



AFRL-AFOSR-VA-TR-2022-0435

Flow Physics and Distillation of the Gust-Induced Stall of a Low Aspect Ratio Wing

Eldredge, Jeffrey
UNIVERSITY OF CALIFORNIA LOS ANGELES
11000 KINROSS AVE STE 102
LOS ANGELES, CA,
US

07/27/2022
Final Technical Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
Air Force Office of Scientific Research
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE 20220727	2. REPORT TYPE Final	3. DATES COVERED	
		START DATE 20180801	END DATE 20211031
4. TITLE AND SUBTITLE Flow Physics and Distillation of the Gust-Induced Stall of a Low Aspect Ratio Wing			
5a. CONTRACT NUMBER	5b. GRANT NUMBER FA9550-18-1-0440	5c. PROGRAM ELEMENT NUMBER 61102F	
5d. PROJECT NUMBER	5e. TASK NUMBER	5f. WORK UNIT NUMBER	
6. AUTHOR(S) Jeffrey Eldredge			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) UNIVERSITY OF CALIFORNIA LOS ANGELES 11000 KINROSS AVE STE 102 LOS ANGELES, CA US			8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research 875 N. Randolph St. Room 3112 Arlington, VA 22203		10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/AFOSR RTA1	11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-AFOSR-VA- TR-2022-0435

12. DISTRIBUTION/AVAILABILITY STATEMENT

A Distribution Unlimited: PB Public Release

13. SUPPLEMENTARY NOTES**14. ABSTRACT**

The unsteady aerodynamics of large-amplitude flow disturbances (i.e., gusts) pose significant challenges for mitigating their effects on agile flight vehicles. In this report, we detail our team's efforts at addressing several aspects of these challenges: investigating the flow physics of gust-wing interactions and developing a suite of tools for practically estimating the aerodynamic flow response to this interaction. We have developed a unique, well-instrumented facility to generate parametric families of transverse gusts. By employing a suction and a controllable system of louvers, the facility generates irrotational gusts, a distinct contrast from other facilities. The irrotational gusts are highly repeatable and are conducive to unsteady aerodynamic modeling. We have used the facility to build a database of experimental results, clearly showing gust signatures in the extensive surface pressure data. To serve as a complement to the experimental studies, we have developed (in concert with a companion program) and utilized a state-of-the-art three-dimensional flow solver that can achieve direct numerical simulation (DNS) predictions at comparable Reynolds numbers to experiments, and can readily introduce irrotational gusts of the form of the experiments into the simulations. We have refined the use of surface pressure sensors to infer the evolving flow state in a gust encounter with the ensemble Kalman filter (EnKF), using inviscid vortex models or coarse-grid CFD as a basis for the dynamical estimate of the flow. In vortex-based estimation, we have found in our demonstrations that we can avoid a direct representation of the gust itself in the model. By leveraging the known relationships between vorticity and pressure, we have developed a low-rank EnKF that identifies and exploits the small number of informative directions for the state update of the flow, thereby utilizing sensor data much more effectively. For coarse-grid CFD estimates, we have also developed a framework for simultaneously estimating the bias error in the state and observations, also by exploiting low-rank representations, reducing error by up to 70 percent.

2

15. SUBJECT TERMS**16. SECURITY CLASSIFICATION OF:****a. REPORT**
U**b. ABSTRACT**
U**c. THIS PAGE**
U**17. LIMITATION OF ABSTRACT**
UU**18. NUMBER OF PAGES**
77**19a. NAME OF RESPONSIBLE PERSON**
GREGG ABATE**19b. PHONE NUMBER (Include area code)**
425-1779

Final Technical Report

**Flow physics and distillation
of the gust-induced stall of a low aspect ratio wing**

Grant Number: FA9550-18-1-0440

1 August 2018 – 31 October 2021

PI: Jeff D. Eldredge
Mechanical & Aerospace Engineering
University of California, Los Angeles
Los Angeles, CA 90095-1597
Tel: 1-310-206-5094
eldredge@seas.ucla.edu

co-PI: Tim Colonius
Mechanical & Civil Engineering
California Institute of Technology
colonius@caltech.edu

co-PI: David Williams
Mechanical, Materials & Aerospace Engineering
Illinois Institute of Technology
williamsd@iit.edu

Prepared for:
Air Force Office of Scientific Research
Engineering and Complex Systems (RTA1)
Unsteady Aerodynamics & Turbulent Flows
875 North Randolph Street, RM 3112
Arlington, VA 22203-1954
Grant Monitor: Dr. Gregg Abate

Abstract

The unsteady aerodynamics of large-amplitude flow disturbances (i.e., gusts) pose significant challenges for mitigating their effects on agile flight vehicles. In this report, we detail our team's efforts at addressing several aspects of these challenges: investigating the flow physics of gust-wing interactions and developing a suite of tools for practically estimating the aerodynamic flow response to this interaction. We have developed a unique, well-instrumented facility to generate parametric families of transverse gusts. By employing a suction and a controllable system of louvers, the facility generates irrotational gusts, a distinct contrast from other facilities. The irrotational gusts are highly repeatable and are conducive to unsteady aerodynamic modeling. We have used the facility to build a database of experimental results, clearly showing gust signatures in the extensive surface pressure data. To serve as a complement to the experimental studies, we have developed (in concert with a companion program) and utilized a state-of-the-art three-dimensional flow solver that can achieve direct numerical simulation (DNS) predictions at comparable Reynolds numbers to experiments, and can readily introduce irrotational gusts of the form of the experiments into the simulations. We have refined the use of surface pressure sensors to infer the evolving flow state in a gust encounter with the ensemble Kalman filter (EnKF), using inviscid vortex models or coarse-grid CFD as a basis for the dynamical estimate of the flow. In vortex-based estimation, we have found in our demonstrations that we can avoid a direct representation of the gust itself in the model. By leveraging the known relationships between vorticity and pressure, we have developed a low-rank EnKF that identifies and exploits the small number of informative directions for the state update of the flow, thereby utilizing sensor data much more effectively. For coarse-grid CFD estimates, we have also developed a framework for simultaneously estimating the bias error in the state and observations, also by exploiting low-rank representations, reducing error by up to 70 percent.

Contents

1	Introduction	4
1.1	Nonlinear gusts and the consequences of losing linearity	4
1.2	A brief review of current efforts	5
1.3	Achievements of this program	6
2	Experimental studies of transverse gust interactions	7
2.1	Development and characterization of the cross-flow wind tunnel	7
2.1.1	Modification of the Andrew Fejer Unsteady Wind Tunnel (AFUWT)	7
2.1.2	Cross flow generation	8
2.1.3	Potential flow model of test section	9
2.1.4	Experimental characterization	12
2.2	Transverse gust interactions with a low aspect ratio wing	18
2.2.1	Test cases	19
2.2.2	Results	20
3	High-fidelity simulations of low aspect ratio aerodynamics	22
4	Estimation of disturbed flows from surface pressure sensors	25
4.1	Ensemble Kalman filtering methodology	28
4.1.1	A review of the filtering problem	28
4.1.2	Additional notation	31
4.1.3	The stochastic ensemble Kalman filter	31
4.1.4	The ensemble transform Kalman filter	33
4.1.5	Covariance inflation	35
4.2	Flow estimation with an ensemble of aggregated vortex models	36
4.2.1	The aggregated vortex model	36
4.2.2	Application to gust-wing interactions	40
4.3	A low-rank ensemble Kalman filter for vortex estimation	47
4.3.1	Low-rank assimilation for the linear–Gaussian case	50
4.3.2	Low-rank assimilation for the nonlinear–Gaussian case	51
4.3.3	Estimation of the leading directions from samples	53
4.3.4	Example: inference of point vortices from wall pressure observations	53
4.4	EnKF-based estimation of bias error in coarse-grid CFD	57
4.4.1	Discretization error as a source of bias	59
4.4.2	Low-rank representation of the bias	60
4.4.3	A Bias-aware EnKF	61
4.4.4	Numerical experiment on flow past an airfoil	63
4.5	Conclusions	66
A	Ensemble Kalman filter algorithms	69
B	References	71

1 Introduction

Studies of the unsteady aerodynamics of wings interacting with various types of atmospheric disturbances (i.e., *gusts*) have been prompted by engineering needs since the early days of human-developed flight. Driven by increasing demand for the development of modern high performance aircraft, unmanned aerial systems, micro aerial vehicles, rotorcraft, and wind turbines, the understanding of the physics underlying wing-gust interactions and the strategies for managing it are central to a variety of pressing technical needs in aerodynamics, including lift stabilization and enhancement, agile maneuverability, and energy extraction and harvesting. Gust alleviation (and exploitation) are essential considerations for active flow control and control-embedded design optimization [57, 58]. In short, it is as critical as ever to understand an aircraft’s response to an incident gust.

1.1 Nonlinear gusts and the consequences of losing linearity

Until recently, most existing frameworks for predicting the response of the aircraft to a gust have been based on a linearized treatment of the problem. In other words, the gust is assumed to be a weak disturbance (compared with the flight speed of the aircraft). Thus, its effect on the aerodynamic force and moment is linearly proportional to its magnitude, so the response to gusts of different strengths are simply re-scalings of a unit response. Another important consequence of this assumption is that the flow induced by a gust arriving at time t has no direct influence on the flow induced by a gust arriving at some later time; the same is true for gusts of different frequencies. Thus, the linearity assumption enables a rich and tractable analytical framework, based fundamentally on the superposition of responses. In the frequency domain, this superposition is captured by a transfer function between the input (gust power spectrum) and the output (force response power spectrum). For example, the relationship between the local sectional lift and the vertical gust velocity for a flat plate at zero mean angle of attack is described by

$$\Phi_l(\Omega, y) = \pi^2 \rho^2 U^2 c^2(y) |H(\Omega c(y))|^2 \Phi_w(\Omega), \quad (1)$$

where the transfer function, $H(\kappa)$, is the Sears function [84]—the response to a sinusoidal vertical gust. For more complex lifting surfaces, the transfer function would be obtained from numerical simulation, for example. The lift power spectrum is used for aircraft certification; see, e.g., FAA regulations for continuous turbulence design criteria (Section 25.341(b) of [39]). For a sinusoidal oscillation in the streamwise component of incident velocity, Greenberg’s analysis provides the analogous transfer function [43]. This superposed aerodynamic response can also be expressed in the time domain, via a Duhamel integral. For example, the lift response at a given spanwise location y to a vertical gust can be computed from

$$F_l(t, y) = \pi \rho U c(y) \int_{-\infty}^t \dot{w}(\tau, y) h[(t - \tau)U/c(y), y] \, d\tau, \quad (2)$$

where $c(y)$ is the local chord length, $\dot{w}$ is the rate of change of the vertical gust velocity, and $h(t^+, y)$ is the indicial response function at y : the sectional lift that results from a unit step change in the vertical velocity—that is, a sharp-edged gust. For a flat plate at zero angle of attack, this indicial response is given by the Küssner function [61]. During the first convective time unit, $0 \leq t^+ < 1$, the leading edge of the gust is still making its way along the wing chord, which explains why this function starts at zero, rather than at $1/2$ as for the Wagner function. The Küssner and Sears functions are equivalent modulo a sine transform. In a more sophisticated lifting surface, there is a similar relationship between the surface’s indicial response and transfer function [28]. Thus, a single response function encodes the complete behavior of the lifting surface to all gusts.

The assumptions that underlie these linear analyses of gusts are questionable in smaller aircraft, particular unmanned vehicles, where the flight speed is comparable to or even smaller than the gust velocity. For example, if one considers a discrete gust described by the simple 1-cos profile, the angle of attack is effectively increased by $\arctan(w(t)/U)$, at a dimensionless pitch rate proportional to $w_0 c/UH$. Thus, early in the encounter—before the aircraft’s own vertical velocity has been affected by the lift induced by the gust—the wing would be prone to exhibit dynamic stall as the gust size H becomes smaller relative to the chord and/or as the gust amplitude becomes large compared to the flight speed. Dynamic stall is characterized by transient flow separation from the leading edge, and this flow response is generally not proportional to the gust amplitude. (For evidence, see [9], and the related problem of wing pitch-up, in which pitch rate acts as a surrogate for gust amplitude [30].) In other words, we can no longer rely on linearity to obtain aerodynamic responses to general gust conditions based on some common indicial response.

The lack of linearity in the response makes the analysis of large-amplitude gust encounters much more challenging. The gusts encountered by an aircraft are numerous in their variety, spanning a broad spectrum of length scales and magnitudes in all three velocity components, and eliciting a range of aerodynamic responses. Without a basic response from which others can be built, one is ostensibly faced with the challenging exercise of subjecting the wing to disturbances sampled from a very large-dimensional parametric space.

1.2 A brief review of current efforts

Because of the challenges of large-amplitude encounters, there have been a significant number of efforts in recent years to experimentally and computationally study gust interactions in a wide variety of forms, including transverse velocity disturbances [15, 85, 75], vortex wake disturbances generated behind an upstream obstacle or other wing [80, 4], and streamwise velocity disturbances [46, 48, 19] encounters, to name a few. The recent paper by Jones et al. [58] provides an excellent and comprehensive review. It should be noted that most experimental facilities devoted to transverse gust encounters have generated these disturbances in a vortical form, generally by creating a jet emerging from the wall of the test section [15, 85]. The current report details our efforts to design, characterize, and utilize a novel *irrotational* transverse gust facility at IIT, using suction through the wall of the test section and louvers to control the gust’s profile. The unique aspects of this facility are that the gust encounters are highly repeatable and the subsequent analysis is simpler compared to vortical disturbances, in which the gust flow is deformed by the wing’s presence.

The work on our experimental effort will be described in Section 2. In that section, we will first discuss the characterization of the facility, and particularly, the calibration of louver motion for obtaining gusts of predictable amplitude and traveling speed [49]. Then we will present our results from using this facility in an experimental investigation of an aspect ratio 2 wing subjected to transverse gusts [47]. The wing in these studies was outfitted with an extensive array of surface pressure transducers, and we will show that the data from these transducers exhibit the signature of the overall flow response.

The multitude of experimental facilities with different strategies for synthesizing gusts creates an underlying challenge for understanding how to compare these facilities’ results and interpret them in the context of real flight. Ostensibly, high-fidelity computations can be used to provide some connective tissue between the various experimental efforts. However, computational tools are not immune from their own challenges in studying the physics of gust encounters. Though a gust’s encounter with a wing may be fairly localized to the wing’s vicinity and wake, the development and larger structure of the gust itself is often important for faithfully predicting this encounter. Therefore, there is an inherent multi-scale aspect to gust encounters. In a conventional grid-based CFD study, the gust must be introduced either through boundary conditions or internal forcing, so

only a limited portion of the gust may be captured. This may miss key aspects of the gust’s behavior in experiments or in free flight. In this report, we describe the application of a new computational framework developed at Caltech, utilizing a lattice Green’s function and adaptive mesh refinement to enable multi-scale direct numerical simulations of flow physics at Reynolds numbers approaching those of experiments, and an immersed boundary projection method to ensure versatile geometries [99, 55]. A summary of this tool’s features and its application to a representative problem of low aspect ratio aerodynamics is presented in Chapter 3.

In spite of the multitude of experimental and computational studies, there remains a considerable gap to practical flight control for an aircraft in uncertain gust conditions. Due to the breadth of responses, a flight controller that would seek to robustly preserve smooth flight would need to react to different flow regimes with different strategies. The flight controller must be able to rapidly detect and classify the disturbance—with a finite number of sensors—and thence determine the appropriate control strategy. As we have discussed, this classification problem is intractably large in its native parameters. Indeed, even for a gust of relatively simple structure, such as a vortex, only in exceptional circumstances can one independently discern its length scale, orientation, position and strength from a small number of surface pressures, and this requires knowing a priori that the disturbance is a vortex [50].

But it is unclear that we need to estimate the gust’s parameters for control to be successful; we may only need to discern the behavior of the wing’s *response*—in the flow field or in the pressure exerted on the wing surface—and this response might be more readily categorized. Therefore, an essential aspect of managing gust interactions is the estimation of the wing’s flow response, using available sensors. In this report, we detail our efforts at UCLA and Caltech to infer the unsteady flow response to disturbances from surface-based pressure sensors [63, 20, 65]. Most of these efforts have been carried out with an eye towards practicable approaches for flight, i.e., fast computations from limited sensors. In particular, we have significantly advanced the state of the art on this estimation problem by identifying and exploiting a low-dimensional subspace that contains the small number “informative” directions of pressure sensors’ updates on the evolving flow state [65]. We have also developed techniques for estimating the bias error in the state [20]. We will present our work on these applications and extensions of the flow estimation problem in Section 4.

1.3 Achievements of this program

Overall, this program has resulted in the following important achievements:

- We have developed a unique, well-instrumented facility at IIT to generate parametric families of transverse and streamwise gusts, both discrete and broadband. The generated gusts are irrotational—a unique feature that distinguishes the facility from others—and therefore highly repeatable with little uncertainty [49].
- We have built a database of experimental results, clearly showing gust signatures in the surface pressure data [47].
- We have developed a state-of-the-art three-dimensional flow solver that can achieve direct numerical simulation (DNS) predictions at comparable Reynolds numbers to experiments. The flow solver can impose a variety of gusts, including irrotational forms in the wind tunnel [55].
- Furthermore, we have also developed an efficient grid-based inviscid vortex modeling framework, based on similar methods to the DNS [3].
- We have refined the use of surface pressure sensors to infer the evolving flow state in a gust encounter, using inviscid vortex models [63] or coarse-grid CFD [20]. By leveraging the known

relationships between vorticity and pressure, we have learned how to use the sensor data much more effectively, enabling us to train low-order interpretable (physics-based) models of the disturbed flow from limited measurement data and modest ensemble sizes [65].

In the following sections, we will describe these achievements in more detail.

2 Experimental studies of transverse gust interactions

In this section we will describe the development of a wind tunnel facility at IIT for generating irrotational cross-flow disturbances and the application of this facility to investigate the aerodynamic response of an aspect ratio 2 rectangular wing.

2.1 Development and characterization of the cross-flow wind tunnel

This section summarizes our work reported in [49] on the development and characterization of a wind tunnel facility for generating irrotational cross-flow gusts.

2.1.1 Modification of the Andrew Fejer Unsteady Wind Tunnel (AFUWT)

The original AFUWT incorporated a set of 4 louvers at the downstream end of the test section to enable control of the freestream flow speed, and generate longitudinal velocity fluctuations with desired velocity profiles or gust spectra (details are described in He et al. [48]). The latest upgrade of the wind tunnel includes the construction of a suction duct onto the top of the test section. The diffuser and the main fan were modified or exchanged to meet the new working requirement, including the structure of the diffuser, flow rate capacity of the fan, as well as fan blade pitch function which allows the blade pitch angle to be adjusted to adapt to low or high flow speed conditions.

Figure 1 shows a schematic diagram of the modified sections of the AFUWT. The features relevant to the main flow passing through the test section were retained from the previous configuration, including the downstream louvers, and the breather (which was always closed in the experiments presented in this report) to the diffuser. The ceiling of the test section was replaced by a sealed duct that covers the entire top of the test section with 10 sub-channels at the inlet, angled at 45° to the horizontal. Each channel consists of a louver mechanism, which through its rotation is able to change the blockage of the channel from completely blocked to fully open. The suction duct returns to the diffuser through a straight channel and a curved exhaust. By sharing a pressure drop that was generated by the main fan, the duct created a certain amount of suction from the top of the test section, hence a vertical velocity component (cross flow) is generated in the test section in addition to the streamwise flow, the principles of which are discussed in Section 2.1.1. The flow that diverges between the test section and suction duct merge in the diffuser, and return to the fan to complete the wind tunnel circuit.

The detailed configuration of the test section and the suction duct are shown in Figure 2. The test section had a length of $LTS = 2.1$ m and cross-section dimensions of $0.61 \text{ m} \times 0.61 \text{ m}$. The coordinates are defined such that the x axis is in the streamwise (horizontal) direction and the y axis is in the vertical direction orientated at the bottom left corner of the test section. The distances in x and y directions are normalized by test-section length LTS and height HTS respectively. A layer of honeycomb was placed in between the test section and the inlet of the suction duct to reduce the turbulent disturbances from the suction. Each louver was individually driven by a Nanotec® PD4-C6018L4204-E-01 servo motor, which allowed for independent control of the blockage area in each sub-channel. As a louver rotates from 0° (parallel to sub-channel walls) to 90° (perpendicular to subchannel walls), the blockage ratio in its sub-channel varies from 0 to 1. The motors were controlled through a dSPACE® DS1104 control box with a sampling rate of $F_s = 100$ samples/s. Despite the potential coupling between the test section and the suction

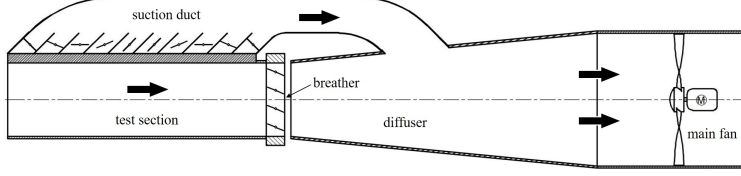


Figure 1: Modified sections of the Andrew Fejer Unsteady Wind Tunnel at IIT.

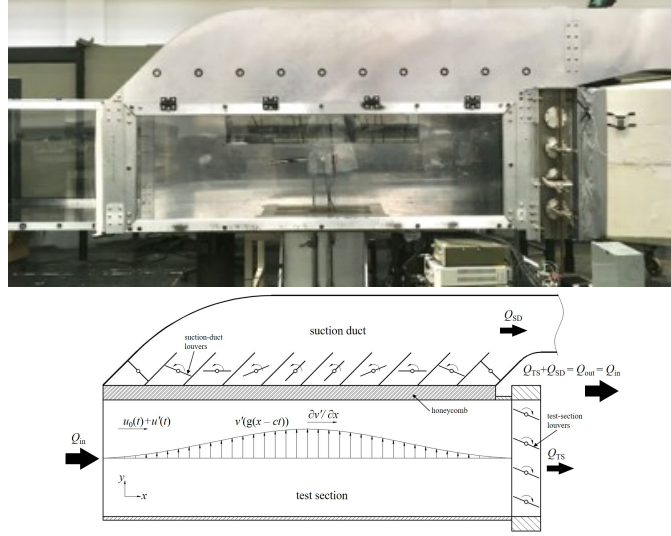


Figure 2: Test section details.

duct (as they share the same pressure source), the test-section (downstream) louvers mainly control the mean flow u_0 and the longitudinal velocity fluctuation u' , while the suction-duct (top) louvers mainly enable temporal and spatial variance of the vertical velocity fluctuation v' . In the results presented in this report, the velocity quantities are normalized by the mean flow speed u_0 such that $u^* = u/u_0$ and $v^* = v/u_0$, the z vorticity is normalized by the test-section length and mean flow speed where $\omega_z^* = \omega_z L_{TS}/u_0$ and time is scaled based on the convective time $t_c = L_{TS}/u_0$. The velocity components in the test section were measured by a Dantec Dynamics® MiniCTA 54T42 x-wire anemometer as well as an Aeroprobe® Air Data Probe (5 holes). The data acquisition was proceeded through a dSPACE® MicroLabBox® system at $F_s = 100$ samples/s.

2.1.2 Cross flow generation

The vertical velocity fluctuation v' in the test section is created by the suction from the suction duct. As the main fan runs, it creates a pressure drop Δp between the inlet of the test section and the diffuser. This pressure drop moves the air in the tunnel from the test section inlet to the diffuser exhaust. When the breather is closed, the flow rate Q_{in} at the inlet of the test section equals to the flow rate Q_{out} where the two branches of flows merge in the diffuser. Since the suction duct constructs a parallel flow path to the test section, the total flow rate Q_{in} is shared by the test section (Q_{TS}) and the suction duct (Q_{SD}), where $Q_{in} = Q_{TS} + Q_{SD}$ (see Figure 2 for parameter locations). While the drop in the test-section flow rate slows the flow speed in the longitudinal direction, the suction generated by the flow that goes into the suction duct drives the flow to move vertically in the test section, in addition to the horizontal movement. As a result, the cross-flow feature of the wind tunnel is achieved by creating a vertical velocity component to the flow velocity. Note that

the vertical velocity fluctuation is one-sided, since the suction on the top can only create positive velocity in the y direction. The major factor that affects the values of Q_{TS} and Q_{SD} is the fluid resistance of the test section R_{TS} and the suction duct R_{SD} , as well as the variable fluid resistance of the downstream louvers R_{LT} and the top louvers R_{LS} . Once the resistance is determined, the flow rates in these two sections are fixed in principle. On the other hand, downstream louvers and top louvers enables us to change the ratio of the resistance $(R_{TS} + R_{LT})/(R_{SD} + R_{LS})$ by changing the blockage areas. As the resistance ratio changes, Q_{TS} and Q_{SD} change accordingly and can be controlled in designated manners within the system limits. Consequently, the u and v velocity components in the test section are controlled. Since each louver is independently driven by a servo motor, this allows for individual control of the louver blockage. For example, the top louvers can rotate nonuniformly as a function of the x coordinate. In other words, the suction/flow rate can be spatially distributed along the suction-duct inlet (the top of the test section) in the streamwise direction as $\tilde{Q}_{SD}(t, x)$, where $\int_0^{L_{TS}} \tilde{Q}_{SD} dx = Q_{SD}$. The distributed flow rate leads to a nonuniform suction for the flow in the test section from the top; therefore, the vertical velocity fluctuation in the test section v' varies as a function of both t and x . If we enforce a traveling-wave function to the louver motions (flow rate distribution), the propagated cross-flow fluctuation can be written in the form of $v'(g(x/c_w t))$, in which $g(x)$ is the wave form and c_w is the wave speed. Further, when a sinusoidal waveform is applied, the one-sided v' is expressed as

$$v' = \frac{1}{2} v_{\max} \left[1 - \cos \left(\frac{2\pi f_w}{c_w} x - 2\pi f_w t \right) \right] \quad (3)$$

where f_w is the wave frequency. Letting $c_w = u_0$, the cross-flow becomes a sinusoidal wave traveling with the mean flow.

2.1.3 Potential flow model of test section

The flow field in the test section can be modeled as a superposition of a uniform flow in the horizontal direction and a distribution of sinks on the top in a channel [49]. Using a methodology similar to that proposed in Greengard [44], the boundary conditions of the channel can be enforced through the use of an infinite number of image sources/sinks above and below the actual cross flow location. Figure 3 shows this formulation with 10 sinks and some of their images, showing the transverse gust velocity field without a background freestream. Working in the complex plane (with complex variable $\tilde{z}$), we consider a channel with lower and upper walls at $\tilde{z} = 0$ and $\tilde{z} = iH_{TS}$ respectively. A source/sink at a location $\tilde{z}_{sk}$ induces a complex velocity

$$w = \frac{\tilde{Q}_{sk}}{2\pi(\tilde{z} - \tilde{z}_{sk})}. \quad (4)$$

We wish to place a row of sinks along the top surface of the channel at location $\tilde{z}_{sk,j} = x_j + iH_{TS}$, with each sink having strength $\tilde{Q}_{sk,j}$. To enforce a no penetration boundary condition at the lower surface, we place image sinks at $\tilde{z}_{sk,j} \pm i2kH_{TS}$, for all integers k . This gives a total complex velocity induced by these sinks of

$$w = u_0 + \frac{1}{4H_{TS}} \sum_{j=1}^n \tilde{Q}_{sk,j} \coth \left(\frac{\pi}{2H_{TS}} (\tilde{z} - \tilde{z}_{sk,j}) \right), \quad (5)$$

after the infinite sums have been evaluated using known Laurent expansions [44]. The sum can be replaced by an integral when the density of sinks increases to infinity (and their individual strengths become infinitesimal). Our potential flow solutions used 105 discrete sinks and their images.

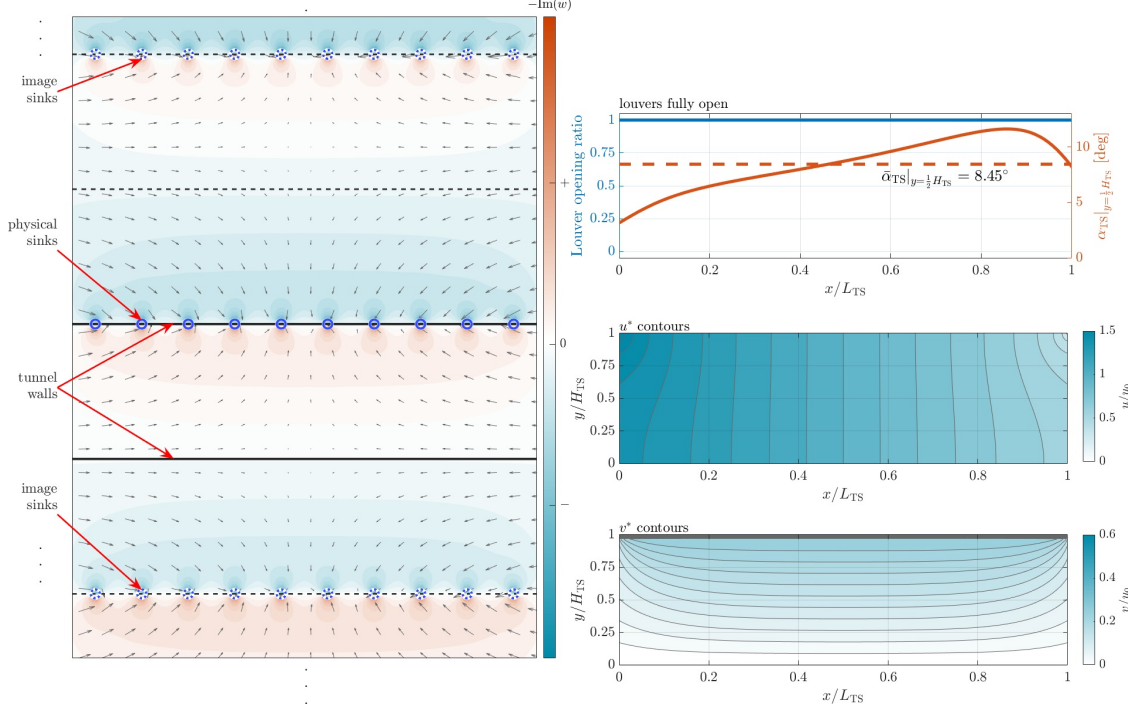


Figure 3: (Left) Example distribution of 10 sinks in potential flow model of test section. (Right) Potential flow solution for $V_{in} = 6$ m/s and louvers fully open.

The right panel of Figure 3 shows the test section flow field given by this analytical solution steady conditions, in which $u_0 = 4.03$ m/s and $\int_0^{L_{TS}} \tilde{Q} dx = Q_{SD}/[m] = 0.73$ m²/s. The top subplot in this panel shows the top louver opening ratio and the flow angle of attack extracted from the potential flow solution. The top louvers are fully opened in this case (the opening ratio is 1 for all x locations), thus $\tilde{Q}$ is uniformly distributed along the top boundary. The mean flow angle of attack at the half test-section height is $\bar{\alpha}_{TS}|_{y=H_{TS}/2} = 8.45^\circ$. The u_0 -normalized velocity components u^* and v^* are shown in contours. The spatial deceleration in u^* (as seen in the middle subplot) causes the flow angle of attack to increase throughout the test section, which is expected since more flow goes into the suction duct as x increases. The v^* contour at the bottom shows that the vertical velocity gradient dv^*/dx is approximately zero, away from the streamwise inlet and outlet regions. Note that the potential flow model may not be accurate near these boundaries, as it does not account for features of the wind tunnel such as the louvers downstream of the test section.

To simulate the traveling-wave flow field with the potential flow model, the flow flux distribution on the top boundary is allowed to vary with respect to both the time and position based on equation 3. Assuming that the vertical velocity at the test-section ceiling/suction-duct inlet is proportional to the local flow flux of the sinks, the traveling cross flow can be determined by appropriately defining the flow flux distribution $\tilde{Q}(t, x)$ based on the blockages of each of the sub-channels in the suction duct. This allows for time-varying solutions for the flow field to be determined for given wave speeds and frequencies. We emphasize that this method assumes that the flow response to changing louver positions is immediate, and thus that the flow field at a given time is only a function of the instantaneous louver position.

Two snapshots of the solution at a wave speed of 6 m/s and a wave frequency of 1.5 Hz are shown

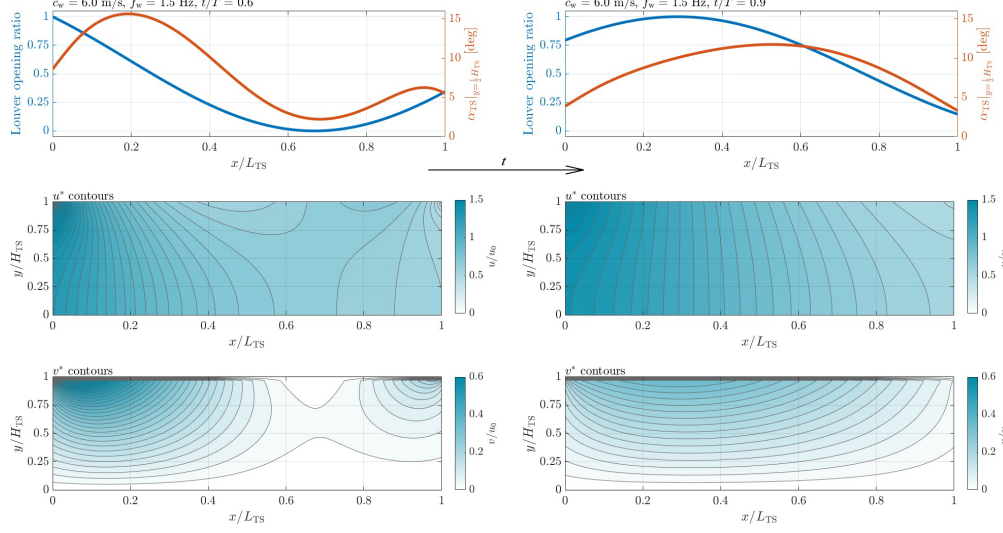


Figure 4: Snapshots of potential flow solution for traveling wave at $c_w = u_0 = 6$ m/s and $f_w = 1.5$ Hz. The top row shows the spatial distribution of the louver opening ratio and flow angle in the middle of the test section, while the remaining subplots show the predicted instantaneous velocity fields.

in Figure 4. Note that the inlet flow speed of the test section V_{in} and the total volume flux at the suction-duct inlet are different from those in the fully-opened solution in Figure 3. The wave speed and frequency are chosen here to match subsequent experiments, which consider a mean flow speed of $u_0 = 6$ m/s and a highest wave frequency of $f_w = 1.5$ Hz. From the left ($t = 0.6T$, where T is the period) to right columns ($t = 0.9T$), the snapshots visually show how traveling characteristics are realized by a spatio-temporally varying sink distribution on the top boundary. The louver opening ratio shown in the top plots demonstrates the sinusoidal form of the sink distribution. Although the correlation between the suction strength and the louver opening ratio may not be fully linear in reality, it is reasonable to assume a linear relation between them in this simplified theoretical analysis (comprehensive studies on the dynamic control of the tunnel could be the subject of future research). At $t = 0.6T$, the zero opening ratio is at $x = 0.67L_{TS}$ while the full opening point has not moved into the test-section domain (the half wavelength in this case $\lambda_w/2 = 2$ m is close to the test-section length since we specified the wave speed to be the same as the mean flow speed). The effect of the spatial suction distribution is visualized in the velocity field contours, especially, the v component contours show the influence of neighboring suction concentrations as one of them is moving out of the test-section exit to the right and the second peak is entering the test-section inlet from the left. When the time reaches $0.9T$, the first peak is completely out of the domain and the second peak has traveled into and dominated the test section. The flow angle of attack also ‘travels’ with the mean flow as a result of the varying suction effect.

Through the flow rates from the lumped element model and the flow field from the potential flow solutions, the theoretical analysis provides some knowledge of what to expect from the suction-driven approach. It also gives preliminary guidance on how to operated the wind tunnel in experiments, which are discussed in the next section. However, we are aware that these theoretical models are simplified descriptions of the wind tunnel, so one should not expect a full replication of the flow parameters or the flow field from the models. Nevertheless, the potential flow model has a good potential to include more comprehensive configurations, e.g., modeling an airfoil encountering

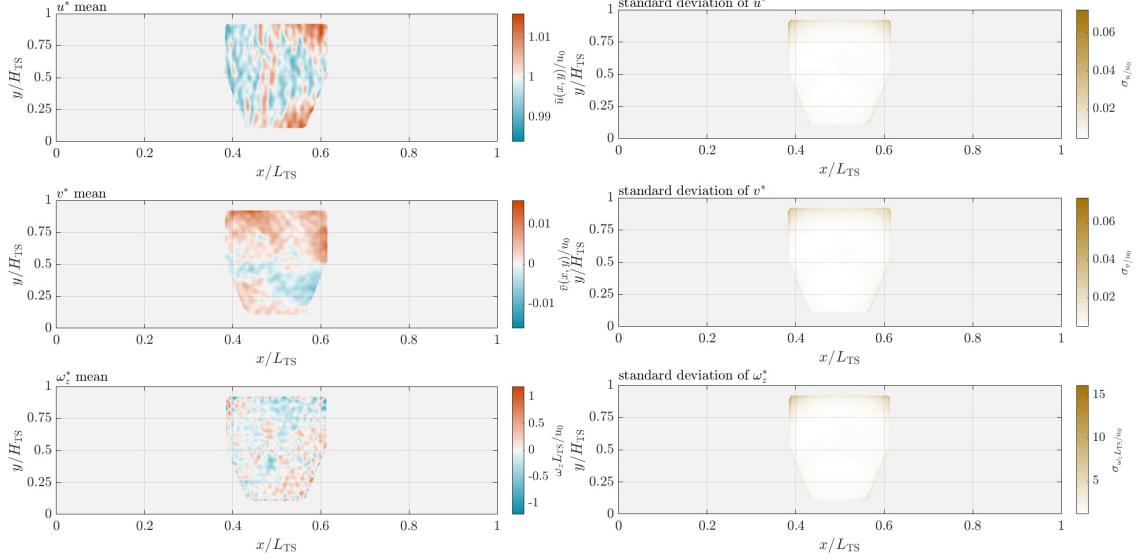


Figure 5: PIV measurement of mid-test-section baseline flow at $u_0 = 3.3$ m/s.

traveling waves, taking advantage of the experimental discovery of this report that the transverse gusts generated are nominally irrotational as a benefit of the low-vorticity flow field created by the suction-driven approach.

2.1.4 Experimental characterization

The test-section flow field is measured using point hot wire/x wire measurements, 5-hole aeroprobe measurements, as well as particle image velocimetry (PIV) in baseline, steady suction, step input, and traveling wave tests. For the baseline tests, the downstream louvers are fully open and the top louvers are fully closed, while the breather remains fully closed. The calibration is obtained through the 5-hole aeroprobe and the single-wire hot wire measurements. The overall flow speed uncertainty is ± 0.08 m/s for the minimum fan blade pitch case and is ± 0.13 m/s for the maximum fan blade pitch. The calibration gives the operational envelope of the flow speed to be 0.3 - 18.0 m/s (50 - 1,000 rpm) and 0.6 - 28.3 m/s (50 - 1,200 rpm) for the minimum and maximum fan blade pitch angles, respectively.

Baseline flow field. Figure 5 shows an example baseline flow field at $u_0 = 3.3$ m/s that is measured by PIV within a sampling window $0.38L_{TS} \leq x \leq 0.62L_{TS}$ longitudinally and $0.15H_{TS} \leq z \leq 0.85H_{TS}$ vertically in the test section. The left panel gives the normalized mean field (averaged over 5 s) of the u^* , v^* components, and the vorticity ω_z^* from top to bottom. The relatively small color map scales make the spatial variations in the velocity and vorticity visible, however, the variations are in fact very small if we examine the absolute values and the corresponding standard deviations which are shown in the right panel. High standard deviation values at the corners of the PIV window indicate that most of the uncertainty comes from the PIV measurement noise at these locations, which raises the overall σ values. As we sort out where the noise comes from, the standard deviation values (in both space and time) in the center region of the test section are $\sigma_{u^*} = 0.0098$, $\sigma_{v^*} = 0.0094$, and $\sigma_{\omega_z^*} = 2.08$. For the z vorticity, particularly, the mean value is $\omega_z^* = -0.003$, which implies that 95.5% of the variation of ω_z^* is within -0.003 ± 4.16 with the PIV measurement noise being taken into account. This background noise level becomes important when we compare the vorticity variations in the crossflow cases later.

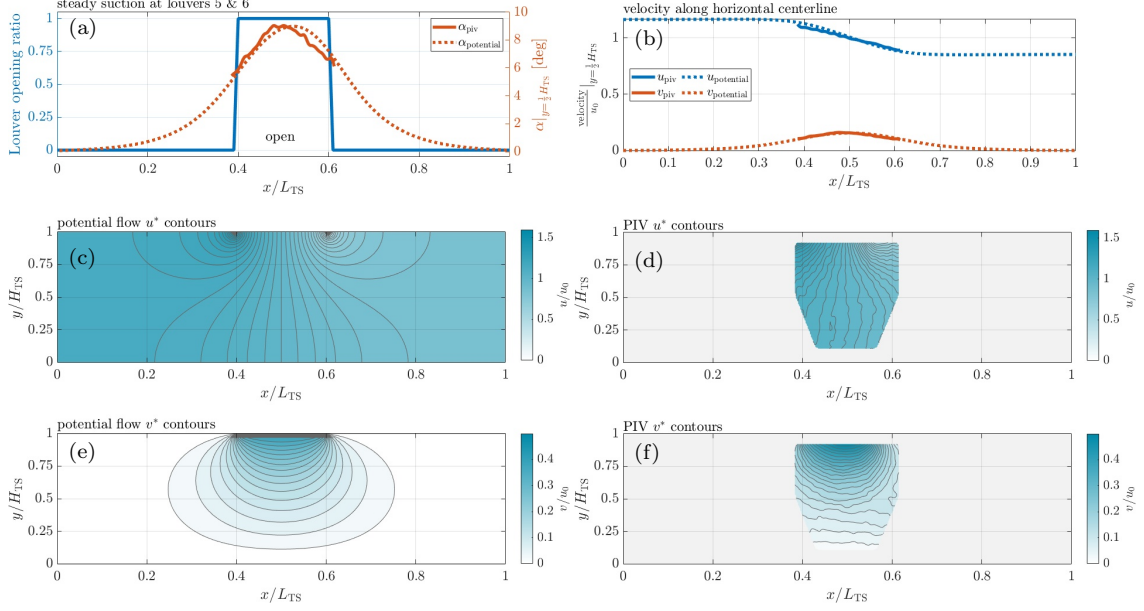
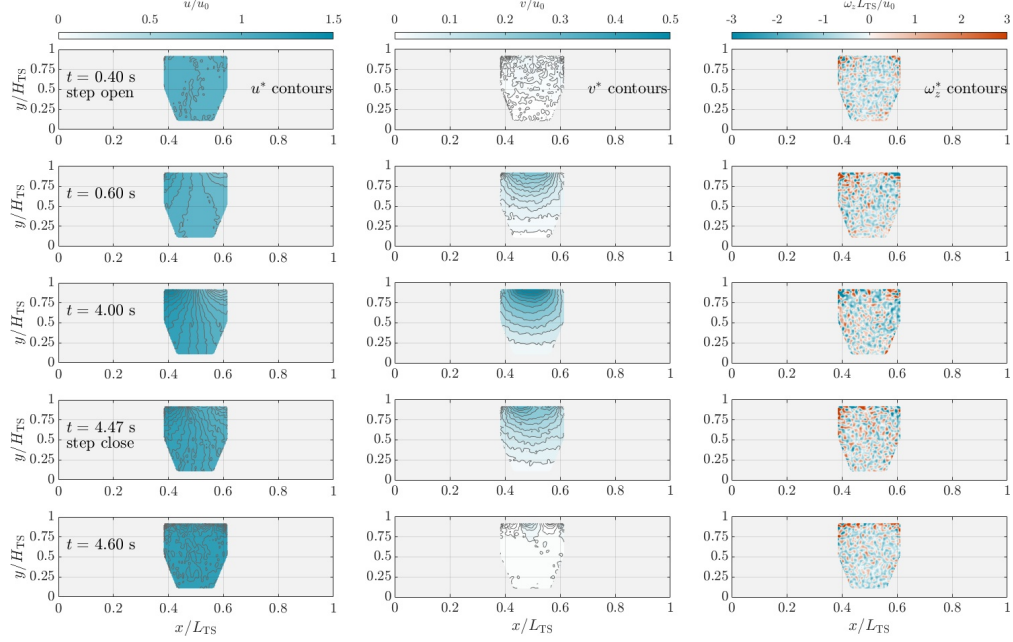


Figure 6: Steady suction test results at $u_0 = 2.0$ m/s, comparing experimental measurements with potential flow model predictions.

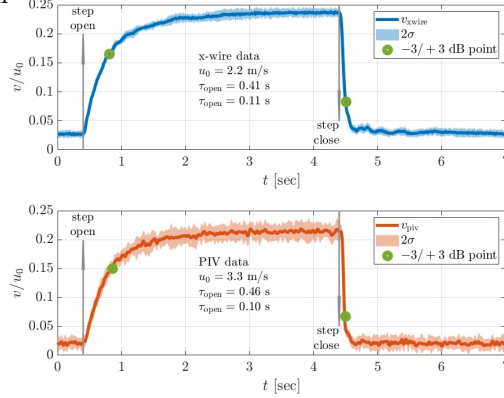
Steady suction. In all transverse flow experiments, the downstream louvers are half open in order to properly distribute the total flow rate between the test section and the suction duct. By fully opening louvers 5 and 6 at the center of the ceiling while keeping the rest of the louvers fully closed, the suction duct generates steady suction from $0.4L_{TS}$ to $0.6L_{TS}$ as shown by the louver opening ratio in Figure 6(a), which deflects the flow upward in the test section. The actual flow field in the PIV window (Figure 6(d,f)) is comparable to the potential flow field in the same window (Figure 6(c,e)). As a result, the experimentally-measured velocity components extracted along the horizontal center line agree between with the potential flow predictions (Figure 6(b)). The spatial distribution of the flow angle of attack at the center line also agrees with the potential flow prediction, reaching a peak of 9° . We conclude that the steady suction configuration gives a flow field that is consistent with the potential flow model with appropriately-defined flow rates.

Transient response. Understanding the transient response of the flow field to the louver input is essential for precise control of the transient cross-flow velocity that are generated. Figure 7a shows the snapshots of the flow field in the transient process, in which the $0.4L_{TS}$ to $0.6L_{TS}$ top louvers are opened or closed by step inputs. The repeated step inputs keep the louvers open for 4 s and then closed for 4 s in sequence. From top to bottom, each row shows the u^* , v^* , and ω_z^* contours from left to right at a certain moment in dimensional time. The flow field at each time corresponds to the louver status which changes from the beginning of the step opening ($t = 0.4$ s), during the transient response to opening ($t = 0.6$ s), approaching a steady state at maximum opening ($t = 4$ s), in the transient related to step closing ($t = 4.47$ s), and to after closing ($t = 4.6$ s). The results show that both u^* and v^* contribute to the evolution of the flow field as the louvers rotate. It is worth noting that the vorticity value in the test section remains at a very low level regardless of the operating status of the louvers/flow conditions (again, the color maps are magnified by the small scale), which will be discussed later in this section.

Point velocity results of the transient response are obtained by measuring the vertical velocity component at the center of the test section ($x = 0.5L_{TS}$, $y = 0.5H_{TS}$, and the lateral location is



(a) Snapshots of transient response of cross flow at $u_0 = 3.3$ m/s



(b) Point velocity measurements in the center of the test section for transient response to louver opening and closing. Top plot $u_0 = 2.2$ m/s; bottom plot $u_0 = 3.3$ m/s.

Figure 7: Transient louver response.

at the middle of the test-section span). The vertical velocity measured by the x wire and PIV are shown in Figure 7b. The top plot in this array shows the phase-averaged x -wire measurements of v^* with $u_0 = 2.2$ m/s over 25 tests and the bottom plot shows the phase-averaged v^* extracted from PIV measurements with $u_0 = 3.3$ m/s over 10 cycles. Both sets of data demonstrate good repeatability of the transient gusts. The non-zero mean v^* value that appears during the louver-closed phase is due to the small amount of leakage which generates low-level suction on the top. While this definitely needs to be compensated in wing/model tests (e.g., through effective angle of attack offset), on the other hand, it also shows how sensitive the test-section flow is to the suction effect. With these data, the time constants of the opening stage τ_{open} and the closing stage τ_{close} are found for the two cases above, where $\tau_{open} = 0.41$ s and $\tau_{close} = 0.11$ s in the $u_0 = 2.2$ m/s case, and $\tau_{open} = 0.46$ s and $\tau_{close} = 0.10$ s for the $u_0 = 3.3$ m/s case. The time constants are comparable between the two cases.

Although the louvers open and close at the maximum rotational rate when a step input is given

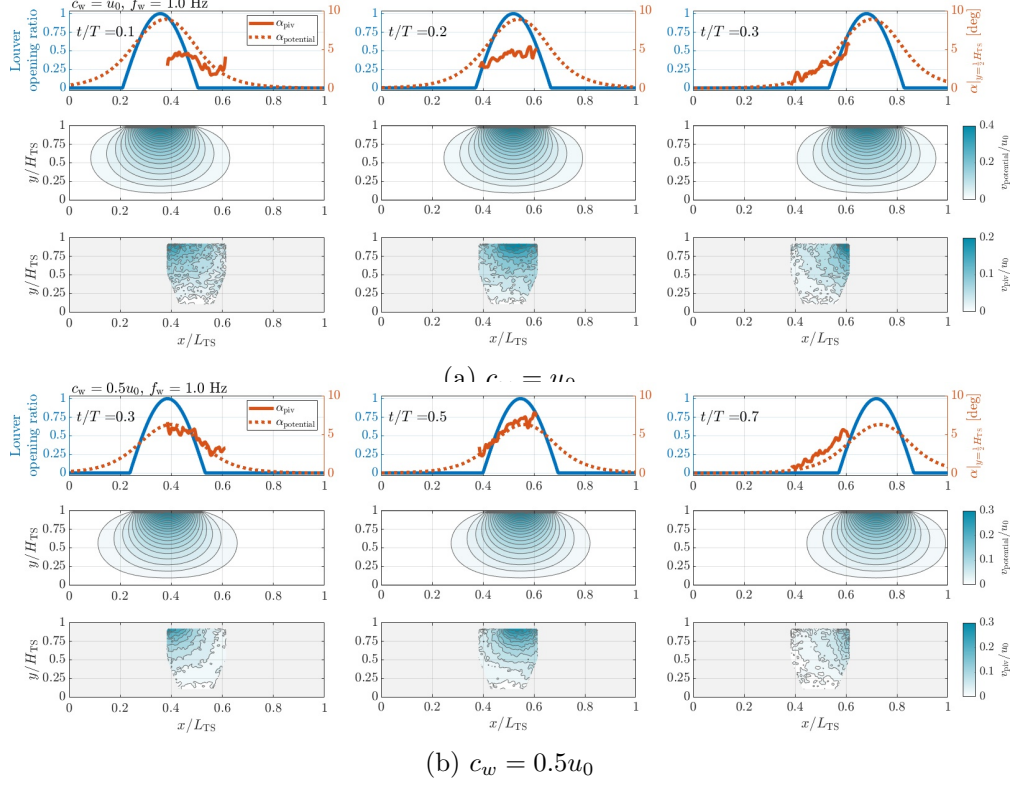


Figure 8: Traveling gust results at three instances of time, comparing flow angle and velocity field measurements with potential flow model predictions.

(in practice, the louvers take approximately 0.03 s to rotate 90° when responding to the step input from minimum to maximum), the flow responds asymmetrically to the two step inputs. The rising time τ_{open} is significantly longer than the falling time τ_{close} , as it requires a longer time for the suction effect to develop due to the fluid inertance in the suction duct. In contrast, when the louvers close, the suction is shutdown directly without much interaction with the suction duct, so that τ_{close} reflects the much shorter time that is required for the mean flow to remove the suction effect. The streamwise velocity will experience a temporary influence due to the coupling in the flow rates and the suction effect. The u^* contours clearly show the acceleration and deceleration regions that are divided by the sink/suction center. Therefore, the rising phase is not a strict ‘step’ in terms of the aerodynamic input to wings/models in the test section. Nevertheless, it does shift the flow condition from one steady state to another in less than half of a L_{TS} -based convective time ($\tau_{open} < 0.5t_c$) or, in terms of the typical wing/model scale, less than 5 chord-based convective times. We therefore expect to observe substantial transient effects during this process.

Traveling gusts. To perform experimental validation of traveling transverse gusts, we first consider isolated, single-peak gusts. The test wave frequency is fixed at $f_w = 1$ Hz, while the wave speed has two values; one equal to the mean horizontal flow speed $c_w = u_0$ and one that is half of the flow speed $c_w = 0.5u_0$. Figures 8a and 8b show snapshots of these two cases respectively, as the gusts travel through the test section with a mean flow speed of $u_0 = 2.84$ m/s. From left to right the snapshots at several points in time are shown as the gusts travel in the streamwise direction. The top rows show the spatial suction distribution on the test-section ceiling via the louver opening ratio (note that the discontinuity due to the finite louver size in the real cases is not represented here) and the model-predicted and the PIV-measured flow angles of attack along the x-direction. The middle

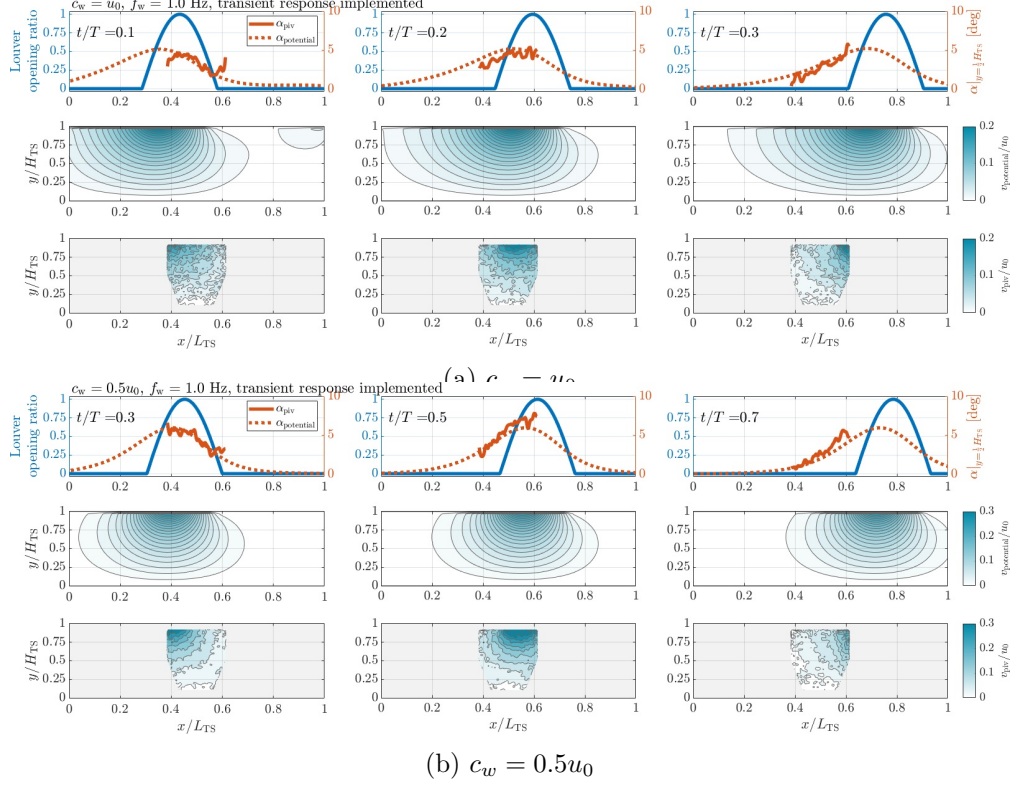


Figure 9: Results for a traveling gust compared with predictions from the potential flow model with a transient response correction.

rows show the v^* contours of the potential flow simulated flow fields and the the bottom rows are the v^* contours of the PIV-measured flow fields. By comparing the two cases, the flow angle of attack shows dependency upon the traveling wave speed. The PIV-measured flow angle of attack does not reach the value that is predicted by the potential flow model when the wave speed is the same as the mean flow speed in the given experimental configuration. Instead, the measured and predicted flow angles of attack have better agreement in the half wave speed ($c_w = 0.5u_0$) case. The difference is also shown by the doubled vertical velocity magnitude in the half wave speed case (note the different color map scales). This phenomenon may relate to 1) the transient behavior of the test-section velocity responding to the louver motions that was discussed in the previous transient study; 2) the split of flow rates in the dynamic cases does not exactly follow the quasi-steady model predictions, of which the limitation was mentioned earlier. On the other hand, we do observe that the velocity gradients dv^*/dx and du^*/dx (not shown in the plots) are generated and carried by the traveling gusts in experiments regardless of the wave speed. For instance, the elliptical contour lines in the PIV-measured v^* fields are well replicated by the potential flow model, which indicates similar velocity gradient patterns between the model predicted and the experimental results.

To improve the accuracy of the potential flow model predictions in dynamic cases, we could implement the transient response of the flow responding to the louver input into the potential flow model. As an initial attempt, the exponential behavior of the vertical velocity component in Figure 7b during the step-open stage can be approximated by a first-order linear dynamical system of the sink strength responding to the louver input, which has a normalized step response $h(t)$ of the form

$$h(t) = 1 - \exp(-\tau_{open}t) \quad (6)$$

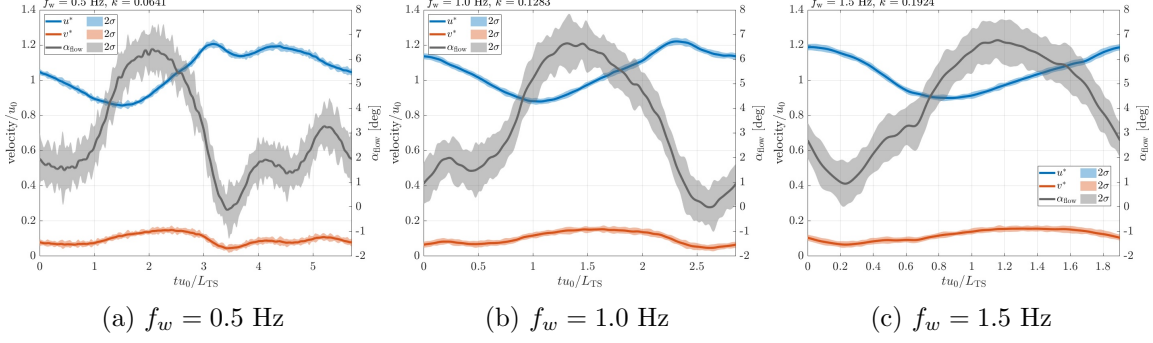


Figure 10: Nominal 1-cos profile, comparing the time evolution of the velocity and flow angle in the center of the test section.

Hence, the flow field response is no longer a function of the instantaneous louver opening ratio, but also depends on its time history. In particular, if the original potential flow model gives a sink strength of $\tilde{Q}_{input}(t)$, the transient correction for each sink is given by

$$\tilde{Q}_{output}(t) = \int_0^t \tilde{Q}_{input}(\tau) h(t - \tau) d\tau. \quad (7)$$

The potential flow fields modified by this transient correction are shown in Figure 9, where we again compare the predicted flow angle of attack and v^* contours to the experimental results. The magnitudes of both the flow angle of attack and v^* (note the change of the color map scale) of the $c_w = u_0$ case (shown in Figure 9a) are corrected since the modified sink strength does not have enough time to rise to the maximum value. Meanwhile, the $c_w = 0.5u_0$ (shown in Figure 9a) flowfield is comparatively unaffected by the transient response modification (the gain and the time constant in the dynamical system are unchanged), which indicates that the transient effect of the flow responding to the louvers is less significant at this lower wave speed. However, we also notice that this initial modification of the potential flow model may not fully represent the transient behaviors of the wind tunnel. For instance, the louver-closing stage is not well modeled as the time constants are indeed different according to the experimental measurements. Therefore, there could be further investigation of combining the transient response and the potential flow model in future work.

Traveling waves in 1-cos form. By enforcing specific wave functions (Eq. 1), the wing is tested under single-sided sinusoidal transverse gusts, of which the propagated vertical disturbances are of the 1-cos form [39]. The control signals are given to each louver independently to enable the louvers to open and close sinusoidally at the wave frequency, with phase delays between each louver signal applying the desired wave speed. For example, when we let $c_w = u_0 = 6$ m/s, the disturbances vary in space and “travel” with a phase speed that matches the mean flow speed. We observe from Figure 10 that the vertical velocity profiles are not single-frequency sinusoidal waves in actual experiments, due to the presence of higher-frequency harmonics. This is because the transverse gusts are only controlled in an open-loop manner, with prescribed louver motions. However, it remains highly repeatable, as shown by the narrow uncertainty band around the results in the figure. The nominal wavelengths of the commanded modes in these cases are $c_w/f_w = 5.72L_{TS}$, $2.86L_{TS}$ and $1.90L_{TS}$. The vertical disturbance amplitude of the actual gust (the gust ratio) varies between $0.08u_0$ - $0.1u_0$, and the instantaneous angle of attack reaches 7 degrees.

Summary of the transverse gust tunnel characterization. We have presented a suction-driven approach to generate transverse gusts for steady cross flow, transient, and traveling wave conditions.

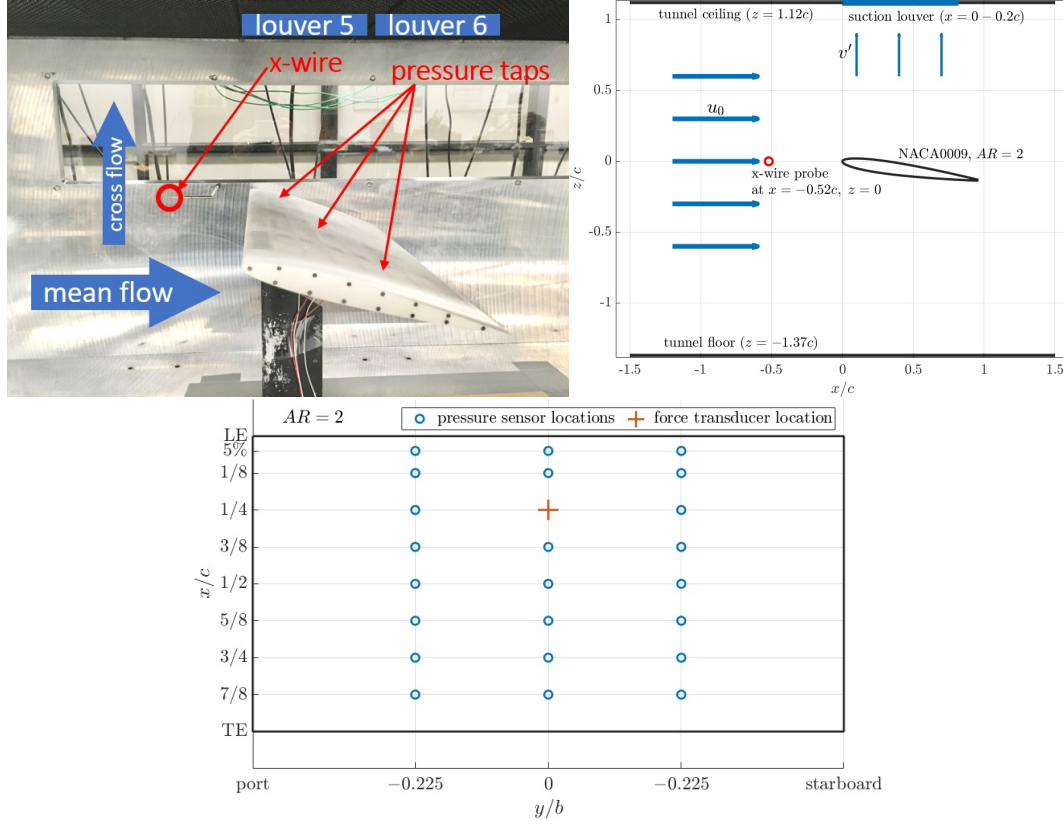


Figure 11: Wing model details.

From the PIV visualizations of the flow fields, we have demonstrated that the wind tunnel can create velocity gradients, especially dv/dx , that can be carried by the traveling wave gusts. The gusts generated by the wind tunnel have low turbulence levels and a high degree of repeatability. Besides the high-quality flow conditions in basic tests, the modified wind tunnel demonstrated promising potential for generating irrotational transverse gusts of different types. The fact that gusts are nominally irrotational allows for accurate potential flow predictions of the flow field, provided the flow rates at the louvers are corrected for time lags (or estimated from pointwise measurements). The vorticity in the test section was kept at background noise level regardless of the flow conditions (steady, transient, or traveling gusts). Given that classical unsteady aerodynamic theories were developed based on potential flow assumptions, the tunnel provides a convenient experimental environment to apply, validate, and extend these classical theories. Though the current open-loop specification of the louver motion does not allow us to generate pure single-frequency gusts, we nonetheless obtain highly repeatable gusts with simple vertical velocity profiles. In the next section, we will explore wing-gust interactions with this feature.

2.2 Transverse gust interactions with a low aspect ratio wing

This section details results reported in [47]. The wing model is setup in the test section as shown in Figure 11. The leading edge of the wing is kept at $0.4L_{TS}$ and $0.55H_{TS}$ of the test section regardless of the change in the angle of attack. The wing has a NACA 0009 cross-section with a chord length $c = 245$ mm and a wing span $b = 490$ mm, which gives an aspect ratio of $A_R = 2$. Twenty three surface pressure sensors (All Sensors® 1-Inch) are instrumented on the top surface of the wing in 3 longitudinal columns as shown in the lower panel of Figure 11. The sensor at

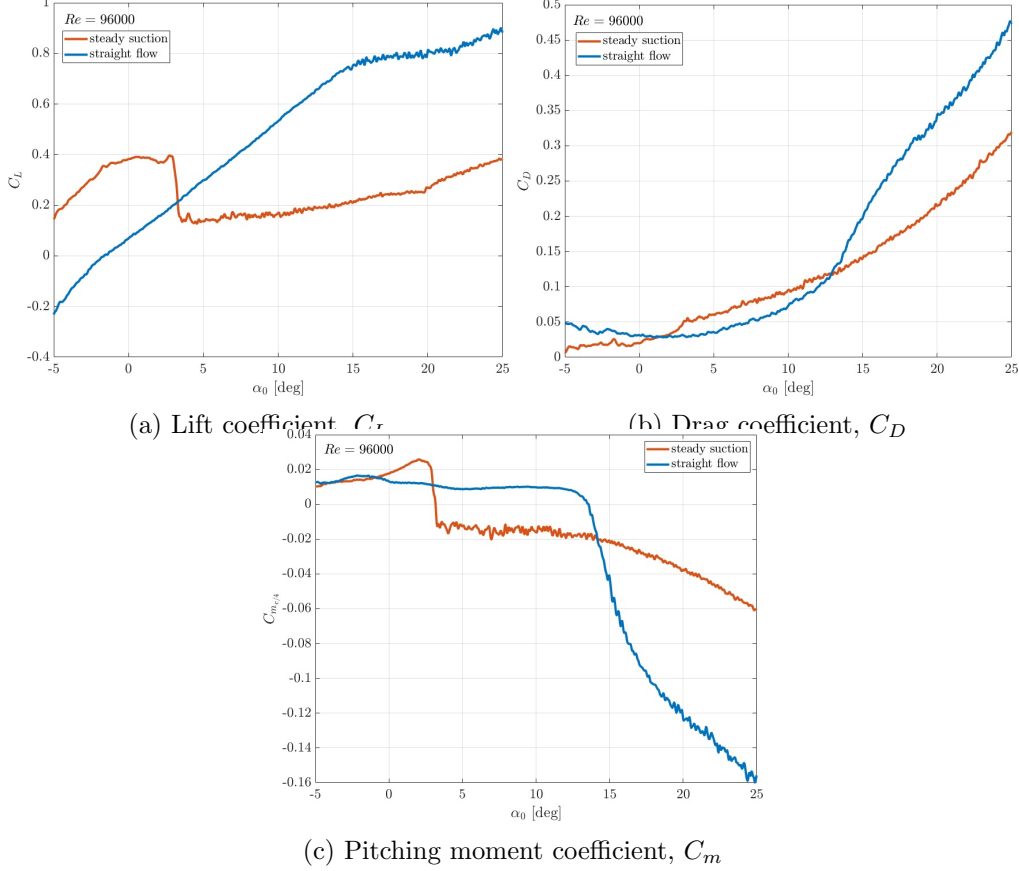


Figure 12: Force measurements from baseline and steady suction conditions.

$x = c/4$, $y = 0$ is missing due to the positioning of the force balance transducer, which is marked as the red cross in the figure. The force transducer is ATI nano 17.

2.2.1 Test cases

Baseline. The baseline forces and moments are measured under a steady straight flow condition (the gust/cross-flow feature is disabled) as the angle of attack of the wing slowly sweeps from -5° to 25° . Therefore, quasi-steady results are obtained for the baseline cases. With a mean flow speed of $u_0 = 6$ m/s, the Reynolds number is $Re = 96,000$.

Transverse gusts. Using the tools demonstrated in the previous section, the vertical velocity fluctuation can be controlled both temporally and spatially, and we are thus able to apply traveling wave disturbances to the wing model and measure the aerodynamic and pressure response. By enforcing specific wave functions, single-sided sinusoidal waves in this report, the propagated cross-flow disturbances are of the traveling wave form (3). From hereon, we set $c_w = u_0$ so that the waves travel with the mean flow speed. The frequencies will be set to $f_w = 0.5, 1.0$, and 1.5 Hz, as in the last tests of the previous section. In that section we found that, even if nominally commanded to be a single frequency mode, the actual vertical velocity profile is not single-frequency sinusoidal waves in the actual experiment due to higher harmonics. However, this does not jeopardize the objectives of the traveling wave tests, since the transverse characteristics are maintained and the results have good repeatability in all cases. With reference to the chord length of the wing, the corresponding reduced frequencies are $k = \pi f_w c / u_0$, are 0.0641, 0.1282 and 0.1923.

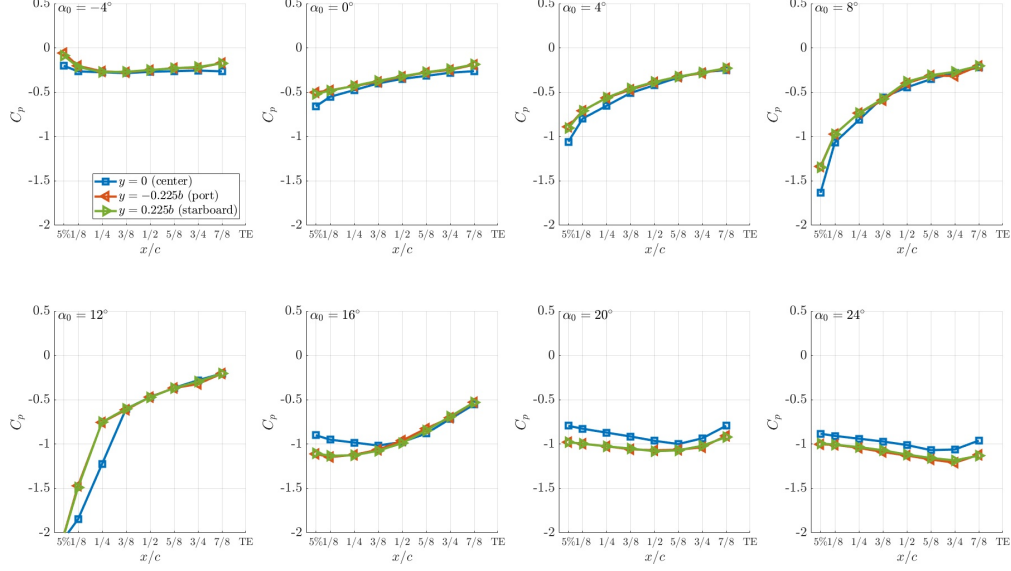


Figure 13: Surface pressure maps, C_p at baseline conditions.

2.2.2 Results

Baseline and steady suction. The quasi-steady baseline measurements are presented in Figure 12. For comparison, these plots also show the corresponding results at steady suction conditions, in which the louvers are steadily half open (similar to Figure 6). In this report, forces are normalized by $q_\infty S$, where $q_\infty = \rho u_0^2/2$ is the dynamic pressure and $S = bc$ is the planform area of the wing; pitching moment is normalized by $q_\infty Sc$, and surface pressure is normalized by q_∞ . The lift curve at baseline conditions is close to the typical ones of wings with similar thickness and aspect ratios. Note that the sting mounting on the lower surface of the wing introduces a flow asymmetry that causes the lift coefficient to be greater than zero when the geometric angle of attack is zero. Combining the lift, drag, and pitching moment information, the wing starts to stall at $\alpha_0 \approx 12^\circ$ and enters full stall above $\alpha_0 \approx 15^\circ$. In addition, the maximum lift to drag ratio (not shown) appears at $\alpha_0 \approx 5^\circ$.

The surface pressure measurements along the three chordwise rows of transducers are shown at different angles of attack in Figure 13. Note that the data points of the third sensor at the center chord $y/b = 0$ are linearly interpolated by the nearby data, since the pressure sensor is missing at this location due to the force transducer. For the full range of angle of attack, the pressure data on the left and right wings show good symmetry, but are not two-dimensional, as the center chord pressure distribution shows disagreement with them near the leading edge. This suggests that the wing-tip vortices do affect the spanwise pressure distribution (and thus, the aerodynamic loads, e.g., roll and yaw moments), albeit evenly on both sides.

Transverse traveling-wave gusts. Two sets of example results of the transverse gust tests at $\alpha_0 = 0^\circ$ and 8° are presented in this section. The wave frequencies are $f_w = 0.5$ Hz ($k = 0.0641$) and $f_w = 1.5$ Hz ($k = 0.1923$), respectively. Figure 14 shows the change in velocity components u and v and the effective angle of attack seen by the wing. The corresponding lift responses are shown in Figure 15, and the surface pressure distributions are mapped in Figure 16, in which the effective angle of attack is plotted versus the dimensionless time, $t^* = u_0 t/c$. The results show that flow disturbances alter the effective angle of attack that is seen by the wing, thereby changing the lift response. In addition to the fluctuations of the effective angle of attack, the cross-flow and

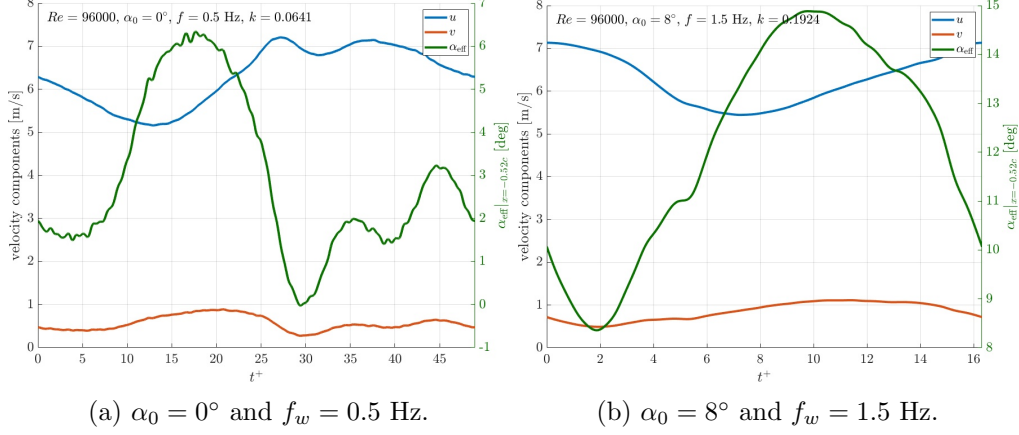


Figure 14: Example pointwise velocity responses in transverse gusts.

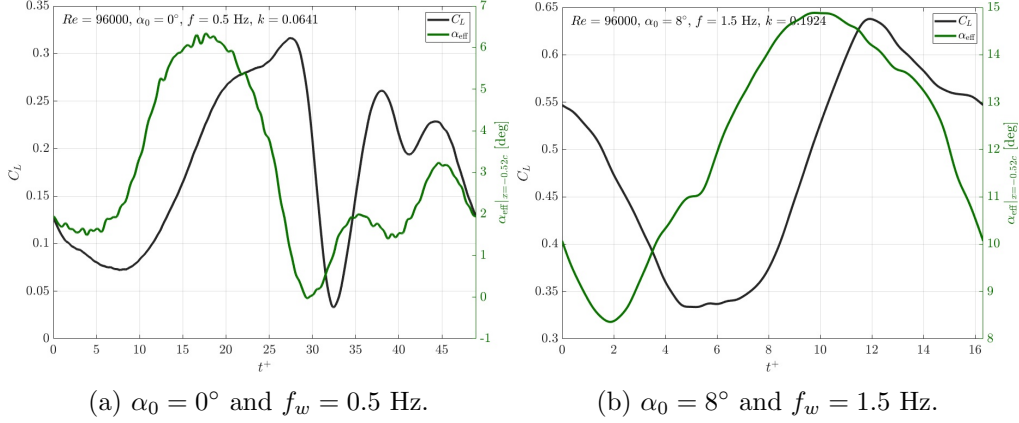


Figure 15: Example lift responses and corresponding effective angle of attack in transverse gusts.

traveling-wave features of the disturbances bring in significant variations to the surface pressure distributions, which further influence the aerodynamic loads, so that they cannot be described by the effective angle of attack alone.

Summary. The abundant characteristics that are contained in the surface pressure distributions obtained in these traveling wave gust studies suggest that one may be able to connect the sparse pressure measurements to the aerodynamic loads. In [47] (under support from this program), we pursued one approach to this end, by using the measurements to generate POD modes. We then integrated the modal decomposition to obtain the force and moment coefficients and compared these with the truth measurements from our force transducers. This approach, which we omit from this report for brevity, showed promising agreement. In Section 4 we will describe another approach, in which we use surface pressure measurements to not only estimate the aerodynamic loads, but also the response of the flow to the incident gust.

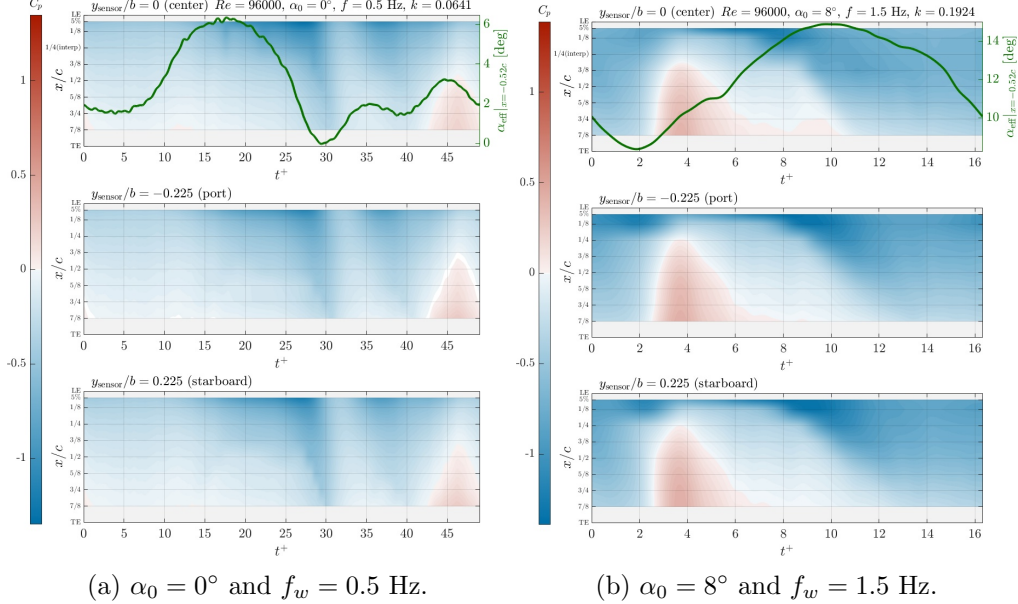


Figure 16: Example surface pressure maps in transverse gusts.

3 High-fidelity simulations of low aspect ratio aerodynamics

High-fidelity simulations form an important companion to the experimental investigations of gust interactions. In work carried out under a related ONR-sponsored project, the Caltech group’s immersed boundary (IB)/lattice Green’s function (LGF) computational framework for simulating three-dimensional, turbulent, external flows, was extended with adaptive mesh refinement [99]. We refer the reader to that paper for details on the methodology. The original IB-LGF method [68] applied an exact far-field boundary condition using fundamental solutions on regular Cartesian grids and allowed active computational cells to be restricted to vortical flow regions in an adaptive fashion as the flow evolves. The combination of spatial adaptivity and regular Cartesian structure led to superior efficiency, scalability, and robustness, but necessitated uniform grid spacing. However, the scale separation associated with thin boundary layers and turbulence at higher Reynolds numbers favors a more flexible distribution of elements/cells. This was achieved by developing a multi-resolution LGF approach that permits block-wise grid refinement while maintaining the important properties of the original scheme. It was shown in [99] that the multi-resolution LGF method can be fruitfully combined with the IB method to simulate external flows around complex geometries at high Reynolds numbers. This novel multi-resolution IB-LGF scheme retains good efficiency, parallel scaling as well as robustness (conservation and stability properties). It has been applied to several problems, including DNS of bluff and streamlined bodies at Reynolds numbers $O(10^4)$. The adaptive mesh reduces the total number of computational cells by up to 99.87%. An example of this application is shown in Figure 17, depicting the start-up of flow past a delta wing at 20 degrees angle of attack and Reynolds number 10^4 .

Under the current project, we used this tool to carry out preliminary simulations of the target problem of this study, the aspect ratio 2 wing, and subjected this wing to a potential flow consisting of distributed sinks at the same relative location as in the cross-flow wind tunnel test section at IIT. The simulations omitted the walls of the test section, however, and thus likely miss some of the blockage effects of the actual geometry. Figure 18 depicts snapshots of the flow past the wing, at the representative angle of 12° but at a lower Reynolds number of $Re = 10^3$. These snapshots

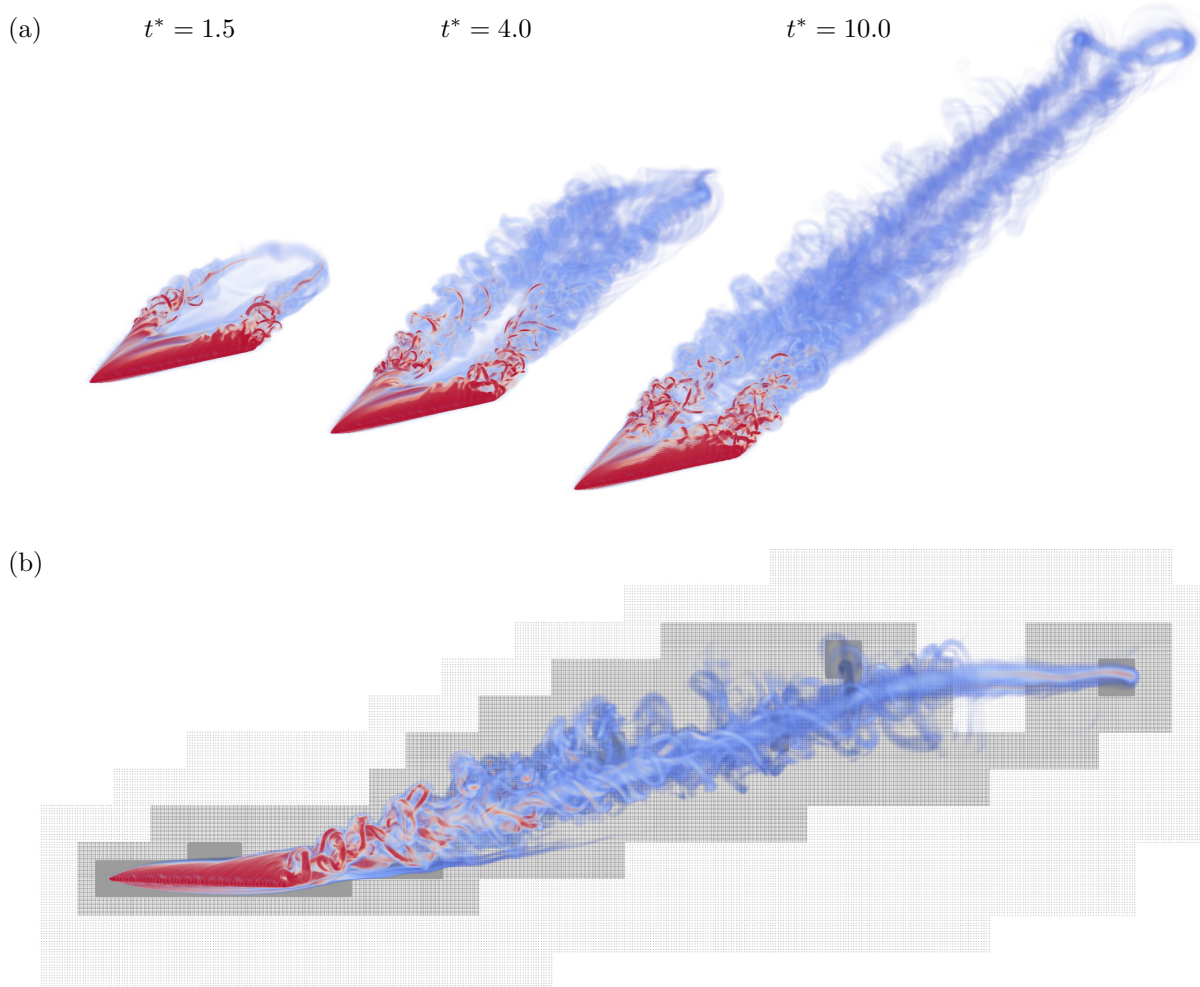


Figure 17: (a) Vorticity visualization at $t^* = t u_0 / D = 1.5, 4.0, 10.0$ for the delta wing at the angle of attack $\alpha = 20^\circ$ and Reynolds number 10^4 . (b) Mesh topology at $t^* = 4.0$. Grid level from 0 to 2 is shown from light to dark gray.

depict an arched vortex emerging from the leading edge, induced by the sink flow’s upwash effect on the wing.

The corresponding lift coefficients from the simulation and the experiment (at $\text{Re} = 95,000$) are shown in Figure 19, along with the time-varying sink strength as measured from the experiment. (The experiment in this case introduced the gust as a transient disturbance, but not as a traveling wave as in the studies in Section 2.2.) Though there is poor quantitative agreement between the simulation and experiment, the general features of a lift enhancement followed by a rapid lift reversal are present in both. There are several likely reasons for the disagreement. The first two have already been alluded to: the Reynolds numbers are widely different, and the simulation lacks the influence from the test section walls. Another important source for error is the strength of the sink. The sink flow used here was determined approximately from the pointwise x-wire velocity measurement, but this measurement cannot distinguish the sink flow from the upstream influence of the wing. It is likely essential to infer this sink measurement through a more detailed assimilation process, using similar tools to those described in this report for inferring flow fields from surface data. This is a subject of our current work.

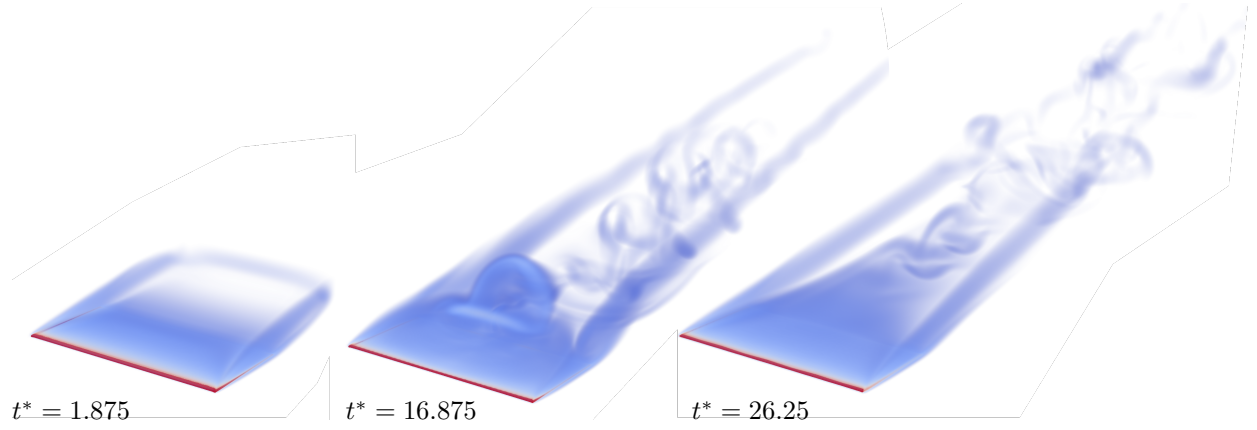


Figure 18: Vorticity visualization at $t^* = t u_0/D = 1.875, 16.875, 26.25$ for the aspect ratio 2 wing at Reynolds number 1000 at angle of attack $\alpha = 12^\circ$.

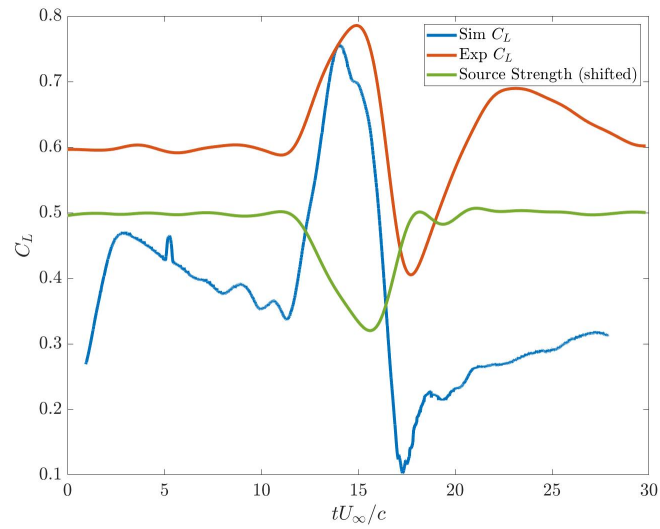


Figure 19: Comparison of lift coefficient for the aspect ratio 2 wing at $\alpha = 12^\circ$ subjected to a time-varying sink flow (with total strength as shown in green). Simulations carried out at Reynolds number 1000; experiments at $Re = 95,000$.

4 Estimation of disturbed flows from surface pressure sensors

The objective of aerodynamic flow estimation is to construct an approximation of the time-varying flow about a lifting surface from the available sensors. This task has a variety of potential uses, principally within the context of an overall closed-loop control strategy, in which the estimation of the evolving flow field enriches the information available to the controller. For example, we may seek to exploit transient flow mechanisms, such as the high lift achievable in dynamic stall [27, 5, 71, 89, 32], or to manage the flow in the presence of unknown atmospheric disturbances (gusts) [60, 74]. In all of these situations, the flow is subjected to transient disturbances—from flow actuators, wing maneuvers, or gusts—and in many instances, the disturbance induces the flow to separate or significantly affects an already separated flow.

The flow estimation problem can be posed in a variety of settings, loosely categorized by the degree to which the estimator is built on a physics-based model. At one extreme, the estimator incorporates little to no physics and a supervised machine learning architecture is trained to map sensor data to some description of the flow field [42, 54]. In contrast, if the estimator relies on a predictive physics-based model, then its job is to use the sensor data to correct that prediction for unrepresented effects. In particular, though the physics model may be outfitted with the geometry and the basic flow state, the sensor data is the only source of information on disturbances. This prediction and measurement update process is the generic task of data assimilation, the focus of this report.

All estimation strategies seek the state vector $\mathbf{x}_k$, which contains sufficient information to specify the system’s state at any discrete time step t_k . In flow estimation, $\mathbf{x}_k$ not only contains a finite-dimensional representation of the flow field, but also any boundary conditions or model parameters that are uncertain. This state vector’s evolution is predicted (*forecast*) with a dynamical operator $\mathbf{f}_k$, based on a time discretization of some underlying physics model,

$$\mathbf{x}_k = \mathbf{f}_k(\mathbf{x}_{k-1}). \quad (8)$$

The dynamical model is paired with an observation operator $\mathbf{h}_k$,

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k), \quad (9)$$

which returns a vector of predicted observations, $\mathbf{y}_k$. The data assimilation framework provides a means of folding new observations from the true system into the model through adjustments of the state vector. Generically, this adjustment (*analysis*) is carried out through an optimization: determine the adjustment that minimizes the difference between new sensor measurements $\mathbf{y}^*$ from the truth system and the model-predicted measurements from (9). Though this optimal adjustment can be derived deterministically, it is important to remember that we anticipate uncertainty in the estimation task due to random noise in both the dynamics and the measurements. Thus, in stochastic estimation, such as in the various forms of Kalman filter, the adjusted state estimate is the one that minimizes its uncertainty after new measurements have been assimilated [59, 2]. The degree to which we favor the dynamical prediction or the measurements is determined by the relative levels of noise we assume for each.

Under the assumption that the noises are drawn from Gaussian distributions, the system’s uncertainty is completely characterized by the covariance matrix. Both the Kalman filter and its variant for non-linear systems, the extended Kalman filter, require that this covariance matrix be stored and propagated along with the mean state itself. For high-dimensional systems such as in fluid dynamics, this storage and propagation of the covariance matrix is computationally intractable. This challenge was addressed by Evensen with the ensemble Kalman filter (EnKF)

[34], a Monte Carlo approach to the two-step filter algorithm. In the EnKF, the system’s mean and covariance are approximated by an ensemble of states randomly sampled from the probability distribution. In the forecast, each state is advanced by the dynamical model; in the analysis step, each ensemble member is updated with the new observation to minimize the posterior covariance. Thus, for state dimension n and ensemble size $M < n$, one need only store Mn state components rather than a $n \times n$ matrix.

The dynamical operator in a flow estimation framework can take various forms, including high-fidelity Navier–Stokes simulation. Indeed, this has been the basis for recent work by da Silva and Colonius [19], carried out under our previous AFOSR-sponsored grant. In that work, they advanced the state vector for flow past an airfoil with an operator derived from an immersed boundary projection method [88]. The state vector contained grid velocity data as well as the value of the freestream velocity and its time derivative. The truth system in this case was generated by the same simulations, but with the time-varying freestream velocity prescribed. However, the EnKF-based estimator only used sparse surface pressure measurements from this truth system, from which it successfully estimated the model’s instantaneous freestream velocity.

Since an ostensible goal for flow estimation is real-time control, dynamical models obtained from high-fidelity CFD will generally not be fast enough to update the state. One strategy might be to utilize data-based modal decompositions, such as proper orthogonal decomposition (POD), which are commonly used to reduce the dimension of grid-based solutions [90]; a Galerkin model based on such a decomposition might provide a suitable approach for computationally-efficient estimation. The resulting reduced-order models (ROM) can be made small enough to allow the use of the standard algorithms, but their well-known fragility to the specification of initial conditions and flow parameters (e.g., Reynolds number) constitutes a major limitation in applications. It would therefore be desirable to seek more robust solutions that combine efficiency with a better representation of the physics.

There are two pathways toward this goal that we pursue in this report. In one approach, developed by da Silva and Colonius [20] in work supported under this program, we utilized a dynamical model for the airfoil flow consisting of cheaper coarse-grid CFD simulations. Their state vector was then augmented with a bias error, which they successfully estimated (in a cheaper low-rank space) along with the coarsened grid state from surface observations of the higher-fidelity truth data. This resulted in a more efficient overall framework than the high-fidelity CFD approach. This will be summarized in Section 4.4.

The second approach that we pursue in this work is to use an alternative low-dimensional representation of the flow, with a dynamical model that is computationally efficient but remains sufficiently rich to describe important physical phenomena. Aerodynamics has long relied on inviscid vortex models for representing and interpreting unsteady flow phenomena [94, 7, 82, 16, 31]. Vortex models have a particular advantage in a flow estimation framework, because their physical fidelity and numerical accuracy are distinct. For example, even a model consisting of a small number of point vortices is still an exact solution (to within time marching error) of the two-dimensional inviscid equations of motion, including the non-linear convection. Thus, a vortex model of any size might serve as a low-dimensional representation of a more complex flow furthermore, it can readily account for the influence of disturbances, e.g., through Biot–Savart interactions.

Viscosity is obviously explicitly absent from an inviscid vortex model. However, among the roles of viscosity in moderate to large Reynolds number external flows, only the generation of vorticity at the surface of a body is truly essential to the flow’s large-scale dynamics. Classically, this vorticity generation was restricted to vortex shedding from the trailing edge, by regularizing to flow at the edge with the Kutta condition [94, 31]. However, in recent years, a number of inviscid vortex models have been developed for predicting aerodynamic flows with leading-edge separation,

at a wide range of angles of attack [95, 98, 78, 22]. Ramesh et al. [78] introduced the concept of the leading-edge suction parameter (LESP), a nondimensional measure of the amount of suction (integrated pressure) at the nose of an airfoil. In their criterion [78], which can be interpreted as a generalization of the Kutta condition [31], vortex elements are only shed from the leading edge if LESP exceeds a critical value, $LESP_c$; if $LESP_c$ is set to zero, then the Kutta condition is enforced at that edge. The circulation of the new vortex element is proportional to the amount by which LESP exceeds $LESP_c$. Thus, to complete a basic inviscid vortex model of a separated flow past an airfoil, one need only specify the value of $LESP_c$. Ramesh et al. [78] hypothesized that this critical value is constant and a function only of the airfoil shape; they tuned its value in a thin airfoil model to match the observed separation in a high-fidelity Navier–Stokes simulation of the airfoil. This hypothesis has been confirmed to hold for a wide variety of unsteady flows generated by an airfoil undergoing unsteady maneuvers [79]. However, for flows subjected to disturbances such as gusts, there is no *a priori* reason to believe that the hypothesis still holds, and, without further knowledge, its value should be assumed time-varying and uncertain. Furthermore, because $LESP_c$ is a measure of a critical behavior in the pressure, its value is likely to be observable from surface pressure measurements.

Thus, an inviscid vortex model outfitted with a leading-edge shedding condition based on $LESP_c$ is a strong candidate for the forecast step in aerodynamic flow estimation. A caveat of the model is that its computational complexity is $O(q^2)$, where q is the number of elements. With elements shed from the edges at every time step, the model can easily lose its computational advantage over CFD-based approaches. To restrain the growth of this dimensionality, vortex elements can be aggregated with various conservation constraints, e.g., based on moments of the vortex system [86], the rate of change of linear impulse [22], or the velocity induced at the leading edge [87]. Aggregation addresses another issue that arises when a vortex model is applied to a flow past an airfoil at moderate angle of attack. Such flows exhibit extensive interactions between shed vorticity and the plate. Singular vortex elements thus remain close to the plate and, without the influence of viscosity, create strong and non-physical pressure disturbances that interfere with the model’s predictive utility. Aggregation tends to move vortex elements away from the plate, thereby mitigating this behavior [22].

To exploit these potential benefits of a vortex-based assimilation procedure, Darakananda et al. [23] developed a data-assimilated vortex model in which the state vector consisted of the positions and strengths of vortex elements and the value of $LESP_c$ for a flat-plate airfoil. The vortices were aggregated by the same approach as in [22], which ensured that the number of elements remained lower than 60 in all cases considered. The vortex model was expressed in an EnKF framework, with the observation vector composed from the jump in surface pressures across the flat plate. By assimilating measurements from a high-fidelity simulation at Reynolds number 500, the estimation framework was able to successfully reproduce the force history on the plate.

The current work extends the estimation framework of Darakananda et al. [23] and is motivated by two observations from that work. In one of the examples in [23], the flat plate was subjected to vortical gusts generated upstream. Remarkably, the estimation framework performed well even with no representation for the gust itself in the state vector. In other words, the dynamics of interaction between the gust and the separated flow was entirely accounted for through adjustments made in the analysis steps to vorticity shed from the plate. The estimator needed no knowledge of the gust’s structure, a great advantage since gusts are, by definition, unknown. In this work, we probe this more deeply by focusing exclusively on flows subjected to incident gusts. We purposely apply gusts that are stronger and in faster sequences than in [23], so that their individual influences on the flow cannot easily be decorrelated. Even if we knew each gust’s parametric form, it would be very difficult to unambiguously disentangle their parameters. In this section, we explore whether

this parameter estimation task can be avoided.

The second observation from [23] was that the estimator’s performance on any given flow varied substantially from one realization to the next, often exhibiting non-physical spikes in the pressure and force at random instants. This lack of consistent behavior is clearly undesirable for practical estimation applications. We attempt to address this by utilizing a variant of the EnKF called the ensemble transform Kalman filter (ETKF). As we review later, the ensemble in the original EnKF (called the stochastic EnKF, or sEnKF) does not reproduce the exact posterior covariance relation of the Kalman filter, causing spurious noise in the resulting estimate. The ETKF, developed by Bishop et al. [6], fixes this with an analysis step that exactly reproduces the expected covariance. We then make a more substantial improvement to the robustness issue in Section 4.3, where we introduce a low-rank EnKF (LREnKF) [65] that identifies the small number of effective (‘informative’) directions in the state update. The LREnKF can use much smaller ensembles and perform much faster state updates to achieve the desired accuracy of the assimilation.

The rest of this section is organized as follows. Section 4.1 introduces the ensemble Kalman filter (EnKF) theory, including both the original sEnKF and the ETKF. Section 4.2.1 reviews the aggregated vortex model and presents a state-space formulation of the model. The results of the data-assimilated aggregated vortex model, using both forms of EnKF, are presented on example problems in 4.2.2 [63]. In Section 4.3, we present the development and preliminary results for our low-rank form of the EnKF, which substantially improves the robustness and efficiency of the estimation framework [64, 65]. Finally, in Section 4.4 we review the EnKF-based estimation of bias error in coarse-grid CFD [20]. For the sake of completeness, pseudo-codes for the two EnKF methods are provided in the Appendix.

4.1 Ensemble Kalman filtering methodology

In this section we describe the filtering methodology that underpins our data assimilation strategies. We start by reviewing the basic filtering problem and the purpose of addressing this problem with an ensemble approach. We then present two flavors of ensemble Kalman filter: the sEnKF as well as a deterministic variant called the ensemble transform Kalman filter (ETKF) (Bishop et al. [6]). Both will be evaluated in the vortex model applications that follow in Section 4.2.2. Then, we will review the important technique of covariance inflation to prevent filter divergence. Finally, we will interpret the data-assimilated vortex model on physical grounds.

4.1.1 A review of the filtering problem

In this section, we present a basic outline of the general filtering problem and its connection with Bayesian inference; details on this can be found in several references, including [2]. Though the variants of the ensemble Kalman filter can be explained without this background, it is useful to discuss it in order to justify the use of the ensemble Kalman filter in a filtering problem and to identify some of its limitations. In this section and those that follow, we use the following font conventions. Serif fonts refer to random variables, e.g. $\mathbf{Q}$ on $\mathbb{R}^n$, or Q on $\mathbb{R}$. Lowercase roman fonts refer to realization of random variables, e.g. q on $\mathbb{R}^n$, or q on $\mathbb{R}$. $\pi_{\mathbf{Q}}$ denotes the probability distribution for the random variable $\mathbf{Q}$, and $q \sim \pi_{\mathbf{Q}}$ means that q is a realization of $\mathbf{Q}$. $\pi_{\mathbf{Q}|\mathbf{R}=\mathbf{r}}$, or $\pi_{\mathbf{Q}|\mathbf{R}}(\cdot|\mathbf{r})$, is the probability distribution of the random variable $\mathbf{Q}$ knowing that random variable $\mathbf{R}$ takes the value $\mathbf{r}$. It is called the *conditional probability* of $\mathbf{Q}$ given $\mathbf{r}$.

For the sake of generality, we frame our discussion on a generic discrete nonlinear state-space model (8), but we will connect the reader throughout to the present case of a low-order vortex model. Formally, the model is described by the pair of random variables $(\mathbf{X}_k, \mathbf{Y}_k)$ for $k \geq 1$, where $\mathbf{X}_k$ is the state variable of a Markov process in $\mathbb{R}^n$ and $\mathbf{Y}_k \in \mathbb{R}^d$ is an observation of the state $\mathbf{X}_k$, assumed to be conditionally independent of the state (detailed later in this section). In our case, the state variable is composed of the LESPc and the positions and strengths of the entire set of

point vortices. The dynamics of the state $\mathbf{X}_k$ is described by the probability distribution for the initial condition $\pi_{\mathbf{X}_0}$ and a propagation equation, also called the dynamical equation:

$$\mathbf{X}_k = \mathbf{f}_k(\mathbf{X}_{k-1}) + \mathbf{W}_k, \quad (10)$$

where $\mathbf{f}_k : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the forward operator and $\mathbf{W}_k$ is an additive process noise. In the present case, $\mathbf{f}_k$ is the time-discretized vortex model. Propagating the dynamical equation for the state variable $\mathbf{X}_{k-1}$ is equivalent to sampling from the transition distribution $\pi_{\mathbf{X}_k | \mathbf{X}_{k-1}}(\cdot | \mathbf{x}_{k-1})$ where $\mathbf{x}_{k-1}$ is one realization of the random variable $\mathbf{X}_{k-1}$.

In general, the state $\mathbf{X}_k$ is only indirectly observed in a noisy and nonlinear fashion through $\mathbf{Y}_k$:

$$\mathbf{Y}_k = \mathbf{h}_k(\mathbf{X}_k) + \mathbf{V}_k, \quad (11)$$

where $\mathbf{h}_k : \mathbb{R}^n \rightarrow \mathbb{R}^d$ is the *observation operator* and $\mathbf{V}_k$ is an additive *measurement noise*. In our case, $\mathbf{h}_k$ applies the unsteady Bernoulli equation to the state vector to obtain a vector of pressure differences, at discrete locations, between the upper and lower surfaces of the plate. The gradient of the observation operator $\mathbf{h}_k$ evaluated at the current state $\mathbf{x}_k$ is called the *tangent linear* of the observation operator, denoted $\mathbf{H}_k$, i.e., $\mathbf{H}_k = \nabla \mathbf{h}_k$. It is important to note that we do not restrict ourselves in this section to Gaussian distributions for the initial condition, process noise, or measurement noise. Furthermore, the forward and observation operators can be time-dependent.

In the filtering problem, we leverage the realizations of the observation process $(\mathbf{y})_{1:k}$ to infer the realization of the state variable $\mathbf{x}$ at time step k . Our main objective is to estimate the posterior distribution of this state—providing us with complete information about the state and its uncertainty—given all of the observations made thus far,

$$\pi_{\mathbf{X}_k | \mathbf{Y}_{1:k}}(\mathbf{x} | \mathbf{y}_{1:k}) = \pi_{\mathbf{X}_k | \mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_k}(\mathbf{x} | \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k). \quad (12)$$

Let us consider the estimation of this posterior distribution from the most recent observation, $\mathbf{y}_k$, which can be computed from Bayes' theorem:

$$\pi_{\mathbf{X}_k | \mathbf{Y}_{1:k}}(\mathbf{x} | \mathbf{y}_{1:k}) = \frac{\pi_{\mathbf{Y}_k | \mathbf{X}_k}(\mathbf{y}_k | \mathbf{x}) \pi_{\mathbf{X}_k | \mathbf{Y}_{1:k-1}}(\mathbf{x} | \mathbf{y}_{1:k-1})}{\pi_{\mathbf{Y}_k}(\mathbf{y}_k)}, \quad (13)$$

where $\pi_{\mathbf{Y}_k | \mathbf{X}_k}(\mathbf{y}_k | \mathbf{x})$ is the likelihood distribution, i.e., the probability of the observation $\mathbf{y}_k$ if we knew the state $\mathbf{x}$; $\pi_{\mathbf{X}_k | \mathbf{Y}_{1:k-1}}(\mathbf{x} | \mathbf{y}_{1:k-1})$ is the prior distribution, our estimate of the state before knowledge of the new observation $\mathbf{y}_k$; and $\pi_{\mathbf{Y}_k}(\mathbf{y}_k)$ is the distribution of the observation. Note that, in the likelihood, we have assumed that the observation errors are independent in time, so it is only conditioned on the current state and not on past observations. Since $\pi_{\mathbf{Y}_k}(\mathbf{y}_k)$ can be thought of as a normalizing constant, we can rewrite (13) as:

$$\pi_{\mathbf{X}_k | \mathbf{Y}_{1:k}}(\mathbf{x} | \mathbf{y}_{1:k}) \propto \pi_{\mathbf{Y}_k | \mathbf{X}_k}(\mathbf{y}_k | \mathbf{x}) \pi_{\mathbf{X}_k | \mathbf{Y}_{1:k-1}}(\mathbf{x} | \mathbf{y}_{1:k-1}). \quad (14)$$

This equation provides us with a means of assimilating a new observation, $\mathbf{y}_k$, into our forecast probability distribution for the state. The forecast itself comes from our dynamical model. Under the Markov assumption for the state dynamics, the distribution of the state at time step k depends only on the state at the previous time step $k-1$, i.e., $\pi_{\mathbf{X}_k | \mathbf{X}_{0:k-1}} = \pi_{\mathbf{X}_k | \mathbf{X}_{k-1}}$. The joint distribution $\pi_{\mathbf{X}_{k-1:k}}$ can thus be factorized into this state transition and the posterior distribution at the end of the previous step:

$$\pi_{\mathbf{X}_{k-1:k} | \mathbf{Y}_{1:k-1}} = \pi_{\mathbf{X}_k | \mathbf{X}_{k-1}} \pi_{\mathbf{X}_{k-1} | \mathbf{Y}_{1:k-1}}. \quad (15)$$

By recursively applying this equation with (14) substituted for the posterior distribution, it is easy to show that

$$\pi_{\mathbf{x}_{0:k} | \mathbf{y}_{1:k}} \propto \pi_{\mathbf{x}_0} \prod_{i=1}^k \pi_{\mathbf{y}_i | \mathbf{x}_i} \pi_{\mathbf{x}_i | \mathbf{x}_{i-1}}. \quad (16)$$

This equation is central in Bayesian inference since it justifies the use of sequential methods to estimate the posterior distribution. We can update our previous estimate with new observations sequentially without having to restart the calculation at every time step.

The prior distribution $\pi_{\mathbf{x}_k | \mathbf{y}_{1:k-1}}$ can be obtained by marginalizing (15) over all the possible realizations of the state at time $k-1$:

$$\pi_{\mathbf{x}_k | \mathbf{y}_{1:k-1}} = \int_{\mathbf{x}_{k-1}} \pi_{\mathbf{x}_k | \mathbf{x}_{k-1}} \pi_{\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}} d\mathbf{x}_{k-1}. \quad (17)$$

This equation is called the Chapman–Kolmogorov equation and corresponds to a direct integration of the state transition kernel.

Thus, Bayes’ theorem (14) coupled with equation (17) provides us with an elegant update rule for incorporating new measurements into the probability distribution, and thus, improving our estimate of the mean and uncertainty of the state. It requires a time-discretized dynamical model (10) for the state transition, $\pi_{\mathbf{x}_k | \mathbf{x}_{k-1}}$, and an observational model (11) for the likelihood $\pi_{\mathbf{y}_k | \mathbf{x}_k}$.

However, there are two primary challenges with applying this in the contexts of interest here. First, we do not generally know the forms of distributions involved in any of these formulas. In what follows in this report, we will make the typical assumption that all errors are drawn from Gaussian distributions with zero mean, but there are good reasons to doubt that this is reasonable in the present nonlinear context of a vortex model. Second, for high-dimensional problems, the integration of the Chapman–Kolmogorov equation (17) is intractable. Indeed, here, as in many physics contexts, the forecast step corresponds to the time advancement of a partial differential equation (the Euler equations), so it is infeasible that we could advance such an equation over all possible values of the state. In our case, it is a time step of the discrete vortex model—lower dimensional than the discretized Navier–Stokes system, to be sure, but still containing tens to hundreds of degrees of freedom.

Both of these challenges motivate our use of ensemble filtering methods in this work to sequentially estimate the posterior distribution. We use a set of M particles $\{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^M\}$ sampled from the state distribution $\pi_{\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}}$ and we aim to construct a particle approximation of the posterior distribution $\pi_{\mathbf{x}_k | \mathbf{y}_{1:k}}$. It should be stressed that the term “particle” is used here in its usual sense in stochastic estimation as a member of the ensemble; it does not denote a vortex particle, a term that we avoid in this report in favor of “vortex element” or “vortex blob”.

To construct this particle approximation of the posterior distribution, we perform the following two steps: a forecast step and an analysis step. In the forecast step, each particle $\mathbf{x}^i$ is propagated through the dynamical equation (8) (the vortex model) and randomly perturbed with noise (called additive covariance inflation, discussed below) to form samples from the prior distribution $\pi_{\mathbf{x}_k | \mathbf{y}_{1:k-1} = \mathbf{y}_{1:k-1}}$. This ensemble forecast step constitutes a Monte Carlo approximation of the Chapman–Kolmogorov equation (17). Given this sampling $\{\mathbf{x}^i\}$ from the prior distribution, one can easily create a sampling $\{\mathbf{y}^i\}$ from the likelihood distribution $\pi_{\mathbf{y}_k | \mathbf{x}_k}$ by evaluating the observation equation (9) (the Bernoulli equation) for each ensemble member $\mathbf{x}^i, i = 1, \dots, M$.

The analysis step then updates the set of ensemble members by assimilating the new realization $\mathbf{y}_k^*$ of the observation variable $\mathbf{Y}_k$; in this report, $\mathbf{y}_k^*$ represents a new pressure difference measured from the truth system, a high-fidelity Navier–Stokes simulation. With the probability distributions unknown, this would require that we estimate the posterior distribution via Bayes’ theorem (14)

given the finite set of samples $\{\mathbf{x}^i, \mathbf{y}^i\}$ from the joint distribution formed by the prior and likelihood. However, the task is made considerably simpler if we assume that the distributions are Gaussian. Under that assumption, and with knowledge of the prior mean and covariance matrix, then Bayes' theorem leads naturally to the analysis step of the classical Kalman filter [59]. In the particle approximation, in which we estimate this mean and covariance from the finite ensemble statistics that emerge from the forecast, we obtain the ensemble Kalman filter [2]. It should be noted, however, that since the underlying distributions are likely non-Gaussian, *our EnKF framework is not expected to converge in the limit of large ensemble size*. In the following sections we will present two forms of this method, but first we present some important notation.

4.1.2 Additional notation

We will use the superscript f to denote prior, forecast quantities and the superscript a to denote analysis, posterior quantities. In particular, the exact prior mean is $\mathbf{x}_k^f = \mathbb{E}[\mathbf{X}_k | \mathbf{Y}_{1:k-1} = \mathbf{y}_{1:k-1}] = \mathbb{E}[\mathbf{X}_k | \mathbf{y}_{1:k-1}]$ and exact prior covariance is $\mathbf{P}_k^f = \text{Cov}[\mathbf{X}_k | \mathbf{y}_{1:k-1}]$. The posterior mean and covariance are given by $\mathbf{x}_k^a = \mathbb{E}[\mathbf{X}_k | \mathbf{y}_{1:k}]$ and $\mathbf{P}_k^a = \text{Cov}[\mathbf{X}_k | \mathbf{y}_{1:k}]$, respectively.

Given an ensemble $(\mathbf{x}^i)$ of size M , we define the ensemble matrix $\mathbf{X} \in \mathbb{R}^{n \times M}$ as:

$$\mathbf{X} = [\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^M]. \quad (18)$$

We use an overbar to denote statistics obtained from the ensemble, such as the sample mean and covariance:

$$\bar{\mathbf{x}} = \frac{1}{M} \sum_{i=1}^M \mathbf{x}^i, \quad \bar{\mathbf{P}} = \frac{1}{M-1} \sum_{i=1}^M (\mathbf{x}^i - \bar{\mathbf{x}})(\mathbf{x}^i - \bar{\mathbf{x}})^\top, \quad (19)$$

where $\top$ denotes the transpose operator. We define the anomaly matrix $\mathbf{X}' \in \mathbb{R}^{n \times M}$ of an ensemble as

$$\mathbf{X}' = \frac{1}{\sqrt{M-1}} [\mathbf{x}^1 - \bar{\mathbf{x}}, \mathbf{x}^2 - \bar{\mathbf{x}}, \dots, \mathbf{x}^M - \bar{\mathbf{x}}]. \quad (20)$$

The anomaly matrix obviously has zero mean. Using $\mathbf{1}$ to denote the vector of ones of length M , we can conveniently define the sample mean, anomaly matrix, and sample covariance from the ensemble matrix:

$$\bar{\mathbf{x}} = \frac{1}{M} \mathbf{X} \mathbf{1}, \quad \mathbf{X}' = \frac{1}{\sqrt{M-1}} \mathbf{X} (\mathbf{I} - \mathbf{1} \mathbf{1}^\top / M), \quad \bar{\mathbf{P}} = \mathbf{X}' \mathbf{X}'^\top. \quad (21)$$

Finally, we have the relation between the ensemble matrix and the anomaly matrix:

$$\mathbf{X} = \bar{\mathbf{X}} + \sqrt{M-1} \mathbf{X}' \quad (22)$$

where $\bar{\mathbf{X}} = \bar{\mathbf{x}} \mathbf{1}^\top = [\bar{\mathbf{x}}, \bar{\mathbf{x}}, \dots, \bar{\mathbf{x}}] \in \mathbb{R}^{n \times M}$. It is easy to verify that $\mathbf{1}^\top \mathbf{1} = M$, $\mathbf{X}' \mathbf{1} = 0$, and $\bar{\mathbf{X}} \mathbf{1} / M = \bar{\mathbf{x}}$.

4.1.3 The stochastic ensemble Kalman filter

The application of the stochastic ensemble Kalman filter for aerodynamic flow estimation was already presented already in [19, 23], but we review it here in order to identify an important deficiency that motivates our current treatment. The forecast step of the sEnKF—and, indeed, of all forms of the EnKF—is given by applying (10) to each ensemble member:

$$\mathbf{x}_k^{f,i} = \mathbf{f}(\mathbf{x}_{k-1}^{a,i}) + \mathbf{w}_k^i \quad \text{for } i = 1, \dots, M, \quad (23)$$

where the random process noise $\mathbf{w}_k^i$ is drawn from the distribution $\mathbf{w}_k^i \sim \pi_{\mathbf{Q}_k}$. It should be noted that this process noise is not inherently part of the dynamical model and must be explicitly introduced; this is discussed in Section 4.1.5.

In order to simplify notation, we drop the time dependence subscript of the variables in the rest of this section, since the analysis step does not involve time propagation. The analysis step of the sEnKF seeks a linear transformation of the form:

$$\mathbf{x}^{a,i} = \mathbf{x}^{f,i} + \mathbf{K}(\mathbf{y}^* + \boldsymbol{\epsilon}^i - \mathbf{h}(\mathbf{x}^{f,i})), \quad \text{for } i = 1, \dots, M, \quad (24)$$

where $\mathbf{x}^{f,i}$ and $\mathbf{x}^{a,i}$ are the i th prior and posterior ensemble member, respectively; $\boldsymbol{\epsilon}^i$ is drawn from the measurement noise distribution $\pi_{\mathbf{V}}$; and $\mathbf{y}^*$ is the realization of the observation variable $\mathbf{Y}$ at the current assimilation time. The expression $\mathbf{y}^* + \boldsymbol{\epsilon}^i - \mathbf{h}(\mathbf{x}^{f,i})$ is called the innovation for the i th ensemble member.

The algorithm of the sEnKF is provided in Algorithm 1. The original analysis step derived by Evensen [34] did not include the measurement noise term, but this was corrected by Burgers et al. [8] to reflect that $\mathbf{y}^*$ is drawn from a random distribution and to avoid some spurious correlations within the ensemble. The matrix $\mathbf{K} \in \mathbb{R}^{n \times d}$ is called the *Kalman gain* and is identical to its form in the standard Kalman filter, derived to minimize the trace of the posterior covariance matrix $\mathbf{P}^a$ (as we will discuss further below):

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^\top (\mathbf{H} \mathbf{P}^f \mathbf{H}^\top + \mathbf{V})^{-1} \quad (25)$$

with $\mathbf{V}$ the covariance matrix of the measurement noise and $\mathbf{H}$ the tangent linear of the observation operator. The exact prior covariance $\mathbf{P}^f$ is approximated by the sample prior covariance $\overline{\mathbf{P}}^f$ formed from the prior ensemble $\{\mathbf{x}^{f,i}\}$ with (21), $\overline{\mathbf{P}}^f = \mathbf{X}^{f'} \mathbf{X}^{f'}^\top$; similarly, the posterior covariance is approximated by $\overline{\mathbf{P}}^a = \mathbf{X}^{a'} \mathbf{X}^{a'}^\top$.

The Kalman gain suggests that we must calculate the tangent linear of the observation operator. Evensen [34] proposed a technique called implicit linearization that approximates $\mathbf{P}^f \mathbf{H}$ and $\mathbf{H} \mathbf{P}^f \mathbf{H}^\top$ given the prior ensemble. First, we construct $\overline{\mathbf{y}}^f$ the mean of the observations $\{\mathbf{h}(\mathbf{x}^{f,i})\}$ for $i = 1, \dots, M$:

$$\overline{\mathbf{y}}^f = \frac{1}{M} \sum_{i=1}^M \mathbf{h}(\mathbf{x}^{f,i}). \quad (26)$$

Let us then define the innovation anomaly matrix $\mathbf{Y}^{f'}$, with i th column given by

$$\mathbf{Y}^{f',i} = \frac{\mathbf{h}(\mathbf{x}^{f,i}) - \overline{\mathbf{y}}^f - \boldsymbol{\epsilon}^i + \bar{\boldsymbol{\epsilon}}}{\sqrt{M-1}}, \quad \text{for } i = 1, \dots, M, \quad (27)$$

and with $\bar{\boldsymbol{\epsilon}}$ the sample mean of $\{\boldsymbol{\epsilon}^i\}$. It should be noted that this mean is itself a random number whose expected value is 0.

The key ingredient of the technique is to make the following approximation:

$$\mathbf{H}(\mathbf{x}^{f,i} - \overline{\mathbf{x}}^f) \simeq \mathbf{h}(\mathbf{x}^{f,i}) - \overline{\mathbf{y}}^f. \quad (28)$$

Using this approximation, it is easy to show that $\mathbf{P}^f \mathbf{H}^\top$ and $\mathbf{H} \mathbf{P}^f \mathbf{H}^\top$ are approximated by $\mathbf{X}^{f'} \mathbf{Y}^{f'}^\top$ and $\mathbf{Y}^{f'} \mathbf{Y}^{f'}^\top - \mathbf{V}$, respectively [2], and thus, the Kalman gain (25) takes the simple form

$$\mathbf{K} = \mathbf{X}^{f'} \mathbf{Y}^{f'}^\top \left(\mathbf{Y}^{f'} \mathbf{Y}^{f'}^\top \right)^{-1}. \quad (29)$$

To gain more insights on this formulation of the filter, we can rewrite the analysis update (24) with ensemble matrix notations. From (22), the update of each posterior ensemble member is

equivalent to a separate update of the posterior mean $\bar{\mathbf{x}}^a$ and the posterior anomaly matrix $\mathbf{X}'^a$. From the linear analysis step (24), the posterior anomaly matrix $\mathbf{X}'^a$ is updated according to ([2]):

$$\mathbf{X}'^a = \mathbf{X}'^f - \mathbf{K}\mathbf{Y}'^f; \quad (30)$$

the update equation for the posterior mean state $\bar{\mathbf{x}}^a$ is obtained by taking the expectation of the analysis step (24):

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{K}(\mathbf{y}^* - \bar{\mathbf{y}}^f), \quad (31)$$

where $\bar{\mathbf{y}}^f$ is defined in (26). Equation (31) is the exact Kalman update equation.

Our derivations thus far have omitted the fact that our actual ensemble is finite sized. Because of the assumed linear form of the analysis step, equation (31) is reproduced almost exactly for finite-sized ensembles, except for the addition of a lingering sample mean of the measurement noise, $\bar{\epsilon}$, in parentheses. However, a more problematic discrepancy lies in the relationship between the posterior and prior covariance matrices. For obtaining this relationship, let us first denote by $\mathbf{W}'$ the anomaly matrix of the measurement noise $\mathbf{W}'^i = (\epsilon^i - \bar{\epsilon})/\sqrt{M-1}$; the product $\mathbf{W}'\mathbf{W}'^\top$ approximates the measurement covariance $\mathbf{V}$. Then, from the update equation (30) and the definitions of the covariances and innovation anomaly matrix, we can construct the equation for the posterior covariance matrix:

$$\bar{\mathbf{P}}^a = (\mathbf{I}_n - \mathbf{K}\mathbf{H})\bar{\mathbf{P}}^f(\mathbf{I}_n - \mathbf{K}\mathbf{H})^\top + \mathbf{K}\mathbf{W}'\mathbf{W}'^\top\mathbf{K}^\top + (\mathbf{I}_n - \mathbf{K}\mathbf{H})\mathbf{X}'^f\mathbf{W}'^\top\mathbf{K}^\top + \mathbf{K}\mathbf{W}'\mathbf{X}'^f(\mathbf{I}_n - \mathbf{K}\mathbf{H})^\top. \quad (32)$$

By taking the expectation of this matrix over the measurement noise process and then minimizing its trace over $\mathbf{K}$, we recover the classical Kalman gain (25) and a simple form of the expected value of the posterior covariance:

$$\mathbb{E}[\bar{\mathbf{P}}^a] = (\mathbf{I}_n - \mathbf{K}\mathbf{H})\bar{\mathbf{P}}^f. \quad (33)$$

Indeed, the classical Kalman filter adopts this expected value in order to propagate the covariance matrix, $\mathbf{P}^a = (\mathbf{I}_n - \mathbf{K}\mathbf{H})\mathbf{P}^f$. However, the finite sample covariance $\bar{\mathbf{P}}^a = \mathbf{X}'^a\mathbf{X}'^a{}^\top$ does not reproduce this ideal relation except in the limit of infinite ensemble sizes. Rather, the covariance of the finite ensemble only strictly satisfies (32), due to spurious correlations between the forecast anomalies and observation noise, $\mathbf{X}'^f\mathbf{W}'^\top$. Thus, the stochastic analysis step of the sEnKF introduces error that degrades the performance of the ensemble Kalman filter. An alternative form of the EnKF will be presented in the next section that addresses this issue.

4.1.4 The ensemble transform Kalman filter

In the previous section, it was shown that finite ensembles do not reproduce the expected value of the posterior covariance matrix $\bar{\mathbf{P}}^a$, and that this can cause the performance of the sEnKF to degrade. To circumvent the issue, Bishop et al. [6] developed the *ensemble transform Kalman filter (ETKF)* that exactly reproduces the ideal propagation equation for the covariance. This has been shown to give better performance in other applications [76, 2]. The ETKF belongs to a more general class of ensemble Kalman filters—called *deterministic ensemble Kalman filters*—that verify exactly the correct propagation equation through various analysis schemes.

To generate a posterior anomaly matrix $\mathbf{X}'^a$, one can start from the desired propagation equation for the covariance and use the factorization of the sample covariance matrix (21):

$$\bar{\mathbf{P}}^a = \mathbf{X}'^a\mathbf{X}'^a{}^\top = (\mathbf{I}_n - \mathbf{K}\mathbf{H})\bar{\mathbf{P}}^f \quad (34)$$

The right-hand side of this equation can also be factorized when $\bar{\mathbf{P}}^f$ is replaced with its own factorization and the ensemble expression (29) of the Kalman gain is introduced. In the following,

it should be noted the innovation anomaly matrix $\mathbf{Y}'^f$ is defined as in (27), but now without the observation noise, so that $\mathbf{Y}'^f = \mathbf{H}\mathbf{X}'^f$:

$$\begin{aligned}\mathbf{X}'^a \mathbf{X}'^a{}^\top &= (\mathbf{I}_n - \mathbf{K}\mathbf{H})\mathbf{X}'^f \mathbf{X}'^f{}^\top \\ &= (\mathbf{I}_n - \mathbf{X}'^f \mathbf{Y}'^f{}^\top (\mathbf{Y}'^f \mathbf{Y}'^f{}^\top + \mathbf{V})^{-1} \mathbf{H})\mathbf{X}'^f \mathbf{X}'^f{}^\top \\ &= \mathbf{X}'^f (\mathbf{I}_M - \mathbf{Y}'^f{}^\top (\mathbf{Y}'^f \mathbf{Y}'^f{}^\top + \mathbf{V})^{-1} \mathbf{Y}'^f) \mathbf{X}'^f{}^\top \\ &= \mathbf{X}'^f \mathbf{G} \mathbf{X}'^f{}^\top.\end{aligned}\tag{35}$$

Matrix $\mathbf{G} \in \mathbb{R}^{M \times M}$ is symmetric positive definite. In order to produce an analysis equation for the posterior anomaly matrix, we seek a square-root factorization of $\mathbf{G}$. In other words, we look for a matrix $\mathbf{G}^{1/2}$ such that $\mathbf{G} = \mathbf{G}^{1/2} \mathbf{G}^{1/2}{}^\top = \mathbf{G}^{1/2}{}^\top \mathbf{G}^{1/2}$. For a positive-definite matrix, the square-root decomposition exists but is not unique. Indeed for an arbitrary orthogonal matrix $\mathbf{U} \in \mathcal{O}(M)$ —i.e., any matrix such that $\mathbf{U}\mathbf{U}^\top = \mathbf{U}^\top \mathbf{U} = \mathbf{I}_M$ —then $\mathbf{G}^{1/2} \mathbf{U}$ is also a square-root of $\mathbf{G}$.

Therefore, the analysis update of the anomaly can be written as a right linear transformation:

$$\mathbf{X}'^a = \mathbf{X}'^f \mathbf{G}^{1/2} \mathbf{U}\tag{36}$$

with some choice of $\mathbf{U} \in \mathcal{O}(M)$, discussed below. Furthermore, the analysis update of the sample mean (30) can be written with the help of (29) as

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{X}'^f \mathbf{Y}'^f{}^\top (\mathbf{Y}'^f \mathbf{Y}'^f{}^\top + \mathbf{V})^{-1} \boldsymbol{\delta},\tag{37}$$

where $\boldsymbol{\delta} = \mathbf{y}^* - \mathbf{h}(\bar{\mathbf{x}}^f)$ is the mean innovation. Equations (36) and (37) are used together, through (22), to update the ensemble.

It is notable that this analysis step assembles both the posterior anomaly and the update to the mean from the columns of $\mathbf{X}'^f$. In other words, the analysis is said to be performed in the *ensemble space*: the update to the state vector is a linear combination of the deviations among the forecast ensemble members from the mean. In our case, for example, the position of a certain vortex element will be updated with a linear combination of the ensemble deviations of that element's position from the ensemble mean, after the ensemble has been advanced by the vortex model. This emphasizes the importance of such variance among ensemble members: without it, any discrepancy $\boldsymbol{\delta}$ between the new observation and the predicted observation is simply ignored.

The form of this update is an attractive property of the ETKF since the ensemble size M —which is on the order of 100—is typically similar to the size of the state n encountered in the vortex model, and certainly much smaller than the state vector associated with a CFD simulation. Using Sherman–Morrison–Woodbury identities, the formula for $\mathbf{G}^{1/2}$ can be simplified:

$$\mathbf{G}^{1/2} = (\mathbf{I}_M - \mathbf{Y}'^f{}^\top (\mathbf{Y}'^f \mathbf{Y}'^f{}^\top + \mathbf{V})^{-1} \mathbf{Y}'^f)^{1/2} = (\mathbf{I}_M + \mathbf{Y}'^f{}^\top \mathbf{V}^{-1} \mathbf{Y}'^f)^{-1/2}.\tag{38}$$

In this work, we choose the symmetric square-root factorization, $\mathbf{G} = \mathbf{G}^{1/2} \mathbf{G}^{1/2}$, among the different possibilities. Given an eigendecomposition $\mathbf{G} = \mathbf{R} \boldsymbol{\Sigma} \mathbf{R}^\top$, the symmetric square root $\mathbf{G}^{1/2}$ is given by $\mathbf{R} \boldsymbol{\Sigma}^{1/2} \mathbf{R}^\top$, with $\boldsymbol{\Sigma}^{1/2}$ the entry-wise positive square root of the diagonal matrix $\boldsymbol{\Sigma}$. Several studies [69, 36, 76] have shown that the symmetric root has useful properties, particularly that it does not introduce a spurious mean in the posterior anomaly matrix: $\mathbf{G} \mathbf{1} = \mathbf{1}$, so $\mathbf{1}$ is an eigenvector of $\mathbf{G}$ with unit eigenvalue; thus, this is also true of $\mathbf{G}^{1/2}$, so $\mathbf{X}'^f \mathbf{G}^{1/2} \mathbf{1} = \mathbf{X}'^f \mathbf{1} = \mathbf{0}$. The overall construction of the size M square-root $\mathbf{G}^{1/2}$ is computationally inexpensive.

Sakov and Oke [81] have found that the choice of the symmetric square root leads consistently to better performance. However, for large ensemble, the symmetric square root can lead to the creation of outliers in the ensemble. One can prevent the appearance of these spurious members by right multiplying $\mathbf{G}^{1/2}$ by a random orthogonal matrix $\mathbf{U}$ from time to time. To avoid bias in the posterior anomaly matrix, we must also ensure that $\mathbf{X}'^a$ has zero mean. To do so, $\mathbf{U}$ must preserve the mean, equivalent to requiring that $\mathbf{U}\mathbf{1} = \mathbf{1}$. Some authors [72, 91] have proposed a scheme based on sampling of the standard normal distribution and use of Householder reflections to construct such mean-preserving random rotations; this algorithm is presented in Algorithm 2. For the other assimilation steps, $\mathbf{U}$ is simply the identity $\mathbf{I}_M$. The overall ETKF is summarized in Algorithm 3.

4.1.5 Covariance inflation

In the different flavors of the EnKF, one seeks to estimate the posterior covariance matrix of a state variable of large dimension n with only a few ensemble members, $M \sim 100$. Therefore, the posterior covariance is usually underestimated with large sampling errors. This rank deficiency typically leads to long-range spurious correlations within the covariance matrix. Over multiple assimilation cycles, the Kalman gain will decrease to zero and the measurements will no longer be assimilated into the ensemble; this is called *filter divergence*. However, efficient regularization techniques such as the *covariance inflation* and *localization* have been developed to mitigate these sampling errors. These techniques underpin the success of the ensemble Kalman filters in high-dimensional filtering problems. Localization is not implemented in this work but will be discussed in section 4.2.2 to circumvent certain flaws of our flow estimator and will be investigated in future work.

Covariance inflation is typically applied between the forecast and analysis steps. The *multiplicative inflation* increases the spread of the ensemble about the sample mean, by rescaling the deviation $\mathbf{x}^{f,i} - \bar{\mathbf{x}}^f$ by the multiplicative factor $\beta > 1$ for each ensemble member:

$$\mathbf{x}^{f,i} \leftarrow \bar{\mathbf{x}}^f + \beta(\mathbf{x}^{f,i} - \bar{\mathbf{x}}^f), \text{ for } i = 1, \dots, M. \quad (39)$$

This form of inflation is equivalent to multiplying the sample prior covariance $\bar{\mathbf{P}}^f$ by β^2 ; the best results are typically found when β is set to 1.01 and 1.03 for the sEnKF and ETKF, respectively.

Additive inflation, in contrast, accounts for the process noise that is inherently missing from the dynamical model; such noise is critical for ensuring that the ensemble maintains a variance among its members. This technique adds to each ensemble member a sample from a Gaussian distribution with zero mean and covariance $\mathbf{Q}$. The additive inflation corresponds to a Tikhonov regularization of the prior covariance matrix:

$$\bar{\mathbf{P}}^f \leftarrow \bar{\mathbf{P}}^f + \mathbf{Q}. \quad (40)$$

Whitaker and Hamill [97] proposed to combine the additive and multiplicative inflations. Indeed, Anderson and Anderson [1] have shown that additive and multiplicative inflation account for distinct kind of errors. The multiplicative inflation tends to account for sampling errors due to the small ensemble size, while the additive inflation tends to correct intrinsic errors of the model. Multiplicative inflation only requires the tuning of one parameter. Adjusting additive inflation requires tuning as many parameters as the dimension of the state. This requires more work and a deeper understanding of the dynamical system. Raanes [76] showed that the additive inflation is a robust technique to account for model errors, more complex inflation schemes can be constructed to reduce further model errors which are not considered in this study [97, 76].

4.2 Flow estimation with an ensemble of aggregated vortex models

4.2.1 The aggregated vortex model

This section provides a summary of the aggregated vortex model, which constitutes the dynamical part of our flow estimator. We focus in this report on the response of a two-dimensional infinitely-thin plate of chord length c , translating impulsively from rest at a velocity U in a fluid of density ρ . Throughout this study, positions will be reported in chord lengths and time will be measured in convective units $t^* = tU/c$.

The flow is modeled by a collection of q regularized point vortices, called blobs, identical to the model used by Darakananda et al. [23]. The vorticity field is given, as usual for a regularized vortex particle method, by

$$\omega(\mathbf{y}, t) = \sum_{j=1}^q \Gamma_j \zeta_\varepsilon(\mathbf{y} - \mathbf{y}_j), \quad (41)$$

where ζ_ε is a blob kernel with radius ε . The purpose of a blob is to de-singularize the Biot–Savart interactions with other nearby elements. Its effect on elements farther than ε is indistinguishable from that of a point vortex. The blob kernel in this work takes the commonly-used algebraic form $\zeta_\varepsilon(\mathbf{y}) = (\varepsilon^2/\pi)(|\mathbf{y}|^2 + \varepsilon^2)^{-2}$. We refer readers to [31] for a review of modeling strategies of two dimensional inviscid flows.

The Kutta condition is applied at the trailing edge. The shedding criterion based on LESPc, developed by Ramesh et al. [78], is applied at the leading edge. The LESPc is included with the vortex elements' positions and strengths as part of the system state vector, $\mathbf{x}$. Following Ramesh et al. [78], the LESPc is forecast to remain constant. Ramesh et al. [79] and Darakananda et al. [23] have shown that LESPc has a strong authority over the vortex dynamics near the leading edge. The true vortex dynamics are encoded in the pressure measurements of the plate, from which we can distill an estimate of the LESPc [54, 23]. In other words, though LESPc is predicted to remain constant, the assimilated pressure measurements will tend to cause it to vary, thereby triggering the release of weaker or stronger vorticity.

We use the same vortex element aggregation scheme developed by Darakananda et al. [23]. By aggregating vortex elements at every time step, the overall population remains modest (smaller than 60 in all cases explored) and the aggregated elements stay relatively farther from the plate than without such treatment, dramatically reducing the occurrence of spurious pressure disturbances.

In our flow estimator, the state variable $\mathbf{x}_k$ contains the positions and circulations of q blobs and the critical leading-edge suction parameter LESPc,

$$\mathbf{x}_k = [x_k^1 \ y_k^1 \ \Gamma_k^1 \ \dots \ x_k^q \ y_k^q \ \Gamma_k^q \ \text{LESPc}_k]^\top. \quad (42)$$

Thus, the state vector dimension is $n = 3q + 1$. The dynamical model (8) applies the following operations:

- (a) Enforce the no-flow-through condition on the plate,
- (b) Introduce new vortices according to the Kutta condition at the trailing edge and the current estimate of LESPc at the leading edge,
- (c) Advect vortices and plate,
- (d) Aggregate vortex elements, and set strengths of aggregated blobs to zero.

More details on $\mathbf{f}_k$ can be found in [23]. The observation operator $\mathbf{h}_k$ in (9) uses the unsteady Bernoulli equation to predict the pressure jump coefficients— $\Delta C_p = 2(p^+ - p^-)/\rho U^2$ where $+$ and

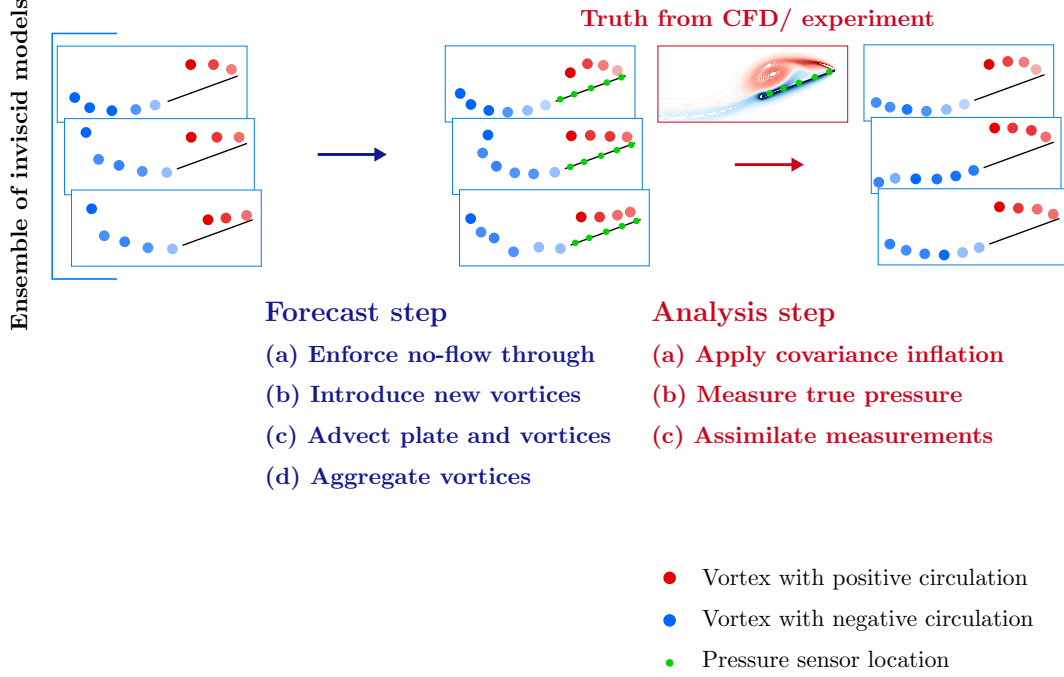


Figure 20: One time step of the data-assimilated aggregated vortex model

— denote the upper and lower side of the plate—at d locations on the plate [21]. In this work, pressure measurements are calculated (and also provided by the truth system) at the following $d = 50$ sampled Chebyshev points:

$$\frac{c}{2} \cos \left(\frac{10i\pi}{512 + 1} \right) \quad \text{for } i = 1, \dots, d. \quad (43)$$

With the aggregated vortex model summarized in Section 4.2.1 and the details of the ensemble filtering methods discussed above, we are now in a position to summarize the overall method. A schematic of this method is presented in Fig. 20. In the forecast, an ensemble of aggregated vortex models are advanced by one time step. At the end of this forecast, multiplicative and additive covariance inflation are applied to each member of the ensemble. The analysis step is then carried out, either by applying Algorithm 1 for the sEnKF or Algorithm 3 for the ETKF.

It is useful to interpret the method physically. Before we discuss the physical roles of each step of the filter, it is important to understand the effect of the ensemble and its propagation on the vorticity field. To demonstrate this effect in the clearest manner, let us assume that the q vortex elements in each ensemble member are singular: point vortices rather than blobs. Their blob form is only used to regularize the Biot–Savart interactions between them, and our discussion here focuses only on the interpretation of the vorticity field itself.

We can then write the vorticity field at location $\mathbf{y}$ and time step k as a function of the random state vector $\mathbf{X}_k$, whose components (aside from LESpc) constitute the strengths and positions of

the singular elements:

$$\omega(\mathbf{y}, \mathbf{X}_k) = \sum_{j=1}^q \Gamma_k^j \delta(\mathbf{y} - \mathbf{y}_k^j), \quad (44)$$

where δ is the Dirac delta function. The expected value of the vorticity field at the end of time step k is given (ideally, for infinite ensemble) in terms of the posterior distribution function

$$\mathbb{E}[\omega(\mathbf{y}, \mathbf{X}_k)] = \int \omega(\mathbf{y}, \mathbf{x}_k) \pi_{\mathbf{X}_k}(\mathbf{x}_k) d\mathbf{x}_k. \quad (45)$$

(For simplicity of notation, we have omitted the fact that the distribution $\pi_{\mathbf{X}_k}$ is conditioned on the observations made thus far.) Under our Gaussian assumption, it is particularly easy to evaluate these integrals, and we arrive at

$$\mathbb{E}[\omega(\mathbf{y}, \mathbf{X}_k)] = \sum_{j=1}^q \frac{\bar{\Gamma}_k^j}{2\pi |\mathbf{P}_k^{y_j}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y} - \bar{\mathbf{y}}_k^j)^\top \mathbf{P}_k^{y_j}^{-1}(\mathbf{y} - \bar{\mathbf{y}}_k^j)\right), \quad (46)$$

where $|\cdot|$ denotes determinant; $\bar{\Gamma}_k^j$ and $\bar{\mathbf{y}}_k^j$ denote, respectively, the mean circulation and position of vortex element j at time step k ; and $\mathbf{P}_k^{y_j}$ is the 2×2 covariance submatrix associated with the position of vortex element j at step k .

In other words, we can interpret each vortex element's uncertainty as defining an elliptically-shaped region, centered at its mean location and endowed with its mean strength. The behavior of this elliptical region's shape over time is determined by the combined influences of three processes in the filter: the forecast, the inflation, and the analysis.

The role of the forecast step is straightforward: it constitutes an inviscid (i.e., advective) advancement of the vortex elements by one step and the creation of new vorticity to satisfy modeled edge conditions. The ensemble of such inviscid models establishes a set of slightly different displacements of each vortex element. Since we interpret this ensemble as approximating a Gaussian distribution both before and after the forecast, the set of displacements of each vortex define a constrained transformation of the vortex's ellipse: a net advection of the center and a stretching and rotation of its shape.

The inflation step imposes two influences on each region's shape. The multiplicative inflation stretches the ellipse uniformly by a small fraction in every direction. The additive inflation, on the other hand, comprises a single step of a random walk. It is well known that a random walk, applied over a large number of steps, approaches a Wiener process, and the associated probability distribution satisfies a linear diffusion equation [73]. This diffusion causes the elliptical region to spread, reminiscent of core spreading in vortex methods [66]. To simulate diffusion of viscosity ν , the random step is chosen from a normal distribution with standard deviation $\sqrt{2\nu\Delta t}$.

In the context here, the additive inflation occurs among other processes, so its interpretation is less clear. Chorin's random vortex method [10], utilizing a large number of overlapping vortex blobs undergoing random walks, is known to converge to the solution of the Navier–Stokes equation as $q^{-1/2} \log q$, where q is the number of blobs [70]. In the EnKF context, it is possible to identify a stochastic differential equation that asymptotically describes the forecast and inflation steps in the limit of large ensemble [35]. We leave a rigorous interpretation of the ensemble of vortex models in this manner for future work. We merely observe that the variance of the distribution from which we select the additive inflation parameter loosely specifies the viscosity of a diffusion process. However, we do not attempt to match this effective viscosity to the actual viscosity in the truth system; rather, we tune the additive inflation to balance the trust between the forecast and analysis steps.

Finally, in order to reconcile the new observations made in this step—namely, the pressures measured on the surface of the wing—with the pressures predicted via the Bernoulli equation, the vortex elements’ positions and strengths (and LESPc) need to be adjusted. The EnKF analysis step provides this adjustment by assembling a minimum least-squares solution to the problem $\mathbf{h}(\bar{\mathbf{x}}^a) = \mathbf{y}^*$, subject to measurement noise $\mathbf{V}$; the minimization is regularized by the forecast $\bar{\mathbf{x}}^f$, with its associated covariance $\bar{\mathbf{P}}^f$, ensuring that the new state vector does not stray far from its forecast. Mathematically, this problem is expressed as [20, 62]:

$$\bar{\mathbf{x}}^a = \underset{\mathbf{x} \in \mathbf{R}^n}{\operatorname{argmin}} \frac{1}{2}(\mathbf{y}^* - \mathbf{h}(\mathbf{x}))^\top \mathbf{V}^{-1}(\mathbf{y}^* - \mathbf{h}(\mathbf{x})) + \frac{1}{2\beta}(\mathbf{x} - \bar{\mathbf{x}}^f)^\top \bar{\mathbf{P}}^{f-1}(\mathbf{x} - \bar{\mathbf{x}}^f), \quad (47)$$

with β the multiplicative covariance inflation factor. The analysis step causes the covariance to shrink, according to (32). The elliptical region associated with each vortex element is translated, stretched, and rotated by the analysis step, but its area shrinks.

It should be noted that the dimension of the state vector, n , changes with each time step of this filtering process: it increases typically by six (three per newly released blob originating from each edge), and occasionally decreases as elements of zero strength are eliminated. There are no inherent restrictions in the EnKF on changes to the dimension of the state. However, in order to keep a consistent state dimension, i.e., the same number of blobs across the different ensemble members, the vortex elements whose circulations are aggregated into another element are not removed, but rather, are simply assigned zero circulation. Similarly, even if the LESP does not exceed the current estimate of the LESPc, a new vortex with zero circulation is still introduced. If the blob has zero circulation across all ensemble members, then it is removed from the state vector.

We discretize the vortex model with the forward Euler time scheme with time step $\Delta t^* = 0.01$. The blob radius δ (normalized by c) is set to 5×10^{-3} and 9×10^{-3} for the sEnKF and the ETKF, respectively. For the aggregated vortex model, the vortex elements are mostly isolated and their interactions require little regularization. The non-zero blob radius primarily regularizes the interactions of the vortex elements soon after their release from the edges. Throughout this study, we use an ensemble of size $M = 50$. At the initial time, no vortices are present in the state vector of the vortex model. The ensemble is initiated with random samples for the LESPc drawn from $\mathcal{N}(0.5, 0.1)$, i.e., a normal distribution with mean 0.5 and covariance 0.1.

We use a different set of inflation parameters for the sEnKF and the ETKF, obtained by trial and error:

- A multiplicative inflation $\beta = 1.01, 1.028$ for the sEnKF and ETKF, respectively.
- An additive inflation for the position of the vortices drawn from $\mathcal{N}(0, 1 \times 10^{-5})$ (normalized by c)
- An additive inflation for the strength of the vortices drawn from $\mathcal{N}(0, 1 \times 10^{-3} \Delta t^*)$ (normalized by Uc)
- An additive inflation for the LESPc drawn from $\mathcal{N}(0, 5 \times 10^{-5})$, $\mathcal{N}(0, 8.5 \times 10^{-5})$ for the sEnKF and ETKF, respectively.
- The measurement noise $\mathbf{V}_k$ is drawn from $\mathcal{N}(0, 1 \times 10^{-8})$ (normalized by ρU^2)

The sEnKF and ETKF share the same additive inflation parameters for the vortex properties. It is important to note that each of these parameters is chosen in order to balance the trust between the vortex model and the analysis step; lower values, for example, lead to a vortex model that is less responsive to measurement innovation. For example, the relatively larger additive inflation for the

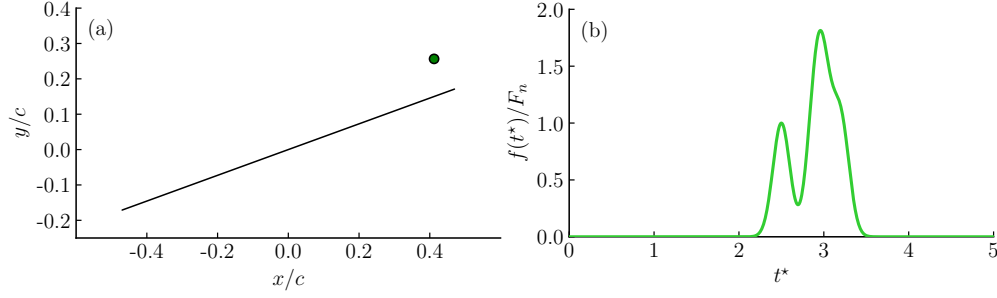


Figure 21: (a): Schematic of the plate at 20° at $t^* = 0$ subject to actuation. Green dot depicts the location of the actuation. (b) Time history of the flow actuation.

LESPc makes this parameter more responsive than the vortex parameters. The same parameters are used for all examples in the next section.

4.2.2 Application to gust-wing interactions

In this section, we present the results of flow estimation for two strongly-perturbed flows about an infinitely thin plate at 20° . In the first case, a sequence of perturbations that mimic pulse actuation is applied near the leading edge of the plate. In the second case, the plate is subject to large scale and coherent perturbations created in the wake of an upstream cylinder. In the discussion of the results, positive and negative vortex elements will refer to vortex elements with positive (counter-clockwise) and negative (clockwise) circulation respectively.

In this study, the true pressure jump measurements are generated from simulations of a flat plate at Reynolds number 500, carried out with a high-fidelity Navier–Stokes solver, based on the immersed boundary projection method with lattice Green’s function [88, 68]. It is important to emphasize that the flow perturbations are only present in the truth system, and their effect is only made available to the vortex model through the assimilation of the true pressure measurements.

Due to the stochastic nature of the filtering problem, results may vary from one simulation to another for the same filter. We estimate the uncertainty in the results in order to draw consistent conclusions about the performance of each filter. For assessment purposes, the results presented in this work have been obtained by running an ensemble of 100 realizations of the same filter on each flow configuration, with each realization consisting of an application of the EnKF (using an $M = 50$ ensemble of vortex models). From this ensemble of realizations, we construct the sample mean and standard deviation of the quantity of interest for comparison with the truth. If we assume that the results obtained over the different runs follow a Gaussian distribution, the 95% confidence interval can be estimated by considering the plus and minus deviation of twice the standard deviation from the mean. It should be noted that the number of vortex elements—and hence, the dimensionality of the state vector—varies from one realization of the filter to the next, so it is not possible to define a mean state from the ensemble of realizations.

As mentioned in Section 4.2.1, the near encounters between the singular vortex elements and the plate can lead to spurious errors in the estimate of the pressure distribution, and hence, in the prediction of overall normal force on the plate. We use a median filter to remove these spurious spikes from our presented results. At each time step, the current value is replaced by the median value of the $l = 7$ previous time steps. The median filter is causal, so it is directly integrated into the estimation sequence.

Translating plate subjected to pulse actuation disturbances In this section, we assess the data-assimilated vortex model on the response of an impulsively translating plate at 20° to disturbances

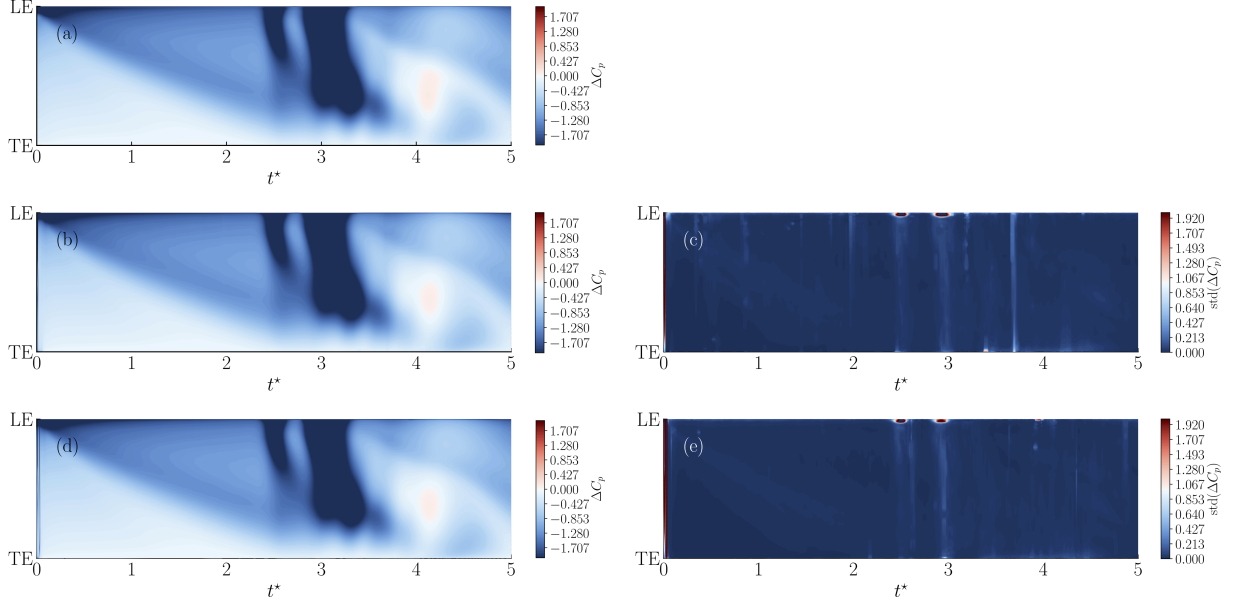


Figure 22: Left column [(a), (b), (d)]: Spatiotemporal map of the pressure coefficient jump for an impulsively translating plate at 20° subjected to pulse actuation disturbance from (a) high-fidelity numerical simulation at Reynolds number 500, and mean over 100 realizations of an inviscid vortex model with (b) the sEnKF and (d) the ETKF. Right column [(c), (e)]: Spatiotemporal map of standard deviation of the pressure coefficient jump over 100 realizations for (c) the sEnKF and (e) the ETKF

applied near the leading edge. We apply the perturbations at a point $0.475c$ from the centroid along the plate and $0.1c$ above it, as shown in Fig. 21(a). The perturbations are introduced as a superposition of vertical body force of nominal strength $F_n = 0.03\rho U^2 c$ at $t^* = 2.5, 2.9, 3.0$ and 3.2 , distributed in Gaussian form in time and space, with temporal standard deviation of $t_{std}^* = 0.1$ and spatial standard deviation $0.1c$. The vertical body force $f(t^*)$ is given by:

$$\frac{f(t^*)}{F_n} = \mathcal{N}(t^*; 2.5, t_{std}^*) + \mathcal{N}(t^*; 2.9, t_{std}^*) + \mathcal{N}(t^*; 3.0, t_{std}^*) + \mathcal{N}(t^*; 3.2, t_{std}^*), \quad (48)$$

where $\mathcal{N}(t^*; \mu, \sigma)$ denotes the temporal Gaussian kernel, with mean μ and standard deviation σ , evaluated at t^* . Fig. 21(b) shows the time history of the force actuation. For reference, the gusts considered in Darakananda et al. [23] were weaker and non-overlapping, applied further upstream of the plate at $t^* = 3, 4$ with nominal amplitude $F_n = 0.01\rho U^2 c$. The disturbance considered here is more challenging due to the presence of strong and overlapping perturbations.

Fig. 22 compares the history of the surface pressure response from the truth with the mean of 100 realizations of the sEnKF and the ETKF. During the first two convective times, we observe the development of the leading-edge vortex. The successive flow actuation disturbances are easily detected by the distinct regions of strongly negative pressure. The first suction region is due to the disturbance applied at $t^* = 2.5$, while the second one is due to the superposed response to the flow disturbances centered at $t^* = 2.9, 3.0$, and 3.2 . Both filters match well with the true pressure distribution. The uncertainty of each filter is characterized by the sample standard deviation of

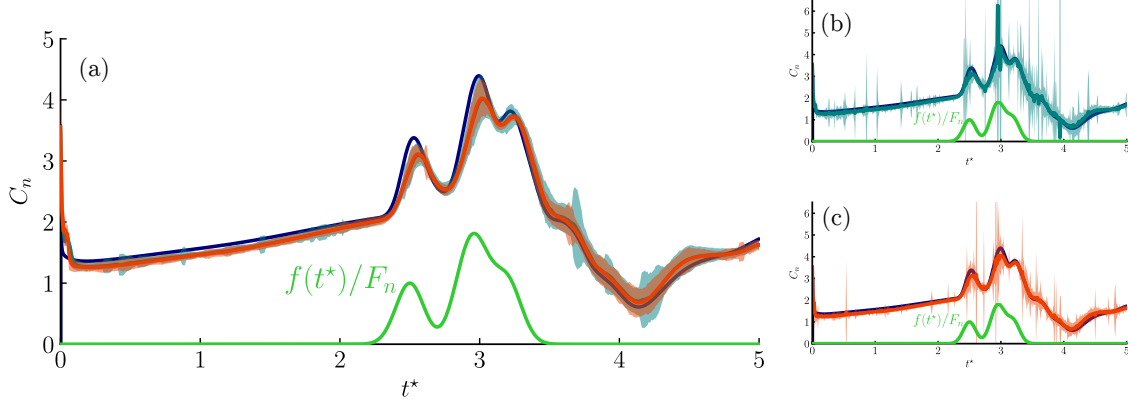


Figure 23: Left panel (a): Normal force coefficient of an impulsively translating plate at 20° subject to actuation, from high-fidelity numerical simulation at Reynolds number 500 ■, mean over 100 realizations of the inviscid vortex model with sEnKF ■, mean over 100 realizations of the inviscid vortex model with ETKF ■. Shaded areas show the 95% confidence interval for the inviscid vortex model with sEnKF and ETKF. Time history of the flow actuation ■. Right panel [(b), (c)]: Comparison with the same normal force coefficient from high-fidelity simulation, but without application of the median filter for the sEnKF (b) and ETKF (c).

the pressure, computed over the 100 runs; these are shown in the right column of Fig. 22. In both filters, two narrow bands of high variance, particularly strong near the leading edge, can be identified at $t^* = 2.5$ and 3.0 . These correspond to the instants of local maxima in the disturbance force. However, the sEnKF has a higher level of variability in the pressure distribution from one run to another. In particular, there is a significant band of high dispersion for the sEnKF at around $t^* = 3.8$, and additional smaller bands at other times. The ETKF does not exhibit such bands, a direct consequence of its statistically consistent analysis update.

The normal force coefficient is the integral of this pressure distribution on the plate. The left panel in Fig. 23 compares the force coefficient obtained from the truth system with the mean of the sEnKF and the ETKF applications over 100 realizations. The mean force predicted by each filter agrees very well with the true force response. The peaks created by the actuation and the subsequent drop of force around $t^* = 4$ are also well predicted. However, consistent with our observations of the surface pressure data, both filters show variability near the peak disturbances; Fig. 23 also shows that the peaks are slightly underpredicted.

The sEnKF exhibits more variability than the ETKF at all times, and this higher variability is apparent in the wider uncertainty envelope for the normal force coefficient, particularly after $t^* > 3.5$. To better appreciate the raw behavior of each filter, the right panels in Fig. 23 show the same results, but without the use of the median filter. The sEnKF, in Fig. 23(b), exhibits frequent spikes throughout the simulation, and particularly so after $t^* > 3.5$. These spikes are not eliminated by the ETKF, but they are much weaker and less frequent.

The origins of this high uncertainty are clear when we examine the flow behavior. In Fig. 24 we show the vorticity distribution from the truth system at three instants, $t^* = 3.0, 4.0$, and 5.0 . The models' vorticity is singular, so it is not possible to make a direct comparison; instead, we plot the locations of the elements for one realization of each filter, with the elements' signed circulations represented in colors consistent with the sign of vorticity in the truth data. Most of the large-scale vortex dynamics are captured by the data-assimilated vortex model. Over the first two convective times, the impulsive translation of the plate creates a leading-edge vortex that grows in size and

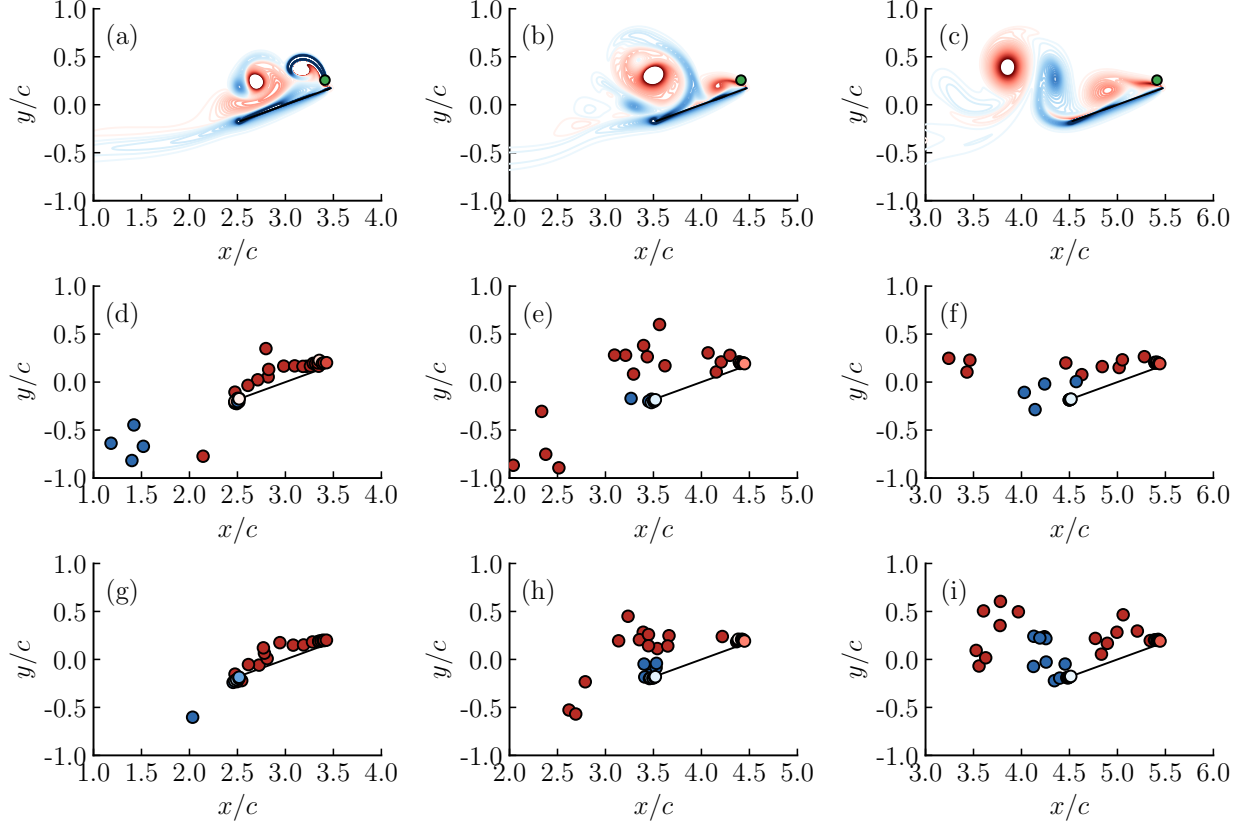


Figure 24: Snapshots of the vorticity distribution at $t^* = 3.0$ (left column), $t^* = 4.0$ (middle column) and $t^* = 5.0$ (right column) for an impulsively translating plate at 20° subject to actuation, predicted from [(a)-(c)] high-fidelity numerical simulation at Reynolds number 500, [(d)-(f)] inviscid vortex model with sEnKF, and [(g)-(i)] inviscid vortex model with ETKF.

strength. This feature is captured by the continuous release of new vortex elements in both filters. The disturbance applied at $t^* = 2.5$ creates a coherent structure that is rapidly advected along the plate and merges with the initial leading-edge vortex. At around $t^* = 3.0$, the resulting vortex lies slightly above the mid chord of the plate, evident in the left column of Fig. 24. Both the sEnKF and ETKF capture this feature with positive vortex elements clustered at a similar position. It is also important to note that the vorticity associated with the disturbance itself, clearly evident at $t^* = 3.0$, is *not* observed in the vortex model results. The absence of this disturbance vorticity is by design; only its effect on vortex elements shed from the edge of the plate is captured by the assimilation of pressure data.

Over the time interval $t^* \in [3.0, 4.0]$, the vortex models predict a separation of a cluster of positive vortex elements from the plate with dynamics similar to the truth flow. The middle column of Fig. 24 compares the results at $t^* = 4.0$. The large coherent structure with positive vorticity is well captured by the inviscid model with both filters, but is more tightly clustered in the case of the ETKF. From $t^* = 4.0$ to 5.0 , the vortex is shed into the wake and triggers a large flux of negative vorticity from the trailing edge. The right column of Fig. 24, for $t^* = 5.0$, shows a good visual agreement between the truth and the prediction of the ETKF. The spatial vortex distribution predicted by the sEnKF is less representative of the true vorticity distribution, however. Overall, the ETKF predicts a more structured and physically consistent spatial distribution of the vortices than the sEnKF. The coherence of these vortex element clusters is consistent with the narrower

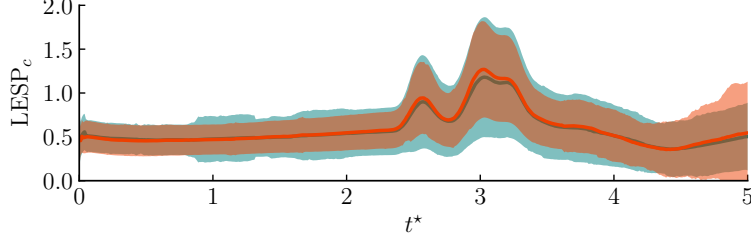


Figure 25: Time history of the ensemble mean value of the LESPc of an impulsively translating plate at 20° subject to actuation, averaged over 100 realizations, from the inviscid vortex model with sEnKF ■ and the inviscid vortex model with ETKF ■. Shaded areas show the 95% confidence interval for the inviscid vortex model with sEnKF and ETKF.

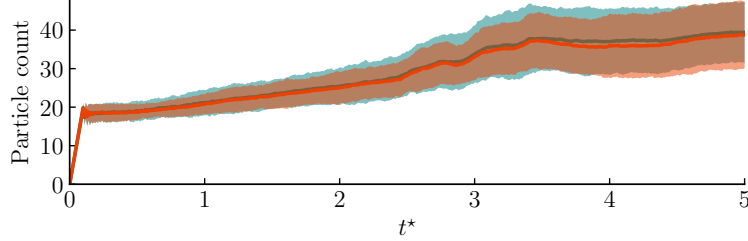


Figure 26: Time history of the particle count for an impulsively translating plate at 20° subject to actuation, averaged over 100 realizations, from the inviscid vortex model with sEnKF ■ and the inviscid vortex model with ETKF ■. Shaded areas show the 95% confidence interval for the inviscid vortex model with sEnKF and ETKF.

uncertainty in the normal force coefficient, and demonstrates a clear advantage of the ETKF over the sEnKF for modeling the flow response to unknown flow perturbations.

In Fig. 25 we compare the time histories of the LESPc estimated by the sEnKF and the ETKF. The LESPc is constrained to remain positive, and the constraint reverts to the Kutta condition if LESPc becomes zero. Before the flow is disturbed, the mean estimate of the LESPc stays on a plateau about 0.5, the mean value in the initial ensemble. This behavior supports the hypothesis of Ramesh et al. [77, 78]: the LESPc remains constant for a given Reynolds number and airfoil section. The time variation of the imposed disturbance is reflected in a similar variation of the LESPc. Indeed, the application of an actuation-like disturbance near the leading edge directly controls the vorticity flux about this edge. Large values of LESPc lead to weaker vorticity, temporarily suppressing the flux into the shear layer. The small decay of LESPc after the first disturbance increases the vorticity flux, triggering the creation of a new coherent structure that merges with the initial leading-edge vortex about the mid-chord. This leading-edge development is then halted after the next disturbance peak, and the leading-edge vortex is shed. The uncertainty envelopes of the two filters are very similar and tend to grow over time. The width of these envelopes is large, reflecting significant variation in the estimated values of LESPc from one realization to the next. This variation indicates a weak physical correlation between this threshold value and the pressure on the plate: this threshold's effect on pressure is only exerted indirectly, through the subsequent release of vorticity. (The LESP itself, in contrast, is more strongly correlated, since shed vortex elements contribute to this value [78, 31].)

The population histories of vortex elements is depicted in Fig. 26. The histories are nearly identical for each filter. In the initial time steps, the number of vortex elements increases linearly from 0 to 20; the model prevents aggregation of elements during this interval. Subsequently, the

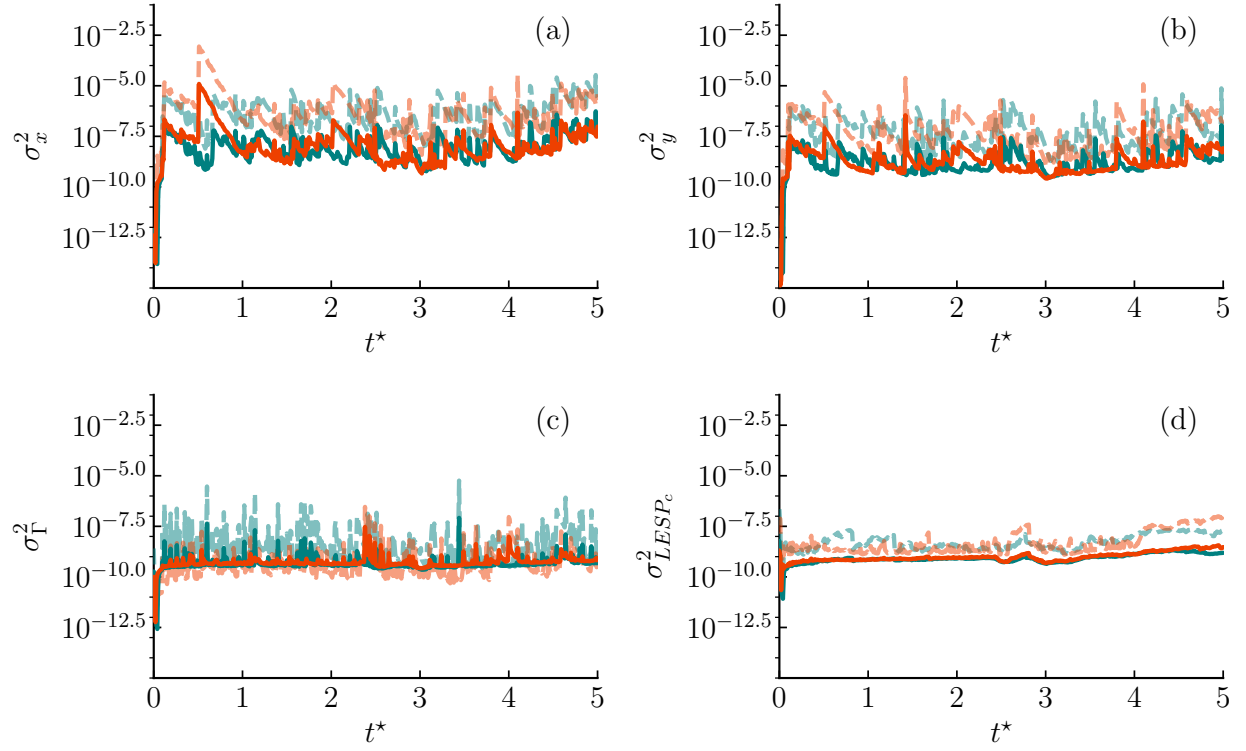


Figure 27: Time history of the ensemble variances of an impulsively translating plate at 20° subject to actuation, averaged over 100 realizations. Mean variances for (a) the x coordinate of the blobs, (b) the y coordinate of the blobs, (c) the circulation of the blobs, and (d) the LESPc from the inviscid vortex model with sEnKF ■ and the inviscid vortex model with ETKF ■. Shaded areas show the 95% confidence interval for the inviscid vortex model with sEnKF and ETKF.

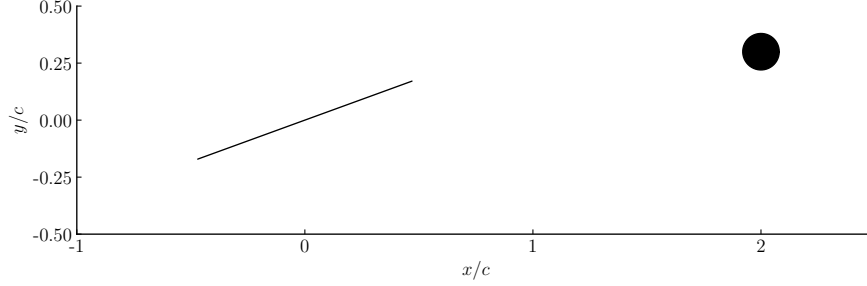


Figure 28: Schematic of a plate at 20° behind a cylinder. The plate and cylinder translate to the right at uniform speed.

population grows slowly up to 35 ± 7 at $t^* = 5$. It should be noted that, without aggregation, this population would be approximately 1000 vortex elements (500 time steps, with two elements shed per step). The variation in population among the different realizations of each filter is attributable to the variation in LESPc, which sets the initial strengths of the elements, which in turn affects their later aggregation. It is interesting to note that this variation of vortex element populations is proportionally larger than the variation in the pressure and normal force, indicating that there is some non-uniqueness in the mapping from surface pressures to vortex element dynamics.

Fig. 27 depicts the ensemble variances of the positions and strengths of the vortex elements and the LESPc estimate, averaged over the 100 realizations of each filter. Each variance is lower-bounded by the additive covariance inflation to avoid filter divergence. The variances of the strengths of the vortices and LESPc are fairly constant (and near the values set by the inflation parameter) while those of the x and y positions are more variable. These values are larger than the variance set by the inflation, likely due to additional error incurred by aggregation.

Plate in the wake of a cylinder In this part, we assess our flow estimator when applied to a plate at 20° angle of attack in the wake of a cylinder. In the truth system, high fidelity simulations are conducted of a cylinder of diameter $0.16c$ is centered 2 chord lengths upstream and 0.3 chord lengths above the plate’s centroid, as shown in Fig. 28. The plate and cylinder are both impulsively set in motion at $t^* = 0$ at speed U , so that their relative configuration remains fixed for all time. The Reynolds number based on the cylinder diameter is 80, sufficiently large that the cylinder’s wake exhibits a von Kármán vortex street. The presence of the plate triggers the wake to break symmetry and achieve this vortex street. This flow configuration distills the main features of a vehicle flying through the wake of a structure, e.g., buildings in an urban environment or other flying vehicles.

It is important to recall that we do not include the cylinder and its wake in our vortex model flow estimator; its effects are only felt through the pressure jump measurements obtained from surface sensors along the plate. Fig. 29 depicts the history of the true surface pressure distribution and those estimated by the sEnKF and ETKF; the associated standard deviations are shown on the right. Both filters estimate pressure fields that agree very well with the true distribution. As expected, it takes about 2 convective times for the cylinder wake to reach and be sensed by the plate. Over this time window, the pressure field is essentially disturbance-free. The encounter of

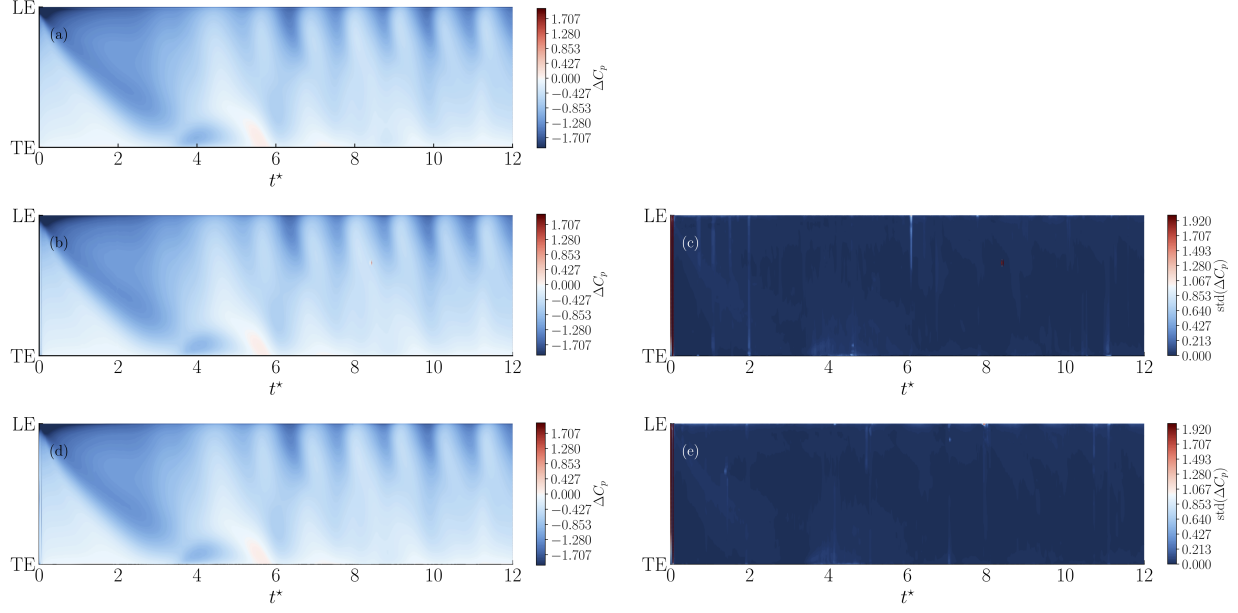


Figure 29: Left column [(a), (b), (d)]: Spatiotemporal map of the pressure coefficient jump for an impulsively translating plate at 20° in a cylinder wake (a) high-fidelity numerical simulation at Reynolds number 500, mean over 100 realizations of an inviscid vortex model with the sEnKF (b) and the ETKF (d). Right column [(c), (e)]: Spatiotemporal map of standard deviation of the pressure coefficient jump over 100 realizations for the sEnKF(c) and ETKF (e)

the vortex structures shed by the cylinder leaves successive short-lived pressure disturbances.

The true vorticity distribution and mean sets of vortex elements are compared in Fig. 30 at $t^* = 3.5$ (soon after the wake has reached the plate), at $t^* = 4.2$ (when the cylinder wake transitions to a von Kármán vortex street) and at $t^* = 10$ (long after the flow has achieved periodic vortex shedding). As expected, the vortex models do not attempt to represent the cylinder wake with vortex elements; rather, the filter accommodates the influence of the wake vorticity by modifying the behavior of the vortex elements shed from the plate. Large scale structures of the flow around the plate are captured and match visually with the true vorticity field.

Concomitantly, the normal force estimate agrees well with the truth, as seen in Fig. 31. The transition from a symmetric cylinder wake to a periodic vortex shedding causes some challenge to the discrete vortex models, as highlighted by the temporary growth of the uncertainty envelope around $t^* = 4.0$. The envelope remains small for both filters after the periodic wake behavior has been established. The history of the LESPc estimate in Fig. 32 reveals the same growth and decay of this critical value observed in the previous example; here, they represent a response to individual cylinder wake vortices passing the leading edge of the plate. These variations are essential to our inviscid framework to control the leading-edge vortex shedding in the presence of flow perturbations. As in the previous example, the weak correlation between the LESPc and measured pressure causes significant volatility in the estimated LESPc among the different realizations. The vortex element population remains small, $O(60)$, after 12 convective time units. Without aggregation, we would have to track 2400 vortex elements.

4.3 A low-rank ensemble Kalman filter for vortex estimation

For typical applications of an EnKF, it is only practical to utilize a *limited* number of ensemble members ($M \sim 100$) compared to the dimensions of the state and the observation. Hence, the

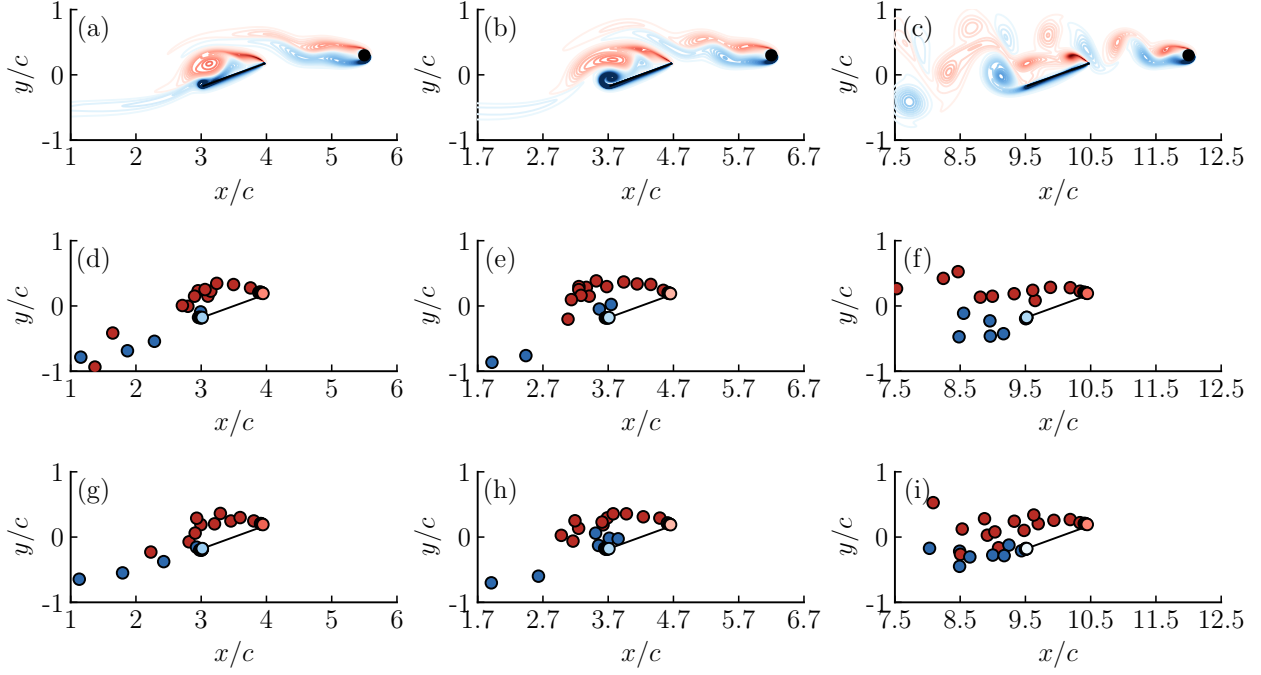


Figure 30: Snapshots of the vorticity distribution at $t^* = 3.5$ (left column), $t^* = 4.2$ (middle column) and $t^* = 10.0$ (right column) for an impulsively translating plate at 20° in a cylinder wake, predicted from [(a)-(c)] high-fidelity numerical simulation at Reynolds number 500, [(d)-(f)] inviscid vortex model with sEnKF, and [(g)-(i)] inviscid vortex model with ETKF.

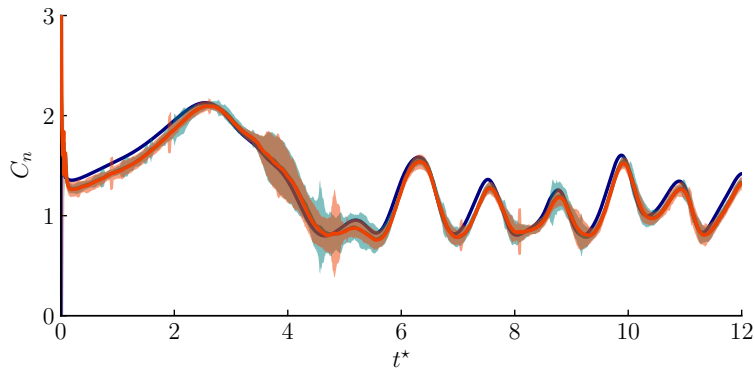


Figure 31: Normal force coefficient of an impulsively translating plate at 20° in a cylinder wake, from high-fidelity numerical simulation at Reynolds number 500 ■, mean over 100 realizations of the inviscid vortex model with sEnKF ■, mean over 100 realizations of the inviscid vortex model with ETKF ■. Shaded areas show the 95% confidence interval for the inviscid vortex model with sEnKF and ETKF

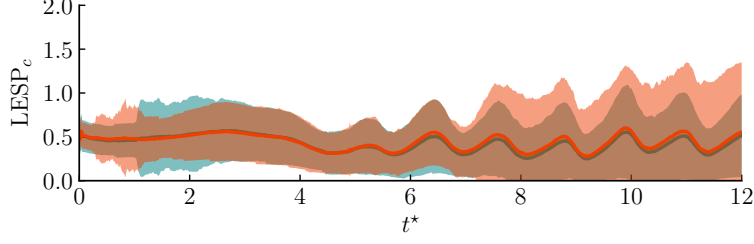


Figure 32: Time history of the ensemble mean value of the LESPc of an impulsively translating plate at 20° in a cylinder wake, averaged over 100 realizations, from the inviscid vortex model with sEnKF ■ and the inviscid vortex model with ETKF ■. Shaded areas show the 95% confidence interval for the inviscid vortex model with sEnKF and ETKF.

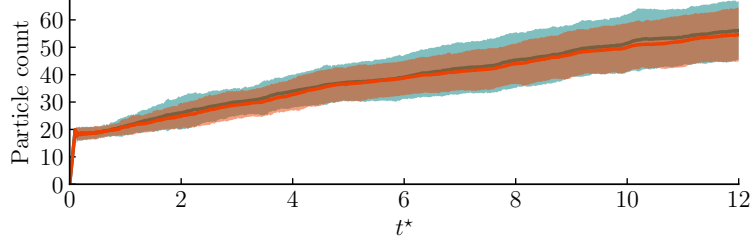


Figure 33: Time history of the particle count for an impulsively translating plate at 20° in a cylinder wake, averaged over 100 realizations, from the inviscid vortex model with sEnKF ■ and the inviscid vortex model with ETKF ■. Shaded areas show the 95% confidence interval for the inviscid vortex model with sEnKF and ETKF.

empirical state and observation covariance matrices are rank-deficient, and suffer from both sampling errors and spurious long-range correlations. While in many high-dimensional problems the EnKF can successfully track the state with very limited samples, this success is predicated on an appropriate regularization of the empirical Kalman gain. Predominant regularization techniques assume that the observations are *local*, i.e., that an observation only provides information about a subset of the state variables which are close in physical distance. This assumption is supported by the rapid decay of the correlations between state and observation variables. In this setting, distance localization cuts off long-range correlations [2].

For elliptic inverse problems, as we encounter with pressure-based observations of the flow, the conditional structure of the joint density of the observation and the state $\pi_{\mathbf{y}_k, \mathbf{x}_k | \mathbf{y}_{1:k-1}}$ is not localized, i.e., there is no rapid decay of the correlations as a function of the distance between variables. Hence, distance based localization in this problem results in biased estimators for the Kalman gain. Instead, we will exploit another kind of low-dimensional structure in the joint distribution. Our treatment draws inspiration from the fast multipole method (FMM) [45], which uses a hierarchical clustering of “source” and “target” elements to accelerate the calculation of the potential field from a large set of point singularities [33]. These clusterings are typically based only on spatial distance, but here we will use information from the prior distribution and observation model (i.e., observation operator and observation noise) to infer these clusters automatically in a way that works independently of the spatial distribution of the singularities and observation locations. To regularize the EnKF, we identify important directions in the state and observation spaces, perform the assimilation in these low-dimensional subspaces, and finally, lift the result to the original state space.

The details of our work are presented in [65]. First, we present the treatment of low-rank

structure in the case of a linear–Gaussian observation model in Section 4.3.1. The main result of this derivation is a factorization of the Kalman gain that exploits the existence of this low-dimensional subspace. Then, we extend this decomposition to the nonlinear–Gaussian setting in Section 4.3.2. For convenience, we continue to omit the time dependence subscripts of the variables in the rest of this section, since the analysis step does not involve time propagation.

4.3.1 Low-rank assimilation for the linear–Gaussian case

In this section, we consider the inference problem for the linear–Gaussian observation model:

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathcal{E}, \quad (49)$$

where the state is given by $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{X}}, \text{Cov}[\mathbf{X}])$, the observational error is given by $\mathcal{E} \sim \mathcal{N}(\mathbf{0}, \text{Cov}[\mathcal{E}])$, and $\mathcal{E}$ is independent of $\mathbf{X}$. The matrix $\mathbf{H} \in \mathbb{R}^{d \times n}$ is called the observation matrix. In order to identify the important assimilation directions, we first define the whitened variables $\tilde{\mathbf{X}} = \text{Cov}[\mathbf{X}]^{-1/2}(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}) \in \mathbb{R}^n$, $\tilde{\mathcal{E}} = \text{Cov}[\mathcal{E}]^{-1/2}\mathcal{E} \in \mathbb{R}^d$. We use $\mathbf{B}^{1/2}$, $\mathbf{B}^{-1/2}$ to denote a square root of the matrix $\mathbf{B}$ and its inverse, respectively. These whitened variables satisfy $\tilde{\mathcal{E}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\tilde{\mathbf{X}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Applying the same whitening to the observation variable, the observation model becomes

$$\tilde{\mathbf{Y}} = \text{Cov}[\mathcal{E}]^{-1/2}(\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}}) = \tilde{\mathbf{H}}\tilde{\mathbf{X}} + \tilde{\mathcal{E}}, \quad (50)$$

where $\boldsymbol{\mu}_{\mathbf{Y}} = \mathbf{H}\boldsymbol{\mu}_{\mathbf{X}}$ and $\tilde{\mathbf{H}} = \text{Cov}[\mathcal{E}]^{-1/2}\mathbf{H}\text{Cov}[\mathbf{X}]^{1/2} \in \mathbb{R}^{d \times n}$ is the whitened observation matrix. In this derivation we assume that $d \leq n$. We can now identify the important assimilation directions with a singular value decomposition (SVD) of the whitened observation matrix $\tilde{\mathbf{H}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^\top$, where $\mathbf{U} \in \mathbb{R}^{d \times d}$ and $\mathbf{V} \in \mathbb{R}^{n \times d}$ are the left and right singular vectors, and $\boldsymbol{\Lambda} \in \mathbb{R}^{d \times d}$ is the diagonal matrix of singular values. From this decomposition, we can rotate and project the whitened state on the subspace spanned by the columns of $\mathbf{V}^\top$ — $\check{\mathbf{X}} = \mathbf{V}^\top\tilde{\mathbf{X}} \in \mathbb{R}^d$ — and similarly rotate the whitened observational error and observation on the subspace spanned by the columns of $\mathbf{U}^\top$ — $\check{\mathcal{E}} = \mathbf{U}^\top\tilde{\mathcal{E}} \in \mathbb{R}^d$, $\check{\mathbf{Y}} = \mathbf{U}^\top\tilde{\mathbf{Y}} \in \mathbb{R}^d$. The SVD of the whitened observation matrix simultaneously identifies a pair of orthogonal bases for the state and observation spaces and an ordering for these directions. In [65] we show that the Kalman gain in the rotated space (where $\check{\mathbf{X}}$ and $\check{\mathbf{Y}}$ live) is given by:

$$\check{\mathbf{K}} = \boldsymbol{\Lambda}(\boldsymbol{\Lambda}^2 + \mathbf{I}_d)^{-1}. \quad (51)$$

Thus, the linear analysis map in the rotated space $\check{\mathbf{T}}$ is given by:

$$\check{\mathbf{T}}(\check{\mathbf{y}}, \check{\mathbf{x}}) = \check{\mathbf{x}} - \check{\mathbf{K}}(\check{\mathbf{y}} - \check{\mathbf{y}}^*) = \check{\mathbf{x}} - \boldsymbol{\Lambda}(\boldsymbol{\Lambda}^2 + \mathbf{I}_d)^{-1}(\check{\mathbf{y}} - \check{\mathbf{y}}^*), \quad (52)$$

where $\check{\mathbf{y}}^* = \mathbf{U}^\top \text{Cov}[\mathcal{E}]^{-1/2}\mathbf{y}^*$ denotes the realization of the assimilated data in the rotated observation space. In these rotated coordinates, the analysis map is local, i.e., each component of the rotated observation only updates one associated component of the rotated state. Using (51) and (52), we show in [65] that the Kalman gain $\mathbf{K} \in \mathbb{R}^{n \times d}$ in the original space factorizes as

$$\mathbf{K} = \text{Cov}[\mathbf{X}]^{1/2}\mathbf{V}\boldsymbol{\Lambda}(\boldsymbol{\Lambda}^2 + \mathbf{I}_d)^{-1}\mathbf{U}^\top \text{Cov}[\mathcal{E}]^{-1/2}. \quad (53)$$

This factorization of $\mathbf{K}$ and its application in the analysis map of the Kalman filter (24) provide a concise summary of the inference process in the low-rank informative subspaces: the innovation term $(\mathbf{y} - \mathbf{y}^*)$ is whitened and rotated by applying $\mathbf{U}^\top \text{Cov}[\mathcal{E}]^{-1/2}$, assimilated in this new set of coordinates using (51), and finally lifted to the original state space by applying $\text{Cov}[\mathbf{X}]^{1/2}\mathbf{V}$. In the whitened space (where $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ live), it is easy to see that $\mathbf{V}\boldsymbol{\Lambda}(\boldsymbol{\Lambda}^2 + \mathbf{I}_d)^{-1}\mathbf{U}^\top$ constitutes the singular value decomposition of the Kalman gain. From the decay of the singular values $\boldsymbol{\Lambda}$, we can instead use a truncated singular value decomposition of $\tilde{\mathbf{H}} \approx \mathbf{U}_r\boldsymbol{\Lambda}_r\mathbf{V}_r^\top$, where $r \leq d$,

$\mathbf{U}_r \in \mathbb{R}^{r \times r}$, $\mathbf{V}_r \in \mathbb{R}^{n \times r}$ are the first r left and right singular vectors, and $\mathbf{\Lambda}_r \in \mathbb{R}^{r \times r}$. A rank- r approximation of the Kalman gain is then given by $\mathbf{K}_r = \text{Cov}[\mathbf{X}]^{1/2} \mathbf{V}_r \mathbf{\Lambda}_r (\mathbf{\Lambda}_r^2 + \mathbf{I}_r)^{-1} \mathbf{U}_r^\top \text{Cov}[\mathcal{E}]^{-1/2}$. We should emphasize that, without the truncation of the SVD, eq. (53) is an exact factorization of the Kalman gain. With the truncated SVD, the state and observation variables are no longer just rotated but also projected to a subspace of dimension $r \leq \min(d, n)$.

4.3.2 Low-rank assimilation for the nonlinear-Gaussian case

In this section, we consider the nonlinear observation model of eq. (11), recalled for reference:

$$\mathbf{Y} = \mathbf{h}(\mathbf{X}) + \mathcal{E}, \quad (54)$$

where the state is $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{X}}, \text{Cov}[\mathbf{X}])$, the observational error is given by $\mathcal{E} \sim \mathcal{N}(\mathbf{0}, \text{Cov}[\mathcal{E}])$, and $\mathcal{E}$ is independent of $\mathbf{X}$. To handle a nonlinear observation operator with the linear prior-to-posterior transformation (24), one simple approximation is to use the Jacobian of the observation operator about the prior mean as the observation matrix [2]. Unfortunately, this treatment is only viable for functions which are well approximated by a linear function over the bulk of the prior distribution. This section generalizes the treatment of the previous section for a nonlinear-Gaussian observation model; see [65] for further details.

In the nonlinear and non-Gaussian setting, Cui et al. [17] showed that the most important assimilation directions in the whitened state space can be identified by the eigenvectors of the *state space Gramian*, which in the case of Gaussian $\mathcal{E}$ reduces to:

$$\mathbf{C}_{\mathbf{X}} = \int \left(\text{Cov}[\mathcal{E}]^{-1/2} \nabla \mathbf{h}(\mathbf{x}) \text{Cov}[\mathbf{X}]^{1/2} \right)^\top \left(\text{Cov}[\mathcal{E}]^{-1/2} \nabla \mathbf{h}(\mathbf{x}) \text{Cov}[\mathbf{X}]^{1/2} \right) d\pi_{\mathbf{X}}(\mathbf{x}) \in \mathbb{R}^{n \times n}, \quad (55)$$

where the expectation is taken over the prior distribution. As expected, eq. (55) reverts to $\tilde{\mathbf{H}}^\top \tilde{\mathbf{H}}$ in the linear-Gaussian case. $\mathbf{C}_{\mathbf{X}}$ is *positive semi-definite (PSD)* and its eigendecomposition can be written as $\mathbf{C}_{\mathbf{X}} = \mathbf{V} \mathbf{\Lambda}_{\mathbf{X}}^2 \mathbf{V}^\top$, where $\mathbf{V} \in \mathbb{R}^{n \times n}$ is an orthogonal basis for the whitened state space with associated eigenvalues $\mathbf{\Lambda}_{\mathbf{X}}^2 \in \mathbb{R}^{n \times n}$.

Inspired by the treatment in the state space, we propose to use the eigenvectors of the *observation space Gramian* $\mathbf{C}_{\mathbf{Y}}$ to select the important assimilation directions in the whitened observation space:

$$\mathbf{C}_{\mathbf{Y}} = \int \left(\text{Cov}[\mathcal{E}]^{-1/2} \nabla \mathbf{h}(\mathbf{x}) \text{Cov}[\mathbf{X}]^{1/2} \right) \left(\text{Cov}[\mathcal{E}]^{-1/2} \nabla \mathbf{h}(\mathbf{x}) \text{Cov}[\mathbf{X}]^{1/2} \right)^\top d\pi_{\mathbf{X}}(\mathbf{x}) \in \mathbb{R}^{d \times d}. \quad (56)$$

The matrix $\mathbf{C}_{\mathbf{Y}}$ is also PSD and has the eigendecomposition $\mathbf{C}_{\mathbf{Y}} = \mathbf{U} \mathbf{\Lambda}_{\mathbf{Y}}^2 \mathbf{U}^\top$, where $\mathbf{U} \in \mathbb{R}^{d \times d}$ is an orthonormal basis for the whitened observation space with associated eigenvalues $\mathbf{\Lambda}_{\mathbf{Y}}^2 \in \mathbb{R}^{d \times d}$. For convenience, we assume that the eigenvectors of $\mathbf{C}_{\mathbf{X}}$ and $\mathbf{C}_{\mathbf{Y}}$ are ordered by decreasing eigenvalues. For a nonlinear observation model, the eigenvalues of the state and observation Gramians can be different. In practice, we use Monte-Carlo approximations of $\mathbf{C}_{\mathbf{X}}$ and $\mathbf{C}_{\mathbf{Y}}$ that are estimated using the prior samples to identify the important subspaces. Depending on the inference problem, the Jacobian $\nabla \mathbf{h} \in \mathbb{R}^{d \times n}$ of the observation operator $\mathbf{h}$ with respect to the state components can be computed either analytically, with automatic differentiation, with complex step differentiation, or, in the worst case, from finite differences. To perform the low-rank assimilation, we only retain the first $r_{\mathbf{X}} \leq n$ eigenmodes for the state space, and $r_{\mathbf{Y}} \leq d$ eigenmodes for the observation space. The ranks $r_{\mathbf{X}}$ and $r_{\mathbf{Y}}$ can be tuned independently based on the decay of $\mathbf{\Lambda}_{\mathbf{X}}^2$ and $\mathbf{\Lambda}_{\mathbf{Y}}^2$; typically we recommend setting these ranks to achieve a threshold $\alpha \in [0, 1]$ for the cumulative normalized energy of the eigenvalue spectra, e.g., in the state space, we set $r_{\mathbf{X}} = \min \{r_{\alpha} \in \llbracket 1, n \rrbracket \mid E_{\mathbf{X}, r_{\alpha}} \geq \alpha\}$, where $E_{\mathbf{X}, r_{\alpha}} = \sum_{i=1}^{r_{\alpha}} \lambda_{\mathbf{X}, i}^2 / \sum_{i=1}^n \lambda_{\mathbf{X}, i}^2$. The new projected state and projected observation variables are

defined as $\check{\mathbf{X}} = \mathbf{V}_{r_{\mathbf{X}}}^\top \tilde{\mathbf{X}} = \mathbf{V}_{r_{\mathbf{X}}}^\top \text{Cov}[\mathbf{X}]^{-1/2}(\mathbf{X} - \mu_{\mathbf{X}}) \in \mathbb{R}^{r_{\mathbf{X}}}$ and $\check{\mathbf{Y}} = \mathbf{U}_{r_{\mathbf{Y}}}^\top \tilde{\mathbf{Y}} = \mathbf{U}_{r_{\mathbf{Y}}}^\top \text{Cov}[\mathbf{Y}]^{-1/2}(\mathbf{Y} - \mu_{\mathbf{Y}}) \in \mathbb{R}^{r_{\mathbf{Y}}}$, where $\mathbf{V}_{r_{\mathbf{X}}}, \mathbf{U}_{r_{\mathbf{Y}}}$ denote the first $r_{\mathbf{X}}, r_{\mathbf{Y}}$ columns of $\mathbf{V}, \mathbf{U}$, respectively.

We now outline how to obtain the factorization of the Kalman gain in the nonlinear setting. In the linear-Gaussian case, the Gramians in (55) and (56) revert to $\mathbf{C}_{\mathbf{X}} = \tilde{\mathbf{H}}^\top \tilde{\mathbf{H}}$ and $\mathbf{C}_{\mathbf{Y}} = \tilde{\mathbf{H}} \tilde{\mathbf{H}}^\top$, respectively. From the definition of the SVD of the whitened observation matrix $\tilde{\mathbf{H}}$, we have $\tilde{\mathbf{H}} \mathbf{v}_i = \lambda_i \mathbf{u}_i$ for some left singular vector $\mathbf{u}_i$, singular value λ_i and right singular vector $\mathbf{v}_i$. This results in the eigendecompositions $\mathbf{C}_{\mathbf{X}} \mathbf{u}_i = \lambda_i^2 \mathbf{u}_i$ and $\mathbf{C}_{\mathbf{Y}} \mathbf{v}_i = \lambda_i^2 \mathbf{v}_i$. Therefore, if the triplet $\{\mathbf{u}_i, \lambda_i, \mathbf{v}_i\}$ is obtained from the eigendecomposition of $\mathbf{C}_{\mathbf{X}}$ and $\mathbf{C}_{\mathbf{Y}}$, the sign of the singular vectors is lost, i.e., we only have $\tilde{\mathbf{H}} \mathbf{v}_i = \pm \lambda_i \mathbf{u}_i$. For a nonlinear observation model, the factorization of the Kalman gain given in eq. (53) is no longer applicable and needs to be generalized. In the rotated and whitened space, the analysis map is given by:

$$\check{\mathbf{T}}(\check{\mathbf{y}}, \check{\mathbf{x}}) = \check{\mathbf{x}} - \check{\mathbf{K}}(\check{\mathbf{y}} - \check{\mathbf{y}}^*), \quad (57)$$

where $\check{\mathbf{K}} = \text{Cov}[\check{\mathbf{X}}\check{\mathbf{Y}}] \text{Cov}[\check{\mathbf{Y}}]^{-1} \in \mathbb{R}^{r_{\mathbf{X}} \times r_{\mathbf{Y}}}$ is the Kalman gain in the informative space with $\text{Cov}[\check{\mathbf{X}}\check{\mathbf{Y}}] \in \mathbb{R}^{r_{\mathbf{X}} \times r_{\mathbf{Y}}}$ and $\text{Cov}[\check{\mathbf{Y}}]^{-1} \in \mathbb{R}^{r_{\mathbf{Y}} \times r_{\mathbf{Y}}}$. In the original space, we get

$$\mathbf{T}(\mathbf{y}, \mathbf{x}) = \mathbf{x} - \text{Cov}[\mathbf{X}]^{1/2} \mathbf{V} \check{\mathbf{K}} \mathbf{U}^\top \text{Cov}[\mathbf{E}]^{-1/2}(\mathbf{y} - \mathbf{y}^*). \quad (58)$$

This new factorization of the Kalman gain nicely generalizes eq. (53) to an inference problem with a nonlinear observation model. We emphasize that the sign issue presented above is obviated by absorbing it into the definition of $\check{\mathbf{K}}$. Let us remark that the definition of $\check{\mathbf{K}}$ reduces to the diagonal matrix $\mathbf{\Lambda}(\mathbf{\Lambda}^2 + \mathbf{I})^{-1}$ in the linear-Gaussian case, where the matrices $\mathbf{U}, \mathbf{V}$ are obtained from the SVD of $\tilde{\mathbf{H}}$. Indeed, eq. (58) constitutes a *change of coordinates for the Kalman gain* between the original space and the informative space. We denote the algorithm applying the linear update in (58) in each analysis step as the low-rank EnKF (LREnKF). A pseudo-code for the proposed low-rank EnKF is presented in [65].

In comparison, the sEnKF estimates the Kalman gain in the original space, a linear operator of dimensions $n \times d$, where n and d are the dimensions of the space and observation spaces. Leveraging the dimension reduction offered by the state and observation Gramians, the LREnKF only has to estimate the Kalman gain in the informative subspaces of $\mathbf{X}$ and $\mathbf{Y}$, whose dimension is only $r_{\mathbf{X}} \times r_{\mathbf{Y}}$, where $r_{\mathbf{X}}$ and $r_{\mathbf{Y}}$ can be significantly smaller than n and d , respectively. To appreciate the benefit of this dimension reduction in the inference problem, it is useful to recall the classical bias-variance trade-off in any statistical learning problem [41]: Given an ensemble size M , the error in the estimation of the Kalman gain $\mathbf{K}$ can be decomposed into a variance term and a bias term. For a limited ensemble size (compared to n and d), the estimated Kalman gain from the vanilla sEnKF will have a large variance, potentially leading to a filter with diverging state estimation error, as is the case for small ensemble size in the example problems. We show later in this paper that with the proposed dimension reduction, the Kalman gain in the informative subspaces $\check{\mathbf{K}}$ contains a much smaller number of entries than $\mathbf{K}$. This results in smaller variance. The bias in computing $\check{\mathbf{K}}$ can be controlled by setting the ranks of the state and observation projections. Due to the rapid decay of the spectrum of the state and observation Gramians, increasing the rank of the projected subspace beyond a certain ratio of the cumulative energy will only marginally reduce the bias, but greatly increase the variance. In other words, we choose the ranks $r_{\mathbf{X}}, r_{\mathbf{Y}}$ to capture the column/row space of the Kalman gain and make the bias small for any finite-sample-estimator of $\mathbf{K}$. Finally, we should stress that without dimension reduction (i.e., $r_{\mathbf{X}} = n$ and $r_{\mathbf{Y}} = d$), the sEnKF and the LREnKF are strictly equivalent. Indeed, the sEnKF estimates the Kalman gain $\mathbf{K} \in \mathbb{R}^{n \times d}$ in the original coordinates, while the LREnKF estimates $\check{\mathbf{K}} \in \mathbb{R}^{n \times d}$ in the informative coordinates and lifts it to the original coordinates according to (58).

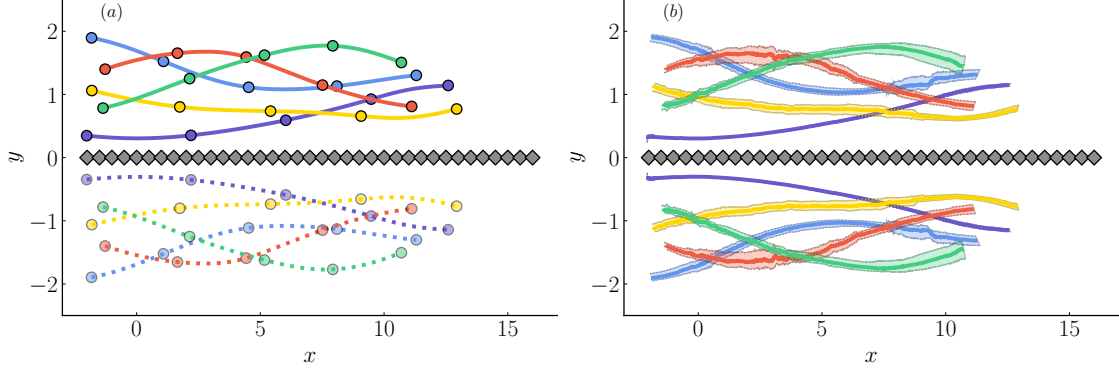


Figure 34: (a): Schematic of the setup. True trajectories of the point vortices (for one realization of the system) are represented by colored lines. The location of the point vortices, sampled every three convective times, is depicted by colored dots. Fainter dashed lines and fainter dots refer to the image point vortices. Location of the pressure sensors are depicted by grey diamonds. (b): Estimation of the trajectories of the vortices with the LREnKF for $M = 40$ with the ranks $r_{\mathbf{x}}$ and $r_{\mathbf{y}}$ set to capture 99% of the cumulative energy spectra. Solid lines depict the time history of the median posterior estimate for the position of the different point vortices. Fainted areas show the 5% and 95% quantiles of the posterior estimate for the position of the point vortices.

4.3.3 Estimation of the leading directions from samples

In practice, we only have access to limited samples $\{\mathbf{x}^1, \dots, \mathbf{x}^M\}$ from the prior distribution to estimate the state and observation Gramians $\mathbf{C}_{\mathbf{x}}, \mathbf{C}_{\mathbf{y}}$ and their associated eigenvectors. To understand how well we can approximate these leading eigenvectors from the empirical Gramians $\hat{\mathbf{C}}_{\mathbf{x}}, \hat{\mathbf{C}}_{\mathbf{y}}$, we can make use of a corollary of the Davis–Kahan theorem [24, 100]. We can apply the Davis–Kahan theorem to the pairs $(\mathbf{\Gamma}, \hat{\mathbf{\Gamma}}) = (\mathbf{C}_{\mathbf{x}}, \hat{\mathbf{C}}_{\mathbf{x}})$, and $(\mathbf{C}_{\mathbf{y}}, \hat{\mathbf{C}}_{\mathbf{y}})$. Two important conclusions can be drawn from this result. First, it follows from the theorem that the error in the subspace spanned by the first r eigenvectors of $\hat{\mathbf{\Gamma}}$ converges to zero at the same rate that $\hat{\mathbf{\Gamma}}$ converges to $\mathbf{\Gamma}$ in the Frobenius norm, as $M \rightarrow \infty$. See Tropp [93] for relevant results on the convergence rates of random matrices. Second, the error in the estimated r -dimensional leading eigenspace depends on the difference between consecutive eigenvalues $\lambda_r^2 - \lambda_{r+1}^2$, which is called the spectral gap. Specifically, the inverse of the spectral gap is a useful indicator of the estimation error in the informative subspaces. Intuitively, two eigenvectors with close eigenvalues are difficult to distinguish. For the fluid dynamics problems considered in this report, the spectra of the state and observation Gramians decay rapidly, and hence the leading eigenvalues are well separated. This suggests that we can estimate the leading eigendirections of $\mathbf{C}_{\mathbf{x}}, \mathbf{C}_{\mathbf{y}}$ with relatively few samples M , and that the row and column spaces of the Kalman gain can be accurately captured using only a small number of such eigendirections.

4.3.4 Example: inference of point vortices from wall pressure observations

In this second example, we compare the LREnKF with the sEnKF for estimating the positions and strengths of a collection of point vortices advecting along a horizontal wall. We rely on pressure observations collected along the wall to estimate the state. These observations were generated from the same observation model used for inference, thereby making this a *twin experiment* [2]. We consider a set of $N = 5$ point vortices located at $\{z_1, \dots, z_N\}$ with circulations $\{\Gamma_1, \dots, \Gamma_N\}$, and a set of d evaluation points located at $\{z'_1, \dots, z'_d\}$, see Fig. 34 (a). The state variable $\mathbf{x}$ contains the positions and circulations of the N point vortices: $\mathbf{x} = [x_1, y_1, \Gamma_1, \dots, x_N, y_N, \Gamma_N]^\top \in \mathbb{R}^{3N}$.

To enforce the no-flow-through condition along the x axis, we use the method of images and aug-

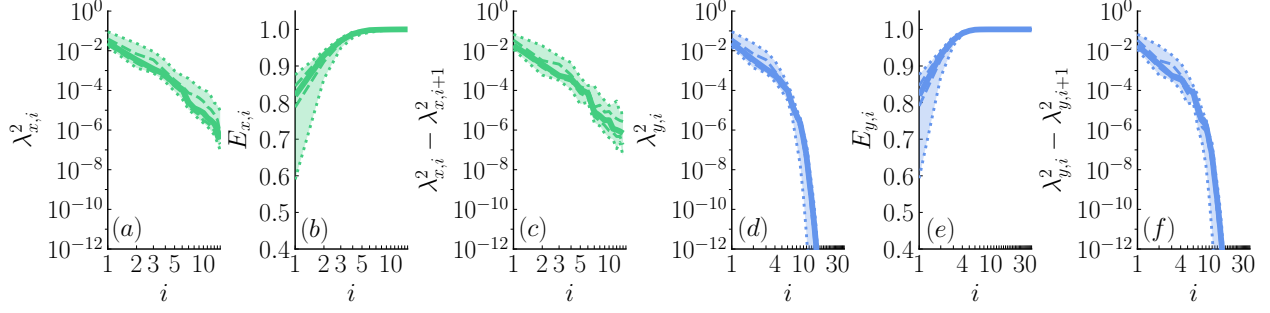


Figure 35: Median spectrum of the state Gramian $\mathbf{C}_{\mathbf{X}}$ (left panels (a)-(c)) and observation Gramian $\mathbf{C}_{\mathbf{Y}}$ (right panels (d)-(f)) over the time interval $[0, 12]$. Statistics are obtained from a run of the sEnKF with $M = 1000$. Panels [(a), (d)]: Median eigenvalues of $\mathbf{C}_{\mathbf{X}}$ and $\mathbf{C}_{\mathbf{Y}}$. Panels [(b), (e)]: Median normalized cumulative energy $E_i = \sum_{j=1}^i \lambda_j^2 / \sum_j \lambda_j^2$ of $\mathbf{C}_{\mathbf{X}}$ and $\mathbf{C}_{\mathbf{Y}}$. Panels [(c), (f)]: Median spectral gap $\lambda_i^2 - \lambda_{i+1}^2$ of $\mathbf{C}_{\mathbf{X}}$ and $\mathbf{C}_{\mathbf{Y}}$. The abscissa axis is in log scale. Dashed lines depict the 25% and 75% quantiles. Dotted lines depict the 5% and 95% quantiles.

ment our collection of vortices with another set of N vortices at the conjugate positions $\{\bar{z}_1, \dots, \bar{z}_N\}$ with opposite circulation $\{-\Gamma_1, \dots, -\Gamma_N\}$. We emphasize that these mirrored vortices are only an artifice to enforce the no-flow-through in the forecast step, and play no role in the analysis step. We add a freestream flow directed along increasing x values with velocity $\mathbf{U}_{\infty} = [1, 0]^{\top}$. The initial position of the J th point vortex is generated randomly with the form $z_J + \rho_r \exp(i\theta)$ where $\{z_1, \dots, z_N\} = \{-2.0 + 0.3i, -1.9 + 1.9i, -1.8 + 1.1i, -1.3 + 1.4i, -1.4 + 0.8i\}$, ρ_r is drawn from $\mathcal{N}(0, 0.1)$ and θ is drawn from the uniform distribution on $[0, \pi]$. The initial circulation of the point vortices is drawn from $\mathcal{N}(0.4, 0.1)$. The collection of point vortices is advanced using a forward Euler scheme with time step $\Delta t = 1 \times 10^{-3}$. The observation vector $\mathbf{y}$ consists of pressure observations at $d = 37$ locations linearly distributed with an interspace of 0.5 along the segment $[-2, 16]$, see Fig. 34 (a). We use a Gaussian observation noise with zero mean and covariance 1×10^{-4} . We obtain the Jacobian of the observation operator used in the state and observation Gramians $\mathbf{C}_{\mathbf{X}}$ and $\mathbf{C}_{\mathbf{Y}}$ by analytical differentiation of the unsteady Bernoulli equation.

We assess the performance of the two filters with a *twin experiment*. We draw a random initial condition for the positions and strengths of the different point vortices. Then we simulate the dynamics of the point vortices over the time interval $[0, 12]$ (corresponding to 12000 assimilation cycles with the choice of time step Δt). The true state at the time step k is denoted by $\mathbf{x}_k^*$. At every time step, we generate the noisy pressure observation at the sample locations based on the true state $\mathbf{x}_k^*$. The realization of the true observation vector at the time step is denoted by $\mathbf{y}_k^*$. In the twin experiment, we seek to estimate the true state $\mathbf{x}_k^*$ only from the knowledge of the noisy and indirect observations $\mathbf{y}_k^*$ and the prior.

First, we assess the spectra and leading subspaces of the state and observation Gramians. Fig. 35 depicts the median eigenvalues of the state and observation Gramians over the time interval $[0, 12]$. We run the sEnKF with ensemble size $M = 1000$ to generate the samples used to compute empirical Gramians in each analysis step. We choose this larger ensemble size to reduce sampling errors. The median ranks $r_{\mathbf{X}}$ to capture 80%, 90%, 95%, and 99% of the cumulative energy of $\mathbf{C}_{\mathbf{X}}$ are 1, 2, 3, and 5, respectively. Similarly, the median ranks $r_{\mathbf{Y}}$ to capture 80%, 90%, and 99% of the cumulative energy of $\mathbf{C}_{\mathbf{Y}}$ are 1, 2, and 4, respectively. The sharp decay of the spectra of $\mathbf{C}_{\mathbf{X}}$ and $\mathbf{C}_{\mathbf{Y}}$ supports our hypothesis of low-rank structure existing in the prior-to-posterior update. A low-dimensional subspace of the pressure observations is only informative along a limited number of directions in the state space. Fig. 36 shows the 1, 2, 3 and 8th eigenvectors of the observation

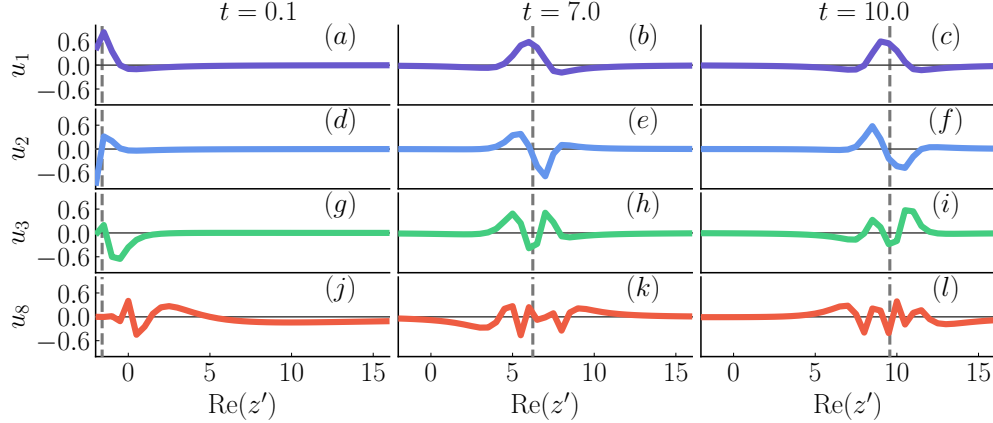


Figure 36: 1, 2, 3 and 8th eigenvectors of the observation Gramian $\mathbf{C}_Y$, i.e., observation modes u_i s, at $t = 0.1$ (left column), $t = 7.0$ (middle column), and $t = 10.0$ (right column). (a)-(c): first mode, (d)-(f): second mode, and (g)-(i): third mode, (j)-(l): eighth mode. The dashed grey vertical line depicts the x component of the position of the centroid of the vortices. The grey horizontal line corresponds to 0 ordinate. Statistics are obtained from a run of the sEnKF with $M = 1000$. The eigenvectors have unit norm.

Gramian at the three times: $t = 0.1, 7.0$, and 10.0 . These modes clearly illustrate how information is extracted from the different pressure discrepancies during the analysis step. The first observation mode corresponds to a spatially weighted average of the pressure discrepancies about the mean x coordinate of the position of the vortices (depicted by the dashed vertical line on the panels of Fig. 36). The second mode captures differences between the pressure observations collected on the left and right side of the centroid location. The higher modes follow the same pattern and act as refined stencils with growing support to extract higher order features from the pressure observations. The leading eigenvectors of the state Gramian are more difficult to interpret, due to the manner in which vortices are coupled in the determination of pressure.

Next, we compare the performance of the LREnKF with the stochastic EnKF (sEnKF). The comparison of these two filters is natural, as the LREnKF without dimension reduction (i.e., for $r_X = n = 3N$ and $r_Y = d$) reverts to the sEnKF. For a given ensemble size M , one could perform a parametric study to determine the ranks r_X and r_Y which give the best performance of the LREnKF. We found, however, that using fixed ranks for all assimilation cycles is suboptimal as we face one of two possible scenarios: either the ranks are too large, leading to additional variance, or the ranks are too small, leading to additional bias. Furthermore, we expect the dimension of the informative subspaces may vary over time as the flow field evolves. Instead, we select the ranks r_X and r_Y adaptively at each time step based on a predetermined ratio of the cumulative energy of the Gramians $\mathbf{C}_X$ and $\mathbf{C}_Y$. The performance of each filter is assessed using the *root mean squared error (RMSE)* for tracking the true state $\mathbf{x}_k^*$. The best estimate of the true state at a given assimilation time is given by the mean of the analysis ensemble at this step, that we denote by $\bar{\mathbf{x}}_k^a$. We define the RMSE at one time between the true state $\mathbf{x}_k^*$ and the estimate $\bar{\mathbf{x}}_k^a$ as $\text{RMSE}_k = \|\mathbf{x}_k^* - \bar{\mathbf{x}}_k^a\|_2 / \sqrt{3N}$, where N is the number of point vortices. The mean RMSE of each filter is computed over the time interval $[8, 12]$ (i.e., the last 4000 assimilation steps), to remove any influence of the initial conditions and to ensure empirical stationarity of the filters. The first 8000 assimilation steps are called the spin-up phase and are discarded. The uncertainty in the RMSE is quantified by the 5%, 25%, 75% and 95% quantiles of the time-averaged RMSE over 50 realizations

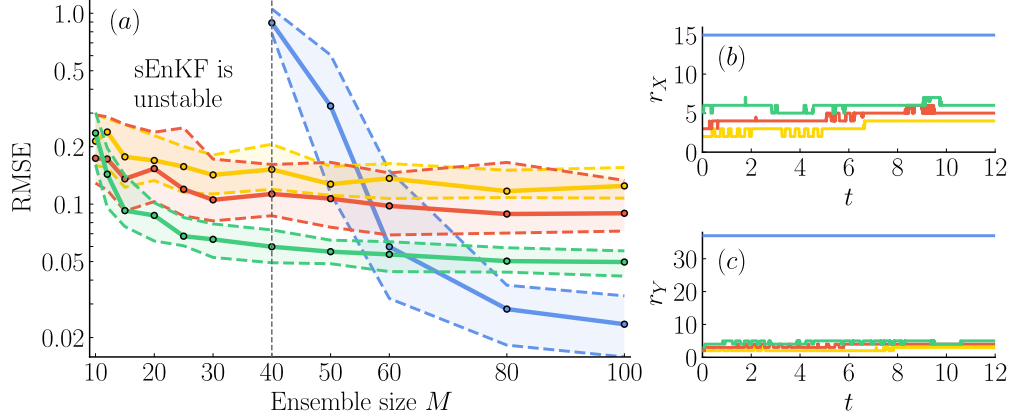


Figure 37: Left column (a): Time-averaged evolution of the median RMSE with the ensemble size M (computed over 50 realizations) with the sEnKF (blue), and with the LREnKF for different ratios of the cumulative energy: 85% (yellow), 95% (orange), and 99% (green). The tracking performance of the sEnKF is unstable for $M < 40$. Dashed lines depict the 25% and 75% quantiles. Right column [(b)-(c)]: Time-history of the median ranks r_X and r_Y of the LREnKF for $M = 30$ (computed over 50 realizations) for different ratios of the cumulative energy of C_X , C_Y : 85% (yellow), 95% (orange), and 99% (green). The dimension of the state and observation spaces, namely n and d , are depicted for comparison in blue.

of the same experiment for different initial ensembles. The computational time for one time step of the sEnKF and LREnKF with $M = 50$ are 2.3 ms and 8.0 ms, respectively. We provide directions to reduce the computational cost of forming the Gramians C_X and C_Y in the conclusion.

Fig. 37(a) shows the time-averaged RMSE for the LREnKF and the sEnKF. We assess the performance of the LREnKF for 85%, 95%, and 99% of the normalized cumulative energy of C_X and C_Y . For $M < 40$, the sEnKF is unstable, while the LREnKF accurately estimates the true state, even for $M = 10$. For $M \in [40, 60]$, the RMSE of the sEnKF significantly increases as M decreases. For $M = 40$, the median RMSE of the sEnKF is 0.9, while the median RMSEs of the LREnKF with the different energy ratios are smaller than 0.16. For large M , the RMSE of the LREnKF decreases with the energy ratio, but remains larger than the RMSE of the sEnKF. The proposed low-rank approximation of the Kalman gain can lead to a small bias for large ensemble sizes. The regularization technique proposed in this work, however, is designed for the small ensemble size regime; one can always increase the energy ratio (beyond 99%) to recover the performance of the sEnKF for large ensemble sizes. The RMSE of the LREnKF with the different energy ratios decreases for $M \in [10, 40]$, then plateaus. In the small ensemble size regime, adding samples improves the estimate of the dominant directions (see the Davis–Kahan theorem [65]), leading to a reduction in variance. As we continue to increase the number of samples for a given number of dimensions to estimate, the variance becomes smaller than the bias from using a truncated basis to approximate the Kalman gain. For $M = 10$, the LREnKF with 85% energy ratio performs better than the LREnKF with 99% energy ratio. For $M = 10$, the ranks needed to achieve 99% is large, leading to additional variance. On the other hand, 85% energy requires a much smaller number of entries to estimate, leading to less variance compared to the bias of this approximation. We recommend choosing the target energy ratio based on the available ensemble size.

Fig. 37(b)-(c) show the history of the median ranks r_X and r_Y (computed over 50 realizations) for the different ratios of the cumulative energy for $M = 30$. The ranks plateau over the time window and remain small compared to the dimension of the state and observation spaces. For the

LREnKF with 99% of the cumulative energy, $r_{\mathbf{x}}, r_{\mathbf{y}}$ are smaller than 8, 6, respectively. Fig. 34 (b) shows one estimate of the trajectories of the vortices over the time interval $[0, 12]$. The results are obtained from the LREnKF with $M = 50$ with the ranks $r_{\mathbf{x}}$ and $r_{\mathbf{y}}$ set to capture 99% of the energy spectra. We quantify the uncertainty with the 5% and 95% quantiles of the posterior for the point vortex positions. We observe an excellent agreement with the true trajectories with a time-averaged RMSE of 0.038.

Earlier in this section, we argued that distance localization schemes will be harmful to regularize inference problems with elliptic observation models, as studied herein. We conclude this section by providing an *a posteriori* justification based on the results of the sEnKF run with a large ensemble ($M = 1000$). Fig. 38 (a)-(c) shows the magnitude of the cross-covariance between the positions and strengths of the different point vortices and the pressure observations at $t = 1.0$. We notice that the cross-covariance entries decay very quickly on a small support about each point vortex. Past a certain radius, however, the cross-covariance entries quickly increase, and beyond this point, they only decay algebraically. More precisely, we observe a decay as the inverse square of the distance between the point vortex and the pressure observation. This decay rate is expected from the pressure field from a Poisson equation (or unsteady Bernoulli equation). In Fig. 38(c), the covariance between the circulation of three point vortices and pressure observations at a distance of 5 and 10 units away is still about 20% and 10% of its maximal value, respectively. This algebraic decay clearly violates the assumption of rapidly decaying correlations of distance localization schemes. The cross-covariance of the components of the position and the strength for the different point vortices can have different algebraic decay rates as a function of the distance. Even for a particular vortex, the variations of the cross-covariance for different components of the position and the strength are drastically different. In unreported results, we observe that the decay rate of the correlations also varies over time as the vortices evolve. In the best case, this suggests that three different localization radius are needed and would have to be tuned independently: an impractical procedure. These different facts support our *a priori* hypothesis that distance localization is not suited to regularize filtering problems with elliptic observation operators.

4.4 EnKF-based estimation of bias error in coarse-grid CFD

The success of a dynamical estimator is contingent on the accuracy of the model chosen to represent the dynamics, and essentially any aspect of the chosen model that falls short of reality is a potential source of bias. In the vortex modeling approach that we have profiled in the previous sections, there are two notable sources of error. The dynamical model is an exact solution, but to the Euler rather than Navier–Stokes equations (and thus misses viscous effects). Furthermore, the vorticity is represented by a finite number of point vortices, a reduced form of the true distributed vorticity. We have not yet investigated the bias errors in our state estimate in this context, and it remains the subject of future work.

In contrast, in a grid-based solution of Navier-Stokes equations, particular model errors include resolution (truncation error), turbulence models, and uncertain initial and boundary conditions. Some of these involve a compromise between accuracy and cost to achieve a certain estimation rate with available computational power. Recently developed multi-fidelity [56] and multi-level [52] Monte Carlo approaches would allow for a more efficient use of the available resources, as they leverage information from a hierarchy of models with different fidelities (and costs) in order to minimize the overall cost of achieving a given estimation accuracy. Although promising, the integration of these techniques to the Kalman filtering framework is still work in progress with associated limitations. In Hoel et al. [52], for instance, they assume that the ensemble size is the same for all resolution levels. Since we almost always use the minimum ensemble size that allow us to capture the most important degrees of freedom of the system, this approach will necessarily

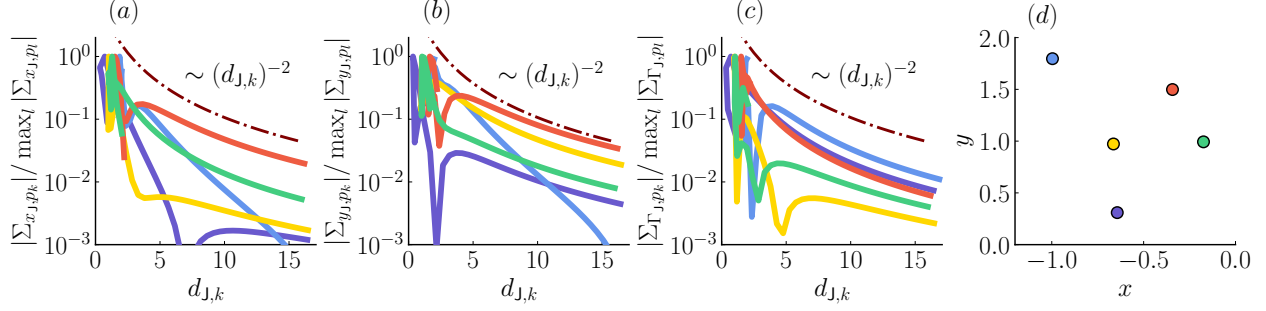


Figure 38: Panel (a), (b), (c) depicts the magnitude of the empirical cross-covariance between the x , y coordinate and strength of the J point vortex with the k th pressure observation at $t = 1.0$, respectively. The curves are plotted against the distance $d_{J,k} = |z_J - z'_k|$, where $z_J = x_J + iy_J$ is the (complex) position of the J th point vortex, and z'_k is the (complex) location of the k th pressure sensor. The magnitude of the cross-covariances are normalized by the maximum cross-covariance (in magnitude) between the x , y coordinate or strength of the J point vortex with the different pressure observation, respectively. The dashed grey vertical line depicts the x component of the position of the centroid of the vortices. Red dashed and dotted curve depicts the algebraic decay $O(1/d_{J,k}^2)$. (d) Mean position of the different point vortices at $t = 1.0$. The same color is used to depict the properties of a point vortex on the panels (a)-(d). The results are obtained from a run of the sEnKF for $M = 1000$.

lead to a more expensive estimator than their single-level counterparts.

Friedland [40] was one of the first to propose a direct treatment of the model error. He proposed a two-stage sequential estimator, termed the Separate-bias Kalman filters (SepKF), in which model state and bias vector were treated independently. Dee and Da Silva [25] provided a rigorous method to independently estimate and sequentially correct for forecast bias. Drecourt et al. [29] compared this method to the colored-noise Kalman filter (ColKF), in which the state vector is augmented to account for noise processes modeled by autoregressive models. Trémolet [92] studied the introduction of bias models to a 3D/4D-Var framework, but assumed a full-rank time-invariant representation of the bias.

These works, however, all assume that the observation model, the function applied to the estimated system state to determine the observable to be compared to measurement, is unbiased. Many of the errors mentioned above can render *both* the dynamics and the observation models biased, with non-zero mean in any ensemble. Biased errors can be particularly harmful in the context of ensemble methods with ensemble size much smaller than the number of degrees of freedom (which is the case for most of the practical applications). Even if the underlying dynamics are inherently low-dimensional, i.e. a few POD modes are responsible for most the system variance, discretization errors may have little support on those modes. Since the corrections applied to forecast state lie in the low-dimensional subspace spanned by the prior ensemble perturbations, large dynamic bias renders the true state of the system unreachable. The observation model is also subject to discretization errors and the other sources of bias mentioned above. The error is essentially one of aliasing: even if the correct flow state is sampled the discretized observation model can be in error, especially when the observable is a nonlinear function of the state.

In this work, a direct treatment of the discretization bias of both forecast and observation models is proposed aiming to equip the standard EnKF framework to deal with imperfect models. In section 4.4.1, we demonstrate that low-resolution models gives rise to biased estimates. In section 4.4.2, these errors are modeled at low rank by representing the relevant features of the bias

as colored-noise processes. The proposed bias-aware methodology is summarized in section 4.4.3. Section 4.4.4 describes an example of a numerical experiment used to assess the performance.

4.4.1 Discretization error as a source of bias

Let $\tilde{\mathbf{f}}(\mathbf{x})$ be the exact transition function of the Navier-Stokes equations, which maps the infinite-dimensional solution $\tilde{\mathbf{x}}_{k-1}$ at time t_{k-1} to the solution $\tilde{\mathbf{x}}_k$ at time t_k , and let $\tilde{\mathbf{h}}$ represent a (potentially nonlinear) function that maps the exact flow state to a p -dimensional vector of observables, y_k ,

$$\tilde{\mathbf{x}}_k = \tilde{\mathbf{f}}(\tilde{\mathbf{x}}_{k-1}) , \quad (59a)$$

$$\mathbf{y}_k = \tilde{\mathbf{h}}(\tilde{\mathbf{x}}_k) + \boldsymbol{\epsilon}_k^m , \quad (59b)$$

where $\boldsymbol{\epsilon}_k^m \sim N(0, \mathbf{V}_k)$ is a d -dimensional random error vector associated with the measurement methodology that is independent of the state and uncorrelated in time.

Since both the state $\tilde{\mathbf{x}}$ and the operators $\tilde{\mathbf{f}}$ and $\tilde{\mathbf{h}}$ are unattainable for practical purposes, we introduce a finite-dimensional approximation for the model and state. We notate their finite n -dimensional approximations with the same symbols but without the tilde. Following Cohn [12], we define a projection operator Π that maps the true state $\tilde{\mathbf{x}}_k$ onto its finite-dimensional representation $\mathbf{x}_k = \Pi\tilde{\mathbf{x}}_k$. The propagation of $\mathbf{x}_k$ can be represented as

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{w}_k , \quad (60)$$

where

$$\mathbf{w}_k = \Pi\tilde{\mathbf{f}}(\tilde{\mathbf{x}}_{k-1}) - \mathbf{f}(\Pi\tilde{\mathbf{x}}_{k-1}) . \quad (61)$$

The forcing term $\mathbf{w}_k$ represents the *model error*, and gathers errors from different sources: discretization error, inaccurate boundary conditions, uncertain forcing, and so on. Analogously, since the continuous state is never available for practical purposes, a discrete version of the observation operator $\mathbf{h}(\cdot)$ needs to be introduced:

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \boldsymbol{\epsilon}_k^m + \boldsymbol{\epsilon}_k^r , \quad (62)$$

where

$$\begin{aligned} \boldsymbol{\epsilon}_k^r &= \tilde{\mathbf{h}}(\tilde{\mathbf{x}}_k) - \mathbf{h}(\Pi\tilde{\mathbf{x}}_k) \\ &= (\tilde{\mathbf{h}}(\tilde{\mathbf{x}}_k) - \tilde{\mathbf{h}}(\Pi\tilde{\mathbf{x}}_k)) + (\tilde{\mathbf{h}}(\Pi\tilde{\mathbf{x}}_k) - \mathbf{h}(\Pi\tilde{\mathbf{x}}_k)) \end{aligned} \quad (63)$$

is the *error of representativeness* [12], which can be viewed as either a discretization error or a modeling error due to the complexity of the mapping between the control space and the observation space. This error can be further split into two contributions: a first term that represents the effect of the exact operator on the unresolved scales (aliasing), and a second term that represents the discretization error of the operator itself.

Because $\mathbf{w}_k$ depends not only on the state but also on the continuous operator $\tilde{\mathbf{f}}$, its value is unknowable from a practical point of view. Therefore, the most common approach is to represent this error as a stochastic process¹ with known bias and covariance. Still, a complete representation of $\mathbf{w}_k$ requires estimation of all $n(n+1)/2$ degrees of freedom of the associated covariance matrix, which is impractical regardless of the estimation procedure used [26]. In the next section we propose a low-rank model for it, as well as for the associated measurement resolution error, $\boldsymbol{\epsilon}_k^r$.

¹ A stochastic process is a indexed family of random variables that is commonly used to represent the time evolution of a random phenomenon.

4.4.2 Low-rank representation of the bias

We propose a model for the error

$$\mathbf{w}_k = \mathbf{\Gamma}_x \boldsymbol{\xi}_k + \boldsymbol{\mu}_k, \quad (64)$$

where the first term comprises the non-zero-mean bias error that represents available deterministic knowledge about the error, and the second term represents a zero-mean random error. Within the first term, $\mathbf{\Gamma}_x \in \mathbb{R}^{n \times n_s}$ is a low-rank representation of spatial distribution of error, determined a priori from available data, and $\boldsymbol{\xi}_k$ are time-varying, time-correlated coefficients that will be estimated online. We anticipate that the first term will represent the slowly-varying time-correlated dynamics of the error whereas the second term will represent the fast-varying time-uncorrelated dynamics, which we model as white noise ($\boldsymbol{\mu}_k \sim N(0, Q_k)$, i.e., $\boldsymbol{\mu}_k$ is a random vector that distributes as a normal distribution with zero mean and covariance matrix Q_k). In meteorology, Dee [26] and Cohn and Parrish [13] have proposed to represent $\mathbf{w}_k$ as a single random variable $\boldsymbol{\mu}_k$ whose covariance matrix Q_k had parameters that could be tuned on-line. It is interesting to note that these models often rely on using the slow modes of the forecast model to achieve a low-order representation of the random part of the bias. This approach correctly recognizes the slowly-varying behavior and the low-dimensionality of the bias, but fails to represent its time correlation and nonzero mean.

The measurement resolution error can be split in two analogous terms:

$$\boldsymbol{\epsilon}_k^r = \mathbf{\Gamma}_y \boldsymbol{\eta}_k + \boldsymbol{\nu}_k, \quad (65)$$

where $\mathbf{\Gamma}_y \in \mathbb{R}^{d \times n_o}$ is a low-rank modal representation of the bias, and $\boldsymbol{\nu}_k \sim N(0, R^r)$ represents random errors in the measurement function, modeled as a zero-mean Gaussian process. The latter can be merged with the measurement accuracy error $\boldsymbol{\epsilon}_k^m$ in equation (62), which in sum is represented as zero-mean Gaussian process with a combined covariance matrix R .

As will be discussed in section 4.4.4, suitable representations for $\mathbf{\Gamma}_x$ and $\mathbf{\Gamma}_y$ can be obtained using data processing tools such as the proper orthogonal decomposition (POD, Holmes et al. [53]) to obtain the spatial modes that best represent the expected variance of the bias. Note that in this formulation all temporal information is contained in the bias parameters $\boldsymbol{\xi} \in \mathbb{R}^{n_s}$ and $\boldsymbol{\eta} \in \mathbb{R}^{n_o}$.

We expect that $\boldsymbol{\xi}_k$ and $\boldsymbol{\eta}_k$ are auto-correlated (and maybe cross-correlated) in time. As a first option, autoregressive models (AR)² are the simplest way of representing this feature in the discrete-time framework. Noise sequences obeying AR models are often referred to as colored-noise sequences [11, 29]. If N_x and N_y are the orders for the autoregressive models for the state and observation bias, respectively, then

$$\boldsymbol{\xi}_k = \bar{\boldsymbol{\xi}} + \sum_{i=1}^{N_x} \boldsymbol{\Phi}_i^x (\boldsymbol{\xi}_{k-i} - \bar{\boldsymbol{\xi}}) + \boldsymbol{\gamma}_{x,k} \quad (66)$$

$$\boldsymbol{\eta}_k = \bar{\boldsymbol{\eta}} + \sum_{i=1}^{N_y} \boldsymbol{\Phi}_i^y (\boldsymbol{\eta}_{k-i} - \bar{\boldsymbol{\eta}}) + \boldsymbol{\gamma}_{y,k}, \quad (67)$$

where $\boldsymbol{\gamma}_{x,k} \sim N(0, \mathbf{Q}^b)$ and $\boldsymbol{\gamma}_{y,k} \sim N(0, \mathbf{R}^b)$, and $\boldsymbol{\Phi}_i^x$ and $\boldsymbol{\Phi}_i^y$ should be determined based on prior knowledge about the system under study. Both $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are assumed to only depend on time, and the bar denotes time-averaged variables. For example, if the Markov condition holds ($N_x = N_y = 1$) and these error terms are supposed to vary in a much slower time scale than the state dynamics, one can choose to use a persistent model and set $\boldsymbol{\Phi}^x = \boldsymbol{\Phi}^y = \mathbf{I}$.

²AR models differ from moving-average (MA) models by the fact that in the former the weights are applied to the previous states, while in the latter the weights are applied to the input (noise).

A second possibility that can be more suitable for periodic flows is to represent the bias as a sum of harmonics of the system's characteristic frequency. In this framework, the columns of $\mathbf{\Gamma}_x$ and $\mathbf{\Gamma}_y$ and the bias coefficient $\boldsymbol{\xi}_k$ and $\boldsymbol{\eta}_k$ are represented by complex vectors, and the bias dynamics are then given by

$$\mathbf{w}_k = \text{Re}(\mathbf{\Gamma}_x \boldsymbol{\xi}_k) + \boldsymbol{\mu}_k \quad (68)$$

$$\boldsymbol{\epsilon}_k^r = \text{Re}(\mathbf{\Gamma}_y \boldsymbol{\eta}_k) + \boldsymbol{\nu}_k, \quad (69)$$

and

$$\boldsymbol{\xi}_k = \exp(\mathbf{\Lambda}_x \Delta t) \boldsymbol{\xi}_{k-1} + \boldsymbol{\gamma}_{x,k} \quad (70)$$

$$\boldsymbol{\eta}_k = \exp(\mathbf{\Lambda}_y \Delta t) \boldsymbol{\eta}_{k-1} + \boldsymbol{\gamma}_{y,k}, \quad (71)$$

where $\mathbf{\Lambda}_x$ and $\mathbf{\Lambda}_y$ are diagonal matrices whose entries are also complex numbers. The real part of these entries represents the mode's growth/decay rate, which is expected to be zero in a purely periodic flow, and the corresponding imaginary part is related to the mode's oscillatory frequency. Note that the noise terms $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are now complex random sequences with zero-mean Gaussian-distributed magnitude, but uniformly-distributed phase. Data processing tools such as the dynamic mode decomposition (DMD), proposed by Schmid [83], can be used to determine suitable matrices $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$ in this case, as will be discussed in section 4.4.4.

In order to take advantage of an eventual nonzero cross-correlation between the bias parameters $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ and the state x in the filtering process, these variables are gathered in an augmented state vector ($\mathbf{z} = [\mathbf{x}^\top \boldsymbol{\xi}^\top \boldsymbol{\eta}^\top]^\top$) that will be estimated using the algorithm proposed in section 4.4.3.

4.4.3 A Bias-aware EnKF

Bias-awareness can be achieved using the augmentation approach: the parameters corresponding to the low-rank bias models developed in the previous section are appended to the original state vector. The two filtering steps introduced in section 4.1.3 become:

Forecast Step: The state of each ensemble member at the next time step is forecast using the (possibly nonlinear) dynamic model:

$$\begin{aligned} \mathbf{z}_k^{f,i} &= \begin{bmatrix} \mathbf{x}_k^{f,i} \\ \boldsymbol{\xi}_k^{f,i} \\ \boldsymbol{\eta}_k^{f,i} \end{bmatrix} = \mathbf{f}^\dagger(\mathbf{z}_{k-1}^i) + \boldsymbol{\mu}_k^{\dagger i} \\ &= \begin{bmatrix} \mathbf{f}(\mathbf{x}_{k-1}^i) \\ \bar{\boldsymbol{\xi}} + \boldsymbol{\Phi}_x(\boldsymbol{\xi}_{k-1}^i - \bar{\boldsymbol{\xi}}) \\ \bar{\boldsymbol{\eta}} + \boldsymbol{\Phi}_y(\boldsymbol{\eta}_{k-1}^i - \bar{\boldsymbol{\eta}}) \end{bmatrix} + \begin{bmatrix} \boldsymbol{\mu}_k^i \\ \boldsymbol{\gamma}_{x,k}^i \\ \boldsymbol{\gamma}_{y,k}^i \end{bmatrix}. \end{aligned} \quad (72)$$

for $i = 1, \dots, M$. Assuming $n_s \ll n$ and $n_o \ll n$, the additional cost associated with the bias dynamics should be negligible.

The prior statistics are defined as

$$\mathbf{Z}_k = \begin{bmatrix} \mathbf{z}_k^{f,1} & \mathbf{z}_k^{f,2} & \dots & \mathbf{z}_k^{f,M} \end{bmatrix}, \quad (73a)$$

$$\bar{\mathbf{z}}_k = \frac{1}{M} \mathbf{Z}_k \mathbf{1}, \quad (73b)$$

$$\bar{\mathbf{P}}_k^{zz} = \frac{1}{M-1} \sum_{i=1}^M \left(\mathbf{z}_k^{f,i} - \bar{\mathbf{z}}_k \right) \left(\mathbf{z}_k^{f,i} - \bar{\mathbf{z}}_k \right)^\top. \quad (73c)$$

where $\mathbf{1}$ represents a M -dimensional vector with unitary entries. The ensemble covariance matrix can also be expressed in terms of the scaled ensemble perturbation matrix

$$\mathbf{Z}'_k = \frac{1}{\sqrt{M-1}} \mathbf{Z}_k \left(\mathbf{I} - \bar{\mathbf{z}}_k \mathbf{1}^\top \right) , \quad (74a)$$

$$\bar{\mathbf{P}}_k^{zz} = \mathbf{Z}'_k \mathbf{Z}'_k{}^\top . \quad (74b)$$

Analysis Step: The optimization framework is used again to obtain the new analysis equations. For each of the ensemble members, provided that both the prior and the measurement distributions are Gaussian, the mode (maximum-likelihood estimate) of the posterior distribution corresponds to the minimizer of the cost function

$$J(\mathbf{z}) = \frac{1}{2\alpha} \left\| \mathbf{L}^{-1} \mathbf{z} - \mathbf{z}_k^{f,i} \right\|_{\bar{\mathbf{P}}_k^{zz}}^2 + \frac{1}{2} \left\| \mathbf{y}_k - \mathbf{h}(\mathbf{x}) - \mathbf{\Gamma}_y \boldsymbol{\eta} - \boldsymbol{\nu}_k^i \right\|_R^2 \quad (75)$$

while restricting $\left[(\mathbf{x} - \mathbf{\Gamma}_x \boldsymbol{\xi})^\top \quad \boldsymbol{\xi}^\top \quad \boldsymbol{\eta}^\top \right]^\top = \mathbf{L}^{-1} \mathbf{z}$ to the affine subset generated by the prior estimate of the state $\mathbf{z}_k^{f,i}$ and the subspace spanned by the scaled perturbation matrix $\mathbf{Z}'_k$, i.e.,

$$\mathbf{z}_k = \underset{\mathbf{z} \in \mathbf{L}(\hat{\mathbf{z}} + \mathbf{Z}'_k)}{\operatorname{argmin}} J(\mathbf{z}) , \quad (76)$$

where

$$\mathbf{L} = \begin{bmatrix} \mathbf{I} & \mathbf{\Gamma}_x & 0 \\ 0 & \mathbf{I} & 0 \\ 0 & 0 & \mathbf{I} \end{bmatrix} . \quad (77)$$

The linear observation function case When the observation function is linear, i.e. $\mathbf{h}(\mathbf{x}) = \mathbf{H}\mathbf{x}$, the observation equation can be written as

$$\mathbf{y}_k^i = \tilde{\mathbf{H}} \mathbf{L} \mathbf{z}_k^{f,i} + \boldsymbol{\nu}_k^i , \quad (78)$$

where

$$\tilde{\mathbf{H}} = [\mathbf{H} \quad 0 \quad \mathbf{\Gamma}_y] . \quad (79)$$

The restriction on the optimization space is enforced by means of a change of variables

$$\begin{aligned} \mathbf{z} &= \mathbf{L} \left(\hat{\mathbf{z}}_k^i + \sqrt{\beta} \mathbf{Z}'_k \mathbf{v} \right) \\ &= \mathbf{L} \hat{\mathbf{z}}_k^i + \sqrt{\beta} \hat{\mathbf{Z}}'_k \mathbf{v} , \end{aligned} \quad (80)$$

where $\mathbf{v} \in \mathbb{R}^M$ is the correction coefficient vector and $\hat{\mathbf{Z}}'_k = \mathbf{L} \mathbf{Z}'_k$ is the bias-corrected ensemble perturbation matrix.

The analysis step objective is then defined as finding

$$\mathbf{v}_k = \underset{\mathbf{v} \in \mathbb{R}^M}{\operatorname{argmin}} J(\mathbf{v}) , \quad (81)$$

where

$$J(\mathbf{v}) = \frac{1}{2} \|\mathbf{v}\|^2 + \frac{1}{2} \left\| \mathbf{y}_k - \tilde{\mathbf{H}} \mathbf{L} \left(\hat{\mathbf{z}}_k^i + \sqrt{\beta} \mathbf{Z}'_k \mathbf{v} \right) - \boldsymbol{\nu}_k^i \right\|_{\mathbf{V}}^2 . \quad (82)$$

Since $J(\mathbf{v})$ is quadratic in $\mathbf{v}$, the solution is unique and corresponds to the root of

$$DJ(\mathbf{v}) = \mathbf{v} - \sqrt{\beta} \left(\tilde{\mathbf{H}} \hat{\mathbf{Z}}'_k \right)^\top \mathbf{V}^{-1} \left(\mathbf{y}_k - \tilde{\mathbf{H}} \mathbf{L} \hat{\mathbf{z}}_k^i - \sqrt{\beta} \tilde{\mathbf{H}} \hat{\mathbf{Z}}'_k \mathbf{v} - \boldsymbol{\nu}_k^i \right) = 0 , \quad (83)$$

which is given by

$$\mathbf{v}_k^i = \sqrt{\beta} \left[\mathbf{I} + \beta \left(\tilde{\mathbf{H}} \hat{\mathbf{Z}}_k' \right)^\top \mathbf{V}^{-1} \left(\tilde{\mathbf{H}} \hat{\mathbf{Z}}_k' \right) \right]^{-1} \left(\tilde{\mathbf{H}} \hat{\mathbf{Z}}_k' \right)^\top \mathbf{V}^{-1} \left(\mathbf{y}_k - \tilde{\mathbf{H}} \mathbf{L} \hat{\mathbf{z}}_k^i - \boldsymbol{\nu}_k^i \right) \quad (84a)$$

$$= \sqrt{\beta} \left(\tilde{\mathbf{H}} \hat{\mathbf{Z}}_k' \right)^\top \left[\mathbf{V} + \beta \left(\tilde{\mathbf{H}} \hat{\mathbf{Z}}_k' \right) \left(\tilde{\mathbf{H}} \hat{\mathbf{Z}}_k' \right)^\top \right]^{-1} \left(\mathbf{y}_k - \tilde{\mathbf{H}} \mathbf{L} \hat{\mathbf{z}}_k^i - \boldsymbol{\nu}_k^i \right), \quad (84b)$$

where we have used the the matrix inversion lemma [51] to obtain the second line. The analysis step can be performed in the ensemble space (solution of a M -by- M matrix - equation 84a), or in the measurement space (solution of a d -by- d matrix - equation 84b), depending on which is smaller.

The final solution is then obtained by projecting these coefficients back to the state space:

$$\mathbf{z}_k^i = \mathbf{L} \hat{\mathbf{z}}_k^i + \beta \hat{\mathbf{Z}}_k' \left[\mathbf{I} + \beta \left(\tilde{\mathbf{H}} \hat{\mathbf{Z}}_k' \right)^\top \mathbf{V}^{-1} \left(\tilde{\mathbf{H}} \hat{\mathbf{Z}}_k' \right) \right]^{-1} \left(\tilde{\mathbf{H}} \hat{\mathbf{Z}}_k' \right)^\top \mathbf{V}^{-1} \left(\mathbf{y}_k - \tilde{\mathbf{H}} \mathbf{L} \hat{\mathbf{z}}_k^i - \boldsymbol{\nu}_k^i \right) \quad (85a)$$

$$= \mathbf{L} \hat{\mathbf{z}}_k^i + \beta \hat{\mathbf{Z}}_k' \left(\tilde{\mathbf{H}} \hat{\mathbf{Z}}_k' \right)^\top \left[\mathbf{V} + \beta \left(\tilde{\mathbf{H}} \hat{\mathbf{Z}}_k' \right) \left(\tilde{\mathbf{H}} \hat{\mathbf{Z}}_k' \right)^\top \right]^{-1} \left(\mathbf{y}_k - \tilde{\mathbf{H}} \mathbf{L} \hat{\mathbf{z}}_k^i - \boldsymbol{\nu}_k^i \right). \quad (85b)$$

Algorithmically, when the inversion is done in the measurement space, instead of solving for $\mathbf{v}_k$, the representer's formulation [38] is used

$$\left[\mathbf{V} + \beta \left(\tilde{\mathbf{H}} \hat{\mathbf{Z}}_k' \right) \left(\tilde{\mathbf{H}} \hat{\mathbf{Z}}_k' \right)^\top \right] \mathbf{b}_k^i = \mathbf{y}_k - \tilde{\mathbf{H}} \mathbf{L} \hat{\mathbf{z}}_k^i - \boldsymbol{\nu}_k^i \quad (86a)$$

$$\mathbf{z}_k^i = \mathbf{L} \hat{\mathbf{z}}_k^i + \beta \hat{\mathbf{Z}}_k' \left(\tilde{\mathbf{H}} \hat{\mathbf{Z}}_k' \right)^\top \mathbf{b}_k^i, \quad (86b)$$

where the columns of $\beta \hat{\mathbf{Z}}_k' \left(\tilde{\mathbf{H}} \hat{\mathbf{Z}}_k' \right)^\top$ are known as the representer's. They correspond to the influence fields of each of the measurement locations onto the corrected solution, and can be used to provide *a posteriori* information about optimal sensor placement [19].

We note, finally, that the non-linear observation function case is described in [20].

4.4.4 Numerical experiment on flow past an airfoil

In our previous study [19] we applied the (bias-blind) EnKF to flow over a flat plate, an airfoil, and a cylinder at Reynolds number $O(100)$, and examined the accuracy of the estimators as a function of ensemble size, initialization scheme, and covariance inflation parameters. We also considered concurrent estimation of unknown/uncertain parameters such as the Reynolds number [18] and a gusting free-stream velocity [19]. In what follows, we base our numerical experiments largely on the knowledge gained from the earlier studies. We restrict our attention here to flow over a flat plate and an airfoil, both at 30 degrees angle of attack, and for both we take the chord Reynolds number to be 200.

The dynamics of the flow are represented by the discretized 2-D incompressible Navier-Stokes equations, and the simulations were carried out using the Immersed Boundary Projection Method [88, 14] enhanced by the Lattice Green's Function formulation [67, 68]. The latter formulation enforces exactly the free-space boundary condition at infinity even though the computation domain is restricted to the relatively compact region of non-zero vorticity near the immersed body. The spatial discretization error is formally first order, but larger errors tend to be confined in the near-surface region and near second-order convergence is observed in regions away from it. Further details of the numerical method and its validation can be found in the references.

The flow state to be estimated consists of the vorticity at each grid point. The discretized surface forces (traction), which comprise the measurement, are an algebraic, nonlinear function

of the vorticity. With the purpose of analyzing the effects of resolution, three meshes with grid Reynolds numbers ($Re_\Delta = Re\Delta x/c$ where $\Delta x = \Delta y$ is the grid spacing) equal to 1, 2 and 4 are used. The surrogate measurements are always drawn from $Re_\Delta = 1$ simulations, while the grid resolution for the estimator model is varied. For the measurements, we measure the pressure at 10 equidistant locations over the surface every 0.05 convective time units.

In our previous studies, we found that an ensemble of 24 members was sufficient to satisfactorily represent the statistics of this flow in the absence of discretization errors ($Re_\Delta = 1$); the flow features periodic vortex shedding, and the modest ensemble size can be interpreted as representing the subspace on which the energetic dynamics are evolving. As discussed below, the bias correction scheme adds 35 additional parameters that must be estimated online, and so we increase the ensemble size to 60 in what follows in order to accommodate for the additional active degrees of freedom of the augmented dynamical system. The initial ensemble is constructed using the sampling scheme proposed by Evensen [37]. First, a dataset of snapshots of the base solution spanning several vortex-shedding cycles is generated, from which the base mean flow $\bar{\mathbf{x}}_b$ and the leading Proper Orthogonal Decomposition (POD) modes are computed. Then, the M ensemble members are randomly sampled from the subspace spanned by the first M POD modes of the data so that the ensemble average is $\bar{\mathbf{x}}_b$ and the norm of ensemble covariance matrix matches the norm of the dataset covariance matrix. We use the relaxation-to-prior spread (RTPS - $\theta = 0.9$) form of multiplicative covariance inflation Whitaker and Hamill [96], as previous studies [19] showed that it outperforms the constant- β model of Anderson and Anderson [1].

Identification of the resolution error basis The numerical error introduced by the different resolution levels is the source of the bias that we will be interested in tracking. State statistics are estimated from a set of base solutions at the different resolutions spanning a sufficiently long time window. Bias statistics can then be estimated using the definitions presented in section 4.4.1:

$$\Delta = [\Pi \mathbf{x}_2^F - \mathbf{f}(\Pi \mathbf{x}_1^F) \quad \cdots \quad \Pi \mathbf{x}_{n_s}^F - \mathbf{f}(\Pi \mathbf{x}_{n_s-1}^F)] \quad (87a)$$

$$\bar{\mathbf{w}} = \frac{1}{n_s - 1} \Delta \mathbf{1} \quad (87b)$$

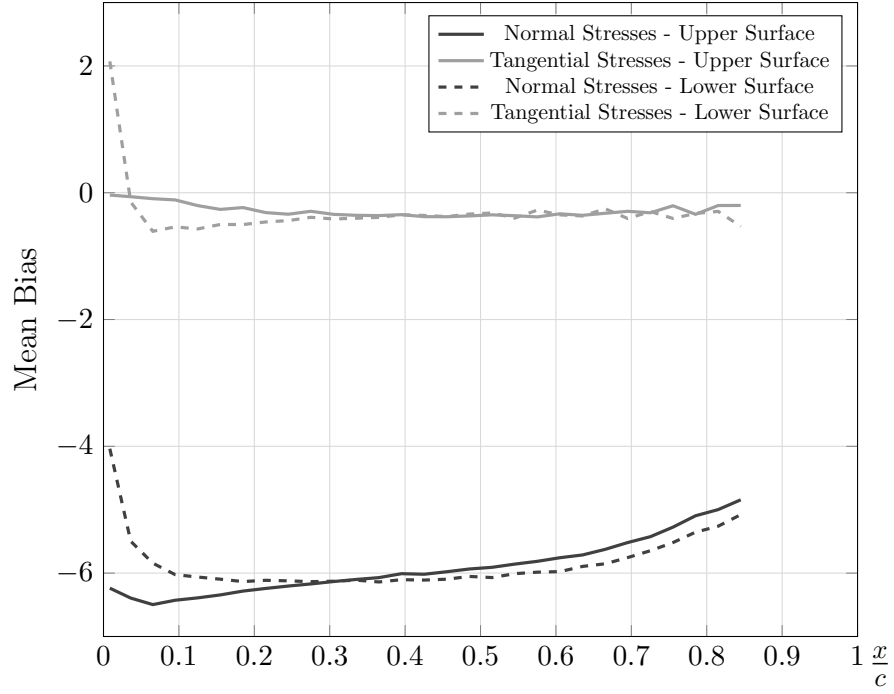
$$\mathbf{E} = [\Pi \mathbf{h}^F(\mathbf{x}_1^F) - \mathbf{h}^C(\Pi \mathbf{x}_1^F) \quad \cdots \quad \Pi \mathbf{h}^F(\mathbf{x}_{n_s}^F) - \mathbf{h}^C(\Pi \mathbf{x}_{n_s-1}^F)] \quad (87c)$$

$$\bar{\epsilon}^r = \frac{1}{n_s - 1} \mathbf{E} \mathbf{1} \quad (87d)$$

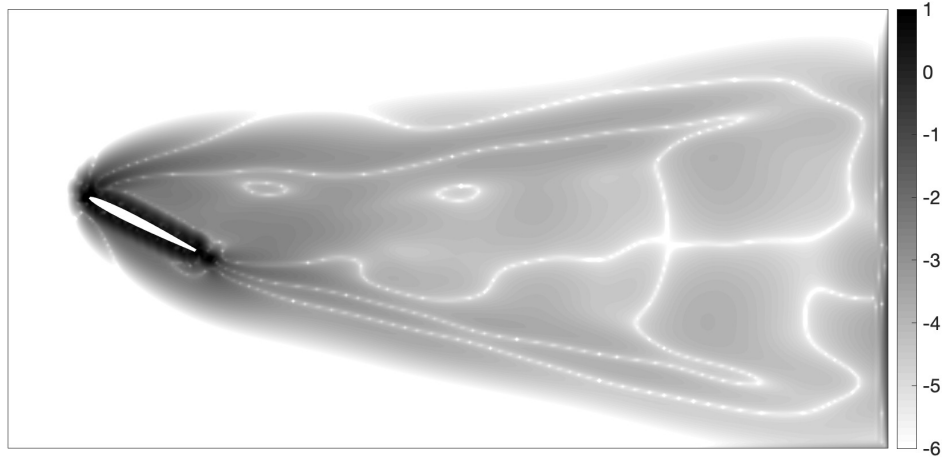
where the superscripts F and C correspond, respectively, to the fine and coarse meshes, and Π is the interpolation operator between the fine and coarse meshes.

Results. Figure 39 shows the temporal mean of the bias fields of both the dynamics ($\mathbf{w}$) and observation model (ϵ^r) between the $Re_\Delta = 4$ and $Re_\Delta = 1$ resolution levels for the airfoil case. The mean bias in the dynamics seems to concentrate near the body where the error introduced by the immersed boundary dominates. Regarding the observation error, there is a pronounced observation bias in the entire surface of the airfoil. The large bias observed in the normal stresses for this case can be explained by the fact that for a closed body the distribution of normal component of the forces acting on it surface is only defined up to a constant (since a constant normal force acting on an immersed body will have zero resultant).

The structure of the corresponding state and observation bias covariance matrices can be analyzed through the POD, i.e. we compute the left singular vectors of $\Delta - \bar{\mathbf{w}}$ and $\mathbf{E} - \bar{\epsilon}^r$, respectively. The low-rank bias representation proposed in section 4.4.2 is justified by the fact that most of the bias variance is restricted to just a few directions in the state space (see analysis in [20]). Retaining the first $n_s = 25$ state POD modes and $n_o = 10$ observation POD modes leaves less than 0.01% of the variance to be modeled as white noise.



(a) Bias field for the observation model ($\bar{\epsilon}^r$).



(b) Bias field for the dynamic model ($\bar{w}$) in log scale.

Figure 39: Temporal average of the bias fields introduced by the resolution error for the flow past an inclined NACA 0009 when comparing the $Re_{\Delta} = 1$ (200 grid points per chord) simulation to the corresponding $Re_{\Delta} = 4$ (50 grid points per chord) simulation.

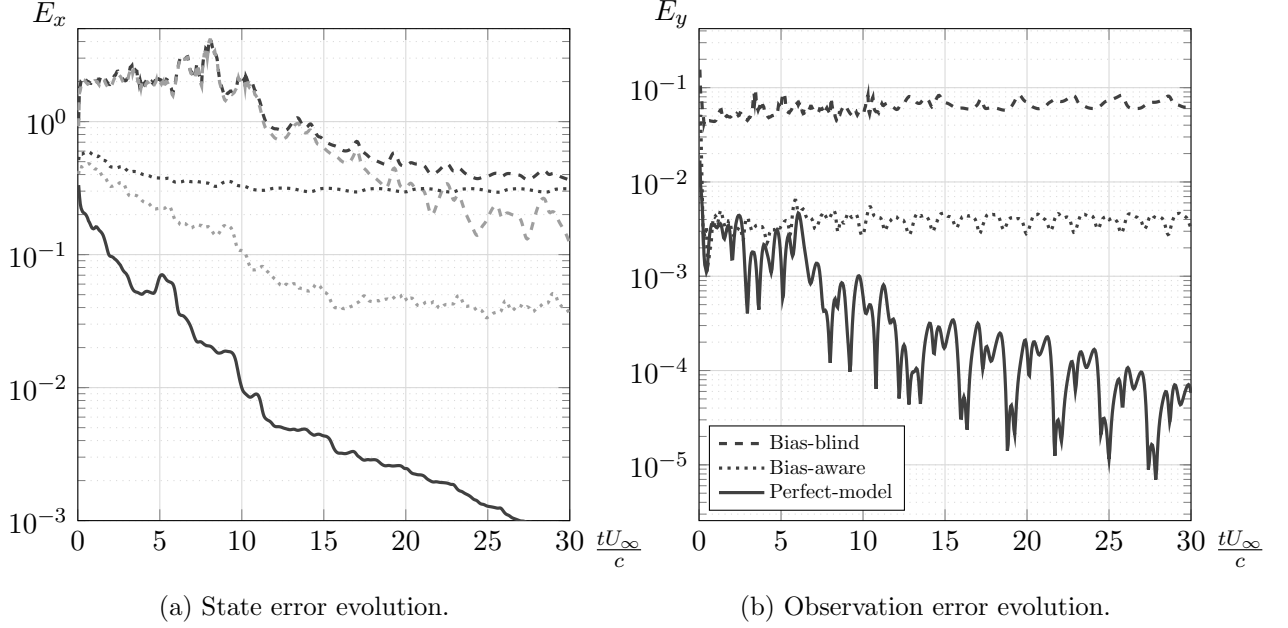


Figure 40: Bias-aware estimator performance ($R = 10^{-4}$, $\lambda_o = 1/10$ and $\lambda_s = 1$) for the inclined NACA 0009 problem. Black lines correspond the standard L2 norm $\|\cdot\|$, while gray lines correspond to the restricted norm $\|\cdot\|_r$.

Therefore, matrices $\mathbf{\Gamma}_x$, $\mathbf{\Gamma}_y$, $\mathbf{\xi}$ and $\mathbf{\eta}$ in equations 66 and 67 can be defined as

$$\mathbf{\Gamma}_x = [\bar{\mathbf{w}} \quad \mathbf{u}_1^s \quad \cdots \quad \mathbf{u}_{n_s}^s] \quad (88a)$$

$$\mathbf{\Gamma}_y = [\bar{\mathbf{\epsilon}}^r \quad \mathbf{u}_1^o \quad \cdots \quad \mathbf{u}_{n_o}^o] \quad (88b)$$

$$\bar{\mathbf{\xi}} = \bar{\mathbf{\eta}} = [1 \quad 0 \quad \cdots \quad 0]^\top, \quad (88c)$$

where $\mathbf{u}_i^s$ and $\mathbf{u}_i^o$ are the i th leading POD modes of $\mathbf{\Delta} - \bar{\mathbf{w}}$ and $\mathbf{E} - \bar{\mathbf{\epsilon}}^r$, respectively, normalized by their respective variances. The process noise $\boldsymbol{\mu}_k$ is sampled from a Gaussian distribution with zero mean and covariance matrix

$$\mathbf{Q}_k = \lambda \bar{\mathbf{P}}_0 + [\mathbf{u}_{n_s+1}^s \quad \mathbf{u}_{n_s+2}^s \quad \cdots] [\mathbf{u}_{n_s+1}^s \quad \mathbf{u}_{n_s+2}^s \quad \cdots]^\top, \quad (89)$$

where λ is a scaling factor. The auto-regressive model parameters are set to $\Phi_1^x = \Phi_1^y = e^{-\Delta t/\tau}$, where Δt is the time interval between two analysis steps, and τ is a reference decorrelation time, here considered to be the vortex shedding period. Alternatively, one could use a least-squares approach to determine the AR coefficients that best fit the data used to construct the low-rank model. The choices for different ARn models were analyzed in [20]. Even though the AR1 coefficients obtained via least-squares differ from the initial guess based on a decorrelation time equal to the vortex shedding period, the differences in performance of the resulting estimator were small. We present the results of applying the bias-aware methodology to the airfoil case. We model the bias dynamics using the POD-AR1 model. Figure 40 shows that the bias-aware estimator greatly improves the accuracy throughout the entire estimation window, achieving a long-term error reduction of 85% for the state and 90% for the measurements.

4.5 Conclusions

This section has presented several advances of the estimation of unsteady aerodynamic flows from surface-based pressure sensors. The data assimilation has been carried out with the ensemble

Kalman filter (EnKF), which relies on a finite-sized ensemble of dynamical models to forecast the state and uncertainty of the system. The EnKF improves the estimate of the state (and lowers its uncertainty) at each step by taking in new sensor data from the true physical process. We have pursued two approaches for the dynamical model: an inviscid vortex model, in which vortex elements are aggregated to keep the degrees of freedom small, and a coarse-grid CFD model. Both of these have advantages of maintaining computational efficiency while ensuring a suitably rich basis for capturing essential flow physics.

In the case of vortex modeling, the unrepresented physical features of the vortex model, e.g., effects of viscosity, are approximately accounted for by the assimilation of pressure measurements obtained from the true physical process [63]. Vorticity flux from the leading edge is obtained by a critical leading-edge suction parameter whose value is also obtained from the measurements. We have interpreted this ensemble forecast physically as an effective translation, stretching, and rotation of the elliptically shaped regions defining each vortex element’s uncertainty region. We have explored two different versions of the EnKF: the stochastic ensemble Kalman filter (sEnKF), which was also used in [23], and the ensemble transform Kalman filter (ETKF). The sEnKF introduces additional numerical noise by not exactly obtaining the correct covariance propagation during the analysis step. In the current vortex estimation procedure, this noise manifests itself as spurious pressures exerted on the surface of the wing. The ETKF, in contrast, reproduces the exact covariance relation even for finite sized ensembles. We have found that this latter filter greatly mitigates the appearance of spurious pressures on the surface, and clusters of vortex elements match better with the true coherent structures. Overall, we have found that the ETKF has better performance and robustness than the sEnKF. In our demonstrations, we have found that the vortex-based estimation framework does not require a model of the disturbance itself. The disturbance’s effect on shed vorticity is accounted for by using measurements to update the state vector in the filter analysis step. This feature is particularly beneficial for a real-time aerodynamic flow estimator, in which the structure of incident flow disturbances such as gusts are unknown.

Generally, ensemble-based filtering introduces some spurious (unphysical) correlations between the state and the observations. One of the challenges of using pressure as a measurement quantity is that it follows from an elliptic relationship with velocity and vorticity, and thus exhibits long-range physical correlations with both. This complicates the usual technique for overcoming the spurious correlations, which involves localizing the state update to just those components that are physically close to the sensor. To overcome that, we have proposed a new regularization method for ensemble Kalman filtering (EnKF) with elliptic observation operators [65]. We have identified a low effective dimension in these cases: low-dimensional projections of the observations strongly inform a low-dimensional subspace of the state space. We derived a low-rank factorization of the Kalman gain based on the spectrum of the Jacobian of the observation operator. The identified eigenvectors generalize the source and target modes of the multipole expansion, independently of the underlying spatial distribution of the problem. Given rapid spectral decay, inference can be performed in the low-dimensional subspace spanned by the dominant eigenvectors. In our test problems (one of which has been shown in this report), we have found that this low-rank EnKF increases the robustness of the estimation and greatly reduces the size of the ensemble.

One drawback of using a low-dimensional vortex model is that its error compared to the true flow state is difficult to assess. In contrast, in the case of prediction with coarse-grid CFD, this error is easier to quantify: discretization errors in the dynamical and observation models lead to biased forecasts and can cause poor estimator performance. We also introduced an approach to use an EnKF framework to simultaneously mitigate the effects of biased forecast and observation models resulting from coarse discretization of the flow [20]. In lieu of treating the bias as a single random variable, we split it into slow and fast components. The fast (incoherent) component was

treated as temporally uncorrelated white noise with a specified covariance, and it can therefore be absorbed in an additive covariance inflation (process noise) scheme. The slow (coherent) part is represented in a low-rank subspace of the measured error between simulations at different grid resolution. The subspace was determined by either POD (or DMD, not shown). The time-correlated modal amplitudes were modeled as an auto-regressive process in the case of POD. The restriction of the bias dynamics to the low-rank subspace that contains most of the variance allows for a more efficient sampling of the state space and enables the use of fewer ensemble members to satisfactorily represent the system statistics. For a small cost increment, the bias-aware estimator reduced the state and observation error by 80% for the airfoil, compared to the bias-blind scheme, and by even larger percentages for the airfoil case where the uncorrected discretization bias was higher. The improvement was similar when the exact error between the truth and the estimator was used to inform the bias statistics, and when the value was inferred by extrapolation from databases with differing intermediate resolutions. Thus the scheme could be used with real measurement data and error data collected *a priori* by running the model at different resolutions.

A Ensemble Kalman filter algorithms

Algorithm 1: Algorithm for the sEnKF, adapted from Asch et al. [2]

Initialize M ensemble member $(\mathbf{x}_0^{a,i}) \sim \mathcal{N}(\bar{\mathbf{x}}_0, \mathbf{P}_0)$
for $k = 1 : K$ **do**
 Compute the ensemble forecast:
 $\mathbf{x}_k^{f,i} = \mathbf{f}(\mathbf{x}_{k-1}^{a,i}) + \mathbf{w}_{k-1}^i$, for $i = 1, \dots, M$
 Sample from the measurement noise distribution $\pi_{\mathbf{v}_k}$:
 $\boldsymbol{\epsilon}_k^i \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_k)$, for $i = 1, \dots, M$
 Compute the sample means:
 $\bar{\mathbf{x}}_k^f = \frac{1}{M} \sum_{i=1}^M \mathbf{x}_k^{f,i}$, $\bar{\boldsymbol{\epsilon}}_k = \frac{1}{M} \sum_{i=1}^M \boldsymbol{\epsilon}_k^i$, $\bar{\mathbf{y}}_k^f = \frac{1}{M} \sum_{i=1}^M \mathbf{h}(\mathbf{x}_k^{f,i})$
 and the anomalies:
 $\mathbf{X}_k'^{f,i} = (\mathbf{x}_k^{f,i} - \bar{\mathbf{x}}_k^f) / \sqrt{M-1}$,
 $\mathbf{Y}_k'^{f,i} = (\mathbf{h}(\mathbf{x}_k^{f,i}) - \boldsymbol{\epsilon}_k^i - \bar{\mathbf{y}}_k^f + \bar{\boldsymbol{\epsilon}}_k) / \sqrt{M-1}$, for $i = 1, \dots, M$
 Invert these equations for $(\mathbf{b}_k^i)$:
 $\mathbf{Y}_k'^f \mathbf{Y}_k'^{f\top} \mathbf{b}_k^i = \mathbf{y}_k^\star + \boldsymbol{\epsilon}_k^i - \mathbf{h}(\mathbf{x}_k^{f,i})$, for $i = 1, \dots, M$
 Update the ensemble:
 $\mathbf{x}_k^{a,i} = \mathbf{x}_k^{f,i} + \mathbf{X}_k'^{f,i} \mathbf{Y}_k'^{f\top} \mathbf{b}_k^i$
end

Algorithm 2: Algorithm to generate mean-preserving random rotations, adapted from [91, 72]

Input: integer M
 Draw $(M-1) \times (M-1)$ samples from an standard normal distribution and store them in $\mathbf{\Omega} \in \mathbb{R}^{(M-1) \times (M-1)}$:
 $\mathbf{\Omega} = \text{randn}(M-1, M-1)$
 SVD decomposition of $\mathbf{\Omega}$:
 $\mathbf{\Omega} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$
 $\mathbf{\Omega} = \mathbf{V}^\top$
 Compute $\mathbf{b} = \mathbf{1}/\sqrt{M}$
 Compute $\mathbf{b}^{sign}$
 $\mathbf{b}^{sign} = \mathbf{b}$
 $\mathbf{b}^{sign}[M] += \text{sign}(\mathbf{b}[M])$ # add $\text{sign}(\mathbf{b}[M])$ to the last entry
 Initialize $\mathbf{B} \in \mathbb{R}^{M \times M}$
 Fill columns of $\mathbf{B}$:
 $\mathbf{B}[:, 1] = \mathbf{b}$
 $\mathbf{B}[:, 2 : M] = -(\mathbf{I}_M + \mathbf{b}^{sign}\mathbf{b}^{sign\top}/(|\mathbf{b}| + 1))[:, 1 : M-1]$
 Initialize $\mathbf{\Lambda}_b \in \mathbb{R}^{M \times M}$
 Fill $\mathbf{\Lambda}_b$
 $\mathbf{\Lambda}_b[1, 1] = 1$
 $\mathbf{\Lambda}_b[1 : M-1, 1 : M-1] = \mathbf{\Omega}$
 return $\mathbf{U} = \mathbf{B}\mathbf{\Lambda}_b\mathbf{B}^\top$

Algorithm 3: Algorithm for the ETKF, adapted from Asch et al. [2]

$\mathbf{U}$ is a mean-preserving rotation generated from time to time, see Algorithm 2, otherwise set to $\mathbf{I}_M$
 Initialize M ensemble member $(\mathbf{x}_0^{a,i}) \sim \mathcal{N}(\mathbf{x}_0, \mathbf{P}_0)$
for $k = 1 : K$ **do**
 Compute the ensemble forecast:
 $\mathbf{x}_k^{f,i} = \mathbf{f}(\mathbf{x}_{k-1}^{a,i}) + \mathbf{w}_{k-1}^i$ for $i = 1, \dots, M$
 Compute the prior mean and anomaly matrix:
 $\bar{\mathbf{x}}_k^f = \frac{1}{M} \sum_{i=1}^M \mathbf{x}_k^{f,i}$, $\mathbf{X}_k'^{f,i} = (\mathbf{x}_k^{f,i} - \bar{\mathbf{x}}_k^f)/\sqrt{M-1}$, for $i = 1, \dots, M$
 Compute the observations:
 $\mathbf{Y}_k^{f,i} = \mathbf{h}(\mathbf{x}_k^{f,i})$, for $i = 1, \dots, M$
 Compute mean of the observations $\bar{\mathbf{y}}_k^f$:
 $\bar{\mathbf{y}}_k^f = \frac{1}{M} \mathbf{Y}_k^f \mathbf{1}$
 Construct the matrices $\mathbf{S}$ and $\mathbf{G}$:
 $\mathbf{S} = \mathbf{V}_k^{-1/2}(\mathbf{Y}_k^f - \bar{\mathbf{y}}_k^f \mathbf{1}^\top)/\sqrt{M-1}$
 $\mathbf{G} = (\mathbf{I}_M + \mathbf{S}\mathbf{S}^\top)^{-1}$
 Construct the vectors $\boldsymbol{\delta}, \mathbf{c}$:
 $\boldsymbol{\delta} = \mathbf{V}_k^{-1/2}(\mathbf{y}_k^* - \bar{\mathbf{y}}_k^f)$
 $\mathbf{c} = \mathbf{G}\mathbf{S}^\top \boldsymbol{\delta}$
 Update the ensemble:
 $\mathbf{X}_k^a = \bar{\mathbf{x}}_k^f \mathbf{1}^\top + \mathbf{X}_k'^{f}(\mathbf{c}\mathbf{1}^\top + \sqrt{M-1}\mathbf{G}^{1/2}\mathbf{U})$
end

B References

- [1] Jeffrey L Anderson and Stephen L Anderson. A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127(12):2741–2758, 1999.
- [2] Mark Asch, Marc Bocquet, and Maëlle Nodet. *Data assimilation: methods, algorithms, and applications*, volume 11. SIAM, 2016.
- [3] Diederik Beckers and Jeff D. Eldredge. Planar potential flow on cartesian grids. arXiv preprint arXiv:2102.11910; accepted for publication in *J. Fluid Mech.*, 2022. URL <https://arxiv.org/abs/2102.11910>.
- [4] Hülya Biler, Girguis Sedky, Anya R Jones, Murat Saritas, and Oksan Cetiner. Experimental investigation of transverse and vortex gust encounters at low reynolds numbers. *AIAA Journal*, 59(3):786–799, 2021.
- [5] James M Birch and Michael H Dickinson. Spanwise flow and the attachment of the leading-edge vortex on insect wings. *Nature*, 412(6848):729–733, 2001.
- [6] Craig H Bishop, Brian J Etherton, and Sharanya J Majumdar. Adaptive sampling with the ensemble transform kalman filter. part i: Theoretical aspects. *Monthly weather review*, 129(3):420–436, 2001.
- [7] C. E. Brown and W. H. Michael. Effect of leading edge separation on the lift of a delta wing. *J. Aero. Sci.*, 21:690–694, 1954.
- [8] Gerrit Burgers, Peter Jan van Leeuwen, and Geir Evensen. Analysis scheme in the ensemble kalman filter. *Monthly weather review*, 126(6):1719–1724, 1998.
- [9] J. Choi, T. Colonius, and D. Williams. Surging and plunging oscillations of an airfoil at low Reynolds number. *J. Fluid Mech.*, 763:237–253, 2015.
- [10] Alexandre Joel Chorin. Numerical study of slightly viscous flow. *Journal of fluid mechanics*, 57(4):785–796, 1973.
- [11] C. K. Chui and G. Chen. *Kalman Filtering With Real Time Applications*. Springer, 2009. doi: 10.1007/978-3-319-47612-4.
- [12] S. E. Cohn. An Introduction to Estimation Theory. *Journal of the Meteorological Society of Japan. Ser. II*, 75(1B):257–288, 1997. doi: 10.2151/jmsj1965.75.1B.257.
- [13] Stephen E. Cohn and David F. Parrish. The behavior of forecast error covariances for a kalman filter in two dimensions. *Monthly Weather Review*, 119(8):1757–1785, 1991. doi: 10.1175/1520-0493(1991)119<1757:TBOFEC>2.0.CO;2.
- [14] Tim Colonius and Kunihiro Taira. A fast immersed boundary method using a nullspace approach and multi-domain far-field boundary conditions. *Computer Methods in Applied Mechanics and Engineering*, 197(25):2131–2146, 2008. doi: <http://dx.doi.org/10.1016/j.cma.2007.08.014>.
- [15] SJ Corkery, Holger Babinsky, and JK Harvey. On the development and early observations from a towing tank-based transverse wing–gust encounter test rig. *Experiments in Fluids*, 59(9):1–16, 2018.

- [16] Georges-Henri Cottet and Petros D. Koumoutsakos. *Vortex methods: theory and practice*. Cambridge university press, 2000.
- [17] Tiangang Cui and Olivier Zahm. Data-free likelihood-informed dimension reduction of bayesian inverse problems. *Inverse Problems*, 37(4):045009, 2021.
- [18] Andre FC da Silva. *An EnKF-based Flow State Estimator for Aerodynamic Problems*. PhD thesis, California Institute of Technology, 2019.
- [19] Andre FC da Silva and Tim Colonius. Ensemble-based state estimator for aerodynamic flows. *AIAA Journal*, 56(7):2568–2578, 2018.
- [20] Andre FC da Silva and Tim Colonius. Flow state estimation in the presence of discretization errors. *Journal of Fluid Mechanics*, 890, 2020.
- [21] Darwin Darakananda. *Vortex Models for Data Assimilation*. PhD thesis, UCLA, 2017.
- [22] Darwin Darakananda and Jeff D Eldredge. A versatile taxonomy of low-dimensional vortex models for unsteady aerodynamics. *Journal of Fluid Mechanics*, 858:917–948, 2019.
- [23] Darwin Darakananda, André Fernando de Castro da Silva, Tim Colonius, and Jeff D Eldredge. Data-assimilated low-order vortex modeling of separated flows. *Physical Review Fluids*, 3(12):124701, 2018.
- [24] Chandler Davis and William Morton Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.
- [25] D. P. Dee and A. M. Da Silva. Data Assimilation in the Presence of Forecast Bias. *Quarterly Journal of the Royal Meteorological Society*, 124(545):269–295, 1998. ISSN 1477-870X. doi: 10.1002/qj.49712454512.
- [26] Dick P. Dee. On-line estimation of error covariance parameters for atmospheric data assimilation. *Monthly Weather Review*, 123(4):1128–1145, 1995. doi: 10.1175/1520-0493(1995)123<1128:OLEOEC>2.0.CO;2.
- [27] Michael H Dickinson and Karl G Gotz. Unsteady aerodynamic performance of model wings at low reynolds numbers. *Journal of experimental biology*, 174(1):45–64, 1993.
- [28] Franklin W. Diederich. The response of an airplane to random atmospheric disturbances. Technical Report 1345, NACA, 1958.
- [29] J.-P. Drecourt, H. Madsen, and D. Rosbjerg. Bias Aware Kalman Filters: Comparison and Improvements. *Advances in Water Resources*, 29(5):707 – 718, 2006. ISSN 0309-1708. doi: 10.1016/j.advwatres.2005.07.006.
- [30] J. Eldredge and C. Wang. High-fidelity simulations and low-order modeling of a rapidly pitching plate. AIAA Paper 2010-4281, 40th AIAA Fluid Dynamics Conference, Chicago, 2010.
- [31] Jeff D. Eldredge. *Mathematical Modeling of Unsteady Inviscid Flows*. Springer International Publishing, 2019.
- [32] Jeff D Eldredge and Anya R Jones. Leading-edge vortices: mechanics and modeling. *Annual Review of Fluid Mechanics*, 51:75–104, 2019.

- [33] Frank Ethridge and Leslie Greengard. A new fast-multipole accelerated poisson solver in two dimensions. *SIAM Journal on Scientific Computing*, 23(3):741–760, 2001.
- [34] Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994.
- [35] Geir Evensen. The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367, 2003. ISSN 16167341. doi: 10.1007/s10236-003-0036-9.
- [36] Geir Evensen. The ensemble kalman filter for combined state and parameter estimation. *IEEE Control Systems Magazine*, 29(3):83–104, 2009.
- [37] Geir Evensen. *Data Assimilation: The Ensemble Kalman Filter*. Springer Science & Business Media, 2009. ISBN 3642037100.
- [38] Geir Evensen and Peter Jan Van Leeuwen. Assimilation of geosat altimeter data for the agulhas current using the ensemble kalman filter with a quasigeostrophic model. *Monthly Weather Review*, 124(1):85–96, 1996.
- [39] *Part 25: Airworthiness Standards: Transport Category Airplanes*. FAA Regulations.
- [40] B. Friedland. Treatment of Bias in Recursive Filtering. *IEEE Transactions on Automatic Control*, 14(4):359–367, 08 1969. ISSN 0018-9286. doi: 10.1109/TAC.1969.1099223.
- [41] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001.
- [42] Kai Fukami, Koji Fukagata, and Kunihiro Taira. Assessment of supervised machine learning methods for fluid flows. *Theoretical and Computational Fluid Dynamics*, pages 1–23, 2020.
- [43] J. M. Greenberg. Airfoil in sinusoidal motion in a pulsating stream. Technical Report TN-1326, NACA, 1947.
- [44] Leslie Greengard. Potential flow in channels. *SIAM journal on scientific and statistical computing*, 11(4):603–620, 1990.
- [45] Leslie Greengard and Vladimir Rokhlin. A fast algorithm for particle simulations. *Journal of computational physics*, 73(2):325–348, 1987.
- [46] Xiaowei He and David R Williams. Unsteady aerodynamic loads on a uas model during a pitch maneuver with roll. AIAA Paper 2020-0822, 2020.
- [47] Xiaowei He and David R Williams. Aerodynamic loads and surface pressure characteristics on a wing in transverse cross-flow gusts. AIAA Paper 2022-0044, 2022.
- [48] Xiaowei He, Katherine J Asztalos, David R Williams, and Kevin Buchert. Tailoring wind tunnel gust spectra with feedback control. AIAA Paper 2020-1557, 2020.
- [49] Xiaowei He, Katherine J Asztalos, James Henry, Scott T Dawson, and David R Williams. Generating traveling cross-flow gusts in a wind tunnel. AIAA Paper 2021-1938, 2021.

- [50] M. Hemati, J. Eldredge, and J. Speyer. Wake sensing for aircraft formation flight. AIAA Guidance, Navigation and Control Conference, Minneapolis, August 2012. AIAA Paper 2012-4768, 2012.
- [51] Harold V Henderson and Shayle R Searle. On deriving the inverse of a sum of matrices. *Siam Review*, 23(1):53–60, 1981. doi: 10.1137/1023004.
- [52] H. Hoel, K. J. H. Law, and R. Tempone. Multilevel Ensemble Kalman Filtering. *SIAM Journal on Numerical Analysis*, 54(3):1813–1839, 2016. doi: 10.1137/15M100955X.
- [53] Philip Holmes, John L. Lumley, Gahl Berkooz, and Clarence W. Rowley. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge Monographs on Mechanics. Cambridge University Press, 2 edition, 2012. doi: 10.1017/CBO9780511919701.
- [54] Wei Hou, Darwin Darakananda, and Jeff D Eldredge. Machine-learning-based detection of aerodynamic disturbances using surface pressure measurements. *AIAA Journal*, 57(12):5079–5093, 2019.
- [55] Wei Hou, Ke Yu, Benedikt Dorschner, and Tim Colonius. An adaptive mesh refinement approach for high reynolds number flows over immersed bodies. In *APS Division of Fluid Dynamics Meeting Abstracts*, pages Q24–006, 2021.
- [56] P. L. Houtekamer and F. Zhang. Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review*, 144(12):4489–4532, 2016. doi: 10.1175/MWR-D-15-0440.1.
- [57] Anya R Jones and Oksan Cetiner. Overview of unsteady aerodynamic response of rigid wings in gust encounters. *AIAA Journal*, 59(2):731–736, 2021.
- [58] Anya R Jones, Oksan Cetiner, and Marilyn J Smith. Physics and modeling of large flow disturbances: Discrete gust encounters for modern air vehicles. *Annual Review of Fluid Mechanics*, 54:469–493, 2022.
- [59] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 1960.
- [60] Wesley Kerstens, Jens Pfeiffer, David Williams, Rudibert King, and Tim Colonius. Closed-loop control of lift for longitudinal gust suppression at low reynolds numbers. *AIAA journal*, 49(8):1721–1728, 2011.
- [61] H. G. Küssner. Zusammenfassender Bericht über den instationären Auftrieb von Flügeln. *Luftfahrtforschung*, 13(12):410–429, 1936.
- [62] Kody Law, Andrew Stuart, and Konstantinos Zygalakis. *Discrete Time: Filtering Algorithms*, pages 79–114. Springer International Publishing, Cham, 2015. ISBN 978-3-319-20325-6. doi: 10.1007/978-3-319-20325-6_4. URL https://doi.org/10.1007/978-3-319-20325-6_4.
- [63] Mathieu Le Provost and Jeff D Eldredge. Ensemble-filtered vortex modeling of strongly disturbed aerodynamic flows. *Phys. Rev. Fluids*, 6(5):050506, 2020. doi: 10.1103/PhysRevFluids.6.050506. URL <https://arxiv.org/abs/2008.11309>. In special issue on "Machine Learning in Fluid Mechanics Invited Papers".

- [64] Mathieu Le Provost, Ricardo Baptista, Youssef Marzouk, and Jeff Eldredge. A low-rank nonlinear ensemble filter for vortex models of aerodynamic flows. AIAA SciTech 2021 Forum. AIAA Paper 2021-1937, 2021. URL https://drive.google.com/open?id=1Vp-AYaks9Yof2nZLQA4Itb3p9c_CTxdG.
- [65] Mathieu Le Provost, Ricardo Baptista, Youssef Marzouk, and Jeff D. Eldredge. A low-rank ensemble Kalman filter for elliptic observations. arXiv preprint arXiv:2203.05120, 2022. URL <https://arxiv.org/abs/2203.05120>.
- [66] A. Leonard. Vortex methods for flow simulation. *J. Comput. Phys.*, 37(3):289–335, 1980.
- [67] Sebastian Liska and Tim Colonius. A parallel fast multipole method for elliptic difference equations. *Journal of Computational Physics*, 278:76–91, 2014. doi: <http://dx.doi.org/10.1016/j.jcp.2014.07.048>.
- [68] Sebastian Liska and Tim Colonius. A fast immersed boundary method for external incompressible viscous flows using lattice green’s functions. *Journal of Computational Physics*, 331:257–279, 2017.
- [69] David M Livings, Sarah L Dance, and Nancy K Nichols. Unbiased ensemble square root filters. *Physica D: Nonlinear Phenomena*, 237(8):1021–1028, 2008.
- [70] Ding-Gwo Long. Convergence of the random vortex method in two dimensions. *Journal of the American Mathematical Society*, 1(4):779–804, 1988.
- [71] Michele Milano and Morteza Gharib. Uncovering the physics of flapping flat plates with artificial evolution. *Journal of Fluid Mechanics*, 534:403–409, 2005.
- [72] Lars Nerger, Tijana Janjić, Jens Schröter, and Wolfgang Hiller. A unification of ensemble square root kalman filters. *Monthly Weather Review*, 140(7):2335–2345, 2012.
- [73] Grigorios A Pavliotis. *Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin equations*. Springer, 2014.
- [74] Gino Perrotta and Anya R Jones. Unsteady forcing on a flat-plate wing in large transverse gusts. *Experiments in Fluids*, 58(8):101, 2017.
- [75] Jens Pfeiffer and Rudibert King. Robust control of drag and lateral dynamic response for road vehicles exposed to cross-wind gusts. *Experiments in Fluids*, 59(3):1–13, 2018.
- [76] Patrick N Raanes. *Improvements to ensemble methods for data assimilation in the geosciences*. PhD thesis, University of Oxford, 2015.
- [77] Kiran Ramesh, Ashok Gopalarathnam, Jack R Edwards, Michael V Ol, and Kenneth Granlund. An unsteady airfoil theory applied to pitching motions validated against experiment and computation. *Theoretical and Computational Fluid Dynamics*, 27(6):843–864, 2013.
- [78] Kiran Ramesh, Ashok Gopalarathnam, Kenneth Granlund, Michael V Ol, and Jack R Edwards. Discrete-vortex method with novel shedding criterion for unsteady aerofoil flows with intermittent leading-edge vortex shedding. *Journal of Fluid Mechanics*, 751:500–538, 2014.

- [79] Kiran Ramesh, Kenneth Granlund, Michael V Ol, Ashok Gopalarathnam, and Jack R Edwards. Leading-edge flow criticality as a governing factor in leading-edge vortex initiation in unsteady airfoil flows. *Theoretical and Computational Fluid Dynamics*, 32(2):109–136, 2018.
- [80] Matthew Rockwood and Albert Medina. Controlled generation of periodic vortical gusts by the rotational oscillation of a circular cylinder and attached plate. *Experiments in fluids*, 61(2):1–13, 2020.
- [81] Pavel Sakov and Peter R Oke. A deterministic formulation of the ensemble kalman filter: an alternative to ensemble square root filters. *Tellus A: Dynamic Meteorology and Oceanography*, 60(2):361–371, 2008.
- [82] Turgut Sarpkaya. An inviscid model of two-dimensional vortex shedding for transient and asymptotically steady separated flow over an inclined plate. *Journal of Fluid Mechanics*, 68(1):109–128, 1975.
- [83] P. J. Schmid. Dynamic Mode Decomposition of Numerical and Experimental Data. *Journal of fluid mechanics*, 656:5–28, 2010.
- [84] W. R. Sears. Some aspects of non-stationary airfoil theory and its practical application. *J. Aeronautical Sci.*, 8(3):104–108, 1941.
- [85] Girguis Sedky, Francis D Lagor, and Anya Jones. Unsteady aerodynamics of lift regulation during a transverse gust encounter. *Physical Review Fluids*, 5(7):074701, 2020.
- [86] Philippe R Spalart. Vortex methods for separated flows. Von Karman Institute for Fluid Dynamics, 1988.
- [87] ArunVishnu SureshBabu, Kiran Ramesh, and Ashok Gopalarathnam. Model reduction in discrete-vortex methods for unsteady airfoil flows. *AIAA Journal*, 57(4):1409–1422, 2019.
- [88] Kuniyiko Taira and Tim Colonius. The immersed boundary method: a projection approach. *Journal of Computational Physics*, 225(2):2118–2137, 2007.
- [89] Kuniyiko Taira and TIM Colonius. Three-dimensional flows around low-aspect-ratio flat-plate wings at low reynolds numbers. *Journal of Fluid Mechanics*, 623:187–207, 2009.
- [90] Kuniyiko Taira, Steven L Brunton, Scott TM Dawson, Clarence W Rowley, Tim Colonius, Beverley J McKeon, Oliver T Schmidt, Stanislav Gordeyev, Vassilios Theofilis, and Lawrence S Ukeiley. Modal analysis of fluid flows: An overview. *Aiaa Journal*, pages 4013–4041, 2017.
- [91] Julian Tödter and Bodo Ahrens. A second-order exact ensemble square root filter for nonlinear data assimilation. *Monthly Weather Review*, 143(4):1347–1367, 2015.
- [92] Yannick Trémolet. Accounting for an imperfect model in 4d-var. *Quarterly Journal of the Royal Meteorological Society*, 132(621):2483–2504, 2006. doi: 10.1256/qj.05.224.
- [93] Joel A Tropp. An introduction to matrix concentration inequalities. *arXiv preprint arXiv:1501.01571*, 2015.
- [94] T. von Kármán and W. R. Sears. Airfoil theory for non-uniform motion. *J. Aeronautical Sci.*, 5(10):379–390, 1938.

- [95] Chengjie Wang and Jeff D Eldredge. Low-order phenomenological modeling of leading-edge vortex formation. *Theoretical and Computational Fluid Dynamics*, 27(5):577–598, 2013.
- [96] Jeffrey S. Whitaker and Thomas M. Hamill. Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, 140(9):3078–3089, 2012. doi: 10.1175/MWR-D-11-00276.1.
- [97] Jeffrey S Whitaker and Thomas M Hamill. Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, 140(9):3078–3089, 2012.
- [98] X. Xia and K. Mohseni. Unsteady aerodynamics and vortex-sheet formation of a two-dimensional airfoil. *J. Fluid Mech.*, 830:439–478, 2017.
- [99] Ke Yu, Benedikt Dorschner, and Tim Colonius. Multi-resolution lattice green’s function method for incompressible flows. *Journal of Computational Physics*, 459:110845, 2022.
- [100] Yi Yu, Tengyao Wang, and Richard J Samworth. A useful variant of the davis–kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015.