JOURNAL OF
ENVIRONMENTAL HYDROLOGY

The Electronic Journal of the International Association for Environmental Hydrology
On the World Wide Web at http://www.hydroweb.com

VOLUME 14 2006

A STOCHASTIC MODEL FOR IDENTIFICATION OF TRENDS IN OBSERVED HYDROLOGICAL AND METEOROLOGICAL DATA DUE TO CLIMATE CHANGE IN WATERSHEDS

A. G. Bobba[1]
R.P. Rudra[2]
J. Y. Diiwu[2]

[1] National Water Research Institute, Burlington, Ontario, Canada
[2] School of Engineering, University of Guelph, Guelph, Ontario, Canada

The development of appropriate public policy responses to impact of climate change on the overall ecosystem would require knowledge of future trends in the natural processes affecting the ecosystem. A stochastic model is therefore formulated for simulating trends in hydrological and meteorological variables. The choice of auto-regressive moving average models of orders p and q (that is, ARMA (p, q) models) is intended to retain any persistence in the natural processes. The model development involved three stages: model identification, parameter estimation, and diagnostic checks. It was found that an ARMA(1, 1) model was adequate for modeling the three variables of temperature, precipitation, and stream flow on a season basis in the Northeast Pond River watershed. Diagnostic checks showed that for each variable the residuals were independent and normally distributed (that is, "white noise"), indicating that the fitted models are the most parsimonious and of best statistical fit.

INTRODUCTION

Hydrological impacts of climate change at watershed and regional scales have been studied in many different parts of the world (Nemeck and Schaake, 1982; Gleick, 1986; Manabe and Weatherald, 1987; Bobba et al., 1997). Also, the modeling of global climate systems has been extensively discussed in the literature (Gleick, 1986; Kite, 1989; Bloomfield, 1992; Bobba et al., 1999). All these studies have shown that relatively small changes in temperature and precipitation can cause large changes in water availability. Since changes in water availability have the potential to impact such human activities as agriculture, energy use, flood control, municipal and industrial water supply, and the overall ecosystem management, the impacts of climate change need to be incorporated in planning and policy formulation.

Development of appropriate public policy responses relevant to human activities it would require some means of identifying and evaluating likely impacts of climate change. However, the ability to assess the impacts of climate change on ecosystems is highly dependent on the ability to evaluate future climatic scenarios (Bobba et al., 1997). In the evaluation process appropriate models would come handy for simulating the relevant time series with the objective of retaining any trends and/or persistence in the series.

In the past various models have been discussed, including water-balance models, for evaluating climate changes and their impacts on water resources (Gleick, 1986; Manabe and Weatherald, 1987; Bobba et al., 1997). Also, general circulation models which model the physical and dynamic climatic processes have been discussed (Manabe, 1982). Gleick (1986) proposed that regional water-balance models be combined with information from general circulation models with the possibility of enhancing the abilities of both. The complexity of general circulation models, combined with the significant amount of computer time they require, limits the applicability of these models for planning and policy formulation. Moreover, due to their global nature results from such models are more applicable at regional scale. For application at local and zonal scales, as is usually the case in specific policy for predicting and reducing the severity of hydrologic changes in a watershed, alternative models are needed.

It is well known that at watershed scale increasing temperature and decreasing precipitation cause decreasing flows and vice versa (Nemeck and Schaake, 1982; Gleick, 1986; Bobba et al., 1997). Evidence of climate change may therefore be shown by positive or negative trends in time series of natural flows. Such trends are usually associated with changes in the structure of the time series caused by cumulative natural or man-made phenomena. The use of a time series modeling approach enables the incorporation of the autocorrelation and cross-correlation structures in the hydrologic variables that are caused by the natural phenomena (Kite, 1989; Hipel and McLoed, 1994). As explained by Hiltunen (1994), more insight into climatic trends may be obtained by combining time series analysis with water-balance approach to climate studies. One advantage of time series models is that they may be applied at any scale; also, such models are computationally convenient. Moreover, the parameters of time series models are not sensitive to the units of measurement. This makes time series models very convenient to use in climate studies.

In an earlier part of this study, observed data on temperature, precipitation, and stream flow from the Northeast Pond River watershed were analyzed with the objective of identifying and modeling trends due to climate change (Diiwu et al., 2000). In the current part of the study a stochastic model is developed which retains the identified trends and any persistence in each of the time series of temperature, precipitation, and stream flow due to climate change. The developed models are intended to be parsimonious (that is, use the minimum number of parameters) and of good statistical fit. The models may be used for forecasting and simulating any of the variables for the purpose of evaluating future impacts of climate change on geophysical and socioeconomic systems. The study watershed is briefly described in the next section.

THE NORTHEAST POND RIVER WATERSHED

The Northeast Pond River watershed (Figure 1) was selected for this study. Observed data on temperature, precipitation, and stream flow from the watershed were used for the identification, parameter estimation and diagnostic checks of the stochastic models. The study watershed, which has an area of 3.90 km^2 is located approximately 20 km west of St. John's in Newfoundland. The climate of the study area is dominated by the Labrador current, which consists primarily of arctic waters. The area has a marine climate, characterized by short but pleasant summers and mild winters (Environment Canada, 1995; Bobba et al., 1997). Monthly mean temperature, precipitation and stream flow were analyzed for the period 1952 to 1983. Daily temperature and precipitation data were obtained from St. John's meteorological station, and daily stream flow data were obtained from the gauge at the watershed outlet at Portugal Cove (Bobba et al., 1997).

MODEL DEVELOPMENT

Development of the model for each variable (temperature, precipitation, and stream flow) was done in three stages. These are model identification, parameter estimation, and diagnostic checks of the fitted models. This approach was taken to satisfy the principle of model parsimony; also the models should be of good statistical fit to the observed data. The theoretical background of these three stages is discussed in the following subsections.

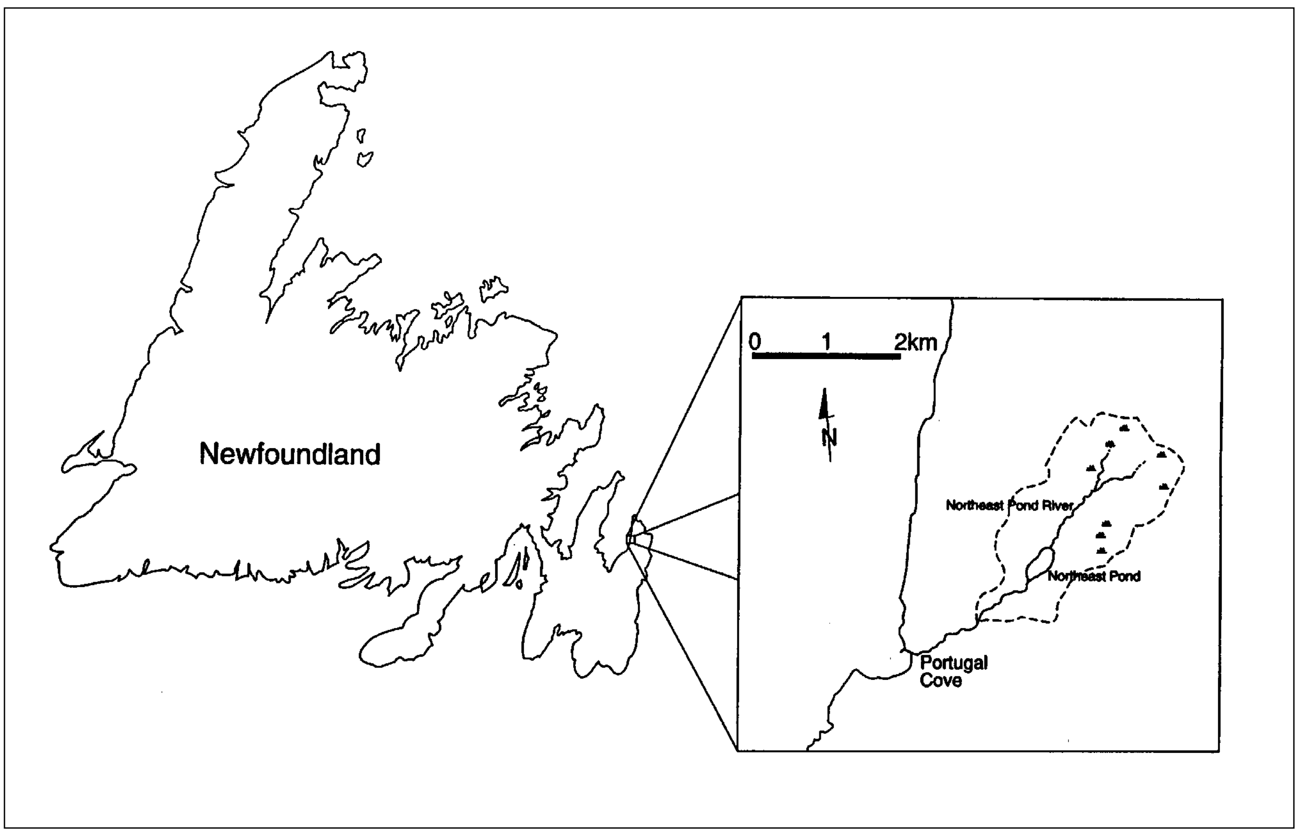

Figure 1. Location of the Northeast Pond river watershed

MODEL IDENTIFICATION

At the model identification stage analysis of the time series was carried out to determine appropriate structures for the models to be fitted. The ability to detect the statistical characteristics of a time series that change significantly over time may only be possible when the annual records cover a sufficiently long time horizon. It is therefore often reasonable to assume that a stationary model would be adequate for annual hydrologic and other natural time series of moderate lengths (Hipel and McLeod, 1994). It has been recommended in earlier studies that hydrologic and geophysical time series that have not been significantly influenced by man-made interactions such as land use changes or catastrophic effects such as earthquakes may be assumed to be stationary (Hipel and McLeod, 1994). The model identification in this study was therefore restricted to stationary models since the observed data being used for the modeling were annual means for a period of 32 years during which no land use or catastrophic changes in the watershed were reported.

Graphical methods of model identification were used. The exploratory data analysis carried out involved the construction of the autocorrelograms and analysis of their trends (Figures 2a, b and c). The sample autocorrelation function used in this study is defined as (Salas, 1993):

$$r_k = C_k / C_0 \tag{1}$$

where C_k denotes the auto-covariance function defined as (Salas, 1993):

$$C_k = \frac{1}{N}\sum_{1}^{N-k}(X_t - \overline{X})(X_{t+k} - \overline{X}) \tag{2}$$

The symbol X_t denotes the time series with its mean $\bar{X}$ and the separation distance k between the realizations of the time series that are being correlated is usually called the lag. It is recommended that k satisfies the relation (Hipel and McLeod, 1994):

$$0 \le k \le \frac{N}{4} \tag{3}$$

To determine which values of the estimated autocorrelation function are significantly different from zero, the 95% confidence limits were drawn on the same axes as the autocorrelogram. A given value of the autocorrelation function is significantly different from zero if that value falls outside the confidence interval. The lag p beyond which r_k is not significantly different from zero, determines the extent to which the time series is auto-correlated. By this the order of the auto-regressive (AR) component of the time series model is obtained as p. If r_k attenuates but does not cut off then this indicates that a moving average (MA) component is needed in the time series model. If however, r_k truncates then an auto-regressive component alone is adequate to model the time series (Hipel and McLeod, 1994).

In case a moving average component is needed the order of that component is determined by examining the sample partial autocorrelogram. For such a time series the partial autocorrelogram would attenuate, and the lag q beyond which the partial autocorrelation values are not significantly different from zero (that is, lie within the 95% confidence limits) is chosen as the order of the moving average component of the model. In that case an ARMA (*p, q*) model would be fitted to the time series.

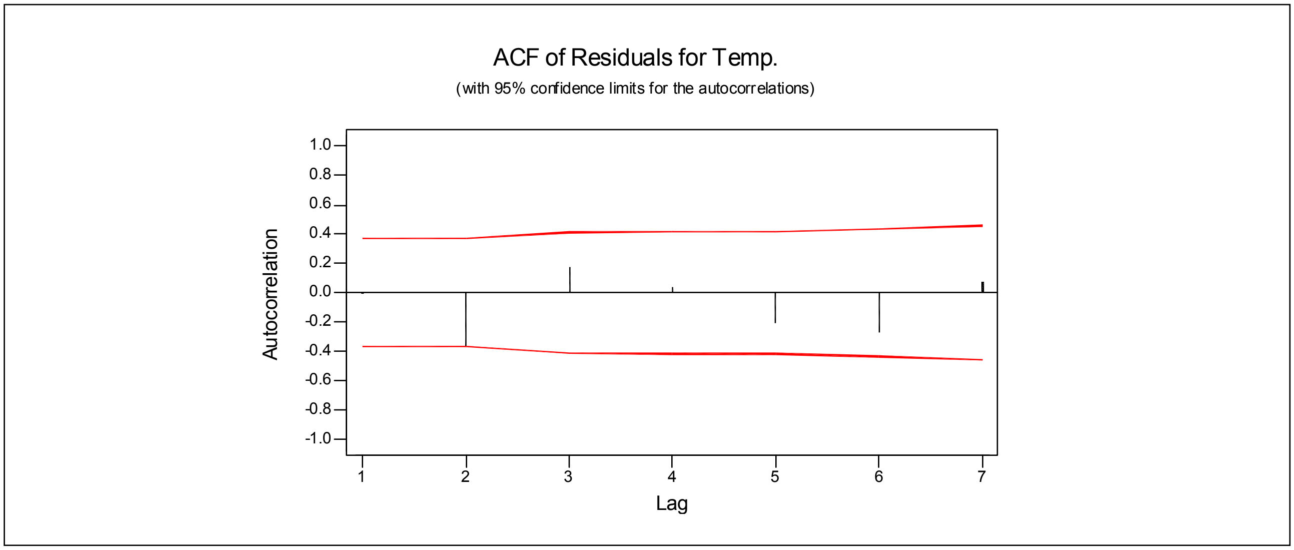

Figure 2a. ACF of residuals for temperature

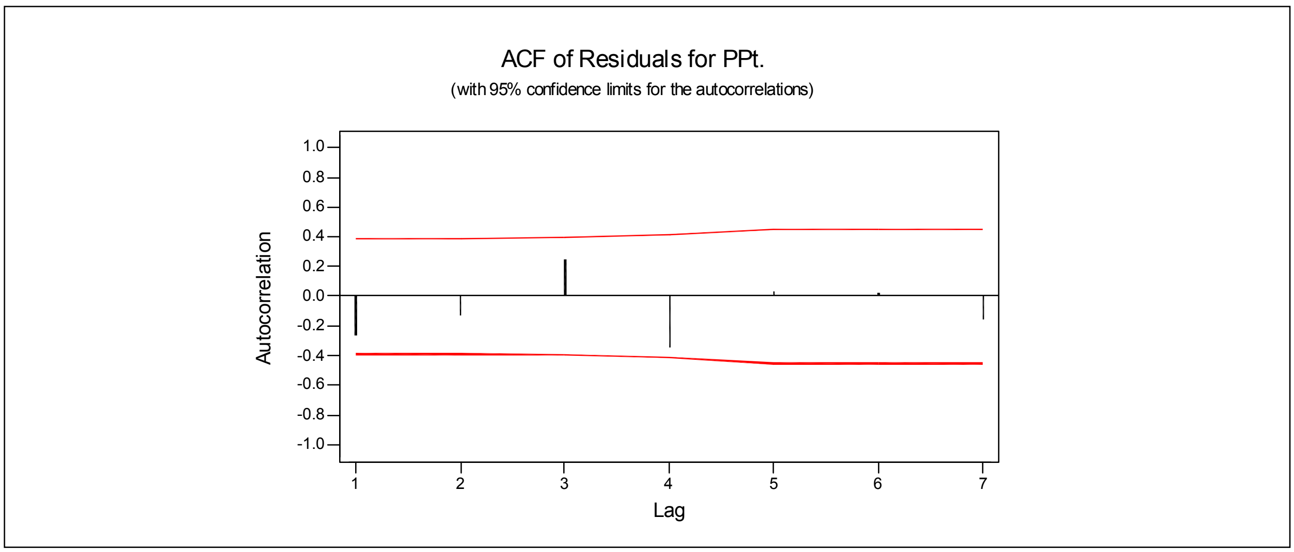

Figure 2b. ACF of residuals for precipitation

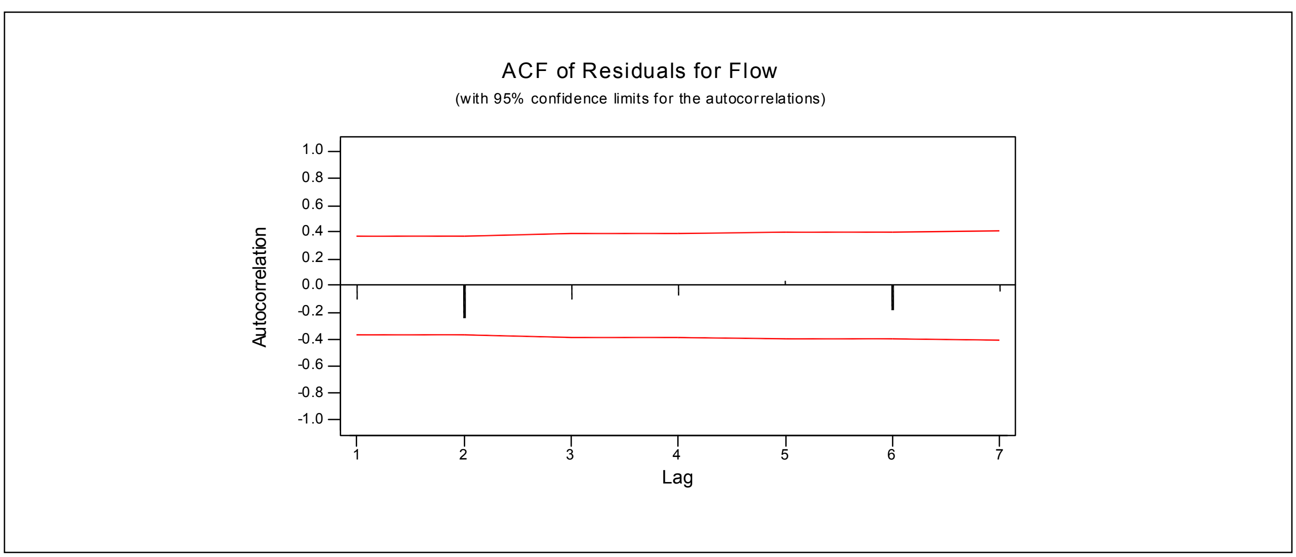

Figure 2c. ACF of residuals for flow

PARAMETER ESTIMATION

The general form of an auto-regressive moving average model ARMA (*p, q*), consisting of an auto-regressive component of order p and a moving average component of order q, is (Salas, 1993):

$$X_t = \mu + \sum_1^p \phi_j (X_{t-j} - \mu) + \varepsilon_t - \sum_1^q \theta_j \varepsilon_{t-j} \tag{4}$$

There are p auto-regressive parameters $\phi_1, \phi_2, \ldots \phi_p$ and q moving average parameters $\theta_1, \theta_2, \ldots \theta_q$.

The symbol μ denotes the mean of the population from which the time series is drawn, and ε_t is called the innovation series (also known as the residual series). The appropriate values of p and q were determined at the model identification stage. If moving average terms are not needed in the model then $q = 0$; on the other hand if auto-regressive terms are not needed then $p = 0$.

In this study the most appropriate model structures were determined by model identification as discussed in the preceding subsection. The parameters in the selected models were then estimated using the maximum likelihood estimation technique. This estimation technique was chosen because it is efficient and usually results in consistent estimates of the parameters (Hipel and McLeod, 1994). For each of the time series of temperature, precipitation, and stream flow the parameters estimated were the mean of the series, auto-regressive parameters, moving average parameters, and the innovation series.

In the maximum likelihood estimation, the objective was to determine the values of the parameters that maximize the likelihood function $L(\beta/\omega)$ or its natural logarithm. Here β denotes a vector consisting of the parameters $\phi_1, \phi_2, \ldots \phi_p, \theta_1, \theta_2, \ldots \theta_q, \sigma_\varepsilon^2$. The symbol ω denotes the set of observations. Such estimates are called maximum likelihood estimates (McCuen, 1993). As a measure of the reliability of the estimated parameters their corresponding standard errors of estimate were computed. Model discrimination can be accomplished by comparing estimated parameters with their corresponding standard errors of estimate. If the absolute value of the estimated parameter is greater than its standard error of estimate, then it can be argued that even at 1% significance level, the estimated parameter is significantly different from zero and should be included in the model. On the other hand, if the standard error of estimate for a parameter is greater than the absolute value of the estimated parameter then for model parsimony the parameter should not be included in the model. If all the standard error of estimate values are large compared to their corresponding estimated parameters, then the fitted model should be used with caution in certain kinds of applications. In this study for those cases for which the standard error of estimate values exceed the absolute values of their corresponding estimated parameters, differencing is applied to the original data prior to fitting the time series model.

DIAGNOSTIC CHECKS

Diagnostic checks are carried out to insure that the underlying model assumptions are satisfied, in which case the fitted model adequately describes the time series under consideration. In this study the method used for diagnostic checks is to determine whether or not the assumptions about the innovation series, that they are independently and normally distributed with constant variance, are satisfied by the residuals of the calibrated model (Figures 3 a, b and c). If the residuals are independent and normally distributed, then they necessarily have constant variance (Hipel and McLeod, 1994).

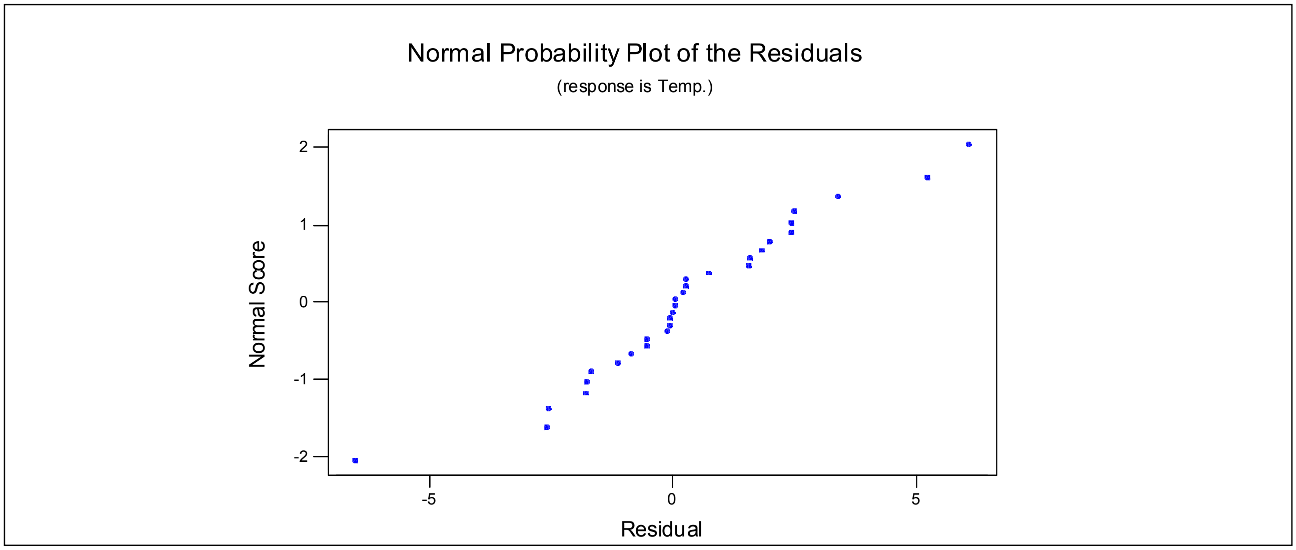

Figure 3a. Normal probability plot of residuals for temperature.

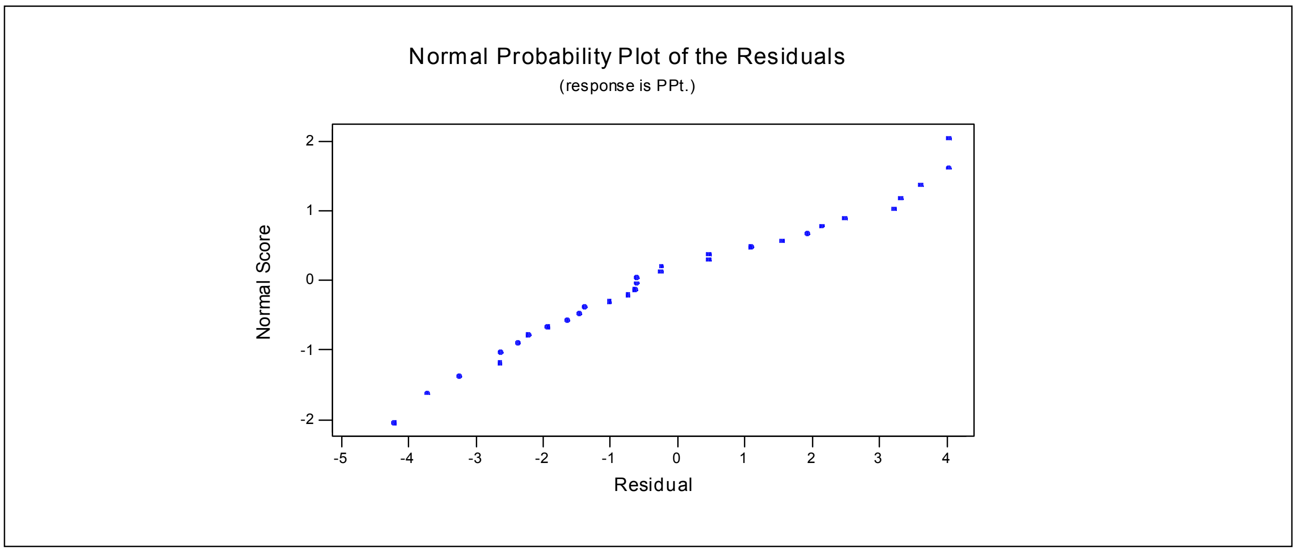

Figure 3b. Normal probability plot of residuals for precipitation.

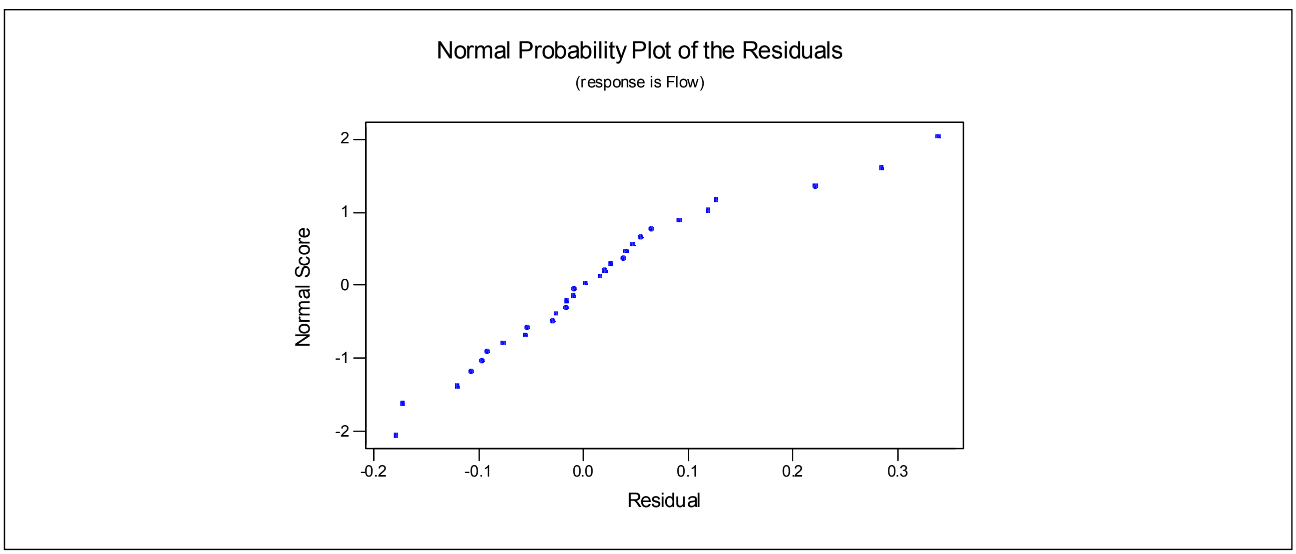

Figure 3c. Normal probability plot of residuals for flow.

Test for Normality

Normal probability plots for the residuals were carried out to test for normality. For normally distributed residuals the plotted points are uniformly distributed about a straight line. In case the probability plot depicts a curve rather than a straight line then the residual series is not normally distributed.

Test for Independence

The innovation series is independently distributed if the estimated residuals of the fitted time series model are uncorrelated, that is the residual series are "white noise". To test if the residuals are uncorrelated, the autocorrelogram of the residuals is examined along with the 95% confidence limits. If none of the autocorrelation values is significantly different from zero, that is all autocorrelation values fall within the 95% confidence limits, then the residuals are "white noise".

RESULTS AND DISCUSSION

The observed data was analyzed on a season basis. Winter was assumed to include the months December, January, and February, while spring was assume to include March, April, and May. Summer months were assumed to be June, July, and August and the fall months were September, October, and November. For each of the variables (temperature, precipitation, and stream flow) the average of the data for the three months of each season was used as the time series data for that season. Model identification, parameter estimation, and diagnostic checks were carried out using the Minitab statistical software (Minitab, 2000). The results of this exercise are reported in the following subsections.

Model Identification

The autocorrelogram and partial autocorrelogram for each variable in each season were plotted on the same axes as their 95% confidence limits**.** The details were presented by Diiwu et. al., 2000, and Bobba and Diiwu, 2002. These graphs (Figures 2a, 2b, and 2c) showed that except for temperature in spring and precipitation in summer, all the autocorrelation and partial autocorrelation values are not significantly different from zero since the entire values lie within the 95% confidence limits.

In the case of temperature in spring it is observed that the autocorrelation and partial autocorrelation values at lag 2 are significantly different from zero since these values lie outside the 95% confidence limits. Similarly, for precipitation in summer the values at lag 1 are significantly different from zero. Hence for calibration purposes a time series model with one auto-regressive and one moving average term, that is ARMA (*1,1*) model, was selected for each variable. The selected ARMA (*1,1*) model to be fitted is of the form:

$$X_t = \mu + \phi_1 (X_{t-1} - \mu) + \varepsilon_t + \theta_1 \varepsilon_{t-1} \quad (5)$$

where μ denotes the population mean, ε_τ denotes the innovation series, ϕ_1 is the auto-regressive parameter, and θ_1 is the moving average parameter.

Parameter Estimation

The selected ARMA (*1,1*) model was fitted to each of the time series by estimating the parameters μ, ε_τ, ϕ_1, and θ_1.

The parameter estimation was carried out by means of the maximum likelihood estimation technique. The corresponding standard error of estimate (SEE) values computed were used to assess the reliability of the estimated parameters of the selected model. For a reliable estimate of a parameter the absolute value of the estimated parameter should exceed the corresponding SEE value. The normal probability plots of the estimated residuals, presented by Diiwu et al. 1999, were also used to confirm the reliability of the estimated parameters. The estimated parameters, along with their SEE values, are presented in Table 1.

The results show that while some of the time series could be directly fitted by an ARMA (*1,1*) model, many of them required one-step differencing before fitting the model and a few of the estimates were even further improved by two-step differencing before fitting the model. For the time series of winter temperature no differencing was required to improve the estimates because

Table 1. Estimated ARMA (1,1) parameters and SEE values (shown in brackets) for different seasons.

	Differencing: 0				Differencing: 1			Differencing: 2		
	ϕ_1	θ_1	ε_t	μ	ϕ_1	θ_1	ε_t	ϕ_1	θ_1	ε_t
Winter										
Temperature	-0.995 (0.052)	-0.831 (0.12)	-6.738 (0.50)	-3.377 (0.25)	-0.372 (0.22)	0.609 (0.19)	0.021 (0.12)	-0.742 (0.14)	0.954 (0.18)	-0.007 (0.026)
Precipitation	-0.608 (0.70)	-0.455 (0.79)	9.181 (0.25)	5.709 (0.16)	-0.482 (0.22)	0.448 (0.22)	0.033 (0.10)	-0.705 (0.14)	0.943 (0.17)	-0.0009 (0.019)
Stream Flow	0.305 (2.32)	0.234 (2.37)	0.104 (0.01)	0.149 (0.01)	0.037 (0.22)	0.956 (0.16)	-0.0003 (0.001)	-0.553 (0.17)	0.945 (0.20)	-0.0001 (0.001)
Spring										
Temperature	-0.996 (0.07)	-0.792 (0.14)	12.214 (0.30)	6.118 (0.15)	-0.370 (0.19)	0.944 (0.17)	0.060 (0.02)	-0.839 (0.14)	0.967 (0.16)	-0.004 (0.011)
Precipitation	0.139 (0.60)	0.447 (0.56)	3.068 (0.11)	3.564 (0.12)	-0.357 (0.19)	0.935 (0.18)	0.037 (0.03)	-0.658 (0.16)	0.950 (0.22)	0.0098 (0.03)
Stream Flow	0.645 (0.22)	0.942 (0.16)	0.059 (0.00)	0.166 (0.00)	-0.174 (0.21)	0.943 (0.17)	-0.0004 (0.001)	-0.583 (0.18)	0.966 (0.22)	0.0002 (0.001)
Summer										
Temperature	0.528 (0.44)	0.750 (0.35)	6.690 (0.04)	14.173 (0.09)	-0.120 (0.21)	0.979 (0.22)	0.0299 (0.02)	-0.482 (0.17)	0.976 (0.17)	0.0052 (0.019)
Precipitation	0.151 (0.37)	-0.390 (0.34)	2.931 (0.27)	3.452 (0.32)	0.374 (0.19)	0.963 0.13)	0.0241 (0.02)	-0.063 (0.22)	1.041 (0.26)	-0.008 (0.032)
Stream Flow	0.114 (0.53)	-0.248 (0.51)	0.063 (0.01)	0.072 (0.01)	0.302 (0.20)	0.970 (0.15)	0.0006 (0.001)	-0.240 (0.20)	0.970 (0.19)	0.0003 (0.001)
Fall										
Temperature	0.015 (2.75)	-0.053 (2.75)	3.098 (0.19)	3.146 (0.19)	0.075 (0.19)	1.062 (0.11)	-0.013 (0.02)	-0.461 (0.18)	0.944 (0.21)	-0.0029 (0.030)
Precipitation	0.557 (0.56)	0.733 (0.47)	2.348 (0.06)	5.303 (0.13)	-0.203 (0.20)	0.947 (0.18)	0.0340 (0.03)	-0.608 (0.16)	0.950 (0.21)	-0.0145 (0.033)
Stream Flow	-0.123 (0.84)	-0.335 (0.80)	0.192 (0.02)	0.171 (0.02)	0.190 (0.21)	0.980 (0.16)	-0.0003 (0.001)	-0.295 (0.20)	0.950 (0.23)	-0.0008 (0.003)

it may be seen that all the SEE values were less than the absolute values of their corresponding estimated parameters. In the case of winter precipitation the absolute values of f_1 and q_1 are exceeded by their corresponding SEE values. Moreover, it may be noticed that the probability plot has curvature at the extremes, negating normality of the residuals. It is noticed that one-step differencing improved the estimates of f_1 and q_1 since the recomputed absolute values of these parameters exceeded their corresponding SEE values. Also, points of the replotted probability plot fell more closely to a straight line than that of the probability plot before differencing. However, two-step differencing did not give better estimates and the probability plot became curved at the extremes, indicating that one-step differencing was adequate to yield good estimates of the parameters.

Diagnostic Checks

The reliability of the estimated parameters alone cannot be used to check the goodness-of-fit of the models. For a complete diagnostic check the model assumptions must also be confirmed to be satisfied by the fitted model. In this study the assumptions of independent and normally distributed innovation (residual) series with constant variance were checked to confirm that the fitted model satisfied them. However, since an independent and normally distributed series necessarily has constant variance (Hipel and McLeod, 1994), only the assumptions of independence and normality of the residuals of the fitted model were checked in this study.

Test for Normality

The normal probability plots carried out are (Diiwu et al. 2000) shown in Figures 3a, 3b and 3c. It is noticed that in some cases (Diiwu et. al. 2000), the points plotted close to a straight line without the need for differencing. In other cases, the points plotted closer to straight lines when two-step differencing was applied prior to fitting the ARMA (1, 1) model (Diiwu et al. 2000). In the rest of the cases, applying one-step differencing prior to fitting the model yielded probability plots in which the points plotted close to straight lines. Hence for all the fitted models, the resulting residuals are normally distributed.

Test for Independence

The autocorrelograms of the estimated residuals presented earlier (Diiwu et al. 1999), shows that except the autocorrelation value at lag zero, all other autocorrelation values are not significantly different from zero since they all lie within the 95% confidence limits. This confirms that the residuals of the fitted models are independent (uncorrelated) within 95% confidence limits. That is, the residuals are “white noise”.

The diagnostic checks have therefore confirmed that for each of the time series of temperature, precipitation, and stream flow in each of the four seasons of winter, spring, summer and fall the fitted time series model adequately describes the time series in terms of parsimony and best statistical fit. The models may therefore be used to simulate future trends in hydrological and meteorological variables for planning and policy formulation purposes.

CONCLUSIONS

Autoregressive moving average (ARMA (*1,1*)) models have been formulated for modeling temperature, precipitation, and stream flow in winter, spring, summer, and fall seasons in the Northeast Pond River watershed. While some of the observed time series needed one or two-step differencing to be applied before estimation of the model parameters, a few of them did not require any differencing. This seems to be attributable to the possibility that trends in the time series due to climate change could introduce non-stationarity to the time series. If even this is the case, the non-stationarity is adequately taken care of by differencing since diagnostic checks show that the fitted models satisfy all the basic assumptions of independent and normally distributed innovations with constant variance. Moreover, the fitted models satisfy the principle of parsimony and are of best statistical fit.

ACKNOWLEDGMENTS

We would like to thank Prof. V. P. Singh, Louisiana State University, Baton Rouge, LA, USA and Prof. T.C. Sharma, University of Saskatoon, Saskatoon, Canada for their constructive comments and suggestions.

REFERENCES

Bloomfield, P. 1992. Trends in global temperature. J. Climate Change, 21: 1-16.

Bobba, A.G., and J.Y. Diiwu. 2002. Stochastic analysis of observed hydrological and meteorological data to identify and model trends due to climate change. Hydrological Science and Technology, 18, 43-54.

Bobba, A.G., V.P. Singh, D.S. Jeffries, and L. Bengtsson. 1997. Application of a watershed runoff model to Northeast Pond River, Newfoundland: to study water balance and hydrological characteristics owing to atmospheric change. Hydrol. Proc., 11: 1573 – 1593

Bobba, A.G., D.S. Jeffries, and V.P. Singh. 1999. Sensitivity of hydrological variables in the Northeast Pond River watershed, Newfoundland, Canada, due to climate change. Water Resour. Mngt., 13: 171 – 188

Diiwu, J.Y., A.G. Bobba, and R.P. Rudra. 2000. Analysis of observed hydrological and meteorological data to identify and model trends due to climate change. Environment Canada Contract No. KW504-9-0119. Burlington, Ontario, Canada.

Environment Canada. 1995. The state of Canada's climate: monitoring variability and change. SOE Report No. 95 – 152. Downsview, Ontario, Canada.

Gleick, P.H. 1986. Methodology for evaluating the regional hydrologic impacts of global climate changes. J. Hydrol. 88, 97-116.

Hiltunen, T. 1994. What do hydrological time series tell about climate change? Publications of Water and Environment Research Institute. National Board of Waters and the Environment, Finland. No. 17: 37-50.

Hipel, K.W., and A.I. McLeod. 1994. Time Series Modelling of Water Resources and Environmental Systems. Elsevier Science B. V. The Netherlands. 1013pp.

Kite, G. 1989. Use of time series analysis to detect climate change. J. Hydrol. 111, 259-279.

Manabe, S.R. 1982. Simulation of climate by general circulation models with hydrologic cycles. In: P. S. Eagleson (ed.), Land Surface Processes in Atmospheric General Circulation Models. Cambridge University Press, 19-66.

Manabe, S.R. and T. Weatherald. 1987. Large scale changes in soil wetness induced by increase in carbon dioxide. J. Atmos. Sci. 44: 1211-1235.

McCuen, R.H. 1993. Microcomputer applications in statistical hydrology. Prentice Hall, New Jersey.

Minitab Inc. 2000. Minitab Statistical Software. Release 13 for Windows. Published by Minitab Inc., USA.

Nemeck, J. and J. Schaake. 1982. Sensitivity of water resource systems to climate variation. Hydrol. Sci. 27: 327-343.

Salas, J.D. (1993). Analysis and modeling of hydrologic time series. Chapter 19 in (Maidment D. R., ed.) Handbook of Hydrology. McGraw-Hill, New York.

ADDRESS FOR CORRESPONDENCE
A. G. Bobba
National Water Research Institute
Burlington
Ontario, Canada L7R 4A6

Email: Ghosh.Bobba@ec.gc.ca