2013-Stell-ATM2013-TN7610

In this paper, multiple regression analysis is used to model the top of descent (TOD) location of user-preferred descent trajectories computed by the flight management system (FMS) on over 1000 commercial flights into Melbourne, Australia. The in- dependent variables cruise altitude, final altitude, cruise Mach, descent speed, wind, and engine type were also recorded or com- puted post-operations. Both first-order and second-order models are considered, where cross-validation, hypothesis testing, and ad- ditional analysis are used to compare models. This identifies the models that should give the smallest errors if used to predict TOD location for new data in the future. A model that is linear in TOD altitude, final altitude, descent speed, and wind gives an estimated standard deviation of 3.9 nmi for TOD location given the trajec- tory parameters, which means about 80% of predictions would have error less than 5 nmi in absolute value. This accuracy is better than demonstrated by other ground automation predictions using kinetic models. Furthermore, this approach would enable online learning of the model. Additional data or further knowl- edge of algorithms is necessary to conclude definitively that no second-order terms are appropriate. Possible applications of the linear model are described, including enabling arriving aircraft to fly optimized descents computed by the FMS even in congested airspace. In particular, a model for TOD location that is linear in the independent variables would enable decision support tool human-machine interfaces for which a kinetic approach would be computationally too slow.

Tenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2013)

Regression Analysis of Top of Descent Location for Idle-thrust Descents

Laurel Stell
Aviation Systems Division
NASA Ames Research Center
Moffett Field, CA, USA
laurel.stell@nasa.gov

Jesper Bronsvoort & Greg McDonald
ATM Automation – ATC Future Systems
Airservices Australia
Melbourne, Australia
jesper.bronsvoort|greg.mcdonald@airservicesaustralia.com

***Abstract*—In this paper, multiple regression analysis is used to model the top of descent (TOD) location of user-preferred descent trajectories computed by the flight management system (FMS) on over 1000 commercial flights into Melbourne, Australia. In addition to recording TOD, the cruise altitude, final altitude, cruise Mach, descent speed, wind, and engine type were also identified for use as the independent variables in the regression analysis. Both first-order and second-order models are considered, where cross-validation, hypothesis testing, and additional analysis are used to compare models. This identifies the models that should give the smallest errors if used to predict TOD location for new data in the future. A model that is linear in TOD altitude, final altitude, descent speed, and wind gives an estimated standard deviation of 3.9 nmi for TOD location given the trajectory parameters, which means about 80% of predictions would have error less than 5 nmi in absolute value. This accuracy is better than demonstrated by other ground automation predictions using kinetic models. Furthermore, this approach would enable online learning of the model. Additional data or further knowledge of algorithms is necessary to conclude definitively that no second-order terms are appropriate. Possible applications of the linear model are described, including enabling arriving aircraft to fly optimized descents computed by the FMS even in congested airspace.**

Keywords – idle-thrust descents; trajectory prediction; top of descent; flight management system

I. Introduction

In congested airspace today, controllers manually direct aircraft to descend in steps in order to merge them into arrival streams. Allowing aircraft to descend smoothly at idle thrust instead would decrease fuel consumption, emissions, and noise to surrounding communities. Such optimized descents are computed by the flight management system (FMS), but increasing the use of these user-preferred trajectories will require a move from tactical clearances to trajectory management. This in turn will require more accurate trajectory prediction, especially around congested airports [1, 2].

To fill this need, many research groups have developed decision support tools, such as the Efficient Descent Advisor (EDA), using kinetic models. Although these methods have resulted in reasonable prediction of time of arrival at specific trajectory points, they have not achieved sufficient accuracy in predicting top of descent (TOD) [4]. A major cause of the error is differences between the aircraft performance models used by the FMS and by the decision support tool.

The purpose of this paper is to use regression analysis to determine the feasibility of modeling TOD location as a first or second order polynomial function of recorded aircraft and trajectory parameters. Over 1000 FMS descents into Melbourne, Australia were used in the analysis. The importance of the various parameters is analyzed using K-fold cross-validation. Since the data analyzed only include flights from one airline arriving at one airport, the coefficients estimated from them cannot be considered definitive, so development of a model suitable for an operational decision support tool will be the subject of future research.

The problem background and related research are described in Sec. II. A statement of the problem and description of the data analyzed are in Sec. III and Sec. IV, respectively. Sec. V presents the regression analysis, concluding with a discussion of the results from a high-level perspective. To indicate how the models presented in this paper can enable more aircraft to fly descents as computed by FMS, Sec. VI describes possible applications. A summary of conclusions is in Sec. VII.

II. Background

Traditional Air Traffic Control (ATC) activities involve the separation and sequencing of aircraft by the controller monitoring the progress of each aircraft and mentally projecting ahead where the aircraft will be in the future. In the descent phase, this projecting ahead leads to a difficult problem for ATC to resolve. Consequently, controllers often direct arriving aircraft to descend in steps so that they can be merged while in level flight. This is particularly true in the US with many climbing, crossing and overflying traffic streams adding complexity to the situation, but this problem is not unique to any location. Australia has designed its airspace for crossing tracks to occur in the cruise phase in preference to the climb and decent phases. When combined with procedural altitude restrictions, arriving aircraft can plan to fly FMS-optimized descents, but this does not remove the need for ATC to have an accurate view of the trajectory to be flown. Improved knowledge of the TOD location by controllers and ground automation would increase the percentage of descents on FMS-computed trajectories.

Many different research groups have developed decision support tools along these lines, including EDA developed by NASA [3]. For a sample of roughly 200 operational descents in four different types of aircraft, over 90% of the metering time predictions from EDA have absolute error less than 30 sec [4]. For the TOD location prediction error, on the other hand, fewer than half the predictions have absolute error less than 5 nmi, which is suspected to be inadequate. References [3, 5–8] also investigated the prediction of the vertical profile using operational data, but none of their sample sizes were large enough to draw conclusions about future prediction error. The ADAPT2 project [9] analyzed prediction of TOD location for 51 commercial flights in B737-600 and B737-800 aircraft. Their results confirmed the difficulty of predicting TOD location within 5 nmi, but they did not indicate a remedy or provide insight into the causes of the large errors. User Request Evaluation Tool (URET) developed by MITRE uses a kinematic model, which was improved by analyzing empirical data [10]; but this paper will show that the use of nominal descent rate and speed for each aircraft type could strongly affect the accuracy of the predicted TOD location if the aircraft follows an FMS-computed descent.

While lack of intent information can result in large TOD prediction errors [2,11], the predictions analyzed in [4] used accurate intent information. Aircraft weights were also available for about 140 of the descents, but using them had a small effect on the TOD prediction error. The large errors were primarily due to differences in aircraft performance models between the FMS and EDA, including Base of Aircraft Data (BADA) family 3 described in [12]. Developing a ground predictor that can accurately predict the FMS-computed descent trajectory is an open problem, and the primary obstacle is that the aircraft performance models used by the FMS are proprietary.

III. Problem Statement

A typical optimized descent is visualized in Fig. 1 and would be performed as follows. At TOD, where the altitude is h_{crz}, the throttle is set to idle and descent is initiated at the cruise Mach number. At crossover altitude when the target descent CAS is reached, the descent is continued at that CAS until the first constraint. Generally, there is some speed constraint within the terminal airspace or below a certain flight level. The International Civil Aviation Organisation (ICAO) specifies the generic terminal airspace speed constraint of 250 KCAS below 10,000 ft. Deceleration is achieved by a shallow segment at idle thrust. This paper only considers the descent from TOD down to altitude h_{f} at the start of this deceleration segment. If this deceleration segment does not exist, the end of the trajectory analyzed here is the first trajectory change point below 10,000 ft. The reason for focusing on the trajectory above h_{f} is that it contains the most uncertainty for controllers, since it is free of ATC procedural constraints and thus can be optimized by the FMS.

In trying to improve the predictions of TOD location from an ATC perspective, it seems beneficial to analyze how the TOD location depends upon the trajectory parameters such as speed profile, cruise and final altitudes, winds, and aircraft mass. The TOD location is determined by integrating the equations of motion along the intended horizontal trajectory. On one hand, this makes it difficult to intuit the relationship between TOD location and the parameters. On the other hand, due to the integration, the TOD location is a smooth function of the trajectory parameters. Therefore, it can be approximated locally by a polynomial, which is much easier to grasp intuitively.

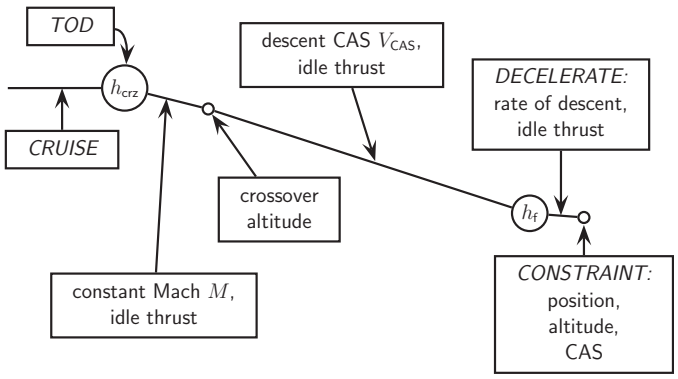

Figure 1. Typical optimized descent trajectory.

The effect of the trajectory parameters on predictions made by EDA was analyzed in detail in [4], but such approximations must be confirmed with operational data. The paper did also report good accuracy for operational data with an approximation linear in descent CAS, aircraft mass, and change in altitude $(h_{\text{crz}} - h_{\text{f}})$. The sample only included about 70 descents for each of two aircraft types — Boeing 757 and Airbus 320 — collected over a two-week period. The TOD locations analyzed were extracted from radar data.

This paper presents the results of regression analysis of the TOD locations computed by the FMS for over 1000 operational descents over 2.5 years in Boeing 737 aircraft. This sample size is acceptable for the number of unknowns in the regression model. This paper also explains details of the regression analysis omitted in [4].

IV. Data Source Description

The trajectory data used in this study were obtained through Automatic Dependent Surveillance Contract (ADS-C). ADS-C is a dependent form of surveillance in which a ground station initiates a contract with an aircraft such that this aircraft will automatically report information obtained from its on-board equipment according to conditions specified in the contract. A little-known feature of ADS-C is the ability to downlink part of the reference trajectory held by the FMS, which is referred to as Intermediate Projected Intent (IPI). IPI consists of up to ten trajectory change points ahead of the aircraft along its intended trajectory. A trajectory change point can be an altitude change, a lateral change, a speed change, or a combination of these.

Between February 2009 and September 2011, Airservices Australia collected data from Boeing 737-800 aircraft equipped with ADS-C, with ADS-C reports every 2 min. The sample includes aircraft with the CFM56-7B24 engine as well as the CFM56-7B26. The flights were inbound to Melbourne, Australia (YMML), whose terminal airspace consists of runway-linked Standard Terminal Arrival Routes (STARs) that are loaded by the crew into the FMS about 45 min prior to arrival. With the lateral path to the runway fully specified, the FMS is then able to calculate an idle-thrust descent as in Fig. 1 and determine the optimal descent point. With the arrival procedure

loaded, the IPI of the ADS-C report reflects this optimized descent and forms the main data for this study.

The IPI is used to determine the TOD altitude h_{crz}, final altitude h_{f}, and horizontal length S_{TOD} of the trajectory between them. The target descent CAS of the aircraft in each sample was estimated by converting Mach numbers from ADS-C state data received below crossover altitude and rounding the average to the nearest 5 kt. Aircraft mass was not available for this study.

The distance S_{TOD} depends upon the wind forecast available to the FMS. The forecast used by the FMS was not available for this study, so it was approximated by forecast winds extracted from the World Area Forecast System in a way similar to airline flight planning applications using the same forecast product [13].

The effect of wind vector $\mathbf{W}$ on the length of the descent as determined by the equations of motion will now be derived. The direction of the intended aircraft velocity $\mathbf{V}$ with respect to the ground is given by the bearing of the IPI segments, so the forecast tailwind w_{tl} and crosswind w_{cr} as a function of altitude can be computed from $\mathbf{V}$ and $\mathbf{W}$. To track a lateral path, the airspeed vector $\hat{\mathbf{V}}$ needs to be headed into the crosswind so that $\mathbf{V} = \hat{\mathbf{V}} + \mathbf{W}$. The true airspeed (TAS), which is denoted V_{TAS}, is the magnitude of $\hat{\mathbf{V}}$. Hence, the effective tailwind $w_{\text{tl\_eff}}$ is

$$w_{\text{tl\_eff}} = w_{\text{tl}} - V_{\text{TAS}} \left[1 - \cos\left(\arcsin\left(\frac{w_{\text{cr}}}{V_{\text{TAS}}} \right) \right) \right], \quad (1)$$

where the trigonometric expression gives the projection of $\hat{\mathbf{V}}$ onto $\mathbf{V}$. If $[t_1, t_2]$ is the time interval over which the descent occurs, the contribution of wind to S_{TOD} is thus $W \equiv \int_{t_1}^{t_2} w_{\text{tl\_eff}} dt$.

Since using the IPI gives V_{TAS}, w_{tl}, and w_{cr} as functions of altitude h, the variable of integration must be changed to h. If γ_{TAS} is the aerodynamic path angle, then

$$\frac{dh}{dt} = V_{\text{TAS}} \sin\gamma_{\text{TAS}}. \quad (2)$$

It is assumed that γ_{TAS} is constant in each IPI segment. As IPI does not include tropopause transition and crossover altitudes, this is not necessarily true for the initial descent stage since h_{crz} may be above the tropopause and the next IPI point may be below crossover. However, with the limited information available, this assumption needs to be made. Then

$$W = \frac{\int_{h_{\text{crz}}}^{h_{\text{f}}} \left[\frac{w_{\text{tl}}}{V_{\text{TAS}}} + \cos\left(\arcsin\left(\frac{w_{\text{cr}}}{V_{\text{TAS}}} \right) \right) - 1 \right] dh}{\sin\gamma_{\text{TAS}}}. \quad (3)$$

The target speeds for the descent are expressed in Mach or CAS rather than TAS. Therefore, the airspeed as a function of altitude in (3) needs to be determined based on the Mach target above crossover altitude, and based on the CAS target below crossover altitude. For these conversions to TAS, International Standard Atmosphere (ISA) conditions were assumed. The term $(\sin\gamma_{\text{TAS}})$ in (3) can be approximated by the averaged value over the IPI segment obtained from integration of (2).

The rate of change of wind will affect S_{TOD} as well if it is taken into account in the equations of motion solved by the FMS trajectory predictor. In particular, it will affect $(\sin\gamma_{\text{TAS}})$ in (3). Assuming that the component of $\mathbf{W}$ parallel to $\hat{\mathbf{V}}$ can be approximated by w_{tl}, the equation of motion for a point-mass system in the direction $\hat{\mathbf{V}}$ becomes

$$\frac{dV_{\text{TAS}}}{dt} = \frac{T - D}{m} - g\sin\gamma_{\text{TAS}} - \frac{dw_{\text{tl}}}{dt}. \quad (4)$$

Again using (2) and rearranging yields

$$\left(\frac{1}{2}\frac{dV_{\text{TAS}}^2}{dh} + g + V_{\text{TAS}}\frac{dw_{\text{tl}}}{dh} \right) \sin\gamma_{\text{TAS}} = \frac{T - D}{m}. \quad (5)$$

Define the modified path angle γ^*_{TAS} by

$$\sin\gamma^*_{\text{TAS}} = \frac{\frac{1}{2}\frac{dV_{\text{TAS}}^2}{dh} + g + V_{\text{TAS}}\frac{dw_{\text{tl}}}{dh}}{\frac{dV_{\text{TAS}}^2}{dh} + g} \sin\gamma_{\text{TAS}}. \quad (6)$$

For an IPI segment that starts at altitude h_1 and ends at altitude h_2, an average value is approximated by

$$\overline{\sin\gamma^*_{\text{TAS}}} = \frac{\sin\gamma_{\text{TAS}}}{h_1 - h_2} \int_{h_2}^{h_1} \left[1 + \frac{\frac{dV_{\text{TAS}}^2}{dh}}{\frac{dV_{\text{TAS}}^2}{dh} + 2g} \right] dh. \quad (7)$$

To account for the rate of change of wind, this is used in place of $\sin\gamma_{\text{TAS}}$ in (3).

A few possible simplifications come to mind. First, $\widehat{W_1}$ might be defined as in (3) without using (7). Second, the effect of the crosswind might also be ignored, which would mean replacing (1) by $w_{\text{tl\_eff}} = w_{\text{tl}}$ to compute $\widehat{W_2}$. Third, the assumption of constant γ_{TAS} on each IPI segment might be extended to constant dh/dt, which could then be directly computed from the details of the IPI segment to obtain $\widehat{W_3}$. The correlation coefficient between W and any $\widehat{W_j}$ is greater than 0.99, so using $\widehat{W_j}$ in the regression analysis in the next section has no discernible effect.

V. Results

Regression analysis was used to model S_{TOD} as a function of the independent variables cruise altitude h_{crz}, final altitude h_{f}, cruise Mach M, descent CAS V_{CAS}, wind contribution W, and engine type. These variables were described in the previous section. The regression analysis used data from 1088 descents. Many of the descents have cruise altitude in the range 380FL–390FL, final altitude around 10,400 ft, and descent speed around 280 KCAS. Overall, though, the values for these variables seem sufficiently spread out for regression analysis. Finally, 11% of the descents were in aircraft with CFM56-7B26 engines, while the others had CFM56-7B24 engines.

The analysis considers models such as

$$S_{\text{TOD}} = \beta_0 + \beta_1 h_{\text{crz}} + \beta_2 h_{\text{f}} + \beta_3 M + \beta_4 V_{\text{CAS}} + \beta_5 W. \quad (8)$$

Of course, the TOD location also depends upon the aircraft performance model, but that will be reflected in the β_i values. Equation (8) assumes S_{TOD} is linear in the independent variables, but regression analysis also allows models that are higher-order polynomials.

Let $\mathbf{y}$ be a column vector containing the FMS-computed values of S_{TOD}, and let X be a matrix containing the values of the independent variables used in the model, with each row containing the values for one descent. In (8), the model includes a constant term β_0, which is handled by putting into X a column of ones. Ideally, the vector of coefficients $\boldsymbol{\beta}$ would satisfy $\mathbf{y} = X\boldsymbol{\beta}$; but there are many more equations than unknowns in this system, and it has no solution. Instead, the multiple regression model is $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ has a Gaussian distribution with mean zero and variance σ^2. The least squares estimate of $\boldsymbol{\beta}$ is the vector $\hat{\boldsymbol{\beta}}$ that minimizes the residual sum of squares

$$RSS = \left\| \mathbf{y} - X\hat{\boldsymbol{\beta}} \right\|^2 = \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} X_{ij}\hat{\beta}_j \right)^2, \qquad (9)$$

where n is the number of observations and p is the number of parameters in the model. The residual is $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - X\hat{\boldsymbol{\beta}}$. If variables, particularly aircraft weight, that affect S_{TOD} are omitted from the data recorded, it will be reflected in $\hat{\boldsymbol{\beta}}$ or $\mathbf{r}$ or both.

A. Variable Selection

Computing $\hat{\boldsymbol{\beta}}$ is straightforward; the results presented below were obtained with the function `lm()` in the R language for statistical computing. The interesting part of regression analysis is choosing the model and checking the underlying assumptions. While (8) is an obvious choice for a model that is linear in the regressors, there are many possible variations as discussed below. Allowing quadratic and cross terms greatly increases the number of possible terms, and it is unlikely that all of them are appropriate. Including extraneous terms leads to overfitting, which increases error if the model is used for predictions in the future. Variable selection is the process of choosing which terms to include. The obvious ways to compare models are by their *RSS* or R^2 values, but that does not indicate whether overfitting is occurring. Hypothesis testing is also commonly used for variable selection, but it has the drawback that a significance level of 0.05, say, results in incorrectly accepting insignificant terms 5% of the time and also incorrectly rejecting significant terms, even in the best of circumstances. Consequently, this analysis uses hypothesis testing in conjunction with cross-validation and additional diagnostics.

In K-fold cross-validation, the samples are randomly divided into K sets $\mathcal{S}_i$ of roughly equal size. For $i = 1, \ldots, K$, let $\hat{\boldsymbol{\beta}}^{(i)}$ be the estimate of the coefficients when fitting a given model to the samples that are *not* in $\mathcal{S}_i$, and let its mean square error be

$$MSE^{(i)} = \frac{1}{n_i} \left\| \mathbf{y}^{(i)} - X^{(i)}\hat{\boldsymbol{\beta}}^{(i)} \right\|^2, \qquad (10)$$

where $\mathbf{y}^{(i)}$ and $X^{(i)}$ contain only the n_i observations that are in $\mathcal{S}_i$. The cross-validation estimate $\widehat{MSE}$ of the mean square error is then the average of $MSE^{(i)}$ and the standard error of $\widehat{MSE}$ is $\sqrt{\text{var}\left(MSE^{(i)}\right)/K}$. Using different samples to compute $MSE^{(i)}$ than were used to compute $\hat{\boldsymbol{\beta}}^{(i)}$ checks for overfitting. For variable selection, a "one-standard error rule" is popular. If model j has cross-validation error $\widehat{MSE}_j$ with standard error se_j and subscript 0 indicates the model with minimum $\widehat{MSE}_j$, then use the simplest model with $\widehat{MSE}_j < \widehat{MSE}_0 + \text{se}_0$. Additional details on cross-validation for variable selection are in [14].

Besides considering second-order terms, a few additional variations on the basic form of (8) are considered. First, because engine type is a categorical variable with two possible values in the data analyzed here, it is included in a model by allowing coefficients to depend upon it, similar to ANOVA. Second, by using (7), W includes the rate of change of wind; so the coefficient of W is theoretically one as explained in [15], provided the wind gradient is included in the equations of motion. Even if the wind gradient is omitted from the equations of motion, the high correlation between W and $\widehat{W}_1$ noted at the end of Sec. IV suggests the coefficient should still be one. Hence, in some models this is assumed. Finally, it is also plausible that the coefficient of h_{f} should be the negative of the coefficient of h_{crz}, which can be assumed by replacing the two separate terms with the single term $\Delta h = h_{\text{crz}} - h_{\text{f}}$.

1) First-order models: Table I lists the models considered that are linear in the independent variables, roughly in decreasing order of model complexity. If W is not listed for a given model, then its coefficient is assumed to be one. Fig. 2 shows the comparison of the linear models using 10-fold cross-validation, with each plot being for a different random seed. For model j, each plot shows $\widehat{MSE}_j$ with error bars denoting $\pm\,\text{se}_j$, and the horizontal dashed line shows $\widehat{MSE}_0 + \text{se}_0$. For reasons explained below, the lengths of the error bars in the two upper plots are noticeably different from the two lower plots. While the choice of random seed changes $\widehat{MSE}_j$ slightly, it has very little effect on the differences in their sizes between models. In short, the differences in $\widehat{MSE}$ between the different models are within the noise of the data and the models, as indicated by the error bars.

Closer inspection of the results reveals that, regardless of random seed, model L2 gives the smallest $\widehat{MSE}_j$, but the first four models and model L7 have roughly the same $\widehat{MSE}_j$. Comparing model L1 with model L2 strongly suggests that S_{TOD} does not depend upon engine type. The results also indicate that M can be dropped and confirm that the coefficient of W can be assumed to be one. While models L5, L6, L8, and L9 that use Δh are worse than those that use h_{crz} and h_{f} separately, they still all have $\widehat{MSE}_j < \widehat{MSE}_2 + \text{se}_2$ or very close to it. This means that the one-standard error rule would select model L9 for any of these seeds. A model that completely ignores h_{crz}, h_{f}, V_{CAS}, or W will have $\widehat{MSE}$ greater than 30 nmi^2, so no further simplification of the model is possible.

The following results of hypothesis testing of the coefficients augment the cross-validation:

- All the terms in model L7 and model L9 have p-values less than 10^{-15}.
- Comparing model L7 to model L9 with the R function `anova()` gives a p-value of 3×10^{-6} for the F-test, which is strong support for separate terms for the two altitudes.

TABLE I. REGRESSION MODELS CONSIDERED.

First-order Models

	terms included
L1	$h_{\text{crz}}, h_{\text{f}}, V_{\text{CAS}}, M, W$ with different coefficients for the two different engine types
L2	$h_{\text{crz}}, h_{\text{f}}, V_{\text{CAS}}, M, W$
L3	$h_{\text{crz}}, h_{\text{f}}, V_{\text{CAS}}, M$
L4	$h_{\text{crz}}, h_{\text{f}}, V_{\text{CAS}}, W$
L5	$\Delta h, V_{\text{CAS}}, M, W$
L6	$\Delta h, V_{\text{CAS}}, M$
L7	$h_{\text{crz}}, h_{\text{f}}, V_{\text{CAS}}$
L8	$\Delta h, V_{\text{CAS}}, W$
L9	$\Delta h, V_{\text{CAS}}$

Second-order Models

	terms included
S1	$h_{\text{crz}}, h_{\text{f}}, V_{\text{CAS}}, M, W$, all interaction terms, all quadratic terms with different coefficients for the two different engine types
S2	$h_{\text{crz}}, h_{\text{f}}, V_{\text{CAS}}, M, W$, all interaction terms, all quadratic terms
S3	same as model S2 but without W^2
S4	$h_{\text{crz}}, h_{\text{f}}, V_{\text{CAS}}, M, W$, all interaction terms, W^2
S5	$h_{\text{crz}}, h_{\text{f}}, V_{\text{CAS}}, M, W$, all quadratic terms
S6	$h_{\text{crz}}, h_{\text{f}}, V_{\text{CAS}}, M, h_{\text{f}}V_{\text{CAS}}, W^2$
S7	$h_{\text{crz}}, h_{\text{f}}, V_{\text{CAS}}, h_{\text{f}}V_{\text{CAS}}, W^2$
S8	$h_{\text{crz}}, h_{\text{f}}, V_{\text{CAS}}, M, h_{\text{f}}V_{\text{CAS}}$
S9	$h_{\text{crz}}, h_{\text{f}}, V_{\text{CAS}}, M, W^2$
S10	$h_{\text{crz}}, h_{\text{f}}, V_{\text{CAS}}, h_{\text{f}}V_{\text{CAS}}$
S11	$h_{\text{crz}}, h_{\text{f}}, V_{\text{CAS}}, W^2$
S12	$\Delta h, V_{\text{CAS}}, M, h_{\text{f}}V_{\text{CAS}}, W^2$

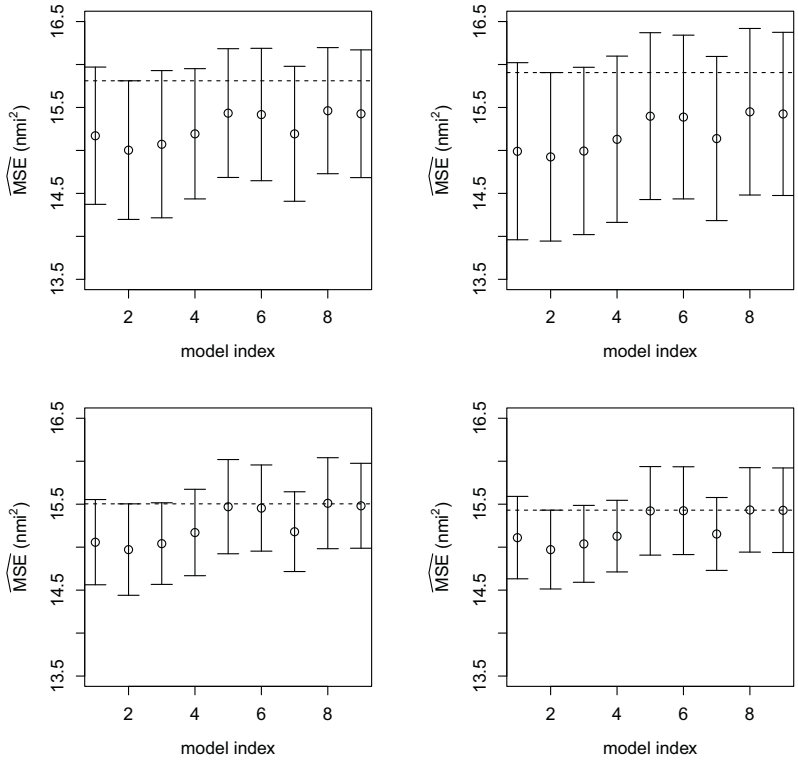

Figure 2. Cross-validation results for first-order models using four different random seeds.

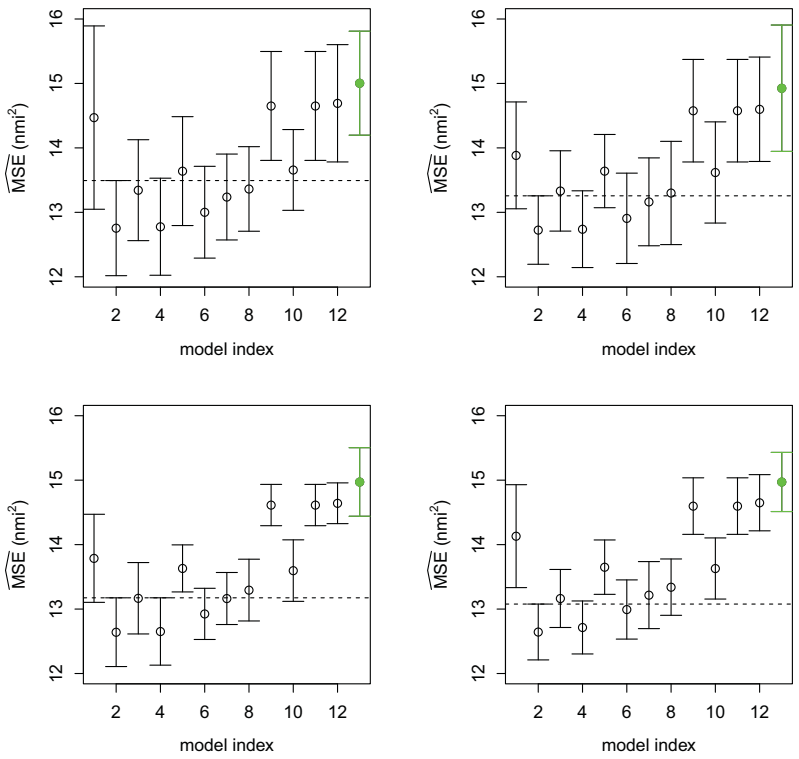

Figure 3. Cross-validation results for second-order models using four different random seeds. Green is for model L2 in Table I.

- The F-test comparing model L4 to model L7 gives a p-value of 0.052, which is only weak evidence for rejecting the hypothesis — based on the equations of motion — that the coefficient of W is one.
- In model L3, the M term has p-value 0.00066, which suggests this model should be chosen over model L7.

Based on the results of both cross-validation and hypothesis testing, the best first-order models are L3, L7, and L9.

2) Second-order models: With over 1000 samples, it is also reasonable to consider models with higher-order terms as listed in Table I. Fig. 3 shows the results of 10-fold cross-validation for the second-order models, but it also includes results for model L2 in green to assist in comparison. Allowing the coefficients to depend upon engine type in model S1 clearly results in overfitting. Comparing model S2 and model S4 strongly suggests that W^2 is the only quadratic term that might be significant. Comparing models S2, S4, and S6 strongly suggests that $h_{\text{f}}V_{\text{CAS}}$ is the only interaction term that might be significant. Finally, comparing model S6 and model S12 shows that using Δh instead of separate h_{crz} and h_{f} terms now gives clearly worse results, although, of course, $\widehat{MSE}$ for model S12 is still less than for model L9 — or even model L2. The one-standard error rule would choose model S6 for one of the random seeds, model S7 for two, and either model S7 or model S8 for the other.

To compare to hypothesis testing, fitting model S6 to all the data and applying t-tests to each of the coefficients gives p-values less than 10^{-5}, which strongly suggests all of them are statistically significant. Understanding these apparently contradictory results requires more detailed analysis of the data.

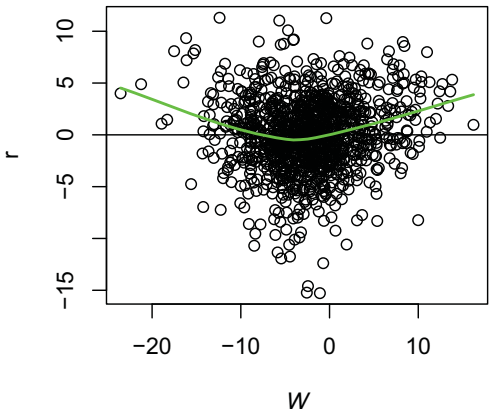

Figure 4. Residuals versus W for model S8.

Models S6 and S8 are the same except that the latter does not include W^2. Fig. 4 shows for each observation the residual $\mathbf{r}$ versus W for model S8 fit to all the data, which helps to explain the difference in $\widehat{MSE}$ between these two models. The green line is an approximation of the mean as a function of W, as computed by the R function `loess.smooth()`. Since the green line has the shape of a quadratic polynomial, including W^2 in the model gives a better fit, but the plot also suggests that the better fit is primarily due to the observations with larger values of $|W|$. Cross-validation using only the 1058 descents with -15 nmi $\leq W \leq 10$ nmi shows that including W^2 in the model is no longer advantageous. Since deleting less than 3% of the data eliminates support for the conclusion that W^2 is significant, it seems best to omit it from the model unless it can be justified physically or procedurally.

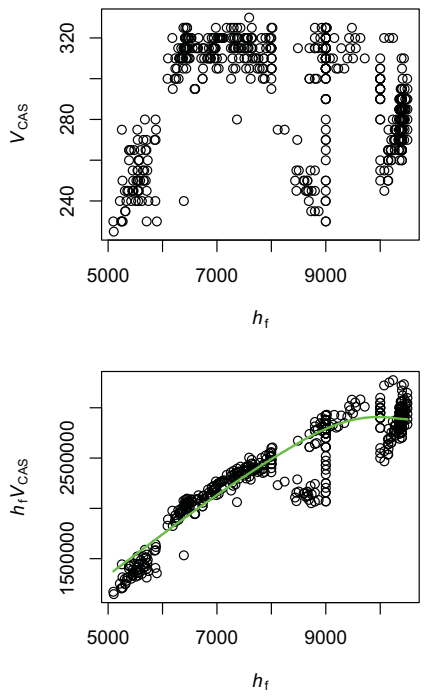

Figure 5. Relationship between observed V_{CAS} and h_{f}.

The next question is whether it is appropriate to include the cross term $h_{\text{f}}V_{\text{CAS}}$ in the model. The upper plot in Fig. 5 shows the relationship between V_{CAS} and h_{f} in the sample descents. The arrangement of the points in clusters can lead to overfitting that might not be detected by cross-validation. Furthermore, the cross term is almost collinear with h_{f}, as indicated by the lower plot in Fig. 5 (where the green line is again obtained by `loess.smooth()`) and by a correlation coefficient of 0.94. This can result in $\hat{\boldsymbol{\beta}}$ having large variance. To check this possibility, Fig. 6 plots, for each of four different models, $(\hat{\beta}_k - \bar{\beta}_k)/\bar{\beta}_k$, using $\hat{\boldsymbol{\beta}}$ from each of the 10 cross-validation fits and each of the four random seeds and $\bar{\boldsymbol{\beta}}$ is the average of the 40 estimates for $\hat{\boldsymbol{\beta}}$. Note that the range of the y axes are much larger for the second-order models than for the first-order models, which means that $\hat{\boldsymbol{\beta}}$ is much more stable for the first-order models. For these reasons, including the cross term in the model does not seem justified.

The preceding discussion suggests that the second-order models carry a high risk of overfitting the data, even though that is not obvious from either cross-validation or hypothesis testing. While one would like cross-validation and hypothesis testing to give clear, consistent results indicating which model to use, that does not occur for reasons that will be discussed further below.

B. Further Assessment of First-order Models

Since the second-order models may be overfitting, the conservative approach is to use one of the first-order models in applications that will predict S_{TOD} for future observations. The next step in the analysis is to perform additional diagnostics and further investigate the estimated prediction error for the most promising models. The remaining analysis discusses the stability of the coefficients, estimates the prediction error, and checks the standard assumptions of multiple regression for models L3, L7, and L9.

1) Stability of coefficients: Recall that the coefficient of M has relatively large variance in the second-order models shown in Fig. 6. This is also true — in fact, worse — for model L3. High leverage points are observations that have a relatively large effect on $\hat{\boldsymbol{\beta}}$; see [16], for example. Most the high leverage points for model L3 have $M < 0.74$. Recreating Fig. 2 using only the 1054 descents with M at least 0.74 shows that including M in the model no longer gives any advantage. Furthermore, its p-value in model L3 is now 0.30. On the other hand, the range of the estimates of the coefficient of M is now even larger, perhaps indicating that the range of M values used in the fit is now too small. In summary, the available data cannot give precise estimates for the coefficient of M, which also affects determining whether or not M is significant and estimating prediction error. All the other coefficients in models L3, L7, and L9 have small variances as in Fig. 6.

2) Goodness of fit: A popular measure of the goodness of fit is the R^2 statistic, which is interpreted as the percentage of variability in the response explained by the predictors. For model L9 fit to all 1088 descents, R^2 is 0.90, which is generally considered a very good fit. As explained in [17], however, the R^2 value can be affected if X fails to have a multivariate normal distribution, which is definitely the case here.

A better indicator of the goodness of fit is the variance σ^2 of the error $\boldsymbol{\epsilon}$. Of course, this cannot be known exactly, but it can be estimated by $\hat{\sigma}^2 = RSS/(n-p)$. For model L9, $\hat{\sigma} = 3.9$ nmi. If the model error ϵ_j has normal distribution with mean zero and standard deviation $\hat{\sigma}$, then $|\epsilon_j| < 5$ nmi for 80% of descents. In fact, $|r_j| < 5$ nmi for 82% of the descents analyzed. The R^2 and $\hat{\sigma}$ values for models L3 and L7 are the same as for model L9 to two decimal places.

The value of $\hat{\sigma}$ also explains the differences between the random seeds in Fig. 2. If cross-validation fold i has m observations, then $\frac{m}{\sigma^2} MSE^{(i)}$ will have a χ^2_m distribution. For 10-fold cross-validation with 1088 descents, $m \approx 109$; so $\sigma = 3.9$ nmi gives a 90%-confidence interval for $MSE^{(i)}$ is $[10.0, 12.1]$ nmi^2. If the folds can be treated as having the same size, then $\frac{n}{\sigma^2}\widehat{MSE}$ will have a χ^2_n distribution. This would imply mean $\widehat{MSE}$ 15.4 nmi^2, standard deviation 0.66 nmi^2, and a 90%-confidence interval for MSE is $[14.3, 16.5]$ nmi^2. The values of $MSE^{(i)}$ and $\widehat{MSE}$ obtained in the cross-validation of the first-order models are consistent with these results.

3) Regression assumptions: To complete this regression analysis, its underlying assumptions should be checked. The first of these is the assumption that the model error is Gaussian. Normal probability plots (or Q-Q plots) of the residuals of models L3, L7, and L9 are very close to each other. They are close to Gaussian but have tails that are a bit too heavy, which is common in real data.

The next diagnostic test is for collinearity. A detailed discussion of diagnosing collinearity, including identification of the collinear terms and the inadequacy of pairwise correlation coefficients, is in [16]. For model L3, the constant term and M are strongly collinear because the values of M do not vary

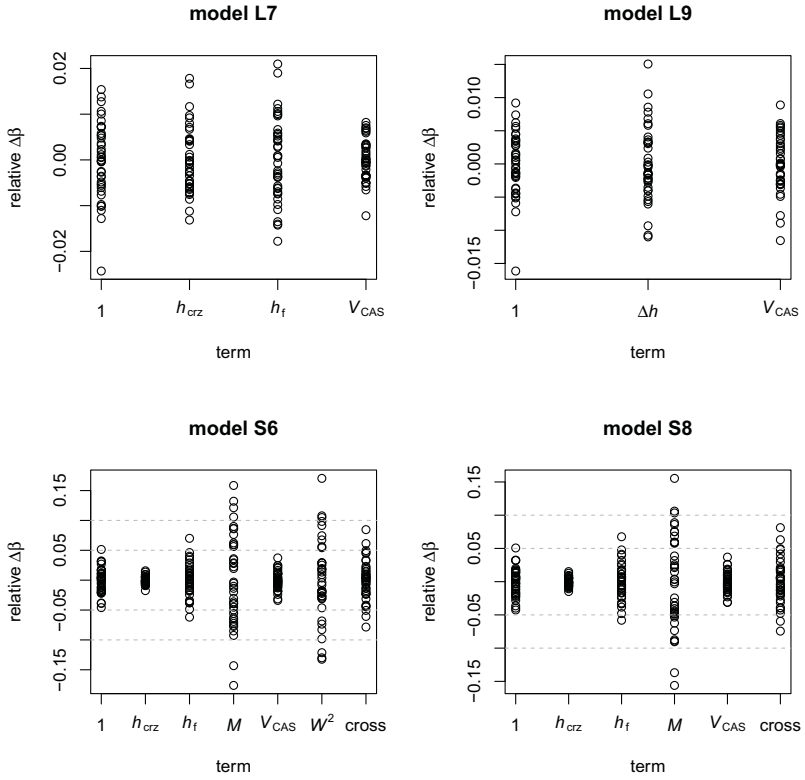

Figure 6. Relative change in $\hat{\boldsymbol{\beta}}$ over cross-validation fits.

much. The primary symptom of collinearity is large variance in estimates of the coefficients, which does indeed occur for the coefficient of M in model L3. Fig. 6 indicates this is not a problem for model L7 or model L9, which both show much weaker symptoms of collinearity.

The ith partial residual vector is

$$\mathbf{r}_i^* = \mathbf{r} + \mathbf{x}_i\hat{\beta}_i = \mathbf{y} - \sum_{j\neq i} \mathbf{x}_j\hat{\beta}_j. \tag{11}$$

The partial residual plot [18] for regressor i shows $\mathbf{r}_i^*$ versus $\mathbf{x}_i$, which indicates the relationship between $\mathbf{x}_i$ and $\mathbf{y}$ *given* the other predictors. Fig. 7 shows the centered partial residual plots for model L3, created by the R function `crPlots` in the `car` package. Each $\mathbf{r}_i^*$ is the residual that would result by omitting $\mathbf{x}_i$ from the model. The red dashed line shows the simple regression fit to the partial residual plot, which has slope $\hat{\beta}_i$ from the original fit. The green solid curve is a local polynomial fit of the points. The envelope around $\mathbf{r}_i^*$ in each plot is roughly independent of $\mathbf{x}_i$, which indicates σ is independent of the values of the independent variables — called homoscedasticity. The partial residual plots also confirm that there is no obvious nonlinearity in the dependence of S_{TOD} on these variables. The slopes and goodness-of-fit of the red lines further confirm that h_{crz}, h_{f}, and V_{CAS} are all important in determining S_{TOD}. For M, however, the red line is very flat, and the green line shows that the points with $M < 0.74$ may have relatively large leverage.

C. Discussion of Results

Two strong conclusions were reached: S_{TOD} does not depend upon engine type, and the coefficient of W can be assumed to equal one. These are supported by cross-validation, hypothesis testing, and the equations of motion — since the two engine types are very similar. Choosing the best first-order model then requires deciding whether M is significant and whether to use Δh instead of two separate altitudes. The cross-validation results indicate the differences between the first-order models are within the noise. Hypothesis testing, however, suggested M is significant and the model should include separate terms for the altitudes. The equations of motion suggest separate altitude terms will be appropriate provided there is "enough" variability in h_{f} since the atmosphere (and hence V_{CAS}) varies with altitude, but it is not clear what is "enough".

While hypothesis testing gives the illusion of clear-cut choices, it would probably be misleading for these data. Even in the best scenarios, a certain percentage of hypothesis tests will give incorrect results. More importantly, the t-test for significance of a regression coefficient is unaware of the distribution of that variable, the distribution of residuals, or of the relationship between residuals and that variable. If only one random seed were used, then strict adherence to the one-standard error rule for cross-validation would also give a clear choice of model. Using multiple random seeds, however, gives a more realistic view. These ideas are illustrated in the next two paragraphs.

First, the cross-validation results indicate the effects of M and W^2 are small relative to the standard errors, even though the t-tests for these coefficients rejected the null hypothesis that

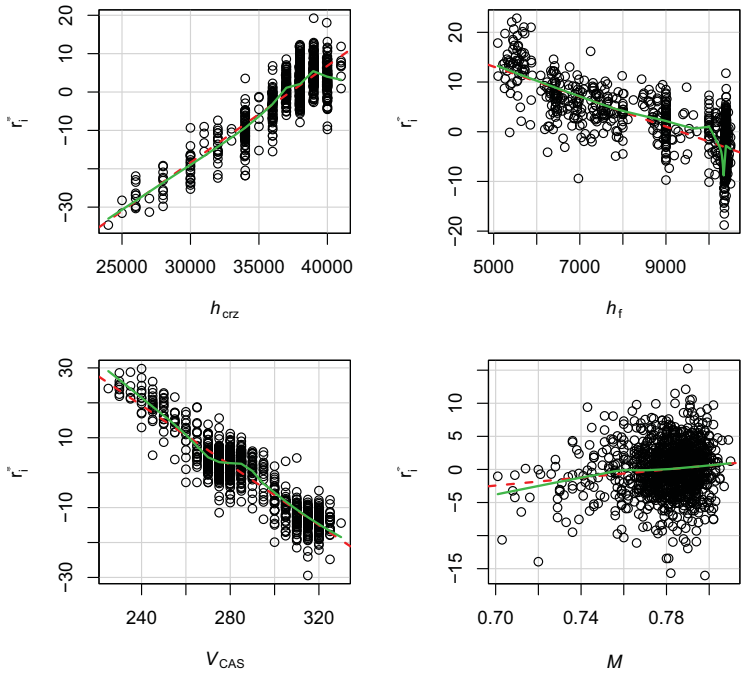

Figure 7. Partial residual plots for model L3.

they were zero. The effect for each of these terms is primarily due to about 3% of observations with values at the edges of the distributions of the independent variable. Since there are so few of these influential observations, the number of them in each of the 10 cross-validation folds is relatively variable between folds and between random seeds, which can be reflected in the variation in standard errors. Determining whether M or W^2 should be included in the model would require collecting enough observations to fill in the distributions where there are currently too few observations. This would probably require on the order of 10,000 observations.

On the other hand, both cross-validation and hypothesis tests indicate that the term $h_\mathrm{f}V_\mathrm{CAS}$ is significant. The primary reason that it might not be appropriate to include this cross term in the model is that the joint distribution of h_f and V_CAS shown in Fig. 5 has a few clusters of points distributed in such a way that the cross term can essentially fit each cluster separately. Nearly all the available observations fit into the clusters in the joint distribution, and each cluster is large enough to be adequately represented in each of the folds. Therefore, cross-validation indicates the cross term should be included, but doing so might produce a model that gives poor predictions for values that are not in these clusters. The explanation for the clusters in Fig. 5 is not completely known. If it were certain that future observations would fit the same pattern, then the cross term could be safely included in the model. Verifying this solely through operational data, however, is difficult. An alternative approach would be to consult subject matter experts in air traffic control procedures and FMS algorithms.

For the best first-order models, $\hat{\sigma} = 3.9$ nmi, which seems large compared to the standard deviation and interquartile range for the observed values of S_TOD, both of which are 12.4 nmi. Even for model S6, $\hat{\sigma}$ is still 3.6 nmi. The most obvious candidate to improve the model is aircraft mass. It would be interesting to be able to repeat the analysis on roughly 1000 observations for which aircraft mass is available in order to determine how much of the remaining variance in S_TOD it explains. It would probably not be of practical utility, though, since aircraft mass is not likely to be available for the potential applications described in the next section.

Despite these issues, the regression analysis has shown that very simple models fit over 80% of the operational S_TOD values within 5 nmi. Models L7 and L9, in particular, seem to satisfy the regression assumptions well, so this should be a good estimate of future prediction accuracy. It is therefore reasonable to use it as the basis for future research into the acceptability of using the models in decision support tools to enable increased use of FMS-optimized descents.

VI. Applications

This section discusses how and why regression models might be used in ATC decision support tools, provided future research finds they are sufficiently accurate. With knowledge of the parameters h_crz, h_f, V_CAS, and W, the ground automation system can estimate the TOD location. If a trajectory with descent length S_TOD and descent speed V_CAS was not acceptable, the controller would like to choose a different descent speed V^*_CAS that would result in an acceptable trajectory. Controllers might wish to see how changing V^*_CAS would affect S^*_TOD, or they might prefer to specify S^*_TOD and have the tool estimate the appropriate V^*_CAS. In either case, the ground automation what-if feature could use the relation $S^*_\mathrm{TOD} \approx S_\mathrm{TOD} - \alpha\,(V_\mathrm{CAS} - V^*_\mathrm{CAS})$, where α is the coefficient of V_CAS estimated by regression of historical data. This is computationally much faster than a kinetic approach, especially when determining V^*_CAS for specified S^*_TOD where a kinetic approach would require iteratively solving the equations of motion with different descent speeds. Hence, a regression model would enable human-machine interfaces such

as a slider bar that might be too slow if based on a kinetic approach.

Another benefit of using regression models is that the application could incorporate aspects of online machine learning to dynamically update the coefficients if appropriate. This might occur if the average (unknown) aircraft mass changes, thereby changing the constant term. The coefficients might also change on a seasonal basis or if an FMS software revision is deployed.

For greatest use of FMS-optimized descents, ATC tools must be able to compute accurately a four-dimensional trajectory, but this capability has not yet been demonstrated on a sufficiently large sample for a variety of aircraft types. The difficulty is that the aircraft performance models used by the FMS are proprietary and unavailable to developers of ground automation. A simple modification to the ground automation's aircraft performance model, such as multiplying $(T - D)$ by a constant [19], might give sufficiently accurate predictions. A trajectory prediction technique developed in Australia uses trajectory information derived from ADS-C data, combined with information on the ground [20]. It has shown that it is possible to tweak the components of the equations of motion in order to obtain a very accurate prediction of the aircraft's trajectory based on information down-linked from the FMS.

VII. Conclusions

Regression analysis of the TOD location for over 1000 commercial descents in B737-800 aircraft to Melbourne, Australia showed that the simple model

$$S_{\text{TOD}} = \beta_0 + \beta_1 h_{\text{crz}} + \beta_2 h_{\text{f}} + \beta_3 V_{\text{CAS}} + W + \epsilon \qquad (12)$$

gives a standard deviation of ϵ of about 3.9 nmi. Over 80% of the residuals are less than 5 nmi in absolute value. Adding a term for M and the second-order term $h_{\text{f}} V_{\text{CAS}}$ gives a standard deviation of ϵ of about 3.6 nmi, but there are indications this model might overfit the data. The first-order models have no sign of overfitting and the estimates of the coefficients are stable. Furthermore, all diagnostics give good results, indicating the assumptions of the regression analysis are reasonable. The model is more accurate than has been demonstrated for any other ground automation predictions of TOD location. This is most likely because the regression coefficients, being obtained from fitting FMS-computed TOD locations, reflect the aircraft performance models used by the FMS; whereas other researchers have not been able to match the FMS performance models. Future research should determine whether incorporating a regression model into a decision support tool to help controllers choose suitable advisories would enable increased use of fuel-efficient descent trajectories computed by the FMS.

References

[1] J. Klooster, S. Torres, M. Castill-Effen, R. Subbu, L. Kanmer, D. Chan, and T. Tomlinson, "Trajectory synchronization and negotiation in trajectory based operations," in *29th Digital Avionics Systems Conference (IEEE DASC 10), Salt Lake City, UT*, October 2010.

[2] J. Bronsvoort, G. McDonald, M. Paglione, C. Garcia-Avello, I. Bayraktutar, and C. M. Young, "Impact of missing longitudinal aircraft intent on descent trajectory prediction," in *30th Digital Avionics Systems Conference (IEEE DASC 11), Seattle, WA*, October 2011.

[3] R. A. Coppenbarger, R. W. Mead, and D. N. Sweet, "Field evaluation of the tailored arrivals concept for datalink-enabled continuous descent approach," *Journal of Aircraft*, vol. 46, no. 4, pp. 1200–1209, 2009.

[4] L. L. Stell, "Prediction of top of descent location for idle-thrust descents," in *9th USA/Europe Air Traffic Management R&D Seminar, Berlin, Germany*, June 2011.

[5] S. M. Green and R. A. Vivona, "Field evaluation of Descent Advisor trajectory prediction accuracy (AIAA 1996-3764)," in *AIAA Guidance, Navigation and Control Conference, San Diego, CA*, July 1996.

[6] D. H. Williams and S. M. Green, "Flight evaluation of Center-TRACON Automation System trajectory prediction process," Tech. Rep. TP-1998-208439, NASA, July 1998.

[7] S. M. Green, M. P. Grace, and D. H. Williams, "Flight test results: CTAS and FMS cruise/descent trajectory prediction accuracy," in *3rd USA/Europe Air Traffic Management R&D Seminar, Napoli, Italy*, June 2000.

[8] T. Courdacher and V. Mouillet, "ADAPT final report," Tech. Rep. TRS T06/22316TC ADAPT-D6-V1.0, EUROCONTROL, August 2008.

[9] R. Christien and A. Marayat, "ADAPT2 data analysis report," tech. rep., EUROCONTROL, December 2009.

[10] W. C. Arthur and M. P. McLaughlin, "User Request Evaluation Tool (URET) interfacility conflict probe performance assessment," in *2nd USA/Europe Air Traffic Management R&D Seminar, Orlando, FL*, December 1998.

[11] M. Paglione, I. Bayraktutar, G. McDonald, and J. Bronsvoort, "Lateral intent error's impact on aircraft prediction," in *8th USA/Europe Air Traffic Management R&D Seminar, Napa, CA*, June 2009.

[12] A. Nuic, "User manual for the base of aircraft data (BADA) revision 3.8," Tech. Rep. 2010-003, EUROCONTROL Experimental Centre, April 2010.

[13] I. C. A. Organization, "Annex 3; meteorological service for international air navigation," tech. rep., ICAO, July 2010.

[14] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. John Wiley & Sons, Inc., 1998.

[15] L. L. Stell, "Analysis of Flight Management System predictions of idle-thrust descents," in *29th Digital Avionics Systems Conference (IEEE DASC 10), Salt Lake City, UT*, October 2010.

[16] D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley & Sons, Inc., 1980.

[17] S. Weisberg, *Applied Linear Regression*. John Wiley & Sons, Inc., 2005.

[18] W. A. Larsen and S. J. McCleary, "The use of partial residual plots in regression analysis," *Technometrics*, vol. 14, pp. 781–790, August 1972.

[19] M. Xue and H. Erzberger, "Improvement of trajectory synthesizer for efficient descent advisor (AIAA 2011-7020)," in *AIAA 11th Aviation, Technology, Integration, and Operations Conference, Virginia Beach, VA*, September 2011.

[20] J. Bronsvoort, G. McDonald, J. Lopez-Leones, and H. Visser, "Improved trajectory prediction for air traffic management by simulation of guidance logic and inferred aircraft intent using existing data-link technology (AIAA 2012-4928)," in *AIAA Guidance, Navigation and Control Conference, Minneapolis, MN*, August 2012.

Author Biographies

Laurel Stell is an Aerospace Engineer at the NASA Ames Research Center, Moffett Field, California, researching aircraft trajectory prediction. She earned her Ph.D. in applied mathematics from Cornell University, Ithaca, New York, in 1993.

Jesper Bronsvoort is an Aerospace Engineer at Airservices Australia, Melbourne. He works on several initiatives involved in the transition to Trajectory Based Operations in Australia. He holds an MSc degree (2011) in aerospace engineering from Delft University of Technology, The Netherlands.

Greg McDonald is an Air Traffic Controller with more than 30 years experience. Since 1998, he has been involved in the Australian ATM Strategic Plan and implementing efficiencies for airlines including AUSOTS flex tracks. His work managing the Tailored Arrivals program for Australia has lead to his interest in improving ground based trajectory prediction.