Investigating the Spatial and Temporal Variability of Precipitation using Entropy Theory

Maya Atieh, Ramesh Rudra, Bahram Gharabaghi and David Lubitz

University of Guelph, Guelph, Ontario

Abstract

This study uses entropy theory to develop a novel application of the apportionment entropy disorder index (AEDI) to capture both spatial and temporal variability in monthly precipitation for various types of hydrologic modeling. In total, 41 Environment Canada stations across Ontario with long term (1955 to 2005) records and a very low percentage of missing data were selected. It was found that the fall and summer seasons are the major contributors to annual precipitation variability. Spatial variability of annual precipitation was observed to be increasing from southern to northern Ontario. The AEDI index map of Ontario, developed in this study, has been successfully integrated into several hydrologic models.

1 Introduction

Prediction of precipitation variability and its influence on the hydrological cycle and water quality is an increasingly active area in hydrological research (Zhang et al. 2000; Zhang et al. 2001; Lemmen et al. 2008; Finney et al. 2010; Ahmed et al. 2013; Disley et al. 2015; Liu et al. 2015; Mattar et al. 2016; Atieh et al. 2017). The distribution of precipitation spatially and temporally directly affects the hydrological cycle within a watershed, altering flow regimes and complicating the accurate predictions of flow (Das et al. 2008; Boyer et al. 2010; Coulibaly 2006; Khan and Coulibaly 2010; Liu and Cui 2011; Trenouth et al. 2013; Trenouth and Gharabaghi 2015; Trenouth et al. 2015; Thompson et al. 2016). An understanding of precipitation variability is also needed for flood and drought preparedness planning. There is a need to better understand the spatial and temporal variability in precipitation to more accurately model the hydrologic system and the health of aquatic life (Perera et al. 2009; Perera et al. 2010; Perera et al. 2013; Betts et al. 2014; Betts et al. 2015; Sabouri et al. 2013; Trenouth and Gharabaghi 2015; Gazendam et al. 2016). Parameters that represent this variability may be important factors to include in modeling studies (Vasiljevic et al. 2012; Rudra et al. 2015).

Understanding precipitation variability is a necessity for water resource management as any change in pattern influences the availability of water resources. As such, studies have used different methods that assess trends in hydrological variables. These methods may be divided into bootstrap (Burn et al. 2010), regression (Cheung et al. 2008), and nonparametric methods (Tabari et al. 2015). Trend and variation analysis of hydrologic parameters including flow and precipitation has commonly been studied by applying statistical analysis methods such as linear regression (Akinremi et al. 1999; Asnaashari et al. 2015; Atieh, Mehltretter et al. 2015). However, the expected accuracy of linear regression analysis is limited because the relationships between hydrologic parameters are generally nonlinear. The nonparametric Mann–Kendall (MK) test has been the most popularly used for trend analysis (Gajbhiye et al. 2016; Palizdan et al. 2014; Sayemuzzaman and Jha 2014). The major limitation to the MK test is its sensitivity to serial correlation in time series data which is available in precipitation data (Palizdan et al. 2014; Sayemuzzaman and Jha 2014). Time series analysis has been the most popular method for hydrological trend analysis as it does not assume linearity (Sang 2013; Yu et al. 2014). The Fourier transform has been widely adopted as a method of distinguishing the high frequency modes within a time series (Fotiadi et al. 1999). Major drawbacks of using the Fourier transform are that it is not well adapted to non-stationary data and does not maintain temporal information (Drago and Boxall 2002; Coulibaly 2006; Palizdan et al. 2014). More recently, wavelet analysis has been adopted for analysing hydrologic parameter trends and variations because it overcomes these limitations of the Fourier transform (Coulibaly 2006; Adamowski et al. 2009; Nalley et al. 2012; Adamowski et al. 2013). Discrete wavelet analyses have also been coupled with the MK test to determine the dominant periodic components affecting precipitation trends (Palizdan et al. 2014). Application of the concept of entropy in water resource engineering is an evolving idea that has attracted more research in the past decade. The application of entropy concepts has facilitated advancements in a range of hydrologic problems including the assessment of model performance, parameter estimation, derivation of functional relationships, development of flow duration curves, streamflow forecasting, uncertainty estimation, and assessment of the efficiency of monitoring networks (Cui and Singh 2016; Hao and Singh 2011; Moramarco and Singh 2010; Singh 1997; Singh 2013;
Singh et al. (2014). Liu et al. (2008) studied the basin-scale spatial patterns of precipitation time series data using Shannon’s entropy theory in the Yellow River basin, China. Entropy was used to detect the spatial structure of precipitation that was found to increase with longitude. Mishra et al. (2009) developed a methodology using entropy theory to investigate the intra-annual and decadal distributions of precipitation across Texas. They were able to determine the spatial variability of precipitation amounts within a year across Texas. Precipitation variability, quantified using entropy, is also an important parameter to include in various hydrologic modeling studies. Atieh, Gharabaghi et al. (2015) and Atieh, Mehlretter et al. (2015) developed an entropy parameter that captures both the temporal and spatial variability of precipitation in models for predicting seasonal and decadal precipitation variability. They were able to determine the spatial variability of intra-annual and decadal distributions of precipitation across the Yellow River basin, China. Entropy was used to detect the spatial structure of precipitation that was found to increase with longitude. This information was then used to determine the average amount of information needed to eliminate uncertainty (Singh 2013).

This study adopts entropy theory as a means to investigate the temporal and spatial variability of precipitation in Ontario, Canada. The uncertainty associated with different time series of precipitation measurements was determined and variation in precipitation was analysed relative to other stations in the area. Precipitation analysis is a common need in the water resource sector, and is necessary for determining the effect of annual, seasonal and decadal variability on the availability of water. This information directly impacts future water resource decision-making. The key objective of this study was to determine temporal and spatial variability of precipitation in Ontario on annual and decadal scales.

2 Methodologies

In this study, temporal variability is measured as the randomness of a time series (annual and decadal) over different time intervals (monthly, seasonal, and yearly). Variability is defined by a disorder index that is the difference between maximum and observed entropy within an individual series. A higher disorder index indicates greater variability.

Precipitation time series were considered individually to determine variability within each time series (temporal variability). Variability was then compared between stations (spatial variability). The time series used in this study are mean daily precipitation, annual precipitation and decadal precipitation time series. The variability of the annual time series was analysed over monthly intervals. The variability of the decadal time series was analysed for monthly, seasonal, and annual time intervals separately. The winter season was considered to include the months of December, January and February. Further, spatial variability in annual and decadal precipitation across Ontario was addressed by analysing average annual and decadal measurements respectively at each location.

2.1 Apportionment Entropy (AE)

Apportionment entropy (AE) is a measure of the temporal variability of monthly precipitation over a year.

\[
AE = -\sum_{i=1}^{12} \left( \frac{a_i}{A} \right) \log_2 \left( \frac{a_i}{A} \right)
\]

\[
A = \sum_{i=1}^{12} a_i
\]

The aggregate precipitation in a year \((A)\) is the sum of the individual aggregate precipitation during each month \((a)\). \(AE\) is highest \((AE = \log_2 12 = 3.585)\) when annual precipitation is equally apportioned between each month (probability 1/12 for each month), while it is lowest \((AE = 0)\) when all precipitation occurs in a single month.
2.2 Decadal Apportionment Entropy (DAE)

Decadal apportionment entropy (DAE) measures the randomness of precipitation over a time interval of a decade:

\[
DAE = -\sum_{i=1}^{10} \left( \frac{d_i}{D} \right) \log_2 \left( \frac{d_i}{D} \right) \\
D = \sum_{i=1}^{10} d_i
\]  

The aggregate decadal precipitation (D) is the sum of the individual aggregate annual precipitations for each year (d). DAE quantifies the variability of the decadal annual precipitation over a 10 y period. DAE is highest when annual precipitation is equally apportioned to each year (DAE = log, 10). Similarly, Equations 4 and 5 may be applied to any time series (monthly, seasonal). For example, Equation 4 could be applied to the aggregate precipitation values for the month of January for a decade, which would allow the determination of the randomness that occurred for that month over the decade.

2.3 Disorder Index (DI)

A disorder index is the entropy calculated, based on the observed precipitation, divided by the maximum possible entropy (even distribution of precipitation throughout the time interval). The disorder index may be calculated using an annual or a decadal dataset, and is thus labelled the annual apportionment entropy disorder index (AEDI) and decadal apportionment entropy disorder index (DEDI) respectively. The disorder index is a measure of the variability within a time series. A mean disorder index is used to compare temporal variability (i.e. from one year to another) and spatial variability (i.e. from one station to another). A long term mean is used as a threshold value to determine high versus low variability.

3 Dataset

This project focused on the province of Ontario, located in east central Canada with a total area of 1,000,000 km². Precipitation is lowest in northwest Ontario and increases in a southeast direction; annual mean precipitation in northern Ontario ranges between 400 mm and 600 mm whereas in the southern parts of Ontario it is between 800 mm and 1,200 mm (Atlas of Canada 2014).

Using the Environment Canada Adjusted Precipitation Dataset stations located in Ontario with total monthly precipitation data available for the 50 y period 1955 to 2005 were selected. This period is similar to that adopted by Nalley et al. (2012) for studying the variability of precipitation in Ontario and Quebec. Kumar et al. (2009) also indicated this period was sufficient for analyzing long term precipitation trends. The 41 stations with the lowest percentages of missing data were selected. Figure 1 shows the location of the selected stations within Ontario. Stations are numbered from the most southern station (station 1) to the most northern one (station 41). Of the 41 stations, 30 had <4% missing data for the 1955 to 2005 period, whereas the highest percentage of missing data among the 41 stations was 22%. This was included in order to allow a wide distribution of stations in northern Ontario. Mishra and Singh (2010) indicated that when missing precipitation data is less than 3%, the data can be considered of good quality. However, they also included in their study a few stations with missing values >3% (Mishra and Singh 2010).

4 Results and Discussion

4.1 Statistical Measures

Using ArcGIS 10.1 and Environment Canada Climate Normals (1971 to 2000), a total of 245 monitoring stations with 30 y of data were selected in Ontario and surrounding provinces to generate a mean annual precipitation (MAP) map using the inverse distance weighted method (Atieh, Gharabaghi et al. 2015). Figure 2 shows that the MAP varies between 500 mm and 1,200 mm across Ontario with the lowest precipitation (500 mm to 700 mm) in northern Ontario. Standard deviation (SD) of the mean daily precipitation was calculated for all stations and was represented as a percentage of MAP. It ranged between a minimum of 22% (northern Ontario) to a maximum of 42% (southern Ontario).
4.2 Variability of Precipitation Distribution Within a Year

Using apportionment entropy, the distribution of annual variability of precipitation in a year was investigated during a 50 y period (1955 to 2005). The apportionment entropy disorder index (AEDI) was used as an indicator of variability of precipitation distribution for different months and different seasons within a year. The higher the AEDI value, the higher the variation within the annual time series. The AEDI was first calculated for each station individually to study temporal variation, after which the stations were compared to determine the spatial distribution of precipitation variability. The degree of variability for specific years and stations can be identified based on a threshold for temporal and spatial variation respectively.

Temporal Variability of Annual Time Series

Apportionment entropy (AE) was calculated at the 41 sites for each month over a year using Equation 2. The AEDI value (Equation 4) was then calculated using the difference between maximum possible apportionment entropy and observed apportionment entropy. The mean AEDI of the stations over 50 y was computed to examine average temporal variation in the region. The long term average of mean AEDI was also computed and used as a threshold indicator to identify the years with high precipitation variability versus those with low variability over a regional basis. Results presented in Figure 3 indicate that within the variability range 0.18 to 0.21, there appear to be a few high and low peaks. Years 1993 and 2000 had the highest annual variability, with a few smaller peaks in 1968, 1977 and 1987. The lowest variation was observed in years 1972, 1975, 1990, 1996, and 2003.

Spatial Variability of Annual Time Series

The means of the computed AEDI for all years across the selected 41 stations for annual and seasonal time series are presented in Figures 4 and 5 respectively to show the spatial variability of annual precipitation time series in Ontario. The long term mean is used as the threshold to determine stations with high variability versus those with low variability.
Figures 4 and 5 both show that stations 1 to 28 were associated with lower variability (both monthly and seasonal) whereas the remaining stations had higher variability. The highest variability was observed at the westernmost stations in Ontario. The trend of spatial variability is similar for both time scales over an annual time series with a few distinctions. An interesting observation is that variability of precipitation within a year increases from southern Ontario to northern Ontario (Figure 6).

Figure 6  Distribution of apportionment entropy disorder index across Ontario.

4.3 Variability of Precipitation Within a Decade

The decadal apportionment entropy disorder index (DEDI) was calculated for each station to investigate the spatial and temporal variability. DEDI was calculated for five decades (1955 to 2005). Results are presented in Figure 7. For the monthly scale, there is a large variation between decades. The highest peak was observed in the decade 1956 to 1965, for the month of October. A high variability peak is observed in February for all decades. These peaks in variability of precipitation may be due to greater ranges of snowfall in February of different years and similar patterns of wide ranges of rainfall in October in Ontario.

![Fig7MeanDEDI5Decades.png](attachment:Fig7MeanDEDI5Decades.png)

Figure 7  Mean DEDI for all stations over 5 decades.

4.4 Applications of AEDI in Hydrologic Modeling

The methodology used in this paper is based on entropy, which is a different perspective on analyzing hydrological trends than has been seen in other papers that used statistical and signal analysis. Using AEDI the variability of the annual and seasonal time series was analyzed. The years with higher variability are determined and any trends throughout the analysed 50 y may be identified. Through collapsing the time scale and analyzing AEDI spatially, across the 41 stations, the spatial distribution of precipitation variability was presented. Similarly, using DEDI, the decadal trends in precipitation was analyzed.

Incorporating precipitation variability in various hydrological models allows a more accurate presentation of the hydrologic system (Hwang et al. 2012; Masinde 2014; Rudra et al. 2015). Different strategies are used to represent precipitation variability within hydrologic models, such as comparing long term precipitation records and dividing precipitation time series into different periods (Masinde 2014; Hiltner et al. 2016; Li et al. 2016). The entropy based methodology presented in this paper generates a simple, easy to use parameter (AEDI) that allows incorporation of temporal and spatial variability of precipitation into any hydrologic model. An AEDI map may be developed for any region in the world using the methodology presented in this paper for generating a spatial AEDI map of Ontario. Atieh, Gharabaghi et al. (2015) and Atieh, Mehltretter et al. (2015) were the first to introduce AEDI as an input parameter to an artificial neural network (ANN) model to more accurately predict flow duration curve (FDC) and sediment rating curve (SRC) parameters at ungauged sites. The results of Atieh, Gharabaghi et al. (2015) indicated that the parameters defining the log normal distribution for the FDCs (location and scale) were highly sensitive to AEDI. Incorporating AEDI in the location and scale ANN models improved prediction performance by 7% and 21%, respectively. Atieh, Mehltretter et al. (2015) reported that removing climatic parameters from the SRC prediction model decreased the correlation coefficient by 40% and the Nash–Sutcliffe coefficient by 42%. This paper recommends using AEDI in future hydrological modeling research such
5 Conclusions

This study presents an alternative method for quantifying spatial and temporal variability of precipitation for hydrologic modeling. The apportionment entropy disorder index (AEDI) was calculated using 50 years (1955 to 2005) historic precipitation records for 41 Environment Canada Stations across Ontario. The AEDI values of the stations were compared to determine the spatial distribution of precipitation variability. Precipitation variability was observed to increase from southern Ontario to northern Ontario.

The decadal entropy disorder index (DEDI) was calculated for 5 decades (1955 to 2005). The highest monthly variability peaks were observed in February. Spatial variability increased from southern Ontario to northern Ontario.

This paper presents a novel approach for a better understanding of the spatial and temporal variability of precipitation which is useful for a variety of hydrologic modeling projects, including flood and drought preparedness planning.

Acknowledgments

The authors would like to acknowledge the generous support of the Natural Sciences and Engineering Research Council of Canada (NSERC) for their support of this research. The authors also acknowledge the support received from Dr. Coulibaly and Dr. Mishra, including constructive comments and the initial Matlab code.

References


