

Identifying Water Quality Sampling Stations in Al-Khobar Water Distribution System, Kingdom of Saudi Arabia

تعيين مواقع رصد نوعية المياه بشبكة التوزيع بمدينة الخبر، المملكة العربية السعودية
محمد الزهراني

Muhammad A Al-Zahrani

King Fahd University of Petroleum & Minerals - Department of Civil Engineering - P.O. Box 686, Dhahran 31261, Saudi Arabia

جامعة الملك فهد للبترول والمعادن - ص.ب 686 - الظهران 31261

المملكة العربية السعودية

Abstract: Water quality monitoring of the distribution system is important in controlling and ensuring that safe water will be delivered to the consumers' taps. The Safe Drinking Water Act requires that water quality in a water distribution network be sampled at locations which are representative of the whole network system. Presently, no guidelines exist describing how these sampling locations are identified. In this paper, a methodology based on Genetic Algorithm (GA) is introduced and applied to a real water distribution network. A prior step before applying the developed method is to simulate the flow within the distribution system. For this purpose, a hydraulic simulation model called EPANET was used. Once the simulation of the water distribution system is performed, the developed GA method is then applied to identify locations of water quality sampling stations for the water distribution network of Al-Khobar City in the Eastern Province of the Kingdom of Saudi Arabia. Final results indicate that the selected water quality monitoring stations identified by the developed model are mostly located in areas within the city where water consumption is very high. The results obtained are compared with the current practice of identifying water sample locations and show more confidence over the current practice.

Keywords: Al-Khobar water distribution network, Water quality monitoring stations, Optimization, Genetic Algorithm (GA), EPANET.

المستخلص: إن رصد التغير النوعي لمياه الشرب المنقولة بشبكات التوزيع يعتبر من الأمور الهامة التي يجب تبنيتها من الجهة المسؤولة عن توزيع المياه، وذلك للتأكد من صلاحية هذه المياه وجودتها ومطابقتها للمواصفات المحلية والعالمية قبل وصولها إلى المستهلك. وللتأكد من ذلك فإنه يجب أخذ عينات للمياه المنقولة عند بعض نقاط الشبكة. وفي هذه الورقة تم تطوير طريقة تتكون من خطوتين تساعد على تحديد المواقع المثلى لأخذ عينات مياه من الشبكة. الخطوة الأولى تعتمد على محاكاة حركة وتدفق مياه الشرب في شبكة التوزيع وذلك باستخدام برنامج (EPANET)، والخطوة الثانية يتم من خلالها استخدام طريقة الحساب الجيني (Genetic Algorithm) لتعيين المواقع المثلى لرصد نوعية المياه بشبكة التوزيع. وللتأكد من فاعلية الطريقة المقترحة فقد تم تطبيقها على شبكة توزيع المياه بمدينة الخبر بالمملكة العربية السعودية. وقد أظهرت النتائج بأن مواقع رصد نوعية المياه المقترحة باستخدام النموذج تقع في أجزاء المدينة الأكثر استهلاكاً للمياه مما يعطي النموذج المطور أفضلية على الطريقة المتبعة حالياً والتي تعتمد على أخذ عينات المياه بشكل عشوائي من نقاط مختلفة من الشبكة.

كلمات مدخلة: شبكة توزيع المياه بمدينة الخبر، محطات الرصد النوعي للمياه، المواقع المثلى، طريقة الحساب الجيني، هيدروليكا.

Introduction

Drinking water quality can deteriorate during distribution to the consumers. Many factors can cause deterioration of water quality in distribution systems. Some of the major causes are source water, treatment processes, operation of systems, transport and transformations, water distribution network condition, and storage.

Water quality of sources may differ (i.e. groundwater, surface water, desalination). Thus, water supplied from multiple sources could undergo water quality changes. For example, introducing a new supply may affect the quality of water in one or more of the following ways:

- The quality of the new supply may be corrosive or may support biological growth.
- The introduction of a new supply is often accompanied by changes in pressures, flow rates and flow direction. These hydraulic changes may disturb existing deposits.
- Blending the new supply with an old one may produce unexpected changes in the water quality.

Treatment processes can also contribute to the deterioration of drinking water quality. For example, while mixing desalinated water with groundwater, disinfection, fluoridation, and pH adjustment, metals are introduced into the finished water (Alam and Sadiq, 1989).

Water in the distribution system may undergo some deterioration before it reaches the consumers. Coliform organisms may enter the distribution system from booster pumps, from the packing used in joints of mains, or from washers on service taps. In addition, the water distribution may become contaminated from outside, for example, through cross-connections, back-syphonage, defective service reservoirs and water tanks, damaged or defective hydrants or washouts, or through inexpert repairs to domestic plumbing systems (WHO, 1971).

The occurrence of corrosion in the distribution system, especially networks with aged pipes, may increase the potential of adding toxic metal to the finished water before it reaches the consumers (Alam and Sadiq, 1989). In addition, leaks in the distribution network could increase the level of concentration of specific contaminants. For example, a study performed by Alabdula'aly (1997) indicates that one major factor that contributes to the high level of nitrate concentration in the water distribution is the existence of leaks, especially in areas where sewer connections do not exist.

Finally, recent studies indicate that storage tanks can affect water quality (Kennedy, *et al.* 1993). Storage tanks and reservoirs can contribute to substantial increase in water age depending on factors such as tank geometry, operation and location.

Importance of Water Quality Monitoring

Parameters related to water quality are measured for a number of reasons. At the treatment plants, the monitoring is undertaken to control the treatment processes, make them efficient and ensure delivery of safe water to the consumers. In the distribution system, monitoring is usually carried out for the chemical analysis of water and to provide supplementary information while assessing its quality. The primary reason for monitoring water quality is to verify whether the observed water quality is suitable for its intended uses. This necessary measure is performed to ensure that the distributed water meets the water quality standards.

The importance of water quality monitoring and assessment arises from a number of different perspectives that may combine the following goals in different ways (Bartram and Balance, 1996):

- To check whether the water meets the users' requirement of quality;
- To measure the impact of different parameters of aquatic environment on water quality;

- To control and regulate water quality;
- To map the variation of water quality from one location to another, in relation to uses and quality influence;
- To check the past trends of water quality influences and policies which led to the present status;
- To determine the adequacy of the distribution system as to whether or not it meets the multi-users' requirements;
- To give an insight into the future trends based on the present and past factors of water quality;
- To check the system performance against the disinfection process;
- To study the feasibility of expansion of the distribution system;
- To study how water quality differs geographically in relation to uses and quality influences.

Sampling Locations and Monitoring

In order to have a general picture of the water quality situation in a water distribution system, sampling locations need to be selected to monitor the water quality parameters. Spatial representation and convenience are the two major factors in selecting the sampling locations (Al-Zahrani and Moied, 2001; Lee, 1990; Lee and Deininger, 1992).

Once the appropriate sampling locations are located, then regular monitoring of water quality at these locations, in addition to monitoring of source quality, is needed. Monitoring should include sufficient parameters to indicate all quality concerns and should be conducted at appropriate locations throughout the source of supply. The monitoring program should include protocols for frequency of sampling and methodology of analysis and should be designed to establish the baseline data indicating both short-term and long-term trends. Such monitoring can serve as a trigger mechanism to detect the occurrence of water contamination problems at their earliest stages (Pontius, 1990).

Identifying Water Quality Monitoring Stations

The United States Environmental Protection Agency (USEPA) requires all drinking water authorities to monitor the water quality in their distribution system to ensure that the water reaching the consumers will be safe to drink and safe for other potable uses. The regulations ("Safe", 1974)

prescribe the sampling frequency and water quality parameters to be monitored. The sampling stations are supposed to represent the whole networking system in terms of quality parameters.

The current practice of water sampling is based on taking water samples from locations that are easy to reach. A survey conducted in the USA showed how water authorities select their sampling stations. Many water authorities choose their monitoring stations by convenience and/or accessibility of the stations. The frequent locations used for collecting water samples include: fire hydrants, storage tanks, pumping stations, commercial buildings, public buildings and private residences (Lee, 1990). The Al-Khobar water authority adopts a similar practice; they do not have any scientific guidelines on how to identify the sampling locations. Thus, a qualitative sampling scheme and/or guidelines need to be developed that would help the water authorities in identifying the proper locations for these sampling stations.

Literature Review

Recent methodologies have been developed to help in locating sampling stations (monitoring stations) based on scientific methods. Some of these recent methodologies are discussed below.

Lee and Deininger (1992) developed a methodology such that when a water sample is taken at a tap and its quality is analyzed, one would know the quality at the closest node of the distribution network. Associated with each node, there is a known demand. For example, if node *i* is sampled with a demand of 10 units out of 100 units, the monitoring station is a representative of 10% of the total demand. If any other station node takes care of more than 10% of the total demand, then this new station will be more representative of the network. So, if it is proposed to build only one monitoring station in the network, then the station covering the maximum flow will be the location of the optimum station. Similarly, if two stations are to be built, then two stations with a maximum coverage will be the desired locations in the network and so on.

The term coverage or covered is used commonly for monitoring stations at a certain node to specify the water quality and the demand at that node. In order to cover the entire demand, practically every node would have to be sampled. However, it is assumed that the water quality at other nodes can be inferred by taking into account the flow patterns in the network. The term coverage or covered denotes that it is possible to infer the water quality at this node based on measurements at some other nodes.

Kumar, *et al.* (1997) and Lee and Deininger (1992) used the percentage proportion contribution as the governing criteria for coverage of an upstream node by monitoring the water quality at the downstream node. The chosen percentage proportion represents the adequate criteria for a particular node being covered by monitoring the water quality of the other node. They assumed that if 60% or more of the flow to a particular monitoring station has passed through an upstream node, then the water quality at this upstream node can be inferred from the quality of the monitored node. Kumar, *et al.* (1997) hydraulically analyzed the network based on the 60% criteria for developing a coverage matrix and came up with four monitoring stations covering 14,000m³/d of water out of 21,300m³/d. The number and locations of the monitoring stations recommended by Kumar, *et al.* (1997) and Lee and Deininger (1992) were found to be identical for the same hypothetical network.

The methodology developed by Lee and Deininger (1992) was suitable for an internal or gradual deterioration of water quality. It was less appropriate for rapid determination of water quality due to an external source of contamination. For example, in the case of a very long pipe with many consumption outlets, Lee's methodology resulted in a single monitoring station at the downstream pipe end. This is true regardless of the location of the intrusion or its time to detection (the time that elapsed until the pollutant has been detected). Kessler, *et al.* (1998) developed a method which identified the best selection of monitoring stations. It allowed detection of an accidental intrusion of contamination within a given level of service (maximum volume of contaminated water consumed prior to detection). Kessler, *et al.* (1998) used 1000ft³ of water as a level of service for detecting the accidental intrusions on municipal water networks. EPANET (Rossman, 1994) was used for the hydraulic simulations.

Bush and Uber (1998) have proposed matrices for water distribution system sampling design and also proposed a methodology to select measurements and locations which are sensitive to changes in parameter values, while ignoring the effects of sensitivity coefficients. Although their model suggested an approximate sampling design method, it gained advantages over statistical design methods which lack physical basis. Bush and Uber (1998) likewise proposed three sensitivity-based methods for ranking potential sampling locations in a water distribution network, namely Max-Sum design, Max-Min design, and Weighted-Sum

design. These methods ranked spatial measurement locations and measurement types according to the measure of work for parameter estimation. They did not consider measurement times and frequency as sampling variables. Max-Sum design ranked the measurements according to summed normalized sensitivity coefficients over the parameters. Max-Min and Weighted-Sum designs were both iterative ranking methods. Max-Min estimated all the parameters accurately while Weighted-Sum ranked the measurements by their contribution to the estimation of individual parameters.

Meier and Barkdoll (2000) proposed a methodology based on Genetic Algorithm to find an optimal location in water distribution systems that causes the water to flow at non-negligible velocities through as much of the pipe network as possible, and to design a program for data collection for network model calibration. Non-negligible flow velocity was used as a criterion for selection, and the Roulette Wheel method was used with uniform crossover and 2% rate mutation.

Al-Zahrani and Moied (2001, 2003) developed a methodology based on Genetic Algorithm (GA) to identify water quality monitoring stations in a water distribution network. Water quality monitoring stations were located based on the flow coverage only.

This paper illustrates the application of the model developed by Al-Zahrani and Moied (2003) to help in identifying the proper locations of the water quality sampling locations in Al-Khobar water distribution network.

Genetic Algorithm (GA)

Genetic Algorithm (GA) is based on the mechanics of natural selection and survival of the fittest (Goldberg, 1989). The basic procedure of GA can be described as follows:

1. Write a fitness function (also called an objective function) that can be used in comparing potential solutions to one another.
2. Randomly generate a set of coded strings representing potential solutions to variables that point to one location in the solution domain.
3. From the initial population, the fittest strings (as measured by their objective function values) are selected to pass their "genetic information" to the next generation. This operation is called "selection", which resembles the survival of the fittest in natural systems. There are many different schemes for selecting survivors; however, all of them share the common goal that more fit members replace the less fitter ones in

the population to advance the searching operation. After selection, the population is on average, more fit than it was before selection.

4. Selection is then accompanied by an operation called "crossover" which creates from the survivors, new strings that contain distinguished properties of the survivors from which they are created. In some cases the new strings will have lower fitness, in other cases they will have higher fitness, and in a certain percentage of cases the children will resemble their parents and thus, have the same fitness values as their parents.
5. Since crossover simply recombines the existing strings into new combinations, successive generations will carry the characteristics contained in the previous populations. It is possible that some desirable strings were not included in the initial (randomly generated) population or might have been lost because individuals possessing those desirable qualities became unfit and disappeared from the population. An operation called "mutation" is therefore used to occasionally alter a string (chromosome) in order to recover desirable qualities or to create new qualities in the strings (chromosomes).

If the GA has been designed well, the population will converge to an optimal solution to the problem. Figure 1 shows the necessary steps involved in applying the developed methodology based on Genetic Algorithm (GA). For detailed application of the developed GA model on a hypothetical network, refer to Al-Zahrani and Moied (2003). (See Fig. 1)

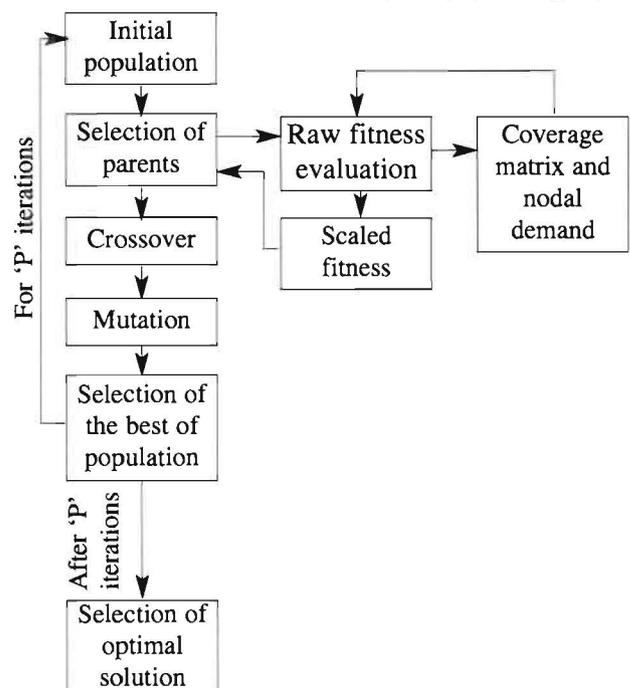


Fig. 1. Genetic Algorithm Flowchart

Table 1 summarizes the locations of the optimal monitoring stations based on the desired number of stations. Figure 3 shows the relation between the maximum scaled fitness values versus the population for different targeted numbers of monitoring stations to be selected or optimized.

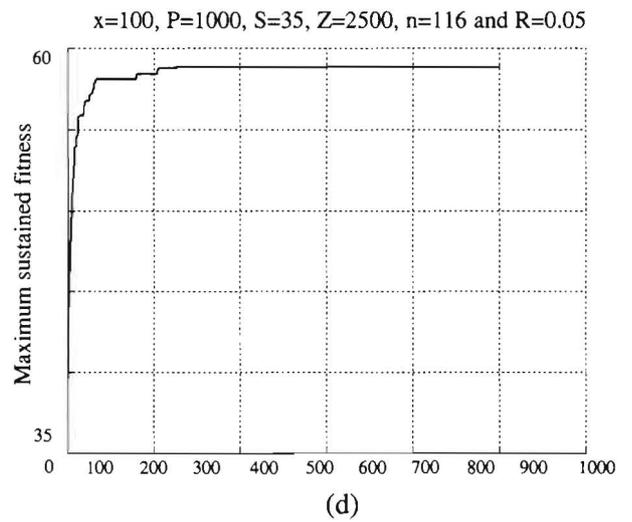
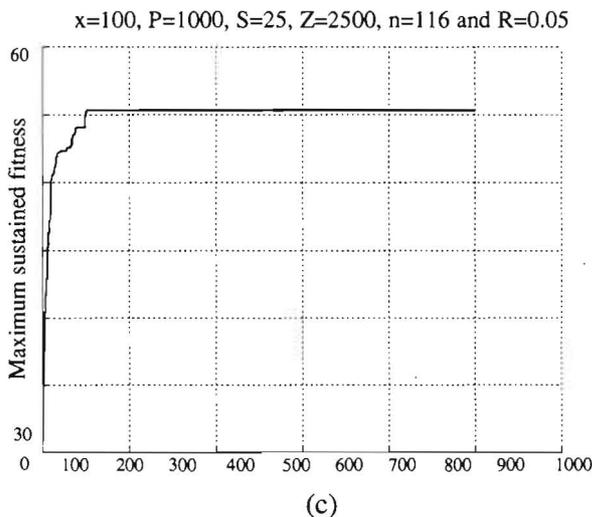
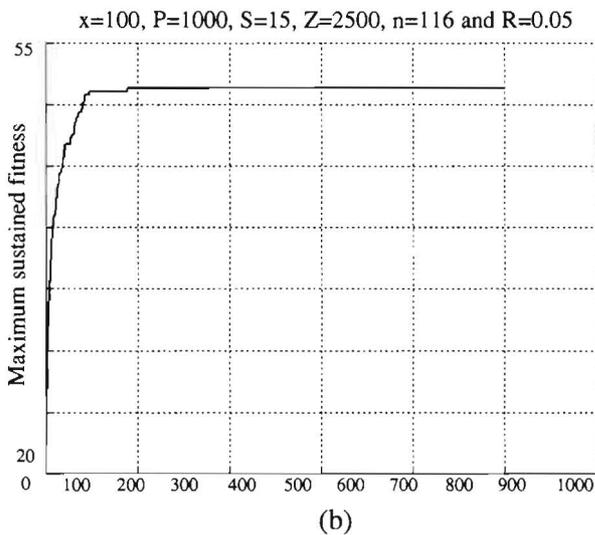
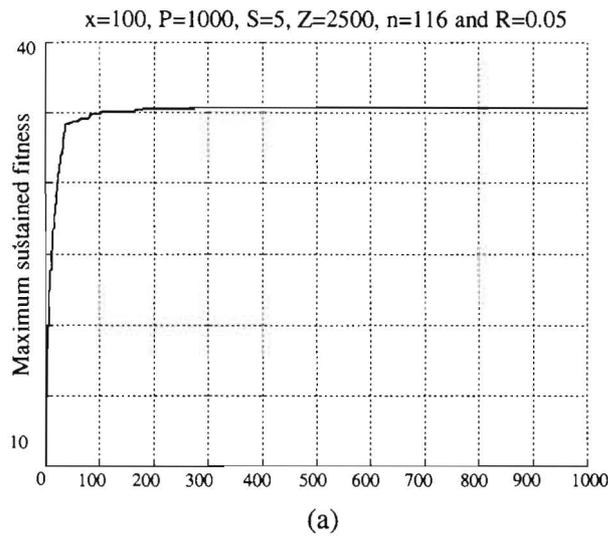


Fig. 3. Maximum fitness value vs. population for (a) 5, (b) 15, (c) 25 and (d) 35 monitoring stations

The results of the optimization model show that most of the simulations converge within 100 to 400 iterations. Steady-state conditions can be observed when exceeding 400 iterations, indicating that an optimal solution was achieved. According to the results, the monitoring stations selected were close to locations where high demand exists.

There is a trade-off between the number of monitoring stations and the demand coverage of the system. In order to investigate this trade-off, curves were plotted between “Demand Coverage” and “Benefit” versus the number of monitoring stations. The benefit represents the information that will be gained when increasing the number of monitoring stations. The demand coverage obtained from the water coverage matrices and the benefit was calculated by taking the difference of the demand coverage between two successive sets of monitoring stations. Figure 4 shows the relation between the demand coverage and benefit versus the number of monitoring stations. The curve shows that the demand coverage of the system increases considerably in the beginning when the number of monitoring stations considered for optimization is low and then it becomes almost steady as the number of monitoring stations increases.

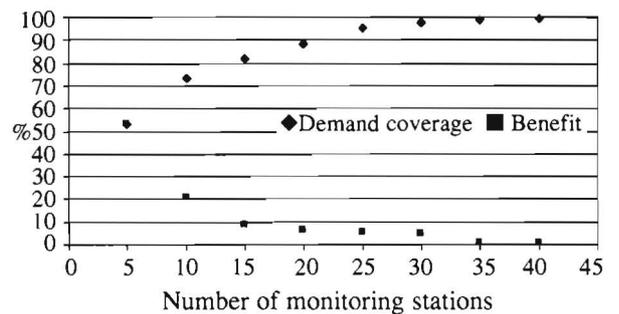


Fig. 4. Demand coverage and benefit vs. number of monitoring stations

This behavior indicates that, as the number of monitoring stations increases, little information or benefit will be gained. This conclusion can be verified from the curves, which show that the benefit of monitoring stations begins to dissipate as the number of monitoring stations increases. Based on this behavior, it is therefore recommended to monitor only 25 locations within the central portion of Al-Khobar water distribution system. It appears from the figure that, when considering 25 monitoring stations, about 94% of the demand will be covered. This value of coverage is quite fair when monitoring water quality in a water distribution system.

By further increasing the number of monitoring stations, almost no extra benefit can be achieved. In addition, increasing the number of monitoring stations will not affect the reliability of the monitoring program.

Figure 5 shows the proposed locations of the 25 monitoring stations identified by the model and the existing water quality monitoring stations identified by the Al-Khobar water authority. The suggested locations of the monitoring stations indicate logical behavior with respect to the investigated area. For five monitoring stations, the locations selected were located in the southeast of the city where high demand exists. As the number of monitoring stations increases, the selected locations start to shift towards the center and then towards the southwest of the city, which is the second high demand area. However, as the number of monitoring stations increases, the locations become more uniformly distributed. At about 25 monitoring stations, the locations identified cover all major parts of the city. On the other hand, the locations identified by the water authority were selected based on the geography of the area, i.e. sampling locations are distributed uniformly over the area, and the selection of the monitoring stations does not depend on the quantity of the flow as in the case of those stations identified by the model. Thus, the model results are more reliable than the current practice of identifying water quality samples.

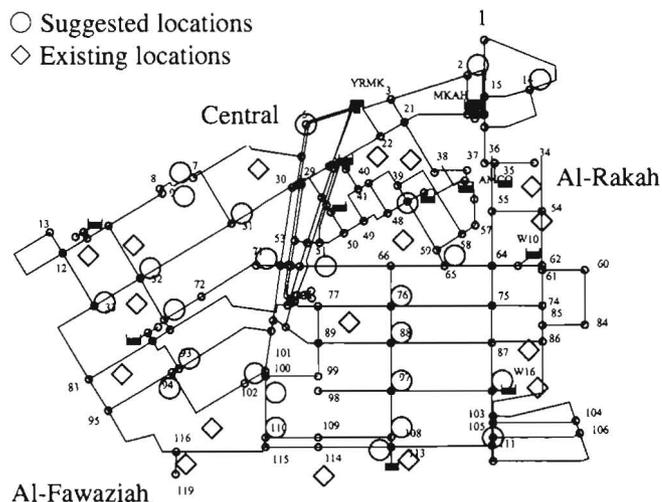


Fig. 5. Locations of suggested and existing water quality monitoring stations within Al-Khobar Water Distribution Network

Conclusion and Recommendations

A methodology based on GA was developed and applied to identify water quality monitoring stations in Al-Khobar water distribution network. The results of this study are compared with the current practice of water quality sampling and show more confidence over the current practice. This type of research can contribute significantly in assuring safe and better water quality to be delivered to the consumers through the water distribution network by identifying proper locations for the monitoring stations over the entire water distribution network.

The model developed in this study is quite flexible and can manage to accommodate additional features associated with water quality monitoring, which are usually governed by any monitoring program. Based on this study, some recommendations are also made as follows:

- A kingdom-wide policy needs to be issued for regular monitoring of water quality in water distribution systems at their optimum locations, in all the metropolitan cities. This will help the operators and water authorities reduce the cost of monitoring and reduce health risks associated with drinking water transported by the distribution network.
- Extension of the current model is suggested in the future by considering multiple objectives of monitoring. These objectives must include chlorine residuals, water age, and multiple sources of water supply.

Acknowledgements

The author expresses his thanks to King Abdulaziz City for Science and Technology (KACST) for its financial support and to King Fahd University of Petroleum and Minerals (KFUPM) for providing the necessary help and research facility to conduct the current study.

References

- Alabdula'aly, A** (1997) Nitrate concentrations in Riyadh, Saudi Arabia drinking water supplies. *Environmental Monitoring and Assessment* **47**: 315-324.
- Alam, I and Sadiq M** (1989) Metal contamination of drinking water from corrosion of distribution pipes. *Environmental Pollution* **57**: 167-178.
- Al-Zahrani, MA and Moied, K** (2001) Locating optimum water quality monitoring stations in water distribution system. *World Water & Environmental Resources Congress*, Orlando, Florida, May 20-24, 2001.
- Al-Zahrani, MA and Moied, K** (2003) Optimizing water quality monitoring stations using genetic algorithms", *The Arabian J for Science and Engineering (AJSE)* **28 (1B)**: 57-75.
- Bartram, J and Balance, R** (1996) *Water quality monitoring: A practical guide to the design and implementation of freshwater quality studies and monitoring programmes.*
- Bush, CA and Uber, JG** (1998) Sampling design methods for water distribution model calibration. *J of Water Resources Planning and Management* **124 (6)**: 334-344.
- Goldberg, DE** (1989). *Genetic Algorithms in search optimization and machine learning.* Addison-Wesley, Reading, Mass.
- Kennedy, MS, Moegling, S, Sarikelle, S and Suravallo, K** (1993) Assessing the effects of storage tank design on water quality. *J of AWWA*, 78-88.
- Kessler, A, Ostfeld, A and Sineri, G** (1998) Detecting accidental contaminations in municipal water networks. *J of Water Resources Planning and Management* **124 (4)**: 192-198.
- Kumar, A, Kansal, ML and Arora, G** (1997) Identification of monitoring stations in water distribution system. *J of Environmental Engineering* **123(8)**: 746-752.
- Lee, B** (1990) *Locating monitoring stations in water distribution networks.* Ph.D. Dissertation, Environmental Health Sciences, University of Michigan.
- Lee, BH and Deininger, RA** (1992) Optimal locations of monitoring stations in water distribution system. *Journal of Environmental Engineering* **118 (1)**: 4-16.
- Meier, RW and Barkdoll, BD** (2000) Sampling design for network model calibration using Genetic Algorithm. *J of Water Resources Planning and Management* **126 (4)**: 245-250.
- Pontius, FW** (1990) *Water Quality and Treatment*, 4th ed. McGraw-Hill, New York.
- Rossman, LA** (1994) *EPANET - User's Manual.* United States Environmental Protection Agency (USEPA), Cincinnati, Ohio.
- Safe Drinking Water Act of 1974** (1974) Federal Registrar, U.S. Environmental Protection Agency, Washington, D.C.
- World Health Organization (WHO)** (1971). *International Standards for Drinking Water.* 3rd ed. Geneva.

Received, 11th May 2004; in revised form; 26th July 2004