Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2018-222-RC2, 2018

Hydrology and Earth System Sciences
Discussions

Interactive comment on "A novel model for simulation of nitrate in aquifers" *by* Roohollah Noori et al.

Anonymous Referee #2

Received and published: 4 August 2018

The paper proposes a reduced-order modeling approach for the simulation of nitrate concentration in groundwater. The proposed ROM should reduce the computational burden associated with the solution of flow and transport equations, while providing accurate results of concentration. Results obtained using a full system model (MODFLOW+MT3DMS) to simulate the nitrate concentration in a relatively small aquifer in Iran (Karaj Aquifer) are compared against ROM results to prove the accuracy of the solution.

I have two main concerns that strongly reduce the impact of the proposed paper.

- The aquifer considered is an unconfined aquifer. However, the model used for the simulations, MODFLOW, is a model for confined aquifers (see eq. 1). The

unsaturated zone above the water table may strongly impact the transport of nitrates from the surface to the groundwater. Moreover, the unsaturated zone might be strongly impacted by evapotranspiration. These modelling details should be clearly stated as limitations of the proposed study. Other modelling details are missing in the description of the transport problem, in particular, the handling of the boundary conditions (Dirichlet or Neumann?), and the assigned nitrate concentrations. From Figure 4 (I assume these are modelling results, but it is not specified), the spatial distribution of the nitrate concentration seems to reflect concentration in the recharge, since I cannot see a clear impact of the transport along the flow velocities. Thus, it is fundamental to clearly show (maybe n the SI) the distribution of concentration in input and its temporal variation.

- The Authors are not using POD to assemble a reduced order model, but to build a regression model. One of the peculiarities of projection-based ROMs is that they are obtained from the physical equations governing the system. In fact, the model equations are typically projected in the space of the principal components (or along the main modes, or basis functions). In this way, the ROM can take into account changes in the input variables (e.g., changes in flow velocities or contaminant concentration) and project them in the reduced space. The method used by the Authors, instead, is not projecting the model equations. During the calibration, the Authors are simply analyzing how well the concentrations can be expressed as a linear combination of the modes: this is evident by the fact that their solution is not linked to the physical variables in input. Of course, few modes are sufficient to reproduce the modeled concentrations in this test case, especially because the spatiotemporal variations in concentration are small (see figure 4). During the projection, the coefficients of the modes, $\tau(t)$, are obtained by a regression on the previously computed coefficients. The proposed novel approach in reality is simply building a statistical spatiotemporal model that approximate past results of the physical model. While this approach may work for

short time windows and systems with small temporal variations in concentration, the proposed methodology cannot consider changes in the input variables during the projection, since the model is not directly based on the projection of the model equations. Thus, possible changes in flow velocities or in the input concentrations cannot be included as part of future scenarios. These points are not clear when reading the manuscript.

The methodology presented is simply a statistical extrapolation of future projections from the full-system model solutions. For this reason, I cannot see the novelty of the proposed approach. Title, abstract and Conclusions should be revised accordingly. The conclusions of the paper would be much stronger if the Authors could use a real ROM, based on the projection of the model equations.

In the following, please find other minor considerations. Most figures should be revised: the figure quality is low and the font size used is too small. Finally, I recommend revising the text with a mother-language English speaker.

Minor comments

Abstract

P1, L15: 'including a simple structure'
what does it means? The goal of ROMs is to pass from high dimensional systems to low systems, possibly linear.

P1, L16: The methodology presented is not 'a solution for the problem', since it does not provide a solution to the transport equation. The proposed model is simply extrapolating future projection from model results. Please revise the sentece.

P1, L18: 'The dominant modes . . . was calculate'
Replace 'was' with 'were'

P1, L17: 'ROMs developed with eigenvectors, to make predictions into the future' Not clear. This sentence should be revised, as the Authors are not using a traditional ROM-based on eigenvectors, but a statistical model.

The abstract should specify that the model used is based on a spatiotemporal regression of model results.

Introduction

P2, L22: 'a ROM based eigenvectors'
It should be 'a ROM based on eigenvectors'

P2, L25: GQSM simply does not provide dominant modes, but the application of POD to its output generates the dominant modes. This is a standard procedure, I do not see any 'modeling challenge' here.

P2 L29: There are many applications of POD to transport problems in groundwater. For example, please see:

- Li et al. 2011, 'Numerical simulation based on POD for two-dimensional solute transport problems', Applied Mathematical Modelling
- Rizzo et al. 2018, 'Adaptive POD model reduction for solute transport in heterogeneous porous media', Computational Geoscience.

2.1 Methods

P3, L19: It would be interesting to see the data (or to have a reference) showing the degradation of the groundwater quality for the studied area.

P3, L20: The observation wells are shown in Figure S2, not S1

P3, L21: in the description of the study area, it would be important to describe the sources of nitrate for this aquifer and show the available concentration data (maybe in the SI).

P4, L4: Please move the part 'mu_i is the seepage or linear pore...' at the end of the sentence, so that it appears just before the associated equation, Eq. 3.

P4, L19: which are the values of the head associated to these Dirichlet boundary conditions? Figure S1 shows that the red boundaries are not along an iso-water table line. Thus, how the imposed head varies along the boundary? As the level table drastically decreased in the past years, are these boundary conditions changing in time? These points should be clarified in the manuscript.

P4, L21: 'For the other boundaries there is no interaction between the adjacent aquifers'. It is more accurate to say that these are 'no flow boundaries conditions'.

P4, L22: 'The spatial distribution hydraulic conductivity'.
Missing 'of the'.

P4, L27: 'The evaporation from groundwater was considered negligible'. This is one of the main limitations in this modelling setup. The Authors are considering an unconfined aquifer, but they are using a model that considers only the saturated part of the aquifer (Eq. 1). Unconfined aquifers can have several meters of variably saturated

soil that strongly interact with precipitation and evapotranspiration and that should be considered for the transport of contaminants.

P5, L5: how was this initial condition determined? Which data on nitrate is available?

P5, L12: 'meteorological, hydrological, geological, hydrogeological, and water quality data'. At which frequency and spatial resolution are these data available? Which are the sources? In particular, which are the hydrological and water quality data used for the calibration and validation of the models?

P5, L28: the flow model equation (eq. 1) used the specific storage S_s. However, here the Authors calibrate the specific yield. Please, modify the text in a consiste way.

P5, L29: It is not clear how the Authors pass from the calibrated recharge to the time-varying recharge for the transient simulation.

P6, L5: it is not clear how the spatiotemporal variation of the nitrate in input is considered. Is this a fixed percentage of the input flows? Are concentration in the recharge and concentration in the domain boundaries modeled in the same way? (Dirichlet or Neumann boundary conditions?

P6, Section 2.4 I find the description of the POD and of the ROM model quite difficult to follow and with many missing details. The steps for assembling the ROM using POD are standard: collection of snapshots, computation of the principal components (or modes) by using SVD of snapshot matrix. Then, the model solution (vector of concentrations depending on time) is approximated by a linear combination of the first modes. The coefficients of this linear combination, which depend on time, are obtained by projecting the model equation (eq. 2) in the space of the principal components. In

the procedure described by the Authors, it is not clear which variables represent the snapshots of concentrations? What is, in formulas, matrix M_i, j? What are matrices Ω and $\Omega^{(i)}$? Different indices should be used to indicate snapshots/eigenvectors/modes. Please, clarify the spatial dimension of each vector.

P6, Eq. 4: variable (x) should appear on both sides of the equation. The index (i) should appear also for the eigenvector . Moreover, it would be useful to introduce an index to link the particular mode Φ to the particular eigenvector .

P6, Eq. 6: What is M_i, j in relation to Ω?

P6, Eq. 7: What is Φ_i? This notation has not been defined.

P7, Eq. 2: Here I am not sure to understand, and I think this is due to the fact that eq. (4) is not clear. The modes are represented by vectors Φ; eigenvectors by vectors . To each eigenvalue in eq. (6) is associated one eigenvector and, thus, one mode by eq. (4). Please clarify the text.

P7, Eq. 8: missing $i = 1$ in the summation

P7, L8-16: in the procedure described up to now, the Authors are not using a real ROM. In fact, they are simply projecting the concentrations from the full model space, to the space of the modes, as it is described in Eq. (7) and (8). Usually ROM models use a projected version of the model equations (Eq. 2), so that the variations in the model input data (recharge, boundary conditions) can be directly considered in the ROM future projections. Here, the real ROM model is provided by a regression model

on the time variation of the coefficient τ (P7, L13). However, there is no information on how this regression is performed and how the input data (e.g. spatiotemporal variation of recharge) are considered in the ROM.

Results

P7, L20: please specify the algorithm was used for calibration of the models.

P7, L22: the two piezometers selected to present the results (Figures 3A, 3B) are really close to the imposed boundary conditions. It would be more interesting to see results for an internal piezometer (e.g., Shah Abassi or Chaman).

P7, L23: change ‘observations well’ in ‘observation wells’.

P8, L9: Figures 4A – 4F show also that there is not much transport of nitrates along the flow direction. It seems that, at least for the temporal scale selected, nitrate concentration in the groundwater reflects the concentrations of the recharge. Is this aspect captured by the model?

P8, L16: In methods, Ω is a function of x, not a matrix.

Conclusions

P10, L5: Following the considerations done in the introduction, the main limitation of GQSMs is the computational burden. The fact that POD is needed to find the modes of the concentration is not a novelty, nor a limitation of GQSMs.

P10, L7-10: the authors are not developing a ROM here. The computed concentrations are projections of the results of the full system model along the main modes. For this reason they have a small error. But there is not a model behind.

P10, L15-16: The ROM developed is simply a statistical model based on the regression on the results of the physically based model. This is not superior to MT3DMS. Moreover, it cannot be applied for the simulation of other models input or parameters, without before re-computing the MT3DMS solutions.

Figures

Figure 1: this figure does not provide useful information to understand the geomorphology of the study area. I suggest replaing this figure by the current Fig S1, with the addiction of a small map showing the basin location within Iran, in the upper left corner.

Figure 2: please revise this figure. The flow chart is quite complex, with too many details that can be avoided. As first, the groundwater and transport part should go on left, as it constitute the first modelling part. The ROM construction should go on the right. The 'START' should be connected only with the flow and transport part, as this is the first modelling step.

- Flow and transport part: the first 6 blocks are standard steps for using MODFLOW/MT2DMS. They can be summarized in one block: 'Preparation of input data for MODFLOW and MT3DMS'. I would remove the alternative calibration for steady-state and transient flow here: they have been used one after the other: before calibration of hydraulic conductivity and recharge using steady state assumption. After calibration of specific yield. The last 3 blocks should be in the part of ROM development, and can be summarized in 'Snapshot selection and

extraction from MT3DMS results'.

- ROM part: the currently first 4 blocks can be summarized in 'Computation of principal modes associated to the snapshots'. The block 'Calculation of temporal and spatial . . .' goes after the selection of the number of modes. 'Developing the ROM for. . .' should be 'Solution of the ROM for. . .'. In the 'no alternative' after 'Satisfying the defined criteria', there should be a block 'Add one mode in the ROM space'. The arrow should connect to the block 'Computation of temporal and spatial terms of ROM . . .'.
- Verification part: what does it means 'development of the ROM?'. The ROM has been developed before, as written in the algorithm: Model Construction. Why do you need to develop it again? And how? This is not explained in the diagram.

Figure 3: together with the temporal variation of the error, it would be interesting to see, for each well, the temporal variation of the data (the Authors can use a different scale on the right of each panel). In this way the readers can better understand is the model is capturing the timing of fluctuations of the data.

Figure 4: the legend for the colorbar is too small. All maps should have the same color scale, so that only one big color bar is needed. Please specify in the caption if these maps are modeling results or interpolation of data-point

Figures SI

Figures S3, S4: missing measuring units.

Figure S5: Why the zones used to assign values of the hydraulic conductivity (in figure S3) are different from the zones used for the specific yield?

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-222, 2018.