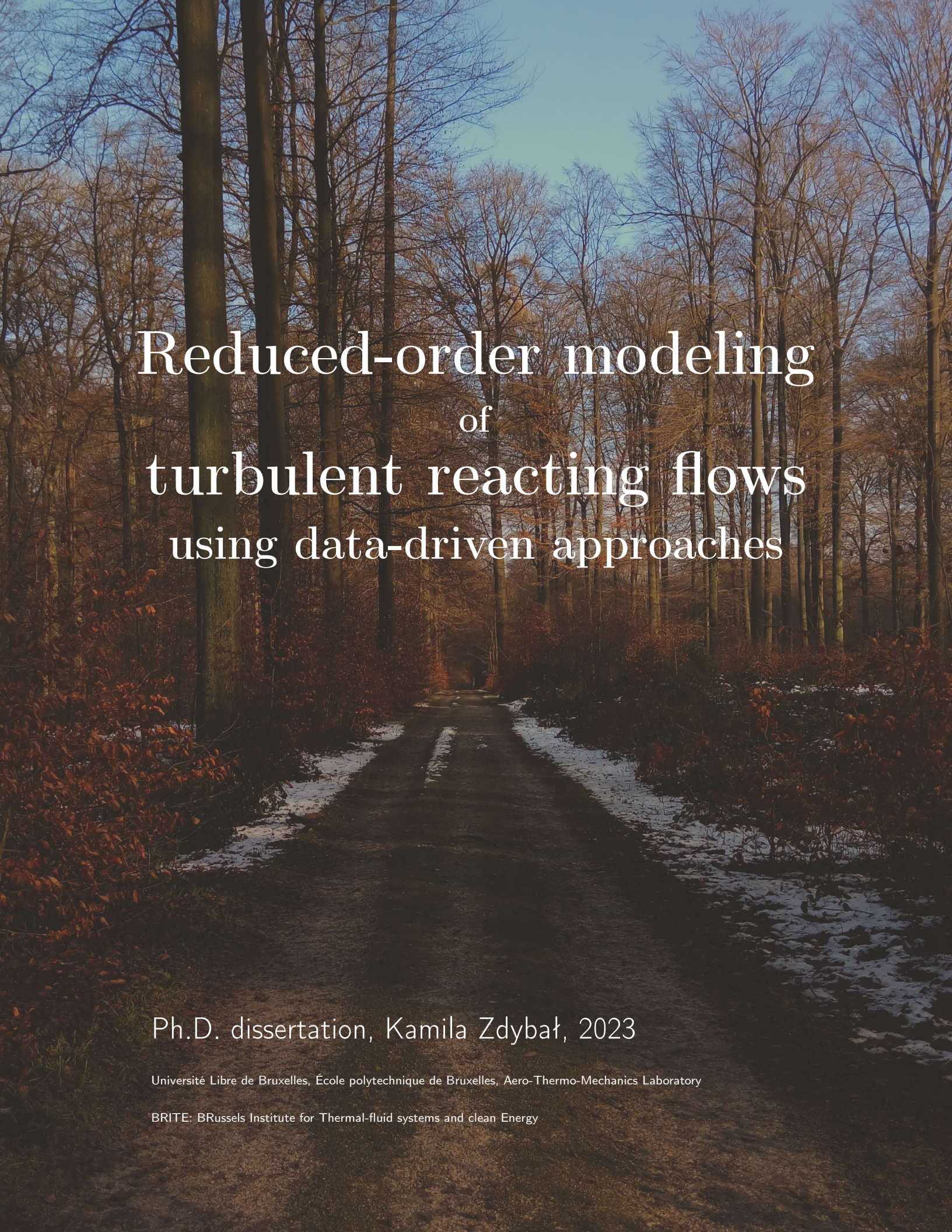

Reduced-order modeling of turbulent reacting flows using data-driven approaches

Ph.D. dissertation, Kamila Zdybał, 2023

Université Libre de Bruxelles, École polytechnique de Bruxelles, Aero-Thermo-Mechanics Laboratory

BRITE: BRussels Institute for Thermal-fluid systems and clean Energy

Abstract

Turbulent multicomponent reacting flows are described by a large number of coupled partial differential equations. With such large systems of equations, the current computational capabilities are insufficient for detailed simulations. At the same time, accurate simulations are crucial to support the rapidly developing combustion technologies. Dimensionality reduction and machine learning approaches appear well-suited for building reduced-order models (ROMs) of complex systems with many degrees of freedom. Dimensionality reduction techniques project a high-dimensional system onto a lower-dimensional basis. Projections can be computed from the available training data and are referred to as low-dimensional manifolds (LDMs). Dimensionality reduction is often coupled with nonlinear regression to bypass the errors associated with the inverse basis transformation. Regression allows to reconstruct the target thermo-chemical state quantities from the LDM parameters. A data-driven ROM workflow provides substantial reduction to the number of transport equations solved in combustion simulations, but the quality of the manifold topology is one of the decisive aspects in successful modeling. Numerous manifold challenges of turbulent combustion have been reported in the literature and ought to be addressed. The present work advances the performance of ROMs of reacting flows. Our main focus is in addressing the outstanding manifold challenges. We provide novel tools and algorithms that can help further reduce the order, and improve the predictive capabilities of the model.

The significant original contribution of this work is the development of tools to quantify the quality of LDMs from the perspective of ROM. We propose a metric that reduces the LDM topology to a single number, based on two aspects that affect modeling in particular: (1) steep gradients and (2) non-uniqueness in dependent quantities of interest (QoIs). Such quantitative tool was not available in the literature thus far. The metric becomes particularly informative when building nonlinear regression models on top of a low-dimensional projection.

We demonstrate that LDM topologies can be improved using our quantitative metric as a cost function in optimization algorithms. The next contribution of this work is development of strategies to improve topologies of low-dimensional data representations. In particular, two new algorithms for variable (feature) selection are developed, that return a subset of the thermo-chemical state vector. The subset is optimized to yield an improved LDM quality once it is projected onto a lower-dimensional basis. We also use our quantitative tools to assess other means of data preprocessing, including data scaling and data sampling. We show that quantitative rankings of various data preprocessing and manifold learning strategies can be created *a priori* at the modeling stage. This allows for automating decisions which thus far had to be performed manually – either through trial and error or using heuristic guidelines. We discover that among many data preprocessing scenarios, adequate data scaling combined with optimized variable selection has the potential to affect the LDM topologies the most. We argue that further improvements in parameterization quality can be achieved in many areas of science and engineering if the low-dimensional parameter space is thoroughly explored and then assessed using the proposed quantitative metric.

While principal component analysis (PCA) has been established in the combustion literature as a dimensionality reduction technique, we develop an alternative approach to obtain LDMs from data. We propose to combine dimensionality reduction and nonlinear regression within an encoder-decoder neural network architecture. Research efforts have thus far considered dimensionality reduction and nonlinear reconstruction as two separate steps. We show significant improvements in LDM topology

when these two steps are allowed to communicate with each other through backpropagation. Data projection becomes directly optimized to represent the QoIs regressed at the output of a decoder. The significant discovery of this work is that a nonlinear reconstruction error optimality promotes finding improved LDM topologies as compared to a linear reconstruction error optimality (*e.g.*, as in PCA). Our approach can become an effective replacement of standalone dimensionality reduction techniques, such as PCA, whenever nonlinear regression is anticipated in the downstream use.

We demonstrate our predictive tools inside a full ROM of a simple system of a zero-dimensional reactor. We first generate a good quality manifold using the proposed tools. We then benchmark several nonlinear regression models: artificial neural networks (ANNs), Gaussian process regression (GPR), kernel regression, and radial basis function (RBF) regression. We show that improved manifold topologies correlate with improved manifold regressibility. We transport the LDM parameters, instead of the high-dimensional thermo-chemical state variables. We demonstrate *a posteriori* insights on the benefits that improved manifold topologies and improved nonlinear regression bring in ROMs. The challenges that remain are linked with nonlinear regression performance, especially at the boundaries of the training manifold. We propose strategies that may help improve kernel-based regression methods. Among these are local kernel rotations based on gradients in QoIs, and local anisotropic bandwidth selections based on local feature sizes in QoIs.

Finally, we provide insights into physical interpretability of low-dimensional data parameterizations obtained using data science tools. We apply local PCA to combustion datasets of varying complexity in order to find settings that support finding physically meaningful information from data. Our approach connects with the recent trends in semi-supervised learning to incorporate any existing information about the system being studied. The results indicate that physics-based knowledge of the system can be used to enhance data-driven algorithms.

Two new Python libraries are developed in this work. The first library is **`PCAfold`**, a Python software package that can be used to generate, analyze and improve low-dimensional data representations. This software is paramount to generating and reproducing results in this dissertation. Each tool developed in this work is available in the **`PCAfold`** library. **`PCAfold`** can be applied broadly in other disciplines of research. The second library, **`multipy`**, has a mostly didactic purpose. This library can accompany and support a graduate course on multicomponent mass transfer. It can become a helpful study tool for students performing research in the area of reacting flows.

Contents