Application of Quantile Regression in Climate Change Studies

By

Reza Tareghian

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

Master of Science

Department of Civil Engineering

University of Manitoba

Winnipeg

Abstract

Climatic change has been observed in many locations and has been seen to have dramatic impact on a wide range of ecosystems. The traditional method to analyse trends in climatic series is regression analysis. Koenker and Bassett (1978) developed a regression-type model for estimating the functional relationship between predictor variables and any quantile in the distribution of the response variable. Quantile regression has received considerable attention in the statistical literature, but less so in the water resources literature. This study aims to apply quantile regression to problems in water resources and climate change studies. The core of the thesis is made up of three papers of which two have been published and one has been submitted. One paper presents a novel application of quantile regression to analyze the distribution of sea ice extent. Another paper investigates changes in temperature and precipitation extremes over the Canadian Prairies using quantile regression. The third paper presents a Bayesian model averaging method for variable selection adapted to quantile regression and analyzes the relationship of extreme precipitation with large-scale atmospheric variables. This last paper also develops a novel statistical downscaling model based on quantile regression. The various applications of quantile regression support the conclusion that the method is useful in climate change studies.

Acknowledgments

First and foremost, I would like to thank my advisor, Dr. Peter Rasmussen, for having given to me such a wonderful opportunity to pursue this research. I am thankful for his inspirational support and critical advice during the course of my research. I am sincerely indebted to Dr. Sung Joon Kim for his support and inspiration during the compilation of this thesis. I am grateful to Dr. Jay Anderson for generously sharing his time and knowledge. I would also like to thank Dr. Attahiru Alfa and Dr. Youssef Loukili for their invaluable time to review my thesis and make suggestions and comments.

I am also sincerely grateful to Manitoba Hydro and the Natural Sciences and Engineering Research Council of Canada (NSERC) for providing financial support of my study which enabled me to undertake a good proportion of my research and to successfully complete my studies.

I want to thank all my friends, especially the Pichak Group who made life in Winnipeg enjoyable for me. A special thanks to my best friends, Alireza Hejazi, Amir Khorsand, Amir Hossein Birjandi, Ehsan Azordegan, Jinke Xu, Maziar Heidari, Gol Azordegan and Pooya Taheri, whose supports and encouragement have made this research possible. I also wish to thank our department caretaker, Peter, who inspired me a lot during the hard days.

Lastly, I am deeply thankful to my parents who gave me love, strength and support in all stages of my life. I would not have been where I am now without their unconditional support. I would also like to thank my sisters, Vajiheh, Hanieh, and Khatereh, for always being there.

To my family

Table of Contents

List of Tables

List of Figures

List of papers included in this thesis

I. Tareghian, R., and Rasmussen, P. 2012. Analysis of Arctic and Antarctic sea ice extent using quantile regression. International Journal of Climatology, DOI: 10.1002/joc.3491.

II. Tareghian, R., Rasmussen, P.F., Anderson, J., and Kim, S.J. 2010. A study of climate extremes changes over the Canadian prairies using quantile regression. Proceedings, Annual Conference - Canadian Society for Civil Engineering, 1, pp. 55-63.

III. Tareghian, R., and Rasmussen, P. 2012. Statistical Downscaling of Precipitation using Quantile Regression. Submitted to Journal of Hydrology.

Chapter 1

Introduction

Climatic change has been observed in many locations and has been seen to have dramatic impact on a wide range of ecosystems (Walther et al., 2002). The Fourth IPCC report showed that human-induced climate change will be a major additional stress in a world where the environment is already being seriously damaged and depleted by increasing resource demands and non-sustainable management practices. The report also indicated that climate change will directly impact the Canadian environment and society, and also indirectly as a result of impacts elsewhere in the world. Major impacts of climate change are expected on water systems, as well as on natural ecological systems, forests, agriculture, and coastal ecosystems. Moreover, there is a serious threat to human health. In many areas, there is a danger of increased exposure to heat stress, weather hazards such as droughts, floods, wildfires, and severe storms, with resulting injuries, deaths, and damage to infrastructure that supports public health (IPCC, 2007).

Observations in many areas have shown that changes in total precipitation are significantly influenced by changes in the tails of precipitation distributions. Similarly, changes in some temperature extremes have been observed. There are some lingering questions about whether such changes are part of decadal fluctuations, or whether they are indicative of long-term trends related to climate change (Easterling et al., 2000). Generally, a significant decrease in the number of days with extreme cold temperatures, an increase in the number of days with extreme warm temperatures, and some detectable increase in the number of extreme wet days have been observed in many parts of the world (Vincent and Mekis, 2006). Whether the observed climate of the Earth is becoming more variable and more extreme is one of the key questions in the review reports of the Intergovernmental Panel on Climate Change (IPCC, 2007).

The traditional method to analyse trends in climatic series, to examine the relationships between climate variables (e.g., link between atmospheric circulation variables and local weather variables, such as precipitation and temperature), or to forecast weather is regression analysis. The purpose of regression analysis is to establish a relationship between a response variable and one or more predictor variables. However, the traditional regression analysis most often used to compute climate trends characterizes only the trend of the mean value of the dependent variable (meteorological variable) for a given set of independent variables. An assumption in traditional regression technique is the so-called homoscedasticity of observational data, i.e. the variance of the error term is assumed to be independent of the value of the covariates which is often not satisfied. If heteroscedasticity is present but is ignored, there will be obvious errors in the conditional distribution.

Another peculiarity of the OLS regression technique when analyzing climate trends is the lack of focus on the entire conditional distribution of meteorological variables for every value of the independent variable. The remark of Mosteller and Tukey (1982) is to the point: "…All that the regression curve can give is the generalization for average values of dependent variable Y distributions corresponding to every value of the independent variable X. One could go further and plot several regression curves, corresponding to various percentile values of distributions, having obtained, ipso facto, more complete picture of available data. It is not done, usually, and, therefore, the regression curve gives rather incompletely such a picture just as the average value gives the incomplete picture of the distribution of values of one variable."

Also, heavy-tailed distributions commonly occur in climate studies (e.g. precipitation modeling), leading to a preponderance of outliers. The conditional mean can then become an inappropriate and misleading measure of central location because it is heavily influenced by outliers.

One common characteristic of climate variables is the existence of unequal variation. In standard statistical methods, unequal variation is considered an inconvenience and transformations are often used to equalize variances (e.g., fourth root transformation for precipitation). An alternative view is that unequal variation in climate variables reflects important processes and should be modeled explicitly.

Koenker and Bassett (1978) developed a regression-type model for estimating the functional relationship between predictor variables and any quantile in the distribution of the response variable. Quantile regression overcomes some of the limitations of standard regression and is suitable for the study of changes in the frequency of environmental extremes over time. Quantile regression produces different rates of change in different quantiles of the response variable, providing a more complete picture of the relationships between variables missed by other regression methods. The particular focus of quantile regression is to estimate quantiles of the conditional distribution of a response variable. It differs from traditional linear regression in a number of ways, as will be explained in Chapter 2, and it largely overcomes the problems with traditional regression enumerated above.

Barbosa (2008) applied quantile regression to characterize long-term sea level variability in the Baltic Sea. She concluded that quantile trends provide a more complete description of regional sea-level long-term variability than OLS regression. Baur et al. (2004) used the conditional quantile regression approach for the interpretation of the nonlinear relationships between daily maximum 1-h ozone concentrations and both meteorological and persistence information at four stations in Greece. They concluded that quantile regression has the advantage of easy implementation and transparency of results, which is in contrast to black box models such as neural networks where the predictor-predictand relationships are less clear. Dunham et al. (2002) analyzed the relationship of the abundance of Lahontan cutthroat trout to the ratio of stream width to depth. They found a negative relationship in upper tails while a least squares regression estimated zero change. They remarked that if they used mean regression estimates, their conclusion would mistakenly be that there is no relation between trout densities and the ratio of stream width to depth. Bremnes (2004) made precipitation forecasts in terms of quantiles to improve or enrich numerical weather prediction outputs. They demonstrated how reliable probabilistic precipitation forecasts in terms of quantiles can be made using quantile regression.

Quantile regression has received considerable attention in the statistical literature, but less so in the water resources literature. This study aims to apply quantile regression to problems in water resources and climate change studies. Special objectives of this study are:

- Present a novel application of quantile regression to analyze the distribution of sea ice extent.
- Investigate changes in temperature and precipitation extremes over the Canadian Prairies using quantile regression.
- Present a Bayesian model averaging method for variable selection adapted to quantile regression.
- Analyze the relationship of extreme precipitation with large-scale atmospheric variables using quantile regression.
- Develop a novel statistical downscaling model based on quantile regression.

Case 1: Antarctic Minimum Sea Ice Extent

In order to investigate changes in climate extremes, the slope of the 20th, 50th, and 80th quantile regression lines of the 32 years (1979-2010) of annual minimum ice extent for Antarctic have been estimated and compared with trends in the mean slope (Figure 1.1). The slope of the 80th percentile line for Antarctic minimum ice extent is much larger than the slope of the mean and lower quantiles, suggesting increased variability.

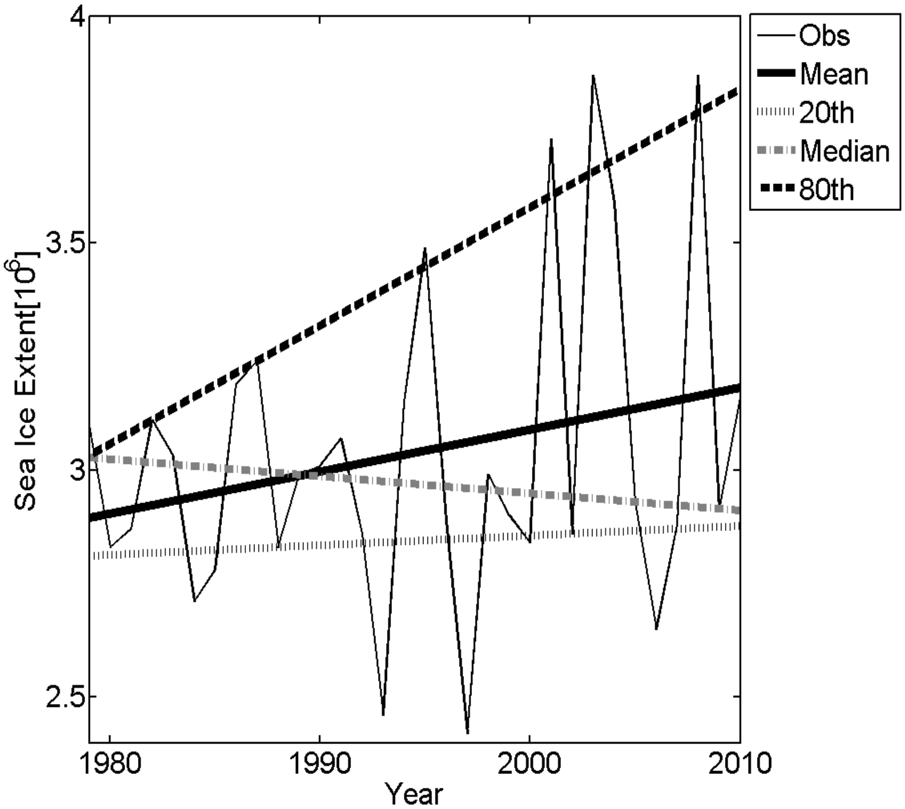

Figure 1.1. Quantile regressions (20th, 50th, and 80th percentiles) and standard linear regression of Antarctic minimum sea ice extent.

In Chapter 3, the paper "Analysis of Arctic and Antarctic Sea Ice Extent using Quantile Regression" is presented (Tareghian and Rasmussen, 2012). Sea ice is an important component of the global climate system. Global climate change affects the Arctic and Antarctic in different ways. The change in the sea ice extent is often assessed using linear trend models estimated by ordinary least squares regression. In this study, a novel application of quantile regression is presented to analyze other aspects of the distribution of sea ice extent.

Case 2: Dauphin Annual Precipitation

The 5th and 95th quantiles of 45 years (1945-1990) of Dauphin annual precipitation have been estimated by quantile regression and compared with changes in the trend of mean levels. As mentioned above, the application of standard regression for modeling extreme events may lead to incorrect conclusions regarding percentiles in the conditional distributions. As illustrated in Figure 1.2, annual precipitation in Dauphin show no particular change in the mean over time, but there is evidence that the precipitation distribution is narrowing over time

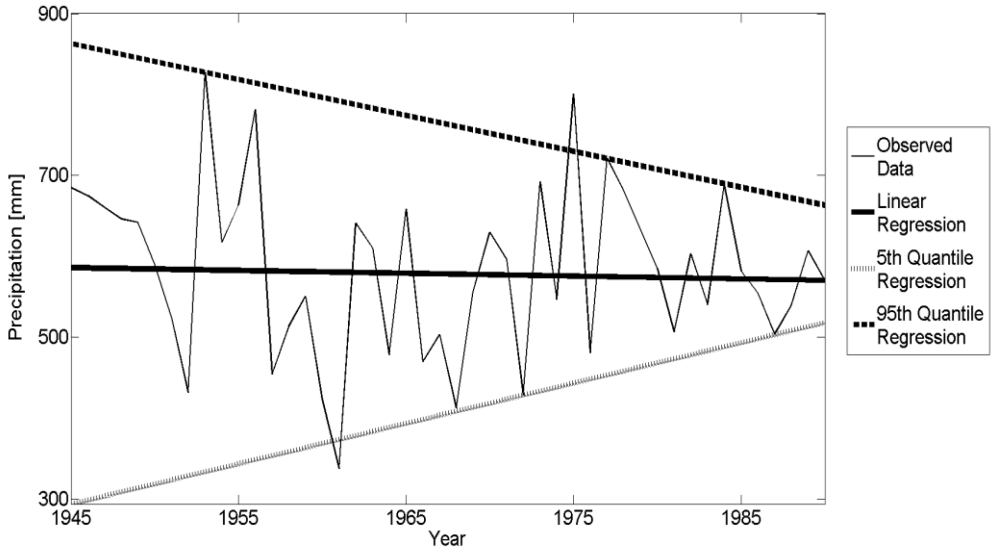

Figure 02.Quantile regressions (5th and 95th percentiles) and standard linear mean regression trends for Dauphin annual precipitation.

In Chapter 4, the paper "A Study of Climate Extremes Changes over the Canadian Prairies using Quantile Regression" is presented (Tareghian et al., 2010). The economic cost of extreme events over the Canadian Prairies, particularly in the agricultural, environmental, and hydroelectric sectors, calls for a detailed investigation of changes in extremes. This study investigates changes in temperature and precipitation extremes over the Canadian Prairies using quantile regression.

Case 3: Comparison of Winnipeg Winter and Summer Precipitation with 500hPa Air Temperature (AIRTEMP500)

Rainfall from thunderstorms is a major contributor to summer precipitation in Winnipeg. Thunderstorms need unstable air, characterized by a temperature profile with warm air near the ground and cold air aloft. As seen in Figure 1.3, the relationships between precipitation and AIRTEMP500 for summer and winter are similar in nature, but the quantile slopes are much higher for summer compared to winter for Winnipeg extreme precipitation (b_1 – values of 2.60 and 0.65 for summer and winter, respectively, in the 98th quantile regression). This conclusion cannot be reached from simply looking at the mean slopes.

In Chapter 5, the paper "Statistical Downscaling of Precipitation using Quantile Regression" is presented (Tareghian and Rasmussen, 2012). Although predictor selection is one of the most important components in the development of any statistical downscaling model, it is often approached in a rather superficial way. In this study, we employed a Bayesian model averaging method, adapted to quantile regression, for variable selection. The advantage of this method is that different predictors can be selected for different parts of the conditional distribution. The variance underestimation and poor representation of extreme events by statistical downscaling methods motivated us to develop a novel statistical downscaling model based on quantile regression. The concluding chapter includes a discussion on how this thesis, with its findings, provides a distinct contribution to knowledge in the research area.

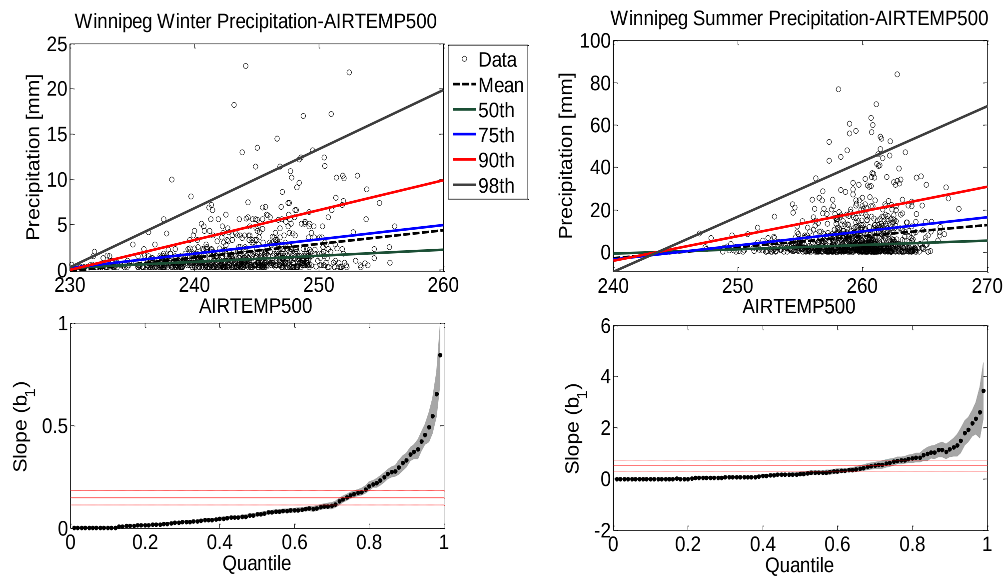

Figure 1.3. Relationship between daily precipitation in Winnipeg and AIRTEMP500 for winter (left) and summer (right). The bottom plots show the slopes of the estimated regression lines. The quantile regression coefficients (black dots) are presented with their 95% confidence bounds (shaded in grey). The least-squares regression coefficients (solid red line) are also given with their 95% confidence bounds (dashed red lines). The vertical axis shows the slopes (%), and the horizontal axis shows the p-value of the quantile (1-99th).

Chapter 2

Quantile Regression

In this chapter, the fundamental principles of quantile regression will be presented. The chapter starts with some basic definitions related to quantile functions. Next, the concepts are extended to a regression setting. Various properties of quantile regression will be illustrated, including, two methods for calculating confidence intervals for quantile regression. Finally, a number of studies on climate change using quantile regression will be listed.

2.1 Basics

2.1.1 Quantiles and Quantile Functions

The median is perhaps the best-known example of a quantile. The median is the middle value of a set of ranked data. In other words, the median (m) splits the ordered data into two parts with an equal number of data points in each. The median of a random variable Y may be defined by the probability statement $P(Y \le m) = P(Y > m) = 1/2$. Any real-valued random variable Y may be characterized by its (right-continuous) distribution function

$$F(y) = P(Y \le y) \tag{2.1}$$

whereas for a proportion p $(0 < p < 1)$,

$$F^{-1}(p) = \inf\{y : F(y) \ge p\} \tag{2.2}$$

is called the pth quantile of Y. The median, $F^{-1}(1/2)$, defines the central location, and other quantiles can be used to describe non-central positions of a distribution.

The general procedure to find the desired quantile of a sample is to sort and rank the observations and then check at which observation the threshold is reached. Koenker and Bassett (1978) showed that determination of a quantile may alternatively be done by optimizing a (weighted) loss function of the form:

$$\mathrm{F}^{-1}(\mathrm{p}) = \underset{\xi_p \in \Re}{\arg\min} \sum_{i \in \{i | y_i \ge \xi_p\}} p\left|y_i - \xi_p\right| + \sum_{i \in \{i | y_i < \xi_p\}} (1-p)\left|y_i - \xi_p\right| \tag{2.3}$$

In other words, the absolute value of the difference between an observation y_i and the unknown optimal value ξ_p is weighted by $(1-p)$ if the observation is below the optimum and by p if the observation is above the optimum (Schulze, 2004). This concept of optimizing a loss function will be employed later in this chapter for estimating the quantile regression coefficients. To illustrate Koenker and Bassett's (1978) new approach to quantile estimation, twenty random observations from a standard normal distribution were drawn and sorted in ascending order (column 2 in Table 2.1). The two sums in

Equation (2.3) were evaluated at each observation, i.e. assuming $\xi_p = y_i$, for p=0.80 and listed in columns 3 and 4. The resulting values of the composite objective function are listed in the last column. Columns 3-5 are shown graphically in Figure 2.1. It can be seen from Table 2.1 and Figure 2.1 that any value in the interval [0.49430, 0.81860] minimizes the objective function in Equation (2.3) and therefore qualifies as an estimate of the 80th sample quantile.

Table 2.1. Evaluation of the objective function in Equation (2.3) for a random sample (n=20) taken from a standard normal distribution for p=0.80.

i	y_i	Leftsum	Rightsum	Objfun
1	-1.48310	22.44736	0	22.44736
2	-1.48140	22.42152	0.00034	22.42186
3	-1.09660	16.88040	0.15426	17.03466
4	-1.02030	15.84272	0.20004	16.04276
5	-0.65680	11.18992	0.49084	11.68076
6	-0.54080	9.79792	0.60684	10.40476
7	-0.44700	8.74736	0.71940	9.46676
8	-0.30860	7.30800	0.91316	8.22116
9	-0.29260	7.15440	0.93876	8.09316
10	-0.29000	7.13152	0.94344	8.07496
11	0.10970	3.93392	1.74284	5.67676
12	0.12690	3.81008	1.78068	5.59076
13	0.15550	3.62704	1.84932	5.47636
14	0.26960	2.98808	2.14598	5.13406
15	0.47540	2.00024	2.72222	4.72246
16	0.49430	1.92464	2.77892	4.70356
17	0.81860	0.88688	3.81668	4.70356
18	1.12870	0.14264	4.87102	5.01366
19	1.17410	0.07000	5.03446	5.10446
20	1.26160	0	5.36696	5.36696

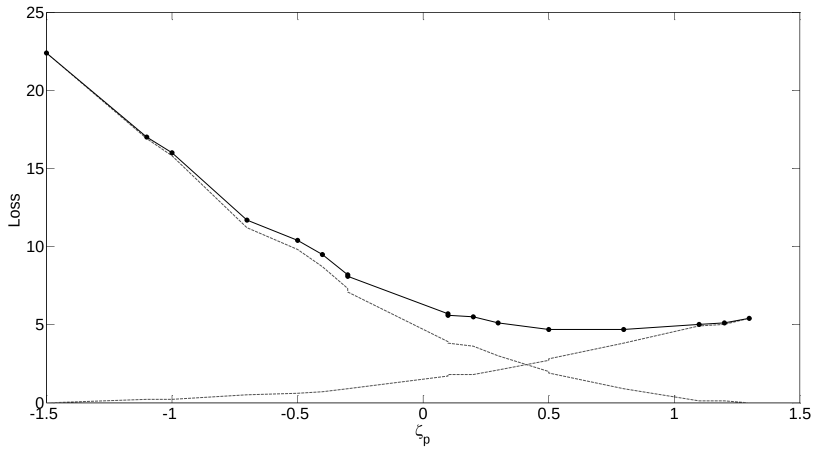

Figure 2.1. The graphical representation of Table 2.1. The two weighted sums of Equation (2.3) are represented by the two dashed lines and the solid line shows the composite objective function, all for p=0.80.

2.1.2 Regression Quantiles

To understand quantile regression, it is useful to contrast it with standard linear regression. The classical simple linear regression model is given by

$$Y = \beta_0 + \beta_1 x + \varepsilon \qquad (2.4)$$

where ε is a random error term, assumed to have zero mean and constant variance, σ^2. The additional assumption that ε is normally distributed is required only for the purpose of statistical hypothesis testing and for calculation of confidence intervals. The fitted regression model is in essence a model for the conditional mean, i.e. $E[Y \mid x] = \beta_0 + \beta_1 x$. Here, the particular interest is conditional quantiles, which for the standard regression model would be calculated as:

$$y_p(x) = \beta_0 + \beta_1 x + \sigma \Phi^{-1}(p) \qquad (2.5)$$

where $y_p(x)$ is the *p*'th quantile of *Y*, conditioned on *x*, and $\Phi^{-1}(p)$ is the *p*'th quantile in a standard normal distribution. This assumes that the error distribution is normal and constant. Both of these assumptions are likely to be violated when studying time series with trends. Nature is rarely normal and if there is a trend in the mean, there is likely to also be a trend in the variance which would violate the assumption of homoscedasticity (constant variance).

Quantile regression attempts to overcome some of the above constraints of standard linear regression by developing specific models for preselected quantiles. In linear quantile regression we seek models of the form

$$y_p(x) = \beta_0^p + \beta_1^p x \tag{2.6}$$

where the parameters β_0^p and β_1^p now specifically define a model for the *p*'th quantile of *Y*, conditioned on *x*. Here, for presentation simplicity, the predictor variable *x* is considered univariate. However, in real applications, it can be multivariate as well. No assumption regarding the error distribution is required. In linear quantile regression, each quantile of the conditional distribution is represented by an individual line. By estimating the model for a range of *p* value, one can obtain a good description of the distribution of *Y* conditional on a given *x*.

In standard linear regression, parameters are estimated by minimizing the sum of squared errors. This ensures that the model is indeed an optimal estimate of the conditional mean. It can be shown that the optimal parameters of a quantile regression model are also the solution to a minimization problem along the same logic as in

Equation (2.3). However, rather than minimizing the sum of squared errors, quantile regression involves the minimization of a weighted average of absolute errors. Specifically, to estimate the parameters of the *p*'th quantile regression model, one must minimize the following objective function:

$$f(\beta_0^p, \beta_1^p) = \sum_{\{i|y_i<\hat{y}_p(x_i)\}} (1-p)\left|y_i - \hat{y}_p(x_i)\right| + \sum_{\{i|y_i>\hat{y}_p(x_i)\}} p\left|y_i - \hat{y}_p(x_i)\right| \tag{2.7}$$

where, in the case of linear quantile regression on one variable, $\hat{y}_p(x_i) = \hat{\beta}_0^p + \hat{\beta}_1^p x_i$. In other words, the absolute value of the difference between an observation y_i and the corresponding *p*'th quantile $\hat{y}_p(x_i)$ is weighted by $(1-p)$ if the observation is below the quantile line and by *p* if the observation is above the line. While perhaps not as intuitive as the method of least squares, the estimation procedure is straightforward to implement. One of the characteristics of the fitted quantile line is that a fraction *p* of observation points will lie below the curve as one would expect. Figure 2.2 shows the result of quantile regression (10th, 50th, and 90th) and standard regression of Winnipeg annual temperature against time.

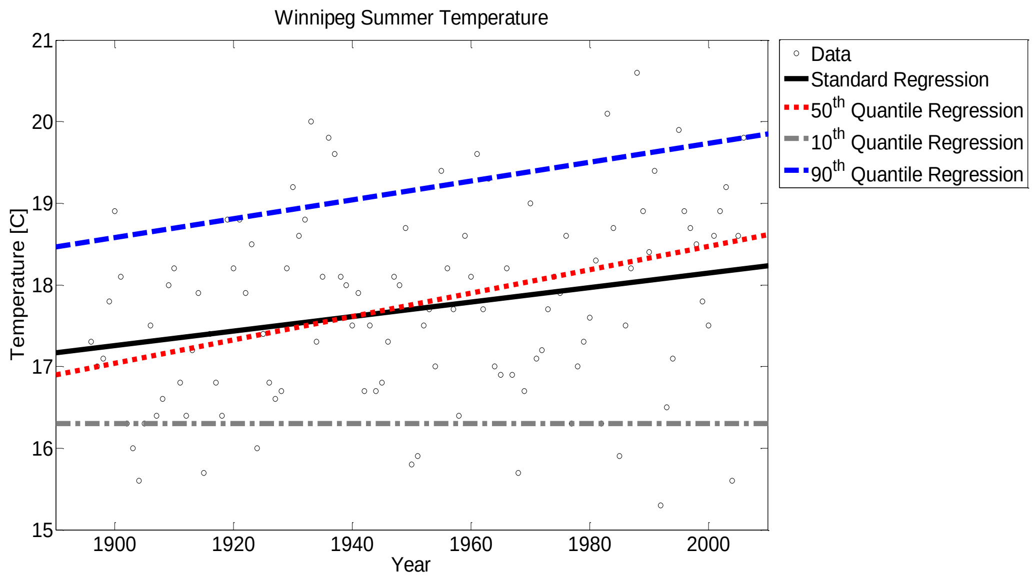

Figure 2.2. Graphical illustration of 10th, 50th, and 90th quantile regression and standard regression.

2.2 Properties

2.2.1 Equivariance and Robustness

Researchers often apply scale transformations to aid interpretation and modeling of a response variable. Equivariance properties of models and estimates refer to situations when, if the data are transformed, the models or estimates go through the same transformation, and the interpretations of the results are invariant (Hao and Naiman, 2007). The following basic equivariance properties of the estimated quantile regression coefficients apply (Koenker and Bassett, 1978):

$$\hat{\beta}_p(\lambda y, x) = \lambda \hat{\beta}_p(y, x) \qquad \lambda \in [0, \infty) \tag{2.8}$$

$$\hat{\beta}_p(-\lambda y, x) = \lambda \hat{\beta}_{1-p}(y, x) \qquad \lambda \in [0, \infty) \tag{2.9}$$

$$\hat{\beta}_p(y + x\gamma, x) = \hat{\beta}_p(y, x) + \gamma \qquad \lambda \in \Re^k \tag{2.10}$$

where $\beta_p(y, x)$ are the parameters of a regression model for the *p*'th quantile of *y* conditioned on *x*. Equations (2.8) and (2.9) express that if the response variable *y* is rescaled by a factor λ, then $\hat{\beta}_p$ is rescaled by the same factor and $\hat{\beta}_p$ is scale equivariant. Equation (2.10) represents location, shift, or regression equivariance. The least squares estimators have the same properties (Schulze, 2004).

The conditional quantiles possess another equivariance property which is much stronger than those in Equations (2.8-2.10), called the monotone property. A transformation *h* is monotone if $h(y) \leq h(y')$ whenever $y \leq y'$. For a monotone function *h*:

$$Q^p(h(y) \mid X) = h(Q^p(y \mid X)) \tag{2.11}$$

For example, if the *h* function is the log transformation, Equation (2.11) can be expressed as:

$$Q^p(\log(y) \mid X) = \log(Q^p(y \mid X)) \tag{2.12}$$

and equivalently,

$$Q^p(y \mid X) = e^{Q^p(\log(y) \mid X)} \tag{2.13}$$

This property allows reinterpreting the fitted quantile regression models for transformed variables (Hao and Naiman, 2007). By contrast, unless *h* is a linear function, the conditional mean does not share this property:

$$E(h(y) \mid X) \neq h(E(y \mid X)) \tag{2.14}$$

In other words, the monotone equivariance property fails to hold for conditional means. The monotone equivariance property is particularly important for studies involving skewed distribution where quantile regression can be used, featuring no loss of information due to the transformation process.

Robustness against outliers (insensitivity to outliers) of the response variable is another important property of quantile regression which may be contrasted to the high sensitivity of the conditional mean to outliers (Buchinsky, 1998). Outliers are defined as observations that lie outside the overall pattern of a distribution. This robustness arises because of the nature of the minimization of the absolute deviations in the quantile regression objective function. This property implies that if $y_i - X\hat{\beta}_p > 0$, then y_i can be made arbitrarily large (up to +∞), or if $y_i - X\hat{\beta}_p < 0$, y_i can be made arbitrarily small (up to -∞) without altering the solution $\hat{\beta}_p$. In other words, if the sign of the residuals does not change, modifying the values of the response variable would not affect the fitted line.

2.2.2 Homogenous and Heterogeneous Models

A basic assumption of the OLS regression technique is the so-called homoscedasticity of observational data, i.e., that the error term is independent of the value of the covariates which is often not satisfied. In this section, two examples will be used to elucidate the difference of homogenous and heterogeneous models, and to highlight the advantage of using quantile regression in the case of heterogeneous models. Two simple bivariate models (i.e. models with one predictor variable) were chosen for illustration of the results.

When the standard deviation of the error term is constant and independent of the x values, the homogenous variance regression model form associated with the OLS regression is appropriate. Two hundred independent and uniformly distributed observations on the interval (0,200) were created as a predictor variable x. The response variable y was generated from a homogenous error (lognormal with median=0 and δ =1) model by:

$$y_i = 5 + 0.05x_i + \varepsilon_i \qquad \text{with} \qquad \varepsilon_i \sim Logn(0,1) \qquad (2.15)$$

It can be seen that the homogeneity of variance assumption is satisfied. The sample data (dots) and the mean regression line (black dashed line) along with the 10th, 25th, 50th (median), 75th, and 90th quantile regression are illustrated in Figure 2.3 (top).

As it was mentioned, the error distribution is log normal and consequently the response distribution is skewed. One method to check that the assumption of normally distributed errors in OLS regression model would not have correct coverage in this case is using prediction intervals. Prediction intervals are directly related to quantiles: the 10th and 90th regression quantiles provide an 80% prediction interval at a specific value of x (predictor variables). Assuming a normal error distribution and using the mean and standard deviation resulting from the least squares estimation, the 80% prediction interval can be calculated using the normal inverse cumulative function at the 0.1 and 0.9 probabilities for OLS regression. For example, at x=160, the width of the 80% prediction interval is 16.65 - 12.14 = 4.51 based on the least squares estimate assuming a normal error distribution, whereas the interval based on the 10^{th} and 90^{th} quantile regression estimates is 15.82 – 13.28 = 2.54. The difference verifies the non-normality of error distribution.

As previously mentioned, the assumption of homoscedasticity is often not satisfied. If heteroscedasticity is present but is ignored, there will be obvious errors in the conditional distribution. In the second example, the response variable *y* generated with a heterogeneous error (normal with μ=0 and δ =$(x+0.5)^{0.5}$) model and with the same sample (Figure 2.3, bottom):

$$y_i = 5 + 0.05x_i + \varepsilon_i \qquad \text{with} \qquad \varepsilon_i \sim N(0,(x_i + 0.5)^{0.5}) \qquad (2.16)$$

It is clear that the classical assumption of independence between the error term and predictor variable is violated. The difference between homogenous and heterogeneous models can be appreciated by comparing Figure 2.3 top and bottom. In Figure 2.3, bottom, the slope estimates differ significantly across the quantiles.

For the heterogeneous models, OLS regression can be used by incorporating weights that are inversely proportioned to the variance function (Neter et al., 1996) or by using generalized linear models based on assuming some kind of distribution, from the exponential family like Poisson, negative binomial, or Gamma distributions to link changes in the variance of *y* with changes in the mean (McCullagh and Nelder, 1989). Although such methods are useful, they are not always entirely successful in resolving a particular issue and may raise a number of other concerns (Cade and Noon, 2003).

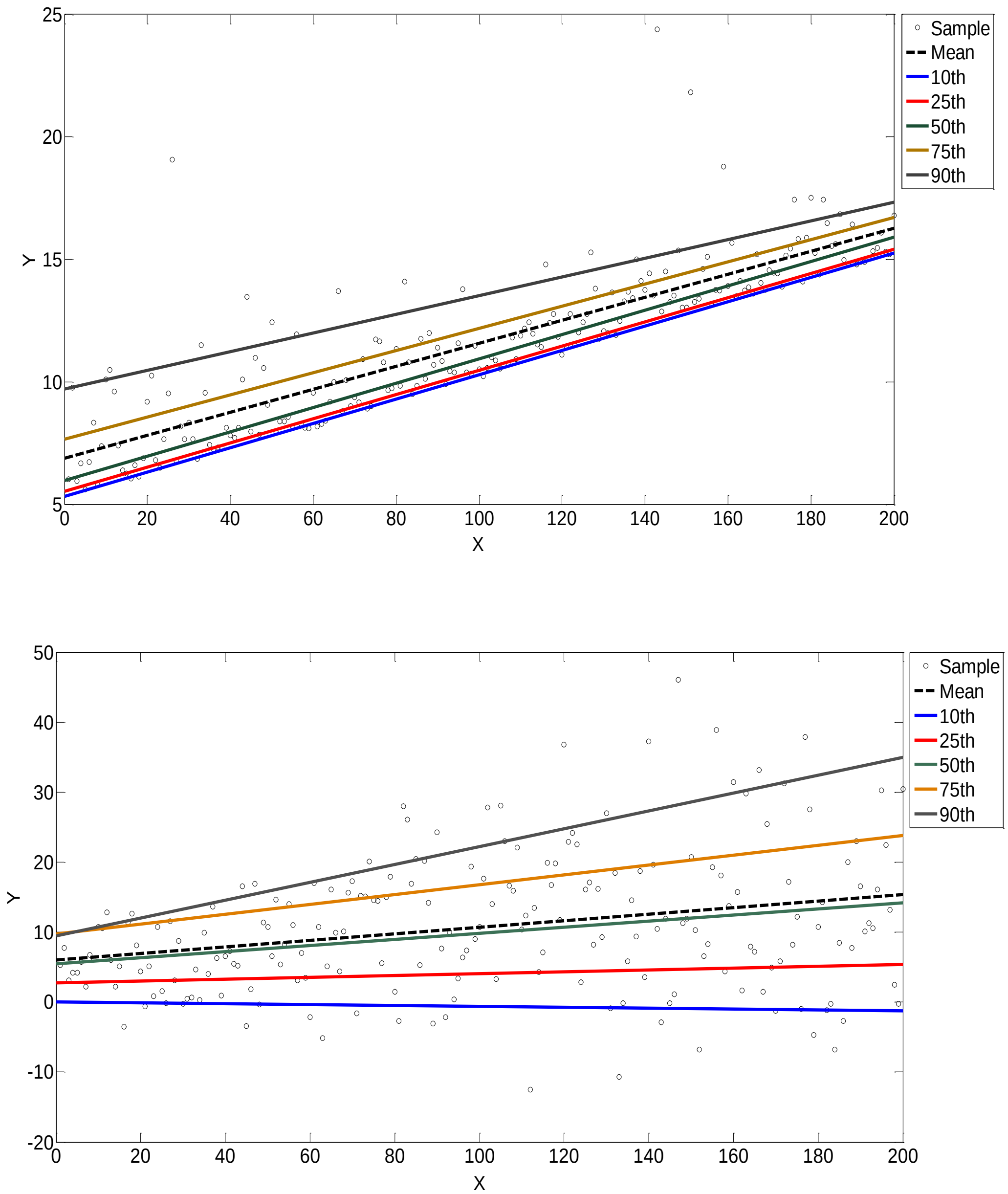

Figure 2.3. A sample (n=200) from a homogenous (top) and heterogeneous (bottom) error models (dots) along with 10th, 25th, 50th, 75th, and 90th quantile regression estimates (solid lines) and OLS regression estimates (dashed line).

Koenker and Machado (1999) in their study on quantile regression inferences highlighted the advantages of quantile regression to model heterogeneous variation models where there is no need to specify how variance changes are linked to mean. They also showed that quantile regression has the ability to detect changes in the shape of the distribution of *y* across the predictor variables.

Instead of only analyzing some selected conditional quantiles, it is also possible to consider the whole range between 0 and 1. Figure 2.4 summarizes the slopes of linear quantile functions as a function of p ($p \in \{0.01,...,0.99\}$). The precision of the estimates are indicated by 95% confidence intervals (see next section). Also shown is the slope of the mean which does not depend on p and therefore appears as a horizontal line. Here, the difference of homogenous and heterogeneous models can again be clearly seen. In the homogenous model (Figure 2.4, top), for most of the quantiles, the trends in quantiles are the same as the mean trend. For some higher quantiles (p>0.8), the slopes are different than the mean slope. This happens because the response distribution is skewed. This difference demonstrates that the quantile regression model provides a more complete view of the changes in the shape of the distribution.

In the heterogeneous model (Figure 2.4, bottom), the slopes of quantiles are significantly different from the mean trend for most of the quantiles. For lower quantiles (p<0.2), the slopes are negative and as the quantiles increases the sign of slopes changes to positive. From the significant change of the slopes of quantile regressions, it can be inferred that the variability of the response variable increases with *x* as it was observed in Figure 2.3 (bottom).

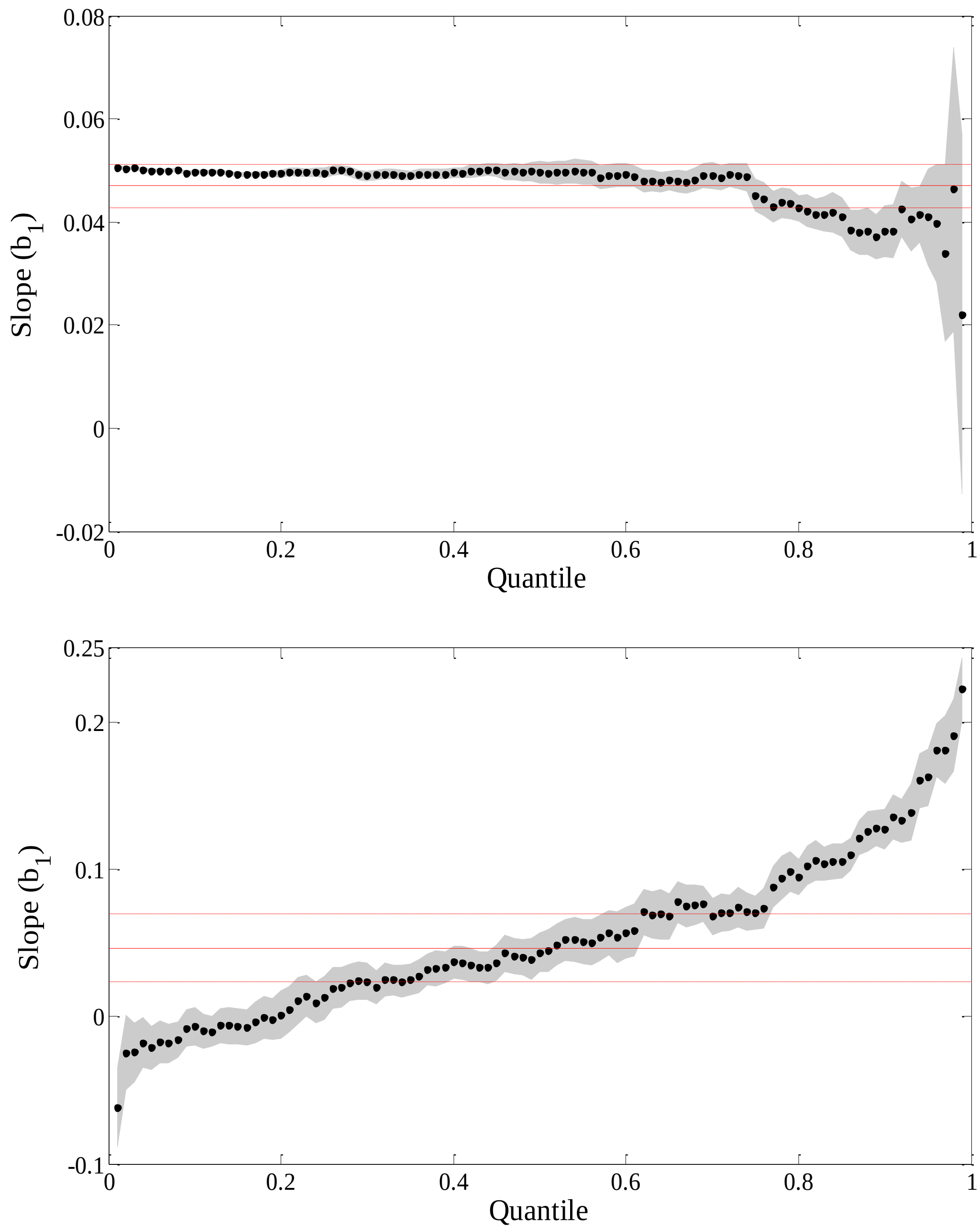

Figure 2.4. Slopes of quantile regression lines. The quantile regression coefficients (black dots) are presented with their 95% confidence bounds (shaded in grey). The least-squares regression coefficients (solid line) are also given with their 95% confidence bounds (dashed lines). The vertical axis shows the slopes (%), and the horizontal axis shows the p-value of the quantile.

2.2.3 The Bootstrapping Method for the Quantile Regression

It is common practice to provide confidence intervals for estimated parameters of a statistical model (Koenker and Hallock, 2001). Suffice it to say that confidence intervals for parameters of quantile regressions have a similar interpretation as confidence intervals for parameters in standard regression models. The bootstrap method, proposed by Efron (1979), is a Monte-Carlo method for estimating the sampling distribution of a parameter estimator that is calculated from a sample of size *n* from some population (Hao and Naiman, 2007). There has been considerable interest in using the bootstrap method to compute standard errors and estimate confidence intervals in quantile regression applications (e.g. Buchinsky, 1995; Kocherginsky et al., 2005).

Efron (1982) suggested the first implementation of the bootstrap method, called the residual bootstrap, for a non-linear median regression problem, and this approach was further developed in several studies for general quantile regression settings (e.g. De Angelis et al., 1993). The method is based on resampling with replacement from the residual vector:

$$\hat{u}_i = y_i - x_i \hat{\beta}_p \qquad i = 1,\dots,n \tag{2.17}$$

Drawing bootstrap samples $u_i^*,\dots,u_n^*$ with replacement from the estimated empirical distribution and setting $y_i^* = x_i \hat{\beta}_p + u_i^*$, a bootstrapped regression coefficient is computed as:

$$\beta_p^* = \arg\min \sum_{\{i|y_i^* < x_i'\beta_p)\}} (1-p)\left|y_i^* - x_i'\beta_p)\right| + \sum_{\{i|y_i^* > x_i'\beta_p)\}} p\left|y_i^* - x_i'\beta_p)\right| \tag{2.18}$$

By repeating this process B times, $\hat{\beta}^*_{p,1},...,\hat{\beta}^*_{p,B}$ are obtained and the asymptotic variance of $\hat{\beta}_p$ can be consistently estimated (Koenker, 2005).

This method is not useful in quantile regression, because it is only valid under the i.i.d error assumption which is rarely satisfied (Koenker, 1994). Alternatively, the (x, y)-pair bootstrap or the so-called design matrix method can be used for independent but not identically distributed error terms. In this method, instead of resampling from the empirical distribution of the residuals, the (x, y) pairs are resampled n times with replacement from the joint empirical distribution of the sample yielding a new sample of size n, (x_i^*, y_i^*). A regression coefficient $\hat{\beta}^*_p$ is estimated for each bootstrap sample of (x_i^*, y_i^*). Repeating the process B times, the asymptotic variance of $\hat{\beta}_p$ can be estimated. The discussion about the proper number of repeations, B, has been covered by Andrews and Buchinsky (2000). Horowitz (1998) suggested to smooth the quantile regression objective function as a refinement of the (x, y)-pair bootstrap method.

2.3 Quantile Regression Model Justification using a Chi-squared (χ^2) Test

In this section, we have adapted a chi-squared (χ^2) test to justify the quantile regression model. The chi-squared (χ^2) test is a simple and common goodness-of-fit test. It essentially compares a data histogram with the probability distribution (for discrete variables) or probability density (for continuous variables) function. The test statistic involves the counts of data values falling into each class in relation to the computed theoretical probabilities,

$$\chi^2 = \sum_{classes} \frac{(\#Observed - \#Expected)^2}{\#Expected} \tag{2.19}$$

If the fitted distribution is very close to the data distribution, the expected and observed counts will be very close for each class, and the squared differences in the numerator of Equation (2.10) will all be very small, yielding a small χ^2. If the fit is not good, at least a few of the classes will exhibit large discrepancies resulting in large values of χ^2. Under the null hypothesis that the data were drawn from the fitted distribution, the sampling distribution for the test statistic is the χ^2 distribution with parameter υ = (number of classes – number of parameters fit - 1) degrees of freedom (Wilks, 2011).

Based on quantile regression definitions, for example for the 80th quantile regression, 80% of observations should be under the 80th quantile regression line and 20% of observations above the line. Now, if we divide the *X* axis into 10 bins with the same number of observations in each and estimate the 20, 40, 60, and 80th quantile regressions, the perfect quantile regression model would include equal observations in each of 10×5 zones which is the null hypothesis for the adapted χ^2 test. So, here two examples of winter and summer precipitation amounts and the 850hPa East component of wind (UWIND850) relationships are presented. For Winnipeg summer case, the total number of observations is 950 and the expected value for each zone is 19. The number of observations in each zone has been counted (Figure 2.5). The calculated χ^2 value is 36.74. As mentioned above, in the traditional χ^2 test, we subtract one degree of freedom because we know the total number of observations, so once we have observed *n-1* bins counts, we will know exactly the number of observations in the last bin. Hence, there is no degree of freedom for the last bin. In our case, we know the total number of observations in 10 bins, so in each bin we will lose a degree of freedom (50-10=40). The critical value for 40 degrees of freedom at the 5% level is 55.8, so the null hypothesis would not be rejected (36.74<55.8) for Winnipeg summer which implies that the quantile regression model is reasonable. For the winter case, the number of observations is 820 and the expected value for each zone is 16.4, and the calculated χ^2 value is 38.64 (Figure 2.6). The null hypothesis would not be rejected (38.64<55.8) for Winnipeg winter as well.

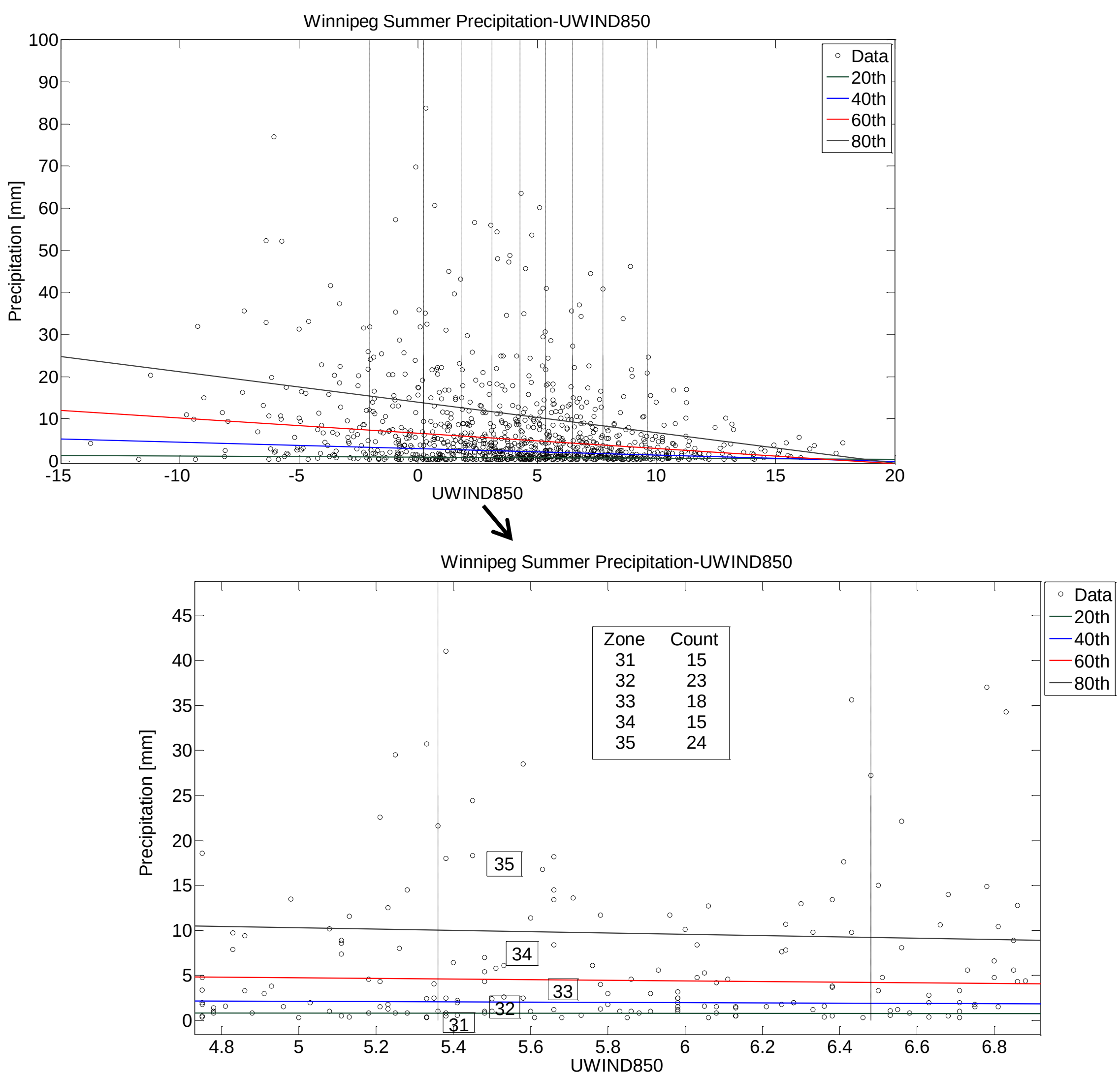

Figure 2.5. Winnipeg summer precipitation-UWIND850 quantile regression trends (top), the bottom figure shows a zoom of a particular region in the top figure.

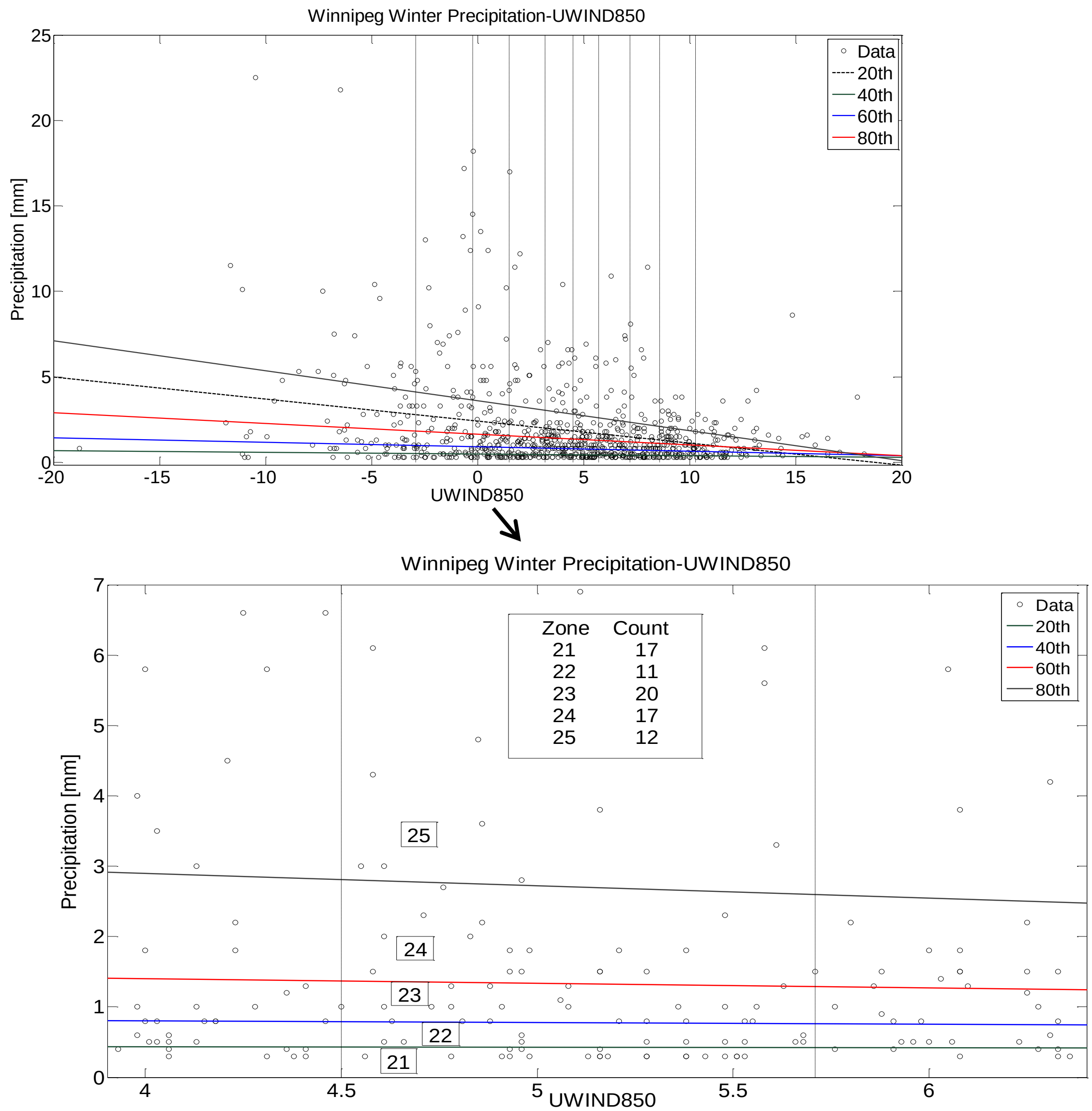

Figure 2.6. Winnipeg winter precipitation-UWIND850 quantile regression trends (top), the bottom figure shows a zoom of a particular region in the top figure.

2.4 Applications

The quantile regression technique has been used widely in economics (e.g. Koenker and Hallock, 2001; Auld and Powell, 2009) and occasionally in ecological and biological studies (e.g. Dunham et al., 2002; Cade et al., 2008). Unequal variation in climate variables (e.g. precipitation) and complex relationships between environmental processes in climate change studies have motivated researchers to look for new techniques like quantile regression to analyze these relationships. Koenker and Schorfheide's (1994) research on global temperature change was one of the first studies on climate change using quantile regression. They concluded that a flexible and reasonably parsimonious quantile regression technique may find useful applications in other aspects of climate change research.

The application of Quantile regression in climate studies can be roughly divided into three categories. These categories are not strictly distinct and some of the studies could be considered in two or three categories.

2.4.1 Trend Detection

Quantile regression has been extensively applied for detecting trends. Two of three papers presented in this thesis can be classified in this category. Providing multiple rates of change for different parts of conditional distribution and enabling researchers to analyze the tails have made this technique applicable for identifying distinct rates of change and quantifying long-term variability in climate variables.

Using quantile regression, Barbosa (2008) characterized long-term sea level variability in the Baltic Sea. She concluded that quantile trends provide a more complete description of regional sea-level long-term variability than OLS regression. Chamaille-James et al.'s (2007) study showed that drought severity increased in Zimbabwe in the course of the 20^{th} century. They suggested the quantile regression technique should be considered more often as a valuable tool to investigate climatic changes in arid and semi-arid regions, because asymmetric trends appear to be a common feature of the recent climate changes. Studies that can be referenced in this category including: tropical cyclones trend changes study (Elsner et al., 2008); annual streamflow distributions analysis (Luce and Holden, 2009); annual rainfall changes over time in Zimbabwe (Mazvimavi, 2010); climate characteristics changes analysis (Teemofeev and Sterin, 2010); air temperature changes over Central Europe (Barbosa et al., 2011); distributional changes detection in environmental processes (Reich, 2012); and spatial patterns of trends in Baltic sea-level variability (Donner et al., 2012).

2.4.2 Interpretation of Non-Linear Relationships

As mentioned previously, quantile regression provides a basis for analyzing the entire conditional distribution of the predictand rather than a single measure of the central tendency of its distribution. Quantile regression's flexibility to allow covariates to have different relationships in different parts of the conditional distribution and the robustness to departures from normality and skewed tails can be an asset when the functional relationship between predictor variables and the response distribution is multifaceted (Mata and Machado, 1996).

Baur et al. (2004) used the conditional quantile regression approach for the interpretation of the nonlinear relationships between daily maximum 1-h ozone concentrations and both meteorological and persistence information at four stations in Greece. They concluded that quantile regression has the advantage of easy implementation and transparency of results, which is in contrast to black box models such as neural networks where the predictor-predictand relationships are less clear. Dunham et al. (2002) analyzed the relationship of the abundance of Lahontan cutthroat trout to the ratio of stream width to depth. They found a negative relationship in upper tails while a least squares regression estimated zero change. They remarked that if they used mean regression estimates, no relation between trout densities and the ratio of stream width to depth would mistakenly be their conclusion. Identification of necessary river flows to protect and enhance migratory birds habitat (Zoellick et al., 2004), the response of nematodes to deep-sea CO_2 sequestration (Fleeger et al., 2010), unravelling the effects of soil properties on water infiltration (Mills et al., 2006), modeling tropical cyclone intensity (Jagger and Elsner, 2009), climate change effects on relationships between forest fine and coarse woody debris carbon stocks (Woodall and Liknes, 2008) are studies that can be classified in this category.

3.4.3 Forecasting

No assumption regarding the error distribution is required in quantile regression. This advantage as well as possession of the monotone property make quantile regression a potentially useful forecasting approach. Bremnes (2004) made precipitation forecasts in terms of quantiles to improve or enrich numerical weather prediction outputs. They demonstrated how reliable probabilistic precipitation forecasts in terms of quantiles can be made using quantile regression. Friederichs and Hense (2007) statistically downscaled extreme precipitation events using a censored quantile regression technique. They concluded that the conditional quantile forecasts are sensitive enough to assess the probability of extreme events. Studies on probabilistic wind power forecasts (Bremnes, 2006), forecast verification scores for extreme value distributions with an application to probabilistic peak wind prediction (Friederichs and Thorarinsdottir, 2012), and estimation of predictive hydrological uncertainty (Weerts et al., 2010) are examples that fall in this category.

Chapter 3

Article1: Analysis of Arctic and Antarctic Sea Ice Extent using Quantile Regression

Abstract

A number of recent studies have examined trends in sea ice cover using ordinary least squares regression. In this study, quantile regression is applied to analyze other aspects of the distribution of sea ice extent. More specifically, trends in the mean, maximum, and minimum sea ice extent in the Arctic and Antarctic are investigated. While there is a significant decreasing trend in mean Arctic sea ice extent of -4.5% per decade from 1979 through 2010, the Antarctic results show a small positive trend of 2.3% per decade. In some cases such as the Antarctic minimum ice cover, selected quantile regressions yield

slope estimates that differ from trends in the mean. It was also found that the variability in Antarctic sea ice extent is higher than in the Arctic sea ice.

Keywords: Sea Ice Extent, Quantile Regression, Arctic, Antarctic.

3.1 Introduction

Sea ice is an important component of the global climate system. The change in sea ice cover has been a central focus of many climate change studies in recent years (e.g., Comiso and Parkinson, 2004; Singarayer et al., 2006), both because of the ice-albedo feedback mechanism that enhances climate response at high altitudes, and because sea ice modifies the exchange of heat, gases, and momentum between the atmosphere and polar oceans (IPCC, 2007).

Among the climatically important characteristics of sea ice (concentration, extent, thickness, velocity, and growth and melt rates), extent is the only variable for which observations are available for more than a few decades. Comiso (2010) found a decreasing trend of -3.8% per decade in Arctic annual average sea ice extent, whereas a small positive trend of 1.2% per decade was observed for Antarctic sea ice during the period 1978-2008. Some studies have used other sources than satellite data to analyze sea ice extent over longer time scales. Rayner et al. (2003) used the Hadley Centre sea ice reanalysis data set for the months of March and September for the 20th century and indicated a sustained decline in arctic ice extent since around the early 1970s. Information from ship reports was used by Vinje (2001) to estimate the April Nordic Seas (i.e. the Greenland, Iceland, Norwegian, Barents, and Western Kara Seas, bounded by

30°W, 70°E, and 80°N) ice extent since around 1860. He reported a generally continuous decline over the period.

Another notable observation in Arctic ice cover is the significant decrease of -11.3% per decade in the summer minimum sea ice extent between 1979 and 2008 (Comiso, 2010). Early onsets of significant melting north of Alaska and Siberia have been observed in 2002 satellite records. In addition, the smallest recovery of winter sea ice cover in the satellite records, and the earliest onset of melt were observed in 2004 and 2005 (NASA, 2005). The sea ice extent in the Arctic in September of 1979 and 2010 are compared in Figure 3.1. September is the month where the annual minimum extent usually occurs. The difference between these two years is significant.

Global climate change affects the Arctic and Antarctic in different ways. Turner and Overland (2009) suggest that topographic factors and the land/sea distribution account for most of the differences. The ice-albedo feedback mechanism operates effectively in the Arctic Ocean because of the high level of solar radiation received in summer. Turner and Overland also concluded that the ozone hole in the Antarctic affects the high latitude ocean and atmosphere significantly, isolating the continent, and increasing the westerly winds over the Southern Ocean, especially during the summer and winter.

The change in sea ice extent is often assessed using linear trend models estimated by ordinary least squares regression. Standard regression models provides information on the slope of the mean, but does not specifically address other aspects of the conditional distribution. Changes in extremes can be equally or even more concerning in climate change studies than trends in the mean. In this paper, a technique known as quantile regression is applied to study changes in the Arctic and Antarctic sea ice extent. Quantile regression extends traditional regression models to the estimation of conditional quantile functions. Time series of monthly mean and annual minimum and maximum sea ice extent are investigated for trends. Trends in the mean of these variables will be compared with trends in low, median, and high (20%, 50%, and 80%) quantiles.

3.2 Quantile Regression

Standard linear regression models are extensively used in statistical analyses. Despite their popularity, these conditional mean models have some limitations. When interest is in the quantiles of the conditional distribution rather than the mean, standard regression models may fail to provide the desired information because the assumption of normally-distributed residuals with constant variance may not be justified. Standard regression models are sensitive to outliers and can lead to unreasonable models if outliers are present in the data set. This is especially a problem if the sample size is moderately small and the error distribution is heavy-tailed (Hao and Naiman, 2007).

Koenker and Bassett (1978) developed a regression-type model for estimating the functional relationship between predictor variables and any quantile in the distribution of the response variable. Quantile regression overcomes some of the limitations of standard regression and is suitable for the study of changes in the frequency of environmental extremes over time. A brief description of the principles of quantile regression is given below.

To understand quantile regression, it is useful to contrast it with standard linear regression. The classical simple linear regression model is given by

$$Y = \beta_0 + \beta_1 x + \varepsilon \tag{3.1}$$

where ε is a random error term, assumed to have zero mean and constant variance, σ^2. The additional assumption that ε is normally distributed is required only for the purpose of statistical hypothesis testing and for calculation of confidence intervals. The fitted regression model is in essence a model for the conditional mean, i.e. $E[Y \mid x] = \beta_0 + \beta_1 x$. The particular interest in the paper is conditional quantiles, which for the standard regression model would be calculated as:

$$y_p(x) = \beta_0 + \beta_1 x + \sigma \Phi^{-1}(p) \tag{3.2}$$

where $y_p(x)$ is the *p*'th quantile of *Y*, conditioned on *x*, and $\Phi^{-1}(p)$ is the *p*'th quantile in a standard normal distribution. This assumes that the error distribution is normal and constant. Both of these assumptions are likely to be violated when studying time series with trends. Nature is rarely normal and if there is a trend in the mean, there is likely to also be a trend in the variance which would violate the assumption of homoscedasticity (constant variance).

Quantile regression attempts to overcome some of the above constraints of standard linear regression by developing specific models for preselected quantiles. In linear quantile regression we seek models of the form

$$y_p(x) = \beta_0^p + \beta_1^p x \tag{3.3}$$

where the parameters β_0^p and β_1^p now specifically define a model for the *p*'th quantile of *Y*, conditioned on *x*. No assumption regarding the error distribution is required. In linear quantile regression, each quantile of the conditional distribution is represented by an individual line. By estimating the model for a range of *p* value, one can obtain a good description of the distribution of *Y* conditional on a given *x*.

In standard linear regression, parameters are estimated by minimizing the sum of squared errors. This ensures that the model is indeed an optimal estimate of the conditional mean. It can be shown that the optimal parameters of a quantile regression model are also the solution to a minimization problem. However, rather than minimizing the sum of squared errors, quantile regression involves the minimization of a weighted average of absolute errors. Specifically, to estimate the parameters of the *p*'th quantile regression model, one must minimize the following objective function:

$$f(\beta_0^p, \beta_1^p) = \sum_{\{i|y_i < \hat{y}_p(x_i)\}} (1-p)\left|y_i - \hat{y}_p(x_i)\right| + \sum_{\{i|y_i > \hat{y}_p(x_i)\}} p\left|y_i - \hat{y}_p(x_i)\right| \tag{3.4}$$

with $\hat{y}_p(x_i) = \hat{\beta}_0^p + \hat{\beta}_1^p x_i$. In other words, the absolute value of the difference between an observation y_i and the corresponding *p*'th quantile $\hat{y}_p(x_i)$ is weighted by $(1-p)$ if the observation is below the quantile line and by *p* if the observation is above the line. While perhaps not as intuitive as the method of least squares, the estimation procedure is straightforward to implement. One of the characteristics of the fitted quantile line is that a fraction *p* of observation points will lie below the curve as one would expect.

Further details about quantile regression models can be found in Cade and Noon (2003), Koenker (2005), and Hao and Naiman (2007). In this study, the quantreg package add-on to the R language combined with Matlab coding was used for quantile regression modeling.

Several studies have employed quantile regression for environmental modeling and climate change impact assessment. Table 3.1 provides a list of some recent studies.

3.3 Data and Analysis

Sea ice extent, defined as the area in which ice concentration is at least 15 percent, is used to quantify the sea ice cover. Monthly sea ice extent data for both northern and southern hemispheres for 32 years (1979-2010) were extracted from the National Snow and Ice Data Center (Fetterer et al., 2002) to examine how the sea ice cover changes in extent during an annual cycle. As seen in Figure 3.2, while the duration of the growth and decay periods are almost the same in the northern hemisphere, the growth and decay seasons last about 7 and 5 months respectively in the southern hemisphere.

3.4 Results

In order to investigate changes in climate extremes, the slope of the 20th and the 80th quantile regression lines of the 32 years of monthly and annual maximum and minimum ice extent for both the northern and southern hemispheres have been estimated and compared with trends in the mean slope. Statistical significance of trends will be discussed later where the results of Figure 3.7 and Table 3.2 presented. As illustrated in Figure 3.3, trends in the quantiles of monthly sea ice extent are almost the same as the trends in the mean for both hemispheres. The results for trend in the mean agrees with the findings in previous studies (e.g., Comiso et al., 2008; Comiso, 2010) that report a significant decreasing trend (-4.5% per decade) in Arctic sea ice and an increasing trend (2.3% per decade) in Antarctic sea ice extent.

Ice that survives the summer, often referred to as the perennial ice cover, consists mainly of relatively thick, multiyear ice floes. The minimum value for each year is a reasonable proxy for estimation of the perennial ice extent. The annual maximum and minimum sea ice extents for both hemispheres have been examined by quantile regression to determine the nature of trend in the tail of these distributions (Figure 3.4 and Figure 3.5). For Arctic sea ice extent, a decreasing trend can be seen for both the maximum and the minimum. Figures 3.4a and 3.4b show that the 20^{th} and 80^{th} percentile lines are almost parallel, suggesting that although the mean may be changing, the variability, as measured by the 20^{th}-to-80^{th} quantile range, in Arctic sea ice extent has not changed dramatically, neither for maximum, nor for minimum ice cover.

Figure 3.5 suggests that the variability in Antarctic minimum sea ice extent has changed over time. The slope of the 80^{th} percentile line for Antarctic minimum ice cover is much larger than the slope of the mean and lower quantiles, suggesting increased variability.

The median regression model may be compared to the mean regression model, obtained from standard linear regression. It is noted that the median regression and mean regression should be similar for symmetrical error distributions (Bancayrin–Baguio, 2009). The difference between mean and median quantile regression in the Antarctic minimum ice cover is notable and shows that Antarctic minimum ice cover is asymmetrical and skewed.

To compare the conditional distributions estimated by standard regression and by quantile regression, the case of sea ice extent for 2010 is considered ($y|_{x=2010}$). In the case of standard regression, residuals (ε) are assumed to be normally distributed around the conditional mean value ($y_{mean}|_{x=2010}$), so the conditional distribution is also normal. The cumulative distribution functions (CDFs) associated with the standard linear regression assuming normal distribution of residuals have been calculated and compared with quantile regression predictions ($y_{0.01,0.02,\ldots,0.99}|_{x=2010}$) for the Arctic and Antarctic maximum and minimum sea ice cover. As it can be seen in Figure 3.6, the 2010 Arctic maximum sea ice cover (Figure 3.6a) and the 2010 Antarctic minimum ice cover (Figure 3.6d) quantile regression predictions are quite different from the results from conventional linear regression, just as one would expect based on the graphs shown in Figure 3.4a and Figure 3.5b. As mentioned above, the increasing trend in high quantiles for the Antarctic minimum ice cover is higher than the mean trend. For example, the quantile regression sea ice prediction for the 80th percentile is 3.75×10^6 km^2 while the standard regression prediction is 3.4×10^6 km^2, a difference of 10%.

It is common practice to provide confidence intervals for estimated parameters of a statistical model (Koenker and Hallock, 2001). Several methods have been proposed to compute approximate confidence intervals for the parameters of quantile regression models, including the rank-score method and bootstrap resampling. The rank-score method, which produces confidence intervals for the estimated parameters by inverting a rank test, has been used in this study. The details of the method are not given here, but can be found in Koenker (1994). Suffice to say that confidence intervals for parameters of quantile regressions have a similar interpretation as confidence intervals for parameters in standard regression models. Figure 3.7 summarizes the slopes of linear quantile functions as a function of *p*. The precision of the estimates are indicated by confidence intervals. Also shown is the slope of the mean which does not depend on *p* and therefore appears as a horizontal line. The quantile regression model provides a more complete view of the changes in the shape of the distribution of sea ice extent. As Figure 3.7 illustrates, in several cases trends in quantiles are different from the mean trends, including the Arctic mean ice extent and the Antarctic minimum ice cover. The advantage of the quantile regression model can be seen in the Arctic mean plot (top left). In the Arctic, it has been observed that the decreasing trend in September sea ice (annual minimum ice cover) is higher than in other months (e.g. Comiso, 2010) and this can be seen in the lower quantiles of Arctic means where the most negative slopes are observed. The ordinary linear regression model on the other hand produces a constant -0.5% trend in all quantiles. The existence of a strong seasonal cycle in the mean plots (top plots) may raise some concern about the appropriateness of quantile regression, or indeed any other type of regression. It can be seen that the confidence envelopes for the slope of the quantile models are wider (statistically less precise estimates) in the central quantiles ($0.4<p<0.6$) compared to the tail quantiles ($p<0.2$ and $p>0.8$) which may at first glance appear somewhat counterintuitive. A more thorough analysis of the data shows that the seasonal cycle and the data dispersion in the central quantiles are the reasons for less precise estimations.

Another observation from Fig 3.7 is that the confidence envelopes in the Arctic do not cross the zero line, confirming that the decreasing trend in the Arctic is statistically significant for all quantiles. This is in contrast to the Antarctic where confidence envelopes in most cases include zero, suggesting that trends are not statistically significant.

A quantile-based measure of skewness (QSK) of the conditional distribution can be defined as:

$$QSK^{(p)} = \frac{y_{(1-p)} - y_{0.5}}{y_{0.5} - y_p} - 1, \quad 0 < p < 0.5 \tag{3.5}$$

where p is preselected value and the specific conditioning on x has been omitted to simplify the notation. This measure represents the ratio of the upper spread to the lower spread minus one. A zero value for QSK over the full range of p-values indicates a symmetric distribution, while positive and negative values indicate right- and left-skewness, respectively (Hao and Naiman, 2007). The QSK measure for p=0.4 are shown in the last column in Table 3.2. Because of the range of slopes changes in the plots in Figure 3.7, the left-skewness can be seen more clearly in Arctic and Antarctic maximum ice cover plots, where the difference between the slopes of the 40th and 50th percentiles is higher than the difference between the slopes of the 60th and 50th percentiles.

The results for changes in sea ice extent based on standard regression and quantile regression are summarized in Table 3.2. The table provides the slope coefficients of mean and quantile regression lines (20th and 80th) and calculated sea ice extent changes from 1979 through 2010 ($y|_{x=2010} - y|_{x=1979}$) for mean and quantile regressions. Some of these changes are statistically significant. For example, the decrease of 2.52×10^6 km^2 (35% of 1979) in the lower quantiles of minimum Arctic sea ice extent from 1979 to 2010 is quite notable.

3.5 Discussion and Conclusions

Temperatures, winds, waves, and currents are some of the factors affecting the Arctic sea ice cover. The warming Arctic temperatures are clearly manifested in the declining sea ice cover. Oscillations within the atmosphere are also associated with sea ice cover and several studies have analyzed the possible connections of the Arctic Oscillation (e.g., Rigor and Wallace, 2004), the North Atlantic Oscillation (e.g., Parkinson, 2000), and periodic changes in wind patterns (e.g., Proshutinsky and Johnson, 1997) with sea ice. Comiso et al. (2008) suggested that despite changes in the modes of Arctic and North Atlantic Oscillations and in the predominant wind patterns, the decline in sea ice cover has continued, implying that the warming conditions may be overriding the oscillations.

The Arctic and Antarctic have experienced different changes in recent years. This study has highlighted the fact that changes are not simply in the mean of the key ice variables, but that their entire distribution shape may be impacted. As discussed by Turner and Overland (2009), the mechanisms behind the different changes in Arctic and Antarctic are complex and not easy to decipher. More detailed analyses on the link between observed changes and the ocean and atmospheric forcings of the sea ice extent should be pursued in the future.

This study found a significant decreasing trend in mean Arctic sea ice extent of -4.5% per decade from 1979 through 2010. The results are even more pronounced for summer minimum sea ice cover where a mean trend of -10.1% per decade in 1979-2010 was observed. These findings suggest that if the warming trend continues, Arctic sea ice may practically disappear some time in the future in summer. While the decreasing trend of sea ice cover in the Arctic is significant, the Antarctic results show a small positive trend of 2.3% per decade in mean sea ice extent for the same period.

The advantage of the quantile regression model can be seen in the Arctic mean sea ice extent (Figure 3.7 top left). In the Arctic, it has been observed that the decreasing trend in September sea ice (annual minimum ice cover) is higher than in other months (e.g. Comiso, 2010) and this can be seen in the lower quantiles of Arctic means where the most negative slopes were observed. In some cases, such as the Antarctic minimum sea ice cover, the trends in the two quantiles considered were found to be different from mean trends and the conditional distributions are asymmetric and skewed (Figure 3.5). It was also found that the Antarctic sea ice variability is higher than the Arctic sea ice.

Standard regression for analyzing sea ice extent might have caused some of the above observations to go undetected. This includes the observed change in variability in the Antarctic minimum ice cover and the left-skewness of conditional distributions. The asymmetric distribution in Antarctic minimum ice cover (Figure 3.5b and Figure 3.7) is a good example of a situation where qquantile regression provides a more detailed picture of the evolution of the conditional distribution.

Quantile regression has the advantage of easy implementation and transparency of results, which is in contrast to black box models such as neural networks (Baur et al., 2004) where the functional relationship between predictor variables and the response distribution is less clear. As noted by Barbosa (2008), quantile regression is a useful technique for identifying distinct rates of change in geophysical time series and should be used for quantifying long-term variability in climatic and oceanographic variables. Future work will explore the role of atmospheric variables on sea ice cover using quantile regression.

Acknowledgments:

Constructive comments by Dr. David Barber and Dr. Ryan Galley on an early draft are greatly appreciated. Two anonymous reviewers provided useful comments that enhanced the presentation. The National Snow and Ice Data Center (NSIDC) is gratefully acknowledged for providing the sea ice extent data. Authors have no conflict of interest to declare.

References:

Barbosa SM. 2008. Quantile trends in Baltic Sea level. Geophysical Research Letters 35 (22): art. No. L22704. DOI: 10.1029/2008GL035182.

Baur D, Saisana M, Schulze N. 2004. Modeling the effects of meteorological variables on ozone concentration - A quantile regression approach. Atmospheric Environment 38 (28): 4689-4699. DOI: 10.1016/j.atmosenv.2004.05.028.

Bancayrin – Baguio C. 2009. Performance of Median and Least Squares Regression for Slightly Skewed Data. World Academy of Science, Engineering and Technology 29: 226-231.

Cade BS, Noon BR. 2003. A gentle introduction to quantile regression for ecologists. Frontiers in Ecology and the Environment 1 (8): 412-420.

Comiso JC. 2010. Variability and Trends of the Global Sea Ice Cover. In: Sea ice, second edition [Thomas, DN, and Dieckmann, GS.]. Wiley-Blackwell, Oxford, UK: 205-246.

Comiso JC, Parkinson CL. 2004. Satellite-observed changes in the Arctic. Physics Today 57 (8): 38-44.

Comiso JC, Parkinson CL, Gersten R, Stock L. 2008. Accelerated decline in the Arctic sea ice cover. Geophysical Research Letters 35 (1): art. no. L01703. DOI: 10.1029/2007GL031972.

Elsner JB, Kossin JP, Jagger TH. 2008. The increasing intensity of the strongest tropical cyclones. Nature 455 (7209): 92-95. DOI: 10.1038/nature07234.

Fetterer F, Knowles K, Meier W, Savoie M. 2002, updated 2009. Sea Ice Index. Boulder, CO: National Snow and Ice Data Center. Digital media.

Fleeger JW, Johnson DS, Carman KR, Weisenhorn PB, Gabriele A, Thistle D, Barry JP. 2010. The response of nematodes to deep-sea CO2 sequestration: A quantile regression approach. Deep-Sea Research Part I: Oceanographic Research Papers 57 (5): 696-707. DOI: 10.1016/j.dsr.2010.03.003.

Hao L, Naiman DQ. 2007. Quantile regression. SAGE: Thousand Oaks, California, USA.

IPCC. 2007. Climate change 2007: The physical science basis, contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press: Cambridge, New York, USA.

Koenker R. 1994. Confidence intervals for regression quantiles. Proceedings of the Fifth Prague Symposium on Asymptotic Statistics, 349-359.

Koenker R. 2005. Quantile Regression. Cambridge University Press: Cambridge, New York, USA.

Koenker R, Bassett G. 1978. Regression quantiles. Econometrica 46 (1): 33-50.

Koenker R, Hallock KF. 2001. Quantile regression. Journal of Economic Perspectives 15 (4): 143-156.

Koenker R, Schorfheide F. 1994. Quantile spline models for global temperature change. Climatic Change 28 (4), 395-404.

Mazvimavi D. 2010. Investigating changes over time of annual rainfall in Zimbabwe. Hydrology and Earth System Sciences 14 (12): 2671-2679. DOI: 10.5194/hess-14-2671-2010.

NASA. 2005. Arctic Sea Ice Continues to Decline, Arctic Temperatures Continue to Rise In 2005 (2005, September 28). NASA, Goddard Space Flight Center. Retrieved January 28, 2011 from http://www.nasa.gov/centers/goddard/news/topstory/2005/arcticice_decline.html

Parkinson CL. 2000. Recent trend reversals in Arctic sea ice extents: Possible connections to the North Atlantic oscillation. Polar Geography 24 (1): 1-12.

Proshutinsky AY, Johnson MA. 1997. Two circulation regimes of the wind-driven Arctic Ocean. Journal of Geophysical Research 102 (C6): 12493-12514.

Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A. 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. Journal of Geophysical Research D: Atmospheres 108 (14): ACL 2-1 - ACL 2-29.

Rigor IG, Wallace JM. 2004. Variations in the age of Arctic sea-ice and summer sea-ice extent. Geophysical Research Letters 31 (9): L09401 1-4. DOI: 10.1029/2004GL019492.

Singarayer JS, Bamber JL, Valdes PJ. 2006. Twenty-first-century climate impacts from a declining Arctic sea ice cover. Journal of Climate 19 (7): 1109-1125. DOI: 10.1175/JCLI3649.1

Turner J, Overland J. 2009. Contrasting climate change in the two polar regions. Polar Research 28 (2): 146-164. DOI: 10.1111/j.1751-8369.2009.00128.x.

Vinje T. 2001. Anomalies and trends of sea-ice extent and atmospheric circulation in the Nordic Seas during the period 1864-1998. Journal of Climate 14 (3): 255-267.

Woodall CW, Liknes GC. 2008. Relationships between forest fine and coarse woody debris carbon stocks across latitudinal gradients in the United States as an indicator of climate change effects. Ecological Indicators 8 (5): 686-690. DOI: 10.1016/j.ecolind.2007.11.002.

Table 3.1. Quantile regression applications in environmental modeling and climate change impact assessment.

Global temperature change over the last century	Koenker and Schorfheide (1994)
Effects of meteorological variables on Ozone concentration	Baur et al. (2004)
Climate change effects on relationships between forest fine and coarse woody debris carbon stocks	Woodall and Liknes (2008)
Tropical cyclones trend changes	Elsner et al. (2008)
Quantile trends in Baltic sea level	Barbosa (2008)
The response of nematodes to deep-sea CO_2 sequestration	Fleeger et al. (2010)
Changes over time of annual rainfall in Zimbabwe	Mazvimavi (2010)

Table 3.2. Slope and sea ice extent changes (1979-2010) of standard and quantile regression.

	Mean		20th Quantile		80th Quantile		QSK
	Slope (%)	Extent Changes (10^6 Km^2)	Slope (%)	Extent Changes (10^6 Km^2)	Slope (%)	Extent Changes (10^6 Km^2)	P=0.4
NH	-0.48[a]	-1.85	-0.49[a]	-1.89	-0.35[a]	-1.33	-4.20
SH	0.23	0.87	0.19	0.73	0.11	0.41	-2.93
NH-Max	-4.14[a]	-1.28	-4.83[a]	-1.50	-3.86[a]	-1.20	-1.04
NH-Min	-8.13[a]	-2.52	-8.12[a]	-2.52	-8.21[a]	-2.55	-4.86
SH-Max	1.09[a]	0.34	1.48	0.46	1.00[a]	0.31	-0.77
SH-Min	0.92	0.29	0.21	0.07	2.60[a]	0.81	-1.21

[a]Statistically significant at the 95% confidence level

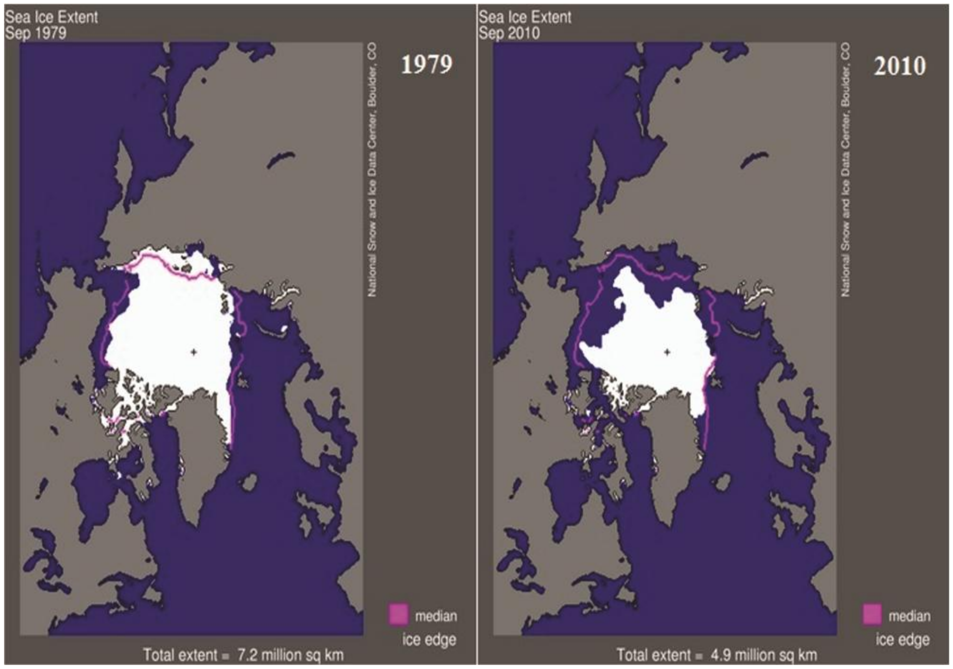

Figure 3.1. Extent of annual minimum sea ice in 1979 and 2010. Source: National Snow and Ice Data Center

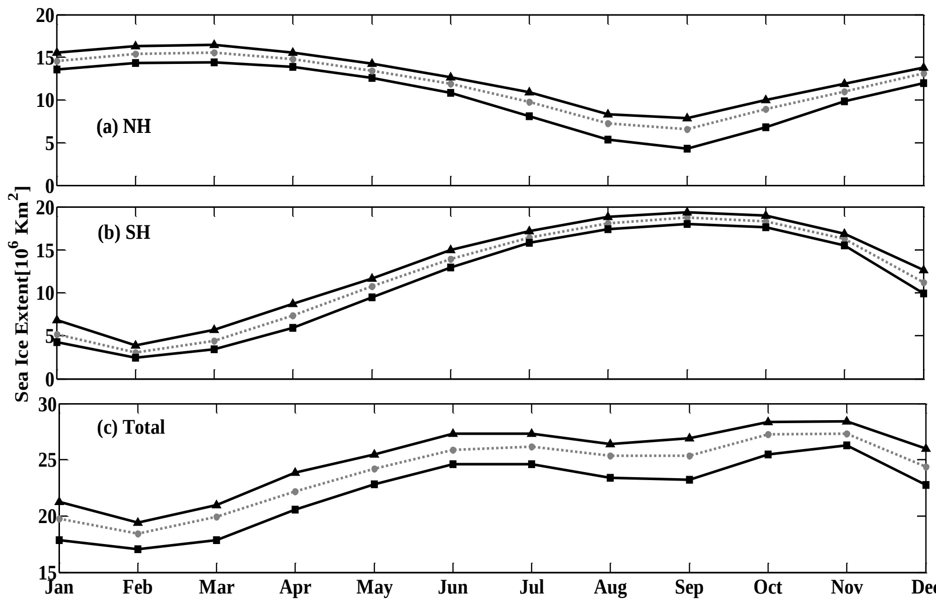

Figure 3.2. Seasonal variation in sea ice extent based on data from 1979 to 2010 in (a) the northern hemisphere, (b) the southern hemisphere and (c) combined northern and southern hemisphere. The middle line represents the climatological means, while the top and bottom lines represent the highest and lowest monthly values for each month.

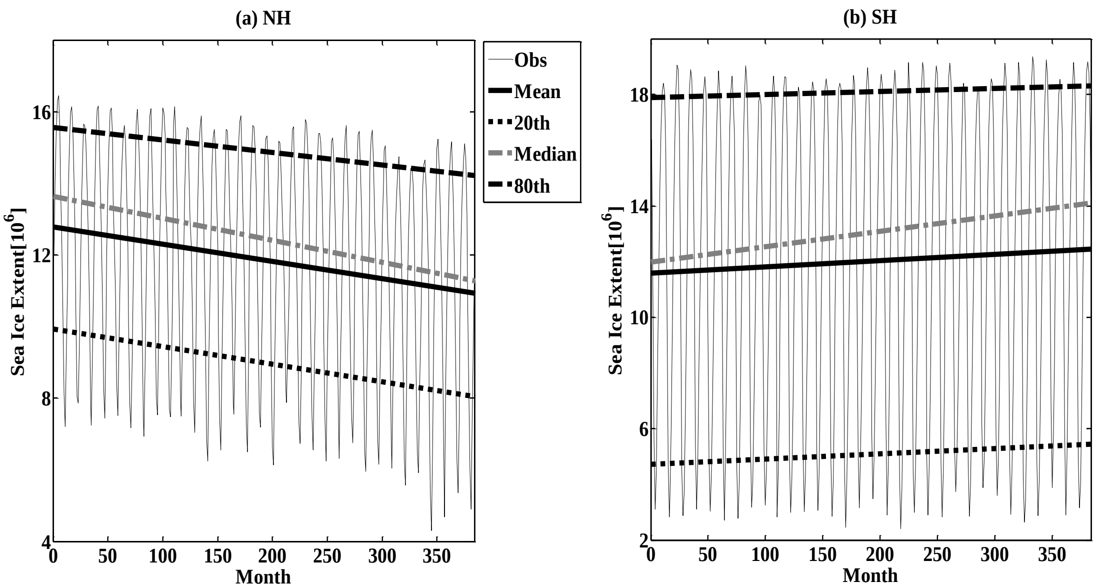

Figure 3.3. Quantile regressions (20th, 50th, and 80th percentiles) and standard linear regression of monthly sea ice extent (1979-2010) for (a) the northern hemisphere, and (b) the southern hemisphere

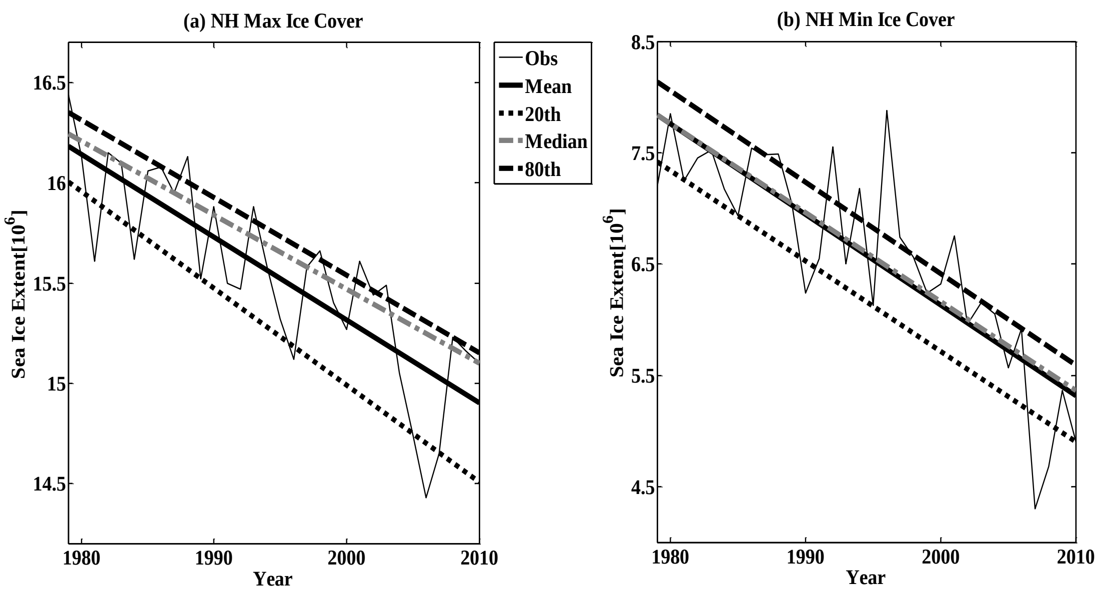

Figure 3.4. Quantile regressions (20th, 50th, and 80th percentiles) and standard linear regression of northern hemisphere sea ice extent for (a) annual maximum ice cover, and (b) annual minimum ice cover.

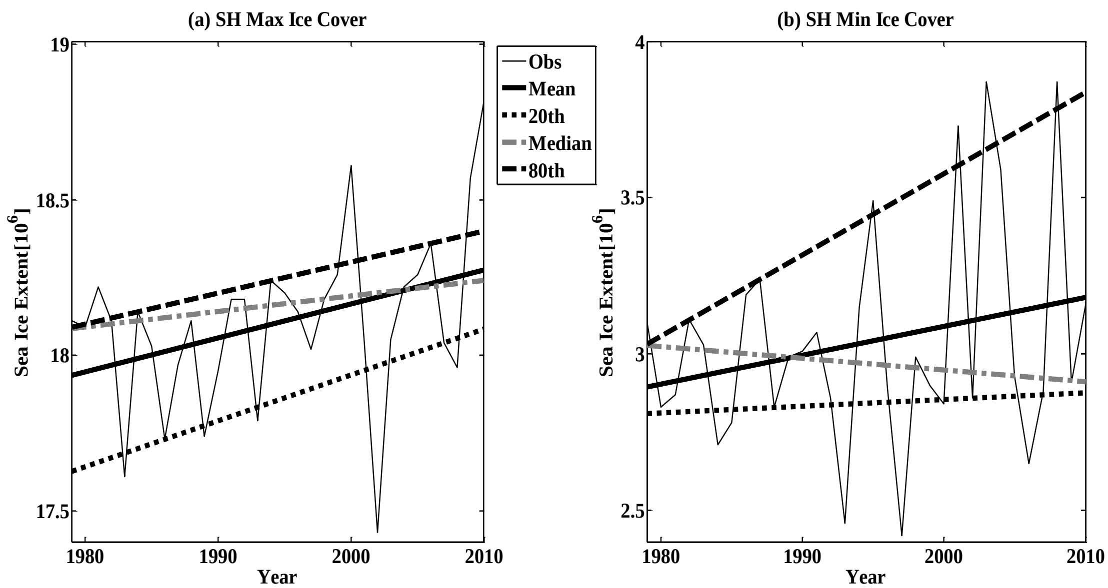

Figure 3.5. Quantile regressions (20th, 50th, and 80th percentiles) and standard linear regression of southern hemisphere sea ice extent for (a) annual maximum ice cover, and (b) annual minimum ice cover.

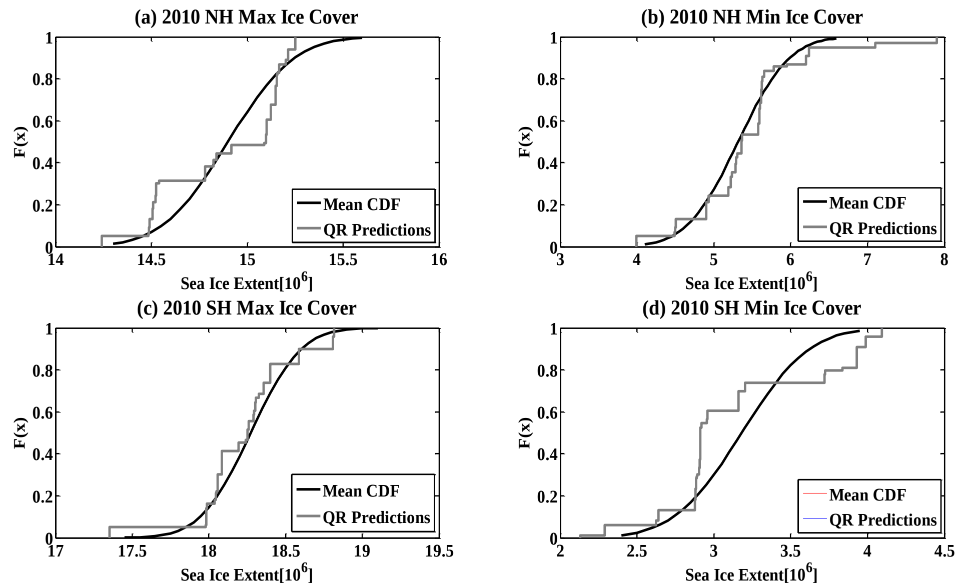

Figure 3.6. Comparison of cumulative distribution function of standard linear regression and quantile regression predictions for 2010 (a) Arctic maximum ice cover, (b) Arctic minimum ice cover, (c) Antarctic maximum ice cover, (d), Antarctic minimum ice cover.

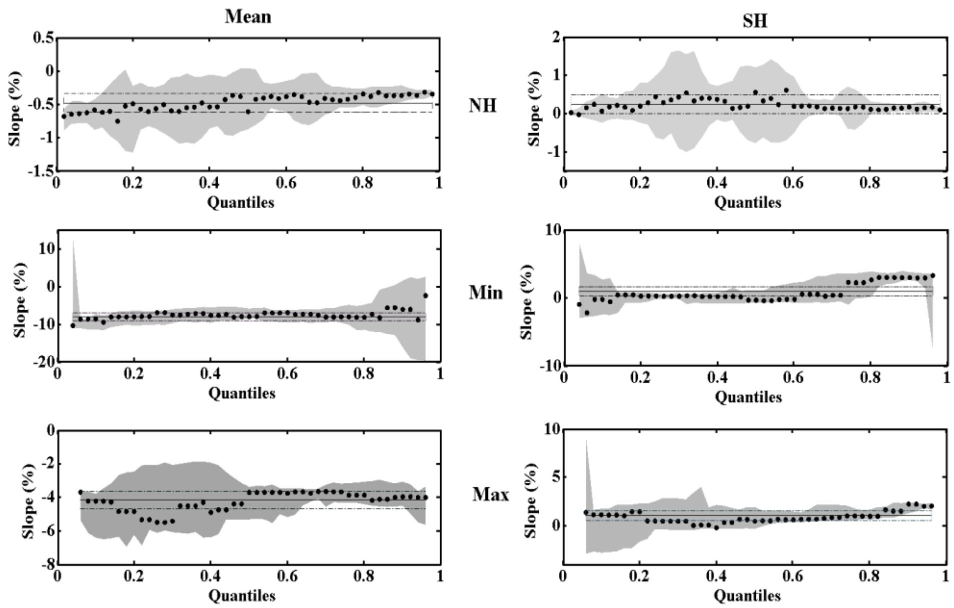

Figure 3.7. Slopes of estimated sea ice extent regression lines. The quantile regression coefficients (black dots) are presented with their 95% confidence bounds (shaded in grey). The least-squares regression coefficients (solid line) are also given with their 95% confidence.

Chapter 4

Article 2: A Study of Climate Extremes Changes over the Canadian Prairies using Quantile Regression

Abstract

Changes in the frequency and intensity of extreme events may have a dramatic impact on society and the natural environment. This study investigates changes in temperature and precipitation extremes over the Canadian Prairies (three stations in each of Alberta and Saskatchewan, four in Manitoba and one in Ontario). Quantile regression analysis has been applied to investigate trends in the annual and seasonal extremes using monthly homogenized data. In order to examine the advantages of quantile regression analysis over standard regression models for modeling extremes, mean trends have been compared with low and high (5% and 95%) quantile regression estimates. The results revealed that extreme temperatures have increased significantly over most of the Canadian Prairies. However, no general pattern can be detected in extreme precipitation.

4.1 Introduction

Extreme weather and climate events have significant impact on both human society and the natural environment. Loss of life and damage costs due to weather hazards have raised attention towards extreme events. In 2005, Hurricane Katrina caused 1500 deaths in New Orleans and Hurricane Mitch caused more than 10,000 deaths Central America. Significant loss of life has occurred due to flooding events in India and Indonesia. In the United States, the Midwest drought of 1988-1989, Hurricane Andrew in South Florida in 1992, and the Midwest flood of 1993 resulted in damages of 39, 30, and 19 billion dollars, respectively (Changnon et al., 2000). Although extreme events in Canada have not caused problems as serious as in the United States, some human life and economic losses have occurred. Four deaths resulted from three weeks of flooding in Calgary, and a debris flow at Five-Mile Creek in Banff National Park blocked the Trans-Canada Highway for several days at the peak of the tourist season (Evans, 2002). Other examples include an Edmonton thunderstorm, hailstorms in Winnipeg, and the Saguenay flood that produced insurance claims of 160, 120, and 350 million dollars (Environment Canada, 2004; White and Etkin, 1997).

Observations in many areas have shown that changes in total precipitation are significantly influenced by changes in the tails of precipitation distributions. Similarly, changes in some temperature extremes have been observed. There are some lingering questions about whether such changes are part of decadal fluctuations, or whether they are indicative of long-term trends related to climate change (Easterling et al., 2000). Generally, a significant decrease in the number of days with extreme cold temperatures, an increase in the number of days with extreme warm temperatures, and some detectable increase in the number of extreme wet days have been observed in many parts of the world (Vincent and Mekis, 2006).

Significant changes in Canada's temperature and precipitation have been observed during the twentieth century. According to Zhang et al. (2000), an increase of about 0.9˚C in the annual mean temperature along with a 12%-increase in mean annual precipitation has been observed in Southern Canada from 1900 to 1998.

Most of southern Canada has been faced with significant increasing trends in the lower and higher percentiles of the daily minimum and maximum temperature distributions (Bonsal et al., 2001). Precipitation extremes show no consistent trends in the number of events or in the intensity (Zhang et al., 2001).

The economic cost of extreme events over the Canadian Prairies, particularly in the agricultural, environmental, and hydroelectric sectors, calls for a detailed investigation of changes in extremes. Lawson (2003) investigated the extreme minimum winter temperatures over the Canadian Prairies during the period 1914-1994. He found a significant downward trend in the occurrence of very cold minimum daily temperatures during January and February. No significant trends were detected for the month of December and only a few sites were found to have significant trends in the month of March. Based on an association of tropical sea surface temperature and atmospheric circulation over the Canadian Prairies, Shabbar and Bonsal (2004) found longer winter warm spells during warm ENSO years across the Central Prairies.

In this paper, quantile regression has been applied to study changes in climate extremes over the Canadian Prairies. Monthly homogenized temperature and precipitation data have been used to investigate the annual and seasonal trends. Mean trends have been compared with low and high (5% and 95%) quantile regression estimates.

4.2 Quantile Regression

Standard regression models are extensively used in statistical analyses. Despite their popularity, these conditional mean models have some limitations. When interest is in the quantiles of the conditional distribution rather than the mean, standard regression models may fail to provide the desired information because the assumption of homoscedasticity may not be justified. Standard regression models can also fail to capture trends in heavy-tailed distributions.

Koenker and Bassett (1978) developed the quantile regression model for estimating the functional relationship between predictor variables and any quantile in the response distribution. Quantile regression overcomes some of the limitations of standard regression and is suitable for the study of changes in the frequency of environmental extremes over time.

Let (x_i, y_i), $i = 1,2,\dots,n$, be pairs of random variables where y_i is a time series of annual or seasonal temperature or precipitation and x_i is a covariate of y_i. The linear conditional quantile function has the following form:

$$Q_y(\tau|x) = \beta_0(\tau) + \beta_1(\tau)x + F^{-1}(\tau) \qquad (4.1)$$

where $\beta_0(\tau)$ and $\beta_1(\tau)$ are the intercept and slope coefficients of the τth quantile regression line. $F^{-1}(\tau)$ is the error distribution function where $E[F^{-1}(\tau)] = 0$. The τth quantile regression, $0 < \tau < 1$, can be defined as any solution to the following minimization function:

$$\hat{Q}_y(\tau) = \arg\min_{Q_y(\tau|x)} \left\{ \sum_{i:y_i \geq Q_y(\tau|x)} \tau \left| y_i - Q_y(\tau|x) \right| + \sum_{i:y_i < Q_y(\tau|x)} (1 - \tau) \left| y_i - Q_y(\tau|x) \right| \right\} \quad (4.2)$$

More information about the quantile regression technique can be found in Buchinsky (1998), Cade and Noon (2003), Yu et al. (2003), Koenker (2005), and Hao and Naiman (2007). A few studies have used this technique for climate modeling and climate change impact assessment (Table 4.1).

The “quantreg” package add-on to the R language (http://www.r-project.org/) can be used for quantile regression analysis. Other packages for quantile regression include Blossom statistical package, Stata and Shazam. In this study, the quantreg package combined with Matlab coding has been used for quantile regression modeling.

4.3 Study Area and Data

The meteorological stations selected for the study are listed in Table 4.2 and their locations are shown in Figure 4.1. The stations cover the southern part of the Canadian Prairies, as well as Churchill in northern Manitoba.

Only stations with 45 years of good records covering the period 1945-1990 were used in this study. Data were extracted from the homogenized Canadian monthly temperature and precipitation archive of weather data. The data have been adjusted to account for inhomogenities caused by station alterations including changes in site exposure, location, instrumentation, observer, observing program, or a combination of the above (Vincent, 1998; Vincent and Gullett, 1999).

4.4 Results

In order to investigate the changes in climate extremes, the 5%- and the 95%-quantiles of annual and seasonal distributions have been estimated by quantile regression and compared with changes in the trend of mean levels. As mentioned above, the application of standard regression for modeling extreme events may lead to incorrect conclusions regarding percentiles in the conditional distributions. As illustrated in Figure 4.2, percentiles in the conditional distribution of temperature and precipitation extremes may be different from trends in the mean. For example, annual precipitation in Dauphin and autumn temperature in Estevan show no particular change in the mean over time, but there is evidence that the precipitation distribution is narrowing over time and there is a significant decreasing trend in temperature extremes as illustrated by the results of quantile regressions of the 5^{th} and 95^{th} percentiles.

Figures 4.3 and 4.4 summarize the results of the estimated trends in the mean, the 5^{th} and the 95^{th} percentiles of annual and winter distributions of temperature and precipitation. Figure 4.3 reveals a significantly increasing trend in annual mean temperatures over all the Prairies. For warm (95^{th} percentiles) and cold (5^{th} percentiles) temperatures, trends of similar magnitude can be seen in the central and eastern part of the Prairies. It is a different story for annual precipitation where no particular spatial pattern of change in mean and percentiles can be found.

In the winter season, there is a warming trend in the mean and the cold and warm temperature percentiles over all of the Prairies except for western Alberta. Most of the southern parts of the Prairies experienced a precipitation decrease, especially in low quantiles. Figure 4.4 shows that winters with high precipitation are becoming less frequent in Manitoba except for Churchill.

As a brief description of the results for other seasons, spring temperatures have similar trends as annual temperatures except over northern Manitoba where warm extremes are becoming colder. The temperature in the central Prairies has increased considerably in summer, while the central and eastern Prairies have experienced a significant negative trend in high precipitation quantiles. Interestingly, most of the stations show a negative trend in mean and extremes temperature in autumn, while there are no discernible trends in extreme precipitation.

4.4.1 Summary of Results

Of the 11 stations studied, 10 showed an increasing annual temperature trend for warm and cold quantiles. Table 4.3 shows where the most significant trends in annual and seasonal temperature have been observed. Most of these stations are from the central Prairies, demonstrating the high impact of climate change over this area. An average increase of 1.8 and 1.6 ˚C in annual warm and cold extremes as well as seasonal increases (e. g., 1.7˚C average increase in winter cold extremes and 1.0 ˚C average increase in spring warm extremes) for the prairie region. In the case of cold extremes (low quantiles), changes of different magnitudes have been observed. In the case of warm extremes, the central and eastern Prairies show a significant positive trend. Western Alberta shows a different temperature trend than other parts of the Prairies which may be because of the influence of the Rocky Mountains.

As seen in Table 4.4, more variable (positive and negative) trends are observed in precipitation changes which suggest that no general pattern of change in precipitation can be detected for the prairie region. Winter dry extremes show a decreasing trend in most of the stations (8 stations), and the eastern Prairies would expect less precipitation in winter wet extremes.

4.5 Discussion and Conclusions

This study has examined the complementary value of quantile regression analysis over standard regression models for modeling temperature and precipitation extremes. The results show that some of the extreme patterns might have gone undetected if a standard regression model had been used. Quantile regression on the other hand provides a more detailed picture of these patterns. We suggest that quantile regression analysis is a valuable tool in climate change studies.

Bonsal et al. (2001) and Vincent and Mekis (2006) indicated that most of southern Canada show significant increasing trends in the lower and higher percentiles of the temperature distribution over the twentieth century. The results from the present study reach similar conclusions for the entire Prairie region. No consistent pattern of trends in precipitation extremes was found for the Prairie region, supporting the conclusions of Zhang et al. (2001) and Vincent and Mekis (2006) for all of Canada.

Lawson (2003) investigated the extreme minimum winter temperatures at selected Prairie locations. He found a significant trend towards fewer occurrences of very cold minimum daily temperature, particularly in the southern, central and western Prairies during the period 1914-94. Quantile regression estimates for winter cold temperatures reveal significant increasing trends for most of the northern and eastern Prairies as well as the central Prairies.

Trends in extreme temperature and precipitation events were investigated over the Canadian Prairies. It was found that temperature has increased significantly over most of the Prairies. Many of the detected trends are expectedly to be associated with the global mean temperature increase.

Our analysis did not demonstrate any significant patterns of trends in precipitation in the Prairie region. The difficulty of detecting trends in precipitation extremes is in some cases caused by the high variability of the extreme events and in other cases by the short records available for analysis. Further investigations on circulation patterns and the relationship between low frequency variability modes like El Niño-Southern Oscillation (ENSO) and the Arctic Oscillation (AO) may provide further information about changes in precipitation extremes over the Canadian Prairies.

References:

Baur, D., Saisana, M., Schulze, N., 2004. Modeling the effects of meteorological variables on ozone concentration - A quantile regression approach. Atmospheric Environment, 38 (28), pp. 4689-4699.

Bonsal, B. R., Zhang, X., Vincent, L. A., Hogg, W. D., 2001. Characteristics of daily and extreme temperatures over Canada. Journal of Climate, 14 (9), pp. 1959-1976.

Bremnes, J. B., 2004. Probabilistic wind power forecasts using local quantile regression. Wind Energy 7 (1), pp. 47-54.

Buchinsky, M., 1998. Recent advances in quantile regression models: A practical guideline for empirical research. Journal of Human Resources, 33 (1), pp. 88-126.

Cade, B. S., Noon, B. R., 2003. A gentle introduction to quantile regression for ecologists. Frontiers in Ecology and the Environment, 1 (8), pp. 412-420.

Changnon, S. A., Pielke Jr., R .A., Changnon, D., Sylves, R. T., Pulwarty, R., 2000. Human factors explain the increased losses from weather and climate extremes. Bulletin of the American Meteorological Society, 81 (3), pp. 437-442.

Easterling, D. R., Meehl, G. A., Parmesan, C., Changnon, S. A., Karl, T. R., Mearns, L. O., 2000. Climate extremes: Observations, modeling, and impacts. Science, 289 (5487), pp. 2068-2074.

Elsner, J. B., Kossin, J. P., Jagger, T. H., 2008. The increasing intensity of the strongest tropical cyclones. Nature, 455 (7209), pp. 92-95.

Environment Canada, 2004. Top ten Canadian weather stories for 2004. Environment Canada, <http://www.msc-smc.ec.gc.ca/media/top10/2004_e.html#topten>, [accessed December 14th, 2008].

Evans, S. G., 2002. Climate change and geomorphological hazards in the Canadian Cordillera; The anatomy of impacts and some tools for adaptation. Natural Resources Canada, Climate Change Impacts and Adaptation Program, Report A099, 14p., <http://adaptation.nrcan.gc.ca/projdb/pdf/71_e.pdf>, [accessed December 14th, 2008].

Friederichs, P., Hense, A., 2007. Statistical downscaling of extreme precipitation events using censored quantile regression. Monthly Weather Review, 135 (6), pp. 2365-2378.

Hao, L., Naiman, D. Q., 2007. Quantile regression. SAGE, Thousand Oaks, California, USA.

Koenker, R., 2005. Quantile Regression. Cambridge University Press, Cambridge, New York, USA.

Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica, 46 (1), pp. 33-50.

Koenker, R., Schorfheide, F., 1994. Quantile spline models for global temperature change. Climatic Change 28 (4), pp. 395-404.

Lawson, B. D., 2003. Trends in winter extreme minimum temperatures on the Canadian Prairies. Atmosphere – Ocean, 41 (3), pp. 233-239.

Sankarasubramanian, A., Lall, U., 2003. Flood quantiles in a changing climate: Seasonal forecasts and causal relations. Water Resources Research 39 (5), pp. SWC41-SWC412.

Shabbar, A., Bonsal, B. 2004. Associations between low frequency variability modes and winter temperature extremes in Canada. Atmosphere - Ocean 42 (2), pp. 127-140.

Tøttrup, A. P., Thorup, K., Rahbek, C., 2006. Patterns of change in timing of spring migration in North European songbird populations. Journal of Avian Biology, 37 (1), pp. 84-92.

Vincent, L. A., 1998. A technique for the identification of inhomogeneities in Canadian temperature series. Journal of Climate 11 (5), pp. 1094-1104.

Vincent, L. A., Gullett, D. W., 1999. Canadian historical and homogeneous temperature datasets for climate change analyses. International Journal of Climatology, 19 (12), pp. 1375-1388.

Vincent, L. A., Mekis, É., 2006. Changes in daily and extreme temperature and precipitation indices for Canada over the twentieth century. Atmosphere – Ocean, 44 (2), pp. 177-193.

White, R., Etkin, D., 1997. Climate change, extreme events and the Canadian Insurance Industry. Natural Hazards, 16 (2-3), pp. 135-163.

Wilson Jr., W. H., 1997. Spring arrival dates of migratory breeding birds in Maine: Sensitivity to climate change. Wilson Journal of Ornithology, 119 (4), pp. 665-677.

Woodall, C. W., Liknes, G. C., 2008. Relationships between forest fine and coarse woody debris carbon stocks across latitudinal gradients in the United States as an indicator of climate change effects. Ecological Indicators, 8 (5), pp. 686-690.

Yu, K., Lu, Z., Stander, J., 2003. Quantile regression: Applications and current research areas. Journal of the Royal Statistical Society Series D: The Statistician, 52 (3), pp. 331-350.

Zhang, X., Hogg, W. D., Mekis, É., 2001. Spatial and temporal characteristics of heavy precipitation events over Canada. Journal of Climate, 14 (9), pp. 1923-1936.

Zhang, X., Vincent, L. A., Hogg, W. D., Niitsoo, A., 2000. Temperature and precipitation trends in Canada during the 20th century. Atmosphere – Ocean, 38 (3), pp. 395-429.

Table 4.1. Quantile regression applications in climate modeling and climate change impact assessment

Global temperature change over the last century	Koenker and Schorfheide (1994)
Estimation of flood quantiles in a changing climate	Sankarasubramanian and Lall (2003)
Forecasting probabilistic wind power	Bremnes (2004)
Modeling the effects of meteorological variables on Ozone concentration	Baur et al. (2004)
Changes in timing of autumn migration in North European songbird populations	Tottrup et al. (2006)
Statistical downscaling of extreme precipitation	Friederichs and Hense (2007)
Analysis of relationship between spring temperature and arrival dates of migratory breeding birds	Wilson Jr. (2007)
Climate change effects on relationships between forest fine and coarse woody debris carbon stocks	Woodall and Liknes (2008)
Tropical cyclones trend changes	Elsner et al. (2008)
Changes of extreme annual rainfall in Zimbabwe	Mazvimavi (2008)

Table 4.2. Location and available data for utilized stations.

Province	Station	Latitude	Longitude	Average Annual Temperature (˚C)	Average Annual Precipitation (mm)
Alberta	Beaverlodge	55.20	-119.40	1.77	518.1
	Calgary	51.12	-114.02	3.74	480.2
	Edson	53.58	-116.47	1.84	606.0
Saskatchewan	Estevan	49.07	-103.00	3.01	493.6
	Klintonel	49.68	-108.92	2.44	494.5
	Prince albert	53.22	-105.68	0.52	486.8
Manitoba	Churchill	58.73	-94.07	-7.23	553.6
	The Pas	53.97	-101.10	1.57	577.7
	Dauphin	51.15	-100.3	-0.41	534.0
	Winnipeg	49.90	-97.23	1.79	604.7
Ontario	Sioux Lookout	50.12	-91.90	1.47	797.6

Table 4.3. Temperature changes (˚C) at stations with the most significant trends

Period	Station	Mean	Quantiles	
			5%	95%
Annual	Dauphin	1.30	2.31	2.57
	Sioux Lookout	1.12	2.31	3.09
	Beaverlodge	1.44	-1.02	-2.83
	Estevan	1.70	2.57	2.44
Winter	Beaverlodge	-2.96	-1.97	-3.30
	Calgary	2.35	1.84	5.12
	Estevan	2.60	2.09	5.89
	Prince Albert	2.80	3.57	5.40
Spring	Sioux Lookout	2.47	3.92	2.52
	Beaverlodge	-2.20	-2.41	-1.41
	Estevan	3.25	3.19	2.30
	Prince Albert	2.95	4.28	0.77
Summer	Winnipeg	0.72	0.13	1.67
	Edson	1.06	1.58	0.32
	Estevan	1.17	1.00	2.63
	Klintonel	1.27	1.59	1.85
Autumn	Churchill	-0.81	-1.01	-1.45
	Buffalo Narrows	-0.53	-1.32	-1.29
	Estevan	-0.27	-2.25	-0.58
	Klintonel	-0.19	-2.08	-2.12

Table 4.4. Precipitation changes (%) at stations with the most significant trends.

Period	Station	Mean	Quantiles 5%	Quantiles 95%
Annual	Dauphin	-2.64	76.34	-23.18
	Edson	20.31	11.24	20.72
	Estevan	-6.50	34.24	25.19
	Klintonel	0.30	-35.67	-8.49
Winter	Churchill	54.05	65.03	39.88
	The Pas	15.83	150.06	-32.25
	Dauphin	-26.89	-49.66	-44.13
	Klintonel	45.53	-21.25	115.94
Spring	Churchill	27.54	16.25	-35.81
	Sioux Lookout	-5.18	-19.85	42.34
	Beaverlodge	-9.91	-64.30	63.21
	Calgary	-7.58	-24.22	-29.37
Summer	Churchill	9.57	81.86	-10.96
	Beaverlodge	23.62	101.72	-23.95
	Estevan	-20.06	359.86	3.75
	Prince Albert	16.74	47.85	-11.96
Autumn	Dauphin	26.30	70.77	54.18
	Calgary	16.97	50.00	40.18
	Estevan	-8.87	-60.38	-30.84
	Klintonel	-43.72	-69.91	-59.53

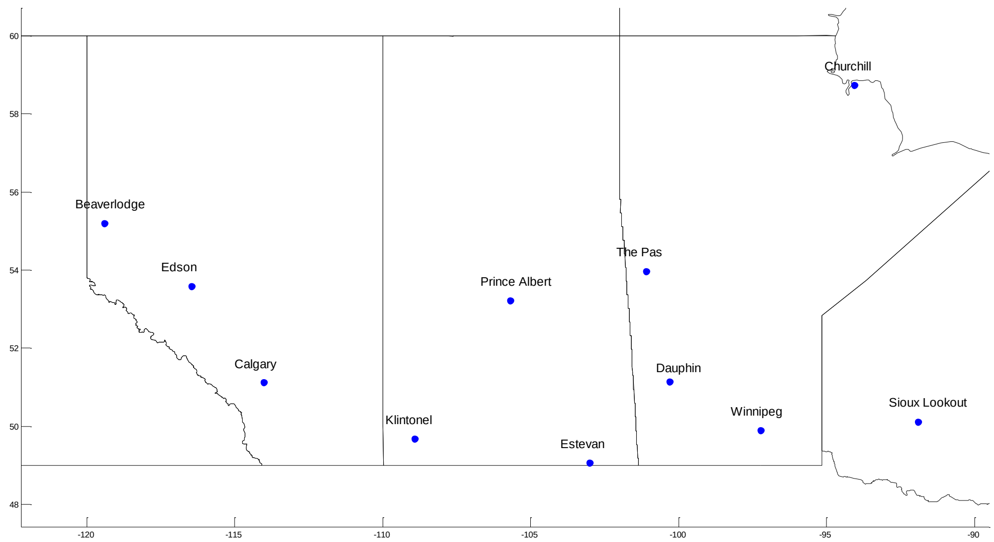

Figure 4.1. Study Area and stations used in the study

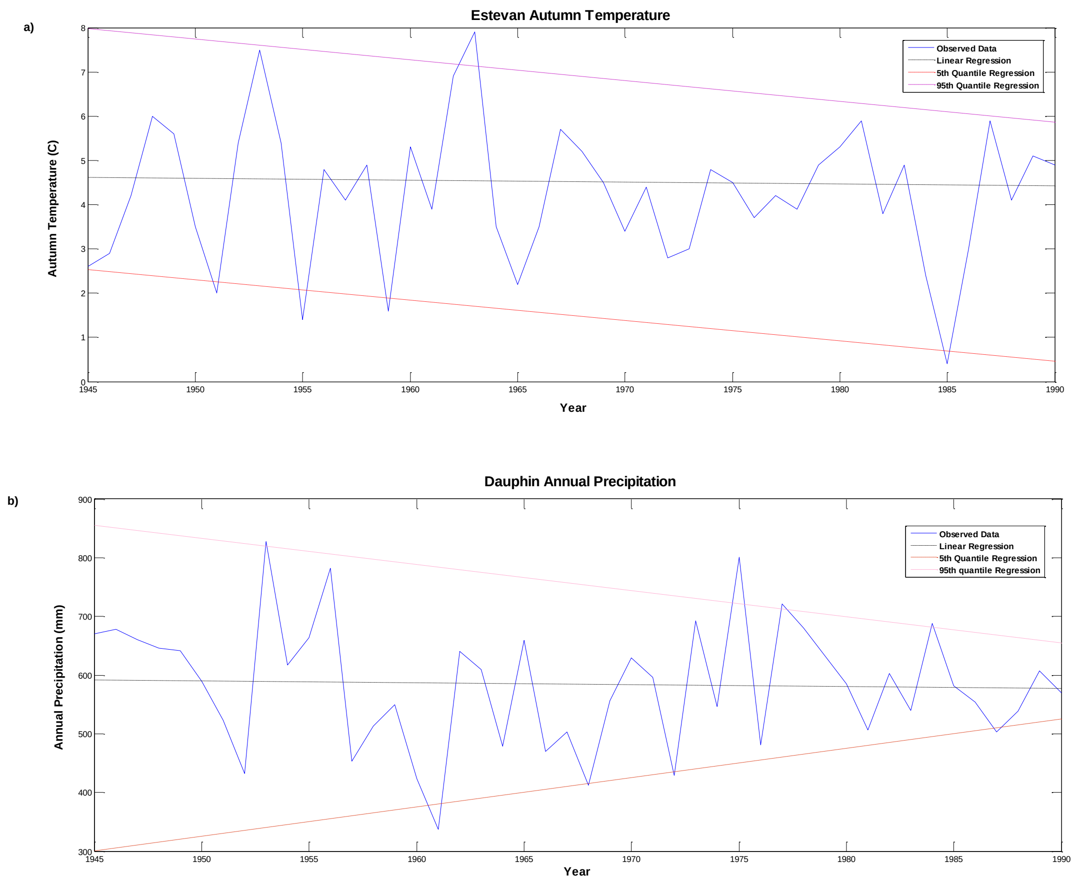

Figure 4.2. Examples of quantile regressions (5th and 95th percentiles) and standard linear mean regression trends for a) Estevan autumn temperature and b) Dauphin annual precipitation.

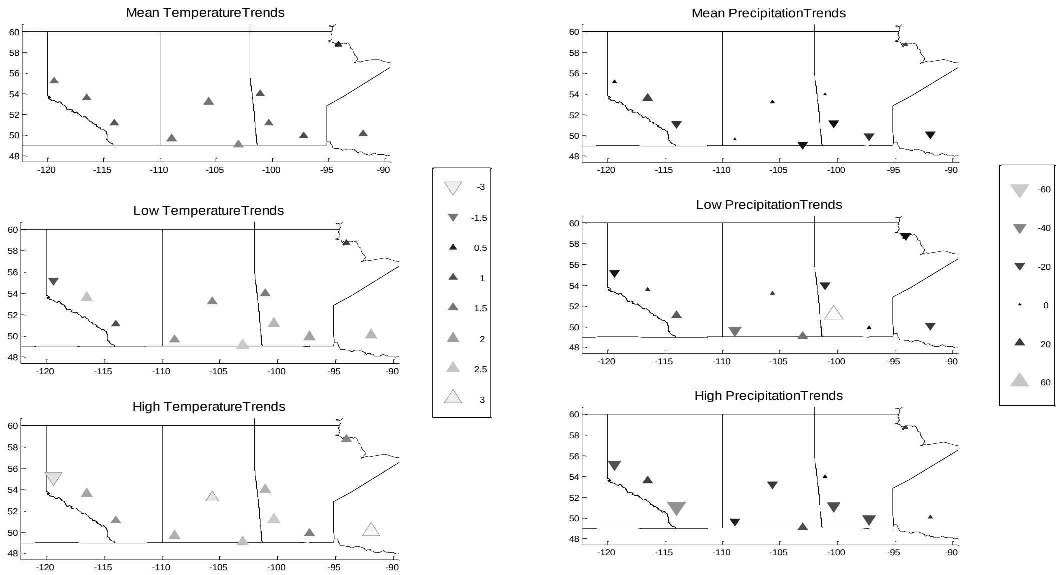

Figure 4.3. Trends in annual mean and extreme temperatures (˚C) and precipitations (%). The numbers represent the changes in temperature and precipitation from 1945 to 1990 based on the slope of the regression line.

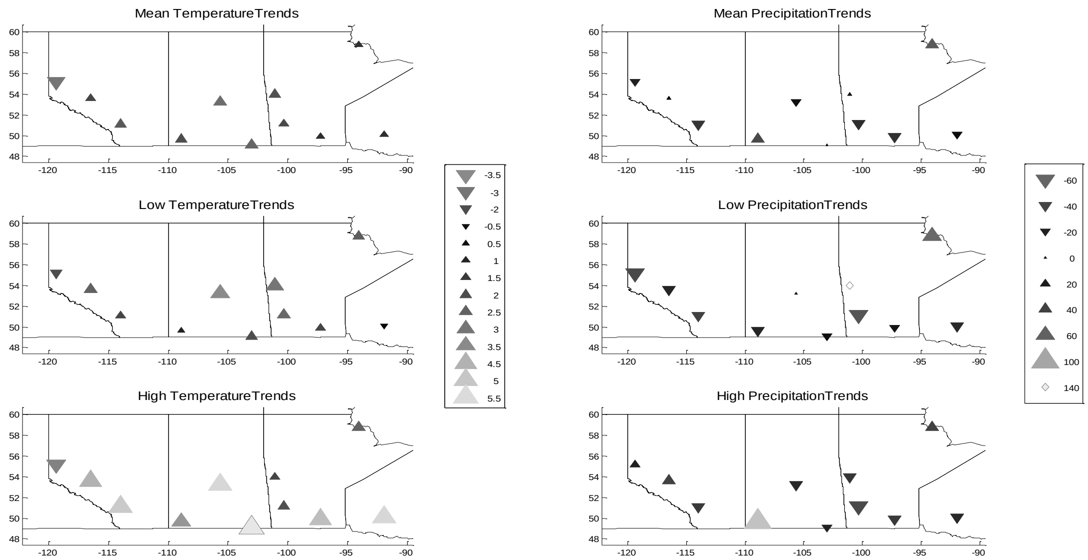

Figure 4.4. Trends in winter mean and extreme temperatures (˚C) and precipitations (%). The numbers represents the changes in temperature and precipitation from 1945 to 1990 based on the slope of the regression line.

Chapter 5

Article 3: Statistical Downscaling of Precipitation using Quantile Regression

Abstract

Statistical downscaling is based on establishing an empirical relationship between predictands (a set of regional climate variables) and predictors (a set of large-scale atmospheric variables). The well-known challenge of predictor selection, variance underestimation, and poor representation of extreme events by statistical downscaling methods motivated us to investigate the use of quantile regression to better predict extreme precipitation. Quantile regression extends traditional regression models to the estimation of conditional quantile functions. A Bayesian method adapted to quantile regression was used to select predictor variables. The advantage of this method is that different predictors can be selected for different parts of the conditional distribution. Fifteen potential predictor variables were used to downscale precipitation amounts at five stations in the Canadian Prairie region. A novel method, involving the fitting of different distributions to predicted regression quantiles, has been introduced to downscale precipitation. The results show the general superiority of the quantile regression statistical downscaling model over the standard regression model, not only in the tails but over the entire distribution for the summer precipitation, while in the case study there was little difference between the two models in the case of winter precipitation.

Keywords: Statistical downscaling, Precipitation, Quantile regression, Variable selection

5.1 Introduction

It is well known that there is a general mismatch between the spatial resolution of output from global climate models (GCM) and the scale of interest in regional assessments of climate change impacts. Various downscaling techniques have been developed in an attempt to bridge this resolution gap. Downscaling methods are usually classified as either dynamical or statistical. Dynamical downscaling involves the use of high-resolution, limited-area climate models within the domain of interest, whereas statistical downscaling use relatively simple statistical models to represent the link between atmospheric circulation variables, presumably well simulated by the GCMs, and local weather variables, such as precipitation and temperature. In a very simplistic form, a statistical downscaling model can be written as $Y = f(\mathbf{X})$ where Y is the predictand such as precipitation and $\mathbf{X}$ is a vector of atmospheric variables such as geopotential height, wind speed, humidity variables, etc. The functional relationship is established using observed time series. The two main challenges in statistical downscaling are the determination of the functional relationship and the identification of the predictor variables than convey the most relevant information about the predictand and the climate change signal.

A range of techniques have been proposed to model the relationship between a predictand and a set of predictor variables. These methods include the classical multiple regression models (e.g. Jeong et al., 2012), techniques using principal components of pressure fields or geopotential heights (e.g. Li and Smith, 2009), and more sophisticated methods such as artificial neural networks (e.g. Tolika et al., 2008), canonical correlation analysis (e.g. Palatella et al., 2010), and singular value decomposition (e.g. Chen et al., 2011).

Fowler et al. (2007) discussed several key assumptions inherent in statistical downscaling techniques. Predictor variables must be physically sensible, realistically modeled by the GCM, and able to fully reflect the climate change signal. For example, the use of only circulation-based predictors may not be a good idea, because the change in atmospheric humidity in a warmer climate might be disregarded. Fowler et al. (2007) also emphasized the inherent assumption that the predictor–predictand relationship remains valid in the future. Wilby (1998) attributed the potential non-stationarity of predictor–predictand relationships in statistical downscaling models to three factors: an inadequate sampling or calibration period, an incomplete set of predictor variables that disregards low-frequency climate behaviour, and, most seriously, temporal change in climate system structures. For example, there is a risk that an important predictor of climate change may appear insignificant when developing a downscaling model under present climate. However, failure to include such a variable may seriously limit the ability to correctly project climate change (Buishand et al., 2004). Although predictor selection is an important component in the development of a statistical downscaling model, it is often approached in a rather superficial way.

Conventional regression models have been used in numerous statistical downscaling studies and are the cornerstone of software packages such as SDSM (Wilby et al., 2002) and ADS (Hessami et al., 2008). Despite their popularity, these conditional mean models have some limitations. When interest is in the quantiles of the conditional distribution rather than the mean, standard regression models may fail to provide the desired information because the assumption of homogenous variance may not be justified. Also, it is common practice to assume that regression residuals are normally distributed but this may not be a valid assumption, even after application of some normalizing transformation. The non-normality of residuals may not be a serious issue if the only interest is in the mean of the conditional distribution. However, when interest is in the tails of the conditional distribution, the distribution of residuals becomes important. We also note that conventional regression models can be sensitive to outliers. While methods are available to deal with outliers, it is an issue that is often not properly dealt with in practice.

In this paper, we address the problem of downscaling daily precipitation from global climate model output using regression models. One common characteristic of predictor–predictand relationships is unequal variation of precipitation in relation to large-scale forcing. To satisfy the assumption of homogeneous variance of regression residuals, it is common practice to transform observed precipitation first, for example by taking the third or fourth root, before estimating the regression model. An alternative point of view is that unequal variation in precipitation reflects important processes that should be modeled explicitly.

Koenker and Bassett (1978) introduced quantile regression for estimating the functional relationship between predictor variables and any quantile in the response distribution. Quantile regression has received considerable attention in the statistical literature, but less so in the water resources literature. Quantile regression overcomes some of the limitations of standard regression models mentioned above and provides a more complete picture of the relationships between variables that may be missed by other regression methods.

On the issue of variable selection, we note that the classical approach of stepwise regression does not always produce satisfactory results. For example, the choice of sub-set variables can in some cases be quite sensitive to small changes in the data, leading to instability in the variable selection (Breiman, 1996). In recent years, penalized likelihood methods for variable selection have attracted much interest. Techniques such as the Lasso (Tibshirani, 1996), the SCAD (Smoothly Clipped Absolute Deviation; Fan and Li, 2001), and the LARS (Least Angle Regression; Efron et al., 2004) have been proposed as alternatives to classical stepwise regression. In this paper, we employ a technique called Bayesian stochastic search variable selection (SSVS) (George and McCulloch, 1993). The SSVS method has been adapted to variable selection in quantile regression by Reed et al. (2009), Li et al. (2010), and Alhamzavi and Yu (2012). In this study, the variable selection model developed by Reed et al. (2009), called SSVSquantreg, is used and will be briefly described later.

Bremnes (2004), Friederichs and Hense (2007), Friederichs (2010), and Cannon (2011) applied quantile regression to downscale precipitation, but their focus was on the capability of quantile regression for forecasting, and not so much the analysis of extreme precipitation and relationships with large-scale variables. In addition to developing a novel statistical downscaling model based on quantile regression, we will specifically focus on the analysis of extreme precipitation. In the above-mentioned studies, sets of 1 to 3 similar predictors were used, but here, a Bayesian method will be applied to select the predictors for each station and season. In our case study, different predictor variables are selected to downscale daily precipitation amounts at five stations in the Canadian Prairie provinces for the winter and summer seasons. A novel method is then introduced to downscale precipitation by fitting probability distributions to predicted regression quantiles. The performance of this new procedure, with special emphasis on extreme precipitation, is evaluated by comparison with traditional regression-based downscaling.

5.2 Methodology

5.2.1 Quantile regression

In this section, we provide a brief description of linear quantile regression, the main tool used in the remainder of the paper for downscaling precipitation. Quantile regression was proposed by Koenker and Bassett (1978). In quantile regression, a regression model is developed for selected quantiles of the conditional distribution of the response variable. Contrary to conventional linear regression, quantile regression does not assume homogenous residual variance and does not make any assumptions about the error distribution. In this sense, it can be considered a more flexible approach than linear regression. In cases where all assumptions of the conventional linear regression model are satisfied, quantile regression should in principle provide answers similar to the traditional model.

While standard regression seeks a model for the conditional expectation of the response variable, quantile regression is concerned with the determination of models for user-selected quantiles in the conditional distribution of the response *y*. The linear conditional quantile function employed in this paper has the form

$$Q_p(y \mid \mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta}_p \tag{5.1}$$

where $\mathbf{x}$ is a vector of explanatory variables and $\boldsymbol{\beta}_p$ is a vector of parameters related to the *p*th quantile regression. The parameters may include an intercept for the regression model in which case the first element of $\mathbf{x}$ is a 1. In linear quantile regression, each quantile of the conditional distribution is represented by an individual hyper-plane. For a given set of observations $(\mathbf{x}_i, y_i),\ i = 1, \ldots, n$, an estimate of the parameters $\boldsymbol{\beta}_p$ is given by:

$$\hat{\boldsymbol{\beta}}_p = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_p(y_i - \mathbf{x}_i^T \boldsymbol{\beta}) \tag{5.2}$$

where the function $\rho_p(\cdot)$ is defined as

$$\rho_p(u) = \begin{cases} u(p-1) & \text{if } u < 0 \\ up & \text{if } u \geq 0 \end{cases} \tag{5.3}$$

In other words, the absolute value of the difference between an observation y_i and the corresponding *p*'th quantile $\hat{y}_p(x_i)$ is weighted by $(1-p)$ if the observation is below the quantile plane and by *p* if the observation is above the plane. As a special case, one sees that the median regression model is obtained by minimizing the sum of absolute errors. This can be compared with the mean regression model which involves the minimization of the sum of squared errors. One of the characteristics of the fitted quantile plane is that a fraction *p* of observation points will lie below the surface, just as one would expect. Further details about quantile regression models can be found in Cade and Noon (2003), Koenker (2005), and Hao and Naiman (2007). In this study, we have employed the R-package quantreg for model estimation.

Table 5.1 lists some recent studies that have employed quantile regression for environmental modeling and climate change impact assessment.

5.2.2 Selection of predictor variables in quantile regression

In regression analyses, one typically starts out with a large number of potential predictor variables and part of the modeling task involves identifying the best sub-set of predictors. Models with an excessive number of predictors tend to have poor prediction accuracy and are difficult to interpret. In traditional regression analysis, stepwise regression is often used although several studies have pointed to a range of shortcomings of this procedure (Harrell, 2001). In recent years, much interest in the statistical literature has focussed on shrinkage methods such as Ridge regression and the Lasso method, see for example Hastie et al. (2001). Most of the methods applicable for traditional linear models cannot be directly applied to quantile regression. In this study, we employ a Bayesian method for variable selection. The general principles of the method were developed for traditional linear models. Madigan et al. (1996) showed that the predictive accuracy of models selected using Bayesian Model Averaging (BMA) is always higher than any single model. The BMA method has been adapted to quantile regression by Reed (2009).

Here we provide only a brief outline of the BMA method for variable selection in quantile regression models. A more detailed description can be found in Reed (2009) and the method is implemented in the R-package SSVSquantreg, an abbreviation of "Stochastic Search Variable Selection for Quantile Regression". The Bayesian analysis implemented in SSVSquantreg aims at producing posterior probabilities for the various model candidates where "model candidate" refers to a particular subset of predictor variables. If there are *k* potential predictor variables, then there are 2^k possible model candidates. The model candidates are identified by the binary vector variable $\gamma = (\gamma_1, \ldots, \gamma_k)$ where $\gamma_i = 1$ if the coefficient of predictor *i* is large and is equal to zero if the coefficient is small. The model parameters are therefore $(\boldsymbol{\beta}, \gamma)$. Starting with a prior distribution $\pi(\boldsymbol{\beta}, \gamma)$ and letting $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ denote the observations, our goal is to determine the posterior distribution $\pi(\boldsymbol{\beta}, \gamma \mid \mathbf{y})$. We are specifically interested in the model probabilities $\pi(\gamma \mid \mathbf{y})$ and $\boldsymbol{\beta}$ must therefore be eliminated by marginalization. There are two main challenges in the Bayesian analysis: (1) Analytical solutions are generally not available so numerical methods must be applied; and (2) an accurate estimate of the posterior probability of the 2^k model candidates can be an overwhelming task even for moderate values of *k*. Both of these obstacles can be overcome using the Gibbs sampler, a well-known Markov Chain Monte Carlo method.

The prior distribution of $(\boldsymbol{\beta}, \gamma)$ can be decomposed as $\pi(\boldsymbol{\beta}, \gamma) = \pi(\boldsymbol{\beta} \mid \gamma)\pi(\gamma)$. For $\pi(\gamma)$ SSVSquantreg assigns a Bernoulli prior to each of the indicator variables γ_j

independently, with prior probability of inclusion π_0. A hierarchical approach is used to specify π_0 which is not given a specific value but is treated as an unknown parameter described by a Beta distribution, $\pi_0 \sim B(a,b)$, where a and b are hyper-parameters. According to Reed (2009), treating π_0 as an unknown parameter leads to a more flexible prior which allows more information to be extracted from the data about the model size.

The prior distribution of the regression parameters $\boldsymbol{\beta}$ is defined independently for each component as

$$\pi(\beta_j \mid \gamma_j) = (1-\gamma_j)\delta_0 + \gamma_j C(0,1) \tag{5.4}$$

where the quantity δ_0 is a degenerate distribution with all probability mass at 0 and $C(0,1)$ is a Cauchy distribution with parameters 0 and 1. The above specification assumes that all predictors have been standardized to have mean value 0 and variance 1. The rationale for using the Cauchy distribution is outlined in Reed (2009).

Bayes' theorem defines the posterior distribution as

$$\pi(\boldsymbol{\beta},\gamma \mid \mathbf{y}) = \frac{f(\mathbf{y} \mid \boldsymbol{\beta},\gamma)\pi(\boldsymbol{\beta},\gamma)}{f(\mathbf{y})} \tag{5.5}$$

where $f(\mathbf{y} \mid \boldsymbol{\beta},\gamma)$ is the density function of the observations given the model and the model parameters, or equivalently the likelihood function when viewed as a function of $(\boldsymbol{\beta},\gamma)$ for given $\mathbf{y}$. A complication in the use of Bayesian analysis in quantile regression is that the distribution of regression errors is not specified and the likelihood function for the parameters is therefore undefined. Yu and Moyeed (2001) proposed the use of a pseudo-likelihood function of the form:

$$\ell(\boldsymbol{\beta},\gamma \mid \mathbf{y}) = f(\mathbf{y} \mid \boldsymbol{\beta},\gamma) = p^n (1-p)^n \exp\left(-\sum_{i=1}^{n} \rho_p (y_i - \mathbf{x}_{i,(\gamma)}^T \boldsymbol{\beta}_{p,(\gamma)})\right) \tag{5.6}$$

where $\mathbf{x}_{i,(\gamma)}$ is the sub-set of predictors for model γ and $\boldsymbol{\beta}_{p,(\gamma)}$ are the corresponding regression coefficients. The above likelihood function is equivalent to assuming that the residuals from the quantile regression have an asymmetric Laplace distribution. This is not an unreasonable approximation and it is used here primarily for implementing the variable selection procedure. The final estimation does not make assumptions about the distribution of residuals. It can be shown that the maximization of (5.6) with respect to the model parameters is equivalent to the solution of (5.2).

The Gibbs sampler can be set up to simulate the posterior distribution. This involves simulating each of the model parameters independently, conditional on previously simulated values of all other parameters. The limiting distribution of this Markov chain converges to the posterior distribution. In particular, the sequence of model candidates that are generated in the Gibbs sampler can be used to estimate the posterior model probabilities. Obtaining an accurate estimate of all 2^k models may require a very large number of iterations of the Gibbs sampler. However, the most likely models will occur frequently and quickly, and since there is limited interest in "unlikely" models, a good estimate of the most relevant probabilities can typically be obtained with a moderate number of iterations. The MC^3 Bayesian stochastic search algorithm is used in the SSVSquantreg model. More detailed information about the MC^3 algorithm can be found in Fernandez et al (2001).

The BMA provides posterior inclusion probabilities of candidate regressors which are quantities of practical interest for variable selection. The sum of the posterior model probabilities of all models containing the candidate regressor j ($\Pr(\beta_j \neq 0 | \gamma)$) is defined as posterior inclusion probabilities. Barbieri and Berger (2004) proposed to retain variables with posterior inclusion probability greater or equal to 1/2. They showed that their model often outperforms the highest posterior probability model. Hence, the median probability model is used in this study for variable selection.

5.2.3 Downscaling precipitation occurrence

Precipitation is an intermittent process with a significant number of dry days. To downscale precipitation from GCMs, we follow a widely-used two-step procedure. In the first step, the occurrence of precipitation is simulated. If a day is determined to be wet, a second step is used to simulate the precipitation amount. The focus of our study is on precipitation amounts, but we will here briefly discuss the issue of downscaling precipitation occurrence.

Following Wilby et al. (1999), we simulate the precipitation occurrence using a linear regression model of the form

$$E[O_i] = \alpha_0 + \alpha_{i-1} O_{i-1} + \alpha_{1i} P_{1i} + ... + \alpha_{ni} P_{ni} \tag{5.7}$$

where O_i is the precipitation occurrence on day *i*, coded as 0 if the day is dry and 1 if it is wet, and *P* are predictors chosen using stepwise regression model. The above model also includes O_{i-1}, the occurrence variable on the previous day, as predictor variable. This improves the ability of the model to simulate realistic sequences of wet and dry spells and embeds some aspects of a Markov chain structure in the model. By using 0 and 1 for the coding of dry and wet days, it is readily seen that $E[O_i]$ represents the probability of day *i* being wet. The downscaled occurrence is therefore random and can be simulated using a random number generator.

It should be noted that although we interpret $E[O_i]$ as a probability, nothing technically prevents this quantity from falling outside the range of 0 to 1 for a given combination of predictor variables. However, this does not preclude us from proceeding with the downscaling. A negative value simply implies a definite dry day while a value greater than 1 implies a definite wet day.

5.2.4 Statistical downscaling of daily precipitation amounts using quantile regression

There are two key issues that we wish to address using quantile regression. First, the best description of the conditional distribution of daily precipitation may require the use of different subsets of predictor variables for different quantiles; and second, the conditional distribution may take different forms (in addition to the mean level) for different predictor conditions. Let $\mathbf{x} = (x_1, x_2, \ldots, x_k)$ denote the set of potential daily predictor variables. Our goal is to derive, for each wet day, the most accurate precipitation distribution, conditioned on the set of predictor variables for the day. This conditional distribution can then be used to determine the expected value of precipitation or to sample a random value to represent the downscaled precipitation for the location. Random sampling from the conditional distribution is common practice, used to ensure that downscaled precipitation time series have realistic variability. However, compared to more conventional approaches to downscaling, we do not limit the error distribution to be Gaussian. To obtain a detailed representation of the conditional distribution, we perform quantile regression for a range of quantiles, for example corresponding to non-exceedance probabilities 0.01 to 0.99 in steps of 0.01. A regression model is developed for each of these quantiles and the different models will typically involve different subsets of predictor variables. To further illustrate this idea, Figure 5.1 shows the results obtained for a particular day, using data described later in the paper. Each probability level (0.01, 0.02,…,0.99) produces a quantile estimate from the corresponding regression model and these values are shown as blue circles in Figure 5.1. Clearly, the estimated points on the cumulative distribution function (CDF) curve do not constitute a monotone sequence as required of a CDF, but this is merely a result of sampling variability and model approximation. The general shape of the CDF is clearly discerned and can be approximated by one of the common probability distributions.

Figure 5.1 shows the fit of a Gamma distribution to the quantile points (dotted line). Also shown is the conditional distribution based on the traditional regression model. For this regression model, precipitation was first transformed by taking the fourth root and the regression model developed for the transformed precipitation, assuming a normal distribution of residuals. The curve in Figure 5.1 corresponds to the back-transformed quantiles (solid line). There are some differences in the two curves, suggesting that one may get different downscaling results depending on which model is used.

5.3 Data and analysis

Daily precipitation data from five weather stations in the central part of Canada were used in our study. The locations of the stations are shown in Figure 5.2 and some additional details are given in Table 5.2. For the application of quantile regression, we considered two seasons, summer and winter, defined respectively as June-July-August and December-January-February. Predictor variables were obtained from the NCEP-NCAR Global Reanalysis 1 (GR1). The data used here covers the period 1961 to 1990. The 30-year record was split into two sub-samples: 1961-1980 was used for model calibration, and 1981-1990 for model validation. A total of 15 candidate predictor variables were selected and extracted for the grids covering each station. The selected GR1 predictors are listed in Table 5.3.

5.4 Results

5.4.1 Selection of predictor variables

The variable selection method described in Section 5.2.2 was used to identify the best sub-set of predictor variables for each quantile level (0.01, 0.02,…,0.99). The computations were done using the R-package SSVSquantreg (http://mcmcpack.wustl.edu/). As examples of results, Tables 5.4 and 5.5 report the posterior inclusion probabilities resulting from the Bayesian analysis of the quantile regression at each decile, as well as the 98th percentile for Churchill summer and winter precipitation. The variables with posterior inclusion probability greater than 0.5 are retained and highlighted. The tables show that some of the variables are important across a range of quantiles (e.g. RHUM500 for Churchill summer precipitation) and others are important only for certain parts of the conditional distribution (e.g. VWIND500 which is significant only for Churchill summer high precipitation). This finding suggests that downscaling precipitation with a fixed set of predictors for the entire conditional distribution of precipitation, as done in traditional statistical downscaling models, may not make optimal use of available information.

A comparison of Tables 5.6 and 5.7 shows that the prediction of summer precipitation is more complex than the prediction of winter precipitation in the sense that more variables are required for summer precipitation. This is particularly apparent in the extreme right tail (98^{th} quantile) of the precipitation distributions where for example 13 predictors for Winnipeg are retained to explain summer precipitation compared with 4 predictors for winter precipitation. Summer precipitation is predominantly convective and more complex in nature whereas most winter precipitation is the result of synoptic-scale systems. A comparison of selected predictors for the 50^{th} and 98^{th} percentiles for Winnipeg summer precipitation shows that upper air temperature is an important factor for extreme precipitation. Rainfall from thunderstorms is a major contributor to summer precipitation in Winnipeg. Thunderstorms need unstable air, characterized by a temperature profile with warm air near the ground and cold air aloft. As seen in Figure 5.3, the relationships between precipitation and AIRTEMP500 for summer and winter are similar in nature, but the slopes are much higher for summer compared to winter for Winnipeg extreme precipitation (b_1 – values of 2.60 and 0.65 for summer and winter, respectively, in the 98^{th} quantile regression). For non-extreme precipitation events, the upper air temperature is less important - an observation that is in agreement with the conclusions of Harnack et al. (1999) and Hellestrom (2005).

Another illustration of the value of quantile regression for analyzing the predictor-predictand relationships can be seen in Figure 5.4. Here, the relationships of precipitation amounts and 500 mb relative humidity in Sioux Lookout and Cold Lake are compared.

The reason why there is a stronger relationship between extreme precipitation (98th quantile regression) and RHUM500 in Sioux Lookout than in Cold Lake is likely because Sioux Lookout is the recipient of southerly flows of warm moist air from the Central United States and the Gulf of Mexico, while in Cold Lake, the Rocky Mountains has a pronounced effect on the climate, and the air masses that do cross the Rockies lose much of their moisture (Vickers et al., 2000).

The standard regression predictors were selected using stepwise regression as implemented in the SDSM model (Wilby et al., 2002) and the ASD model (Hessami et al., 2008).

5.4.2 Downscaling precipitation occurrence

The first step in the downscaling process is to simulate the days on which precipitation occurs. The predictors required in Equation (5.7) were chosen using stepwise regression. Generally, due to the large number of data points, stepwise regression tends to include variables that are only very vaguely related to the predictand and it may be appropriate to use the R-square value or the error variance to terminate the variable selection.

Equation (5.7) provides unbiased estimates of the frequency of wet and dry days, but in general the prediction skill is not very good. This is a well-known problem in statistical downscaling. The quality of the downscaling of precipitation occurrence may be measured by one of several skill scores proposed for 2-class categorical forecasts. Table 5.8 shows the Heidke Skill Score (HSS) for the five sites and two seasons. The HSS is defined as

$$HSS = \frac{P_C - P_E}{1 - P_E} \tag{5.8}$$

where P_C is the observed probability of correct forecasts, i.e. the frequency of cases where the downscaled occurrence matches the observed occurrence, and P_E is a reference value, equal to the probability of a correct guess if the downscaled values are completely random. A perfect downscaling model will have a HSS value of 1 and a completely random model will have a value of 0. Table 5.8 shows that the HSS for the various sites and seasons are in the range of 0.15 to 0.25 which is rather low but still considerably better than random sampling.

It may be of interest to examine more complex properties of the downscaled occurrence sequences. Figure 5.5 shows observed and simulated wet and dry spell lengths for the Winnipeg location. These are properties of practical interest in hydrology. The plots indicate that the downscaling models tend to shift the probability distributions of wet and dry spell lengths towards shorter durations for both summer and winter seasons. The results are similar to those obtained by Wilby et al. (2002) using SDSM.

5.4.3 Downscaling of daily precipitation amounts

For each of the five sites and two seasons, daily precipitation on wet days was generated using the procedure described in Section 5.2.4. More specifically, on any given day, quantile regression was used to produce quantiles of the conditional distribution corresponding to different p-values. Initially, the Gamma, the Generalized Extreme Value (GEV), the Generalized Pareto, and the Lognormal distributions were considered candidates for representing the quantiles. Several options can be considered for fitting the distributions. If one considers, as we did, p-values from 0.01 to 0.99 in steps of 0.01, then it is not unreasonable to view the quantiles as a sample of pseudo-random variables from the conditional distribution, and conventional methods such as maximum likelihood or the method of moments can be used to estimate the distribution parameters. We used the method of maximum likelihood. If one calculates only a few quantiles that are not equally spaced in probability space, then it may be more appropriate to use some curve-fitting technique.

Although in principle one could conduct a goodness-of-fit test to determine which distribution type best fits the set of quantiles on a given day, we simplified the procedure by using a single distribution type for every day and every site. In a set of preliminary runs, the four candidate distributions were therefore applied consistently to determine which distribution overall provides the best result. Figure 5.6 shows the downscaling results for the summer season in Winnipeg in the form of quantile-quantile (Q-Q) plots. The Gamma distribution provides excellent results while the other three distributions produce some unrealistically large precipitation values. For example, the GEV distribution is clearly not suitable. This fact can be understood when recalling that no transformation of precipitation was done, so the conditional distribution tends to be highly skewed with a lower bound at zero. The GEV simply cannot accommodate this shape whereas the Gamma, which has a lower bound at zero, can. The log-normal and Generalized Pareto distributions do better than the GEV, but still produces some unreasonable values of large precipitation. Based on the above evaluation, the Gamma distribution was retained for the further use.

To assess the performance of the proposed QR model, we compare it to traditional downscaling based on multiple regression. One could compare the two models using scatter plots of observed versus downscaled values, but because there is a significant random component in both models, it can be difficult to appreciate differences in model performance using scatter plots. Therefore model performance is evaluated on the basis of Q-Q-plots which better reveals differences, especially in the right tail of precipitation distributions. The Q-Q plots of quantile regression and standard regression for the five stations in summer and winter are compared in Figures 5.7 and 5.8, respectively, for the calibration period.

For the summer season, there appears to be an advantage of quantile regression for some of the sites. In particular, the precipitation distributions at Winnipeg and Churchill appear to be better simulated by the QR downscaling model than by the traditional model. This is generally true not only for extreme right tail of the distribution but also for the central part of the distribution. For winter precipitation, there appears to be no significant difference between the two downscaling approaches (Figure 5.8). This may be because winter precipitation tend to be less extreme and easier to model with a conventional regression model. The conditional distribution of summer precipitation on the other hand may be heavily skewed and heavy-tailed. For example, for the Cold Lake station, the maximum value of winter precipitation is 13.5 mm and the coefficient of skewness is 2.4, while the maximum is 93.7 mm and the skewness is 3.8 for summer precipitation amounts.

The results for the validation period follow the same trend as calibration period (Figure 5.9 and Figure 5.10). The superiority of the QR statistical downscaling model for summer and similar results of quantile and standard regression models for winter (Figure 5.10) can be seen in the Q-Q plots comparisons.

To further compare the two modeling approaches, the mean value, standard deviation, and three STARDEX indices of extreme precipitation are listed in Table 5.9. The STARDEX indices (Goodess, 2003) are the 90th percentile of rain day amounts (pq90), the percentage of total rainfall from events greater than the long-term 90th percentile (pfl90), and the percentage of the of the number of days with more than 20 mm for summer (psh20) and more than 5 mm for winter (psh5). The results in Table 5.9 corroborate the conclusions from the Q-Q plots that the QR model performs better than the standard model in summer for most of the cases, especially in terms of standard deviation, and the pq90, pfl90, and psh20 indices. For winter precipitation, no particular method stands out as superior, although quantile regression outperforms the standard regression at more stations and for more indices.

The bar plot in Figure 5.11 illustrates the advantage of the QR downscaling model in summer for simulating the percentage of days with more than 20 mm of precipitation. There is no significant difference between standard and quantile regression performance for simulating the percentage of days with more than 5 mm in winter.

Comparison of the inverse cumulative distribution function of the two model simulations at Prince Albert for summer (Figure 5.12 left) and for winter (Figure 5.12 right) further highlights the advantage of the OR downscaling model for summer, especially in the extremes.

5.4 Discussion and conclusions

Bremnes (2004) argued that a flexible predictor selection model would possibly improve forecasts from a precipitation downscaling model. In this study, we have taken advantage of some recent developments in quantile regression and have employed a flexible predictor selection model called SSVSquantreg developed by Reed et al. (2009). Using a Bayesian framework, this model provides a basis for selecting different predictors for different quantiles, resulting in what we believe is a more flexible and more accurate description of the conditional distribution. One of the issues of regression-based statistical downscaling models is the potential for over-fitting (Wetterhall et al., 2007; Liu et al., 2011). Traditional models like SDSM are susceptible to this problem, and typically one should not use more than say 5 or 6 predictors to avoid over-fitting. The developed QR statistical downscaling model in this study has the advantage that different numbers of predictors can be used for different part of the conditional distribution.

In addition to its general downscaling application, quantile regression can provide valuable information for analyzing the relationships between extreme precipitation and atmospheric variables, for example by highlighting the different processes that govern summer and winter precipitation.

In our comparison study, the proposed QR model showed better performance for summer precipitation than the standard regression model. However, the two models produced fairly similar results for winter precipitation. It is not surprising that there is more difference in the summer where precipitation processes are far more complex and the distribution of daily precipitation is highly skewed and heavy-tailed.

As noted by Timofeev and Sterin (2010), quantile regression provides a more detailed understanding of processes in the climate system, such as surface and tropospheric warming, stratospheric cooling, changes in climate variability and extreme characteristics. Baur et al. (2004) argued that quantile regression has the advantage of easy implementation and transparency of results, which is in contrast to black box models such as neural networks where the functional relationship between predictor variables and the response distribution is less clear. The model developed in this study has this advantage that it requires no strong assumptions and it can be easily applied to other variables such as temperature and wind speed.

Acknowledgments

The financial support of the Natural Sciences and Engineering Council of Canada and Manitoba Hydro is gratefully acknowledged. Constructive comments by Dr. Jay Anderson and Dr. Tim Papakyriakou on an early draft are greatly appreciated.

References

Alhamzawi, R., and Yu, K. 2012. Variable selection in quantile regression via Gibbs sampling. Journal of Applied Statistics, 39 (4), 799-813.

Barbieri, M.M., and Berger, J.O. 2004. Optimal predictive model selection. Annals of Statistics, 32 (3), 870-897.

Barbosa, S.M. 2008. Quantile trends in Baltic Sea level. Geophysical Research Letters, 35 (22), art. No. L22704.

Baur, D., Saisana, M., and Schulze, N. 2004. Modeling the effects of meteorological variables on ozone concentration - A quantile regression approach. Atmospheric Environment, 38 (28), 4689-4699.

Breiman, L. 1996. Heuristics of instability and stabilization in model selection. Annals of Statistics, 24 (6), 2350-2383.

Bremnes, J.B. 2004. Probabilistic forecasts of precipitation in terms of quantiles using NWP model output. Monthly Weather Review, 132 (1), 338-347.

Buishand, T.A., Shabalova, M.V., and Brandsma, T. 2004. On the choice of the temporal aggregation level for statistical downscaling of precipitation. Journal of Climate, 17 (9), 1816-1827.

Cade, B.S., and Noon, B.R. 2003. A gentle introduction to quantile regression for ecologists. Frontiers in Ecology and the Environment, 1 (8), 412-420.

Cannon, A.J. 2011. Quantile regression neural networks: Implementation in R and application to precipitation downscaling. Computers and Geosciences, 37 (9), pp. 1277-1284.

Chen, S.-T., Tseng, H.-W., Lin, C.-Y., Yang, T.-C., and Yu, P.-S. 2011. Hydrological drought in Tseng-Wen reservoir basin under climate change scenarios. Journal of Taiwan Agricultural Engineering, 57 (3), 44-60.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., Ishwaran, H., Knight, K., Loubes, J.-M., Massart, P., Madigan, D., Ridgeway, G., Rosset, S., Zhu, J.I., Stine, R.A., Turlach, B.A., Weisberg, S., Hastie, T., Johnstone, I., and Tibshirani, R. 2004. Least angle regression. Annals of Statistics, 32 (2), 407-499.

Elsner, J.B., Kossin, J.P., and Jagger, T.H. 2008. The increasing intensity of the strongest tropical cyclones. Nature, 455 (7209), 92-95.

Fan, J., and Li, R. 2001. Variable selection via non-concave penalized likelihood and its oracle properties. Journal of the American Statistical Association, 96 (456), 1348-1360.

Fernández, C., Ley, E., and Steel, M.F.J. 2001. Model uncertainty in cross-country growth regressions. Journal of Applied Econometrics, 16 (5), 563-576.

Fowler, H.J., Blenkinsop, S., and Tebaldi, C. 2007. Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling. International Journal of Climatology, 27 (12), 1547-1578.

Friederichs, P. 2010. Statistical downscaling of extreme precipitation events using extreme value theory. Extremes, 13 (2), 109-132.

Friederichs, P., and Hense, A. 2007. Statistical downscaling of extreme precipitation events using censored quantile regression. Monthly Weather Review, 135 (6), 2365-2378.

George, E.I., and McCulloch, R.E. 1993. Variable selection via Gibbs sampling. Journal of the American Statistical Association, 88, 881-889.

Goodess, C.M. 2003. Statistical and Regional dynamical Downscaling of Extremes for European regions: STARDEX. Eggs, 6, 25-29.

Hao, L., and Naiman, D.Q. 2007. Quantile Regression. SAGE: Thousand Oaks, California.

Harnack, R.P., Apffel, K., and Cermak III, J.R. 1999. Heavy precipitation events in New Jersey: Attendant upper-air conditions. Weather and Forecasting, 14 (6 PART 2), 933-954.

Harrell, F. E. 2001. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. Springer, New York.

Hastie, T., Tibshirani, R., and Friedman, J. 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York.

Hellström, C. 2005. Atmospheric conditions during extreme and non-extreme precipitation events in Sweden. International Journal of Climatology, 25 (5), 631-648.

Hessami, M., Gachon, P., Ouarda, T.B.M.J., and St-Hilaire, A. 2008. Automated regression-based statistical downscaling tool. Environmental Modelling and Software, 23 (6), 813-834.

Jeong, D.I., St-Hilaire, A., Ouarda, T.B.M.J., and Gachon, P. 2012. Comparison of transfer functions in statistical downscaling models for daily temperature and precipitation over Canada. Stochastic Environmental Research and Risk Assessment, 26(5), 633-653.

Koenker, R. 2005. Quantile Regression. Cambridge University Press, New York.

Koenker, R., and Bassett, G. 1978. Regression quantiles. Econometrica 46 (1), 33-50.

Koenker, R., and Schorfheide, F. 1994. Quantile spline models for global temperature change. Climatic Change, 28 (4), 395-404.

Li, Q., Xiy, R., and Lin, N. 2010. Bayesian regularized quantile regression. Bayesian Analysis, 5 (3), 533-556.

Li, Y., and Smith, I. 2009. A statistical downscaling model for Southern Australia winter rainfall. Journal of Climate, 22 (5), 1142-1158.

Liu, Z., Xu, Z., Charles, S.P., Fu, G., and Liu, L. 2011. Evaluation of two statistical downscaling models for daily precipitation over an arid basin in China. International Journal of Climatology, 31 (13), 2006-2020.

Madigan, D., Raftery, A.E., Volinsky, C.T., and Hoeting, J.A. 1996. Bayesian model averaging. Integrating Multiple Learned Models (IMLM-96), 77-83.

Mazvimavi, D. 2010. Investigating changes over time of annual rainfall in Zimbabwe. Hydrology and Earth System Sciences, 14 (12), 2671-2679.

Palatella, L., Miglietta, M.M., Paradisi, P., and Lionello, P. 2010. Climate change assessment for Mediterranean agricultural areas by statistical downscaling. Natural Hazards and Earth System Science, 10 (7), 1647-1661.

Reed, C., Dunson, D., and Yu, K. 2009. Bayesian variable selection in quantile regression. Technical Report, Department of Mathematical Sciences, Brunel University.

Tareghian, R., and Rasmussen, P. 2012. Analysis of Arctic and Antarctic sea ice extent using quantile regression. International Journal of Climatology, DOI: 10.1002/joc.3491.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B, 58 (1), 267-288.

Timofeev, A.A., and Sterin, A.M. 2010. Using the quantile regression method to analyze changes in climate characteristics. Russian Meteorology and Hydrology, 35 (5), 310-319.

Tolika, K., Anagnostopoulou, C., Maheras, P., and Vafiadis, M. 2008. Simulation of future changes in extreme rainfall and temperature conditions over the Greek area: A comparison of two statistical downscaling approaches. Global and Planetary Change, 63 (2-3), 132-151.

Vickers, G., Buzza, S., Schmidt, D., and Mullock, J. 2000. The Weather of the Canadian Prairies. NAV CANADA report.

Villarini, G., Smith, J.A., Napolitano, F., and Baeck, M.L. 2011. Hydrometeorological analysis of the December 2008 flood in Rome. Hydrological Sciences Journal, 56 (7), 1150-1165.

Weerts, A.H., Winsemius, H.C., and Verkade, J.S. 2011. Estimation of predictive hydrological uncertainty using quantile regression: Examples from the National Flood Forecasting System (England and Wales). Hydrology and Earth System Sciences, 15 (1), 255-265.

Wetterhall, F., Halldin, S., and Xu, C.-Y. 2007. Seasonality properties of four statistical-downscaling methods in central Sweden. Theoretical and Applied Climatology, 87 (1-4), 123-137.

Wilby, R.L. 1998. Statistical downscaling of daily precipitation using daily airflow and seasonal teleconnection indices. Climate Research, 10 (3), 163-178.

Wilby, R.L., Dawson, C.W., and Barrow, E.M. 2002. SDSM - A decision support tool for the assessment of regional climate change impacts. Environmental Modelling and Software, 17 (2), 147-159.

Wilby, R.L., Hay, L.E., and Leavesley, G.H. 1999. A comparison of downscaled and raw GCM output: Implications for climate change scenarios in the San Juan River Basin, Colorado. Journal of Hydrology, 225 (1-2), 67-91.

Yu, K., and Moyeed, R.A. 2001. Bayesian quantile regression. Statistics and Probability Letters, 54 (4), pp. 437-447.

Table 5.1. Quantile regression applications in environmental modeling and climate change impact assessment.

Global temperature change over the last century	Koenker and Schorfheide (1994)
Effects of meteorological variables on Ozone concentration	Baur et al. (2004)
Tropical cyclones trend changes	Elsner et al. (2008)
Quantile trends in Baltic sea level	Barbosa (2008)
Changes over time of annual rainfall in Zimbabwe	Mazvimavi (2010)
Estimation of predictive hydrological uncertainty	Weerts et al. (2010)
Hydrometeorological analysis of the December 2008 flood in Rome	Villarini et al. (2011)
Analysis of Arctic and Antarctic sea ice extent	Tareghian and Rasmussen (2012)

Table 5.2. Station information.

Station	Province	Average Annual Precipitation (mm)	Average Annual Temperature(˚C)
Cold Lake	Alberta	427	1.7
Prince Albert	Saskatchewan	487	0.5
Winnipeg	Manitoba	605	1.8
Churchill	Manitoba	554	-7.2
Sioux Lookout	Ontario	716	1.5

Table 5.3. Selected NCEP predictor variables

Predictor variable	Abbreviation	Predictor variable	Abbreviation
500hPa Air temperature	AIRTEMP500	Specific humidity at 2-meter height	SQ
850hPa Air temperature	AIRTEMP850	E-component of wind at 10-meter height	SUWIND
500hPa Geopotential height	PHI500	V-component of wind at 10-meter height	SVWIND
850hPa Geopotential height	PHI850	500hPa East component of wind	UWIND500
Mean sea level pressure	PMSL	850hPa East component of wind	UWIND850
500hPa Relative Humidity	RHUM500	500hPa North component of wind	VWIND500
850hPa Relative Humidity	RHUM850	500hPa North component of wind	VWIND850
1000hPa Relative Humidity	RHUM1000		

Table 5.4. Posterior inclusion probabilities across quantiles for Churchill summer precipitation.

Variable	10^{th}	20^{th}	30^{th}	40^{th}	50^{th}	60^{th}	70^{th}	80^{th}	90^{th}	98^{th}
(Intercept)	0.97	0.44	0.39	0.36	0.34	0.49	0.55	0.54	0.61	0.57
PHI500	0.00	0.00	0.01	0.08	0.10	0.13	0.16	0.31	0.10	0.21
PHI850	0.00	0.00	0.00	0.61	0.94	0.96	0.99	1.00	0.99	0.96
PMSL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.01
RHUM500	0.00	0.59	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RHUM850	0.01	0.01	0.04	0.02	0.02	0.27	0.28	0.22	0.02	0.17
RHUM1000	0.01	0.00	0.01	0.01	0.01	0.03	0.06	0.24	0.95	1.00
SQ	0.12	0.13	0.19	0.28	0.34	0.46	0.53	0.54	0.52	0.53
UWIND500	0.00	0.00	0.00	0.01	0.01	0.04	0.05	0.08	0.38	0.19
UWIND850	0.00	0.00	0.01	0.01	0.02	0.78	1.00	0.93	0.40	0.34
VWIND500	0.00	0.04	0.48	0.34	0.11	0.99	1.00	1.00	1.00	1.00
VWIND850	0.01	0.14	0.39	0.68	0.97	0.08	0.63	0.49	0.15	0.34
SUWIND	0.00	0.02	0.01	0.03	0.08	0.97	1.00	1.00	0.99	0.61
SVWIND	0.00	0.02	0.03	0.02	0.04	0.14	0.25	0.36	0.18	0.48
AIRTEMP500	0.01	0.00	0.01	0.59	0.91	0.87	0.86	0.73	0.77	0.65
AIRTEMP850	0.01	0.01	0.01	0.07	0.08	0.14	0.14	0.30	0.27	0.45

Table 5.5. Posterior inclusion probabilities across quantiles for Churchill winter precipitation.

Variable	10th	20th	30th	40th	50th	60th	70th	80th	90th	98th
(Intercept)	0.99	0.94	0.65	0.12	0.07	0.19	0.35	0.75	0.64	0.31
PHI500	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.28	0.23
PHI850	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.68	0.67	0.68
PMSL	0.00	0.00	0.00	0.00	0.00	0.00	0.65	0.28	0.06	0.00
RHUM500	0.00	0.01	0.26	0.81	0.67	0.76	1.00	1.00	1.00	0.99
RHUM850	0.00	0.01	0.07	0.18	0.43	0.82	0.24	0.01	0.01	0.02
RHUM1000	0.01	0.04	0.03	0.01	0.00	0.00	0.00	0.01	0.01	0.02
SQ	0.12	0.12	0.12	0.13	0.13	0.21	0.24	0.29	0.29	0.30
UWIND500	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.03	0.10
UWIND850	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.08
VWIND500	0.00	0.00	0.00	0.00	0.01	0.02	0.01	0.01	0.01	0.26
VWIND850	0.00	0.00	0.00	0.01	0.04	0.07	0.01	0.01	0.05	0.08
SUWIND	0.00	0.00	0.01	0.01	0.03	0.05	0.02	0.03	0.03	0.10
SVWIND	0.00	0.00	0.00	0.01	0.01	0.01	0.03	0.03	0.08	0.21
AIRTEMP500	0.00	0.00	0.01	0.00	0.00	0.07	0.75	0.98	0.99	0.32
AIRTEMP850	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.03	0.04	0.66

Table 5.6. Selected predictor variables for standard regression (mean) and 25th, 50th, 75th, and 98th quantiles for the five stations – Summer season.

Station	Quantile	Selected Predictors
Churchill	Mean	PMSL,RHUM500,RHUM1000,VWIND500,SUWIND
	25th	RHUM500, VWIND850
	50th	PHI850,RHUM500,VWIND850,AIRTEMP500
	75th	PHI850,RHUM500,UWIND850,VWIND500,VWIND850,SUWIND,AIRTEMP500
	98th	PHI850,RHUM500,RHUM1000,VWIND500,AIRTEMP500
Cold Lake	Mean	PMSL,RHUM500,SQ,UWIND850
	25th	RHUM500, SUWIND
	50th	PMSL,RHUM500,UWIND850,AIRTEMP500
	75th	(Intercept),RHUM500,RHUM850,RHUM1000,SQ,VWIND500,SUWIND,SVWIND
	98th	(Intercept),PHI850,SQ,UWIND500,UWIND850,VWIND500,SUWIND,AIRTEMP500, AIRTEMP850
Prince Albert	Mean	PMSL,RHUM500,SQ,UWIND850,AIRTEMP850
	25th	(Intercept),UWIND850
	50th	PHI500,RHUM850,UWIND850,VWIND850,AIRTEMP850
	75th	(Intercept),PHI500,PMSL,RHUM500,RHUM1000,SQ,UWIND850,VWIND500,SUWIND,AIRTEMP500,AIRTEMP850
	98th	(Intercept),PHI850,SQ,UWIND850,VWIND500,SUWIND,SVWIND,AIRTEMP500
Sioux Lookout	Mean	RHUM500,SQ,VWIND500,SUWIND
	25th	RHUM500,VWIND850
	50th	(Intercept),RHUM500,SQ,VWIND500,SUWIND
	75th	(Intercept),PHI850,RHUM500,SQ,VWIND850,SUWIND,SVWIND,AIRTEMP500, AIRTEMP850
	98th	RHUM500,SQ,VWIND850,SUWIND

Winnipeg	Mean	RHUM500,RHUM1000,VWIND500,SUWIND,AIRTEMP500
	25th	RHUM1000,VWIND850
	50th	RHUM500,RHUM1000,VWIND500,SUWIND
	75th	(Intercept),RHUM500,RHUM850,SQ,UWIND500,UWIND850,VWIND500,SUWIND, SVWIND
	98th	(Intercept),PHI500,PHI850,RHUM500,RHUM1000,SQ,UWIND500,UWIND850, VWIND500,VWIND850,SUWIND,SVWIND,AIRTEMP500,AIRTEMP850

Table 5.7. Selected predictor variables for standard regression (mean) and 25th, 50th, 75th, and 98th quantiles for five stations – Winter season.

Station	Quantile	Selected Predictors
Churchill	Mean	PMSL,RHUM500,UWIND500,SVWIND,AIRTEMP500
	25th	(Intercept),SQ
	50th	RHUM500,SUWIND
	75th	RHUM500,SUWIND
	98th	PHI850,RHUM500,AIRTEMP500
Cold Lake	Mean	RHUM500,SUWIND,SVWIND,AIRTEMP500
	25th	(Intercept),SQ
	50th	RHUM500,SUWIND
	75th	RHUM500,SUWIND
	98th	RHUM500,SUWIND
Prince Albert	Mean	PHI500,PMSL,VWIND500,AIRTEMP500
	25th	(Intercept),RHUM500
	50th	RHUM500,VWIND850
	75th	(Intercept),RHUM500,VWIND500,AIRTEMP500
	98th	(Intercept),VWIND500,AIRTEMP500
Sioux Lookout	Mean	PHI850,SQ,VWIND850,SVWIND,AIRTEMP850
	25th	RHUM500,SUWIND
	50th	RHUM500,SUWIND
	75th	(Intercept),PHI850,RHUM500,SQ,VWIND850,SUWIND,AIRTEMP500
	98th	(Intercept),RHUM500,VWIND850,SVWIND,AIRTEMP500
Winnipeg	Mean	PHI500,RHUM500,UWIND500,VWIND500,SUWIND,SVWIND, AIRTEMP500
	25th	(Intercept),SUWIND
	50th	RHUM500,SUWIND

	75^{th}	(Intercept),VWIND850,SVWIND,AIRTEMP500
	98^{th}	(Intercept),VWIND500,SUWIND,SVWIND,AIRTEMP500

Table 5.8. Heidke skill score (Equation 5.8) for downscaling precipitation occurrence.

	Heidke Skill Score	
Station	Summer	Winter
Churchill	0.15	0.22
Cold Lake	0.24	0.20
Prince Albert	0.17	0.20
Sioux Lookout	0.23	0.23
Winnipeg	0.25	0.19

Table 5.9. Comparison of standard regression (SR) and quantile regression (QR) with observed (OBS) precipitation statistics. The STARDEX indices Pq90, Pfl90, and Psh20 are defined in the text.

Indices	Station	Summer			Winter		
		Obs	SR	QR	Obs	SR	QR
Mean	Churchill	4.26	4.18	4.65	1.41	1.32	1.23
	Cold Lake	5.82	5.14	6.63	1.85	1.79	1.65
	Prince Albert	5.75	5.46	6.03	1.54	1.50	1.50
	Sioux Lookout	6.98	6.49	6.82	2.35	2.16	2.60
	Winnipeg	6.95	6.48	6.84	1.85	1.69	1.84
Standard Deviation	Churchill	5.83	4.90	5.87	1.81	1.14	1.19
	Cold Lake	8.05	6.61	8.17	2.14	1.84	1.51
	Prince Albert	8.09	6.59	7.69	1.85	1.40	1.50
	Sioux Lookout	8.86	7.55	8.38	3.21	2.26	2.98
	Winnipeg	10.09	8.32	9.22	2.48	1.71	2.20
Pq90	Churchill	11.20	9.68	11.15	3.00	2.81	2.61
	Cold Lake	14.20	13.02	17.16	4.60	4.10	3.49
	Prince Albert	14.18	14.40	14.47	3.60	3.19	3.18
	Sioux Lookout	17.30	15.21	17.04	5.60	5.06	6.21
	Winnipeg	17.80	16.15	16.71	4.80	3.88	4.42
Pfl90	Churchill	42.51	36.86	39.64	39.11	29.50	31.79
	Cold Lake	42.05	41.26	39.69	35.14	34.21	30.54
	Prince Albert	44.72	38.90	39.82	36.06	31.03	33.00

	Sioux Lookout	40.19	36.32	39.48	41.26	34.32	36.60
	Winnipeg	44.56	39.84	41.64	41.22	33.62	38.63
Psh20 (S) Psh5 (W)	Churchill	2.59	1.90	2.60	4.15	1.50	1.60
	Cold Lake	5.84	4.60	6.90	7.73	5.60	3.40
	Prince Albert	6.65	4.20	5.30	4.67	2.80	3.30
	Sioux Lookout	8.06	5.00	7.50	12.10	10.40	15.10
	Winnipeg	8.47	6.70	6.90	9.14	5.70	8.00

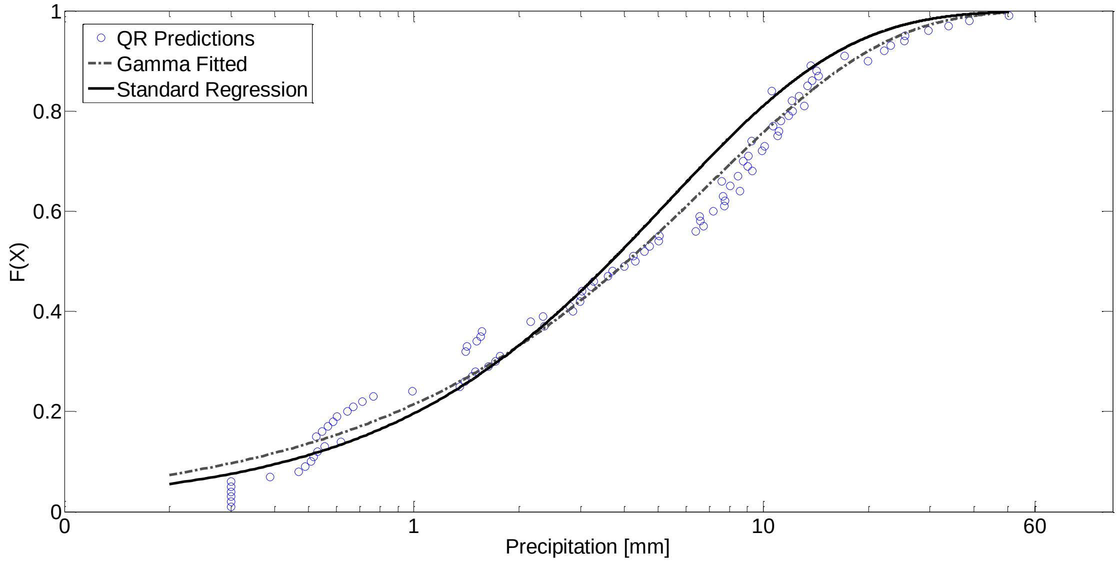

Figure 5.1. Quantile regression predictions along with a fitted Gamma distribution and standard regression predictions for Winnipeg summer precipitation.

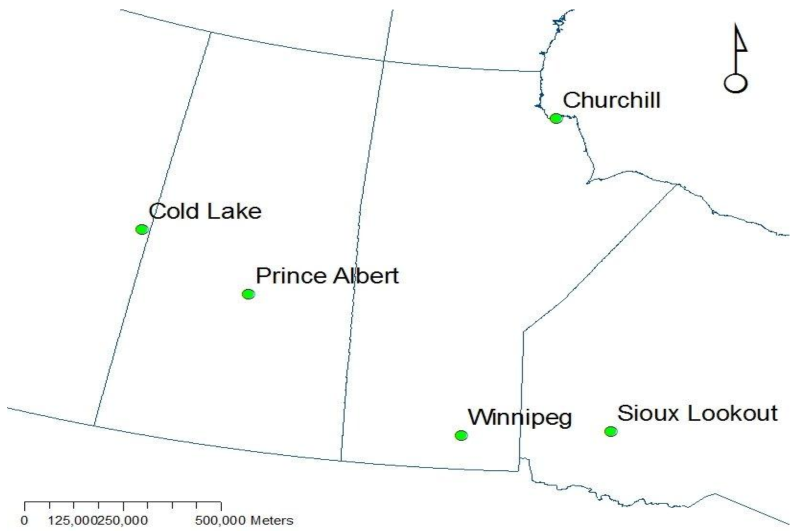

Figure 5.2. Location of stations used in the study.

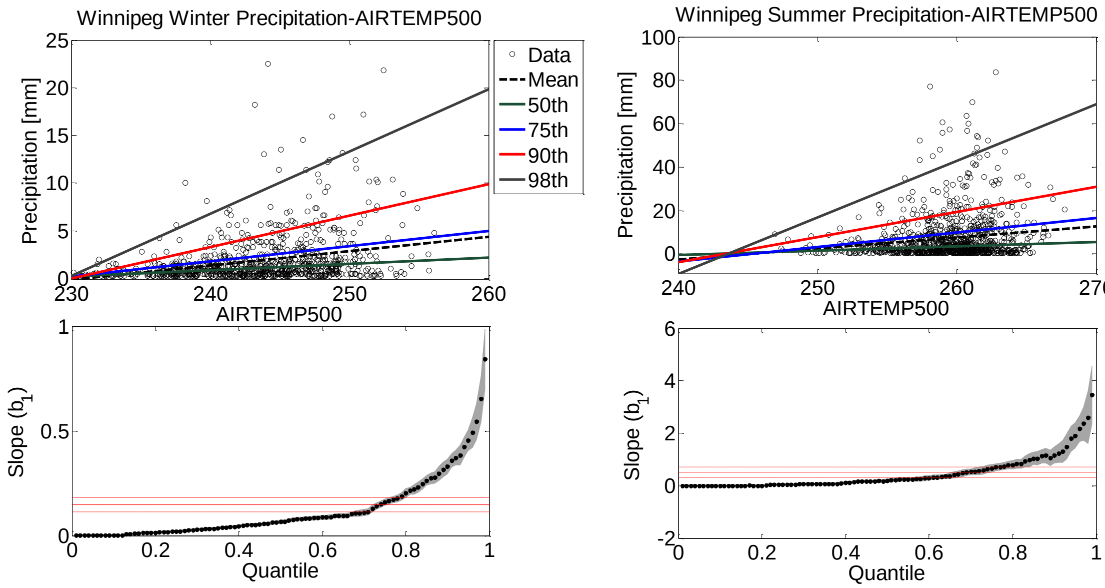

Figure 5.3. Relationship between daily precipitation in Winnipeg and AIRTEMP500 for winter (left) and summer (right). The bottom plots show the slopes of the estimated regression lines. The quantile regression coefficients (black dots) are presented with their 95% confidence bounds (shaded in grey). The least-squares regression coefficients (solid red line) are also given with their 95% confidence bounds (dashed red lines). The vertical axis shows the slopes (%), and the horizontal axis shows the p-value of the quantile (1-99th).

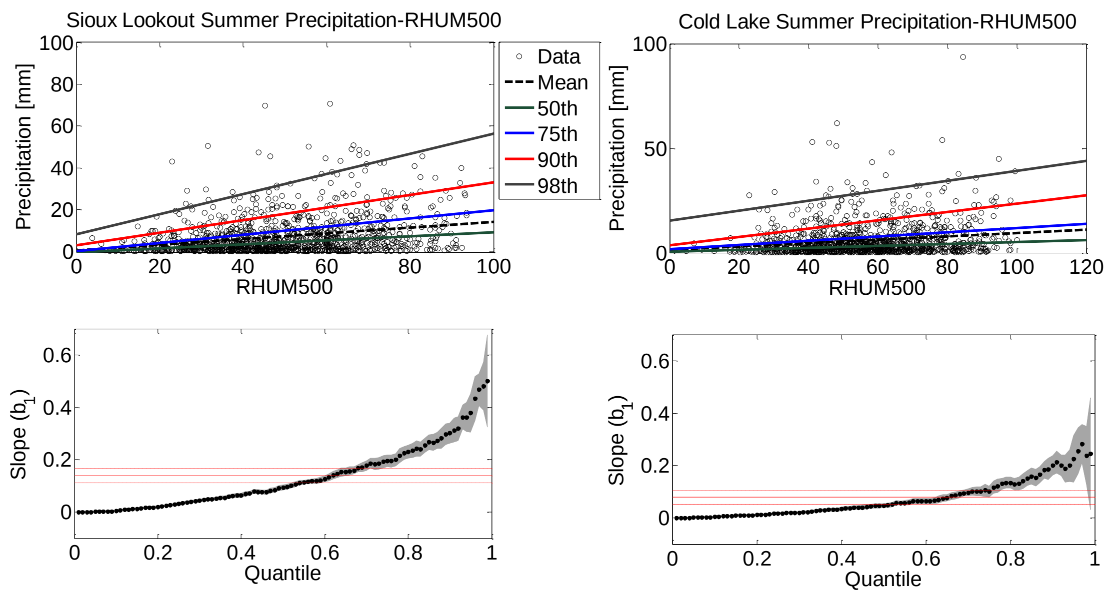

Figure 5.4. Relationship of RHUM500 and daily summer precipitation in (left) Sioux Lookout and (right) Cold Lake. The bottom plots are the same as in Figure 5.3.

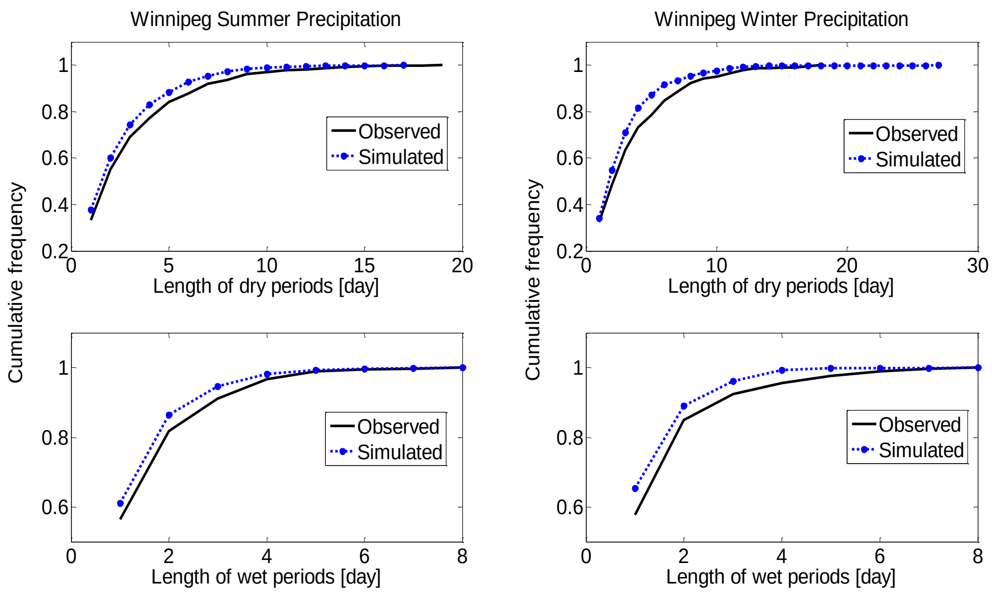

Figure 5.5. Frequency distribution of dry spell lengths (top) and wet spell lengths (bottom) of observed and simulated precipitation at the Winnipeg station for (left) summer and (right) winter seasons.

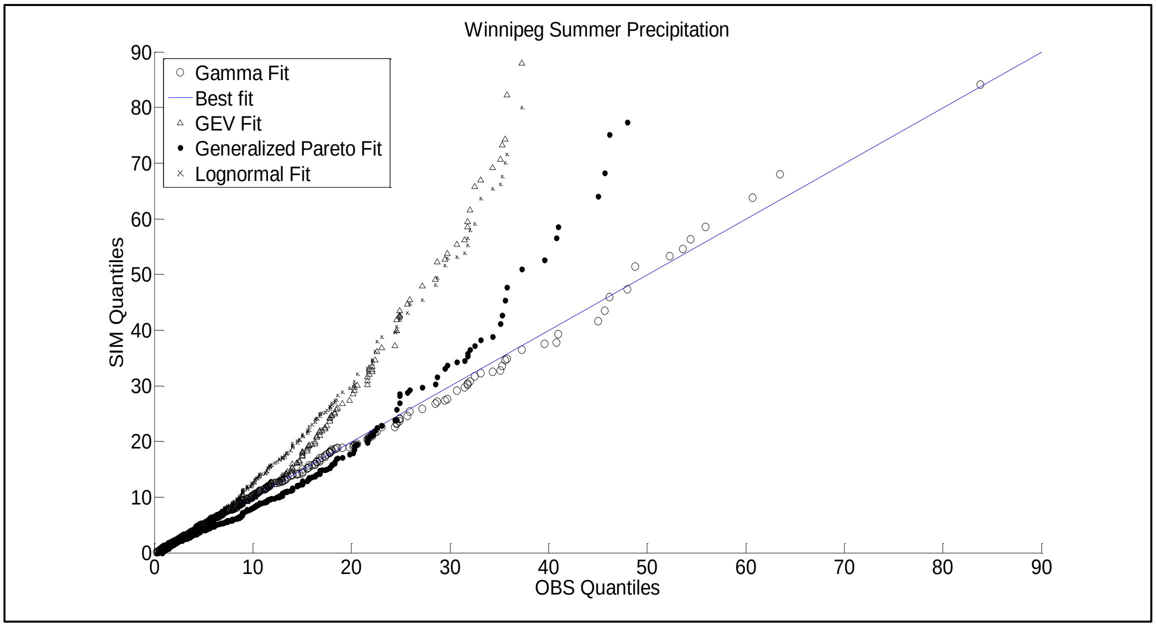

Figure 5.6. Q-Q plot comparison of different fitted distributions.

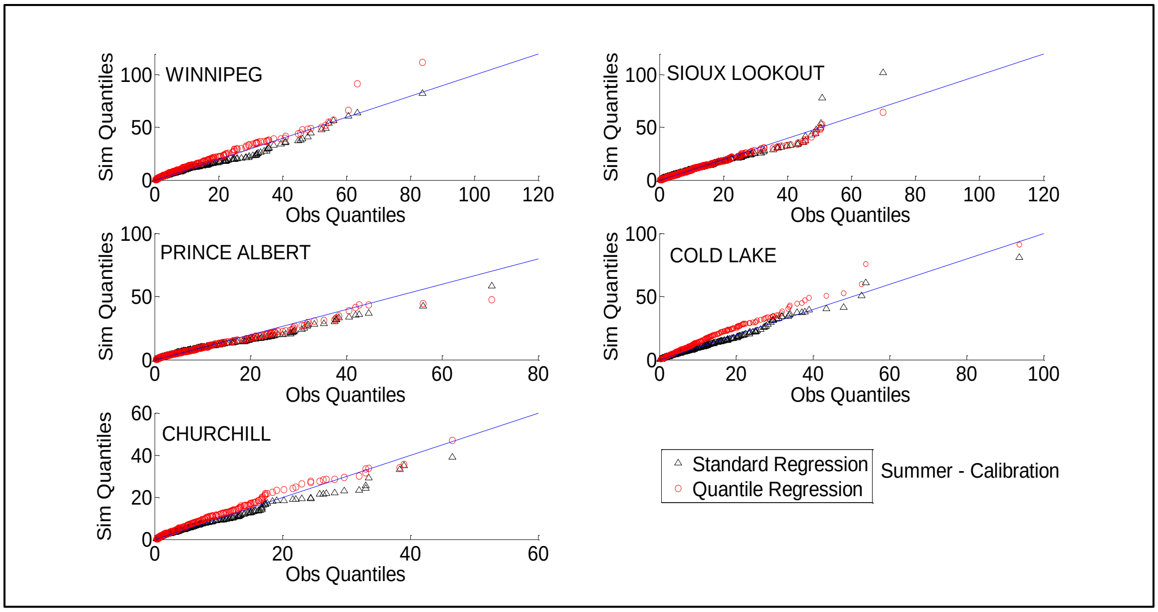

Figure 5.7. Q-Q plot comparison of simulated summer precipitation with the quantile regression and standard regression statistical downscaling models for the five stations in the calibration period.

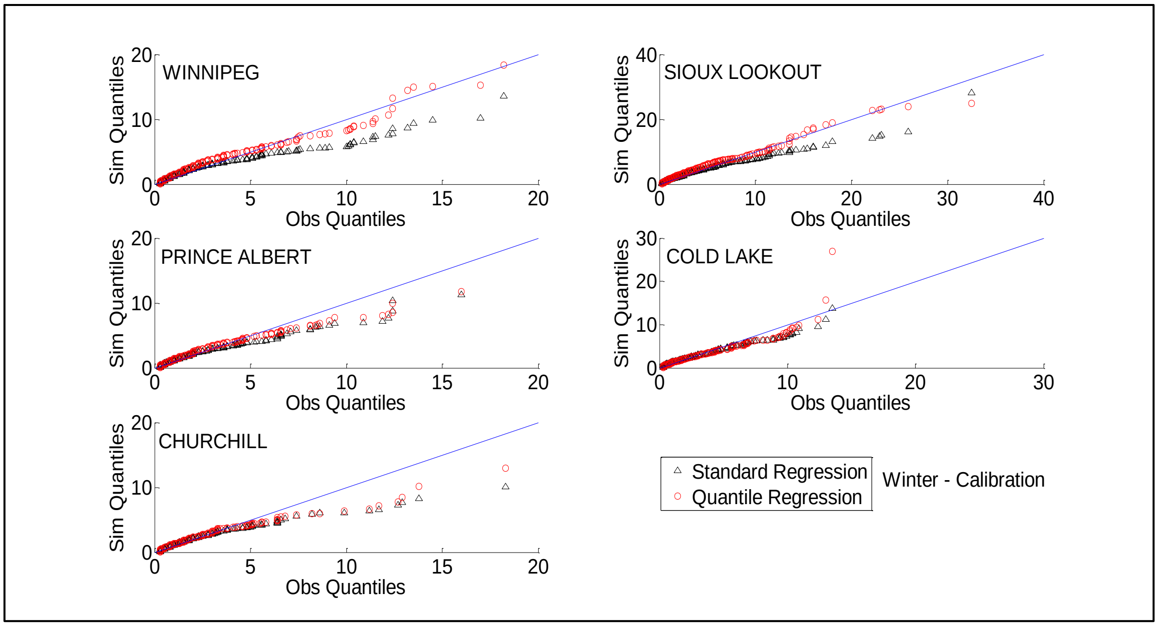

Figure 5.8. Q-Q plot comparison of simulated winter precipitation with the quantile regression and standard regression statistical downscaling models for the five stations in the calibration period.

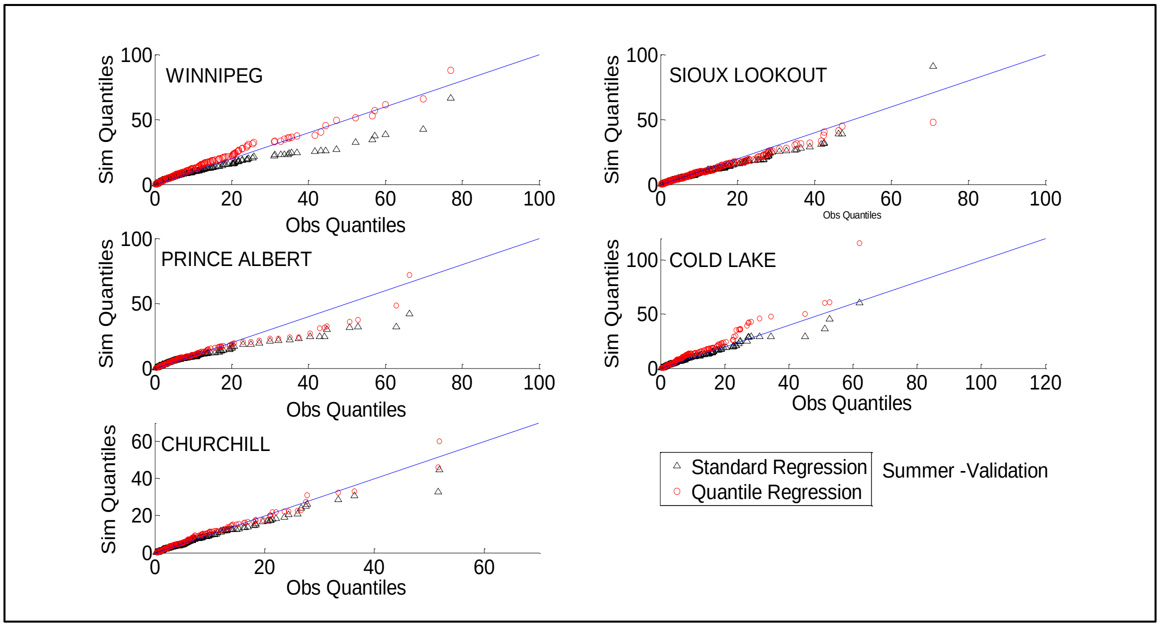

Figure 5.9. Q-Q plot comparison of simulated summer precipitation with the quantile regression and standard regression statistical downscaling models for the five stations in the validation period.

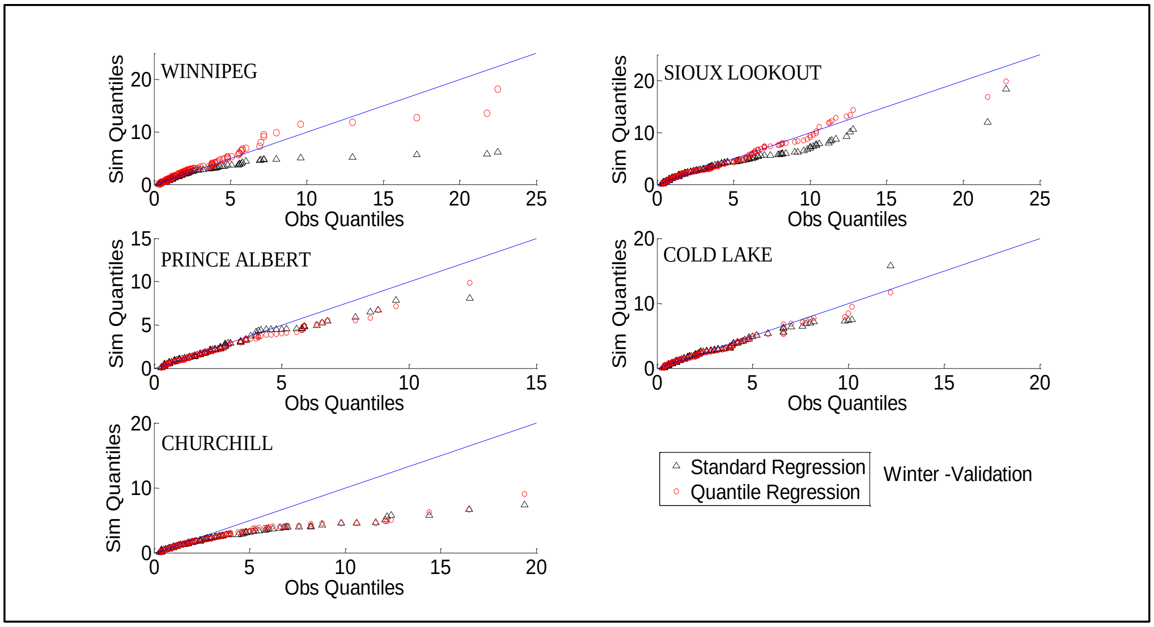

Figure 5.10. Q-Q plot comparison of simulated winter precipitation with the quantile regression and standard regression statistical downscaling models for the five stations in the validation period.

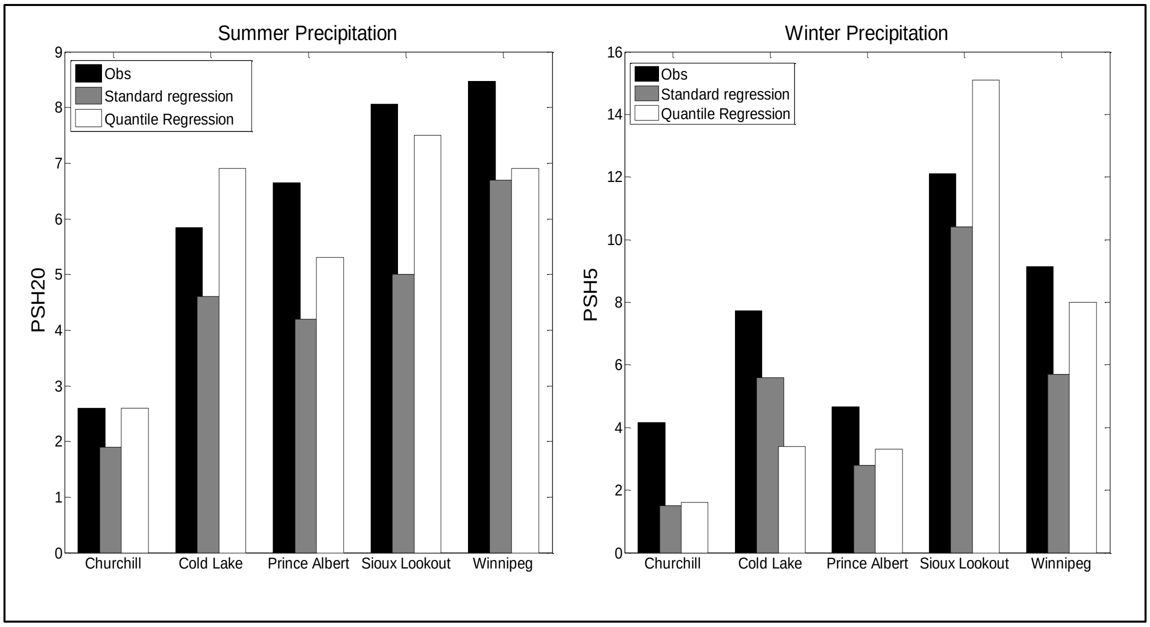

Figure 5.11. Comparison of quantile regression and standard regression statistical downscaling models for simulating the percentage of summer days with more than 20 mm precipitation (left) and the percentage of winter days with more than 5 mm precipitation (right).

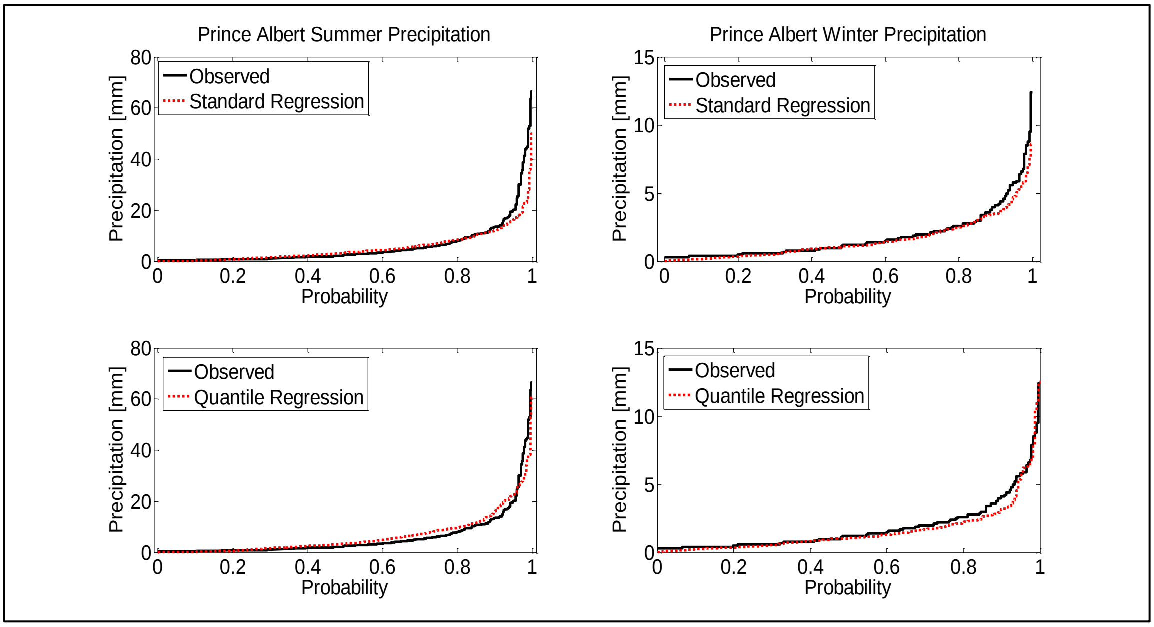

Figure 5.12. Inverse cumulative distribution function for the standard regression (top) and the quantile regression (bottom) statistical downscaling models for the Prince Albert. Left panels are summer precipitation; right panels are winter precipitation.

Chapter 6

Discussion and Conclusions

In this thesis, the advantages of quantile regression over OLS regression have been discussed and illustrated by different applications. In summary, quantile regression provides a more detailed description of the conditional distribution which enables researchers to analyze the tails and identify trends not only in mean values but also different quantiles of climate variables. Flexibility to allow covariates to have different relationships in different parts of the conditional distribution, the robustness to departures from normality, and the invariance under monotone transformation are some of the other properties that make quantile regression a useful statistical approach.

As discussed in Section 2.4, three application categories were defined for quantile regression: detection of trends, interpretation of non-linear relationships, and forecasting. The three papers presented in this thesis cover all three applications. The first two papers can be categorized in the trend detection class, and the third paper covers the interpretation of non-linear relationships and forecasting classes.

The Arctic and Antarctic have experienced different climatic changes in recent years. A number of recent studies have examined trends in sea ice cover using ordinary least squares regression (e.g., Comiso, 2010). Sea ice is important in the global climate system and there is a need for studies analyzing the trend of the whole distribution of sea ice extent. In our study, quantile regression was applied to analyze other aspects of the distribution of sea ice extent than the mean value. More specifically, trends in the mean, maximum, and minimum sea ice extent in the Arctic and Antarctic were investigated. We found a significant decreasing trend in mean Arctic sea ice extent of -4.5% per decade from 1979 through 2010, while the Antarctic results showed a small positive trend of 2.3% per decade. We showed that the changes are not simply in the mean of the key ice variables, but that their entire distribution shape may be impacted. The advantage of the quantile regression model was observed particularly in cases with asymmetric response distributions such as Antarctic minimum ice cover. We also found that the variability in Antarctic sea ice extent is higher than in the Arctic sea ice.

Observations in many areas have shown that changes in total precipitation are significantly influenced by changes in the tails of precipitation distributions. Also changes of daily temperatures are much more evident in the tails of the distributions (hot and cold days) than in the average temperatures. Changes in the frequency and intensity of extreme events may have a dramatic impact on society and the natural environment. In this study, we investigated changes in temperature and precipitation extremes over the Canadian Prairies using quantile regression. The results revealed that extreme temperatures have increased significantly over most of the Canadian Prairies. However, no general pattern can be detected in extreme precipitation. We also found that some of the extreme patterns might have gone undetected if a standard regression model had been used.

Statistical downscaling use relatively simple statistical models to represent the link between atmospheric circulation variables, presumably well simulated by the GCMs, and local weather variables, such as precipitation and temperature. As discussed by Wilby et al. (2002), the selection of appropriate downscaling predictor variables is one of the most challenging stages in the developing of any statistical downscaling models. The inflexibility of predictor selection methods for standard regression models is one of the concerning issues. In this study, we used a Bayesian method adapted to quantile regression to select predictor variables. The advantage of this method is that different predictors can be selected for different parts of the conditional distribution. We found that downscaling precipitation with a fixed set of predictors for the entire conditional distribution of precipitation, as done in traditional statistical downscaling models, may not make optimal use of available information. The results showed that more predictor variables were retained to explain summer precipitation than winter precipitation. This occurs because summer precipitation is predominantly convective and more complex in nature whereas most winter precipitation is a result of synoptic-scale systems. Using quantile regression for analyzing the relationships between extreme precipitation and atmospheric variables provided valuable information by highlighting the different processes that govern summer and winter precipitation.

In order to downscale precipitation, in the first step, the occurrence of precipitation was simulated using a linear regression model. To address the issue of variance underestimation and poor representation of extreme events by statistical downscaling models, a novel method, involving the fitting of different distributions to predicted regression quantiles, was introduced to downscale precipitation, and the performance of this new procedure, with special emphasis on extreme precipitation, was evaluated by comparison with traditional regression-based downscaling. Initially, the Gamma, the Generalized Extreme Value (GEV), the Generalized Pareto, and the Lognormal distributions were considered candidates for representing the response distribution. The Gamma distribution provided the best results compared with the other distribution candidates. Comparison of our proposed quantile regression model and the standard regression model demonstrated the superiority of quantile regression for modelling summer precipitation. However, the two models produced fairly similar results for winter precipitation. These results could be anticipated since summer precipitation processes are far more complex and the distribution of daily precipitation is highly skewed and heavy-tailed.

Quantile regression has the advantage of easy implementation and transparency of results, which is in contrast to black box models such as neural networks (Baur et al., 2004) where the functional relationship between predictor variables and the response distribution is less clear. As noted by Barbosa (2008), quantile regression is a useful technique for identifying distinct rates of change in geophysical time series and should be used for quantifying long-term variability in climatic and oceanographic variables.

References:

Andrews, D.W.K., and Buchinsky, M. 2000. A three-step method for choosing the number of bootstrap repetitions. Econometrica, 68 (1), pp. 23-51.

Auld, M.C., and Powell, L.M. 2009. Economics of food energy density and adolescent body weight. Economica, 76 (304), pp. 719-740.

Barbosa, S.M. 2008. Quantile trends in Baltic Sea level. Geophysical Research Letters 35 (22): art. No. L22704.

Barbosa, S.M., Scotto, M.G., and Alonso, A.M. 2011. Summarising changes in air temperature over Central Europe by quantile regression and clustering. Natural Hazards and Earth System Science, 11 (12), pp. 3227-3233.

Baur, D., Saisana, M., and Schulze, N. 2004. Modeling the effects of meteorological variables on ozone concentration - A quantile regression approach. Atmospheric Environment 38 (28): 4689-4699.

Bremnes, J.B. 2004. Probabilistic forecasts of precipitation in terms of quantiles using NWP model output. Monthly Weather Review, 132 (1), pp. 338-347.

Bremnes, J.B. 2006. A comparison of a few statistical models for making quantile wind power forecasts. Wind Energy, 9 (1-2), pp. 3-11.

Buchinsky, M. 1995. Estimating the asymptotic covariance matrix for quantile regression models a Monte Carlo study. Journal of Econometrics, 68 (2), pp. 303-338.

Buchinsky, M. 1998. Recent advances in quantile regression models: A practical guideline for empirical research. Journal of Human Resources, 33 (1), pp. 88-126.

Cade, B.S., Terrell, J.W., and Porath, M.T. 2008. Estimating fish body condition with quantile regression. North American Journal of Fisheries Management, 28 (2), pp. 349-359.

Chamaillé-Jammes, S., Fritz, H., and Murindagomo, F. 2007. Detecting climate changes of concern in highly variable environments: Quantile regressions reveal that droughts worsen in Hwange National Park, Zimbabwe. Journal of Arid Environments, 71 (3), pp. 321-326.

Chernozhukov, V., Hong, H., and Tamer, E. 2007. Estimation and confidence regions for parameter sets in econometric models. Econometrica, 75 (5), pp. 1243-1284.

Comiso, JC. 2010. Variability and Trends of the Global Sea Ice Cover. In: Sea ice, second edition [Thomas, DN, and Dieckmann, GS.]. Wiley-Blackwell, Oxford, UK: 205-246.

De Angelis, D., Hall, P., and Young, G.A. 1993. Analytical and bootstrap approximations to estimator distributions in L regression1. Journal of the American Statistical Association, 88, pp. 1310-1316.

Donner, R.V., Ehrcke, R., Barbosa, S.M., Wagner, J., Donges, J.F., and Kurths, J. 2012. Spatial patterns of linear and nonparametric long-term trends in Baltic sea-level variability. Nonlinear Processes in Geophysics, 19 (1), pp. 95-111.

Easterling, D.R., Meehl, G.A., Parmesan, C., Changnon, S.A., Karl, T.R., and Mearns, L.O. 2000. Climate extremes: Observations, modeling, and impacts. Science, 289 (5487), pp. 2068-2074.

Efron, B. 1979. Bootstrap methods: Another look at the jackknife. Annals of Statistics, 7, pp. 1-26.

Efron, B. 1982. The jackknife, the bootstrap and other resampling plans. Society for Industrial and Applied Mathematics, 38.

Elsner, J.B., Kossin, J.P., and Jagger, T.H. 2008. The increasing intensity of the strongest tropical cyclones. Nature 455 (7209): 92-95.

Fleeger, J.W., Johnson, D.S., Carman, K.R., Weisenhorn, P.B., Gabriele, A., Thistle, D., and Barry, J.P. 2010. The response of nematodes to deep-sea CO2 sequestration: A quantile regression approach. Deep-Sea Research Part I: Oceanographic Research Papers, 57 (5), pp. 696-707.

Friederichs, P. and T. Thorarinsdottir. 2012. Forecast verification scores for extreme value distributions with an application to peak wind prediction. Environmetrics (submitted, arXiv)

Friederichs, P., and Hense, A. 2007. Statistical downscaling of extreme precipitation events using censored quantile regression. Monthly Weather Review, 135 (6), pp. 2365-2378.

Gutenbrunner, C., and Jureckova, J. 1992. Regression rank scores and regression quantiles. Ann. Statist., 20 (1), pp. 305-330.

Gutenbrunner, C., Jureckova, J., Koenker, R., and Portnoy, S. 1993. Tests of linear hypotheses based on regression rank scores. Journal of Nonparametric Statistics, 2, pp. 307-331.

Hájek, J., and Šidák, Z. 1967. Theory of Rank Tests. Prague: Academia.

Hao L., and Naiman D.Q. 2007. Quantile regression. SAGE: Thousand Oaks, California, USA.

Horowitz, J.L. 1998. Bootstrap methods for median regression models. Econometrica, 66 (6), pp. 1327-1351.

Hušková, M. 1994. Some sequential procedures based on regression rank scores. Nonparametric Statistics, 3, pp. 285-298.

IPCC. 2007. Climate change 2007: The physical science basis, contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press: Cambridge, New York, USA.

Jagger, T.H., and Elsner, J.B. 2009. Modeling tropical cyclone intensity with quantile regression. International Journal of Climatology, 29 (10), pp. 1351-1361.

Kocherginsky, M., He, X., and Mu, Y. 2005. Practical confidence intervals for regression quantiles. Journal of Computational and Graphical Statistics, 14 (1), pp. 41-55.

Koenker R. 1994. Confidence intervals for regression quantiles. Proceedings of the Fifth Prague Symposium on Asymptotic Statistics, 349-359.

Koenker R. 2005. Quantile Regression. Cambridge University Press: Cambridge, New York, USA.

Koenker, R., and Bassett, G. 1978. Quantile Regression. Econometrica, 46 (1), pp. 33-50.

Koenker, R., and Hallock, K.F. 2001. Quantile regression. Journal of Economic Perspectives, 15 (4), pp. 143-156.

Koenker, R., and Machado, J.A.F. 1999. Goodness of Fit and Related Inference Processes for Quantile Regression. Journal of the American Statistical Association, 94 (448), pp. 1296-1310.

Koenker, R., and Schorfheide, F. 1994. Quantile spline models for global temperature change. Climatic Change 28 (4), 395-404.

Luce, C.H., and Holden, Z.A. 2009. Declining annual streamflow distributions in the Pacific Northwest United States, 1948-2006. Geophysical Research Letters, 36 (16), art. no. L16401.

Mata, J., and Machado, J.A.F. 1996. Firm start-up size: A conditional quantile approach. European Economic Review, 40 (6), pp. 1305-1323.

Mazvimavi, D. 2010. Investigating changes over time of annual rainfall in Zimbabwe. Hydrology and Earth System Sciences 14 (12): 2671-2679.

McCullagh, P., and Nelder, J. 1989. Generalized linear regression models. New York: Chapman and Hall.

Mills, A.J., Fey, M.V., Gröngröft, A., Petersen, A., and Medinski, T.V. 2006. Unravelling the effects of soil properties on water infiltration: Segmented quantile regression on a large data set from arid south-west Africa. Australian Journal of Soil Research, 44 (8), art. no. SR05180, pp. 783-797.

Mosteller, F., and Tukey, J.W. 1977. Data analysis and regression: a second course in statistics. Addison-Wesley Pub., Reading, Mass.

Neter, J., Kutner, M.H., Nachtsheim, C.J., and Wasserman, W. 1996. Applied linear regression models. Chicago, IL: IRWIN.

New, M., Hewitson, B., Stephenson, D.B., Tsiga, A., Kruger, A., Manhique, A., Gomez, B., Coelho, C.A.S., Masisi, D.N., Kululanga, E., Mbambalala, E., Adesina, F., Saleh, H., Kanyanga, J., Adosi, J., Bulane, L., Fortunata, L., Mdoka, M.L., and Lajoie, R. 2006. Evidence of trends in daily climate extremes over southern and west Africa. Journal of Geophysical Research D: Atmospheres, 111 (14), art. no. D14102.

Reich, B.J. 2012. Spatiotemporal quantile regression for detecting distributional changes in environmental processes. Journal of the Royal Statistical Society. Series C: Applied Statistics. Article in Press.

Schulze, N. 2004. Applied Quantile Regression: Microeconometric, Financial and Environmental Analysis (PhD dissertation). Eberhard-Karls-Universität Tübingen, Tübingen.

Tareghian, R., and Rasmussen, P. 2012. Analysis of Arctic and Antarctic sea ice extent using quantile regression. International Journal of Climatology, DOI: 10.1002/joc.3491.

Tareghian, R., and Rasmussen, P. 2012. Statistical Downscaling of Precipitation using Quantile Regression. Submitted to Journal of Hydrology.

Tareghian, R., Rasmussen, P.F., Anderson, J., and Kim, S.J. 2010. A study of climate extremes changes over the Canadian prairies using quantile regression. Proceedings, Annual Conference - Canadian Society for Civil Engineering, 1, pp. 55-63.

Timofeev, A.A., and Sterin, A.M. 2010. Using the quantile regression method to analyze changes in climate characteristics. Russian Meteorology and Hydrology, 35 (5), pp. 310-319.

Vincent, L.A., and Mekis, É. 2006. Changes in daily and extreme temperature and precipitation indices for Canada over the twentieth century. Atmosphere - Ocean, 44 (2), pp. 177-193.

Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J.C., Fromentin, J.-M., Hoegh-Guldberg, O., and Bairlein, F. 2002. Ecological responses to recent climate change. Nature, 416 (6879), pp. 389-395.

Weerts, A.H., Winsemius, H.C., and Verkade, J.S. 2011. Estimation of predictive hydrological uncertainty using quantile regression: Examples from the National Flood Forecasting System (England and Wales). Hydrology and Earth System Sciences, 15 (1), pp. 255-265.

Wilby, R.L., Dawson, C.W., Barrow, E.M. 2002. SDSM - A decision support tool for the assessment of regional climate change impacts. Environmental Modelling and Software, 17 (2), pp. 147-159.

Wilks, D.S. 2011. Statistical methods in the atmospheric sciences. 3rd ed. Elsevier. Boston, USA.

Woodall, C.W., and Liknes G.C. 2008. Relationships between forest fine and coarse woody debris carbon stocks across latitudinal gradients in the United States as an indicator of climate change effects. Ecological Indicators 8 (5): 686-690.

Zoellick, B.W., Ulmschneider, H.M., Cade, B.S., and Stanley, A.W. 2004. Isolation of snake river islands and mammalian predation of waterfowl nests. Journal of Wildlife Management, 68 (3), pp. 650-662.