

Sources of Uncertainty and Error in the Simulation of Flow in Porous Media

James Glimm

Department of Applied Mathematics and Statistics
SUNY at Stony Brook, Stony Brook NY 11794-3600
Center for Data Intensive Computing
Brookhaven National Laboratory, Brookhaven, NY

Shuling Hou

Theoretical Division,
Los Alamos National Laboratory Los Alamos, NM 87545

Yoon-ha Lee

Department of Applied Mathematics and Statistics
SUNY at Stony Brook, Stony Brook NY 11794-3600

David H. Sharp

Applied Physics Division and Theoretical Division,
Los Alamos National Laboratory Los Alamos, NM 87545

Kenny Ye

Department of Applied Mathematics and Statistics
SUNY at Stony Brook, Stony Brook NY 11794-3600

March 29, 2003

JG is supported by the MICS program of the U.S. Department of Energy DE-FG02-90ER25084, DE-AC02-98CH10886, grant DAAL-03-91-0027 and the National Science Foundation, grant DMS-01022480. S. H. and D. H. S

are supported by the U. S Department of Energy. YL is supported by the Department of Energy DE-FG02-90ER25084.

Abstract

We are concerned here with the analysis and partition of uncertainty into component pieces, for a model prediction problem for flow in porous media.

1 Introduction

In previous papers we have developed a Bayesian approach to uncertainty quantification [8, 9, 10] and the analysis of errors [5, 4, 6, 3, 2, 7] in numerical simulations. Prediction depends on both inverse and forward solutions, the former to fix the model and its parameters and the latter to solve the model and make predictions. In the Bayesian framework, the indeterminacy potentially inherent in the inverse solution is resolved by the use of a probability framework. In this framework, the probabilities of observation and solution errors are the leading contributions to the likelihood of an observation, which is a key factor in the definition of the posterior. The posterior is a probability distribution for models and parameters as constrained by observations. Both the inverse and the forward steps depend on the forward simulation, errors in which introduce error and uncertainty into the analysis.

The purpose of this paper is to analyze and partition prediction uncertainty into component pieces associated with substeps of the prediction process. As with earlier papers, we illustrate our ideas with a simplified study of prediction for an idealized petroleum reservoir. We suppose that we have observations of oil production from an early time period (the “past”) and we use this information to constrain the prediction of the production for a later period (the “future”). The past production is known but the geology which gives rise to it is not. The errors or uncertainty in predictions of future production depend on geological, physical, and numerical parameters, *e.g.* the geology correlation length, the oil to water viscosity ratio, and the coarseness of the mesh used in simulating either the forward or inverse problems.

Our goal is model transparency. We want an uncertainty model, that is simple to describe in an understandable and plausible manner the separate contributions to uncertainty, but accurate enough that it does not add to or increase the uncertainty being explained. Not only is such a model of benefit

to the user. More importantly, such a model has a higher chance of being transferable from one problem to another. Because error models require extensive computation to validate and calibrate, the ability to transfer them (from a simple to a more realistic context) is important for the practical application of uncertainty analysis in general.

Our predictions yield a probability distribution for future oil production, based on observations of past oil production. The point of the uncertainty analysis is to assess the spread (uncertainty) contained in this distribution. For example, a probability concentrated at a single value for future production would indicate zero uncertainty. To assess the overall prediction uncertainty, we consider the standard deviation of the ensemble of prediction errors; the ensemble corresponds to different choices for the reservoir which generates the observations of the past production. Each prediction error has been non-dimensionalized by dividing it by its prediction and this normalization has been performed before the standard deviation is computed. The standard deviation σ is estimated by a Least Squares approximation of the slope of a normal plot of prediction errors, as distributed over the ensemble. The conventional 5%, 95% confidence intervals are then given as the mean prediction $\pm 1.96\sigma$.

The main result of this paper is to estimate the separate contributions to σ . The observations are insufficient to specify the geology. As a result there is an inherent uncertainty in the problem specification. This contribution to σ is labeled σ_{geology} . In addition there is uncertainty associated with the use of approximate numerics to solve the inverse and forward problems, which we denote by σ_{inverse} and σ_{forward} . Finally, since we consider simplified (approximate) statistical methodologies, we introduce $\sigma_{\text{statistics}}$ to quantify the effects of these approximations. Use of finite sized ensembles and observational error introduce other uncertainties, not, considered here.

If the various sources of uncertainty were independent, then the σ would combine according to the rule

$$\sigma_{\text{total}}^2 = \sum_i \sigma_i^2 \tag{1}$$

While this formula is useful qualitatively, we observe deviations from it in the range of 30% - 50%, indicating significant correlation among the different contributions to the uncertainty. We have been able to observe directly only some of the individual σ_i . For the others, some version of (1) is used to *define* the missing σ_i , tacitly assuming independence.

A second result is to determine the dependence of the various σ_i on the parameters defining the problem. These parameters represent geological, numerical, and physical information (explanatory variables) which may be assumed to be known. Even in our simple study, the explanatory variables take on 60 distinct values in total. Thus it is important to compress and synthesize this information, so that useful and comprehensible trends in the dependence of the various σ on the explanatory variables can be understood.

Finally, we observe a general ordering in the magnitudes of the various σ :

$$\sigma_{\text{geology}} \gg \sigma_{\text{forward}} \gg \sigma_{\text{inverse}} \gg \sigma_{\text{statistics}} \quad (2)$$

which must be a consequence of the specific features of porous media flow adopted in this work.

2 The Problem Formulation

The present paper and earlier ones in this series [5, 4, 6, 3] serve to establish a proto-type model for solution errors for flow in petroleum reservoirs. In order to focus on the uncertainty quantification issues, we have examined somewhat idealized reservoir descriptions. The idealized Darcy and Buckley-Leverett equations

$$\begin{aligned} v &= -K\Lambda\nabla p ; & \nabla \cdot v &= 0 \\ s_t + v \cdot \nabla f &= 0 \end{aligned}$$

are solved for a total seepage velocity v and oil saturation s . Here $K = K(x, z)$ is the random total permeability, Λ the relative transmissivity and f the fractional flux. The permeability K is sampled from a lognormal distribution. The flow is nondimensionalized to lie in the unit square $0 \leq x, z \leq 1$. See also [4, 6] for a more detailed specification of the simulations.

Our error model is based on arrival time differences for solution flux values (the observed oil to water production data, or oil cut), and thus measures errors in the propagation wave speeds, or saturation level sets, as the relative oil flux is a (nonlinear) function of the saturation. The arrival time errors are non-dimensionalized through their expression as a fraction of the actual arrival time intervals. These response variables parameterize the observations made in the past. We parameterize future observations with the total future oil production. We consider the dependence of the response variables upon

the scaleup mesh spacing, the geology heterogeneity length scale and the hydrocarbon *vs.* injected fluid viscosity ratio. These are the explanatory variables. The scaled up mesh spacing [12, 13, 1, 11, 14] uses 5, 10 and 20 cells between wells, with a fine grid geology model defined by 100 cells between wells. The horizontal correlation length of the permeability field varies by a factor of five, from the fraction $\lambda = 0.2$ of the inter well spacing to the entire inter well spacing $\lambda = 1.0$, with the values $\lambda = 0.2, 0.4, 0.6, 0.8, 1.0$. The vertical correlation length is fixed at 0.02 as a fraction of the vertical reservoir height. The viscosity ratio ν , needed in the definitions of Λ and f , varies from 5 to 40 with values $\nu = 5, 10, 20, 40$. Thus values typical of both water flooding and oil-gas systems are included.

The difference between the fine grid and the coarse grid solution is the solution error, which we study here. We only examine the oil cut, so the solution and the error are time series. We are not using actual production data in this study. We therefore model the real problem by selecting a particular geology K_{i_0} as the “correct” one, and consider the fine grid solution oil cut s_{i_0} as a stand in for the observed oil production \mathcal{O} .

We assume we have production data up to the present time T_0 , defined in terms of an oil cut level $s_{i_0}(T_0)$, here selected as 0.8, 0.6, or 0.4. For each choice of horizontal correlation length λ 50 geologies, sampled from a given log normal distribution, define the prior distribution $P(K)$. According to Bayes theorem, the posterior $P(K|\mathcal{O})$

$$P(K|\mathcal{O}) = \frac{P(\mathcal{O}|K)P(K)}{\int P(\mathcal{O}|K)P(K)dK} .$$

$P(K|\mathcal{O})$ is defined in terms of a likelihood, $P(\mathcal{O}|K)$, for the observation \mathcal{O} to occur assuming the geology K is correct. Let s_f and s_c denote the fine and coarse (upscaled) solutions, respectively. As explained above, we choose s_{f,i_0} as a stand in for an actual observation \mathcal{O} . In evaluating the likelihood, we take $\mathcal{O} = s_{f,i_0}$ and compare to $s_{c,j}$, $j \neq i_0$. The likelihood is thus the probability of an error in $s_{c,j}$ sufficient to produce the discrepancy

$$s_{f,i_0}(t) - s_{c,j}(t), \quad 0 \leq t \leq T_0 .$$

Since we assume K_j is correct in evaluating the likelihood, the likelihood specifies the probability of a solution error of a given size, to which we apply our Gaussian error model. See [4] for more detail.

We specify the arrival time error model with five degrees of freedom. They are the breakthrough time and incremental elapsed time at oil cut levels of

0.8, 0.6, 0.4 and 0.2. To avoid an undue influence from numerical diffusion in the coarse grid simulations, we define the breakthrough time in terms of an oil cut value of 0.95. That is $\Delta(t_l) = t(S_l) - t(S_{l-1})$, where

$$t(S_l) = \sup_t \{s(t) \geq S_l\}, \quad (3)$$

and $S_l = 1 - 0.2 \cdot l$, $0 \leq l \leq N$, and $\Delta(S_0) = t(S_0)$. Thus, the errors to be modeled are $e(S_l) = \Delta_f(S_l) - \Delta_c(S_l)$, the time difference between a given change in the oil cut level for the fine and the coarse grid solutions. The mean and covariance matrix of the error are estimated by

$$\overline{e(l)} = \frac{1}{n} \sum_i e_i(l) \quad (4)$$

$$C^s(l, m) = \frac{1}{n-1} \sum_i \left(e_i(l) - \overline{e(l)} \right) \left(e_i(m) - \overline{e(m)} \right) \quad (5)$$

This elapsed time error model is used only to model the past, $t \leq T_0$, while future oil production

$$\int_{\text{present}}^{\text{final time}} s_i(t) dt \quad (6)$$

is used to describe the future. The final time is set at 1.4 (PVI) in (6) and s_i might denote the fine or coarse grid solution. We correct the course grid solution for the mean solution error in all of its uses.

3 Findings

The standard deviation for the relative error in prediction provides a convenient method for quantitative assessment of the accuracy of a prediction. The standard deviation is computed from the the Bayesian posterior and the error model for forward simulations. They are computed for each choice of exact geology i_0 and then averaged (RMS) over i_0 to yield a final value.

This method is not applicable for fine grid solutions. For the fine grids, we specify an arrival time window (size 0.03) for past observations of the oil cut. Simulations passing through this window define the posterior.

3.1 Approximate independence of distinct error sources

We make precise the definitions of the different σ_i which will enter into our analysis. σ_{geology} represents the uncertainty inherent in the prediction problem. σ_{geology} originates in the incomplete set of observations (the oil cut) used to characterize the reservoir. It is defined as the standard deviation associated with the following prediction problem: the posterior is determined by fine grid solutions, using the windowing method, while the future is also simulated using the fine grid. σ_{forward} is defined as the standard deviation of the prediction error resulting from use of the upscaled solution operator for forward predictions, starting from a knowledge of the exact geology. σ_{inverse} , the uncertainty in the selection of the posterior distribution, *i.e.* the inverse problem, is not directly observable, so we define it in terms of an RMS difference

$$\sigma_{\text{Bayes ff}}^2 = \sigma_{\text{geology}}^2 + \sigma_{\text{inverse}}^2 \quad (7)$$

where $\sigma_{\text{Bayes ff}}$ is defined prediction σ for prediction based on the Bayesian posterior (e.g. course grid solutions for the inverse problem) and the fine grid solutions for the forward simulation. The statistical uncertainty is also not directly observable, and is defined in terms of an RMS difference

$$\sigma_{\text{Bayes ff stat}}^2 = \sigma_{\text{Bayes ff}}^2 + \sigma_{\text{stat}}^2 \quad (8)$$

where $\sigma_{\text{Bayes ff stat}}$ is the standard deviation associated with the Bayesian prediction using a fine grid solution for the forward step, and an approximate statistical model for the coarse grid error covariance used for the inverse problem in the definition of the Bayesian posterior. We consider two approximate statistical models: a diagonal covariance and a simple parametric model in which the diagonal covariance entries satisfy a linear dependence on explanatory variables. See [7] for details. For comparison with (1), we define $\sigma_{\text{Bayes stat}}$ as the standard deviation associated with the Bayesian prediction using a coarse grid for both the inverse and forward steps and an approximate statistical model for the solution error covariance.

Eq. (1) asserts that

$$\sigma_{\text{Bayes stat}}^2 \approx \sigma_{\text{total}}^2 \equiv \sigma_{\text{geology}}^2 + \sigma_{\text{forward}}^2 + \sigma_{\text{inverse}}^2 + \sigma_{\text{stat}}^2 \quad (9)$$

Eliminating the trivial combinations from this formula which collapse by definition, we are asserting that

$$\sigma_{\text{Bayes stat}}^2 \approx \sigma_{\text{total}}^2 \equiv \sigma_{\text{Bayes ff stat}}^2 + \sigma_{\text{forward}}^2 \quad (10)$$

There are actually three versions of this formula, with the full statistics covariance (no approximations), the diagonal covariance and the parametric model diagonal covariance. The formula asserts independence of prediction errors associated with the inverse and forward (past and future) aspects of prediction.

To assess the accuracy of (10), we perform an RMS average over all explanatory parameter values of σ_{total} as defined by the formula and of $\sigma_{\text{Bayes stat}}$ as defined directly. We find that (10) explains about 2/3 of the value of σ_{total} compared to $\sigma_{\text{Bayes stat}}$ in the sense of an RMS average of the dependence of both sides of the equation on the explanatory parameters.

3.2 Dependence of error on problem parameters

Here we develop a parametric model for the dependence of the various σ_i on the explanatory variables. We propose a relation of the form

$$\ln \sigma_i^2 = a_0 + a_{\text{geology}} x_{\text{geology}} + a_{\text{mesh}} x_{\text{mesh}} + a_p x_p \quad (11)$$

Here the a 's are coefficients to be determined and the x 's represent values for the explanatory variables. Thus $x_{\text{mesh}} = \Delta x/D$ where D is the interwell spacing and Δx is the scaled up mesh spacing. Also x_p is the natural logarithm of the ratio of the displaced to displacing viscosities, and x_{geology} is the horizontal correlation length, expressed as a fraction of D . These definitions are modified from related definitions in [7]. With these definitions, our main result for this section is a table of the a 's for each σ_i , together with a percentage of the RMS parametric variability captured by this formula. The a 's are obtained from a nonlinear regression, to minimize errors in a fit to the exponential of (11). See Table 1.

We now discuss these results qualitatively. σ_{geology} increases strongly with correlation length and viscosity ratio. This dependence is shown graphically in Fig. 1. Note that although the viscosity ratio has nothing to do with the geology, it affects the flow, and the degree to which the flow is predictable or unpredictable. High viscosity ratios make the flow less predictable. σ_{forward} has a strong dependence on the degree of scale up (mesh spacing), moderate dependence on the correlation length and very little dependence on the viscosity ratio. See Fig. 2. $\sigma_{\text{Bayes ff}}$ (fine grid forward simulation) is remarkably independent of the degree of scale up in the inverse problem, shows only a slight increase with viscosity ratio, but is strongly increasing with correlation

	a_0	a_{geology}	a_{mesh}	a_p	% predictivity
$\sigma_{\text{Bayes stat}}$	-6.09	1.27	1.51	0.12	91 %
σ_{total}	-6.16	1.40	2.53	0.24	93 %
σ_{geology}	-7.19	1.64	0.0	0.44	92 %
σ_{forward}	-8.14	1.55	6.62	0.31	90 %
σ_{inverse}	-4.31	0.09	3.50	-1.08	55 %
σ_{stat}	-8.54	1.11	3.73	-0.39	33 %

Table 1: Parametric model coefficients and RMS per cent predictivity of the model. The values given here utilize the parametric model for the diagonal approximation to the covariance to the error.

length. σ_{inverse} , with its indirect definition, is somewhat less well predicted by the parametric model, but shows significant decrease with increasing viscosity ratio, and a possible increase with correlation length. The statistical standard deviation σ_{stat} for both diagonal covariance and for the parametric model for the diagonal covariance is small, so that the parametric dependence and the relative error in these quantities is not of much significance. We note also a small correction to Tables 3 and 4 of [7]: the full error model confidence intervals should be incremented by about one percentage point. This being done, the three different statistical models (exact and two approximations) are nearly identical, again showing that a σ_{stat} or confidence interval inferred from this data will be very small.

3.3 Ordering of error sources by magnitude

The main results of this section are summarized in Table 2, which justifies the ordering given in (2). Each entry is defined as an RMS average over the 60 explanatory parameter values, or over those values for fixed Δx .

References

- [1] M. A. CHRISTIE, T. C. WALLSTROM, L. J. D. S. HOU, D. H. SHARP, AND Q. ZOU, *Effective medium boundary conditions in upscaling*, in Proceedings of the 7th European Conference on the Mathematics of Oil Recovery, Baveno, Italy, Sept. 5-8, 2000.

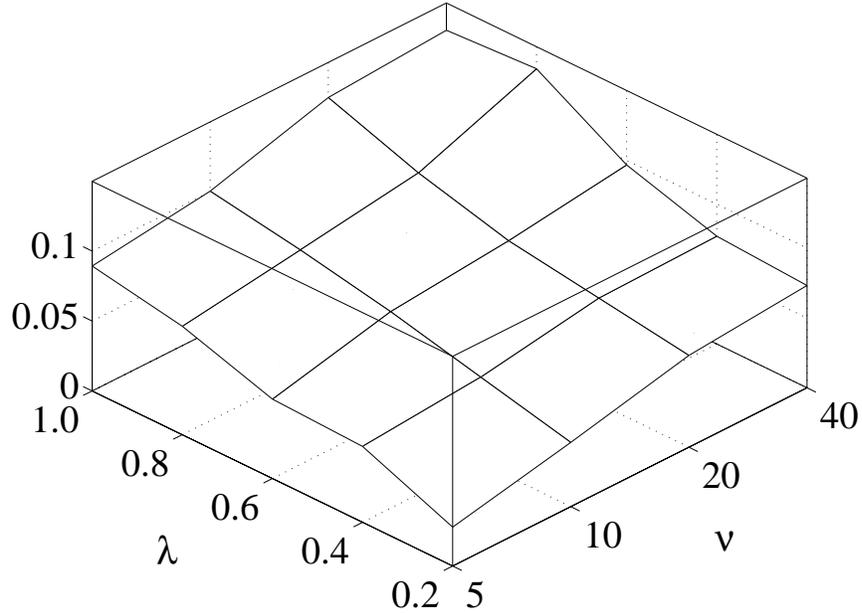


Figure 1: Dependence of σ_{geology} on viscosity ratio ν and correlation length λ .

	Average	$\Delta x/D$		
	(RMS)	0.2	0.1	0.05
σ_{total}	0.117	0.129	0.113	0.108
$\sigma_{\text{Bayes stat}}$	0.092	0.098	0.089	0.090
σ_{geology}	0.087	0.087	0.087	0.087
σ_{forward}	0.067	0.084	0.061	0.053
σ_{inverse}	0.041	0.047	0.039	0.036
σ_{stat}	0.015	0.017	0.014	0.014

Table 2: Magnitudes of distinct contributions to the standard deviation for relative error in predictions. We tabulate in the first column the RMS average over all explanatory parameters, and in the next three columns the RMS average with Δx held fixed.

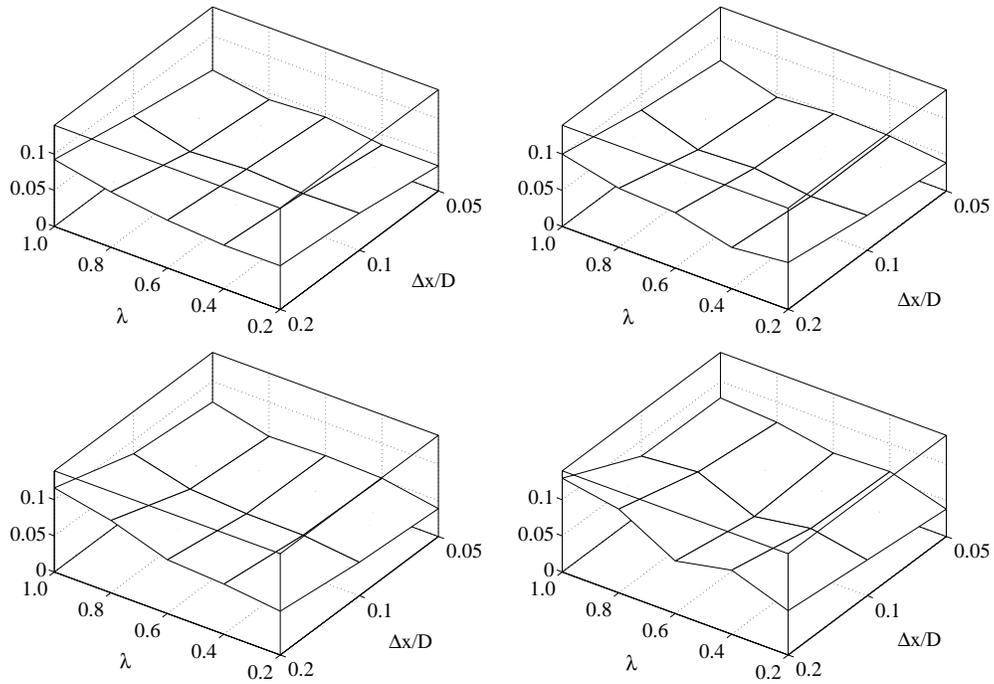


Figure 2: Dependence of σ_{forward} on degree of upscaling, correlation length and the viscosity ratio. The viscosity ratio, which has the weakest influence, is varied in the four frames shown above ($\nu = 5, 10, 20, 40$, in the order upper left, upper right, lower left, lower right) while each of the frames shows the dependence on the most important variables, the up scaling $\Delta x/D$ and the correlation length λ .

- [2] B. DEVOLDER, J. GLIMM, J. W. GROVE, Y. KANG, Y. LEE, K. PAO, D. H. SHARP, AND K. YE, *Uncertainty quantification for multiscale simulations*, Journal of Fluids Engineering, 124 (2002), pp. 29–41.
- [3] J. GLIMM, Y. HA LEE, AND K. YE, *A simple model for scale up error*, Contemporary Mathematics, 295 (2002), pp. 241–251.
- [4] J. GLIMM, S. HOU, H. KIM, Y. LEE, D. SHARP, K. YE, AND Q. ZOU, *Risk management for petroleum reservoir production: A simulation-based study of prediction*, J. Comp. Geosciences, 5 (2001), pp. 173–197.
- [5] J. GLIMM, S. HOU, H. KIM, D. SHARP, AND K. YE, *A probability model for errors in the numerical solutions of a partial differential equation*, CFD Journal, 9 (2000).
- [6] J. GLIMM, S. HOU, Y. LEE, D. SHARP, AND K. YE, *Prediction of oil production with confidence intervals*, SPE 66350, Society of Petroleum Engineers, 2001. SPE Reservoir Simulation Symposium held in Houston, Texas, 11-14 Feb.
- [7] ———, *Solution error models for uncertainty quantification*, Contemporary Mathematics, (2003). Submitted. SUNYSB preprint 02-16. LANL preprint LA-UR: 02-5987.
- [8] J. GLIMM AND D. H. SHARP, *Stochastic partial differential equations: Selected applications in continuum physics*, in Stochastic Partial Differential Equations: Six Perspectives, R. A. Carmona and B. L. Rozovskii, eds., Mathematical Surveys and Monographs, American Mathematical Society, Providence, 1997.
- [9] ———, *Stochastic methods for the prediction of complex multiscale phenomena*, Quarterly J. Appl. Math., 56 (1998), pp. 741–765.
- [10] ———, *Prediction and the quantification of uncertainty*, Physica D, 133 (1999), pp. 152–170.
- [11] T. WALLSTROM, M. A. CHRISTIE, L. J. DURLOFSKY, AND D. H. SHARP, *Effective flux boundary conditions for up scaling porous media equations*, Transport in Porous Media, 46 (2000), pp. 139–153.

- [12] T. WALLSTROM, S. HOU, M. A. CHRISTIE, L. J. DURLOFSKY, AND D. H. SHARP, *Accurate scale up of two phase flow using renormalization and nonuniform coarsening*, Computational Geoscience, 3 (1999), pp. 69–87.
- [13] —, *Application of a new two-phase upscaling technique to realistic reservoir cross sections*, Proceedings of the SPE 15th Symposium on Reservoir Simulation, (1999), pp. 451–462. SPE 51939.
- [14] T. WALLSTROM, S. HOU, M. A. CHRISTIE, L. J. DURLOFSKY, D. H. SHARP, AND Q. ZOU, *Application of effective flux boundary conditions to two phase upscaling in porous media*, Transport in Porous Media, 46 (2000), pp. 155–178.