ISSN: 2835-7914

Research Article

Biomedical Science and Clinical Research

Use of Segmented Linear Regression Under a Bayesian Approach to Detect Climate Change in Different Regions of the World

Emerson Barili*, Jorge Alberto Achcar

Medical School, University of São Paulo, Ribeirão Preto, São Paulo, Brazil

***Corresponding Author**
Emerson Barili, Medical School, University of São Paulo, Ribeirão Preto, São Paulo, Brazil.

Submitted: 16 Feb 2023; **Accepted:** 13 Mar 2023; **Published:** 20 Mar 2023

Citation: Barili, E., Achcar, J.A. (2023). Use of Segmented Linear Regression Under a Bayesian Approach to Detect Climate Change in Different Regions of the *World. Biomed Sci Clin Res*, 2(1), 149-158.

Abstract
In this work we study the behavior of some climate data (annual temperature and precipitation averages) obtained from climate stations in eleven countries in different regions of the world. One of the goals of the study is to determine whether the climate variables have change-points that could indicate the possible beginning of a change in climate. Another goal is to analyze the possible changes detected by the change-points in terms of the linear trends of the climate variables under investigation. Based on the information provided, differences between different regions in terms of the locations of the change-points and the changes they produce may also be inferred. The data sets used in the study consist of the annual averages of the twelve monthly temperature averages and the annual averages of the total rain precipitation observed in each one of the twelve months of the year obtained over a period of time from the end of the 19th century to the end of the 20th century. Segmented linear regression models are used to study the existence of possible changes in the behavior of climatic variables, as well as the types of changes produced.

Keywords: Climate Data, Segmented Linear Regression Models, Change-Points, Annual Temperature And Precipitation Average

Introduction

In the past few decades, different effects of climate change, particularly changes in temperature and precipitation, have been observed around the world. These effects include shrinkage of glaciers, earlier melting of ice in rivers and lakes, changes in plant and animal habitats, and earlier flowering of trees. As pointed out by the Intergovernmental Panel on Climate Change, when we look at the data collected by different climate agencies around the world, it is also possible to observe a loss of sea ice, accelerated sea level rise, and worldwide longer and more intense heat waves (i.e., periods of unusually hot weather lasting from days to weeks).

Human-induced temperature and precipitation changes have not been uniform over time, across the planet, and even in different regions of the same country. For instance, the average rainfall in the United States of America (USA) has increased since 1900, but some areas of the country experienced increases greater than the national rainfall average while other areas had increases that were lower than the national average. Throughout the twenty-first century, precipitation during winter and spring is projected to be higher for the northern part of the USA and lower for the southwest. Additionally, for many regions of the planet there are predictions that heat waves will become more intense and cold waves will be less intense everywhere.

Another example is changes occurring in the global temperature. For instance, the global average temperature remained at around 13◦C (56.7◦F) throughout the nineteenth century and the first decade of the twentieth (http://www.currentresults.com/

Environment-Facts/changes-in-earth-temperature.php, accessed on 01 July 2021). Between 1920 and 1940, the average temperature increased approximately 0.1◦C (0.18◦F) per decade. Even though during the 1980s the average temperature stabilized around 14◦C (57.2◦F), since then, it has mainly increased at a rate of approximately 0.2◦C (0.36F) per decade. From 2000 to 2009 the annual average global temperature has been 0.61◦C (1.1◦F) higher than the period ranging from 1950 to 1980. If this rate of increase is maintained, the average world temperature is expected to increase in the twenty-second century an additional 2◦C (3.6◦F) from its twenty-first century average.

For all the reasons mentioned so far, it is therefore important to study the changes occurring in the global temperature and precipitation averages. This becomes even more important due to the different behaviors of the changes in different parts of the world, since some regions may be more greatly affected than others.

In recent years, we have seen the publication of a large number of scientific works related to climate change events, such as, change in precipitation, temperature, sea levels, among others, and their implications. For instance, [1] study the impact of climate change on water resources and flooding; [1, 2] deal with the relation between climate change and health effects; [3] perform an analysis of the changes in global temperature considering information since the preindustrial era; [4] study the relationship between global warming and climate change; [5] analyze the impact of climate change on the coastal areas of Bangladesh; [6] deal with sea level changes in connection to global warming; [7]. study the impact of climate change on migration; [8]. describes the developments in the understanding of how temperature and humidity have changed; [9] study the impact of climate change on the marine life; [10] analyze the impact of climate change on the sub-Saharan Africa; [11] study the health effects of future food production under climate change [12]. analyze the threat posed on ecosystems by climate change; [13]. present an analysis of the relationship between temperature increase and crop production; [14] analyze the changes in extreme temperature data; [15] study the joint change in temperature and precipitation from multiple climate models using a Bayesian point of view.

Using the annual temperature and precipitation averages (annual averages of the twelve monthly temperature averages and the annual averages of the observed total rain precipitation in each month) collected in eleven climate stations from different continents of the world from the late nineteenth-century until recent years, one of the goals of this study is to identify by statistical analysis, possible changes in the behavior of the global annual temperature and precipitation averages. Possible changes in climatic variables will be identified by detecting change points using the statistical model. Another objective is to analyze what types of changes are observed in the linear trends of the two assumed climatic variables. Among the many works that discuss the presence of change-points in time series are the studies of [16,17] where Bayesian analyses of change-point problems are presented; [16-18] with an analysis of the presence of change-points in auto-regressive series; [19-20], in which change-points detection is also applied to temperature and precipitation data; and, in which a genetic algorithm is used to detect change- points with an application to climate data [21]. In the present work, the estimation of the location of the possible change-points and the changes in behavior they might register will be performed using segmented linear regression models. Additionally, once the possible change-points are identified, an analysis is performed in order to see whether the changes they detect are statistically significant.

Segmented regression (piecewise regression) models are used when the dependent variable is analyzed through partitions of the time interval and a separate line segment is fitted to each one of them. Segmented regression is useful in climate data, since the response (dependent) variables are clustered in different time periods. This allows the analysis of the different relationships between the response variables and the time (considered as an explanatory variable). The boundaries between the segments are known as change-points. In general, in segmented regression, the point in the explanatory variable at which the functional form of the response variable changes is of great interest in climate change studies. This importance resides in the fact that this point in time may represent a shift in the climate behavior at a specified region of the world. Some models consider that adjacent regression segments meet at the change-point and others assume a discontinuity. The latter will be considered in the present work.

Since models in the presence of change-points usually have some complex forms for the likelihood function, many authors consider a Bayesian approach when estimating the parameters present in those models. In particular, the Markov chain Monte Carlo (MCMC) methods are useful in the analysis (see, for example [22-26]). A Bayesian approach is also considered in this study in order to estimate the regression parameters, the change-points, and the variances of the error terms present in the regression model. Other statistical models could also be used in the analysis of climate data to obtain the inferences of interest; see, for instance [27-31].

This work is organised as follows. Section 2 presents the model considered in the analysis. Section 3 gives the methodology used to estimate the parameters. Section 4 presents a brief description of the data used. In Section 5, we give the results obtained. Finally, in Section 6 we present a discussion of the results and their implications. This work also contains one supplementary information files, presented as online resources, in which further information regarding the model and estimation of the parameters are given.

Description of the Model

The segmented regression models used in the present work will allow the presence of one or more change-points. These change-points may represent changes in the linear trend behavior of the annual temperature and precipitation averages data collected from several climate stations in different parts of the world.

To begin describing the model, let $T > 0$ be an integer indicating the number of observed values in a given data set, i.e., the length of the follow-up period (in the present case, T represents the number of years of climate series monitoring). Let $Z(t)$, $t = 1, 2, \dots, T$, indicate the elements of this data set. When we use linear regression models and the dependent variable $Z(t)$ is different of zero from all t, $t = 1, 2, \dots, T$, we may assume transformations of $Z(t)$, $t = 1, 2, \dots, T$, to satisfy the assumptions required for the usual linear regression models (i.e., normality and constant variance of the so-called residuals). In our case, we assume a log-transformation of $Z(t)$, i.e., we define $Y(t) = \log Z(t)$, t = 1, 2, . . . , T. Hence, $Y = \{Y(1), Y(2), \dots, Y(T)\}$ will denote the observed data. In the model considered here, we have the time as a covariate (explanatory variable) indicated by $R(t) = t$, $t = 1, 2, \dots, T$. When there is only one change-point, denoted by τ_1, we could assume the model,

$$Y(t) = \begin{cases} \alpha_1 + \beta_1 t + \epsilon, & t \le \tau_1, \\ \alpha_2 + \beta_2 t + \epsilon, & \tau_1 < t \le T, \end{cases} \quad (1)$$

where α_i and β_i, $i = 1, 2$, are the regression parameters, and ϵ is the error parameter assumed to have normal distribution with mean zero and variance σ^2, i.e., $\epsilon \sim N(0, \Sigma)$. If the regression is continuous, we need to have $\alpha_1 + \beta_1 \tau_1 + \epsilon = \alpha_2 + \beta_2 \tau_1 + \epsilon$. Replacing the expression for α_2 in (1) gives the model, $Y(t) =$ $\alpha_1 + \beta_1 t + \epsilon$, $t \le \tau_1$,

$$Y(t) = \begin{cases} \alpha_1 + \beta_1 t + \epsilon, & t \le \tau_1, \\ \alpha_1 + \beta_1 \tau_1 + \beta_2 (t - \tau_1) + \epsilon, & \tau_1 < t \le T. \end{cases} \quad (2)$$

Nonlinear least squares regression techniques are usually used to fit model (2) for the data set. RETIRAR

The approach followed here differs from (2). We consider a simpler linear regression model borrowed from the BUGGS manual (see, for instance, http://pmean.com /11/segmented. html, accessed on 13 September 2021). If we assume the possible presence of $J \geq 1$ change-points, then the general form of the model may be described, for j = 1, 2, . . . , J, as,

$$Y(t) = \begin{cases} \alpha_j + \beta_{j1}\,[R(t) - \tau_j] + \epsilon_j, & \tau_{j-1} < t \leq \tau_j, \\ \alpha_j + \beta_{j2}\,[R(t) - \tau_j] + \epsilon_j, & \tau_j < t \leq T, \end{cases} \tag{3}$$

where we take $\tau_0 = 0$, and $Y(t)$ denotes the climate response of interest at time t (mean temperature/precipitation) in logarithmic scale; $R(t) = t$ is the independent variable (time measured in years here) defined on the interval [1, T]; $\tau = (\tau_1, \tau_2, \ldots, \tau_J)$ is the vector of change-points; the error terms ϵ_j are considered unobserved random variables with a normal distribution $N(0, \sigma^2)$; and α_j, β_{j1}, and β_{j2}, $j = 1, 2, \ldots, J$, are the regression coefficients. (Note that the variance σ^2 may not be the same for a values of j.)

Hence, when we have J change-points, the parameters to be estimate are α_j, β_{j1}, β_{j2}, σ^2, and τ_j, $j = 1, 2, \ldots, J$, with the estimation performed sequentially. That is, we start by considering the first change-point τ_1. In this case, we will need to estimate α_1, β_{11}, β_{12}, σ_1, and τ_1. Once τ_1 is located and the remaining parameters are estimated, we keep the estimated regression formula for all $t \leq \tau_1$. If there is no need for a second change-point, then we also keep the estimated regression formula for $\tau_1 < t \leq T$. Suppose there is the need of a second change-point. Then, we consider the covariate $R(t)$ defined only in the interval (τ_1, T) and use (3) with $j = 2$. In this case, the parameters to be estimated are α_2, β_{21}, β_{22}, σ_2, and τ_2. If there are no more change-points, then we keep this estimated regression in the interval (τ_1, T). If there is a third change-point, then we keep the estimated regression in the case of $j = 2$ in the time interval (τ_1, τ_2) and consider the covariate $R(t)$ defined now on (τ_2, T). The procedure is repeated for the case $j = 3$, and the parameters are estimated accordingly and successively unti we find all possible change-points.

Estimation of the Parameters

The vector of parameters of the model is estimated using the Bayesian approach. Hence, the information provided by its posterior distribution will be used. If θ is the vector of parameters of a model describing a given set **D** of observed data, then its posterior distribution, denoted by $\mathbf{P}(\theta \mid \mathbf{D})$, is such that $P(\boldsymbol{\theta} \mid \mathbf{D}) \propto L(\mathbf{D} \mid \theta) P(\theta)$, where $L(\mathbf{D} \mid \theta)$ is the so-called likelihood function of the model and $P(\theta)$ is the prior distribution of θ. (Here, "∝" indicates "proportional to.") Therefore, we need to specify the likelihood function of the model and the prior distribution of the vector of parameters.

The Bayesian methodology used in this work combines the prior information, usually elicited from climate experts, with the likelihood function to obtain the joint posterior distributions of all parameters. This replaces the usual maximum likelihood inference approach, which uses classical inference methods.

The likelihood function of the model at the step whose parameters are α_j, β_{j1}, β_{j2}, σ_j, and τ_j, is given by,

$$L(\mathbf{Y} \mid \alpha_j, \beta_{j1}, \beta_{j2}, \sigma_j^2, \tau_j) = \prod_{t=\tau_{j-1}+1}^{\tau_j} \frac{1}{\sqrt{2\,\pi\,\sigma_j^2}} \exp\left(-\frac{1}{2\sigma_j^2}\,[Y(t) - \alpha_j - \beta_{j1}\,(t-\tau_j)]^2\right)$$
$$\times \prod_{t=\tau_j+1}^{T} \frac{1}{\sqrt{2\,\pi\,\sigma_j^2}} \exp\left(-\frac{1}{2\sigma_j^2}\,[Y(t) - \alpha_j - \beta_{j2}\,(t-\tau_j)]^2\right). \tag{4}$$

The other elements to be specified are the prior distributions of the parameters. Hence, at the jth step ($j = 1, 2, \ldots, J$), the regression parameters α_j, β_{j1}, and β_{j2} will have normal prior distributions $N(0, a)$ with possibly different values for the hyperparameter a for different parameters (a is taken sufficiently large in order to have approximately non-informative prior distributions); the parameter $1/\sigma^2$ is assumed either with a gamma Gamma(a_1, a_2) prior distribution with mean a_1/a_2 and variance $a_1/(a_2)^2$ or with a uniform distribution defined on an appropriate interval. The first change-point is assumed with a uniform prior distribution $U(0, T)$ (an approximately non-informative prior). The second change-point will have uniform prior distribution defined on $(\tau 1, T)$. In general, the jth change-point will have its prior distribution a uniform distribution defined on (τ_{j-1}, T), $j = 1, 2, \ldots, J$. The specific values of the hyperparameters are given in the file Online-Resource-1.pdf given as supplementary information.

In practical work we could use more informative prior distributions, especially for the change-points, by observing the plots of the climate time series and seeing the possible candidates for the change-points.

This approach leads to better inference results and guarantees convergence of the MCMC simulation algorithm.

We further assume prior independence of the parameters.

Once the regression coefficients are estimated and the possible change-points are located, an analysis is performed in order to assess whether the changes detected by the change-points are statistically significant. As a consequence, information related to the significance of the change-points is also provided. This analysis is performed by constructing the 95% credible intervals of the differences $\Delta^{(j)} = \beta_{j2} - \beta_{j1}$, where β_{j1} and β_{j2} are, respectively, the slopes of the linear regression models before and after the jth change-point. If zero is part of a 95% credible interval, then that is an indication that the two slopes are not significantly statistically different. Hence, we also use a Bayesian inference approach to see whether an estimated change- point detects significantly statistically different linear trends. This approach replaces standard hypothesis tests usually used in classical inference to test whether two parameters are equal. The values used to produce the 95% credible intervals are those generated by the MCMC algorithm.

Data Sets Used

The data sets considered in this study consist of the annual averages of the twelve monthly temperature averages and the annual averages of the total rain precipitation observed in each one of the twelve months of the year calculated from the monthly temperature averages and monthly total rain precipitation values extracted from the Research Data Archive managed by the Data Engineering and Curation Section of the Computational and Information Systems Laboratory at the National Center for Atmospheric Research, United States of America. This site contains a large and diverse collection of meteorological and oceanographic observations, operational and reanalysis model outputs, and remote sensing data sets to support atmospheric and geosciences research (see, for example, https://rda.ucar.edu/index.html?hash=data user& action=register and https://rda.ucar.edu/datasets/ds570.0/#!subset.html, both accessed on 01 July 2021). It contains data from more than 4700 different climate stations (2600 in more recent years) from all around the world.

Different follow-up periods are given for the different climate stations, and collection of data for some of them goes as far back as the mid-1700s.

The primary data sets consist of monthly average temperature and total monthly rain precipitation. Since the data sets have many missing observations (months with no data), these were replaced by the monthly averages of the available data for that month. For instance, if in a given year we have missing data for the month of April, we fill the hole in the data by assigning to that month the mean obtained using the values for the month April from all the years in which they are available.
The data used in our calculations were the annual temperature and precipitation averages.

In this study, we have considered data from eleven climate stations in countries located in the North and South America, Europe, Asia, and Oceania extracted from the file ds570 of the Research Data Archive.

The climate stations, observational periods, the number T of observed data are: Station 30910 in Aberdeen, United Kingdom (UK) (1872–2020; $T = 149$); Station 16 67000 in Geneva, Switzerland (1826–2020; $T = 195$); Station 171300 in Ankara, Turkey (1826–2020; $T = 195$); Station 276120 in Moscow, Russia (1881–2020; $T = 140$); Station 416400 in Lahore, Pakistan (1876–2020; $T = 145$); Station 428070 in Calcutta, India (1878–2020; $T = 143$); Station 476620 in Tokyo, Japan (1876–2020; $T = 145$); Station 607150 in Tunis, Tunisia (1896– 2020; $T = 124$); Station 722080 in Charleston, USA (1832–2020; $T = 189$); Station 837810 in São Paulo, Brazil (1887–2019; $T = 133$); and Station 945780 in Brisbane, Australia (1887–2020; $T = 134$).

Figures 1 and 2 show plots of the annual temperature averages and Figures 3 and 4 present plots of the annual precipitation averages. All values are the logarithms of the measurements reported at the different climate stations during their corresponding observational periods. Looking at these figures, we see the possible presence of change-points indicating changes from increasing/decreasing trends to decreasing/increasing.

In the case of the temperature data, Figures 1 and 2, we see that in the final years of the follow-up period, for each climate station, there is an indication of an increasing trend in the annual temperature averages. In some cases, the variations throughout the observational period are very large, for instance, in the case of Moscow in the years close to 1940 and 1980, and in Lahore at several points during the observational period. We also see large variations in the cases of Geneva, Charleston, Ankara, and Brisbane.

If we look at the precipitation data in Figures 3 and 4, we see a decrease in the annual precipitation averages in the final years of the observation period in stations such as Geneva and Brisbane. Increasing trends in the final years of the observational period are detected in the remaining stations with a more pronounced increase in some (see, for instance, the data from the station located in Tokyo).

When we look at Figures 3 and 4, we also see some patterns of large variations in the annual precipitation averages. In some stations, for instance, Aberdeen and Geneva around the year 1925, Lahore and Brisbane around 1900, Calcutta around 1940, and Charleston around 1850, we have some sudden drops in the annual precipitation averages. In others, a steady oscillation between large and small values may be detected.

As we see in Figures 1, 2, 3, and 4, the annual temperature and precipitation averages behave differently for different stations located in different parts of the world. Therefore, analyzing the trends in these measurements is important for determining some of the consequences of climate changes in particular parts of the globe.

Results Obtained from the Segmented Regression Models

We see in Figures 1, 2, 3, and 4 that in all data sets, it is possible to detect at least one possible change-point. Hence, depending on the data set, we assume the presence of either one, two, or three change-points. The analysis is split into the two cases of temperature and precipitation. We start with the estimation of the regression parameters and change-points in the case of the temperature and then proceed to the case in which the precipitation data are analyzed. Once an estimation of the parameters is made, we proceed to analyze the significance of the detected change-points. Estimation of the parameters will be performed using the OpenBugs software (Spiegelhalter et al. 2007). Convergence of the MCMC algorithm is monitored using the traceplots of the generated Gibbs samples.

Average Temperature

When we look at Figures 1 and 2, we see that when data from Aberdeen and Geneva are used, perhaps we should allow the possible presence of two change-points. For the remaining stations one change-point might suffice. The corresponding prior distributions of the parameters are given in Table 1 in the file Online-Resource-1.pdf. Estimation of the parameters was made using a sample of size 1000 in the case of Moscow, 4000 in the case of Ankara, and 11000 in the case of Geneva (first change-point). The sample size when we consider data from Aberdeen (second change point), Calcutta, Tokyo, Tunis, and Brisbane is 5000, and it is 6000 in the remaining cases. In order to have approximately uncorrelated values, samples were obtained using values taken every 100th iteration of the algorithm. These were taken after burn-in periods of 911000 and 611000 steps in the case of Geneva's first and second change-points, respectively.

In the remaining cases, the burn-in period was 411000 iterations. Table 2 in the file Online-Resource-1.pdf gives the values of the estimated regression coefficients, variances, and change- points. Looking at that Table 2 we see that the estimated change-points in the case of stations in Ankara, Moscow, Lahore, Calcutta, Tokyo, Charleston, Tunis, and São Paulo are as follows. Those corresponding to Moscow and Lahore are located in the years 1972 and 1976, respectively; those related to Ankara and Tokyo are both in the year 1987; in Tunis and Charleston they are located in 1954 and 1966, respectively; in Calcutta it is located in the year 1965; and that of São Paulo is located in 1914. The two estimated change-points related to the data from stations in Aberdeen and Geneva are as follows. Those corresponding to Aberdeen are located in the years 1927 and 1949. In the case of Geneva, they are located in the years 1968 and The estimated change-point corresponding to the station in Brisbane is 2012.

The estimated segmented regression models for the annual mean temperatures are shown in Figures 1 and 2. We observe in those figures that the estimated segmented regressions capture very well the annual averaged temperature trends. We also see that the increasing trends in the average annual temperatures after the last change-points are captured accurately by the model.

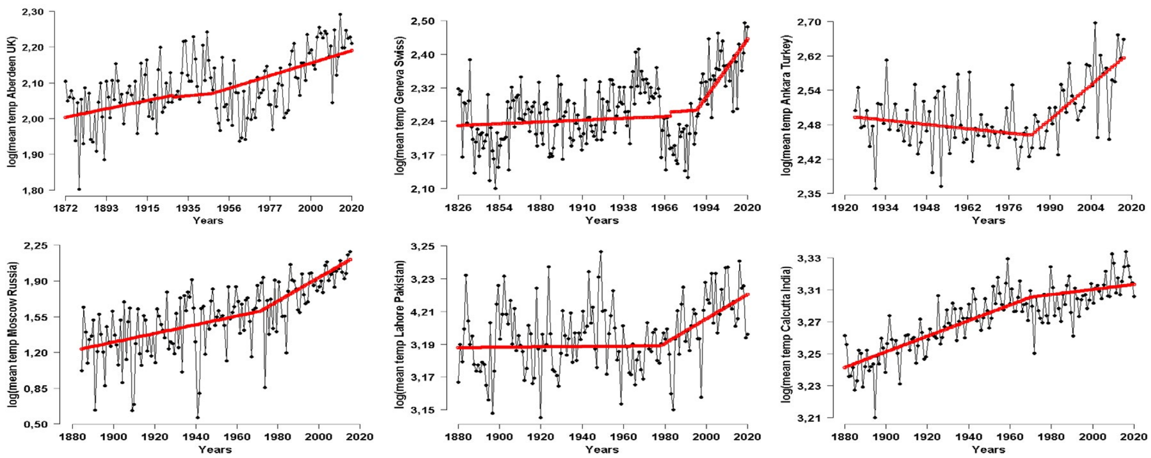

Figure 1: Annual average temperature and fitted segmented linear regressions (Aberdeen, Geneva, Ankara, Moscow, Lahore, and Calcuta). The jagged lines represent the average temperature, while the heavy lines are the fitted linear regressions.

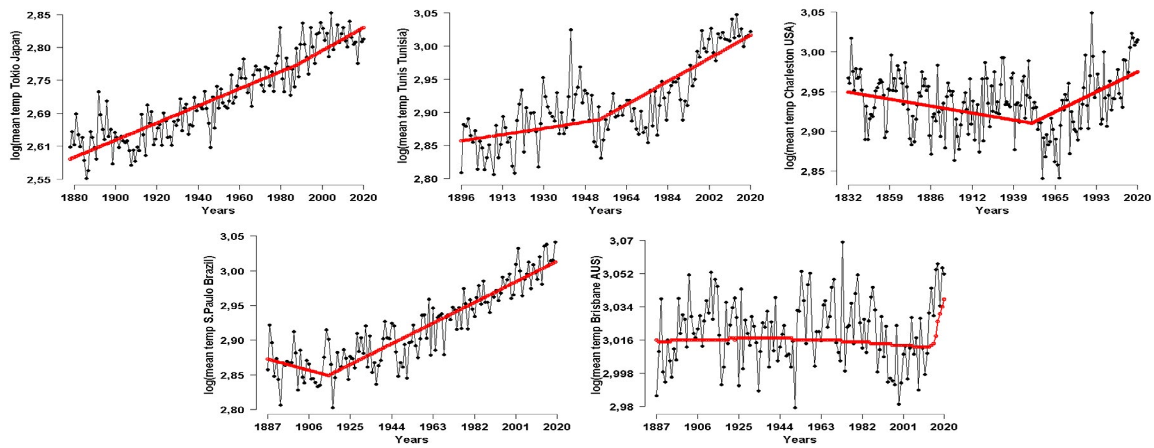

Figure 2: Annual average temperature and fitted segmented linear regressions (Tokyo, Tunis, Charleston, S˜ao Paulo, and Brisbane). The jagged lines represent the average temperature, while the heavy lines are the fitted linear regres- sions.

Averaged Precipitation

When we look at the case of annual precipitation averages given by Figures 3 and 4, we see the possible presence of two change-points in the case of Moscow and Calcutta; in the case of Brisbane, we have three possible candidates; and in all the remaining stations, we detect only one possible change-point.

The prior distributions of the regression parameters, errors variances, and change-points are given in Table 3. In all cases, estimation of the parameters was performed using a sample of size 5000 collected every 100th iteration of the MCMC algorithm after a burn-in period of 411000 iterations.

In Table 4 in$_i$ the file Online-Resource-1.pdf file we have the Bayesian estimates of the regression parameters α_j , β_{j1}, and β_{j2}, of the variance σ_2, $j = 1, 2, \ldots, J$, as well as the corresponding estimated change-points. Looking at this table, we see that the change-points corresponding to data from stations in Aberdeen, Geneva, Ankara, Lahore, Tokyo, Tunis, Charleston, and São Paulo are as follows. Those of São Paulo and Aberdeen are located, respectively, in the years 1889 and 1886; and that corresponding to the station in Ankara is located in 1935. Lahore and Tunis have change-points located, respectively, in the years 1906 and 1924. The change-points corresponding to Tokyo and Geneva are located in the years 1977 and 2008, respectively. Stations located in Moscow, Calcutta, and Charleston that have two possible change-points have them located as follows. In Charleston, we have them in the years 1843 and 1895; in Moscow, in 1900 and also in 1936. When we look at the station in Calcutta, we see that the change-points are in the years 1943 and 2015. The three change-points occurring in Brisbane are located in the years 1900, 1925, and 1970.

The estimated segmented regression models are shown in Figures 3 and 4.

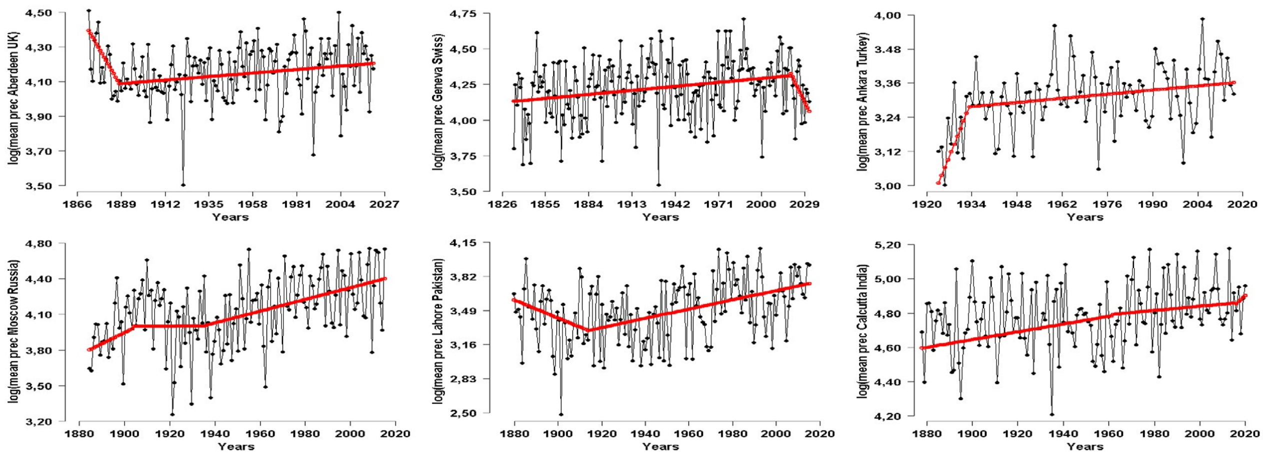

Figure 3: Annual average rain precipitation and fitted segmented linear regressions (Aberdeen, Geneva, Ankara, Moscow, Lahore, and Calcuta). The jagged lines represent the average rain precipitation, while the heavy lines are the fitted linear regressions.

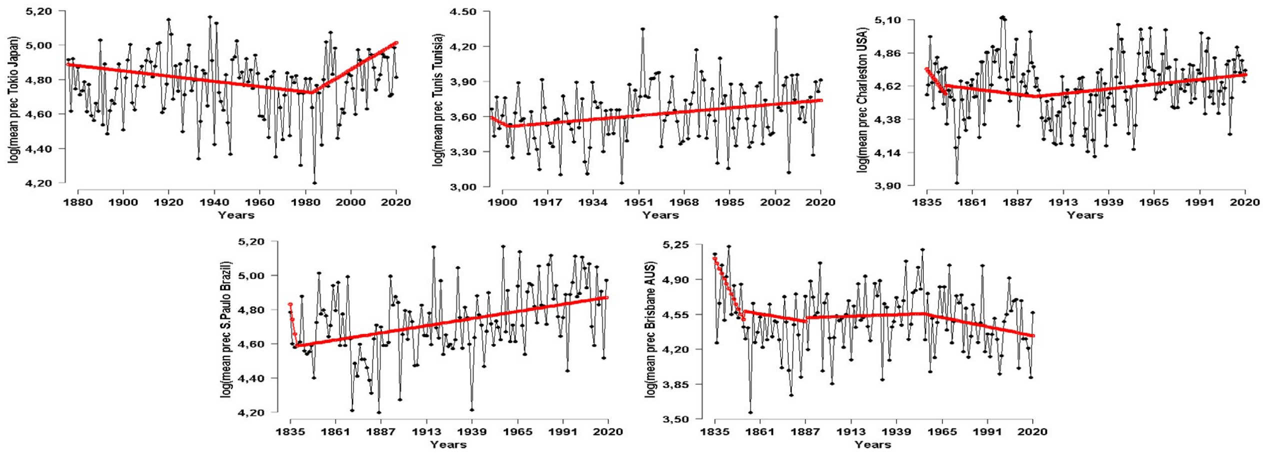

Figure 4: Annual average rain precipitation and fitted segmented linear regressions (Tokyo, Tunis, Charleston, S˜ao Paulo, and Brisbane). The jagged lines represent the average rain precipitation, while the heavy lines are the fitted linear regressions.

When we look at those figures we see that the estimated linear regressions capture very well the trends of the observed data. In the present case, we also have that the steep changes in the trend are captured as well; see for instance, the final years in Geneva and Calcutta, where, respectively, a steep decrease and increase in the annual averaged precipitation occurred. The steep decrease and increase in the beginning of the observational period in the data from the stations in Brisbane and Ankara, respectively, were also captured accurately.

Discussion

In this work we have used a simple segmented linear regression model to estimate the linear trends as well as the points in time where changes in the trends of the annual temperature and precipitation averages might have occurred. We could see from Figures 1, 2, 3, and 4 that, in general, the changes in the trends as well as the change-points were estimated accurately.

When we look at Figures 1 and 2, we see that, in all stations, trends indicating an increase in the annual mean temperature in the last decades are present. These linear trends are captured by the model in all cases. Additionally, changes that occurred throughout the observational period, again with few exceptions–see, for instance, Geneva–were also captured by the model. If we turn our attention to Figures 3 and 4 where we have the plots related to the average annual precipitation data and the corresponding estimated linear regressions, we observe that results and data variations are more heterogeneous than those in the case of the average annual temperature.

In Table 1 we present a summary of the types of changes in the estimated linear trends of the annual temperature and precipitation averages detected by the estimated change-points.

Once the locations of the possible change-points have been estimated, we need to assess the significance of the detected changes. The estimated 95% credible intervals of $\Delta^{(j)}$ are given in Table 2. Those intervals not containing zero indicate that the changes detected by the corresponding change-points as well as change-points themselves are significant.

We also see in that table that in the case of the annual temperature averages, the changes detected by the change-points in the data from Lahore, Calcutta, Tokyo, and the first change-point in the Aberdeen data, are deemed not significant. That means that even though the estimated values of the slopes β_{j1} and β_{j2} are different, this difference is not statistically significant.

Hence, these detected change-points are not considered significant, as well as the changes detected by them. Notice that, of these non-significant changes only in Lahore did the values of β_{j1} and β_{j2} produce changes in the trend (from decreasing to increasing); in the other cases, the changes were only in the values of estimated parameters. When we look at the results given by the data from the remainder of the stations, the changes detected by the estimated change-points as well as the change-points themselves may be considered statistically significant.

When we consider the case of the annual precipitation averages, we may see that with the exception of the data from the stations in Aberdeen, Ankara, and Lahore, all the changes detected by the change-points are deemed statistically non-significant.

Focusing only on the statistically significant detected changes, we have that in the case of the annual temperature averages, we see a change from decreasing trends to increasing in the data from Ankara, São Paulo, Charleston, and Brisbane. In the remaining cases, the changes are only in the value of the slope. The behavior is kept the same (see, for instance, the cases of Moscow, Tunis, Geneva, and Aberdeen second change-point). If we consider the cases of Moscow and Tunis, the values of β change from smaller to larger, hence producing a steeper trend increase after the change-point (see Figure 1) and, therefore, a more rapid increase in the annual temperature averages. In the case of Geneva, the values of the slopes also change from smaller to larger producing steeper and steeper increasing trends as we move across the change-points. In the case of Aberdeen, the changes are also from smaller to larger slopes, producing a steeper increasing trend after the second estimated change-point.

Table 1: Types of behaviors detected by the change-points for each station and response variable. Here, ↓ and ↑ indicate a decreasing and an increasing trend, respectively.

Station	Temperature	Precipitation
Ankara	(↓, ↑)	(↑, ↑)
Moscow	(↑, ↑)	(↑, ↑, ↑)
Lahore	(↓, ↑)	(↓, ↑)
Calcutta	(↑, ↑)	(↑, ↑, ↑)
Tokyo	(↑, ↑)	(↓, ↑)
Tunis	(↑, ↑)	(↓, ↑)
Charleston	(↓, ↑)	(↓, ↓, ↑)
São Paulo	(↓, ↑)	(↓, ↑)
Aberdeen	(↑, ↑, ↑)	(↓, ↑)
Geneva	(↑, ↑, ↑)	(↑, ↓)
Brisbane	(↓, ↑)	(↓, ↓, ↑, ↓)

Table 2: 95% credible intervals for the differences $\Delta^{(j)} = \beta_{j2} - \beta_{j1}$ of the fitted linear regression models delimited by the corresponding change-points.

	Temperature		Precipitation	
Station	change-point	95% Cred. Int	change-point	95% Cred. Int
Ankara	$\tau_1 = 1987$	$(-7.113\text{E}-03;\ -3.05\text{E}-03)$	$\tau_1 = 1935$	$(1.59\text{E}-03;\ 0.1951)$
Moscow	$\tau_1 = 1970$	$(-1.685\text{E}-02;\ -1.303\text{E}-04)$	$\tau_1 = 1900$	$(-6.645\text{E}-03;\ 2.327\text{E}-02)$
	–	–	$\tau_2 = 1936$	$(-6.517\text{E}-03;\ 4.543\text{E}-04)$
Lahore	$\tau_1 = 1976$	$(-1.582\text{E}-02;\ 1.081\text{E}-04)$	$\tau_1 = 1906$	$(-5.4575\text{E}-02;\ -4.991\text{E}-03)$
Calcutta	$\tau_1 = 1965$	$(-4.226\text{E}-04;\ 1.086\text{E}-03)$	$\tau_1 = 1943$	$(-5.577\text{E}-02;\ 5.969\text{E}-03)$
	–	–	$\tau_2 = 2015$	$(-0.1116;\ 9.019\text{E}-02)$
Tokyo	$\tau_1 = 1987$	$(-1.842\text{E}-03;\ 9.136\text{E}-04)$	$\tau_1 = 1977$	$(-3.476\text{E}-02;\ 3.243\text{E}-03)$
Tunis	$\tau_1 = 1954$	$(-2.242\text{E}-03;\ -2.096\text{E}-04)$	$\tau_1 = 1924$	$(-0.1972;\ 0.1675)$
Charleston	$\tau_1 = 1966$	$(-2.212\text{E}-03;\ -1.191\text{E}-03)$	$\tau_1 = 1843$	$(-0.1424;\ 0.1152)$
	–	–	$\tau_2 = 1895$	$(-5.594\text{E}-03;\ 4.099\text{E}-04)$
São Paulo	$\tau_1 = 1914$	$(-3.326\text{E}-03;\ -1.663\text{E}-03)$	$\tau_1 = 1889$	$(-0.327;\ 6.6605\text{E}-02)$
Aberdeen	$\tau_1 = 1927$	$(-1.263\text{E}-03;\ 6.941\text{E}-04)$	$\tau_1 = 1886$	$(-6.348\text{E}-02;\ -4.0\text{E}-03)$
	$\tau_2 = 1948$	$(-1.941\text{E}-03;\ -3.899\text{E}-05)$	–	–
Geneva	$\tau_1 = 1968$	$(-3.404\text{E}-03;\ -1.65\text{E}-03)$	$\tau_1 = 2008$	$(-3.596\text{E}-03;\ 7.143\text{E}-02)$
	$\tau_2 = 1957$	$(-6.372\text{E}-03;\ -3.064\text{E}-05)$	–	–
Brisbane	$\tau_1 = 2012$	$(-1.106\text{E}-02;\ -2.918\text{E}-03)$	$\tau_1 = 1900$	$(-0.2552;\ 2.167\text{E}-03)$
	–	–	$\tau_2 = 1925$	$(-1.433\text{E}-02;\ 4.514\text{E}-03)$
	–	–	$\tau_3 = 1970$	$(-2.285\text{E}-03;\ 1.295\text{E}-02)$

Hence, considering the significant changes detected by the change-points and looking at the detected behavior after the last statistically significant detected change-point, we have that in all cases, an increasing linear trend is detected in the annual temperature averages.

If we turn our attention to the statistically significant changes in the case of the precipitation data, for Aberdeen and Lahore we detect changes from decreasing to increasing, while a change only in the value of the slope occurs in the case of Ankara.

Note that this change in values is from a larger to a smaller value; hence the speed at which the annual precipitation average increases decreases after the change-point, i.e., the trend is still increasing, but now at a lower speed.

Note that in the cases in which the change-points did not indicate statistically significant changes in trends of the annual averages, the estimated parameters, and consequently the fitted linear regression curve, described accurately the general trend of the data in all cases. Taking this into account, we see that in more recent years of the corresponding observational periods, there is a steeper increase in the annual temperature averages in Geneva, Ankara, Lahore, Charleston, and Brisbane.

We also have a consistent increase during the whole observational period in the case of Tokyo and for most of the observational period in the case of São Paulo. In addition to that, after the last detected change-point, we have a more gradual increase in the annual temperature averages in the cases of Calcutta and Tokyo. Therefore, even though changes were not considered statistically significant, they do represent the trend behavior of the data. For the annual precipitation averages, looking at Figure 3 and 4, we observe different behaviors for the eleven climate stations, but in general, the change-points do not indicate significant differences between the slopes of the linear regression models for most of the cases (with the exception of Ankara and Aberdeen). It is important to point out that the effects of climate change in precipitation could be more localized in time, such as occurrences of large precipitation values in short periods of time and long dry periods; hence the annual precipitation averages might not be affected significantly.

Therefore, based on the results presented here, we observe that the annual averages are changing worldwide in different forms in different parts of the world. The segmented regression model considered in this work captures fairly well those differences.

Detecting statistically significant points in time where changes in the behavior of the annual temperature and precipitation averages occur as well as the type of changes is very important, since these may affect food production, the occurrences of natural disasters, and deterioration of human health, among many other hazardous events. Additionally, if the changes are not considered significant, the estimated behavior of the data during the observational period may provide information that may be of use in implementing measures to decrease any negative impact they may produce. For instance, the series of changes in Brisbane pointing to a decrease in the annual precipitation average may produce an alert of possible occurrences of prolonged dry seasons, and this may affect food production and/or the occurrences of severe fires, causing human health problems as well as economic losses, among other nuisances, whereas as in the case of Calcutta, the steep increase in the annual precipitation average detected at the end of the observation period may also affect food production, loss of human life due to flooding, as well as economic loss and disease. On the other hand, the steep and consistent drop in the annual average precipitation in the final years of the observational period at the station in Geneva, may also produce problems.

In the case of the annual temperature averages, the consistent increase in their values at the end of all observational periods, may accelerate the melting of the ice cap at the poles as well as in glaciers present in some of the regions where data used here were collected. For instance, the consistent increase in the annual temperature averages seen in Moscow could also be happening in Siberia, and changes in that ecosystem may affect wildlife present in the region and even worldwide, as well as human health, since the melting of permafrost may result in the release into the atmosphere of greenhouse gases as well as pathogens to which humans may be susceptible.

It is important to point out that the detected change-point for the São Paulo station, from which there is a continuous increase in the annual mean temperature, could be related to global climate change. However, this city had a large population growth in the last 100 years. This increase in the population may have affected the local climate, contributing to the continuous growth in the annual temperature averages. That should be considered in addition to other factors caused by human behavior that are changing the climate of the entire planet. The more recent change-point year (1987) for Ankara could be related to the elevation of this city (894 meters), where the influence of global climate change might be not so strong.

Hence, considering the information provided by statistical analyses, and in particular the information provided by models similar to those used in the present study, governments, and the population in general, could take actions to mitigate the potential impact produced by those changes in the climate variables considered here.

The changes detected in earlier time intervals of the observational periods could serve as an indicator of what could happen if they were to repeat in the future. That could be done by taking advantage of the information we have regarding the consequences of the sudden drop in temperature averages and/or sudden increase in the precipitation averages that occurred in the past. The segmented regression models could be of use in all these cases.

References

1. Arnell, N. W., & Lloyd-Hughes, B. (2014). The global-scale impacts of climate change on water resources and flooding under new climate and socio-economic scenarios. Climatic Change 122, 127-140.
2. Costello, A., Abbas, M., Allen, A., Ball, S., & Bell, S. (2009). Bellamy, r. Friel, S., Groce, n., Johnson, A., Kett, M., lee, M., levy, C., Maslin, M., McCoy, D., McGuire, B., Montomery, H., napier, D., Pagel, C., Patel, J., Antonio, J., de oliveira, P., redclift, n., rees, H., rogger, D., Scott, J., Stephenson, J., Twigg, J., Wolff J., and Patterson, C, 1693-1733.
3. Hawkins, E., Ortega, P., Suckling, E., Schurer, A., Hegerl, G., Jones, P., ... & Van Oldenborgh, G. J. (2017). Estimating changes in global temperature since the preindustrial period. Bulletin of the American Meteorological Society, 98(9), 1841-1856.
4. Lineman, M., Do, Y., Kim, J. Y., & Joo, G. J. (2015). Talking about climate change and global warming. PloS one, 10(9), e0138996.
5. Kabir, R., Khan, H. T., Ball, E., & Caldwell, K. (2016). Climate change impact: the experience of the coastal areas of Bangladesh affected by cyclones Sidr and Aila. Journal of Environmental and Public Health, 2016.
6. Levermann, A., Clark, P. U., Marzeion, B., Milne, G. A., Pollard, D., Radic, V., & Robinson, A. (2013). The multimillennial sea-level commitment of global warming. Proceedings of the National Academy of Sciences, 110(34), 13745-13750.
7. Kaczan, D. J., & Orgill-Meyer, J. (2020). The impact of climate change on migration: a synthesis of recent empirical insights. Climatic Change, 158(3-4), 281-300.
8. Matthews, T. (2018). Humid heat and climate change. Progress in Physical Geography: Earth and Environment, 42(3), 391-405.
9. Poloczanska, E. S., Brown, C. J., Sydeman, W. J., Kiessling, W., Schoeman, D. S., Moore, P. J., ... & Richardson, A. J. (2013). Global imprint of climate change on marine life. Nature Climate Change, 3(10), 919-925.
10. Serdeczny, O., Adams, S., Baarsch, F., Coumou, D., Robinson, A., Hare, W., ... & Reinhardt, J. (2017). Climate change impacts in Sub-Saharan Africa: from physical changes to their social repercussions. Regional Environmental Change, 17, 1585-1600.
11. Springmann, M., Mason-D'Croz, D., Robinson, S., Garnett, T., Godfray, H. C. J., Gollin, D., ... & Scarborough, P. (2016). Global and regional health effects of future food production under climate change: a modelling study. The Lancet, 387(10031), 1937-1946.
12. Turner, M. G., Calder, W. J., Cumming, G. S., Hughes, T. P., Jentsch, A., LaDeau, S. L., ... & Carpenter, S. R. (2020). Climate change, ecosystems and abrupt change: science priorities. Philosophical Transactions of the Royal Society B, 375(1794), 20190105.
13. Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., ... & Asseng, S. (2017). Temperature increase reduces global yields of major crops in four independent estimates. Proceedings of the National Academy of sciences, 114(35), 9326-9331.
14. Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A. M. G., ... & Vazquez-Aguirre, J. L. (2006). Global observed changes in daily climate extremes of temperature and precipitation. Journal of Geophysical Research: Atmospheres, 111(D5).
15. Tebaldi, C., & Sansó, B. (2009). Joint projections of temperature and precipitation change from multiple climate models: a hierarchical Bayesian approach. Journal of the Royal Statistical Society: Series A (Statistics in Society), 172(1), 83-106.
16. Barry, D., & Hartigan, J. A. (1993). A Bayesian analysis for change point problems. Journal of the American Statistical Association, 88(421), 309-319.
17. Carlin, B. P., Gelfand, A. E., & Smith, A. F. (1992). Hierar-

chical Bayesian analysis of changepoint problems. Journal of the Royal Statistical Society series C (applied statistics), 41(2), 389-405.
18. Robbins, M. W., Gallagher, C. M., & Lund, R. B. (2016). A general regression changepoint test for time series data. Journal of the American Statistical Association, 111(514), 670-683.
19. Gallagher, C., Lund, R., & Robbins, M. (2012). Changepoint detection in daily precipitation data. Environmetrics, 23(5), 407-419.
20. Gilks, W. R., Richardson, S., & Spiegelhalter, D. (Eds.). (1995). Markov chain Monte Carlo in practice. CRC press.
21. Li, S., & Lund, R. (2012). Multiple changepoint detection via genetic algorithms. Journal of Climate, 25(2), 674-686.
22. Achcar, J. A., & Bolfarine, H. (1989). Constant hazard against a change-point alternative: a Bayesian approach with censored data. Communications in Statistics-Theory and Methods 18(10), 3801-3819.
23. Bacon, D. W., & Watts, D. G. (1971). Estimating the transition between two intersecting straight lines. Biometrika, 58(3), 525-534.
24. Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. Econometrica, 47-78.
25. Gelfand, A. B. AFM Smith. 1990. Samping-based approaches to calculating marginal densities. Journal of the American Statistical Association, 85, 398-4.
26. Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). Bayesian Data Analysis Chapman & Hall. CRC Texts in Statistical Science, 136.
27. Powell, B. (2016). Statistical modelling of climate change. In Climate Change (pp. 341-354). Elsevier.
28. Diaz-Nieto, J., & Wilby, R. L. (2005). A comparison of statistical downscaling and climate change factor methods: impacts on low flows in the River Thames, United Kingdom. Climatic Change, 69(2-3), 245-268.
29. Ribes, A., Zwiers, F. W., Azais, J. M., & Naveau, P. (2017). A new statistical approach to climate change detection and attribution. Climate Dynamics, 48(1), 367-386.
30. Vanem, E. (2016). Joint statistical models for significant wave height and wave period in a changing climate. Marine Structures, 49, 180-205.
31. Costello, A., Abbas, M., Allen, A., Ball, S., Bell, S., Bellamy, R., ... & Patterson, C. (2009). Managing the health effects of climate change: lancet and University College London Institute for Global Health Commission. The Lancet 373(9676), 1693-1733.

Copyright: *©2023 Emerson Barili. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*