Journal of Environmental Sciences Vol. 14, No. 3, pp. 367—374, 2002 ISSN 1001-0742 CN11—2629/X

Article ID: 1001-0742(2002)03-0367-08 **CLC number**: X823; X51 **Document code**: A

A genetic algorithm based stochastic programming model for air quality management

MA Xiao-ming, ZHANG Fan

(Center for Environmental Sciences, Peking University, Beijing 100871, China. E-mail: zhangfan@water.pku.edu.cn)

Abstract: This paper presents a model that can aid planners in defining the total allowable pollutant discharge in the planning region, accounting for the dynamic and stochastic character of meteorological conditions. This is accomplished by integrating Monte Carlo simulation and using genetic algorithm to solve the model. The model is demonstrated by using a realistic air urban-scale SO_2 control problem in the Yuxi City of China. To evaluate effectiveness of the model, results of the approach are shown to compare with those of the linear deterministic procedures. This paper also provides a valuable insight into how air quality targets should be made when the air pollutant will not threat the residents' health. Finally, a discussion of the areas for further research are briefly delineated.

Keywords: stochastic model; genetic algorithms; air quality management; optimization

Introduction

Starting with early work of Teller(Teller, 1968) and Kohn (Kohn, 1971) beginning in the seventies and continuing to the present, a host of mathematical and decision support techniques have been developed and employed to aid in forming air quality planning optimization models. According to the treatment of parameters in these models, they can be classified into two major categories. The first category examines deterministic models in which parameters are assumed to be known with certainty in advance. The second one discusses the stochastic approaches which explicitly account for the uncertainties in the parameters. These parameters are random variables involved in the estimation of a pollutant's impact, such as the source emission rates, background deposition rates, cost-removal functions. But what is most emphasized is the great variability in meteorological conditions.

The purpose of this paper is to provide a natural setting to view the development from deterministic models to stochastic models and the shortcomings in these earlier modeling efforts, and then develop and apply a modeling approach aimed at overcoming these limitations of present air planning. In this paper, a chance-constrained programming(CCP) based regional air quality model that allows random, statistically dependent transfer coefficients of any distribution is developed and illustrated. Both long-term and short-term air quality standards have been integrated into the requirement constraints with Monte Carlo simulation. Dramatic increases in computational power over the past decade and the evolution of genetic algorithms(GAs) actualize the solution.

The remainder of the paper is organized as follows. First, background information and a literature review are given below. Then the model is described, and the solution algorithm is outlined. The approach is illustrated using a realistic air quality management problem. Finally, results and conclusions from the application of the model are discussed.

1 Background and literature review

1.1 Deterministic models

Deterministic model presuppose that the coefficients are known in advance with certainty and has a constant value. The following is the simplest possible deterministic air quality optimization model. It is to minimize the abatement quantity while subject to the non-violation of the air quality standard.

$$\mathrm{Min}: f = \sum_{j=1}^{n} E_j R_j , \tag{1}$$

$$\text{s.t}, \sum_{j=1}^{n} E_j(1 - R_j)t_{ij} + B_i \leqslant D_i, \quad i = 1,2,\cdots,m \tag{2}$$
$$0 \leqslant R_j \leqslant 1 \qquad j = 1,2,\cdots,n$$

where t_{ij} represent transfer coefficient for source j and receptor i, m is the number of receptors, n is the number of sources, R_j is the removal rate at source j, E_j represents the emissions from source j, B_i represents the background concentration at receptor i and D_i is the ambient limit at receptor i.

In order to describe the complex stochastic phenomena in a deterministic way, two methods have been used. One takes the most undesirable air condition as the constant constraint value. The second has all the possible meteorological conditions weighted by their probabilities of occurrence and then gets a frequency-weighted meteorological input.

While being simple in solution, the shortcomings of deterministic models are obvious. A worst-case analysis tends to yield strategies that may be more costly than necessary. Since expensive controls may be required during days when the same amount of pollutants have little impact on air quality because of the favorable air condition. When using the second method, complex meteorological scenarios are often extrapolated from limited measurements and a small number of sources. The resulting strategies lead either to violation of the annual air quality standard or to unnecessarily high costs for pollution abatement.

1.2 Stochastic models

In the late seventies and early eighties, researchers began to recognize shortcomings in deterministic models. This led to stochastic approaches that mainly deal with great variability in meteorological conditions. Transfer coefficients began to be treated as having random components which mainly depend on meteorological variables such as wind direction and speed. There are several solutions suggested for the stochastic program(Fronza, 1984; Ellis, 1985; Fuessle, 1987). Among these research efforts, chance-constrained programming is most developed and widely used.

A typical chance-constrained program based approach consists of transforming the stochastic program into the following chance-constrained program.

$$\text{Min}: f = \sum_{j=1}^{n} E_j R_j,$$
$$\text{s.t.Pr}\left\{ \sum_{j=1}^{n} E_j(1 - R_j)t_{ij} + B_i \leqslant D_i, \right\} \geqslant \alpha_i \quad i = 1,2,\cdots,m,$$
$$0 \leqslant R_j \leqslant 1; \qquad j = 1,2,\cdots,n.$$

This model requires that the probability of meeting ambient standard D_i is greater than or equals to a preassigned reliability level α_i. The transfer coefficients t_{ij} are functions of randomly distributed annual frequencies of meteorological conditions and are therefore themselves random variables.

The traditional method for solving CCP model is to formulate deterministic equivalents to the chance constraints and then apply an appropriate algorithm to the resulting deterministic optimization model. Initial work was illustrated by Charnes and Cooper(Charnes, 1959; 1961), later, Ellis(Ellis, 1985) and Guldmann (Guldmann, 1988). Such a method is also based on the worst possible distribution, therefore tending to yield constraints that may be considerably tighter than necessary too. What's more, assumption of zero-order decision rules and normal distributions are commonly used to obtain deterministic equivalents. When the normal distribution assumption cannot be satisfied, the deterministic equivalent becomes infeasible. Since random variables are sometimes non-normal and statistically dependent, this standard technique of CCP is limited in real problems.

In previous work the exogenously specified air quality standards are almost long-term. The thought of integrating short-term standards in a probabilistic programming framework(eg. 1-hours, 3-hours, 24-hours and monthly average) had been put up by Jean-Michel Guldmann (Guldmann, 1986). He provided two methods to actualize the solution. Real-time control and dynamic approach to compute the concentration of every short time. But with the statistical and computation difficulties, he did not actualize the solution.

1.3 Model objects

Most of the previous work placed economic efficiency as the foremost objective with preemptive priorities to guide the development of models and the choices of solution(Teller, 1968; Guldmann, 1986; Fronza, 1984; Loughlin, 2000; Ellis, 1985). "Health effects" associated with pollution were less considered. Kohn and Burlingame's LP model(Kohn, 1971) is the first one to explicitly include health effects. But it did not account for the variability of meteorology conditions and the correlation between observed concentration values and mean hospital stay is difficult to be precisely determined.

2 Model description

The model in this paper consists of a linear program and an air quality simulation. The linear program is chance-constrained. The objective is to determine the maximum annual aggregate emissions subject to achieving exogenously specified both long-term and short-term air quality standards. Gaussian air model was used as the modeling algorithm that connects the linear chance-constrained program and the simulation.

If emission sources are distributed continuously, the model can be formulated as follows:

$$\text{Max} \iint Q \mathrm{d}x \mathrm{d}y,$$

$$\text{s.t.} \begin{cases} E[\iint (a_{ij}x_j) + B_i] \leqslant S_i, \\ P[\iint (a_{ij}x_j) + B_i \leqslant S'_i] > \alpha_i, \end{cases}$$

$$j = 1,2,\cdots,n, i = 1,2,\cdots,m.$$

In real planning work, we usually select discrete emission sources as control points, so the above model can be written as:

$$\text{Max} Q = \sum_{j=1}^{n} x_j,$$

$$E\left[\sum_{j=1}^{n}(a_{ij}x_j) + B_i\right] \leqslant S_i, \tag{3}$$

$$s.t. \quad P\left[\sum_{j=1}^{n}(a_{ij}x_j) + B_i \leqslant S'_i\right] > \alpha_i, \tag{4}$$

$$j = 1,2,\cdots,n, i = 1,2,\cdots,m,$$

where, Q is the maximum allowable total annual emission in the control region(t/y), n is the number of emission sources in the inventory, j is the source index; m is the number of receptors (air quality checkpoints), i is the receptor index; x_j is the maximum allowable emission(t/y) at control site j; E is the expected annual ambient pollutant concentration in the control region; P represents probability, a_{ij} is the transfer coefficient in terms of $\mu g/m^3$ of pollutant contributed to receptor i from emission source j. a_{ij} is a random variable and is a function of the following random variables: ξ_{1i} is the wind speed; ξ_{2i} is the wind direction, ξ_{3i} is the total sky cover/low sky cover category, ξ_{4i} is the observation ordinal number. On the basis of these four random variables, we can obtain meteorological inputs to the Gaussian diffusion model to determine the value of a_{ij}. The four meteorological variables: ξ_1, ξ_2, ξ_3, ξ_4 were treated as discrete variables respectively and may follow any distribution. They may be got from the time sequence meteorological statistics by Monte Carlo simulation. B_i is the background pollutant concentration for receptor i. S_i is the maximum allowable average annual ambient pollutant concentration for receptor i in terms of $\mu g/m^3$. The value of S_i varies from low to high according to the different function of the region where receptor i is located. S_i' is the threshold guaranteeing no air pollution threat to human's health. α_1 is a preassigned constraint reliability level, which can also be defined as the meaning of risk——the probability of occurrence of a high popularity of disease among sensitive populations.

Eq.(3) and (4) contain air quality constraints assuring that annual and instant pollutant concentration regulations are satisfied respectively. The left-hand side of Eq.(3) estimates the expected emissions under different meteorological situations. Chance-constraint set of Eq.(4) requires that the allowable instant concentration S_i' must be satisfied at receptor i at least percent of the time.

The goal of ensuring public health weighs heavily in the model when judging potential strategies. In our research in Yuxi, it is discovered that air pollution damage varies not only with the length of exposure, namely the value of annual or a long period of average pollutant concentration, but related closely with the frequency of the occurring of pollutant concentrations that exceeds certain threshold. So in this model (Yuxi air quality planning report, 1997), we not only require the achievement of an average annual level of air quality to prevent chronic and long-term effects, we also integrate short-term standards as certain threshold to minimize the threat to human health in a probabilistic programming framework to confine the frequency of short-term acute effects.

This model emphasizes three aspects of considerations: the assimilative capacity of the air, air quality and it's implication on human health.

3 The algorithm

3.1 Simulate the probability distribution of meteorological variables

The four meteorological variables: ξ_1, ξ_2, ξ_3, ξ_4 were treated as discrete variables respectively and may follow any distribution.

Observation ordinal number is a parameter that has no relationship to the meteorological condition. For a long and stable observation, it can be formulated by uniform distribution, where is a random variable.

$$f(x) = \frac{1}{b-a}, \qquad a \leqslant x \leqslant b.$$

Wind direction, total sky cover and low sky cover are purely meteorological statistics, which can be processed as discrete variables in stochastic simulation.

According to the previous research work, wind speed can be fitted by the method of interpolation such as Log-normal distribution, which can be expressed as follows:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x}\exp\left[-\frac{(\ln x-\mu)^2}{2\sigma^2}\right], x > 0,$$

where μ and α are parameters, and σ is a random variable.

3.2 Generating random numbers from probability distribution functions

Supposing random variables have density matrix $\begin{bmatrix} x_0 x_1 x_2 \cdots \\ p_0 p_1 p_2 \cdots \end{bmatrix}$ and event $A_i = (\xi = x_i), i = 1, 2, \cdots$ satisfy the complete condition, then we can simulate by generating random from uniform distribution $U(0, 1)$. If $\sum_{i=0}^{k-1} p_i \leqslant \eta \leqslant \sum_{i=0}^{k} p_i$, then the random variable is x_k accordingly.

3.3 Computing the expectation function and probability function

According to the theory of stochastic simulation, the expectation of the left side of the Eq.(3) can be estimated by the follows:

$$\frac{1}{N}\sum_{i=1}^{N} C(x, y, \xi_i),$$

where ξ_i is the random vector of the 4 variables, $\xi_{1i}, \xi_{2i}, \xi_{3i}, \xi_{4i}$, which are random variables extracting from the distribution $\Phi(\xi_1), \Phi(\xi_2), \Phi(\xi_3), \Phi(\xi_4)$ separately, N is the number of samplings.

Similarly, generate N random vectors ξ_i according to the probability distribution of each variable. Supposing that there are N' random vectors satisfying Eq.(4), then the probability of the left side of Eq.(4) can be estimated by the limit theorem using the following formulation:

$$\theta = \frac{N'}{N}.$$

3.4 Genetic algorithm

GAs can be applied to obtain good solutions for many problems to which traditional optimization approaches have not proven successful. Several GA approaches have been developed and applied into the water management field (Scott, 1995; Wang, 1991; Liong, 1995; Brian, 1994). However, there are almost no such cases in air quality management.

GAs are a class of probabilistic procedures that search for good solutions to problems by emulating the "survival to the fittest" concept seen in nature. The principle idea of the GAs can be summarized as follows.

In a GA, a potential solution to a problem is most often represented as a vector of values or genes. In the context of this model, each gene may represent the allowable emission level at a controlling emission source in the study region. In GA the set of potential strategies are also called a population, generally consisting of about 50 to 200 strategies, which are generated at random or seeded with good solutions. The problem is subjected to several probabilistic operators that are analogous to natural selection, mating (including genetic combination) and mutation. In the selection step, pairs of strategies are selected for reproduction from the population in a manner such that fitter strategies are selected more frequently. Each pair of strategies may then undergo mating or crossover to form two new strategies. The new strategies are then ordered to create a new population. The selective and mating steps continue until the new population is the same size as the current population. The performance of each strategy in the population is characterized by a fitness value. After the new population has been generated, use Monte Carlo simulation to evaluate the fitness of a strategy. For example, the vector representing an air control strategy is first decoded to determine which emission level are allowable at each source. An air quality model is then run to determine the resulting air quality. The strategy is assigned a fitness that is a function of how effectively it meets the ambient target as well as other modeled objectives and constraints. After repeating the process for required number of generations, best strategies are sure to be found.

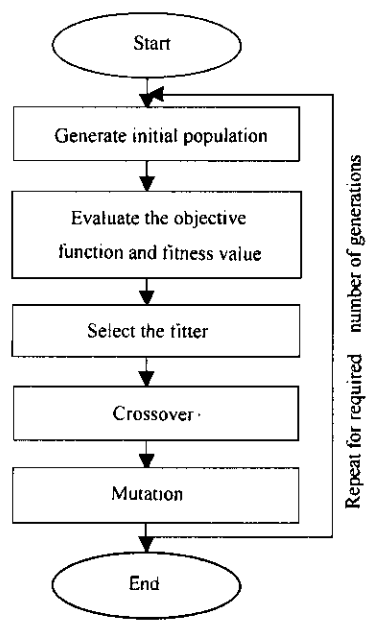

Fig.1 The flowchart of GA

The GA search process is depicted in Fig. 1.

Monte Carlo simulation is used to compute the value of expectation and probability in Eqs. (3) and (4) during selection and mutation.

4 Case study

The CCP model based on Monte Carlo simulation and GA algorithm has been applied to a case study involving SO_2 control for Yuxi City. Although there are regulations for several air pollutants, the total and subregional impacts of SO_2 from power plants are the focus of model application in this study.

4.1 Data and parameteres

The study region includes seven county in Yuxi, covering about 757.2 km^2. According to the different application purposes and natural enviroumental conditions, it can be divided into two control levels which are shown in Fig. 2. The first class area include Chunhe tourist area, Dongfeng dam water resourse conservation and ecological environmental protection zone, covering about 19 km^2.

The national ambient standard for annual average cocentration of SO_2 is 0.02 mg/m^3 for the first class control level and 0.06 mg/m^3 for the second class control level. The national ambient standard for 1h concentration of SO_2 is 0.15 mg/m^3 for the first class control level and 0.50 mg/m^3 for the second class control level.

In this study, we select seven aggregated emission sources and seven sensitive receptor locations as control sites, which are also indicated in Fig. 2. Receptors are mainly geometrical centers of each county. They can be used to reperesent the different control requirement of each functional area. Emission sources represent the average emission power of certain area. They are choosed according to different natural conditions and the density of pollutant emissions.

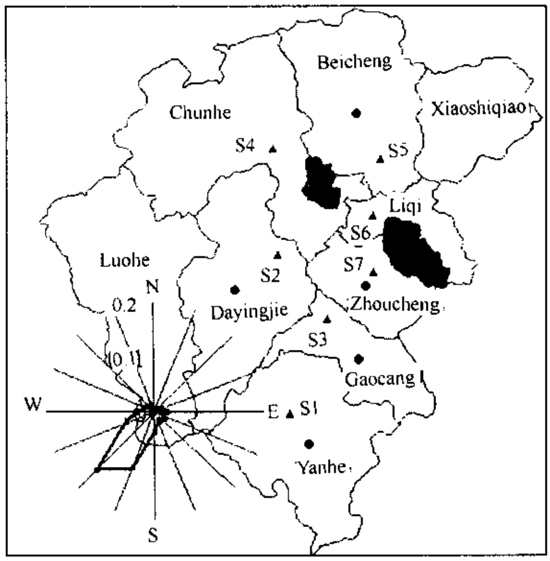

Fig.2 Boundary of study region, involving seven county in Yuxi, Yunnan. Triangles and circles on the map represent the emission sources and receptors respectively. The first class area is represented by dark-colored area. The air mass movement is from the northeast to the southwest

The meteorological data used in this study are drawn from a series of observations made at the local weather service station from 1994 to 1998, including wind speed, wind direction (specified according to 16 sectors), total sky cover and low sky cover categories. We state the occurrence probability of these meteorological parameters (Table 1 – 3).

Table 1 The wind speed

Wind speed, m/s	≤1.9	2 – 2.9	3 – 4.9	5 – 5.9	≥6
Frequency of occurence	0.715	0.13	0.12	0.018	0.017

Table 2 The wind direction

Wind direction	NNE	NE	ENE	E	ESE	SE	SSE	S	SSW
	0.004	0.009	0.01	0.022	0.01	0.02	0.018	0.04	0.111
Wind direction	SW	WSW	W	WNW	NW	NNW	N	C	
Freq. of occur.	0.154	0.052	0.036	0.014	0.013	0.005	0.004	0.478	

Table 3 Total sky cover/low sky cover categories

Total/low	≤4/≤4	5 – 7/≤4	≥8/≤4	≥5/5 – 7	≥8/≥8
Freq. of occur.	0.65	0.019	0.038	0.055	0.238

Accounting for the fact that polluting souces of Yuxi are mainly area sources, the physical stack height of each hypothetical source is taken as 15m. Backgrond concentration of SO_2 is 0.01 mg/m^3. The upper bound of searching area Q_0 is 50000 ton roughly set according to the present pollution level and controlling emission goal in Yuxi. Based on the results of trials using different

$P_{crossover}$ and P_{mute}, crossover frequency and mutation frequency are selected to be 0.6 and 0.2 respectively.

4.2 Model results

The CCP model was solved by VB programs. The resulting optimal emission and allocation between areas are presented in Table 4.

The total permitting annual emission of the study region is 22700 ton, slightly below that based upon deterministic LP model (1) - (2) which is 34000 ton. This demonstrates that regulation put aside by short-term air quality standard is stricter.

Table 4 Optimal emission and allocation (10 thousand ton)

Emission sources	S1	S2	S3	S4	S5	S6	S7
Maximum emission	0.78	2.85	3.06	4.12	4.23	3.87	3.7

The allowable maximum emission varies among areas, with an obvious increasing trend in diffusion from south to north. This matches with the fact that the local predominant wind pattern is S, SSW and SW, which verify the rationality of the CCP model to some extent.

5 Conclusions

There are also some extensions that can be suggested as areas for further research.

In this model the acceptable violation probability was set accoring to the national short-term air quality standard. However, in order to set more suitable and empirical value estimation, long-term times-series observation on receptor exposures to pollution and resulting damages is required.

The use of transfer coefficents discussed above are all in a linear programming contex. However, if the pollutant is nonlineraly reactive for which the source-receptor relationship is non-linear, then non-linear optimizing models are required.

In this model, we only considered the random variables related to meterology. However, the pollution generation and treatment such as sulfur content and heating value etc. all display stochastical characteristics. Another factor: emission rate, which always be taken as constant inputs to the Gaussian diffusion model, in fact also changes from time to time. These stochastical factors should all be accounted for and incorporated as random variables into the future model.

Time varable character of meterological conditions has been accounted for in planning models but not in real-time control. That is to say, planners cannot get information about how to dynamically allocate the total emissions between desiarable and undesirable weather conditions. For example, we know in Yuxi, emission in winter should be much lower as compared to that in summer. But what is exactly the variable allocation strategie and what is the emission ratios between monthes during a year, this can be realized by developing other dynamic real-time control air planning models.

Computational burden is a great concern, requiring totally 8 hours computing time in our study. Since in most GAs, every strategy in the population must be tested in each generation of the algorithm. For an air quality problem, this means that an air quality model must be run thousands or even tens of thousands of times during the search process.

References:

Brian J R, Wayland E, Ranjithan S, 1994. Using genetic algorithms to solve a multiple objective groundwater pollution containment problem [J]. Water Resources Research, 30(5):1589—1603.

Charnes A, W W Cooper, 1959. Chance-constrained programming[J]. Management Science, 6: 73—79.

Charnes A, W W Cooper, 1961. Deterministic equivalents for optimizing and satisfying under change constraints[J]. Journal of the American Statistical. Association, 57: 134—148.

Daniel H L, S Ranji R, John W B *et al.*, 2000. Application of genetic algorithms for the design of ozone control strategies[J]. Journal of the Air & Waste Management Association, 50(6): 1050—1063.

Daniel S E, D C Diakoulaki, C P Pappis, 1997. Operations research and environmental planning[J]. European Journal of Operational

Research, 102: 248—263.

Ellis J H, E A Mcbean, G J Farquhar, 1985. Chance-constrained/ stochastic linear programming model for acid rain abatement-I. Complete colinearly and noncolinaerity[J]. Atmosphere Environment, 19(6): 925—937.

Ellis J H, E A Mcbean, G J Farquhar, 1986. Chance-constrained /stochastic linear programming model for acid rain abatement-II. Limited colinerarity[J]. Atmosphere Environment, 20(3), 501—511.

Fronza G, P Melli, 1984. Assignment of emission abatement levels by stochastic programming[J]. Atmosphere Environment, 18(3): 531—535.

Goldberg D E, 1989. Genetic algorithms in search, optimization, and machine learing[M]. Reading, MA: Addison-Wesley.

Greenberg H J, 1995. Mathematical programming models for envrionmental quality control[J]. Operations Research, 43 (4):579—593.

Guldmann Jean-Michel, 1986. Interactions between weather stochasticity and the locations of pollution sources and receptors in air quality planning: a chance-constrained approach[J]. Geographical Analysis, 18(3): 199—214.

Guldmann Jean-Michel, 1988. Chance-constrained dynamic model of air quality management[J]. Journal of Environmental Engineering, 114 (5):1116—1133.

Hugh E, Michael L B, 1994. Critical loads and development of acid rain control options[J]. Journal of Environmental Engineering, 120(2): 273—290.

Iwamura K, Liu B, 1996. A genetic algorithm for chance constrained programming[J]. Journal of Information Optimization Sciences, 17(2): 40—47.

Kohn R E, 1971. Optimal air quality standards[J]. Econometrica. 39(6): 983—995.

Kohn R E, 1971. Application of linear programming to a controversy of air pollution control[J]. Management Science, 17: 609—621.

Kohn R E, Burlingame D E, 1971. Air quality control model combining data on morbidity and pollution abatement[J]. Decision Science, 2: 300—310.

Liong Shie-Yui, Weng Tat Chan, Jaya Shree Ram,1995. Peak-flow forecasting with genetic algorithm and SWMM[J]. Journal of Hydraulic Engineering, 121(8): 613—617.

Liu B A, Zhao R Q, 1998. Stochastic planning and fuzzy programming[M]. Beijing: Tsinghua University Press.

Michael B M, Edward S R, 1985. A linear programming model for acid rain policy analysis[J]. Journal of the Air Pollution Control Association, 35(11): 1137—1148.

Robert W F, M ASCE, Downey E B *et al*., 1987. Air quality planning: a general chance-constraint model[J]. Journal of Environmental Engineering, 113(1): 107—123.

Rockafellar R T, Roger J B, 1991. Scenarios and policy aggregation in optimization under uncertainty[J]. Mathematics of Operations Research, 16(1): 119—147.

Scott E C, J Wayland Eheart, S Ranjithan, 1995. Using genetic algorithms to solve a multiobjective groundwater monitoring problem[J]. Water Resources Research, 31(2): 399—409.

Teller A, 1968. The use of linear programming to estimate the cost of some alternative air pollution abatement policies[Z]. IBM: 345—353.

Tsunemi Watanabe, Hugu Ellis, 1993. Stochastic programming models for air quality management[J]. Computers Ops Res, 20(6): 651 — 663.

Tsunemi Watanabe, Hugh Ellis, 1993. Robustness in stochastic programming models[J]. Appl Math Modelling, 17: 547—554.

Wang Q J, 1991. The genetic algorithm and its application to calibrating conceptual rainfall-runoff models[J]. Water Resources Research, 27 (9): 2467—2471.

(Received for review May 30, 2001. Accepted June 19, 2001)