ScienceDirect

Atmospheric Environment
Volume 42, Issue 36, November 2008, Pages 8464-8469

A Bayesian hierarchical model for urban air quality prediction under uncertainty

Yong Liu [a], Huaicheng Guo [a], Guozhu Mao [b], Pingjian Yang [a]

Show more

Outline | Share | Cite

https://doi.org/10.1016/j.atmosenv.2008.08.018
Get rights and content

Abstract

Urban air quality is subject to the increasing pressure of urbanization, and, consequently, the potential impact of air quality changes must be addressed. A Bayesian hierarchical model was developed in this paper for urban air quality predication. Literature data on three pollutants and four external driving factors in Xiamen City, China, were studied. The air quality model structure and prior distributions of model parameters were determined by multivariate statistical methods, including correlation analysis, classification and regression trees (CART), hierarchical cluster analysis (CA), and discriminant analysis (DA). A multiple linear regression (MLR) equation was proposed to measure the relationship between pollutant concentrations and driving variables; and Bayesian hierarchical model was introduced for parameters estimation and uncertainty analysis. Model fit between the observed data and the modeled values was demonstrated, with mean and median values and two credible levels (2.5% and 97.5%). The average relative errors between the observed data and the mean values of SO_2, NO_x, and dust fall were 6.81%, 6.79%, and 3.52%, respectively.

Previous Next

Keywords

Bayesian hierarchical model; Markov Chain Monte Carlo (MCMC); Urban air quality; Multiple linear regression (MLR)

1. Introduction

Recent years have seen rapid urbanization worldwide, especially in developing countries. Urban air quality is under significant threat from increased populations, traffic, industrialization, and energy use (Mage et al., 1996, Zárate et al., 2007). Some serious issues, such as air pollution, acid rain, ecological deterioration, and direct public health and monetary loss, have emerged. Consequently, it is urgent that cities adopt integrated urban air quality management strategies to promptly respond to the stresses of pollution and to effectively manage emission sources. Prediction of the potential impact of air pollution is a fundamental basis of urban air quality management.

Simulation models have been widely used to estimate the potential impact of air pollution in urban environments; such models combine weather conditions and emission sources (Gidhagen et al., 2005). A number of methods have been developed for model parameter estimation and pollutant concentration prediction, such as statistical analysis (Huang, 1992), the artificial neural network (ANN) model (Dorling et al., 2003, Hooyberghs et al., 2005, Pérez-Roa et al., 2006), and the hybrid linear–nonlinear method (Chelani and Devotta, 2006). Huang (1992) proposed a stepwise cluster analysis method for predicting urban air quality based on the improvement of the automatic interaction detection algorithm. This method can successfully address continuous and discrete variables, as well as the nonlinear relations among the variables. The urban district of Xiamen, China, was used as a case study, and the model performed well. Hooyberghs et al. (2005) developed a neural network model to successfully forecast daily average PM_{10} concentrations in Belgium. Chelani and Devotta (2006) used a hybrid method combining linear and nonlinear structures of the time series to forecast air quality. The NO_2 concentration in a sampling site in Delhi from 1999 to 2003 was the case study, and results showed that the hybrid model outperformed the individual linear and nonlinear models. Zárate et al. (2007) introduced two versions of the emission inventory into a mesoscale air quality model in Bogota, Colombia, and found that the simulated concentrations fit the observed values well. In addition, model-based air quality forecasting systems have recently been developed for urban air pollution management (Finardi et al., 2008).

The previous models are mainly process based and focus on predicting accurate pollutant concentrations at various temporal and spatial scales; thus, having numerous meteorological data, a clear understanding of pollutant dispersion processes, and a large capacity for computation is essential to obtaining results. However, for air quality management on an urban scale, sometimes it is more useful to determine potential air quality changes from related driving factors. Huang (1992) showed the practical application of urban-scale air quality prediction from related contributing variables.

Uncertainty is an important issue to address both when developing and applying models (Liu et al., 2007). A practical air quality management strategy must consider the uncertainties in (1) the data set, (2) the model structure and parameters, and (3) the model output. Decision makers and stakeholders should be informed of the uncertainties in a straightforward and direct manner.

Using Bayesian statistics is a proper way to address uncertainty, and hence it has been commonly applied to modeling since the 1990s (Qian et al., 2003). Keats et al. (2007) proposed a Bayesian model to successfully address source determination problems in an urban environment. In Bayesian statistics, all unknown parameters are treated as random variables, and their distributions are derived from known information (Borsuk et al., 2001). Researchers can use results of previous studies, thus performing a rigorous uncertainty analysis and presenting the key information for management decision making (Reckhow, 1994). Here we applied a Bayesian hierarchical model (BHM) to predict urban air quality to support air quality management under uncertainty. The air quality described by Huang (1992) for urban Xiamen City, China, was used as the case study.

2. Materials and methods

2.1. Data source and multivariate statistical analysis

To demonstrate the proposed method, we used the air quality data of Huang (1992), which consisted of 31 sampling sites and grid squares ($1 \times 1\ km^2$) covering the entire urban district of Xiamen City, China (Table S1 in Supporting Materials). The air pollutants were SO_2, NO_x, and dust fall (DF); the contributing variables were industrial coal consumption (x_1), population density (x_2), traffic flow (x_3), and shopping density (density of retail stores and shopping malls, x_4).

Some multivariate statistical methods were conducted before the application of the BHM, including correlation analysis, classification and regression trees (CART), hierarchical cluster analysis (CA) and discriminant analysis (DA). Correlation analysis was conducted to provide a basis for the air quality model structure. The results could also help formulate the prior and likelihood functions in the Bayesian model. CART is a classification method that uses historical data to construct decision trees. There are two important steps in CART: construction of the maximum tree and pruning for the right tree size (Breiman et al., 1984). CART classifies the most important properties from a large number of variables and was used in this study to estimate which variables were most important for the pollutants.

CA is an unsupervised pattern recognition method. It is often used to detect similarities and to divide a large group into smaller ones (Almeida et al., 2007). Hierarchical CA, the most common type of CA (Shrestha and Kazama, 2007), is often used in the analysis of varieties (Gidhagen et al., 2005). It was used here to analyze differences among the raw data. In hierarchical CA, using Euclidean distance is the easiest and most intuitive way to mathematically define the similarity between different objects (Almeida et al., 2007). We conducted hierarchical CA on the data using Ward's method with squared Euclidean distances as a measure of similarity (Astela et al., 2006). Euclidean distance can be calculated using the following equation (Kannel et al., 2007):

$$d\left(x,y\right)=\sum_{m=1}^{p}\left(x_m-y_m\right) \tag{1}$$

where, $d(x,y)$ is the Euclidean distance between two items represented by x_m and y_m; p is the dimensional space of the variables. Analysis of variance (ANOVA) is used to calculate the distances between clusters in Ward's method, which is proved to minimize the sum of squares of any two possible clusters at each step (Ward, 1963). $D_{\text{link}}/D_{\text{max}}$ is usually used for exploring the linkage distance, which means the quotient between the linkage distances for a particular case divided by the maximal distance (Singh et al., 2005).

Hierarchical CA gives primary grouping results for the spatial variation of the raw data. DA is then often used to confirm the groups clustered by CA (Kannel et al., 2007). We applied DA based on the CA results, with the aim of determining the significance of different variables and minimizing the errors of these classifications. DA is a method of analyzing dependence that is a special case of canonical correlation. One of its objectives is to determine the significance of different variables, which can allow for the separation of two or more naturally occurring groups. The key step in DA is the construction of discriminant functions (DFs) for each group as follows (Kannel et al., 2007):

$$f\left(G_n\right)=c_n+\sum_{s=1}^{r}w_{ns}\cdot p_{ns} \tag{2}$$

where n is the number of groups (G), c_n is the constant inherent to each group, r is the number of parameters used to classify a set of data into a given group, w_s is the weight coefficient, assigned by DA to a given selected parameter (p_s). A standard mode was applied in this study to construct DFs to assess spatial variations, using raw data as the requirements of DA.

2.2. Multiple linear air quality model

The basic air quality model applied in this study is summarized as follows:

$$C=f\left(k,x\right)+\varepsilon \tag{3}$$

where C is the predicted pollutant concentration, k is the model parameter, x is the driving variable, and $\varepsilon$is the error. Correlation analysis, CART, hierarchical CA, and DA were used to determine the model structure of Eq. (3).

The correlation analyses of the raw data and the log-transformed data were reported (Table S2 in Supporting Materials). The Pearson correlation results showed the following: (a) Generally, there was a linear correlation between the predicted pollutant concentrations and the driving variables, except for x_3; (b) the linear correlation between x_2 and x_4 was high, with a correlation coefficient of 0.852, indicating a potential interaction between the two variables;

and (c) the raw data could be used in Eq. (3) instead of the log-transformed data because there were higher correlation coefficients among the raw data. CART showed that the most influential predictors were x_2 and x_4 for SO_2, x_2 for NO_x, and x_1 for DF (Fig. S1 in Supporting Materials). According to hierarchical CA, most of the sampling sites could be assigned to one group (Fig. S2 in Supporting Materials). DA showed that 93.3% of the originally grouped cases were correctly classified (Table S3 in Supporting Materials), and No. 28 should be assigned to Group B. The CA and DA results demonstrated that there was little variance in the spatial distribution. Thus, a multiple linear regression (MLR) equation was established to reflect the relationship between the specific pollutant concentrations and the four driving variables, as follows:

$$C = f(k, x) + \varepsilon = k_0 + k_1 \cdot x_1 + k_2 \cdot x_2 + k_3 \cdot x_3 + k_4 \cdot x_4 + k_5 \cdot x_2 x_4 + \varepsilon \tag{4}$$

Specifically, for air quality prediction at sampling site $i\,(i = 1, \ldots, 31)$, Eq. (3) can be described in detail as follows:

$$\underbrace{\begin{bmatrix} C_{1,j} \\ C_{2,j} \\ \ldots \\ C_{i,j} \\ \ldots \end{bmatrix}}_{\text{concentration}} = f(K_{m,j}, X_{i,q}) + \varepsilon = \overbrace{\begin{bmatrix} k_{0,j} \\ k_{0,j} \\ \ldots \\ k_{0,j} \\ \ldots \end{bmatrix}}^{\text{intercept}}$$
$$+ \underbrace{\begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} & x_{1,2}x_{1,4} \\ x_{2,1} & x_{2,2} & x_{2,1} & x_{2,4} & x_{2,2}x_{2,4} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ x_{i,1} & x_{i,2} & x_{i,3} & x_{i,4} & x_{i,2}x_{i,4} \\ \ldots & \ldots & \ldots & \ldots & \ldots \end{bmatrix}}_{\text{driving factors}} \overbrace{\begin{bmatrix} k_{1,j} \\ k_{2,j} \\ k_{3,j} \\ k_{4,j} \\ k_{5,j} \end{bmatrix}}^{\text{coefficient}} + \varepsilon \tag{5}$$

where j represents the three predicted pollutant concentrations (j = 1, 2, 3 represents SO_2, NO_x, and DF, respectively); q = 1, 2, 3, and 4 represents coal consumption, population density, traffic flow, and shopping density, respectively ; $m\,(m = 1, \ldots, 5)$ is the total number of intercept and slopes for each pollutant. A normally distributed error term was introduced in Eq. (5), with zero mean and a variance of σ^2, $\varepsilon \sim N\left(O, \sigma^2 I\right)$. There will probably be correlations among the regression intercept and slopes. Thus, there is a $m \times m$ covariance matrix Σ_K for the intercept and slopes, with diagonal elements $\Sigma_{mm} = \sigma_m^2$ and off-diagonal-elements $\Sigma_{ml} = \rho_{ml}\sigma_m\sigma_l\,(l = 1, \ldots, 5)$. The specification of Σ_K is a key issue for the regression modeling. Scaled inverse-Wishart distribution method was recommended in Gelman and Hill (2007) to obtain ρ_{ml} and Σ_K. Alternatively, the predicators can be centered $\left(x'_{i,q}\right)$ to eliminate the correlations of model coefficients as $x'_{i,q} = x_{i,q} - \left(\sum_{i=1}^{31} x_{i,q}\right)/31$(Gelman and Hill, 2007).

Conventional multilevel regression modeling is often used to solve model predictions and parameter estimations such as those in Eq. (5). There are two points of note regarding conventional regression models: (a) parameters identifiable as a necessary condition, and (b) the danger of underestimating the uncertainties of model predictions (Omlin and Reichert, 1999). For an uncertainty analysis, a conventional multilevel regression model can provide uncertainties of parameter values by standard errors (SE), which will sometimes distort the true distributions of the parameter errors and model errors. The Bayesian approach can address this by using a conventional multilevel regression model; however, we knew little about the parameters in Eq. (5). The non-informative prior distribution would be a challenge for a Bayesian application. Thus, we used a combination of a conventional multilevel regression model and a Bayesian approach. The results from the multilevel regression model could provide hints on how to reasonably determine the prior distributions of the parameters, thus avoiding the large standard deviations that usually occur with non-informative prior distributions. The R package 'lem4' was applied for parameters estimation in conventional multilevel regression modeling (Gelman and Hill, 2007). Here we centered the predicators to eliminate the coefficients correlations, using $x'_{i,q}$ to substitute $x_{i,q}$ in modeling.

2.3. Bayesian hierarchical model

In the Bayesian approach, the entire posterior distribution of parameter values is used for parameter estimation (Qian and Shen, 2007, Liu et al., 2008). Thus, the Bayesian method can involve uncertainty and parameter estimation in the modeling process, which can be useful for urban air quality management. The Bayesian inference was based on Bayes' theorem as follows (Gill, 2002):

$$\underbrace{p(\theta|y)}_{\text{posterior}} = \frac{\overbrace{p(\theta)}^{\text{prior}}\ \overbrace{p(y|\theta)}^{\text{likelihood}}}{\underbrace{p(y)}_{\text{probability of } y}} \propto p(\theta)\,p(y|\theta)\,, \tag{6}$$

where $\theta$is the parameter; $p(\theta|y)$ is the posterior probability of θ, which is the conditional distribution of the parameters after the observed data; $p(\theta)$is the prior probability of θ; and $p(y|\theta)$ is the likelihood function. We applied the specific hierarchical model considering the common characteristics of the three air pollutants on source and dispersion, which also provided the possibility of parameter pooling in the BHM. The BHM responds to model complexity by decomposing the problem into levels. Assuming a model with more than one unknown parameter, such as $\{\theta_1, \theta_2\}$, then given a BHM, Eq. (6) can be modified as follows:

$$p(\theta_1, \theta_2|y) \propto p(y|\theta_1, \theta_2)\,p(\theta_1|\theta_2)\,p(\theta_2) \tag{7}$$

The likelihood function for Eq. (5) then becomes:

$$\prod_{j=1}^{3}\prod_{i=1}^{31}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[\frac{\left(C^o_{i,j}-(k_{0,j}+k_{1,j}\cdot x_{i,1}+k_{2,j}\cdot x_{i,2}+k_{3,j}\cdot x_{i,3}+k_{4,j}\cdot x_{i,4}+k_{5,j}\cdot x_{i,2}x_{i,4})\right)^2}{-2\sigma^2}\right] \tag{8}$$

where $C^o_{i,j}$ is the observed value for the three pollutants in the 31 sampling sites.

The Monte Carlo method is widely used to assist with numerical summarization in Bayesian methods; and the Markov Chain Monte Carlo (MCMC) is a popular Monte Carlo algorithm (Qian et al., 2003). Three steps are usually involved in using the MCMC sampling method in a BHM (Qian et al., 2005, Malve and Qian, 2006): prior probability distribution formulation, likelihood function specification, and MCMC sampling. Usually, Gibbs sampling and Metropolis-Hastings algorithms are used to obtain the sampling for the posterior probability distributions. The model can be used for prediction when convergence is reached after sufficient burning in.

The hierarchical structure of the model can be described as follows after Malve and Qian (2006):

$$C^o_{i,j} \sim N\left[f\left(K_{m,j}, X'_{i,q}\right), \sigma^2\right] \tag{9}$$

$$K_{m,j} \sim N\left(k_m, \sigma \cdot k_m^2\right) \tag{10}$$

$$k_m \sim \text{unif}\left(0, \text{Upp} \cdot k_m\right) \tag{11}$$

$$\sigma \cdot k_m \sim \text{unif}\left(0, \text{Upp} \cdot \sigma \cdot k_m\right) \tag{12}$$

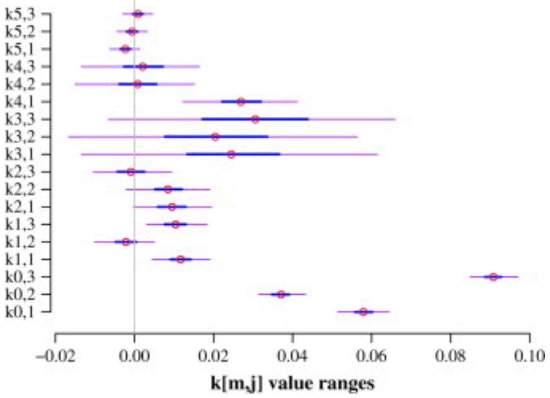
$$\sigma \sim \text{unif}\left(0, 0.2\right) \tag{13}$$

where Upp$\cdot k_m$ and Upp$\cdot\sigma\cdot k_m$ are decided from the conventional multilevel regression model results. σ, k_m, and $\sigma\cdot k_m$ are the hyper-parameters of the model, which can be estimated by using Eqs. (9), (10), (11), (12), (13).

The practical implementation of MCMC, using Gibbs sampling in this study, is based on a specialized and free software, WinBUGS (Lunn et al., 2000) running from *R* package R2WinBUGS (Gelman and Hill, 2007). The inference of this study was based on random 1000 taken from the posterior distribution after a sufficient “burn-in” and the MCMC algorithm had converged. A potential scale reduction factor, *Rhat*, was produced in package R2WinBUGS to determine the model convergence (at convergence, *Rhat*≈1.0).

3. Results and discussion

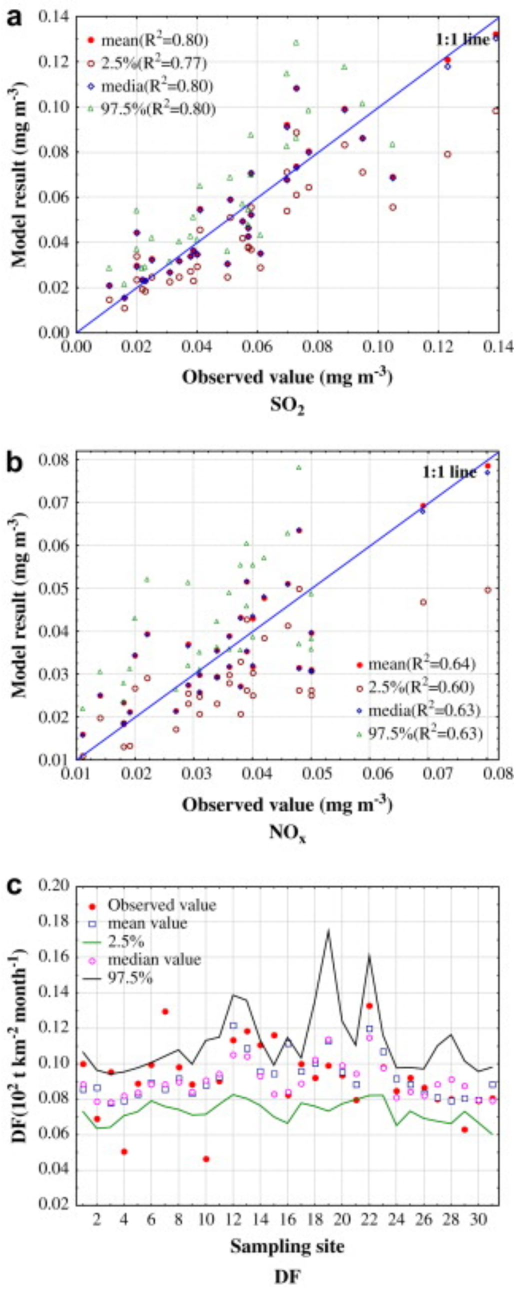
The MLR results from the conventional multilevel regression model were calculated using the *R* package ‘lem4’ (Table S4 in Supporting Materials), which provided useful information about the Bayesian approach. Upp·k_m and Upp·σ·k_m were then decided using 0.10, which provides a reasonable and practically “flat” region for parameter estimation. The MCMC simulations were carried out in WinBUGS with four chains, each with 20,000 iterations (first 10,000 discarded after model convergence), and 1000 samples for each unknown quantity were taken from the next 10,000 iterations to reduce autocorrelations of the data after Malve and Qian (2006). The *Rhat* values (≈1.0), MCMC trace plot of σ, and model variance (Figs. S3 and S4 in Supporting Materials) showed that the sequences converged rapidly. Fig. 1 describes the posterior distributions of $k_{m,j}$, including the mean value and 50% and 95% posterior credible intervals. The MLR model and $k_{m,j}$ results could then be used in Eq. (5) for air quality prediction and management given different uncertainty levels.



Download : Download full-size image

Fig. 1. The posterior distributions of $k_{m,j}$ in the MLR model. Circles are estimated posterior means, short thick lines are the 50% posterior credible intervals, and the long thin lines are the 95% posterior credible intervals.

Fig. 2 shows the fit between the observed data and the modeled values, with mean and median values and two credible levels (2.5% and 97.5%). The average relative errors between the observed data and the mean values from the BHM for SO_2, NO_x, and DF, respectively, were 6.81%, 6.79%, and 3.52%; for the 2.5% credible level, the relative errors were 16.13%, 18.00%, and 16.85%; for the 97.5% credible level, the relative errors were −32.59%, −34.96%, and −27.08%; and for median values, the relative errors were −6.27%, −6.38%, and −2.91%. The following factors contributed to the errors in the MLR model: (a) errors in monitoring and investigation; (b) the data used here were averaged values, and spatial accuracy was ignored; (c) we used an easily understood linear equation to estimate air quality, but despite the linear relationship between the driving factors and the air quality variables, the equation reflected only part of the true air quality because the urban air environment is such a complex and uncertain system. We also presented the correlations matrix among the model coefficients when before and after centering the predicators, which showed that centering the predictors eliminated the correlations (Table S5 in Supporting Materials). In addition, centering also makes the intercept (k_{0j}) meaningful and easy to interpret and verify the model's reliability. When industrial coal consumption, population density, traffic flow and shopping density were at their mean level, the concentration of SO_2, NO_x, and DF would reach 0.058 mg m^{-3}, 0.037 mg m^{-3} and 9.1 t km^{-2} mon^{-1}, respectively.



Download : Download full-size image

Fig. 2. The model fitting results for SO_2, NO_x and DF, between observed data and the modeled mean, median and 2.5% and 97.5% credible level values. R^2 is the coefficient of determination.

A Bayesian hierarchical model was applied, in combination with conventional multilevel regression model, to predict air quality in an urban environment. Compared to other methods, the BHM can provide probability levels for the estimation, so there is more room for related decision making under uncertainty. Model accuracy can be improved with the collection of more data, thereby changing the model equation to a nonlinear style.

4. Conclusion

A Bayesian hierarchical model was applied in this study for air quality predication in urban environment under uncertainty. The air quality reported by Huang (1992) for Xiamen City, China, was used as a case study, and the data included 31 sampling sites and three air quality variables. A linear equation was determined between the air quality variables (SO_2, NO_x, and DF) and four driving factors. An MCMC trace plot showed that the BHM sequences converged rapidly. The posterior distributions of $k_{m,j}$ at different credible levels were obtained from the model. The average relative errors between the observed data and the mean values from the BHM for SO_2, NO_x, and DF were 6.81%, 6.79%, and 3.52%, respectively. This case study shows that BHM is useful for predicting urban air quality; and that air quality can be predicted from related contributing factors on an urban scale.

Acknowledgments

The author would like to appreciate the editors, Dr. Song S. Qian, and the anonymous reviewers for their helpful comments on the paper. This paper was supported by the "China National Water Pollution Control Program" (2008ZX07102-001) and "National Basic Research (973) Program" Project (No.2005CB724205).

Appendix. Supplementary data

Download : Download Word document (5MB)

Recommended articles

References

Almeida et al., 2007 J.A.S. Almeida, L.M.S. Barbosa, A.A.C.C. Pais, S.J. Formosinbo
Improving hierarchical cluster analysis: a new method with outlier detection and automatic clustering
Chemometrics and Intelligent Laboratory Systems, 87 (2) (2007), pp. 208-217
View PDF View article View in Scopus Google Scholar

Astela et al., 2006 A. Astela, M. Biziukb, A. Przyjaznyc, J. Namieśnikb
Chemometrics in monitoring spatial and temporal variations in drinking water quality
Water Research, 40 (2006), pp. 1706-1716
Google Scholar

Borsuk et al., 2001 M.E. Borsuk, D. Higdon, C.A. Stow, K.H. Reckhow
A Bayesian hierarchical model to predict benthic oxygen demand from organic matter loading in estuaries and coastal zones
Ecological Modelling, 143 (2001), pp. 165-181
View PDF View article View in Scopus Google Scholar

Breiman et al., 1984 L. Breiman, J. Friedman, R. Olshen, C. Stone
Classification and Regression Trees
Wadsworth, California (1984)
p. 358
Google Scholar

Chelani and Devotta, 2006 A.B. Chelani, S. Devotta
Air quality forecasting using a hybrid autoregressive and nonlinear model

Atmospheric Environment, 40 (2006), pp. 1774-1780

View PDF View article View in Scopus Google Scholar

Dorling et al., 2003 S.R. Dorling, R.J. Foxall, D.P. Mandic, G.C. Cawley
Maximum likelihood cost functions for neural network models of air quality data
Atmospheric Environment, 37 (2003), pp. 3435-3443

View PDF View article View in Scopus Google Scholar

Finardi et al., 2008 S. Finardi, D.R. Maria, A. D'Allura, C. Cascone, G. Calori, F. Lollobrigida
A deterministic air quality forecasting system for Torino urban area, Italy
Environmental Modelling & Software, 23 (2008), pp. 344-355

View PDF View article View in Scopus Google Scholar

Gelman and Hill, 2007 A. Gelman, J. Hill
Data Analysis Using Regression and Multilevel/Hierarchical Models
Cambridge University Press, New York (2007)

Google Scholar

Gidhagen et al., 2005 L. Gidhagen, C. Johansson, J. Langner, V.L. Foltescu
Urban scale modeling of particle number concentration in Stockholm
Atmospheric Environment, 39 (2005), pp. 1711-1725

View PDF View article View in Scopus Google Scholar

Gill, 2002 J. Gill
Bayesian Methods: a Social and Behavioral Sciences Approach
Chapman & Hall/CRC, Boca Raton, Florida (2002)

Google Scholar

Hooyberghs et al., 2005 J. Hooyberghs, C. Mensink, G. Dumont, F. Fierens, O. Brasseur
A neural network forecast for daily average PM_{10} concentrations in Belgium
Atmospheric Environment, 39 (2005), pp. 3279-3289

View PDF View article View in Scopus Google Scholar

Huang, 1992 G.H. Huang
A stepwise cluster analysis method for predicting air quality in an urban environment
Atmospheric Environment, 26 (3) (1992), pp. 349-357

View PDF View article View in Scopus Google Scholar

Kannel et al., 2007 P.R. Kannel, S. Lee, S.R. Kanel, S.P. Khan
Chemometric application in classification and assessment of monitoring locations of an urban river system
Analytica Chimica Acta, 582 (2) (2007), pp. 390-399

View PDF View article View in Scopus Google Scholar

Keats et al., 2007 A. Keats, E. Yee, F.S. Lien
Bayesian inference for source determination with applications to a complex urban environment
Atmospheric Environment, 41 (2007), pp. 465-479

View PDF View article View in Scopus Google Scholar

Liu et al., 2007 Y. Liu, H.C. Guo, Z.X. Zhang, L.J. Wang, Y.L. Dai, Y.Y. Fan
An optimization method based on scenario analyses for watershed management under uncertainty
Environmental Management, 39 (5) (2007), pp. 678-690
CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Liu et al., 2008 Y. Liu, P.J. Yang, C. Hu, H.C. Guo
Water quality modeling for load reduction under uncertainty: a Bayesian approach
Water Research, 42 (13) (2008), pp. 3305-3314
View PDF View article View in Scopus ↗ Google Scholar ↗

Lunn et al., 2000 D.J. Lunn, A. Thomas, N. Best, D. Spiegelhalter
WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility
Statistics and Computing, 10 (2000), pp. 325-337
View in Scopus ↗ Google Scholar ↗

Mage et al., 1996 D. Mage, G. Ozolins, P. Peterson, A. Webster, R. Orthofer, V. Vandeweerd, M. Gwynne
Urban air pollution in megacities of the world
Atmospheric Environment, 30 (1996), pp. 681-686
View PDF View article View in Scopus ↗ Google Scholar ↗

Malve and Qian, 2006 O. Malve, S.S. Qian
Estimating nutrients and chlorophyll a relationships in Finnish lakes
Environmental Science and Technology, 40 (24) (2006), pp. 7848-7853
CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Omlin and Reichert, 1999 M. Omlin, P. Reichert
A comparison of techniques for the estimation of model prediction uncertainty
Ecological Modelling, 115 (1999), pp. 45-59
View PDF View article View in Scopus ↗ Google Scholar ↗

Pérez-Roa et al., 2006 R.J. Pérez-Roa, H. Jorquera, J.R. Pérez-Correa, V. Vesovic
Air-pollution modelling in an urban area: correlating turbulent diffusion coefficients by means of an artificial neural network approach
Atmospheric Environment, 40 (2006), pp. 109-125
View PDF View article View in Scopus ↗ Google Scholar ↗

Qian et al., 2003 S.S. Qian, C.A. Stow, M.E. Borsuk
On Monte Carlo methods for Bayesian inference
Ecological Modelling, 159 (2–3) (2003), pp. 269-277
View PDF View article View in Scopus ↗ Google Scholar ↗

Qian et al., 2005 S.S. Qian, K.H. Reckhow, J. Zhai, G. McMahon
Nonlinear regression modeling of nutrient loads in streams: a Bayesian approach
Water Resources Research, 41 (2005), p. W07012, 10.1029/2005WR003986 ↗
Google Scholar ↗

Qian and Shen, 2007 S.S. Qian, Z. Shen
Ecological applications of multilevel analysis of variance

Ecology, 88 (10) (2007), pp. 2489-2495

CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Reckhow, 1994 K.H. Reckhow
Importance of scientific uncertainty in decision-making
Environmental Management, 18 (1994), pp. 161-166

View in Scopus ↗ Google Scholar ↗

Shrestha and Kazama, 2007 S. Shrestha, F. Kazama
Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan
Environmental Modelling & Software, 22 (4) (2007), pp. 464-475

View PDF View article View in Scopus ↗ Google Scholar ↗

Singh et al., 2005 K.P. Singh, A. Malik, S. Sinha
Water quality assessment and apportionment of pollution sources of Gomti River (India) using multivariate statistical techniques: a case study
Analytica Chimica Acta, 538 (2005), pp. 355-374

View PDF View article View in Scopus ↗ Google Scholar ↗

Ward, 1963 J.H. Ward
Hierarchical grouping to optimize an objective function
Journal of the American Statistical Association, 58 (1963), pp. 236-244

View in Scopus ↗ Google Scholar ↗

Zárate et al., 2007 E. Zárate, L.C. Belalcázar, A. Clappier, V. Manzi, H. Van den Bergh
Air quality modelling over Bogota, Colombia: combined techniques to estimate and evaluate emission inventories
Atmospheric Environment, 41 (2007), pp. 6302-6318

View PDF View article View in Scopus ↗ Google Scholar ↗

Cited by (38)

Examining the effects of flood damage, federal hazard mitigation assistance, and flood insurance policy on population migration in the conterminous US between 2010 and 2019

2022, Urban Climate

Show abstract

Time series-based PM<inf>2.5</inf> concentration prediction in Jing-Jin-Ji area using machine learning algorithm models

2022, Heliyon

> *Citation Excerpt :*
>
> ...Epidemiological and experimental evidences have proven it to be associated with respiratory and cardiovascular mortality and morbidity rates, life expectancy (Burnett et al., 2014; Xing et al., 2016; Apte et al., 2018; Al-Hemoud et al., 2019; Diao et al., 2020; Bu

et al., 2021; Geng et al., 2021), and the threat to public health may remain even when its concentration is at low levels (Feng et al., 2016; Ouyang et al., 2020; Yu et al., 2020). Traditional statistical models such as partial least squares regression model (Polat and Gunay, 2015), generalized Markov model (Sun et al., 2013; Alyousifi et al., 2019), Bayesian method (Riccio et al., 2006; Liu et al., 2008; Faganeli Pucer et al., 2018), etc., are often used for the prediction of air pollutant concentration on time series. However, because these models all have the shortcoming of over-simplified, they inherently have difficulties in unraveling the nonlinear interaction relationship between multivariate factors and PM2.5 concentration, so that the favorable factors for PM2.5 prediction cannot be fully utilized (Ni et al., 2017)....

Show abstract

Spatio-temporal modelling of PM<inf>10</inf> daily concentrations in Italy using the SPDE approach

2021, Atmospheric Environment

Show abstract

An LSTM-based aggregated model for air pollution forecasting

2020, Atmospheric Pollution Research

Citation Excerpt :

...Due to uncertainty among studies on PM 2.5 by machine learning (Barai et al., 2007), the Bayesian hierarchy model was used to forecast urban air quality in the context of air quality management. The results show that certain gasses such as SO2, NO2 are useful for air quality prediction (Liu Y. et al., 2008). The evaluation model still has scope for improvement through predictive methods such as Multiple Linear Regression (MLR), ARIMA, and the Generalised Regression Neural Network (GRNN) for PM2.5....

Show abstract

The air quality index trend forecasting based on improved error correction model and data preprocessing for 17 port cities in China

2020, Chemosphere

Citation Excerpt :

...For example, Mishra and Goyal (2016) used artificial neural network (ANN), multivariate linear regression (MLR), artificial intelligence-based neural blur (NF), and other individual models to forecast air pollution. Feng et al. (2013) used wavelet transform and neural network integration model Yong et al. (2008) used a Bayesian hierarchical model. Kumar and Goyal (2011) used the principal component regression (PCR)....

Show abstract

Explore a Multivariate Bayesian Uncertainty Processor driven by artificial neural networks for probabilistic PM<inf>2.5</inf> forecasting

2020, Science of the Total Environment

Show abstract

View all citing articles on Scopus ↗

View Abstract