Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Stable Gray-Box Surrogate for Unsteady Aerodynamic Coefficients in Real-Time Simulation

Massimo Di Pierro

massimo.dipierro@gmail.com

University of California

Research Article

Keywords: unsteady aerodynamics, dynamic stall, aerodynamic coeffcient prediction, real-time flight simulation, reduced-order model, Wiener model, recurrent neural network, benchmark

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11178486/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Stable Gray-Box Surrogate for Unsteady Aerodynamic Coefficients in Real-Time Simulation

Massimo Di Pierro
University of California, Santa Cruz, CA 95064, USA

Abstract

Real-time flight simulation needs aerodynamic loads at every step; coefficient tables are quasi-steady and grow exponentially with their arguments. We study a compact gray-box surrogate: a bank of learnable second-order lags (digital filters whose response always decays), feeding a small neural network that supplies the nonlinear, stall-like map, plus a linear term for rate effects. It is benchmarked against twelve alternatives, including calibrated Hansen–Gaunaa–Madsen (HGM) and Goman–Khrabrov dynamic-stall models and GRU and LSTM networks, on measured two-dimensional airfoil data: 216 pitch-oscillation runs of three airfoils under cross-validation, and ramps of three further airfoils after training on oscillations only. On oscillations the surrogate is significantly more accurate than quasi-static tables and linear models; for unseen Reynolds numbers it is also more accurate than the calibrated semi-empirical models, but for unseen reduced frequencies it only matches them and is outperformed by rate-augmented memoryless maps. On ramps it outperforms those memoryless maps and predicts dynamic-stall peak loads better than the calibrated semi-empirical models, which nevertheless have the lower time-history error. Small gated recurrent networks are more accurate, by 7–53%, without a stability guarantee on their internal dynamics. The learned filters act as a stable basis of lags rather than as identified physical time constants.

Keywords: unsteady aerodynamics; dynamic stall; aerodynamic coefficient prediction; real-time flight simulation; reduced-order model; Wiener model; recurrent neural network; benchmark.

1 Introduction

A real-time vehicle simulation must compute the aerodynamic force and moment on the airframe at every integration step, typically hundreds of times per second, and never slower than wall-clock time, so that the loop can close around a pilot, a flight computer or hardware in the loop. Resolving the flow with computational fluid dynamics (CFD) at each step is many orders of magnitude too slow for this budget, so the aerodynamics are reduced to a cheap surrogate evaluated from a handful of state variables.

The standard surrogate is a table of the dimensionless force and moment coefficients (C_D, C_L, C_Y and the rolling, pitching and yawing moments) indexed by angle of attack α, sideslip β, Mach number, body rates and control deflections, scaled to dimensional loads by dynamic pressure and reference geometry. Tables are fast and transparent, but they have two structural limitations. Their memory grows exponentially with the number of indexing arguments, and they are *quasi-steady*: they return a coefficient for the instantaneous state, whereas the real loads depend on the recent *history* of the motion. Lift and moment respond to a lagged effective incidence; near and beyond stall the response forms hysteresis loops and the transient overshoots of dynamic stall. Adding rate or history arguments to a table compounds the memory problem, and a feed-forward neural network of the instantaneous state has no memory either.

These limitations are no longer academic. Flight simulators used for stall and upset-recovery training must now reproduce the post-stall and unsteady aerodynamics of the extended envelope [1]; rotorcraft and wind-turbine simulations routinely operate in dynamic stall; and the developers of flight software increasingly test it against simulated aerodynamics in closed loop. The alternatives to tables each have a gap (Section 2). Semi-empirical dynamic-stall models of the Leishman–Beddoes family are physically grounded and cheap, but need per-airfoil calibration and structural assumptions that fail in deep stall. Rational-function approximations (RFA) from aeroelasticity are linear. Recurrent neural networks such as long short-term memory (LSTM) and gated recurrent unit (GRU) networks are flexible, but their internal dynamics carry no stability guarantee, their parameters have no physical meaning, and in the aerodynamic literature they are rarely compared with a calibrated semi-empirical model on held-out measured data.

This paper studies a model that sits between these classes, built the same way as the semi-empirical models: lagged internal variables followed by a static nonlinear characteristic. The lags are a bank of learnable second-order discrete-time filters ("biquads"), each a damped second-order response similar to the lags of Jones's approximation but able also to overshoot; the static characteristic is a small neural network (a multilayer perceptron, MLP) that is fitted instead of prescribed; and an explicit linear term carries the near-linear rate effects. In system-identification terms this lag-then-static-map arrangement is a *Wiener* model. The filters are parameterized so that their response always decays, whatever values training gives the parameters; each carries two past values as its state, and the cost per time step is fixed and grows linearly with the number of inputs. It can be read as a trainable nonlinear extension of the RFA, or as a data-driven generalization of the "static characteristic at lagged internal variables" structure shared by the semi-empirical dynamic-stall models.

The questions the paper answers are empirical: *does this structure predict measured unsteady airfoil loads, including dynamic stall, better than the methods it would replace, and does it do so robustly across flow conditions, reduced frequencies, airfoils and motion types?* The contributions are:

1. a formulation of the biquad-Wiener surrogate for unsteady aerodynamic coefficients, with its stability parameterization and an exact, loop-free training scheme for multi-run data (Sections 3–4);

2. a benchmark on *measured* unsteady data: 216 pitch-oscillation runs of three wind-turbine and general-aviation airfoils from the Ohio State University (OSU) campaign [2, 3, 4], under leave-one-Reynolds-number-out and leave-one-frequency-out cross-validation, against twelve alternatives spanning look-up tables, a classical RFA, calibrated Hansen–Gaunaa–Madsen and Goman–Khrabrov dynamic-stall models, and memoryless, finite-memory and recurrent neural networks, with three training seeds and paired significance tests (Section 7);

3. a motion-transfer test on the University of Glasgow dynamic-stall database [5], in which every model is trained on sinusoidal oscillations only and evaluated on measured constant-pitch-rate ramps of three airfoils, a motion family that separates models carrying a state from rate-augmented static maps (Section 8);

4. an analysis showing that the learned filters act as a stable basis of lags rather than identifying physical time constants, a comparison of nominal computational cost, and a discussion of the conditions in which the approach does and does not help (Sections 9–10).

All experiments use two-dimensional wind-tunnel sections; application to a full six-degree-of-freedom aircraft database is discussed but not tested (Section 10). The model is implemented as one of the

models of the HiroSim simulator (https://hirosim.com), where it runs as an inference-only plugin (Section 5); a synthetic verification case with known dynamics is retained in condensed form (Section 6).

2 Background and related work

This section reviews the four bodies of work the proposed model draws on or competes with: the aerodynamic databases of flight simulation, the physics and semi-empirical modeling of unsteady and separated airfoil flow, reduced-order models from aeroelasticity, and data-driven sequence models. Table 1 summarizes how the model classes compare on the properties that matter for a real-time simulator.

2.1 Aerodynamic models in flight simulation

The aerodynamic model of a flight simulator has traditionally been a map from the instantaneous state to force and moment coefficients. Modern engineering and training simulators store that map as multidimensional tables indexed by angle of attack, sideslip, Mach number, control deflections and configuration, supplemented by dynamic (rate) derivatives. The tables are populated from wind-tunnel campaigns, CFD, and flight-test system identification [6], and their fidelity is limited by grid resolution, by the number of independent arguments that can be afforded, and by the quasi-steady assumption itself.

The quasi-steady assumption fails exactly where recent regulation has pushed simulators. Tobak and Schiff [7] showed that the aerodynamic response in aircraft dynamics is a functional of the motion history, of which rate derivatives are only the first-order approximation; near and beyond stall the history dependence becomes nonlinear and produces hysteresis. Goman and Khrabrov [8] made this tractable with a single internal state, a flow-separation variable obeying a first-order lag, and such stall models have been identified directly from flight-test data [9]; Greenwell [10] reviews the resulting family of models for manoeuvring aircraft. Loss-of-control research and the U.S. rule requiring full-flight simulators used for stall and upset-recovery training to be qualified for the extended envelope from March 2019 [1], have turned post-stall and unsteady aerodynamic modeling from a research topic into an operational requirement. Each added history or rate argument, however, multiplies the size of a table, which is the scaling problem this paper addresses.

2.2 Unsteady aerodynamics and dynamic stall

For attached flow the unsteady response is linear and classical: Wagner's indicial function [11] and Theodorsen's function [12] give the lift response to a step and to harmonic motion, and Jones's approximation [13] turns the former into two first-order lags with time constants of about 22 and 3.3 semichords of travel. The unsteady effects are governed by the reduced frequency $k = \omega c/(2U)$; the flow is commonly regarded as quasi-steady for $k < 0.05$ [14]. Past the static stall angle the flow develops *dynamic stall*: separation is delayed, a leading-edge vortex forms and convects over the chord producing lift and nose-down moment overshoots, and reattachment lags on the downstroke, so that the loads trace wide hysteresis loops [15].

Engineering practice relies on *semi-empirical* models that add a few lag states to a measured static polar. The Leishman–Beddoes model [16] combines an indicial attached-flow response, a leading-edge pressure lag, a Kirchhoff trailing-edge separation point with its own lag, and a vortex-lift module; the state-space formulation of Hansen, Gaunaa and Madsen (HGM) [17] keeps the first three and is widely used in wind-energy codes; the Goman–Khrabrov model [8] reduces the

separation dynamics to a single lag. These models reproduce measured loops but need per-airfoil calibration, and they deviate from experiment in deep stall [18]. They share the structure exploited here: a static nonlinear characteristic evaluated at *lagged* internal variables, the lags being first-order filters whose time constants are fixed in convective time.

2.3 Reduced-order models from aeroelasticity

Aeroelastic analysis needs the unsteady aerodynamic loads in the time domain, which led to rational-function approximations (RFA) of frequency-domain aerodynamic influence coefficients, from Roger's method [19] to extensions that optimize the lag roots [20]. An RFA is a linear state-space model whose poles ("lag roots") are chosen by the analyst or optimized, and whose residues are fitted to data. Nonlinear extensions include Volterra series [21]; Kou and Zhang [22] review the data-driven reduced-order models of unsteady aerodynamics. Block-oriented models, in which a linear dynamic block and a static nonlinearity act in series, are a natural compromise; Wiener-type models have been applied to transonic unsteady aerodynamics [23], and their identification is surveyed in [24].

2.4 Machine-learning models of unsteady loads

Neural networks were applied to unsteady separated flow as early as the work of Faller, Schreck and Luttges [25]. Recurrent networks, in particular long short-term memory (LSTM) networks, have since been used for unsteady aerodynamics across flow conditions [26], including transonic buffet on pitching and plunging airfoils [27] and dynamic stall loads over the full operating range of a pitching airfoil [28]; Brunton, Noack and Koumoutsakos [29] review machine learning for fluid mechanics. The most common architectures, LSTM and GRU networks [30, 31], carry an internal state that is updated at each step through learned nonlinear "gates" deciding how much of the past to keep. That state stays bounded, but its dynamics are not guaranteed to be stable and the network weights have no physical meaning. In addition, many studies train and test on a single airfoil and motion family, often from CFD, and comparisons against a calibrated semi-empirical model on held-out experimental conditions are rare.

2.5 Structured linear recurrences

Machine learning and signal processing have converged on models that pair *linear lags whose stability is built into their parameterization* with a static nonlinearity. Trainable second-order digital filters of this kind were introduced in audio processing, for equalizers [32] and for compact models of nonlinear audio effects [33]; in system identification, dynoNet [34] uses trainable transfer functions as blocks of lag-plus-nonlinearity models; and in sequence modeling, structured state-space models [35] show that banks of stable linear lags followed by a nonlinear read-out can match or exceed gated recurrent networks. The model of this paper is a simple member of this family. Its contribution lies in the specialization to unsteady aerodynamic coefficients, the interpretation as a trainable nonlinear RFA, the implementation in a deterministic simulator and, most of all, the benchmark against established methods on measured unsteady data.

Table 1: Qualitative comparison of model classes for real-time unsteady aerodynamic coefficients. "Stable by construction" refers to the internal recurrence for every parameter value, not to closed-loop stability.

Class	Memory	Nonlinear map	Stable by construction	Parameters interpretable	Data needed
Quasi-static table	none	tabulated	n/a	yes	static polars
Rate-augmented table	rates only	tabulated	n/a	yes	dynamic tests
Linear RFA	fixed lags	none	if poles constrained	lag roots	frequency response
Semi-empirical DS model	2–12 fixed lags	static polar + rules	yes	yes	polar + calibration
Gated RNN (GRU/LSTM)	learned	learned	no	no	large sets
Proposed Wiener model	stable lag basis	learned	yes	no (Sec. 9)	time histories

3 Model

3.1 Structure

Let the input at discrete time n (time step Δt) be $u[n] \in \mathbb{R}^N$ (for an airfoil, e.g. the angle of attack α, the reduced pitch rate $\hat{q} = \dot{\alpha}c/(2U)$ and the airspeed U; for an aircraft, any set of state and control channels) and the target the coefficient vector $C[n] \in \mathbb{R}^K$, e.g. (C_L, C_D, C_m). (Capitals denote aircraft or generic coefficients; the airfoil data of Sections 7–8 use the section lift C_l, the pressure drag $C_{d,p}$, the normal force C_n and the quarter-chord moment C_m.) The model has three parts (Fig. 1): a bank of lags, a static nonlinear map, and a linear term.

Input scaling. Each input is first made dimensionless and of order one, $\hat{u}_i[n] = (u_i[n] - \mu_i)/\sigma_i$, with the mean μ_i and standard deviation σ_i of the training data (stored with the model), so that no input dominates the fit merely because of its units.

Lags. Each scaled input feeds M second-order filters (biquads). Filter f, acting on input $c(f)$, updates its output g_f once per time step from the current and two previous inputs and its own two previous outputs:

$$g_f[n] = b_{0f}\hat{u}_{c(f)}[n] + b_{1f}\hat{u}_{c(f)}[n-1] + b_{2f}\hat{u}_{c(f)}[n-2] - a_{1f}g_f[n-1] - a_{2f}g_f[n-2], \tag{1}$$

This is the discrete-time form of a second-order linear system: the a coefficients set how fast the response decays and whether it oscillates (Section 3.2), the b coefficients its gain and shape. A filter with $a_2 = 0$ and one b is a first-order lag, the building block of Jones's approximation and of the semi-empirical models. The $F = NM$ filters act side by side on the inputs, giving F lagged versions of the motion history.

Static nonlinear map and linear term. The lagged signals and the scaled inputs themselves form the vector

$$v[n] = (g_0[n], \ldots, g_{F-1}[n], \hat{u}_0[n], \ldots, \hat{u}_{N-1}[n]) \in \mathbb{R}^D,$$

($D = F + N$ entries), which a one-hidden-layer MLP maps to the coefficients. The MLP is a flexible curve fit: each of its H "hidden units" is a smooth step, tanh of a weighted sum of the entries of

v, and each coefficient is a weighted sum of these steps, which is enough to represent a stall break or a drag rise. A linear term L adds a direct linear dependence on the inputs:

$$h_j[n] = \tanh\Big(b^1_j + \textstyle\sum_d W^1_{jd} v_d[n]\Big), \qquad o_k[n] = b^2_k + \textstyle\sum_j W^2_{jk} h_j[n] + \sum_i L_{ki} \hat{u}_i[n], \tag{2}$$

each converted back to physical units as $C_k = o_k s_k + m_k$. The weights W, b and L are fitted to data. The linear term carries the near-linear effects, most importantly the rate (damping) derivatives, so that the MLP models the residual nonlinear, history-dependent part rather than competing with a large linear term. The filter bank is shared by all coefficients: one internal unsteady state drives every coefficient through a per-coefficient read-out, as a single separation state drives lift, drag and moment in the semi-empirical models. Passing the scaled inputs directly to the MLP (the "skip" path of Fig. 1) lets a purely static input, such as a configuration flag or the airspeed, shape the map without passing through the lags.

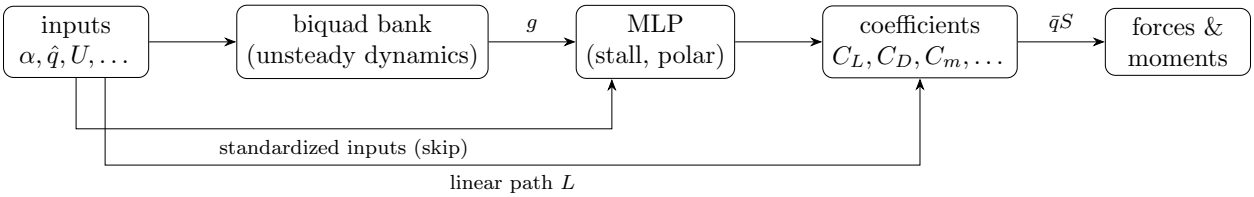


Figure 1: The biquad-Wiener surrogate: learnable biquads (unsteady dynamics) feed a small MLP (static stall and polar map); the standardized inputs also reach the MLP directly and, through the linear path, the coefficients. The coefficients are scaled to dimensional loads by dynamic pressure and reference geometry.

3.2 Stability by construction

The response of filter (1) to a single pulse of input decays like r^n and oscillates with a phase advance θ per step: r is the fraction of the response that survives each time step, and $\theta = 0$ means no oscillation. The filter is stable, i.e. its response dies out instead of growing, if and only if $r < 1$ (in signal-processing terms, its two "poles" $re^{\pm i\theta}$ lie inside the unit circle). Training could in principle push the coefficients a_1, a_2 into the unstable region, so they are not fitted directly: each filter carries two free parameters (ρ_f, θ_f) from which they are computed as

$$r_f = \sigma(\rho_f) \in (0, 1), \quad a_{1f} = -2 r_f \cos\theta_f, \quad a_{2f} = r_f^2, \tag{3}$$

where $\sigma(\rho) = 1/(1 + e^{-\rho})$ lies between 0 and 1 for every ρ. Hence $r_f < 1$ whatever values the optimizer chooses, and every filter is stable without any constraint during training; $\theta_f \to 0$ gives a non-oscillating, critically damped lag. A serialized coefficient file is checked against the discrete-time stability condition again before a simulator accepts it. Since stable filters give bounded outputs for bounded inputs, and the MLP maps bounded values to bounded values, the whole model returns bounded coefficients for any bounded input history (it is "bounded-input bounded-output"). A general recurrent neural network has no such guarantee: its internal state is updated by a nonlinear function of itself, and small perturbations of that state can grow even when its outputs remain bounded.

Throughout the paper, "stable by construction" means exactly this property of the surrogate's internal lags, and the resulting bounded response. It does not imply stability of the closed loop the surrogate forms with the vehicle dynamics and a flight-control system (a stable quasi-static table

can also destabilize a closed loop through wrong damping), nor robustness to inputs outside the training envelope, nor numerical robustness on a particular embedded target, nor acceptability for simulator qualification; these are system-level properties to be established by conventional means. What the property removes is one failure mode specific to learned recurrent surrogates, an internal state that grows or oscillates on its own, and it can be checked by inspecting the poles of a coefficient file.

3.3 From coefficients to loads

The outputs are dimensionless. Dimensional loads follow from dynamic pressure $\bar{q} = \frac{1}{2}\rho V^2$ and reference geometry (area S, chord c, span b):

$$\text{force} = \bar{q}\, S\, C, \qquad \text{moment} = \bar{q}\, S\, \ell\, C, \tag{4}$$

with $\ell = c$ for pitch and $\ell = b$ for roll and yaw. In a simulator $\bar{q}$ is read from a channel or computed from airspeed and density; the network itself predicts only dimensionless coefficients.

4 Training

Fitting criterion and data. Training adjusts all the filter and MLP parameters to minimize the mean-square difference between predicted and measured coefficients. The data are time histories of inputs and measured coefficients, generally several separate runs (wind-tunnel records, flight-test manoeuvres, CFD simulations) of different lengths and sampling rates. The model is advanced on a fixed simulator grid of step Δt, the rate at which it will run in the simulator, and its outputs are interpolated linearly to the measurement instants, where the loss

$$\mathcal{L} = \frac{1}{JK} \sum_{\text{runs}} \sum_{j} \sum_{k=1}^{K} \big(o_k(t_j) - y_k(t_j)\big)^2 \tag{5}$$

is evaluated with each coefficient scaled like the inputs (J the number of measured samples). Every run starts from zero filter state; to keep the arbitrary initial condition out of the loss, the inputs are extended backwards by a pre-roll (one period of the motion for periodic tests, a hold at the initial state for transients) during which no loss is accumulated. Models are always checked on whole runs withheld from training ("validation" runs), never on randomly chosen points of the training runs, which the model would partly have seen through its memory.

Computing the fit. Minimizing $\mathcal{L}$ requires its derivatives with respect to every parameter. For the MLP these are standard; through the filters they are obtained by propagating sensitivities backwards in time ("back-propagation through time"). The canonical trainer of the simulation stack does this with a hand-derived recursion whose derivatives agree with finite differences to better than 10^{-9}. For the benchmark, a faster equivalent is used: the output of a linear filter is the superposition of its responses to each past input sample, i.e. the convolution of the input with its pulse response, which for a biquad has the closed form

$$h_f[n] = b_{0f} p_f[n] + b_{1f} p_f[n{-}1] + b_{2f} p_f[n{-}2], \qquad p_f[n] = r_f^n \, \frac{\sin\big((n{+}1)\theta_f\big)}{\sin\theta_f}. \tag{6}$$

Evaluating the convolution by fast Fourier transform processes all time steps at once, and standard automatic-differentiation software then supplies the derivatives. The two implementations agree to 3×10^{-13}, and a model trained this way and exported to the simulator's coefficient format reproduces its predictions when run step by step (Section 7.4).

Optimization. The parameters are updated iteratively with the Adam gradient method [36] (step size 10^{-2}, reduced gradually during training; gradient norm limited to 10), each iteration ("epoch") using all training runs. Training runs for a fixed number of epochs, and the parameters retained are those with the lowest error on the validation runs ("early stopping"), which guards against fitting measurement noise. The poles are initialized log-uniformly in time constant between 2 and 60 simulator steps with small θ, and the numerators so that every filter starts with a unit steady-state gain. Training is offline; the runtime is inference only.

5 Real-time deployment

At each simulator step the runtime reads the declared input channels in physical units, standardizes them, advances every biquad once by (1), evaluates (2), publishes the dimensionless coefficients, and optionally the dimensional loads (4). The runtime state is the two-sample input and output history of each biquad; it is the only quantity to save and restore at simulator checkpoints. The model dimensions are fixed when the coefficient file is loaded, and the forward pass uses fixed-size buffers with no allocation and no interpolation search. The denominator coefficients are checked for stability before use and, if dimensional loads are requested, the configuration must provide a valid dynamic-pressure source and positive reference geometry. The per-step cost of the configuration benchmarked below ($N = 3$, $M = 4$, $H = 16$, $K = 3$) is 357 multiply–adds and 16 hyperbolic tangents, an estimated 0.7 μs at 1 GHz (Section 9), and its state is 48 numbers; Table 6 compares it with the alternatives. By contrast, a single coefficient tabulated over d arguments at g points per axis stores g^d entries (a coarse $g = 20$ grid over five arguments is 3.2×10^6 entries), and each history or rate argument multiplies that again.

6 Verification on a synthetic unsteady airfoil

Before turning to measured data, the structure was verified on a synthetic generator with known dynamics, mapping (α, flap, q) to (C_D, C_L, C_m): lift and moment respond to a first-order-lagged effective angle of attack, lift saturates through a tanh, drag follows a quadratic polar and the moment includes pitch-rate damping; the excitation is a random sequence of smoothed steps, not a periodic motion. A memoryless MLP given the same three inputs, including the pitch rate, reaches a held-out standardized error of 1.3×10^{-1}; the biquad-Wiener model reaches 2.0×10^{-3} with only 60 more parameters (Table 2). Driven through a sinusoidal α sweep with flap and rate held at zero, the full model traces the C_L–α and C_m–α hysteresis loops (Fig. 2), while the memoryless model collapses each to a curve. In this generator the lagged-α contribution is only about 4% of the variance of C_m, the rest being pitch-rate damping; the explicit linear path and input standardization are what allow that small history-dependent component to be recovered next to the large rate term. In closed-loop use in the simulator, with an α sweep and an airspeed step from 40 to 60 m/s, all dimensional loads scale by $(60/40)^2$ and the pitching moment equals $\bar{q}Sc\,C_m$ to machine precision. This case verifies the structure, the gradients and the runtime; it is not aerodynamic validation, which is the subject of the next two sections.

Table 2: Synthetic unsteady airfoil ($N = 3$ inputs, $K = 3$ coefficients, 4000 samples): held-out standardized MSE, averaged over the three coefficients, on the final 20% of the record.

Configuration	Biquads	Trainable params	Held-out MSE
Memoryless MLP (inputs include q)	0	316	1.3×10^{-1}
Biquad-Wiener model	12	376	2.0×10^{-3}

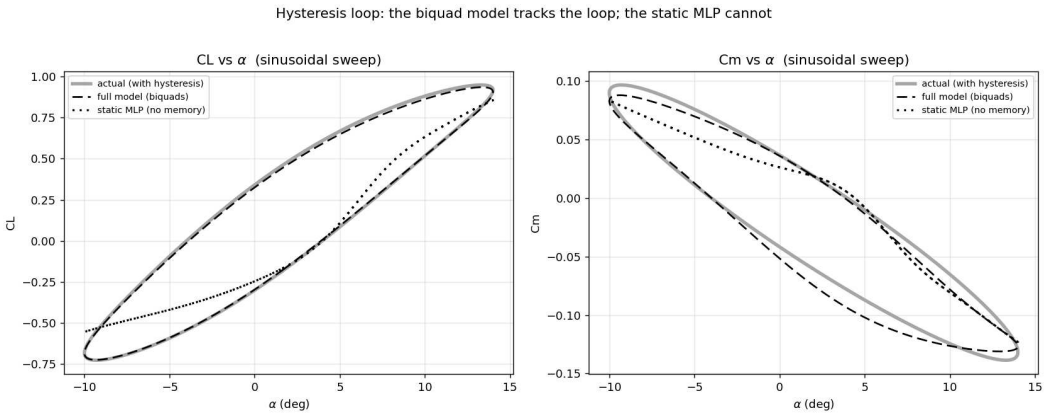


Figure 2: Synthetic verification: predicted and true hysteresis on a sinusoidal α sweep (flap and pitch rate zero, so the loop is carried entirely by the lagged effective incidence). The true response (thick grey) is reproduced by the biquad-Wiener model (dashed), while the memoryless ablation (dotted) collapses each loop to a single-valued curve.

7 Benchmark on measured oscillating-airfoil data

7.1 Data

The Ohio State University (OSU) campaign for the U.S. National Renewable Energy Laboratory measured the S809, S801 and NACA 4415 sections [2, 3, 4] in a 3×5 ft tunnel: 18-in-chord ($c = 0.457$ m) models pitched about the quarter chord by a cam, with the loads integrated from surface pressures. For each airfoil the clean-surface data comprise two amplitudes ($\pm 5.5°$, $\pm 10°$), three oscillator frequencies (about 0.6, 1.2 and 1.8 Hz), four Reynolds numbers (0.75 to 1.5 million) and three mean angles (8°, 14°, 20°): 72 runs per airfoil and 216 in total, with k from 0.017 to 0.11 and α from $-2°$ to 31°, spanning attached flow to deep dynamic stall. Each run holds 120 unaveraged samples (about 33 per cycle) of α, C_l, the pressure drag $C_{d,p}$ and C_m; the static polars of the same campaign are used by the table and semi-empirical models. All models are driven by the same smooth kinematics, a six-harmonic Fourier fit of the measured incidence (residual 0.15° RMS, the level of its 0.1° quantization), which also supplies $\dot{\alpha}$ and the pre-roll. Because the runs are not phase-averaged, their cycle-to-cycle scatter sets a *repeatability floor*, the RMS deviation of each run's samples from its own phase average, below which no deterministic model can go (Table 3).

7.2 Compared models

All models receive the same information: the kinematics $\alpha(t)$ and $\hat{q}(t) = \dot{\alpha}c/(2U)$, the airspeed U and, for those that use one, the measured static polar at the nearest Reynolds number. Recurrent and filter-based models are advanced on the same $\Delta t = 10$ ms (100 Hz) grid and scored at the

measurement instants. The alternatives are:

- **Quasi-static table**: linear interpolation of the static polar at $\alpha(t)$, which is current practice in most simulators.
- **Table in** $(\alpha, \hat{q})$: a rate-augmented table, fitted to the unsteady training runs by least squares on a bilinear grid (2° in α, seven levels of $\hat{q}$) with a second-difference smoothness penalty that fills cells the data do not reach; this is the "dynamic derivative" extension of a table, with 462 entries for three coefficients.
- **RFA, fixed lags**: a Roger-type rational-function approximation in the time domain [19, 20]: each coefficient is a linear combination of α, $\hat{q}$, U and first-order aerodynamic lags of α and $\hat{q}$ with lag roots $b_i \in \{0.0455, 0.1, 0.3, 0.6\}$ per semichord (Jones's two roots among them), fixed in convective time and therefore airspeed-scheduled; residues by ridge least squares.
- **HGM**: the Hansen–Gaunaa–Madsen state-space Beddoes–Leishman model [17] (lift and drag; it has no moment equation), driven by the static polar through a Kirchhoff separation point, with Jones's attached-flow constants, a pressure lag T_p and a boundary-layer lag T_f. It is run with the published constants ($T_p = 1.5$, $T_f = 6$) and, separately, with T_p and T_f *calibrated* on the training runs of each fold by Nelder–Mead minimization of the standardized error.
- **Goman–Khrabrov** [8]: the separation point obeys $\tau_1 \dot{x} + x = f_{\mathrm{st}}(\alpha - \tau_2 \dot{\alpha})$ and $C_l = C_{l\alpha}(\alpha - \alpha_0)\big((1 + \sqrt{x})/2\big)^2$, with τ_1 and τ_2 (in convective units) calibrated on the training runs (lift only).
- **Memoryless MLP**: the proposed model with the filters removed, i.e. the MLP and linear path on the instantaneous $(\alpha, \hat{q}, U)$.
- **Learned-pole RFA**: the proposed filter bank with a linear read-out instead of the MLP.
- **Tapped-delay MLP** (time-delay network): an MLP fed with 16 past values of the inputs spanning 0.6 s, i.e. a memory of fixed, finite length.
- **GRU** and **LSTM** [31, 30], the two standard gated recurrent networks (Section 2), with an 8-dimensional state (parameter counts comparable to the proposed model), and a GRU with a 32-dimensional state (ten times more parameters), each with a linear output layer and the same linear input term.
- **Proposed** biquad-Wiener model with $M = 4$ biquads per input and $H = 16$ hidden units (376 parameters); as sensitivity variants, a larger model ($M = 6$, $H = 32$) and a model with airspeed-scheduled poles, $r_f(U) = r_f^{U/U_{\mathrm{ref}}}$ and $\theta_f(U) = \theta_f U/U_{\mathrm{ref}}$, which fixes the filter time constants in convective rather than physical time and keeps every filter stable.

Every learned model is trained with the procedure of Section 4 (2000 epochs, early stopping on a validation subset of about one sixth of the training runs), with identical inputs, targets, splits and budget. The learned models are trained with three random seeds; the sensitivity variants, the learned-pole RFA and the 32-unit GRU with one. No model's hyperparameters were tuned independently: all learned models share the optimizer, learning rate (10^{-2}; 5×10^{-3} for the 32-unit GRU), budget and stopping rule. This keeps the comparison controlled, but equal budgets do not make the models equally well tuned, and the ranking among the learned models, including the proposed one, could change under per-model tuning.

Implementation details. All 72 clean runs of each OSU airfoil are used; every record was checked to hold its 120 samples, and none is missing or excluded. The kinematic fit refines the oscillator frequency within ±5% of the nominal value (grid search, then golden-section search) before the six-harmonic least-squares fit. Inputs and outputs are standardized with the means and standard deviations of the training runs of each fold. Static polars are interpolated linearly in α and held constant beyond their range; the lift slope and zero-lift angle are fitted over $-4^\circ \le \alpha \le 6^\circ$, where the Kirchhoff separation point is set to one, and it is forced to decrease monotonically outside that range. The semi-empirical time constants are calibrated in logarithmic coordinates, without bounds, by Nelder–Mead from the published values and from one perturbed start, keeping the better of the two (at most 120 iterations each). The Glasgow selection rules and smoothing (Savitzky–Golay, window 9, cubic) are given in Section 8.

7.3 Protocols and metrics

The models are tested by cross-validation on whole flow conditions: the runs are split into groups ("folds"), each group is predicted by models trained on all the other groups, and the process is repeated so that every run is predicted exactly once without having been used for training. Two groupings are used:

- **leave-one-Reynolds-number-out** (4 folds): generalization to an unseen Reynolds number, and hence airspeed, at otherwise familiar motions;
- **leave-one-frequency-out** (3 folds): generalization to an unseen oscillator frequency, i.e. to unseen reduced frequencies, the defining parameter of unsteady aerodynamics. Holding out the lowest or highest frequency makes this an extrapolation test.

The primary metric is the pooled root-mean-square error (RMSE) of each coefficient over all test samples, reported as mean ± standard deviation over seeds. For paired comparisons, the per-run C_l RMSE of the proposed model is compared with that of every alternative over the 216 runs with the Wilcoxon signed-rank test, a standard paired test of whether one model's per-run errors are systematically lower than another's, which does not assume normally distributed errors. We report the fraction of runs in which the proposed model is more accurate and the median ratio of the per-run errors as effect sizes. The p-values are not adjusted for multiple comparisons; with Holm's correction over the 14 comparisons of a protocol, every $p < 10^{-4}$ remains significant at the 5% level; the few larger p-values in Table 5 are those whose significance depends on the correction.

7.4 Results

Figures 3 and 4 give the held-out errors of the two protocols and Table 5 the paired comparisons over all 216 runs. The quasi-static table and both linear models (the fixed- and learned-lag RFAs) have lift errors 2.2–3.4 times the best: a table cannot represent the loop, and a linear model cannot represent the stall break. The semi-empirical models close much of that gap once calibrated (HGM with its published constants is clearly worse), and calibrated Goman–Khrabrov is the best physics-based lift model on all three airfoils. For unseen Reynolds numbers the proposed model is more accurate than all of these, with a lower per-run lift error than calibrated HGM in 78% of runs and than calibrated Goman–Khrabrov in 72% ($p < 10^{-4}$), and than the tapped-delay network in 82%. It is only marginally more accurate than the memoryless MLP and the table in $(\alpha, \hat{q})$ (54–59% of runs), and the gated networks are better than it in 68–76% of runs, with a median error 9–12% lower; they are the most accurate models in lift on every airfoil, within 1.2–1.4 times the repeatability floor (Table 3). For unseen reduced frequencies the proposed model generalizes

less well: its advantage over calibrated HGM is small (60% of runs), it is on a par with calibrated Goman–Khrabrov (51%, $p = 0.31$), the memoryless MLP and the $(\alpha, \hat{q})$ table are more accurate than it in 76% and 64% of runs, and the gated networks in 95–98%, with a median error 29–35% lower. The errors of all models grow from 8° to 20° mean incidence (Table 4), most for the semi-empirical models, whose single separation lag describes the vortex-shedding and reattachment phases of deep stall least well (Fig. 5).

That a memoryless map of $(\alpha, \hat{q}, U)$ can match or beat the proposed model here is a property of the test rather than evidence against memory: for periodic motion $(\alpha, \hat{q})$ fixes the phase of the cycle, so any periodic loop is a function of the instantaneous state. The oscillation benchmark alone cannot separate models that carry a state from models that see the rate; this is the purpose of the ramp test of Section 8.

The larger variant of the proposed model is worse on almost every airfoil, presumably overfitting the 45–50 training runs of a fold. The airspeed-scheduled poles do not help across Reynolds numbers, but are slightly more accurate than the base model for unseen frequencies (60% of runs). The proposed model's seed-to-seed spread is larger than that of the gated networks, particularly for unseen frequencies (Fig. 4). A trained model exported to the runtime coefficient format and streamed through the runtime recursion reproduces the batch predictions to 5×10^{-15}, so the benchmarked model is exactly the deployed one.

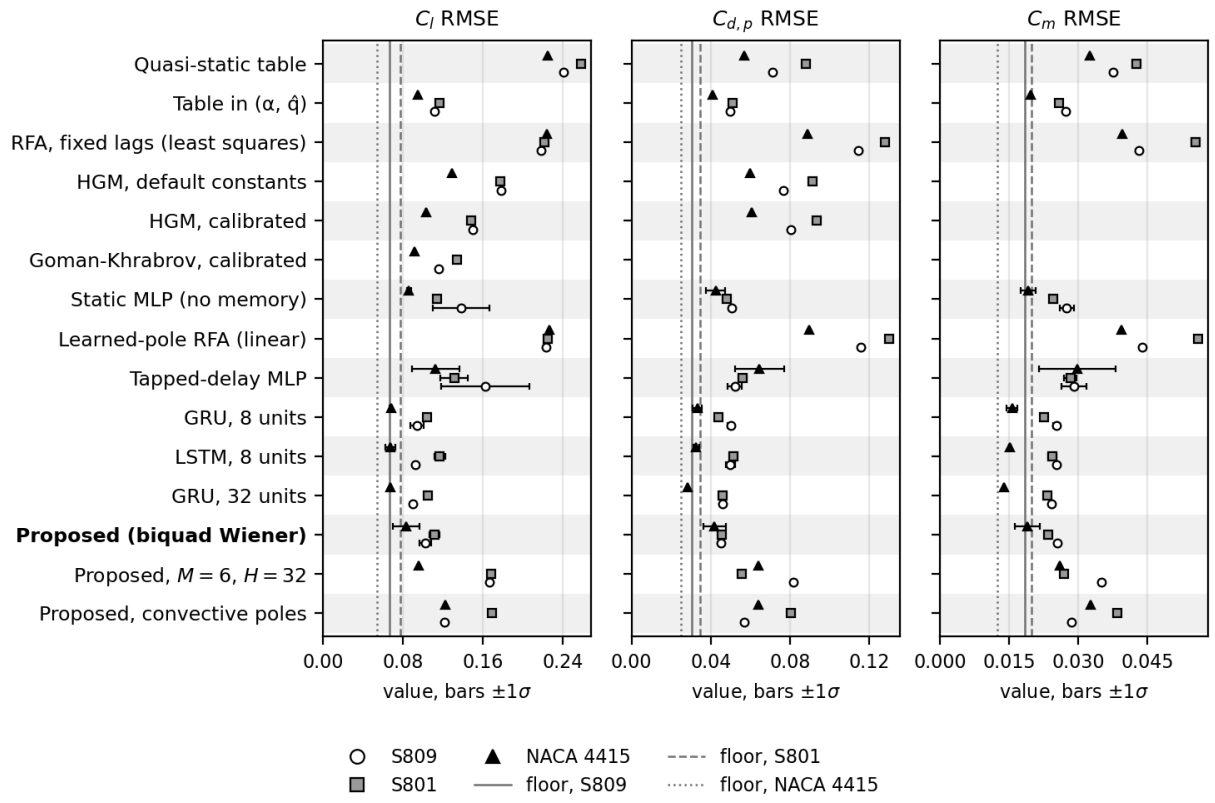


Figure 3: OSU benchmark, leave-one-Reynolds-number-out: held-out RMSE of each model (rows) for each coefficient (panels) and airfoil (markers). Horizontal bars: standard deviation over three training seeds (learned models; the variants and the 32-unit GRU have one seed). Vertical lines: repeatability floor of each airfoil. HGM predicts lift and drag, Goman–Khrabrov lift only.

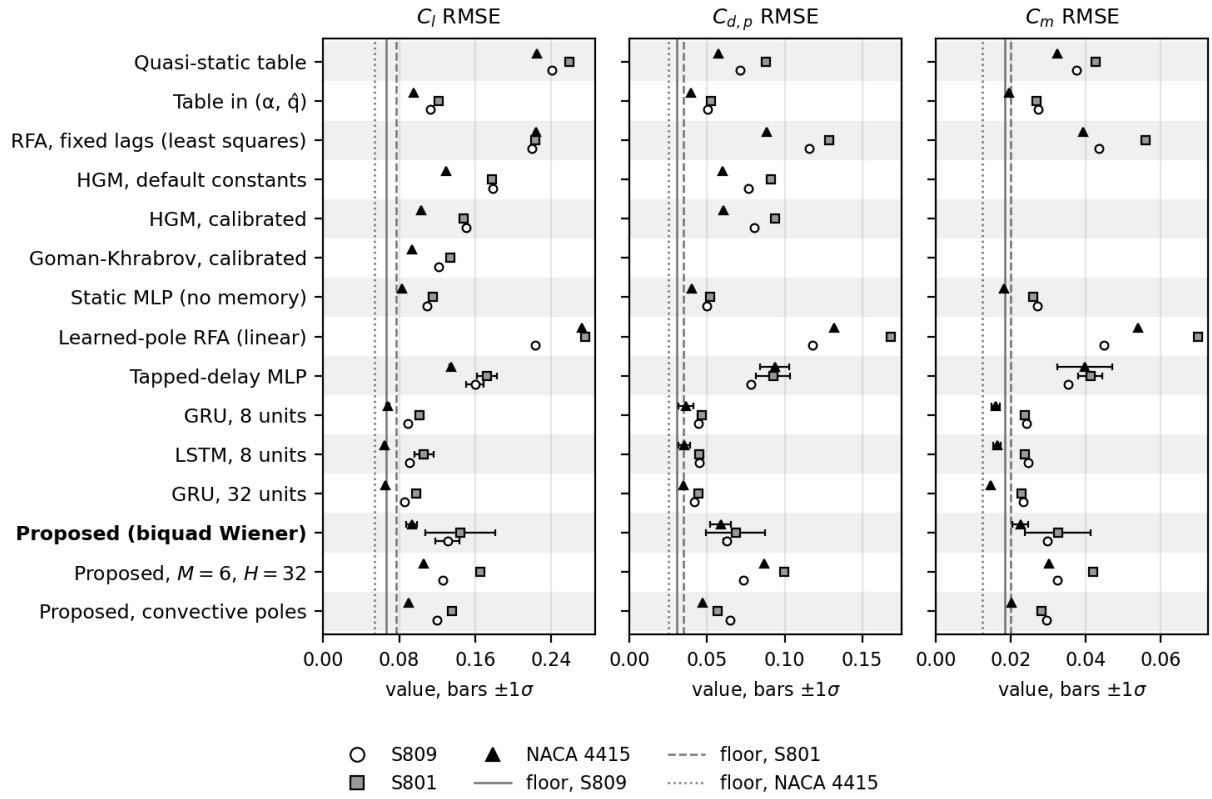


Figure 4: OSU benchmark, leave-one-frequency-out (unseen reduced frequency): held-out RMSE, as in Fig. 3.

Table 3: Repeatability floor: RMS cycle-to-cycle scatter of the measurements about each run's phase average.

Airfoil	C_l	$C_{d,p}$	C_m
S809	0.066	0.0302	0.018
S801	0.077	0.0348	0.020
NACA 4415	0.054	0.0251	0.012

Table 4: Lift RMSE by mean incidence, Reynolds protocol, averaged over the three airfoils.

Model	8°	14°	20°
Quasi-static table	0.153	0.257	0.292
Table in $(\alpha, \hat{q})$	0.062	0.111	0.136
RFA, fixed lags (least squares)	0.266	0.179	0.211
HGM, default constants	0.080	0.154	0.219
HGM, calibrated	0.076	0.112	0.188
Goman–Khrabrov, calibrated	0.060	0.108	0.153
Static MLP (no memory)	0.077	0.114	0.138
Learned-pole RFA (linear)	0.271	0.182	0.213
Tapped-delay MLP	0.118	0.137	0.148
GRU, 8 units	0.048	0.089	0.115
LSTM, 8 units	0.049	0.093	0.120
GRU, 32 units	**0.045**	**0.087**	**0.115**
Proposed (biquad Wiener)	0.060	0.100	0.125
Proposed, M=6, H=32	0.124	0.147	0.157
Proposed, convective poles	0.112	0.135	0.160

Table 5: Paired comparison of per-run lift RMSE, all 216 runs: fraction of runs in which the proposed model is *more* accurate than the alternative ("better"), median ratio of the proposed model's per-run error to the alternative's ("ratio", below one when the proposed model is more accurate), and unadjusted Wilcoxon signed-rank p-value.

	Reynolds protocol			Frequency protocol		
Alternative	better	ratio	p	better	ratio	p
Quasi-static table	98%	0.44	$< 10^{-4}$	88%	0.53	$< 10^{-4}$
Table in $(\alpha, \hat{q})$	59%	0.95	$< 10^{-3}$	36%	1.10	$< 10^{-4}$
RFA, fixed lags (least squares)	100%	0.44	$< 10^{-4}$	95%	0.53	$< 10^{-4}$
HGM, default constants	85%	0.65	$< 10^{-4}$	73%	0.77	$< 10^{-4}$
HGM, calibrated	78%	0.80	$< 10^{-4}$	60%	0.94	0.008
Goman–Khrabrov, calibrated	72%	0.85	$< 10^{-4}$	51%	0.99	0.308
Static MLP (no memory)	54%	0.98	0.022	24%	1.17	$< 10^{-4}$
Learned-pole RFA (linear)	100%	0.44	$< 10^{-4}$	98%	0.46	$< 10^{-4}$
Tapped-delay MLP	82%	0.76	$< 10^{-4}$	80%	0.87	$< 10^{-4}$
GRU, 8 units	24%	1.14	$< 10^{-4}$	2%	1.41	$< 10^{-4}$
LSTM, 8 units	32%	1.10	$< 10^{-4}$	5%	1.47	$< 10^{-4}$
GRU, 32 units	30%	1.14	$< 10^{-4}$	3%	1.53	$< 10^{-4}$
Proposed, M=6, H=32	75%	0.75	$< 10^{-4}$	53%	0.98	0.003
Proposed, convective poles	79%	0.85	$< 10^{-4}$	40%	1.04	0.003

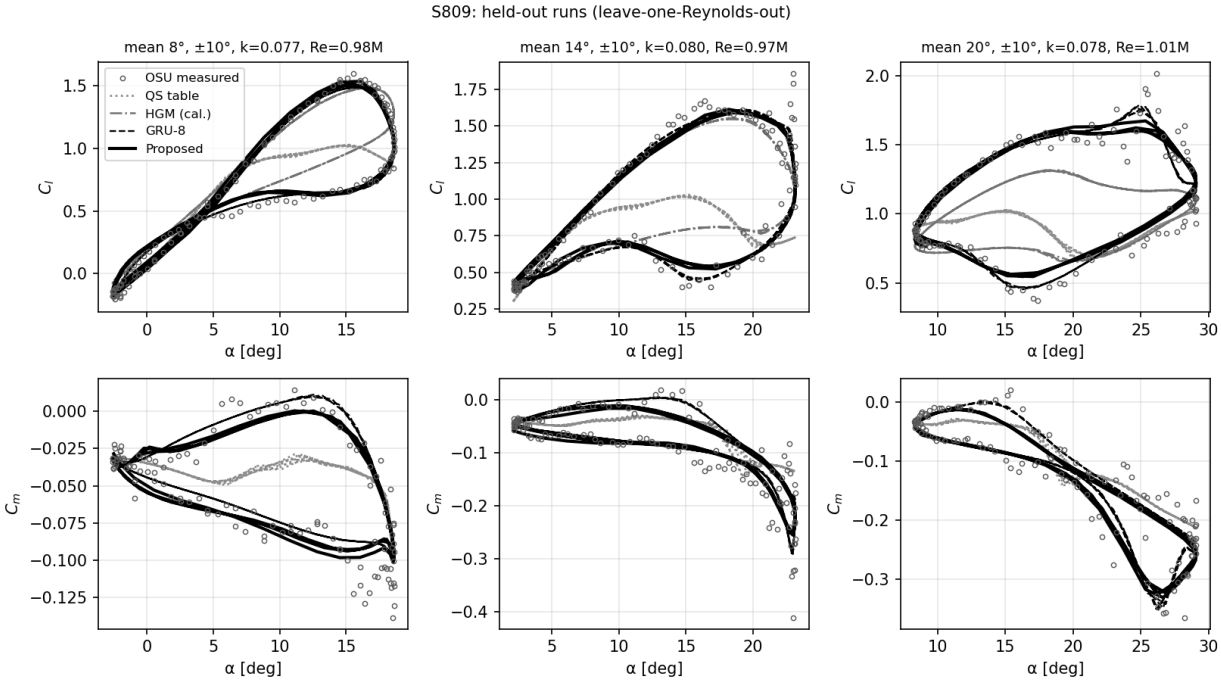


Figure 5: Held-out S809 runs ($\pm 10°$, $k \approx 0.08$, Re $\approx 1.0 \times 10^6$, a Reynolds number excluded from training) at the three mean angles: measured lift and moment (circles, 3.6 cycles, not averaged), quasi-static table (dotted), calibrated HGM (dash-dot; no moment model), 8-unit GRU (dashed) and proposed model (thick solid).

8 Motion transfer: from oscillations to ramps

8.1 Why a second test is needed

For periodic motion, the pair $(\alpha, \hat{q})$ determines the phase of the cycle uniquely. A static map of $(\alpha, \hat{q})$, such as a rate-augmented table or a memoryless network given the pitch rate, can therefore represent any periodic hysteresis loop without carrying a state, and on oscillating-airfoil data it can look deceptively good. It cannot represent a response that depends on history in any other way: after a ramp stops, $\hat{q} = 0$ and α is constant, yet the loads keep evolving as the dynamic-stall vortex convects and the flow relaxes to its static state. A simulator that flies arbitrary manoeuvres needs exactly that capability. The second test therefore trains every model on sinusoidal oscillations only and evaluates it on measured constant-pitch-rate ramps, a motion family absent from training.

8.2 Data and protocol

The University of Glasgow dynamic-stall database [5] (CC BY 4.0) contains ensemble-averaged surface-pressure histories of 2-D models of $c = 0.55$ m pitching about the quarter chord in a 5×7 ft low-speed tunnel, mostly at $U \approx 42$ m/s ($\mathrm{Re} \approx 1.5 \times 10^6$), under sinusoidal oscillations (averaged over ten cycles) and ramp-up and ramp-down motions at constant pitch rate (averaged over five hold–ramp–hold cycles). Three sections are used: the NACA 0012 and NACA 0015 and the cambered NACA 23012B. The normal-force and quarter-chord moment coefficients C_n and C_m are integrated from the 30 pressure taps with the trapezoidal rule, closing both surfaces to the trailing-edge pressure (they agree with the coefficients exported by the database's own viewer to about 0.01 in C_n).

Training uses the sinusoidal runs with mean incidence below 30° and a period of at most 1.25 s (reduced frequencies 0.04–0.50, amplitudes up to ±33°; 175–244 runs per section); testing uses every ramp-up and ramp-down of at most 2.3 s with reduced pitch rate $r = \dot{\alpha}c/(2U) \geq 0.005$ (up to 0.19; 91–147 runs per section). The quasi-static ramps ($r < 0.001$) of the same section provide the static polar of the table and semi-empirical models. Sinusoids are driven by their Fourier fit with a one-period pre-roll; ramps by Savitzky–Golay-smoothed (window 9, cubic) measured kinematics with a 0.2 s hold at the initial incidence. All models run on a 4 ms (250 Hz) grid (a fast ramp reaches 30° in under 0.1 s), with the inputs $(\alpha, \hat{q}, U)$ and outputs (C_n, C_m); the semi-empirical models predict C_n from the C_n polar. The model set and training procedure are those of Section 7, with three seeds.

Most ramps end at 30–40°, beyond the largest incidence of the training sinusoids of some sections. Errors are therefore reported twice: over the whole record, and over the portion of each ramp before α first leaves the incidence range covered in training ("in envelope"), which separates motion transfer from extrapolation in α.

8.3 Results

Figure 6 gives the ramp errors of the models trained on oscillations only. The picture differs from the oscillation benchmark in the way the argument of Section 8 predicts. The static maps of $(\alpha, \hat{q})$, namely the rate-augmented table and the memoryless MLP, which matched or beat the proposed model on oscillations, are now clearly worse than it within the envelope on all three airfoils (for the NACA 0012, by more than a factor of two), and the tapped-delay network, whose memory is limited to its window, falls between them. The proposed model thus demonstrably uses its state. It does not have the lowest time-history error, however: the calibrated Goman–Khrabrov and HGM models have a lower C_n RMSE on all three sections, and the gated networks on the two symmetric

ones, while on the cambered NACA 23012B the proposed model is better than the gated networks. Outside the training envelope all learned models degrade, as expected of any data-driven model extrapolated in incidence.

Fig. 7 shows why the RMSE ranking does not tell the whole story. HGM, which has no vortex-lift term, saturates at the static-like level and misses the dynamic-stall overshoot of C_n that follows a fast ramp; the proposed model and the GRU reproduce the overshoot, but a small error in its timing costs them heavily in RMSE. On the peak normal force of each ramp-up (Fig. 6, right), the maximum load that a structural or handling-qualities analysis would use, the proposed model is better than calibrated HGM and calibrated Goman–Khrabrov on all three airfoils; the LSTM is best on the symmetric sections.

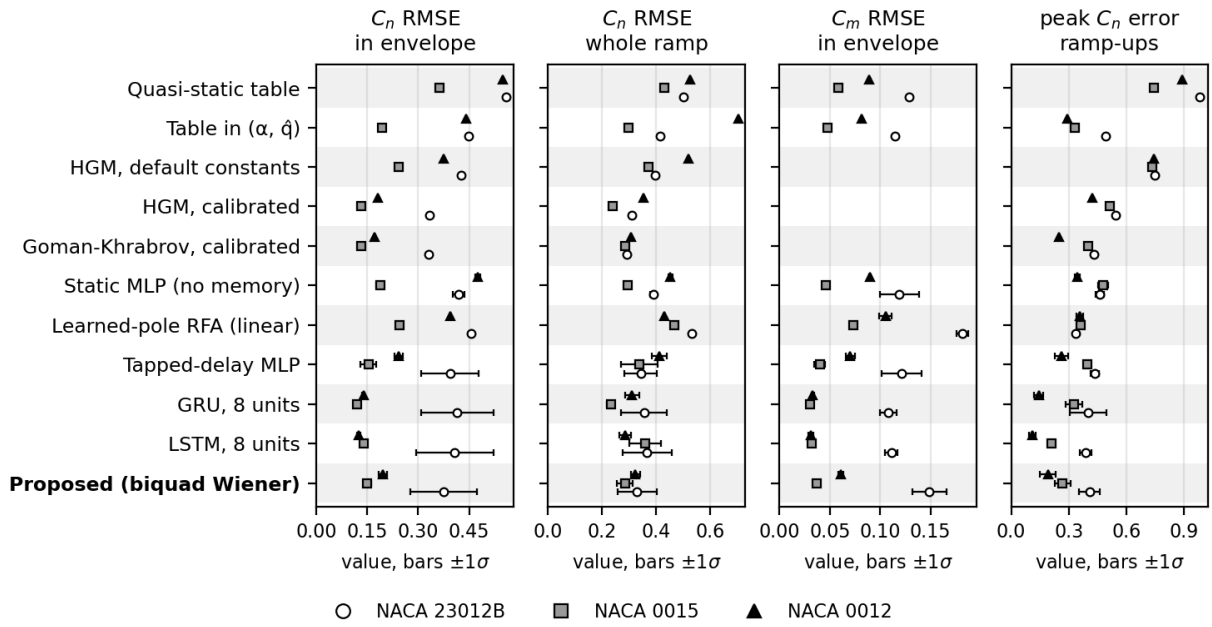


Figure 6: Glasgow ramps, models trained on sinusoids only: RMSE of C_n within the training incidence envelope and over the whole record, RMSE of C_m within the envelope, and mean absolute error of the peak normal force of the ramp-ups, for each airfoil (markers). Horizontal bars: standard deviation over three training seeds. The data are ensemble-averaged, so no repeatability floor is shown.

9 Learned time scales and cost

A filter pole of radius r at step Δt has the convective time constant

$$\hat{\tau} = -\frac{2U\Delta t}{c \ln r} \tag{7}$$

in semichords of travel, the unit of the classical lags (Jones: about 22 and 3.3; HGM: $T_p = 1.5$, $T_f = 6$). One might hope to read physical time constants off the trained model; the benchmark does not support that. Comparing, for every α- and $\hat{q}$-channel filter of every trained model, the convective time constant after training with its value at initialization, the median filter moves by less than 20%, and only 7–15% move by more than a factor of two. On records of a few cycles

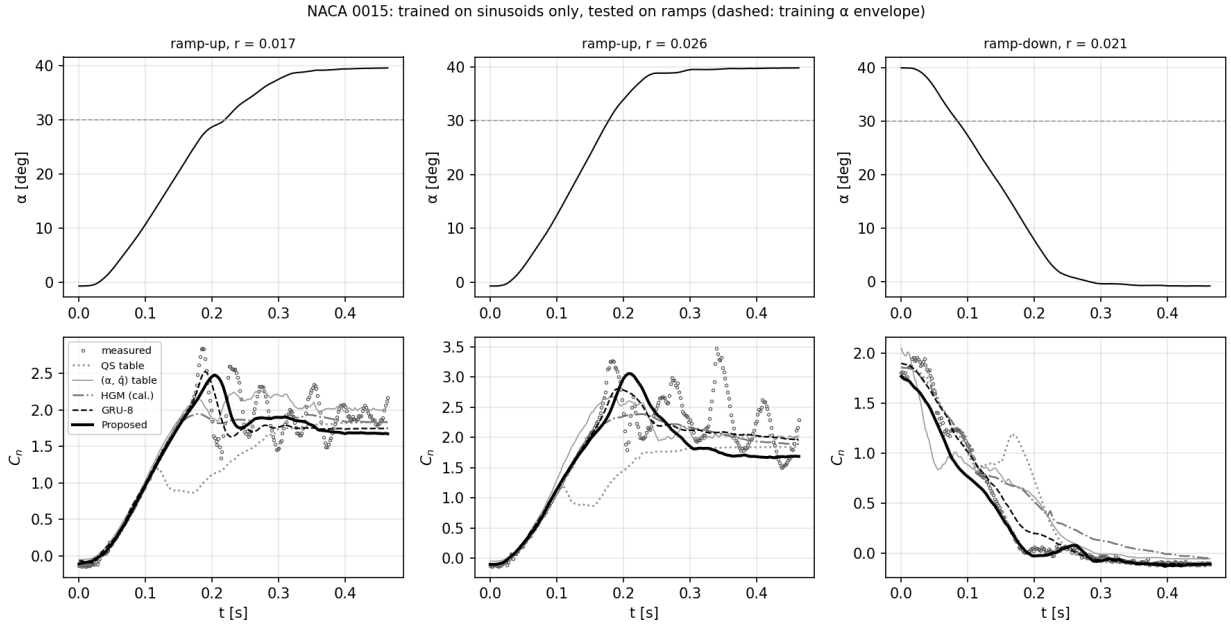


Figure 7: NACA 0015, models trained on sinusoids only: two measured ramp-ups and a ramp-down (top: incidence, dashed lines the training envelope; bottom: normal force): measured (circles), quasi-static table (dotted), $(\alpha, \hat{q})$ table (thin grey), calibrated HGM (dash-dot), 8-unit GRU (dashed) and proposed model (thick solid).

Table 6: Parameters (S809), state size, estimated time per step at 1 GHz (from operation counts, $N = 3$ inputs, $K = 3$ outputs; see text), and whether the internal recurrence is stable for every parameter value.

Model	Class	Params	State	ns/step	Stable by constr.
Quasi-static table	Table	0	–	45	–
Table in $(\alpha, \hat{q})$	Table	462	–	75	–
RFA, fixed lags (least squares)	Linear ROM	36	8	52	yes
HGM, default constants	Semi-empirical	0	4	140	yes
HGM, calibrated	Semi-empirical	2	4	140	yes
Goman–Khrabrov, calibrated	Semi-empirical	2	1	43	yes
Static MLP (no memory)	Learned	124	–	425	–
Learned-pole RFA (linear)	Learned	108	48	105	yes
Tapped-delay MLP	Learned	844	183	1145	yes (FIR)
GRU, 8 units	Learned	348	8	777	no
LSTM, 8 units	Learned	452	16	1185	no
GRU, 32 units	Learned	3660	32	5385	no
Proposed (biquad Wiener)	Learned	376	48	677	yes
Proposed, M=6, H=32	Learned	902	72	1507	yes
Proposed, convective poles	Learned	376	48	677	yes

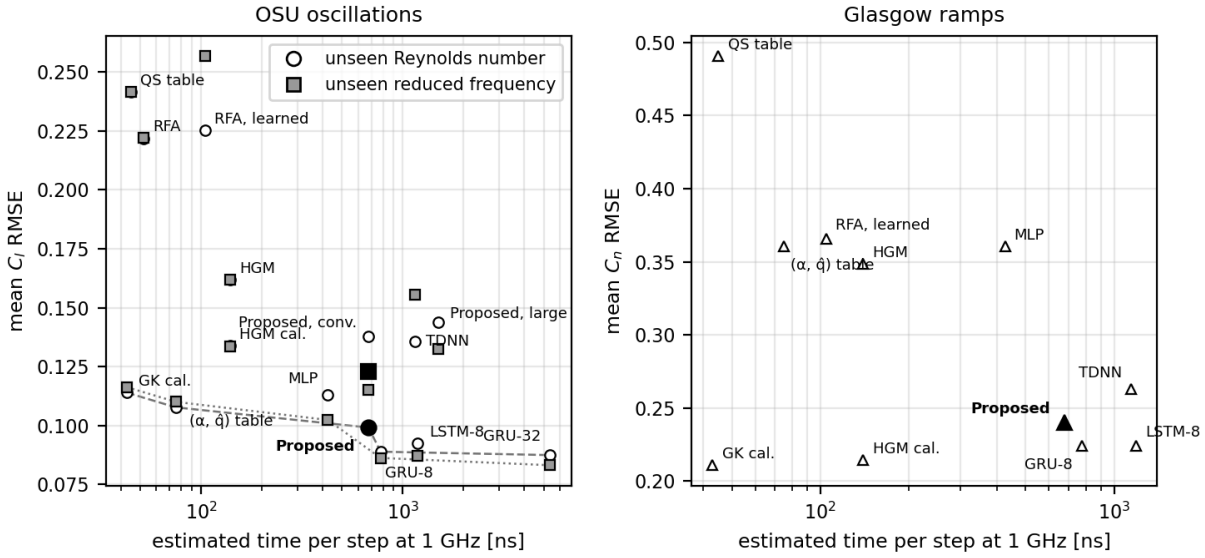


Figure 8: Accuracy versus estimated cost. Left: OSU held-out lift RMSE averaged over the three airfoils, for unseen Reynolds numbers (circles) and unseen reduced frequencies (squares); right: Glasgow ramp C_n RMSE within the training envelope, averaged over the three airfoils (the semi-empirical models predict C_n only). Filled markers: proposed model. Lines: Pareto fronts (dashed, circles; dotted, squares).

sampled at ≈ 33 points per cycle, the loss is only weakly sensitive to individual pole locations, and the optimizer adapts the numerators and the read-out instead. In practice the filter bank acts as a stable, well-spread basis of lags, much like the fixed lag roots of an RFA, rather than as an identifier of physical time constants; interpreting individual learned poles physically would require richer excitation (e.g. step or ramp responses) or regularization toward a physical prior.

Table 6 summarizes the cost of each model as a single number: the time per step estimated from its operation counts for a scalar processor at 1 GHz, assuming one cycle per multiply–add, 20 cycles per hyperbolic tangent, sigmoid or square root, 15 cycles per one-dimensional table look-up (interval search and interpolation) and 25 per bilinear look-up. This is an estimate, not a measurement: at this size memory access and the implementation of the transcendental functions matter, so the figures indicate relative cost and order of magnitude only. The proposed model needs about 0.7 μs per step, similar to an 8-unit GRU, less than an LSTM or the tapped-delay network, and about one eighth of the 32-unit GRU; the tables and semi-empirical models are an order of magnitude cheaper still.

Figure 8 places every model in the accuracy–cost plane. For unseen Reynolds numbers the proposed model lies on the Pareto front: nothing cheaper is more accurate, and the gated networks that are more accurate cost 15% to eight times more. For unseen reduced frequencies it is dominated by the cheaper memoryless MLP. On the ramps, the calibrated Goman–Khrabrov and HGM models are both the cheapest and, in time-history error, the most accurate on average, while the proposed model sits with the gated networks at a lower cost. Of the models with a learned recurrent state, the proposed one is the only one whose recurrence is stable for every parameter value.

10 Discussion

Summary of the results. On measured unsteady loads of six two-dimensional airfoil sections, in two independent facilities, the biquad-Wiener surrogate is consistently and significantly more accurate than the methods it is meant to replace in a real-time simulator: quasi-static tables and linear rational-function models. Against the calibrated semi-empirical models the answer depends on the test and on the quantity of interest: on oscillating airfoils at unseen Reynolds numbers the surrogate has the lower time-history error, at unseen reduced frequencies the two are on a par, and on ramps the calibrated models have the lower time-history error while the surrogate predicts the dynamic-stall peak loads better. On periodic motion a rate-augmented table and a memoryless network match it, and beat it at unseen frequencies; unlike them, it transfers to motions it has not seen. It is not the most accurate model tested: small gated recurrent networks have a lower error on most airfoils.

Practical value. The model offers a combination of properties: (i) it is more accurate than quasi-static tables and RFAs on all measured data, and than calibrated semi-empirical models in time-history error at unseen Reynolds numbers and in peak load on ramps; (ii) it is learned directly from time histories, without a static polar, separation-point construction or hand-tuned constants, and applies to coefficients for which no semi-empirical model exists (here the pitching moment, which the HGM and Goman–Khrabrov formulations do not predict); (iii) its internal recurrence is stable by construction, with a bounded, fixed-cost, allocation-free forward pass; and (iv) its lift error exceeds that of the best recurrent network by 7–24% for unseen Reynolds numbers and by 44–53% for unseen reduced frequencies, at a similar nominal cost. Where the last few percent of accuracy matter most, a gated network is the better choice on this evidence. Where a learned surrogate must be admitted into a closed-loop simulator, a recurrence whose stability can be checked by inspecting its poles, and a short inference kernel, are easier to verify than the nonlinear state dynamics of a gated network. As stated in Section 3, this is a property of the surrogate alone: it does not establish the stability of the closed loop with the vehicle and its flight controls, nor the acceptability of the model for simulator qualification, which remain to be shown by the usual means.

Extension to six-degree-of-freedom databases. The experiments concern two-dimensional sections with three inputs; whether the approach carries over to an aircraft database is untested and is the main open question. The architecture itself scales linearly. A typical database has about $N = 12$ inputs (α, β, Mach number, the three body rates, three to five control surfaces and a configuration variable) and six force and moment coefficients; with $M = 4$ filters per input and $H = 64$ hidden units give about 4 000 multiply–adds per step, whereas a table over the same arguments is infeasible. Only channels with unsteady dynamics need filters (Mach number or configuration can bypass the bank through the skip path), and the cross-coupling between coefficients enters only through the shared MLP. The open issues are elsewhere: the data needed to cover a coupled multi-dimensional envelope with time histories (flight test, dynamic wind-tunnel tests or unsteady CFD), extrapolation outside that envelope, and whether a single filter bank with poles fixed in seconds can serve an envelope spanning a wide range of airspeeds; in the present tests, scheduling the poles with airspeed did not help across Reynolds numbers.

Limitations. Besides the two-dimensional, low-Mach scope (Reynolds numbers of about 1–2 million), the OSU records are short and coarsely sampled, which limits what any model can identify and explains the weakly constrained poles; the model is therefore best understood as a stable, structured dynamical basis with a nonlinear read-out, not as an identification procedure for aerodynamic

time constants. All learned models were trained with an identical, fixed budget and without per-model tuning; on the Glasgow data the validation loss of every learned model was still decreasing at the end of the budget. As with any data-driven model, extrapolation in incidence or rate beyond the training envelope is unreliable (Fig. 6, whole-ramp panel) and must be guarded in a simulator by envelope checks or a fallback model.

11 Conclusion

A compact Wiener surrogate, a bank of learnable biquads with a recurrence stable for every parameter value, feeding a small MLP with an explicit linear path, was formulated for unsteady aerodynamic coefficients in real-time simulation and benchmarked on measured two-dimensional airfoil data: 216 oscillating-airfoil runs of three sections under condition-level cross-validation, and a transfer test from oscillations to ramps on three further sections, against tables, rational-function approximations, calibrated semi-empirical dynamic-stall models and memoryless, finite-memory and recurrent neural networks. The surrogate is significantly more accurate than quasi-static and linear models. Against calibrated semi-empirical models it has the lower time-history error at unseen Reynolds numbers, is on a par at unseen reduced frequencies, and predicts dynamic-stall peak loads better on ramps, where their time-history error is lower. Rate-augmented static maps match or beat it on periodic motion but, unlike it, fail to generalize from periodic to transient motion. Gated recurrent networks remain somewhat more accurate at similar nominal cost, without a stability guarantee on their internal dynamics.

Natural extensions for future work include measured execution times (mean and worst case) and memory footprints on desktop and embedded processors, with accuracy–cost trade-off comparisons; per-model hyperparameter tuning and longer training budgets; ablations of the individual architectural choices (linear term, skip inputs, learned versus fixed poles, second- versus first-order sections) on measured data; bootstrap confidence intervals for the paired comparisons; and an application to a full six-degree-of-freedom aircraft database in closed-loop simulation.

References

[1] Federal Aviation Administration, "Flight simulation training device qualification standards for extended envelope and adverse weather event training tasks," Final rule, 14 CFR Part 60, Amdt. No. 60-4, *Federal Register*, vol. 81, no. 61, pp. 18178–18388, 2016.

[2] R. Reuss Ramsay, M. J. Hoffmann, G. M. Gregorek, "Effects of grit roughness and pitch oscillations on the S809 airfoil," NREL/TP-442-7817, National Renewable Energy Laboratory, Golden, CO, 1995, doi:10.2172/205563.

[3] R. Reuss Ramsay, M. J. Hoffmann, G. M. Gregorek, "Effects of grit roughness and pitch oscillations on the S801 airfoil," NREL/TP-442-7818, National Renewable Energy Laboratory, Golden, CO, 1996, doi:10.2172/204225.

[4] M. J. Hoffmann, R. Reuss Ramsay, G. M. Gregorek, "Effects of grit roughness and pitch oscillations on the NACA 4415 airfoil," NREL/TP-442-7815, National Renewable Energy Laboratory, Golden, CO, 1996, doi:10.2172/266691.

[5] R. B. Green, M. Giuni, "Dynamic stall database R and D 1570-AM-01: Final report," University of Glasgow, 2017 [data collection, CC BY 4.0], doi:10.5525/gla.researchdata.464.

[6] V. Klein, E. A. Morelli, *Aircraft System Identification: Theory and Practice*, AIAA, Reston, VA, 2006, doi:10.2514/4.861505.

[7] M. Tobak, L. B. Schiff, "On the formulation of the aerodynamic characteristics in aircraft dynamics," NASA TR R-456, 1976.

[8] M. Goman, A. Khrabrov, "State-space representation of aerodynamic characteristics of an aircraft at high angles of attack," *Journal of Aircraft*, vol. 31, no. 5, pp. 1109–1115, 1994, doi:10.2514/3.46618.

[9] D. Fischenberg, "Identification of an unsteady aerodynamic stall model from flight test data," AIAA Paper 95-3438, 1995, doi:10.2514/6.1995-3438.

[10] D. I. Greenwell, "A review of unsteady aerodynamic modelling for flight dynamics of manoeuvrable aircraft," AIAA Paper 2004-5276, 2004, doi:10.2514/6.2004-5276.

[11] H. Wagner, "Über die Entstehung des dynamischen Auftriebes von Tragflügeln," *ZAMM*, vol. 5, no. 1, pp. 17–35, 1925, doi:10.1002/zamm.19250050103.

[12] T. Theodorsen, "General theory of aerodynamic instability and the mechanism of flutter," NACA Report 496, 1935.

[13] R. T. Jones, "The unsteady lift of a wing of finite aspect ratio," NACA Report 681, 1940.

[14] J. G. Leishman, *Principles of Helicopter Aerodynamics*, 2nd ed., Cambridge University Press, Cambridge, UK, 2006.

[15] W. J. McCroskey, "The phenomenon of dynamic stall," NASA TM-81264, 1981.

[16] J. G. Leishman, T. S. Beddoes, "A semi-empirical model for dynamic stall," *Journal of the American Helicopter Society*, vol. 34, no. 3, pp. 3–17, 1989, doi:10.4050/JAHS.34.3.3.

[17] M. H. Hansen, M. Gaunaa, H. Aa. Madsen, "A Beddoes–Leishman type dynamic stall model in state-space and indicial formulations," Risø-R-1354(EN), Risø National Laboratory, Roskilde, Denmark, 2004.

[18] S. Gupta, J. G. Leishman, "Dynamic stall modelling of the S809 aerofoil and comparison with experiments," *Wind Energy*, vol. 9, no. 6, pp. 521–547, 2006, doi:10.1002/we.200.

[19] K. L. Roger, "Airplane math modeling methods for active control design," in *Structural Aspects of Active Controls*, AGARD-CP-228, pp. 4.1–4.11, 1977.

[20] S. H. Tiffany, W. M. Adams, Jr., "Nonlinear programming extensions to rational function approximation methods for unsteady aerodynamic forces," NASA TP-2776, 1988.

[21] W. Silva, "Identification of nonlinear aeroelastic systems based on the Volterra theory: progress and opportunities," *Nonlinear Dynamics*, vol. 39, no. 1–2, pp. 25–62, 2005, doi:10.1007/s11071-005-1907-z.

[22] J. Kou, W. Zhang, "Data-driven modeling for unsteady aerodynamics and aeroelasticity," *Progress in Aerospace Sciences*, vol. 125, art. 100725, 2021, doi:10.1016/j.paerosci.2021.100725.

[23] J. Kou, W. Zhang, M. Yin, "Novel Wiener models with a time-delayed nonlinear block and their identification," *Nonlinear Dynamics*, vol. 85, no. 4, pp. 2389–2404, 2016, doi:10.1007/s11071-016-2833-y.

[24] M. Schoukens, K. Tiels, "Identification of block-oriented nonlinear systems starting from linear approximations: a survey," *Automatica*, vol. 85, pp. 272–292, 2017, doi:10.1016/j.automatica.2017.06.044.

[25] W. E. Faller, S. J. Schreck, M. W. Luttges, "Neural network prediction and control of three-dimensional unsteady separated flowfields," *Journal of Aircraft*, vol. 32, no. 6, pp. 1213–1220, 1995, doi:10.2514/3.46866.

[26] K. Li, J. Kou, W. Zhang, "Deep neural network for unsteady aerodynamic and aeroelastic modeling across multiple Mach numbers," *Nonlinear Dynamics*, vol. 96, no. 3, pp. 2157–2177, 2019, doi:10.1007/s11071-019-04915-9.

[27] R. Zahn, C. Breitsamter, "Airfoil buffet aerodynamics at plunge and pitch excitation based on long short-term memory neural network prediction," *CEAS Aeronautical Journal*, vol. 13, no. 1, pp. 45–55, 2022, doi:10.1007/s13272-021-00550-6.

[28] A. Mohamed, D. Wood, "Deep learning predictions of unsteady aerodynamic loads on an airfoil model pitched over the entire operating range," *Physics of Fluids*, vol. 35, no. 5, art. 053113, 2023, doi:10.1063/5.0139907.

[29] S. L. Brunton, B. R. Noack, P. Koumoutsakos, "Machine learning for fluid mechanics," *Annual Review of Fluid Mechanics*, vol. 52, pp. 477–508, 2020, doi:10.1146/annurev-fluid-010719-060214.

[30] S. Hochreiter, J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997, doi:10.1162/neco.1997.9.8.1735.

[31] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proc. EMNLP*, pp. 1724–1734, 2014, doi:10.3115/v1/D14-1179.

[32] S. Nercessian, "Neural parametric equalizer matching using differentiable biquads," in *Proc. 23rd Int. Conf. Digital Audio Effects (DAFx2020)*, pp. 265–272, 2020.

[33] S. Nercessian, A. Sarroff, K. J. Werner, "Lightweight and interpretable neural modeling of an audio distortion effect using hyperconditioned differentiable biquads," in *Proc. IEEE ICASSP*, pp. 890–894, 2021, doi:10.1109/ICASSP39728.2021.9413996.

[34] M. Forgione, D. Piga, "dynoNet: a neural network architecture for learning dynamical systems," *International Journal of Adaptive Control and Signal Processing*, vol. 35, no. 4, pp. 612–626, 2021, doi:10.1002/acs.3216.

[35] A. Gu, K. Goel, C. Ré, "Efficiently modeling long sequences with structured state spaces," in *International Conference on Learning Representations (ICLR)*, 2022.

[36] D. P. Kingma, J. Ba, "Adam: a method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, 2015, arXiv:1412.6980.

Declarations

Funding. No external funding was received for this work.

Competing interests. The author declares no competing interests.

Data availability. The OSU oscillating-airfoil data are publicly available from the U.S. National Renewable Energy Laboratory (NREL/TP-442-7815, -7817, -7818) [2, 3, 4]; the Glasgow dynamic-stall database is available under a CC BY 4.0 licence [5]. The synthetic verification data are generated by the equations of Section 6.

Code availability. The model is implemented in the HiroSim simulator (`https://hirosim.com`); the implementation is not publicly available. The inference procedure is fully specified by Eqs. (1)–(4), and the benchmark protocols and all baseline models by Sections 7–8.

Author contributions. The sole author conceived the method, implemented the prototype, designed and performed the experiments, and wrote the manuscript.

Use of AI tools. Anthropic's Claude assisted with the literature search and reference verification, with writing the benchmark scripts, and with drafting and editing the manuscript. The author reviewed all such contributions and is responsible for the entire content of the paper.