Support Vector Machines (SVMs) for Monitoring Network Design

by Tirusew Asefa[1,2], Mariush Kemblowski[1], Gilberto Urroz[1], and Mac McKee[1]

Abstract

In this paper we present a hydrologic application of a new statistical learning methodology called support vector machines (SVMs). SVMs are based on minimization of a bound on the generalized error (risk) model, rather than just the mean square error over a training set. Due to Mercer's conditions on the kernels, the corresponding optimization problems are convex and hence have no local minima. In this paper, SVMs are illustratively used to reproduce the behavior of Monte Carlo–based flow and transport models that are in turn used in the design of a ground water contamination detection monitoring system. The traditional approach, which is based on solving transient transport equations for each new configuration of a conductivity field, is too time consuming in practical applications. Thus, there is a need to capture the behavior of the transport phenomenon in random media in a relatively simple manner. The objective of the exercise is to maximize the probability of detecting contaminants that exceed some regulatory standard before they reach a compliance boundary, while minimizing cost (i.e., number of monitoring wells). Application of the method at a generic site showed a rather promising performance, which leads us to believe that SVMs could be successfully employed in other areas of hydrology. The SVM was trained using 510 monitoring configuration samples generated from 200 Monte Carlo flow and transport realizations. The best configurations of well networks selected by the SVM were identical with the ones obtained from the physical model, but the reliabilities provided by the respective networks differ slightly.

Background

The design of ground water pollution monitoring networks entails the selection of sampling points (spatial) and sampling frequency (temporal) to determine physical, chemical, and biological characteristics of ground water (Loaiciga et al. 1992). The current approach to ground water quality monitoring network design requires that consideration be given to (1) the spatial and temporal coverage of the monitoring sites; (2) the competing objectives of a monitoring program; (3) the complex nature of geologic, hydrologic, and other environmental factors; (4) the stochastic character of transport parameters

(geologic, hydrologic, and environmental) used in the design process; and (5) the risk posed to society (failure to detect, poor characterization, etc.). Based on design objectives one can identify three categories of monitoring networks: (1) leak detections; (2) characterization; and (3) long-term monitoring. In the following sections we focus our attention on leak-detection monitoring networks. Interested readers may refer to the works of Hudak and Loaiciga (1992), Datta and Dhiman (1996), Molina et al. (1996), Mahar and Datta (1997), Montas et al. (2000), and Reed et al. (2000) for design of characterization and long-term monitoring networks.

Leak–Detection Monitoring Networks

Networks in this category enable one to detect unexpected leaks before the contaminants reach a compliance boundary, which is usually located at some relatively short distance, say 100 m, from a landfill. Solutions to this problem are complicated due to the facts that the location and magnitude of leaks are random and contaminants travel within systems characterized by poorly defined boundaries and hydrologic conditions. Massmann

[1]Department of Civil and Environmental Engineering and Utah Water Research Laboratory, Utah State University, 8200 Old Main Hill, Logan, UT 84321.

[2]Corresponding author: Tampa Bay Water, 2535 Landmark Drive, Suite 211, Clearwater, FL 33761: fax (435) 797-3663; tasefa@tampabaywater.org

and Freeze (1987a, 1987b) presented a comprehensive framework and application for the design of landfill operation. The objective was to maximize the net present value of stream of costs, benefits, and risks. Risk is associated with the cost of failure, with failure being defined as the situation in which the concentration measured is in excess of a standard at the compliance surface. The decision analysis will then compare different alternative monitoring networks by calculating the reduction of the probability of failure. Meyer and Brill (1988) analyzed the same problem through an integer-programming model of facility location (Church and ReVelle 1974). Their objective was to maximize the detection of plumes that exceed a concentration standard. Random plumes were generated from parameters sampled from longitudinal and transverse dispersivity distributions, which were assumed to be positively correlated with a normal distribution and uniform velocity vector directions. For a given number of monitoring wells (representing cost constraint), the integer programming would select locations that resulted in a higher probability of plume detection. Meyer et al. (1994) extended the work of Meyer and Brill (1988) to include an additional objective of minimizing the area of contamination at the time of detection. Storck et al. (1997) extended the work of Meyer et al. (1994) to three-dimensional flow and transport problems and ignored dispersion transport. Instead, they used an advective particle-tracking approach.

Angulo and Tang (1999) used similar objective criteria and uncertainty analysis as in Meyer et al. (1994) but formed the objective function through "utility criteria" for a system that consists of a single row of wells perpendicular to the flow direction. For a given number of wells at a given distance from the landfill, the total cost of the system was found to be the sum of system construction and monitoring costs, plus the cost of aquifer remediation associated with expected volumes of contaminated aquifer given detection and no detection. Morisawa and Inoue (1991) presented a methodology for designing an optimal monitoring network based on fuzzy utility functions. A stochastic simulation of both critical and precursor indicator chemicals resulted in a mathematical description of a four-attribute design problem. An optimum monitoring network was then defined as a network having maximum total utility, which was evaluated as the fuzzy expectation of weighted arithmetic sums of the four utilities. All the aforementioned studies assumed that there is continuous monitoring at potential monitoring well locations. Monitoring frequency was examined by Jardine et al. (1996) within a decision analysis framework for designing monitoring networks for fractured rock.

Methodology

In this paper we present a methodology for the application of support vector machines (SVMs) in the design of ground water contamination detection monitoring networks. The objective of the modeling exercise is to design an optimal monitoring network for an initial detection of ground water contamination that maximizes the probability of detection while minimizing cost (i.e., number of

monitoring wells). A given topology of monitoring network (number and locations) has a corresponding probability of plume detection (reliability). If the number and/or location of a well changes, the monitoring network will have a different reliability. With an increased number of wells, one can increase the reliability of a monitoring network, but this in turn would result in higher cost. While the cost of monitoring wells in a network can be easily quantified in monetary values, it is not simple to quantify the change in reliability (probability of detection) in monetary terms. Therefore, we set our objective herein getting trade-off curves between monitoring network sizes and corresponding reliabilities. SVMs will be trained to provide network reliabilities for an arbitrary configuration of monitoring wells so that one would be able to select those providing the highest reliabilities without running time-consuming flow and transport models.

Model Domain and Uncertainty

The aquifer is assumed to be represented by a rectangular box in which a two-dimensional flow takes place. The overall dimensions of the model were 1000 by 500 m (Figure 1). Model cell sizes were 20 by 20 m. The boundary conditions for the steady-state flow model were zero flux at $y = 0$ m and $y = 500$ m and constant head along the other two boundaries. These boundary conditions imply that the regional ground water flow direction is relatively well known. The head values at $x = 0$ m and $x = 1000$ m were chosen to result in an average gradient over the domain of 0.01. We considered only two types of uncertainties: source location and the spatial distribution of the hydraulic conductivity. The method can readily be extended to incorporate additional uncertain parameters (e.g., boundary condition, recharge). The leak is assumed to be equally probable at any location in the landfill cell (equivalent to a numerical cell). The source of hydrogeological uncertainty is limited to spatial variability in the hydraulic conductivity field. The natural logarithm of the hydraulic conductivity, $Y = \ln(K)$, was modeled as a Gaussian second-order stationary stochastic process with a mean value of $\mu_Y = 0.79$. The variance of Y was set at $\sigma^2_Y = 0.96$, with correlation scale of $l = 100$ m and exponential covariance function. Meyer et al. (1994) also used a similar case study. The method presented by Tompson et al. (1989), based on the turning bands algorithm, was used to generate a stationary, correlated two-dimensional hydraulic conductivity field. Note that this simple representation of the aquifer is not a limitation of the application of SVMs, but rather a simplification of numerical modeling to save computational time. The procedure developed here can easily be extended to more complicated site conditions.

Flow and Transport Simulations

The USGS finite-difference MODFLOW code (McDonald and Harbaugh 1988; Harbaugh and McDonald 1996) was used to simulate the steady-state ground water flow problem. MODFLOW was chosen because it has been widely used, and its use has been extensively documented in the technical literature. Transport was simulated using a particle-tracking approach. MODPATH (Pollock 1988, 1989), which is designed to use the head calculated
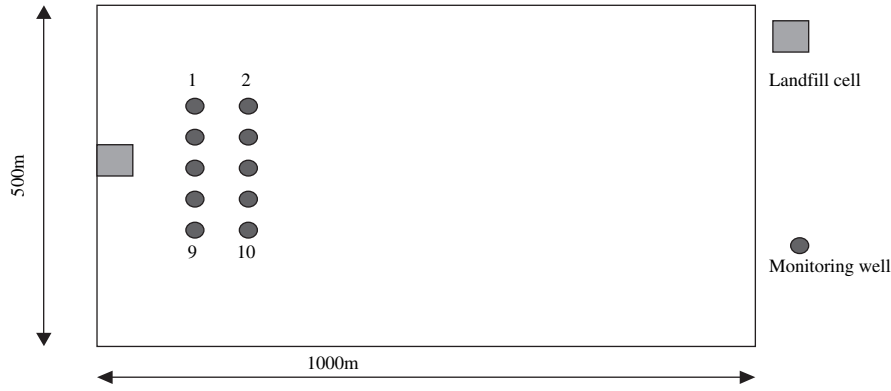
Figure 1. Generic site conceptualization.

from MODFLOW, was used in this study. MODPATH tracking code ignores dispersion in transport. Some previous work on monitoring network design also ignored dispersion (e.g., Massmann and Freeze 1987b; James and Gorelick 1994; Jardine et al. 1996; Storck et al. 1997). Massmann and Freeze (1987b) assumed that the advective component dominated contaminant migration and the effect of dispersion on detection probability would be similar to that of increasing the contaminant source area. Storck et al. (1997) assumed that, because dispersion acts to increase the volume that is contaminated by a plume, ignoring the effect of dispersion results in smaller plumes that are more difficult to detect; this is a conservative assumption that agrees with a pessimistic (worst case) design philosophy. They reported that a network designed by the dispersion-less method detected even more plumes if its performance was evaluated with respect to a set of plumes generated with dispersion. Hence, we will also neglect dispersion transport in the present approach to simplify computation. Note that the method presented here can easily be extended to include dispersion and other forms of transport. MODFLOW and MODPATH were run in a consecutive manner with a command interface made of C/C++ processing programs and UNIX shell commands before the results were passed to the SVMs.

Background on SVM

The support vector methodology (Vapnik 1995, 1998) estimates a functional dependency, $f(\mathbf{x})$, between monitoring well locations $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L\}$ taken from $\mathbf{x} \in \mathbf{R}^K$ and their corresponding reliabilities $\{y_1, y_2, \ldots, y_L\}$ with $y \in \mathbf{R}$ drawn from a set of L independent and identically distributed monitoring networks/reliabilities observations by minimizing the following regularized functional:

$$\text{Minimize } \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{L}\left(\xi_i + \xi_i^*\right) \qquad (1a)$$

$$\text{subject to } \begin{cases} y_i - \sum_{j=1}^{K}\sum_{i=1}^{L}w_j x_{ji} - b \leq \varepsilon + \xi_i \\ \sum_{j=1}^{K}\sum_{i=1}^{L}w_j x_{ji} + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \qquad (1b)$$

to obtain $f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x}\rangle + b$ with $\mathbf{w} \in X, b \in R$

$$\text{or } f(x) = \sum_{j=1}^{K}w_j x_j + b \qquad (1c)$$

$\langle \mathbf{w}, \mathbf{x}\rangle$ denotes the inner product or dot product between $\mathbf{w}$ and $\mathbf{x}$, b is bias, and K is an input dimension (in this case the maximum number of potential monitoring well locations). The estimated function, $f(\mathbf{x})$, would then calculate the probability of contaminant detection for an arbitrary monitoring configuration $\mathbf{x}$. The first term in the objective function is a stabilizer that is a prior on the regression function. It minimizes the complexity of the function f (i.e., the estimated function will always tend to be flat, avoiding overfitting). The second term (together with the constraints) represents the ε-insensitive loss function depicted in Figure 2. As shown in the figure, the parameters ξ_i, ξ_i^* are slack variables that determine the degree to which samples with error more than ε be penalized. In other words, any error smaller than ε does not require ξ_i, ξ_i^* and hence does not enter the objective function because these data points have a value of zero for the loss function. The constant $C > 0$ determines the trade-off between the complexity of function f and the amount up to which deviations larger than ε are tolerated. Usually, Equation 1 is solved in its dual form using Lagrange multipliers. Writing Equation 1 in its dual form and differentiating with respect to primal variables ($\mathbf{w}$, b, ξ_i, ξ_i^*) and rearranging gives:

Maximize $W(\alpha^*, \alpha)$

$$= -\varepsilon\sum_{i=1}^{L}\left(\alpha_i + \alpha_i^*\right) + \sum_{i=1}^{L}y_i(\alpha_i - \alpha_i^*)$$

$$-\frac{1}{2}\sum_{i,j=1}^{L}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(\mathbf{x}_i, \mathbf{x}_j) \qquad (2a)$$

subject to the constraints $\sum_{i=1}^{L}(\alpha_i^* - \alpha_i) = 0, 0 \leq \alpha_i, \alpha_i^* \leq C$

$$\qquad (2b)$$

$$\text{to obtain } f(\mathbf{x}) = \sum_{i=1}^{N}(\alpha_i^* - \alpha_i)k(\mathbf{x}, \mathbf{x}_i) + b \qquad (2c)$$
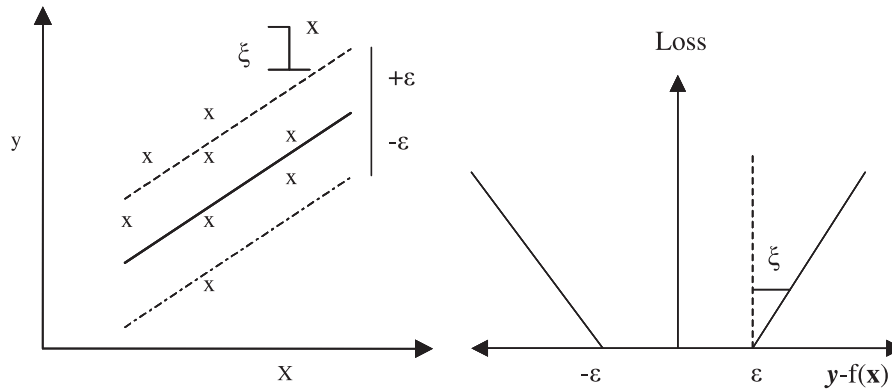
Figure 2. ε-insensitive loss function Γ.

where $k(\mathbf{x},\mathbf{x}_i)$ is a kernel function that replaces the dot products of input examples, and α_i^* and α_i are Lagrange multipliers. From the Kuhn-Tucker (KT) condition it follows that the Lagrange multipliers may be nonzero only for $|f(\mathbf{x}_i) - y_i| \geq \varepsilon$. In other words, for all samples inside the ε-tube, the α_i, α_i^* vanish. The samples that have no vanishing coefficient are called support vectors, hence the name support vector machines. Intuitively, one can imagine the support vectors as sample points that support the "decision surface" or hyperplane.

In its present form, SVMs for classification problems were developed at AT&T Bell Laboratory by Vapnik and coworkers in the early 1990s (e.g., Boser et al. 1992). SVMs for regression were first introduced in Vapnik (1995), and earlier applications were reported in the late 1990s (e.g., Vapnik et al. 1997). Despite enjoying success in other fields (Schölkopf et al. 1999b), there are few applications of SVM in hydrology. Dibike et al. (2001) applied SVM successfully in both remotely sensed image classification and regression (rainfall/runoff modeling) problems and reported a superior performance over the traditional artificial neural networks (ANNs). Kaneviski et al. (2002) used SVM for mapping soil pollution by Chernobyl radionuclide Sr90 and concluded that the SVM was able to extract spatially structured information from the row data. Liong and Sivapragasam (2002) also reported a superior SVM performance compared to artificial neural net in forecasting flood stage.

SVM Hyperparameters

Two of the SVM hyperparameters are ε, which defines the ε-insensitive loss function, and the capacity C. The parameter C sets an upper bound on the support vector coefficients (αs). It controls the trade-off between minimizing the loss function (satisfying the constraints) and minimizing the regularizer (complexity). The lower the value of C, the more the weight given to the regularizer. As C approaches infinity, the problem tends to be unconstrained and also unstable. Cristianini and Shawe-Taylor (2000) suggest the use of a range of values of C and some validation techniques for selecting the optimal value of this parameter. A rule of thumb suggested by Saunders et al. (1998) for selecting parameter C is to use a value that is slightly lower than the largest coefficient, or α value,

obtained from training with C equal to infinity. The idea behind this approach is justified since choosing a value higher than the largest coefficient will obviously have no effect because the box constraint (Equation 2b) will never be violated.

Using the ε-insensitive loss function enables one to optimize the generalization bound. This relies on defining a loss function that ignores errors located within a certain distance of the true value. Selecting an appropriate ε is largely a heuristic exercise. Schölkopf et al. (1999a) present a modified SVM algorithm that automatically calculates ε given the fraction of data points outside the ε-tube, a parameter referred to as υ. Then again, υ is determined heuristically. Mattera and Haykin (1999) propose to choose the value of ε so that the percentage of support vectors is about 50% of the number of samples. Smola et al. (1998) and Kowk (2001) proposed asymptotically optimal ε-values proportional to input noise level.

Selection of Suitable Kernels

The problem of choosing a suitable architecture for a neural network application is similar to the problem of choosing suitable kernels for SVM. Conceptually, the choice of a kernel, among other things, corresponds to choosing a similarity measure of the data because kernels are defined as an inner product of a mapping function in feature space, and inner products measure similarity of data. Being able to compute inner products means being able to make geometric constructions in terms of angles and lengths of input vectors. Since one is interested in predicting y from known $\mathbf{x}$, one would like to have some form of measure that relates $(\mathbf{x},y)$ to the training examples. Informally, this means that similar inputs lead to similar outputs. In the output space, similarity is measured by a loss function. In the case of pattern recognition, the only possible outputs are "similar" and "different."

One can construct kernels from basic simple functions or as a combination of other simple kernels (see Vapnik 1995, 1998; Schölkopf and Smola 2002 for different methods of kernel construction). In this research we have tested several commonly used kernels for the SVMs, which are presented in the Appendix. However, we only report results obtained using Strongly Regularized Fourier and spline kernels.

An SVM Algorithm

Figure 3 contains a graphical overview of the different steps in regression SVM. The input patterns (in this case, monitoring well locations for which a prediction of reliabilities should be made) are mapped into feature space by a mapping function Φ. One does not have to explicitly know the mapping function, but it is sufficient that the dot products that correspond to evaluating kernel k functions at monitoring locations $k(\mathbf{x}_i, \mathbf{x})$ are known. Finally, the dot products are added up using the weights $\beta_i = (\alpha_i - \alpha_i^*)$, which are nonzero Lagrange multipliers. This result is then added to the bias b to yield the final network reliability predictions.

Implementation of SVMs

The implementation of SVMs for the design of initial detection monitoring networks may be generalized by the following eight steps: (1) generating a random leak location; (2) generating a random (although correlated) hydraulic conductivity field; (3) solving a steady-state ground water flow model to obtain ground water head for the given hydraulic conductivity field; (4) transport simulation of the resultant contaminant plume until it reaches model boundaries; (5) keeping track of simulated plumes exceeding regulatory standards at monitoring wells; (6) developing a training and testing set (input being network configurations, while output is the reliability provided by a given monitoring network); (7) using the trained SVM to predict the reliability provided by a set of monitoring wells; and (8) using the flow and transport code to verify if the SVM-recommended network configuration provides the required probability of detection.

Illustrative Example

Figure 1 depicts the generic problem considered here. Ten potential monitoring well locations are considered (the first column of monitoring wells being located 100 m from the landfill cell). For each random leak location and spatially variable hydraulic conductivity field, the steady-state ground water flow problem is first solved, followed by the solution of the advective transport problem using a particle-tracking approach.

Five thousand particles uniformly generated over the landfill cell were used to represent the random leak. Each particle was advected until all particles left the model boundary (compliance boundary). The contaminated plume was defined by the model cells through which at least one particle passed. Detection at the monitoring locations was assumed to occur if the concentration at the wells reached 10% of the concentration introduced at the source cell. Given a unit mass of contaminant for each particle in the source cell, the concentration at the monitoring location will be equal to the ratio of the number of particles that pass through that location to the number of particles introduced at the source cell, divided by the flow passing the monitoring well. Storck et al. (1997) also used a similar detection criterion. Continuous sampling at the monitoring wells was also assumed. Two hundred Monte Carlo ground water flow and transport runs were conducted. For each Monte Carlo run, detection was noted at each potential monitoring well location in order to calculate reliability as follows:

$$\text{Reliability} = \frac{\text{Number of Detected Plumes}}{\text{Total Number of Generated Plumes}} \quad (2c)$$

The training and testing sets were randomly drawn from possible combinations of the 10 potential monitoring wells.

$$\text{Sample Points} = \sum_{i=1}^{10} \text{Combination}(10, i) \quad (3)$$

Therefore, the input vector has 10 dimensions corresponding to the 10 potential monitoring locations. The output will be the reliability (probability of detection) of the given monitoring configuration. Table 1 shows 10 samples of training patterns. In the table, the value 1 or 0 defines the existence/nonexistence of a potential monitoring well at a given monitoring network. For example, sample number 1 consists of well numbers 5, 6, 8, and 9 (Figure 1). These wells detected 148 of the randomly generated plumes, and hence their reliability is 0.74. The reliabilities of 1023 (Equation 3) monitoring networks were similarly calculated constituting the total training and testing patterns.
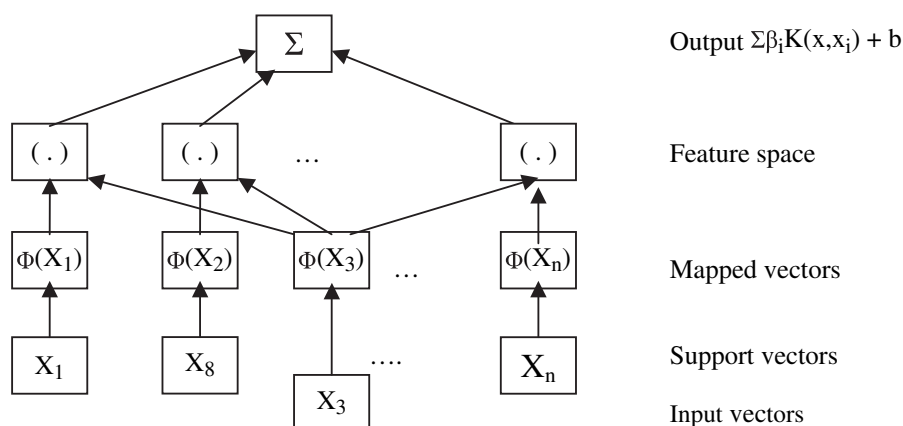

Figure 3. Architecture of an SV regression machine.

Table 1
Training Patterns

No.	Input Pattern										Output Pattern
1	0	0	0	0	1	1	0	1	1	0	0.74
2	0	0	0	0	1	1	0	1	1	1	0.74
3	0	0	0	0	1	1	1	0	0	0	0.67
4	0	0	0	0	1	1	1	1	0	0	0.67
5	0	0	0	0	1	1	1	1	0	1	0.71
6	0	0	0	0	1	1	1	1	1	0	0.74
7	0	0	0	1	0	0	0	0	1	0	0.64
8	0	0	0	1	0	0	0	0	1	1	0.67
9	0	0	0	1	0	0	0	1	0	0	0.72
10	0	0	0	1	0	0	1	0	0	0	0.77

Figure 5. Range of absolute deviations for network of a given size. The dots represent mean absolute deviations.

Results and Discussions

The SVM software developed by Royal Holloway University of London and AT&T Speech and Image Processing Service Research Lab (Saunders et al. 1998) was used in this study. There is also a wide array of other SVM software available online (e.g., http://www.kernel-machines.org).

Five hundred and ten sample patterns generated from possible combinations of 10 potential monitoring wells and corresponding probabilities of detection were used to train the SVM, and the rest (513) of the generated samples were used for testing. Two types of performance criteria are used here: the root mean square error (RMSE) and the absolute mean deviation, as defined below:

$$RMSE = \sqrt{\frac{\sum (\text{Predicted Value} - \text{Simulated Value})^2}{n}}$$

$$AMD = \frac{\sum |(\text{Predicted Value} - \text{Simulated Value}|}{n}$$

where n is sample size. One can also use other performance measures.

Figure 4 shows plots of performance of the SVM on training and testing patterns for the best performance over the testing pattern. It is important to note that the tot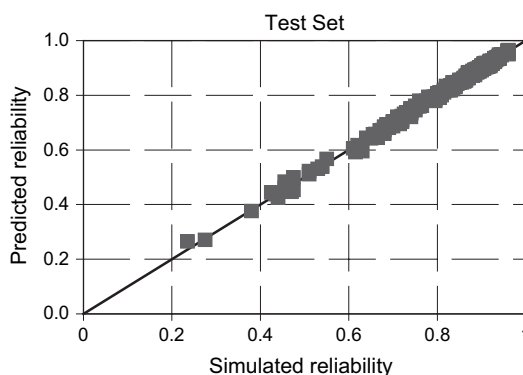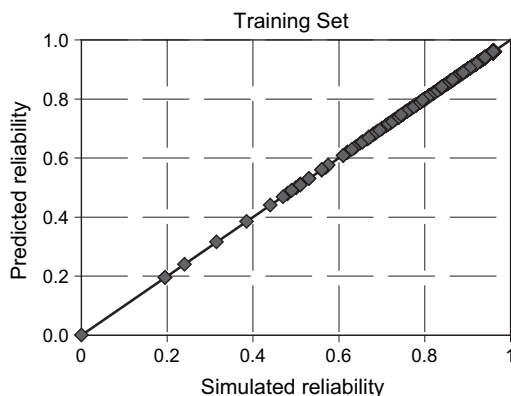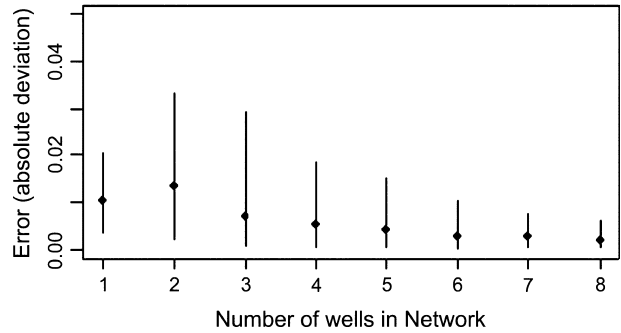al training and testing data came from a unique combination of the 10 potential monitoring well locations. Therefore, network configurations used in testing have not been seen before by the SVM (during training). Similar reliabilities may be obtained with different numbers and positions of monitoring wells. Therefore, Figure 4 shows the combined performance of the SVM over all sizes of networks (in this case, 1 to 10). Figure 5 shows the range of prediction error of probability of leak detection for networks of different sizes during testing. The parameters for Figure 5 are $\varepsilon = 0.001$, $C = 90$, and a strongly regularized Fourier kernel of $\gamma = 5$ as given by Vapnik (1998). The resulting number of support vectors was 456. For a pre-specified value of $\varepsilon = 0.001$, good performance was obtained for a value of C between 10 and 100, the best generalization being at $C = 90$. Figure 6 shows the RMSE for the two types of kernels chosen.

The support vectors seem to converge to a constant value as C increases. This is shown in Figure 7. Another result to note here is the change in the number of support vectors with change in value of ϵ. With a decrease in ε, the number of support vectors increases (Figure 8). The fact that ε controls the number of support vectors is expected and can be explained as follows (Vapnik 1998). Points that lie on or outside the ε-tube (Figure 2) define the support vectors, and it is necessary to note that we obtained our approximation by solving the optimization problem given by Equation 2 for which the KT conditions hold. The Lagrange multipliers can be regarded as forces,
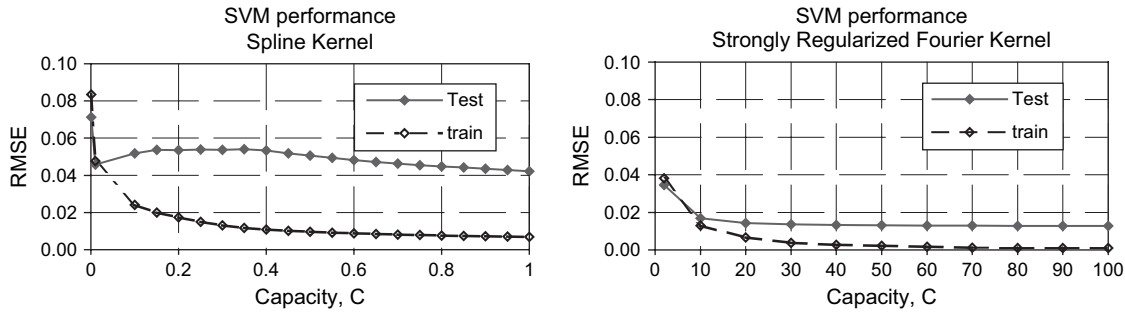

Figure 4. SVM performance in training and testing data.

Figure 6. Spline and strongly regularized Fourier kernels.

and the approximation corresponds to a flexible rod one would like to fit into the ε-tube. The kernel function k defines the law of elasticity. By narrowing the width of the tube the number of points that would lie on or outside the ε-tube increases. Conversely, by setting the number of support vector, one can calculate the corresponding ε (Schölkopf et al. 1999a).

Comparison with ANNs

ANNs are by far the most popular data-driven models that have been used extensively in the past couple of decades in approximating physically based hydrological models. Interested readers are pointed to the review on applications of ANNs in modeling hydrological processes by the American Society of Civil Engineers (ASCE) Task Committee (ASCE 2000).

We trained ANNs using the same training data explained in the previous section. Figure 9 compares the fraction of plumes detected by the SVMs, ANNs, and the physical models during testing for the best-performing monitoring networks. Even though both SVMs and ANNs perform well compared to the physical model, improvements are shown by the SVMs over the ANNs. As shown in the figure, with increase in the network size, the probability of detecting more and more plumes increases. Table 2 presents the locations as well as reliabilities provided by the top two best-performing networks of SVMs, ANNs, and the physical models. Columns 3, 5, and 7 show the locations of these wells as selected by these models. It can be seen that the reliability estimations of SVMs are better (closer to the physical models) than those of ANNs. It is also important to note that the best network

configurations identified by both SVMs and the physical models are identical, whereas differences exist with that of ANNs and the physical models. Columns 9 and 10 represent percentage change in network reliabilities by both ANNs and SVMs as compared to the physical models. Once again, reliabilities estimated by SVMs are closer to the physical models than that of ANNs.

It is interesting to note that wells that have good performance individually may not necessarily be in the best-performing group. For example, well 6 has individually the best detection performance, but it is not in the highest detecting groups according to both SVM and the physical model. Based on budgetary limits one can select the monitoring network size. The higher the number of wells within a network, the higher the degree of reliability these wells provide and the higher the cost. The decision problem is then to select the number and locations of the wells that have the least cost for a given reliability level. Alternatively, for a given number of wells, one may select the network that provides the highest reliability.

Conclusions

We have shown a possible application of SVMs in learning the relationship between network configurations and ground water contamination detection reliabilities that was obtained by ground water flow and transport models. The flow and transport simulations were kept relatively simple to save computational time. This is not a limitation of the SVM approach. The approach shown here can easily be extended to more complex site conditions and flow and transport simulations. A continuous
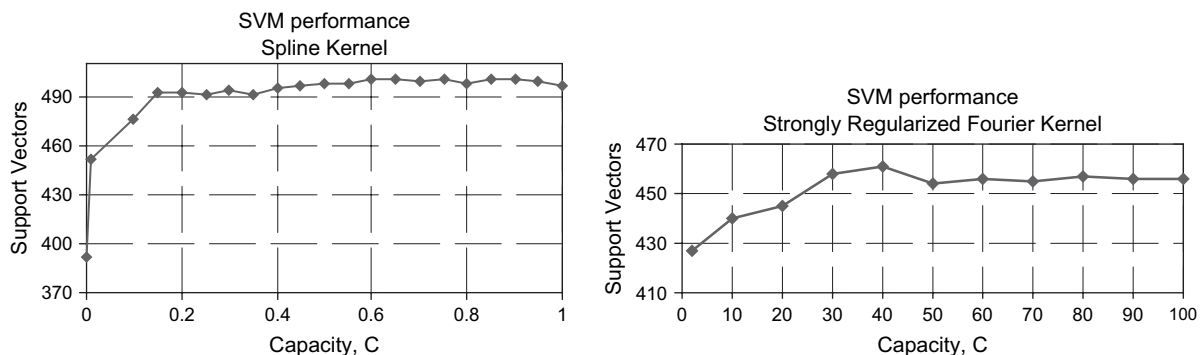

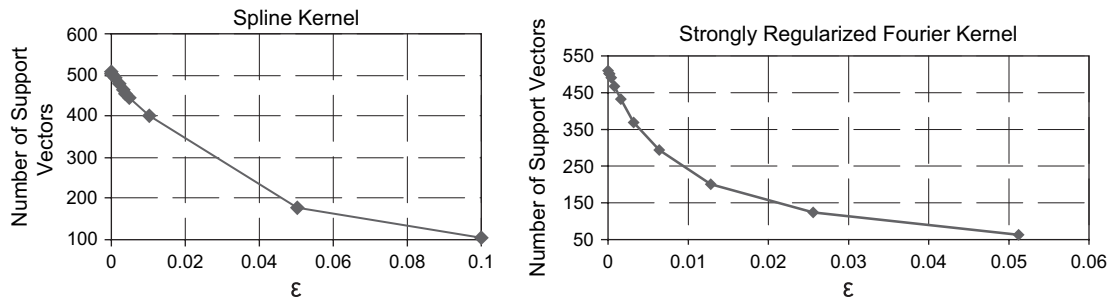**Figure 7. Change of number of support vectors vs. *C*.**

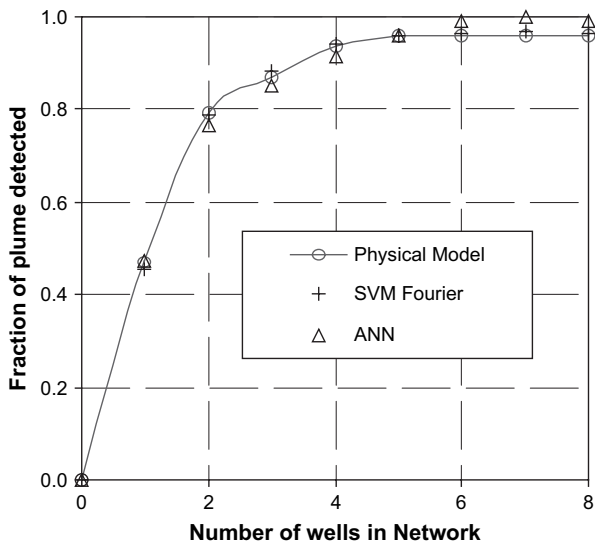**Figure 8. Change in number of support vectors with ε.**

Figure 9. Best SVM performance compared to ANN and physical model.

sampling scheme is also assumed, while in practice one may need to consider the frequency of sampling. More than 1000 samples generated from 10 possible monitoring well locations and the corresponding reliabilities are used for training and testing the SVM. Two hundred

Monte Carlo simulations were used to produce the reliability values. Our initial test suggested that this number is enough for the value of the reliabilities to be independent of the number of simulations. A larger number of potential monitoring locations can also be considered, but a larger number of potential monitoring locations will need a larger number of training and testing sets, which can become computationally costly for both the physical model and SVMs.

SVMs' selected combination of best-performing wells for all sizes of networks was identical with the one obtained from the physical model. Not only were these configurations identical, but they also gave very close estimates of reliabilities. These results are also better than that of the most commonly used ANNs. The performance of the strongly regularized Fourier kernel was slightly better than that of the spline kernel. For a prespecified error level of 0.001, the best value of capacity, C, seems to occur between 30 and 100. The number of support vectors remains more or less the same at about 456 (from a training size of 510). The selection of the SVM hyperparameters is still a heuristic exercise.

Based on the results of this study, SVMs have shown remarkable performance in learning network configuration/reliability relationships obtained by the physical ground water flow and transport models. The main contribution of

Table 2
Selected Monitoring Wells Based on SVM, ANN, and Flow and Transport Simulation

| Network Size | Rank | Physical Model | | Trained SVM | | Trained ANN | | Percentage Change | |
		Well Nos.	Reliability	Well Nos.	Reliability	Well Nos.	Reliability	ANN	SVM
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	1	6	0.475	6	0.455	4	0.474	0.21	4.21
	2	4	0.44	4	0.426	6	0.439	0.23	3.18
2	1	3, 7	0.795	3, 7	0.79	3, 5	0.766	3.65	0.63
	2	1, 5	0.76	1, 5	0.78	1, 5	0.687	9.61	2.63
3	1	3, 5, 10	0.87	3, 5, 10	0.884	3, 5, 10	0.849	2.41	1.61
	2	1, 5, 8	0.84	4, 6, 9	0.84	4, 6, 7	0.824	1.90	0.00
4	1	1, 4, 5, 9	0.935	1, 4, 5, 9	0.94	1, 4, 5, 9	0.913	2.35	0.53
	2	3, 5, 8, 9	0.93	3, 5, 8, 9	0.935	3, 6, 7, 9	0.906	2.58	0.54
5	1	1, 3, 5, 7, 9	0.96	1, 3, 5, 7, 9	0.96	1, 3, 5, 7, 9	0.958	0.21	0.00
	2	1, 4, 6, 8, 9	0.96	1, 4, 6, 8, 9	0.954	1, 3, 5, 9, 10	0.954	0.63	0.63

the study lies in reducing the huge efforts required in using physically based flow and transport models. One important application of the present research is within the remediation optimization framework. Once properly trained, SVMs will effectively replace cumbersome and time-consuming flow and transport codes that have to be run a relatively large number of times in order to obtain the best remediation strategy (Smalley et al. 2000). For example, Rogers and Dowla (1994) and Johnson and Rogers (2000) used ANNs to substitute time-consuming flow and transport models in an optimization study of ground water remediation. Others have used linear and nonlinear regression tools to substitute transport models (Lefkoff and Gorelick 1990; Ejaz and Peralta 1995).

SVMs are based on state-of-the-art machine learning techniques that have a solid statistical background, and they show good promise in applications in hydrology in general and in ground water in particular. We have shown a forward process approximation application of SVMs. Future work will focus on extending SVMs to address hydrologic inverse problems. There are encouraging preliminary applications in this area. For example, density estimation (Vapnik 1998; Weston et al. 1999; Mukherjee and Vapnik 1999) is a classical inverse problem that has been tackled successfully using SVMs.

Acknowledgements

We are grateful for the thoughtful review and suggestions by three anonymous reviewers.

Appendix: Kernels

The following are commonly used kernels for SVMs:

1. Simple dot product

$$K(x_i, x_j) = x_i \bullet x_j$$

2. Radial basis function

$$K(x_i, x_j) = \exp\left(-\gamma |x_i - x_j|^2\right)$$

where γ is user defined

3. Linear spline with an infinite number of points

For a one-dimensional case,

$$1 + x_i x_j + x_i x_j \min(x_i, x_j) - \frac{x_i + x_j}{2}(\min(x_i, x_j))^2 + \frac{(\min(x_i, x_j))^3}{3}$$

For a multidimensional case,

$$K(x_i, x_j) = \prod_{k=1}^{n} K_k(x_i^k, x_j^k)$$

4. Strongly regularized Fourier kernel

For the one-dimensional case,

$$\frac{1 - \gamma^2}{2(1 - 2\gamma \cos(x_i - x_j) + \gamma^2)}$$

where γ is user defined

For the multidimensional case,

$$K(x_i, x_j) = \prod_{k=1}^{n} K_k(x_i^k, x_j^k)$$

References

Angulo, M., and W.H. Tang. 1999. Optimal ground water detection monitoring system design under uncertainty. *Geotechnical and Geoenvironmental Engineering* 125, no. 6: 510–517.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (Rao Govindaraju). 2000. Artificial neural networks in hydrology. II: Hydrologic applications. *Journal of Hydrologic Engineering* 5, no. 2: 124–137.

Boser, B.E., I. Guyon, and V. Vapnik. 1992. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth ACM Workshop on Computational Learning Theory*, ed. D. Haussler, 144–152. Pittsburg, Pennsylvania: ACM.

Church, R.L., and C. ReVelle. 1974. The maximal covering location problem. *Papers of the Regional Science Association* 32, 101–118.

Cristianini, N., and J. Shawe-Taylor. 2000. *An Introduction to Support Vector Machines and Other Kernel Based Learning Methods*, 1st ed. Cambridge, UK: Cambridge University Press.

Datta, B. and S.D. Dhiman. 1996. Chance-constrained optimal monitoring network design for pollutants in ground water. *Journal of Water Resources Planning and Management* 122, no. 3: 180–188.

Dibike, B.Y., S. Velickov, D. Solomatine, and B.M. Abbot. 2001. Model induction with support vector machines: Introduction and applications. *Journal of Computing in Civil Engineering* 15, no. 3: 208–216.

Ejaz, M.S., and R.C. Peralta. 1995. Modeling for optimal management of agricultural and domestic waste water loading to streams. *Water Resources Research* 31, no. 4: 1087–1096.

Harbaugh, A.W., and M.G. McDonald. 1996. User's documentation for MODFLOW-96, an update to the U.S. Geological Survey modular finite-difference ground-water flow model. USGS Open-File Report 96–485.

Hudak, P.F., and H. Loaiciga. 1992. A location modeling approach for ground water monitoring network augmentation. *Water Resources Research* 28, no. 3: 643–649.

James, B.R., and S.M. Gorelick. 1994. When enough is enough: The worth of monitoring data in aquifer remediation design. *Water Resources Research* 30, no. 12: 3499–3513.

Jardine, K., L. Smith, and T. Clemo. 1996. Monitoring networks in fractured rocks: A decision analysis approach. *Ground Water* 34, no. 3: 504–518.

Johnson, V.M., and L.L. Rogers. 2000. Accuracy of neural network approximators in simulation-optimization. *Journal of Water Resources Planning and Management* 126, no. 2: 48–56.

Kaneviski, M., A. Pozdnukhov, S. Canu, and M. Maignan. 2002. Advanced spatial data analysis and modeling with support vector machines. *International Journal of Fuzzy Systems* 4, no. 1: 606–615.

Kowk, J.T. 2001. Linear dependency between ε and the input noise in ε-support vector regression. In *ICANN 2001, LNCS2130*, ed. G. Dororffner, H. Bishof, and K. Hornik, 405–410.

Lefkoff, L.J., and S.M. Gorelick. 1990. Simulating physical processes and economic behavior in saline, irrigated agriculture: Model development. *Water Resources Research* 26, no. 7: 1359–1369.

Liong, S.Y., and C. Sivapragasam. 2002. Flood stage forecasting with SVM. *Journal of the American Water Resources Associations* 38, no. 1: 173–186.

Loaiciga, H.A., R.J. Charbeneau, L.G. Everett, G.E. Fogg, B.F. Hobbs, and S. Rouhani. 1992. Review of ground water

quality monitoring network design. *Journal of Hydrologic Engineering* 118, no. 1: 11–37.

Mahar, P.S., and B. Datta. 1997. Optimal monitoring network and ground water pollution source identification. *Journal of Water Resources Planning Management* 123, no. 4: 199–207.

Massmann, J., and R.A. Freeze. 1987a. Ground water contamination from waste management sites: The interaction between risk-based engineering design and regulatory policy, 1. Methodology. *Water Resources Research* 23, no. 2: 351–367.

Massmann, J., and R.A. Freeze. 1987b. Ground water contamination from waste management sites: The interaction between risk-based engineering design and regulatory policy, 2. Results. *Water Resources Research* 23, no. 2: 368–380.

Mattera, D., and S. Haykin. 1999. Support vector machines for dynamic reconstruction of a chaotic system. In *Advances in Kernel Methods: Support Vector Learning*, ed. B. Schölkopf, C.J.C. Burges, and A.J. Smola, 211–242. Cambridge, Massachusetts: MIT Press.

McDonald, M.G., and A.W. Harbaugh. 1988. A modular three-dimensional finite-difference ground-water flow model. Techniques of water resources investigations. USGS 06–A1.

Meyer, P.D., and E.D. Brill. 1988. Method for locating wells in a ground water monitoring network under conditions of uncertainty. *Water Resources Research* 24, no. 8: 1277–1282.

Meyer, P.D., A.J. Valocchi, and J.W. Eheart. 1994. Monitoring network design to provide initial detection of ground water contamination. *Water Resources Research* 30, no. 9: 2647–2659.

Molina, G.R., J.J. Beauchamp, and T. Wright. 1996. Determining an optimal sampling frequency for measuring bulk temporal changes in ground water quality. *Ground Water* 34, no. 4: 579–587.

Montas, H.J., R.H. Mohtar, A.E. Hassan, and F. AlKhad. 2000. Heuristic space-time design of the monitoring wells for contaminant plume characterization in stochastic flow fields. *Journal of Contaminant Hydrology* 43: no. 3–4: 271–301.

Morisawa, S., and Y. Inoue. 1991. Optimum allocation of monitoring wells around a solid-waste landfill site using precursor indicators and fuzzy utility functions. *Journal of Contaminant Hydrology* 7: no. 4: 337–370.

Mukherjee, S., and V. Vapnik. 1999. Support vector method for multivariate density estimation. Center for Biological and Computational Learning. Department of Brain and Cognitive Sciences, MIT.C. B.C.L. No. 170.

Pollock, D.W. 1989. Documentation of computer programs to compute and display pathlines using results from USGS modular three-dimensional finite-difference ground water model. USGS Open-File Report 89–389.

Pollock, D.W. 1988. Semianalytical computation of path lines for finite difference models. *Ground Water* 26, no. 6: 743–750.

Reed, P., B. Minsker, and A.J. Valocchi. 2000. Cost-effective long-term ground water monitoring design using a genetic algorithm and global mass interpolation. *Water Resources Research* 36, no. 12: 3731–3741.

Rogers, L.L., and F.U. Dowla. 1994. Optimization of ground water remediation using artificial neural networks *Water Resources Research* 30, no. 2: 457–481.

Saunders, C., M.O. Stitson, J. Weston, L. Bottou, B. Schölkopf, and A. Smola. 1998. Support vector machine reference manual. Royal Holloway Technical Report CSD-TR-98-03. Royal Holloway.

Schölkopf, B., P.L. Bartlett, A. Smola, and R. Williamson. 1999a. Shrinking the tube: a new support vector regression algorithm. In *Advances in Neural Information Processing Systems 11*, ed. M.S. Kearns, S.A. Solla, and D.A. Cohn, 330–336. Cambridge, Massachusetts: MIT Press.

Schölkopf, B., J.C., Burges, and A. Smola. 1999b. *Advances in Kernel Methods: Support Vector Learning*. Cambridge, Massachusetts: MIT Press.

Schölkopf, B., and A. Smola. 2002. *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. Cambridge, Massachusetts: MIT Press.

Smalley, J.B., B.S. Minsker, and D.E. Goldberg. 2000. Risk-based in situ bioremediation design using noisy genetic algorithm. *Water Resources Research* 36, no. 20: 3043–3052.

Smola, A., B. Murata, B. Schölkopf, and K.R. Müller. 1998. Asymptotically optimal choice of epsilon-loss for support vector machines. In *Proceedings of ICANN '98, Perspectives in Neural Computing*, ed. L. Niklasson, M. Bodén, and T. Ziemke, 105–110. Berlin, Germany: Springer Verlag.

Storck, P., J.W. Eheart, and A.J. Valocchi. 1997. A method for the optimal location of monitoring wells for detection of ground water contamination in three dimensional heterogeneous aquifers. *Water Resources Research* 33, no. 9: 2081–2088.

Tompson, A.F.B., R. Ababou, and L.W. Gelhar. 1989. Implementation of the 3-dimensional turning bands random field generator. *Water Resources Research* 25, no. 10: 2227–2243.

Vapnik, V. 1998. *Statistical Learning Theory*. New York: Wiley.

Vapnik, V. 1995. *The Nature of Statistical Learning Theory*. New York: Springer.

Vapnik, V., S. Golowich, and A. Smola. 1997. Support vector method for function approximation, regression estimation, and signal processing. In *Advances in Neural Information Processing Systems 9*, ed. M. Mozer, M. Jordan, and T. Petsche, 281–287. Cambridge, Massachusetts: MIT Press.

Weston, J., A. Gammerman, M.O. Stitson, V. Vapnik, V. Vovk, and C. Watkins. 1999. Support vector density estimation. In *Advances in Kernel Methods: Support Vector Learning*, ed. B. Schölkopf, J.C. Burges, and A. Smola, 293–305. Cambridge, Massachusetts: MIT Press.