


# GUST ERROR METHOD (GEM): PARAMETER IDENTIFICATION FOR AIRCRAFT SUBJECT TO VON KÁRMÁN PROCESS NOISE

Jeremy Cox[1], KC Wong[1] & Gregory Chamitoff[2]

[1]The University of Sydney
[2]Texas A&M University College of Engineering

## Abstract

Identifying an aircraft's stability and control parameters from flight data is confounded by atmospheric gusts, a coloured process noise: output error methods ignore it and become biased and overconfident, while filter error methods, as usually applied, do not pool the heterogeneous responses whose shared parameters separate airframe inertia from the aerodynamic moments. From such a set of heterogeneous manoeuvres on a sub-scale aircraft we recover the roll-axis inertia and aerodynamic moments jointly, and establish a two-pole actuator lag model. This is enabled by the Gust Error Method (GEM) proposed here for linear dynamic systems: the coloured process noise is decomposed by the Karhunen-Loève transform into a low-order basis, whose coefficients are estimated as random effects in a nonlinear mixed effects model. Treating each response's gust as a random effect lets GEM pool heterogeneous responses sharing a parameter set, with calibrated confidence intervals under coloured noise; on synthetic data it recovers parameters across a range of gust intensities with good coverage, matching the filter error method on a single response while extending to the multi-response setting the application requires.

**Keywords:** system identification; process noise; Karhunen-Loève transform; mixed-effects models; von Kármán turbulence

## 1. List of Symbols

| Symbol | Description |
|---|---|
| **Indices** | |
| $i$ | Indexes over sampling units. |
| $j$ | Indexes over observation channels within a sampling unit. |
| $k$ | Indexes over observation timestamps within an observation channel. |
| $m$ | Index variable that does not share meaning outside its local context. |
| **Counts** | |
| $n$ | Number of observations. |
| $M$ | Number of sampling units in an experiment. |
| $N$ | Total number of observations across all sampling units, $N = \sum_i n_i$. |
| $q$ | Number of random effects per sampling unit (dimension of $\vec{b}_i$). |
| **Scalars** | |
| $t$ | Time during a dynamic system response. |
| $\tau$ | Temporal displacement or latency. |
| $b$ | A single random effect variable. |

| Symbol | Description |
|---|---|
| $\lambda$ | An eigenvalue of a matrix. |
| $\sigma^2$ | A variance. |
| $\sigma^2_{\mathrm{PN}}$ | Process-noise intensity (variance) parameter of the roll-mode simulation studies. |
| $\sigma^2_{\mathrm{gyro}}$ | Roll-rate (gyro) measurement-noise variance. |
| $\ell^*$ | Profiled log likelihood for the MLA method. The asterisk indicates profiling out a scale variance $\sigma^2$. |
| $g^*$ | Per-sampling-unit weighted sums of squared residuals and squared random effects. |
| $\eta$ | Mode information ratio (MIR). |
| $p$ | Aircraft roll model, roll rate. |
| $\delta_a$ | Aircraft roll model, aileron control input. |
| $L_p$ | Aircraft roll model $L_p = \frac{d\dot{p}}{dp}$. |
| $L_\delta$ | Aircraft roll model $L_\delta = \frac{d\dot{p}}{d\delta_a}$. |
| $I_T$ | Total body roll-axis inertia for a configuration, $I_T = I_{b,0} + \beta\, I_b$. |
| $I_{b,0}$ | Baseline body roll inertia (no ballast). |
| $I_b$ | Known wingtip-ballast roll-inertia increment (flight 3). |
| $\beta$ | Ballast indicator: $0$ (flights 1, 2), $1$ (flight 3). |
| $\omega_p$ | Propeller angular rate (signed by rotation sense). |
| $V$ | Segment-mean measured true airspeed; fixed per response. |
| $\rho$ | Air density. |
| $S$ | Wing reference area. |
| $b_w$ | Wing span. |
| $C_{\ell_p}$ | Non-dimensional roll-damping derivative. |
| $C_{\ell_{\delta_a}}$ | Non-dimensional aileron roll-control derivative. |
| $k_Q$ | Propeller reaction-torque (roll) coefficient. |
| $p_w$ | Rolling-gust angular rate. |
| $\delta_0$ | Zero-roll aileron neutral: location of the trim band (airframe asymmetry). |
| $\delta_s$ | Servo-arm deflection: inner actuator state, observed by the wing potentiometer. |
| $\delta_c$ | Aerodynamic control-surface deflection: outer actuator state, generates the roll moment. |
| $T$ | Lumped actuator time constant of the one-pole model. |
| $T_s$ | Servo (electromechanical) lag: first pole of the two-pole actuator. |
| $T_a$ | Linkage/aerodynamic lag: second pole of the two-pole actuator. |
| $g_p$ | Potentiometer affine gain (volts per PWM count). |
| $v_0$ | Potentiometer affine offset (volts). |
| $\sigma^2_{\mathrm{pot}}$ | Potentiometer measurement-noise variance. |

**Vectors**

| | |
|---|---|
| $\vec{\cdot}$ | Notation modifier for vector-valued quantities. |
| $\vec{X}$ | The state vector of a dynamic system. |
| $\vec{X}_0$ | The state vector at the start of a system response. |
| $\dot{\vec{X}}$ | The state vector's time derivative. |
| $\vec{Y}$ | The observation model vector of a dynamic system. |
| $\vec{U}$ | The vector of control inputs. |
| $\vec{F}$ | The vector of process noise inputs. |
| $\vec{Z}$ | Vector of measurements. |
| $\vec{\varepsilon}$ | Vector of additive measurement noise. |
| $\vec{r}$ | Vector of residuals $\vec{r} = \hat{\vec{\varepsilon}} = \vec{Z} - \vec{Y}$. |
| $\vec{\phi}$ | A random function's vector-valued KLT bases. |

| Symbol | Description |
|---|---|
| $\vec{b}$ | Vector of random effects. |
| $\theta$ | The vector of dynamic system parameters. |
| $\zeta$ | The vector of variance parameters. |
| $\Theta$ | Vector containing system and variance parameters $\Theta = [\theta^T, \zeta^T]^T$. |
| $\vec{e}$ | Standard basis vector ($1$ in the position indicated by the subscript, $0$ elsewhere). |
| **Matrices** | |
| $\mathbf{I}$ | Identity matrix. |
| $\mathbf{A}$ | The sensitivity of the state derivative to the state vector. |
| $\mathbf{B}$ | The sensitivity of the state derivative to the control vector. |
| $\mathbf{G}$ | The sensitivity of the state derivative to the process noise vector. |
| $\mathbf{C}$ | The sensitivity of the observation function to the state vector. |
| $\mathbf{D}$ | The sensitivity of the observation function to the control vector. |
| $\mathbf{H}$ | The sensitivity of the observation function to the process noise vector. |
| $\mathbf{R}$ | Covariance matrix of measurement noise $\vec{\varepsilon}$. |
| $\Delta$ | Covariance matrix of $\vec{b}$. |
| $\kappa$ | Matrix-valued cross-autocovariance function. Operates on two arguments: two coordinates within a random function's domain. |
| **Operators** | |
| $p(x\|y)$ | Probability density of argument $x$ given argument $y$. Argument $x$ contributes both meaning (distribution) and value (realised value) to the operator. Argument $y$ is optional and specifies only value. |
| $E[\cdot]$ | Expectation operator. $E[x] = \int p(x)\, x\, dx$. |
| $\|\cdot\|^2_{\cdot}$ | $\|\vec{x}\|^2_{\mathbf{A}} = \vec{x}^T \mathbf{A} \vec{x}$. Squared energy norm of vector $\vec{x}$ for symmetric positive definite matrix $\mathbf{A}$. $\mathbf{A} = \mathbf{I}$ if $\mathbf{A}$ is omitted. |
| $\mathrm{tr}(\cdot)$ | Matrix trace operator. $\mathrm{tr}(A) = \sum_i A_{ii}$. |
| **Distributions** | |
| $\mathcal{N}(\cdot,\cdot)$ | 'Gaussian' or 'normal' distribution. First argument is the mean vector and the second argument is the covariance matrix. |
| $\chi^2(k)$ | Chi-squared distribution with $k$ degrees of freedom; the reference distribution for the Wald statistic. |

## 2. Introduction

Estimation of dynamic system parameters is frequently desired for assessment, validation or to provide a system model for control purposes. For system parameter estimation purposes, an aircraft will be flown with sensors to observe various quantifiable properties of the aircraft's state, and the data is analysed to estimate the system parameters that explain the motion. Aircraft system parameter identification techniques largely build on early maximum likelihood (ML) estimation by Iliff and Maine [1]. Three main categories of time-domain techniques have arisen: Equation Error Method (EEM), Output Error Method (OEM) and Filter Error Method (FEM). As the pattern in their names suggests, these techniques are different approaches to modeling the errors present in dynamic system responses. The reader is directed to relevant texts and literature for a full summary of these methods [2, 3, 4].
Critically with the existing formulations of EEM, OEM and FEM, a technique to estimate parameters from flight data affected by both sensor noise and process noise (gust input) remains an area of difficulty. EEM fits reconstructed state derivatives rather than the measured data directly, which amplifies measurement noise and biases the estimates through correlation between the noisy regressors and the equation error. OEM is susceptible to estimation errors and overconfidence when used to analyse system responses affected by process noise, because the effect of process noise is bundled in with sensor noise and assumed to be temporally uncorrelated when it is not. FEM addresses process noise by modelling it as white noise additive to the states and estimating its covariance through a Kalman filter. This is effective but carries practical restrictions: it assumes a steady-state (single

optimal Kalman gain) applied to one homogeneous response, models the disturbance as white, and does not extend straightforwardly to coloured gusts or to multiple heterogeneous operating conditions. Both [2] and [3] present FEM as a specialised tool reserved for cases where process noise genuinely cannot be neglected.

In [3] Jategaonkar emphatically states an important principle of data analysis, that *"If it is not in the data, it can not be identified."* (p. 103). This principle suggests a complementary perspective that any phenomenon leaving a discernible imprint on the data is, in principle, amenable to detection and modelling. It is this perspective that motivates the ideas explored in this study. An alternative technique is presented to model the combination of sensor noise and process noise separately, which we term the Gust Error Method (GEM). As the following sections show, GEM delivers calibrated confidence intervals under coloured gusts and, by pooling many heterogeneous manoeuvres, recovers parameters that no single response could identify. We demonstrate this on flight data from a sub-scale aircraft, jointly identifying its roll-axis inertia and aerodynamic moment derivatives.

## 3. Methodology

Two key ideas are necessary for the GEM algorithm. Firstly, a low-rank representation of the process noise sequence as latent random variables must be developed. Secondly, a mixed effects model is used to estimate the low-rank process noise object as a multidimensional random effect jointly with the system parameters and variance parameters.

### 3.1 Linear Time Invariant Model

The proposed technique is applied to linear time-invariant (LTI) systems of the form

$$\dot{\vec{X}}(t) = \mathbf{A}\vec{X}(t) + \mathbf{B}\vec{U}(t) + \mathbf{G}\vec{F}(t) \tag{1}$$

$$\vec{Y}(t) = \mathbf{C}\vec{X}(t) + \mathbf{D}\vec{U}(t) + \mathbf{H}\vec{F}(t) \tag{2}$$

$$\vec{Z}(t_k) = \vec{Y}(t_k) + \vec{\varepsilon}_k, \quad \vec{\varepsilon}_k \sim \mathscr{N}\left(\vec{0}, \mathbf{R}_j(\zeta)\right), \tag{3}$$

A sampling unit contains all of the data which belong to a particular system response whose process noise (and therefore state trajectory and modeled observations) is independent of all other sampling units. Because the likelihood structure to be developed is entirely independent between sampling units, the dynamical system in a sampling unit may take a completely different form.

Frequently, dynamic systems are observed by a heterogeneous collection of sensors that do not sample concurrently, or even at the same rate. Traditional analysis techniques address this by using filtering and interpolation schemes to estimate the sensor values on a common synchronous grid of timestamps. Filtering, interpolation and resampling introduces correlation in the regressands that violates the assumptions in the error model. The software developed to support this work supports sampling multiple different kinds of sensors on independent and ragged timestamp grids. Observations are divided into vector-valued channels, where each channel may have its own definition for the matrices $\mathbf{C}_i$, $\mathbf{D}_i$ and $\mathbf{H}_i$ in $\vec{Y}_i(t)$ in terms of the system parameters $\theta$, and its own measurement noise model $\mathbf{R}_j(\zeta)$.

## 4. Karhunen-Loève Transform

Atmospheric turbulence is a continuous random function in $\mathbb{R}^\infty$. The random function's value in effect is drawn from a multivariate normal distribution over a Hilbert space. In practice, to deal with this kind of object in a finite digital computer, the Hilbert space representation is approximated by evaluating the random function on a finite grid of coordinates within the random function's domain. The Karhunen-Loève Transform (KLT) describes how an ordered coordinate space can be found where the coordinate directions are ordered in descending order of 'importance'. The effect of this ordering is that the rank-$m$ truncation is optimal in the mean-square sense: over all rank-$m$ representations it minimises the expected $L^2$ error $E\left[\left\|\vec{F} - \vec{F}_m\right\|^2\right]$ between the process and its truncation $\vec{F}_m$, the residual being the sum of the discarded eigenvalues.

The KLT approximates a random function such as the stochastic process $\vec{F}(t)$ as

$$\vec{F}(t) = \sum_{m=1}^{\infty} b_m^* \vec{\phi}_m^*(t), \tag{4}$$

where the experimental domain coordinate $t$ may be understood as either continuous (analytical treatment) or discrete (evaluation in a finite digital computer), and the use of the symbol $b^*$ foreshadows the role that the KLT coefficients will take later in the mixed effects modeling. To obtain the KLT's optimal low dimensional approximation of $\vec{F}(t)$, the sum in (4) is truncated to a finite number of terms. The order of bases in $\vec{\phi}^*(t)$, and the distribution of the corresponding $b^*$ is the critical contribution that the KLT makes. The KLT selects the orthogonal basis $\vec{\phi}^*(t)$ as the eigenfunctions, and $b_m^*$ as independently distributed with variance equal to the eigenvalues $\lambda_m$, of the autocovariance $\kappa(t,t')$ over the random function domain, ordered by decreasing eigenvalue. If the domain coordinate $t$ is instead treated as discrete, $\vec{\phi}^*(t)$ are the eigenvectors of the autocovariance matrix of the random function over the discrete domain, and the eigenvalues' role in the distribution of $b_m^*$ is the same as in the continuous case. In the case of matrix-valued cross-autocovariance $\kappa(t,t')$ (for example for a vector valued process noise), the eigenbases $\vec{\phi}^*(t)$ comprise vector-valued eigenfunctions in the continuous case, or vector-valued elements of the eigenvectors in the discrete case.
In the mixed effects modeling to come, it is convenient to bundle the magnitude of the random coefficients into the basis functions, so that the random coefficients can be treated as having equal magnitude. The decomposition in (4) can be rewritten

$$\vec{F}(t) = \sum_{m=1}^{\infty} b_m \vec{\phi}_m(t), \tag{5}$$

$$\vec{\phi}_m(t) = \sqrt{\lambda_m/\sigma_{\mathrm{PN}}^2}\, \vec{\phi}_m^*(t), \tag{6}$$

so that the distribution of the coefficients is $\vec{b} \sim \mathcal{N}\left(\vec{0}, \Delta\right)$, where $\Delta$ is diagonal and the process noise variance scale factors out so that $\Delta = \sigma_{\mathrm{PN}}^2 \mathbf{I}$. In the flight mechanics context the process noise variance scale parameter $\sigma_{\mathrm{PN}}^2$ will not in general be known, so it must be kept accessible within the variance model for later estimation.
The eigendecomposition underlying this expansion is guaranteed by Mercer's theorem [5, 6]: a continuous, symmetric positive-semidefinite kernel (here the cross-autocovariance $\kappa(t,t')$, with $\kappa(t,t') = \kappa(t',t)^T$ ensuring a real spectrum) admits a convergent eigenexpansion. Karhunen [7] and Loève [8] (see [9]) showed that a random function can be expanded in these eigenfunctions with uncorrelated random coefficients. On a finite grid this reduces to the discrete principal-component form actually computed here.
Optimality in the process-noise $L^2$ norm (above) does not by itself guarantee that a given rank is adequate for estimation: that depends on the system, because a mode is informative only through the system's response to it. Adequacy requires that the high-order, high-frequency modes become progressively less informative, which holds whenever the product of the noise colour and the system's response to the process noise decays with frequency. A dynamic system cannot respond instantaneously to a disturbance, so its response to process noise is attenuated at high frequency. The method fails when the noise colour rises with frequency fast enough to overcome that attenuation; the modes remain informative without bound and no truncation rank suffices. Directly measuring the process noise ($\mathbf{H} \neq \mathbf{0}$) strengthens this requirement, since the system's low-pass action no longer renders the high-order modes uninformative. A band-unlimited measured disturbance admits no adequate truncation. Process noise whose power does not fall off at high frequency is, however, physically uncommon.

## 5. Mixed Effects Model

A mixed effects model is used to model the likelihood of data that is partitioned into sampling units, where each partition is affected by a particular realisation of the latent random effect $\vec{b}_i$. The latent random effect could be from any distribution, but the distribution dealt with in this work is

$\vec{b}_i \sim \mathcal{N}\left(\vec{0}, \Delta_i(\zeta)\right)$ as described in §4.. This distribution is usually treated as identical between sampling units, but the likelihood is not inherently encumbered so.
The likelihood of normally distributed $\vec{b}$ has the same form as the likelihood of normally distributed errors $\vec{\varepsilon}$, and the total likelihood is their product. Estimating $\zeta$, $\theta$ and $\vec{b}$ jointly is the focus of this work and is addressed in §6.; the remainder of this section presents the probability of the data given the model.
If the measurement model is Gaussian (which in atmospheric flight mechanics, is the prevailing assumption) conditional on the random effects, then the probability density of the data is

$$p\left(\vec{Z}_i \middle| \vec{b}_i, \Theta\right) = \frac{1}{(2\pi)^{n/2} \left|\mathbf{R}_i(\zeta)\right|^{1/2}} \exp\left[-\frac{1}{2}\left\|\vec{Z}_i - \vec{Y}(\theta, \vec{b}_i)\right\|^2_{\mathbf{R}_i(\zeta)^{-1}}\right]. \tag{7}$$

A Gaussian prior on $\vec{b}$ gives the probability density of $\vec{b}$

$$p\left(\vec{b}_i \middle| \Theta\right) = \frac{1}{(2\pi)^{q/2} \left|\Delta_i(\zeta)\right|^{1/2}} \exp\left[-\frac{1}{2}\left\|\vec{b}_i\right\|^2_{\Delta_i(\zeta)^{-1}}\right]. \tag{8}$$

The marginal likelihood is the product over sampling units of the random-effects integral given by the law of total probability

$$p\left(\vec{Z} \middle| \Theta\right) = \prod_{i=1}^{M} \int p\left(\vec{Z}_i \middle| \vec{b}_i, \Theta\right) p\left(\vec{b}_i \middle| \Theta\right) d\vec{b}_i. \tag{9}$$

### 5.1 Maximisation Tooling

Software libraries that solve mixed-effects maximum-likelihood problems exist for general use, but their assumptions are incompatible with the structural features identified above. `lme4` [10] in the R Statistical Computing Environment [11] provides analytical gradients only where the model function is closed-form or amenable to parameter-by-parameter finite-differencing; the dynamics constraint (1) requires forward integration of an ODE at every evaluation, with derivatives propagated through it, which is outside its native scope. Template Model Builder (TMB) [12] compiles the model to C++ and differentiates it with CppAD, but CppAD requires a fixed computation graph whose operations cannot depend on the parameters being optimised; the latency on some channels makes the graph parameter-dependent in a way the backend cannot interpret. TMB's state-space extension assumes additive white process noise at each timestep [13], and representing coloured noise would require augmenting the state with a shaping filter whose dimension depends on the rationality of the noise power spectral density. For irrational PSDs such as the von Kármán model no finite-state shaping filter exists; rational approximations introduce approximation error and lose the natural mode-truncation interpretation the KL expansion provides.
For these reasons, a dedicated software implementation was developed, using the MATLAB [14] mathematical computing environment.

## 6. Lindstrom and Bates' Alternating Algorithm vs Modified Laplace Approximation

Mixed effects modeling for nonlinear models is not a class of problems with any exact closed form solution.
Lindstrom and Bates' alternating algorithm (LBA) [15] estimates model parameters for general nonlinear mixed models, requiring only first derivatives of the model function. The method works by alternating between a penalised nonlinear least squares (PNLS) minimisation problem (solved iteratively), and a linearisation of the nonlinear model (first order Taylor expansion in the parameters and the random effect variables) that allows estimation of variance parameters using a linear mixed effects model (also solved iteratively). Under the first order Taylor expansion of the model in the parameters, nonlinearity of the model is neglected in informing the likelihood. Readers are referred to [15], or [16] for a detailed description of the method. Numerical testing (see §8.) LBA has shown bias in parameter estimates for linear dynamic systems, so LBA is not treated in detail here. LBA requires only the first derivatives of the observations $\vec{Y}_i(\theta, \vec{b}_i)$ with respect to $\theta$ and $\vec{b}_i$ for each system

response, so the model evaluation burden is comparable to output error methods. Because preliminary work indicated that LBA was less suitable than MLA, confidence in variance parameters was not implemented.

A second approach approximates the random-effects marginal integral by Laplace's method [17]. Applied to the nonlinear mixed-effects likelihood by Pinheiro and Bates [18], this yields a modified Laplace approximation (MLA), which in their benchmark of likelihood maximisation schemes provided a better balance of accuracy and performance than LBA. The MLA formulation results in the nested optimisation problem for the logarithm of the likelihood (9)

$$\hat{\Theta} = \arg\max_{\Theta} \ell^* \tag{10}$$

$$\ell^* = -\frac{1}{2}\sum_{i=1}^{M} n\left[1+\log(2\pi)+\log(\hat{\sigma}^2)\right] + \log|\Delta_i(\zeta)| + \log\left|g_i^{*\prime\prime}(\Theta,\vec{Z}_i,\hat{\vec{b}}_i)\right| + \log|\mathbf{R}_i(\zeta)| \tag{11}$$

$$g_i^*(\Theta,\vec{Z}_i,\vec{b}_i) = \left\|\vec{Z}_i-\vec{Y}_i(\theta,\vec{b}_i)\right\|^2_{\mathbf{R}_i(\zeta)^{-1}} + \left\|\vec{b}_i\right\|^2_{\Delta_i(\zeta)^{-1}} \tag{12}$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{M} g_i^*(\Theta,\vec{Z}_i,\hat{\vec{b}}_i), \quad \hat{\vec{b}}_i = \arg\min_{\vec{b}_i} g_i^*(\Theta,\vec{Z}_i,\vec{b}_i), \quad N=\sum_{i=1}^{M} n, \tag{13}$$

which arises by taking a second order Taylor series approximation of the observations' dependence on $\vec{b}$, so that the integrand's dependence on $\vec{b}$ becomes an integrable Gaussian expression.

## 7. MLA Simplifications for Linear Dynamic Systems

The process noise input is decomposed by (5) into a sum of independent components, and a property of linear dynamic systems is that the state trajectory response to a sum of inputs is the sum of state trajectory responses to the individual input components. Explicitly, the state trajectory is linear in $\vec{b}_i$ as

$$\vec{X}(t,\vec{b}_i) = \vec{X}(t)|_{\vec{F}(t)=\vec{0}} + \sum_{m=1}^{q} b_m \vec{X}(t)|_{\vec{X}_0=\vec{0},\vec{U}(t)=\vec{0},\vec{F}(t)=\vec{\phi}_m(t)}, \tag{14}$$

and the observation model is therefore also linear in $\vec{b}_i$ as

$$\vec{Y}(t,\vec{b}_i) = \mathbf{C}\vec{X}(t,\vec{b}_i) + \mathbf{D}\vec{U}(t) + \mathbf{H}\sum_{m=1}^{q} b_m\vec{\phi}_m(t) \tag{15}$$

$$= \vec{Y}(t)|_{\vec{F}(t)=\vec{0}} + \sum_{m=1}^{q_i} b_m \vec{Y}(t)|_{\vec{X}_0=\vec{0},\vec{U}(t)=\vec{0},\vec{F}(t)=\vec{\phi}_m(t)}. \tag{16}$$

The dynamic system parameters determine the linear system sensitivity matrices, which may be unique functions for each $i$ and indeed, the dynamic system may even differ in the number of states, control input dimensions and process noise input dimensions for each $i$. The linear system matrices may be any $\mathscr{C}^2$ continuous function of $\theta$ as $\mathbf{A}_i(\theta)$, $\mathbf{B}_i(\theta)$, $\mathbf{G}_i(\theta)$, $\mathbf{C}_i(\theta)$, $\mathbf{D}_i(\theta)$, and $\mathbf{H}_i(\theta)$. The truncation rank $q_i$ may take a unique value for each $i$. The process noise dimension may differ between sampling units, and the basis functions $\vec{\phi}_{i,m}(t)$ may also be unique to each sampling unit, and the random function bases may each derive from unique cross-autocovariances $\kappa_i(t,t')$.

The important consequences of linearity of $\vec{Y}(t,\vec{b}_i)$ with respect to $\vec{b}_i$ are twofold. First, the Hessian of $g_i^*$ with respect to $\vec{b}_i$ is

$$g_i^{*\prime\prime} = 2\frac{d\vec{Y}(t,\vec{b}_i)}{d\vec{b}_i}^T \mathbf{R}_i^{-1}\frac{d\vec{Y}(t,\vec{b}_i)}{d\vec{b}_i} - 2\frac{d^2\vec{Y}(t,\vec{b}_i)}{d\vec{b}_i^2}^T \mathbf{R}_i^{-1}\left[\vec{Z}_i(t)-\vec{Y}(t,\vec{b}_i)\right] + 2\Delta_i(\zeta)^{-1}, \tag{17}$$

but the second derivative term $\frac{d^2\vec{Y}(t,\vec{b}_i)}{d\vec{b}_i^2}$ is zero. Secondly, the term $\frac{d\vec{Y}(t,\vec{b}_i)}{d\vec{b}_i}$ is independent of $\vec{b}_i$ so that the Hessian $g_i^{*\prime\prime}(\Theta,\vec{Z}_i,\hat{\vec{b}}_i)$ is constant across the inner optimisation that finds $\hat{\vec{b}}_i$. This linearity and Hessian invariance allows a simpler form for the gradient of (11), detailed in §A.
For a linear dynamic system the modified Laplace approximation is, in fact, not an approximation. Because the observation model is linear in the random effects, the marginalisation over them in (9) is an exact Gaussian integral (the reader is referred to [18, 17] for the construction); the modified Laplace likelihood is therefore the exact marginal likelihood for the linear models considered here, and the finite Karhunen-Loève truncation is the only approximation in the likelihood. This isolates the truncation as the single controllable source of modelling error, and confirms that the variance-parameter bias discussed in §8.4 is a property of the estimator rather than of the likelihood approximation.

### 7.1 Mode Information Ratio

Using either MLA or LBA, a method to determine what rank to truncate the process noise at is needed. The fit could be rerun with higher truncation rank, but this defeats the computational benefit of truncating. Each element of $\vec{b}$ can be assessed using the 'mode information ratio' (MIR) $\eta$, which compares the weight of information from the data with the weight of information from the prior, in $g_i^{*\prime\prime}$. Concretely, MIR is

$$\eta_{i,m} = \frac{\left[\frac{d\vec{Y}_i}{d\vec{b}_i}^T \mathbf{R}_i^{-1} \frac{d\vec{Y}_i}{d\vec{b}_i}\right]_{m,m}}{\left[\Delta_i^{-1}\right]_{m,m}}, \tag{18}$$

so that $\eta_{i,m} \gg 1$ indicates the mode is primarily informed by the data and $\eta_{i,m} \ll 1$ indicates the mode is primarily informed by the prior. The quantities to compute MIR are readily available from the optimisation routine. Because the KLT orders the modes in order of descending importance, the least significant mode will usually be the highest order retained mode, and the most significant discarded mode will usually be the least order truncated mode. This ordering is a heuristic, not a theorem: a mode's data information is the system's response to that mode's basis function, so $\eta$ need not decrease monotonically with mode index. If the truncation falls where the system's frequency response to the process noise has a trough, $\eta$ may dip well below unity and rise again at higher order; the practical safeguard is to inspect the $\eta$ profile, or its maximum over retained modes, rather than the last retained mode alone.

## 8. Simulation Studies

Simulation studies have been performed to validate parameter recovery performance under a range of conditions, using synthetic datasets generated according to a known truth.

### 8.1 Aircraft Roll Mode Model

An aircraft roll mode model was used to create system responses, based on the roll mode simulation study in [4]. The roll mode system is a single state dynamic system, with dimensional parameters $L_p$ and $L_\delta$, which are normally the parameters of primary interest for estimation. The model also contains the variances $\sigma^2_{\text{gyro}}$ and $\sigma^2_{\text{PN}}$. The state derivative was modeled as

$$\dot{p}(t) = L_p p(t) + L_\delta \delta_a(t) + \dot{p}_{\text{gust}}(t), \tag{19}$$

where the process noise $\dot{p}_{\text{gust}}(t)$ has been explicitly modeled as stochastic angular acceleration with the colour of von Kármán roll angular rate turbulence using the power spectral density in [19], with variance scale parameter $\sigma^2_{\text{PN}}$. The roll gust model assumed an altitude of $1000\,\text{ft}$, an airspeed of $50\,\text{ft/s}$, and a wingspan of $6\,\text{ft}$. The measurement model is

$$\dot{p}_{\text{measured}}(t_k) = p(t_k) + \vec{\varepsilon}(t_k), \tag{20}$$

with $\vec{\varepsilon}(t_k) \sim \mathcal{N}\left(\vec{0}, \sigma^2_{\text{gyro}}(\zeta)\right)$.

The synthetic experiment assumed that the observations were uniformly spaced at $100\,\mathrm{Hz}$, with an applied aileron doublet input during a 10 second test segment. Fitting was conducted with the true parameter values as starting estimates, for speed. Convergence behaviour was generally well-conditioned from informed starts, though convergence from arbitrary initial estimates is not guaranteed and a systematic basin-of-convergence study is left to future work.

## 8.2 Rank Truncation Sweep

To characterise the adequacy of a chosen rank truncation and the MIR as the discriminator of useful modes and irrelevant modes, a simulation study was performed sweeping over values of the rank truncation. The roll mode experiment was simulated 1024 times and fit using a range of fitting techniques to estimate the parameter vector $\Theta = \{L_p,\, L_\delta,\, \sigma^2_{\mathrm{PN}},\, \sigma^2_{\mathrm{gyro}}\}$. Each of the $N = 1024$ datasets comprised a single sampling unit (one system response), and the initial roll rate $p_0$ was held at its true value rather than estimated; the per-unit initial state is estimated only in the multi-response sweeps of § 8.3 and §8.4. The FEM baseline fitting technique was applied to both white noise and coloured von Kármán roll angular rate turbulence, which was the only place in the truncation rank sweep where the dataset generation differed between fitting techniques. The FEM technique presented in [4] requires that the measurement noise be provided, so $\sigma^2_{\mathrm{gyro}}$ was fixed at the true value for the FEM fits and the MLA ($\sigma^2_{\mathrm{gyro}}$ fixed) fit, but not in any other case. The system parameters' fit quality is shown in Figure 1, the variance parameters fit quality is shown in Figure 2 and the MIR and timing is shown in Figure 3. Results at rank truncation of 128 are shown in Table 1. In all cases, the optimiser completed without errors and reported that the solver halted having found a local optimum to within tolerance.

| | $L_p$ (truth $-2$) | | | $L_\delta$ (truth $-10$) | | |
|---|---|---|---|---|---|---|
| Fitter | Bias | RMSE | Cov | Bias | RMSE | Cov |
| FEM, white truth | $+0.36$ | $0.49$ | $79\%$ | $+1.11$ | $2.02$ | $84\%$ |
| FEM, vK truth | $+0.16$ | $0.29$ | $55\%$ | $+0.53$ | $1.11$ | $61\%$ |
| MLA ($\sigma^2_{\mathrm{gyro}}$ fixed) | $-0.04$ | $0.25$ | $94\%$ | $-0.11$ | $0.88$ | $95\%$ |
| MLA ($\sigma^2_{\mathrm{gyro}}$ fitted) | $-0.04$ | $0.25$ | $93\%$ | $-0.09$ | $0.89$ | $94\%$ |
| LBA ($\sigma^2_{\mathrm{gyro}}$ fitted) | $+0.22$ | $0.31$ | $84\%$ | $+0.55$ | $1.01$ | $89\%$ |
| MLA, $b_w$-misspec | $-0.24$ | $0.37$ | $95\%$ | $-0.43$ | $1.07$ | $93\%$ |
| LBA, $b_w$-misspec | $+0.23$ | $0.33$ | $94\%$ | $+0.38$ | $1.02$ | $93\%$ |

Table 1 – Parameter-recovery results at the adequate rank truncation ($q = 128$, $N = 1024$ datasets). Bias, RMSE and 95% confidence-interval coverage for the system parameters $L_p$ (truth $-2$) and $L_\delta$ (truth $-10$). FEM is shown on both correctly specified (white) and misspecified (von Kármán) process noise; MLA and LBA fit the correct von Kármán colour, with two cases ($b_w$-misspec) fitting at twice the true gust wingspan.

Smaller truncation rank results (not shown) exhibited very poor estimation performance, particularly as the rank truncation was swept through eigenvalues whose eigenvectors exhibited sinusoid-like behaviour at timescales comparable to the timescales of the dynamic system's behaviour. This is attributable to the process noise being allowed to asymmetrically explain parts of the response that were close to directly confounded with the unforced response. At those low truncation ranks where poor estimation was observed, $\eta \gg 1$ in all cases.

In Figures 1 and 2, the estimates, RMSE and coverage all reach a steady state from a rank truncation of 128 and higher. Figure 3 shows that $\eta$ has descended below 1 at 96 modes retained, where all parameters except $\hat{\sigma}^2_{\mathrm{gyro}}$ are steady at that truncation rank and higher, with only $\hat{\sigma}^2_{\mathrm{gyro}}$ requiring more retained modes. The behaviour is consistent with the proposed metric of $\eta \ll 1$ for the smallest retained mode as suitable for assessing adequacy of the rank truncation for estimation performance. Ensuring that the last $\eta \ll 1$ was used to determine suitable rank truncation for all fits in the rest of this work.

The FEM estimates are independent of truncation rank, so are shown as a level within the figures. The

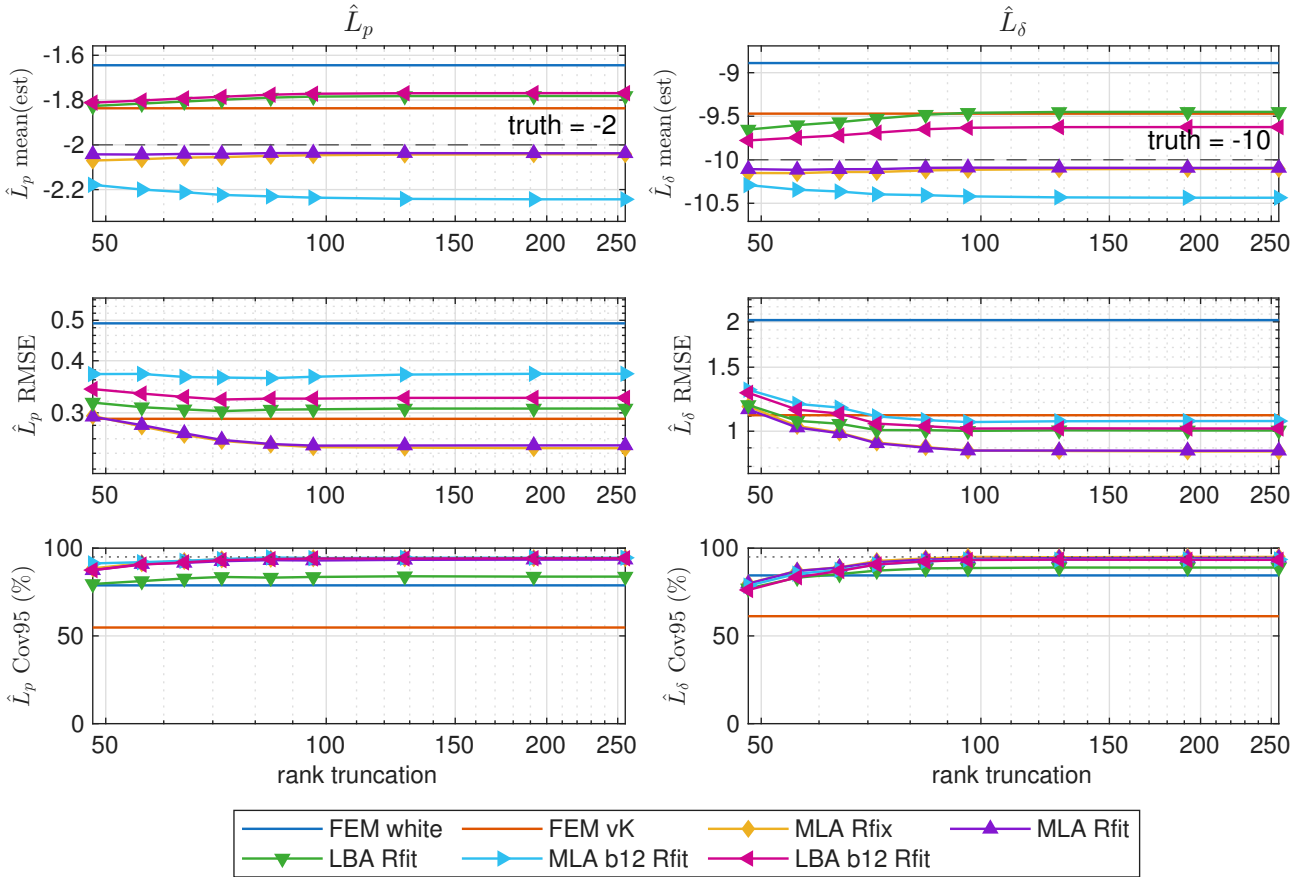



Figure 1 – Simulation study showing system parameter fit quality vs truncation rank.

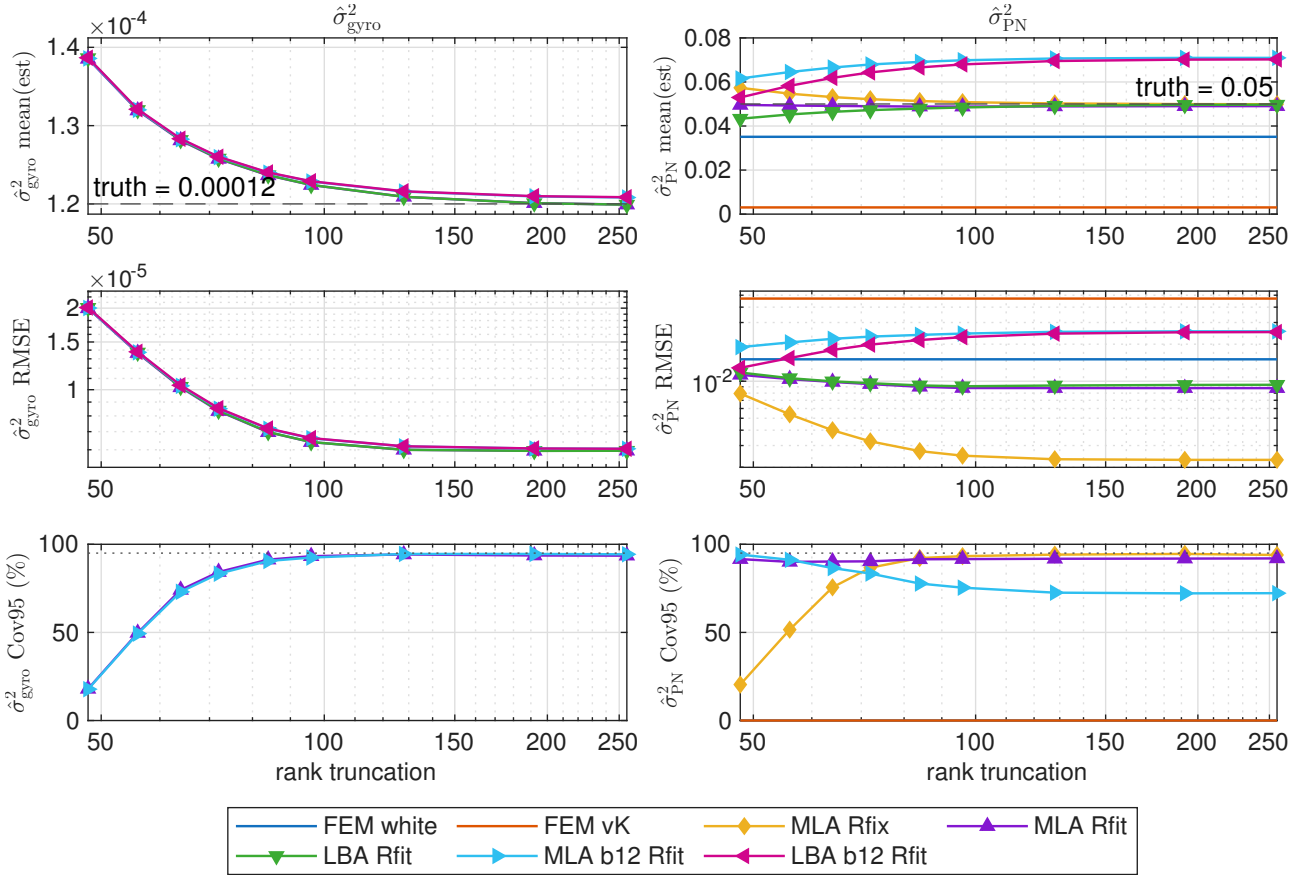



Figure 2 – Simulation study showing variance parameter fit quality vs truncation rank.

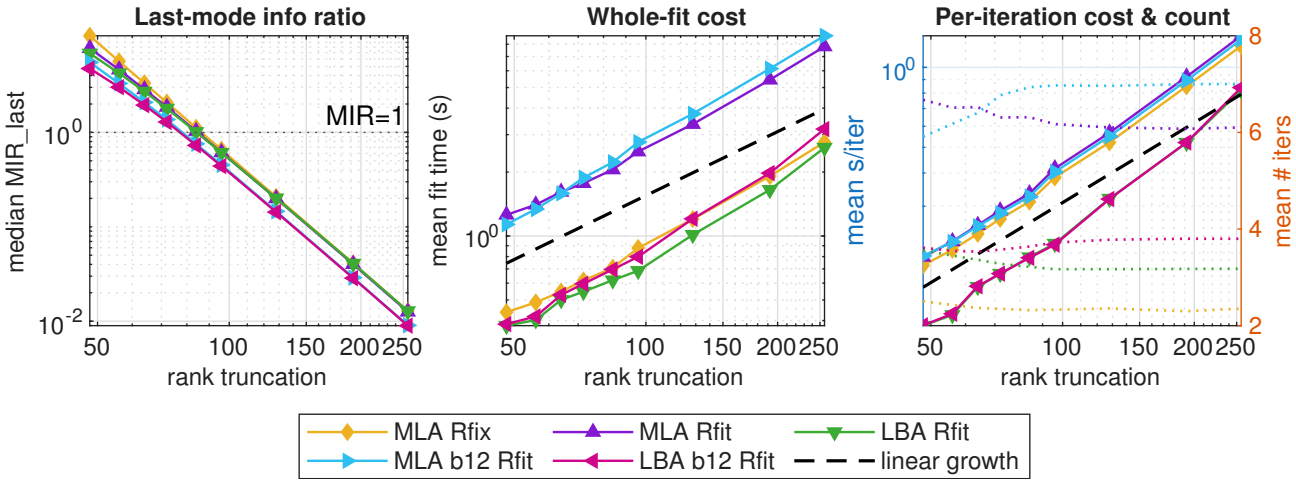



Figure 3 – MIR and timing information, for the rank truncation sweep.

FEM baseline here is supplied the true measurement-noise covariance and fits a single response in which the process noise is modelled as white. Under these conditions its system-parameter estimates are comparable to MLA's but with higher RMSE, and its 95% confidence intervals are optimistic: coverage falls below nominal because the coloured gust is absent from the error model. This is the expected behaviour of any estimator that treats a temporally correlated disturbance as white, not a deficiency peculiar to FEM; MLA attains nominal coverage because it represents the gust structure explicitly. The residual bias of the FEM estimates in the correctly-specified (white-noise) case is consistent with finite-sample bias under the weakly-informative single short response considered here.

We do not claim a decisive single-response advantage. Coloured process noise could in principle be brought into FEM by augmenting the state with a shaping filter matched to the gust spectrum, and for a rational spectrum (as in the roll-rate case here) that filter is exact. The standard treatment [3] nonetheless advises against extending the model to track noise colour, holding predictive capability to be primary; the FEM baseline used here reflects that recommended practice and the published formulation of [4]. For the general aircraft case, where the gust spectrum is irrational and no finite shaping filter exists, GEM instead represents the disturbance natively through the Karhunen-Loève expansion.

More fundamentally, estimation from a single response conditions all inference on one realisation of the stochastic disturbance. Without replication the disturbance cannot be averaged out, so its projection onto the parameter directions enters as bias that lengthening the response only attenuates. Replication across independent responses is what removes it: the sampling-unit study of §8.4 shows the system-parameter bias decaying as responses accumulate, and the flight application of §9.2, with 18 independent responses, yields estimates that are consistent across two datasets. This replication-based consistency, rather than any single-response margin over FEM, is the substantive advantage of treating the process noise as a random effect.

As truncation rank increases $\hat{\sigma}^2_{\text{gyro}}$ reduces, as higher order truncated modes allow variance in the data to be explained as the effect of process noise. Coverage of $\hat{\sigma}^2_{\text{gyro}}$ is poor and its estimate upward-biased at low truncation rank; the mechanism behind this variance mis-attribution is examined in § 8.4. The LBA method provides comparable estimates in the rank truncation sweep, with similar dependence on rank truncation as the MLA estimates. LBA exhibits inferior bias behaviour and RMSE as compared to MLA for the correctly specified models. The MLA and LBA estimates with misspecified process noise colour show deterioration in the estimates' bias and RMSE. Confidence interval coverage does not significantly deteriorate for the parameters of primary interest, though coverage deterioration due to bias may be masked by the wide estimate variance due to modest information content in the synthetic dataset considered here. Process noise intensity estimates obtain significant bias when the process noise model is misspecified, though this is not surprising.

The computational cost of MLA is much larger than that of LBA, which is itself much larger than that

of FEM. The cost of MLA grows linearly with the number of parameters, system response duration and number of records, and slightly faster than linearly with the rank truncation: the per-iteration inner solve over the random effects, together with the associated log-determinant and trace terms, carries an $O(q^3)$ cost from factorising the $q \times q$ random-effects Hessian, which adds to the otherwise linear scaling and dominates at high rank. One caveat is that as a system response grows longer, the required truncation rank will grow, as there are more resolvable details of the stochastic input in a longer system response. The computational cost's linear growth with the number of responses and faster than linear growth with the response duration incentivises an approach to experiment design with many information-dense short responses, whereas FEM's applicability mainly to single system responses motivates creating a single response containing all of the information required for the model parameters to be estimated. Many short responses also de-risk the experiment operationally, because a single long manoeuvre must hold the linear regime over its full duration (tolerating airspeed drift and altitude loss) and invites conditioning the result on one favourably-behaved response, whereas with short manoeuvres additional information is simply a matter of repeating the manoeuvre. The simulation study was conducted on a desktop PC with an AMD Ryzen 9 7950X3D CPU and $64\,\mathrm{GB}$ of system RAM, using MATLAB parallel execution (via `parfeval()`) across the independent replicates; each individual fit ran on a single worker, so the reported per-fit timings reflect single-threaded performance.

## 8.3 Process Noise Intensity Sweep

To investigate the effect of process noise intensity, a simulation study of N=1024 datasets each containing 8 system responses was conducted. The initial state for each response was estimated, so that the whole vector of estimated parameters was $\Theta = \{L_p,\ L_\delta,\ \sigma^2_{\mathrm{PN}},\ \sigma^2_{\mathrm{gyro}}\} \cup \{p_{0,i}\}_{i=1}^{8}$. The mean estimate, RMSE and coverage are shown in Figure 4.

The LBA estimates' quality shows significant deterioration with increasing process noise intensity, with increasing bias and deteriorating coverage for the system parameters. In contrast, the MLA estimates do not show significant bias or undercoverage for the system parameters. In both cases, RMSE increases (except for relative $\sigma^2_{\mathrm{PN}}$ RMSE) as the process noise intensity increases. The system parameters' RMSE increases with $\sigma^2_{\mathrm{PN}}$ because a process noise sequence that mimics the effect of any of the other parameters is proportionately more likely. The measurement noise RMSE increases because it confounds more directly with higher order process noise modes. Conversely, because a larger process noise intensity has a larger effect on the data it is more informed by the relevant modes and there are more relevant modes, reducing its RMSE.

MLA coverage for $\hat{\sigma}^2_{\mathrm{gyro}}$ deteriorates substantially as process noise intensity increases. Measurement noise is not usually a parameter of primary interest, so this is not critical for aircraft parameter estimation.

## 8.4 Sampling Units Sweep

To compare MLA and LBA performance with larger datasets, a simulation study across the number of system responses was conducted. The estimated parameters in this case were $\Theta = \{L_p,\ L_\delta,\ \sigma^2_{\mathrm{PN}},\ \sigma^2_{\mathrm{gyro}}\} \cup \{p_{0,i}\}_{i=1}^{M}$, and N=1024 simulated datasets and fits were conducted for each case. The process noise was set to $\sigma^2_{\mathrm{PN}} = 0.1\,\mathrm{rad}^2/\mathrm{s}^4$, and parameter recovery is shown in Figure 5.

As more sampling units are included, the inadequacy of LBA becomes apparent. The system parameter estimates remain biased and coverage declines to zero, as the parameter variance about the biased mean is increasingly resolved as distinct from the true value. The process noise intensity sweep of §8.3 shows that LBA accumulates bias as the process noise increases, and the sampling unit sweep in this section shows that this bias does not reduce when a larger dataset is available. The conclusion for linear dynamic system parameter identification therefore is that LBA is unsuitable due to both bias and undercoverage, even though it alleviates the need to compute mixed second derivatives of the modeled observations.

MLA system parameter estimate performance behaves desirably as the dataset grows. Estimate bias reduces, RMSE improves and coverage remains steady. Variance parameters show less desirable behaviour, with bias in the mean and undercoverage as the dataset grows. Despite the coverage decline of $\hat{\sigma}^2_{\mathrm{gyro}}$ for fits with more sampling units, the RMSE decreases while bias remains steady. This

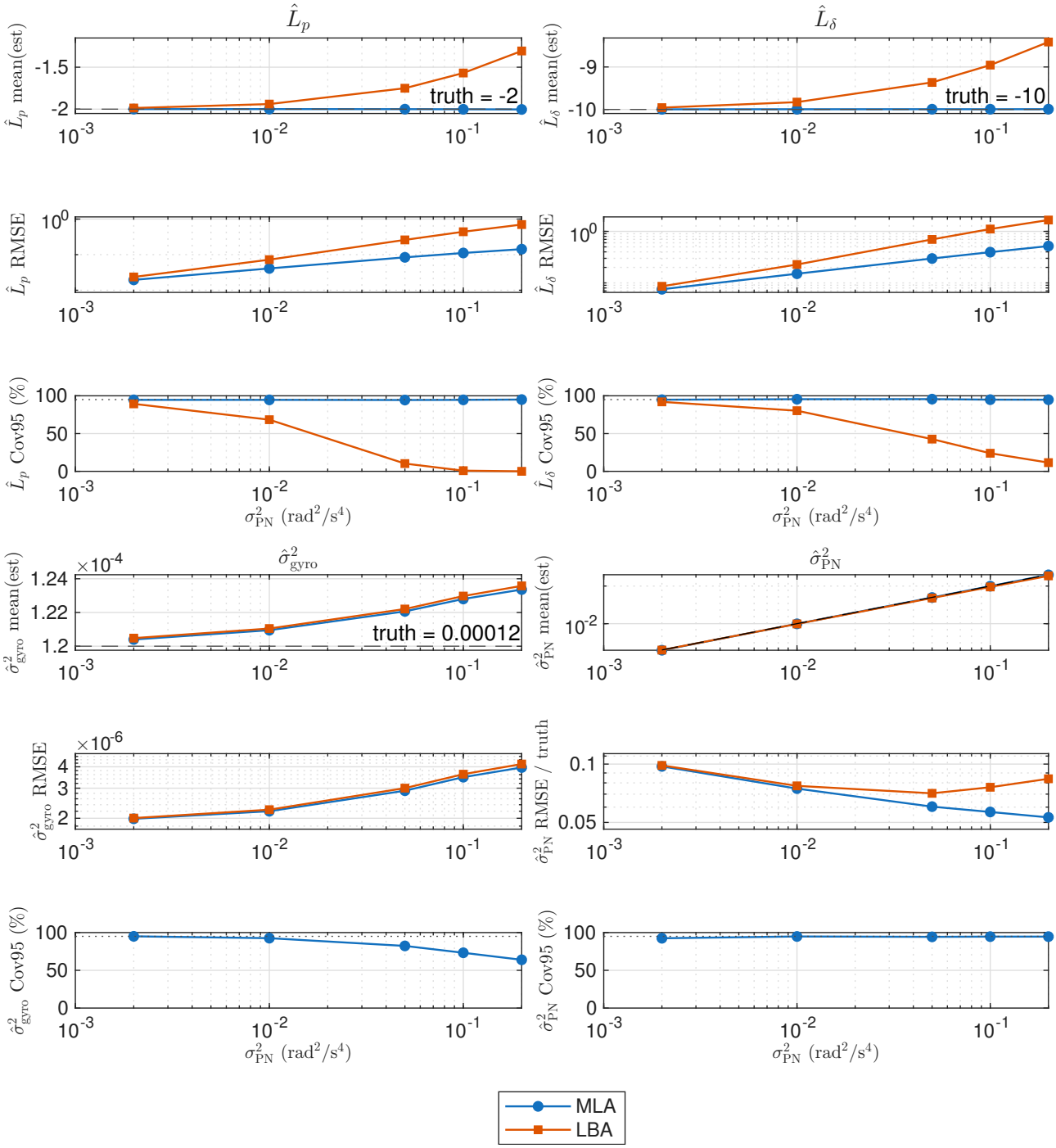



Figure 4 – Simulation study showing parameter fit quality vs process noise intensity: system parameters (top) and variance parameters (bottom).

indicates that the estimation of $\hat{\sigma}^2_{\mathrm{gyro}}$ is becoming more tightly distributed around the biased value, but the biased estimate is not departing further from the true $\sigma^2_{\mathrm{gyro}}$ as the dataset grows. Coverage for $\hat{\sigma}^2_{\mathrm{PN}}$ remains nominal, with similar behaviour for the estimate bias and RMSE.

The persistent upward bias of $\hat{\sigma}^2_{\mathrm{gyro}}$ is not a finite-sample effect that more responses remove, nor an artefact of the likelihood approximation (exact for a linear system, §7.); it is a property of the maximum-likelihood variance estimate. It arises from a likelihood-geometry asymmetry. Setting a weakly-informed high-order KL mode to zero returns its contribution to the residuals, but that contribution is an oscillatory, high-frequency signal: being sign-balanced, its first-order effect on the Gaussian likelihood cancels and only a small second-order penalty remains, which a slight inflation of the measurement variance readily absorbs. Moving the high-order modes away from zero, by contrast,

incurs the large cost of their strong regularisation. The estimator therefore prefers to estimate these modes as essentially zero and to charge the variance they would have carried to the measurement-noise model. The effect is not removed by the rank truncation; it is expected instead to attenuate with longer individual responses, where more residuals must absorb each equally weakly-informed mode at a higher likelihood cost, though this is not characterised here. Equivalently, the rank-truncated estimator slightly under-attributes variance to the random effect and over-attributes it to the measurement noise: the process-noise intensity $\hat{\sigma}^2_{\mathrm{PN}}$ shows a slight downward bias across this sweep, of the character carried by maximum-likelihood variance components (and addressed by REML), whereas the upward bias of $\hat{\sigma}^2_{\mathrm{gyro}}$ is amplified by the truncation and would persist under REML.

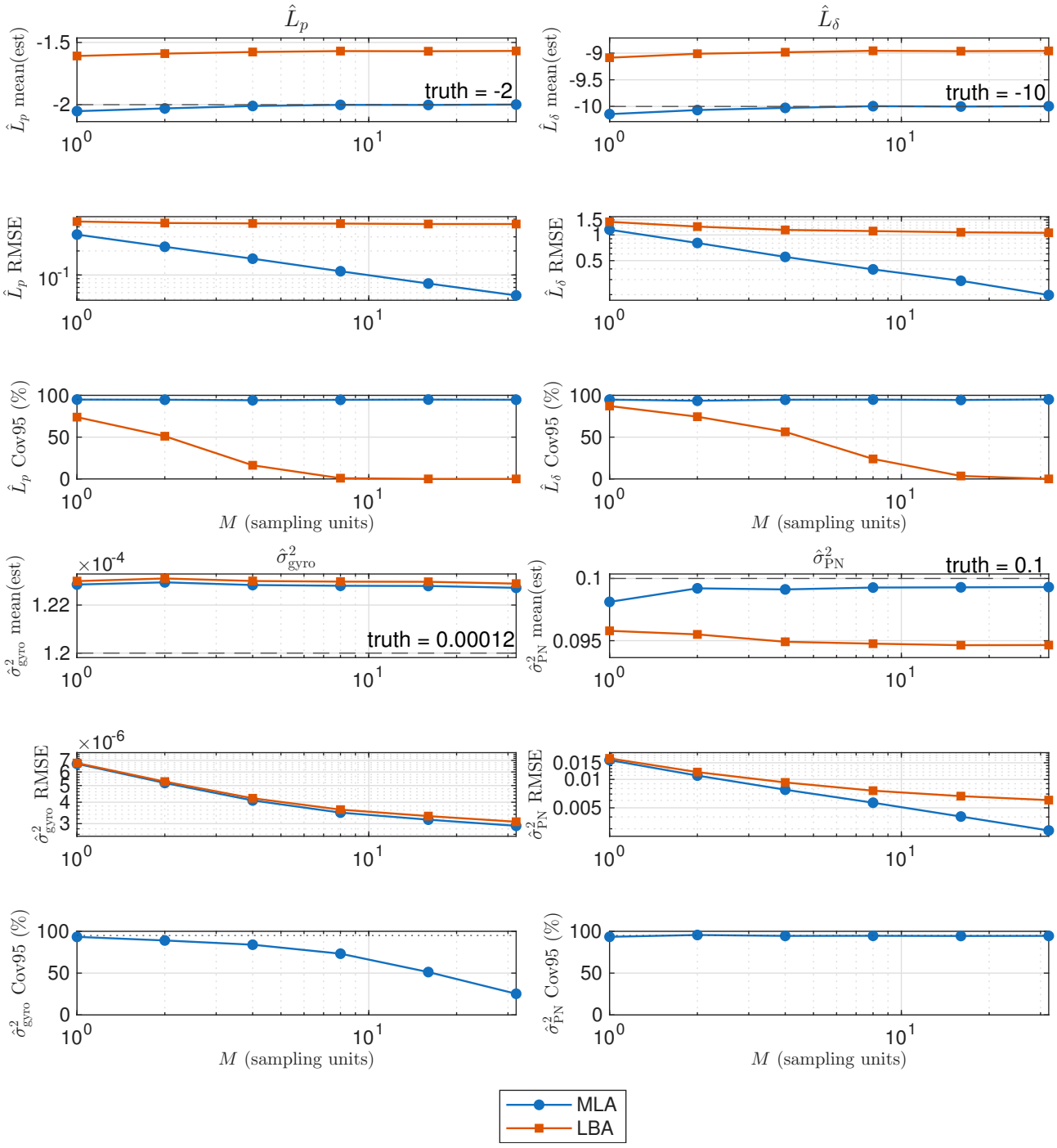



Figure 5 – Simulation study showing parameter fit quality vs number of sampling units: system parameters (top) and variance parameters (bottom).

### 8.5 Simulation Study Conclusion

The simulation studies presented show that within GEM, the MLA algorithm provides superior estimates to LBA, that are consistent with maximising the log likelihood 11. Similarly to the conclusion drawn by [18], we feel that the improved estimation performance of MLA relative to LBA justifies its increased computational cost in all applications to real data, though LBA may be useful to provide initial parameter estimates to warm-start MLA.

## 9. Application

### 9.1 Test Platform and Flight Testing

A test platform based on the ZOHD Nano-Talon EVO has been developed for flight dynamics research. The platform uses ArduPilot [20] software on a MATEK H743WING computer to provide flight stabilisation, navigation, autonomy and logging. A LUA script to facilitate open-loop command injection during automated waypoint flight has been developed, enabling large numbers of doublet inputs to be conducted.

**Sensing** The flight computer incorporates an inertial measurement unit (IMU). The aircraft's servos have been modified so that the internal position sensing potentiometer (pot) can be logged. The servo's internal potentiometer has substantial noise and experiences large deviations in voltage when the servo is under load, likely due to changes in the pot supply voltage as described in [21]. A pitot tube measures the airspeed, though in the present study this data is only used to set the mean flight speed during an aileron response. The pitot tube was calibrated prior to data collection for this study. The aircraft is also fitted with a GPS, magnetometer and three aerodynamic vanes, though these are not used in the present study.

**Experiment** Three flights were conducted in varying configurations. The first flight was a baseline configuration with no ballast and a 6x4E APC prop. The second flight used a reverse direction propeller (6x4EP APC prop), and no ballast. The third flight used the reverse direction propeller, and included $20\,\mathrm{g}$ of lead ballast $70\,\mathrm{mm}$ inboard of each wing tip. In each flight, a waypoint mission was conducted with positive and negative doublets of the aileron servo command at three airspeed setpoints. The entire waypoint mission was completed twice in each flight, to form the 'Sweep 1' and 'Sweep 2' datasets. Similar doublets in each flight on the elevator, rudder and throttle controls were conducted, but are not used in this analysis. A tight time window of data was selected from each aileron doublet, to minimise the effect of velocity variation and slower Dutch roll or other dynamic behaviours.

### 9.2 Model

**Configurations** Three flights share one airframe; the fit is joint across all aileron-doublet segments, with config entering only through $I_T$ and the propeller input sign:

$$I_T = I_{b,0} + \beta\, I_b, \qquad \begin{array}{ll} \text{flt 1:} & \beta = 0, \quad \text{APC 6x4E } (+) \\ \text{flt 2:} & \beta = 0, \quad \text{APC 6x4EP } (-,\ \text{reversed}) \\ \text{flt 3:} & \beta = 1, \quad \text{APC 6x4EP } (-,\ \text{reversed}) \end{array} \tag{21}$$

**Dynamics** The quantity of interest is the body roll rate $p$, which the rate gyro observes directly. Writing $\delta_{\text{eff}}$ for the effective control-surface deflection (set by the actuator model below) and $p_w$ for the rolling gust (a rotational velocity process noise, represented by the Karhunen-Loève expansion of §4. ), the roll-moment balance is

$$\dot{p} = \underbrace{L_p\,(p - p_w)}_{\text{aero damping}} + \underbrace{L_\delta\,(\delta_{\text{eff}} - \delta_0)}_{\text{aileron}} + \underbrace{\frac{k_Q}{I_T}\,\omega_p^2}_{\text{prop reaction torque}}, \qquad \vec{Z}_{\text{gyro}} = p + \vec{\varepsilon},\ \ \vec{\varepsilon} \sim \mathcal{N}\left(0, \sigma_{\text{gyro}}^2\right), \tag{22}$$

with the dimensional roll derivatives expressed through the fitted non-dimensional coefficients and the per-response airspeed $V$,

$$L_p = \frac{\rho V S b_w^2 C_{\ell_p}}{4 I_T}, \qquad L_\delta = \frac{\rho V^2 S b_w C_{\ell_{\delta_a}}}{2 I_T}. \tag{23}$$

The aero damping acts on the airframe rate relative to the gust, $p - p_w$. The rolling gust represents the roll-axis rotational rate produced by the spanwise gradient of the vertical gust velocity, with the colour of the von Kármán rolling-rate spectrum [19]. Over the short aileron-doublet window it is the dominant roll-axis disturbance, and the translational ($\beta$, $\alpha$) gust components are not modelled. The single-state roll model is a deliberate simplification that neglects the airframe's lateral-directional (Dutch-roll) coupling. The tight aileron-doublet window was chosen partly to limit it.

**Actuator models** The commanded aileron $\delta_a$ reaches the surface through the servo and linkage. We compare three nested representations of $\delta_{\text{eff}}$, all sharing the input latency $\tau_a$ on the command $\delta_a(t-\tau_a)$:

- *No-lag:* $\delta_{\text{eff}}(t) = \delta_a(t-\tau_a)$, where the command drives the surface directly (gyro channel only).
- *One-pole:* a single first-order actuator state $\delta_s$,

$$T\,\dot{\delta}_s = \delta_a(t-\tau_a) - \delta_s, \qquad \delta_{\text{eff}} = \delta_s. \tag{24}$$

- *Two-pole:* a servo stage feeding a linkage/aerodynamic stage,

$$T_s\,\dot{\delta}_s = \delta_a(t-\tau_a) - \delta_s, \qquad T_a\,\dot{\delta}_c = \delta_s - \delta_c, \qquad \delta_{\text{eff}} = \delta_c. \tag{25}$$

The one- and two-pole models add a second observation channel: the internal servo potentiometer (pot) reads the servo deflection through an unknown affine map, at its own observation latency $\tau_p$,

$$\vec{Z}_{\text{pot}} = g_p\,\delta_s + v_0 + \varepsilon_{\text{pot}}, \qquad \varepsilon_{\text{pot}} \sim \mathcal{N}\left(0,\, \sigma_{\text{pot}}^2\right). \tag{26}$$

The affine map neglects the load-dependent supply variation noted above, which is treated here as part of the pot measurement noise. The models are strictly nested: no-lag is the one-pole limit $T \to 0$, and the one-pole model is the two-pole limit $T_a \to 0$ (the servo state then both drives the moment and is observed by the pot). The one- vs two-pole comparison is therefore a nested test on a *shared* (gyro + pot) observation set, with the no-lag fit serving as a latency-only baseline.

**Propeller-sign data convention** The reversed propeller (flights 2 and 3) is handled in the data by negating both propeller input entries, $\omega_p \to -\omega_p$ and $\omega_p^2 \to -\omega_p^2$ (an explicit signed square, not a re-squaring), leaving $L_p, L_\delta, k_Q$ config-independent.

**Pre-Processing** The data were not filtered, interpolated, or resampled to a uniform grid. Such pre-processing is common in flight-test parameter estimation, but it alters the measurement-noise statistics and breaks the independence and covariance assumptions the error model relies on. Because the present method models the measurement error explicitly and operates directly on the flight computer's reported, non-uniformly-spaced timestamps, this pre-processing is unnecessary and is deliberately avoided. To examine the effect of dataset size, the gyro data were decimated (by masking to every $m$-th sample) to nominal rates of 150, 300, 500 and $1000\,\text{Hz}$.

**Parameter inventory**

- Global $\theta$: aerodynamics $\{C_{\ell_p},\ C_{\ell_{\delta_a}},\ k_Q\}$, inertia $I_{b,0}$, input latency $\tau_a$; plus the actuator set (none, $\{T\}$, or $\{T_s, T_a\}$) and, for the pot models, $\{g_p, v_0, \tau_p\}$.
- Local, per response $i$: $V_i$ (fixed $=$ segment-mean measured airspeed), the initial roll rate $p_{0,i}$; for the two-pole model the free aero-surface initial state $\delta_{c0,i}$ (the servo initial state is held at the command, because the preceding $100\,\mathrm{ms}$ of the flight held the aileron command fixed).
- Known/fixed: $I_b,\ \delta_0,\ \rho,\ S,\ b_w,\ \beta_i$.
- Variance $\zeta$: gyro $\sigma^2_{\mathrm{gyro}}$, gust intensity $\sigma^2_{\mathrm{PN}}$ (§4.); and, for the pot models, $\sigma^2_{\mathrm{pot}}$.

## 9.3 Results and Discussion

The proposed models were fitted to the data using MLA, and the primary interest parameters estimated for the two-pole model are shown in Table 2. Table 3 shows the likelihoods for the one-pole and two-pole models, ΔAIC and ΔBIC to discriminate them, as well as a Wald test for consistency in the primary-interest parameters between the two-pole models identified for Sweep 1 and Sweep 2 datasets. Figure 6 shows the process noise realisations from the three fitted models in the Sweep 1 fit using $1000\,\mathrm{Hz}$ gyro data. Figure 7 shows the state trajectory and comparison with measurements, for a selection of positive doublet responses from the two-pole Sweep 1 fit. Figure 8 shows correlation information for the parameter estimates' variance.

| Parameter | Sampling rate (Hz) | | | |
|---|---|---|---|---|
| | 150 | 300 | 500 | 1000 |
| *Sweep 1* | | | | |
| $C_{\ell_p}$ | $-0.500\,(0.103)$ | $-0.431\,(0.063)$ | $-0.441\,(0.064)$ | $-0.447\,(0.063)$ |
| $C_{\ell_{\delta_a}}$ (/PWM, $\times 10^{-4}$) | $-3.24\,(0.66)$ | $-2.82\,(0.40)$ | $-2.89\,(0.41)$ | $-2.92\,(0.40)$ |
| $k_Q$ (N m/(rad/s)$^2$) | $-0.0246\,(0.0147)$ | $-0.0208\,(0.0106)$ | $-0.0211\,(0.0109)$ | $-0.0216\,(0.0109)$ |
| $I_{b,0}$ (kg m$^2$) | $0.0079\,(0.0021)$ | $0.0064\,(0.0012)$ | $0.0068\,(0.0013)$ | $0.0069\,(0.0012)$ |
| $\tau_a$ (ms) | $-30.1\,(2.9)$ | $-28.1\,(1.8)$ | $-28.8\,(1.7)$ | $-29.0\,(1.5)$ |
| $T_s$ (ms) | $16.7\,(2.5)$ | $16.4\,(2.4)$ | $16.4\,(2.4)$ | $16.3\,(2.3)$ |
| $T_a$ (ms) | $10.3\,(5.6)$ | $13.8\,(4.4)$ | $12.5\,(3.9)$ | $12.5\,(3.6)$ |
| $\tau_p$ (ms) | $11.6\,(3.8)$ | $13.8\,(2.7)$ | $13.1\,(2.6)$ | $13.0\,(2.5)$ |
| $\sigma^2_{\mathrm{gyro}}$ (rad$^2$/s$^2$) | $0.01319^{+0.00098}_{-0.00092}$ | $0.01291^{+0.00059}_{-0.00057}$ | $0.01264^{+0.00046}_{-0.00045}$ | $0.01228^{+0.00031}_{-0.00030}$ |
| $\sigma^2_{\mathrm{PN}}$ (rad$^2$/s$^2$) | $0.220^{+0.062}_{-0.049}$ | $0.163^{+0.039}_{-0.031}$ | $0.165^{+0.037}_{-0.030}$ | $0.164^{+0.035}_{-0.029}$ |
| *Sweep 2* | | | | |
| $C_{\ell_p}$ | $-0.374\,(0.047)$ | $-0.384\,(0.049)$ | $-0.393\,(0.050)$ | $-0.407\,(0.052)$ |
| $C_{\ell_{\delta_a}}$ (/PWM, $\times 10^{-4}$) | $-2.43\,(0.30)$ | $-2.50\,(0.31)$ | $-2.56\,(0.32)$ | $-2.65\,(0.33)$ |
| $k_Q$ (N m/(rad/s)$^2$) | $-0.0201\,(0.0080)$ | $-0.0210\,(0.0085)$ | $-0.0213\,(0.0088)$ | $-0.0223\,(0.0093)$ |
| $I_{b,0}$ (kg m$^2$) | $0.0050\,(0.0010)$ | $0.0054\,(0.0010)$ | $0.0057\,(0.0010)$ | $0.0060\,(0.0010)$ |
| $\tau_a$ (ms) | $-27.7\,(1.9)$ | $-28.5\,(1.8)$ | $-29.0\,(1.7)$ | $-29.3\,(1.5)$ |
| $T_s$ (ms) | $16.0\,(2.6)$ | $16.1\,(2.7)$ | $16.1\,(2.7)$ | $16.1\,(2.6)$ |
| $T_a$ (ms) | $16.3\,(5.5)$ | $14.4\,(5.1)$ | $13.6\,(4.5)$ | $13.2\,(4.0)$ |
| $\tau_p$ (ms) | $13.7\,(2.8)$ | $12.9\,(2.9)$ | $12.4\,(2.8)$ | $12.1\,(2.6)$ |
| $\sigma^2_{\mathrm{gyro}}$ (rad$^2$/s$^2$) | $0.01512^{+0.00108}_{-0.00101}$ | $0.01380^{+0.00063}_{-0.00061}$ | $0.01349^{+0.00050}_{-0.00048}$ | $0.01302^{+0.00033}_{-0.00032}$ |
| $\sigma^2_{\mathrm{PN}}$ (rad$^2$/s$^2$) | $0.130^{+0.035}_{-0.028}$ | $0.143^{+0.036}_{-0.029}$ | $0.149^{+0.035}_{-0.029}$ | $0.155^{+0.034}_{-0.028}$ |

Table 2 – Two-pole actuator-model parameter estimates. System parameters: 95% half-widths ($1.96\,\mathrm{SE}$, parenthesised). Variance parameters ($\sigma^2_{\mathrm{gyro}}, \sigma^2_{\mathrm{PN}}$): asymmetric 95% log-normal intervals ($v\,e^{\pm 1.96\,\mathrm{SE}(\log v)}$, shown as $^{+}_{-}$ offsets) reflecting the exponential variance parameterisation. The variance estimates carry a small rank-truncation bias (§8.4); their wide intervals indicate the variance error here is dominated by information deficiency rather than that bias. Columns are nominal gyro sampling rates; the two airspeed sweeps are stacked vertically.

The dynamic model was developed progressively, and the use of many similar manoeuvres allowed comparing estimated process-noise realisations directly for any discernible nonrandomness. If the model is misspecified, then the estimated process-noise realisations may contain features that are

| Sweep | Rate (Hz) | $q$ | $n_{\text{obs}}$ | $k_{1\text{p}}$ | $k_{2\text{p}}$ | $-\ell^*_{1\text{-pole}}$ | $-\ell^*_{2\text{-pole}}$ | $\Delta$AIC | $\Delta$BIC | $\max \eta$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 150 | 30 | 2578 | 30 | 31 | −2018.8 | −2022.9 | −6.3 | −0.5 | 0.23 |
| | 300 | 36 | 4864 | 30 | 31 | −3618.4 | −3633.5 | −28.3 | −21.8 | 0.21 |
| | 500 | 40 | 6974 | 30 | 31 | −5179.7 | −5194.9 | −28.4 | −21.5 | 0.19 |
| | 1000 | 48 | 13310 | 30 | 31 | −9999.7 | −10019.2 | −37.0 | −29.5 | 0.19 |
| 2 | 150 | 30 | 2578 | 30 | 31 | −1932.6 | −1947.4 | −27.6 | −21.7 | 0.16 |
| | 300 | 36 | 4864 | 30 | 31 | −3486.2 | −3500.8 | −27.1 | −20.6 | 0.19 |
| | 500 | 40 | 6979 | 30 | 31 | −4971.6 | −4988.5 | −31.7 | −24.9 | 0.19 |
| | 1000 | 48 | 13315 | 30 | 31 | −9623.1 | −9644.2 | −40.3 | −32.8 | 0.19 |

| Rate (Hz) | $\chi^2$ (7 dof) | $p$ |
|---|---|---|
| 150 | 8.1 | 0.320 |
| 300 | 2.2 | 0.946 |
| 500 | 2.6 | 0.916 |
| 1000 | 1.9 | 0.963 |

Table 3 – Two-pole (servo+aero) vs one-pole actuator model. $-\ell^*$ is the profile negative log-likelihood; $\Delta$ = two-pole − one-pole, so negative favours two poles. Both models share the bounds-fixed pot observation set ($n_{\text{obs}}$ identical per cell). $k_{1\text{p}}, k_{2\text{p}}$ are the free-parameter counts; the two-pole model adds a single parameter (the second actuator pole; the per-response surface-deflection initial state is held fixed), so its BIC penalty $k \log n_{\text{obs}}$ is marginally the larger. $q$ is the KL truncation rank (common across units); $\max \eta$ is the mode-information ratio of the least-significant retained mode ($\eta_{i,q_i}$, worst case over sampling units) for the two-pole fit – $\ll 1$ confirms the truncation has reached the prior-dominated regime. Lower panel: cross-sweep consistency of the two-pole core parameter vector (Wald $\chi^2$, 7 dof).

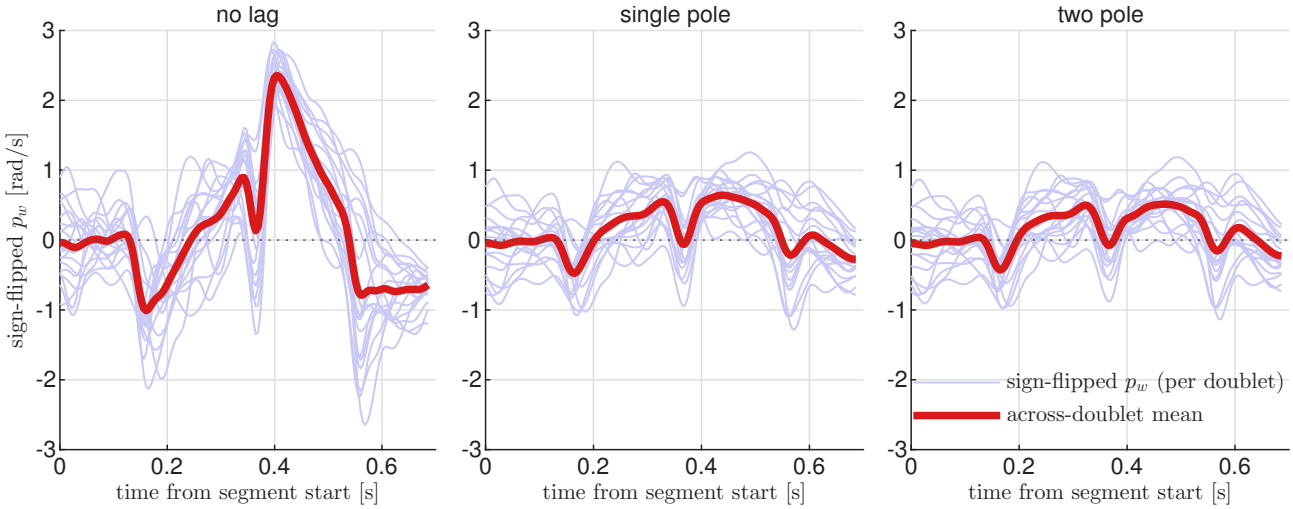



Figure 6 – Estimated process noise realisations, with a factor $-1$ applied to negative doublet inputs, for the three models applied to Sweep 1 dataset with 1000 Hz gyro data. The mean process noise (red line) reveals any consistent behaviour, indicating model deficiency that is being absorbed into the process-noise estimate.

similar across responses, which account for the same model misspecification affecting the responses. The no-pole estimated process-noise realisations in Figure 6 demonstrate a consistent behaviour that is apparent in the mean process noise, with a strong negative peak at $t = 0.15\,\text{s}$ and strong positive peak at $t = 0.4\,\text{s}$. These peaks are consistent with an unmodeled transient behaviour shortly after the aileron movements during the doublet, and this phenomenon motivated the one-pole and two-pole models during model development. The similar shape of the peak after $t = 0.4\,\text{s}$ in the one-pole estimated process-noise realisations motivated the addition of the second pole in the actuator's lag model. The two-pole lag model shows the least coherent estimated process-noise realisations, though three sharp negative peaks remain and a slow positive peak from $t = 0.2\,\text{s}$ to $t = 0.7\,\text{s}$. The

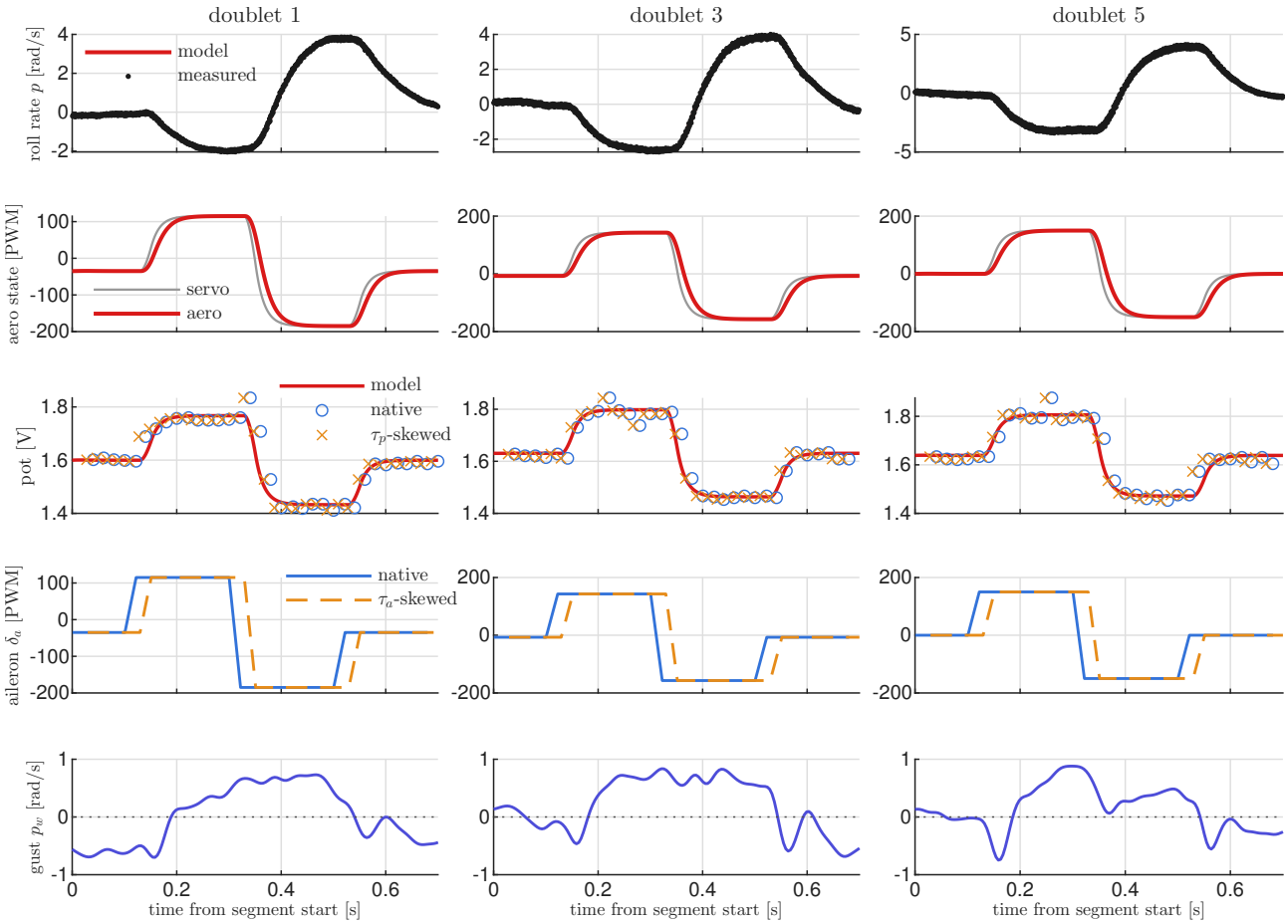



Figure 7 – Model and data for the first three positive doublets from Sweep 1 fit on 1000 Hz data.

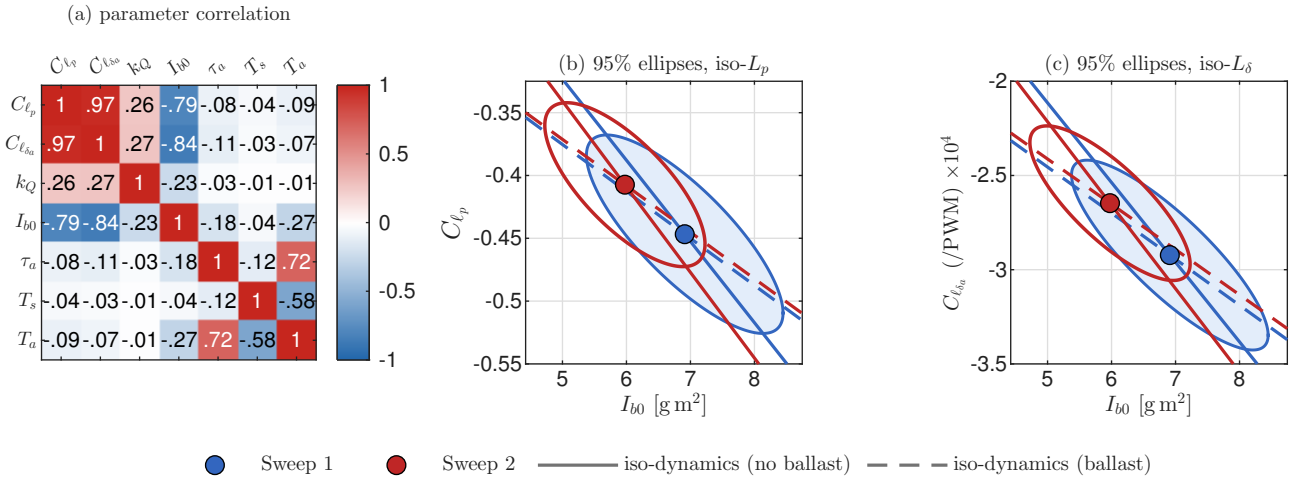



Figure 8 – Parameter estimates' correlation in Sweep 1 fit on 1000 Hz data. Subfigures b and c visualise the correlated parameters' relation to the dimensional dynamics ($L_p$ and $L_\delta$ for (b) and (c) respectively), for both sweeps' fits on 1000 Hz data.

sharp negative peaks coincide with the three aileron movements during each doublet, suggesting that some transient unmodeled dynamics remain. The slower positive peak may be attributable to unmodeled Dutch roll dynamics. All modelling is an approximation, and the two-pole lag model identified for the aileron actuator dynamics was deemed sufficient for the demonstration of GEM using MLA.

Formal discrimination requires a likelihood criterion evaluated on a common observation set. The no-pole model can already be set aside on the coherent process-noise realisations alone, but because

it omits the potentiometer channel its likelihood is computed on a different observation set and is not AIC/BIC-comparable. Among the one- and two-pole models, the likelihood improvement justifies the additional lag; both AIC and BIC favour two poles strongly in every case except the 150Hz case for Sweep 1 where the comparison is marginal. The AIC and BIC comparison is made at a common truncation rank $q$ for both models at each sampling rate and sweep, so their likelihoods are comparable. Likelihoods at different truncation ranks are not comparable, because the rank enters through $\log|\Delta_i|$, $\log|g_i^{*\prime\prime}|$ and $\hat{\sigma}^2$.

Because both sweeps measure the same airframe, the primary-interest parameters $\{C_{\ell_p},\ C_{\ell_{\delta_a}},\ k_Q,\ I_{b,0},\ \tau_a,\ T_s,\ T_a\}$ should agree across them. Because Sweep 1 and Sweep 2 comprise disjoint sets of doublets, their estimates are statistically independent, so the per-rate Wald tests (Table 3) use the summed covariance to test for a discrepancy exceeding the joint sampling uncertainty; none is found at any sampling rate ($p = 0.32$–$0.96$). This is a consistency check rather than a positive demonstration of reproducibility: given the modest per-parameter confidence the test has limited power, so a non-significant result indicates the absence of evidence of inconsistency rather than evidence of agreement. A significant result, by contrast, would have been a clear warning of model misspecification or an estimation pathology; its absence is necessary, but not sufficient, for confidence in the estimates.

The parameter estimates shown in Table 2 show differences in the primary interest parameters that are moderate relative to their magnitudes, but small relative to the joint sampling uncertainty (the Wald tests above). Whilst the confidence in parameter estimates is modest, the intent of MLA is to correctly model the experimental error, including process noise, to produce accurate confidence intervals. The contrast experiment that is enabled by jointly estimating across different parameterised linear dynamics models allows the aircraft body inertia to be estimated alongside the aerodynamic coefficient derivatives, whereas $\{C_{\ell_p},\ C_{\ell_{\delta_a}},\ I_{b,0}\}$ would be unidentifiable from any single system response. Similarly, because the dynamics are parameterised in terms of the airspeed-independent non-dimensional coefficients, a single global parameter set describes linear dynamic system responses at every airspeed. However, despite the modest confidence in parameters, the strong correlation (see Figure 8a) between $C_{\ell_p}$, $C_{\ell_{\delta_a}}$ and $I_{b,0}$ means that the ratio of these parameters is more tightly determined than the single-parameter confidence intervals might suggest, as visualised in Figures 8b and 8c. The extremely strong correlation between $C_{\ell_p}$ and $C_{\ell_{\delta_a}}$ reflects the fact that the roll rate does not exhibit significant behaviour in any of the system responses except as induced by the aileron, and the steady roll rate is well informed. The cluster of correlation in the timing parameters reflects the weakly informed trade-off between lag attributable to the servo dynamics, and lag attributable to the linkage-aero dynamics.

The confidence bounds presented in Table 2 show that increased sample rate provides only a modest improvement in parameter confidence. The information-deteriorating effect of process noise is uniform across gyro sampling rates, and higher gyro sampling rates mainly serve to provide a better resolved picture of the process noise, but do not de-confound the process noise's potential to explain what might otherwise be attributed to a system parameter. The timing parameters $\{\tau_a,\ T_a,\ \tau_p\}$ confidence improves the most at higher sample rate, reflecting their genuine higher-bandwidth behaviour. The $T$ timing parameter confidence improves only modestly, because it is informed primarily by the potentiometer data which are included at their native nominal $50\,\mathrm{Hz}$ rate in all fits.

## 10. Conclusion

In this paper we have proposed the Gust Error Method (GEM) as an alternative approach to aircraft system parameter identification subject to coloured process noise, to complement existing EEM, OEM and FEM techniques. The primary insight that enables GEM is representing the process noise realisation as a series expansion in Gaussian random coefficients that multiply basis functions determined by the Karhunen-Loève transform of the process' cross-autocovariance, and treating the coefficients as random effects in a mixed effects model. To apply the technique to flight data problems, an optimisation routine using the modified Laplace approximation was developed, making use of the linearity of a linear dynamic system with respect to input signals. Because the Karhunen-Loève decomposition of the process noise relies on evaluating the process' cross-autocovariance over the experiment domain, the technique is applicable to power spectral densities with fractional power laws,

such as von Kármán gust spectra. GEM expands on the scope of problems that FEM is applicable to, to include parameterised linear models, so that responses with differing linear systems but shared underlying parameters can be fit jointly in the presence of coloured process noise. Because it pools information across independent responses, each carrying its own gust realisation as a random effect, GEM also recovers the parameter consistency that single-response estimation, conditioned on one realisation of the disturbance, cannot attain.

The modified Laplace approximation formulation has been validated on synthetic data for comparison against a FEM baseline and the alternative Lindstrom and Bates alternating algorithm for solving mixed effects models. The performance of the estimation technique with increased process noise intensity and increased number of system responses was also investigated, affirming the selection of modified Laplace approximation formulation against the alternative. The mode information ratio was presented as a means of assessing the rank truncation for adequacy in capturing the modeled noise, given the available information in the data.

Across two independent datasets, GEM provided consistent parameter estimates. Despite modest parameter confidence, strong correlation in parameter estimates confines them to a narrow range of dynamic behaviour. The application showed that parameter confidence in estimates improved only modestly with higher sampling rate data about the same flight manoeuvres, because process noise provides an information-deteriorating effect that cannot be overcome by better resolving it.

The present formulation of GEM using modified Laplace approximation estimates process noise realisations using a rank-truncated approximation of the true cross-autocovariance of the process over the experiment domain. The Karhunen-Loève rank truncation provides optimality in the process noise space, but not in the measurement space where the dynamic system is actually observed. GEM is currently restricted to linear dynamic systems, though they may be parameterised across responses according to the operating point of each response. This linear restriction is shared with filter error method as presented in the literature, where the nonlinear filtering the general case requires makes the estimation considerably harder. Output error method, by contrast, applies directly to nonlinear dynamic systems, which remains a genuine advantage over GEM. Future improvements to GEM could include expansion to apply to nonlinear dynamic systems, alternative treatment of the rank truncation, a formulation that incorporates noise colour parameters, optimisation to reduce the computational cost, and application to more real world data.

## 11. Acknowledgements

The authors acknowledge helpful discussions with Dan Newman and Jeremy Randle. The authors acknowledge remote piloting by Jeremy Randle.

The authors acknowledge the statistical workshops and workflows provided by the Sydney Informatics Hub, a Core Research Facility of the University of Sydney.

The authors acknowledge the open source enabling technology of ArduPilot and related open source projects, and the software developers and community for helpful discussions.

The authors disclose use of LLM AI to assist with software development, mathematical calculations, and manuscript editing. All outputs were reviewed and validated by the authors. We take full responsibility for the accuracy and integrity of the results presented.

## 12. Contact Author Email Address

Correspondence related to this work should be directed to the first author, at jcox8640@uni.sydney.edu.au.
Flight data used in this work are available on reasonable request by email.

## 13. Copyright Statement

The authors confirm that they, and/or their company or organization, hold copyright on all of the original material included in this paper. The authors also confirm that they have obtained permission, from the copyright holder of any third party material included in this paper, to publish it as part of their paper. The authors confirm that they give permission, or have obtained permission from the copyright holder of this paper, for the publication and distribution of this paper as part of the ICAS proceedings or as individual off-prints from the proceedings.

## References


[1] K. W. Iliff and R. E. Maine. *Practical aspects of a maximum likelihood estimation method to extract stability and control derivatives from flight data*. Technical Note NASA-TN-D-8071. NASA, 1976.

[2] V. Klein and E. A. Morelli. *Aircraft system identification: theory and practice*. 1st. American Institute of Aeronautics and Astronautics, 2006.

[3] R. V. Jategaonkar. *Flight vehicle system identification: a time-domain methodology*. 2nd. American Institute of Aeronautics and Astronautics, 2015.

[4] J. A. Grauer and E. A. Morelli. "Aircraft parameter estimation considering process and measurement noise". In: *Journal of Aircraft* (2024).

[5] James Mercer. "Functions of Positive and Negative Type, and their Connection with the Theory of Integral Equations". In: *Philosophical Transactions of the Royal Society A* 209 (1909), pp. 415–446.

[6] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. Cambridge, MA: MIT Press, 2006. ISBN: 978-0-262-18253-9. URL: http://www.gaussianprocess.org/gpml/.

[7] K. Karhunen. *Über lineare Methoden in der Wahrscheinlichkeitsrechnung*. Annales Academiae Scientiarum Fennicae: Ser. A 1. Universität Helsinki, 1947.

[8] M. Loève. *Fonctions aléatoires du second ordre...* 1965.

[9] Roger G. Ghanem and Pol D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Revised. Mineola, NY: Dover Publications, 2003. ISBN: 978-0-486-42818-5.

[10] Douglas Bates et al. "Fitting Linear Mixed-Effects Models Using lme4". In: *Journal of Statistical Software* 67.1 (2015), pp. 1–48. DOI: 10.18637/jss.v067.i01.

[11] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2025. URL: https://www.R-project.org/.

[12] Kasper Kristensen et al. "TMB: Automatic Differentiation and Laplace Approximation". In: *Journal of Statistical Software* 70.5 (2016), pp. 1–21. DOI: 10.18637/jss.v070.i05.

[13] Uffe H. Thygesen et al. "Validation of ecological state space models using the Laplace approximation". In: *Environmental and Ecological Statistics* (2017), pp. 1–23. DOI: 10.1007/s10651-017-0372-4.

[14] The MathWorks, Inc. *MATLAB*. The MathWorks, Inc., Natick, Massachusetts. Versions R2020a–R2025a, including Optimization Toolbox and Statistics and Machine Learning Toolbox. URL: https://www.mathworks.com.

[15] M. J. Lindstrom and D. M. Bates. "Nonlinear mixed effects models for repeated measures data". In: *Biometrics* (1990).

[16] J. C. Pinheiro and D. M. Bates. *Mixed-effects models in S and S-PLUS*. Springer, 2000.

[17] Luke Tierney and Joseph B Kadane. "Accurate approximations for posterior moments and marginal densities". In: *Journal of the american statistical association* 81.393 (1986), pp. 82–86.

[18] José C. Pinheiro and Douglas M. Bates. "Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model". In: *Journal of Computational and Graphical Statistics* 4.1 (1995), pp. 12–35. DOI: 10.1080/10618600.1995.10474663. eprint: https://www.tandfonline.com/doi/pdf/10.1080/10618600.1995.10474663. URL: https://www.tandfonline.com/doi/abs/10.1080/10618600.1995.10474663.

[19] U.S. Department of Defense. *Military Specification: Flying Qualities of Piloted Airplanes*. Military Specification MIL-F-8785C. U.S. Department of Defense, Nov. 1980.

[20] ArduPilot Development Team. *ArduPilot: Open Source Autopilot*. https://ardupilot.org. Accessed: 2025-11-01. 2025.

[21] Matthew James Lindsay Anderson. *A Methodology for Aerodynamic Parameter Estimation of Tail-Sitting Multi-Rotors (PhD Thesis)*. 2017.

# A Gradient and Hessian of the Log Likelihood

## 1.1 Gradient With Respect To Observation Model and Variance Model

The gradients $\frac{d\ell^*}{d\Theta}$ ($\frac{d\ell^*}{d\theta}$ and $\frac{d\ell^*}{d\zeta}$) and $\frac{d\ell^*}{d\vec{b}_i}$ are required for gradient-based minimisation of (11). First, define four sub-terms

$$\ell^* = -\frac{1}{2}\sum_{i=1}^{M} \overbrace{n_i\left[1+\log(2\pi)+\log(\hat{\sigma}^2)\right]}^{T_1} + \underbrace{\log|\Delta_i(\zeta)|}_{T_2} + \underbrace{\log\left|g_i^{*\prime\prime}(\Theta,\vec{Z}_i,\hat{\vec{b}}_i)\right|}_{T_3} + \underbrace{\log|\mathbf{R}_i(\zeta)|}_{T_4}, \tag{27}$$

which contribute their per-sampling-unit derivatives additively to the $\ell^*$ gradient and therefore can be examined independently.

### *Derivative of $T_1$*

By the envelope theorem, since $\hat{\vec{b}}_i = \arg\min_{\vec{b}_i} g_i^*$ implies $\frac{dg_i^*(\hat{\vec{b}}_i)}{d\vec{b}_i} = \vec{0}$ (so that the implicit contribution to $\frac{d\hat{\sigma}^2}{dx}$ through $\frac{d\hat{\vec{b}}_i}{dx}$ vanishes) the gradient of $T_1$ for $x \in \{\theta, \zeta, \vec{b}_i\}$ reduces to

$$\frac{dT_1^{(i)}}{dx} = \frac{n_i}{\hat{\sigma}^2}\frac{d\hat{\sigma}^2}{dx}, \tag{28}$$

$$\frac{d\hat{\sigma}^2}{dx} = \frac{1}{N}\sum_{i=1}^{M}\frac{dg_i^*(\hat{\vec{b}}_i)}{dx}. \tag{29}$$

The per-sampling-unit $g_i^*$ derivatives are

$$\frac{dg_i^*}{d\theta} = -2\vec{r}_i^T\,\mathbf{R}_i^{-1}\frac{d\vec{Y}_i(\theta,\hat{\vec{b}}_i)}{d\theta}, \tag{30}$$

$$\frac{dg_i^*}{d\zeta} = -\vec{r}_i^T\,\mathbf{R}_i^{-1}\frac{d\mathbf{R}_i(\zeta)}{d\zeta}\mathbf{R}_i^{-1}\vec{r}_i - \hat{\vec{b}}_i^T\,\Delta_i(\zeta)^{-1}\frac{d\Delta_i(\zeta)}{d\zeta}\Delta_i(\zeta)^{-1}\hat{\vec{b}}_i, \tag{31}$$

$$\frac{dg_i^*}{d\vec{b}_i} = -2\frac{d\vec{Y}_i(\theta,\vec{b}_i)}{d\vec{b}_i}^T\mathbf{R}_i^{-1}\vec{r}_i + 2\Delta_i(\zeta)^{-1}\vec{b}_i, \tag{32}$$

with $\vec{r}_i = \vec{Z}_i - \vec{Y}_i(\theta,\vec{b}_i)$. The gradient with respect to $\vec{b}_i$ vanishes at the inner optimum $\hat{\vec{b}}_i$ but is non-zero for general $\vec{b}_i$, as required for joint optimisation over $\Theta$ and $\vec{b}_i$.

### *Derivatives of $T_2$ and $T_4$*

$T_2 = \log|\Delta_i(\zeta)|$ and $T_4 = \log|\mathbf{R}_i(\zeta)|$ are log-determinants of covariances that depend only on $\zeta$, so $\frac{dT_2^{(i)}}{d\theta} = \frac{dT_4^{(i)}}{d\theta} = 0$ and $\frac{dT_2^{(i)}}{d\vec{b}_i} = \frac{dT_4^{(i)}}{d\vec{b}_i} = \vec{0}$. By Jacobi's formula for the derivative of the log-determinant,

$$\frac{dT_2^{(i)}}{d\zeta} = \mathrm{tr}\left(\Delta_i(\zeta)^{-1}\frac{d\Delta_i(\zeta)}{d\zeta}\right), \tag{33}$$

$$\frac{dT_4^{(i)}}{d\zeta} = \mathrm{tr}\left(\mathbf{R}_i(\zeta)^{-1}\frac{d\mathbf{R}_i(\zeta)}{d\zeta}\right). \tag{34}$$

### *Derivative of $T_3$*

For a linear dynamic system, the Hessian $g_i^{*\prime\prime}$ simplifies to

$$g_i^{*\prime\prime} = 2\frac{d\vec{Y}_i(\theta,\vec{b}_i)}{d\vec{b}_i}^T \mathbf{R}_i(\zeta)^{-1}\frac{d\vec{Y}_i(\theta,\vec{b}_i)}{d\vec{b}_i} + 2\Delta_i(\zeta)^{-1}, \tag{35}$$

which is independent of $\vec{b}_i$ (the constant-Hessian property described in §7.) but varies with $\Theta$. Jacobi's formula gives

$$\frac{dT_3^{(i)}}{dx} = \operatorname{tr}\left(\left(g_i^{*\prime\prime}\right)^{-1}\frac{dg_i^{*\prime\prime}}{dx}\right), \qquad x \in \{\theta,\zeta\}, \tag{36}$$

$$\frac{dT_3^{(i)}}{d\vec{b}_i} = \vec{0}, \tag{37}$$

with the Hessian derivatives

$$\begin{aligned}\frac{dg_i^{*\prime\prime}}{d\theta} &= 2\frac{d^2\vec{Y}_i}{d\vec{b}_i\,d\theta}^T \mathbf{R}_i(\zeta)^{-1}\frac{d\vec{Y}_i}{d\vec{b}_i} \\ &\quad + 2\frac{d\vec{Y}_i}{d\vec{b}_i}^T \mathbf{R}_i(\zeta)^{-1}\frac{d^2\vec{Y}_i}{d\vec{b}_i\,d\theta},\end{aligned} \tag{38}$$

$$\begin{aligned}\frac{dg_i^{*\prime\prime}}{d\zeta} &= -2\frac{d\vec{Y}_i}{d\vec{b}_i}^T \mathbf{R}_i(\zeta)^{-1}\frac{d\mathbf{R}_i(\zeta)}{d\zeta}\mathbf{R}_i(\zeta)^{-1}\frac{d\vec{Y}_i}{d\vec{b}_i} \\ &\quad - 2\Delta_i(\zeta)^{-1}\frac{d\Delta_i(\zeta)}{d\zeta}\Delta_i(\zeta)^{-1}.\end{aligned} \tag{39}$$

### *Assembled gradient*

The gradient of $\ell^*$ aggregates the per-term contributions with the overall $-\frac{1}{2}$ factor from (27):

$$\frac{d\ell^*}{dx} = -\frac{1}{2}\sum_{i=1}^{M}\left[\frac{dT_1^{(i)}}{dx} + \frac{dT_2^{(i)}}{dx} + \frac{dT_3^{(i)}}{dx} + \frac{dT_4^{(i)}}{dx}\right]. \tag{40}$$

For a linear dynamic system, the per-parameter contributions specialise:

$$\frac{d\ell^*}{d\theta} = -\frac{1}{2}\sum_{i=1}^{M}\left[\frac{dT_1^{(i)}}{d\theta} + \frac{dT_3^{(i)}}{d\theta}\right], \tag{41}$$

$$\frac{d\ell^*}{d\zeta} = -\frac{1}{2}\sum_{i=1}^{M}\left[\frac{dT_1^{(i)}}{d\zeta} + \frac{dT_2^{(i)}}{d\zeta} + \frac{dT_3^{(i)}}{d\zeta} + \frac{dT_4^{(i)}}{d\zeta}\right], \tag{42}$$

$$\frac{d\ell^*}{d\vec{b}_i} = -\frac{1}{2}\frac{dT_1^{(i)}}{d\vec{b}_i}. \tag{43}$$

The assembled gradient $\frac{d\ell^*}{d\Theta}$, including the $T_3$ mixed-derivative term, was verified against a real central-difference of the profile log-likelihood.

## 1.2 Finite Differencing the Dynamic System Model

MATLAB's native complex type allows complex step differencing of the observation model evaluation. Complex step differencing requires the observation to be analytic in $\theta$; a linear dynamic system satisfies this when its matrices are analytic and the integrator uses no non-analytic operations. In this work, the entire state trajectory is integrated and observations evaluated using imaginary perturbations to $\theta$. The second mixed derivative required to evaluate $\frac{d^2\vec{Y}_i}{d\vec{b}_i\,d\theta}$ can be exactly evaluated by

noticing that an imaginary step with respect to $\theta$ and a real step with respect to $\vec{b}_i$ gives the Taylor expansion

$$\vec{Y}\left(\Theta + ih\vec{e}_p,\, \vec{b} + k\vec{e}_m\right) = \left(\vec{Y} + ih\,\partial_{\Theta_p}\vec{Y} - \tfrac{h^2}{2}\,\partial^2_{\Theta_p}\vec{Y} - i\tfrac{h^3}{6}\,\partial^3_{\Theta_p}\vec{Y}\right) + k\,\partial_{\vec{b}_m}\left(\vec{Y} + ih\,\partial_{\Theta_p}\vec{Y} - \tfrac{h^2}{2}\,\partial^2_{\Theta_p}\vec{Y} - i\tfrac{h^3}{6}\,\partial^3_{\Theta_p}\vec{Y}\right) + \mathscr{O}(h^4), \tag{44}$$

which leads to the finite differencing stencil

$$\frac{\partial^2\vec{Y}}{\partial\Theta_p\,\partial\vec{b}_m} = \frac{\Im\left[\vec{Y}\left(\Theta + ih\vec{e}_p,\, \vec{b} + k\vec{e}_m\right) - \vec{Y}\left(\Theta + ih\vec{e}_p,\, \vec{b}\right)\right]}{hk} + \mathscr{O}(h^2). \tag{45}$$

The model evaluation $\vec{Y}\left(\Theta + ih\vec{e}_p,\, \vec{b}\right)$ can be reused from the complex step evaluation for the first derivative of the system model. Because $\vec{Y}_i$ is exactly linear in $\vec{b}_i$, the real step $k$ enters the stencil exactly: there is no $\mathscr{O}(k)$ truncation error, and $k$ may be taken of order unity (indeed $k = 1$) rather than small. This is also the numerically prudent choice: the cross term is recovered from the difference of two nearly-equal imaginary parts, so a small $k$ would reintroduce the subtractive cancellation that complex step differencing otherwise avoids, whereas an $\mathscr{O}(1)$ step keeps the difference well above the round-off floor.

## 1.3 Variance Parameterisation

Both the measurement-noise and random-effects covariances use an exponential parameterisation, so each variance enters positively without constraining $\zeta$. The measurement-noise covariance is block-diagonal across observation channels, with the channel-$j$ block

$$\mathbf{R}_{i,j}(\zeta) = \exp(\zeta_m)\,\mathbf{I}_{n_{i,j}}, \tag{46}$$

where $\zeta_m$ is the entry of $\zeta$ assigned to channel $j$. This block is self-deriving ($\partial_{\zeta_m}\mathbf{R}_{i,j} = \mathbf{R}_{i,j}$, and zero for any other entry of $\zeta$), and the random-effects covariance $\Delta_i(\zeta)$ takes the same form with the same identity. Variance parameters may be shared across channels, for example two channels that are identical sensors, or two process-noise channels that are gust components sharing a variance scale.

## 1.4 Hessian and Standard Errors

The standard errors reported in this work are obtained from the Hessian of the profile log-likelihood with respect to the full parameter vector $\Theta$, including the profiled scale variance $\sigma^2$, formed by the method described in [16].

Because the gradient $\frac{d\ell^*}{d\Theta}$ is evaluated to machine precision (the observation-model first derivatives by complex-step differentiation and the mixed second derivative $\frac{d^2\vec{Y}_i}{d\vec{b}_i\,d\theta}$ by the imaginary-and-real stencil described above, both free of subtractive cancellation), this Hessian can be formed by ordinary real finite-differencing of the gradient.