

# Cellular time series: a data structure for spatio-temporal analysis and management of geoscience information

Sina Nabaei and Bahram Saghafian

## ABSTRACT

Geoscientists are continuously confronted by difficulties involved in handling varieties of data formats. Configuration of data only in time or space domains leads to the use of multiple stand-alone software in the spatio-temporal analysis which is a time-consuming approach. In this paper, the concept of cellular time series (CTS) and three types of meta data are introduced to improve the handling of CTS in the spatio-temporal analysis. The data structure was designed via Python programming language; however, the structure could also be implemented by other languages (e.g., R and MATLAB). We used this concept in the hydro-meteorological discipline. In our application, CTS of monthly precipitation was generated by employing data of 102 stations across Iran. The non-parametric Mann–Kendall trend test and change point detection techniques, including Pettitt's test, standard normal homogeneity test, and the Buishand range test were applied on the generated CTS. Results revealed a negative annual trend in the eastern parts, as well as being sporadically spread over the southern and western parts of the country. Furthermore, the year 1998 was detected as a significant change year in the eastern and southern regions of Iran. The proposed structure may be used by geoscientists and data providers for straightforward simultaneous spatio-temporal analysis.

**Key words** | big data, cellular time series, meta data, precipitation, Python, spatio-temporal

### Sina Nabaei

Young Researchers and Elites Club, Science and Research Branch, Islamic Azad University, Tehran, Iran

### Bahram Saghafian (corresponding author)

Department of Civil Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran  
E-mail: [b.saghafian@gmail.com](mailto:b.saghafian@gmail.com)

## INTRODUCTION

The field of big data has dramatically expanded in the field of information and communications technology (ICT) since 2010; it has evolved in early stages in water and climate sciences (Chen & Han 2016). The term 'big data' appears simple, but its technical meaning is somewhat ambiguous. It is commonly used to refer to massive datasets, as well as to complexity beyond the capacity of conventional computing tools (Snijder *et al.* 2012).

The big data characteristics include 'high volume' (e.g., data volume, collection, retention), 'high velocity' dealing with speed of data collection and processing (e.g., data caches, point-to-point data routing), and 'high variety' dealing with categories (e.g., data formats and meta data) (Laney 2009). Discovery of knowledge from a high volume dataset is a challenging systematic issue (e.g., utilizing

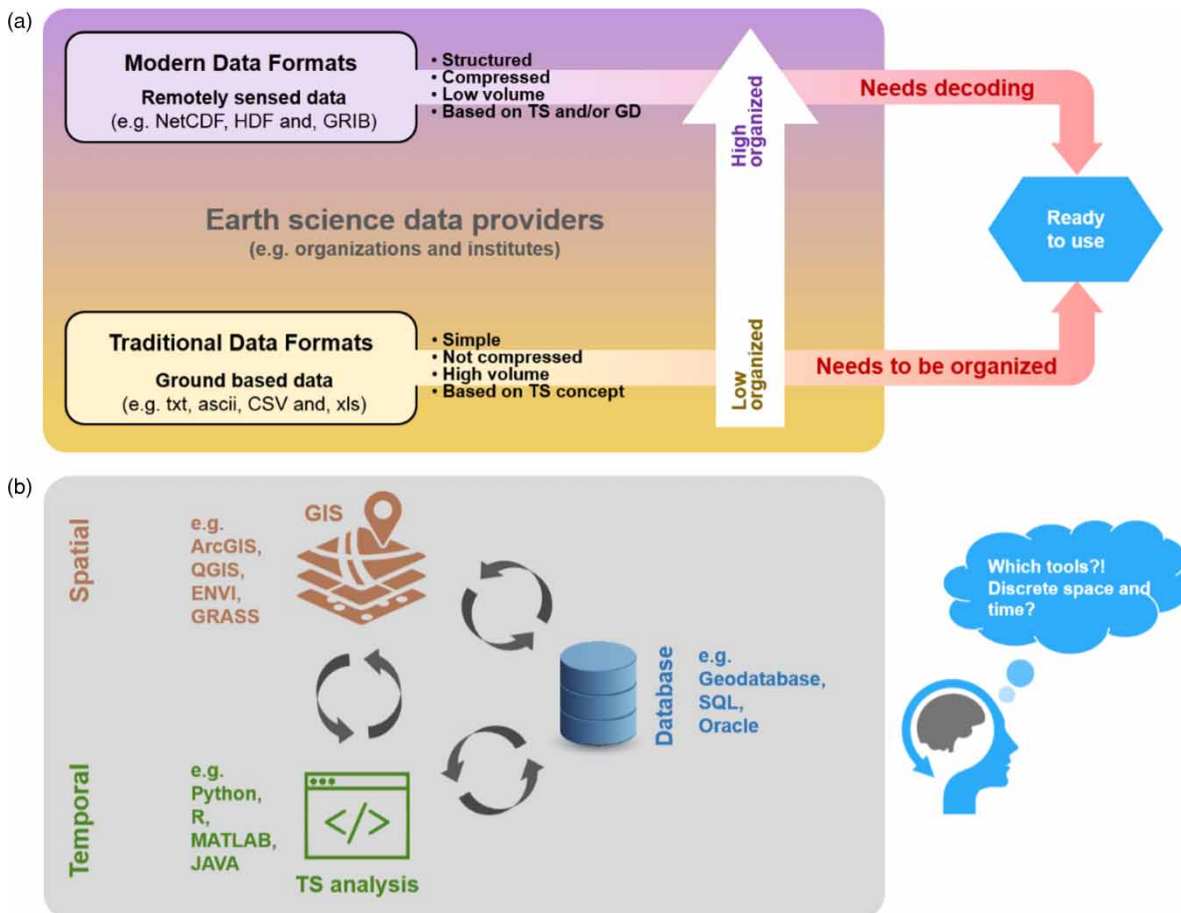
the existing hardware, data formats, models, and methods) (National Research Council 2013).

Some disciplines, such as environmental and earth sciences, natural resources, meteorology and hydrometeorology involve dynamic variables (e.g., Simonovic 2009; Abuzied & Mansour 2018; Nodoushan & Shakibaeinia 2018). In this regard, the spatio-temporal analysis of well-structured data remains a challenging issue (Harpham & Danovaro 2015). Furthermore, large volumes of hydrometeorologic data have been recorded and released (e.g., data from sensors, remote sensing (RS), Earth observation, and internet of things) in recent decades and push the water and climate sectors into the inevitable big data field. Accordingly, for analysis of big data, there is need for a structure that, in addition to

handling the time dimension, can accommodate the space dimension (Liu *et al.* 2016).

In recent years, global RS data have been stored by a number of structured data models, such as advanced self-describing network Common Data Form (netCDF), Hierarchical Data Format (HDF), and Gridded Binary (GRIB) that handle time and space dimensions (Chen & Han 2016). Recent generations of these data models (e.g., netCDF-4 and HDF5) can store more complex datasets (high volume and different variety) (Rew *et al.* 2006). netCDF and HDF are commonly used for storing big gridded data (GD), but to extract the values of variables, there is the need for additional time-consuming processing (decoding). In addition, the main problem (in data handling) arises when the spatio-temporal analysis of geoscience data is commonly performed by the end-users who are not trained ICT scientists (Ward *et al.* 2018).

RS data assessment leads to a new generation of challenges. Data providers publish modern geoscience datasets (e.g., satellite products) in high-compressed formats (e.g., HDF). For the RS product assessment (e.g., precipitation), there is a need to compare them with ground-based measures. On the other hand, historical ground-based data have been recorded and published since much earlier times and are usually accessible in traditional formats (e.g., txt, CSV, xls, and xlsx). Thus, there is a big gap (in terms of published formats) between structured data (e.g., RS and global reanalyzed datasets shared in netCDF or HDF formats) and ground-based data (Figure 1(a)). Consequently, synchronization and sorting of multiple data formats has become a crucial and inevitable time-consuming step in the process of RS data assessment. Hence, there is a need for a common flexible solution which covers the variety of traditional and modern data formats.



**Figure 1** | (a) Varieties of data formats and challenges in data analysis. (b) Unstructured traditional procedure for simultaneous space and time analysis.

Figure 1(b) addresses another issue. Traditional spatio-temporal approaches (combining the stand-alone software) and transferring the results among them are not sufficient for huge data analysis with speed and accuracy. Geographic information systems (GIS) software (e.g., ArcGIS) are mainly based on the space domain (GD structure). Also, time series (TS) analysis software (e.g., SPSS) assists researchers to investigate the dynamic of data along the time dimension. Nevertheless, there is not a clear concept/guideline for how to combine the bunch of software due to spatio-temporal analysis (Figure 1(b)). To improve the procedure of spatio-temporal analysis, a new structure may be required. One of the main objectives of this paper is to provide a data structure/concept to improve the speed of spatio-temporal analysis. In this study, we focus on the ground-based data.

Different studies have been conducted in the field of data structure. Kruger *et al.* (2011) studied the utility of meta data in the query of NEXRAD data. They showed this is an efficient approach for selecting data subsets in time series. It is based on the Hydro-NEXRAD system, which contains a vast archive of online radar-centric data. However, their study was limited to a specific online dataset and not customizable by end-users independently. Liu *et al.* (2016) proposed the SciDB server-based data query structure. SciDB was developed for the NetCDF-4 format. It is implemented on the online platform and assists data query procedure. SciDB is a server-based structure and not developed for offline data sources (and does not cover traditional formats). Seo *et al.* (2019) studied NEXRAD dataset and introduced a pilot infrastructure for data management. It is based on meta data query and facilitates searching/converting procedures in precipitation datasets. This is a specified study on NEXRAD data and not developed for all types of geo-datasets (e.g., offline and user-defined).

The analysis of climatic and meteorological fluctuations are being conducted across different regions around the world (e.g., Yao *et al.* 2014; Pascale *et al.* 2015; Shirvani 2015; Abidi *et al.* 2017; Gundogdu 2017; Aminyavari *et al.* 2018; Kolachian & Saghafian 2019). As climatic variables fluctuate over time and in space, numerous studies have been reported to explore the spatio-temporal nature of climate (e.g., Toros 2012; Sen Roy & Rouault 2013; Delavau *et al.* 2015; Luković *et al.* 2015). In addition, climate variables affect hydrology and water resources. With the development

of better computer-based algorithms and methods (such as machine learning, data mining, and artificial intelligence), many studies have been performed to simulate and forecast hydro-climatic phenomena (e.g., Chuntian & Chau 2002; Wu & Chau 2011; Chau 2017; Ali Ghorbani *et al.* 2018; Moazenzadeh *et al.* 2018; Nabaei *et al.* 2019; Yaseen *et al.* 2019).

Trend and change point tests are common temporal TS analysis, prevalently conducted on point data (e.g., Shi *et al.* 2014; Khalili *et al.* 2016). However, the study of trend and change point across space enhances our understanding of how TS behave spatially. At a glance, spatio-temporal climate data studies could be categorized into four groups based on the type of input data and results representation. The first group, with a huge number of reported studies, deals with *in-situ* (point) data. Inputs are typically provided by meteorological stations (e.g., Duhan & Pandey 2013; Santos & Fragoso 2013; de Luis *et al.* 2014; Goyal 2014; Hosseinzadeh Talaei *et al.* 2014; Shi *et al.* 2015; Singh *et al.* 2015; Ghasemi 2015; Li *et al.* 2015; Longobardi & Mautone 2015; Powell & Keim 2015; Javari 2017; Roushangar *et al.* 2018). The results of such studies lack spatial dimension and are specific to a certain location.

In the second study group, station-based TS analysis is transposed into a spatially continuous domain. This is usually carried out via interpolation techniques applied on output statistics (e.g., Dinpashoh 2006; Fathian *et al.* 2014; Hosseinzadeh Talaei *et al.* 2014; Abolverdi *et al.* 2016; Javari 2016; Minaei & Irannezhad 2016). The third type regionalizes a study area by inputting main (e.g., precipitation) or auxiliary (e.g., outputs of statistical analysis) variables via application of different classification techniques such as clustering (e.g., Dinpashoh *et al.* 2004; Modarres 2006; Raziie *et al.* 2008; Modarres & Sarhadi 2011; Fazel *et al.* 2017). The last study group deploys RS (e.g., radar and imagery; Khodadoust Siuki *et al.* 2017; Sharifi *et al.* 2018) or GD for spatio-temporal analysis (e.g., Sarmadi & Shokoochi 2015; Fallah *et al.* 2017). This procedure is more recent and outcome is presented in GD forms.

Efficient management of geoscience data is one of the major concerns of data scientists. As mentioned previously, most of the studies were limited to a particular format (e.g., NetCDF) or developed based on a specific platform (e.g., Hydro-NEXRAD or online services). It can be hindered in achieving a common data structure to the vast range

of applications and multi-formats. To overcome these obstacles/solidity, a new data concept with the following specifications could be advantageous: (1) simple enough and applicable by a wide range of researchers; (2) flexible enough to cover the multi-format datasets (i.e., filling the gap between *in-situ* and RS data); (3) customizable (i.e., extendable) structure for generating and managing dataset; (4) could be implemented in a single platform due to integrating all the procedures of spatio-temporal analysis (not rely on multiple stand-alone software).

In this study, we used Python 2.7 which is compatible with the GNU General Public License (GPL). Python is one of the high-level languages and is well-suited for geoscience disciplines. Other languages (e.g., R, MATLAB, Java, and Scala) could also be used to generate and apply this concept. To cover the desired goals, we introduce the structure/concept of cellular time series (CTS), spatial meta data (SMD), temporal meta data (TMD), and meta CTS (mCTS) in this paper. Herein, CTS is defined as a set of connected cells over space, each of which carry a time series.

## METHODS

At first, the generation procedure of CTS is outlined. Next, three types of meta data are proposed for managing (query) of CTS information. Also, some well-known statistical analysis is reviewed which we applied in this paper across Iran. The details of Mann–Kendall (MK) trend test are described, followed by three different change point detection methods including Pettitt's test, standard normal homogeneity test (SNHT), and the Buishand range test.

### Generation of CTS

This paper proposes a conceptual structure for spatio-temporal applications where one needs to convert *in-situ* time series into CTS. Figure 2(c) shows three dimensions of the CTS that include X (longitude), Y (latitude), and T (time). Numerous data generation/prediction methods (e.g., IDW, Bayesian kriging, co-kriging, spline, and also numerical weather generation models) could be implied to create spatial (X-Y) layouts (Figure 2(a)). Finally, two-dimensional (2D) matrices in space were merged to form a three-dimensional (3D) CTS (Figure 2(b) and 2(c)). In this study, however,

ground-based data were the basis for CTS as such data are presumed as a reference data source to assess/bias correct other global datasets (e.g., RS and reanalyzed data). Similarly, RS and reanalyzed data (usually with NetCDF, GRIB, and HDF formats) could also be used to generate CTS.

### Spatial meta data

This is a 2D matrix in space dimension (X and Y). In this case, the SMD values are constant over time and attributed to all the temporal slices involved in the CTS. For example, digital elevation model, coordinates (longitude and latitude), the surface of a region (e.g., states, catchments, and climate zones), and land use can be categorized as SMD. SMD can help to perform spatial analysis on the CTS, for example, the query of the CTS values in a particular area. Afterwards, any spatial calculations (e.g., sum, average, standard deviation, and entropy) can be performed via SMD (Figure 2(d) and 2(g)).

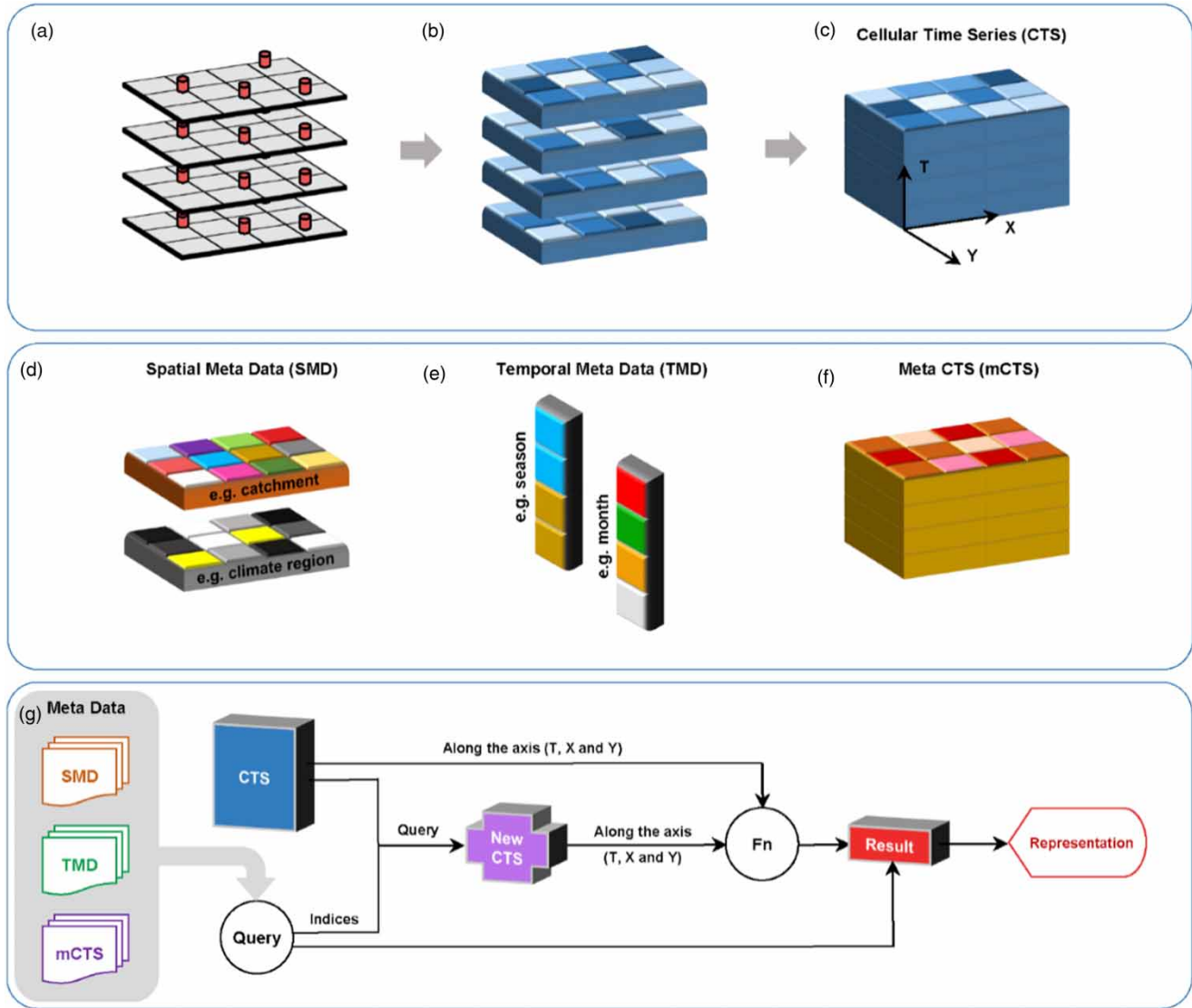
### Temporal meta data

This is a vector (along the time dimension). The TMD values are constant over the space and attributed to all the spatial slices involved in the CTS. For example, year, month, day, hour, minute, seasons, and different calendars (e.g., solar, Julian, and lunar) may be categorized as TMD. TMD can help to perform temporal analysis on the CTS. For example, extracting the values of specific days and converting daily precipitation to monthly could be performed via TMD (Figure 2(e) and 2(g)).

### Meta CTS

This is a 3D matrix (X, Y, and T) and could contain SMD and TMD features. The mCTS values can vary in both time and space. Any dynamic (in time and space) variable which could be used in the query of CTS information is assumed as mCTS. Different variables (e.g., temperature and humidity) and drought indices (e.g., SPI, SPEI, Palmer) may be inferred as mCTS. For example, the query of drought events (based on a specific index threshold) could be performed by mCTS (Figure 2(f) and 2(g)).

Figure 2(g) illustrates the general procedure (interactions) of CTS and meta data application. The procedure



**Figure 2** | (a) *In-situ* data in time layers, (b) spatially gridded maps in time, and (c) assembled 3D CTS, (d) spatial meta data, (e) temporal meta data, (f) meta CTS, and (g) a general procedure of using CTS and meta data query for spatio-temporal analysis.

may be applied in most of the programming language. It is based on the meta data queries, extracting a new CTS, and applying functions ( $F_n$ ) along the desired dimension.

### MK trend test

MK trend test (Mann 1945; Kendall 1975) is a rank-based non-parametric test, developed to recognize monotonic increasing or decreasing trend. MK statistics could be used in a hypothesis test to determine significant trend. There are two main statistics in the MK. The first is  $S$  and

calculated by inputting  $x_i$  and  $x_j$  as a sequence of TS values as follows:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i) \quad (1)$$

$$\text{sgn}(x_j - x_i) = \begin{cases} +1 & \text{if } x_j - x_i > 0 \\ 0 & \text{if } x_j - x_i = 0 \\ -1 & \text{if } x_j - x_i < 0 \end{cases} \quad (2)$$

$\text{Sgn}$  is the sign function and returns conditional values. The second statistic is  $Z$  which refers to the standard

normal deviate and is calculated as:

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

$Var(S)$  is the variance of  $S$  and is determined as follows:

$$Var(S) = \frac{n(n - 1)(2n + 5) - \sum_{i=1}^g t_i(t_i - 1)(2t_i + 5)}{18} \quad (4)$$

In the above equation,  $n$ ,  $g$ , and  $t_i$  are the number of samples, number of tied groups, and number of data values in the  $i$ th group, respectively (see Hirsch *et al.* 1982).

Two tail test with null  $H_0$  (no trend) and alternative  $H_A$  (rejected  $H_0$ ) hypothesis is then applied. The standard normal cumulative values are determined for  $\alpha = 5\%$  and  $10\%$  significant levels. Owing to the two tail test, if the  $Z$  is greater or smaller than critical values ( $Z_{1-\alpha/2}$  and  $-Z_{1-\alpha/2}$ , respectively), the  $H_0$  is rejected and trend is significant.

### Pettitt's test

Pettitt (1979) proposed a non-parametric test to detect a single change point which is commonly applied on climate and hydrologic TS. The null hypothesis is  $H_0$  (no change) while the alternative is defined as  $H_A$  (rejected  $H_0$ ). The change point occurs at  $t$ . The statistic  $K_T$  refers to the detected change point.

$$K_T = \max|U_{t,T}| \quad (5)$$

$$U_{t,T} = \sum_{i=1}^t \sum_{j=t+1}^T \text{sgn}(x_i - x_j) \quad (6)$$

$T$  is number of total samples in the series. The variable  $t$  divides the series into two subseries and the set of  $U_{t,T}$  is determined.

$$p \approx 2 \exp\left(\frac{-6K_T^2}{T^3 + T^2}\right) \quad (7)$$

If  $p$  is smaller than confidence level,  $K_t$  is significant and the corresponding  $t$  is detected as a change point. In this

study, the 5% confidence level is selected for two tail hypothesis test.

### SNHT

SNHT, proposed by Alexandersson (1986), detects a single change point. Statistics are computed as:

$$T_k = kZ_1^2 + (n - k)Z_2^2 \quad (1 \leq k \leq n) \quad (8)$$

$$Z_1 = \frac{1}{k} \sum_{i=1}^k \frac{x_i - \bar{x}}{\sigma} \quad (9)$$

$$Z_2 = \frac{1}{n - k} \sum_{i=k+1}^n \frac{x_i - \bar{x}}{\sigma} \quad (10)$$

The variable  $k$  refers to the change point and  $n$  is the total sample size. Maximum value of  $T_k$  is a candidate of significant change point and should be checked by hypothesis test.  $\bar{x}$  and  $\sigma$  are mean and standard deviation of series.

### Buishand's range test

Buishand (1982) used cumulative deviation from mean ( $\bar{x}$ ) to detect change point ( $k$ ). Adjusted partial sum ( $S_k$ ) is calculated as follows:

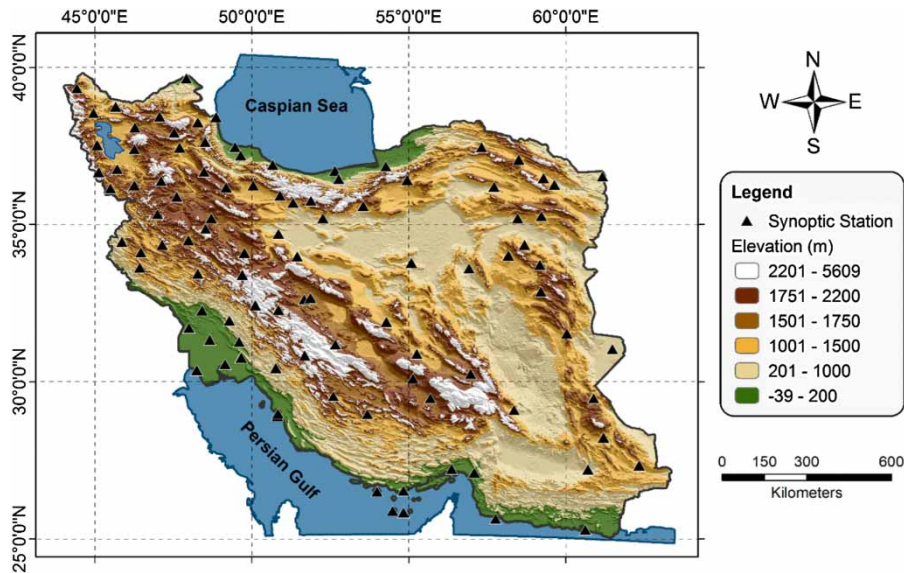
$$S_K = \sum_{i=1}^k (x_i - \bar{x}) \quad (1 \leq i \leq n) \quad (11)$$

$$R = \frac{\text{Max}(S_k) - \text{Min}(S_k)}{\sigma} \quad (12)$$

The change point was evaluated by comparing  $R/\sqrt{n}$  with the critical value at the significant level.

### APPLICATION OF CTS

Iran is a country in western Asia between 44°00'E–63°25'E longitude and 25°00'N–38°40'N latitude with an area of about 1,648,000 km<sup>2</sup>. The Caspian Sea in the north and the Persian Gulf in the south border some 2,000 km of coast-line in total (Figure 3). Zagros and Alborz are the two main



**Figure 3** | Topographic map and location of selected synoptic stations in Iran.

mountain ranges in the country. Also, two vast central deserts are known as the Dasht-e Lut and the Dasht-e Kavir. The climate varies greatly across the country, depending on topography among other things.

Monthly precipitation of some 102 synoptic stations were obtained from the Iranian Meteorological Organization. The location of the stations is shown in Figure 3. The study period covers 1987–2014 as the common period of quality precipitation data. In order to provide a basis for seasonal analysis, December is set as the first month and so on; accordingly, the seasons are set as follows: winter = DJF (Dec, Jan, and Feb), spring = MAM (Mar, Apr, and May), summer = JJA (Jun, Jul, and Aug), and autumn = SON (Sep, Oct, and Nov). The CTS precipitation was generated using 336 monthly precipitation layers at  $0.5^\circ \times 0.5^\circ$  spatial resolution. Precipitation maps were generated via inverse distance weighted (IDW) interpolation method with exponent 2. CTS is not limited to IDW interpolation and could be loaded with many other techniques, even optimization tools. Afterward, the spatio-temporal trend and change point analyses were performed on the generated CTS.

## RESULTS AND DISCUSSION

In this section, trend and change point analyses were applied on the CTS at different time scales. Results,

extracted in cellular form, are illustrated at different significant levels.

### Precipitation distribution

Spatial distribution of annual and seasonal average precipitation are shown in Figure 4. Similarly, monthly precipitation maps were also generated and are summarized in Table 1.

As shown in the maps, northern and western Iran (along Alborz and Zagros mountain chains) receive the highest precipitation depths (Figure 4). The average annual precipitation varies from 20 mm (in Dasht-e Lut desert) to 1,410 mm (along the northern coastline). The overall average values closely match those reported by previous studies (e.g., Dinpashoh *et al.* 2004; Modarres & Sarhadi 2011; Raziéi *et al.* 2014).

Summer rain mainly occurs along the Caspian Sea in the north where over 80% of the total summer precipitation in Iran is received (Figure 4). Average annual precipitation is 242.9 mm over the study period, 75% of which occurs in winter and spring (Table 1). Total precipitation from November to April constitutes about 83% of annual precipitation (Figure 5). Clearly, central and southeastern regions receive the lowest precipitation (Figure 4) and are classified as arid and semi-arid climates (Modarres & Sarhadi 2009).

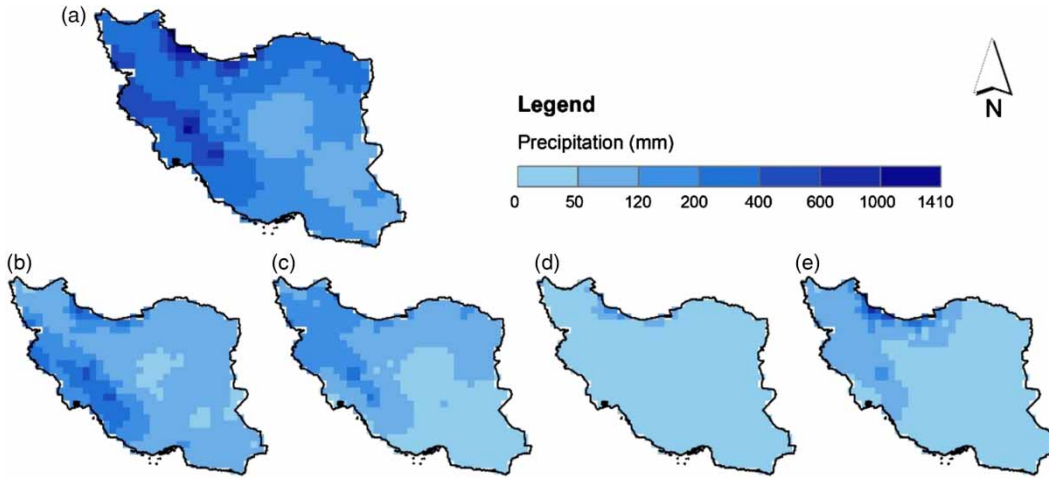


Figure 4 | Spatial distribution of long-term average precipitation over Iran at different time scales: (a) annual, (b) winter, (c) spring, (d) summer, and (e) autumn.

Table 1 | Characteristics of average annual, seasonal, and monthly precipitation in Iran

Annual Precipitation (mm)	Seasonal			Monthly			
	Season	Precipitation (mm)	Contribution (%)	Month	Precipitation (mm)	Contribution per season (%)	Contribution per year (%)
242.9	Winter	109.5	45.1	Dec	37.4	34.1	15.4
				Jan	38.1	34.8	15.7
				Feb	34.0	31.1	14.0
				Sum	109.5	100.0	-
	Spring	78.9	32.5	Mar	40.6	51.4	16.7
				Apr	25.7	32.6	10.6
				May	12.6	16.0	5.2
				Sum	78.9	100.0	-
	Summer	10.4	4.3	Jun	4.2	40.4	1.7
				Jul	3.0	28.8	1.2
				Aug	3.2	30.8	1.3
				Sum	10.4	100.0	-
	Autumn	44.1	18.1	Sep	5.3	12.0	2.2
				Oct	13.3	30.1	5.5
				Nov	25.5	57.9	10.5
				Sum	44.1	100.0	-

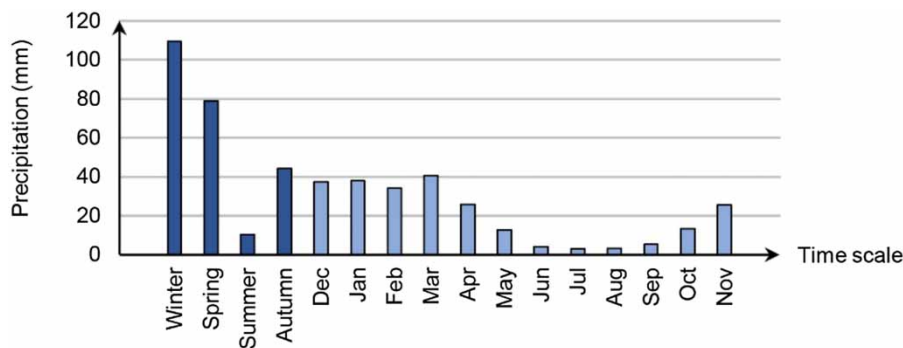
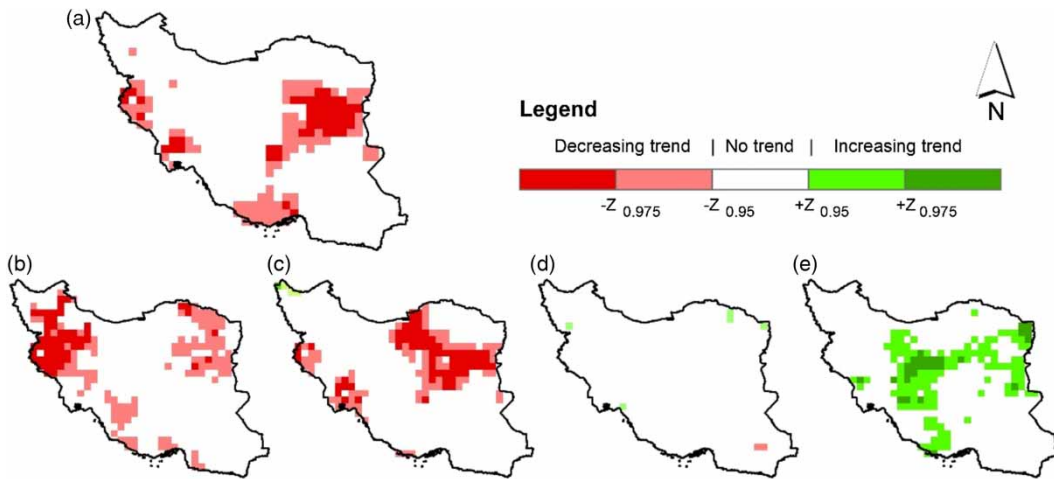


Figure 5 | Spatially averaged precipitation at seasonal and monthly time scales.





**Figure 6** | Spatial distribution of precipitation trend at 5% and 10% significant levels based on the MK test at different time scales: (a) annual, (b) winter, (c) spring, (d) summer, and (e) autumn.

### Significant trend

Hypothesis tests were applied to detect significant trend at 5% and 10% levels. Two tail standard normal cumulative values are  $Z_{0.975}$  and  $Z_{0.95}$ , respectively. The Z values were applied as a threshold for the P-values of MK statistic to detect cells with significant trend. Spatial distribution of significant trend in annual and seasonal time scales are shown in Figure 6. Also, the MK test was applied at monthly time scale as shown in Figure 7.

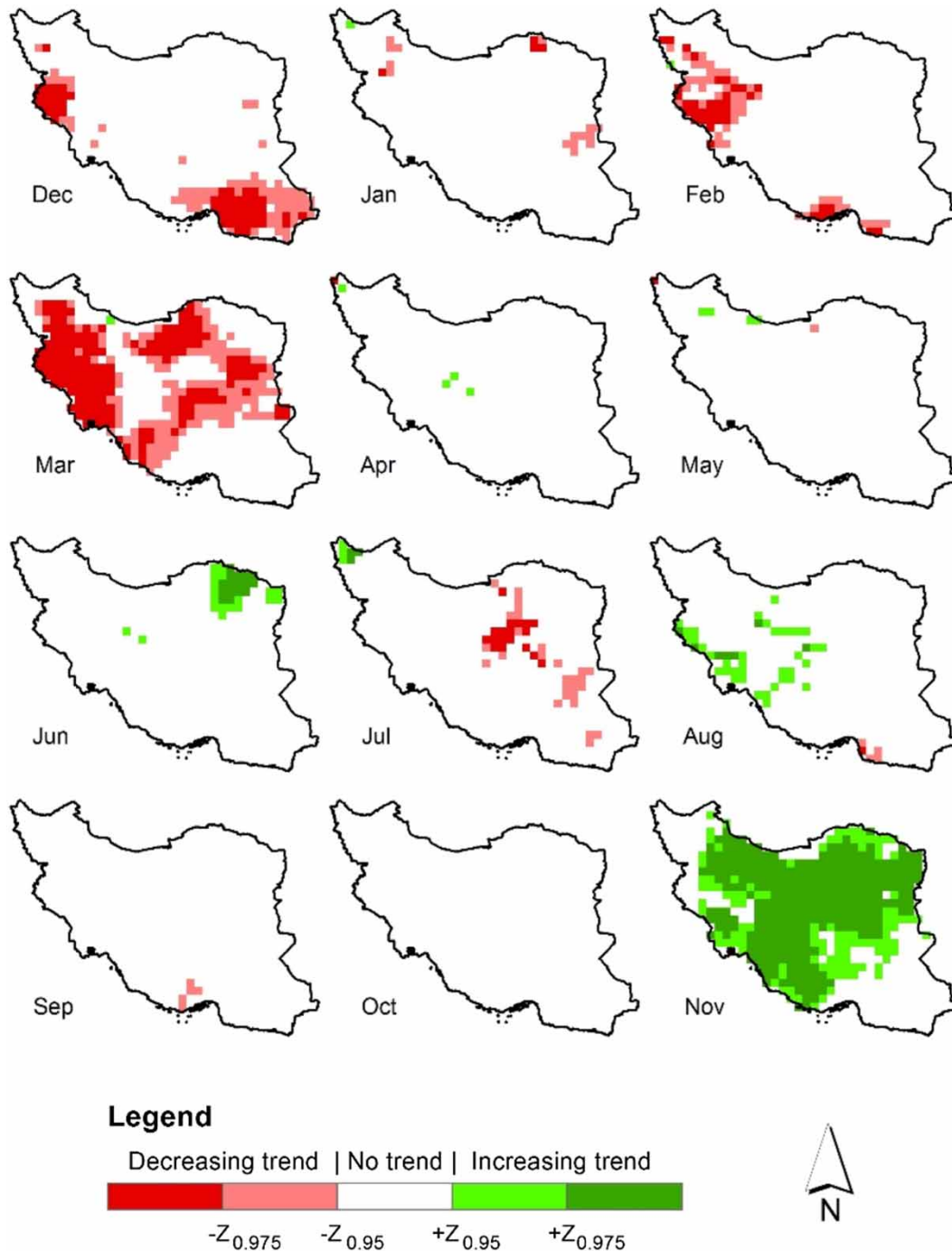
Annual precipitation trend decreases in the east, and in small portions of the west and south (Figure 6(a)). Decomposition of annual trend revealed that winter and spring fluctuations are the main sources of downward trend (Figure 6(b) and 6(c)), while the insignificant contribution of summer precipitation (about 4%) and upward incremental trend in autumn precipitation, appearing in the central and sporadic regions in eastern and southern regions (Figure 6(e)), balance the overall annual trend.

The monthly precipitation trend maps (Figure 7) shows that downward winter trend occurred in December (mostly in the southeast) and February (mostly in the west). Moreover, spring is affected by a downward March precipitation trend. More detailed analysis on March precipitation revealed that areas with 5% significant trend are concentrated in the west of Iran. Also, there are areas in the north, centre, and east with decreasing trend. The increasing autumn trend is mainly affected by November precipitation

(Figure 7). Trends are observed in most regions at 5% significance level, except in parts of the northwest and southeast.

Figure 8 summarizes the trend analysis, in terms of the percentage of the country affected, at 5% and 10% significant levels. About 22.4% of areas have negative annual trend, out of which 7.9% is significant at 5% level. Areas subject to winter, spring, summer, and autumn precipitation trend cover 26.7% (decreasing), 23.7% (decreasing), nearly zero, and 23.4% (increasing) of the total study area, respectively. The downward precipitation trend in March covers 49.3% of Iran, out of which 27.7% of area is affected by 5% significant level. The dominant increasing trend occurs in November and covers 66.8% of Iran (Figure 8). These two observations are in light of the fact that March and November receive 16.7% and 10.5% of total annual precipitation (Table 1).

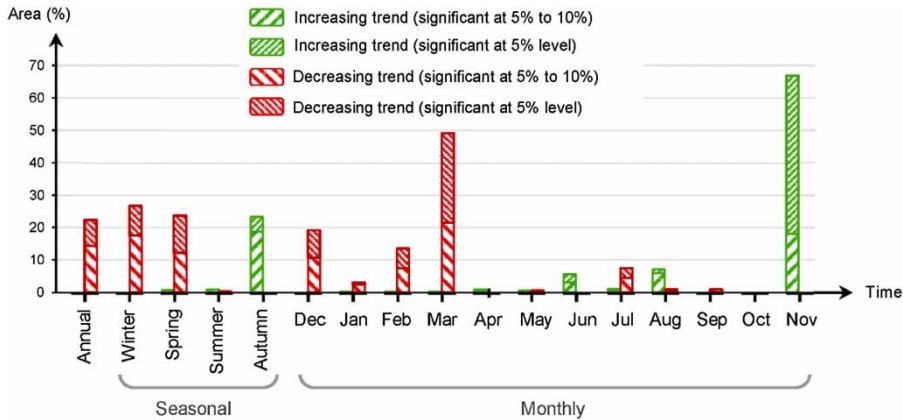
Other researchers have also analyzed the precipitation trend in Iran. Talaei (2014) studied trend in precipitation of seven rain gauges in Hamedan (a western province in Iran) and found that monthly trend analysis (from 1969 to 2009) via the MK test revealed decreasing and increasing trend in March and November, respectively (conforms with Figure 7 in Hamedan province). However, in February, the positive MK statistics found in Hamedan are in contrast with our results shown in Figure 7 (western regions). The main reason for this inconsistency is due to the difference between the length of the study periods. In other words, incorporation of more recent data in our study resulted in a downward trend in February.



**Figure 7** | Spatial distribution of monthly precipitation trend at 5% and 10% significant levels based on the MK test.

Some'e *et al.* (2012) analyzed *in-situ* precipitation data in Iran. They applied the MK test on 28 synoptic stations in the period of 1967–2006. Three northwest stations (Khoy, Urmia, and Tabriz) were found to have significant negative

trend at annual scale. Seasonal trend analysis revealed downward spring trend in three of the studied stations (in central, eastern, and southeast regions) and upward summer trend in the northeast. No significant trend was found in autumn

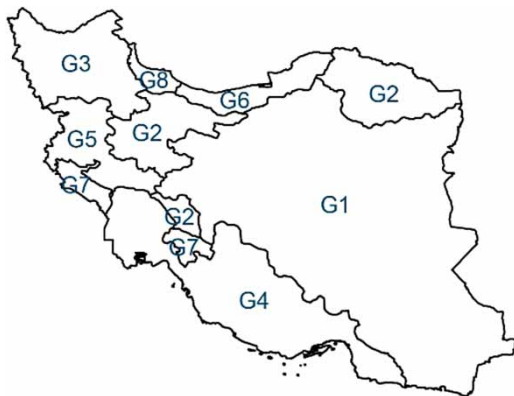


**Figure 8** | Percentage of area with increasing/decreasing precipitation trend at 5% and 10% significant levels at various time scales.

while negative winter trend was detected in the northwest and along the coasts of the Caspian Sea and in one station in the southeast region. Comparison of their results with our study indicates some similarity in spring (eastern) and winter (northwest) trends. The incompatibility of the trend in some areas could be due to differences in the study period as well as different spatial scales.

Tabari *et al.* (2011) investigated annual precipitation trend in 13 stations over the 1966–2005 period along the Zagros mountain chain, stretching in western Iran towards the southeast. Results showed a significant decreasing trend (at 95% level) in one southern and two western stations. This is in general agreement with this study although the study periods were not quite the same and that study focused on point analysis.

In another study, Modarres & Sarhadi (2009) investigated annual precipitation trend in eight regions of Iran (Figure 9). They showed increasing and decreasing annual



**Figure 9** | Precipitation regions (Modarres 2006).

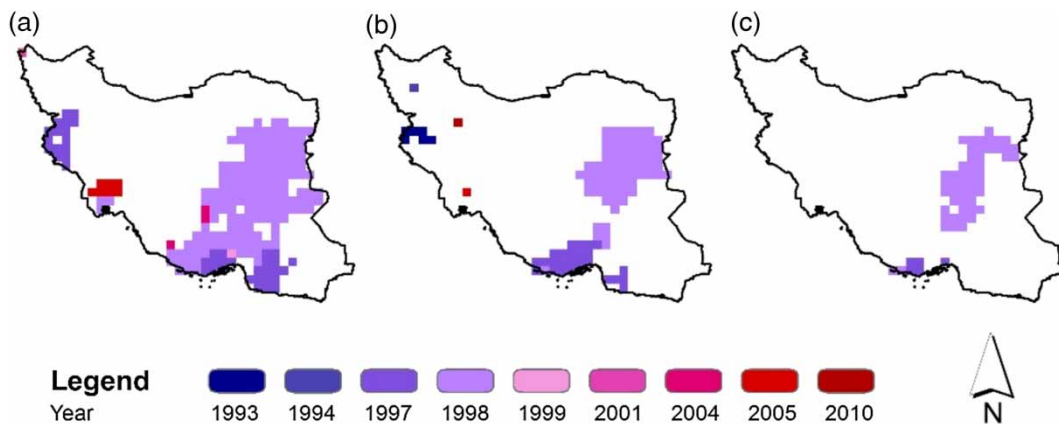
precipitation trend, respectively, in G2 and G8 regions for the 1951–2000 study period. In our study, MK statistic ( $Z$ ) and  $P$ -value were calculated for all precipitation classes (Table 2) over the 1987–2014 period. However, no significant trend was detected in annual precipitation (second column) in G2 and G8 regions. Statistics were also studied at seasonal scale that indicated significant downward trend in winter in G5 and G7 regions (Table 2). The contrast in some of our results with those of Modarres & Sarhadi (2009) stems from the difference in the duration of the study periods such that more recent data could have been affected by climate change. Our study added 14 years of recent years in comparison with Modarres & Sarhadi's (2009) study.

### Significant change point

The change point was detected by Pettitt, SNHT, and Buis-hand tests with two tail hypothesis at 5% significant level (Figure 10). The year 1998 was detected as the predominant change point in the east and south of Iran by almost all tests. Results showed higher sensitivity of Pettitt's test in comparison with SNHT and Buis-hand test. Pettitt's test revealed a broad range of change years (1997–2005) although the year 1998 was the predominant change point in areas where detection was positive. Interestingly, a similar spatial pattern in detection of decreasing trend in annual precipitation (Figure 6(a)) and the change point tests (Figure 10) could be observed.

**Table 2** | MK statistic (Z) and P-values over precipitation regions of Iran

Precipitation region	Annual		Winter		Spring		Summer		Autumn	
	Z	P-value	Z	P-value	Z	P-value	Z	P-value	Z	P-value
G1	-1.05	0.35	-1.04	0.36	-1.28	0.27	-0.32	0.61	1.28	0.26
G2	-0.71	0.50	-1.57	0.15	-0.86	0.42	0.46	0.58	1.29	0.23
G3	-0.18	0.55	-1.24	0.30	0.11	0.59	0.28	0.64	0.40	0.69
G4	-1.41	0.20	-1.55	0.13	-1.16	0.30	0.46	0.52	1.43	0.17
G5	-1.40	0.21	-2.14	0.06*	-1.49	0.16	0.41	0.62	0.90	0.38
G6	-0.34	0.75	-0.08	0.70	-1.47	0.17	-0.39	0.68	0.08	0.70
G7	-1.44	0.21	-1.90	0.09*	-1.55	0.14	1.04	0.35	1.34	0.20
G8	-0.30	0.74	-1.08	0.33	-0.14	0.66	-0.31	0.77	0.14	0.73

\*Significant at  $\alpha = 10\%$ .**Figure 10** | Significant change point in precipitation detected by: (a) Pettitt's test, (b) SNHT, and (c) Buishand test.

## CONCLUSIONS

The traditional TS structure focuses only on the time dimension. On the other hand, 'GD', which is increasingly heard of after launching satellite sensors in the RS era, deals with regular spatial data (mesh) in a snapshot. In the current paper, the concepts of CTS, SMD, TMD and, mCTS are introduced. We believe that CTS can bridge the traditional geo-datasets (especially ground-based data; e.g., binary, txt, ascii, csv, excel) to the real-world dimensions (time and space) in order to facilitate data processes (in scripting environments) for geoscientists.

In this study, we applied trend and change point detection methods on the CTS, which is a well-suited data structure to perform any time series analysis and probability-based analysis (e.g., copula) or any other methods

applicable to *in-situ* TS (e.g., wavelet). Furthermore, the CTS is perfectly adaptable for spatial analysis (e.g., entropy and homogeneity) or different machine learning approaches (e.g., principal component analysis, clustering, bagging, and boosting). We chose the trend and change point methods because they are well-practiced and provide pre-requisite information for climate change studies.

Numerous studies have focused on trend and change point occurrence within precipitation gauge-based data series. However, due to the lack of explicit spatial dimension in such studies, spatial inferences of the results are not straightforward. Accordingly, this study focused on the spatial distribution of precipitation trend and change point in Iran in the context of CTS database. The precipitation data of some 102 synoptic stations were employed to generate 25 CTS-based maps at annual, seasonal, and monthly

scales via Python programming. Moreover, temporally averaged precipitation maps at different time scales were constructed across Iran as the basis for further analysis.

Trend analyses were applied on the CTS at various time scales. The MK test revealed significant trend at 5% and 10% levels. Downward trend was identified in annual, winter and spring seasonal maps while upward trend in autumn precipitation was detected. Furthermore, monthly precipitation maps showed negative and positive trends, respectively, in March and November precipitation. Comparison of the cellular trend results with other similar studies at point scale were mostly in agreement except when the duration of the data were different.

Change point analysis conducted by Pettitt's test, SNHT, and the Buishand test revealed that the year 1998 was a significant change point in precipitation. This entirely coincides with the annual downward precipitation trend in the east and south of Iran.

All in all, the proposed integrated CTS, SMD, TMD, and mCTS concept is well-suited to load in a matrix-based programming language (e.g., Python, R, MATLAB). Thus, there is good potential to define and use different numerical and statistical functions (in this study: MK, Pettitt, SNHT, and Buishand) in the scripting environments. CTS enhances the functionality of these languages due to spatio-temporal analysis in a single environment. This concept could be put in use to link time series with geographic data and enable managers to support spatial decision-making.

### Suggestions and future directions

Different hydro-climatologic variables (e.g., temperature, solar radiation, humidity, and wind speed) could be produced in the CTS format. Also, different interpolation methods (e.g., kriging, co-kriging, and spline) could be studied and implemented. CTS and meta data are flexible joint structures that facilitate scripting for spatio-temporal analysis. Different cell size (i.e., resampling) could be generated by spatial quarry (via SMD). The 3D analysis could be applied to CTS (time and space simultaneously) to investigate the dynamic of climatic variables. Data providers could serve more online options including CTS in publishing *in-situ*, RS, and reanalyzed data.

### REFERENCES

- Abidi, S., Hajji, O. & Habaieb, H. 2017 Study of rainfall variations in Tessa subwatershed of Medjerda River in Tunisia. In: *Water Resources in Arid Areas: The Way Forward* (O. Abdalla, A. Kacimov, M. Chen, A. Al-Maktoumi, T. Al-Hosni & I. Clark, eds). Springer, Cham, Switzerland, pp. 59–74.
- Abolverdi, J., Ferdosifar, G., Khalili, D. & Kamgar-Haghighi, A. A. 2016 *Spatial and temporal changes of precipitation concentration in Fars province, southwestern Iran. Meteorology and Atmospheric Physics* **128** (2), 181–196. doi:10.1007/s00703-015-0414-0.
- Abuzied, S. M. & Mansour, B. M. H. 2018 *Geospatial hazard modeling for the delineation of flash flood-prone zones in Wadi Dahab basin, Egypt. Journal of Hydroinformatics* **21**, 180–206.
- Alexandersson, H. 1986 *A homogeneity test applied to precipitation data. Journal of Climatology* **6** (6), 661–675. doi:10.1002/joc.3370060607.
- Ali Ghorbani, M., Kazempour, R., Chau, K.-W., Shamshirband, S. & Taherei Ghazvinei, P. 2018 *Forecasting pan evaporation with an integrated artificial neural network quantum-behaved particle swarm optimization model: a case study in Talesh, Northern Iran. Engineering Applications of Computational Fluid Mechanics* **12** (1), 724–737. doi:10.1080/19942060.2018.1517052.
- Aminyavari, S., Saghafian, B. & Delavar, M. 2018 *Evaluation of TIGGE ensemble forecasts of precipitation in distinct climate regions in Iran. Advances in Atmospheric Sciences* **35** (4), 457–468.
- Buishand, T. A. 1982 *Some methods for testing the homogeneity of rainfall records. Journal of Hydrology* **58** (1–2), 11–27. doi:10.1016/0022-1694(82)90066-X.
- Chau, K. 2017 *Use of meta-heuristic techniques in rainfall-runoff modelling. Water* **9** (3), 186. doi:10.3390/w9030186.
- Chen, Y. & Han, D. 2016 *Big data and hydroinformatics. Journal of Hydroinformatics* **18**, 599–614.
- Chuntian, C. & Chau, K. W. 2002 *Three-person multi-objective conflict decision in reservoir flood control. European Journal of Operational Research* **142** (3), 625–631. doi:10.1016/S0377-2217(01)00319-8.
- de Luis, M., Čufar, K., Saz, M. A., Longares, L. A., Ceglar, A. & Kajfež-Bogataj, L. 2014 *Trends in seasonal precipitation and temperature in Slovenia during 1951–2007. Regional Environmental Change* **14** (5), 1801–1810. doi:10.1007/s10113-012-0365-7.
- Delavau, C., Chun, K. P., Stednyk, T., Birks, S. J. & Welker, J. M. 2015 *North American precipitation isotope ( $\delta^{18}\text{O}$ ) zones revealed in time series modeling across Canada and northern United States. Water Resources Research* **51** (2), 1284–1299. doi:10.1002/2014WR015687.
- Dinpashoh, Y. 2006 *Study of reference crop evapotranspiration in I.R. of Iran. Agricultural Water Management* **84** (1–2), 123–129. doi:10.1016/j.agwat.2006.02.011.
- Dinpashoh, Y., Fakheri-Fard, A., Moghaddam, M., Jahanbakhsh, S. & Mirnia, M. 2004 *Selection of variables for the purpose of*

- regionalization of Iran's precipitation climate using multivariate methods. *Journal of Hydrology* **297** (1–4), 109–123. doi:10.1016/j.jhydrol.2004.04.009.
- Duhan, D. & Pandey, A. 2013 Statistical analysis of long term spatial and temporal trends of precipitation during 1901–2002 at Madhya Pradesh, India. *Atmospheric Research* **122**, 136–149. doi:10.1016/j.atmosres.2012.10.010.
- Fallah, B., Sodoudi, S., Russo, E., Kirchner, I. & Cubasch, U. 2017 Towards modeling the regional rainfall changes over Iran due to the climate forcing of the past 6000 years. *Quaternary International* **429**, 119–128. doi:10.1016/j.quaint.2015.09.061.
- Fathian, F., Morid, S. & Kahya, E. 2014 Identification of trends in hydrological and climatic variables in Urmia Lake basin, Iran. *Theoretical and Applied Climatology* **119** (3–4), 443–464. doi:10.1007/s00704-014-1120-4.
- Fazel, N., Berndtsson, R., Uvo, C. B., Madani, K. & Kløve, B. 2017 Regionalization of precipitation characteristics in Iran's Lake Urmia basin. *Theoretical and Applied Climatology*. doi:10.1007/s00704-017-2090-0.
- Ghasemi, A. R. 2015 Changes and trends in maximum, minimum and mean temperature series in Iran. *Atmospheric Science Letters* **16** (3), 366–372. doi:10.1002/asl2.569.
- Goyal, M. K. 2014 Statistical analysis of long term trends of rainfall during 1901–2002 at Assam, India. *Water Resources Management* **28** (6), 1501–1515. doi:10.1007/s11269-014-0529-y.
- Gundogdu, I. B. 2017 Usage of multivariate geostatistics in interpolation processes for meteorological precipitation maps. *Theoretical and Applied Climatology* **127** (1–2), 81–86. doi:10.1007/s00704-015-1619-3.
- Harpham, Q. & Danovaro, E. 2015 Towards standard metadata to support models and interfaces in a hydro-meteorological model chain. *Journal of Hydroinformatics* **17** (2), 260–274.
- Hirsch, R. M., Slack, J. R. & Smith, R. A. 1982 Techniques of trend analysis for monthly water quality data. *Water Resources Research* **18** (1), 107–121. doi:10.1029/WR018i001p00107.
- Hosseinzadeh Talaei, P., Kouchakzadeh, M. & Shifteh Some'e, B. 2014 Homogeneity analysis of precipitation series in Iran. *Theoretical and Applied Climatology* **118** (1–2), 297–305. doi:10.1007/s00704-013-1074-y.
- Javari, M. 2016 Spatial-temporal variability of seasonal precipitation in Iran. *The Open Atmospheric Science Journal* **10** (1), 84–102. doi:10.2174/1874282301610010084.
- Javari, M. 2017 Assessment of temperature and elevation controls on spatial variability of rainfall in Iran. *Atmosphere* **8** (3), 45. doi:10.3390/atmos8030045.
- Kendall, M. G. 1975 *Rank Correlation Methods*, 4 edn. C. Griffin, London, UK.
- Khalili, K., Tahoudi, M. N., Mirabbasi, R. & Ahmadi, F. 2016 Investigation of spatial and temporal variability of precipitation in Iran over the last half century. *Stochastic Environmental Research and Risk Assessment* **30** (4), 1205–1221.
- Khodadoust Siuki, S., Saghafian, B. & Moazami, S. 2017 Comprehensive evaluation of 3-hourly TRMM and half-hourly GPM-IMERG satellite precipitation products. *International Journal of Remote Sensing* **38** (2), 558–571.
- Kolachian, R. & Saghafian, B. 2019 Deterministic and probabilistic evaluation of raw and post processed sub-seasonal to seasonal precipitation forecasts in different precipitation regimes. *Theoretical and Applied Climatology* **137**, 1479–1493.
- Kruger, A., Krajewski, W. F., Domaszczynski, P. & Smith, J. A. 2011 Hydro-NEXRAD: metadata computation and use. *Journal of Hydroinformatics* **13** (2), 267–276. doi:10.2166/hydro.2010.057.
- Laney, D. 2009 3D data management: controlling data volume, velocity, and variety. *Information and Software Technology* **51** (4), 769–784. doi:10.1016/j.infsof.2008.09.005.
- Li, X., Zhai, G., Gao, S. & Shen, X. 2015 Decadal trends of global precipitation in the recent 30 years. *Atmospheric Science Letters* **16** (1), 22–26. doi:10.1002/asl2.514.
- Liu, H., Van Oosterom, P., Hu, C. & Wang, W. 2016 Comparing NetCDF and a multidimensional array database on managing and querying large hydrologic datasets: a case study of SciDB. *Procedia Engineering* **154**, 207–214.
- Longobardi, A. & Mautone, M. 2015 Trend analysis of annual and seasonal air. In: *Engineering Geology for Society and Territory, Vol. 3: River Basins, Reservoir Sedimentation and Water Resources* (G. Lollino, M. Arattano, M. Rinaldi, O. Giustolisi, J.-C. Marechal & G. E. Grant, eds). Springer International Publishing, Cham, Switzerland, pp. 501–504.
- Luković, J., Blagojević, D., Kilibarda, M. & Bajat, B. 2015 Spatial pattern of North Atlantic Oscillation impact on rainfall in Serbia. *Spatial Statistics* **14**, 39–52. doi:10.1016/j.spasta.2015.04.007.
- Mann, H. B. 1945 Nonparametric tests against trend. *Econometrica: Journal of the Econometric Society* **13**, 245–259.
- Minaei, M. & Irannezhad, M. 2016 Spatio-temporal trend analysis of precipitation, temperature, and river discharge in the northeast of Iran in recent decades. *Theoretical and Applied Climatology* **131**, 167–179. doi:10.1007/s00704-016-1963-y.
- Moazenzadeh, R., Mohammadi, B., Shamshirband, S. & Chau, K. 2018 Coupling a firefly algorithm with support vector regression to predict evaporation in northern Iran. *Engineering Applications of Computational Fluid Mechanics* **12** (1), 584–597. doi:10.1080/19942060.2018.1482476.
- Modarres, R. 2006 Regional precipitation climates of Iran. *Journal of Hydrology (NZ)* **45**, 15–29.
- Modarres, R. & Sarhadi, A. 2009 Rainfall trends analysis of Iran in the last half of the twentieth century. *Journal of Geophysical Research* **114** (D3), D03101. doi:10.1029/2008JD010707.
- Modarres, R. & Sarhadi, A. 2011 Statistically-based regionalization of rainfall climates of Iran. *Global and Planetary Change* **75** (1–2), 67–75. doi:10.1016/j.gloplacha.2010.10.009.
- Nabaei, S., Sharafati, A., Yaseen, Z. M. & Shahid, S. 2019 Copula based assessment of meteorological drought characteristics: regional investigation of Iran. *Agricultural and Forest Meteorology* **276–277**, 107611. doi:10.1016/j.agrformet.2019.06.010.

- National Research Council 2013 *Frontiers in Massive Data Analysis*. The National Academies Press, Washington, DC, USA. doi:10.17226/18374.
- Nodoushan, E. J. & Shakibaeinia, A. 2018 Multiphase mesh-free particle modeling of local sediment scouring with  $\mu$  (I) rheology. *Journal of Hydroinformatics* **21**, 279–294.
- Pascale, S., Lucarini, V., Feng, X., Porporato, A. & Hasson, S. 2015 Analysis of rainfall seasonality from observations and climate models. *Climate Dynamics* **44** (11–12), 3281–3301. doi:10.1007/s00382-014-2278-2.
- Pettitt, A. N. 1979 A non-parametric approach to the change-point problem. *Applied Statistics* **28**, 126–135.
- Powell, E. J. & Keim, B. D. 2015 Trends in daily temperature and precipitation extremes for the Southeastern United States: 1948–2012. *Journal of Climate* **28** (4), 1592–1612. doi:10.1175/JCLI-D-14-00410.1.
- Raziei, T., Bordi, I. & Pereira, L. S. 2008 A precipitation-based regionalization for Western Iran and regional drought variability. *Hydrology and Earth System Sciences* **12** (6), 1309–1321. doi:10.5194/hess-12-1309-2008.
- Raziei, T., Daryabari, J., Bordi, I. & Pereira, L. S. 2014 Spatial patterns and temporal trends of precipitation in Iran. *Theoretical and Applied Climatology* **115** (3–4), 531–540. doi:10.1007/s00704-013-0919-8.
- Rew, R., Hartnett, E. & Caron, J. 2006 NetCDF-4: Software implementing an enhanced data model for the geosciences. In: *22nd International Conference on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology*, Atlanta, GA, USA.
- Roushangar, K., Alizadeh, F. & Nourani, V. 2018 Improving capability of conceptual modeling of watershed rainfall–runoff using hybrid wavelet-extreme learning machine approach. *Journal of Hydroinformatics* **20** (1), 69–87.
- Santos, M. & Fragoso, M. 2013 Precipitation variability in Northern Portugal: data homogeneity assessment and trends in extreme precipitation indices. *Atmospheric Research* **131**, 34–45. doi:10.1016/j.atmosres.2013.04.008.
- Sarmadi, F. & Shokoohi, A. 2015 Regionalizing precipitation in Iran using GPCC gridded data via multivariate analysis and L-moment methods. *Theoretical and Applied Climatology* **122** (1–2), 121–128. doi:10.1007/s00704-014-1292-y.
- Sen Roy, S. & Rouault, M. 2013 Spatial patterns of seasonal scale trends in extreme hourly precipitation in South Africa. *Applied Geography* **39**, 151–157. doi:10.1016/j.apgeog.2012.11.022.
- Seo, B.-C., Keem, M., Hammond, R., Demir, I. & Krajewski, W. F. 2019 A pilot infrastructure for searching rainfall metadata and generating rainfall product using the big data of NEXRAD. *Environmental Modelling & Software* **117**, 69–75. doi:10.1016/J.ENVSOFT.2019.03.008.
- Sharifi, E., Steinacker, R. & Saghafian, B. 2018 Multi time-scale evaluation of high-resolution satellite-based precipitation products over northeast of Austria. *Atmospheric Research* **206**, 46–63.
- Shi, P., Qiao, X., Chen, X., Zhou, M., Qu, S., Ma, X. & Zhang, Z. 2014 Spatial distribution and temporal trends in daily and monthly precipitation concentration indices in the upper reaches of the Huai River, China. *Stochastic Environmental Research and Risk Assessment* **28** (2), 201–212.
- Shi, P., Wu, M., Qu, S., Jiang, P., Qiao, X., Chen, X., Zhou, M. & Zhang, Z. 2015 Spatial distribution and temporal trends in precipitation concentration indices for the southwest China. *Water Resources Management* **29** (11), 3941–3955. doi:10.1007/s11269-015-1038-3.
- Shirvani, A. 2015 Change point analysis of mean annual air temperature in Iran. *Atmospheric Research* **160**, 91–98. doi:10.1016/j.atmosres.2015.03.007.
- Simonovic, S. P. 2009 A new method for spatial and temporal analysis of risk in water resources management. *Journal of Hydroinformatics* **11** (3–4), 320–329.
- Singh, D., Gupta, R. D. & Jain, S. K. 2015 Statistical analysis of long term spatial and temporal trends of temperature parameters over Sutlej river basin, India. *Journal of Earth System Sciences* **124** (1), 17–35. doi:10.1007/s12040-014-0530-0.
- Snijder, C., Matzat, U. & Reips, U.-D. 2012 ‘Big data’: big gaps of knowledge in the field of internet science. *International Journal of Internet Science* **7** (1), 1–5. doi:10.3923/ijds.2012.1.10.
- Some'e, B. S., Ezani, A. & Tabari, H. 2012 Spatiotemporal trends and change point of precipitation in Iran. *Atmospheric Research* **113**, 1–12. doi:10.1016/j.atmosres.2012.04.016.
- Tabari, H., Some'e, B. S. & Zadeh, M. R. 2011 Testing for long-term trends in climatic variables in Iran. *Atmospheric Research* **100** (1), 132–140. doi:10.1016/j.atmosres.2011.01.005.
- Talae'e, P. H. 2014 Iranian rainfall series analysis by means of nonparametric tests. *Theoretical and Applied Climatology* **116** (3–4), 597–607. doi:10.1007/s00704-013-0981-2.
- Toros, H. 2012 Spatio-temporal variation of daily extreme temperatures over Turkey. *International Journal of Climatology* **32** (7), 1047–1055. doi:10.1002/joc.2325.
- Ward, S., Scott, B., Kabo-bah, A., Fatawu, A. N. & Mwinkom, X. F. 2018 Water resources data, models and decisions: international expert opinion on knowledge management for an uncertain but resilient future. *Journal of Hydroinformatics* **21**, 32–34.
- Wu, C. L. & Chau, K. W. 2011 Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis. *Journal of Hydrology* **399** (3–4), 394–409. doi:10.1016/J.JHYDROL.2011.01.017.
- Yao, Z., Liu, Z., Huang, H., Liu, G. & Wu, S. 2014 Statistical estimation of the impacts of glaciers and climate change on river runoff in the headwaters of the Yangtze River. *Quaternary International* **336**, 89–97. doi:10.1016/j.quaint.2013.04.026.
- Yaseen, Z. M., Sulaiman, S. O., Deo, R. C. & Chau, K.-W. 2019 An enhanced extreme learning machine model for river flow forecasting: state-of-the-art, practical applications in water resource engineering area and future research direction. *Journal of Hydrology* **569**, 387–408. doi:10.1016/J.JHYDROL.2018.11.069.