

$b_0 = K_p(1 + NT_s) + K_i T_s(1 + NT_s) + K_d N$

$b_1 = -(K_p(2 + NT_s) + K_iT_s + 2K_dN)$

$b_2 = K_p + K_dN$

$a_0 = (1 + NT_s)$

$a_1 = -(2 + NT_s)$

$a_2 = 1$

which can be rearranged to:

$$u[k] = -\frac{a_1}{a_0}u[k-1] - -\frac{a_2}{a_0}u[k-2] + \frac{b_0}{a_0}e[k] + \frac{b_1}{a_0}e[k-1] + \frac{b_2}{a_0}e[k-2] \tag{3}$$

Equation (3) can be used with identification techniques in order to obtain the parameters that multiply u(k-1), u(k-2), e(k), e(k-1) and e(k-2).

In order to use the discrete PID , a model with anti-windup was implemented and a global optimization algorithm was used to try to fit the data to the structure, identifying the values of Kp, Ki and Kd. The anti windup was implemented by checking if the control action has reached the saturation value, and if it did, zeroing the integrator input.

The results of the identification based in the model structures aforementioned are presented in the Section 3.

3. RESULTS AND DISCUSSION

In this section, the results from the identification procedures are shown and discussed. The two data sets depicted in Figure 3 were used in the process, one for identification and the other for validation of the model.

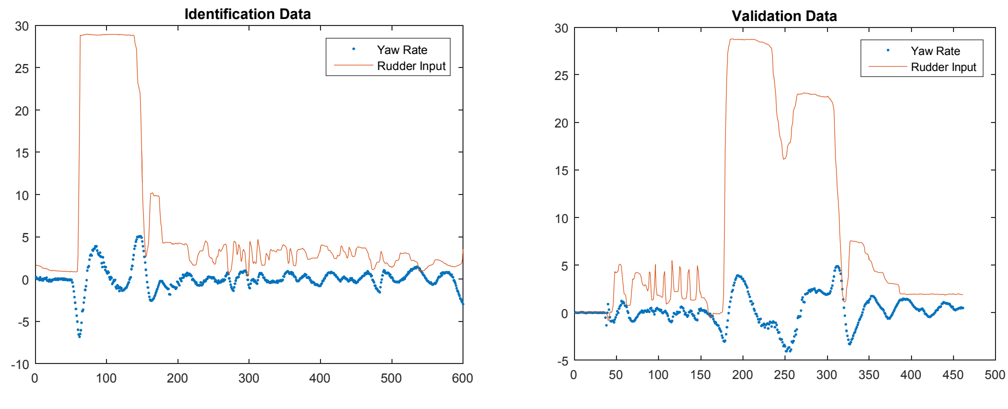

Figure 3: Datasets used in the identification process.

Non Linear Hammerstein-Wiener (NLHW) model

The initial NLHW model prompted good results with fits of almost 80% as can be seen in Figure 4.

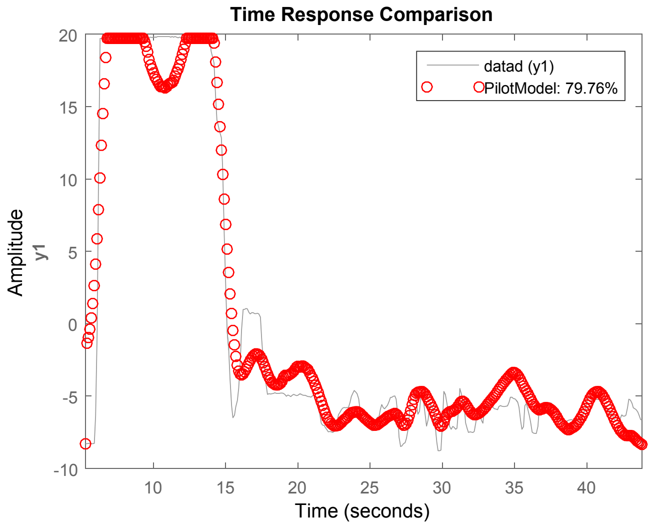

Figure 4: NLWH model (3/2 order).

The orders of the model above were selected to comply to Equation (3), but the coefficients did not match the structure – the gains identified with the least squares method did not reflect the relation between the coefficients a_0, a_1, a_2, b_0, b_1 and b_2 (notice that all coefficients are divided by a_0). This means that even though the identified model may have a good fit to the data (meaning it does represent the behavior), it does not necessarily behave as a PID controller.

Table 2. Fitting characteristics with different orders.

NLHW Compared with Identification data			
Orders 3/2	Orders 4/3	Orders 5/4	Orders 6/5
Correlation=0.9800 R^2=0.9590 Fit=79.76%	Correlation=0.9812 R^2=0.9610 Fit=80.24%	Correlation=0.9915 R^2=0.9829 Fit=86.91%	Correlation=0.9821 R^2=0.9624 Fit=80.61%

Since the PID structure was not necessarily followed, an attempt to raise the order of the model was made and better results were attained as can be seen in Figure 5. Raising the model order increased the fit up to a certain point (fifth order for B(z) and fourth order for F(z)), and from that point on the fit would start to decrease again as can be seen in table 2.

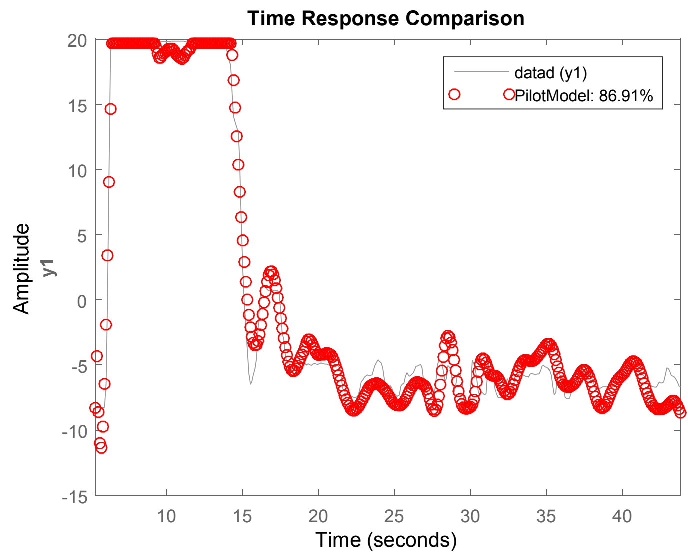

Figure 5: Higher order (5/4) NLWH model.

When the model identified was applied to the validation dataset, the fit decreased considerably. Figure 6 shows the higher order model from Figure 5 applied to the dataset of Test 3.

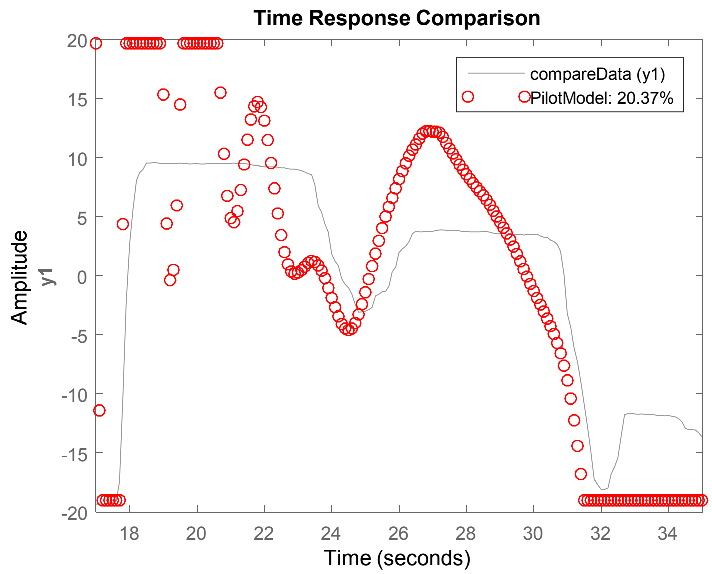

Figure 6: Higher order NLWH model from Figure 5 applied to validation data.

The decrease in the model fit is expected due to many reasons such as signal noise and fitting issues, but considering that the object being modeled is a human, it is highly expected that the reaction would differ slightly from one situation to another, as the human reaction is not deterministic like a mathematical model. The open questions are: first, if this variation is representative of that particular subject, and secondly, if it is worth including a stochastic portion in the model to account for this kind of variations.

The answer to those questions depends greatly on the application intended for the model.

PID anti-windup Model

Identifying a PID anti-windup model prompted good results like those obtained with the NLHW model as can be seen in Figure 7.

The identified gains were Kp = -2.331226, Ki=-1.633316 and Kd=-1.471163.

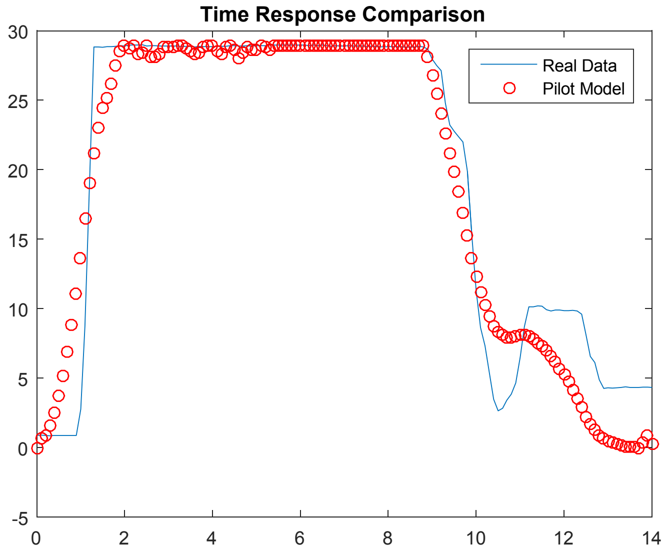

Figure 7: PID anti windup model

When the model was applied to the validation dataset, both the correlation and the R^2 decreased to 0.1926 and near zero (Figure 8). This may indicate that linear PID models do not cope with significant data variations, at least initially it appears that NLHW models are slightly more robust. It is interesting to validate this hypothesis in the future with simulations. Table 2 compares the performance of the two models.

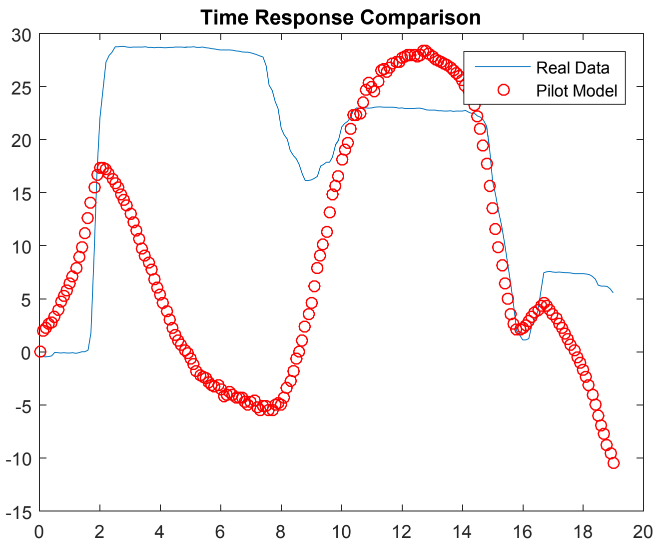

Figure 8: PID anti windup model identified applied to validation data.

Table 3. Models Comparison.

NLHW (Higher order)		PID - Anti Windup	
Compared with Identification data	Compared with Validation data	Compared with Identification data	Compared with Validation data
Correlation=0.9915	Correlation=0.8417	Correlation=0.9646	Correlation=0.2764
R^2=0.9829	R^2=0.3658	R^2=0.9245	$R^2 \approx 0$
Fit=86.91%	Fit=20.37%	Fit=72.52%	Fit=-(63,54)%

4. CONCLUSIONS

This paper has presented a brief rationale supporting the development of mathematical representations of pilot behavior. It tested different representations to obtain the models abstracted from the pilot´s cognitive process to identify inputs and outputs and using system identification techniques to obtain the mathematical model.

The difficulties in pilot modeling are evident due to the complexity of the brain, lack of repeatability in behavior and the great number of variables that can affect the human performance.

Within the modeling scenario, adaptability concerns coping with dynamic and uncertain conditions and feasibility refers to the models contribution to an applied context. Regarding feasibility, the VMCG testing application has been shown to be a viable way of using this kind of model in the development process. Adaptability appears to be more difficult to reproduce as could be seen in the model application to different datasets. Even though the two datasets used for estimation and validation were obtained using the same test pilot and the correlation metrics indicate some level of model adherence, the identified models could be further improved, most prominently in the validation phase. Moreover, it seems that the pilot did not behave in the same manner in different occasions, what could explain the worse performance in the validation phase. This raises the discussion regarding human intra-subject variability, which means the same pilot exposed to the same situation may react differently in different days, due to a variety of reasons (physiological, psychological, social etc.). Hollnagel and Woods (2005) have proposed four modes of interaction applicable to pilot and the cockpit, namely, scrambled, tactical, opportunistic and strategic. Working with more data would make it possible to fit the observed behavior in one of these modes. Another relevant question is if these variations are in the scope of interest in which the model is used. In the proposed application, maybe this kind of accuracy is not needed to see if a regular pilot would be able to perform that task. Maybe other applications may require that kind of behavior.

This paper supports the Tustin´s hypothesis that the human behavior can be approximated by a PID and that apparently there is no need to stick strictly to the PID structure, as became evident during the identification of the NLWH model, when raising the order of the model did improve the fitting results. Table 3 clearly shows that the NLWH model is more robust to changes in the simulation scenario than the PID model, as it can be seen that all the validation metrics decrease considerably less in the NLWH model when the model is tested against validation data (specially the correlation with the value of 0.8417). This confirms in part that the cognitive architecture abstraction is important to identify inputs and outputs, but not necessarily regarding the mathematical representation and that system identification is a viable tool to obtain those models.

Further research

In the field of identification and model validation there are vast expansions to this initial work, such as improving the experiment execution and investigating which variables do impact on the model qualities and characteristics. In addition, validation of the model with more simulations of different situations, with different speeds, aircraft models, pilots etc.

Using different mathematical representations is also a fertile field of research. Neural Networks are a potential candidate, but considering that fairly good results could be obtained with linear models, best linear approximations (Castro-Garcia et all, 2015) in addition to the nonlinear dynamic structures could be a good tradeoff between simplicity and performance.

Regarding future applications for such models it possible to use this or similar models for detection and identification of an event (in the studied scenario this is when the engine failure causes the need for control input, or that a change in the design has created a undesirable handling condition), as well as compensating/controlling based on the model.

REFERENCES

Book: Aguirre, L.A. (2015). Introdução a identificação de sistemas: técnicas lineares e não lineares: teoria e aplicação. Editora UFMG. Belo Horizonte. In Portuguese.

Book: Billings, S.A. (2013). Nonlinear system identification: NARMAX methods in the time, frequency and spatio-temporal domains. Chichester: John Wiley & Sons.

Book: Hollnagel, E. & Woods, D. D. (2005). Joint cognitive systems: Foundations of cognitive systems engineering. Boca Raton, FL: Taylor & Francis.

Book: Ljung, L. (1999). System Identification: Theory for the User. 2nd ed., Upper Saddle River: Prentice Hall, N.J.

Book: Luenberger, D. G. (1996). Optimization by Vector Space Methods. Wiley-Interscience.

Book: Montgomery, D. C. (2001). Design and Analysis of Experiments. John Wiley & Sons Ltd. New York.

Book: Ogata, K. (1998). Discrete-Time Control Systems. Prentice-Hall International.

Conference: Castro-Garcia, R. Tiels, K. Schoukens, J. Suykens, J.A.K. (2015). Incorporating Best Linear Approximation within LS-SVM-Based Hammerstein System Identification IEEE Conference on Decision and Control (CDC), Osaka.

Conference: McRuer, D. T. and Jex, H. R. (1967). Review of Quasi-Linear Pilot Models IEEE Transactions on Human Factors in Electronics, Vol. HFE-8, No. 3.

Journal: Hess, R. (1979). A rationale for human operator pulsive control behavior . Journal of Guidance and Control.

Journal: Kleinman, D, Baron, S, Levison, W. (1970). An optimal control model of human response, Part I: Theory and validation. Automatica.

Journal: Lone. M and Cooke, (2014). A. Review of pilot models used in aircraft flight dynamics. Aerospace Science and Technology.

Journal: Wierenga, R. (1969). An evaluation of a pilot model based on Kalman filtering and optimal control , IEEE Transactions on Man Machine Systems.

Journal: Tustin, A. (1947). The Nature of the Operator's Response in Manual Control, and its Implications for Controller Design. Convention on Automatic Regulators and Servo Mechanisms.

Software: Mathworks. (2015). MATLAB 2015b, The MathWorks Inc., Natick, MA.

Thesis: Alm, T. (2007). Simulator-based design: Methodology and vehicle display applications. Doctoral dissertation (No. 1078). Linköping, Sweden: Linköping University.

Thesis: Turetta, F.M.S. (2014). Analysis of Pilot´s reaction time with alerting systems characteristics and Workload. Mater´s dissertation. São José dos Campos, Brazil: Aeronautics Institute of Technology.

Web page: http://www.ecfr.gov/. Federal Aviation Regulations (FAR): Sec. 25.149, 2002. Accessed in 8/06/2016.