



Harnessing Instability Mechanisms in Airfoil Flow for Data-Driven Forecasting of Extreme Events

Benedikt Barthel*^① and Themistoklis P. Sapsis[†]^②

Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

<https://doi.org/10.2514/1.J062992>

For certain Reynolds numbers, airfoils are subject to sporadic high-amplitude fluctuations in the aerodynamic forces. These extreme excursions may be seen as prototypical of the kind of unsteady and intermittent dynamics relevant to the flow around airfoils and wings in a variety of real-world applications. Here we investigate the instability mechanisms at the heart of these extreme events, and how they may be harnessed for efficient data-driven forecasting. Through a wavelet and spectral analysis of the pressure and vorticity, we find that the extreme events arise due to the instability of a specific frequency component distinct from the vortex shedding mode. During these events, this extreme event frequency draws energy from the energetically dominant vortex shedding flow and undergoes an abrupt transfer of energy from small to large scales. We propose a preprocessing algorithm to extract this extreme event frequency from the surface pressure data, which in conjunction with an extreme event-tailored loss function, allows us to avoid the commonly used long short-term memory architecture in favor of a simple feed-forward network—a significant reduction in cost over the previous state-of-the-art. Our model requires only three pressure sensors, and it is robust to their location—showing promise for the use of our model in dynamically varying applications. Finally, we show that relying solely on the statistics of the pressure and drag data for optimal sensor placement fails to improve model prediction over uniform or random sensor placement.

Nomenclature

a	=	acquisition function
\mathcal{F}	=	Fourier transform operator
f	=	temporal frequency
I	=	mutual information
k_s	=	spatial wavenumber w.r.t s
P	=	airfoil surface pressure
\tilde{P}	=	filtered airfoil surface pressure
\hat{P}	=	wavelet transformed airfoil surface pressure
q	=	filtered drag coefficient
\hat{q}	=	model prediction of filtered drag coefficient
r	=	temporal correlation function
s	=	arc length along airfoil surface
$\mathbf{u}, \boldsymbol{\omega}, p$	=	fluid velocity, vorticity, and pressure
\mathcal{W}	=	wavelet transform operator
α^*	=	maximum adjusted area under precision-recall curve
Γ	=	vector of γ and its derivative
γ	=	wavelet transformed airfoil surface pressure evaluated at extreme event frequency
σ_q	=	model uncertainty in prediction of filtered drag coefficient
τ	=	lead time for prediction of q
χ	=	extreme event rate
$\tilde{\Omega}$	=	wavelet transformed vorticity
$\hat{\Omega}$	=	Fourier and wavelet transformed vorticity

Subscripts

EE	=	conditioned on extreme events
e	=	evaluated at extreme event frequency
iu	=	integrated uncertainty

j	=	pressure sensor index
ow	=	output weighted
s	=	evaluated at airfoil surface
v	=	evaluated at vortex shedding frequency

I. Introduction

MANY engineering systems are subject to rare high-amplitude fluctuations commonly referred to as *extreme events* [1]. While here we focus primarily on fluid–structure interactions, such events occur in a wide variety of systems ranging from climate systems to stock markets. Although rare, events like gusts or rogue waves have a disproportionate effect on the fatigue life of aircraft, naval vessels, or marine infrastructure. Due to their rare nature, the prediction of these events is inherently challenging, especially as they often occur in complex systems where the physical mechanisms are unknown [2]. Various authors have ventured to address this problem through strategies such as optimal sampling [3–5] or training strategies which preferentially amplify rare events [6–8].

Two-dimensional airfoil flow is one of the canonical test cases for the dynamics of fluid–structure interaction and has been the subject of extensive study for decades [9,10]. However, the advent of machine learning and data-driven techniques has unlocked new lenses to study this and other classical problems in the field of fluid dynamics [11–13]. In particular, several authors, including Rudy and Sapsis [8], Gomez et al. [14], and Maulik et al. [15], have proposed neural network models for the reconstruction of the flow from sparse measurements of a range of flow observables. However, data-driven predictions based on the airfoil surface pressure have been of particular interest due to their practical measurability. In particular, the development of effective strategies for the optimal distribution of pressure sensors is of critical importance, as the design constraints on aircraft necessitate the sparse distribution of such sensors. To this end, a range of methods have been investigated for the purpose of data assimilation from sparse measurements and the development of predictive tools, such as linear systems theory [16] and neural networks [8,17–19].

Oscillator flows such as airfoil flows are generally insensitive to noise and exhibit multiple characteristic time scales [20,21]. This makes them an ideal candidate for the study of multiscale slow–fast-type extreme events—as classified by Farazmand and Sapsis [22]. From this perspective, the Reynolds number regime $\mathcal{O} \sim (10^4)$ is of particular interest. This regime lies between steady laminar and fully

Received 14 March 2023; revision received 26 June 2023; accepted for publication 6 July 2023; published online 31 July 2023. Copyright © 2023 by the American Institute of Aeronautics and Astronautics, Inc. All rights reserved. All requests for copying and permission to reprint should be submitted to CCC at www.copyright.com; employ the eISSN 1533-385X to initiate your request. See also AIAA Rights and Permissions www.aiaa.org/randp.

*Post-Doctoral Associate, 77 Massachusetts Avenue; bbarthel@mit.edu (Corresponding Author).

[†]Professor of Mechanical and Ocean Engineering, 77 Massachusetts Avenue.

turbulent regimes, and it is especially susceptible to highly nontrivial dynamics, which depend significantly on angle of attack and Reynolds number [23–25]. One tool for the study of such systems is the continuous wavelet transform, which, in a manner analogous to the short-time Fourier transform, quantifies the time-varying strength of a signal's frequency components [26,27]. Wavelet analysis has previously been used in the context of extreme events by Cousins and Sapsis [28,29], Bayındır [30], and Srirangarajan et al. [31] for the detection of rogue waves and turbulent bursts in pipe flow, respectively. The majority of these studies have focused on spatial wavelet transforms, while here we perform a temporal analysis. Additionally, the combination of wavelet analysis with machine learning models using the type of output weighted strategies discussed above remains largely unexplored.

This work conducts an investigation of the physical mechanisms driving the extreme bursting events observed in the flow over a two-dimensional airfoil at $R = 17,500$ at constant angle of attack. We first analyze the statistical connection between various observables with the aim of identifying the causal mechanisms driving the observed extreme events. We analyze basic correlations as well as mutual information, which is a measure of mutual dependence between two signals [7,32,33]. We connect the results of this analysis to further spectral analysis of the data with the aim of shedding light on the spatiotemporal scales driving the extreme event dynamics. We then build on previous work by Rudy and Sapsis [8], who studied the data-driven reconstruction of this flow using a range of neural network architectures. We exploit the findings of our analysis to design extreme event tailored—also referred to as output-weighted [4]—data processing and training strategies for the efficient data-driven prediction of extreme events from optimal sparse sampling of the surface pressure. For a broad discussion of output-weighted strategies for neural networks applied to this and other systems, see Rudy and Sapsis [34].

The rest of the paper is organized as follows. In Sec. II we describe the problem under investigation. In Sec. III we perform a statistical analysis of the data and describe the physical mechanisms driving the extreme events, and how these manifest in the observed data. In Sec. IV we discuss the limitations of offline optimal sensing algorithms. The main results of this work are then presented in Sec. V, and we provide some discussion of our findings in Sec. VI.

II. Problem Description

We consider a two-dimensional direct numerical simulation of an incompressible flow around a NACA 4412 airfoil at an angle of attack $\alpha = 5^\circ$ and a chord-length-based Reynolds number $R = 17,500$. The flow is governed by the Navier–Stokes and continuity equations:

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} - \frac{1}{R} \nabla^2 \mathbf{u} + \nabla p = 0 \quad (1)$$

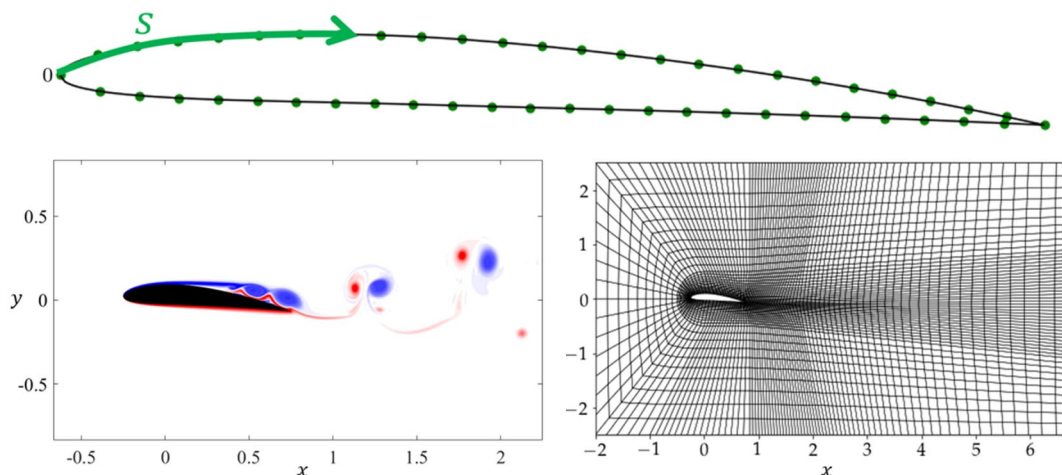


Fig. 1 Airfoil geometry, with arclength measure s (dots represent every other sensor location), snapshot of vorticity, and computational grid.

$$\nabla \cdot \mathbf{u} = 0 \quad (2)$$

where $\mathbf{u} \equiv [u(x, y, t), v(x, y, t)]$ is the velocity, $p(x, y, t)$ is the pressure field, t is time, and x, y are the spatial dimensions parallel and perpendicular to the freestream, respectively. The simulation is carried out using the open source spectral element code Nek5000 developed by Fischer et al. [35] with 4368 elements at spectral order 7 and a convective outflow boundary condition [36]. We use the same data set as Rudy and Sapsis [8], who report that further refinement of the numerical grid did not meaningfully alter the results. At this Reynolds number the flow is susceptible to intermittent, yet nonperiodic turbulent bursts which manifest as high-amplitude fluctuations in a range of observables, most notably the drag. These events may be seen as prototypical of the kind of unsteady dynamics faced in a variety of aerospace applications. In this work we focus on two observables: the surface pressure and the drag coefficient. The former is a practically measurable quantity and will serve as the input to our model, while the latter acts as the extreme event indicator and will serve as the model output.

Throughout this work we define s as a generalized measure of arc length measured clockwise from the leading edge (as shown in Fig. 1). For example, $s \in [0, 0.5)$ refers to the upper surface of the airfoil and $s \in [0.5, 1)$ refers to under side. The surface pressure is saved at 100 equally spaced locations around the airfoil surface. A visualization of the airfoil flow, the simulation grid, and the arc length measure are summarized in Fig. 1, and we refer the interested reader to Rudy and Sapsis [8] for a more detailed discussion of the numerical method.

The aerodynamic forces are computed using skin friction and surface pressure according to

$$\mathbf{F}(t) = \oint_s (\boldsymbol{\tau}(s, t) + nP(s, t)) ds = D(t)\hat{\mathbf{e}}_x + L(t)\hat{\mathbf{e}}_y \quad (3)$$

Here t is time, and x and y represent the directions parallel and normal to the freestream, respectively. The lift and drag coefficients are then defined as

$$C_D(t) \equiv \frac{D(t)}{\rho U_\infty^2 c}, \quad C_L(t) \equiv \frac{L(t)}{\rho U_\infty^2 c} \quad (4)$$

To distinguish the extreme events from the background vortex shedding, we apply a Gaussian smoothing operation to the time series of the drag coefficient to extract the nonperiodic behavior:

$$q(t) \equiv (K * C_D)(t) \quad (5)$$

where $K(t') \propto \exp(-(t'/2f_v)^2)$ is a Gaussian smoothing kernel and $f_v = 1.44$ is the most energetic frequency corresponding to periodic vortex shedding. Moving forward we simply refer to $q(t)$ as the drag. In addition to the raw pressure signal, we also consider a version of the pressure with the same Gaussian filter applied:

$$\tilde{P}(s, t) \equiv (K * P(s, t))(s, t) \quad (6)$$

We refer to $P(s, t)$ and $\tilde{P}(s, t)$ as the raw and filtered pressures, respectively. In general we will consider the pressure measured at subset of discrete sensor locations, and thus treat the surface pressure as a vector valued quantity $\mathbf{P}(t) \in \mathbb{R}^n$, where n is the number of sensors. An illustrative example of the drag as well as the raw and filtered pressures is shown in Fig. 2.

This flow was previously investigated by Rudy and Sapsis [8] using a deep long short-term memory (LSTM) network. Those authors considered a variety of input observables, and found that the extreme events could be predicted from a range of different observables including full- and reduced-order descriptions of the flowfield as well as surface pressure. This suggests that the extreme events are a result of an underlying physical instability inherent in the governing equations. Practically, we aim to exploit this mechanism to predict future extreme events observed in the drag from sparse measurements of the surface pressure as efficiently as possible. In other words we seek a data-driven map

$$\mathbf{P}(t) \rightarrow q(t + \tau) \quad (7)$$

for maximum lead time τ , with minimal $\dim(\mathbf{P})$, and at minimal computational cost. All the code used to generate the results presented in this work can be found at https://github.com/ben-barthel/Airfoil_EE_2023.

III. Statistical Analysis and Physical Mechanisms of Extremes

To gain a deeper insight into the dynamics of the flow and mechanisms driving the extreme events, we first perform a detailed statistical analysis of the data. We first analyze the surface pressure, as this will serve as the basis of our modeling efforts, then in Sec. III.C we analyze the vorticity field—both globally and locally along the airfoil surface—to further probe the extreme event dynamics. Before presenting our results, we first review some definitions we use throughout the following sections. For a signal $x(t)$ with discrete values x_i and distribution X , we define the mean μ_x , variance σ_x^2 , and the probability density function $f_X(x)$. For two signals $x(t)$ and $y(t)$ the covariance is defined as

$$\sigma_{xy} \equiv \text{cov}(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y) \quad (8)$$

To further quantify the connection between two signals we also make use the mutual information defined as

$$I(X, Y) \equiv \int_y \int_x f_{X,Y}(x, y) \log \left(\frac{f_{X,Y}(x, y)}{f_X(x)f_Y(y)} \right) dx dy \quad (9)$$

where $f_{X,Y}$ is the joint probability density function of X and Y . The mutual information is the Kullback–Leibler (KL) divergence between the joint probability distribution and the product of the marginal probability distributions—it quantifies the error in the assumption that two distributions X and Y are uncorrelated. We also propose the “extreme event conditioned mutual information,” defined as the mutual information integrated only over values of the output greater than two standard deviations from the mean: $y > 2\sigma_y$ (all values of the input X are included):

$$I_{EE}(X, Y) \equiv \int_x \int_{y > 2\sigma_y} f_{X,Y}(x, y) \log \left(\frac{f_{X,Y}(x, y)}{f_X(x)p_Y(y)} \right) dx dy \quad (10)$$

We choose a cutoff of two standard deviations; however, we found that the results were not sensitive to changes of $\pm\sigma$. For all results presented in this work the probability density functions are approximated using Monte Carlo estimation, and the relevant integrals are carried out using trapezoidal integration.

A. Covariance and Mutual Information Structure

To investigate the spatial dynamics of the surface pressure we compute the covariance and mutual information matrices: $\text{cov}(P(s, t), P(s', t))$ and $I(P(s, t), P(s', t))$ for both the raw pressure P and the filtered pressure \tilde{P} . These quantify the information shared between different locations along the airfoil. The covariance matrices and mutual information matrices are shown in Figs. 3a and 3b, respectively. The left plot shows the raw pressure signal and the right shows the filtered pressure. We notice that the results for the raw and filtered pressure are qualitatively similar, and thus the following discussion applies to both.

These results reveal three distinct regions. First, the underside of the airfoil, $0.5 < s < 1.0$. This region displays a high degree of mutual information and strong correlation. Second, the front section of the upper surface, $0 < s < 0.3$. This region exhibits similar features as the underside: strong mutual information and correlation; however, in this region the mutual information drops off much more quickly with separation between the sensor locations. These results imply that these regions are amenable to sparse sensor distribution, since any

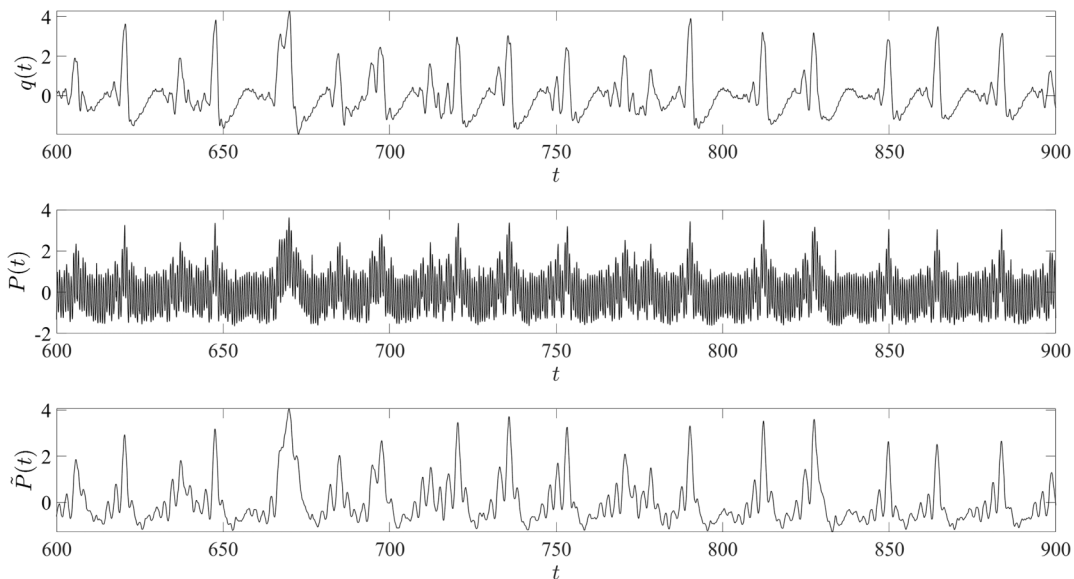


Fig. 2 From top to bottom: filtered drag coefficient, raw pressure signal, and filtered pressure signal. Pressure data are taken at a single representative sensor location 25% of the way along the upper surface.

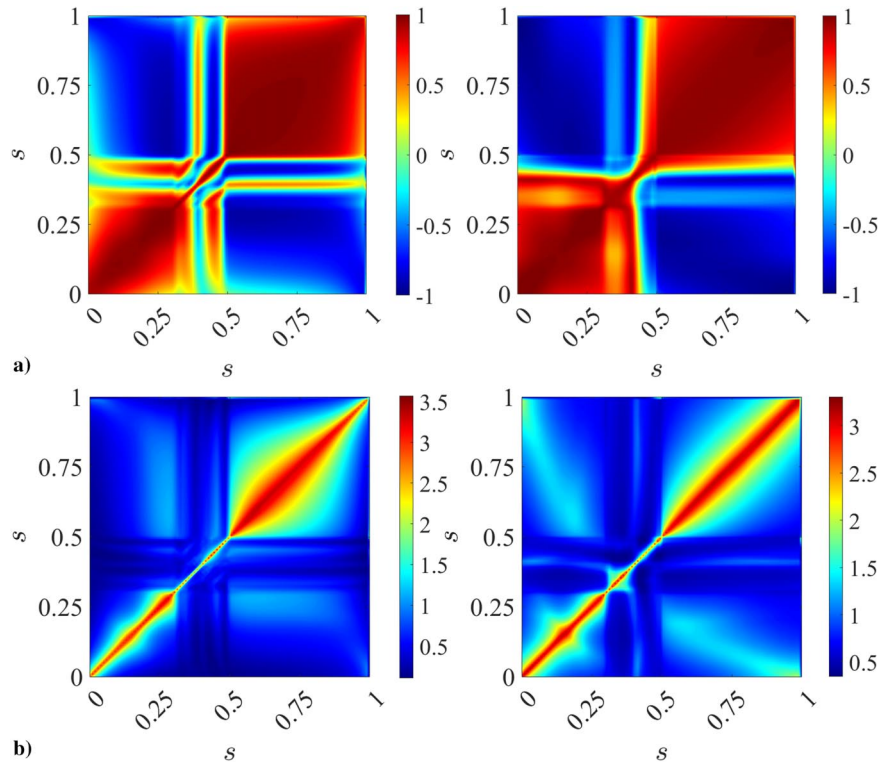


Fig. 3 Pressure covariance matrix $\text{cov}(P(s), P(s'))$ (a) and mutual information matrix $I(P(s), P(s'))$ (b). Raw pressure (left); filtered pressure (right). Here s is the generalized arc length.

additional sensor is unlikely to contribute new information. Note also that due to the airfoil having a nonzero angle of attack we see strong negative correlation between the upper and lower surface pressures. However, there is little mutual information between the upper and lower surfaces. This implies that measurements on one surface do not necessarily provide information about the other. The exception to this is the rear part of the upper surface, $0.3 < s < 0.5$. In this region there is little to no mutual information and significant variation in the correlation. As a result, this region likely requires relatively higher sensor density. We note that the transition point between the first two regions, $s = 0.3$, coincides with the point of flow separation (see Fig. 1). Therefore the increased disorder observed in Sec. III is likely due to the complexity and increased unsteadiness of the flow in this region.

We also compute the standard and extreme event mutual information between the surface pressure and the drag coefficient:

$I(P(s, t), q(t + \tau))$ and $I_{EE}(P(s, t), q(t + \tau))$. These are plotted in Fig. 4 for the raw and filtered pressure signals for a range of τ . As we are interested in the spatial variation of these quantities, to ease comparison we normalize each by its maximum value. In all cases we do not observe strong dependence on the lead time τ . As with the intrapressure mutual information we see strong spatial dependence in the region $0.3 < s < 0.5$. For both the raw and filtered pressure signal, the extreme event mutual information is (locally) peaked in this region. We note that the local peak is less pronounced for the raw data, because in that case the vortex shedding frequency overwhelms all other frequency content. On the other hand, for the standard mutual information this region lies in the trough of the spatial distribution. This suggests that the mechanisms driving the extreme events are strongest in the separation region. However, due to their rarity, this connection is not reflected in the standard mutual information profile. Next, we analyze this extreme event

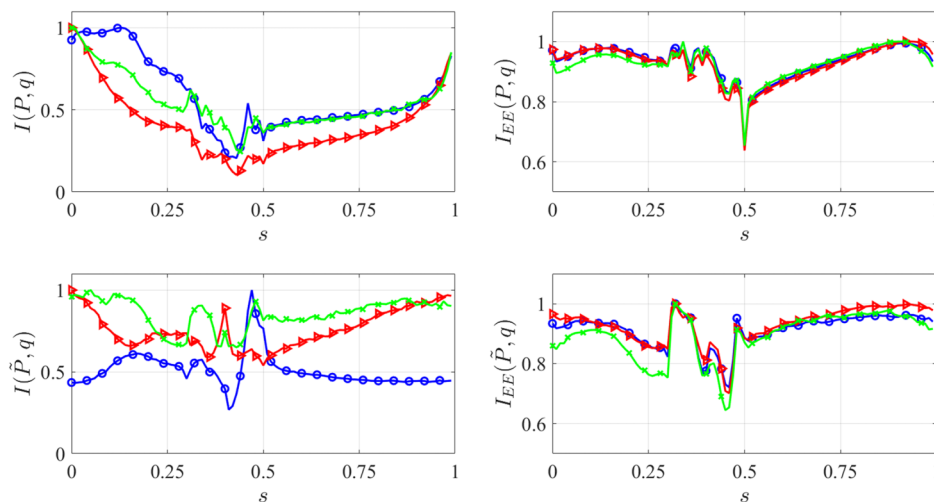


Fig. 4 Normalized mutual information (left) and extreme event mutual information (right) between raw (top row) and filtered (bottom row) pressure signal as a function of arc length s for $\tau = 0, 3, 7$ (blue circles, red triangles, green crosses).

mechanism, and its connection to the extreme event mutual information in more detail.

B. Extreme Event Mechanisms

Extreme events occurring in dynamic systems are known to arise due to a variety of factors—not all of which are fully understood. One class of dynamic system known to give rise to extreme events is slow-fast multiscale systems [2]. In such cases, the system evolves on two or more manifolds that have significant separation of characteristic time scales. At most times, the system evolves along the slower manifold. Occasionally, the trajectory may encounter instability of this slow manifold, resulting in the trajectory rapidly approaching the fast manifold. Once the unstable region has passed the system relaxes back to the slow manifold. Such phenomena are often observed as sporadic high-amplitude bursts [2].

Airfoil flow is an example of such a multiscale system. Such systems have multi-peaked spectral content—or in other words they have multiple characteristic frequencies. At this Reynolds number there are two high frequencies (fast manifolds), the vortex shedding frequency, $f_v = 1.44$, corresponding to the energetically dominant oscillatory flow, and a second frequency corresponding to the extreme event dynamics, $f_e = 0.4$. Figure 5a shows the standard and premultiplied temporal Fourier power spectrum of the filtered surface pressure defined as

$$\mathcal{P}(s, f) \equiv \int P(s, t) e^{-ift} dt \quad (11)$$

and

$$\mathcal{P}_{pm}(s, f) \equiv f\mathcal{P}(s, f) \quad (12)$$

respectively. The latter is useful for visualizing higher frequency content as it de-emphasizes the slow dynamics ($f \rightarrow 0$). In the standard power spectrum there is a clear maximum close to $f = 0$, corresponding to the slow dynamics. The extreme event frequency is also evident in the plain spectrum, but is best seen in premultiplied spectrum, which exhibits a clear peak around $f = 0.4$. We limit the frequency axis to $f \in [0, 1]$ —which does not include the vortex

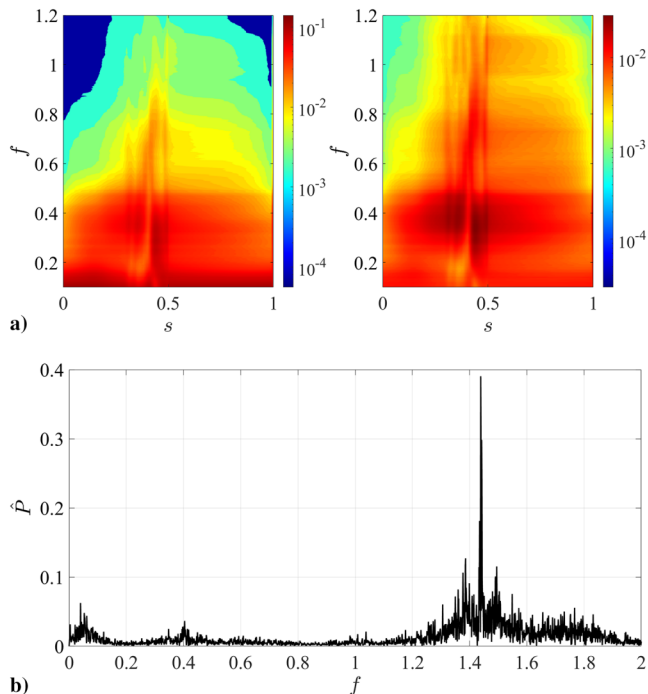


Fig. 5 Standard (left) and premultiplied (right) power spectrum of the surface pressure (a). Power spectrum of surface pressure at a single location (b).

shedding frequency, $f = 1.44$ —as it is much stronger than either the slow or extreme event dynamics and obscures these when included in the figure. Note also the increased magnitude in the region $0.3 < s < 0.5$ consistent with the results of Sec. III.A. To illustrate the relative strengths of f_v and f_e , Fig. 5b shows the power spectrum of the surface pressure for $f \in [0, 1.5]$ at a single location along the airfoil.

The connection between this frequency and the extreme events is best interpreted through the wavelet transform. The wavelet transform allows for the visualization of the time-varying strength of a signal’s frequency content. The wavelet transform has been used in the past to identify extreme events by for example Srirangarajan et al. [31] to detect bursts in pipe flow and Cousins and Sapsis [28,29] for the early detection of rogue waves. The continuous wavelet transform of a signal $x(t)$ is defined as

$$\hat{X}(f, t) \equiv \mathcal{W}(x(t)) = \sqrt{\frac{f}{f_c}} \int_{-\infty}^{\infty} x(s) \psi\left(f \frac{s-t}{f_c}\right) ds \quad (13)$$

Here $\psi(t)$ is the wavelet function, and f_c is the wavelet specific center frequency. The wavelet function is not unique, but must satisfy several conditions, including finite energy and localized support [26]. Here we use the Morlet wavelet

$$\psi(t) = e^{-t^2/2} \cos(5t) \quad (14)$$

Moving forward we refer to the wavelet transform of the pressure signal as $\hat{P}(s, f, t)$, where f is the frequency. The wavelet transform of the pressure signal at $s = 0.34$ is shown in the upper panel of Fig. 6. This location corresponds to the peak in the spatial power spectrum in Fig. 5. Figure 6 clearly shows the bursts of energy at $f = f_e = 0.4$.

We define the extreme event indicator γ as the wavelet coefficient that maximizes the spectrogram of the filtered pressure signal, i.e., for $f = f_e$

$$\gamma(s, t) \equiv |\hat{P}(s, f_e, t)| \quad (15)$$

In the lower panel of Fig. 6 we show the clear correlation between γ and the extreme drag events. These results suggest that the extreme events observed in this flow are indeed of the multiscale system type. Furthermore, viewing them in relation to the relative global energy content of f_v and f_e illustrated in Fig. 5b highlights the limits of energy-based reduced-order modeling strategies such as principle component analysis (PCA) or proper orthogonal decomposition (POD) when used in the context of extreme event prediction in such systems. These methods rank modes based on total energy content, a criterion that in this case would significantly underestimate the role of the extreme event frequency and necessitate retaining a large number of modes in order to capture the extreme event dynamics. This is consistent with the findings of Rudy and Sapsis [8], who observed very slow convergence with rank when approximating the flowfield using POD. In Sec. V we

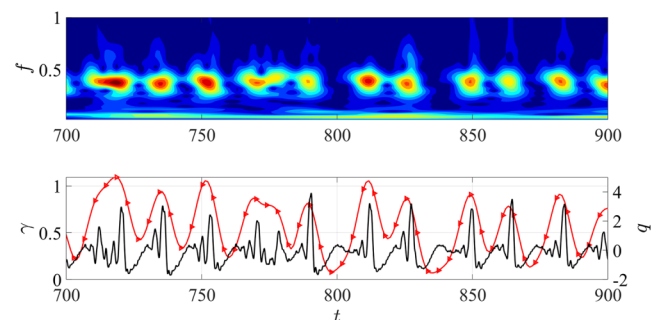


Fig. 6 Absolute value of wavelet-transformed pressure signal at $s = 0.34$ (upper panel). Extreme event indicator $\gamma = |\hat{P}|_{f=f_e}$ at same location (red triangles) compared to drag coefficient (black line) (lower panel).

show that preprocessing the pressure signal to extract this extreme event frequency allows for a drastic reduction in the model complexity and data required for accurate forecasting.

C. Flowfield Analysis

In Secs. III.A and III.B we identified the unstable extreme event frequency and its connection to the separation region of the airfoil. Here we zoom out and analyze the full boundary layer to better understand the mechanisms at the heart of the instability and the subsequent extreme drag events. To this end we follow the time evolution of the vorticity field

$$\omega(x, y, t) \equiv \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \quad (16)$$

over the course of a single extreme event from $t = 911$ – 931 , with a peak at $t = 922$. The top panel of Fig. 7 shows three snapshots of the vorticity field over this time interval—for clarity we focus on the region near the boundary layer. The corresponding values of the instantaneous drag coefficient are shown in the small inset of the same figure. The red markers in the latter represent the time instances of the three vorticity snapshots. We see clear evidence of boundary-layer separation/disorder during the time instance corresponding to the peak in drag coefficient. Although we show only a single extreme event here, this disordered behavior of the boundary layer was observed during the peak of all extreme events.

The transient dynamics observed in the full vorticity field are subtle, and thus to better understand these transient dynamics we compute the wavelet transform of the entire vorticity field:

$$\hat{\Omega}(x, y, t, f) = \mathcal{W}(\omega(x, y, t)) \quad (17)$$

This allows us to investigate the component of the vorticity evolving with the extreme event frequency identified in Sec. III.B—which is not the dominant energetic contributor to the full field, and is thus liable to be obscured in the snapshots of the full solution.

1. Global Dynamics

To analyze the dynamics of the boundary layer we first consider the wavelet transform of the full flowfield—focusing on the extreme event frequency

$$\hat{\Omega}_e(x, y, t) \equiv \hat{\Omega}(x, y, t, f) \Big|_{f=f_e=0.4} \quad (18)$$

and the vortex shedding frequency

$$\hat{\Omega}_v(x, y, t) \equiv \hat{\Omega}(x, y, t, f) \Big|_{f=f_v=1.44} \quad (19)$$

The wavelet components associated with the extreme event frequency (18) and the vortex shedding frequency (19) are plotted in the center and lower panels of Fig. 7, respectively. In the former, we see clear evidence of a coherent structure with relatively small characteristic spatial length scale that undergoes a transient instability, resulting in a temporary loss of coherence during the extreme event before recovering as the drag coefficient returns to its nominal state. The vortex shedding mode has a much larger characteristic length scale—on the order of the vortical structures observed in the full vorticity field—and does not appear to undergo any significant changes in its dominant length scale during the extreme event.

To further illustrate the dynamics of these two frequency components we compute the temporal correlation function

$$r_\alpha(t_0, t) \equiv \left| \iint \hat{\Omega}_\alpha^*(x, y, t_0) \hat{\Omega}_\alpha(x, y, t) dx dy \right| \quad (20)$$

where the superscript “*” denotes the complex conjugate and $\alpha = e, v$ and the integration is performed over the entire domain. For the special case where $t_0 = t$ this is equivalent to the L_2 norm of the wavelet mode

$$\|\hat{\Omega}_\alpha\|^2(t) \equiv \iint |\hat{\Omega}_\alpha(x, y, t)|^2 dx dy \quad (21)$$

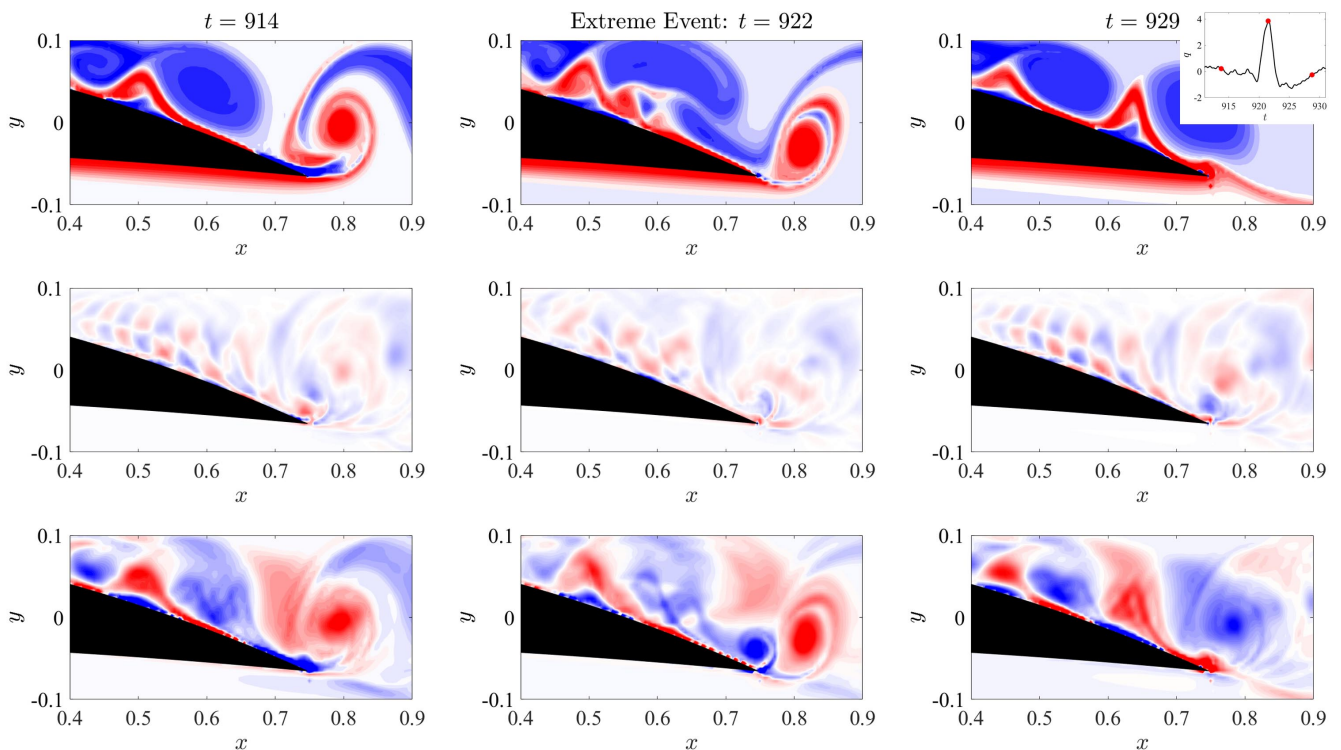


Fig. 7 Snapshots of vorticity traversing a single extreme event. Top panel: full vorticity ω . Center and low panels: wavelet coefficient evaluated at the extreme event frequency $\hat{\Omega}_e$ and vortex shedding frequency $\hat{\Omega}_v$, respectively.

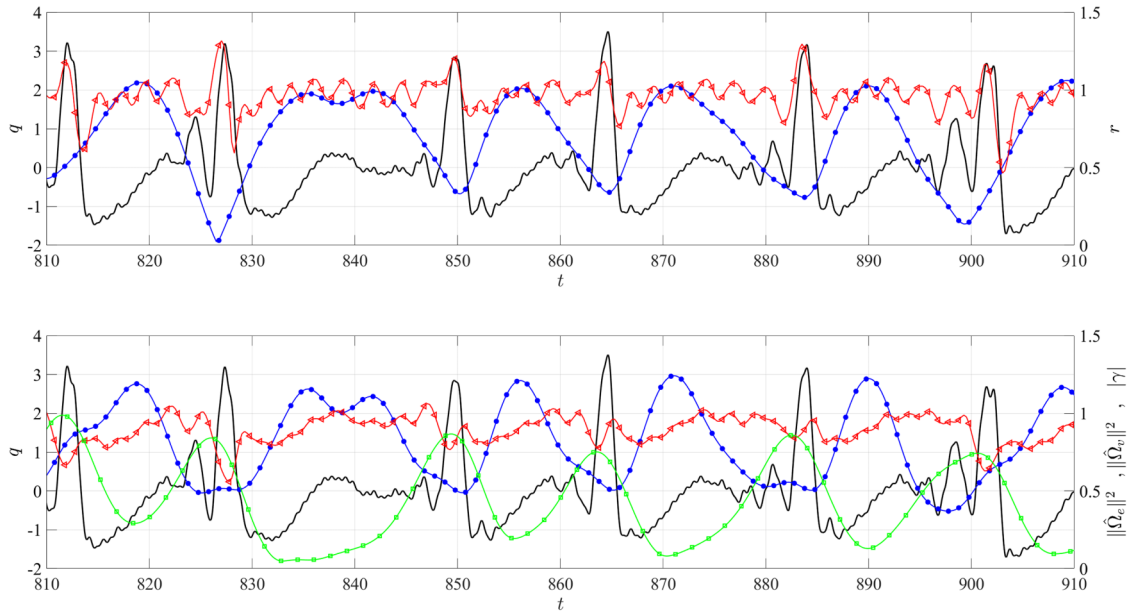


Fig. 8 Top panel: normalized correlation function (20) for $t_0 = 911$ for $\hat{\Omega}_e$ (blue circles) and $\hat{\Omega}_v$ (red triangles). Bottom panel: normalized vorticity wavelet coefficient norms and normalized surface pressure wavelet coefficient $\gamma(s,t)_{s=0.34}$ (15) (green squares). Drag coefficient $q(t)$ is shown in black.

These metrics, respectively, quantify the temporal evolution of the shape (length scale) and magnitude of the vorticity at a specific temporal frequency.

The correlation (20) is plotted in the upper panel of Fig. 8 for both $\hat{\Omega}_e$ (blue circles) and $\hat{\Omega}_v$ (red triangles). We fix $t_0 = 911$ as a representative snapshot corresponding to the vorticity structure during the quiescent periods—however, any quiescent time instance could be used. We clearly observe a systematic and drastic loss of coherence in the extreme event mode during the spikes in the drag coefficient. Inspection of the center panel of Fig. 7 suggests that this is at least in part due to an increase in the dominant spatial length scale. The coherence of the vortex shedding mode actually fluctuates—with what further analysis reveals to be at close to the extreme event frequency—with an amplitude that slightly increases during the extreme events; however, no drastic loss of coherence is observed. The significance of this fluctuation is not immediately clear, but it suggests some interaction between the two frequency components—the investigation of which is the focus of ongoing research.

The lower panel of Fig. 8 shows the time evolution of the norm (21) of the extreme event mode (blue circles) and vortex shedding mode (red triangles). To ease comparison, both norms are normalized by

their value at t_0 . For reference we also plot the drag coefficient (black) and the normalized extreme event wavelet coefficient (15), evaluated at the same representative location discussed previously, $s = 0.34$ (green squares). While as previously noted, the magnitude of γ peaks in sync with the extreme drag events, the global norm of the extreme event component of the vorticity, $\hat{\Omega}_e$, drops in magnitude during the same time intervals. In contrast to both of these, the dominant vortex shedding mode, Eq. (19), is significantly more stable and exhibits a much smaller relative drop in magnitude during the extreme events.

2. Dynamics on the Airfoil Surface

To better understand the instability of the extreme event mode and the associated transfer of energy, we also compute the spatial Fourier transform of the temporally wavelet transformed vorticity evaluated at the airfoil surface. In other words we compute

$$\begin{aligned} \tilde{\Omega}_s(t, f, k_s) &\equiv \left| \mathcal{F}(\hat{\Omega}(x(s), y(s), t, f)) \right| \\ &= \left| \int \hat{\Omega}(x(s), y(s), t, f) e^{-ik_s s} ds \right| \end{aligned} \quad (22)$$

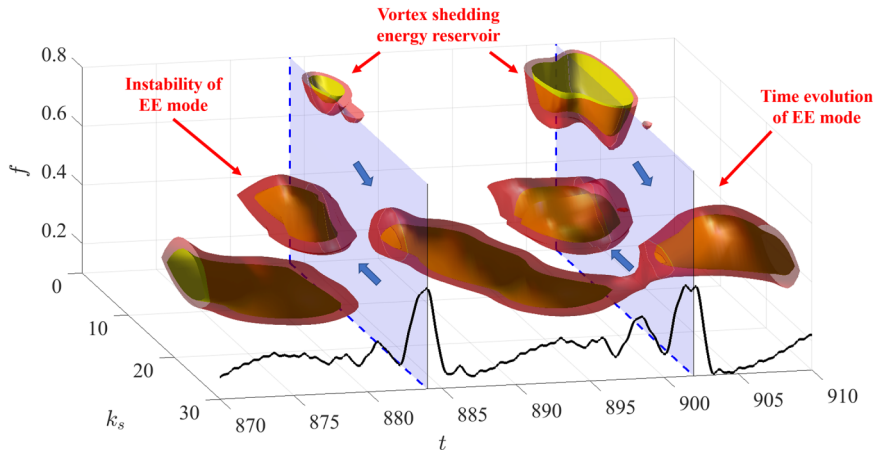


Fig. 9 Isocontours of vorticity amplitude on the airfoil surface as a function of temporal frequency f , spatial wavenumber k_s , and time t . Blue arrows represent direction of energy transfer.

where k_s is the spatial wavenumber with respect to the arclength s along the airfoil surface defined in Sec. II, and $x(s)$ and $y(s)$ are the coordinates of that surface. This reduces the spatial dimensions from two to one, and thus allows us to visualize the transfer of energy between various spatial and temporal scales as a function of time. The isocontours of this quantity are plotted in Fig. 9 over a time horizon covering two extreme events—the drag coefficient is also plotted for reference. This plot succinctly summarizes the observations discussed above. First we see that during the quiescent periods, the energy of the extreme event mode, which actually seems to meander slightly about $f = 0.4$, is concentrated at a wavenumber $k_s \approx 20$. Then during the extreme events, energy is drawn from the higher frequency vortex shedding mode, which serves as an energy reservoir, leading to the instability of the extreme event mode that abruptly transfers its energy to a lower wave number $k_s \approx 10$. This appears in Fig. 9 as the “pinching off” of the isocontours during the spikes in the drag. This is the manifestation of loss of coherence and increase in spatial scale of the extreme event mode observed in center panel of Fig. 7 and quantified in the upper panel of 8. We note in closing that this phenomenon of the extreme events evolving on a manifold distinct from the energetically dominant one has been observed in other systems, such as Kolmogorov flow [37,38]. In that case the energy of the flow is dominated by a specific triad of spatial wave-numbers; however, projecting the flow onto this triad fails to predict the extreme energy dissipation events observed in that flow. Similarly, in the case of the airfoil flow considered here, extracting only the dominant vortex shedding dynamics would miss the extreme event dynamics entirely. However, our findings are contrary to the far more common phenomenon of instabilities transferring energy from large to small scales.

IV. Offline Sparse Sensor Placement

The analysis of mutual information structure described in Sec. III indicates that certain sections of the airfoil are statistically more informative of the drag coefficient. To test the practical implications of this discovery, we first propose an offline strategy to optimally select sensor locations. Such an algorithm does not require actually training the neural network. Therefore, it can be thought of as a preprocessing step that allows us to optimally design the network architecture before training. At each iteration, the sampling algorithm, which is outlined in Algorithm 1, selects the next best sensor location by maximizing a cost function referred to as an acquisition function. Throughout this work we use the term “acquisition function” strictly in connection with such a sampling strategy, and the term “cost function” to refer to the cost function used to train a given model.

For a given application, the choice of acquisition function is not obvious; see, for example, Chaloner and Verdinelli [39], Sapsis [40], and Yang et al. [41]. In this framework, sensor locations are selected sequentially, and therefore we seek locations that are maximally informative of the drag coefficient and minimally redundant with respect to the previously placed sensors. Thus, we propose the following two acquisition functions based on the previously defined standard and extreme event mutual information:

$$a_{j+1}^1(s, s_j^*) = \frac{I(P(s), q)}{(1/j) \sum_{k=1}^j I(P(s), P(s_k^*))} \quad (23)$$

$$a_{j+1}^2(s, s_j^*) = \frac{I_{EE}(P(s), q)}{(1/j) \sum_{k=1}^j I(P(s), P(s_k^*))} \quad (24)$$

where s^* is the vector of optimal sensor locations that have already been placed. The numerator—the mutual information between the pressure signal with the drag coefficient—rewards predictive capability. The denominator—the average of the intrapressure sensor mutual information—penalizes redundancy. This second condition ensures that sensors are not placed in locations that do not contribute information not already encoded in previously placed sensors.

Algorithm 1: Optimal placement of N sensors

```

while  $j < N$  do
   $s_{j+1}^* = \operatorname{argmax}_s a_{j+1}(s, s_j^*)$ 
  update  $s_{j+1}^s = [s_j^*, s_{j+1}^*]$ 
end while

```

There is no unique way to quantify the connection between a prospective sensor location and the previously placed sensors, and the average used here is only one option. Therefore, we also considered a second acquisition function where the arithmetic mean in the denominator of Eqs. (23) and (24) is replaced with a geometric mean, but we did not observe significantly different results. A more exhaustive study of candidate functions is beyond the scope of this work, and so for the sake of brevity we restrict ourselves to Eqs. (23) and (24). Going forward we refer to any results obtained using this algorithm as offline-mutual-information (OMI_N), where N is the number of sensors.

A. Results: Sensor Placement

We apply Algorithm 1 with acquisition functions (23) and (24) to our data set to compute the first six optimal sensor locations. Because the results of Fig. 3b indicate that the general behavior of the mutual information is not dependent on the lead time τ , we fix $\tau = 0$. Additionally, in order to facilitate comparison with Rudy and Sapsis [8], we consider only the raw pressure signal. The acquisition function landscape at each iteration is plotted in Fig. 10. The optimal sensor locations after each iteration are then summarized in Fig. 11.

The globally optimal sensor locations are simply the points of maximum standard and extreme event mutual information—these are located at approximately $s = 0.15$ for Eq. (23) and $s = 0.3$ for Eq. (24), respectively. However, inspection of Fig. 10 indicates that in the latter case the acquisition function landscape does not display any significant variation along the airfoil, calling into question the viability of Eq. (24) as a practical metric for optimal sensor placement. For the standard mutual information case, Eq. (23), at iterations 2–4 the acquisition function exhibits multiple local maxima of roughly equal value in the region $0.3 < s < 0.5$. These multiple peaks are sequentially “picked off” throughout the iterations 2–4. This phenomenon is also observed, but to a slightly lesser extent, in the extreme event mutual information case (24). The similarity of these local maxima mean that these four sensor locations should be thought of as an “optimal grouping” rather than a strict ranking, since measurement noise or numerical errors could affect their ordering. However, we note that adding small amounts of noise did not significantly impact the qualitative features of the results. Figure 11 highlights that while the ordering of the sensors varies between the standard and extreme event versions of the model, the final distribution of the optimal sensors is qualitatively very similar.

The underside of the airfoil ($0.5 < s < 1$) is completely ignored by the algorithm until iteration 5 for the standard case and iteration 6 for the extreme event case, where a strong maximum is observed just downstream of the leading edge. This solitary underside sensor near the leading edge is consistent with the mean pressure profile observed in the flow over an inclined airfoil. The mean pressure gradient (w.r.t. arc length) is generally significant along the upper surface but relatively weak along the lower surface. Therefore, a single sensor can capture a significant amount of the information of the pressure field along the lower surface, since once the jump in pressure across the leading edge is established there is not much more to be gained from further probing the pressure along the underside of the airfoil.

B. Results: Evaluation

To test the efficacy of the proposed offline sensor placement algorithm we train the same LSTM network described in Rudy and Sapsis [8] using the first five optimal sensor locations predicted using

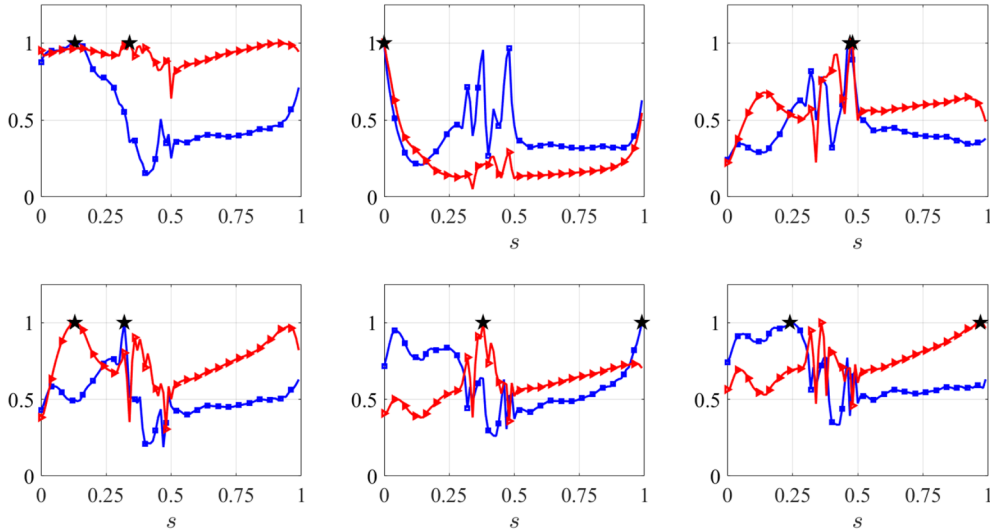


Fig. 10 OMI acquisition function landscape for sensors 1–6 (left to right, top to bottom)—maxima indicated by stars. Standard mutual information (23) (blue squares); extreme event mutual information (24) (red triangles).

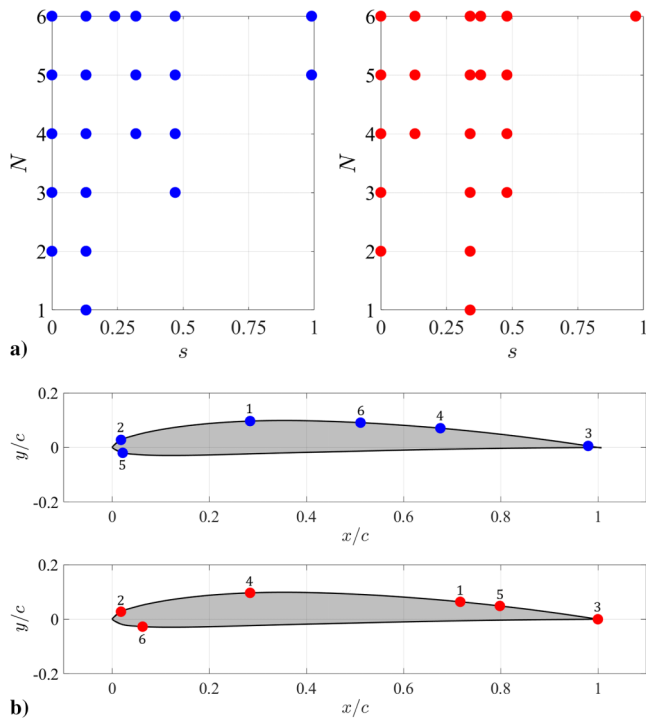


Fig. 11 Optimal sensor locations after each iteration N (a). Optimal sensor locations on airfoil: standard mutual information (blue) and extreme event mutual information (red) (b).

Algorithm 1 with acquisition function (23). Due to the similarities of the sensor locations predicted by Eqs. (23) and (24) and the high computational cost of training the network, we omit the predictions of Eq. (24) from this analysis. We compare those results to those in Rudy and Sapsis [8] using 50 sensors spaced equally around the airfoil. The network architecture is

$$\begin{aligned} &\rightarrow \text{FC32} \rightarrow \text{LSTM32} \rightarrow \text{LSTM32} \rightarrow \text{FC32} \rightarrow \text{FC16} \rightarrow \text{FC8} \\ &\rightarrow \text{FC4} \rightarrow \text{FC1} \rightarrow \end{aligned} \quad (25)$$

where FC and LSTM stand for “fully connected feed forward” and “long short-term memory,” respectively, and the swish activation function is applied between each layer [42]. To isolate the effects of the sensor placement, we make no changes to the architecture or

Table 1 Summary of neural net models compared in Fig. 12

Name	Sensor placement	Input observable	No. of sensors (N)
OMI_N	OMI algorithm	$P(t)$	N
uni_5	Uniform	$P(t)$	1.5
$uni_{.50}$	Uniform	$P(t)$	50

other than the input dimension and utilize the same training strategies as Rudy and Sapsis [8]. Training is conducted using 70% of the data, with the remaining 30% split evenly between validation and testing. The model was trained over three restarts using 140 history points and a mean square error (MSE) loss function until the validation error failed to decrease for 10 epochs—no regularization was used. The interested reader is referred to Rudy and Sapsis [8] for a more detailed description of the network architecture and training strategy.

We compare three different models: the reference case from Rudy and Sapsis [8] using 50 sensor locations, the OMI_5 model, as well as a second reference case using five uniformly spaced sensors—all three use the raw pressure data as an input. The last case is included to verify that any potential benefit of our algorithm is actually due to the algorithmic placement of the sensors and not simply a reflection of oversampling by Rudy and Sapsis [8]. The three models are summarized in Table 1.

To compare the predictive capabilities of the models we compute both the MSE of the model prediction as well as the “maximum adjusted area under the precision-recall curve”—a metric introduced by Guth and Sapsis [6] that quantifies the accuracy of extreme event prediction. The area under the precision-recall curve is then defined as

$$\alpha(\chi) = \int_0^1 S(R, \chi) dR \quad (26)$$

where the event rate χ is defined as the probability that the output exceeds some threshold, the precision S is the ratio of correct event predictions to total event predictions, and the recall R is the ratio of correct event predictions to the actual number of events. The maximum adjusted value is then defined as

$$\alpha^* = \max_{\chi \in [0,1]} (\alpha(\chi) - \chi) \quad (27)$$

When the value of α^* is large (approaches unity) the model is very good at predicting rare events; alternatively, when the value

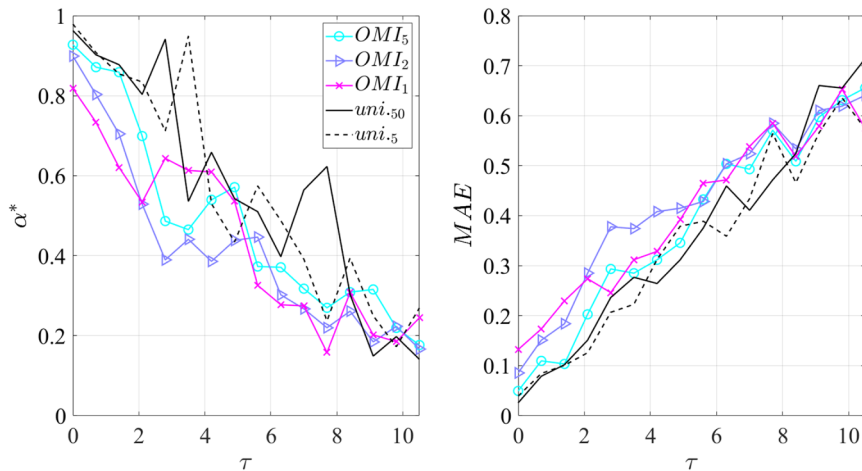


Fig. 12 Maximum adjusted area under the precision recall curve, α^* , and mean absolute error, MAE , for the models summarized in Table 1 as a function of the lead time τ .

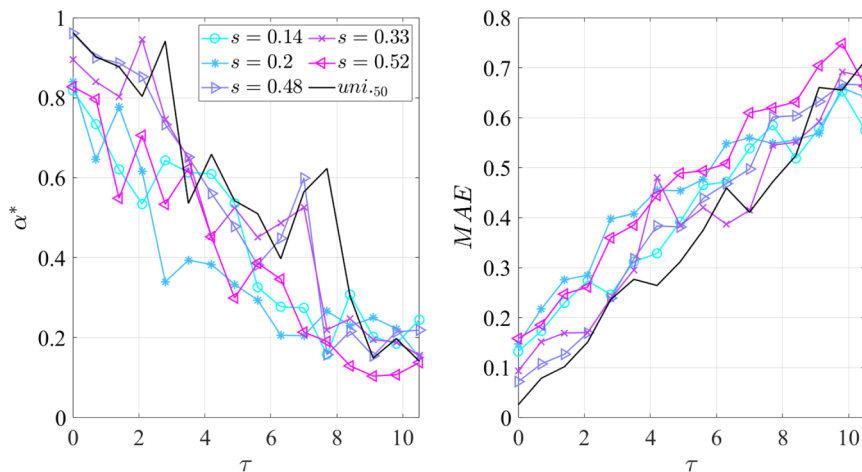


Fig. 13 Maximum adjusted area under the precision recall curve, α^* , and mean absolute error, MAE , for models trained using a single sensor as a function of the lead time τ .

approaches zero a model does no better than a guess based on the aggregate frequency of extreme events.

Figure 12 compares the mean absolute error (MAE) and α^* of the various models for a range of lead times τ . As expected, for all cases MAE increases and α^* decreases with τ —it is more difficult to predict the far future. Comparing the models, we first observe that there is no significant difference between the two reference cases with 5 and 50 sensors, suggesting that the neural network model developed by Rudy and Sapsis [8] is amenable to far more sparse sensor distributions than is suggested by those authors. Additionally, we see no clear distinction between the results of the model trained using our predicted optimal sensor locations and the uniformly sampled reference cases. This highlights the limitations of the mutual information as a practical tool for engineering design. To further highlight this limitation and exclude the possibility of our conclusions being influenced by oversampling, we train the model using the five optimal sensor locations individually. In other words, we train five models, each with a single sensor (the n th optimal OMI prediction) as its input. The same error metrics for this experiment are plotted in Fig. 13. Again we see that the optimal sensor location $s = 0.14$ performs no better, and in many cases worse, than the suboptimal locations—see, for example, the value of α^* at $\tau = 0$ and 7. These results strongly suggest that optimal sensing based purely on mutual information does not adequately capture the extreme event mechanism identified in Sec. III.B and thus fails as a practical tool for optimal sensing.

V. Wavelet Preprocessing for Extreme Event Prediction

Despite its robustness to sensor location, the LSTM network considered in Sec. IV is expensive to train, so here we explore a different avenue of model reduction: preprocessing the data through offline identification of the extreme event dynamics. In Sec. III.B we find that the extreme event dynamics are directly related to the dynamics of a single frequency component. Here we show that exploiting this observation through the event indicator (15) allows for the forecasting of the extreme events using very simple network models.

A. Methods

The extreme event indicator γ defined in Eq. (15) is not only highly correlated with the bursting events, as it represents an isolated frequency component, but also free of noise. This makes it amenable to accurate numerical differentiation. We therefore define the following transformation:

$$D_n(\gamma) : \gamma \rightarrow [\gamma, \dot{\gamma}, \dots, \gamma^{(n)}] \quad (28)$$

which allows us to track not only the value of γ , but also its growth rate. This is crucial as we seek to forecast bursting for nonzero lead times τ , and therefore it is imperative for the model to observe growth and not just magnitude. The differentiation operation in Eq. (28) is essentially a phase shift of the signal and thus aids the forecasting capabilities of the model, i.e., the predictions for $\tau > 0$.

We find that, for this flow, a single derivative ($n = 1$) is sufficient, and including a second derivative did not meaningfully improve results. Note that this differentiation is applied offline, and thus does not affect the computational cost of training the model. We propose a preprocessing procedure described in Algorithm 2, in which we replace Eq. (7) with

$$\Gamma(t) \rightarrow q(t + \tau) \quad (29)$$

Here the input data are defined as

$$\Gamma \equiv D_1(\gamma) = [\gamma, \dot{\gamma}] \quad (30)$$

and γ is defined in Eq. (15). We utilize a fully connected feed forward neural network $f: \mathbb{R}^n \rightarrow \mathbb{R}^1$ with layers

$$\rightarrow \text{FC8} \rightarrow \text{FC16} \rightarrow \text{FC16} \rightarrow \text{FC8} \rightarrow \text{FC1} \rightarrow \quad (31)$$

with the swish activation function [42] applied between each layer. A regularization constant of 0.01 was applied to the activation layers—no regularization was applied to the kernels. Note that, unlike the LSTM network (25) utilized by Rudy and Sapsis [8], this network does not map sequences to sequences; it simply maps values of Γ at time t to values of q at time $t + \tau$. To train the model we use both a standard and output-weighted MAE loss function:

$$MAE = \sum_j |\hat{q}_j - q_j| \quad (32)$$

$$MAE_{OW} = \sum_j \frac{|\hat{q}_j - q_j|}{f_q(q_j)} \quad (33)$$

Here \hat{q}_j is the model prediction, q_j is the training data, and $f_q(q_j)$ is the probability density function of the training data evaluated at q_j . Rudy and Sapsis [34] found that output-weighted loss functions significantly improve prediction of outlier events in a variety of flows, including airfoil and Kolmogorov flow. While those authors use the MSE, we find that in our case the MAE consistently performed slightly better. Our model is summarized graphically in Fig. 14.

To quantify the uncertainty of our model, we perform an ensemble analysis resulting in a mean prediction $\hat{q}(t)$ and variance $\sigma_q(t)$. We set aside 80% of the data set for training, and for each iteration of the ensemble we randomly select 75% of that training data (60% of the total) to use for training. The remaining 20% of the data are used for testing. All results presented here are computed exclusively using these test data.

Algorithm 2: Wavelet preprocessing

1. Input data: $P(t)$
2. Compute spectrogram: $|\hat{P}(t)|$
3. Select maximum: $\gamma(t) \equiv |\hat{P}(f, t)|_{f=f_e}$
4. Differentiate: $\Gamma = [\gamma, \dot{\gamma}]$
5. Output: $\Gamma(t)$

B. Results: Basis Comparison

To illustrate the advantages of the wavelet basis and the output-weighted loss function, we compare the model predictions using the three basis types P , \tilde{P} , and $[\gamma, \dot{\gamma}]$ and the two loss functions MAE and MAE_{OW} . In all cases we use four evenly spread sensor locations at $s = 0.05, 0.35, 0.65, \text{ and } 0.95$. The first two sensor locations represent areas identified in Secs. III.A and III.B as predictive of the drag. The latter two locations are chosen as to not neglect the underside of the airfoil. In all cases we train an ensemble of 10 models for 250 epochs.

Figures 15 and 16 compare the model prediction of the drag coefficient to the true value for $\tau = 0$ and $\tau = 7$. At $\tau = 0$, the filtered pressured model and the wavelet model, shown in blue and red, respectively, perform well, with only the raw pressure model, shown in green, suffering from significant noise corruption. In fact, in this case the smooth pressure model slightly outperforms the wavelet basis—best seen by comparing the predicted probability density functions in Fig. 17. This is because there exists an accurate linear mapping from the smooth pressure $\tilde{P}(t)$ to the drag $q(t)$, and in taking the wavelet transform of the filtered pressure some information is lost—limiting the potential accuracy of the wavelet model. See the Appendix for a brief discussion on this linear mapping. Furthermore, we observe that for zero lead time the models trained using the MAE loss function perform slightly better than those trained using the MAE_{OW} loss—The benefit is most pronounced for the raw pressure data.

The benefits of the wavelet basis and the output weighted loss functions do not become apparent until considering nonzero lead times—a reflection of the nonlinearity of the time shift operation $q(t) \rightarrow q(t + \tau)$. In this case, when using the standard MAE loss, all three models entirely fail to capture the extreme events. However, when we use the output-weighted loss, the wavelet basis retains much of the accuracy observed for $\tau = 0$, while the performance of the raw and filtered pressure models deteriorates significantly. The filtered pressure model still traces the occurrence of the extreme events, but suffers from significant noise corruption, leading to a number of false-positive predictions. This phenomenon is even more pronounced for the raw pressure model, which, as expected, suffers from even greater noise corruption. The distinction between the three basis types is less evident in the predicted probability density functions shown in 17. Here we see again that the models trained using the standard MAE loss fail to capture the tails of the distribution entirely; however, with the MAE_{OW} loss both the filtered pressure and wavelet models capture the general shape of the distribution. The wavelet model does, however, capture the small peaks around $q = -1.5$ and $q = 3$ slightly better than the others.

To quantify the forecasting capabilities of each model, we track the number of extreme events predicted as a function of time. For this purpose we define an extreme event as a local maximum, whose value is more than 2 standard deviations greater than the mean. Thus, a time instant t_j is considered to represent an extreme event t_{EE} if it satisfies the following conditions:

$$t_{EE} : t_j \text{ s.t. } \left[\frac{\partial q}{\partial t} \Big|_{t_j} = 0 \ \& \ q(t_j) > 2\sigma_q \right] \quad (34)$$

We then define the number of extreme events $N_{EE}(t_1, t_2)$ as the number of extreme events in the interval, t_1 to t_2 , or more explicitly,

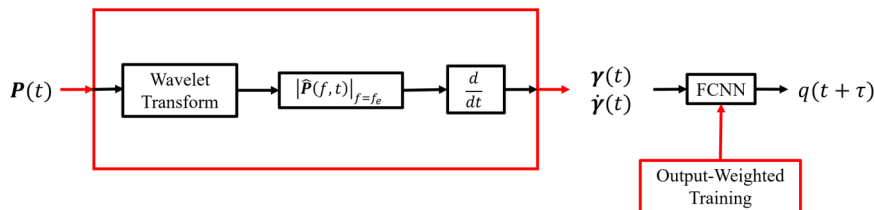


Fig. 14 Illustration of wavelet preprocessing Algorithm 2. Elements outlined in red represent the contributions of this work.

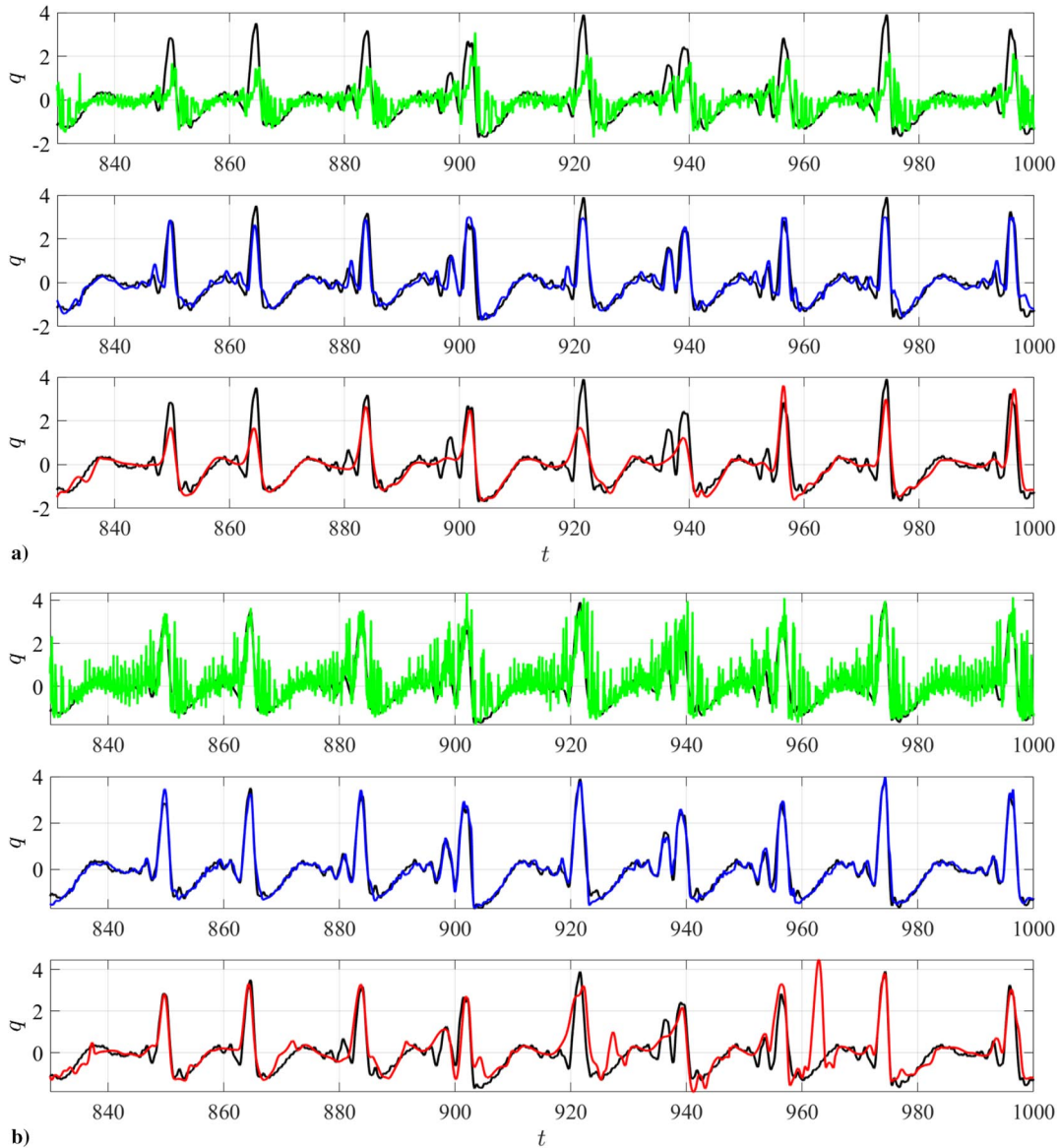


Fig. 15 Time series of predictions for raw pressure (green), filtered pressure (blue), and wavelet (red) models, compared to true values (black) for $\tau = 0$. Loss function: MAE (a); MAE_{OW} (b).

$$N(t_n, t_0) = \sum_{j=J_0}^{j_n} \delta_{t_j, t_{EE}} \quad (35)$$

where t_j is treated as a discrete series and j_0 and j_n are the indices of t_n and t_0 , respectively. To avoid overpenalizing noise, we enforce a minimum separation of the identified extreme events equal to the characteristic period of the extreme event frequency: $T_{EE} = 1/f_{EE}$. While it relies on two user-defined parameters (the extreme event threshold and the minimum peak separation), this metric provides a useful quantification of the forecasting capabilities of each model.

This metric is plotted in Fig. 18 for the filtered pressure and wavelet models for $\tau = [0, 3, 7, 10]$ —we omit the raw pressure model due to its poor performance. Since the models trained using the MAE loss perform so poorly for $\tau > 0$, we plot only the results obtained using the MAE_{OW} loss. Again we see that for $\tau = 0$ both models perform similarly, with the filtered pressure model slightly outperforming the wavelet model. However, as the lead time τ increases, the wavelet model retains much of its accuracy, while the filtered pressure model, on the other hand, dramatically overestimates the number of extreme events. This is due to the significant noise in the filtered pressure model, the magnitude of which is often comparable to the underlying signal.

We also compute the MAE , MSE , and α^* , defined in Eq. (27), as well as the error in the total number of extreme events predicted. These are plotted in Fig. 19. Again, both pressure models overestimate the number of extreme events defined by Eq. (34)—we again omit the raw pressure model due to its poor performance. However, for the aggregate error metrics, MAE , MSE , and α^* , the differences are less pronounced. Both the filtered pressure and wavelet model significantly outperform the raw pressure model, but the difference between those two is minimal.

The discrepancy in performance between the three basis types can be understood through the simple nature of the model architecture. The large amplitude of the high-frequency fluctuations in the raw pressure is comparable and sometimes even larger than the bursting amplitude—see Fig. 2—therefore a one-to-one map is destined to fail. This phenomenon is mitigated by filtering the vortex shedding frequency out of the pressure data—in this case there is indeed a linear map for zero lead time. However, for nonzero lead times the amplitude of the fluctuations at the extreme event frequency are significant enough to introduce significant ambiguity in a one-to-one map. Conversely, the wavelet basis is free of noise and fluctuates on a time scale associated with the mean time between extreme events, thereby greatly improving the feasibility of such a simple mapping.

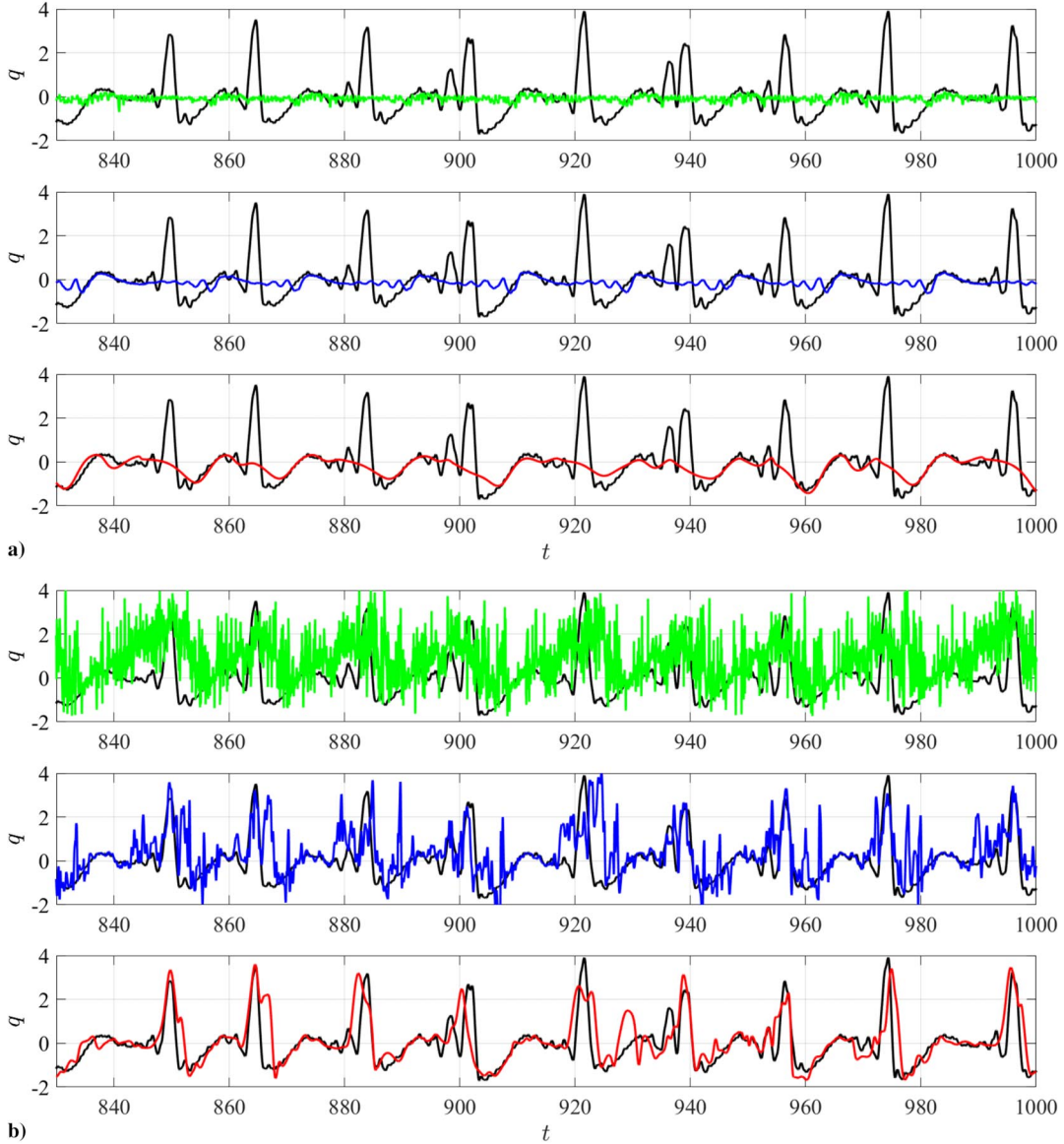


Fig. 16 Time series of predictions for raw pressure (green), filtered pressure (blue), and wavelet (red) models, compared to true values (black) for $\tau = 7$. Loss function: MAE (a); MAE_{OW} (b).

C. Results: Optimal Sensing

We now assess the potential of using the wavelet preprocessing algorithm in the context of optimal sensor selection. For the sake of brevity we focus exclusively on $\tau = 7$ and forgo comparison with the raw and filtered pressure models. As in Sec. IV we use Algorithm 1; however, in this case the acquisition function requires training the network (and is minimized, not maximized). In particular we consider the following two acquisition functions, which penalize uncertainty in the model prediction:

$$a_{iu} = \frac{1}{T} \int_0^T \sigma_q(t) dt = \frac{1}{N} \sum_{j=1}^N \sigma_{q,j} \quad (36)$$

$$a_{ow} = \frac{1}{T} \int_0^T \frac{\sigma_q(t)}{f_q(q(t))} dt = \frac{1}{N} \sum_{j=1}^N \frac{\sigma_{q,j}}{f_q(q_j)} \quad (37)$$

We refer to these as “integrated uncertainty” (iu) and “output weighted” (ow), respectively. To distinguish the effects of the loss and acquisition functions, we perform three iterations of Algorithm 1 with each acquisition function with both the MAE and MAE_{OW} loss functions. As the acquisition functions need to be evaluated at every sensor location, we use a slightly reduced ensemble of seven models

and train over only 70 epochs during the active search. Once the optimal sensor locations are found, we retrain an ensemble of 10 models using those optimal sensors locations for 200 epochs. Due to the increased cost of performing the search algorithm (as compared to the offline search in Sec. IV) and our observations that beyond three sensors the marginal improvement due to each additional sensor drops off significantly, we restrict our analysis to three sensors. The resulting optimal sensor locations are summarized in Fig. 20.

Figure 21 compares the output probability density function and time series of the mean predictions of the model trained with each loss function and each acquisition function—each using their respective optimal sensor locations. As a comparison we also include the predictions of a reference model with three evenly spaced sensors. For both the time series and the pdf, we observe that regardless of the acquisition function the models trained using the standard MAE loss fail, while the models trained with the output-weighted MAE_{OW} predict the bursting events relatively accurately. To further compare the models we plot the time series of the uncertainty bound, $\hat{q}(t) \pm \sigma_q$ in Fig. 22. Consistent with the results of Sec. IV, we observe the model to be robust to specific sensor locations, with little distinction between the three sensor distributions.

Interestingly, inspection of Figs. 21a and 22 indicates that the output-weighted acquisition function (37) performs slightly worse

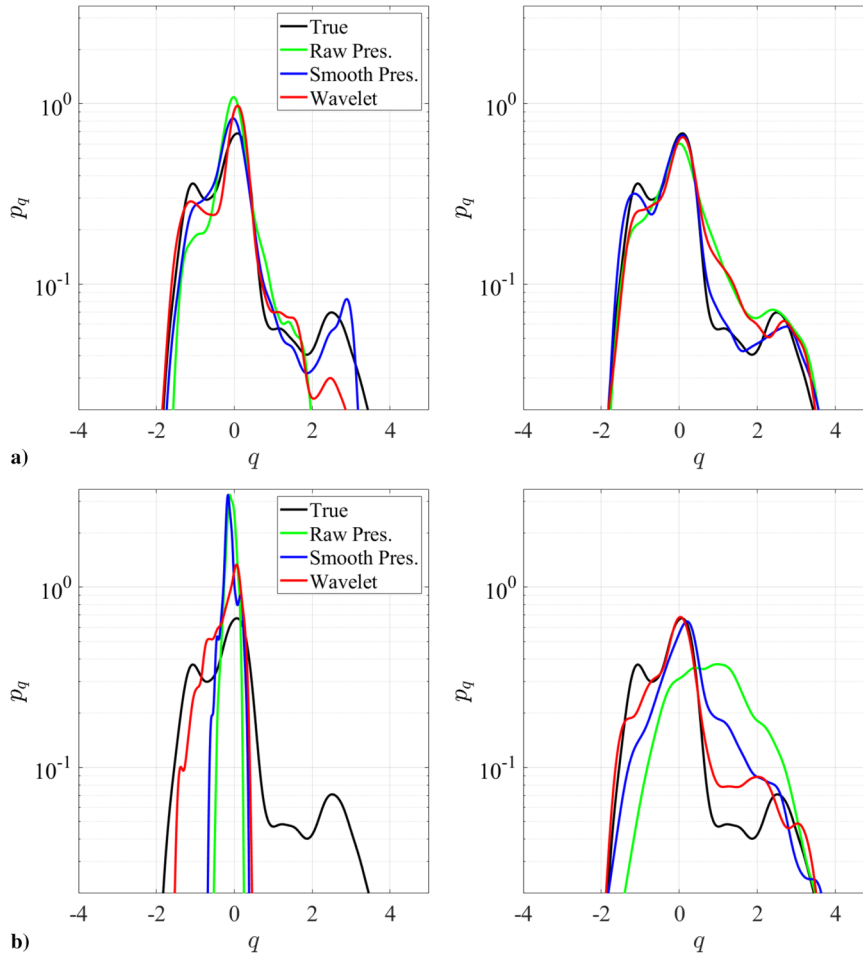


Fig. 17 Probability density function of predictions for raw pressure (green), filtered pressure (blue), and wavelet (red) models, compared to true values (black) for $\tau = 0$ (a) and $\tau = 7$ (b). Loss function: MAE (left); MAE_{OW} (right).

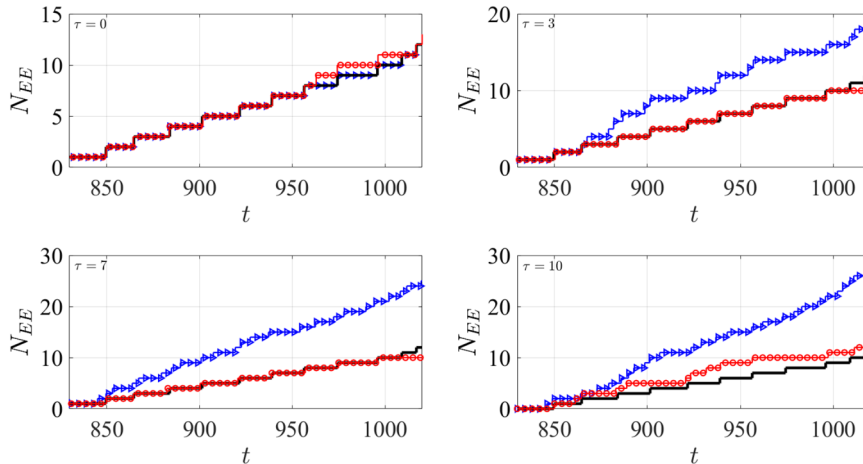


Fig. 18 Predicted number of extreme events by filtered pressure model (blue) and wavelet model (red) compared to true number (black) for from top left $\tau = [0, 3, 7, 10]$.

than the others—exhibiting some false-positive fluctuations between $t = 920$ and $t = 960$. Furthermore, the results using the non-output-weighted acquisition function (36) do not exhibit any meaningful improvement over the reference case. These findings indicate that the specific locations of the sensors are of secondary importance when compared to the effects of the output-weighted loss function and the wavelet preprocessing. The latter of which extracts the bursting events from the input data a priori. These results indicate that further emphasizing extreme events through strategies such as output-weighted optimal sensing is not only unnecessary, but

could result in a loss of accuracy during the quiescent periods. Our findings here and in Sec. IV are consistent with recent work by Zhong et al. [19], who also found that sensor location had only minor effects when using surface pressure measurements for the data-driven reconstruction of unsteady airfoil flow.

From a practical point of view, the similarity of these results is significant. The active regions of the flow, and thus the optimal sensor locations predicted here and in Sec. IV, are likely to vary with Reynolds number and angle of attack. However, aircraft experience a wide range of flow speeds and orientations, making the robustness

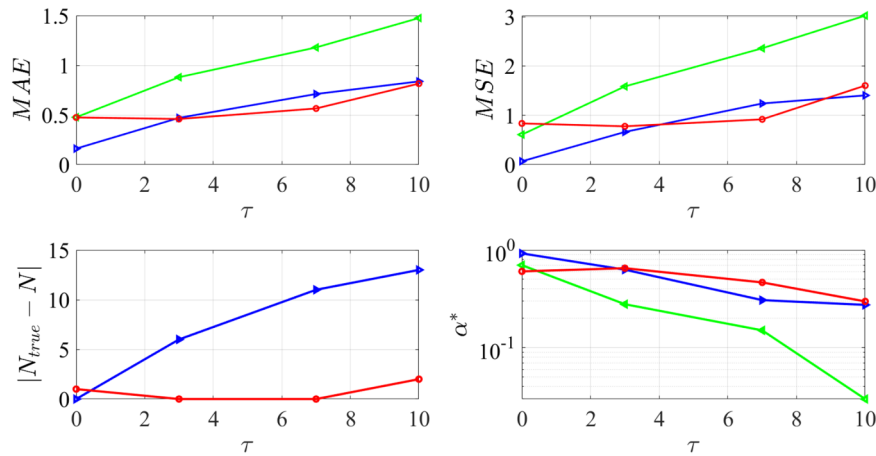


Fig. 19 From top left: MAE , MSE , error in the number of predicted extreme events, and α^* . Models: raw pressure (green), filtered pressure (blue), and wavelet (red).

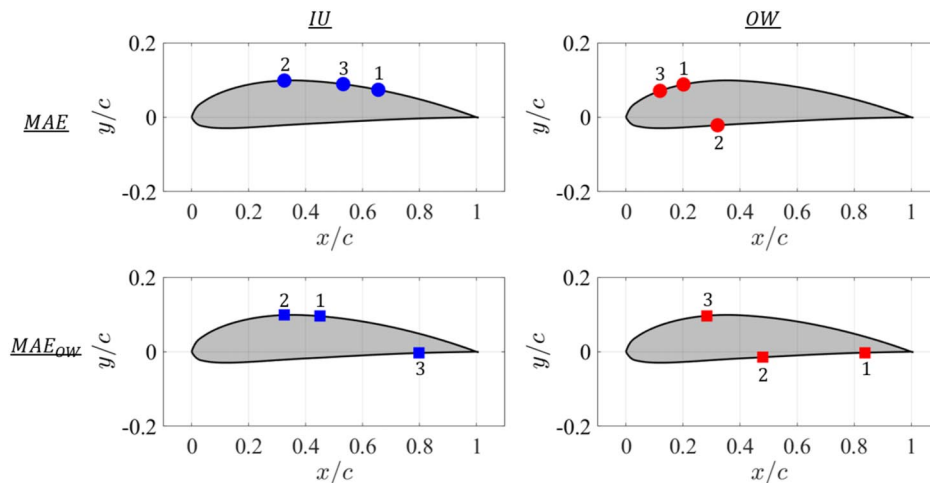


Fig. 20 Optimal sensor locations predicted using Algorithm 1 with IU (left column) and OW (right column) acquisition functions with MAE (top row) and MAE_{OW} (bottom row) loss functions.

to sensor location a valuable asset. These results support the possibility of a sparse sensing strategy that is applicable for a wide range of airfoil designs and is robust to dynamic changes in angle of attack. A parameter study over Reynolds number and flow geometry to confirm this hypothesis is beyond the scope of the present work, but is the topic of ongoing research.

VI. Conclusions

We have investigated the mechanisms driving the nonperiodic bursting phenomena observed in the two-dimensional flow over a NACA 4412 airfoil at finite angle of attack. We have conducted a detailed analysis of the spatiotemporal statistics of the airfoil surface pressure and its connection to the extreme events observed in the drag force. Through a wavelet analysis we found that the surface pressure exhibits multiscale behavior with three distinct time scales. In addition to the dominant vortex shedding frequency, the flow exhibits a slowly varying quiescent time scale and a second energetic frequency component—at approximately one-third the vortex shedding frequency. We established that the extreme excursions of the drag first observed by Rudy and Sapsis [8] correspond to instabilities of this latter frequency component in the surface pressure.

These findings were corroborated by an analysis of the wavelet transformed vorticity field. This analysis revealed that during quiescent times the extreme event frequency evolves independently of the vortex shedding frequency; however, occasionally the extreme event frequency undergoes a transient instability that links the fortunes of

these two generally disparate time scales. This instability is composed of two steps: first the extreme event frequency draws energy from the higher-frequency vortex shedding flow, and then at the extreme event frequency there is an abrupt nonlinear energy transfer from smaller to larger spatial length scales. Interestingly these findings are contrary to the far more common situation where linear instabilities transfer energy from a slowly evolving (or stationary) mean flow to faster time scales and smaller length scales. Therefore, while we have identified the slow-fast system at the heart of the bursting events, the exact mechanism by which the instabilities in the pressure and vorticity are translated to the aerodynamic forces is not yet clear, and remains the topic of ongoing research. For example, it is still unclear what causes the global (integrated over the full domain) magnitude of the extreme event mode to decrease during the extreme events—see the lower panel of Fig. 8—or why the temporal correlation of the vortex shedding mode exhibits fluctuations resembling the extreme event frequency—see the upper panel of the same figure. Furthermore, here we have considered only a single angle of attack, and further study is required to establish how the orientation of the flow impacts both the active regions of the airfoil and the extreme event frequency.

From a modeling perspective, we pursued two separate strategies. First, in Sec. IV we investigated the implications of these results for the existing LSTM architecture developed by Rudy and Sapsis [8]—which takes raw pressure as its input. We considered an optimal sensing strategy based purely on the statistics of the data, and therefore did not require the computationally costly step of

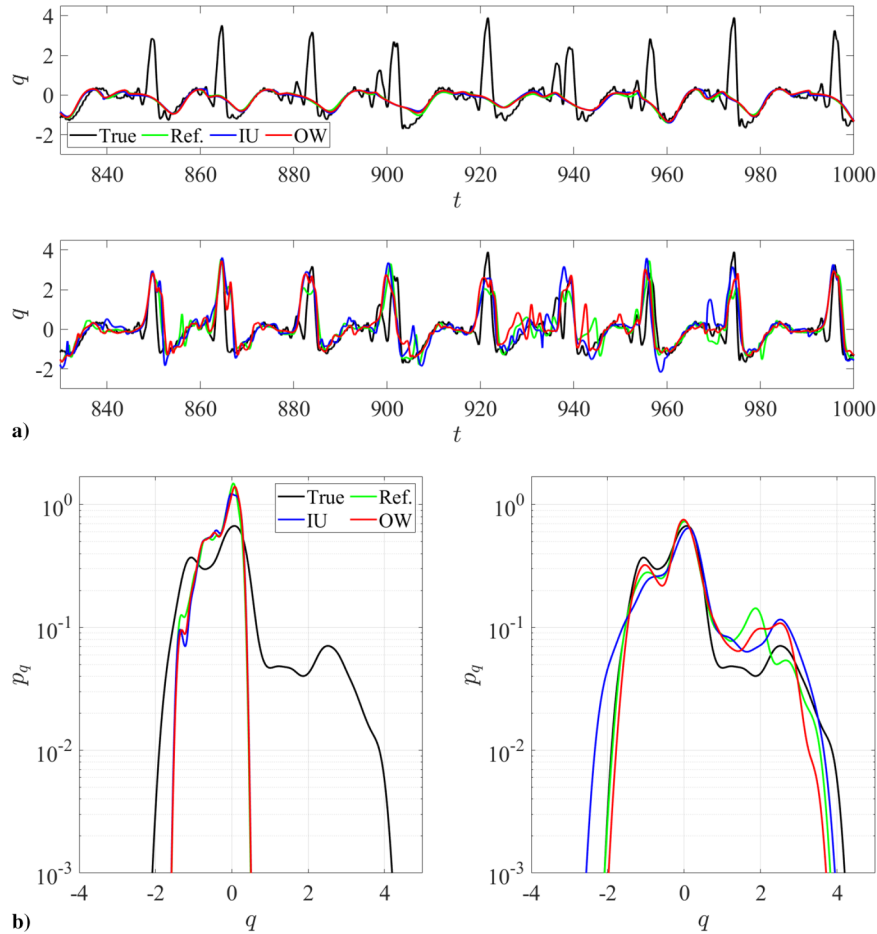


Fig. 21 Predicted time series (a) and probability density functions (b) for reference (green), *IU* (blue), and *OW* (red) acquisition functions to true values (black). *MAE* (top/left) and *MAE_{OW}* (bottom/right).

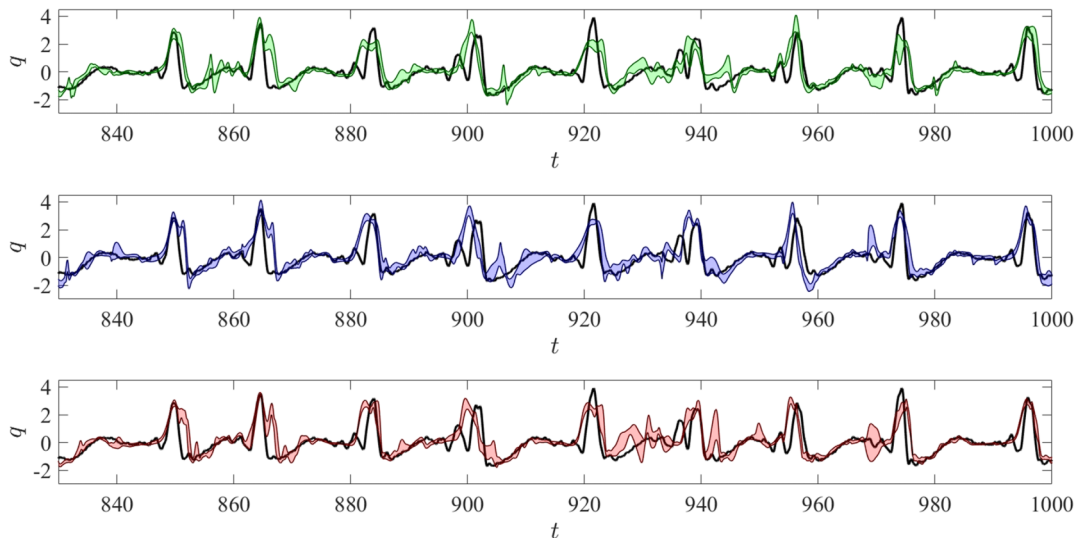


Fig. 22 Predictions for reference (green), *IU* (blue), and *OW* (red) acquisition functions to true values (black). Shaded area represents mean \pm one standard deviation of ensemble prediction.

training model. Using the LSTM model this mutual-information-based algorithm failed to predict sensor locations that performed better than a simple uniform sensor distribution. This failure of the purely mutual-information-based sensor placement demonstrated the limitations of a purely statistical offline sampling strategy and highlighted the limitations of mutual information as a practical tool. Additionally, the model complexity incurred by the LSTM layers needed to process rapidly varying time series such as the fluctuating

surface pressure remains cumbersome regardless of the sparsity of the sensor array.

Second, we also developed a preprocessing algorithm (see Fig. 14) to extract the time-varying magnitude of the extreme event frequency component from the raw pressure signal. By isolating the wavelet coefficient associated with the extreme event frequency, we eliminate the high-frequency fluctuations resulting in a signal that slowly fluctuates on a time scale associated with the mean time between extreme

events—which are by definition rare. This enables the (approximate) one-to-one mapping of the wavelet coefficient, which is free of rapid high-amplitude noise, to the drag for lead times $\tau > 0$. This then eliminates the need for a costly LSTM architecture and allows for accurate prediction using a simply connected feed forward neural network. These results are consistent with findings of Cousins and Sapsis [28,29], who used a spatial wavelet transform to extract unstable spatial length scales to efficiently predict rogue waves in variety of dynamic systems, including the Majda–McLaughlin–Tabak model and the modified nonlinear Schrödinger equation.

While this preprocessing drastically reduces the noise in the signal, it can, and in our case does, eliminate some potentially useful information as well. As noted in Sec. V.B and in the Appendix, for $\tau = 0$ there exists an accurate linear map from the surface pressure to the drag coefficient. By isolating a single wavelet coefficient, some of the information in the pressure signal is lost, leading to the wavelet model performing slightly worse than the filtered pressure model for $\tau = 0$. However, the predictions from the wavelet model are far more robust to increasing values of τ . At $\tau = 7$ the predictions of the wavelet model have degraded only slightly, while the raw and filtered pressure models exhibit significant noise corruption. The higher the frequency of oscillation, the more nonlinear the transformation $q(t) \rightarrow q(t + \tau)$ —resulting in the degradation of the filtered and raw pressure models for $\tau > 0$.

This preprocessing alleviates the need for recursive or convolutional network architectures as used by authors such as Rudy and Sapsis [8] and Hou et al. [17]. However, even with this highly extreme event-targeted algorithm, we found that training the model using an output-weighted loss function is necessary for accurate predictions. Most interestingly, we find that with these training interventions the specific locations of the sensors is of secondary importance. This is incredibly advantageous as it suggests that the predictive capabilities of our approach are robust to dynamic changes in angle of attack or freestream velocity—however, this requires further study. Our findings suggest that improving the prediction of rare events does not necessarily require more complex models, but can be achieved by identifying observables that reflect the underlying physical mechanisms and through tailored training strategies—as also discussed by various authors, including Sapsis [4], Blanchard and Sapsis [5], Rudy and Sapsis [34], and Farazmand and Sapsis [37]. We believe that the herein-proposed wavelet-based analysis is applicable to a wide range of slow-fast systems and remains the topic of ongoing research.

Appendix: Linear Mapping

Here we illustrate the linear map from the vector-valued filtered pressure signal $\tilde{P}(t)$ to the scalar drag coefficient $q(t)$. Let $\tilde{P} \in \mathbb{R}^{N \times 100}$ and $q \in \mathbb{R}^{N \times 1}$ be their discrete representations— N is the number of data points (time steps). We then seek a linear representation $\hat{q} = \tilde{P}a$ that minimizes the L_2 norm $\|q - \hat{q}\|^2$. The optimal coefficient vector is given by

$$a^* = \tilde{P}_{\text{train}}^+ q_{\text{train}} \quad (\text{A1})$$

where the superscript “+” denotes the pseudo-inverse, and the subscript “train” refers to the subset of data used for training. Here we use the first 10% of the data to fit the regression (38) and the last 20% for testing. Figure A1 compares the predictions of the linear regression to the truth for $\tau = 0$ and $\tau = 7$. For $\tau = 0$ the linear prediction is indistinguishable from the truth, while for $\tau = 7$ the linear model completely fails. This is a reflection of the highly nonlinear nature of the time shift operation $q(t) \rightarrow q(t + \tau)$.

Acknowledgments

We acknowledge support from the Army Research Office (Grant No. W911NF-17-1-0306) and the Air Force Office of Scientific Research (Grant No. MURI FA9550-21-1-0058).

References

- [1] Sapsis, T. P., “Statistics of Extreme Events in Fluid Flows and Waves,” *Annual Review of Fluid Mechanics*, Vol. 53, No. 1, 2021, pp. 85–111. <https://doi.org/10.1146/annurev-fluid-030420-032810>
- [2] Farazmand, M., and Sapsis, T. P., “Extreme Events: Mechanisms and Prediction,” *Applied Mechanics Reviews*, Vol. 71, No. 5, 2019, pp. 1–19. <https://doi.org/10.1115/1.4042065>
- [3] Mohamad, M. A., and Sapsis, T. P., “Sequential Sampling Strategy for Extreme Event Statistics in Nonlinear Dynamical Systems,” *Proceedings of the National Academy of Sciences*, Vol. 115, No. 44, 2018, pp. 11,138–11,143. <https://doi.org/10.1073/pnas.1813263115>
- [4] Sapsis, T. P., “Output-Weighted Optimal Sampling for Bayesian Regression and Rare Event Statistics Using Few Samples,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 476, No. 2234, 2020, p. 20190834. <https://doi.org/10.1098/rspa.2019.0834>
- [5] Blanchard, A., and Sapsis, T., “Output-Weighted Optimal Sampling for Bayesian Experimental Design and Uncertainty Quantification,” *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 9, No. 2, 2021, pp. 564–592. <https://doi.org/10.1137/20M1347486>
- [6] Guth, S., and Sapsis, T. P., “Machine Learning Predictors of Extreme Events Occurring in Complex Dynamical Systems,” *Entropy*, Vol. 21, No. 10, 2019, p. 925. <https://doi.org/10.3390/e21100925>
- [7] Qi, D., and Majda, A. J., “Using Machine Learning to Predict Extreme Events in Complex Systems,” *Proceedings of the National Academy of Sciences*, Vol. 117, No. 1, 2020, pp. 52–59. <https://doi.org/10.1073/pnas.1917285117>
- [8] Rudy, S. H., and Sapsis, T. P., “Prediction of Intermittent Fluctuations from Surface Pressure Measurements on a Turbulent Airfoil,” *AIAA Journal*, Vol. 60, No. 7, 2022, pp. 4174–4190. <https://doi.org/10.2514/1.J.061163>
- [9] Lissaman, P. B. S., “Low-Reynolds-Number Airfoils,” *Annual Review of Fluid Mechanics*, Vol. 15, No. 1, 1983, pp. 223–239. <https://doi.org/10.1146/annurev.fl.15.010183.001255>

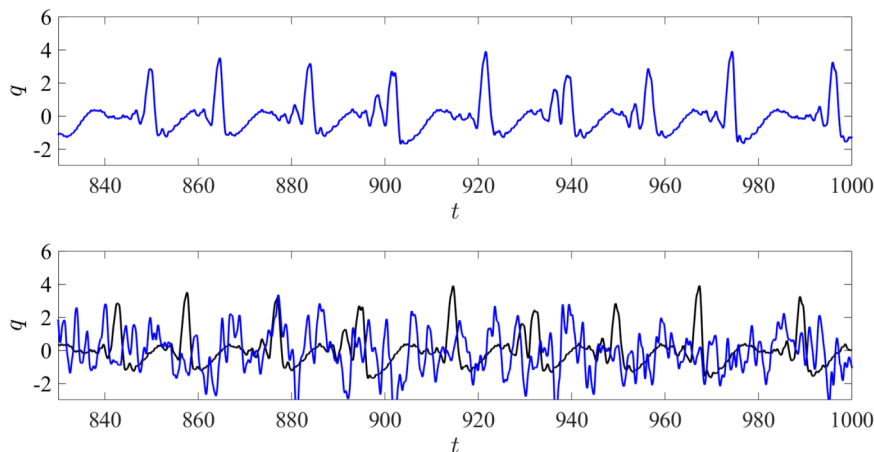


Fig. A1 Prediction of drag coefficient from linear regression model (blue) compared to true value (black) [$\tau = 0$ (upper panel); $\tau = 7$ (lower panel)].

- [10] Kim, D.-H., Chang, J.-W., and Chung, J., “Low-Reynolds-Number Effect on Aerodynamic Characteristics of a NACA 0012 Airfoil,” *Journal of Aircraft*, Vol. 48, No. 4, 2012, pp. 1212–1215. <https://doi.org/10.2514/1.C031223>
- [11] Brenner, M. P., Eldredge, J. D., and Freund, J. B., “Perspective on Machine Learning for Advancing Fluid Mechanics,” *Physical Review Fluids*, Vol. 4, No. 10, 2019, Paper 100501. <https://doi.org/10.1103/PhysRevFluids.4.100501>
- [12] Fukami, K., Fukagata, K., and Taira, K., “Assessment of Supervised Machine Learning Methods for Fluid Flows,” *Theoretical and Computational Fluid Dynamics*, Vol. 34, No. 4, 2020, pp. 497–519. <https://doi.org/10.1007/s00162-020-00518-y>
- [13] Brunton, S. L., Noack, B. R., and Koumoutsakos, P., “Machine Learning for Fluid Mechanics,” *Annual Review of Fluid Mechanics*, Vol. 52, No. 1, 2020, pp. 477–508. <https://doi.org/10.1146/annurev-fluid-010719-060214>
- [14] Gomez, D. F., Lagor, F., Kirk, P. B., Lind, A., Jones, A. R., and Paley, D. A., “Unsteady DMD-Based Flow Field Estimation from Embedded Pressure Sensors in an Actuated Airfoil,” *AIAA Scitech 2019 Forum*, AIAA Paper 2019-0346, 2019.
- [15] Maulik, R., Fukami, K., Ramachandra, N., Fukagata, K., and Taira, K., “Probabilistic Neural Networks for Fluid Flow Surrogate Modeling and Data Recovery,” *Physical Review Fluids*, Vol. 5, No. 10, 2020, Paper 104401. <https://doi.org/10.1103/PhysRevFluids.5.104401>
- [16] Dawson, S. T., Schiavone, N. K., Rowley, C. W., and Williams, D. R., “A Data-Driven Modeling Framework for Predicting Forces and Pressures on a Rapidly Pitching Airfoil,” *45th AIAA Fluid Dynamics Conference*, AIAA Paper 2015-2767, 2015.
- [17] Hou, W., Darakananda, D., and Eldredge, J. D., “Machine-Learning-Based Detection of Aerodynamic Disturbances Using Surface Pressure Measurements,” *AIAA Journal*, Vol. 57, No. 12, 2019, pp. 5079–5093. <https://doi.org/10.2514/1.J058486>
- [18] Le Provost, M., Hou, W., and Eldredge, J., “Deep Learning and Data Assimilation Approaches to Sensor Reduction in Estimation of Disturbed Separated Flows,” *AIAA Scitech 2020 Forum*, AIAA Paper 2020-0799, 2020.
- [19] Zhong, Y., Fukami, K., An, B., and Taira, K., “Sparse Sensor Reconstruction of Vortex-Impinged Airfoil Wake with Machine Learning,” *Theoretical and Computational Fluid Dynamics*, Vol. 37, April 2023, pp. 269–287.
- [20] Williamson, C. H. K., “Vortex Dynamics in the Cylinder Wake,” *Annual Review of Fluid Mechanics*, Vol. 28, No. 1, 1996, pp. 477–539. <https://doi.org/10.1146/annurev.fl.28.010196.002401>
- [21] Symon, S., Rosenberg, K., Dawson, S. T. M., and McKeon, B. J., “Non-Normality and Classification of Amplification Mechanisms in Stability and Resolvent Analysis,” *Physical Review Fluids*, Vol. 3, No. 5, 2018, Paper 053902. <https://doi.org/10.1103/PhysRevFluids.3.053902>
- [22] Farazmand, M., and Sapsis, T. P., “Closed-Loop Adaptive Control of Extreme Events in a Turbulent Flow,” *Physical Review E*, Vol. 100, No. 3, 2019, Paper 033110. <https://doi.org/10.1103/PhysRevE.100.033110>
- [23] Wang, S., Zhou, Y., Alam, M. M., and Yang, H., “Turbulent Intensity and Reynolds Number Effects on an Airfoil at Low Reynolds Numbers,” *Physics of Fluids*, Vol. 26, No. 11, 2014, Paper 115107. <https://doi.org/10.1063/1.4901969>
- [24] Gopalakrishnan Meena, M., Taira, K., and Asai, K., “Airfoil-Wake Modification with Gurney Flap at Low Reynolds Number,” *AIAA Journal*, Vol. 56, No. 4, 2018, pp. 1348–1359. <https://doi.org/10.2514/1.J056260>
- [25] Menon, K., and Mittal, R., “Aerodynamic Characteristics of Canonical Airfoils at Low Reynolds Numbers,” *AIAA Journal*, Vol. 58, No. 2, 2020, pp. 977–980. <https://doi.org/10.2514/1.J058969>
- [26] Addison, P. S., *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance*, 2nd ed., CRC Press, Boca Raton, FL, 2016, Chap. 2.
- [27] Mojahed, A., Bergman, L. A., and Vakakis, A. F., “New Inverse Wavelet Transform Method with Broad Application in Dynamics,” *Mechanical Systems and Signal Processing*, Vol. 156, July 2021, Paper 107691. <https://doi.org/10.1016/j.ymsp.2021.107691>
- [28] Cousins, W., and Sapsis, T. P., “Quantification and Prediction of Extreme Events in a One-Dimensional Nonlinear Dispersive Wave Model,” *Physica D*, Vol. 280, July 2014, pp. 48–58. <https://doi.org/10.1016/j.physd.2014.04.012>
- [29] Cousins, W., and Sapsis, T. P., “Reduced-Order Precursors of Rare Events in Unidirectional Nonlinear Water Waves,” *Journal of Fluid Mechanics*, Vol. 790, March 2016, pp. 368–388. <https://doi.org/10.1017/jfm.2016.13>
- [30] Baymndr, C., “Early Detection of Rogue Waves by the Wavelet Transforms,” *Journal of Signal Processing Systems*, Vol. 380, Nos. 1–2, 2016, pp. 156–161.
- [31] Srirangarajan, S., Allen, M., Preis, A., Iqbal, M., Lim, H. B., and Whittle, A. J., “Wavelet-Based Burst Event Detection and Localization in Water Distribution Systems,” *Journal of Signal Processing Systems*, Vol. 72, No. 1, 2013, pp. 1–16. <https://doi.org/10.1007/s11265-012-0690-6>
- [32] Rossi, F., Lendasse, A., François, D., Wertz, V., and Verleysen, M., “Mutual Information for the Selection of Relevant Variables in Spectrometric Nonlinear Modelling,” *Chemometrics and Intelligent Laboratory Systems*, Vol. 80, No. 2, 2006, pp. 215–226. <https://doi.org/10.1016/j.chemolab.2005.06.010>
- [33] Shakibian, H., and Moghadam Charkari, N., “Mutual Information Model for Link Prediction in Heterogeneous Complex Networks,” *Scientific Reports*, Vol. 7, No. 1, 2017, Paper 44981. <https://doi.org/10.1038/srep44981>
- [34] Rudy, S., and Sapsis, T., “Output-Weighted and Relative Entropy Loss Functions for Deep Learning Precursors of Extreme Events,” *Physica D: Nonlinear Phenomena*, Vol. 443, Jan. 2017, Paper 133570.
- [35] Paul, F., Fischer, J. W. L., and Kerkemeier, S. G., “Nek5000 Webpage,” 2008, <http://nek5000.mcs.anl.gov>.
- [36] Dong, S., “A Convective-Like Energy-Stable Open Boundary Condition for Simulations of Incompressible Flows,” *Journal of Computational Physics*, Vol. 302, Dec. 2015, pp. 300–328. <https://doi.org/10.1016/j.jcp.2015.09.017>
- [37] Farazmand, M., and Sapsis, T. P., “A Variational Approach to Probing Extreme Events in Turbulent Dynamical Systems,” *Science Advances*, Vol. 3, No. 9, 2017, Paper e1701533. <https://doi.org/10.1126/sciadv.1701533>
- [38] Wan, Z. Y., Vlachas, P., Koumoutsakos, P., and Sapsis, T., “Data-Assisted Reduced-Order Modeling of Extreme Events in Complex Dynamical Systems,” *PLoS ONE*, Vol. 13, No. 5, 2018, Paper e0197704. <https://doi.org/10.1371/journal.pone.0197704>
- [39] Chaloner, K., and Verdinelli, I., “Bayesian Experimental Design: A Review,” *Statistical Science*, Vol. 10, No. 3, 1995, pp. 273–304. <https://doi.org/10.1214/ss/1177009939>
- [40] Sapsis, T., “Output-Weighted Optimal Sampling for Bayesian Regression and Rare Event Statistics Using Few Samples,” *Proceedings of the Royal Society A*, Vol. 476, No. 2234, 2020, Paper 2019-0834. <https://doi.org/10.1098/rspa.2019.0834>
- [41] Yang, Y., Blanchard, A., Sapsis, T., and Perdikaris, P., “Output-Weighted Sampling for Multi-Armed Bandits with Extreme Payoffs,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 478, No. 2260, 2022, Paper 20210781. <https://doi.org/10.1098/rspa.2021.0781>
- [42] Ramachandran, P., Zoph, B., and Le, Q. V., “Searching for Activation Functions,” arXiv preprint arXiv:1710.05941, Oct. 2017. <https://doi.org/10.48550/arXiv.1710.05941>

K. Taira
Associate Editor