Black-Box Reduced-Order Prediction by Proper Orthogonal Decomposition with Interpolation

S. Ali* [a], and M. Damodaran* [b],

*Singapore-Massachusetts Institute of Technology Alliance (SMA)
Nanyang Technological University, Nanyang Avenue, Singapore 639798

[a]Department of Mathematics and Science
Singapore Polytechnic, 500 Dover Road, Singapore 139651

[b]School of Mechanical and Aerospace Engineering
Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798

Abstract:

Many commercial and third-party flow simulation solvers which are often used for flow modelling in engineered systems can be regarded as black box solvers because the internal structures of the solvers are not transparent to the user. This work attempts to show that a proper orthogonal decomposition (POD) method with interpolation methods to form a black-box reduced-order model solver which does not require any *a-priori* knowledge of the actual system. The combination of POD and interpolation methods is demonstrated as an efficient and economical tool for computing approximations of higher dimensional solutions of black box flow solvers. This capability of the reduced-order model to successfully approximate the higher dimensional solutions is illustrated with a couple of applications on engineered systems. In this paper, this efficient and economical reduced-order model solver is shown to be an effective tool in the field of inverse design of the hard disk drive and uncertainty optimization of the groundwater contaminant transport model. It is also shown that the presence of non-linearities in the

[a] SMA Research Fellow; Currently Lecturer, Singapore Polytechnic; Email: shidrati@sp.edu.sg; Tel: (65) 68704688.
[b] Associate Professor and SMA Faculty Fellow; Email: mdamodaran@ntu.edu.sg; Tel: (65) 67905599.

higher dimensional solutions corresponding to these examples is captured by the non-linear interpolation methods used in the reduced-order approximations.

Keywords: Proper Orthogonal Decomposition, Optimization, CFD, Hard Disk Drives, Groundwater Contaminant Transport.

1. Introduction

Although advances in high performance computing technologies have rendered the solution of high fidelity simulation models for engineered systems, the use of these methods for routine design of such systems requires enormous computational effort, remains prohibitively expensive and results in poor turnover times. Reduced-order modeling is an attractive alternative approach for developing low order computational models which not only capture the richness of the solutions from high fidelity models but also permit efficient solutions and quick turnover which are attractive for designers. A plethora of reduced-order models such as the reduced-basis methods [1 - 3], Krylov subspace based methods [4 - 6], and proper orthogonal decomposition techniques [7 - 9] have been developed and applied to practical engineered systems to obtain low-order approximations of full high fidelity solutions in many fields. The construction of low-order approximation models basically assumes that the engineered systems can be described by fewer degrees of freedom as compared with the high fidelity models while maintaining accuracy of the solutions and resulting in computational economy. While research in computational science and engineering has resulted in the development of a number of well established high fidelity engineering simulation software such as *Fluent* [10], *Abaqus* [11], *Modflow* [12] which are routinely used in both industries and research facilities, most of these softwares function as black boxes. The users usually interact with these software via scripts and rarely have access to the source codes which implement the algorithms for the simulation models. In such cases, it is difficult to ob-

tain low-order models using existing model reduction methods, which typically employ a Galerkin projection technique that necessitates knowledge of the underlying mathematical model and algorithms. Ly and Tran [13] proposed a different technique of obtaining low-order models by combining the use of proper orthogonal decomposition (POD) and interpolation techniques, called the PODI technique. They have successfully demonstrated the capability of this technique on several problems where the mathematical models are assumed to be unknown. Bui-Than [14] noted that a limitation of this method is that coefficients have to be smooth functions of the parameters in order for the interpolated results to be reliable.

Assuming the coefficients are smooth functions of the parameters, it is then necessary to estimate the number of POD modes necessary to achieve a desired accuracy. Ly and Tran [13] proposed to use the relative energy of the system, as captured by the ratio between the eigenvalues of selected POD modes and the total sum of the eigenvalues of all the POD modes, in order to choose the number of modes needed to achieve the desired accuracy. The coefficients of each mode for different parameter values can then be evaluated using the interpolation method. The number of modes needed to achieve the desired accuracy is usually much smaller that the dimension of the high-dimensional problem and thus, when combined with the interpolation method, yields an economical low-order approximation to the high-dimensional solution. Thus the reduced-order model enables a rapid computation of solutions which is necessary for facilitating quick design optimization practice. In this paper, a simple inverse design problem and an uncertainty optimization problem are explored to illustrate the effective application of the PODI method in different optimization problems governed by differing physical systems. In the inverse design problem considered, the physical system is represented by flow characteristics within a hard disk drive (HDD) enclosure which are governed by the

Navier-Stokes equations while in the uncertainty optimization problem, the physical system is represented by the groundwater contaminant transport equations. The governing equations are used only for the purpose of generating the ensemble of snapshots. The PODI technique is then used to generate the reduced-order models to be used by the optimization solvers. The paper is organized as follows. The basic PODI method is first outlined in section 2. Description of the two physical problems, their corresponding design optimization problems and the results are addressed in section 3, before concluding.

2. Proper Orthorgonal Decomposition with Interpolation (PODI)

The method of snapshots [8] is used in this paper to construct the POD modes. In this method, the POD modes are expressed as a linear combination of the "snapshots" of state variables of the high-fidelity or experimental system. We consider a high-fidelity computational model with state variables contained in the vector $U(\vec{x},\mu)$, which has dimension n. The ensemble of snapshots in the parameter space, $\mu_i \in \Re^b$, can be expressed as $\{U(\vec{x};\mu_i) \in \Re^n, i=1,\ldots,M\}$, where M is the number of snapshots. Each POD mode can be expressed as linear combination of the snapshots, i.e.

$$\phi_j = \sum_{i=1}^{M} \beta_j^i U_i \tag{1}$$

where ϕ_j is the jth POD mode and $U_i = U(\vec{x},\mu_i)$ denotes the ith snapshot. It can be shown [12] that the set of coefficients, $\beta_j = [\beta_j^1, \beta_j^2, \ldots, \beta_j^M]$, satisfies the eigenproblem

$$R\beta_j = \lambda_j \beta_j \tag{2}$$

where $R_{i,j} = (1/M)(U_i, U_j)$ and $(\cdot,\cdot)$ denotes an inner product. The jth eigenvalue of R, λ_j, determines the importance of the jth POD mode. The relative "energy" (in the 2-norm sense) of POD mode i is given by $\lambda_i / \sum_{j=1}^{M} \lambda_j$, and is often used as a criterion for selecting

N, the number of basis functions to retain in the low-order model. Then the *N*th-order POD approximation of the state $U(\vec{x};\mu_k)$ can be expressed as

$$U_N(\vec{x};\mu_k) = \sum_{i=1}^{N} \alpha_i(\mu_k)\phi_i(\vec{x}) \tag{3}$$

for some $N < M$, where the coefficient $\alpha_i(\mu_k)$ gives the contribution of the *i*th POD mode.

Each POD modal coefficient, α_i, for all *N* coefficients, corresponding to the ensemble of snapshots are easily computed as

$$\alpha_i(\mu_k) = (\phi_i, U(\vec{x};\mu_k)), \quad \text{k = 1,…,M} \tag{4}$$

Then in the POD with interpolation (PODI) method, each POD modal coefficient, α_i, at a given μ_{new}, $\alpha_i(\mu_{new})$, can be obtained by an interpolation method on the values of $\{\alpha_i(\mu_k), k = 1,\ldots,M\}$ and thus $U(\vec{x};\mu_{new})$ can be approximated by $\hat{U}_N = \sum_{i=1}^{N} \hat{\alpha}_i(\mu_{new})\phi_i$ where $\hat{\alpha}_i(\mu_{new})$ is the interpolated value of $\alpha_i(\mu_{new})$. For this investigation, three different interpolation methods are used, namely linear, cubic Hermite interpolation, and cubic spline interpolation [15 – 16]. The relative energy, given by $\lambda_i / \sum_{j=1}^{M} \lambda_j$, is used to determine the relevance of POD modes used. Only POD modes with energy higher than a given tolerance, and their coefficients, are used in evaluating the approximation of the new $U(\vec{x};\mu)$.

3. Description and results of the optimal design problem for the physical systems

Optimal design in engineering problems often involves numerous evaluations of the physical systems in the form of the high-dimensional mathematical models. Even with the advance of computers and computing power, the computational costs and time for evaluation these high-dimensional models are prohibitive in optimal design. The use of

reduced-order models is therefore a welcome substitution in optimal design problems. In this paper, two examples of optimal design problem in engineering in the form of flow characteristics in hard disk drives and groundwater contamination transport are used to demonstrate the effectiveness of using the PODI model in substitution of the high-dimensional models. In the following subsections, the mathematical model of the physical systems and their corresponding optimization problems will be described briefly but for the purposes of the current study, these mathematical models are assumed to be unknown. The results of the PODI model and the optimization problems are also discussed.

3.1 Flow Characteristics in the Hard Disk Drive (HDD)

The airflow within Hard Disk Drive (HDD) enclosures as shown in Figure 1 is described by incompressible Navier-Stokes equations:

$$\frac{\partial U_i}{\partial x_i} = 0, \tag{5a}$$

$$\frac{\partial U_i}{\partial t} + U_j \frac{\partial U_i}{\partial x_j} = -\frac{1}{\rho}\frac{\partial P}{\partial x_i} + \nu \frac{\partial^2 U_i}{\partial x_j \partial x_j}, i, j = 1,2,3 \tag{5b}$$

where ν is the kinematic viscosity. The incompressible Navier-Stokes equations are an adequate representation of the flow within the enclosure as the Knudsen number is less than 0.01 and the Mach number less than 0.1 as discussed in Ali. et. al. [17]. The problem is treated as an isothermal problem. The Navier-Stokes equations are then solved using the commercial code, *Fluent* [10]. Since the exact codes and algorithms used to solve for the airflow within the HDD enclosures are unavailable to the user, it can be assumed that the mathematical models and algorithms are unknown.

The airflow in the HDD enclosures is driven by the rotation of the media disk used to store data. As such, the single parameter, $\mu \in \Re^1$, for this problem is the rotational speed of the media disk. The field variables of interest in this study are the pressure and

the flow velocities. For the HDD enclosure problem, one of the concerns of the industry is the airflow induced vibration of the actuator arm carrying the read-write head which traverses in the radial direction across the disk surface. Since one of the factors causing the vibrations of the arm is the pressure acting on the arm, a simplified inverse design problem is considered by specifying a target pressure distribution on the arm and requiring the determination of the rotational speed of the media disk that will yield this target pressure. This inverse design problem can then be formulated as a least squares minimization where the objective function, $J(\mu) = \left\| P_{desired} - \hat{P}(\mu) \right\|_2^2$, subject to $\mu \in [0.0, 7000.0]$ rpm where $\hat{P}(\mu)$ is evaluated using the PODI approximation method.

3.1.1 PODI approximation to the high dimensional model

For the high dimensional model, 212,460 nodes are used in the mesh to model the HDD enclosure. The boundary conditions are prescribed such that the no slip boundary condition applies at all the surfaces of the enclosure. The media disk is then set to spin at a prescribed rotational speed until steady state has been reached. An ensemble of 15 snapshots ranging between 0.0 rpm and 7000.0 rpm in steps of 500 rpm for the rotational speed is constructed by numerically solving the Navier-Stokes equations using *Fluent* [10] with a convergence criteria of $O(10^{-3})$ for the residual errors.

Figure 2 shows the contour plots of the pressure and velocity distributions in the HDD enclosure for two of the 15 snapshots in the ensemble corresponding to 4000 and 6500 rpm. Using the 15 snapshots, the POD modes are obtained. Based on the relative energies of the POD modes, $\lambda_i / \sum_{j=1}^{M} \lambda_j$, the first two modes are able to capture 99.9997% of the relative energy. For the present study in this section, the first three modes are then used to obtain the low-order solution as they represent the most significant component of the snapshot data. The assumption of smoothness of the coefficients, necessary for

the PODI model, holds in this example as shown in Figure 3 where the variations of the first three coefficients in response to the change in rotational speeds are plotted. From Figure 3, the first three coefficients of the pressure and velocity are observed to be smooth functions of the rotational speed, i.e. $\alpha_i \in C^2[\mu_{\min}, \mu_{\max}]$. As such, the PODI method can be used to predict the field variables at various rotational speeds in the range of $[0.0,7000.0]$.

Table 1 shows the percentage relative errors of the low-order results vis-à-vis the high-dimensional results obtained from Fluent for the rotational speeds of 100, 1200, 3600, 5400, and 6750 rpm. Three different interpolation methods, namely linear, cubic spline, and cubic Hermite interpolations are used to predict the low-order approximations and their errors are tabulated in Table 1 for the above rotational speeds. It is observed that the spline and Hermite polynomials yield more accurate results than those predicted by linear interpolation in the case of the pressure than in the case of flow velocity. Part of the explanation for this error pattern could be deduced from Figure 3 where it can be seen that the coefficients of the first pressure mode, α_1, is slightly non-linear with respect to the rotational speed in Figure 3(a) unlike the linear slope of the coefficients of the first y-velocity mode in Figure 3(b). The linear interpolation is likely to work better with the linear coefficients of the flow velocity mode than the non-linear coefficients of the pressure mode.

The results in Table 1 also demonstrate that using only the first three modes, which capture more than 99.997% of the relative energy is not enough to decrease the errors. The relative errors, given as $E_{true} = \dfrac{\left\| U - \hat{U}_N \right\|_\infty}{\left\| U \right\|_\infty}$, are still very large for some of the rotational speeds such as 100 and 5400 rpm. The question then remains as to the number of

modes needed to reduce the relative errors to an acceptable level. Using cubic Hermite interpolation and different numbers of POD modes, the plot of percentage relative errors for the low-order solution of pressure against N, the number of POD modes, for 100, 1200, and 3600 rpm are presented in Figure 4. One observes in Figure 4(a) that the increase in the number of POD modes used results in a decrease in relative error until the sixth mode. After the sixth mode, the insignificant decrease in relative error indicates that the contribution of the higher modes than the sixth mode to the accuracy of the reduced-order model is negligible for the approximation of the pressure within the HDD in Figure 4(a). However, in Figure 4(b), there are still significant reductions in relative errors of the velocity approximations of the PODI method as the number of POD modes used are increased to the ninth mode. Therefore, based on the relative errors plotted in Figure 4, it can be seen that the inverse design problem uses 10 modes to obtain the low-order solutions of the pressure to ensure better accuracy overall.

3.1.2 Inverse design problem

The desired pressures corresponding to rotational speeds of 250, 1750, 4250, and 6250 rpm are chosen for $P_{desired}$ and the least squares minimization problem, $J(\mu) = \left\| P_{desired} - \hat{P}(\mu) \right\|_2^2$, with the linear constraints, $\mu \in [0.0, 7000.0]$, for these desired pressures are solved. The least squares minimization problem is a convex optimization problem and can be readily solved using Levenberg-Marquardt method [18] which uses a search direction that is a cross between the Gauss-Newton direction and the steepest descent.

The results of these optimization problems are tabulated in Table 2. Most of the optimal results obtained are very close to the desired rotational speeds with less than 1% percentage errors. This relatively good accuracy is achieved even with the errors in the

PODI approximations. Such accuracies signify that the PODI technique for this example problem is adequate for approximating the solutions of the high-dimensional model. Using Levenberg-Marquardt method to solve for the least squares minimization problems, the number of evaluations of $\hat{P}(\mu)$ for the four desired pressures ranges from 10 to 24, depending on how close the starting point is to the optimal point. Based on the average time given in Table 2, the PODI method takes on average 10 seconds for each evaluation of $\hat{P}(\mu)$. As each evaluation of $\hat{P}(\mu)$ using *Fluent* [10] will take about 4 hours, it can be seen that the reduced-order PODI model is able to achieve a considerable reduction in computational costs while yielding good optimal results.

3.2 Groundwater Contaminant Transport

The groundwater contaminant transport model based on an example in Chiang [19], is constructed within the confines of a valley where a river flows through the valley, surrounded in the north and south by impermeable granite hills as shown in the map in Figure 5. The dimensions of the valley are specified in Table 3. Three pumping wells, each operating to deliver 400 m^3 of water per day, shown in Figure 5 are used to supply water to the inhabitants of the valley while two observation wells, located near the riverbanks, are used to monitor the quality of the groundwater. The river, which flows from the west (upstream) to the east (downstream), is part of the first permeable unconfined aquifer layer. Below the first layer is a confined aquifer and separating the two is a thin silty layer. All three layers are defined by their respective horizontal hydraulic conductivities, vertical hydraulic conductivities, and effective porosities given in Table 3. Table 3 also shows other salient properties and dimensions of all thre three layers. The properties of the river are given in Table 4. The flow processes in each of the three layered groundwater flow model can be described by

$$k_x{}^i \frac{\partial^2 h_i}{\partial x^2} + k_x{}^i \frac{\partial^2 h_i}{\partial y^2} + k_z{}^i \frac{\partial^2 h_i}{\partial z^2} - G^i = 0; \quad i = 1,2,3, \tag{6}$$

where h_i is the hydraulic head, $k_x{}^i$ is the horizontal hydraulic conductivity, $k_z{}^i$ is the vertical hydraulic conductivity, and G^i is the source/sink term at layer *i*, such as the three pumping wells and the river. Constant hydraulic heads are found both upstream (west) and downstream (east) and are denoted in Figure 5 as the blue bands.

Pollutants with certain concentration levels are released upstream of the river and their spread through the valley can be described by

$$D_x{}^i \frac{\partial^2 C_i}{\partial x^2} + D_x{}^i \frac{\partial^2 C_i}{\partial y^2} + D_z{}^i \frac{\partial^2 C_i}{\partial z^2} - v_x{}^i \frac{\partial C_i}{\partial x} - v_y{}^i \frac{\partial C_i}{\partial y} - v_z{}^i \frac{\partial C_i}{\partial z} + F_i = 0; \quad i = 1,2,3, \tag{7}$$

where C_i is the concentration of the pollutants dissolved in the groundwater, $D_x{}^i$ is the horizontal hydrodynamic dispersion coefficient, $D_z{}^i$ is the hydrodynamic dispersion coefficient, and F_i is the source/sink term at layer *i*. The seepage velocities, given by v_x, v_y, and v_z, are evaluated using

$$v_x{}^i = k_x{}^i \frac{\partial h_i}{\partial x}, v_y{}^i = k_x{}^i \frac{\partial h_i}{\partial y}, v_z{}^i = k_z{}^i \frac{\partial h_i}{\partial z}, \qquad i = 1,2,3 \ , \tag{8}$$

where the hydraulic head for each layer *i*, h_i, are obtained from Equation (6).

The groundwater flow model described by Equation (6) is solved using *PMWinPro* solver [19] with 6240 nodes with a convergence criteria of $O(10^{-3})$ for the residual errors while the pollutant transport model described in Equation (7) is solved using *MT3D* solver [20] with a similar convergence criteria of $O(10^{-3})$ for the residual errors. Based on Equations (6) – (8), the spread and level of the pollution concentration in the valley area are indirectly dependent on the hydraulic conductivities amongst other parameters. Therefore the field variable of interest in this example is the concentration of the pollutants dispersed in the valley and the set of parameters, μ, chosen are the initial pollutant

concentration released upstream of the river and the vertical hydraulic conductivities, k_z^i, at level i=1,2,3.

It is often difficult to know the exact parameters such as the hydraulic conductivities, the hydrodynamic dispersion coefficients, and porosities of the soil. The parameters are often estimated based on measurable data such as water levels in the observation wells. The uncertainty in the parameter values are further compounded by the noise in the data obtained. For exact values of the parameters, the pollutant concentrations in the vicinity of the pumping wells can easily be evaluated using either the high-fidelity solver or the PODI method. However with the uncertainty in the parameters, the pollutant concentrations in the vicinity of the pumping wells have to be approximated by a maximization problem that accounts for this uncertainty.

Thus in this study, a simple uncertainty optimization problem is constructed. Assuming that all other parameters are know and are constants for the different soil layers and types and only the vertical hydraulic conductivities are unknown, pollutant concentrations data from the observation wells are measured for different pollutant concentrations released into the river, upstream. Using the statistical mean and standard deviations, the data is able to yield the possible range of vertical hydraulic conductivities for the soil layers implicitly. For a given new pollutant concentration released into the river, the average pollutant concentration for this unknown vertical hydraulic conductivity is approximated by the maximum average pollutant concentration in the vicinity of the three pumping wells for this possible range of vertical hydraulic conductivities. The uncertainty optimization problem can therefore be written as

$$\min \quad \gamma \tag{9}$$

s.t.

$$\begin{gathered} C_1\left(C_0^{new}, k_z{}^1, k_z{}^2, k_z{}^3\right) \le \gamma \\ C^-\left(C_0^j\right) \le C_2\left(C_0^j, k_z{}^1, k_z{}^2, k_z{}^3\right) \le C^+\left(C_0^j\right), \quad j = 1,..., J \\ k_z{}^i \in \Re, \quad i = 1,2,3 \end{gathered}$$

where C_0^j, and C_0^{new} are the pollutant concentrations released into the river, upstream, C^- and C^+ are obtained from the confidence intervals based on the mean value of the measured pollutant concentrations, $\overline{C}$ and their standard deviation, σ, using the formula $\overline{C} - 0.02\sigma$ and $\overline{C} + 0.02\sigma$ respectively. The average pollutant concentration at the three pumping wells is denoted by C_1 while the average pollutant concentration at the observation wells are denoted by C_2. The variable γ is used instead of the actual objective function, C_1, in order to convert the problem into a minimization problem with only a single variable to be optimized as opposed to optimizing C_1 over all possible values of $k_z{}^i$.

3.2.1 PODI approximation to the high dimensional model

The high dimensional model solved using *PMWinPro* solver [19] and *MT3D* [20] uses 6240 hexahedral cell elements within the 3D domain of the soil layers in valley. 15 snapshots were obtained by solving the high dimensional model using the initial pollutant concentration released upstream of 10,000 to 30,000 $\mu g / m^3$ in steps of 10,000 $\mu g / m^3$ and the vertical hydraulic conductivities ranging from 0.1 to 1.0 in steps of 0.2 for the first aquifer layer, 0.01 to 0.1 in steps of 0.02 for the silt layer and 0.2 to 2.0 in steps of 0.4 for the second aquifer layer. Figure 6 shows pollutant concentrations within the valley for two of the 15 snapshots where the pollutants are mostly confined within the river and riverbanks. The contour plots of the concentrations in the first layer for the

two different sets of parameters are shown in Figure 6(a) and (b). Similarly, the contour plots of the concentrations in the second layer are shown in Figure 6(c) and (d) while the contour plots of the concentrations in the third layer are shown in Figure 6(e) and (f). From the contour plots in Figure 6, the pollutants do not penetrate much into the third layer due to the silt layer, confining the spread of the pollutants mainly to the first aquifer layer.

The POD modes are constructed based on the 15 snapshots. The relative energies of the POD modes are plotted in Figure 7. The first six POD modes have energies which are above $0.01\,\mu$ and therefore are likely to contribute the most to the snapshot data. The seventh and larger POD modes have energies that are less than 10^{-13} which means that the eigenvalues of these modes are essentially zero. Using the POD modes in tandem with cubic spline interpolation, the relative errors for the PODI results, $E_{true} = \dfrac{\left\| C - \hat{C}_N \right\|_\infty}{\left\| C \right\|_\infty}$, for several different sets of parameters are plotted in Figure 8. The first set of the parameters marked by a solid line and '+' in Figure 8 is included in the ensemble of snapshots data while the second parameter set in Figure 8, which is marked by a dotted line and 'x', is not included in the ensemble but is within the bounds of the range of parameters of the ensemble data set. The third parameter set in Figure 8, which is marked by a solid line and '*x*', is not within the range of parameters of the ensemble data set. The relative errors plotted in Figure 8 reflect the choice of the parameter set as the relative error of the first parameter set decreases most rapidly with increasing number of POD modes while the relative error of the third parameter set is least responsive to the increasing number of POD modes. As the POD modes mainly contain information acquired from the ensemble snapshot, it is clear that the PODI approximations for parameter sets within the ensemble would yield the most accurate solutions.

Based on Figure 8, the relative errors show very minute changes after the sixth POD modes. This observation tallies with the relative energies plotted in Figure 7. Therefore, the first 6 POD modes are used in tandem with cubic spline interpolation for the uncertainty optimization problem.

3.2.2 Uncertainty Optimization Problem

The uncertainty problem defined by Equation (9) is a semi-infinite programming problem [21, 22] as the uncertainty in the actual values of the vertical hydraulic conductivities perforce the evaluation of the constraints stated in Equation (9) over all possible values (infinite) of $k_z^{\,i}$, i = 1,2,3. In this example, the actual values of the vertical hydraulic conductivities are 0.5, 0.05, and 1.0 for the respective layers but they are assumed to be unknown. Based on the known pollution concentrations at the two observation wells for three different pollutant concentrations released into the river, 5,000, 10,000, and 30,000 $\mu g/m^3$, the values of C^- and C^+ for these three different pollutant concentrations are evaluated. The semi-infinite programming problem in Equation (9) is then solved using *MATLAB* Optimization Toolbox [23] function, *fseminf*, for several different values of pollutant concentrations released into the river upstream to obtain the maximum average amount of pollutant concentration in the vicinity of the three pumping wells for each of the released pollutant concentration. The evaluations of the constraint functions, C_1 and C_2, in Equation (9), are obtained from the PODI model with 6 POD modes and cubic spline interpolation.

The results of the maximum average pollutant concentration in the vicinity of the three pumping wells, $\max C_1 = \gamma$, and the actual average concentration based on the actual vertical hydraulic conductivities are plotted against the initial pollutant concentra-

tion, C_0, in Figure 9. The relative error between the actual C_1 and the maximum C_1 can be expressed as $E = \frac{(\text{actual}\, C_1 - \max C_1)}{\text{actual}\, C_1}$ and is also plotted in Figure 9. While some of the results obtained from solving the semi-infinite programming overestimate the actual average concentrations, others underestimate the actual concentrations as seen in Figure 9. It would have been preferable that all the optimization results overestimate the actual average concentrations to ensure that the results, however, most of the negative relative errors are quite small, less than 5%, as shown in Figure 9. Part of the reason for the underestimation of the actual average concentrations is the error propagation from the PODI approximations. The PODI approximations of C_1 and C_2 for a given set C_0 and k_z^i, i = 1,2,3, could actually be less than the actual C_1 and C_2, thus causing the underestimation of the actual average concentrations. The relative errors shown in Figure 9 indicate that for those C_0 values which are within the range of the parameter set of the ensemble snapshots, the optimization results, which are dependent on the PODI approximations, are closer to the actual average concentrations. For C_0 values which are not within this range, such as 5,000 $\mu g / m^3$, the relative errors are larger due to the larger difference between the PODI approximations and the true values.

At least 8 function calls to the objective and constraint functions are made in solving the semi-infinite programming problem. Since there are 3 C_2 evaluations and 1 evaluation of C_1 for all possible values of k_z^i at each constraint function call and the range of possible values of k_z^i is discretised into more than 400 sections, more than 12,800 PODI evaluations of C_1 and C_2 are made in solving the semi-infinite programming problem. The computational cost of each PODI approximations is in the order of $\mathrm{O}(Nn + NW^2)$ operations where n is the number of high dimensional elements, N is the

number of POD modes, and W is the number of finite support for the cubic-spline bases used for the interpolation while the computational cost of the high dimensional model is about $O(n^3)$ for each solution of Equations (6) and (7). Since N and W are much smaller than n, the computational cost of 12,800 PODI evaluations is effectively much smaller than the computational cost of 12,800 solutions of Equations (6) and (7). Thus, in such uncertainty optimization, the usage of reduced-order model such as PODI allows for great savings in the computational costs.

5. Conclusion

In this paper, the PODI method has been shown to work effectively as a black-box reduced-order model in solving two problems with different mathematical models and optimization problems. In the HDD flow model, the PODI method is able to predict the airflow characteristics such as pressure and velocity governed by the Navier-Stokes Equations while in the groundwater contamination model, the PODI method predicted the pollutant concentration levels governed by Darcy's Law and the transport equation. In both problems, the PODI method uses interpolation methods coupled with POD modes to find an approximation to the true solutions without any prior knowledge of the mathematical models. Furthermore, the PODI method is shown to have much less computational cost and took less computational time than the high dimensional solvers used in the two examples, demonstrating that it is an efficient and economical tool to be used in the optimization problems posed.

The errors of the PODI approximations within the range of parameter set used for the ensemble of snapshots are relatively small for a high enough number of POD modes used. The errors of the PODI predictions outside this range of parameter set are higher as the POD modes are unable to capture the components of the actual high dimensional system as accurately, which might affect the optimization results as shown in the

groundwater contaminant transport example. The relative energies measured by the eigenvalues of the POD modes were used to select the number of modes used in the PODI method but the question remains as to the optimal number of modes that could reduce the actual error to an acceptable level. More attention should also be given to developing an error estimate for the PODI method to enhance the performance of the PODI technique in optimization problems.

Acknowledgement

This work has been partially supported by Seagate Technologies International and Singapore-MIT Alliance Graduate Program in Computational Engineering. The authors would like to acknowledge Mr. Elson Goh of Seagate Technologies International for his help with the CAD models of the HDD enclosure.

REFERENCES

1. A.K. Noor and J.M. Peters, "Reduced basis technique for nonlinear analysis of structures." *AIAA Journal*, vol. 18, no. 4, pp. 455–462, 1980.
2. J.P. Fink and W.C. Rheinboldt, "On the error behavior of the reduced basis technique for nonlinear finite element approximations." *Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 63, no. 1, pp. 21–28, 1983.
3. T.A. Porsching and M. Lin Lee, "The reduced basis method for initial value problems," *SIAM Journal on Numerical Analysis*, vol. 24, no. 6, pp. 1277–1287, 1987.
4. P. Feldmann and R.W. Freund, "Efficient linear circuit analysis by Pade approximation via the Lanczos process," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 14, no. 5, pp. 639-649, 1995.

5. R.W. Freund and P. Feldmann, "Reduced-order modeling of large passive linear circuits by means of the *SyPVL* algorithm," in *Proceedings of the 1996 IEEE/ACM International Conference on Computer-Aided Design*, pp. 280-287, 1996.

6. K. Gallivan, E. Grimme, and P. Van Dooren, "Asymptotic waveform evaluation via a Lanczos method," *Applied Mathematics Letters*, vol. 7, no. 5, pp. 77-80, 1994.

7. J.L. Lumley, "The structure of inhomogeneous turbulence," in *Atmospheric Turbulence and Wave Propagation*, Yaglom and V.I. Tatarski (eds). Nauka, Moscow, pp. 166-178, 1967.

8. L. Sirovich and M. Kirby, "Low dimensional procedure for the characterization of human faces," *Journal of the Optical Society of America A*, vol. 4, no. 3, pp. 519-524, 1987.

9. L. Sirovich, "Turbulence and the dynamics of coherent structures, parts I-III" *Quarterly of Applied Mathematics*, vol. 65, pp. 561-590, 1987.

10. FLUENT Inc., "Flow simulation software user guide," Fluent 6.2.16, 2004.

11. ABAQUS Inc., "Abaqus documentation, Version 6.5.4," 2005.

12. M.G. MacDonald and A.W. Harbaugh, "MODFLOW, A modular three-dimensional finite difference ground water flow model," *U. S. Geological Survey*, Open file report, pp. 83-875, Chapter A1, 1988.

13. H.V. Ly and H.T. Tran, "Modeling and Control of Physical Processes Using Proper Orthogonal Decomposition," *Mathematical and Computer Modelling*, 33 (2001) 223-236.

14. T. Bui-Thanh, "Proper Orthogonal Decomposition Extensions and Their Applications in Steady Aerodynamics," Masters Thesis, Singapore-MIT Alliance, NTU, 2003.

15. C. Bajaj, "Multi-dimensional Hermite Interpolation and Approximation for Modeling and Visualization," *Proceedings of The IFIP TC5/WG5.2/WG5.10 CSI International Conference on Computer Graphics, ICCG93, IFIP Transactions B-9. Graphics, Design and Visualization,* (Edited by S. Mudur, S. Pattanaik), 335-348, North Holland, 1993.

16. R.H. Bartels, J.C. Beatty, and B.A. Barsky, "Hermite and Cubic Spline Interpolation," In *An introduction to Splines for Use in Computer Graphics and Geometric Modelling* (Edited by Morgan Kaufmann), 9-17, San Francisco, CA, 1998.

17. S. Ali, M. Damodaran, and Q.Y. Ng, "Computational Models for Predicting Airflow Induced Particle Contamination in Hard Disk Drive Enclosures." In *Proceedings of 17th AIAA Computational Fluid Dynamics Conference,* **AIAA-2005-5342**, Toronto, Jun 2005.

18. K. Levenberg, "A Method for the Solution of Certain Problems in Least Squares," *Quart. Appl. Math*., 2, 164-168, 1944.

19. W.H. Chiang, "3D-Groundwater Modeling with PMWIN: A Simulation System for Modeling Groundwater Flow and Transport Processes," 2nd Edition, *Springer-Verlag*, Berlin Heidelberg, 2005.

20. C. Zheng, "MT3D: A Modular Three-Dimensional Transport Model for Simulation of Advection, Dispersion and Chemical Reaction of Contaminants in Groundwater Systems," Report for *Robert S. Kerr Environmental Research Laboratory, Environmental Protection Agency*, October 1990.

21. L.D. Coope and G.A. Watson, “Algorithm for semi-infinite programming,” *Math. Programming*, 32, 337-356, 1985.

22. R. Reemtsen and S. Gorner, “Numerical Methods for Semi-Infinite Programming: A Survey,” *NonCovex Optimization and Its Applications*, Kluwer Academic Publishers, 195-275, 1998.

23. T. Coleman, M. Branch, and A. Grace, *Optimization Toolbox for Use with MATLAB*, The MathWorks Inc., 1999.

LIST OF FIGURES

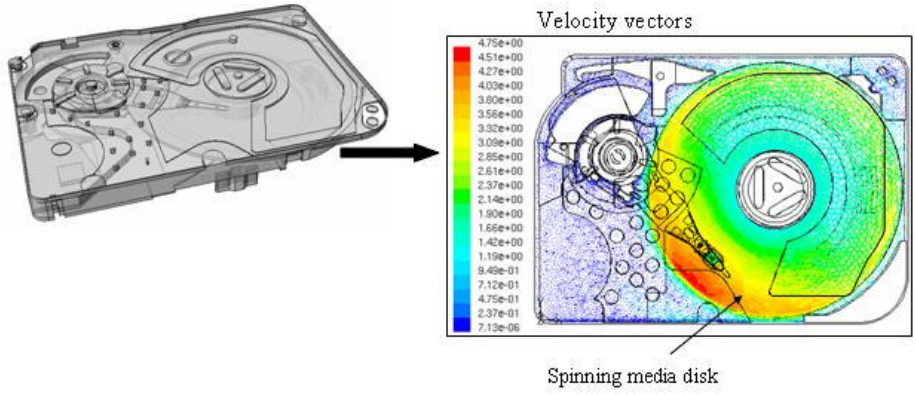
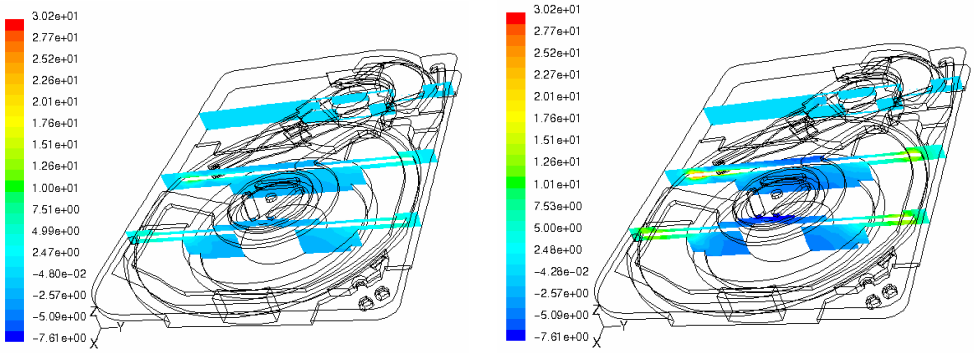
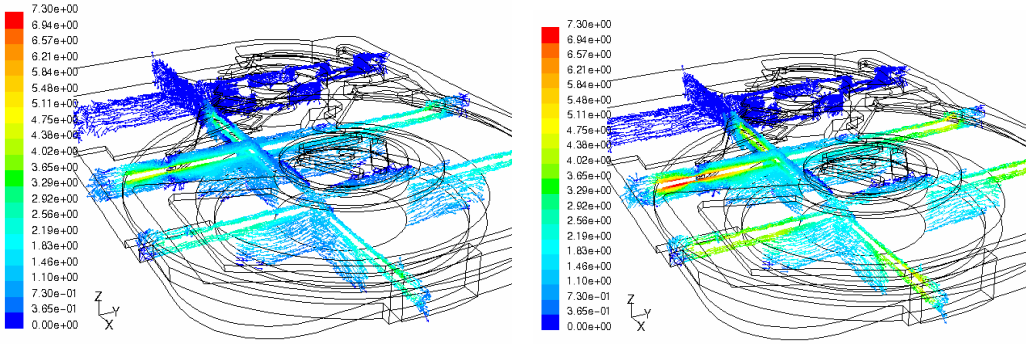


Figure 1: Airflow in a one-inch HDD enclosure



(a) 4000 rpm (b) 6500 rpm

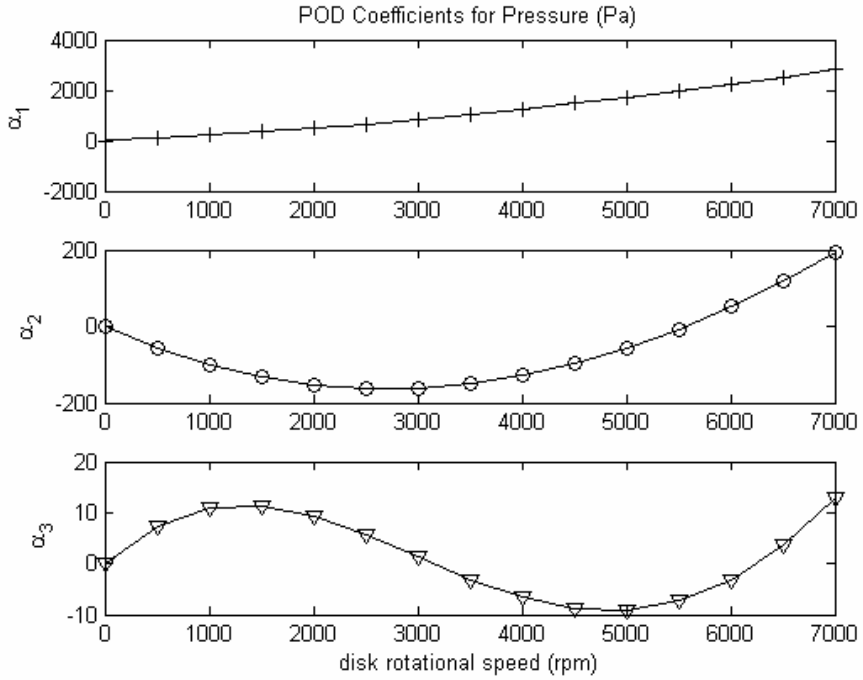
(i) Pressure



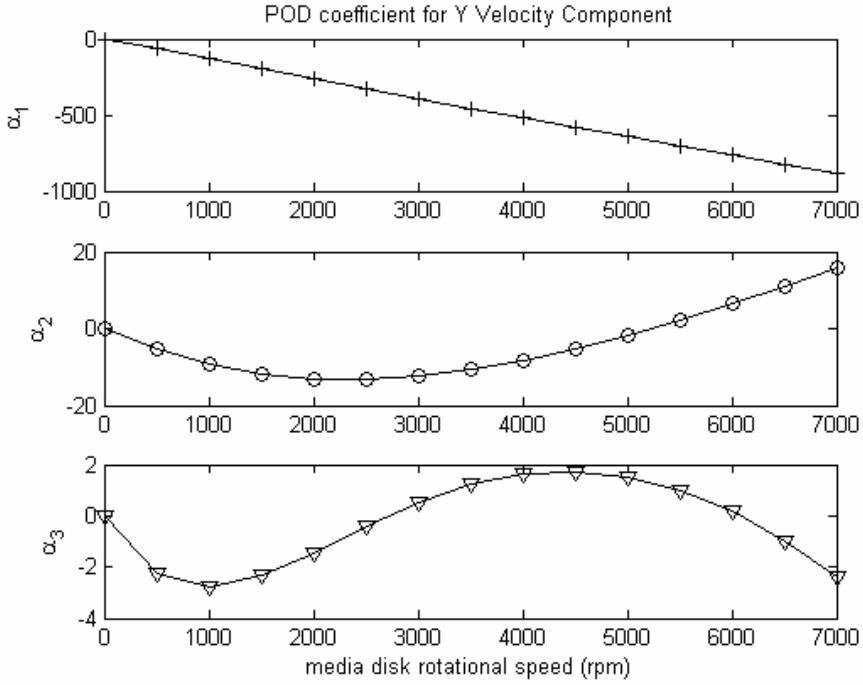
(a) 4000 rpm (b) 6500 rpm

(ii) Velocity

Figure 2: Two snapshots of the pressure and flow velocity in the HDD enclosure

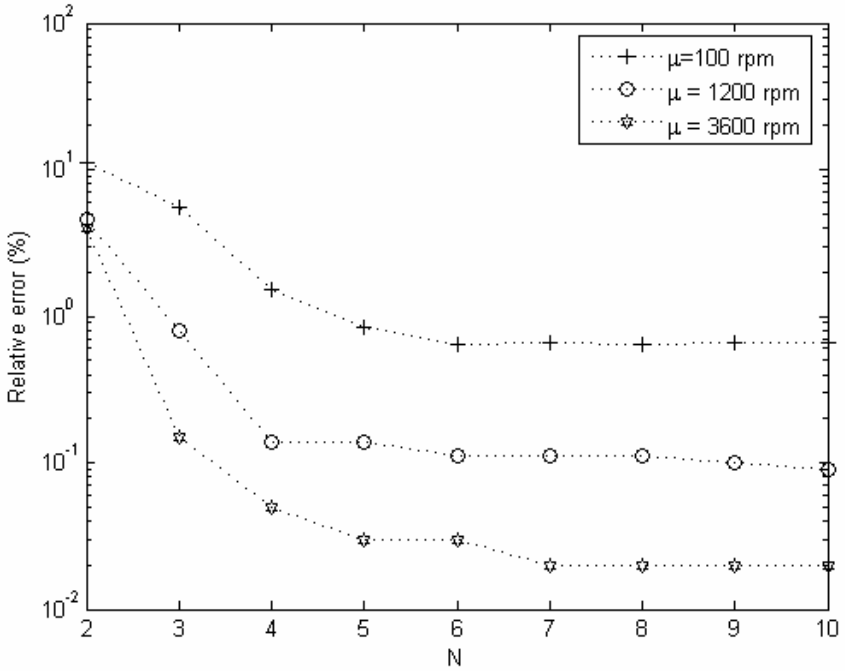


(a) Coefficients of Pressure

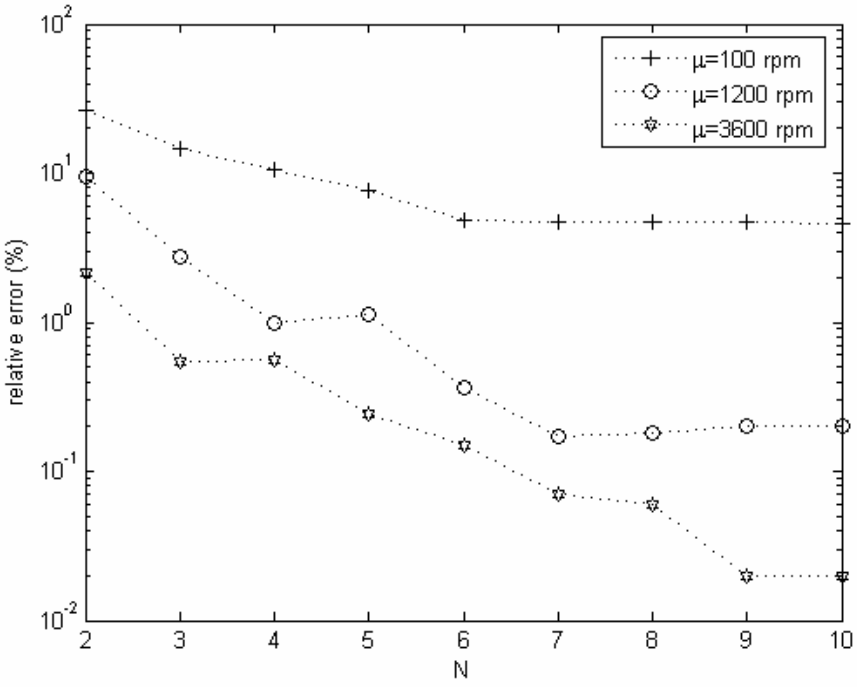


(b) Coefficients of U_2

Figure 3: Plots of the first three POD coefficients, α_i, with respect to μ in the first example of flow in HDD enclosures



(a) Relative error for Pressure



(b) Relative error for X-Velocity

Figure 4: Plot of relative error of pressure and x-velocity component in percentage against N, the number of POD modes

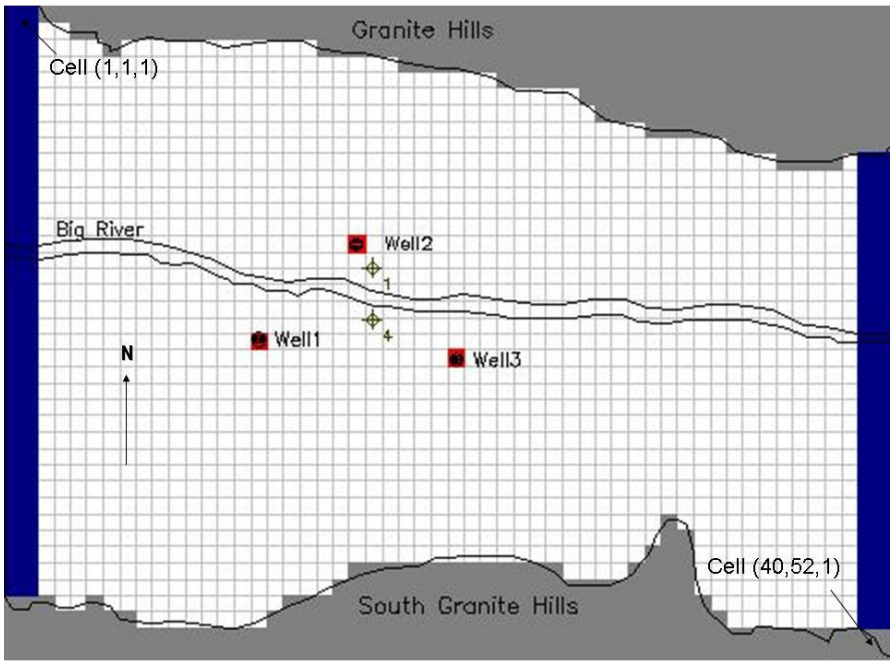


Figure 5: Map of the valley through which a river is flowing. The 3 pumping wells used to supply water to the inhabitants of the valley are marked in red. The observation wells, marked by the crosses, are located near to the riverbanks.

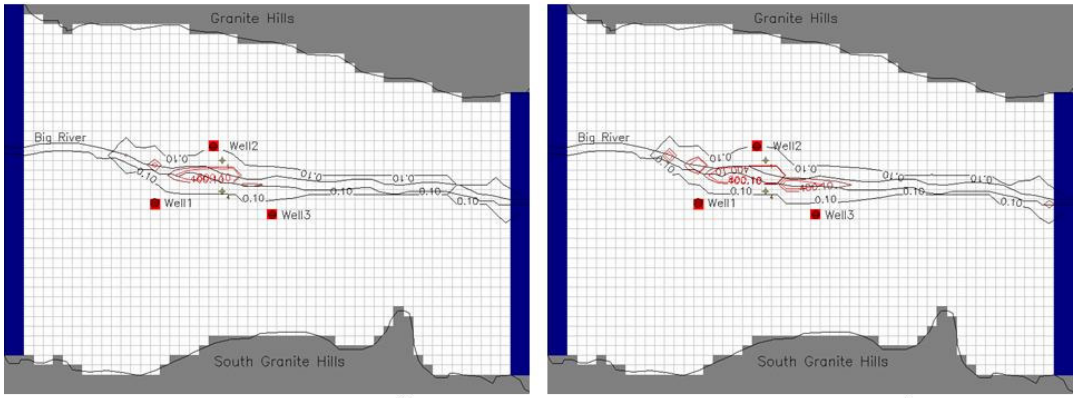


Figure 6: Contour plots of two of the 15 snapshots used in the construction of the POD modes.

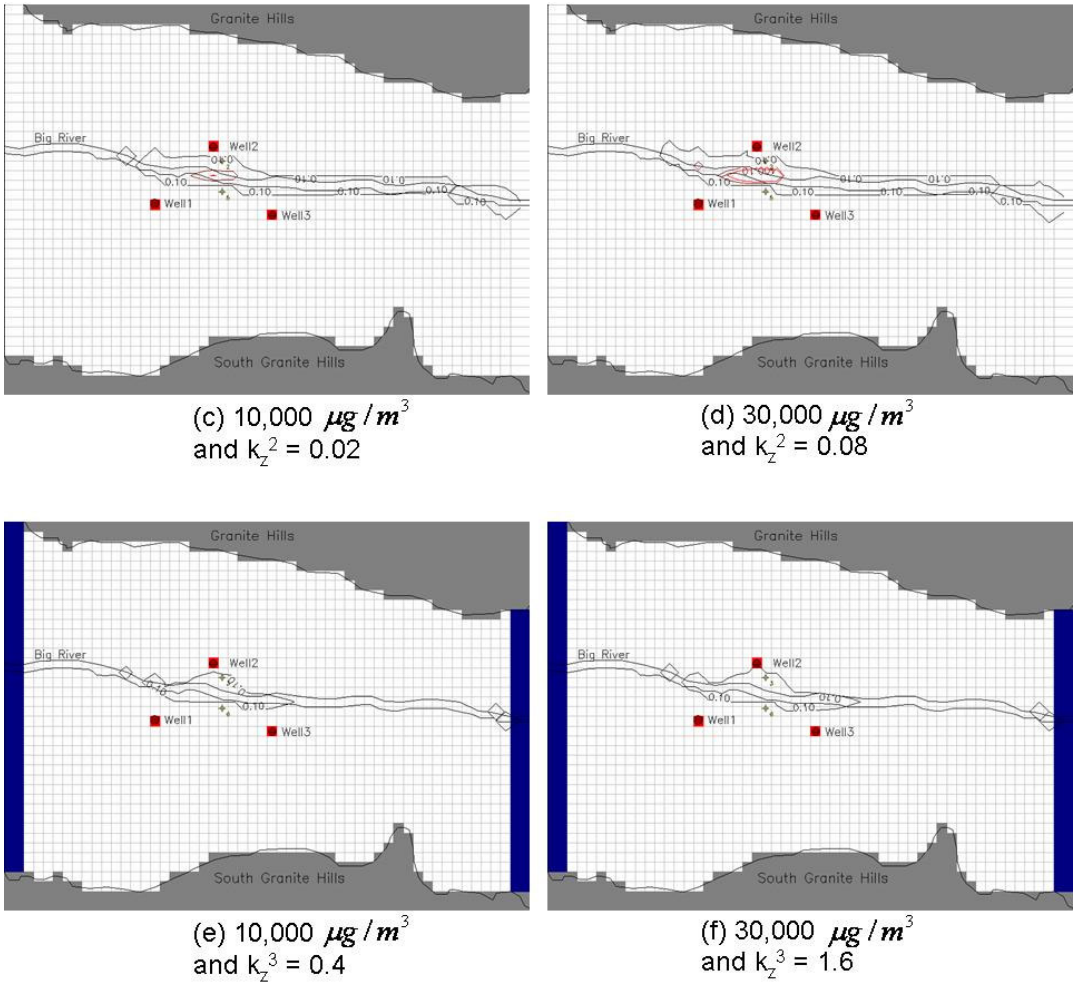


Figure 6 (cont'd): Contour plots of two of the 15 snapshots used in the construction of the POD modes for the three layers.

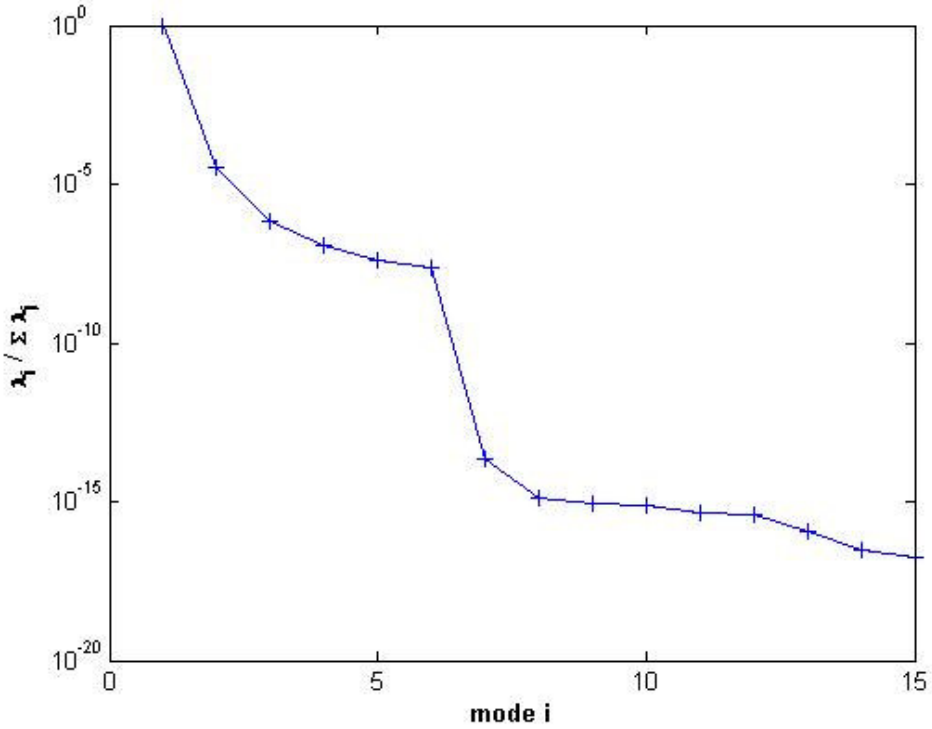


Figure 7: The plot of relative energies of the POD modes for the groundwater contaminant transport problem.

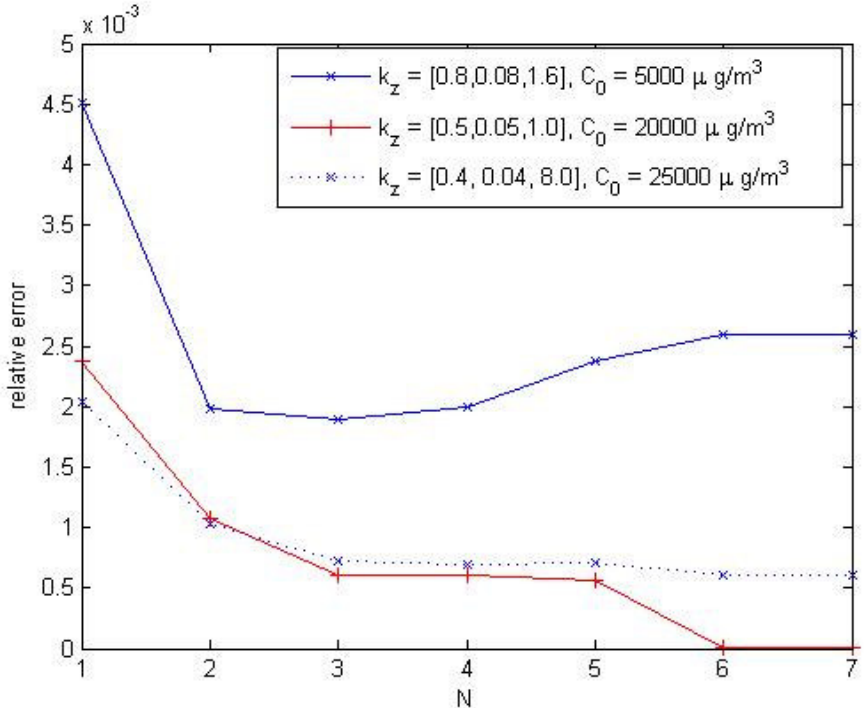


Figure 8: Plot of relative errors of the PODI results for different parameters against the number of POD modes, *N*.

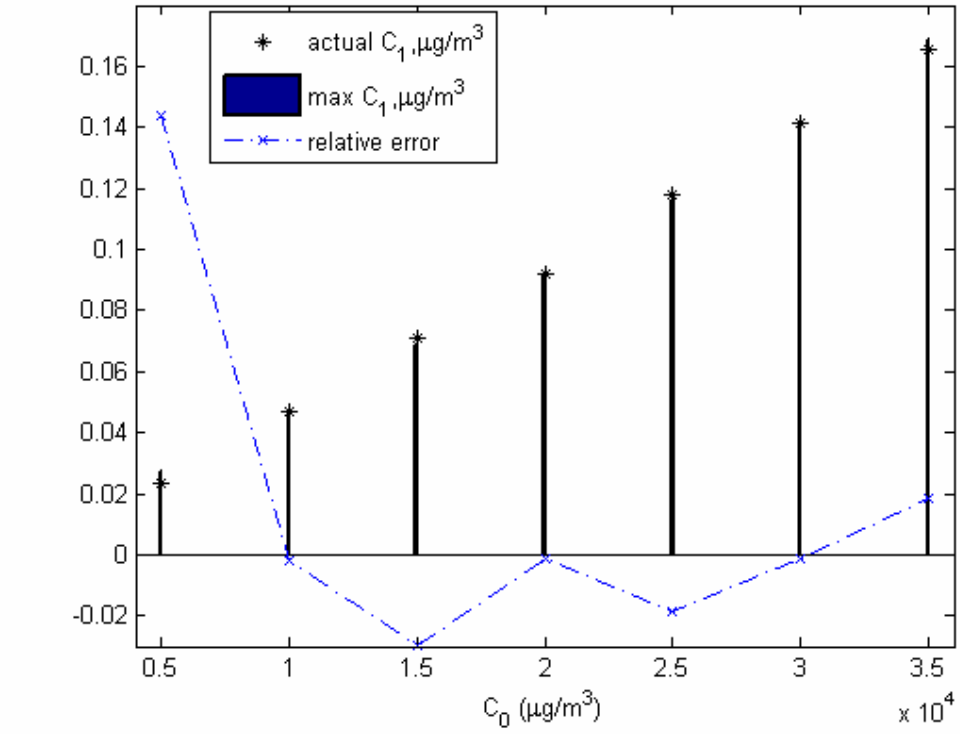


Figure 9: Plot of the actual average pollutant concentration in the vicinity of the three pumping wells, the maximum average pollutant concentration in the vicinity based on the uncertainty optimization problem, and the relative error between the actual average and maximum average against the initial pollutant concentration released into the river.

LIST OF TABLES

Table 1: Percentage error for the PODI solutions of the flow in the HDD enclosure

(based on 3 POD modes)

μ (rpm)	U	$E_{true} = \frac{\left\|U - \hat{U}_N\right\|_\infty}{\left\|U\right\|_\infty}$ (%)		
		Linear	**Spline**	**Hermite**
100	Pressure	7.72	5.71	5.5
	X-Velocity	12.39	15.19	14.62
	Y-Velocity	14.25	19.76	19.55
	Z-Velocity	22.15	18.63	18.69
1200	Pressure	1.01	0.77	0.8
	X-Velocity	2.6	2.68	2.71
	Y-Velocity	2.58	2.87	2.86
	Z-Velocity	1.32	1.11	1.10
3600	Pressure	0.16	0.15	0.15
	X-Velocity	0.6	0.55	0.55
	Y-Velocity	1.05	1.02	1.02
	Z-Velocity	0.31	0.33	0.33
5400	Pressure	10.78	10.74	10.74
	X-Velocity	8.88	8.87	8.87
	Y-Velocity	13.51	13.51	13.51
	Z-Velocity	9.63	9.64	9.63
6750	Pressure	0.09	0.036	0.037
	X-Velocity	0.37	0.37	0.37
	Y-Velocity	0.6	0.59	0.59
	Z-Velocity	0.13	0.14	0.14

Table 2: Optimal Solution of the Inverse Design Problem (starting point = 5000 rpm)

Desired rotational speed (rpm)	**Calculated rotational speed (rpm)**	**Percentage Error (%)**	**No. of function calls (PODI)**	**Average Time (min) (over 4 runs)**
250	256.6	2.6	24	3.94
1750	1751.1	0.06	16	2.68
4250	4249.8	0.005	10	1.74
6250	6249.7	0.005	14	2.29

Table 3: Properties of the aquifer and silt layers

Horizontal hydraulic conductivity (m/day)	**Layer 1**	5.0
	Layer 2	0.5
	Layer 3	2.0
Effective Porosity	**Layer 1**	0.20
	Layer 2	0.25
	Layer 3	0.25
Elevation of aquifer tops (m)	**Layer 1**	20.0
	Layer 2	8.5
	Layer 3	8.0
Elevation of aquifer bottoms (m)	**Layer 1**	8.5
	Layer 2	8.0
	Layer 3	0.0
Dimensions of the valley and surrounding hills (m)	**Length**	6750
	Width	5000

Table 4: Big river properties

Upstream river stage (m)	**Downstream river stage (m)**	**River width (m)**	**Riverbed hydraulic conductivity (m/day)**	**Riverbed thickness (m)**	**Upstream riverbed bottom elevation (m)**	**Downstream riverbed bottom elevation (m)**
19.4	17.0	100	2.0	1.0	17.4	15.0

: