AAPG2022	CONMAN		JCJC
Coordinated by	Jean-Christophe LOISEAU	48 months	238 000 €
CE46 – Modèles numériques, simulation, applications			

Data-driven CONtrol and MANifold interpolation

Summary table of persons involved in the project:

Partner	Name	First name	Position	Role	Involvement
Arts et Métiers	**LOISEAU**	**Jean-Christophe**	**Maître de Conférences**	**Principal Investigator**	**20 PM***
Arts et Métiers	MERLE	Xavier	Maître de Conférences	Task Leader	6PM*
Arts et Métiers	ROBINET	Jean-Christophe	Professeur des Universités	Advisor	2PM*

* Being *Enseignants-Chercheurs*, all participants can spend at most 50% of their time to the project, the remaining 50% being for teaching activities. Dedicating all of their available research time to the project thus amounts to a maximal involvement of 24 person-month (PM), the project running over a duration of 48 months.

Any changes that have been made in the full proposal compared to the pre-proposal The requested funds have been slightly increased (+3000€, +1.3% compared to the pre-proposal) to account for the administrative management costs. Other than that, no specific change has been made that could modify the eligibility of the current project since drafting the pre-proposal.

---

1 Proposal's context, positioning and objective(s)

Applications of closed-loop flow control have epic proportions: drag reduction, lift increase, mixing enhancement, or noise mitigation. It is however challenged by strong nonlinearities, partial state information, parametric dependencies, or time delays. The present project aims at tackling one of these problems, the parametric dependencies. It forms the first step in a long research program crafted by the principal investigator.

1.1 Objectives and research hypothesis

Deep learning has a transformative impact in many fields. Current state-of-the-art for controlling complex systems is formed by *deep reinforcement learning* (DRL) methods. The most prominent examples include Alpha Go Zero [57] for board games or Alpha Star [63] for the real-time-strategy game StarCraft II. Recently, DeepMind has also been successful in controlling the nuclear fusion plasma in a tokamak [20]. DRL applied to flow control is however a burgeoning field with contrasted successes, mostly on the cylinder flow [46, 59, 23]. A literature review of applications in fluid dynamics is provided in [24]. Despite these achievements, DRL still is too expensive, too complicated or not understood well enough for practical applications. In situations of industrial interest, the gold standard is formed by methods from linear optimal, robust or model-predictive control. While having a long and well-established history, these techniques still suffer from certain limitations, most notably when parametric dependencies need to be accounted for. Recent advances in data-driven approximations of high-dimensional linear operators and differential geometry may provide a well-grounded theoretical framework to overcome these limitations. This will be explored and benchmarked in the present project. Although the methodology proposed is fairly general, a particular emphasis will be given to control-oriented reduced-order models.

Objectives of the project and research hypothesis Many practical models in engineering sciences belong to the class of *generalized linear models*

$$\boldsymbol{y} = \mathbf{K}\mathbf{x} + \varepsilon$$

with $\boldsymbol{y} \in \mathbb{C}^m$, $\boldsymbol{x} \in \mathbb{C}^n$ and $\boldsymbol{K} \in \mathbb{C}^{m \times n}$ a linear mapping from $\boldsymbol{x}$ to $\mathbf{y}$. In system identification, it includes for instance OKID where $\boldsymbol{y}$ is the response of the unknown linear system, $\boldsymbol{K}$ a Toeplitz matrix constructed from the input sequence and $\boldsymbol{x}$ is the vector of unknown Markov parameters of the system. Similarly, *linear stochastic estimation* or the *linear deconvolution problem* can also be cast as generalized linear models.

Of interest to us are situations where $\boldsymbol{K}$ is unknown. Given training pairs $(\boldsymbol{x}_i, \boldsymbol{y}_i)$, an ordinary least-squares regression can be formulated to identify $\boldsymbol{K}$. Yet, for typical engineering problems $\boldsymbol{x}$ and/or $\boldsymbol{y}$ are high-dimensional vectors. Hence, we are unlikely to have sufficient data to obtain a good statistical estimate of $\boldsymbol{K}$. It can however be assumed to be low-rank, an assumption often verified for high-dimensional dissipative systems. A good estimate can be obtained by solving the following rank-constrained problem

$$\begin{aligned} \underset{\boldsymbol{P},\boldsymbol{Q}}{\text{minimize}} \quad & \|\boldsymbol{M}^{1/2}(\boldsymbol{Y} - \mathbf{PQ}^H\boldsymbol{X})\|_F^2 \\ \text{subject to} \quad & \boldsymbol{P}^H\boldsymbol{MP} = \boldsymbol{I}_r \end{aligned}$$

where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are data matrices, $\boldsymbol{M}$ a positive-definite mass matrix, and r the rank of the desired approximation. Although non-convex, these rank-constrained problems admit a closed-form solution [19, 26, 38]. Most modal decompositions also fall into this framework. POD and PCA are recovered if $\boldsymbol{X} = \boldsymbol{Y}$ and $\boldsymbol{P} = \boldsymbol{Q}$. Similarly, DMD [54, 55] is obtained for $\boldsymbol{y}_i = \boldsymbol{x}_{i+1}$. **One major objective of this project is thus to unify most data-driven linear models (i.e. modal decompositions and system identification techniques) into this framework**.

Another key objective is to explore the combination of this reduced-rank regression framework with recent advances in differential geometry to propose a methodology for parameterizing data-driven linear models. Low-rank modal decompositions such as POD, BPOD or DMD are often used to formulate a Galerkin expansion of the state vector of the system. A reduced-order model can then be obtained by projecting the governing equations onto the span of these modes, or identified from time-series of the modes' amplitudes as in Galerkin regression [39, 38, 14]. In either case, the resulting low-order model has a limited dynamical range, e.g. its validity as the Reynolds number varies is very narrow. This results from the inability of these decompositions to account for the spatial deformation of the modes as the parameters of the system change. Similar drawbacks are encountered for classical system identification techniques such as ERA [32] or N4SID [61]. This is particularly detrimental in feedback control applications. As an example, the performances of the low-dimensional LQG controller designed for the canonical cylinder flow at $\mathrm{Re} = 50$ rapidly degrade as the Reynolds number is varied by even a few percent. Accounting for these parametric dependencies is thus of paramount importance.

One benefit of the reduced-rank regression framework is that the singular value decomposition of the matrix $\boldsymbol{K} = \boldsymbol{U\Sigma V}^T$ can easily be computed. Given models obtained at different operating conditions μ_i, i.e. $\boldsymbol{K}_i = \boldsymbol{K}(\mu_i)$, the matrices $\boldsymbol{U}_i$, Σ_i and $\boldsymbol{V}_i$ can be interpolated at a new operating condition. Yet, $\boldsymbol{U}_i$ and $\boldsymbol{V}_i$ being orthogonal matrices, this interpolation cannot however proceeds in an entry-wise way. To preserve this fundamental structure, it needs to occur on the Grassmann or Stiefel manifolds [69]. These are equipped with a natural notion of distance and gradient which can be leveraged to design efficient sampling strategy in the parameter space. High-order interpolation schemes on these manifolds have moreover been proposed by Ralf Zimmerman [69]. **Combining reduced-rank regressions and tools from differential geometry, another key objective is thus to develop a natural framework for parameterizing data-driven linear models**.

Finally, an aspect often overlooked in data-driven models is their robustness to the training data and their sensitivity to the operating conditions. A first step in this direction was achieved by Hay *et al.* [25], deriving the parametric sensitivity of POD modes and POD eigenvalues. Including the sensitivity of the modes to the projection basis in a Galerkin projection procedure was shown to increase the range of validity of the reduced-order model. **POD being a special instance of the more general framework presented earlier, one of the final aim of this project is thus to generalize the sensitivity analysis from [25] to all models falling into the reduced-rank regression framework**. Regarding the model's robustness, a methodology based on ensembling and bagging was proposed by Sashidhar & Kutz [53], specifically for DMD. Their ensembling procedure however relies on simple averaging not necessarily consistent with the low-rank structure of the model. Note that each model generated during this ensembling can be associated with a point on the Stiefel manifold. Once again, this manifold provides a natural framework in which to formulate the different statistical objects needed to quantify the models uncertainties while preserving their fundamental low-rank structure. **Leveraging once again the properties of the Stiefel manifold, the last theoretical objective of this project is thus to propose a well-grounded methodology to equip data-driven linear models with uncertainty quantification capabilities**. This last aspect can be of utmost importance when trying to control a nonlinear system with an otherwise imperfect reduced-order model.

Technical barriers Data-driven models have emerged as a powerful paradigm over the past decade. However, explicitly accounting for their parametric dependencies in a consistent, rigorous, and computationally efficient way still is challenging. Yet, exploiting properties of the Grassmann and Stiefel manifolds (the mathematical objects at the core of this project), Zimmermann [67, 68, 69] has recently proposed a set of methods to perform high-order interpolation on these manifolds. These theoretical and algorithmic developments served as the basis for two proofs of concept by the principal investigator and collaborators. In [40], the Grassmann manifold was used to parameterize the deformation of the instability modes into the POD ones for the canonical two-dimensional cylinder flow. In [52], parameterization of DMD-based state observers for an airfoil in transonic buffeting conditions with a varying angle of attack was conducted on the Stiefel manifold. Despite the novelty of the mathematical approach undertaken in this project, very few barriers, whether technical or theoretical, have thus been identified. The few that have relate to the proper definition of statistical quantities on the Stiefel manifold involved for uncertainty quantification. However, these potential barriers concern only a single task in this project.

Expected results Focusing on data-driven models for control purposes, the main theoretical and numerical developments to be conducted in this project are

- Implementation of current state-of-the-art techniques for linear optimal control of high-dimensional systems [51, 9, 56] to establish baseline performances against which to compare the data-driven models developed later on.
- Development of a theoretical framework based on reduced-rank regression to unify most modal decomposition and linear system identification techniques used in our community at large.
- Combine this theoretical framework with recent advances in differential geometry to propose a well-grounded methodology for parameterizing data-driven linear models. It also includes the derivation of the parametric sensitivity of the different models to enable high-order interpolation for the parameterization.
- Exploit ensembling and bagging techniques from statistical learning to obtain more robust and more generalizable data-driven models as well as providing some elements to quantify their prediction uncertainties.

Each of these major actions constitutes one work package described later in the document.

Regarding the outcomes and results expected by the end of the project, different objectives are being pursued.

- Despite the wide adoption of data-driven models, even more so now with the large number of publications using deep learning techniques, very few quantitative benchmarks on well described test cases and reference implementations are available. Using the two-dimensional cylinder and shear-driven flows as examples, one of our objectives is to rigorously establish these baseline performances and provide open-access implementations of the best performing models for other researchers to objectively compare the performances of their new approaches.
- As a proof of concept, the control strategy proposed in this work (including the parameterization and uncertainty quantification) will be applied to a three-dimensional flow past a slender body in a chaotic regime [49, 11, 50]. The objective is to showcase the capabilities of the proposed methodology to a flow configuration of both academic and industrial interest.
- Finally, all of the numerical developments will be integrated either in `nekStab`, an open-source toolbox for large-scale instability and bifurcation analysis for the spectral element solver `Nek5000`, or in an another open-source toolbox dedicated to data-driven linear modeling that will be developed during the project.

Theoretical and numerical results will be published in top-tier applied mathematics or fluid dynamics journals, while the different toolboxes will have a companion paper published in the Journal of Open Source Software.

1.2 Originality and relevance in relation to the state of the art

Parameterizing reduced-order models, even for linear systems, is a long standing problem in computational sciences. In fluid dynamics, one major reason is the continuous deformation of the coherent structures forming the backbone of the reduced representation as the operating conditions change. One approach is to collect snapshots of the dynamics of the system at different points in the parameter space. After stacking all these snapshots into a data matrix, a low-rank basis globally valid over the whole range of sampled operating conditions is typically obtained using Proper Orthogonal Decomposition. It is then used to build a parameterized reduced order model based on a Galerkin projection of the governing equations. This approach is typically known as *Reduced Basis*.

In this greedy approach, the dimension of the reduced basis is closely related to the concept of the *Kolmogorov n-width*. It provides an upper bound for the dimension of the Euclidean space needed to embed the solution manifold for a given tolerance. Yet, this manifold might be of much lower topological dimension than its embedding dimension. Reduced basis techniques thus introduce artificial degrees of freedom in the reduced-order model to compensate for the mismatch between the topological dimension of the solution manifold and the dimension of its embedding subspace. Introducing these artificial degrees of freedom may however have a detrimental effect on the stability and accuracy of the reduced-order model as discussed by Lee & Carlberg [37] or more recently by the principal investigator in [14]. Instead of using a global basis, an alternative is to construct multiple local ones (i.e. one at each point sampled in the parameter space) along with an interpolant. Doing so, the dimension of the low-rank basis (and hence of the reduced-order model) tends to be closer to the topological dimension of the solution manifold. A first step in this direction was achieved by Amsallem & Farhat

[1, 2] using the so-called Grassmann manifold for projection-based reduced-order models for aeroacoustic applications. Being a Riemannian manifold, one major benefit of using the Grassmann manifold is that it is equipped with a natural notion of distance. The distances between low-rank bases associated to different operating conditions can be used to determine the next point to be sampled in the parameter space or to construct the interpolant. Given only two points on this manifold, this interpolation reduces to *subspace angle interpolation*. If more points are available, a higher-order interpolant can be obtained using e.g. Lagrange interpolation. As discussed in Amsallem & Farhat [1, 2], this interpolant needs to be constructed within the tangent subspace of the manifold. However, one drawback of the Grassman manifold is that it is primarily concerned with the span of the low-rank basis rather the basis itself. Two orthogonal matrices $\boldsymbol{X}$ (e.g. POD basis) and $\boldsymbol{Y}$ (e.g. DMD basis) with the same span correspond to the same point on the Grassmann manifold. As such, physical interpretation of the low-dimensional state vector might be lost upon interpolation of the low-rank bases.

If the interpretability of the low-rank basis is equally important, two alternatives can be considered. In the first one, the interpolation still is conducted on Grassmann manifold. After having interpolated the low-rank bases, these are then re-aligned with respect to a reference point by formulating an *orthogonal Procustes problem*. A second alternative, advocated for by Zimmermann [67, 68, 69] is to replace the Grassman manifold by the Stiefel one. Two orthogonal matrices $\boldsymbol{X}$ (e.g. POD basis) and $\boldsymbol{Y}$ (e.g. DMD basis) with the same span no longer correspond to the same point on the Stiefel manifold, unless $\boldsymbol{X} = \boldsymbol{Y}$ (compared to $\text{span}(\boldsymbol{X}) = \text{span}(\boldsymbol{Y})$ for the Grassmann manifold). For theoretical reasons, interpolation on the Stiefel manifold is more computationally intensive than on the Grassmann one. Yet, if the gradient of the low-rank bases with respect to the operating conditions is available, an accurate cubic Hermite interpolant can be constructed [67, 68, 69], leading to more accurate estimates for new operating conditions. In 2015, Benner, Gugercin and Wilcox [7] published a survey of projection-based model reduction methods for parametric dynamical systems with an exhaustive list of references, including both the reduced basis approach and the matrix manifold interpolation techniques. A slightly more recent literature review on the subject has also been conducted by Zimmermann, Peherstorfer and Wilcox [70].

Compared to projection-based reduced-order models, the literature on parameterized data-driven models for dynamical systems is more recent and much scarcer. Despite the well-grounded theoretical framework described in the previous paragraph, most of the works on parameterized data-driven reduced-order models have turned their attention to deep learning techniques. A large fraction of this body of work is moreover dedicated to implicitly parameterizing turbulence models, whether RANS-like or subgrid-scale models for LES. For relatively simple dynamical systems, Kalia et al. [33] have used auto-encoder networks to identify parameterized normal forms based on simulated data of relatively simple dynamical systems. Based on a probabilistic point of view, Morton et al. [44] have proposed to use variational auto-encoders conditioned on the parameters of the systems to construct sequential generative models of fluid flows. They illustrated their approach using the flow past two side-by-side co- or counter-rotating cylinders with a two-dimensional parameter space.

The wealth of deep learning models published recently in the fluid dynamics literature is only paralleled by the rising popularity of *Dynamic Mode Decomposition* [54, 55] and its variants. These include *Exact DMD* [60], Extended DMD [64], online DMD [66], compressed DMD [12, 6], DMD with control [48], Recursive DMD [45], optimized DMD [3, 53], Forward-Backward DMD [18], Total Least Squares DMD [28], high-order DMD [36], ioDMD [8], sparsity-promoting DMD [31], Bayesian DMD [58], multi-resolution DMD [34], randomized DMD [10, 22], kernel DMD [65, 27], low-rank DMD [26], consistent DMD [4], physics-informed DMD [5], and a few more. Despite this exhaustive list, none of these groups have addressed the problem of parameterizing these data-driven models. To the best of our knowledge, a similar remark about the lack of parameterization holds true for most linear system identification techniques proposed in the literature. These include EigenRealization Algorithm [32], OKID [62], N4SID [61], MOESP [30], CVA [35] and others.

Leveraging the theoretical framework being developed for parameterizing projection-based reduced-order models, the aim of the CONMAN project is to extend it to data-driven models, with a particular emphasis on control-oriented applications. As discussed in §1.1, a first step to achieve this goal is to unify most data-driven linear and system identification techniques in the common reduced-rank regression framework. Preliminary work in this direction is currently being conducted by the principal investigator of the project in collaboration with Steven Brunton from the University of Washington (Seattle, USA). Using the Grassmann and Stiefel manifold interpolation techniques developed by Amsallem [1, 2] and Zimmermann [67, 70], a second step is to develop a priori error estimates and parameter space sampling strategies to reduced the offline computational cost of building these parameterized models. Preliminary results have already been obtained for the transients and post-transients dynamics of the canonical cylinder flow at $\text{Re} = 100$ in [40] and for an airfoil in transonic buffeting conditions with varying angle of attack in [52] by the principal investigator and his collaborators. Finally, following Sashidhar & Kutz [53], the project also aims at equipping these data-driven models with uncertainty quantification capabilities. This is of utmost importance in practical applications, particularly when imperfect models are used for state- or output-feedback control. Focusing on fluid dynamics applications, these data-driven control-oriented models will be tested and validated on two standard benchmarks, namely the canonical cylinder and shear-driven cavity flows with varying Reynolds numbers. To showcase the utility and performances of these models on a realistic flow configuration, the flow past a three-dimensional slender body in a chaotic regime will be considered. This particular flow configuration has already been studied by Rigas et al. in [49, 11, 50, 15]. Along with the unification, parameterization and uncertainty quantification methodologies, we expect this project to provide the much needed, yet currently lacking, benchmark performances against which to compare more advanced control strategies, most notably those based on deep learning techniques.

1.3 Methodology and risk management

The project runs for four years. The main participants are Jean-Christophe Loiseau (JCL), a post-doctoral researcher, and a Ph.D. candidate. Professor Robinet, head of DynFluid, will provide feedback about project management and student supervision, while uncertainty quantification aspects will be investigated with the help of Xavier Merle. The project also benefits from ongoing collaborations with Professor Steve Brunton. Moreover, JCL plans to defend his *Habilitation à diriger des recherches* shortly after the start of the project. This will ease the supervision of the Ph.D. candidate. If the defense is delayed or the supervision by the principal investigator is denied by the Arts et Métiers doctoral school, Professor Robinet will assure the administrative supervision of the Ph.D. student.

1.3.1 Overview of the project's organization

The project is structured in four technical work packages (WP1 to WP4) and two organizational ones (WP0 and WP5). The organizing tasks are devoted to project coordination (WP0) and scientific communication and dissemination (WP5). A Gantt diagram describing the planning of the project over the course of the four years is provided below.

WP	Task	Planning			
		Year 1	Year 2	Year 3	Year 4
x	Post-doctoral researcher				
	PhD candidate				
1	**Task 1.1** – Adjoint of the Direct-Adjoint				
	Task 1.2 – Adjoint of the Adjoint-Direct				
	Task 1.3 – Balanced Proper Orthogonal Decomposition				
	Task 1.4 – Optimal actuator and sensor placement				
2	**Task 2.1** – Unifying framework for data-driven models				
	Task 2.2 – Development of an open-source toolbox				
	Task 2.3 – Benchmarking data-driven models				
	Task 2.4 – Data-driven sensors and actuators placement				
3	**Task 3.1** – Parameterization of BPOD				
	Task 3.2 – Parameterization of ADA and AAD				
	Task 3.3 – Parameterization of data-driven models				
4	**Task 4.1** – Sensitivity of data-driven models				
	Task 4.2 – UQ on the Stiefel manifold				
	Task 4.3 – Application to a 3D slender body flow				

Table 1. Gantt diagram associated with the four technical work packages for the duration of the project.

In the rest of this document, the project is assumed to start on January 1st 2023. Contacts with possible candidates for the post-doctoral position will be established during the early summer 2022. Given the academic calendar, a successful Ph.D. candidate will be recruited in September 2023. The Ph.D. thesis duration covers the last three years of the project.

1.3.2 Description by work package

In this section, each work package (objectives, tasks, tentative duration, and outcomes) is described. A clear overview of the different milestones and deliverable is also given at the end.

WP 0 – Project management

Leader	Jean-Christophe Loiseau
Contributing participants	All participants
Duration	4 years

This work package deals with the coordination of the project and its management. Its main goal is to ensure good coordination and transfer of knowledge between the participants. All participants will meet twice a month to discuss progress made and difficulties encountered. Thrice a year, meetings with Jean-Christophe Robinet will take place. Their goal is to ensure that the project does not deviate from its tracks and discuss any problem related to management or human resources. The main deliverable will be the biannual activity and final reports.

WP 1 – Baseline models

Leader	Jean-Christophe Loiseau
Contributing participants	Principal Investigator, Post-doctoral researcher
Duration	12 months

The parameterization strategy proposed herein applies to a wide range of linear models. Yet, our main focus is on control-oriented data-driven models. To assess the performances and robustness of these models, reference data must be obtained. Running over the first year of the project, this WP deals with the implementation of state-of-the-art linear methods for flow control. These developments will happen within `nekStab`, an open-source toolbox for large-scale stability and bifurcation analysis. Relying on the spectral element solver `Nek5000`, it is being developed by the principal investigator. Of interest for this work package are Nek5000's capabilities in solving the direct and adjoint linearized Navier-Stokes equations and the coupling with `nekStab`'s large-scale eigenvalue and singular value solvers.

Consider a linear time-invariant dynamical system

$$
\begin{aligned}
\dot{\boldsymbol{x}} &= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \\
\boldsymbol{y} &= \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u}
\end{aligned}
$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is the state vector, $\boldsymbol{u} \in \mathbb{R}^p$ represents the p input to the system, and $\boldsymbol{y} \in \mathbb{R}^q$ the q output. A feedback control law of the form $\boldsymbol{u} = -\mathbf{K}\mathbf{x}$ can stabilize the otherwise unstable dynamics. For *linear quadratic regulators* (resp. *linear quadratic estimator*), synthesizing $\boldsymbol{K} \in \mathbb{R}^{p \times n}$ (resp. $\boldsymbol{L} \in \mathbb{R}^{n \times q}$) requires the solution of an algebraic Ricatti equation. Due to the sheer size of the discretization in fluid dynamics, using a direct Ricatti solver is however computationally intractable, even for moderately complex flow configurations. Two main strategies [51, 9, 56] have been proposed to overcome this limitation. The implementation of these strategies forms the main tasks of this work package.

Task 1.1 – Adjoint of the Direct-Adjoint (ADA)

Assigned to	Post-doctoral researcher (100%)
Duration	4 months
Risk	Low

In a feedback control problem, synthesizing the control gain $\boldsymbol{K} \in \mathbb{R}^{p \times n}$ requires the solution of an algebraic Ricatti equation. A first step toward high-dimensional systems was achieved by Bewley *et al.* [9], introducing the *Adjoint of the Direct-Adjoint*. Originally limited to single input-single output (SISO) systems or decentralized control schemes, ADA has been extended by Semeraro & Pralits [56] to accommodate for multiple input and multiple output. The aim of this task is to implement the ADA algorithm [56] into `nekStab`. All the computational routines are already available. Hence, the implementation should be relatively fast. In the rest of this project, controllers synthesized using ADA, either with the LQR (a special case of $\mathcal{H}_2$ synthesis) or robust control ($\mathcal{H}_\infty$ synthesis) paradigms, will serve as our baseline references to evaluate the performances of the data-driven models developed in WP2 to WP4.

Task 1.2 – Adjoint of the Adjoint-Direct (AAD)

Assigned to	Post-doctoral researcher (100%)
Duration	4 months
Risk	Low

For optimal performances, feedback control requires full-state information. Yet, in practical applications, only limited sensor measurements are available and the state $\boldsymbol{x}$ of the system needs to be estimated. As for the control synthesis, the Kalman gain $\boldsymbol{L} \in \mathbb{R}^{n \times q}$, enabling the estimation of the state $\boldsymbol{x}$ from the limited measurements $\boldsymbol{y}$, is solution to a Ricatti equation. Following [56], a similar approach, the *Adjoint of the Adjoint-Direct*, can be used for high-dimensional systems. In parallel to Task 1.1, the goal of the present one is the implementation of AAD into `nekStab`. As for ADA, all the computational components are already available to the project, and state estimators obtained by AAD will serve as our baseline references.

Task 1.3 – Balanced Proper Orthogonal Decomposition

Assigned to	Principal Investigator (100%)
Duration	2 months
Risk	None

BPOD [51] aims at approximating *balanced truncation* [43], otherwise intractable for large-scale systems. BPOD does not try to approximate directly the high-dimensional LQR and LQE gain matrices. Instead, it first builds a reduced-order model of the system based on snapshots from direct and adjoint impulse response simulations. Relying on a low-rank approximation of the cross Gramian, BPOD ensures that the constructed reduced-order model captures most of the input-output properties of the original high-dimensional system. The reduced-order model being of modest dimension, a standard Ricatti solver is then used to compute the linear quadratic regulator and estimator for feedback control purposes.

In its current form, `nekStab` already contains all the tools and routines needed for BPOD. Hence, the implementation of BPOD will be fast. Because of its good theoretical and computational properties, combined with its relative ease of implementation, BPOD will not only be one of the baseline models for comparisons later in the project but also serves as a fallback strategy if Tasks 1.1 and 1.2 are delayed due to unforeseen difficulties.

Task 1.4 – Optimal sensor and actuator placement

Assigned to	Principal investigator (50%) and Post-doctoral researcher (50%)
Duration	4 months
Risk	Moderate

In the previous tasks, the matrix $\boldsymbol{B}$ describing the actuators and the matrix $\boldsymbol{C}$ describing the available measurements are fixed *a priori*. Actuators and sensors are determined prior to reduced-order modeling, either based on technical constraints or physical knowledge about the system. If very few sensors or actuators can be afforded, fine-tuning their locations or parameters can provide the extra authority needed to control the system at a reduced cost. It can be formulated as an optimization problem and solved using various techniques, e.g. gradient-based optimization [16], evolutionary algorithms [21] or simple heuristics [29]. To our knowledge, the current state-of-the-art method has been proposed by Chen & Rowley [16] using gradient-based minimization of the $\mathcal{H}_2$ norm of the controlled system. To date, this approach has only been illustrated on relatively simple one-dimensional partial differential equations for which the Ricatti and Lyapunov equations involved in the computation of the gradient can be solved using standard direct solvers.

The aim of this task is thus to adapt the framework proposed in [16] to a high-dimensional setup. While matrix-free/time-stepper solvers for the Ricatti equations will already have been implemented in Tasks 1.1 and 1.2, similar developments are needed for the Lyapunov equation. These implementations will provide the basic routines to evaluate the gradient of the cost function. Regarding the optimizer, Rowley & Chen [16] suggest using a conjugate gradient method. Even though a similar strategy may be employed eventually, our first implementation will rely on a sequential quadratic programming (SQP) relaxation with a BFGS solver. SQP can easily accommodate for constraints on the locations of both sensors and actuators. A dedicated paper presenting the algorithm and its performances on the different test cases will be written and submitted to one of the journals mentioned in WP5.

Test cases and applications All the algorithms developed in this work package and the subsequent ones will be tested and evaluated on two standard benchmarks, namely the canonical cylinder and shear-driven cavity flows. In both cases, the controllers and estimators will be synthesized based on a linearization of the Navier-Stokes around the actual fixed point of the systems and the time-averaged solution of the nonlinear equations. While the latter is more debatle from a mathematical point of view, it provides a more realistic setup, unstable base flows being hardly achievable in actual experiments. It also is a first step toward the applications to turbulent flows. In the second half of the project, application to a three-dimensional slender body at moderately high Reynolds numbers, similar to the one in [49, 11, 50], will be considered. Being of high industrial and aerodynamics relevance, it will serve as a proof of concept of the applicability of the methodologies developed during this project.

WP 2 – Data-driven linear modeling

Leader	Jean-Christophe Loiseau
Contributing participants	All participants
Duration	14 months

This second work package forms one of the main contributions of this project. It covers most of the first year of the PhD candidate. Given two sets of data, $\boldsymbol{X} \in \mathbb{C}^{n \times k}$ and $\boldsymbol{Y} \in \mathbb{C}^{m \times k}$, most data-driven linear models can be formulated as the reduced-rank regression problem mentioned in §1. Of particular interest here is that many linear system identification techniques (e.g. ERA [32], N4SID [61], DMDc [48], or ioDMD [8]) can also be cast into this framework.

Task 2.1 – Unifying framework for data-driven linear modeling

Assigned to	Principal Investigator (20%) and PhD candidate (80%)
External collaborator	Steven Brunton
Duration	4 months
Risk	Low

Based on an exhaustive literature review, the goal of this task is to unify the most widely used data-driven linear modeling techniques into a single framework. In the statistics community, a first step toward this aim was done by De la Torre [19]. Developing such a unifying theoretical framework will provide a better understanding of the mathematical limitations and statistical interpretation of this class of models when applied to large-scale dynamical systems. Being the first task to be tackled by the PhD candidate, it will give them the opportunity to get up to speed with the current literature. It will also be the occasion for the PhD candidate to implement these models on toy problems, getting familiar with both the theoretical and numerical aspects (e.g. regularized regression, truncated singular value decomposition, etc).

Task 2.2 – Development of an open-source toolbox

Assigned to	Principal Investigator (20%), PhD candidate (50%), and Post-doctoral researcher (30%)
Duration	6 months
Risk	Low

In this second task, the first version of the toolbox will be developed, implementing all of the models unified in Task 2.1. It will form the foundation of the developments to come in WP3 and WP4. The core of this toolbox will be written in Julia. Getting inspiration from `PySR` [17], the toolbox will also come with a Python front-end for increased usability. It will moreover expose an API consistent with `scikit-learn` [47], the most widely used package for Machine Learning, to benefit from its wide ecosystem. The toolbox will be released under the MIT license, and a public GitHub repository and website (hosted using GitHub pages) will be created. A paper presenting the toolbox will be submitted for publication in the Journal of Open Source Software.

Task 2.3 – Benchmarking data-driven models

Assigned to	PhD candidate (60%) and Post-doctoral researcher (40%)
Duration	6 months
Risk	Low

The aim of this task is to benchmark the performances of the data-driven models. These results will be compared to the performances of the high-fidelity models developed in WP1. With practical situations in mind, emphasis will be given to models that do not explicitly need the adjoint solver or prior knowledge of the governing equations. For most models (e.g. ERA, N4SID, etc), an extensive body of literature already exists and no new results are expected. The goal is to set the reference performances against which parameterized versions of the models will be compared in WP3. For models based on variants of DMD (e.g. DMDc [48] or ioDMD [8]), the situation is different. It has been shown in [38, 26, 5] that recasting DMD into the reduced-rank regression framework improves the model's accuracy and generalizability. Similar improvements are expected for the control-oriented versions of DMD. No such results are currently available in the literature. By the end of this task, preliminary results regarding the control of the flow past the three-dimensional slender body in chaotic regime studied in [50] will be produced.

Task 2.4 – Data-driven optimal sensors and actuators placement

Assigned to	Principal Investigator + PhD candidate
External collaborator	Krithika Manohar
Duration	4 months
Risk	Moderate

This last task echoes Task 1.4. The optimization problem in Task 1.4 is a formidable task. To be useful in practical situations, it however needs good *a priori* estimates of the optimal sensors and actuators. Yet, selecting *a priori* these sensors and actuators from a set of candidates is a rapidly intractable combinatorial problem. Recently, Manohar *et al.* [42] have proposed a greedy algorithm to tackle this issue. Based on a balanced realization using the whole set of possible sensors and actuators, the best candidates are selected sequentially while ensuring that the performances of the model are gracefully degraded. As for [16], this strategy has only been applied to relatively simple one-dimensional partial differential equations. This task has two main objectives. First, it aims at showcasing the applicability of the data-driven approach to sensor and actuator placement by Manohar *et al.* [42] to high-dimensional systems resulting from the discretization of the Navier-Stokes equations. Once a good set of sensors and actuators has been selected, their positions and parameters will be fine-tuned using the optimization problem from Task 1.4. This last step aims at restoring part of the control authority lost when reducing the number of sensors and actuators available.

WP 3 – Parameterization and interpolation on the Grassman and Stiefel manifolds

Leader	Jean-Christophe Loiseau
Contributing participants	All participants
Duration	12 months

This third work package constitutes the major contribution of this project: the parameterization of data-driven models with respect to the operating conditions. For simplicity, parameterization with respect to a single parameter (e.g. the Reynolds number) will be considered, albeit the proposed methodology is more general. Given training data at different operating conditions, a data-driven linear model can be identified for each sampled point in the parameter space using the methodology developed in WP2. Yet, interpolating the models matrices in an entry wise fashion for a new operating condition cannot be used as it would not preserve the underlying low-rank structure of the problem. This structured interpolation needs to take place on a particular matrix manifold, either the Grassman or Stiefel manifold. Because they are Riemannian manifolds, both of these are equipped with an intrinsic notion of distance between the models.

Based on prior theoretical works [1, 2, 67, 70, 68, 69], this work package aims at exploring the possibility to use these tools from differential geometry to obtain parameterized data-driven linear models. First results on an airfoil in transonic buffeting conditions with varying angle of attack have already been obtained by the principal investigator and his collaborators in [52]. Our attention will be focused on two particular points:

- How to properly sample the parameter space for the particular problem of control-oriented models ?

- Defining *a priori* error estimates for the accuracy of the interpolated models.

As for WP2, the performances of the parameterized data-driven models will be benchmarked and compared against those of the high-fidelity models. The different algorithms for the parameterization will be implemented in the open-source toolbox. This work package mostly covers the third year of the project, corresponding to the second year of the PhD thesis.

Task 3.1 – Parameterization of BPOD on the Grassman manifold

Assigned to	Principal Investigator
Duration	3 months
Risk	Low

This task follows the work of David Amsallem [1, 2] on parameterized reduced-order modeling for aero-elastic applications. Given a parametric high-fidelity linear model

$$\begin{aligned} \dot{\boldsymbol{x}} &= \mathbf{Ax} + \mathbf{Bu} \\ \boldsymbol{y} &= \mathbf{Cx} + \mathbf{Du} \end{aligned}$$

where $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$ can depend explicitly on the parameter, the aim of BPOD is to find two low-rank bases $\boldsymbol{V}$ and $\boldsymbol{W}$ such that the reduced-order model

$$\begin{aligned} \boldsymbol{W}^H \boldsymbol{V} \dot{\boldsymbol{z}} &= \boldsymbol{W}^H \mathbf{AVz} + \boldsymbol{W}^H \mathbf{Bu} \\ \hat{\boldsymbol{y}} &= \mathbf{CVz} + \mathbf{Du} \end{aligned}$$

captures most of the input-output properties of the high-dimensional system. For a new operating condition, the matrices describing the system and their parametric dependencies being known ahead of time, only the matrices $\boldsymbol{V}$ and $\boldsymbol{W}$ need to be interpolated. Being only interested in their span rather than the particular bases, this interpolation needs to happen on the Grassmann manifold.

The aim of this task is to implement a parameterized version of BPOD inside `nekStab`. A particular emphasis will be given to the choice of the sampling points in the parameter space as to reduced the computational cost of the offline stage. For that purpose, dedicated *a priori* error estimators will be developed. This task will serve as a reference against which to compare the performances of other parameterized models. BPOD being well-established in the control community and no particular difficulties in its parameterization being foreseen, it will once again serve as a fallback strategy.

Task 3.2 – Parameterization of ADA and AAD on the Stiefel manifold

Assigned to	Principal Investigator (50%) and PhD candidate (50%)
Duration	3 months
Risk	Low

Although ADA and AAD form the current state-of-the-art for linear controller and estimator synthesis for high-dimensional systems, the computational cost remains relatively high. Given a collection of controllers $\{\boldsymbol{K_1}, \cdots, \boldsymbol{K_n}\}$ or Kalman gains $\{\boldsymbol{L_1}, \cdots, \boldsymbol{L_n}\}$ obtained in an offline stage at different operating conditions, this task will propose a methodology to obtain good estimates $\hat{\boldsymbol{K}}$ and $\hat{\boldsymbol{L}}$ for a new operating condition without having to solve the ADA or AAD problems. Entries of $\boldsymbol{K}$ (and similarly for $\boldsymbol{L}$) are not independent as there exists an all-to-all coupling due to being solution of a Ricatti equation. Obtaining estimates of these matrices at a new operating condition using interpolation techniques thus need to preserve this coupling. Once again, interpolation needs to act globally on $\boldsymbol{K}$ (or $\boldsymbol{L}$) rather than in an element-wise fashion. Given the economy-sized SVD factorization $\{\boldsymbol{U}_1 \Sigma_1 \boldsymbol{V}_1^T, \cdots, \boldsymbol{U}_n \Sigma_n \boldsymbol{V}_n^T\}$ of the controllers and Kalman gains obtained from ADA and AAD at different operating conditions, $\hat{\boldsymbol{U}}$ and $\hat{\boldsymbol{V}}$ can be obtained from interpolating the $\boldsymbol{U_i}$'s and $\boldsymbol{V}_i$'s on the dedicated Stiefel manifold while $\hat{\Sigma}$ only needs linear interpolation. The estimate of the controller for the new operating condition is then obtained as $\hat{\boldsymbol{K}} = \hat{\boldsymbol{U}} \hat{\Sigma} \hat{\boldsymbol{V}}^T$. A similar strategy will be used for the Kalman gain $\hat{\boldsymbol{L}}$. As for the other work packages, the performances of these parameterized high-fidelity models will be evaluated and compared against the baseline models developed in WP1.

Task 3.3 – Parameterization of data-driven models on the Stiefel manifold

Assigned to Principal Investigator (20%) and PhD candidate (80%)
Duration 6 months
Risk Moderate

As stated before, most data-driven linear models can be cast into the following optimization problem

$$\begin{aligned} \underset{\boldsymbol{P},\boldsymbol{Q}}{\text{minimize}} \quad & \|\boldsymbol{M}^{1/2}(\boldsymbol{Y} - \mathbf{P}\mathbf{Q}^H\boldsymbol{X})\|_F^2 \\ \text{subject to} \quad & \boldsymbol{P}^H\boldsymbol{M}\boldsymbol{P} = \boldsymbol{I}_r \end{aligned}$$

where the exact meaning of $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{P}$ and $\boldsymbol{Q}$ depend on the particular class of model considered. The operator $\boldsymbol{L} = \mathbf{P}\mathbf{Q}^T$, from which the matrices of the reduced-order model can be constructed, has a low-rank structure. Given models obtained at different operating conditions, interpolating its matrices at a new point in the parameter space thus needs to preserve this low-rank structure. Given $\boldsymbol{P}$ and $\boldsymbol{Q}$, the economy-sized SVD factorization $\boldsymbol{L} = \boldsymbol{U}_L\Sigma_L\boldsymbol{V}_L^T$ can easily be computed. As for Task 3.2, the model at a new operating condition can be obtained form the previously computed models based on the interpolation of the sets $\{\boldsymbol{U}_1, \cdots, \boldsymbol{U}_n\}$ and $\{\boldsymbol{V}_1, \cdots, \boldsymbol{V}_n\}$ on the Stiefel manifold. Due to the unifying framework developed in Task 2.1, this procedure is generic and can be applied to all models having been cast in this scope. A first of proof of concept has been proposed by the principal investigator and collaborators in [52] to develop state observers from DMD models parameterized by the angle of attack of an airfoil in transonic buffeting conditions. As for the rest of the project, attention will be given to control-oriented models although other models (e.g. LSE or SPOD) might be considered. The accuracy and performances of these parameterized models will be compared against the baseline references produced in WP1.

WP 4 – Sensitivity, uncertainty quantification, and applications

Leader Jean-Christophe Loiseau & Xavier Merle
Contributing participants Principal Investigator + Xavier Merle + PhD candidate
Duration 12 months

Except for Task 4.3, applying the methodology developed in the previous work packages to a fully three-dimensional chaotic flow in the wake of a slender body, Tasks 4.1 and 4.2 are more prospective. An aspect often overlooked in data-driven models applied to fluid dynamics is their robustness and statistical significance. Assessing the prediction uncertainty is also critical, particularly for feedback control using an imperfect model. While techniques falling under the umbrella of *robust control* exist, a different route will be explored, combining ideas from sensitivity analysis and uncertainty quantification with statistics on the Stiefel manifold.

Task 4.1 – Sensitivity of data-driven models to varying parameters

Assigned to Principal Investigator (50%) + PhD candidate (50%)
Duration 4 months
Risk Low

So far, parameterizing the models relies on interpolation along the Stiefel manifold solely based on position information. High-order interpolation schemes can be used if one has access to the gradient of the models with respect to the control parameter. In 2009, Hay *et al.* [25] derived the parametric sensitivity of POD modes and POD eigenvalues. POD being a particular instance of the more general optimization problem explored in WP2, the aim of the present task is to extend the work of Hay *et al.* [25] to all data-driven models falling into the unifying framework at the center of this project. Such sensitivity analysis serves multiple purposes. First, including the sensitivity modes into the projection basis for models relying on (Petrov-) Galerkin projection (e.g. BPOD) was shown to increase the range of validity of the reduced-order model, see [25]. It can also serve to inform the sampling strategy, concentrating the computational cost into regions of the parameter space exhibiting high sensitivity. High-order interpolation schemes (e.g. cubic Hermite interpolation on Riemannian manifolds) also benefit from these sensitivities, see [69]. Finally, it can be used to perform between-models comparisons, those with high sensitivity being possibly more accurate but less likely to generalize, and hence less robust.

Task 4.2 – Uncertainty quantification on the Stiefel manifold

Assigned to	Xavier Merle (30%) + PhD candidate (70%)
Duration	6 months
Risk	Moderate

An aspect often overlooked in data-driven models is their robustness with respect to the data itself. This is however critical, particularly for models used for feedback control purposes. The main goal of this task is to robustify the data-driven models and equip them with uncertainty quantification capabilities. Robustifying the identified model with respect to the training data can be performed using an *ensembling* strategy. While approaches such as *Spectral Proper Orthogonal Decomposition* are naturally formulated using such an ensembling, most other data-driven models are not. A first step in this direction has been proposed by Sashidhar & Kutz [53]. One limitation of their approach is that their ensembling procedure relies on simple averaging not necessarily consistent with the low-rank structure of the model. It should be emphasized however that each model generated during this ensembling can be associated with a point on the Stiefel manifold. Once again, this particular manifold provides a natural framework in which to formulate the different statistical tasks needed to quantify the models uncertainties while preserving their fundamental low-rank structure. This task will be conducted in close collaboration with Xavier Merle, a *Maître de Conférences* in DynFluid with a strong expertise in uncertainty quantification techniques.

Task 4.3 – Application to the 3D flow past a slender body

Assigned to	PhD candidate (100%)
External collaborator	Georgios Rigas
Duration	3 years
Risk	Moderate

Although included in WP4, this last task runs over the whole duration of the PhD thesis. After having benchmarked the different techniques on the two test cases (e.g. the canonical cylinder and shear-driven cavity flows), the end goal of this project is to showcase their applicability on a realistic configuration of both academic and industrial interest. For that purpose, the chaotic flow past a three-dimensional slender body already investigated in [49, 11, 50] is considered. Data needed to train the models will be generated using the open-source spectral element solver `Nek5000`, a highly parallel CFD solver for which the DynFluid laboratory has a long and recognized expertise. This task will benefit from on-going collaborations with Georgios Rigas [13, 15] and might see, by then end of the project, the deployment of the control-oriented models into an actual experiment conducted at Imperial College.

WP 5 – Scientific communication and dissemination

Leader	Jean-Christophe Loiseau
Contributing participants	All participants
Duration	4 years

This last work package includes everything related to the communication and dissemination. It includes publication of scientific articles, participation to national and international conferences or workshops, as well as public outreach and the provision of open-source tools.

- **Publications** – Different journals might be targeted depending on the applications. Regarding the theoretical aspects, these include SIAM journal on applied dynamical systems, SIAM journal on mathematics of data science, Journal of Nonlinear Science or Nonlinearity. Applications coming primarily from fluid dynamics, tentative journals include Journal of Fluid Mechanics, Theoretical and Computational Fluid Dynamics, Physical Review Fluids, or Physics of Fluids.

- **Communication** – The following conferences will be targeted: APS Annual Meeting of the Division of Fluid Dynamics, European Fluid Mechanics Conference, 26th International Congress of Theoretical and Applied Mechanics, Bifurcation and Instabilities in Fluid Dynamics, or the SIAM Conference on Applications of Dynamical Systems. Workshops will include the ERCOFTAC SIG33 workshop or relevant EUROMECH Colloquia.

- **Public outreach** – The principal investigator is a regular contributor to the blog Towards Data Science. Short articles describing the major breakthroughs for a non-technical audience might be published on this platform. A dedicated website, hosted using the GitHub Pages system, will also be published online and regularly updated with the latest news regarding the project. Finally, through on-going collaborations with Steven Brunton and regular visits of the principal investigator to Seattle, the public outreach aspect of this project will benefit graciously from their studio and LightBoard (see link for an example).

- **Open-source tools** – One major outcome of this project will be the development of an open-source and user-friendly package for the creation of parameterized data-driven linear models. Following what is being undertaken by the principal investigator on `nekStab`, a dedicated web page will be created and a companion paper submitted to the Journal of Open Source Software.

Along with these well-identified scientific journals, conferences and actions, every opportunity to give seminars in national and international research institutes will be taken.

Milestones and deliverable

Please find below a table summarizing the major deliverable and their expected due dates.

WP	Deliverable	Due date
0	Kick-off Meeting	T0
	D0.1 – D0.7 : Activity reports	Every semester.
	D0.8 : Final report	T0 + 48
1	D1.1 : Implementation of BPOD in `nekStab`	T0 + 3
	D1.2 : Implementation of ADA in `nekStab`	T0 + 7
	D1.3 : Implementation of AAD in `nekStab`	T0 + 10
	D1.4 : Optimal sensor/actuator placement in `nekStab`	T0 + 13
2	D2.1 : Version 1.0 of the open-source package for data-driven linear models.	T0 + 18
	D2.2 : Online publication of the benchmarks	T0 + 24
	D2.3 : Baseline references of the uncontrolled 3D slender body configuration	T0 + 24
3	D3.1 : Parameterization of BPOD models	T0 + 24
	D3.2 : Parameterization of ADA and AAD	T0 + 28
	D3.3 : Version 2.0 of the open-source package including the parameterization of the data-driven linear models which have been cast into the unifying reduced-rank regression framework	T0 + 36
	D3.4 : Online publication of the extended benchmarks, including the parameterization	T0 + 36
4	D4.1 : Implementation of the sensitivity analysis in `nekStab`	T0 + 38
	D4.2 : Version 3.0 of the open-source package with uncertainty quantification capabilities	T0 + 42
	D4.3 : Final results on the 3D slender body flow configuration with control	T0 + 48

Table 2. List of the different deliverable and expected due date.

2 Organization and implementation of the project

2.1 Scientific coordinator and its team

Name	Person-month	Call, agency	Project's title	Scientific coordinator	Start-End
Jean-Christophe LOISEAU	4	CleanSky 2	PERSEUS	Nicolas Mazllier	02/20 – 03/23
Jean-Christophe ROBINET	6	CleanSky 2	PERSEUS	Nicolas Mazellier	02/20 – 03/23
	10	DGAC	MAMBO	J.-Ch. Robinet	09/21 – 09/26
Xavier MERLE					

Table 3. Implication of the scientific coordinator in on-going projects.

Jean-Christophe LOISEAU Principal investigator, he graduated in 2010 from the International Master Program in Fluid dynamics proposed by Université Pierre et Marie Curie and Ecole Polytechnique. During this time, he received a grant from the Undergraduate Research Opportunity Program of Imperial College to study the local stability of two-phase flows with Professor Spelt. He then enrolled in a PhD program at Arts et Métiers with Professors Robinet (Arts et Métiers, Paris) and Leriche (Université Lille 1, Villeneuve d'Ascq). He defended his thesis, entitled *Dynamics and global stability analysis of three-dimensional flows*, in May 2014. From June to December 2014, he obtained a post-doctoral scholarship at DIMEG, Politecnico di Bari, working with Professors Cherubini and Di Palma on nonlinear optimal perturbations in canonical wall-bounded shear flows. From March 2015 to August 2016, he obtained a post-doctoral researcher position at KTH with Professor Luca Brandt. Hired in September 2016 by Arts et Métiers as *Enseignant-Chercheur contractuel*, he was promoted to Maître de Conférences in September 2017. His research activities are evenly divided between *i*) elucidating the physical mechanisms responsible for transition to turbulence in three-dimensional flows, and *ii*) the development and use of data-driven techniques for reduced-order modeling in the physical and engineering sciences.

As of March 2022, he has published more than 20 scientific papers, the most relevant ones for this project being

- [52] A. Sansica, **J.-Ch. Loiseau**, M. Kanamori, A. Hashimoto and J.-Ch. Robinet. *System Identification of two-dimensional transonic buffet*. AIAA Journal, p. 1–17, 2022.
- [14] J. L. Callaham, S. L. Brunton and **J.-Ch. Loiseau**. *On the role of nonlinear correlations in reduced-order modelling*. Journal of Fluid Mechanics, vol. 938, 2022.
- [40] **J.-Ch. Loiseau**, S. L. Brunton and B. R. Noack. *From the POD-Galerkin method to sparse manifold models*. Handbook of Model Order Reduction, vol. 3, 2021.
- [38] **J.-Ch. Loiseau**. *Data-driven modeling of the chaotic thermal convection in an annular thermosyphon*. Theoretical and Computational Fluid Dynamics, 34(4), p. 339–365, 2021.
- [41] **J.-Ch. Loiseau**, B. R. Noack and S. L. Brunton. *Sparse reduced-order modelling: sensor-based dynamics to full-state estimation*. Journal of Fluid Mechanics, 844, p. 459–490, 2018.
- [39] **J.-Ch. Loiseau** and S. L. Brunton. *Constrained sparse Galerkin regression*. Journal of Fluid Mechanics, 838, p. 42–67, 2018.

Since 2017, he has co-supervised one PhD thesis (C. Tarsia Morisco, *Dynamique nonlinéaire et stabilité linéaire d'une tuyère sur-détendue*, funded by CNES, 2020) and is actively co-supervising one with Professor Robinet (R. Schuch Frantz, *Instabilities and transition to turbulence in periodic flows*, defense in April 2022), as well as with Professors Antoine Dazin and Francesco Romano (Arts et Métiers, Lille) on machine learning techniques for turbo-machines (started in Sep. 2021). Finally, he is involved in the CleanSky 2 project PERSEUS, co-supervising a post-doctoral researcher working on adjoint-based sensitivity methods for flow control applications.

While the activities on hydrodynamic instabilities have profited from a *bourse ministérielle*, those on data-driven modeling rely exclusively on collaborations with A. Sansica (JAXA, Japan), B. R. Noack (Harbin Institute of Technology, China), G. Rigas (Imperial College, UK), J. N. Kutz, and S. Brunton (University of Washington, USA). The project will benefit from continuous feedback from these researchers. Even though already enjoying international exposure (seminars in European and oversea institutes, invited researcher at the Institute for Pure and Applied Mathematics at UCLA, invited speaker at the spring school *Outstanding challenges in nonlinear dynamics* in Les Houches), the PhD thesis and post-doctoral position included in this project will enable him to start his own research group and secure long-term funding dedicated to the development of these activities. Finally, plans are being made for him to defend his *Habilitation à diriger des recherches* shortly after the start of the project to ease the supervision of the PhD candidate.

Jean-Christophe ROBINET Head of DynFluid since 2019, he obtained his *Habilitation à diriger des recherches* in 2008 from Université Pierre et Marie Curie. His main research activities focus on linear and nonlinear instabilities, both in the compressible and incompressible regimes. He has been involved in numerous ANR projects, the most recent ones being SICOGIF (2009-2013), DECOMOS (2010-2014), and ASCA (2018-2022). His involvement in this project is primarily limited to that of advisor. Given his expertise with ANR projects, his role will be to ensure the correct progress of the project through periodic meetings. If the HDR defense of the principal investigator is delayed or the supervision denied by Arts et Métiers doctoral school, he will also ensure the administrative supervision of the Ph.D. candidate.

Xavier MERLE After having defended his Ph.D. thesis in 2009 dedicated to global stability analyses, he obtained a position as Maître de Conférences at DynFluid where his main research activities focus on the development of uncertainty quantification techniques for the simulation of turbulent flows. He currently co-supervises two Ph.D. theses on the subject in collaboration with Prof. Paola Cinnella and funded by SAFRAN Tech. In this project, he will lead the theoretical and numerical developments grouped under Task 4.2 related to the formulation of an uncertainty quantification framework for data-driven linear models.

2.2 Implemented and requested resources to reach the objectives

A detailed breakdown of the different expenses is given below, along with a summary table. These informations are also available from the dedicated ANR website.

Staff expenses Theses expenses form the largest fraction of the requested funding and will cover the salary of the Ph.D. candidate and post-doctoral researcher. Following Arts et Métiers guidelines, the monthly wage for the Ph.D. candidate is set to 3350 € (including taxes) for a duration of 36 months, amounting to 120 600 €. Similarly, the monthly wage for the post-doctoral researcher is set to 4000 € (including taxes) for a duration of 18 months, amounting to 72 000 €.

Instruments and material costs Being of mathematical and numerical nature, the project incurs relatively little instruments and material costs. Most large scale computations will be run on the national high-performance computing facilities through a dedicated application. These costs are thus limited to the acquisition of a workstation with a sufficient number of core to run all the test cases locally and conduct preliminary analyses for the three-dimensional flow computations. A tentative description of this workstation is provided below

Vendor	DELL
Number of cores	2×24
Hard drive	4To
Cost	$\sim$ 10 000 €

Building and ground costs None.

Outsourcing/subcontracting None.

General and administrative costs & other operating expenses These expenses include the administrative management costs, as well as the travel costs. It also cover expenses related to the participation to national and international conferences. The administrative management fees applied by Arts et Métiers account for 13% of the total requested funds, amounting to 30 000€. Over the duration of the project, it is anticipated that the post-doctoral researcher will attend one international conference, while the Ph.D. candidate will participate in two to three of them. For each participation, expenses are estimated on average to 2000 € (including conference fees and travel costs), totalling to 8000€. Finally, 2000€ are also budgeted for miscellaneous travel costs, most notably the expenses for organizing the Ph.D. defense.

<table>
<tr><th></th><th colspan="2">Arts et Métiers</th></tr>
<tr><td rowspan="2">Staff expenses</td><td>Ph.D. thesis (duration : 36 months)</td><td>120 600 €</td></tr>
<tr><td>Post-doctoral position (duration : 18 months)</td><td>72 000 €</td></tr>
<tr><td>Instruments and material costs (including scientific consumables)</td><td colspan="2">10 000 €</td></tr>
<tr><td>Building and ground costs</td><td colspan="2">x</td></tr>
<tr><td>Outsourcing and subcontracting</td><td colspan="2">x</td></tr>
<tr><td rowspan="3">General and administrative costs, and other operating expenses</td><td>Travel costs and missions</td><td>2000 €</td></tr>
<tr><td>Conferences and seminars</td><td>6000 €</td></tr>
<tr><td>Administrative management and structure cost**</td><td>27 378 €</td></tr>
<tr><td>Sub-total</td><td colspan="2">237 978 €</td></tr>
<tr><td>Requested funding</td><td colspan="2">237 978 €</td></tr>
</table>

Table 4. Requested means by item of expenditure and by partner*.

* The amount indicated here must be strictly identical to those entered on the website. If both information are not consistent, if they were badly filled or in lacking, the information entered online will prevail on those reported in the submission form/scientific document.

** For marginal cost beneficiaries, these costs will be a package of 13% of the eligible expenses. For full cost beneficiaries, these costs will be a sum of max. 68% of staff expenses and max. 7% of other expenses.

3 Impact and benefits of the project

Applications of closed-loop flow control have epic proportions: drag reduction, lift increase, mixing enhancement, or noise mitigation. In this context, data-driven models have emerged as a powerful paradigm over the past decade. Yet, despite the large body of literature, these techniques still haven't reach the level of maturity of projection-based models, most notably for parametric dynamical systems. The aim of this project is thus to close this gap by leveraging recent progresses in high-dimensional rank-constrainted regressions problems and differential geometry.

Scientific impact Being able to parameterize generalized linear models offer now possibilities. Linear state-space models are the workhorses in the control community. The EigenRealization algorithm (ERA) for instance identifies a state-space model from the low-rank factorization of the Hankel matrix constructed from impulse response data. The range of validity of the model as the system's parameters vary is however limited. By collecting impulse reponse of the system at different operating conditions, the methodology proposed in this project will allow for a better estimate of the model at a new operating condition while limiting the required computational cost. This translate to a better synthesized controller which, in turn, implies better performances of the control system, e.g. more lift enhancement or drag reduction.

Although a strong emphasis throughout the project is given to control-oriented models, the proposed methodology is fairly general. It is applicable to most data-driven models resulting from the formulation of a rank-constrainted least-squares problem. Two prominent such models are POD and DMD. They are used across all scientific disciplines, from fluid dynamics to cognitive and computational neurosciences, for tasks as diverse as reduced-order modeling, coherent structures extraction, or simply for compression. Combining the ability to parameterize these models along with smart sampling strategies of the parameter space can drastically reduced the number of experiments to be conducted in order to obtain a good description of an otherwise complex system. Likewise, non-negative matrix factorization (NNMF) is a widespread algorithm in computer vision. NNMF is another special instance of reduced-rank regression problem to which the proposed methodology might be extended. Similarly, Gaussian mixtures are widely used to model complex probability distributions such as the ones arising in molecular dynamics. Extension of the interpolation scheme to the manifold of symmetric positive definite matrices exists. Doing so would provide researchers with the ability to parameterize these probability densities based on the temperature of the thermal bath in which the system evolves. Because of its generality, the methodology proposed in this project thus far-reaching applications, not only in engineering and control systems but also in data science and machine learning in general.

In the same vein, uncertainty quantification techniques have gained a renewed interest in recent years. This was made possible by new algorithms with improved performances and the ever increasing computational power allowing the application of uncertainty quantification techniques to increasingly complex systems. Extending naturally data-driven linear models with built-in uncertainty quantification capabilities is thus timely. In the machine learning community, linear dimensionality reduction techniques are often used as pre-processing steps prior to fitting a regression or classification model. Being able to quantify the uncertainties in this initial pre-processing step might be extremely valuable for downstream tasks. Once obtained, they can be propagated through the machine learning model to obtain better and more reliable estimates of the uncertainties of the output, both for regression or classification problems. For control applications, the main subject of interest in this project, these extensions would allow to design a control law being robust to epistemic uncertainties.

Societal impacts and industrial applications Although the overall methodology developed during the project is quite general, a strong emphasis is given to control-oriented models. As stated before, applications of closed-loop flow control have epic proportions. Despite the emergence of deep reinforcement learning techniques, the gold standard in industry still relies on linear time-invariant or parameter-varying state-space models combined with optimal or model predictive control. These methods are however challenged by high-dimensionality, uncertainties and parametric dependencies, all three aspects being at the center of this project. Being able to parameterize these models by efficiently sampling the parameter space might thus drastically reduce the computational cost of building them while, at the same time, increasing their range of validity. With the same overarching goal, introducing built-in uncertainty quantification capabilities into these models will lead to more robust controllers and state estimators even in the presence of epistemic uncertainties.

In order to illustrate the industrial relevance of our methodology, a three-dimensional flow of industrial interest is considered as a demonstrator. It exhibits all of the key physical phenomena currently challenging feedback control in industrial setups. These include turbulence, high-dimension, bi-stability, and vortex shedding. All of these phenomena are crucially relevent in the automotive industry. Being able to provide robust low-order models for real-time feedback control targetting drag reduction will thus have an enormous impact, both economically and environmentally. Reliably reducing drag even by a few percent at the scale of a single car or truck directly translates to gigantic savings in term of fuel consumption at the scale of the worlwide population. Additionally, reducing fuel consumption subsequentially implies reduced greenhouse gas emission, a timely objective given the current climatic and geopolitic situation. Finally, even with the transition to electric vehicles and renewable energy, being able to reduced drag on road vehicles still is extremely relevant in order to increase the distance such a vehicle can travel using a limited battery capacity.

Strategy for disseminating and exploiting the results The dissemination strategy relies on different actions. As for most scientific project, results will be published in top-tier journals in applied mathematics (for the theoretical aspects) or fluid dynamics (for the applications). These include journals listed in WP5. The most impactful results will also be presented to the community by participating to national and international conferences as well as through seminars and online webinars. A dedicated website will also be published online using the GitHub Pages system. Along with providing a description of the project, this website will also host periodic updates on the project's progress, the biannual reports, as well as an exhaustive list of the scientific production with links to open-access versions of the published papers whenver possibles (notably by hosting them on arXiv).

Because we value open-source softwares and reproducible research, all the numerical methods developed during the project will be implemented in open-source packages. The first one, `nekStab`, is an open-source toolbox written in Fortran for large-scale bifurcation and stability analysis tailored to the open-source spectral element solver `Nek5000`. It will host all of the algorithmic developments from WP1 relating to high-dimensional Ricatti and Lyapunov solvers as well as the optimization algorithm for sensors and actuators placement. Although it reached version 1.0 only recently, `nekStab` has been actively developed by the principal investigator of this project over the course of the past ten years. Developments associated to WP2 to WP4, related to data-driven models and their parameterization, will rely on a second open-source package which has yet to be developed. As stated earlier in the document, the computational core of this package will be written in Julia. This relatively new programming language, first published in 2011, has already established itself as a serious contender in the field of scientific computing by combining the flexibility of Python and numerical performances of compiled languages such C or Fortran. Another key aspect is its extremely high composability. In particular, our yet-to-come package will leverage some capabilities from `Manifold.jl`, a Julia package providing a unified interface for operations on differential manifolds. The package will moreover benefit from a front-end written in Python to increase its reach. This front-end will expose an API compatible with `scikit-learn`, the most widely used package for Machine Learning in Python, to benefit directly from its wide ecosystem.

Finally, public outreach for a non-technical audience or students is also a key aspect of our dissemination strategy. It will rely on different channels. The principal investigator being a regular contributor to the blog Towards Data Science (see here for the list of contributions), blog posts presenting theoretical aspects which might be useful to practionners outside the field of fluid dynamics will be published via this platform (current contributions collect more than 3500 views per month on average). These posts will also be advertised on Twitter. Finally, Steve Brunton, with whom the principal investigator has a long standing collaboration, has graciously offered that we use his LightBoard and recording studio to produce videos presenting the most important results of this project.

Bibliography

To help the referees navigate the bibliography, the most important contributions from the principal investigator are highlighted in red. The most relevant ones regarding the mathemtical framework described in this project are shown in bold.

[1] David Amsallem and Charbel Farhat. Interpolation Method for Adapting Reduced-Order Models and Application to Aeroelasticity. *AIAA Journal*, 46(7):1803–1813, 2008. Publisher: American Institute of Aeronautics and Astronautics _eprint: https://doi.org/10.2514/1.35374.

[2] David Amsallem and Charbel Farhat. An Online Method for Interpolating Linear Parametric Reduced-Order Models. *SIAM Journal on Scientific Computing*, 33(5):2169–2198, jan 2011. Publisher: Society for Industrial and Applied Mathematics.

[3] Travis Askham and J. Nathan Kutz. Variable Projection Methods for an Optimized Dynamic Mode Decomposition. *SIAM Journal on Applied Dynamical Systems*, 17(1):380–416, jan 2018. Publisher: Society for Industrial and Applied Mathematics.

[4] Omri Azencot, Wotao Yin, and Andrea Bertozzi. Consistent Dynamic Mode Decomposition. *SIAM Journal on Applied Dynamical Systems*, 18(3):1565–1585, jan 2019. Publisher: Society for Industrial and Applied Mathematics.

[5] Peter J. Baddoo, Benjamin Herrmann, Beverley J. McKeon, J. Nathan Kutz, and Steven L. Brunton. Physics-informed dynamic mode decomposition (piDMD). *ArXiv:2112.04307 [physics]*, dec 2021. ArXiv: 2112.04307.

[6] Zhe Bai, Eurika Kaiser, Joshua L. Proctor, J. Nathan Kutz, and Steven L. Brunton. Dynamic Mode Decomposition for Compressive System Identification. *AIAA Journal*, 58(2):561–574, 2020. Publisher: American Institute of Aeronautics and Astronautics _eprint: https://doi.org/10.2514/1.J057870.

[7] Peter Benner, Serkan Gugercin, and Karen Willcox. A Survey of Projection-Based Model Reduction Methods for Parametric Dynamical Systems. *SIAM Review*, 57(4):483–531, 2015. Publisher: Society for Industrial and Applied Mathematics.

[8] Peter Benner, Christian Himpe, and Tim Mitchell. On reduced input-output dynamic mode decomposition. *Advances in Computational Mathematics*, 44(6):1751–1768, dec 2018.

[9] Thomas Bewley, Paolo Luchini, and Jan Pralits. Methods for solution of large optimal control problems that bypass open-loop model reduction. *Meccanica*, 51(12):2997–3014, dec 2016.

[10] Diana Alina Bistrian and Ionel Michael Navon. Randomized dynamic mode decomposition for nonintrusive reduced order modelling. *International Journal for Numerical Methods in Engineering*, 112(1):3–25, 2017. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.5499.

[11] R. D. Brackston, J. M. García de la Cruz, A. Wynn, G. Rigas, and J. F. Morrison. Stochastic modelling and feedback control of bistability in a turbulent bluff body wake. *Journal of Fluid Mechanics*, 802:726–749, sep 2016. Publisher: Cambridge University Press.

[12] Steven L. Brunton, Joshua L. Proctor, Jonathan H. Tu, and J. Nathan Kutz. Compressed sensing and dynamic mode decomposition. *Journal of Computational Dynamics*, 2(2):165, 2015. Company: Journal of Computational Dynamics Distributor: Journal of Computational Dynamics Institution: Journal of Computational Dynamics Label: Journal of Computational Dynamics Publisher: American Institute of Mathematical Sciences.

[13] J. L. Callaham, J.-C. Loiseau, G. Rigas, and S. L. Brunton. Nonlinear stochastic modelling with Langevin regression. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 477(2250):20210092, jun 2021.

[14] Jared L. Callaham, Steven L. Brunton, and Jean-Christophe Loiseau. On the role of nonlinear correlations in reduced-order modelling. *Journal of Fluid Mechanics*, 938:0, may 2022.

[15] Jared L. Callaham, Georgios Rigas, Jean-Christophe Loiseau, and Steven L. Brunton. An empirical mean-field model of symmetry-breaking in a turbulent wake. *ArXiv:2105.13990 [physics]*, may 2021. ArXiv: 2105.13990.

[16] Kevin K. Chen and Clarence W. Rowley. H2 optimal actuator and sensor placement in the linearised complex Ginzburg–Landau system. *Journal of Fluid Mechanics*, 681:241–260, aug 2011. Publisher: Cambridge University Press.

[17] Miles Cranmer. MilesCranmer/PySR v0.2. sep 2020.

[18] Scott T. M. Dawson, Maziar S. Hemati, Matthew O. Williams, and Clarence W. Rowley. Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. *Experiments in Fluids*, 57(3):42, feb 2016.

[19] Fernando De la Torre. A Least-Squares Framework for Component Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(6):1041–1055, jun 2012. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[20] Jonas Degrave, Federico Felici, Jonas Buchli, Michael Neunert, Brendan Tracey, Francesco Carpanese, Timo Ewalds, Roland Hafner, Abbas Abdolmaleki, Diego de las Casas, Craig Donner, Leslie Fritz, Cristian Galperti, Andrea Huber, James Keeling, Maria Tsimpoukelli, Jackie Kay, Antoine Merle, Jean-Marc Moret, Seb Noury, Federico Pesamosca, David Pfau, Olivier Sauter, Cristian Sommariva, Stefano Coda, Basil Duval, Ambrogio Fasoli, Pushmeet Kohli, Koray Kavukcuoglu, Demis Hassabis, and Martin Riedmiller. Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature*, 602(7897):414–419, feb 2022. Number: 7897 Publisher: Nature Publishing Group.

[21] K. D. Dhuri and P. Seshu. Multi-objective optimization of piezo actuator placement and sizing using genetic algorithm. *Journal of Sound and Vibration*, 323(3):495–514, 2009.

[22] N. Benjamin Erichson, Lionel Mathelin, J. Nathan Kutz, and Steven L. Brunton. Randomized Dynamic Mode Decomposition. *SIAM Journal on Applied Dynamical Systems*, 18(4):1867–1891, jan 2019. Publisher: Society for Industrial and Applied Mathematics.

[23] Dixia Fan, Liu Yang, Zhicheng Wang, Michael S. Triantafyllou, and George Em Karniadakis. Reinforcement learning for bluff body active flow control in experiments and simulations. *Proceedings of the National Academy of Sciences*, 117(42):26091–26098, oct 2020.

[24] Paul Garnier, Jonathan Viquerat, Jean Rabault, Aurélien Larcher, Alexander Kuhnle, and Elie Hachem. A review on deep reinforcement learning for fluid mechanics. *Computers & Fluids*, 225:104973, 2021.

[25] Alexander Hay, Jeffrey T. Borggaard, and Dominique Pelletier. Local improvements to reduced-order models using sensitivity analysis of the proper orthogonal decomposition. *Journal of Fluid Mechanics*, 629:41–72, jun 2009. Publisher: Cambridge University Press.

[26] Patrick Héas and Cédric Herzet. Low-Rank Dynamic Mode Decomposition: An Exact and Tractable Solution. *Journal of Nonlinear Science*, 32(1):8, dec 2021.

[27] Patrick Héas, Cédric Herzet, and Benoit Combès. Generalized Kernel-Based Dynamic Mode Decomposition. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3877–3881. May 2020. ISSN: 2379-190X.

[28] Maziar S. Hemati, Clarence W. Rowley, Eric A. Deem, and Louis N. Cattafesta. De-biasing the dynamic mode decomposition for applied Koopman spectral analysis of noisy datasets. *Theoretical and Computational Fluid Dynamics*, 31(4):349–368, aug 2017.

[29] Akhtar Imran, Jeff Borggaard, John Burns, Haroon Imtiaz, and Lizette Zietsman. Using functional gains for effective sensor location in flow control: A reduced-order modelling approach. *Journal of Fluid Mechanics*, 781:622–656, 2015.

[30] I. W. Jamaludin, N. A. Wahab, N. S. Khalid, S Sahlan, Z. Ibrahim, and M F. Rahmat. N4SID and MOESP subspace identification methods. In *2013 IEEE 9th International Colloquium on Signal Processing and its Applications*, pages 140–145. Mar 2013.

[31] Mihailo R. Jovanović, Peter J. Schmid, and Joseph W. Nichols. Sparsity-promoting dynamic mode decomposition. *Physics of Fluids*, 26(2):24103, feb 2014. Publisher: American Institute of Physics.

[32] Jer-Nan Juang and Richard S. Pappa. An eigensystem realization algorithm for modal parameter identification and model reduction. *Journal of Guidance, Control, and Dynamics*, 8(5):620–627, 1985. Publisher: American Institute of Aeronautics and Astronautics _eprint: https://doi.org/10.2514/3.20031.

[33] Manu Kalia, Steven L. Brunton, Hil G. E. Meijer, Christoph Brune, and J. Nathan Kutz. Learning normal form autoencoders for data-driven discovery of universal,parameter-dependent governing equations. *ArXiv:2106.05102 [cs, math]*, jun 2021. ArXiv: 2106.05102.

[34] J. Nathan Kutz, Xing Fu, and Steven L. Brunton. Multiresolution Dynamic Mode Decomposition. *SIAM Journal on Applied Dynamical Systems*, 15(2):713–735, jan 2016. Publisher: Society for Industrial and Applied Mathematics.

[35] W.E. Larimore. Canonical variate analysis in identification, filtering, and adaptive control. In *29th IEEE Conference on Decision and Control*, pages 596–604. Dec 1990.

[36] Soledad Le Clainche and José M. Vega. Higher Order Dynamic Mode Decomposition. *SIAM Journal on Applied Dynamical Systems*, 16(2):882–925, jan 2017. Publisher: Society for Industrial and Applied Mathematics.

[37] Kookjin Lee and Kevin T. Carlberg. Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. *Journal of Computational Physics*, 404:108973, mar 2020.

[38] Jean-Christophe Loiseau. Data-driven modeling of the chaotic thermal convection in an annular thermosyphon. *Theoretical and Computational Fluid Dynamics*, 34(4):339–365, aug 2020.

[39] Jean-Christophe Loiseau and Steven L. Brunton. Constrained sparse Galerkin regression. *Journal of Fluid Mechanics*, 838:42–67, mar 2018. Publisher: Cambridge University Press.

[40] Jean-Christophe Loiseau, Steven Brunton, and Bernd Noack. From the POD-Galerkin Method to Sparse Manifold Models. 2021. Publisher: De Gruyter.

[41] Jean-Christophe Loiseau, Bernd R. Noack, and Steven L. Brunton. Sparse reduced-order modelling: sensor-based dynamics to full-state estimation. *Journal of Fluid Mechanics*, 844:459–490, jun 2018. Publisher: Cambridge University Press.

[42] Krithika Manohar, J. Nathan Kutz, and Steven L. Brunton. Optimal Sensor and Actuator Selection using Balanced Model Reduction. *IEEE Transactions on Automatic Control*, pages 1–1, 2021. Conference Name: IEEE Transactions on Automatic Control.

[43] B. Moore. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Transactions on Automatic Control*, 26(1):17–32, feb 1981. Conference Name: IEEE Transactions on Automatic Control.

[44] Jeremy Morton, Mykel J. Kochenderfer, and Freddie D. Witherden. Parameter-Conditioned Sequential Generative Modeling of Fluid Flows. *AIAA Journal*, 59(3):825–841, 2021. Publisher: American Institute of Aeronautics and Astronautics _eprint: https://doi.org/10.2514/1.J059315.

[45] Bernd R. Noack, Witold Stankiewicz, Marek Morzyński, and Peter J. Schmid. Recursive dynamic mode decomposition of transient and post-transient wake flows. *Journal of Fluid Mechanics*, 809:843–872, dec 2016. Publisher: Cambridge University Press.

[46] Romain Paris, Samir Beneddine, and Julien Dandois. Robust flow control and optimal sensor placement using deep reinforcement learning. *Journal of Fluid Mechanics*, 913, apr 2021. Publisher: Cambridge University Press.

[47] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine Learning in Python. *The Journal of Machine Learning Research*, 12(null):2825–2830, nov 2011.

[48] Joshua L. Proctor, Steven L. Brunton, and J. Nathan Kutz. Dynamic Mode Decomposition with Control. *SIAM Journal on Applied Dynamical Systems*, 15(1):142–161, jan 2016.

[49] G. Rigas, A. R. Oxlade, A. S. Morgans, and J. F. Morrison. Low-dimensional dynamics of a turbulent axisymmetric wake. *Journal of Fluid Mechanics*, 755, sep 2014. Publisher: Cambridge University Press.

[50] Georgios Rigas, Lucas Esclapez, and Luca Magri. Symmetry breaking in a 3D bluff-body wake. *ArXiv:1703.07405 [physics]*, mar 2017. ArXiv: 1703.07405.

[51] C. W. Rowley. Model reduction for fluids, using balanced proper orthogonal decomposition. *International Journal of Bifurcation and Chaos*, 15(03):997–1013, mar 2005. Publisher: World Scientific Publishing Co.

[52] Andrea Sansica, Jean-Christophe Loiseau, Masashi Kanamori, Atsushi Hashimoto, and Jean-Christophe Robinet. System Identification of Two-Dimensional Transonic Buffet. *AIAA Journal*, pages 1–17, feb 2022. Publisher: American Institute of Aeronautics and Astronautics.

[53] Diya Sashidhar and J. Nathan Kutz. Bagging, optimized dynamic mode decomposition (BOP-DMD) for robust, stable forecasting with spatial and temporal uncertainty-quantification. *ArXiv:2107.10878 [cs, math]*, jul 2021. ArXiv: 2107.10878.

[54] Peter J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, aug 2010. Publisher: Cambridge University Press.

[55] Peter J. Schmid. Dynamic Mode Decomposition and Its Variants. *Annual Review of Fluid Mechanics*, 54(1):225–254, 2022. _eprint: https://doi.org/10.1146/annurev-fluid-030121-015835.

[56] Onofrio Semeraro and Jan O. Pralits. Full-order optimal compensators for flow control: the multiple inputs case. *Theoretical and Computational Fluid Dynamics*, 32(3):285–305, jun 2018.

[57] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis. Mastering the game of Go without human knowledge. *Nature*, 550(7676):354–359, oct 2017. Number: 7676 Publisher: Nature Publishing Group.

[58] Naoya Takeishi, Yoshinobu Kawahara, Yasuo Tabei, and Takehisa Yairi. Bayesian Dynamic Mode Decomposition. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pages 2814–2821. Melbourne, Australia, aug 2017. International Joint Conferences on Artificial Intelligence Organization.

[59] Hongwei Tang, Jean Rabault, Alexander Kuhnle, Yan Wang, and Tongguang Wang. Robust active flow control over a range of Reynolds numbers using an artificial neural network trained through deep reinforcement learning. *Physics of Fluids*, 32(5):53605, may 2020. Publisher: American Institute of Physics.

[60] Jonathan H. Tu, Clarence W. Rowley, Dirk M. Luchtenburg, Steven L. Brunton, and J. Nathan Kutz. On dynamic mode decomposition: Theory and applications. *Journal of Computational Dynamics*, 1(2):391, 2014. Company: Journal of Computational Dynamics Distributor: Journal of Computational Dynamics Institution: Journal of Computational Dynamics Label: Journal of Computational Dynamics Publisher: American Institute of Mathematical Sciences.

[61] Peter Van Overschee and Bart De Moor. N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30(1):75–93, jan 1994.

[62] Francesco Vicario, Minh Q. Phan, Raimondo Betti, and Richard W. Longman. OKID as a Unified Approach to System Identification. Pages 3443–3460, 2014.

[63] Oriol Vinyals, Igor Babuschkin, Wojciech M. Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H. Choi, Richard Powell, Timo Ewalds, Petko Georgiev, Junhyuk Oh, Dan Horgan, Manuel Kroiss, Ivo Danihelka, Aja Huang, Laurent Sifre, Trevor Cai, John P. Agapiou, Max Jaderberg, Alexander S. Vezhnevets, Rémi Leblond, Tobias Pohlen, Valentin Dalibard, David Budden, Yury Sulsky, James Molloy, Tom L. Paine, Caglar Gulcehre, Ziyu Wang, Tobias Pfaff, Yuhuai Wu, Roman Ring, Dani Yogatama, Dario Wünsch, Katrina McKinney, Oliver Smith, Tom Schaul, Timothy Lillicrap, Koray Kavukcuoglu, Demis Hassabis, Chris Apps, and David Silver. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 575(7782):350–354, nov 2019. Number: 7782 Publisher: Nature Publishing Group.

[64] Matthew O. Williams, Ioannis G. Kevrekidis, and Clarence W. Rowley. A Data–Driven Approximation of the Koopman Operator: Extending Dynamic Mode Decomposition. *Journal of Nonlinear Science*, 25(6):1307–1346, dec 2015.

[65] Matthew O. Williams, Clarence W. Rowley, and Ioannis G. Kevrekidis. A Kernel-Based Approach to Data-Driven Koopman Spectral Analysis. *ArXiv:1411.2260 [math]*, jul 2015. ArXiv: 1411.2260.

[66] Hao Zhang, Clarence W. Rowley, Eric A. Deem, and Louis N. Cattafesta. Online Dynamic Mode Decomposition for Time-Varying Systems. *SIAM Journal on Applied Dynamical Systems*, 18(3):1586–1609, jan 2019. Publisher: Society for Industrial and Applied Mathematics.

[67] Ralf Zimmermann. A Matrix-Algebraic Algorithm for the Riemannian Logarithm on the Stiefel Manifold under the Canonical Metric. *SIAM Journal on Matrix Analysis and Applications*, 38(2):322–342, jan 2017. Publisher: Society for Industrial and Applied Mathematics.

[68] Ralf Zimmermann. Manifold interpolation and model reduction. *ArXiv:1902.06502 [cs, math]*, sep 2019. ArXiv: 1902.06502.

[69] Ralf Zimmermann. Hermite Interpolation and Data Processing Errors on Riemannian Matrix Manifolds. *SIAM Journal on Scientific Computing*, 42(5):0, jan 2020. Publisher: Society for Industrial and Applied Mathematics.

[70] Ralf Zimmermann, Benjamin Peherstorfer, and Karen Willcox. Geometric Subspace Updates with Applications to Online Adaptive Nonlinear Model Reduction. *SIAM Journal on Matrix Analysis and Applications*, 39(1):234–261, jan 2018. Publisher: Society for Industrial and Applied Mathematics.