

## Planning for water supply projects in Kolkata, India

H. Guha<sup>1</sup>, B. Thakur<sup>2\*</sup>, T.S. Konar<sup>3</sup>, P.P. Biswas<sup>4</sup>

<sup>1</sup>Practicing Consultant, Kolkata, INDIA

<sup>2\*</sup>Department of Civil Engineering, Meghnad Saha Institute of Technology, Kolkata, INDIA

<sup>3</sup>Department of Civil Engineering, Camellia School of Engineering and Technology Kolkata, INDIA

<sup>4</sup>Department of Construction Engineering, Jadavpur University, Kolkata, INDIA

\*Corresponding Author: e-mail: b2981975@yahoo.co.in, Tel +91-9830241217, Fax.+91-33-24431032

---

### Abstract

Mega cities in developing countries like Kolkata in India are often dependent on international agencies for financing its civic infrastructures like water supply network. These funding agencies would require cost benefit analysis for appraisal of the projects. Such study would demand rapid cost estimation as well as the optimization of the project operations. Traditional methods namely multiple regressions and mathematical programming that are generally used for these studies seem to be relatively inefficient for cost benefit analysis of water supply projects. On the other hand, the performance of nature inspired techniques has been found to provide encouraging results mainly because of its ability to manage nonlinear, non-convex and stochastic dataset. Gene expression programming has been used in lieu of multiple regressions for developing the capacity-cost curves. Genetic algorithm has been employed in lieu of common mathematical programming for improving the optimization process. Applications of evolutionary algorithms have allowed investigating the issues to an extent that was not possible with the traditional methods. The key issues and difficulties for financing water supply projects to the southern fringe of Kolkata have been identified with the help of evolutionary algorithm even without the detail cost benefit analysis as prescribed by the funding agencies.

*Keywords:* Gene expression programming, Genetic algorithm, Water supply network, Developing country.

---

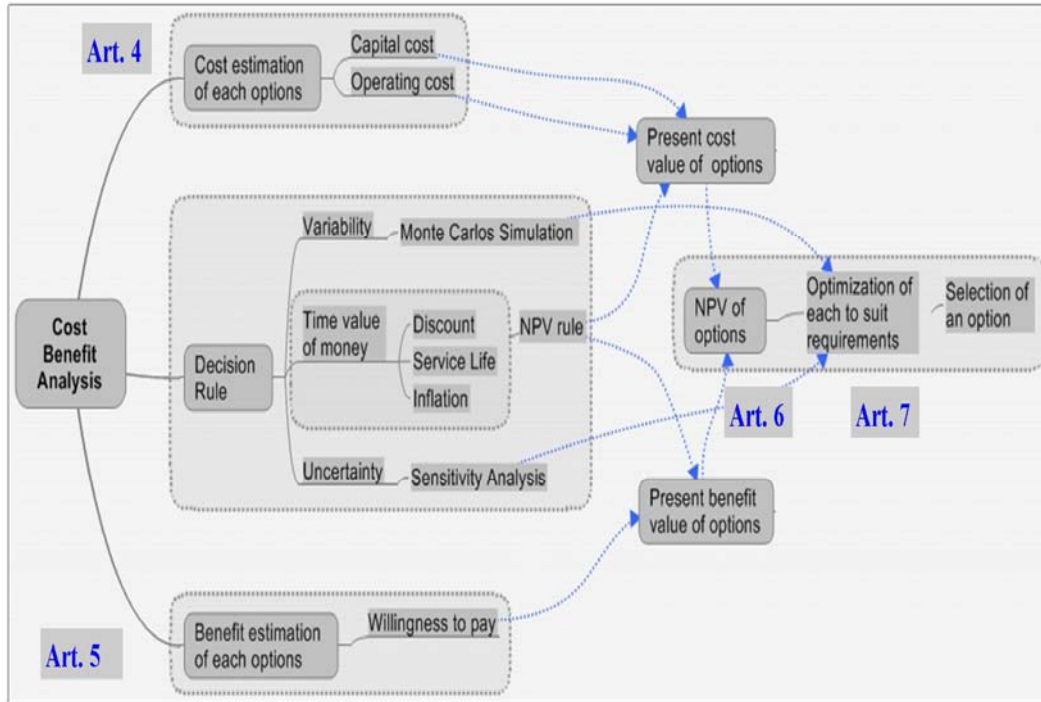
### 1. Introduction

Population in Indian cities like Kolkata is increasing rapidly. It has increased from 11.02 million in 1991 to 13 million in 2001 indicating about 1.8% annual growth rate (Bannerjee and Das, 2006). High rate of urbanization has generated a tremendous pressure on civic facilities like water supply. The first water works in Kolkata was established in 1909 and water services are provided free of cost since then. In addition, majority of the population is too poor to pay any fees at all. Water charges in Kolkata have been included in the property tax (Ruet 2002). The residential property tax in Kolkata has been found to be iniquitous (Nath 1987). Roy et al. (2003) reported that Kolkata Municipal Corporation grants more than 75% subsidy to its water supply costs. Politicians in Kolkata are now unwilling to change a century of old practice of free water supply. In such scenario, internal resources are scarce for financing water supply projects. Water supply projects have been taken up only sporadically in ad hoc basis to cater for immediate problems. Lately, overseas agencies like World Bank, Asian Development Bank etc are providing soft loans for infrastructure developments in Kolkata. These banks stipulate cost benefit analysis (CBA) for water supply projects. CBA is a quantitative method that strives to estimate the cost and benefits of water supply projects in monetary terms. Detailed estimation of costs and benefits for a future project is an expensive and time consuming process. In practice, it is nearly impossible to calculate detailed estimate of each of the options of a CBA for want of resources. Statistical methods like multiple regression based cost curves are often used for estimating costs. In western countries like UK, standard cost curves for water works components have been proposed (Ford and Warford 1969) for many years that are yet to be developed in India. The relationships among the factors resulting in the monetary values of costs or benefits of a water supply system are nonlinear and complex. The correlations obtained from regression based cost curves have not been found to be good enough for a CBA in Kolkata. The results were improved after application of bio-inspired technique. Similar observations have been also found for

optimizing the water supply planning. Results obtained from linear programming have been found to be improved when nature inspired methods like genetic algorithm is employed.

## 2. Methodology

2.1 CBA: The flow diagram of a typical CBA that is used for selection, planning and approval for international funding of a water supply project is presented in Figure 1.



**Figure 1.** Schematic CBA flow diagram

It has three major steps namely cost estimation, benefit estimation and decision rules. Since the cost and benefits occur at different time periods the time value of money principle is introduced in the decision rules (Hanley and Splash 2003). The inputs and formulation for the costs and benefits are not precise in most of the cases. Risk studies with the help of Monte Carlo's simulation are usually prescribed in standardized CBA guidelines for exploring the variability of results or risk of the analysis (TCB, Canada 2007), (EU Jaspers 2008). The uncertainty is a situation when the levels of risks are not known. Sensitivity analysis is often used to investigate the uncertainty in a CBA (Rocquigny, Devictor and Tarantola 2008). The cost of water supply systems are generally categorized to two types namely capital cost and operating cost. For CBA, the popular approach is to estimate the cost by means of an existing cost curve or relationship (Mays 2008). Since the costs occur throughout the life span of the facility present value rules as presented in Equation 1 are applied.

$$C_p = \frac{\sum_{t=1}^L \frac{C_t(1+r)^t}{(1+i)^t}}{\sum_{t=1}^L Q_t} + \frac{\sum_{t=1}^L \frac{O_t(1+r)^t}{(1+i)^t}}{\sum_{t=1}^L Q_t} \quad (1)$$

Where,  $C_p$  = Present value of cost in INR/Kl,  $i$  = Discount rate at year  $t$ ,  $C_t$  = Acquisition costs in Indian rupees (INR) incurred at year  $t$ ,  $O_t$  = Sustenance costs in INR incurred at year  $t$ ,  $Q_t$  = Quantity of water in Kilo liter (Kl) generated at year  $t$ ,  $L$  = Expected life of the facilities,  $i$  = Discount rate at year  $t$  and  $r$  = inflation rate at year  $t$ . The benefit of a proposed water supply project could be hard to define in monetary units. However, willingness to pay (WTP) method is often used in CBA of water supply projects for estimating the benefits (Bockstael and McConnell). WTP method is a direct survey among the beneficiaries about the quantum of money they are willing to pay for water supply service. The benefits of a project also occur in different time periods. The

principle of time value of money as presented in Equation 2 is therefore used to compress the benefit to the present value. The difference between the cost and benefit is the net present value (NPV) and is presented in Equation 3.

$$WTP_p = \frac{\sum_{t=1}^{75} \frac{WTP_0(1+r)^t}{(1+i)^t}}{t} \quad (2)$$

$$NPV = WTP_p - C_p \quad (3)$$

Where,  $WTP_p$  = Present value of WTP for each Kl of water,  $t$  = Expected life of the water treatment facility for which CBA is conducted. The NPV of each option provide the idea about its financial suitability. The options are optimized for attaining the objectives in the proposed situation. The risks of the options are estimated generally with the help of Monte Carlos simulation. Sensitivity analysis is normally used for exploring the uncertainty. The final selection of the project is made based on the NPV, optimized performance, risks, uncertainty etc.

**2.2 Criticisms:** The water supply CBA has faced criticisms since early days. For example, Hanke and Walker (1974) criticized the CBA of a US federal government project executed in 1967 and pointed out a number of anomalies. Mathematical programming for cost allocation in water supply CBA is also not new. Young et al. (1980) published detailed treatment of linear programming for water resources cost allocation among Swedish municipalities. Efforts have been made over time to standardize the CBA. For example, Office of Management and Budget (OMB) in US has issued guidelines for CBA and risk assessment (Kopp et al., 1997). They found that the willingness to pay is the most successful method for valuation. European Union started TECHNEAU, an integrated project funded by the European Commission (Baffoe-Bonnie et al., 2006). One of the mandates was to develop and demonstrate CBA including practices for risk management. They recommended economic valuation, decision criteria and risk management in CBA. In addition, they emphasized on sensitivity analysis to evaluate the potential effects of key uncertainties in the decision making process. Similar to US administration European Union published guides for CBA (Florio, 2006). It gives advises for steps like collection of territorial data, option analysis, financial analysis, economic analysis, evaluation criteria, risk and sensitivity analysis etc. Many of the criticisms in CBA are directed towards the accuracy of the estimation and allocation methods (Kennedy 1981), (Pearce 1998). Bio-inspired techniques could improve such accuracy to a considerable extent.

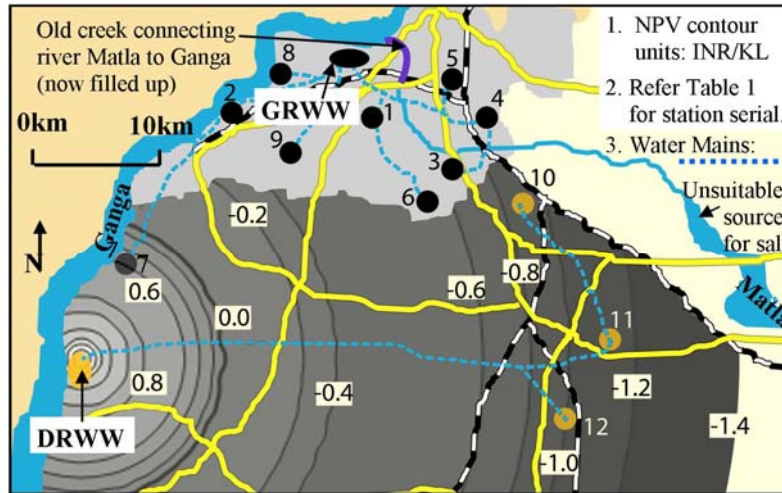
**2.3 Bio-inspired programs:** Evolutionary algorithms are based on processes similar to those that occur during biological evolution. The optimization problem in nature normally differs from the engineering problems. In nature, all species are subject to simultaneous evolution, and to concurrent changes in the environment. By contrast, in engineering problems, the desired goal is normally fixed and specified in advance. Nevertheless, in view of the ingenuity of the designs found in nature, there is sufficient reason for developing optimization algorithms based on natural evolution. While classical optimizations are generally deterministic and is better suited to convex type problems bio-inspired optimizations are stochastic and is also efficient for non-convex problems (Wahde 2008). The issues arising out of the present study seem to be more suited to bio-inspired methods. Two techniques namely gene expression programming (GEP) and genetic algorithm (GA) have been employed in the present study. The results appear to be better than that of the classical methods.

### 3. Study Area

The water supply network consists of trunk pipe lines from the treatment plant to booster stations. These facilities have been financed by the international agencies in Kolkata. The trunk lines and booster stations are the subject matter of the present study. The main water treatment plants and small diameter pipes from booster stations to individual households belong to different financing categories and have not been included in the present study. The CBA for booster pumping stations in southern water districts of Kolkata has been analyzed. The map of the area is presented in Figure-2. The urban areas are supplied from Garden Reach Water Works (GRWW) and the peri-urban / rural areas are served by Dakshin Roypur Water works (DRWW). Intermediate booster stations as shown in Table 1 and marked in Figure 2 according to serial numbers in Table 1 are used for optimizing the supply. It has been hypothesized that the tributary serving population, land rent of the booster station facilities and distance from the water treatment plants should be the main factors for ascertaining costs. The salient features of the water service networks have been presented in Table 1 (Dutta-Roy et al 2010/1&2).

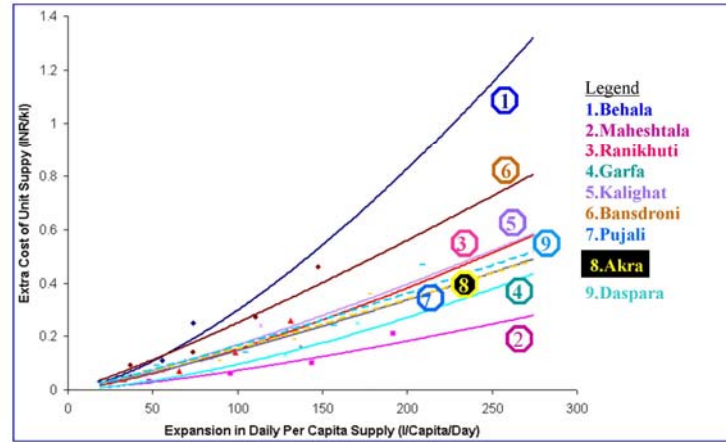
**Table 1.** Water pumping stations in Kolkata and southern fringes

Supply	Sl	Station Name	Tributary pop. (000)	Distance (KM)	Ln Rent (Mil/yr)	Tank (MG)	Supply (MGD)
GRWW Existing		GRWW Direct	579.6			NA	60.0
	1	Behala	550.9	7.00	6.57	4.5	9.0
	2	Mahestala	236.1	5.00	1.82	5.0	10.0
	3	Ranikuthi	241.3	12.00	4.38	3.5	7.0
	4	Garfa	185.9	14.00	1.84	3.5	7.0
	5	Kalighat	319.7	10.00	6.08	4.0	8.0
	6	Bansdroni	122.7	11.00	1.04	2.0	4.0
GRWW Planned	7	Pujali	34.7	15.00	0.41	0.7	2.1
	8	Akra	230.0	5.00	5.77	3.0	6.0
DRWW Existing	9	Das para	130.0	12.00	1.03	2.0	4.0
		DRWW Direct	757.8			NA	21.0
	10	Kamalgazi	341.8	57.92	0.60	1.0	3.0
	11	Begampur	419.7	42.48	0.34	1.0	3.0
	12	Krishnamohanpur	474.5	41.15	0.24	1.0	3.0

**Figure 2.** Study area map

#### 4. Cost estimation

**4.1 Estimation methods:** The traditional capacity-cost curves as presented in Figure-3 are developed for nine numbers booster pumping stations under GRWW using standard unit rate estimating method. Such parametric method or unit-rate estimate requires complete technical design, drawing and specifications for each booster station for preparation of bill of materials and there by the cost of the project. Preparation of these details is expensive and time consuming. In practice, it is not feasible to estimate such details for each option in a CBA. In order to develop an alternative rapid cost estimating routine two prediction techniques viz. standard multiple regression model and GEP model have been investigated using the existing data of the nine numbers booster stations developed by Dutta-Roy et al. (2010/1). It has been assumed that daily supply, land rent and treatment plant distance should be the major factors contributing towards the unit costs of water supply and are therefore considered as the input data in the above stated prediction techniques (Dutta-Roy et al. 2010/1). The models are described in subsequent sections.



**Figure 3.** Capacity-cost curves

**4.2 Regression model:** In order to develop an empirical relationship among these factors a multivariate regression (MR) model has been attempted as the first trial. The capacity-cost relationship of the nine numbers existing booster stations in the GRWW network was developed in parametric method by Dutta-Roy et al (2010/1). In Figure 3, the capacity-cost relationships for each of these nine stations with five supply quantities (existing quantity and 25%, 50%, 75% and 100% increment over the existing quantity) are presented. The data has been used to generate a best fit regression equation as shown in Equation-4.

$$C_p = 1.32 - 2.64 \times 10^{-6} (DS) + 0.11(Di) + 6.12 \times 10^{-8} (LR) \quad (4)$$

Where,  $C_p$ = Cost of supply in INR /Kl, DS= Daily Supply in KL, Di= Distance in KM from GRWW, LR= Land rent in INR/Year. The co-efficient of determination ( $R^2$ ) is found to be only 0.5383. It is felt that further improvement of the model is essential for its effective use.

**4.3 Gene expression programming:** The GEP is a computer program related to artificial intelligence that in principle mimics the natural genetic algorithm as it uses populations of individuals, selects these according to fitness, and introduces genetic variation using one or more genetic operators (Mitchell 1996). Dawkins (1995) noted that the fundamentals of a life system would require a self-copying replicator like a chromosome in which there is a hereditary variation. Secondly, it must survive by casual effects on something else which is called a phenotype. A basic example of a replicator/phenotype system is DNA/protein system which is the foundation of life in earth. Ferreira (2001) proposed a similar adoptive algorithm for data analysis. The idea was later expanded in a text book (Ferreira 2006) and a software system christened GeneXproTools (Gepsoft 2010) has been developed. The software has been utilized for the present analysis in investigating the cost prediction model. The supply costs are predicted for nine numbers of existing booster pumping stations under GRWW for five supply quantities as before. The independent variables are the same as that of the traditional MR model namely daily supply, distance from GRWW and land rent. GeneXProTools (2010) software has been used for analysis.

In GEP model the highest unit is chromosome, which consists of genes. A gene is considered to have a head and a tail. The length of head is prefixed based on which the tail length is determined by Equation (5) as follows:

$$t = (n - 1)h + 1 \quad (5)$$

Where,  $t$ = Tail Length,  $h$ = Head Length,  $n$ = Number of Independent Variables. Every unit length of head consists of either a dependent variable or an operator, whereas a tail consists of dependent variables only. The length of a gene is the summation of head length and tail length, and length of chromosome is the total of all the gene lengths. In the current study the GEP model is considered to have 50 such chromosomes, each having 4 genes with head length 8. In course of simulation (referred as change in generation), every chromosome alters its configuration independently and comes up with an equation. The equation, producing the closest result based on some specified fitness function, is considered as the desired equation. The fitness function used in the present case is root relative square error (RRSE) and the number of generations of simulation is 1500. The expression for supply cost is found to be as expressed in Equation (6).

$$C_p = \ln \left\{ 1.84 + Di + \sin \left( \frac{2.22 + LR}{Di} \right) \right\} + \frac{0.85DS \times \ln(DS)}{DS + LR} + \frac{1}{3} \cos(0.774353LR) + \cos\{\sin(Di)\} - 0.794 \quad (6)$$

Where,  $C_p$ = Cost of supply in INR /kl, DS= Daily Supply in KL,  $D_i$ = Distance in KM from source plant, LR= Land rent in INR/Year. (All angles under trigonometric functions are in radians). The graphical nature of the expression as obtained from the GEP model is presented in Figure-4. The model predicted values seems to have followed the actual that is the target values satisfactorily.

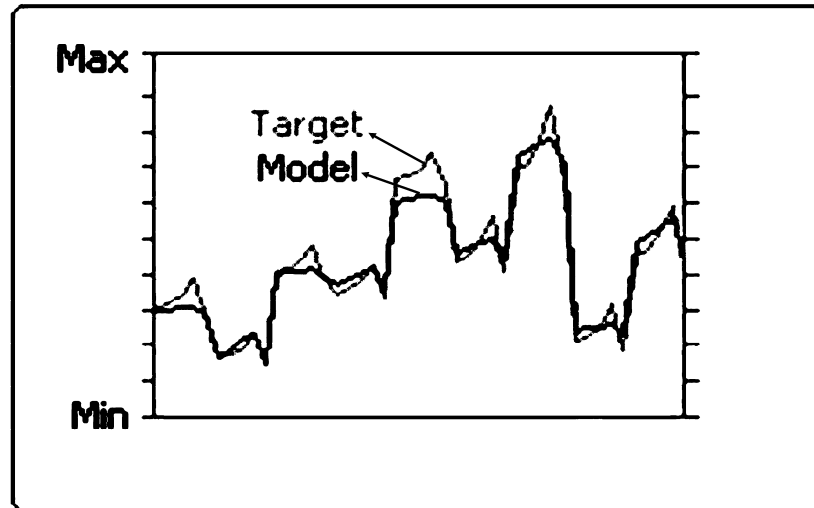


Figure 4. Comparison of actual (target) and GE model predicted values

The prediction performance of the GEP model is compared with that of the traditional MR model. The best fit straight lines between predicted costs and actual costs (the parametric estimates of costs for all the booster pumping stations) are drawn for both of the MR and GEP models as shown in Figure-5. For MR model the points are much more scattered as compared to the GEP model, indicating better correlation in the later case. The co-efficient of determination ( $R^2$ ) is also observed to be improved from 0.5383 in case of MR model to 0.9513 for GEP model showing better prediction potential of the GEP model.

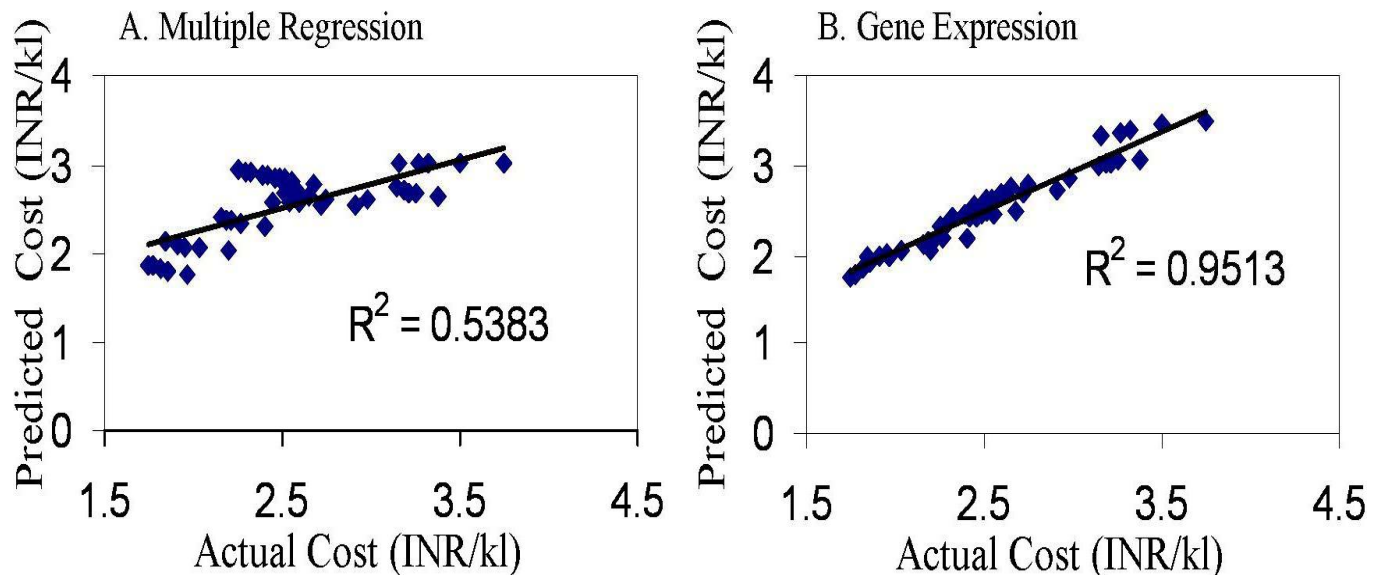


Figure 5. Comparison of the MR and GEP models

**4.4 Cost estimate for DRWW network:** The GEP has been found to be appropriate for estimating costs. The input values namely tributary population, supply quantity and distance from the plant for the three numbers booster stations in DRWW network have been collected from field survey and is presented in Table 2. The cost estimate for booster pumping stations in the DRWW network is estimated from Equation 6 and is presented in Table 2.

**Table 2.** Estimated costing for DRWW booster stations

Booster Stations	Sl	Estimated Population	Increased Quantity (KL/Day)	DRWW distance (KM)	Land rents (INR/Yr)	Cost (Cp) in (INR/Kl)	NPV (INR/Kl)
Kamalgazi	10	419,718	62957.7	42.48	341,032	4.65	-0.4460
Begampur	11	474,514	71177.1	41.15	243,594	6.46	-0.5847
Krishnamohanpur	12	341,803	51270.45	57.92	608,986	4.51	-2.3521

**5. Benefit Estimation**

The benefit of water supply projects is often measured by WTP method. Several WTP studies conducted in Kolkata for water supply projects have been presented in Table 3.

**Table 3.** Water supply project WTP in Kolkata

Sl	Reference	Mean WTP (INR/Kl)
1	World bank (2001)	3.63
2	Roy et al (2003)	9.37
3	Mazumdar and Gupta (2006)	3.18
4	Guha (2007)	11.70
5	Dutta-Roy et al (2010/1)	4.83
6	Dutta-Roy et al (2010/2)	3.30

A review of Table 3 would indicate considerable variations of WTP in different studies. For example, the WTP estimated by World Bank is more or less the same that of Dutta-Roy et al (2010/2) despite the fact that it was conducted a decade later and is subjected to higher inflation. The Government of India (RBI 2010) publishes whole sale price index which is a measure of inflation. Based on RBI (2010) bulletins, the inflation index has increased by about 32 percent from 2001 to 2010. Apparently, the inflation has no impact on the WTP between the said two studies. Such anomaly is caused by the different survey samples in each study. The year 2001 study was conducted in well settled boroughs of Kolkata where average per capita income was higher than that of 2010 study which was conducted in the peri-urban areas where the poorer sections of the population live. The results of the later study (Dutta-Roy et al., 2010/2) served as the basis of the present CBA analysis. The WTP survey was carried out among 201 sample households in the study area. Data about other variables such as family size and age, educational level, income level, quality and quantity of water used and awareness were also collected and correlated with the WTP. The WTP is found to be positively correlated to income.

It has been assumed that the average working class consumers should be reflected in the present CBA study. The WTP estimate (Dutta-Roy et al., 2010/2) is required to be adjusted for working class. There is no recent official data on spatial income distribution in Kolkata. The Labour Bureau (2010/1) published a 2001 sample survey result based on 3024 working class families. The per capita income was found to be INR1329.65 per month. The Labour Bureau (2010/2) published the consumer price index which is a measure of inflation for the consumers. It has been used to project the per capita income from 2001 to 2009. The said price index is found to be 157 and the revised per capita average income is estimated as INR2087. As already stated the WTP is positively correlated to income and the WTP survey estimated the following relationship between them as presented in Equation 7 (Dutta-Roy et al., 2010/2).

$$WTP_m = 0.0211PC\_Incm + 6.1137 \tag{7}$$

Where,  $WTP_m$ = Monthly per Family WTP (INR/Family/Month) and  $PC\_Incm$ = Monthly per family income (INR/Family/Month). The observed mean WTP of INR 3.30 (Dutta-Roy 2010/2) has been adjusted for average working class consumers having an average monthly income of INR2087 with the help of Equation 7 and is estimated as INR2.41 per KL which has been used for optimization of water supply network.

**6. NPV for DRWW Network**

The relationships obtained for costs and benefits have been used for estimating the net present value. The 75 years simulated NPV with actual present values of the variables for three numbers booster stations in the DRWW network have also been



presented in Table 2. The NPV of the stations are negative mainly because of its distance from the treatment plant. A contour map of NPV has been developed for showing the spatial variations of NPV within the DRWW supply areas. The typical size of a booster pumping station is assumed as suitable for fifty thousand people and the land rent is assumed as INR500,000 per year. The NPV for the DRWW network is estimated according to Equation 3. The cost ( $C_p$ ) has been estimated from equation 6 and WTP is assumed as INR2.41 per KL as estimated from equation 7. The contours of the NPV are presented in Figure 2.

## 7. Optimization of Supply Quantity

**7.1 Optimization objective:** The water supply quantity is optimized to suit the objectives of the CBA. The supplies to individual booster stations are maximized subject to three main constraints namely (i) total supply available from GRWW, (ii) per capita supply in each station should vary between 140-160 lpd for equity among consumers and (iii) the cost should not exceed the benefit i.e. the WTP. The optimization scheme is formulated as:

Maximize;

$$\sum Q_i \quad (8)$$

Subject to:

$$0 \leq \sum Q_i \leq Q_{\max} \quad (9)$$

$$\sum C_i \leq 0.5(WTP \times \sum P_i) \quad (10)$$

$$140 \leq \frac{Q_i \times 1000}{P_i} \leq 160 \quad (11)$$

Where,  $Q_i$  = Daily Water Supply from the  $i$ th Supply Facility (kl),  $Q_{\max}$  = Maximum Possible Daily Supply from GRWW (kl) ( $Q \times \eta$ ),  $Q$  = Daily Production from GRWW (120MG),  $\eta$  = Supply Efficiency after Incurring Losses (80%),  $C_i$  = Supply cost from  $i^{\text{th}}$  facility for supplying  $Q_i$  as per capacity cost curves in Figure-3.  $f(Q_i)$ , WTP = Consumer Willingness to Pay (INR/Person/Day) (0.7225±07867 INR/Person/Day),  $P_i$  = Tributary population of the  $i^{\text{th}}$  facility.

The optimization scheme has two complicating issues namely stochastic inputs and power nature of the capacity cost curves. It has been investigated with two optimization techniques namely conventional linear programming (LP) and bio-inspired genetic algorithm (GA). The results are compared in the following.

**7.2 Linear programming:** For the conventional LP method, Risk Optimizer software (DTS 2010) has been used for optimization. It is designed for managing stochastic input for the LP model. In the optimization model, the WTP value is inserted as a normal distribution with mean and standard deviation. The mean, maximum and minimum values of the costs are obtained from the regression equations of the capacity cost curves presented in Figure 3 and inserted as triangular distributions in the model. The Monte Carlo sampling type was chosen. The stopping condition for the optimization was set when the best results converge and does not change by more than 0.01% for last 100 iterations.

The Risk optimization module continued for about 35 minutes in a Pentium PC but could not reach to a converging result when it was manually abandoned. Out of the 340 completed simulations, only 54 runs were found to be valid satisfying all the constraints of the optimization problem. The model used standard linear programming technique. It seems that the linear programming method did not perform satisfactorily due to its inability to manage the power nature of the capacity cost curves.

**7.3 Genetic algorithm:** GA mimics Darwinian principles of natural selection by creating an environment where hundreds of possible solutions to a problem can compete with one another, and only the “fittest” survive. Just as in biological evolution, each solution can pass along its good “genes” through “offspring” solutions so that the entire population of solutions will continue to evolve better solutions.

To perform the optimization through GA the Evolver software (DTS 2010) was used in the present case. Evolver, which uses GA to solve optimization problems, finds global optimal solutions to linear, non-linear, table-based, or stochastic numerical problems and can solve problems with any combination of these qualities. It can also generate permutations of existing values, re-order the values, or group the values together in different ways to find the optimum solution. The non-linear power nature of the capacity cost curves was supposed to be effectively negotiated through GA using Evolver.

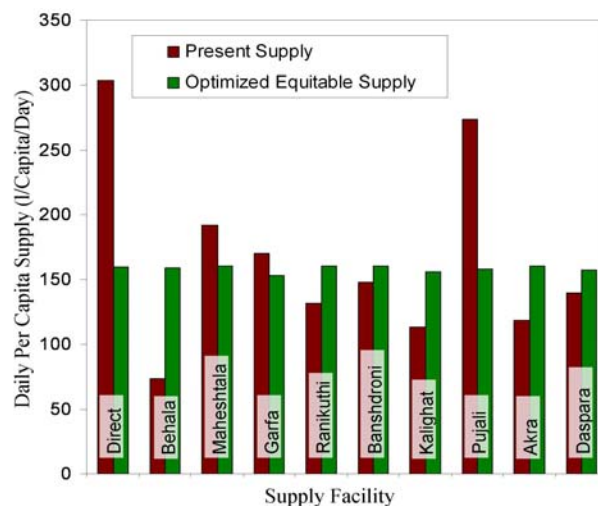
In the current study the basic recipe solving method was used. While creating a new solution, two parents are chosen from the current population of solutions. Solutions with high fitness scores are more likely to be chosen as parents. In Evolver, parents are



chosen with a rank-based mechanism. The basic recipe solving method performs crossover using a uniform crossover routine. Two groups are formed by randomly selecting items to be in one group or another. The uniform crossover method is considered better at preserving schema, and can generate any schema from the two parents. It also performs mutation by looking at each variable individually. A random number between 0 and 1 is generated for each of the variables in the organism, and if a variable gets a number that is less than or equal to the mutation rate, then that variable is mutated. The amount and nature of the mutation is automatically determined by a proprietary algorithm of Evolver. Mutating a variable involves replacing it with a randomly generated value (within its valid min-max range). Evolver uses a rank-ordered rather than generational replacement method and the worst-performing organisms are always replaced with the new organism that is created by selection, crossover, and mutation, regardless of its fitness “score”. While dealing with the Hard constraints in an optimization problem, if a new offspring violates some externally imposed constraints, Evolver backtracks towards one of the parents of the child, changing the child until it falls within the valid solution space.

In the optimization model, the inputs for WTP and capacity cost curves were used in the same manner as of the earlier case. The model was set to stop when the best results converge and do not change by more than 0.01% for the last 100 iterations like the previous case. The population size was 50 and the optimization random number seed (chosen randomly) was 10963935. Basic recipe solving method was used and the crossover rate and mutation rate was found to be 0.5 and 0.1 respectively.

The best results were obtained in the same Pentium PC within 33 seconds after running 781 simulations. Total optimization time was only 70 seconds within which a total 1964 trials were performed with 233 valid ones satisfying all optimization constraints before the optimization stopped satisfying the stopping condition. Thus, it is observed that the GA is capable to manage both the stochastic nature of the inputs and the power nature of the capacity cost curves efficiently. The optimum total supply was found to be 416676.34 kl/Day (91.98MGD).



**Figure 6.** Comparison of the optimized supply with present supply

## 8. Risk and Uncertainty

The risk is the variation due to stochastic nature of the inputs of the optimization problem. The optimization problem has been efficiently dealt with the Evolver software (DTS 2010). Monte Carlos simulation has been used to estimate the risk profile. The total cost of supply for the optimized supply condition is simulated using the @Risk simulation tool (DTS 2010). The simulation was run for 10000 times. The mean, standard deviation and the distribution of the simulated total cost is presented in Figure-7.

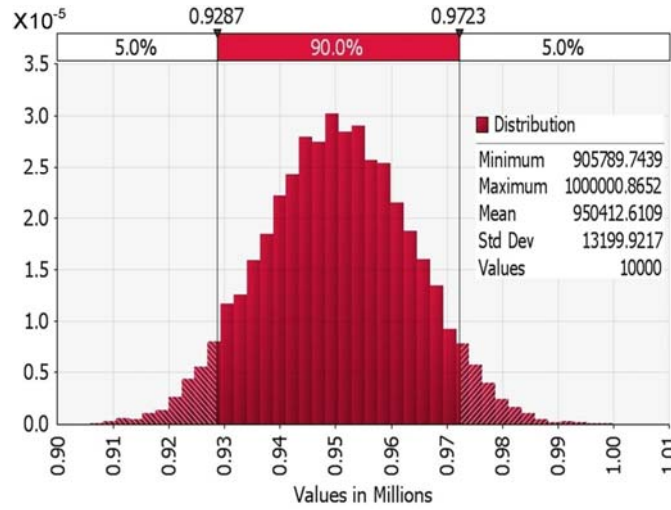


Figure 7. Distribution of the simulated total cost for maximum supply

In order to investigate the uncertainty sensitivity analysis has been performed. The sensitivity of the total cost on different stochastic input parameters are represented in the tornado diagrams presented in Figure-8 developed with the help of @Risk software (DTS 2010). The total cost of supply seems to be most sensitive on the cost of supply from Kalighat pumping station and least sensitive on the cost of supply from Pujali pumping station.

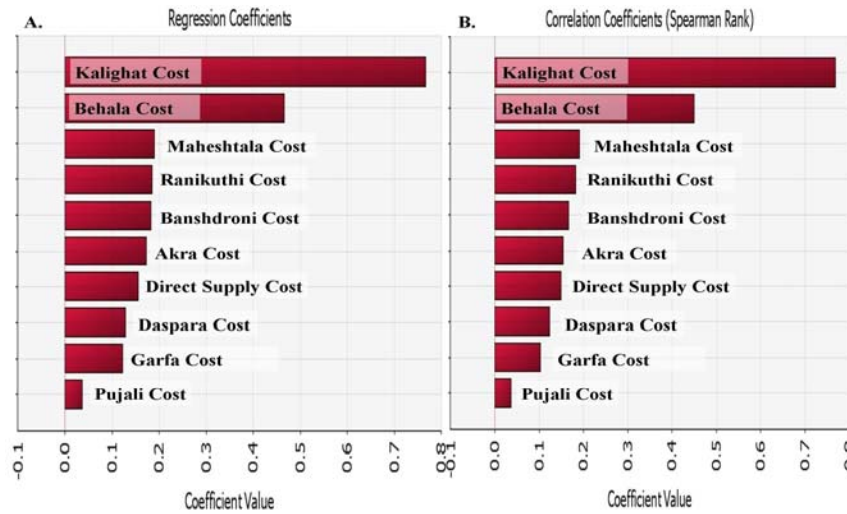


Figure 8. Tornado diagrams

9. Discussions

The present article explores the CBA of two water supply services namely GRWW and DRWW networks. The GRWW which is an urban water supply network in the southern side of Kolkata has enough data available for developing a capacity-cost relationship of the booster stations as shown in Figure-3. Numbers of WTP studies as presented in Table 3 have been also conducted within the GRWW network for estimating the benefits. Such cost-benefit data have allowed performing the CBA and estimating the NPV as detailed elsewhere (Dutta-Roy 2010/1) for the GRWW network. However, the water services within the network have remained sub-optimal as shown in Figure-6. The traditional optimizing tools like LP did not perform satisfactorily as shown in article 7.2. The optimization of the options which is a necessary part of a CBA as shown in Figure-1 could not be done with the traditional tools alone.

The evolutionary algorithm like GA was necessary for optimizing the water supply network. The optimized per capita supplies from different facilities are compared in Figure-6. It may be noted that direct supply from GRWW and Pujali was economized considerably after optimization of the GRWW network. The saved water was redirected to stations like Behala, Akra etc for

improving the objective of the network. Apparently, the stochastic input data and power nature of the cost curves have prevented the satisfactory performance of the traditional tools.

DRWW is for supplying water to the peri-urban areas of the southern boundaries of Kolkata. The traditional CBA method was found to be inadequate because of unavailability of capacity cost curves. The capacity cost curves based on parametric estimates were available only for the stations shown in Figure 3. The traditional regression analysis on the parametric data had the coefficient of determination ( $R^2$ ) of only 0.5383 as shown in article 4.2. Capacity-cost curves based from such regression analysis would not be adequate for prediction. The evolutionary algorithm like GEP has predicted the trend with a  $R^2$  of more than 0.9 as shown in article 4.3. The higher  $R^2$  value of GEP has permitted estimating the costs of booster stations in DRWW network relatively accurately. The predicted costs for DRWW booster stations were estimated in Table 2. The NPV for booster stations were also presented in Table 2.

The NPV for standardized booster stations were estimated in article 6. The iso-NPV line or the contours of NPV of standardized booster stations are presented in Figure 2 with respect to some typical values of other non spatial explanatory variables of cost estimator. In that case the contours in INR per Kiloliter as marked in Figure 2 would indicate that the NPV has become zero only at a distance of about 12.5 KM distance from the treatment plant. The three existing stations in DRWW network marked 10, 11 and 12 in Figure 2 are all located in the negative NPV zone. However the actual NPV (which are again negative) of each existing stations in DRWW network are presented in Table-2. It may be noted that international funding agency like Asian Development Bank would provide soft loans to water supply projects that result in positive NPV according to CBA executed as per their standard guidelines (ADB 1999). Any booster stations in the negative NPV zone in Figure 2 would have negative NPV in terms of banks' guidelines and would not qualify for grants from such funding agencies. In practical terms, the authority has to source funds elsewhere for such projects. The application of evolutionary algorithm has allowed the decision makers to learn about the future booster stations that would not qualify for grants from international agencies even before a detailed CBA.

## 10. Conclusions

International funding agencies like World Bank, Asian Development Bank (ADB) etc insist for CBA before funding a water supply project. Evolutionary algorithm applications have only made it possible to draw contours up to a distance of more than 50KM from the treatment plant as presented in Figure-2. It has also provided further insight about the issues related to long term water resources planning in Kolkata. Most of the locations within the DRWW command area are presently forced to draw ground water for potable supply. The area is infested with the arsenic in subsurface water. Many people in the area are affected from arsenic poisoning (Das, Samanta, Mandal et al 1996). Therefore, treated surface water is vitally important for settlements in these areas.

The map in Figure 2 would indicate that one should normally plan for water treatment plants in the river Matla in east of Kolkata for reducing distance from the source. A treatment plant in the river Matla in lieu of river Ganga would have reduced the supply cost to booster stations like 10, 11 and 12. However, the water from the river Matla as indicated in the map in Figure 2 is no longer fit for treatment as the result of an environmental disaster occurred more than two centuries ago. The river Matla was originally connected to the river Ganga. The water course ran through Kolkata as shown in a violet line in the map in Figure 2. The part of the creek shown in violate line in figure 2 was filled up with the development of the city. The process started about two hundred years ago. Kolkata is in estuarine zone and all rivers are subjected to tide. After disconnection from the perennial river Ganga the river Matla lost the inflow of river water and carried only sewage and runoff from Kolkata city and surroundings. In effect, it became an inlet of the sea and became salty as a result of tide. The river is no longer suitable potable supply.

The settlements in the DRWW area are therefore subjected to multiple problems. The ground water has arsenic. The eastern river Matla is salty. The settlements are relatively away from the river Ganga causing an expensive water supply system. The water supply projects in these areas shall not be approved for soft loans in CBA as stipulated by the international funding agencies because of negative NPV. The water supply to these areas shall therefore remain costly in future. The evolutionary algorithms have allowed noting about these issues before the growth of dense settlements.

## References

- (ADB) Asian Development Bank. 1999. Handbook for the Economic Analysis of Water Supply Projects. [http://www.adb.org/documents/handbooks/water supply projects/default.asp](http://www.adb.org/documents/handbooks/water%20supply%20projects/default.asp) (last access: 30 March 2010).
- Baffoe-Bonnie, B., Harle, T., Glenne, E., et al.: Framework for Operational Cost Benefit Analysis in Water Supply, TECHNEAU, contract number 018320, 2006.
- Bocksteal N. and McConnell K. E. 2007. Environmental and Resource Valuation with Revealed Preferences: A Theoretical Guide to Empirical Models. Springer, Dordrecht, Netherlands.
- Das D., Chatterjee A., Mandal B.K., Samanta G., Chakraborti D., Chanda B. 1996. Arsenic in groundwater in six districts of West Bengal, India. *Environmental Geochemistry and Health*, Vol. 18, pp. 5-15.
- Dawkins R. 1995. River Out Of Eden: A Darwinian View of Life, Basic Books, New York, USA.

- (DTS) Decision Tools Suite. 2010. Palisade software. Palisade Corporation, Ithaka, New York, USA, <http://www.palisade.com/> (last access: 13 Oct 2010).
- Dutta-Roy K., Thakur B., Konar T.S., Chakrabarty S.N., 2010/1. Rapid evaluation of water supply project feasibility in Kolkata, India, *Drinking Water Engineering Science*, Vol. 3, No. 1, pp. 29-42.
- Dutta-Roy K., Thakur B., Konar T.S., Chakrabarty S.N., 2010/2. Water supply project feasibilities in fringe areas of Kolkata, India, *Drinking Water Engineering Science*. Discussion, Vol. 3, No. 2, pp. 199–249.
- (EU) JASPERS 2008. Guidelines for Cost Benefit Analysis of Water and Wastewater Projects to be Supported by the Cohesion Fund and the European Regional Development Fund in 2007-2013, European Commission, Brussels, Belgium.
- Ferreira C. 2000. Gene Expression Programming: A New Adaptive Algorithm for Solving Problems. *Complex Systems*, Vol. 13, No. 2, pp. 87-129.
- Ferreira C., 2006. Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence (Studies in Computational Intelligence). Springer-Verlag, Berlin, Germany.
- Florio, M., Finzi, U., Genco, M., et al. 2006. Guide to cost-benefit analysis of investment projects, Evaluation Unit, DG Regional Policy, European Commission, Brussels, Belgium.
- Ford J.L. and Warford J. J. 1969. Cost Functions for the Water Industry. *The Journal of Industrial Economics*, Vol. 18, No 1, pp. 53-63.
- Gepsoft, GeneXproTools 4.0, <http://www.gepsoft.com/>
- Hanke, S. and Walker, R. 1974. Benefit Cost Analysis Reconsidered: An Evaluation of the Mid State Project. *Water Resources Research*, Vol. 10, No. 5, pp. 898–908.
- Hanley N. and Splash C. L. 2003 Cost Benefit Analysis and the Environment, Edward Elgar, Cheltenham, UK.
- Kennedy D. 1981 Cost-Benefit Analysis of Entitlement Problems: A Critique. *Stanford Law Review*, Vol. 33 pp: 387.
- Kopp, R. J., Krupnick, A. J., and Toman, M. 1997, Cost-Benefit Analysis and Regulatory Reform: An Assessment of the Science and the Art, Discussion Paper 97-19, Resources for the Future, Washington, DC, USA, 1997.
- Labour Bureau 2010/1. Report on Working Class Family Income & Expenditure Survey 1999-2000. Government of India <http://labourbureau.nic.in/CPI-MF-Kolkata.htm> (last access: 15 Oct 2010).
- Labour Bureau 2010/2. Previous Years/Months Index Numbers, All-India Average Consumer Price Index Numbers for Industrial Workers, Government of India <http://labourbureau.nic.in/indexes.htm> (Last access: 15 Oct 2010).
- Mays L. 2008. Urban Water Supply Handbook: History Strategic Planning, And Outsourcing, McGraw-Hill, New York, USA.
- Mitchell M., An Introduction to Genetic Algorithms 1996. MIT Press, Boston, USA.
- NeuroDimensions Inc.: Neural network software. 2010. <http://www.nd.com/> (last access: 10 October 2010).
- Pearce D., 1998 Cost-Benefit Analysis and Environmental Policy, *Oxford Review of Economic Policy*, Vol. 14, No. 4, pp. 84-100.
- Rocquigny E., Devictor N. and Tarantola S. 2008 Uncertainty in Industrial Practice: A guide to Quantitative Uncertainty Management. John Wiley & Sons Ltd, Chichester, UK.
- Roy, J., Chattopadhyay, S., Mukherjee, S., et al. 2003, An Economic Analysis of Demand for Water Quality: Case of Kolkata, *Economic and Political Weekly*, Vol. 39, No.2 (Jan. 10-16, 2004), pp. 186–192.
- Ruet, J. 2002. Water supply & sanitation as “urban commons” in Indian metropolis: how redefining the State/Municipalities relationships should combine global and local de facto ‘commoners’, Centre de Sciences Humaines, N. Delhi, India.
- Shyam Nath 1987. Is Residential Property Tax Equitable?: A Case Study of Calcutta. *Economic and Political Weekly*, Vol. 39, No. 34, pp. 3818-3826.
- (TBC) Treasury Board of Canada. 2007. Canadian Cost-Benefit Analysis Guide Regulatory Proposals, President of the treasury Board Canada, Catalogue No. BT58-5/2007, ISBN 978-0-662-05039-1
- Wahde M. 2008. Biologically Inspired Optimization Methods An Introduction. 1st ed. WIT Press. Southampton, UK.
- Young, H. P., Okada, N., and Hashimoto, T. Cost Allocation in Water Resources Development – A Case Study of Sweden, RR-32, International Institute For Applied Systems Analysis, Laxenburg, Austria, 1980.
- Zurada M., 1992. An Introduction to Artificial Neural Systems. PWS Publishing Company, Mumbai, India.

### Biographical notes

**Dr. Himadri Guha** is a graduate in Civil Engineering. He did his masters in Civil Engineering in Oklahoma and MBA in Chicago. He completed his PhD from Jadavpur University, India. His current research fields are in pollution control, engineering systems and information technology applications in civil engineering. He is presently self-employed as a consultant.

**Biswajit Thakur** is working as the Assistant Professor in the department of civil engineering of Meghnad Saha Institute of Technology in Kolkata, India for about five years. Before joining MSIT he worked as a senior research fellow in the civil engineering department of Jadavpur University, Kolkata. He received a bachelor degree (1999) and a master degree (2001) in civil engineering and is currently engaged in his doctoral research. His research areas are pollution control, sustainable development and collaboration in engineering design. He has more than twenty papers published in research journals.

**Tuhin Subhra Konar** is an Assistant Professor in the department of civil engineering of Camellia School of Engineering and Technology in Kolkata. He received a bachelor degree (2008) and a master degree (2010) in civil engineering. His research fields are pollution control, engineering systems and application of information technology in civil engineering. He has six articles published in journals and symposiums.

**Dr. Partha Pratim Biswas** is a graduate in Civil Engineering. He has two post graduate degrees in soil mechanics and engineering mechanics respectively. He completed his PhD in engineering from Jadavpur University, India. He is a Professor of the Department of Construction Engineering, Jadavpur University. He is also interested in management issues related to water supply and drainage in urban areas. He has numerous research articles.

Received October 2010

Accepted April 2011

Final acceptance in revised form April 2011