



HAL
open science

Quantifying truncation-related uncertainties in unsteady fluid dynamics reduced order models *

Valentin Resseguier, Agustin M Picard, Etienne Mémin, Bertrand Chapron

► To cite this version:

Valentin Resseguier, Agustin M Picard, Etienne Mémin, Bertrand Chapron. Quantifying truncation-related uncertainties in unsteady fluid dynamics reduced order models *. 2021. hal-03169957v1

HAL Id: hal-03169957

<https://hal.science/hal-03169957v1>

Preprint submitted on 15 Mar 2021 (v1), last revised 23 Apr 2021 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Quantifying truncation-related uncertainties in unsteady fluid dynamics reduced order models *

Valentin Resseguier[†], Agustin M. Picard[†], Etienne Mémin[‡], and Bertrand Chapron[§]

Abstract. In this paper, we present a new method to quantify the uncertainty introduced by the aggressive dimensionality reduction commonly practiced in the field of computational fluid dynamics, the ultimate goal being to simulate accurate priors for real-time data assimilation. Our key ingredient is a stochastic Navier-Stokes closure mechanism that arises by assuming random unresolved flow components. This decomposition is carried out through Galerkin projection with a Proper Orthogonal Decomposition (POD-Galerkin) basis. The residual velocity fields, model structure and evolution of coefficients of the reduced order’s solutions are used to compute the resulting multiplicative and additive noise’s correlations. The low computational cost of these consistent correlation estimators makes them applicable to the study of turbulent fluid flows. This stochastic POD-ROM is applied to 2D and 3D DNS wake flows at Reynolds 100 and 300, respectively, with Uncertainty Quantification (UQ) and forecasting outside the learning interval in mind. The proposed stochastic POD-ROM approach is shown to stabilize the unstable temporal coefficients and to maintain their variability under control, while exhibiting an impressively accurate predictive capability.

Key words. Fluid dynamics, reduced order model, uncertainty quantification, stochastic closure, proper orthogonal decomposition

AMS subject classifications. 60H35, 65M60, 65M75, 76M35, 93B11

1. Introduction. The industrial application of partial differential equation (PDE)-driven processes – fluid dynamics, for instance – can be a daunting task, mainly due to the computational complexity associated with its simulations, to be carried out in real-time for some applications such as flow control. To tackle this difficulty, reduced order models (ROM) are common means to speed up deterministic and stochastic design simulations [8, 51, 19], or optimal control problem [32, 17, 40]. A vast community also proposed a plethora of different algorithms to reduce the computational cost of stochastic PDEs for uncertainty quantification (UQ) applications [39, 49, 50, 63, 18, 71].

In turbulent fluid dynamics, the system’s energy usually spreads out over many degrees of freedom. This prevents low-dimensional approximation from being sufficiently accurate, but rough approximations can be sufficient for specific industrial applications, especially when the quantity of interest (QoI) is a spatial average (e.g. lift and drag). However, severe modal truncation often usually end up destabilizing the system and overdamping some of the stable coefficients of the reduced order solution [63]. Consequently, to stabilize the ROM, authors introduce an additional deterministic term (typically an eddy viscosity term) [2, 14, 73], some of them fitting its parameterization, with a possible calibration stage on the available data [11, 20, 74]. This calibration procedure can be extended to the complete set of the ROM’s parameters [56, 66, 1, 9]. ROM performances have mainly been evaluated on low Reynolds-number flows (say $Re \leq 100$) [e.g. 9, 69, 74], while for flows at much greater Reynolds numbers, evaluations close to the learning time interval and reduced dynamics for two-dimensional flow observables (e.g. particle image velocimetry) have been

*Submitted to the editors on July 22, 2020.

Funding: This work was supported by the ERC EU project 856408-STUOD, the ESA DUE GlobCurrent project, the “Laboratoires d’Excellence” CominLabs, Lebesgue and Mer through the SEACS project.

[†]Lab, SCALIAN DS, Rennes, France (valentin.resseguier@scalian.com).

[‡]Fluminance team, Inria, Rennes, France .

[§]LOPS, Ifremer, Plouzané, France.

usually considered [e.g. 11, 1, 20, 69]. Alternatively, longer ROM predictions have been performed with substantial number of modes [e.g. 13, 68]. Yet, turbulent flow ROMs remain inexact and uncontrolled in the long run, owing an intrinsic chaotic nature and the growth of accumulated error along time. Straying from the learning time interval, the determination of accurate parameterizations, in terms of both initial and forcing conditions, becomes more and more difficult.

Ensemble-based data assimilation – such as particle filters or ensemble Kalman filters – can alleviate these issues by forecasting an ensemble of simulations covering the most likely future states of the system, while sequentially constraining them with the on-coming measurements [24, 25]. Still, this necessitates accurate quantification of their associated simulation’s uncertainties. In this context, the aim is not to reduce the dimensionality of UQ, but rather to quantify the uncertainty introduced by the dimensionality reduction during the construction of the ROM.

Note, UQ is a recurrent issue in applied fluid dynamics, and many strategies are generally proposed to incorporating randomness in the physical models through some of their parameters [e.g. 39, 42]. However, the error introduced by these noisy parameters is not *a priori* tied to the reduction of dimensionality and to the unresolved model contribution. In particular, in fluid dynamics, random initial conditions have first been widely used for both UQ and predictability studies [e.g. 46]. It was later demonstrated this yields under-dispersive quantification, i.e. it has a tendency to underestimate the error associated to the dimensionality reduction [7, 48, 31, 28].

Alternatively, authors considered additive noise, most likely beginning with the introduction of EDQNM [52, 41]. Without special cares, one may then rapidly face loss of energy conservation as well as stability issues.

The Modified Quasilinear Gaussian (MQG) method [65, 64, 63] approximates the third-order moment, to help redistribute the right amounts of energy between coefficients of the reduced solution. Finally, several techniques build on averaging and homogenization theory to establish a time-scale separation hypothesis [38, 53, 48, 30], one of the most notable being the MTV model [44]. This latter approach can reproduce intermittency and extreme events alike, thanks to its correlated additive and multiplicative noises. However, besides potential energy conservation issues, the noise covariance is often not explicit enough, and has to be simplified and estimated using the available data. Interested readers can refer to [58], and references herein, for more detailed reviews on model error specification in coarse-scale computational fluid dynamics (CFD). For ROM UQ, [67] proposes distributions and efficient sampling methods for the projection matrices in Galerkin-projection-based dimensionality reduction methods. Although apparently more pertinent than methods based on randomized parameters, the relation of this technique to errors associated to mode truncation and turbulent chaotic behaviour remains unclear.

In this paper, we propose to quantify the uncertainty introduced by modal dimensionality reduction through the so-called dynamics under location uncertainty (LU) [45, 59]. Specifically, we adapt the aforementioned stochastic closure to the Galerkin-projection-based ROM. Inspired from the theoretical works of [10, 47], the LU closure relies on the stochastic transport of the flow variables, together with a decorrelation assumption of the unresolved fluctuations with respect to the resolved slow/large scales. More precisely, the residual velocity – i.e. the difference between the usual Navier-Stokes solution \mathbf{v} and some large-scale velocity component \mathbf{w} – is assumed to be time-decorrelated at the characteristic time of the large-scale processes. This residual velocity is informally denoted $\mathbf{v}' = \sigma \dot{\mathbf{B}} = \sigma d\mathbf{B}_t/dt$ where $t \mapsto \sigma \mathbf{B}_t$ is a \mathbf{Q} -Wiener process [23, 57], and hence, Gaussian in nature. Note this apparent simplified Gaussian assumption leads to, as we will see it,

92 a non-Gaussian multiplicative noise in the dynamics. Spatial correlations of the residual
 93 velocity are specified through the Hilbert-Schmidt integral operator σ with a C^2 kernel
 94 $\check{\sigma}: \bar{\Omega}^2 \rightarrow \mathbb{R}^{d \times d}$, which can be modeled or learned on data:

$$95 \quad (1.1) \quad \sigma(\mathbf{x})d\mathbf{B}_t \triangleq \int_{\Omega} \check{\sigma}(\mathbf{x}, \mathbf{z})d\mathbf{B}_t(\mathbf{z})d\mathbf{z} \quad \forall(\mathbf{x}, t) \in \Omega \times [0, T].$$

96 \mathbf{B}_t is an \mathbb{I}_d -cylindrical Wiener process and plays the role of spatio-temporal white noise.
 97 This definition enables us to characterize the way physical quantities are transported by
 98 the stochastic flow:

$$99 \quad (1.2) \quad d\mathbf{X}_t = \mathbf{w}(\mathbf{X}_t, t)dt + \sigma(\mathbf{X}_t)d\mathbf{B}_t \quad \forall t \in [0, T],$$

100 This resembles the expression for transport in classical fluid dynamics. Material deriva-
 101 tives and other differential operations of fluid dynamics are then derived through the use
 102 of stochastic PDEs (SPDE), in particular, by applying the Itô-Wentzell formula [37] and
 103 a stochastic version of the Reynolds transport theorem [35, 45, 59]. As such, LU models
 104 can successfully apply to model error quantification [16, 60, 62, 58], to improve large-scale
 105 simulations [6, 5, 35, 60, 62], for reduced order modeling and data analysis [61] or for
 106 data assimilation purposes [15, 75] in geophysical fluid dynamics and CFD. To note, in
 107 the geometric mechanics community [33, 21], the Stochastic Advection by Lie Transport
 108 (SALT) method has been derived for large-scale modeling and data assimilation [22]. Both
 109 frameworks have been compared, numerically [6, 58] and conceptually, [5, 62], with LU and
 110 SALT exhibiting different conservation properties, namely energy preservation and circula-
 111 tion conservation, respectively. Applied to the barotropic Quasi-Geostrophy application,
 112 LU leads to improved accuracy when compared to a classical large-scale deterministic
 113 framework or to the circulation conservation stochastic setup [5, 6]. The LU setting also
 114 fully captures the structural deformation of the large-scale flow component by the spatial
 115 inhomogeneity of the small-scales[5]. Very importantly, these properties are independent
 116 of the stochastic integral used.

117 In this paper, our focus is to analyse the LU setting to help define efficient highly
 118 reduced order models for real time data assimilation or control applications. For sake of
 119 simplicity simplicity, we will deal with Proper Orthogonal Decomposition (POD) [43] as
 120 a dimensionality reduction technique – where time and space dependency are separated –
 121 but the proposed methodology applies to all types of modal decompositions.

122 In section 2, we introduce the stochastic fluid dynamics closure we will employ through-
 123 out the paper, as well as its algebraic structure, followed by the Galerkin projection of this
 124 SPDE to derive our stochastic ROM, with a brief recall of the principle behind POD-
 125 ROMs in section 3. Several estimations needed to complete our stochastic POD-ROM are
 126 detailed in 4, as well as the efficient and consistent estimators we exploit to help rely both
 127 on data and on the closure’s physical grounding. In section 5, we discuss the conservative
 128 properties of LU dynamics and of its reduced versions, and finally, section 6 numerically
 129 evaluates the UQ capabilities of our ROM.

130 **2. Navier-Stokes model under location uncertainty.** Galerkin projections of Navier-
 131 Stokes equations do not specifically take into account residual velocity contributions, and
 132 thus a precise quantification of their induced errors. To help remedy this issue, we propose
 133 to directly project SPDEs – i.e. the LU Navier-Stokes representation – instead of the
 134 classical Navier-Stokes equations. The resulting ROM is expected to describe the same
 135 physical system, as we postulate that the solution of the SPDE to be statistically similar
 136 to some large-scale component of the original (deterministic) equation’s solution.

137 **2.1. The random physical model.** Let us denote by Ω an open bounded subset of
 138 \mathbb{R}^d and $T \in \mathbb{R}_+^*$. In the case of incompressible fluids, the LU Navier-Stokes equations on
 139 $\Omega \times [0, T]$ read:

$$140 \quad (2.1) \quad \underbrace{\mathbb{D}_t \mathbf{w}}_{\text{Stochastic transport}} = \underbrace{-\nabla(p \, dt + \mathbf{p}_\sigma d\mathbf{B}_t)}_{\text{Pressure forcing}} + \underbrace{\frac{1}{Re} \Delta(\mathbf{w} dt + \boldsymbol{\sigma} d\mathbf{B}_t)}_{\text{Molecular viscous dissipation}},$$

$$141 \quad (2.2) \quad 0 = \underbrace{\nabla \cdot (\mathbf{w}^* dt + \boldsymbol{\sigma} d\mathbf{B}_t)}_{\text{Mass conservation}},$$

142 where, for every smooth-enough function $\mathbf{q} : \Omega \times [0, T] \rightarrow \mathbb{R}^d$ and for every linear integral
 143 operators ζ, ξ defined on a subspace of L^2 -valued stochastic process $\mathcal{L}^2 \triangleq (\mathcal{L}^2(\Omega))^d$ by
 144 their kernels $\check{\zeta}, \check{\xi} \in (\mathcal{L}^2(\Omega^2))^{d \times d}$, we denote:

$$145 \quad (2.3) \quad (\mathbb{D}_t \mathbf{q})_k \triangleq \underbrace{d_t q_k}_{\substack{\triangleq q_k(\mathbf{x}, t+dt) - q_k(\mathbf{x}, t) \\ \text{Time increment}}} + \underbrace{(\mathbf{w}^* dt + \boldsymbol{\sigma} d\mathbf{B}_t) \cdot \nabla q_k}_{\text{Advection}} - \underbrace{\nabla \cdot (\frac{1}{2} \mathbf{a} \nabla q_k)}_{\text{Turbulent diffusion}} dt,$$

$$146 \quad (2.4) \quad \mathbf{w}^* \triangleq \mathbf{w} - \frac{1}{2} (\nabla \cdot \mathbf{a})^T,$$

$$147 \quad (2.5) \quad \mathbf{a} \triangleq \mathbb{E} \{ (\boldsymbol{\sigma} d\mathbf{B}_t) (\boldsymbol{\sigma} d\mathbf{B}_t)^T \} / dt = \boldsymbol{\sigma} \otimes \boldsymbol{\sigma},$$

$$148 \quad (2.6) \quad (\zeta \otimes \xi)(\mathbf{x}) \triangleq \int_{\Omega} \check{\zeta}(\mathbf{x}, \mathbf{z}) \check{\xi}^T(\mathbf{x}, \mathbf{z}) d\mathbf{z} \quad \forall \mathbf{x} \in \Omega,$$

149 and Re corresponds to the Reynolds number.

150 To fully characterize the above SPDEs, initial and boundary conditions must be defined,
 151 as well as an appropriate functional space for $\mathbf{w} : \Omega \times [0, T] \rightarrow \mathbb{R}^d$. This will allow to
 152 formally derive a stochastic closure mechanism, with a focus on its use on reduced order
 153 models. Theoretical foundations to analyze SPDEs in more general cases are outside of
 154 the scope of this paper. Interested readers can refer to [47, 27]. The stochastic transport
 155 operator \mathbb{D}_t involves the usual terms of a deterministic material derivative, on top of
 156 three additional new terms appear: an advecting velocity correction (\mathbf{w}^* instead of \mathbf{w}),
 157 a heterogeneous and anisotropic turbulent diffusion, and a multiplicative noise. This last
 158 term corresponds to the advection by the unresolved velocity $\boldsymbol{\sigma} \dot{\mathbf{B}}$. Finally, we can recover
 159 the classical Navier-Stokes equations by setting the residual velocity to zero – i.e. $\boldsymbol{\sigma} =$
 160 0. Let us highlight that for conserved scalar this operator corresponds to the material
 161 derivative – i.e. the derivative along the stochastic flow $d(\mathbf{w}(\mathbf{X}_t, t))$ [59]. This stochastic
 162 Navier-Stokes model is generic, and depending on the application, forces and boundary
 163 conditions, it may be modified to other incompressible flow experiments. For compressible
 164 flows, some new terms appear. Interested readers can refer to [59] for the application of
 165 the LU setting to geophysical flows.

166 In addition to the classical physical assumptions pertaining to establishments of the
 167 Navier-Stokes equations for incompressible fluids, the main assumption of LU modeling
 168 setting is to consider the unresolved fluctuation velocity component uncorrelated in time.
 169 To note, the incompressible character of the random fluctuations can be easily relaxed at
 170 the price of additional terms in the transport operator [45, 59]

171 **2.2. Algebraic structure of the model.** The algebraic properties of the different terms
 172 can be quickly described. We can formally rewrite the velocity evolution law (2.1) for
 173 $t \in [0, T]$ as follows:

$$174 \quad (2.7) \quad d_t \mathbf{w} = (d\mathbb{M})(\mathbf{w}) \triangleq ((\mathbf{L} + \mathbf{F})(\mathbf{w}) + \mathbf{C}(\mathbf{w}, \mathbf{w})) dt + (\mathbf{G} d\mathbf{B}_t)(\mathbf{w}) + (\mathbf{H} d\mathbf{B}_t) + d\mathbf{P},$$

175 where \mathbf{L} represents the molecular viscosity term, $\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2$, the turbulent diffusion
 176 plus the advecting velocity correction, and $\mathbf{G}d\mathbf{B}_t$, the advection by the random residual
 177 velocity. All of them are linear differential operators, while \mathbf{C} , the term representing the
 178 usual non-linear advection effect, is a bilinear differential operator. The additive noise
 179 $\mathbf{H}d\mathbf{B}_t$ corresponds to the molecular viscous dissipation of the time-uncorrelated velocity
 180 component, $\boldsymbol{\sigma}\dot{\mathbf{B}}$, while the last term on the right-hand side, $d\mathbf{P}$, is the pressure forcing.

181 Additionally, under suitable boundary conditions, the algebraic structures of the dif-
 182 ferent operators can be further detailed. For instance, \mathbf{L} and \mathbf{F}_1 are symmetric negative
 183 operators. If we consider some additional incompressibility conditions on the effective drift
 184 term $-\frac{1}{2}(\nabla \cdot \mathbf{a})^T$, we also obtain $\nabla \cdot \mathbf{w} = 0$; operator \mathbf{C} is skew-symmetric with respect
 185 to the second argument – *i.e.* $\mathbf{g} \mapsto \mathbf{C}(\mathbf{f}, \mathbf{g})$ is skew-symmetric –, and \mathbf{F}_2 and $\mathbf{G}d\mathbf{B}_t$ are
 186 also skew-symmetric operators. As a matter of fact, if $\langle \boldsymbol{\zeta}, \boldsymbol{\xi} \rangle \triangleq \int_{\Omega} \boldsymbol{\zeta} \cdot \boldsymbol{\xi}$ denotes the scalar
 187 product of L^2 , for every $\boldsymbol{\zeta}$ and $\boldsymbol{\xi} \in \mathcal{H}_0^1(\Omega)$ the space of $H_0^1(\Omega)$ -valued processes (see A
 188 for a precise definition), with $H_0^1(\Omega)$ denoting the Sobolev space $H_0^1(\Omega) = \{f \in L^2(\Omega) : \partial f / \partial x_i \in L^2 \ \forall i = 1, \dots, d; \ f|_{\partial\Omega} = 0\}$, with partial derivative taken in the weak sense,
 189 an integration by parts gives:

$$191 \quad (2.8) \quad \langle \boldsymbol{\zeta}, (\mathbf{G}d\mathbf{B}_t)(\boldsymbol{\xi}) \rangle = \sum_{k=1}^d \int_{\Omega} \zeta_k (\boldsymbol{\sigma}d\mathbf{B}_t \cdot \nabla) \xi_k = \sum_{k=1}^d \int_{\Omega} \zeta_k \nabla \cdot (\boldsymbol{\sigma}d\mathbf{B}_t \xi_k),$$

$$192 \quad (2.9) \quad = - \sum_{k=1}^d \int_{\Omega} (\boldsymbol{\sigma}d\mathbf{B}_t \cdot \nabla \zeta_k) \xi_k = - \langle (\mathbf{G}d\mathbf{B}_t)(\boldsymbol{\zeta}), \boldsymbol{\xi} \rangle.$$

193 Moreover, the turbulent diffusion \mathbf{F}_1 is related to the random skew-symmetric operator
 194 $\mathbf{G}d\mathbf{B}_t$. Indeed, for every $H^2(\Omega)$ -valued processes $\boldsymbol{\xi}$ in $\mathcal{H}^2(\Omega)$,

$$195 \quad (2.10) \quad (\mathbf{F}_1(\boldsymbol{\xi}))_k \triangleq \nabla \cdot (\frac{1}{2} \mathbf{a} \nabla \xi_k) = \nabla \cdot (\frac{1}{2} (\boldsymbol{\sigma} \otimes \boldsymbol{\sigma})^T \nabla \xi_k) = \frac{1}{2} (\boldsymbol{\sigma}^T \nabla) (\otimes (\boldsymbol{\sigma}^T \nabla) \xi_k),$$

$$196 \quad (2.11) \quad = \frac{1}{2} \mathbf{G} (\otimes \mathbf{G}(\xi_k)) = (\frac{1}{2} \mathbf{G} (\otimes \mathbf{G}(\boldsymbol{\xi})))_k = (-\frac{1}{2} \mathbf{G}^* (\otimes \mathbf{G}(\boldsymbol{\xi})))_k,$$

197 where \mathbf{G}^* denotes the adjoint of \mathbf{G} . It also shows that $\mathbf{F}_1(\mathbf{w})dt = \frac{1}{2}d \langle \mathbf{G}(\mathbf{w}), \mathbf{B}_t \rangle$, where
 198 $\langle \boldsymbol{\zeta}, \boldsymbol{\xi} \rangle$ denotes the quadratic covariation of $\boldsymbol{\zeta}$ and $\boldsymbol{\xi}$. The diffusion term explicitly appears
 199 when working with the Itô stochastic integral and is only implicitly taken into account
 200 with Stratonovich integral [5, 62].

201 As discussed in more detail in section 5, these algebraic properties make the LU Navier-
 202 Stokes model – and to a certain extent, its reduced order versions – conservative (up to
 203 molecular viscosity and boundary conditions effects).

204 **3. Galerkin projection.** To sample good priors for future Bayesian estimation algo-
 205 rithms, we aim at deriving a computationally efficient fluid dynamics ROM able to quantify
 206 its own errors with respect to the true fluid dynamics (*i.e.* the Navier-Stokes equations).
 207 As previously mentioned, standard Galerkin techniques – even with the best deterministic
 208 closures – are incapable of such a goal as they were not originally designed for it. Hence,
 209 we propose to perform Galerkin projections on the LU Navier-Stokes model instead and
 210 to study its appropriateness for this sort of tasks.

211 **3.1. A ROM with correlated additive and multiplicative noise.** Let \mathbf{v} be the real
 212 velocity field (the Navier-Stokes equation's solution) and $\boldsymbol{\phi}_0$, a background velocity field,
 213 typically the velocity temporal mean $\bar{\mathbf{v}} \triangleq \frac{1}{T} \int_0^T \mathbf{v}$. To reduce the state space dimension,
 214 we project the fluid velocity anomaly, $\mathbf{v} - \boldsymbol{\phi}_0$, in a subspace spanned by a number of

215 orthonormal spatial modes $(\phi_i)_{1 \leq i \leq n}$.

$$216 \quad (3.1) \quad \mathbf{v}(\mathbf{x}, t, \omega) = \underbrace{\sum_{i=0}^n b_i(t, \omega) \phi_i(\mathbf{x})}_{\triangleq \mathbf{w}^R} + \underbrace{\text{Residual}}_{\triangleq \mathbf{v}'} \quad \forall (\mathbf{x}, t) \in \Omega \times [0, T],$$

217 where $b_0 = 1$ and $\phi_0 = \bar{\mathbf{v}}$ by convention. The associated temporal coefficients, b_i , are
 218 possibly random and depend on a realization ω of a sample space $\tilde{\Omega}$ whereas the reduced
 219 basis functions ϕ_i , are assumed to be deterministic and stationary.

220 As it is typically done while working with ROMs, we aim at specifying the evolution of
 221 the projected velocity field \mathbf{w}^R . A standard technique for that is the Galerkin projection of
 222 the physical PDE – here, the Navier-Stokes equations – onto the reduced basis' functions
 223 ϕ_i , where the resolved component \mathbf{w}^R is approximated by the solution of these projected
 224 equations. As discussed in section 1, mode truncation can create many problems. For
 225 moderately turbulent to turbulent flows and small dimension n , a closure model to handle
 226 the truncated modes is unavoidable. With the LU setting, described in section 2, an elegant
 227 stochastic alternative for this problem can be derived. Accordingly, we will (i) assume that
 228 the residual velocity \mathbf{v}' is time-decorrelated and is denoted $\sigma d\mathbf{B}_t/dt$ and (ii) approximate
 229 the resolved component \mathbf{w}^R to a realization of the solution of the Galerkin projection of the
 230 stochastic Navier-Stokes representation (2.1)-(2.2). The former hypothesis is a debatable
 231 choice with respect to some ROM's applications. Its pertinence and limitation will be
 232 discussed further in section 4.5.

233 To obtain this stochastic ROM, the SPDE (2.7) is first projected onto the divergence-
 234 free function space through the non-local Leray operator $\mathcal{P} = \mathbb{I}_d - \nabla \nabla^T \Delta^{-1}$. This projec-
 235 tion, which requires the resolution of a Poisson equation is used to simplify the system in
 236 removing the pressure term. Since divergence-free solution is considered (see section 3.2),
 237 the resolved velocity component w^R is naturally incompressible and we get

$$238 \quad (3.2) \quad \mathcal{P} d_t \mathbf{w}^R = d_t \mathbf{w}^R, \quad \mathcal{P} \mathbf{L}(\mathbf{w}^R) = \mathbf{L}(\mathbf{w}^R), \quad \mathcal{P} d\mathbf{P} = 0.$$

239 Moreover, $\mathcal{P}(\mathbf{H}d\mathbf{B}_t) = (\mathbf{H}d\mathbf{B}_t)$ because of the necessary incompressibility of the Brow-
 240 nian term in the continuity equation (2.2)¹. Then, the resulting SPDE is projected onto
 241 each one of the reduced basis' functions:

$$242 \quad (3.3) \quad db_i = (\phi_i, d_t \mathbf{w}^R) = (d\mathbb{M}_i^R)(\mathbf{b}) \triangleq (\phi_i, \mathcal{P}(d\mathbb{M})(\mathbf{w}^R)), \quad 1 \leq i \leq n,$$

243 where $\mathbf{b} = (b_i)_{0 \leq i \leq n}$ and

$$244 \quad (d\mathbb{M}_i^R)(\mathbf{b}) = \sum_{p=0}^n \underbrace{(\phi_i, \mathbf{L}(\phi_p))}_{\triangleq l_{pi}} b_p dt + \sum_{p=0}^n \underbrace{(\phi_i, \mathcal{P}\mathbf{F}(\phi_p))}_{\triangleq f_{pi}} b_p dt + \sum_{p,q=0}^n \underbrace{(\phi_i, \mathcal{P}\mathbf{C}(\phi_p, \phi_q))}_{\triangleq c_{pqi}} b_p b_q dt$$

$$245 \quad (3.4) \quad + \underbrace{(\phi_i, (\mathbf{H}d\mathbf{B}_t))}_{\triangleq (\theta_{i\bullet} \bullet d\mathbf{B}_t)} + \sum_{p=0}^n \underbrace{(\phi_i, \mathcal{P}(\mathbf{G}d\mathbf{B}_t)(\phi_p))}_{\triangleq (\alpha_{pi\bullet} \bullet d\mathbf{B}_t)} b_p, \quad 1 \leq i \leq n.$$

246 The terms $(\alpha_{pi\bullet} \bullet d\mathbf{B}_t)_{1 \leq p, i \leq n}$ and $((\theta_{i\bullet} \bullet \alpha_{0i\bullet}) d\mathbf{B}_t)_{1 \leq i \leq n}$ correspond to a Gaussian skew-
 247 symmetric matrix and a Gaussian vector respectively, both with correlated coefficients. If
 248 the functions ϕ_i were spatial Fourier modes associated with small wave-numbers, the ROM
 249 (3.3) would be a (stochastic) LES-like model expressed in Fourier space and \mathbf{b} would be
 250 the set of Fourier coefficients of the solution, but we will be focusing on the POD here.

¹which is coherent with the fact that $\nabla \cdot \mathbf{v}' = 0$ from (3.1) because $\nabla \cdot \mathbf{v} = 0$.

251 **3.2. Proper Orthogonal Decomposition.** In the POD framework, the reduced basis'
 252 functions are computed through a set of velocity snapshots $(\mathbf{v}_{\text{obs}}(\bullet, t_i))_{0 \leq i \leq N-1}$. More
 253 precisely, they are a solution to the constrained optimization problem:

$$254 \quad (3.5) \quad \underset{(\phi_i)_{1 \leq i \leq n}}{\text{Maximize}} \sum_{i=1}^n \overline{(\phi_i, \mathbf{v}_{\text{obs}} - \bar{\mathbf{v}})^2} \quad \text{subject to } (\phi_i, \phi_j) = \delta_{ij}, \quad 1 \leq i, j \leq n.$$

255 Reduced basis functions ϕ_i are thus the n orthonormal functions which can best explain
 256 the snapshots' temporal variability. Similarly to a principal component analysis (PCA),
 257 the solutions of this optimization problem are the eigenfunctions of the velocity anomalies'
 258 $(\mathbf{v}_{\text{obs}} - \bar{\mathbf{v}})$ spatial covariance. Numerically, this matrix is extremely large, and we generally
 259 opt to solve the dual problem instead: we seek the eigenvalue decomposition of the velocity
 260 anomalies' temporal covariance:

$$261 \quad (3.6) \quad C_{ij}^v = ((\mathbf{v}_{\text{obs}} - \bar{\mathbf{v}})(\bullet, t_i), (\mathbf{v}_{\text{obs}} - \bar{\mathbf{v}})(\bullet, t_j)), \quad 0 \leq i, j \leq N-1.$$

262 This method is often referred to as the snapshot method. In the POD framework, the
 263 mode $\phi_0 = \bar{\mathbf{v}}$ is set to the time averaged velocity, and the temporal coefficients energies
 264 $\overline{b_i^2}$ are denoted λ_i . Furthermore, if the snapshots describe a divergence-free velocity field
 265 \mathbf{v}_{obs} , the spatial bases ϕ_i are divergence-free as well.

266 **4. Estimations of subgrid terms.** To close our stochastic ROM (3.3)-(3.4), we need
 267 to estimate the variance tensor \mathbf{a} (involved in the ROM matrix \mathbf{f}) as well as the ROM
 268 noise variances and correlations. Firstly, if we recall that $b_0 = 1$, we note that:

$$269 \quad (4.1) \quad (\alpha_{\bullet i \bullet} d\mathbf{B}_t)^T \mathbf{b} + (\theta_{i \bullet} d\mathbf{B}_t) = \sum_{k=1}^n (\alpha_{ki \bullet} d\mathbf{B}_t) b_k + ((\theta_{i \bullet} + \alpha_{0i \bullet}) d\mathbf{B}_t), \quad 1 \leq i \leq n.$$

270 The multiplicative and additive noises of the ROM correspond to the first and second term
 271 of the right-hand side respectively. To simplify notations, we write:

$$272 \quad (4.2) \quad \tilde{\alpha}_{ji \bullet} = \alpha_{ji \bullet} + \delta_{j0} \theta_{i \bullet}, \quad 1 \leq i, j \leq n.$$

273 To fully specify the ROM, the following correlations must be estimated:

$$274 \quad (4.3) \quad \begin{cases} \mathbf{a}(\mathbf{x}) &= \mathbb{E} \{ (\boldsymbol{\sigma}(\mathbf{x}) d\mathbf{B}_t) (\boldsymbol{\sigma}(\mathbf{x}) d\mathbf{B}_t)^T \} / dt \quad \forall \mathbf{x} \in \Omega, \\ \Sigma_{pi, qj}^{\alpha} &= \mathbb{E} \{ (\tilde{\alpha}_{pi \bullet} d\mathbf{B}_t) (\tilde{\alpha}_{qj \bullet} d\mathbf{B}_t) \} / dt, \quad 1 \leq i, j \leq n, \quad 0 \leq p, q \leq n. \end{cases}$$

275 **4.1. The curse of dimensionality.** Computing correlations of Gaussian noises $(\tilde{\alpha} d\mathbf{B}_t)$
 276 involves the two-point quadratic cross-variation tensor of the small-scale velocity:

$$277 \quad (4.4) \quad \mathbf{Q}(\mathbf{x}, \mathbf{y}) = \mathbb{E} \{ (\boldsymbol{\sigma}(\mathbf{x}) d\mathbf{B}_t) (\boldsymbol{\sigma}(\mathbf{y}) d\mathbf{B}_t)^T \} / dt \quad \forall \mathbf{x}, \mathbf{y} \in \Omega.$$

278 The coefficients (i, j) of the covariance matrix of the Gaussian vector $(\boldsymbol{\theta} d\mathbf{B}_t)$ can be ex-
 279 pressed as follows:

$$280 \quad (4.5) \quad \mathbb{E} \{ (\theta_{i \bullet} d\mathbf{B}_t) (\theta_{j \bullet} d\mathbf{B}_t) \} = \mathbb{E} \{ (\phi_i, \nu \nabla^2 \boldsymbol{\sigma} d\mathbf{B}_t) (\nu \nabla^2 \boldsymbol{\sigma} d\mathbf{B}_t, \phi_j) \},$$

$$281 \quad (4.6) \quad = \iint_{\Omega^2} \phi_i(\mathbf{x})^T \nu^2 \nabla_x^2 \nabla_y^2 \mathbf{Q}(\mathbf{x}, \mathbf{y}) dt \phi_j(\mathbf{y}) d\mathbf{x} d\mathbf{y}.$$

282 Since by definition, $\forall \mathbf{x} \in \Omega$, $\mathbf{a}(\mathbf{x}) = \mathbf{Q}(\mathbf{x}, \mathbf{x})$, the whole stochastic ROM is closed by
 283 specifying the tensor \mathbf{Q} . In practice, this tensor is often heterogeneous – i.e. $\boldsymbol{\sigma} \mathbf{B}_t$ is
 284 non-stationary in space. Thus, its spatially-discretized version is expected to be so large

285 that its estimation or even its storage becomes prohibitive. To overcome this difficulty,
 286 a first strategy is to assume a model structure for the covariance as per [36, 60, 62, 58].
 287 If the small-scale velocity is observed, other techniques can be considered. Indeed, one
 288 can build a POD representation of the small-scale velocity $\sigma\dot{\mathbf{B}}$, learned from available
 289 snapshots or realizations as in [6, 21, 62, 58]. Depending on the desired accuracy for the
 290 noise representation, a great number of modes would have to be estimated and the number
 291 of coefficients involved in the ROM can quickly grow out of control. Here, we rely on
 292 a method specifically devised for these kind of ROM frameworks: the noise structure is
 293 again learned from observed residual velocity snapshots, but without passing through the
 294 covariance \mathbf{Q} . It enables us to directly estimate the correlations of the random ROM's
 295 coefficients, as it will be explained later on.

296 **4.2. Variance tensor estimation.** Considering the variance tensor \mathbf{a} stationary leads
 297 to a simple estimator through time averaging as in [61]:

$$298 \quad (4.7) \quad \hat{\mathbf{a}} = \Delta t \overline{\mathbf{v}'_{\text{obs}} (\mathbf{v}'_{\text{obs}})^T},$$

299 where

$$300 \quad (4.8) \quad \mathbf{v}'_{\text{obs}} = \mathbf{v}_{\text{obs}} - \mathbf{w}_{\text{obs}}^R = \mathbf{v}_{\text{obs}} - \sum_{i=0}^n b_i^{\text{obs}} \phi_i,$$

301 is the observed POD's residual velocity. Since $\mathbf{a}dt = d\langle \sigma\mathbf{B}, (\sigma\mathbf{B})^T \rangle$, the estimator (4.7)
 302 is consistent thanks to the quadratic covariation definition.

303 **4.3. Noise correlation estimation.** Besides the variance tensor, there are $n^2(n+1)^2$
 304 correlations to estimate according to (4.3). For any function ξ in $H^2(\Omega)$, let us introduce
 305 the linear functional:

$$306 \quad (4.9) \quad K_{jq}[\xi] \triangleq (\phi_j, -\mathcal{P}[(\xi \cdot \nabla) \phi_q] + \delta_{q0} \nu \Delta \xi), \quad 1 \leq j \leq n, \quad 0 \leq q \leq n.$$

307 Using this notation, the noise's covariance can be estimated as follows:

$$308 \quad (4.10) \quad \widehat{\Sigma_{pi,qj}^{\alpha}} = \frac{\Delta t}{\lambda_p^{\text{obs}}} K_{jq} \left[\overline{b_p^{\text{obs}} \left(\frac{\Delta b_i^{\text{obs}}}{\Delta t} \right)'' \mathbf{v}'_{\text{obs}}} \right], \quad 1 \leq i, j \leq n, \quad 0 \leq p, q \leq n,$$

309 where $b_0^{\text{obs}} = \lambda_0^{\text{obs}} = 1$ and for $1 \leq i \leq n$,

$$310 \quad (4.11) \quad b_i^{\text{obs}} = (\phi_i, \mathbf{v}_{\text{obs}}),$$

$$311 \quad (4.12) \quad \lambda_i^{\text{obs}} = \overline{(b_i^{\text{obs}})^2},$$

$$312 \quad (4.13) \quad \left(\frac{\Delta b_i^{\text{obs}}}{\Delta t} \right)'' = \left(\frac{\Delta b_i^{\text{obs}}}{\Delta t} \right)' - \overline{\left(\frac{\Delta b_i^{\text{obs}}}{\Delta t} \right)'},$$

$$313 \quad (4.14) \quad \left(\frac{\Delta b_i^{\text{obs}}}{\Delta t} \right)' = \left(\frac{\Delta b_i^{\text{obs}}}{\Delta t} \right) - \left((\mathbf{b}^{\text{obs}})^T (\mathbf{l} + \check{\mathbf{f}})_{\bullet i} + (\mathbf{b}^{\text{obs}})^T \mathbf{c}_{\bullet \bullet i} \mathbf{b}^{\text{obs}} \right),$$

$$314 \quad (4.15) \quad \left(\frac{\Delta b_i^{\text{obs}}}{\Delta t} \right) (t_k) = \frac{b_i^{\text{obs}}(t_k + \Delta t) - b_i^{\text{obs}}(t_k)}{\Delta t}, \quad 0 \leq k \leq N-1.$$

315 To ensure the noise's covariance matrix to have the desired symmetric non-negative struc-
 316 ture, we only keep the symmetric part of the estimated tensor (4.10) and set its possible
 317 negative eigenvalues to zero. The part inside the functional K_{jq} in the estimator (4.10)
 318 is inspired from [29], where products of martingale time increments are projected onto
 319 orthogonal functions of $\mathcal{L}^2([0, T])$. For ROMs with a small number of dimensions n , the

estimator's computational cost is remarkably low as the observed coefficients of the reduced solution \mathbf{b}^{obs} were already computed by the method of snapshots (see section 3.2). Hence, the computational cost for $b_p^{\text{obs}} \left(\frac{\Delta b_i^{\text{obs}}}{\Delta t} \right)''$ is negligible considering that the part inside of K_{jq} only involves a projection of the residual velocity along the time dimension. Finally, the $n(n+1)$ differential operators K_{jq} are computed on only $n(n+1)$ functions that do not depend on the time variable.

We prove the consistency of this estimator in Appendix A by capitalising on the quadratic covariation's definition and the orthogonality of the observed temporal coefficients b_i^{obs} . In addition, this can also be extended to non-orthogonal coefficients by solving the linear system engendered by the matrix $(\bar{b}_p \bar{b}_k)_{pk}$.

4.4. Noise dimension reduction. Using the Cholesky decomposition $\boldsymbol{\sigma}^\alpha$ of the noise covariance tensor:

$$(4.16) \quad \Sigma_{pi,qj}^\alpha = \sum_{lk} \sigma_{pi,lk}^\alpha \sigma_{qj,lk}^\alpha, \quad 1 \leq i, j \leq n, \quad 0 \leq p, q \leq n,$$

and $O(n^2)$ independent white noises, one can sample realizations of the ROM Gaussian noise terms $\tilde{\boldsymbol{\alpha}} d\mathbf{B}_t$. However, $\boldsymbol{\sigma}^\alpha$ has $O(n^4)$ coefficients, while the ROM works with only n modes and involves – leaving the noise terms aside – $O(n^3)$ coefficients. Therefore, we propose to reduce the noise dimension through a tensorial PCA of $\boldsymbol{\Sigma}^\alpha$, eventually only keeping the n first eigenvectors. This leads to the following sampling strategy:

$$(4.17) \quad \tilde{\boldsymbol{\alpha}} d\mathbf{B}_t \approx \sum_{k=1}^n \tilde{\boldsymbol{\alpha}}_k^R d\beta_t^{(k)},$$

where $(\tilde{\boldsymbol{\alpha}}_k^R)_k \in \mathbb{R}^{(n+1) \times n}$ are the matrix forms of the first n eigenvectors (weighted by the corresponding eigenvalues' square roots) and $(\beta^{(k)})_k$ are n independent one-dimensional Brownian motions. Since $\tilde{\boldsymbol{\alpha}} d\mathbf{B}_t$ is a multiplicative noise and the temporal coefficients b_i have various amplitudes $\sqrt{\lambda_i}$, the covariance matrix $\boldsymbol{\Sigma}^\alpha$ is adequately re-normalized by the amplitudes $\sqrt{\lambda_i}$ before applying the PCA.

It is important to note that the methodology described here is different from a more usual methodology based on PCA decomposition of the residual velocity \mathbf{v}' keeping n modes and assuming that the corresponding temporal coefficients $(b_i)_{n+1 \leq i \leq 2n}$ are time-decorrelated. The complexity of the final ROM is the same in both methodologies but our ROM maximizes the noise's variance instead of the residual velocity's variance. Thus, our method is better in terms of ROM UQ.

4.5. Time down-sampling rate. Under the LU Navier-Stokes model hypothesis, the unresolved term of the velocity field must be time-decorrelated noise. This assumption is consistent with the fact that the higher-order coefficients of the reduced order solution often tend to have shorter correlation time in fluid dynamics systems. However, in practice, this assumption is not found to be true, and it is a recurrent issue for data-driven modeling of systems combining fast and slowly evolving components [3, 4, 55, 54]. Consequently, a time down-sampling scheme is proposed to force the noise terms to be as decorrelated as possible. An estimation of the down-sampling rate is thus proposed.

In particular, by assuming that the spatially averaged covariance function has a Gaussian form with a standard deviation equal to the correlation time τ , a simple expression allows us to compute it. For a given unresolved velocity correlation matrix we write:

$$(4.18) \quad C_{ij}^{v'} = (\mathbf{v}'_{\text{obs}}(\bullet, t_i), \mathbf{v}'_{\text{obs}}(\bullet, t_j)) = C_{ij}^v - \sum_{k=1}^n b_k^{\text{obs}}(t_i) b_k^{\text{obs}}(t_j), \quad 0 \leq i, j \leq N-1,$$

362 and its associated stationary covariance function

$$363 \quad (4.19) \quad \text{Cov}_s(t_p) = \frac{1}{N-p} \sum_{q=0}^{N-1-p} C_{q,q+p}^{w'}, \quad 0 \leq p \leq N-1,$$

364 We propose the following correlation time estimation:

$$365 \quad (4.20) \quad \hat{\tau} = \sqrt{2 \frac{\overline{\text{Cov}_s^2}}{(\frac{\Delta \text{Cov}_s}{\Delta t})^2}},$$

366 using a forward Euler temporal discretization of the stationary covariance:

$$367 \quad (4.21) \quad \frac{\Delta \text{Cov}_s}{\Delta t}(t_p) \triangleq \frac{\text{Cov}_s(t_p + \Delta t) - \text{Cov}_s(t_p)}{\Delta t}, \quad 0 \leq p \leq N-1.$$

368 These estimations follow over-simplified assumptions known to have a restricted validity.
 369 Experimentally, they still systematically provided the best simulation results when com-
 370 pared to other more complex estimators (whose derivation is outside the scope of this
 371 paper). Moreover, with a white unresolved velocity with a dirac stationary covariance
 372 function, we obtain $\hat{\tau} = dt$. This result is exactly expected from a white unresolved velocity,
 373 to prevent an overly aggressive down-sampling.

374 Before computing the estimations presented in sections 4.2, 4.3 and 4.4, we thus use
 375 this estimated correlation time $\hat{\tau}$ to down-sample both the entire dataset and the observed
 376 coefficients of the reduced order solution, leaving us with a time step $\Delta t \approx \hat{\tau}$. On top of
 377 the considerable gain in accuracy, this time down-sampling reduces the amount of data to
 378 process, and hence, makes the offline ROM building process faster.

379

380 Having estimated and specified all of the stochastic ROM's parameters in equations
 381 (3.3) and (3.4), it is now possible to forecast ensembles of realizations of the ROM through
 382 Monte-Carlo simulations. But before presenting numerical results, we discuss some impor-
 383 tant properties of LU models in their reduced order versions.

384 **5. Kinetic energy budget.** First, the conservative properties of the LU Navier-Stokes
 385 representation are recalled, followed by a proof that by combining Galerkin projections and
 386 the advecting velocity correction compressibility, an intrinsic energy dissipation appears.

387 **5.1. Full-order model budget.** As derived in [59, 62], since the pressure does not
 388 influence the energy budget, by neglecting the molecular viscosity, the divergence of \mathbf{w} and
 389 boundary conditions effects, and by applying Itô's lemma, the expression for the kinetic
 390 energy budget writes:

$$391 \quad (5.1) \quad \frac{d}{dt} \left(\frac{1}{2} \|\mathbf{w}\|_{\mathcal{L}^2}^2 \right) = \underbrace{\left(-\frac{1}{2} \mathbf{G}^* (\otimes \mathbf{G}(\mathbf{w})), \mathbf{w} \right)}_{\text{Loss by diffusion}} + \underbrace{\frac{1}{2} \|\mathbf{G}(\mathbf{w})\|_{\text{HS}}^2}_{\text{Energy flux from the noise}} = 0 \quad \forall t \in [0, T],$$

392 where $\|\mathbf{f}\|_{\mathcal{L}^2}^2 = (\mathbf{f}, \mathbf{f})$ is the squared norm of L^2 and

$$393 \quad (5.2) \quad \|\zeta\|_{L^2}^2 = \iint_{\Omega^2} \sum_{i,j=1}^d \zeta_{i,j}^2(\mathbf{x}, \mathbf{z}) d\mathbf{x} d\mathbf{z}$$

394 is the squared Hilbert-Schmidt norm of the integral operator ζ . This enables us to state
 395 that the energy is conserved for each realization of the stochastic process, and as a direct

consequence of this, the energy intake of the noise and the dissipation by the turbulent diffusion must exactly compensate each other. The latter dissipates the kinetic energy of the mean $\|\mathbb{E}\{\mathbf{w}\}\|_{L^2}^2$, while the former only releases random energy $\|\mathbf{w} - \mathbb{E}\{\mathbf{w}\}\|_{L^2}^2$ into the system. Thus, the time-uncorrelated component of the velocity field drains energy from the mean field to the random component of \mathbf{w} .

We can also express this energy transfer with the expectation of equation (5.1):

$$(5.3) \quad \frac{d}{dt} \int_{\Omega} \text{Var}(\mathbf{w}) = \frac{d}{dt} \mathbb{E} \|\mathbf{w} - \mathbb{E}\{\mathbf{w}\}\|_{L^2}^2 = -\frac{d}{dt} \|\mathbb{E}\{\mathbf{w}\}\|_{L^2}^2 \quad \forall t \in [0, T].$$

Besides the physical relevance of energy conservation, variance inflation and its relation to the mean field are also of primary interest for data assimilation or ensemble forecasting issues.

5.2. Reduced-order model budget. Following (5.1), the full-order LU Navier-Stokes model (2.1)-(2.2) conserves the kinetic energy, up to molecular viscosity and boundary condition effects. With reduced order models, the advecting velocity correction is expected however to create either energy compression or dilation, and the mode truncation, to introduce a small energy leak. The ROM (3.3) does not exactly solve the global LU Navier-Stokes model (2.1)-(2.2) but the Galerkin projection of the divergence-free component of equation (2.1) :

$$(5.4) \quad d_t \mathbf{w}^R = \sum_{i=0}^n db_i \phi_i = \sum_{i=1}^n (d\mathbb{M}_i^R)(\mathbf{b}) \phi_i = \Pi_{\phi} [\mathcal{P}(d\mathbb{M})(\mathbf{w}^R)] \quad \forall t \in [0, T],$$

where Π_{ϕ} is the projection onto the reduced subspace. Specifically, for any function $\zeta \in \mathcal{L}^2$, the projection Π_{ϕ} is defined as $\Pi_{\phi}[\zeta] \triangleq \sum_{p=1}^n (\phi_p, \zeta) \phi_p$. As previously stated, we can evaluate the variation of kinetic energy with the Itô formula. The following result is proved in Appendix B for every $t \in [0, T]$:

$$(5.5) \quad \frac{d}{dt} \left(\frac{1}{2} \|\mathbf{w}^R\|_{\mathcal{L}^2}^2 \right) = \underbrace{-\frac{1}{2} \|\mathbf{G}(\mathbf{w}^R)\|_{\text{HS}}^2}_{\text{Loss by diffusion}} + \underbrace{\frac{1}{2} \|\Pi_{\phi} [(\mathbf{G})(\mathbf{w}^R)]\|_{\text{HS}}^2}_{\text{Energy flux from the noise}} - \underbrace{\frac{1}{2} (\nabla \cdot (\nabla \cdot \mathbf{a})^T, \|\mathbf{w}^R\|^2)}_{\text{Advecting velocity correction compressibility}},$$

$$(5.6) \quad = -\frac{1}{2} (\nabla \cdot (\nabla \cdot \mathbf{a})^T, \|\mathbf{w}^R\|^2) - \underbrace{\frac{1}{2} \|\Pi_{\phi}^{\perp} [(\mathbf{G})(\mathbf{w}^R)]\|_{\text{HS}}^2}_{<0},$$

where $\|\bullet\|$ stands for the Euclidean norm on \mathbb{R}^d and $\Pi_{\phi}^{\perp} = \mathbb{I}_d - \Pi_{\phi}$ is the projector onto the orthogonal complement of the space spanned by the reduced basis' functions.

The first term of (5.6) corresponds to the fluid compression/dilatation created by the velocity correction $\mathbf{w}^* - \mathbf{w} = -\frac{1}{2} (\nabla \cdot \mathbf{a})^T$. By construction, the ROM meets the incompressible condition – i.e. $\nabla \cdot \mathbf{w} = 0$ – but unlike the LU Navier-Stokes model, it does not satisfy the finite-variation part of the mass conservation (2.2): $\nabla \cdot \mathbf{w}^* = 0$. The velocity correction divergence is associated to turbulence heterogeneity [5, 15, 61], corresponding to spatial maxima (positive energy fluxes) and minima (negative energy fluxes) of the turbulence's kinetic energy $\|\mathbf{v}'\|_2^2$.

The second term of (5.6) informs about the energy flux due to mode truncation. The ROM's subgrid diffusion extracts energy from some temporal coefficients of the reduced order solution, while the multiplicative noise distributes it to others. These stabilizing and destabilizing effects recreate a large part of the energy fluxes between coefficients of the reduced solution, otherwise lost in deterministic ROMs. Thus, the decorrelated velocity

434 component drains energy from the coefficients of the reduced solution to give it back to
 435 the temporal coefficients of the truncated modes – i.e. modes orthogonal to the reduced
 436 space. This energy flux is exactly characterized by the second term of (5.6). But, our
 437 stochastic ROM cannot transfer energy from the truncated modes to the reduced solution’s
 438 coefficients, as the former cannot be specified due to the truncation.

439 Such a dissipation could be perfectly prevented by considering $-\frac{1}{2}(\Pi_\phi[\mathbf{G}])^*(\otimes\Pi_\phi[\mathbf{G}])$
 440 represented in the ROM by $\left(-\frac{1}{2}\sum_{i=1}^n\Sigma_{pi,qi}^\alpha\right)_{p,q}$ (see equations (B.11) and (B.12) in Ap-
 441 pendix B), instead of the reduced version of the full turbulent diffusion operator
 442 $\left(-\frac{1}{2}(\phi_p, \mathbf{G}^*(\otimes\mathbf{G}(\phi_q)))\right)_{p,q}$. This correction would have been implicit using the Galerkin
 443 projection with Stratonovich calculus. As demonstrated in [5, 58, 62], using Stratonovich
 444 integral, the turbulent dissipation term does not explicitly appear, neither when deriving
 445 the ROM from scratch nor when switching notations after the Galerkin projection. Let
 446 us stress, that the modified advection is still present in the Stratonovich form of the LU
 447 model. Changing the ROM notations from Itô to Stratonovich [37], the following correction
 448 term appear:

$$449 \quad (5.7) \quad \frac{1}{2}d\langle \mathbf{b}^T \boldsymbol{\alpha}_{\bullet p \bullet}, \mathbf{B} \rangle = \sum_{q=0}^n \left(\frac{1}{2} \sum_{i=1}^n \Sigma_{ip,qi}^\alpha \right) b_q dt = \sum_{q=0}^n \left(-\frac{1}{2} \sum_{i=1}^n \Sigma_{pi,qi}^\alpha \right) b_q dt.$$

450 In Stratonovich form the ROM also still implicitly include the diffusion term ensuing
 451 from the modes truncature. Notwithstanding, this additional dissipation in (5.6) is not a
 452 problem. It enables the setting of an energy dissipation representative of a direct energy
 453 cascade in the truncated modes’ coefficients. Hence, for very turbulent flows described by
 454 few modes, the energy leak appears to be necessary as our main concern is to restore energy
 455 fluxes between the coefficients of the reduced order system. Indeed, those energy fluxes
 456 can be very difficult to model correctly in reduced versions of non-linear systems. As an
 457 example, [63] observes that a basis encoding 50% of the energy can lead to a ROM missing
 458 more than 98% of energy transfers. These missing energy fluxes create instabilities in some
 459 coefficients of the reduced solution (missing negative energy fluxes) while overdamping
 460 others (missing positive energy fluxes). Restoring the energy fluxes between coefficients of
 461 the reduced system is a significant challenge LU models tackle through this natural energy
 462 leak

463 To note, the noise dimensionality reduction of section 4.4 introduces an additional
 464 energy leak due to the fact that we are only keeping n eigenvectors over $n(n+1)$ in the
 465 noise covariance matrix’s PCA. This introduces another negative energy flux as the noise
 466 variance is reduced while the dissipation is maintained.

467 **6. Ensemble forecasting.** As a means to measure the performance of the proposed
 468 ROM in UQ scenarios, ensemble simulations were carried out using the data available
 469 from DNS simulations of wake flows at Reynolds 100 and 300, the former being quasi
 470 two dimensional while the latter, is fully three dimensional. Wake flows are well-studied,
 471 non-linear, oscillatory flows which are physically produced by a uniform-velocity inflow
 472 facing a solid obstacle – here, a cylinder. Vortices are thus created behind the obstacle and
 473 periodically detach from it.

474 The dynamics, and in particular its intrinsic dimensionality, strongly differ between
 475 Reynolds 100 and 300. Indeed, as illustrated by figure 1, at Reynolds 100 most of the
 476 energy is concentrated in just a few coefficients of the reduced solution. In contrast, at
 477 Reynolds 300, the two first modes are meaningful for a rough approximation, but energy
 478 spreads over many degrees of freedom. The top panels of figures 3 and 7 confirm a strong
 479 difference in complexity between the two flows.

480 The ROM's results \mathbf{w}^R will be compared directly to DNS simulations \mathbf{v}_{ref} on a time
 481 interval outside intervals on which the ROM's coefficients and basis functions have been
 482 estimated

483 **6.1. Baseline ROMs.** To better appreciate our stochastic ROM's capabilities, results
 484 are compared to two different state-of-the-art algorithms.

485 **6.1.1. Deterministic baseline ROM.** The first state-of-art ROM is a prototype of
 486 deterministic ROM widely used in fluid dynamics. After a POD-Galerkin on the classical
 487 Navier-Stokes equations, the molecular viscosity coefficient $1/Re$ is replaced by an eddy
 488 viscosity coefficient $1/Re^{\text{ev}} \geq 1/Re$ [14]:

$$489 \quad (6.1) \quad \frac{d}{dt} b_i = \frac{Re}{Re^{\text{ev}}} \mathbf{b}^T \mathbf{l}_{\bullet i} + \mathbf{b}^T \mathbf{c}_{\bullet \bullet i} \mathbf{b}, \quad 1 \leq i \leq n.$$

490 The eddy viscosity coefficient is typically fitted by least squares using the dataset b^{obs} .
 491 Without that correction, some coefficients of the reduced solution might become unstable
 492 because of the mode truncation destabilizing processes described in 5.2. The eddy viscosity
 493 term generally manages to stabilize them, but this does not always make this data-driven
 494 method accurate. In order to emulate this behavior, we also simulate a simple DNS POD-
 495 Galerkin with no eddy viscosity model, i.e. the ROM (6.1) with $Re^{\text{ev}} = Re$.

496 **6.1.2. Stochastic baseline ROM.** As a deterministic model, the ROM above can
 497 hardly be used for dynamics-error UQ nor ensemble-based data assimilation. As a matter
 498 of fact, a simple randomisation of the initial conditions are known to usually lead to error
 499 underestimation (see section 1). So, to keep things simple, we define a second baseline
 500 ROM by adding a white noise term to the first baseline ROM. This constitutes a simple
 501 *ad hoc* randomization techniques of a given dynamical system through a Gaussian additive
 502 forcing. Despite its potential lack of physical relevance such a strategy is very often adopted
 503 in data-assimilation applications [16].

$$504 \quad (6.2) \quad db_i = \left(\frac{Re}{Re^{\text{ev}}} \mathbf{b}^T \mathbf{l}_{\bullet i} + \mathbf{b}^T \mathbf{c}_{\bullet \bullet i} \mathbf{b} \right) dt + \sigma_{i \bullet}^{\text{ev}} d\mathbf{W}_t, \quad 1 \leq i \leq n,$$

505 where $\sigma^{\text{ev}} \in \mathbb{R}^{n \times n}$ is the Cholesky decomposition of the the ROM's noise covariance
 506 $\Sigma^{\text{ev}} = \sigma^{\text{ev}} (\sigma^{\text{ev}})^T$ and $t \mapsto \mathbf{W}_t \in \mathbb{R}^n$ is a vector of n independent Brownian motions. Since
 507 no prior physical information is available for these baseline models, the noise's covariance
 508 is also learned from data as follows:

$$509 \quad (6.3) \quad \widehat{\Sigma}_{ij}^{\text{ev}} = \Delta t \overline{\left(\frac{\Delta b_i^{\text{obs}}}{\Delta t} \right)'' \left(\frac{\Delta b_j^{\text{obs}}}{\Delta t} \right)''}, \quad 1 \leq i, j \leq n,$$

510 where for $1 \leq i \leq n$,

$$511 \quad (6.4) \quad \left(\frac{\Delta b_i^{\text{obs}}}{\Delta t} \right)''_{\text{ev}} = \left(\frac{\Delta b_i^{\text{obs}}}{\Delta t} \right)'_{\text{ev}} - \overline{\left(\frac{\Delta b_i^{\text{obs}}}{\Delta t} \right)'_{\text{ev}}},$$

$$512 \quad (6.5) \quad \left(\frac{\Delta b_i^{\text{obs}}}{\Delta t} \right)'_{\text{ev}} = \left(\frac{\Delta b_i^{\text{obs}}}{\Delta t} \right)' - \left(\frac{Re}{Re^{\text{ev}}} (\mathbf{b}^{\text{obs}})^T \mathbf{l}_{\bullet i} + (\mathbf{b}^{\text{obs}})^T \mathbf{c}_{\bullet \bullet i} \mathbf{b}^{\text{obs}} \right).$$

513 As for the LU ROM, the implicit assumption of white noise residual $\left(\frac{\Delta b_i^{\text{obs}}}{\Delta t} \right)''_{\text{ev}}$ may not be
 514 valid. Thus, similarly to section 4.5 an optimal subsampling time step is estimated with
 515 $\sum_{k=1}^n \left(\frac{\Delta b_k^{\text{obs}}}{\Delta t} \right)''_{\text{ev}}(t_i) \left(\frac{\Delta b_k^{\text{obs}}}{\Delta t} \right)''_{\text{ev}}(t_j)$ instead of C_{ij}' before the noise covariance estimation.

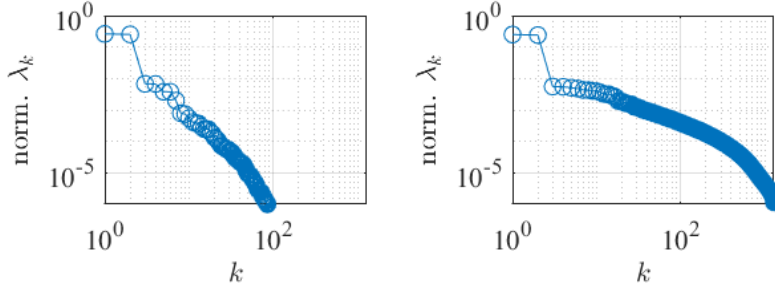


Figure 1. Energies of the reduced solution coefficients normalized by the total solution energy temporal mean $\|\mathbf{v}\|_{\mathcal{L}^2}^2$ for the 2D flow at Reynolds 100 (left panel) and the 3D flow at Reynolds 300 (right panel).

516 **6.2. 2D wake flow at Reynolds 100.** Starting with the simpler of the two test cases,
 517 800 seconds of DNS simulation data (about 160 pseudo-periods) were generated for a 2D
 518 wake flow behind a circular cylinder at $Re = 100$. These data were then split into train-
 519 ing and test sets, the former (140 pseudo-periods) being utilized for the construction of
 520 the ROMs, and the latter (20 pseudo-periods), as a reference to which the ROMs will be
 521 compared. Performance will thus be measured through four different metrics: the error in-
 522 curred by the model on each temporal coefficient in the time domain, by the reconstructed
 523 velocity fields with respect to the reference reconstructed field and the ground truth simu-
 524 lation, and the global prediction accuracy as measured by the evolution in time of RMSE,
 525 bias and ensemble minimum error.

526 **6.2.1. Temporal coefficients forecast.** The 2-dimensional LU ROM of the Reynolds-
 527 100 flow is simulated 100 times. In figure 2, the coefficients of the reduced solution $b_i^{(k)}$
 528 (where k stands for the k -th simulation) are compared to the coefficients of the reference
 529 solution $b_i^{\text{ref}} = (\phi_i, \mathbf{v}_{\text{ref}})$ (plotted in black). The ensemble mean (in green) follows almost
 530 perfectly the phase's reference and exhibits a slightly damped magnitude. The realization
 531 we have (randomly) singled-out shows a slowly divergent behavior. The damping effect can
 532 be seen as a consequence of the exchange of energy between the mean and the variance (see
 533 equation (5.3)), (this interplay is also evidenced by the growth of the confidence interval
 534).

535 **6.2.2. Velocity forecasts.** Having analyzed the model's forecast capabilities in terms
 536 of the ROMs's temporal coefficient, we now focus on the analysis of the velocity fields'
 537 forecast and compare them to the reference – i.e. the fullDNS simulation : $\tilde{\Pi}_\phi[\mathbf{v}_{\text{ref}}] \triangleq$
 538 $\tilde{\Pi}_\phi[\mathbf{v}_{\text{ref}} - \bar{\mathbf{v}}] + \bar{\mathbf{v}} = \sum_{i=0}^n b_i^{\text{ref}} \phi_i$ (with $b_0^{\text{ref}} = 1$). Here our goal will be to assess qualitatively
 539 the potential limitation of the ROM. Through equation (3.1), we will compute the velocity
 540 field for the mean and the lowest error realization, as well as, for comparizon purpose, the
 541 prediction generated by the eddy viscosity ROM (6.1) of same dimension. At each time
 542 step, we define the lowest-error realization as follow:

$$543 \quad (6.6) \quad \mathbf{b}^{\min} = \underset{(\mathbf{b}^{(k)})_k}{\text{argmin}} \left\| \mathbf{w}^R - \tilde{\Pi}_\phi[\mathbf{v}_{\text{ref}}] \right\|_{\mathcal{L}^2}^2 = \underset{(\mathbf{b}^{(k)})_k}{\text{argmin}} \|\mathbf{b}^{(k)} - \mathbf{b}^{\text{ref}}\|^2.$$

544 We plot in figure 3 the vorticity fields of the ROMs of dimension $n = 8$. A system of
 545 dimension 8 enable to reproduce fairly well the topology of the velocity fields From figure
 546 2 we conclude that our model's best realization is capable of staying in phase with the
 547 reference, even 50 seconds into the validation set, while the ensemble realizations mean
 548 and the deterministic baseline's start losing accuracy.

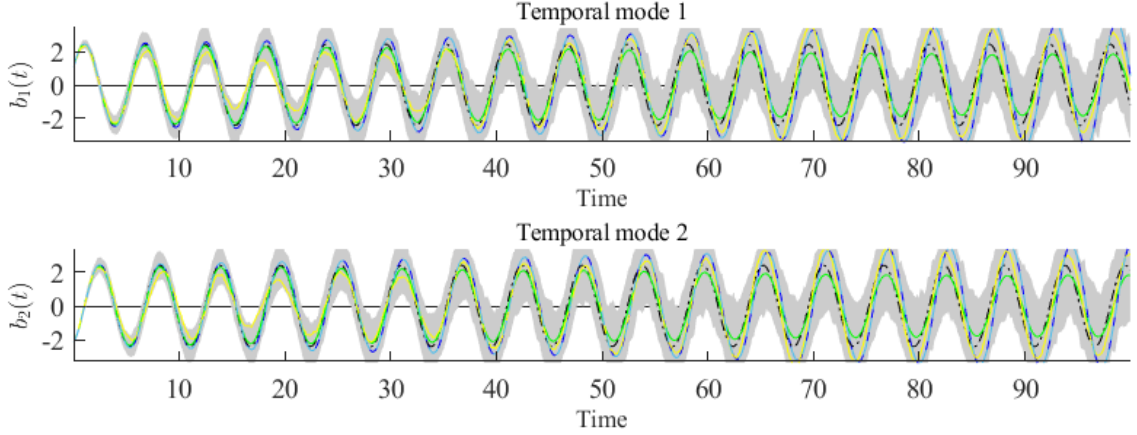


Figure 2. LU POD-ROM forecast for $n = 2$ coefficients of the reduced solution of a 2D wake flow at Reynolds 100 : ensemble mean (green line), one random realization (yellow line), confidence interval (gray shade). Blue lines correspond to deterministic ROMs : baseline eddy-viscosity POD-ROM (6.1) (light-blue dashed line) and DNS POD-Galerkin (dark-blue solid line). The dash-dot black plots are the observed references.

549 **6.2.3. Global prediction accuracy.** In order to compare more precisely the perfor-
 550 mance of the proposed ROM, we plot systematically the RMSE , as well as the ensemble’s
 551 bias, variance and minimum error. Those quantities will also be compared to those ob-
 552 tained by the stochastic baseline ROM (6.2). This should enable us to inspect the UQ
 553 capabilities of the proposed ROM.

554 Although random energy transfers is a necessary feature when the number of modes
 555 is not large enough, it increases the temporal coefficients variance in keeping the biases
 556 constant , which yields an increase of the Root Mean Square Error (RMSE) of the ensemble:

$$557 \quad (6.7) \quad \text{RMSE} \triangleq \hat{\mathbb{E}} \left\| \mathbf{w}^R - \tilde{\Pi}_\phi[\mathbf{v}_{\text{ref}}] \right\|_{\mathcal{L}^2}^2 = \left\| \hat{\mathbb{E}}\{\mathbf{w}^R\} - \tilde{\Pi}_\phi[\mathbf{v}_{\text{ref}}] \right\|_{\mathcal{L}^2}^2 + \int_{\Omega} \widehat{\text{Var}}(\mathbf{w}^R).$$

558 In all the following normalized error plots, the blue plots correspond to the solution
 559 computed using a deterministic POD-Galerkin ROMs (with and without eddy viscosity)
 560 and the red one, to the RMSE. The green line represents the ensemble bias, whereas the
 561 magenta line is the error incurred by the ensemble’s solution closest to the reference (6.6),
 562 and the gray shade corresponds to $1.96 \times$ the standard-deviation. Those are computed
 563 for the whole set of generated realizations. The initial condition is common to all and the
 564 values are normalized by the square root of the solution’s energy averaged over the training
 565 set : $(\|\bar{\mathbf{v}}\|_{\mathcal{L}^2}^2 + \sum_{i=1}^n \lambda_i)^{1/2}$.

566 In figure 4, the aforementioned error curves generated by the LU POD-ROM are com-
 567 pared to the stochastic POD with eddy viscosity model (6.2) for $n = \{2, 4, 8\}$. The LU
 568 POD-ROM best realization incurs rather low error, and for the first two cases, its bias too.
 569 The same cannot be said for the stochastic eddy viscosity model, which sees its variance
 570 grows out of control at the very beginning of the simulation, making the model’s predic-
 571 tions diverge almost instantly. As for the $n = 8$ simulation, we observe that it differs
 572 considerably from the other two by having unstable realizations, thence the sudden peaks
 573 in the bias and mean curves. When a realisation diverges, we re-sample it uniformly from
 574 the other members of the ensemble.

575 Focusing on our stochastic POD-ROM, in all three cases, even though the ensemble
 576 bias grows as time passes by, the realization with the lowest error does not. This is a crucial

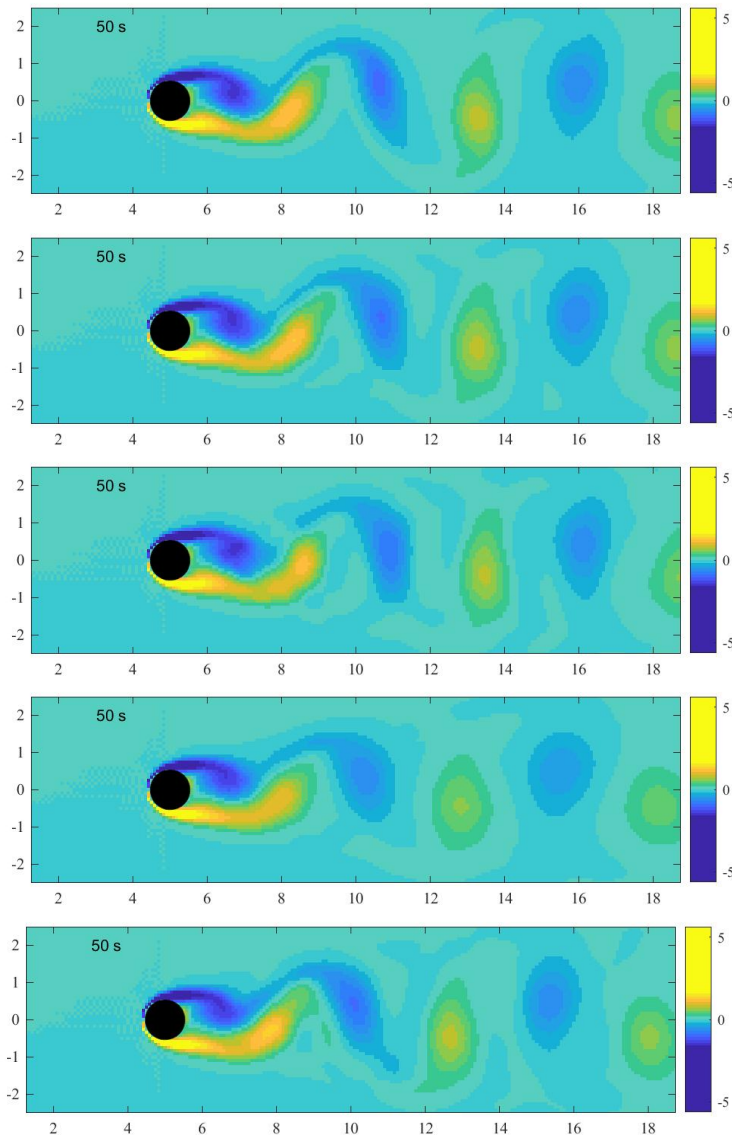


Figure 3. Vorticity fields – 10 vortex shedding cycles after the learning period – from (from top to bottom) the reference simulation \mathbf{v}_{ref} (2D DNS at Reynolds 100: state space dimension of about 10^4), its projection onto the POD basis $\tilde{\Pi}_\phi[\mathbf{v}_{ref}]$, the LU POD-ROM model best prediction, the LU POD-ROM model ensemble mean, and the prediction of the deterministic baseline POD-ROM (6.1) (fitted eddy viscosity) (ROM state spaces are of dimension 8). Vorticity is the velocity curl and is usual and convenient way for visualizing 2D flows and their vortices.

577 property for a UQ model to have, as this means that an appropriate filtering technique could
 578 eventually retrieve it at each observation, leading to a stable, low error data assimilation
 579 system.

580 **6.3. 3D wake flow at Reynolds 300.** For a more challenging test, a 3D wake flow
 581 behind a circular cylinder at $Re = 300$ was simulated, constructing the ROM just like
 582 in the case of the wake flow at Reynolds 100. The 440 seconds of DNS simulation data
 583 (about 88 vortex shedding pseudo-periods) were split into training (80 vortex shedding
 584 pseudo-periods) and test sets (8 pseudo-periods). Similarly, the ROM was subjected to
 585 the same benchmarks as before, with the equivalent eddy viscosity UQ enabled model for
 586 comparison. As it is a more complex flow, we expect the different temporal coefficients to

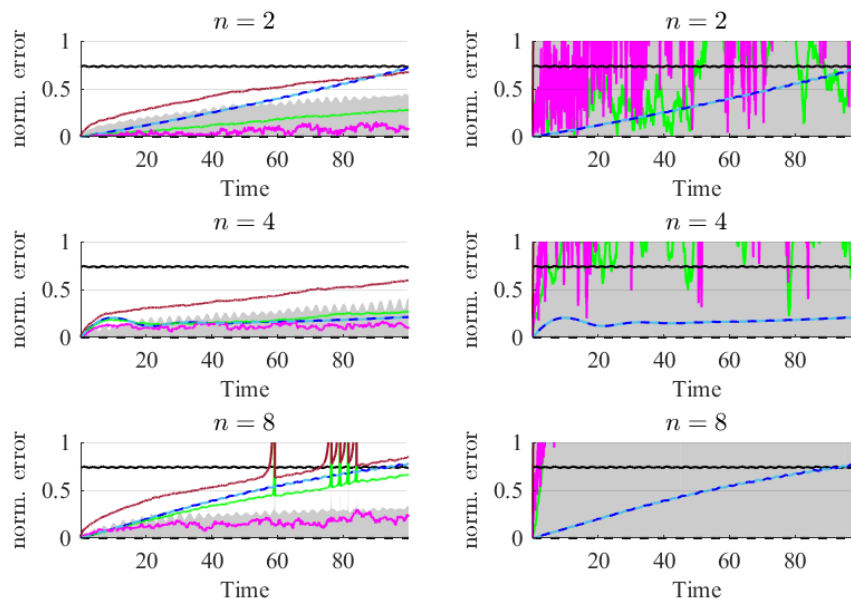


Figure 4. Normalized error for $n = 2, 4$ and 8 coefficients of the reduced solution of a 2D wake flow at Reynolds 100 – with the projection of the DNS onto the POD basis $\tilde{\Pi}_\phi[\mathbf{v}_{ref}]$ as reference : RMSE (brown line), bias (green line), ensemble minimum error (magenta line) and $1.96 \times$ the standard-deviation (shaded gray) for the LU POD-ROM (left panel) and stochastic baseline POD-ROM (right panel) . Blue lines correspond to deterministic ROMs : baseline eddy-viscosity POD-ROM (6.1) (light-blue dashed line) and DNS POD-Galerkin (dark-blue solid line). The black solid line at the top is the error considering only the time mean velocity, i.e. $b_i = 0, \forall i > 0$.

587 be harmonically richer to capture as much of the small-scale interactions as possible, but
 588 yet not to be able to fully reproduce the intricacies of the DNS-simulated flow.

589 **6.3.1. Temporal coefficients forecast.** This new scenario being fully 3D and more
 590 complex, the advantages of the proposed model are clear. This can be seen from the
 591 temporal coefficient forecasts, where the proposed LU stochastic ROM manages to stabilize
 592 the system while the deterministic PODs starts diverging after 5 seconds (i.e. before
 593 one complete vortex shedding), as evidenced by the first temporal coefficients in figure
 594 5. Intermittency probably fools the learning procedure of the eddy viscosity method and
 595 makes it less robust and not adapted to the test set. In contrast, the LU learning procedure
 596 of section 4 shows a greater robustness and leads to more accurate results. It is also
 597 interesting to note how the random energy contribution increases with the order of the
 598 temporal coefficient in the form of the ensemble variance. This is attributed to energy
 599 transfers between temporal coefficients, facilitated by the interplay of multiplicative noise
 600 and turbulent diffusion. This effect is in fact amplified in the 8-dimension ROM in figure 6,
 601 where the temporal coefficients' means get more damped as the amount of random energy
 602 in the system increases. This energy dissipation mechanism allows the system to remain
 603 bounded, unlike its deterministic counterparts.

604 Interestingly enough, in the $n = 8$ dimension ROM, even when the ensemble variance
 605 increases with the temporal coefficient order, coefficients 5, 6, 7 and 8 of the random
 606 realization (yellow line) also attempt to follow the right reference amplitude and variability,
 607 as it can be observed in figure 6.

608 **6.3.2. Velocity forecasts.** As in the previous case, we now analyze qualitatively the

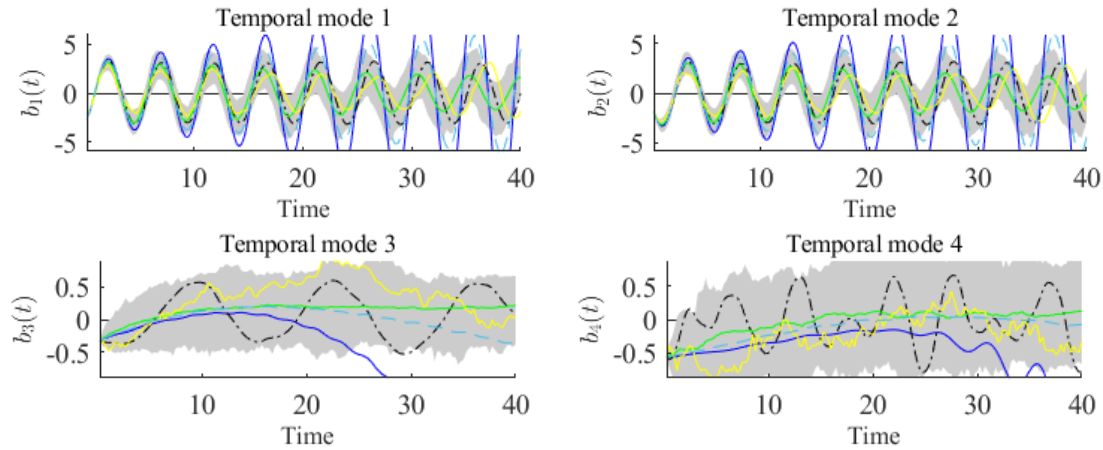


Figure 5. *LU POD-ROM forecast for $n = 4$ coefficients of the reduced solution of a 3D wake flow at Reynolds 300 : ensemble mean (green line), one random realization (yellow line), confidence interval (gray shade). Blue lines correspond to deterministic ROMs : baseline eddy-viscosity POD-ROM (6.1) (light-blue dashed line) and DNS POD-Galerkin (dark-blue solid line). The dash-dot black plots are the observed references.*

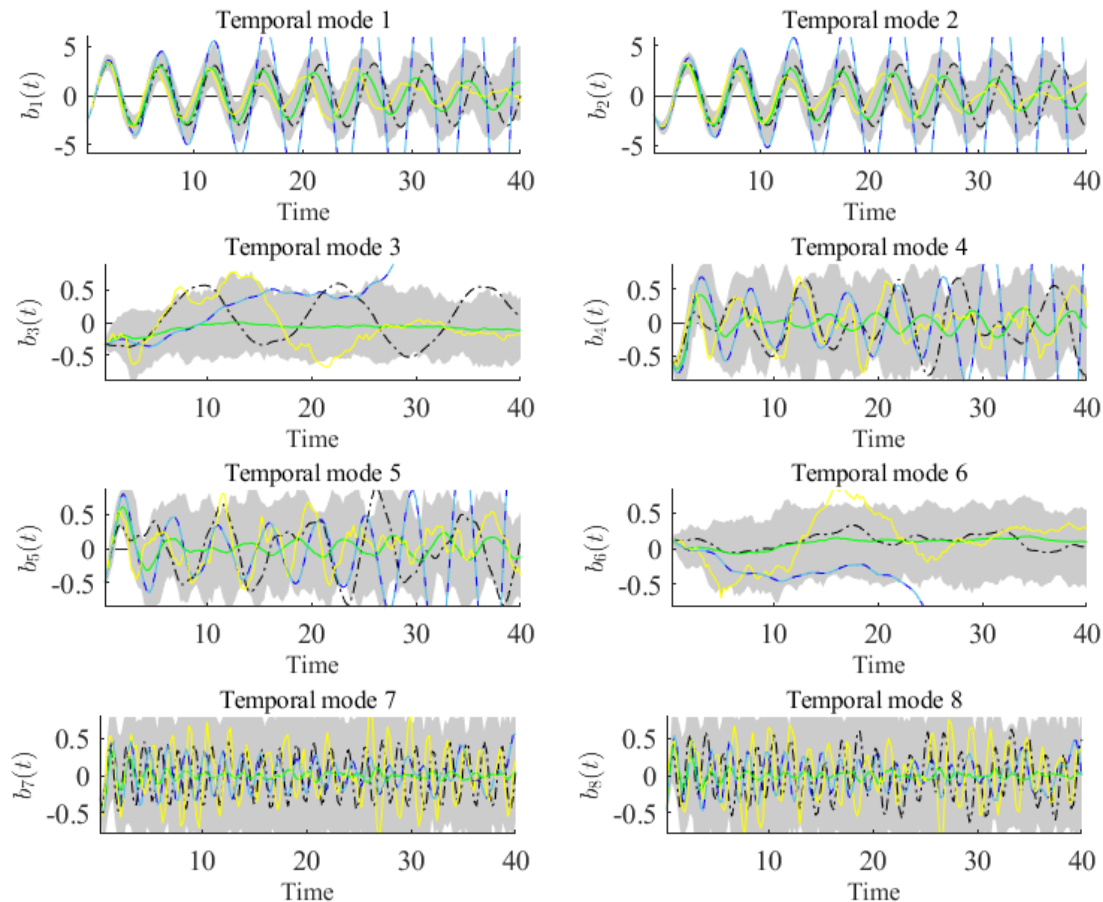


Figure 6. *LU POD-ROM forecast for $n = 8$ coefficients of the reduced solution of a 3D wake flow at Reynolds 300 : ensemble mean (green line), one random realization (yellow line), confidence interval (gray shade). Blue lines correspond to deterministic ROMs : baseline eddy-viscosity POD-ROM (6.1) (light-blue dashed line) and DNS POD-Galerkin (dark-blue solid line). The dash-dot black plots are the observed references.*

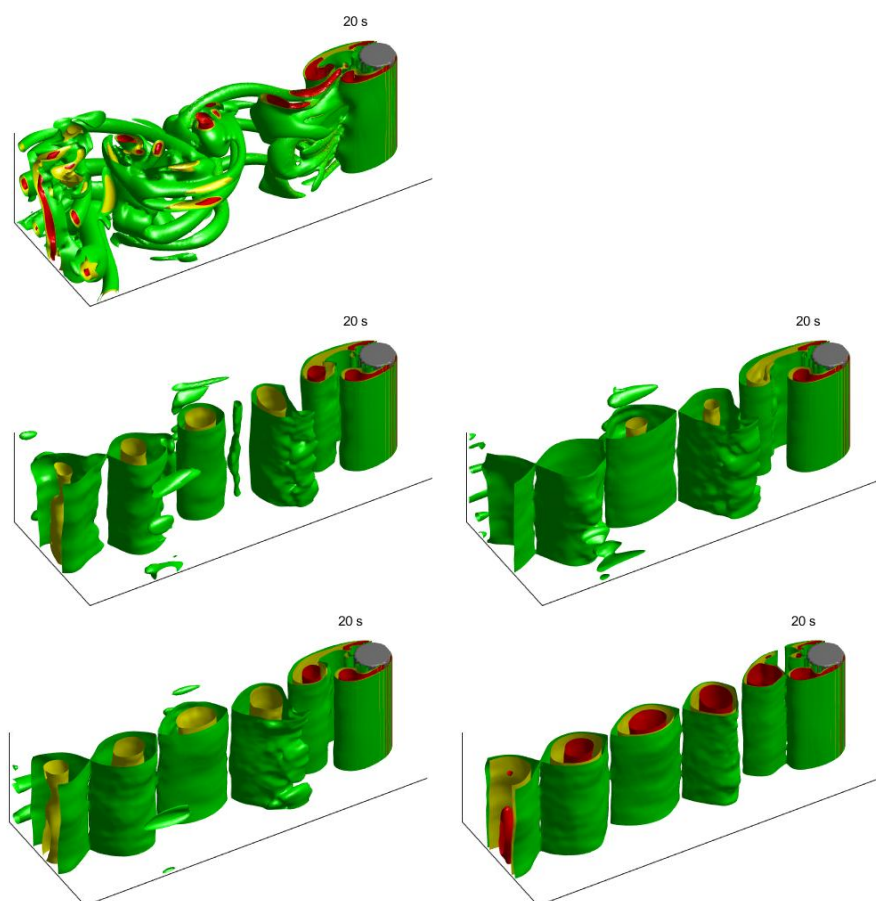


Figure 7. *Q-criterion iso-surfaces – 4 vortex shedding cycles after the learning period – from (from left to right and from top to bottom) the reference simulation \mathbf{v}_{ref} (3D DNS at Reynolds 300: state space dimension of about 10^7), its projection onto the POD basis $\tilde{\Pi}_\phi[\mathbf{v}_{ref}]$, the LU POD-ROM model ensemble mean, the LU POD-ROM model best prediction, and the prediction of the deterministic baseline POD-ROM (6.1) (fitted eddy viscosity) (POD-ROM state spaces are of dimension 8). *Q-criterion* [34] is a quadratic function of the velocity gradient and is usual and convenient way for visualizing 3D flows and their vortices.*

609 performance of our proposed method in in terms of velocity field predictions. As can be
 610 observed from figure 7, the best prediction remains stable and manages to stay close to the
 611 theoretically optimal solution in 8-dimensions on a period of four shedding cycles after the
 612 learning period. Conversely, the baseline model starts to diverge, as evidenced by the red
 613 zones inside the vortices. This result is quite impressive as, even if the mean velocity field
 614 is not as close to the reference as one would like, it still reveals the potential of a method
 615 combining our model with a data assimilation technique to retrieve the best realization
 616 and provide corrections on the fly.

617 **6.3.3. Global prediction accuracy.** Just like in the two-dimensional flow, the error
 618 curves are plotted and compared to the results obtained with the stochastic eddy viscosity
 619 model (6.2). The curves in figure 8 evidence the predictive power of our proposed model, is
 620 capable of great stability even after 40 seconds while the deterministic and eddy viscosity
 621 models prove to be quite unstable, with a rapid divergence. This property is showcased by
 622 the bias and mean curves that converge to the zero-temporal-coefficient error curve instead
 623 of growing in an unbound manner, while still having at least one realization with relatively
 624 low error that could eventually be identified through data assimilation techniques.

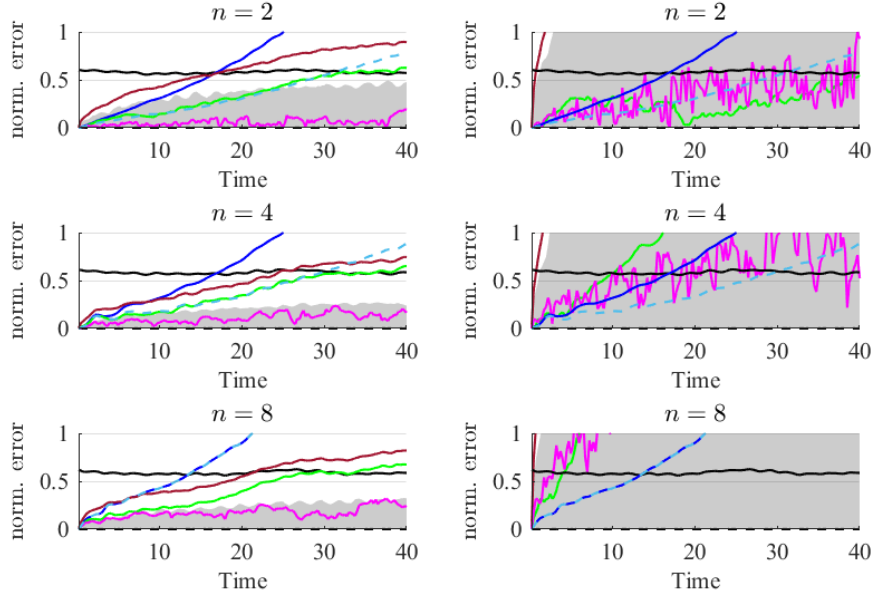


Figure 8. Normalized error for $n = 2, 4$ and 8 coefficients of the reduced solution of a 3D wake flow at Reynolds 300 – with the projection of the DNS onto the POD basis $\bar{\Pi}_\phi[\mathbf{v}_{ref}]$ as reference : RMSE (red line), bias (green line), ensemble minimum error (magenta line) and $1.96 \times$ the standard-deviation (shaded gray) for the LU POD-ROM (left panel) and stochastic baseline POD-ROM (right panel) . Blue lines correspond to deterministic ROMs : baseline eddy-viscosity POD-ROM (6.1) (light-blue dashed line) and DNS POD-Galerkin (dark-blue solid line). The black solid line at the top is the error considering only the time mean velocity, i.e. $b_i = 0, \forall i > 0$.

625 **7. Conclusion.** This paper proposed a a stochastic ROM derived from a stochastic
 626 fluid flow dynamics modelling setting, called dynamics under location uncertainty (LU),
 627 which formulates unresolved small-scale parameterization through SPDEs. The stochastic
 628 ROM is then obtained through a classical POD-Galerkin projection of these SPDEs, and
 629 the basis function are defined from high resolution simulations of the target flow. The
 630 resulting model bears similarities with the Navier-Stokes equations, to now encompasses
 631 an advection velocity correction, a turbulent diffusion and a skew-symmetric multiplicative
 632 noise term. From the statistics of the POD’s residuals – i.e. the data component orthogonal
 633 to the POD’s modes –, these new terms can be fully characterized.

634 The implementation necessitates additional offline computations on top of the clas-
 635 sical POD-Galerkin procedure: the estimation of a variance tensor and of the reduced
 636 multiplicative noise’s covariance matrix. The former, proportional to the point-wise $d \times d$
 637 covariance matrix of that POD’s residual, is readily computed. However, the latter inher-
 638 ently constitutes a formidable computational challenge, due to the enormous size of the
 639 dataset’s elements, to hamper the full usage of the stochastic closure mechanism. Also,
 640 with a fully data-driven method, the risk of over-fitting is very high. To circumvent these
 641 issues, an intermediate solution is proposed with an easy-to-compute estimator, along with
 642 consistency proofs under LU setting assumptions. Since the latter may not always be met
 643 – specifically, the time decorrelation assumption of the POD residual –, measures are taken
 644 to enforce them on the model, as part of its construction process. Namely, to force the
 645 decorrelation – and in doing so, improve the accuracy of the estimators and the ROM as
 646 a whole –, a down-sampling of the dataset is applied at a rate equal to the this residual’s
 647 correlation time. Finally, as a means to restrict the number of coefficients needed to char-

648 acterize the ROM to the usual $O(n^3)$, a technique to reduce the dimensionality of the noise
649 covariance matrix was also presented.

650 The conservative properties of the LU closure are discussed, and we demonstrate that,
651 when performing Galerkin projections of the It \bar{o} form of LU SPDEs, we end up forfeiting
652 these properties. In the reduced order version, two energy fluxes appear: the first one is
653 attributed to the possible advecting velocity correction divergence, whilst the second one
654 is negative and is directly associated with mode truncation. We argue that this energy
655 loss is understandable, and even desirable, with an interplay between noise and turbulent
656 diffusion to maintain non-linear energy fluxes between coefficients of the reduced solution
657 despite mode truncation.

658 Numerical comparisons are performed between our stochastic ROM and state-of-the-art
659 deterministic and stochastic ROMs. As test cases, we chose a two-dimensional wake flow
660 at Reynolds 100 with few degrees of freedom, and a more complex three-dimensional wake
661 flow at Reynolds 300 with many more degrees of freedom. Deterministic Navier-Stokes
662 simulations stood for references and all of the ROMs were initialized with their values, and
663 the forecasts, compared to them. For the stochastic ROMs, ensembles of 100 realizations
664 were forecast. The state-of-the-art stochastic ROM's solutions quickly diverged in time,
665 whilst the temporal coefficients of our ROM proved to be neither unstable nor overdamped.
666 The LU POD-ROM solution's biases were even found to be smaller than those of each one
667 of the other stochastic ROMs we have considered. Moreover, at any given moment, the
668 LU POD-ROM ensembles managed to remain very close to the reference, suggesting that
669 they could be efficient priors for Bayesian inverse problems.

670

671 For upcoming work, an application of LU POD-ROM in conjunction with particle
672 filter algorithms will be considered to estimate velocity flows in real-time from few local
673 measurements.

674 Finally, when working at very large Reynolds numbers, DNS simulations are no longer
675 an option, but LES and RANS (Reynolds Averaged Navier Stokes) can still provide use-
676 ful data to build ROMs on, at the cost of some precision. To tackle this, the closure
677 mechanism's estimation procedure may be adjusted to address the small-scale velocity
678 statistics neglected by LES-like approaches. Another approach is to consider the Galerkin
679 projection on a LU version of LES-like PDEs. Still a gap will appear between neglected
680 small-scale velocities and the real measurements of the flow when applied to the forth-
681 coming data assimilation procedures. Besides, in realistic applications, Reynolds number,
682 initial and boundary conditions, as well as a plethora of other parameters, are often poorly
683 known. So, combinations with other state-of-the-art dimensionality reduction methods are
684 probably necessary. For the moment, the on-going implementation of LU terms in the
685 ITHACA-FV library [69, 70] will surely aid on these practical aspects.

686 At last, relying purely on data can offer the advantage directly tackle very high
687 Reynolds number conditions. It requires to filter noisy measurements and likely, to rely
688 on simplified (often 2D) data-driven models [1, 12, 26, 72]. From this latter point of view,
689 modeling under location location offers a great flexibility which should of major interest.

690 **Appendix A. Estimation formulas.** In this appendix, we consider a a full probability
691 space (A, \mathcal{F}, P) , and a filtration of σ -algebra $\{\mathcal{F}_t\}_{t \geq 0}$. For a given Hilbert space $H(\Omega)$ of real
692 functions, we understand $\mathcal{H}(0, T; \Omega, \mathcal{F})$ as the space of $H(\Omega)$ -valued, strongly measurable,
693 $\{\mathcal{F}_t\}_{t \geq 0}$ adapted processes such that

694 (A.1)
$$\mathbb{E} \int_0^T \|u(s)\|_H^2 ds < \infty.$$

695 In the following we assume that the POD modes ϕ_i belong to $H^2(\Omega)$ while $(\mathbf{x} \mapsto \mathbf{v}'(\mathbf{x}, t))$
 696 and $(\mathbf{x} \mapsto \boldsymbol{\sigma}\mathbf{B}(\mathbf{x}, t))$ belong to $\mathcal{H}^2(0, T; \Omega, \mathcal{F})$; we assume also that the POD modes'
 697 gradients are uniformly bounded on Ω . We recall below a classical proposition related to
 698 quadratic variation process:

699 *Proposition: Stochastic integration and quadratic variations.* If M is a continuous mar-
 700 tingale and $X \in \mathcal{L}^2(0, T; \Omega, \mathcal{F})$ then there exists a unique bounded continuous martingale
 701 $\int_0^t X dM$ such that for every continuous martingale N (with zero initial condition)

$$702 \quad (\text{A.2}) \quad \left\langle \int_0^t X dM, N \right\rangle = \int_0^t X d\langle M, N \rangle.$$

703 Also, after defining boundary conditions, \mathcal{P} and \mathbf{K} are well defined on $\mathcal{H}^1(0, T, \Omega, \mathcal{F})$
 704 and $\mathcal{H}^2(0, T; \Omega, \mathcal{F})$ respectively. Then, for every $\zeta \in \mathcal{H}^1(0, T; \Omega, \mathcal{F})$, $\|\mathcal{P}\zeta\|_{\mathcal{L}^2} \leq \|\zeta\|_{\mathcal{L}^2}$ since
 705 \mathcal{P} is a projection and then, for every $\boldsymbol{\xi} \in \mathcal{H}^2(0, T; \Omega, \mathcal{F})$, by Cauchy-Schwarz and triangular
 706 inequalities:

$$707 \quad (\text{A.3}) \quad |K_{jq}(\boldsymbol{\xi})| \leq \|\phi_j\|_{\mathcal{L}^2} (\|\nabla \phi_q^T\|_{\infty} + \nu) \|\boldsymbol{\xi}\|_{\mathcal{H}^2},$$

708 which leads to the continuity of \mathbf{K} . We also assume that the observed coefficients of the
 709 reduced solution $b_i^{\text{obs}} = (\phi_i, \mathbf{v}_{\text{obs}})$ are continuous semi-martingales and solutions of the
 710 ROM (3.3)-(3.4). From there, the orthogonality of the coefficients of the reduced solution
 711 yields, for $1 \leq i, j \leq n$ and $0 \leq p, q \leq n$:

$$712 \quad (\text{A.4}) \quad \int_0^T b_p d\langle b_i, \int_0^t (\tilde{\alpha}_{qj} \bullet d\mathbf{B}_s) \rangle = \int_0^T \sum_{k=0}^n b_p d\langle \int_0^t (\tilde{\alpha}_{ki} \bullet d\mathbf{B}_s) b_k, \int_0^t \tilde{\alpha}_{qj} \bullet d\mathbf{B}_s \rangle,$$

$$713 \quad (\text{A.5}) \quad = \sum_{k=0}^n \left(\int_0^T b_p b_k \right) \Sigma_{ki, qj}^{\alpha},$$

$$714 \quad (\text{A.6}) \quad = T \lambda_p \Sigma_{pi, qj}^{\alpha}.$$

715 Now, let us note that for $1 \leq j \leq n$ and $0 \leq q \leq n$:

$$716 \quad (\text{A.7}) \quad \tilde{\alpha}_{qj} \bullet d\mathbf{B}_s = (\phi_j, -\mathcal{P}[(\boldsymbol{\sigma} d\mathbf{B}_s \cdot \nabla) \phi_q] + \delta_{q0} \nu \Delta \boldsymbol{\sigma} d\mathbf{B}_s) = d(K_{jq}[\boldsymbol{\sigma}\mathbf{B}]).$$

717 Then, with the definition of the quadratic covariation and the increment notation $\Delta\xi(t_k) =$
 718 $\xi(t_{k+1}) - \xi(t_k)$, we obtain the estimator's expression and its consistency for every $1 \leq i, j \leq$
 719 n and $0 \leq p, q \leq n$:

$$720 \quad (\text{A.8}) \quad \Sigma_{pi, qj}^{\alpha} = \frac{1}{\lambda_p T} \int_0^T b_p d\langle b_i, \int_0^t (\tilde{\alpha}_{qj} \bullet d\mathbf{B}_s) \rangle,$$

$$721 \quad (\text{A.9}) \quad = \frac{1}{\lambda_p T} \int_0^T b_p d\langle b_i, K_{jq}[\boldsymbol{\sigma}\mathbf{B}] \rangle,$$

$$722 \quad (\text{A.10}) \quad = \frac{1}{\lambda_p T} \mathbb{P}\text{-}\lim_{\Delta t \rightarrow 0} \sum_{t_k=0}^T b_p(t_k) (\Delta b_i)(t_k) K_{jq}[\boldsymbol{\sigma} \Delta \mathbf{B}_{t_k}],$$

$$723 \quad (\text{A.11}) \quad = K_{jp} \left[\frac{1}{\lambda_p T} \mathbb{P}\text{-}\lim_{\Delta t \rightarrow 0} \sum_{t_k=0}^T b_p(t_k) (\Delta b_i)(t_k) \boldsymbol{\sigma} \Delta \mathbf{B}_{t_k} \right],$$

724 where the continuity of the operator K_{jp} on $\mathcal{H}^2(0, T; \Omega, \mathcal{F})$ and the previous proposition
 725 enabled us to switch the limit in probability and the operator K_{jp} . The martingale flow

726 increments $\sigma \Delta \mathbf{B}_{t_k}$ are approximated by $v'(\bullet, t_k) \Delta t$. In practice, for each $1 \leq i \leq n$ we
 727 can replace Δb_i by

$$728 \quad (\text{A.12}) \quad \Delta b_i'' = \Delta b_i' - \overline{\Delta b_i'},$$

729 where

$$730 \quad (\text{A.13}) \quad \Delta b_i' = \Delta b_i - \left(\mathbf{b}^T (\mathbf{l} + \check{\mathbf{f}})_{\bullet i} + \mathbf{b}^T \mathbf{c}_{\bullet \bullet i} \mathbf{b} \right) \Delta t.$$

731 Mathematically, this is still correct for very large values of T . Indeed,

$$732 \quad \overline{\Delta b_i'} = \frac{b_i(T+\Delta T) - b_i(0)}{T} \Delta t \xrightarrow{T \rightarrow \infty} 0 \text{ and } b_i' - b_i \text{ has finite variations. Thus, } b_i'' - b_i \text{ approaches}$$

733 a finite variation process (for large T). Numerically, it is more accurate as it allows us to
 734 remove the smooth-in-time part of Δb_i , thus minimizing estimation errors as well.

735 **Appendix B. Energy dissipation.**

736 Hereafter, we neglect the boundary conditions and the viscosity ($\mathbf{L} = 0$), and we
 737 assume σ to be Hilbert-Schmidt and the modes to be orthonormal. We also assume that
 738 the reduced basis' functions are in $H_0^2(\Omega)$ and $\mathbf{x} \mapsto \sigma \mathbf{B}(\mathbf{x}, t)$ in $\mathcal{H}_0^2(0, T; \Omega, \mathcal{F})$. Both are
 739 divergence-free (since they are learned from a set of incompressible velocity fields). With
 740 the same assumption as in appendix A, we have $\forall i \leq n$, $\|\nabla \phi_i^T\|_\infty = \sup_\Omega \|\nabla \phi_i^T\| < \infty$

741 (where $\|\bullet\|$ stands for the Euclidean norm on $\mathbb{R}^{d \times d}$) and thus,

$$742 \quad \|\mathbf{G}(\phi_i)\|_{\text{HS}}^2 \leq \|\nabla \phi_i^T\|_\infty^2 \|\sigma\|_{\text{HS}}^2 < \infty, \text{ i.e. } \mathbf{G}(\phi_i) \text{ is Hilbert-Schmidt. This makes } \mathbf{G}(\mathbf{w}^R)$$

743 and $\Pi_\phi [\mathbf{G}(\mathbf{w}^R)]$ also Hilbert-Schmidt.

744 Using the projected Navier-Stokes model (5.4), for every $t \in [0, T]$ we can formally
 745 remove the orthogonal projection by moving it into the divergence-free functions space \mathcal{P}
 746 through integration by parts:

$$747 \quad (\text{B.1}) \quad d_t \mathbf{w}^R = \sum_{i=1}^n (\phi_i, \mathcal{P}(d\mathbb{M})(\mathbf{w}^R)) \phi_i = \sum_{i=1}^n (\mathcal{P}\phi_i, (d\mathbb{M})(\mathbf{w}^R)) \phi_i,$$

$$748 \quad (\text{B.2}) \quad = \sum_{i=1}^n (\phi_i, (d\mathbb{M})(\mathbf{w}^R)) \phi_i = \Pi_\phi [(\mathbf{C}(\mathbf{w}^R, \mathbf{w}^R) + \mathbf{F}(\mathbf{w}^R)) dt + (\mathbf{G}d\mathbf{B}_t)(\mathbf{w}^R)].$$

749 Then, upon applying the Itô formula to the local kinetic energy we obtain

$$750 \quad (\text{B.3}) \quad d \left(\frac{1}{2} \|\mathbf{w}^R\|_{\mathcal{L}^2}^2 \right) = \int_\Omega (d_t(\mathbf{w}^R)^T \mathbf{w}^R + \frac{1}{2} d_t((\mathbf{w}^R)^T, \mathbf{w}^R)) \quad \forall t \in [0, T].$$

751 To remove the (orthogonal) projection operator Π_ϕ from the first term, we exploit its
 752 symmetry, and afterwards, the fact that \mathbf{w}^R is already in the reduced subspace:

$$753 \quad (\text{B.4}) \quad \int_\Omega d_t(\mathbf{w}^R)^T \mathbf{w}^R = (\Pi_\phi [(d\mathbb{M})(\mathbf{w}^R)], \mathbf{w}^R),$$

$$754 \quad (\text{B.5}) \quad = ((d\mathbb{M})(\mathbf{w}^R), \Pi_\phi [\mathbf{w}^R]),$$

$$755 \quad (\text{B.6}) \quad = ((d\mathbb{M})(\mathbf{w}^R), \mathbf{w}^R),$$

$$756 \quad (\text{B.7}) \quad = (((\mathbf{F}_1 + \mathbf{F}_2)(\mathbf{w}^R) + \mathbf{C}(\mathbf{w}^R, \mathbf{w}^R)) dt + (\mathbf{G}d\mathbf{B}_t)(\mathbf{w}^R), \mathbf{w}^R),$$

$$757 \quad = \underbrace{\left(-\frac{1}{2} \mathbf{G}^* (\otimes \mathbf{G}(\mathbf{w}^R)), \mathbf{w}^R \right) dt}_{\text{from } (\mathbf{F}_1(\mathbf{w}^R), \mathbf{w}^R) \text{ using (2.11)}} + \underbrace{\left(\frac{1}{2} ((\nabla \cdot \mathbf{a}) \nabla) \mathbf{w}^R, \mathbf{w}^R \right)}_{\text{from } (\mathbf{F}_2(\mathbf{w}^R), \mathbf{w}^R)}$$

$$758 \quad (\text{B.8}) \quad + \underbrace{\left(\mathbf{C}(\mathbf{w}^R, \mathbf{w}^R) dt + (\mathbf{G}d\mathbf{B}_t)(\mathbf{w}^R), \mathbf{w}^R \right)}_{=0 \text{ by skew-symmetry of } \xi \mapsto \mathbf{C}(\mathbf{w}^R, \xi) \text{ and } \mathbf{G}d\mathbf{B}_t},$$

$$759 \quad (\text{B.9}) \quad = \frac{1}{2} \|\mathbf{G}(\mathbf{w}^R)\|_{\text{HS}}^2 - \frac{1}{2} (\nabla \cdot (\nabla \cdot \mathbf{a})^T, \|\mathbf{w}^R\|^2),$$

760 where the second term comes from integration by parts. Besides, the Itô term of the energy
761 budget is straightforward to compute from equation (B.2):

$$762 \quad (B.10) \quad \int_{\Omega} \frac{1}{2} d_t \langle (\mathbf{w}^R)^T, \mathbf{w}^R \rangle = \frac{1}{2} \|\Pi_{\phi} [\mathbf{G}(\mathbf{w}^R)]\|_{\text{HS}}^2 dt,$$

$$763 \quad (B.11) \quad = \sum_{p,q=0}^n \left(\frac{1}{2} \int_{\Omega} \Pi_{\phi} [\mathbf{G}(\phi_p)] \otimes \Pi_{\phi} [\mathbf{G}(\phi_q)] \right) b_p b_q dt.$$

764 From the definition of the projection operator Π_{ϕ} and using the extended notation $(\zeta \otimes \xi) \triangleq$
765 $\int_{\Omega} \check{\zeta}(\mathbf{z}) \check{\xi}^T(\mathbf{z}) d\mathbf{z}$, we can express the above quadratic operator with the noise statistics as
766 follows:

$$767 \quad (B.12) \quad \int_{\Omega} \Pi_{\phi} [\mathbf{G}(\phi_p)] \otimes \Pi_{\phi} [\mathbf{G}(\phi_q)] = \sum_{i=1}^n (\phi_i, \mathbf{G}(\phi_p)) \otimes (\phi_i, \mathbf{G}(\phi_q)) = \sum_{i=1}^n \Sigma_{pi,qi}^{\alpha}.$$

768 Finally, by orthogonality, the kinetic energy budget (B.3) simplifies to:

$$769 \quad (B.13) \quad \frac{d}{dt} \left(\frac{1}{2} \|\mathbf{w}^R\|_{\mathcal{L}^2}^2 \right) = \frac{1}{2} \|\mathbf{G}(\mathbf{w}^R)\|_{\text{HS}}^2 - \frac{1}{2} (\nabla \cdot (\nabla \cdot \mathbf{a})^T, \|\mathbf{w}^R\|^2) - \frac{1}{2} \|\Pi_{\phi} [\mathbf{G}(\mathbf{w}^R)]\|_{\text{HS}}^2,$$

$$770 \quad (B.14) \quad = -\frac{1}{2} \|\Pi_{\phi}^{\perp} [\mathbf{G}(\mathbf{w}^R)]\|_{\text{HS}}^2 - \frac{1}{2} (\nabla \cdot (\nabla \cdot \mathbf{a})^T, \|\mathbf{w}^R\|^2),$$

771 where $\Pi_{\phi}^{\perp} = \mathbb{I}_d - \Pi_{\phi}$ is the projector on the orthogonal complement of the reduced subspace.

772 **Acknowledgments.** We warmly thank Pranav Chandramouli for the generation of the
773 three-dimensional wake flow data, Pierre-Louis Lee for his help in the generation of the
774 two-dimensional wake flow data, and Reda Bouaida for his help in the computation of the
775 ROM matrix \mathbf{f} . The authors also acknowledge the support of the ERC EU project 856408-
776 STUOD, the ESA DUE GlobCurrent project, the ‘‘Laboratoires d’Excellence’’ CominLabs,
777 Lebesgue and Mer through the SEACS project. Finally, we also thank Darryl D. Holm,
778 Dan Crisan, Wei Pan and Igor Shevchenko for the insightful discussions.

779 References.

- 780 [1] G. ARTANA, A. CAMMILLERI, J. CARLIER, AND E. MÉMIN, *Strong and weak con-*
781 *straint variational assimilations for reduced order fluid flow modeling*, Journal of Com-
782 *putational Physics*, 231 (2012), pp. 3264–3288.
- 783 [2] N. AUBRY, P. HOLMES, J. LUMLEY, AND E. STONE, *The dynamics of coherent*
784 *structures in the wall region of a turbulent boundary layer*, J. Fluid Mech., 192 (1988),
785 pp. 115–173.
- 786 [3] R. AZENCOTT, A. BERI, A. JAIN, AND I. TIMOFEYEV, *Sub-sampling and parametric*
787 *estimation for multiscale dynamics*, Communications in Mathematical Sciences, 11
788 (2013), pp. 939–970.
- 789 [4] R. AZENCOTT, A. BERI, AND I. TIMOFEYEV, *Adaptive sub-sampling for parametric*
790 *estimation of Gaussian diffusions*, Journal of Statistical Physics, 139 (2010), pp. 1066–
791 1089.
- 792 [5] W. BAUER, P. CHANDRAMOULI, B. CHAPRON, L. LI, AND E. MÉMIN, *Deciphering*
793 *the role of small-scale inhomogeneity on geophysical flow structuration: a stochastic*
794 *approach*, Journal of Physical Oceanography, 50 (2020), pp. 983–1003.
- 795 [6] W. BAUER, P. CHANDRAMOULI, L. LI, AND E. MÉMIN, *Stochastic representation of*
796 *mesoscale eddy effects in coarse-resolution barotropic models*, Ocean Modelling, (2020).
- 797 [7] J. BERNER, S.-Y. HA, J. HACKER, A. FOURNIER, AND C. SNYDER, *Model un-*
798 *certainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics*
799 *representations*, Monthly Weather Review, 139 (2011), pp. 1972–1995.

- 800 [8] S. BOYAVAL, C. LE BRIS, T. LELIEVRE, Y. MADAY, N. C. NGUYEN, AND A. T.
801 PATERA, *Reduced basis techniques for stochastic problems*, Archives of Computational
802 methods in Engineering, 17 (2010), pp. 435–454.
- 803 [9] S. L. BRUNTON, J. L. PROCTOR, AND J. N. KUTZ, *Discovering governing equations*
804 *from data by sparse identification of nonlinear dynamical systems*, Proceedings of the
805 national academy of sciences, 113 (2016), pp. 3932–3937.
- 806 [10] Z. BRZEŹNIAK, M. CAPIŃSKI, AND F. FLANDOLI, *Stochastic partial differential equa-*
807 *tions and turbulence*, Mathematical Models and Methods in Applied Sciences, 1 (1991),
808 pp. 41–59.
- 809 [11] M. BUFFONI, S. CAMARRI, A. IOLO, AND M. V. SALVETTI, *Low-dimensional*
810 *modelling of a confined three-dimensional wake flow*, Journal of Fluid Mechanics, 569
811 (2006), pp. 141–150.
- 812 [12] A. CAMMILLERI, F. GUENIAT, J. CARLIER, L. PASTUR, E. MÉMIN, F. LUSSEYRAN,
813 AND G. ARTANA., *POD-spectral decomposition for fluid flow analysis and model re-*
814 *duction*, Theor. and Comp. Fluid Dyn., (2013).
- 815 [13] K. CARLBERG, C. BOU-MOSLEH, AND C. FARHAT, *Efficient non-linear model re-*
816 *duction via a least-squares Petrov–Galerkin projection and compressive tensor approx-*
817 *imations*, International Journal for Numerical Methods in Engineering, 86 (2011),
818 pp. 155–181.
- 819 [14] W. CAZEMIER, R. VERSTAPPEN, AND A. VELDMAN, *Proper orthogonal decomposi-*
820 *tion and low-dimensional models for driven cavity flows*, Physics of fluids, 10 (1998),
821 pp. 1685–1699.
- 822 [15] P. CHANDRAMOULI, D. HEITZ, S. LAIZET, AND E. MÉMIN, *Coarse large-eddy sim-*
823 *ulations in a transitional wake flow with flow models under location uncertainty*, Com-
824 puters & Fluids, 168 (2018), pp. 170–189.
- 825 [16] B. CHAPRON, P. DÉRIAN, E. MÉMIN, AND V. RESSEGUIER, *Large-scale flows under*
826 *location uncertainty: a consistent stochastic framework*, Quarterly Journal of the Royal
827 Meteorological Society, 144 (2018), pp. 251–260.
- 828 [17] G. CHEN, J. SUN, AND Y.-M. LI, *Adaptive reduced-order-model-based control-law*
829 *design for active flutter suppression*, Journal of Aircraft, 49 (2012), pp. 973–980.
- 830 [18] P. CHEN, A. QUARTERONI, AND G. ROZZA, *Reduced basis methods for uncertainty*
831 *quantification*, SIAM/ASA Journal on Uncertainty Quantification, 5 (2017), pp. 813–
832 869.
- 833 [19] F. CHINESTA, P. LADEVEZE, AND E. CUETO, *A short review on model order reduc-*
834 *tion based on proper generalized decomposition*, Archives of Computational Methods
835 in Engineering, 18 (2011), pp. 395–404.
- 836 [20] L. CORDIER, B. R. NOACK, G. TISSOT, G. LEHNASCH, J. DELVILLE, M. BALA-
837 JEWICZ, G. DAVILLER, AND R. K. NIVEN, *Identification strategies for model-based*
838 *control*, Experiments in fluids, 54 (2013), p. 1580.
- 839 [21] C. COTTER, D. CRISAN, D. D. HOLM, W. PAN, AND I. SHEVCHENKO, *Numerically*
840 *modeling stochastic lie transport in fluid dynamics*, Multiscale Modeling & Simulation,
841 17 (2019), pp. 192–232.
- 842 [22] C. COTTER, D. CRISAN, D. D. HOLM, W. PAN, AND I. SHEVCHENKO, *Sequential*
843 *monte carlo for stochastic advection by Lie transport (SALT): A case study for the*
844 *damped and forced incompressible 2d stochastic euler equation*, Journal on Uncertainty
845 Quantification, (under review).
- 846 [23] G. DA PRATO AND J. ZABCZYK, *Stochastic Equations in Infinite Dimensions*, Ency-
847 clopedia of Mathematics and its Applications, Cambridge University Press, 1992.
- 848 [24] A. DOUCET AND A. JOHANSEN, *A tutorial on particle filtering and smoothing: Fifteen*

- 849 *years later*, Handbook of Nonlinear Filtering, 12 (2009), pp. 656–704.
- 850 [25] G. EVENSEN, *Data assimilation: The ensemble Kalman filter*, Springer-Verlag, New-
851 york, 2006.
- 852 [26] L. FICK, Y. MADAY, A. T. PATERA, AND T. TADDEI, *A stabilized pod model for*
853 *turbulent flows over a range of reynolds numbers: Optimal parameter sampling and*
854 *constrained projection*, Journal of Computational Physics, 371 (2018), pp. 214–243.
- 855 [27] F. FLANDOLI, *The interaction between noise and transport mechanisms in PDEs*,
856 Milan Journal of Mathematics, 79 (2011), pp. 543–560.
- 857 [28] C. FRANZKE, T. O’KANE, J. BERNER, P. WILLIAMS, AND V. LUCARINI, *Stochastic*
858 *climate theory and modeling*, Wiley Interdisciplinary Reviews: Climate Change, 6
859 (2015), pp. 63–78.
- 860 [29] V. GENON-CATALOT, C. LAREDO, AND D. PICARD, *Non-parametric estimation*
861 *of the diffusion coefficient by wavelets methods*, Scandinavian Journal of Statistics,
862 (1992), pp. 317–335.
- 863 [30] G. GOTTWALD, D. CROMMELIN, AND C. FRANZKE, *Stochastic climate theory*, in
864 Nonlinear and Stochastic Climate Dynamics, Cambridge University Press, 2015.
- 865 [31] G. GOTTWALD AND J. HARLIM, *The role of additive and multiplicative noise in fil-*
866 *tering complex dynamical systems*, Proceedings of the Royal Society A: Mathematical,
867 Physical and Engineering Science, 469 (2013), p. 20130096.
- 868 [32] M. GUNZBURGER AND J. MING, *Optimal control of stochastic flow over a backward-*
869 *facing step using reduced-order modeling*, SIAM Journal on Scientific Computing, 33
870 (2011), pp. 2641–2663.
- 871 [33] D. HOLM, *Variational principles for stochastic fluid dynamics*, Proceedings of the
872 Royal Society of London A: Mathematical, Physical and Engineering Sciences, 471
873 (2015).
- 874 [34] J. HUNT, A. WRAY, AND P. MOIN, *Eddies, stream, and convergence zones in turbu-*
875 *lent flows*, Center for turbulence research report CTR-S88, (1988), pp. 193–208.
- 876 [35] S. KADRI HAROUNA AND E. MÉMIN, *Stochastic representation of the Reynolds*
877 *transport theorem: revisiting large-scale modeling*, Computers & Fluids, 156 (2017),
878 pp. 456–469.
- 879 [36] R. KRAICHNAN, *Small-scale structure of a scalar field convected by turbulence*, Physics
880 of Fluids (1958-1988), 11 (1968), pp. 945–953.
- 881 [37] H. KUNITA, *Stochastic flows and stochastic differential equations*, vol. 24, Cambridge
882 university press, 1997.
- 883 [38] T. KURTZ, *A limit theorem for perturbed operator semigroups with applications to*
884 *random evolutions*, Journal of Functional Analysis, 12 (1973), pp. 55–67.
- 885 [39] O. LE MAITRE, M. REAGAN, H. NAJM, R. GHANEM, AND O. KNIO, *A stochastic*
886 *projection method for fluid flow. II. random process*, Journal of Computational Physics,
887 181 (2002), pp. 9–44.
- 888 [40] C. LECLERCQ, F. DEMOURANT, C. POUSSOT-VASSAL, AND D. SIPP, *Linear iterative*
889 *method for closed-loop control of quasiperiodic flows*, Journal of Fluid Mechanics, 868
890 (2019), pp. 26–65.
- 891 [41] C. LEITH, *Atmospheric predictability and two-dimensional turbulence*, Journal of the
892 Atmospheric Sciences, 28 (1971), pp. 145–161.
- 893 [42] V. LUCARINI, R. BLENDER, C. HERBERT, F. RAGONE, S. PASCALE, AND
894 J. WOUTERS, *Mathematical and physical ideas for climate science*, Reviews of Geo-
895 physics, 52 (2014), pp. 809–859.
- 896 [43] J. L. LUMLEY, *Coherent structures in turbulence*, in Transition and turbulence, Else-
897 vier, 1981, pp. 215–242.

- 898 [44] A. MAJDA, I. TIMOFEYEV, AND E. VANDEN EIJNDEN, *A mathematical framework*
899 *for stochastic climate models*, Communications on Pure and Applied Mathematics, 54
900 (2001), pp. 891–974.
- 901 [45] E. MÉMIN, *Fluid flow dynamics under location uncertainty*, Geophysical & Astrophysical
902 Fluid Dynamics, 108 (2014), pp. 119–146.
- 903 [46] O. MÉTAIS AND M. LESIEUR, *Statistical predictability of decaying turbulence*, Journal
904 of the atmospheric sciences, 43 (1986), pp. 857–870.
- 905 [47] R. MIKULEVICIUS AND B. ROZOVSKII, *Stochastic Navier–Stokes equations for turbu-*
906 *lent flows*, SIAM Journal on Mathematical Analysis, 35 (2004), pp. 1250–1310.
- 907 [48] L. MITCHELL AND G. GOTTWALD, *Data assimilation in slow-fast systems using ho-*
908 *mogenized climate models*, Journal of the atmospheric sciences, 69 (2012), pp. 1359–
909 1377.
- 910 [49] A. NOUY, *Generalized spectral decomposition method for solving stochastic finite el-*
911 *ement equations: invariant subspace problem and dedicated algorithms*, Computer
912 Methods in Applied Mechanics and Engineering, 197 (2008), pp. 4718–4736.
- 913 [50] A. NOUY, *Recent developments in spectral stochastic methods for the numerical solu-*
914 *tion of stochastic partial differential equations*, Archives of Computational Methods in
915 Engineering, 16 (2009), pp. 251–285.
- 916 [51] A. NOUY, *A priori model reduction through proper generalized decomposition for solv-*
917 *ing time-dependent partial differential equations*, Computer Methods in Applied Me-
918 chanics and Engineering, 199 (2010), pp. 1603–1626.
- 919 [52] S. ORSZAG, *Analytical theories of turbulence*, Journal of Fluid Mechanics, 41 (1970),
920 pp. 363–386.
- 921 [53] G. PAPANICOLAOU AND W. KOHLER, *Asymptotic theory of mixing stochastic ordinary*
922 *differential equations*, Communications on Pure and Applied Mathematics, 27 (1974),
923 pp. 641–668.
- 924 [54] A. PAPAVALIOU, G. PAVLIOTIS, AND A. STUART, *Maximum likelihood drift estima-*
925 *tion for multiscale diffusions*, Stochastic Processes and their Applications, 119 (2009),
926 pp. 3173–3210.
- 927 [55] G. PAVLIOTIS AND A. STUART, *Parameter estimation for multiscale diffusions*, Jour-
928 nal of Statistical Physics, 127 (2007), pp. 741–781.
- 929 [56] L. PERRET, E. COLLIN, AND J. DELVILLE, *Polynomial identification of POD based*
930 *low-order dynamical system*, Journal of Turbulence, (2006), p. N17.
- 931 [57] C. PRÉVÔT AND M. RÖCKNER, *A concise course on stochastic partial differential*
932 *equations*, vol. 1905, Springer, 2007.
- 933 [58] V. RESSEGUIER, L. LI, G. JOUAN, P. DÉRIAN, E. MÉMIN, AND C. BERTRAND,
934 *New trends in ensemble forecast strategy: uncertainty quantification for coarse-grid*
935 *computational fluid dynamics*, Archives of Computational Methods in Engineering,
936 (2020), pp. 1–82.
- 937 [59] V. RESSEGUIER, E. MÉMIN, AND B. CHAPRON, *Geophysical flows under location*
938 *uncertainty, part I random transport and general models*, Geophysical & Astrophysical
939 Fluid Dynamics, 111 (2017), pp. 149–176.
- 940 [60] V. RESSEGUIER, E. MÉMIN, AND B. CHAPRON, *Geophysical flows under location*
941 *uncertainty, part II quasi-geostrophy and efficient ensemble spreading*, Geophysical &
942 Astrophysical Fluid Dynamics, 111 (2017), pp. 177–208.
- 943 [61] V. RESSEGUIER, E. MÉMIN, D. HEITZ, AND B. CHAPRON, *Stochastic modelling*
944 *and diffusion modes for proper orthogonal decomposition models and small-scale flow*
945 *analysis*, Journal of Fluid Mechanics, 826 (2017), pp. 888–917.
- 946 [62] V. RESSEGUIER, W. PAN, AND B. FOX-KEMPER, *Data-driven versus self-similar*

- 947 *parameterizations for stochastic advection by lie transport and location uncertainty*,
948 *Nonlinear Processes in Geophysics*, 27 (2020), pp. 209–234.
- 949 [63] T. SAPSIS AND A. MAJDA, *Blending modified Gaussian closure and non-Gaussian re-*
950 *duced subspace methods for turbulent dynamical systems*, *Journal of Nonlinear Science*,
951 23 (2013), pp. 1039–1071.
- 952 [64] T. SAPSIS AND A. MAJDA, *Statistically accurate low-order models for uncertainty*
953 *quantification in turbulent dynamical systems*, *Proceedings of the National Academy*
954 *of Sciences*, 110 (2013), pp. 13705–13710.
- 955 [65] T. SAPSIS AND A. MAJDA, *A statistically accurate modified quasilinear Gaussian clo-*
956 *sure for uncertainty quantification in turbulent dynamical systems*, *Physica D: Non-*
957 *linear Phenomena*, 252 (2013), pp. 34–45.
- 958 [66] P. SCHMID, *Dynamic mode decomposition of numerical and experimental data.*, *J.*
959 *Fluid Mech.*, 656 (2010), pp. 5–28.
- 960 [67] C. SOIZE AND C. FARHAT, *A nonparametric probabilistic approach for quantifying*
961 *uncertainties in low-dimensional and high-dimensional nonlinear models*, *International*
962 *Journal for Numerical Methods in Engineering*, 109 (2017), pp. 837–888.
- 963 [68] G. STABILE, F. BALLARIN, G. ZUCCARINO, AND G. ROZZA, *A reduced order varia-*
964 *tional multiscale approach for turbulent flows*, *Advances in Computational Mathemat-*
965 *ics*, (2019), pp. 1–20.
- 966 [69] G. STABILE, S. HIJAZI, A. MOLA, S. LORENZI, AND G. ROZZA, *POD-Galerkin*
967 *reduced order methods for CFD using finite volume discretisation: vortex shedding*
968 *around a circular cylinder*, *Communications in Applied and Industrial Mathematics*,
969 8 (2017), pp. 210–236.
- 970 [70] G. STABILE AND G. ROZZA, *Finite volume POD-Galerkin stabilised reduced order*
971 *methods for the parametrised incompressible navier–stokes equations*, *Computers &*
972 *Fluids*, 173 (2018), pp. 273–284.
- 973 [71] D. TORLO, F. BALLARIN, AND G. ROZZA, *Stabilized weighted reduced basis meth-*
974 *ods for parametrized advection dominated problems with random inputs*, *SIAM/ASA*
975 *Journal on Uncertainty Quantification*, 6 (2018), pp. 1475–1502.
- 976 [72] A. TOWNE, O. SCHMIDT, AND T. COLONIUS, *Spectral proper orthogonal decom-*
977 *position and its relationship to dynamic mode decomposition and resolvent analysis*,
978 *Journal of Fluid Mechanics*, 847 (2018), p. 821–867.
- 979 [73] Z. WANG, I. AKHTAR, J. BORGGAAARD, AND T. ILIESCU, *Proper orthogonal de-*
980 *composition closure models for turbulent flows: a numerical comparison*, *Computer*
981 *Methods in Applied Mechanics and Engineering*, 237 (2012), pp. 10–26.
- 982 [74] X. XIE, M. MOHEBUJAMAN, L. G. REBHOLZ, AND T. ILIESCU, *Data-driven fil-*
983 *tered reduced order modeling of fluid flows*, *SIAM Journal on Scientific Computing*, 40
984 (2018), pp. B834–B857.
- 985 [75] Y. YANG AND E. MÉMIN, *High-resolution data assimilation through stochastic subgrid*
986 *tensor and parameter estimation from 4D EnVar* , *Tellus A: Dynamic Meteorology and*
987 *Oceanography*, 69 (2017), p. 1308772.