

# A Reduced Order Model for Porous Media Flows which Couples Linear Regression with Principal Component Analysis (PCA) using SVD or HOSVD: Upscaling the Permeability Field for Multiphase Reservoir Flow Simulation

B. Lashore<sup>a,\*</sup>, K. Christou<sup>a</sup>, J.L.M.A. Gomes<sup>a</sup>

<sup>a</sup>*Mechanics of Fluids, Soils & Structures Research Group  
School of Engineering, University of Aberdeen, UK*

---

## Abstract

Representation of heterogeneous properties (*i.e.*, permeability in this case) within a large system, is a challenge that cuts across many industries that deal with flows in porous media. This is because it is impossible to obtain the value of permeability at every point in realistic domains. And, even if this was possible, it would be extremely challenging to get the computing resources to make use of all available data values. This challenge gave birth to upscaling, a model order reduction method.

In this paper, a single realization of a permeability field was upscaled by first projecting the permeability field into a principal component analysis space (using SVD for 2D field and HOSVD for 3D fields). Then linear regression was employed within each of the principal component spaces to obtain approximate representations of the initial principal component. The new principal components are then recombined to form a new permeability field with a reduced order. This PCA-linear regression method was compared to traditional industry-standard upscaling techniques (*e.g.*, arithmetic and harmonic averaging) and to a stochastic-based (probability density function, PDF) method of upscaling, to highlight its benefits/performance.

It was shown that the new upscaling technique retained the heterogeneous

---

\*Corresponding author.

Email address: lashorebabatunde@yahoo.com (B. Lashore)

nature of the BaseCase (*i.e.*, high-resolution configuration). Additionally, the upscaling technique did not require *a priori* understanding of similar permeability blocks within the existing domain.

*Keywords:* Principal Component Analysis (PCA) and Linear Regression, Singular Value Decomposition (SVD) and Higher Order Singular Value Decomposition (HOSVD), Upscaling, Reduced Order Models (ROM), Permeability field, Heterogeneity

---

## 1. Introduction

Naturally occurring rock formations are inherently heterogeneous, which means that rock properties such as permeability, pore spaces (*i.e.*, porosity) and pore throat vary spatially at different length scales. Heterogeneity of rock formations is relevant in calculations of transport and storage of fluids in porous media (*i.e.*, multiphase flow in subsurface rocks). Therefore, the representation of such heterogeneous properties is of great importance to oil and gas (*i.e.*, reservoir management), carbon capture and storage (CCS, *i.e.*, CO<sub>2</sub> transportation and storage in subsurface rocks) and waste management (*i.e.*, remediation of contaminated soil).

In essence, rock heterogeneity makes it impossible to obtain reliable and accurate spatial information about geological properties throughout the domain (*i.e.*, uncertainty problem). Additionally, due to limitations in computing resources, it is neither efficient nor possible to use exact values of geological properties at all spatial coordinates of the domain to perform fine-grid simulations (*i.e.*, scaling problem) [10, 33, 41]. This is particularly true for large geological domains which span several kilometres. It is possible to overcome these challenges by a family of techniques called *upscaling*. Upscaling techniques solve both uncertainty and scaling problems by replacing discrete geological and fluid properties of detailed high resolution domains with coarse descriptions (low resolution) of these properties [50].

Upscaling techniques which best preserve statistical properties and flow dynamics behaviour of the high resolution domain give the best representation of heterogeneous properties. There are several upscaling techniques which fall under different categories [22, 41, 47]. Traditional upscaling techniques are deterministic in nature, however over the last decade, techniques classed as stochastic have received great attention from academic and industrial porous media communities worldwide [21, 40, 51]. One of the reasons for

this is because stochastic upscaling techniques are robust enough to handle multiphase flow in porous media and they are able to address scaling and uncertainty problems.

*Upscaling* is a term commonly used in the oil and gas sector and is often referred, throughout the applied mathematics and engineering communities, as a form of model order reduction (or reduced order models, MOR/ROM). Depending on the context in which MOR is used, it can simply be referred as dimensionality reduction. The definition of MOR also depends on the context it is being used, but it suffices to state that MOR captures the essential features of a system, structure or parameter [45]. This implies that an original system exists with more detailed information. Figure 1 presents an interpretation of MOR concept in which a graphical representation of a rabbit is reconstructed with a reduced number of facets (rhs of this illustration).

Schilders [45] provided a historical perspective on the development of Model Order Reduction, from the work of Fourier in 1807 which approximated a function with a few trigonometric terms, to the work of Odabasoglu et al. [36] in 1998 which introduced Passive Reduced-order Interconnect Macromodeling Algorithm (PRIMA). Truncated Balanced Realization is another MOR method developed by Moore [34], which introduced the principal component analysis (PCA) [24] and Golub and Reinsch's [19] algorithm for solving singular value decomposition (SVD). Hankel-norm reduction [18] and Proper Orthogonal Decomposition (POD) [46] are also MOR methods discussed in the references published in 1984 and 1987, respectively. The asymptotic waveform evaluation (AWE) [39] was the first MOR method based on Krylov subspace technique, this method was followed by Pade-via-Lanczos method [16] and the PRIMA. In this work, upscaling is performed through linear regression in the PCA space which is obtained with SVD or HOSVD for 2D and 3D cases, respectively.

This work focuses on reducing the order of permeability distribution in porous media flow simulations. Section 2 introduces the mathematical model used to describe the relevant physics for multi-fluid flow through porous media and highlights the relevance of the permeability field. The mathematical model is discretised by a novel high-order accurate control volume finite element method (CVFEM) [20] which couples both the fine resolution and upscaled representation of the permeability field. Pre-processing statistical properties and post-simulation multiphase flow behaviour are investigated to assess the various upscaling techniques used. Four upscaling techniques

are investigated in this work and they are named Arithmetic mean, Harmonic mean, probability density function (PDF) and PCA (using SVD for 2D models and HOSVD for 3D model) coupled with linear regression.

In Section 3, fundamental mathematics of SVD and associated properties are presented. Then, the concept of spaces is explained to give the reader the understanding necessary to appreciate the transformations and projections involved in this work. Finally, Section 3 describes the reduced order model reduction for the permeability field within the principal component space. Section 4 is dedicated to model description and simulation. It starts by describing the high resolution (*i.e.*, base case) on which the four upscaling techniques are applied to. A brief summary of the pre-processing step for each of the models is also presented. Section 5 provides the results and further discussion on the results of the simulations, it also gives further interpretation to the result obtained from the SVD upscaling technique. Finally, Conclusions are drawn in Section 6.

### *1.1. Brief Overview of Upscaling*

Upscaling methods replace discrete geological and fluid properties of detailed high resolution domains with coarse descriptions (low resolution) of these properties [50]. This subject of upscaling heterogenous properties has interested industries (*i.e.*, oil and gas, CCS and waste management) as well as the academic community for over 7 decades.

Those interested in large-scale reservoirs or ground water generally take a two-stage geostatistical approach. In the first stage, information from seismic data, well data and analogous outcrops is used to model the large scale heterogeneities associated with facies. In the second stage, statistical models such as continuous multi-variate Gaussian field is used to model the rock properties of the facies [15]. In addition to seismic and well data used in the previous stage, core data can be used as a sources of information for the statistical properties used to determine the random field in this stage.

Several studies were carried out on the upscaling of permeability fields [11, 15, 22, 25, 29, 50, 52, 54]. Cardwell and Parsons [9] investigated arithmetic and weighted averages as upscaling techniques for calculating a single equivalent permeability for oil reservoirs with varying permeabilities. The general conclusion reached from this study was that the equivalent permeability lies between the harmonic and the arithmetic average.

Yeo and Zimmerman [54] discussed the accuracy of renormalization method for upscaling 2D hydraulic conductivity fields for two cases. Both cases em-

played the 4-quadrant model (see Section 4.1 for details) set-up. The first case had low permeability in all the blocks except the bottom right block which had high permeability, while the second case employed the checkboard style described in Section 4.1. Different conductivity ratios were also investigated for the two cases. The conclusion drawn was that renormalization worked fairly well for the first case, but very poorly for the second case.

Renard and de Marsily [41] provided a comprehensive review of the various techniques for upscaling. They found out that the techniques classed broadly under deterministic, stochastic and heuristic complemented each other rather than antagonising each other. Renard and de Marsily [41] then recommended that it is better to select a technique which provides block permeabilities instead of one which provides uniform effective permeability.

Christie and Blunt [11] provided 2 sets of problems to compare the upscaling and upgridding techniques of different simulators. The paper does not describe in any detail the upscaling techniques used by the different simulators, but it provides a complete dataset for running simulations.

### *1.2. PCA, SVD, POD, KLT and Reservoir Characterisation*

PCA, SVD and POD are closely related. In the petroleum engineering community where these techniques are used for reservoir characterization, they are sometimes used interchangeably, and are normally referred to as the Karhunen-Loeve Transformation (KLT) [28].

A lot of the recent research in the field of petroleum engineering field which use these techniques, focus on the problem of history matching [2, 26, 28, 48, 53]. History matching is the process of building or adjusting one or more sets of models for a reservoir, such that the model(s) closely reproduces the past behaviour (flow rate, production volume or pressures) of a reservoir, and hopefully predict its future behaviour as well [42]. It is important to note that the model(s) referred to here could be a numerical model representing the system of parameters (*i.e.*, variables) which characterise the reservoir, or simply a model for parameterizing one of more components (*i.e.*, variables) of the system.

Jafarpour and McLaughlin [28] compares KLT to discrete cosine transform (DCT) and goes on to state that the DCT is computationally more efficient than the KLT while being almost as accurate. Tavakoli and Reynolds [48] noted that history matching calculation of sensitivities for all production data with respect to system parameters (*i.e.*, variables) is not feasible. Hence, they (Tavakoli and Reynolds [48]) explored a new parameterization

for reducing the number of variables based on the principal right singular vectors of the dimensionless sensitivity matrix.

Perhaps, the works of Afra and Gildin [2], Afra and Gildin [3], Afra et al. [4], Insuasty et al. [26] are more closely related to the one performed here. However, similar to the work of Jafarpour and McLaughlin [28], they are more related to comparing several realizations than reducing the dimensionality of the dataset for a single realization. The work of Xiao et al. [53] is also closely related to that performed here and interestingly, they used the same CVFEM simulator used in this work. However, like the work of Tavakoli and Reynolds [48], the Xiao et al. were more interested in reducing the number of variable used to describe the system.

Essentially, the recent developments in the field of petroleum engineering described here make use of SVD to select the most suitable realization from multiple realizations or to reduce the representative system variables. However, the work described here is concerned with reducing dimensionality of the dataset for a single realization of one of the system variable (*i.e.*, permeability).

## 2. Multi-Fluid Flow Model

In this work, a control volume finite element method (CVFEM) formulation is used to discretise and solve the set of multi-fluid flow equations. Continuity equations are embedded into the pressure equation to enforce mass conservation and the exact force balance (extended Darcy equations). The numerical formulation employs an implicit algorithm with respect to time which is less restrictive than the traditional industry-standard implicit-pressure-explicit-saturation (IMPES) scheme [6, 17].

The formulation used in this work uses a dual consistent CVFEM representation which is embedded in novel families of triangular and tetrahedral finite element-pairs,  $P_n$ DG- $P_m$  and  $P_n$ DG- $P_m$ DG. In this element-pairs velocity is represented by  $n^{\text{th}}$ -order polynomials that are discontinuous across elements, while pressure is represented by  $m^{\text{th}}$ -order polynomials that may be either continuous ( $P_n$ DG- $P_m$ ) or discontinuous ( $P_n$ DG- $P_m$ DG) across elements. Most properties of  $P_n$ DG- $P_m$ DG element-pairs are similar to those in the  $P_n$ DG- $P_m$  element-pairs but allows a representation in which pressure, saturation and other solution variables (*e.g.*, temperature, concentrations etc) are discontinuous across finite element boundaries [1, 43].

Immiscible multi-fluid flows in porous media may be represented by the extended Darcy equation,

$$\mathbf{q}_\alpha = \mathbf{u}_\alpha S_\alpha = -\frac{\mathcal{K}_{r_\alpha} \mathbf{K}}{\mu_\alpha} (\nabla p_\alpha - \mathbf{s}_{u_\alpha}), \quad (1)$$

where  $\alpha$  designates a phase,  $\mathbf{q}_\alpha$  is the  $\alpha$ -th phase Darcy velocity and  $\mathbf{u}_\alpha$  is the saturation-weighted Darcy velocity.  $\mu_\alpha$ ,  $p_\alpha$ ,  $\rho_\alpha$ , and  $\mathbf{s}_{u_\alpha}$  are the phase dynamic viscosity, pressure, density and source term, which may include gravity and/or capillarity, respectively.  $\mathcal{K}_{r_\alpha}(S_\alpha)$  is the phase relative permeability, which is a function of the phase saturation  $S_\alpha(\mathbf{r}, t)$ , which in turn is a function of the vector position  $\mathbf{r}$ , and time,  $t$ .

In addition to the extended Darcy equation (Eqn. 1), saturation conservative equations,

$$\phi \frac{\partial S_\alpha}{\partial t} + \nabla \cdot (\mathbf{u}_\alpha S_\alpha) = s_{cty, \alpha}, \quad (2)$$

are discretised in space with CV basis functions, and in time with the  $\theta$ -method [37]. In Eqn. 2,  $\phi$  is the porosity and  $s_{cty}$  is a source term.

The discretised saturation and Darcy equations are solved using a multigrid-like approach as described by Pavlidis et al. [38]. The numerical formulation is fully described by Gomes et al. [20] (see also [1, 44]). These numerical methods are embedded in the next-generation flow simulator Fluidity/IC-FERST model software<sup>1</sup> (a full description of the model can be found in [20, 27]). Although, the numerical formulation used here is relatively new, it is not the focus of this work. The result and discussion section (Section 5) refers to Dawe and Grattoni [14] for validation of the numerical formulation with respect to the 2D simulation. However, more extensive validation tests can be found in the aforementioned references [1, 20, 27, 38, 44].

Finally,  $\mathbf{K}(\mathbf{r})$ , the focus of this work, is the absolute permeability tensor of the porous medium and it is a function of the vector position  $\mathbf{r}$ .  $\mathbf{K}$  is prescribed in the system of governing conservative equations, *i.e.*, it is an independent input variable. Since  $\mathbf{K}$  is a petrophysical rock property that varies spatially, it cannot be determined at every spatial coordinate. And, even if this information is available at every spatial coordinate, it would require immense computer resources to process during flow simulations. This is the

---

<sup>1</sup><http://multifluids.github.io>

reason upscaling is required for processing permeability variables. Upscaling can be performed from the continuum scale, across the micro scale ( $mm$ ), local scale and meso scale ( $m$ ) to the field scale ( $km$ )[23]. The upscaling discussed here is from the micro scale also known as the Darcy scale.

### 3. PCA coupled with Linear Regression

#### 3.1. Principal Component Spaces and Model Order Reduction for the permeability field

##### 3.1.1. Understanding the permeability space

Permeability can be represented as tensor quantity, but for the purpose of the following explanation (and the remaining of this paper), it is assumed that permeability is a scalar quantity. Furthermore, the simulator prescribes the permeability field by assigning permeability values to the centre of element-pairs [12] which discretises the simulation domain.

Permeability values assigned in the centre of the element/cell represents values of permeability for the associated element (*i.e.*,  $P_0$ DG element-pair). For simplicity, in the remaining part of this sub-section, a 2D geometry is assumed where the geometry is discretized by a square grid with the value for each square element assigned in the centre of the element. This explanation allow one to easily visualize this permeability space because it is defined by the cartesian coordinates system.

The ROM method discussed here is concerned with reducing the volume of permeability data within the permeability space. Consider for instance a small square which is discretized to contain four squares such that 4 data points are defined in the middle of the four squares. The permeability space of 4 can be reduced to a smaller value of 1 through an averaging process such as arithmetic, harmonic or geometric mean. This type of reduction is the focus of this paper, albeit the main focus of this paper is achieving this reduction through linear regression in the principal component spaces.

##### 3.1.2. Understanding the principal component spaces

Essentially, the existing permeability space is transformed with SVD for 2D permeability field and HOSVD for 3D permeability fields, to principal component spaces (PCS). Within the PCS, traditional PCA dimensionality reduction can be optionally performed (*i.e.*, feature selection or elimination, see Section 3.1.3 for more details), before interpolation is used to implement further data reduction in the principal component spaces. The reduced data



in the PCS is then transformed back into the permeability space to obtain a reduced permeability space. An algorithm which impliments the procedure described here, is presented in Section 4.1.3

The PCS is a bit challenging to visualize because it represents projections which are embedded in the SVD/HOSVD factors of the original permeability field. An aid for visualizing the PCS for a 2D permeability field is to write out the  $\mathbf{U}$ ,  $\mathbf{\Sigma}$  and  $\mathbf{V}^\top$  matrices (or SVD factors) and consider the singular vectors (*i.e.*, columns in  $\mathbf{U}$  and rows in  $\mathbf{V}^\top$ ) as separate PCS. A similar intuition can be extended to HOSVD for 3D permeability fields. The decomposition of a matrix into its factors using SVD [13, 49] or HOSVD [30, 32] is covered in details in several published materials. However, for completeness, the appendix contains a brief introduction to SVD (Appendix A) and HOSVD (Appendix B).

### 3.1.3. Reduced Order Model/Principal Component Space Reduction

For the PCA-Linear regression (PCA-LinReg) reduced order model of the permeability field, all the work is done in the principal component spaces. This can be completed in two steps, which is exemplified for the 2D model thus:

1. The first step is the principal component analysis (PCA) [24, 49] using the SVD method, where the singular values are examined and, based on a pre-defined criteria,  $z$  singular values are selected, starting from  $\sigma_1$  and in a decreasing order, where  $z \leq r$ .  
Note that, in the case with  $z = r$ , it can be assumed that step 1 has been ignored. However, when  $z < r$ , the first  $z$  columns of  $\mathbf{U}$  and the first  $z$  rows of  $\mathbf{V}^\top$  are extracted from  $\mathbf{U}$  and  $\mathbf{V}^\top$  to form new matrices  $\mathbf{U}_{\text{new}}$  and  $\mathbf{V}_{\text{new}}^\top$  respectively. The selected  $z$  singular values also form a new diagonal matrix,  $\mathbf{\Sigma}_{\text{new}}$ .
2. The second step involves interpolation within each singular vector (column in  $\mathbf{U}$  and row in  $\mathbf{V}^\top$ ) to reduce the data points in each column of  $\mathbf{U}$  and in  $\mathbf{V}^\top$ . From this step,  $\mathbf{U}_{\text{new2}}$  and  $\mathbf{V}_{\text{new2}}^\top$  are obtained. The reduced permeability field is calculated by multiplying  $\mathbf{U}_{\text{new2}}$ ,  $\mathbf{\Sigma}_{\text{new}}$  and  $\mathbf{V}_{\text{new2}}^\top$ .

It is trivial to show that these two steps can be replicated for a 3-D model using HOSVD to obtain the PCA space.

Finally on the PCS, it is worth noting that two methods exist for obtaining PCA space[49] , namely, the co-variance method and the SVD method. Tharwat [49] goes on to describe both methods in details. In this work, the authors have opted to use the SVD method to obtain the PCA space. This is because the singular vectors which make up the PCS have a strong correlation to the coordinate axis they represent. This was exploited in making the assumption that a linear relationship exist for the data in each of the PC spaces, hence the linear regression performed to reduce the data in the PCS.

### 3.2. Principal Component Analysis Versus Linear Regression

Ng [35] was quoted to have said “PCA is not linear regression, and despite some cosmetic similarity, these are actually totally different algorithms”. The main differences between these two mathematical methods is summarised in the Table 1 below and can also be deduced from the figure 2.

	PCA	Linear Regression
1	Determines the best vector (or set of vectors) which compresses the data points by minimizing the orthogonal distance to the points	Approximates the best fit line for the points by minimizing the error
2	Essentially, it forms a projection of the original points onto the new vector space	Predicts $y$ from $x$ through a function $f(x_1, x_2, \dots, x_n) = y$
3	Not used for predictions, as original feature space ( <i>i.e.</i> , $x_1$ and $x_2$ ) give way to a new reduced dimension. Coordinates $x_1, x_2, \dots, x_n$ give way to $z_1, \dots, z_k$ where $k \leq n$	

Table 1: Table comparing PCA with Linear Regression

## 4. Model Description and Simulation Setup

As previously mentioned, four upscaling techniques are investigated in this work. Arithmetic and harmonic means are used to obtain the first two up-scaled representation and these two are classed as deterministic techniques.

The third is a randomly generated permeability field with Gaussian distribution, prescribed by a probability density function (PDF) which was obtained from a domain discretised with high-resolution mesh with known permeability distribution (*i.e.*, base case). The fourth upscaling technique is the crux of this research, it introduces a reduced order model which couples PCA and linear regression.

#### 4.1. Model Description

The BaseCase (*i.e.*, fine mesh resolution) model for the permeability field builds on the  $2 \times 2$  block, 4 quadrant model used initially by Cardwell and Parsons [9] and later by Yeo and Zimmerman [54] and Dawe and Grattoni [14]. These authors designed the permeability field for the quadrants in a checkerboard style which means that diagonally opposite blocks retained the same permeability values with one set of diagonally opposite blocks having a higher permeability value compared to the other set. Hence, one set of diagonally opposite blocks are referred to as high permeability blocks while the other set is referred to as low permeability blocks.

The BaseCase permeability model used here retains the checkerboard designed used by the recently named authors. Albeit, the permeability dataset for each block in the set of diagonally opposite blocks have similar value (*i.e.*, not exactly the same as in the papers referred to), although one set can still be referred to as a set of high permeability blocks while the other set is referred to as a set of low permeability blocks. Furthermore, within each block, the permeability field varies within a pre-defined range and consists of 1600 permeability values. More details on the design of the permeability field for the BaseCase is available in section 4.1.1. Figures 4a and 4b give a pictorial representation of the permeability field for the base case with the mesh discretization visible and invisible, respectively. Using the labelling scheme used by Dawe and Grattoni [14], the quadrant model block  $K1$ ,  $K2$ ,  $K3$  and  $K4$  are top left, bottom left, top right and bottom right, respectively.

As already mentioned, four upscaled permeability field models were obtained from the high-resolution (*i.e.*, the BaseCase) permeability field. When compared to the BaseCase model, all the upscaled models were upgridded to reduce the resolution of the permeability field (Fig. 3). The first two upscaled models are arithmetic and harmonic averaging methods (referred to as ArithMean and HarmMean, respectively).

In both sets of test-cases, averages are calculated for each block using the 1600 data points from the blocks of the BaseCase to obtain a permeability

value for each block in the upscaled models. This effectively homogenizes the permeability field in each block. The upscaled permeability field for the third and fourth cases are referred to as PDFCase and PCA-LinReg Case, respectively. And, as can be deduced from the names, these cases are based on probability density function (PDF) and SVD, respectively. The description for these permeability field models are presented in Sections 4.1.2 and 4.1.3 respectively. It is appropriate to note that although these models (PDFCase and PCA-LinReg Case) are upscaled, they retain heterogeneity within the blocks.

#### 4.1.1. BaseCase Pre-processing Algorithm

1. A function is used to randomly generate uniform (*i.e.*, Gaussian) dataset for each block. The range for each block is as below:
  - a) **K1** : 700 – 900 mD
  - b) **K2** : 50 – 150 mD
  - c) **K3** : 100 – 200 mD
  - d) **K4** : 1000 – 1200 mD
2. The number of data points generated for each block is 1600 and the data values are randomly allocated within each block in a structured pattern. Albeit, the simulation mesh is unstructured.
3. Merge the dataset from different blocks into a single file retaining the original “structured pattern” (in step 1 above) for the randomly positioned dataset for each block. This file will be used by the CVFEM-based flow simulator to prescribe the permeability field for the base case.

#### 4.1.2. PDFCase Pre-processing Algorithm

1. The datasets from step one of the pre-processing algorithm for the base case (Section 4.1.1) are used to calculate mean and standard deviation for each block of the permeability field.
2. A function which generates a Gaussian dataset using arithmetic mean and standard deviation is used to stochastically produce a quarter of the data points from the BaseCase for each of the four blocks (*i.e.*, 400 data points for each block).
3. The data values in the dataset for each block is randomly positioned within each block in a structured pattern. As in the base case, the simulation grid is unstructured.
4. Finally, the structured/arranged dataset from each block is combined into a single file and then used by the flow simulator, which merge the permeability mapping to the mesh with the appropriate element-pairs.

#### 4.1.3. 2-D PCA (i.e., using SVD) & LinReg Case Pre-processing Algorithm

1. The permeability field obtained in step 3 of the pre-processing algorithm for the base case is factorized using SVD to obtain  $\mathbf{U}$ ,  $\Sigma$ , and  $\mathbf{V}$ .
2. A selected number ( $z$ ) of singular values are retained in decreasing order of magnitude to form a new diagonal matrix,  $\Sigma_{new}$ . (This step is optional, it is the conventional PCA dimensionality reduction)
3. The first  $z$  columns of  $\mathbf{U}$  and  $\mathbf{V}$  are selected to form new truncated matrices call  $\mathbf{U}_{new}$  and  $\mathbf{V}_{new}$  respectively. (This step is necessary only if the optional step in 2 above is executed)
4. Within each column of  $\mathbf{U}_{new}$  and  $\mathbf{V}_{new}$ , a linear interpolation is performed to form a new column with the number of data points in each column reduced to half of the original column. These new columns are recombined to form  $\mathbf{U}_{new2}$  and  $\mathbf{V}_{new2}$ .
5. The permeability field used for the SVD case is the matrix  $\mathbf{A}_{new2}$ , which is defined as the product of  $\mathbf{U}_{new2}\Sigma_{new}\mathbf{V}_{new2}^T$ . It is a permeability field with a quarter of the data values in the base case and therefore a permeability field with reduced order.

#### 4.1.4. 3-D PCA (i.e., using HOSVD) & LinReg Case Pre-processing Algorithm

The algorithm for the 3-D case is similar to that of the 2-D case described above. However:

1. In step 1 above, HOSVD is used for factorization instead of SVD. Therefore, a  $3 \times 3$  core tensor is obtained in place of the  $\Sigma$  term, while 3 factor matrices are obtained in place of  $\mathbf{U}$  and  $\mathbf{V}$ .
2. It goes without saying that the core tensor is treated as the  $\Sigma$  term, while the 3 factor matrices are treated similarly to  $\mathbf{U}$  and  $\mathbf{V}$  as described in Section 4.1.3 above.

The authors found the updated Tensorly library package [31] helpful in implementing and executing the algorithm for HOSVD.

#### 4.2. Simulations

For the simulations, each of the model's domain is initially fully saturated with a fluid (Phase 2), and a wetting phase (1) fluid is driven into the domain from the left hand side at a constant mass flow rate. No-flux boundary conditions were imposed on the upper and lower borders of each domain,

while mixed fluids are recovered from the right-hand face of the domain. Relative permeability,  $\mathcal{K}_{r\alpha}$  (Eqn. 1), which is often expressed as a function of local, residual and maximum phase saturations prescribed in the pore rock matrix. In the simulations conducted for this work, the modified Brooks and Corey [7] model was used [5],

$$\mathcal{K}_{rw}(S_w) = \mathcal{K}_{rw}^\circ \left[ \frac{S_w - S_{w,irr}}{1 - S_{w,irr} - S_{nw,r}} \right]^{n_w}, \quad (3)$$

$$\mathcal{K}_{rnw}(S_{nw}) = \mathcal{K}_{rnw}^\circ \left[ \frac{S_{nw} - S_{nw,r}}{1 - S_{w,irr} - S_{nw,r}} \right]^{n_{nw}}, \quad (4)$$

where subscripts  $w$  and  $nw$  stand for wetting and non-wetting phases, respectively.  $\mathcal{K}_{rw}^\circ$  and  $\mathcal{K}_{rnw}^\circ$  are end-point relative permeability to wetting and non-wetting phases,  $S_{w,irr}$  and  $S_{nw,r}$  are irreducible wetting and residual non-wetting phase saturations, respectively. Exponents  $n_w$  and  $n_{nw}$  are both set to 2.

Figure 4 shows permeability mapping for all cases, while Figs. 5-9 show phase saturation distribution during fluid injection for time,

$$t = 0.15, 0.30, 0.50, 1.15, 1.75 \text{ and } 2.95 \text{ s for all cases.}$$

## 5. Results and Discussion

Mesh resolution for the the BaseCase and upscaled cases are shown in Fig. 3. Domain of the BaseCase (*i.e.*, high-resolution mesh) was discretised with 4112 P<sub>1</sub>DG-P<sub>1</sub> triangular element-pairs, whilst in upscaled cases the domain was discretised with 728 P<sub>1</sub>DG-P<sub>1</sub> triangular element-pairs. Therefore the resolution for the upscaled cases is reduced by a factor of approximately 5.6.

Figure 4 compares the permeability distribution for all the models. As indicated in the figure's caption, the permeability legend in Fig. 4(a) is representative for all the other figures (*i.e.*, Fig. 4(b) to (f)). Additionally, Fig. 4(b) to (f) show the cases without the element edges which discretize the domain.

This highlights the fact that the permeability distribution of the Arith-Mean and HarmMean cases, Fig. 4(c) and (d), are homogeneous within each block, while PDFCase and PCA-LinReg cases, Fig. 4(e) and (f), are heterogeneous. Given that the BaseCase is also heterogeneous, it can be stated that

the model for the PDFCase and the PCA-LinReg Case give a better representation of the BaseCase. An additional consequence of the homogeneity in the ArithMeanCase and the HarmMeanCase is that the upgrid size has no influence on the outcome of their permeability distribution. Whereas, for the PDFCase and the PCA-LinReg Case, the permeability field will change if the upgrid resolution changes.

With respect to the PDFCase and for this specific resolution, it is important to note that Fig. 4(e) is only one realization of infinitely many possible representations of the PDFCase. Meanwhile, the PCA-LinReg Case will retain this specific (*i.e.*, Fig. 4(e)) permeability field for this specific resolution (*i.e.*, if linear interpolation is used, but this digresses from the main point and will be discussed in future work). Therefore, the permeability value at a specific point in the PDFCase is not directly related to the data value at the nearest point in the BaseCase because the randomly generated PDF dataset is also randomly arranged within each block. However, the data value at a point in the PCA-LinReg Case is directly related to the nearest data values with respect to the same point in the BaseCase, through the interpolation which takes place in the PCS.

Furthermore, prior to the pre-processing required for evaluating the ArithMeanCase, HarmMeanCase and the PDFCase it is necessary to have an understanding of the whole domain and pre-define blocks with similar properties (*i.e.*, blocks  $K1$ ,  $K2$ ,  $K3$  and  $K4$ ) in this case. However, with the PCA-LinReg Case, knowledge of blocks with similar properties is not required prior to the upscaling. Albeit, one has to select the upgrid resolution carefully, if the resolution is too large then there is the risk of losing relevant data in the interpolation. The limits of the techniques used for the PCA-LinReg Case will be discussed in the author's future work.

Before going on to review the results of the simulations, it is beneficial to briefly consider the Buckley-Leverett and cross-flow behaviour which are used to validate the results from this work.

Buckley and Leverett [8] in 1942 derived the Buckley-Leverett equation, an analytical (albeit simplified) two phase flow equation. The equation determined that a plane of constant saturation progresses uniformly through the domain during the flow simulation. The equation is referred to as "simplified" because of the assumptions made while deriving it, some of which were ignored in this work. Principally, this work ignores the fact that the equation deals with injection into an homogeneous permeability field while this work deals injection into two blocks ( $K1$  and  $K2$ ) with different perme-

abilities. Additionally, the permeability field in each block is heterogeneous. However, the general validation test or expectation from Buckley-Leverett's influence on this work, is that, the injected fluid saturation in  $K1$  and  $K2$  is expected to have a piston-shape front within those blocks, subject some inconsistency at the boundary between the two blocks, and also within the blocks due to the heterogeneity.

Dawe and Grattoni [14] conducted experiments to investigate cross-flow, and as previously mentioned, the permeability field for the experimental work were somewhat replicated here for the simulations. Therefore, the work of Dawe and Grattoni [14] can be used to validate the resulting simulations from this work. In particular, it will be noted that the simulations matches the experiment at the boundary between  $K1$  and  $K2$ . Additionally, the cross-flow from  $K1$  to  $K4$  is well represented in the simulations.

Figures 5-9 show snapshots of phase 1 (*i.e.*, injected or displacing fluid) saturation field for all studied cases. In Figs. 5-9(a) and (b), even though fluid is injected at the same rate across the left face it travels faster in  $K1$  compared to  $K2$ . Figures 5-9 (c) show that fluid experience a resistance to flow at the interface between  $K1$  and  $K3$  which encourages the flow into  $K4$  (Figs. 5-9 (d) and (e)). Figures 5-9 (f) indicates preferential flow in  $K4$  compared to  $K3$ .

Figure 10 compares phase 1 saturation distribution for all the cases at  $t = 1.15s$ . Generally, saturation distribution for all cases closely resemble each other with arithmetic and harmonic mean (Figs. 10(b) and (c)) being the most identical. Albeit, closer inspection will reveal that there is more displacement in Fig. 10(b) compared to (c). This is logical giving that the block mean values calculated for arithmetic mean is higher than that for harmonic mean. However, the important feature of Figs. 10(b) and (c) is that they poorly represent the BaseCase (Fig. 10(a)) at the top left hand corner of block  $K2$ .

Figure 10(d) also shows an inconsistency with the BaseCase (Fig. 10(a)) where it shows phase 1 moving from block  $K2$  into  $K4$ . This inconsistency is with respect to this realization of the PDFCase. Hence, PCA-LinReg Case in Fig. 10(e) give the best representation of the BaseCase in *i.e.*, Fig. 10(a)

## 6. Conclusion

Upscaling is a form of reduced order model, and a method of implementing upscaling by coupling linear regression with PCA is introduced in this work.



For a 2-D model, SVD is used to decompose the existing permeability field to obtain the PCA space, while HOSVD is used for 3-D models. Within, the PCA space, linear regression (or interpolation) is used to reduce the data in the PCA space. The reduced data in the principal component space is then recombined to form a reduced data of the permeability field.

Results show that the permeability field data reduction can be specifically defined for the PDFCase and the PCA-LinReg Case but not for the ArithMeanCase and the HarmMeanCase. This is because the latter cases will produce the same permeability field irrespective of the upgrid resolution. The PCA-LinReg Case is the only technique that does not require a prior knowledge of blocks with similar permeability values before upscaling. This is a huge advantage as it saves time during the data pre-processing stage, because sorting and calculations for each distinct region is avoided.

The author's future work will provide a quantitative analysis of the results obtained from the coupled PCA-Linear regression method. Additionally, future work will present other methods which follow a similar principle but also seek to optimize the quantitative results obtained.

## References

- [1] Adam, A., Pavlidis, D., Percival, J., Salinas, P., Xie, Z., Fang, F., Mugeridge, A., Jackson, M., 2016. Higher-order conservative interpolation between control-volume meshes: Application to advection and multi-phase flow problems with dynamic mesh adaptivity. *Journal of Computational Physics* 321, 512–531.
- [2] Afra, S., Gildin, E., 12 2013. Permeability parametrization using higher order singular value decomposition (hosvd). *Proceedings - 2013 12th International Conference on Machine Learning and Applications, ICMLA 2013* 2, 188–193.
- [3] Afra, S., Gildin, E., 06 2016. Tensor based geology preserving reservoir parameterization with higher order singular value decomposition (hosvd). *Computers & Geosciences* 94, 110–120.
- [4] Afra, S., Gildin, E., Tarrahi, M., 06 2014. Heterogeneous reservoir characterization using efficient parameterization through higher order svd (hosvd). *Proceedings of the American Control Conference*, 147–152.
- [5] Alpak, F., Lake, L., Embid, S., 1999. Validation of a modified carman-kozeny equation to model two-phase relative permeabilities. In: *SPE Annual Technical Conference and Exhibition*. SPE, Texas, SPE56479.
- [6] Aziz, K., Settari, A., 1986. *Fundamentals of reservoir simulation*. Elsevier Applied Science Publishers, New York.
- [7] Brooks, R., Corey, A., 1964. *Hydrology Papers*. Vol. 3. Colorado State University Press, Ch. Hydraulic properties of porous media.
- [8] Buckley, S., Leverett, M., 1942. Mechanism of fluid displacement in sands. *Transactions of the American Institute of Mining* 146, 107–116.
- [9] Cardwell, W.T., J., Parsons, R., Dec 1945. Average permeabilities of heterogeneous oil sands. *Society of Petroleum Engineers SPE-945034-G* (160(1)), 34–42.  
URL <https://doi.org/10.2118/945034-G>
- [10] Chen, Z., Huan, G., Ma, Y., 2006. *Computational Methods for Multi-phase Flows in Porous Media*. SIAM.

- [11] Christie, M., Blunt, M., 2001. Tenth spe comparative solution project: A comparison of upscaling techniques. Society of Petroleum Engineers SPE-66599-MS.
- [12] Christou, K., Randunz, W., Lashore, B., Oliveira, F., Gomes, J., 2018. Numerical investigation of viscous flow instabilities in multiphase heterogeneous porous media. *Advances in Water Resources* (Accepted for publication).
- [13] Cline, A. K., 2007. Computation of the Singular Value Decomposition. Chapman & Hall/CRC, Ch. 47, pp. 45.1–45.13.
- [14] Dawe, R., Grattoni, C., 2008. Experimental displacement patterns in a 2 x 2 quadrant block with permeability and wettability heterogeneities - problems for numerical modelling. *Transport in Porous Media* 71, 5–22.
- [15] Ewing, R. E., 1997. Aspects of upscaling in simulation of flow in porous media. *Advances in Water Resources* 20 (5), 349 – 358.  
URL <http://www.sciencedirect.com/science/article/pii/S0309170896000528>
- [16] Feldmann, P., Freund, R. W., May 1995. Efficient linear circuit analysis by pade approximation via the lanczos process. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 14 (5), 639–649.
- [17] Geiger, S., Roberts, S., Matthäi, S., Zoppou, C., A. Burri, A., 2004. Combining finite element and finite volume methods for efficient multiphase flow simulations in highly heterogeneous and structurally complex geologic media. *Geofluids* 4 (4), 284–299.
- [18] GLOVER, K., 1984. All optimal hankel-norm approximations of linear multivariable systems and their  $l_1$ -error bounds. *International Journal of Control* 39 (6), 1115–1193.  
URL <https://doi.org/10.1080/00207178408933239>
- [19] Golub, G. H., Reinsch, C., Apr 1970. Singular value decomposition and least squares solutions. *Numerische Mathematik* 14 (5), 403–420.  
URL <https://doi.org/10.1007/BF02163027>

- [20] Gomes, J., Pavlidis, D., Salinas, P., Percival, J., Melnikova, Y., Pain, C., Jackson, M., 2017. A force-balanced control volume finite element method for multi-phase porous media flow modelling. *International Journal For Numerical Methods in Fluids* 83, 431–445.
- [21] Guilleminot, J., Soize, C., Ghanem., R., 2012. Stochastic representation for anisotropic permeability tensor random field. *International Journal for Numerical and Analytical Methods in Geomechanics* 36(13), 1592–1608.
- [22] Hastings, J., Muggeridge, A., 2001. Upscaling uncertainty permeability using small cell renormalization. *Mathematical Geology* 33(4), 491–502.
- [23] Helmig, R., Miller, C., Jakobs, H., Class, H., Hilpert, M., Kees, C., Niessner, J., 2004. Multiphase flow and transport modeling in heterogeneous porous media. In: Bucchianico, A., Mattheij, R., Peletier, M. (Eds.), *Mathematics in Industry 8*. Vol. 8 of *Water Flow*. The European Consortium For Mathematics in Industry, Springer, pp. 449–488.
- [24] Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 24 (6), 417–441.  
URL <https://doi.org/10.1037/h0071325>
- [25] Indelman, P., Dagan, D., 1993. Upscaling of permeability of anisotropic heterogeneous formations - 2 general structure and small perturbation analysis. *Water Resources Research* 29(4), 925–933.
- [26] Insuasty, E., den Hof, P. M. J. V., Weiland, S., Jansen, J. D., 2017. Low-dimensional tensor representations for the estimation of petrophysical reservoir parameters. *Society of Petroleum Engineers (SPE) SPE-182707-MS*.
- [27] Jackson, M., Gomes, J., Mostaghimi, P., Percival, J., Tollit, B., Pavlidis, D., Pain, C., El-Sheikh, A., Muggeridge, A., Blunt, M., 2013. Reservoir modeling for flow simulation using surfaces, adaptive unstructured meshes, and control-volume-finite-element methods. In: *SPE Reservoir Simulation Symposium*.

- [28] Jafarpour, B., McLaughlin, D. B., 2009. Reservoir characterization with the discrete cosine transform. Society of Petroleum Engineers (SPE) SPE-106453-PA.
- [29] King, P. R., Jun 1996. Upscaling permeability: Error analysis for renormalization. *Transport in Porous Media* 23 (3), 337–354.  
URL <https://doi.org/10.1007/BF00167102>
- [30] Kolda, T. G., Bader, B. W., 2009. Tensor decompositions and applications. *SIAM Review* 51, 455–500.
- [31] Kossaifi, J., Panagakis, Y., Anandkumar, A., Pantic, M., 2018. Tensorly: Tensor learning in python. CoRR abs/1610.09555.
- [32] Lathauwer, L. D., Moor, B. D., Vandewalle, J., 2000. A multilinear singular value decomposition. *SIAM J. Matrix Analysis Applications* 21, 1253–1278.
- [33] Miller, C., Christakos, G., Imhoff, P., McBride, J., Pedit, J., 1998. Multiphase flow and transport modeling in heterogeneous porous media: Challenges and approaches. *Advances in Water Resources* 21(2), 77–120.
- [34] Moore, B., February 1981. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Transactions on Automatic Control* 26 (1), 17–32.
- [35] Ng, A., 2018. Machine learning. Online Course video on Coursera.org titled Principal Component Analysis Problem Formulation, last checked: 01.12.2018.  
URL <https://www.coursera.org/learn/machine-learning/lecture/GBFTt/principal-component-analysis-problem-formulation>
- [36] Odabasioglu, A., Celik, M., Pileggi, L. T., Aug 1998. Prima: passive reduced-order interconnect macromodeling algorithm. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 17 (8), 645–654.
- [37] Pavlidis, D., Gomes, J., Xie, Z., Percival, J., Pain, C., Matar, O., 2014. Compressive-advection and multi-component methods for interface-capturing. *Journal of Computational Physics*. In Press.

- [38] Pavlidis, D., Gomes, J., Xie, Z., Percival, J., Pain, C., Matar, O., 2016. Compressive advection and multi-component methods for interface-capturing. *International Journal for Numerical Methods in Fluids* 80 (4), 256–282.
- [39] Pillage, L. T., Rohrer, R. A., Apr 1990. Asymptotic waveform evaluation for timing analysis. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 9 (4), 352–366.
- [40] Ravalec-Dupin, M. L., 2010. Pilot block method methodology to calibrate stochastic permeability fields to dynamic data. *International Association for Mathematical Geosciences* 42, 165–185.
- [41] Renard, P., de Marsily, G., 1997. Calculating equivalent permeability: A review. *Advances in Water Resources* 20(5-6), 253–278.
- [42] Rwechungura, R. W., Dadashpour, M., Kleppe, J., 2011. Advanced history matching techniques reviewed. *Society of Petroleum Engineers (SPE) SPE-142497-MS*.
- [43] Salinas, P., Pavlidis, D., Xie, Z., Osman, H., Pain, C., Jackson, M., 2018. A discontinuous control volume finite element method for multiphase flow in heterogeneous porous media. *Journal of Computational Physics* 352, 602–614.
- [44] Salinas, P., Percival, J., Pavlidis, D., Xie, Z., Gomes, J., Pain, C., Jackson, M., et al., 2015. A discontinuous overlapping control volume finite element method for multi-phase porous media flow using dynamic unstructured mesh optimization. In: *SPE-173279-MS, Reservoir Simulation Symposium*. Society of Petroleum Engineers.
- [45] Schilders, W., 2008. *Introduction to Model Order Reduction*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 3–32.  
URL [https://doi.org/10.1007/978-3-540-78841-6\\_1](https://doi.org/10.1007/978-3-540-78841-6_1)
- [46] Sirovich, L., Oct. 1987. Turbulence and the dynamics of coherent structures. I - Coherent structures. II - Symmetries and transformations. III - Dynamics and scaling. *Quarterly of Applied Mathematics* 45, 561–571.
- [47] Szymkiewicz, A., 2013. *Modelling Water Flow in Unsaturated Porous Media*. Springer.

- [48] Tavakoli, R., Reynolds, A. C., 2010. History matching with parametrization based on the svd of a dimensionless sensitivity matrix. Society of Petroleum Engineers (SPE) SPE-118952-PA.
- [49] Tharwat, A., 01 2016. Principal component analysis - a tutorial. *International Journal of Applied Pattern Recognition* 3, 197.
- [50] Vereecken, H., Kasteel, R., Vanderborght, J., Harter, T., 2007. Upscaling hydraulic properties and soil water flow processes in heterogeneous soils: A review. *Vadose Zone Journal* 6, 1–28.
- [51] Verwoerd, W., 2009. New stochastic model for dispersion in heterogeneous porous media: 1 application to unbounded domains. *Applied Mathematical Modelling* 33, 605–625.
- [52] Wen, X., Gómez-Hernández, J., 1996. Upscaling hydraulic conductivities in heterogeneous media: An overview. *Journal of Hydrology* 183, ix–xxxii.
- [53] Xiao, D., Fang, F., Pain, C., Navon, I., Salinas, P., Wang, Z., 2018. Non-intrusive model reduction for a 3d unstructured mesh control volume finite element reservoir model and its application to fluvial channels. *International Journal of Oil, Gas and Coal Technology* 19 (3), 316–339.
- [54] Yeo, I.-W., Zimmerman, R. W., Oct 2001. Accuracy of the renormalization method for computing effective conductivities of heterogeneous media. *Transport in Porous Media* 45 (1), 129–138.  
URL <https://doi.org/10.1023/A:1011849804979>

## List of Figures

1	Graphical illustration of model order reduction (initially used by Schilders [45], with graphics credited to Harvard University, Microsoft Research.) . . . . .	25
2	Graphical comparison between linear regression and PCA (extracted from Ng [35]). . . . .	25
3	Mesh grid used in the performed numerical simulations: (a) high-resolution and (b) upscaled cases with 4112 and 728 triangular elements, respectively. . . . .	26
4	Permeability field for the BaseCase and upscaled configurations (permeability legend in Fig. 4(a) is representative for all the other figures). . . . .	27
5	Simulations for the BaseCase (Saturation legend in Fig. 5(a) is representative for all the other figures - <i>i.e.</i> , Fig. 5(b) to (f))	28
6	Simulations for the ArithMean Case (Saturation legend in Fig. 6(a) is representative for all the other figures - <i>i.e.</i> , Fig. 6(b) to (f)) . . . . .	29
7	Simulations for the HarmMean Case (Saturation legend in Fig. 7(a) is representative for all the other figures - <i>i.e.</i> , Fig. 7(b) to (f)) . . . . .	30
8	Simulations for the PDFCase (Saturation legend in Fig. 8(a) is representative for all the other figures - <i>i.e.</i> , Fig. 8(b) to (f))	31
9	Simulation for the SVDCase (Saturation legend in Fig. 9(a) is representative for all the other figures - <i>i.e.</i> , Fig. 9(b) to (f)) .	32
10	Comparing phase 1 saturation distribution for all the models at $t = 1.15$ s (Saturation legend in Fig. 10(a) is representative for all the other figures - <i>i.e.</i> , Fig. 10(b) to (e)) . . . . .	33
11	Mesh grid used in the performed numerical 3-D simulations: (a) high-resolution and (b) upscaled cases with 2019 and 303 tetrahedral elements respectively . . . . .	34
12	Permeability field for the 3D base case as well as the HOSVD upscaled cases (Permeability legend in Fig. 12(a) is representative for all the other figures - <i>i.e.</i> , Fig. 12(b) to (d)) . . . . .	35
13	Comparing phase 1 saturation distribution for the 3-D Cases at $t = 0.18, 0.28$ and $0.42$ s (Saturation legend in Fig. 13(a) is representative for all the other figures - <i>i.e.</i> , Fig. 13(b) to (e))	36



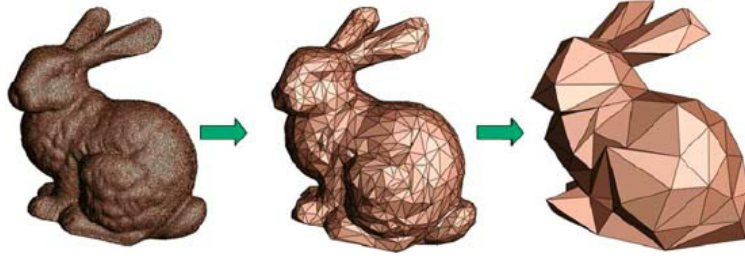


Figure 1: Graphical illustration of model order reduction (initially used by Schilders [45], with graphics credited to Harvard University, Microsoft Research.)

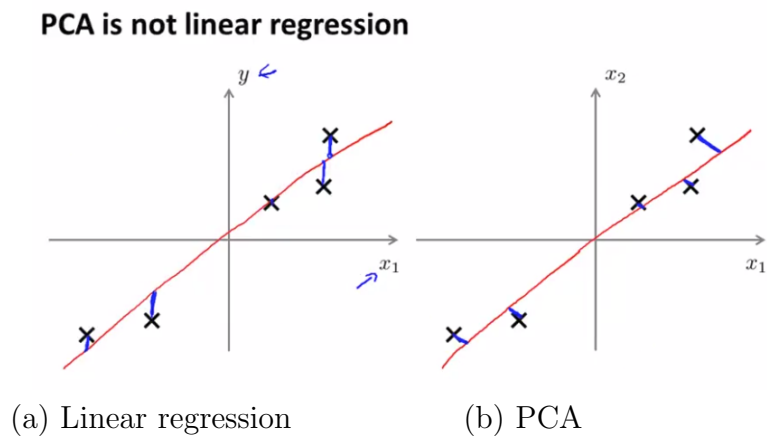
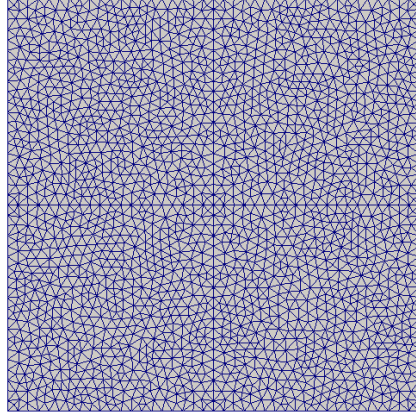
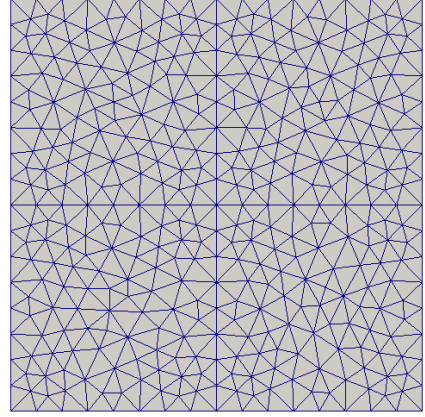


Figure 2: Graphical comparison between linear regression and PCA (extracted from Ng [35]).

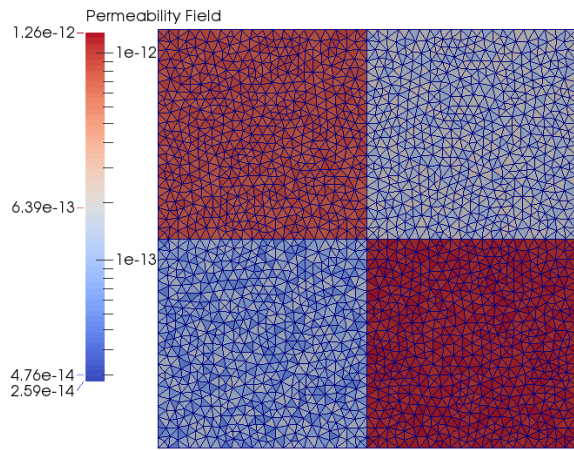


(a) High Resolution Grid (BaseCase)

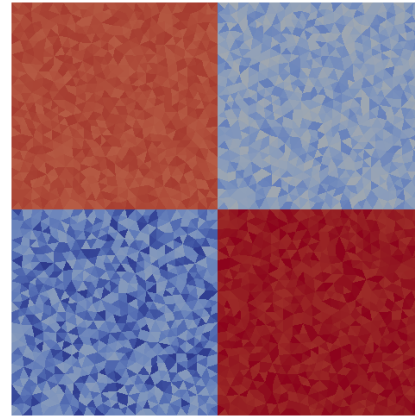


(b) Low Resolution Grid (Upscaled Cases)

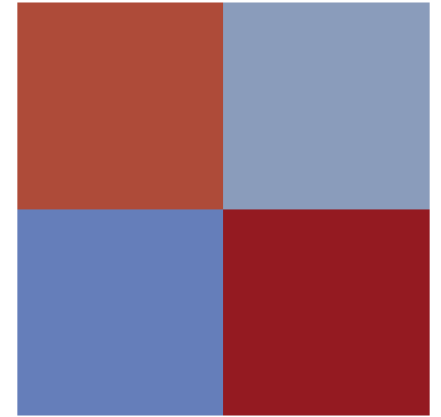
Figure 3: Mesh grid used in the performed numerical simulations: (a) high-resolution and (b) upscaled cases with 4112 and 728 triangular elements, respectively.



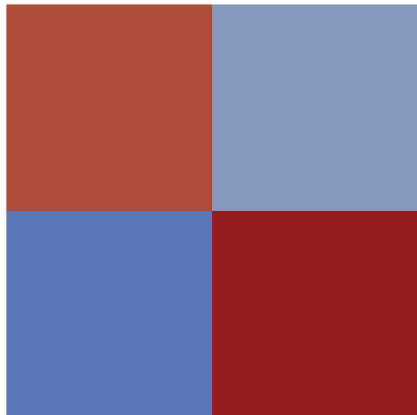
(a) BaseCase (overlapped with the mesh)



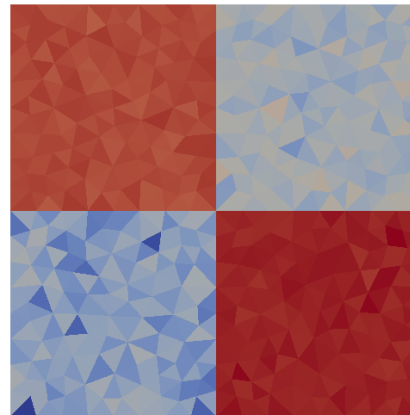
(b) BaseCase (without Mesh)



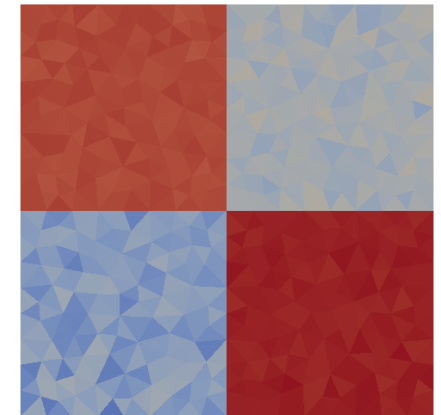
(c) ArithMean Case



(d) HarmMean Case

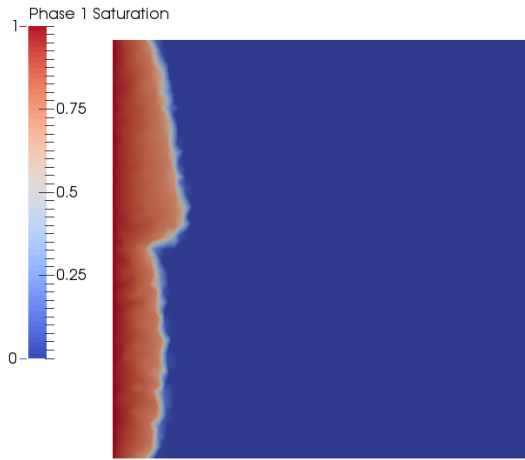


(e) PDFCase



(f) SVDCase

Figure 4: Permeability field for the BaseCase and upscaled configurations (permeability legend in Fig. 4(a) is representative for all the other figures).



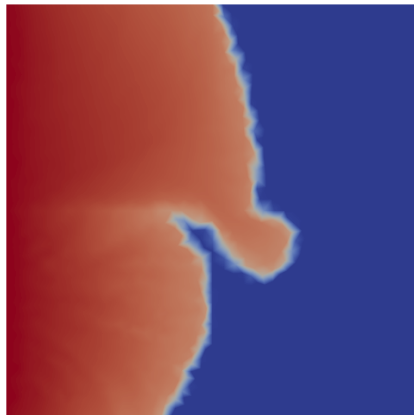
(a) Phase 1 Saturation at  $t=0.15s$



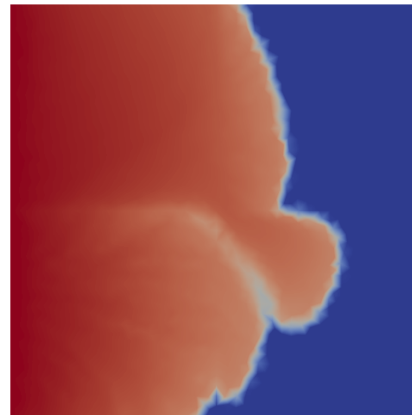
(b)  $t=0.30s$



(c)  $t=0.50s$



(d)  $t=1.15s$

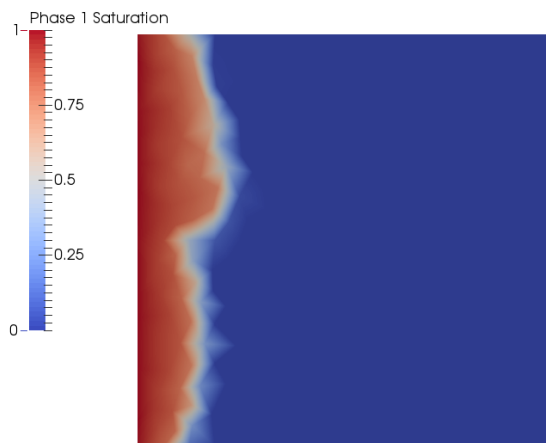


(e)  $t=1.75s$



(f)  $t=2.95s$

Figure 5: Simulations for the BaseCase (Saturation legend in Fig. 5(a) is representative for all the other figures - *i.e.*, Fig. 5(b) to (f))



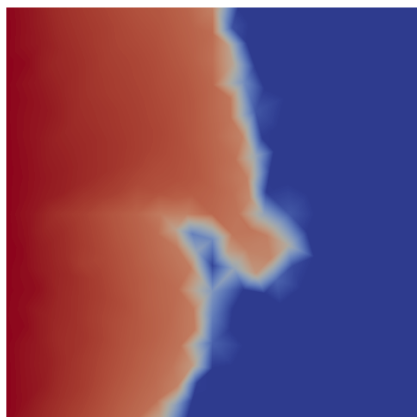
(a) Phase 1 Saturation at  $t=0.15\text{s}$



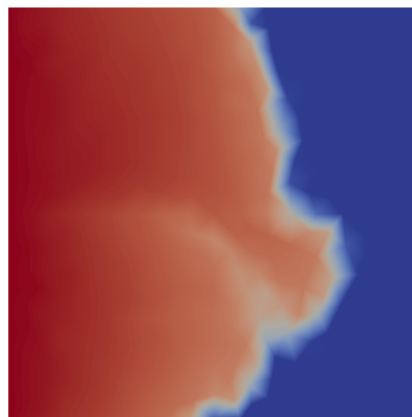
(b)  $t=0.30\text{s}$



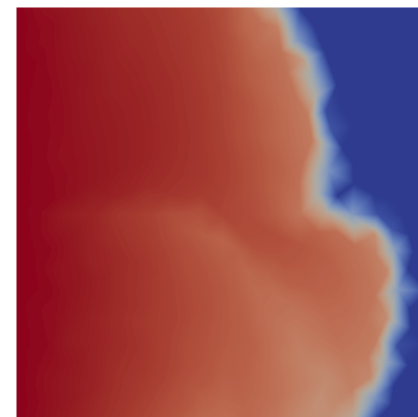
(c)  $t=0.50\text{s}$



(d)  $t=1.15\text{s}$

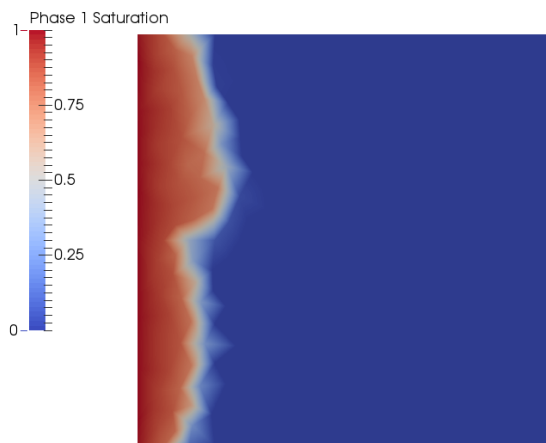


(e)  $t=1.75\text{s}$



(f)  $t=2.95\text{s}$

Figure 6: Simulations for the ArithMean Case (Saturation legend in Fig. 6(a) is representative for all the other figures - *i.e.*, Fig. 6(b) to (f))



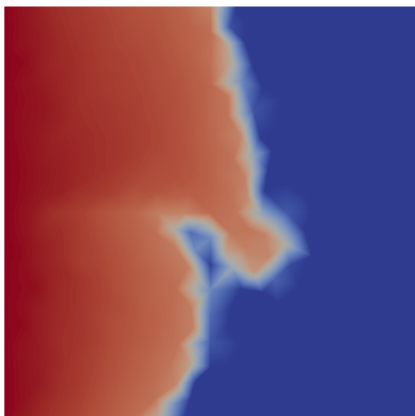
(a) Phase 1 Saturation at  $t=0.15$ s



(b)  $t=0.30$ s



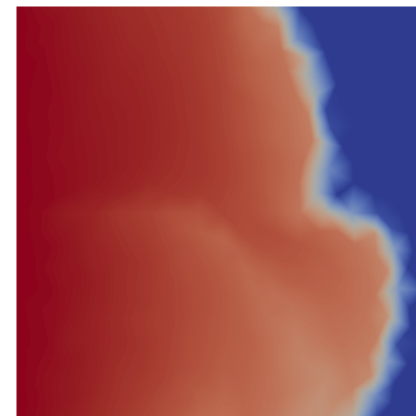
(c)  $t=0.50$ s



(d)  $t=1.15$ s

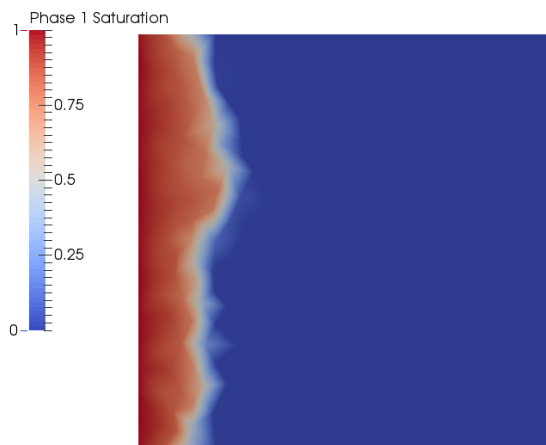


(e)  $t=1.75$ s



(f)  $t=2.95$ s

Figure 7: Simulations for the HarmMean Case (Saturation legend in Fig. 7(a) is representative for all the other figures - *i.e.*, Fig. 7(b) to (f))



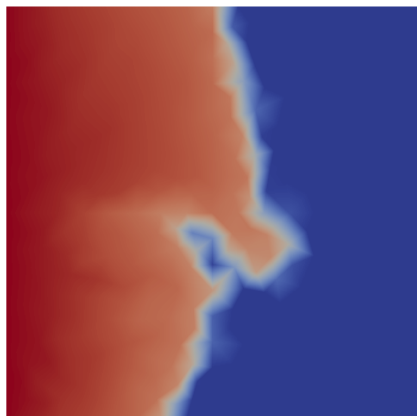
(a) Phase 1 Saturation at  $t=0.15$ s



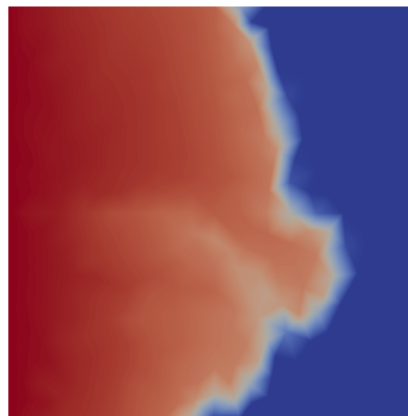
(b)  $t=0.30$ s



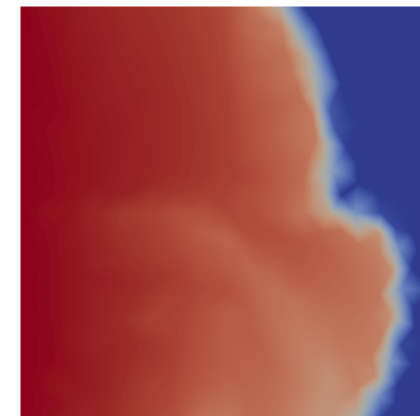
(c)  $t=0.50$ s



(d)  $t=1.15$ s

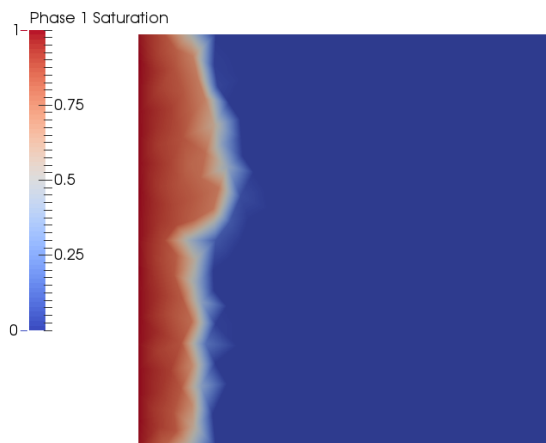


(e)  $t=1.75$ s



(f)  $t=2.95$ s

Figure 8: Simulations for the PDFCase (Saturation legend in Fig. 8(a) is representative for all the other figures - *i.e.*, Fig. 8(b) to (f))



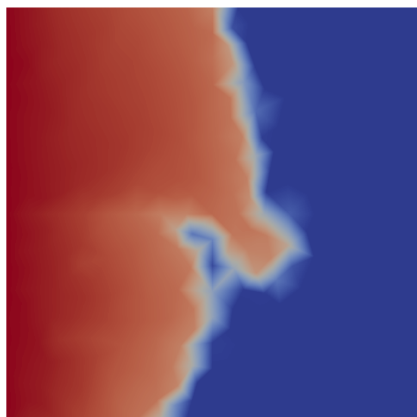
(a) Phase 1 Saturation at  $t=0.15s$



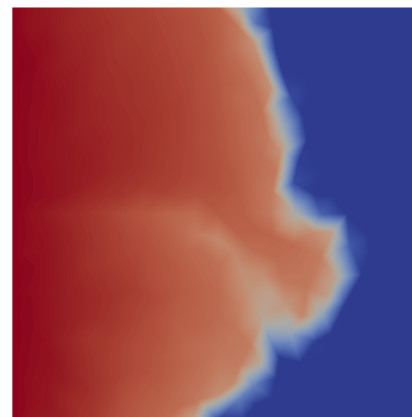
(b)  $t=0.30s$



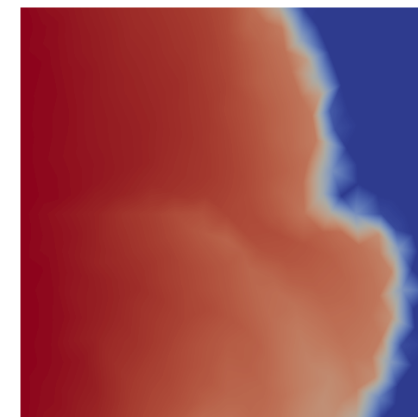
(c)  $t=0.50s$



(d)  $t=1.15s$



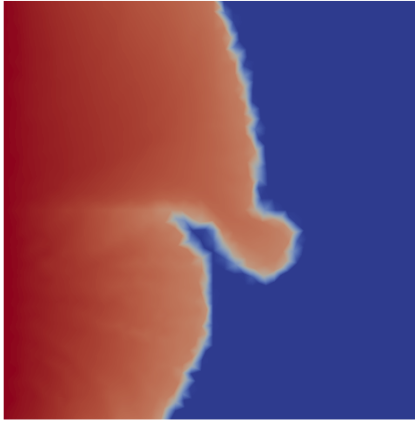
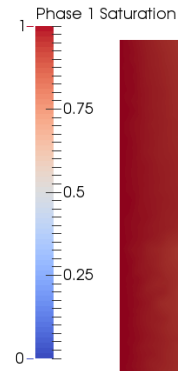
(e)  $t=1.75s$



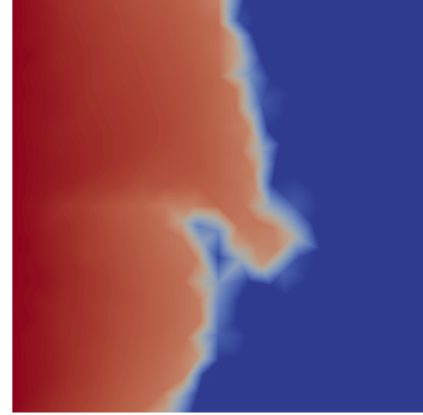
(f)  $t=2.95s$

Figure 9: Simulation for the SVDCase (Saturation legend in Fig. 9(a) is representative for all the other figures - *i.e.*, Fig. 9(b) to (f))

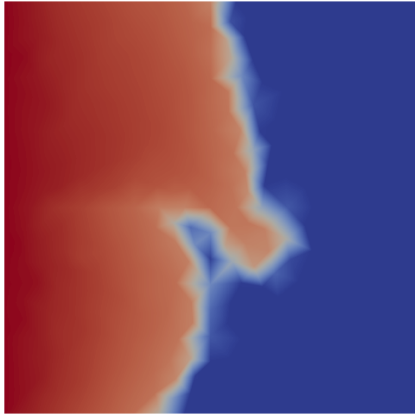




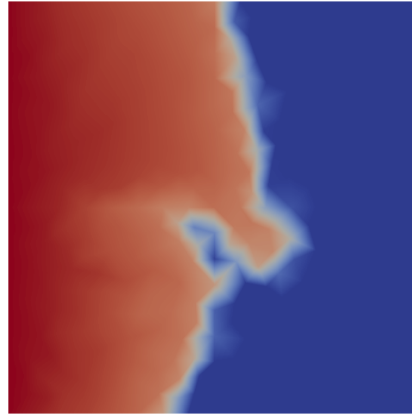
(a) BaseCase



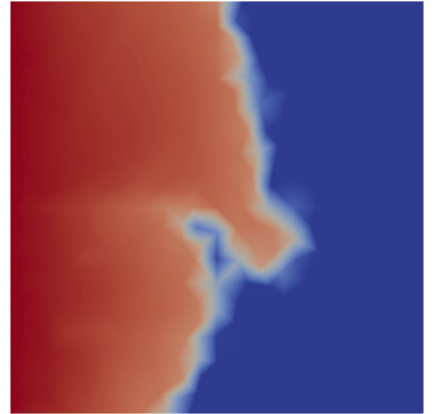
(b) ArithMeanCase



(c) HarmMeanCase

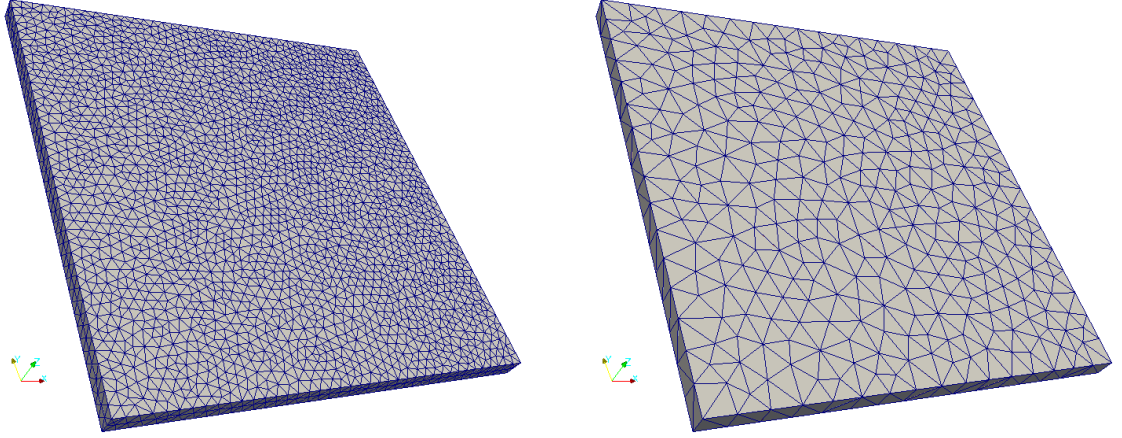


(d) PDFCase



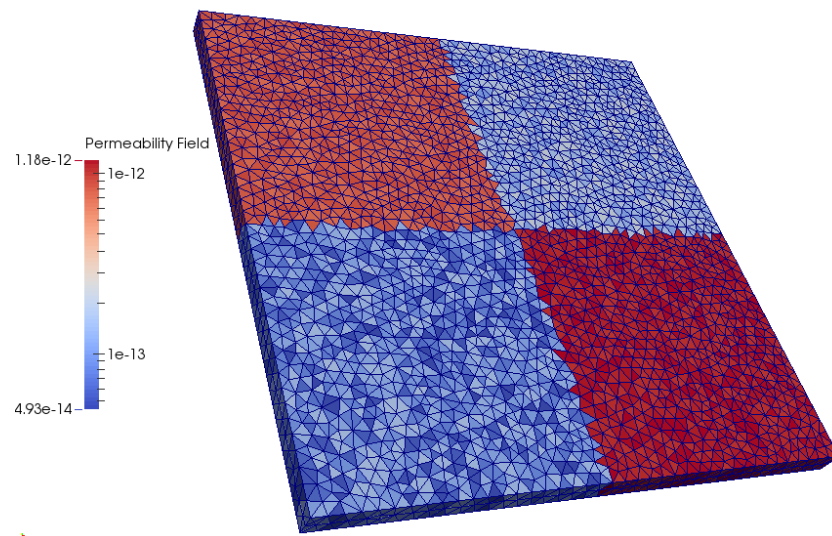
(e) SVDCase

Figure 10: Comparing phase 1 saturation distribution for all the models at  $t = 1.15$  s (Saturation legend in Fig. 10(a) is representative for all the other figures - *i.e.*, Fig. 10(b) to (e))

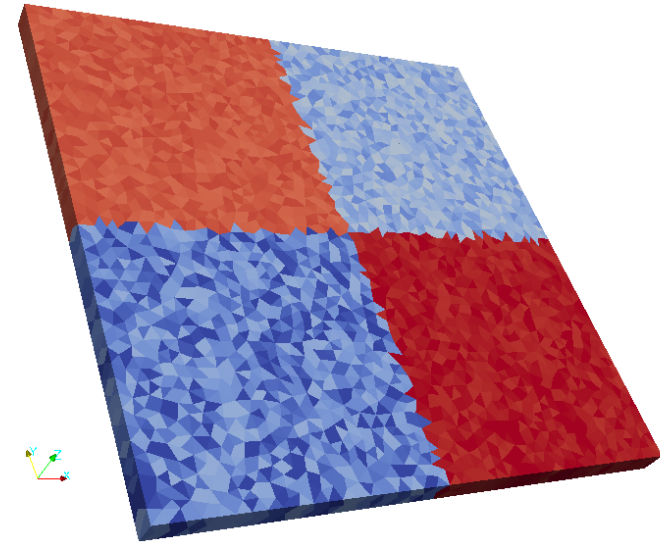


(a) 3-D High Resolution Grid (BaseCase) (b) 3-D Low Resolution Grid (Upscaled Cases)

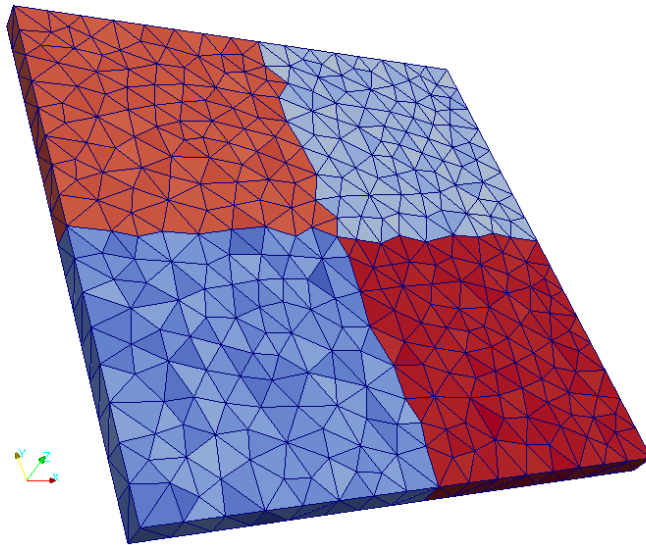
Figure 11: Mesh grid used in the performed numerical 3-D simulations: (a) high-resolution and (b) upscaled cases with 2019 and 303 tetrahedral elements respectively



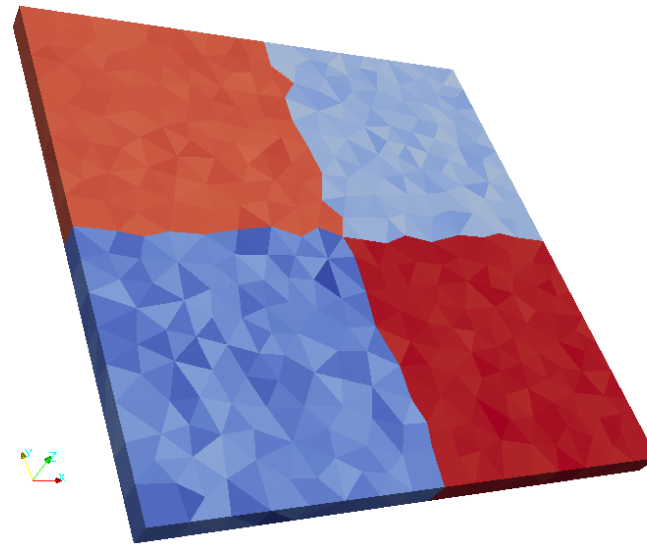
(a) 3-D BaseCase (overlapped with the mesh)



(b) 3-D BaseCase (without Mesh)



(c) 3-D HOSVD Case (overlapped with the mesh)



(d) 3-D HOSVD Case (without Mesh)

Figure 12: Permeability field for the 3D base case as well as the HOSVD upscaled cases (Permeability legend in Fig. 12(a) is representative for all the other figures - *i.e.*, Fig. 12(b) to (d))

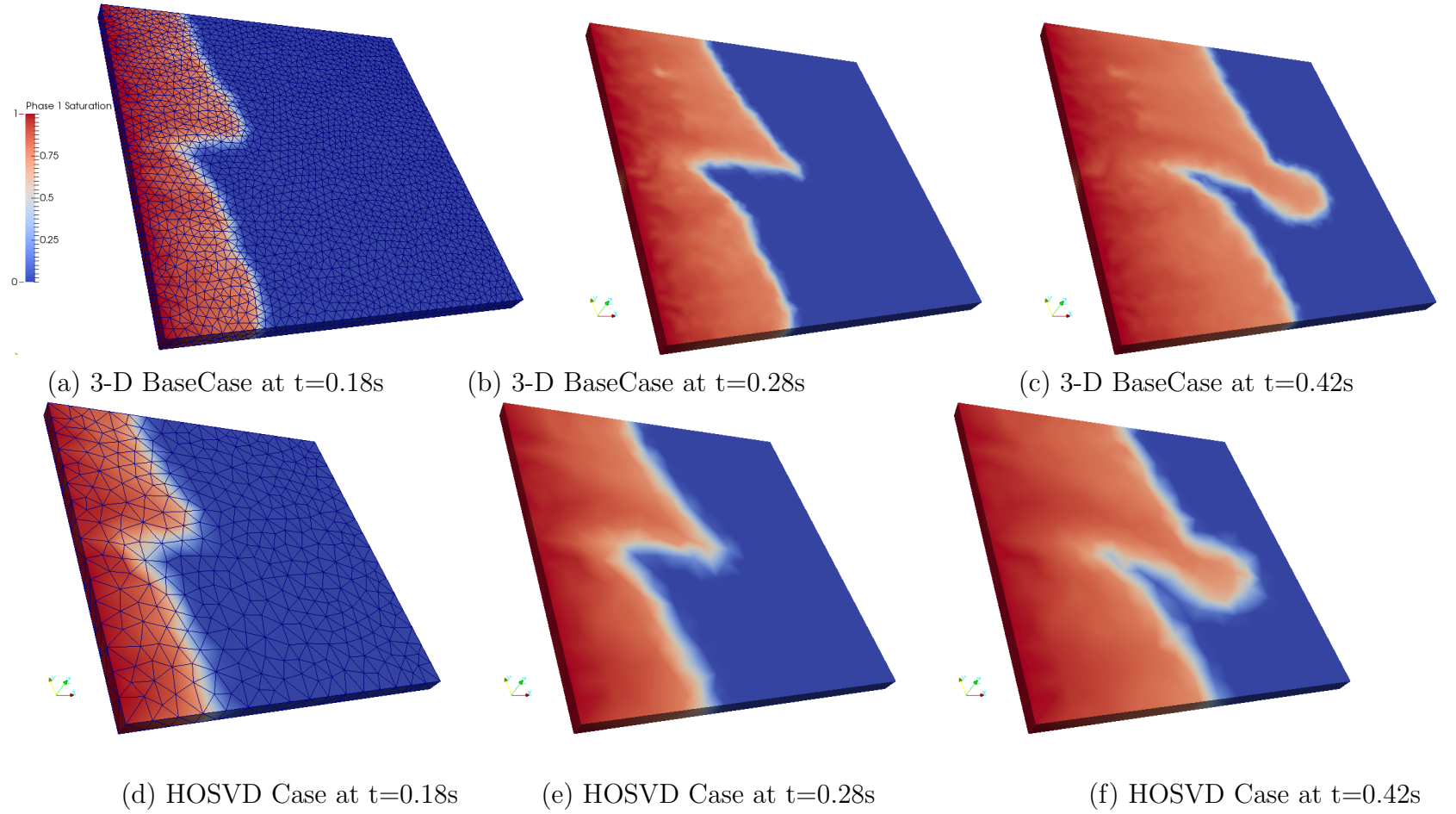


Figure 13: Comparing phase 1 saturation distribution for the 3-D Cases at  $t = 0.18, 0.28$  and  $0.42$  s (Saturation legend in Fig. 13(a) is representative for all the other figures - *i.e.*, Fig. 13(b) to (e))

## Appendix A.

### A Brief Introduction to SVD

Given a matrix  $\mathbf{A} \in \mathcal{R}^{m \times n}$ , the singular value decomposition (SVD) is a method for factorizing  $\mathbf{A}$ , such that,

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top. \quad (\text{A.1})$$

where  $\mathbf{U} = [\underline{u}_1, \underline{u}_2, \dots, \underline{u}_m] \in \mathcal{R}^{m \times m}$  and  $\mathbf{V} = [\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n] \in \mathcal{R}^{n \times n}$  are orthogonal but not necessarily the same.  $\mathbf{\Sigma}$  is a diagonal matrix, *i.e.*,

$$\mathbf{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r) \in \mathcal{R}^{m \times n},$$

and  $r = \min\{m, n\}$  is the rank of  $\mathbf{A}$ , with  $\sigma_1 \geq \sigma_2 \geq \dots \sigma_r \geq 0$ . The diagonal entries of  $\mathbf{\Sigma}$  (*i.e.*,  $\sigma_i, i = 1, r$ ) are called singular values of  $\mathbf{A}$ . If  $\mathbf{A}$  is a square matrix then  $\mathbf{U}$ ,  $\mathbf{\Sigma}$  and  $\mathbf{V}$  are also square matrices. Also, as  $\mathbf{U}$  and  $\mathbf{V}$  are orthogonals and unitary,

$$\mathbf{U}^\top \mathbf{U} = \mathbf{V}^\top \mathbf{V} = \mathbf{I}, \quad (\text{A.2})$$

where  $\mathbf{I}$  is an identity matrix. However, if  $\mathbf{A}$  is an  $m \times n$  matrix,  $\mathbf{U}$  and  $\mathbf{V}$  still remain as square matrices, albeit with different matricial sizes.  $\mathbf{U}$  is an  $m \times m$  while  $\mathbf{V}$  is an  $n \times n$ . Another relevant property of  $\mathbf{U}$  and  $\mathbf{V}$  since they are square (irrespective of whether  $\mathbf{A}$  is square or rectangular) is that

$$\mathbf{U}^{-1} = \mathbf{U}^\top \quad \text{and} \quad \mathbf{V}^{-1} = \mathbf{V}^\top, \quad (\text{A.3})$$

and if  $\mathbf{A}$  is an  $m \times n$  matrix thus  $\mathbf{\Sigma}$  is a diagonal  $m \times n$  matrix.

For a rectangular matrix, the singular values fill the first  $r$  places on the main diagonal of  $\mathbf{\Sigma}$ . The SVD benefits from the orthogonality of the columns in  $\mathbf{U}$  and  $\mathbf{V}$ . The columns of  $\mathbf{U}$  are known as the left singular vectors while the columns in  $\mathbf{V}$  are known as the right singular vectors. A simple consideration of the benefit of the orthogonal factors,  $\mathbf{U}$  and  $\mathbf{V}$ , is the ease with which they allow the inverse of the square matrix  $\mathbf{A}$  to be calculated. This in turn makes it easy to solve the linear equations,  $\mathbf{A}x = b$  with the SVD factorization,

$$\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top x = b \implies x = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^\top b, \quad (\text{A.4})$$

with benefit of the orthogonal property already discussed.

## Appendix B.

### A Brief introduction to HOSVD

Using the tensor notation from Kolda and Bader [30], given a tensor  $\mathcal{X} \in \mathcal{R}^{N_1 \times N_2 \times \dots \times N_m}$ , the higher order singular value decomposition (HOSVD) is a method for factorizing  $\mathcal{X}$ , such that,

$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{U}^1 \times_2 \mathbf{U}^2 \times \dots \times_m \mathbf{U}^m. \quad (\text{B.1})$$

where tensor  $\mathcal{G} \in \mathcal{R}^{N_1 \times N_2 \times \dots \times N_m}$  and is normally referred to the core tensor. Sub-tensors of  $\mathcal{G}$  are orthogonal and ordered.

And matrices  $\mathbf{U}^m \in \mathcal{R}^{N_m \times N_m}$  are orthogonal factor matrices which could be thought of as the principal components in each mode