

Water Resources Research



RESEARCH ARTICLE

10.1029/2021WR029986

Key Points:

- We present a quantitative method for model-to-experiment comparison of gas migration in porous media
- We use an Invasion Percolation model and light-transmission-based experiments for demonstration
- Comparison method is automated, objective and allows scale-dependent conclusions

Correspondence to:

I. Banerjee,
ishani.banerjee@iws.uni-stuttgart.de

Citation:

Banerjee, I., Guthke, A., Van De Ven, C. J. C., Mumford, K. G., & Nowak, W. (2021). Overcoming the model-data-fit problem in porous media: A quantitative method to compare invasion-percolation models to high-resolution data. *Water Resources Research*, 57, e2021WR029986. <https://doi.org/10.1029/2021WR029986>

Received 15 MAR 2021
Accepted 23 JUN 2021

Overcoming the Model-Data-Fit Problem in Porous Media: A Quantitative Method to Compare Invasion-Percolation Models to High-Resolution Data

I. Banerjee¹ , A. Guthke¹ , C. J. C. Van De Ven^{2,3} , K. G. Mumford² , and W. Nowak¹

¹Institute for Modelling Hydraulic and Environmental Systems (IWS)/LS3, University of Stuttgart, Stuttgart, Germany, ²Department of Civil Engineering, Queen's University, Kingston, Ontario, Canada, ³Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, British Columbia, Canada

Abstract Invasion percolation (IP) models offer a computationally inexpensive way to simulate multiphase flow in porous media, but only very few studies have compared their results to actual laboratory experimental image data. One reason might be the difficulty in quantitative assessment: IP models do not have a notion of experimental time but only have an integer counter for simulation steps that imply a time order. Previous experiments-to-model comparison studies have either used perceptual similarity or spatial moments as measures of comparison. In this work, we present an objective and quantitative comparison method that overcomes the limitations of the traditional approaches. First, we perform a volume-based time matching between real-time experiments and IP model results. Then, we evaluate the quality of fit using a diffused version of the so-called Jaccard coefficient, which is known from image recognition. We demonstrate our method's applicability on a laboratory-scale experimental video of gas injection in homogeneous, saturated sand, comparing it to a Macro-IP model's simulation results. We consider random realizations of the initial entry pressure field to capture the sand's inherent pore-scale heterogeneity. We find that our proposed method is reliable and intuitive in identifying realistic model realizations. The "strictness" of the method can be adjusted to relevant scales of interest via the blurring (diffusion) radius of the compared images. Beyond the application presented here, our comparison method can be used to compare any high-resolution space-time model output to experimental data given as raster images, thus providing valuable insights for model development in many research areas.

Plain Language Summary Computer models are becoming popular tools to gain a better understanding of processes happening in soils. Gas flow in the subsurface is relevant for many engineering problems. Models need to be compared to experimental or field data to test how well they represent the truth. For very large datasets such as time-series of images, this experiment-to-model comparison is often problematic. On the one hand, manually identifying similarities between the model and experimental images is a tedious method that can lead to errors (e.g., two researchers might arrive at two different conclusions). On the other hand, using automated "summary metrics" can lead to loss of information from the images, and again may trigger wrong conclusions. In this work, we present an experiment-to-model comparison method for raster-image type datasets that overcome these limitations. We demonstrate our method using simple, fast, and computationally cheap invasion percolation models for gas flow in soil. In our method, we use a metric to quantify perceptual similarity along with a "switch" that allows the user to control the application-specific strictness of the match. We find that our automated method is reliable and intuitive and can be used for gaining insights for model development in many research areas.

1. Introduction

Numerical models are used to simulate multiphase flow in porous media to gain further information about the systems and to make useful predictions for real-world engineering applications. All numerical models have underlying assumptions. Therefore, they need to be compared to actual experimental or field data to test, calibrate, and validate the models. Additionally, such comparisons lead to increased process understanding and therefore yield improved model formulations.

© 2021. The Authors.
This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Gas migration in the subsurface is a specific example of multiphase flow in porous media. It is an essential phenomenon for several reasons. Gas migration from leaky oil and gas wells can reduce the quality of shallow groundwater and can cause safety concerns due to the release of combustible gas and greenhouse gas emissions to the atmosphere (Vengosh et al., 2013, 2014). In the case of subsurface carbon dioxide storage, gas migration through fractures in cap-rock can not only contaminate shallow aquifers (Sakaki et al., 2013) but also decrease the storage efficiency. Similar is the concern for geological storage of hydrogen (Woods & Norris, 2016). For the example applications mentioned above, knowledge of gas flow mechanisms in the subsurface is necessary to establish monitoring or investigative networks. Besides, in the remediation of volatile organic compounds (VOCs) and non-aqueous phase liquids (NAPLs) in the subsurface, methods like in-situ air-sparging (Brooks et al., 1999) or thermal remediation (Hegele & Mumford, 2014) require an understanding of gas migration behavior for efficient capture and removal of contaminants. The probable flow path of gas and the degree of contact between gas and water in the subsurface is essential to be known for all these applications. However, it is difficult to predict such gas migration paths because small-scale variations in porous media govern the processes involved in gas migration.

Both continuum modeling and stochastic rule-based modeling approaches are used to describe such multiphase flows in porous media systems. Continuum models are a near-accurate representation of the physical, real-world system, but they are complex, time-consuming and computationally intensive. On the other hand, stochastic rule-based models are a simplified representation of reality, based on many assumptions. They are simple, fast, and computationally cheap.

Perturbations at the pore scale highly influence gas migration processes. The different length scales involved in these processes increase the difficulty to use the already computationally expensive continuum models for such systems (Glass et al., 2001). Undoubtedly, a fine-mesh discretization or multi-scale schemes are required to incorporate pore-scale heterogeneities correctly for these models, but this increases the computation time even further. If coarser mesh discretizations are used to decrease computational effort, artifacts of numerical dispersion smooth out the pore-scale heterogeneities (Geistlinger et al., 2009; Glass et al., 2001). Thus, the prediction capability of such models for gas migration processes is reduced, and it contradicts their claim to be “physically correct”.

Several stochastic rule-based models for the different regimes of immiscible multiphase flow in porous media were suggested in the works of Lenormand et al. (1988). Out of those, Invasion Percolation (IP) models with a representation of gravity forces have been used for upward migration of a low-density fluid (like gas) displacing a high-density fluid (like water) in a porous medium (Frette et al., 1992). Many studies have used IP models with variations to model different types of gas flow systems in porous media (e.g., Ewing & Berkowitz, 2001; Glass et al., 2001; Mumford et al., 2010, 2015; Wagner et al., 1997). Out of these variations, Macro-IP models (e.g., Glass et al., 2001; Mumford et al., 2015) conceptualize the porous medium as a group of internally homogeneous isotropic blocks consisting of sub-networks of pores instead of individual pores as in pore-scale IP models that consider individual pores (e.g., Wagner et al., 1997; Wilkinson & Willemssen, 1983). Macro-IP models are a reasonable choice for scales greater than or equal to the experimental data used in this study because using a pore-scale IP model would require a very high computational effort for typical sandbox experiments. Hereafter, we will use the term “IP-type models” to refer to the entire class of IP model variations. These models’ simplicity and low computational cost make them an ideal choice for describing such systems stochastically, which allows quantifying uncertainties at bearable computing times.

While IP-type models have the advantages mentioned above, it is challenging to calibrate, test or validate them for two reasons:

1. There is no time parameter in IP-type models but only a model loop counter. This unawareness of the model towards time makes it hard to compare the models to real-time experimental or field data. Also, the missing time parameter makes it challenging to use IP-type models along with other models like reactive transport models, which are important for subsurface carbon dioxide storage application (Oldenburg et al., 2016).
2. Metrics for quantitative comparison of IP-type models to actual experimental or field data are lacking. Various authors have presented many IP-type models (e.g., Berkowitz & Ewing, 1998; Glass et al., 2001; Kueper & McWhorter, 1992; Tsimpanogiannis & Yortsos, 2004; Wagner et al., 1997; Wilkinson &

Willemsen, 1983; Wilkinson, 1984). However, only a few studies have compared their models to actual (field or experimental) data. In the few studies of model-to-experimental data comparison, the comparison has been made either based on perceptual similarity (like the shape of gas clusters, channels or pools) (Glass et al., 2001; Mumford et al., 2010; Trevisan et al., 2017) or by comparing spatial moments (Mumford et al., 2015). A perceptual comparison is subjective, very time consuming due to its manual character, and maybe deceitful, and when comparing moments, there is a loss of information from the data due to the aggregation of detailed images to just a few summary pictures. Also, when using several moments, it is unclear how to combine them into a unique metric.

Our research aims to develop and demonstrate a comparison method for IP-type models and experimental data that solves the time-matching problem and is automated, objective, and quantitative. We propose a volume-based time matching followed by evaluating a metric: the Jaccard coefficient from image recognition, to quantify the degree of similarity between experimental and model images. While having IP-type models with their specific characteristics in mind, the proposed metric, in fact, allows for quantitative comparison of arbitrary model results to high-resolution data. Hence, our contribution generalizes to any highly spatially (and temporally) resolved model predictions in various disciplines.

The presentation of our method is structured as follows. Section 2 discusses the traditional approaches of comparison for IP-type models to experiments (Sections 2.1 and 2.2) and describes our comparison method in detail (Sections 2.3 and 2.4). To demonstrate the applicability of our method, we use it to compare Macro-IP model simulation data to gas-injection experiments in homogeneous sand in Section 3. This demonstration helps us assess our method's strengths and weaknesses, discussed in Section 4. Section 5 summarizes our research findings and provides a future scope of research.

2. Methods of Comparison

It is often difficult to obtain direct measurements in the subsurface to ascertain the behavior of the phases. So, indirect measurements using laboratory experiments in a representative porous medium are often used as an alternative. Non-invasive imaging methods like optical imaging using UV or visible light, dual-energy gamma radiation, X-ray microtomography, and magnetic resonance imaging are a few popular choices for such laboratory experiments (Oostrom et al., 2007; Werth et al., 2010). The time problem of IP-type models, as discussed above, is traditionally handled in one of the following ways:

1. The experimental image and the model output image are compared at characteristic/specific time points (like breakthrough time) (Birovljev et al., 1991; Glass et al., 2001; Mumford et al., 2015; Trevisan et al., 2017).
2. When an IP-type model is used in combination with a continuum model for transport (heat or mass), that transport controls the time scale, and the continuum model's time steps are used to compare to the real-time experimental data (Molnar et al., 2019; Mumford et al., 2010).

This section introduces and discusses existing methods for comparing IP-type models to experimental data via imaging methods (like perceptual comparison and spatial moments comparison). After that, we describe our proposed comparison method in detail, with an optional time-matching step aiming to solve the time problem associated with IP-type models. The proposed metric for quantitative comparison of model results to high-resolution data itself is general and could also be applied to arbitrary other types of models.

2.1. Perceptual Comparison

Perceptual comparison is the method of visually comparing experimental and model data for similarities. For example, Birovljev et al. (1991) compared the width of the fronts between the two phases from their experiments and IP simulations. In the work of Glass et al. (2001), the length of gas clusters, pool height, as well as saturation distribution images from injection experiments and IP model results were perceptually compared. Trevisan et al. (2017) compared the experimental image against a combination of model realizations, indicating the probabilities of gas position. Molnar et al. (2019) visually compared the structure of gas fingers from their Macro-IP simulations to the experiments of Hegele and Mumford (2014).

This method's primary advantage is that no pixel-based information is lost from the highly resolved data because the images are visually compared on a pixel-to-pixel basis. Also, multiple global attributes of an image (width of a finger, tortuosity and so on) from simulation and the experiment can be compared simultaneously. A valuable (but immeasurable) strength of the method is that it intuitively applies the user's expert knowledge to the judgment. However, the method of perceptual comparison is non-quantitative. Although it includes no computational effort, with the increase in the number of images to compare visually, the task may require an enormous effort from the user. Thus, it can become time-consuming, cumbersome, and non-objective.

Consider a situation where a user compares the breakthrough image (image when the invading phase has percolated across the entire defending phase-saturated porous medium) from the experiment and an IP type model (e.g., Macro-IP) visually. Initially, the user compares only two images. Now, since IP type models rely on stochastic simulation, the user runs 1,000 entry pressure realizations for the model to determine the near-accurate entry pressure field as in the experiment. The number of images to compare increases from 2 to 1,001. Next, for this Macro-IP model, the user wishes to calibrate model parameters like the saturation of the defending phase and porosity and so decides to use saturation values and porosity values from 0 to 1 in increments of 0.1 excluding 0 and 1. Thus, for each entry pressure field for the model, the user runs the model 18 times to fit these two parameters. Now, the user has to compare 1.8×10^4 model images to 1 experimental image. The user is further interested in comparing various versions of the IP model (say four different versions) including the calibration of the two parameters mentioned above for each model version. Now, the user visually compares 7.2×10^4 model images against one experimental image. Worse still, the user intends to visually inspect all images from the time of injection until the breakthrough. We realize: the effort of comparison amplifies extremely. In short, perceptual comparison can be too tedious and subjective for stochastic analysis and also where the experimental data is spatially and temporally intensive.

2.2. Comparison of Spatial Moments

Spatial moment methods have often been used to quantitatively describe both experimental data and numerical simulation output for multiphase flow in porous media (e.g., Essaid & Hess, 1993; González-Nicolás et al., 2017; Jawitz et al., 2003; Kueper & Gerhard, 1995; Pantazidou & Liu, 2008; Trevisan et al., 2015, to name just a few). This method was used to compare Macro-IP models and gas-injection experiments in the work of Mumford et al. (2015). In a 2D water-saturated porous medium invaded by a gas phase, the method involves the calculation of zeroth (M_{00}), first (M_{10} , M_{01}) and second moments (M_{20} , M_{02}) to describe the spatial distribution of the gas:

$$M_{ij} = \int_{z=-\infty}^{\infty} \int_{x=-\infty}^{\infty} \phi \rho_g S_g(x, z) x^i z^j dx dz \quad (1)$$

Here, ϕ is the porosity, ρ_g is the density of the gas, S_g is the gas saturation value, x and z are horizontal and vertical dimensions in the 2D space. These moments represent the mass, position and spread of the gas, respectively. Further, the centroid position of gas (X_c , Z_c) and its spatial extent as variance (σ_{xx}^2 , σ_{zz}^2) are calculated from these moments:

$$\begin{aligned} X_c &= \frac{M_{10}}{M_{00}} & \sigma_{xx}^2 &= \frac{M_{20}}{M_{00}} - X_c^2 \\ Z_c &= \frac{M_{01}}{M_{00}} & \sigma_{zz}^2 &= \frac{M_{02}}{M_{00}} - Z_c^2 \end{aligned} \quad (2)$$

Theoretically, the spatial moments of experimental gas saturation data should be reproducible by the model simulation output.

The method of spatial moments comparison is quantitative and objective. Also, it requires less time and human effort than the perceptual comparison method. This method's main disadvantage is the loss of information from the data (for images: pixel information) by the aggregation to a handful of summary statistics before comparison, which can be relevant for process understanding.

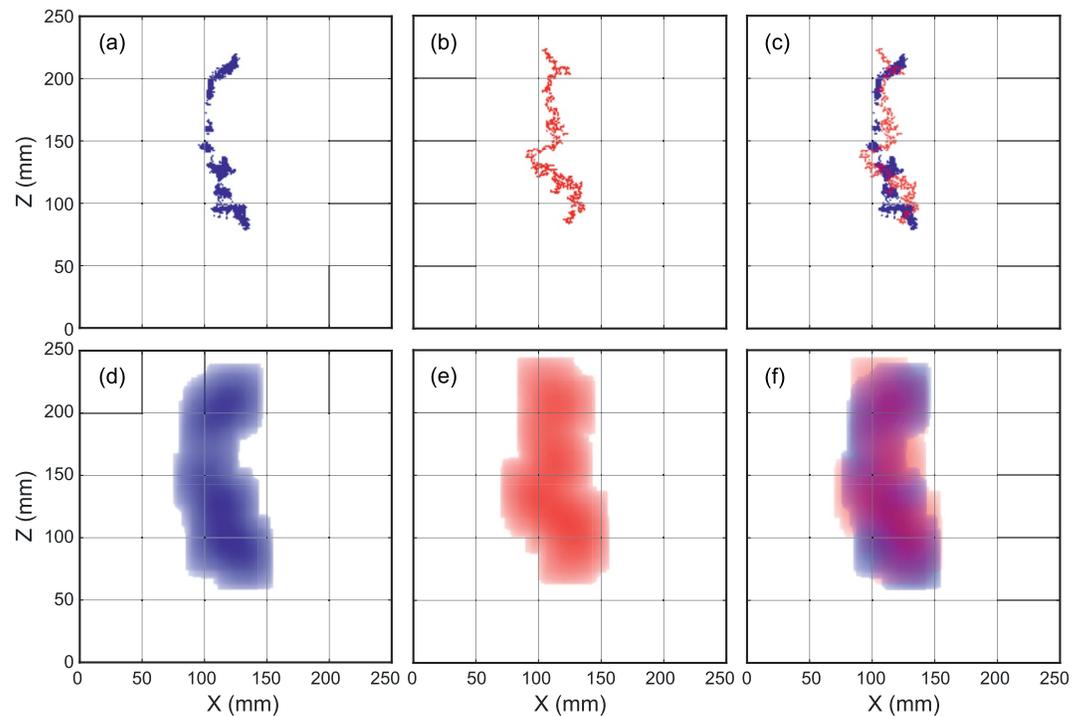


Figure 1. Visualization of intersection evaluated in the (Diffused) Jaccard coefficient through exemplary images. (a) Experimental image; (b) time-matched model realization; (c) intersection of images (purple) evaluated by Jaccard coefficient (here: $J = 0.12$); (d) and (e) corresponding blurred experimental and model images (4% domain size blur radius); (f) intersection of blurred images (purple) evaluated by Diffused Jaccard coefficient (here: $J_d = 0.54$). The experiment and the model are explained in Sections 3.1 and 3.2, respectively.

Also, in 2D space, we evaluate five spatial moment values, which individually comment on the quality of fit. However, there exists no standard method to combine the information from all of them in one unique metric to identify the best-fit realization. One reason for this is the difference in magnitude of the different spatial moment values. Moreover, for any arbitrary choice of combination of the spatial moment values, the conclusions obtained would be ambiguous. For example, any combined-moment metric would produce the best possible value for an exact mirrored copy of the experimental image (same first and second moments). When interested in predicting preferential flow pathways of a phase, identifying a mirrored image might be completely misleading. Due to this reason, the comparison of spatial moments does not enable any unique decision between competing models or in the calibration of models. In Section 4, we will further show how the loss of information of pixel data in the moments method can yield misleading results.

2.3. Proposed Metric of Comparison: (Diffused) Jaccard Coefficient

We propose to compute the Jaccard coefficient (Tan et al., 2005) to quantify the similarity between experimental and simulated images. The metric is widely used in the context of object identification or image recognition. Per set theory, it is defined as the size of the intersection between two sets A and B divided by the size of the union of these sets.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

The Jaccard coefficient ranges between zero and one, with zero meaning no similarity and one meaning complete similarity. To understand how we compute the Jaccard coefficient for binary (black/white) images, consider Figure 1 (top row) as a schematic illustration of the intersection between the binary-color coded (colored block = 1, white block = 0) data sets “experimental image” (Figure 1a) and “model image”

(Figure 1b) without taking into account the experimental setup at the moment (this will be explained in Section 3.1). The intersected blocks ($|A \cap B|$) are colored purple (Figure 1c). We count the number of purple blocks and divide by the total number of colored blocks ($|A \cup B|$) in the images of Figures 1a and 1b without double counting the pixels that agree to be colored in both of them (purple blocks of Figure 1c), to compute the Jaccard coefficient. We can automate this to be calculated for a time-series of images.

A pixel-by-pixel comparison would reject even a perfect model in a scenario where the point of gas injection in the experiment is not precisely known, leading to an offset between experiment and model. In many real-world applications, this offset would be of no concern; on the contrary, one would wish to identify such a perfect model run. Thus, we propose to blur the images from both the experiment and the model before computing the Jaccard coefficient. Since the blurring produces non-binary values, we use a slightly adapted implementation of the Jaccard coefficient for sets $A = \{a_i : a \in R, i = 1, 2, \dots, n\}$ and $B = \{b_i : b \in R, i = 1, 2, \dots, n\}$, which is also known as the Ruzicka similarity coefficient (Deza & Deza, 2016):

$$J_d(A, B) = \frac{\sum_i^n \min(a_i, b_i)}{\sum_i^n \max(a_i, b_i)} \quad (4)$$

We call this metric J_d the *Diffused Jaccard coefficient*. We illustrate its evaluation in the bottom row of Figure 1.

We blur the images using the 2D-Gaussian blur function:

$$G(x, z) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+z^2}{2\sigma^2}} \quad (5)$$

Mathematically, the blurred images are produced by convolution with a Gaussian kernel of specified width (standard deviation σ). Hence, the radius of blurring is increased or decreased by altering the σ value in Equation 5. We use the kernel size relative to the original domain size as a unit for the images' blur-radius. The blurring results in non-binary images (see Figures 1d–1f) with values between 0 and 1. These represent a spatial tolerance in matching and do not refer to intermediate values of gas saturation.

2.4. Volume-Based Time Matching for IP-Type Models

Before using our proposed metric on IP-type models, we need an additional step of time matching because of the lack of real-time notion. That means we need to know which images of the experimental time-series and the model iterations to compare with each other. So, we perform a volume-based time-matching between the experiments and the IP-type models as a first step. We use Equations 6 and 7 to compute the gas volume in the experiment and the model, respectively.

$$V_{exp}(t) = \sum_{t=t_{exp}}^{t=t_{end}} Q_{exp} \times t; t = t_{exp}, 2 \cdot t_{exp}, 3 \cdot t_{exp}, \dots, t_{end} \quad (6)$$

Here, Q_{exp} is the rate of injection of gas in the experiment [volume/time], t_{exp} is the time step in between the capture of two successive images in the experiment and t_{end} is the end time of the experiment.

$$V_{model}(n) = \sum_{n=1}^{n=n_{top}} n_{blocks} \times \phi \times S_g \times V_{block}; n = 1, 2, 3, \dots, n_{top} \quad (7)$$

Here, n_{blocks} is the number of blocks invaded per loop counter n in the Macro-IP model described in Section 3.2, n_{top} is the model loop counter when the gas reaches the top of the domain, and V_{block} is the volume of each discretized block of the Macro-IP model. The porosity ϕ of the sand used in the experiment is assumed to be uniform, and S_g is the gas saturation value assigned to the entire gas cluster.

After that, for all the time-wise elements in the V_{exp} and V_{model} data vectors, the Euclidean distance is computed. Based on this distance value, each element of the V_{exp} vector gets a nearest neighboring element (minimum Euclidean distance) in the V_{model} vector. For all those nearest-neighbor pairs, we assign the experimental time t to the corresponding model loop counter n . We may now compare the respective time-matched images based on our proposed Diffused Jaccard Metric (Section 2.3). In our specific implementation, we use the MATLAB inbuilt function *knnsearch* with the *exhaustive* search algorithm to efficiently conduct the nearest neighbor search.

Depending on the type of model, the model loop counter may increase without an increase in volume because of a fragmentation process. A fragmentation process refers to a combination of an imbibition (of water) step and an invasion (of gas) step. If this imbibition occurs at the bottom of a gas cluster, mobilization of the gas cluster occurs. If the imbibition occurs in any other block along the gas cluster, it can lead to snap-off events, dividing the cluster into two discrete clusters. Hence, this process does not add gas to the system but is actually a re-arrangement of the gas-occupied blocks (Birovljev et al., 1995; Mumford et al., 2009; Wagner et al., 1997). In such a situation, all these model loop counters will match the same single experimental time (because each experimental image by definition shows an increase in volume due to the fixed gas injection rate Q_{exp}). We recommend assigning the experimental time to the last matching model loop counter in that case because this model state represents the completed processes at the matched volume.

3. Demonstration of the Method

In this section, we present a demonstration of our comparison method on a real experiment of gas injection into homogeneous sand. First, we describe the experiment (Section 3.1) and the model (Section 3.2) chosen for our comparative study. After that, we list the steps we conducted in this experiment-to-model comparison (Section 3.3).

3.1. Experimental Data

We use experimental data obtained from an air injection experiment in homogeneous water-saturated sand of 0.7 mm average grain size. The experimental setup and the data processing method are described in detail in Van De Ven and Mumford (2019). Here, we present a brief description of the experiment (Experiment 0.1-A) used for this study.

Air was injected in a thin acrylic glass cell of size 250 mm \times 250 mm \times 10 mm, filled with homogeneous water-saturated sand (porosity = 0.366) at an injection rate of 0.1 ml min⁻¹. At this injection rate, the gas flow is observed as discontinuous clusters that fragment and coalesce while migrating vertically. The occurrence of discontinuous gas flow was further verified by continuous gas pressure measurements at the gas injection point (Van De Ven & Mumford, 2019). For more details on the gas flow regimes, we recommend referring to the work of Geistlinger et al. (2006). The data is obtained as a time series of 2D images collected at the rate of 30 frames per second for a total experimental time of 330 s, using the light transmission technique (Kechavarzi et al., 2000; Niemet & Selker, 2001; Tidwell & Glass, 1994).

Individual pixel intensity values of these raw images are averaged over a block size of 1 \times 1 mm, and the block is assigned the average intensity value. Then, optical density values are computed per block, and a detection limit of 0.02 is set (Van De Ven & Mumford, 2019). Optical density is defined as the negative logarithm transform of the ratio of the transmitted to incident light intensity (Kechavarzi et al., 2000). Wherever the block intensity exceeds the detection limit, gas is considered to be present. This results in a time series of binary images (Van De Ven & Mumford, 2019).

3.2. Model

In this work, we use the Macro-IP conceptual model for simulation of gas (air) flow in water-saturated sand, as in Mumford et al. (2015), which is based on the Near-Pore Modified Macro Invasion Percolation model (NP-MMIP) of Glass et al. (2001).

Initially, the defending phase (here: water) occupies all the blocks. Then, we place the invading phase (here: air) in the block that contains the gas injection point. The model operates at a block-based spatial discretization of 1×1 mm in accordance with the processed experimental conditions. We evaluate the blocks connected to the invaded block for their invasion thresholds (T_e) using:

$$T_e = P_e + P_w \quad (8)$$

Here, P_e is the local entry pressure, which is the capillary pressure required by air to percolate a water-occupied block, and P_w is the pressure of the water phase. Then, we invade the neighboring block with the lowest T_e .

The term P_w in Equation 8 incorporates the buoyancy forces to more accurately simulate gas migration in a water-saturated porous medium. Hydrostatic water pressure is assumed, and therefore P_w is calculated using:

$$P_w = \rho_w g z \quad (9)$$

Here, ρ_w is the water density, g is the acceleration due to gravity, and z is the height from the top of the acrylic glass cell (along Z axis).

The P_e field of a porous medium depends on the pore-scale arrangement of the solid and its interaction with the fluids. Thus, a precise measurement of the P_e field at our experimental scale is difficult and highly computationally expensive. Since P_e is a point on the capillary pressure (P_c)–saturation (S) curve, we sample the P_e values that we assign individually to all model blocks, using the Brooks-Corey $P_c - S$ relationship (Brooks & Corey, 1964) for our material of interest:

$$S_e = \frac{S_w - S_r}{1 - S_r} = \left(\frac{P_c}{P_d} \right)^{-\lambda} \quad (10)$$

The term $\frac{S_w - S_r}{1 - S_r}$ is called the effective wetting phase saturation (S_e), where S_w is the wetting phase saturation and S_r is the residual wetting saturation. Also, P_c is the capillary pressure, P_d is the macroscopic displacement pressure which we used to generate the cumulative distribution function of P_c , and λ is the pore-size distribution index. We sample the P_e values from the inverse of the cumulative distribution function of P_c (using Equation 10):

$$P_c = P_d u^{\frac{1}{\lambda}} \quad (11)$$

Here, u is a random number from the standard uniform distribution in the interval $[0, 1]$. This sampling method is called the Inverse Transform sampling method, which has been used in the works of Glass et al. (2001) and Mumford et al. (2015). Note that there exists no spatial correlation between the P_e values assigned to the blocks, but this extension could easily be achieved via geostatistical simulation.

The Macro-IP model from Mumford et al. (2015) also includes an additional rule for fragmentation and mobilization of gas clusters based on the work of Wagner et al. (1997). In Wagner et al. (1997), the re-invasion of water (causing fragmentation and mobilization) into the gas-filled pores is realized by a withdrawal pressure threshold at the pore scale. At the scale of the Macro-IP model, the threshold for re-invasion, also known as the terminal threshold (T_t), is calculated as the summation of the terminal pressure (P_t) and the hydrostatic pressure (P_w).

$$T_t = P_t + P_w \quad (12)$$

The values of P_t are calculated based on a P_e - to $-P_t$ ratio obtained from the characteristic drainage and imbibition curves for the porous medium that accounts for the capillary-pressure hysteresis (Gerhard & Kueper, 2003).

Table 1
Macro-IP Model Parameter Values

Parameter	Symbol	Values	Units
Density of water	ρ_w	1,000	kg/m ³
Acceleration due to gravity	g	9.82	m/s ²
Average $P_t - P_e$ ratio	α	0.6	- (Mumford et al., 2009)
Displacement pressure	P_d	8.66	cm of H ₂ O (Schroth et al., 1996)
Pore-Size distribution index	λ	5.57	- (Schroth et al., 1996)
Model domain size	x-z	25 × 25	cm ²
Block discretization	-	1 × 1	mm ²

$$P_t = \alpha P_e \quad (13)$$

A fragmentation or mobilization event of a gas cluster occurs when:

$$T_{t,g} > T_{e,w} \quad (14)$$

where g and w stand for gas- and water-occupied blocks, respectively (Mumford et al., 2015). Re-invasion of water into a gas-occupied block causes a complete expulsion of gas from that block. If the re-invasion of water occurs at sites adjacent to a gas cluster, the gas cluster's mobilization occurs. If the re-invasion causes a disconnection in the gas cluster, fragmentation occurs. After each fragmentation or mobilization step, clusters are re-identified to account for coalescence events. A gas cluster's growth occurs only when (re-) connected to the injection block.

Table 1 lists the parameter values taken for the Macro-IP model. We run the model 1,000 times, each one differing in the random spatial arrangement of the P_e values, thus also differing in the terminal pressure's spatial arrangement. These model runs represent the uncertainty in these parameter values, which is difficult (if not even impossible) to determine with experimentation. For each of these 1,000 model runs, the simulation ends when the gas reaches any block on the top of the model domain.

In our Macro-IP version, we assume a constant gas saturation value of 0.2, based on Van De Ven et al. (2020), for all the gas clusters (Mumford et al., 2015). This approach may not be an accurate representation of reality. However, we make this assumption to retain the model's simplicity and hence reduce its computational cost, and this assumption does not affect the discussion of our comparison method.

3.3. Steps of the Model-Experiment-Comparison

First, we carry out the experiment-to-model time matching (discussed in Section 2.4) using Equations 6 and 7. The model simulation starts from the time gas first percolates into the water-saturated sand in the experiment. Then, we compute the Jaccard coefficient between the time series of experimental images and the corresponding model images across all the model realizations (see Section 2.3). Further, we compute the Diffused Jaccard coefficient. We alter the radius of blurring from 0% to 50% of the domain size in steps of 1% by changing the standard deviation value.

For comparing our method to the existing method of spatial moments, we also calculate the difference between centroid and variance values (Section 2.2) of the final experimental image and of the corresponding time-matched model output images. We normalize this difference by the centroid and variance values of the experimental image and then identify the minimum of the normalized differences of each of the centroid and variance values in X - Z axes. Comparing zeroth moments M_{00} is unnecessary because our time matching is based on a volume matching, so M_{00} is always accurate.

Additional to the model images, we also translate an experimental image in the X - Z space to compare with the original experimental image. This facilitates additional discussions for our method. On this image, too,

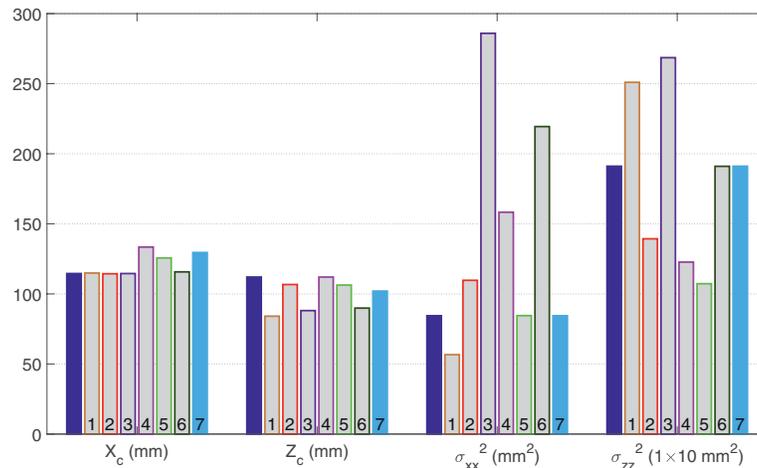


Figure 2. Bar chart of spatial moments in X-Z space (computed using Equation 2) for experimental image and its shifted copy (solid blue bars) and model realizations (gray bars) as listed in Table 2; numbers in the bars denote respective realization number.

we calculate the Jaccard coefficient, the Diffused Jaccard coefficient and the spatial moments. We present and discuss the results of this application in Section 4.

4. Results and Discussion

In this section, we present the results obtained from our case-study in Section 3. Section 4.1 discusses results based on the experimental image at the end of the experiment and corresponding time-matched images of all model realizations. Section 4.2 focuses on using the proposed metric on the time-series of the time-matched images from the experiment and model.

4.1. Comparison Based on Final Experimental Image

We can compute the (Diffused) Jaccard coefficient for any of the time-matched images of the experimental data time-series. Here, we begin our discussion by picking the experimental image at the end of the experiment (at $t = 330s$) and the corresponding time-matched model image per model realization for clarity. The quality-of-fit (no matter if assessed perceptually, by spatial moments or by our proposed metric) varies significantly across the 1,000 model realizations. For example, the values for J vary between 0.003 and 0.17, and those of J_d vary between 0.01 and 0.79. Therefore, we select a set of model realizations for discussion that show a reasonable agreement with the experimental image. We will motivate our choice of individual realizations in Section 4.1.1.

4.1.1. Spatial Moments Can Be Misleading

We start by comparing the spatial moments of seven arbitrary (time-matched) images to the final experimental image (Figure 2).

By comparing the height of the gray bars (model realizations) against the height of the solid dark blue bars (experimental data) in Figure 2, we see that none of the six model realizations clearly outperforms the others in all spatial moments; rather, performance varies significantly across the four measures. Indeed, we have chosen four of the realizations (numbers 3, 4, 5, and 6) to exactly match X_c , Z_c , σ_{xx}^2 and σ_{zz}^2 , respectively (see Figure 2). Recall that all of our time-matched realizations by definition match the zeroth moment, which is equivalent to the volume in the domain. Table 2 lists the characteristics of the seven realizations chosen for detailed analysis and discussion here.

The final model images of these realizations are visually displayed in Figure 3. Realization 7 in Figure 3h might look familiar to the reader—it is, in fact, the original experimental image, but translated 15 mm to the right and 10 mm upwards in X-Z space. This realization represents a typical situation where the actual

Table 2
Characteristics of the Seven Realizations Selected for Detailed Analysis Along With Their Ranking Obtained From Their Jaccard Coefficient (J) and Diffused Jaccard Coefficient (J_d) Values at 4% of Domain Size Blur Radius

Realization number	Description	Rank by J	Rank by J_d
1	Model realization with maximum Jaccard coefficient value for final image	1	2
2	Model realization with fastest increase of diffused Jaccard coefficient with increase in blurring radius	2	1
3	Model realization with minimum normalized difference to experimental centroid in X direction for final image	6	4
4	Model realization with minimum normalized difference to experimental centroid in Z direction for final image	4	7
5	Model realization with minimum normalized difference to experimental variance in X direction for final image	5	6
6	Model realization with minimum normalized difference to experimental variance in Z direction for final image	3	3
7	Experimental image translated in X-Z space (“shifted copy of experimental image”)	7	5

Note. The individual spatial moment values for the realizations can be read off from Figure 2. The values for J are provided in Figure 3, and those for J_d are provided in Figure 6.

injection point is not precisely known. Consistently, this realization meets the second spatial moments (spread) in X and Z but fails in reproducing the exact position of the gas (first moments). Hence, if we relied on a spatial moments comparison, we might fail in identifying this realization as an (almost) perfect fit. In real applications, identification of this realization as the true one is essential because it would tell us that the model we use is, in fact, correct, but only the injection point is assumed at the wrong position. We cannot afford to *not* detect a perfect model realization. Therefore, we strongly recommend using our proposed (Diffused) Jaccard coefficient as a metric of comparison, as discussed in the following sections.

4.1.2. Jaccard Coefficient Yields More Conclusive Ranking

Realization 1 is chosen as the realization that has scored the best Jaccard coefficient (cf. Table 2). When assessed through the eyes of spatial moments, this realization does not look especially convincing (the gray bar for realization 1 is of the same height as the solid dark blue experimental bar for X_c , but shorter for Z_c and σ_{xx}^2 and taller for σ_{zz}^2 in Figure 2). However, from Figure 3b we see that, perceptually, it does achieve a good fit, which confirms that comparison on a pixel-to-pixel basis as done with the Jaccard coefficient yields more conclusive results than the spatially-aggregated moments method. We observe that the gas finger of realization 1 with the highest J value extends farther to the top than the original experimental image (Figure 3a). This is not a systematic result, but this particular randomized entry pressure field yielded the highest pixel-by-pixel agreement with the experimental image. Suppose the distance of the gas finger to the top of the glass cell was of particular interest to the modeler. In that case, one could modify the Jaccard coefficient calculation such that it only compares, for example, the blocks of the outer gas finger boundary.

Realizations 3–6 (Figures 3d–3g) have a poor perceptual fit to the experimental image, and this is reflected in their Jaccard coefficient values. However, a combination of spatial moments of these realizations cannot lead us to this conclusion. For example, in Figure 2, compared to the height of the experimental dark blue bars, the height of realization 3 gray bars vary slightly for Z_c value but highly for σ_{xx}^2 and σ_{zz}^2 values. Simultaneously, the height of realization 5 gray bars vary slightly for X_c and Z_c values but highly for σ_{zz}^2 value. Thus, when we compare the height of the gray bars for realizations 3–6 to the height of the dark blue experimental bars in Figure 2, we cannot identify any pattern in their variations. In contrast, the Jaccard coefficient provides a single value and a clear, precise ranking of the realizations.

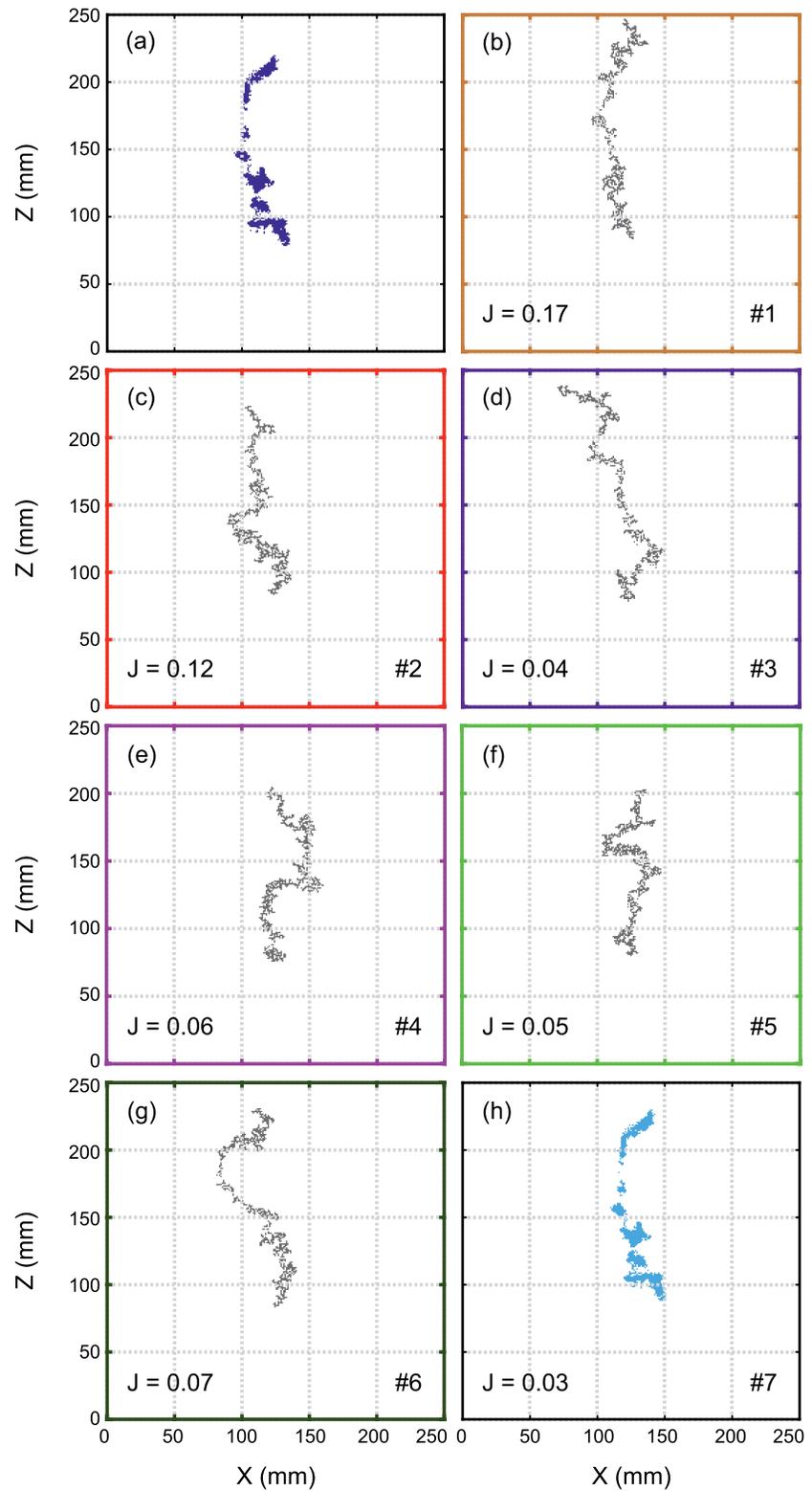


Figure 3. Images at the end of the experiment as a basis for perceptual comparison. (a) Final experimental image (b - g) time-matched model realizations and (h) shifted copy of the experimental image as listed in Table 2, with box colors and numbers in figures representing the respective realization.

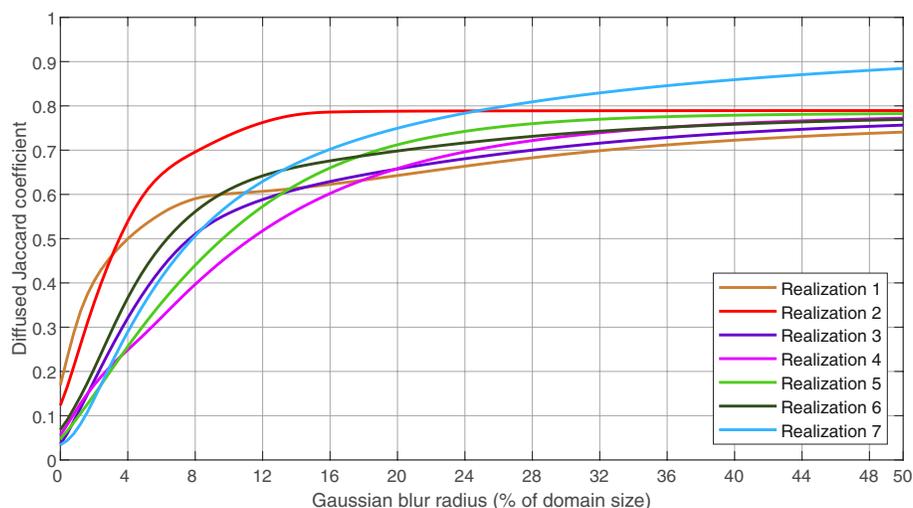


Figure 4. Diffused Jaccard coefficient as a function of Gaussian blur radius for the realizations listed in Table 2.

Yet, we observe that Realization 7 has a very low J value of 0.03, although it is just a translation of the original experimental image. Therefore, we recommend using the diffused version of the Jaccard coefficient (see next Section).

4.1.3. Diffused Jaccard Coefficient Provides Most Insightful Ranking

We have already schematically illustrated the evaluation of the Diffused Jaccard coefficient with Figure 1 in Section 2.3. In fact, the experimental image shown in Figure 1 corresponds to the final experimental image (Figure 3a), and the shown model realization corresponds to realization 2 (Figure 3c) from Table 2. When we blur the images as in Figures 1d and 1e (here, with 4% domain size blurring), we compare their similarity (using J_d) on a scale larger than that of the individual pixels.

Figure 4 summarizes the performance of the seven realizations of Table 2 as a function of the Gaussian blur radius varied from a detailed pixel level of 0% of the domain size (non-diffused Jaccard coefficient value) to a very summarized level of 50% of the domain size.

In Figure 4, upon increasing the blurring radius, the J_d values for the shifted copy of the experimental image (realization 7) keep increasing and go up to 0.9 (cyan blue plot line). Thus, the problem with using the Jaccard coefficient mentioned in the previous section for realization 7 is resolved using the Diffused Jaccard coefficient. So, for cases like the shifted copy of the experimental image, while spatial moments lead to inconclusive results and the non-diffused Jaccard coefficient evaluates it as a poor fit, the use of the Diffused Jaccard coefficient is a safer choice. Also, we picked realization 2 (see Table 2) for having the steepest increase of J_d values upon increasing blur radius in Figure 4 and we see that it is a good perceptual match to the experimental image (see Figure 3c). Recall that Figure 1 shows the intersection of realization 2 with the experimental image in a blurred and non-blurred state. Indeed we find that both realizations look very similar “in nature,” although the pixel-by-pixel comparison would not see that because they differ slightly in position and tortuosity.

Further, note that in Figure 4, the J_d values of realizations 3–6 do not improve as fast as realizations 1 and 2 with increased blurring. This slower rate of increase implies that the Diffused Jaccard metric does not favor perceptually different realizations as much as it favors realizations with small offsets to the original experimental image. To further demonstrate that blurring does not favor all realizations equally, we show seven arbitrary model realizations at a blur radius of 4% of the domain size in Figure 5. At this blur radius, the value of J_d amongst the 1,000 model realization images varies between 0.05 and 0.54. From Figure 5, it is evident that blurring does fade out the information in the pixels; however, for a given blur radius, differences between good and bad realizations are revealed reliably.

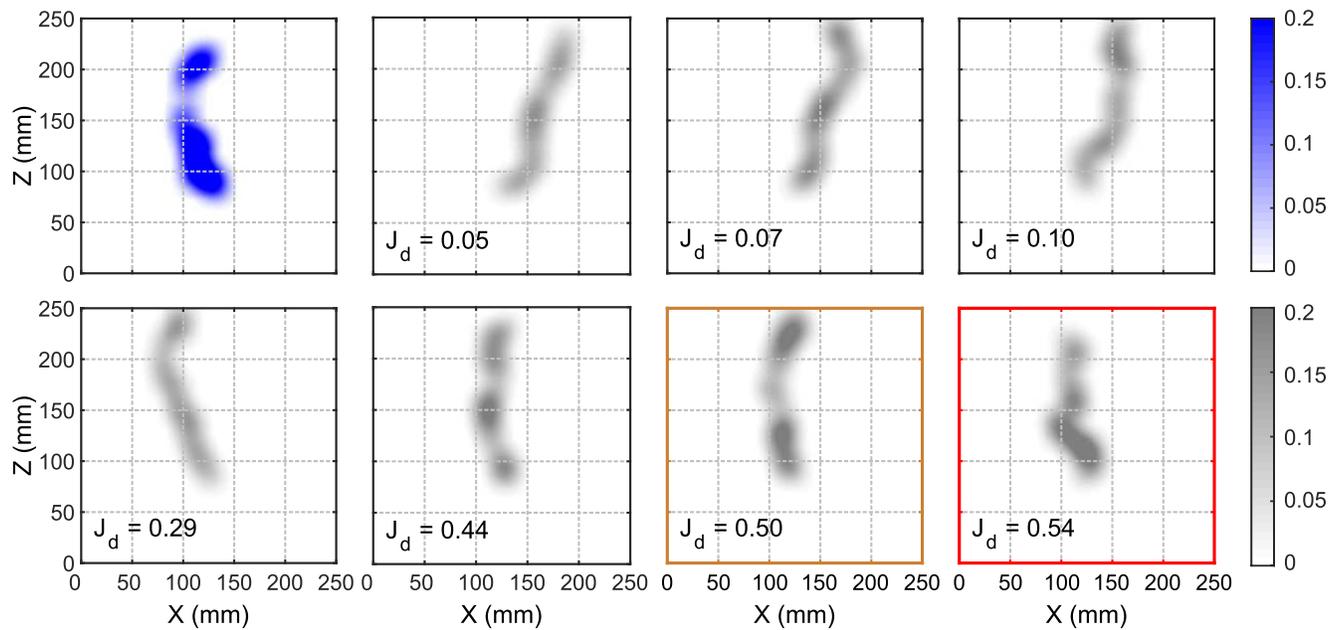


Figure 5. A set of arbitrarily selected model realizations along with the experimental image to present the range of the Diffused Jaccard coefficient J_d at 4% of the domain size blur radius; the model realizations in mustard yellow box and red box are realization 1 and realization 2 from Table 2, respectively.

To investigate to what extent we can blur the images for a meaningful comparison, we show our selection of seven realizations from Table 2 along with the experimental image at different blur radii in Figure 6. It is evident from this figure that the images lose more and more pixel details with increasing blur radius. We notice that, at blur-radius values of 20% and 40% of domain size, the images from the experiment and realizations 1–7 look almost the same. This explains why the J_d values of the different realizations do not improve significantly after approximately 20% of domain size blur radius in Figure 4. Thus, the upper limit of the blur radius would depend on the application for which the experiments and modeling are done. However, in general, it seems safe to say that any radius of blurring exceeding 50% of the domain size is questionable.

4.2. Comparison Based on Experimental Time Series

This section explores the possibility of computing the (Diffused) Jaccard coefficient for the time-series of time-matched images. Figure 7 summarizes the achieved diffused Jaccard values J_d of the six model realizations of Table 2 as a function of experimental time at a blur radius of 4% of the domain size. This plot helps evaluate the model performance over the entire process of evolution of the gas finger instead of only assessing the final gas finger at the end of the experiment. Note that we refrain from computing the spatial moments over time because that would give us four time-series per model realization; we would not know how to combine them into a single meaningful measure for model realization ranking. Remember that realization 7 is just the final experimental image shifted in space, so we do not show its temporal evolution here.

All the time-series plotted in Figure 7 show some common trends. As examples, we pick two instances of time for further investigation. First, the value of J_d for all realizations drops at a time of around 15s. Second, all the realizations show an abrupt increase at around 240s.

When we zoom into the gas cluster at 15s (column 1 of Figure 8), we see that the experiment gas cluster has far more gas-occupied blocks than the corresponding model realizations. This is why, even upon diffusion, the J_d value decreases. The smaller number of gas-occupied blocks in the model images can be due to an over-estimation in the gas saturation value used for volume-based time matching (Section 2.4). Alternative-

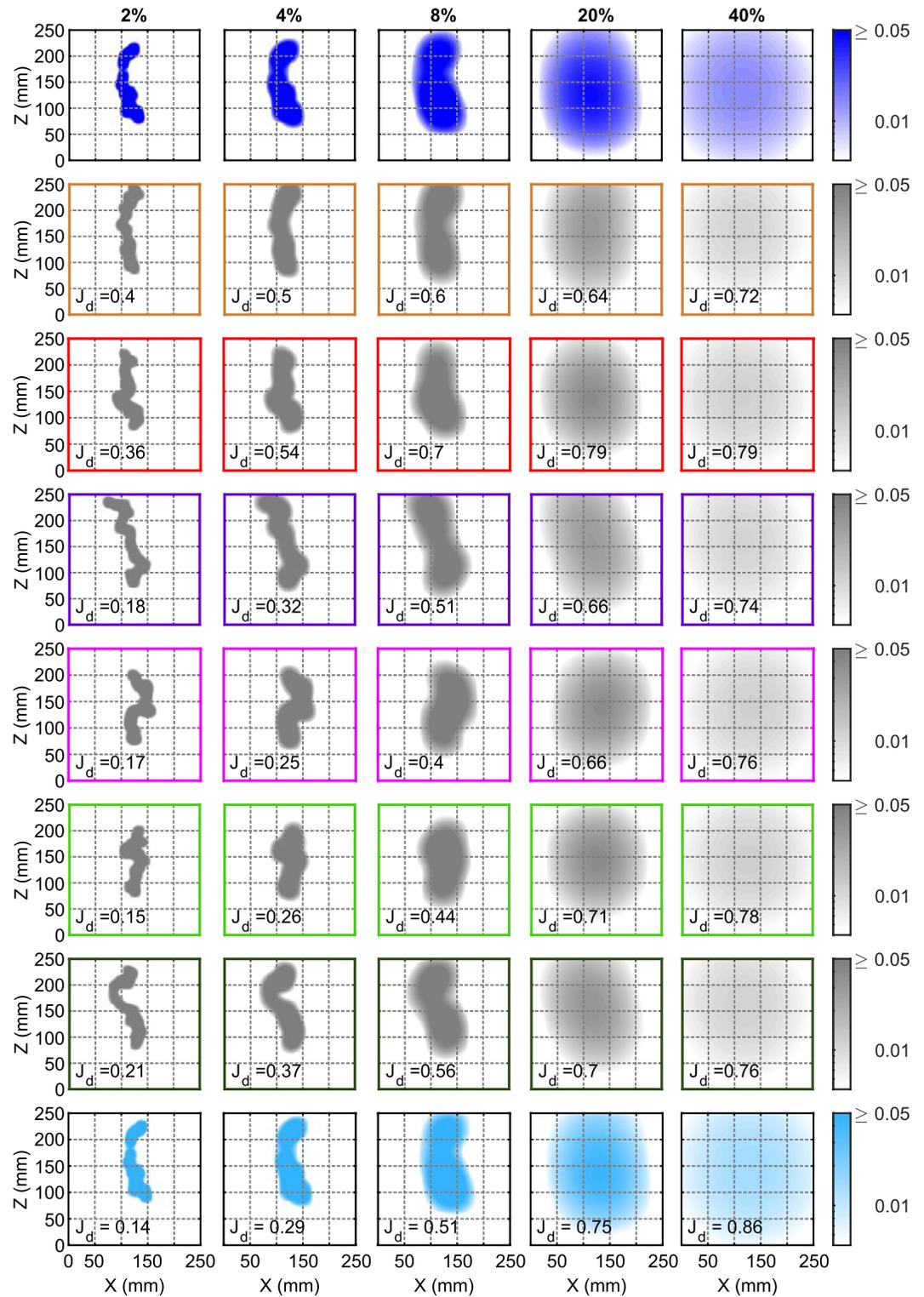


Figure 6. Experimental image and corresponding realization images (from top row to bottom row: Realizations 1–7 of Table 2) with blur-radius values of 2%, 4%, 8%, 20% and 40% of the domain size in logarithmic color-scale. Box colors represent respective realizations; cyan blue pixels are used for the shifted copy of the experimental image (realization 7).

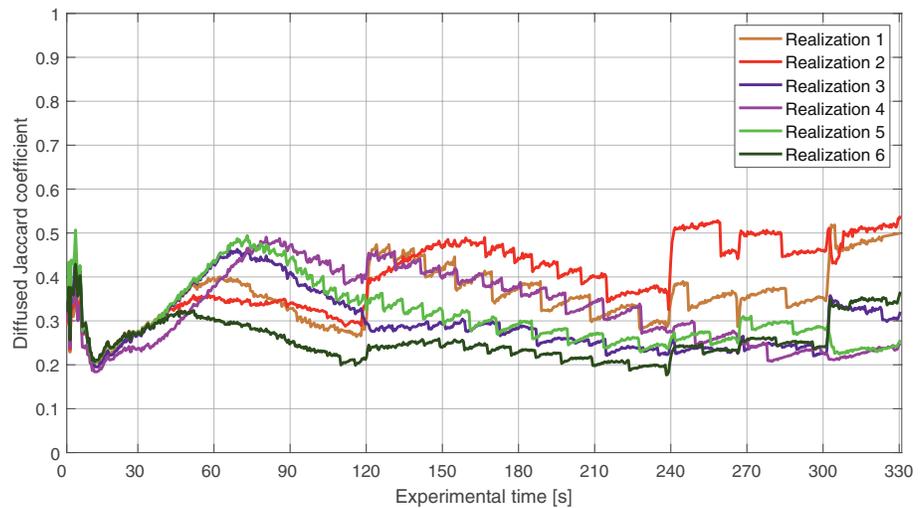


Figure 7. Diffused Jaccard coefficient (J_d) over all time-matched images with a blur radius of 4% of the domain size, for model realizations 1–6 from Table 2.

ly, the higher number of gas-occupied blocks in the experimental image can be due to noisy pixels wrongly identified as gas-occupied pixels.

To investigate the abrupt rise in the J_d value at around 240s, we zoom into the gas finger images at time 239 and 241s in the second and third column of Figure 8, respectively. At both times, when we compare the experimental finger to the model fingers, perceptually, they are different. But we notice that the experimental image (dark blue gas finger) at 239s, with more gas accumulated toward the bottom, is much shorter in height than at 241s. The model tends to expand toward the top, as is seen in all the model realization images of columns 2 and 3 of Figure 8. So, when the gas finger in the experimental image at 241s suddenly moves up, the J_d value increases abruptly for all the model realizations. Note here that the increase in J_d value at around 240s is not the same across all model realizations in Figure 7. Realization 2 (third row, red box in Figure 8) has the closest resemblance to the experimental image at 241s. Accordingly, the maximum increase in J_d value is seen for realization 2 (red plot line) in Figure 7.

On the one hand, looking at the temporal evolution of J_d for specific realizations could help us improve the model structure (identify “wrong decisions”, common errors to all realizations, and so on). On the other hand, the experimental team can use this temporal evolution of J_d as a diagnostic tool to detect unexpected behavior in the experimental data set. Such behavior detection could help them identify and resolve problems with experimental setup conditions or data processing techniques.

4.3. Summary of Findings

We summarize that spatial moments are hard to aggregate into a single meaningful measure; they are inconclusive and can even be misleading (Section 4.1.1). They would be a poor choice for time-series evaluations. Perceptual comparison takes into account our expert intuition about the quality-of-fit, but it becomes an impossible task for the number of realizations (and potentially even timesteps) that we would like to investigate (Section 2.1). The Jaccard coefficient quantifies pixel-by-pixel agreement and is hence close to a “measure of perception”, but it evaluates the perfect-but-shifted model realization only as mediocre (Section 4.1.2). The Diffused Jaccard solves this problem: with increasing blur radius, this image emerges as the best one. We can understand the blur radius as a switch that helps the user control the extent of details to be retained in the images. So, the most meaningful blur radius for a specific application has to be decided by the user; it can reflect the intended purpose of identification (Section 4.1.3). For example, suppose the user wants to know the radius of a gas contamination zone in the subsurface to protect the groundwater table. In that case, the images can have a relatively high blur radius. Our recommendation is to try increasing blur

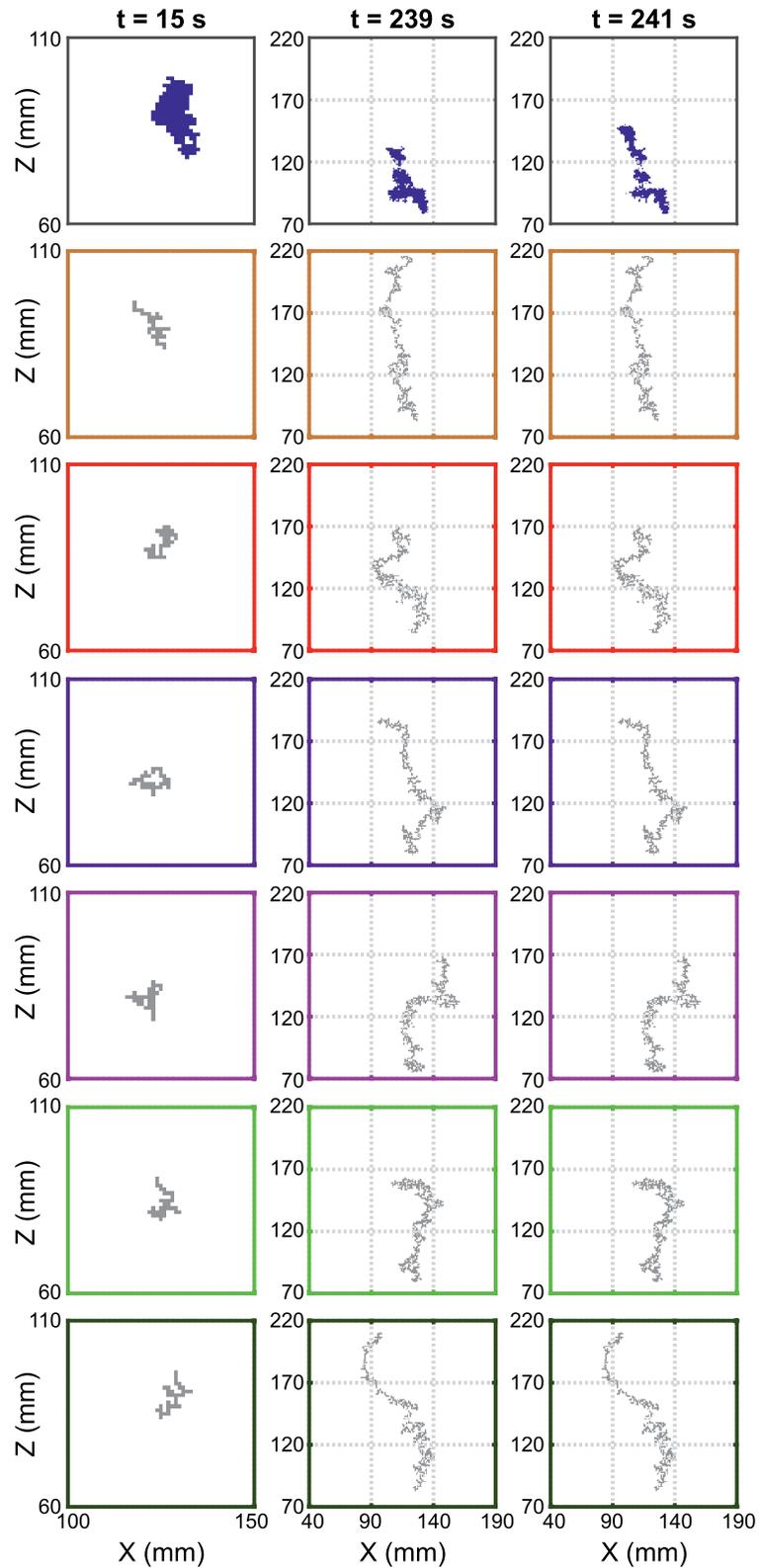


Figure 8. Zoomed images at specific times per column (axis limits are different for each column). From top row to bottom row: experimental images and images of model realization 1–6 from Table 2; box colors represent respective realizations.

radii to observe such effects as shown here synthetically with the translated experimental image and then decide its value. We recommend to always report Diffused Jaccard coefficient value together with its blur radius (with an appropriate unit; here, we have used the percentage of the domain size) for transparency.

5. Summary, Conclusions and Outlook

Our proposed volume-based time-matching method enables us to compare IP-type models with no sense of real-world time to real-time experimental data. To avoid subjective and tedious perceptual comparison of images, we propose to compute the Jaccard coefficient as a quantitative, automated and therefore objective metric for comparison. Applying the proposed method to a real experiment of gas injection in homogeneous saturated sand and Macro-IP model has shown that the Jaccard coefficient provides intuitive results, while spatial moments (as an alternative for an objective method) may be inconclusive or misleading. We also designed a synthetic test case by slightly translating the original final experimental image in space - a situation that might be realistic in many cases. The binary Jaccard coefficient is not able to identify this almost-perfect model due to its pixel-by-pixel strictness. We, therefore, improve on that by using the Diffused Jaccard coefficient. The Diffused Jaccard coefficient can be seen as sliding between rigorous pixel-by-pixel assessment to aggregated (blurred) assessment over space as done in spatial moments. The advantage is that we have a single insightful number for ranking and that we have this slider that can be used for model investigation and improvement.

Our comparison method can be used to compare models other than IP models, and space-time discretized experimental data-sets not limited to light-transmission images. The method can be used to compare any high-resolution model output and experimental data given as raster images.

Beyond the scope of the present research, we anticipate that our proposed metric will be handy for model calibration (identification of best-fit parameter values) and for comparison of alternative model types, for example, to predict under different multiphase flow regimes in porous media, but also, again, in any other discipline that works with raster model output and data.

Data Availability Statement

The experimental data set used in this study is made available by Van De Ven et al. (2021) at Scholars Portal Dataverse: <https://doi.org/10.5683/SP2/RQKOCN>. The modeling data and codes used for this study are made available in the DaRUS dataverse: <https://doi.org/10.18419/darus-1776> (Banerjee et al., 2021).

Acknowledgments

The authors would like to thank the German Research Foundation (DFG) for financial support of this project within the Research Training Group GRK1829 “Integrated Hydrosystem Modeling” and the Cluster of Excellence EXC 2075 “Data-integrated Simulation Science (SimTech)” at the University of Stuttgart under Germany’s Excellence Strategy - EXC 2075 - 39074001. The authors would also like to thank Prof. Dr. Olaf A. Cirpka, University of Tübingen, for fruitful discussions on an early version of this manuscript.

References

- Banerjee, I., Guthke, A., & Nowak, W. (2021). Replication data for: Overcoming the model-data-fit problem in porous media: A quantitative method to compare invasion-percolation models to high-resolution data. *DaRUS*. <https://doi.org/10.18419/darus-177610.18419/darus-1776>
- Berkowitz, B., & Ewing, R. P. (1998). Percolation theory and network modeling applications in soil physics. *Surveys in Geophysics*, *19*, 23–72. <https://doi.org/10.1023/A:1006590500229>
- Birovljev, A., Furuberg, L., Feder, J., Jssang, T., Mly, K. J., & Aharony, A. (1991). Gravity invasion percolation in two dimensions: Experiment and simulation. *Physical Review Letters*, *67*, 584–587. <https://doi.org/10.1103/PhysRevLett.67.584>
- Birovljev, A., Wagner, G., Meakin, P., Feder, J., & Jøssang, T. (1995). Migration and fragmentation of invasion percolation clusters in two-dimensional porous media. *Physical Review E*, *51*, 5911–5915. <https://doi.org/10.1103/PhysRevE.51.5911>
- Brooks, M. C., Wise, W. R., & Annable, M. D. (1999). Fundamental changes in situ air sparging flow patterns. *Ground Water Monitoring and Remediation*. <https://doi.org/10.1111/j.1745-6592.1999.tb00211.x>
- Brooks, R., & Corey, A. (1964). Hydraulic properties of porous media. *Hydrology Papers* (Vol. 4). Colorado State University.
- Deza, M. M., & Deza, E. (2016). Encyclopedia of distances. *Springer-Verlag*. <https://doi.org/10.1007/978-3-662-52844-0>
- Essaid, H. I., & Hess, K. M. (1993). Monte Carlo simulations of multiphase flow incorporating spatial variability of hydraulic properties. *Ground Water*, *31*, 123–134. <https://doi.org/10.1111/j.1745-6584.1993.tb00836.x>
- Ewing, R. P., & Berkowitz, B. (2001). Stochastic pore-scale growth models of DNAPL migration in porous media. *Advances in Water Resources*, *24*, 309–323. [https://doi.org/10.1016/S0309-1708\(00\)00059-2](https://doi.org/10.1016/S0309-1708(00)00059-2)
- Frette, V., Feder, J., Jøssang, T., & Meakin, P. (1992). Buoyancy-driven fluid migration in porous media. *Physical Review Letters*, *68*, 3164–3167. <https://doi.org/10.1103/PhysRevLett.68.3164>
- Geistlinger, H., Krauss, G., Lazik, D., & Luckner, L. (2006). Direct gas injection into saturated glass beads: Transition from incoherent to coherent gas flow pattern. *Water Resources Research*, *42*(7), 1–12. <https://doi.org/10.1029/2005WR004451>
- Geistlinger, H., Lazik, D., Krauss, G., & Vogel, H. J. (2009). Pore-scale and continuum modeling of gas flow pattern obtained by high-resolution optical bench-scale experiments. *Water Resources Research*, *45*. <https://doi.org/10.1029/2007WR006548>

- Gerhard, J. I., & Kueper, B. H. (2003). Capillary pressure characteristics necessary for simulating DNAPL infiltration, redistribution, and immobilization in saturated porous media. *Water Resources Research*, 39. <https://doi.org/10.1029/2002WR001270>
- Glass, R. J., Conrad, S. H., & Yarrington, L. (2001). Gravity-destabilized nonwetting phase invasion in macroheterogeneous porous media: Near-pore-scale macro modified invasion percolation simulation of experiments. *Water Resources Research*, 37(5), 1197–1207. <https://doi.org/10.1029/2000WR900294>
- González-Nicolás, A., Trevisan, L., Illangasekare, T. H., Cihan, A., & Birkholzer, J. T. (2017). Enhancing capillary trapping effectiveness through proper time scheduling of injection of supercritical CO₂ in heterogeneous formations. *Greenhouse Gases: Science and Technology*, 7, 339–352. <https://doi.org/10.1002/ghg.1646>
- Hegele, P. R., & Mumford, K. G. (2014). Gas production and transport during bench-scale electrical resistance heating of water and trichloroethene. *Journal of Contaminant Hydrology*, 165, 24–36. <https://doi.org/10.1016/j.jconhyd.2014.07.002>
- Jawitz, J. W., Annable, M. D., Demmy, G. G., & Rao, P. S. (2003). Estimating nonaqueous phase liquid spatial variability using partitioning tracer higher temporal moments. *Water Resources Research*, 39. <https://doi.org/10.1029/2002WR001309>
- Kechavarzi, C., Soga, K., & Wiart, P. (2000). Multispectral image analysis method to determine dynamic fluid saturation distribution in two-dimensional three-fluid phase flow laboratory experiments. *Journal of Contaminant Hydrology*, 46, 265–293. [https://doi.org/10.1016/S0169-7722\(00\)00133-9](https://doi.org/10.1016/S0169-7722(00)00133-9)
- Kueper, B. H., & Gerhard, J. I. (1995). Variability of point source infiltration rates for two-phase flow in heterogeneous porous media. *Water Resources Research*, 31, 2971–2980. <https://doi.org/10.1029/95WR02329>
- Kueper, B. H., & McWhorter, D. B. (1992). The use of macroscopic percolation theory to construct large-scale capillary pressure curves. *Water Resources Research*, 28, 2425–2436. <https://doi.org/10.1029/92WR01176>
- Lenormand, R., Touboul, E., & Zarcone, C. (1988). Numerical models and experiments on immiscible displacements in porous media. *Journal of Fluid Mechanics*, 189, 165–187. <https://doi.org/10.1017/S0022112088000953>
- Molnar, I. L., Mumford, K. G., & Krol, M. M. (2019). Electro-thermal subsurface gas generation and transport: Model validation and implications. *Water Resources Research*, 55, 4630–4647. <https://doi.org/10.1029/2018WR024095>
- Mumford, K. G., Dickson, S. E., & Smith, J. E. (2009). Slow gas expansion in saturated natural porous media by gas injection and partitioning with non-aqueous phase liquids. *Advances in Water Resources*, 32(1), 29–40. <https://doi.org/10.1016/j.advwatres.2008.09.006>
- Mumford, K. G., Hegele, P. R., & Vandenberg, G. P. (2015). Comparison of two-dimensional and three-dimensional macroscopic invasion percolation simulations with laboratory experiments of gas bubble flow in homogeneous sands. *Vadose Zone Journal*, 14(11). <https://doi.org/10.2136/vzj2015.02.0028>
- Mumford, K. G., Smith, J. E., & Dickson, S. E. (2010). The effect of spontaneous gas expansion and mobilization on the aqueous-phase concentrations above a dense non-aqueous phase liquid pool. *Advances in Water Resources*, 33(4), 504–513. <https://doi.org/10.1016/j.advwatres.2010.02.002>
- Niemet, M. R., & Selker, J. S. (2001). A new method for quantification of liquid saturation in 2D translucent porous media systems using light transmission. *Advances in Water Resources*, 24(6), 651–666. [https://doi.org/10.1016/S0309-1708\(00\)00045-2](https://doi.org/10.1016/S0309-1708(00)00045-2)
- Oldenburg, C. M., Mukhopadhyay, S., & Cihan, A. (2016). On the use of Darcy's Law and invasion-percolation approaches for modeling large-scale geologic carbon sequestration. *Greenhouse Gases: Science and Technology*, 6(1), 19–33. <https://doi.org/10.1002/ghg.1564>
- Oostrom, M., Dane, J. H., & Wietsma, T. W. (2007). A review of multidimensional, multifluid, intermediate-scale experiments: Flow behavior, saturation imaging, and tracer detection and quantification. *Vadose Zone Journal*, 6, 610–637. <https://doi.org/10.2136/vzj2006.0178>
- Pantazidou, M., & Liu, K. (2008). DNAPL distribution in the source zone: Effect of soil structure and uncertainty reduction with increased sampling density. *Journal of Contaminant Hydrology*, 96, 169–186. <https://doi.org/10.1016/j.jconhyd.2007.11.002>
- Sakaki, T., Plampin, M. R., Pawar, R., Komatsu, M., & Illangasekare, T. H. (2013). What controls carbon dioxide gas phase evolution in the subsurface? Experimental observations in a 4.5m-long column under different heterogeneity conditions. *International Journal of Greenhouse Gas Control*, 17, 66–77. <https://doi.org/10.1016/j.ijggc.2013.03.025>
- Schroth, M. H., Istok, J. D., Ahearn, S. J., & Selker, J. S. (1996). Characterization of miller-similar silica sands for laboratory hydrologic studies. *Soil Science Society of America Journal*, 60, 1331–1339. <https://doi.org/10.2136/sssaj1996.03615995006000050007x>
- Tan, P.-N., Steinbach, M., & Kumar, V. (2005). Association analysis: Basic concepts and algorithms. *Introduction to Data mining*. <https://doi.org/10.1111/j.1600-0765.2011.01426.x>
- Tidwell, V. C., & Glass, R. J. (1994). X ray and visible light transmission for laboratory measurement of two-dimensional saturation fields in thin-slab systems. *Water Resources Research*, 30(11), 2873–2882. <https://doi.org/10.1029/94WR00953>
- Trevisan, L., Illangasekare, T. H., & Meckel, T. A. (2017). Modeling plume behavior through a heterogeneous sand pack using a commercial invasion percolation model. *Geomechanics and Geophysics for Geo-Energy and Geo-Resources*, 3(3), 327–337. <https://doi.org/10.1007/s40948-017-0055-5>
- Trevisan, L., Pini, R., Cihan, A., Birkholzer, J. T., Zhou, Q., & Illangasekare, T. H. (2015). Experimental analysis of spatial correlation effects on capillary trapping of supercritical CO₂ at the intermediate laboratory scale in heterogeneous porous media. *Water Resources Research*, 51, 8791–8805. <https://doi.org/10.1002/2015WR017440>
- Tsimpanogiannis, I. N., & Yortsos, Y. C. (2004). The critical gas saturation in a porous medium in the presence of gravity. *Journal of Colloid and Interface Science*, 270, 388–395. <https://doi.org/10.1016/j.jcis.2003.09.036>
- Van De Ven, C., Mumford, K., & Banerjee, I. (2021). Replication Data for: Overcoming the model-data-fit problem in porous media: A quantitative method to compare invasion-percolation models to high-resolution data. *Scholars portal dataverse*. <https://doi.org/10.5683/SP2/RQKOCN>
- Van De Ven, C. J., Abraham, J. E., & Mumford, K. G. (2020). Laboratory investigation of free-phase stray gas migration in shallow aquifers using modified light transmission. *Advances in Water Resources*, 139, 103543. <https://doi.org/10.1016/j.advwatres.2020.103543>
- Van De Ven, C. J., & Mumford, K. G. (2019). Characterization of gas injection flow patterns subject to gravity and viscous forces. *Vadose Zone Journal*, 18(1), 1–11. <https://doi.org/10.2136/vzj2019.02.0014>
- Vengosh, A., Jackson, R., Warner, N., Darrah, T., & Kondash, A. (2014). A critical review of the risks to water resources from unconventional shale gas development and hydraulic fracturing in the United States. *Environmental Science & Technology*, 48, 8334–8348. <https://doi.org/10.1021/es405118y>
- Vengosh, A., Warner, N., Jackson, R., & Darrah, T. (2013). The effects of shale gas exploration and hydraulic fracturing on the quality of water resources in the United States. *Procedia Earth and Planetary Science*, 7, 863–866. <https://doi.org/10.1016/j.proeps.2013.03.213>
- Wagner, G., Meakin, P., Feder, J., & Jøssang, T. (1997). Buoyancy-driven invasion percolation with migration and fragmentation. *Physica A: Statistical Mechanics and its Applications*, 245(3–4), 217–230. [https://doi.org/10.1016/S0378-4371\(97\)00324-5](https://doi.org/10.1016/S0378-4371(97)00324-5)

- Werth, C. J., Zhang, C., Brusseau, M. L., Ostrom, M., & Baumann, T. (2010). A review of non-invasive imaging methods and applications in contaminant hydrogeology research. *Journal of Contaminant Hydrology*, *113*, 1–24. <https://doi.org/10.1016/j.jconhyd.2010.01.001>
- Wilkinson, D. (1984). Percolation model of immiscible displacement in the presence of buoyancy forces. *Physical Review A*, *30*, 520–531. <https://doi.org/10.1103/PhysRevA.30.520>
- Wilkinson, D., & Willemsen, J. F. (1983). Invasion percolation: A new form of percolation theory. *Journal of Physics A: Mathematical and General*. <https://doi.org/10.1088/0305-4470/16/14/028>
- Woods, A. W., & Norris, S. (2016). Dispersion and dissolution of a buoyancy driven gas plume in a layered permeable rock. *Water Resources Research*, *52*(4), 2682–2697. <https://doi.org/10.1002/2015wr018159>