This is a repository copy of *A reduced order model for sea water intrusion simulation using proper orthogonal decomposition*.

Article:

A Reduced Order Model for Sea Water Intrusion Simulation Using Proper Orthogonal Decomposition

by Mohammadali Geranmehr[1] , Domenico Bau[2] , Alex S. Mayer[3] , and Weijiang Yu[2]

Abstract

Sea water intrusion (SWI) simulators are essential tools to assist the sustainable management of coastal aquifers. These simulators require the solution of coupled variable-density partial differential equations (PDEs), which reproduce the processes of groundwater flow and dissolved salt transport. The solution of these PDEs is typically addressed numerically with the use of density-dependent flow simulators, which are computationally intensive in most practical applications. To this end, model surrogates are generally developed as substitutes for full-scale aquifer models to trade off accuracy in exchange for computational efficiency. Surrogates represent an attractive option to support groundwater management situations in which fast simulators are required to evaluate large sets of alternative pumping strategies. Reduced-order models, a sub-category of surrogate models, are based on the original model equations and may provide quite accurate results at a small fraction of computational cost. In this study, a variable-density flow reduced-order model based on proper orthogonal decomposition (POD) and utilizing a fully coupled flow and solute-transport model is implemented with a finite-difference (FD) approach for simulating SWI in coastal aquifers. The accuracy and computational efficiency of the FD-POD approach for both homogeneous and—more realistic—heterogeneous systems are investigated using test cases based on the classic Henry's problem (Henry 1964). The findings demonstrate that the combined FD-POD approach is effective in terms of both accuracy and computational gain and can accommodate the output of the most popular variable-density flow models, such as those from USGS's MODFLOW family.

Introduction

Coastal regions globally face mounting challenges in managing groundwater resources. These aquifers are increasingly stressed due to escalating demand and the compounding effects of climate change, sea-level rise,

[1]Corresponding author: Department of Civil and Structural Engineering, The University of Sheffield, Sheffield, S1 3JD, UK; mgeranmehr@gmail.com

[2]Department of Civil and Structural Engineering, The University of Sheffield, Sheffield, UK; mgeranmehr@gmail.com; d.bau@sheffield.ac.uk; weijiang.yu@sheffield.ac.uk

[3]University of Texas at El Paso, El Paso, Texas; amayer2@utep.edu

and variations in groundwater recharge patterns (Ketabchi et al. 2016; Alfarrah and Walraevens 2018; Befus et al. 2020; Panthi et al. 2022). Among the most pressing issues confronting coastal aquifers is SWI, a phenomenon where saline water infiltrates freshwater aquifers driven by density variations. This intrusion not only jeopardizes water quality but also escalates management costs and threatens the sustainability of groundwater resources.

To address the complexities of coastal groundwater systems, groundwater models have emerged as indispensable tools for water administrators and stakeholders. Models such as SUTRA (Voss 1984), MODFLOW 6 (MF6) (Langevin et al. 2017), and SEAWAT (Guo and Langevin 2002; Langevin et al. 2008), which is based on MODFLOW-2005 (Harbaugh 2005) and MT3DMS (Zheng and Wang 1999), play a pivotal role by simulating groundwater flow and solute transport. Notably, these models are considered "coupled" variable-density models, as the migration of salt affects density, consequently impacting flow dynamics, which in turn affects salt transport. This intricate relationship underpins the coupling between flow and transport processes in coastal aquifers.

Achieving solutions to the coupled PDEs underlying groundwater flow and solute transport relies on

sophisticated numerical techniques. Discretization methods, including FD, used in SEAWAT, finite elements (FE), used in SUTRA, and finite volumes (FV), used in MF6, are employed to transform PDEs into systems of ordinary differential equations (ODEs). However, the computational demands of these models, necessitating fine grid resolution and iterative coupling, present significant challenges. For example, the computational intensity and data requirements of groundwater models often hinder their application in practical management applications. Ideally, models should balance computational efficiency with accuracy to enable repeated simulations for optimization, risk analysis, and decision support. To address these challenges, surrogate models have emerged as viable alternatives, trading off accuracy for computational speed. These models can be grouped into three main categories: data-driven surrogates, multi-fidelity surrogates, and model-driven surrogates. A comprehensive review of surrogates may be found in Asher et al. (2015).

Data-driven surrogates rely on "observations" to generate predictions based on statistical or machine-learning techniques to learn patterns directly from the data. The definition of observations is somewhat loose in that these can be compiled from actual field data or generated from full-scale model training runs. For example, in the case of field observations, machine learning techniques have been widely applied to water level modeling in large-scale aquifers. These techniques include, among others, artificial neural networks (ANNs) (Coulibaly et al. 2001; Krishna et al. 2008; Mohanty et al. 2010), adaptive neuro-fuzzy inference systems (ANFIS) (Jalalkamali et al. 2010; Emamgholizadeh et al. 2014), genetic programming (Fallah-Mehdipour et al. 2013), and support vector machines (SVMs) (Rajaee et al. 2019). In the construction of surrogates for full-scale models, data-driven approaches have been widely adopted for applications in contaminant transport simulation (Shekofteh et al. 2012; Arabgol et al. 2016) and water quality simulation (Haggerty et al. 2023). These techniques have also been widely adopted for variable-density flow models, most typically in simulation-optimization approaches to assist coastal groundwater management (Kourakos and Mantoglou 2009; Sreekanth and Datta 2011; Ataie-Ashtiani et al. 2014; Roy and Datta 2017; Christelis et al. 2018; Lal and Datta 2018; Ranjbar and Mahjouri 2018; Song et al. 2018; Kopsiaftis et al. 2019; Yang et al. 2021) and also to conduct sensitivity analyses for uncertainty quantification (Rajabi et al. 2015; Koohbor et al. 2019; Rajabi 2019).

Multi-fidelity surrogate models (Asher et al. 2015; Zhou et al. 2016) are a particular category of physically based models that combine information from computationally efficient low-fidelity numerical models (i.e., with reduced numerical resolution, larger tolerance, and/or simplified physics), with "data" from higher-fidelity models to construct a surrogate that integrates both types of data. Multi-fidelity surrogates have also been applied to SWI simulation (Kerrou and Renard 2010; Christelis and Mantoglou 2016; Christelis and Mantoglou 2019;

Dey and Prakash 2020; Christelis 2021; Christelis and Hughes 2022).

Reduced-order models (ROMs) constitute a third important category of physically based surrogates, which most often rely on projection methods (Asher et al. 2015) to capture the essential modes of the system dynamics to mitigate computational complexity. One of the earlier applications of projection techniques to groundwater flow models was reported by Vermeulen et al. (2004), who used State-Space Projection and Galerkin Projection techniques based on empirical orthogonal functions (EOFs) and proper orthogonal decomposition (POD) to extract dominant modes of variability from high-dimensional data using the Karhunen–Loève expansion (Newman 1996). Numerous studies have since highlighted the versatility of POD in engineering applications (Lu et al. 2019), and particularly for subsurface flow simulation, which makes it a powerful tool for groundwater management (Mcphee and Yeh 2008). Approaches for optimizing the selection of dominant modes (Siade et al. 2010; Di et al. 2012), implementing ROMs for Monte Carlo simulation (Pasetto et al. 2011), and estimating hydraulic conductivity spatial distributions in inverse problems (Winton et al. 2011) are among the earlier reported applications of POD for developing surrogates of groundwater flow simulators.

More recently, Boyce and Yeh (2014) applied POD for parameter-independent model reduction in transient groundwater flow in confined aquifers. Boyce et al. (2015) extended this method to unconfined groundwater flow based on Galerkin projection and the Newton formulation of MODFLOW (Niswonger et al. 2011). Of related interest are the works of Stanko et al. (2016), who combined POD and discrete empirical interpolation methods to create a reduced model implemented within the MODFLOW framework, Ushijima and Yeh (2017), who developed system matrices in a parameterized approach for creating a reduced-order groundwater model, and Gosses et al. (2018), who developed POD based reduced-order models that can handle complex Dirichlet, Neumann, and Cauchy boundaries conditions. More recently, Dey and Dhar (2020) applied POD to simulate flow in randomly heterogeneous confined aquifers, Xia et al. (2020) combined POD with moment equations for Monte Carlo flow simulation, Ushijima et al. (2021) coupled POD with metaheuristic algorithms to optimize design parameters in aquifers, and Dey and Dhar (2023) integrated POD within the OpenFOAM framework (Jasak et al. 2007) for groundwater flow applications.

POD-based model reduction techniques have been applied for subsurface solute-transport modeling as well. For example, Li et al. (2013) used POD to construct ROMs of transient mass transport in heterogeneous media, Dehghan and Abbaszadeh (2018) used POD with local radial basis functions and differential quadrature, Rizzo et al. (2018) applied POD to model solute transport in heterogeneous porous media, and Stanko and Yeh (2019) used POD to model solute transport in the presence of nonlinear sorption. POD was used for nitrate transport simulation by Noori et al. (2020), who

developed PODMT3DMS, a modified version of the popular transport model MT3DMS (Zheng and Wang 1999), and by Dehghan et al. (2022), who combined POD with a finite-element (FE) reactive transport model under advection-dominated flow conditions.

In the simulation of variable-density flow in coastal aquifers, POD approaches to model reduction have received limited attention, likely due to the computational challenges posed by the coupling between the flow and the transport processes. Li and Hu (2013) applied POD to a Galerkin FE model, to simulate SWI in coastal aquifers. They used singular value decomposition (SVD) to create basis functions for reducing the model size and demonstrated that POD is a promising method for developing surrogate variable-density models.

This study aims to further investigate to application of POD-based projection methods for developing ROMs that can support SWI management. The coupling of POD with FD solution of the coupled flow and transport equation is proposed to address the computational challenges posed by the complex dynamics of variable-density flow observed in coastal aquifers. The model framework is tested by focusing on the classic Henry's problem (Henry 1964), under the hypothesis of advection–diffusion salt transport in heterogeneous systems. This structured approach aims to enhance the computational efficiency of the SWI simulation while minimizing the accuracy loss with respect to the full-scale model, thereby providing an ideal tool to assist the sustainable management of coastal groundwater resources.

This article is structured as follows: Section "Methodology" outlines the methodology devised to construct, the FD-based variable-density flow model, which is then reduced by means of the POD scheme. In section "Numerical Experiments," Henry's problem is defined and solved using the FD-POD approach, in the case of both homogeneous and heterogeneous porous media. Results of these investigations are then reported and discussed comprehensively. Finally, section "Conclusions" offers a conclusion summarizing the key findings of the study.

Methodology

Variable Density Flow Model

Variable-density groundwater flow models rely on solving two coupled PDEs. The first PDE governs density-dependent flow in saturated porous media and is written as (Langevin et al. 2008):

$$\nabla \left[\rho \mathbf{K} \left(\nabla h + \frac{\rho - \rho_f}{\rho_f} \nabla z \right) \right] = \rho S_s \frac{\partial h}{\partial t} + \theta \frac{\partial \rho}{\partial c} \frac{\partial c}{\partial t} - \rho_s q_s \tag{1}$$

where ∇ is the gradient operator (1/m), h is the equivalent freshwater head (m), ρ is the generic water density (kg/m^3), which depends on the salt concentration c (kg/m^3), ρ_f is the freshwater density (kg/m^3), ρ_s is the sink and source density (kg/m^3), q_s is the volumetric flow rate per unit volume, representing sink and source terms

(1/day), θ is the porosity, S_s is the specific elastic storage (1/m), and $\mathbf{K}$ is the hydraulic conductivity tensor (m/day).

The second PDE is the solute-transport equation, which reads (Bedekar et al. 2016):

$$\frac{\partial c}{\partial t} = \nabla(\mathbf{D}\nabla c) - \nabla(\nu c) - \frac{q_s}{\theta} c_s \tag{2}$$

where ν is the pore velocity vector (m/day), calculated as $\nu = -\mathbf{K}\nabla h/\theta$ through Darcy's law, and $\mathbf{D}$ is the hydrodynamic dispersion tensor (m^2/day). In this study, no mechanical dispersion is considered, so that $\mathbf{D}$ is a diagonal matrix that accounts solely for chemical diffusion and is obtained by multiplying the effective molecular diffusion coefficient $D*$ by the 3×3 identity matrix.

Hence the simulation of variable density flow is based on the solution of two PDEs (Equations 1 and 2), which are "coupled" through the two dependent variables h and c. The solution proposed here relies on an FD discretization over the aquifer domain, which transforms the two PDEs into the two following systems of ordinary differential equations (ODEs):

$$\mathbf{\Lambda} \cdot \mathbf{h} + \mathbf{f} = \boldsymbol{\beta} \cdot \frac{d\mathbf{h}}{dt} \tag{3}$$

$$\mathbf{\Lambda}' \cdot \mathbf{c} + \mathbf{f}' = \boldsymbol{\beta}' \cdot \frac{d\mathbf{c}}{dt} \tag{4}$$

In Equations 3 and 4, $\mathbf{h}$ and $\mathbf{c}$ are the vectors of head and concentration at the model cells, $\mathbf{\Lambda}$ and $\boldsymbol{\beta}$ are the stiffness and the capacity matrices for the hydraulic head, respectively, $\mathbf{\Lambda}'$, $\boldsymbol{\beta}'$ are the stiffness and the capacity matrices for the concentration, respectively, and $\mathbf{f}$ and $\mathbf{f}'$ represent generic sink and source terms for water and salt, respectively. The two ODE systems (Equations 3 and 4) are solved by discretizing the time-derivatives at the right-hand side by finite differences, that is:

$$\frac{d\mathbf{h}}{dt} \cong \frac{\mathbf{h}_{t+\Delta t} - \mathbf{h}_t}{\Delta t} \tag{5}$$

$$\frac{d\mathbf{c}}{dt} \cong \frac{\mathbf{c}_{t+\Delta t} - \mathbf{c}_t}{\Delta t} \tag{6}$$

where Δt represents a time step. Following an implicit scheme, the estimation of $\mathbf{h}$ and $\mathbf{c}$ at the generic time $t + \Delta t$, given their values at time t, is obtained by solving the following systems of equations:

$$\left(\frac{\boldsymbol{\beta}}{\Delta t} - \mathbf{\Lambda} \right) \cdot \mathbf{h}_{t+\Delta t} = \frac{\boldsymbol{\beta}}{\Delta t} \cdot \mathbf{h}_t + \mathbf{f} \tag{7}$$

$$\left(\frac{\boldsymbol{\beta}'}{\Delta t} - \mathbf{\Lambda}' \right) \cdot \mathbf{c}_{t+\Delta t} = \frac{\boldsymbol{\beta}'}{\Delta t} \cdot \mathbf{c}_t + \mathbf{f}' \tag{8}$$

which can be expressed in a simpler form as:

$$\mathbf{A} \cdot \mathbf{h}_{t+\Delta t} = \mathbf{b} \tag{9}$$

$$\mathbf{A}' \cdot \mathbf{c}_{t+\Delta t} = \mathbf{b}' \tag{10}$$

In Equations 9 and 10, $\mathbf{A}$ and $\mathbf{A}'$ are square matrices of size N equal to the number of FD cells, and $\mathbf{b}$ and $\mathbf{b}'$ are $N \times 1$ column vectors. The systems (Equations 9 and 10) are generally non-linear and coupled to one another due to the dependency of $\mathbf{A}$ and $\mathbf{b}$ on $\mathbf{c}_{t+\Delta t}$, and the dependency of $\mathbf{A}'$ and $\mathbf{b}'$ on $\mathbf{h}_{t+\Delta t}$. After imposing the prescribed initial and boundary conditions for the flow and the transport models, their solution is tackled by devising an iterative Picard procedure. In essence, the system flow component (Equation 9) is solved using first the concentration evaluated at the previous time step $\mathbf{c}_t$ to calculate $\mathbf{A}$ and $\mathbf{b}$. The obtained head field $\mathbf{h}_{t+\Delta t}$ is then used to update the velocity field using Darcy's law to calculate $\mathbf{A}'$ and $\mathbf{b}'$, for the transport solution (Equation 10). The obtained concentration field $\mathbf{c}_{t+\Delta t}$ is next used to update the density field, which is fed back into the system (Equation 9). This process is repeated iteratively until numerical convergence is reached for both $\mathbf{h}_{t+\Delta t}$ and $\mathbf{c}_{t+\Delta t}$. FD discretization details are presented in the Appendix A section.

Proper Orthogonal Decomposition

The POD approach relies upon converting a system of linear equations of size N to a smaller system of linear equations of size R ($R \leq N$). To apply this method, the use of orthogonal basis functions is essential. The main concept behind the POD is to separate, or decompose, the spatial and temporal behaviors of the system. The basis functions represent the spatial nature of the system, which depends on its spatial properties and are key elements for reducing the system's complexity. The temporal components can be calculated once the basis functions are identified. This decomposition is mathematically formulated by the Karhunen–Loève theorem (Newman 1996), according to which, the hydraulic head $h(x,y,z,t)$ and the concentration $c(x,y,z,t)$ are expressed as (Vermeulen et al. 2004):

$$h(x, y, z, t) = h_0(x, y, z) + \sum_{k=1}^{\infty} U_k(x, y, z) H_k(t) \quad (11)$$

$$c(x, y, z, t) = c_0(x, y, z) + \sum_{k=1}^{\infty} W_k(x, y, z) C_k(t) \quad (12)$$

where h_0 and c_0 are "steady-state" functions, $H_k(t)$ and $U_k(x, y, z)$ are time and space-dependent functions for the hydraulic head, and $C_k(t)$ and $W_k(x, y, z)$ are time and space-dependent functions for the concentration, respectively. Both sets of U_k and W_k space-dependent functions are orthogonal (i.e., the spatial integral of their product equals zero). By selecting a finite number R of representative basis functions, and discretizing h and c on the N cells of the FD grid, the hydraulic head and concentration vectors may be approximated as (Li et al. 2013):

$$\mathbf{h}(t) = \mathbf{h}_0 + \sum_{k=1}^{R} \mathbf{U}_k \cdot \mathrm{H}_k(t) = \mathbf{h}_0 + \mathbf{U} \cdot \mathbf{H}_t \quad (13)$$

$$\mathbf{c}(t) = \mathbf{c}_0 + \sum_{k=1}^{R} \mathbf{W}_k \cdot \mathrm{C}_k(t) = \mathbf{c}_0 + \mathbf{W} \cdot \mathbf{C}_t \quad (14)$$

where $\mathbf{U}$ and $\mathbf{W}$ are $N \times R$ matrices, whose columns include the basis function values at the FD grid cells, whereas $\mathbf{H}_t$ and $\mathbf{C}_t$ are $R \times 1$ column vectors including the time functions $H_k(t)$ and $C_k(t)$, where $k = 1, 2, R$.

Capturing the spatial basis functions can be achieved by constructing "snapshot sets," that is, system states ($\mathbf{h}$ and $\mathbf{c}$), evaluated using the full-scale model at different times and for different dynamic simulation scenarios for sink/source terms and/or boundary conditions. For example, in aquifers subject to pumping and spatio-temporally variable groundwater recharge, these scenarios may include different well locations, extraction rates and schedules, and different groundwater recharge patterns. Notably, the quantity of snapshots and the selection of representative simulation scenarios is crucial as they affect the quality of the basis functions, which in turn accuracy of the POD model. If "unseen" pumping schemes and recharge patterns fall within the range of behaviors represented by these scenarios, the ROM is likely to closely reproduce the results of the full-scale model. Otherwise, new scenarios are required and the ROM will need to be updated to include the corresponding snapshot sets. This flexibility is a key feature of the POD approach.

In this work, a "snapshot" represents the $N \times 1$ vectors $\mathbf{h}$ and $\mathbf{c}$ at a specific time. With T time steps, the collection of snapshots over time creates an $N \times T$ matrix for both h and c. For Q simulations, representing different dynamic simulation scenarios, snapshots are collected for each scenario, resulting in as many $N \times T$ matrices, which are then concatenated horizontally into the two $N \times M$ matrices ($M = Q \cdot T$) $\mathbf{H}_{obs}$ and $\mathbf{C}_{obs}$.

Once these matrices are assembled, the basis function can be derived using the SVD algorithm. The SVD (Golub and Kahan 1965), a widely used method for identifying important patterns and structures within data, relies on extracting and compressing information from data into linear combinations of orthogonal eigenfunctions multiplied by low-order time-dependent weights (Golub and Van Loan 2013). With the SVD, the matrices $\mathbf{H}_{obs}$ and $\mathbf{C}_{obs}$, are factorized as:

$$\mathbf{H}_{obs} = \mathbf{U}_h \cdot \mathbf{\Sigma}_h \cdot \mathbf{V}_h^{\mathrm{T}} \quad (15)$$

$$\mathbf{C}_{obs} = \mathbf{U}_c \cdot \mathbf{\Sigma}_c \cdot \mathbf{V}_c^{\mathrm{T}} \quad (16)$$

In Equations 15 and 16, $\mathbf{U}_h$ and $\mathbf{U}_c$ are orthonormal $N \times N$ matrices (with rows and columns forming a set of orthogonal unit vectors), $\mathbf{\Sigma}_h$ and $\mathbf{\Sigma}_c$ are $N \times M$ diagonal matrices containing "singular values," and $\mathbf{V}_h$ and $\mathbf{V}_c$ are orthonormal $M \times M$ matrices. $\mathbf{U}_h$ and $\mathbf{U}_c$ contain the eigenvectors of the matrices $\mathbf{H}_{obs} \cdot \mathbf{H}_{obs}^{\mathrm{T}}$ and $\mathbf{C}_{obs} \cdot \mathbf{C}_{obs}^{\mathrm{T}}$, respectively, whereas $\mathbf{V}_h$ and $\mathbf{V}_c$ contain eigenvectors of the matrices $\mathbf{H}_{obs}^{\mathrm{T}} \cdot \mathbf{H}_{obs}$ and $\mathbf{C}_{obs}^{\mathrm{T}} \cdot \mathbf{C}_{obs}$, respectively. The "singular" values in $\mathbf{\Sigma}_h$ and $\mathbf{\Sigma}_c$ are the square root of the M eigenvalues of $\mathbf{H}_{obs}^{\mathrm{T}} \cdot \mathbf{H}_{obs}$ and $\mathbf{C}_{obs}^{\mathrm{T}} \cdot \mathbf{C}_{obs}$, respectively. Ranking these eigenvalues in decreasing order allows the identification of the dominant modes of variability in the data, which is instrumental for dimensionality reduction. In practice, this is achieved

by selecting the R larger singular values ($R \leq M$), and since in Equations 15 and 16 $\mathbf{U}_h$ and $\mathbf{U}_c$ include spatial components and $\mathbf{\Sigma}_h \cdot \mathbf{V}_h{}^T$ and $\mathbf{\Sigma}_c \cdot \mathbf{V}_c{}^T$ include time components of the head and concentration snapshots, the basis function matrices $\mathbf{U}$ and $\mathbf{W}$ (Equations 13 and 14) are formed by selecting the first R columns of $\mathbf{U}_h$ and $\mathbf{U}_c$.

Variable Density Flow Reduced Order Model

The FD model reduction is achieved by substituting the head and concentration vectors $\mathbf{h}$ and $\mathbf{c}$ given by Equations 13 and 14 into the ODE systems (Equations 3 and 4), and pre-multiplying them by the transpose matrices $\mathbf{U}^T$ and $\mathbf{W}^T$, respectively. This leads to expressing these in terms of the R time functions $H_k(t)$ and $C_k(t)$ ($k = 1, \ldots, R$). Using an FD discretization for these as in Equations 5 and 6, and adopting implicit schemes as in Equations 7 and 8, leads ultimately to transforming the coupled systems (Equations 9 and 10) into:

$$\left[\mathbf{U}^T \cdot \mathbf{A} \cdot \mathbf{U}\right] \cdot \mathbf{H}_{t+\Delta t} = \mathbf{U}^T \cdot \mathbf{b} \tag{17}$$

$$\left[\mathbf{W}^T \cdot \mathbf{A}' \cdot \mathbf{W}\right] \cdot \mathbf{C}_{t+\Delta t} = \mathbf{W}^T \cdot \mathbf{b}' \tag{18}$$

Equations 17 and 18 represent a system of $2 \cdot R$ equations in $2 \cdot R$ unknowns, which needs to be solved at each time step. The computational gain of this approach stems from reducing the number of unknowns from $2 \cdot N$ in Equations 9 and 10 down to $2 \cdot R$. Once the $\mathbf{H}$ and $\mathbf{C}$ solutions are calculated, the hydraulic head and the concentration vectors $\mathbf{h}$ and $\mathbf{c}$ are estimated using Equations 13 and 14.

The ROM framework is described in Figure 1. The initial step consists of the compilation of the snapshot datasets $\mathbf{H}_{obs}$ and $\mathbf{C}_{obs}$. This step relies on the use of a full-scale variable density-flow simulation model, which, from a computational perspective, is the most expensive part of the process. Next, an SVD algorithm is applied to reconstruct the basis function matrices $\mathbf{U}$ and $\mathbf{W}$, which are then applied to solve the reduced order systems (Equations 17 and 18).

Similar to the full-scale systems (Equations 9 and 10), the reduced order systems (Equations 17 and 18) are nonlinear and coupled to one another due to the dependencies of $\mathbf{A}$ and $\mathbf{b}$ on $\mathbf{c}_{t+\Delta t}$, and $\mathbf{A}'$ and $\mathbf{b}'$ on
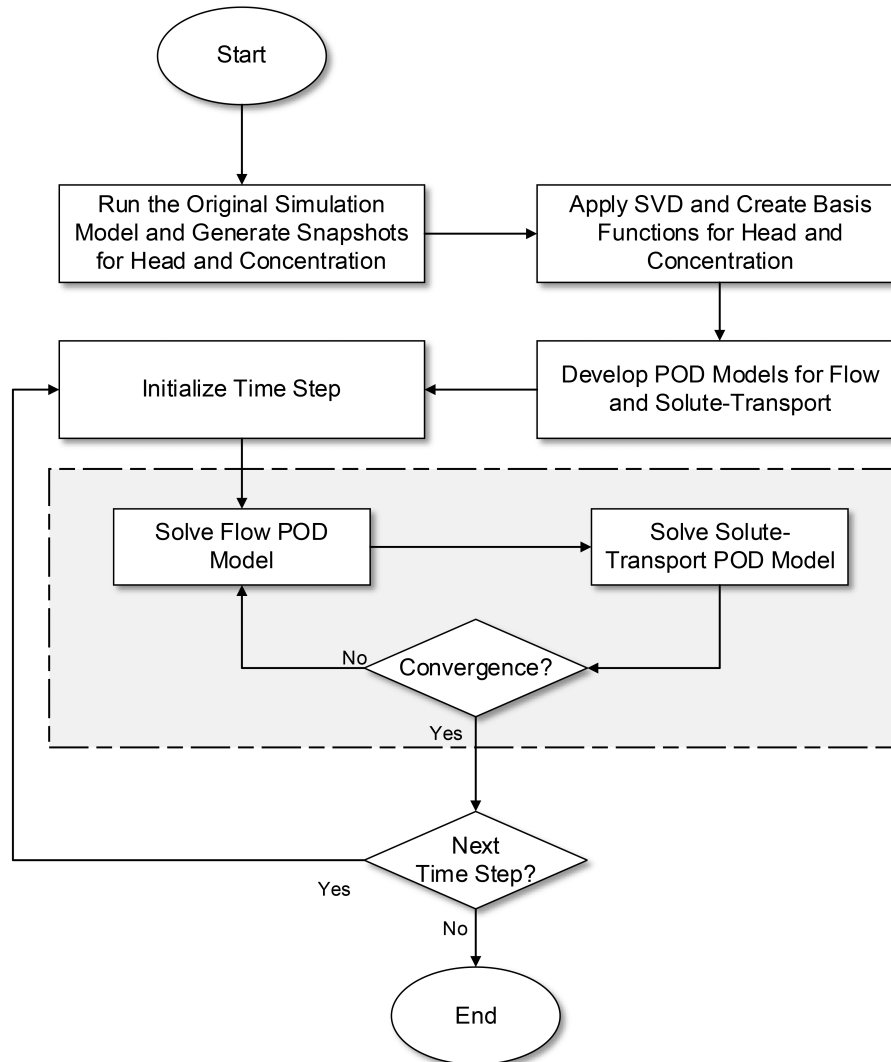

Figure 1. Flowchart for variable-density flow reduced order model.

M. Geranmehr et al. Groundwater

$\mathbf{h}_{t+\Delta t}$ their solution at any given time step is thus carried out using the Picard iteration scheme described in section "Variable Density Flow Model." Convergence is ensured through a Picard iteration scheme, where the solution is iteratively refined until the changes between successive iterations fall below a predefined tolerance (Algorithm 1). The pseudocode of the proposed model is as follows:

It is important to point that the snapshot datasets $\mathbf{h}_{obs}$ and $\mathbf{c}_{obs}$ that are used to reconstruct the basis function matrices $\mathbf{U}$ and $\mathbf{W}$ should not necessarily be assembled using the same FD variable density flow model (Equations 9 and 10) that is then used to construct the reduced order systems of Equations 17 and 18. Indeed, it is possible to use any other variable density flow model, so long as this complies with the same spatial and temporal discretization, input parameters, initial and boundary conditions, and sources and sink inputs as the

FD variable density flow model presented in section "Variable Density Flow Model."

Numerical Experiments

Henry's problem is a classical benchmark for testing SWI models. This problem (Henry 1964) considers a confined porous medium characterized by a 2×1 $(m \times m)$ rectangular domain, whose vertical cross-section is depicted in Figure 2. Flow-wise, both the top and bottom boundaries of the domain are assumed to be impermeable. At the left boundary, a constant and uniform specified flux is assumed, mimicking inland boundary conditions. At the right boundary, a uniform and constant head is prescribed, representing a hydrostatic pressure distribution at the seashore interface. Top and bottom boundaries are considered impermeable also transport-wise. The salt concentration is fixed at 35 g/L

Algorithm 1: POD Model for Variable-Density Flow

---

1 BEGIN.
2   // Step 1: Initialization.
3   Set *model*   // Initialize model domain, parameters, etc.
4   Get R // Order of POD model.
5   Get *tolerance*   // Tolerance for Picard iteration scheme.
6   // Step 2: Compute Snapshots.
7   **for** t in range(T):   // Loop through time steps.
8   **while** $|\mathbf{c}_{t+1} - \mathbf{c}_{t+1,\text{old}}| > tolerance$ // Picard iteration for snapshots.
9   get $\mathbf{h}_{t+1}$ by Equation 9   // Compute head at $t+1$.
10   set $\mathbf{c}_{t+1,\text{old}} = \mathbf{c}_{t+1}$ // Update previous concentration.
11   get $\mathbf{c}_{t+1}$ by Equation 10   // Compute concentration at $t+1$.
12   // Step 3: Compute Basis Functions.
13   get $\mathbf{U}_h$ by Equation 15   // Compute basis functions for head.
14   get $\mathbf{U}_c$ by Equation 16   // Compute basis functions for concentration.
15   // Step 4: Construct POD model.
16   **for** t in range(T):   // Loop through time steps for POD mode.
17   **while** $|\mathbf{c}_{t+1} - \mathbf{c}_{t+1,\text{old}}| > tolerance$ // Picard iteration for POD model.
18   get $\mathbf{H}_{t+1}$ by Equation 17   // Compute head time function at $t+1$.
19   get $\mathbf{h}_{t+1}$ by Equation 13   // Compute head at $t+1$.
20   set $\mathbf{c}_{t+1,\text{old}} = \mathbf{c}_{t+1}$ // Update previous concentration.
21   get $\mathbf{C}_{t+1}$ by Equation 18   // Compute concentration time function at $t+1$.
22   get $\mathbf{c}_{t+1}$ by Equation 14   // Compute concentration at $t+1$.
23   // Step 5: Results Output.
24   Print $\mathbf{h}$ and $\mathbf{c}$ // Print final head and concentration matrices.
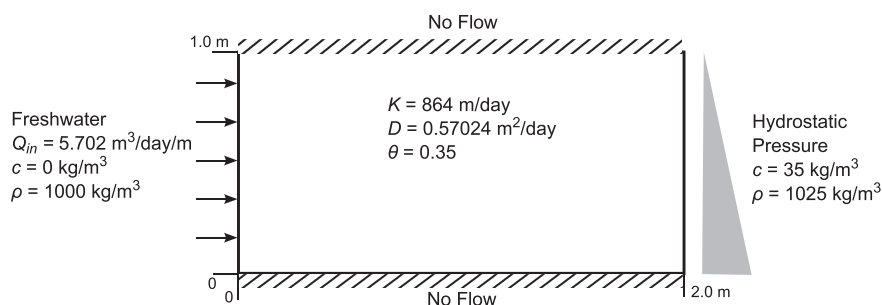25 END.

Figure 2. Henry's benchmark problem: vertical cross-section of the model domain including flow and transport boundary conditions.

at the right boundary, and to zero at the left boundary, signifying influx of freshwater.

Henry (1964) provided a semi-analytical solution for this problem, laying the foundation for understanding SWI in coastal aquifers. Henry's solution assumes salt migration is driven solely by advection and diffusion processes, excluding hydrodynamic dispersion due to heterogeneity in the hydraulic conductivity. The solution describes a steady-state density stratification, with denser salt water infiltrating below the lighter freshwater, forming a characteristic sea water wedge that mediates the advection of freshwater entering the aquifer from the left boundary with the inland diffusion of sea water originating from the right boundary, which modifies the water density.

The model parameters adopted in this work are those taken from (Guo and Langevin 2002). The domain is discretized with a uniform 2D grid of 200×100 grid blocks ($N = 20{,}000$), each of size $0.02\,\text{m} \times 0.02\,\text{m}$. The effective molecular diffusion D^* is assumed to be equal to $0.57024\,\text{m}^2/\text{d}$ and the porosity θ is 0.35.

While the Henry's problem considers steady-state conditions, the numerical tests presented in the following are conducted in transient state, with steady-state achieved at "large" time, which, for the problem at hand, is practically reached at $t = 500\,\text{min}$. In all simulations, initial conditions (at $t = 0$) assume that the freshwater hydraulic head h is equal to 1.0 m and the solute concentration C is equal to zero throughout the domain.

The snapshot datasets $\mathbf{h}_{obs}$ and $\mathbf{c}_{obs}$ (section "Proper Orthogonal Decomposition") are assembled by using the h and C model output at each time step, with $\Delta t = 1\,\text{min}$, so that both matrices have a size $N \times M$ of $20{,}000 \times 500$.

Scenario 1: Homogenous System

The first testing scenario is based on the original Henry's problem, in which the system is assumed to be homogeneous with a hydraulic conductivity K of 864 m/d. Four different solvers are applied and compared: (a) FD: the developed FD full-scale model (Equations 9 and 10); (b) FD-POD: a POD-based reduced order FD model (Equations 17 and 18) with snapshots generated using the FD full-scale model; (c) MF6: the FV-based MODFLOW 6 code; and (d) FD-POD-MF6: a POD-based reduced order FD model (Equations 17 and 18) with snapshots generated using MODFLOW 6.

Simulation results are presented in Figures 3 and 4, which display maps of the hydraulic head and the solute concentration, respectively, for the case in which $R = 10$ is selected.

Results in Figures 3 and 4 show that, at all assessed times (10, 50, 100, 500 min), the FD model (Section "Variable Density Flow Model") produces results that are close but do not perfectly match MF6's results, which are considered here as the benchmark solution. This is not surprising as the two models adopt different numerical schemes. The freshwater head from the FD and POD-based FD model match very closely. In addition, it is observed that the POD-based FD model obtained using snapshots generated by MF6, produces freshwater head distributions that match MF6's better than when
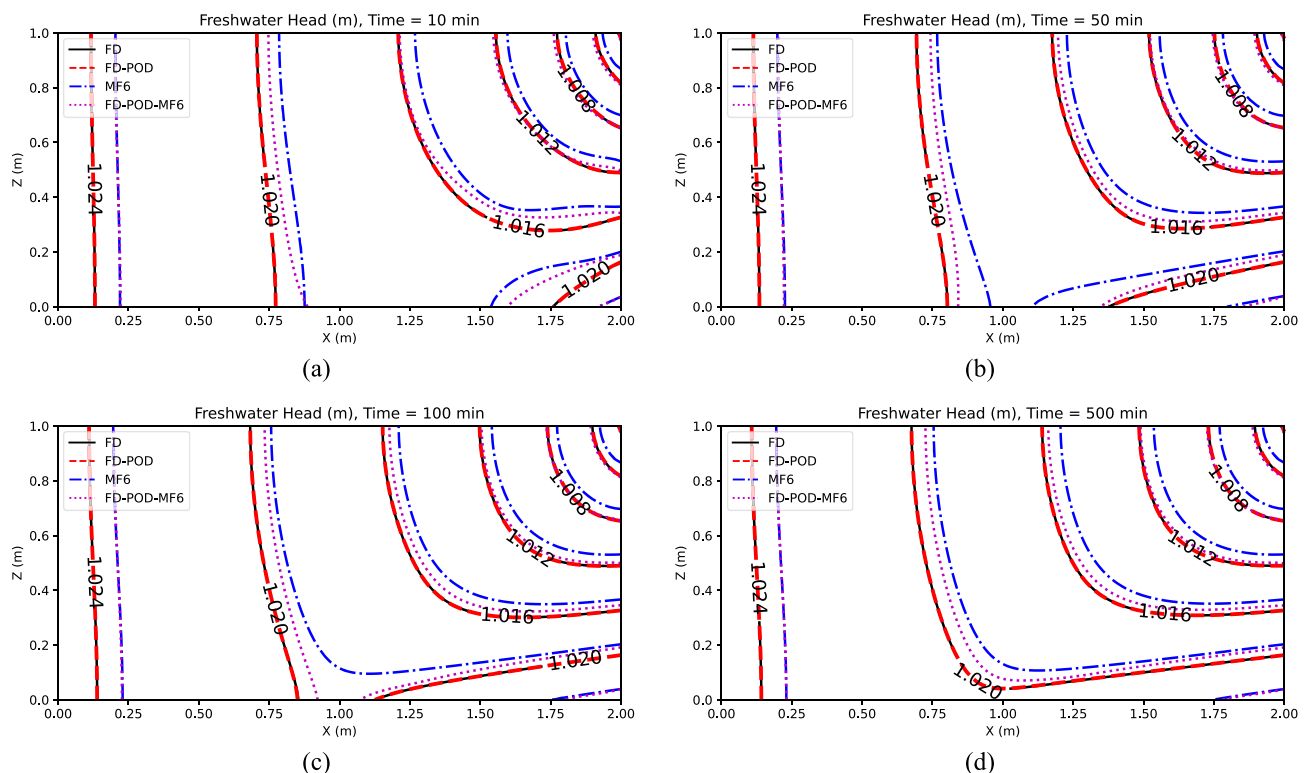

Figure 3. Freshwater hydraulic head for Henry's Problem (Scenario 1), after (a) 10 min, (b) 50 min, (c) 100 min, and (d) 500 min.
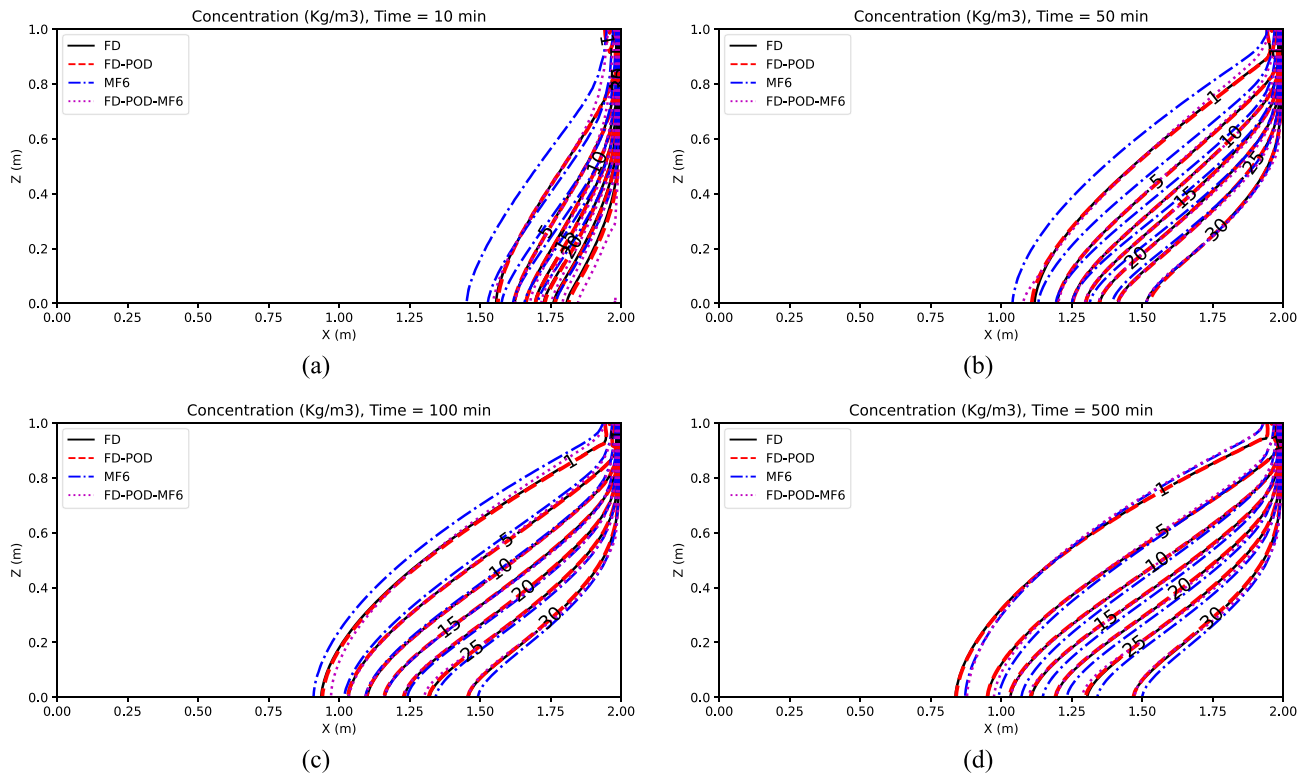
Figure 4. Concentrations for Henry's Problem (Scenario 1), after (a) 10 min, (b) 50 min, (c) 100 min, and (d) 500 min.

Table 1
Comparison of POD Models for the Original Henry's Problem at $t = 500$ min

Snapshot Generation	POD Rank R	Freshwater Head (m)		Concentration (kg/m^3)	
		r^2	RRMSE	r^2	RRMSE
FD	10	0.9999988	10.12×10^{-6}	0.9999639	7.33×10^{-6}
	20	0.9999991	8.47×10^{-6}	0.9999709	6.65×10^{-6}
	50	0.9999992	8.39×10^{-6}	0.9999716	6.53×10^{-6}
MF6	10	0.9913467	8.37×10^{-4}	0.9956631	7.94×10^{-4}
	20	0.9914066	8.34×10^{-4}	0.9958612	7.72×10^{-4}
	50	0.9918227	8.14×10^{-4}	0.9962977	7.31×10^{-4}

these snapshots are generated with the FD model. As for salt concentration, results from all four models match generally well particularly at larger times, with MF6 results being slightly higher than other models, particularly at lower times.

Table 1 summarizes model comparisons from the application of the POD-based RO approaches for increasing values of R. These comparisons are conducted in terms of the coefficient of determination r^2 and the relative root mean squared error (RRMSE) calculated for freshwater head and concentration by considering the outputs of the full-scale models versus those from the POD-based ROMs at $t = 500$ min. Results show a slight general improvement for values of R that increase from 10 to 50, indicating that selecting R equal to 10 is acceptable in this case. In terms of computational gain, the reduction in model size is observed to lead to a drastic

reduction in calculation time, which results at least one order of magnitude smaller. Because with the same R for flow and solute transport model, not equal precision is achieved, which is likely due the complexity of the solute transport PDE, in comparison with the flow PDE.

Scenario 2: Heterogeneous System

To test the POD-base model reduction approaches under more complex conditions, a heterogeneous spatial distribution of the hydraulic conductivity K is hypothesized (Figure 5). This is generated as a single realization of stochastic log-normal process characterized by an average $\mu_{\ln K} = 6.761$ (i.e., $\ln(864)$) and an anisotropic exponential covariance function with variance $\sigma_{\ln K^2} = 1$, and correlation scales $\lambda_x = 2$ m and $\lambda_z = 0.1$ m in the horizontal and vertical directions. As in Scenario 1, the solutions from four different models are compared.
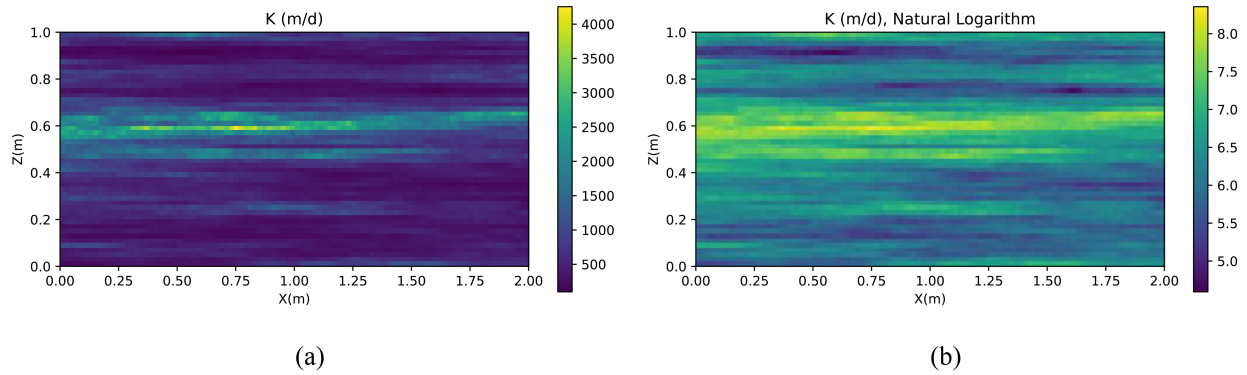
(a)

(b)

Figure 5. Heterogeneous K field, (a) exact values, (b) natural logarithmic values.
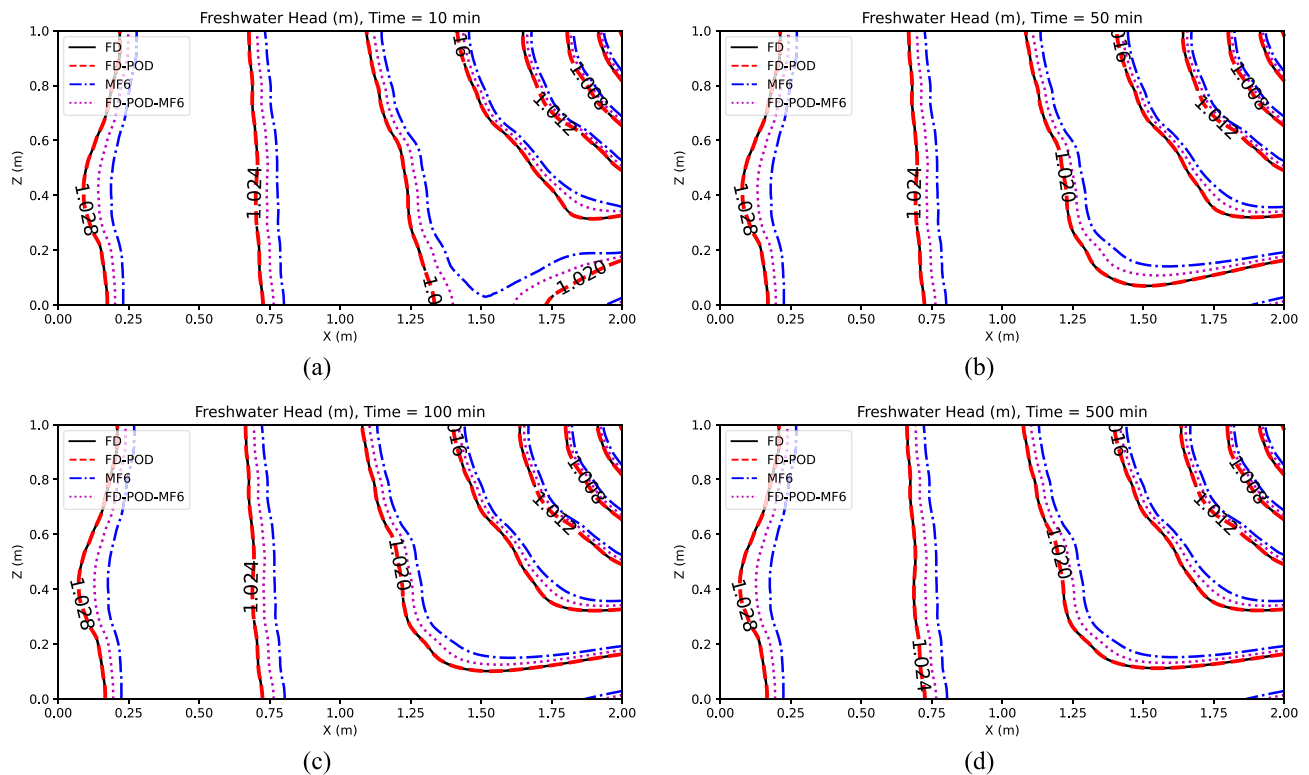

(a)

(b)

(c)

(d)

Figure 6. Freshwater hydraulic head for Henry's Problem in the case of a heterogeneous aquifer, after (a) 10 min, (b) 50 min, (c) 100 min, and (d) 500 min.

Figures 6 and 7 display model results for freshwater head and concentration, respectively. These results show a general good agreement for such a complex stochastic K field. The FD model and the FD-POD model with $R = 10$ show h and C outputs that are always practically the same but differ, however slightly, from those calculated with MF6 and the POD reduced model constructed using MF6 snapshots with $R = 10$. Such a difference is generally more pronounced for the h field than for the C field.

Table 2 displays r^2 and RRMSE calculated for h and C at $t = 500$ min by comparing the POD-based reduced order model outputs to those from the full-scale models. Results show the FD-POD model matches the FD model results quite well already with R equal to 10. The FD-POD model with snapshots generated by MF6

does not perform as well in terms of accuracy with respect to MF6, indicating that values of R of the order of 50 or larger are preferable.

Considering the more complex heterogeneous Henry's problem, the challenges associated with accurately capturing SWI dynamics become apparent. Heterogeneity introduces irregularities in the aquifer properties, requiring a larger number of POD components to accurately represent the spatial patterns. Despite this, the FD-POD approach demonstrates promising results, with high r^2 values and relatively low RRMSE values for both the freshwater head and concentration compared with the MF6 models.

Overall, these results highlight an interesting feature: by generating a set of snapshots with MF6, a POD model
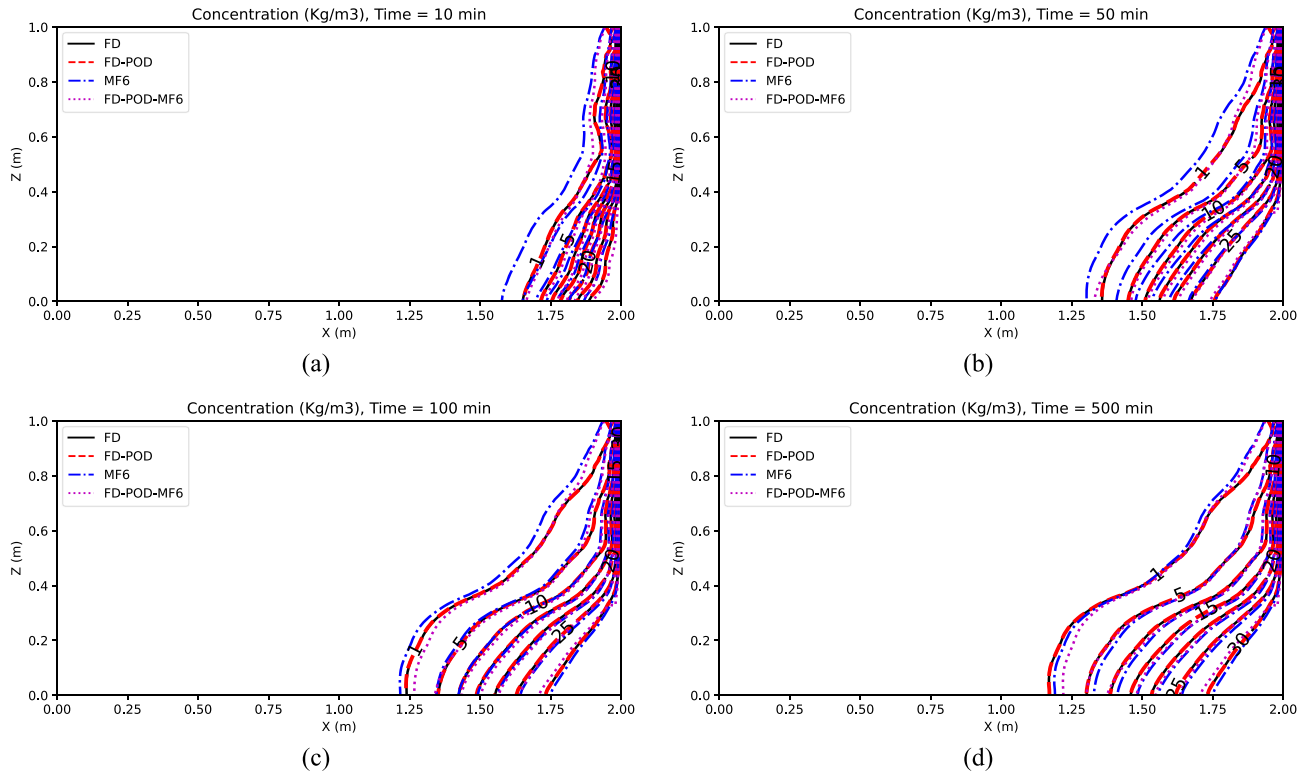
Figure 7. Concentrations for heterogeneous Henry's Problem, after (a) 10 min, (b) 50 min, (c) 100 min, and (d) 500 min.

Table 2
Comparison of POD Models for the Heterogeneous Henry's Problem at $t = 500$ min

Snapshot Generation	POD Rank R	Freshwater Head (m) r^2	Freshwater Head (m) RRMSE	Concentration (kg/m^3) r^2	Concentration (kg/m^3) RRMSE
FD	10	0.99999977	6.04×10^{-6}	0.9997752	1.25×10^{-5}
	20	0.99999983	5.12×10^{-6}	0.9999181	1.18×10^{-5}
	50	0.99999984	5.01×10^{-6}	0.9999343	1.06×10^{-5}
MF6	10	0.99589707	9.45×10^{-4}	0.9895655	13.16×10^{-4}
	20	0.99687873	8.29×10^{-4}	0.9961524	7.99×10^{-4}
	50	0.99706948	8.07×10^{-4}	0.9973246	6.66×10^{-4}

can be constructed as a faster surrogate, replacing the original MF6 model. This capability significantly reduces computational costs, making the FD-POD approach an attractive option for SWI simulations.

Discussion

The essence of the POD model lies in the reduced computational cost that can be achieved by retrieving the approximated models (Equations 13 and 14), which ultimately allows for the solution of the reduced-order systems (Equations 17 and 18) instead of the full-scale systems (Equations 9 and 10). The ROM accuracy depends ultimately on the quality of basis functions stored in the matrices **H** and **W**. Such a quality depends on the representativeness of the boundary condition scenarios considered in the calculation of the snapshot sets, referred to as

$\mathbf{H}_{obs}$ and $\mathbf{C}_{obs}$, and the reduced number R of orthonormal basis functions selected to approximate the state functions $h(x,y,z,t)$ and $c(x,y,z,t)$. While the selection of R can be somehow "automated" by selecting the higher eigenvalues (e.g., by selecting only those exceeding a prescribed threshold fraction [e.g., 0.01] of the maximum eigenvalue), the selection of representative boundary condition scenarios, depends on the complexity of the boundary conditions themselves. Nonetheless when addressing the balance between efficiency and precision, of course, one cannot run the full-scale model for any unseen boundary condition set as this would basically void the computational gain of the ROM approach. However, it seems reasonable to assess the accuracy of the POD-based model through comparison with a smaller set of full-scale simulations. For example, if the ROM was used to substitute a full-scale model in a simulation-optimization framework,

the smaller set of validating simulations could be chosen from the optimal sets of solutions or in the "surrounding" them. If the POD model is deemed to be sufficiently accurate, then it can be trusted for further simulations. Regarding "model complexity," the tests presented in our article generally show that larger R values are needed in the case of highly heterogeneous systems, but this does not appear to reduce efficiency significantly. In terms of the size of snapshot sets, adequate values should be determined based on the range of conditions (e.g., pumping rates, boundary conditions) captured in the simulations. Note that in our test only one scenario was tested (boundary conditions are constant and constant), and additional snapshots should be considered, for example, if also effects of pumping wells on SWI were to be considered.

Conclusions

In this study, a novel approach has been introduced to merge POD with FD methods to address the computational challenges associated with variable-density flow in coastal aquifers. Focusing on simulating SWI, particularly addressing Henry's problem in both homogeneous and heterogeneous media, the aim was to enhance computational efficiency while maintaining accuracy compared with full-scale models.

The significance of combating SWI challenges in coastal aquifers cannot be overstated. The proposed FD-POD approach emerged as a potent tool for groundwater management, leveraging reduced-order models to navigate the complexities of variable-density flow dynamics. By emphasizing the balance between efficiency and precision, the approach offers promising solutions for sustainable coastal groundwater resource management.

Benchmarking the FD-POD model against established methods like MF6 reveals compelling findings. The model yields results comparable to traditional methods, showcasing its efficacy in capturing the essential dynamics of density-dependent flow and solute transport. Moreover, while minor discrepancies exist, the overall computational gains outweigh these, emphasizing the model's ability to strike a balance between efficiency and accuracy.

The strengths of the POD-based reduced-order model are manifold. Its computational efficiency, evidenced by significant reductions in model size and calculation time, makes it a cost-effective solution for simulating SWI in coastal aquifers. Moreover, its accuracy ensures reliable simulations while simplifying the computational complexity of variable-density flow dynamics. For instance, when a model needs to be run several times, such as in optimization-simulation approaches, the POD presents a unique advantage. A set of snapshots could be generated with MF6 to create basis functions, and subsequently, a POD model could be constructed as a faster surrogate model, replacing the original MF6 model.

While FD-POD may provide a computationally efficient and accurate surrogate model, its effectiveness hinges on the diversity and representativeness of the snapshots used in its construction. For example, in simulation-optimization applications for groundwater management in coastal aquifers, it is crucial to generate snapshots across a broad spectrum of potential well configurations, pumping rates, and schedules. This ensures that the ROM can accurately capture the system's behavior under various operating conditions.

Practically, the model facilitates rapid decision-making in coastal groundwater management, allowing for timely responses to changing environmental conditions. Its adoption supports sustainable resource management by assessing the impact of SWI on water quality and aquifer sustainability. Moreover, it enhances decision support systems, aiding in the development of effective strategies to mitigate intrusion and protect freshwater resources.

Looking forward, several avenues for future research and development emerge. Refinement of model calibration processes could enhance accuracy under varying hydrogeological conditions. Integration of uncertainty quantification techniques could assess the impact of uncertainties on model predictions. Validation in diverse real world scenarios is imperative for understanding performance under complex real world conditions. Integration with decision support systems and exploration of multi-objective optimization techniques could further enhance practical utility and identify optimal solutions for coastal aquifer management.

In addressing these limitations and exploring future research avenues, the FD-POD approach stands poised to evolve and meet the evolving challenges of groundwater management and sustainability in coastal regions. Its potential to revolutionize SWI simulation in coastal aquifers is evident, offering a comprehensive framework for efficient and accurate decision-making in resource management.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Alfarrah, N., and K. Walraevens. 2018. Groundwater overexploitation and seawater intrusion in coastal areas of arid and semi-arid regions. *Water* 10, no. 2: 143. https://doi.org/10.3390/w10020143

Arabgol, R., M. Sartaj, and K. Asghari. 2016. Predicting nitrate concentration and its spatial distribution in groundwater resources using support vector machines (SVMs) model. *Environmental Modeling & Assessment* 21, no. 1: 71–82. https://doi.org/10.1007/s10666-015-9468-0

Asher, M.J., B.F.W. Croke, A.J. Jakeman, and L.J.M. Peeters. 2015. A review of surrogate models and their application to groundwater modeling. *Water Resources Research* 51, no. 8: 5957–5973. https://doi.org/10.1002/2015WR016967

Ataie-Ashtiani, B., H. Ketabchi, and M.M. Rajabi. 2014. Optimal management of a freshwater lens in a small island using surrogate models and evolutionary algorithms. *Journal of Hydrologic Engineering* 19, no. 2: 339–354. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000809

Bedekar, V., E.D. Morway, C.D. Langevin, and M.J. Tonkin. 2016. *MT3D-USGS Version 1: A U.S. Geological Survey Release of MT3DMS Updated With New and Expanded Transport Capabilities for Use With MODFLOW. U.S. Geological Survey Techniques and Methods. 6-A53*. Reston, VA: USGS. https://doi.org/10.3133/tm6A53

Befus, K.M., P.L. Barnard, D.J. Hoover, J.A. Finzi Hart, and C.I. Voss. 2020. Increasing threat of coastal groundwater hazards from sea-level rise in California. *Nature Climate Change* 10, no. 10: 946–952. https://doi.org/10.1038/s41558-020-0874-1

Boyce, S.E., T. Nishikawa, and W.W.G. Yeh. 2015. Reduced order modeling of the Newton formulation of MODFLOW to solve unconfined groundwater flow. *Advances in Water Resources* 83, no. 1: 250–262. https://doi.org/10.1016/j.advwatres.2015.06.005

Boyce, S.E., and W.W.G. Yeh. 2014. Parameter-independent model reduction of transient groundwater flow models: Application to inverse problems. *Advances in Water Resources* 69, no. 1: 168–180. https://doi.org/10.1016/j.advwatres.2014.04.009

Christelis, V. 2021. Surrogate-based optimization methods for coastal aquifer management. Ph.D. Thesis, National Technological University of Athens.

Christelis, V., and A.G. Hughes. 2022. Multifidelity surrogate models for efficient uncertainty propagation analysis in salars systems. *Frontiers in Water* 4, no. 1: 1–15. https://doi.org/10.3389/frwa.2022.827036

Christelis, V., and A. Mantoglou. 2019. Pumping optimization of coastal aquifers using seawater intrusion models of variable-fidelity and evolutionary algorithms. *Water Resources Management* 33, no. 2: 555–568. https://doi.org/10.1007/s11269-018-2116-0

Christelis, V., R.G. Regis, and A. Mantoglou. 2018. Surrogate-based pumping optimization of coastal aquifers under limited computational budgets. *Journal of Hydroinformatics* 20, no. 1: 164–176. https://doi.org/10.2166/hydro.2017.063

Christelis, V., and A. Mantoglou. 2016. Coastal aquifer management based on the joint use of density-dependent and sharp interface models. *Water Resources Management* 30, no. 2: 861–876. https://doi.org/10.1007/s11269-015-1195-4

Coulibaly, P., F. Anctil, R. Aravena, and B. Bobée. 2001. Artificial neural network modeling of water table depth fluctuations. *Water Resources Research* 37, no. 4: 885–896. https://doi.org/10.1029/2000WR900368

Dehghan, M., B. Hooshyarfarzin, and M. Abbaszadeh. 2022. Numerical simulation based on a combination of finite-element method and proper orthogonal decomposition to prevent the groundwater contamination. *Engineering with Computers* 38, no. 4: 3445–3461. https://doi.org/10.1007/s00366-021-01439-y

Dehghan, M., and M. Abbaszadeh. 2018. A combination of proper orthogonal decomposition–discrete empirical interpolation method (POD–DEIM) and meshless local RBF-DQ approach for prevention of groundwater contamination. *Computers & Mathematics with Applications* 75, no. 4: 1390–1412. https://doi.org/10.1016/j.camwa.2017.11.012

Dey, S., and A. Dhar. 2023. A reduced-order framework for three-dimensional-equivalent confined groundwater modeling with emphasis on well-boundary implementation. *Hydrogeology Journal* 31, no. 7: 1883–1902. https://doi.org/10.1007/s10040-023-02674-x

Dey, S., and A. Dhar. 2020. On proper orthogonal decomposition (POD) based reduced-order modeling of groundwater flow through heterogeneous porous media with point source singularity. *Advances in Water Resources* 144, no. 1: 103703. https://doi.org/10.1016/j.advwatres.2020.103703

Dey, S., and O. Prakash. 2020. Managing saltwater intrusion using conjugate sharp interface and density dependent models linked with pumping optimization. *Groundwater for Sustainable Development* 11, no. 1: 100446. https://doi.org/10.1016/j.gsd.2020.100446

Di, Z., Z. Luo, Z. Xie, A. Wang, and I.M. Navon. 2012. An optimizing implicit difference scheme based on proper orthogonal decomposition for the two-dimensional unsaturated soil water flow equation. *International Journal for Numerical Methods in Fluids* 68, no. 10: 1324–1340. https://doi.org/10.1002/fld.2610

Emamgholizadeh, S., K. Moslemi, and G. Karami. 2014. Prediction the groundwater level of Bastam plain (Iran) by artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS). *Water Resources Management* 28, no. 15: 5433–5446. https://doi.org/10.1007/s11269-014-0810-0

Fallah-Mehdipour, E., O. Bozorg Haddad, and M.A. Mariño. 2013. Prediction and simulation of monthly groundwater levels by genetic programming. *Journal of Hydro-Environment Research* 7, no. 4: 253–260. https://doi.org/10.1016/j.jher.2013.03.005

Golub, G.H., and C.F. Van Loan. 2013. *Matrix computations*. Baltimore, MD: John Hopkins University Press.

Golub, G., and W. Kahan. 1965. Calculating the singular values and pseudo-inverse of a matrix. *Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis* 2, no. 2: 205–224.

Gosses, M., W. Nowak, and T. Wöhling. 2018. Explicit treatment for Dirichlet, Neumann and Cauchy boundary conditions in POD-based reduction of groundwater models. *Advances in Water Resources* 115, no. 1: 160–171. https://doi.org/10.1016/j.advwatres.2018.03.011

Guo, W., and C.D. Langevin. 2002. *User's Guide to SEAWAT; a Computer Program for Simulation of Three-Dimensional Variable-Density Ground-Water Flow. Open-File Report 2001-434*. Tallahassee, FL: USGS. https://doi.org/10.3133/ofr01434

Haggerty, R., J. Sun, H. Yu, and Y. Li. 2023. Application of machine learning in groundwater quality modeling – A comprehensive review. *Water Research* 233, no. 1: 119745. https://doi.org/10.1016/j.watres.2023.119745

Harbaugh, A.W. 2005. *MODFLOW-2005, The U.S. Geological Survey Modular Ground-Water Model—The Ground-Water Flow Process. U.S. Geological Survey Techniques and Methods. 6-A16*. Reston, VA: USGS. https://doi.org/10.3133/tm6A16

Henry, H.R. 1964. Effects of dispersion on salt encroachment in coastal aquifers. *U.S. Geological Survey Water-Supply Paper* 1613-C, no. 1: C71–C84.

Jalalkamali, A., H. Sedghi, and M. Manshouri. 2010. Monthly groundwater level prediction using ANN and neuro-fuzzy models: A case study on Kerman plain, Iran. *Journal of Hydroinformatics* 13, no. 4: 867–876. https://doi.org/10.2166/hydro.2010.034

Jasak, H., A. Jemcov, and Ž. Tuković. 2007. OpenFOAM: A C++ library for complex physics simulations. In *International Workshop on Coupled Methods in Numerical Dynamics*. Dubrovnik: IUC.

Kerrou, J., and P. Renard. 2010. A numerical analysis of dimensionality and heterogeneity effects on advective dispersive seawater intrusion processes. *Hydrogeology Journal* 18, no. 1: 55–72. https://doi.org/10.1007/s10040-009-0533-0

Ketabchi, H., D. Mahmoodzadeh, B. Ataie-Ashtiani, and C.T. Simmons. 2016. Sea-level rise impacts on seawater intrusion in coastal aquifers: Review and integration. *Journal of Hydrology* 535, no. 1: 235–255. https://doi.org/10.1016/j.jhydrol.2016.01.083

Koohbor, B., M. Fahs, B. Ataie-Ashtiani, B. Belfort, C.T. Simmons, and A. Younes. 2019. Uncertainty analysis for seawater intrusion in fractured coastal aquifers: Effects of fracture location, aperture, density and hydrodynamic parameters. *Journal of Hydrology* 571, no. 1: 159–177. https://doi.org/10.1016/j.jhydrol.2019.01.052

Kopsiaftis, G., E. Protopapadakis, A. Voulodimos, N. Doulamis, and A. Mantoglou. 2019. Gaussian process regression tuned by Bayesian optimization for seawater intrusion prediction. *Computational Intelligence and Neuroscience* 2019, no. 1: 2859429. https://doi.org/10.1155/2019/2859429

Kourakos, G., and A. Mantoglou. 2009. Pumping optimization of coastal aquifers based on evolutionary algorithms and surrogate modular neural network models. *Advances in Water Resources* 32, no. 4: 507–521. https://doi.org/10.1016/j.advwatres.2009.01.001

Krishna, B., Y.R. Satyaji Rao, and T. Vijaya. 2008. Modelling groundwater levels in an urban coastal aquifer using artificial neural networks. *Hydrological Processes* 22, no. 8: 1180–1188. https://doi.org/10.1002/hyp.6686

Lal, A., and B. Datta. 2018. Development and implementation of support vector machine regression surrogate models for predicting groundwater pumping-induced saltwater intrusion into coastal aquifers. *Water Resources Management* 32, no. 7: 2405–2419. https://doi.org/10.1007/s11269-018-1936-2

Langevin, C.D., J.D. Hughes, E.R. Banta, R.G. Niswonger, S. Panday, and A.M. Provost. 2017. *Documentation for the MODFLOW 6 Groundwater Flow Model. Techniques and Methods. 6-A55*. Reston, VA: USGS. https://doi.org/10.3133/tm6A55

Langevin, C.D., D.T. Thorne, A.M. Dausman, M.C. Sukop, and W. Guo. 2008. *SEAWAT Version 4: A Computer Program for Simulation of Multi-Species Solute and Heat Transport. Techniques and Methods. 6-A22*. Reston, VA: USGS. https://doi.org/10.3133/tm6A22

Li, X., X. Chen, B.X. Hu, and I.M. Navon. 2013. Model reduction of a coupled numerical model using proper orthogonal decomposition. *Journal of Hydrology* 507, no. 1: 227–240. https://doi.org/10.1016/j.jhydrol.2013.09.011

Li, X., and B.X. Hu. 2013. Proper orthogonal decomposition reduced model for mass transport in heterogenous media. *Stochastic Environmental Research and Risk Assessment* 27, no. 5: 1181–1191. https://doi.org/10.1007/s00477-012-0653-2

Lu, K., Y. Jin, Y. Chen, Y. Yang, L. Hou, Z. Zhang, Z. Li, and C. Fu. 2019. Review for order reduction based on proper orthogonal decomposition and outlooks of applications in mechanical systems. *Mechanical Systems and Signal Processing* 123, no. 1: 264–297. https://doi.org/10.1016/j.ymssp.2019.01.018

Mcphee, J., and W.W.G. Yeh. 2008. Groundwater management using model reduction via empirical orthogonal functions. *Journal of Water Resources Planning and Management* 134, no. 2: 161–170. https://doi.org/10.1061/(ASCE)0733-9496(2008)134:2(161)

Mohanty, S., M.K. Jha, A. Kumar, and K.P. Sudheer. 2010. Artificial neural network modeling for groundwater level forecasting in a river island of eastern India. *Water Resources Management* 24, no. 9: 1845–1865. https://doi.org/10.1007/s11269-009-9527-x

Newman, A.J. 1996. *Model Reduction Via the Karhunen-Loeve Expansion Part I: An Exposition. Technical Research Report. T.R. 96-32*. College Park, MD: University of Maryland.

Niswonger, R.G., S. Panday, and M. Ibaraki. 2011. *MODFLOW-NWT, A Newton Formulation for MODFLOW-2005. Techniques and Methods. 6-A37*. Reston, VA: USGS.

Noori, R., F. Hooshyaripor, S. Javadi, M. Dodangeh, F. Tian, J.F. Adamowski, R. Berndtsson, A. Baghvand, and B. Klöve. 2020. PODMT3DMS-tool: Proper orthogonal decomposition linked to the MT3DMS model for nitrate simulation

in aquifers. *Hydrogeology Journal* 28, no. 3: 1125–1142. https://doi.org/10.1007/s10040-020-02114-0

Panthi, J., S.M. Pradhanang, A. Nolte, and T.B. Boving. 2022. Saltwater intrusion into coastal aquifers in the contiguous United States—A systematic review of investigation approaches and monitoring networks. *Science of the Total Environment* 836, no. 1: 155641. https://doi.org/10.1016/j.scitotenv.2022.155641

Pasetto, D., A. Guadagnini, and M. Putti. 2011. POD-based Monte Carlo approach for the solution of regional scale groundwater flow driven by randomly distributed recharge. *Advances in Water Resources* 34, no. 11: 1450–1463. https://doi.org/10.1016/j.advwatres.2011.07.003

Rajabi, M.M. 2019. Review and comparison of two meta-model-based uncertainty propagation analysis methods in groundwater applications: Polynomial chaos expansion and Gaussian process emulation. *Stochastic Environmental Research and Risk Assessment* 33, no. 2: 607–631. https://doi.org/10.1007/s00477-018-1637-7

Rajabi, M.M., B. Ataie-Ashtiani, and C.T. Simmons. 2015. Polynomial chaos expansions for uncertainty propagation and moment independent sensitivity analysis of seawater intrusion simulations. *Journal of Hydrology* 520, no. 1: 101–122. https://doi.org/10.1016/j.jhydrol.2014.11.020

Rajaee, T., H. Ebrahimi, and V. Nourani. 2019. A review of the artificial intelligence methods in groundwater level modeling. *Journal of Hydrology* 572, no. 1: 336–351. https://doi.org/10.1016/j.jhydrol.2018.12.037

Ranjbar, A., and N. Mahjouri. 2018. Development of an efficient surrogate model based on aquifer dimensions to prevent seawater intrusion in anisotropic coastal aquifers, case study: The Qom aquifer in Iran. *Environmental Earth Sciences* 77, no. 11: 418. https://doi.org/10.1007/s12665-018-7592-2

Rizzo, C.B., F.P.J. De Barros, S. Perotto, L. Oldani, and A. Guadagnini. 2018. Adaptive POD model reduction for solute transport in heterogeneous porous media. *Computational Geosciences* 22, no. 1: 297–308. https://doi.org/10.1007/s10596-017-9693-5

Roy, D.K., and B. Datta. 2017. Fuzzy C-mean clustering based inference system for saltwater intrusion processes prediction in coastal aquifers. *Water Resources Management* 31, no. 1: 355–376. https://doi.org/10.1007/s11269-016-1531-3

Shekofteh, H., M. Afyuni, M.A. Hajabbasi, B.V. Iversen, H. Nezamabadi-Pour, F. Abassi, and F. Sheikholeslam. 2012. Nitrate leaching from a potato field using adaptive network-based fuzzy inference system. *Journal of Hydroinformatics* 15, no. 2: 503–515. https://doi.org/10.2166/hydro.2012.075

Siade, A.J., M. Putti, and W.W.G. Yeh. 2010. Snapshot selection for groundwater model reduction using proper orthogonal decomposition. *Water Resources Research* 46, no. 8: 1–13. https://doi.org/10.1029/2009WR008792

Song, J., Y. Yang, J. Wu, J. Wu, X. Sun, and J. Lin. 2018. Adaptive surrogate model based multiobjective optimization for coastal aquifer management. *Journal of Hydrology* 561, no. 1: 98–111. https://doi.org/10.1016/j.jhydrol.2018.03.063

Sreekanth, J., and B. Datta. 2011. Comparative evaluation of genetic programming and neural network as potential surrogate models for coastal aquifer management. *Water Resources Management* 25, no. 13: 3201–3218. https://doi.org/10.1007/s11269-011-9852-8

Stanko, Z.P., and W.W.G. Yeh. 2019. Nonlinear model reduction of solute transport models. *Advances in Water Resources* 130, no. 1: 157–171. https://doi.org/10.1016/j.advwatres.2019.05.028

Stanko, Z.P., S.E. Boyce, and W.W.G. Yeh. 2016. Nonlinear model reduction of unconfined groundwater flow using

POD and DEIM. *Advances in Water Resources* 97, no. 1: 130–143. https://doi.org/10.1016/j.advwatres.2016.09.005

Ushijima, T.T., W.W.G. Yeh, and W.K. Wong. 2021. Constructing robust and efficient experimental designs in groundwater modeling using a Galerkin method, proper orthogonal decomposition, and metaheuristic algorithms. *PLoS One* 16, no. 8: e0254620. https://doi.org/10.1371/journal.pone.0254620

Ushijima, T.T., and W.W.G. Yeh. 2017. A proposed fast algorithm to construct the system matrices for a reduced-order groundwater model. *Advances in Water Resources* 102, no. 1: 68–83. https://doi.org/10.1016/j.advwatres.2017.01.010

Vermeulen, P.T.M., A.W. Heemink, and C.B.M. Te Stroet. 2004. Reduced models for linear groundwater flow models using empirical orthogonal functions. *Advances in Water Resources* 27, no. 1: 57–69. https://doi.org/10.1016/j.advwatres.2003.09.008

Voss, C.I. 1984. *A Finite-Element Simulation Model for Saturated-Unsaturated, Fluid-Density-Dependent Ground-Water Flow With Energy Transport or Chemically-Reactive Single-Species Solute Transport. Water-Resources Investigations Report. 84-4369*. Reston, VA: USGS. https://doi.org/10.3133/wri844369

Winton, C., J. Pettway, C.T. Kelley, S. Howington, and O.J. Eslinger. 2011. Application of proper orthogonal decomposition (POD) to inverse problems in saturated groundwater flow. *Advances in Water Resources* 34, no. 12: 1519–1526. https://doi.org/10.1016/j.advwatres.2011.09.007

Xia, C.-A., D. Pasetto, B.X. Hu, M. Putti, and A. Guadagnini. 2020. Integration of moment equations in a reduced-order modeling strategy for Monte Carlo simulations of groundwater flow. *Journal of Hydrology* 590, no. 1: 125257. https://doi.org/10.1016/j.jhydrol.2020.125257

Yang, Y., J. Song, C.T. Simmons, B. Ataie-Ashtiani, J. Wu, J. Wang, and J. Wu. 2021. A conjunctive management framework for the optimal design of pumping and injection strategies to mitigate seawater intrusion. *Journal of Environmental Management* 282, no. 1: 111964. https://doi.org/10.1016/j.jenvman.2021.111964

Zheng, C., and P. Wang. 1999. *MT3DMS: A Modular Three-Dimensional Multispecies Transport Model for Simulation of Advection, Dispersion, and Chemical Reactions of Contaminants in Groundwater Systems; Documentation and User's Guide. Contract Report SERDP-99-1*. Vicksburg, MS: U.S. Army Engineer Research and Development Center.

Zhou, Q., X. Shao, P. Jiang, Z. Gao, H. Zhou, and L. Shu. 2016. An active learning variable-fidelity metamodelling approach based on ensemble of metamodels and objective-oriented sequential sampling. *Journal of Engineering Design* 27, no. 4–6: 205–231. https://doi.org/10.1080/09544828.2015.1135236

Appendix A

Finite Difference Discretization

Considering density-dependent flow PDE as:

$$\frac{\partial}{\partial x}\left(\rho K^x \frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(\rho K^y \frac{\partial h}{\partial y}\right)$$
$$+ \frac{\partial}{\partial z}\left(\rho K^z \left\{\frac{\partial h}{\partial z} + \frac{\rho - \rho_f}{\rho_f}\right\}\right)$$
$$= \rho S_s \frac{\partial h}{\partial t} + \theta \frac{\partial \rho}{\partial c}\frac{\partial c}{\partial t} - \rho_s q_s \tag{A1}$$

By applying an FD spatial discretization Equation A1 becomes:

$$\frac{1}{\Delta x}\left(\left[\rho K^x \frac{\partial h}{\partial x}\right]_{i+0.5} - \left[\rho K^x \frac{\partial h}{\partial x}\right]_{i-0.5}\right)$$
$$+ \frac{1}{\Delta y}\left(\left[\rho K^y \frac{\partial h}{\partial y}\right]_{j+0.5} - \left[\rho K^y \frac{\partial h}{\partial y}\right]_{j-0.5}\right)$$
$$+ \frac{1}{\Delta z}\left(\left[\rho K^z \left\{\frac{\partial h}{\partial z} + \frac{\rho - \rho_f}{\rho_f}\right\}\right]_{k+0.5}\right.$$
$$\times \left. - \left[\rho K^z \left\{\frac{\partial h}{\partial z} + \frac{\rho - \rho_f}{\rho_f}\right\}\right]_{k-0.5}\right)$$
$$= \rho S_s \frac{\partial h}{\partial t} + \theta \frac{\partial \rho}{\partial c}\frac{\partial c}{\partial t} - \rho_s q_s \tag{A2}$$

where i, j, and k represent cell indices in x, y, and z directions, respectively. The index $i+0.5$ and $i-0.5$ represent a property related to the cell i to $i+1$ and $i-1$ to i, respectively. Similarly, $j+0.5$, $j-0.5$, $k+0.5$, and $k-0.5$ are defined. K^x, K^y, and K^z represent hydraulic conductivity in x, y, and z directions. Multiplying both sides by the cell volume and using an FD approximation in time with an implicit approach yield:

$$\Delta y \Delta z \rho_{i+0.5,j,k} K^x_{i+0.5,j,k} \frac{h^{t+1}_{i+1,j,k} - h^{t+1}_{i,j,k}}{\Delta x}$$
$$- \Delta y \Delta z \rho_{i-0.5,j,k} K^x_{i-0.5,j,k} \frac{h^{t+1}_{i,j,k} - h^{t+1}_{i-1,j,k}}{\Delta x}$$
$$+ \Delta x \Delta z \rho_{i,j+0.5,k} K^y_{i,j+0.5,k} \frac{h^{t+1}_{i,j+1,k} - h^{t+1}_{i,j,k}}{\Delta y}$$
$$- \Delta x \Delta z \rho_{i,j-0.5,k} K^y_{i,j-0.5,k} \frac{h^{t+1}_{i,j,k} - h^{t+1}_{i,j-1,k}}{\Delta y}$$
$$+ \Delta x \Delta y \rho_{i,j,k+0.5} K^z_{i,j,k+0.5} \left(\frac{h^{t+1}_{i,j,k+1} - h^{t+1}_{i,j,k}}{\Delta z}\right.$$
$$\left. + \frac{\rho_{i,j,k+0.5} - \rho_f}{\rho_f}\right) - \Delta x \Delta y \rho_{i,j,k-0.5} K^z_{i,j,k-0.5}$$
$$\times \left(\frac{h^{t+1}_{i,j,k} - h^{t+1}_{i,j,k-1}}{\Delta z} + \frac{\rho_{i,j,k-0.5} - \rho_f}{\rho_f}\right)$$
$$= (\Delta x \Delta y \Delta z)\left(\rho S_s \frac{h^{t+1}_{i,j,k} - h^t_{i,j,k}}{\Delta t} + \theta E \frac{\partial c}{\partial t} - \rho_s q_s\right) \tag{A3}$$

where t is the time step index, $E = \partial \rho / \partial c = 0.7143$ (Guo and Langevin 2002) and Δx, Δy, and Δz are the dimensions of each cell in x, y, and z directions. Equation A3 can be expressed as:

$$d_1 h^{t+1}_{i-1,j,k} + d_2 h^{t+1}_{i+1,j,k} + d_3 h^{t+1}_{i,j-1,k} + d_4 h^{t+1}_{i,j+1,k}$$
$$+ d_5 h^{t+1}_{i,j,k-1} + d_6 h^{t+1}_{i,j,k+1} + d_7 h^{t+1}_{i,j,k} = d_0 \tag{A4}$$

where d_0 to d_7 are as follows:

$$d_0 = -\frac{\Delta x \Delta y}{\rho_f}\rho_{i,j,k+0.5}K^z_{i,j,k+0.5}\left(\rho_{i,j,k+0.5} - \rho_f\right)$$
$$- \frac{\Delta x \Delta y}{\rho_f}\rho_{i,j,k-0.5}K^z_{i,j,k-0.5}\left(\rho_{i,j,k-0.5} - \rho_f\right)$$
$$+ \left(\theta E\frac{\partial c}{\partial t} - \frac{\rho S_s}{\Delta t}\left(h^t_{i,j,k}\right) - \rho_s q_s\right)$$
$$\times (\Delta x \Delta y \Delta z) \tag{A5}$$

$$d_1 = \frac{\Delta y \Delta z}{\Delta x}\rho_{i-0.5,j,k}K^x_{i-0.5,j,k} \tag{A6}$$

$$d_2 = \frac{\Delta y \Delta z}{\Delta x}\rho_{i+0.5,j,k}K^x_{i+0.5,j,k} \tag{A7}$$

$$d_3 = \frac{\Delta x \Delta z}{\Delta y}\rho_{i,j-0.5,k}K^y_{i,j-0.5,k} \tag{A8}$$

$$d_4 = \frac{\Delta x \Delta z}{\Delta y}\rho_{i,j+0.5,k}K^y_{i,j+0.5,k} \tag{A9}$$

$$d_5 = \frac{\Delta x \Delta y}{\Delta z}\rho_{i,j,k-0.5}K^z_{i,j,k-0.5} \tag{A10}$$

$$d_6 = \frac{\Delta x \Delta y}{\Delta z}\rho_{i,j,k+0.5}K^z_{i,j,k+0.5} \tag{A11}$$

$$d_7 = -(d_1 + d_2 + d_3 + d_4 + d_5 + d_6)$$
$$- \frac{\rho_{i,j,k}S_s}{\Delta t}(\Delta x \Delta y \Delta z) \tag{A12}$$

When applied to each of the FD grid cells, Equation A4 produces a system of equations, which may be written in matrix form as Equation 9.

The solute transport PDE with no mechanical dispersivity is written as:

$$\frac{\partial c}{\partial t} = D\left(\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2} + \frac{\partial^2 c}{\partial z^2}\right) - \frac{\partial}{\partial x}\left(v^x c\right)$$
$$- \frac{\partial}{\partial y}\left(v^y c\right) - \frac{\partial}{\partial z}\left(v^z c\right) + \frac{q_s c_s}{\theta} \tag{A13}$$

Using an FD method with an implicit approach Equation A13 is transformed to:

$$\frac{c^{t+1}_{i,j,k} - c^t_{i,j,k}}{\Delta t} = D\left(\frac{c^{t+1}_{i-1,j,k} - 2c^{t+1}_{i,j,k} + c^{t+1}_{i+1,j,k}}{\Delta x^2}\right.$$
$$+ \frac{c^{t+1}_{i,j-1,k} - 2c^{t+1}_{i,j,k} + c^{t+1}_{i,j+1,k}}{\Delta y^2}$$
$$+ \left.\frac{c^{t+1}_{i,j,k-1} - 2c^{t+1}_{i,j,k} + c^{t+1}_{i,j,k+1}}{\Delta z^2}\right)$$

$$- \frac{v^x_{i+1,j,k}c^{t+1}_{i+1,j,k} - v^x_{i-1,j,k}c^{t+1}_{i-1,j,k}}{2\Delta x}$$
$$- \frac{v^y_{i,j+1,k}c^{t+1}_{i,j+1,k} - v^y_{i,j-1,k}c^{t+1}_{i,j-1,k}}{2\Delta y}$$
$$- \frac{v^z_{i,j,k+1}c^{t+1}_{i,j,k+1} - v^z_{i,j,k-1}c^{t+1}_{i,j,k-1}}{2\Delta z}$$
$$+ \frac{q_s c_s}{\theta} \tag{A14}$$

which can be expressed as:

$$e_1 c^{t+1}_{i-1,j,k} + e_2 c^{t+1}_{i+1,j,k} + e_3 c^{t+1}_{i,j-1,k} + e_4 c^{t+1}_{i,j+1,k}$$
$$+ e_5 c^{t+1}_{i,j,k-1} + e_6 c^{t+1}_{i,j,k+1} + e_7 c^{t+1}_{i,j,k} = e_0 \tag{A15}$$

where the coefficients e_0 to e_7 are:

$$e_0 = -\frac{1}{\Delta t}c^t_{i,j,k} - \frac{q_s c_s}{\theta} \tag{A16}$$

$$e_1 = \frac{D}{\Delta x^2} + \frac{v^x_{i-1,j,k}}{2\Delta x} \tag{A17}$$

$$e_2 = \frac{D}{\Delta x^2} - \frac{v^x_{i+1,j,k}}{2\Delta x} \tag{A18}$$

$$e_3 = \frac{D}{\Delta y^2} + \frac{v^y_{i-1,j,k}}{2\Delta y} \tag{A19}$$

$$e_4 = \frac{D}{\Delta y^2} - \frac{v^y_{i+1,j,k}}{2\Delta y} \tag{A20}$$

$$e_5 = \frac{D}{\Delta z^2} + \frac{v^z_{i-1,j,k}}{2\Delta z} \tag{A21}$$

$$e_6 = \frac{D}{\Delta z^2} - \frac{v^z_{i+1,j,k}}{2\Delta z} \tag{A22}$$

$$e_7 = -\frac{2D}{\Delta x^2} - \frac{2D}{\Delta y^2} - \frac{2D}{\Delta z^2} - \frac{1}{\Delta t} \tag{A23}$$

When applied to each FD grid cell, Equation A15 generates a system of equations, which may be written in matrix form as Equation 10.