

DATA SCIENCE AND APPLICATIONS SESSION

# THE MANN-KENDALL TEST FOR WEATHER FORECASTING IN SMART CITIES

Miloš Todorov<sup>1\*</sup>, [0000-0001-6938-6974]

Ninoslava Tihi<sup>2</sup>, [0000-0001-8657-423X]

Marko Pavlović<sup>3</sup>, [0000-0002-5293-9533]

Suad Suljević<sup>3</sup> [0000-0003-2382-4734]

<sup>1</sup>The Faculty of Mathematics and Computer Science, Alfa BK University, Belgrade, Serbia

<sup>2</sup>The Higher Education Technical School of Professional Studies, University of Novi Sad, Novi Sad, Serbia

<sup>3</sup>MB University, Belgrade, Serbia

### Correspondence:

Miloš Todorov

e-mail: milos.todorov@alfa.edu.rs

#### Abstract:

In smart cities, predicting and responding to the coming of severe weather is one of the most important advantages that can be gained from precise weather forecasting. As a direct consequence of climate change, the planet is experiencing an increased frequency of heatwaves, storms, and floods more often than in the past. To be able to respond promptly, local government agencies must have access to precise weather predictions, which would help them in decision making and effectively delivering emergency aid, and, if necessary, evacuating areas that are at a larger risk of danger. In addition to this, people can use mobile apps to get weather forecasts in a short amount of time. This article employs the Mann-Kendall test to analyse monthly time series data, specifically focusing on weather parameters like temperature, soil moisture, and precipitation. This test will certainly provide useful insights to choose the best appropriate model for predicting drought-related parameters, which will be later visualized and presented to decision makers.

#### Keywords:

Statistical Analysis, The Mann-Kendall Test, Weather Forecasting, Smart Cities.

## INTRODUCTION

Time series forecasting is a statistical technique used to make predictions about future values of a time-dependent data set based on past observations [1]. It is commonly used in various fields such as economics, finance, and meteorology to forecast trends and make informed decisions. To integrate the prediction accuracies of machine learning methodologies, the framework for forecasting precipitation in smart cities utilises fuzzy logic. This procedure, which is referred to as fusion, was proposed by Atta-ur Rahman [2]. In Europe, the effects of soil moisture on seasonal temperature and precipitation prediction scores have been shown by Van Den Hurk [3].

One popular method for time series analysis is the Mann-Kendall test, which is a non-parametric test used to detect trends in time series data [4], [5]. This test analyses the data for monotonic trends, which are trends that consistently increase or decrease over time. The non-parametric Mann-Kendall test is suggested due to its robustness and lack of assumptions on data distribution.

It is used to ascertain the direction of the trend in a time series, whether it is upward or downward. It is a process that uses ranks instead of raw data values. It is suitable for data that does not have a normal distribution, has outliers or exhibits non-linear trends [4],[5]. This methodology is often used to estimate trends in a time series of mean, high and low flows, capturing their temporal variations and facilitating the prediction of future trends.

Another important technique in time series analysis is the Theil-Sen trend method, which is a robust method for estimating trends in time series data. This method is resistant to outliers and does not assume a specific distribution of the data, making it a reliable option for analysing time series data with irregular patterns or anomalies. The Theil-Sen trend method calculates the median of all possible slopes between data points, providing a robust estimate of the trend without being influenced by extreme values. Overall, time series trend analysis techniques, such as the Mann-Kendall test and Theil-Sen trend method, play a crucial role in analysing and predicting future trends in time-dependent data sets, helping researchers and decision-makers make informed decisions based on historical patterns and trends [6].

In addition, the Theil-Sen estimator is often used as a method for calculating trend slope often in combination with other trend analysis techniques, particularly the Mann-Kendall test [4], [5], [6], [7]. The statistically significant trends are determined using the Mann-Kendall test at a certain level of significance. The trend quantification and extension are thereafter carried out using the Theil-Sen estimator. This estimator is regarded as a reliable mathematical tool for estimating the extent of a trend. It does this by calculating the median value over all possible combinations of pairings within the given time series [7]. The Mann-Kendall test is particularly useful for identifying trends in time series data that may not be linear or follow a specific pattern. It is widely used in environmental studies to detect changes in variables such as precipitation patterns, temperature fluctuations and air pollution levels. The test calculates the Kendall rank correlation coefficient, which measures the strength and direction of any trend present in the data. A positive coefficient indicates an increasing trend, while a negative coefficient indicates a decreasing trend [4], [5].

The Theil-Sen estimator is a commonly used methodology for determining the slope of a trend, generally employed in combination with other methods of trend research, particularly the Mann-Kendall test[6]. The Mann-Kendall test and the Theil-Sen estimator are often used in many applications to estimate gradual trends, which are then utilized in the forecast of different weather parameters [8], [9], [10], [11].

In this paper, the application of the Mann Kendall test on soil moisture, precipitation and temperature will be discussed.

## 2. THE MANN- KENDALL TEST AND THEIL-SEN TREND ESTIMATOR

In this study, a two-stage technique is used to estimate trends. The first phase involves applying the nonparametric Mann-Kendall test to find trends. The second step involves modelling the trends using the Theil-Sen trend estimator. Figure illustrates this process.

Therefore, a two-step procedure for trend estimation is applied at the beginning of this research:

(Step 1) Identifying trends with the the Mann-Kendall test; (Step 2) Modelling trends using the Theil-Sen trend estimator.



Figure 1. A chart diagram for the trend detection in modelling the temperature time series

#### Step 1 – The Mann Kendall test

The Mann-Kendall trend test utilizes Kendall's statistic (S) to score each data point throughout the whole time series [4],[5], [12]:

$$S = \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} \operatorname{sign}(Q_i - Q_j)$$
(1)

where n represents the total number of time series elements and Q denotes the mean monthly or annual flow at time increments i and j, respectively. Furthermore, the value of sign() is determined by whether the time series member at time step i is greater than the corresponding member at time step j. If so, sign() returns +1; otherwise, it returns -1. If the time series member at time step i is equal to the corresponding member at time step j, sign() returns 0.

Variance of Kendall's statistics Var(S) is calculated as [4], [5], [12]:

$$Var(S) \begin{cases} \frac{1}{18} [n \cdot (n-1) \cdot (2n+5)], & if \ n = l \\ \frac{1}{18} [n \cdot (n-1) \cdot (2n+5) - \sum_{q=1}^{l} t_q \cdot (t_q-1) \cdot (2t_q+5)], & else \end{cases}$$
(2)

where *n* is a total number of time series members, *l* is number of groups which represent the set of sample data with the same values and  $t_q$  is number of observations (frequencies) in the  $q^{th}$  group.

Once Kendall statistic is determined, the Mann-Kendall test statistic (Z) for the time series is calculated as follows [4],[5], [12]:

$$Z = \begin{cases} \frac{S-1}{\sqrt{Var(S)}}, & \text{if } S > 0\\ 0, & \text{if } S = 0\\ \frac{S+1}{\sqrt{Var(S)}}, & \text{if } S < 0 \end{cases}$$
(3)

Since the Mann-Kendal statistic (Z) conforms to a normal distribution, the probability density function for a normal distribution N(0,1) is defined as:

$$f(Z) = \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-Z^2}{2}} \tag{4}$$

The null hypothesis is accepted when the probability value p is greater than 0.05 and there is no discernible trend in the time series. Conversely, at a significance level of p≤0.05 the alternative hypothesis is accepted indicating the presence of a trend in the time series which may be negative or positive.

#### Step 2 - The Theil-Sen trend estimator

If the Mann-Kendall test statistic (Step 1) suggests that a substantial trend continues to exist in the time series, the Theil-Sen trend estimator is utilized to calculate the linear trend parameters including the intercept and slope. Consequently, the extent of the linear trend in time series data is determined by averaging the pairwise slopes [6], [12].

$$f(Z) = \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{Z^2}{2}} \tag{5}$$

However, when there are statistically non-significant trends, there is no need to use the Thel-Sen trend estimator since the linear trend slope is 0 (Figure ).

```
#Importing libraries
import numpy as np
import pandas as pd
import pymannkendall as mk
#Reading datasets
padavine=pd.read_excel('padavine.xlsx')
padavine.head()
temperatura=pd.read_excel('temperatura.xlsx')
temperatura.head()
vlaznost_zem=pd.read_excel('vlaznost_zem.xlsx')
vlaznost_zem.head()
#Applying Mann-Kendall test
padavineTest = mk.original_test(padavine['Vrednost'])
print(padavineTest)
temperaturaTest = mk.original_test(temperatura['Vrednost'])
print(temperaturaTest)
vlaznost_zemTest = mk.original_test(vlaznost_zem['Vrednost'])
print(vlaznost_zemTest)
```

#### Listing 1. A Python program for applying the Mann-Kendall test

## 3. DATASET AND FRAMEWORK FLOW

Data were gathered from a weather station located near Novi Sad from January 2014 to December 2020. The data collected by this network station comes from three sensors that measure the temperature of the air, the amount of precipitation, and the moisture content of the soil. Every hour of the day, measurements are carried out.

Following the retrieval of the data in Excel format from the web portal, the data are transformed and then stored in a tabular data structure called data frame for further manipulation. The preparation process continues with filtering and aggregating using the "mean" and "sum" functions to acquire the monthly average values and finally sorting the obtained data by months and years and exporting them in an Excel file.

#### 3.1. CODE ALGORITHM IN PYTHON

Several Python programming modules that are suited for this research were used to construct specialised scripts that were produced expressly for the aim of this study. It was through the use of the Python program script that integrates a few Python packages that the outcomes were obtained [12].

Listing 1 shows an example of a small program written in the Python programming language. It contains importing libraries, reading dataset and applying the Mann-Kendall test over loaded data.

## 4. RESULTS AND ANALYSIS

The following table presents the outcomes derived from the Python script.

The Mann-Kendall test provide a combination that includes the following components:

- Trend: indicates the trend (whether it is rising, decreasing, or there is no trend at all);
- **p-value**: p-value of the significance test;
- Slope: Theil-Sen estimator/slope;

• Intercept: the intercept of the Kendall-Theil Robust Line, for the seasonal test, with the whole period cycle regarded as a unit time step.

As shown in Table 1, trend analysis provides results for three meteorological variables: temperature, precipitation and soil moisture. No significant trend is present in the temperature time series according to the obtained p value from the Mann-Kendal test (p=0.6569), meaning there's no clear increasing or decreasing pattern. No significant trend has also been found in the precipitation time series (p= 0.1529) and therefore it can be concluded that this time series is relatively stable over time. However, in the soil moisture time series the Mann-Kendall test has detected a decreasing trend with a highly significant p value (p=3.28×10-5), indicating a real downward trend in the soil moisture levels. The slope represents the rate of change of each variable over time. The temperature and precipitation time series have a slight increase but not very significant slope, but the soil moisture has a significant decrease meaning the soil moisture is consistently declining over time. The intercept represents the expected value of the variable at the start of the dataset. Temperature started around 12.37 °C, precipitation started around 39.05 mm while soil moisture started around 25.54% but it is decreasing significantly.

## 5. CONCLUSION

Smart cities benefit most from accurate weather forecasting because they can prepare for severe weather. Heatwaves, storms, and floods have risen globally owing to climate change. Local government agencies can plan, give emergency help, and evacuate high-risk regions with accurate weather forecasts. Mobile applications can quickly provide weather predictions, allowing homeowners to change their plans and take safety precautions.

Accurate weather forecasting in intelligent cities goes beyond crisis management. Transportation, tourism, and agriculture depend on accurate weather predictions. Agricultural professionals can predict crop yields, control irrigation, and safeguard crops from weather.

Table 1. Output results of applying the Mann-Kendall test in Python code

Time Series	Trend	p-value	Slope	Intercept
Temperature	No trend	0.6568582113023143	0.0138589659774	12.36958549592
Precipitation	No trend	0.1528558365907872	0.1939764492753	39.05002264492
Soil moisture	Decreasing	3.284048567153697e-05	0.0627024599945	25.53967055026

By optimising routes and schedules, transport firms can reduce delays and improve customer satisfaction. Being aware helps people plan for outside events and activities, reducing weather-related difficulties.

The Mann-Kendall test is an extremely useful instrument that can be used to recognise monotonic patterns in time series data. This makes it very pertinent for the field of environmental research. When it comes to estimating trends in time series data, the Theil-Sen trend technique offers a straightforward and trustworthy methodology. After conducting the tests, the lack of a trend in the temperature and precipitation time series suggests that other external factors (e.g., local climate variability) may be influencing them. The decrease in soil moisture could indicate increasing drought conditions, reduced groundwater recharge, or changes in land use affecting water retention.

# REFERENCES

- K. C. Tan, 'Trends of rainfall regime in Peninsular Malaysia during northeast and southwest monsoons', *J. Phys. Conf. Ser.*, vol. 995, p. 012122, Apr. 2018, doi: 10.1088/1742-6596/995/1/012122.
- [2] A. Rahman et al., 'Rainfall Prediction System Using Machine Learning Fusion for Smart Cities', *Sensors*, vol. 22, no. 9, p. 3504, May 2022, doi: 10.3390/ s22093504.
- B. Van Den Hurk, F. Doblas-Reyes, G. Balsamo, R. D. Koster, S. I. Seneviratne, and H. Camargo, 'Soil moisture effects on seasonal temperature and precipitation forecast scores in Europe', *Clim. Dyn.*, vol. 38, no. 1–2, pp. 349–362, Jan. 2012, doi: 10.1007/s00382-010-0956-2.
- [4] H. B. Mann, 'Nonparametric Tests Against Trend', *Econometrica*, vol. 13, no. 3, p. 245, Jul. 1945, doi: 10.2307/1907187.
- [5] M. G. Kendall, 'A NEW MEASURE OF RANK CORRELATION', *Biometrika*, vol. 30, no. 1–2, pp. 81–93, Jun. 1938, doi: 10.1093/biomet/30.1-2.81.
- [6] P. K. Sen, 'Estimates of the Regression Coefficient Based on Kendall's Tau', J. Am. Stat. Assoc., vol. 63, no. 324, pp. 1379–1389, Dec. 1968, doi: 10.1080/01621459.1968.10480934.
- [7] I. Krtolica, M. Todorov, S. Prohaska, and M. Stojković, 'Annual and Low-Flow Trends in Serbia', *J. Hydrol. Eng.*, vol. 29, no. 3, p. 05024008, Jun. 2024, doi: 10.1061/JHYEFF.HEENG-6030.
- [8] M. Stojković, S. Prohaska, and J. Plavšić, 'Stochastic structure of annual discharges of large European rivers', *J. Hydrol. Hydromech.*, vol. 63, no. 1, pp. 63–70, Mar. 2015, doi: 10.1515/johh-2015-0009.

- [9] M. Gocic and S. Trajkovic, 'Analysis of changes in meteorological variables using Mann-Kendall and Sen's slope estimator statistical tests in Serbia', *Glob. Planet. Change*, vol. 100, pp. 172–182, Jan. 2013, doi: 10.1016/j.gloplacha.2012.10.014.
- [10] K. H. Hamed and A. Ramachandra Rao, 'A modified Mann-Kendall trend test for autocorrelated data', *J. Hydrol.*, vol. 204, no. 1–4, pp. 182–196, Jan. 1998, doi: 10.1016/S0022-1694(97)00125-X.
- [11] H. X. Do, S. Westra, and M. Leonard, 'A global-scale investigation of trends in annual maximum streamflow', *J. Hydrol.*, vol. 552, pp. 28–43, Sep. 2017, doi: 10.1016/j.jhydrol.2017.06.015.
- [12] Md. Hussain and I. Mahmud, 'pyMannKendall: a python package for non parametric Mann Kendall family of trend tests.', *J. Open Source Softw.*, vol. 4, no. 39, p. 1556, Jul. 2019, doi: 10.21105/joss.01556.