

Data Analysis in Nonstationary State

Sadık Alashan¹ 

Received: 28 July 2017 / Accepted: 6 February 2018 /

Published online: 8 March 2018

© Springer Science+Business Media B.V., part of Springer Nature 2018

Abstract Climate change impact must be taken into account in any water resources planning management studies, because it does not allow future occurrences to be repeated as a replicate of the past. The stationarity is no longer valid, because the climate change plays significant role on ascending or descending trend components in any hydro-meteorological events such as temperature, precipitation, evaporation, runoff and discharge. The identification of trends can be represented by well-known methodologies, such as the Mann-Kendall and sequential Mann-Kendall. These methodologies require restrictive assumptions such as data length, serial independence and Gaussian (normal) probability distribution function (PDF). On the other hand, Innovative trend analysis (ITA) method proposed by Şen is helpful to identify even visually with direct interpretations without restrictive assumptions. The PDF or cumulative distribution function (CDF) is effective tool for risk level determination but it cannot tell anything about the trend in a given hydro-meteorological data. The PDF (CDF) does not yield any clue about the trend possibility, and hence, it's alone use in any water resources structure design may lead to erroneous planning studies. In this study, nonstationary nature of a given monthly hydro-meteorological data is examined by trend determination procedures. For the application, monthly averages of maximum daily temperatures are used on Oxford station, UK. It is observed that the temperature values of each month have a positive trend and the nonstationary empirical cumulative frequency curves on first half group match better all data group than the stationary state.

Keywords Climate change · Trend analysis · Innovative trend analysis · Frequency analysis

1 Introduction

The risk analysis using to extreme value determination is achieved by the probability or cumulative distribution functions (PDF or CDF) in cases of stationary hydro-meteorological records. Extreme values help to design water resources systems properly, and also for their

✉ Sadık Alashan
sadikalashan@bingol.edu.tr

¹ Civil Engineering Department, Hydraulics and Water Resources Division, Bingol University, 12000 Bingol, Turkey

optimal management studies. Due to especially climate change impact, hydro-meteorological variables show increasing or decreasing trends, which render them into a nonstationary structure. For this objective, trend analyses are achieved mostly by the Mann-Kendall, Spearman's Rho and sequential Mann-Kendall with Sen's slope estimator trend tests (Kendall 1975; Mann 1945; Obeysekera and Salas 2016; Sen 1968; Spearman 1904). Each one of these methods is based on assumptions such as enough data length, serial independence, normal (Gaussian) PDF. On the other hand, Sen (2017) proposed a new trend determination approach without any assumption, namely, innovative trend analysis (ITA) method. It is originated from that a time series (parent time series) is considered as two half in ITA and this enables to researchers a comparison of second half of parent time series according to its first half unlike traditional methods. The traditional methods treat a time series as whole and there is no comparison between the first half and second. To find the comparison, they need an assumption which the time series or errors fit normal distribution. The normal distribution is universal set of the time series and the comparison is made with it. Also the normal distribution requires enough data length and serial independence assumptions. There are ITA implementation studies by different researchers showing that the method is efficient. Sonali and Kumar (2013) applied ITA on monthly, seasonal and annual maximum and minimum temperatures over all regions in India and explained that ITA outcomes match other trend test results such as Mann-Kendall, Sen's slope, Spearman's Rho and their derivatives. Elouissi et al. (2016) investigated Macta watershed monthly precipitation records to identify climatologically tendencies using ITA and expressed that ITA provides qualitative interpretations. Güçlü et al. (2016) used ITA and the trend possibilities of precipitation records of the Florya station in Istanbul were searched and then quantified using intensity-duration-frequency curves. Oztopal and Sen (2017) applied ITA for determining trends in seven different sub-climatological region of Turkey and mentioned that ITA enables partial trend determination in low, medium and high data ranges unlike monotonic trend. On the other hand, Salas and Obeysekera (2014); Bayazit (2015) and Obeysekera and Salas (2016) discussed nonstationary analysis usage so as to include trends in data analysis with PDF and CDF. As human intervention and anthropogenic climate change influences are becoming evident, nonstationary statistics would be better to represent the stochastic properties of the hydro-meteorological records (Rehan and Hall 2014). The nonstationary analysis is based on time varying probabilities that cannot be accounted by stationary processes. Lins and Cohn (2011) accepted the nonstationarity as a characteristic of the natural world.

The main purpose of this paper is to show the difference between the stationary and nonstationary behaviour of the hydro-meteorological data records. Stationarity works are rather old fashion methodologies and in the future the nonstationarity of the records must be taken into consideration through trend analyses.

2 Methodology

Knowledges on the temporal variation of hydrological characteristics and parameters gives one the opportunity to build a nonstationarity model, where the available information allows one to reduce the bias of the predictions (Montanari and Koutsoyiannis 2014). Herein, Innovative trend analysis (ITA) method is used to calculate the nonstationary changes in hydro-meteorological variables over time. The parent time series in hydro-meteorological record is divided into two halves and then the each half is sorted into ascending order for the

implementation of the ITA as explained by Sen (2017). The first half is plotted on the horizontal and the second on vertical axis and the result is the scatter of points. Furthermore, the first half of parent time series must be plotted on the horizontal axis because the second half occurs after the first half changes with time. If the scatter of points remains wholly or partially above (below) the 1:1 (45°) straight-line then there is an increasing (decreasing) trend component in the hydro-meteorological record structure (Fig. 1a). As known, if a value on horizontal axis smaller (bigger) than a vertical value then a scatter point occurs below (above) the 45° straight-line on Cartesian coordinate system (see Fig. 1). If they are equals then the scatter point occurs over the 45° straight-line and there is no trend. In case of an increasing (decreasing) tendency on trend there is a steady increasing (decreasing) trend on data group (Fig. 1b).

Figure 2 shows the application of ITA based on a hypothetical data which it contains a certain trend quantity and no measurement error to test the method's efficient. To find the trend amount, first and second halves averages (μ_f, μ_s) are calculated (Fig. 2a) and the trend slope, a , is calculated by dividing the arithmetic averages difference by the half-length ($n/2$) of the data. Figure 2b shows the trend slope (a) and intersection point (b). All these calculations are performed using the following set of equations. In here, x_i (y_i) is the first (second) half group value and t_i is time which is i th number. More detailed information about ITA is provided by Sen (2017).

$$\mu_f = \frac{1}{(n/2)} \sum_{i=1}^{n/2} x_i \tag{1}$$

$$\mu_s = \frac{1}{(n/2)} \sum_{i=1}^{n/2} y_i \tag{2}$$

$$a = \frac{2 * (\mu_s - \mu_f)}{n} \tag{3}$$

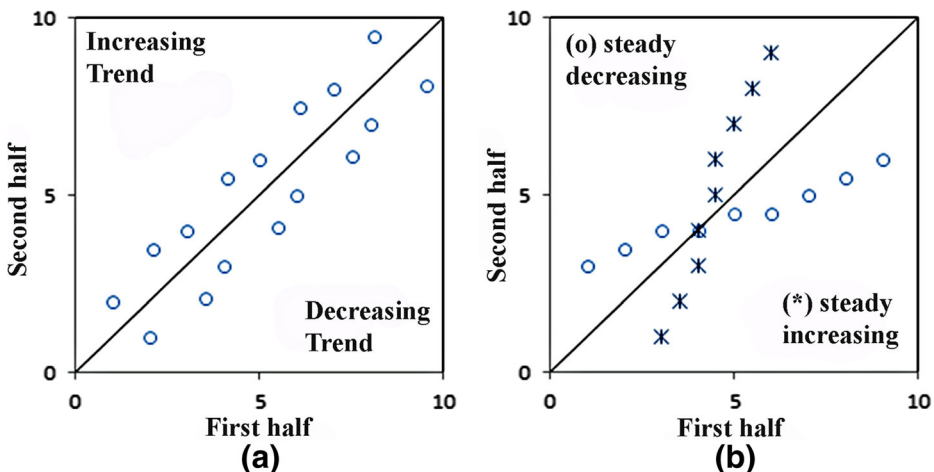


Fig. 1 a Monotonic trend. b Changing trend

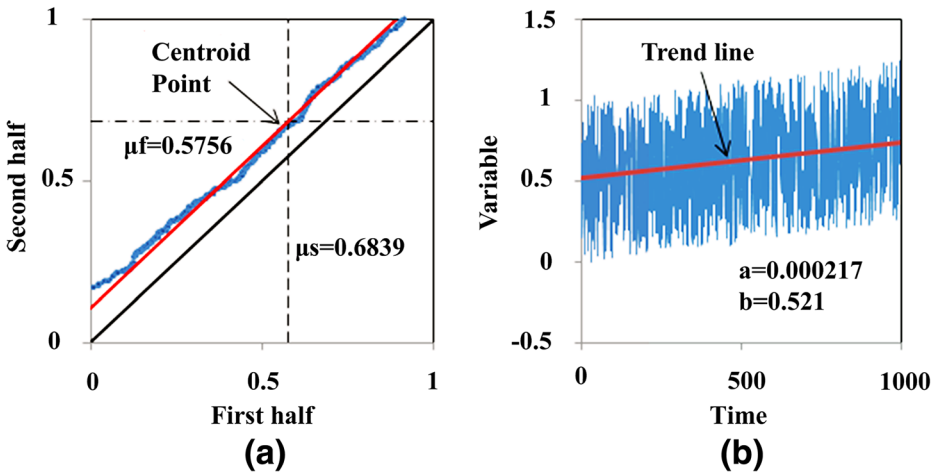


Fig. 2 Random stochastic processes (a) ITA graph (b) Time series graph

$$y_i = b + a * t_i \tag{4}$$

$$b = \mu_s - a * \bar{t} \tag{5}$$

Frequency is the occurrence probability of an event within a certain variable range. Let the first half data in order $x_1 < x_2 < \dots < x_{n/2}$ with the average, μ_f , and standard deviation, σ_f (Eqs. 1 and 6). Using Eq. (7), empirical cumulative distribution function in a stationary state (ECDF_S), $F(x_i, s)$, is simply obtained without regarded any trend and/or shift (jump, sudden change). An empirical equation is proposed in Eq. (8), where α , the trend in percent, a , the trend and μ_f , the mean of first half series (Eqs. 1 and 3). The trend existence in the data set implies nonstationary state. If the trend has a positive (negative) trend then with respect to the first half, the second half will have higher (smaller) extreme values. Thus a similar probability gives larger (smaller) a data value in the positive (negative) trend or the same data gives smaller (larger) a probability value. In order to import the effect of trend into the ECDF_S, these frequency values must be multiplied by $(1 + \alpha)$ and $(1 - \alpha)$ as in Eq. (9) for positive and negative trend components, respectively. The equation gives an empirical cumulative frequency distribution in nonstationary state (ECDF_NS). In case of $1 \pm \alpha = 1$, stationary case is valid without any trend component in the time series.

$$\sigma_f = \sqrt{\frac{1}{n/2} \sum_{i=1}^{n/2} (\mu_x - x_i)^2} \tag{6}$$

$$F(x_i, s) = \frac{2i}{n} \tag{7}$$

$$\alpha = \frac{a}{\mu_f} \tag{8}$$

$$F(x_i, ns) = (1 + \alpha) \frac{2i}{n} \tag{9}$$

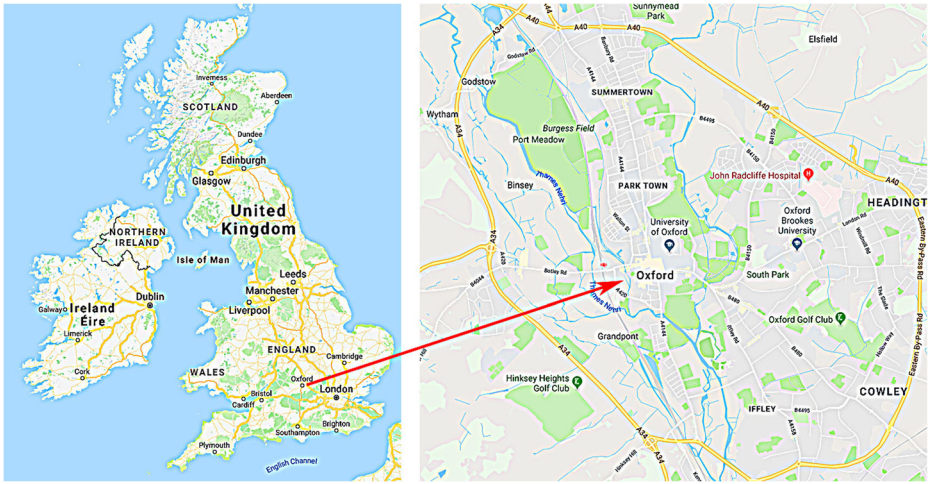


Fig. 3 The location of Oxford, UK

3 Study Area and Application

The methodology presented in the previous section is applied in the Oxford City, UK (Fig. 3). Monthly averages of daily maximum temperatures between 1853 and 2016 years are taken into consideration for implementing ITA and investigation of nonstationarity. A total of $164 \times 12 = 1968$ monthly temperature values are used and the coldest month is January (6.76 °C) and the warmest July (21.81 °C). The statistical information of the data is given in Table 1.

Average of the each half is calculated with Eqs. (1) and (2) which gave $\mu_f = 6.54$ C and $\mu_s = 6.98$ C for January and the same procedure is repeated for the other months. As an example, in January trend slope is calculated as $a = 2 \times (6.98 - 6.54) / (2016 - 1853 + 1) = 0.0054$ °C/year and trend in percent as $\alpha = a / \mu_f = 0.0054 / 6.54 = 8.205 \times 10^{-4}$ per year or $8.205 \times 10^{-4} \times 82 = 0.067 = 6.7\%$ per period. The trend slope ranges from 0.002 °C/year in February to 0.013 °C/year in November. However, the trend in percent changes from 0.04 in August to 0.12 in November. The biggest trend is in November and the lowest trend takes place in February and August considered both indicators. It is observed that there is a bigger

Table 1 A statistical summary of the monthly mean values of daily maximum temperatures(°C) for Oxford

Months	Measured years	Minimum	Mean	Maximum	Std. Dev.	Skew
January	1853–2016	−0.20	6.76	10.40	2.00	−0.58
February	1853–2016	−0.20	7.45	11.50	2.23	−0.79
March	1853–2016	6.00	10.00	15.50	1.86	0.09
April	1853–2016	9.50	13.22	19.40	1.62	0.62
May	1853–2016	12.90	16.73	20.40	1.48	0.00
June	1853–2016	15.90	19.92	24.40	1.57	−0.05
July	1853–2016	17.50	21.81	27.10	1.95	0.34
August	1853–2016	17.10	21.27	26.40	1.73	0.40
September	1853–2016	15.20	18.55	23.20	1.54	0.39
October	1853–2016	11.00	14.19	18.90	1.37	0.23
November	1853–2016	6.10	9.70	13.50	1.44	−0.05
December	1853–2016	0.20	7.35	13.40	1.92	−0.52

change in daily maximum temperatures in March, August, September, October and November compared to the minimum changes in February and May months (Fig. 4). Trends in winter are small amounts despite large percentages due to small temperatures. These different temperature trends can cause seasonal shifts in the future. The reason of different trends amounts among months may be local concentration differences of greenhouse gases such as water

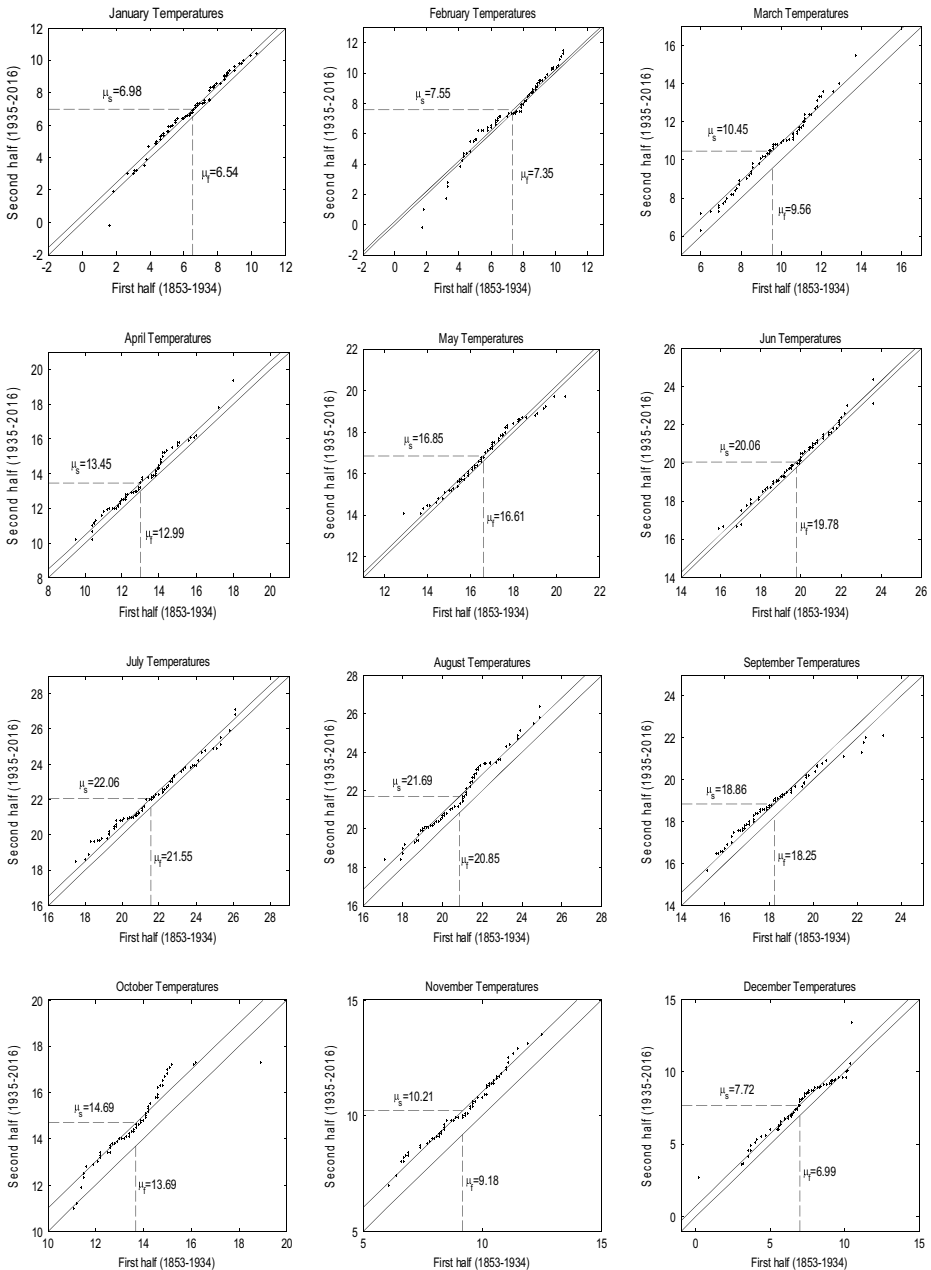


Fig. 4 Trend in monthly averages of daily maximum temperatures (°C)

vapor, C₂O, N₂O, NO and CH₄ among months. These greenhouse gases restrain reflection of shortwave solar radiation from the earth surface to the atmosphere (Toprak 2016). That is, the concentration differences may cause different trends on the temperatures among months.

The ECDF_NS is obtained for January as $F(x_i, ns) = (1 - 0.067) \times F(x_i, s)$ is used according to Eq. (9). These processes are repeated for other months and ECDFs are plotted for Oxford daily maximum temperatures (see Fig. 5). Horizontal axis represents maximum temperatures and

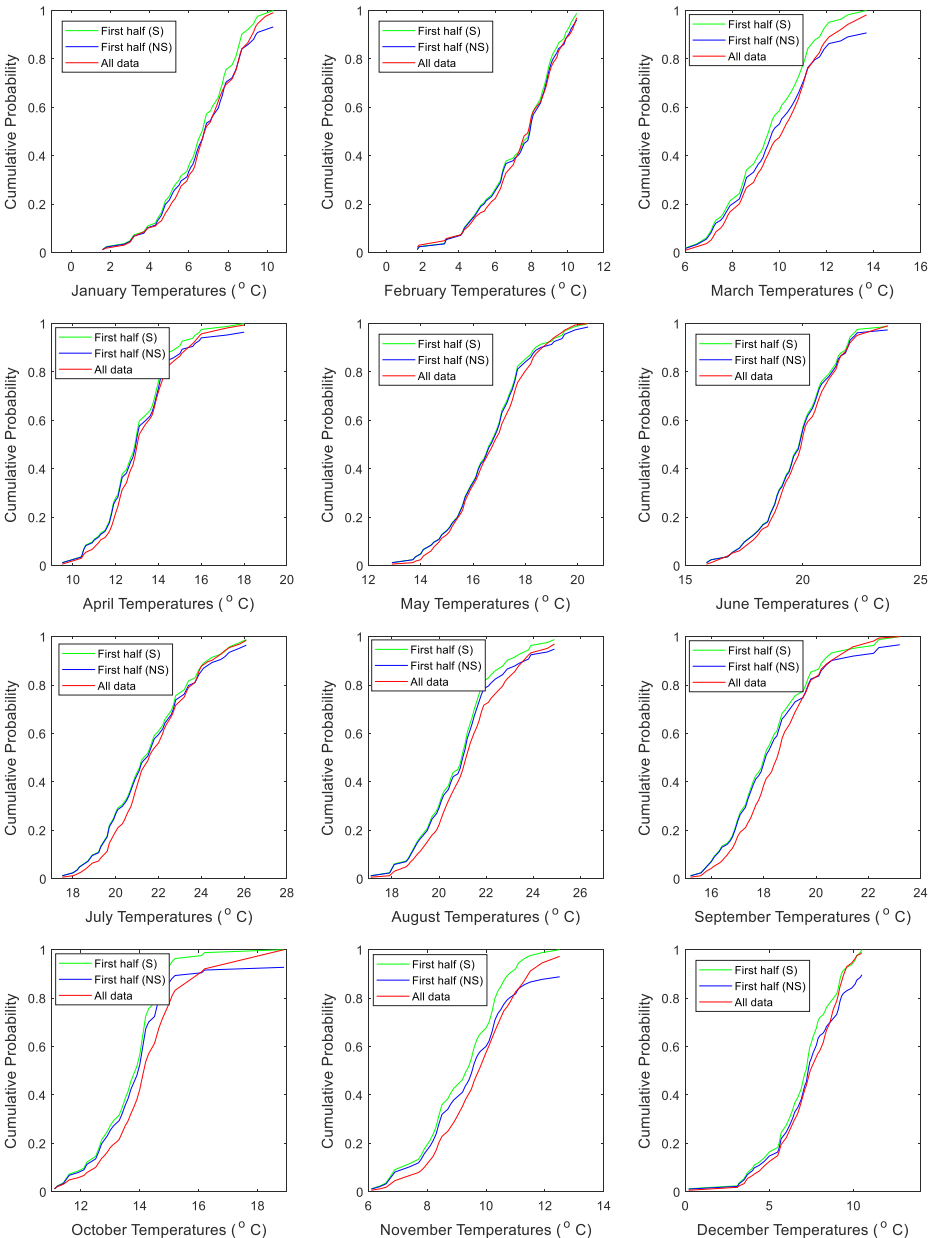


Fig. 5 Empirical cumulative distribution functions for monthly averages of maximum daily temperatures on Oxford

Table 2 Nonstationary data analysis test results and error amounts

	Jan	Feb	March	April	May	June	July	Aug	Sep	Oct	Nov	Dec
First half mean, μf ($^{\circ}C$)	6.54	7.35	9.56	12.99	16.61	19.78	21.55	20.85	18.25	13.69	9.18	6.99
Second half mean, μs ($^{\circ}C$)	6.98	7.55	10.45	13.45	16.85	20.06	22.06	21.69	18.86	14.69	10.21	7.72
Trend ($^{\circ}C$ /year)	0.005	0.002	0.011	0.006	0.003	0.003	0.006	0.010	0.007	0.012	0.013	0.009
Trend in percent	0.067	0.027	0.093	0.035	0.014	0.014	0.024	0.04	0.033	0.073	0.112	0.104
RMSE between ECDF _S and ECDF _A	0.043	0.027	0.073	0.047	0.034	0.031	0.046	0.073	0.082	0.13	0.107	0.066
RMSE between ECDF _{NS} and ECDF _A	0.021	0.02	0.034	0.032	0.028	0.025	0.04	0.054	0.07	0.091	0.056	0.044
Improvement in the error	50.4%	26.8%	53.1%	32.0%	18.4%	19.7%	14.1%	26.4%	14.9%	30.1%	47.8%	33.6%

vertical axis shows ECDF_S, ECDF_NS and ECDF_A values ($F(x_{i,s})$ and $F(x_{i,ns})$). In here, green, blue and red line represents first half data in stationary state, first half data in nonstationary state and all data in stationary state. It is seen that ECDF_S values are different than empirical cumulative distribution function values in all data state (ECDF_A) and this difference is bigger on March, August, September, October and November months than the other months. This difference is smaller between ECDF_NS and ECDF_A according to ECDF_S and ECDF_A. The minimum changes among ECDF_S and ECDF_A are seen in January, February and June months. Root mean square error (RMSE) method is used for testing the method's accuracy (Eq. 10). In here, obs and calc represents observed and calculated value, and n is data length. RMSE between ECDF_NS and ECDF_A ranges from 0.02 to 0.091 whereas ECDF_S and ECDF_A from 0.027 to 0.13. The nonstationary approach makes it possible to reduce errors among 14.1% and 53.1% (Table 2).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (obs(i) - calc(i))^2}{n}} \quad (10)$$

4 Conclusions

The main purpose of this paper is to use the innovative trend analysis (ITA) method and propose an update of the stationary frequency calculations by inclusion of the trend, and hence, nonstationary character of hydro-meteorological records. Oxford City, UK, monthly mean of daily maximum temperatures are used for testing cumulative nonstationary frequency analysis. The method is applied to monthly averages of daily maximum temperature values in Oxford, but can be used for other data and study areas. The nonstationary state is considered for frequency analysis using trend in percent. The trend obtained with ITA is added to empirical cumulative distribution function values in stationary state (ECDF_S) which are used for comparison of stationary and nonstationary states. Nonstationary empirical cumulative distribution frequency values (ECDF_NS) in first half data fits better to empirical cumulative distribution frequency in all data (ECDF_A) according to ECDF_S. Among ECDF_S and ECDF_A values, the difference is small, in where the trends have small values. This state shows that trend quantity lead to nonstationarity. Difference between ECDF_NS and ECDF_A is smaller than ECDF_S and ECDF_A values. The RMSE results support these findings and give smaller error amounts in the nonstationary state according to stationary state. It is concluded that nonstationary state gives more realistic results than stationary state. Nonstationary analysis enables qualitative forecasting of the future and thus, long-term planned hydraulic structures can be planned efficiently and resources are used more efficiently. Also, ITA can be useful for policy makers for determining risky areas with regard to flooding, drought, air temperature, sea level etc.

References

- Bayazit M (2015) Nonstationarity of hydrological records and recent trends in trend analysis: a state-of-the-art review. *Environ Processes* 2:527–542. <https://doi.org/10.1007/s40710-015-0081-7>

- Elouissi A, Sen Z, Habi M (2016) Algerian rainfall innovative trend analysis and its implications to Macta watershed. *Arab J Geosci* 9. <https://doi.org/10.1007/S12517-016-2325-X>
- Güçlü YS, Şişman E, Yeleğen MÖ (2016) Climate change and frequency-intensity-duration (FID) curves for Florya station. *Istanbul J Flood Risk Manage* 11:S403–S418. <https://doi.org/10.1111/jfr3.12229>
- Kendall MG (1975) Rank correlation methods. Oxford University Press
- Lins HF, Cohn TA (2011) Stationarity: wanted dead or alive? *J Am Water Resour As* 47:475–480. <https://doi.org/10.1111/j.1752-1688.2011.00542.x>
- Mann HB (1945) Nonparametric tests against trend. *Econometrica* 13:245–259. <https://doi.org/10.2307/1907187>
- Montanari A, Koutsoyiannis D (2014) Modeling and mitigating natural hazards: stationarity is immortal! *Water Resour Res* 50:9748–9756. <https://doi.org/10.1002/2014WR016092>
- Obeysekera J, Salas JD (2016) Frequency of recurrent extremes under nonstationarity. *J Hydrol Eng* 21:21. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001339](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001339)
- Oztopal A, Sen Z (2017) Innovative trend methodology applications to precipitation records in Turkey. *Water Resour Manag* 31:727–737
- Rehan BM, Hall JW (2014) Flood risk management decision analysis with finite historical records and highly variable climate effects. In: *Vulnerability, Uncertainty, and Risk*. doi:<https://doi.org/10.1061/9780784413609.289>
- Salas JD, Obeysekera J (2014) Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. *J Hydrol Eng* 19:554–568. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000820](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000820)
- Sen PK (1968) Estimates of the regression coefficient based on Kendall's Tau. *J Am Stat Assoc* 63:1379–1389. <https://doi.org/10.2307/2285891>
- Sen Z (2017) Innovative trend significance test and applications. *Theor Appl Climatol* 127:939–947. <https://doi.org/10.1007/s00704-015-1681-x>
- Sonali P, Kumar DN (2013) Review of trend detection methods and their application to detect temperature changes in India. *J Hydrol* 476:212–227
- Spearmen C (1904) The proof and measurement of association between two things. *Am J Psychol* 15:72–101. <https://doi.org/10.2307/1412159>
- Toprak ZF (2016) What is the global climate change? With the perspective of the people who live international. *J Environ Prot* 6:134–137. <https://doi.org/10.5963/IJEP0601015>