International Journal of
Civil Engineering (IJCE)
ISSN(P): 2278–9987; ISSN(E): 2278–9995
Vol. 10, Issue 2, Jul–Dec 2021; 13–22

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

PROJECTION OF FUTURE CLIMATE DATA FOR TORONTO CITY USING GENERALIZED EXTREME VALUE THEORY

Mostafa Hassan[1], Lamya Amleh[2]& Hesham Othman[3]

[1] Research Assistant, Department of Civil Engineering, Ryerson University, Toronto, Ontario, Canada

[2] Professorof Structure Engineering, Department of Civil Engineering, Ryerson University, Toronto, Ontario, Canada

[3] Assistant Professor, Department of Civil Engineering, Ryerson University-Cairo Campus/ Universities of Canada in Egypt, New Administrative Capital, Egypt

ABSTRACT

Climate change is now recognized as one of the most critical impact on infrastructure.The objective of this research is to predict the future minimum temperature, hourly mean temperature and wind hourly mean data using generalized extreme value theory (GEVT) for next (50 -100 years) return period. In this research, the historical minimum temperature, hourly mean temperature and wind hourly mean data for Toronto city have been gathered annually from Canada weather statistics. The full characterization for the minimum temperature, hourly mean temperature and wind hourly mean data have been performed. The analysis was done by different software programs such as (SPSS and MATLAB).The minimum temperature, hourly mean temperature and wind hourly mean data were tested using Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test and Augmented Dickey–Fuller test (ADF) to recognize the data type (i.e., stationary or nonstationary).Based on the results of this study,it is observed that the minimum temperature, hourly mean temperature and wind hourly mean data are nonstationary data.Additionally, the return level for minimum temperature, hourly mean temperature and wind hourly mean data were obtained for the next 50-100 years return period.

KEYWORDS:*Generalized Extreme Value Theory, Return Level, Non Stationary, Minimum Temperature, Wind Hourly Mean, Hourly Mean Temperature*

Article History

Received: 07 Oct 2021 | Revised: 11 Oct 2021 | Accepted: 16 Oct 2021

INTRODUCTION

Extreme weather events are the major problem occurs to the infrastructure. As a result of climate change, their frequency and intensity may increase and infrastructure may be subjected to severe damage [1], [2]. It is important to have models of extreme events which are able to consider future trends due to climate change. The climate data are time series which is a sequential set of data points, measured typically over successive times. Climate data is mathematically defined as a set of vectors x (t), t = 0, 1, 2, where (t) represents the time elapsed [3],[4], [5]. In general, time series data consistof four main components which can be separated from the observed data. The components are trend, cyclical, seasonal and irregular components [6]. In time series forecasting, past observations are collected and analyzed to develop an appropriate mathematical model which captures the underlying data generating process for the series [7]. A time series is non-deterministic in nature. The general assumption for time series variables x_t is independent and identically distributed (i.i.d) following the normal distribution. Climate researchers expect future climate change in Canada and other Arctic places to be more pronounced than it is elsewhere in the world. Thus, Canada's infrastructures could be at risk if actions are not taken in order to consider or adapt the climate change.

Methods for modeling extremes of nonstationary processes are similar to those of stationary ones and include generalized extreme value theory (GEVT) distribution for block maxima; generalized pareto distribution (GPD) for threshold exceedances and point process characterizations of Extremes (PPE) [8]. The GEVT is the classical extreme value theory which provides a strict framework for the analysis of climate extremes and their return period. In addition, GEVT distribution assumed data to be independent, identically distributed according to Leclerc and Ouarda [9]. The GEVT distribution is widely employed in the environmental sciences and for modelling extremes [10]. In nonstationary GEVT distribution [11], the mean parameter is expressed as a function of time (t) as in equation (1) and possibly other covariates [3]. Caroniet al., (2015) assumed that the location (μ) is function of time (t) as shown in equation (2), while the scale and shape parameters aren't function of time (t) [12].

$$F(y;\mu(t),\sigma,\varepsilon) = exp\left\{-\left[1+\varepsilon\left(\frac{y-\mu(t)}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}}\right\} \quad (1)$$

$$\mu(t) = \mu_0 + \mu_1 t + \mu_2\, t^2 + \mu_3\, t^3 \quad (2)$$

The cornerstone of extreme values theory is the extremely types of theorem [3], providing the convergence of the distribution for one of the maximum three limiting distributions usually referred as Gumbel or extreme value type I distribution [13], Fréchet or extreme value type II distribution [14], and Weibull or extreme value type III distribution [15]. The extreme value Parameters location μ, scale σ, and shape ε satisfy the inequalities: $-\infty < \mu(t) < \infty$; $\sigma > 0$; $-\infty < \varepsilon < \infty$. The type II (Fréchet) and type III (Weibull) classes of extreme value distributions correspond respectively to the cases $\varepsilon > 0$ and $\varepsilon < 0$, and the subset of the GEVT family with $\varepsilon = 0$ can be interpreted as the limit of equation (1) where ε tends to 0, correspond to the type I (Gumbel) with cumulative distribution function.

The model of extreme value theory focuses on the statistical behavior of $M_n = \max\{X_1,,..,, X_n\}$, where $(X_1,.,.., X_n)$, is a sequence of independent random variables having a common distribution function F. So, M_n represents the maximum of the process over n time units of observation. If n is the number of observations in a year, then M_n corresponds to the annual maximum. In theory the distribution of M_n can be derived exactly for all values of n as in equation (3) [16], [17].

$$\Pr\{M_n \le Z\} = \Pr(X_1 \le Z, \ldots\ldots, X_n \le Z) = \{F(z)\}^n \quad (3)$$

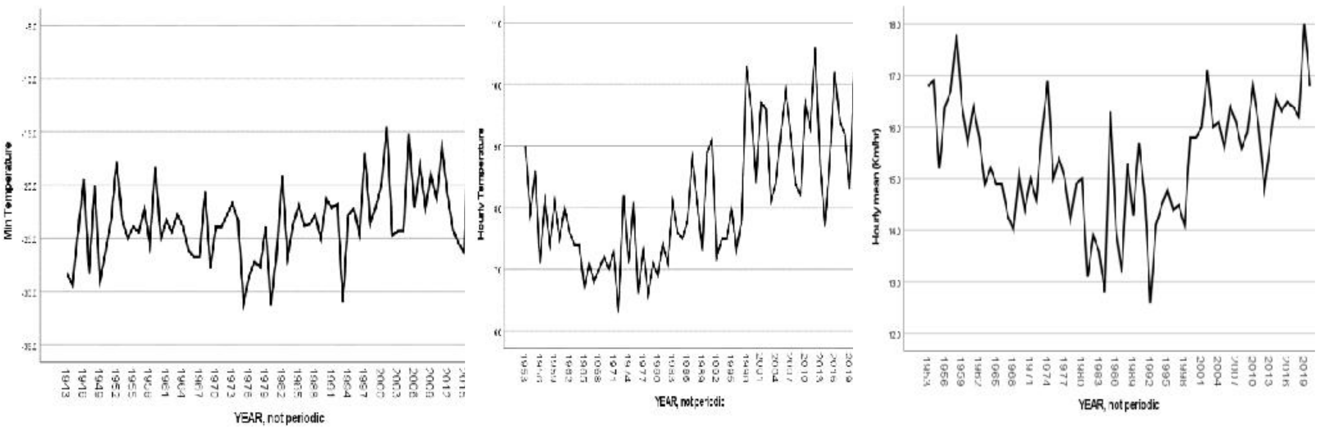
OBJECTIVE OF THIS RESEARCH

The aim for this research is to predict the future minimum temperature, hourly mean temperature and wind hourly mean for Toronto city in the future for a return period which is equal to 50-100 years return period using generalized extreme value theory (GEVT) (Block maxima). Moreover, minimum temperature, hourly mean temperature and wind hourly mean were modelled as nonstationary data. These future data can help in checking the applicability of the current climate loads of the Design Code to model future climatic actions.

In the research methodology, the prediction for the climate data for Toronto city was conducted using numerical method by different software such as MATLAB, SPSS for statistics. In addition, the statistical tests were applied on different climate data in order to check the data is stationary or nonstationary using MATLAB. Moreover, prediction of the minimum temperature, hourly mean temperature and wind hourly mean for Toronto city was conducted using MATLAB software program.

Historical data for minimum temperature, hourly mean temperature and wind hourly mean for Toronto city were collected annually from 1943-2020, 1953-2020 and 1953-2020 respectively as shown in Figure 1. The prediction for minimum temperature, hourly mean temperature and wind hourly mean in the future will be achieved using generalized extreme value theory for next 50 and 100 years return period.

The descriptive statistics for minimum temperature, hourly mean temperature and wind hourly mean are shown in Table 1. The mean for minimum temperature, hourly mean temperature and wind hourly mean are -23.195 degrees Celsius, 8.1 degrees Celsius and 15.384 Km/hr respectively. In addition, the standard deviation for minimum temperature, hourly mean temperature and wind hourly mean for Toronto city are 4.0931, 1.0411 and 1.1553 respectively as shown in Table 1.



(a). Minimum Temperature.(b). Hourly Mean Temperature. (c) Wind Hourly Mean.

Figure 1: Time Series for Different Climate Data for Toronto City.

Table 1: Descriptive Statistics for Different Climate Data

Statistics Data for Maximum Temperature	N=78 (Minimum Temperature)	N=68 (Hourly Mean Temperature)	N=68 (Wind Hourly Mean)
Mean	-23.195	8.11	15.384
Median	-23.65	8.0	15.5
Mode	-23.9	8.1	16.4
Std. Deviation	4.0931	1.0411	1.1553
Variance	16.753	1.084	1.335
Skewness	1.006	0.566	-0.216
St. Error of Skewness	0.272	0.291	0.291
Kurtosis	3.143	-0.536	-0.22
Std. Error of Kurtosis	0.538	0.574	0.574
Range	25.1	4.3	5.4
Minimum	-31.3	6.3	12.6
Maximum	-6.2	10.6	18

Types of Test Applied on the Data

This research conducts two types of tests applied on data. It includes Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) and Augmented Dickey-Fuller (ADF) tests were applied on the data points for minimum temperature, wind hourly mean and mean and hourly mean temperature for Toronto city in Canada in order to check if the data is stationary or nonstationary.

Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) Test

This test assesses the null hypothesis that a univariate time series is trend stationary against the alternative that it is a nonstationary unit root process as defined in MATLAB [18], [19], [20].

Augmented Dickey-Fuller Test for a Unit Root

The Augmented Dickey – Fuller test for a unit root assesses the null hypothesis of a unit root using the model as defined as defined in MATLAB coding program.

It was concluded that the minimum temperature indicates that its p-value for different lags is less than the critical value (5 %) from KPSS Test, Test; therefore, the time series for minimum temperature is nonstationary. Moreover, the hourly mean temperature and wind hourly mean indicate also p-value for different lags is less than the critical value (5 %) which means that data are non-stationary. Moreover, it was observed that the minimum temperature, hourly mean temperature and wind hourly mean indicate failure to reject the unit root null for ADF test, therefore the time series for minimum temperature, hourly mean temperature and wind hourly mean are nonstationary.

ANALYSIS OF RESULTS

Predicting return level values for different climate data was conducted through generalized extreme value theory in order to predict the rare event in the future. Based on the extreme value theory that derives the GEVT distribution, it can be used to fit a sample of extremes to the GEVT distribution in order to obtain the parameters that best explain the probability distribution of the extremes. The estimation of the parameters for the shape, scale and location for the minimum temperature data were calculated using MATLAB function. It was concluded that the shape parameter is -0.09 (Weibull Distribution), the scale is 3.5677 and the location is -24.9046 for maximum likelihood parameter estimates. Probability density and cumulative distribution functions for minimum temperature data were plotted as shown in Figure2 and Figure3 respectively and cumulative distribution function for GEVT (Block maxima) was calculated from equation (4), by substituting in equation (4) by the previous calculated parameters, Thus CDF curve is obtained where; $\sigma > 0$, $-\infty < \mu, \xi < \infty$.

$$F(X; \mu,\sigma,\varepsilon)=\exp\{-[1+\varepsilon((X-\mu)/\sigma)]^{(-1/\varepsilon)}\} \quad (4)$$

Using the GEVT distribution, Extreme value analysis is used to compute the return levels of extremes for corresponding different return periods as in equations (5), (6). In this approach, return levels are expressed as a function of the return period T [21], [22].

$$[T = \frac{1}{1-\text{p}}] \quad (5)$$

where p is the non-exceedance probability of occurrence in a given year. The P return level (q_p) derived from the GEVT distribution can be expressed as in equation (6) [3], [22].

$$q_p = ((\frac{-1}{\ln p})^{\varepsilon} - 1) * \frac{\sigma}{\varepsilon} + \mu \quad (6)$$

The return value is defined as a value that is expected to be equal or exceeded on average once every interval of time (T) with a probability of 1/T. Therefore, once obtaining the equation for CDF of the GEVT distribution as in (equation (4)) = (1-(1/T)). Then, calling “T" on the right-hand side of this equation as a return period, and "X" in equation (4) (left hand side) is the return value. The return value can be calculated by solving this equation (by inverting the GEVT distribution). In this recent work, return values were calculated for 5, 10, 15, 20, 25, 50, 70 and 100 years return period.

Return level values for minimum temperature are considered as nonstationary for different return periods (5-10-50-100) were obtained using inverse function for GEVT in MATLAB in order to obtain the value of minimum temperature in the future as shown in Figure 4. It was observed that the minimum temperature for 50 years return period will be -10.29 degrees Celsius and for 70 years return period, the minimum temperature will be -8.31 degrees Celsius. Moreover, it was observed that the minimum temperature for 100 years return period will be -5.7 degrees Celsius. Moreover, all the return level values for minimum temperature in the future are considered after year 2020.

The estimation of the parameters for the shape, scale and location for the hourly mean temperature data were calculated using MATLAB function. It was concluded that the shape parameter is -0.0333(Weibull Distribution), the scale is 0.8477and the location is 7.6397for maximum likelihood parameter estimates. Moreover, Probability density and cumulative distribution functions for hourly mean temperature data are plotted in Figure 5 and Figure 6 respectively.

Return level values for hourly mean temperature at different return periods (5-10-50-100) were obtained using the inverse function for GEVT in MATLAB in order to obtain the return level value of hourly mean temperature in the future as shown in Figure 7. It was observed that the hourly mean temperature for 50 years return period will be 12.36 degrees Celsius and for 70 years return period, the hourly mean temperature will be 13.26 degrees Celsius. Moreover, it was observed that the hourly mean temperature for 100 years return period will be 14.5 degrees Celsius. Moreover, all the return level values for hourly mean temperature in the future are considered after year 2020.

The estimation of the parameters for the shape, scale and location for wind hourly mean data were calculated using MATLAB function. It was concluded that the shape parameter is -0.3327 (Weibull Distribution), the scale is 1.1821 and the location is 15.0048. Moreover, Probability density and cumulative distribution functions for wind hourly mean data are plotted in Figure 8 and Figure 9 respectively.

It was observed that the wind hourly mean for next 50 years return period, the return level value will be 18 Km/hr and for 70 years return period, the return level value for wind hourly mean will be 18.279 Km/hr. In addition, the return level value for wind hourly mean will be 18.6 Km/hr at 100 years return period as shown in Figure 10. In addition, all the return level values for wind hourly mean in the future are considered after year 2020.

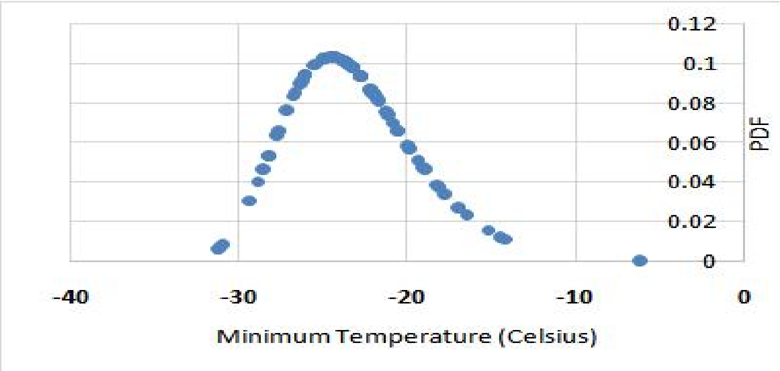


Figure 2: Probability Density Function Using GEVT for Minimum Temperature.

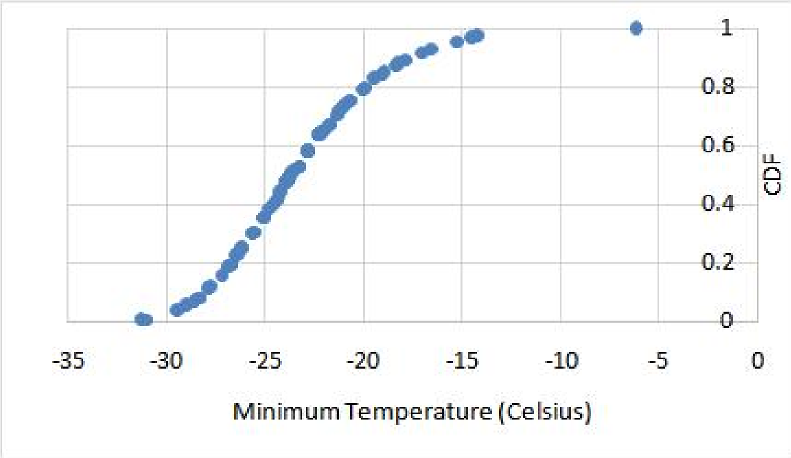


Figure 3: Cumulative Distribution Function Using GEVT for Minimum Temperature.

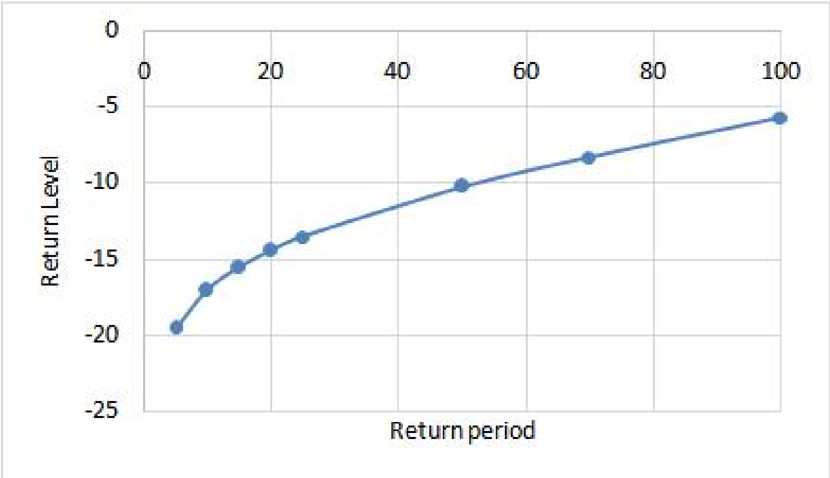


Figure 4: Return Level for Minimum Temperature versus Different Return Period for Toronto City (Non-Stationary).

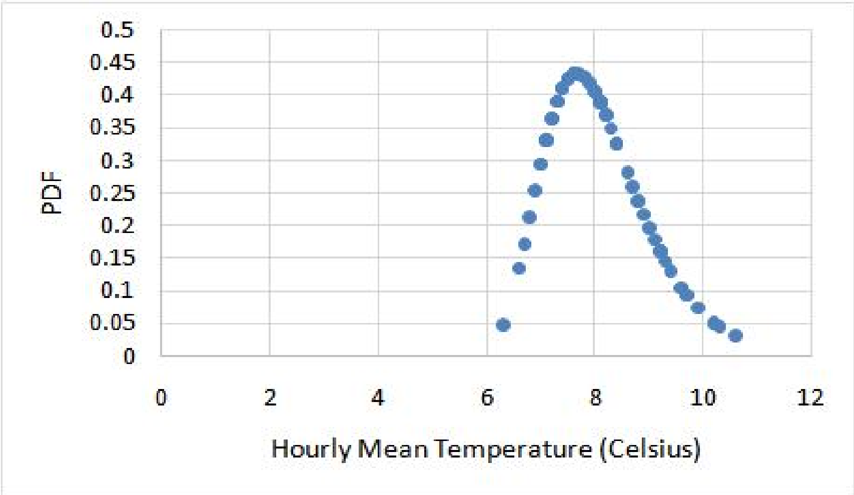


Figure 5: Probability Density Function for Hourly Mean Temperature for Toronto City.

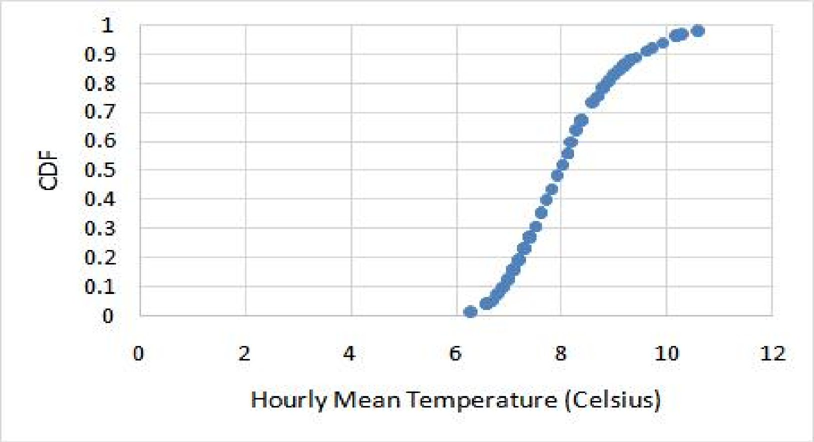


Figure 6: Cumulative Distribution Function for Hourly Mean Temperature for Toronto City.

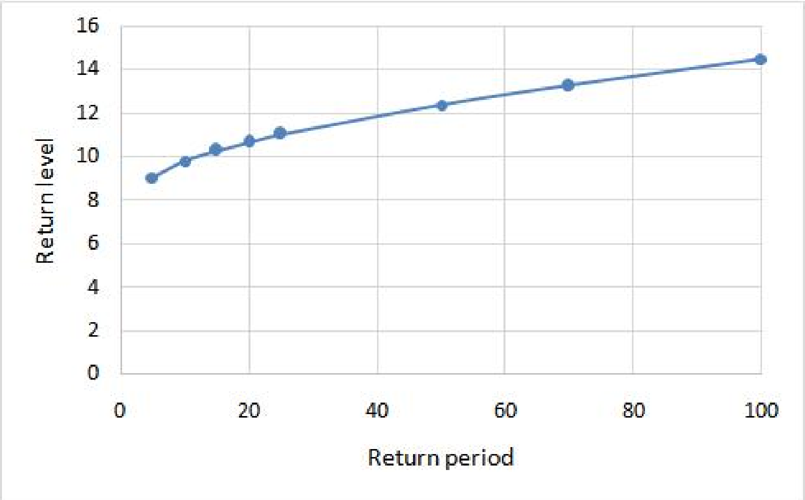


Figure 7: Return Level for Hourly Mean Temperature Versus Return Period for Toronto City (Non-Stationary).

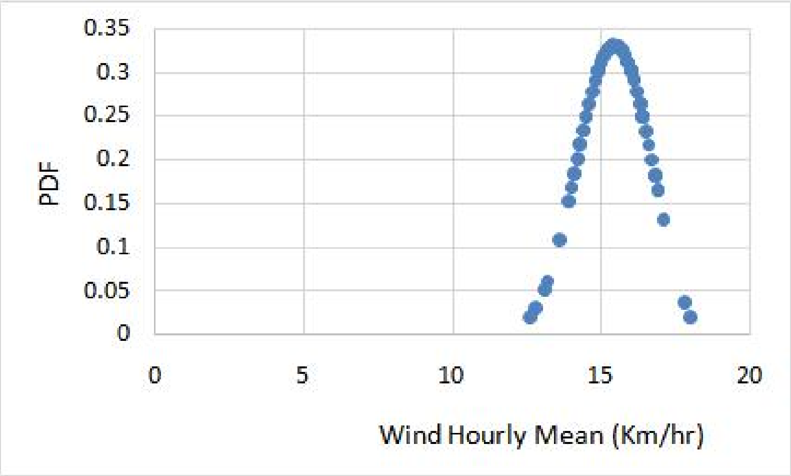


Figure 8: Probability Density Function for Wind Hourly Mean for Toronto City.

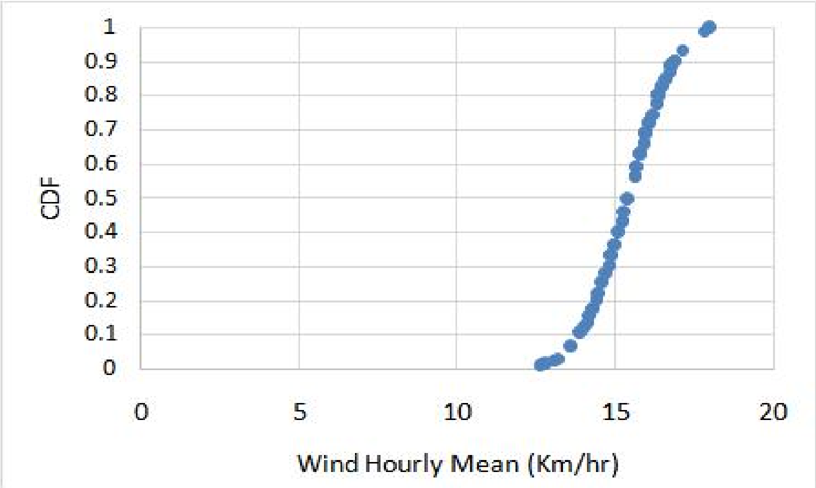


Figure 9: Cumulative Distribution Function for Wind Hourly Mean for Toronto City.

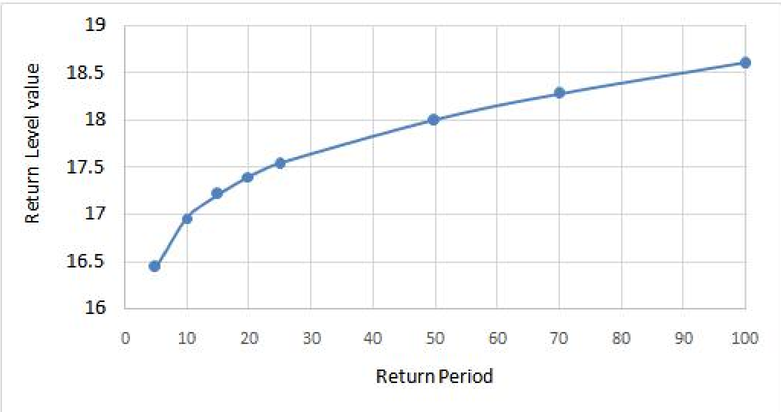


Figure 10: Return Level for Wind Hourly Mean (Non-Stationary).

CONCLUSIONS

The historical data for minimum temperature, hourly mean temperature and wind hourly mean were gathered annually from Canada weather statistics. In addition, the descriptive statistics for different climate data are concluded from the past observation data. The KPSS and ADFtestswere conducted for the time series analysis in order to test the observed time series either it is stationary or nonstationary. The minimum temperature, hourly mean temperature and wind hourly mean are nonstationary data for Toronto city in Canada according to KPSS and ADF test. Generalized extreme value theory was applied on data and Weibull distribution is fitted for the data of minimum temperature, hourly mean temperature and wind hourly mean. In addition, Predicting the return level values for next 50 years for minimum temperature, hourly mean temperature and wind hourly mean are -10.29 degrees Celsius, 12.36 degrees Celsius and 18 Km/hr respectively using generalized extreme value theory for Toronto city which is modelled as non -stationary data. Furthermore, Predicting the return level values for next 100 years for minimum temperature, hourly mean temperature and wind hourly mean are -5.7 degrees Celsius, 14.5 degrees Celsius and 18.6 Km/hr respectively using generalized extreme value theory for Toronto city which is modelled as nonstationary data.

ACKNOWLEDGMENTS

The authors would like to acknowledge the National Research Council Canada (NRCC) in partnership with Ryerson University, which have funded this and related research.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. *M. G. Stewart and X. Deng. Climate Impact Risks and Climate Adaptation Engineering for Built Infrastructure, American Society of Civil Engineers, 1(2014), DOI: 10.1061/AJRUA6.0000809.*
2. *M. S. Ali and M. S. Mirza. Impact of Climate Change on Civil Infrastructure – A State-of-the-Art Review, EIC Climate Change Technology Conference, (2015).*
3. *S. Coles. An Introduction to statistical Modeling of extreme values, Department of Mathematics, University of Bristol springer series in statistics, Springer Series in Statistics, 2001. 45-73.*
4. *K. W. Hipel and A. I. McLeod. Time Series Modelling of Water Resources and Environmental Systems, Elsevier,45, 45 (1994). eBook ISBN: 9780080870366.*
5. *J. H. Cochrane. Time Series for Macroeconomics and Finance, Graduate School of Business, 1997, University of Chicago.*
6. *M.I. Ullah. Component of time series, Basic statistics and Data analysis. Statistics Lecture Notes, Online MCQs, 2020. https://itfeature.com/component-of-time-series.*
7. *G. P. Zhang. A neural network ensemble method with jittered training data for time series forecasting. Information Sciences, 177 (2007), 5329–5346.DOI:10.1016/j.ins.2007.06.015*
8. *M. G. Stewart, D. V. Val, E. B. Arteaga, A. O. Connor, X. Wang. Climate Adaptation Engineering and Risk Based Design and Management of Infrastructure, In Maintenance and Safety of Aging Infrastructure, Chapter 21(2014).641-684, DOI: 10.1201/b17073-22.*
9. *M. Leclerc, and T. Ouarda. Non-stationary regional flood frequency analysis at ungauged sites. Journal of Hydrology, 343 (2007), 254–265, DOI:10.1016/j.jhydrol.2007.06.021*
10. *R. Reiss and M. Thomas. Statistical analysis of extreme values: with Applications to Insurance, Finance, Hydrology and Other Fields, ISBN 978-3764372309, New York. 2007.*
11. *S. E. Adlouni, T. B. M. J. Ouarda, X. Zhang, R. Roy, B. Bobe'e. Generalized maximum likelihood estimators for the nonstationaryno stationary generalized extreme value model, Water Resources Research, 43 (2007), doi:10.1029/2005WR004545.*
12. *C. Caroni, D. Panagoulia, P. Economou. Stationary and no stationary generalized extreme value modelling of extreme precipitation over a mountainous area under climate change, Environmetrics, 25 (2015), 29-43, https://doi.org/10.1002/env.2252.*

13. E.J. Gumbel. Statistics of Extreme, Columbia University Press, New York, NY, USA, ISBN: 0231021909978 0231021906. 1958.

14. Fréchet. M. Sur la loi de probabilité de l'écart maximum, Ann. Soc. Polon. 6 (1927), 93–116.

15. W. Weibull. A statistical distribution function of wide applicability, J. Appl. Mech.-Trans, ASME, 18 (1951), 293–297.

16. A.Ferreira, and L. de Haan. Extreme value theory, Springer, (2006), 37-63. https://doi.org/10.1007/0-387-34471-3

17. J. Galambos. The asymptotic theory of extreme order statistics, Wiley, New-York, (1st ed.), Krieger, Florida (2nd ed.), 1985.

18. D. Kwiatkowski, P. C. B. P. Phillips, Schmidt, Shin Y. et al. Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root, Journal of Econometrics, 54 (1992), 159–178.

19. J.D. Hamilton. Time Series Analysis, Princeton, NJ: Princeton University Press, ISBN 0-691-04289-6,1994.

20. W.K.Newey, and K. D. West. A Simple Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica, 55 (1987),703–708.https://doi.org/10.2307/1913610.

21. D. Cooley. Return Periods and Return Levels Under Climate Change, Extremes, 65(2013),97-114.DOI: 10.1007/978-94-007-4479-0_4.

22. L. Cheng, A. Agha Kouchak, E. Gilleland, R. W. Katz, et al. Non-stationary extreme value analysis in a changing climate, Climatic change, 127(2014), 353-369.https://doi.org/10.1007/s10584-014-1254-5.