1 Motivation

1.1 Darcy's Law

In 1856, French engineer Henri Darcy concluded a series of experiments by which he deduced a simple linear relationship between the instantaneous discharge rate through a porous material and the pressure gradient over a certain distance.

$$Q = k(\kappa)\Delta P$$

Where Q is instantaneous flow, $\Delta P = P_l - P_0$ is the pressure gradient and k is a function of permeability, κ. Such simple relationships are abundant: this formula is analogous to Ohm's Law of electrical current, Fourier's Law of heat conduction or Fick's law of diffusion across a permeable membrane. Preserved are qualities which we intuitively wish to embody: if there is no pressure gradient over a distance, no flow occurs; if a pressure gradient does exist, flow will occur from high to low pressures; the greater the pressure gradient, the greater the discharge rate; the discharge rate will be different through different materials, even if the pressure gradient is the same.

Though Darcy derived his formula experimentally, later, it was theoretically derived from the Navier-Stokes equation via homogenization.

$$Q = \frac{A\kappa}{\mu L}\Delta P$$

Where μ is the viscosity, L is length and A is the cross-sectional area to flow. Thus, Darcy's Law replaces the description of a heterogeneous material with a homogeneous one with an identical discharge rate for a particular pressure gradient.

1.2 Objective

Although more accurate methods do exist for computing permeability, limited data and resources may make the application of Darcy's law the most efficient option. Petroleum engineering and geosciences make extensive use of this relationship when approximation and educated hypothesis are satisfactory. As a general rule, for a hydrocarbon or groundwater resource to be exploitable without stimulation, the permeability must be greater than 100 mD, where 1 darcy is $10^{-12}m^2$. Attempting to calculate the true value of permeability for an area as vast as the site of an oil reservoir would be impractical, if not impossible. Rather, geoscientists will calculate the permeability of a handful of samples taken at different depths and estimate the order of magnitude, thereby generalizing for the site.

What we seek is an methodology for selecting a value for k, the function of permeability, from a small sampling of data which maximally represents the system as a whole.

2 Numerical Solutions

Ideally, we would have access not only to the data derived from experimentation, but also to knowledge of the underlying reality: a mapping of every particle in the system with which we could validate a mathematical model by re-running the experiment, controlling for κ and letting Q vary. This is an unreasonable request; however, we can simulate it by generating a geometry and solving for fluid flows across a pressure gradient. In work done by Graveleau[3], the finite element package, Comsol Metaphysics, was used in tandem with Matlab's statistical toolboxes to just this end. Large domains, as seen below, were defined by

$$f(x) = \begin{cases} -0.5 & -2.5 < x < -1.8 \\ -0.15x^4 + 0.13x^3 + 0.65x^2 - 0.3^x + 0.5 & -1.8 < x < 1.8 \\ 0.5 & 1.8 < x < 2.5 \end{cases}$$

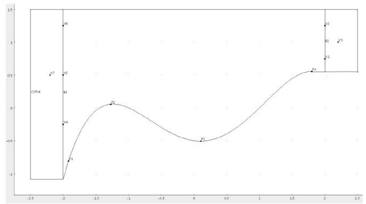
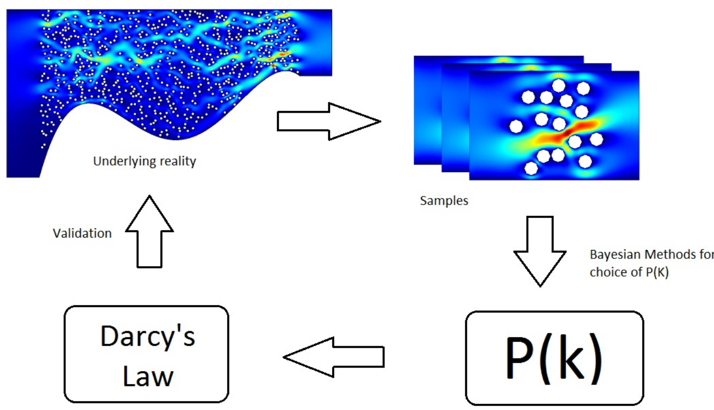


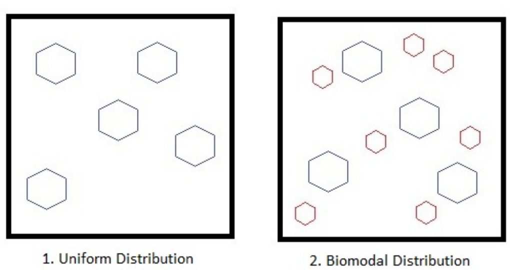
Figure 1: Defined Domain

Inside this surface, geometries were generated, allowing particles to overlap boundaries. Subjected to different pressure gradients, flow was computed at the end regions with *no-slip* boundary conditions (in which velocity is set to zero) at the top and bottom. From this domain, we sample small, rectangular units and perform similar computations. Analogous to sample matter collected in-field, we hope to determine what the large scale permeability will be by observing these units. Unlike experimentation, however, we will then be able to verify our claims by directly solving for flow. The process can be illustrated thusly



3 Methodology

For this report, three different geometries will be considered: uniformly distributed particles and a mixture of two different particles, for each of which we will have twenty five samples at three different pressure gradients (75 each, 150 total).



Although we have access to the underlying reality, for every sample we will behave as though the Comsol routine is a black box which outputs a set of data, Q, given three different pressure gradients, $\Delta P = 9; 999; 99,999$. These values were chosen for ease of computation, but also in an attempt to capture the linear relationship between pressure gradient and instantaneous flow. At large pressure gradients, $\Delta P = 99,999$, we expect to see characteristics of turbulence.

We begin by the straightforward calculation of vector k,

$$k = Q^{-1}\Delta P$$

To get any further we must make hypothesis about the system. At the heart of Darcy's law is the idea of replacing of a heterogeneous medium for a homogenous one for which the flow is equivalent. Therefore we seek a single value for k. Yet we know this value will have an associated error when attempting to generalize about the whole system, therefore we seek a range of values that k might reasonably hold. This suggests a distribution. How should we approximate k? One approach would be to take the sample mean and variance, but we will see that this proves to be too simplistic after attempting to validate the model using the large-scale domain. It would be better to define k as coming from some distribution. And what to make of the vector Q? Does the experimenter have some knowledge of the margin of error in his apparatus, σ_e? Is there a significant ratio between the error margins of k and the experimental error?

I will apply Bayesian statistical methods to the following scenarios:

- k is a parameter, σ_e is some known value
- k is a parameter, σ_e is a parameter
- k is normally distributed, $\bar{k}$, σ_k and σ_e are parameters
- k is derived from multiple Gaussians (in the case of bimodal particles)

4 Results and Analysis

4.1 Uniformly Distributed Particles

4.1.1 Trial 1

In this case, we assume that there is some experimental error for each of the data points. Since we are blinded to any additional information about the methodology under which the data was collected, we begin with the simplest case in which the error is additive. We give our experimenter the benefit of the doubt when he quotes to us a known error for his apparatus, σ_e. Therefore we write

$$Q^d = k\Delta P + E$$

Where

- Q^d is a vector containing the experimental flow data
- k is a parameter
- $E \sim N(0, \sigma_e)$ with σ_e known

Ultimately, we seek a value for k. By Bayes Theorem,

$$P(k|Q^d) \propto P(Q^d|k)P(k)$$

Since we have no expectation as to what values k may occupy,nor its order of magnitude, we will use a log uniform prior, $\frac{1}{k}$. To determine the distribution of $P(Q^d|k)$, we will derive its characteristic function using Fourier Transforms

$$\begin{aligned} F(P(Q|k)) &= F(P(k\Delta P + E|k)) \\ &= F(P(k\Delta P))F(P(E)) \end{aligned}$$

The characteristic functionsof Gaussian distributions have the form $\Phi_x = e^{i\mu - \frac{1}{2}\sigma^2 t^2}$.

Therefore,since we know $E \sim N(0, \sigma_e)$,

$$\begin{aligned} &= F(P(k\Delta P)) * e^{-\frac{1}{2}\sigma_e^2 t^2} \\ &= e^{ik\Delta P} e^{-\frac{1}{2}\sigma_e^2 t^2} \\ &= e^{ik\Delta P - \frac{1}{2}\sigma_e^2 t^2} \end{aligned}$$

We can conclude that $P(Q|k) \sim N(k\Delta P, \sigma_e)$. Therefore,

$$\begin{aligned}
P(k|Q^d) &\propto P(Q^d|k)P(k) \\
&\propto \prod_{n=1}^{N} P(Q_i|k)P(k) \\
&\propto \prod_{n=1}^{N} \left[\frac{1}{\sqrt{2\pi}\sigma_e} e^{\frac{-1}{2}\left(\frac{Q_i - k\Delta P}{\sigma_e}\right)^2} \right] \frac{1}{k} \\
&\propto e^{\frac{-1}{2}\sum_{n=1}^{N}\left(\frac{Q_i - k\Delta P}{\sigma_e}\right)^2} \frac{1}{k}
\end{aligned}$$

Now, let's return to the matter of choosing an appropriate σ_e. The most natural approach is to take the sample variance. This value is very large with respect to most of the values of Q_i, which undermines our data set. It would be better to take different values for σ_e for different pressure gradients, or even the maximal element of these. Erring on the side of caution, we will not consider taking the minimal value. The normalized posterior distributions and corresponding peak values are as follows. Note that, although Bayes prohibits quoting the maximal value as the correct answer, our ultimate goal is to settle on a single value for k, a "best guess". Although simplistic, this trial can give us some insight to the physical nature of the system. Using

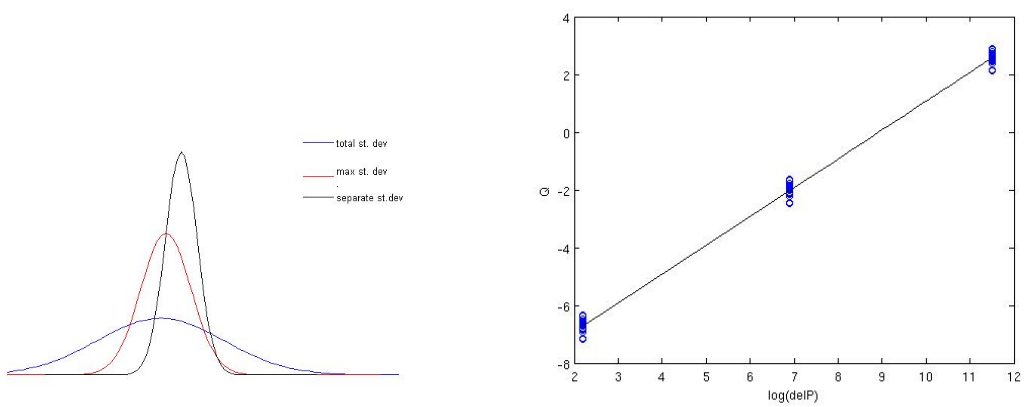


Figure 2: Posterior Distributions for k, Fitted Curve

our estimated value for k to suggest an expected flow Q^{est}, we see that the data is well represented for the smaller pressure gradients. However, once we reach $\Delta P = 99,999$, the differences are erratic. This could well be the effects of turbulent flow, a phenomenon not well captured by such a simplistic, linear system. In subsequent trials, we will accommodate this hypothesis by comparing probabilities in which the high-pressure flow is eliminated with those in which it is remains incorporated.

4.1.2 Trial 2

Perhaps we do not have such good faith in the value for σ_e quoted by our experimenter, and wish to estimate it as well. In this case,

$$Q^d = k\Delta P + E$$

Where

- Q^d is a vector containing the experimental flow data
- k is a parameter
- $E \sim N(0, \sigma_e)$ with σ_e as a parameter

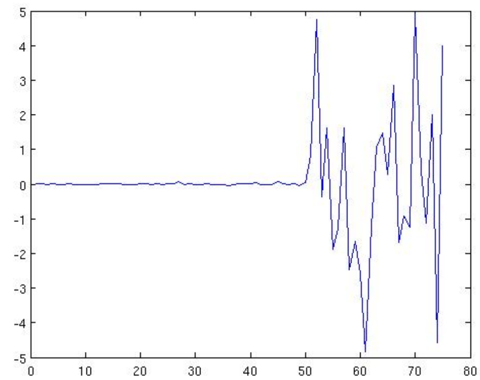


Figure 3: Q(data)-Q(estimate) at increasing pressure gradients

Taking the Fourier Transform as before,

$$\begin{aligned} P(k|Q^d) &\propto P(Q^d|k)P(k)P(\sigma_e) \\ &\propto \prod_{n=1}^{N} P(Q_i|k)P(k)P(\sigma_e) \\ &\propto \prod_{n=1}^{N} \left[\frac{1}{\sqrt{2\pi}\sigma_e} e^{\frac{-1}{2}\left(\frac{Q_i - k\Delta P}{\sigma_e}\right)^2} \right] * \frac{1}{k} * \frac{1}{\sigma_e} \end{aligned}$$

This time, we cannot absorb σ_e into the proportionality. Instead, we will use the following trick to incorporate σ_e terms into the exponential:

$$\sigma_e^{-N} = e^{-N \log \sigma_e} = e^{-\frac{1}{2}N 2 \log \sigma_e} \approx e^{-\frac{1}{2}\sum_{n=1}^{N} 2 \log \sigma_e}$$

Therefore,

$$P(k|Q^d) \propto e^{\frac{-1}{2}\sum\left[\left(\frac{Q_i - k\Delta P_i}{\sigma_e}\right)^2 + 2\log \sigma_e\right]} * \frac{1}{k} * \frac{1}{\sigma_e}$$

To find the best linear fit to our data set, we seek to maximize $P(Q^d|k)$, thereby minimizing the power of the exponential. We expect that the presence of the logarithmic term will prevent σ_e from becoming too large. However, we have no real interest in its actual value; we only care about the posterior distribution of k. Therefore, we will marginalize over σ_e. As suggested in the first trial, we have experimented with the idea

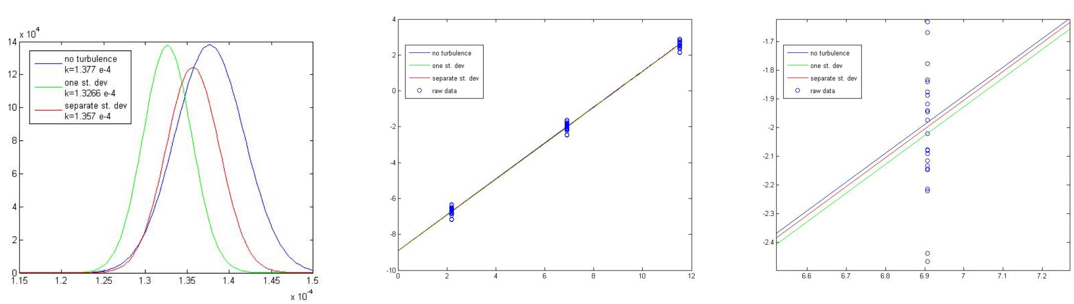


Figure 4: Posterior Distributions for k, Fitted Curve

that σ_e could be constant for the whole system, or take on different values at different pressure gradients. Additionally, we have incorporated the hypothesis that trimming high pressure flows from the data set will alter our estimation of k. Fitting curves to the data set, we can see that the scenarios produce very similar results. It is only on a very small scale that we perceive differences in these lines, potentially negligible in the fields of application for this experiment.

4.1.3 Trial 3

Darcy's Law homogenizes a material to a single value of k, however, it seems plausible that the porosities of each sample could be drawn from a distribution. After all, we do not expect a mineral deposit to be truly homogenous; only that its overall properties can be estimated as such on large enough scales. Thus, we take

$$Q^d = k\Delta P + E$$

Where

- Q^d is a vector containing the experimental flow data
- $k \sim N(\bar{k}, \sigma_k)$
- $E \sim N(0, \sigma_e)$

Deriving the characteristic function via Fourier Transforms,

$$\begin{aligned} F(P(k\Delta P + E)) &= F(P(k\Delta P))F(P(E)) \\ &= e^{i\bar{k}\Delta P - \frac{1}{2}\sigma_k^2 \Delta P^2 t^2 - \frac{1}{2}\sigma_e^2 t^2} \\ &= e^{i\bar{k}\Delta P - \frac{1}{2}(\sigma_k^2 \Delta P^2 + \sigma_e^2) t^2} \end{aligned}$$

And we conclude that

$$P(Q|k) \sim N(\bar{k}\Delta P, \sqrt{(\sigma_k^2 \Delta P^2 + \sigma_e^2)}) = N(\bar{k}\Delta P, \sigma_Q)$$

$$\begin{aligned} P(k, \sigma_Q | Q^d) &\propto P(Q^d | \bar{k}, \sigma_Q) P(k) P(\sigma_Q) \\ &\propto \prod_{n=1}^{N} P(Q_i|k) P(k) P(\sigma_Q) \\ &\propto \prod_{n=1}^{N} \left[\frac{1}{\sqrt{2\pi}\sigma_Q} e^{\frac{-1}{2}\left(\frac{Q_i - k\Delta P}{\sigma_Q}\right)^2} \right] * \frac{1}{k}\frac{1}{\sigma_Q} \\ &\propto e^{\frac{-1}{2}\sum\left[\left(\frac{Q_i - k\Delta P_i}{\sigma_Q}\right)^2 + 6\log\sigma_Q\right]} * \frac{1}{k} * \frac{1}{\sigma_Q} \end{aligned}$$

We will keep the hypotheses of the second trial, marginalizing over σ_Q to produce the following

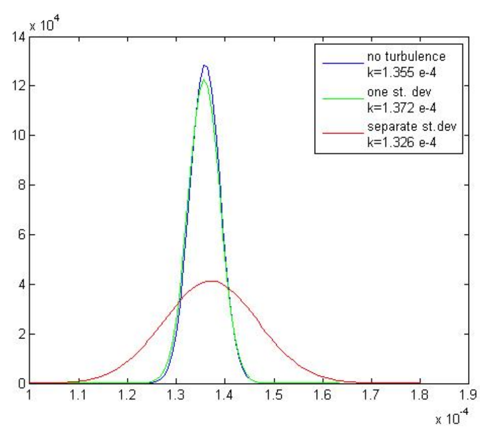


Figure 5: Posterior Distribution of K

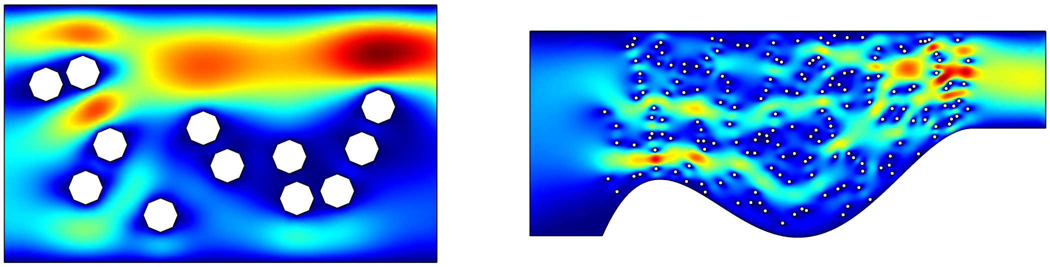


Figure 6: Resulting Flows of A Sample and the Large Domain

4.1.4 Large Scale Comparison

Till now, we have been working with small samples of twelve particles in a rectangular domain. How well do the posterior distributions of these samples relate to the large-scale domain, the behavior of which is our ultimate interest? Staying true to our experimental scenario, we will compare our predicted flows of trial 3 to the "real" outcomes. Giving Comsol the large domain, specifying the number of particles and their geometries, and setting $\Delta P = 999$, we have the following flows:

The resulting value for flow is $Q = 0.09745$, which corresponds to a porosity of $k = 9.74e-5$ a slightly lower value than what we were expecting. As we may have guessed, the best prediction came from the model which incorporated different standard deviations for each pressure gradient.

4.2 Bimodal Distribution of Particles

Now we will consider the case in which the particles are bimodal in their distribution. Since we our blinding ourselves from this knowledge, we will keep the posterior probability distributions as calculated before:

$$P(k|Q^d) \propto e^{\frac{-1}{2}\sum_{n=1}^{N}\left(\frac{Q_i-k\Delta P}{\sigma_e}\right)^2}\frac{1}{k}$$

for k as a parameter, σ_e known,

$$P(k|Q^d) \propto e^{\frac{-1}{2}\sum\left[\left(\frac{Q_i-k\Delta P_i}{\sigma_e}\right)^2+2\log\sigma_e\right]} * \frac{1}{k} * \frac{1}{\sigma_e}$$

for both k and σ_e as parameters, and

$$P(Q^d|k) \propto e^{\frac{-1}{2}\sum\left[\left(\frac{Q_i-\bar{k}\Delta P_i}{\sigma_Q}\right)^2+6\log\sigma_Q\right]} * \frac{1}{k} * \frac{1}{\sigma_Q}$$

for $k \sim N(\bar{k}, \sigma_k)$, $E \sim N(0, \sigma_e)$, σ_k,σ_e,$\bar{k}$ are parameters.
Following similar procedures as before, we obtain results of the same nature from the new data set.

Let's say that we have reason to believe that our samples do, in fact, come from some heterogeneous

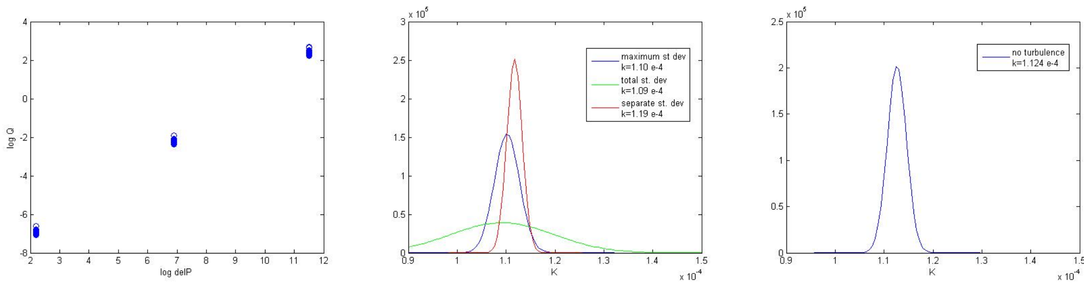


Figure 7: Sampling of results from Bimodal Trials

mixture. This is a realistic expectation of our real-world scenario, since many varieties of shale are composed of unevenly distributed clay minerals and quartz. We may hypothesize the following.

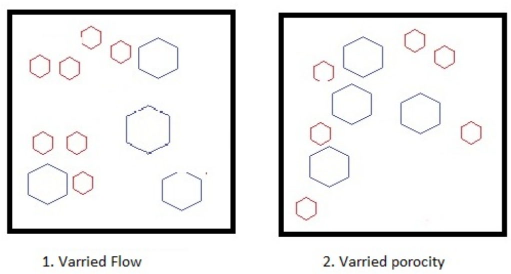


Figure 8: Illustrations of Hypotheses

If the particles are unevenly distributed throughout the material, while keeping porosity constant, the flow will be different along the sample, though the final computation constant. Therefore, flow would best be described as a series of Q_is. Conversely, if the particles are distributed in groupings, the porosity may vary through the sample as some k_is, though the flow will, ultimately, remain the same. To reflect this hypothesis, we will attempt to fit multiple Gaussians to our data, utilizing EM methods for a set number of iterations. Taking a two Gaussian mixture for our model captures a very important physical point. Without

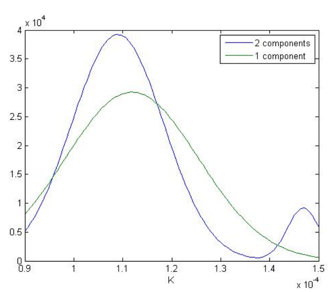


Figure 9: Varying K

prior understanding of turbulence, we might conjecture that, at high pressure gradients, fluid particles will be as much a hindrance to flow as the solid particles and therefore expect higher values of k in the higher values ΔP. This idea is not supported by the data. Rather, we see relatively even ranges of k for all three pressure gradients. For the case of flow, we see a dramatic difference in the expected distributions of Q for a

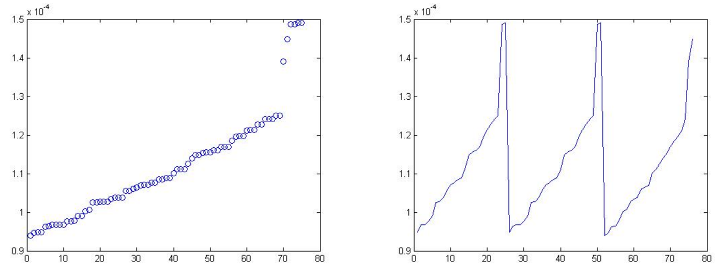


Figure 10: Sorted values of K from data set

single component, but very little in the cases above two. Fitting multiple Gaussians pushes this value higher than we might expect, and goes against our instincts about the nature of Darcy's Law.

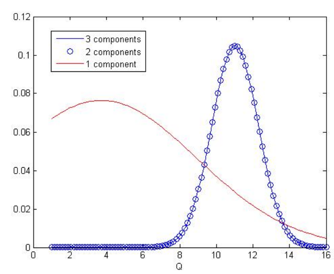


Figure 11: Varying Q

4.2.1 Large Scale Comparison

As with the uniform case, after designating $\Delta P = 999$ and the geometries and number of particles (but not their densities), we pass the problem to Comsol to produce the following flows.

Comparing the outcomes of the sample and large-scale domain, we get insight for interpreting the results

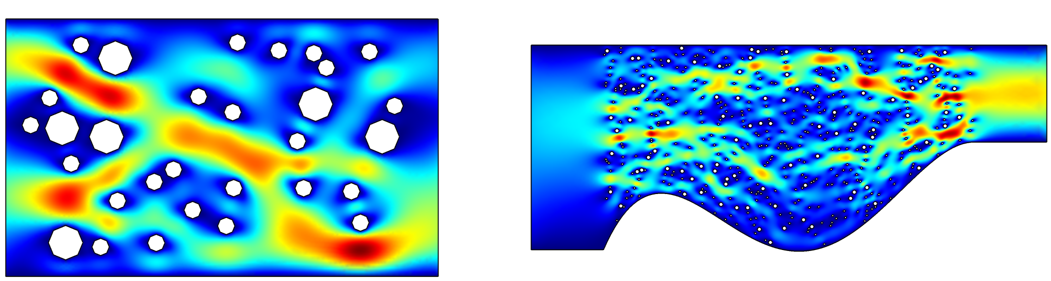


Figure 12: Resulting Flows from Sample and Large Domain

of the EM methodology as discussed. In the sample flow, we observe paths of intense,unbroken flow through

empty channels where no particles are placed. This would seem to give merit to the idea of fitting multiple Gaussians. However, in the case of the large scale, these paths become broken, as the channels are split by a meshwork of particles. The data we were given did, infact, suggest the possibility of a two-Gaussian fit, but the larger physical system simply does not support it.
Intuitively, we believe that the materials to be studied will not have an even distribution of particles, that the porosity will vary throughout the domain. This is supported by the data and our analysis of the Comsol flows.

5 References

1. Régis Cottereau, J. Tinsley Oden, Todd, A Oliver, Ernesto Preudencio and Serge Prudhomme. *Discussions on model errors and model validation.*2004.

2. Daniela Calvetti and Erkki Somersalo. *Survey and Tutorials in the Applied Mathematics Sciences, Vol 2, Introduction to Bayesian Scientific Computing.* 2007 Ed Springer.

3. Mattieu Graveleau, Serge Prudhomme, Corey Bryant. *On how to include Bayesian inference in mechanical computation.* 2010.

4. Freeze, Allen. *Henry Darcy and the Fountains of Dijon***Groundwater vol 32 no 1**. 1994 (23-30)

5. Masa Prodanovic. *Applications of Momentum balance to flow in porous media* **Transport Phenomena** Spring 2010.