ISSN 2616-5872 Vol.7, Issue 4: 101-106, DOI: 10.25236/AJEE.2025.070413

Variability of Annual Daily Maximum Rainfall Data of South Australia

Zhuodi Yao

Shaanxi Agricultural Development Group Xi'an Branch, Xi'an, 710000, China

Abstract: *Under the influence of the Mediterranean climate, South Australia (SA) has significant extreme rainfall and temperature events and annual variability. In order to investigate the relationship between annual daily maximum rainfall (ADMR) and annual daily maximum temperature (ADMT) in SA, a series of statistical tests were carried out. Two kinds of null and alternative hypotheses were established based on the difference of weather stations and related altitudes. After analyzing descriptive statistics and using normality test, Mann-Whitney test, Kruskal-Wallis test and correlation and regression analysis were applied. In different weather stations of SA, ADMR/ADMT data vary significantly. Nonetheless, in different stations of SA with different altitudes, ADMR/ADMT data are not significantly different. Furthermore, there is a weak correlation between selected ADMR and ADMT data in SA. Thus, it is unsuitable to directly predict ADMR based on ADMT in SA, and vice versa. For policymakers of SA, they need to adjust measures to local conditions when making ADMR/ADMT related policies.*

Keywords: *Annual Daily Maximum Rainfall; Annual Daily Maximum Temperature; Mann-Whitney Test; Kruskal-Wallis Test; Correlation Analysis; Regression Analysis*

1. Introduction

It is commonly known that rainfall varies greatly in space and time, particularly in the area of South Australia (SA) having the Mediterranean climate, which has significant extreme rainfall events and annual variability[3,10]. Under the influence of the Mediterranean climate, most rainfall events of Annual rainfall occur in the winter half year (i.e., May to October) of SA, showing the imbalance of rainfall[16]. Previous studies show that the variability of rainfall has a series of influences on SA. Based on the impact on river flow and other catchments, rainfall variability is essential for water balance of SA[22]. Notably, flood risk in Adelaide was proved to be mainly related to annual daily maximum rainfall (ADMR) rather than other factors such as ground features, which is essential for other areas[3]. In addition to events of flooding, the yield of various crops such as wheat in SA is also associated with rainfall[6]. After investigating the interaction between rainfall, wheat yield and nitrogen supply, appropriate rainfall and soil was proved to increase up to the yield of wheat by half[17]. However, since uncertainty remains to understand the mechanism responsible for rainfall variability[13], rainfall variability of SA, especially variability of ADMR, still needs further investigation.

Nowadays, lots of researchers investigate the relationship between oceanic-atmospheric phenomena and rainfall. The interactions of the southern Oscillation Index and the Indian Ocean Dipole result in complex variability of rainfall in SA[9,13]. In order to provide guidance to human activities, Australian Bureau of Meteorology (BoM) developed a dynamical system to predict rainfall by modelling of ocean atmosphere[12]. Even so, better prediction is required for preventing droughts and flooding[9]. Annual daily maximum temperature (ADMT), which is related to oceanic-atmospheric phenomena, is possible to be associated with rainfall (ADMR). Previous studies show that there are lack of evidence to prove the correlation between rainfall varieties and values of temperature in SA[9,20]. Therefore, further effort is required to understand the relationship of ADMT and ADMR, two typical climate extremes of SA[4].

For better understanding historical rainfall data, statistical hypothesis tests are used to carry out data analysis. When the dependent variable is unsure or turns out to be non-normal, parametric t-test and analysis of variance (ANOVA) are not appropriate for data analysis[15]. In this case, the Mann-Whitney test (MWt) and the Kruskal-Wallis test (KWt) are suitable for comparing two groups and two or more groups, respectively. As shown in Fig. 1, both tests have been widely used in recent years, which are applied in many areas such as social science and medical science[1,15]. Nonetheless, some problematic disadvantages cannot be overlooked. For example, when carrying out the KWt, privacy concerns of biomedical research remain because of the confidential information in related data[8]. In order to

ISSN 2616-5872 Vol.7, Issue 4: 101-106, DOI: 10.25236/AJEE.2025.070413

overcome existing shortcomings of the MWt and the KWt, much work so far has focused on improving these tests. In a case of using mammography to diagnose breast cancer, the uFilter was applied to improve the MWt for reduction of dimensionality and ranking features[14]. In another case of solving two-sample problem under dependent censoring, a new estimator for MWt effect was proposed[5]. In order to determine association and summarize relationship between two variables, correlation and regression analysis are useful for data analysis when independent and dependent variables are both in scale. Notably, the causation is not determined by correlation. In addition to conventional time-series or cross-section data, panel data is also suitable for correlation and regression analysis[19]. Nowadays, correlation and regression analysis have been widely used in most aspects of data science, especially in economics and business[11,21].

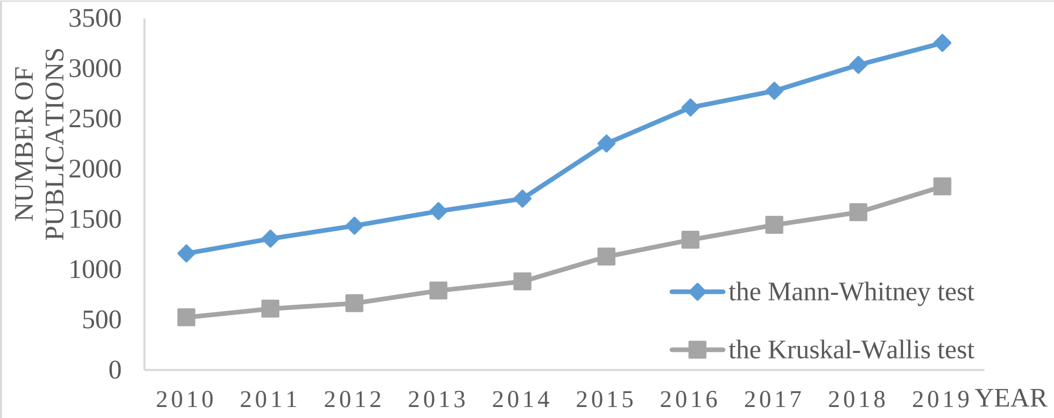

Fig.1 Evolution of published works concerning MWt or KWt (Source: Web of Science 2020)

This research paper is a case study of analyzing the relationship between ADMR and ADMT in SA based on non-parametric hypothesis tests. The purpose of this study is to understand the variety of rainfall and temperature extremes in SA, providing guidance for prediction of serious consequences caused by rainfall and temperature extremes in SA.

2. Study area and methodology

As shown in Table 1 and Fig. 2, four weather stations in SA were selected for this research. Depending on the altitude, the four stations can be divided into two groups: stations at relatively low altitudes (i.e., Adelaide Airport and Kent Town) and stations at relatively high altitude (i.e., Rosedale and Mount Barker). ADMR and ADMT data from 1990 to 2019, which were collected from online climate data of BoM (2020).

Table 1 Information of selected weather stations (Source: BoM 2020)

No.	Station	Number	Latitude (°S)	Longitude (°E)	Altitude (m)
1	Adelaide Airport	023034	34.94	138.53	2
2	Kent Town	023090	34.92	138.62	48
3	Rosedale	023343	34.55	138.83	116
4	Mount Barker	023733	35.07	138.85	359

In this study, the Statistical Package for the Social Sciences (SPSS) (IBM Corporation 2017) was used for data analysis. A flow chart of methodology is shown in Fig. 3. Since the reasons of selecting hypothesis tests are shown in Fig. 4, appropriate hypothesis tests can be applied in different situations[2]. On the one hand, selected data was divided into four groups based on the difference of weather stations (i.e., 1 = Adelaide Airport, 2 = Kent Town, 3 = Rosedale, 4 = Mount Barker). In this case, the null hypothesis (Ho) is that no association exists between ADMR/ADMT and weather stations; the alternative hypothesis (HA) is that an association exists between ADMR/ADMT and weather stations. On the other hand, elected data was divided into two groups based on the difference of altitudes (i.e., 1 = Adelaide Airport and Kent Town, 2 = Rosedale and Mount Barker). In this case, the null hypothesis (Ho) is that no association exists between ADMR/ADMT and altitude; the alternative hypothesis (HA) is that an association exists between ADMR/ADMT and altitude.

ISSN 2616-5872 Vol.7, Issue 4: 101-106, DOI: 10.25236/AJEE.2025.070413

Fig.2 Map of selected weather stations (Source: Google Map 2020)

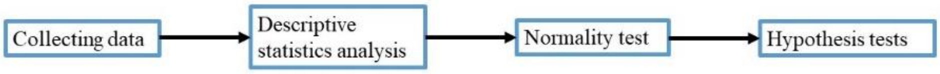

Fig.3 A flow chart of methodology

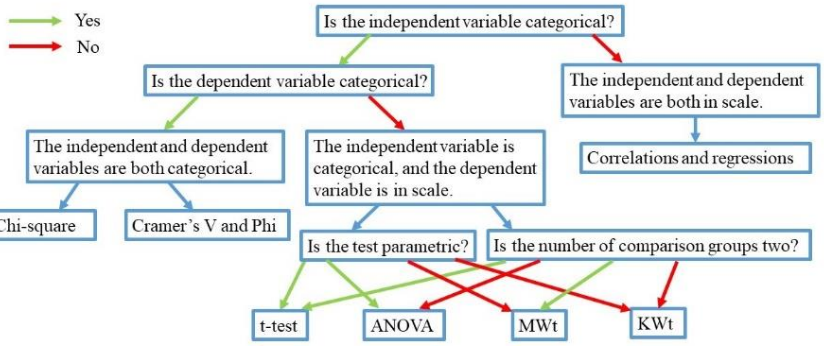

Fig.4 Appropriate hypothesis tests for data analysis

3. Results and discussions

3.1 Descriptive statistics

Descriptive statistics are shown in Table 2. Notably, the skewness of ADMR (i.e., 1.05) is relatively large compared with that of ADMT (i.e., 0.25), showing relatively large variability of ADMR.

Table 2 Descriptive statistics

Parameters	ADMR	ADMT
Mean	37.34	42.25
Median	34.60	42.10
Mode	22.60[a]	40.00[a]
Std. Deviation	14.13	1.98
Skewness	1.05	0.25
Std. Error of Skewness	0.22	0.22
Kurtosis	1.18	-0.35
Std. Error of Kurtosis	0.44	0.44

Note: "[a]" means multiple modes exist and the smallest value is shown.

3.2 Normality test

Since both sample sizes of ADMR and ADMT are more than 50, Kolmogorov-Smirnov test is more

ISSN 2616-5872 Vol.7, Issue 4: 101-106, DOI: 10.25236/AJEE.2025.070413

suitable for this research than Shapiro-Wilk test. In Kolmogorov-Smirnov test, data is normal when significant value is more than 0.05, and vice versa. As shown in Table 3, the significant values of ADMR and ADMT data in Kolmogorov-Smirnov test are 0.00 and 0.20, respectively. Thus, selected ADMR data are normally distributed[7], while selected ADMT data are not. The Q-Q plots also show same results. Data are distributed normally when data points are close to the diagonal line (see ADMT), and vice versa (see ADMR). The result of this normality test supports previous points in a published literature, which also investigates the relationship between ADMR and ADMT in SA[20]. Since the t-test and ANOVA are not suitable when the dependent variable is non-normal, KWt and MWt were used in subsequent data analysis.

Table 3 Summary of normality test

Data	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
ADMR	0.15	120	0.00	0.93	120	0.00
ADMT	0.04	120	0.20*	0.99	120	0.37

Notes: "*" means this is a lower bound of the true significance; "a" means Lilliefors Significance Correction.

3.3 KWt

As shown in Table 4 and Fig. 5 and 6, the results of KWt are statistically significant and the Ho of ADMR and ADMT are rejected. After calculation, the significant values of ADMR and ADMT (see Table 4, i.e., 0.004 and 0.000, respectively) are both less than 0.05. Meanwhile, according to Fig. 5, the mean values of each station are not in similar level, so the means are significantly different with each other. Furthermore, Thus, there is an association exists between ADMR/ADMT and weather stations. In different weather stations of SA, ADMR/ADMT data are significantly different.

Table 4 Summary of using KWt

No.	Null Hypothesis	Test	Sig.	Decision
1	The distribution of ADMR is the same across categories of station.	Independent-Samples KWt	0.004	Reject the null hypothesis.
2	The distribution of ADMT is the same across categories of station.	Independent-Samples KWt	0.000	Reject the null hypothesis.

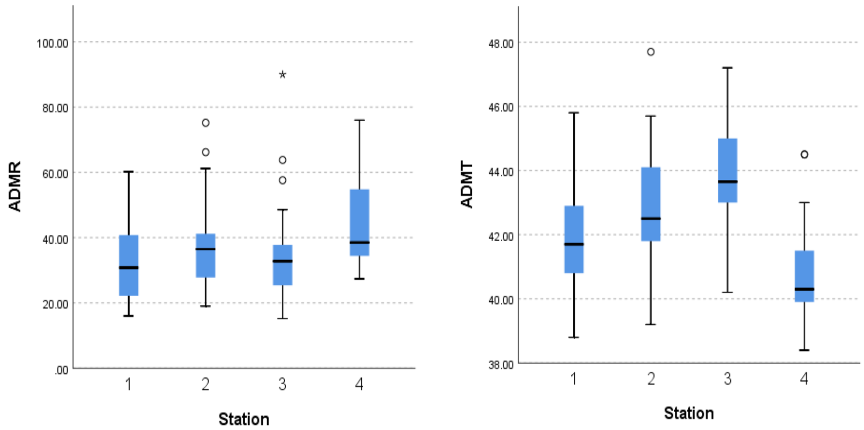

Fig.5 Results of KWt for ADMR and ADMT

3.4 MWt

As shown in Table 5 and Fig. 6, the results of MWt are not statistically significant and the Ho of ADMR and ADMT are retained. After calculation, the significant values of ADMR and ADMT (see Table 5, i.e., 0.155 and 0.955, respectively) are both more than 0.05. Meanwhile, according to Fig.6, the mean ranks are in similar level. Thus, there is no association exists between ADMR/ADMT and altitude of selected weather station. In different stations of SA with different altitudes, ADMR/ADMT data are not significantly differen[18] t. The results of KWt and MWt show that there are other factors besides altitude that contribute to the difference in ADMR and ADMT between different weather stations. For example,

ISSN 2616-5872 Vol.7, Issue 4: 101-106, DOI: 10.25236/AJEE.2025.070413

the distance from the weather stations to the sea was proved to have influence on ADMR/ADMT[20].

Table 5 Summary of using MWt

No.	Null Hypothesis	Test	Sig.	Decision
1	The distribution of ADMR is the same across categories of altitude.	Independent-Samples MWt	0.155	Retain the null hypothesis.
2	The distribution of ADMT is the same across categories of altitude.	Independent-Samples MWt	0.955	Retain the null hypothesis.

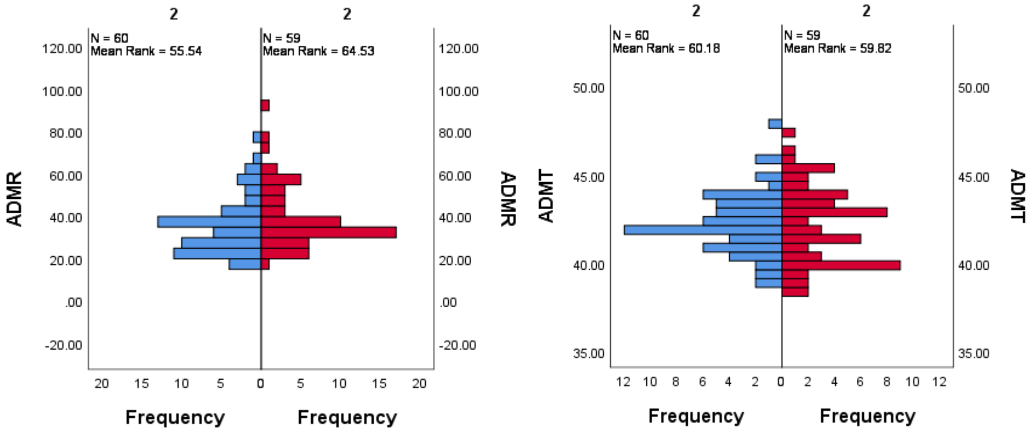

Fig.6 MWt results of ADMR and ADMT for altitude category

3.5 Correlation and regression analysis

The relationship between ADMR and ADMT was investigated by using correlation and regression analysis. Since Spearman' rank requires at least one variable is non-parametric, Pearson's product moment was used in this research. As shown in Table 6, the correlation coefficient is -0.170, which is close to zero. Furthermore, the significant value (i.e., 0.063) is larger than 0.01. The above results show that there is weak correlation between selected ADMR and ADMT data in SA, and the test is not statistically significant. Considering that sometimes ADMR and ADMT did not happen in the same month, the result is acceptable.

Table 6 Correlation between ADMR and ADMT

Model	Unstandardized coefficients		Standardized coefficients	t	Significance
	B	Std. error	Beta		
Constant	43.143	0.509		84.835	0.000
ADMR	-0.024	0.013	-0.170	-1.875	0.063

4. Conclusions

In order to analyze the relationship between ADMR and ADMT in SA, KWt, MWt and Correlation tests were performed based on a case study. In different weather stations of SA, ADMR/ADMT data vary significantly. Thus, policymakers of SA need to adjust measures to local conditions when making ADMR/ADMT related policies. Nonetheless, in different stations of SA with different altitudes, ADMR/ADMT data are not significantly different. This phenomenon shows that there are other factors besides altitude that contribute to the difference in ADMR and ADMT between different weather stations in SA. Furthermore, there is weak correlation between selected ADMR and ADMT data in SA. It is unsuitable to directly predict ADMR based on ADMT in SA, and vice versa. Further effort is required to analyze the relationship between ADMR and ADMT in SA based on more weather stations. Meanwhile, other factors besides altitude that contribute to the difference in ADMR and ADMT in SA remain to be explored.

References

[1] Ai, C, Huang, L & Zhang, Z 2020, 'A Mann–Whitney test of distributional effects in a multivalued

ISSN 2616-5872 Vol.7, Issue 4: 101-106, DOI: 10.25236/AJEE.2025.070413

treatment', Journal of Statistical Planning and Inference, vol. 209, pp. 85–100.
[2] Bureau of Meteorology 2018, Bureau of Meteorology, Australian Government, viewed 10 June 2020, <http://www.bom.gov.au/climate/data/?ref=ftr>.
[3] Ashcroft, L, Karoly, DJ & Dowdy, AJ 2019, 'Historical extreme rainfall events in southeastern Australia', Weather and Climate Extremes, vol. 25, p. 100210.
[4] Avila, FB, Dong, S, Menang, KP, Rajczak, J, Renom, M, Donat, MG & Alexander, LV 2015, 'Systematic investigation of gridding-related scaling effects on annual statistics of daily temperature and precipitation maxima: A case study for south-east Australia', Weather and Climate Extremes, vol. 9, pp. 6–16.
[5] Emura, T & Hsu, J-H 2020, 'Estimation of the Mann–Whitney effect in the two-sample problem under dependent censoring', Computational Statistics & Data Analysis, vol. 150, p. 106990.
[6] Evans, FH, Guthrie, MM & Foster, I 2020, 'Accuracy of six years of operational statistical seasonal forecasts of rainfall in Western Australia (2013 to 2018)', Atmospheric Research, vol. 233, p. 104697.
[7] Google Map 2020, Google Map, Google LLC (United States), viewed 10 June 2020, <https://www.google.com/maps>.
[8] Guo, S, Zhong, S & Zhang, A 2013, 'Privacy-preserving Kruskal–Wallis test', Computer Methods and Programs in Biomedicine, vol. 112, no. 1, pp. 135–145.
[9] Kamruzzaman, M, Beecham, S, Metcalfe, AV & Cai, W 2019, 'Granger causal predictors for maximum rainfall in Australia', Atmospheric Research, vol. 218, pp. 1–11.
[10] L. & Nicótina 2008, 'On the impact of rainfall patterns on the hydrologic response', Water Resources Research.
[11] Li, M & Gençay, R 2017, 'Tests for serial correlation of unknown form in dynamic least squares regression with wavelets', Economics Letters, vol. 155, pp. 104–110.
[12] Lim, EP, Hendon, HH, Anderson, DLT, Charles, A & Alves, O 2011, 'Dynamical, Statistical–Dynamical, and Multimodel Ensemble Forecasts of Australian Spring Season Rainfall', Monthly Weather Review, vol. 139, no. 3, pp. 958–975.
[13] Montazerolghaem, M, Vervoort, W, Minasny, B & McBratney, A 2016, 'Long-term variability of the leading seasonal modes of rainfall in south-eastern Australia', Weather and Climate Extremes, vol. 13, pp. 1–14.
[14] Pérez, NP, Guevara López, MA, Silva, A & Ramos, I 2015, 'Improving the Mann–Whitney statistical test for feature selection: An approach in breast cancer diagnosis on mammography', Artificial Intelligence in Medicine, vol. 63, no. 1, pp. 19–31.
[15] Ruxton, GD & Beauchamp, G 2008, 'Some suggestions about appropriate use of the Kruskal–Wallis test', Animal Behaviour, vol. 76, no. 3, pp. 1083–1087.
[16] Sadras, V, Roget, D & Krause, M 2003, 'Dynamic cropping strategies for risk management in dry-land farming systems', Agricultural Systems, vol. 76, no. 3, pp. 929–948.
[17] Sadras, VO, Lawson, C, Hooper, P & McDonald, GK 2012, 'Contribution of summer rainfall and nitrogen to the yield and water use efficiency of wheat in Mediterranean-type environments of South Australia', European Journal of Agronomy, vol. 36, no. 1, pp. 41–54.
[18] Web of Science 2020, Web of Science, Clarivate Analytics (United States), viewed 10 June 2020, <https://apps-webofknowledge-com>.
[19] Wu, J & Zhu, L 2011, 'Testing for serial correlation and random effects in a two-way error component regression model', Economic Modelling, vol. 28, no. 6, pp. 2377–2386.
[20] Ye, Q & Ahammed, F 2020, 'Quantification of relationship between annual daily maximum temperature and annual daily maximum rainfall in South Australia', Atmospheric and Oceanic Science Letters, pp. 1–8.
[21] Yin, Y 2020, 'Model-free tests for series correlation in multivariate linear regression', Journal of Statistical Planning and Inference, vol. 206, pp. 179–195.
[22] Zhao, F, Zhang, L, Chiew, FHS, Vaze, J & Cheng, L 2013, 'The effect of spatial rainfall variability on water balance modelling for south-eastern Australian catchments', Journal of Hydrology, vol. 493, pp. 16–29.