~~A Reduced Order Model for Permeability Fields Using Singular Value Decomposition~~A Reduced Order Model for Porous Media Flows: Coupling Linear Regression with Principal Component Analysis (PCA) using SVD or High-Order Singular Value Decomposition (HOSVD) in Upscaling Calculations for Permeability in Reservoir Simulations

B. Lashore[a,*], K. Christou[a], J.L.M.A. Gomes[a]

[a]*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 SVD-based method was developed to reduce the order of porous media permeability field (*i.e.,* upscale)~~ 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 for a new permeability field with a reduced order. This PCA-linear re-

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

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

Keywords: Principal Component Analysis (PCA), Linear Regression, Higher-Order Singular Value Decomposition (HOSVD), Upscaling, Reduced Order Models (ROM), Permeability.

---

1. Introduction

Naturally occuring rock formations are inherently heterogeneous, which means that ~~rock~~ geophysical 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_2 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) [11, 34, 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 (*i.e.,* low-resolution) of these properties [50].

Upscaling techniques which best preserve statistical ~~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~~ geophysical and thermo-physical properties of high-resolution domains result in optimal prediction of fluid flow dynamics. Over the past 60 years, several upscaling techniques were developed (see [23, 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 [22, 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 uncertainity problems.statistical

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 be simply ~~be~~ referred as 'dimensionality reduction'. The definition of MOR also depends on the context ~~it is being~~ in which it is used, but it suffices to state that ~~MOR captures~~ MOR's are able to capture 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 Odabasioglu et al. [37] in 1998, which introduced the passive reduced-order interconnect macromodeling algorithm (PRIMA). Truncated balanced realization is another MOR method developed by Moore [35], which introduced the principal component analysis (PCA) [25] and Golub and Reinsch's [19] algorithm for solving singular value decomposition (SVD). Hankel-norm reduction [18] and proper orthogonal decompositions (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 [17] and the PRIMA. ~~In this work, upscaling is performed by SVD in principal~~

In this work, upscaling is performed through linear regression in the PCA space which is obtained with high-order singular value decomposition (HOSVD/SVD) in 2-/3-D problems. This work focuses on reducing the order

of permeability distribution for porous media flow simulations.

~~Four upscaling techniques are investigated in this work, arithmetic and harmonic means are used to obtain the first two upscaled 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 singular value decomposition (SVD) as a method of upscaling using the concept of principal component space to reduce the order of the permeability distribution.~~

A brief overview of current upscaling methods in reservoir simulation is described in Section 2. Section 3 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 high-order accurate control volume finite element method (CVFEM) [20]. 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 arithmetric mean, harmonic mean, probability density function (PDF) and PCA (using SVD and HOSVD methods) coupled with linear regression. In Section 4, 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 4 describes the reduced order model reduction for the permeability field within the principal component space. Section 5 is dedicated to model description and simulation. It starts by describing the high-resolution problem (*i.e.,* 'base case') in which the four upscaling techniques are applied to. A brief summary of the pre-processing steps for each of the models is also presented. Section 6 provides the results and further discussion of the results, it also gives further interpretation to the result obtained from the SVD upscaling technique. Finally, conclusions are drawn in Section 7.

2. Upscaling

2.1. Brief Overview of Upscaling

~~Section 1 briefly introduced the rational behind upscaling~~ Upscaling methods replace discrete geological and fluid properties of detailed (high-resolution) domains with coarse descriptions (low-resolution) of these properties [50]. Upscaling of heterogeneous properties has attracted the interest of several porous media flow communities (*e.g.,* oil and gas, CCS and waste management) worldwide for over 7 decades. Those interested in large-scale reservoirs or groundwater generally take a two-stage geo-statistical 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 rock properties of the facies [16]. In addition to seismic and well data used in the previous stage, core data can be used as a source of information for the statistical properties used to determine random fields in this stage.

A number of studies were carried out on upscaling of subsurface permeability (see [12, 16, 23, 26, 30, 50, 52, 54] for a few key references). 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 harmonic and arithmetic averages.

Yeo and Zimmerman [54] discussed the accuracy of renormalization method for upscaling 2D hydraulic conductivity fields for two cases. Both cases employed the 4-quadrant model (see Section 5.1 for details) set-up. The first case had low permeability in all blocks except the bottom right block which had high permeability, while the second case employed the check-board style described in Section 5.1. 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. They also recommended to select a technique which provides block permeabilities instead of one which

provides uniform effective permeability. Finally, Christie and Blunt [12] provided 2 sets of problems to compare the upscaling and upgridding (*i.e.,* mesh coarsening) 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 flow simulation models' benchmarking.

2.2. 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 interchangable, and are normally referred to as Karhunen-Loeve Transformations (KLT) [29].

A number of recent research in reservoir engineering focus on history matching (HM) problems (see [2, 27, 29, 48, 53]). History matching is the process of building or adjusting one or more sets of reservoir configurations (or realisations), such that reservoir models are able to closely reproduces past flow behaviour (*e.g.,* flow rate, production volume, pressure etc) in a geological formation, aiming to predict future flow behaviour with good accuracy [42]. It is important to note that reservoir configurations could be either 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 [29]compares KLT and discrete cosine transform (DCT) and concluded that the DCT is computationally more efficient than the KLT while being almost as accurate. Tavakoli and Reynolds [48] noted that HM calculation of sensitivities for all production data with respect to system parameters is not feasible. Hence, they explored a new parameterization for reducing the number of variables based on the principal right singular vectors of the dimensionless sensitivity matrix.

Perhaps, work performed by Afra and Gildin [2, 3], Afra et al. [4] and Insuasty et al. [27]are more closely related to the one performed here. However, similar to the work of Jafarpour and McLaughlin [29], 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 flow simulator model used in this work. However, similarly to Tavakoli and Reynolds [48], Xiao et al. focused on reducing the number of variable used to describe the system.

Essentially, recent developments in the field of reservoir simulation described in this section make use of SVD theory 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 a single system variable (*i.e.,* permeability).

3. 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, 10].

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, $\mathrm{P}_n\mathrm{DG}$-P_m and $\mathrm{P}_n\mathrm{DG}$-$\mathrm{P}_m\mathrm{DG}$. In this element-pairs velocity is represented by n^{th}-order polynomials that are discontinuous across elements, while pressure is represented by m^{th}-order polynomials that may be either continuous ($\mathrm{P}_n\mathrm{DG}$-P_m) or discontinuous ($\mathrm{P}_n\mathrm{DG}$-$\mathrm{P}_m\mathrm{DG}$) across elements. Most properties of $\mathrm{P}_n\mathrm{DG}$-$\mathrm{P}_m\mathrm{DG}$ element-pairs are similar to those in the $\mathrm{P}_n\mathrm{DG}$-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,

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

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

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

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

are discretised in space with CV basis functions, and in time with the θ-method [21]. In Eqn. 2, ϕ 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[1] (a full description of the model can be found in [20, 28, 43]).Although, the numerical formulation used here is relatively new, it is not the focus of this work. The result and discussion section (Section 6) refers to Dawe and Grattoni [15] for validation of the numerical formulation with respect to the 2D simulation. However, extensive model and software quality assurance (*i.e.,* verification and validation tests) can be found in the aforementioned references [1, 20, 28, 38, 44].

Finally, $\mathbf{K}(\mathbf{r})$, the focus of this work, is prescribed in the system of governing conservative equations, *i.e.,* it is an independent input variable. Since $\mathbf{K}$ is a petrophysical property of geological formations that varies spatially, it cannot be determined at every grid point (*i.e.,* 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 reason upscaling is required for processing permeability spatial variability. Upscaling can be performed from the continuum scale, across the micro-scale (mm), local- and meso-scale (m) to field-scale (km)[24]. The upscaling discussed here is for micro-scale, also known as the Darcy scale.

[1] http://multifluids.github.io

4. Singular Value Decomposition (SVD) and Application for Reducing the Order of Permeability FieldsCoupled PCA and Linear Regression

4.1. 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,

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\intercal}. \tag{3}$$

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

4.2.1. Understanding the permeability space

Permeability can be represented as either tensor or scalar quantity, however for the purpose of the following explanation (and the remaining of this work), permeability is assumed as a scalar quantity. In addition, in the numerical simulations performed for this work, permeability is discretised element-wise with projection to the control volume space within each element-pairs (see [13]). Permeability values are assigned to the centre of the FE (*i.e.,* P_0DG element-pair). For simplicity, in the remaining part of this sub-section, a $2D$ geometry is assumed which is discretized with triangular elements (P_1DG-P_2 and P_1DG-P_1DG element-pairs). This explanation allows one to easily visualize this permeability space because it is defined by Cartesian coordinate 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.

4.2.2. Understanding the principal component spaces

Essentially, the existing permeability space is transformed (with SVD and HOSVD for 2- and 3-D permeability data, respectively) to principal

component spaces (PCS). Within the PCS, traditional PCA dimensionality reduction can be optionally performed (*i.e.,* feature selection or elimination, Section 4.2.3), before interpolation is used to implement further data reduction in PCS. 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 5.1.3

The PCS is a particularly challenging to visualize as it represents projections which are embedded in SVD/HOSVD factors of the original permeability dataset. An aid for visualizing the PCS for a 2D permeability data is to write out $\mathbf{U}$, $\mathbf{\Sigma}$ and $\mathbf{V}^\intercal$ matrices (or SVD factors) and consider the singular vectors (*i.e.,* columns in $\mathbf{U}$ and rows in $\mathbf{V}^\intercal$) as separate PCS.A similar procedure can be extended to HOSVD for 3D permeability dataset. The decomposition of a matrix into its factors using SVD [14, 49] or HOSVD [31, 33] is covered in details in several published materials. However, for completeness, Appendix A and Appendix B contain brief overviews of SVD and HOSVD, respectively.

4.2.3. Reduced Order Model/Principal Component Space Reduction

For the PCA-Linear regression reduced order model (PCA-LinReg) of the permeability data, all work is performed in PCS. This can be conducted in two steps (for 2-D models), thus:

Step 1: Perform PCA [25, 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 σ_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 $\mathbf{V}^\intercal$ are extracted from $\mathbf{U}$ and $\mathbf{V}^\intercal$ to form new matrices $\mathbf{U_{new}}$ and $\mathbf{V}^\intercal_{\mathbf{new}}$ respectively. The selected z singular values also form a new diagonal matrix, $\mathbf{\Sigma_{new}}$.

Step 2: Interpolation within each singular vector (column in $\mathbf{U}$ and row in $\mathbf{V}^\intercal$) is conducted to reduce the number of data points in each column of $\mathbf{U}$ and in $\mathbf{V}^\intercal$. From this step, $\mathbf{U_{new2}}$ and $\mathbf{V}^\intercal_{\mathbf{new2}}$ are obtained. The reduced permeability field is calculated by multiplying $\mathbf{U_{new2}}$, $\mathbf{\Sigma_{new}}$ and $\mathbf{V}^\intercal_{\mathbf{new2}}$.

These two steps can be readily replicated for 3-D models using HOSVD to obtain the PCA space.

Finally on the PCS, it is worth noting that two methods exist for obtaining PCA space: co-variance and SVD methods (see [49] detailed description of both methods). In this work, the authors have opted to use the SVD method to obtain the PCA space, as singular vectors which make up the PCS have 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.

4.3. Principal Component Analysis Versus Linear Regression

Ng [36] 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 and can also be deduced from 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 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 $f(x_1, x_2, ..., x_n) = y$
3	Not used for predictions, as original feature space (*i.e.,* x_1 and x_2) give way to a new reduced dimension. Cordinates $x_1, x_2, ..., x_n$ give way to $z_1, ..., z_k$ where $k \leq n$	

Table 1: Table comparing PCA with Linear Regression.

5. Model Description and Simulation Setup

As previosly 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.

5.1. Model Description

The BaseCase (*i.e.,* high-resolution mesh) configuration for the permeability dataset builds on the 2×2 block, *i.e.,* 4 quadrant model orginally used by Cardwell and Parsons [9] and later by Yeo and Zimmerman [54] and Dawe and Grattoni [15]. In the domain designed by these authors, permeability is represented as checker-board in which 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 desgined used by these references. 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). Furthermore, within each block, permeability dataset varies within a pre-defined range and consists of 1600 permeability values. Details on the design of the permeability field for the BaseCase are described in Section 5.1.1. Figures 4a-b show permeability distribution for the BaseCase with the associated mesh resolution. Using the labelling scheme used by Dawe and Grattoni [15], the quadrant model block $K1$, $K2$, $K3$ and $K4$ are top left, bottom left, top right and bottom right, respectively.

As previously mentioned, four upscaled techniques applied to permeability dataset were obtained from the high-resolution (*i.e.,* the BaseCase) configuration. Prior to upscaling, mesh resolution of the BaseCase configuration was coarsened, Fig. 3, in which the number of triangular elements was reduced from 4112 to 728. Arithmetic (ArithMean) and harmonic (HarmMean) averaging methods were initially applied to the BaseCase configuration. 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 permeability values for each block in the upscaled models. This effectively homogenized the

permeability dataset in each block. Permeability dataset was also upscaled using probability density function (PDFCase) and SVD (PCA-LinReg Case). Description of these models are introduced in Sections 5.1.2-5.1.3. Different from ArithMean and HarmMeanand, PDFCase and PCA-LinReg Case are able to retain heterogeneity within each block.

5.1.1. BaseCase Pre-processing Algorithm

Step 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;

Step 2: The number of data points generated for each block is 1600, and data is randomly distributed within each block in a structured pattern. Albeit, the mesh is unstructured;

Step 3: Merge the dataset from different blocks into a single file retaining the original "structured pattern" (*Step 1*) for the randomly positioned dataset for each block. This data-structure will be used by the CVFEM-based flow simulator to prescribe the permeability dataset for the BaseCase configuration.

5.1.2. PDFCase Pre-processing Algorithm

Step 1: Datasets from *Step 1* of Section 5.1.1 are used to calculate mean and standard deviation for each block of permeability data;

Step 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 configuration for each of the four blocks (*i.e.,* 400 data points for each block);

Step 3: Data values in the dataset for each block is randomly positioned within each block in a structured pattern. Similarly to the BaseCase configuration, simulation mesh grid is also unstructured;

Step 4: Finally, the structured/arranged dataset from each block is combined into a single data-structure and then used by the flow simulator, which merge the permeability mapping to the mesh with the appropriate element-pairs (in simulations shown here, P_1DG-P_2 and P_1DG-P_1DG element-pairs are used).

5.1.3. 2-D PCA (SVD) & LinReg Case Pre-processing Algorithm

Step 1: Permeability data obtained in *Step 3* of Section 5.1.1 is factorized using SVD method to obtain $\mathbf{U}, \Sigma$, and $\mathbf{V}$;

Step 2: A selected number (z) of singular values are retained in decreasing order of magnitude to form a new diagonal matrix, Σ_{new} (this step is optional as it is the conventional PCA dimensionality reduction);

Step 3: First z columns of $\mathbf{U}$ and $\mathbf{V}$ are selected to form new truncated matrices called $\mathbf{U_{new}}$ and $\mathbf{V_{new}}$, respectively. This step is necessary only if the optional step in 2 above is executed;

Step 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 $\mathbf{V_{new2}}$;

Step 5: Permeability dataset used for the PCA-LinReg Case is the matrix $\mathbf{A_{new2}}$, which is defined as the product of $\mathbf{U_{new2}\Sigma_{new}V_{new2}^{\intercal}}$. It is a permeability dataset with a quarter of the data values of the BaseCase, *i.e.,* a permeability dataset with reduced order.

5.1.4. 3-D PCA (i.e., using HOSVD) & LinReg Case Pre-processing Algorithm

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

1. In *Step 1* of Section 5.1.3, HOSVD method is used for factorization instead of SVD. Therefore, a 3×3 core tensor is obtained in place of the $\mathbf{\Sigma}$ term, while 3 factor matrices are obtained in place of $\mathbf{U}$ and $\mathbf{V}$;
2. The core tensor is treated as the $\mathbf{\Sigma}$ term, while the 3 factor matrices are treated similarly to $\mathbf{U}$ and $\mathbf{V}$ as described in Section 5.1.3.

The authors found that the Tensorly library package [32] helpful for implementation of the HOSVD algorithm.

5.2. Simulations

In all performed numerical simulations, the 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. A no-flux boundary conditions were imposed on upper and lower borders of the domain, while mixed fluids left the domain from the right-hand border 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, is described in the pore rock matrix by the modified Brooks and Corey [7] (see [5]),

$$\mathcal{K}_{rw}\left(S_w\right) = \mathcal{K}^{\circ}_{rw}\left[\frac{S_w - S_{w,irr}}{1 - S_{w,irr} - S_{nw,r}}\right]^{n_w}, \tag{4}$$

$$\mathcal{K}_{rnw}\left(S_{nw}\right) = \mathcal{K}^{\circ}_{rnw}\left[\frac{S_{nw} - S_{nw,r}}{1 - S_{w,irr} - S_{nw,r}}\right]^{n_{nw}}, \tag{5}$$

where subscripts w and nw stand for wetting and non-wetting phases, respectively. $\mathcal{K}^{\circ}_{rw}$ and $\mathcal{K}^{\circ}_{rnw}$ 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 the 2-D cases, whereas Figs. 5-9 show phase saturation distribution during fluid injection for time,

$$t = 0.15,\ 0.30,\ 0.50,\ 1.15,\ 1.75 \text{ and } 2.95 \text{ s for all the 2-D cases.}$$

Additionally, the number of P_1DG-P_2 triangular element-pairs mesh resolution was reduced from 4112 (BaseCase, Fig, 3a) to 728 (upscaled cases, Fig, 3b), a reduction by a factor of approximately 5.6.

6. Results and Discussion

After reduction of dimensionality of the permeability dataset using the methods described in Sections 4.2 and 5.1), numerical simulations were performed with Fluidity/IC-FERST model (see Section 3 and [20, 28, 43]).

6.1. 2-D Models and Simulations

Figure 4 shows permeability distribution for all the 2-D cases and highlights homogeneous representation resulting from arithmetic and harmonic

averaged, Figs. 4c-d, whereas PDFCase and PCA-LinReg cases (Figs. 4e-f) are inherently heterogeneous. Given that the BaseCase is also heterogeneous, it can be stated that models for the PDFCase and PCA-LinReg Case give a qualitatively better representation of the permeability dataset within the BaseCase. An additional consequence of the homogeneity in the ArithMeanCase and the HarmMeanCase is that the coarsening has no influence on the resulting permeability distribution. However, in both, PDFCase and PCA-LinReg Case, permeability distribution will change as mesh resolution is reduced.

With respect to the PDFCase and for this specific resolution, it is important to note that Fig. 4(e) is only one realiztion of infinitely many possible representations of the PDFCase. Meanwhile, the PCA-LinReg Case will retain this specific permeability distribution (*i.e.,* Fig. 4(e)) for such mesh resolution (*i.e.,* if linear interpolation is used, but this digresses from the aim of this work and it will be discussed in future publications). 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, throught 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 future work.

Before discuss flow simulation results with permeability distribution resulted from methods described in Sections 4.2 and 5.1), it is beneficial to briefly consider the Buckley-Leverett flow model 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 permeabilities. 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 [15] conducted experiments to investigate cross-flow, and as previously mentioned, the permeability field for the experimental work were replicated here for the simulations. Therefore, the work of Dawe and Grattoni [15] can be used to qualitatively 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. 5-9a-b, even though fluid is injected at the same rate across the left face it travels faster in $K1$ compared to $K2$. Figures 5-9c show that the fluid experiences a resistance to flow at the interface between $K1$ and $K3$ (due to lower permeability at $K2$ and $K3$) which encourages the flow into $K4$ (*i.e.,* bottom-right block, Figs. 5-9d-e). Figures 5-9 (f) indicates preferential flow in $K4$ compared to $K3$.

Figure 10 compares phase 1 saturation distribution for all the 2-D cases at $t = 1.15s$. Generally, saturation distribution for all the 2-D cases closely resemble each other with arithmetic and harmonic mean (Figs. 10b-c) being nearly identical. Albeit, closer inspection will reveal that there is more displacement in Fig. 10b compared to Fig. 10c. 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. 10b-c is that they poorly represent the BaseCase (Fig. 10a) at the top left hand-side corner of block $K2$.

Figure 10d also shows an inconsistency with the BaseCase (Fig. 10a) 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. 10e gives the best representation of the BaseCase in *i.e.,* Fig. 10a.

6.2. 3-D Models and Simulations

Following on from the work described for the 2-D models and simulations, Figure 11(a) and (b) presents the mesh grid for building the 3-D models. Figure 11(a), with 2109 tetrahedral elements, is the high resolution grid and is used to run the simulation for the 3-D base case. While Figure 11(b), with 303 tetrahedral elements, is the low resolution grid used to run the simulation for the 3-D PCA(using HOSVD)-LinReg case. The number of tetrahedral elements for the low grid resolution grid is approximately 7 times less than that for the high resolution grid.

Figure 12 shows the permeability distribution for the 3-D cases. The aim here is to show that the PCA-LinReg model, previously obtained with SVD for the 2-D case is obtainable with HOSVD for a 3-D model. Therefore, we only concentrate on one upscaling case for the 3-D model, that is the PCA (using HOSVD) LinReg Model which is in Figure 12(b). When this model is compared to the base case, Figure 12(a), the same advantages of PCA-LinReg model from the 2-D case are observable. For example:

1. The model retains heterogeneity within each region
2. A single realization will be obtained for the same upgrid resolution
3. The correlation to local heterogeniety in the base case is maintained

Figure 13(a)-(c) and Figure 13(d)-(f) show the phase 1 saturation distribution for the base case at PCA(using HOSVD)-LinReg case for time,

$$t = 0.18,\ 0.28 \text{ and } 0.42 \text{ s}.$$

Figure 13(a) and (d), which represent the phase 1 saturation distribution for the base case and the 3-D PCA-LinReg case respectively at time $t = 0.18$, retain the mesh grid to give a more details view of the flow pattern. Comparing, Figure 13(b) and (e) as well as Figure 13(c) and (f) indicates that the 3-D PCA-LinReg model gives a good representation of the flow pattern. A detailed quantitative analysis of this upscaled model will be presented in the authors' future work.

6.3. Computational Cost

An accurate measure of the CPU cost is not available at this time, however a general note can be made regarding the savings in computational costs with respect to the upgridding resolution.

The 2-D base case which had 4112 elements was run for about 7 hours while the upscaled 2-D case were completed under 20 minutes. For the 3-D case, the base case which had 2109 elements required about 6 days to run while the upscaled 3-D case was completed in under 6 hours.

7. Conclusion

Upscaling is a form of reduced order model. 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 data in PCA space. The reduced data in the PCS is then recombined to form a reduced data of the permeability dataset.

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 dataset irrespective of the coarser resolution. The PCA-LinReg Case is the only technique that does not require a prior knowledge of blocks with similar permeability distribution before upscaling. This is a huge advantage as it saves time during data pre-processing stages 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., Muggeridge, A., Jackson, M., 2016. Higher-order conservative interpolation between control-volume meshes: Application to advection and multiphase 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., 2007. Reservoir Simulation: Mathematical Techniques in Oil Recovery. Siam.

[11] Chen, Z., Huan, G., Ma, Y., 2006. Computational Methods for Multiphase Flows in Porous Media. SIAM.

[12] Christie, M., Blunt, M., 2001. Tenth spe comparative solution project: A comparison of upscaling techniques. Society of Petroleum Engineers SPE-66599-MS.

[13] 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).

[14] Cline, A. K., 2007. Computation of the Singular Value Decomposition. Chapman & Hall/CRC, Ch. 47, pp. 45.1–45.13.

[15] 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.

[16] 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

[17] 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.

[18] GLOVER, K., 1984. All optimal hankel-norm approximations of linear multivariable systems and their l, -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] Gomes, J., Tollit, B., Milles, B., Pain, C., 2013. Fluidization Engineering – Principles and Practice. National University of Singapore, Singapore, Ch. Multi-Physics Flow Modelling for Nuclear Applications, http://serve.me.nus.edu.sg/arun/file/Publications/books/FluidizationEngineering-Practice.pdf.

[22] 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.

[23] Hastings, J., Muggeridge, A., 2001. Upscaling uncertainity permeability using small cell renormalization. Mathematical Geology 33(4), 491–502.

[24] 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 Mathemetics in Industry, Springer, pp. 449–488.

[25] 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

[26] 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.

[27] 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.

[28] 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.

[29] Jafarpour, B., McLaughlin, D. B., 2009. Reservoir characterization with the discrete cosine transform. Society of Petroleum Engineers (SPE) SPE-106453-PA.

[30] 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

[31] Kolda, T. G., Bader, B. W., 2009. Tensor decompositions and applications. SIAM Review 51, 455–500.

[32] Kossaifi, J., Panagakis, Y., Anandkumar, A., Pantic, M., 2018. Tensorly: Tensor learning in python. CoRR abs/1610.09555.

[33] Lathauwer, L. D., Moor, B. D., Vandewalle, J., 2000. A multilinear singular value decomposition. SIAM J. Matrix Analysis Applications 21, 1253–1278.

[34] 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.

[35] Moore, B., February 1981. Principal component analysis in linear systems: Controllability, observability, and model reduction. IEEE Transactions on Automatic Control 26 (1), 17–32.

[36] 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

[37] 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.

[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

[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

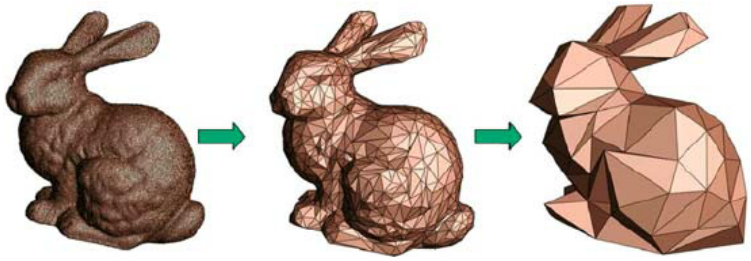
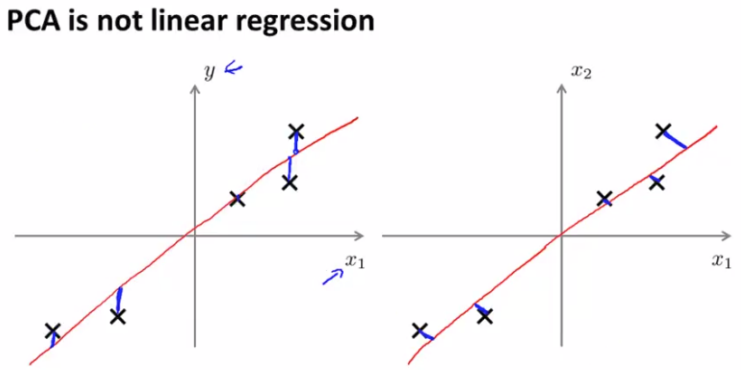
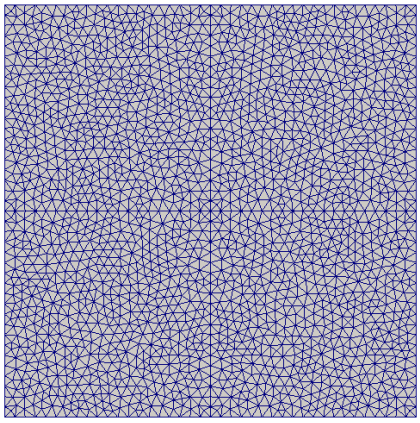


Figure 1: Graphical illustration of model order reduction (initially used by Schilders [45], with graphics credited to Harvard University, Microsoft Research.)

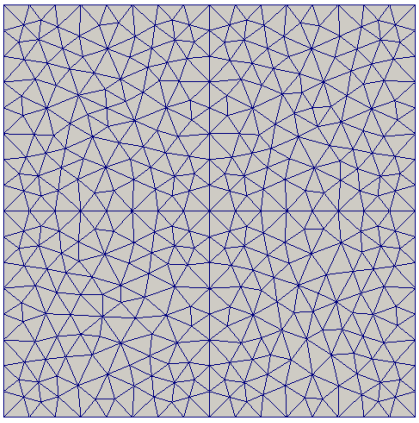


(a) Linear regression (b) PCA

Figure 2: Graphical comparision between linear regression and PCA (extracted from Ng [36]).

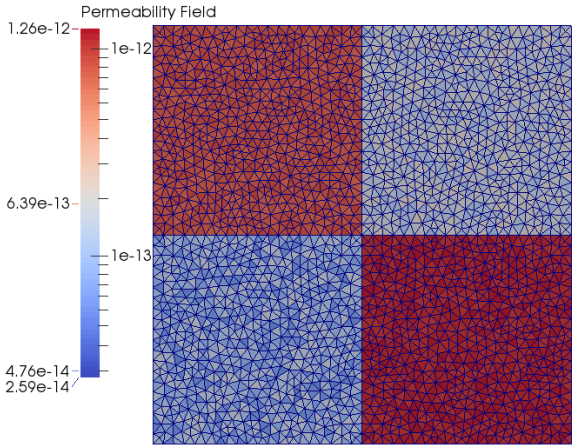


(a) High Resolution Grid (BaseCase)

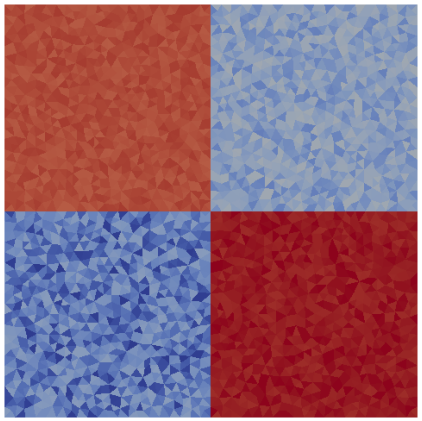


(b) Low Resolution Grid (Upscaled Cases)

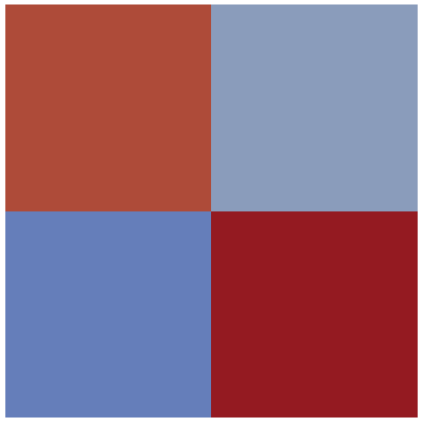
Figure 3: Mesh grid used in the performed numerical simulations: (a) high-resolution and (b) upscaled cases with 4112 and 728 triangular elements, respectively.



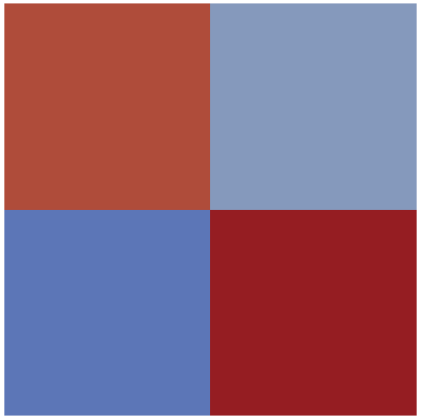
(a) BaseCase (overlapped with the mesh)



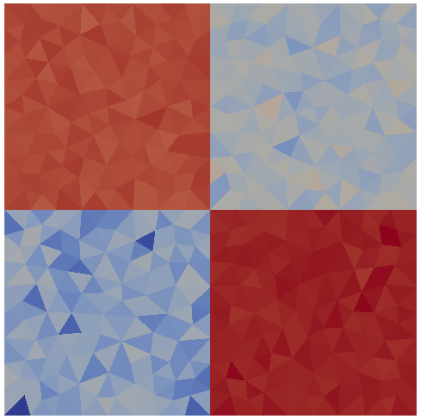
(b) BaseCase (without Mesh)



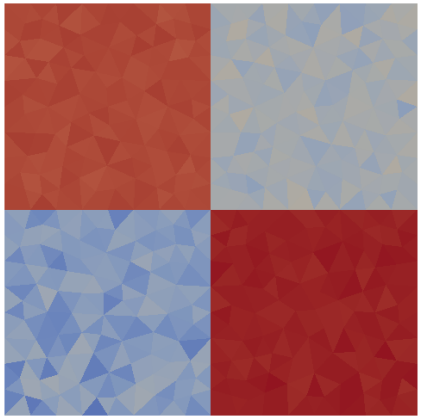
(c) ArithMean Case



(d) HarmMean Case

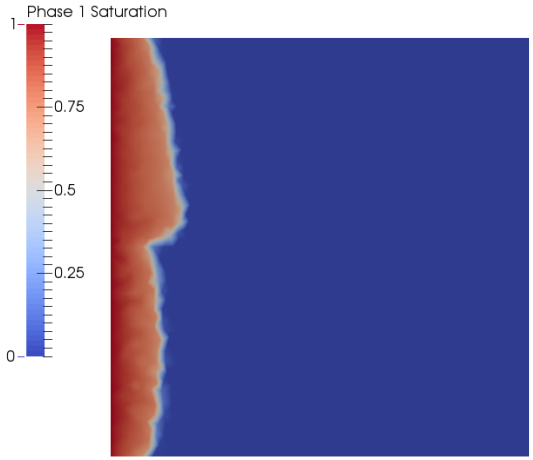


(e) PDFCase



(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

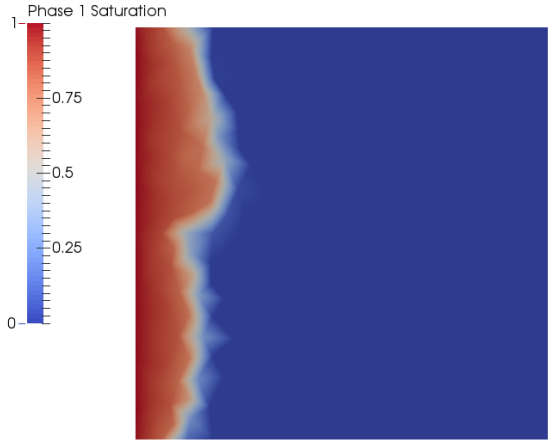
(c) t=0.50s

(d) t=1.15s

(e) t=1.75s

(f) t=2.95s

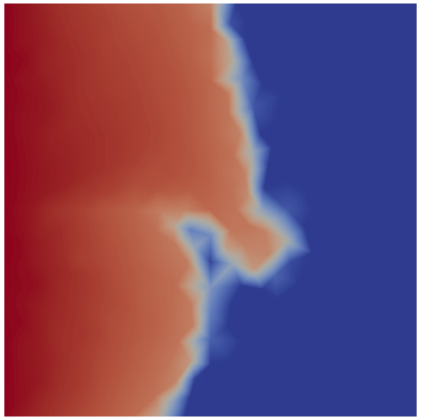
Figure 5: Simulations for the BaseCase (Saturation legend in Fig. 5(a) is representative for all the other figures - *i.e.,* Fig. 5(b) to (f))



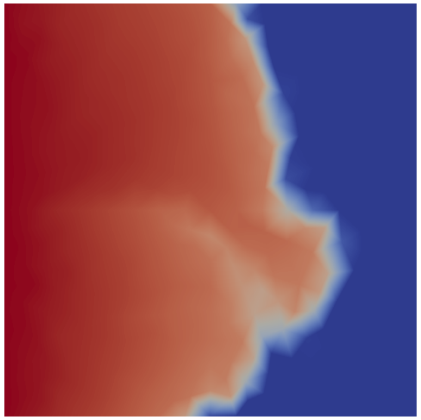
(a) Phase 1 Saturation at t=0.15s

(b) t=0.30s

(c) t=0.50s



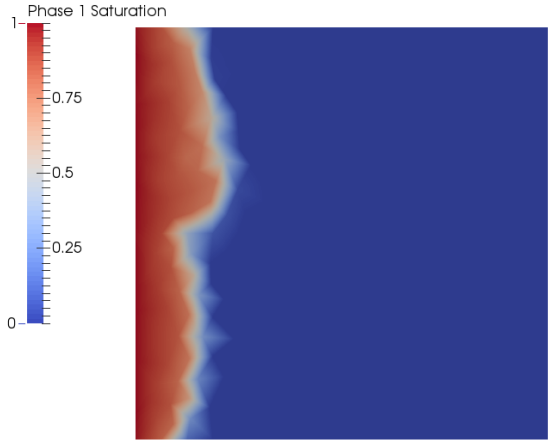
(d) t=1.15s



(e) t=1.75s

(f) t=2.95s

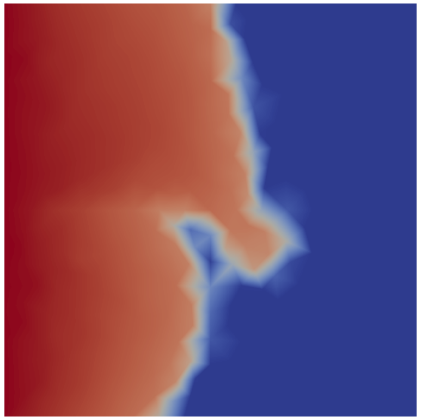
Figure 6: Simulations for the ArithMean Case (Saturation legend in Fig. 6(a) is representative for all the other figures - *i.e.*, Fig. 6(b) to (f))



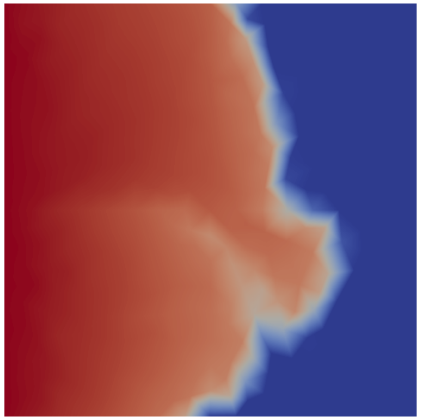
(a) Phase 1 Saturation at t=0.15s

(b) t=0.30s

(c) t=0.50s



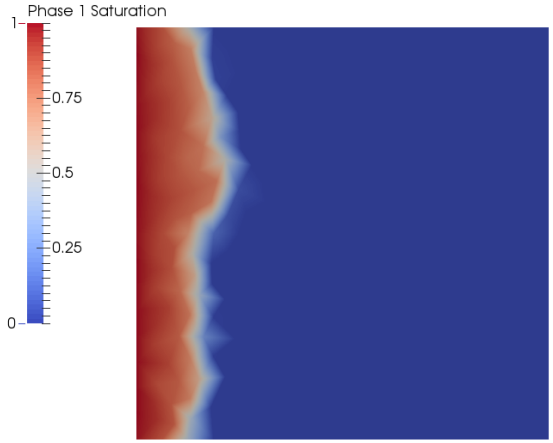
(d) t=1.15s



(e) t=1.75s

(f) t=2.95s

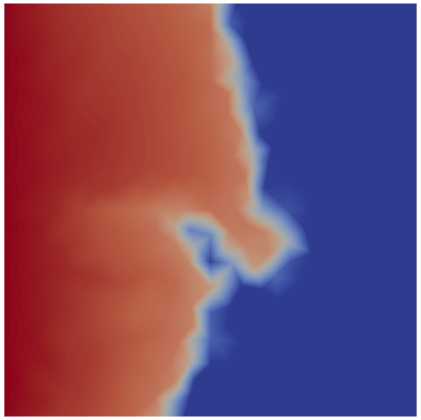
Figure 7: Simulations for the HarmMean Case (Saturation legend in Fig. 7(a) is representative for all the other figures - *i.e.*, Fig. 7(b) to (f))



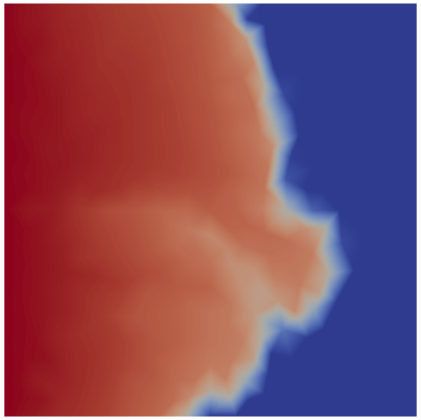
(a) Phase 1 Saturation at t=0.15s

(b) t=0.30s

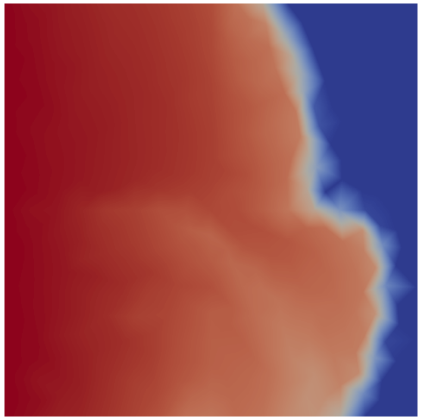
(c) t=0.50s



(d) t=1.15s



(e) t=1.75s



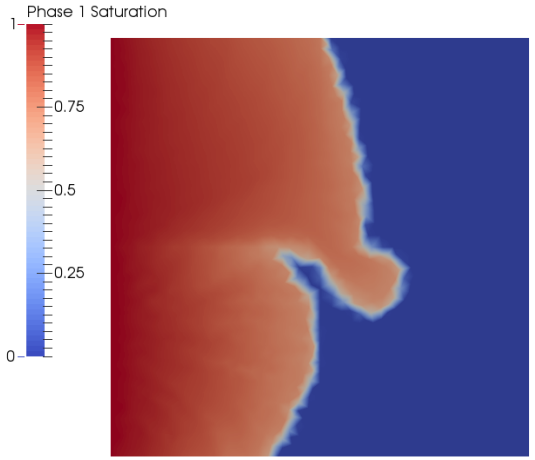
(f) t=2.95s

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))

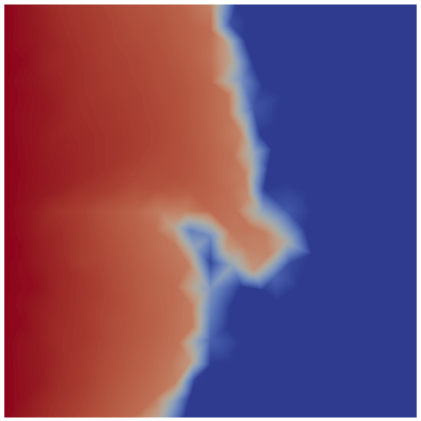
(a) Phase 1 Saturation at t=0.15s (b) t=0.30s (c) t=0.50s

(d) t=1.15s (e) t=1.75s (f) t=2.95s

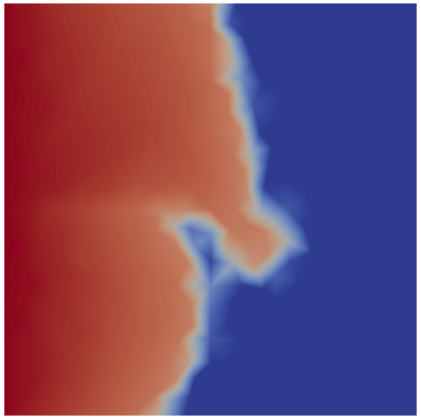
Figure 9: Simulation for the SVDCase (Saturation legend in Fig. 9(a) is representative for all the other figures - *i.e.,* Fig. 9(b) to (f))



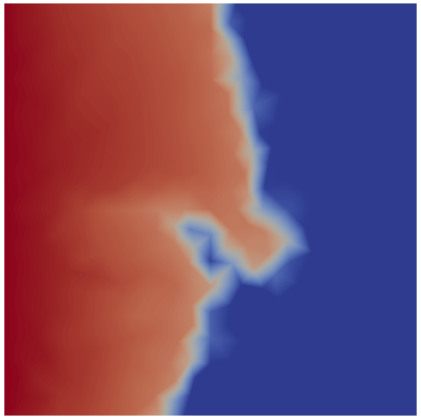
(a) BaseCase



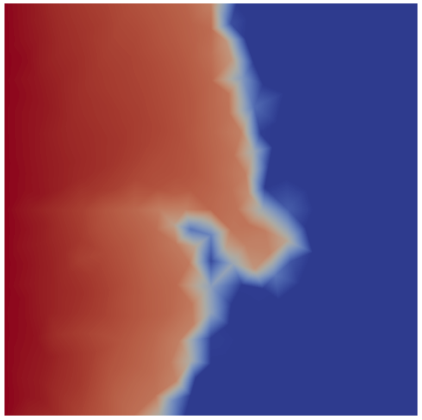
(b) ArithMeanCase



(c) HarmMeanCase

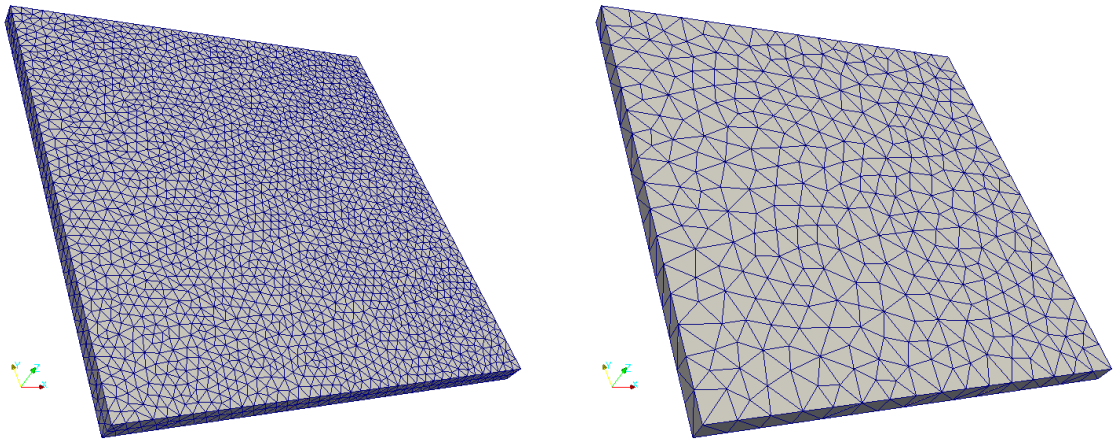


(d) PDFCase



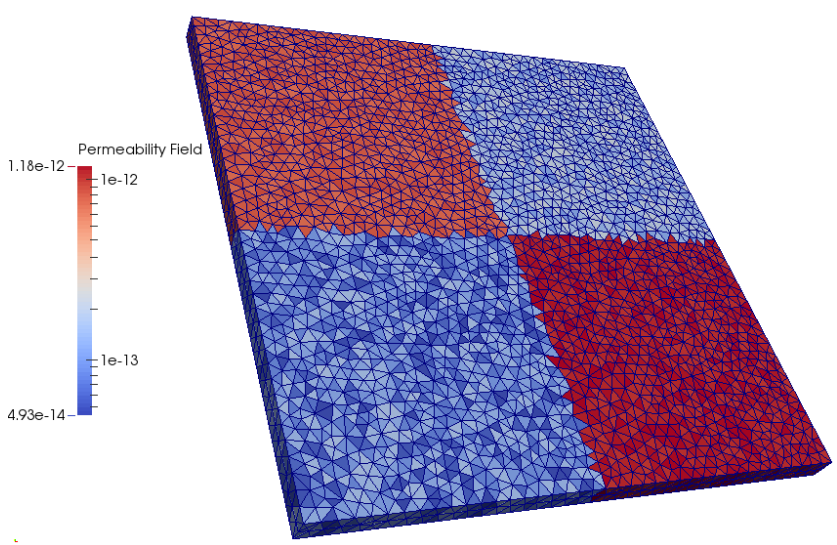
(e) SVDCase

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))

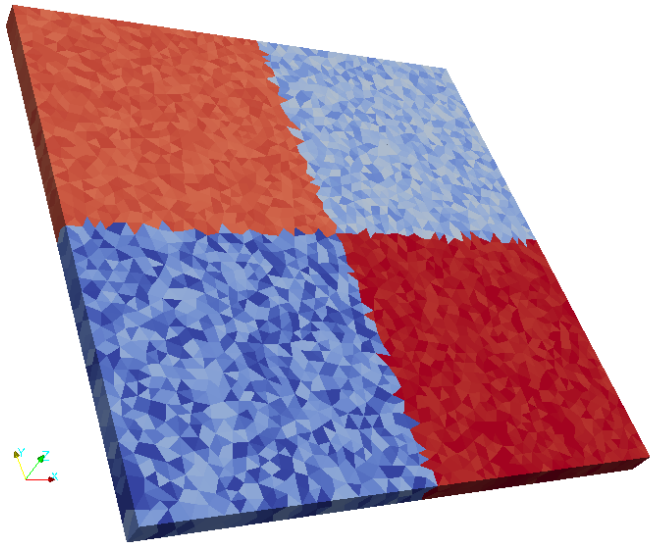


(a) 3-D High Resolution Grid (BaseCase) (b) 3-D Low Resolution Grid (Upscaled Cases)

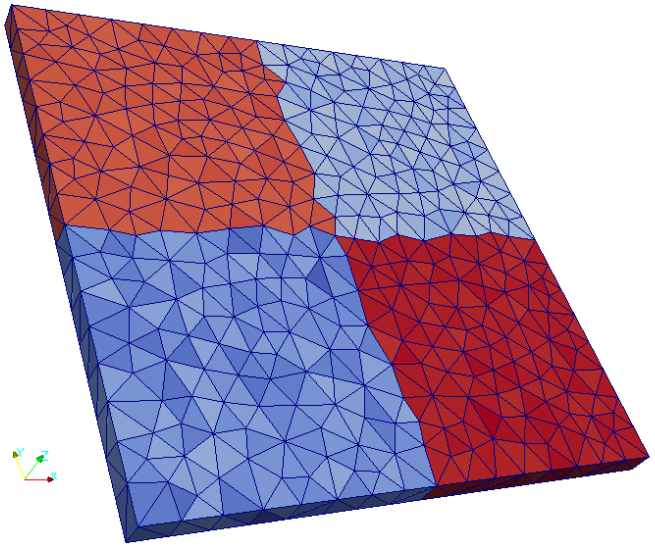
Figure 11: Mesh grid used in the performed numerical 3-D simulations: (a) high-resolution and (b) upscaled cases with 2019 and 303 tetrahedral elements respectively



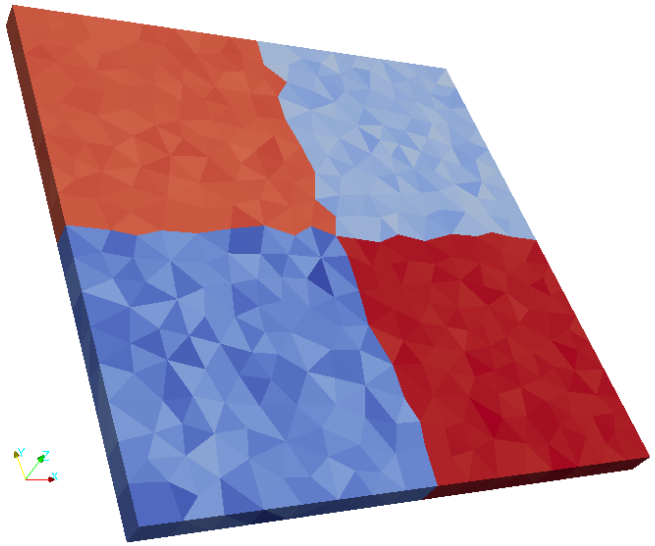
(a) 3-D BaseCase (overlapped with the mesh)



(b) 3-D BaseCase (without Mesh)

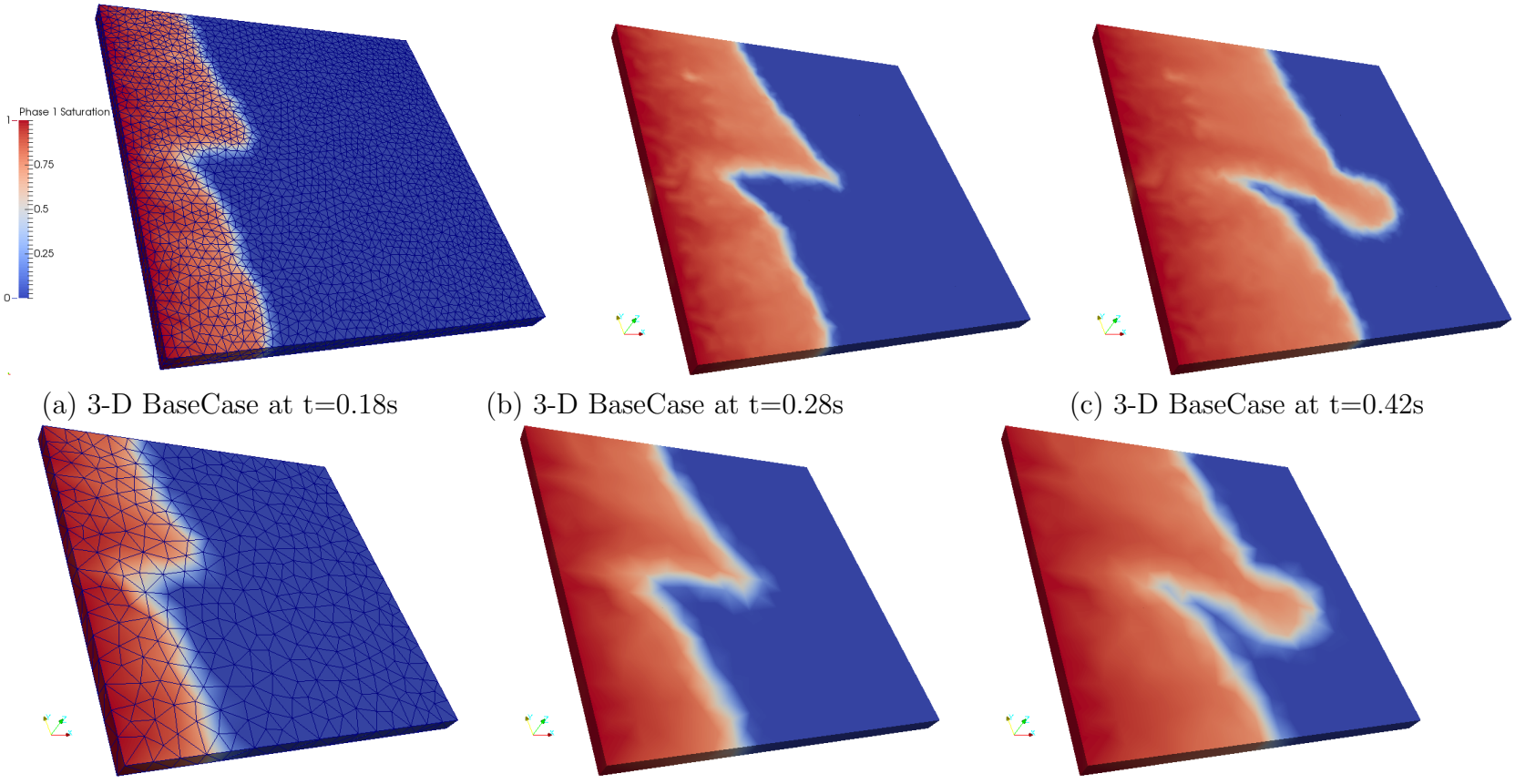


(c) 3-D HOSVD Case (overlapped with the mesh)



(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))



(a) 3-D BaseCase at t=0.18s (b) 3-D BaseCase at t=0.28s (c) 3-D BaseCase at t=0.42s

(d) HOSVD Case at t=0.18s (e) HOSVD Case at t=0.28s (f) HOSVD Case at t=0.42s

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,

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\intercal}. \tag{A.1}$$

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

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

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

$$\mathbf{U}^{\intercal}\mathbf{U} = \mathbf{V}^{\intercal}\mathbf{V} = \mathbf{I}, \tag{A.2}$$

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

$$\mathbf{U}^{-1} = \mathbf{U}^{\intercal} \quad \text{and} \quad \mathbf{V}^{-1} = \mathbf{V}^{\intercal}, \tag{A.3}$$

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

For a rectangular matrix, the singular values fill the first r places on the main diagonal of Σ. 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,

$$\mathbf{U}\Sigma\mathbf{V}^{\intercal}x = b \implies x = \mathbf{V}\Sigma^{-1}\mathbf{U}^{\intercal}b, \tag{A.4}$$

with benefit of the orthogonal property already discussed.

Appendix B.
A Brief introduction to HOSVD

Using the tensor notation from Kolda and Bader [31], given a tensor $\mathcal{X} \in \mathcal{R}^{N_1 \times N_2 \times \cdots \times N_m}$, the higher order singular value decomposition (HOSVD) is a method for factorizing $\mathcal{X}$, such that,

$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{U}^1 \times_2 \mathbf{U}^2 \times \cdots \times_m \mathbf{U}^m. \tag{B.1}$$

where tensor $\mathcal{G} \in \mathcal{R}^{N_1 \times N_2 \times \cdots \times N_m}$ 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.