



MOX-Report No. 09/2016

**Relative impact of advective and dispersive processes on
the efficiency of POD-based model reduction for solute
transport in porous media**

Rizzo, C.B.; de Barros, F.P.J.; Perotto, S.; Oldani, L.;
Guadagnini, A.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it

<http://mox.polimi.it>

Relative impact of advective and dispersive processes on the efficiency of POD-based model reduction for solute transport in porous media^{*}

Calogero B. Rizzo^b, Felipe P. J. de Barros^{*}, Simona Perotto[‡],
Luca Oldani^b, Alberto Guadagnini[‡]

February 19, 2016

^b Politecnico di Milano, Milano, Italy

^{*} Sonny Astani Department of Civil and Environmental Engineering
University of Southern California, Los Angeles, California, USA

[‡] MOX, Dipartimento di Matematica, Politecnico di Milano, Milano, Italy

[‡] Dipartimento di Ingegneria Civile e Ambientale, Politecnico di Milano, Milano, Italy
calogero.rizzo@mail.polimi.it, fbarros@usc.edu, simona.perotto@polimi.it,
luca.oldani@mail.polimi.it, alberto.guadagnini@polimi.it

Keywords: Proper Orthogonal Decomposition, Advection-Dispersion Equation, Model Order Reduction

Abstract

We study the applicability of a model order reduction technique to the cost-effective solution of transport of passive scalars in porous media. Transport dynamics is modeled through the advection-dispersion equation (ADE) and we employ Proper Orthogonal Decomposition (POD) as a strategy to reduce the computational burden associated with the numerical solution of the ADE. Our application of POD relies on solving the governing ADE for selected time intervals, termed snapshots. The latter are then employed to achieve the desired model order reduction. The problem dynamics require alternating, over diverse time scales, between the solution of the full numerical transport model, as expressed by the ADE, and its reduced counterpart, constructed through the selected snapshots. We explore the way the selection of these time scales is linked to the Péclet number (Pe) characterizing transport under steady-state flow conditions

^{*}Funding from the European Unions Horizon 2020 Research and Innovation programme (Project Furthering the knowledge Base for Reducing the Environmental Footprint of Shale Gas Development FRACRISK - Grant Agreement No. 640979) and from MIUR (Italian Ministry of Education, Universities and Research - PRIN2010-11; project Innovative methods for water resources under hydro-climatic uncertainty scenarios) is acknowledged.

taking place in two-dimensional homogeneous and heterogeneous porous media. We find that the length of the time scale within which the POD-based reduced model solution provides accurate results tends to increase with decreasing Pe . This suggests that the effects of local scale dispersive processes facilitate the POD method to capture the salient features of the system dynamics embedded in the selected snapshots. Since the dimension of the reduced model is much lower than that of the full numerical model, the methodology we propose enables one to accurately simulate transport at a markedly reduced computational cost.

1 Introduction

Improving our ability to provide reliable and computationally efficient approaches to quantify transport process in porous media is of major importance to a wide range of applications. Most notably, these include the characterization of the feedback between anthropogenic activities and the subsurface environment, with direct implications on the assessment of water quality which is key to physically based development of modern strategies addressing the water-energy-food nexus.

Analysis of flow and transport phenomena in porous media often involves investigations within domains of large extent [26, 27]. Practical difficulties associated with the way one can rigorously include pore-scale modeling in the analysis of typical laboratory and field-scale settings lead to depicting of solute transport in such media through effective models [23, 29, 24, 7, 14]. Several alternative modeling options are available in this context, a common choice being a continuum-based representation grounded on the standard advection-dispersion equation (ADE) on which we focus here. Evolution of transport scenarios of practical interest to industrial and environmental applications typically occurs over temporal scales associated with the duration of several days, months and sometimes years. High quality numerical approximations are required to capture extreme values of chemical concentrations, i.e., peak values and low concentrations, constituting key environmental performance metrics with severe implications on human health through the proper quantification of the exceedance of toxicity thresholds at sensitive locations in the system [5]. Accurate numerical solution of the ADE typically requires considerable computational time and the use of appropriate discretization techniques [9]. Computational time requirements become a major challenge when several transport scenarios needs to be evaluated, as is typical in modern probabilistic risk assessment approaches [20]. It then becomes relevant to develop methods to decrease the complexity of the discrete model associated with the governing process equations [31, 30], while preserving accuracy of the solution.

Here, we focus on the analysis of the Proper Orthogonal Decomposition (POD) technique [17] and explore its ability to reduce the computational burden associated with the solution of the ADE under a variety of transport conditions characterized by diverse relative strengths of advective and dispersive

processes. POD is one of the most widely used model order reduction techniques and has been applied to a wide range of problems in the fields of data analysis, statistics and/or dynamic systems [4, 10, 19, 11]. The approach has been recently employed for the solution of fully saturated flow and transport in subsurface reservoirs [15, 18, 13, 12] and it has been shown to have the potential to considerably decrease the discrete problem size while maintaining a relatively high accuracy with respect to the solution of the original governing equation. It has also been employed in the context of preliminary studies addressing uncertainty quantification of flow in groundwater systems to speed up numerical Monte Carlo simulations of flow in the presence of random forcing and system parameters [21, 22].

The key feature of model reduction techniques is that the solution of a given partial differential equation (PDE) can be well approximated by a linear combination of a limited number of basis functions which can be conveniently selected. The coefficients of this linear combination can then be computed through the solution of the reduced system obtained via a Galerkin projection of the governing PDE. In the POD framework, the basis functions identifying the finite subspace where the model is projected are the eigenfunctions of an integral operator whose kernel is given by the spatial correlation function of the state variable of interest. A critical feature of the application of the POD technique to dynamically evolving scalar fields of the kind described by an ADE is related to the way this finite subspace is built. This is typically accomplished through (a) solving the governing PDE at a set of predefined time steps termed snapshots, (b) applying Principal Component Analysis (PCA) to the snapshots, and (c) employing the resulting principal components as basis functions to characterize the above mentioned finite subspace [3]. The proper selection of the snapshots is a critical point in the workflow, because it controls the errors associated with the model reduction. Examples of studies providing guidelines for snapshot selection can be found in [16, 2, 28] where POD is applied to subsurface flow or in [1] for simulations related to oceanic flows. The problem is particularly challenging in the presence of diverse competing physical processes driving the evolution of the system, such as in the case of the ADE where the space-time distribution of dissolved chemical concentration is governed by the relative importance of advective and dispersive/diffusive processes. We investigate the manner in which the interplay between these processes impacts the effectiveness of POD-based model reduction techniques for the computationally efficient and accurate numerical solution of the ADE in a porous medium under typical subsurface environmental conditions. We introduce a new approach, termed *Snapshot Splitting Technique* (SST), for the selection of the snapshots. This strategy leads to a marked improvement of the quality of the reduced problem solution, resulting in the ability of the reduced model to provide accurate solutions for a sustained period of time while maintaining the same number of snapshots.

The structure of this work is as follows. Section 2 includes a brief presentation of the ADE model, the POD method and the mixed Finite Element-based

POD method (FE-POD) we employ. In section 3 we illustrate the use of the mixed FE-POD method to the solution of solute transport in homogeneous and heterogeneous media. Section 4 is devoted to the presentation and discussion of our model reduction strategy and algorithm in the presence of advective and diffusive/dispersive processes of competing strength. Section 5 includes our conclusions and outlines of future developments.

2 Reduction of the ADE through Proper Orthogonal Decomposition

We introduce here the advection-dispersion equation (ADE), which we employ to describe solute transport in a porous medium at the continuum scale, and illustrate the theoretical basis and workflow for the application of the POD for the model reduction. We perform our analysis by relying on a dimensionless formulation which enables us to discriminate through the Péclet number (Pe) the relative importance of advection and diffusion/dispersion processes on the effectiveness of the POD approach to model reduction.

2.1 Dimensionless problem formulation

The general format of the ADE employed to depict the evolution of chemical concentration following a continuous point injection within a Darcy-scale velocity field \mathbf{u} is:

$$\frac{\partial c^*}{\partial t^*} - \nabla \cdot (\mathbf{D}\nabla c^*) + \mathbf{u} \cdot \nabla c^* = c_{\text{in}}^* \delta(\mathbf{0}, t^*), \quad (1)$$

where $c^* : \mathbb{R}^n \rightarrow \mathbb{R}$ is the solute concentration, \mathbf{D} is a tensor embedding the effects of diffusive and local-scale dispersive processes, c_{in}^* is the solute concentration injected per unit time, δ is the Dirac delta function and t^* is the time. We consider a steady-state velocity field \mathbf{u} given by Darcy's law:

$$\mathbf{u} = -\frac{\mathbf{K}}{\phi} \nabla h. \quad (2)$$

Here, ϕ and \mathbf{K} are the porosity and the hydraulic conductivity tensor of the porous medium, respectively. The steady-state spatial distribution of hydraulic head h is governed by:

$$\nabla \cdot [\mathbf{K}\nabla h] = 0 \quad (3)$$

equipped with proper boundary conditions.

In a two-dimensional system (i.e., $n = 2$) with x and y denoting principal directions, we have:

$$\mathbf{D} = \begin{bmatrix} D_x & 0 \\ 0 & D_y \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} u_x \\ u_y \end{bmatrix}, \quad (4)$$

where D_x and D_y are constant. Here, we consider the average flow direction to be aligned along the x -axis and define the Péclet number as:

$$Pe = \frac{\langle u_x \rangle L}{D_x}, \quad (5)$$

where $\langle u_x \rangle$ is the average value of u_x , the component of \mathbf{u} along x , and L is a characteristic length scale of the system. Next, we introduce the following dimensionless quantities:

$$\begin{aligned} x &= \frac{x^*}{L}, & y &= \frac{y^*}{L}, & c &= \frac{c^* u_x}{c_{\text{in}}^* L}, \\ t &= \frac{t^* u_x}{L}, & \mathcal{D} &= \frac{D_y}{D_x}, & \mathbf{v} &= \frac{\mathbf{u}}{\langle u_x \rangle}. \end{aligned} \quad (6)$$

The above dimensionless quantities allow recasting (1) as:

$$\frac{\partial c}{\partial t} - \frac{1}{Pe} \left(\frac{\partial^2 c}{\partial x^2} + \mathcal{D} \frac{\partial^2 c}{\partial y^2} \right) + v_x \frac{\partial c}{\partial x} + v_y \frac{\partial c}{\partial y} = \delta(\mathbf{0}, t). \quad (7)$$

Finally, we present the algebraic formulation of problem (7) using a finite element (FE) method and define the solution by $c_f(\cdot)$ and the corresponding vector by \mathbf{c}_f . We can formulate the *full discrete problem* as follows: For each $k \geq k_0$ find $\mathbf{c}_f^k \in \mathbb{R}^{n_f}$ such that

$$\left(\frac{\mathbf{M}}{\Delta t} + \mathbf{A} \right) \mathbf{c}_f^{k+1} = \mathbf{f}^{k+1} + \frac{\mathbf{M}}{\Delta t} \mathbf{c}_f^k, \quad (8)$$

with \mathbf{c}_f^k the solution at time t_k and $\mathbf{c}_f^{k_0} = \mathbf{c}_{f,0}^{k_0}$. The matrix \mathbf{M} is the mass matrix, the matrix \mathbf{A} is the stiffness matrix which is linked to \mathcal{D} , Pe and \mathbf{v} , \mathbf{f} is a known vector and Δt is the time step.

2.2 Proper Orthogonal Decomposition method

The basic idea of Proper Orthogonal Decomposition (POD) is to reduce the original problem into a new simpler problem where the solution lies in a subspace of reduced dimensionality, as compared to the space where the full model is defined. The construction of such a subspace is a challenging issue in the model development. A commonly adopted technique relies on solving the full model for a given number of time intervals, usually termed snapshots and then relying on the information embedded in them to build the desired subspace. In this context, our strategy starts from the computation of the snapshots through the solution of the full model (7) from the initial system state up to a given simulation time t_0 . A key point in this approach is then to select the snapshots and t_0 in a way that the salient features of the early system dynamics are adequately captured.

These are then employed to advance the solution of the ensuing reduced model beyond t_0 , as the concentration further progresses in the domain. In this sense, the construction of the solution subspace can be performed by establishing a suitable relationship between a proper space V_{n_f} in which the solution of the full model lies and where each element can be represented by a n_f -dimension vector, and a reduced subspace V_{n_p} with dimension $n_p \ll n_f$. Given a vector $\mathbf{x} \in \mathbb{R}^{n_f}$ that represents an element of V_{n_f} , one can always find its counterpart $\hat{\mathbf{x}} \in \mathbb{R}^{n_p}$ as:

$$\hat{\mathbf{x}} = \mathbf{P}^T \mathbf{x}, \quad (9)$$

where the columns of the n_f -by- n_p projection matrix \mathbf{P} are an orthonormal basis spanning the subspace V_{n_p} .

From an operational standpoint, computation of \mathbf{P} requires constructing the n_p basis vectors that span the subspace V_{n_p} . We do so by selecting n_p snapshots $(\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_{n_p})$ where $\mathbf{c}_i \in \mathbb{R}^{n_f}$ and defining a matrix:

$$\mathbf{X} = [\mathbf{s}_1 \quad \mathbf{s}_2 \quad \dots \quad \mathbf{s}_{n_p}], \quad (10)$$

where \mathbf{s}_i is the i -th normalized snapshot \mathbf{c}_i . We define:

$$\mathbf{C}_s = \mathbf{X}^T \mathbf{X}, \quad (11)$$

where \mathbf{C}_s is an n_p -by- n_p symmetric matrix. Using the spectral theorem, we can rewrite \mathbf{C}_s as:

$$\mathbf{C}_s = \mathbf{G} \mathbf{\Lambda} \mathbf{G}^T, \quad (12)$$

where \mathbf{G} is the orthogonal matrix of the eigenvectors of \mathbf{C}_s and $\mathbf{\Lambda}$ is a diagonal matrix whose entries are the eigenvalues of \mathbf{C}_s . Since the dimension of \mathbf{C}_s is usually small, the computation of matrices \mathbf{G} and $\mathbf{\Lambda}$ requires a small amount of time compared to the requirement for the solution of the full numerical model.

By exploiting the well-known relation between the spectral decomposition in (12) and the singular value decomposition of X [13, 12] we obtain that:

$$\mathbf{P} = \mathbf{X} \mathbf{G} \mathbf{\Lambda}^{-1/2}. \quad (13)$$

so that $\mathbf{P}^T \mathbf{P} = \mathbf{I}$ with \mathbf{I} the identity matrix of order n_p .

The matrix \mathbf{P} contains all information about the relationship between the two spaces V_{n_f} and V_{n_p} and the columns of the projection matrix $\tilde{\mathbf{p}}_i$ are a set of basis vectors of V_{n_p} . Note that in general:

$$\mathbf{P} \mathbf{P}^T \neq \mathbf{I}, \quad (14)$$

where the identity matrix \mathbf{I} is here of order n_f . Equation (14) is key to characterize the POD dynamics. A vector $\mathbf{v} \in V_{n_f}$ can be split as $\mathbf{v} = \tilde{\mathbf{v}} + \mathbf{v}_0$, where $\tilde{\mathbf{v}}$ and \mathbf{v}_0 belong to the subspace generated by \mathbf{P} and to the associated orthogonal subspace, respectively. Therefore $\mathbf{v}_0 \cdot \phi = 0$ for every ϕ linear combination of

the POD basis vectors. Using the properties of the projection matrix \mathbf{P} , we can write:

$$\tilde{\mathbf{v}} = \mathbf{P}\mathbf{P}^T\mathbf{v}. \quad (15)$$

The matrix $\mathbf{P}\mathbf{P}^T$ acts as a filter extracting the component of a function belonging to the POD subspace. We note that both vectors $\tilde{\mathbf{v}}$ and \mathbf{v} have a dimension equal to n_f . In other words, increasing the process dynamics captured by the POD subspace (e.g., through an appropriate selection of a given number of snapshots or by increasing the number of snapshots) leads to a decreased error induced by the model reduction. Given the dynamics of the system considered, it is then relevant to derive a flexible formulation which enables us to increase the dimension of the POD subspace by relying on a fixed number of snapshots, i.e., on a given computational effort. We do so in the following section and then assess the benefit of such a formulation by a suite of computational examples.

In practice, the POD based reduced model is obtained by projecting equations (8) over the subspace induced by the matrix \mathbf{P} , i.e., the subspace containing all the possible linear combinations of the columns of \mathbf{P} . We introduce the solution of the reduced problem $c_p(\cdot)$ with its corresponding vector \mathbf{c}_p . The *reduced problem* using the POD method reads:

For each $k \geq k_0$ find $\mathbf{c}_p^k \in \mathbb{R}^{n_p}$ such that

$$\mathbf{P}^T \left(\frac{\mathbf{M}}{\Delta t} + \mathbf{A} \right) \mathbf{P} \mathbf{c}_p^{k+1} = \mathbf{P}^T \mathbf{f}^{k+1} + \mathbf{P}^T \left(\frac{\mathbf{M}}{\Delta t} \right) \mathbf{P} \mathbf{c}_p^k, \quad (16)$$

with $\mathbf{c}_p^{k_0} = \mathbf{P}^T \mathbf{c}_{f,0}^{k_0}$ the projection of the initial value.

2.3 Snapshot Splitting Technique (SST) for increasing the POD subspace dimension

Next, we describe a new approach to improve the classic POD method. Given n_p snapshots $(\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_{n_p})$ where $\mathbf{c}_i \in \mathbb{R}^{n_f}$, for each snapshot we define n_l new snapshots \mathbf{k}_{ij} such that:

$$\mathbf{k}_{ij} = g_j(\mathbf{c}_i), \quad (17)$$

where $g_j(\cdot) : \mathbb{R}^{n_f} \rightarrow \mathbb{R}^{n_f}$ for $i = 1, \dots, n_p$ and $j = 1, \dots, n_l$. If the vectors associated with each snapshot are independent, then n_l basis vectors are obtained for each snapshot, i.e., the POD subspace will have a dimension of $n_p \times n_l$. In such a case, the columns of matrix (10) are the vectors $\mathbf{k}_{ij} \in \mathbb{R}^{n_f}$ subject to appropriate normalization. The POD method can then be applied as illustrated in 2.2.

The strategy for the selection of the functions $g_j(\cdot)$ is not unique, as long as the generated vectors are independent. The key concept we employ here for the definition of $g_j(\cdot)$ is that the domain can be partitioned into areas where the solution dynamics are similar. We define the maximum coefficient of the snapshots set as:

$$c_{\max} = \max_i \max_j c_{ij}, \quad (18)$$

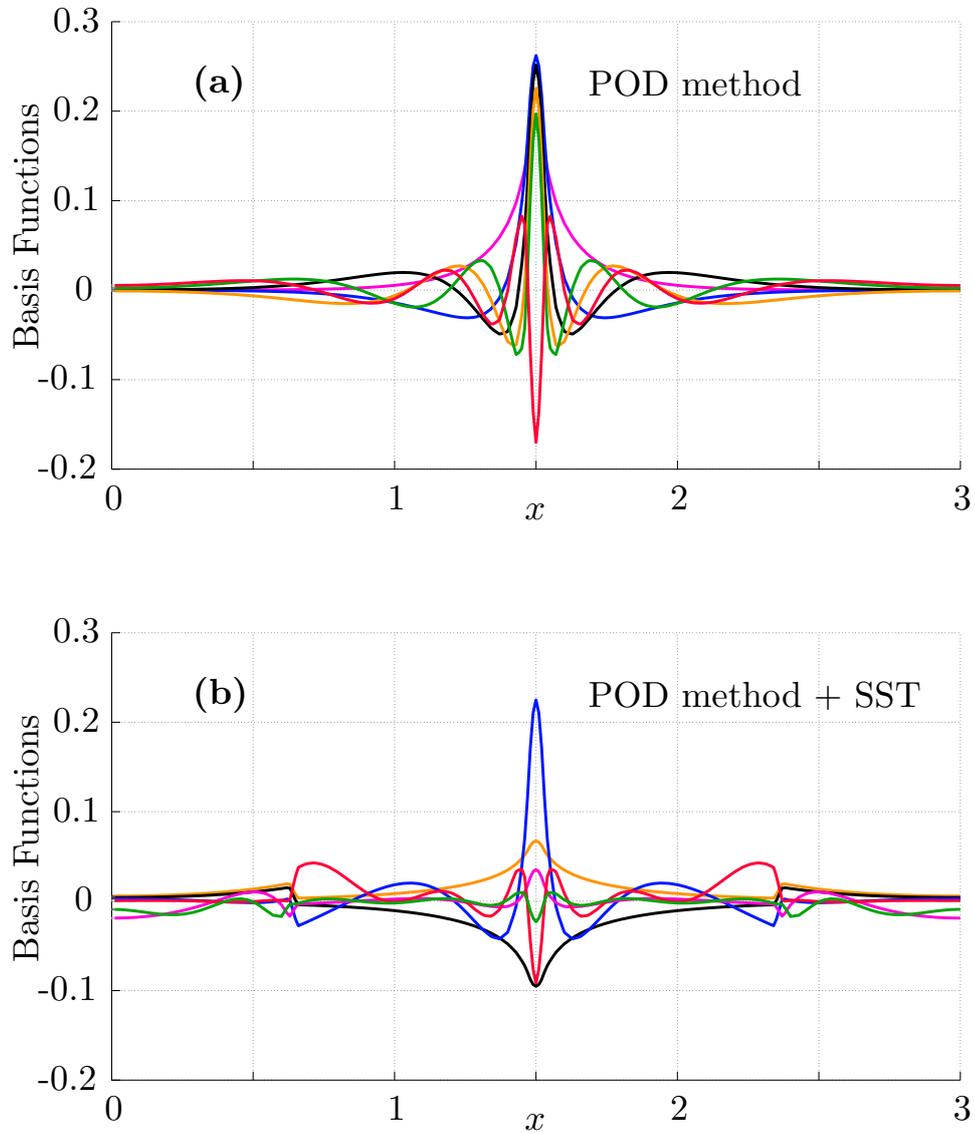


Figure 1: Distribution along a longitudinal profile (i.e., aligned in the x direction) of the basis functions $\tilde{\mathbf{p}}_i$ ($i = 1, \dots, 7$) obtained (a) without and (b) with the *Snapshot Splitting Technique* (SST) for a purely diffusive setting. The solute source is located at $x = 1.5$.

where $i = 1, \dots, n_p$ and c_{ij} is the j -th component of the vector \mathbf{c}_i . We partition $[0, c_{\max}]$ into n_l intervals introducing $n_l + 1$ values ξ_i with $i = 0, \dots, n_l$ such that:

$$0 = \xi_0 < \xi_1 < \xi_2 < \dots < \xi_{n_l-1} < \xi_{n_l} = c_{\max}. \quad (19)$$

Finally, we define the functions $g_j(\cdot)$ such that:

$$[g_j(\mathbf{v})]_m = \begin{cases} v_m & \text{if } v_m < \xi_j, \\ 0 & \text{otherwise,} \end{cases} \quad (20)$$

where $j = 1, \dots, n_l$, $m = 1, \dots, n_f$ and v_m is the m -th component of the vector $\mathbf{v} \in \mathbb{R}^{n_f}$. In our application, we select a logarithmic distribution for ξ_i , other distribution being admissible. Since we are studying the case of continuous injection, the value c_{\max} occurs at the source location. A logarithmic distribution of ξ_i takes into account that the fast solution dynamics are close to the source. As a consequence, our technique allows enriching the POD subspace in these critical zones of the evolving solute concentration.

To highlight the differences between the way the POD subspace is constructed by employing the typical approach illustrated in Section 2.2 and our proposed technique to correct such construction by relying on snapshots defined in (17), we term our approach as *Snapshot Splitting Technique* (SST) and exemplify its beneficial effect by considering a purely diffusive problem (i.e., $\mathbf{u} = \mathbf{0}$ in (1)) with a continuous point injection located at $x = y = 1.5$ in a 3×3 rectangular domain representing a homogeneous porous medium (see also Section 3 for details about the numerical solution of the full model). Figure 1 depicts the distribution along a longitudinal profile (i.e., aligned in the x direction) of the basis functions obtained without (Figure 1a) and with (Figure 1b) the application of the SST. In this illustrative example we select the snapshots \mathbf{c}_i from the first 10 time steps within which the full model is solved. Using the approach illustrated in Section 2.2, the results clearly show that all basis functions are significantly different from zero solely at locations very close to the source. This implies that (a) employing the POD subspace generated via these basis functions does not lead to an accurate representation of the system dynamics at locations far away from the source, so that (b) the ensuing reduced model can propagate in time only the information associated with system states close to the source. Otherwise, application of the SST to increase the dimension of the POD subspace (while relying on a fixed number of snapshots) allows constructing a set of basis functions which carry significant information at a variety of locations in the domain.

3 Application of POD to Advective-Dispersive Transport driven by Continuous Solute Injection

In this section we apply the methods discussed in Section 2 to reduce the dimensionality of a transport problem described by equation (7). We start by consid-

ering a homogeneous porous medium where a uniform velocity field ($v_x = 1$ and $v_y = 0$) takes place. We employ a Finite Element (FE) approach ($\mathbb{P}1$ elements) for spatial discretization and the Backward Euler scheme for time discretization with discretization step $\Delta t = 0.05$. We recall that while we ground our examples on FE, POD reduction can also be applied in the presence of diverse numerical methods, e.g., Finite Volumes. The computational domain is a 5×3 rectangle with a structured triangular mesh. Each length unit is divided into 35 segments, the computational grid being then composed by 36750 triangles. The number of degrees of freedom of the $\mathbb{P}1$ space with open boundary conditions coincides with the number of nodes of the numerical grid, i.e., 18656. A continuous point injection source is placed at $x = y = 1.5$ in all settings we examine here. When the SST illustrated in Section 2.3 is employed for the construction of the POD subspace, we choose $n_l = 10$ with a logarithmic distribution for the ξ_j as explained in 2.3.

Without loss of generality, we set $\mathcal{D} = 1$ so that the only unfixed parameter is the Péclet number (Pe). To explore the effect of Pe on the behavior of the reduced model, as compared against the full numerical solution \mathbf{c}_f^k of the model (8), we use the following steps for the computation of \mathbf{c}_p^k at time step k :

- (S1) For the first N_t time steps ($k \leq N_t$), \mathbf{c}_p^k is computed by solving the full numerical model (8), so $\mathbf{c}_p^k = \mathbf{c}_f^k$.
- (S2) We compute the projection matrix (13) via the POD method upon relying on the solution computed at step S1. The snapshots are taken as \mathbf{c}_f^k at time step $k \leq N_t$.
- (S3) Using the projection matrix and the discrete problem defined in (16), we compute \mathbf{c}_p^k for additional N_a time steps.

In all our computational examples we consider $N_t = 10$ and $N_a = 90$. Note that the choice of these values is not influential for the demonstration of our model reduction strategy. A key requirement to verify the efficiency of the method is that the final number of reduced model steps is larger than the final number of the full model steps, while keeping a satisfactory computational accuracy. The influence of N_t and N_a on the optimization of the procedure is case dependent and we leave this analysis to future studies.

We perform our computations with the reduced model in the absence or in the presence of the SST introduced in Section 2.3. When the SST is employed during step S2, the solution will be denoted as c_{p+SST}^k or \mathbf{c}_{p+SST}^k .

The quality of the ensuing solutions obtained with the reduced model is compared by computing the relative error E_s^k between the full numerical problem solution c_f and the reduced problem solutions, i.e., c_p or c_{p+SST} , which is

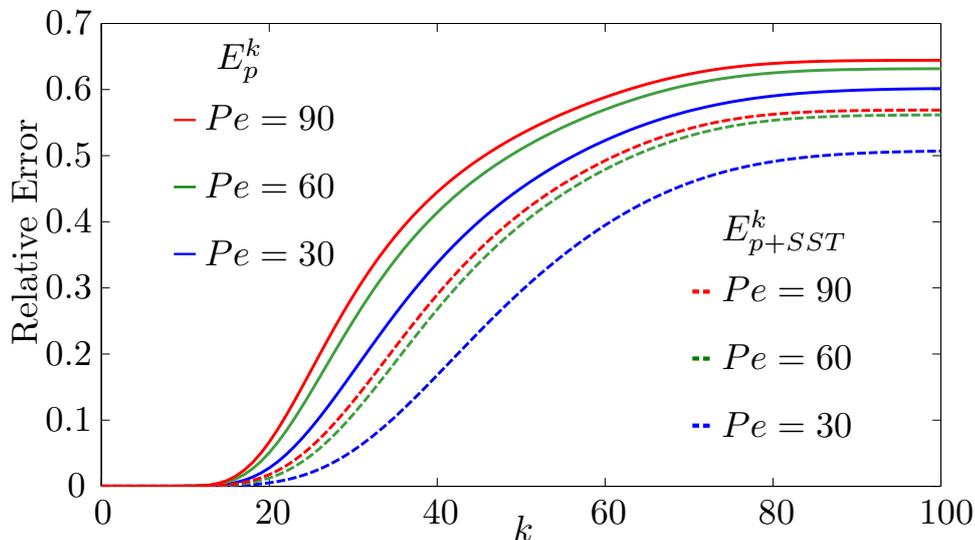


Figure 2: Effect of the *Snapshot Splitting Technique* (SST) on the evolution of the relative error (21) with the number of time steps, k , after the reduced model is employed (full model is run for $k \leq 10$) for a homogeneous system with diverse values of Pe .

computed for each time step k as:

$$E_s^k = \frac{\|c_f^k - c_s^k\|_{L^2(\Omega)}}{\|c_f^k\|_{L^2(\Omega)}}, \quad (21)$$

where subscript $s = p + SST$ or p depending on whether the SST is employed or not in the construction of the reduced model.

Figure 2 depicts the relative error (21) as a function of the number of time steps k and for three selected Péclet numbers, i.e., $Pe = 30, 60, 90$, identifying mildly to highly advective transport settings. The relative error increases in a nonlinear fashion with Pe and with the time elapsed since the last snapshot is computed from the full numerical model ($k = 10$ in the example). The plateau to which the relative errors tends is due to the open boundary conditions that allow solute mass to exit the domain at the downstream boundary. Figure 2 reveals that employing the proposed SST enables us to increase the effectiveness of the information content embedded in the snapshots extracted from the solution of the full model, this leading to a remarkable reduction in the associated relative error.

As a complement to Figure 2, Figure 3 depicts the spatial distributions of solute concentration c_f^k , c_p^k and c_{p+SST}^k along the longitudinal cross-section at

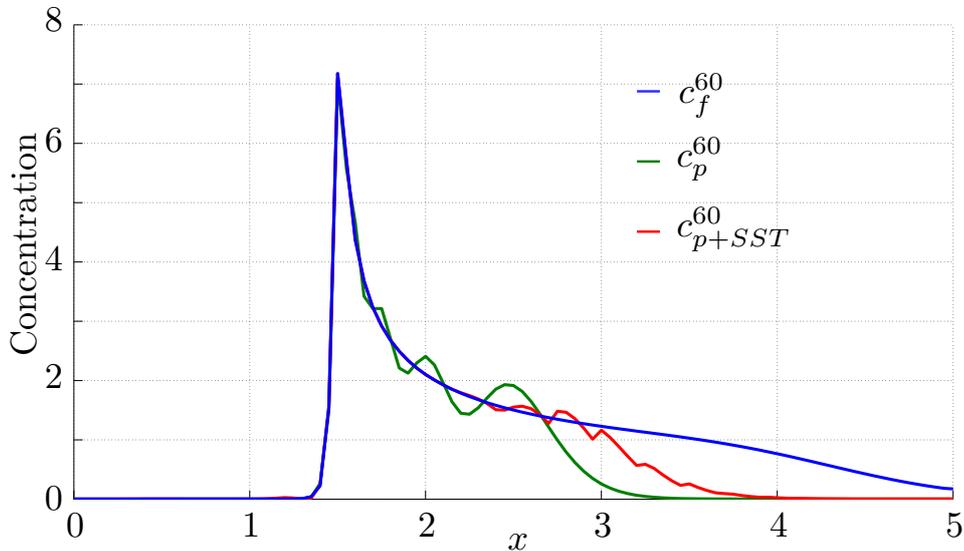


Figure 3: Spatial distributions of solute concentration computed by the full model and the reduced model with and without activating the *Snapshot Splitting Technique* (SST) at simulation time step $k = 60$. Results are depicted for the homogeneous test setting along the longitudinal cross-section passing through the injection point for $Pe = 30$.

$y = 1.5$ (i.e., passing through the injection point) for $Pe = 30$ and time step $k = 60$, i.e., after 50 time steps from the time at which the latest snapshot from the full model has been taken. Notably, the reduced solution associated with the SST, c_{p+SST}^k , displays much less severe spatial fluctuations than its counterpart c_p^k . This result confirms the beneficial effect of the proposed SST to the reduced model even after a considerable time has elapsed since the observation time associated with the last snapshot computed via the full model.

Figures 4a, 4b and 4c respectively depict the spatial distributions of solute concentration obtained by the full model solution and the POD approximation without and with the use of SST. Numerical results are illustrated for the observation time corresponding to $k = 60$ at which the relative error is high and the approximation is not so reliable in both cases. Even as severe spatial fluctuations can be observed, it is clear that employing the SST tends to drive the solution of the reduced model towards the behavior displayed by the true system dynamics, as represented more faithfully by the full model solution. It can also be noted that application of the SST has a beneficial effect and renders an approximation of improved quality in a larger area around the point source, when compared against the traditional application of the POD method.

As an additional and challenging test case, we then consider the performance

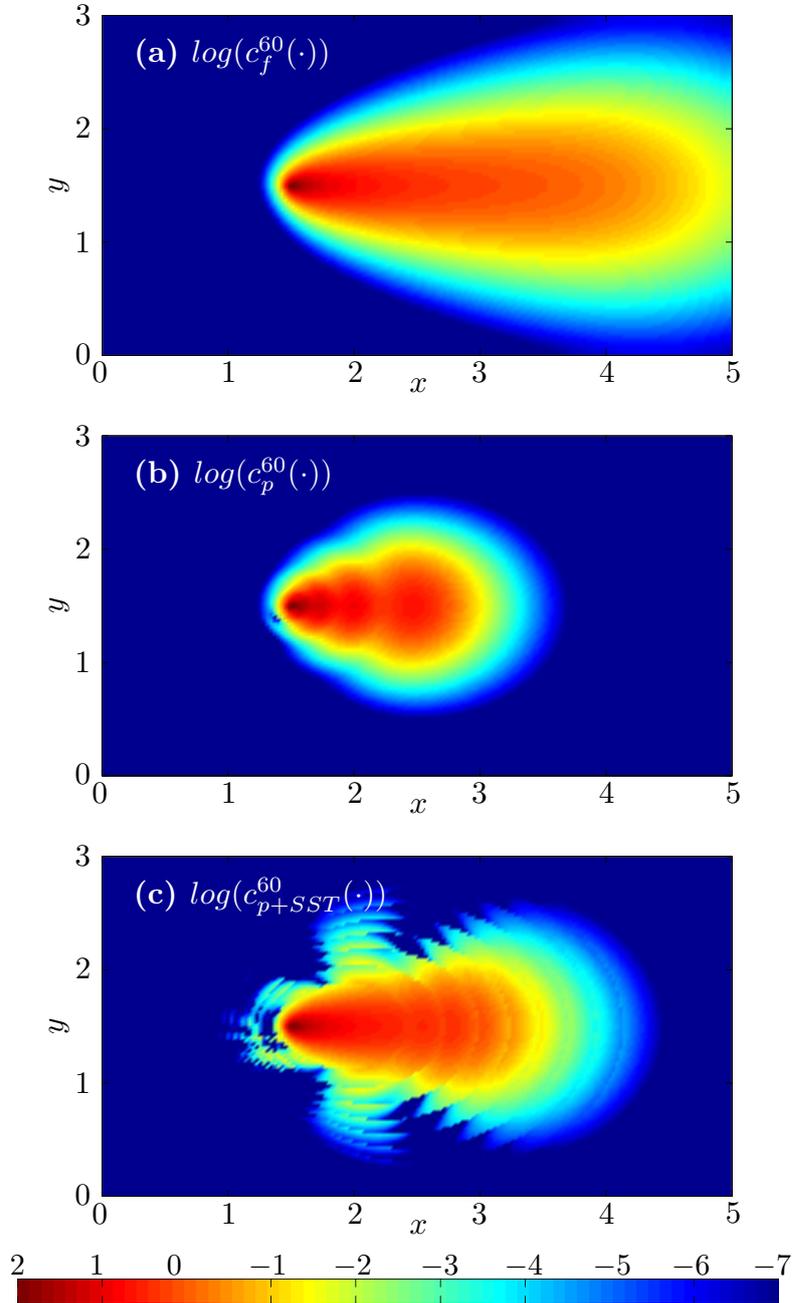


Figure 4: Spatial distributions of solute concentration obtained by (a) the full model solution and the POD approximation (b) without and (c) with the use of the *Snapshot Splitting Technique* (SST). Results correspond to the longitudinal profiles depicted in Figure 3. Color images are in logarithmic scales.

of the reduced model in a heterogeneous porous medium. Here we consider heterogeneity to stem from the spatial variability of the hydraulic conductivity field, which we model as a stationary random process of space. We generate a random realization of an isotropic $\mathbf{Y} = \log \mathbf{K}$ field, where \mathbf{Y} is the standard normal distribution, with correlation scale $\lambda = 2L$ and exponential isotropic covariance model using the widely tested SGeMS software [25].

In practical cases, the Péclet number is computed using the \mathbf{K} correlation scale as characteristic length. Therefore, we introduce the effective Péclet number \widehat{Pe} . The relation between the effective Péclet number \widehat{Pe} and the Péclet number defined in (5) is:

$$\widehat{Pe} = \frac{\langle \mathbf{u} \rangle \lambda}{D} \simeq 2Pe, \quad (22)$$

where $\langle \mathbf{u} \rangle$ is the average velocity magnitude over the domain and D is the diffusion coefficient of the system (in this case $D = D_x = D_y$).

The steady-state Darcy velocity field \mathbf{u} is obtained by solving the governing equations (2) and (3) upon setting permeameter-like boundary conditions to the system, corresponding to a unit head drop across the domain.

Following the procedure used for the homogeneous case, the transport problem is solved by employing the FE method ($\mathbb{P}1$ elements) for space discretization and the Backward Euler scheme for time discretization with $\Delta t = 0.05$. The spatial distributions of c_f^k, c_p^k are then calculated for the same value of Pe considered in the homogeneous setting, corresponding to $\widehat{Pe} = 60, 120, 180$. Figure 5 depicts the temporal evolution of the relative error (21), suggesting that, while the overall quality of the reduced model solution deteriorates with time, it does so to a lesser extent than the corresponding solution associated with the homogeneous set-up (see also Figure 2). This result might be related to the observation that the solute plume tends to follow the high conductivity paths in the system [8] and the effect of these paths close to the source tend to be influential to the plume behavior over time [6]. As such, capturing these features by the POD techniques enables the reduced solution to mimic the full model for longer time than in the homogeneous setting, where no preferential conductivity path is present. The beneficial effects of the SST for the enrichment of information embedded in the POD subspace is clear also in the presence of heterogeneity (Figure 5). This notwithstanding, it can be noted that one cannot propagate the reduced model solution in time indefinitely, without incurring in significant approximation errors. This is clearly observed in Figure 6, where the spatial distribution of c_f^k, c_p^k and c_{p+SST}^k are depicted at time step $k = 60$ for $Pe = 30$.

4 Projection Subspace Update

The results of Section 3 clearly show that the error between the reduced and full model solutions tends to increase with time after the last snapshot has been observed and tends to increase until it becomes not negligible after a certain

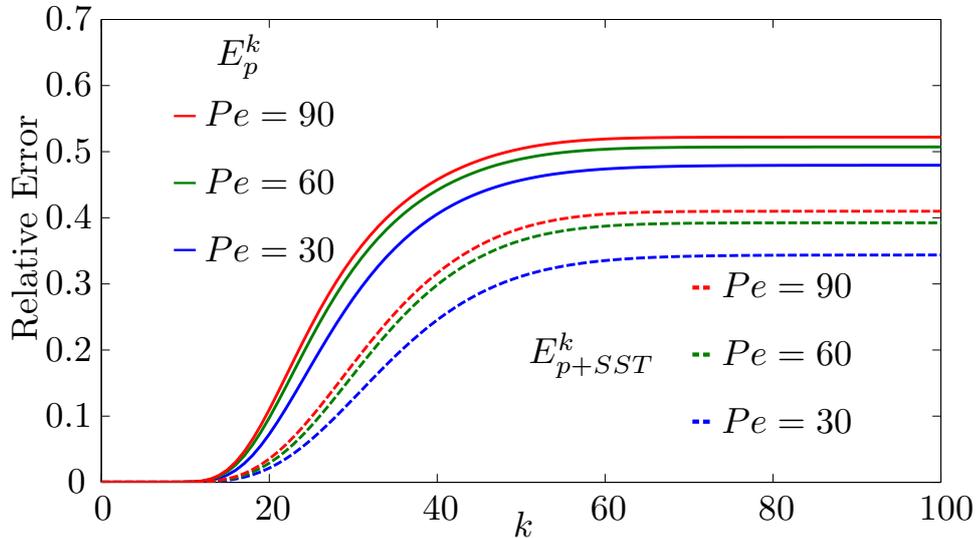


Figure 5: Effect of the *Snapshot Splitting Technique* (SST) on the evolution of the relative error (21) with the number of time steps, k , after the reduced model is employed (full model is run for $k \leq 10$) for a heterogeneous system with diverse values of Pe .

time has elapsed. This is due to the fact that the information content in the generated POD subspace (even by relying on SST) is not large enough to capture all of the key details of the dynamically evolving concentration field on the basis of a limited number of snapshots. As such, the quality of the approximation tends to deteriorate as time advances since the solution significantly changes compared to the early time states. The strategy we follow here, in order to make optimal use of the POD model reduction strategy in these types of dynamically evolving systems, is based on alternating between the full model FE and POD-FE solution of the transport problem. By doing so, the POD subspace is dynamically updated over discrete temporal intervals where snapshots are extracted from the FE full problem solution and processed to construct the reduced space basis.

For convenience, and given the results of Section 3, we illustrate our findings by relying on the projection matrix (13) constructed through the SST introduced in 2.3. Figure 7 depicts the temporal evolution of the error (21) obtained by solving the full model in the heterogeneous setting previously described for a fixed number $N_t = 10$ of time steps and alternating the use of the full model with the reduced POD-FE solution for diverse numbers of iterations, i.e., $n_{POD} = 10, 20, 30$, and for $Pe = 30, 60, 90$. It can be noted that alternating between the two solutions enables one to update the POD basis in time so that the updated POD subspace can follow closely the system dynamics. The global error is seen to

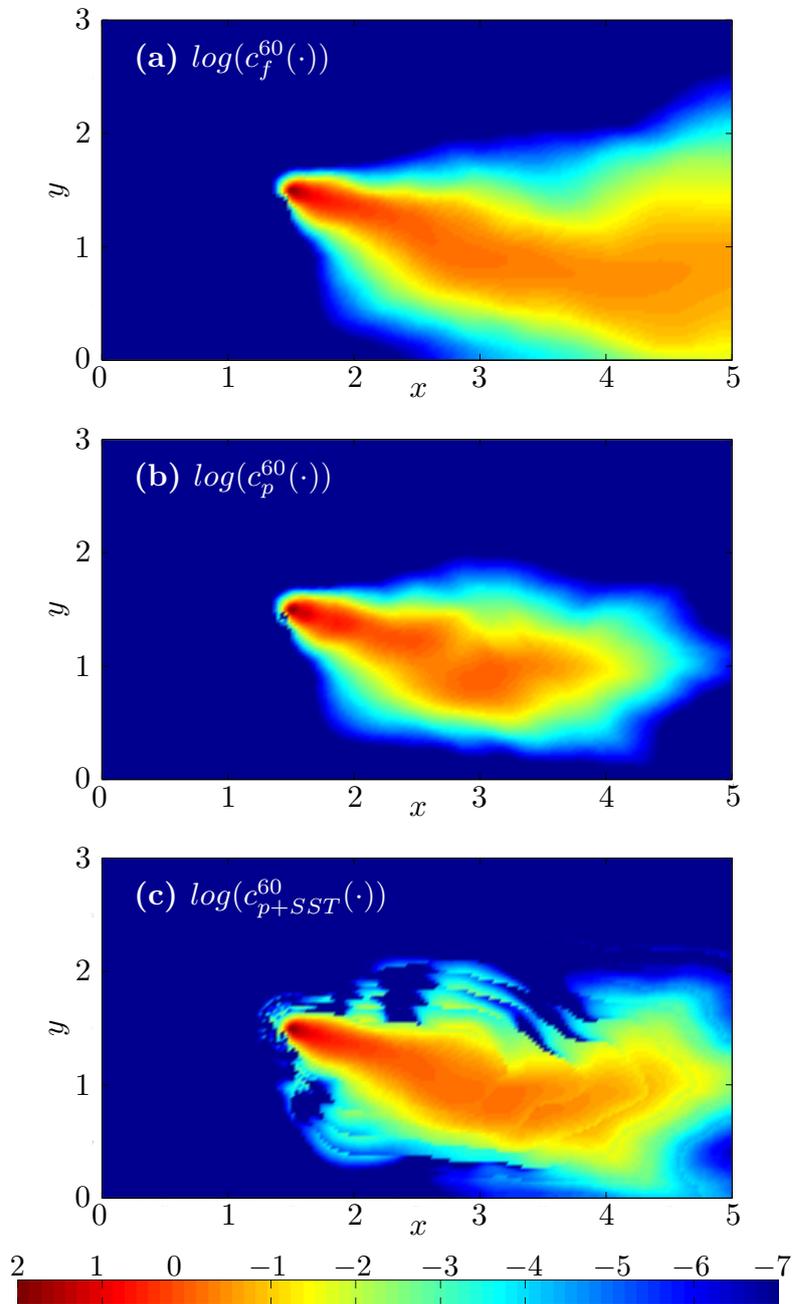


Figure 6: Spatial distributions of solute concentration obtained by (a) the full model solution and the POD approximation (b) without and (c) with the use of the *Snapshot Splitting Technique* (SST). Color images are in logarithmic scales.

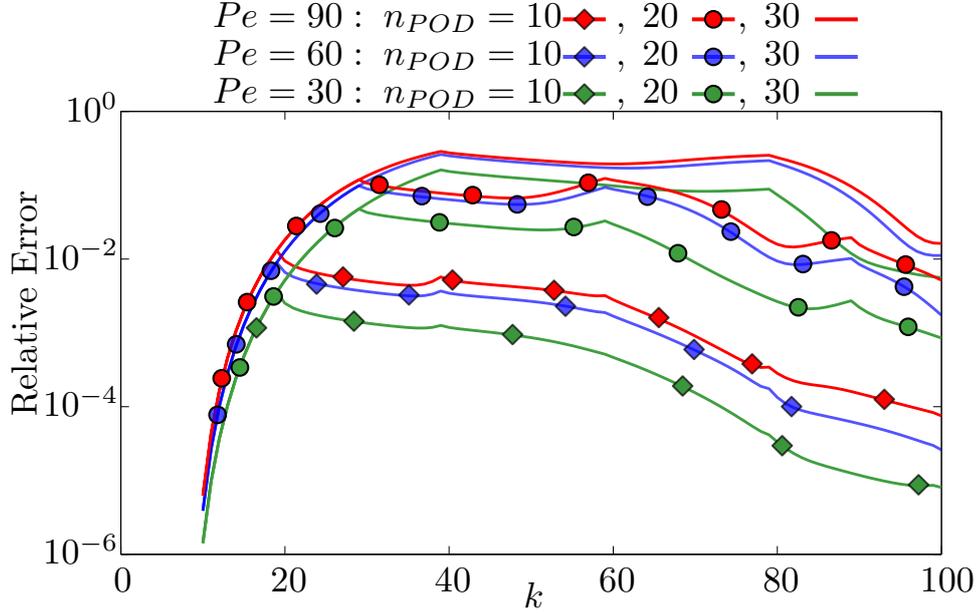


Figure 7: Temporal evolution of the relative error E^k (21) for diverse values of Pe and employing a subspace update after a fixed interval n_{POD} .

consistently decrease after each full problem iteration cycle. This highlights the need for an update in time of the POD basis to obtain a reduced model yielding a good accuracy and suggests that there is a beneficial cumulative effect to the quality of the reduced model solution of subsequent basis updates.

The results shown in Figure 7 suggest that setting the number of iterations n_{POD} a priori does not allow a good adaptation of the POD algorithm to the system dynamics. For example, it is seen that the concentration configuration in the domain changes rapidly after a few time steps from the beginning of solute injection so that the relative distance between the full model solution and the POD subspace increases sharply within few time steps and a model update is required. Otherwise, the variability of concentration in space tends to be smooth out after a few time steps so that a reduced model based on this type of information is prone to retain a relatively high global accuracy for a sustained period of time. This behavior can be quantified and embedded in the dynamic POD update procedure through the definition of the following residual for each time step:

$$\mathcal{R}_k = \frac{\|(\frac{\mathbf{M}}{\Delta t} + \mathbf{A})\mathbf{P}\mathbf{c}_p^k - (\mathbf{f}^k + \frac{\mathbf{M}}{\Delta t}\mathbf{P}\mathbf{c}_p^{k-1})\|_2}{\|\mathbf{f}^k\|_2}, \quad (23)$$

where $\|\cdot\|_2$ is the Euclidean norm. Note that (23) does not depend of the full model solution but relies solely on the reduced problem. We then employ (23) to

design the adaptive algorithm for the update of the POD subspace basis which is described in the following:

- (U1) A solution \mathbf{c}_p^k is computed by solving the full discrete model, i.e., $\mathbf{c}_p^k = \mathbf{c}_f^k$, for the first N_t time steps ($N_t = 10$ in our example, i.e., $k \leq 10$).
- (U2) The projection matrix is computed via the POD method relying on the full model solution computed in U1, \mathbf{c}_p^k representing a snapshot.
- (U3) The model is projected to the POD subspace yielding the reduced discrete model which is solved until the residual \mathcal{R}_k in (23) attains a given threshold value, ϵ .
- (U4) Starting from the last solution obtained in U3, we switch to the solution of the full discrete model; in our example, we solve the full model for additional N_t time steps, a full analysis of the feedback between N_t and the transport setting being outside the scope of this work.
- (U5) The solution computed in U4 together with the old snapshots are employed to update the projection matrix used in the POD reduction method. Transport simulation progresses from step U3 with the newly constructed reduced model.

Figure 8 depicts the temporal evolution of the global relative error (21) as a function of k and for different choices of Pe and of the selected threshold ϵ , for the same problem setting associated with Figure 7. It can be noted that the error is consistently low, due to the improved efficiency according to which the subspace updating is performed. As expected, the largest global relative errors are related to the selected threshold. Lowering the latter yields to a frequent update, resulting in an increased computational time.

Finally, using the subspace update strategy presented in this section, we can simulate a dispersion and advection problem for a given time interval using a reduced amount of computational resources and being able to control the error through the value of the threshold ϵ . With reference to this point, Figure 9 clearly shows that concentration values computed at $k = 60$ using the alternating algorithm with $Pe = 30$ and $\epsilon = 10^{-1}$ is almost identical to the solution resulting from the full numerical problem shown in Figure 6a. It is remarkable to note that the results for 70 time steps out of 100 have been obtained by way of the reduced model.

5 Summary and Conclusions

Our work leads to the following major conclusions:

1. We present a new efficient algorithm for improving the performance of the numerical solution of transport problems associated with migration of

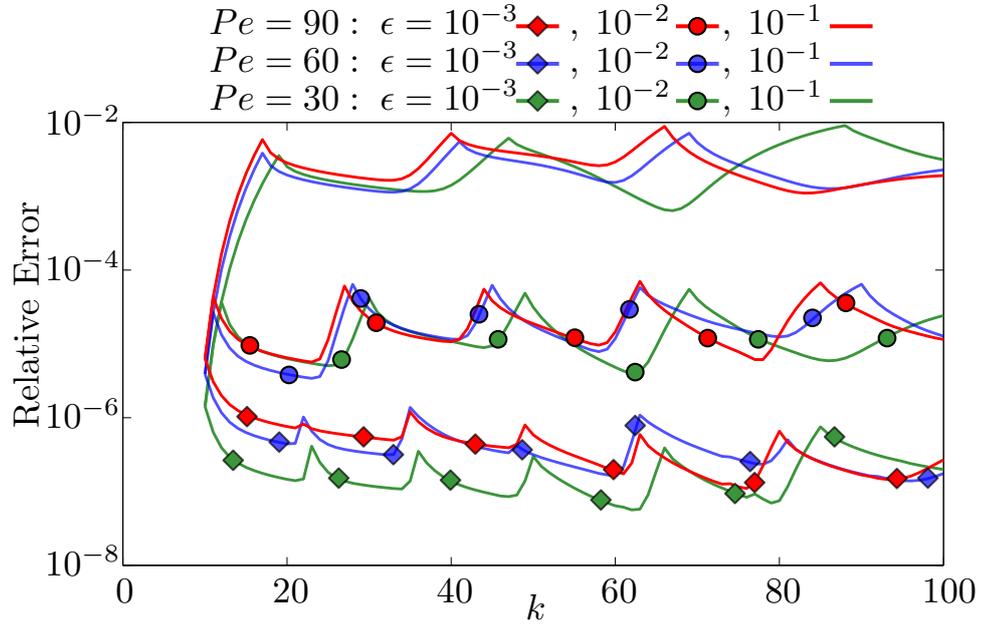


Figure 8: Temporal evolution of the relative error E^k (21) for diverse values of Pe and imposing a different threshold on the residual \mathcal{R}_k (23).

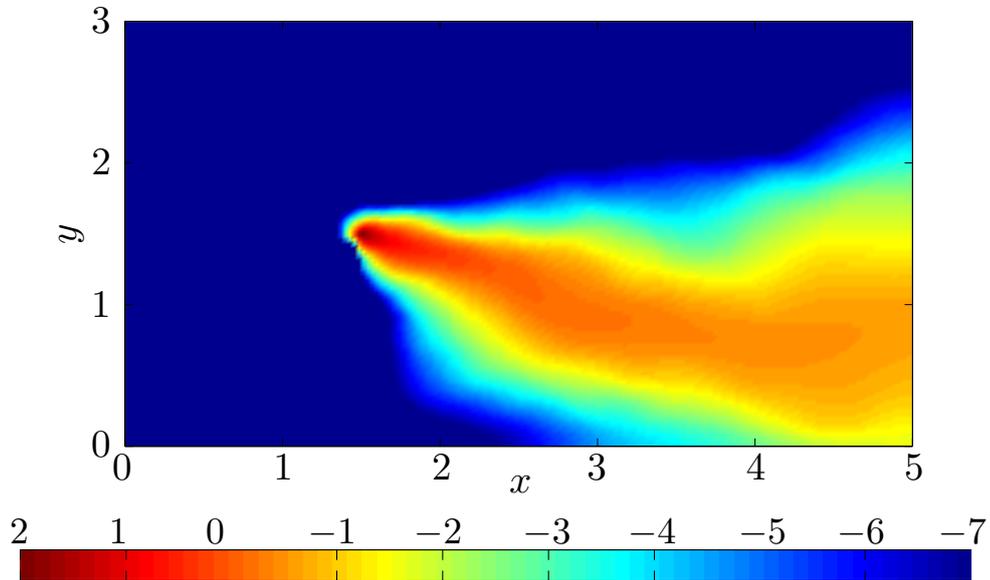


Figure 9: Spatial distributions of solute concentration obtained by the alternating algorithm solution. Color images are in logarithmic scales.

conservative dissolved chemicals in a porous medium under the action of advective and diffusive/dispersive processes. The proposed algorithm is based on a POD model reduction approach and employs a combination of the full and reduced model, the solution of which is alternated in time. The adaptive model reduction strategy we propose relies on (i) employing a *Snapshot Splitting Technique* (SST), which enables us to enrich the information content associated with the basis employed to construct the POD subspace on the basis of a given number of snapshots, and (ii) alternating between the solution of the full and reduced model through the definition of a threshold on the algebraic residual (23). We remark that the optimization of the model alternation sequence as a function of the domain heterogeneity and transport setting is outside the scope of the present contribution, which is keyed to illustrate and demonstrate the viability of the proposed strategy. As such, our work shows that using our subspace updating strategy enables us to simulate an advection-dispersion problem for a given time interval using a reduced amount of computational resources and with the ability to control the computational error through a desired threshold ϵ .

2. We explore the effect of the Péclet number on the quality of the reduced model approximation. We find that the POD method is associated with an improved performance for low Péclet numbers. This is related to the observation that the snapshots employed to build the basis vectors are taken from the full model solution and the concentration dynamics for low Péclet numbers are strongly linked to the past system states.
3. The improvement associated with the proposed SST enables us to increase the number of steps within which the reduced model can provide a viable solution of the system behavior in the time-alternating algorithm, yielding a numerical solution associated with increased computational efficiency.

The *Snapshot Splitting Technique* and the time alternating model reduction algorithm has the potential to be extended to diverse types of problems governing scalar transport in different types of flow fields and media. Cycling between the solution of the full and reduced models can provide a marked decrease of the complexity of a discrete problem characterized by a considerably high physical dimension according to which an ad hoc fast solver can be built.

References

- [1] Y. Cao, J. Zhu, Z. Luo, and I. Navon. Reduced-order modeling of the upper tropical pacific ocean model using proper orthogonal decomposition. *Computers and Mathematics with Applications*, 52(8-9):1373–1386, 2006.

- [2] M. Cardoso, L. Durlafsky, and P. Sarma. Development and application of reduced-order modeling procedures for subsurface flow simulation. *International Journal for Numerical Methods in Engineering*, 77(9):1322–1350, 2009.
- [3] A. Chatterjee. An introduction to the proper orthogonal decomposition. *Current Science*, 78(7):808–817, 2000.
- [4] D. Crommelin and A. Majda. Strategies for model reduction: Comparing different optimal bases. *Journal of the Atmospheric Sciences*, 61(17), 2004.
- [5] F. P. J. de Barros, S. Ezzedine, and Y. Rubin. Impact of hydrogeological data on measures of uncertainty, site characterization and environmental performance metrics. *Advances in Water Resources*, 36:51–63, 2012.
- [6] F. P. J. de Barros and W. Nowak. On the link between contaminant source release conditions and plume prediction uncertainty. *Journal of Contaminant Hydrology*, 116(1):24–34, 2010.
- [7] M. Dentz, T. Le Borgne, A. Englert, and B. Bijeljic. Mixing, spreading and reaction in heterogeneous media: A brief review. *Journal of Contaminant Hydrology*, 120:1–17, 2011.
- [8] Y. Edery, A. Guadagnini, H. Scher, and B. Berkowitz. Origins of anomalous transport in heterogeneous media: Structural and dynamic controls. *Water Resources Research*, 50(2):1490–1505, 2014.
- [9] B. Esfandiar, G. Porta, S. Perotto, and A. Guadagnini. Impact of space-time mesh adaptation on solute transport modeling in porous media. *Water Resources Research*, 51(2):1315–1332, 2015.
- [10] I. Jolliffe. *Principal component analysis*. Wiley Online Library, 2005.
- [11] M. E. Kowalski and J.-M. Jin. Model-order reduction of nonlinear models of electromagnetic phased-array hyperthermia. *Biomedical Engineering, IEEE Transactions on*, 50(11):1243–1254, 2003.
- [12] K. Kunisch and S. Volkwein. Galerkin proper orthogonal decomposition methods for parabolic problems. *Numerische Mathematik*, 90(1):117–148, 2001.
- [13] K. Kunisch and S. Volkwein. Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics. *SIAM Journal on Numerical Analysis*, 40(2):492–515, 2002.
- [14] T. Le Borgne, M. Dentz, and J. Carrera. Lagrangian statistical model for transport in highly heterogeneous velocity fields. *Physical Review Letters*, 101(9):090601, 2008.

- [15] H. Li, Z. Luo, and J. Chen. Numerical simulation based on pod for two-dimensional solute transport problems. *Applied Mathematical Modelling*, 35(5):2489–2498, 2011.
- [16] X. Li and B. X. Hu. Proper orthogonal decomposition reduced model for mass transport in heterogenous media. *Stochastic Environmental Research and Risk Assessment*, 27(5):1181–1191, 2013.
- [17] J. L. Lumley. The structure of inhomogeneous turbulent flows. *Atmospheric Turbulence and Radio Wave Propagation*, pages 166–178, 1967.
- [18] Z. Luo, H. Li, Y. Zhou, and Z. Xie. A reduced finite element formulation based on pod method for two-dimensional solute transport problems. *Journal of Mathematical Analysis and Applications*, 385(1):371–383, 2012.
- [19] H. V. Ly and H. T. Tran. Proper orthogonal decomposition for flow calculations and optimal control in a horizontal cvd reactor. Technical report, DTIC Document, 1998.
- [20] M. Moslehi, R. Rajagopal, and F. P. J. de Barros. Optimal allocation of computational resources in hydrogeological models under uncertainty. *Advances in Water Resources*, 83:299–309, 2015.
- [21] D. Pasetto, A. Guadagnini, and M. Putti. Pod-based monte carlo approach for the solution of regional scale groundwater flow driven by randomly distributed recharge. *Advances in Water Resources*, 34(11):1450–1463, 2011.
- [22] D. Pasetto, M. Putti, and W. W.-G. Yeh. A reduced-order model for groundwater flow equation with random hydraulic conductivity: Application to monte carlo methods. *Water Resources Research*, 49(6):3215–3228, 2013.
- [23] G. Porta, B. Bijeljic, M. Blunt, and A. Guadagnini. Continuum-scale characterization of solute transport based on pore-scale velocity distributions. *Geophysical Research Letters*, 2015.
- [24] G. Porta, J.-F. Thovert, M. Riva, A. Guadagnini, and P. Adler. Microscale simulation and numerical upscaling of a reactive flow in a plane channel. *Physical Review E*, 86(3):036102, 2012.
- [25] N. Remy, A. Boucher, and J. Wu. *Applied geostatistics with SGeMS: a user’s guide*. Cambridge University Press, 2009.
- [26] Y. Rubin. *Applied stochastic hydrology*. Oxford Univ. Press, New York, 2003.
- [27] M. Sahimi. *Flow and transport in porous media and fractured rock: from classical methods to modern approaches*. John Wiley & Sons, 2011.

- [28] A. J. Siade, M. Putti, and W. W.-G. Yeh. Snapshot selection for ground-water model reduction using proper orthogonal decomposition. *Water Resources Research*, 46(8), 2010.
- [29] M. Siena, A. Guadagnini, M. Riva, B. Bijeljic, J. P. Nunes, and M. Blunt. Statistical scaling of pore-scale lagrangian velocities in natural porous media. *Physical Review E*, 90(2):023013, 2014.
- [30] J. van Doren, R. Markovinovic, and J. D. Jansen. Reduced-order optimal control of waterflooding using pod. In *9th European Conference on the Mathematics of Oil Recovery*, 2004.
- [31] J. F. van Doren, R. Markovinović, and J.-D. Jansen. Reduced-order optimal control of water flooding using proper orthogonal decomposition. *Computational Geosciences*, 10(1):137–158, 2006.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

- 07/2016** Pacciarini, P.; Gervasio, P.; Quarteroni, A.
Spectral Based Discontinuous Galerkin Reduced Basis Element Method for Parametrized Stokes Problems
- 08/2016** Dassi, F.; Perotto, S.; Si, H.; Streckenbach, T.
A priori anisotropic mesh adaptation driven by a higher dimensional embedding
- 05/2016** Alfio Quarteroni, A.; Lassila, T.; Rossi, S.; Ruiz-Baier, R.
Integrated Heart - Coupling multiscale and multiphysics models for the simulation of the cardiac function
- 06/2016** Micheletti, S.; Perotto, S.; Signorini, M.
Anisotropic mesh adaptation for the generalized Ambrosio-Tortorelli functional with application to brittle fracture
- 02/2016** Crivellaro, A.; Perotto, S.; Zonca, S.
Reconstruction of 3D scattered data via radial basis functions by efficient and robust techniques
- 01/2016** Domanin, M.; Buora, A.; Scardulla, F.; Guerciotti, B.; Forzenigo, L.; Biondetti, P.; Vergara, C.
Computational fluid-dynamic analysis of carotid bifurcations after endarterectomy: closure with patch graft versus direct suture
- 03/2016** Tarabelloni, N.; Ieva, F.
On Data Robustification in Functional Data Analysis
- 04/2016** Pettinati, V.; Ambrosi, D; Ciarletta P.; Pezzuto S.
*Finite element simulations of the active stress in the imaginal disc of the *Drosophila Melanogaster**
- 62/2015** Signorini, M.; Zlotnik, S.; Díez, P.
Proper Generalized Decomposition solution of the parameterized Helmholtz problem: application to inverse geophysical problems.
- 63/2015** Lancellotti, R.M.; Vergara, C.; Valdetaro, L.; Bose, S.; Quarteroni, A.
Large Eddy Simulations for blood fluid-dynamics in real stenotic carotids