CSMA 2019
14ème Colloque National en Calcul des Structures
13-17 Mai 2019, Presqu'île de Giens (Var)

A non-intrusive reduced order data assimilation method applied to the monitoring of urban flows

J. K. Hammond [1], R. Chakir[2]

[1] *INRIA, 2 Rue Simone IFF, 75012 France*
[2] *Université Paris Est, IFSTTAR, 10-14 Bd Newton, Cité Descartes, 77447 Marne La Vallée Cedex, France*

Résumé — In this work we investigate a variational data assimilation method to rapidly estimate urban pollutant concentration around an area of interest using measurement data and CFD based models in a non-intrusive and computationally efficient manner. In case studies presented here, we used a sample of solutions from a dispersion model with varying meteorological conditions and pollution emissions to build a Reduced Basis approximation space and combine it with concentration observations. The method allows to correct for unmodeled physics, while significantly reducing online computational time.
Mots clés — Reduced Basis Methods, Variational Data Assimilation Methods, CFD.

1 Introduction

As the population increases, cities must constantly reassess their urban planning. However, this must be done in such a way to preserve the quality of life of its inhabitants. Energy saving, sustainable water and air quality are some of the important challenges associated with growing cities. In this context, the monitoring of the different urban flows (pollution, heat) is very important. For instance data assimilation approaches can be used in monitoring. These methods incorporate available measurement data and mathematical model to provide improved approximations of the physical state. The effectiveness of modeling and simulation tools is essential. Advanced physically based models could provide spatially rich small-scale solution, however the use of such models is challenging due to explosive computational times in real-world applications. Beyond computational costs, physical models are often constrained by available knowledge on the physical system. To overcome these difficulties, we resort to a new technique combining Model Order Reduction (MOR) and variational data assimilation known as PBDW state estimation and introduced in [10]. The PBDW formulation combines a Reduced Basis (RB) [4, 8, 11, 12] from the physically based model and the experimental observations, in order to provide a real-time state estimate in a non-intrusive manner. The RB is used to diminish the cost of using a high-resolution model by exploiting the parametric structure of the governing equations. In addition, variational data-assimilation techniques are used to correct the model error. In this work we extend the PBDW method previously applied to small-scale experimental problems to the monitoring of urban pollution as an important test case for practical applications, but also as an example of the very generic approach that proves well suited to online monitoring of urban flows over large scales. Our focus here is a problem of pollutant dispersion at the urban scale which can provide insight on how to treat the practical problems associated to MOR and data assimilation of complex flows involved in many sophisticated methods of urban air quality modelling.

2 The PBDW : a variational reduced order data assimilation method

We consider data assimilation methods that provide an estimate of the true physical state $c^{true}(\mathcal{S})$ in a configuration $\mathcal{S}$ of the physical system by combining the information of a physical model and experimental data. We want to find the best possible approximation of the physical system being studied while expending minimal resources, which translates in practice to using the best model possible and available data without requiring excessive computational investment to solve the problem, focusing here on methods combining reduction and data assimilation in a non-intrusive procedure.

Given a(n unknown) parameter configuration $\mathbf{p} \in \mathcal{D}$ representing the physical system $\mathcal{S}$ (where $\mathcal{D} \in \mathbb{R}^{N_p}$ is the parameter domain and N_p is the number of parameters), we consider models in the form of a problem $\mathcal{P}$

$$\mathcal{P} : \Omega \times \mathcal{D} \to \mathbb{R}$$

associated to a parameterized PDE : find $c(\mathbf{p}) \in \mathcal{X}$ such that

$$\mathcal{L}(\mathbf{p})\, c(\mathbf{p}) = 0 \quad \text{in } \Omega \qquad + \qquad \text{Boundary conditions on } \partial\Omega,$$

where $\Omega \subset \mathbb{R}^d$ is a bounded domain, $d = 2$ or 3 and $\mathcal{X}$ some suitable Banach space. And given M observations, we assume our data y_m^{obs}, $1 \leq m \leq M$, are of the form

$$y_m^{obs} = \ell_m(c^{true}(\mathcal{S})). \tag{1}$$

where $\ell_m \in \mathcal{X}'$ are linear functionals representing the sensors.

The PBDW method is a non-intrusive reduced order method of data assimilation for parameterized PDEs that belongs to the family of Reduced Basis (RB) methods. Standard reduced basis methods are projection-based model reduction methods relying on the relatively small dimension of the solution manifold $\mathcal{M}$ associated to the problem $\mathcal{P}$ for parameter configurations $\mathbf{p} \in \mathcal{D}$ (we note that not all problems have a low-dimensional solution manifold). If the manifold of solutions is of relatively small dimension, it can be approximated by a finite set of well-chosen solutions of $\mathcal{P}$, $(c(\mathbf{p}_1), \cdots, c(\mathbf{p}_N))$, generating an N-dimensional space, called the RB space. This space is then used as approximation space in the discrete method of solving $\mathcal{P}$, for example replacing the large number of simple basis functions generating a finite element space with N solutions $c(\mathbf{p}_i)$ $1 \leq i \leq N$, to $\mathcal{P}$, each providing information on the solution manifold $\mathcal{M}$. The idea of reduced basis methods is to compute an inexpensive and accurate approximation, $c_N(\mathbf{p})$, of the solution $c(\mathbf{p})$ to problem $\mathcal{P}$ for any $\mathbf{p} \in \mathcal{D}$ by seeking a linear combination of the particular solutions :

$$c_N(\mathbf{p}) = \sum_{i=1}^{N} \alpha_i(\mathbf{p})\, c(\mathbf{p}_i). \tag{2}$$

Efficient implementation of traditional RBMs requires construction of all parameter-independent quantities during a prior *offline* stage, which implies modifying the calculation code, an intrusive procedure. The methods explored in this work take advantage of the reduction capacity of RBMs, but utilize the RB space in a non-intrusive manner. The PBDW method considers our mathematical model to be the "best-knowledge" model $\mathcal{P}^{bk}$ (i.e. the best adapted model available for the problem $\mathcal{P}$), and the set of admissible parameters $\mathcal{D}^{bk}$. The PDE model $\mathcal{P}^{bk}$ is used to build an N-dimensional RB *background* space, $\mathcal{Z}_N$, representing solutions to the known problem, designed to handle parametric uncertainty. Information on physical location and form of the M sensors providing the data is used to build an M-dimensional *update* space, $\mathcal{U}_M$, representing the information gathered by the sensors. The PBDW solution, noted $c_{M,N}(\mathbf{p})$ is built from the two approximation spaces, $\mathcal{Z}_N$ and $\mathcal{U}_M$. We thus aim to approximate the true physical state $c^{true}(\mathcal{S})$ by

$$c_{M,N}(\mathbf{p}) = c_N^{bk}(\mathbf{p}) + \eta_M \tag{3}$$

where $\eta_M \in \mathcal{U}_M$ is an *update* correction term associated to the experimental observations, and $c_N^{bk}(\mathbf{p}) \in \mathcal{Z}_N$ is a reduced basis approximation of the solution to the model $\mathcal{P}^{bk}$. The PBDW problem, as with many data assimilation methods, is posed as a minimization problem, in which we minimize the *update* contribution, keeping our approximation close to the solution manifold $\mathcal{M}^{bk}$ associated to $\mathcal{P}^{bk}$ for $\mathcal{D}^{bk}$, and imposing experimental observation values at the sensor points.

$$\left\| \begin{array}{l} \text{Find } (u_{N,M}, z_N, \eta_M) \text{ such that} \\ (u_{N,M}, z_N, \eta_M) \;=\; \underset{\substack{\tilde{c}_{N,M} \in \mathcal{X} \\ \tilde{z}_N \in \mathcal{Z}_N \\ \tilde{\eta}_M \in \mathcal{U}_M}}{\operatorname{arginf}} \left\{ \|\tilde{\eta}_M\|_{\mathcal{X}}^2 \middle| \begin{array}{l} \langle \tilde{c}_{N,M} - \tilde{z}_N, v \rangle_{\mathcal{X}} = \langle \tilde{\eta}_M, v \rangle_{\mathcal{X}}, \forall\, v \in \mathcal{X} \\ \langle \tilde{c}_{N,M}, \phi \rangle_{\mathcal{X}} = \langle c^{true}, \phi \rangle_{\mathcal{X}}, \forall \phi \in \mathcal{U}^M \end{array} \right\}. \end{array} \right. \tag{4}$$

We rely on the Euler-Lagrange equations, derived from the minimization problem (4), to find a linear system of size $(M+N) \times (M+N)$ for non-iterative solution of the problem. The procedure is decomposed into *offline* and *online* stages, where the approximation space and linear system construction is done offline, allowing a very efficient online stage.

3 Applications to the monitoring of urban flows

We want to apply these methods to estimate urban pollution using concentration data and a dispersion model. Our dispersion model $\mathcal{P}^{bk}$ relies on computational fluid dynamic modeling. The wind field carrying the pollution is the solution of an incompressible Navier-Stokes equation with $k-\varepsilon$ turbulent closure, and simplifying for unknown physics, the pollutant concentration is the solution to an advection-diffusion equation given by

$$\overbrace{\rho\,\vec{v}\cdot\nabla c}^{\text{transport}} - \overbrace{div((\underbrace{\varepsilon_{mol}+\varepsilon_{turb}}_{\varepsilon_{tot}})\nabla c)}^{\text{diffusion}} = \overbrace{\rho F_{src}}^{\text{source}}, \tag{5}$$

along with appropriate boundary conditions for an exterior calculation domain. We chose a particulate pollutant $PM_{2.5}$ (particulate matter of diameter $d \leq 2.5\mu m$) in this study, which on the short term can be considered to have negligible reaction. We set inflow velocities $\mathbf{p}_v$ (in a fixed direction $(1,1)^T$) within the *calm* and *light air* categories of the Beaufort scale (from $0.1\frac{m}{s}$ to $1.3\frac{m}{s}$) and set source intensity $\mathbf{p}_s$ (from of 1×10^{-3} to $1\times10^{-2}\ \frac{mg}{m^3\cdot s}$) representing varying traffic based on reports made available to the public by the U.S. EPA and on municipality websites [6, 13, 3].
The velocity field $\vec{v}$ and turbulent diffusion field ε_{turb} can be seen as parameters of the advection-diffusion equation (5) which allow us to decouple the computation of the wind field. Additionally, given the large scale of air quality modeling problems, and the numerical problems caused by different orders of terms in the PDE (5), we also want to consider a dimensionless approach. A dimensionless approach generalizes the problem ; it can give insight into which parameters may be of lesser importance and may be approximated or ignored, and can help scale the problem if the values of certain terms vary significantly from others. In advection-diffusion problems the important physical quantities are the velocity, diffusion, and concentration. We thus a-dimensionalize with respect to these variables, and the spatial variable by a characteristic length, and consider a dimensionless problem $\mathcal{P}_{adim}$ over a dilation Ω_0 of the domain Ω by the characteristic length. Then in order to compromise between accuracy, numerical stability, and computational time we use pseudo-steady-state CFD wind fields, solutions to Reynolds-Averaged Navier-Stokes with $k-\varepsilon$ turbulence by *Code_Saturne* [1] (a general purpose finite-volume CFD software), and solved the dimensionless problem $\mathcal{P}_{adim}$ by finite elements with a Streamline Upwind Petrov-Galerkin (SUPG) stabilization scheme [2, 7] using FreeFem++ [5] for ease of implementation.

We begin by computing a set of *training* solutions to the model $\mathcal{P}^{bk}$ over the parameter set $\mathcal{D}^{bk}$ and selecting the generators of a reduced basis. We compute two different Update spaces, from sensors placed randomly and sensors selected by a GEIM-based Greedy algorithm[9]. Next, in order to evaluate the capacity of our data assimilation method to treat imperfect models, specifically models which may not account for all physical processes, we used a shifted model $\mathcal{P}^{trial}$ to compute synthetic data representing a "true" solution used in our case studies. In the shifted model $\mathcal{P}^{trial}$ pollution's concentration are solution to the following advection-diffusion-reaction :

$$\rho\vec{v}\cdot\nabla c - div((\varepsilon_{mol}+\varepsilon_{turb})\nabla c) + \rho R c = \rho F_{src}, \tag{6}$$

where ρRc represents a linear reaction term with coefficient R for approximate total change from production and loss during reaction processes. We next provide PBDW state estimation results comparing to sets of *trial* solutions of the advection-diffusion-reaction shifted model $\mathcal{P}^{trial}$. We take parameters $\mathbf{p}\in\mathcal{D}^{bk}$ (but different from the solutions generating $\mathcal{Z}_N$) and each *trial* set corresponds to a different model shift with $R\in(0,\ 10^{-3},\ 10^{-4})$.

3.1 Case study in exterior air quality

We first set a case study for a relatively simple (with respect to the complexity of a real-world case at urban scale with precise geometry and varying conditions) domain of dimensions $75m\times120m$, representing a small residential neighborhood polluted by traffic on a street and by combustion sources (not shown here) in residential yards (see FIGURE 1). Examples for various $\mathbf{p}=(\mathbf{p}_v,\mathbf{p}_s)$ of solution $c^{bk}(\mathbf{p})$ to $\mathcal{P}^{bk}$, can be seen in FIGURE 2.

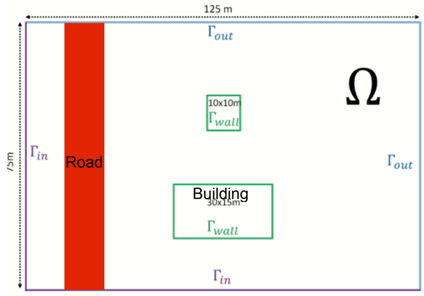

FIGURE 1 – Calculation domain (with urban obstacles) and street pollution source $\mathbf{p}_s$

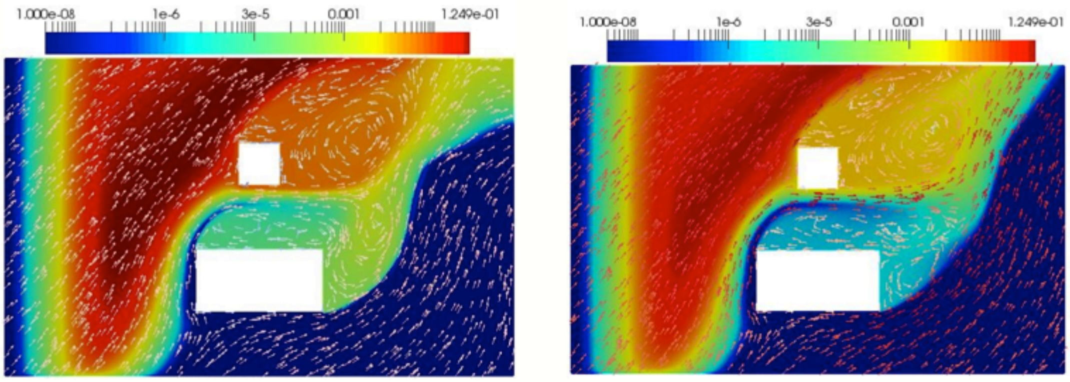

FIGURE 2 – Concentration solution (logarithmic scale) over velocity field. Left : with $\mathbf{p}_v = 0.1\frac{m}{s}$ and $\mathbf{p}_s = 1 \times 10^{-3}\frac{mg}{m^3}$. Right : with $\mathbf{p}_v = 1.3\frac{m}{s}$ and $\mathbf{p}_s = 1 \times 10^{-2}\frac{mg}{m^3}$.

In FIGURE 3 we consider the case of significant model error (by an added reaction term of $R = 0.001$). We see significant improvement between $N = 2$ and $N = 6$ (the lowest contour line shows 1% error). We see that with $N = 6$ and $M = 15$ the error is under 7% everywhere, and often under 1%.

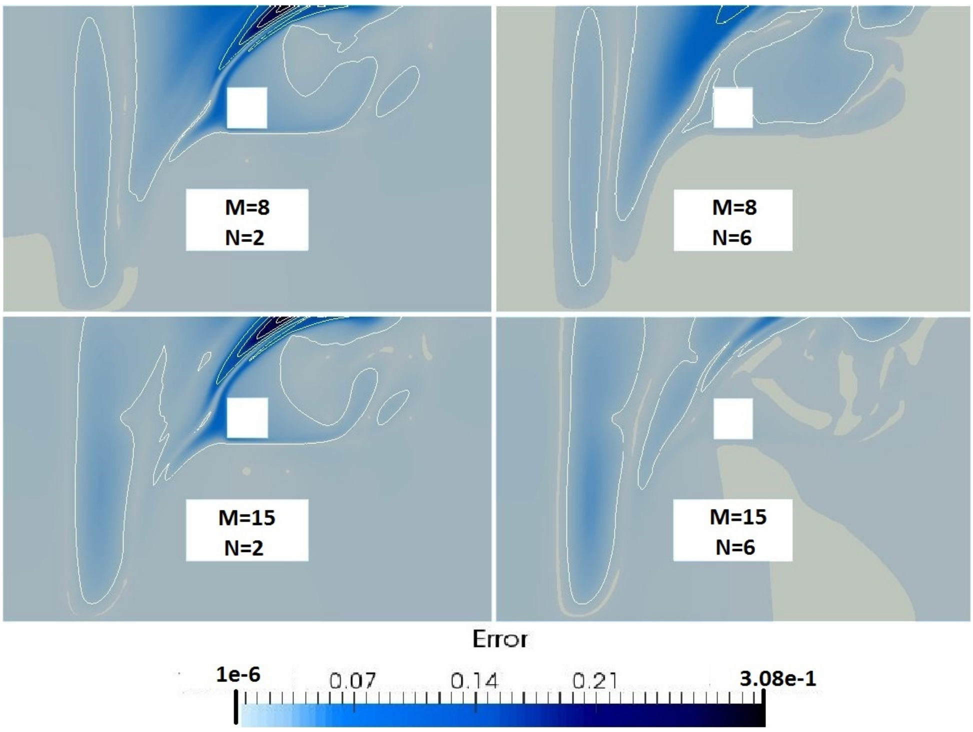

FIGURE 3 – Relative mean pointwise PBDW approximation error maps for $N = 2$ (left), $N = 6$ (right), and for $M = 8$ (top) and $M = 13$ (bottom), over $\mathbf{p} \in \mathcal{D}^{trial}$ with model error.

3.2 A Real-World Application

We extend our study to a real-world application over Fresno, California, city affected by particularly high pollutant concentrations. This application is in view of epidemiology exposure assessments employed by a research team at UC Berkeley (UCB). The long-term goal is to improve the methods for estimating individual exposures and expand the ability of current UCB epidemiological studies to evaluate the association of these exposures to various health conditions. We aim to extend reduced order data assimilation methods for deterministic PDE-based models to a real-world inspired case study in the hopes of showing the feasibility of these methods in real applications. In FIGURE 4 we can see a 3D geometric representation of a neighborhood in Fresno, used as calculation domain in our study.

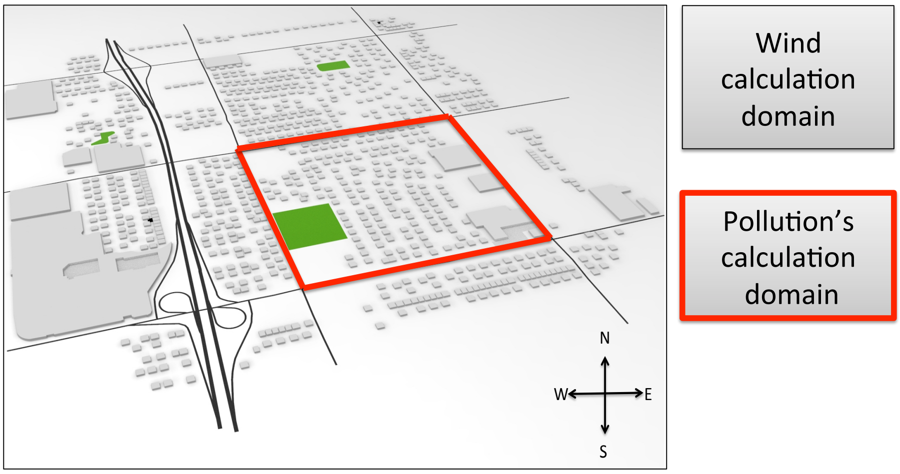

FIGURE 4 – Neighborhood in Fresno over which a wind field was computed using *Code_Saturne* and domain used to study pollutant concentrations (in red).

In FIGURE 5 we see a wind field with $\vec{v}_{in} = 1.3 * z^{0.4}$ in **SE** direction (308 deg) corresponding to real meteorological conditions on April 1, 2001, and an associated dimensionless trial solution to $\mathcal{P}^{trial}$ with $R = 0$ and $R = 0.001$, where pollution sources are taken to be two streets.

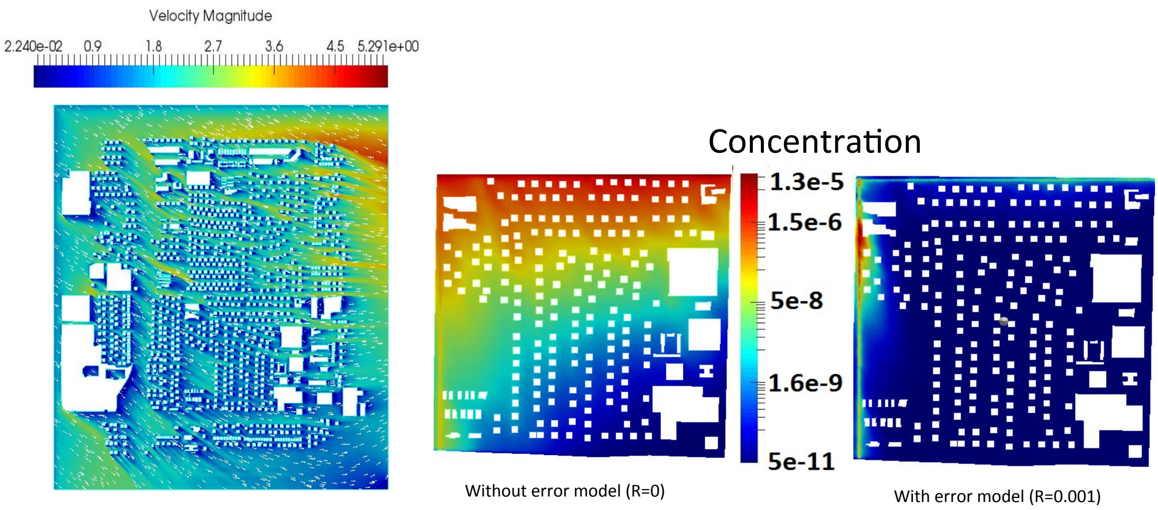

FIGURE 5 – Wind field corresponding to conditions over Fresno on April 1, 2001 (left). Dimensionaless concentration with two pollution sources solution (right).

In FIGURE 6 we show relative average PBDW approximation errors without and with model error (for $R = 0.001$) plotted over the calculation domain using a set of 8 trial solutions to $\mathcal{P}^{trial}$. We can see for this simple first test on our real-world case study, with non-negligible model error by an added reaction term, we can reconstruct the concentration field with under 1% error nearly everywhere, a promising result for future application of these methods.

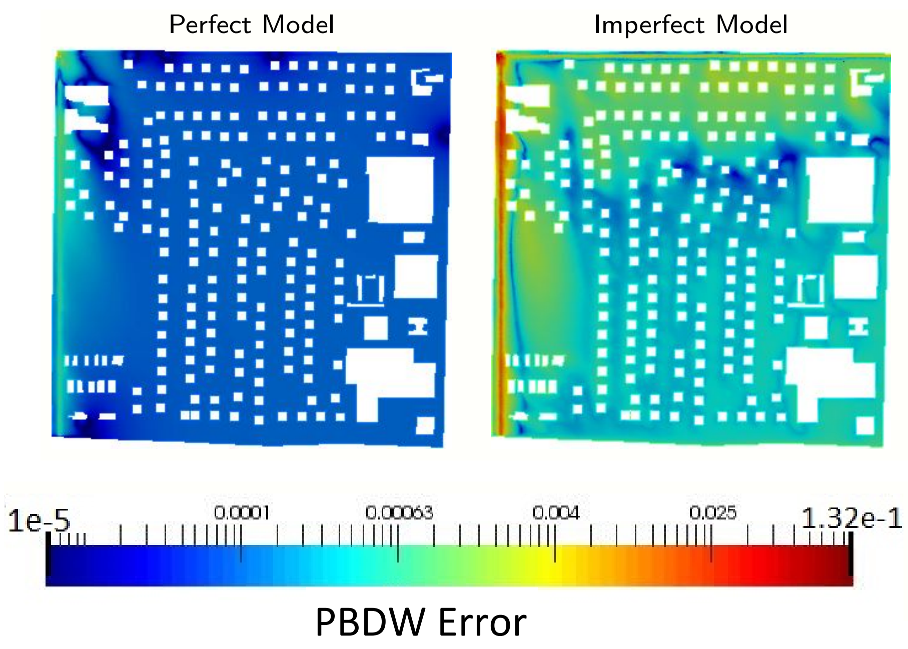

FIGURE 6 – Relative mean PBDW approximation error over a set of 8 trial solutions to $\mathcal{P}^{trial}$ with no model error (left) and with $R = 0.0001$ (right). Here $M = 8$ and $N = 3$.

In TABLE 2 we give computational times required for this first study of our data assimilation method over our real-world computational domain. We give offline computational time in TABLE 1 for the calculation of a concentration solution only given the CFD wind field, as the CFD field cost depends highly on computational power of the machine.

FEM CPU Times	
FEM $\mathbb{P}_1$ - SUPG	31min

TABLE 1 – Computational times of the FEM approximation of $\mathcal{P}^{bk}$. Average over the set of training solutions considered here.

PBDW CPU Times	Online Stage (average CPU times) Reconstruction of the full solution $c_{M,N}$
$M = 8, N = 3$	$7.1s$
$M = 15, N = 3$	$12.2s$
$M = 15, N = 5$	$13.3s$

TABLE 2 – Computational times of the PBDW state estimation for various M and N values. Average over the set of trial solutions considered here.

4 Conclusion

In this paper we studied the implementation of the PBDW using a simplified dispersion model with source terms and boundary conditions informed by literature. We examined the results of the PBDW method using synthetic data from a shifted model $\mathcal{P}^{trial}$ and a shifted parameter set Ξ^{trial} to study the stability of and validate the method in our case studies. These results show promise in the expansion of the PBDW reduced basis data assimilation method from relatively small domains with simple geometry, as has been studied in previous works, toward a large domain with highly complex geometry, and over complex physical phenomena depending on turbulent velocity fields. While the extension to application over the full city of Fresno and use with real observational data will require more study, we believe this first step demonstrates the feasibility of non-intrusive reduced order variational data assimilation methods as the PBDW in urban-scale real-world scenarios.

4.1 References

Références

[1] F. Archambeau, N. Mehitoua, M. Sakiz, *Code Saturne : a Finite Volume code for the computation of turbulent incompressible flows,* Int. J. Finite Volumes, 2004.

[2] A. Brooks, T. Hughes *Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations*, Computer Methods in Applied Mechanics and Engineering, p199–259, 1982.

[3] City of Calgary, *Vehicle Kilometres Travelled in Calgary : Proposed Methodology*, 2010.

[4] M. Grepl, Y. Maday, N. Nguyen, and A.T Patera. *Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations*, M2AN, Vol 41, No 3, p575-605 (2007).

[5] F. Hecht, *New development in FreeFem++*, Journal of Numerical Mathematics, p251–265, 2012.

[6] J. Houck, J. Crouch, R. Huntley, *Review of Wood Heater and Fireplace Emission Factors*, 10th Annual Emission Inventory Meeting, 30th April-3rd May, 2001.

[7] T. Hughes, A. Brooks, *A multidimensional upwind scheme with no crosswind diffusion*, Finite element methods for convection dominated flows, vol 34, p19–35, 1979.

[8] Y. Maday. *Reduced basis method for the rapid and reliable solution of partial differential equations.* in International Congress of Mathematicians, Vol. III, p1255-1270, Eur. Math. Soc. Zurich (2006).

[9] Y. Maday, O. Mula, *A generalized empirical interpolation method : application of reduced basis techniques to data assimilation*, Analysis and numerics of partial differential equations, p221–235, 2013.

[10] Y. Maday, A. Patera, J. Penn, M.Yano, *A Parameterized-Background Data-Weak Approach to Variational Data Assimilation : Formulation, Analysis, and Application to Acoustics*, Int. J. Numer. Meth. Engng, vol 102, no. 5, p933 ?965. 2014.

[11] C. Prud'homme, D. Rovas, K. Veroy, L. Machiels, Y. Maday, A. Patera, and G. Turinici. *Reliable real-time solution of parametrized partial differential equations : Reduced-basis output bound methods.* J Fluids Engineering, Vol 124 p70-80 (2002).

[12] A. Quarteroni, G. Rozza, and A. Manzoni. *Certified Reduced Basis Approximation for Parametrized Partial Differential Equations and Applications*. Journal of Mathematics in Industry 2011 Vol 1, No 3 (2011).

[13] USEPA, *Emission Factor Documentation for AP-42, Section 13.2.1 Paved Roads*, 2011.