

A Bayesian Quantile Regression Approach for Nonstationary Frequency Analysis of Annual Maximum Sea Level in a Changing Climate

Authors: Uranchimeg, Sumiya, Kim, Jang-Gyeong, Kim, Jin-Guk, Kwon, Hyun-Han, and Lee, Seung Oh

Source: Journal of Coastal Research, 85(sp1) : 536-540

Published By: Coastal Education and Research Foundation

URL: <https://doi.org/10.2112/SI85-108.1>

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

A Bayesian Quantile Regression Approach for Nonstationary Frequency Analysis of Annual Maximum Sea Level in a Changing Climate

Sumiya Uranchimeg[†], Jang-Gyeong Kim, Jin-Guk Kim[†], Hyun-Han Kwon^{†*}, and Seung Oh Lee[§]



www.cerf-jcr.org

[†] Department of Civil Engineering
Chonbuk National University, Korea

[§] Department of Civil Engineering, Hongik
University, Seoul, Republic of Korea

ABSTRACT



www.JCRonline.org

Uranchimeg, S.; Kim, J.-G.; Kim, J.-G.; Kwon, H.-H., and Lee, S.-O., 2018. A Bayesian Quantile Regression Approach for Nonstationary Frequency Analysis of Annual Maximum Sea Level in a Changing Climate. *In*: Shim, J.-S.; Chun, I., and Lim, H.S. (eds.), *Proceedings from the International Coastal Symposium (ICS) 2018* (Busan, Republic of Korea). *Journal of Coastal Research*, Special Issue No. 85, pp. 536–540. Coconut Creek (Florida), ISSN 0749-0208.

Sea level rise is primarily caused by global warming and has been a key consideration in design practices in coastal engineering. The design of coastal structures is currently based on a frequency analysis of the local sea level under the stationary assumption, meaning that the maximum sea level will not vary significantly over time. However, the stationary assumption for sea levels might not be valid in a changing climate. In such a context, this study proposes a systematic approach to investigate nonstationarity in annual maximum sea levels (AMSLs) and offers estimates of design water levels for coastal structures using a non-crossing quantile regression-based nonstationary frequency analysis model within a fully Bayesian framework. The AMSLs for 20 tide gauge stations, each with more than 28 years of hourly records, are considered and compiled in this study. The nonstationarity in the AMSLs are explored by focusing on the change in the scale and location parameter of the probability distributions. The majority of the stations (three-fourths) are found to have an upward-convergent/divergent pattern in the distribution, and the distribution changes are confirmed by significance tests. This study determines an overly simple nonstationary frequency analysis (NSFA) approach with a time-dependent mean value might lead to underestimation of the AMSLs, which results in an increase the failure risk in coastal structures. A more detailed discussion of the characteristics of the distribution changes for the design water level is provided in the paper.

ADDITIONAL INDEX WORDS: *Nonstationarity, annual maximum sea levels, distribution, Bayesian quantile regression, design water level and uncertainty.*

INTRODUCTION

Global warming is expected to continue and is one of the primary contributors to the observed sea level rise (SLR), caused by the thermal expansion of ocean water, the melting of mountain glaciers, and the melting of parts of the Greenland ice sheet. Considering all of these factors, the IPCC published that the sea level will rise by approximately 100 cm or more by 2100 if the concentration increase in carbon dioxide continues as expected, as published in the Climate Change 2014 Synthesis Report (Pachauri *et al.*, 2014). The consequences of SLR have led to increases in the frequency and intensity of extreme water levels in coastal areas. The local sea level is typically the most critical factor for many coastal applications, including coastal mapping, marine boundary delineation, coastal zone management, coastal flood defense and engineering, insurance, and design of sustainable habitat restoration, which are all exacerbated by SLR.

A reliable estimation of extreme sea level events is needed to mitigate the hazardous impact of extreme water level conditions in coastal areas. In the past, many engineering practices such as

water resource engineering and coastal engineering were based on stationary assumptions such as the stationary extreme value analysis approach. However, the assumption of stationarity is untenable in most cases due to existing trends in the mean and variability of annual maximum sea levels (AMSLs) (Khaliq *et al.*, 2006). Thus, an advanced statistical model that considers time-varying changes in the data is needed to address such issues.

Given diverse evidence of climate change, it is unlikely that the assumption of stationarity in hydrologic data is sound. Therefore, advanced methods in extreme value analysis must be developed and applied (Khaliq *et al.*, 2006). In recent years, the concept of nonstationary extreme value analysis has been developed and applied in various fields of study (Butler *et al.*, 2007; Hundedeha *et al.*, 2008; Katz; Parlange and Naveau, 2002; Lee; Kwon and Kim, 2010, 2012). In the nonstationary approach, the parameters of the distribution function are replaced by time-dependent parameters, such that the results of the extreme value analysis also vary with time.

Time-dependent models of the generalized extreme value (GEV) distribution for determining return periods have been studied recently and applied to hydro-meteorological data (Katz; Parlange and Naveau, 2002; Kim *et al.*, 2016; Méndez *et al.*, 2007). Prior to an extreme frequency analysis, the stationarity of the data must be confirmed with methods such as the Mann-

DOI: 10.2112/SI85-108.1 received 30 November 2017; accepted in revision 10 February 2018.

*Corresponding author: hkwon@jbnu.ac.kr

©Coastal Education and Research Foundation, Inc. 2018

Kendall (MK) test, Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, or the Augmented Dickey-Fuller (ADF) test. However, these methods are generally based on the mean of the response variables and might not be appropriate for extreme events that substantially deviate from the mean (Katz and Brown, 1992).

The aim of this paper is to provide a brief introduction to a novel nonstationary extreme value analysis method based on a Bayesian quantile regression. The novel approach is described in order to estimate the future design water level for coastal engineering tasks. The approach described here provides an advanced investigation into return levels and distributional changes over extended time horizons. An estimate of a nonstationary GEV distribution extreme analysis is adopted in order to compare a more traditional method with the estimates of novel approaches. These approaches are applied to the AMSLs of tide gauge stations in South Korea.

METHODS

Local Sea Level Observations

In this study, we first extracted AMSLs from tide gauge stations. The tide gauge data was obtained from the Korean Hydrographic and Oceanographic Administration (KHOA, <http://www.khoa.go.kr>). A total of 45 tide gauge stations are operated, and the KHOA performs careful quality control on the data, including eliminations of any datum shift errors and abnormal values. Additionally, this study considered the record length and the proportion of values missing as the quality assessment criteria. Stations with less than 5% missing values and a record length greater than 28 years were chosen for analysis. The missing values are distributed in an unsystematic way and differs at each stations, therefore the missing data are not being considered in this study. Out of 45 tide gauge stations, 20 were selected which covers all coastal areas in South Korea. The spatial distribution of the tide gauge stations is shown in Figure 1.

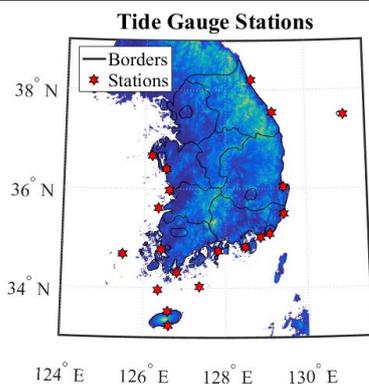


Figure 1. Spatial distribution of selected tide gauge stations.

Nonstationary Frequency Analysis (NSFA)

Several extreme value analysis techniques were developed to estimate the return periods of extreme sea levels (Bernardara; Andreewsky and Benoit, 2011; Lee; Kwon and Kim, 2010, 2012; Woodworth and Blackman, 2002). Many studies have suggested the use of the GEV approach to estimate the extreme sea levels (Kim *et al.*, 2016; Zhang and Sheng, 2013). Before applying NSFA, the data should be tested against trends or changes/shifts

in the mean and variability in order to assess its stationarity or the nonstationarity (Hawkes *et al.*, 2008). In this study, the trend and stationarity of the data were tested by the MK test, ADF test, and KPSS test prior to NSFA.

In this study, the nonstationary form of the GEV distribution was selected and is generally expressed as Equations (1) and (2);

$$GEV(x, t) = \exp \left[- \left(1 + \xi \times \frac{x - \mu_{GEV}(t)}{\sigma} \right)^{-\frac{1}{\xi}} \right] \quad (1)$$

$$\mu(t) = \alpha * t + \beta \quad (2)$$

where x is the independent value (*e.g.*, water level), $\mu_{GEV}(t)$ the time-dependent location parameter, σ is the scale parameter, and ξ is the shape parameter. α and β are regression coefficients.

Bayesian Quantile Regression (BQR)

For the analysis of distributional changes in AMSLs, this study proposes the following quantile regression model. Let $y_{i,p}$ denote the extreme sea level for year i , and p denotes the quantile level $p \in (0, 1)$; then, the generic structure of the model is given by (Equation 3)

$$y_{i,p} = \eta_{i,p} + \epsilon_{i,p} = x_i' \beta_p + \epsilon_{i,p} \quad (3)$$

where x_i' is a vector of regressors, β_p is a vector of regression coefficients for the p^{th} conditional quantile, $\epsilon_{i,p}$ is an unspecified quantile-specific error term, for which no distribution is specified other than the constraint $Q_{\epsilon_{i,p}}(p) = 0$, and is estimated by minimizing the asymmetrically weighted absolute deviations through linear programming.

A Bayesian inference commonly requires a likelihood; thus, a typical approach to Bayesian quantile regression is to assume an Asymmetric Laplace Distribution (ALD) for ϵ , which enables the maximization of a likelihood function of an independently distributed ALD. The ALD is characterized by a set of parameters such as location (μ), precision (δ^2), and skewness ($0 < p < 1$) by letting $\mu = 0$ to ensure $Q_{\epsilon}(p) = 0$.

According to the features of ALD (Yue and Rue, 2011), purposed quantile regression model can be rewritten as following

$$y_{i,p} = \eta_{i,p} + \frac{1-2p}{p(1-p)} w_{i,p} + \sqrt{\frac{2w_{i,p}}{\delta^2 p(1-p)}} z_i \quad (4)$$

where \mathbf{W} is an exponentially distributed random variable with rate parameter δ^2 , and \mathbf{Z} has a standard normal distribution $\mathbf{N}(\mathbf{0}, \mathbf{1})$. Posterior estimates can be subsequently obtained using Bayesian updates conditioned on the exponential random variable \mathbf{W} . And $w_{i,p}$ and $z_{i,p}$ are subject-specific values of \mathbf{W} and \mathbf{Z} , respectively, and $\eta_{i,p}$ is defined in Equation (3).

The Bayesian inference framework was adopted to enable the inclusion and improved consideration of the uncertainties in the proposed quantile regression model parameters. The posterior inference for the desired quantiles, p_j , simultaneously proceeds via data augmentation by introducing observation-specific latent weights, w_{i,p_j} , as specified in Equation (4).

Analytical integration of the joint distribution and sampling was done by Markov Chain Monte Carlo (MCMC) method in WinBUGs model with self-written script.

Distributional Change Detection

A Bayesian quantile regression-based distributional change detection approach in AMSLs was proposed based on a previous study (Shiau and Huang, 2015). Figure 2 shows an example of the basic concept, which exemplifies the detection of a distribution change using Bayesian quantile regression.

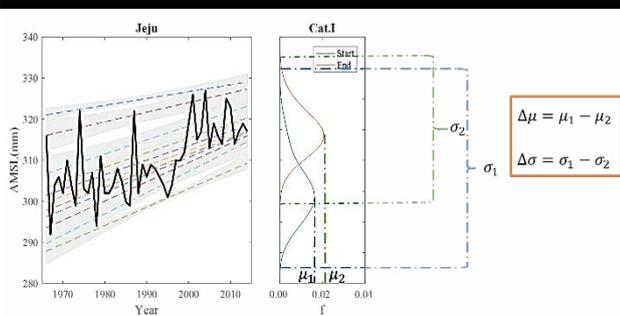


Figure 2. An illustration of construction of an empirical PDF of the AMSLs through quantile regression for a wide range of quantiles

Since the sign of the slope parameter of the AMSL can vary with quantile, different behaviors in probabilistic distributions can be produced from year to year. Detecting the distribution changes in AMSL over time is thus made possible by comparing the shapes of the derived empirical probability density functions (PDFs), as illustrated in Figure 2. To compare the shapes of the empirical PDFs, a simplified comparison is performed to illustrate changes in the location (mean) and scale (dispersion) of the PDFs. Thus, there are nine possible categories in terms of distributional changes depending on both changes of scale and location of PDFs, as identified by Shiau and Huang (2015).

RESULTS

Trends and Stationary Tests

As mentioned in the methodology section, well known statistical methods are conducted, such as the MK test, for determining the existence of linear trends in AMSLs, and the KPSS test and ADF test are employed to determine the existence of stationarity in AMSLs. The results are illustrated in Figure 3.

The MK test detected significant linear trends in the AMSL at 13 tide gauge stations, which were distributed over all coastal areas of South Korea. The ADF test and KPSS test detected statistically significant nonstationary in AMSLs at 13 and 18 tide gauge stations, respectively.

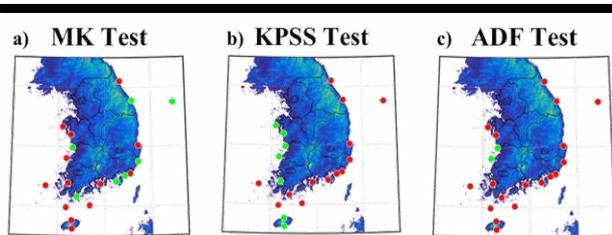


Figure 3. Trends and stationarity test results. The green dots indicate no linear trends/stationarity, and red dots indicate statistically significant trend/nonstationarity in AMSLs. The panel indicates MK test (a), ADF test (b), and KPSS test (c).

Distributional Changes in the AMSLs and their Relations to Design Water Level

The purpose of the study was to identify the distributional changes in the desired quantiles, inferred from the time-varying mean and variance using the entire posterior distributions. The BQR model was applied to the AMSLs for a wide range of quantiles (*i.e.*, 0.05, 0.15, 0.25, 0.35, 0.45, 0.50, 0.65, 0.75, 0.85, and 0.95) for all tide gauge stations.

Table 1. Categorization of distributional changes in the AMSLs and summary of the significance test for the difference in the distribution. (·) indicates p-value and test statistics in the two-sample KS and AD test, respectively

Station Name	Category	Significant Change in Distribution	
		Two-Sample KS test	Two-Sample AD test
1 Anheung	I	Yes(0)	Yes(18.992)
2 Boryeong	I	Yes(0)	Yes(14.733)
3 Busan	III	Yes(0)	Yes(15.794)
4 Chujado	I	Yes(0)	Yes(11.777)
5 Heuksando	I	Yes(0)	Yes(21.324)
6 Gadukdo	III	Yes(0.001)	Yes(5.588)
7 Geomundo	III	Yes(0)	Yes(11.711)
8 Gunsan	III	Yes(0)	Yes(10.191)
9 Jeju	I	Yes(0)	Yes(28.826)
10 Mokpo	I	Yes(0)	Yes(38.634)
11 Mukho	VI	No	Yes(2.901)
12 Pohang	I	Yes(0)	Yes(28.132)
13 Seogwipo	III	Yes(0)	Yes(24.536)
14 Sokcho	I	Yes(0)	Yes(16.344)
15 Tongyoung	VI	No	No(1.196)
16 Ulleungdo	VI	No	No(1.652)
17 Ulsan	VI	No	Yes(2.528)
18 Wando	IX	Yes(0)	Yes(7.588)
19 Wido	III	Yes(0)	Yes(12.947)
20 Yeosu	IX	Yes(0)	Yes(9.704)

Four categories (*i.e.*, Categories I, III, VI, and IX) of distributional change in tide gauge stations were identified, and two-sample KS and AD tests were applied to assess the significance of differences in the estimated distributions in the initial and final years of the records for all stations, as summarized in Table 1. The two-sample KS test showed that 16 out of 20 stations have significantly different distributions at the 10% significance level, while the two-sample AD tests detected 18 stations with AD statistics over the critical value of 2.492, indicating that they have significantly different distributions. Categories I and III showed an increasing trend in the mean value of AMSLs, while Categories VI showed no change in the mean value of AMSLs, and Category IX showed a decreasing trend in the mean value of AMSLs, with associated variances.

A reliable design water level estimation is vital for the effective performance of coastal structures. Changes in design water level for a set of representative stations under the identified categories are illustrated in Figure 4. In Figure 4, the design water levels for the first and final years of the observation period were compared with the design water level estimated by the NSFA method.

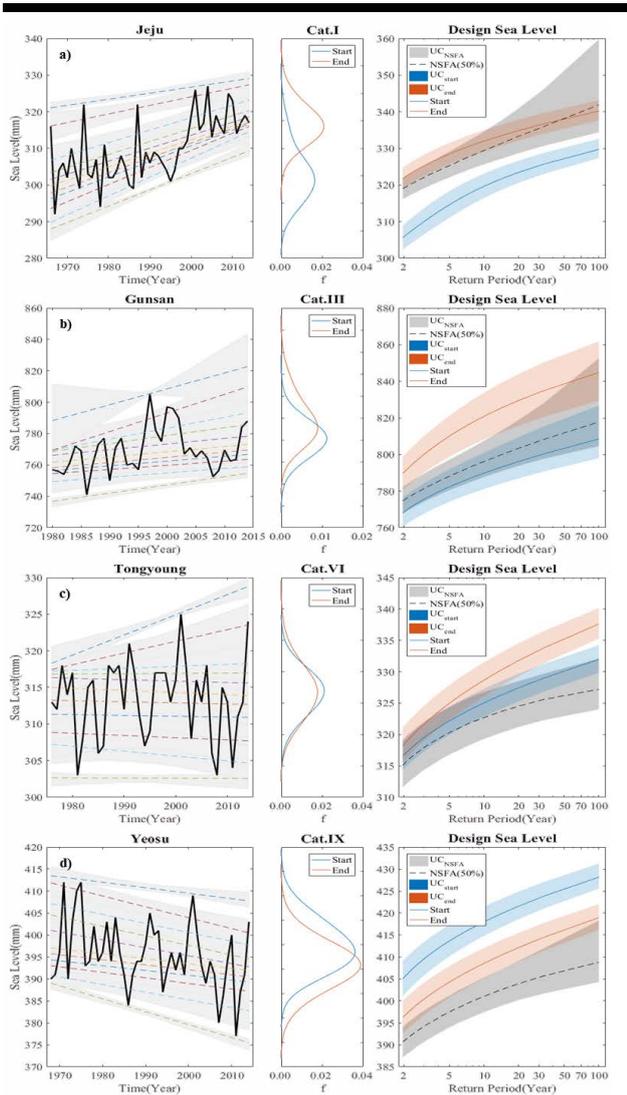


Figure 4. Distributional changes in the AMSLs compared with design sea level estimates of the NSFA approach. The left panel represents the time series of AMSL and its quantiles with Bayesian credible bounds. The middle panel represents the estimated distribution of quantiles in the first and last years of the records. The right panel represents nonstationary design rainfall estimates, corresponding to different return periods with those of the NSFA approach. The stations are a) Jeju, b) Gunsan, c) Tongyoung, d) Yeosu, which are selected from different categories.

DISCUSSION

Assessment of distributional changes in the AMSLs over the various quantiles contribute to identifying the trends involved in different aspects of coastal structure and engineering. Specially, information on the distributional changes for a specific quantiles such as higher tail of the AMSL distribution, supports estimating design water level due to the fact that appropriate design water level required for coastal defense structure and engineering in

contests of protecting and reducing risks for people who lives in coastal zones.

Traditional nonstationary extreme sea level analysis is considered when the existence of time-varying change is detected in the time series. Thus, this study adopted an MK test for determining the existence of linear trends in AMSLs and KPSS test and ADF test for existence of stationary in AMSLs. 13 tide gauge stations out of 20 were detected to have a significant linear trend in their AMSLs and there was no spatial characteristics (Figure 3a). But stationary test results found more statistically significant nonstationary tide gauge stations in their AMSLs. 13 and 18 tide gauge stations were detected by KPSS test and ADF test, respectively (Figure 3b and 3c).

Four stations were selected to represent the different distributional changes for each category in order to demonstrate the behavior of trend and design water level: Jeju (Category I), Gunsan (Category III), Tongyoung (Category VI), and Yeosu (Category VII), as shown in Figure 4.

The spatial distribution of the identified categories associated with the distributional changes in the AMSLs is illustrated in Figure 5. The Category I and III stations are distributed at all coastal zones of South Korea, while Category VI and IX are detected in the South and East Coast regions. Interestingly, tide gauge stations on the west coast all show increasing trends (Categories I and III). The changes in the distributions are significant, except at 4 stations with the two-sample KS test and 2 stations with the two-sample AD test.

The spatial distribution of the identified categories associated with the distributional changes in the AMSLs is illustrated in Figure 5. The Category I and III stations are distributed at all coastal zones of South Korea, while Category VI and IX are detected in the South and East Coast regions. Interestingly, tide gauge stations on the west coast all show increasing trends (Categories I and III). The changes in the distributions are significant, except at 4 stations with the two-sample KS test and 2 stations with the two-sample AD test.

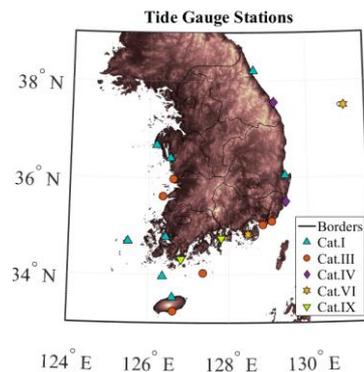


Figure 5. Spatial distribution of the categories associated with the distributional changes in AMSLs.

The spatial distribution of the identified categories associated with the distributional changes in the AMSLs is illustrated in Figure 5. The Category I and III stations are distributed at all coastal zones of South Korea, while Category VI and IX are detected in the South and East Coast regions. Interestingly, tide

gauge stations on the west coast all show increasing trends (Categories I and III). The changes in the distributions are significant, except at 4 stations with the two-sample KS test and 2 stations with the two-sample AD test.

CONCLUSIONS

This study proposed a BQR model to assess nonstationarity by exploring distributional changes in the AMSLs of tide gauge stations over the observation period. More specifically, the distributional changes were categorized into a number of classes in which nonstationarity was largely defined by changes in location and scale parameters of the probability distribution. The BQR model-based nonstationarity detection scheme was then utilized to understand the impact of distributional changes on estimates of the design water level.

The proposed BQR model provides reliable estimates for detecting distribution changes by simultaneously fitting models over the quantiles found to be consistent with the monotonic response desired under the non-crossing constraints within a fully Bayesian framework. Four categories of distributional change were found in the AMSLs. Most of the stations were classified as Category I, which is characterized by an upward-divergent pattern in the distribution. Six stations were classified as Category III.

The BQR model-based nonstationarity frequency analysis approach provided a method for understanding the key attributes of the dynamic distributional changes in the AMSLs in a changing climate. In order to focus more clearly on nonstationarity and its direct impact on design water level estimates, this study explored changes in design water level estimates for different nonstationarity categories. For Categories I, III, and IV, a noticeable increase in design sea level was observed, while Categories IV and IX showed no evidence of association with risk of increased extreme sea levels.

ACKNOWLEDGMENTS

This research was supported by a grant (17AWMP-B079625-04) from the Water Management Research Program funded by the Ministry of Land, Infrastructure and Transport of the government of Korea.

LITERATURE CITED

- Bernardara, P.; Andreewsky, M., and Benoit, M., 2011. Application of regional frequency analysis to the estimation of extreme storm surges. *Journal of Geophysical Research: Oceans*, 116(C2).
- Butler, A.; Heffernan, J. E.; Tawn, J. A., and Flather, R. A., 2007. Trend estimation in extremes of synthetic North Sea surges. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 56(4), 395-414.
- Hawkes, P. J.; Gonzalez-Marco, D.; Sánchez-Arcilla, A., and Prinos, P., 2008. Best practice for the estimation of extremes: A review. *Journal of Hydraulic Research*, 46(S2), 324-332.
- Hundecha, Y.; St-Hilaire, A.; Ouarda, T.; El Adlouni, S., and Gachon, P., 2008. A nonstationary extreme value analysis for the assessment of changes in extreme annual wind speed over the Gulf of St. Lawrence, Canada. *Journal of Applied Meteorology and Climatology*, 47(11), 2745-2759.
- Katz, R. W., and Brown, B. G., 1992. Extreme events in a changing climate: variability is more important than averages. *Climatic Change*, 21(3), 289-302.
- Katz, R. W.; Parlange, M. B., and Naveau, P., 2002. Statistics of extremes in hydrology. *Advances in water resources*, 25(8), 1287-1304.
- Khaliq, M.; Ouarda, T.; Ondo, J.-C.; Gachon, P., and Bobée, B., 2006. Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. *Journal of Hydrology*, 329(3), 534-552.
- Kim, Y.-T.; Cho, H.-R.; Uranchimeg, S.; Lee, S.-O., and Kwon, H.-H., 2016. A Hierarchical Bayesian Model based Nonstationary Frequency Analysis of Extreme Sea Level under Climate Change along the Shorelines in South Korea. *Journal of Coastal Research*, 75(sp1), 745-749.
- Lee, J.-J.; Kwon, H.-H., and Kim, T.-W. (2010). Concept of trend analysis of hydrologic extreme variables and nonstationary frequency analysis. Paper presented at the Proceedings of the Korea Water Resources Association Conference.
- Lee, J.-J.; Kwon, H.-H., and Kim, T.-W., 2012. Spatio-temporal analysis of extreme precipitation regimes across South Korea and its application to regionalization. *Journal of Hydro-environment Research*, 6(2), 101-110.
- Méndez, F. J.; Menéndez, M.; Luceño, A., and Losada, I. J., 2007. Analyzing monthly extreme sea levels with a time-dependent GEV model. *Journal of Atmospheric and Oceanic Technology*, 24(5), 894-911.
- Pachauri, R. K.; Allen, M. R.; Barros, V. R.; Broome, J.; Cramer, W.; Christ, R., . . . Dasgupta, P. (2014). *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change: IPCC.*
- Shiau, J.-T., and Huang, W.-H., 2015. Detecting distributional changes of annual rainfall indices in Taiwan using quantile regression. *Journal of Hydro-environment Research*, 9(3), 368-380.
- Woodworth, P. L., and Blackman, D. L., 2002. Changes in extreme high waters at Liverpool since 1768. *International journal of climatology*, 22(6), 697-714.
- Yue, Y. R., and Rue, H., 2011. Bayesian inference for additive mixed quantile regression models. *Computational Statistics & Data Analysis*, 55(1), 84-96.
- Zhang, H., and Sheng, J., 2013. Estimation of extreme sea levels over the eastern continental shelf of North America. *Journal of Geophysical Research: Oceans*, 118(11), 6253-6273.