A novel approach combining bootstrapped non-intrusive reduced order models and unscented transform for the robust and efficient CFD analysis of accidental gas releases in

A novel approach combining bootstrapped non-intrusive reduced order models and unscented transform for the robust and efficient CFD analysis of accidental gas releases in congested plants

Nicolò Abrate[a], Alberto Moscatello[a], Gianmario Ledda[a], Nicola Pedroni[a], Federica Carbone[a], Emanuela Maffia[a], Andrea Carpignano[a]

[a] *Politecnico di Torino, Energy Department, Corso Duca degli Abruzzi, 24 - 10129 Turin (Italy)*

Abstract

The risk assessment for safety-critical, complex systems is a very challenging computational problem when it is performed with high-fidelity models, e.g. CFD, like in the case of accidental gas releases in congested systems. Within this framework, a novel CFD approach, named Source Box Accident Model, has been recently proposed to efficiently model such phenomena by splitting the simulation of the gas release and its subsequent dispersion in the system in two steps. In this view, the present paper proposes a non-intrusive, Proper Orthogonal Decomposition-Radial Basis Functions reduced order model that exploits the two-step nature of the SBAM approach, to mimic the behaviour of the original, long-running CFD model code at a significantly lower computational cost. Moreover, the paper presents a methodology combining the bootstrap and unscented transform approaches to efficiently assess the ROM uncertainty in the safety-critical simulation output quantities of interest, e.g. the flammable volume. The results obtained in a test case involving a high pressure, accidental gas release in an off-shore Oil & Gas plant are in very

Email addresses: nicolo.abrate@polito.it (Nicolò Abrate), alberto.moscatello@polito.it (Alberto Moscatello), gianmario.ledda@polito.it (Gianmario Ledda), nicola.pedroni@polito.it (Nicola Pedroni), federica.carbone@studenti.polito.it (Federica Carbone), emanuela.maffia@studenti.polito.it (Emanuela Maffia), andrea.carpignano@polito.it (Andrea Carpignano)

satisfactory agreement with those produced by CFD, with a relative error smaller than 10% and a reduction in the computational time of about three orders of magnitude.

Keywords: reduced order models; radial basis functions; proper orthogonal decomposition; computational fluid-dynamics; high-pressure gas release; ANSYS Fluent;

Acronyms

CFD Computational Fluid Dynamics

FOM Full Order Model

LFL Low Flammability Limit

LOOCV Leave-One-Out Cross Validation

NIROM Non-Intrusive Reduced Order Model

POD Proper Orthogonal Decomposition

QRA Quantitative Risk Assessment

RANS Reynolds Averaged Navier-Stokes

RBF Radial Basis Functions

RMSE Root Mean Square Error

ROM Reduced Order Model

SB Source Box

SBAM Source Box Accident Model

UFL Upper Flammability Limit

UT Unscented Transform

1. Introduction

In the last decades, the attention to major hazards, especially in the industrial field, has considerably increased. Some regulations have been introduced to improve the safety of risk-relevant industrial plants, like Oil & Gas, chemical and nuclear ones, which involve dangerous substances related to flammability, toxicity and radioactivity, and dangerous equipment like highly pressurized tanks, pipelines etc. A failure in this kind of systems could lead to major accidents Casal (2008), which, although very rare, must be considered in the risk assessment by evaluating the possible consequences and related damage level. The Quantitative Risk Assessment (QRA) is defined as the appropriate methodology for the safety demonstration of major accident involving industrial plants by the European Union (EU) Offshore Safety Directive 2013/30/EU Parliament (2013). The QRA consists of the analysis of hundreds of plausible accidental scenarios, which consequences must be evaluated through suitable simulations tools. The state-of-the-art of QRA methodologies entails the employment of simple integral or semi-empirical models Vinnem and Røed (2014), due to the necessity to analyze a large number of events in a time compatible with the design project schedule of the plant. In fact, the aforementioned models can provide results in a few minutes and do not require any particular theoretical background or expertise level by the user. In the past, empirical or integral models were formulated for several phenomena as turbulent free-jets Chen and Rodi (1980); Becker et al. (1967), gas dispersions Davidson (1989), jet fires TNO (2005); Zamejc (2014), explosions TNO (2005), Vapor Cloud Explosion (VCE) Baker et al. (1996) and were implemented in software like PHAST DNV or ALOHA US, providing simple comprehensive tools to companies. On the other hand, the range of validity of these models is constrained by their assumptions and empirical tests conditions, therefore their application cannot be generalized to any kind of phenomena. Moreover, their accuracy is so limited that excessively large over estimation of the threat zones related to the accidental event often occur Dasgotra et al. (2018); Derudi et al. (2014); Schleder et al. (2015); Zuliani et al. (2016); Pontiggia et al. (2014). Finally, they completely neglect the actual geometrical features of the domain, i.e., only free jets and fires are considered, and the interaction with the surrounding equipment is not accounted for. This could have a strong impact on the final design of the safety systems, especially considering high-pressure gas leak scenarios in congested domains where the gas-objects interaction plays a key role on the

final gas cloud configuration. For these reasons, and more in general due to the increase of technology complexity and safety standards, novel QRA approaches are needed to enhance the accuracy of the risk estimation. Computational Fluid Dynamics (CFD) methods can represent a possible strategy to overcome this issue, although they require a high computational effort. Hence, the employment of the CFD in a QRA is still prohibitive, and its use is limited for the verification of the most critical scenarios. This work proposes an alternative strategy to reduce the computational effort related to the high-fidelity simulation of high-pressure gas releases in congested plants. These kind of scenarios are characterized by a highly compressible flow near the leak point and a subsonic incompressible flow far from that point, i.e., in the largest part of the domain, and as a matter of fact, the main issue in simulating such events is mainly related to the formation of an under-expanded jet Franquet et al. (2015) near the leak source, which is likely to interact with an obstacle. The resolution of the steep flow field variables gradient appearing in that region require the generation of a dense mesh, which translates in an unacceptable computational effort. Several authors proposed innovative strategies to handle this situation. For instance, in Colombini et al. (2021, 2022a,b) the authors propose an empirical-CFD combined approach to develop a novel assessment tool to estimate the extent of a high-pressure methane jet impinging on different shaped obstacles (cylinder, sphere, pipe rack). Despite this methodology permits a relevant simplification of the simulation and computational cost reduction since the initial jet expansion is accounted for by the Birch model Birch et al. (1984), it can be applied only if the obstacle is placed far enough from the release point to permit a complete expansion of the gas jet. Full CFD approaches are also proposed, as in Liu et al. (2014) where a CO_2 high pressure leakage from a pipeline is studied, or as in de Souza et al. (2019) where predictions for hazardous area classification is presented analyzing a set of scenarios. Nevertheless, in all these cases only free-jets are considered. Within this framework, a computationally efficient approach, named Source Box Accident Model (SBAM) Moscatello et al. (2021) has been recently proposed to significantly mitigate the issue of the computational cost. It consists in splitting the accident evolution in two steps: i) the *release* phase, which concerns a small volume around the break, named Source Box (SB), where the underexpanded jet occur and compressibility effects are relevant and ii) the *dispersion* phase, where the flow can be considered incompressible and buoyancy forces are relevant. In this last step, where the SB flow profiles and gas concentrations are taken as input,

the analysis is extended to the full spatial scale of the plant (e.g. an offshore platform, see figure 6 later), in order to evaluate some safety-critical parameters, e.g., the flammable volume. This two-step approach allows to fairly reduce the computational burden with respect to a monolithic CFD simulation, as different numerical settings, e.g., the computational mesh, can be tailored according to the physical phenomena involved. Decoupling the two physical stages implies that also the parameter space can be partitioned. For example, the break size and release pressure are relevant in the gas release, while the wind direction is not. This means that it is possible to compute the concentration and flow fields for several SB characterised by different combinations of parameters and use them in the dispersion scenarios, whenever they are needed. The numerical features and performances of SBAM have been presented in Moscatello et al. (2021) while its performance against experimental data in Moscatello et al. (2022). This procedure would lend itself to the generation of a SB library as a ready-to-use input to the dispersion simulation, which is strongly case-dependent. Relying on data libraries is a common approach in the framework of the risk analysis, but in this case it would jeopardise the advantages of using a CFD approach. For example, the risk analyst would not be free to select the flow field profiles for an arbitrary set of parameters, losing the CFD flexibility. Moreover, a thorough evaluation of the different SB scenarios would be still quite computationally expensive. In this paper, which is intended as a follow-up of the work presented in Moscatello et al. (2021) and in Moscatello et al. (2022), we present a strategy which relies on machine learning algorithms aiming at further reducing the computational burden associated with the SB simulation and at the same time providing an high level of flexibility. In fact, machine learning algorithms, such as neural networks, are increasingly used in the safety field to predict gas dispersion paths with relatively low computational efforts Song et al. (2021). In particular, a strategy relying on Reduced order models (ROMs) Pedroni and Zio (2017); Pedroni (2022) is employed. This kind of approach reduces both the memory consumption and the calculation time, introducing some controllable approximations Benner et al. (2015) in the model response with respect to the high-fidelity model, the CFD model in this case. ROMs can be roughly divided in physics-driven and data-driven methods. The first ones yield an approximated model by manipulating directly the high-fidelity model Lorenzi et al. (2016, 2017). However, on top of the manifest difficulty in accessing the high-fidelity model equations in most of the commercial codes, these modifications would require a new code verification

and validation (V&V) phase. On the contrary, the second ones consider the code as a black-box providing the input-output data examples used to actually train the empirical regression model. The data-driven nature of such methods makes them non-intrusive and application-independent, hence they can be employed effectively even with commercial, validated codes Casenave et al. (2020). However, this aspect also brings some drawbacks. First, the data selection may bias the model prediction capabilities. Second, since no information on the full-order model (FOM) is explicitly available except its responses, the lack of important training data can seriously limit the ROM prediction capability Lassila et al. (2014). Third, it is very difficult to obtain an *a priori* error prediction Rahman et al. (2018). Nevertheless, the advantages of these methods are so attractive with respect to their shortcomings that they are becoming very popular. Adouze et al. were among the first ones to present a non-intrusive reduced order model (NIROM) for the solution of parametrized partial differential equations Audouze et al. (2013), but, to the best of our knowledge, the first NIROM application to the Navier-Stokes equations is due to Xiao Xiao et al. (2015), who proposed a method relying on Taylor expansion and another one based on sparse grid collocation. Since then, many approaches have been proposed to perform non-intrusive model reduction, especially for CFD applications Kumar et al. (2016); Demo et al. (2019). Therefore, the aim of the paper is to assess the NIROM effectiveness in reducing the computational cost associated with the SB simulations (i.e., to the generation of the spatial SB flow profiles). A successful model reduction of the release phase would represent the first step for a computationally smart strategy for safety-oriented, CFD-based analyses. Thanks to its non-intrusiveness, this approach could be applied to any validated code for the QRA of different industrial congested systems, like the forthcoming nuclear fusion plants and hydrogen installations. In this paper, our case of study concerns the accidental release of CH_4 on an Oil & Gas off-shore platform, simulated using the ANSYS Fluent code. Since the objective of our NIROM application is to approximate the behaviour of a Source-Box (SB) and to provide a set of flow fields for a subsequent (dispersion) simulation, it is of paramount importance to quantify the ROM approximation error and propagate it on the dispersion simulation output, i.e. risk oriented quantities such as the flammable volume, etc. In the works mentioned above, except for Xiao (2019), a little effort was devoted to this perspective. Therefore, in this work we propose a novel combination of the Proper Orthogonal Decomposition-Radial Basis Functions (POD-RBF) approach with

two statistical approaches, namely the bootstrap method Efron (1981); Kohavi (1995) and the unscented transform Julier and Uhlmann (1997), for an estimation of confidence intervals on the final output (i.e. the irreversible injuries area). The techniques to estimate the error bounds are quite general and could be applied to other NIROMs. In synthesis, from the methodological viewpoint the novel contribution of the present paper is twofold: i) the bootstrap method is combined for the first time (to the best of the authors' knowledge) with a (fast-running) POD-based NIROM to obtain an empirical (bootstrapped) distribution of a functional output, which allows building confidence in the ROM estimates (i.e., in the gas flow spatial fields); ii) the unscented transform is employed to propagate - with a relatively small number of long-running CFD simulation - the uncertainties in the spatial fields onto the final simulation output (i.e., released pollutant mass). From the applicative viewpoint, the efficient combination of statistical techniques mentioned above is employed for the first time in the quantitative analysis of a congested Oil & Gas off-shore platform. As a final remark, it is worth acknowledging that we do not aim at carrying out a complete, full-scale QRA of an energy plant, but rather at developing a versatile tool that can support it, in an efficient and computationally tractable fashion. The paper is organised as follows. In Section 2 we first discuss the main features of SBAM, then we present the POD-RBF method and the methodology to estimate the confidence interval on the output of the whole calculation chain. Then, in Section 3 we discuss the main simulation settings and we presents our main results and the relative discussion, as well. Finally, some concluding remarks and future perspectives are reported in Section 4.

2. The non-intrusive POD-RBF model

Basically all the physico-mathematical problems that may be posed in the field of physics and engineering can be compactly represented as the action of a certain model $\mathcal{M}$ on a P-dimensional input vector parameter $\vec{p}$,

$$\vec{y} = \mathcal{M}(\vec{p}), \tag{1}$$

where $\vec{y}$ is the m-dimensional output response, named *snapshot*. The model $\mathcal{M}$ hides both the model variables (e.g., space, time, energy...) and the different modelling and numerical approximations and complexities. Hence, it may be though as a black-box taking a certain input from the outside and

responding with a certain output. Incidentally, this is exactly the way any computer code aiming at solving any practical problem works. Therefore, in the following the notions of *physico-mathematical model* and *computer model* are assumed to be interchangeable.

2.1. CFD modeling: Source Box Accident Model

In our case, the computational model is a CFD approach to study the accidental high-pressure gas releases in industrial, congested plants, named SBAM. It was firstly introduced in Moscatello et al. (2021) and its performances against experimental data were presented in Moscatello et al. (2022). In order to focus on the peculiarities featuring this approach and the implications to realise a ROM simulation framework, this method will be briefly discussed in the following.

High-pressure gas releases, i.e. from 10 bar onward, in large, open environments (e.g. 30x20x5 m platform deck, see figure 6) always lead to a complex multi-scale and multi-physics phenomenon. In such cases, an under-expanded jet is generated near the release point Franquet et al. (2015), i.e. a supersonic highly compressible flow (Ma >0.3) (Munson et al., 2010, chapter 9), characterised by a strong discontinuity in the flow-field quantities. Such discontinuities are located at a specific distance from the release point, where a Mach disk appears: this is a normal shock where a steep variation of velocity, density, temperature and pressure is expected. In this region, near the release point, the flow is dominated by inertial effects, and buoyancy forces are negligible. If the high-pressure release occurs in a large environment, there is enough space for the gas to slow down from supersonic to subsonic velocity. Hence, far from the release point, the flow reaches a subsonic flow regime (Ma <0.3). At this point, the flow can be considered incompressible and it is no more inertia-dominated, i.e. buoyancy forces can be relevant.

These two phases can be named *release* (compressible, inertia-dominated) and *dispersion* (incompressible, subsonic). In Moscatello et al. (2021) it is demonstrated that splitting the entire phenomenon in this two pieces has some advantages from a modelling point of view, especially if the CFD simulation is QRA-oriented. This is the basis of SBAM, in which the release phase is simulated in a small domain, the SB, sized in a proper way to contain all the compressibility effects, and the dispersion in the environment under analysis, e.g., an off-shore Oil & Gas platform. The coupling of the two simulations is realised imposing the velocity flow field ($\tilde{\mathrm{v}}(\mathrm{x},\mathrm{y},\mathrm{z})$) and

gas concentration ($C_{gas}(x, y, z)$) distributions obtained on the SB faces as boundary conditions in the dispersion simulation.

This last step, as sketched in figure 1, allows to evaluate the mass and volume of the gas cloud resulting from the dispersion. Thanks to these results, some safety related quantities can be evaluated. In the case considered in this paper, which concerns a CH_4 release, the interests safety parameters are:

- Total dispersed CH_4 mass [kg], M_{CH4}
- CH_4 mass in the flammable cloud [kg], $M_{CH4,flam}$
- Flammable cloud volume [m^3], V_{flam}
- Irreversible Injuries (II) volume [m^3], V_{II}
- Irreversible Injuries area at 1.5 m height [m^3], A_{II}

Apart from the first parameter, the other ones are related to the flammability which is defined by the LFL (Low Flammability Limit) and the UFL (Upper Flammability Limit). For CH_4, these quantities are equal to 5 % and 15 % by volume of air, respectively Scheiner et al. (1985). $M_{CH4,flam}$ is a relevant parameter as it is a measure of the potential energy released in case of explosion. V_{flam} (which can be defined also as the High Lethality zone according to Italian regulations Dec) represents the region in which the gas concentration is between the LFL and UFL and a gas ignition can occur causing serious damage to the equipment and people death. The V_{II} and A_{II} are defined according to Italian regulations Dec, and are the one with a gas concentration above the 0.5·LFL, and in which some damages to equipment and people are expected. Since these scalar quantities are relevant for the QRA, they will be considered as the final output of the whole calculation chain. Nevertheless, in order to avoid any loss of generality, they will be indicated as $\vec{z}$, in order to remark that the methodology presented in the paper can be applied to spatial and/or dynamic distributions as well. This computational strategy allows to simplify the simulation of this complex, multi-physics phenomena, allowing to save considerably computational time when several scenarios are needed for a QRA. This is possible because each SB simulation available can be employed for several dispersion scenarios. As it will be explained later, the SB simulation can be performed by the ROM model, dramatically reducing the computational cost. In addition, the SB

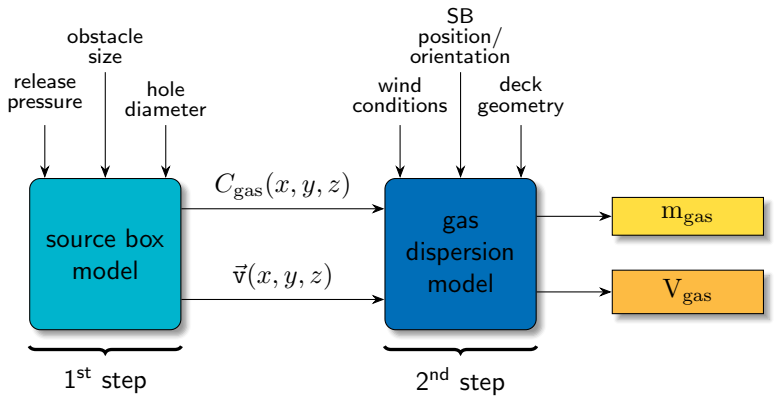

Figure 1: Sketch of the SBAM approach.

and the dispersion simulations are affected by different set of parameters. The first one is affected mainly by the gas properties, its pressure, the release hole size and the shape and orientation of the obstacles near the break, while the second depends on the congested plant configuration, the release position and direction and the wind velocity magnitude and direction. This suggested the development of a surrogate model to mimic the SB behavior, thus reducing dramatically the computational cost associated with SBAM.

This paper aims at proving the effectiveness of the NIROM approach to maximise the computational gain of this decoupled simulation approach, hence, it stands as a further development of SBAM, a model previously investigated in Moscatello et al. (2021, 2022). As this is a proof-of-concept, the pressure will be considered as the only SB varying parameter. It could be argued that the choice of only one out of all the possible varying parameters may jeopardise the considerations drawn throughout the paper about the accuracy, the robustness and the reliability of the proposed method. The gas release pressure, however, is the most relevant parameter from both the physical and the computational point of view, ranging continuously from 10 to 80 bar Vivalda et al. (2018), which is a very large interval from the fluid-dynamics perspective. On the contrary, the break size and the SB obstacle dimensions usually assume only few discrete values in the QRA framework, thus limiting the interest for such parameters from a ROM perspective. Moreover, the focus of our analysis is more related to prove the methodology proposed to quantify and propagate the NIROM uncertainty rather than

focusing on the parameter space sampling.

2.2. Reduced order model overview: offline and online phases

The reduced order model presented in this paper aims at approximating (with a reduced computational burden) the response $\vec{y}$ of a general physico-mathematical model $\mathcal{M}$ depending on a certain vector parameter $\vec{p}$. As previously mentioned, this is achieved with a data-driven approach. First, according to a suitable sampling strategy, the FOM is sampled to gather the high-fidelity responses, whose dimensionality is reduced via POD. The POD coefficients, representing the solutions in a reduced order space, are then used to train the RBF network. Afterwards, the ROM is validated on some untrained parameter values, in order to verify that the approximation error of the ROM is acceptable for the intended application. Finally, the bootstrap method is applied to propagate the ROM approximation error induced by the training samples choice. This study is performed by constructing a set of different ROMs, each trained with data sampled with replacement from the original training set.

Once this computationally expensive phase is completed, the ROM can be used as a fast-running tool to approximate the FOM responses on new parameter values. If the ROM response is employed as input for another model $\mathcal{M}'$, in our case the dispersion simulation, an estimate of the ROM approximation error can be obtained via the unscented transform (UT). This method estimates the uncertainty in the response of $\mathcal{M}'$ by means of a limited number of model evaluations. These steps are summarised in algorithm 1 and in figures 2-3.

2.3. Model sampling and reduction

The performances of data-driven models are strongly dependent on the quality of the training data. Therefore, whatever is the reduced order modelling approach, one of the most important steps is sampling the FOM.

One of the most popular deterministic sampling strategies in this framework is the adoption of sparse grids Smolyak (1963) since, with a proper choice of the quadrature rule used to map the input parameter space, different levels of nested samples can be obtained. The use of nested sets of points can help to efficiently train the ROM, allowing an *a posteriori* refinement of the parameter samples. Since this paper aims to present and test a non-intrusive reduced order modelling approach, the application presented later deals only with the most relevant input parameter, i.e. the release pressure.

Algorithm 1: POD-RBF with uncertainty estimation

Offline procedures

1. define the p-dimensional parameter space $\mathbb{R}^p$;
2. select a parameter space sampling strategy (i.e. sparse grids, random sampling...);
3. generate full-order model snapshots $\vec{y}_i \in \mathbb{R}^m$ for each parameter sample $\vec{p}_i$;
4. divide the data into the training and the validation sets;
5. reduce dataset dimensionality, using POD (see algorithm 2);
6. train the RBF net with the POD coefficients $\vec{a}_i \in \mathbb{R}^t$ (see algorithm 3);
7. apply the bootstrap method (see algorithm 5) to generate a set of ROMs;
8. compute the error distribution of the ROMs on the validation set;
9. validate the ROM on the validation set. If the average error between validation data and the set of bootstrapped ROMs is not acceptable, go back to step 3, adding more training points;

Online procedure

1. interpolate with RBF the POD coefficients over a new point (i.e. not used during training) $\vec{p}_j$ in $\mathbb{R}^p$;
2. back-project the POD coefficients $\vec{a}_j \in \mathbb{R}^t$ to get the approximated snapshot in the original space, $\vec{y}_j \in \mathbb{R}^m$.
3. if the ROM response is used as input for another model, apply the UT (algorithm 6) to estimate the confidence interval.

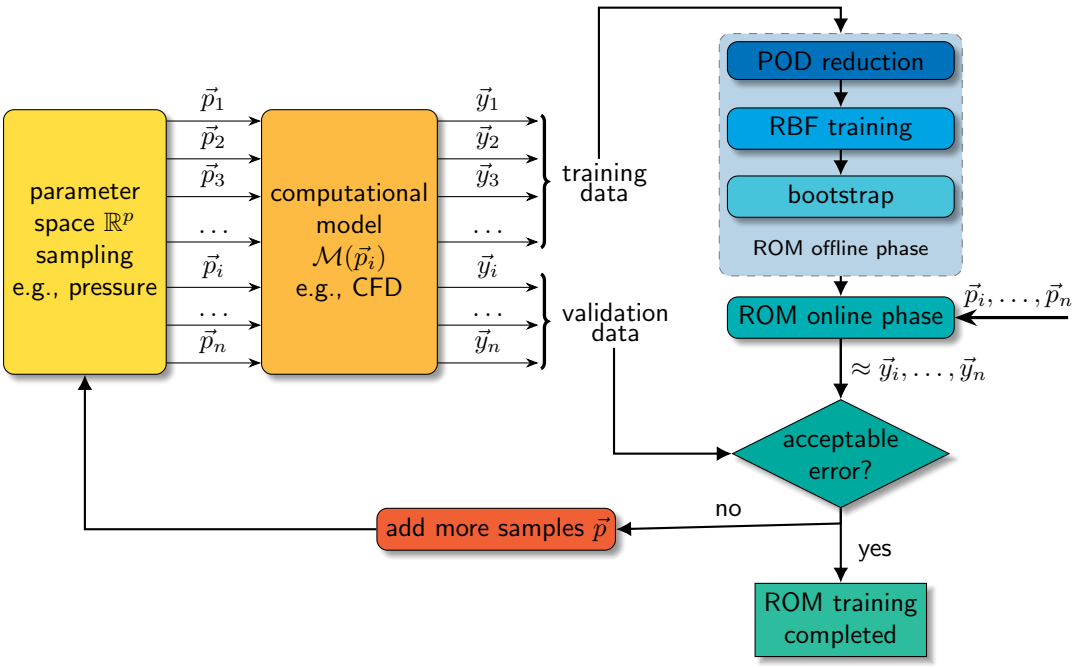

Figure 2: Offline phase procedure workflow.

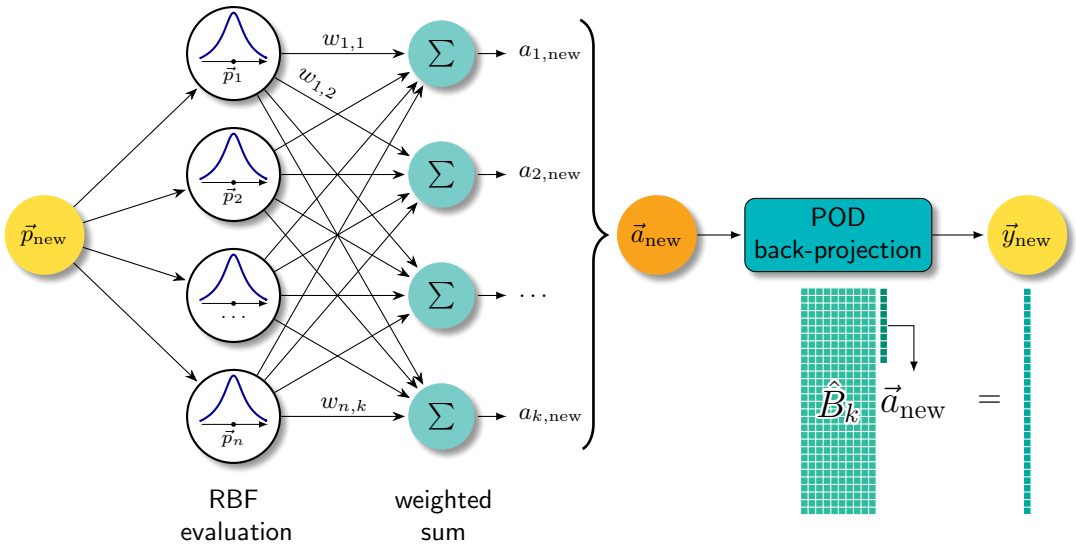

Figure 3: Online phase procedure workflow when a new parameter value $\vec{p}_{\text{new}}$ is provided to the ROM. The set of weights $w_{i,j}$ is computed in the training phase, see algorithm 3.

Therefore, a simple nested Newton-Cotes quadrature rule is employed for an adaptive selection of the training points.

After the FOM sampling, the data are reduced by means of the POD. The core idea of the POD method, described in algorithm 2 (see algorithm 2 in appendix Appendix A) and in fig. B.16, is to express the original data as an expansion of basis functions, known as POD modes, extracted with a singular value (eigenvalue) decomposition of the FOM snapshot matrix (correlation matrix). The main advantage of this technique is that the first modes retain most of the FOM dataset information. Therefore, truncating the POD expansion usually implies a limited loss of information Volkwein (2011).

The usual figure of merit employed to assess the number of POD basis functions needed to achieve a certain accuracy is the POD energy $\mathcal{E}_k$ (see algorithm 2). The calculation of this quantity is simple and cheap, but often not enough in order to fully characterise the POD expansion accuracy. Therefore, in this paper, the root mean squared error between the FOM output and the ROM estimates will be evaluated as well to have an additional figure of merit.

2.4. Model training and tuning

Once the reduction step is completed, the parameter-dependent POD coefficients are employed to train a network of RBFs, which can be then adopted in the online phase to interpolate the FOM solution on new parameter values.

Among the different types of radial basis functions, in this paper we choose to rely on the inverse multi-quadrics formulated by Hardy Hardy (1971). The value of these functions depend on two parameters. The first one is the euclidean distance between the centers, i.e. the training parameter values $\vec{p}_i$, and the collocation points, i.e. the new parameter $\vec{p}_j$, while the second one is the so-called hyperparameter σ, which is a free parameter determining the shape of the RBFs. The choice of σ is a very delicate aspect of the training phase, as it strongly affects both the interpolation accuracy and its numerical stability. Usually, the choice of this parameter is performed in order to minimise the interpolation error on some test points, which do not belong to the training set. Due to the large computational cost associated with the CFD simulations, this approach is too computationally expensive. Therefore, among the various approaches suggested in the literature, we follow the one which seems the most general, i.e. the Leave-One-Out-Cross-Validation (LOOCV) technique, which consists of training the meta-model with each of the N sets obtained taking N-1 samples and using the left one as a test point. In this framework, the best hyperparameter is the one minimizing the root-mean square error (RMSE) computed from the N trained models.

The RBF training phase is summarised in algorithm 3 (see appendix Appendix A), while the hyperparameter optimisation is presented in algorithm 4 (see appendix Appendix A).

2.5. Model validation and error estimation by a bootstrap-based ensemble of ROMs

When the training phase is completed, the model needs a validation on new parameter values to verify its accuracy and consistency with respect to the FOM solution. An extensive validation would require lots of FOM evaluations, jeopardising the overall computational efficiency of the meta-model. A common strategy to overcome this issue consists in being satisfied with the selection of a few, significant new parameter values not too close to the training points. This is the most popular approach in the literature when the NIROMs do not allow an *a priori* estimation of their confidence interval Rahman et al. (2018); Xiao et al. (2017).

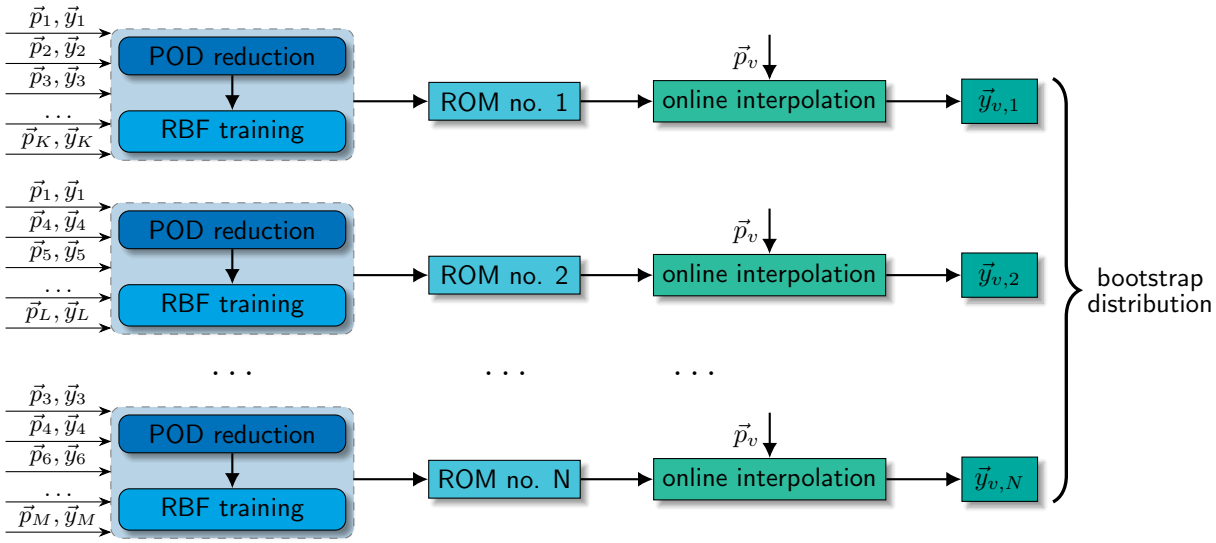

Figure 4: Sketch of the generation of the ROM distribution, and, consequently, of the output spatial fields via bootstrapping.

In this paper we employ the same approach, albeit trying to complement this limited validation with a statistical sensitivity study concerning the training set. As a matter of fact, the ROM responses are biased by the selection of the training points, whatever is the sampling strategy of the parameter space. Therefore, as a complement of the validation phase, we apply the bootstrap method for a full exploitation of the available training set, maximising the amount of information extracted. The idea of this non-parametric statistical method, presented in algorithm 5 (see appendix Appendix A) and sketched in fig. 4, consists in training a large number of ROMs using different training sets, obtained by resampling with replacement the original training set. In this way, it is possible to build an ensemble of meta-models, which can be used to construct a distribution of output responses: in the end, this can be employed to estimate the error distribution for each validation point Zio (2006); Secchi et al. (2008); Zio et al. (2010); Pedroni et al. (2010); Marelli and Sudret (2018).

The training phase of the NIROM is usually much cheaper than the offline phase, yet it may be still quite time consuming if the number of training parameters and/or the number of reduced order coefficients are large, because

of the RBF tuning process. However, since each model reboot is independent, the boostrap procedure can be massively parallelised. Due to the features of the POD-RBF ROM approach, some precautions are needed. First of all, since the RBF kernel becomes singular if the same training data is repeated, each time that a training case is resampled, it is deliberately ignored, thus the resampled set is always poorer than the original one. Then, due to the interpolatory nature of the RBF approach, some care should be devoted when the meta-model is tested on the validation set, because one or more validation samples could fall outside the sampled parameter space, i.e. the convex hull obtained from the sampled points. In these situations, the RBF net extrapolates rather than interpolating. In order to better understand the model behaviour in such cases, the bootstrap is applied in two different ways in our calculations. The first time, each validation point is accepted and evaluated by the ROM, which thus could extrapolate parameter values falling outside the training points, while the second time the validation points falling outside the training range are discarded. In this last case, no extrapolation occurs for parameter values drawn outside the parameter "box".

2.6. Unscented Transform and POD for uncertainty propagation and confidence interval estimation

As mentioned above, bootstrapping the ROM generates an ensemble of meta-models. If the ROM response is the final goal of its application, confidence intervals can be extracted from the ensemble distributions. However, when the ROM response $\vec{y}_{ROM}$ is an input for another model $\mathcal{M}'$, this operation may not be trivial, especially when $\vec{y}_{ROM}$ is a spatial field. In this case, the uncertainty in the ROM responses should be propagated through $\mathcal{M}'$ in order to obtain a confidence interval for the final output of the calculation chain. To avoid many queries of the model, it is proposed to approximate the confidence interval with the so-called Unscented Transform (UT) method Julier and Uhlmann (1997).

This method, which is an extension of the Kalman filter Kalman (1960) to non-linear models, approximates the original m-dimensional statistical distribution with a set of $2m+1$ specific samples, identified as *sigma points*, that are then transformed using the non-linear model, i.e. $\mathcal{M}'$ in this case. The transformed sigma points can be used to estimate the mean and the variance of the non-linear model due to the uncertain input.

In this case, the choice of the sigma points is not trivial, since the input distribution dimension is proportional to the number of volumes defining the

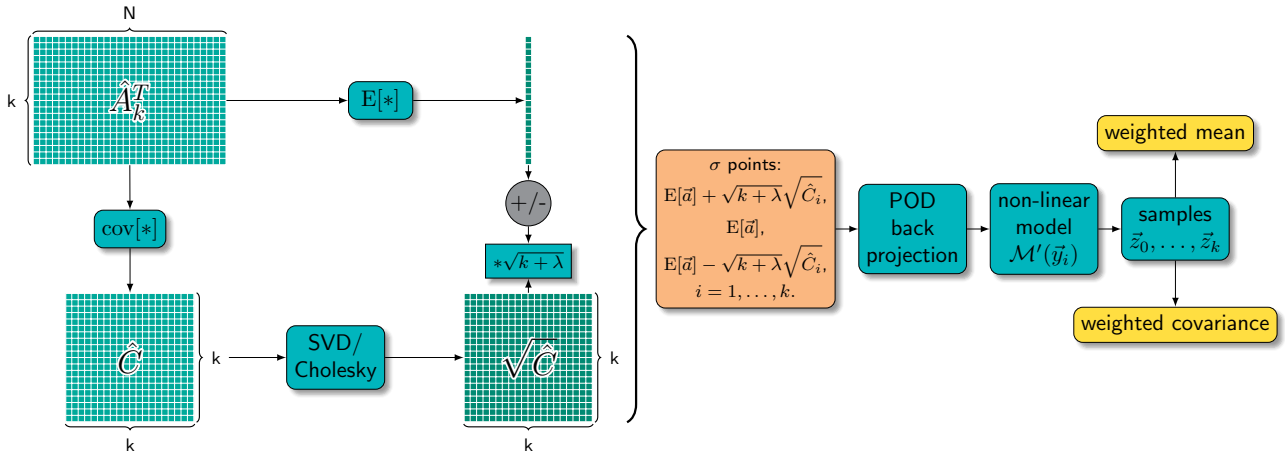

Figure 5: Sketch of the POD-UT approach for uncertainty quantification.

computational mesh of the CFD calculation. To overcome this issue, the POD algorithm is applied again to reduce the dimensionality of the ROM ensemble, similarly to Foad et al. (2020). The UT-POD procedure is presented in algorithm 6, while a conceptual scheme is presented in Figure 5.

3. Study case: accidental gas release in a congested environment

In this section the results of the NIROM application to the QRA-oriented, SBAM approach are presented and discussed, with a specific focus on the peculiarities featuring this case of study.

3.1. Numerical setup of the case study

The considered case study is a high-pressure methane release in an offshore platform deck under moderate wind conditions. The domain is visible in Figure 6 and its dimensions are 30x20x5 m. The gas release position is indicated by the blue box in Figure 6 which is placed at an height of 2.5 m from the ground, while, the release direction is described by the blue arrow near the box (along the x-axis). The possible release pressure (p_{rel}) range is 10-80 bar while the release diameter is fixed to 1 cm. These values are taken from Vivalda et al. (2018), where the plausible values for an high pressure gas release in an offshore environment are furnished basing on the review

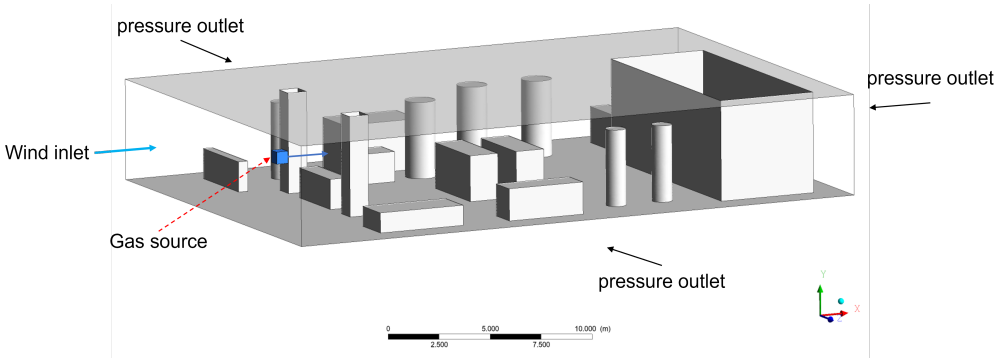

Figure 6: CAD of the case study domain with the representation of a box where the SB profiles are imposed, i.e. representing the gas release source

.

of loss of containment accident reports and statistical analyses. A uniform wind velocity profile is assumed, which intensity is equal to 6 m/s, the most frequent value in the Adriatic Sea RSE, and which direction is shown in figure 6 by the cyan arrow. This last is chosen in order to maximize the gas spreading on the platform, thus considering one of the worst scenarios.

The SB is dimensioned as a cube whose length (L_{SB}) is such that all the compressibility effects are exhausted in its volume Moscatello et al. (2021). Inside the cube, an obstacle is present to have an impinging jet. To ensure a robust coupling with the dispersion model, a fixed reference coordinate system and a fixed denomination for the SB faces, visible in Figure 7 (left), is employed. Within this reference frame, the release point is always positioned at the centre of the back face and the release direction is always parallel to x-axis in the SB. Moreover, since two symmetry planes can be defined, the simulation is carried out only on one quarter of the cube, using the denomination in Figure 7 (right). The obstacle inside the SB is featured by a diameter equal to 20 cm and a distance of 30 cm.

The SB dimensions, in principle, would change in function of the release pressure. The different size of the SB, however, can be an issue when the interesting profiles are extracted and used to train a NIROM. At first, the SB is dimensioned following the procedure explained in Moscatello et al. (2021), considering p_{rel}=80 bar (the largest possible). The resulting characteristic length of the SB is L_{SB}=0.6 m. A non-uniform tetrahedral mesh is realized in ANSYS meshing. The simulation is performed in steady-state using a k-ω SST formulation of the RANS equations. The mixture CH_4-Air is loaded

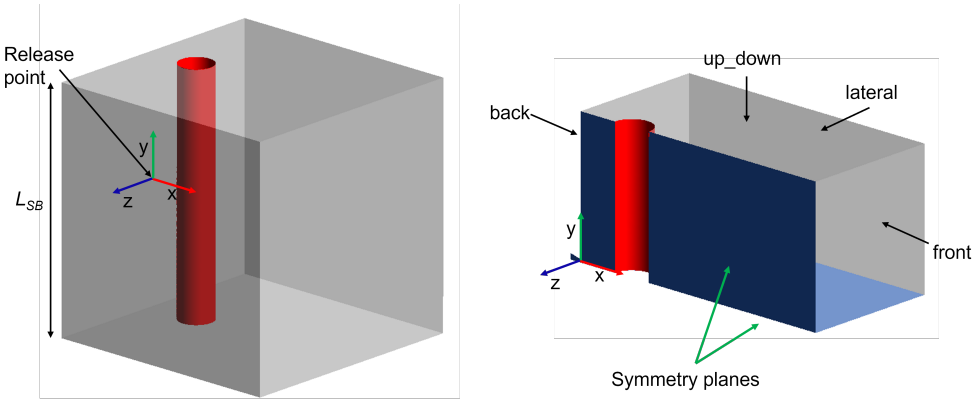

Figure 7: Source Box with reference coordinate frame and face names.

in the Fluent setup and the "Species Transport" model without chemical reaction is used. A CH_4 mole fraction equal to 1 is imposed at the domain inlet and a mass flow inlet is set considering a chock mass flow rate due to the pressure release condition. A pressure outlet set at atmospheric pressure is imposed on all the external SB surfaces back, up_down, front, lateral to reproduce the open environment around. A wall with no-slip condition is imposed on the obstacle surface and a symmetry condition is imposed on the symmetry planes.

In principle, the SB size scales as the square root of the release pressure. However, in order to have snapshots of the same dimensionality, each calculation was run using the same SB dimensions, the ones related to the maximum pressure. Then, the different velocity profile components (along x, y and z directions) and CH_4 mass fraction profiles on the faces delimiting the SB (back, lateral, front, up-down) were exported from Fluent on a fixed cartesian grid consisting of 1000 sample points for each direction. These profiles are used as boundary conditions in the dispersion simulation, where they are imposed at the faces of the box (see blue box in figure 6) representing the gas source in the plant. On this box, a sufficiently refined mesh to maintain the spatial information of the profiles is realized. Each dispersion simulation requires about 4 hours on a Precision Dell Tower 7820 with a Intel Xeon Gold 6136 CPU (3.00GHz) and 64 RAM.

The dispersion simulation is performed in steady-state with a k-ω standard formulation of the RANS equation. A wind with intensity of 6 m/s and direction along the x-axis is modelled on the face indicated in figure 6. On

the other lateral faces of the domain, a pressure outlet with ambient pressure is imposed. The blue box represents the dispersion source: on its faces the velocity and CH_4 mass fraction profiles obtained by the SB simulations are loaded as boundary conditions. All the other platform surfaces, i.e. the objects, the floor and the ceiling, are modelled as walls with no slip condition.

3.2. Sampling strategy and CFD dimensionality reduction

Despite in this application the parameter space is one-dimensional, its variation is remarkable from both the physical and the numerical point of view. Increasing the pressure requires more computational resources, as the formation of vortexes is enhanced. Therefore, the sampling strategy should be carefully selected to adequately cover the release pressure range (10-80 bar), which is expected to induce very large variations in the resulting flow fields. As mentioned in the previous Section, the parameter values for the CFD snapshot generation were chosen with the Newton-Cotes rule.

It is important to remark that to carry out SB simulations at high pressure, the mass flow inlet boundary conditions must be used instead of imposing directly the pressure. Hence, mass flows are evaluated consistently with the needed release pressures considering the chocked flow conditions. Since the pressure is the reference engineering parameter for QRA, the mass flow was sampled, but the corresponding value of the pressure, which is unique due to the sonic flow regime, was considered as the free parameter $\vec{p}$, exploiting the fact that the CFD model is considered as a black-box.

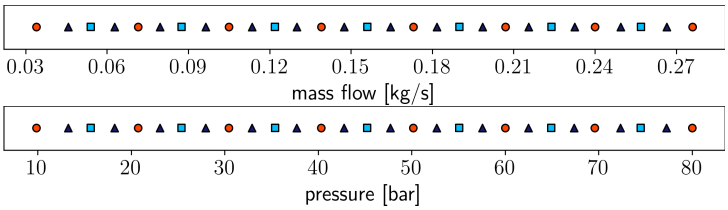

Figure 8: High-fidelity model samples generated with the Newton-Cotes rule. The orange circles are the first level, the light blue squares are the second level and the dark blue triangles are the third level.

Figure 8 provides a sketch of the distribution of the inlet mass flow samples and of the corresponding pressure samples. As a starting point, an initial level was generated dividing the range to have intervals of about 10 bar each (red dots). Then, the range was partitioned to have intervals of approximately 5 bar each, so that the previous level was included (light blue

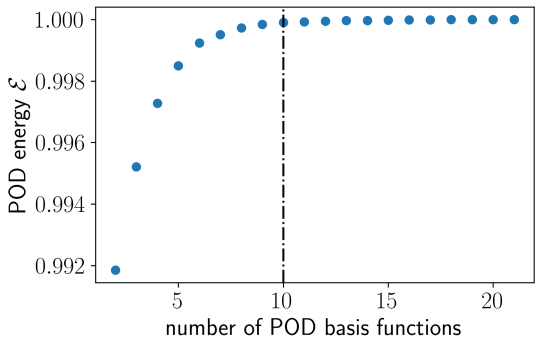

Figure 9: POD basis energy, computed as in algorithm 2.

squares). Finally, the intervals width was halved again, in order each simulation to span 2.5 bar.

Concerning the dimensionality reduction, a sensitivity study on the approximation error induced by the POD expansion truncation shows that a relatively small number of basis functions is sufficient to adequately represent the original gas concentration and flow field, catching the 99.99 % of the POD energy $\mathcal{E}$, whose convergence trend is provided in Figure 9. Figure B.17 shows the Root-Mean-Square Error (RMSE) between the CFD snapshots and its POD representation (left) and the RMSE between the CFD snapshot and its ROM approximation (right) as a function of different truncation orders, for different values of the training pressures. The RMSE on the left is useful to highlight that the truncation error approaches zero as the number of basis functions increases, as reasonably expected looking at the trend in Figure 9. On the contrary, the second one shows that, after a certain expansion order, the error stabilises on a certain value. This behaviour is explained by the fact that, in addition to the vanishing truncation error, the NIROM snapshot is also affected by the interpolation error due to the RBF interpolation. This trend has an important implication related to the NIROM approximation capabilities, i.e. the truncation error is negligible with respect to the interpolation error, provided that a sufficient number of basis functions is employed to represent the original dataset.

Throughout the paper all the POD expansions are truncated at least at the 10^{th} term, unless differently specified, as indicated by the dashed black line in Figures 9-B.17.

3.3. Model training

Exploiting the fact that the samples are nested, the NIROM was initially trained with the first level of points and validated with the points belonging to the second level.

Figures 10 to B.20 provide an overview of the percentage relative error between NIROM and CFD evaluations of the output distributions of interest, namely the CH_4 concentration and the x, y and z components of the velocity field. The relative percentage error, computed for each face of the SB (see Figure 7), is evaluated as follows

$$\epsilon_i = 100 \frac{||y_{\mathrm{i,CFD}}(\vec{r}) - y_{\mathrm{i,ROM}}(\vec{r})||}{||y_{\mathrm{i,CFD}}(\vec{r})||}, \tag{2}$$

where $\vec{r}$ indicates the couple of spatial coordinates pertaining to the i-th face. It should be noticed that, consistently with the notation used in the paper, "y" indicates the model response, while "y" is the usual symbol for one of the three spatial variables of the cartesian reference system.

The black stars visible in figs. 10 to B.20 are the training points belonging to the first level of sampled points, while the dots represent the validation points. The size and the colour of these points is proportional to the magnitude of the error.

By inspection of these figures, it is possible to notice the presence of significant errors on the back face, where the profiles are featured by some oscillations due to the gas entrainment with the air which causes a steep velocity variation near the leak source. Due to the relative error definition given in eq. (2), it may happen that the error may get very large values, exceeding 100%, when $||y_{\mathrm{i,CFD}}(\vec{r})||$ is very close to 0, as for the y and z components of the velocity in the back face. Despite their magnitude, these errors do not affect the overall quality of the NIROM approximations, because the back plane, which is tangent to the source point, has a negligible contribution to the overall mass flowing out of the sourcebox. In spite their lower physical importance for the dispersion phase, the contributions for this face dominates the overall error behaviour, making the L_2 norm on the whole snapshot unreliable. To overcome this issue, the face-wise errors are weighted with the mass fraction flowing from each face,

$$\varepsilon = \sum_{i=1}^{faces} \frac{||y_{\mathrm{CFD,i}}(\vec{r}) - y_{\mathrm{ROM,i}}(\vec{r})||}{||y_{\mathrm{CFD,i}}(\vec{r})||} w_i, \tag{3}$$

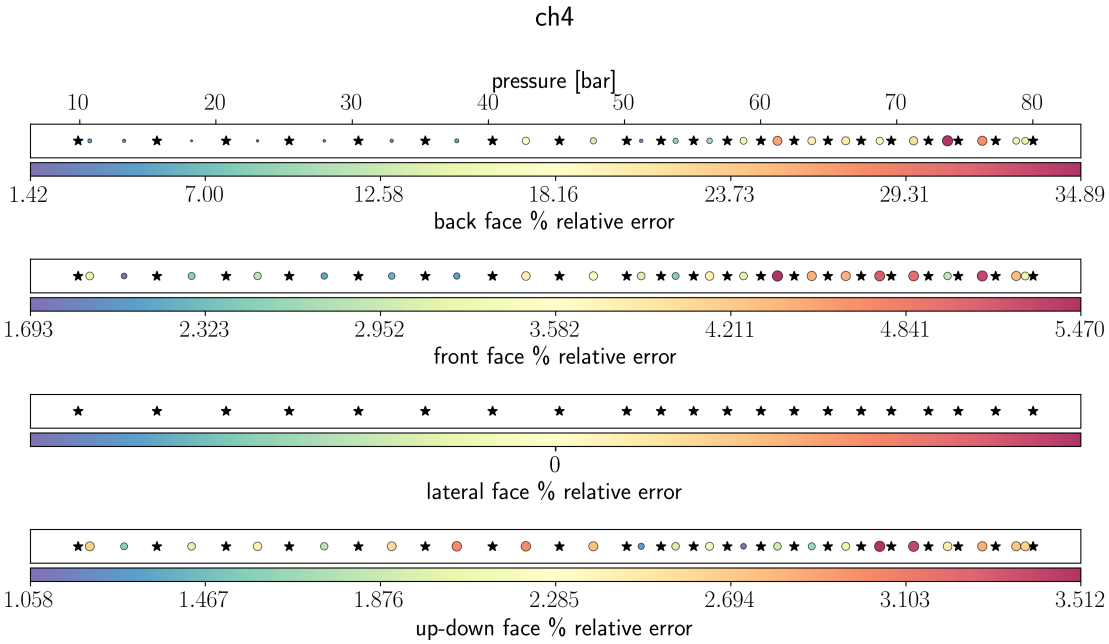

Figure 10: Face-wise relative L_2 error for the CH_4 concentration. The black stars represent the training cases, while the dots represent the validation cases and their color is related to the relative error.

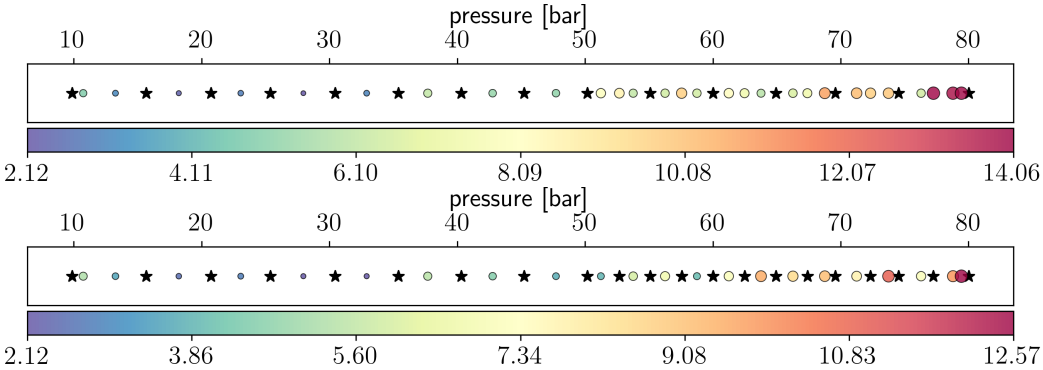

Figure 11: Weighted percentage relative error between NIROM and CFD using two different sets of training points in the two graphs. The black stars represent the training cases, while the dots represent the validation cases and their color is related to the relative error.

where the weights are defined as

$$w_i = \frac{\dot{m}_i}{\dot{m}_{tot}}. \tag{4}$$

Concerning the CH_4 concentration profile, it is important to notice here that the error on the lateral face cannot be measured since no gas reaches this face.

Exploiting the estimator given by Equation 3, it is possible to combine the relative errors for the different profiles on each face to get a more realistic, meaningful and physically reliable overview of the NIROM accuracy with respect to the reference solution. The values of this estimator is reported in Figure 11. The graph on the top represents the weighted relative error obtained using the first level of the sampled points as training and the other levels as validation, while the graph on the bottom provides the same estimator using the first level and some points of the second for training and the remaining points for validation. As it can be noticed, more training points were taken above 50 bar, in order to reduce the relative error in this region. If the parameter space would be high-dimensional, a more rigorous adaptive selection technique could be employed to spare some computational time, as in Alsayyari et al. (2021).

An example of the ROM surrogate ability to mimic the response of the SBAM model is visible in Figures 12, B.21 and B.22, where it is possible to see the CH_4 mass fraction on the front face and the normal velocity component on

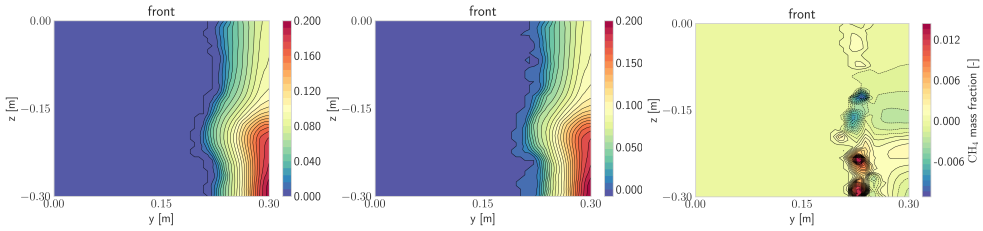

Figure 12: CH_4 mass fractions for CFD (left) and NIROM (centre) and their difference (right) on the front face for the validation case with p_{rel}=51.237 bar.

the front and up_down faces, respectively. Each Figure provides the reference profile, computed with CFD, the NIROM surrogate profile and the local difference between the two. By direct inspection, it is possible to conclude that the NIROM is able to reproduce fairly well the main spatial features of the flow field, with negligible errors except for a few, small regions on some faces, where oscillations around zero due to entrainment may occur, as already discussed for the back one.

In light of these considerations and having in mind that are referred to the worst case, the orders of magnitude of the errors reported in Figure 11 are judged sufficiently low for the purpose of QRA. In what follows, the analysis is thus focused on another source of (model) uncertainty in the ROM estimation, i.e. the error variability due to the training point selection.

3.4. ROM error estimation by bootstrap

The error distribution for each validation case is constructed rebooting the NIROM 500 times, by sampling with replacement the initial set of 21 training points.

As mentioned in Section 2.5, some care should be used during the evaluation of the bootstrapped error distribution, checking whether the validation case falls in the training range or not. In the following, only the interpolation situation is examined, since the NIROMs, due to their data-driven nature, are usually employed as interpolating tools. It should be remarked here that, with respect to the deterministic rule used to generate the training and validation samples for our application, there is no guarantee that the bootstrap random resampling covers uniformly the parameter space, therefore this approach should yield, in this particular case, conservative confidence intervals.

Figures B.23 and B.24 show the sample distributions of the weighted percentage relative error and RMSE for the validation cases featured by

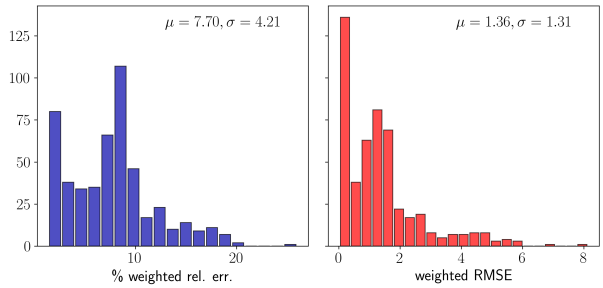

Figure 13: Weighted relative percentage error (left) and weighted RMSE (right) computed for the validation case p_{rel}=51.237 bar.

the minimum and maximum pressure, respectively, i.e. 10.374 and 79.442 bar. These graphs are quite informative about the method robustness: the distributions for both the cases are highly skewed towards lower bounds of the error, suggesting that the NIROM is weakly sensitive to the selection of the finite-sized set of training points. Figure 13 provides the distributions for the validation case p_{rel}=51.237 bar, which is featured by the largest variance and mean error. Due to its larger sensitivity to the training sample choice with respect to the others, this case is identified as the worst one, therefore it will be used in the following section to propagate the ROM approximation error through the dispersion calculation chain by means of the UT. It should be noticed that, despite in fig. 11 the cases featured by the largest weighted relative error are those falling in the interval 70-80 bar, the bootstrap analysis shows that the validation case which is the most sensitive to the training sample selection is the one featured by p_{rel}=51.237 bar. This case is thus addressed as the worst one in the following section.

3.5. Uncertainty propagation to the dispersion simulation

Propagating the bootstrap-evaluated uncertainty characterising the profiles computed by the NIROM onto the output of the successive CFD dispersion simulation is not trivial. In addition to the large dimensionality of the uncertain input (each snapshot contains $> 6 \cdot 10^6$ elements), the bootstrapped profiles are not associated with real, physical parameters, but only to the set of training points. Thus, the empirical nature of this distribution makes difficult to adopt a smart sampling strategy. A possible option could be using a brute force technique, evaluating the dispersion model with each bootstrapped ROM response. However, despite the dispersion simulations

are faster than the SB calculations (∼5 hours against ∼25), this would not be practically feasible.

To overcome these issues, the procedure illustrated in algorithm 6 is employed. The POD technique helps reducing the dimensionality of the SB gas flow profiles, providing a set of t scalar coefficients for each profile, while the UT allows to select only the most relevant coefficients for the final uncertainty estimation. Consistently with the POD reduction for the training phase, also in this case the POD is truncated at the 10^{th} order, yielding $k = 2t = 20$ sigma points. For each of this point, a dispersion calculation is performed. In particular, the scenario with the largest variance and mean error (p_{rel}=51.237 bar) is investigated. In the following, some safety-critical output parameters estimated by a dispersion calculation are analysed, comparing the results obtained using the high-fidelity, CFD profiles in one case and the approximate, ROM profiles in the other one. Figure 14 shows the irreversible injuries volumes obtained using respectively the CFD SB profiles and the NIROM SB profiles are represented. At a first look, no differences can be appreciated in the two pictures. In both cases the gas cloud tends to split in two portions along the vertical direction, and the same platform components are invested by the gas. Some small discrepancies can be appreciated in one of the two extremity of the upper region and it is difficult to observe any other difference in the shape of the two clouds. In fact, in the case of the CFD profile, a V_{II} equal to 22.2 m^3 is obtained, while in the NIROM case is 22.4 m^3, confirming that the difference is negligible ($\sim 0.9\%$).

This qualitative comparison can be helpful in verifying if there are some relevant differences in the dangerous cloud shape: however, a more detailed analysis needs to be carried out by comparing some safety related quantities. For the purposes of a QRA, the evaluation of the mass and volumes involved in the accident is fundamental for the estimation of the energy that can be released in case of fire or explosion. In addition, a QRA requires to estimate the dangerous zones extension in terms of volumes and areas. For these reasons, the parameters introduced in section 2.1 are evaluated. Table 1 provides the main output quantities computed within the dispersion calculation using the original CFD profiles, provided by the SB simulation, and the surrogate profiles computed with the NIROM. With the aim of providing a visual representation of these quantities and their uncertainties, the same quantities are normalised with respect to the CFD case and shown in fig. 15. The results in the NIROM column, i.e. the mean and the standard deviation of each response, are obtained from the 20 dispersion calculations, exploiting

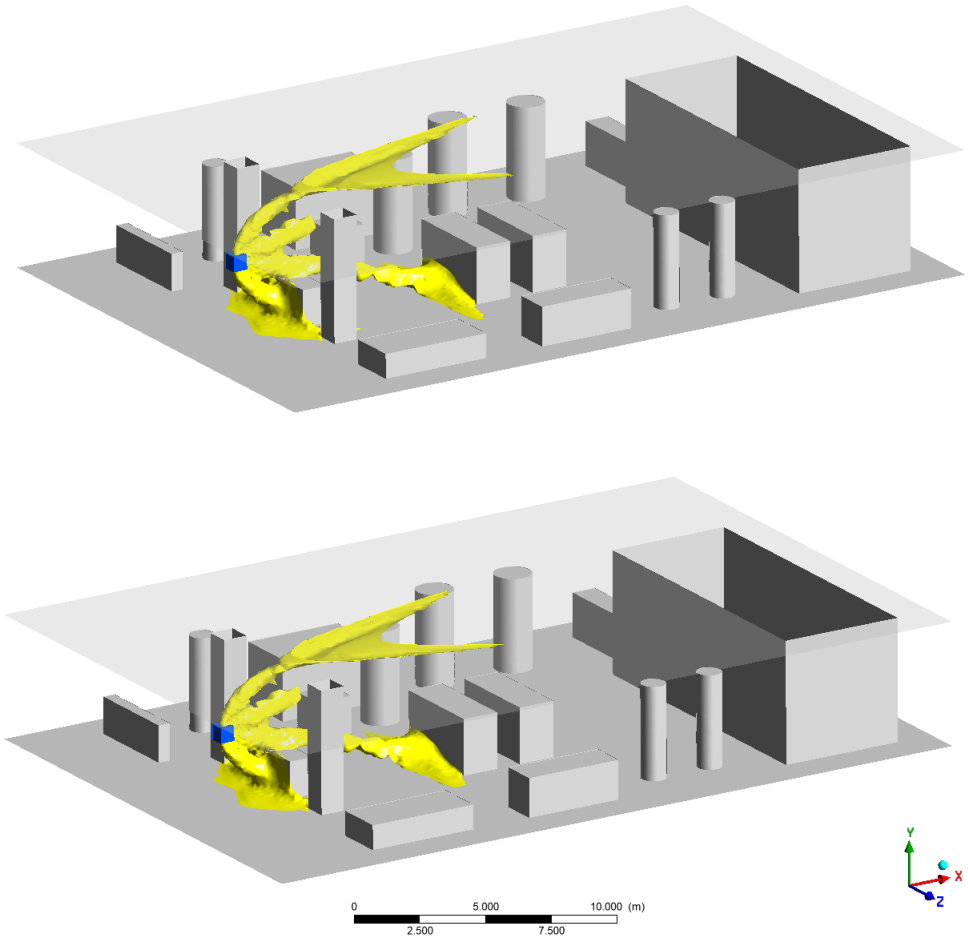

Figure 14: Irreversible Injuries volume obtained using the CFD SB profiles (top) and of the NIROM SB profiles (bottom).

Table 1: Dispersion calculation outputs computed with the original CFD profiles and with the surrogate ones given by the NIROM. The results of the second column are provided with an uncertainty, put in parentheses, given in terms of 1 standard deviation. The last column shows the absolute value of the relative difference per each parameter.

	CFD profile	ROM profile	Rel. difference
M_{CH4} [kg]	5.959	5.6(2)	∼6 %
$M_{CH4,flam}$ [kg]	0.066	0.064(4)	∼3 %
V_{flam} [m^3]	1.925	1.8(1)	∼6.5 %
V_{II} [m^3]	22	23(1)	∼4.5 %
A_{II} [m^2]	4.20	4.5(3)	∼7 %

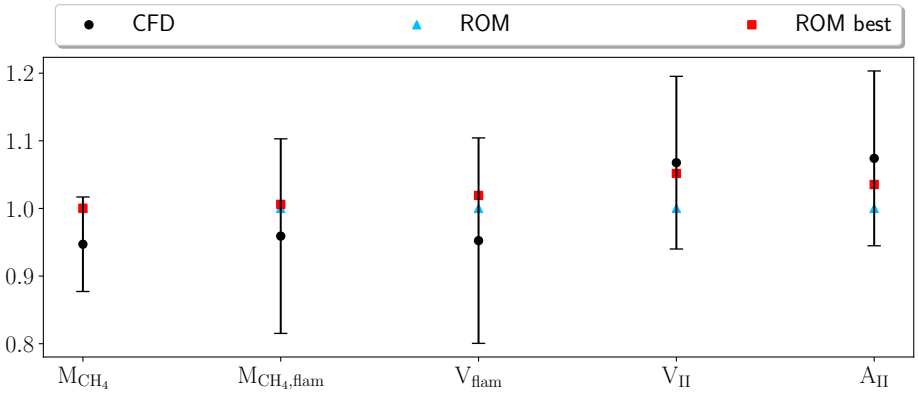

Figure 15: Expected value and standard deviation for some safety-relevant quantities, i.e., the total dispersed mass of CH_4, the mass of CH_4 in the cloud, the flammable volume, the irreversible injuries volume and the irreversible injuries area at 1.5 m. Each data is normalised with respect to the CFD case.

the UT. The results obtained using the NIROM SB profiles are surprisingly similar to the CFD related ones. The relative difference in the mean values is always below 7%, and this is a remarkable result if we consider the dramatic computational cost reduction. In fact, to obtain the CFD SB profiles, almost 24 h are employed, while the ROM profiles are obtained in few seconds. This improvement permits to cut the time necessary for the SB library generation, since it needs only the simulation of a reduced set of cases for the training of the ROM. Moreover, since some parameters are slightly underestimated but from a safety point of view an overestimation of the accident consequences is preferable, a safety coefficient can be applied to the results to assure a conservative estimation.

4. Conclusions and future perspectives

In this paper we presented a POD-RBF framework for the non-intrusive reduced order modelling of QRA-oriented CFD simulations. This class of NIROMs has been recently applied in many research fields, yet most of the applications do not show satisfactory assessment and propagation of the ROM approximation errors, in particular in the presence of functional (e.g., time- and/or space-dependent) outputs.

In this respect, since the final aim is to employ these surrogate models to reduce the computational burden associated with QRA activities (where it is of paramount importance to endow the simulation model responses with confidence intervals), we proposed a methodology for the assessment of the ROM uncertainty in the estimation of functional (in this case, space-dependent) safety oriented quantities. To test this methodology and prove its effectiveness, we applied the ROM to the case of a methane high pressure, accidental gas release in an off-shore Oil & Gas plant, where the output quantities of interest, from a QRA perspective, are volumes, areas and masses related to flammability of the gas.

First, the original CFD model was run on a set of nested parameter samples, which allowed to adaptively refine the training dataset. Afterwards, the model evaluations were used to train the ROM: first, the dimensionality of the spatial fields was reduced thanks to the POD algorithm; then, a net of RBFs was constructed by tuning the RBF hyperparameter with the LOOCV technique. A first assessment of the model accuracy was conducted comparing the ROM and the CFD evaluations on a set of points not belonging to the training set. The validation error, computed as a weighted sum of the SB face-wise errors, was considered sufficiently low ($<15\%$) using the first level of training cases plus additional cases above 50 bar.

Then, to estimate the model response variability to the training samples, the bootstrap method was employed to obtain the statistical distributions from the ensemble of ROM responses. Since, in our application, the ROM response is used as input for the CFD dispersion simulation, a quantitative assessment of the impact of the ROM approximation error was considered mandatory.

The propagation of the uncertainty from the ROM response distribution to the dispersion model output was carried out coupling the POD and the UT techniques. The first allowed to reduce the input data dimensionality, while the second was used to select the input data for the dispersion calculations

and finally to estimate the dispersion output confidence intervals. The main quantities of interest in such calculations, pertinent to the QRA analysis, are in very good agreement with the same results obtained using the SB profiles computed by CFD, with a relative error between the two approaches below 7% and a reduction of the computational time of about three orders of magnitude, suggesting that the POD-RBF NIROM is adequate to obtain fast yet very accurate results. It should be remarked that a QRA study has an intrinsically high level of uncertainty, which makes the additional 7% introduced by the ROM model acceptable. Moreover, this error is only a quantification of the impact of the NIROM integration in the numerical model SBAM, and it is not representative of the simulation error. In fact, for this last, the uncertainties related to the spatial discretization, round-off errors and numerical schemes should be quantified.

In the future, we plan to increase the number of training points considering also other input parameters affecting the SB, namely the break size and the obstacle features. Since these parameters have a strong influence on the SB dimension, a more sophisticated strategy should be devised in order to handle the snapshots defined on a different spatial domain. Moreover, adaptive sparse sampling techniques should be employed in order to progressively refine the parameter space where the error is not acceptable.

In parallel to these activities, a surrogate model for the dispersion phase should be trained as well. Such a model could be efficiently coupled with the first NIROM, allowing to realise a real-time simulation framework (featured by the capability of providing confidence intervals on the main results thanks to the combination of bootstrapping and UT), setting the basis for an effective CFD-QRA integration.

Credit authorship contribution statement

Nicolò Abrate: Conceptualization, Methodology, Software, Investigation, Writing - original draft, Writing- Review & Editing Visualization, Data curation. **Alberto Moscatello**: Conceptualization, Methodology, Software, Writing - original draft, Data curation. **Gianmario Ledda**: Conceptualization, Methodology, Software, Writing - original draft, Data curation. **Nicola Pedroni**: Conceptualization, Methodology, Supervision, Writing- Review & Editing. **Federica Carbone**: Software, **Emanuela Maffia**: Software, **Andrea Carpignano**: Conceptualization, Resources, Writing- Review & Editing, Supervision, Project administration, Funding acquisition

Acknowledgements

The research presented in this paper has been sponsored by the Italian Ministry of Economic Development's General Directorate for Infrastrucutres and Safety of Energy and Geomining Systems (DGISSEG).

References

Decreto del presidente del consiglio dei ministri. linee guida per la predisposizione del piano d'emergenza esterna di cui all'articolo 20, comma 4, del decreto legislativo 17 agosto 1999, n. 334e. 2005. https://www.anci.it/wp-content/uploads/linee-guida-piano-emergenza-esterna.pdf. Accessed: 2022-12-19.

Atlante eolico. http://atlanteeolico.rse-web.it/. Accessed: 2021-11-10.

F. Alsayyari, Z. Perkó, M. Tiberga, J. L. Kloosterman, and D. Lathouwers. A fully adaptive nonintrusive reduced-order modelling approach for parametrized time-dependent problems. *Computer Methods in Applied Mechanics and Engineering*, 373:113483, 2021. ISSN 00457825. doi: 10.1016/j.cma.2020.113483. URL https://doi.org/10.1016/j.cma.2020.113483.

C. Audouze, F. De Vuyst, and P. B. Nair. Nonintrusive reduced-order modeling of parametrized time-dependent partial differential equations. *Numerical Methods for Partial Differential Equations*, 29:1587–1628, 2013. doi: https://doi.org/10.1002/num.21768. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/num.21768.

Q. A. Baker, M. J. Tang, E. A. Scheier, and G. J. Silva. Vapor cloud explosion analysis. *Process Safety Progress*, 15:106–109, 1996.

H. A. Becker, H. C. Hottel, and G. C. Williams. The nozzle-fluid concentration field of the round, turbulent, free jet. *Journal of Fluid Mechanics*, 30(2):285–303, 1967. doi: 10.1017/S0022112067001430.

P. Benner, S. Gugercin, and K. Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review*, 57:483–531, 2015. doi: 10.1137/130932715. URL https://doi.org/10.1137/130932715.

A. D. Birch, Brown D. R., M. G. Dodson, and F. Swaffield. The structure and concentration decay of high pressure jets of natural gas. *Combustion Science and Technology*, 36(5-6):249–261, 1984. doi: 10.1080/00102208408923739. URL https://doi.org/10.1080/00102208408923739.

J. Casal. *Industrial safety Series: Evaluation of the Effects and Consequences of Major Accidents in Industrial Plants*, volume 8. Elsevier Science, 2008.

F. Casenave, A. Gariah, C. Rey, and F. Feyel. A nonintrusive reduced order model for nonlinear transient thermal problems with nonparametrized variability. *Advanced Modeling and Simulation in Engineering Sciences*, 7, 2020. ISSN 22137467. doi: 10.1186/s40323-020-00156-3.

C. J. Chen and W. Rodi. Vertical turbulent buoyant jets: a review of experimental data. *NASA Sti/Recon Technical Report A*, 80:23073, 1980.

C. Colombini, G. Maugeri, G. Zanon, R. Rota, and V. Busini. Unignited high-pressure methane jet impinging a pipe rack: practical tools for risk assessment. *Journal of Loss Prevention in the Process Industries*, 69:104378, 2021.

C. Colombini, E. Carminati, A. Parisi, R. Rota, and V. Busini. Safety evaluations on unignited high-pressure methane jets impacting a spherical obstacle. *Journal of Loss Prevention in the Process Industries*, 74:104631, 2022a.

C. Colombini, C. Iannantuoni, R. Rota, and V. Busini. Unignited high-pressure methane jet impacting a cylindrical obstacle: An assessment tool for consequences analysis. *Journal of Loss Prevention in the Process Industries*, 76: 104593, 2022b.

A. Dasgotra, G.V.V. Varun Teja, A. Sharma, and K. B. Mishra. CFD modeling of large-scale flammable cloud dispersion using FLACS. *Journal of Loss Prevention in the Process Industries*, 56:531–536, 2018. ISSN 0950-4230. doi: https://doi.org/10.1016/j.jlp.2018.01.001. URL https://www.sciencedirect.com/science/article/pii/S0950423017301596.

G.A. Davidson. Simultaneous trajectory and dilution predictions from a simple integral plume model. *Atmospheric Environment (1967)*, 23:341–349, 1989. ISSN 0004-6981. doi: https://doi.org/10.1016/0004-6981(89)90582-9. URL https://www.sciencedirect.com/science/article/pii/0004698189905829.

A. O. de Souza, A. M. Luiz, A. T. P. Neto, A. C. B. de Araujo, H. B. da Silva, S. K. da Silva, and J. J. N. Alves. CFD predictions for hazardous area classification. *Chinese journal of chemical engineering*, 27:21–31, 2019.

N. Demo, M. Tezzele, and G. Rozza. A non-intrusive approach for the reconstruction of POD modal coefficients through active subspaces. *Comptes Rendus - Mecanique*, 347:873–881, 2019. ISSN 16310721. doi: 10.1016/j.crme.2019.11.012. URL https://doi.org/10.1016/j.crme.2019.11.012.

M. Derudi, D. Bovolenta, V. Busini, and R. Rota. Heavy gas dispersion in presence of large obstacles: selection of modeling tools. *Industrial & Engineering Chemistry Research*, 53:9303–9310, 2014.

DNV. PHAST. https://www.dnv.com/software/services/plant/consequence-analysis-phast.html. [Online; accessed 21-June-2022].

B. Efron. Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*, 68:589–599, 1981. ISSN 00063444. URL http://www.jstor.org/stable/2335441.

B. Foad, A. Yamamoto, and T. Endo. Efficient uncertainty quantification for PWR during LOCA using unscented transform with singular value decomposition. *Annals of Nuclear Energy*, 141:107341, 2020. ISSN 18732100. doi: 10.1016/j.anucene.2020.107341. URL https://doi.org/10.1016/j.anucene.2020.107341.

E. Franquet, V. Perrier, S. Gibout, and P. Bruel. Free underexpanded jets in a quiescent medium: A review. *Progress in Aerospace Sciences*, 77:25–53, 2015. ISSN 0376-0421. doi: https://doi.org/10.1016/j.paerosci.2015.06.006. URL https://www.sciencedirect.com/science/article/pii/S0376042115000548.

R. L. Hardy. Multiquadric equations of topography and other irregular surfaces. *Journal of Geophysical Research (1896-1977)*, 76:1905–1915, 1971. doi: https://doi.org/10.1029/JB076i008p01905. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JB076i008p01905.

S. J. Julier and J. K. Uhlmann. New extension of the Kalman filter to nonlinear systems. In Ivan Kadar, editor, *Signal Processing, Sensor Fusion, and Target Recognition VI*, volume 3068, pages 182–193. International Society for Optics and Photonics, SPIE, 1997. doi: 10.1117/12.280797. URL https://doi.org/10.1117/12.280797.

R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82:35–45, 1960. ISSN 0021-9223. doi: 10.1115/1.3662552. URL https://doi.org/10.1115/1.3662552.

R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'95, page 1137–1143, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc. ISBN 1558603638.

D. Kumar, M. Raisee, and C. Lacor. An efficient non-intrusive reduced basis model for high dimensional stochastic problems in CFD. *Computers and Fluids*, 138:67–82, 2016. ISSN 00457930. doi: 10.1016/j.compfluid.2016.08.015. URL http://dx.doi.org/10.1016/j.compfluid.2016.08.015.

T. Lassila, A. Manzoni, A. Quarteroni, and G. Rozza. Model Order Reduction in Fluid Dynamics: Challenges and Perspectives. In Quarteroni A., Rozza G., editor, *Reduced Order Methods for Modeling and Computational Reduction. MS&A - Modeling, Simulation and Applications*, volume 9, chapter 10. Springer, Cham, 2014. doi: https://doi.org/10.1007/978-3-319-02090-7_9.

X. Liu, A. Godbole, C. Lu, G. Michal, and P. Venton. Source strength and dispersion of co2 releases from high-pressure pipelines: Cfd model using real gas equation of state. *Applied Energy*, 126:56–68, 2014.

S. Lorenzi, A. Cammi, L. Luzzi, and G. Rozza. POD-Galerkin method for finite volume approximation of Navier–Stokes and RANS equations. *Computer Methods in Applied Mechanics and Engineering*, 311:151–179, 2016. ISSN 0045-7825. doi: https://doi.org/10.1016/j.cma.2016.08.006. URL https://www.sciencedirect.com/science/article/pii/S0045782516308829.

S. Lorenzi, A. Cammi, L. Luzzi, and G. Rozza. A reduced order model for investigating the dynamics of the Gen-IV LFR coolant pool. *Applied Mathematical Modelling*, 46:263–284, 2017. ISSN 0307-904X. doi: https://doi.org/10.1016/j.apm.2017.01.066. URL https://www.sciencedirect.com/science/article/pii/S0307904X17300720.

S. Marelli and B. Sudret. An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis. *Structural Safety*, 75:67–74, 2018. ISSN 0167-4730. doi: https://doi.org/10.1016/j.strusafe.2018.06.003. URL https://www.sciencedirect.com/science/article/pii/S0167473017302977.

A. Moscatello, A. C. Uggenti, R. Gerboni, and A. Carpignano. A novel approach to high-pressure gas releases simulations. *Journal of Loss Prevention in the Process Industries*, 72:104531, 2021. ISSN 0950-4230. doi: https://doi.org/10.1016/j.jlp.2021.104531. URL https://www.sciencedirect.com/science/article/pii/S0950423021001406.

A. Moscatello, R. Gerboni, G. Ledda, A. C. Uggenti, A. Piselli, and A. Carpignano. CFD gas release model performance evaluation through wind tunnel experiments. *Journal of Loss Prevention in the Process Industries*, 75:104715, 2022.

ISSN 0950-4230. doi: https://doi.org/10.1016/j.jlp.2021.104715. URL https://www.sciencedirect.com/science/article/pii/S0950423021003193.

B. Munson, D. Young, T. Okiishi, W. Huebsch, and A. Rothmayer. *Fluid Mechanics*. Wiley, 2010.

EU Parliament. Directive 2013/30/eu of the european parliament and of the council of 12 june 2013 on safety of offshore oil and gas operations and amending directive 2004/35/ec. *Official Journal of the European Union*, 2013.

N. Pedroni. Computational methods for the robust optimization of the design of a dynamic aerospace system in the presence of aleatory and epistemic uncertainties. *Mechanical Systems and Signal Processing*, 164:108206, 2022. ISSN 0888-3270. doi: https://doi.org/10.1016/j.ymssp.2021.108206. URL https://www.sciencedirect.com/science/article/pii/S0888327021005811.

N. Pedroni and E. Zio. An adaptive metamodel-based subset importance sampling approach for the assessment of the functional failure probability of a thermal-hydraulic passive system. *Applied Mathematical Modelling*, 48:269–288, 2017. ISSN 0307-904X. doi: https://doi.org/10.1016/j.apm.2017.04.003. URL https://www.sciencedirect.com/science/article/pii/S0307904X17302652.

N. Pedroni, E. Zio, and G. Apostolakis. Comparison of bootstrapped artificial neural networks and quadratic response surfaces for the estimation of the functional failure probability of a thermal-hydraulic passive system. *Reliability Engineering and System Safety*, 95:386–395, 2010. URL https://hal-supelec.archives-ouvertes.fr/hal-00609171.

M. Pontiggia, V. Busini, M. Ronzoni, G. Uguccioni, and R. Rota. Effect of large obstacles on high momentum jets dispersion. In *Conference on Safety & Environment in Process & Power Industry-CISAP6*, pages 523–528, 2014.

S. M. Rahman, O. San, and A. Rasheed. A hybrid approach for model order reduction of barotropic quasi-geostrophic turbulence. *Fluids*, 3, 2018. ISSN 2311-5521. doi: 10.3390/fluids3040086. URL https://www.mdpi.com/2311-5521/3/4/86.

S. Rippa. An algorithm for selecting a good value for the parameter c in radial basis function interpolation. *Advances in Computational Mathematics*, 34:105–126, 2011. ISSN 10197168. doi: 10.1007/s10444-010-9146-3.

B. J. Scheiner, C.E. Jordan, J.M. Kuchta, K. C. Dean, M. H. Stanczyk, M. Deul, P. B. Altringer, W. C. McBee, A. G. Kim, A. G. Smelley, et al. *Investigation of*

fire and explosion accidents in the chemical, mining, and fuel-related industries: a manual. Number 678-688. US Department of the Interior, Bureau of Mines, 1985.

A. M. Schleder, E. Pastor, E. Planas, and M. R. Martins. Experimental data and CFD performance for cloud dispersion analysis: The USP-UPC project. *Journal of Loss Prevention in the Process Industries*, 38:125–138, 2015.

P. Secchi, E. Zio, and F. Di Maio. Quantifying uncertainties in the estimation of safety parameters by using bootstrapped artificial neural networks. *Annals of Nuclear Energy*, 35:2338–2350, 2008. ISSN 0306-4549. doi: https://doi.org/10.1016/j.anucene.2008.07.010. URL https://www.sciencedirect.com/science/article/pii/S0306454908002119.

S. A. Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. *Dokl. Akad. Nauk SSSR*, 148:1042–1045, 1963.

Dooguen Song, Kwangho Lee, Chuntak Phark, and Seungho Jung. Spatiotemporal and layout-adaptive prediction of leak gas dispersion by encoding-prediction neural network. *Process Safety and Environmental Protection*, 151:365–372, 2021. ISSN 0957-5820. doi: https://doi.org/10.1016/j.psep.2021.05.021. URL https://www.sciencedirect.com/science/article/pii/S0957582021002615.

TNO. *Methods for the calculation of physical effects, Yellow Book*. The Hague, 2005.

EPA US. ALOHA software. https://www.epa.gov/cameo/aloha-software. [Online; accessed 21-June-2022].

J. Vinnem and W. Røed. *Offshore Risk Assessment Vol. 2*. Springer, 2014.

C. Vivalda, R. Gerboni, and A. Carpignano. A practical approach to risk-based gas monitoring system design for oil and gas offshore platforms. In *Proceedings of the 14th Probabilistic Safety Assessment and Management Conference*, 2018.

S. Volkwein. Model reduction using proper orthogonal decomposition. *Lecture notes, Institute of Mathematics and Scientific Computing, University of Graz. see http://www. uni-graz. at/imawww/volkwein/POD. pdf*, 1025, 2011.

D. Xiao. Error estimation of the parametric non-intrusive reduced order model using machine learning. *Computer Methods in Applied Mechanics and Engineering*, 355:513–534, 2019. ISSN 00457825. doi: 10.1016/j.cma.2019.06.018. URL https://doi.org/10.1016/j.cma.2019.06.018.

D. Xiao, F. Fang, A.G. Buchan, C.C. Pain, I.M. Navon, and A. Muggeridge. Non-intrusive reduced order modelling of the navier–stokes equations. *Computer Methods in Applied Mechanics and Engineering*, 293:522–541, 2015. ISSN 0045-7825. doi: https://doi.org/10.1016/j.cma.2015.05.015. URL https://www.sciencedirect.com/science/article/pii/S0045782515001887.

D. Xiao, F. Fang, C. C. Pain, and I. M. Navon. A parameterized non-intrusive reduced order model and error analysis for general time-dependent nonlinear partial differential equations and its applications. *Computer Methods in Applied Mechanics and Engineering*, 317:868–889, 2017. ISSN 00457825. doi: 10.1016/j.cma.2016.12.033. URL http://dx.doi.org/10.1016/j.cma.2016.12.033.

E. Zamejc. Api standard 521 new alternative method to evaluate fire relief for pressure relief device sizing and depressuring system design. *Journal of Loss Prevention in the Process Industries*, 27:21–31, 2014. ISSN 0950-4230. doi: https://doi.org/10.1016/j.jlp.2013.10.016. URL https://www.sciencedirect.com/science/article/pii/S0950423013002088.

E. Zio. A study of the bootstrap method for estimating the accuracy of artificial neural networks in predicting nuclear transient processes. *IEEE Transactions on Nuclear Science*, 53:1460–1478, 2006. doi: 10.1109/TNS.2006.871662.

E. Zio, G.E. Apostolakis, and N. Pedroni. Quantitative functional failure analysis of a thermal–hydraulic passive system by means of bootstrapped artificial neural networks. *Annals of Nuclear Energy*, 37:639–649, 2010. doi: https://doi.org/10.1016/j.anucene.2010.02.012. URL https://www.sciencedirect.com/science/article/pii/S0306454910000666.

C. Zuliani, C. De Lorenzi, and S. Ditali. Application of CFD simulation to safety problems–challenges and experience including a comparative analysis of hot plume dispersion from a ground flare. *Chemical Engineering Transactions*, 53:79–84, 2016.

Appendix A. POD-RBF algorithms

The appendix reports the detailed algorithms adopted throughout the paper.

Algorithm 2: Proper Orthogonal Decomposition algorithm

Input

1. snapshot matrix $\hat{Y} = [\vec{y}_1, \vec{y}_2, \ldots, \vec{y}_n]$, with $\vec{y}_i \in \mathbb{R}^m$;
2. snapshot matrix rank r;
3. truncation error ε;

Output

1. POD basis $\hat{B}_k = [\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_k]$, with $\vec{b}_i \in \mathbb{R}^m$;
2. POD eigenvalues $\lambda_i, \forall i = 1, \ldots, k$;
3. POD energy $\mathcal{E}_k$;
4. POD coefficients $\hat{A}_k = [\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_k]$, with $\vec{a}_i \in \mathbb{R}^t$;

try:
  $[\hat{\Psi}, \hat{\Sigma}, \hat{\Phi}] = \texttt{svd}(\hat{Y})$ # *Singular Value Decomposition*
  for $i = 1, \ldots, r$ **do**
    $\vec{b}_i = \vec{\psi}_i$ # $\vec{\psi}_i$ *is the* $\hat{\Psi}$ *i-th column*
    $\lambda_i = \sigma_i^2$ # σ_i *is the i-th diagonal entry of* $\hat{\Sigma}$
  end
except *Memory Error*:
  if $n > m$ **then**
    $[\hat{\Psi}, \hat{\Lambda}] = \texttt{eig}(\hat{Y}\hat{Y}^{\mathrm{T}})$ # *Eigenvalue Decomposition*
    for $i = 1, \ldots, r$ **do**
      $\vec{b}_i = \vec{\psi}_i$ # $\vec{\psi}_i$ *is the i-th column of* $\hat{\Psi}$
      $\lambda_i = \hat{\Lambda}_{i,i}$ # $\hat{\Lambda}_{i,i}$ *is the i-th diagonal entry of* $\hat{\Lambda}$
    end
  else
    $[\hat{\Phi}, \hat{\Lambda}] = \texttt{eig}(\hat{Y}^{\mathrm{T}}\hat{Y})$ # *Eigenvalue Decomposition*
    for $i = 1, \ldots, r$ **do**
      $\vec{b}_i = \frac{\hat{Y}\vec{\phi}_i}{\sqrt{\lambda_i}}$ # $\vec{\psi}_i$ *is the i-th column of* $\hat{\Psi}$
      $\lambda_i = \hat{\Lambda}_{i,i}$ # $\hat{\Lambda}_{i,i}$ *is the i-th diagonal entry of* $\hat{\Lambda}$
    end
  end
end
choose k such that $1 - \mathcal{E}_k < \varepsilon$
$\hat{A}_k = \hat{Y}^{\mathrm{T}}\hat{B}_k$ # *compute reduced order coefficients*

Algorithm 3: Radial Basis Function training algorithm.

Input

1. data reduced via POD, $\hat{A}_k \in \mathbb{R}^{n \times k}$;
2. parameter matrix $\hat{P} = [\vec{p}_1, \vec{p}_2, \ldots, \vec{p}_n]$, with $\vec{p}_i \in \mathbb{R}^p$;
3. RBF type f (e.g. $f = 1/\sqrt{||\vec{p}_1 - \vec{p}_j||_2 + \sigma^2}$);
4. hyperparameter σ via algorithm 4 or Hardy's formula Hardy (1971);

Output

training matrix $\hat{W} \in \mathbb{R}^{n \times k}$

 # loop over each column of $\hat{A}_k$, $\vec{a}_i \in \mathbb{R}^n$

for $i = 1, \ldots, k$ **do**

  choose hyperparameter σ_i (optimal selection with algorithm 4)

  # compute distance matrix $\hat{D} = [\vec{d}_1, \ldots, \vec{d}_n] \in \mathbb{R}^{n \times n}$

  for $j = 1, \ldots, n$ **do**

    $\vec{d}_j = [f(||\vec{p}_1 - \vec{p}_j||_2, \sigma_i), \ldots, f(||\vec{p}_n - \vec{p}_j||_2, \sigma_i)]$

  end

  $\vec{w}_i = \hat{D}^{-1}\vec{a}_i$ *#* $\vec{w}_i$ *is the* i*-th column of* $\hat{W}$

end

Algorithm 4: Optimal hyperparameter selection.

Input

1. data reduced via POD, $\hat{A}_k \in \mathbb{R}^{n\times k}$;
2. parameter matrix $\hat{P} = [\vec{p}_1, \vec{p}_2, \ldots, \vec{p}_n]$, with $\vec{p}_i \in \mathbb{R}^p$;
3. RBF type f (e.g. $f = 1/\sqrt{||\vec{p}_1 - \vec{p}_j||_2 + \sigma^2}$);
4. initial hyperparameter guess $\sigma_0 = \dfrac{0.815}{n}\displaystyle\sum_{i=1}^{n} ||\vec{p}_i - \vec{p}_\ell||_2$, where ℓ is the closest point to i in the parameter space Hardy (1971).

Output

1. Optimal hyperparameter σ_{opt}

Optimise σ **minimising** RMSE

Select σ # *e.g. via conjugate gradient method*
Compute $\hat{D}$ and $\vec{w}_i$ as in algorithm 3
use Rippa's algorithm Rippa (2011) and LOOCV
$[\hat{L}, \hat{U}] = \mathtt{lu}(\hat{D})$ # *Compute LU decomposition*
for *j=1, …, n* **do**
  $\vec{x} = \hat{U}^{-1}\hat{L}^{-1}\vec{e}_j$ # $\vec{e}_j$ *is the j-th column of the identity matrix*
  $\delta_j = \left\| \dfrac{w_{i,j}}{x_j} \right\|$ # *compute interpolation error with j-th components of* $\vec{x}$ *and* $\vec{w}_i$
end
$\text{RMSE} = \sqrt{\dfrac{1}{n}\displaystyle\sum_{j=1}^{n} \delta_j^2}$
end

Algorithm 5: Bootstrap method

Input

1. number of reboots N_r;
2. training parameter values $\hat{P} = [\vec{p}_1, \vec{p}_2, \ldots, \vec{p}_n]$;
3. training snapshots $\hat{Y} = [\vec{y}_1, \vec{y}_2, \ldots, \vec{y}_n]$;

Output

Ensemble of ROMs, $\mathcal{R}_1, \ldots, \mathcal{R}_{N_r}$

 # this can be done in parallel

for *i=1, ..., N_r* **do**
  for *j=1, ..., N* **do**
    Sample one parameter point $\vec{p}_j$ from $[\vec{p}_1, \vec{p}_2, \ldots, \vec{p}_n]$ with replacement;
    if $\vec{p}_j$ exists **then**
      Discard $\vec{p}_j$ to avoid singular RBF kernel;
  end
  Reduce dataset with algorithm 2;
  Train and tune the RBF network with algorithms 3, 4;
end

Algorithm 6: Unscented Transform and POD for uncertainty propagation

Input

ensemble of ROM responses $\hat{Y}_{\mathcal{R}} = [\vec{y}_{\mathcal{R},1}, \ldots, \vec{y}_{\mathcal{R},N_r}]$ (spatial fields) for a certain parameter value $\vec{p}_{\text{new}}$;

Output

mean $\mathrm{E}[\vec{z}]$ and covariance $\mathrm{cov}[\vec{z}]$ estimates for the response z of the model $\mathcal{M}'$;

 # choose sigma points

Compute reduced order coefficients $\vec{a} \in \mathbb{R}^k$ via POD (algorithm 2) of $\hat{Y}_{\mathcal{R}}$;

Compute mean $\vec{\mu} = \mathrm{E}[\vec{a}]$ and covariance $\hat{C} = \mathrm{cov}[\vec{a}]$;

Compute Cholesky or SVD decomposition to obtain $\hat{L}\hat{L}^{\intercal} = k\hat{C}$;

Apply the scaling factor $s = k + \lambda$, $\hat{S} = \sqrt{s}\sqrt{\hat{C}}$; *# λ is a free parameter (here λ=0)*

Determine weights as $w_i = 1/2s$, $i = 1, \ldots, k$ and $w_0 = \lambda/s$ (associated with $\vec{\mu}$);

 # this can be performed in parallel

for *i=1, ..., k* **do**

$\vec{\sigma}_{UT,i} = \vec{\mu} - \hat{S}_i$ *# $\hat{S}_i$ is the i-th column of the $\hat{S}$ matrix*

$\vec{z}_i = \mathcal{M}'(\vec{\sigma}_{UT,i})$ *# apply non-linear model*

$\vec{\sigma}_{UT,i+k} = \vec{\mu} + \hat{S}_i$

$\vec{z}_{i+k} = \mathcal{M}'(\vec{\sigma}_{UT,i+k})$ *# apply non-linear model*

end

compute $\mathrm{E}_w[\vec{z}] = \sum_{i=0}^{2k+1} w_i \vec{z}_i$ and

$\mathrm{cov}_w[\vec{z}] = \sum_{i=0}^{2k+1} w_i (\vec{z}_i - \mathrm{E}_w[\vec{z}])(\vec{z}_i - \mathrm{E}_w[\vec{z}])^{\mathrm{T}}$;

Appendix B. Additional figures

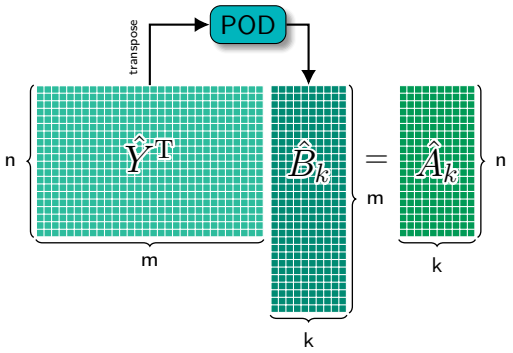

Figure B.16: POD-driven dimensionality reduction for the snapshot matrix containing the training CFD solutions.

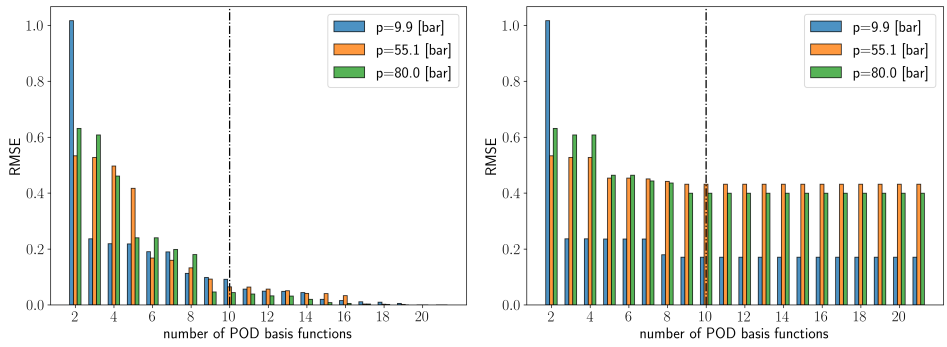

Figure B.17: Root-Mean Squared-Error between the original snapshot and its truncated POD representation (left) and Root-Mean-Squared Error between the original snapshot and the ROM reconstruction (right) for some training pressures.

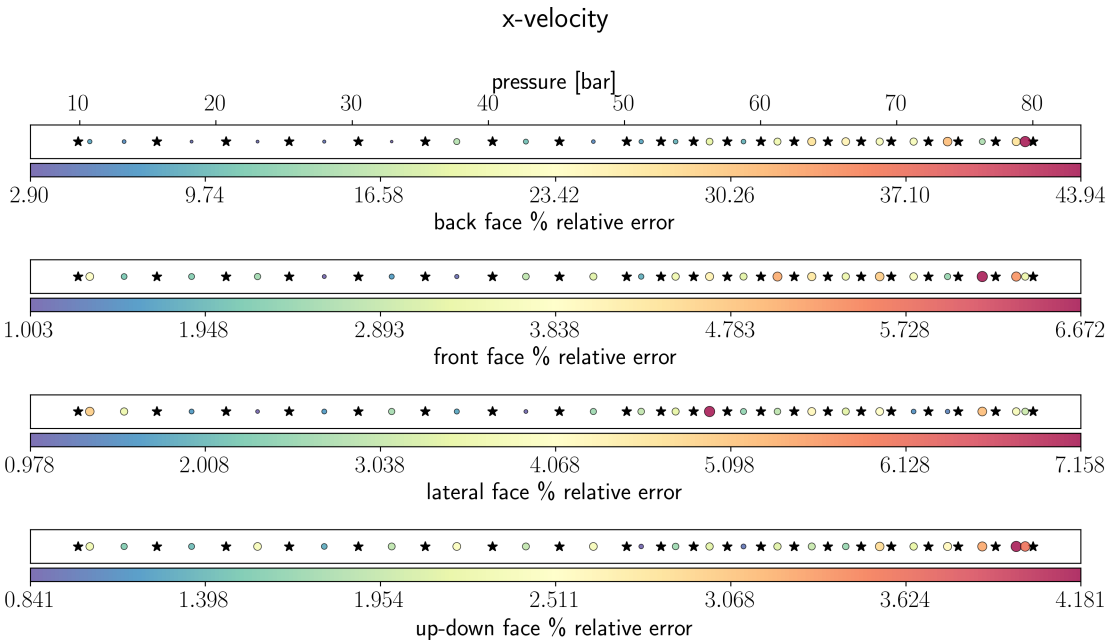

Figure B.18: Face-wise relative L_2 error for the x-component of the velocity. The black stars represent the training cases, while the dots represent the validation cases and their color is related to the relative error.

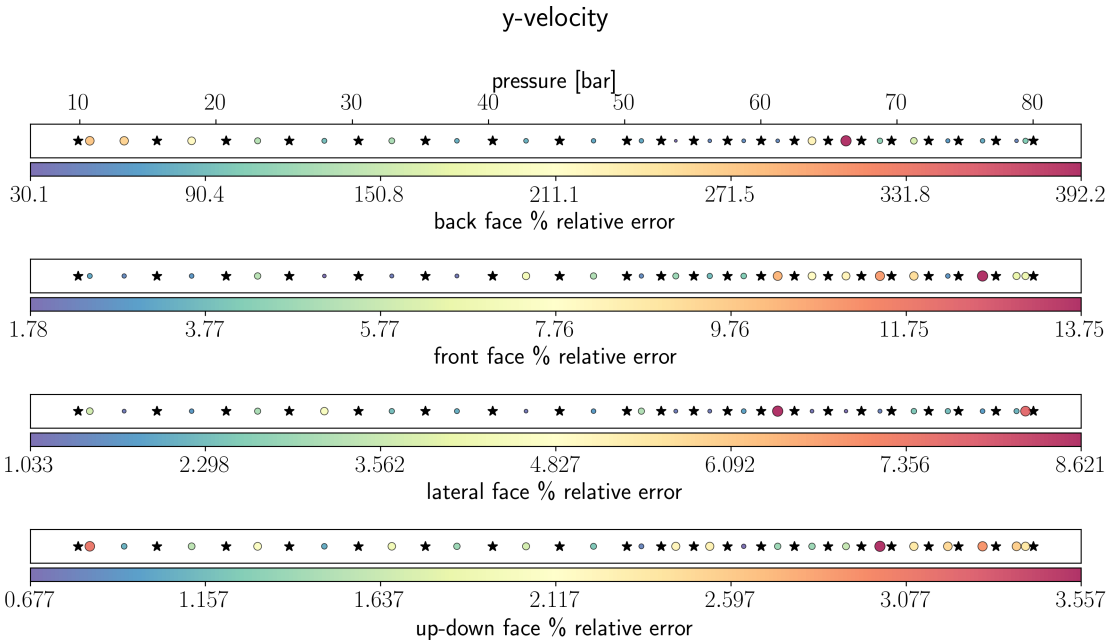

Figure B.19: Face-wise relative L_2 error for the y-component of the velocity. The black stars represent the training cases, while the dots represent the validation cases and their color is related to the relative error.

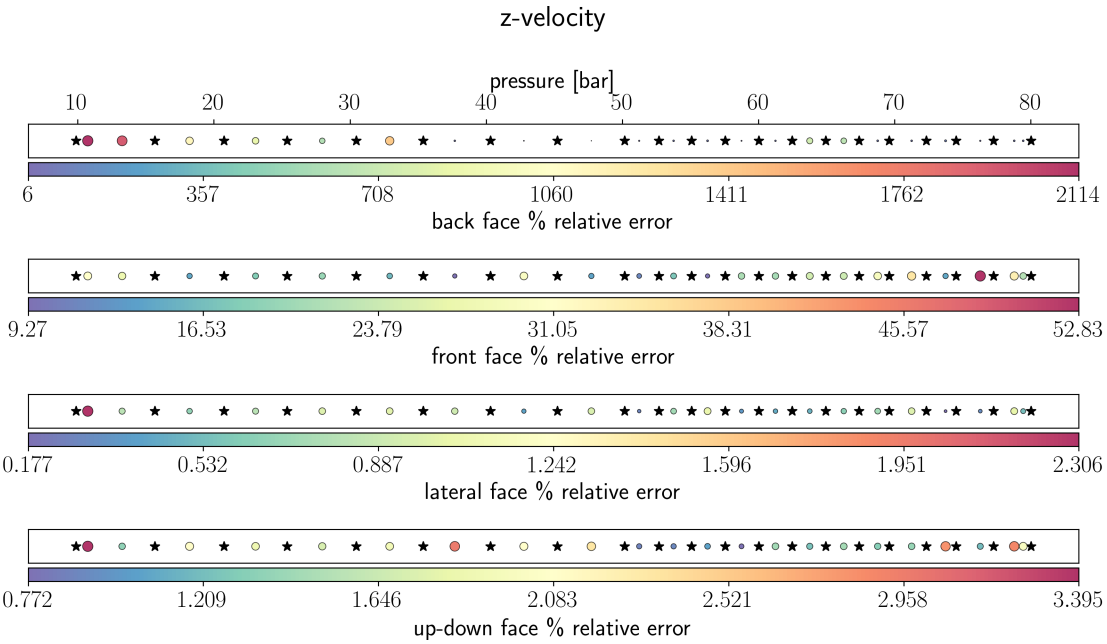

Figure B.20: Face-wise relative L_2 error for the z-component of the velocity. The black stars represent the training cases, while the dots represent the validation cases and their color is related to the relative error.

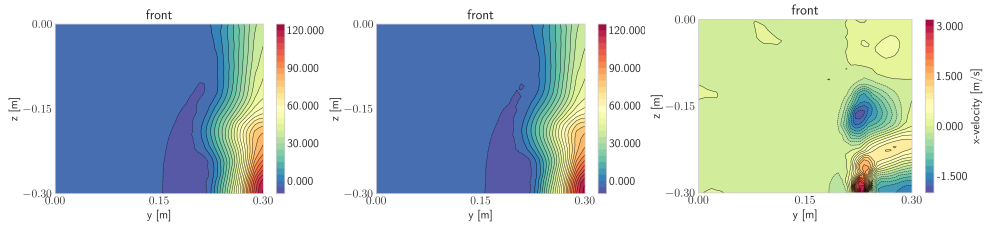

Figure B.21: x-component of the velocity field for CFD (left) and NIROM (centre) and their difference (right) on the front face for the validation case with p_{rel}=51.237 bar.

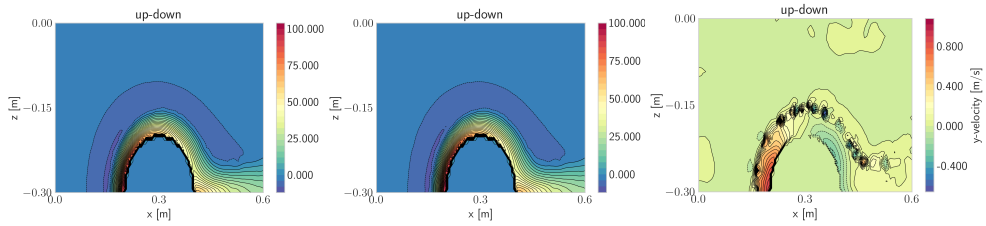

Figure B.22: y-component of the velocity field for CFD (left) and NIROM (centre) and their difference (right) on the up-down face for the validation case with p_{rel}=51.237 bar.

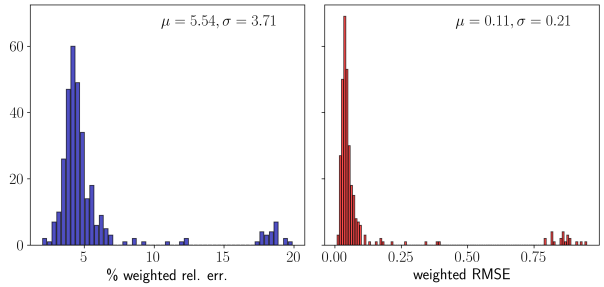

Figure B.23: Weighted relative percentage error (left) and weighted RMSE (right) computed for the validation case p_{rel}=10.739 bar.

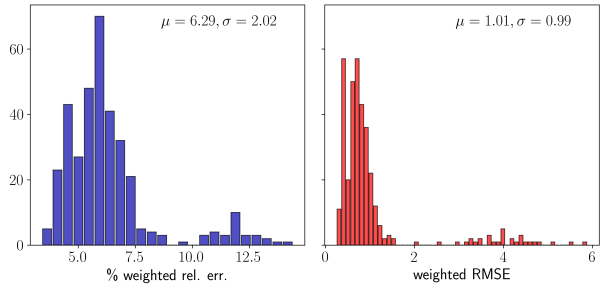

Figure B.24: Weighted relative percentage error (left) and weighted RMSE (right) computed for the validation case p_{rel}=79.442 bar.