UK/KRCEE Doc #: P12.3 2007

Using the Box Complex Method to Optimize Pumping Strategies at the PGDP

Prepared by
Kentucky Research Consortium for Energy and Environment
233 Mining and Minerals Building
University of Kentucky, Lexington, KY 40506-0107

Prepared for
United States Department of Energy Portsmouth/Paducah Project Office
Acknowledgment: This material is based upon work supported by the Department of Energy under Award Number DE-FG05-03OR23032.

May 2007

UK/KRCEE Doc #: P12.3 2007

Using the Box Complex Method to Optimize Pumping Strategies at the PGDP

By

Joshua D. Kopp
University of Kentucky
College of Engineering
Department of Civil Engineering
Lexington, Kentucky

May 2007

CONTENTS

FIGURES

TABLES

1. INTRODUCTION

1.1 BACKGROUND

Over the last several decades numerous hazardous waste sites have taken a heavy toll on the nation's environment. Through the release of toxic chemicals or other pollutant such sites have contaminated either the air, soil, or water beneath them. Depending on the type of contaminant, these areas may not be safe for human habitation. Any site that has been contaminated with a hazardous waste and poses a risk to human health or the environment has been classified as a Superfund site by the EPA (EPA 2007). Those sites that pose the greatest environmental risk have been classified as national priority list (NPL) sites, and are eligible for federal clean-up dollars.

Of the 15 active NPL sites in Kentucky, the Paducah Gaseous Diffusion Plant (PGDP) is contaminated the worst. The PGDP is an active uranium enrichment facility located in approximately 10 miles West of Paducah, Kentucky and 3.5 miles south of the Ohio River (KRCEE 2007).

The area of study for this project is the Paducah Gaseous Diffusion Plant (PGDP) and surrounding areas that are enclosed by the DOE Water Policy Boundary (Figure 1). The Water Policy Boundary was defined by DOE as the area that contains or has potential to contain properties overlying the contamination plume (KRCEE 2007). The PGDP site is located on land owned by the DOE. Other property in the water policy boundary is owned by the Tennessee Valley Authority (TVA), the West Kentucky Wildlife Management Area (WKWMA), and private owners.

At the PGDP site, soil and groundwater has been contaminated with trichloroethylene (TCE). TCE is a volatile organic chemical (VOC) and is part of a family of synthetic chlorinated hydrocarbons. It has been manufactured as a solvent with its greatest appeal being a reduced potential for fire or explosion (Ensley, 1991). TCE is typically a colorless or blue organic liquid (EPA 2007) with an odor like chloroform and a sweet, burning taste. TCE was used as a solvent in the degreasing of metal parts at the PGDP site.

A common method of TCE entering the environment is by leaching into the soil. TCE has a tendency to stick to soil particles and remain there for a long time (ASTDR 2007). This will lead to TCE contaminating the groundwater and potentially nearby surface water. TCE does not last long in surface water and will evaporate quickly so it is commonly found as a vapor in the air (ATSDR 2007). TCE at the PGDP site has leached into the soil and reached the groundwater. Currently groundwater seepage is transporting the TCE towards the Ohio River. TCE has also been found in drinking water wells around the PGDP site.

The long term health effects associated with exposure to TCE are not yet completely understood. However, the Environmental Protection Agency (EPA) has set a Maximum

Contaminant Level Goal (MCL) for TCE of 5 parts per billion (ppb) or 5 μg/L. This is the value at which none of the potential health problems caused by TCE should occur.

TCE has been determined to be *"probably carcinogenic to humans*" by the International Agency on Research for Cancer (IARC) (ATSDR 2001). Therefore it poses a potential health risk to the local population. Drinking water with amounts over the MCL for an extended period of time could result in liver and kidney damage (ATSDR 2007). There is also some evidence suggesting that TCE can impair fetal development in pregnant women (ATSDR 2007).

Some residents around the PGDP site get drinking water from wells which could be contaminated with TCE. These wells have now been identified and the users given a municipal supply of drinking water (KRCEE 2007). These residents have agreed to not drill any more wells, however future residents may still drill wells which could lead to possible human exposure to TCE contamination (KRCEE 2007).

1.2 PROBLEM STATEMENT

Since 1997 a pump-and-treat (P&T) operation has been used to try and contain the spread of the existing TCE plume. Extraction wells placed around the site extract groundwater to the surface. Once the contaminated groundwater has reached the surface it is treated by air-stripping to remove the TCE. The location of the P&T wells currently in operation is shown in Figure 2. The theoretical P&T wells shown in the figure represent potential wells that could be added to the system to increase the removal of contaminated groundwater. Observations wells have been drilled to measure TCE concentrations down gradient of the plant.

To determine the future extent of the TCE contamination plume an artificial neural network (ANN) has been developed. This model has inputs of pumping rates at extraction wells and will forecast TCE concentrations in an observation well for future years. This ANN model developed using the theoretical and existing extracting wells as inputs, and the TCE concentrations at Observation Well Bayou-1 for years 2009, 2015, 2021, and 2027 as outputs.

If there is a need to reduce the TCE concentration below a maximum concentration (i.e. the MCL), the pumping rates (inputs) of this model can be changed until this concentration is obtained. Numerous combinations of pumping rates are available to achieve the concentration. Since all extraction wells cost money to operate, the question then becomes, what is the optimum combination of pumping rates to minimize the cost? To determine this optimum combination of pumping rates, the ANN model will be coupled with a box-complex optimization technique. The optimization aspect will be to minimize the cost of the pumping while not validating maximum concentration constraint. Only year 2027 will be evaluated for this study.

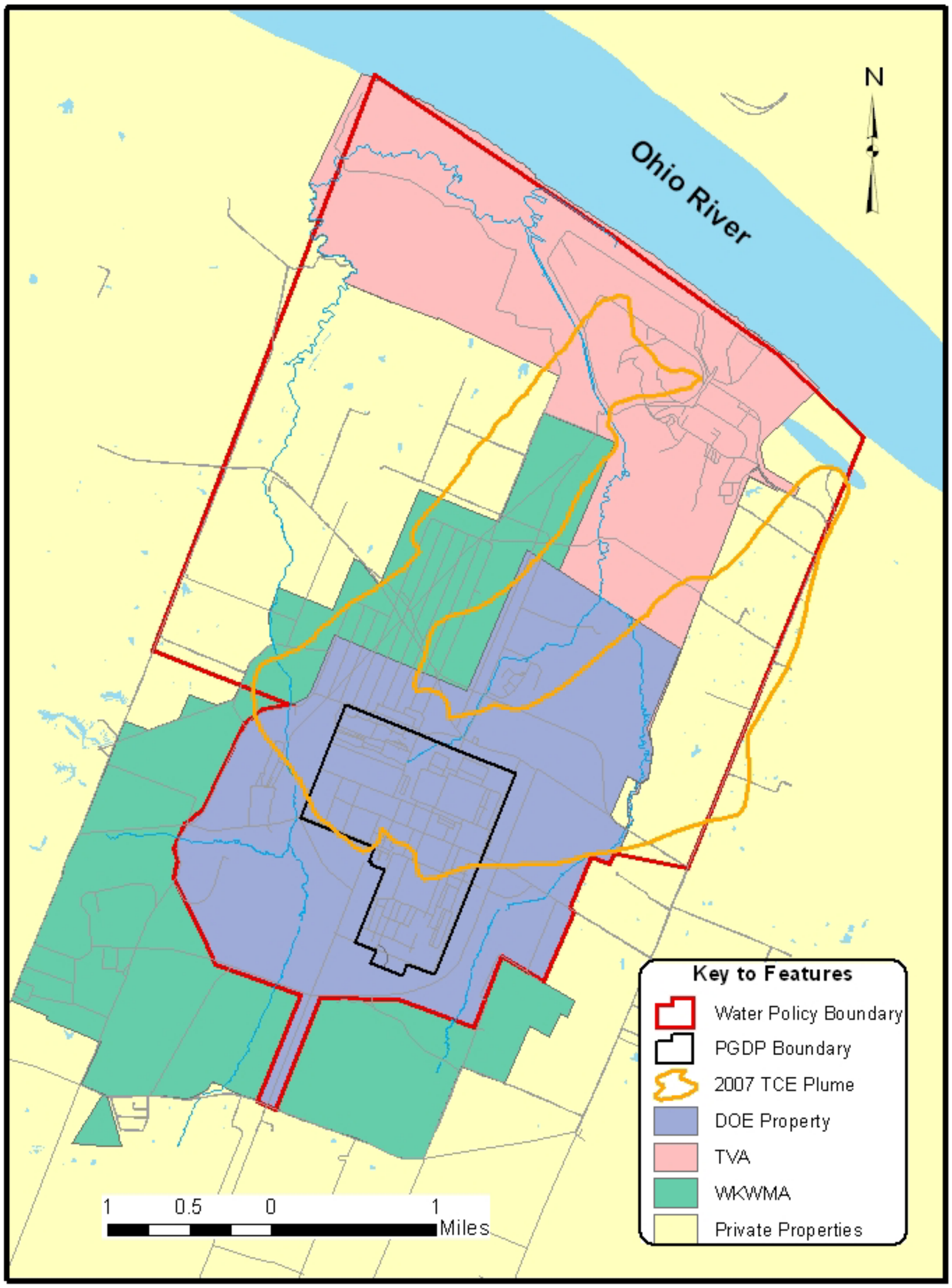

Figure 1: Layout of the PGDP and Surrounding Areas Including the Modeled Existing TCE Plume

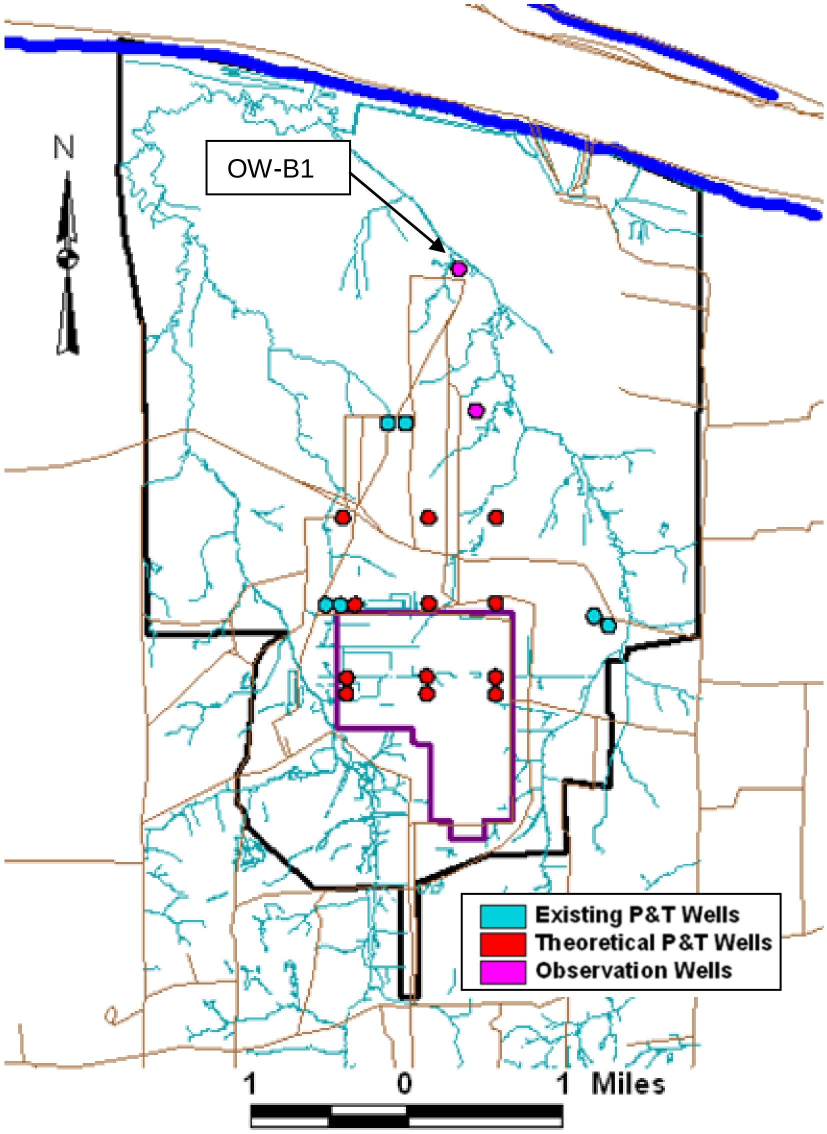

Figure 2: Location of Wells in PGDP Site used in ANN Models

2. MATHEMATICAL FORMULATION OF PROBLEM

2.1 OBJECTIVE FUNCTION

The objective of this optimization is to minimize the pumping costs of the extraction wells at the PGDP. The only costs considered are the cost per unit discharge. Capital costs of pumps are not considered. Cost for each pump was arbitrarily determined for the purpose of this project. The arbitrarily chosen values of cost for each well are shown in Table 1. The objective function can be stated as follows:

$$Min(z) = \sum_{i=1}^{n} q_i C_i \qquad \textbf{Equation 1}$$

where z is the total cost for all wells per unit time, q_i is the pumping rate at the ith well and C_i is the cost per unit volume of the ith well. The total cost for the length of the P&T process can then be found by multiplying z by the length of time that the pumps will be in operation. It was assumed for this study that the wells would be operating twelve hours per day, 365 days per year, for a total of 20 years. So the total cost evaluated will be the value of z multiplied by 20 years.

Table 1: Pumping Costs (arbitrarily chosen)

Extraction Well	Cost ($/gal.)
1	0.05
2	0.06
3	0.04
4	0.05
5	0.06
6	0.04
7	0.05
8	0.06
9	0.04
10	0.05
11	0.06
12	0.04
13	0.05
14	0.06
15	0.04
16	0.05
17	0.06
18	0.04

2.2 CONSTRAINTS

An implicit constraint to the box-complex is that the pumping rates must decrease the TCE concentration in observation well Bayou-1 below 30 ug/L. This constraint will be evaluated based on an artificial neural network model developed by Kopp (2007). This ANN forecasts TCE concentrations for the year 2027 when given the pumping rates of the 18 wells in the PGDP area. When a point in the complex is determined it will be input into the ANN model to determine the concentration at the observation well. If this concentration is greater than 30 ug/L, the point is considered infeasible and a new point will be generated. This constraint's practical use is that it ensures progress is being made in reducing the concentration of TCE due to pumping and treating of contaminated groundwater. The ANN was embedded in the optimization to make it easier to evaluate.

There is an explicit constraint on the range of the pumping rates. They must be between 0 and 100 gpm. This is the range of pumping rates used to train the ANN model. This is also the reason that the constraint of 30 ug/L was used rather than the MCL of 5 ug/L. This constraint must be within the TCE values that the ANN model was trained with, which was about 25 to 80 ug/L. Therefore, 30 ug/L was chosen as the reduced concentration after 20 years of pumping. The MCL (5ug/L) was not used as the constraint since the ANN forecasts that even with the maximum pumping rate at all wells the MCL can still not be obtained.

3. DISCUSSION OF METHODS USED

3.1 ARTIFICIAL NEURAL NETWORKS

Artificial neural networks (ANN) are an inductive modeling technique used in many fields of research. ANNs are popularly applied in forecasting, pattern recognition, and classification problems. An ANN serves as an alternative to linear and non-linear regression and is very useful when the actual physical relationship between two or more variables is unknown. Thus they are considered a black box method where input is given to the model and an output and is then obtained.

A supervised training approach was used for this study. Supervised training is where inputs as well as known outputs are required to perform the training of the model. During the training process the ANN acquires the hidden knowledge by studying the inter-relationship between the input and output data. The goal of the training process is to minimize the error between the observed and the predicted output of the model.

A feed-forward multi-layered perceptron (MLP) was used as the model architecture. In this architecture neurons are arranged in layers. Input layer neurons are buffers. The hidden layer and output layer neurons are defined as activation functions. An activation function will transform the input and pass it to the subsequent layer. The result of the output neuron will be the output of the model to the user.

Neurons in different layers are interconnected by weights. The information is passed from left to right (i.e. input to hidden layer, hidden to output layer). The trained knowledge of the model is stored in these weights. Equation 3 shows how the trained knowledge of the weights as it is applied to the inputs.

$$\sum_{i=1}^{n} x_i w_{1i} = net_j$$ **Equation 2**

where x_i is the input from the *i*th neuron in the preceding layer and w_{1i} is the weight interconnecting the *i*th neuron in the preceding layer to neuron 1 in the hidden layer (see Figure 3). Net_j is the weighed sum and this information is then transformed by the activation function as shown in Equation 4.

$$f(net_j) = \frac{1}{1 + e^{-\lambda net_j}}$$ **Equation 3**

where λ is a learning rate which determines how much the function will transform. The interconnecting weights are redefined by the training algorithm and represent the knowledge gained by the ANN model. A new set of inputs can then be used and the process repeated.

A back-propagation training algorithm was used. The mathematical equivalent of this is a steepest descent algorithm. In this training algorithm the weights of the model are adjusted after each dataset is run through the model. An epoch is known as one complete run of the datasets through the ANN model. Multiple epochs are performed in the model to properly obtain all the knowledge.

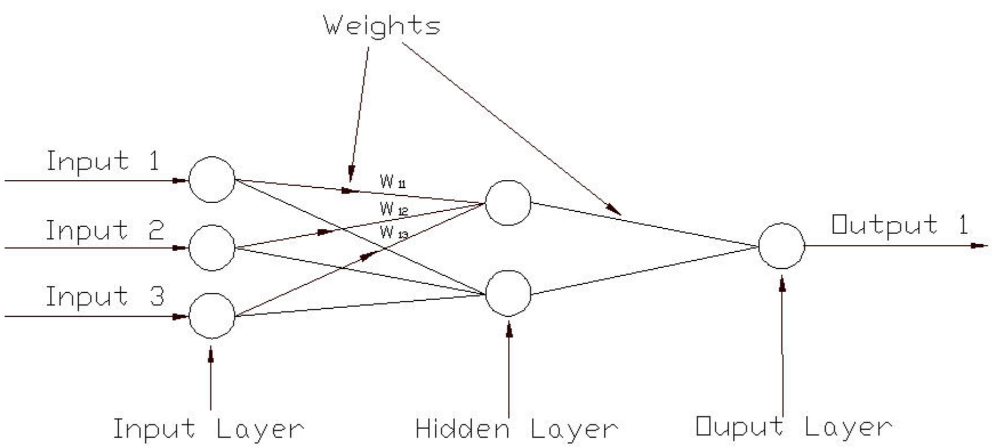

Figure 3: General ANN Model Architecture

Each ANN model will be different and will depend on the data available to train with, the desired output, and the architecture of the model used. The architecture of the ANN model is dependent highly on the type of problem being considered (Maier and Dandy, 1999). Numerous studies have shown that the best setup for an ANN consists on one input, one hidden, and one output layer. However this is not always the case, as some functions may prove to be difficult to approximate with one hidden layer thus requiring an additional layer of hidden nodes (Cheng and Titterington, 1994). The number of input nodes is fixed to the number of model inputs while the number of output nodes is fixed to the number of model outputs. The number of nodes in the hidden layer is critical since it will determine the number of connection weights (Maier and Dandy 1999).

3.1.1 Training, Testing, and Validation

The more datasets that are available to train and validate the model, the more accurate the model will be. After training has finished a model needs to be validated to ensure that it is a robust model and that it is not over training the data. Validation needs to occur with a separate group of datasets that has not been used in training or testing process (Maier and Dandy 1999).

Caution needs to be taken to prevent a biased model. The most standard way to do this is to divide the datasets into three sub-sets: a training set, a testing set, and a validation set. Literature typically suggests division of data into training/testing and validation of 80% vs. 20% or 70% vs. 30%. The training and testing data is then further divided by the same percentage as the previous division. Each sub-set of data must be representative of the entire dataset to ensure good training (Maier and Dandy 1999). Obviously the training set is used to train the ANN model. The testing set should not be used in training so that the data is new to the model. It can then be simulated in the model and the error results from training and testing can be compared. If the final error indices are similar for training and testing data then the model can be considered robust.

If the training and testing errors are significantly different, then memorization most probably took place during training and a better model will need to be developed. Too many connection weights will allow overtraining of the data which is where the model has learned the idiosyncrasies of the training set (memorization of data) and thus a loss of ability to generalize (Maier and Dandy 1999). When memorization occurs, the model has been over-trained and has captured noise from the dataset. Memorization of the model can be detected by a continual reduction in the training set error while the testing set error remains the same or becomes worse (Maier and Dandy 1999).

3.1.2 Activation Functions

Types of activation functions commonly used are logistic sigmoid (unipolar activation) with an output variation of 0 to 1, hyperbolic tangent sigmoid (bipolar activation) with an output variation of -1 to 1, and linear that only has values of 0 and 1. Maier and Dandy (1999) found that other transfer functions may be used as long as they are differentiable. Normalization of the data must take place to ensure that values of net_j stay with the range of the function Typically normalization occurs by dividing a category of inputs by the maximum value in that category. The unipolar activation function was used for the ANN models in this study. The graph of this function is shown in Figure 4.

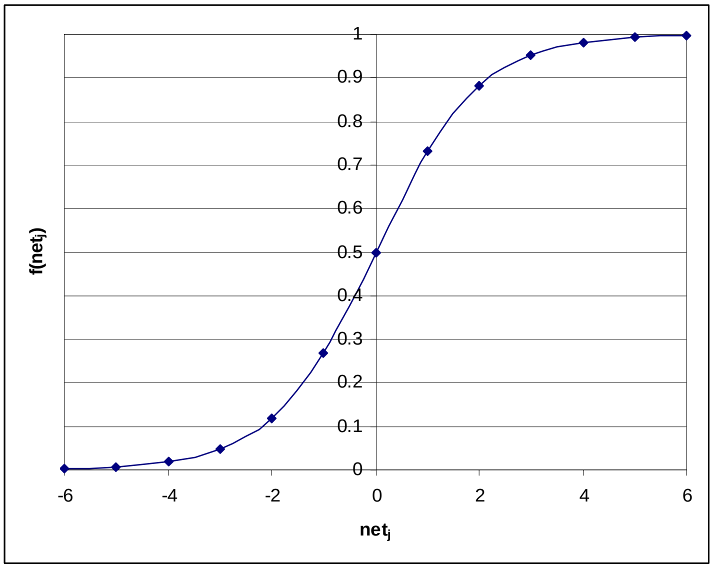

Figure 4: Graph of the Unipolar Activation Function

3.1.3 Learning Rate and Momentum Parameter

Parameters other than the activation function need to be considered as well depending on the training algorithm chosen. For the back propagation algorithm, the two most common are the learning rate and momentum parameter. In this algorithm, the learning rate will dictate the magnitude of the weight changes. Values range from 0 to 1 and choosing a learning rate for the ANN model will have a significant impact on the results. Larger learning rates will move the algorithm too quickly and possibly skip the optimal solution. Small learning rates increase the computational time of the model. The momentum parameter is meant to improve the BPA by allowing for a larger learning rate that will result in faster convergence of the model but will minimize the tendency to bypass the optimal solution (Rumelhart et al., 1986).

3.2 BOX-COMPLEX METHOD

The Box-Complex method is an algorithm used to determine a set of decision variables to optimize an objective function developed by Box (1965). A complex is a flexible mathematical figure made up of at least $n+1$ points where n is the number of variables (Ormsbee 1981). The complex lies in n dimensional space. Each point consists of coordinates which corresponds to individual variables of the objective function (Ormsbee

1981). The complex moves around the solution space by expanding in contracting in any direction as long as it is feasible.

The generation of the initial complex begins with determining a feasible initial point that satisfies both explicit and implicit constraints. Implicit constraints are those that limit the value of some group of variables (i.e. F(x)<0) and explicit constraints limit the values of an individual variable (i.e. 0<Xi<100). Once this initial feasible point has been determined, a random number generator is used to obtain the remaining points of the initial complex. The random number generator should be set up to return variables within the range of the explicit constraints. It is then necessary to check and see if the point satisfies the implicit constraints.

If an infeasible point is generated the following process will move it back towards feasibility. First, determine the centroid of the feasible points already determined (including the initial point). Move the infeasible point halfway towards this centroid. If the point is still infeasible continue moving it half the remaining distance towards the centroid until is becomes feasible. Continue this process until *n+1* feasible points have been generated to form the initial complex.

Expansion and contraction of the complex may now take place. Compute the value of the objective function at each point in the complex. Determine the point that produces the worst results (P_{worst}) (worst is defined as opposite the goal of the objective function). A new point (P_{new}) is then determined by going a specific distance away from P_{worst} in the direction of the centroid of the remaining feasible points, $P_{centroid}$.

$$P_{new} = (1+\alpha)P_{centroid} - \alpha P_{worst} \quad \textbf{Equation 4}$$

The value α is an expansion coefficient and Box recommended a value of 1.3. Evaluate the objective function at P_{new} and determine if it is better than P_{worst}. If P_{new} is better, P_{worst} is disregarded and P_{new} becomes part of the complex. If P_{new} is worse than P_{worst}, then a new point P_{new2} is contracted back towards the centroid at another specified distance based on the contraction coefficient.

$$P_{new2} = \omega P_{new} + (1+\omega)P_{centroid} \quad \textbf{Equation 5}$$

A value of 0.5 is recommended as this contraction coefficient (Tufail 2007). This continues until a P_{new} is obtained that produces a better value of the objective function than P_{worst}. This process shifts the complex towards better values of the objective function.

Eventually this process of expansion and contraction will shrink the complex near the optimal values of the objective function. It will terminate after consecutive objective functions give the same result, indicating that the complex has converged on the centroid (Ormsbee 1981). For a more in depth description of the Box-Complex method, please see Box (1965), Ormsbee (1981), and Tufail (2007).

3.2.1 Computer Code

To perform the box-complex algorithm, a C++ computer code was used. This code was originally developed by Tufail (2007) for two variables and was adjusted to fit the problem of this report with 18 variables.

4. RESULTS AND DISCUSSION

The box complex program was run for a number of different iterations. Table 2 shows each run with the number of iterations, the constraint value, and the maximum cost for 20 years of operation. There was minimal improvement after 20,000 iterations, but it was not significant enough to display. The final set of pumping rates from the 20,000 iteration run is displayed in Table 3.

Table 2: Total Pumping Cost

Iterations	2027 TCE Concentration (ug/L)	Total Cost (millions of dollars)
200	29.43	$136.4
500	29.89	$126.1
750	30.00	$123.4
1000	29.96	$121.4
1500	29.97	$121.0
2000	29.96	$120.8
5000	29.92	$120.2
10000	29.89	$120.0
20000	29.82	$119.4

Table 3: Pumping Rate at Each Well for Optimal Strategy

Extraction Well	Pumping Rate (gpm)	Cost per gpm
1	60.9	$0.05
2	25.8	$0.06
3	0.0	$0.04
4	54.6	$0.05
5	1.4	$0.06
6	35.8	$0.04
7	0.0	$0.05
8	0.0	$0.06
9	0.0	$0.04
10	72.9	$0.05
11	0.0	$0.06
12	5.5	$0.04
13	12.8	$0.05
14	50.5	$0.06
15	0.5	$0.04
16	66.8	$0.05
17	49.5	$0.06
18	0.0	$0.04

It is obvious that there are some wells that should be doing most of the pumping and others that should not be pumping at all. One interesting trend is that the wells with the lowest cost per gpm (i.e. $0.04) were not being used as much and the more expensive wells were being used more. This may be an indication that the influence of a well over the TCE concentration is more important than the unit cost per well. This shows that the total number of wells being used can be reduced by using wells that have a greater influence over the TCE concentration regardless of the unit cost per well. The results of the optimization show that by using this approach a lower cost can be achieved.

Table 2 shows another trend in the data that sticks out. After 1500 iterations, the cost and maximum value of the constraint both begin to decrease. This trend continues up to and past the 20,000 iterations. However, after 20,000 iterations it becomes insignificant. The reason for this may be the extremely high number of combinations that are possible meaning that there are local minimum that the optimization is picking up. It requires more iterations to get out of that local minimum to find the global minimum.

The program used allowed pumping values to reach up to 6 decimal places and obviously this would not be a practical value for a pump to operate. This was a limitation of the program and it could be fixed easily by an experienced programmer. However, for this report it was not fixed, thus pumping rates were computed for up to 6 decimal places and then rounded to one decimal place for presentation. If pumping rates were limited to integer values then the results would most likely be different. Wells with small pumping rates would most likely not be used for practical purposes (i.e. Extraction Wells 5, 12, 13, and 15). It may be better to completely eliminate these wells from the optimization and only used wells that are greater contributors.

5. CONCLUSIONS

The pumping strategy listed above will give the optimum performance to limit the total cost. However, it is not practical and must be adjusted. The adjustments can be made by making two simple adjustments to the program 1) limit the pumping rates to integer values and 2) require a minimum pumping rate (i.e. 15 gpm). These two adjustments will bring the optimization solution much closer to a practical pumping strategy that can reduce the TCE concentration below a set level.

WORKS CITED

ATSDR, 2007. Retrieved from (http://www.atsdr.cdc.gov/tfacts19.html#bookmark03)

ATSDR, 2001. Public Health Assessment, Paducah Gaseous Diffusion Plant, retrieved from (http://www.atsdr.cdc.gov/HAC/PHA/paducah/pad_toc.html)

Box, M. J. "*A New Method of Constraint Optimization and a Comparison with Other Methods.*"Computer J. 8(1965):42-52.

Cheng. B., and Titterington, D.M. 1994. *Neural Networks: A review from a statistical perspective*. Statistical Science 9 (1). 2-54.

Ensley, B.D. 1991. *Biochemical Diversity of Trichloroethylene Metabolism*. Annu. Rev. Microbiol. (45) 283-99.

EPA 2007. Superfund Frequently Asked Questions, retrieved from (http://www.epa.gov/superfund/contacts/index.htm)

Krishnapura, V.G. and Jutan, A. 1997. "*ARMA neuron networks for modeling nonlinear dynamical systems."* Canadian Journal of Chemical Engineering. 75 (3), 574–582.

KRCEE (Kentucky Research Consortium for Energy and Environment). 2007. *Property Acquisition Study for Areas near the Paducah Gaseous Diffusion Plant, Paducah, Kentucky*

Maier, H.R., and Dandy, C.D. 1999. "*Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications*." Environmental and Modelling Software. 15, 101-124.

Ormsbee, L. 1981. "*Optimization of Hydraulic Networks*." International Symposium on Urban Hydrology, Hydraulics, and Sediment Control. Lexington, KY. July 27-30.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. *"Learning internal representation by back- propagating errors."* In: Rumelhart, D.E., McCleland, J.L., the PDP Research Group (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition. MIT Press, MA.

Zhang, G., Patuwo, B.E., and Hu, M.Y. 1998. *"Forecasting with artifical neural networks: The state of the art."* International Journal of Forecasting. 14, 35-62.