UNIVERSITÉ DU QUÉBEC
Institut National de la Recherche Scientifique
Centre Eau Terre Environment

Modélisation GARCH multivariée pour les variables climatiques et hydrologiques

Par

REZA MODARRES

Thèse présentée pour l'obtention du grade de Philosophiae Doctor (Ph.D.) en sciences de l'eau

Jury d'évaluation:

Examinateur externe	Ali A. Assani Université du Québec à Trois-Rivières
Examinatrice externe	Jan Adamowski Université McGill
Examinatrice interne	André St-Hilaire INRS-ETE
Directeur de recherche	Taha B.M.J. Ouarda INRS-ETE

Abstract

Hydrologic time series modeling usually includes linear approaches which model the time varying mean or the conditional mean of the hydrologic variables. However, most of the hydrologic variables show nonlinear variations through time. The nonlinear modeling of hydrologic variables has received considerable attentions in recent decades. Although a number of nonlinear models have been presented in the literature, the nonlinear time series models have not been sufficiently applied in hydrology and climatology. As the hydroclimatic variables change and influence each other within a temporal and spatial scale, it is essential to apply the appropriate multivariate models which take into account the nonlinear relationships between hydrologic variables through space and time.

The main goal of this study is to propose and develop a class of multivariate time series models called 'Multivariate Generalized Autoregressive Conditional Heteroscedasticity' (MGARCH) model, usually applied in financial time series modeling, for different hydrologic and climatic variables. The MGARCH modeling approach is used to model the conditional variance-covariance or volatility-covolatility of hydroclimatic variables. This study presents different types of univariate asymmetric GARCH models such as EGARCH, PGARCH and TGARCH models and multivariate GARCH models such as VECH, BEKK, CCC and DCC models to consider this time varying conditional variance-covariance relationship between different hydrologic variables. Moreover, different stationarity and nonlinearity tests are also applied in this study to test and compare different hydrologic and climatic variables and their variance-covariance structure.

The asymmetric GARCH models for streamflow heteroscedastict modeling indicate a better performance for exponential GARCH (EGARCH) model than the ARIMA models while other asymmetric models (PGARCH, TGARCH) did not show a better performance. However, it is also observed that the adding a GARCH model to the SARIMA model for rainfall time series modeling does not improve the accuracy of estimation, especially when the Box-Cox transformation is applied on rainfall time series. The univariate GARCH model for testing the volatility change of SOI shows a remarkable change in the short run persistency of the conditional variance of SOI and shows more extreme conditional variances in recent decades.

The diagonal VECH and CCC models adapted and developed to investigate the effect of the variance of rainfall on the streamflow show that rainfall has a strong conditional variance while runoff shows a short run conditional variance. The covariance between rainfall and runoff shows a long run characteristic and a high degree of nonlinearity. This characteristic may be due to the effect of physical catchment features on rainfall-runoff process. It seems that the CCC model which assumes a constant rainfall-runoff correlation is not valid for rainfall-runoff process. It is also observed that the MGARCH(1,1) model is sufficient for conditional variance-covariance modeling comparing to higher order models, i.e MGARCH(2,2) model.

The advantage of developing the MGARCH approach for drought analysis is also investigated in this research. Drought is a climate phenomenon usually related to large atmospheric circulations. The diagonal VECH and BEKK approaches showed that the

covariance structure between drought and atmospheric oscillations (NAO and SOI) is not strong and mostly related to the cross products of shocks rather than the covariances at the previous time steps. The time varying conditional correlation between drought and atmospheric indices do not show a significant change and trend during 1954-2010.

The MGARCH approach is also adapted for modeling the variance-covariance structure between temperature and output of GCM models which are applied for downscaling. The diagonal VECH and DCC model indicate short run persistence between GCM predictors and temperature time series. Except some GCMs such as specific humidity and 2m temperature, which have a strong covariance association with maximum and minimum temperature, other GCMs do not influence the variance of temperature data. The conditional correlation between GCMs and temperature time series do not show a significant upward or downward trend during 1980 to 2000.

In the field of social and public health and medical treatment, hip fracture is assumed to be largely related to different climate conditions. Adapting the CCC MGARCH method in the present study show a high impact of severe weather condition on hip fracture rate in Montreal region. It is observed that the snow depth, minimum temperature and day length are the most effective weather factors on hip fracture. It can be observed that the association between hip fracture incidence and climate variables is very weak or linear for small numbers of hip fracture incidences while this association (climate effect on hip fracture rate) increases exponentially and in a nonlinear fashion for the higher hip fracture rate values and harsh weather conditions.

This research also shows that the hydrologic and climatologic variables exhibit nonlinear temporal variation which the MGARCH model seems to be an interesting approach to be developed, investigated and applied in order to capture this nonlinear characteristic of hydrologic and climatic variables. We can see that daily time series show a higher degree of nonlinearity and the rainfall-runoff process indicates the highest nonlinearity among all hydroclimatic process in this study. In addition, the conditional variance-covariance structures show stationarity for all process. However, some trend nonstationarity is observed for some time series such as temperature and their association to other variables.

Finally, the proposed methods in this study give us the opportunity to have a closer look at the time varying second order moment of different hydrologic and climatic variables and to develop our understanding of their relationship. However, the univariate GARCH models show both advantage and disadvantage over univariate linear models such as ARIMA and SARIMA models. A high number of parameters also remains the main disadvantage of multivariate GARCH models.

Foreword

This thesis presents the research conducted during my doctoral studies. The structure of this thesis follows the standard structure of INRS-ETE theses. The first part of the thesis includes a general summary of the work performed. The summary aims to review succinctly the main results obtained and discuss their significance. The second part of the thesis contains seven articles, published (2) or submitted (5).

Articles and authors contributions

1- **Modarres, R.,** Ouarda, TBMJ. 2012. Generalized autoregressive conditional heteroscedasticity modeling of hydrologic time series. ACCEPED for publication in **Hydrological Processes.**

2- **Modarres, R.,** Ouarda, TBMJ. 2013. Modeling heteroscedasticty of streamflow time series. **Hydrological Sciences Journal, 58(1), 54-64.**

3- **Modarres, R.,** Ouarda, TBMJ. 2013. Testing and modeling the volatility change of SOI. To be PREPARED and SUBMITTED to **Climatic Change.**

4- **Modarres, R.,** Ouarda, TBMJ. 2013. Modeling rainfall-runoff relationship using multivariate GARCH model. To be SUBMITTED to **Journal of Hydrology.**

5- Modarres, R., Ouarda, TBMJ., Vanasse, A., Orzanco, MG., Gosselin, P. 2013. Modeling climate effects on hip fracture rate by multivariate GARCH model in Montreal region, Canada. To be SUBMITTED to **International Journal of Biometeorology.**

6- **Modarres, R.,** Ouarda, TBMJ. 2013. A Generalized Conditional Heteroscedastic model for temperature downscaling. SUBMITTED to **Journal of Climate.**

7- **Modarres, R.,** Ouarda, TBMJ. 2013. Modeling the relationship between climate oscillations and drought by a multivariate GARCH model. SUBMITTED to **Journal of Hydrometeorology.**

In the first article, R. Modarres presented a comparison between univariate linear (SARIMA) and nonlinear (SARIMA-GARCH) approaches for rainfall time series modeling. R. Modarres and T.B.M. J. Ouarda have discussed on the work. T.B.M. J Ouarda revised the article.

In the second article, R. Modarres proposed asymmetric GARCH models for modeling heteroscedasticity of streamaflow. This work was carried out based discussions between R. Modarres and T.B.M. J. Ouarda who also assisted to revise the manuscript.

In the third article, R. Modarres investigated the conditional variance of the southern Oscillation Index (SOI) and its change in recent decades in the context of climate change. This work was done with the assist of T.B.M. J. Ouarda and his revision and comments on the manuscript.

In the forth paper, R. Modarres developed a multivariate nonlinear model for investigating the heteroscedasticity of the rainfall-runoff relationship using the VECH model. R. Modarres carried out the analysis and wrote the paper. T.B.M. J. Ouarda assisted with comments and revised the article.

In the fifth article, the multivariate model was applied to model the link between climate variables and hip fracture in over Montreal region by R. Modarres. Alain Vanasse, Maria Gabriela Orzanco and Pierre Gosselin provided the data and discussion on the epidemiological aspects of the results. T.B.M. J. Ouarda assisted to design the procedure and revised the manuscript.

In the sixth article, the variance-covariance structure of CGCM and temperature time series was estimated and discussed. R. Modarres and T.B.M. J Ouarda had discussions throughout the work and T.B.M. J. Ouarda revised the manuscript.

In the seventh article, R. Modarres addressed the nonlinear association of two atmospheric circulations and standardized precipitation index using the VECH and BEKK models. The meetings and discussion between R. Modarres and T.B.M. J Ouarda were also held to do this part of research. T.B.M. J. Ouarda also assisted to revise the manuscript.

Dedication

This thesis is dedicated to my parents, who provided me the chance to pursue my higher educations, to my wife, Shiva, who gracefully supported and encouraged me during my PhD and my sister and brother, Roya and Hassan.

Acknowledgements

I would like to show my deepest gratitude to my supervisor Prof. Taha B.M.J. Ouarda for giving me the opportunity to do this PhD-research and giving me the freedom to try out new approaches in hydrology. I take this opportunity to express my sincere appreciation to Prof. Ouarda for his critical yet constructive comments and unending source of support, feedback and advice.

I owe my thankfulness to Hydroclimate modeling Group and my friends for their support and friendship during my PhD study at INRS-ETE. I am also grateful to the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Canada Research Chair (CRC) Program who provided the financial support for this research. My thanks also go to the examination committee members for reviewing my thesis and their helpful comments.

Table of Contents

List of Tables

List of Figures

PART I: THESIS SUMMARY

1. Introduction

1.1. Présentation de la thèse

Au cours des dernières décennies, il y a eu un intérêt croissant pour l'étude, le développement et l'application de modèles de séries chronologiques en hydrologie, en gestion les ressources hydrologie et en sciences du climat et de l'environnement. La modélisation de la dépendance des termes des séries temporelles hydrologiques a commencé à partir de début 1960, parmi les travaux pertinents, on trouve Thomas et Fiering (1962) et Matalas (1967).

Après ces deux travaux, l'utilisation des modèles de séries chronologiques a connu une attention considérable dans l'analyse et la prévision hydrologique. De nombreux chercheurs ont tenté de développer ces modèles pour différentes variables hydrologiques et de développer des méthodes alternatives pour estimer les paramètres des modèles et de sélectionner des tests pour la modélisation de séries chronologiques (Salas, 1993). Bien que les modèles de séries temporelles univariées soient très communs en hydrologie, la modélisation par les approches multidimensionnelles des séries temporelles n'ont pas été assez développées.

Les méthodes univariées et multivariées appliquées à la modélisation des séries chronologiques hydrologiques appartiennent généralement à la classe des approches qui tiennent compte des parties linéaires des processus hydrologiques.

Cependant, la plupart des processus régissant l'évolution temporelle des variables hydrologiques comprennent principalement des composantes non-linéaires. Les méthodes actuelles de modélisation des séries temporelles ne sont pas suffisantes pour capter cette non-linéarité. La relation temporelle linéaire entre les variables hydrologiques et

climatiques n'a pas été modelisée par les approches non linéaires multivariées temporelles malgré qu'il s'agisse d'un sujet approprié.

L'un des modèles non linéaire multivarié populaire est utilisé dans la modélisation de séries chronologiques financières qui estime le moment du second ordre ou la variance des variables financières. Cette étude tente d'introduire et de développer ce type de modèle dans le domaine de l'hydrologie et de la climatologie.

1.2. Problématique et l'hypothèse

Dans la plupart des modèles de séries temporelles hydrologiques et climatiques, la moyenne conditionnelle des variables hydrologiques est prise en considération.

Cela implique que l'hypothèse principale de ces modèles est que la variation temporelle d'une variable est régie par un processus linéaire. Cependant, de nombreuses études récentes ont mis en évidence l'existence de la non-linéarité et la non-stationnarité des processus régissant la variation temporelle et spatiale des variables hydrologiques.

Les modèles des séries temporelles actuel utilisés en hydrologie ignorent la variation temporelle des moments plus élevés des variables hydrologiques telles que la variance et la covariance.

Par conséquent, la première hypothèse de cette étude est que le second moment ou la variance des variables climatiques et hydrologiques varie dans le temps et la connection entre les varainces des variables hydrologiques et climatiques, ou la covariance, varie aussi dans le temps.

On suppose également que cette covariance est forte et a une mémoire à long terme dans le temps. En outre, on suppose que la variance-covariance conditionelle des variables hydrologiques et climatiques est fortement non linéaire et non stationnaire dans le temps.

Par conséquent, il est important d'appliquer des modèles alternatifs qui peuvent capter la variation du moment de second ordre des variables hydrologiques et climatiques et leur relation à travers la structure de covariance.

Il est également important de tester la non-linéarité et la non stationnarité des variables hydrologiques et climatiques et leur association à travers la variation temporelle de la variance-covariance. Ces tests sont très utiles pour observer l'influence du changement climatique sur l'association des différentes variables hydrologiques et climatiques dans le temps.

1.3. Les objectifs

Les objectifs de la présente recherche sont les suivants : Le premier consiste à introduire une classe de modèle non linéaire de séries multivariées temporelles, appelée «Autorégressive Conditionnelle Hétéroscédastique Généralisée Multivariée, ACHGM», couramment utilisée dans la modélisation des séries temporelles financières pour être adaptée et développée dans le domaine de l'hydrologie, des ressources de l'eau et des sciences du climat et de tout autre domaine connexe ou la relation multivariée temporelle entre les moments de second ordre est intéressante.

Le second objectif est de tester la non-linéarité et la non-stationnarité des variables hydrologiques et climatiques et de leur structure de variance-covariance développées par les modèles Autorégressifs Conditionnels Hétéroscédastiques Généralisés (ACHG) Multivariés.

Plus précisément, cette étude cherche à atteindre les objectifs suivants:

- Proposer, adapter et développer des modèles ACHG multivariés pour la modélisation du comportement non linéaire des variables hydrologiques et climatiques afin de poursuivre la première et la deuxième hypothèse de cette étude .
- Tester la non-linéarité et la non-stationnarité des variables hydrologiques et climatiques et leur association afin de poursuivre la troisième hypothèse de cette recherche
- étudier les avantages / désavantages des modèles ACHGM

1.4. L'organisation du document

Outre l'introduction, la partie I de la thèse est composé des sections suivantes. La prochaine section présentera l'état de l'art (revue de littérature) de la modélisation des séries temporelles hydrologiques. La méthodologie proposée et ses détails sont donnés dans la section 3. Les résultats et la discussion sont présentés dans la section 4. Les conclusions de cette thèse et des recommandations pour les travaux futurs sont présentés dans la section 5. Finalement, les articles issus de cette recherche sont présentés dans la partie II de la thèse.

2. Les Séries chronologiques hydrologiques: Revue de littérature

Dans les dernières décennies, les approches de modélisation des séries temporelles ont été principalement limitées à des méthodes connues en matière de modélisation

hydrologique. Les concepts des approches de la modélisation de séries temporelles ont été couramment utilisés dans la modélisation des séries chronologiques financières et économiques avant 1970. Cependant, l'application des modèles des séries chronologiques à l'analyse et la à modélisation hydrologiques a commencé à partir des années 1970. Parmi les premiers travaux de modélisation de la variation dans le temps des données hydrologiques, on peut citer les travaux de Quimpo (1971) et Klemes (1973) sur la modélisation stochastique et l'étude de Spoila et Chander (1974) sur la modélisation de l'écoulement de l'eau de surface avec un modèle autorégressif à moyenne mobile (ARMM).

Cependant, depuis la publication du livre de Box et Jenkins (1970) sur la modélisation et la prévision des séries temporelles, le nombre d'études sur la modélisation hydrologique des séries chronologiques a augmenté rapidement. Leur méthode était appelée la méthode Box-Jenkins de la modélisation des séries chronologiques et est rapidement devenue populaire chez les hydrologues.

La dépendance entre les séries temporelles est l'une des caractéristiques les plus importantes des variables hydrologiques qui a été mentionnée par Matalas (1967) en étudiant les données hydrologiques mensuelles et annuelles. La variation temporelle des caractéristiques statistiques des données des débits telles que la moyenne et la variance a été étudiée par Moreau et Pyatt (1970) à des échelles de temps hebdomadaires et mensuelles de Haw River en Caroline du Nord.

L'autocorrélation des débits mensuels liés aux précipitations a été supposée linéaireet les premiers coefficients d'autocorrélation variaient de manière saisonnière (e.g Moss et Bryson (1974)). Dans une autre étude, McKerchar et Delleur (1974) illustrent l'avantage

de modèles de séries chronologiques linéaires pour la prévision et la génération de la séquence d'un ensemble de données synthétiques.

Ils ont décrit l'application du modèle autorégressif à moyenne mobile intégrée (ARMMI) saisonnière pour modéliser le débit mensuel. Le travail de McKerchar et Delleur (1974) était une des plus importantes études qui illustrent la justesse de la méthodologie Box et Jenkins pour analyser et faire des prévisions dans le domaine de l'hydrologie et des ressources en eau.

La méthodologie de Box et Jenkins a été révisée et appliquée dans les travaux de Hipel et al., (1977) et McLeod et al., (1977) et l'avantage du modèle ARMMI saisonnier a été illustré dans la modélisation temporelles des séries des débits de la rivière , à Ogdensburg, New York. Le modèle linéaire ARMMI a été appliqué dans de nombreuses autres études. Rao et al.(1982) ont discuté de la procédure de la modélisation et les avantages apportés pour les variables hydrologiques et climatiques. La description de la méthode optimale pour l'estimation des paramètres des modèles linéaires et l'identification du meilleur modèle ont été examinées par Salas et al. (1982). D'autres chercheurs tels que Onof et. al., (1996) et Montanari et al. (2000) ont appliqué des modèles linéaires aux séries des précipitations et des séries de débits. De plus, ils ont montré l'application du modèle linéaire à la modélisation et à la prévision des séries chronologiques hydrologiques.

Benyahya et al. (2007) ont également utilisé des modèles linéaires des séries chronologiques pour modélisation de la température de l'eau.

Toutefois, ces dernières années, la modélisation du processus non linéaire en hydrologie et en climatologie a reçu une attention considérable. Bien que l'on savait déjà que la plupart des variables hydrologiques sont régies par un processus non linéaire, la complexité de ces processus à être capturée par des méthodes mathématiques et statistiques.

Les différents mécanismes non linéaires qui agissent dans les processus hydrologiques à différentes échelles spatiales et temporelles qui ont reçu beaucoup d'attention au cours des dernières deux décennies (par exemple, Rao et Yu, 1990; Chen et Rao, 2003).

À l'aide du test de Hinich bispectral, on peut étudier les caractéristiques de la linéarité et de la normalité des séries chronologiques des différentes variables hydro-climatologiques. Rao et Yu (1990) ont détecté la non-linéarité dans les séries météorologiques quotidiennes dans le nord-est des États-Unis. Les résultats de l'étude de Chen et Rao (2003) ont également indiqué la non-linéarité de la série des débits mensuels à la mi-ouest des États-Unis.

En dépit de tous les intérêts et les progrès dans la recherche sur les caractéristiques non linéaires des processus hydrologiques. Il n' y a peu d'études sur la nature de la non-linéarité. Bien que certaines méthodes de contrôle de la linéarité ou non-linéarité dans les séries chronologiques sont disponibles (Wang et al. 2006), il existe aussi peu d'études sur l'utilisation des approches qui sont capables de modéliser la non-linéarité des variables hydrologiques et climatiques.

L'une des approches les plus intéressantes pour la modélisation de la non-linéarité des séries chronologiques est utilisée dans la modélisation de séries temporelles financières et de l'économétrie. Cette approche qui traite de la modélisation de la variance, ou le

moment du second ordre d'une variable et son évolution temporelle est appelée «Autorégressive Conditionnelle Hétéroscédastique Généralisée» ou modèle «ACHG» et a été introduie par Engle (1982) et développée par Bollerslev (1986).

Le modèle ACHG suppose que la variance d'une variable évolue dans le temps et sa nouvelle valeur dans la nouvelle étape est liée à la variance dans des étapes précédentes.

Ce type d'écart variant dans le temps est appelé «variance conditionnelle» ou «hétéroscédasticité». ».

Le modèle (univariée) ACHG a été utilisé dans le domaine de l'hydrologie et de la climatologie par Wang et al. (2005), Romilly (2005) et Chen et al., (2008). Wang et al. al., (2005) ont appliqué le modèle ACHG pour les séries débits quotidiens et mensuels de la rivière Jaune supérieure à Tangnaihai, en Chine.

Ils ont conclu que les modèles traditionnels des séries chronologiques linéaires, le modèle autorégressif et le modèle autorégressif avec moyennes mobiles désaisonnalisées, ne sont pas suffisants pour décrire la variance en fonction du temps des débits et un modèle ACHG doit être appliqué sur les résidus d'un modèle ARMM pour capturer la variation dans le temps du le comportement de la variance des débits.

Chen et al., (2008) ont appliqué les modèles ARMM linéaire et non-linéaire à des approches ACH modèlisant les débits de moyenne mobile de 10 jours de la rivière Wu-Shi en Taiwan et a vérifié que les modèles non linéaires de séries temporelles sont bien adaptés par rapport aux approches traditionnelles linéaires comme le modèle ARMM.

En comparant le modèle ARMM et ACH, Chen et al., (2008) ont signalé une augmentation du coefficient d'efficacité (CE) de 0,28 pour le modèle ARMM et de 0,76

pour le modèle ACH tandis que l'erreur absolue moyenne (EAM) se réduit de 60,45 m^3/s pour le modèle ARMM à 41,35 m^3/s pour le modèle ACH.

Dans une autre application du modèle ACHG, Romilly (2005) a appliqué le modèle ARMMI saisonnier à la température mondiale moyenne mensuelle et a signalé l'existence d'hétéroscédasticité dans les résidus du modèle. Cette étude a montré que le modèle ARMMI saisonnier a une performance inférieure à celle du modèle ACHG pour la modélisation temporelle de la température moyenne globale.

Malgré ces quelques applications d'un modèle ACHG univarié, à l'hydrologie et des avantages de ce modèle mentionnés par ces études, il est clair que dans le cas de la plupart des autres variables hydroclimatiques telles que les précipitations, l'évaporation, la vitesse du vent, l'humidité, etc, le modèle ACHG n'a pas été appliqué et les avantages / inconvénients du modèle ACHG n'ont pas été abordés clairement.

Comme mentionné précédemment, les variables hydrologiques et climatiques sont corrélées les unes aux autres dans l'espace et dans le temps. Par conséquent, on s'attend à supposer que les caractéristiques statistiques des variables hydroclimatiques varient dans le temps et dans l'espace par rapport aux caractéristiques statistiques des autres variables. La variation temporelle de cette relation peut être modélisée par un modèle de série chronologique multivariée. Bien que le nombre d'études sur les techniques d'analyse multivariée en matière d'hydrologie et de climatologie a augmenté très rapidement, les modèles de séries temporelles multivariées n'ont pas été appliqués à une grande échelle.

Peu d'études ont appliqué les modèles de séries chronologiques multivariées dans le domaine de l'hydrologie, les ressources en eaux et la climatologie. Parmi ces études, on

trouve le travail effectué par Cooper et Wood (1982) sur la modélisation séries chronologiques de pluie-débit.

Ledolter (1978) a discuté les modèles stochastiques pour les processus vectoriels et a donné l'exemple d'un modèle ARMM multivarié pour l'estimation des débits. Salas et al. al., (1985) ont fourni une discussion et ont suggéré des approches alternatives pour des modèles de séries chronologiques multivariées pour les systèmes des ressources en eau. Ils ont indiqué les propriétés d'une approche de modélisation ARMM multivariée et l'un des modèles de sous-ensembles intéressants d'un modèle ARMM multivarié appelé le modèle contemporain ARMM (ARMMC). Le modèle ARMMC peut être utilisé pour la modélisation de deux ou plusieurs séries temporelles qui sont statistiquement liées les unes aux autres simultanément. Autrement dit, il n'y a pas d'interactions temporelles entre les variables. Cela rendra le modèle plus simple que le modèle ARMM multivarié complet. L'application du modèle ARMMC en hydrologie a été réalisée par Camacho et al. (1987), Bartolini et al. (1988) et Rasmussen et al. (1996). Toutes ces études ont discuté des avantages des modèles de séries temporelles multivariés, mais ils supposent toujours la linéarité des processus hydrologiques et climatiques. Cette hypothèse n'est pas vraie pour la plupart des variables hydrologiques et climatiques qui évoluent dans l'espace et dans le temps à travers un processus non linéaire.

Les approches multivariées non linéaires de modélisation des séries temporelles qui peuvent capter ce processus non linéaire n'ont pas été appliquées en hydrologie et en climatologie encore. D'autre part, les modèles multivariés non-linéaires sont très populaires dans la modélisation des séries chronologiques financières et économiques. La variation non linéaire des données financières a été reconnue et traitée pendant une

longue période. Pour la modélisation de séries temporelles financières, il est supposé que le moment du second ordre, ou la variance, des différents rendements des actifs financiers des différents marchés sont corrélés les uns aux autres. Par conséquent, les modèles non linéaires de séries temporelles sont très populaires en finance. Il est maintenant largement admis que les volatilités financières, ou des groupes de variances élevées ou basses, se déplacent ensemble dans le temps entre actifs et marchés. Il est reconnu qu'un cadre de modélisation multivariée conduit à une modélisation plus pertinente et la prévision des changements dans la volatilité, ou la variance, des séries temporelles financières. L'approche la plus intéressante utilisée pour la modélisation multivariée de la variance (volatilité) et la covariance (co-volatilité) entre les différents marchés est le modèle ACHG multivarié (Bauwens et. al., 2006).

Les concepts de la modélisation ACHG univarié des actifs des marchés ont d'abord été introduits par Engle (1982) et développés par Bollerslev (1986). Depuis lors, un grand nombre d'études ont appliqué le modèle univarié et multivarié ACHG dans la modélisation des séries chronologiques financière et économique. Par exemple, Karolyi (1995) a appliqué un modèle ACHGM pour tester la dynamique à court terme des rendements et de la volatilité des titres négociés aux bourses de New York et de Toronto. Les co-mouvements des rendements boursiers entre deux marchés en Chine ont été examinés par Tsui et Yu (1999) en utilisant le modèle bivarié ACHGM. Kim (2000) a appliqué le modèle ACHGM afin de relier la volatilité des mois productions industrielles des États-Unis et du Japon et la volatilité des taux de change yen / dollar. Choudhry (2003) a présenté une étude empirique sur l'effet de la volatilité des marchés boursiers sur les dépenses des consommateurs américains en utilisant un modèle ACHG multivarié.

Fong et al. (2006) ont appliqué le modèle ACHGM pour examiner l'effet de trois actifs financiers, des taux de change du mark allemand, du dollar américain et du yen japonais les uns sur les autres. Plus récemment, Hakim et McAleer (2009) ont appliqué un modèle ACHG multivarié pour prévoir la corrélation conditionnelle entre les actifs financiers de trois catégories d'organisations internationales, à savoir les échanges boursiers, obligataires et à l'étranger. Ils ont trouvé des preuves de retombées de la volatilité et l'effet asymétrique de stocks négatifs et positifs sur la variance conditionnelle dans la plupart des paires de séries.

Toutes les études mentionnées ci-dessus et d'autres travaux ont appliqué et développé les modèles ACHG multivariés pour les séries chronologiques financières et aucun développement ou application de cette méthode multivariée ne peut être trouvé dans les autres domaines d'études, en particulier pour la modélisation de séries temporelles multivariées de la variance des variables hydrologiques et climatiques malgré le fait qu'il est généralement admis que les changements dans le temps peuvent être associés à la variance des autres variables dans l'espace et le temps.

Par conséquent, la nécessité d'adapter et de développer cette méthode en hydrologie et en climatologie est nécessaire, afin de clarifier la variation temporelle et spatiale de la variance et de covariance des variables hydrologiques et climatiques.

3. Méthodologie

3.1. La Modèle ACHG univarié

Le modèle univarié ACHG non linéaire appliqué à la modélisation des séries temporelles financières considère la volatilité ou la variance conditionnelle (σ_t) des rendements des actifs simples et tente de modéliser le processus de volatilité d'une variable, Y_t.

Dans ce cas, nous avons

$$Y_t = \sigma_t \varepsilon_t \qquad \varepsilon_t \sim \mathrm{N}(0,1) \qquad (1)$$

Où $\sigma_t^2 = \mathrm{E}(\mathrm{Y}_t^2 | F_{t-1})$ désigne la variance conditionnelle. Cette variance peut être notée comme suit par un modèle ACHG d'ordre V et M:

$$\sigma_t{}^2 = w + \sum_{i=1}^{V} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{M} \beta_j \sigma_{t-j}^2 \qquad (2)$$

La première partie du modèle ci-dessus a été introduite par Engle (1982) et est appelée Modèle Autorégressif Conditionnel Hétéroscédastique (ACH) d'ordre (V):

$$\sigma_t{}^2 = \omega + \sum_{i=1}^{V} \alpha_i Y_{t-i}^2 = \omega + \alpha_1 Y_{t-1}^2 + \alpha_2 Y_{t-2}^2 + \ldots + \alpha_V Y_{t-V}^2 \qquad (3)$$

Bollerslev (1986) a ajouté une variance conditionnelle retardée au modèle ACH, qui agit comme un terme de lissage et ce modèle est alors appelé le modèle ACH généralisée (ACHG) d'ordre (V, M).

En hydrologie, cependant, l'approche ACHG est généralement appliquée pour la modélisation de la variance conditionnelle ou l'hétéroscédasticité restant dans les résidus des modèles linéaires. Ce type de modèle est généralement désigné comme un modèle d'erreur ARMM-ACHG. Dans ce cas, si la moyenne conditionnelle d'une série temporelle

est modélisée par un modèle linéaire comme un processus ARMM, ARMMI ou SARMMI, l'ordre V de l'autorégressive conditionnelle hétéroscédastique pour la variance conditionnelle des résidus d'un modèle modèle linéaire, (σ_t) est définie comme suit :

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + ... + \alpha_V \varepsilon_{t-V}^2 = \omega + \sum_{i=1}^{V} \alpha_i \varepsilon_{t-i}^2 \qquad (4)$$

$$\varepsilon_t = \sigma_t e_t \qquad e_t \sim Normal(0,1) \qquad (5)$$

Ici, ε_t indique les résidus du modèle linéaire qui ne sont pas corrélés, mais qui ont des variances qui changent au fil du temps, e_t représente une variables aléatoire indépendante et identiquement distribuées (iid) de moyenne 0 et de variance 1, indépendante des réalisations passées ($e_{t-1}, e_{t-2}, \cdots$), $\alpha_1, ..., \alpha_V$ sont les paramètres du modèle ACH et ω est une constante (Wei, 2006). Dans ce modèle, la variance de l'erreur est variable dans le temps et dépend des V erreurs du passé ε_{t-1}^2, $\varepsilon_{t-2}^2, \ldots, \varepsilon_{t-V}^2$.

Le modèle Autorégressif Conditionnel Hétéroscédastique Généralisé, ACHG (V, M) est défini comme suit

$$\sigma_t^2 = \omega + \sum_{i=1}^{V} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{M} \beta_j \sigma_{t-j}^2 \qquad (6)$$

$$\varepsilon_t = \sigma_t e_t \qquad e_t \sim Normal(0,1) \qquad (7)$$

Où $\alpha_1, ..., \alpha_V$ et $\beta_1, ..., \beta_M$ sont les paramètres du modèle ACHG (V, M). Par conséquent, l'ARMM- ACHG (V, M) indique que la moyenne conditionnelle est décrite par un modèle ARMM (ou d'autres modèles linéaires tels que ARMMI ou ARMMI saisonnier), tandis que sa variance conditionnelle est décrite par un modèle ACHG (V, M).

La modélisation d'une variable hydrologique avec un modèle d'erreur ARMM-ACHG exige que nous approuvions l'existence de la variance conditionnelle dans les résidus des modèles linéaires ajustés sur la série chronologique hydrologique. Par conséquent, nous avons besoin d'un test pour vérifier si la variance des résidus est fonction du temps.

Similaire au test quant à la pertinence d'un modèle linéaire, Bollerslev (1986) a démontré que le fonction d'autocorrélation (FAC) des carrés des résidus standardisés (RSC, ci-après) est utile pour identifier la variance conditionnelle dans les résidus. Afin de vérifier si la variance des résidus est conditionnelle à son histoire passée ou, en d'autres termes, si les résidus présentent un effet ACH, nous pouvons appliquer le test de Ljung-Box et Engle du multiplicateur de Lagrange fréquemment utilisé, dans le cas des RSC.

Le test d'ajustement de Ljung-Box est calculé pour la série chronologique RSC (ε^2)

$$Q = N(N+2)\sum_{k=1}^{L}(N-k)^{-1}r_k^2(\varepsilon^2) \qquad (8)$$

Sous l'hypothèse nulle d'absence d'effet ACH dans les résidus, la statistique de test est asymptotiquement distribuée selon une loi du khi-carré.

En plus du FAC des RSC, le test de multiplicateur Engle et de Lagrange pour l'effet ACH proposé par Engle (1982) est également utilisé pour tester l'effet ACH. La statistique de test est donnée par NR2, où R est le coefficient de corrélation calculé à partir de plusieurs échantillons de la régression de ε_t^2 sur une constante et $\varepsilon_{t-1}^2,...,\varepsilon_{t-V}^2$ et N est la taille de l'échantillon. L'hypothèse nulle d'absence d'effet ACH est acceptée si la statistique de test est asymptotiquement distribuée comme une distribution de khi-carré avec des *k-v* degrés de liberté. Le test peut également être utilisé pour étudier l'effet ACHG de RSC (Bollerslev, 1986).

3.2. Les Différents types de modèle ACHG univarié

La formulation générale d'un modèle ACHG a été présentée dans la section précédente. Dans la modélisation de séries temporelles financières, il ya quelques variations du modèle ACHG qui ont été développées pour une meilleure modélisation et l'estimation de la volatilité des actifs et des rendements financiers. Les trois principales variations d'un modèle ACHG sont développées dans le domaine de l'hydrologie et la climatologie.

3.2.1. Le modèle Puissance ACHG (PACHG)

Le modèle PACHG (Ding, et al, 1993) peut être défini comme suit:

$$\sigma_t^{\delta} = \omega + \sum_{i=1}^{V} \alpha_i (|\varepsilon_{t-i}|)^{\delta} + \sum_{j=1}^{M} \beta_j \sigma_{t-j}^{\delta} \qquad (9)$$

Où $\delta > 0$. Le cas $\delta = 2$ correspond à la forme habituelle de modèle ACHG. La spécification d'un exposant variable δ a le mérite de permettre de modéliser avec une «mémoire longue» dans les chocs de la variance conditionnelle (voir Ding et al. 1993 pour une discussion générale).

3.2.2. Le modèle Seuil ACHG (SACHG)

Dans la modélisation de séries temporelles financières, il est supposé que les rendements positifs et négatifs ($\varepsilon_{t-i} > 0$, ou les bonnes nouvelles et $\varepsilon_{t-i} < 0$, ou les mauvaises nouvelles) ont des effets différents sur la variance conditionnelle. Pour examiner ce comportement asymétrique des rendements, le modèle SACHG (Zakoïan, 1994) est utilisé, qui peut être spécifié comme suit :

$$\sigma_t^{2} = \omega + \sum_{i=1}^{V} \alpha_i \varepsilon_{t-j}^{2} + \sum_{j=1}^{M} \beta_j \sigma_{t-j}^{2} + \sum_{k=1}^{r} \gamma_k \varepsilon_{t-k} I_{t-k}^{-} \qquad (10)$$

Où $I_{t-k}^{-}=1$ si $\varepsilon_t<0$ et 0 sinon. Ici, si $\gamma_i>0$, les mauvaises nouvelles augmentent la volatilité et si $\gamma\neq 0$, l'impact des nouvelles est asymétrique.

3.2.3. Le modèle Exponentielle ACHG (EACHG)

Le modèle EACHG a été proposé la première fois par Nelsen (1991) et est rapidement devenu populaire dans les applications financières comme la variance conditionnelle est une fonction exponentielle dans ce domaine et assure avoir une variance positive conditionnelle:

$$Log(\sigma_t^2)=\omega+\sum_{i=1}^{V}(\alpha_i\left|\frac{\varepsilon_{t-i}}{\sigma_{t-i}}\right|)+\sum_{j=1}^{M}\beta_j\log(\sigma_{t-j}^2) \qquad (11)$$

Nelson (1991) a décrit les contraintes de la forme générale du modèle ACHG comme la positivité de la persistance et la variance conditionnelle de chocs, ce qui peut être résolu en utilisant le logarithme de la variance dans le modèle EACHG. Les preuves de l'avantage d'un modèle EACHG par rapport a un modèle ACHG peuvent être trouvées dans Nelson (1991).

3.3. Les Modèles ACHG multivariés

L'extension multivariée des modèles ACH et ACHG peut être définie en principe de manière similaire à un vecteur ARMM (VARMM). Nous considérons d'abord le modèle ACH multivarié, puis le modèle ACHG multivarié.

Ayant une moyenne de zéro à K dimensions, le processus non corrélé en série $u_t=(u_{1t},...,u_{Kt})'$, peut être représenté par

$$u_t = \sum_{t|t-1}^{1/2} \varepsilon_t \qquad (12)$$

Où ε_t est un bruit blanc i.i.d. k-dimensionnel, $\varepsilon_t \sim i.i.d(0, I_K)$, et $\sum_{t|t-1}$ est la matrice de covariance conditionnelle de u_t , donné, u_{t-1}, u_{t-2} .... La $\sum_{t|t-1}^{1/2}$ est le symétrique positive racine carrée définie de $\sum_{t|t-1}$.

En définissant la distribution conditionnelle de u_t avec la forme suivante

$$\Omega_{t-1} = \{u_{t-1}, u_{t-2}, ...\} \qquad (13)$$

$$u_t \Big| \Omega_{t-1} \sim (0, \sum_{t|t-1}) \qquad (14)$$

Ils présentent une analyse multivariée ACH (q) si

$$VECH(\sum_{t|t-1}) = \gamma_0 + \Gamma_1 vech(u_{t-1}u'_{t-1}) + ... + \Gamma_v vech(u_{t-v}u'_{t-v}) \qquad (15)$$

Où VECH désigne l'opérateur demi-vectorisation qui empile les colonnes d'une matrice carrée de diagonale vers le bas à une vecteur, γ_0 est un vecteur de dimension $\frac{1}{2}K(K+1)$ de constantes et les Γ_j sont des matrices de coefficients de dimension $(\frac{1}{2}K(K+1) \times \frac{1}{2}K(K+1))$.

A titre d'exemple, un processus bivarié (K = 2) ACH(1) peut être défini de la manière suivante :

$$vech\begin{bmatrix} \sigma_{11,t|t-1} & \sigma_{12,t|t-1} \\ \sigma_{12,t|t-1} & \sigma_{22,t|t-1} \end{bmatrix} = \begin{bmatrix} \sigma_{11,t|t-1} \\ \sigma_{12,t|t-1} \\ \sigma_{22,t|t-1} \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \\ \gamma_{30} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{21} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{bmatrix} \begin{bmatrix} u^2_{1,t-1} \\ u_{1,t-1}u_{2,t-1} \\ u^2_{2,t-1} \end{bmatrix} \qquad (16)$$

Le modèle ACH multivarié peut-être tout simplement généralisé de la même manière que dans le cas univarié. Dans l'analyse ACHG multivariée, le modèle (ACHGM) a des matrices de covariance conditionnelles qui ont la forme suivante

$$VECH(\sum_{t|t-1}) = \gamma_0 + \sum_{i=1}^{v} \Gamma_j vech(u_{t-j}u'_{t-j}) + \sum_{j=1}^{m} G_j vech(\sum_{t-j|t-j-1}) \qquad (17)$$

Où les G_j sont également fixés avec $\frac{1}{2}K(K+1)\times\frac{1}{2}K(K+1))$ matrices de coefficients.

Par exemple, un modèle bivarié ACHG (1,1) est

(18)

Cette forme générale du modèle ACHGM est appelée le modèle VECH complet et dispose de 21 paramètres à estimer dans le cas de deux variables (k = 2) modélisées VECH (1,1).

3.4. Les Différents types de modèle ACGHM

Il est clair que, même pour un modèle ACHGM simple pour un cas bivarié, un grand nombre de paramètres doivent être estimés, ce qui le rend difficile à manipuler. En particulier, les implications d'un modèle général de ce type de relation entre les variables et leurs propriétés de moments plus élevés ne sont pas évidentes. Par conséquent, les modèles plus restreints doivent être proposés. Certains de ces modèles ACHGM sont présentés dans les sections suivantes.

3.4.1. Le modèle VECH Diagonale

Bollerslev et al., (1988) ont considéré comme des processus ACH où les matrices Γ_j sont toutes diagonales. Dans le cas du premier ordre, le modèle est de la forme

$$diagonal\ vech\begin{bmatrix}\sigma_{11,t|t-1} & \\ \sigma_{12,t|t-1} & \sigma_{22,t|t-1}\end{bmatrix}=\begin{bmatrix}w_{11} & \\ w_{21} & w_{22}\end{bmatrix}+\begin{bmatrix}\gamma_{11} & \\ \gamma_{21} & \gamma_{22}\end{bmatrix}\begin{bmatrix}u_{1,t-1}^2 & \\ u_{1,t-1}u_{2,t-1} & u_{2,t-1}^2\end{bmatrix}$$

(19)

De la même façon, nous pouvons écrire le modèle ACHG diagonal comme suit

$$diagonal\ VECH\begin{bmatrix}\sigma_{11,t|t-1} & \\ \sigma_{12,t|t-1} & \sigma_{22,t|t-1}\end{bmatrix}=\begin{bmatrix}w_{11} & \\ w_{21} & w_{22}\end{bmatrix}+\begin{bmatrix}a_{11} & \\ a_{21} & a_{22}\end{bmatrix}\begin{bmatrix}u_{1,t-1}^2 & \\ u_{1,t-1}u_{2,t-1} & u_{2,t-1}^2\end{bmatrix} \quad (20)$$
$$+\begin{bmatrix}g_{11} & \\ g_{21} & g_{21}\end{bmatrix}\begin{bmatrix}\sigma_{11,t-1|t-2} & \\ \sigma_{21,t-1|t-2} & \sigma_{22,t-1|t-2}\end{bmatrix}$$

3.4.2. Les modèles BEKK et BEKK diagonal

L'un des problèmes liés au modèle ACHGM est que les paramètres doivent être tels que les matrices de covariance conditionnelles, $\sum_{t|t-1}$ sont définies positives. Afin de garantir cette propriété, Engle et Kroner (1995) ont étudié la variante suivante du modèle ACH multivarié à la suite des travaux de Baba et al (1990)

$$\sum_{t|t-1} = \Gamma_0^* + \Gamma_1^{*'} u_{t-1}u_{t-1}'\Gamma_1^* + \ldots + \Gamma_q^{*'} u_{t-q}u_{t-q}'\Gamma_q^* \qquad (21)$$

Avec les Γ_j^* qui sont chacun des matrices ($K \times K$). Ici, les $\sum_{t|t-1}$ sont définies positives si Γ_0^* a cette propriété qui peut être exécutée en l'inscrivant dans une forme de produit, $\Gamma_0^* = C_0^{*'} C_0^*$ avec une matrice triangulaire.

L'avantage de ce modèle est qu'il est relativement parcimonieux. Par exemple, pour un processus bivarié avec K = 2 et q = 1, il n'y a que sept paramètres, tandis que le modèle ACH pleinement développé a 12 coefficients. En outre, contrairement au modèle en diagonale, il peut produire des interactions très riches entre les moments conditionnels du second ordre.

Alternativement, le modèle BEKK-ACHG complet peut être écrit comme ceci

$$\sum_{t|t-1} = C_0^{*\prime} C_0^* + \sum_{n=1}^{N}\sum_{j=1}^{q} \Gamma_{jn}^{*\prime} u_{t-j} u_{t-1} u_{t-j}' \Gamma_{jn}^* + \sum_{n=1}^{N}\sum_{j=1}^{m} G_{jn}^{*\prime} \sum_{t-j|t-j-1} G_{jn}^* \qquad (22)$$

Où C_0^* est une matrice triangulaire ($K \times K$) dimension et les matrices de coefficients Γ_{jn}^* et G_{jn}^* ont également de dimension ($K \times K$).

Par conséquent, le modèle BEKK (1,1) peut s'écrire sous forme matricielle comme suit

$$\begin{bmatrix} \sigma_{11,t|t-1} & \sigma_{12,t|t-1} \\ \sigma_{12,t|t-1} & \sigma_{22,t|t-1} \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \\ w_{12} & w_{22} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}' \begin{bmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}\varepsilon_{1,t-1} & \varepsilon_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}$$
$$+ \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}' \begin{bmatrix} \sigma_{11,t-1|t-2} & \sigma_{12,t-1|t-2} \\ \sigma_{12,t-1|t-2} & \sigma_{22,t-1|t-2} \end{bmatrix} \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \qquad (23)$$

Ce modèle a 12 paramètres à estimer, ce qui est encore relativement élevé. Une autre spécification du modèle BEKK est le modèle diagonal ci-dessous:

$$\begin{bmatrix} \sigma_{11,t|t-1} & \sigma_{12,t|t-1} \\ & \sigma_{22,t|t-1} \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \\ & w_{22} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \\ & \gamma_{22} \end{bmatrix}' \begin{bmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ & \varepsilon_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} \gamma_{11} & \\ & \gamma_{22} \end{bmatrix}$$
$$+ \begin{bmatrix} g_{11} & \\ & g_{22} \end{bmatrix}' \begin{bmatrix} \sigma_{11,t-1|t-2} & \sigma_{12,t-1|t-2} \\ & \sigma_{22,t-1|t-2} \end{bmatrix} \begin{bmatrix} g_{11} & \\ & g_{22} \end{bmatrix} \qquad (24)$$

Ce modèle a seulement sept paramètres à estimer et peut donner l'équation suivante pour la variance-covariance conditionnelle entre deux variables (Baur, 2006)

$$\sigma_{11,t} = w_{11} + a_{11}^2 \varepsilon_{1,t-1}^2 + g_{11}^2 \sigma_{1,t-1} \quad (25)$$

$$\sigma_{21,t} = w_{11} + a_{11} a_{22} \varepsilon_{1,t-1} + g_{11} g_{22} \sigma_{21,t-1} \quad (26)$$

$$\sigma_{2,t} = w_{11} + a_{22}^2 \varepsilon_{1,t-1}^2 + g_{22}^2 \sigma_{1,t-1} \quad (27)$$

3.4.3. Le modèle Constant et dynamique de la Corrélation conditionnelle

Bollerslev (1990) a proposé un modèle de classe ACHGM dans lequel les corrélations conditionnelles sont constantes et donc les covariances conditionnelles sont proportionnelles au produit de l'écart type correspondant. Cette restriction réduit considérablement le nombre de paramètres inconnus et simplifie l'estimation.

La variance conditionnelle dans le modèle CCC est définie comme

$$\sigma_t = D_t R D_t = (\rho_{ij} \sqrt{\sigma_{iit} \sigma_{jjt}}) \quad (28)$$

Où

$$D_t = diag(\sigma_{11t}^{1/2} \ldots \sigma_{NNt}^{1/2}) \quad (29)$$

σ_{iit} peut être défini comme un modèle ACHG univarié et

$$R = (\rho_{ij}) \quad (30)$$

est une matrice symétrique définie positive avec $\rho_{ii} = 1$, $\forall i$

R est une matrice contenant les corrélations constantes conditionnelles

Par exemple, le modèle ACHG (1,1) pour spécifier chaque variance conditionnelle en D_t est

$$\sigma_{iit} = \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i \sigma_{ii,t-1} \quad (31)$$

Le modèle CCC dispose d'un total $k(k+5)/2=12$ de paramètres.

Le principal inconvénient du modèle CCC est l'hypothèse de corrélation invariante qui n'est pas une hypothèse valable pour les variables hydrologiques et climatiques, en particulier dans le contexte des changements et des variations climatiques. Pour rendre la matrice de corrélation conditionnelle en fonction du temps, Engle (2002) a proposé une corrélation conditionnelle dynamique (CCD) où le modèle, basé sur l'équation (28) et ayant une corrélation qui varie dans le temps, la covariance conditionnelle peut être écrite comme suit

$$\sigma_t = (1-\theta_1-\theta_2)\bar{\sigma} + \theta_1 \varepsilon_{t-1}\varepsilon_{t-2} + \theta_2 \sigma_{t-1} \qquad (32)$$

Où θ_1 et θ_2 sont des paramètres scalaires pour capturer les effets des chocs précédents et les corrélations dynamiques précédentes conditionnelles à la corrélation conditionnelle dynamique actuelle et de satisfaire $\theta_1+\theta_2<1$ce qui implique que $\sigma_t>0$ et $\bar{\sigma}$ est la matrice de variance inconditionnelle de dimension $k\times k$. Lorsque $\theta_1=\theta_2=0$, l'Eq. (32) est équivalente au modèle CCC.

3.5. Les tests de non-stationnarité et la non-linéarité

Dans la plupart des modèles hydrologiques, la stationnarité et la non-linéarité sont les hypothèses de base . Il est donc nécessaire de tester la stationnarité et la non-linéarité des variables hydrologiques. Dans cette étude, une série de tests de stationnarité et un test non-linéarité robustes sont appliquées pour vérifier ces hypothèses pour les variables hydrologiques et climatiques et leur structure de variance-covariance, de même que leur association (corrélation). Ces méthodes pour tester la stationnarité comprennent le test de

Dickey-Fuller augmenté (ADF) (Dickey et Fuller (1979), les tests KPSS (Kwiatkowski et al., 1992) et PP (Phillips et Perron, 1988) et des tests de racine unitaire. Par ailleurs, le test BDS (Brock et al., 1996) est utilisé pour tester la non-linéarité des variables hydrologiques.

3.6. La sélection des données et des processus hydroclimatiques

Afin de démontrer les procédures ACHG univariée et multivariés, différentes variables et processus importantes des hydrologiques et climatiques ont été prises en compte dans cette étude. Les variables hydro-climatiques prises en compte dans cette étude sont présentées dans le tableau 1 pour les procédures ACHG à la fois univariée et multivariée. Ces variables sont sélectionnées pour examiner la structure de la variance-covariance conditionnelle des différents sujets climatiques et hydrologiques. Par conséquent, la sélection des variables permet à différents modèles qui seront développés pour différentes échelles temporelles telles que l'échelle journalière et mensuelle de temps.

Par exemple, il convient de noter que les modèles ACHG n'ont pas été développés avant pour les séries chronologiques des précipitations (mensuelles). En outre, les modèles asymétriques sont conçus pour les séries chronologiques des débits pour la première fois d'étudier les caractéristiques asymétriques d'hétéroscédasticité des séries chronologiques des débits. La variance conditionnelle de l'indice d'oscillation australe (IOA) comme une variable de déclenchement pour le système hydroclimatique mondiale est également important d'être évalué.

En outre, les différents processus populaires et importants à des quotidiens et mensuelles échelles de temps ont été choisis pour étudier la variation et la résistance de la structure de covariance conditionnelle en hydrologie et en climatologie.

Les relations précipitations-débit est le processus hydroclimatiques le plus intéressant et important dans la gestion des ressources en eau. L'anlayse de la sécheresse et de sa relation avec la circulation atmosphérique est très important pour atténuer et gérer la sécheresse. En outre, la réduction d'échelle de température est l'un des quest importants pour l'analyse du climat dans le contexte du changement climatique. La dernière partie de l'élaboration de modèles de ACHGM est consacrée aux problèmes de santé qui est important dans le contexte du changement climatique sur la santé publique. Différents modèles de ACHGM comme diagonale VECH, BEKK diagonale, CCC et DCC, sont donc élaborés pour chaque processus selon les caractéristiques de chaque processus.

Les modèles de ACHGM lient la covariance entre les différents variables hydrologique et climatiques dans les différents processus qui sont étudiés dans cette recherche.

Tableau 1 les Données et les processus hydrologiques et climatiques de l'étude

Panneau A : études de cas univariés			
Variables/ processus	***L'échelle de temps***	***l'emplacement***	***le modèle***
Série chronologiques des précipitations	Mensuel	les stations de Campsie (Canada) et Isfahan (Iran)	ARMMIS- ACHG
Séries chronologiques indice d'oscillation australe (IOA)	Mensuel	N/A	ARMM-ACHG
Séries chronologiques des débits	Quotidien	la rivière Matapédia (Canada)	ARMMI-ACHG asymétrique
Panneau B: études de cas multivariés			
Relation de pluie-débit	Quotidien	Saint-Laurent bassin (Canada)	Diagonal VECH et CCC
Relation de Indice standardisé de précipitation (ISP) et indice d'oscillation australe (IOA) et Oscillation Nord Atlantique	Mensuel	Les Stations Oroomieh et Shiraz (Iran)	Diagonal VECH et Diagonal BEKK
relation de fracture de la hanche et des conditions de temps	Quotidian	Région de Montréal (Canada)	ARMMX-ACHG et CCC
relation de MCG et température	Quotidian	La station de l'Aéroport de Bagotville (Canada)	Diagonal VECH et CCD

4. Résultats et discussions

Les résultats de cette étude sont résumés ici en 3 sections principales. La première section présente les résultats des modèles de variance conditionnelle montés sur les différentes variables hydrologiques et climatiques. Dans la section 2, les résultats des modèles de covariance conditionnelles sont présentés et discutés. Finalement, les résultats des tests de non-linéarité et non-stationnarité des différentes variables et leur structures variance-covariance conditionnelle sont brièvement discutés.

4.1. Résultats sur la variance conditionnelle

La variance conditionnelle estimée par les modèles ACHG montre différentes caractéristiques pour les différentes variables hydrologiques et climatiques. Par exemple, le tableau 2 montre les paramètres ACHG pour certaines variables hydrologiques et climatiques considérés dans cette étude.

On constate que les précipitations, l'Oscillation Nord Atlantique (ONA) et certaines séries chronologiques de CMCG ont un grand nombre de paramètres ACHG qui présentent une mémoire forte pour la variance conditionnelle. Ceci indique que la variance à chaque pas de temps est fortement dépendante de la variance au pas de temps précédent. En revanche, la sécheresse, l'indice d'oscillation australe (IOA) et la plupart des séries chronologiques de CMCG n'indiquent pas une persistance à long terme dans la variance conditionnelle. Parmi ces variables, les débits, la température minimum et de 12-mois séries chronologiques de indice standardisé de précipitation (ISP) montrent le plus grand degré de volatilité comme paramètre ACH (α) est relativement plus élevé que le paramètre ACH pour d'autres variables hydrologiques et climatiques. Le tableau 2 indique également une forte densité persistance de la plupart des variables dans le cas où les mesures de la persistance de clôture sont $\alpha + \beta = 1$.

De plus, plus dans le long et le court terme de la persistance de la variance conditionnelle des différentes variables, certaines autres caractéristiques importantes sont observées par l'application du modèle ACHG dans cette étude. Par exemple, la performance du modèle ARMMIS pour la modélisation des séries temporelles des précipitations ne montre pas une amélioration significative après ajustement du modèle ACHG pour les résidus obtenus avec un modèle d'erreur ARMMIS- ACHG.

Tableau 2. Estimations de modèle ACHG univariées pour certaines variables hydrologiques et climatiques dans cette étude

Les données	Les séries	Les paramètres			la mesure de persistance
		ω	α	β	$\alpha+\beta$
Indice atmosphérique	IOA	0,34	0,39	0,34	0,73
	ONA	0,12	0,05	0,82	0,87
sécheresse	3-mois ISP	0,51	0,50	0,07	0,57
	12-mois ISP	0,05	0,89	0,02	0,91
Débit	la station de Beaurivage	0,10	0,71	0,21	0,92
	la station de Bras D'Henri	0,13	0,86	0,11	0,97
précipitation	la station de St, Pierre	0,81	0,05	0,93	0,98
	la station de St, severin	0,42	0,04	0,94	0,98
Température maximale	La station de l'Aéroport de Bagotville	8,06	0,46	0,50	0,96
Température minimale	La station de l'Aéroport de Bagotville	6,82	0,63	0,33	0,96
modèles de circulation générale (MCG)	humidité spécifique	0,02	0,19	0,79	0,98
	MSLP	0,55	0,40	0,06	0,46
	La vitesse du vent	0,71	0,16	0,11	0,27
	La composant U	0,62	0,34	0,01	0,35

Le modèle ACHG semble supprimer l'hétéroscédasticité restant dans les résidus du modèle SARMMI mais pas nécessairement pour améliorer les performances du modèle linéaire. En outre, la transformation de Box-Cox (1964) pour la stabilisation de la variance semble être en mesure de retirer la fonction de variance temporelle des séries de précipitations pour que le modèle ACHG, avec plus de paramètres, ne soit pas nécessaire pour une modélisation plus poussée.

D'autre part, la modélisation hétéroscédastique montre l'avantage du modèle ACHG exponentiel asymétrique pour améliorer l'efficacité d'un modèle ARMMI pour la modélisation des séries chronologiques des débits. Cependant, d'autres modèles ACHG asymétrique ne présentent pas cet avantage. La figure 2 montre le débit observé contre le débit estimé par les différents modèles ACHG asymétriques. Il est clair que le modèle ACHG exponentiel surpasse les modèles ARMMI et les autres pour la modélisation des séries temporelles de débit.

Un autre résultat important est associé aux modèles ACHG consacrés à la modélisation ainsi que le test de la variance conditionnelle de l'indice d'oscillation australe (IOA). La variance conditionnelle de l'IOA est illustrée à la figure 2. Cette figure révèle un changement significatif dans la variance conditionnelle après 1976 (qui est identifié comme le point de changement). On peut voir que la variance conditionnelle de l'IOA est 2 à 3 fois plus élevée après 1976. Le paramètre ACH pour IOA a augmenté de 0,29 à 0,42 alors que le paramètre ACHG a diminué de 0,42 à 0,35. En d'autres termes, la volatilité est de plus en plus forte alors que la mémoire de la variance va en s'affaiblissant au cours de ces dernières années.

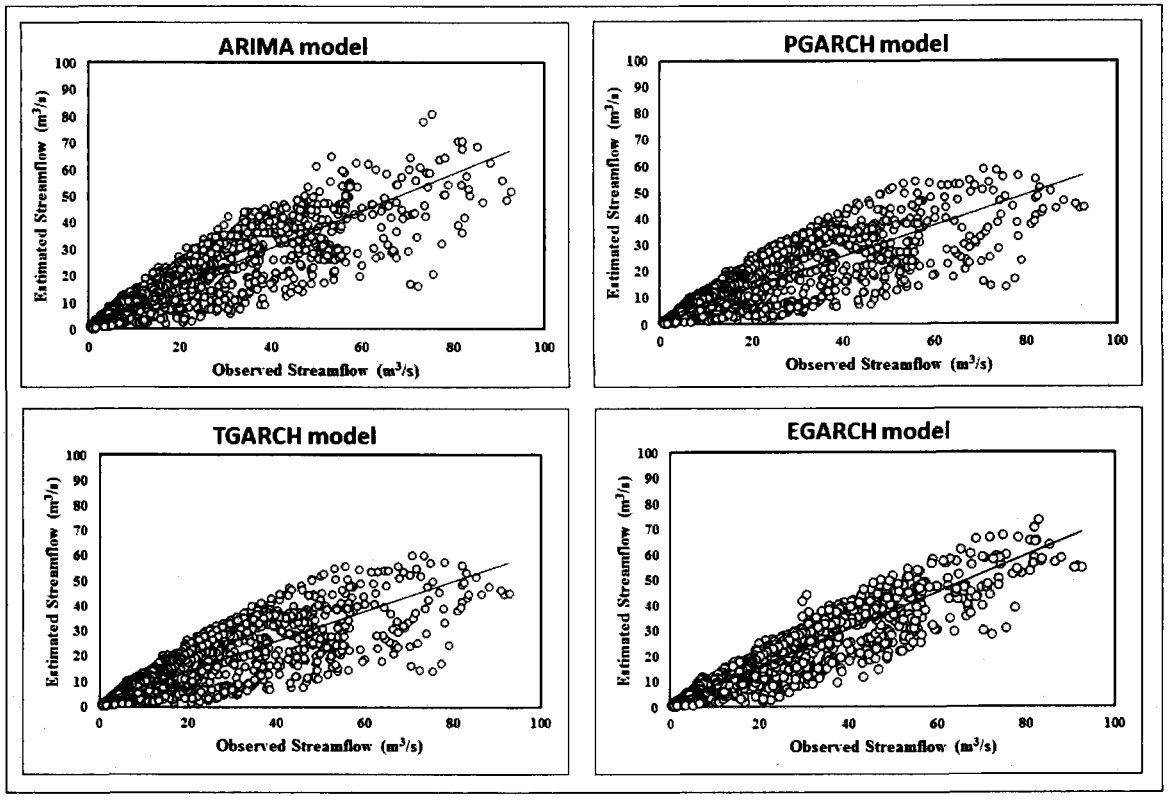


Figure 1. Diagramme de dispersion des débits observés contre estimés pour les différents modèles GARCH

4.2. Résultats sur la covariance conditionnelle

Le principal objectif de cette recherche est de développer et d'estimer la covariance conditionnelle entre les différentes variables hydrologiques et climatiques. Dans ce qui suit, nous décrivons les principaux résultats de la modélisation des différents processus ACHGM pour les variables hydroclimatiques prises en compte dans cette recherche.

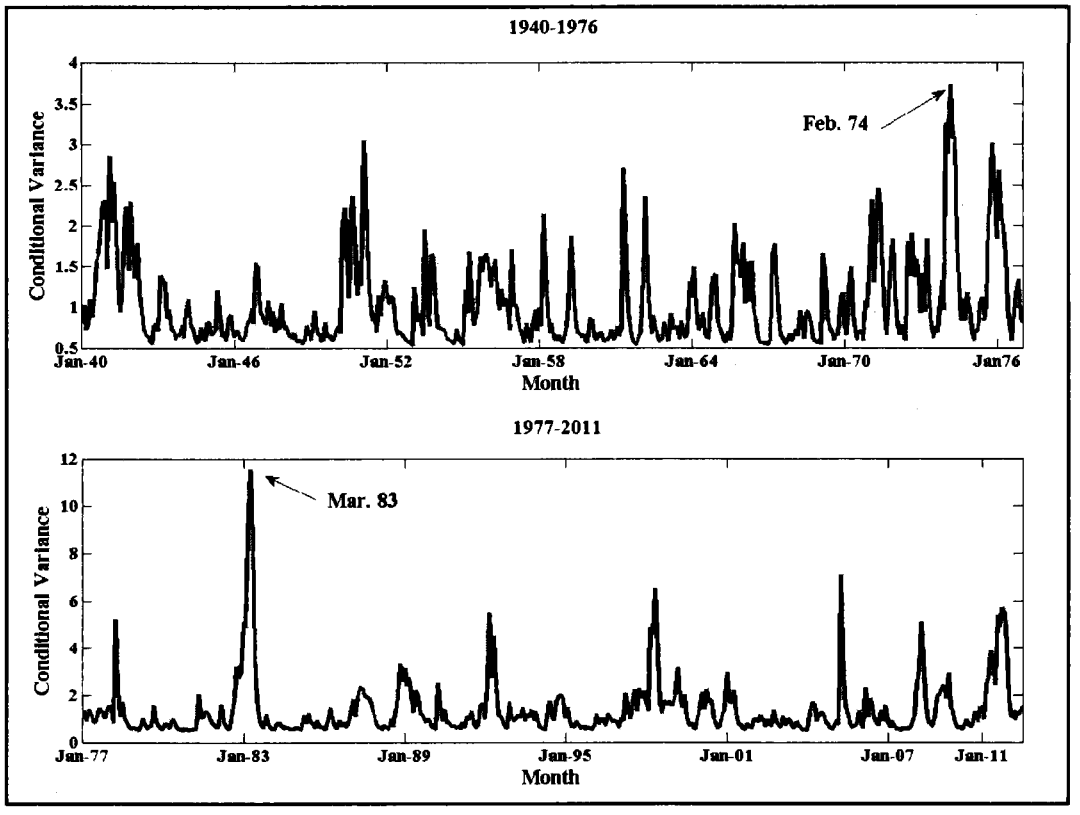


Figure 2. Les séries variances conditionnelles temporelles de SOI avant et après 1976

4.2.1. La Covariance conditionnelle de processus précipitation-débit

Les modèles VECH bivariés (1,1) et CCC (1,1) ont été développés pour étudier la structure de covariance conditionnelle dans un processus pluie-débit pour un sous-bassin du bassin versant du Saint-Laurent (sud-est), dans la province de Québec. Le modèle diagonal VECH indique que la persistance à court terme ou co-volatilité entre les précipitations et les débits n'est pas forte alors que le paramètre ACHGM indique une mémoire forte de la covariance entre les précipitations et les débits qui en fait une mesure forte de la persistance.

La variance conditionnelle de précipitations est beaucoup plus forte que celle des débits, ce qui suggère que la force de la covariance conditionnelle peut provenir de la forte variance des précipitations. On observe que la covariance conditionnelle n'est pas restée élevée pendant une longue période. L'effet des caractéristiques physiques du bassin versant telles que la superficie du bassin versant sur la covariance conditionnelle devrait être mentionné comme facteurs de modification.

Dans la figure 3, la matrice conditionnelle de variance-covariance d'un processus pluie-débit, dérivé du modèle diagonal VECH, est illustrée à titre d'exemple. La série temporelle de covariance conditionnelle pour un processus pluie-débit permet de constater que la covariance ne peut pas rester élevée pendant une longue période avant de retomber à un niveau plus bas.

Les résultats du modèle CCC sont identiques aux résultats du modèle VECH diagonal, ce qui implique une forte mémoire de la covariance conditionnelle entre les précipitations et le débit. Par ailleurs, un exemple de la corrélation conditionnelle entre les précipitations et le ruissellement est présenté à la figure 4. Cette corrélation conditionnelle a été estimée au moyen du modèle VECH diagonal.

La figure 4 montre que la corrélation entre les précipitations et les débits varie rapidement dans le temps et n'a pas tendance à rester à un niveau haut ou bas pendant une longue période. Cette fluctuation montre également que l'hypothèse de la corrélation conditionnelle constante (CCC) pour un processus pluie-débit n'est pas réaliste et le modèle CCC ne peut pas être approprié pour la modélisation de la covariance conditionnelle pour un processus pluie-débit.

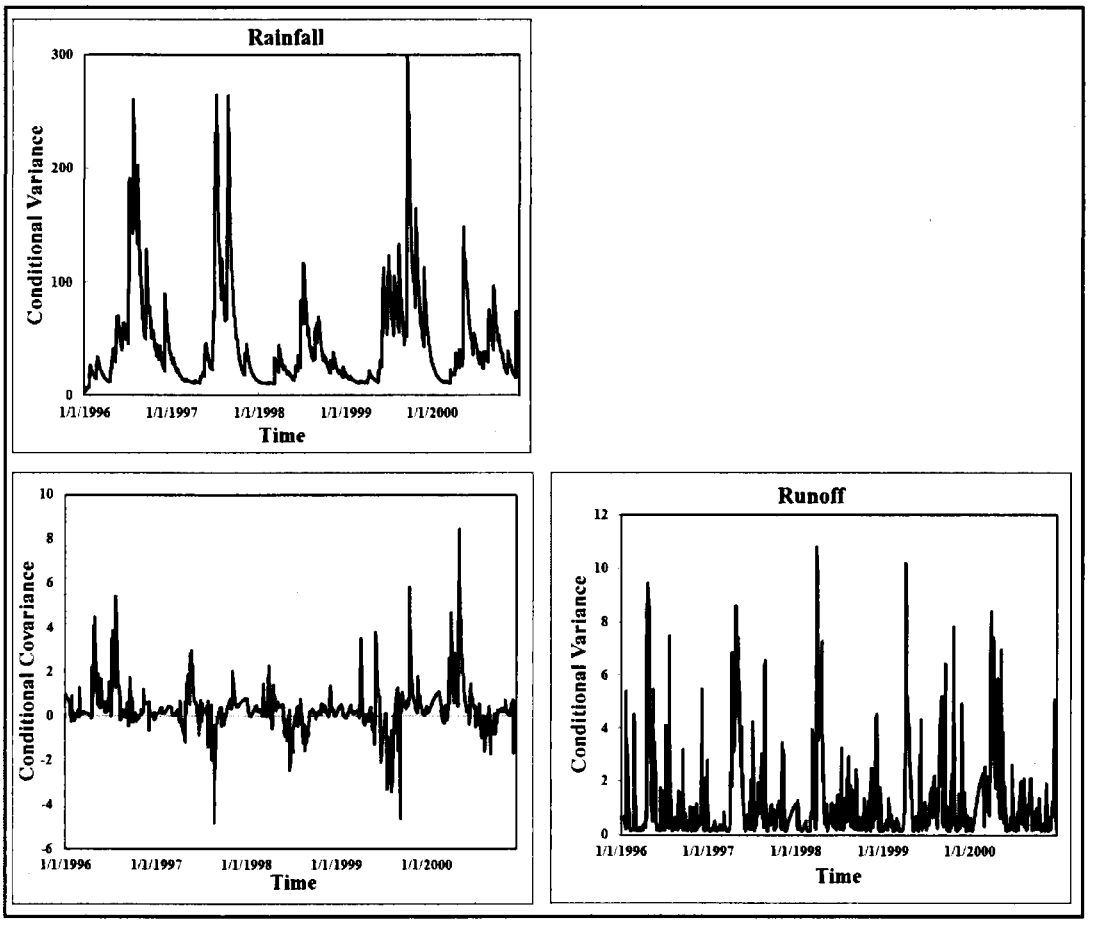


Figure 3. La matrice de variance-covariance conditionnelle entre les précipitations à la station Saint-Séverin et les débits à la station de Beaurivage

Finalement, l'expérience de simulation révèle que le ACHGM (1,1) est suffisant pour la modélisation de la structure de covariance des processus précipitations- débit et les modèles d'ordre supérieur tels que ACHGM (2,2) n'améliorent pas l'efficacité du modèle tandis qu'ils augmentent le nombre de paramètres.

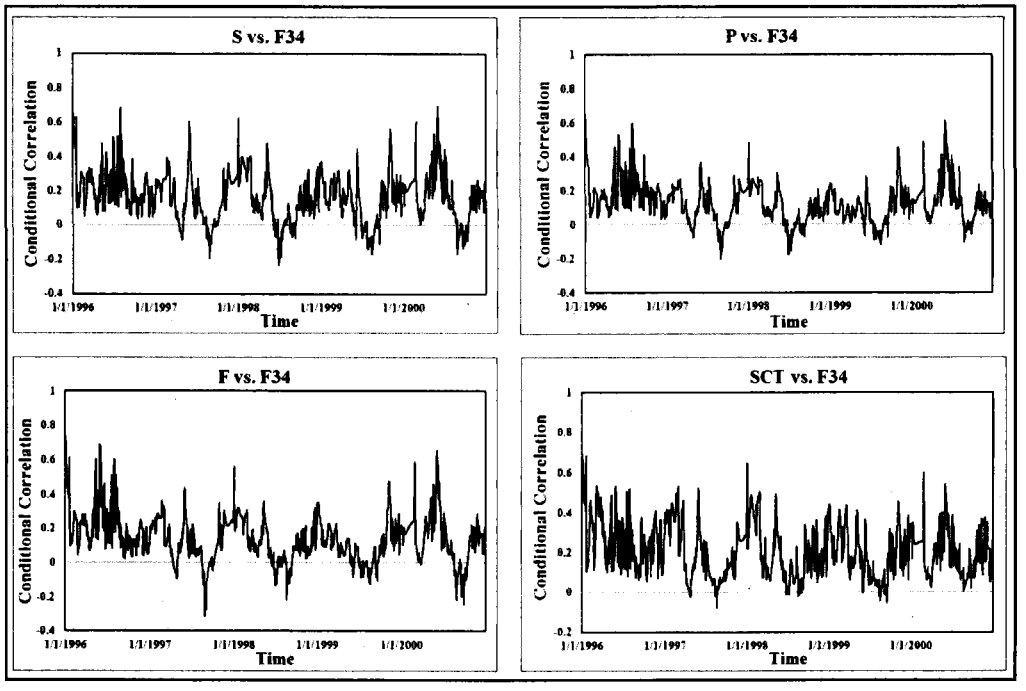


Figure 4. La corrélation conditionnelle entre les débits à la station de Beaurivage et des séries chronologiques des précipitations

4.2.2. La Covariance conditionnelle pour les effets du climat sur les fractures de la hanche

Le modèle CCC indique une relation non linéaire entre l'incidence des fractures de la hanche et les variables climatiques sur la région de Montréal, Canada. Parmi les différentes variables climatiques, la hauteur de neige, la température minimale et la longueur du jour sont les plus importants facteurs météorologiques régissant la fracture de la hanche chez les deux sexes (hommes et femmes) et les deux groupes d'âge (40-74 et 75 +) à l'étude. La covariance conditionnelle entre les fractures de la hanche et les variables climatiques est fonction du taux de fracture de la hanche chez les deux sexes et les groupes d'âge, à titre d'exemple dans les figures 5 et 6 pour les séries chronologiques de durée 3 jours de la fracture de la hanche.

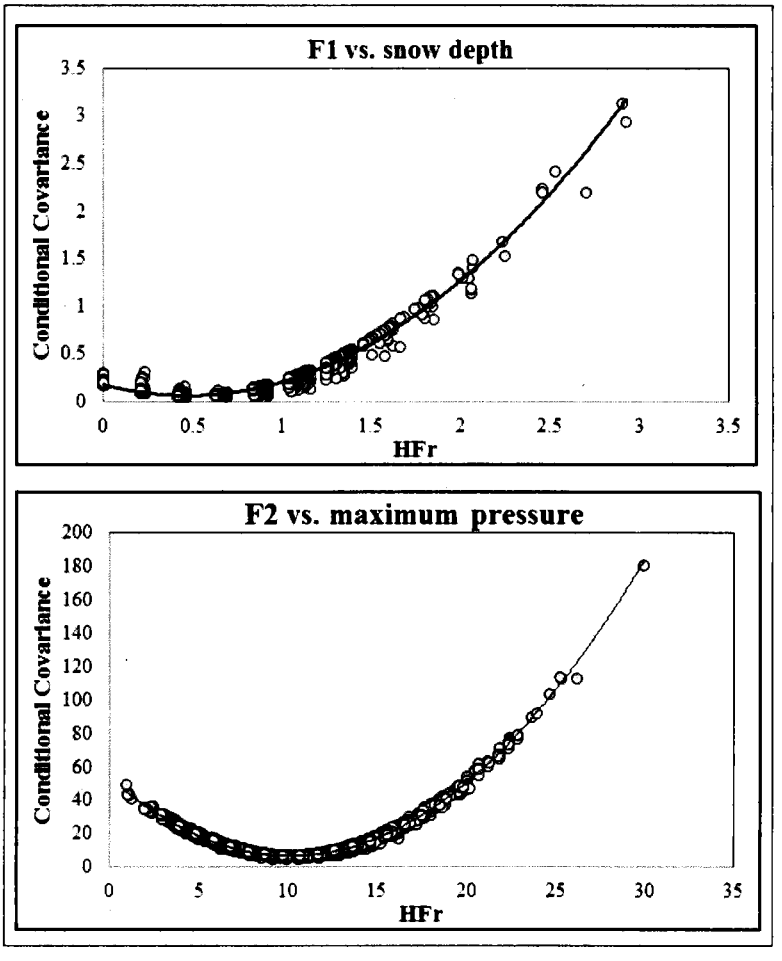


Figure 5. Covariance conditionnelle en fonction du taux de fracture de durée 3 jours de la hanche (HFr). L'exemple est pour les hommes en âge de 40-74 (F1) et les femmes de 75 ans et plus (F2)

Ces figures indiquent une forte relation non linéaire entre le taux de fracture de la hanche et de la variance des variables climatiques. On observe que la covariance entre l'incidence des fractures de la hanche et les variables climatiques est très faible ou linéaire pour un petit nombre de fractures de la hanche tandis que l'incidence de cette association augmente rapidement et de façon non linéaire pour des valeurs de la hanche plus élevées.

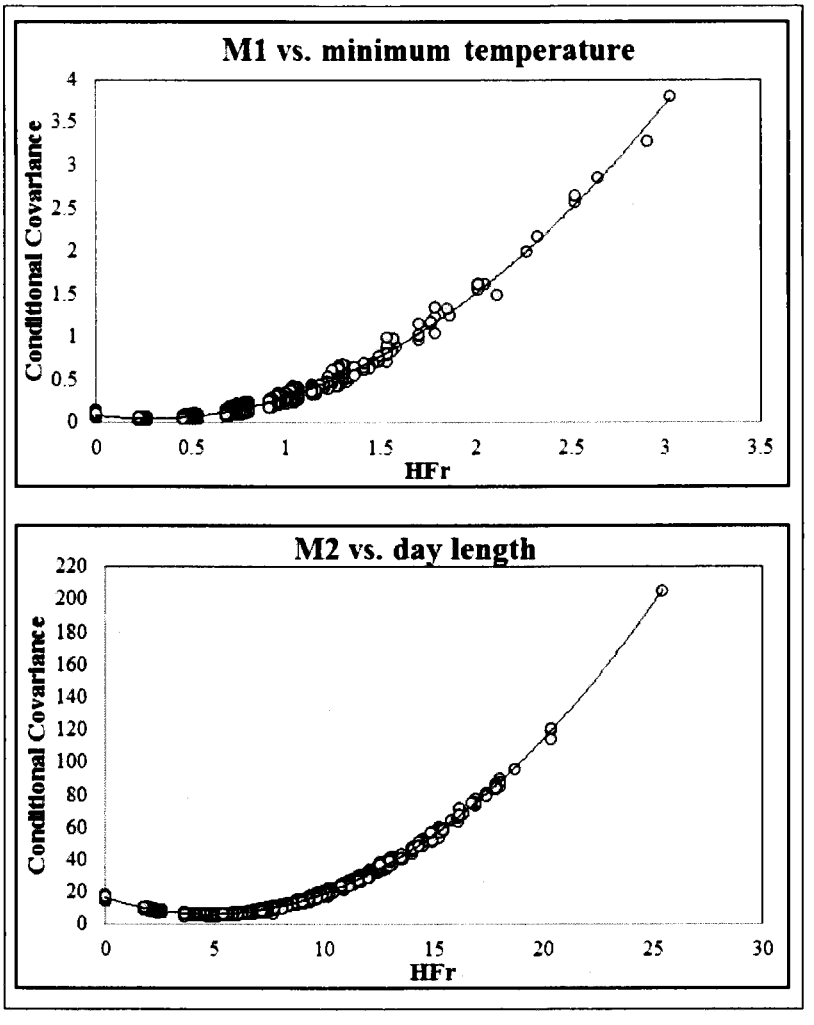


Figure 6. La covariance conditionnelle en fonctinon du taux de fracture de la hanche de durée 3 jours (HFr). L'exemple est pour les hommes de 40 à 74 ans (M1) et des hommes de l'âge 75 + (M2)

Par exemple, la covariance conditionnelle entre F1 et la profondeur de la neige est presque constante ou nulle pour le taux de fracture de la hanche <1 et elle commence à augmenter de façon exponentielle par la suite. Ces résultats montrent des fortes associations non linéaires entre les variances des variables climatiques et les taux de fractures de la hanche, ce qui signifie que le risque de fracture élevé dépend plus

fortement des conditions météorologiques sévères que des conditions moyennes. En conséquence, le risque de fracture de la hanche semble augmenter lorsque les conditions climatiques spécifiques, tels que les fortes chutes de neige ou de basses température se produisent et perdurent pour une période de temps. (Par exemple deux ou trois jours). Le modèle non linéaire appliqué dans cette étude justifie sans conteste cette affirmation.

4.2.3. La Covariance conditionnelle de MCG et de la température

Les Modèles climatiques globaux (MCG) sont considérés comme des outils appropriés pour reproduire les variables atmosphériques. Dans l'approche de la mise à l'échelle statistique (ES), les variables climatiques à l'échelle locale peuvent être obtenues à partir des prédicteurs de MCG à l'échelle globale. Cette recherche développe la covariance conditionnelle entre les prédicteurs de MCG des séries temporelles de température maximale et minimale.

Les modèles VECH diagonal et DCC révèlent la structure de covariance entre les différents prédicteurs des MCG et des séries chronologiques de la température de l'air. On observe que la covariance entre les prédicteurs de MCG et des séries chronologiques de température (maximale et minimale) dépend fortement du produit croisé des chocs. Cela donne à penser que les chocs de la volatilité des prédicteurs à chaque pas de temps dépendent de la volatilité précédente. Selon les modèles de covariance, les MCG peuvent être divisés en trois groupes :

Le premier groupe a une relation caractérisée par une forte covariance avec la température comme l'humidité spécifique. Le deuxième groupe comprend les MCG qui ont à la fois une persistance à court terme et à long terme avec des paramètres GARCH positifs tels que la vitesse du vent et la hauteur géopotentielle. Le troisième groupe est

caractérisé par les paramètres ACHG négatifs et montre une relation négative entre la variance de la MCG et la température.

La covariance conditionnelle entre l'humidité spécifique et les séries chronologiques de température est donnée à la figure 7 à titre d'exemple. La covariance conditionnelle entre les prédicteurs des MCG et la température montre une saisonnalité hiver-été pour la plupart des MCG à l'exception de l'humidité spécifique et la température à une hauteur de 2 mètres (T_2) qui montrent une variation temporelle de la covariance différente de la saisonnalité.

Les modèles de DCC révèlent aussi une persistance significative à court terme pour la covariance conditionnelle pour tous les modèles de circulation générale tandis que les paramètres de covariance ne sont pas significatifs pour certains MCG. Ceci suggère l'hypothèse irréaliste du modèle de corrélation conditionnelle constante. La variation quotidienne de corrélation conditionnelle entre certaines variables de sortie des MCG et des séries chronologiques de température est donnée dans les figures 8 et 9 pour chaque jour de l'année au cours de 1980-2000.

Les résultats montrent que les coefficients de corrélation (en moyenne) sont peu élevés, montrant une faible association entre la plupart des prédicteurs MCG et la température. De plus ils restent au même niveau au cours de l'année, sauf pour certaines variables issues des MCG tels que l'humidité spécifique, (T_2), la hauteur géopotentielle et la direction du vent avec les variations saisonnières et des coefficients relativement élevés.

En outre, la variation annuelle des coefficients de corrélation conditionnelle pendant 1980-2000 ne révèle aucun changement significatif pour l'association entre les variables de sorties de MCG et la température au cours de la période d'étude de 20 ans.

4.2.4. La Covariance conditionnelle pour l'analyse de la sécheresse

Les deux principaux indices océaniques et atmosphériques, l'Indice d'Oscillation Australe (IOA) et l'Oscillation Nord Atlantique (ONA), sont censés influencer les conditions climatiques dans le monde. Cette recherche étudie la covariance et la corrélation conditionnelle entre ces indices et l'indice de précipitations standardisée (IPS) dans les stations Oroomieh et Shiraz en Iran.

La covariance conditionnelle estimée par les modèles VECH diagonal et BEKK diagonal est faible et la plupart du temps négative, entre les indices atmosphériques et la sécheresse. On observe que la persistance à court terme n'est pas très forte, sauf pour la persistance à court terme entre le IOA et l'indice de sécheresse pour les deux stations. Cela implique que la variation des indices IOA et ONA peut avoir une influence rapide à court terme sur les variations de sécheresse dans les stations visées dans cette étude.

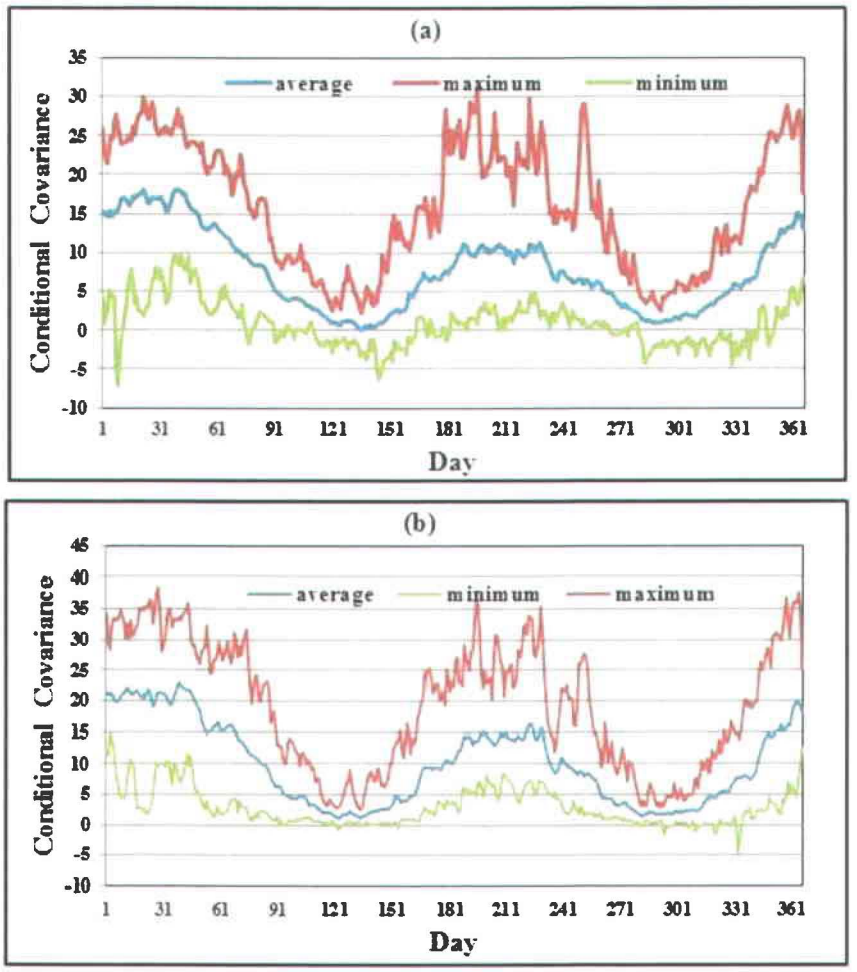


Figure 7. La covariance conditionnelle entre l'humidité et a) la température maximale et b) la température minimale.

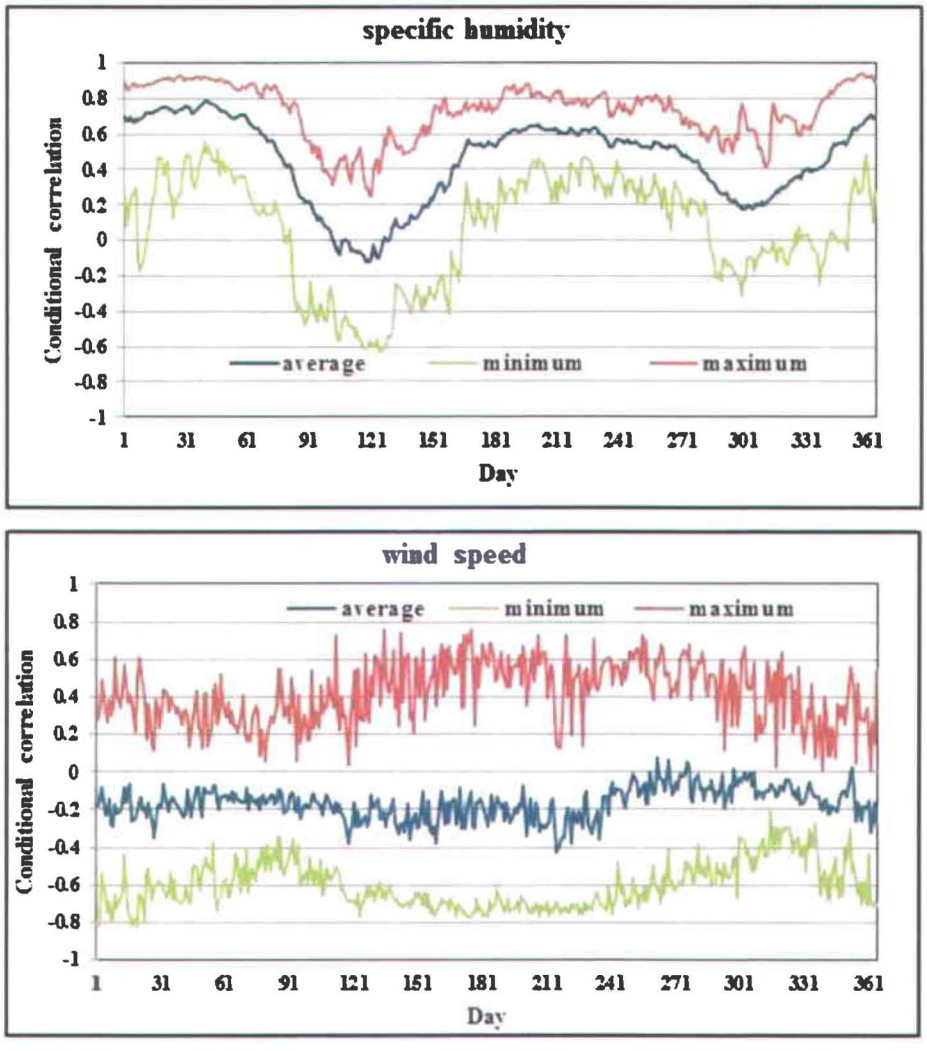


Figure 8. Les Exemples de corrélation conditionnelle quotidienne entre certains prédicteurs de MCG (humidité spécifique, la vitesse du vent) et de la température maximale.

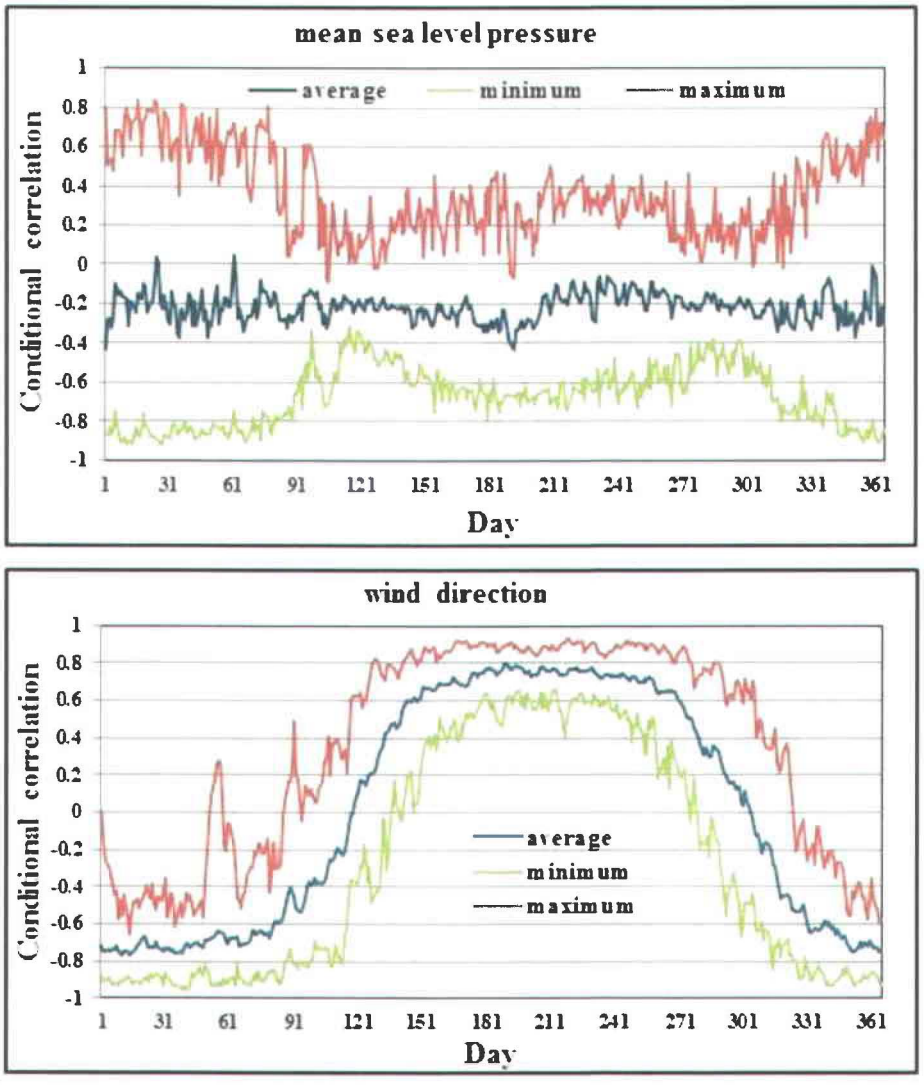


Figure 9. Les exemples de corrélation conditionnelle quotidienne entre certains prédicteurs de MCG (température moyenne du niveau des mers et la direction du vent) et la température minimale.

Bien que les deux modèles diagonaux VECH et BEKK montrent la même structure de la covariance conditionnelle, la simulation montre que le mode diagonal VECH est plus performant que le modèle BEKK pour estimer la covariance conditionnelle. Les tableaux 3 et 4 indiquent les critères de performance du modèle pour les deux modèles à des

stations sélectionnées. A titre d'exemple, la covariance et la corrélation conditionnelle IPS entre 12 mois (IPS12) et des indices atmosphériques sont donnés dans les figures 10 et 11, en utilisant le modèle diagonal VECH.

Ces figures montrent que les covariances et les corrélations conditionnelles n'ont pas tendance à rester dans un niveau haut ou bas pour longtemps et leurs fluctuations sont élevées. Cela implique que l'association entre la sécheresse et les indices atmosphériques a fortement fluctué et non persistante au fil du temps pour notre étude de cas. En outre, aucune variation saisonnière remarquable n'est observée pour cette association. La variation annuelle des coefficients de corrélation ne montrent pas de tendance significative ni de tendance dans la relation entre les indices atmosphériques et de l'indice de sécheresse dans les stations sélectionnées au cours de la période de 1954 à 2010.

4.3. La stationnarité et la non-linéarité

Il existe deux principales hypothèses pour l'analyse et la modélisation hydrologique, stationnarité et la non-linéarité. Cette étude présente une enquête sur la non-linéarité et la stationnarité des différentes variables hydroclimatiques et leur structure de variance-covariance.

Les résultats des tests KPSS, PP et ADF ont révélé la stationnarité pour toutes les variables hydroclimatiques et leurs associations. Cependant, certaines preuves de tendance non-stationnaire sont observées pour l'association des MCG et des séries chronologiques de la température. Cette non-stationnarité est généralement associée à l'existence de valeurs aberrantes dans une série chronologique.

Tableau 3. Estimations des critères de performance des modèles de covariance conditionnelle à la station Oroomieh (Iran)

série de covariance	VECH diagonal		BEKK diagonal		statistique DM
	BIAS normalisé	RMSE normalisé	BIAS normalisé	RMSE normalisé	
IOA-IPS3	0.42	7.97	0.64	3.86	-4.19*
ONA -IPS3	0.82	12.94	-1.5	20.94	-17.4*
IOA-IPS12	-0.48	6.1	0.61	8.3	3.16
ONA-IPS12	-0.63	17.08	-0.8	8.7	-7.31*

* Significatif au niveau de 5% et mieux

Tableau 4. Estimations critères de performance des modèles de covariance conditionnelle à la station Shiraz (Iran)

Les série de covariance	VECH diagonal		BEKK diagonal		statistique DM
	BIAS normalisé	RMSE normalisé	BIAS normalisé	RMSE normalisé	
IOA-IPS3	0.24	8.32	0.62	17.45	-3.14*
ONA-IPS3	0.45	9.8	-0.41	51.87	-8.02*
IOA-IPS12	-0.14	12.98	-1.64	54.36	-10.7*
ONA-IPS12	-0.29	13.1	-2.09	12.52	-2.72*

* Significatif au niveau de 5% et mieux

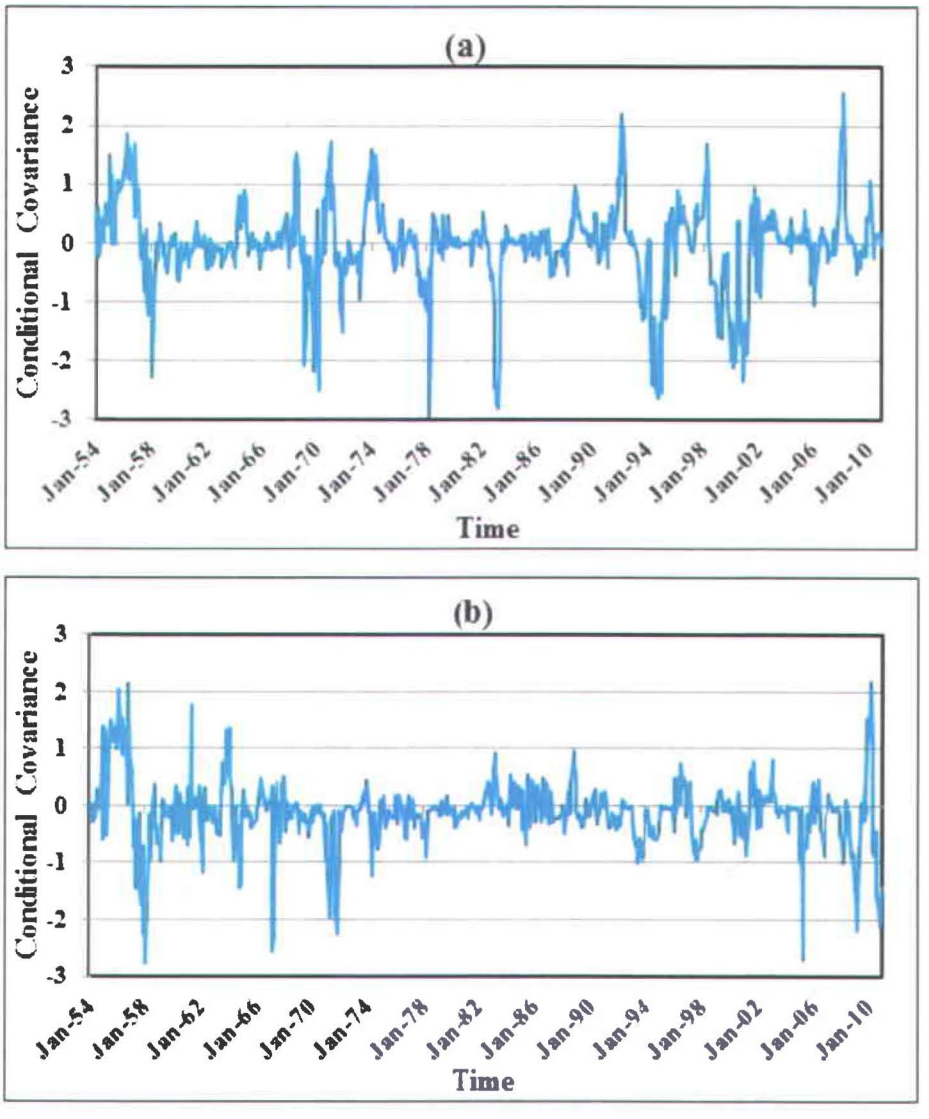


Figure 10. La covariance conditionnelle entre IOA et IPS12 aux stations Oroomieh (a) et Shiraz (b).

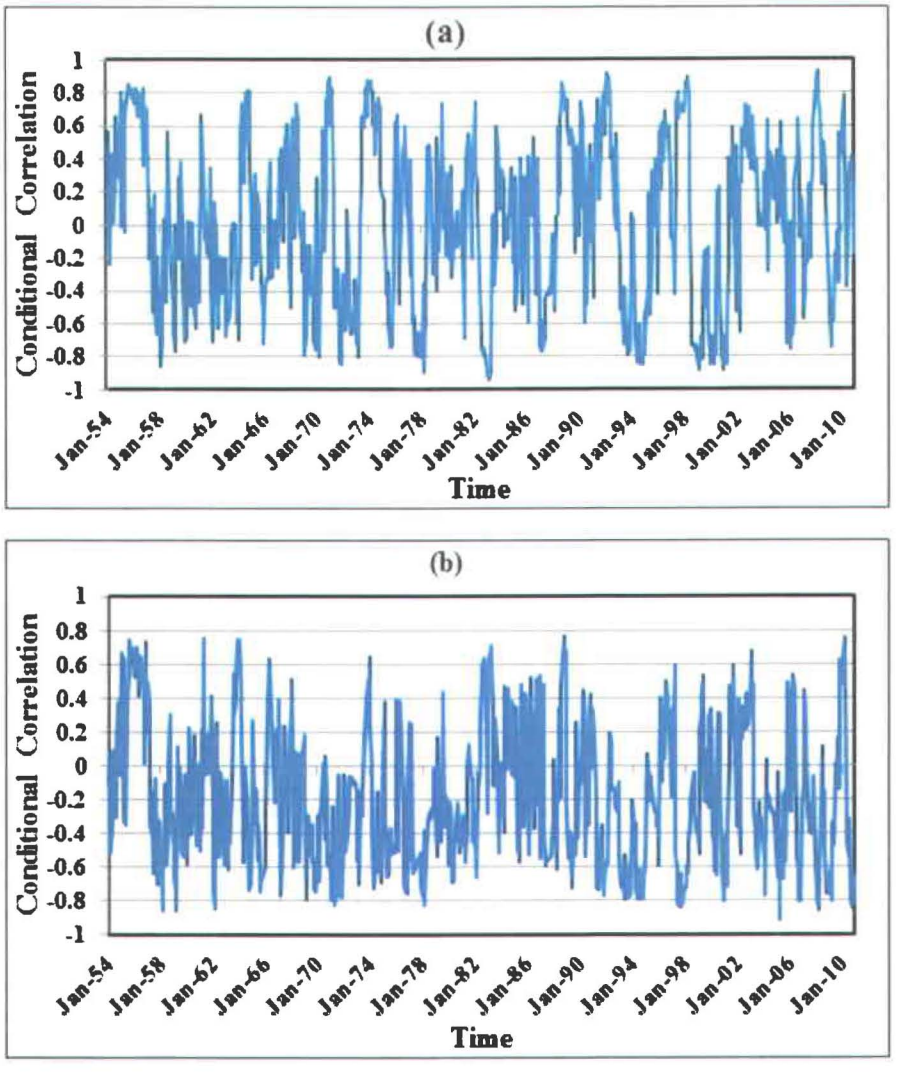


Figure 11. La corrélation conditionnelle entre IOA et IPS12 aux stations Oroomieh (a) et Shiraz (b).

En outre, la méthode BDS montre un haut degré de la non-linéarité pour tous les covariances conditionnelles et la corrélation entre les variables hydroclimatiques. Les résultats du test BDS montre qu'il existe des mécanismes de non-linéarité plus forts et plus compliqués agissant sur la covariance conditionnelle pour les séries chronologiques

quotidienne que les covariances conditionnelles pour les séries chronologiques mensuelles entre les différentes variables hydrologiques et climatiques.

Cependant, il semble que le processus pluie-débit a le plus grand degré de la non-linéarité. Ceci suggère l'effet du processus de captage physique derrière la relation non linéaire pluie-débit et sa structure de covariance. Le tableau 5 montre les résultats du test BDS pour quelques exemples de processus hydroclimatiques dans cette recherche.

Tableau 5. Exemples de résultats de test BDS pour les différents processus hydrologiques et climatiques

Processus	m=2		m=3		m=4	
	résultats	p-valeur	résultat	p-valeur	résultat	p-valeur
précipitation-débit	0,15	0	0,25	0	0,31	0
	0,15	0	0,26	0	0,33	0
	0,14	0	0,24	0	0,30	0
	0,14	0	0,24	0	0,30	0
MGC-Température	0,17	0	0,28	0	0,35	0
	0,04	0	0,07	0	0,08	0
	0,13	0	0,21	0	0,26	0
	0,15	0	0,25	0	0,32	0
	0,05	0	0,08	0	0,09	0
	0,02	0	0,04	0	0,06	0
	0,05	0	0,08	0	0,09	0
	0,18	0	0,30	0	0,40	0
	0,12	0	0,19	0	0,23	0
	0,03	0	0,06	0	0,08	0
	0,03	0	0,06	0	0,08	0
	0,05	0	0,09	0	0,12	0
IOA-IPS3	0,04	0	0,07	0	0,09	0
	0,10	0	0,18	0	0,22	0
ONA -IPS	0,01	0	0,02	0	0,03	0
	0,05	0	0,10	0	0,13	0

5. Conclusions

L'approche de la modélisation des séries temporelles est une méthode populaire en hydrologie, climatologie, pour la gestion des ressources en eau et les sciences de l'environnement. Bien que les modèles de séries chronologiques furent appliqués depuis des décennies en hydrologie, les modèles non linéaires de séries chronologiques n'ont pas été inspectés et mis au point. Cependant, la relation non linéaire entre les phénomènes hydrologiques dans l'espace et le temps est devenue plus intéressant au cours des dernières années. De nombreux chercheurs ont essayé d'appliquer différentes méthodes pour modéliser cette relation non linéaire.

Dans cette thèse, une approche de la classe non linéaire multivariée temporelle est introduite pour le développement, l'adaptation et l'application en hydrologie et en climatologie. Les modèles univariés et multivariés autorégressifs conditionnels hétéroscédastiques généralisés sont utilisés pour modéliser la variation temporelle du moment de second ordre de différentes variables hydroclimatiques et leur relation à travers la structure de covariance conditionnelle.

La plupart des variables de cette étude montrent une faible mémoire de la variance temporelle dans le temps ou la variance conditionnelle, sauf les précipitations, l'ONA et certaines séries chronologiques MCG.

Les autres séries hydrologique et climatiques ont montré un paramètre ARC fort ou assez fort, qui est un indicateur de groupement de la volatilité de ces variables. Cependant, on peut également constater que la mesure de la persistance ($\alpha + \beta$) reste élevée pour la plupart des variables considérées dans cette étude, ce qui indique que le moment de second ordre des variables hydroclimatiques n'est pas constant dans le temps, si la

dépendance augmente sur des pas de temps précédents ou si la volatilité des processus aléatoire régissant leurs innovations varie.

La structure de covariance conditionnelle entre les différentes variables de l'atmosphère, climatiques et hydrologiques ont également montré des structures positives et négatives. La plus forte covariance positive conditionnelle a été observée pour les processus pluie-débit et la covariance entre certaines variables issues des sorties de MCG telles que l'humidité spécifique et des séries chronologiques de la température.

Cependant, cette recherche a montré qu'il n'y a pas une forte association entre les moments du second ordre de la plupart des variables hydroclimatiques. Il semble que la covariance conditionnelle entre les variables hydroclimatiques dépend en grande partie des produits croisés de chocs (variable aléatoire ou innovations).

En d'autres termes, le regroupement de co-volatilité et la volatilité est plus forte que la covariance entre les variables climatiques et hydrologiques, du moins pour ceux étudiés dans cette recherche. La mémoire dans la matrice de covariance, donc, repose en grande partie sur la volatilité du choc précédent plutôt que la variance au pas de temps précédent.

En outre, si nous pouvons interpréter la non-linéarité dans la structure de variance-covariance comme une valeur au carré, la comparaison de la non-linéarité des variables hydroclimatiques indique une forte différence et le degré de non-linéarité. Les résultats des tests nous disent que la variance-covariance quotidienne a une forte non-linéarité temporelle pour les séries chronologiques mensuelles.

Il a également été observé que les processus pluie-débit ont le plus haut degré de non-linéarité des différents processus hydroclimatiques examinés dans cette étude. Les autres variances-covariances entre les variables climatiques montrent presque le même degré de

non-linéarité à l'exception de la covariance entre certains MCG et des séries chronologiques de la température. En outre, le test de stationnarité de la variance-covariance conditionnelle a révélé une stationnarité pour la plupart des associations entre les différentes variables. Les résultats ont indiqué la stationnarité pour tous les niveaux, alors que l'hypothèse nulle de stationnarité de tendance de certaines associations de variance-covariance peut être rejetée.

L'application des modèeles ACHG et ACHGM dans cette étude a également montré quelques avantages et certains inconvénients pour l'hydrologie et la climatologie. Le principal avantage des méthodes proposées est d'explorer la structure de covariance entre les différentes variables hydrologique et climatiques et est signalé pour la première fois dans cette étude. Les modèles ACHGM peut aussi estimer la variation temporelle des covariances et des coefficients de corrélation pour l'ensemble des séries chronologiques. Cette capacité offre l'avantage de permettre d'enquêter sur le changement (ou tendance) dans l'association entre différentes variables hydrologique et climatiques dans les perspectives locales, mondiales et des changements climatiques. Les résultats indiquent également l'avantage des modèles ACHG asymétriques pour augmenter les performances des modèles traditionnels de séries chronologiques, i. e ARMMI, pour la modélisation moyenne conditionnelle de débit. Cependant, le modèle ACHG symétrique n'a pas montré d'avantage sur le modèle SARMMI pour la modélisation de séries de précipitations mensuelles. Il semble que le modèle ACHGM symétrique supprime uniquement l'hétéroscédasticité des résidus du modèle linéaire SARMMI. En règle générale, les approches ACHG univariées et multivariées sont des méthodes utiles pour développer et appliquer dans les domaines de l'hydrologie et de la climatologie. Ces

modèles peuvent capter la non-linéarité univariée et multivariée existant dans les variables hydrologiques et leurs processus multidimensionnels non linéaires dans l'espace et le temps. L'approche ACHGM, issue de l'économétrie et de la finance semble être une approche intéressante pour l'analyse, modélisation et simulation non linéaire de phénomènes hydrologiques et climatologiques.

6. Les Recommandations pour les travaux futurs

Bien que cette étude a élaboré des modèles ACHGM multivariés pour les différentes variables hydroclimatiques, il reste encore des sujets intéressants pour les travaux futurs. Par exemple, les structures de variance-covariance de certaines autres variables telles que la vitesse du vent, la température, l'évapotranspiration, d'autres prédicteurs MCG, la température de l'eau, etc, peuvent être considérés pour de futures études. Par ailleurs, l'étude des variables physiques qui influencent la covariance conditionnelle et les paramètres des modèles ACHGM dans les applications hydrologiques reste un défi important pour des études futures. La répartition spatiale de la structure de variance-covariance entre les variables hydrologiques et climatiques est un élément important à étudier. Dans le contexte du changement climatique à venir, le changement dans le moment du second ordre de différentes variables climatiques et leur impact sur la variance des autres variables climatiques et hydrologiques est un sujet important pour de futures études.

Bibliographie

[1]. Adeloye, A. J., and Montaseri, M. (2002). Preliminary streamflow data analyses prior to water resources planning study. Hydrol. Sci. J., 47(5), 679–692.

[2]. Astel, A., Mazerski, J., Polkowska, Z., and Namieśnik, J. (2004), Application of PCA and time series analysis in studies of precipitation in Tricity (Poland). Adv. Environ. Res. 8, 337–349.

[3]. Baba, Y., Engle, R.F. Kraft, D. and Kroner, K. (1990). Multivariate simultaneous generalized ARCH, unpublished manuscript, University of California, San Diego

[4]. Bartolini, P., Salas, J.D. and Obeysekera, J. T.B. (1988). Multivariate Periodic ARMA(1,1) Processes. Water Res. Res., 24, 1237-1248

[5]. Baur, D. (2006). A flexible dynamic correlation model. In: Econometric analysis of financial and economic time series. Eds: Terrel, D and Fomby TB. Elsevier publications. Netherlands, 380p

[6]. Bauwens, L., Laurent, S., and Rombouts, V.K. (2006). Multivariate GARCH models: A survey. J. Appl. Econometrics, 21, 79-109

[7]. Benyahya, L., A. St-Hilaire, T.B.M.J. Ouarda, B. Nedushan et B. Bobée. (2007). Prediction of water temperature based on stochastic approaches. Case study of the Deschutes River (Oregon, U.S.A.) Journal of Environmental Engineering and Science 6: 437-448.

[8]. Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity, J. Econometrics 31, 307–327.

[9]. Bollerslev, T. (1990). Modeling the coherence in short-run nominal exchanges rates: a multivariate generalized ARCH model, Review of economics and statistics 72, 498-505.

[10]. Box, G.E., and Cox, D.R. (1964). An analysis of transformed data. J. Roy. Statist. Soc. B 39, 211–252

[11]. Box, G. E. P. and Jenkins, G.M. (1970). Time series analysis, forecasting and control, Holden – Day, San Francisco, California.

[12]. Brock, W.A., Dechert, W.D., Scheinkman, J. A., and LeBaron, B. (1996). A test for independence based on the correlation dimension, Econ. Rev,. 15 (3), 197–235.

[13]. Camacho, F., McLeod, A. I., and Hipel, K.W. (1987). Multivariate contemporaneous ARMA model with hydrological applications. Stoch. Hydrol. Hydraul., 1, 14-154

[14]. Chen, C. H., Liu, C. H., and Su, H .C. (2008). A nonlinear time series analysis using two-stage genetic algorithms for streamflow forecasting. Hydrol. Process. 22, 3697–3711

[15]. Chen, H.L., and Rao, A.R. (2003). Linearity analysis on stationary segments of hydrologic time series. J. Hydrol. 277, 89–99.

[16]. Choudhry, T. (2003). Stock market volatility and the US consumer expenditure. J. Macroeconomics, 25, 367-385

[17]. Cooper, D. M., and Wood, E. F. (1982). Identification of multivariate time series and multivariate input-output models. Water. Res. Res., 18, 937-946

[18]. Ding, Z., Granger, C. W. J., and Engle, R. F. (1993), A Long Memory Property of Stock Market Returns and a New Model, J. Empiric. Finance, 1, 83–106.

[19]. Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation, Econometrica 50, 987-1008.

[20]. Engle, R. F. (2002). Dynamic Conditional Correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticty models. J. buss. Economy. Stat., **20**, 339-350

[21]. Fong, P.W., Li, W.K., An, H. Z. (2006). A simple multivariate ARCH model specified by random coefficients. Computational statistics and data analysis, 51, 1779-1802

[22]. Hakim, A., McAleer, M. (2009). Forecasting conditional correlations in stock, bond and foreign exchange markets. Math. Comput. Simul. 79, 2830-2846

[23]. Hipel, KW., Mcleod, AJ., Lennox, WC. (1977). Advances in Box-Jenkins modeling 1. Model construction. Water Res. Res., 13, 567-575

[24]. Karolyi, G. A. (1995). A Multivariate GARCH Model of International Transmissions of Stock Returns and Volatility. J. Bus. Econ. Stat., 13, 11-25

[25]. Kim, J. (2000). The relationship between the monetary regime and output volatility: a multivariate GARCH-M model of the Japanese experience, 1919-1996. Japan and the world economy, 12, 49-69

[26]. Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., and Shin, Y. (1992). Testing the null of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? J. Econometrics, 54, 159–178.

[27]. Klemes, V. (1973). Watershed as semi-infinite storage reservoirs, J. Irrig. Drain. Div., Am. Soc. Civ. Eng., 99(IR4), 477-491.

[28]. Ljung G.M., and Box G.E.P. (1978). On a Measure of Lack of Fit in Time Series Models. Biometrika, 65: 297-303.

[29]. Ledolter, J. (1978). The analysis of multivariate time series applied to problems in hydrology. J. Hydrol., 36, 327-352

[30]. Matalas, N. C. (1967). Time series analysis, Water Res. Res., Vol. 3, 817-829.

[31] McKerchar, A. I. and Delleur, L. W. (1974). Application of seasonal parametric linear stochastic models to monthly flow data, Water Res. Res., 10, 246-255.

[32]. Mcleod, A.J., Hipel K.W., Lennox, W.C. (1977). Advances in Box-Jenkins modeling 2. Applications. Water Res. Res., 13, 577-586

[33]. Montanari, A., Rosso R., and Taqqu, M. S. (2000). A seasonal fractional ARIMA model applied to the Nile river monthly flows at Aswan, Water Res. Res., 36, 1249-1259.

[34]. Moreau, D. H., and Pyatt, E. E. (1970). Weekly and monthly flows in synthetic hydrology, Water Res. Res., 6(1), 53-61, 1970.

[35]. Moss, M. E. and Bryson, M. C. (1974). Autocorrelation structure of monthly streamflow, Water Res. Res., 15, 734-744.

[36]. Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. Econometrica, 69, 1519–1554.

[37]. Onof, C., and Wheater, H. S. (1996). Modeling of the time series of spatial coverage of British rainfall field. J. Hydrol. 2, 115-131.

[38]. Phillips, P. C.B., and Perron, P. (1988). Testing for a unit root in time series regression. Biometrika, 75, 335-334

[39]. Quimpo, R. (1971). Structural relation between parametric and stochastic hydrology models, International Symposium on Mathematical Models in Hydrology, Int. Assoc. of Hydrol. Sci., Warsaw.

[40] Rao, A. R., Kashyap, R. L., and Mao, L. T. (1982). Optimal choice of type and order of models for river flow time series, Water Res. Res., 18, 1097-1109.

[41]. Rao, A.R., Yu, G.-H. (1990). Gaussianity and linearity tests of hydrologic time series. Stochast. Hydrol. Hydraul. 4, 121–134.

[42]. Rasmussen, P.F., Salas, J.D., Fagherazzi, L., Rassam, J.C., and Bobée, B. (1996). Estimation and validation of contempronous PARMA models for streamflow simulation. Water Res. Res., 32, 3151-3160

[43]. Romilly, P. (2005). Time series modeling of global mean temperature for managerial decision-making. J. Environ. Manage., 76, 61-70

[44]. Salas, J. D. (1993). Analysis and modeling of hydrologic time series, In: D. R. Maidment (Ed) Handbook of Hydrology, 19.1-19.72, McGraw Hill, New York.

[45]. Salas, J. D. and Obeysekera, J. T. B. (1982). ARMA model identification of hydrologic time series, Water Res. Res., 18, 1011-1021.

[46]. Salas. J. D., Tabios, G. Q., Ill, & Bartolini, P. (1985) Approaches to multivariate modeling of water resources time series. Water Res. Bull., 24, 683-708.

[47]. Spoila, S. K and Chander, S. (1974). Modeling of surface runoff systems by an ARIMA model. J. Hydrol. 22, 317-332.

[48]. Thomas, H. A and Fiering, M. B. (1962). Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation, Harward University Press, Cambridge.

[49]. Tsui, A. K., and Yu, Q. (1999). Constant conditional correlation in a bivariate GARCH model: evidence from the stock markets of China. Math. Comput. Simul. 48, 503-509

[50]. Wang W, Vrijling J.K., van Gelder P.H.A.J., and Ma J., (2006). Testing for nonlinearity of streamflow processes at different timescales. J. Hydrol. 322: 247–268

[51]. Wang, W., J. K.Vrijling, P. H. A. J. M. Van Gelder and Ma, J. (2005). Testing and modeling autoregressive conditional heteroskedasticity of streamflow processes. Nonlinear Proc. Geoph. 12, 55-66

[52]. Zakoïan, J. M. (1994). Threshold Heteroskedastic Models, J. Econ. Dyn. Cont., 18, 931-944.

PART II: ARTICLES

Article 1. Generalized autoregressive conditional heteroscedasticity modeling of hydrologic time series

Generalized autoregressive conditional heteroscedasticity modeling of hydrologic time series

R. Modarres[a*], T. B. M. J. Ouarda[a,b],

[a] Canada Research Chair on the Estimation of Hydrometeorological variables, INRS-ETE, 490 De La Couronne, Québec, QC, Canada, G1K 9A9

Tel: +1 418 654-3842, Fax: +1 418 654-2600,

[b] Masdar Institute of Science and Technology, P.O. Box 54224, Abu Dhabi, UAE

E-mail: Reza.Modarres@ete.inrs.ca

[*] Corresponding author

Accepted for publication in Hydrological Processes

Abstract

The existence of time dependent variance or conditional variance, commonly called heteroscedasticity, in hydrologic time series has not been thoroughly investigated. This paper deals with modeling the heteroscedasticity in the residuals of the seasonal autoregressive integrated moving average (SARIMA) model using a Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model. The model is applied to two monthly rainfall time series from humid and arid regions. The effect of Box-Cox transformation and seasonal differencing on remaining the seasonal heteroscedasticity in the residuals of the SARIMA model is also investigated. It is shown that the seasonal heteroscedasticity in the residuals of the SARIMA model can be removed using Box-Cox transformation along with seasonal differencing for the humid region rainfall. On the other hand, a transformation and a seasonal differencing could not remove heteroscedasticity from the residuals of the SARIMA model fitted to rainfall data in the arid region. Therefore, the GARCH modeling approach is necessary to capture the heteroscedasticity remaining in the residuals of a SARIMA model. However the evaluation criteria do not necessarily show that the GARCH model improves the performance of the SARIMA model.

Keywords: nonlinear time series, Heteroscedasticity, GARCH, Engle's test, SARIMA model, seasonality, Box-Cox transformation

1. Introduction

During the last decades, hydrological time series analysis and modeling have received considerable attention for hydrologic prediction, simulation and forecasting. One of the main concerns of hydrological time series modeling is whether the hydrologic variable is governed by a linear or a nonlinear process through time.

The application of the linear time series models has been widely discussed in the literature for different hydrologic and climatic variables such as rainfall (Astel et al., 2004, Machiwal and Jha, 2008), streamflows (Ouarda et al., 1997; Adeloye and Montaseri, 2002), floods (Toth et al., 1999), droughts (Modarres, 2007) and water quality variables (Worrall and Burt, 1999).

Although there is a growing interest in nonlinear analysis of hydrologic time series, relatively few efforts have been done in this field. The Generalized Autoregressive Conditional Heteroscedasticity (GARCH) approach which is commonly used in modeling the time variation of the second order moment or the variance of financial time series can be an appropriate method for nonlinear modeling of hydrologic time series. Though the time variation of the variance of hydrologic variables has been mentioned in the literature, few studies have applied the GARCH approach to model this phenomenon in hydrologic variables.

In a pioneering study, Wang et. al., (2005) applied the GARCH approach to model the heteroscedasticity of daily and monthly streamflow time series of the upper Yellow River at Tangnaihai in China. They concluded that the conventional linear time series models, Autoregressive model and the deseasonalized Autoregressive Moving Average (ARMA)

model, are not sufficient to describe the time dependent variance of streamflow and a GARCH model needs to be fitted to the residuals of an ARMA model to capture the time variation behavior of the streamflow variance. Chen et al., (2008) applied linear ARMA and nonlinear ARCH models to model 10-day streamflows of the Wu-Shi River in Taiwan and verified that nonlinear time series models are superior to the traditional linear approaches such as an ARMA model. They reported an increase of the coefficient of efficiency (CE) from 0.28 for an ARMA model to 0.76 for an ARCH model while the mean absolute error (MAE) reduced from 60.45 for an ARMA model to 41.35 for an ARCH model. However, no seasonal or integrated time series model was applied to compare with a GARCH model in their study.

In another application of the GARCH model, Romilly (2005) fitted a Seasonal AutoRegressive Integrated Moving Average (SARIMA) model to the global mean monthly temperature and mentioned the existence of heteroscedasticity in the residuals and the need for applying a GARCH model to remove it from the residuals. However, the model comparison in this study indicated that the SARIMA model performs slightly better than the GARCH model for the global mean temperature time series modeling according to R^2 =0.772, mean absolute error (MAE=0.084) and the root mean squared error (RMSE=0.111) for the SARIMA model against R^2 =0.771, MAE= 0.084 and RMSE=0.112 for the GARCH model.

In addition to the fact of rare GARCH model applications in hydrology, rainfall is a missing variable in heteroscedastic time series modeling. Rainfall plays a vital role for a successful development and implementation of water resources tools to assess engineering and environmental problems such as flood control, reservoir operation,

hydropower generation, water quality control, etc. Henceforth, the efficient rainfall estimation, modeling and forecasting is a critical mission in hydrology, water resources, and environmental sciences.

The main objective of this study is therefore, to illustrate the advantages/disadvantages of a GARCH approach for modeling the monthly rainfall time series against the SARIMA model commonly used for monthly hydrologic time series modeling. In addition, the effect of Box-Cox transformation on stabilizing the variance of rainfall data and its effect on the advantages/disadvantages of a GARCH model is also investigated.

2. Methods

2.1. SARIMA model

The SARIMA time series model has a general multiplicative form, SARIMA(p,d,q)×(P,D,Q) (Hipel and McLeod, 1996). The first set of brackets contains the order of the nonseasonal parameters while the orders of the seasonal parameters are listed inside the second set of brackets.

The mathematical formulation of a SARIMA(p,d,q)×(P,D,Q) model can be written as the following:

$$\phi_p(B)\Phi_P(B^S)\nabla^d\nabla_S{}^D Y_t = \theta_q(B)\Theta_Q(B^S)a_t \qquad (1)$$

where Y_t is the observed time series, $\phi_p(B)$ is a polynomial of order p, $\theta_q(B)$ is a polynomial of order q, Φ_P and Θ_Q are the seasonal polynomials of order P and Q, ∇^d and ∇_S^D are the nonseasonal and seasonal differencing operators, respectively, B is the backward operator, S represents the number of seasons per year and a_t is an independent

identically distributed (i.i.d) normal error with a zero mean and standard deviation σ_a respectively.

The SARIMA model in eq. (1) is referred to as a multiplicative model as the nonseasonal and seasonal Autoregressive operators are multiplied together on the left hand side while the two Moving Average operators are multiplied together on the right hand side (Hipel and McLeod, 1996). Building the above SARIMA model from the observations includes three steps; model identification, parameter estimation and goodness-of-fit test or checking model adequacy. When the order of parameters of an initial model is identified according to the structure of the autocorrelation function (ACF) and its (significant) parameters are estimated using a method of estimation such as the method of moments or the method of maximum likelihood, the model adequacy should be checked. The details of model building can be found in Hipel and McLeod (1996).

For checking the adequacy of a model, the autocorrelation function (ACF) of the residuals of a SARIMA model (ε) is first inspected. It is well-known that for random and independent series of length n, the lag k autocorrelation coefficient is normally distributed with a mean zero and a variance 1/n and the 95% confidence limits are given by $\pm 1.96/\sqrt{n}$. If all autocorrelation coefficients fall within the confidence limits, the adequacy of the time series model is accepted.

More formally, the Ljung-Box lack-of-fit test (commonly called the portmanteau lack-of-fit test) is also used to test the adequacy of the SARIMA model. The Portmanteau lack-of-fit-test (Ljung and Box, 1978) computes a statistic Q, which is χ^2 distributed with $(L-p-q)$ degrees of freedom and is given by:

$$Q = N(N+2)\sum_{k=1}^{L}(N-k)^{-1}r_k^2(\varepsilon) \qquad (2)$$

Where N is the sample size, L is the number of autocorrelations of the residuals included in the statistic, which can be 15 to 25 for nonseasonal models and 2S to 4S for seasonal models (Hipel and McLeod, 1996) , and r_k is the sample autocorrelation of the residual time series, ε, at lag k. If the probability of Q is higher than $\alpha = 0.05$, there is a strong evidence that the residuals are time independent and the model is adequate. If this probability is less than $\alpha = 0.05$, it is reasonable to conclude that the residuals are time dependent and the model is inadequate and we need to repeat the process of model building to achieve an adequate model (Hipel and McLeod, 1996).

2.2. GARCH modeling approach

By applying a SARIMA model, the dependence of an observation at time t, Y_t, to the previous observations, Y_{t-1}, Y_{t-2}, …, Y_{t-p}, or the conditional mean, is considered for time series modeling. Though, the errors or the residuals of a SARIMA model may exhibit the adequacy, conditional variance or time dependent variance may exist in the residuals which can be captured by a GARCH model. This type of model is also called SARIMA-GARCH error model as the conditional variance of the residuals of a SARIMA model will be modeled by a GARCH approach. To describe a GARCH modeling approach, the Autoregressive Conditional Heteroscedasticity (ARCH) model is first described.

If the conditional mean is described by a SARIMA model, the Vth order of the Autoregressive Conditional Heteroscedasticity or the ARCH(V) model for the conditional variance of the residuals of a SARIMA model, (σ_t), is defined as the following (Engle, 1982):

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + ... + \alpha_V \varepsilon_{t-V}^2 = \omega + \sum_{i=1}^{V} \alpha_V \varepsilon_{t-V}^2 \qquad (3)$$

$$\varepsilon_t = \sigma_t e_t \qquad e_t \sim Normal(0,1) \qquad (4)$$

Where, ε_t indicates the residuals of the SARIMA model which are uncorrelated but have variances that change over time, e_t denotes a real valued i.i.d random variable with mean 0 and variance 1, independent of past realizations $(e_{t-1}, e_{t-2}, ...)$, $\alpha_1, ..., \alpha_V$ are the parameters of the ARCH model and ω is a constant (Wei, 2006). In this model, the variance of the error is time varying and depends on the V past errors, ε_{t-1}^2, $\varepsilon_{t-2}^2, \ldots, \varepsilon_{t-V}^2$ through their squares.

The generalized ARCH (GARCH) model, introduced by Bollerslev (1986), improves the original specification by adding lagged conditional variance, which acts as a smoothing term.

The Generalized Autoregressive Conditional Heteroscedasticity model, GARCH(V,M), is defined as follows :

$$\sigma_t^2 = \omega + \sum_{i=1}^{V} \alpha_V \varepsilon_{t-V}^2 + \sum_{j=1}^{M} \beta_M \sigma_{t-M}^2 \qquad (5)$$

$$\varepsilon_t = \sigma_t e_t \qquad e_t \sim Normal(0,1) \qquad (6)$$

Where $\alpha_1, ..., \alpha_V$ and $\beta_1, ..., \beta_M$ are the parameters of the GARCH(V,M) process. Therefore, the SARIMA-GARCH(V,M) model is one in which the conditional mean is described by a SARIMA model while the conditional variance is described by a GARCH(V,M) model.

2.3. Tests for the ARCH effect

Although the ACF of the residuals and the Ljung-Box lack-of-fit test are usually used for time series model adequacy analysis, the variance time dependent characteristics of the residuals can not be inspected by the ACF of the residuals. Bollerslev (1986) stated that the ACF of the standardized squared residuals (SSRs, hereafter) is useful for identifying and checking the heteroscedasticity of residuals. In order to check if the variance of the residuals is conditional on its past history or, in other words, if the residuals exhibit an ARCH effect, we apply the commonly used Ljung-Box and Engle's Lagrange Multiplier tests for the SSRs.

The Ljung-Box lack-of-fit test for SSRs is calculated through equation (2) but for the SSRs time series (ε^2) :

$$Q = N(N+2)\sum_{k=1}^{L}(N-k)^{-1}r_k^2(\varepsilon^2) \qquad (7)$$

Under the null hypothesis of no ARCH effect in the residuals, the test statistic is asymptotically chi-square distributed.

The Engle's Lagrange Multiplier test for the ARCH effect was proposed by Engle (1982). The test statistic is given by NR^2, where R is the sample multiple correlation coefficient computed from the regression of ε_t^2 on a constant and $\varepsilon_{t-1}^2,\ldots,\varepsilon_{t-V}^2$ and N is the sample size. The null hypothesis of no ARCH effect is accepted if the test statistic is asymptotically distributed as a chi-square distribution with v degrees of freedom. The test can also be used to investigate the GARCH effect of SSRs (Bollerslev, 1986).

2.4. Data transformation

Although a few studies have applied GARCH models in hydrology, the existing of conditional variance in hydrologic variables has already been addressed for a long time

and some methods have been used to stabilize this variance. The Box-Cox class of power transformation technique (Box and Cox, 1964) is the most common method for variance stabilization. The Box-Cox transformation can be expressed by the following equation:

$$Y_i^{B-C} = \begin{cases} \dfrac{Y_i^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(Y_i) & \text{if } \lambda = 0 \end{cases} \qquad (8)$$

Where Y_i^{B-C} is the Box-Cox transformed data, Y_i is the original time series and λ is the power parameter chosen to ensure that the transformed data are approximately Gaussian. The Box-Cox transformation is assumed to stabilize the variance of a time series so that no further ARCH effect is observed in the residuals of a time series model. The optimum value of λ is chosen based on Box-Cox normality plot which indicates the variation of λ against the correlation coefficient of a normal probability plot. The value of λ corresponding to the maximum correlation is then the optimal choice for λ (Lye, 1993).

2.5. Comparison approach

The time series modeling approaches used in this study are compared through a multi-criteria comparison by applying a set of evaluation metrics (Modarres, 2009). The evaluation metrics can be classified into three groups, metrics which calculate the absolute error, metrics which calculate the relative error and the dimensionless metrics (Dawson et al., 2007). In this study, the following evaluation metrics of the above categories are used:

- metrics for calculating absolute errors

Absolute Maximum Error (AME)

$$AME = \max\left(\left|Q_i - \hat{Q}_i\right|\right) \qquad (9)$$

Peak Difference (PDIFF)

$$PDIFF = \max(Q_i) - \max(\hat{Q}_i) \tag{10}$$

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum\left|Q_i - \hat{Q}_i\right| \tag{11}$$

Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{n}} \tag{12}$$

- metrics for calculating relative errors

Relative Absolute Error

$$RAE = \frac{\sum_{i=1}^{n}\left|Q_i - \hat{Q}_i\right|}{\sum_{i=1}^{n}\left|Q_i - \overline{Q}_i\right|} \tag{13}$$

Mean Relative Error

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Q_i - \hat{Q}_i}{Q_i}\right) \tag{14}$$

- Dimensionless metrics

Coefficient of determination (R-squared)

$$R^2 = \left[\frac{\sum_{i=1}^{n}(Q_i - \overline{Q})(\hat{Q}_i - \tilde{Q})}{\sqrt{\sum_{i=1}^{n}(Q_i - \overline{Q})^2 \sum_{i=1}^{n}(\hat{Q}_i - \tilde{Q})^2}}\right]^2 \tag{15}$$

Coefficient of Efficiency (CE)

$$CE = 1 - \frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{n}(Q_i - \overline{Q})^2} \tag{16}$$

Index of Agreement (IoAd)

$$IoAd = 1 - \frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{n}\left(\left|\hat{Q}_i - \overline{Q}\right| + \left|Q_i - \overline{Q}\right|\right)^2} \tag{17}$$

In these equations, Q_i is the observed time series, $\hat{Q}_i$ is the predicted time series, $\overline{Q}$ and $\tilde{Q}$ are the mean of the observed and predicted time series, respectively.

Thought the above criteria allow us to sort models based on their accuracy, it is also important to test whether the improvement of model accuracy between two competing models is statistically significant. To address this issue, a test statistic proposed by Diebold and Mariano (1995) which is already used in financial time series, is applied and reported for the first time in hydrological sciences. Although this test is usually used for out-of-sample forecasting accuracy analysis (i. e. Mohammadi and Su, 2010), here we use it for within-sample prediction accuracy testing. In our case, this test can indicate if there is a significant difference between SARIMA and SARIMA-GARCH's performances for predicting (estimating) rainfall time series.

Formally, let $e_{1,t}$ and $e_{2,t}$, t=1,…,n, denote model errors from SARIMA and SARIMA-GARCH models and $g(e_{1,t})$ and $g(e_{2,t})$ are their associated loss functions and $d_t = g(e_{1,t}) - g(e_{2,t})$. Diebold and Mariano (1995) defined the B statistics

$$B = \frac{\overline{d}}{\sqrt{\frac{s}{n}}} \sim N(0,1) \qquad (18)$$

Where $\overline{d}$ is the sample mean, s is the variance of loss differential and n is the number of observations. Under null hypothesis of zero mean loss differential, the null hypothesis of equal modeling accuracy is rejected if the test statistic is negative and statistically significant.

3. Applications and case studies

In this section we illustrate the application of the above methods for two monthly rainfall time series, one from a humid and the other from an arid region. These two rainfall data sets are selected to investigate and compare the ARCH effect of rainfall in the two regions and to evaluate the GARCH model efficiency and advantage for modeling the rainfall heteroscedasticity in the humid and arid regions.

3.1. Rainfall for a humid region

3.1.1. Data

The first case study includes the monthly time series of the Campsie station, ($54^{\circ}07'$N latitude and $114^{\circ}40'$W longitude), Alberta province (station elevation: 670.6 m), Canada. The mean annual rainfall in the Campsie station is 343 mm for the period of 1950-2008. The monthly rainfall time series of the Campsie station is illustrated in Figure (1a). The monthly distribution of the mean and standard deviation (STDEV) of rainfall (Figure (1b)) indicates seasonality in both rainfall depth and variance in the Campsie station. In the following sections, the results of time series models fitted to both original and transformed data are given and discussed.

3.1.2. Results for modeling original data

Since the ACF of rainfall time series (Figure (2a)) indicates a strong seasonal structure, the SARIMA$(1,0,1)_{12}$ model is fitted to rainfall data. This model includes two seasonal autoregressive and moving average parameters of order 1 which are significant at 95% level and does not include any non-seasonal parameters.

The adequacy of the SARIMA$(1,0,1)_{12}$ model is verified using the Ljung-Box statistics and the ACF of the residuals. The ACF of the residuals (Figure (2b)) indicates no autocorrelation structure in the residuals as all coefficients fall within the confidence limit. No seasonality is observed in the residuals as the coefficients at lags k=12, 24, 36,... also fall within the confidence interval.

The results of the Ljung-Box test (Figure (3a)) also show the adequacy of the model as the p-values of Ljung-Box statistic (Q) exceed the critical value ($\alpha = 0.05$). Therefore, the null hypothesis of no autocorrelation structure in the residuals can not be rejected.

Although the residuals seem statistically uncorrelated according to the ACF and no seasonal structure is observed in the residuals, the ACF of the SSRs is investigated to check the presence of the ARCH effect. The ACF of the SSRs has been given in Figure (3b) which illustrates the existence of an ARCH effect in the SSRs of the SARIMA$(1,0,1)_{12}$ model. A seasonal heteroscedasticity is also observed in the SSRs of the SARIMA$(1,0,1)_{12}$ model as the autocorrelation coefficients exceed the confidence limit at lags k=12, 24, 36, ... The p-values of the Engle's test for the ARCH effect of SSRs of the SARIMA$(1,0,1)_{12}$ model are presented in Figure (3c) which illustrate the presence of heteroscedasticity in the SSRs of the SARIMA$(1,0,1)_{12}$ model. For all lags, the p-values of the test are less than the critical value ($\alpha = 0.05$) and the null hypothesis of no ARCH effect is rejected. In other words, the heteroscedasticity of the residuals of SARIMA$(1,0,1)_{12}$ model is significant and we need to apply a GARCH approach to remove it from the residuals of the SARIMA$(1,0,1)_{12}$ model.

As seasonality is also observed in the SSRs of the SARIMA$(1,0,1)_{12}$ model, we first try to remove this seasonal variance by adding a seasonal differencing parameter and fitting a

SARIMA$(1,1,2)_{12}$ model to rainfall data. This model includes a seasonal autoregressive parameter, two seasonal moving average parameters and a seasonal differencing operating, $\nabla_S{}^D$.

The ACF of the residuals is first investigated for validating the adequacy of the model. The ACF of the residuals (Figure (4a)) indicates the appropriateness of the model as the autocorrelation coefficients are within the confidence interval of ACF. The Ljung-Box test for the residuals also indicates the adequacy of the SARIMA$(1,1,2)_{12}$ model for the rainfall time series of Campsie station (Figure (4b)). All p-values of the test lie outside the critical level ($\alpha = 0.05$) and reject the existence of an autocorrelation structure in the residuals.

For testing the ARCH effect, the Engle's test is used for the SSRs of the SARIMA$(1,1,2)_{12}$ model. The p-values of the test (Figure (4c)) show that the null hypothesis of no ARCH effect in the SSRs of the SARIMA$(1,1,2)_{12}$ model is rejected and no improvement of the seasonal heteroscedasticity stabilization is also observed in the SSRs of the SARIMA$(1,1,2)_{12}$ model. Therefore, adding a seasonal differencing parameter can not remove the seasonal heteroscedasticity and we try fitting a GARCH model to the residuals of both SARIMA$(1,0,1)_{12}$ and SARIMA$(1,1,2)_{12}$ models to stabilize the heteroscedasticity.

A GARCH(1,1) model is first fitted to the residuals of the SARIMA$(1,0,1)_{12}$ model. To test the existence of an ARCH effect, the ACF of the SSRs of the SARIMA$(1,0,1)_{12}$–GARCH(1,1) (given in Figure (5a)) is inspected. The figure indicates some reduction in the autocorrelation structure of the residuals comparing to the ACF of SSRs for SARIMA$(1,0,1)_{12}$ model given in Figure (3b). For example, the autocorrelations at lags

k=1 to k=10 and k=13 to k=20 in the ACF of the SSRs of SARIMA$(1,0,1)_{12}$-GARCH(1,1) (Figure (5a)) are within the confidence interval while these autocorrelations are significant in Figure (3b). However, the seasonality still exists in the SSRs of SARIMA$(1,0,1)_{12}$-GARCH(1,1).

The p-values of the Engle's test for the ARCH effect are also shown in Figure (5b). It can be seen that except for the first four lags (k=1- 4), all p-values are significantly below the critical value, ($\alpha = 0.05$), and the null hypothesis of no ARCH effect is rejected. Therefore the SARIMA$(1,0,1)_{12}$-GARCH(1,1) model is not able to remove the conditional variance, specially the seasonal conditional variance, of the monthly rainfall of Campsie station.

In the following, a GARCH(1,2) model is fitted to the residuals of the SARIMA$(1,1,2)_{12}$ model. The ACF of the SSRs of the SARIMA$(1,1,2)_{12}$-GARCH(1,2) model is presented in Figure (6a). This figure illustrates no significant autocorrelation structure as all the coefficients at different lags are within the confidence interval up to lag k=39. The significant autocorrelation at lag k=39 could be considered as a random effect in the heteroscedasticity of rainfall.

The major feature of Figure (6a) is that the seasonal autocorrelation coefficients at lags k=12, 24, 36 and 48 are not significant. This is an important advantage of using a seasonal ARIMA model with an appropriate differencing parameter together with a GARCH approach to remove the seasonal heteroscedasticity from the seasonal hydrologic time series. The p-values of the Engle's test given in Figure (6b) also verifies the advantage of GARCH modeling approach together with a seasonal differencing to stabilize the conditional variance of rainfall time series. All p-values are larger than the

critical value ($\alpha = 0.05$) and the null hypothesis of no ARCH effect in the residuals of SARIMA$(1,1,2)_{12}$–GARCH (1,2) model can not be rejected.

3.1.3. Results for modeling transformed data

In addition to the investigation of the effect of seasonal differencing on stabilizing the seasonal heteroscedasticity, the Box-Cox transformation is also used to stabilize the variance of the rainfall time series. The power parameter of the Box-Cox transformation is $\lambda = 0.112$.

To ensure investigating the effect of a seasonal differencing on removing the heteroscedasticity from residuals, two SARIMA models are fitted to the transformed data. In the first model no seasonal differencing parameter is added while for the second model a seasonal differencing parameter is also included.

The first SARIMA model for Box-Cox transformed rainfall time series of Campsie station includes two seasonal and two nonseasonal parameters of order 1 to have the SARIMA(1,0,1) ×(1,0,1) model. The ACF and the p-values of Ljung-Box test for the residuals of the model, given in Figures 7(a) and 7(b), respectively, indicate the acceptance of the null hypothesis of model adequacy. However, if the SSRs of the model are examined for an ARCH effect, a seasonal heteroscedasticity can be observed in the residuals of SARIMA(1,0,1)×(1,0,1) model (Figure 7(c)) as the autocorrelation coefficients at lags k=12, 24 and 36 are significant while other autocorrelation coefficients are within the confidence level. The p-values of Engle's test for SSRs are also given in Figure (7d). It can be seen that most of the p-values, but not all of them, are larger than the critical value ($\alpha = 0.05$). Some p-values at lags k=12 to k=16 are less

than the critical value and suspect the existence of an ARCH effect in the residuals. The Engle's test indicates that the Box-Cox transformation has stabilized the heteroscedasticity of the residuals to a very low level but some seasonal conditional variances still remain in the residuals. This suggests the need of a GARCH model to stabilize the remaining heteroscedasticity of the SARIMA(1,0,1)×(1,0,1) model fitted to the Box-Cox transformed data. Therefore, we fit a GARCH(1,1) to the residuals of SARIMA(1,0,1)×(1,0,1) model. The residuals of the SARIMA(1,0,1)×(1,0,1)-GARCH(1,1) model are then checked for the ARCH effect inspecting the ACF of SSRs and the p-values of the Engle's test. The results are given in Figure (8). This figure illustrates that the autocorrelation coefficients are within the confidence level of ACF and are not significant. The seasonal autocorrelation coefficients at lags k=12, 24, 36,... are also insignificant. The result of the Engle's test also indicates no ARCH effect in the SSRs of the GARCH model. Therefore, the SARIMA(1,0,1)× (1,0,1)-GARCH(1,1) is sufficient for modeling the conditional variance of rainfall time series of Campsie station. From the above analysis, a GARCH model seems to be required for modeling the heteroscedasticity in the residuals of the SARIMA(1,0,1)× (1,0,1) model fitted to the Box-Cox transformed rainfall without using a seasonal differencing parameter. However, the seasonal differencing parameter is also included in the model to check if the seasonal differencing can remove the heteroscedasticity from residuals so that no ARCH effect remains to be modeled by a GARCH approach. Adding a seasonal differencing parameter to the SARIMA(1,0,1)×(1,0,1) model, we fit a SARIMA(1,0,1)×(1,1,1) model to the Box-Cox transformed rainfall time series. The adequacy of the model is confirmed by inspecting the ACF of the residuals and using the Ljung-Box test.

The ACF and the p-values of the Engle's test of SSRs (Figure 9) indicate that if an appropriate seasonal differencing is used with a Box-Cox transformation, the seasonal heteroscedasticity will be stabilized. Neither nonseasonal nor seasonal heteroscedasticity can be observed in the ACF of the SSRs of SARIMA(1,0,1)×(1,1,1) model fitted to Box-Cox transformed rainfall data. The p-values of Engle's test are also larger than the critical value and no ARCH effect exists in the residuals. Therefore, we agree that an appropriate transformation and deseasonlization approach would result in removing the ARCH effect from the residuals of the SARIMA model for monthly rainfall of Campsie station and no further GARCH model is necessary for stabilizing the conditional variance of rainfall data.

3.2. Rainfall for an arid region

3.2.1. Data

The second case study deals with modeling the monthly rainfall time series of the Isfahan station ($32^{\circ}37'$N latitude and $51^{\circ}40'$E longitude), Isfahan province (Elevation: 1550.4 m), Iran. The station is located in the semi arid region of Iran with the mean annual rainfall of 122.8 mm for the period 1951-2005. The Monthly rainfall time series of the Isfahan station and the monthly distribution of the rainfall mean and standard deviation are given in Figure (10a) and Figure (10b), respectively.

In the following two sections, we present the results of time series modeling for the original and Box-Cox transformed rainfall data of Isfahan station to discuss the ARCH effect of the rainfall in an arid region.

3.2.2. Results for modeling original data

The ACF of the rainfall (Figure (11a)) indicate the seasonal behavior of rainfall in the Isfahan station. Therefore, the SARIMA(0,0,1)×(1,0,1) model is first fitted to rainfall data. This model has three parameters, one nonseasonal moving average parameter and two seasonal autoregressive and seasonal moving average parameters, all of order 1. The ACF of the residuals of this model (Figure 11b) and the p-values of Ljung-Box test (Figure 11c) indicate the adequacy of the model. The existence of an ARCH effect in the SSRs of the above model is investigated through inspecting the ACF of SSRs and using the Engle's test. The ACF (Figure 12a) indicates neither nonseasonal nor seasonal heteroscedasticity in the residuals as the autocorrelation coefficients are within the confidence limits. However, at lags k=8 and k=20, two autocorrelation coefficients fall outside the confidence level on the other hand. The p-values of Engle's test (Figure 12b) proves no ARCH effect in the SSRs of the SARIMA(0,0,1)×(1,0,1) model and suggests that the significant autocorrelation coefficients at lags k=8 and k=20 could be considered as random effects on conditional variance of rainfall in Isfahan station.

Adding a seasonal differencing to the above model to have a SARIMA(0,0,1)×(1,1,1) model shows the same adequacy of the model according to the ACF and p-values of Ljung-Box test and the same ARCH structure in the residuals as those for the SARIMA (0,0,1)×(1,0,1) model. No seasonal heteroscedasticity is observed and the autocorrelation coefficients are still significant at lags k=8 and k=20.

The significant autocorrelation coefficients at lags k=8 and k=20 could be considered as random conditional variances of rainfall in Isfahan station. These random effects could be related to an uncertainty in climate conditions in the arid regions. The timing and amount

of precipitation indicate a high irregular fluctuation in the arid regions and may impose random effects on the conditional variance of rainfall.

One can accept that the SARIMA(0,0,1)×(1,1,1) is enough for modeling the rainfall time series of the Isfahan station at this level considering the lags k=8 and k=20 as random ARCH effects and does not feel necessary to use a GARCH model to remove these random effect. Nevertheless, we continue our modeling approach by fitting a GARCH model to the residuals of the SARIMA model in order to investigate the advantage/disadvantage of the GARCH models.

Trying to remove the random ARCH effect from the residuals by fitting a GARCH model, the SARIMA(0,0,1)×(1,0,1)–GARCH(0,2) and SARIMA(0,0,1)×(1,1,1)–GARCH(0,2) are used. As the ACF and the Engle's test for the residuals of both above models are the same, we illustrate the ACF of the SSRs and the p-values of Engle's test for SARIMA(0,0,1)×(1,1,1)–GARCH(0,2) model in Figure 13(a) and 13(b), respectively. No ARCH effect is observed in the residuals of SARIMA(0,0,1)×(1,1,1)-GARCH (0,2) model as all autocorrelation coefficients fall within the confidence interval of the ACF and all p-values are larger than the critical value. It is important to note that neither nonseasonal nor seasonal heteroscedasticity is observed in the residuals of the GARCH model and the random conditional variances at lags k=8 and k=20 have also been removed from the residuals. This suggests the appropriateness of a GARCH approach for modeling heteroscedasticity of rainfall time series in the arid region where rainfall indicates irregular temporal fluctuations and variation.

3.2.3. Results for modeling transformed data

Time series modeling of the original rainfall of Isfahan indicates the advantage of using a GARCH model to remove conditional variance of the residuals of a SARIMA model. However, we also apply a transformation method to check if it will reduce the heteroscedasticity of rainfall data that no ARCH effect remains to be modeled by a GARCH approach.

The Box-Cox transformation method with $\lambda = 0.056$ is applied to transform rainfall of the Isfahan station. Two SARIMA models with and without differencing parameter are then fitted to transformed data in order to see the adequacy of Box-Cox transformation for removing heteroscedasticity from the residuals of the SARIMA models.

The SARIMA model fitted to the transformed data without applying seasonal differencing is SARIMA(0,0,2)×(1,0,2). The model is adequate according to the ACF (Figure 14a) and p-values of Ljung-Box (Figure 14b) of the residuals.

The heteroscedasticty remaining in the residuals of the above model is investigated by checking the ACF of SSRs (Figure 15(a)) and using the Engle's test for an ARCH effect (Figure15(b)). The ACF of the SSRs indicates the existence of both nonseasonal and seasonal heteroescedasticity. The results of Engle's test also indicate an ARCH effect in the residuals of SARIMA(0,0,2)×(1,0,2) model fitted to the Box-Cox transformed data.

By adding a seasonal differencing operator to the above model, the effect of seasonal differencing on stabilizing the seasonal heteroscedasticity is examined. The SARIMA(0,0,2)×(1,1,2) model is also adequate according to Ljung-Box test for the residuals. However, both seasonal and nonseasonal heteroscedasticity are still observed in the SSRs (Figure 16(a)). The Engle's test indicates some improvement of modeling

heteroscedasticity of the residuals by adding a seasonal differencing to the model since some of p-values are larger than critical value ($\alpha = 0.05$) in Figure 16(b). However, ARCH effect is still remaining in the residuals and we need to apply a GARCH approach to model this heteroscedasticity in the residuals.

The GARCH model is then fitted to the residuals to have a SARIMA(0,0,2)×(1,0,2)–GARCH(2,3) model. To check the ARCH effect of the residuals, the ACF of SSRs and the p-values of Engle's test are given in Figure (17). The ACF of SSRs (Figure 17(a)) indicates the adequacy of the GARCH approach for modeling the heteroscedasticity of rainfall time series. However, a very weak seasonal heteroscedasticity is observed as the seasonal autocorrelation coefficients at lags k=12 and k=24 fall outside the confidence level. If we fit a GARCH model to the residuals of the SARIMA(0,0,2)×(1,1,2) model to have a SARIMA(0,0,2)×(1,1,2)-GARCH(2,2) model, the evidence of the weak heteroscedasticity is also removed. The ACF and the p-values of the Engle's test for the above model are given in Figure (18a) and Figure 18(b), respectively. It is clear that the GARCH model has removed both seasonal and nonseasonal heteroscedasticity from the residuals.

Modeling the transformed data suggests that Box-Cox transformation and seasonal differencing can not remove conditional variance from the residuals of Isfahan station without applying a GARCH model.

4. Model comparison

This section provides the criteria for model performance evaluation and cross-comparison of the SARIMA and GARCH models fitted to rainfall data at the Campsie and Isfahan stations.

Firstly, the criteria for the Campsie station are considered (Table 1). For the original rainfall data, the SARIMA model with seasonal differencing parameter, SARIMA$(1,1,2)_{12}$, indicates a slightly better performance than other models according to RMSE, AME and PDIFF. It suggests the SARIMA$(1,1,2)_{12}$ is a better model for predicting peak rainfall in the Campsie station than other models. Though, the performance of the SARIMA and GARCH models look almost the same based on the error and dimensionless metrics, the DM statistic indicates a significant difference between SARIMA and SARIMA-GARCH's performances with and without seasonal differencing. Nevertheless, the model with seasonal differencing parameter indicates a more negative DM statistics comparing to the model without differencing.

On the other hand, for Box-Cox transformed data, the SARIMA and GARCH models seem to relatively outperform the SARIMA and GARCH models fitted to the original rainfall data according to dimensionless metrics. However, no improvement in model accuracy is observed comparing the GARCH model to the SARIMA model for Box-Cox transformed rainfall time series of the Campsie station according to DM statistic. This suggests that the application of a GARCH model for rainfall of the Campsie station, as a sample of humid region, does not necessarily improve the performance of a SARIMA model after Box-Cox transformation

Looking at the criteria for the Isfahan station in Table 2 also suggests that a GARCH model does not improve the performance of the SARIMA model fitted to the original rainfall time series.

However, for Box-Cox transformed rainfall time series, the GARCH models perform better than the SARMA models. Both global metrics, R^2 and RMSE, show the best performance for SARIMA(0,0,2)×(1,1,2)-GARCH(2,2) model. Based on other error metrics, the GARCH model gives also the lowest error. The dimensionless metrics also suggest a better agreement between the GARCH modeled and observed rainfall than that of the SARIMA model. However, the null hypothesis of equal accuracy is rejected for all SARIMA-GARCH against SARIMA models. This suggests that there is not a remarkable difference in the performance of the SARIMA-GARCH against a SARIMA model for Isfahan rainfall time series. It is also important to note that the performance of the time series models fitted to transformed rainfall data, both in arid and humid regions, is much better than the models fitted to the original data, especially for rainfall of the Isfahan station, for which the performance of the model is doubled after Box-Cox transformation according to R^2 and CE.

Nevertheless, the DM test results indicate that the null hypothesis of equal modeling accuracy is only rejected for SARIMA-GARCH models for original rainfall time series at Campsi station. Therefore, one can see that the main advantage of a GARCH model is the ability of capturing the heteroscedasticity of the residuals of a SARIMA model (as was also reported by Romilly (2005) for global mean temperature time series modeling) and there is no guarantee to have a better performance by a GARCH model, at least for our case study.

5. Discussion

The volatility, the time-dependent variance or the conditional heteroscedasticity, is usually neglected in hydrologic time series modeling. The present study shows that although the seasonal ARIMA model seems to be adequate for modeling the conditional mean of the monthly rainfall time series and its seasonal variation, a seasonal ARCH effect remains in the residuals. Fitting a SARIMA model to the sample rainfall data set used in this study we noticed the inadequacy of SARIMA models for capturing the seasonal conditional variance of monthly rainfall for both sample data sets coming from the humid and arid regions even by adding a seasonal differencing parameter to the model.

Adding a seasonal differencing parameter to a SARIMA model does not remove the seasonal heteroscedasticity from the residuals of the model. This phenomenon is observed for both rainfall data from humid and arid regions. However, the Box-Cox transformation together with a seasonal differencing seemed to be an efficient approach for capturing seasonal heteroscedasticity of rainfall data of the humid region (the Campsie station) but not a sufficient method for rainfall data from the arid region (the Isfahan station). However, these results are explanatory and future analyses are necessary to generalize them.

This suggests that there may be a very irregular conditional variance in rainfall data in arid regions which can not be removed by a seasonal differencing and a Box-Cox transformation. The remaining of an ARCH effect in the residuals of the SARIMA model for Box-Cox transformed rainfall data in the arid region may be due to the perturbations of temperature and evapotranspiration fluctuations and non-equilibrium interaction of the

earth and atmospheric components in the arid regions. The reasons for the irregular ARCH effect in the rainfall of the Isfahan station and its relationship to (local or regional) climate fluctuations should be carefully examined using a larger database in different arid regions of the world.

The investigation of the causes of the seasonal ARCH effect and the causes of the inadequacy of the linear models commonly used for seasonal hydrologic time series modeling is out of the scope of the present study. However, it can be assumed that the seasonal heteroscedasticity may be the result of the seasonality of the atmospheric and climatic factors and components which influence the seasonal variance of hydrological variables. The irregular pattern of the climate fluctuations in the arid regions may result in an irregular conditional variance of rainfall in these regions as illustrated by the Isfahan rainfall case study. The above statement on irregular conditional variance requires a more careful investigation which may be the topic of future studies.

Although the GARCH model indicates the capability of modeling conditional variance, it does not improve the efficiency of the SARIMA models, especially for rainfall data in the arid region. However, the performance of both SARIMA and GARCH models fitted to transformed rainfall data in the arid region is twice better than that of the models fitted to the original rainfall data according to multi-criteria error evaluation. This can be an important aspect of the rainfall time series modeling in the arid regions that should be carefully considered. The Box-Cox transformation seems to be an effective method to reduce the (hidden) variance of rainfall in the arid regions which also improves the performance of time series models.

It should also be mentioned that the Diebold-Mariano (DM) test did not show a notable difference between SARIMA and SARIMA-GARCH prediction performances, except for the non-transformed rainfall time series from humid region.

The results of the present study suggest that using the GARCH approach together with an appropriate transformation technique may increase the performance of the time series models in some cases, such as our example of the Campsie station and may stabilize the heteroscedasticity of rainfall time series. However, from parsimonious point of view, and as a disadvantage of GARCH modeling approach, it seems that adding more parameters into rainfall time series models by a GARCH model may not guarantee achieving better rainfall prediction accuracy.

6. **Conclusions and future work**

This study illustrated the GARCH modeling approach for a rainfall time series with seasonal variation and showed the advantage of a GARCH approach to model the conditional variance of rainfall data. In order to better understand the advantage/disadvantage of the GARCH modeling approach for hydrologic time series modeling, especially for rainfall in arid regions, the use of a larger data base from different regions of the world is strongly recommended in future studies. Simulation-based studies are also necessary to confirm and generalize the results of the present study. Space limitations prevented this from being done in the current study, and this should be the topic of the future research efforts. In addition, the advantage/disadvantage of GARCH models over SARIMA models for out-of sample forecasting would be important to be considered in the future investigations.

An interesting topic for future GARCH modeling efforts deals with the modeling non-stationary hydrological time series through climate change conditions, especially time series with change in higher order moments such as the variance, skewness and kurtosis. The application of multivariate GARCH models in order to investigate the shift of the conditional variance from one variable such as rainfall into another variable such as streamflow is also an interesting topic for future works. Using a combination of GARCH models with other time series modeling approaches such as fractional and periodic time series models would be useful for modeling persistence and periodicity in the variance of different hydrologic and climatic variables. Coupling GARCH models with smoothing approaches and nonparametric methods would also helpful in modeling conditional variance of the hydrologic time series.

Acknowledgements

The financial support for this study was graciously provided by the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Canada Research Chair (CRC) Program. The authors also thank the anonymous reviewer for useful comments which improved the quality of the paper.

References

Adeloye AJ. Montaseri M. 2002. Preliminary streamflow data analyses prior to water resources planning study. Hydrological Sciences Journal **47:** 679–692.

Astel A, Mazerski J, Polkowska Z, Namieśnik J. 2004. Application of PCA and time series analysis in studies of precipitation in Tricity (Poland). Advances in Environmental Research **8:** 337–349.

Bollerslev T. 1986. Generalized Autoregressive Conditional Heteroscedasticity. Journal of Econometrics **31**: 307–327.

Box GE, Cox DR. 1964. An analysis of transformed data. Journal of Royal Statistical Society B **39**: 211–252

Chen CH, Liu CH, Su HC. 2008. A nonlinear time series analysis using two-stage genetic algorithms for streamflow forecasting. Hydrological Processes **22**: 3697–3711

Dawson CW, Abrahart RJ, See LM. 2007. Hydrotest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. Environmental Modeling and software, **22**: 1034-1052

Diebold FX, Mariano, RS. 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics. 13: 253–263.

Engle RF. 1982. Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation. Econometrica **50**:987-1008.

Hipel KW, McLeod AE. 1996. Time series modeling of water resources and environmental systems. Elsevier, Amsterdam, the Netherlands

Ljung GM. Box GEP. 1978. On a Measure of Lack of Fit in Time Series Models. Biometrika **65**: 297-303.

Lye, L.M. 1993. A technique for selecting the Box-Cox transformation in flood frequency analysis. Canadian Journal of Civil Engineering. 20: 706-766

Machiwal D. Jha M. 2008. Comparative evaluation of statistical tests for time series analysis: application to hydrological time series. Hydrological Sciences Journal. **53**: 353-366

Mohammadi H, Su L. 2010. International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models. Energy Economics. 32: 1001-1008.

Modarres R. 2007. Streamflow drought time series forecasting. Stochastic Environmental Research and Risk Assessment. **21**: 223–233

Modarres R. 2009. Multi-criteria validation of artificial neural network rainfall-runoff modeling. Hydrology and Earth System Sciences. **13**: 411-421

Ouarda TBMJ. Labadie JW. Fontane G. 1997. Indexed sequential hydrologic modeling for hydropower capacity estimation. Journal of the American Water Resources Association. **33**: 1-13.

Romilly P. 2005. Time series modeling of global mean temperature for managerial decision-making. Journal of Environmental Management. **76**: 61-70

Toth E. Montanari A. Brath A. 1999. Real-time flood forecasting via combined use of conceptual and stochastic models. Physics and Chemistry of the Earth (B). **24**: 793-798.

Wang W, Vrijling JK. Van Gelder PHAJM. Ma J. 2005. Testing and modeling autoregressive conditional heteroskedasticity of streamflow processes. Nonlinear Processes in Geophysics. **12**: 55-66

Wei WWS. 2006. Time series analysis, univariate and multivariate methods. Pearson Education, Inc. USA

Worrall F. Burt TP. 1999. A univariate model of river water nitrate time series. Journal of Hydrology. **214**: 74-90

TABLE CAPTION

Table 1. Model criteria for the Campsie rainfall time series

Table 2. Model criteria for the Isfahan rainfall time series

FIGURE CAPTIONS

Figure 1. a) monthly rainfall time series and b) monthly distribution of rainfall mean and standard deviation of the Campsie station for 1956-2008

Figure 2. a) ACF of monthly rainfall time series and b) ACF of the residuals of SARIMA$(1,0,1)_{12}$ model for the Campsie station

Figure 3. a) p-values of Ljung-Box test of the residuals, b) Autocorrelation function and c) p-values of Engle's test for the SSRs of SARIMA$(1,0,1)_{12}$ model

Figure 4. a) Autocorrelation function of the residuals, b) p-values of Ljung-Box test of the residuals and c) p-values of Engle's test for the SSRs of SARIMA$(1,1,2)_{12}$ model

Figure 5 a) Autocorrelation function and b) p-values of Engle's test for the SSRs of SARIMA$(1,0,1)_{12}$–GARCH(1,1) model for rainfall of the Campsie station

Figure 6. a) Autocorrelation function and b) p-values of Engle's test for the SSRs of SARIMA$(1,1,2)_{12}$–GARCH(1,2) model for rainfall of the Campsie station

Figure 7. a) Autocorrelation function of the residuals b) p-values of Ljung-Box test for the residuals c) Autocorrelation function of SSRs and d) p-values of Engle's test of SSRs of SARIMA$(1,0,1)\times(1,0,1)$ model for Box-Cox transformed rainfall of the Campsie station

Figure 8. a) Autocorrelation function and b) p-values of Engle's test of SSRs of SARIMA$(1,0,1)\times(1,0,1)$–GARCH(1,1) model for Box-Cox transformed rainfall of Campsie station

Figure 9. a) Autocorrelation function and b) p-values of Engle's test of SSRs of SARIMA$(1,0,1)\times(1,1,1)$ model for Box-Cox transformed rainfall of the Campsie station

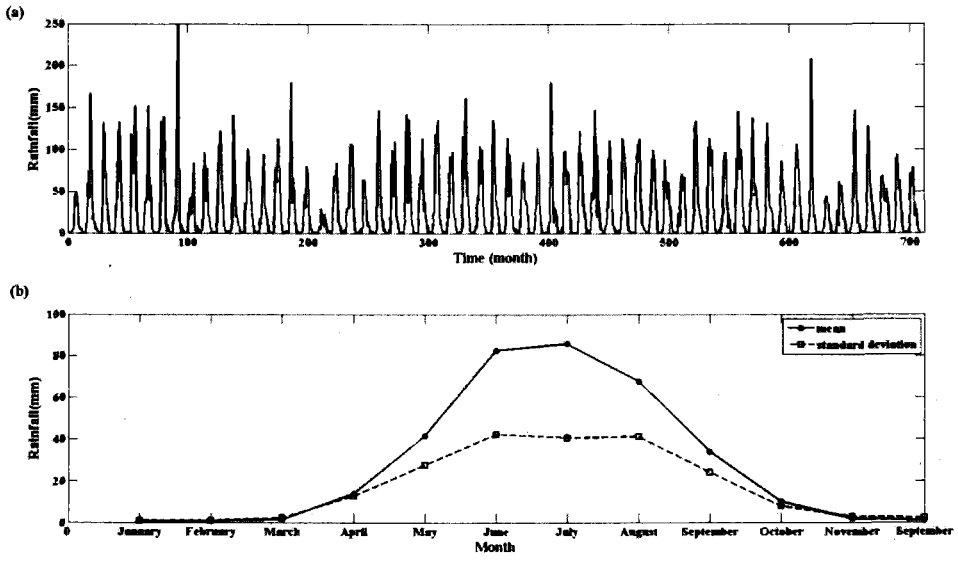


Figure 1. a) monthly rainfall time series and b) monthly distribution of rainfall mean and standard deviation of the Campsie station for 1956-2008

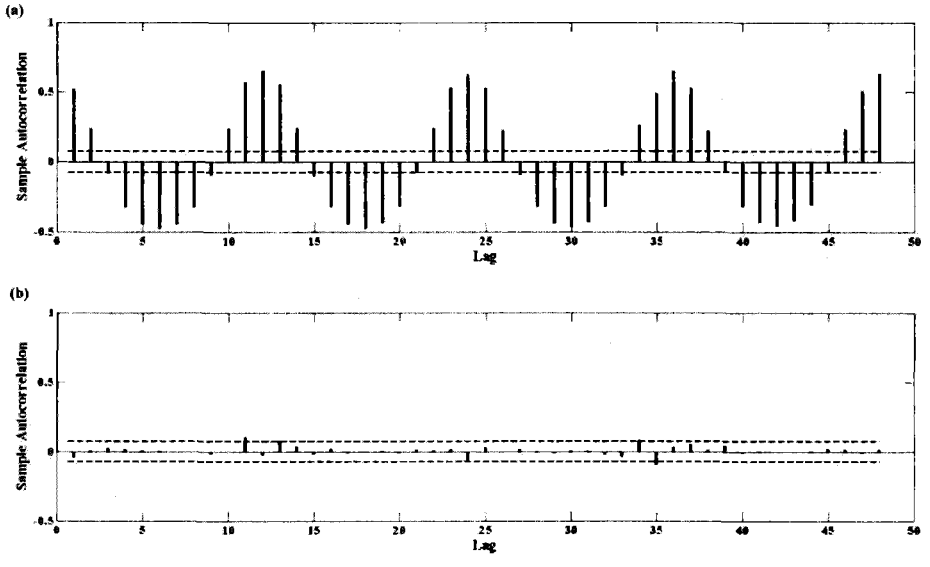


Figure 2. a) ACF of monthly rainfall time series and b) ACF of the residuals of SARIMA(1,0,1)12 model for the Campsie station

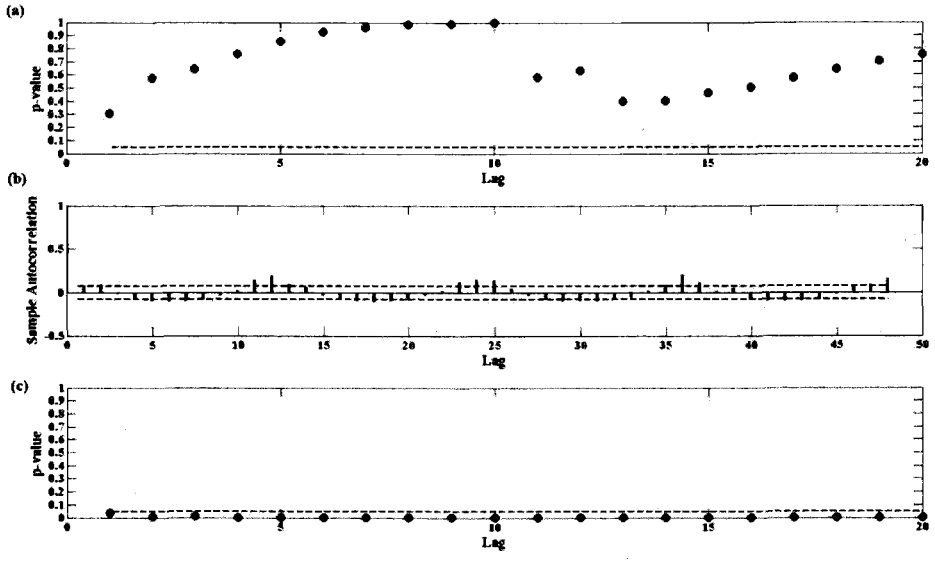


Figure 3. a) p-values of Ljung-Box test of the residuals, b) Autocorrelation function and c) p-values of Engle's test for the SSRs of SARIMA(1,0,1)12 model

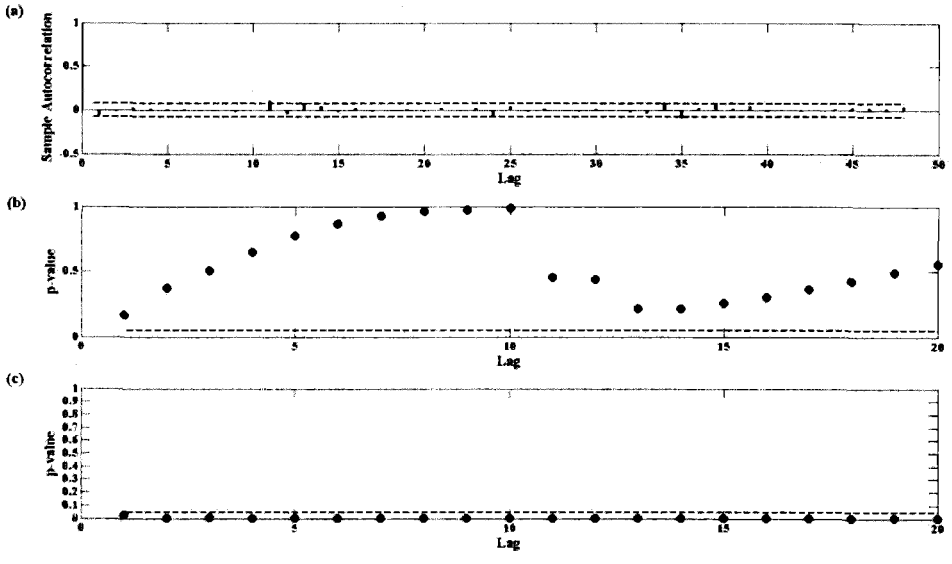


Figure 4. a) Autocorrelation function of the residuals, b) p-values of Ljung-Box test of the residuals and c) p-values of Engle's test for the SSRs of SARIMA(1,1,2)12 model

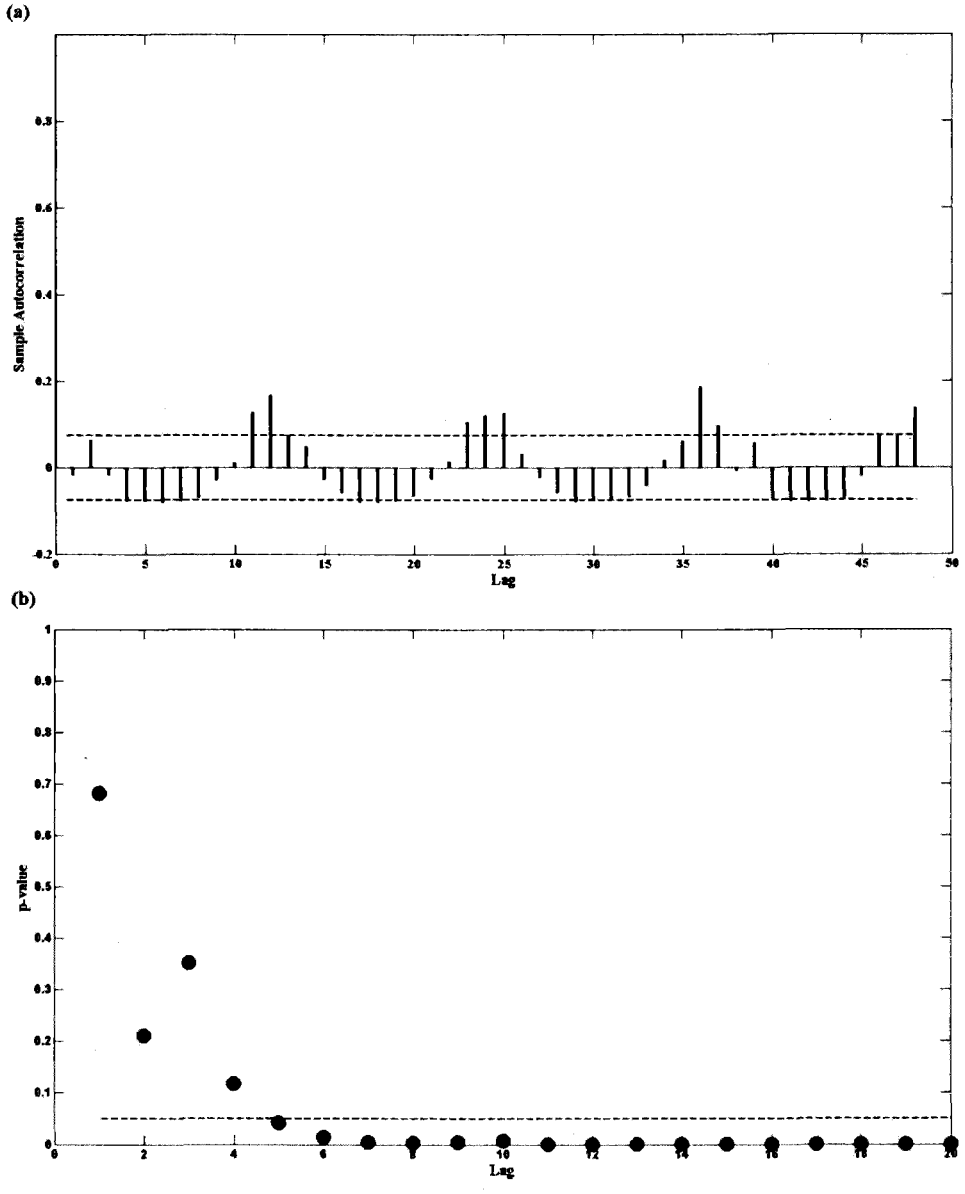


Figure 5. a) Autocorrelation function and b) p-values of Engle's test for the SSRs of SARIMA(1,0,1)12–GARCH(1,1) model for rainfall of the Campsie station

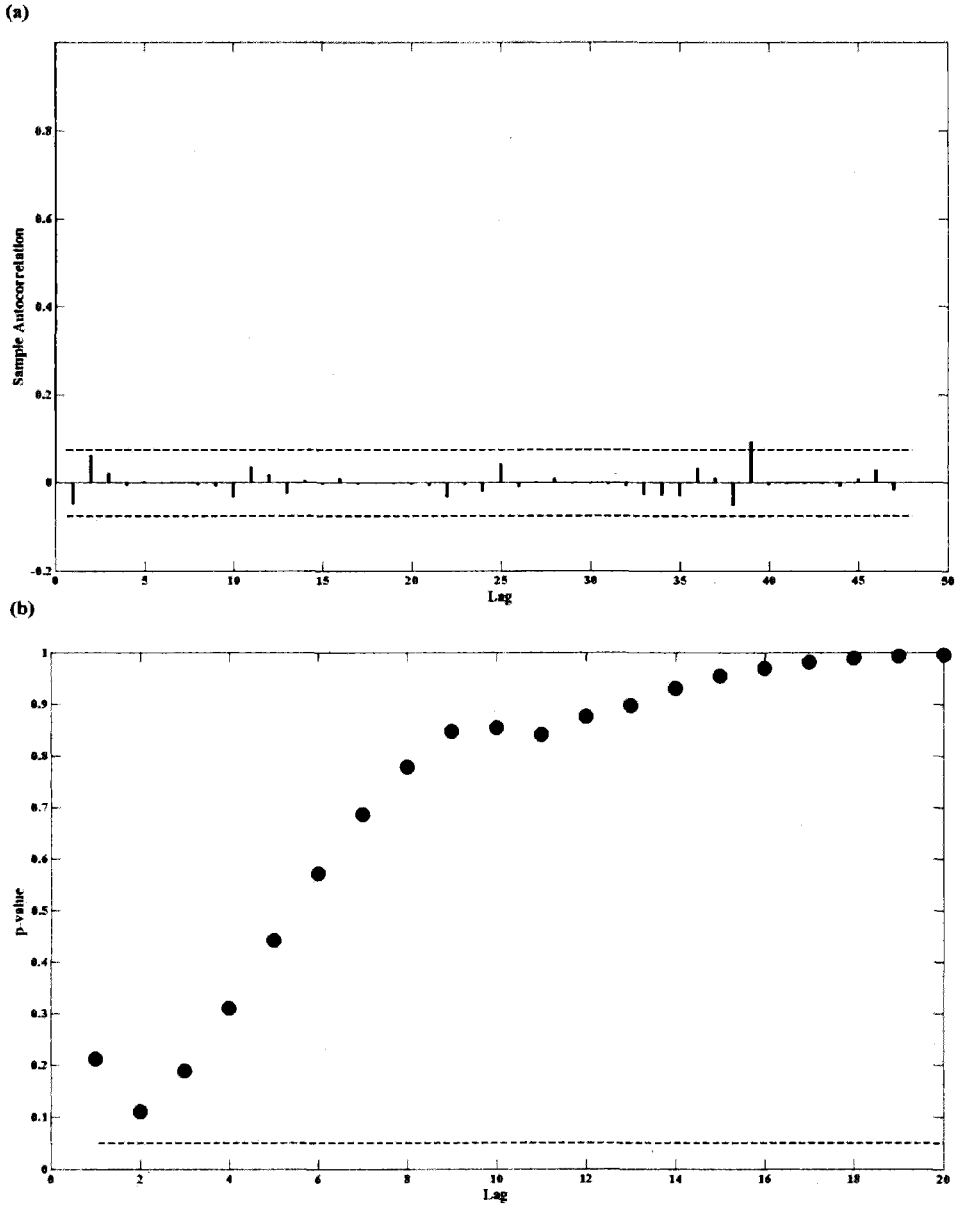


Figure 6. a) Autocorrelation function and b) p-values of Engle's test for the SSRs of SARIMA(1,1,2)12–GARCH(1,2) model for rainfall of the Campsie station

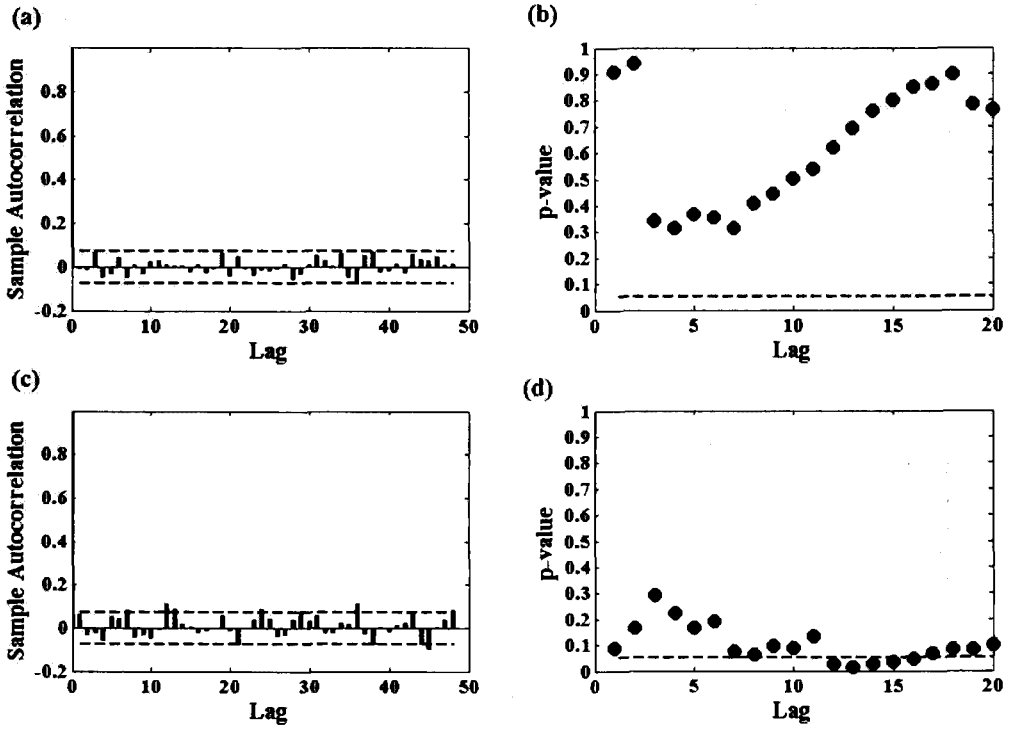


Figure 7. a) Autocorrelation function of the residuals b) p-values of Ljung-Box test for the residuals c) Autocorrelation function of SSRs and d) p-values of Engle's test of SSRs of SARIMA(1,0,1)×(1,0,1) model for Box-Cox transformed rainfall of the Campsie station

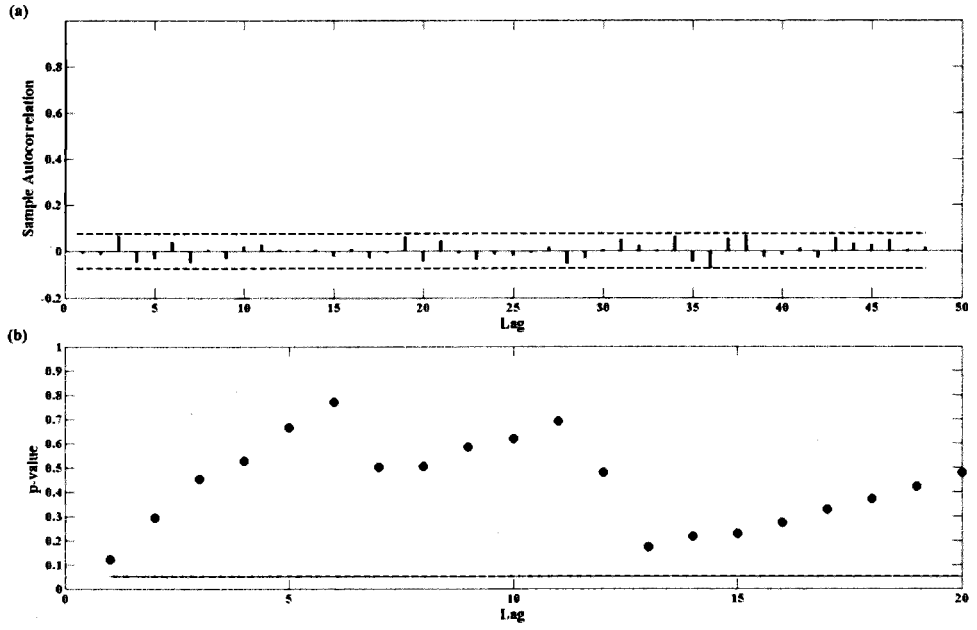


Figure 8. a) Autocorrelation function and b) p-values of Engle's test of SSRs of SARIMA(1,0,1)×(1,0,1)–GARCH(1,1) model for Box-Cox transformed rainfall of Campsie station

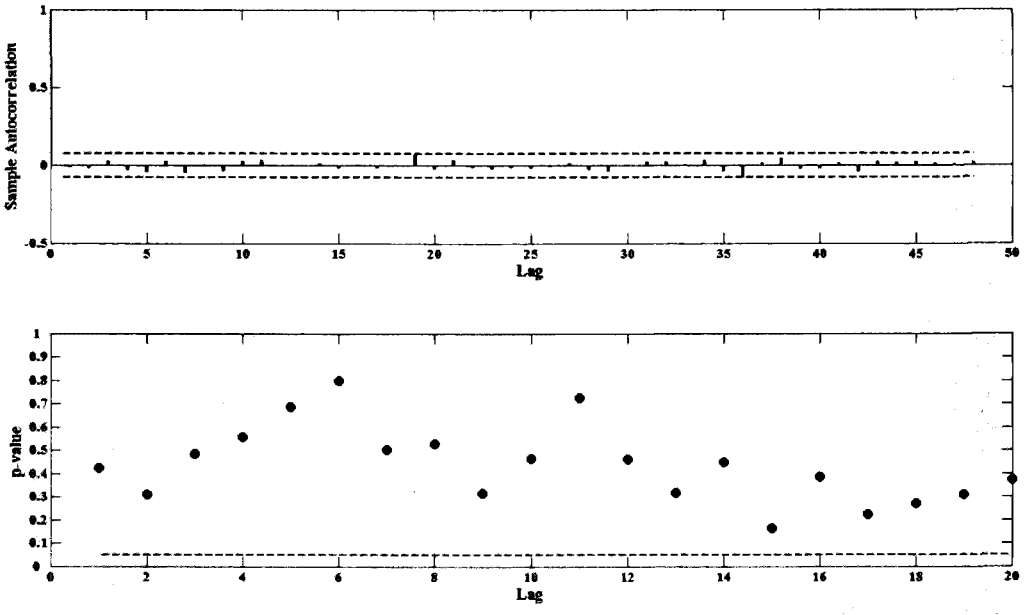


Figure 9. a) Autocorrelation function and b) p-values of Engle's test of SSRs of SARIMA(1,0,1)×(1,1,1) model for Box-Cox transformed rainfall of the Campsie station

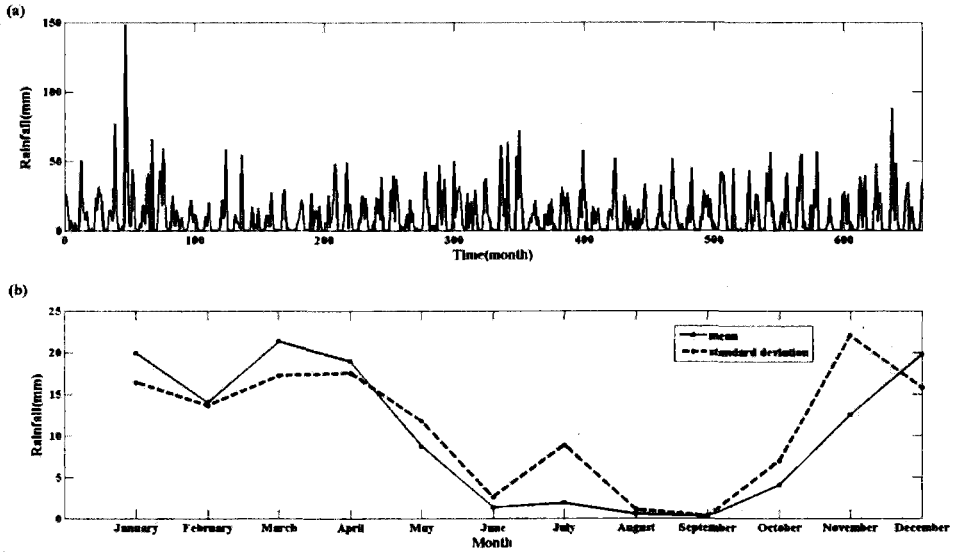


Figure 10. a) monthly rainfall time series and b) monthly distribution of rainfall mean and standard deviation of Isfahan station for the 1951-2005

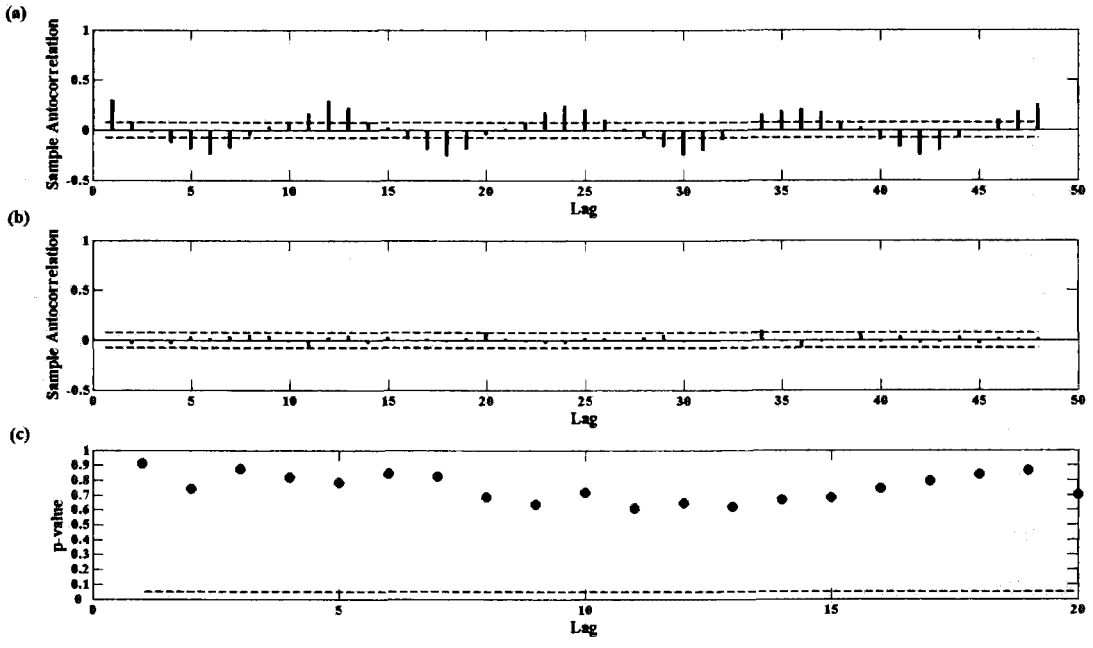


Figure 11. a) ACF of Isfahan rainfall time series, b) ACF of the residuals and c) p-values of the Ljung-Box test of SARIMA(0,0,1)×(1,0,1) model for the Isfahan station

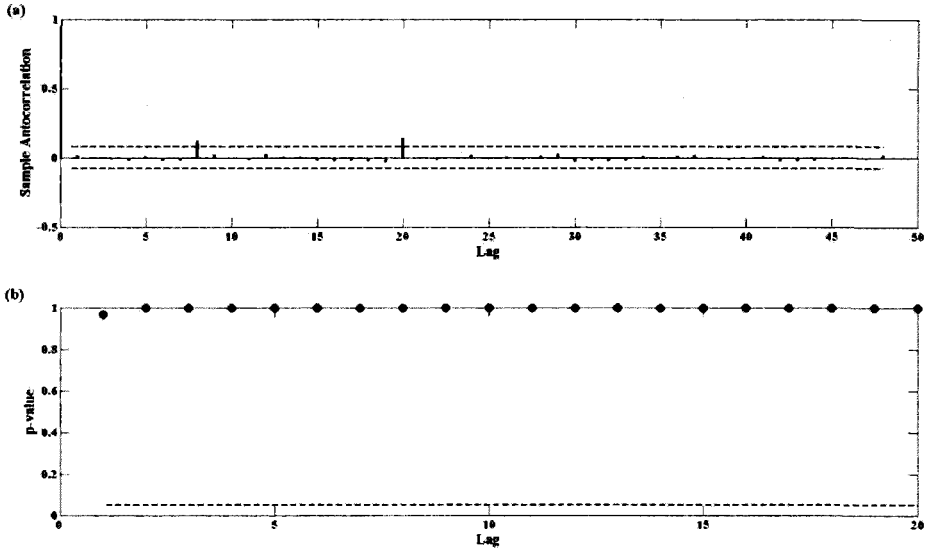


Figure 12. a) Autocorrelation function and b) p-values of Engle's test of SSRs of SARIMA(0,0,1)×(1,0,1) model for the Isfahan station

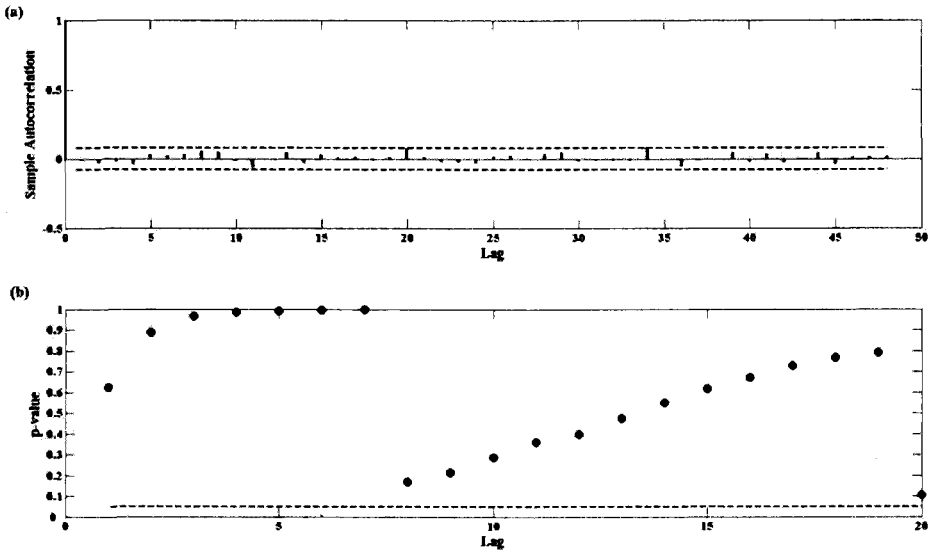


Figure 13. a) Autocorrelation function and b) p-values of Engle's test of SSRs of SARIMA(0,0,1)×(1,1,1)–GARCH(0,2) model for the Isfahan station

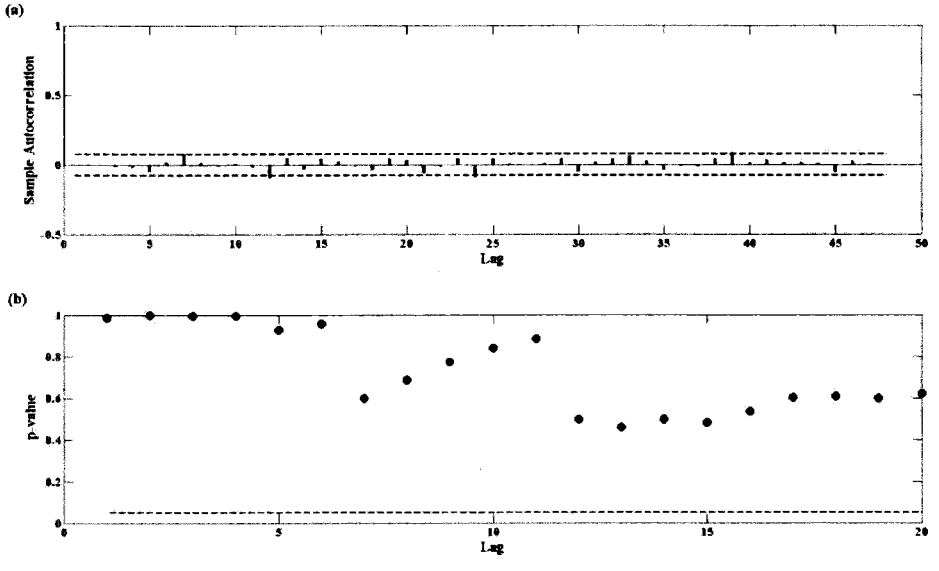


Figure 14. a) ACF of the residuals and b) p-values of the Ljung-Box test of SARIMA $(0,0,2)\times(1,0,2)$ model for Box-Cox transformed rainfall of the Isfahan station

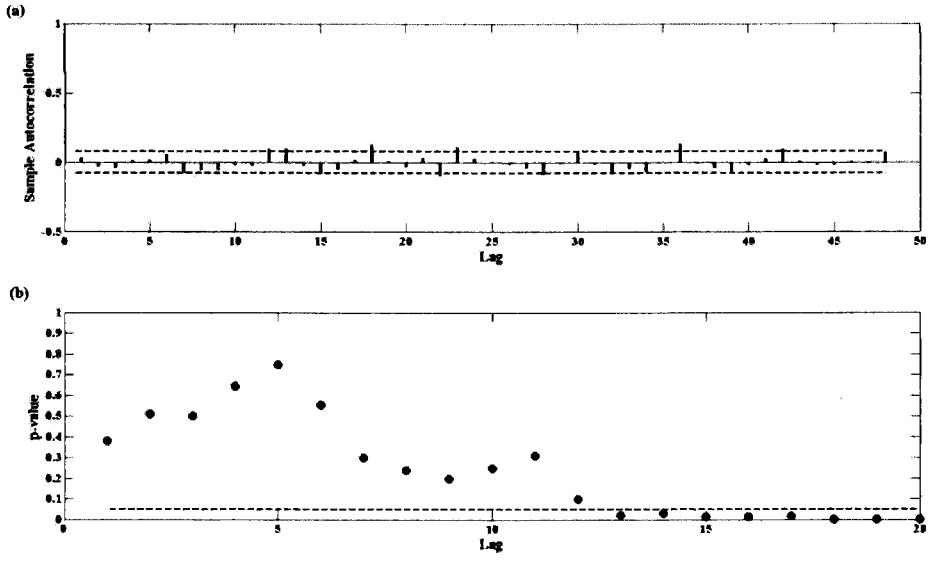


Figure 15. a) Autocorrelation function and b) p-values of Engle's test of SSRs of SARIMA(0,0,2)×(1,0,2) model for Box-Cox transformed rainfall of the Isfahan station

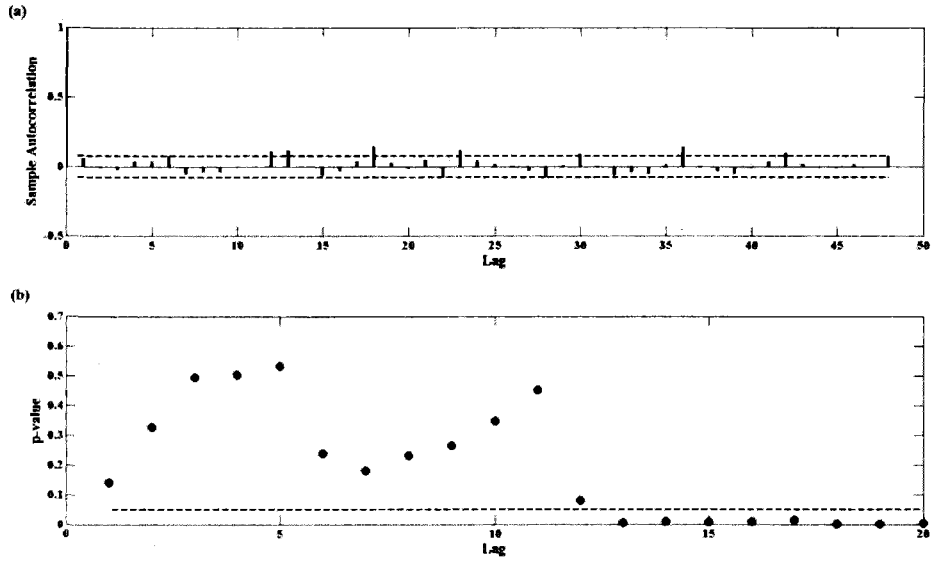


Figure 16. a) Autocorrelation function and b) p-values of Engle's test of SSRs of SARIMA(0,0,2)×(1,1,2) model for Box-Cox transformed rainfall of the Isfahan station

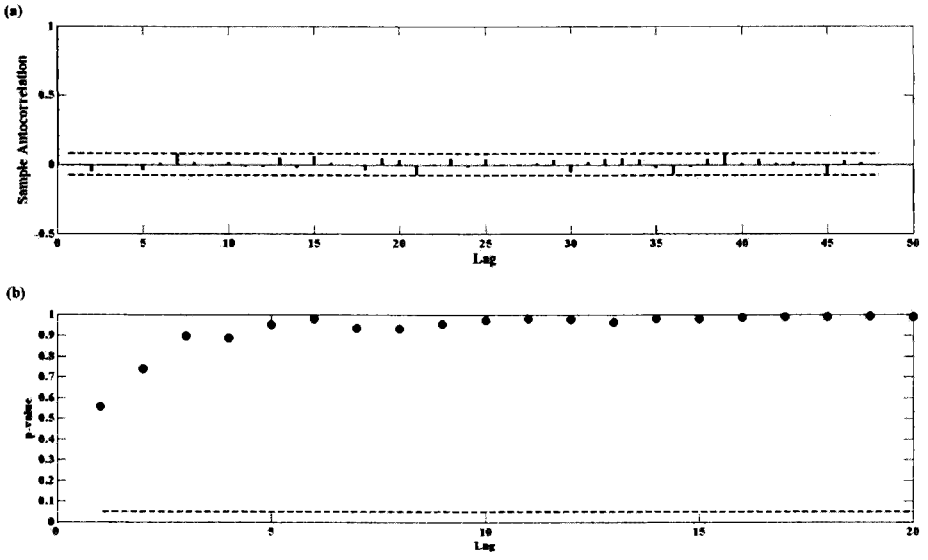


Figure 17 a) Autocorrelation function and b) p-values of Engle's test of SSRs of SARIMA(0,0,2)×(1,0,2)–GARCH(2,3) model for Box-Cox transformed rainfall of the Isfahan station

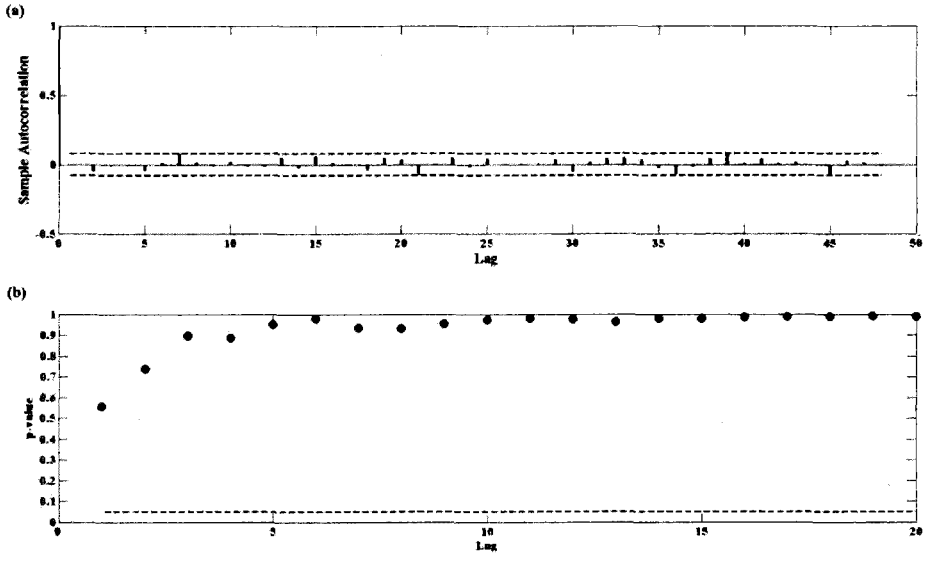


Figure 18. a) Autocorrelation function and b) p-values of Engle's test of SSRs of SARIMA(0,0,2)×(1,1,2)–GARCH(2,2) model for Box-Cox transformed rainfall of the Isfahan station

Table 1. Model criteria for the Campsie rainfall time series

Variable	Model	AME	PDIFF	MAE	RMSE	RAE	MRE	R^2	CE	IoAd	DM
original rainfall time series	$SARIMA(1,0,1)_{12}$	181.18	156.12	13.36	23.35	0.43	-1.76	0.65	0.65	0.88	-14.7*
	$SARIMA(1,0,1)_{12}$-GARCH(1,1)	181.79	157.51	13.23	23.36	0.42	-1.22	0.65	0.65	0.88	
	$SARIMA(1,1,2)_{12}$	174.77	156.01	13.30	23.33	0.42	-1.70	0.66	0.66	0.89	-18.01*
	$SARIMA(1,1,2)_{12}$-GARCH(1,2)	182.48	162.53	13.18	23.36	0.42	-0.71	0.66	0.64	0.87	
Box-Cox transformed rainfall	SARIMA(1,0,1)×(1,0,1)	0.43	0.20	0.09	0.12	0.33	-0.01	0.85	0.85	0.95	371.1
	SARIMA(1,0,1)×(1,0,1)-GARCH(1,1)	0.44	0.19	0.09	0.12	0.32	-0.01	0.85	0.85	0.96	
	SARIMA(1,0,1)×(1,1,1)	0.43	0.19	0.09	0.12	0.32	0.0	0.86	0.86	0.96	-

Notes: * significant at 5% level or better.

Table 2. Model criteria for the Isfahan rainfall time series

Variable	Model	AME	PDIFF	MAE	RMSE	RAE	MRE	R^2	CE	IoAd	DM
Original rainfall	SARIMA(0,0,1)×(1,0,1)	134.13	107.80	7.85	13.03	0.70	-7.01	0.27	0.27	0.64	0.05
	SARIMA(0,0,1)×(1,0,1)-GARCH(0,2)	133.16	109.24	7.98	12.94	0.72	-7.07	0.28	0.28	0.65	
	SARIMA(0,0,1)×(1,1,1)	134.43	109.49	8.01	13.15	0.71	-7.38	0.27	0.27	0.64	0.12
	SARIMA(0,0,1)×(1,1,1)-GARCH(0,2)	134.78	110.52	8.05	13.14	0.72	-7.44	0.27	0.27	0.64	
Box-Cox transformed rainfall	SARIMA(0,0,2)×(1,0,2)	5.49	1.94	1.04	1.40	0.53	0.30	0.55	0.55	0.84	-0.07
	SARIMA(0,0,2)×(1,0,2)-GARCH(2,3)	5.38	1.83	1.04	1.39	0.53	0.31	0.56	0.55	0.85	
	SARIMA(0,0,2)×(1,1,2)	5.25	1.85	1.02	1.39	0.51	0.32	0.57	0.57	0.85	0.15
	SARIMA(0,0,2)×(1,1,2)-GARCH(2,2)	5.19	1.62	1.00	1.36	0.51	0.31	0.58	0.58	0.86	

Article 2. Modeling heteroscedasticty of streamflow times series

Modeling heteroscedasticty of streamflow times series

R. Modarres[a*], T. B. M. J. Ouarda[b,c],

[a] Hydroclimate modeling group, INRS-ETE, 490 De La Couronne, Québec, QC, Canada, G1K 9A9

[b] Masdar Institute of Science and Technology, P.O. Box 54224, Abu Dhabi, UAE

Tel: +1 418 654-3842, Fax: +1 418 654-2600,

E-mail: Reza.Modarres@ete.inrs.ca or touarda@masdar.ac.ae

[*] Corresponding author

Hydrological Sciences Journal

2013, **58**(1): 54-64

Abstract

Time series modeling approaches are useful tools for simulating and forecasting hydrologic variables and their change through time. Although linear time series models are common in hydrology, the nonlinear time series model, the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model, has been rarely used in hydrology and water resources engineering. The GARCH model considers the conditional variance remaining in the residuals of the linear time series models such as an ARMA or an ARIMA model. In the present study, the advantages of a GARCH model against a linear ARIMA model are investigated using three classes of a GARCH approach namely Power GARCH, Threshold GARCH and Exponential GARCH models. A daily streamflow time series of the Matapedia River, Quebec, Canada, is selected for this study. It is shown that the ARIMA (13,1,4) model is adequate for modeling streamflow time series of Matapedia River but the Engle's test shows the existence of heteroscedasticity in the residuals of the ARIMA model. Therefore, an ARIMA (13,1,4)-GARCH (3,1) error model is fitted to the data. The residuals of this model are examined for the existence of heteroscedasticity. The Engle's test indicates that the GARCH model has considerably reduced the heteroscedasticity of the residuals. However, the Exponential GARCH model seems to completely remove the heteroscedasticity from the residuals. The multi-criteria evaluation for model performance also proves that the Exponential GARCH model is the best model among ARIMA and GARCH models. Therefore, the application of a GARCH model is strongly suggested for hydrologic time series modeling as the conditional variance of the residuals of the linear models can be removed and the efficiency of the model will be improved.

Keywords—Heteroscedasticity, GARCH, Engle's test, Exponential GARCH, streamflow time series

1. Introduction

Time series modeling has been an important topic in hydrology, water resources and climate sciences for decades as it is a useful tool in hydrologic analysis, forecasting and simulation. The linear time series models have been widely applied for different hydrologic and climatic variables such as rainfalls (Machiwal and Jha, 2008), streamflows (Ouarda et al., 1997), floods (Toth et al., 1999), droughts (Modarres, 2007) and water quality variables (Kurunc et. al, 2005). Time series models in hydrology are often linear models that usually focus on modeling and predicting the mean behavior, or the first moment of the variable. These models are usually insufficient for capturing the nonlinear properties of the processes governing the temporal variation of hydrologic variables.

Recently, there has been a growing interest in applying nonlinear time series models in hydrology. Investigations on nonlinearity and applications of nonlinear models to hydrologic variables have received considerable attentions in the recent decades. For example, Rao and Yu (1990) investigated the nonlinear characteristics of the annual streamflow and daily rainfall and temperature time series. They reported the existence of nonlinearity in daily meteorological series but nonlinearity was not observed in the annual streamflow time series. Chen and Rao (2003) investigated the nonlinear behavior of monthly hydrologic time series and indicated that all of the stationary segments of standardized monthly temperature and precipitation series are either Gaussian or linear and some of the standardized monthly streamflows are nonlinear. Wang et. al., (2006) indicated that annual streamflow time series are linear while daily streamflow time series are nonlinear. This nonlinearity behavior of streamflows weakens as the time scale of the data increases from a day to a year.

Although the linearity in the mean or the first moment of hydrologic variables can be captured by linear time series models, the variance or the second order moment of hydrologic variables, which may be responsible for nonlinearity is rarely considered in time series modeling. The Generalized AutoRegressive Conditional Heteroscedasticity model (GARCH), originating from econometrics, provides an appropriate framework for focusing on the nonlinearity behavior of the second order moment of hydrologic variables. This type of model was proposed by Engle (1982) for modeling the conditional variance, or time varying variance, of financial time series and developed by Bollerslev (1986). It represents a nonlinear model which applies the past variance in the explanation of the future variance. This type of model has not received

considerable attention so far by the hydrologic community. Few studies such as Wang et. al., (2005), Wang (2006) and Chen et. al., (2008) have applied and discussed the advantage of GARCH-type model in streamflow time series modeling.

The aim of the present paper is to apply the GARCH-type models in streamflow time series modeling in order to illustrate the advantages of the nonlinear models against the linear ARIMA model usually applied in hydrologic time series modeling. We also introduce and apply three sub classes of the GARCH model in the present work.

2. Methods

2.1. ARIMA model

The linear time series model used in this study is the popular Autoregressive Integrated Moving Average model which may be applied for seasonal and nonseasonal time series. The ARIMA model in this study does not include the seasonal parameters and can be written as the follows (Hipel and McLeod, 1996):

$$\phi_p(B)\nabla^d Y_t = \theta_q(B)\varepsilon_t \qquad (1)$$

where Y_t is the observed time series, $\phi_p(B)$ is a polynomial of order p, $\theta_q(B)$ is a polynomial of order q, ∇^d is the nonseasonal differencing operator, B is the backward operator and ε_t is an independent identically distributed (i.i.d) normal error with a zero mean and standard deviation σ. The nonseasonal differencing operator which is selected just large enough to remove all of nonseasonal nonstationarity is given by the following equation

$$\nabla^d Y_t = (1-B)^d Y_t \qquad (2)$$

Linear model building has three steps, namely model identification, model estimation and model diagnostic checking. For the sake of brevity, the reader is referred to Hipel and McLeod (1996) for more details.

2.2. GARCH model

2.2.1. General theory

The nonlinear model used in this study includes the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model and its varieties commonly applied in econometrics. This approach considers the volatility or the conditional variance (σ_t) of a single asset return and tries to model the volatility process of a univariate variable, Y_t. In this case we have

$$Y_t = \sigma_t \varepsilon_t \qquad \varepsilon_t \sim \mathrm{N}(0,1) \tag{3}$$

Where $\sigma_t^2 = \mathrm{E}(\mathrm{Y}_t^2 | F_{t-1})$ denotes the conditional variance that satisfies the equation

$$\sigma_t^{\,2} = \omega + \sum_{i=1}^{V} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{M} \beta_j \sigma_{t-j}^2 \tag{4}$$

The first part of the above model first introduced by Engle (1982) is called an Autoregressive Conditional heteroscedasticity (ARCH) model of order (V):

$$\omega + \sum_{i=1}^{V} \alpha_i Y_{t-i}^2 = \omega + \alpha_1 Y_{t-1}^2 + \alpha_2 Y_{t-2}^2 + \ldots + \alpha_V Y_{t-V}^2 \tag{5}$$

Bollerslev (1986) added a lagged conditional variance (σ_{t-j}^2) to the ARCH model which acts as a smoothing term and this model (Eq. 4) was then called the Generalized ARCH (GARCH) model of order (V,M).

In hydrology, however, the GARCH approach is applied to model conditional variance or the heteroscedasticity remaining in the residuals of a linear model (ε_t). This type of model is usually referred to as an ARIMA-GARCH error model. Therefore, the Generalized Autoregressive Conditional Heteroscedasticity model, GARCH(V,M), for the residuals of an ARIMA model is defined as the following:

$$\sigma_t^2 = \omega + \sum_{i=1}^{V} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{M} \beta_j \sigma_{t-j}^2 \qquad (6)$$

$$\varepsilon_t = \sigma_t e_t \qquad e_t \sim Normal(0,1) \qquad (7)$$

Where $\alpha_1,...,\alpha_V$ and $\beta_1,...,\beta_M$ are the parameters of the GARCH(V,M) process. Therefore, the ARMA-GARCH(V,M) indicates that the conditional mean is described by an ARMA model (or any other linear models such as an ARIMA or SARIMA model) while its conditional variance is described by a GARCH(V,M) model.

2.2.2. Types of GARCH models

The general formulation of a classic GARCH model was presented in the previous section. However, the classical model has an important drawback; the conditional variance only depends on the modulus of the past variables: past positive and negative innovations have the same effect on the current variance. However, it is not always a valid hypothesis in finance. Therefore, in financial time series modeling, a number of GARCH models have been developed for modeling and estimating asymmetric volatility in the financial assets and returns (Francq and Zakoian, 2010). In this study we introduce the following three main versions of the asymmetric GARCH model for the application in the field of hydrology and water resources.

i. Power GARCH (PGARCH) model

The PGARCH model (Ding, et. al., 1993) can be defined as the following:

$$\sigma_t^{\delta} = \omega + \sum_{i=1}^{V} \alpha_i (|\varepsilon_{t-i}|)^{\delta} + \sum_{j=1}^{M} \beta_j \sigma_{t-j}^{\delta} \qquad (8)$$

where $\delta > 0$. The case $\delta = 2$ with ε_t standard normal corresponds to the classical GARCH model. The specification of a varying exponent δ has the merit of following 'long memory' in the shocks of conditional variance and increases the flexibility of GARCH-types models (see Ding et. al., 1993 and Francq and Zakoian (2010) for a broad discussion).

ii. Threshold GARCH (TGARCH) model

In financial time series modeling, it is assumed that positive and negative returns ($\varepsilon_{t-i} > 0$ or good news and $\varepsilon_{t-i} < 0$ or bad news) have different effects on conditional variance. To consider this asymmetric behavior of the returns, the TGARCH model (Zakoïan, 1994) is used which can be specified as the following

$$\sigma_t^2 = \omega + \sum_{i=1}^{V} \alpha_i \varepsilon_{t-j}^2 + \sum_{j=1}^{M} \beta_j \sigma_{t-j}^2 + \sum_{k=1}^{r} \gamma_k \varepsilon_{t-k} I_{t-k}^- \qquad (9)$$

where $I_{t-k}^- = 1$ if $\varepsilon_t < 0$ and 0 otherwise. Here, if $\gamma \neq 0$, the news impact on volatility is asymmetric. For example, if $\gamma > 0$, bad news increases volatility and if $\gamma < 0$ good news increase volatility. In other words, the current volatility, or conditional variance, depends on both modulus and the sign of the past observations.

iii. Exponential GARCH (EGARCH) model

The EGARCH model was proposed first by Nelsen (1991) and soon became popular in financial applications as the conditional variance is an exponential function in this case and ensures having a positive conditional variance:

$$Log(\sigma_t^2) = \omega + \sum_{i=1}^{V}(\alpha_i \left|\frac{\varepsilon_{t-i}}{\sigma_{t-i}}\right|) + \sum_{j=1}^{M}\beta_j \log(\sigma_{t-j}^2) \qquad (10)$$

Nelson (1991) described the constraints of the general form of the GARCH model such as nonnegativity of ω and the persistency of shocks to the conditional variance, which can be solved by using the logarithm of the variance in the EGARCH model. The proofs of the advantage of an EGARCH model over a GARCH model can be found in Nelson (1991) and Francq and Zakoian (2010).

2.2.3. Test for ARCH effect

By using a linear time series models, it is assumed that the residuals of the model are time independent. However, the squared residuals sometimes remain autocorrelated. This feature is called an ARCH effect and, if it exists in the residuals, the variance of the residuals is assumed heteroscedastic.

To test the ARCH effect in the residuals, the Engle's Lagrange Multiplier test for the ARCH effect proposed by Engle (1982) is used. The test statistic is given by NR^2, where R is the sample multiple correlation coefficient computed from the regression of ε_t^2 on a constant and $\varepsilon_{t-1}^2,\ldots,\varepsilon_{t-V}^2$ and N is the sample size. The null hypothesis of no ARCH effect is accepted if the test statistic is asymptotically distributed as a chi-square distribution with v degrees of freedom. The test can also be used to investigate the GARCH effect (Bollerslev, 1986).

3. Model comparison

Time series models used in this study are evaluated trough a multi-purpose multi-criteria comparison by applying parametric criteria and a graphical test which take into account the absolute and relative model errors and the global dimensionless measurements (Modarres, 2009).

3.1. Parametric criteria

The parametric criteria include evaluation metrics that compare the observed and estimated variables. The following metrics are applied in this study (Dawson et al., 2007):

- Absolute Maximum Error (AME)

$$AME = \max\left(\left|Q_i - \hat{Q}_i\right|\right) \tag{11}$$

- Peak Difference (PDIFF)

$$PDIFF = \max(Q_i) - \max(\hat{Q}_i) \tag{12}$$

- Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum\left|Q_i - \hat{Q}_i\right| \tag{13}$$

- Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{n}} \tag{14}$$

- Relative Absolute Error

$$RAE = \frac{\sum_{i=1}^{n}\left|Q_i - \hat{Q}_i\right|}{\sum_{i=1}^{n}\left|Q_i - \overline{Q}_i\right|} \tag{15}$$

- Coefficient of determination (R-squared)

$$R^2 = \left[\frac{\sum_{i=1}^{n}(Q_i - \overline{Q})(\hat{Q}_i - \tilde{Q})}{\sqrt{\sum_{i=1}^{n}(Q_i - \overline{Q})^2 \sum_{i=1}^{n}(\hat{Q}_i - \tilde{Q})^2}}\right]^2 \tag{16}$$

- Index of Agreement (IoAd)

$$IoAd = 1 - \frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{n}\left(\left|\hat{Q}_i - \overline{Q}\right| + \left|Q_i - \overline{Q}\right|\right)^2} \tag{17}$$

In these equations, Q_i is the observed time series, $\hat{Q}_i$ is the predicted time series, $\overline{Q}$ and $\tilde{Q}$ are the mean of the observed and predicted time series, respectively.

3.2. quantile-quantile plot (qq plot)

Drawing a Quantile-Quantile plot (qqplot) is a common graphical way for checking if the distributions of two data sets are different. The qqplot is a plot of the quantiles of the first data set against the quantiles of the second data set. If the two data sets come from a population with the same distribution, the points should fall approximately along a 45-degree line. The greater the departure from this line, the greater the evidence that the two data sets have different distributions. The qqplot is used to check the efficiency of the proposed method to keep the distribution of the estimated streamflows the same as the distribution of the observed streamflow time series.

4. Data set

The data set used in the present study includes a 12-year daily streamflow time series of the Matapedia River, near the Amqui basin (station code: 01BD008, 48° 29' N latitude and 67° 27' W longitude, drainage area: 558 km^2) in the province of Quebec, Canada.

5. Results and discussion

5.1. ARIMA model

To build an ARIMA model for selected data, we first inspect the ACF of the data. The inspection of ACF is done to find an initial model and the initial order of the parameters. The ACF of the

daily time series is given in Figure (1a). As the daily streamflow indicates a high persistence and nonstationarity, we apply a logarithmic transformation and d=1 differencing operator to reduce this high persistence characteristic of the data. The ACF of the logarithmized time series is given in Figure (1b). This figure indicates a significant reduction in nonstationarity and persistence of daily streamflows. Therefore, we try to fit an ARIMA model to the logarithmized time series. Trying different models with different parameter orders, the best model is selected based on the minimum Akaike Information Criterion (AIC) and testing the residuals of the model for adequacy. Therefore, an ARIMA(13,1,4) model is fitted to daily time series of the Matapedia River. The p-values of the Ljung-Box goodness-of-fit test (Figure 1(c)) indicate the adequacy of the above model as all p-values are above the critical level ($\alpha = 0.05$). Then, we apply the Engle's test for testing the ARCH effect in the residuals. As shown in Figure (1(d)) for the first 20 lag times, all p-values of the Engle's test are less than the critical level and therefore the existence of an ARCH effect in the residuals of the ARIMA(13,1,4) model is verified. Hence, a GARCH model should be applied to capture and remove the heteroscedasticity remaining in the residuals of the ARIMA(13,1,4) model.

5.2. GARCH model

It was illustrated in the previous section that the heteroscedasticity of the residuals of the ARIMA model is significant and we need to fit a GARCH model to capture the conditional variance of the residuals. The three types of GARCH models presented in section 2.2 are fitted to the residuals of the ARIMA model.

5.2.1. Power GARCH model

The first model fitted to the residuals of the ARIMA model is the PGARCH (3,1) model. The PGARCH(3,1) model for the residuals of the ARIMA model can be written as the following:

$$\sigma_t^{2.35} = 0.92(|\varepsilon_{t-1}|)^{2.35} - 0.51(|\varepsilon_{t-2}|)^{2.35} + 0.04(|\varepsilon_{t-3}|)^{2.35} + 0.58(|\sigma_{t-1}|)^{2.35} \qquad (18)$$

Therefore, we have an ARIMA(13,1,4)-PGARCH(3,1) error model for streamflow data with $\delta = 2.35$. For the above model, all parameters are significant at 95% level and the model is selected according to the minimum AIC.

In order to check the ARCH effect of the residuals of the above model, the Engle's test is applied. The p-values of the test are given in Figure 2. The figure illustrates the advantage of a PGARCH model against the ARIMA model as most of the p-values are larger than the critical value and the ARCH effect has been significantly reduced. However, some p-values are still below the critical value for some lags which indicates some ARCH effect in the residuals. Therefore, one can conclude that the PGARCH model has not completely removed the heteroscedasticity from the residuals of the ARIMA model.

5.2.2. Threshold GARCH model

The threshold GARCH model can take into account the effect of the negative and positive errors of the ARIMA model on the conditional variance of the residuals. The TGARCH(3,1) model can be fitted to the residuals of the ARIMA model according to the minimum AIC which can be written as follows:

$$\sigma_t^2 = 0.86\varepsilon_{t-1}^2 - 0.58\varepsilon_{t-2}^2 + 0.02\varepsilon_{t-3}^2 + 0.69\sigma_{t-1}^2 + 0.05\varepsilon_{t-1} \qquad (19)$$

In this model, as $\gamma \neq 0$, we can conclude that the negative error (observed-estimated < 0) and the positive error (observed-estimated > 0) of the ARIMA model have an asymmetric impact on the

volatility or the heteroscedasticity of the residuals. On the other hand, it is said in finance that as $\gamma > 0$, bad news, or $\varepsilon < 0$, increases the volatility. In our case as an example in hydrology, we can say that over-estimation of an ARIMA model, or negative errors when 'observed – estimated = residuals < 0', will increase the heteroscedasticity in the residuals. Looking at the residuals of the ARIMA model indicates 61% of the errors are over-estimations (negative errors) and 39% of the errors are under-estimations (positive errors) which may result in the asymmetric characteristic of the residuals for which a TGARCH has been fitted. In other words, the heteroscedasticity of the residuals of the ARIMA model of streamflow time series comes from over-estimation rather than under- estimation.

To check the efficiency of heteroscedasticity modeling, the ARCH effect of the residuals of the ARIMA-TGARCH model is examined by the Engle's test. The p-values of the test are given in Figure 3. Similar to the PGARCH model, this figure also illustrates a significant reduction in the heteroscedasticity of the ARIMA-TGARCH model compared to the ARIMA model. However, some p-values are still below the critical value at some lag times (e.g. K=1, 6, 8, 9, 14, 17 and 20) and it seems that the heteroscedasticity still exists in the residuals and the TGARCH model is not adequate for capturing all heteroscedasticity of the residuals.

5.2.3. Exponential GARCH model

The EGARCH model is a GARCH model for the logarithm of the conditional variance. Fitting an EGARCH(3,1) to the residuals of the ARIMA(13,1,4) model will give the following model

$$Log(\sigma_t^2) = -0.98 + 1.38\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| - 0.45\left|\frac{\varepsilon_{t-2}}{\sigma_{t-2}}\right| - 0.11\left|\frac{\varepsilon_{t-3}}{\sigma_{t-3}}\right| + 0.91(\log(\sigma_{t-j}^2)) \qquad (20)$$

All the parameters of the model are significant at the 95% level. The existence of an ARCH effect in the residuals of the ARIMA(13,1,4)-EGARCH(3,1) model is examined by the Engle's

test. The p-values of the test are given in Figure 4. It is clear that all p-values are above the critical level of rejecting the null hypotheses of the ARCH effect and one can conclude that no ARCH effect is remaining in the residuals of an ARIMA-EGARCH error model. We can also see that the EGARCH model is much better than the PGARCH and TGARCH model in capturing the conditional variance of the residuals of the ARIMA model. This may be due to fact that the EGARCH model considers the logarithm of the variance of the residuals.

5.3. Model evaluation

The performance of time series models presented in the previous sections is evaluated here by applying different criteria which evaluate the errors of the model. Before looking at the criteria, we first look at the scatter plots and time series plots of the observed against estimated streamflow time series in Figures 5 and 6.

The scatter plots of the observed against estimated streamflow time series (Figure 5) indicate that all models are weak in estimating the high flows but their performance is good in estimation low flows. The cloud of observations of high flows indicates that the performance of the ARIMA model and PGRCH and TGARCH models are almost the same. The variation of high flow observations for the EGARCH model, however, reveals the best performance among the models as the distribution of the observations around the best fit line is less than the other models. The time series plots of the observed and estimated streamflows, given in Figure 6, also confirm that the EGARCH model has the best performance among the models. However, the numerical criteria are used and discussed to evaluate the model performance.

5.3.1. Multi-criteria evaluation

The evaluation and comparison of the time series models is carried out through the application of the criteria listed in section 3.1 and are presented in Table 1. It is observed that the performance

of an ARIMA model is better than the PGARCH and TGARCH models according to both the error and dimensionless criteria. The error in peak flow estimation for the ARIMA model is also less than that of the PGARCH and TGARCH models which implies that the ARIMA model is more efficient for estimating the maximum flow than the GARCH models. It is observed that the EGARCH model is the best model among the models fitted to streamflow time series. All criteria, except the PDIFF, demonstrate the better performance of the EGARCH model. The error criteria, MAE, RAE and RMSE, indicate relatively less error for the EGARCH model than other models. According to R^2, the EGARCH model describes 93% of the variation of the observed streamflow. Therefore, the EGARCH model can be considered as a 'very satisfactory' model, based on model efficiency classification of R^2 given by Dawson et. al., (2007). The other models are considered as 'good' models according to their R^2 values.

The PDIFF is the only measurement for which the ARIMA model performs better than the EGARCH model. This measurement indicates the difference between a (single) maximum value in the observed data set and a (single) maximum value in the estimated data set and cannot reflect the overall performance of a model. The scatter plots show also that the general agreement between observed and estimated high flows is better in the EGARCH model than the ARIMA model. Therefore, the EGARCH model is selected as the best model among the time series models applied in this study.

5.3.2. qq-plot

The qq-plots of the time series models are given in Figure 7. These plots show that all models underestimate the high streamflow quantiles, but the low streamflow quantiles are better estimated. It is worth noting that the departure of estimated quantiles from the observed ones is almost the same for the ARIMA, PGARCH and TGARCH models. All models show a difference

between observed and estimated streamflow for values that range between 40 and 60 m^3/s. This difference increases for higher quantiles (streamflow > 60 m^3/s). For the EGARCH model, however, the departure of quantiles from y=x line is smaller than the other models for both medium and high quantiles. This represents one of the advantages of using the appropriate GARCH model: reduction of the variation of the model output and reduction of the uncertainty of model estimation. The small departure of the EGARCH model may also be due to the logarithm transformation of the variance of the residuals of the ARIMA model.

6. Conclusions

The conditional variance or the time-dependent variance of hydrologic variables is usually ignored in hydrologic time series modeling. This study illustrates the advantages of the application of a nonlinear GARH approach for modeling the variance of streamflow time series. The ARMA-GARCH error model presented in this study includes the common ARMA or ARIMA model for modeling the conditional mean of streamflows while the GARCH approach models the conditional variance remaining in the residuals of the ARMA or ARIMA model. The Engle's test confirms that the variance of the residual of the linear model is not homoscedastic and shows time varying features.

Three types of asymmetric GARCH models namely Power GARCH, Threshold GARCH and exponential GARCH models, already applied in financial sciences, are introduced an applied here for modeling the time varying variance of daily streamflows. These three models are based on asymmetric behavior of the conditional variance of a financial variable.

Though the GARCH models show the advantage of reducing the heteroscedasticity of the residuals of the ARIMA model, the performances of the Power GARCH and Threshold GARCH

models are not better than the linear ARIMA model. However, the exponential GARCH model outperforms the ARIMA model and other GARCH models. The best performance of the EGARCH model may be due to the fact that the exponential GARCH model considers the logarithm of the variance of the residuals which may lead to a stabilization of variance and reduces the error of the model. Apart from the effect of logarithm transformation, no specific reason was indicated in financial literature, for example Tavares et. al (2008) and Miron and Tudor (2010), for a better performance of the EGARCH model against PGARCH and TGARCH models fitted to different financial time series, except that the data "speaks" and help us to decide which model is the best model (Curto, 2012, personal discussion). Therefore, we need to apply different hydrologic variables to examine the suitability of different asymmetric GARCH models in hydrology, climatology and water resources.

Another important result is related to the significant parameter of a Threshold GARCH model which verifies the asymmetric effect of negative and positive errors on the remaining heteroscedasticity in the residuals of an ARIMA model. The effect of asymmetry on the heteroscedasticity is a well-known concept in financial time series modeling but it has not received enough attention in hydrology. In this study it was observed that the negative errors of the ARIMA model or the over- estimation will increase the heteroscedasticity in the residuals of a linear model. In other words, over-estimation causes more heteroscedastic residuals than under-estimation. Consequently GARCH models can be considered as a serious alternative in hydrology when over-estimation happens. The asymmetric effect of the residuals of the ARIMA model on the conditional variance of the residuals could also be a promising topic for future studies in hydrology and water resources.

Finally, the results of case study suggest the advantages of (an appropriate) of some GARCH type models for both capturing the heteroscedasticity in the residuals of the linear time series models and improving the efficiency of the models. As a few studies have applied nonlinear GARCH models in hydrology, it is hoped that future studies will focus on the application of these models to other hydrologic and climatic variables such as rainfall and temperature at different time scales using different GARCH models. Although the PGARCH and TGARCH models do not show an advantage over ARIMA model for our streamflow data, they may show their advantages for other hydrologic time series with different conditional variance temporal characteristics.

Acknowledgement

The authors would like to thank the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Canada Research Chair (CRC) Program for financially supporting this work.

References

Bollerslev, T., 1986. Generalized Autoregressive Conditional Heteroscedasticity. Journal of Econometrics 31, 307–327.

Chen, C. H., Liu, C.H. and Su, H.C., 2008. A nonlinear time series analysis using two-stage genetic algorithms for streamflow forecasting. Hydrological Processes, 22, 3697–3711

Chen, H.L., & Rao, A.R., 2003. Linearity analysis on stationary segments of hydrologic time series. Journal of Hydrology, 277, 89–99.

Ding, Z., Grangerand, C.W.J and Engle, R. F., 1993. A Long Memory Property of Stock Market Returns and a New Model, Journal of Empirical Finance, 1, 83–106.

Engle, R. F., 1982. Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation, Econometrica, 50, 987-1008.

Francq, C., Zakoian, J.M., 2010. GARCH models structure, statistical inference and finacial applications. John wiley & Sons Ltd. UK. 489 pp.

Hipel K. W., and McLeod, A.E., 1996. Time series modeling of water resources and environmental systems. Elsevier, Amsterdam, The Netherlands

Tavares, A.B., Curto, J.D., and Tavares, G.N., 2008. Modeling heavy tails and asymmetry using ARCH-Type models with stable Paretian distribution. Nonlinear Dynamics., 51, 231-243

Kurunc A, Yurekli, K., and Cevik, O., 2005. Performance of two stochastic approaches for forecasting water quality and streamflow data from Yesilirmak River, Turkey. Environmental Modeling and software, 20, 1195-1200

Machiwal, D., and Jha, M., 2008. Comparative evaluation of statistical tests for time series analysis: application to hydrological time series. Hydrological Sciences Journal, 53, 353-366

Miron, D., and Tudor C., 2010. Asymmetric conditional volatility models: empirical estimation and comparison of forecasting accuracy. Romanian Journal of Economic Forecasting, 13, 3 74-92

Modarres, R., 2007. Streamflow drought time series forecasting. Stochastic Environmental Research and Risk Assessment, 21(3), 223–233

Modarres, R., 2009. Multi-criteria validation of artificial neural network rainfall-runoff modeling. Hydrology and Earth System Sciences, 13, 411-421

Nelson, D. B., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica, 69, 1519–1554.

Ouarda, T. B. M. J., Labadie, J.W. and Fontane, G., 1997. Indexed sequential hydrologic modeling for hydropower capacity estimation. Journal of the American Water Resource Association. 33, 1-13.

Rao, A.R., and Yu, G. H. 1990. Gaussianity and linearity tests of hydrologic time series. Stochastic Hydrology and Hydrauics, 4, 121–134.

Toth, E., Montanari, A., and Brath, A., 1999. Real-time flood forecasting via combined use of conceptual and stochastic models. Physics and Chemistry of the Earth. (B) 24, 793-798.

Wang, W., 2006. Stochasticity, Nonlinearity and Forecasting of Streamflow Processes, Deft University Press, Amsterdam, 2006, ISBN 1-58603-621-1

Wang, W., Vrijling, J. K., Van Gelder, P. H. A. J. M., and Ma, J., 2005. Testing and modeling autoregressive conditional heteroskedasticity of streamflow processes. Nonlinear Process in Geophysics. 12, 55-66

Wang, W., Vrijling, J. K., Van Gelder, P. H. A. J. M. and Ma, J., 2006. Testing for nonlinearity of streamflow processes at different timescales. Journal of Hydrology, 322, 247–268

Zakoïan, J. M. 1994. Threshold Heteroskedastic Models, Journal of Economic Dynamics Control, 18, 931-944.

TABLE CAPTIONS

Table 1 Evaluation criteria for time series models

FIGURE CAPTIONS

Figure 1. a) ACF of daily streamflows, b) ACF of the lag-1 differenced logarithmic daily streamflow, c) p-values of the Ljung-Box of the residuals and d) p-values of the Engle's test for SSRs of ARIMA (13,1,4) model

Figure 2. p-values of the Engle's test for SSRs of ARIMA-PGARCH error model

Figure 3. p-values of the Engle's test for SSRs of ARIMA-TGARCH error model

Figure 4. p-values of the Engle's test for SSRs of ARIMA-EGARCH error model

Figure 5. Scatter plot of observed against estimated streamflows for different models

Figure 6. Time series plot of observed and estimated streamflows for different models

Figure 7 qq plots of time series models

Table 1 Evaluation criteria for time series models

criteria	Models			
	ARIMA	PGARCH	TGARCH	EGARCH
AME	54.34	58.23	58.33	**42.11**
PDIFF	**11.44**	33.70	32.25	19.33
MAE	2.46	2.93	2.92	**2.03**
RMSE	5.66	7.15	7.10	**4.86**
RAE	0.27	0.32	0.32	**0.22**
R^2 (%)	86	81	81	**93**
IoAd (%)	94	89	90	**96**

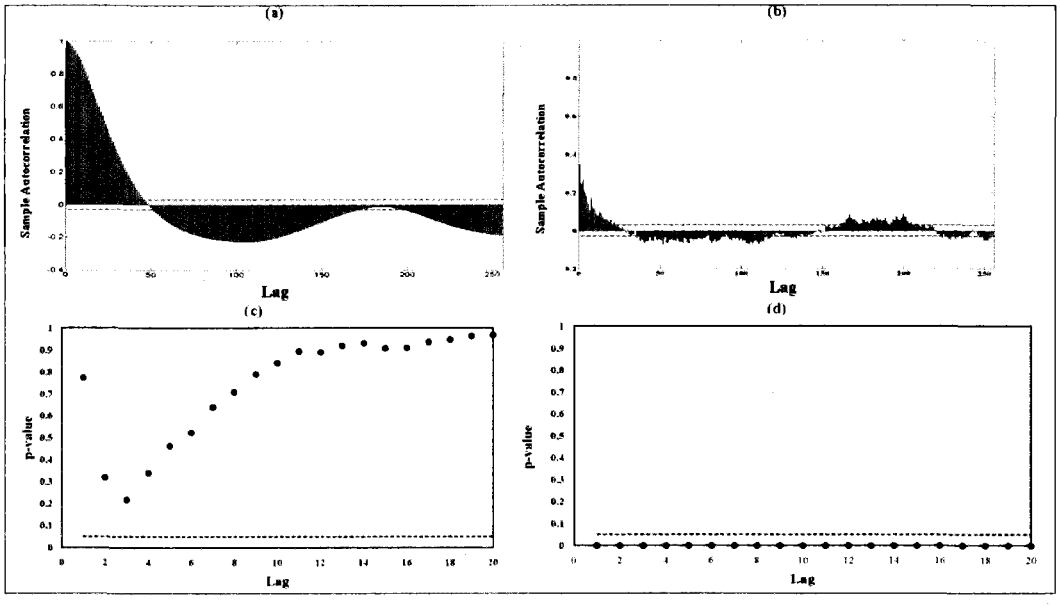


Figure 1

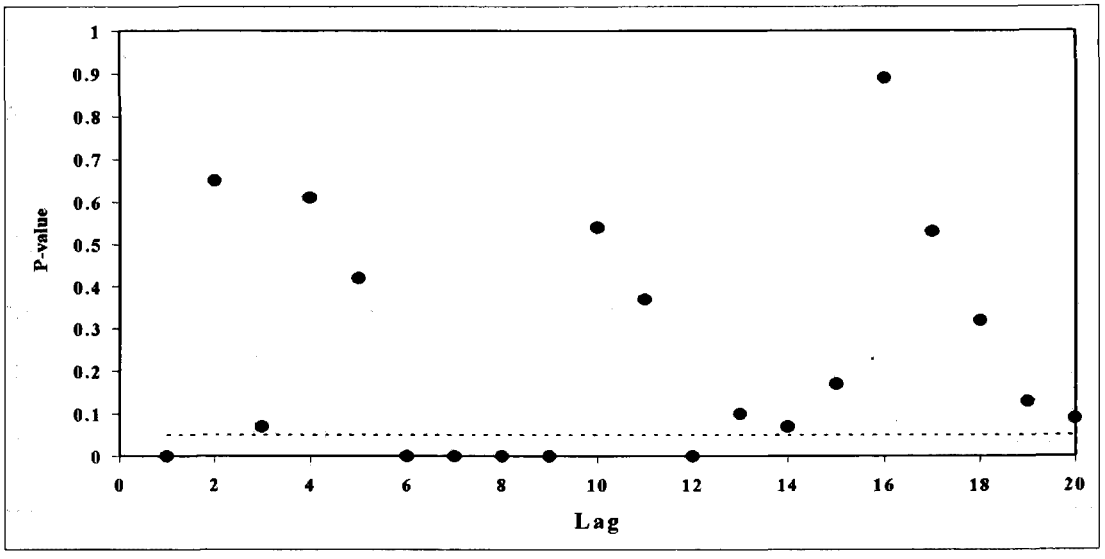


Figure 2

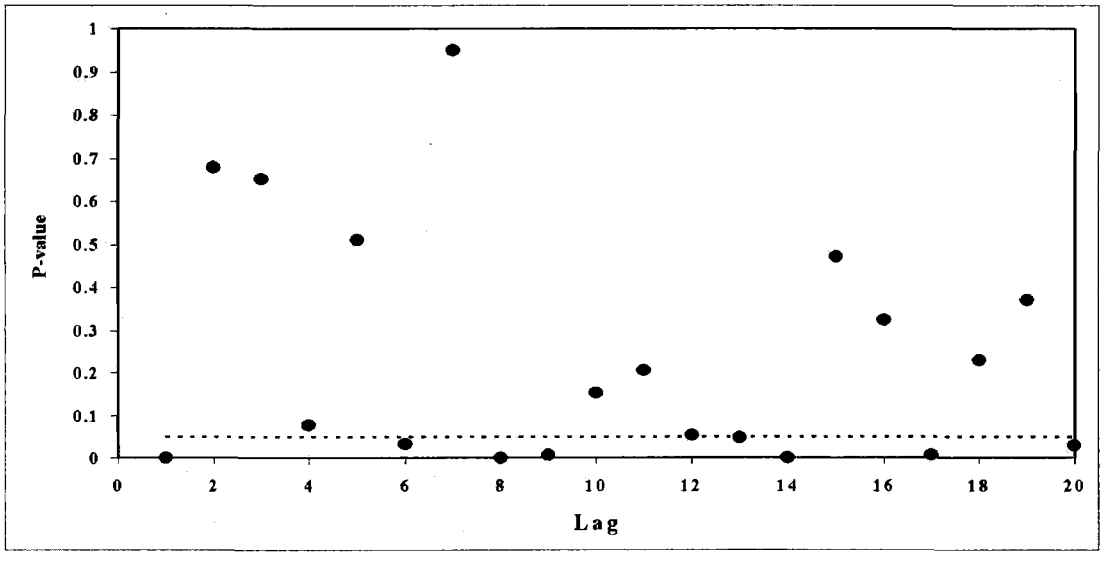


Figure 3

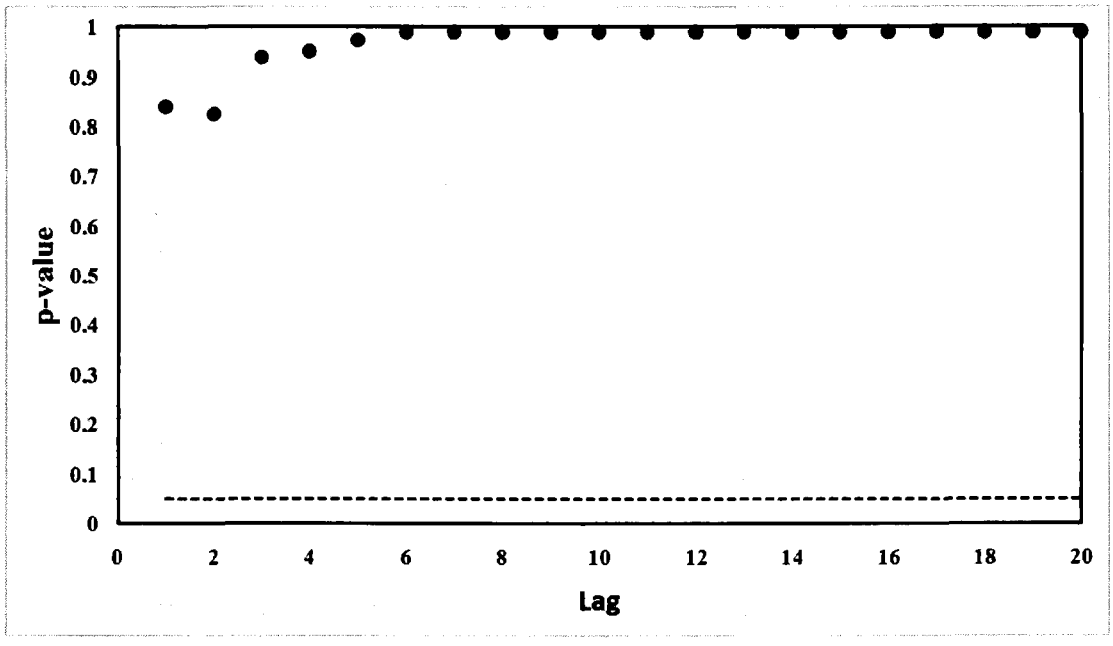


Figure 4

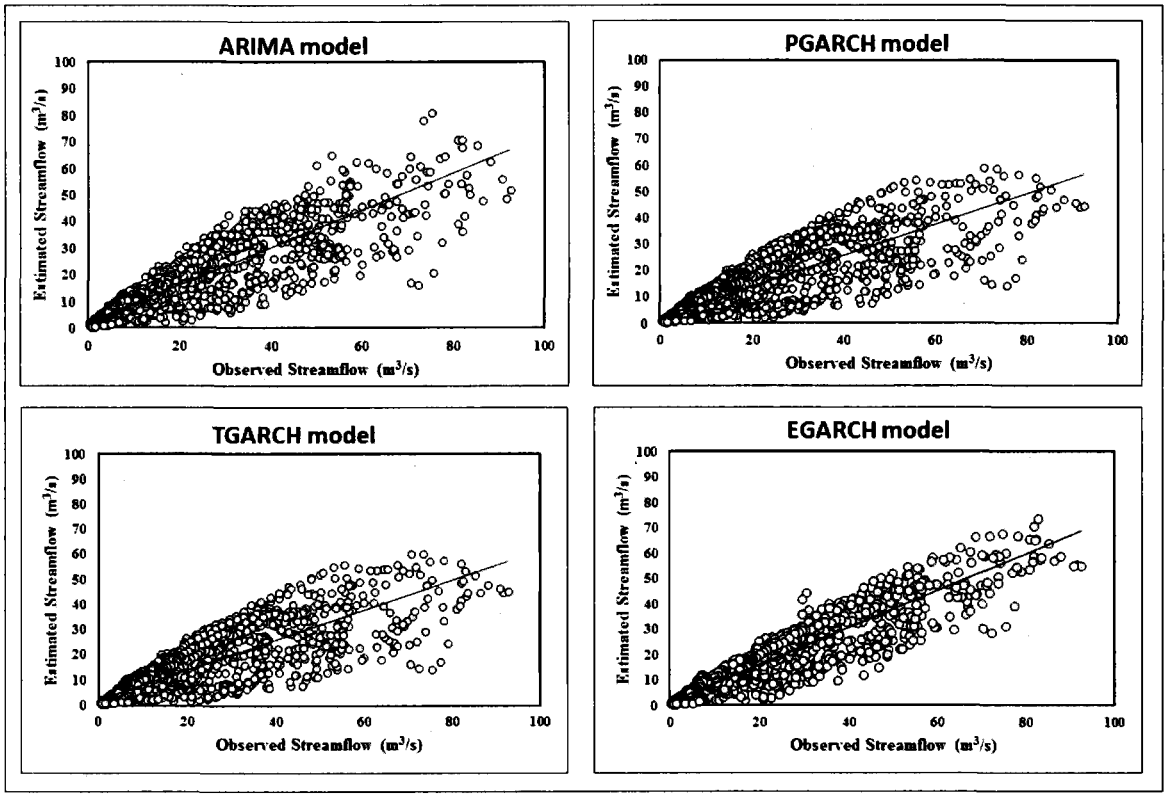

ARIMA model
PGARCH model
TGARCH model
EGARCH model
Estimated Streamflow (m³/s)
Observed Streamflow (m³/s)
0
10
20
30
40
50
60
70
80
90
100

Figure 5

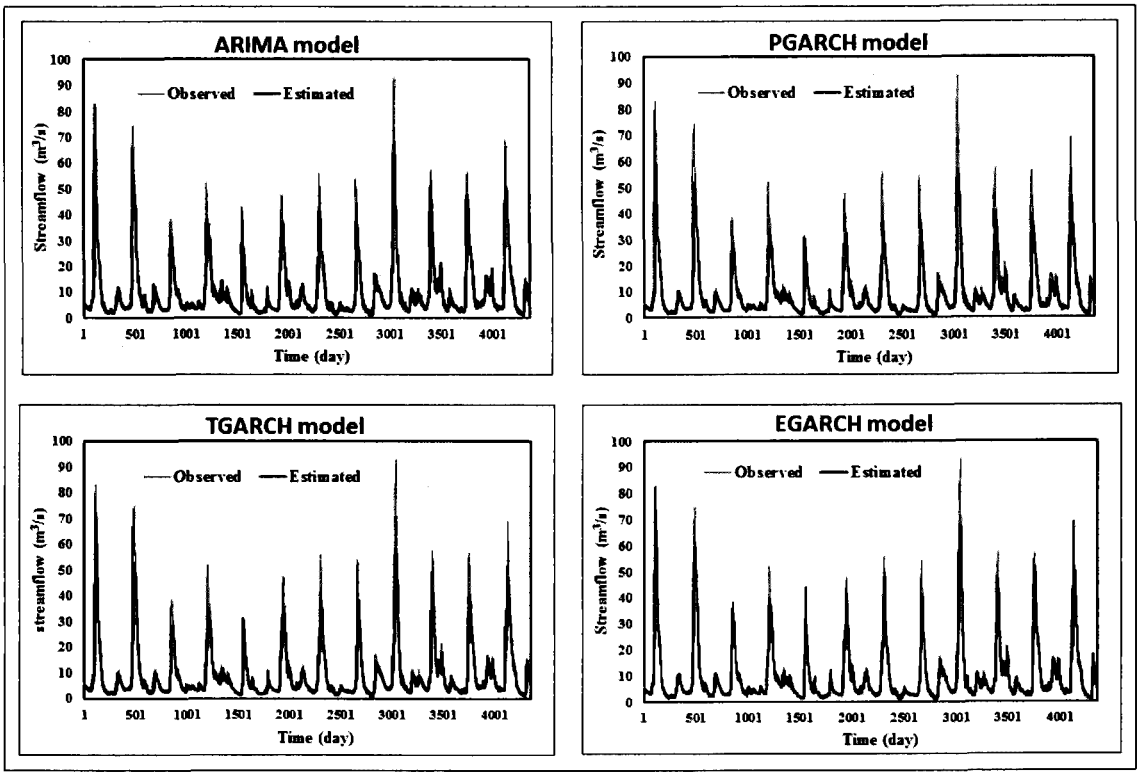

ARIMA model
PGARCH model
TGARCH model
EGARCH model
Observed
Estimated
Streamflow (m³/s)
streamflow (m³/s)
Time (day)
0
10
20
30
40
50
60
70
80
90
100
1
501
1001
1501
2001
2501
3001
3501
4001

Figure 6

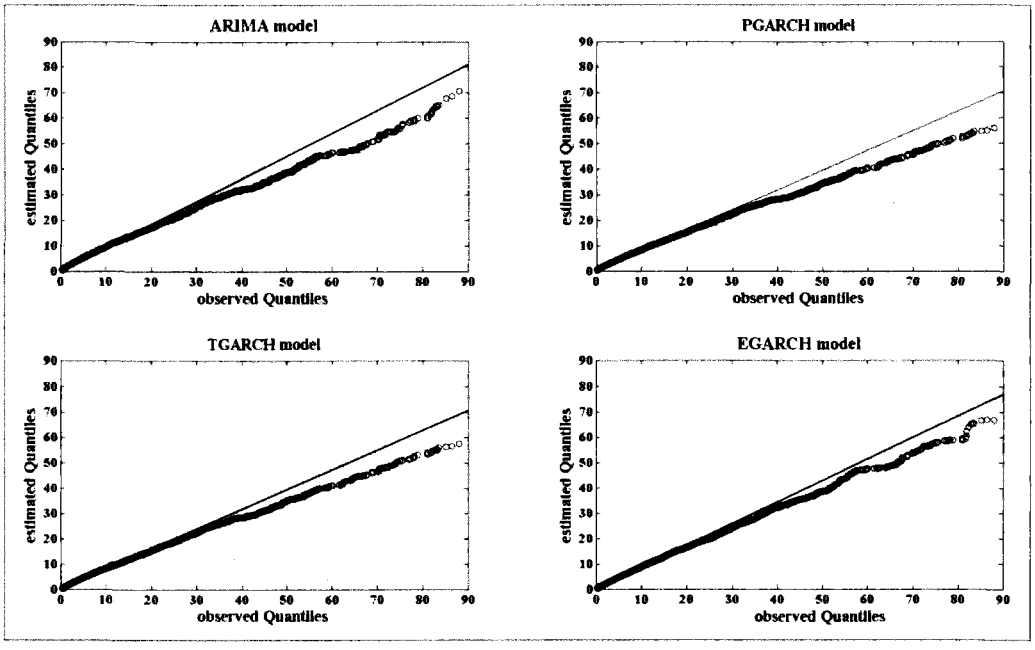

ARIMA model
PGARCH model
TGARCH model
EGARCH model
estimated Quantiles
observed Quantiles
0
10
20
30
40
50
60
70
80
90

Figure 7

Article 3. Testing and modeling the volatility change of SOI

Testing and modeling volatility change of ENSO

R. Modarres[a*], T. B. M. J. Ouarda[a,b],

[a] Hydroclimate modeling group, INRS-ETE, 490 De La Couronne, Québec, QC, Canada

[b] Masdar Institute of Science and Technology, P.O. Box 54224, Abu Dhabi, UAE

Tel: +1 418 654-3842, Fax: +1 418 654-2600,

E-mail: Reza.Modarres@ete.inrs.ca or touarda@masdar.ac.ae

[*] Corresponding author

Submitted for publication to Climatic Change

Abstract

El/Nino-Southern Oscillation (ENSO) is by far the most energetic climate signal. Any change in ENSO characteristics will have serious consequences on the global climate system. This work suggests a different view at the change in ENSO properties in addition to change in its descriptive statistics. The volatility or the conditional variance of ENSO is tested and modeled by two approaches, ARMA-GARCH error and GARCH models, to investigate the change in short-run and long run persistency of the second order moment of ENSO before and after a change point which is detected by the Bayesian change point detection method. Nonparametric tests first revealed a significant change in statistical characteristics such as the mean, the (unconditional) variance and the probability distribution of ENSO after the change point in 1975. The heteroscedasticity in the residuals of the ARMA model fitted to SOI time series is not significant before 1975 while it appears after this change point. The GARCH model indicates an increasing short run persistency after 1975 and decreasing long run persistency. A seasonal shift of extreme heteroscedasticity is observed from summer to winter season. In addition, the nonlinearity and nonstationarity of the SOI volatility have increased in recent decades. This may be due to an increase in frequency and magnitude of extreme volatilities after 1975. The results of this study indicate that ENSO has become more dynamic and uncertain in recent decades. The increase in frequency of extreme events together with extreme conditional variance after 1975 may increase the prediction uncertainty of ENSO driven climate phenomenon.

Keywords: SOI, GARCH, volatility, climate change, conditional variance, Bayesian change point detection

1. Introduction

El Nino/Southern Oscillation (ENSO) phenomenon is a strong climate dominant mode in the Pacific which has major effects on global climate and the ecosystem as well as significant socio-economic consequences around the globe (e.g. Chen et al., 2002) . Therefore, any change in ENSO characteristics and ENSO variation in the past, present and future has been a topic of interest in recent decades. Many studies have reported changes in the ENSO. For example, it is reported that the change in statistical moments of ENSO is an important indicator of human induced climate change (Timmermann, 1999). The period and growth rate of the ENSO, itself, can also be affected by the strength of the ocean-atmospheric feedbacks (Wang, 2007). Over the last century, ENSO characteristics underwent significant changes (e.g. An and Wang 2000). Qian et al., (2011) applied empirical mode decomposition (EMD) and investigated changes in frequency of ENSO and indicated 30% increase in amplitude of ENSO interannual variability around 1937. More recently, Wu et al., (2013) also showed the change of ENSO by EMD method.

One of the important aspects of ENSO characteristics is nonlinearity. The nonlinearity of ENSO is important because, as a global climate driven phenomenon, any change in nonlinearity of ENSO may have big unpredictable and complex influences on the global atmospheric cycle. Different studies have mentioned nonlinearities in ENSO cycle from different views (e.g. Philip and van Oldenborgh, 2009; Boucharel et al., 2011) and have tried to discuss the influencing factors on the nonlinearity of ENSO.

While changes in the properties of ENSO have already been examined, very few studies have paid attention to the second order moment or the variance of ENSO. For example, Cobb et al., (2013) showed that the twentieth-century ENSO variance is significantly higher than average

fossil Coral ENSO variance. Xue et al., (2003) showed the interdecadal change of standard deviation (SD) of SST normals. Stahle et al., 1998, showed the long-term heteroscedasticty (conditional variance) in reconstructed winter SOI.

The present study aims to look at the change of ENSO from a new perspective; the change in the time varying second order moment, the conditional variance or the heteroscedasticity of ENSO. The investigation of this change is carried out based on the Southern Oscillation Index (SOI) time series during 1940-2011 (Figure1), calculated from the monthly or seasonal fluctuations in the air pressure between Tahiti and Darwin, and obtained from the national weather service.

2. Methodology

2.1. Change point detection

Because of the growing evidence of climate change, change point analysis in hydrologic and climatic time series requires revision with new methods. One of these new methods is the Bayesian change point detection (Seidou and Ouarda, 2007). While classical statistics may give the most probable position of the change point, the advantage of Bayesian method is that it provides a full posterior probability distribution of its position. By applying the Bayesian change point detection, the posterior distribution of probability of the number of changes is first estimated. The number of detected change point with the highest probability of occurrence is chosen. Then, the Bayesian inference provides the time position of each selected change point and their respective distribution of probability of occurrence (see more details in Ehsanzadeh et al., 2011). In this study, we apply this method to find a change point in SOI time series in order to compare the statistical and stochastic features of SOI before and after that change point.

2.2. Volatility modeling

The volatility or the heteroscedasticity or the time varying variance of a time series, Y_t, is estimated using a Generalized Autoregressive Conditional Heteroscedasticity (GARCH) approach (Engle, 1982, Bollerslev, 1986). The GARCH(v,m) model can be specified as follows:

$$h_t^2 = \omega + \sum_{i=1}^{v} \alpha_i Y_{t-v}^2 + \sum_{j=1}^{m} \beta_j h_{t-j}^2 \qquad (1)$$

Where ω is a constant, α and β are parameters of the model to be estimated, v and m are the order of the model, h_t^2 is the conditional variance at time t and h_{t-j}^2 is the conditional variance at time step t-j.

In this model, the short-run persistency in conditional variance is defined by the ARCH parameter (α) while the long-run persistency in conditional variance is defined based on the GARCH parameter (β). The high value of $\alpha + \beta$ indicates a high intensity of persistence in the conditional variance of the time series.

In hydrology and climatology, the GARCH approach is usually applied to model and remove the heteroscedasticity in the residuals of the Autoregressive Moving Average (ARMA) model. This type of model is called ARMA-GARCH error model. In other words, in this model, the conditional mean of a variable is modeled by an ARMA model while the conditional variance remaining in the residuals is modeled and removed by a GARCH approach (Modarres and Ouarda, 2012).

Therefore, an ARMA(p,q) model for SOI time series can be written as follows

$$\varphi_p(B) SOI_t = \theta_q(B) \varepsilon_t \qquad (2)$$

Where $\varphi_p(B)$ is the polynomial of order p, $\theta_q(B)$ is a polynomial of order q and B is a backward operation. The heteroscedasticity of in the residuals, ε_t, of the ARMA model or $\sigma^2_{\varepsilon_t}$, can be modeled by a GARCH approach.

It is therefore important, first, to test the existing of the time varying variance, or the ARCH effect, in the residuals of an ARMA model. The main approach to test the ARCH effect is the Engle's Lagrange Multiplier test (Engle, 1982). The test statistic is given by NR^2, where R is the sample multiple correlation coefficient computed from the regression of ε_t^2 on a constant and $\varepsilon^2_{t-1,,,}\varepsilon^2_{t-v}$ and N is the sample size. The null hypothesis of no ARCH effect cannot be rejected if the test statistic is asymptotically distributed as a chi-squared distribution with v degrees of freedom. We consider p=0.05 as the rejection level of the null hypothesis of no ARCH effect in the residuals of the ARMA model.

2.3. Test methods for statistical comparison

In order to compare the change in SOI statistical characteristics before and after change point, three nonparametric tests are used (Conover, 1999); 1) the Wilcoxon test which is a robust test to construct the hypothesis of equality of the mean of two different populations; 2) the Levene's test which hypothesizes the equality of two population variances; and 3) the Kolmogorov-Smirnov (KS) test which hypothesizes the equality of cumulative distribution functions (CDFs) of two populations.

Moreover, to test the change in stationarity and nonlinearity of the SOI and its heteroscedasticity, the Augmented Dickey Fuller (ADF) test (Dickey and Fuller, 1979) for stationarity and the BDS test (Brock et al., 1996) for nonlinearity are applied to compare the SOI and its volatility before and after the change point. For the details of formulations of these tests the reader is referred to Wang (2006).

3-Results and Discussion

3.1. Bayesian change point detection

Using the Bayesian method, one change point in SOI time series can be detected in 1975. Partitioning time series based on the detected change point and fitting a regression line to each segment (Figure 1), SOI is divided into two segments, 1940-1975 and 1976-2011. This change point indicates a shift in the magnitude of SOI time series during the study period. The next step is to compare statistical and stochastic features of SOI before and after 1975.

It should be noted that Latif and Keenlyside (2009) also noticed a slight inter-annual variability during the last century or the so-called regime shift in the mid-1970s in the Tropical Pacific climate system according to SST anomalies along the pacific equator. However, they did not mention the exact change point of SOI. Robbins et al., (2011) showed a possible shift point for the mean (first order moment) of SOI in 1978 by applying SOI time series during 1950 to 1987, which seems to be too short for investigating change point in SOI.

3.2. Statistical comparison of SOI before and after 1975

Considering 1975 as a change point in SOI time series, the change in the descriptive statistics of SOI is first investigated. It is observed in Table 1 that the descriptive statistics such as the (unconditional) mean, standard deviation, skewness and kurtosis show remarkable change after 1975. The average SOI has changed from zero to -0.22 and the (unconditional) variance has increased. Moreover, the distribution of SOI has become more skewed and peaked. These changes in symmetric characteristics of ENSO are reported to be a reflection of nonlinearity of the tropical Pacific system (Hannachi et al., 2003).

Figure 2 shows the explanatory data analysis before and after 1975. The normal histograms (Figure 2a) and quantile-quantile plots (Figure 2b) reveal a change in the lower tail of SOI after 1975 where extreme El Nino events are observed for the second period. The Jarque-Bera test for

normality (results not shown) indicates that the null hypothesis of normality cannot be rejected for the first period while it can be strongly rejected for the second period. This suggests the effect of extreme negative values (El Nino events) on the asymmetry, or the non-normality, of SOI frequency distribution in the second period.

The autocorrelation function (ACF) of SOI (Figure 2c) suggests no difference between two periods while the ACF for the squared or the (unconditional) variance of SOI (Figure 2d) shows a change in ACF in the second period. It is observed that the autocorrelation coefficients at lag times k=1 and k=2 are higher for the second period than those for the first period. This suggests that the time varying variance or the conditional variance is becoming stronger after 1975. Finally, the monthly SOI boxplot (Figure 2e) reveals no significant change in the monthly variation of SOI after 1975, except for January and February. Extreme El Nino events are observed in January and February in the second period while the extreme events in the first period are observed in summer season.

Although the graphical methods and descriptive statistics indicate changes in SOI characteristics in recent years, the nonparametric tests (Wilcoxon, Levene's, and Kolmogorov-Smirnov tests) are also applied to test the validity of the explanatory analysis for the statistical features of SOI time series. The results of nonparametric test are given in Table 2. The Wilcoxon test shows a strong difference (significant at 1% level) in the mean value of SOI before and after 1975. The standard deviation of SOI shows a slight change at 5% significant level according to the Levene's test. Finally, the nonparametric KS test reveals a significant change in the distribution function of SOI after 1975 at 5% significant level.

Finally, the stationarity and nonlinearity of SOI time series before and after change point are evaluated using ADF and BDS tests. The results are presented in Table 3. The ADF statistics

indicate both SOI time series are stationary and we cannot reject the stationarity at 1% significance level. The more negative ADF statistic indicates a stronger rejection of the hypothesis of a unit root. Therefore, one can see that the stationarity of the SOI time series for 1976-2011 is less than that for 1940-1975. In addition, the results of BDS test also reveal the increasing nonlinearity after 1975. For all dimensions (m1-m5), the nonlinearity is significant for both periods but the test statistics are larger for the second period which implies the existence of a higher degree of nonlinearity in the SOI time series after 1975.

3.3. ARMA and ARMA-GARCH error models for SOI

The parameters of the ARMA model for two SOI time series, 1940-1975 and 1976-2011, are given in Table 4. The ARMA(1,1) model is selected based on the minimum Akaike Information Criteria (AIC) in favor of higher order ARMA models.

The p-values of Engle's test results for ARCH effect (Figure 3) indicates that there is no ARCH effect in the residuals of the ARMA(1,1) model for the first SOI time series, 1940-1975, as none of the p-values falls below the critical p=0.05 threshold. On the other hand, it is observed that the first p-value of the test (at lag k=1) for the second SOI time series, 1976-2011, is less than the critical value. This implies that the null hypothesis of no ARCH effect in the residuals of the ARMA(1,1) model for the second SOI time series cannot be rejected. Therefore, it is clear that the time varying variance, or the heteroscedasticity, has appeared in the residuals of the ARMA model after 1975. To remove this heteroscedasticity, a GARCH model is fitted to the residuals of the ARMA model to obtain an ARMA-GARCH error model. The parameters of the ARMA-GARCH error model for second period of SOI are given in Table 4. It can be seen that the ARCH parameter (α) is statistically significant while the GARCH parameter (β) is not statistically significant and can be eliminated from the model. This implies the rising of a

heteroscedasticity in the residuals, white noise or innovations, of the SOI in recent decades. However, the intensity of persistence $(\alpha + \beta)$ of the volatility seems not to be very strong $(\alpha + \beta = 0.25)$.

3.4. GARCH models for SOI

In addition to investigate the ARCH effect in the residuals of an ARMA model, we fit a GARCH model to SOI time series before and after 1975 to evaluate the change in heteroscedastic feature of SOI.

The parameters of the GARCH model for SOI time series (Table 5) are statistically significant for both periods. The GARCH(1,1) model has been selected based on the minimum AIC in favor of higher order models.

The conditional variance of SOI time series are also illustrated in Figure 4 in addition to monthly distribution of conditional variance through a box-plot illustration (Figure 5). The model parameters suggest a significant change in short-run persistence characteristic of SOI after 1975 where the ARCH parameter has increased from 0.29 to 0.43. The degree of long run persistence shows a reduction from 0.43 to 0.32 and the intensity of persistence, $\alpha + \beta$, has increased slightly. It can be seen that the degree of variability (volatility) of SOI has increased while the memory of the second order moment of SOI has decreased due to reduction of β. In other words, the dependence of SOI variance on previous time steps is decreasing and the degree of instability (volatility) of the variance has been increasing during recent decades.

The Conditional variance time series (Figure 4) shows a 2 to 3 times augmentation in the second period, especially in 1982-83 when the unprecedented ENSO event has occurred. Although the monthly average of conditional variance remains around 1 for both periods (Figure 5), the extreme values of conditional variance show a remarkable increase for the second period.

For both periods, the extreme heteroscedasticities (outside $\pm 2.7\sigma$) are observed but the number of extreme hetereoscedasticities has increased on the second period for winter and autumn seasons and decreased for summer. This seasonal shift of conditional variance can also be detected using the QQ plot of heteroscedasticity (Figure 6). If the conditional variances before and after 1975 have the same distribution function, the points should fall approximately along the 45-degree reference line. The greater the departure from the reference line, the greater the evidence that the two conditional variances have different distribution functions.

Except for December, September and November, all other months show a remarkable departure from the reference line. The upper tail of quantile function of heteroscedasticity departs from the 45-degree reference line for most of the month, but the change in lower parts (quantiles) is not significant. It is observed that this departure tends to move above the 45-degree reference line after 1975 for January to April (and also for October) and to move below the 45-degree reference line from May to August. This implies a seasonal shift of the extreme volatility of SOI from summer to winter season after 1975.

Finally, in Table 6, the stationarity and nonlinearity of conditional variance of SOI are presented. From the ADF test result, the stationarity of the conditional variance is clear while the level of stationarity for the second period is lower than that for the first period. Moreover, the conditional variance for both periods demonstrates nonlinearity while the degree of nonlinearity is higher for the second period. These results indicate a notable change in the second order moment of SOI toward more instability, variability and nonlinearity in recent decades.

4. Summary and Conclusions

In this paper, the conditional variance or volatility of ENSO is tested and modeled and its changes in recent decades are explored for the first time. Using a Bayesian change point method,

a significant change point was detected in 1975. Comparing SOI time series before and after 1975 shows a remarkable change in descriptive statistics, stationarity and nonlinearity properties of SOI time series after the change point in 1975. It was shown that the mean and (unconditional) variance of SOI has increased and the distribution of SOI has become non-normal and skewed with higher peakedness after 1975. It was also observed that the SOI has become less stationary and more nonlinear after 1975. The major reason for becoming more asymmetric seems to be due to more (negative) extreme events after 1975.

The heteroscedasticity was not significant in the residuals of the ARMA model for the first period (1940-1975) while it has become significant in the residuals of the model after the change point. It is also observed that the volatility of SOI has increased both in magnitude and frequency of extreme volatilities. The GARCH parameters indicated an increasing in the volatility and decreasing in the memory of the variance of SOI. This augmentation in the volatility and second order moment may lead to an increasing frequency of extreme events such as extreme precipitation and temperature (Higgins et al., 2001; Arblaster and Alexander, 2012) or extreme drought conditions (Rajagopalan et al., 2000, Balling and Goodrich, 2007) in relation with extreme ENSO phases. These changes may also lead to an increase in ENSO driven climate uncertainty and variability and devastate the potential climate predictability in both regional and global context (Tang and Deng, 2010).

5. Perspectives for future work

For future works, it would be very interesting to investigate the change in ENSO volatility for longer historical SOI dataset and also upon the change in El Nino and La Nina phases separately using asymmetric GARCH models. It is also recommended to apply and test volatility of other elements of the tropical ocean-atmosphere system in future studies so that the source of volatility

changes and feedbacks are revealed. It will be interesting to investigate the ability of the ensemble global climate models for simulation and projections of the future changes in ENSO volatility. The seasonal shift of the extreme ENSO condition and the shift in the seasonal volatility are also interesting topics for future studies.

Acknowledgement
The authors thank Natural Sciences and Engineering Research Council (NSERC) of Canada and the Canada Research Chair (CRC) Program for financial supports.

References

Arblaster JM, Alexander LV (2012) The impact of the El Niño-Southern Oscillation on maximum temperature extremes, Geophys Res Lett 39, L20702, doi:10.1029/2012GL053409,

An SI, wang B (2000) Interdecadal change of the structure of the ENSO mode and its impact on the ENSO frequency. J Clim 13: 2044-2055

Baling RC, Goodrich GB (2007) Analysis of drought determinants for the Colorado River basin. Clim. cHange. 82: 179-194

Bollerslev T (1986) Generalized Autoregressive Conditional Heteroscedasticity. J Econometrics, 31: 307–327.

Boucharel J, Dewitte B, du Penhoat Y, Garel B, Yeh SW, Kug JS (2011) ENSO nonlinearity in a warming climate. Clim Dyn 37: 2045-2065. Doi: 10/1007/s00382-011-1119-9

Brock WA, Dechert WD, Scheinkman JA, LeBaron B (1996) A test for independence based on the correlation dimension. Econ Rev. 15: 197–235.

Chen CC, Mccarl B, Hill H (2002) Agricultural value of Enso information under alternative phase definition. Clim Change. 54: 305-325

Cobb KM, Westphal N, Sayani HR, Watson JT, Di Lorenzo E, Cheng H, Edwards RL, Charles CD (2013) Highly variable El Nino-Southern oscillation throughout the Holocene. Sceince 339: 67-70

Conover WJ (1999) Practical Nonparametric Statistics, 3rd, Wiley, NY, USA, 592 pp.

Dickey DA, Fuller WA (1979). Distribution of the estimators for autoregressive time series with a unit root. J Am Stat Assoc 74: 423–431.

Ehsansazeh E, Ouarda TBMJ, Saley HM (2011) A simultaneous analysis of gradual and abrupt changes in Canadian low streamflows. Hydrol Process 24: 727-739

Engle RF (1982) Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation. Econometrica 50: 987-1008

Hannachi A, Stephenson DB, Sperber KR (2003) Probability-based methods for quantifying nonlinearity in the ENSO. Clim Dyn doi: 10.1007/s00382-002-0263-7

Higgins RW, Kousky VE, Xie P (2011). Extreme Precipitation Events in the South-Central United States during May and June 2010: Historical Perspective, Role of ENSO, and Trends. J Hydromet 12: 1056-1070

Latif M, Keenlyside NS (2009). El Nino/Southern Oscillation response to global warming. PNAS 106: 20578-20583, doi: 10.1073/pnas.0710860105

Modarres R, Ouarda TBMJ (2012) Generalized autoregressive conditional heteroscedasticity modeling of hydrologic time series. (In press) Hydrol Process DOI: 10.1002/hyp.9452.

Philip S, van Oldengorgh GJ (2009) Significant Atmospheric Nonlinearities in the ENSO Cycle. J Clim 22: 4014-4028 DOI: 10.1175/2009JCLI2716.1

Qian C, Wu Z, Congbin F, Wamg D (2011) On Changing El Nino: A View from Time-Varying Annual Cycle, Interannual Variability, and Mean State. J Clim 24: 6486-6500.

Rajagopalan B, Cook E, Lall U, Ray BK (2000) Spatiotemporal variability of ENSO and SST teleconnections to summer drought over the United States during the twentieth century. J Clim 13: 4244-4255

Robbins M, Gallagher C, Lund R, Aue A (2011) Mean shift testing in correlated data. J Time Series Anal 32: 498–511

Seidou O, Ouarda TBMJ (2007). Recursion-based multiple change point detection in multiple linear regression and application to river streamflows. Water Res Res 43: W07404, doi:10.1029/2006WR005021

Stahle DW, D Arrigo RD, Krusic PJ, cleaveland MK, Cook ER, Allan RJ, Cole JE, Dunbar RB, Therrekk MD, gay DA, Moore MD, Stokes MA, burns BT, Villanueva-Diaz J, Thompson LG (1998) Experimental Dendroclimatic Reconstruction of the Southern Oscillation. Bulletin of the American Meteorological Society 79 (10): 2137-2152.

Tang Y, Deng Z (2010) Low-dimensional nonlinearity of ENSO and its impact on nonlinearity. Physica *D* 239: 258-268

Timmermann A (1999) Detecting the nonstationary response of ENSO to greenhouse warming. J Atmos Sci 56: 2313-2325

Xue Y, Smith TM, Reynolds RW (2003) Interdecadal changes of 30-yr SST normals during 1871–2000. J Clim 16: 1601–1612.

Wang F (2007). Investigating ENSO sensevity to mean climate in an intermediate model using a novel statistical technique. Geophys Res Lett 34, doi: 10.1029/2007GL029348

Wang W (2006) Stochasticity, Nonlinearity and Forecasting of Streamflow Processes, Deft University Press, Amsterdam, 2006, ISBN 1-58603-621-1

Wu Z, Huang NE, Long SR, Peng C.-K (2007). On the trend, detrending, and variability of nonlinear and nonstationary time series. Proc. Natl. Acad. Sci. USA, 104: 14 889–14 894.

Table 1 descriptive statistics of SOI time series

SOI Series	Mean	(Unconditional) variance	Skewness	Kurtosis	maximum	minimum
1940-2011	-0.11	1.18	-0.19	3.54	2.9	-4.6
1940-1975	-0.0014	1.02	0.078	-0.137	2.9	-2.7
1976-2011	-0.22	1.34	-0.31	0.756	2.9	-4.6

Table 2 nonparametric test results

SOI Series	Wilcoxon	Levene	Kolmogorov-Smirnov
statistic	-2.62	4.03	1.48
p-value	0.009	0.04	0.024

Table 3 Stationarity and non-linearity test results for SOI

Test	ADF		BDS							
			1940-1975				1976-2011			
	1940-1975	1976-2011	m2	m3	m4	m5	m2	m3	m4	m5
statistic	-6.21	-5.61	0.05	0.08	0.10	0.11	0.06	0.11	0.13	0.14
p-value	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

Table 4 ARMA-GARCH models for SOI time series (*p*-values of *t*-statistics for each parameter are given in parenthesis)

SOI Series	φ	θ	α	β
1940-1975	0.89 (0.001)	-0.50 (0.001)	-	-
1976-2011	0.87 (0.001)	-0.39 (0.001)	0.25 (0.001)	0.03 (0.82)

Table 5 GARCH models for SOI time series (p-values of t-statistics for each parameter are given in parenthesis)

SOI Series	ω	α	β	$\alpha + \beta$
1940-1975	0.28 (0.007)	0.29 (0.007)	0.43 (0.008)	0.72
1976-2011	0.33(0.001)	0.43 (0.001)	0.32 (0.008)	0.75

Table 6 Stationarity and nonlinearity test results for conditional variance

Test	ADF		BDS							
			1940-1975				1976-2011			
	1940-1975	1976-2011	m2	m3	m4	m5	m2	m3	m4	m5
statistic	-8.73	-7.77	0.08	0.13	0.16	0.17	0.10	0.16	0.20	0.22
p-value	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

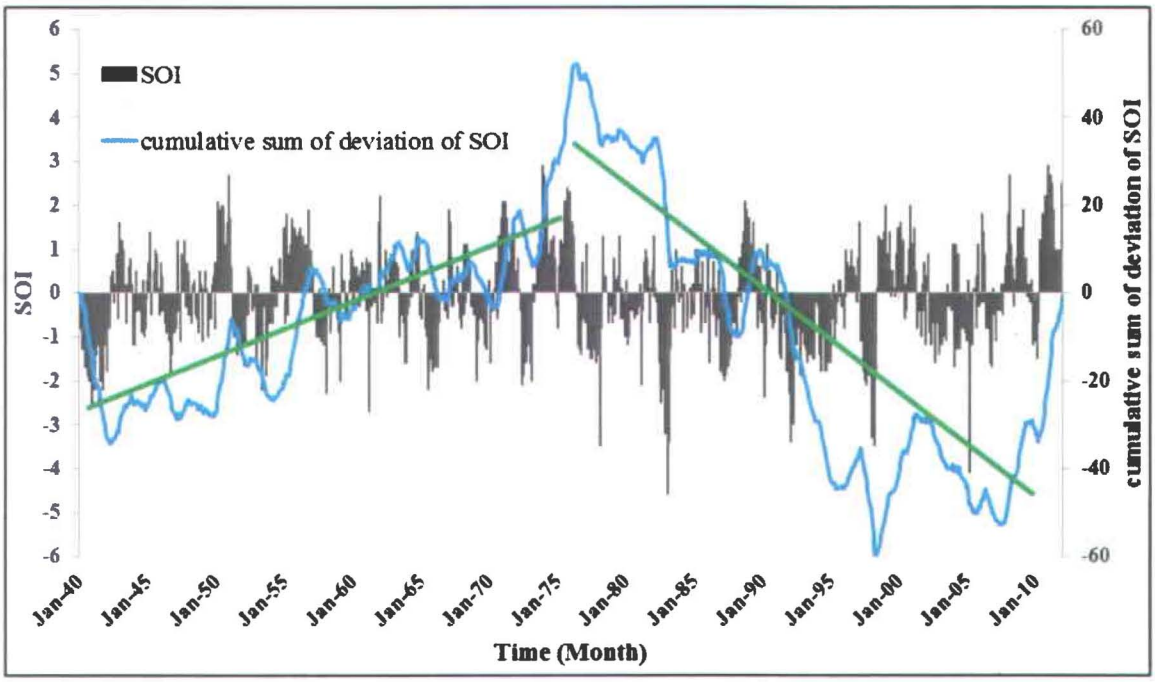


Figure 1. Illustration of Bayesian change point detection analysis for SOI time series (1940-2011). Green lines show regression lines for different segments discriminated by the detected change point

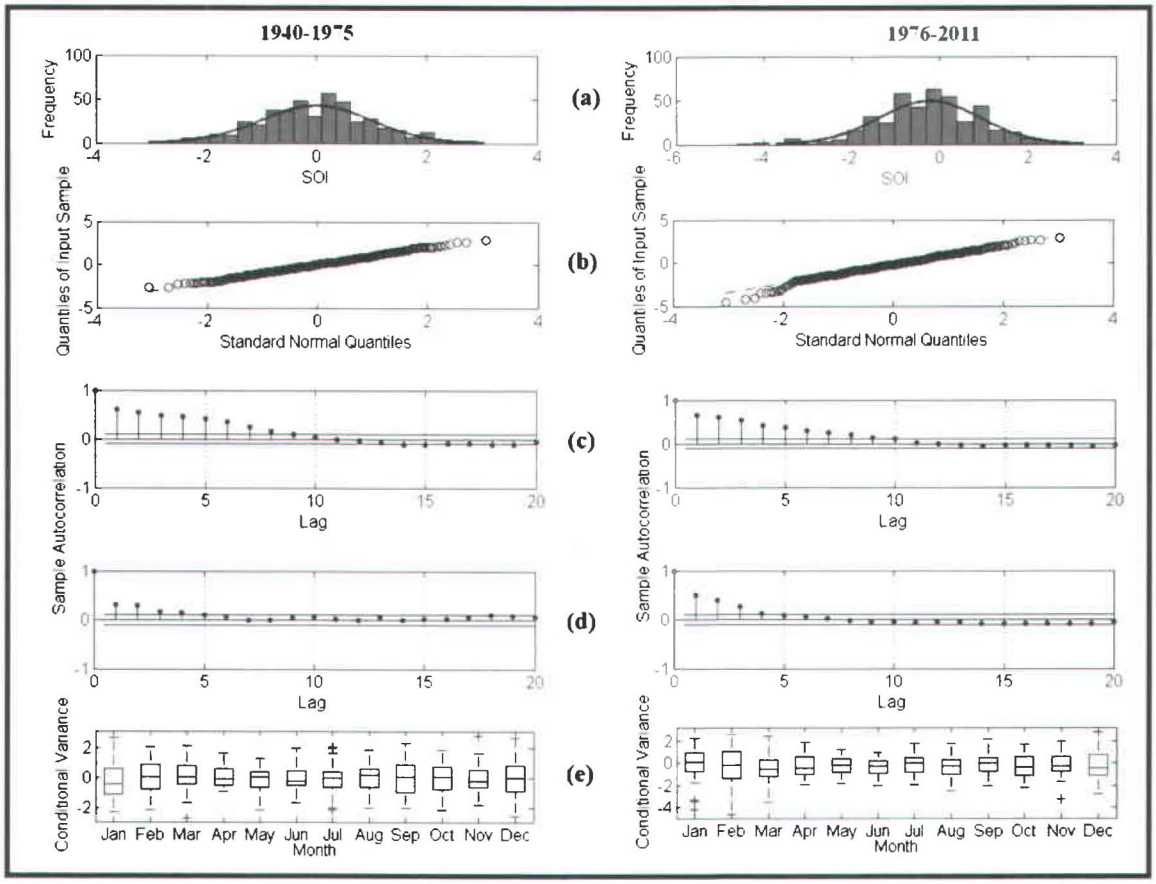


Figure 2. Exploratory data analysis plot for SOI time series, Normal histogram (a), quantile quantile plot (b), ACF for SOI (c), ACF for squared SOI (d) and monthly boxplot (e)

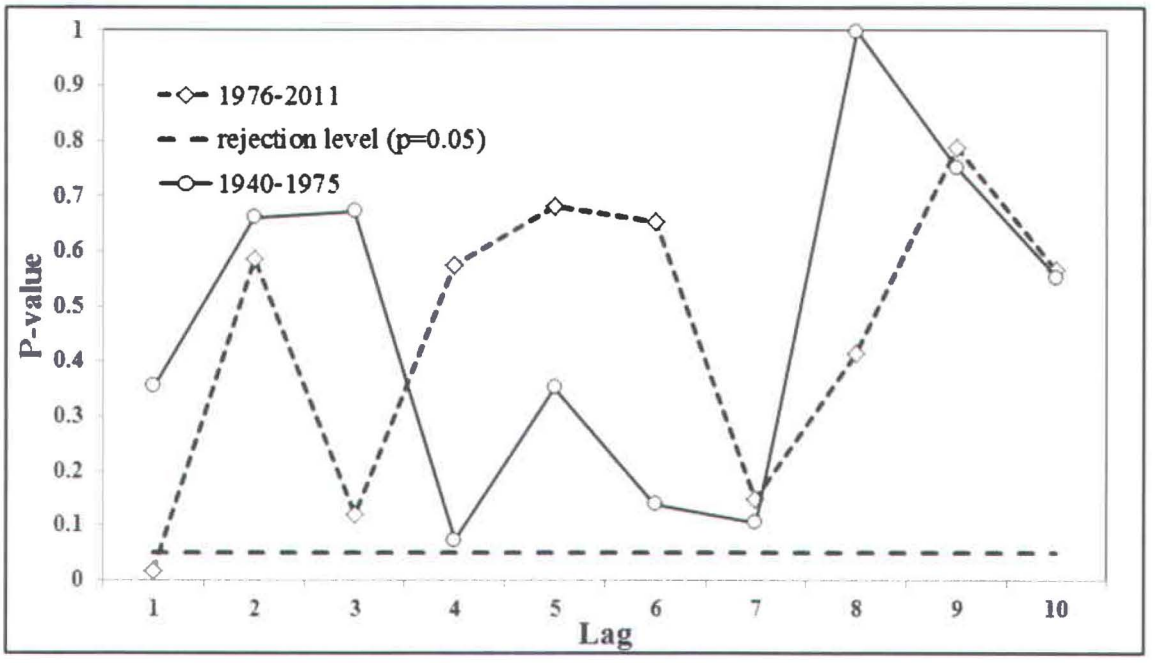


Figure 3. *p*-values of the Engle's test for ARMA models

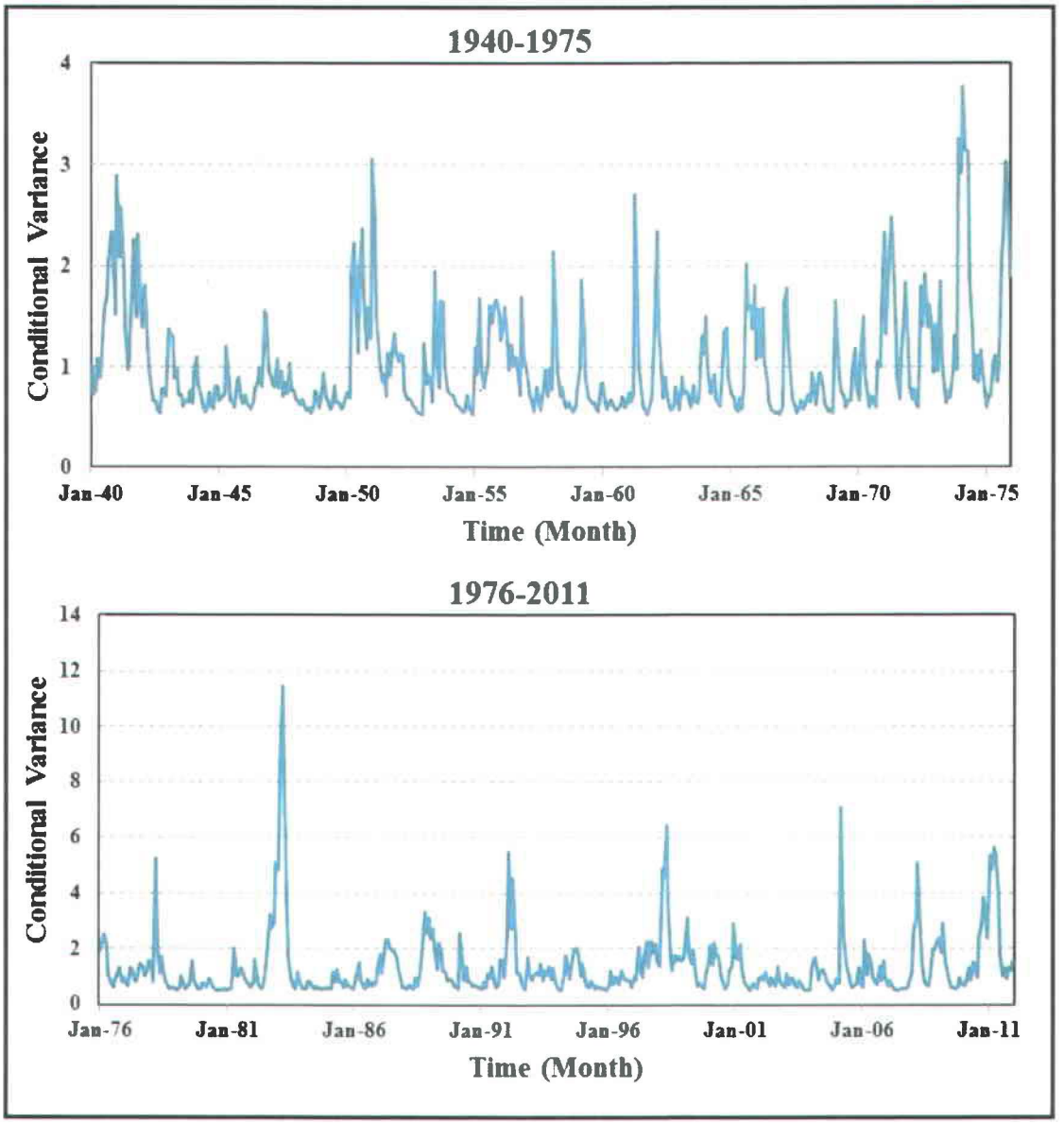


Figure 4. Conditional variance Time series

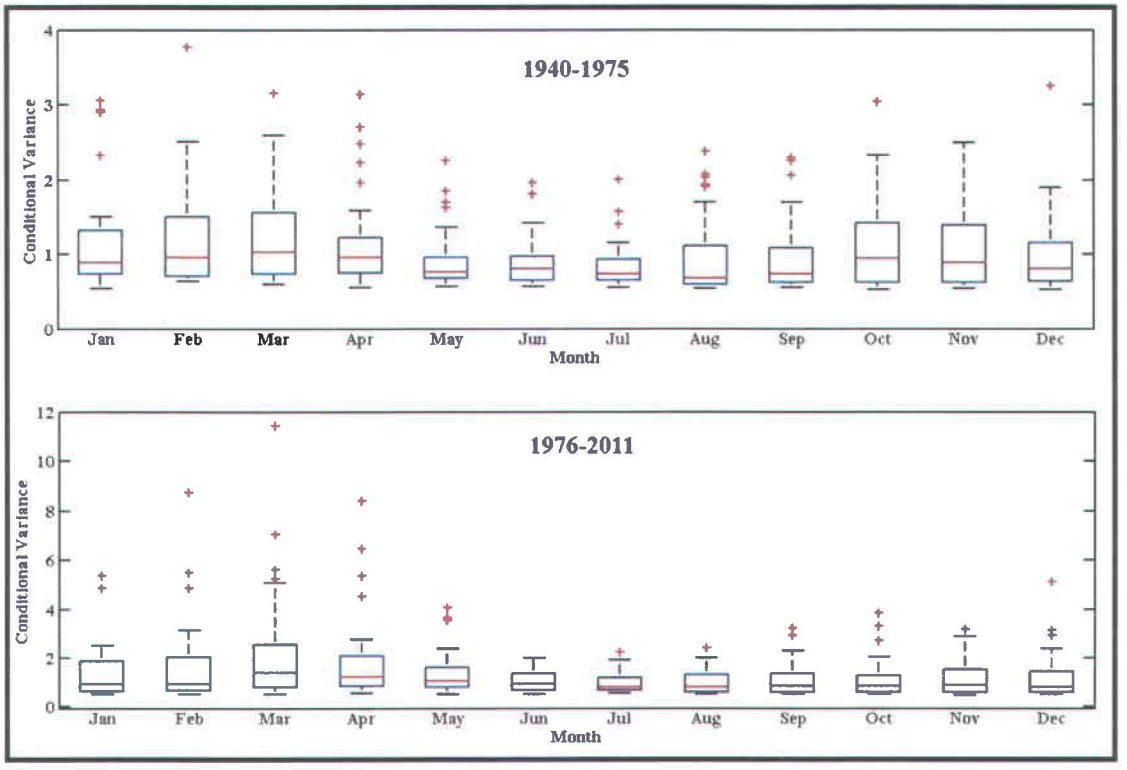


Figure 5. Box plot for monthly conditional variance of SOI

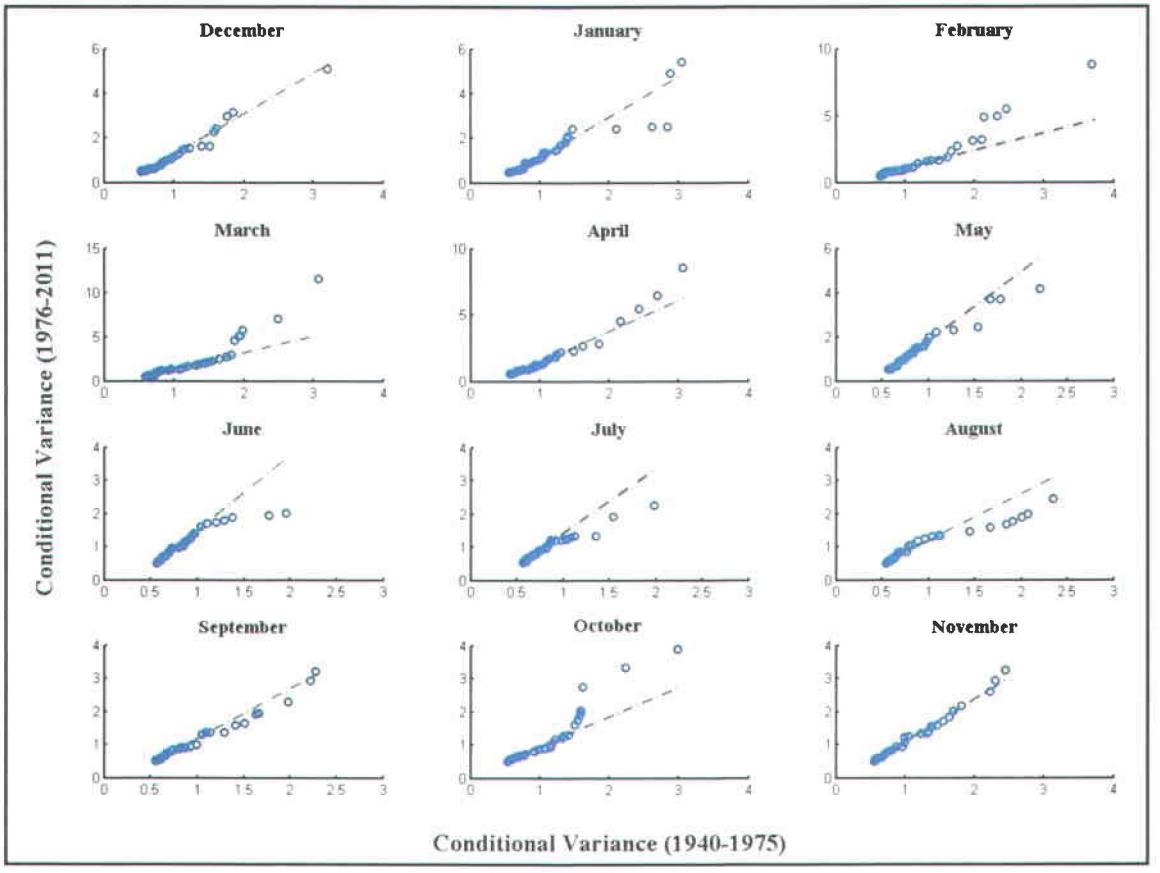


Figure 6. Monthly quantile-quantile plots for conditional variances before (*X*-axis) and after (*Y*-axis) 1975

Article 4. Modeling rainfall-runoff relationship using multivariate GARCH model

Modeling rainfall-runoff relationship using multivariate GARCH model

R. Modarres[a*], T. B. M. J. Ouarda[a,b],

[a] Hydroclimate modeling group, INRS-ETE, 490 De La Couronne, Québec, QC, Canada

[b] Masdar Institute of Science and Technology, P.O. Box 54224, Abu Dhabi, UAE

Tel: +1 418 654-3842, Fax: +1 418 654-2600,

E-mail: Reza.Modarres@ete.inrs.ca or touarda@masdar.ac.ae

[*] Corresponding author

Submitted for publication to Journal of Hydrology

Abstract

This paper aims to introduce a multivariate GARCH (MGARCH) modeling approach by developing and applying the bivariate diagonal VECH, the Constant Conditional Correlation (CCC) specifications and the ARMAX-GARCH model to estimate the time varying variance-covariance relationship of the rainfall-runoff processes for a sub-basin in the Saint Laurent (south-east) catchment, Quebec province, Canada. The stationarity and nonlinearity of the rainfall-runoff's conditional variance-covariance process are also tested. The proposed methods indicate that the residuals of a linear ARMAX model show a Heteroscedastic effect which can be removed by a GARCH model. However, the nonlinearity still exists in the residuals. The bivariate diagonal VECH-GARCH(1,1) and CCC-GARCH(1,1) models indicated both short-run and long-run persistency in the conditional variance-covariance matrix of the rainfall-runoff process. The conditional variance of rainfall appears to have a stronger persistency, especially long-run persistency, than the conditional variance of streamflow which shows a short-lived drastic increasing pattern and a stronger short-run persistency. The conditional covariance and conditional correlation coefficients seem to be different for each bivariate rainfall-runoff process with different degrees of stationarity and nonlinearity. The spatial and temporal pattern of variance-covariance features may reflect the signature of different physical and hydrological variables such as drainage area, topography, soil moisture and ground water fluctuations on the strength, stationarity and nonlinearity of the conditional variance-covariance for a rainfall-runoff process.

Key words: Rainfall-Runoff, Bivariate GARCH, ARMAX-GARCH, conditional covariance, diagonal VECH, Conditional Correlation.

1. Introduction

Hydro-meteorological processes have a dynamic and complex structure in space and time and most of their properties are not well understood. One of these properties is the nonlinearity of the hydrologic variables in space and time. A major concern in hydrologic modeling is, therefore, whether the underlying process should be modeled as linear or nonlinear.

The rainfall-runoff (R-R) process is a well-known highly complex and nonlinear phenomenon in hydrology with obvious relevance for water resources management. A considerable effort has been devoted to develop models for runoff estimation from rainfall data and to understand the dynamic features of the R-R process. Deterministic (physical), conceptual and parametric (empirical) models are the terms used for R-R model classification (Dawson and Wilby, 2001). The first model is based on physical laws of mass and energy transfer while the second provides a simplified representation of hydrological processes. In practice, the lack of sufficient (physical) data often impedes the application of these types of models. The third category represents the mathematical transfer functions to transfer meteorological variables into runoff. These methods are popular and commonly used for modeling hydrological processes such as the R-R process as they are often cheaper, simpler to implement and require less data than other models.

Statistical methods such as multivariate regression models (Wang et al., 2008; Hundecha et al., 2008; McIntyre and Al-Qurashi, 2009, among others), artificial neural networks (Kumar et al., 2005; Nayak et al., 2007; Machado et al., 2011, among others) and multivariate time series models have been considered in the third category for R-R modeling.

Although time series models are well known and applied in hydrology, climatology and water resources management, the multivariate time series approaches are rarely applied for the modeling of hydrologic processes such as the R-R process (e.g. Cooper and Wood, 1982;

Camacho et al., 1987). More recently, Niedzielski (2007) applied a multivariate Autoregressive (MAR) technique for regional scale R-R process. However, nonlinear multivariate time series approaches have not been investigated and applied for R-R modeling.

One of the most common nonlinear time series models is the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model which has been very popular in finance and econometrics since the seminal papers of Engle (1982) and Bollerslev (1986) which extended traditional ARMA models for the conditional mean to essentially analogous models for conditional variance. These models are now commonly used to describe changes in the volatility of financial time series. Further extension of this univariate model to a multivariate model, usually analogous to the extension from the ARMA to the vector ARMA was proposed by Engle and Kroner (1995). The Multivariate GARCH (MGARCH) models have then been commonly applied to study the relations between the volatilities and co-volatilities of several markets (e.g. De Goeij and Marquering, 2004; Serban et al., 2007; Cai et al., 2009; Syllignakis and Kouretas, 2011).

The objectives of the present paper are twofold. First, it aims to fill the gap of developing nonlinear time series models in hydrology and water resources using MGARCH model for a rainfall-runoff process. The MGARCH approach is not a rainfall-runoff model such as physically based models or artificial neural networks which are already applied in hydrology. This approach aims to model volatility and co-volatility or the conditional variance-covariance structure existing in a rainfall-runoff process. In addition to introducing MGARCH models, this study also investigates and compares the stationarity and nonlinearity of the conditional variance covariance for different rainfall-runoff processes in a catchment scale.

In order to pursue the above objectives, the paper is organized as follows: after the introduction section, the model specifications of both linear and nonlinear multivariate time series models and the testing materials are presented in sections 2 and 3. The paper continues to section 4 with an example of conditional variance and covariance modeling of the R-R process. The conclusions and suggestions for future work are finally presented in sections 5 and 6.

2. Model specifications

2.1. ARMAX model with multiple inputs

The general Autoregressive Moving Average model with multiple eXogenous variables for R-R modeling (adapted from Hipel and McLeod, 1996) can be expressed as:

$$\phi_p(B)Y_t = \omega(B)X_t + \theta_q(B)\varepsilon_t \qquad (1)$$

where Y_t is the observed runoff time series, X_t is the observed rainfall time series, $\phi_p(B)$ is a polynomial of order p, $\theta_q(B)$ is a polynomial of order q, ω is the transfer function, B is the backward operator, ε_t is an independent identically distributed (i.i.d) normal error with a zero mean and variance σ_ε^2 respectively.

The basic assumption of the above model is that the R-R process is a linear function and the residuals of the model are time independent. Although this assumption is usually tested and verified in time series model building, the time varying variance, σ_ε^2, is often ignored. In the next section, a procedure is presented for testing the existence of a time varying variance in the residuals of a linear model.

2.2. Testing for ARCH effect

Testing the existence of a time varying variance, or heteroscedasticity, in the residuals of an ARMAX model or testing for an ARCH (Autoregressive Conditional Heteroscedasticity) effect is usually carried out by the Engle's Lagrange Multiplier test.

The Engle's Lagrange Multiplier test for the ARCH effect was proposed by Engle (1982). The test statistic is given by NR^2, where R is the sample multiple correlation coefficient computed from the regression of ε_t^2 on a constant and $\varepsilon_{t-1}^2,\ldots,\varepsilon_{t-V}^2$ and N is the sample size. The null hypothesis of no ARCH effect is accepted if the test statistic is asymptotically distributed as a chi-square distribution with v degrees of freedom. The test can also be used to investigate the Generalized ARCH (GARCH) effect in the residuals (Bollerslev, 1986).

In the presence of an ARCH effect or time varying variance in the residuals of a (linear) time series model, an ARCH or a GARCH model is fitted to the squared residuals of the model in order to remove the heteroscedasticity from the residuals. In the next section, the details of the ARCH/GARCH models are presented.

2.3. Univariate GARCH model

Assume a univariate serially uncorrelated, zero mean process, u_t. The u_t is said to follow an Autoregressive Conditionally Heteroscedastic process of order v (or ARCH(v)) if the conditional distribution of u_t, given its past observations, u_{t-1}, u_{t-2}, …, has zero mean and conditional variance

$$\sigma_t^2|t-1 = Var(u_t|\Omega_{t-1}) = E(u_t^2|\Omega_{t-1}) = \alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_v u_{t-v}^2 = \alpha_0 + \sum_{i=1}^{v} \alpha_i u_{t-i}^2 \quad (1)$$

$$u_t|\Omega_{t-1} \sim N(0, \sigma_t) \quad (2)$$

Where Ω_{t-1} is the conditional distribution function, σ_t^2 is the conditional variance and α is the model parameter. This ARCH process was first introduced by Engle (1982) and allows the conditional variance change over time as a function of past observations.

The above ARCH model was then developed by Bollerslev (1986) to the Generalized ARCH (GARCH) model by entering the lagged conditional variance. The GARCH model is then written as follows

$$\sigma_t^2|t-1 = \alpha_0 + \sum_{i=1}^{V} \alpha_i \, u_{t-i}^2 + \sum_{j=1}^{M} \beta_j \, \sigma_{t-j}^2 \qquad (3)$$

Where $\alpha_i \ and \ \beta_j$ are the model parameters to be estimated. For the above GARCH model, the higher the $\alpha + \beta$, the higher the intensity of persistence in volatility of the time series. It should also be noted that the ARCH parameter, α, and the GARCH parameter, β, indicate short-run and long run persistence, respectively.

The model in which the conditional mean is modeled by an ARMAX model and the conditional variance remaining in the residuals of the ARMAX model (ε_t) is captured by the GARCH model is referred to as ARMAX-GARCH error (or ARMAX-GARCH) model. This model has two advantages. First, the heteroscedasticity has been eliminated from the residuals and second, the effect of eXogenous variables (X_t) on the dependent variable (Y_t) has also been eliminated. Therefore, the residuals of an ARMAX-GARCH model are supposed to be innovation or white noise of the dependent variable free from the effects of the eXogenous variables.

2.4. Multivariate GARCH model

2.4.1. Background

The multivariate GARCH model is the extension of the univariate model for modeling the temporal relationship between the conditional variance of two variables. For example in our

case, we are interested in understanding if there is a link between the variance of rainfall and streamflow and how this link changes through time.

Having a K-dimensional zero mean, serially uncorrelated process $u_t = (u_{1t}, \ldots, u_{kt})'$, which can be represented as

$$u_t = \sum_{t|t-1}^{1/2} \varepsilon_t \tag{4}$$

Where ε_t is a k-dimensional i.i.d. white noise, $\varepsilon_t \sim i.i.d\ (0, I_k)$ and $\sum_{t|t-1}$ is the conditional covariance matrix of u_t, given u_{t-1}, u_{t-2},....

By defining the conditional distribution of u_t in the following form

$$\Omega_{t-1} = \{u_{t-1}, u_{t-2}, \ldots\} \tag{5}$$

$$u_t \left| \Omega_{t-1} \sim (0, \sum_{t|t-1}) \right. \tag{6}$$

We can then define the MGARCH model as the following

$$VECH(\sum_{t|t-1}) = \omega_0 + \sum_{i=1}^{v} \Gamma_j vech(u_{t-j} u'_{t-j}) + \sum_{j=1}^{m} G_j vech(\sum_{t-j|t-j-1}) \tag{7}$$

Where VECH denotes the half-vectorization operator which stacks the columns of a square matrix from the diagonal downwards in a vector, ω_0 is a $\frac{1}{2}K(K+1)$- dimension vector of constants and the Γ_j's and G_j's are $(\frac{1}{2}K(K+1) \times \frac{1}{2}K(K+1))$ coefficient matrices.

For example, a bivariate GARCH(1,1) model is

$$VECH\begin{bmatrix}\sigma_{11,t|t-1} & \sigma_{12,t|t-1}\\ \sigma_{12,t|t-1} & \sigma_{22,t|t-1}\end{bmatrix}=\begin{bmatrix}\sigma_{11,t|t-1}\\ \sigma_{12,t|t-1}\\ \sigma_{22,t|t-1}\end{bmatrix}=\begin{bmatrix}\omega_{11}\\ \omega_{12}\\ \omega_{22}\end{bmatrix}+\begin{bmatrix}\gamma_{11} & \gamma_{12} & \gamma_{13}\\ \gamma_{21} & \gamma_{21} & \gamma_{23}\\ \gamma_{31} & \gamma_{32} & \gamma_{33}\end{bmatrix}\begin{bmatrix}u_{1,t-1}^2\\ u_{1,t-1}u_{2,t-1}\\ u_{2,t-1}^2\end{bmatrix} \quad (8)$$
$$+\begin{bmatrix}g_{11} & g_{12} & g_{13}\\ g_{21} & g_{21} & g_{23}\\ g_{31} & g_{32} & g_{33}\end{bmatrix}\begin{bmatrix}\sigma_{11,t-1|t-2}\\ \sigma_{12,t-1|t-2}\\ \sigma_{22,t-1|t-2}\end{bmatrix}$$

The above vech parameterization of MGARCH model has an important drawback. The number of parameters in Eq. (8) equals (K(K+1)/2)(1+2(K(K+1)/2) which becomes excessively large as K increases. For example, the above bivariate GARCH(1,1) model has 21 parameters. To overcome such a problem, some specifications of a VECH model have been introduced. In this study, we introduce and apply the two following model specification for R-R process.

2.4.2. Diagonal VECH model

Bollerslev et al., (1988) suggested a diagonal MGARCH model called diagonal VECH (DVECH) model. The diagonal model constrains the matrices Γ_j and G_j to be diagonal. In this case, the conditional covariance between $u_{1,t}^2$ and $u_{2,t}^2$ depends on lagged cross-products of the two random processes (variables) involved and lagged values of the covariance itself.

The number of parameters in the diagonal GARCH(1,1) equals 3(K(K+1)/2). For the bivariate case, therefore, we only have 9 parameters to be estimated instead of the 21 parameters in the VECH representation. The bivariate DVECH(1,1) model can then be written as follows:

$$vech\begin{bmatrix}\sigma_{11,t|t-1} & \\ \sigma_{12,t|t-1} & \sigma_{22,t|t-1}\end{bmatrix}=\begin{bmatrix}\omega_{11} & \\ \omega_{21} & \omega_{22}\end{bmatrix}+\begin{bmatrix}\gamma_{11} & \\ \gamma_{21} & \gamma_{22}\end{bmatrix}\begin{bmatrix}u_{1,t-1}^2 & \\ u_{1,t-1}u_{2,t-1} & u_{2,t-1}^2\end{bmatrix} \quad (9)$$
$$+\begin{bmatrix}g_{11} & \\ g_{21} & g_{21}\end{bmatrix}\begin{bmatrix}\sigma_{11,t-1|t-2} & \\ \sigma_{21,t-1|t-2} & \sigma_{22,t-1|t-2}\end{bmatrix}$$

It should be noted that this matrix form of a DVECH model is defined using the Hadamard or element-by-element product. The above matrix notation can be rewritten as follows to get the conditional variance-covariance time series:

$$\sigma_{11,t} = \omega_{11} + \gamma_{11}\, u_{1,t-1}^2 + g_{11}\, \sigma_{11,t-1|t-2} \quad (10)$$

$$\sigma_{21t} = \omega_{21} + \gamma_{21}\, u_{1,t-1}\, u_{2,t-1} + g_{21}\, \sigma_{21,t-1|t-2} \quad (11)$$

$$\sigma_{22,t} = \omega_{22} + \gamma_{22}\, u_{2,t-1}^2 + g_{22}\, \sigma_{22,t-1|t-2} \quad (12)$$

Here, $\sigma_{11,t}$ and $\sigma_{22,t}$ are the conditional variance of rainfall and runoff respectively and σ_{21t} indicates the conditional covariance or co-volatility between rainfall and runoff time series.

2.4.3. Constant Conditional Correlation model

Bollerslev (1990) proposed a class of the MGARCH model in which the conditional covariances are proportional to the product of the corresponding conditional standard deviation and the constant conditional correlation (CCC). The CCC-MGARCH model was then widely applied in empirical research mostly because of its computational simplicity (Tse and Tsui, 2002).

By applying the model, the conditional covariance between two time series, i and j, is defined as

$$\sigma_{ijt} = D_t R D_t = \rho_{ij}\sqrt{\sigma_{iit}\sigma_{jjt}} \quad (13)$$

Where

$$D_t = diag(\sigma_{11t}^{\frac{1}{2}} \dots \sigma_{KKt}^{\frac{1}{2}}) \quad (14)$$

Where σ_{iit} and σ_{jjt} can be defined by any univariate GARCH model and (ρ_{ij}) is the constant conditional correlation. This model is applied in this study to estimate the conditional correlation and covariance between rainfall and runoff.

3. Simulation experiment

In order to verify the empirical models, we simulated conditional covariances for both DVECH(1,1) and CCC(1,1), and DVECH and CCC(2,2) models. We consider the parameters of the empirical models for simulation and compare the simulated and experimental conditional covariances using two criteria, Normalized Root Mean Square Error (NRMSE) and normalized BIAS. These criteria can be written as follows

$$NBIAS = \frac{1}{k}\sum_{i=1}^{k}\frac{sim_{CCOV}-emp_{CCOV}}{emp_{CCOV}} \quad (15)$$

$$NRMSE = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(\frac{sim_{CCOV}-emp_{CCOV}}{emp_{CCOV}})^2} \quad (16)$$

Where sim_{CCOV} and emp_{CCOV} denote the simulated and empirical conditional covariances and k is the number of estimates.

4. Testing procedures

As the rainfall-runoff process is usually perceived as a nonlinear or nonstationary process, investigation on the non-stationarity and nonlinearity of the second order moment and the variance-covariance structure of the R-R process are examined using the following tests:

4.1. Test for Stationary

The stationary test of the variance-covariance structure of the R-R process is carried out using two methods, the Augmented Dickey and Fuller (ADF) unit root test (Dickey and Fuller, 1979) which tests for the presence of a unit root in the series; and the KPSS test proposed by Kwiatkowski et al. (1992), which tests for stationarity around a deterministic trend and around a fixed level.

4.2. Test for Nonlinearity

The nonlinearity in the variance-covariance structure is tested by the Brock-Dechert-Scheinkman (BDS) test (Brock et al., 1996). The BDS statistic can be used as a diagnostic tool to approve a deterministic nonlinear dynamic in time series. The null hypothesis is that the time series sample comes from an independent identically distributed (i.i.d) process. The BDS test embeds the scalar time series into an m-dimensional space and calculates the correlation integral. Under the null hypothesis, Brock et al., (1996) exploited the asymptotic normality of the correlation integral and calculated the BDS statistic. For the sake of brevity, the mathematical formulation of the BDS test is not presented here and the reader is referred to Wang et al., (2006) for detailed information. In this study, we consider m=2, 3 and 4 to test the nonlinearity in our time series.

5. Case Study

To illustrate the use of the above tools, we consider a catchment with two streamflow and four rainfall gauging stations in a hydrologic region called south-east Saint-Laurent watershed in Quebec Province, Canada. The study watershed and the location of the stations are shown in Figure 1.

The two neighboring streamflow stations are Bras D'Henri and Beaurivage which are called "F34' and "F7" in the rest of the paper. The four rainfall stations are St. Flavien, St. Pierre, St. Severin and Scott stations (Table 1) which are called "F", "P", "S" and "SCT" in the following sections. Five years (1996-2000) of continued concurrent daily streamflow and rainfall time series have been used in this study to illustrate the conditional variance-covariance structure of the R-R process in the basin.

5.1. ARMAX-GARCH modeling

5.1.1. Estimating the ARMAX model

The first step in the estimation of the model is to determine the order of the model through investigating the autocorrelation function (ACF) of streamflow and cross correlation between rainfall and streamflow time series. The cross correlation functions between streamflow and rainfall time series are given in Figure 2 for different lag times. This figure illustrates the significant, but not very strong, correlation between rainfall and runoff at lags k=1 to k=5. This indicates that the model may include one to five past rainfall observations as the eXogenous variables. The order of the ARMA model for streamflow is identified based on the ACF and PACF (not shown in the figures). In this study, the logarithmic transformed streamflow is used for ARMAX modeling.

Different ARMAX models are then developed for modeling the R-R time series relationship using different rainfall time series as eXogenous variables for each streamflow time series. The parameters of the models are estimated by the maximum likelihood method and the ACFs and PACFs of the residuals are investigated to be time independent for testing model adequacy. In the case of more than one adequate model, the best model is selected according to the Akaike

information Criteria (AIC) for each streamflow station. The order of the model and the rainfall time series included in each model are given in Table 2 for each streamflow time series.

5.1.2. Testing ARCH effect in the residuals

In accordance to the ARMAX models, the residuals (or innovations) of each model are tested for the time dependency and ARCH effect using ACF and Engle's test. The ACF of the residuals and squared residuals of the ARMAX models are given in Figure 3. The figure suggests the adequacy of the ARMAX model as the autocorrelation coefficients are within the confidence level. The ACF of the squared residuals shows a significant coefficients at lag k=1 for "F7" station and lag k=1 as well as higher lags (k=6, k=20) for "F34" station.

In addition, the results of Engle's test are also given in Figure 4 for k=1-15 lag times. The p-values of the test indicate the heteroscedasticity in the residuals of the ARMAX model. It is observed that the test's p-values are below the critical level ($\alpha = 0.05$) at k= 1 for "F7" and at lags k=1 and k=6 for "F34". Therefore, the ARCH effect exists in the residuals of the ARMAX model which can be removed by applying a GARCH model.

5.1.3. Estimating the GARCH model

As the ARMAX model for the R-R process fails to incorporate the nonlinear structure of the residuals, a GARCH model will be fitted to the squared residuals to take into account the time varying higher order moment of the R-R process. The order of the GARCH model fitted to the residuals of the ARMAX model is given in Table 2. The conditional variance for the residuals of the ARMAX models for "F7"and "F34"stations can therefore be written as follows:

F7: $\sigma_t^2 = 0.09 + 0.58\varepsilon_{t-1}^2 + 0.41\sigma_{t-2}^2$ (17)

F34: $\sigma_t^2 = 0.056 + 0.46\varepsilon_{t-1}^2 + 0.21\varepsilon_{t-2}^2$ (18)

The GARCH models fitted to the residuals of the ARMAX model are supposed to remove heteroscedasticity from the residuals. The p-values of the Engle's test for the residuals of ARMAX-GARCH model for the R-R process are given in Figure 5. The p-values suggest that the null hypothesis of no ARCH effect in the residuals cannot be rejected and no heteroscedasticity remains in the residuals of the ARMAX-GARCH model.

Moreover, in order to evaluate the empirical results of ARMAX-GARCH modeling, we generate 2000 realizations of the ARMAX-GARCH model for each streamflow time series and estimate their (theoretical) parameters. The box-plots of the simulated parameters and the empirical parameters are given in Figures 6 and 7 for "F7" and "F34" time series respectively. The simulation results indicate that the empirical models are valid for the estimation of the volatility of streamflow time series as the parameters of the models (both mean and variance equations' parameters) fall within the 25 and 75 percentiles of the simulated parameters.

Referring to the above GARCH models (Eq. 15), one can see both short-run and long-run persistence effect in the residuals of the streamflow at Beaurivage gauging station ("F7") as both ARCH and GARCH parameters are significant in the model. On the other hand, only short-run persistence of the heteroscedasticity is observed for residuals of streamflow at Bras D' Henri station ("F34") as the model includes only positive and significant α parameters. The measure of the persistence in volatility $(\alpha + \beta)$ therefore confirms that heteroscedastic persistency is stronger for "F7" time series.

We can also check this persistency by plotting the lagged residuals against conditional variance (Figure 8). As if u_t process is time independent no pattern will appear, this figure shows that the conditional variance at each time step depends on the previous residuals for both streamflow time series. As the ARMAX model is supposed to filter the effect of rainfall by the regression

process, Figure 8 suggests that the heteroscedasticity remaining in the runoff process is influenced by other factors such as catchment properties rather than rainfall heteroscedasticity.

5.1.4. Stationarity and nonlinearity test

The results of ADF, KPSS and BDS tests for the residuals of the ARMAX and ARMAX-GARCH model as well as the conditional variance of the streamflow's residuals are given in Table 3.

All the residuals and conditional variance time series appear to be stationary according to the ADF test and pass the KPSS stationarity test at 10% level, which shows no significant change in mean. The KPSS results for trend stationarity test also indicate no significant trend in the residuals and conditional variance time series.

The BDS test reveals a significant nonlinearity for all time series for all dimensions, m=2, 3 and 4. The results show no sharp difference between nonlinearity of the residuals of the models for the two streamflow time series, "F7"and "F34" while the nonlinearity for "F34" is slightly stronger than that for "F7".

5.2. Multivariate GARCH models

The above analysis showed the heteroscedasticity in the streamflow innovations. In this section, we develop a bivariate GARCH model to illustrate the heteroscedastic features of R-R process.

5.2.1. Bivariate GARCH model for "F7" streamflow time series

5.2.1.1. DVECH model estimation

To develop the bivariate GARCH (1,1) model, we consider the streamflow time series as a dependent variable and the rainfall time series as an input variable and estimate 4 models for heteroscedastic R-R relationship for each streamflow time series. The parameters of each

bivariate DVECH(1,1) model are estimated by the maximum likelihood method. These parameters are significant according to asymptotic t-statistics (Table 4). For example, the conditional variance-covariance matrix for "S" and "F7" time series can be written as follows

$$\sigma_{11,t} = 0.421 + 0.048 u_{1,t-1}^2 + 0.948 \sigma_{1,t-1} \quad (19)$$

$$\sigma_{21t} = 0.031 + 0.055 u_{1,t-1} u_{2,t-1} + 0.862 \sigma_{21,t-1} \quad (20)$$

$$\sigma_{22,t} = 0.105 + 0.708 u_{2,t-1}^2 + 0.216 \sigma_{2,t-1} \quad (21)$$

By applying the above model, the conditional variance for "S" and "F7" time series and their co-variance are estimated and illustrated in Figure 9. The conditional covariance between "F7" and other rainfall time series, "F", "P" and "SCT", are shown in Figure 10.

According to table 4, the measure of intensity of persistence, $\alpha + \beta$, for R-R covariance between rainfall and "F7" streamflow time series are 0.917, 0.890, 0.864 and 0.864 for "S", "P", "F" and "SCT", respectively. These values suggest the highest covariance persistency for "S"-"F7" process. The results show a high persistency in R-R covariance structure as the persistency measurements are relatively high. Calculating the persistency measurement for rainfall conditional variance also reveals the highest persistency for "S" and "P" time series followed by "F", "SCT". On the other hand, the persistency of conditional variance of runoff time series is less than those for all rainfall time series, though it is still relatively high. Figure 9 also shows that the conditional variance of rainfall seems to be more dynamic (more volatile) than the conditional variance of runoff.

The conditional covariances between rainfall time series and "F7" (Figures 9 and 10) show a low level of temporal fluctuation. The magnitude of covariance is different from station to station and from time to time. It can also be seen that the conditional covariance tends to decrease rapidly after a major increase and remains usually low for a long time before another peak. The

conditional covariance is almost at the lowest level during 1997-1998 and becomes stronger during 1999-2000 for all R-R processes.

We next move to estimate the conditional correlation coefficients in order to examine how the R-R correlation coefficient changes over time. Figure 11 shows the evolution of the estimated conditional correlation coefficients between "F7" and rainfall time series. The analysis of conditional correlation patterns reveals low level of rainfall association to direct runoff generation. The correlation coefficients usually remain near or below the average value during 1996-2000, particularly during 1997-1999. The streamflow seems to have a higher correlation with "SCT" and "S" than other rainfall time series.

Similar to conditional covariance, the conditional correlation does not remain in a constant position and tends to change from high to low values and vice versa. It is also interesting to note that the ratio of the frequency of positive to negative correlation coefficients is decreasing form "SCT", to "S", "P" and "F", respectively.

5.2.1.2. CCC model

The bivariate CCC-GARCH(1,1) model estimation is carried out using the maximum likelihood method. The parameters of the conditional variance for the rainfall time series are illustrated in Table 5. In addition, the constant conditional correlations (ρ_{ij}) between rainfall and runoff time series, as well as the unconditional correlation coefficient, are presented in this table. These coefficients also indicate a weak correlation between rainfall and runoff at Beaurivage River (station "F7"). Both constant conditional and unconditional correlation coefficients indicate the highest association between "SCT" and "F7" followed by "S", "P" and "F" stations (the same order as showed by DVECH model).

The parameters of each univariate GARCH model for rainfall and runoff time series are also given in this table. The parameters of the GARCH models indicate the higher persistency level for rainfall time series as the persistency measurement is higher for rainfall GARCH models. It is also observed that the long-run persistence of rainfall is stronger than that for streamflow time series in contrast to short-run persistence. These results are in agreement with the persistency characteristics obtained by DVECH model (Table 4).

The R-R conditional covariances obtained by the CCC model are illustrated in Figure 12. A similar temporal pattern to the conditional variance of the runoff (see R-R variance-covariance matrix in Figure 9) is observed for conditional covariances. Indeed, it is an expected feature as the conditional covariance is a product of constant conditional correlation and conditional variance of rainfall. Furthermore, this figure illustrates a weaker covariance structure for R-R process than those estimated by DVECH model (Figure 10).This phenomenon may also be related to the weak constant conditional correlation between rainfall and runoff process.

5.2.1.3. Model verification

In order to verify the empirical MGARCH models, the conditional covariance for both low order (1,1) and high order (2,2) model are simulated for DVECH and CCC models. The two evaluation criteria, given in Table 6, show that the empirical DVECH(1,1) model is quite better than DVECH (2,2) model for all R-R conditional covariance. This implies that the R-R covariance at time t depends significantly on the variance-covariance structure at time t-1 but not on t-2.

The results also show that CCC(1,1) and CCC(2,2) models do not give significantly different R-R conditional covariances. This is perhaps due to the constant conditional correlation assumption and perhaps because the low order GARCH(1,1) models, fitted to streamflow and rainfall time series separately, are enough for conditional variance of rainfall and runoff. The low criteria

values for CCC models also suggest that the GARCH(1,1) model is appropriate for estimating the conditional variance of rainfall and streamflow time series and no higher order models are necessary.

5.2.2. Bivariate GARCH model for "F34" streamflow time series

5.2.2.1. DVECH model

The same procedure is followed for estimating the R-R conditional variance-covariance matrix for "F34" station. Table 7 shows the estimated parameters for the bivariate GARCH(1,1) model. It is observed that these parameters are all statistically significant and indicate a meaningful covariance between the rainfall and runoff time series. For example, the conditional variance-covariance matrix between rainfall at station "SCT" and "F34" streamflow time series is presented in Figure 13. The conditional R-R covariance can therefore be written as follows:

$$\sigma_{21t} = 0.061 + 0.022\, u_{1,t-1}\, u_{2,t-1} + 0.918\, \sigma_{21,t-1} \qquad (20)$$

According to the parameters for bivariate GARCH models, the intensity of persistency for the conditional covariance between rainfall and ''F34'' time series are 0.979, 0.986, 0.966 and 0.940 for "S", "P", "F" and "SCT", respectively.

Based on the conditional covariance models, the temporal variation of conditional covariance between "F34" and rainfall time series are estimated and illustrated in Figure 14. The temporal pattern of conditional covariance does not show a high level of fluctuation. It shows some sharp short-lived positive covariances only in 1996 and 2000 and a sharp negative covariance in 1997 and usually remains low before increasing.

On the other hand, we look at the time varying conditional correlation coefficients (Figure 15). The figure suggests a high time varying and fluctuation in conditional correlation coefficients. They tend to change rapidly from one status (high or low value) to the opposite status through

time. Visually, some regular but weak fluctuation is also observed for conditional correlation coefficients between rainfall and runoff at Bras D'Henri basin.

5.2.2.2. CCC model

The bivariate CCC-GARCH(1,1) model parameters and the constant conditional (ρ_{ij}) and unconditional correlation between rainfall and runoff time series for "F34" time series are presented in Table 8. Both constant conditional and unconditional correlation coefficients indicate the highest correlation between "SCT" and "F34" time series followed by "S", "P" and "F". Moreover, one can see, again, that the persistency of the conditional variance is stronger for rainfall time series while the conditional variance of streamflow shows a strong short-run persistence in contrast with the rainfall time series.

The R-R conditional covariances estimated by the CCC model are presented in Figure 16. The same features as the conditional covariances between rainfall and "F7" time series (Figure 12) can be observed here.

5.2.2.3. Model verification

The criteria for verifying the empirical models for conditional covariance between "F34" and rainfall time series are given in Table 9. The NRMSE and NBIAS indicate that both DVECH(1,1) and DVECH(2,2) models have small and similar errors. However, following the parsimonious principle, the DVECH(1,1) model is a better choice. The DVECH(1,1) model has 9 parameters while DVECH(2,2) model has 15 parameters. In addition, no significant difference is also observed between CCC(1,1) and CCC(2,2) models for estimating conditional covariance, regarding the error criteria. Therefore, the empirical (1,1) models seem to be appropriate and enough for estimating R-R conditional variance-covariance relationship.

5.2.3. Stationarity and nonlinearity test

The stationarity and nonlinearity of the conditional covariance between rainfall and runoff time series are discussed in this section. According to Table 10, we cannot accept the unit root hypothesis with ADF test at 1% significance level for conditional covariances of the R-R process. It is observed that the stationarity of the conditional covariances between "P", "S" and "F34" is stronger than those corresponding to "F7" while the conditional covariances between "F", "SCT" and "F7" seem to be more stationary than those corresponding to "F34". The stationarity in the mean and trend of conditional covariance is not rejected either according to KPSS level stationary tests.

On the other hand, the BDS test rejects the linearity hypothesis in conditional covariance for all R-R processes in the basin. It can be seen that the nonlinearity degree of conditional covariance is stronger between "P", "S" and "F7" than between "F", "SCT" and "F7" for different dimensions and an opposite condition is observed for "F34". This suggests the reverse action of the nonlinearity and stationarity for the conditional covariance of the R-R process and the impact of scale (drainage area) and distance between rainfall and runoff stations on the degree of nonlinearity and stationarity for our case study.

6. Summary and concluding remarks

Multivariate nonlinear time series models have not yet been applied to investigate the time varying second order moment or conditional covariance for multivariate processes in hydrology, climatology and water resources. This study developed and applied a multivariate nonlinear GARCH time series modeling approach with the objective of understanding the R-R process

from a new perspective. From the ARMAX-GARCH and multivariate GARCH models for the R-R example in this study, the following conclusions can be derived:

6.1. From the linear ARMAX model, with rainfall time series as the multiple eXogenous inputs, the existence of an ARCH effect or the time varying variance in the residuals is evident according to the Engle's test. The GARCH model removes the heteroscedasticity but the nonlinearity still remains in the residuals of the ARMAX-GARCH model after removing conditional variance and rainfall effects. This may imply the effect of watershed factors on the nonlinearity remaining in the residuals among which the scaling variable or the drainage area is the most important one (Sivapalan et al., 2002).

6.2. The conditional variance of the streamflow's residuals appears to be stationary. The stationarity and persistency of the conditional variance for the larger basin is stronger than those for a smaller basin. These findings may show that in a large basin, the conditional variance of runoff is affected by the watershed features such as the drainage area and river network, surface abstraction and infiltration, time of concentration, etc, which attenuate the response of catchment to magnitude of the conditional variance of rainfall. A higher persistency and stationarity in the innovations of ARMAX-GARCH model at larger basin is also observed. The BDS test for nonlinearity also shows a relatively higher level of nonlinearity in the runoff of the smaller basin where short-run memory of the conditional variance is observed. This is in agreement with Wang et al., (1981) who indicated that the catchments become more linear with increasing catchment size.

6.3. Using the diagonal bivariate GARCH model, it was revealed that the conditional variance of rainfall and runoff are dynamic through time. However, the conditional variance of rainfall seems to have a long-run memory while the conditional variance of runoff shows a strong

short-run memory which is probably indicated by a sudden sharp increase in the conditional variance.

6.4. The conditional covariance between rainfall and runoff shows a significant memory and persistence. However, it is observed that the conditional covariance did not remain high for a long time for both gauging stations. This phenomenon was also seen for the conditional correlation between rainfall and runoff time series. It may be attributed to the driving factors behind excess runoff production such as the catchment storage properties and soil moisture which delay the excess runoff to be generated and reach to the drainage network. The interaction between surface runoff, ground water and physical characteristics may also result in lower level of fluctuation of conditional covariance in the larger basin.

6.5.It seems that the hypothesis of the constant correlation between rainfall and runoff is not valid and the CCC model is not an appropriate approach to show the time varying R-R variance-covariance relationship in a catchment scale. In the other word, the runoff coefficient varies in time and space as the results of catchment physical and scale heterogeneity.

6.6. The conditional and unconditional correlation between rainfall and runoff differs from station to station. Both coefficients are larger for "F34" station which drains a (much) smaller basin than "F7" station. Therefore, one can conclude that the drainage area of a basin has an important role on the conditional correlation and covariance level for a R-R process from the results of this study. However, this needs to be examined and generalized by a larger data base.

6.7. The measurement of persistency indicates a high level of persistency in the R-R conditional covariance. It is observed that the persistency is higher between runoff and rainfall stations

located upstream of the basins. This is probably an indicator of the effect of drainage area on the persistency of the R-R covariance process. Nonetheless, this statement needs also to be generalized by future careful investigations.

6.8. The stationarity and nonlinearity of the conditional covariance are different among different R-R processes. For example, the stationarity is stronger between ''F34'' and upstream rainfall stations while this stationarity becomes weaker for "F7". The intensity of stationarity may be related to the intensity of nonlinearity (as also stated by Wang et al., 2006) but our results indicate that the stationarity and nonlinearity may vary in the opposite direction.

6.9. The simulation experiment showed an important fact that the (1,1) models seem to be sufficient for modeling R-R conditional variance covariance structure. Though the simulation is not carried out in empirical finance, the MGARCH(1,1) model is believed to be sufficient in almost all financial cases. In addition, sufficient low order models are important from parsimonious point of view, especially for MGARCH models for which the number of parameters increases rapidly with the order of the model.

Generally, this study provides a nonlinear modeling and testing approach for conditional variance-covariance structure in R-R processes. The results indicate a strong nonlinearity in runoff time series which is related to rainfall nonlinearity which passes through the nonlinear catchment system and is recorded at the gauging stations.

7. Challenges for future work

This study presented a bivariate GARCH model for R-R process. The R-R process includes different nonlinear spatial and temporal phases and different effective parameters all of which were not considered in this study. Therefore, one can improve the variance–covariance matrix of the R-R process by adding other variables such as temperature, ground water or soil moisture time series for future studies. The effect of upstream flow on downstream flow is not considered in this study which may be a good topic for generalizing the results of this study. However, from the theory of MGARCH approach, the number of parameters of a model will increase rapidly by adding a variable into the model.

Finally, as most of the hydrologic processes are nonlinear and nonstationary due to climate change, the development and application of these models are strongly suggested for other multivariate processes in hydrology and climatology in the changing environment.

Acknowledgement

The authors acknowledge the financial support graciously provided by the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Canada Research Chair (CRC) Program. The authors wish also to thank Alain Rousseau for providing the map of study area.

References

Bollerslev, T.,1986. Generalized Autoregressive Conditional Heteroscedasticity, J. Econometrics 31, 307–327.

Bollerslev, T., 1990. Modeling the coherence in short-run nominal exchanges rates: a multivariate generalized ARCH model, Review of economics and statistics 72, 498-505.

Bollerslev, T., Engle, R.F. and Wooldridge, J.M., 1988. A capital asset pricing model with time varying covariances, J. Pol. Econ., 96, 116–31

Brock, W.A., Dechert, W. D., Scheinkman, J. A., and LeBaron, B. 1996. A test for independence based on the correlation dimension, Econ. Rev,. 15 (3), 197–235.

Cai, Y., Yeutien Chou, R., and Li, D., 2009. Explaining international stock correlations with CPI fluctuations and market volatility. J. Banking and Finance, 33, 2026-2035

Camacho, F., McLeod, A. I., and Hipel, K. W. 1987. Multivariate contemporaneous ARMA model with hydrological applications, Stoch. Hydrol. Hydraul., 1, 14-154

Cooper, D. M., and Wood, E.F., 1982. Identification of multivariate time series and multivariate input-output models, Water. Resour. Res., 18, 937-946

Dawson C.W., and Wilby, R. L., 2001. Hydrological modeling using artificial neural networks, Prog. Phys. Geog,. 25, 80–108.

De Goeij, P and Marquering, W., 2004. Modeling the conditional covariance between stock and bond returns: a multivariate GARCH approach, J. Fin. Econom., 2, 531-564

Dickey, D.A., and Fuller, W. A. , 1979. Distribution of the estimators for autoregressive time series with a unit root, J. Am. Stat. Assoc., 74, 423–431.

Engle, R. F., 1982. Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation, Econometrica, 50, 987-1008

Engle, R. F. and Kroner, K. F. , 1995. Multivariate simultaneous multivariate ARCH. Econom. theory, 11, 122-150

Goodrich, D. C., Lane, L. J., Shillito, R. M., Miller, S. N., Syed, K. H., and Woolhiser, D. A. .1997. Linearity of basin response as a function of scale in a semi-arid watershed, Water Resour. Res., 33(12), 2951– 2965

Hipel, K. W., and McLeod, A. E., 1996. Time series modeling of water resources and environmental systems. Elsevier, Amsterdam, The Netherlands

Hundecha, Y., Ouarda., T.B.M.J., and Bardossy, A. , 2008. Regional estimation of parameters of a rainfall-runoff model at ungauged watersheds using the "spatial" structures of the parameters within a canonical physiographic-climatic space. Water. Res. Res., 44, W01427, doi:10.1029/2006WR005439

Kumar, A.R.S., Sudheer, K.P., Jain, S.K., and Agarwal, P.K., 2005. Rainfall-runoff modelling using artificial neural networks: comparison of network types, Hydrol. Process., 19, 1277-1291, doi: 10.1002/hyp.5581

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y., 1992. Testing the null of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? J. Econometrics 54, 159–178.

McIntyre, N., and Al-Qurashi, A., 2009. Performance of ten rainfall-runoff models applied to an arid catchment in Oman, Environ. model. Software. 24, 126-738, doi: 10.1016/j.envsoft.2008.11.001

Machado, F., Mine, M., Kaviski, E., and Fill, H., 2011. Monthly rainfall-runoff modelling using artificial neural networks, Hydrol. Sci. J., 56, 349-361, doi: DOI: 10.1080/02626667.2011.559949

Nayak, P. C., Sudheer, K.P., and Jain, S.K., 2007. Rainfall-runoff modeling through hybrid intelligent system, Water Resour. Res., 43, W07415, doi: 10.1029/2006WR004930

Niedzielski, T. 2007. A data-based regional scale autoregressive rainfall-runoff model: a study from the Odra River, Stoch. Environ. Res. Risk Assess., 21, 649-664, doi: 10.1007/s00477-006-0077-y

Quible, R., Rousseau, A. N., Duchemin, M., Poulin, A., Gangbazo, G., and Villeneuve, J.P. 2006. Selecting a calculation method to estimate sediment and nutrient loads in streams: Application to the Beaurivage River (Quebec, Canada). J. Hydrol., 326, 295-310

Serban, M., Brockwell, A., Lehoczky, J., and Srivastava, S., 2007. Modelling the dynamic dependence structure in multivariate financial time series. J. Time Series A., 28, 763-782, doi: 10.1111/j.1467-9892.2007.00543.x

Sivapalan, M., Jothityangkoon, C., and Menabde, M. 2002. Linearity and nonlinearity of basin response as a function of scale: Discussion of alternative definitions, Water Resour. Res., 38(2), 1012, doi:10.1029/2001WR000482.

Syllignakis, M. M., and Kouretas, G.P., 2011. Dynamic correlation analysis of financial contagion: Evidence from the Central and Eastern European markets, Int. Rev. Economy and Fin., 20, 717-732, Doi: 10.1016/j.iref.2011.01.006

Tse, Y. K., and Tsui, A. K.C., 2002. A multivariate generalized autoregressive conditional heteroscedasticity model with time varying correlations. J. Business And economic statistics, 20, 351-362

Wang, Y. C., Chen, S.T., Yu, P. S., and Yang, T. C., 2008. Storm-event rainfall-runoff modelling approach for ungauged sites in Taiwan, Hydrol. Process., 22, 4322-4330, doi: 10.1002/hyp.7019

Wang C. T., Gupta, V.K., and Waymire, E., 1981. A geomorphologic synthesis of nonlinearity in surface runoff. Water Resour. Res., 19(3), 545-554

Wang W. Vrijling J.K. Van Gelder P.H.A.J.M. and Ma, J., 2006. Testing for nonlinearity of streamflow processes at different timescales. J. Hydrol., 322, 247-268

List of Tables

Figure Captions

Figure 1 Location of the study basin (source: Quible et al., 2006)

Figure 2 Cross correlation function between rainfall and runoff time series for different lag times

Figure 3 ACF of the residuals and squared residuals of the ARMAX models

Figure 4 p-values of the Engle's test for ARMAX models

Figure 5 p-values of the Engle's test for ARMAX-GARCH models

Figure 6 Boxplot of parameters for simulated ARMAX-GARCH models for "F7" time series. The gray circles show empirical parameters

Figure 7 Same as Figure 6 but for "F34" time series

Figure 8 Empirical conditional variance of innovation as a function of previous value

Figure 9 Conditional variance-covariance matrix for "F7" and "S" time series by DVECH model

Figure 10 Estimated conditional covariance between "F7" and rainfall time series by DVECH model

Figure 11 Estimated conditional correlation coefficients between rainfall and "F7" time series by DVECH model

Figure 12 Estimated conditional covariance between "F7" and rainfall time series by CCC model

Figure 13 Conditional variance-covariance matrix for "F34" and "SCT" time series by DVECH model

Figure 14 Same as Figure 10 but for "F34" and rainfall time series

Figure 15 Same as Figure 11but for "F34" streamflow time series

Figure 16 Same as Figure 12 but for "F34" and rainfall time series

Table 1 List of streamflow and rainfall gauging stations

	Station / River name	Station code in text	Latitude	Longitude	Drainage area (Km^2)
Rainfall stations	St. Pierre	"P"	46°15 N	71°13 W	-
	St. Severin	"S"	46°20 N	71°03 W	-
	St. Flavien	"F"	46°30 N	71°35 W	-
	Scott	"SCT"	46°30 N	71°05 W	-
Gauging stations	Beaurivage	"F7"	46°39 N	71°17 W	709
	Bras D'Henri	"F34"	46°32 N	71°20 W	137

Table 2 ARMAX -GARCH model parameters and variables for streamflow time series

streamflow series	eXogeneous Rainfall series	Mean equation	Variance equation
"F7"	"F", "P", "SCT"	$c = 0.07$ $\emptyset_1 = 0.93$ $\theta_1 = -0.09$	$\alpha_0 = 0.09$ $\alpha_1 = 0.58$ $\beta_1 = 0.41$
"F34"	"F", "S", "SCT"	$c = -0.019$ $\emptyset_1 = 0.41$ $\emptyset_2 = 0.45$ $\emptyset_3 = 0.077$ $\theta_1 = 0.56$	$\alpha_0 = 0.056$ $\alpha_1 = 0.46$ $\alpha_2 = 0.21$

Table 3 Stationarity and nonlinearity test results for ARMAX-GARCH models

series	Streamflow series	KPSS level stationary test		KPSS trend stationary test		ADF unit root test	
		Results	p-value	Results	p-value	Results	p-value
ARMAX residuals	"F7"	0.014	>0.1	-0.0036	0.77	-11.44	0
	"F34"	0.037	>0.1	-0.0012	0.38	-9.19	0
ARMAX-GARCH residuals	"F7"	0.021	>0.1	-0.0069	0.63	-10.83	0
	"F34"	0.019	>0.1	-0.0056	0.69	-10.72	0
Conditional variance	"F7"	0.067	>0.1	-0.0010	0.59	-9.75	0
	"F34"	0.048	>0.1	-0.0017	0.17	-8.44	0
BDS test							
series	station	m=2		m=3		m=4	
		statistic	p-value	statistic	p-value	statistic	p-value
ARMAX residuals	"F7"	0.035	0	0.064	0	0.082	0
	"F34"	0.042	0	0.069	0	0.083	0
ARMAX-GARCH residuals	"F7"	0.040	0	0.071	0	0.081	0
	"F34"	0.046	0	0.077	0	0.094	0
Conditional variance	"F7"	0.055	0	0.089	0	0.104	0
	"F34"	0.083	0	0.139	0	0.167	0

Table 4. Bivariate DVECH-GARCH (1,1) parameters for R-R model at "F7" station (p-values of t-statistic are given in parenthesis)

Rainfall station	DVECH Parameters								
	ω_{11}	ω_{21}	ω_{22}	γ_{11}	γ_{21}	γ_{22}	g_{11}	g_{21}	g_{22}
"S"	0.421	0.031	0.105	0.048	0.055	0.708	0.948	0.862	0.216
	(0.001)	(0.03)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
"P"	0.811	0.025	0.105	0.051	0.064	0.714	0.936	0.826	0.212
	(0.001)	(0.014)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
"F"	0.98	0.038	0.107	0.054	0.062	0.716	0.928	0.802	0.206
	(0.001)	(0.075)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
"SCT"	0.51	0.068	0.101	0.048	0.058	0.72	0.947	0.806	0.216
	(0.001)	(0.011)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Table 5 Bivariate CCC-GARCH (1,1) parameters for R-R model at "F7" (p-values of t-statistic are given in parenthesis)

series	GARCH parameters			Constant Conditional Correlation	Unconditional Correlation
	α_0	α	β		
"S"	0.410 (0.001)	0.047 (0.001)	0.950 (0.001)	0.095	0.036
"P"	0.783 (0.001)	0.049 (0.001)	0.938 (0.001)	0.078	0.042
"F"	0.833 (0.001)	0.115 (0.001)	0.820 (0.001)	0.061	0.027
"SCT"	0.526 (0.001)	0.046 (0.001)	0.947 (0.001)	0.123	0.067
"F7"	0.110 (0.001)	0.732 (0.001)	0.188 (0.001)	-	-

Table 6 Estimation of NBIAS and NRMSE for empirical against simulated conditional covariances for "F7"

Rainfall series	Model							
	DVECH				CCC			
	(1,1)		(2,2)		(1,1)		(2,2)	
	NRMSE	NBIAS	NRMSE	NBIAS	NRMSE	NBIAS	NRMSE	NBIAS
"S"	13.12	-0.19	22.1	-0.37	0.52	-0.06	0.52	-0.08
"P"	11.8	0.11	25.8	0.63	0.46	0.02	0.47	0.04
"F"	6.32	0.33	42.12	-2.19	0.51	-0.03	0.53	-0.05
"SCT"	9.08	0.15	24.6	-0.23	0.47	-0.007	0.48	0.02

Table 7 Same as Table 4 but for "F34" station

Rainfall station	DVECH Parameters								
	ω_{11}	ω_{21}	ω_{22}	γ_{11}	γ_{21}	γ_{22}	g_{11}	g_{21}	g_{22}
"S"	0.421	0.497	0.127	0.049	0.048	0.849	0.949	0.931	0.144
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
"P"	0.815	0.418	0.137	0.049	0.043	0.861	0.938	0.943	0.113
	(0.001)	(0.001)	(0.001)	(0.001)	(0.008)	(0.001)	(0.001)	(0.001)	(0.001)
"F"	0.795	0.031	0.147	0.051	0.027	0.862	0.937	0.939	0.087
	(0.001)	(0.009)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
"SCT"	0.517	0.061	0.154	0.047	0.022	0.887	0.948	0.918	0.065
	(0.001)	(0.034)	(0.001)	(0.001)	(0.025)	(0.001)	(0.001)	(0.001)	(0.012)

Table 8 Same as Table 5 but for "F34" station

series	GARCH parameters			Constant Conditional Correlation	Unconditional Correlation
	α_0	α	β		
"S"	0.409 (0.001)	0.047 (0.001)	0.950 (0.001)	0.138	0.0.74
"P"	0.782 (0.001)	0.048 (0.001)	0.939 (0.001)	0.118	0.070
"F"	0.859 (0.001)	0.053 (0.001)	0.927 (0.001)	0.101	0.050
"SCT"	0.524 (0.001)	0.045 (0.001)	0.948 (0.001)	0.182	0.107
"F34"	0.168 (0.001)	0.902 (0.001)	0.023 0.33	-	-

Table 9 Same as Table 6 but for "F34"

Rainfall series	Model							
	DVECH				CCC			
	(1,1)		(2,2)		(1,1)		(2,2)	
	NRMSE	NBIAS	NRMSE	NBIAS	NRMSE	NBIAS	NRMSE	NBIAS
"S"	0.48	0.47	0.53	0.52	0.47	0.47	0.54	0.53
"P"	0.44	0.44	0.50	0.49	0.44	0.44	0.51	0.50
"F"	0.45	0.44	0.52	0.50	0.44	0.44	0.51	0.51
"SCT"	0.52	0.51	0.57	0.55	0.51	0.51	0.56	0.55

Table 10 Stationarityy and nonlinearity test results for conditional covariances

Rainfall series	Streamflow series	KPSS level stationary test		KPSS trend stationary test		ADF unit root test	
		Results	p-value	Results	p-value	Results	p-value
"P"	"F7"	0.17	>0.05	-0.75	0.43	-11.29	0
	"F34"	0.20	>0.01	-0.16	0.85	-29.93	0
"S"	"F7"	0.15	>0.05	-0.54	0.58	-9.58	0
	"F34"	0.18	>0.01	-0.16	0.87	-27.14	0
"F"	"F7"	0.11	>0.05	-1.49	0.13	-12.10	0
	"F34"	0.20	>0.01	-0.45	0.65	-5.71	0
"SCT"	"F7"	0.19	>0.01	-1.03	0.29	-12.95	0
	"F34"	0.24	>0.01	-0.31	0.75	-7.58	0
BDS test							
Rainfall series	Streamflow series	m=2		m=3		m=4	
		statistic	p-value	statistic	p-value	statistic	p-value
"P"	"F7"	0.15	0	0.25	0	0.31	0
	"F34"	0.08	0	0.13	0	0.17	0
"S"	"F7"	0.15	0	0.26	0	0.33	0
	"F34"	0.08	0	0.14	0	0.17	0
"F"	"F7"	0.14	0	0.24	0	0.30	0
	"F34"	0.17	0	0.27	0	0.37	0
"SCT"	"F7"	0.14	0	0.24	0	0.30	0
	"F34"	0.17	0	0.28	0	0.36	0

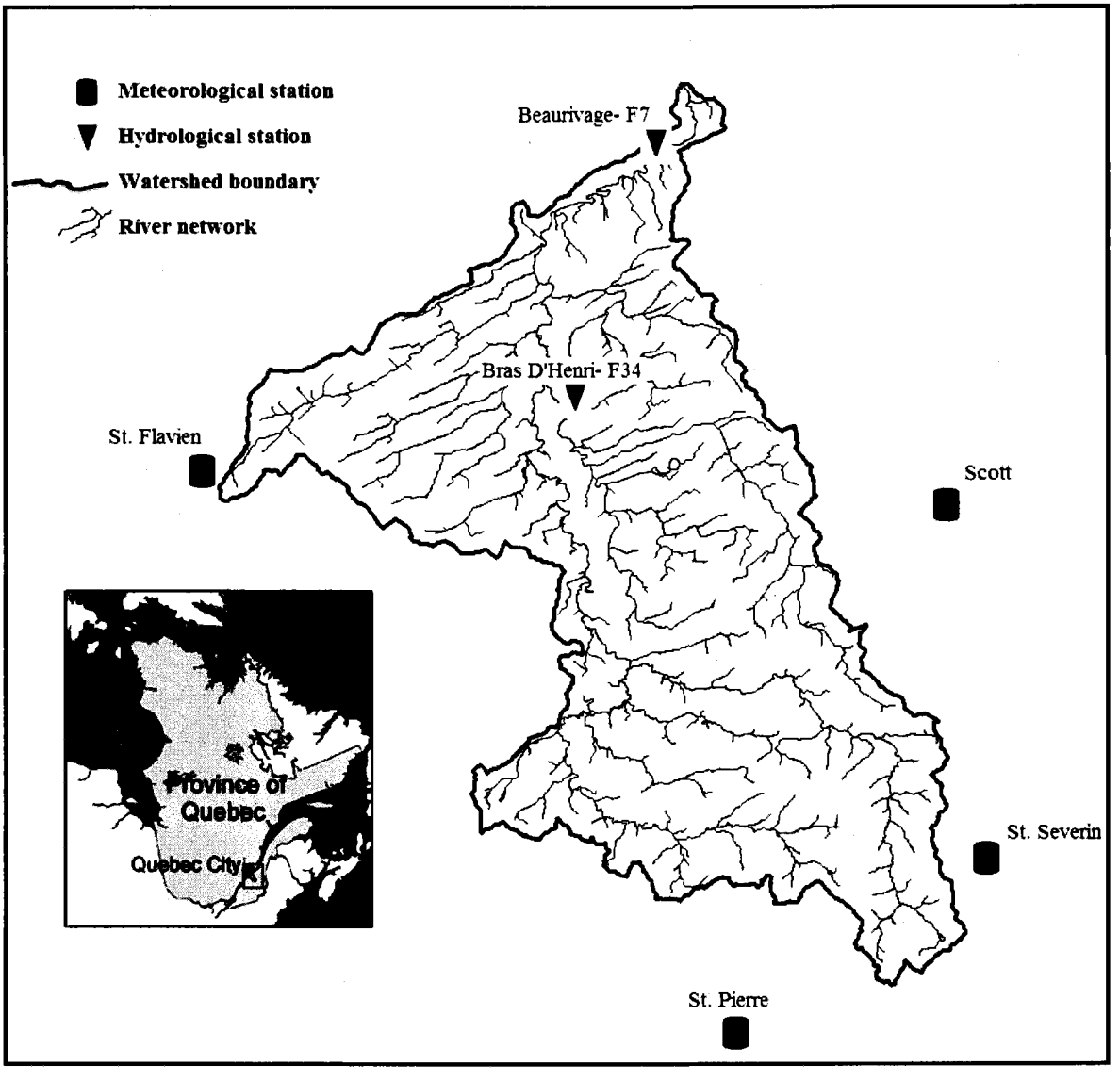

Meteorological station
Hydrological station
Watershed boundary
River network
Beaurivage- F7
Bras D'Henri- F34
St. Flavien
Scott
St. Severin
St. Pierre
Province of Quebec
Quebec City

Figure 1

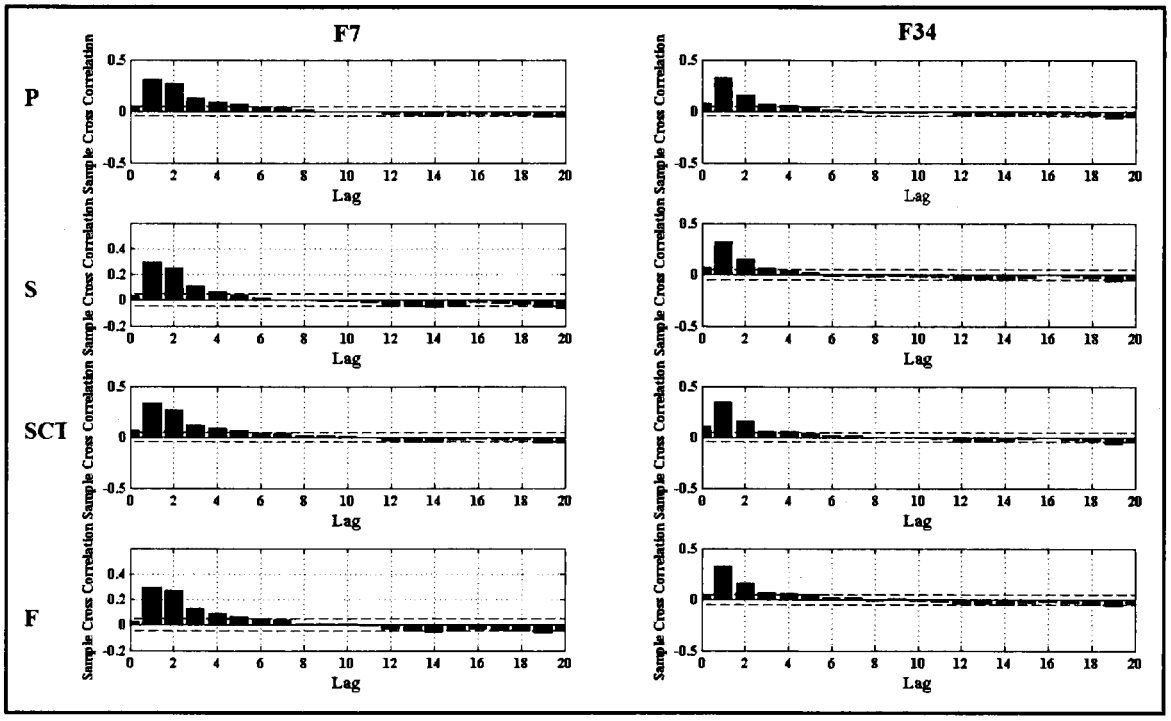

F7
F34
P
S
SCI
F
Sample Cross Correlation Sample Cross Correlation Sample Cross Correlation
Lag

Figure 2

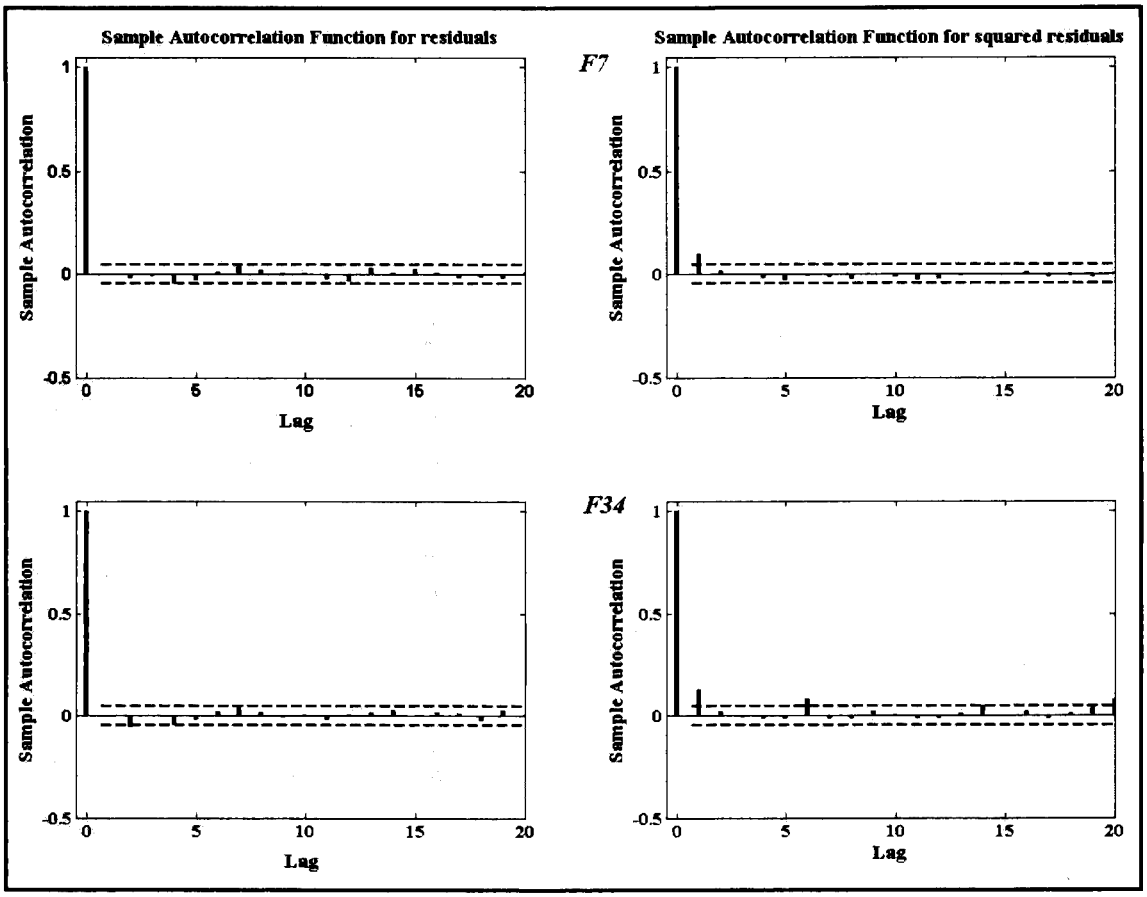

Sample Autocorrelation Function for residuals
Sample Autocorrelation Function for squared residuals
F7
F34
Sample Autocorrelation
Lag
1
0.5
0
-0.5
0
5
10
15
20

Figure 3

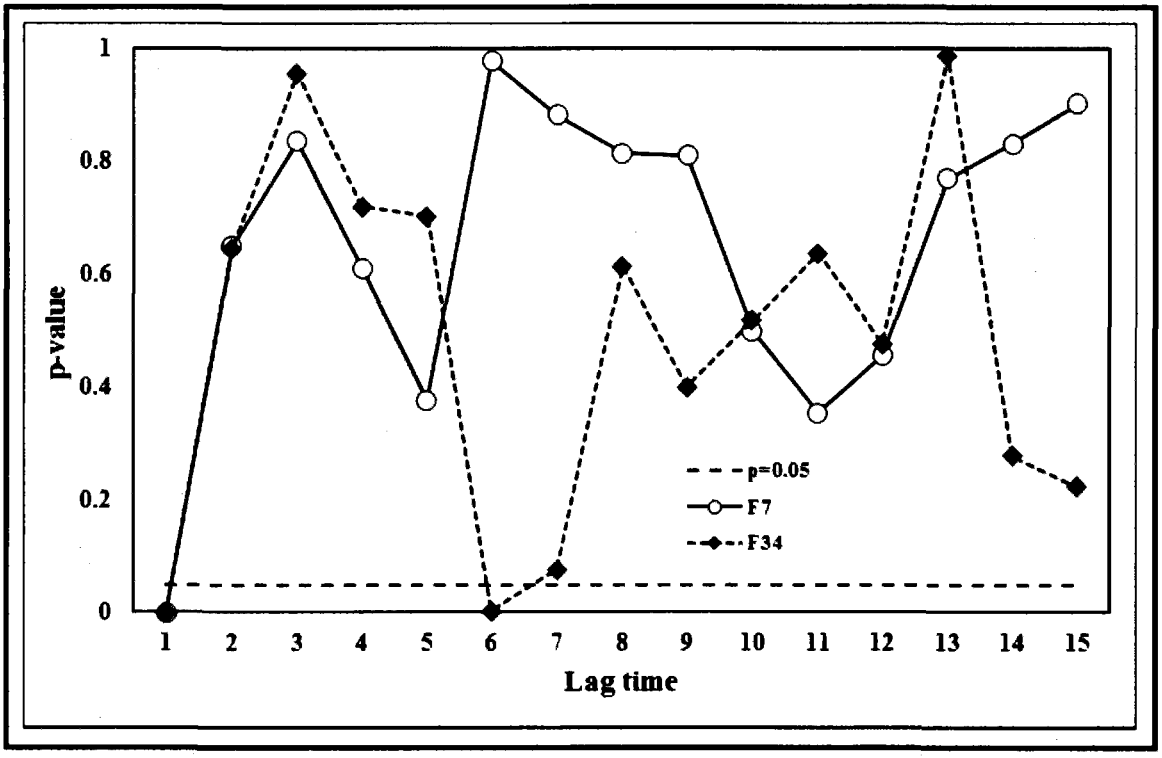

1
0.8
0.6
0.4
0.2
0
p-value
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Lag time
p=0.05
F7
F34

Figure 4

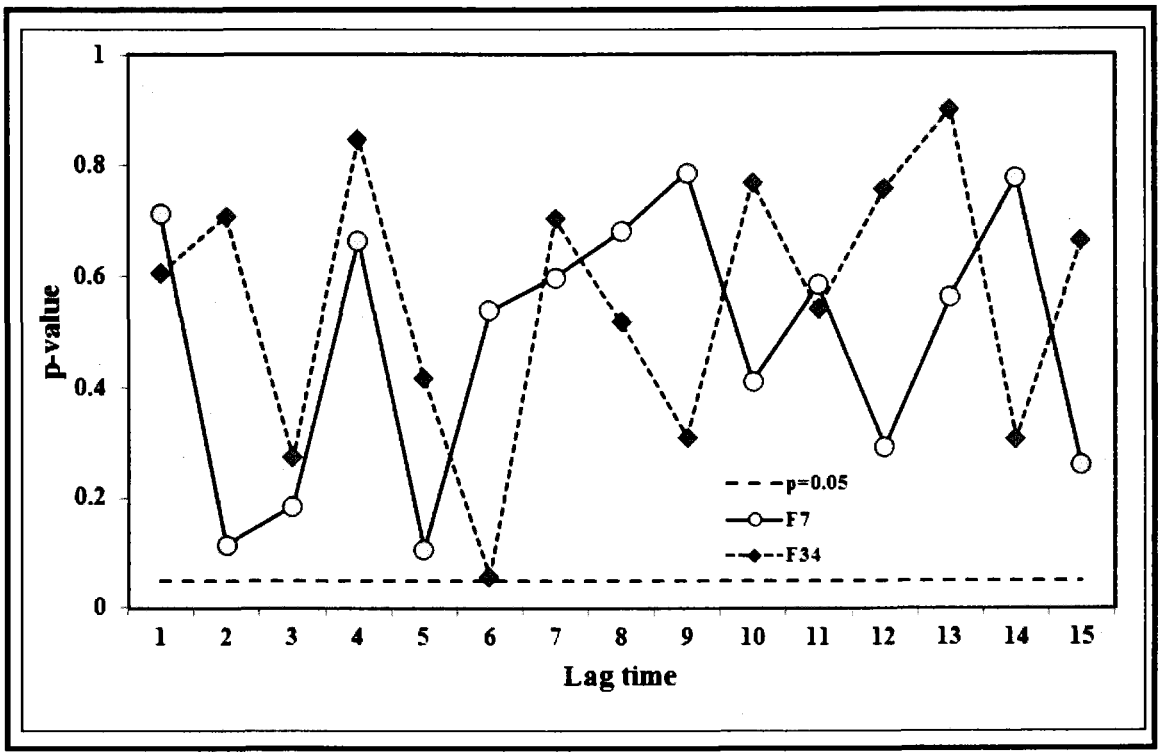

1
0.8
0.6
0.4
0.2
0
p-value
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Lag time
p=0.05
F7
F34

Figure 5

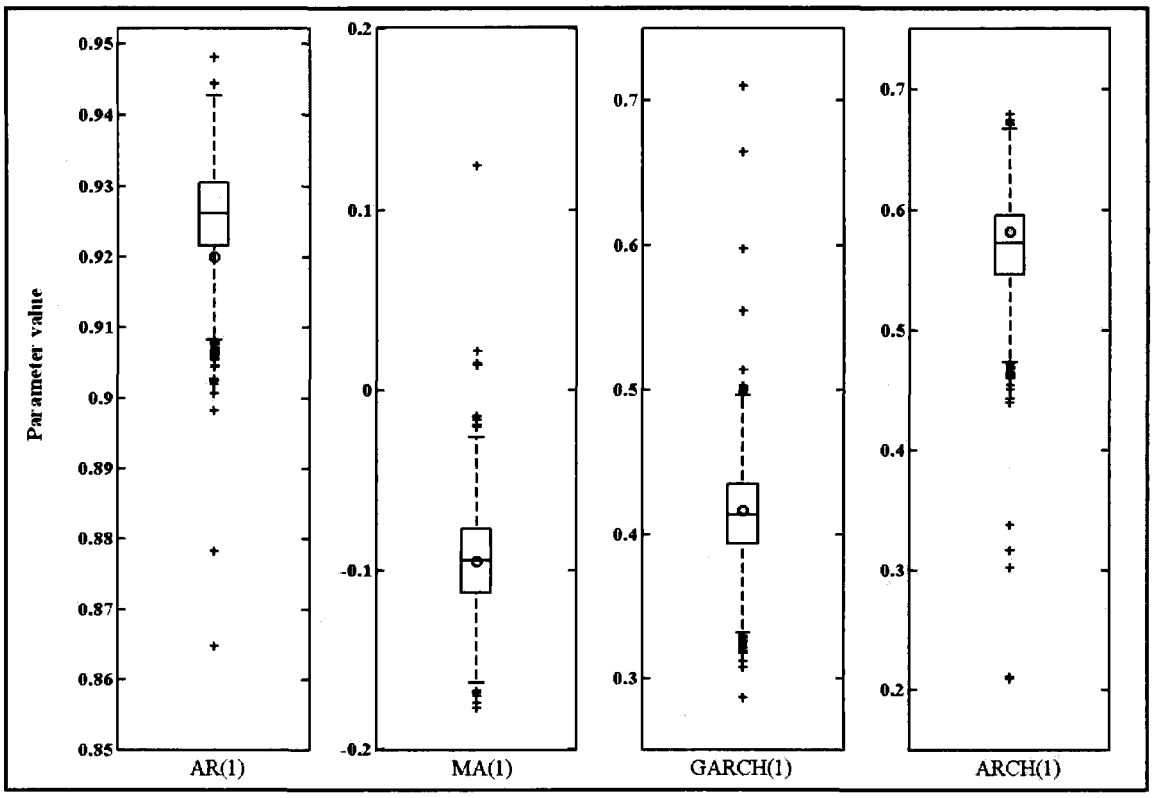

Parameter value
0.95
0.94
0.93
0.92
0.91
0.9
0.89
0.88
0.87
0.86
0.85
AR(1)
0.2
0.1
0
-0.1
-0.2
MA(1)
0.7
0.6
0.5
0.4
0.3
GARCH(1)
0.7
0.6
0.5
0.4
0.3
0.2
ARCH(1)

Figure 6

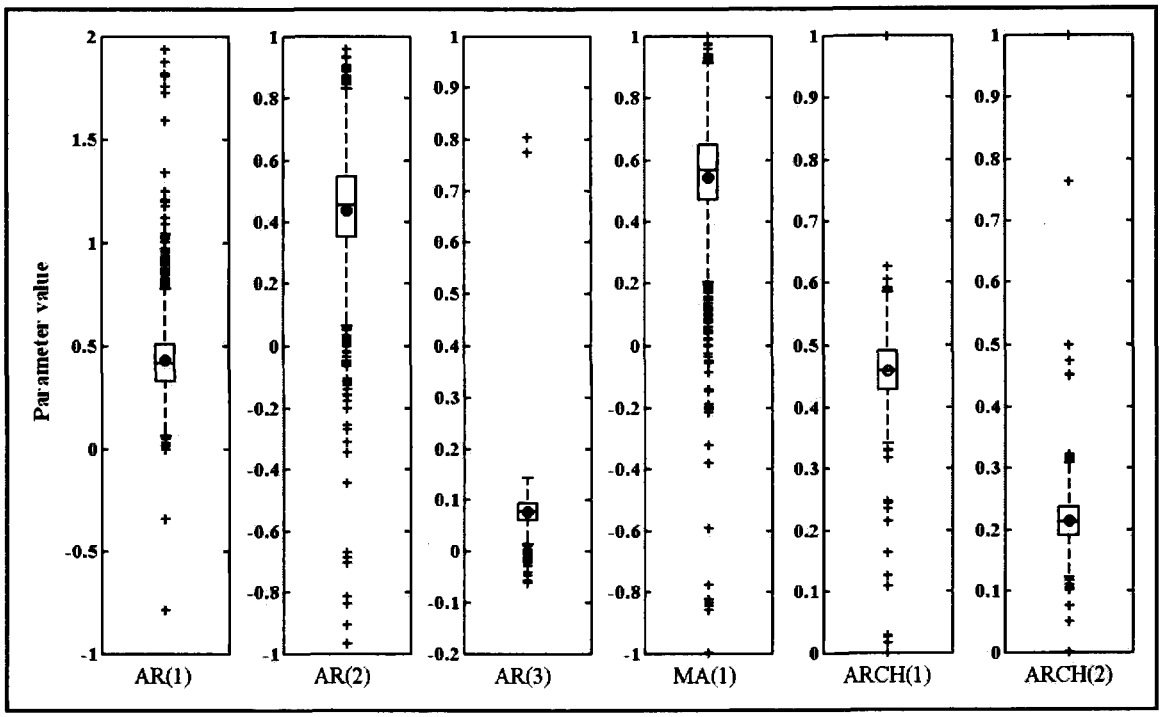


Figure 7

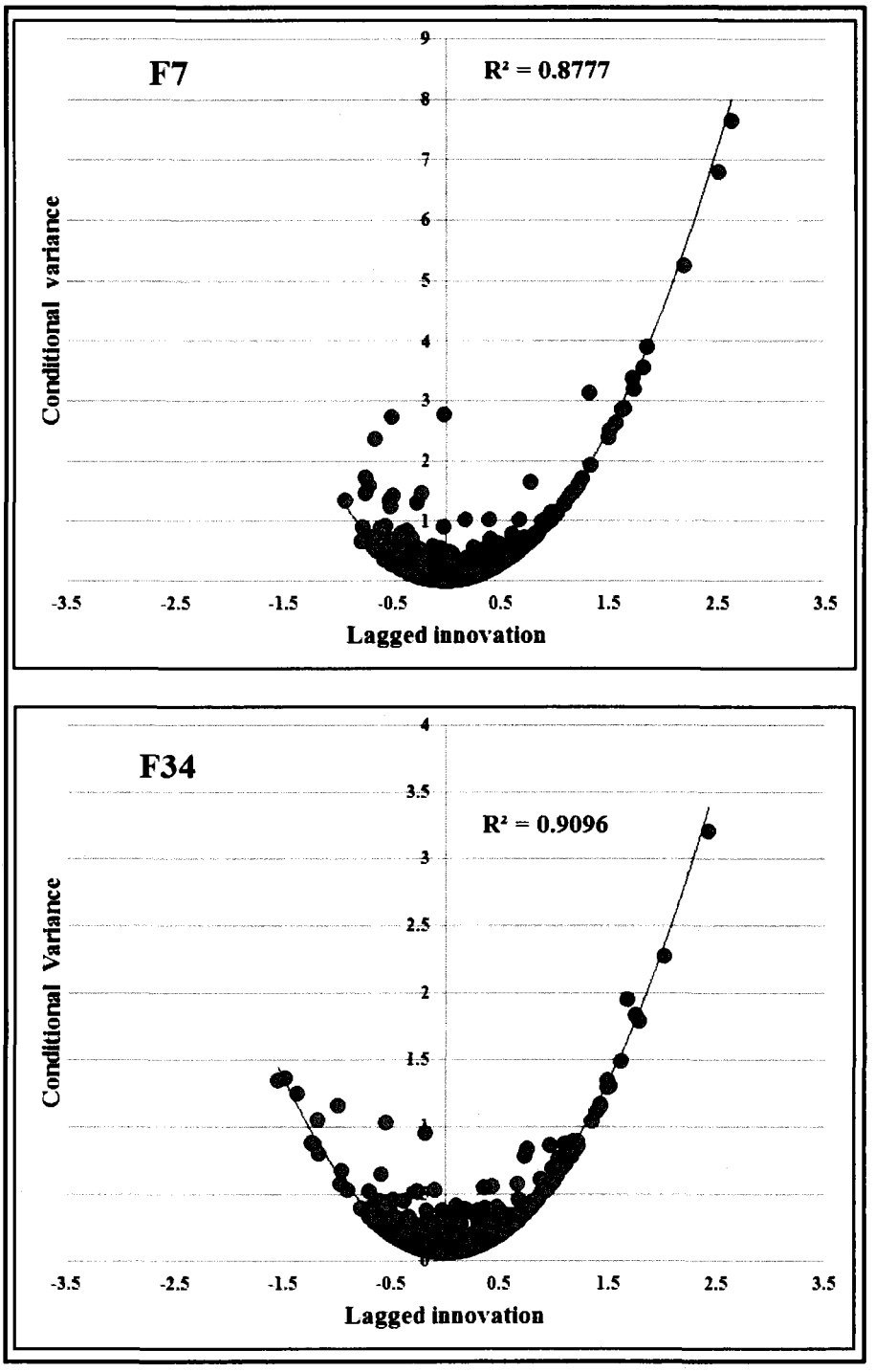

F7
R² = 0.8777
Conditional variance
Lagged innovation
F34
R² = 0.9096
Conditional Variance
Lagged innovation

Figure 8

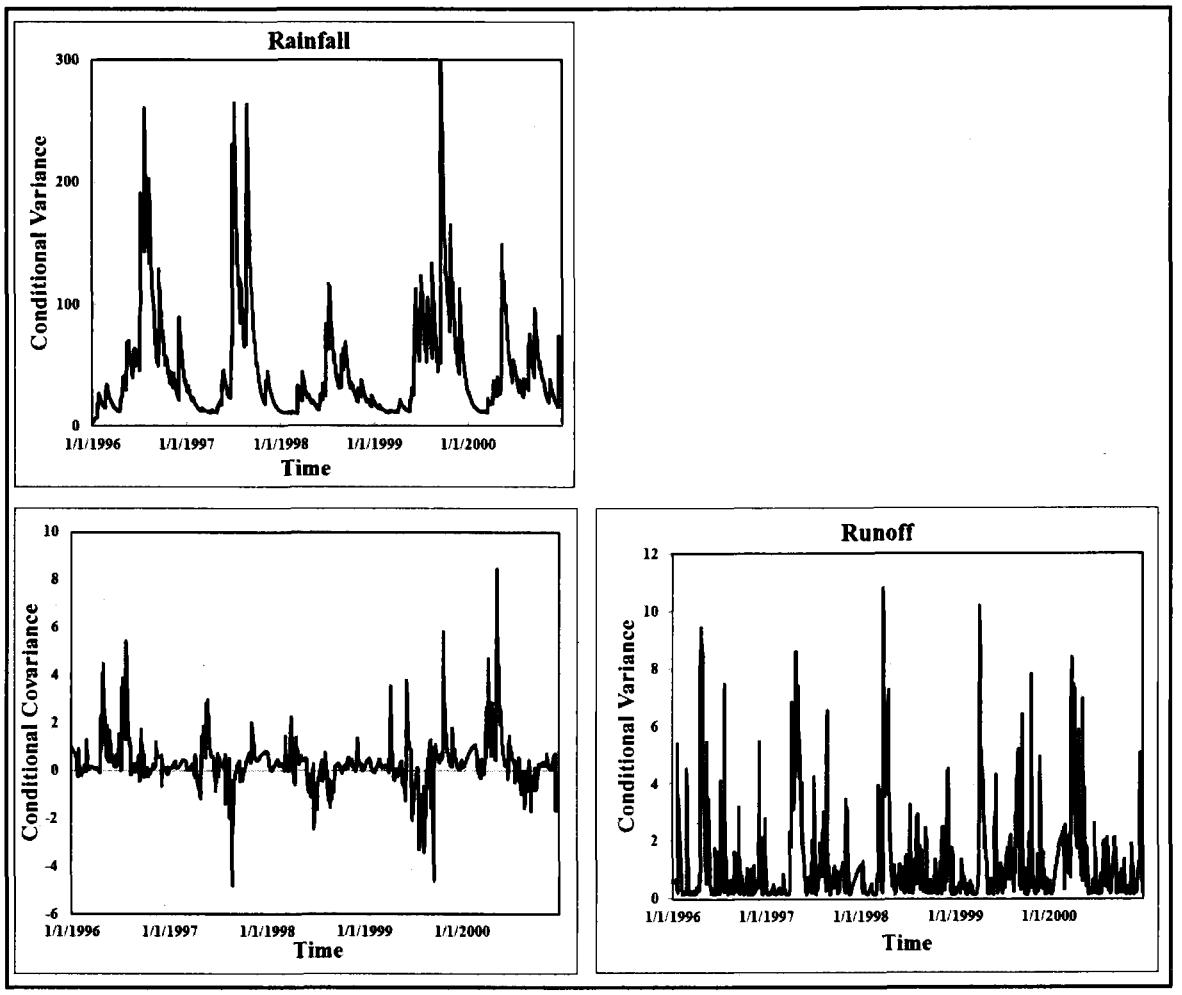

Rainfall
Conditional Variance
300
200
100
0
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000
Time
Conditional Covariance
10
8
6
4
2
0
-2
-4
-6
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000
Time
Runoff
Conditional Variance
12
10
8
6
4
2
0
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000
Time

Figure 9

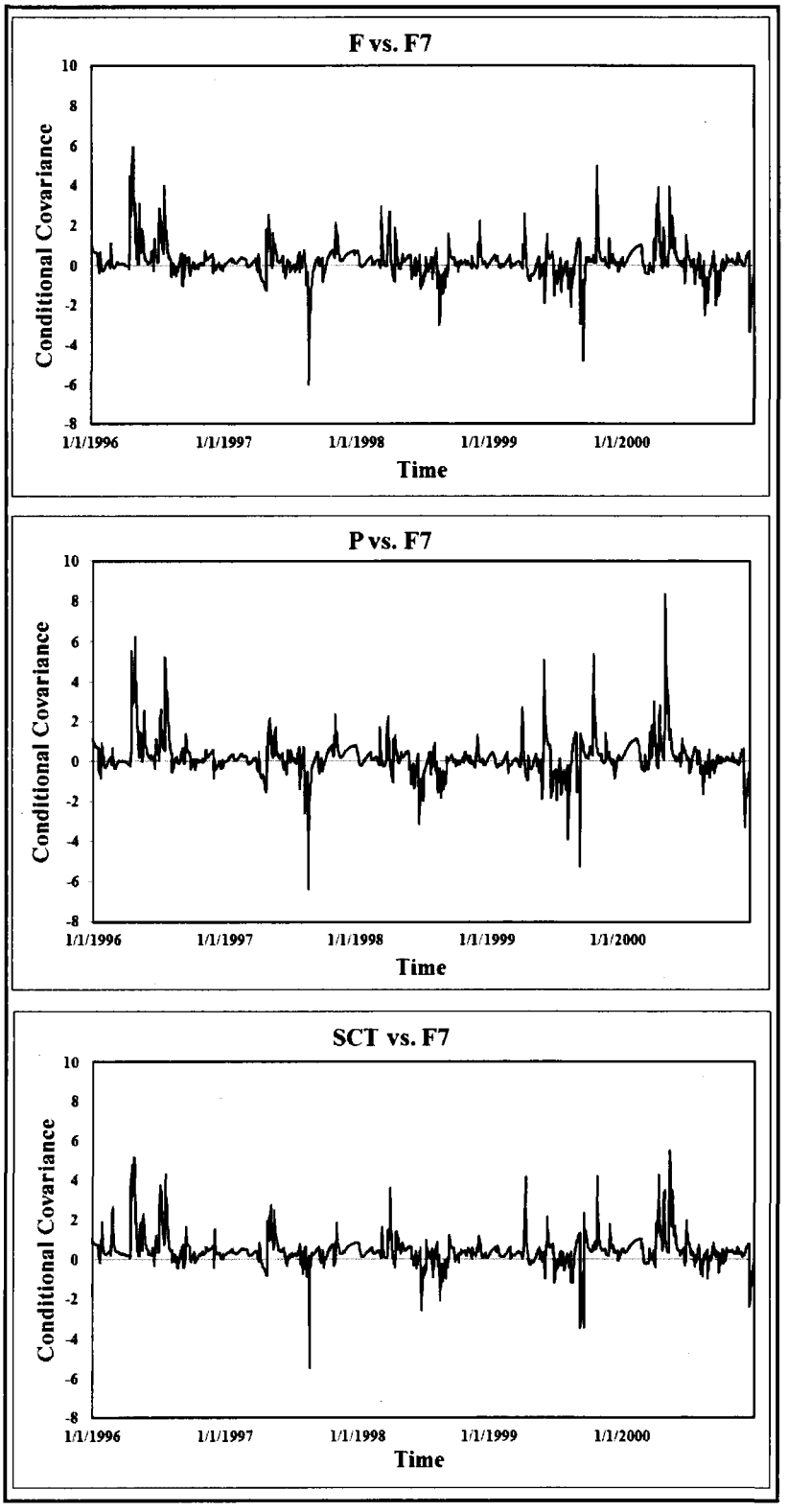

F vs. F7
Conditional Covariance
10
8
6
4
2
0
-2
-4
-6
-8
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000
Time
P vs. F7
Conditional Covariance
10
8
6
4
2
0
-2
-4
-6
-8
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000
Time
SCT vs. F7
Conditional Covariance
10
8
6
4
2
0
-2
-4
-6
-8
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000
Time

Figure 10

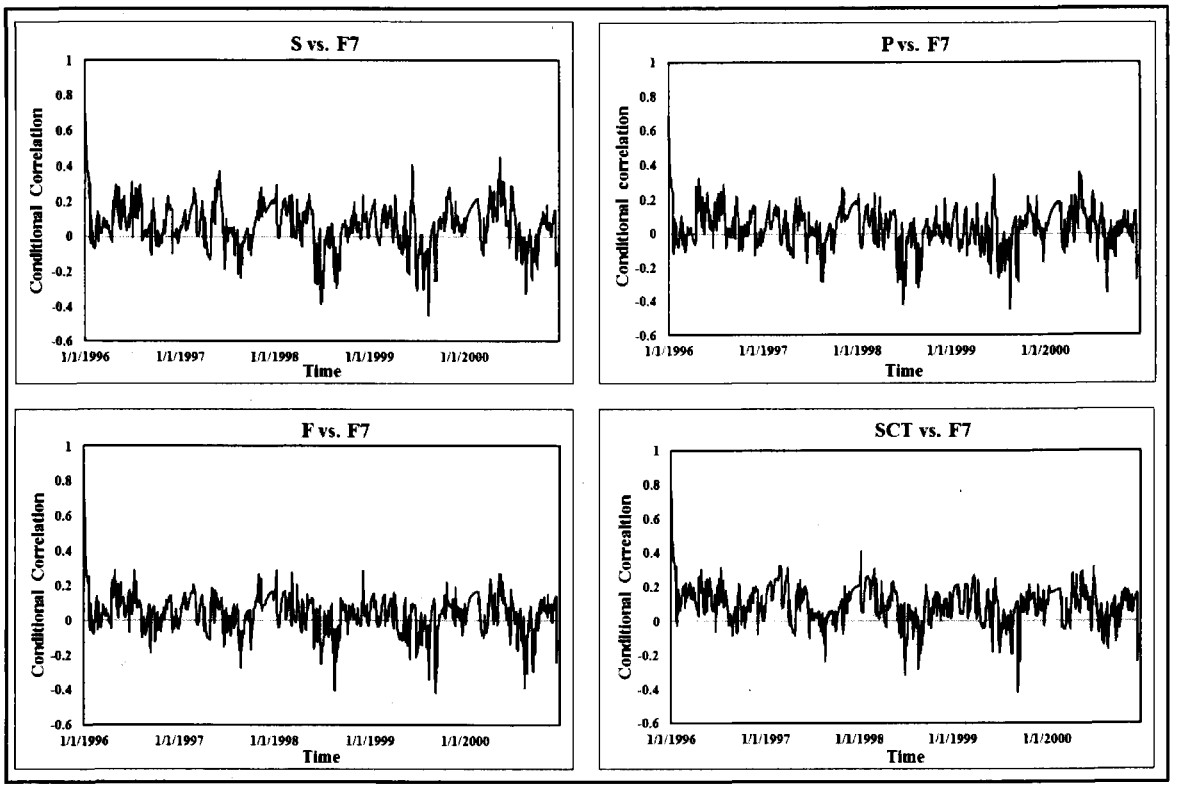

S vs. F7
Conditional Correlation
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000
Time
P vs. F7
Conditional correlation
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000
Time
F vs. F7
Conditional Correlation
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000
Time
SCT vs. F7
Conditional Correaltion
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000
Time

Figure 11

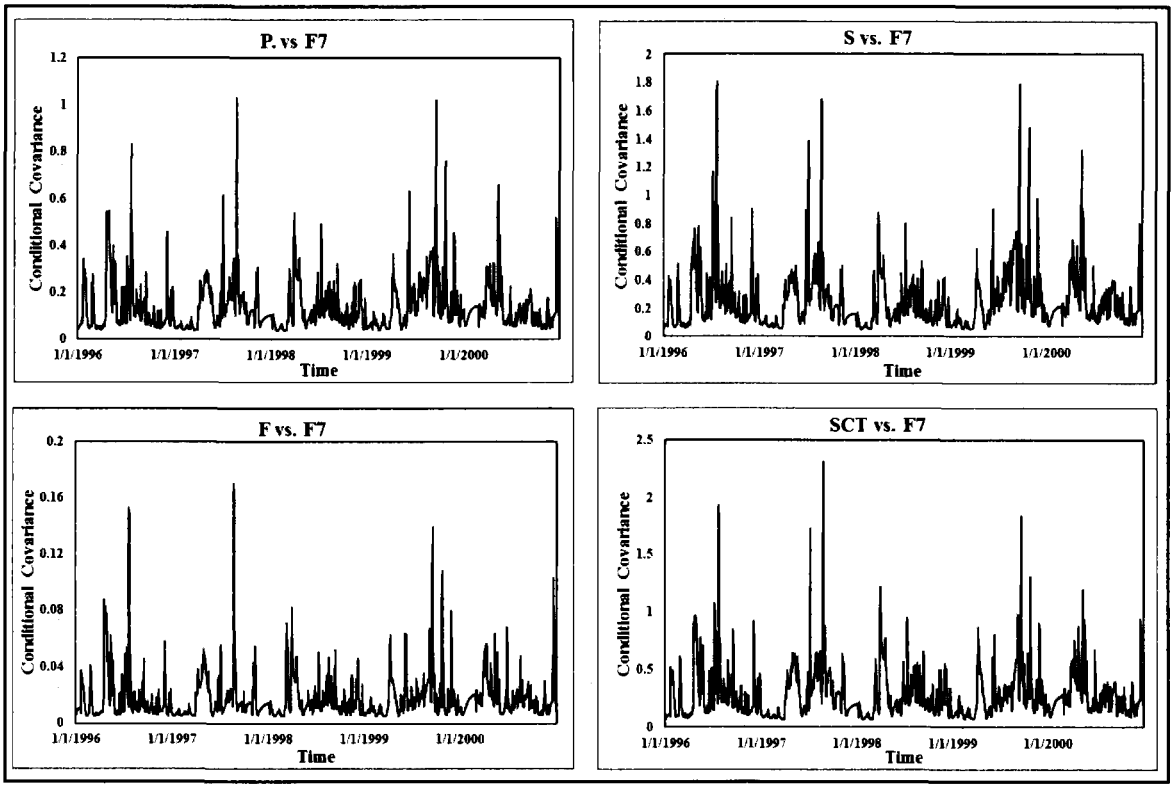

P. vs F7
Conditional Covariance
1.2
1
0.8
0.6
0.4
0.2
0
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000
Time
S vs. F7
Conditional Covariance
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000
Time
F vs. F7
Conditional Covariance
0.2
0.16
0.12
0.08
0.04
0
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000
Time
SCT vs. F7
Conditional Covariance
2.5
2
1.5
1
0.5
0
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000
Time

Figure 12

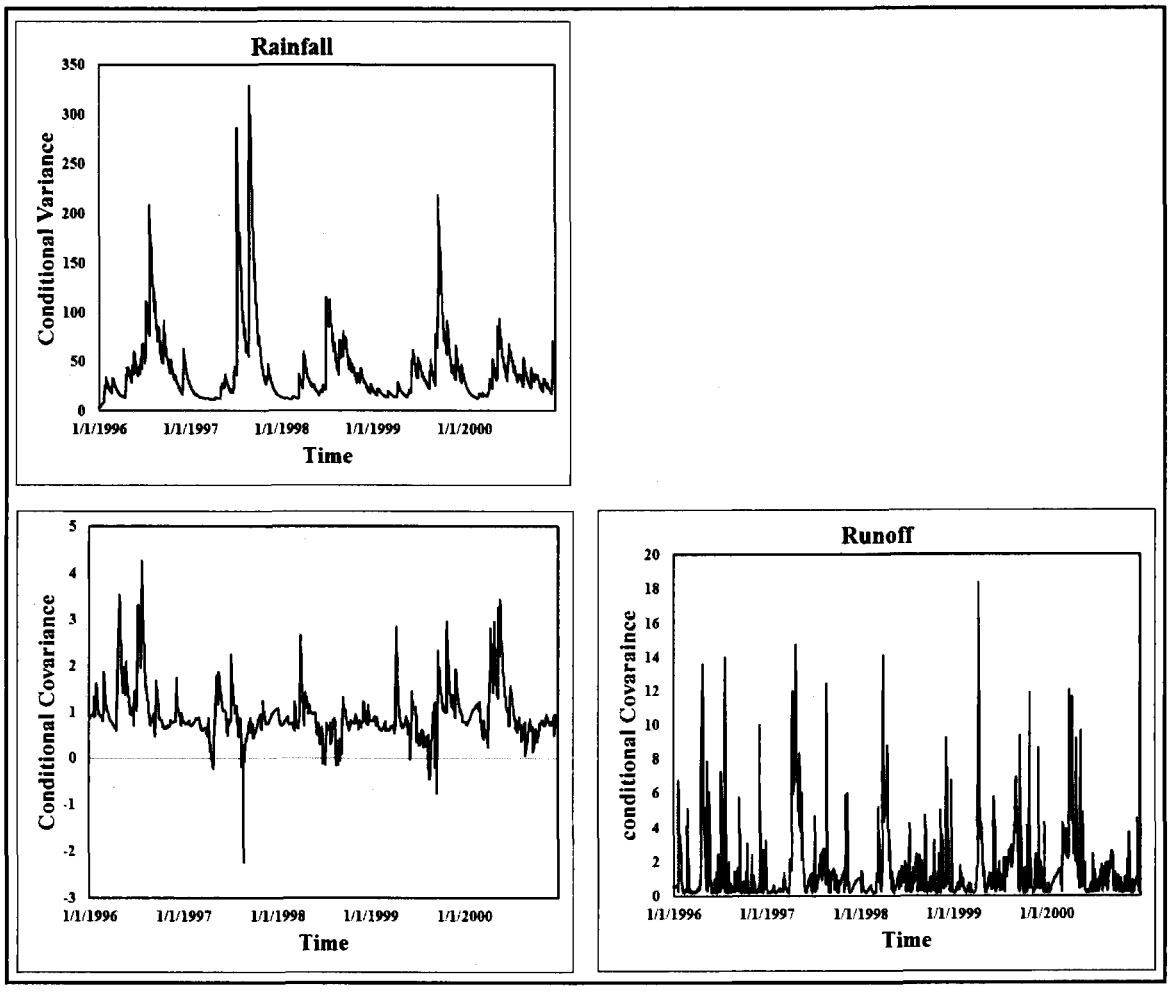

Rainfall
Conditional Variance
350
300
250
200
150
100
50
0
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000
Time
Conditional Covariance
5
4
3
2
1
0
-1
-2
-3
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000
Time
Runoff
conditional Covaraince
20
18
16
14
12
10
8
6
4
2
0
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000
Time

Figure 13

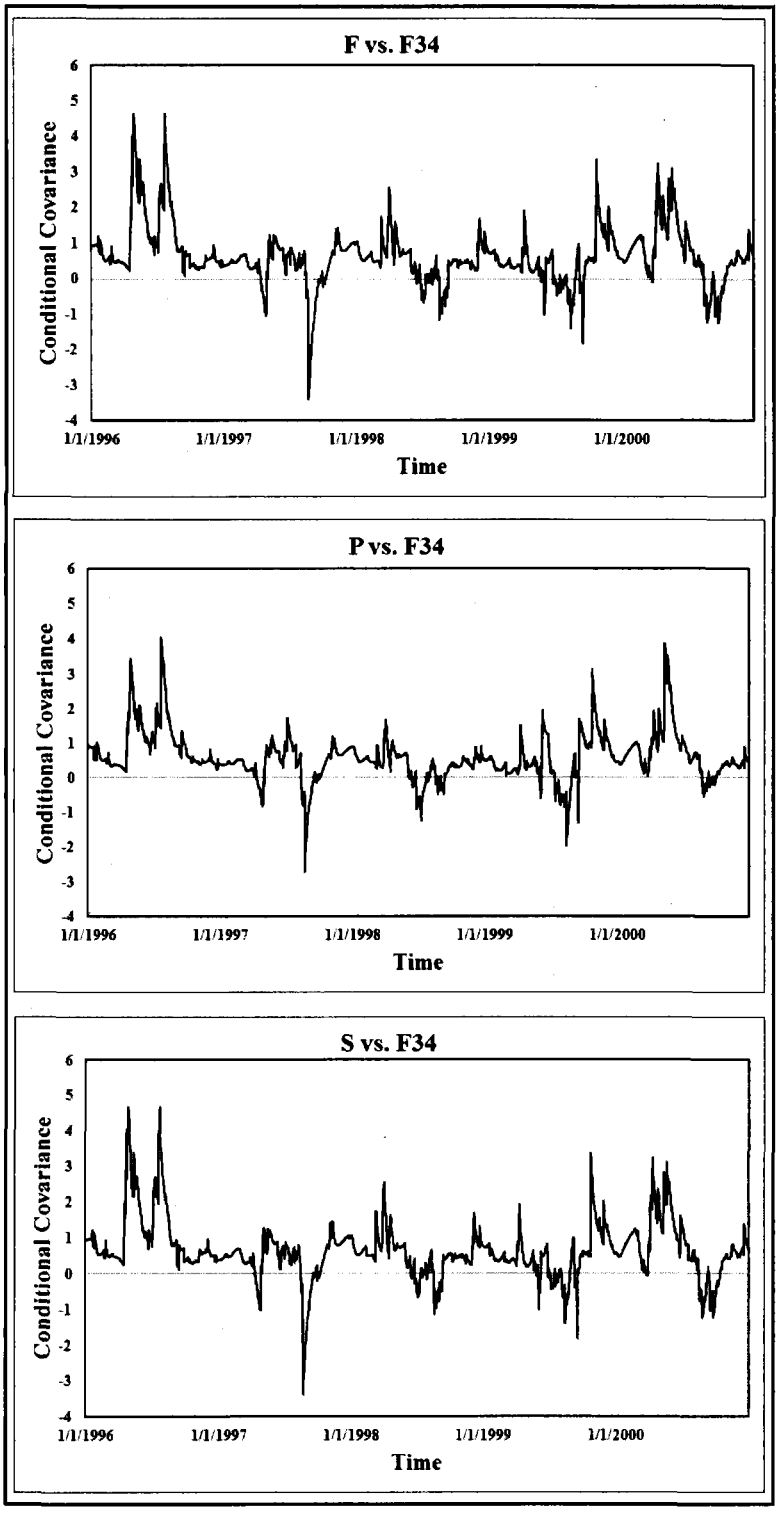

F vs. F34
Conditional Covariance
6
5
4
3
2
1
0
-1
-2
-3
-4
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000
Time
P vs. F34
Conditional Covariance
6
5
4
3
2
1
0
-1
-2
-3
-4
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000
Time
S vs. F34
Conditional Covariance
6
5
4
3
2
1
0
-1
-2
-3
-4
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000
Time

Figure 14

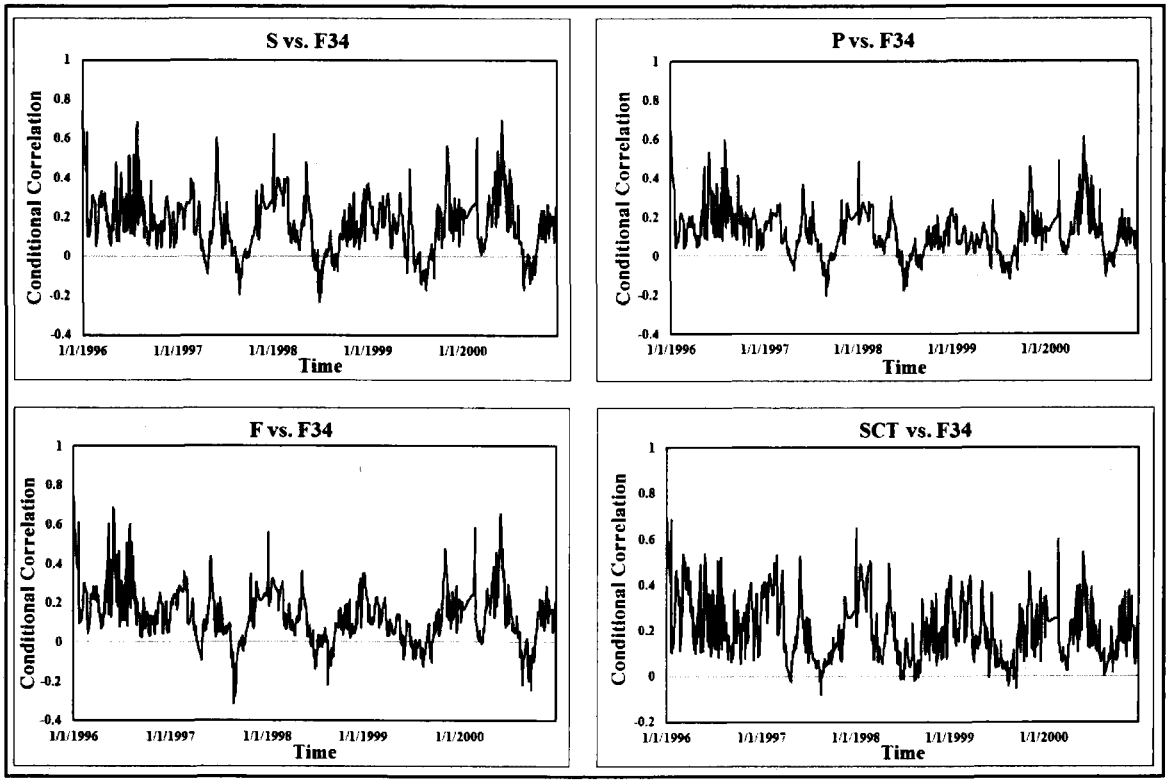

S vs. F34
P vs. F34
F vs. F34
SCT vs. F34
Conditional Correlation
Time
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000

Figure 15

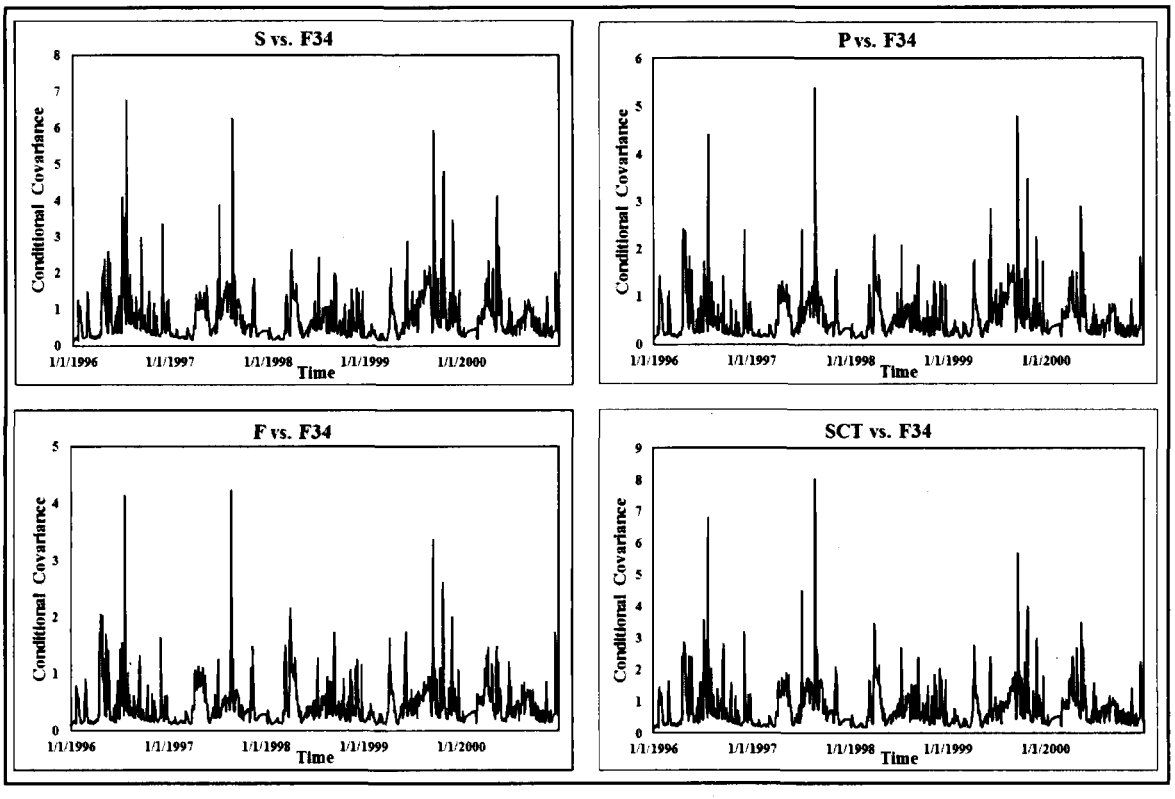

S vs. F34
P vs. F34
F vs. F34
SCT vs. F34
Conditional Covariance
Time
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000

Figure 16

Article 5. Modeling climate effects on hip fracture rate by multivariate GARCH model in Montreal region, Canada

Modeling climate effects on hip fracture rate by multivariate GARCH model in Montreal region, Canada

Reza Modarres[a*], Taha B. M. J. Ouarda[a,b], Alain Vanasse[c], Maria Gabriela Orzanco[c], Pierre Gosselin[d]

[a]HYdrocliamte modeling group anada Research Chair on the Estimation of Hydrometeorological Variables, INRS-ETE, 490 de la Couronne, Quebec, Qc, Canada, G1K 9A9

Tel: +1 418 654-3842, Fax: +1 418 654-2600,

E-mail: Reza.Modarres@ete.inrs.ca

[b] Masdar Institute of Science and Technology, P.O. Box 54224, Abu Dhabi, UAE

E-mail: touarda@masdar.ac.ae

[c]Groupe de recherche PRIMUS, Université de Sherbrooke 3001, 12e avenue Nord, Sherbrooke, QC, Canada, J1H 5N4

Tel : +1 418 346-1110, Fax : +1 819 564-5424,

E-mail : Alain.Vanasse@USherbrooke.ca, Maria.Gabriela.Orzanco@USherbrooke.ca

[d]Changements climatiques, Unité Santé et environnement, Institut National de santé Publique du Québec (INSPQ), 945 rue Wolf, Québec, Qc, Canada, G1V 5B3

Tel : +1 418 650-5115, Fax : +1 418 654-3144,

E-mail : pierre.gosselin@inspq.qc.ca

* Corresponding author

Submitted for publication to Intenational Journal of biometeorology

Abstract

The change in extreme meteorological variables and demographic shift towards older population has made it important to investigate the association of climate variables and hip fracture by advanced methods in order to find the most effective climate variables on hip fracture incidence. The nonlinear Autoregressive Moving Average with eXogenous variable-Generalized Autoregressive Conditional Heteroscedasticity (ARMA*X*-GARCH) and multivariate GARCH (MGARCH) time series approaches are applied to investigate the nonlinear association between hip fracture rate in female and male patients of the ages 40-74 and 75+ and climate variables in the period of 1993-2004, in Montreal, Canada. The models describe 50% to 56% of daily variation of hip fracture rate and identify the snow depth, air temperature, day length and air pressure as the influencing variables on the time varying mean and variance of the hip fracture rate. The conditional covariance between climate variables and hip fracture rate is exponentially increasing which shows that the effect of climate variables on hip fracture rate is most acute when rates are high and climate conditions are at their worst. In Montreal, Climate variables, particularly snow depth and air temperature, appear to be important predictors of the hip fracture incidence. The association of climate variables and hip fracture does not seem to linearly change in time and increases exponentially with the harsh climate conditions. The results of this study can be used to provide an adaptive climate-related public health program and allocation services for avoiding hip fracture risk.

Keywords: hip fracture, ARMA*X*-GARCH, multivariate GARCH, conditional covariance, climate

Introduction

Hip fracture is already a cause of mortality and morbidity and a growing problem in the western countries along with overall aging of societies. Hip fracture constitutes a significant economic burden in developed countries and reduces the quality of life and makes 20% reduction in expected survival (Benetos et al, 2007). It was also reported that mortality rates after hip fracture continue to rise over the subsequent months and peak at 1 year, with a rate of 36% for men and 21% for women (Harvey and Dennison, 2010). The importance of preventing at least some of the clinical and economic consequences of hip fracture has led researchers to try to identify the factors affecting hip fracture incidence and risk.

Various factors affect hip fracture risk and incidence including genetics, demography (increasing age and gender), chronic diseases such as osteoporosis and diabetes, medication (hormone replacement therapy and drug altering the central nervous system), medical and gynaecological history (past fractures), anthropometric variables(increased or decreased body weight), life style factors such as physical inactivity or activity and exposure to sunlight as well as nutrition factors (calcium and vitamin D intake, Benetos et al, 2007). Nursing home design and care procedures are also important (Quigley et al. 2013) as well as urban design (Cummings and Melton, 2002). Lan et al. (2010) suggested multiple risk factors for hip fracture incidence such as low consumption of milk, grip strength and bone -mineral density. However, these are more clinical –level risk factors and seem to ignore climate factors, with the exception of the effect of sunlight (Bulajic-Kopjar, 2010; Edvardsen et al. 2009) and its indirect effect on physical exercise.

Many studies have indicated the association between climate and hip fracture from which some of them are presented here. For example, the effect of cold climate on the increasing rate of hip fracture in Oslo, Norway, has been reported (Lofthus et al, 2001) as well as the effect of freezing

rain in Montreal, Canada (Levy et al., 1998). A recent study in Valencia (Spain) showed a significant impact of meteorological variables on hip fracture, especially in cold seasons (Tenias et al., 2009). The climate and its seasonal variation influence the risk of hip fracture incidence for Scandinavian countries (Gullberg et al., 1993; Gronskag et al., 2010). The effect of day length, sunshine duration and temperature on daily activity of elderly people and high hip fracture risk was indicated by Mirchandari et al. (2005) and Sumukadas et al. (2009). After adjustment for season, trend, day-of-week and autocorrelation, Turner et al. (2011) indicated the effect of low temperature on fall-related hip fracture hospitalization in New South Wales, Australia, using a Poisson regression model.

Time series methods are a particularly promising statistical approach for modeling the variation of different variables and their influence on each other through time. However, these (linear) models are rarely applied to address epidemiologic time-varying variables such as hip fracture and very few studies can be found in literature (Lin and Xiraxagar, 2006). The linear assumptions of these models are sometimes not enough to model the nonlinearity embedding the hip fracture association with climate variables.

The main objective of this study is to introduce a nonlinear financial time series modeling approach, Generalized AutoRegressive Conditional Heteroscedasticity (GARCH), for modeling the nonlinear relationship between hip fracture and climate variables. The hypothesis here is that the change in the hip fracture association to climate variation is partially nonlinear and could be modeled by a nonlinear approach. Moreover, this study aims to model and predict daily variation of hip fracture and its association to weather conditions.

Methods

Data collection

The data set used in this study includes the daily hip Fracture incidences over the Montreal region, Quebéc province, Canada which are population-standardized for 100000 person-day from 1993-2004. The data set includes hospital discharge data of the patients who are from the Montreal region and are aged of ≥40 years (total population ≥40 in mid-period census of 2001 was about 900 000) and the main diagnosis of the admission is a hip fracture (ICD-9 codes 820); and excludes patients with injury causes other than "accidental fall" (ICD-9 codes E880 to E888) or "accidents due to natural and environmental factors" (ICD-9 codes E900 to E909). Daily hospitalisations events extracted from Québec's hospital database (MED-ÉCHO) were supplied by the *Institut national de la santé publique du Québec* (INSPQ) and the *Ministère de la Santé et des Services Sociaux du Quebec* (MSSS) (Lambert et al. 2010). The reliability of the assessment of patient morbidities using Quebec's medico-administrative data has been rated as very high for various diagnoses (levy et al. 1999).

These hip fracture rates are calculated for two age groups, 40-74 years and 75 years and older, of the female and male patients separately. This decision was made in order to maintain sufficient numbers of fractures in gender groups allowing for a meaningful analysis. These female and male groups are called F1, F2, M1 and M2 hereafter in this study. In order for modeling the relationship between climate variables and hip fracture rate in this study, 12 climate variables from 9 stations within the Montreal region in daily time scale were gathered first. Then, the daily time series are averaged over the region to obtain one single time series for each climate variable. Figure 1 shows the geographic location of Montreal region and the location of the meteorological stations.

Methods

Univariate ARMA*X*-GARCH model

The univariate nonlinear time series modeling approach includes modeling the conditional mean by the autoregressive moving average model (ARMA*X*(p,q,s) where p and q are the order of the autoregressive and moving average parameters and the eXogeneos variable, respectively) and modeling the conditional variance of the residuals (errors) by a GARCH(v,m) model where v and m are the order of the parameters of GARCH model. This type of model is called ARMA*X*-GARCH error model (Hamilton, 1994).

The univariate linear ARMA*X* time series model has the following general form (Hipel and Mc Leod (1996)

$$\phi_p(B)Y_t = \omega_s(B)x_t + \theta_q(B)\varepsilon_t \qquad (1)$$

where Y_t is the observed (hip fracture rate) time series, $\phi_p(B)$ is a polynomial of order p, $\theta_q(B)$ is a polynomial of order q, B is the backward operator, and ε_t is an independent identically distributed (i.i.d) normal error with a zero mean and standard deviation σ_a. The ω_s is the parameter of the model and X is the independent, input, regressor or eXogeneous (climate) variable. The order of the ARMA model is identified by the Autocorrelation function (ACF) of the sample.

Using independent climate variables, many models can be developed with different climate variables. We follow a stepwise procedure to select the final model. The stepwise procedure considers the climate variables with the highest correlation coefficients to include in the model. In the following, another single climate variable is added in order to increase the R^2 of the model.

This procedure is repeated for different climate variables as far as the highest R^2 is obtained and the parameters of the models are significant at 5% level.

The nonlinear GARCH model is then used to model the conditional variance remaining in the residuals (ε_t) of the ARMA*X* model. The conditional variance (σ_t) is estimated by the following equation (Hamilton, 1994):

$$\sigma_t^{\,2} = w + \sum_{i=1}^{V} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{M} \beta_j \sigma_{t-j}^2 \qquad (2)$$

Where α and β are the parameters of the model and V and M are the order of the GARCH(V,M) model.

Multivariate GARCH model

The relationship between conditional variance of two or more variables is modelled by a multivariate GARCH (MGARCH) model (Engle and Kroner, 1995). The MGARCH model is a popular model in finance and econometrics. In a univariate GARCH model, the past values of the residuals, ε, makes a sigma field, $\Im_{t-1}$ with measurable σ_t^2. In other words, the current conditional variance, σ_t, depends on the previous variances and the residuals.

The extension from univariate GARCH model to an n-variate model requires allowing the conditional variance-covariance matrix of n-dimensional zero mean random variables, ε, to depend on the elements of the sigma field. Letting σ_t be measurable with respect to the sigma field, the multivariate GARCH is $\varepsilon | \Im_{t-1} \sim N(0, \sigma_t)$ where N indicates a normal distribution.

The parameterization for σ_t as a function of $\Im_{t-1}$ allows each element of σ_t to depend on v lagged values of the squares and cross-products of ε_t and the m lagged values of the elements of

σ_t. The number of these lags shows the order of the MGARCH (*v, m*) model. In this study, we propose a class of MGARCH model called the Constant Conditional Correlation (CCC) model (Bollerslev, 1990).

The conditional variance in the CCC model is defined as the following for two time series, *i* and *j*,

$$\sigma_t = D_t R D_t = (\rho_{ij} \sqrt{\sigma_{iit} \sigma_{jjt}}) \quad (3)$$

Where

$$D_t = diag(\sigma_{11t}^{1/2} \ldots \sigma_{NNt}^{1/2}) \quad (4)$$

σ_{iit} and σ_{jjt} can be defined as any univariate GARCH model and

$$R = (\rho_{ij}) \quad (5)$$

ρ_{ij} is the constant conditional correlation.

Results

Preliminary data analysis

A total number of 22,850 hip fracture incidences were observed during 1993-2004 in Montreal region among which 75.8% were females and 24.20% were male patients. The age-standardized hip fracture incidences for 1993-2004 period are 66, 284, 193 and 626 per 100000-person-year for F1, M1, F2 and M2 groups, respectively. The female to male ratio of hip fracture rate is 1.52 and 2.3 for age groups 40-74 and 75+, respectively. A seasonal variation of hip fracture rate for the groups is observed where the hip fracture rate begins increasing from November to January and then decreasing to the summer time (Modarres et al., 2012).

The relationship between hip fracture rate and climate variables is then investigated through simple Pearson product correlation coefficient [see Supplemental Material, Table 1]. One can see that these correlations are weak and show that the fluctuation of climate variables may not influence the HF incidences in a daily time scale. Looking at the autocorrelation functions (ACF) of the hip fracture rate time series [Supplemental Material, Figure 1] indicates that the coefficients are statistically insignificant (fall within the 95% confidence level). Thus, the daily information data set is insufficient for establishing MGARCH models.

Therefore, we transform daily time series into two new time series, in order to increase the autocorrelation structure. Then, in the following step, daily hip fracture rate time series are extracted from the output of the models fitted to new time series, to show the effect of climate variable on hip fracture rate in daily time scale.

The new time series are created by adding up the daily hip fracture rate in 3- and 5-day periods. The 3- and 5-day aggregations are selected to have a short time scale for our analysis. The climate variables are also transformed to 3- and 5-day variables to keep the uniformity of hip fracture rate-climate time series. This transformation increases 12 climate variables to 16 variables. The Pearson product correlation coefficients between aggregated hip fracture rate and climate variables are given in Tables 1 and 2. These tables show a significant association between climate and hip fracture rate time series. It also indicates stronger correlation coefficients than the coefficients between climate variables and hip fracture rate in daily time scale. Therefore, the aggregated hip fracture rate time series are used for modeling in the next sections.

From the ACF of the aggregated time series [see Supplemental Material, Figures 2 and 3], it is clear that the autocorrelation coefficients of aggregated 3- and 5-day time series are higher than

those for daily time series and remains significant to the higher lags. Therefore, we use the 3-day and 5-day data set instead of daily time series for modeling the relationship between climate variation and hip fracture rate in Montreal region using the proposed approaches.

ARMA*X*-GARCH models

Using aggregated hip fracture rate time series and 16 independent climate variables for hip fracture rate time series, it was observed that the GARCH(1,1) model is enough for the conditional variance remaining in the residuals of both time series. The final selected ARMA*X*-GARCH models and the climate variables included in each model as the e*X*ogenous variables are given in Tables 3 and 4 for 3-day and 5-day hip fracture rate time series, respectively.

It can be seen that for the younger groups (F1 and M1) the snow depth and temperature are the most important climate variables which describe 59% and 62% of the temporal variation of 3-day hip fracture rate, respectively. For the older groups (F2 and M2) of 3-day time series the maximum pressure and day length play the most important roles. These variables describe 57% and 56% of the variation of hip fracture incidences. It is also clear that the length of the day is the most frequent climate variables governing the hip fracture incidence in Montreal.

For the 5-day hip fracture rate time series, the models (Table 4) indicate that the temperature and snow depth are also the most important climate variables for the younger groups (F1 and M1 groups). The wind speed and maximum pressure also appear in the models. The ARMA*X*-GARCH models capture 76% and 78% of the temporal variation of 5-day hip fracture incidence for the younger groups. On the other hand, for F2 and M2 groups the maximum pressure seems to be the most important factor while the rainfall depth and length of the day also appear in the

models. These climate variables define 73% variation of 5-day hip fracture incidences for the older groups in Montreal region.

Daily hip fracture estimation

In this section, the daily hip fracture rate time series are extracted from the model-predicted 3- and 5- day aggregated time series to show the model performance for modeling daily hip fracture rate. The predicted daily time series are plotted against the observed daily hip fracture rate in Figures 2 and 3. The figures indicate that the model can describe (in average) 50% of daily variation of hip fracture rate in Montreal region in related to climate variables. Figures 2 and 3 also appear to show that the error of the model decreases while the number of hip fracture rate increases. In other words, the performance of the model is better for high values of hip fracture rate than for low values. This suggests that the association of climate variables to the hip fracture incidence is much stronger when hip fracture rate is high and the influence of climate conditions on small number of hip fracture incidence seems to be insignificant. The small number of incidences which are not related to climate conditions may be due to indoor incidences or other factors of falling, especially for 75+ people. In the next section, this statement is examined by the multivariate model.

Conditional variance and covariance

Here, the dynamic variances and covariances of hip fracture and climate variables based on the CCC-GARCH model are given to show how the predictive value of climate variables for hip fracture rate changes across the full range of hip fracture rate values.

The first step for a CCC model is to estimate the conditional variance for each time series based on a GARCH model. The conditional variance allows us to test how the variance itself changes over time by comparing the variability of different hip fracture rate time series.

For the conditional variances of the 3- and 5-day hip fracture rate [see Supplemental Material, Figures 4 and 5] the peak variances for F1 group are observed in 1994, 1997, 1998 and 1999. All the peak variances are observed in January except the high value in April, 1997. The highest variances for F2 group are observed in March 1993 and January, 1998. For M1 group, the peaks are observed in January, 1998, 1999 and 2002 while for M2 group, the highest variance is observed in January 1994.

These figures also show that the conditional variances are not constant over time and the variance of the older groups (F2 and M2 groups) visually seems to have more variability and fluctuation than the younger groups (F1 and M1 groups). This suggests that the hip fracture rate for F2 and M2 groups changes in a larger range than F1 and M1 groups

Moreover, a seasonality of conditional variance can also be observed. The monthly average and standard deviation of the conditional variance of the 3-day hip fracture rate is illustrated as an example in Figures 4(a) and 4(b).

The highest average of conditional variance is observed in autumn-winter season for the younger group (Figure 4a) while a sharp seasonal variation is not evident for F2 and M2 groups. On the other hand, the monthly change of the standard deviation of the conditional variance (Figure 4b) for the younger groups (F1 and M1) is consistent with the monthly change in the average (Figure 4a) while the standard deviation for the elder group (M2 and F2) shows high monthly variation for both winter and summer months. This suggests a higher instability in the variance of the elder group. The seasonality of the mean was also reported by Modarres et al (2012). Therefore, one

can see that both linear and nonlinear characteristics of hip fracture rate time series show seasonality.

It is also observed that the standard deviation of the conditional variance of winter season is higher than the summer time for all groups. The higher dispersal of the variance can also be observed in the boxplot of the monthly conditional variance which is given for F1 group as an example in Figure 4c. The boxplots also reveals higher variability and range of the variance in winter time, especially in December and January when temperature and the length of the day are decreasing and snow depth is increasing.

The time varying covariance between hip fracture incidences and climate conditions is also estimated by applying equations 3 and 4. For the CCC models, we use the climate variables determined by the ARMA*X*-GARCH model for each hip fracture rate time series. Plotting the conditional covariance against the hip fracture rate time series shows how the conditional variance of the climate variables (included in the model) influences the temporal variation of the hip fracture rate incidences.

The conditional covariance between different hip fracture groups and climate variables are plotted against the hip fracture rate (Figures 5 and 6) for 3-day hip fracture time series. Other figures related to 3-day hip fracture are presented in appendix Figures 6 and 7. All these figures indicate a strong and nonlinear relationship between the time varying hip fracture variances and the variances of the climate variables. It can be observed that the association between hip fracture incidence and climate variables is very weak or linear for small numbers of hip fracture incidences while this association (climate effect on hip fracture rate) increases rapidly and in a nonlinear fashion for the higher hip fracture rate values. For example, the conditional covariance between F1 and snow depth is almost constant or zero for hip fracture rate < 1 and begins

increasing exponentially afterward. Similar results are observed for 5-day hip fracture rate data and for other groups and climate variables [see Supplemental Material, Figures 8 to 11).

Discussion

This study implemented a novel multivariate GARCH model to show how climate variables affect hip fracture rates in Montreal, Canada. We estimated the time varying second order moment or the conditional variance of the hip fracture rate and its relationship to the conditional variance of climate variables. These models demonstrate strong nonlinear associations between the variances of climate variables and hip fracture rates, implying that high hip fracture risk depends more strongly on severe weather conditions than on average conditions. Accordingly, the risk of hip fracture incidences appears to increase when specific climate conditions, such as heavy snowfall or low temperature, remain for a time period. (e.g. a couple of days). The nonlinear model applied in this study clearly supports this statement.

The nonlinear ARMA*X*-GARCH models not only identify the most important climate variables on hip fracture incidence but also show the time-varying variance of the hip fracture rate in contrast to linear time series models applied by Lin and Xiraxagar (2006) and Modarres et al. (2012) or the Poisson regression model (turner et al., 2011; Bischoff-Ferrari et al. 2007) which only considered the conditional mean (first order moment) of hip fracture and its temporal variation. The climate variables included in the models, explain 56% to 78% variation of aggregated hip fracture rate and almost 50% variation of daily HF incidence for both age groups. Furthermore, the proposed model building procedure based on estimation daily hip fracture rate from short-period aggregated time series suggests a good method for future studies on daily hip fracture time series modeling in the absence of significant autocorrelation. However, it should be noted that the climate variables have been averaged over the study region. Future studies should

consider more dense meteorological data to achieve a better understanding of the association between climate variation and hip fracture in daily time scale.

The multivariate CCC model shows a nonlinear relationship between hip fracture rate and climate variables for both gender and age groups. The covariances between hip fracture rate and climate variables are strongly significant and increase exponentially. The insignificant covariance between low hip fracture rate and climate variables probably suggests that the variation in the background of low hip fracture rate may not depend on weather conditions and we should seek for other explanations such as indoor falls or drug consumption. Future studies should separate indoor and outdoor incidences in order to establish a more precise climate-hip fracture association. Although examining simultaneously weather variables the elderly may prefer to stay at home during adverse climate conditions, the MGARCH approach shows that the risk of hip fracture incidences may increase when the adverse climate conditions remain for several days. This suggests future investigation of the dynamic and social behaviour of people through another study, for example a questionnaire-based study, to understand the preferences of the elderly for staying at home or going outside during and after harsh climate conditions. This could lead to free automated alert systems like those already found in the field of air quality in countries such as Canada (Sante et service sociaux Quebec, http://www.msss.gouv.qc.ca/sujets/santepub/environnement/index.php?cote-air-sante-abonnement-aux-alertes-en), the United States (Airalert service for Herts and beds, https://www.airalert.info/hertsbeds/registration.aspx), the United Kingdom and France (Nort east London and city (NHS), http://www.onel.nhs.uk/for-health-professionals/severe-weather-warnings.htm).

Conclusions

This study is the first to apply nonlinear univariate and multivariate GARCH model to show the influence of climate conditions on hip fracture incidence. The proposed approaches for modeling hip fracture rate-climate association seem to effectively establish both linear and nonlinear effects of climate variables on hip fracture incidences. The findings of this study indicate that the effect of the climate condition on hip fracture incidence in the Montreal region appears to be a nonlinear process, exponentially increasing and becoming more significant with adverse weather conditions. Given that the association of small number of hip fracture rate and climate variables is insignificant or linear for both gender and age groups under study, and given that the majority of hip fracture incidences are caused by a fall (Masud et al., 2001), this low baseline hip fracture rate is most likely related to indoor falls or other known factors such as prescription drug (as well as alcohol or other drugs) consumption affecting the central nervous system or higher prevalence of cognitive impairment in elderly people. In addition, other possible confounders such as people's behavior, weekends and holiday vocations should be considered in the future studies.

Many extreme weather conditions have shown an increase in frequency, intensity or duration in Quebec Province (Lemmen et al. 2001) and will most likely follow a similar trend in the future. It then becomes important to plan future medical services based on the most robust models that would include climate determinants for the best approximation of the expected incidence of such severe and debilitating injuries as hip fractures, especially a for short-period time scale. The approach proposed here will support informed decisions for a better allocation of scarce resources and provide a climate-related preventive message to the people at risk for hip fracture. Some climate change adaptation programs already begin to use weather data for their services planning currently (Oven Et al. 2012).

Acknowledgement

The work was supported by the Institut national de santé publique du Québec (INSPQ) and the Institut national de la recherche scientifique – Centre Eau, Terre et Environnement (INRS-ETE), the Green Fund in the framework of Action 21 of the government of Québec's 2006-2012 Climate Change Action Plan and the Network of Centers of Excellence GEOIDE and the Fonds de Recherche en Santé du Québec (FRSQ).

References

Benetos LS, Babis GC, Zoubos AB, Benetou V, Soucacos, PN (2007) Factors affecting the risk of hip fractures. Injury- International Journal of the care of the injured 38: 735-744.

Bischoff-Ferrari HA, Orav JE, Barrett JA, and Baron JA (2007) Effect of seasonality and weather on fracture risk in individuals 65 years and older. Osteoporosis International 18:1225–33

Bollerslev T (1990) Modeling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. Review of Economics and Statistics 72: 498-505

Bulajic-Kopjar M (2000) Seasonal variations in incidence of fractures among elderly people. Injury Preventation 6: 16-19

Cummings SR, Melton III JR (2002) Epidemiology and outcomes of osteoporotic fractures. Lancet 359: 1761–1767.

Edvardsen K, Engelsen O, Brustad M (2009) Duration of vitamin D synthesis from weather model data for use in prospective epidemiological studies. Int J Biometeorol., 53: 451-459

Engle RF, Kroner KF (1995) Multivariate simultaneous generalized ARCH. Econometrics Theory 11: 122-50

Gullberg B, Duppe H, Nilsson B, Redlund-Johnell I, Sernbo I, Obrant K, et al. (1993) Incidence of hip fractures in Malmo, Sweden (1950-1991). Bone. Suppl 1:S23-S29.

Gronskag AB, Siri F, Romundstad P, Langhammer A, Schei B (2010) Incidence and seasonal variation in hip fracture incidence among elderly women in Norway. The HUNT Study. Bone. 46: 1294-1298

Hamilton, J.D (1994) Time series analysis. Princeton University Press.

Hipel KW, McLeod AE (1996) Time series modeling of water resources and environmental systems. Elsevier, Amsterdam, The Netherlands

Harvey N, Dennison E (2010) Osteoporosis: impact on health and economics. Nature reviews rheumatology 6: 99-105.

Lambert LJ, Blais C, Hamel D, Brown KA, Giguere M, et al. (2010) reliability of the assessment of pattern comorbidities using Quebec 's medico-administrative data. Canadian Journal of cardiology 26: supplement: D 29D-29D.

Lan TY, Hou CY, Chen CY, Chang WC, Lin J, Lin CC, et al. (2010) Risk factures for hip fracture in older adults: a case-control study in Taiwan. Osteoporosis International 21: 773-784.

Lemmen DS, Warren FJ, Lacroix J, Bush E (2011) From Impacts to Adaptation: Canada in a Changing Climate 2007. Ottawa: Government of Canada; 2008. http://adaptation.nrcan.gc.ca/assess/2007/index_e.php. Accessed July 28, 2011.

Levy AR, Bensimon DR, Mayo NE, Leighton HG (1998) Inclement weather and the risk of hip fracture. Epidemiology 9: 172-177.

Levy AR, Tamblyn RM, Fitchett D, McLeod PJ, Hanley JA (1999) Coding accuracy of hospital discharge data for elderly survivors of myocardial infarction. Canadian Journal of cardiology 15:1277-1282

Lin HC, Xiraxagar S (2006) Seasonality of hip fracture and estimates of season attributable effects: a multivariate ARIMA analysis of population-based data. Osteoporosis International 17: 795–806.

Lofthus CM, Osnes EK, Falch JA, Kaasad TS, Kristiansen IS, Nordsletten L (2001) Epidemiology of hip fractures in Oslo, Norway. Bone 29: 413-418

Masud T. Morris RO (2001) Epidemiology of falls. Age and ageing 30: supplement 3-7

Mirchandani S, Aharonoff GB, Hiebert R, Capla EL, Zuckerman JD, Koval KJ (2005) The effects of weather and seasonality on hip fracture incidence in older adults. Orthopedics 28:149– 155

Modarres R, Ouarda TBMJ, Vanasse A, Orzanco MG, Gosselin P (2012) Modeling seasonal variation of hip fracture in Montreal, Canada. Bone. 50: 909-16

Oven KJ, Curtis SE, Reaney S, Riva M, stewart MG, et al. (2012) Climate change and health and social care: defining future hazard, vulnerability and risk for infrastructure systems supporting older people 's health care in England. Applied Geography 33, 16-24

Quigley PA, Campbell RR, Bulat T, Olney RL, Buerhaus P, Needleman J (2012) Incidence and Cost of Serious Fall-Related Injuries in Nursing Homes. Clinical nursing research 21: 10-23.

Sumukadas D. Witham M, Struthers A and Mcmurdo M (2009) Day length and weather conditions profoundly affect physical activity levels in older functionally impaired people. Journal of Epidemiology and Community Health 63: 305-309.

Tenias JM, Estarlich M, Fuentes-Leonarte V, Iniguez C, Ballester F (2009) Short-term relationship between meteorological variables and hip fractures: An analysis carried out in a health area of the Autonomous Region of Valencia, Spain (1996-2005). Bone 45: 794-98

Turner RM, Hayen A, Dunsmuir WT, Finch CF (2011) Air temperature and the incidence of fall-related hip fracture hospitalisations in older people. Osteoporosis International 22:1183-9.

Table 1 Correlation coefficients between 3-day hip fracture rate and climate variables

Climate variables	F1r	F2r	M1r	M2r
Maximum temperature ($^{\circ}C$)	-0.107*	-0.128*	-0.172*	-0.137*
Minimum temperature ($^{\circ}C$)	-0.122*	-0.131*	-0.200*	-0.143*
Mean Temperature ($^{\circ}C$)	-0.114*	-0.128*	-0.188*	-0.141*
Rainfall depth (mm)	-0.029	-0.046*	-0.027	-0.072*
Snow depth (mm)	0.125*	0.064*	0.097*	0.072*
Precipitation depth (mm)	0.013	-0.017	0.018	-0.045*
Number of snowy days	0.121*	0.105*	0.163*	0.097*
Number of rainy days	-0.092*	-0.079*	-0.118*	-0.096*
Number of days with precipitation	-0.028	-0.031	-0.020	-0.040*
Maximum snow depth (mm)	0.130*	0.079*	0.198*	0.082*
Maximum wind speed (km/s)	0.017	0.020	0.060*	-0.021
Mean Wind speed (km/s)	0.010	0.015	0.047*	-0.033
Day Length (hr)	-0.022	-0.077*	-0.069*	-0.069*
Maximum pressure (hp)	0.061*	0.064*	0.107*	0.069*
Minimum pressure (hp)	-0.004	-0.015	-0.033	-0.023
Mean pressure (hp)	0.035	0.024	0.056*	0.048*

* $p<0.01$

Table 2 Correlation coefficients between 5-day hip fracture rate and climate variables

Climate variables	F1r	F2r	M1r	M2r
Maximum temperature ($^{\circ}C$)	-0.150*	-0.162*	-0.239*	-0.177*
Minimum temperature ($^{\circ}C$)	-0.170*	-0.164*	-0.277*	-0.186*
Mean Temperature ($^{\circ}C$)	-0.159*	-0.162*	-0.261*	-0.181*
Rainfall depth (mm)	-0.057*	-0.048*	-0.072*	-0.095*
Snow depth (mm)	0.187*	0.099*	0.159*	0.100*
Precipitation depth (mm)	0.017	-0.001	0.003	-0.057*
Number of snowy days	0.168*	0.151*	0.245*	0.137*
Number of rainy days	-0.137*	-0.103*	-0.190*	-0.119*
Number of days with precipitation	-0.040*	-0.027	-0.040*	-0.039
Maximum snow depth (mm)	0.182*	0.101*	0.265*	0.109*
Maximum wind speed (km/s)	0.064*	0.029	0.039*	-0.017
Mean Wind speed (km/s)	0.046*	0.018*	0.027	-0.015
Day Length (hr)	-0.055*	-0.130*	-0.114*	-0.105*
Maximum pressure (hp)	0.081*	0.086*	0.157*	0.094*
Minimum pressure (hp)	-0.026	-0.044*	-0.059*	-0.048*
Mean pressure (hp)	0.038	0.042*	0.070*	0.054*

* $p<0.01$

Table 3 time series models for 3-day hip fracture rate time series

Groups	model	Climate variables included	R^2
M1	ARMAX(4,1,1)- GARCH(1,1)	Snow depth, minimum temperature, Day length	0.624
M2	ARMAX(4,1,1)- GARCH(1,1)	Maximum pressure, Day length	0.562
F1	ARMAX(4,1,1)- GARCH(1,1)	Snow depth, minimum temperature	0.596
F2	ARMAX(4,1,1)- GARCH(1,1)	Day length, maximum pressure	0.578

Table 4 time series models for 5-day HFr time series

groups	model	Climate variables included	R^2
M1	ARMAX(6,1,1)- GARCH(1,1)	minimum temperature, maximum snow depth, maximum pressure	0.782
M2	ARMAX(6,1,1)- GARCH(1,1)	Maximum pressure, rainfall depth	0.736
F1	ARMAX(6,1,1)- GARCH(1,1)	Snow depth, minimum temperature, maximum wind speed	0.767
F2	ARMAX(6,1,1)- GARCH(1,1)	Maximum pressure, day length	0.733

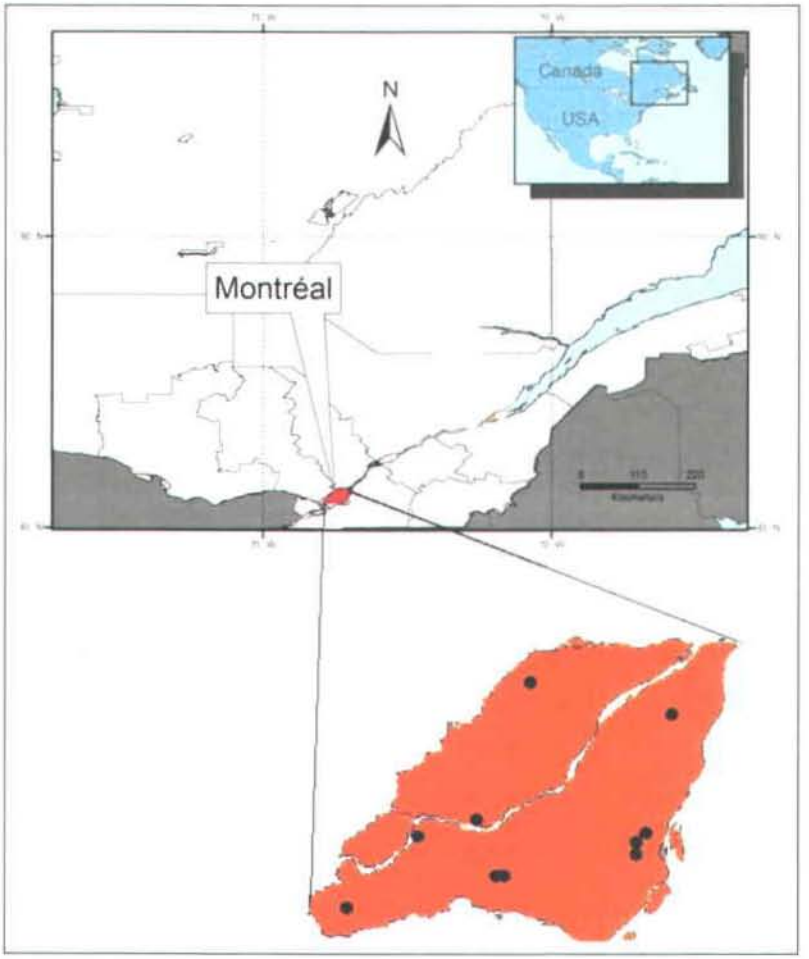


Figure 1. Geographical location of study area and meteorological stations

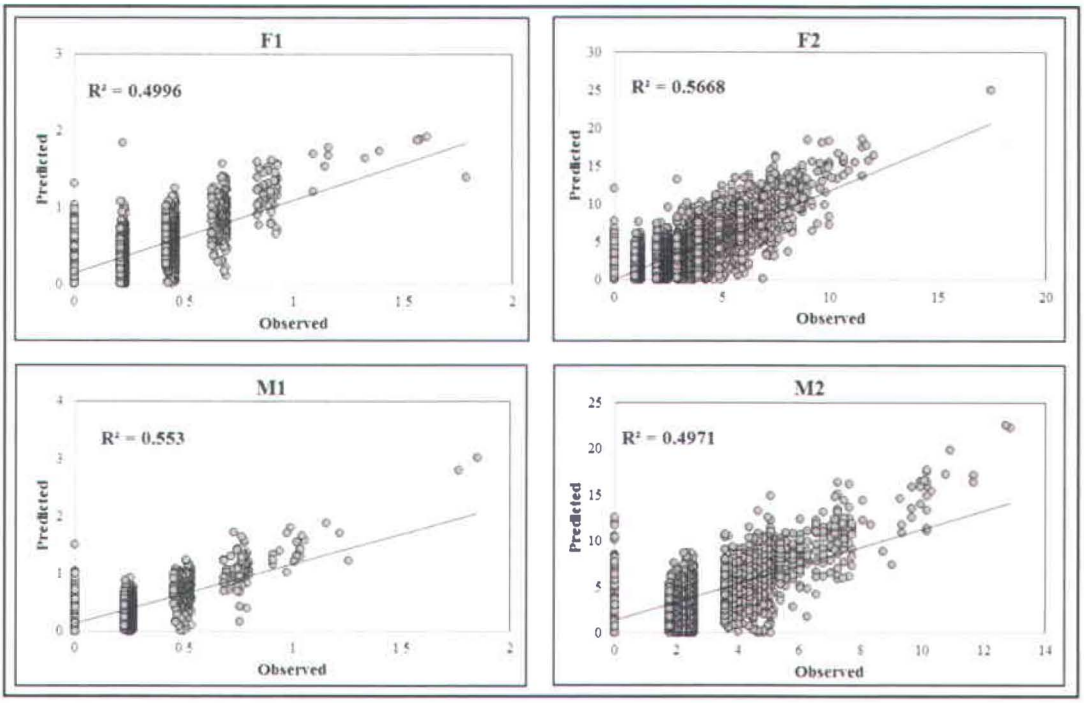


Figure 2 Observed against estimated daily hip fracture rate time series extracted from 3-day hip fracture rate time series

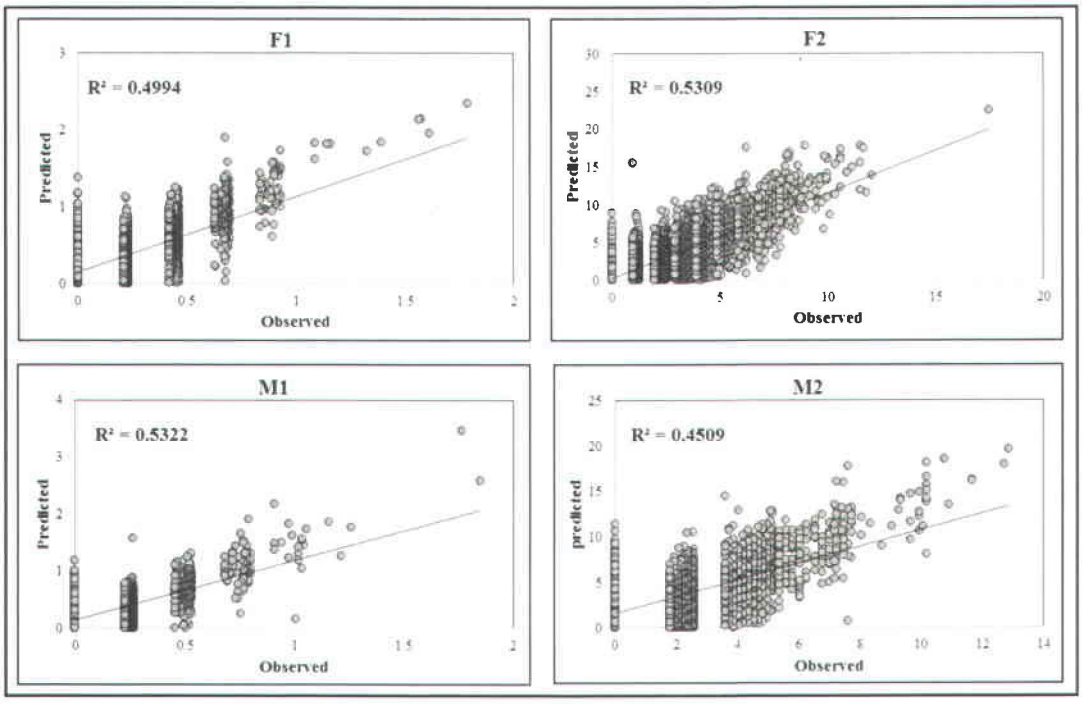


Figure 3 Observed against estimated daily hip fracture rate time series extracted from 5-day hip fracture rate time series

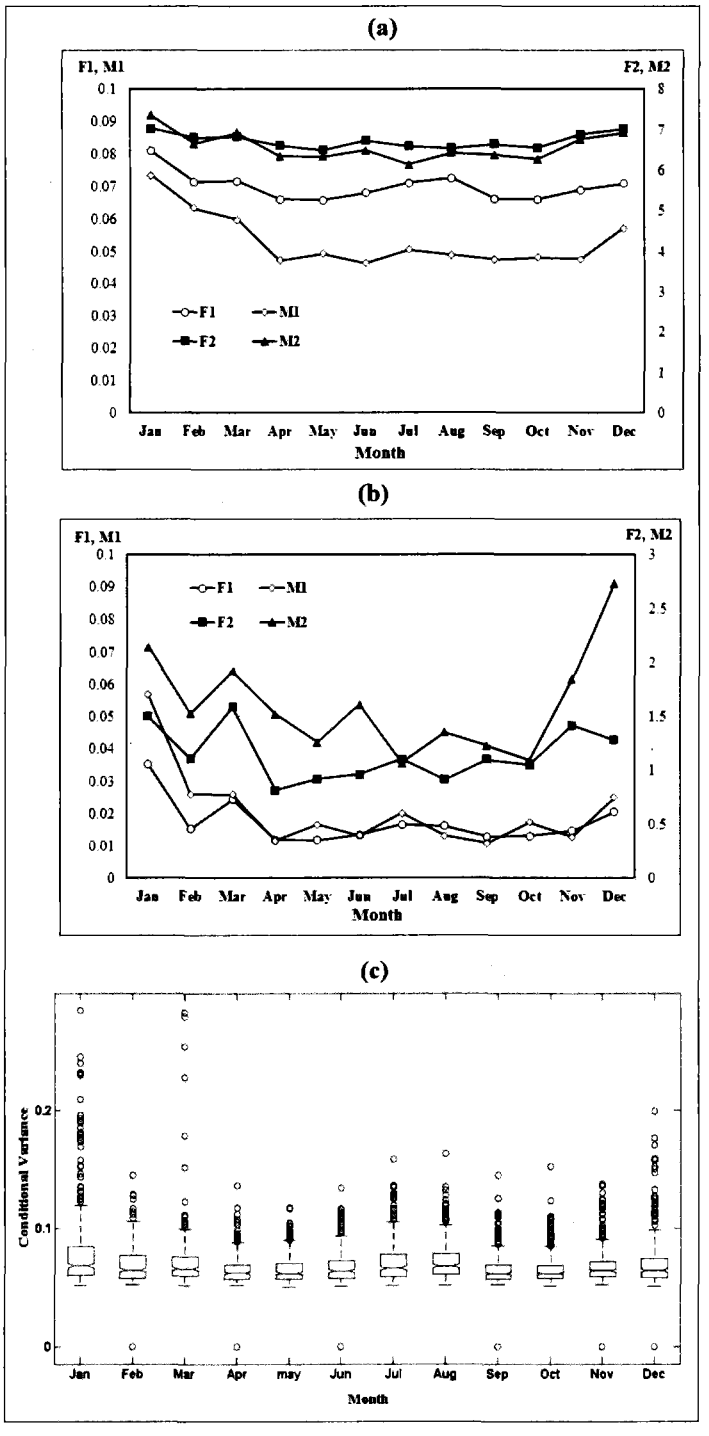


Figure 4 seasonality of the average and variance of the conditional variance of 3-day time series

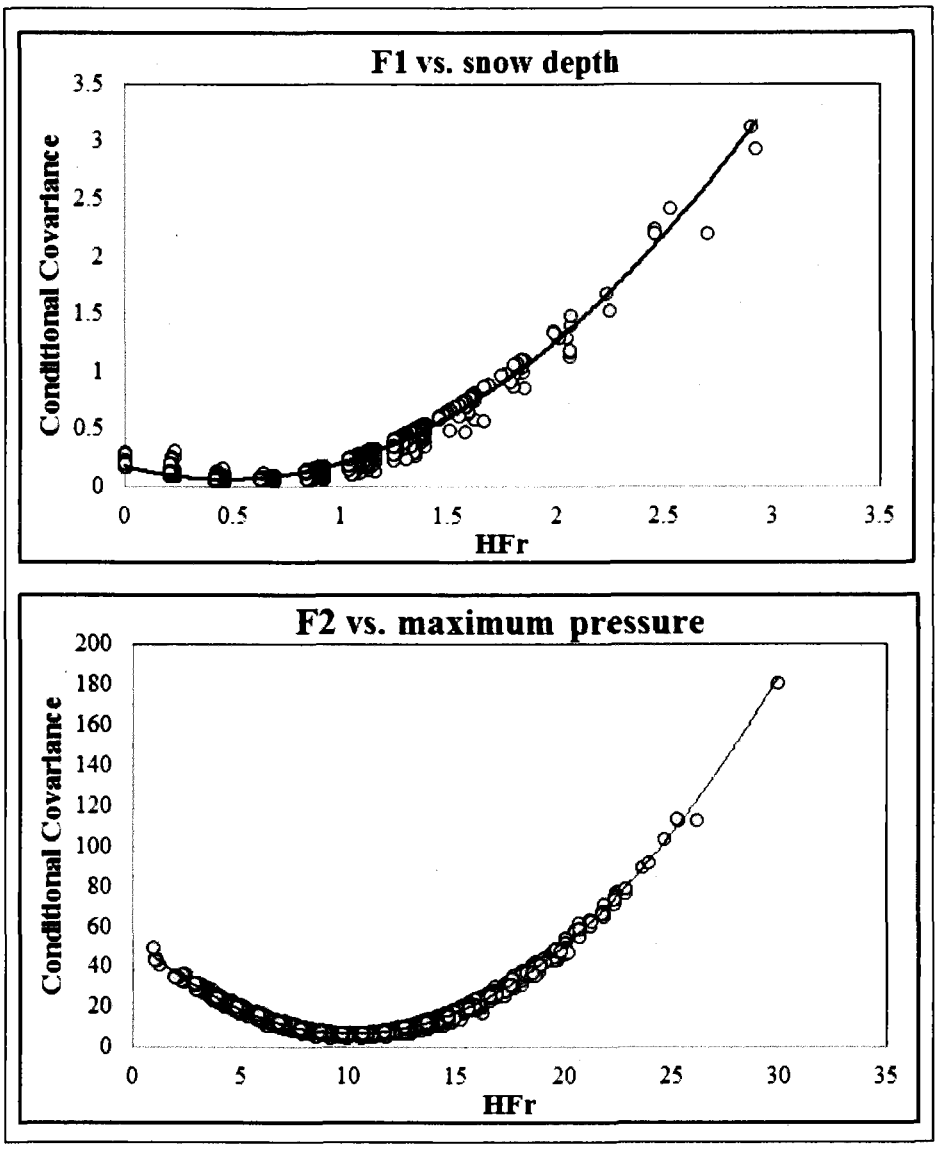


Figure 5. Conditional Covariance against 3-day hip fracture rate. Example for F1 and F2 groups

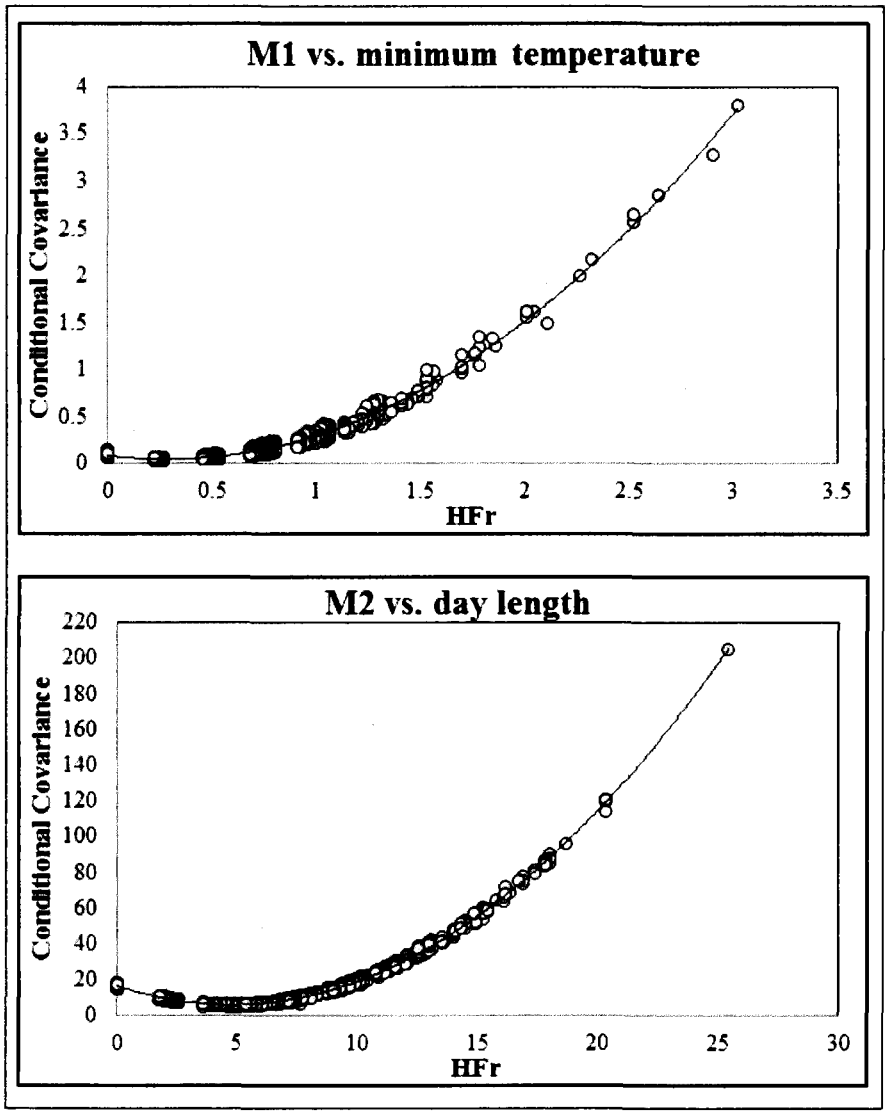


Figure 6. Conditional Covariance against 3-day hip fracture rate. Example for M1 and M2 groups

Supplemental Material

Modeling climate effects on hip fracture rate by multivariate GARCH model in Montreal region, Canada

Reza Modarres[a], Taha B. M. J. Ouarda[a,b], Alain Vanasse[c], Maria Gabriela Orzanco[c], Pierre Gosselin[d]

[a]Hydroclimate modeling group, INRS-ETE, 490 de la Couronne, Quebec, Qc, Canada, G1K 9A9

[b] Masdar Institute of Science and Technology, P.O. Box 54224, Abu Dhabi, UAE

[c]Groupe de recherche PRIMUS, Université de Sherbrooke 3001, 12e avenue Nord, Sherbrooke, QC, Canada, J1H 5N4

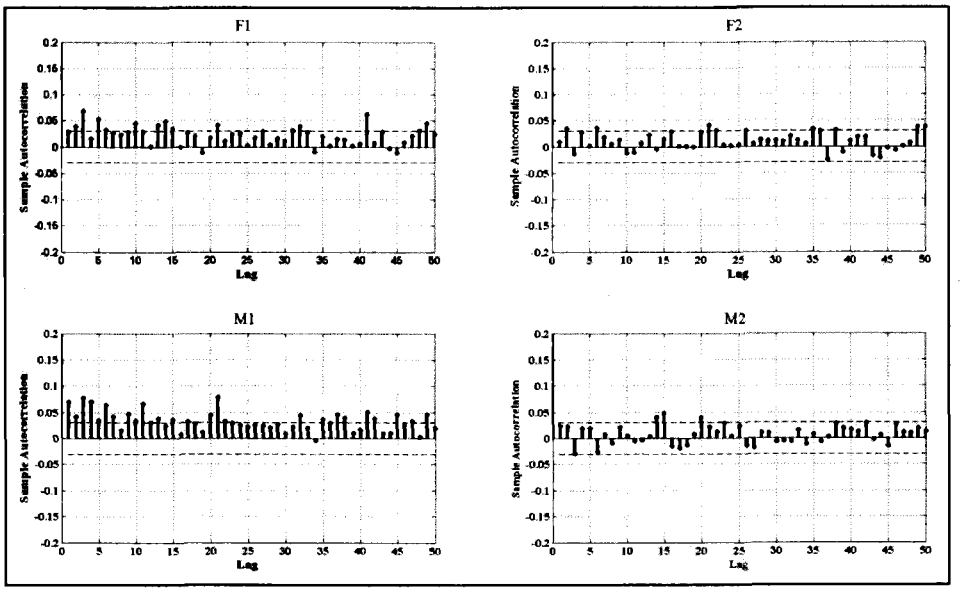
[d]Changements climatiques, Unité Santé et environnement, Institut National de santé Publique du Québec (INSPQ), 945 rue Wolf, Québec, Qc, Canada, G1V 5B3

Appendix A. Additional Analyses

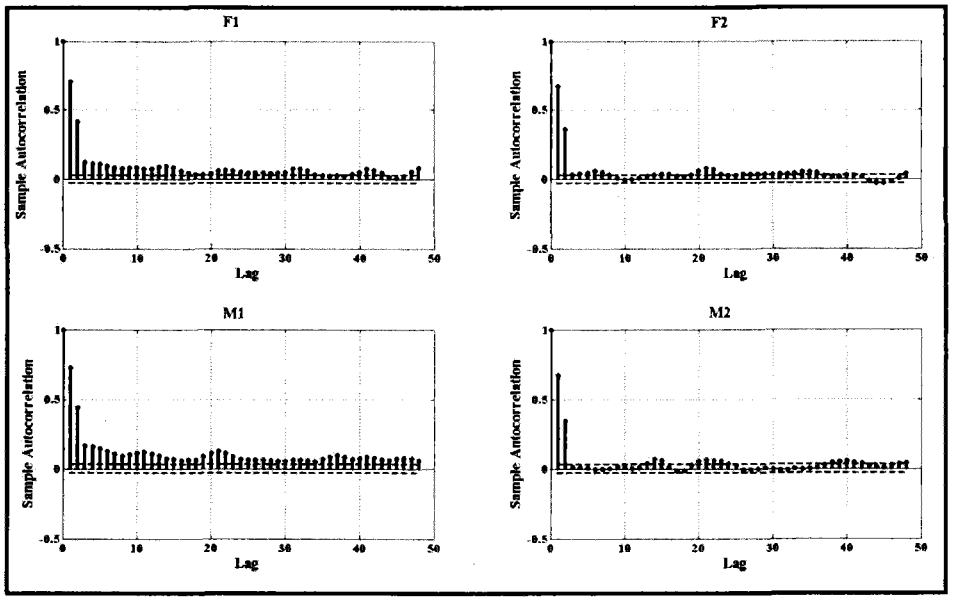
Appendix Table 1. Correlation coefficients between daily HFr and climate variables

Climate variables	F1r	F2r	M1r	M2r
Maximum temperature ($^{\circ}C$)	-0.027	-0.064***	-0.028	-0.029
Minimum temperature ($^{\circ}C$)	-0.038	-0.070***	-0.034	-0.032
Mean Temperature ($^{\circ}C$)	-0.032	-0.067***	-0.031	-0.031
Rainfall depth (mm)	-0.016	-0.028	-0.010	-0.013
Snow depth (mm)	0.031	0.054***	-0.004	0.008
Precipitation depth (mm)	-0.008	-0.004	-0.012	-0.012
Maximum snow depth (mm)	0.029	0.048***	0.015	0.000
Mean Wind speed (km/s)	0.003	0.018	0.021	-0.012
Day Length (hr)	-0.023	-0.038	-0.31	-0.031
Maximum pressure (hp)	0.020	0.018	-0.012	-0.05
Minimum pressure (hp)	0.019	-0.002	-0.012	0.004
Mean pressure (hp)	0.022	0.014	-0.013	0.000

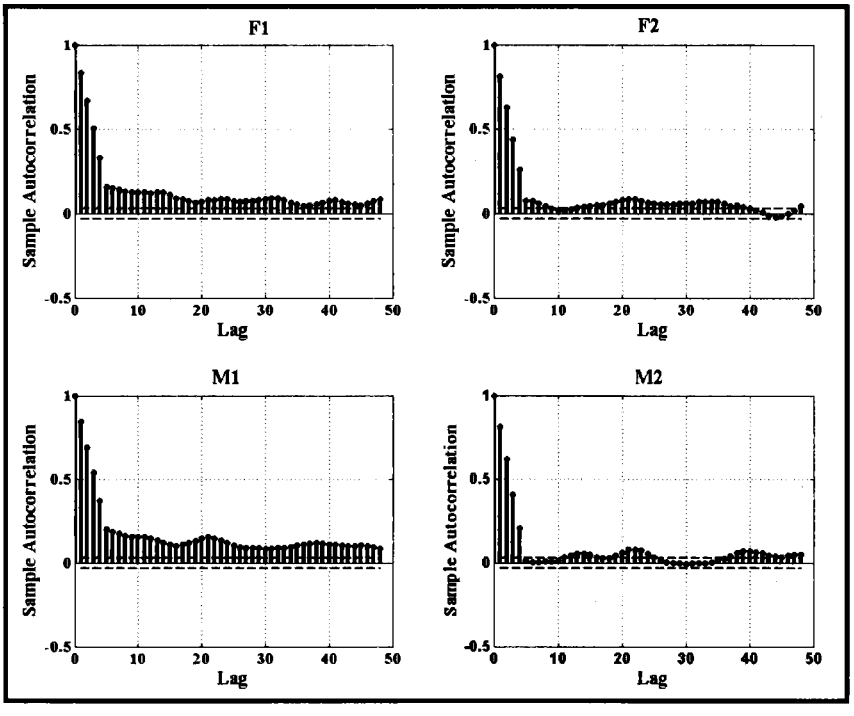
*** $p<0.01$,** $p<0.05$,* $p<0.10$



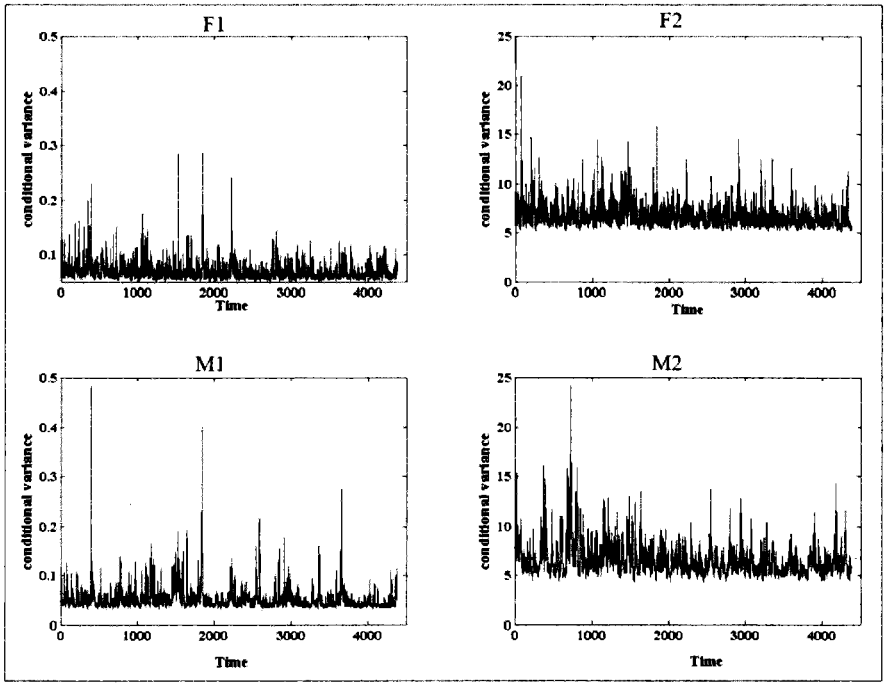
Appendix Figure 1. Autocorrelation functions of daily HFr time series of different age and gender groups



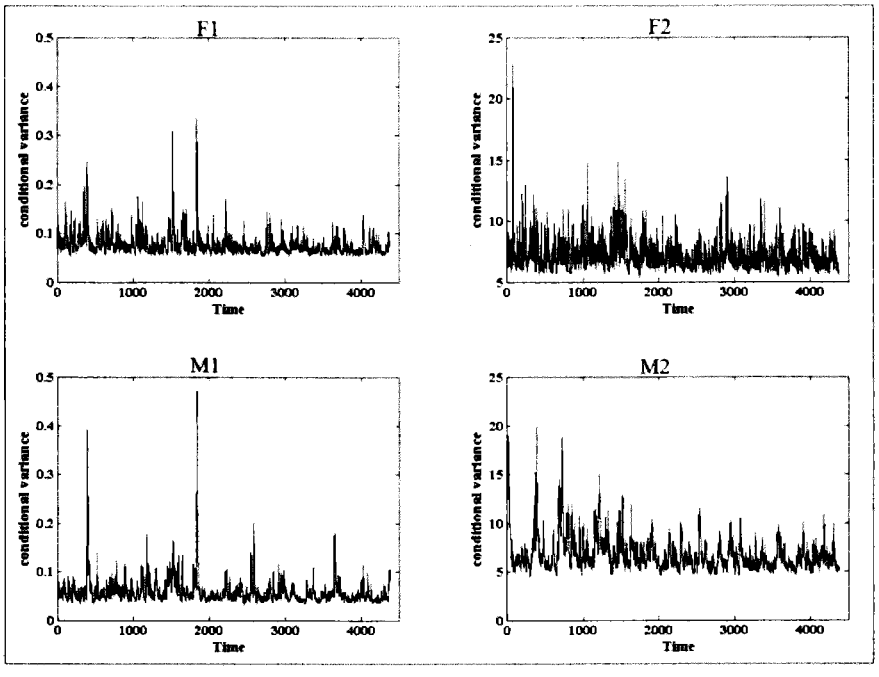
Appendix Figure 2. Autocorrelation functions of 3-day HFr time series of different age and gender groups



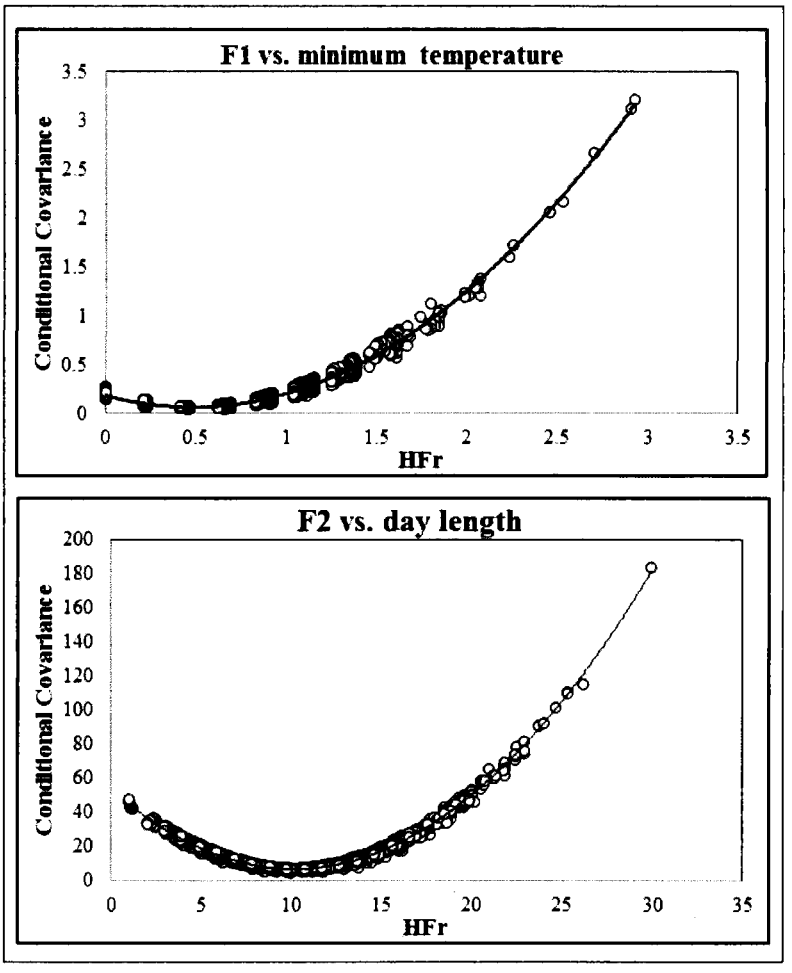
Appendix Figure 3. Autocorrelation functions of 5-day HFr time series of different age and gender groups



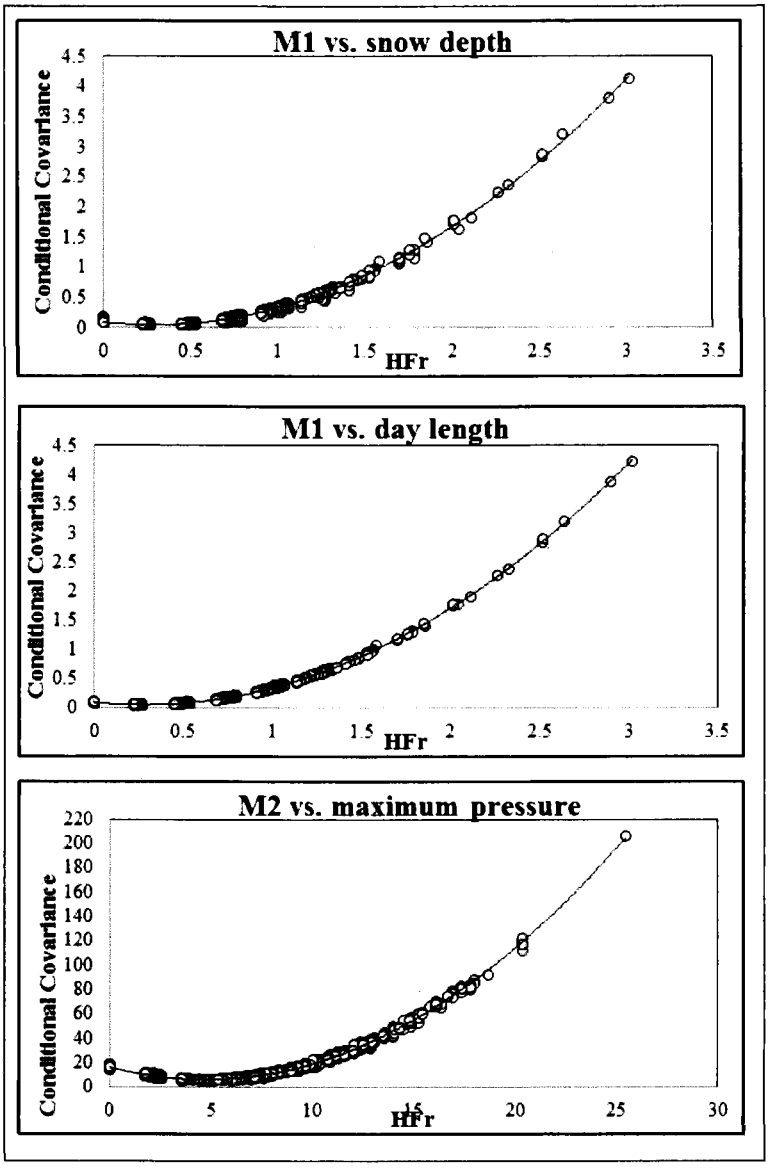
Appendix Figure 4. Conditional variance of 3-day HFr time series



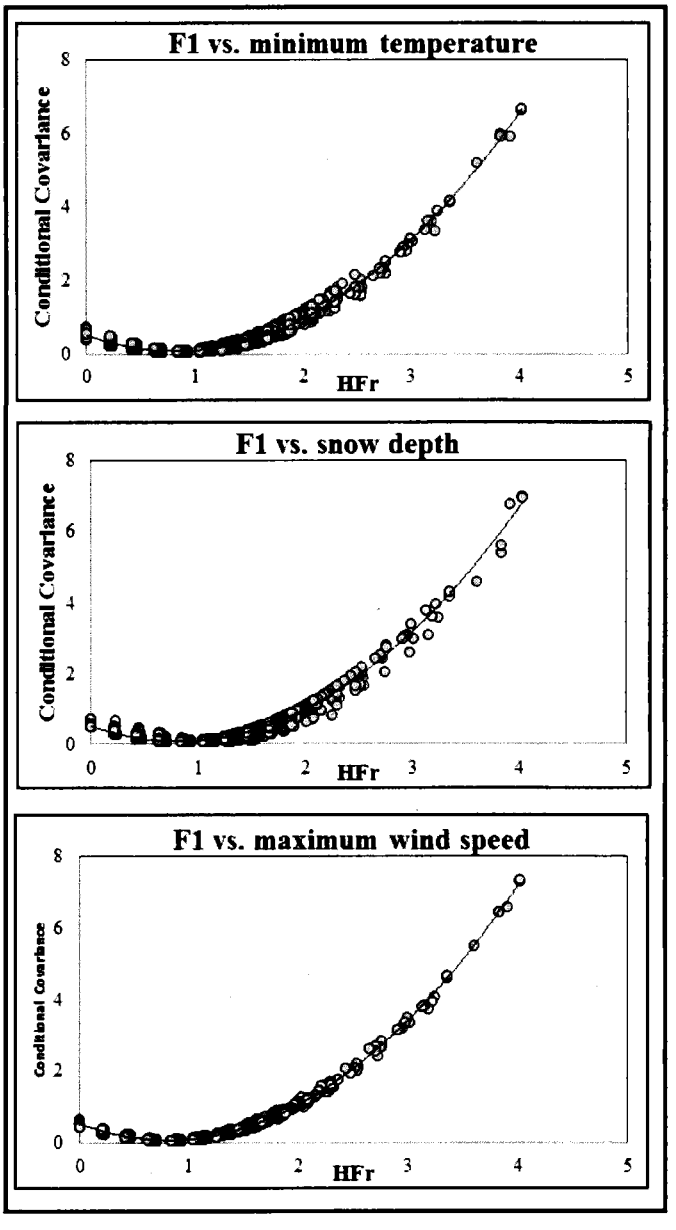
Appendix Figure 5. Conditional variance of 5-day HFr time series



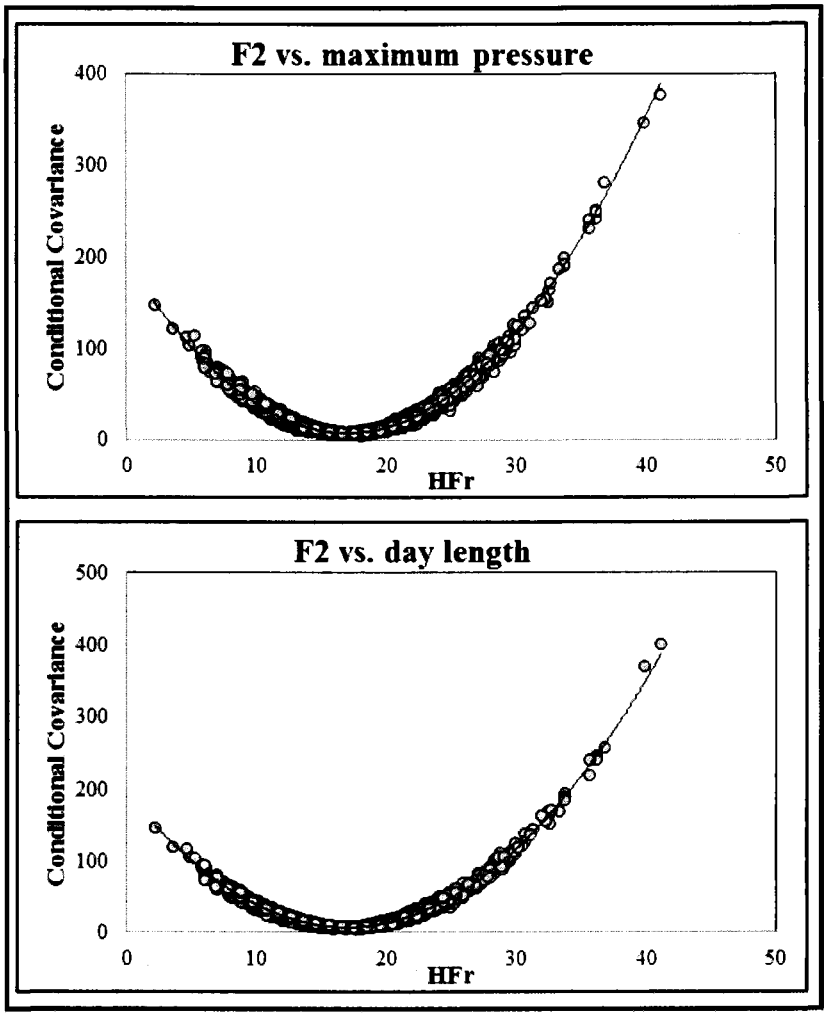
Appendix Figure 6. Conditional Covariance against 3-day hip fracture rate. Example for F1 and F2 groups



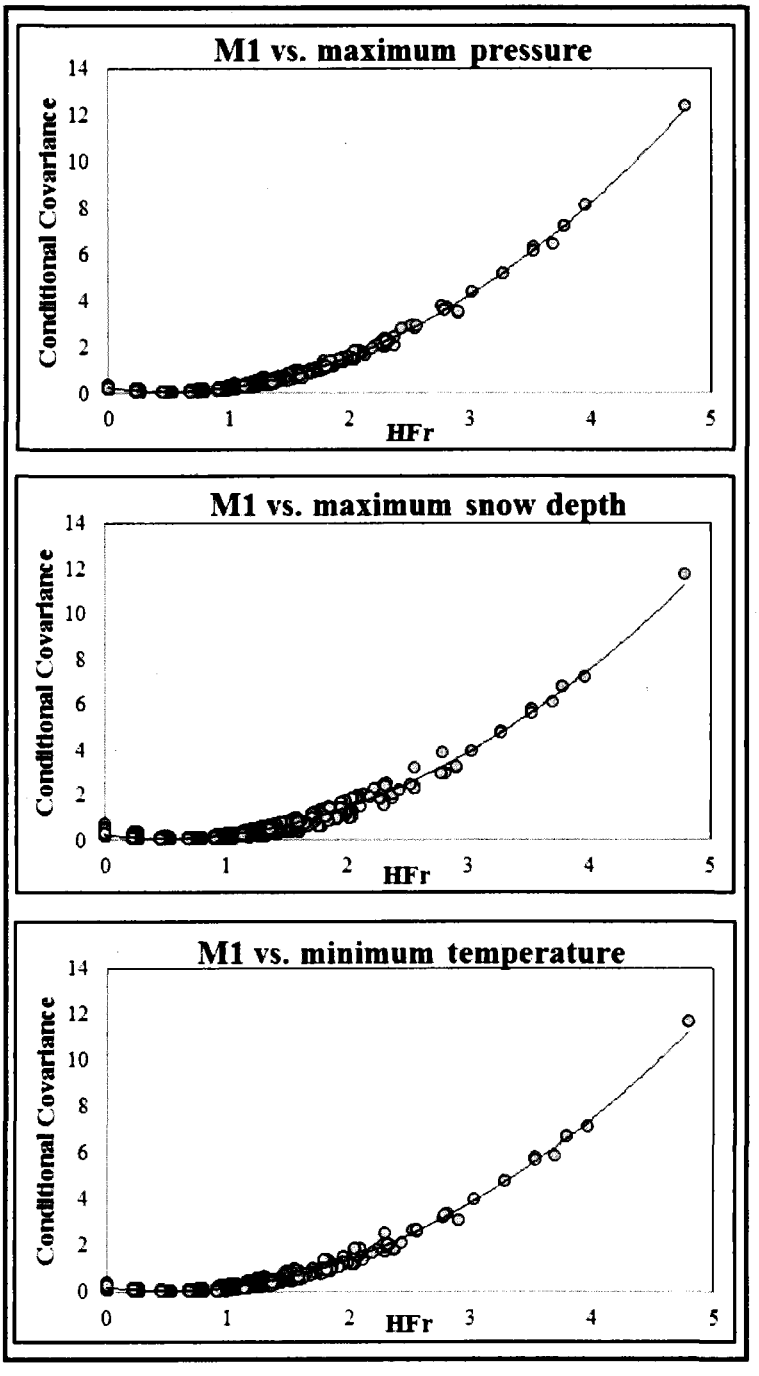
Appendix Figure 7. Conditional Covariance against 3-day hip fracture rate. Example for M1 and M2 groups



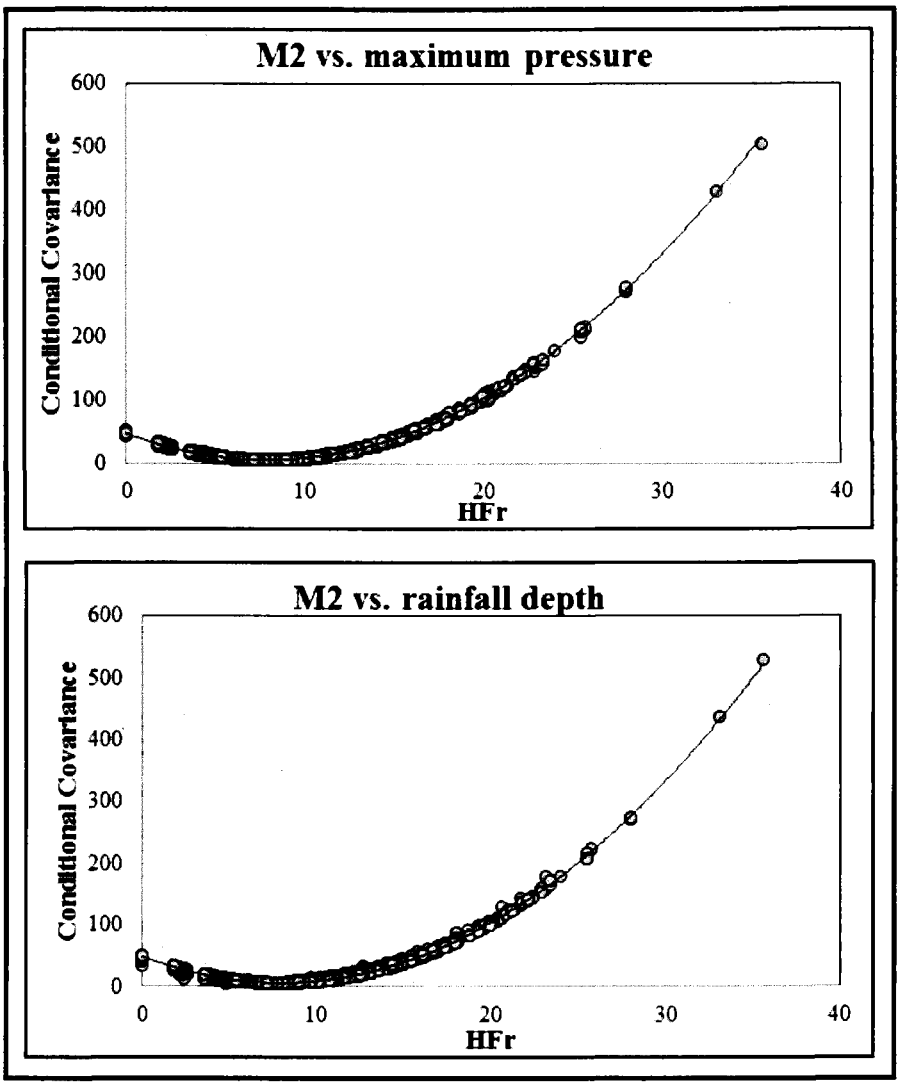
Appendix Figure 8. Conditional Covariance between 5-day F1 time series and climate variables



Appendix Figure 9. Conditional Covariance between 5-day F2 time series and climate variables



Appendix Figure 10. Conditional Covariance between 5-day M1 time series and climate variables



Appendix Figure 11. Conditional Covariance between 5-day M2 time series and climate variables

Article 6. A Generalized Conditional Heteroscedastic model for temperature downscaling

A Generalized Conditional Heteroscedastic model for temperature downscaling

R. Modarres[a*], T. B. M. J. Ouarda[a,b]

[a] Hydroclimate modeling group, INRS-ETE, 490 De La Couronne, Québec, QC, Canada

Tel: +1 418 654-3842, Fax: +1 418 654-2600,

E-mail: Reza.Modarres@ete.inrs.ca or taha_ouarda@ete.inrs.ca

[b] Masdar Institute of Science and Technology, P.O. Box 54224, Abu Dhabi, UAE

[*] Corresponding author

Submitted for publication to Journal of Climate

Abstract

This study describes a method for deriving the time varying second order moment, or heteroscedasticity, of local daily temperature and its association to large CGCM predictors. This is carried out by applying a multivariate Generalized Autoregressive Conditional Heteroscedasticity (MGARCH) approach to construct the conditional variance-covariance structure between GCM predictors and maximum and minimum temperature time series during 1980-2000. Two MGARCH specifications namely diagonal VECH and Dynamic Conditional Correlation (DCC) are applied and 25 GCM predictors were selected for a bivariate temperature heteroscedastic modeling. It is observed that the conditional covariance between predictors and temperature is not very strong and mostly depends on the interaction between the random process governing temporal variation of predictors and predictants. The DCC model reveals a time varying conditional correlation between GCM predictors and temperature time series. No remarkable increasing or decreasing change is observed for correlation coefficients between GCM predictors and observed temperature during 1980-2000 while weak winter-summer seasonality is clear for both conditional covariance and correlation. Furthermore, the stationarity and nonlinearity KPSS and BDS tests showed that GCM predictors, temperature and their conditional correlation time series are nonlinear but stationary during 1980-2000 according to BDS and KPSS test results. However, the degree of nonlinearity of temperature time series is higher than most of the GCM predictors.

Keywords: Multivariate GARCH, temperature, diagonal VECH, DCC, conditional covariance, conditional correlation, nonlinearity, stationarity,

1. Introduction

Downscaling techniques have been developed for years to generate temporal and spatial variability of local scale climate information from large scale atmospheric variables simulated by Atmosphere-Ocean Global Climate Models (AOGCMs). Downscaling large-scale GCM outputs to a finer spatial resolution are carried out through two fundamental approaches called dynamic and statistical approaches. While dynamic downscaling refers to the use of Regional Climate Models (RCMs), Statistical downscaling methods (SDM) seek the relationship between variables well simulated by Global Climate Models (GCMs, usually large-scale fields) and regional or local surface climate variables. In other words, SDMs assume that there is an empirical relationship between upper-air circulation variables or predictors and local climate variables or predictants. In contrast to dynamical downscaling methods, SDMs need lower computation requirements, are easier for implementation, are able to provide point climate observations (Wilby et al., 2002, Herrera et al., 2006), can be easily transferred to other regions and are able to directly incorporate observations into method. On the other hand, these methods usually require long and reliable observation data and are dependent upon choice of predictors (Fowler et al., 2007).

Generally speaking, SDMs may be divided into three main approaches called transfer functions, stochastic weather generators and weather typing approaches. The most common and relatively simpler transfer function methods which derive an empirical relationship between predictors and predictants include linear models such as regression-based models (Hessami et al., 2008, Fasbender and Ouarda, 2010; Jeong et al., 2012b, Hammami et al., 2012), Principle Component Analysis (Huth, 2004), Canonical Correlation Analysis (Busuioc et al., 2006) and Singular Value

Decomposition (Oshima et al., 2002). These methods have an important drawback, the assumption of a linear association between predictors and predictants, which is usually violated as a nonlinear mechanism governs the climate system. However, the nonlinearity in this association between climate variables using nonlinear methods has not been fully investigated in the literature. A few nonlinear transfer function methods have been applied for downscaling, among which Artificial Neural Networks are the most common approach. For example, the superiority of nonlinear ANNs against multiple linear regression was demonstrated by Miksovsky and Raidl (2005) who used NCEP/NCAR reanalysis data for temperature downscaling. They showed a strong detectable non-linearity for winter season whereas the relationship between predictants and predictors were reported rather linear for summer. The ANNs were also applied by Schoof and Pryor (2001) and Cannon (2007) for precipitation and temperature downscaling. Another example of application non-linear methods for downscaling was presented by Vrac et al., (2007) who used Generalized Additive Models for precipitation and temperature downscaling in Western Europe. Fischer et al., (2004) introduced a multivariate non-linear model called Multivariate adaptive Regression Splines (MARS) for rainfall anomaly induced by El Nino/Southern Oscillation (ENSO).

All the above methods have applied non-linear approaches to capture the non-linear associations between large atmospheric predictors and regional or local predictants. This study aims to have a new look at the nonlinear association between predictors and predictants through their second order moment or their variance by introducing a class of multivariate nonlinear methods commonly used in financial time series modeling and known as the Multivariate Autoregressive Generalized Conditional Heteroscedasticity (MGARCH) models. The most obvious application

of MGARCH models is the study of the conditional covariance structure between different markets.

Although time series modeling approaches such as univariate and multivariate autoregressive moving average (ARMA) models are very popular in hydrologic and climatic sciences, they have rarely applied for downscaling (e.g. Laux et al. 2011). Traditional time series approaches assume that the expected value or the mean of a process at each time step, t, is conditioned on the previous time steps, $t\text{-}1$, …, $t\text{-}k$, where k is the lag time. This expected value or the mean is called the conditional mean. In a GARCH approach, it is assumed that the variance of a process at each time step, t, depends on the previous time steps. This variance which varies in time is called conditional variance or heteroscedasticty. As mentioned above, the multivariate GARCH approach has been developed to establish the relationship or the co-movement of the variance of two processes or the conditional covariance. This study tries to look at the conditional variance-covariance structure between Global Climate Model (GCM) outputs and the local or regional predictants which can be called as Heteresocsedastic Downscaling. To illustrate this new perspective, the MGARCH models are first introduced in the next section. The methods are followed by an illustrative example for heteroscedastic modeling. The conclusions and challenges for future work are given afterwards.

2. Heteroscedastic procedures

2.1. Univariate model

We begin with the univariate Autoregressive Generalized Conditional Heteroscedasticity ,GARCH(v,m) model proposed by Engle (1982) and generalized by Bollerslev (1986). Here, the term Autoregressive implies that the variance of a variable at a time t depends on the variance at

time *t-1*, *t-2*, …,*t-m*. In other words, we have conditional heteroscedasticity or time varying variance. In this case, GARCH model is written as follows

$$\sigma_t^2 = \omega + \sum_{i=1}^{v} \alpha_i \varepsilon_{t-v}^2 + \sum_{j=1}^{m} \beta_j \sigma_{t-m}^2 \qquad (1)$$

Here σ_t^2 is the conditional variance of the variable at time *t*, ω, $\alpha\ (i = 0, \dots, v)$ and $\beta (j = 1, \dots, m)$ are the parameters of the model to be estimated, σ_{t-m}^2 is the conditional variance at time *t-m*, ε_{t-v}^2 is the noise or the shock of the process at time *t-v* and v and m are the order of the model, or the number of parameters in the model. For example, a GARCH(1,1) model has one ALPHA or ARCH parameter and one BETA or GARCH parameter.

The parameters of the model should be positive to ensure that the conditional variance is always positive and $\alpha + \beta < 1$; where $\alpha + \beta$ captures the persistence of volatility or the persistence of conditional variance. It should be noted that α, and β indicate the degree of a short-term and long-term persistence, respectively. The intensity of persistency is therefore equal to $\alpha + \beta$.

2.2. Multivariate model

2.2.1. Overview

The Multivariate GARCH model is the extension of the univariate model to the bivariate form where the heteroscedasticity of one process is assumed to transfer to the other process and increase or decrease its volatility.

In a multivariate model, the conditional variance, σ_t^2 depends on the lagged shocks $\boldsymbol{\varepsilon_{t-i}}$, $i = 1, \dots, v$ and on the lagged conditional covariance matrix, $\boldsymbol{\sigma_{t-j}}$, $j = 1, \dots, m$. Therefore, the general form of a bivariate MGARCH (1,1) model with two series ($k = 2$) can be given as follows

$$VECH(\boldsymbol{\sigma_t^2}) = \boldsymbol{\omega} + \sum_{i=1}^{V} \boldsymbol{A_1} VECH(\boldsymbol{\varepsilon_{t-1}} \acute{\boldsymbol{\varepsilon}}_{t-1}) + \sum_{j=1}^{M} \boldsymbol{B_1} VECH(\boldsymbol{\sigma_{t-1}^2}) \qquad (2)$$

Where $\boldsymbol{\omega}$ is a $\frac{1}{2}k(k+1)\times 1$ vector and $\boldsymbol{A_1}$ and $\boldsymbol{B_1}$ are $\left(\frac{1}{2}k(k+1)\times\frac{1}{2}k(k+1)\right)$ parameter matrices. The VECH(.) shows the operator which stacks the lower portion of a matrix in a vector. The above bivariate model can easily be written in the following matrix format

$$VECH\begin{bmatrix}\sigma^2_{11,t} & \sigma^2_{12,t}\\ \sigma^2_{21,t} & \sigma^2_{22,t}\end{bmatrix}=\begin{bmatrix}\sigma^2_{11,t}\\ \sigma^2_{12,t}\\ \sigma^2_{22,t}\end{bmatrix}=$$

$$\begin{bmatrix}\omega_{10}\\ \omega_{20}\\ \omega_{30}\end{bmatrix}+\begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix}\begin{bmatrix}\varepsilon^2_{1,t-1}\\ \varepsilon_{1,t-1}\varepsilon_{2,t-1}\\ \varepsilon^2_{2,t-1}\end{bmatrix}+\begin{bmatrix}b_{11} & b_{12} & b_{13}\\ b_{21} & b_{22} & b_{23}\\ b_{31} & b_{32} & b_{33}\end{bmatrix}\begin{bmatrix}\sigma^2_{11,t-1}\\ \sigma^2_{12,t-1}\\ \sigma^2_{22,t-1}\end{bmatrix} \qquad (3)$$

Where $\sigma^2_{11,t}$ and $\sigma^2_{22,t}$ are the conditional variance and $\sigma^2_{12,t}$ is the conditional covariance between the two time series (predictor and predictant time series in our case).

The above bivariate MGARCH specification is called the full bivariate VECH(1,1) model which has 3+9+9=21 parameters to be estimated. To reduce the number of parameters of the model, some specifications have been introduced from which, two specifications, the diagonal VECH model and Dynamic Conditional Correlation (DCC) models, are introduced and applied here in this study. The diagonal VECH model estimates the time varying conditional covariance between CGCMs and temperature time series while the DCC model gives the time varying correlation between them.

2.2.2. Diagonal VECH model

The diagonal VECH model was introduced by Bollerslev et al. (1988) to reduce the number of parameters of the full VECH model. The diagonal model constrains the matrices $\boldsymbol{A_1}$ and $\boldsymbol{B_1}$ to

be diagonal. In this case, the number of parameters of the bivariate MGARCH(1,1) reduces to $3(k(k+1)/2)$ or 9 parameters which is relatively less than the full model. The diagonal VECH model can therefore be written as follows

$$dVECH\begin{bmatrix}\sigma_{11,t}^2 & \\ \sigma_{21,t}^2 & \sigma_{22,t}^2\end{bmatrix} =$$

$$\begin{bmatrix}\omega_{11} & \\ \omega_{21} & \omega_{22}\end{bmatrix} + \begin{bmatrix}a_{11} & \\ a_{12} & a_{22}\end{bmatrix}\begin{bmatrix}\varepsilon_{1,t-1}^2 & \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2\end{bmatrix} + \begin{bmatrix}b_{11} & \\ b_{21} & b_{22}\end{bmatrix}\begin{bmatrix}\sigma_{11,t-1}^2 & \\ \sigma_{12,t-1}^2 & \sigma_{22,t-1}^2\end{bmatrix} \quad (4)$$

Using the Hadamard or element-by-element product of the above matrix, the conditional variance-covariance structure between the two time series can be as follows

$$\sigma_{11,t}^2 = \omega_{11} + a_{11}\varepsilon_{1,t-1}^2 + b_{11}\,\sigma_{11,t-1}^2 \quad (5)$$

$$\sigma_{21,t}^2 = \omega_{21} + a_{21}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + b_{21}\sigma_{21,t-1} \quad (6)$$

$$\sigma_{22,t}^2 = \omega_{22} + b_{22}\varepsilon_{2,t-1}^2 + b_{22}\sigma_{22,t-1}^2 \quad (7)$$

Where $\sigma_{11,t}^2$ and $\sigma_{22,t}^2$ denote the conditional variance of the predictor and predictant and $\sigma_{12,t}$ is the conditional covariance between them, respectively.

2.2.3. Dynamic Conditional Correlation

The dynamic conditional correlation (DCC) model was introduced by Engle (2002) to generalize the constant conditional correlation (CCC) model (Bollerslev, 1990). The CCC model is a specification for the general MGARCH model which aims to reduce the number of parameters of the model and assumes that the conditional correlations between the elements of $\boldsymbol{\varepsilon}_t$ are time-invariant. This implies that the conditional covariance between two time series is given by

$$\sigma_{ijt}^2 = D_t R_{ij} D_t \quad (8)$$

Where

$$D_t = diag(\sigma_{11t}^2 \dots \sigma_{22t}^2) \quad (9)$$

and R_{ij} is the constant conditional correlation between the conditional variances of the two time series (σ_{11t}^2 and σ_{22t}^2, equations 5 and 7). The CCC model implies that the conditional covariance is proportional to the conditional variances of the two time series. Therefore, equation (8) can be re-written in the following form

$$\sigma_{12,t}^2 = R_{12} \times \sigma_{11t}^2 \times \sigma_{22t}^2 \quad (10)$$

The main drawback of the CCC model is the assumption of invariant conditional correlation which seems unrealistic for many hydro-climatic variables, especially in the context of climate science and climate change. Therefore, we need to apply a model which allows R to be time varying. For estimating time varying R, Engle (2002) introduced the DCC model. The conditional correlation in a DCC(1,1) model can be written as follows

$$R_t = diag(\sigma_{11,t}^{\frac{1}{2}}, \dots, \sigma_{kk,t}^{\frac{1}{2}}) Q_t diag(\sigma_{11,t}^{\frac{1}{2}}, \dots, \sigma_{kk,t}^{\frac{1}{2}}) \quad (11)$$

Where the $k \times k$ symmetric positive definite matrix $Q_t = (\sigma_{ij,t})$ is given by

$$Q_t = (1 - \theta_1 - \theta_2)\bar{Q} + \theta_1 \varepsilon_{t-1} \acute{\varepsilon}_{t-1} + \theta_2 Q_{t-1} \quad (12)$$

$\bar{Q}$ is the $k \times k$ unconditional variance matrix of ε_t and θ_1 and θ_2 are non-negative parameters satisfying $\theta_1 + \theta_2 < 1$ which allows to capture the effects of previous shocks and dynamic conditional correlations on the current dynamic conditional correlation. It is obvious that when $\theta_1 + \theta_2 = 0$ the DCC is equivalent to a CCC model.

3. Test methods

An important issue in the failure of linear models for an accurate downscaling is the nonlinearity and non-stationarity in the climate variables. Although numerous studies have applied the nonlinear methods for downscaling (as some of them were mentioned above), there are very few studies to test the nonlinearity (e.g. Miksovsky and Raidl, 2006 by the method of surrogate data) and non-stationarity of GCM predictors and their link to surface climate variables.

This study applies the BDS test (Brock et al., 1996) to investigate the nonlinearity of GCM predictors and their link to surface temperature. The test has its origin in chaos theory and is useful to detect both deterministic chaos and test the goodness-of-fit of a model. It is a nonparametric method for testing for serial independence and nonlinear structure in the time series based on the correlation integral of the series. The BDS statistic for $m > 1$ is defined as

$$BDS_{m,M}(r) = \sqrt{M}\frac{C_m(r)-C_1^m(r)}{\sigma_{m,M}(r)} \qquad (13)$$

Where $M = N - (m-1)\tau$ is the number of embedded points in an m-dimensional space, r is radius of a sphere centered on X_i, $H(u)$ is the Heaviside function, $X_i = (x_t, x_{t-\tau}, x_{t-(m-1)\tau})$ is a new series generated from the main time series (x_t), $c_{m,M}(r)$ counts up the number of points in the m-dimensional space that lies within a hypercube of radius r and σ is the standard deviation of the points in this space. Under the null hypothesis, $\{X_t\}$ is an i.i.d process and the BDS statistic converges to a unit normal as $M \to \infty$. The BDS test in this study is applied to investigate and compare the degree of nonlinearity of different predictors and predictants and their conditional variance-covariance structure.

The KPSS test (Kwiatkowski et al.,1992) is also applied for investigating the stationarity. It should be noted that the KPSS test examines the stationarity around the deterministic trend (trend

stationarity) and the stationarity around a fixed level (level stationarity) and is more powerful than other methods such as the Augmented Dickey-Fuller (ADF) test (Lahrech and Sylwester (2011). Assume we can decompose $\{X_t\}$ into the sum of a deterministic trend, a random walk and a stationary error in the following regression model

$$x_t = r_t + \beta_t + \varepsilon_t \quad (14)$$

Where $r_t = r_{t-1} + u_t$ and u_t is i.i.d $\sim N(0, \sigma_h^2)$, β_t is a deterministic trend and ε_t is a stationarity error. By these assumptions, the null hypothesis of level stationarity cannot be rejected if $\beta = 0$ and alternatively, the null hypothesis of trend stationarity cannot be rejected if $\sigma_u^2 = 0$. The reader is referred to Wang (2006) for more details.

4. Illustrative example

4.1. Data description

To illustrate the above procedure for variance-covariance modeling, the following predictor and predictant data sets are applied. Daily maximum and minimum temperature data series (T_{max} and T_{min}, hereafter) from Bagotville Airport station ($71.00^{\circ}W, 48.33^{\circ}N$) in the Quebec province, Canada, for the period 1980-2000, are selected as predictants. The temperature data are obtained from Environment Canada, and have been homogenized by Vincent et al., (2002). The predictor data set includes the Canadian Coupled Global Climate Model (CGCM3 T47) model outputs. The CGCM3 model is an improved Canadian AOGCM models (i.e CGCM1 and CGCM2) with new components. As the proposed model considers a bivariate conditional variance-covariance structure between a predictor and a predictant, 25 predictors from one near grid point are selected to illustrate the MGARCH approach. The CGMs data set used in this study is based on the CGCM3 run at T47 resolution which provides a $3.75^{\circ} \times 3.75^{\circ}$ surface grid resolution and covers

the period 1961-2000. All the CGCM3 data sets can be found in the DAI CGCM3 predictors (2008) documentation. Table 1 shows the predictor variables of NCEP/NCAR interpolated on a CGCM3 T47 grid. The location of grid points and surface sites are also illustrated in Figure 1. It should be noted that the CGCM predictors have a range of weak and moderate (unconditional) correlation with the temperature time series according to Jeong et al. (2012a).

4.2. Data screening

Before developing MGARCH models for heteroscedastic analysis, we look at the temporal variation of the mean and unconditional variance of the predictors. Here, the unconditional variance refers to the squared temperature time series. The daily evolution of the mean and variance of the observed and squared (or the unconditional variance) temperature time series are given in Figure2. The daily average and variance of original and squared temperature have been calculated for each calendar day during 1980-2000.

It can be seen that the mean and variance of temperature (Figure 2a) show strong winter-summer seasonality. However, it is observed that the squared temperature (or the unconditional variance) time series shows a different temporal variation from the original temperature time series. This may imply that the temporal variation of the second order moment, or the variance, of temperature is different from the mean or the first order moment. This feature has not been investigated in previous studies.

Next, we examine the correlation between the original and squared predictors and predictant time series. Figure 3 indicates the monthly (Pearson product) correlation coefficients between some selected predictors and temperature data. This figure indicates different (monthly) correlation

structure between original (not squared) temperature and squared temperature (or the unconditional variance) and GCM predictors. This phenomenon was neither investigated on previous studies. For example, the correlation coefficient between temperature and MSLP moves completely in a different direction of the variation of the correlation between squared temperature and MSLP. This feature can also be seen for C_{10} predictors. It is observed that the correlation between squared (unconditional variance) temperature and predictors are relatively weaker than that for original (non-transformed) time series. One can also see that the difference in correlation coefficient is remarkable for winter and fall seasons while for the summer, there is not a significant divergence between the two correlation coefficients. All these suggest that the predictor-predictant variance relationship is different from the relationship between their means, which is considered in different statistical downscaling (SD) procedures.

Moreover, the stationarity and nonlinearity test results for selected time series are presented in Table 2. The results show a significant level stationarity for all predictors and predictants according to the KPSS test. The temperature time series indicate a significant trend non-stationarity during 1980-2000. Among the GCM predictors, H_5, U_{10}, X_8, Z_5 and Z_8 also indicate significant trend non-stationarity.

The BDS test reveals nonlinearity for all temperature and GCM predictors. The temperature time series shows a higher degree of nonlinearity than most of the GCM predictors, except for T_2, specific humidity and Z_5 at all dimensions. All other predictors indicate almost similar degree of nonlinearity for all m dimensions. We also do not observe a significant nonlinearity difference between different geopotential heights.

4.3. Univariate GARCH models

Before applying the bivariate GARCH models for heteroscedastic GCM predictors-temperature variance-covariance modeling, we look at the conditional variance of the GCM predictors and temperature time series. The parameters of the GARCH models for conditional variance are presented in Table 3. We can see that all ARCH parameters and most of the GARCH are significant while the GARCH parameters for a few GCMs are not significant.

Based on the GARCH models, the GCM predictors can be categorized into three groups. The first group includes the predictors with insignificant or very small GARCH parameters such as C_{10}, U_8, U_{10}, X_5 *and* Z_8 predictors. These predictors do not have a strong memory or persistence in the variance and conditional variance greatly depends on the previous shocks (random process or innovation). In other words, the short-term persistence is stronger than the long-term persistence for the first group. The second group includes the predictors with significant large GARCH parameters such as specific humidity (H_5, H_8, H_{10}) and T_2. These predictors have a high degree of conditional variance memory and persistency and small ARCH parameters (short-term persistence). The third group includes other GCM predictors which show relatively higher ARCH and GARCH parameters than the first group. The conditional variance in this group depends on both previous shocks and variances at (more or less) the same degree.

The conditional variances of some selected GCM predictors and temperature time series are given in Figure 4. This figure shows the daily average conditional variance during 1980-2000. Among the predictors, the temperature at 2-m height (T_2), specific humidity (H_8) and wind direction (X_8) show a remarkable different temporal (daily) variation. Other predictors show almost the same temporal variation with winter-summer seasonality in the conditional variance.

It is also clear that all conditional variances show a high level of fluctuation and do not tend to stay at a high (low) level for a long time before changing to another low (high) level.

4.4. Multivariate model for maximum temperature

4.4.1. VECH model

In order to consider the heteroscedastic model for maximum temperature (T_{max}), the bivariate diagonal VECH model is first developed and estimated. The parameters of the conditional covariance between predictors and maximum temperature are given in Table 4. The covariance parameters show a relatively low level of the relationship between the variance of the predictors and predictants. Except for specific humidity and T_2 which show a large positive GARCH parameter (b_{21}), other predictants have relatively small or negative GARCH parameters. Moreover, the ARCH parameters are also relatively small and vary between $a_{21} = 0.15$ for H_5 to $a_{21} = 0.60$ for U_5. A negative GARCH parameter is also observed for some of the GCM-T_{max} covariance relationship. The covariance structure between GCM predictors and maximum temperature, therefore, seem to greatly depend on the cross products of shocks, signifying that shocks in the predictors volatility impact on future volatility of T_{max}. The variance of T_{max} does not seem to be largely influenced by the past variances of (most of) the GCM predictors and most the GCM predictors show a negative influence on future variance of the surface T_{max} at the selected station.

Therefore, based on the covariance structure, the GCM predictors can be divided into three groups; the first group with a high covariance relationship includes specific humidity (H_5, H_8, H_{10}) and T_2; the second group with both short-term and long-term persistence with positive covariance relationship with T_{max} including wind speed (S_5, S_8, S_{10}), U_5, X_8 and

geopotential height ($Z_5, Z_8,$); and the third group with the same short-term persistence influence but negative covariance effect on T_{max} including the rest of GCM predictors. The time varying covariance between GCM predictors and maximum temperature are illustrated in Figure 5 for some typical GCM predictors. This figure shows the average, maximum and minimum conditional covariance for Julian days during 1980-2000. The conditional covariance between GCM predictors and T_{max} shows winter-summer seasonality for most GCM predictors such as those illustrated in Figure 5 except for the specific humidity and T_2 which show a different temporal covariance variation and seasonality.

4.4.2. DCC model

We next move to estimate conditional correlation between predictors and maximum temperature using the DCC model. The parameters of the DCC model are given in Table 5. The estimates of θ_1 are statistically significant for all GCM predictors while the estimates of θ_2 are not significant for some of the GCM predictors-T_{max} relationship. The DCC results indicate that the assumption of constant conditional correlation (CCC) between predictors and predictants are not supported empirically. The effect of short-term persistence of shocks on dynamic conditional correlations is greatest for U_5, MSLP and Z_8 at 0.65, 0.64 and 0.62, respectively. These results shows that the short-term variation of maximum temperature is greatly dependent on U_5, MSLP and Z_8.

The largest long-term influence of shocks on the conditional correlation is observed for V_{10}, specific humidity and T_2. In addition some other GCM predictors such as geopotential height and wind direction have also long-term influence on maximum temperature.

Using the DCC model, the daily conditional correlation coefficient can be estimated for the whole 1980-2000 period. In order to illustrate the daily fluctuation of conditional correlation, the average, maximum and minimum estimates of the correlation coefficients for each Julian day are given in Figure 6 for some typical GCM predictors. It can be seen that (average) correlation coefficients show a weak association between most of the GCM predictors such as MSLP, C_{10} , D_5, V_5, S_8 and T_{max} and remain in the same level during the year. Some other GCM predictors such as specific humidity, T_2, geopotential height and wind direction have large correlation with a dynamic seasonal fluctuation.

Having daily time series of conditional correlation, the annual variation of conditional correlation coefficients during 1980-2000 is also illustrated in boxplot format in Figure 7 for some GCM predictors. In this figure, each boxplot contains 365 daily conditional correlation coefficients (366 for leap years). These predictors (and others which are not illustrated in this figure) reveal no remarkable change in the association of GCM predictors and maximum temperature during the 20-year study period at the selected station. In other words, this figure shows no trend or significant periodicity in the association between GCM predictors and maximum temperature.

4.4.3. Testing conditional correlation

Finally, we look at the stationarity and nonlinearity test results for the correlation coefficients between GCM predictors and temperature in Table 6. The KPSS test shows level stationarity for correlation coefficients between all predictors and maximum temperature and a trend stationarity for most of the GCM predictors. However, strong (1% significant level) trend non-stationarity for the correlation between T_{max} and some GCM predictors such as T_2, specific humidity, U_5, X_5

and Z_5 and some weak (5% significant level) trend non-stationarity for C_{10}, D_{10}, S_5, S_{10} are observed. This suggests the existence of outlier correlation coefficients which make the whole correlation coefficient time series not stationarity around such a basic trend line.

The BDS test exhibits a significant nonlinear structure for conditional correlation between the predictants and maximum temperature. It is observed that the degree of nonlinearity in the correlation coefficients is almost identical among most of GCM predictors. However, the correlation coefficients between some GCM predictors such as T_2, C_{10}, specific humidity, V_{10}, X_5, X_{10} and Z_{10} and maximum temperature are showing a much larger nonlinearity than those for other GCM predictors. It can be seen that these GCM predictors with high nonlinearity show also the trend non-stationairty according to KPSS test.

4.5. Multivariate model for minimum temperature

4.5.1. VECH model

In this section, the bivariate GARCH model for minimum temperature heteroscedastic analysis is developed. Following diagonal VECH procedure, the estimated models are given in Table 7. The GARCH model for minimum reveals a relatively high degree of short-term persistency and low level of memory in conditional covariance for minimum temperature as the ARCH parameter are usually larger than the GARCH parameter. Both negative and positive GARCH parameters are observed. The intensity of persistency, however, is low for most of the GCM predictors. This suggests a high degree of volatility variation in minimum temperature and less remembering the past variances. It is important to note that the short-term and long-term persistencies are almost similar to what were observed for maximum temperature (see Table 4).

The conditional covariance between some typical GCM predictors and T_{min} are illustrated in Figure 8. This figure (and the other covariances not given in this figure) shows different temporal (daily and seasonal) covariance structure between GCM predictors and minimum temperature. It seems that there is not a big difference for temporal variation of the conditional covariance between T_{max} and T_{min} and most effective GCM predictors, i.e T2 and specific humidity (see also Figure 5). Similar to conditional covariance for T_{max}, a winter-summer difference is also observed for the covariance structure between GCM predictorss and T_{min}.

4.5.2. DCC model

Next step in the heteroscedastic analysis for GCM predictors-minimum temperature relationship includes the estimation of dynamic conditional correlation between predictors and T_{min}. The parameters of DCC estimations and the constant conditional correlations (CCC) are given in Table 8. For the DCC models, the strength of the persistency and the parameters are almost the same as those estimated for T_{max} (see Table 5). The DCC parameters indicate that the assumption of constant conditional correlation is violated as the parameters θ_1 and θ_2 are significant. However, for most of the GCM predictors, θ_2 is not significant or is relatively small. This suggests that the correlation between T_{min} and GCM predictors changes in time but it depends largely on the cross-products of the shocks and the covariance has an insignificant and small dependence on previous covariances. This table also shows that the constant conditional correlations are negative and relatively small for most of the GCM predictors.

The largest long-term influence of shocks to the conditional correlation is observed for T_2 and H_5, X_5 and X_8. In addition some other GCM predictors such as geopotential height, wind direction, vorticity and V component have long-term influence on minimum temperature.

Using the DCC model, the daily time varying conditional correlations between GCM predictors and minimum temperature are given in Figure 9 for some typical GCM predictors.

This figure shows the average DCC remains around a fixed value for most of the GCM predictors but the daily maximum and minimum observed DCC for 1980-2000 shows that a high level of association between predictors and minimum temperature also exists. These conditional correlations usually show winter-summer seasonality. However, a few GCM predictorss such as T_2, specific humidity, wind direction and geopotential height have a different temporal (seasonality) association to minimum temperature.

Figure 10 illustrates the annual evaluation of conditional correlation coefficients between some typical GCM predictors and minimum temperature. Similar to figure 7, each boxplot contains 365 daily conditional correlation coefficients (366 for leap years). The annual evaluation of conditional correlation does not show any significant change (trend) or significant regular fluctuations during 1980-2000. However, the extreme (outlier) coefficients are observed for some GCM predictors such as MSLP, Z_8, etc.

4.5.3. Testing conditional correlation

The stationarity and nonlinearity test results for the conditional correlation are reported in Table 9. This table demonstrates level-stationarity for all predictors-predictants associations. However, KPSS test cannot pass the null hypothesis of trend-stationarity for the conditional correlation coefficient between T_{min} and some GCM predictors (i.e. H_5, H_{10}, T_2, U_5, X_5 and X_{10}).

This table also tells the existence of nonlinarity in the association of all GCM predictors and minimum temperature according to BDS test results. Similar to the results for Tmax, one can see that the degree of nonlinearity is, more or less, identical for all GCM predictors except some (i.e.

T_2, X_5,X_{10} and Z_5) which show a higher degree of nonlinearity for all *m=2,..,4* dimensions. It is also observed that a few GCM predictors with trend non-stationarity (i.e. T_2, X_5,X_{10}) show a higher degree of nonlinearity than other GCM predictors. This is in contrast to what is observed for T_{max} (see Table 6) for which the GCM predictors with non-stationarity also indicated a higher degree of nonlinearity.

5. Summary and Conclusions

In the state-of-the-art of statistical downscaling, the first order moment or the mean of large atmospheric variables are transferred to the first order moment of the surface variables such as temperature and rainfall. Although in the downscaling literature, the (unconditional) correlation between the predictors and surface observations are always reported, the link between their second order moments (or variance) is not examined. This study showed that the association between the variances of the GCM predictors and temperature is different (both in magnitude and temporal variation) from the association in their mean which is applied in a downscaling. For modeling the relationship between the second order moment of predictors and predictants, this study developed and applied an econometric approach. Modeling the conditional variance-covariance structure showed that except a few GCM predictors such as specific humidity and T_2 with strong variance persistency, other GCM predictors showed both short-term and long-term variance persistency. The conditional variance of the predictors has relatively large ARCH parameter and small GARCH parameters. This suggests the volatility of most of the GCM predictors is dynamic and the memory in the variance is usually weak. The conditional variances of temperature time series are also the same as most of the GCM predictors and show both relative dynamic volatility and memory in the variance.

The proposed multivariate GARCH procedures including the diagonal VECH and dynamic conditional correlation (DCC) specifications show that the conditional covariance structure between the GCM predictors and surface temperature is not very strong and does not have a strong persistence for most of the predictors.

The diagonal VECH model showed a short-term covariance or comovement in predictor-predictant relationship which is largely dependent on the cross-product of the shocks of the GCM predictors and temperature (maximum and minimum) time series. The GARCH parameters, on the other hand, reveal a very weak degree of the covariance structure for GCM predictors-temperature relationship. Both position and negative GARCH parameters are observed in this study.

Furthermore, dynamic conditional correlation (DCC) showed that the constant conditional correlation (CCC) between GCM predictors and temperature is unrealistic. The same as the VECH model, the DCC model also reveals that the co-volatility structure between GCM predictor and temperature is not strong and the correlation depends strongly on cross products of shocks rather than covariance between them. The time varying conditional correlation estimated by the DCC procedure showed no significant change in the association between predictors and temperature during 1980-2000 while a weak seasonality is observed for the conditional correlation coefficients between some of the GCM predictors and temperature data.

This study also investigated the stationarity and nonlinearity embedding in the predictors and predictants and their relationship applying two popular KPSS and BDS tests. The results showed level stationarity in all predictors and predictants time series during 1980-2000. However, it was observed that temperature time series and some GCM predictors are illustrating a significant non-stationarity around a deterministic trend which can usually be referred to extreme events. It

seems that the probability of extreme event occurrence is tending to increase from 1980 to 2000. However, this needs to be investigated by future studies.

Furthermore, the nonlinearity tests indicated that all predictors and predictants are exhibiting nonlinearity but the temperature and a few GCM predictors have an exceptional larger nonlinearity than other GCM predictors. These GCM predictors have usually more influence of conditional mean and variance of temperature. It may suggest that the nonlinearity and non-stationarity of the temperature time series are originating from the non-linear and non-stationarity variation of these GCM predictors.

The stationarity test also revealed no change in the association (correlation) between GCM predictors and surface temperature during 1980-2000, though a trend non-stationarity is observed for the correlation coefficients between some GCM predictors and temperature time series. The correlation coefficients also reveal non-linearity while the highest degree of non-linearity is observed for the correlation between temperature and those non-stationarity GCM predictors. All these tests show that the non-linearity and non-stationarity of temperature time series may largely depend on some particular GCM predictors such as specific humidity, wind direction, geopotential height and T_2 which show both non-stationarity and nonlinearity.

Finally, it has to be noted that the proposed method seems to have some advantages and disadvantages. The first advantage of MGARCH approach is modeling the variance-covariance relationship between the GCM predictors and surface variables which are not considered in other models, especially those currently used for downscaling. It is believed that the predictor-predicatnt relationship is not linear and nonlinear methods should be applied for downscaling to describe this relationship. The MGARCH models bring this opportunity to downscale the variance of GCM predictors to the variance of the local climate variables.

Another major advantage of the MGARCH model is that it can estimate the time varying association (covariance and correlation) between predictors and predictants. In other words, the method allows us having a step-by-step variation of this association. For example in our study, a day-to-day (conditional) correlation variation was obtained. This advantage would be interesting for two reasons; First, for the selection of the GCM predictors for downscaling purposes and second for investigating the time varying change in the correlation between predictors and predictants in the context of climate change. The results of our case study revealed no significant change for the correlation coefficients between GCM predictors and local (maximum and minimum) temperature at the selected surface station during 1980-2000. However, some extreme correlation coefficients are observed in this relationship. For both maximum and minimum temperature time series, no trend is observed for conditional correlations during 1980-2000. On the other hand, the monthly variation in conditional correlation indicated a weak seasonality for almost all cases.

However, the MGARCH models have two main disadvantages. The first global disadvantage is the number of parameters which increases rapidly with increases of the order and the variables included in the model. Though the diagonal VECH model reduces the number of parameters, it seems that the number of parameters is still relatively high and the parameter estimation may be a real problem.

Another specific disadvantage of MGARCH models for downscaling is that although these models can be developed and applied for more than two variables, the conditional covariance can only be estimated between "one" GCM input and "one" output variable. In other words, dissimilar to multivariate regression model or other multivariate models, the MGARCH models give the conditional covariance between two variables. Therefore, it is not possible to get the

effect of a couple of variables (i.e GCM predictors) on a climate variable (e.g temperature or rainfall) in one single model, such as other typical downscaling methods (i.e. multivariate regression and canonical correlation analysis).

6. Recommendations for future work

The proposed method in this study illustrates an example of the variance-covariance link between Global Climate Models (GCM predictors) and maximum and minimum temperature. It is interesting to investigate time varying variances of other GCM predictors with different spatial resolutions for future studies.

The GCM predictors of this study cover the period 1980-2000. It would be necessary and interesting to investigate the conditional variance of predictors covering future period (e.g. 2001-2100) in order to explore future heteroscedastic change in GCM predictors. As this study shows no significant change in the correlation coefficients between GCM predictors and temperature, it would be interesting to estimate and examine the correlation change during last decades of 20th century for other GCM predictors and RCMs not included in this study. Another interesting topic is to test the spatial variation of the conditional variance of GCM predictors and spatial variation of the covariance between GCM predictors and surface climate variables. As a final point, the MGARCH models are strongly suggested to apply for variance-covariance modeling between GCM predictors and precipitation time series. Finally, it is important to investigate the physical basis of the conditional variance-covariance structure between different predictors and predictant in future studies.

Acknowledgement

The financial support of this study from the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Canada Research Chair (CRC) Program is highly acknowledged.

References

Bollerslev, T., 1986: Generalized Autoregressive Conditional Heteroscedasticity, J. Econometrics., **31**, 307–327.

Bollerslev, T., R.F. Engle, and J.M. Wooldridge, 1988: A capital asset pricing model with time varying covariances, *J. Pol. Econ.*, **96**, 116–31

Bollerslev, T., 1990: Modeling the coherence in short-run nominal exchanges rates: a multivariate generalized ARCH model, *Rev. Econom. Stat.*, **72**, 498-505.

Brock, W.A., W. D. Dechert, J. A. Scheinkman, and B. LeBaron (1996), A test for independence based on the correlation dimension, *Econ. Rev,. 15 (3),* 197–235.

Busuioc, A., F. Giorgi, X. Bi, M. Ionita, 2006: Comparison of regional climate model and statistical downscaling simulations of different winter precipitation change scenarios over Romania. *Theor. Appl. Clim.*, **86**,101-123

Cannon A. J., 2007, Nonlinear analog predictor analysis: A coupled neural network/analogmodel for climate downscaling. *Neural networks*, **20**,444-453.

DAI CGCM3 Predictors (2008) Sets of predictor variables derived from CGCM3 T47 and NCEP/NCAR reanalysis, version 1.1 April 2008, Montreal, 15 pp.

Engle, R. F., 2002: dynamic Conditional Correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticty models. *J. buss. Economy. Stat.*, **20**, 339-350

Engle, R. F., 1982: Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation, *Econometrica,* **50**, 987-1008

Fasbender, D., Ouarda, T.B.M.J. (2010). Spatial Bayesian Model For statistical downscaling of AOGCM to minimum and maximum daily temperatures. J. Climate, 23, 5222-5242 DOI: 10.1175/2010JCLI3415.1

Fischer, M., Dewitte, B., and Maitrepierre L., A non-linear statistical downscaling model: El Nino/ Southern Oscillation impact on precipitation over New Caledonia. Geophys. Res. Lett., 31, L16204, doi:10.1029/2004GL020112, 2004

Fowler, H. J., Blenkisop, S., and Tebaldi C. 2007. Review: Linking climate change modeling to impact studies: recent advances in downscaling techniques for hydrological modeling. *Int. J. Clim.*, **27**, 1547-1578

Hammami, D., Lee, T.S., Ouarda, T.B.M.J., and Lee, J. (Predictor selection for downscaling GCM data with LASSO). J. Geophys. Res. Atmosph., 117, DOI: 10.1029/2012JD017864

Herrera, E., Ouarda. T.B.M.J., and Bobée, B. 2006. Méthodes de désagrégation appliquées aux Modèles du Climat Global Atmosphère-Océan (MCGAO). Revue des sciences de l'eau, 19, 297-312.

Hessami, M., P. Gachon, T.B.M.J. Ouarda, A. St-Hilaire, 2008: Automated regression-based statistical downscaling tool. *Environ. Model. Soft*, **23**, 813-834

Huth, R. 2004. Sensitivity of local daily temperature change estimates to the selection of downscaling models and predictors. J. Clim., 17, 649-652

Jeong, D.I., A. St-Hilaire, T.B.M.J. Ouarda, P. Gachon, 2012a. CGCM3 predictors used for daily temperature and precipitation downscaling in Southern Quebec, Canada. *Theor. Appl. Clim.*, 107,389-406

Jeong, D.I., A. St-Hilaire, T. B. M. J. Ouarda and P. Gachon, 2012b. Multisite statistical downscaling model for daily precipitation combined by multivariate multiple linear regression and stochastic weather generator. *Clim. Change* (In press)

Kwiatkowski, D., P.C.B. Phillips, , P. Schmidt and Y. Shin, 1992: Testing the null of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *J. Econometrics* **54**, 159–178.

Lahrech, A., and K. Sylwester, 2011: U.S. and Latin American stock market linkages. J. Int. Money and Fin., **30**, 1341-1357

Laux, P., S. Vogl, W. Qiu, H. R. Knoche, and H. Kunstmann, 2011: Copula- based statistical re_nement of precipitation in RCM simulations over complex terrain. Hydrol. Earth Syst. Sci., 15, 2401-2419.

Miksovsky, J., and A. Raidl, 2005: Testing the performance of three nonlinear methods of time series analysis for prediction and downscaling of European daily temperatures. *Nonlinear process. Geophys.*, **12**,979-991

Miksovsky, J., and A. Raidl, 2006: Testing for nonlinearity in European climatic time series by the method of surrogate data. *Theor. Appl. Clim.,* 83, 21–33

Oshima, N., H. Kato, S. Kadokura, 2002: An application of statistical downscaling to estimate surface air temperature in Japan. J. geophys. Res. Atm., 107, DOI: 10.1029/2001JD000762

Schoof, J.T., S.C. Pryor, 2001, Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks. Int. J. Clim., **7**, 773-790

Vincent LA, Zhang X, Bonsal BR, Hogg WD (2002) Homogenization of daily temperatures over Canada. *J Clim.*, 15:1322–1334

Vrac, M.,P. Marbaix,D. Paillard, and P. Naveau, 2007: Non-linear statistical downscaling of present and LGM precipitation and temperature over Europe. *Climt., Past.*, **4**, 669-682

Wang, W., 2006. Stochasticity, Nonlinearity and Forecasting of Streamflow Processes, Deft University Press, Amsterdam, 2006, ISBN 1-58603-621-1

Wilby, R. L., C. W. Dawson, and E. M. Barrow, 2002: SDSM - a decision support tool for the assessment of regional climate change impacts. *Environ. Model. Soft.*, **17**, 147-159

List of TABLES

FIGURE captions

Figure 1. Location of CGCM3 grid points, selected grid point and surface observation site for this study

Figure 2 Evaluation of the mean and (unconditional) variance for a) original and b) squared temperature time series

Figure 3 seasonal (unconditional) correlation coefficients between original and squared predictors and temperature data

Figure 4 Daily (averaged) conditional variance for some CGCM predictors and temperature time series (1980-2000)

Figure 5 Daily average, maximum and minimum conditional covariance between CGCM predictors and T_{max} time series during 1980-2000

Figure 6 Daily average, maximum and minimum conditional correlation coefficients between typical CGCM predictors and T_{max} during 1980-2000

Figure 7 Annual variation of conditional correlation between T_{max} and some typical CGCM predictors during 1980-2000

Figure 8 Daily average, maximum and minimum conditional covariance between typical CGCM predictors and T_{min} time series during 1980-2000

Figure 9 Daily average, maximum and minimum conditional correlation coefficients between typical CGCM predictors and T_{min} during 1980-2000

Figure 10 Annual and seasonal conditional correlation between T_{min} and CGCM predictors during 1980-2000

Table 1 NCEP and CGCM3 predictors on the selected grid point

Number	NCEP/NCAR predictors (unit)	abbreviation
1	500 hPa vorticity (s^{-1})	C_5
2	850 hPa vorticity (s^{-1})	C_8
3	1000 hPa vorticity (s^{-1})	C_{10}
4	500 hPa divergence (s^{-1})	D_5
5	850 hPa divergence (s^{-1})	D_8
6	1000 hPa divergence (s^{-1})	D_{10}
7	500 hPa specific humidity (kg/kg)	H_5
8	850 hPa specific humidity (kg/kg)	H_8
9	1000 hPa specific humidity (kg/kg)	H_{10}
10	Mean Sea Level Pressure (Pa)	MSLP
11	500 hPa wind speed (m/s)	S_5
12	850 hPa wind speed (m/s)	S_8
13	1000 hPa wind speed (m/s)	S_{10}
14	Temperature at 2 m (°C)	T_2
15	500 hPa U component (m/s)	U_5
16	850 hPa U component (m/s)	U_8
17	1000 hPa U component (m/s)	U_{10}
18	500 hPa V component (m/s)	V_5
19	850 hPa V component (m/s)	V_8
20	1000 hPa V component (m/s)	V_{10}
21	500 hPa wind direction (m/s)	X_5
22	850 hPa wind direction (m/s)	X_8
23	1000 hPa wind direction (m/s)	X_{10}
24	500 hPa geopotential height (m^2/s^2)	Z_5
25	850 hPa geopotential height (m^2/s^2)	Z_8

Table 2 Stationary and nonlinearity test results for predictors and predictants

Series	*KPSS level stationary test*		*KPSS trend stationary test*		*BDS test*					
					m=2		m=3		m=4	
	Results	*p*-value	Results	*p*-value	Results	*p*-value	Results	*p*-value	Results	*p*-value
T_{max}	0.01	>0.1	0.0001	0.02	0.14	0.001	0.23	0.001	0.30	0.001
T_{min}	0.01	>0.1	0.0002	0.001	0.14	0.001	0.24	0.001	0.31	0.001
C_5	0.07	>0.1	1.61×10^6	0.75	0.01	0.001	0.02	0.001	0.03	0.001
C_8	0.7	>0.1	-2.76×10^6	0.58	0.01	0.001	0.02	0.001	0.02	0.001
C_{10}	0.26	>0.05	-1.92×10^6	0.71	0.01	0.001	0.02	0.001	0.02	0.001
D_5	0.04	>0.1	9.84×10^7	0.85	0.02	0.001	0.04	0.001	0.05	0.001
D_8	0.02	>0.1	2.60×10^6	0.61	0.02	0.001	0.03	0.001	0.04	0.001
D_{10}	0.04	>0.1	1.16×10^6	0.75	0.02	0.001	0.03	0.001	0.04	0.001
H_5	0.01	>0.1	1.12×10^5	0.03	0.07	0.001	0.11	0.001	0.13	0.001
H_8	0.009	>0.1	1.36×10^6	0.79	0.09	0.001	0.16	0.001	0.20	0.001
H_{10}	0.008	>0.1	4.51×10^6	0.37	0.13	0.001	0.23	0.001	0.29	0.001
MSLP	0.18	>0.05	2.00×10^6	0.69	0.04	0.001	0.07	0.001	0.08	0.001
S_5	0.14	>0.05	1.42×10^6	0.78	0.05	0.001	0.07	0.001	0.08	0.001
S_8	0.02	>0.1	-5.65×10^7	0.91	0.02	0.001	0.03	0.001	0.03	0.001
S_{10}	0.08	>0.1	-7.99×10^7	0.87	0.01	0.001	0.02	0.001	0.02	0.001
T_2	0.008	>0.1	6.47×10^5	0.20	0.13	0.001	0.23	0.001	030	0.001
U_5	0.14	>0.05	2.71×10^6	0.61	0.07	0.001	0.11	0.001	0.13	0.001
U_8	0.10	>0.1	-4.36×10^6	0.41	0.05	0.001	0.07	0.001	0.08	0.001
U_{10}	0.07	>0.1	-1.21×10^5	0.02	0.02	0.001	0.04	0.001	0.05	0.001
V_5	0.05	>0.1	4.96×10^7	0.92	0.02	0.001	0.04	0.001	0.05	0.001
V_8	0.07	>0.1	3.11×10^6	0.55	0.02	0.001	0.03	0.001	0.04	0.001
V_{10}	0.07	>0.1	1.31×10^6	0.80	0.01	0.001	0.02	0.001	0.02	0.001
X_5	0.31	>0.01	0.0005	0.13	0.07	0.001	0.10	0.001	0.13	0.001
X_8	0.29	>0.01	0.0011	0.009	0.06	0.001	0.09	0.001	0.11	0.001
X_{10}	0.06	>0.1	0.0006	0.23	0.03	0.001	0.04	0.001	0.04	0.001
Z_5	0.01	>0.1	1.41×10^5	0.005	0.12	0.001	0.20	0.001	0.20	0.001
Z_8	0.03	>0.1	1.15×10^5	0.02	0.08	0.001	0.13	0.001	0.15	0.001

Table 3 GARCH parameters for selected predictors and predictants

Variables	ω	α	β	Variables	ω	α	β
T_{max}	**8.06**	**0.46**	**0.50**	S_{10}	**0.62**	**0.13**	**0.21**
T_{min}	**6.82**	**0.63**	**0.33**	T_2	**0.01**	**0.20**	**0.78**
C_5	**0.74**	**0.13**	**0.14**	U_5	**0.44**	**0.45**	**0.06**
C_8	**0.90**	**0.13**	**0.51**	U_8	**0.62**	**0.34**	0.01
C_{10}	**0.81**	**0.18**	0.002	U_{10}	**0.70**	**0.24**	0.04
D_5	**0.61**	**0.25**	**0.15**	V_5	**0.55**	**0.27**	**0.19**
D_8	**0.59**	**0.23**	**0.18**	V_8	**0.66**	**0.23**	**0.13**
D_{10}	**0.37**	**0.14**	**0.49**	V_{10}	**0.88**	**0.15**	0.03
H_5	**0.02**	**0.09**	**0.89**	X_5	**1487.3**	**0.77**	0.03
H_8	**0.02**	**0.19**	**0.79**	X_8	**3713.5**	**0.44**	**0.35**
H_{10}	**0.01**	**0.27**	**0.66**	X_{10}	**10881**	**0.30**	**0.29**
MSLP	**0.55**	**0.40**	**0.06**	Z_5	**0.10**	**0.44**	**0.43**
S_5	**0.58**	**0.32**	**0.10**	Z_8	**0.38**	**0.46**	**0.04**
S_8	**0.71**	**0.16**	**0.11**				

Note: Entries in bold indicate that the null hypothesis is rejected at the 5% level.

Table 4 Bivariate diagonal VECH parameters for conditional covariance between T_{max} and CGCM predictors

Predictors	ω_{21}	a_{21}	b_{21}	Predictors	ω_{21}	a_{21}	b_{21}
C_5	**1.32**	**0.29**	**-0.10**	T_2	**0.12**	**0.19**	**0.77**
C_8	**0.14**	**0.25**	**-0.24**	U_5	**-0.51**	**0.60**	**0.05**
C_{10}	**-1.13**	**0.23**	-0.06	U_8	-0.03	**0.51**	-0.03
D_5	**-0.20**	**0.38**	**-0.14**	U_{10}	**-0.63**	**0.37**	**-0.28**
D_8	**-1.16**	**0.33**	**-0.27**	V_5	0.001	**0.40**	**-0.10**
D_{10}	**-1.7**	**0.25**	**-0.39**	V_8	**1.15**	**0.30**	**-0.27**
H_5	**0.05**	**0.15**	**0.83**	V_{10}	**1.65**	**0.30**	**-0.11**
H_8	**0.09**	**0.19**	**0.78**	X_5	-5.02	**0.43**	-0.05
H_{10}	**0.18**	**0.27**	**0.68**	X_8	**11.7**	**0.44**	**0.48**
MSLP	**-0.41**	**0.51**	**-0.13**	X_{10}	**-55.8**	**0.27**	-0.004
S_5	0.01	**0.49**	**0.14**	Z_5	**0.77**	**0.46**	**0.43**
S_8	**-0.38**	**0.33**	**0.08**	Z_8	**1.32**	**0.57**	**0.06**
S_{10}	**-1.1**	**0.22**	**0.15**				

Note: Entries in bold indicate that the null hypothesis is rejected at the 5% level.

Table 5 DCC parameters for conditional covariance between T_{max} and CGCM predictors

Predictors	θ_1	θ_2	$\theta_1 + \theta_2$	CCC	Predictors	θ_1	θ_2	$\theta_1 + \theta_2$	CCC
C_5	**0.37**	0.001	0.37	0.22	T_2	**0.21**	**0.76**	0.97	0.91
C_8	**0.23**	**0.31**	0.56	-0.01	U_5	**0.65**	0.03	0.68	-0.23
C_{10}	**0.07**	**0.88**	0.95	-0.21	U_8	**0.56**	0.001	0.56	-0.06
D_5	**0.43**	0.001	0.43	-0.06	U_{10}	**0.40**	0.001	0.40	-0.09
D_8	**0.30**	**0.63**	0.93	-0.07	V_5	**0.51**	0.001	0.51	-0.04
D_{10}	**0.22**	0.001	0.22	-0.2	V_8	**0.36**	0.001	0.36	0.16
H_5	**0.11**	**0.86**	0.97	0.51	V_{10}	**0.05**	**0.90**	0.95	0.26
H_8	**0.18**	**0.78**	0.96	0.74	X_5	**0.42**	**0.50**	0.90	-0.07
H_{10}	**0.21**	**0.76**	0.97	0.88	X_8	**0.23**	**0.31**	0.54	-0.01
MSLP	**0.64**	0.001	0.64	-0.12	X_{10}	**0.33**	**0.61**	0.94	-0.08
S_5	**0.59**	0.001	0.59	-0.14	Z_5	**0.36**	**0.58**	0.94	0.80
S_8	**0.40**	0.001	0.40	-0.14	Z_8	**0.62**	**0.07**	0.67	-0.01
S_{10}	**0.27**	0.02	0.29	-0.23					

Note: Entries in bold indicate that the null hypothesis is rejected at the 5% level.

Table 6 Stationary and nonlinearity test results for conditional correlations between T_{max} and CGCM predictors

CGCM series	*KPSS level stationary test*		*KPSS trend stationary test*		*BDS test*					
					m=2		m=3		m=4	
	Results	*p*-value	Results	*p*-value	Results	*p*-value	Results	*p*-value	Results	*p*-value
C_5	0.11	>0.1	-1.40×10^6	0.34	0.02	0.001	0.04	0.001	0.06	0.001
C_8	0.04	>0.1	-2.39×10^6	0.17	0.04	0.001	0.07	0.001	0.09	0.001
C_{10}	0.02	>0.1	-1.96×10^6	0.05	0.15	0.001	0.25	0.001	0.32	0.001
D_5	0.03	>0.1	-9.43×10^6	0.56	0.03	0.001	0.05	0.001	0.06	0.001
D_8	0.06	>0.1	2.21×10^6	0.22	0.05	0.001	0.08	0.001	0.09	0.001
D_{10}	0.09	>0.1	1.56×10^6	0.07	0.02	0.001	0.04	0.001	0.06	0.001
H_5	0.02	>0.1	-4.12×10^6	0.007	0.16	0.001	0.27	0.001	0.34	0.001
H_8	0.01	>0.1	-7.76×10^6	0.001	0.16	0.001	0.27	0.001	0.34	0.001
H_{10}	0.02	>0.1	-7.51×10^6	0.001	0.17	0.001	0.28	0.001	0.36	0.001
MSLP	0.08	>0.1	-2.99×10^6	0.20	0.04	0.001	0.07	0.001	0.08	0.001
S_5	0.07	>0.1	-4.31×10^6	0.05	0.04	0.001	0.07	0.001	0.08	0.001
S_8	0.02	>0.1	2.37×10^6	0.14	0.03	0.001	0.04	0.001	0.06	0.001
S_{10}	0.03	>0.1	2.10×10^6	0.05	0.02	0.001	0.05	0.001	0.06	0.001
T_2	0.04	>0.1	6.54×10^6	0.001	0.17	0.001	0.28	0.001	0.35	0.001
U_5	0.02	>0.1	-4.91×10^6	0.03	0.05	0.001	0.08	0.001	0.09	0.001
U_8	0.03	>0.1	-7.63×10^7	0.71	0.03	0.001	0.05	0.001	0.06	0.001
U_{10}	0.04	>0.1	3.82×10^7	0.80	0.03	0.001	0.05	0.001	0.06	0.001
V_5	0.05	>0.1	1.14×10^6	0.55	0.03	0.001	0.05	0.001	0.06	0.001
V_8	0.05	>0.1	-7.81×10^7	0.56	0.03	0.001	0.04	0.001	0.06	0.001
V_{10}	0.10	>0.1	1.32×10^6	0.15	0.14	0.001	0.23	0.001	0.28	0.001
X_5	0.01	>0.1	7.63×10^6	0.03	0.17	0.001	0.28	0.001	0.36	0.001
X_8	0.04	>0.1	-2.39×10^6	0.17	0.04	0.001	0.07	0.001	0.08	0.001
X_{10}	0.01	>0.1	4.89×10^6	0.09	0.17	0.001	0.29	0.001	0.37	0.001
Z_5	0.06	>0.1	-4.56×10^6	0.008	0.13	0.001	0.21	0.001	0.26	0.001
Z_8	0.06	>0.1	-2.56×10^6	0.18	0.05	0.001	0.08	0.001	0.09	0.001

Table 7 Bivariate diagonal VECH parameters for conditional covariance between T_{min} and CGCM predictors

Predictors	ω_{21}	a_{21}	b_{21}	Predictors	ω_{21}	a_{21}	b_{21}
C_5	**0.49**	**0.31**	**-0.01**	T_2	**0.16**	**0.28**	**0.67**
C_8	**-0.6**	**0.25**	**-0.23**	U_5	**-0.28**	**0.65**	0.006
C_{10}	**-1.15**	**0.21**	-0.02	U_8	**0.20**	**0.55**	-0.03
D_5	**-0.54**	**0.40**	**-0.10**	U_{10}	**-0.22**	**0.36**	**-0.23**
D_8	**-1.02**	**0.31**	**-0.25**	V_5	**0.45**	**0.42**	**-0.09**
D_{10}	**-1.24**	**0.73**	**-0.39**	V_8	**1.12**	**0.28**	**-0.25**
H_5	**0.09**	**0.18**	**0.78**	V_{10}	**1.27**	**0.30**	-0.007
H_8	**0.09**	**0.24**	**0.74**	X_5	**8.98**	**0.45**	-0.05
H_{10}	**0.07**	**0.29**	**0.69**	X_8	**-1.56**	**0.38**	0.03
MSLP	**-0.98**	**0.54**	**-0.11**	X_{10}	**-30.1**	**0.29**	0.01
S_5	**-0.11**	**0.53**	**0.06**	Z_5	**0.83**	**0.66**	**0.13**
S_8	**-0.12**	**0.36**	0.05	Z_8	**0.45**	**0.65**	0.003
S_{10}	**-0.48**	**0.24**	**0.12**				

Note: Entries in bold indicate that the null hypothesis is rejected at the 5% level.

Table 8 DCC parameters for conditional covariance between T_{min} and CGCM predictors

Predictors	θ_1	θ_2	$\theta_1 + \theta_2$	CCC	Predictors	θ_1	θ_2	$\theta_1 + \theta_2$	CCC
C_5	**0.30**	0.001	0.30	0.15	T_2	**0.12**	**0.86**	0.98	0.92
C_8	**0.23**	0.001	0.23	-0.11	U_5	**0.59**	0.04	0.63	-0.21
C_{10}	**0.72**	**0.26**	0.96	-0.27	U_8	**0.50**	0.001	0.50	-0.11
D_5	**0.39**	0.001	0.39	-0.07	U_{10}	**0.18**	0.02	0.20	-0.07
D_8	**0.28**	0.001	0.28	-0.18	V_5	**0.41**	0.001	0.41	0.04
D_{10}	**0.16**	0.001	0.16	-0.19	V_8	**0.27**	0.001	0.27	0.18
H_5	**0.74**	**0.24**	0.98	0.57	V_{10}	**0.11**	0.02	0.13	0.27
H_8	**0.16**	0.001	0.49	0.83	X_5	**0.31**	**0.65**	0.96	-0.01
H_{10}	-0.003	**0.10**	0.10	0.89	X_8	**0.24**	0.001	0.24	-0.11
MSLP	**0.48**	0.001	0.48	-0.19	X_{10}	**0.26**	**0.70**	0.96	-0.06
S_5	**0.52**	**0.06**	0.58	-0.17	Z_5	**0.33**	**0.54**	0.87	0.76
S_8	**0.35**	0.01	0.36	-0.11	Z_8	**0.50**	**0.19**	0.45	0.45
S_{10}	**0.20**	**0.09**	0.29	-0.17					

Table 9 Stationary and nonlinearity test results for conditional correlations between T_{min} and CGCM predictors

CGCM series	*KPSS level stationary test*		*KPSS trend stationary test*		*BDS test*					
					m=2		m=3		m=4	
	Results	*p*-value	Results	*p*-value	Results	*p*-value	Results	*p*-value	Results	*p*-value
C_5	0.09	>0.1	-3.12×10^7	0.78	0.03	0.001	0.06	0.001	0.08	0.001
C_8	0.16	>0.05	-9.03×10^7	0.31	0.03	0.001	0.06	0.001	0.09	0.001
C_{10}	0.03	>0.1	-2.30×10^6	0.17	0.06	0.001	0.09	0.001	0.12	0.001
D_5	0.06	>0.1	-1.78×10^6	0.21	0.04	0.001	0.07	0.001	0.09	0.001
D_8	0.06	>0.1	-8.62×10^7	0.41	0.04	0.001	0.07	0.001	0.10	0.001
D_{10}	0.09	>0.1	2.51×10^7	0.69	0.04	0.001	0.08	0.001	0.10	0.001
H_5	0.06	>0.1	-7.28×10^6	0.01	0.05	0.001	0.08	0.001	0.09	0.001
H_8	0.03	>0.1	-2.16×10^7	0.22	0.03	0.001	0.05	0.001	0.06	0.001
H_{10}	0.03	>0.1	8.33×10^8	0.02	0.03	0.001	0.06	0.001	0.08	0.001
MSLP	0.03	>0.1	-2.30×10^6	0.17	0.05	0.001	0.09	0.001	0.12	0.001
S_5	0.01	>0.1	-3.21×10^6	0.09	0.05	0.001	0.08	0.001	0.10	0.001
S_8	0.03	>0.1	1.25×10^6	0.33	0.03	0.001	0.06	0.001	0.08	0.001
S_{10}	0.03	>0.1	9.35×10^7	0.24	0.04	0.001	0.07	0.001	0.10	0.001
T_2	0.04	>0.1	2.36×10^6	0.001	0.17	0.001	0.29	0.001	0.36	0.001
U_5	0.09	>0.1	-4.44×10^6	0.02	0.06	0.001	0.09	0.001	0.11	0.001
U_8	0.05	>0.1	-3.21×10^9	0.98	0.04	0.001	0.07	0.001	0.09	0.001
U_{10}	0.06	>0.1	7.45×10^7	0.45	0.04	0.001	0.08	0.001	0.11	0.001
V_5	0.03	>0.1	1.5×10^6	0.30	0.03	0.001	0.07	0.001	0.09	0.001
V_8	0.06	>0.1	9.35×10^8	0.93	0.04	0.001	0.07	0.001	0.10	0.001
V_{10}	0.10	>0.1	7.17×10^7	0.17	0.03	0.001	0.07	0.001	0.09	0.001
X_5	0.01	>0.1	1.01×10^5	0.005	0.18	0.001	0.31	0.001	0.39	0.001
X_8	0.16	>0.05	-9.03×10^7	0.31	0.03	0.001	0.06	0.001	0.08	0.001
X_{10}	0.01	>0.1	8.16×10^6	0.01	0.18	0.001	0.30	0.001	0.40	0.001
Z_5	0.04	>0.1	-7.67×10^7	0.45	0.12	0.001	0.19	0.001	0.23	0.001
Z_8	0.02	>0.1	-1.36×10^6	0.41	0.07	0.001	0.11	0.001	0.12	0.001

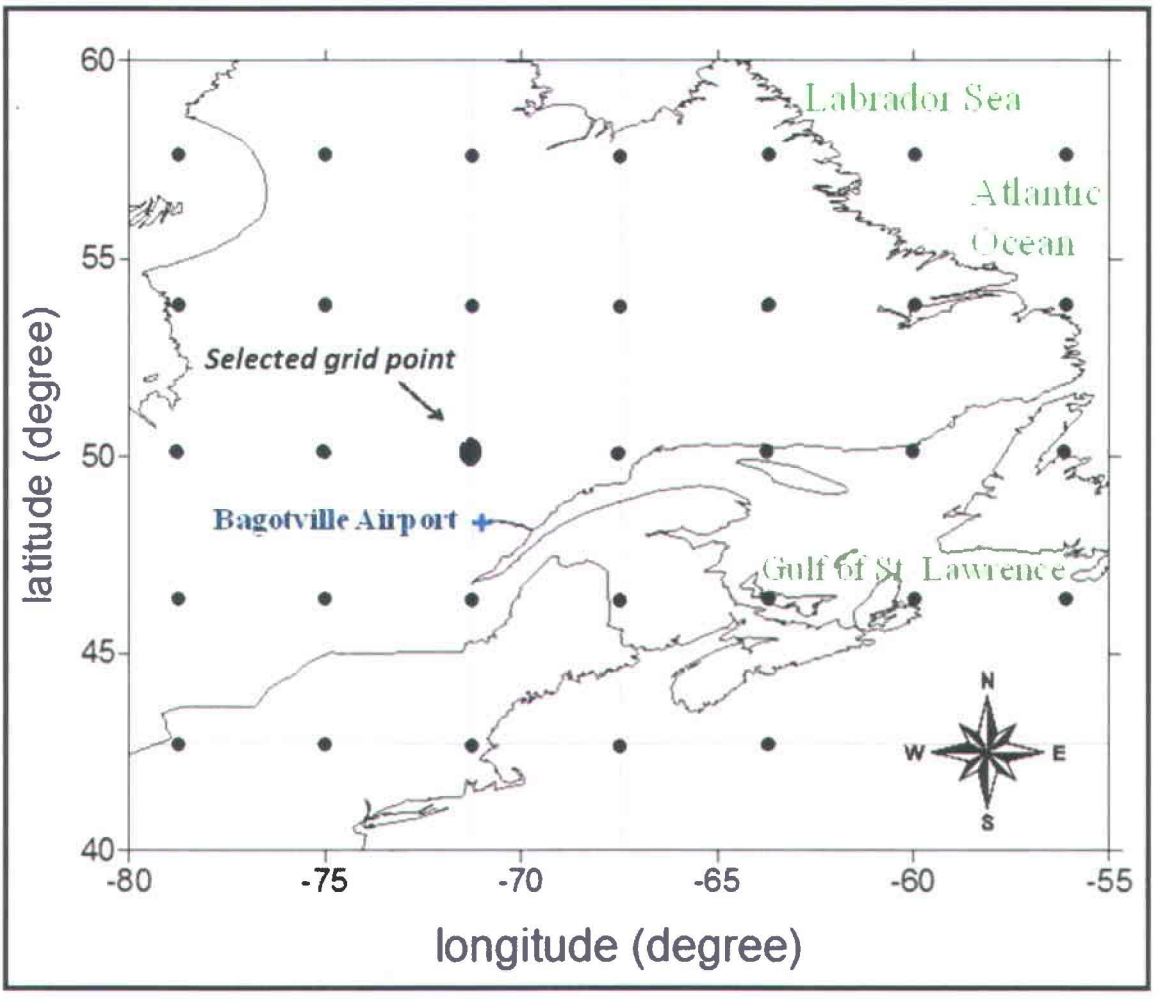


Figure 1. Location of CGCM3 grid points, selected grid point and surface observation site for this study

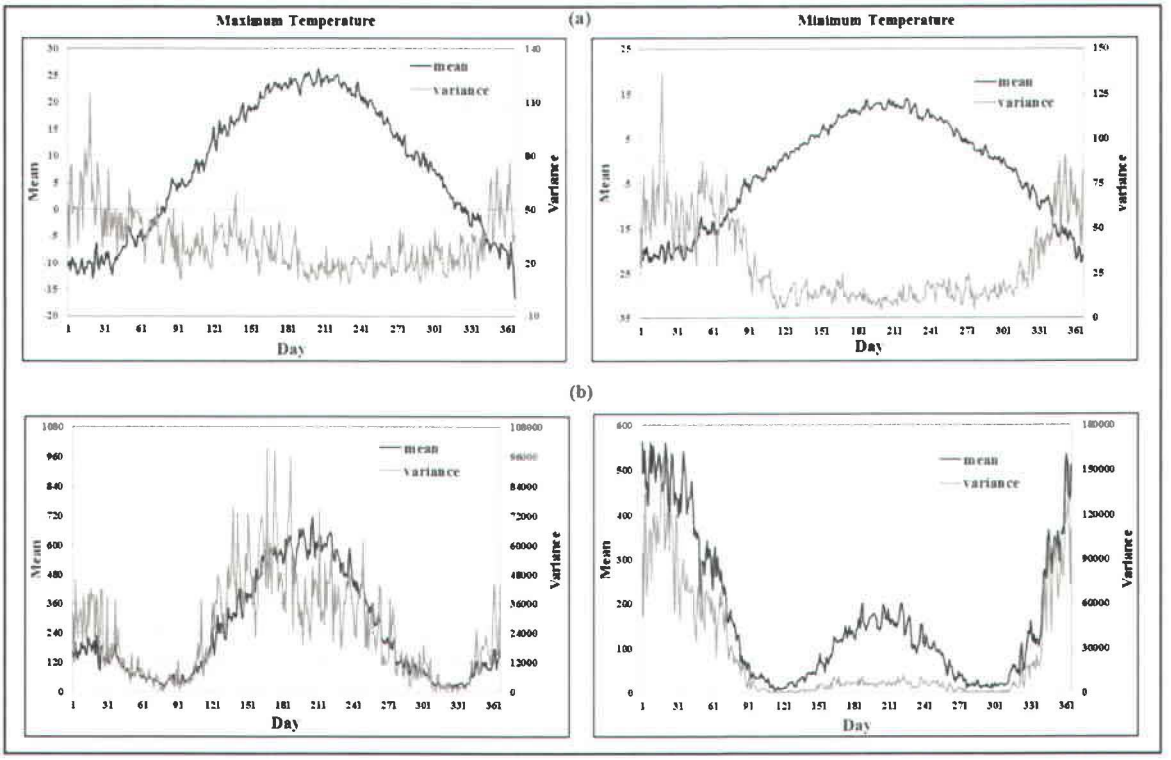


Figure 2 Daily evolutions of the mean and variance for a) original and b) squared temperature time series during 1980-2000

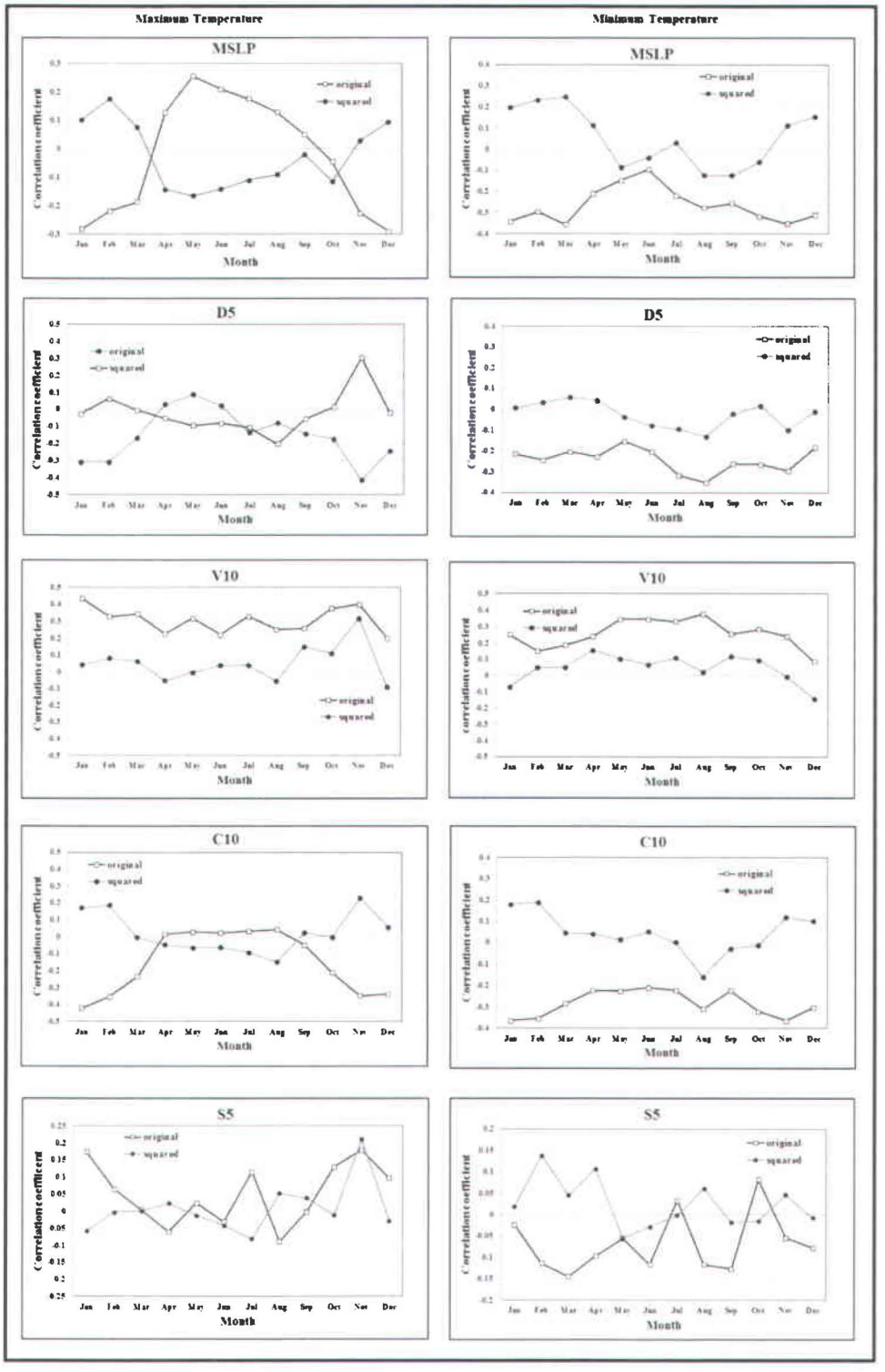


Figure 3 seasonal correlation coefficients between original and squared predictors and temperature data

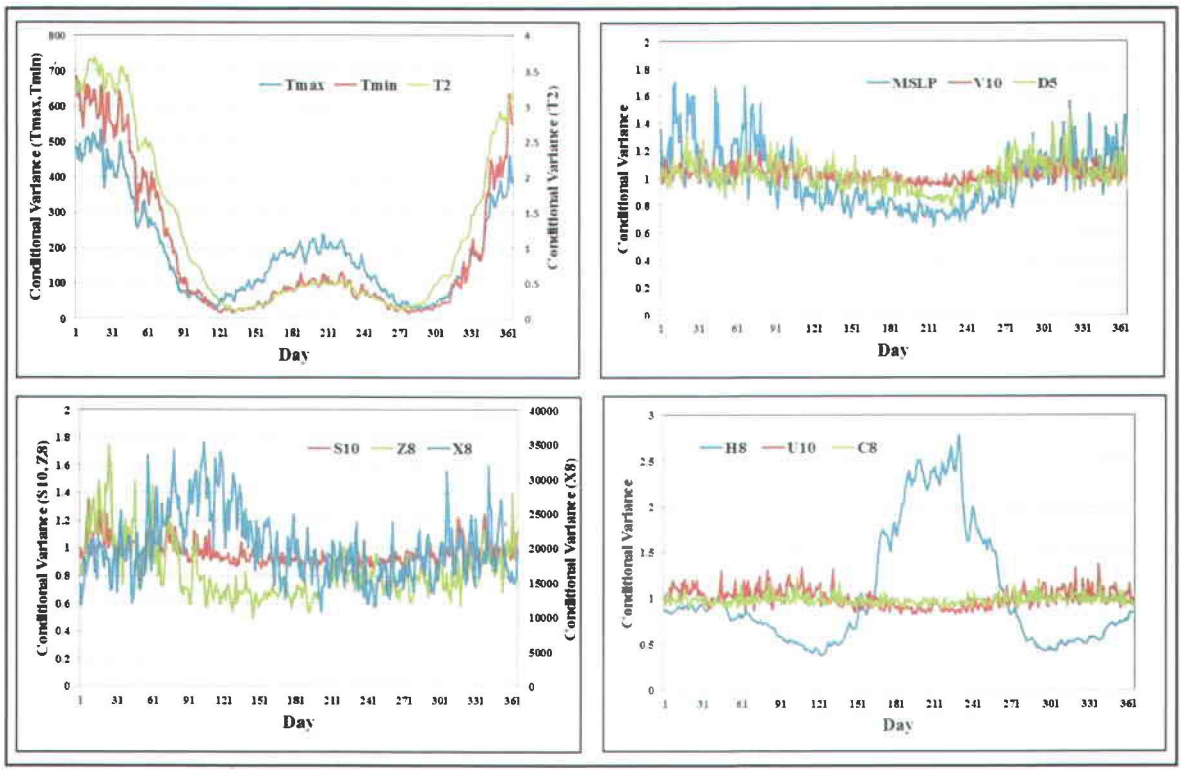


Figure 4 Average conditional variance for some GCM predictors and temperature time series (1980-2000) in a calendar time scale

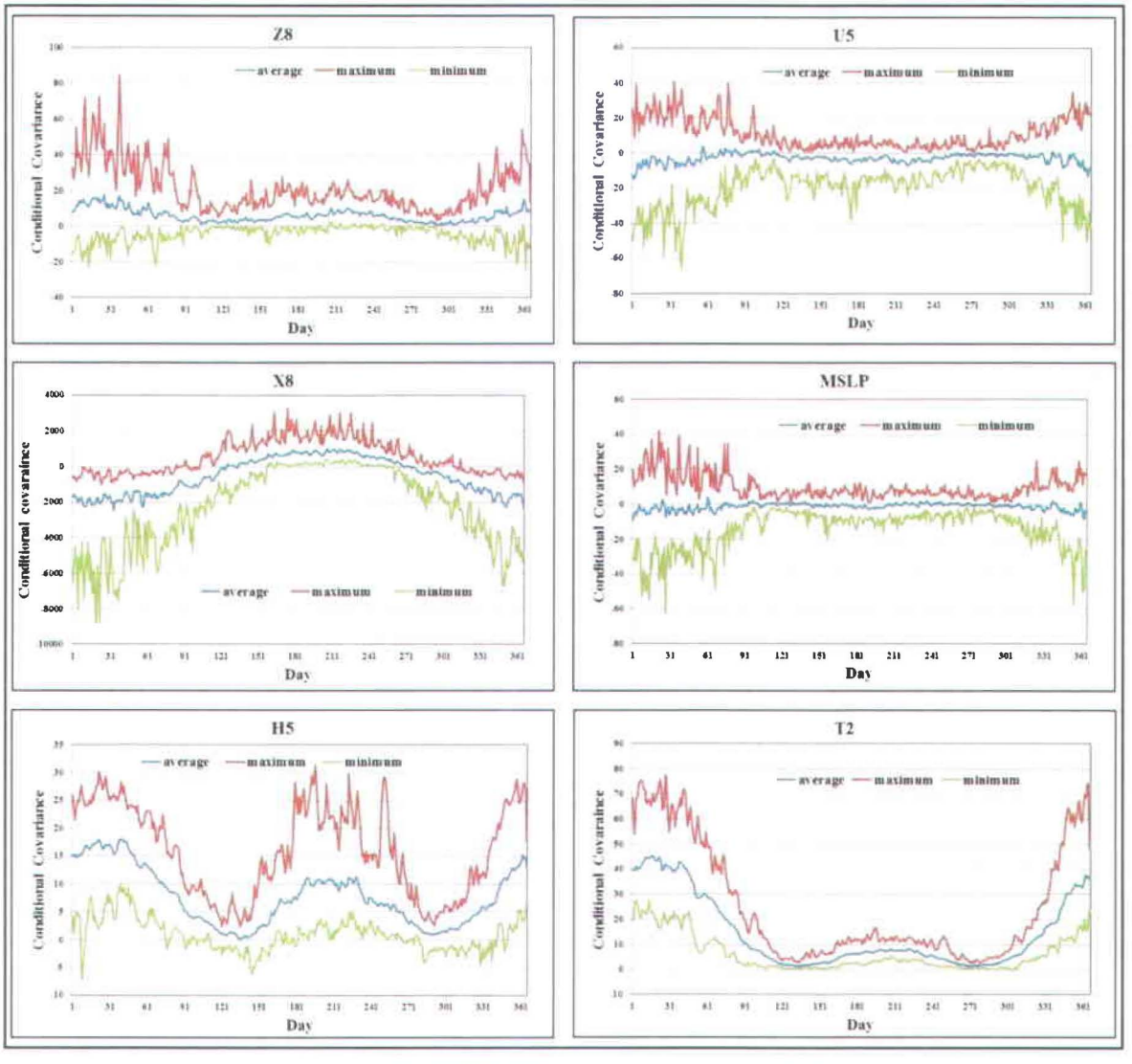


Figure 5 Daily average, maximum and minimum conditional covariance between GCM predictors and T_{max} time series (1980-2000)

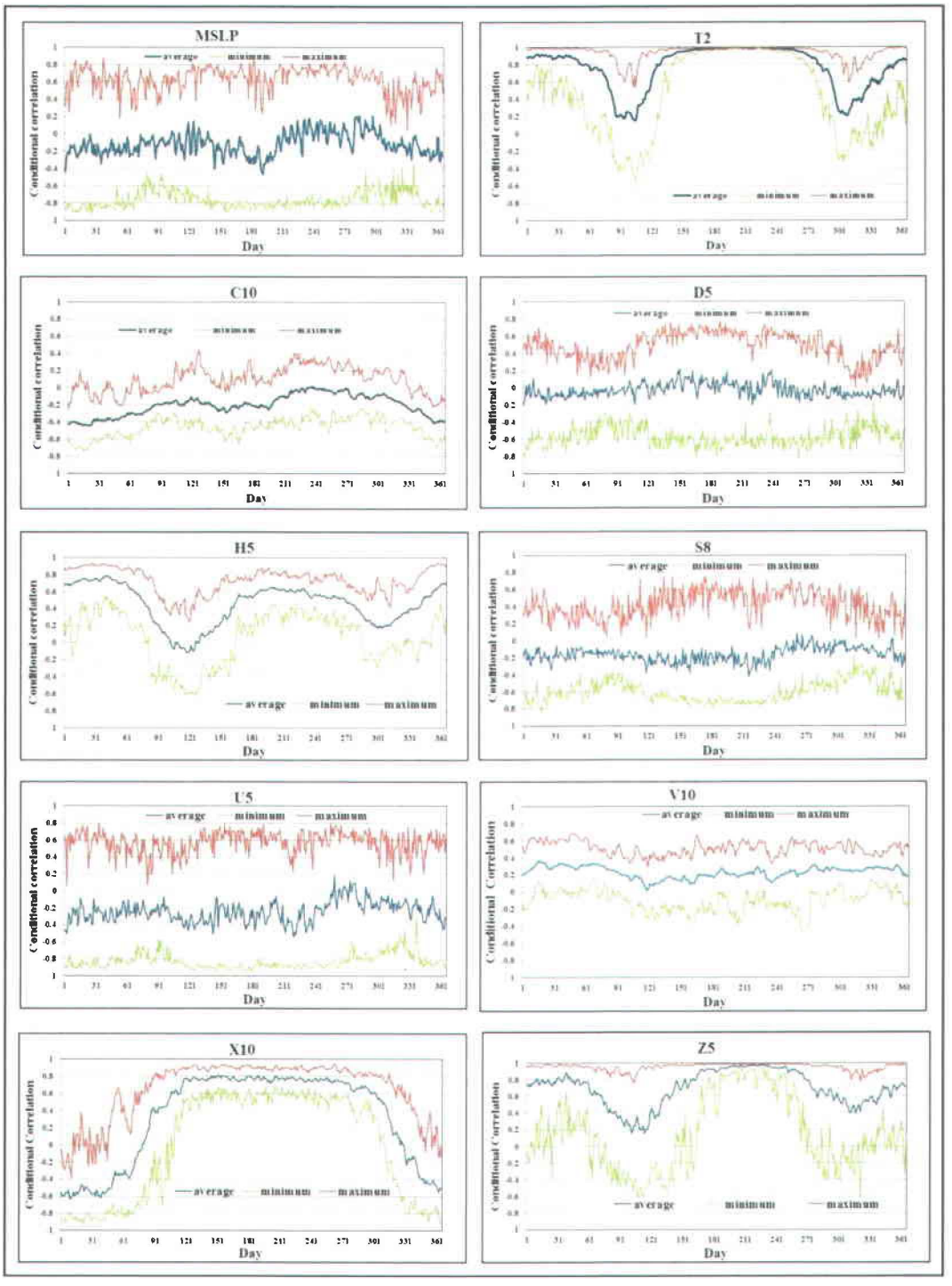


Figure 6 Daily average, maximum and minimum conditional correlation coefficients between typical GCM predictors and T_{max} (1980-2000)

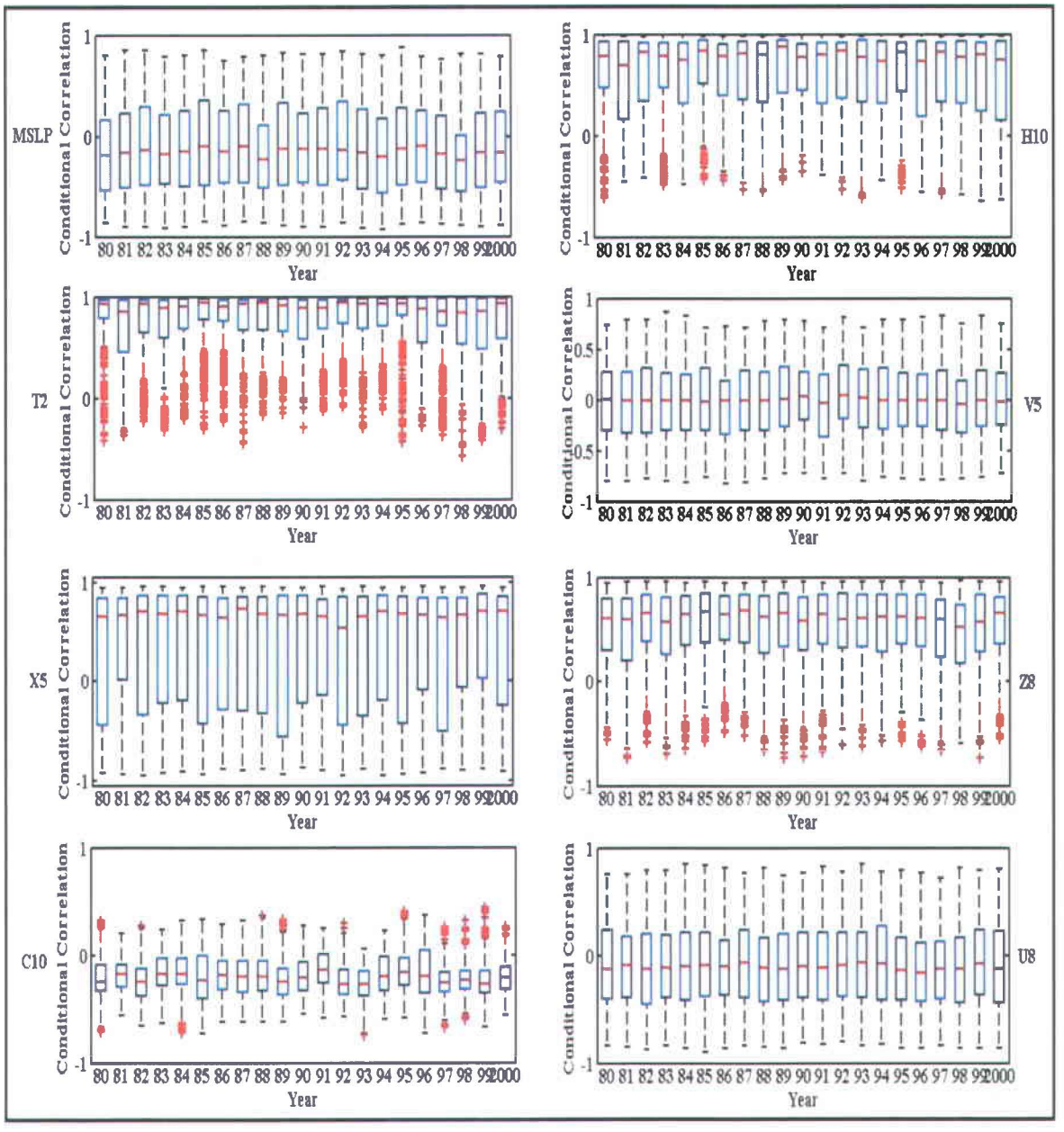


Figure 7 Annual variation of conditional correlation between T_{max} and some typical GCM predictors (1980-2000)

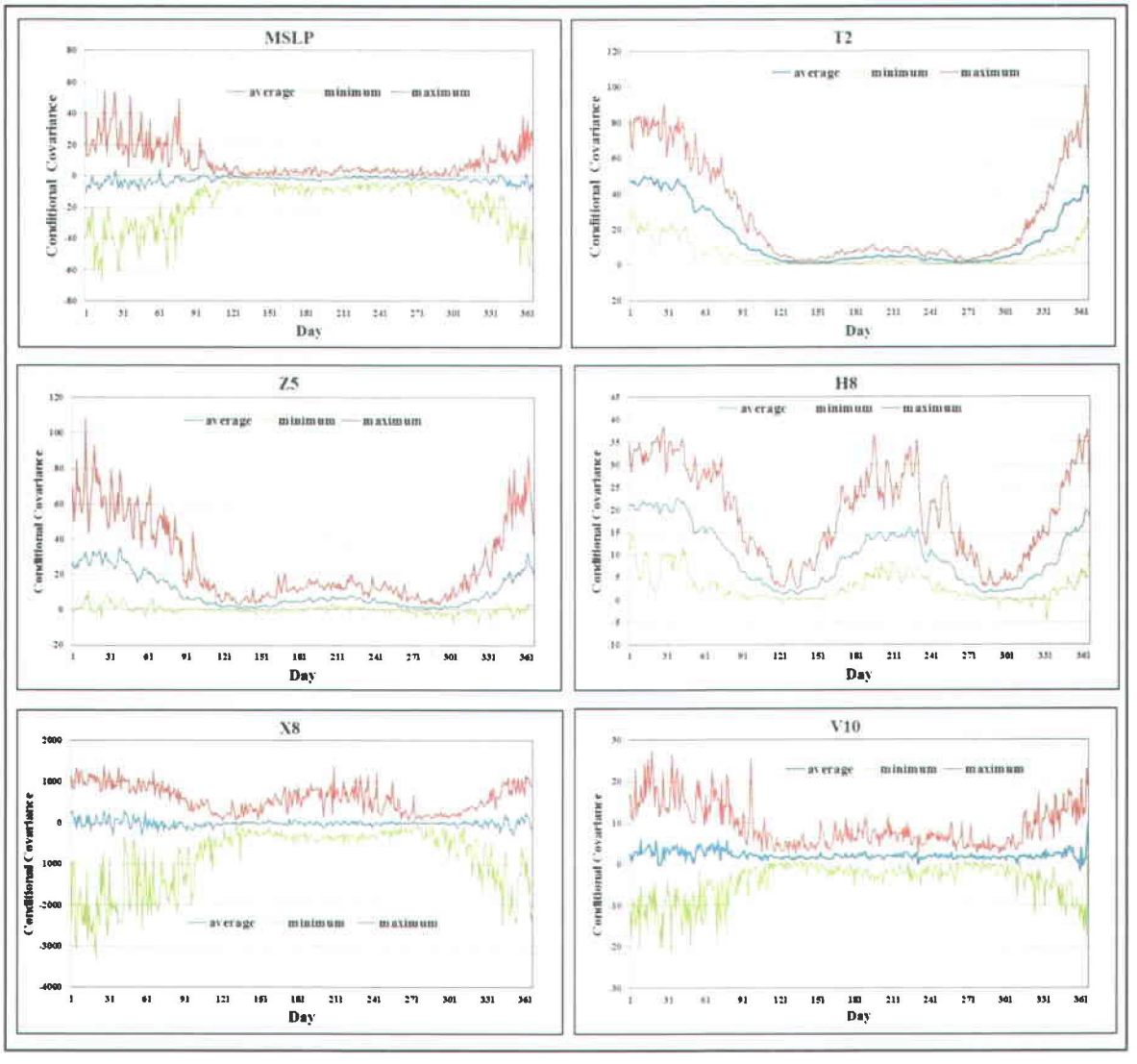


Figure 8 Daily average, maximum and minimum conditional covariance between typical GCM predictors and T_{min} time series (1980-2000)

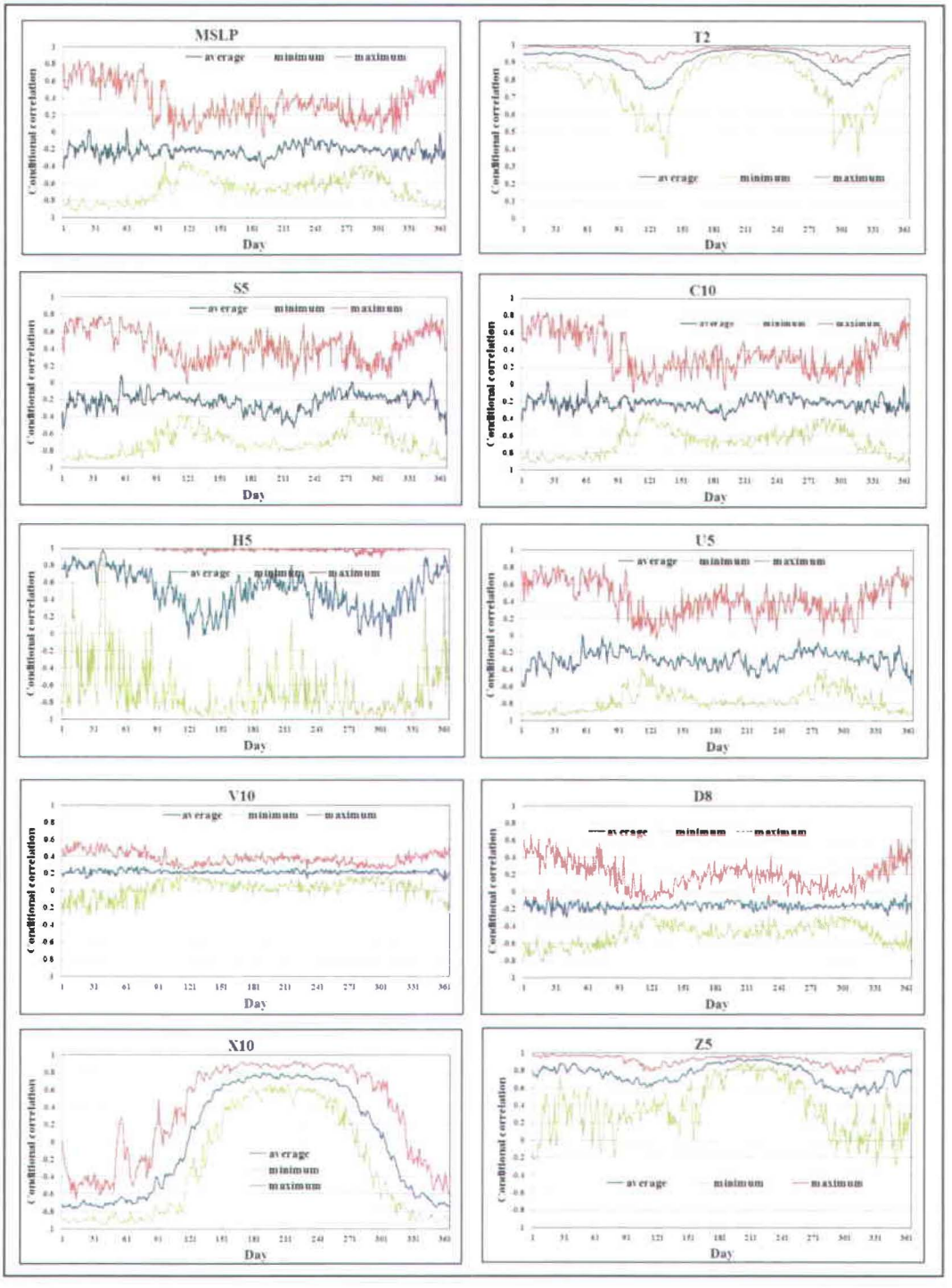


Figure 9 Daily average, maximum and minimum conditional correlation coefficients between typical GCM predictors and T_{min} (1980-2000)

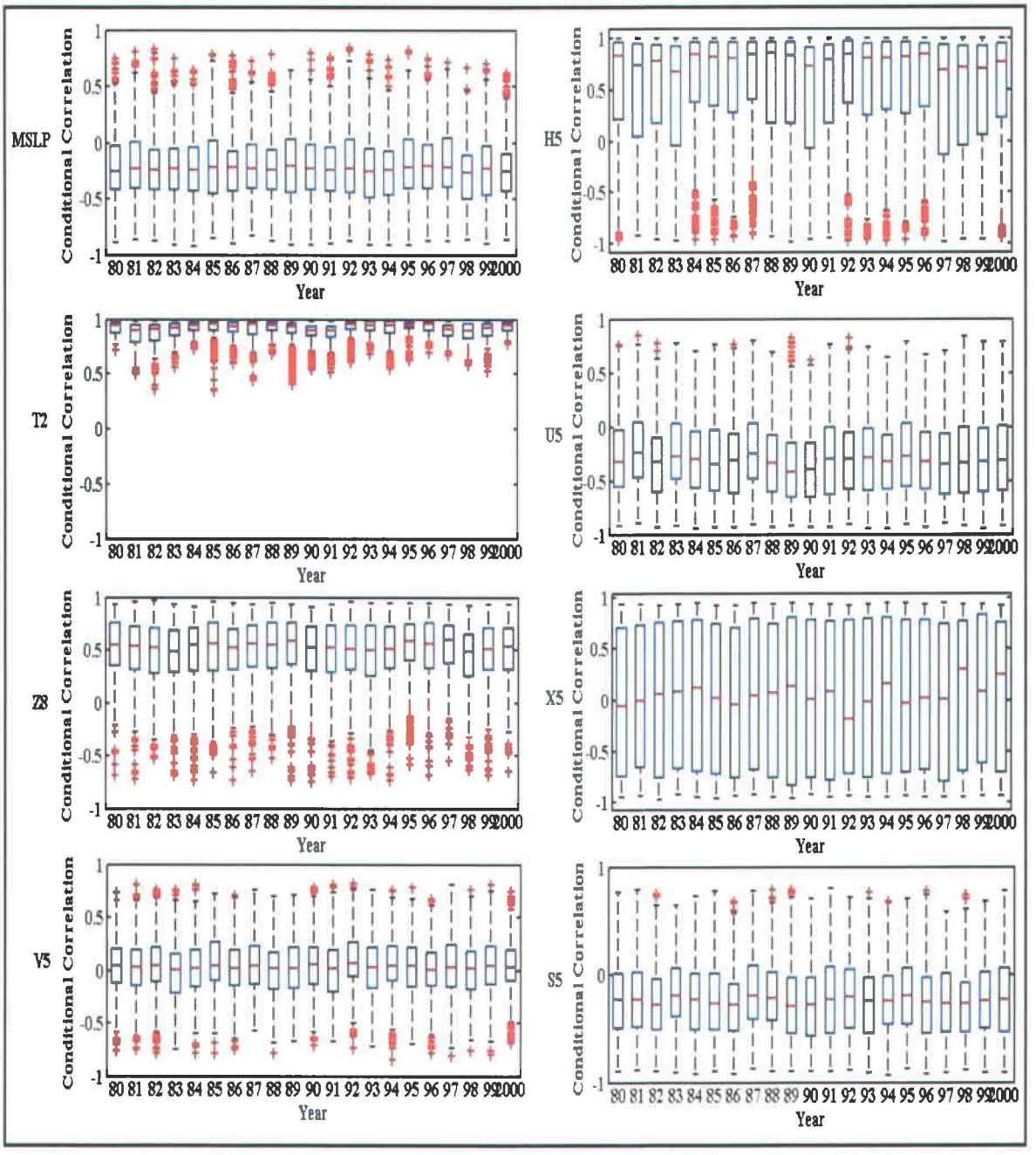


Figure 10 Annual and seasonal conditional correlation between T_{min} and GCM predictors (1980-2000)

Article 7. Modeling the relationship between climate oscillations and drought by a multivariate GARCH model

Modeling the relationship between climate oscillations and drought by a multivariate GARCH model

R. Modarres[a*], T. B. M. J. Ouarda[a,b]

[a] Hydroclimate modeling group, INRS-ETE, 490 De La Couronne, Québec, QC, Canada

Tel: +1 418 654-3842, Fax: +1 418 654-2600,

[b] Masdar Institute of Science and Technology, P.O. Box 54224, Abu Dhabi, UAE

E-mail: Reza.Modarres@ete.inrs.ca or touarda@masdar.ac.ae

[*] Corresponding author

Submitted for publication to Journal of Hydrometeorology
2013

Abstract
Typical multivariate time series models may exhibit co-movement in mean but not in variance of hydrologic and climatic variables. This paper introduces Multivariate Generalized Autoregressive Conditional Heteroscedasticity (MGARCH) models to capture the co-movement of the variance or the conditional covariance between two hydrologic time series. The diagonal VECH and diagonal BEKK models are developed to evaluate the covariance link between drought and two atmospheric circulations, Southern Oscillation Index (SOI) and North Atlantic Oscillation (NAO) time series during 1954-2000. The univariate GARCH model first indicates a strong persistency level in conditional variance for NAO and a moderate level for SOI. The conditional variance of short-term drought index indicates low level of persistency while the long-term index drought indicates high level of persistency in conditional variance. The estimated conditional covariance between drought and atmospheric indices is shown to be weak and negative. It is also observed that the covariance between drought and atmospheric indices is largely dependent on short-run variance of atmospheric indices rather than their long-run variance. The nonlinearity and stationarity tests show that the conditional covariances are nonlinear but stationary. However, the degree of nonlinearity is higher for the covariance between long-term drought and atmospheric indices. It is also observed that the nonlinearity of NAO is higher than that for SOI, in contrast to the stationarity which is stronger for SOI time series.

Key words: Drought, Bivariate GARCH, conditional covariance, diagonal VECH, diagonal BEKK, SOI, NAO, Nonlinearity

1. Introduction

Among natural hazards and disasters, drought is perhaps the most complex but least understood phenomenon with different characteristics in space and time which prohibit us to define its beginning and end. Drought spatial progress is slow and usually takes a long-time period to pass by a region. In addition, drought direct and indirect impacts on economic, social and environmental systems are destructive. These characteristics may have been the reasons for the development and application of a number of methods and approaches for drought definition, monitoring, modeling and forecasting over the past decades. A huge number of studies dealing with drought characterization can be seen in the literature. Four main types of droughts, namely meteorological, hydrological, agricultural and socio-economic have been discussed in the literature and a number of indices and methods have been developed to identify drought conditions and characteristics for each type of drought.

Among different methods used for drought characterization, linear time series approaches such as Autoregressive Moving Average (ARMA) and related models are very popular in hydrology. The characteristic of seasonality in hydrologic variables, such as droughts, and an inherent advantage of having a model with a few parameters but a reasonable result have made the time series approaches popular for drought time series modeling (e.g. Mishra and Desai, 2005; Durdu, 2010).

In spite of the popularity of multivariate analysis such as multivariate frequency distribution functions (copula functions) for hydrologic and drought probabilistic analysis (e.g Chebana and Ouarda, 2007; Shiau and Modarres, 2009; Shih-Chieh and Govindaraju, 2010), multivariate time series modeling approaches have not been reasonably investigated for drought modeling and

forecasting. The recent review by Mishra and Singh (2011) on drought modeling approaches indicates no multivariate time series model, such as vector ARMA model, application in the literature for drought analysis, though a few applications of simple multivariate autoregressive (MAR) time series model can be found for other hydrological variables such as streamflow (Niedzielski, 2007; Sohail et al., 2008; Chaleeraktrakoon, 2009). Another gap in drought research is that neither univariate nor multivariate "nonlinear" time series models which take into account the time varying variance-covariance or the conditional variance-covariance structure of drought, have been developed and applied in hydrological sciences or in drought research.

The nonlinear time series modeling approach usually refers to a popular econometrics Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model. The GARCH model is widely used in finance for investigating the volatility and time varying risk of the assets, stock markets and returns. The theoretical aspects of the model were first introduced by Engle (1982) and developed by Bollerslev (1986). The GARCH model has rarely been applied for hydrologic and climatic variables. Wang et al., (2005) and Chen et al., (2008) showed the advantage of univariate GARCH models over linear models. More recently, Modarres and Ouarda (2012a) indicated that the GARCH model does not have superiority over Seasonal Autoregressive Moving average (SARIMA) model for rainfall time series modeling except for removing heteroscedasticity from the residuals of the linear model.

Based on the literature review and different categories of time series methods (Figure1), the multivariate GARCH (MGARCH) models which are very popular in financial time series modeling have not been applied in hydrology yet. The main application of MGARCH models in econometrics is the study of the relationship between the conditional variance or volatility of different markets (Bauwens et al., 2006).

The aim of this study is to introduce and develop univariate and multivariate GARCH models for drought conditional variance-covariance, or volatility-covolatility, relationship with atmospheric indices. Furthermore, this study also examines and compares the nonlinearity and non-stationarity of drought and its link to atmospheric circulations.

This paper is organized as follows: the theoretical background of the univariate and multivariate GARCH models is given in the following section. The test methods applied in this study are presented in section 3. Section 4 is devoted to an example of the models for drought analysis. The last sections are devoted to concluding remarks and recommendations for future work.

2. GARCH models

Time series models can be classified based on the space of variables (univariate or multivariate) and the hypothesis of the underlying process (linear or nonlinear). According to this perspective, we have four types of time series models (Figure 1). The focus of this paper is on univariate and multivariate nonlinear GARCH time series modeling approaches which have not been used for drought analysis yet. These models are described in the following sections.

2.1. Univariate model

The univariate nonlinear model was first introduced by Engle (1982) as a class of Autoregressive Conditional heteroscedasticity (ARCH) model to capture the volatility clustering of financial time series. In an ARCH model, the conditional variance of the shocks (h_t^2) that occurs at time t is a function of the squares of past shocks ($\varepsilon_{t-1}^2, \ldots, \varepsilon_{t-v}^2$). Therefore, the ARCH model of order v or ARCH(v) can be written as follows:

$$h_t^2 = \omega + \sum_{i=1}^{v} \alpha_i \varepsilon_{t-v}^2 \qquad (1)$$

Bollerslev (1986) suggested adding lagged conditional variance to the ARCH model to generalize the effect of past variances on the current variance, h_t^2 in addition to the previous shocks. This model or the Generalized Autoregressive Conditional heteroscedasticity (GARCH(v,m)) model can then be specified as follows:

$$h_t^2 = \omega + \sum_{i=1}^{v} \alpha_i \varepsilon_{t-v}^2 + \sum_{j=1}^{m} \beta_j h_{t-m}^2 \qquad (2)$$

Where ω is constant and α and β are parameters of the model to be estimated. In this model, the short-run persistency in conditional variance is defined by the ARCH parameter (α) while the long-run persistency in conditional variance is defined by (β) parameter. The high value of ($\alpha + \beta$)indicates a high intensity of persistence in the conditional variance of the time series.

2.2. Multivariate Model

2.2.1. Overview

Having a *K*-dimensional zero mean, serially uncorrelated process $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t},...,\varepsilon_{Kt})'$ represented as:

$$\boldsymbol{\varepsilon}_t = \mathbf{H}_{t|t-1}^{1/2} z_t \qquad (3)$$

where z_t is a k-dimensional i.i.d. white noise, $z_t \sim i.i.d(0, I_K)$, then we have $\mathbf{H}_{t|t-1}$ as the conditional covariance matrix of $\boldsymbol{\varepsilon}_t$, given ε_{t-1}, ε_{t-2},... and $E[\boldsymbol{\varepsilon}_t|\Omega_{t-1}] = \mathbf{0}$ and $E[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\Omega_{t-1}] = \mathbf{H}_t$.

The above definition of conditional covariance matrix $\mathbf{H}_t$ needs to be parameterized now.

Remember the univariate GARCH (*v,m*) model, in a multivariate case, one may want to allow $\mathbf{H}_t$ to depend on lagged shocks $\boldsymbol{\varepsilon}_{t-i}$, $i = 1,\ldots,v$(or the ARCH process of order *v*) and on lagged

conditional covariance matrices $\mathbf{H}_{t-i}$, $i=1,\ldots,m$ (or the GARCH process of order m). Therefore, the general form of a MGARCH model is written as follows:

$$vech(\mathbf{H}_t)=\mathbf{W}+\mathbf{A}_1 vech(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1})+\mathbf{B}_1 vech(\mathbf{H}_{t-1}) \quad (4)$$

Where $\mathbf{W}$ is a $\frac{1}{2}K(K+1)\times 1$ vector and $\mathbf{A}_1$ and $\mathbf{B}_1$ are $(\frac{1}{2}K(K+1)\times\frac{1}{2}K(K+1))$ parameter matrices. The VECH(.) denotes the operator which stacks the lower portion of a matrix in a vector. As the conditional covariance matrix is symmetric, VECH($\mathbf{H}_t$) contains all unique elements of $\mathbf{H}_t$ and can therefore be written in a matrix form as follows:

$$vech\begin{bmatrix} h^2_{11,t|t-1} & h^2_{12,t|t-1} \\ h^2_{12,t|t-1} & h^2_{22,t|t-1}\end{bmatrix}=\begin{bmatrix} h^2_{11,t|t-1} \\ h^2_{12,t|t-1} \\ h^2_{22,t|t-1}\end{bmatrix}=\begin{bmatrix} w_{10} \\ w_{20} \\ w_{30}\end{bmatrix}+\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33}\end{bmatrix}\begin{bmatrix} \varepsilon^2_{1,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon^2_{2,t-1}\end{bmatrix} \quad (5)$$
$$+\begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33}\end{bmatrix}\begin{bmatrix} h^2_{11,t-1|t-2} \\ h_{12,t-1|t-2} \\ h^2_{22,t-1|t-2}\end{bmatrix}$$

The number of parameters of the VECH model (a_{ii} *and* b_{jj}) is equal to $(v+m)\left(\frac{K(K+1)}{2}\right)^2+K(K+1)/2$.

The main drawback of the VECH specification is that the number of parameters will become excessively large as K and the order of the model (v and m) increase. For example the above bivariate GARCH(1,1) model has 21 parameters while the trivariate model has 78 parameters. Estimation of this general model may therefore be quite problematic. Therefore, some specific diagonal parameterizations are introduced to reduce the number of parameters (Frances and van Dijk, 2000). Among these specifications, in this study we develop and apply the diagonal VECH and diagonal BEKK models and for simplicity, we discuss only the case m=v=1.

2.2.2. Diagonal VECH

One of the main disadvantages of the full VECH model is the high number of parameters, as mentioned above. To overcome this problem, Bollerslev et al., (1988) suggested a model by constraining the matrices $\mathbf{A}_1$ and $\mathbf{B}_1$ in (4) to be diagonal. In this case, the conditional covariance between $\varepsilon_{i,t}$ and $\varepsilon_{j,t}$, depends only on lagged cross-products of the two shocks involved and lagged values of the covariance itself

$$\mathbf{H}_t = \mathbf{W} + \mathbf{A}_1 \otimes (\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1}) + \mathbf{B}_1 \otimes \mathbf{H}_{t-1} \qquad (6)$$

Where $\otimes$ denotes the Hadamard or element-by-element product. This model is then called a diagonal VECH model and has $3(\frac{k(k+1)}{2})$ parameters. For example, the bivariate diagonal VECH(1,1) model can be given as follows:

$$vech\begin{bmatrix} h^2_{11,t|t-1} & \\ h^2_{12,t|t-1} & h^2_{22,t|t-1} \end{bmatrix} = \begin{bmatrix} W_{11} & \\ W_{21} & W_{22} \end{bmatrix} + \begin{bmatrix} a_{11} & \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \varepsilon^2_{1,t-1} & \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon^2_{2,t-1} \end{bmatrix} + \begin{bmatrix} b_{11} & \\ b_{21} & b_{21} \end{bmatrix} \begin{bmatrix} h^2_{11,t-1|t-2} & \\ h^2_{21,t-1|t-2} & h^2_{22,t-1|t-2} \end{bmatrix} \qquad (7)$$

This diagonal VECH model has therefore 9 parameters to be estimated which is much less than full VECH model with 21 parameters. To get the conditional variance-covariance equation from the above specification we can write

$$h^2_{11,t} = W_{11} + a_{11}\varepsilon^2_{1,t-1} + b_{11}h^2_{1,t-1} \qquad (8)$$

$$h^2_{21t} = W_{21} + \varepsilon_{1,t-1}\varepsilon_{2,t-1} + b_{21}h^2_{21,t-1} \qquad (9)$$

$$h^2_{22,t} = W_{22} + a_{22}\varepsilon^2_{1,t-1} + b_{22}h^2_{1,t-1} \qquad (10)$$

2.2.3. Diagonal BEKK

The BEKK is the acronym for the work by Baba, Engle, Kraft, and Kroner which was the early version of Engle and Kroners' paper (1995). The diagonal BEKK model is an alternative for the diagonal VECH presentation.

In this case, the diagonal BEKK(1,1) model can be written in a diagonal form where the off-diagonal elements are all equal to zero (apart from the constant term):

$$\begin{bmatrix} h^2{}_{11,t|t-1} & h^2{}_{12,t|t-1} \\ & h^2{}_{22,t|t-1} \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \\ & w_{22} \end{bmatrix} + \begin{bmatrix} a_{11} & \\ & a_{22} \end{bmatrix}' \begin{bmatrix} \varepsilon^2_{1,t-1} & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ & \varepsilon^2_{2,t-1} \end{bmatrix} \begin{bmatrix} a_{11} & \\ & a_{22} \end{bmatrix} \quad (11)$$
$$+ \begin{bmatrix} b_{11} & \\ & b_{22} \end{bmatrix}' \begin{bmatrix} h^2{}_{11,t-1|t-2} & h^2{}_{12,t-1|t-2} \\ & h^2{}_{22,t-1|t-2} \end{bmatrix} \begin{bmatrix} b_{11} & \\ & b_{22} \end{bmatrix}$$

And therefore the conditional variance-covariance equations can be written as the followings (Baur, 2006):

$$h^2{}_{11,t} = W_{11} + a^2_{11}\varepsilon^2_{1,t-1} + \mathrm{b}^2_{11}h^2{}_{1,t-1} \quad (12)$$

$$h^2{}_{21t} = W_{12} + a_{11}a_{22}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + \mathrm{b}_{11}b_{22}h^2{}_{21,t-1} \quad (13)$$

$$h^2{}_{22,t} = W_{22} + a^2_{22}\varepsilon^2_{1,t-1} + b^2_{22}h^2{}_{1,t-1} \quad (14)$$

In this study, therefore, we develop and apply the two diagonal specifications to establish the co-volatility relationship between drought and atmospheric indices.

3. Simulation and model verification

As MGARCH specifications have been developed to estimate the conditional covariance between two time series, different specifications are supposed to result in a similar covariance estimation. The performance of the MGARCH models for estimating conditional covariance is

analyzed and compared using simulation procedures and three performance criteria, normalized BIAS (NBIAS), normalized root mean square error (NRMSE) and Diebold and Mariano (DM) criteria.

The NBIAS and NRMSE criteria allow one to sort among models based on the covariance estimating accuracy. These criteria can be written as follows

$$NBIAS = \frac{1}{k}\sum_{i=1}^{k}\frac{s_{CCOV}-e_{CCOV}}{e_{CCOV}} \quad (15)$$

$$NRMSE = \sqrt{\frac{1}{k}\sum_{i=1}^{k}\left(\frac{s_{CCOV}-e_{CCOV}}{e_{CCOV}}\right)^2} \quad (16)$$

Where s_{CCOV} and e_{CCOV} denote simulated and empirical conditional covariances and k is the number of estimations.

However, these criteria do not test if the improvement in conditional covariance estimation between different models is statistically significant or not. To address this issue, the DM statistic is applied. The DM test (Diebold and Mariano, 1995) is a common test in financial time series modeling to compare different models with a basic model to evaluate if their outputs are different from the basic model or not (e. g. Mohammadi and Su, 2010). The DM test is applied in this study in order to evaluate if the conditional covariance estimated by diagonal VECH and diagonal BEKK are statistically different from the simulated conditional covariance.

Having $e_{1,t}$ and $e_{2,t}$, t=1, …n, as the errors between simulations and diagonal VECH and diagonal BEKK estimations, respectively, and $g(e_{1,t})$ and $g(e_{2,t})$ as their loss functions and $d_t = g(e_{1,t}) - g(e_{2,t})$ as the loss functions, Diebold and Mariano (1995) define the following statistic:

$$B = \frac{\bar{d}}{\sqrt{\frac{s}{n}}} \sim N(0,1) \qquad (17)$$

Where $\bar{d}$ is the sample mean, s is the variance of loss differential and n is the number of observations. The null hypothesis of zero mean loss differential or the equal of conditional covariance of different models is rejected if the test statistic is negative and statistically significant.

4. Testing procedures

4.1. Test for stationarity

Stationarity is one of the important characteristics of hydrologic variables. It is usually the basic assumption for a number of hydrologic modeling approaches and further tests such as linearity/nonlinearity testing. The stationary test is carried out in this study for drought and atmospheric indices as well as the conditional variance-covariance between them using two tests: Augmented Dickey Fuller (ADF) and Phillips-Perron (PP) unit root tests (Dickey and Fuller, 1979; Phillips and Perron, 1988). The PP test can examine the statioanrity around a deterministic trend (tend stationarity) and stationarity around a fixed level (level stationarity).

4.2. Test for nonlinearity

Natural systems, such as atmospheric processes, are commonly perceived as nonlinear. The nonlinear mechanism acting on drought and atmospheric circulations as well as their link is investigated in this study using the BDS test (Brock et al., 1996) which has its roots in chaos theory. It is based on the m-dimensional correlation integral where m represents the embedding

space in the new series $\{X_t\}$, $X_t = (x_t, x_{t-\tau}, \dots, x_{t-(m-1)\tau})$, which is generated from a scalar time series $\{Y_t\}$ of length N and then we have:

$$C_{m,M}(r) = \binom{M}{2}^{-1} \sum_{1 \le i < j \le M} H(r - \|X_i - X_j\| \qquad (18)$$

Where $M = N - (m-1)\tau$ is the number of embedded points in m-dimensional space, r is radius of a sphere centered on X_i and $H(u)$ is the Heaviside function (Abramowitz and Stegun, 1972, P.1020) and therefore, The BDS statistic for $m > 1$ is defined as:

$$BDS_{m,M}(r) = \sqrt{M} \frac{C_m(r) - C_1^m(r)}{\sigma_{m,M}(r)} \qquad (19)$$

where σ is the standard deviation of the points in the embedded m-dimensional space. Under the null hypothesis, $\{X_t\}$ is an i.i.d process and the BDS statistic converges to a unit normal as $M \to \infty$. This convergence requires large samples for values of embedding dimension much larger than m=2, so m is usually restricted to the range from 2 to 5 (Wang et al., 2006).

5. Applications

5.1. Data description

5.1.1. Drought index

Different indices have been developed for drought analysis among which the Standardized Precipitation Index (SPI) introduced by McKee et al. (1993) has received widespread applications. The SPI can quantify the precipitation deficit for different time scale and therefore is a flexible index to show the impact of drought on different types of water resources in space and time. SPI has also the advantages of statistical consistency and the ability to describe both short and long-term impacts of drought on water resources (Hayes et al. 1999). Therefore, many

studies on modeling drought characteristics such as severity, duration and frequency as well as the studies on drought forecasting and drought links to atmospheric and climatic indices have applied SPI as a drought index (e.g. Bordi and Sutera, 2001; Vicente-Serrano, 2005, Ozger et al., 2012).

The drought data set in this study includes the 3-month and 12-month Standardized Precipitation Index (SPI3 and SPI12, hereafter) time series for 2 stations in the northwestern and southwestern territories of Iran, namely, Oroomieh and Shiraz stations during 1954-2010. The location of these stations is illustrated in Figure 2 together with the climate zones of Iran classified based on the UNEP aridity index (Raziei and Pereira, 2012). The SPI time series of these stations are illustrated in Figure 3. The SPI3 and SPI12 time series are selected in order to compare the heteroscedastic characteristics of both short and long-term drought indices.

5.1.2. Atmospheric indices

In this study, two major atmospheric indices which are widely mentioned to influence the precipitation of Iran, especially over the western and southwestern territories (e.g Nazemosadat and Ghasemi, 2004; Raziei et al., 2009), the Southern Oscillation Index (SOI) and North Atlantic Oscillation (NAO) are used for drought time series modeling. The monthly time series of atmospheric indices during 1954-2010 are given in Figure 4.

5.2. Preliminary data analysis

The link between atmospheric indices (SOI and NAO) and drought conditions for the selected stations is first investigated through a descriptive analysis including the unconditional correlation coefficients and cross-correlation coefficients in different lag times. The (unconditional) monthly

correlation coefficients between oscillations and SPI time series are given in Figure 5. As the figure indicates, a negative correlation between SOI and SPI for both stations and both SPI time series is observed. However, the negative correlation is stronger for the winter season than the other seasons and some positive (but weak) relationship is observed for the summer season. This is in agreement with Nazemosadat and Ghasemi (2004) who indicated the low intensity of winter drought during El Nino events and also in agreement with the weak correlation coefficients between SOI and SPI reported by Raziei et al., (2009).

On the other hand, the NAO-SPI link shows both positive and negative associations for all seasons. It is observed that the NAO-SPI association is temporally irregular compared to the SOI-SPI association and does not show a strong seasonality. However, the winter drought (negative SPI) seems to be related to the negative NAO phase while summer drought indicates both positive and negative association with NAO for both stations and time scales.

In addition, the cross correlation coefficients given in Figure 6, panel A, indicate a weak lag time effect in SOI-SPI association while it is almost insignificant for the NAO-SPI association. The association is usually weakening and becoming insignificant for lag times $k > 2$. This suggests a short-run memory in SOI-SPI and NAO-SPI relationship. On the other hand, the cross correlation between squared SOI and SPI time series (Figure 6, Panel B) shows a weaker relationship than for the original (non- squared) data. The most interesting feature in Figure 6 belongs to SPI time series of Oroomieh station. For example, the negative insignificant correlation coefficients between the original SPI3 and SOI time series have become positive and significant for the squared data, while it is not observed for SOI-SPI12 association. The same condition is observed for NAO-SPI at Oroomieh station where negative correlations for the original data have become positive for squared time series. For Shiraz station, it is observed that

the correlation coefficients between squared time series are almost weaker than those for the original time series. This suggests that the second order moment or the variance of atmospheric indices may have a different association to drought than the first order moment or the mean. This phenomenon has not yet been considered in previous drought studies and would be interesting to investigate.

5.3. Conditional variance models

The univariate GARCH model is developed for drought and atmospheric indices and the parameters are estimated using the maximum likelihood method. The order of the GARCH models and their parameters are illustrated in Table 1.

This table shows that the conditional variances of atmospheric indices are different from each other. It can be see that the short-run persistence of SOI is much stronger than that for NAO while in the opposite, NAO shows a long run persistence as the β parameter is large. One can also see that NAO has a stronger intensity of persistence and variance memory than SOI as $\alpha + \beta$ is larger for NAO. The conditional variance time series of atmospheric indices are illustrated in Figure 7. This figure shows that the conditional variance of SOI is larger than that for NAO and extreme conditional variances are observed for SOI time series. It should also be noted that no seasonality is observed for the conditional variances of atmospheric indices. However, extreme conditional variances are usually observed in the winter season.

Table 1 also shows that the short run persistence is dominant for drought time series where the GARCH parameter (β) is much smaller than the ARCH parameter (α) (except for SPI12 at Shiraz station) in the models. The ARCH parameters also indicate a stronger short-run persistency in conditional variance for SPI12 than for SPI3 time series, implying volatility

clustering in long-term drought. The conditional variance for SPI12 at both stations shows a high degree of intensity of persistency according to $\alpha+\beta$ measurement. The conditional variances of drought time series are illustrated in Figure 8. This figure shows that the conditional variance of short-term drought varies rapidly through time while the conditional variance of long-term drought shows some sudden drastic increase and less fluctuation. In addition, no sharp seasonality is observed for drought conditional variances.

We next test the conditional variance of atmospheric and drought indices for stationarity and nonlinearity using the ADF, PP and BDS tests. The results are given in Table 2. Both ADF and PP tests indicate stationary conditional variance for all data series. The results for atmospheric indices show less stationarity for NAO time series than for SOI according to the ADF test statistic. It can also be seen that neither level nor trend non-stationarity is observed for the conditional variance of atmospheric indices. The BDS test statistics show a higher nonlinearity degree for the conditional variance of NAO for all dimensions.

Moreover, it is observed that the stationarity for the conditional variance of SPI3 is stronger than that for SPI12 at both stations while the stationarity is stronger at Shiraz station. On the contrast, the nonlinearity of SPI12 time series is stronger according to the BDS test. However no significant difference is observed between drought nonlinearity degrees at the two stations.

5.4. Conditional covariance models

The bivariate model for conditional covariance between SPI and atmospheric indices is developed using two types of MGARCH model diagonal specifications, the diagonal VECH and the diagonal BEKK models. The results for the two stations are given in the following sections.

5.4.1. Models for Oroomieh station

The estimates of the parameters of the diagonal VECH (1,1) model for Oroomieh station are given in Table 3. These parameters are estimated using the maximum likelihood method. It is observed that the elements of the matrices, W, A and B, are all statistically significant for SOI-SPI relationship. However, some parameters for NAO-SPI relationship in matrix B are not significant. These estimations lead to write the equations for conditional covariances between drought and atmospheric indices (given below Table 3).

The diagonal VECH models show that the conditional covariances depend greatly on the cross products of the lagged shocks rather than the lagged covariances.

We can see that β parameters are negative, except for SOI-SPI12, implying that the covariance at each time step, *t*, has a negative association to the covariance at time step *t-1*. The highest (negative) covariance link is observed for NAO-SPI3 with $\beta = -0.80$. The highest intensity of persistency is observed for SOI-SPI12 covariance where $\alpha + \beta = 0.8$ and the lowest belongs to NAO-SPI12 covariance with $\alpha + \beta = 0.08$.

We next move to estimation of the diagonal BEKK model and its 7 parameters which are illustrated in Table 4. The diagonal matrices, A and B, are all significant indicating both SOI and NAO influence the conditional variance of the SPI time series. The conditional covariance between atmospheric indices and drought at Oroomieh station using the diagonal BEKK model are given below table 4.

Similar to the diagonal VECH model, these equations indicate also that short run persistency is much stronger than long run persistency in the covariance structure between drought and atmospheric indices. The largest persistency is observed between SOI and SPI12 ($\alpha + \beta =$

0.79). It is also observed that the covariance between drought and NAO is negative for both short and long-term drought time series. The only difference between BEKK and VECH estimation is the β parameter for Eq. (26) which is much smaller than the estimation of the VECH model.

In order to verify and select a MGARCH model among diagonal models, we apply a simulation procedure to simulate the conditional covariance between SPI and atmospheric indices.

The criteria for conditional covariance simulations are given in Table 5. The NBIAS indicates that the diagonal VECH model performs relatively better than the diagonal BEKK model for estimating the conditional covariance. However, the NRMSE shows a better performance for the diagonal BEKK model to estimate the SOI-SPI3 and NAO-SPI12 conditional covariances. It is also observed that the uncertainty of the covariance estimation is relatively higher for the NAO-SPI relationship than that for SOI-SPI according to both criteria. This may be due to the weak covariance between NAO and drought at both stations.

On the other side, the DM statistics reveal that the estimated and simulated conditional covariances are statistically different among the models, except for the covariance between SOI and SPI12. In other words, the difference between diagonal VECH estimation and simulation (e_1) and the difference between diagonal BEKK estimation and simulation (e_2) are statistically significant. Therefore, there is a significant difference between the two models for estimating the conditional covariance and the VECH model seems to give better covariance estimation.

According to simulation results, the diagonal VECH model is used to plot the time varying conditional covariances and correlations between drought and atmospheric indices for Oroomieh station (Figure 9). This figure illustrates that SOI has a larger covariance with drought than the NAO for both short and long-term drought time series. The conditional correlation between

drought and SOI is much stronger than that between drought and NAO. The conditional correlation between SOI and SPI are usually falling within ± 0.40 while they are usually falling within ± 0.2 for NAO.

In addition, the seasonal variation in the link between drought and atmospheric indices is investigated through drawing monthly boxplots of conditional correlation coefficients (Figure 10). In this figure, each boxplot includes 57 correlation coefficients. Figure 10 indicates no sharp seasonality in the correlation coefficients. However, a few extreme (out of the 75% Quantiles) correlation coefficients between NAO and SPI3 are observed from August to January. It is also observed that these extreme coefficients are mostly observed for the short-term drought index (SPI3) while long-term drought (SPI12) does not show extreme (positive or negative) correlation coefficients. This suggests that the link between drought and atmospheric circulation becomes stronger than normal condition, mostly for short-term drought events at Oroomieh station.

The annual variation of conditional correlation between SPI and atmospheric indices are given in Figures 11 and 12 for SOI and NAO, respectively. In this figure each boxplot includes 12 correlation coefficients for 12 months. This figure does not show a strong fluctuation for SPI3 but the link between SPI12 and atmospheric indices shows a weak 3-5 years periodicity, especially for SPI12-SOI link.

Finally, we come to compare the degree of nonlinearity and stationarity of conditional covariances using BDS, PP and ADF tests. The results are given in Table 6. It is seen that the conditional covariances are stationary regarding both ADF and PP test results. It can also be seen that the covariance between SPI and NAO is more stationary than the covariance between SPI and SOI for both SPI3 and SPI12. However, a weak trend non-stationarity in the covariance of

NAO-SPI12 structure should be noticed where the p-value of the test statistic is on the level of hypothesis rejection (p=0.05).

The BDS test indicates that the conditional covariance between SPI12 and the atmospheric indices has a higher degree of nonlinearity than the covariance between SPI3 and the atmospheric indices for all dimensions. Similarly to conditional variance, the high the stationarity the less the nonlinearity is observed for conditional covariance.

5.4.2. Model development for Shiraz station

The same procedure is followed to estimate the conditional covariance and correlation between SPI and the atmospheric indices for Shiraz station. The estimated diagonal VECH and 9 parameters are shown in Table 7. It is clear that the conditional variance of SPI depends on their own lags, lagged cross-products of the shocks and lagged conditional covariance. However, the conditional covariance parameters are not significant for NAO-SPI12 link (same as for Oroomieh station). Based on the parameters, the conditional covariances for drought at Shiraz station (Table 7) show the same covariance structure as the Oroomieh station. It is seen that the β parameters are negative for drought and NAO relationship and the persistency of the covariance structure is not very strong for NAO. However, the covariance between drought and SOI is significant and positive. The largest intensity is observed for SOI-SPI12 where $\alpha + \beta = 0.74$ which is relatively high but not very strong ($\alpha + \beta < 0.90$).

In the following, we look at the estimates of the diagonal BEKK models for Shiraz station(see Table 8). The 2×2 parameter matrices are almost all significant for both drought time series. According to the BEKK models in this table, a weak covariance structure between drought and atmospheric indices at Shiraz station is observed. The SOI-SPI connection is stronger than NAO-

SPI link. The covariance between SOI and SPI seems to depend relatively on cross- products of shocks and weakly on lagged covariances (which is relatively small for SOI12 ($\beta = 0.19$) and almost null for SPI3 ($\beta = 0.02$)). The covariance structure between NAO and SPI also depends dimly on cross-products of shocks. However, the lagged covariance structure between NAO and SPI is negative and null ($\beta = -0.06$). Similar to Oroomieh station, the largest long run persistency is observed between SOI and 12-month SPI ($\alpha + \beta = 0.72$).

To select between VECH and BEKK models, the performance criteria of the models are examined (Table 9). It is clear from the DM statistics that the accuracy of the two models is statistically different by comparing the simulated conditional covariance. Following the DM test results, both NBIAS and NRMSE criteria show a much better performance of diagonal VECH model against diagonal BEKK models for estimating the conditional covariance between drought and atmospheric indices. Therefore, the diagonal VECH model is applied to present the conditional covariance at Shiraz station.

Using the diagonal VECH model, we give the time varying conditional covariances and correlations for Shiraz station in Figure 13. This figure shows no significant temporal variation difference between SOI-SPI3 and NAO-SPI3. However, the covariance for SOI-SPI12 is much stronger than that for NAO-SPI12. The correlation coefficients show a much larger association between SOI and drought than that between NAO and drought at Shiraz station.

The monthly and annual variation of conditional correlation coefficients is also illustrated in Figures 14, 15 and 16 through boxplots. The monthly distribution of conditional correlation coefficients is almost identical to that observed for Oroomieh station. However, the annual variation of the correlation coefficient seems to be more irregular than that for Oroomieh station.

This may suggest that the atmospheric circulations have more regular relationship with drought at Oroomieh station than with drought at Shiraz station.

The stationarity and nonlinearity results for the conditional covariance are given in Table 10. The results indicate stationary covariances between drought and atmospheric indices regarding ADF and PP tests. However a non-stationarity trend is also observed at 5% significance level for the SOI-SPI3 and NAO-SPI12 covariances. This is perhaps due to extreme drought events which are strongly influenced by atmospheric indices.

On the other hand, all covariances are nonlinear according to BDS test results. In contrast to Oroomieh station, it is observed that the degree of nonlinearity is more or less similar among different covariances. In other words, the covariance structure between drought an atmospheric indices seem to be more linear than that at Oroomieh station.

6. Summary and Conclusions

This paper developed and applied univariate and multivariate GARCH approaches namely diagonal VECH and diagonal BEKK, to investigate the association between drought and atmospheric indices through a new conditional variance-covariance perspective. This study provides this new outlook by an example of the association of SOI and NAO to short-term and long-term SPI time series at two stations in Iran.

We showed that the diagonal VECH model with 9 parameters has less biased covariance estimations than the diagonal BEKK model with 7 parameters. Both of these models have reasonably fewer parameters than the full VECH model with 21 parameters, but the diagonal VECH model seems to give more accurate estimations for conditional covariance regarding the

simulation experiment. Therefore, the diagonal VECH model outputs on the variance-covariance link were given and discussed.

From the example presented in this study our findings on the development of a bivariate heteroscedastic model for drought analysis can be summarized in the following items:

5.1. The conditional variance of atmospheric indices demonstrated different behaviors. While NAO shows a high degree of memory in the conditional variance, SOI does not show either a strong long run or short run memory in the conditional variance. However, the intensity of persistency ($\alpha + \beta$) does not show a big difference between the two indices.

5.2. The GARCH models for SPI time series indicated stronger short run persistence (ARCH effect) than long run persistence (GARCH effect) for both short and long-term SPI time series. The intensity of persistency is, however, higher for SPI12 time series (very close to $\alpha + \beta = 1$) than that for SPI3 time series at both stations.

5.3. The conditional variance of the atmospheric and drought indices seems to be stationary but nonlinear. The degree of stationarity is higher for long-term SPI than that for short-term SPI time series while the degree of nonlinearity is higher for short term SPI time series. The same condition for atmospheric indices is observed which suggests an inverse relationship between intensity of nonlinearity and stationarity. This was also reported for conditional variance of rainfall and runoff time series (Modarres and Ouarda, 2012b).

5.4. The multivariate GARCH model which establishes the conditional covariance structure between two time series shows a low level of covariance interaction between atmospheric circulations (SOI and NAO) and SPI time series for our two examples in Iran. Both diagonal VECH and diagonal BEKK models show that the link between the second order moment of atmospheric circulations and drought and the long run persistency of the conditional covariance

between them are weak. On the other hand, the short run interaction is much stronger between drought and atmospheric circulations which indicates a significant relationship between the lagged cross-products of the shocks, or random process, of them. This implies that the variation of SOI and NAO may have a rapid and short-run influence on drought variation at the stations under investigation.

5.5. The conditional correlation coefficients show a range of high to low association between atmospheric circulation and drought time series. It is observed that the correlation coefficients do not remain at the same level (high or low) for the long time before changing to the next (low or high) status. This also suggests a low level of memory in the link between atmospheric indices and drought for our examples. It was also seen that the correlation coefficients do not show a sharp seasonality and trend during 1954-2010.

5.6. The results of the sationarity test for conditional covariances reveal a stationary covariance between drought and atmospheric indices for most cases during 1954-2010. It may indicate that the change in the link between atmospheric circulation and drought does not show a significant change during 1954-2010. However, some covariances cannot pass the trend stationarity test. Although the non-stationarity around a trend could be over-affected by some outlier data which makes the whole series non-stationary around such a basic trend, further tests and careful analysis are necessary to reveal the exact reason for non-stationarity around the trend for the connection between drought and atmospheric indices.

It should be noted that (econometric) MGARCH models show some advantages and disadvantages over hydrological models. The MGARCH models give the opportunity to estimate the link between second order moments of hydrological variables. These models also provide a time varying, step by step, link between two variables by providing the conditional correlation

and covariance estimation for each time step. Therefore, one can see the temporal variation of the association between two variables very easily. However, the number of parameters grows rapidly with the order of the model and parameter estimation becomes a real problem. It is observed from econometric literature that the MGARCH(1,1) model satisfies the conditional covariance structure between financial data. The hydrologic example in this study also indicates a satisfactory MGARCH(1,1) model as the link between second order moments of atmospheric and drought indices is not very strong. The satisfactory MGARCH(1,1) model was also indicated for rainfall-runoff models with a stronger conditional covariance structure (Modarres and Ouarda2012b).

7. Future challenges

First, it is recommended to investigate the relationship between other atmospheric indices and other hydrologic and climatic variables such as other types of drought, rainfall or streamflow. It is also important examine the MGARCH models for hydrologic drought and its relationship to meteorological drought or other atmospheric indices. It would be interesting to apply the MGARCH approach to investigate the effect of the time varying variance of different variables such as rainfall, streamflow, temperature, wind speed, evaporation, etc., on each other. The investigation and evaluation of the volatility and co-volatility between climate and hydrologic variables in the context of climate change is also vital as the second order moment of hydrologic variables may show a higher degree of fluctuations and nonlinearity in the future. This change in a climate variable may influence the other climatic and hydrologic variables in an exponential manner in future. Moreover, the physical variables influencing conditional covariance and the

parameters of the MGARCH models in hydrologic applications remains an important challenge for further studies.

Acknowledgements

The authors thank the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Canada Research Chair (CRC) Program for providing the financial support for this study.

References

Abramowitz, M. and I.A. Stegun, (Eds.), 1972: *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th printing. New York, Dover, 1046 pp.

Baur, D., 2006: A flexible dynamic correlation model. *Econometric analysis of financial and economic time series.* D. Terrel, and T. B. Fomby, Eds., Elsevier publications. Netherlands, 3-31

Bauwens, L., Laurenti, S., and , J. V.K., Rombouts, 2006: Multivariate GARCH models: a survey. *J. Appl. Econom.,* **21**, 79-109

Bollerslev, T., 1986: Generalized Autoregressive Conditional Heteroscedasticity, *J. Econometrics*, **31**, 307–327.

Bollerslev, T., R.F. Engle, and J. M. Wooldridge, 1988: A capital asset pricing model with time varying covariances, *J. Polit. Econ.,* **96,** 116–31.

Bordi, I., and A. Sutera, 2001: Fifty years of precipitation: Some spatially remote teleconnnections. *Water Res. Manage.,* **15,** 247-280.

Brock, W.A., W. D. Dechert, J.A. Scheinkman, and B. LeBaron, 1996: A test for independence based on the correlation dimension, *Econ. Rev.*, **15 (3),** 197–235.

Chaleeraktrakoon, C., 2009: Parsimonious SVD/MAR(1) Procedure for Generating Multisite Multiseason Flows. *J. Hydrol. Eng.*, **14,** 516-527,

Chebana F., T. B. J. Ouarda, 2007: Multivariate L-moment homogeneity test. *Water Res. Res.,* 43, W08406 DOI: 10.1029/2006WR005639

Chen C.H., C. H. Liu, and H. C. Su, 2008: A nonlinear time series analysis using two-stage genetic algorithms for streamflow forecasting. *Hydrol. Processes.* **22,** 3697–3711

Dickey, D.A., and W. A. Fuller, 1979: Distribution of the estimators for autoregressive time series with a unit root, *J. Am. Stat. Assoc.,* **74,** 423–431.

Diebold F.X., and R. S. Mariano, 1995: Comparing predictive accuracy. *J. Business Econ. Stat.,* **13,** 253–263.

Durdu, O. F. 2010: Application of linear stochastic models for drought forecasting in the Buyuk Menderes river basin, western Turkey. *Stoch. Env. Res. Risk. Assess.,* **24,** 1145-1162.

Engle, R. F. 1982: Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation, *Econometrica,* **50:** 987-1008.

Engle, R.F., and K. F. Kroner, 1995: Multivariate simultaneous generalized ARCH, *Econom. Theory* ., **11,** 122–50

Frances, P.H., and D. van Dijk, 2000: Nonlinear time series models in empirical finance. Cambridge University Press. 280p.

Hayes, M.J., M. D. Svoboda, D.A. Wilhite, and O.V. Vanyarkho, 1999: Monitoring the 1996 drought using the Standardized Precipitation Index. *Bull Am. Meteorol. Soc.*, **80(2),** 429–438.

McKee, T.B., N.J. Doesken, and J. Kliest, 1993: The relationship of drought frequency and duration to time scales. In: Proceedings of the eighth conference on applies climatology, Anaheim, CA, 17–22. January 1993. American Meteorological Society, Boston, MA, USA, pp 179–184

Mishra A.K., and V. Desai, 2005: Drought forecasting using stochastic models. *Stoch. Env. Res. Risk. Assess.,* **19,** 326-339.

Mishra, A.K., and V. P. Singh, 2011: Drought modeling - A review. *J. Hydrol.,* **403,** 157-175.

Modarres, R., and T. B. M. J. Ouarda, 2012a. Generalized autoregressive conditional heteroscedasticity modeling of hydrologic time series. *Hydrol. Process.,* (In press Article)

Modarres, R., T. B. M. J. Ouarda, 2012b: Modeling rainfall-runoff relationship using multivariate GARCH model. *J. Hydrol.* (submitted article)

Mohammadi, H., and L. Su, 2010: International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models. *Energy Economic,*. **32,** 1001-1008.

Nazemosadat M.J., and A. R. Ghasemi, 2004: Quantifying the ENSO-related shifts in the intensity and probability of drought and wet periods in Iran. *J. Climate.,* **17,** 4005-4018. DOI: 10.1175/1520-0442(2004)017

Niedzielski, T., 2007: A data-based regional scale autoregressive rainfall-runoff model: a study from the Odra River. *Stoch. Env. Res. Risk. Assess.,* **21,** 649-664. DOI: 10.1007/s00477-006-0077-y

Ozger, M., A.K. Mishra, and V.P. Singh, 2012: Long Lead Time Drought Forecasting Using a Wavelet and Fuzzy Logic Combination Model: A Case Study in Texas. *J. Hydrometeor.,***13**, 284-297, DOI: 10.1175/JHM-D-10-05007.1

Phillips, P.C.B., and P. Perron, 1988: Testing for a unit root in time series regression. *Biometrika,* **75,** 335-334

Raziei, T., B. Saghafian, A. A. Paulo, L. S. Pereira, and I. Bordi, 2009: Spatial Patterns and Temporal Variability of Drought in Western Iran. *Water Res. Manage.,* **23,** 439-455.

Raziei T. and L. S. Pereira, 2012: Estimation of ETo with Hargreaves-Samani and FAO-PM temperature methods for a wide range of climates in Iran. *Agricultural Water Management*, Doi: 10.1016/j.agwat.2012.12.019.

Shiau J. T., and R. Modarres, 2009: Copula-based drought severity-duration-frequency analysis in Iran . *Meteo. Appl.,* **16,** 481-489.

Shih-Chieh K., and R. S. Govindaraju, 2010: A copula-based joint deficit index for droughts. *J. Hydrol.,* **380,** 121-134

Sohail A., K. Watanabe, and S. Takeuchi, 2008: Runoff analysis for a small watershed of Tono area Japan by back propagation artificial neural network with seasonal data. *Water Res. Manage.,* **22,** 1-22.

Vicente-Serrano S.M., 2005: El Niño and La Niña influence on droughts at different timescales in the Iberian Peninsula. *Water Res. Res.,* **41,** W12415 DOI: 10.1029/2004WR003908

Wang W. Vrijling J.K., P.H.A.J.M., Van Gelder and J. Ma, 2006:. Testing for nonlinearity of streamflow processes at different timescales. *J. Hydrol.,* **322,** 247-268

Wang W, J. K. Vrijling, P.H.A.J.M., Van Gelder, and J. Ma, 2005: Testing and modeling autoregressive conditional heteroskedasticity of streamflow processes. *Nonlinear Processes Geophys.* **12,** 55-66

List of Tables

Figure captions

Figure 15 Annual conditional correlation boxplots between SPI3 and SOI (a) and NAO (b) for Shiraz station

Figure 16 same as figure 14 but for SPI12

Table 1 Univariate GARCH model estimations for selected time series

Data	Series	Parameters			Persistency measurement	Order
		ω	α	β	$\alpha + \beta$	
Atmospheric Index	SOI	0.34	0.39	0.34	0.73	GARCH(1,1)
	NAO	0.12	0.05	0.82	0.87	GARCH(1,1)
Oroomieh drought	SPI3	0.51	0.50	0.07	0.57	GARCH(1,1)
	SPI12	0.05	0.89	0.02	0.91	GARCH(1,1)
Shiraz drought	SPI3	0.47	0.49	0.08	0.57	GARCH(1,1)
	SPI12	0.04	0.70	0.29	0.99	GARCH(1,1)

Table 2 Stationarity and Nonlinearity test results for conditional variance

Data	series	PP level stationary test		PP trend stationary test		ADF unit root test	
		Results	*p*-value	Results	*p*-value	Results	*p*-value
Atmospheric Indices	SOI	-12.5	>0.1	0.0001	0.29	-9.57	0
	NAO	-8.03	>0.1	-0.0002	0.28	-6.49	0
Oroomieh	SPI3	-17.37	>0.1	0.0003	0.78	-9.99	0
	SPI12	-4.13	>0.1	0.0003	0.65	-3.65	0.02
Shiraz	SPI3	-14.21	>0.1	-0.0008	0.25	-14.88	0
	SPI12	-5.73	>0.1	-0.001	0.22	-6.07	0
BDS test							
Data	series	m=2		m=3		m=4	
		statistic	*p*-value	statistic	*p*-value	statistic	*p*-value
Atmospheric Indices	SOI	0.06	0	0.10	0	0.12	0
	NAO	0.11	0	0.18	0	0.22	0
Orromieh	SPI3	0.05	0	0.08	0	0.10	0
	SPI12	0.16	0	0.27	0	0.34	0
Shiraz	SPI3	0.07	0	0.10	0	0.11	0
	SPI12	0.16	0	0.27	0	0.34	0

Table 3 diagonal VECH(1,1) estimates for Oroomieh station

Panel a: SOI	***Parameters***			***AIC***
	W	**A**	**B**	
SPI3	$\begin{bmatrix} \mathbf{0.37} & \mathbf{-0.08} \\ & \mathbf{0.44} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.37} & \mathbf{0.48} \\ & \mathbf{0.42} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.32} & \mathbf{-0.13} \\ & \mathbf{0.15} \end{bmatrix}$	5.49
SPI12	$\begin{bmatrix} \mathbf{0.31} & 0.004 \\ & \mathbf{0.04} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.42} & \mathbf{0.56} \\ & \mathbf{0.85} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.33} & \mathbf{0.24} \\ & \mathbf{0.12} \end{bmatrix}$	4.58
Panel b: NAO				
SPI3	$\begin{bmatrix} 0.11 & 0.01 \\ & \mathbf{0.48} \end{bmatrix}$	$\begin{bmatrix} 0.04 & \mathbf{0.10} \\ & \mathbf{0.50} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.84} & \mathbf{-0.80} \\ & 0.03 \end{bmatrix}$	5.56
SPI12	$\begin{bmatrix} \mathbf{0.04} & 0.02 \\ & \mathbf{0.05} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.09} & \mathbf{0.18} \\ & \mathbf{0.86} \end{bmatrix}$	$\begin{bmatrix} -0.13 & -0.10 \\ & 0.05 \end{bmatrix}$	4.86

Note: Entries in bold are significant at the 10% level and less

Substituted coefficients SOI

$$H_{SOI\_SPI3} = -0.08 + 0.48\varepsilon_{1,t-1}\varepsilon_{2,t-1} - 0.13H_{SOI\_SPI3,t-1}$$

$$H_{SOI\_SPI12} = 0.004 + 0.56\varepsilon_{1,t-1}\varepsilon_{2,t-1} + 0.24H_{SOI\_SPI12,t-1}$$

Substituted coefficients NAO

$$H_{NAO\_SPI3} = 0.01 + 0.10\varepsilon_{1,t-1}\varepsilon_{2,t-1} - 0.80H_{NAO\_SPI3,t-1}$$

$$H_{NAO\_SPI12} = 0.02 + 0.18\varepsilon_{1,t-1}\varepsilon_{2,t-1} - 0.10H_{NAO\_SPI12,t-1}$$

Table 4 diagonal BEKK(1,1) estimates for Oroomieh station

Panel a: SOI	***Parameters***			***AIC***
	W	**A**	**B**	
SPI3	$\begin{bmatrix} \mathbf{0.37} & \mathbf{-0.08} \\ & \mathbf{0.49} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.64} \\ & \mathbf{0.68} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.53} \\ & \mathbf{-0.21} \end{bmatrix}$	5.49
SPI12	$\begin{bmatrix} \mathbf{0.32} & 0.004 \\ & \mathbf{0.04} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.64} \\ & \mathbf{0.88} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.58} \\ & \mathbf{0.38} \end{bmatrix}$	4.57
Panel b: NAO	***Parameters***			
SPI3	$\begin{bmatrix} \mathbf{0.11} & \mathbf{0.02} \\ & \mathbf{0.45} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.21} \\ & \mathbf{0.71} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.91} \\ & \mathbf{-0.26} \end{bmatrix}$	5.56
SPI12	$\begin{bmatrix} \mathbf{0.13} & 0.02 \\ & 0.05 \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.19} \\ & \mathbf{0.93} \end{bmatrix}$	$\begin{bmatrix} \mathbf{-0.91} \\ & \mathbf{0.21} \end{bmatrix}$	4.85

Note: Entries in bold are significant at the 10% level and less

Substituted coefficients SOI

$$H_{SOI\_SPI3} = -0.08 + 0.41\varepsilon_{1,t-1}\varepsilon_{2,t-1} - 0.11H_{SOI\_SPI3,t-1}$$

$$H_{SOI\_SPI12} = 0.004 + 0.57\varepsilon_{1,t-1}\varepsilon_{2,t-1} + 0.22H_{SOI\_SPI12,t-1}$$

Substituted coefficients NAO

$$H_{NAO\_SPI3} = 0.02 + 0.15\varepsilon_{1,t-1}\varepsilon_{2,t-1} - 0.23H_{NAO\_SPI3,t-1}$$

$$H_{NAO\_SPI12} = 0.02 + 0.18\varepsilon_{1,t-1}\varepsilon_{2,t-1} - 0.19H_{NAO\_SPI12,t-1}$$

Table 5 Criteria estimates for conditional covariance at Oroomieh station

Covariance series	Diagonal VECH		Diagonal BEKK		DM statistic
	NBIAS	NRSME	NBIAS	NRSME	
SOI-SPI3	0.42	7.97	0.64	3.86	-4.19*
NAO-SPI3	0.82	12.94	-1.5	20.94	-17.4*
SOI-SPI12	-0.48	6.1	0.61	8.3	3.16
NAO-SPI12	-0.63	17.08	-0.8	8.7	-7.31*

*Significant at 5% level and better

Table 6 Stationary and nonlinearity test results for conditional covariances at Oroomieh station

Indices series	Drought series	PP level stationary test		PP trend stationary test		ADF unit root test	
		Results	*p*-value	Results	*p*-value	Results	*p*-value
SOI	SPI3	-17.4	>0.1	0.002	0.84	-15.1	0
	SPI12	-9.7	>0.1	-0.0006	0.47	-7.8	0
NAO	SPI3	-29.2	>0.1	0.004	0.17	-29.4	0
	SPI12	-26.1	>0.1	0.006	0.05	-26.1	0
BDS test							
Indices series	Drought series	m=2		m=3		m=4	
		statistic	*p*-value	statistic	*p*-value	statistic	*p*-value
SOI	SPI3	0.04	0	0.07	0	0.09	0
	SPI12	0.10	0	0.18	0	0.22	0
NAO	SPI3	0.01	0	0.02	0	0.03	0
	SPI12	0.05	0	0.10	0	0.13	0

Table 7 diagonal VECH(1,1) estimates for Shiraz station

Panel a: SOI	***Parameters***			***AIC***
	W	**A**	**B**	
SPI3	$\begin{bmatrix} \mathbf{0.38} & \mathbf{-0.08} \\ & \mathbf{0.49} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.37} & \mathbf{0.38} \\ & \mathbf{0.47} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.30} & 0.06 \\ & \mathbf{-0.08} \end{bmatrix}$	5.27
SPI12	$\begin{bmatrix} \mathbf{0.39} & -0.03 \\ & \mathbf{0.04} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.38} & \mathbf{0.50} \\ & \mathbf{0.84} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.28} & \mathbf{0.24} \\ & \mathbf{0.08} \end{bmatrix}$	4.53
Panel b: NAO	***Parameters***			
SPI3	$\begin{bmatrix} 0.15 & 0.05 \\ & \mathbf{0.47} \end{bmatrix}$	$\begin{bmatrix} 0.04 & \mathbf{0.15} \\ & \mathbf{0.47} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.79} & \mathbf{-0.37} \\ & -0.06 \end{bmatrix}$	5.31
SPI12	$\begin{bmatrix} \mathbf{0.97} & -0.01 \\ & \mathbf{0.05} \end{bmatrix}$	$\begin{bmatrix} 0.07 & \mathbf{0.16} \\ & \mathbf{0.95} \end{bmatrix}$	$\begin{bmatrix} -0.04 & \mathbf{-0.14} \\ & -0.01 \end{bmatrix}$	4.73

Note: Entries in bold are significant at the 10% level and less

Substituted coefficients SOI

$$H_{SOI\_SPI3} = -0.08 + 0.38\varepsilon_{1,t-1}\varepsilon_{2,t-1} + 0.06H_{SOI\_SPI3,t-1}$$

$$H_{SOI\_SPI12} = -0.03 + 0.50\varepsilon_{1,t-1}\varepsilon_{2,t-1} + 0.24H_{SOI\_SPI12,t-1}$$

Substituted coefficients NAO

$$H_{NAO\_SPI3} = 0.05 + 0.15\varepsilon_{1,t-1}\varepsilon_{2,t-1} - 0.37H_{NAO\_SPI3,t-1}$$

$$H_{NAO\_SPI12} = -0.01 + 0.16\varepsilon_{1,t-1}\varepsilon_{2,t-1} - 0.14H_{NAO\_SPI12,t-1}$$

Table 8 diagonal BEKK(1,1) estimates for Shiraz station

Panel a: SOI	***Parameters***			***AIC***
	W	**A**	**B**	
SPI3	$\begin{bmatrix} \mathbf{0.38} & \mathbf{-0.08} \\ & \mathbf{0.44} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.59} & \\ & \mathbf{0.67} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.56} & \\ & 0.04 \end{bmatrix}$	5.27
SPI12	$\begin{bmatrix} \mathbf{0.39} & \mathbf{-0.03} \\ & \mathbf{0.04} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.58} & \\ & \mathbf{0.91} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.56} & \\ & \mathbf{0.34} \end{bmatrix}$	4.59
Panel b: NAO	***Parameters***			
SPI3	$\begin{bmatrix} 0.15 & 0.04 \\ & \mathbf{0.43} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.21} & \\ & \mathbf{0.68} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.89} & \\ & -0.07 \end{bmatrix}$	5.31
SPI12	$\begin{bmatrix} 0.11 & -0.02 \\ & 0.04 \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.17} & \\ & \mathbf{0.97} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.92} & \\ & -0.06 \end{bmatrix}$	4.73

Note: Entries in bold are significant at the 10% level and less

Substituted coefficients SOI

$H_{SOI\_SPI3} = -0.08 + 0.40\varepsilon_{1,t-1}\varepsilon_{2,t-1} + 0.02H_{SOI\_SPI3,t-1}$

$H_{SOI\_SPI12} = -0.03 + 0.53\varepsilon_{1,t-1}\varepsilon_{2,t-1} + 0.19H_{SOI\_SPI12,t-1}$

Substituted coefficients NAO

$H_{NAO\_SPI3} = 0.03 + 0.15\varepsilon_{1,t-1}\varepsilon_{2,t-1} - 0.06H_{NAO\_SPI3,t-1}$

$H_{NAO\_SPI12} = -0.02 + 0.17\varepsilon_{1,t-1}\varepsilon_{2,t-1} - 0.06H_{NAO\_SPI12,t-1}$

Table 9 Criteria estimates for conditional covariance at Shiraz station

Covariance series	Diagonal VECH		Diagonal BEKK		DM statistic
	NBIAS	NRSME	NBIAS	NRSME	
SOI-SPI3	0.24	8.32	0.62	17.45	-3.14*
NAO-SPI3	0.45	9.8	-0.41	51.87	-8.02*
SOI-SPI12	-0.14	12.98	-1.64	54.36	-10.7*
NAO-SPI12	-0.29	13.1	-2.09	12.52	-2.72*

*Significant at 5% level and better

Table 10 Stationary and nonlinearity test results for conditional covariances at Shiraz station

Indices series	Drought series	PP level stationary test		PP trend stationary test		ADF unit root test	
		Results	*p*-value	Results	*p*-value	Results	*p*-value
SOI	SPI3	-17	>0.1	-0.0001	0.02	-17.3	0
	SPI12	-22.9	>0.1	-0.0003	0.25	-22.9	0
NAO	SPI3	-22.9	>0.1	-0.0003	0.25	-22.9	0
	SPI12	-23.8	>0.1	0.0008	0.02	-23.5	0
BDS test							
Indices series	Drought series	m=2		m=3		m=4	
		statistic	*p*-value	statistic	*p*-value	statistic	*p*-value
SOI	SPI3	0.04	0	0.06	0	0.08	0
	SPI12	0.03	0	0.04	0	0.04	0
NAO	SPI3	0.03	0	0.04	0	0.04	0
	SPI12	0.04	0	0.09	0	0.13	0

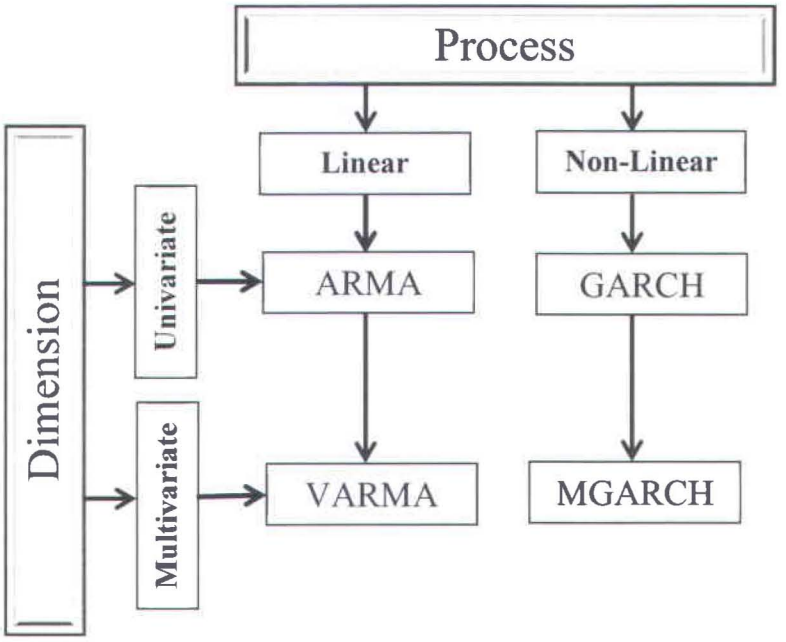

Process
Linear
Non-Linear
Dimension
Univariate
Multivariate
ARMA
GARCH
VARMA
MGARCH

Figure 1

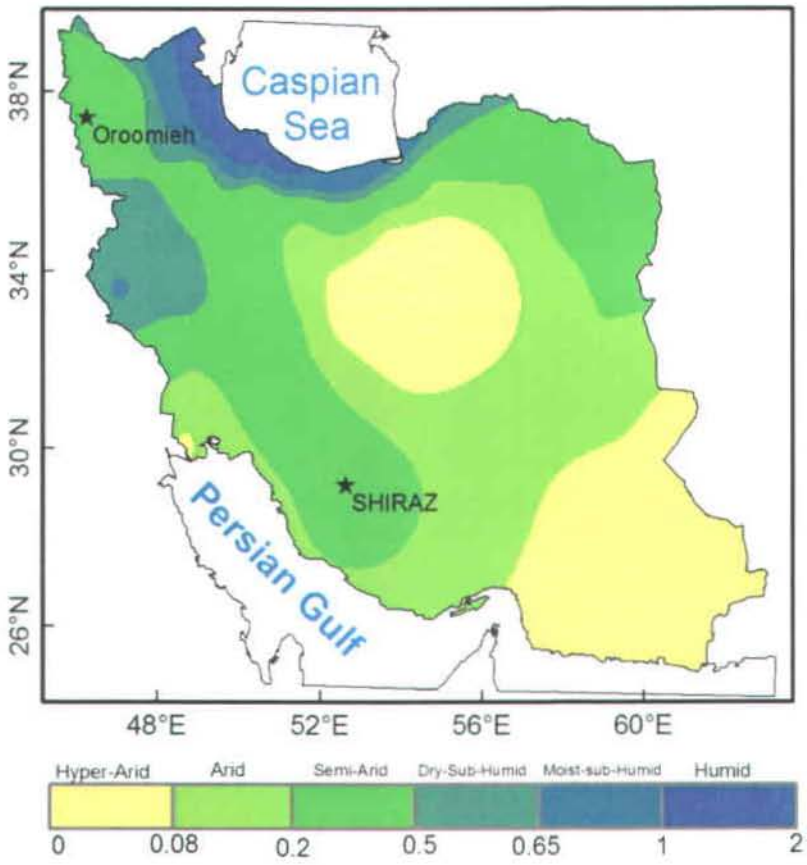

Caspian Sea
Oroomieh
SHIRAZ
Persian Gulf
38°N
34°N
30°N
26°N
48°E
52°E
56°E
60°E
Hyper-Arid
Arid
Semi-Arid
Dry-Sub-Humid
Moist-sub-Humid
Humid
0
0.08
0.2
0.5
0.65
1
2

Figure 2

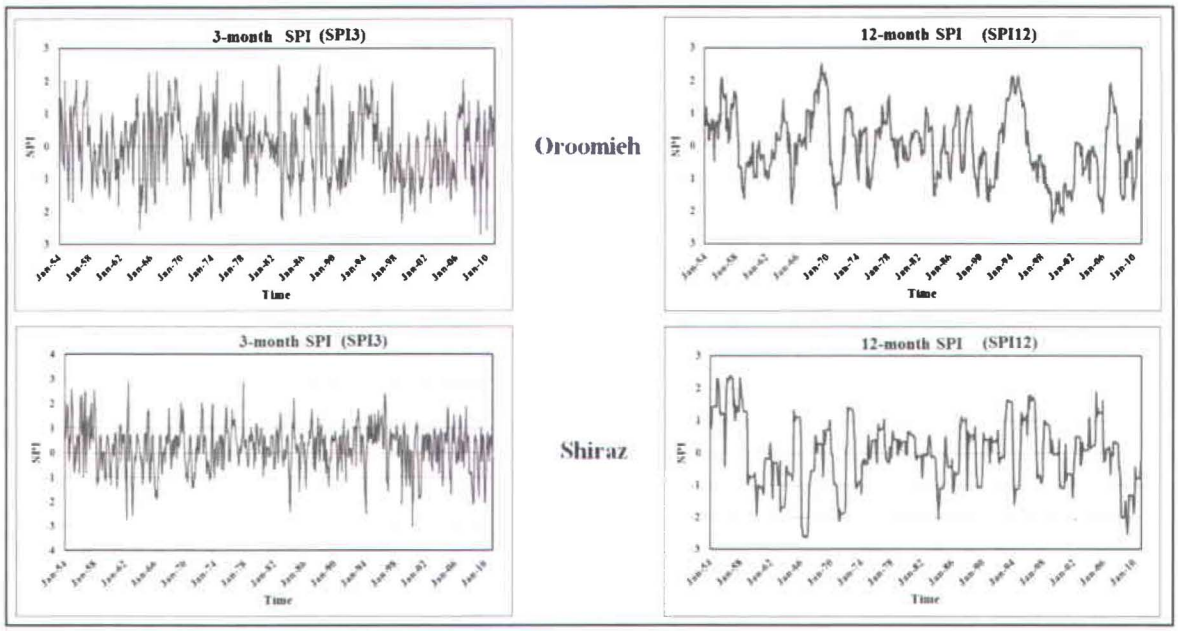

3-month SPI (SPI3)
12-month SPI (SPI12)
Oroomieh
3-month SPI (SPI3)
12-month SPI (SPI12)
Shiraz
SPI
Time

Figure 3

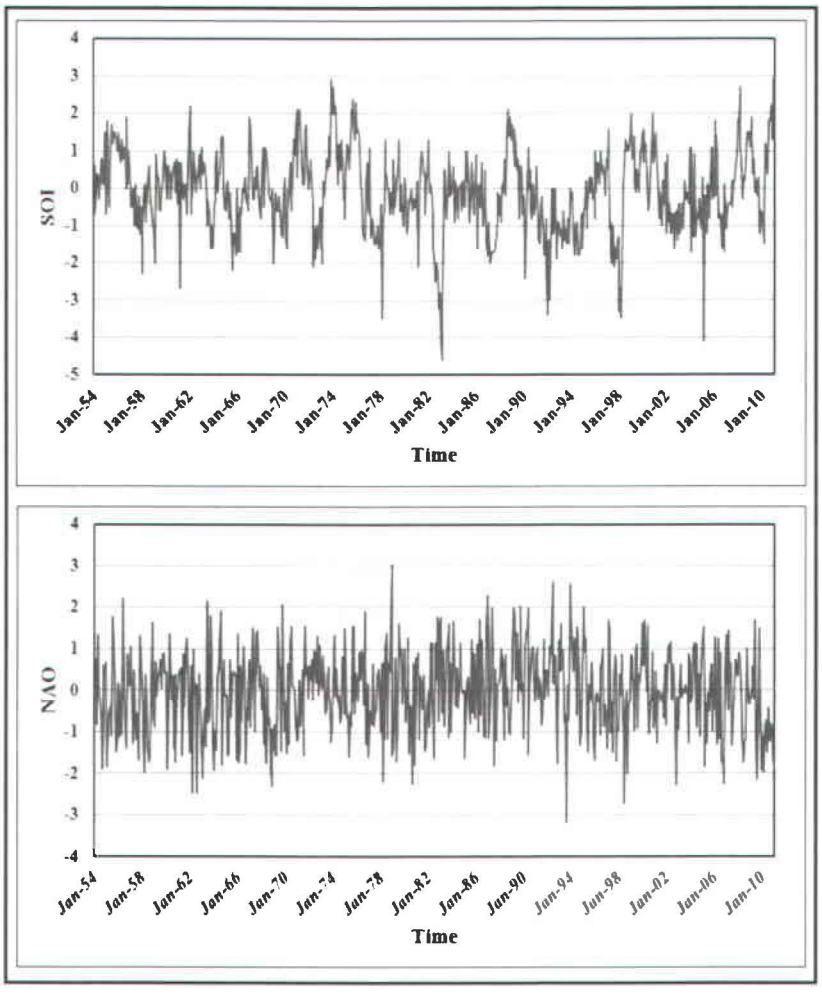

SOI
4
3
2
1
0
-1
-2
-3
-4
-5
Jan-54
Jan-58
Jan-62
Jan-66
Jan-70
Jan-74
Jan-78
Jan-82
Jan-86
Jan-90
Jan-94
Jan-98
Jan-02
Jan-06
Jan-10
Time
NAO
4
3
2
1
0
-1
-2
-3
-4
Jan-54
Jan-58
Jan-62
Jan-66
Jan-70
Jan-74
Jan-78
Jan-82
Jan-86
Jan-90
Jan-94
Jan-98
Jan-02
Jan-06
Jan-10
Time

Figure 4

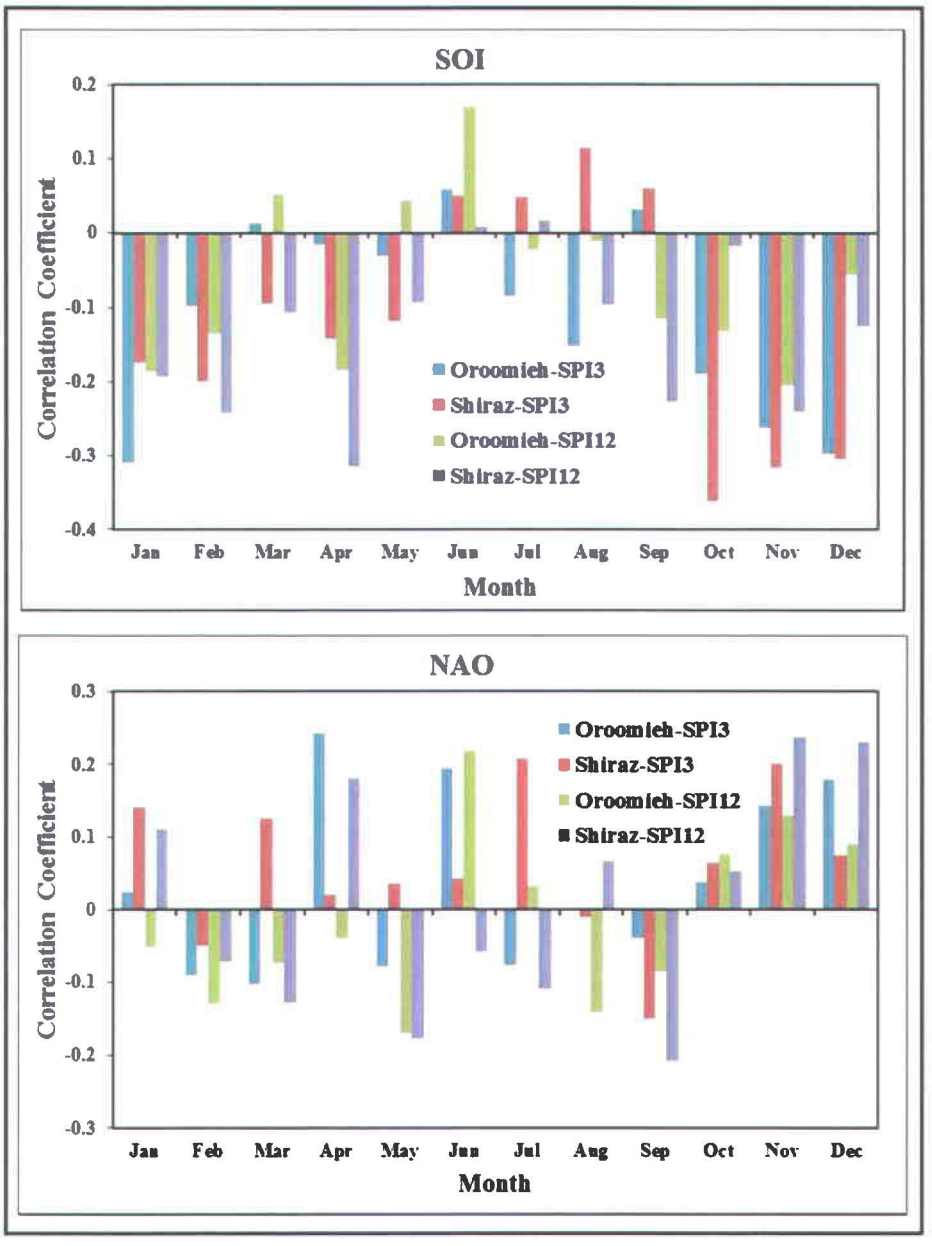

SOI
Correlation Coefficient
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
Oroomieh-SPI3
Shiraz-SPI3
Oroomieh-SPI12
Shiraz-SPI12
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
NAO
Correlation Coefficient
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
Oroomieh-SPI3
Shiraz-SPI3
Oroomieh-SPI12
Shiraz-SPI12
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month

Figure 5

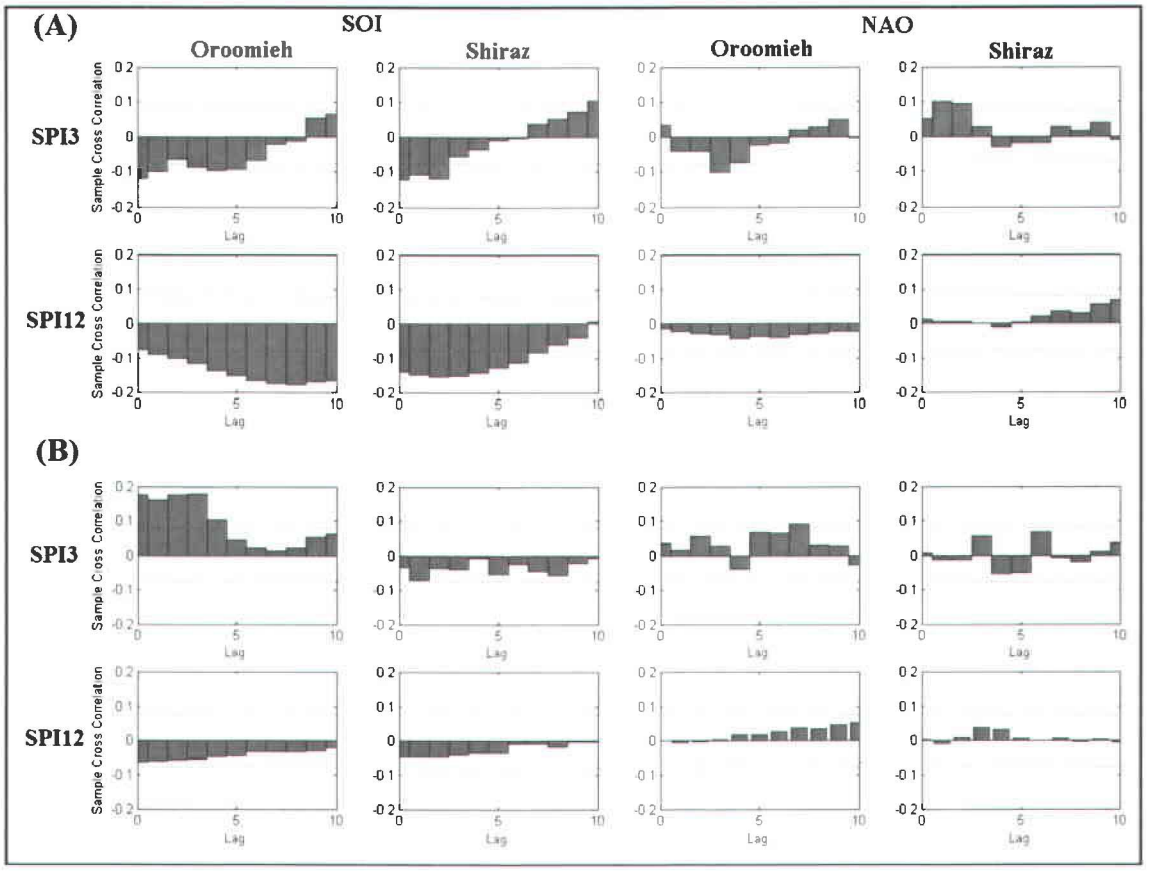

(A)
SOI
NAO
Oroomieh
Shiraz
Oroomieh
Shiraz
SPI3
SPI12
Sample Cross Correlation
Lag
(B)
SPI3
SPI12
Sample Cross Correlation
Lag

Figure 6

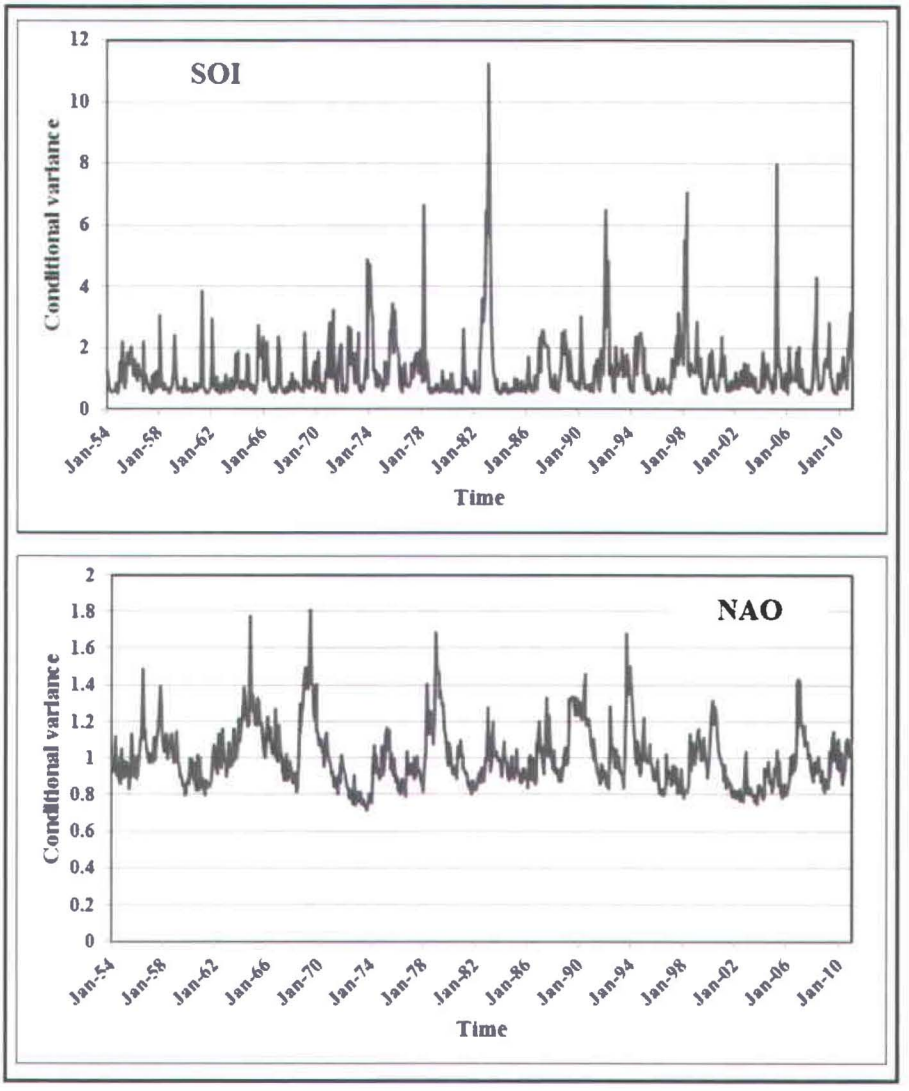

SOI
Conditional variance
12
10
8
6
4
2
0
Jan-54
Jan-58
Jan-62
Jan-66
Jan-70
Jan-74
Jan-78
Jan-82
Jan-86
Jan-90
Jan-94
Jan-98
Jan-02
Jan-06
Jan-10
Time
NAO
Conditional variance
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Jan-54
Jan-58
Jan-62
Jan-66
Jan-70
Jan-74
Jan-78
Jan-82
Jan-86
Jan-90
Jan-94
Jan-98
Jan-02
Jan-06
Jan-10
Time

Figure 7

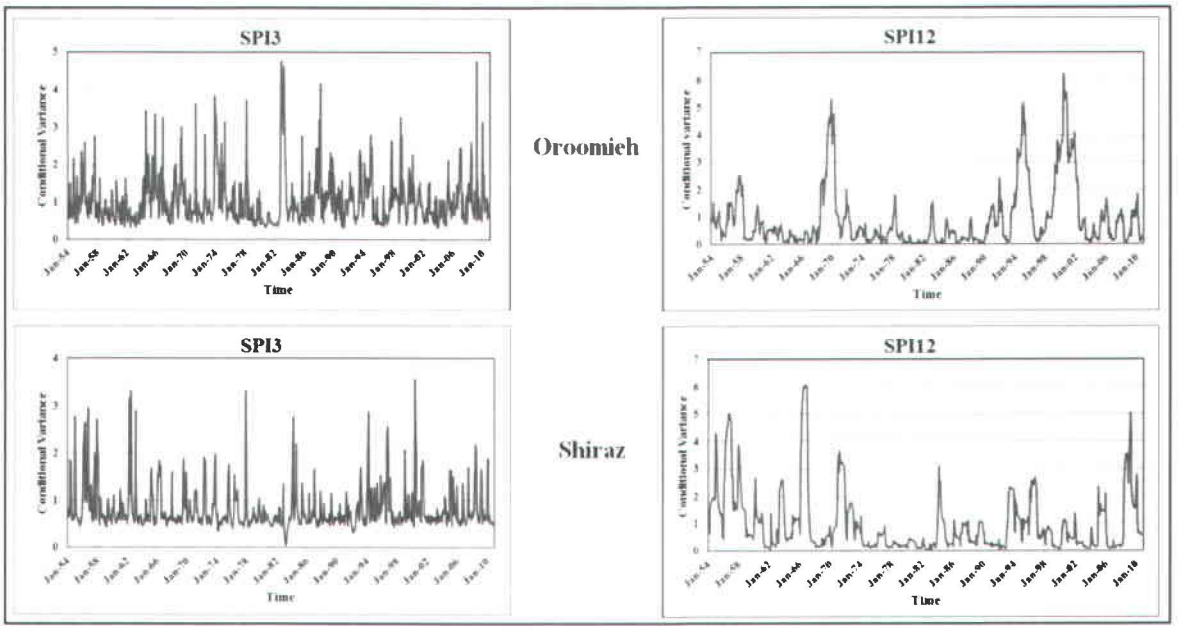

SPI3
SPI12
Oroomieh
SPI3
SPI12
Shiraz
Conditional Variance
Time

Figure 8

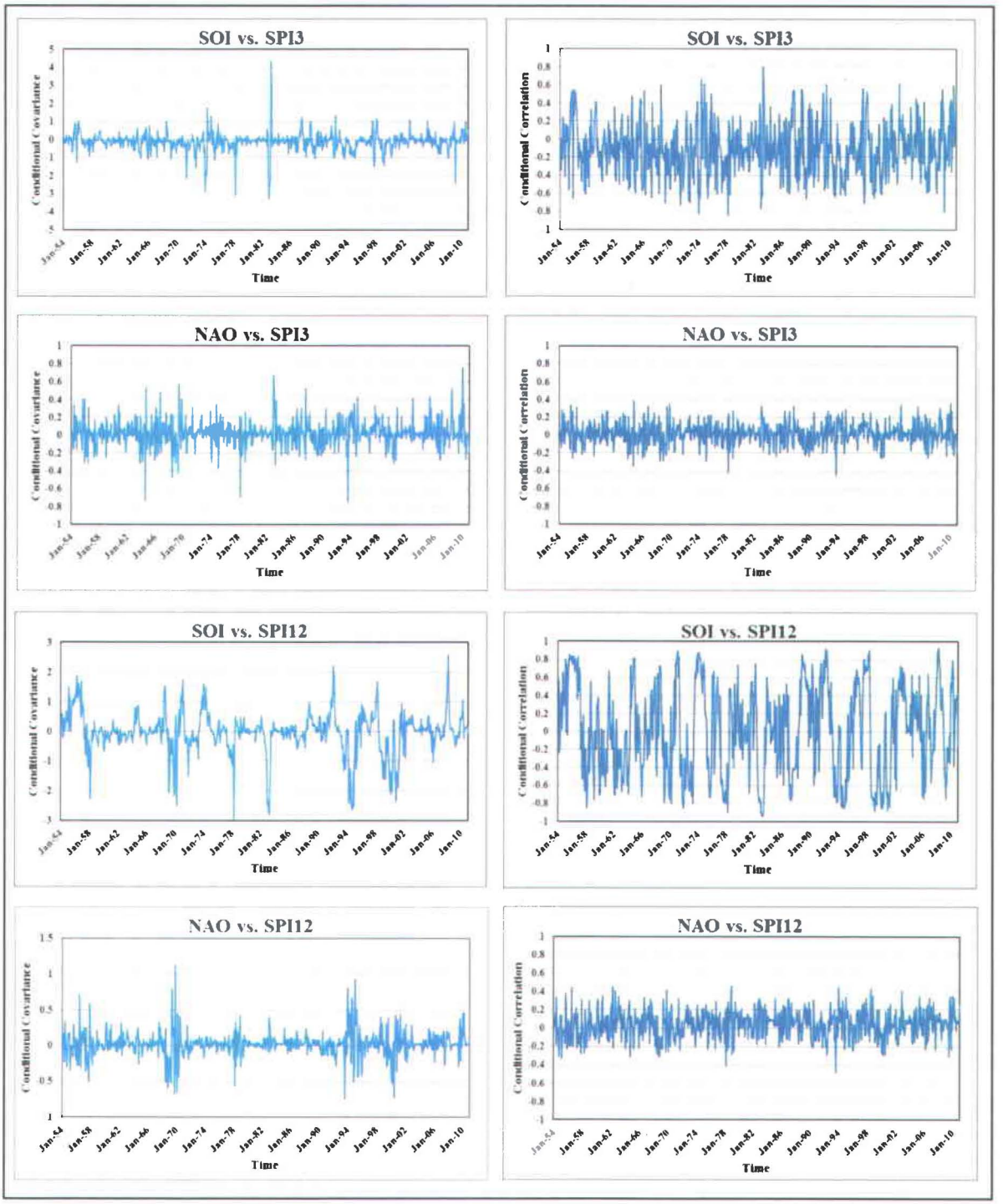

SOI vs. SPI3
Conditional Covariance
Time
SOI vs. SPI3
Conditional Correlation
Time
NAO vs. SPI3
Conditional Covariance
Time
NAO vs. SPI3
Conditional Correlation
Time
SOI vs. SPI12
Conditional Covariance
Time
SOI vs. SPI12
Conditional Correlation
Time
NAO vs. SPI12
Conditional Covariance
Time
NAO vs. SPI12
Conditional Correlation
Time

Figure 9

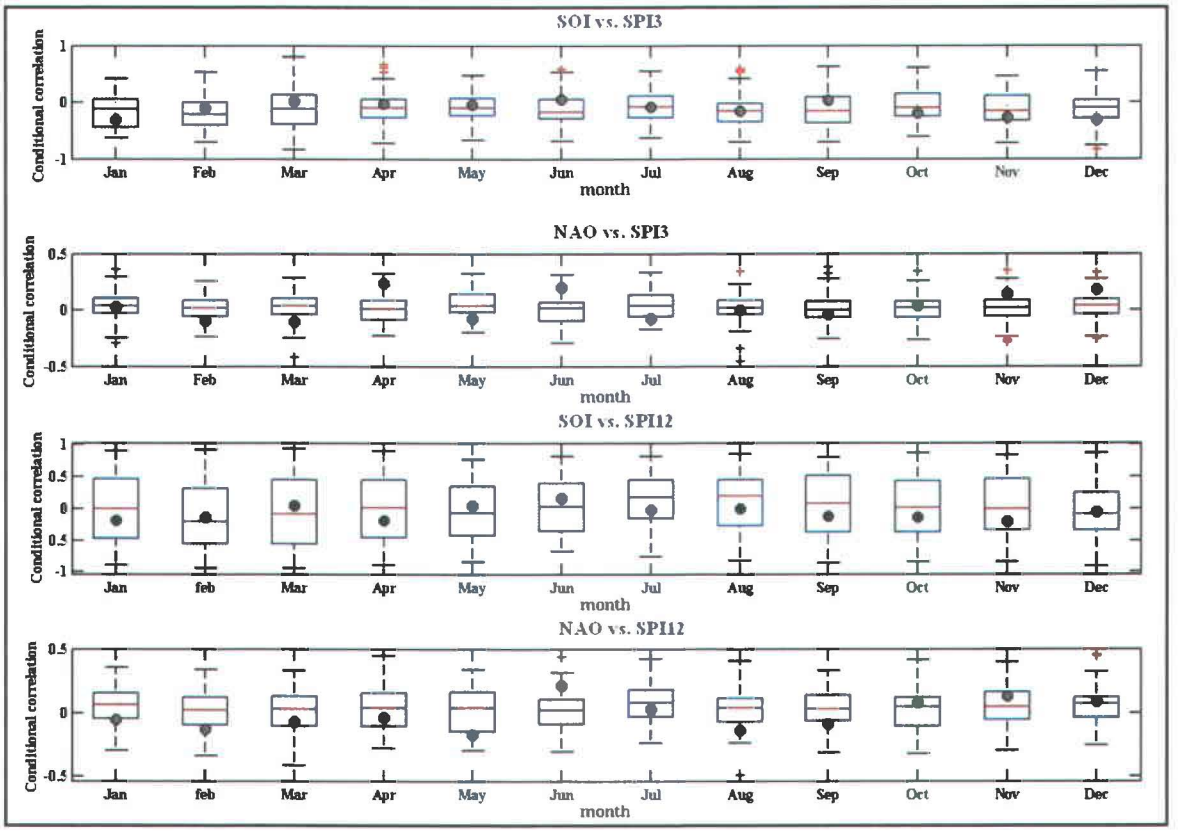

SOI vs. SPI3
Conditional correlation
1
0
-1
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
month
NAO vs. SPI3
Conditional correlation
0.5
0
-0.5
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
month
SOI vs. SPI12
Conditional correlation
1
0.5
0
-0.5
-1
Jan
feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
month
NAO vs. SPI12
Conditional correlation
0.5
0
-0.5
Jan
feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
month

Figure 10

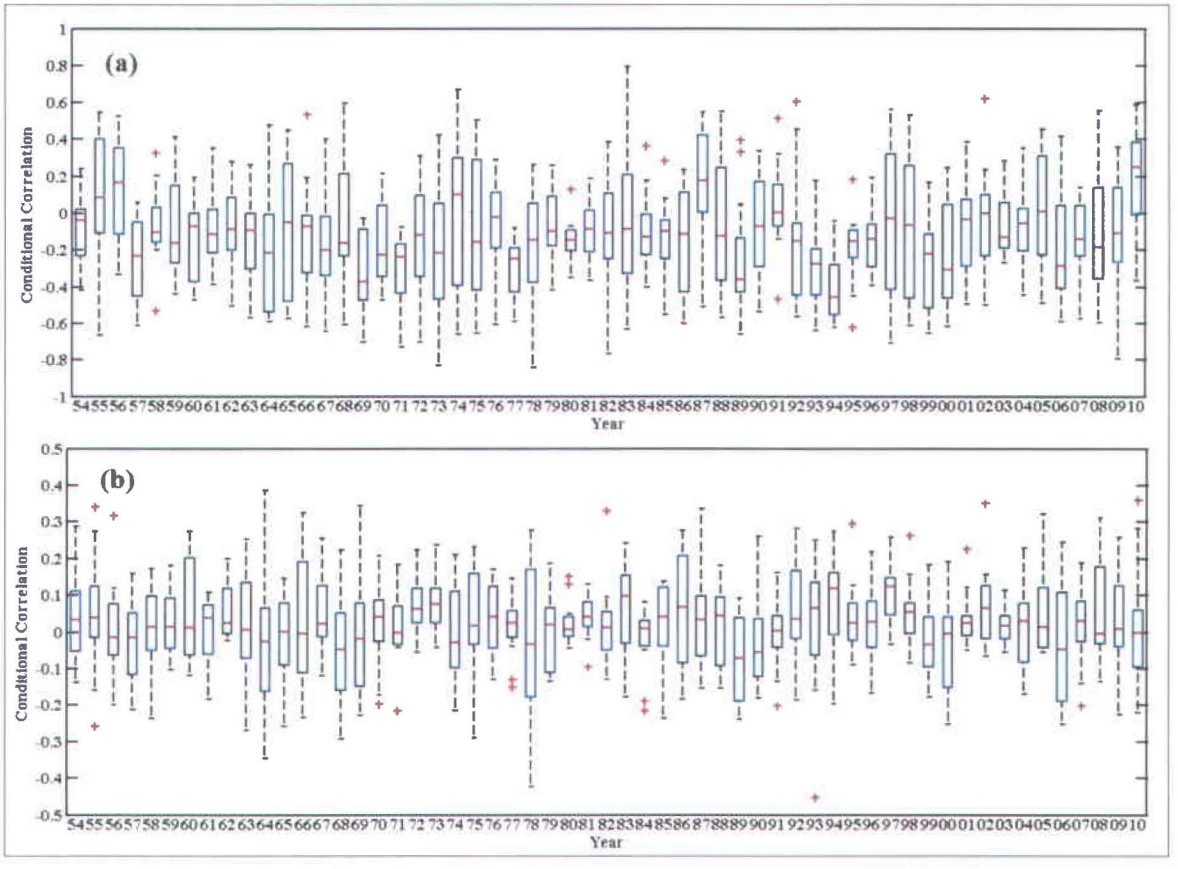

(a)
Conditional Correlation
Year
(b)
Conditional Correlation
Year

Figure 11

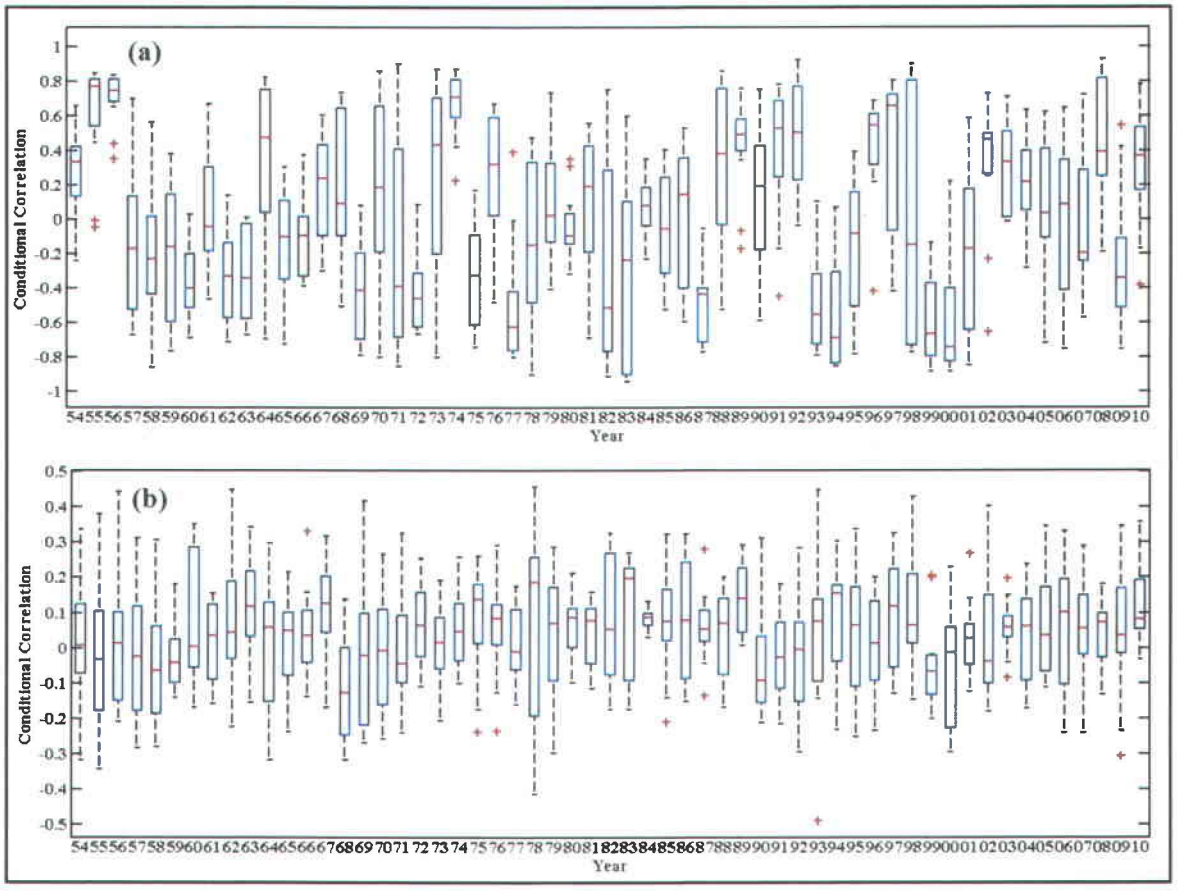

(a)
Conditional Correlation
Year
(b)
Conditional Correlation
Year

Figure 12

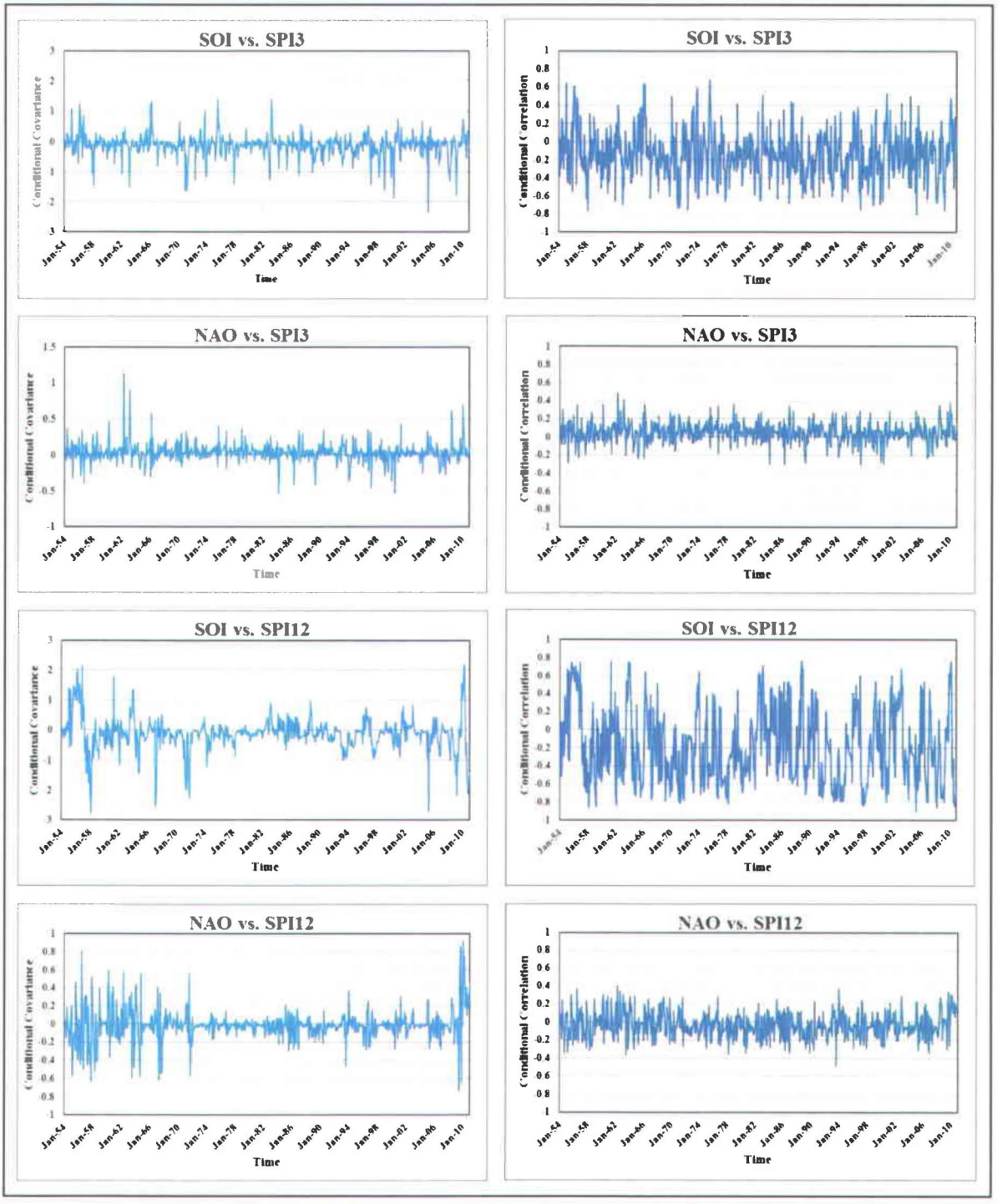

SOI vs. SPI3
Conditional Covariance
Time
SOI vs. SPI3
Conditional Correlation
Time
NAO vs. SPI3
Conditional Covariance
Time
NAO vs. SPI3
Conditional Correlation
Time
SOI vs. SPI12
Conditional Covariance
Time
SOI vs. SPI12
Conditional Correlation
Time
NAO vs. SPI12
Conditional Covariance
Time
NAO vs. SPI12
Conditional Correlation
Time

Figure 13

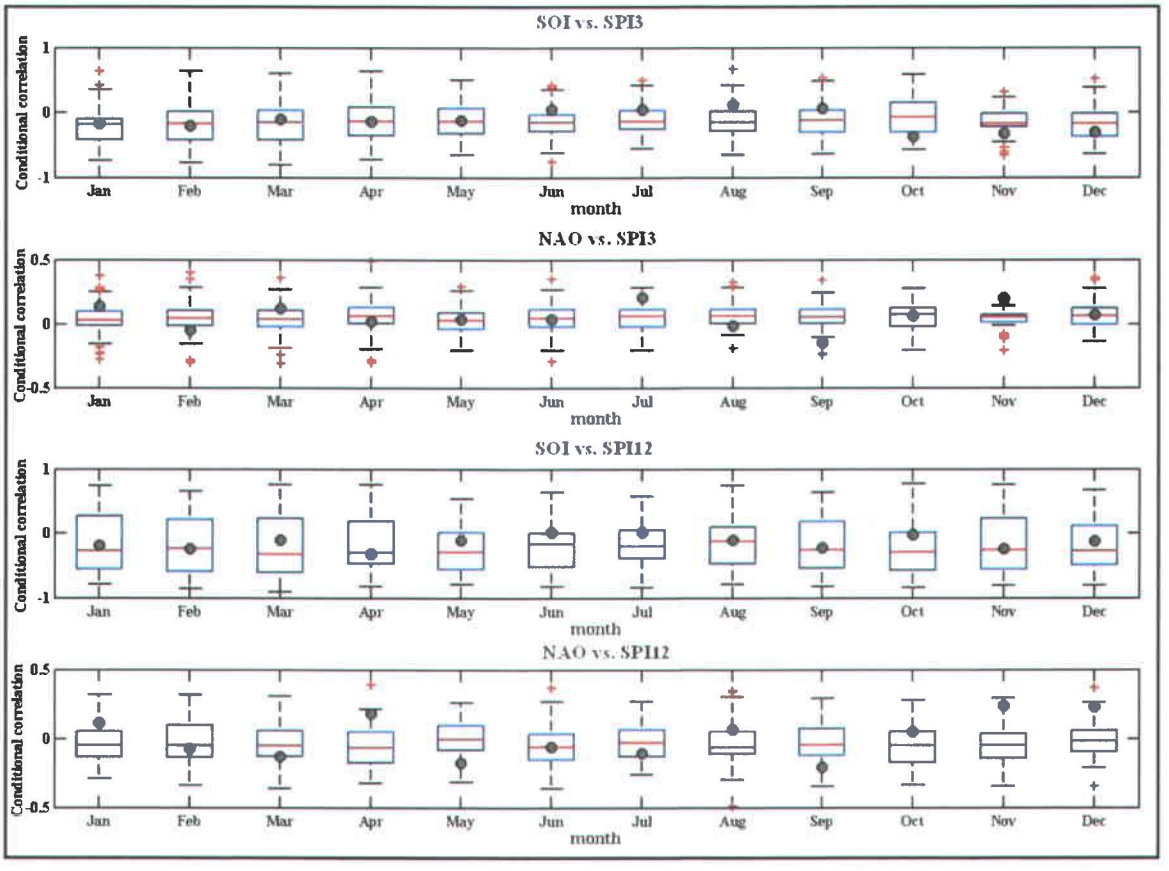

SOI vs. SPI3
Conditional correlation
1
0
-1
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
month
NAO vs. SPI3
Conditional correlation
0.5
0
-0.5
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
month
SOI vs. SPI12
Conditional correlation
1
0
-1
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
month
NAO vs. SPI12
Conditional correlation
0.5
0
-0.5
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
month

Figure 14

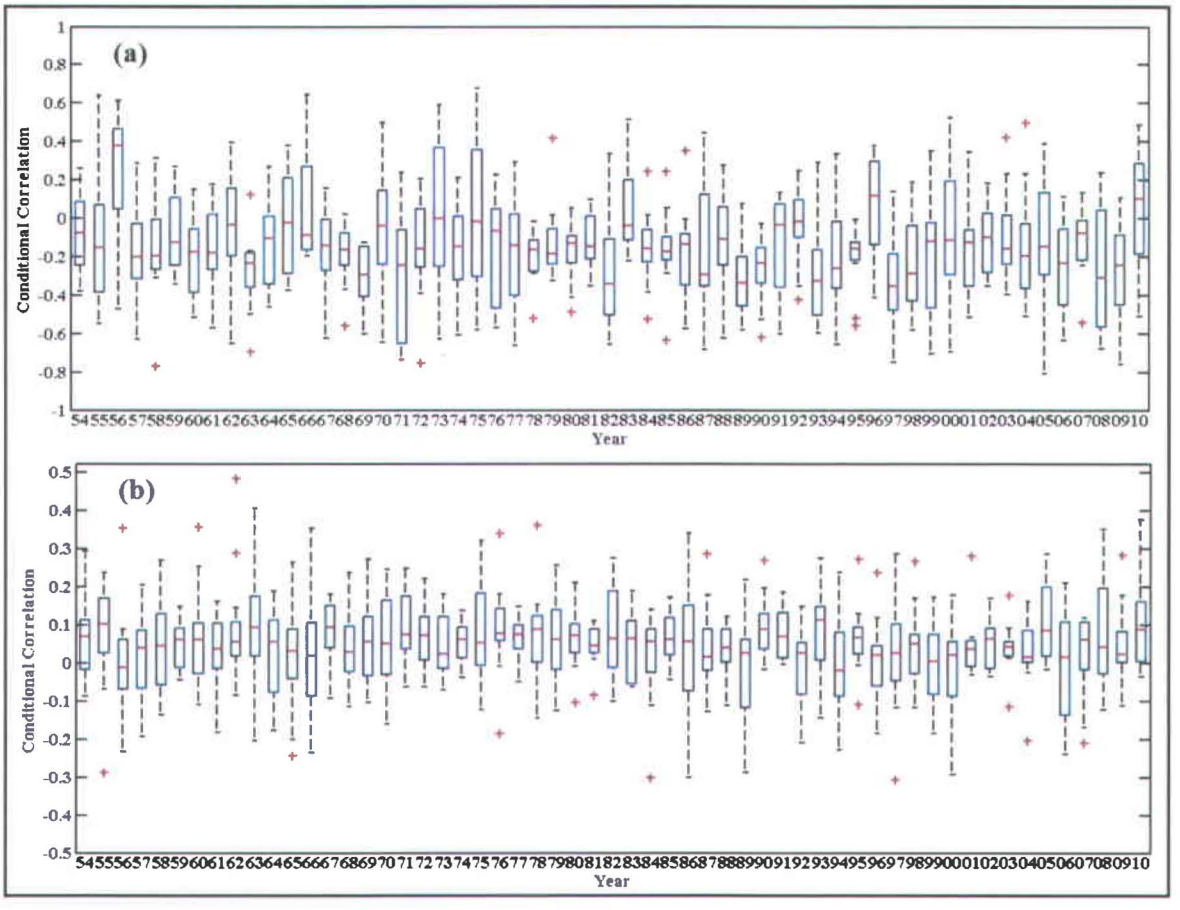

(a)
Conditional Correlation
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
Year
(b)
Conditional Correlation
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
Year

Figure 15

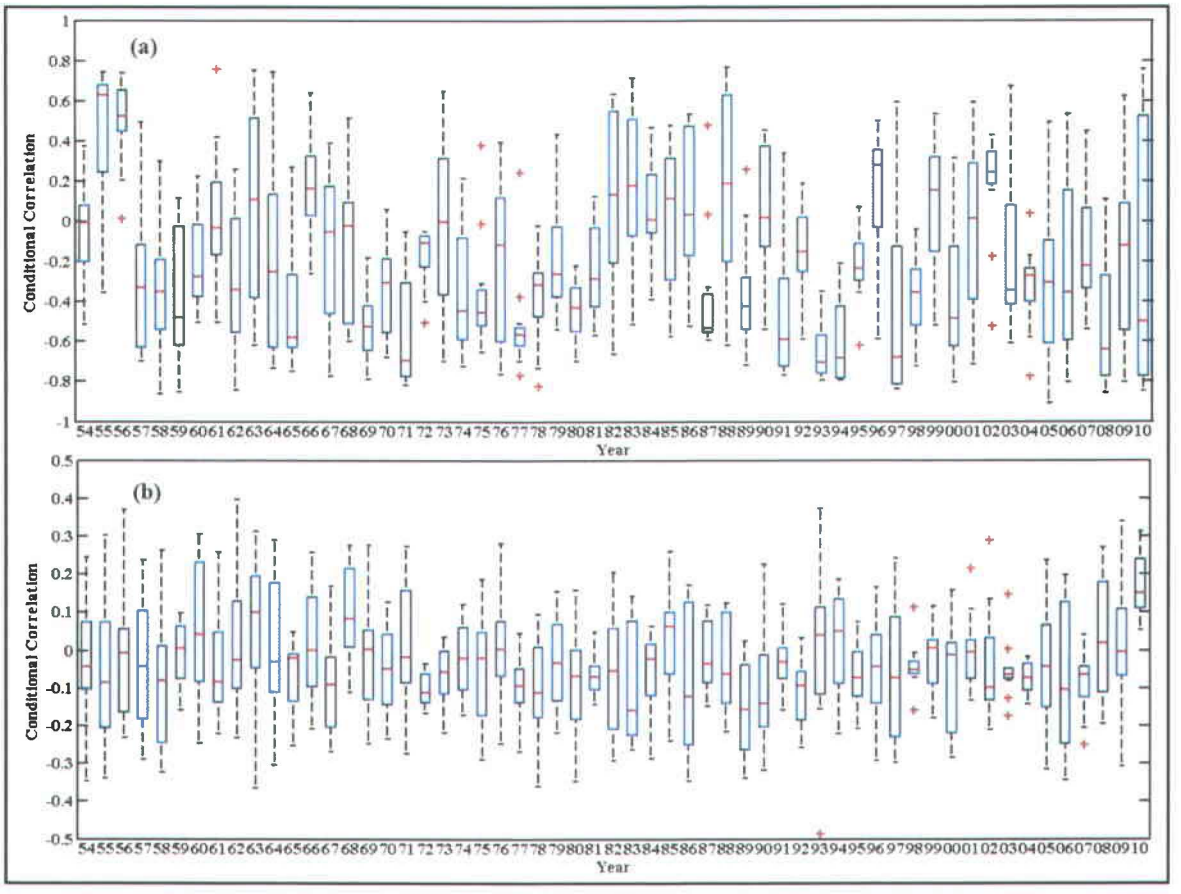

(a)
Conditional Correlation
Year
(b)
Conditional Correlation
Year

Figure 16