

Variational System Identification of Aircraft

Dimas Abreu Archanjo Dutra*

West Virginia University, Morgantown, West Virginia, 26506-6106

Variational system identification is a new formulation of maximum likelihood for estimation of parameters of dynamical systems subject to process and measurement noise, such as aircraft flying in turbulence. This formulation is an alternative to the filter-error method that circumvents the solution of a Riccati equation and does not have problems with unstable predictors. In this paper, variational system identification is demonstrated for estimating aircraft parameters from real flight-test data. The results show that, in real applications of practical interest, it has better convergence properties than the filter-error method, reaching the optimum even when null initial guesses are used for all parameters and decision variables. This paper also presents the theory behind the method and practical recommendations for its use.

Nomenclature

Mathematical Symbols

$\mathbb{R}$	=	Set of real numbers
$\delta(t)$	=	Dirac's delta
δ_{ij}	=	Kronecker's delta
I	=	Identity matrix
d	=	Differential operator
$E[\cdot]$	=	Expectation operator
$\mathcal{N}(x; \mu, \Sigma)$	=	Multivariate normal density at x
$h_q[X]$	=	Differential entropy of X under density q
t	=	Time
T	=	Sampling period
k	=	Sample index
n_x	=	Number of states
n_u	=	Number of external inputs
n_y	=	Number of measured outputs
n_θ	=	Number of unknown parameters
$N + 1$	=	Number of time samples
$x(t)$	=	State vector
$u(t)$	=	External input vector
y_k	=	Output vector
v_k	=	Measurement noise
$W(t)$	=	n_x -dimensional Wiener process
$w(t)$	=	weak derivative of $W(t)$
θ	=	Unknown parameters
$f(x, u, \theta)$	=	SDE drift
$G(\theta)$	=	SDE diffusion matrix
$h(x, u, \theta)$	=	Output function
$R(\theta)$	=	Measurement noise covariance
$Q(\theta)$	=	Process noise covariance
$q(x_{0:N})$	=	Assumed density
$\mathcal{Q}$	=	Search space of assumed densities q
$p_\theta(a b)$	=	Probability density of the random variable A at the value a , given $B = b$, under the parameter θ

*Teaching Assistant Professor, Department of Mechanical, Materials, and Aerospace Engineering, 1306 Evansdale Drive.

const = Constant terms which do not change the location of maxima

Superscripts

A^T = A transpose
 A^{-1} = A inverse

Subscripts

x_k, u_k = $x(kT), u(kT)$ evaluation at a sample index
 $x_{i:j}$ = $x_i, x_{i+1}, \dots, x_{j-1}, x_j$ sequence range

Acronyms

DLR *Deutsches Zentrum für Luft- und Raumfahrt*
ELBO Evidence Lower Bound
FEM Filter-Error Method
KLD Kullback–Leibler Divergence
MAP Maximum *a Posteriori*
ML Maximum Likelihood
OEM Output Error Method
SDE Stochastic Differential Equation
VI Variational Inference

I. Introduction

AIRCRAFT system identification, the art and science [1] of determining mathematical models of aircraft from flight-test data, has become increasingly important in the development of new aircraft [2]. As the field evolves [3–7] the problems that can be solved at the state-of-the-art are often limited by what estimation methods can be run at the computational hardware available.

Maximum likelihood (ML) approaches such as the equation-error, output-error, and filter-error methods are widely used in aircraft system identification due to their asymptotic efficiency properties. When both process noise and measurement noise are considered, the ML estimator is the filter-error method (FEM), which employs a Kalman filter to approximate the prior predictive distribution of each measurement, given the previous measurements up to that point. There are multiple formulations of this method, but they either require solution of a Riccati equation, the most common approach in the aircraft literature [8–10, Chap. 5]; or they require the estimation of the filter gain as an independent parameter [11], the most common approach for black-box estimation, which can lead to unstable predictors in the optimization process. A consequence of these limitations is that the FEM requires good initial parameter estimates for convergence to occur.

Recently [12, 13], a new formulation of the ML estimator has been proposed for systems subject to process and measurement noise, using variational inference (VI). This approach was then reframed as unconstrained optimization and reparameterized for steady-state problems by the author in Ref. [14], reducing the number of decision variables and complexity of the estimator. VI uses a different paradigm than the FEM for performing ML. The FEM uses a filter to approximate the prior predictive distribution and, assuming that distribution is correct, computes the likelihood. In VI, the estimation problem is augmented with the smoothing posterior distribution of the states, and a lower bound on the likelihood function is optimized instead. The approximation error between the candidate posterior and the true one is encoded in the objective function.

In this paper, the application of variational system identification is demonstrated for estimating aircraft parameters from flight-test data, an important use case of practical interest [2]. This is also the first application of the method in non-simulated data. Our goal is to show that the estimator’s usefulness and its results as a drop-in replacement for the FEM and OEM. The results show that it is as good as, if not better than, those methods. In particular, convergence is attained with null initial guesses for all system parameters, in all examples tested in this paper.

II. Problem Formulation

The goal of ML estimators for continuous-time dynamical systems is to obtain the vector $\theta \in \mathbb{R}^{n_\theta}$ of unknown parameters of a system modeled by the stochastic differential equation (SDE)

$$dx(t) = f(x(t), u(t), \theta) dt + G(\theta) dW(t), \quad (1)$$

where $x: \mathbb{R} \rightarrow \mathbb{R}^{n_x}$ are the states of the system, $u: \mathbb{R} \rightarrow \mathbb{R}^{n_u}$ are the known external inputs, and the model functions f and G are given. The SDE formulation is a more formal mathematical representation of the differential equation

$$\dot{x}(t) = f(x(t), u(t), \theta) + G(\theta)w(t), \quad (2)$$

where $w(t)$ is the weak derivative of the Wiener process $W(t)$, i.e., white Gaussian noise with spectral density

$$\mathbb{E} [w(t)w(\tau)^\top] = \mathbf{I} \delta(t - \tau). \quad (3)$$

We consider a continuous–discrete formulation where the states are indirectly measured at discrete time instants through the output vector

$$y_k = h(x(kT), u(kT), \theta) + v_k, \quad k = 0, 1, \dots, N; \quad (4)$$

where k is the sample index, v_k is the measurement noise, T is the sampling period, and the output function h is given. The measurement noise is assumed to be a zero-mean, white, Gaussian noise process, independent from the process noise, with constant covariance $R(\theta)$:

$$\mathbb{E} [v_i v_j^\top] = R(\theta) \delta_{ij}, \quad i, j = 0, 1, \dots, N. \quad (5)$$

To apply the estimator, the SDE is discretized using a method such as the Euler–Maruyama scheme, chosen here for its simplicity. Given a prior distribution for the initial state x_0 , the joint density of the discretized state path $x_0, x_1, \dots, x_N$, written in shorthand as $x_{0:N}$, is

$$p_\theta(x_{0:N}) = p_\theta(x_0) \prod_{k=1}^N p_\theta(x_k | x_{k-1}), \quad (6)$$

where $p_\theta(x_0)$ is the initial-state prior density, and the state transition density under the Euler–Maruyama scheme is the multivariate normal with the following parameters:

$$p_\theta(x_k | x_{k-1}) = \mathcal{N}(x_k; F(x_{k-1}, u_{k-1}, \theta), Q(\theta)), \quad (7)$$

$$F(x_{k-1}, u_{k-1}, \theta) = x_k + T f(x_{k-1}, u_{k-1}, \theta), \quad (8)$$

$$Q(\theta) = T G(\theta) G(\theta)^\top. \quad (9)$$

Similarly, Eqs. (4)–(5) imply that

$$p_\theta(y_k | x_k) = \mathcal{N}(y_k; h(x_k, u_{k-1}, \theta), R(\theta)), \quad (10)$$

$$p_\theta(y_{0:N} | x_{0:N}) = \prod_{k=0}^N p_\theta(y_k | x_k), \quad (11)$$

laying the groundwork for applying the estimator.

III. Variational Inference

In this section, the variational inference formulation of the maximum likelihood problem is derived, in a manner similar to [12, 14]. Starting from Bayes’ rule,

$$p_\theta(x_{0:N} | y_{0:N}) = \frac{p_\theta(x_{0:N}, y_{0:N})}{p_\theta(y_{0:N})}, \quad (12)$$

we have that for every $x_{0:N}$ in the support of the posterior, the likelihood $p_\theta(y_{0:N})$ of a parameter vector θ is the ratio between the density of the complete data and the posterior:

$$p_\theta(y_{0:N}) = \frac{p_\theta(x_{0:N}, y_{0:N})}{p_\theta(x_{0:N}|y_{0:N})}. \quad (13)$$

This equation is not of practical use at any single state path $x_{0:N}$, but can be used to bound the likelihood when evaluated at a population of paths. For any density $q(x_{0:N})$ whose support lies in the support of the posterior*, we have that

$$\log p_\theta(y_{0:N}) = \log \left(\frac{p_\theta(x_{0:N}, y_{0:N})}{q(x_{0:N})} \times \frac{q(x_{0:N})}{p_\theta(x_{0:N}|y_{0:N})} \right) = \log \left(\frac{p_\theta(x_{0:N}, y_{0:N})}{q(x_{0:N})} \right) + \log \left(\frac{q(x_{0:N})}{p_\theta(x_{0:N}|y_{0:N})} \right). \quad (14)$$

Taking the expectation with respect to q , we have that

$$\log p_\theta(y_{0:N}) = \int \log \left(\frac{p_\theta(x_{0:N}, y_{0:N})}{q(x_{0:N})} \right) q(x_{0:N}) dx_{0:N} + \int \log \left(\frac{q(x_{0:N})}{p_\theta(x_{0:N}|y_{0:N})} \right) q(x_{0:N}) dx_{0:N}. \quad (15)$$

The last term on the right-hand side of Eq. (15) is the Kullback–Leibler divergence (KLD) of q from the posterior, which is always non-negative

$$\text{KL}[q(x_{0:N})||p_\theta(x_{0:N}|y_{0:N})] = \int \log \left(\frac{q(x_{0:N})}{p_\theta(x_{0:N}|y_{0:N})} \right) q(x_{0:N}) dx_{0:N} \geq 0. \quad (16)$$

The KLD, also known as relative entropy, quantifies the difference between two probability distributions, being zero if and only if they are equal almost everywhere. This means that the first term on the right-hand side of Eq. (15) is a lower bound on the likelihood, or model evidence, which we will denote the evidence lower bound (ELBO),

$$\mathcal{L}(\theta, q) := \int \log \left(\frac{p_\theta(x_{0:N}, y_{0:N})}{q(x_{0:N})} \right) q(x_{0:N}) dx_{0:N}. \quad (17)$$

Substituting Eqs. (16)–(17) into Eq. (15), it is evident that the bound gets tighter as q approaches the posterior, as quantified by the KLD:

$$\log p_\theta(y_{0:N}) = \mathcal{L}(\theta, q) + \text{KL}[q(x_{0:N})||p_\theta(x_{0:N}|y_{0:N})] \geq \mathcal{L}(\theta, q). \quad (18)$$

This lays out the basis of variational inference for maximum likelihood estimation: by searching for both the parameter vector θ and the density q which maximize the ELBO $\mathcal{L}(\theta, q)$, we are simultaneously maximizing the likelihood function and minimizing the divergence from q to the true posterior. The density q , called the assumed density, should be chosen from a family of distributions $\mathcal{Q}$ that is general enough to yield tight bounds, but still be tractable enough to compute and optimize the ELBO. The ML estimation problem using VI is then

$$\underset{q \in \mathcal{Q}, \theta \in \mathbb{R}^{n_\theta}}{\text{maximize}} \mathcal{L}(\theta, q). \quad (19)$$

In situations where the true posterior is contained in $\mathcal{Q}$, this formulation is equivalent to ML estimation. When the likelihood function is intractable to compute, as it is in parameter estimation of nonlinear systems, VI with its assured lower bound offers stronger theoretical guarantees than when approximations like the extended Kalman filter or particle smoothers are used to compute and maximize the likelihood [15, 16].

VI can also be understood as a regularized generalization of maximum *a posteriori* (MAP) estimation over a population of points [see 14, Sec. 3.2]. Eq. (17) can be split into the expected complete-data log-density and the differential entropy of the assumed density q :

$$\mathcal{L}(\theta, q) = \int \log p_\theta(x_{0:N}, y_{0:N}) q(x_{0:N}) dx_{0:N} - \int \log q(x_{0:N}) q(x_{0:N}) dx_{0:N}. \quad (20)$$

$$= \int \log p_\theta(x_{0:N}, y_{0:N}) q(x_{0:N}) dx_{0:N} + h_q[X_{0:N}]. \quad (21)$$

In MAP estimation, a single state path is chosen as the most probable, given the observations. In VI, a population of paths, defined by q , is chosen to take into account the uncertainty in the states. The differential entropy of the population of paths $h_q[X_{0:N}]$, which increases as the population is spread out over a larger region, is included to prevent the population from collapsing at a single point, the posterior mode.

*Or, equivalently, the measure associated with q is absolutely continuous with respect to the posterior probability measure.

IV. Variational System Identification

As discussed in the previous section, to apply VI for solving a ML estimation problem, we need to define both the underlying dynamic system and noise model, as in Sec. II, as well as the search space $\mathcal{Q}$ of the assumed density q . For the problem defined in Sec. II, a key property for making the underlying optimization problem sparse is to choose q from a space of densities that can be represented by a factor graph with the same topology as that of the true posterior. This amounts to choosing q under which the process x_k has the Markov property [see 14, Sec. 5].

For tractability in computing the ELBO in Eq. (21), we also choose the assumed density q to be multivariate normal, as it has many tractable methods for computing expectations with respect to it. The differential entropy of the multivariate normal has a simple closed-form expression. As shown in Ref. [14, Sec. 5], a multivariate normal distribution $q(x_{0:N})$ with the Markov property can be represented as

$$q(x_{0:N}) = q(x_0) \prod_{k=1}^N q(x_k | x_{k-1}), \quad (22)$$

where each term is given by

$$q(x_0) = \mathcal{N}(x_0; \mu_0, \Sigma_0), \quad q(x_k | x_{k-1}) = \mathcal{N}(x_k; \mu_k + \Sigma_{k,k-1} \Sigma_{k-1}^{-1} (x_{k-1} - \mu_{k-1}), \Sigma_{k|k-1}). \quad (23)$$

In linear systems and nonlinear systems operating at small deviations form an equilibrium point, as is often the case in aircraft system identification, the posterior covariances converge to steady-state values when the data record is long enough. Under this simplification, the covariances for all $k = 0, \dots, N$ are given by

$$\Sigma_{k|k-1} = \Sigma_{\text{cond}}, \quad \Sigma_{k,k-1} = \Sigma_{\text{cross}}, \quad \Sigma_k = \Sigma_{\text{cond}} + \Sigma_{\text{cross}} \Sigma_{\text{marg}}^{-1} \Sigma_{\text{cross}}^T, \quad (24)$$

and $\mu_{0:N}$, $\Sigma_{\text{cond}} \succ 0$, and Σ_{cross} are the parameters of q . This is the steady-state parameterization of Ref. [14], which is an analog of the FEM using a steady-state Kalman filter. We also note that this representation has a very simple differential entropy, as the chain rule for differential entropy gives

$$h_q[X_{0:N}] = h_q[X_0] + \sum_{k=1}^N h_q[X_k | X_{k-1}] = \frac{1}{2} \log \det \Sigma_{\text{marg}} + \frac{N}{2} \log \det \Sigma_{\text{cond}} + \text{const}, \quad (25)$$

where const denotes constant terms which do not change the location of maxima.

With the parameterization of the assumed density defined, we can return to the objective function defined in Eq. (21) and substitute Eqs. (6) and (11) to obtain

$$\begin{aligned} \mathcal{L}(\theta, q) = & \int \log p_\theta(x_0) q(x_0) dx_0 + \sum_{k=1}^N \int \log p_\theta(x_k | x_{k-1}) q(x_{k-1:k}) dx_{k-1:k} \\ & + \sum_{k=0}^N \int \log p_\theta(y_k | x_k) q(x_k) dx_k + h_q[X_{0:N}]. \end{aligned} \quad (26)$$

Instead of having an expectation over a high-dimensional space, this form of the ELBO only requires expectation over the states at each time sample or at consecutive time-sample pairs. This allows integration using sigma points, which is exact for linear systems and a reasonable approximation for nonlinear systems under small deviations from trim.

For simplicity, in this paper we use sigma points without the center one, which gives all remaining points equal weight. If we let $x_k^{(i)}$ denote the sigma points associated with $q(x_k)$ or $q(x_{k-1:k})$, the objective function is then

$$\begin{aligned} \mathcal{L}(\theta, q) = & \frac{1}{2n_x} \sum_{i=1}^{2n_x} p_\theta(x_0^{(i)}) + \frac{1}{4n_x} \sum_{k=1}^N \sum_{i=1}^{4n_x} \log p_\theta(x_k^{(i)} | x_{k-1}^{(i)}) + \frac{1}{2n_x} \sum_{k=0}^N \sum_{i=1}^{2n_x} \log p_\theta(y_k | x_k^{(i)}) \\ & + \frac{1}{2} \log \det \Sigma_{\text{marg}} + \frac{N}{2} \log \det \Sigma_{\text{cond}} + \text{const}. \end{aligned} \quad (27)$$

This function can then be maximized using standard nonlinear programming packages. Note that $x_k^{(i)}$ depend on q and its underlying parameters. Changes in $\mu_{0:N}$ translate all sigma points by some amount, while increasing Σ_{marg} increases their spread, decreasing the average log-density but increasing the entropy. The optimum balances these competing criteria.

V. Results

This section describes the application of the variational system identification to real flight-test data. Freely accessible data from Refs. [10, 17] was used to demonstrate the method. The data are of a HFB-320 Hansa Jet, a VFW-Fokker 614, and a DHC-6 Twin Otter. These represent a wide range of aircraft: a business jet, a jet airliner, and a STOL utility aircraft. The models used in the textbook examples of Refs. [10, 17] were used in this paper.

The open source Python package `visid`, developed by the author and released[†] on Github, was used to implement the variational estimator. The code for the experiments in this paper is also released[‡] on Github as open source software. The underlying optimization problem was solved with the open-source nonlinear program solvers in `scipy.optimize.minimize`, part of the SciPy package [18]. Automatic differentiation with JAX [19] was used to compute the objective function's gradient and Hessian. The initial guess of all optimization decision variables was zero, except for the HFB-320, which had the smoother mean $\mu_{0:N}$ initially set to the measurements of the states in $y_{0:N}$.

As argued in Ref. [9, Sec. IV.B], care must be taken when evaluating the results of the FEM, as the filter corrections can yield a very good match to the data. The same applies to the variational method with respect to the smoother output. Because of this, the optimal parameters θ^* estimated by variational system identification were evaluated using the following criteria:

- 1) The output associated to the smoother mean, that is, $y_k^* = h(\mu_k, u_k, \theta^*)$.
- 2) The output of the free simulation using θ^* .
- 3) The one-step-ahead predictions of the steady-state Kalman filter associated to θ^* .
- 4) The difference between the smoother-mean finite differences, $\frac{\mu_{k+1} - \mu_k}{T}$ and the model drift function $f(\mu_k, u_k, \theta^*)$.

These quantities show the two things the estimator seeks to minimize: the output error and the equation error. When two repetitions of the same maneuver are available, one dataset is used for estimation and another for validation. If only one repetition is available, no validation is performed. For validation, the VI method is run with the estimated model parameters θ^* fixed, for estimation of the assumed density q only (smoothing).

A. Results for the HFB-320

This example corresponds to an elevator 3-2-1-1 input followed by a pulse on a HFB-320 of the *Deutsches Zentrum für Luft- und Raumfahrt* (DLR). The application of both the OEM and FEM to this dataset is reported in Ref. [10, Sec. 5.11.2]. The collocation-based OEM was also applied to this dataset in Ref. [20, Sec. III.A]. A nonlinear longitudinal model with 4 states, 7 outputs and 2 inputs was fit to the data. Please refer to Refs. [10, 20] or the accompanying software of this paper and its references for the model equations.

The model fit to the estimation data is shown in Figs. 1–2. A single dataset was provided, there is no independent validation dataset. As noted in Ref. [10], there is considerable atmospheric turbulence in this record, as can be seen in the measurements of α and a_z in Fig. 1. The free simulation shows a similar fit than the OEM in Refs. [10, 20], with most error in V_T and θ where the low damping and frequency of the phugoid accumulates perturbations due to turbulence.

The smoother output and KF predictor are very close to the FEM results in Ref. [10]. A closer inspection of Fig. 2 shows that the derivative of the smoothed state mean is very noisy for V_T and α . This is likely a result of excessive smoother corrections due to overconfidence in the measured values. The measurements of V_T , a_x , and a_z have visible quantization levels and the estimated measurement noise standard deviations are well below this interval. No bounds were placed on R or Q in this example, however, for comparison with the other references.

B. Results for the DHC-6 Lateral-Directional Motion

This example corresponds to aileron and rudder multi-step inputs on a DHC-6 Twin Otter of the NASA Glenn Research Center. The application of both the OEM and EEM to this dataset is reported in Ref. [17, Examples 5.1, 6.1, and 7.1]. A linear lateral directional model with 4 states, 5 outputs and 2 inputs was fit to the data, parameterized by the nondimensional stability and control derivatives. Please refer to Ref. [17] or the accompanying software of this paper and its references for the model equations.

The model fit to the estimation data is shown in Figs. 3–6. There is little turbulence in the records and an excellent match is obtained for the estimation (Figs. 3–4) and validation (Figs. 5–6) datasets. The divergence of the roll angle ϕ in the free simulation of the validation dataset is attributed to different values of the bias parameter in that record and the accumulation of error in the slow spiral mode.

[†]<https://github.com/dimasad/visid>

[‡]<https://github.com/dimasad/scitech-2025-code>

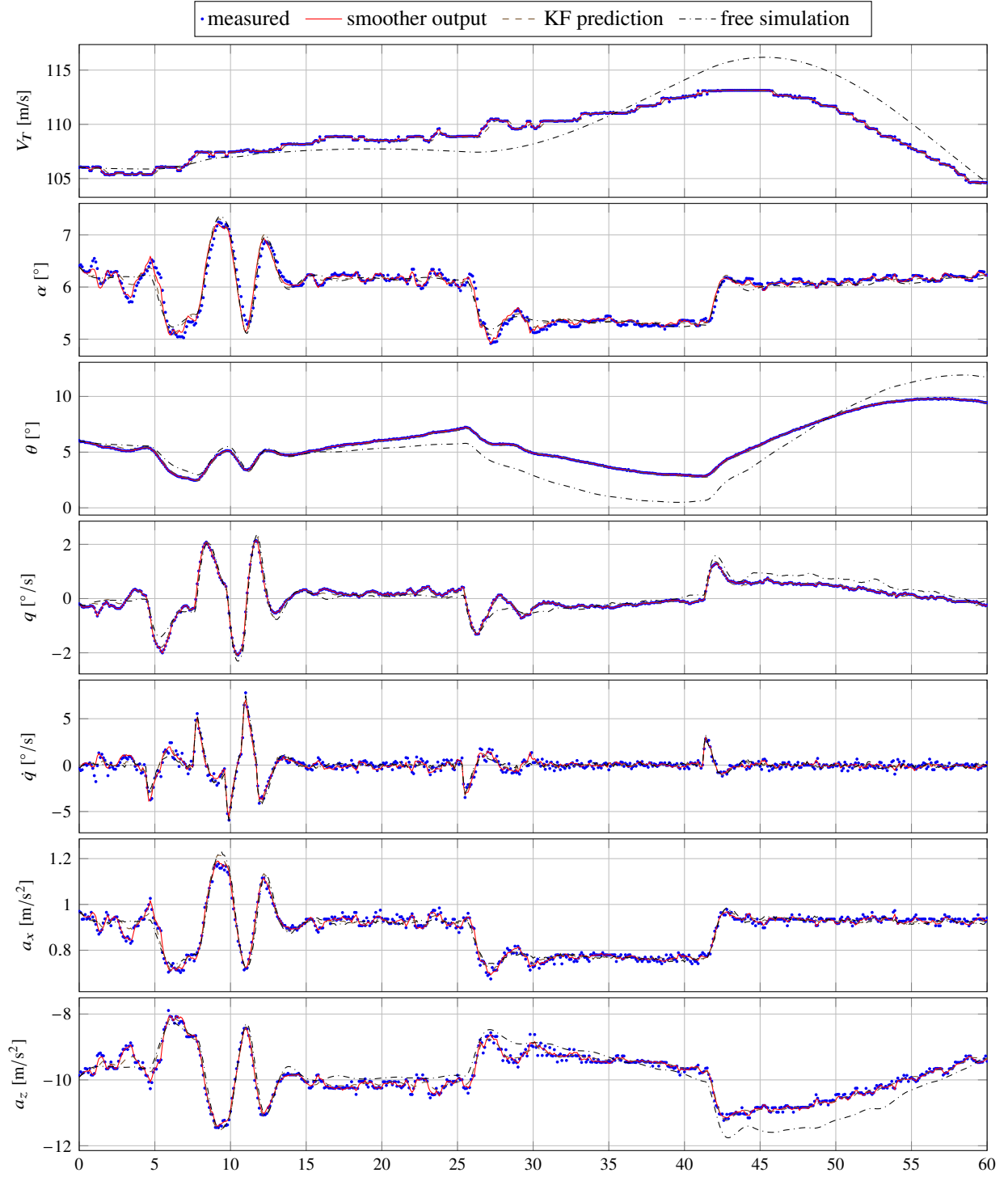


Fig. 1 Evaluation of model outputs for the optimal parameters estimated by variational system identification on the HFB-320 flight-test data. This is the same dataset that was used to estimate the model.

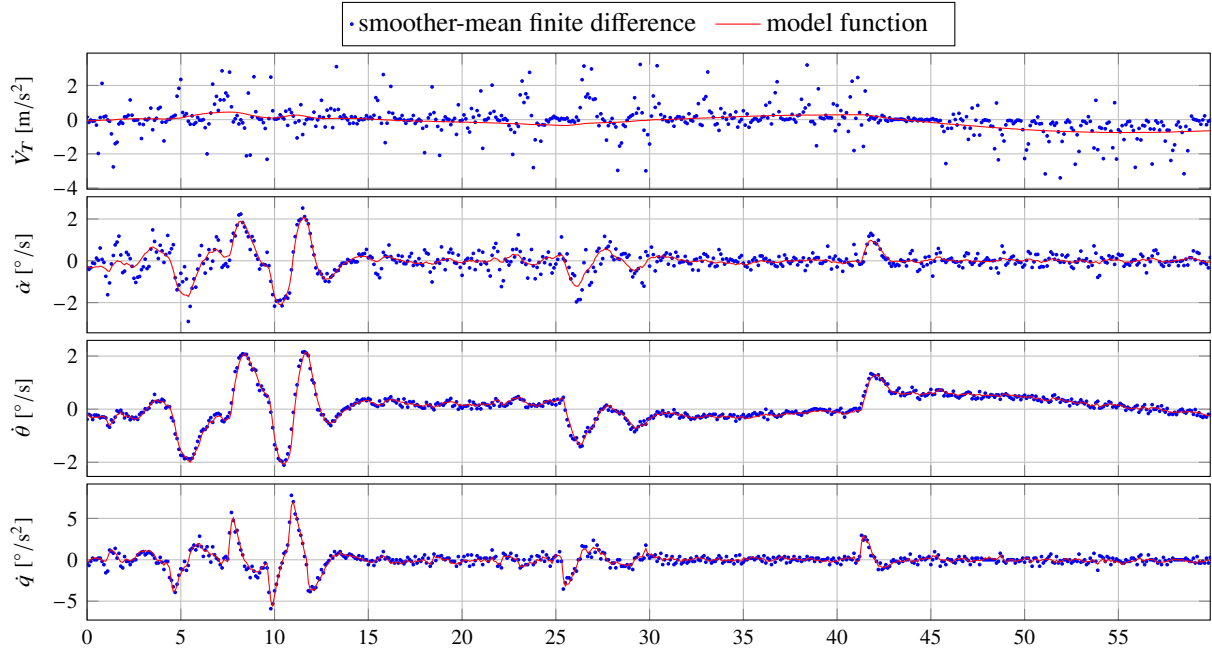


Fig. 2 Evaluation of state derivatives for the optimal parameters estimated by variational system identification on the HFB-320 flight-test data. This is the same dataset that was used to estimate the model.

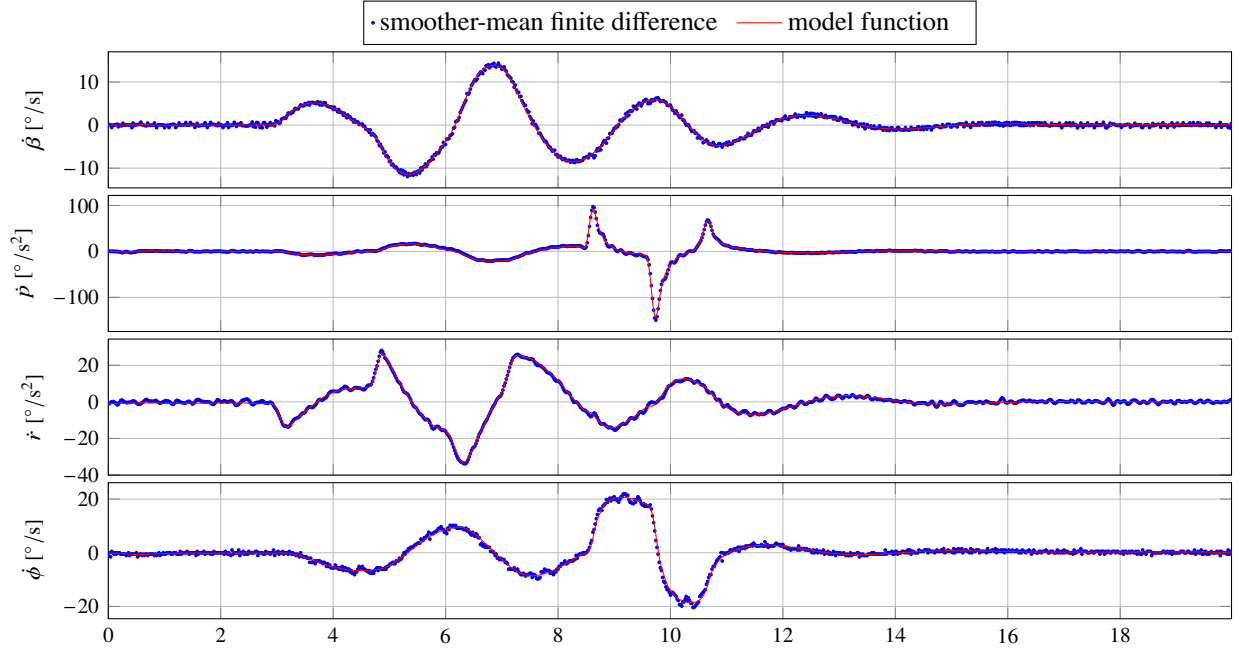


Fig. 3 Evaluation of state derivatives for the optimal parameters estimated by variational system identification on the DHC-6 lateral-directional motion flight-test data. This is the same record that was used to estimate the model.

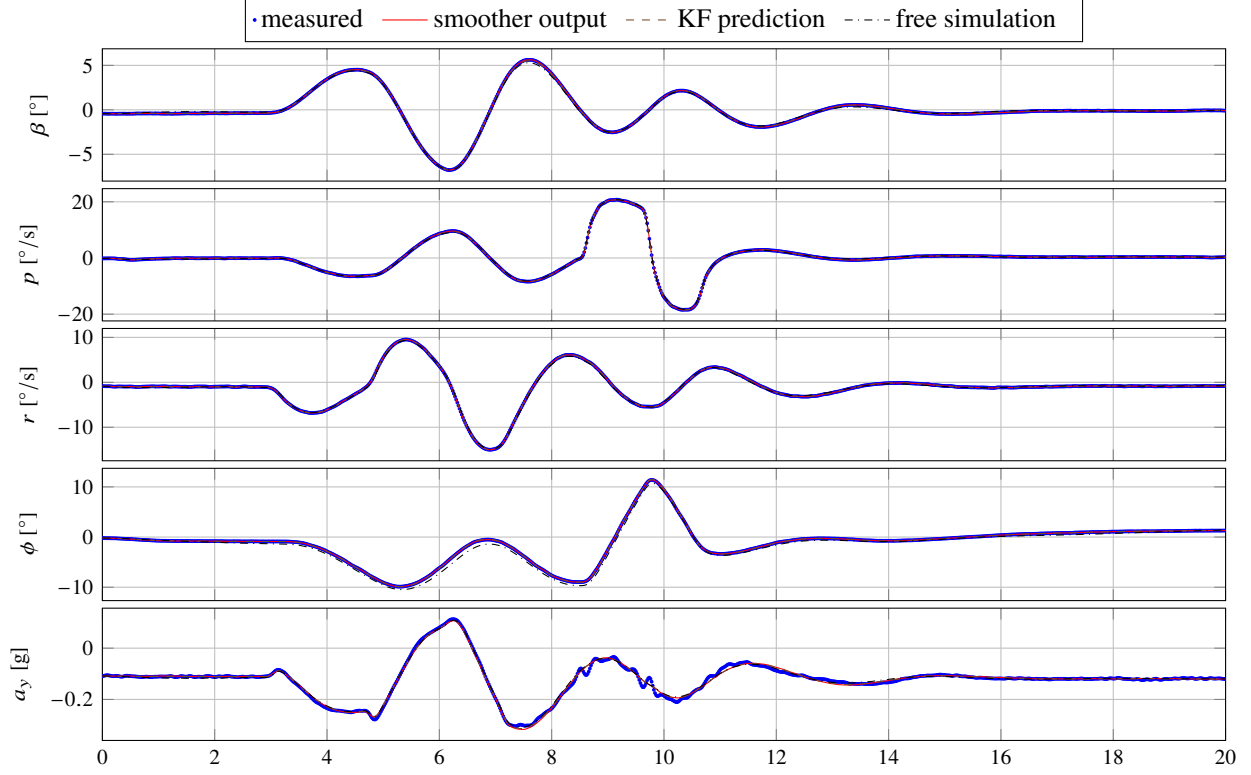


Fig. 4 Evaluation of model outputs for the optimal parameters estimated by variational system identification on the DHC-6 lateral-directional motion flight-test data. This is the same record that was used to estimate the model.

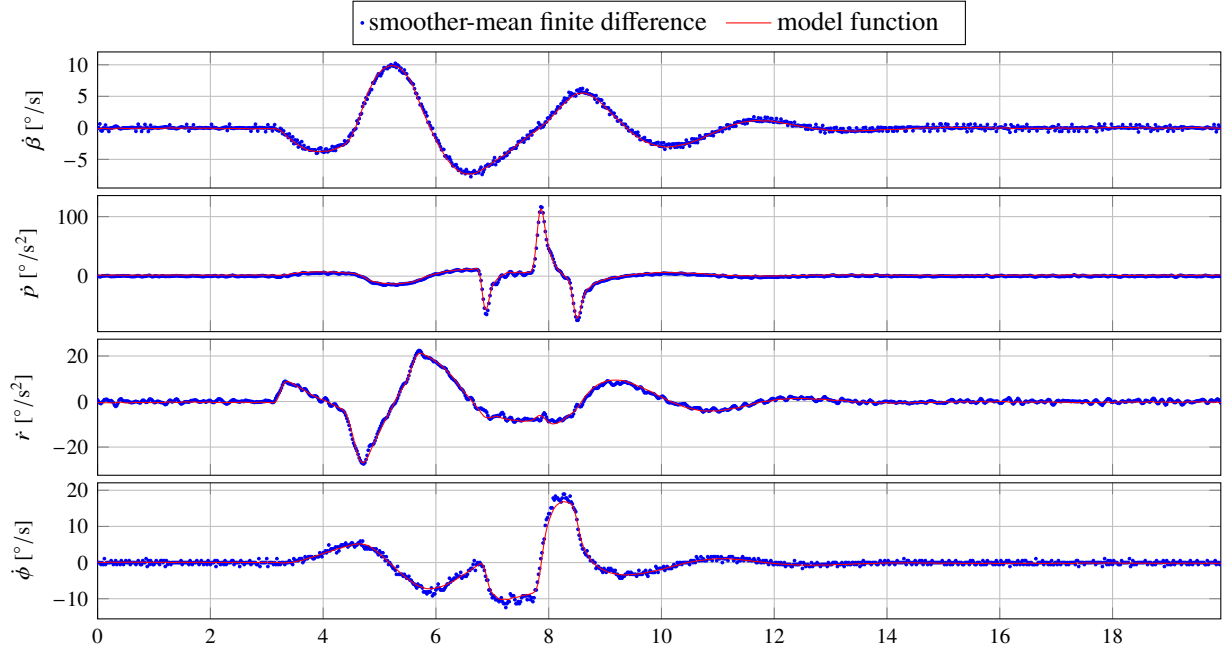


Fig. 5 Evaluation of state derivatives for the optimal parameters estimated by variational system identification on the DHC-6 lateral-directional motion validation flight-test data. This is a different record than the one used to estimate the model.

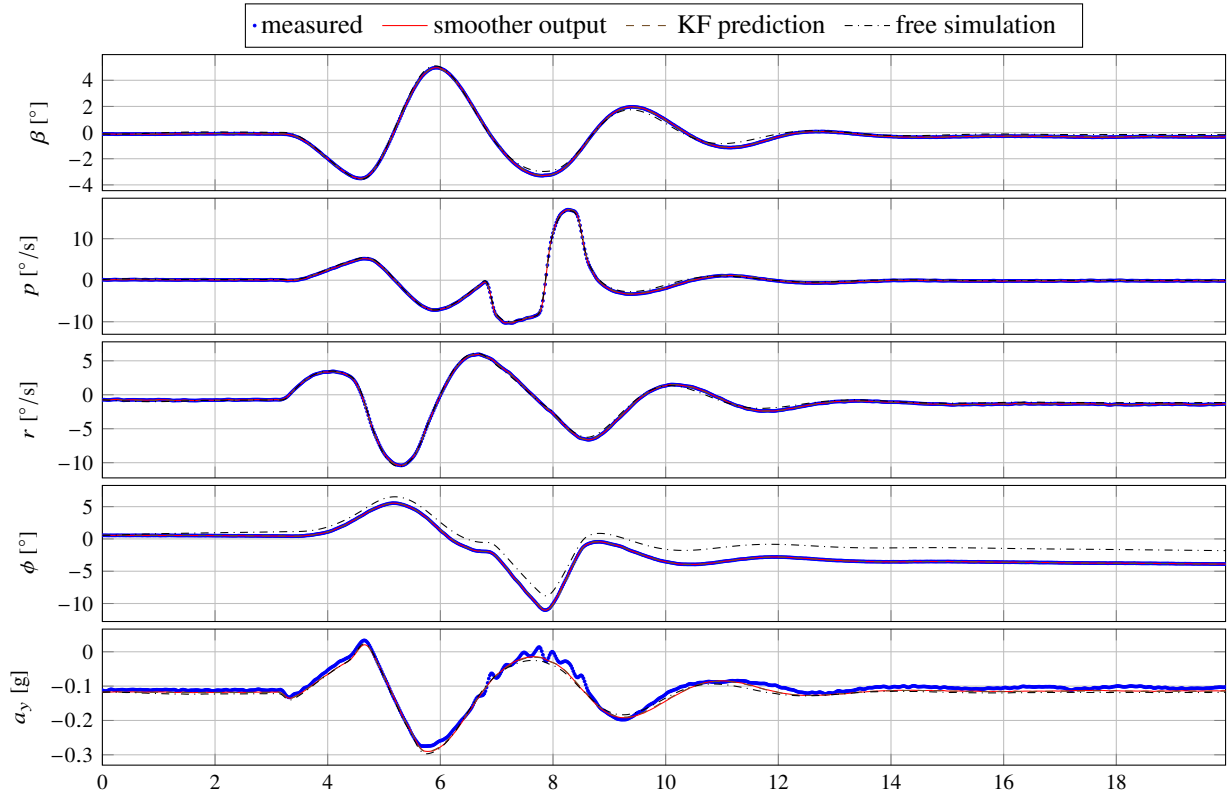


Fig. 6 Evaluation of model outputs for the optimal parameters estimated by variational system identification on the DHC-6 lateral-directional motion validation flight-test data. This is a different record than the one used to estimate the model.

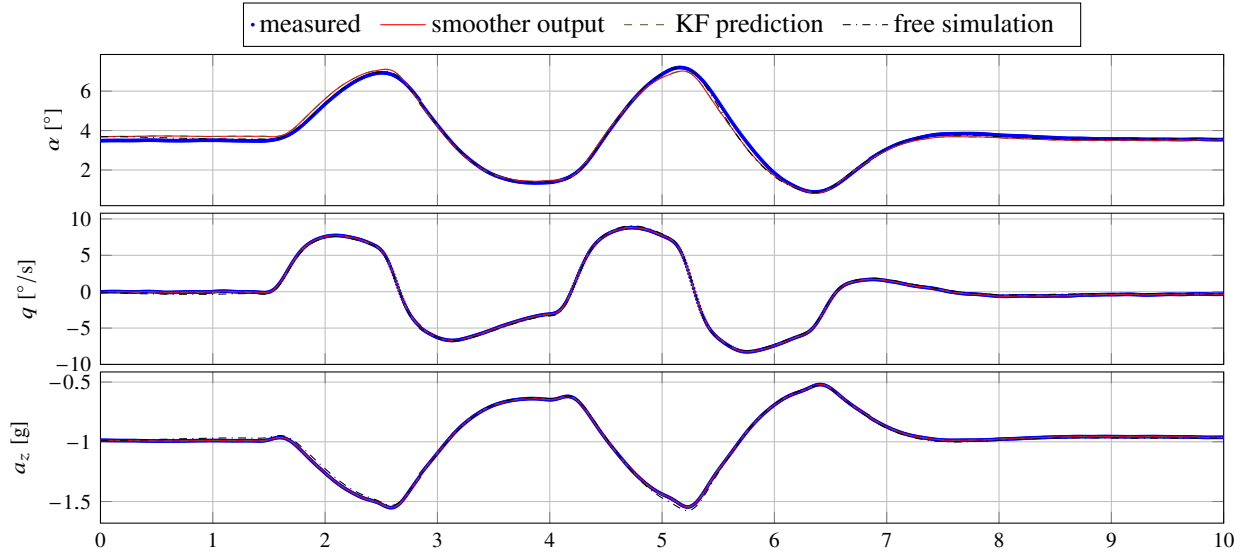


Fig. 7 Evaluation of model outputs for the optimal parameters estimated by variational system identification on the DHC-6 short period motion flight-test data. This is the same record that was used to estimate the model.

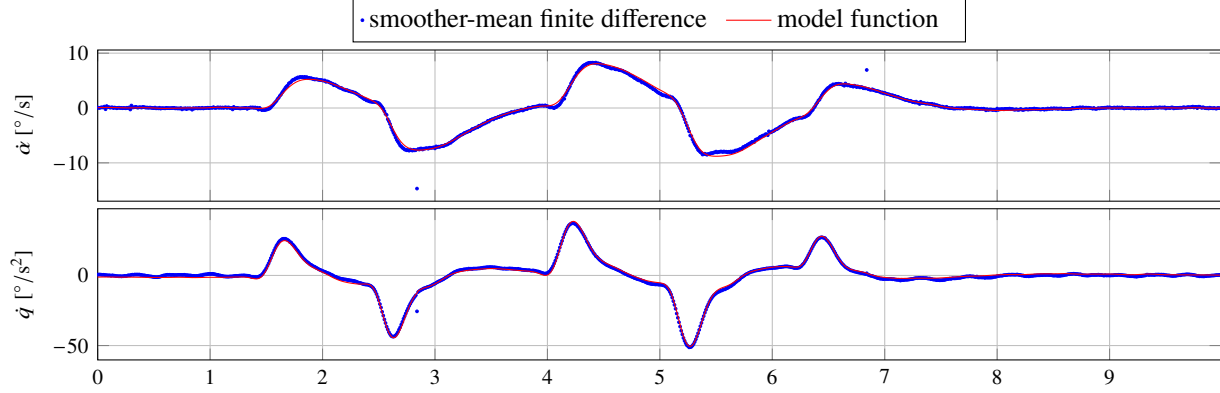


Fig. 8 Evaluation of state derivatives for the optimal parameters estimated by variational system identification on the DHC-6 short period motion flight-test data. This is the same record that was used to estimate the model.

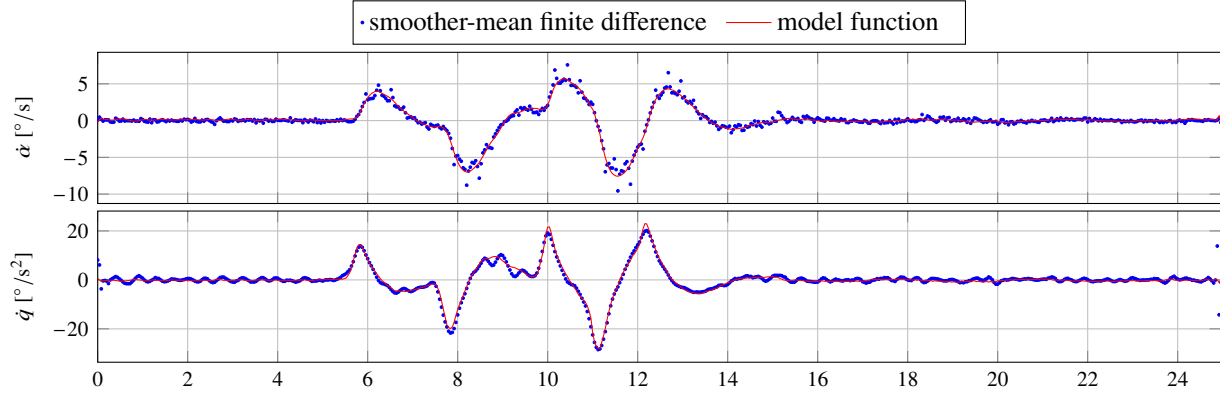


Fig. 9 Evaluation of state derivatives for the optimal parameters estimated by variational system identification on the VFW-Fokker 614 short period motion flight-test data. This is the same record that was used to estimate the model.

Some quantization effects can be noticed in the first and last plots of Figs. 3 and 5, but the signal to noise ratio is much larger than in the example of Sec. V.A due to higher resolution. The smoother is overcorrecting to β , p , and r , but by a small amount. No bounds were placed on R or Q in this example as well, for comparison with the other references.

C. Results for the DHC-6 short period motion

This example corresponds to an elevator multi-step input on a DHC-6 Twin Otter of the NASA Glenn Research Center. The application of the OEM to this dataset is reported in Ref. [17, Example 6.3]. A linear short period motion model with 2 states, 3 outputs and 1 input was fit to the data, parameterized by the dimensional stability and control derivatives. Please refer to Ref. [17] or the accompanying software of this paper and its references for the model equations.

The model fit to the estimation data is shown in Figs. 7–8. There is little turbulence in the records and an excellent match is obtained for the vertical acceleration a_z and the pitch rate q . A low-frequency error is present on the angle of attack α due to unmodelled phugoid dynamics. A single dataset was provided, so there is no independent validation dataset.

D. Results for the VFW-Fokker 614 Short Period Motion

This example corresponds to an elevator 3-2-1-1 input on a VFW-Fokker 614 of the DLR. The application of the OEM and nonlinear filters to this dataset is reported in Ref. [10, Sec. 7.5.2]. A linear longitudinal model with 2 states, 2 outputs and 1 input was fit to the data, parameterized by the dimensional stability and control derivatives. Please refer

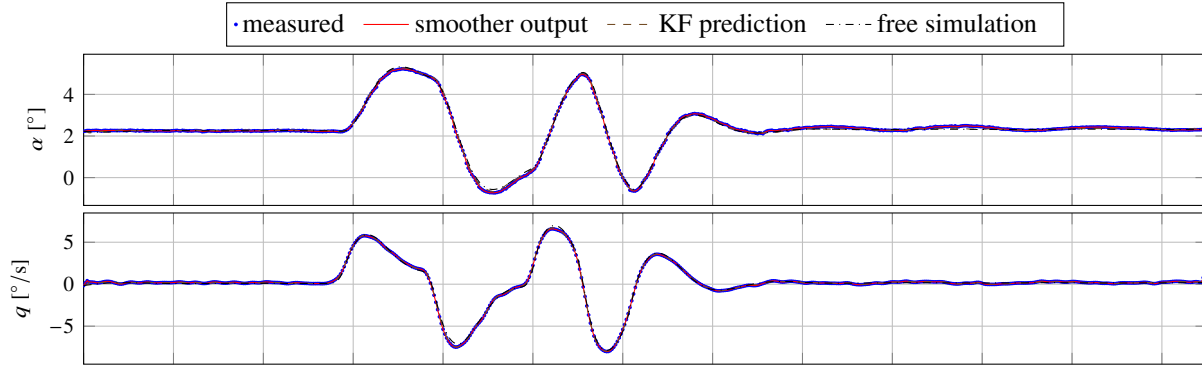


Fig. 10 Evaluation of model outputs for the optimal parameters estimated by variational system identification on the VFW-Fokker 614 short period motion flight-test data. This is the same record that was used to estimate the model.

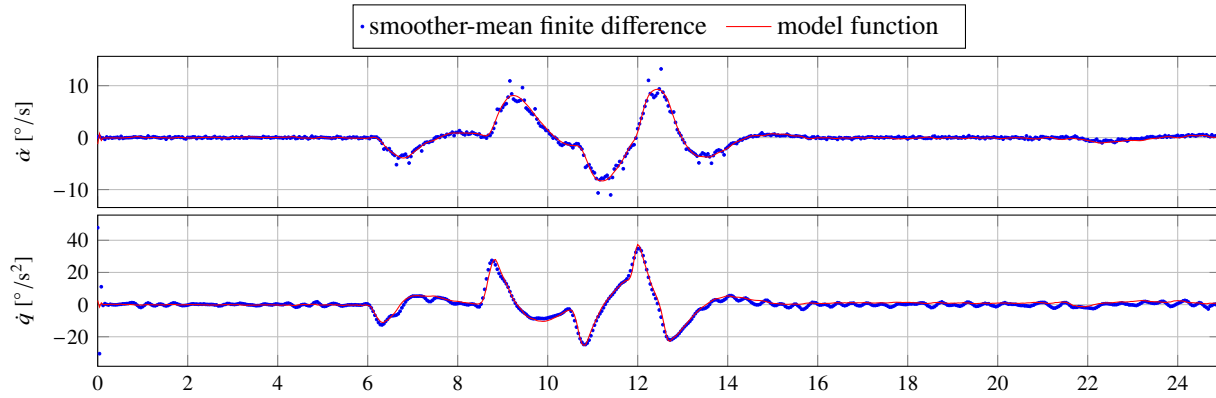


Fig. 11 Evaluation of state derivatives for the optimal parameters estimated by variational system identification on the VFW-Fokker 614 short period motion validation flight-test data. This is a different record than the one used to estimate the model.

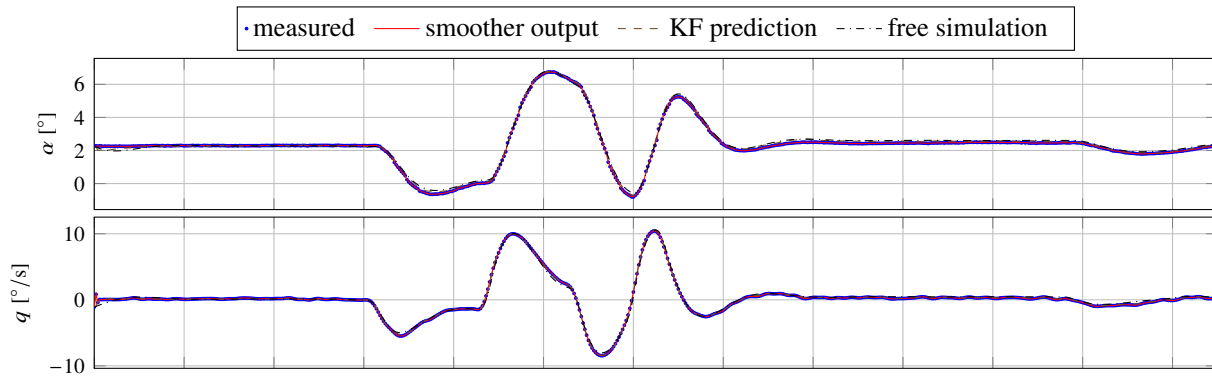


Fig. 12 Evaluation of model outputs for the optimal parameters estimated by variational system identification on the VFW-Fokker 614 short period motion validation flight-test data. This is a different record than the one used to estimate the model.

to Ref. [10] or the accompanying software of this paper and its references for the model equations.

The model fit to the estimation data is shown in Figs. 9–12. There is little turbulence in the records and an excellent match is obtained for the estimation (Figs. 9–10) and validation (Figs. 11–12) datasets. The free simulation of the estimated model is very close to the results of the OEM in Ref. [10].

VI. Conclusion and Future Work

This paper showed the application of variational system identification for a wide range of models and aircraft. For flights in turbulence, like shown in Sec. V.A, the results of the method are very similar to those of the FEM. A major practical advantage of variational system identification over the OEM and FEM, however, is that the estimator converges to the optimum, for all examples shown, with the initial guess of all system parameters set to zero.

Beyond the applications shown here, variational system identification has the potential to enable treatment of both process and measurement noise in problems where the FEM or OEM are not currently applicable. As shown in [14], variational system identification can be used for processing long records of data, to estimate parameters from models with a large number of states, and to estimate models for which good initial parameter estimates are not available. The computation of the objective function and its derivatives is embarrassingly parallel and can benefit from great speedups when computed in modern graphical processing units (GPUs). The strong theoretical support of VI also makes it a great candidate for global nonlinear modeling and treatment of non-Gaussian noise.

Acknowledgments

This research was made possible by the NASA Established Program to Stimulate Competitive Research, Grant # 80NSSC20M0055. Computational resources were provided by the WVU Research Computing Thorny Flat and Dolly Sods HPC clusters, which are funded in part by NSF OAC-1726534 and NSF OAC-2117575.

References

- [1] Jategaonkar, R. V., “Flight Vehicle System Identification-Engineering Utility,” *Journal of Aircraft*, Vol. 42, No. 1, 2005, pp. 11–11. <https://doi.org/10.2514/1.15766>.
- [2] Grauer, J. A., and Morelli, E. A., “Introduction to the Advances in Aircraft System Identification from Flight Test Data Virtual Collection,” *Journal of Aircraft*, Vol. 60, No. 5, 2023, pp. 1329–1330. <https://doi.org/10.2514/1.C037583>.
- [3] Jategaonkar, R. V., Fischenberg, D., and Gruenhagen, W., “Aerodynamic Modeling and System Identification from Flight Data: Recent Applications at DLR,” *Journal of Aircraft*, Vol. 41, No. 4, 2004, pp. 681–691. <https://doi.org/10.2514/1.3165>.
- [4] Wang, K. C., and Iliff, K. W., “Retrospective and Recent Examples of Aircraft Parameter Identification at NASA Dryden Flight Research Center,” *Journal of Aircraft*, Vol. 41, No. 4, 2004, pp. 752–764. <https://doi.org/10.2514/1.332>.
- [5] Morelli, E. A., and Klein, V., “Application of System Identification to Aircraft at NASA Langley Research Center,” *Journal of Aircraft*, Vol. 42, No. 1, 2005, pp. 12–25. <https://doi.org/10.2514/1.3648>.
- [6] Deiler, C., Mönnich, W., Seher-Weiß, S., and Wartmann, J., “Retrospective and Recent Examples of Aircraft and Rotorcraft System Identification at DLR,” *Journal of Aircraft*, 2023, pp. 1–27. <https://doi.org/10.2514/1.C037262>.
- [7] Morelli, E. A., and Grauer, J. A., “Advances in Aircraft System Identification at NASA Langley Research Center,” *Journal of Aircraft*, Vol. 60, No. 5, 2023, pp. 1354–1370. <https://doi.org/10.2514/1.C037274>.
- [8] Maine, R. E., and Iliff, K. W., “Formulation and Implementation of a Practical Algorithm for Parameter Estimation with Process and Measurement Noise,” *SIAM Journal on Applied Mathematics*, Vol. 41, No. 3, 1981, pp. 558–579. <https://doi.org/10.1137/0141045>.
- [9] Grauer, J. A., and Morelli, E. A., “A New Formulation of the Filter-Error Method for Aerodynamic Parameter Estimation in Turbulence,” *AIAA Atmospheric Flight Mechanics Conference*, American Institute of Aeronautics and Astronautics, Dallas, TX, 2015. <https://doi.org/10.2514/6.2015-2704>.
- [10] Jategaonkar, R., *Flight Vehicle System Identification*, Second ed., AIAA, Reston, VA, 2015. <https://doi.org/10.2514/4.102790>.
- [11] Wills, A., and Ninness, B., “On Gradient-Based Search for Multivariable System Estimates,” *IEEE Transactions on Automatic Control*, Vol. 53, No. 1, 2008, pp. 298–306. <https://doi.org/10.1109/TAC.2007.914953>.

- [12] Courts, J., Hendriks, J., Wills, A., Schön, T. B., and Ninness, B., “Variational State and Parameter Estimation,” *IFAC-PapersOnLine*, Vol. 54, No. 7, 2021, pp. 732–737. <https://doi.org/10.1016/j.ifacol.2021.08.448>.
- [13] Courts, J., Wills, A. G., Schön, T. B., and Ninness, B., “Variational system identification for nonlinear state-space models,” *Automatica*, Vol. 147, 2023, p. 110687. <https://doi.org/10.1016/j.automatica.2022.110687>.
- [14] Dutra, D. A. A., “Parameterizations for Large-Scale Variational System Identification Using Unconstrained Optimization,” *Automatica*, 2024. <https://doi.org/10.48550/arXiv.2404.10137>, (to be published).
- [15] Kristensen, N. R., Madsen, H., and Jørgensen, S. B., “Parameter estimation in stochastic grey-box models,” *Automatica*, Vol. 40, No. 2, 2004, pp. 225–237. <https://doi.org/10.1016/j.automatica.2003.10.001>.
- [16] Schön, T. B., Wills, A., and Ninness, B., “System identification of nonlinear state-space models,” *Automatica*, Vol. 47, No. 1, 2011, pp. 39–49. <https://doi.org/10.1016/j.automatica.2010.10.013>.
- [17] Morelli, E. A., and Klein, V., *Aircraft System Identification: Theory and Practice*, Second ed., Sunflyte Enterprises, Williamsburg, VA, 2016.
- [18] Virtanen, P. et al., “SciPy 1.0: fundamental algorithms for scientific computing in Python,” *Nature Methods*, Vol. 17, No. 3, 2020, pp. 261–272. <https://doi.org/10.1038/s41592-019-0686-2>.
- [19] Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., and Zhang, Q., “JAX: composable transformations of Python+NumPy programs,” , 2018. URL <http://github.com/google/jax>, v0.3.13.
- [20] Dutra, D. A. A., “Collocation-Based Output-Error Method for Aircraft System Identification,” *AIAA Aviation 2019 Forum*, AIAA, 2019. <https://doi.org/10.2514/6.2019-3087>.