Theor Appl Climatol
DOI 10.1007/s00704-017-2082-0

ORIGINAL PAPER

Estimating missing daily temperature extremes in Jaffna, Sri Lanka

A. Thevakaran[1] **· D.U.J. Sonnadara**[2]

Received: 22 March 2016 / Accepted: 13 February 2017

Abstract The accuracy of reconstructing missing daily temperature extremes in the Jaffna climatological station, situated in the northern part of the dry zone of Sri Lanka, is presented. The adopted method utilizes standard departures of daily maximum and minimum temperature values at four neighbouring stations, Mannar, Anuradhapura, Puttalam and Trincomalee to estimate the standard departures of daily maximum and minimum temperatures at the target station, Jaffna. The daily maximum and minimum temperatures from 1966 to 1980 (15 years) were used to test the validity of the method. The accuracy of the estimation is higher for daily maximum temperature compared to daily minimum temperature. About 95% of the estimated daily maximum temperatures are within ±1.5 °C of the observed values. For daily minimum temperature, the percentage is about 92. By calculating the standard deviation of the difference in estimated and observed values, we have shown that the error in estimating the daily maximum and minimum temperatures is ±0.7 and ±0.9 °C, respectively. To obtain the best accuracy when estimating the missing daily temperature extremes, it is important to include Mannar which is the nearest station to the target station, Jaffna. We conclude from the analysis that the method can be applied successfully to reconstruct the missing daily temperature extremes in Jaffna where no data is available due to frequent disruptions caused by civil unrests and hostilities in the region during the period, 1984 to 2000.

✉ D.U.J. Sonnadara
upul@phys.cmb.ac.lk

[1] Department of Physics, University of Jaffna, Jaffna, Sri Lanka

[2] Department of Physics, University of Colombo, Colombo, Sri Lanka

1 Introduction

Serially incomplete climatic datasets create problems in many climatological studies and applications. Especially when large datasets are needed for the analysis of climatic parameters and other factors affected by climate conditions (Kiraly et al. 2006; Efstathiou et al. 2011; Varotsos et al. 2012), this becomes an issue. The existence of missing data may have been caused by a number of factors such as failure of equipment due to malfunctioning, extreme weather conditions such as hurricanes or floods, human-induced factors such as civil unrests or even accidental loss of data files in computers (Elshorbagy et al. 2000). A number of techniques have been developed in the past to estimate missing data based on regression analysis, time series analysis, artificial neural networks and interpolation techniques (Kemp et al. 1983; DeGaetano et al. 1995; Eischeid et al. 1995; Elshorbagy et al. 2000; Islam et al. 2000; Allen and DeGaetano 2001). The accuracy of these techniques may depend on the region, time of the year or spatial distribution of stations.

Estimation of missing temperature data is important for meteorologists, hydrologists and environment protection workers all over the world. The literature dealing with the estimation of missing daily temperature discusses a number of methods or procedures to resolve this problem. The methods adopted in the past can be classified into three broad categories: (1) within the station, (2) between stations and (3) regression-based techniques (Allen and DeGaetano 2001). The within-station method which can be considered as one of the simplest methods of estimating missing observations is based on taking the average of the temperature records of the previous and following days. The other two methods use data from one or more neighbouring stations to estimate the missing values at the

Published online: 24 February 2017

target stations. Kemp et al. (1983) assessed missing daily extremes in Idaho and British Columbia by a method based on the data either from the station itself or from neighbouring stations. They used seven different methods to estimate missing observations. The mean square error between the estimated and actual values is used to test the accuracy of the results. They concluded that the type of data, the length of the missing weather records and also the data in surrounding stations that are available for use in the analysis all contribute to the selection of the most appropriate technique. The authors reported that the between-station regression techniques generate significantly smaller errors compared to other methods.

Huth and Nemesova (1995) used a method to estimate the missing temperature based on the classification method by using principal component analysis and cluster analysis. Weather observations at close meteorological stations are utilized in the classification and the coefficients of the regression equations are determined for each class to estimate the temperature from a nearby station. They concluded that weather classification improved temperature estimation. Xia et al. (1999) applied six methods to estimate missing climatological data at three forest climate stations and six weather stations. Their study shows that multiple regression analysis using the closest stations produces the best results. A least squares regression approach to estimate missing daily maximum and minimum temperatures was presented by Allen and DeGaetano (2001). They have compared their method with two other popular estimation methods based on temperatures from neighbouring stations using regression and temperature departure-based approaches. They concluded that the accuracy of individual daily maximum and minimum temperature estimates is similar between the methods.

This paper presents a temperature departure-based approach of estimating the missing daily extreme temperatures at the Jaffna climatological station in the northern peninsula of Sri Lanka. Estimating missing daily data is particularly important for the dry zone covering northern and eastern regions of Sri Lanka since they have only a few climatological stations to cover a large extent of land mass. The recording of data was affected by frequent hostilities in these regions during the period 1984 to 2000, leading to long periods of missing data. In an earlier work, a temperature departure-based method was applied to estimate the missing monthly mean temperature records and found to be successful (Thevakaran and Sonnadara 2013). The purpose of the present paper is to extend the work carried out previously and to adopt the method to reconstruct missing daily maximum and minimum temperature records of the Jaffna climatological station, to determine the accuracy of the estimation and to suggest ways of improving the estimates.

2 Data sample

The data sample used in this work consisted of daily maximum and minimum temperature records of five stations (Target station—Jaffna; Neighbouring stations—Mannar, Anuradhapura, Puttalam and Trincomalee) during a 15-year period obtained from the Department of Meteorology, Sri Lanka. All stations were selected from the dry zone in the northern part of Sri Lanka within the same geographical region which experience similar climatic conditions (see Fig. 1). The weather in the northern part of Sri Lanka is usually hot during the dry season (February to October) and slightly cooler during the wet season (November to January). Although Sri Lanka is affected by two monsoons (southwest monsoon and northeast monsoon), the selected region is considered as semi-arid since most of the rain is received during the few months of the northeast monsoon period leaving the rest of the year with very little precipitation. Further details of the demarcation of these climatological zones and their rainfall and temperature characteristics have been given in several previous studies (Domroes and Ranatunge 1993; Puvaneswaran and Smithson 1993).

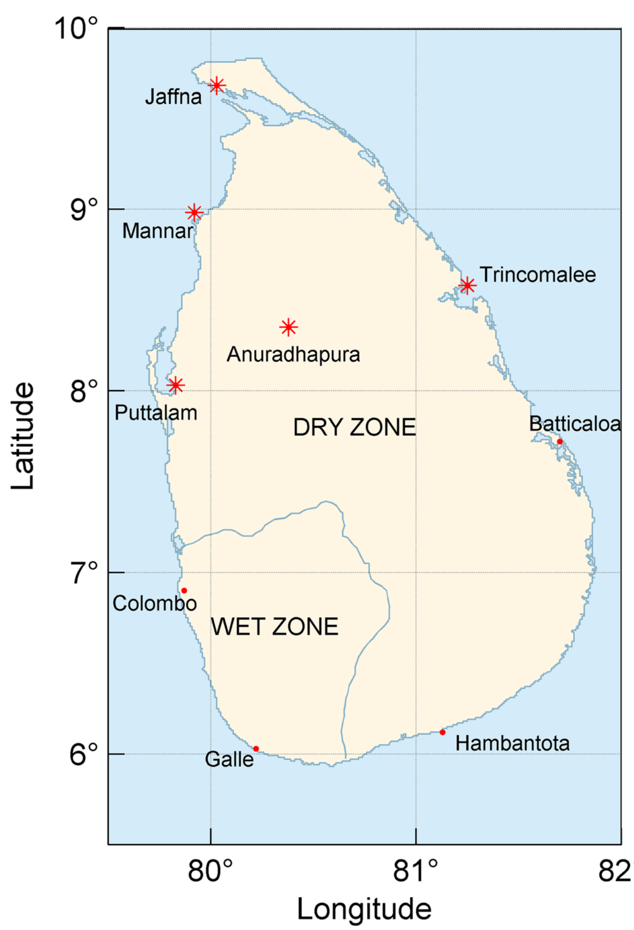


Fig. 1 Map of Sri Lanka indicating the locations of the selected stations: Target station—Jaffna; Neighbouring stations—Mannar, Puttalam, Anuradhapura and Trincomalee

The target station Jaffna is one of the main stations in Sri Lanka having weather records dating back to 1869. During the period where the data are missing, the selected neighbouring stations were in operation without any interruptions and thus can be used in the missing temperature reconstruction work. Ideally, data from the climatological station at Mulative which is on the northeast coast parallel to Mannar on the northwest coast could have been used. However, due to missing temperature data, it was not selected as a neighbouring station. The analysis in this study is based on daily maximum and minimum temperature data of five first-order stations (given in Table 1) recorded during the period 1966 to 1980. The missing temperature records during this period were very few (range from 0.6 to 1.2%). To create a serially complete dataset, isolated missing values were filled by using the measured temperature records of the adjacent days at the same station. The estimated error in this procedure is approximately ±1 °C. Since the region selected for this work is mostly a flat terrain without hills, the influence of elevation on temperature (DeGaetano and Belcher 2007) is minimal. On the other hand, the literature indicates that the impact of coastal effects is high up to 10 km inland from the coast (Daly 2006), and four of the five stations selected for this work are within this boundary. However, coastal effects are most noticeable when the water temperature is significantly different from the land temperature such as during the winter season. In our situation, since the temperature difference between land and water during different seasons is small, the effect on the reconstruction is minimal. However, the seasonal temperature in the inland station may behave differently to the other four stations which are in the close proximity to the coast. The lengths of data records together with the relevant station information (latitude, longitude and altitude) and percentage of missing records in each station are given in Table 1.

3 Methodology

Estimation of missing daily minimum and maximum temperatures for the target station Jaffna was done by using a refined form of the method originally developed by Steurer (1985, unpublished), based on the between-station technique. An assumption is made in this method that for a given day, the temperatures experienced by stations having similar geographical conditions will deviate by similar amounts from their standard normal values.

Thus, for daily maximum/minimum temperatures, it was assumed that for the five selected stations, the temperatures will deviate from their standard normal by similar amounts. For daily temperatures, the standard departure Z for station j was defined as,

$$Z_j = \frac{T_j - \overline{T}_j}{S_{j.}}$$

where j represents any station in the same climate division, T_j is the daily maximum/minimum temperature, $\overline{T}_j$ is the mean of the daily maximum/minimum temperatures and S_j is the standard deviation of the daily maximum/minimum temperatures. The average daily standard departure at the target station can be calculated by using the standard departures from the neighbouring stations excluding the target station, provided these stations are from the same climate division as the target station. However, the daily temperature in stations of close proximity to the target station may show higher correlation compared to distant stations. In addition, the variation in the seasonal temperature cycle between a pair of stations during a year may not follow a linear pattern. Since this method is based on daily standard departures of temperature values, seasonal variability in temperature will not be an issue. In order to favour the stations having better correlation with the target station, when combining the standard departures, separate weights are assigned for each station based on the inverse square of the standard deviation of the difference between the estimated and observed values, for the given pair of stations. Thus,

$$Z_t = \frac{\sum w_j \times Z_j}{\sum w_j}$$

Table 1 Summary of daily maximum and minimum temperature data used in this study

	Station name	Period (years)	Latitude N°	Longitude E°	Altitude (m)	Length (years)	Missing records
1	Jaffna	1966–1980	9.68	80.03	3.1	15	0.55%
2	Mannar	1966–1980	8.98	79.92	3.6	15	0.60%
3	Anuradhapura	1966–1980	8.35	80.38	92.5	15	0.58%
4	Puttalam	1966–1980	8.03	79.83	2.1	15	1.02%
5	Trincomalee	1966–1980	8.58	81.25	23.9	15	1.22%

and the weights w_j are estimated by taking,

$$w_j = \left(\frac{1}{\sigma_j}\right)^2$$

where the σ_j are the standard deviations of the difference between the estimated and observed values of the target and neighbouring stations. In this work, only four neighbouring stations of the target station were considered. The minimum requirement is to have data from at least one neighbouring station although three is recommended. Thus, the missing daily maximum/minimum temperature at the target station can be estimated by,

$$T_t = Z_t \times S_t + \overline{T}_t$$

where t represents the target station, and $\overline{T}_t$ and S_t are the mean daily maximum/minimum temperatures and their standard deviations. The accuracy of the estimates is verified by calculating the standard deviation of the difference between the estimated values and observed values.

4 Results and discussion

Jaffna is located in the northern peninsula which belongs to the dry zone of Sri Lanka. It has one of the highest average temperatures in the country. In this part of the country, the daily maximum temperature increases steadily from January to April and then decreases gradually during the remaining months up to December. The daily minimum temperature also follows a similar pattern with the peak month being May instead of April. During the period under consideration, the highest daily temperature was observed during the month of April (36.1 °C) and the lowest daily temperature was observed during the month of January (18.3 °C). The standard deviation of daily maximum and minimum temperatures varies by approximately 2.1 and 3.2 °C, respectively. As shown in Table 2, the temperature variation in the neighbouring stations selected for this work is similar to the variation seen in Jaffna.

In order to explore the link between the seasonal variation of daily maximum and minimum temperatures between stations, the correlation between the daily mean maximum/minimum temperatures at the target station and the neighbouring stations was studied. Figure 2a, b shows the seasonal variation of daily mean maximum temperature at Jaffna and two neighbouring stations, namely Mannar and Trincomalee. For comparison, the same is shown in Fig. 2c, d for the daily mean minimum temperature. The data show that, from the selected stations, Mannar which is in the northwest coastal belt and the station in close proximity to Jaffna closely follows the seasonal variation of maximum temperature in Jaffna. Trincomalee, which is on the northeast coastal belt and located far from Jaffna, shows a different seasonal pattern. In contrast, the minimum temperature for the same two stations shows a different relationship although the correlation between Mannar and Jaffna appears to be slightly better than that between Trincomalee and Jaffna. Thus, for minimum temperature, there is only a slight advantage by choosing Mannar over Trincomalee. Comparatively, the correlation of minimum temperature between stations may not be strong and the estimation could be difficult due to the effect of urban warming which is connected to the local climatic conditions of the stations. In all selected stations, the mean daily maximum/minimum temperature increases from the beginning of the year and reaches a peak towards mid-year and then decreases towards the end of the year. The cycle of temperature change during the year is depicted by the smooth line drawn through the daily mean temperature variations.

Initial work showed that the standard departures of daily maximum/minimum temperature of the target station Jaffna and the selected neighbouring stations are correlated at different degrees. Figures 3 and 4 show the histograms of difference between the standard departure of daily maximum/minimum temperature at the target station and neighbouring stations. In order to obtain the best result, measurements where the difference of the Z value is greater than ±2 can be adjusted when estimating the temperature at Jaffna from neighbouring stations. The difference in the Z value of maximum/minimum

Table 2 Daily temperature variability (1966–1980)

	Station name	Maximum temperature (°C)	Minimum temperature (°C)	Maximum temperature variability (°C)	Minimum temperature variability (°C)
1	Jaffna	36.1	18.3	±3.21	±2.14
2	Mannar	36.7	17.7	±2.70	±1.68
3	Anuradhapura	38.7	15.7	±3.44	±1.82
4	Puttalam	36.6	14.9	±1.80	±2.18
5	Trincomalee	37.9	19.9	±3.76	±1.36

Fig. 2 Seasonal correlation of daily maximum and minimum temperature. Tmax: **a** Jaffna vs. Mannar and **b** Jaffna vs. Trincomalee. Tmin: **c** Jaffna vs. Mannar and **d** Jaffna vs. Trincomalee. *Smooth line* is drawn through the 10-day moving average values calculated with daily mean values

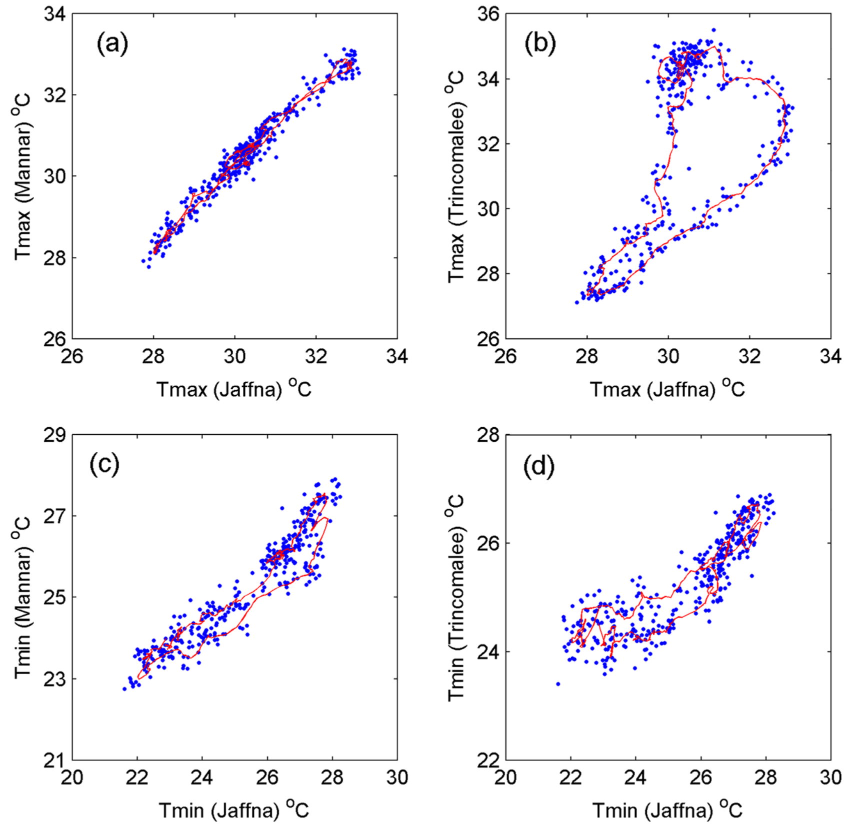
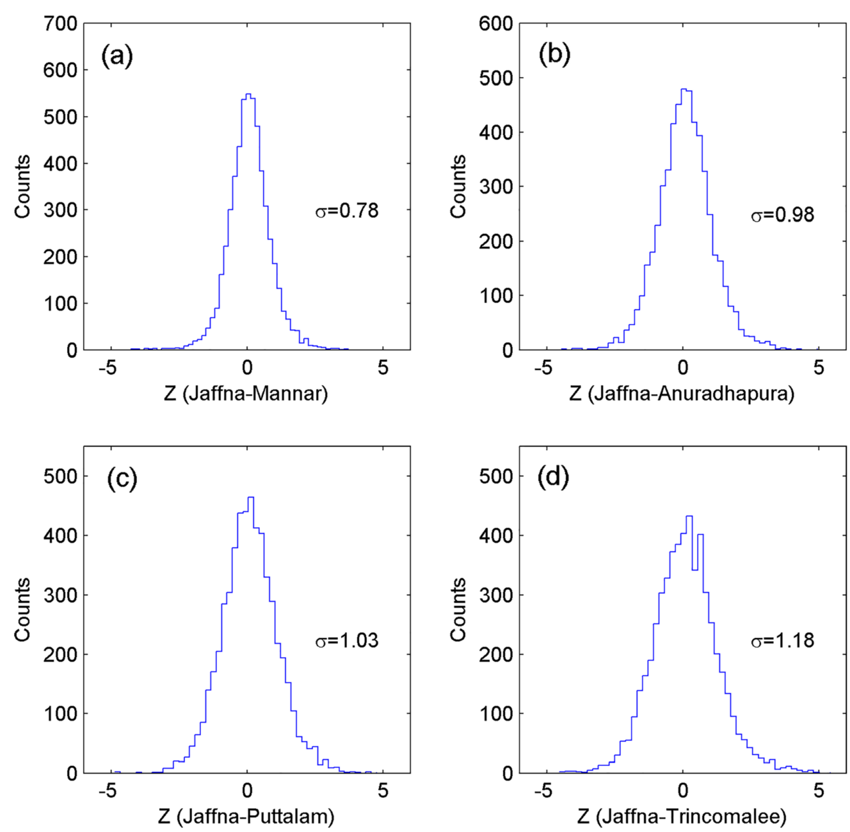


Fig. 3 Histograms of the difference between standard departures estimated based on daily maximum temperature values at Jaffna and four neighbouring stations



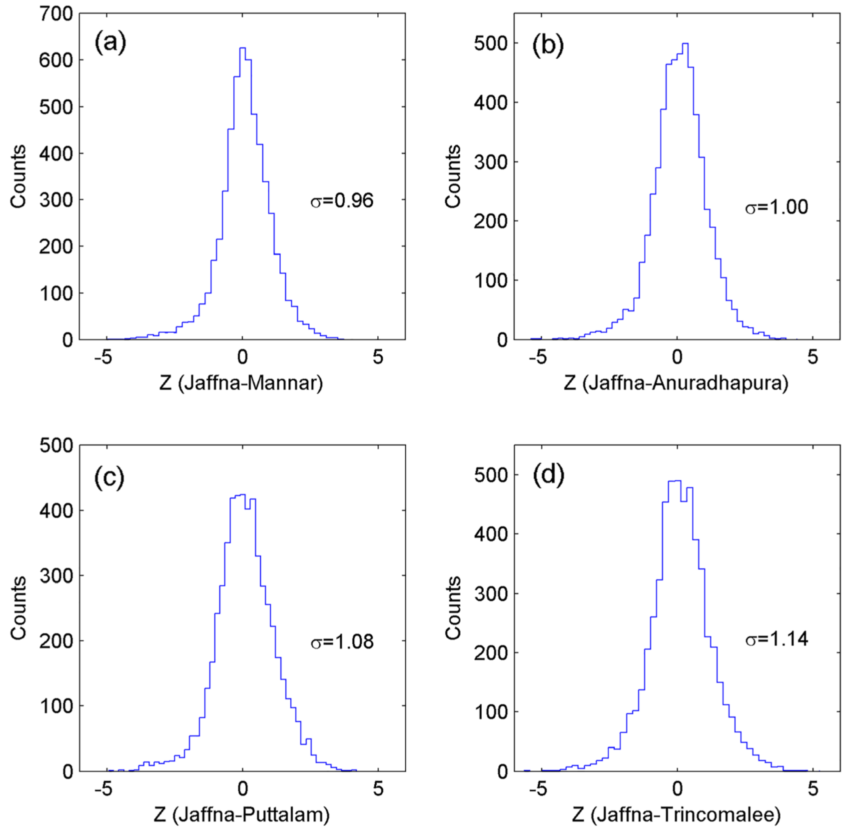


Fig. 4 Histograms of the difference between standard departures estimated based on daily minimum temperature values at Jaffna and four neighbouring stations

temperatures between Jaffna and the neighbouring stations shows that there are only a few outliers outside ±2 (about 2% to 8%). Since the outliers could be due to real shifts in temperature, in this analysis, no adjustments were applied. The corresponding weights can be estimated by computing the difference between the measured and the estimated daily maximum/minimum temperature at Jaffna based on standard departures of daily maximum/minimum temperatures calculated from each neighbouring station.

Table 3 shows the linear correlation coefficients, standard deviations and corresponding weights between the target station and the four neighbouring stations separately for the daily maximum and minimum temperatures. As expected, a strong correlation (r = 0.90) is observed for daily maximum temperature between Jaffna and the neighbouring station Mannar. The other three neighbouring stations, Anuradhapura, Puttalam and Tricomalee, are also correlated, but the relationship is not as strong as Mannar. This is expected since Mannar is the station closest to the target station compared to other stations. Unlike the daily maximum temperature, the correlation coefficients of daily minimum temperatures are not very different between the stations (0.82 to 0.88). Based on the estimated standard deviations between Jaffna and the neighbouring stations, the corresponding weights are calculated separately for daily maximum and minimum temperatures and shown in the last two columns of Table 3. Thus, stations having smaller standard deviations indicating a better correlation will contribute to a higher weight when the missing values are calculated.

Table 3 Linear correlation coefficients, standard deviations and corresponding weights between the target station Jaffna and the four neighbouring stations

Neighbouring stations	Correlation coefficient		Standard deviation (σ_j)		Weights (w_j)	
	Maximum	Minimum	Maximum	Minimum	Maximum	Minimum
Mannar	0.9034	0.8798	0.711	1.075	1.98	0.87
Anuradhapura	0.8405	0.8715	0.914	1.112	1.20	0.81
Puttalam	0.8282	0.8564	0.949	1.175	1.11	0.72
Trincomalee	0.7534	0.8197	1.136	1.317	0.77	0.58

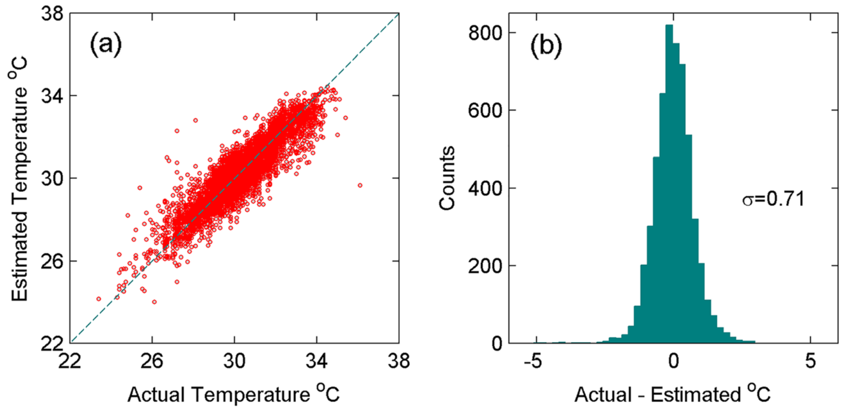


Fig. 5 Comparison between the actual and estimated values for daily maximum temperature at Jaffna for the period 1966 to 1980. **a** Correlation. **b** Difference

The correlation between the observed and estimated daily maximum temperatures in Jaffna is shown in Fig. 5. The same is shown for daily minimum temperatures in Fig. 6. For both maximum and minimum temperatures, the linear correlation coefficient is 0.9 which is significant at $p < 0.01$. The gradient of each of the fitted lines is approximately 1. About 95% of the estimated values of daily maximum temperature are within ±1.5 °C of the actual values, while for daily minimum temperature, it is about 92%. In general, the accuracy of estimated maximum temperatures is higher than minimum temperatures. The standard deviation of the difference between the actual and predicted value is ±0.71 °C for the maximum temperature and ±0.93 °C for the minimum temperature. There are a few outliers substantially deviating from the expected $y = x$ line. In principal, these could be due to actual shift in the temperature values in the target station. Furthermore, no significant systematic temperature-dependent bias was detected in the data. The accuracy was also tested by using three stations instead of four stations and systematically removing one station at a time. When the Mannar station was not included in the calculation, the error in reconstruction was increased to ±0.85 and ±0.98 for daily maximum and minimum temperatures, respectively. Removing other stations did not change the error in reconstruction by an appreciable amount. Thus, even with the weighted average, to obtain the best accuracy, it is important to include the station nearest to the target station (in this case Mannar) situated within the same climatological division in reconstructing the missing daily extreme temperature data.

The main assumptions made when extending the technique discussed above to the calculation of missing maximum/minimum temperatures are that the long-term mean and standard deviation of maximum/minimum temperatures calculated from the period where the data is available would remain the same during the period where the data is missing. The linear trend of monthly mean temperatures at the target station Jaffna during the last 100 years varies from −0.06 to +0.04 °C per decade. Thus, the shift in mean temperature for a 15-year period is less than ±0.1 °C. The error of the mean value ($S_t/\sqrt{15}$) of the daily maximum/minimum temperatures estimated for the period 1966–1980 is approximately ±0.25 °C which indicates that the variability in the daily maximum/minimum temperature during a 15-year period is quite small.

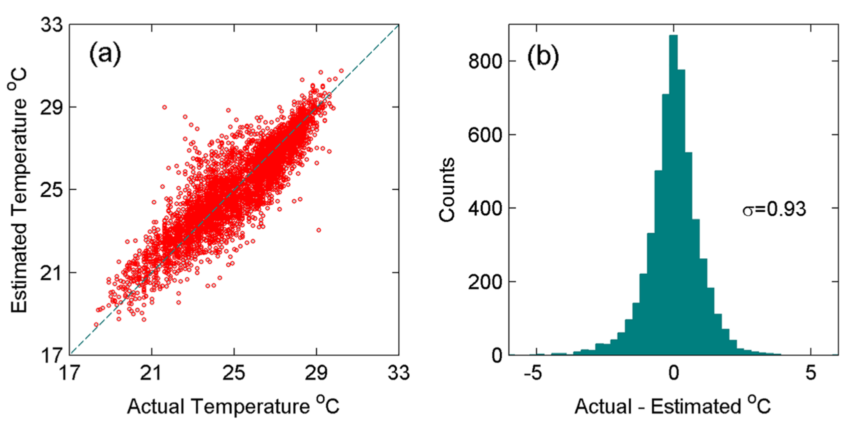


Fig. 6 Comparison between the actual and estimated values for daily minimum temperature at Jaffna for the period 1966 to 1980. **a** Correlation. **b** Difference

5 Conclusions

The results of a statistical method of estimating the missing daily maximum and minimum temperature in Jaffna, which is situated in the dry zone of Sri Lanka, is presented. The daily maximum and minimum temperatures from 1966 to 1980 (15 years) were used to test the accuracy of the method. Four neighbouring stations having serially complete temperature data records from the same climatological division were selected to estimate the daily extreme temperature at the target station. There is a strong correlation between the target station Jaffna and the neighbouring station Mannar for the daily maximum temperature. Among the selected stations, weakest correlation was found between Jaffna and Trincomalee. In contrast, the correlation between the daily minimum temperature in Jaffna and neighbouring stations has less diversity. The standard deviation of the difference between the target station and the neighbouring stations varied from 0.7 to 1.1 °C for the daily maximum temperature, while the variation for the daily minimum temperature was 1.1–1.3 °C.

The accuracy of estimating maximum temperatures is higher than minimum temperatures. The accuracy of the temperature reconstruction can be evaluated by calculating the standard deviation of the difference between the actual and estimated values. The estimated accuracy is ±0.71 °C for the daily maximum temperature and ±0.93 °C for the daily minimum temperature. About 95% of the estimated values of daily maximum temperatures are within ±1.5 °C of the actual values, while for daily minimum temperatures, it is about 92%. Especially for the maximum temperature, it is important to include stations in close proximity such as Mannar to obtain the best results. In fact, the accuracy of the reconstruction of the daily maximum temperature can be improved slightly by removing distant stations having relatively weak correlations, such as Trincomalee.

The level of accuracy obtained in this work is comparable with published literature and perhaps sets the lower limit that can be achieved in temperature reconstruction for this region. In moving forward, the technique applied in this work can be extended to improve the quality of temperature data. Especially given the fact that there are no terrain effects in the dry zone, this technique can assist in developing a high-quality data set for the dry zone of Sri Lanka or at least improve the data collected during the periods of civil unrest.

Acknowledgments Financial support by Higher Education for the Twenty First Century (HETC) Project assisted by the World Bank (IDA Credit 49190-LK) is acknowledged.

References

Allen RJ, DeGaetano AT (2001) Estimating missing temperature extremes using an optimized regression approach. Int J Climatol 21: 1305–1319. doi:10.1002/joc.679

Daly C (2006) Guidelines for assessing the suitability of spatial climate data sets. Int J Climatol 26:707–721. doi:10.1002/joc.1322

DeGaetano AT, Eggleston KL, Knapp WW (1995) A method to estimate missing maximum and minimum temperature observations. J Appl Meteorol 34(2):371–380

DeGaetano AT, Belcher BN (2007) Spatial interpolation of daily maximum and minimum air temperature based on meteorological model analyses and independent observations. J Appl Meteor Climatol 46(11):1981–1992

Domroes M, Ranatunge E (1993) A statistical approach towards a regionalization of daily rainfall in Sri Lanka. Int J Climatol 13(7):741–754

Efstathiou MN, Tzanis C, Cracknell AP, Varotsos CA (2011) New features of land and sea surface temperature anomalies. Int J Remote Sens 32:3231–3238

Eischeid JK, Baker CB, Karl TR, Diaz HF (1995) The quality control of long-term climatological data using objective data analysis. J Appl Meteorol 34:2787–2795

Elshorbagy A, Panu SU, Simonovic SP (2000) Group-based estimation of missing hydrological data: I. Approach and general methodology. J Hydrol Sci 45(6):849–866

Huth R, Nemesova I (1995) Estimation of missing daily temperature: can a weather categorization improve its accuracy. J Clim 8(7):1901–1916

Islam N, Doscher C, Davies T (2000) Estimating daily maximum and minimum temperature for mount cook, South Island, New Zealand using statistical model and 'aiNet' neural network models. J Hydrol 39(2):83–106

Kemp WP, Burnell DG, Everson DO, Thomson AJ (1983) Estimating missing daily maximum and minimum temperatures. J Climate Appl Meteor 22(9):1587–1593

Kiraly A, Bartos I, Janosi IM (2006) Correlation properties of daily temperature anomalies over land. Tellus A 58:593–600

Puvaneswaran KM, Smithson PA (1993) An objective classification of homogeneous rainfall regimes in Sri Lanka. Theo Appl Climatol 48(2–3):133–145

Thevakaran A, Sonnadara DUJ (2013) Reconstruction of missing monthly temperature observations in Jaffna-Sri Lanka. J National Sci Foundation 41(1):21–28

Varotsos C, Efstathiou M, Tzanis C, Deligiorgi D (2012) On the limits of the air pollution predictability: the case of the surface ozone at Athens, Greece. Env Sci Poll Res 19:295–300

Xia Y, Fabian P, Stohl A, Winterhalter M (1999) Forest climatology: estimation of missing values for Bavaria, Germany. Agri Forest Meteorol 96(1–3):131–144