6-22-2011

Drought Regionalization of Brazos River Basin Using an Entropy Approach

Deepthi Rajsekhar
Texas A & M University - College Station, deepthir86@tamu.edu

Ashok Mishra
Texas A & M University - College Station, amishra@tamu.edu

Vijay P. Singh
Texas A & M University - College Station, vsingh@tamu.edu

Drought Regionalization of Brazos River basin using an entropy approach

Deepthi Rajsekhar
Graduate student
236 Scoates Hall
Texas A&M University
(732)-306-0739
deepthir86@tamu.edu

Ashok Mishra
Post doctoral Associate
321 E Scoates Hall
Texas A&M University
(979)-661-6430
amishra@tamu.edu

Vijay Singh
Professor
321 Scoates Hall
Texas A&M University
(979)-845-7028
vsingh@tamu.edu

ABSTRACT

Assessment and understanding of past climate is an important step for drought mitigation and water resources planning. In this study, streamflow simulation obtained from the variable infiltration capacity (VIC) model was used for drought characterization, and subsequently regionalization was done based on the annual severity level, for the Brazos basin in Texas over a time span of 1949-2000. It is important to study drought characteristics within a regional context. Hence, identification of homogenous drought regions is a prerequisite, so that the drought characteristics can be studied within each of these regions. In this study, the concept of entropy was used for formation of homogenous regions based on drought severity. A standardized version of mutual information known as directional information transfer was used for station grouping. Results obtained were compared with the conventional k-means clustering method. The regions obtained were similar in both cases.

Keywords

Drought Regionalization, Entropy, Directional information transfer, K-means clustering.

1. INTRODUCTION

In a very general sense, drought is an extended period of time which experiences a deficiency in precipitation. It is a normal, recurring feature of climate which is inevitable. It is a very gradual phenomenon, and often it is difficult to identify the beginning or end of a drought (Wilhite and Glantz, 1985). A drought can extend for just a few months, or it may persist for several years. There is no universally accepted definition for droughts and they are classified into meteorological, hydrological, agricultural and socioeconomic droughts (Mishra and Singh, 2010).

Because of the impact of droughts on society, adequate monitoring and planning is required for effective mitigation of the same. Similar water management schemes and drought planning can be adopted for homogenous regions. Also, identification of homogenous drought regions is needed for regional frequency analysis of droughts.

A few works focussing on the regional analysis of droughts include Nathan and McMahon (1990), Tallaksen and Hisdal (1997), and Byzedi and Saghafian (2009).

The objectives of the paper are: (1) Application of VIC model for streamflow drought analysis (2) regionalisation of Brazos basin based on the annual drought severity levels, and (3) identification of critical regions within the basin using entropy and comparison with the k means clustering method.

2. STUDY AREA

The area considered for this study is the Brazos River basin in Texas. The coordinates of the source and mouth are $33^0$16'07''N $100^0$0'37''W and $28^0$52'33''N $95^0$22'42''W respectively. The basin has an area of 116,000km^2 and an average discharge of 300.2m^3/s (www.wikipedia.org). Figure 1 shows station locations used for validation of VIC results. The basin crosses over three climatic regions of Texas: continental, sub-tropic semi-humid and sub-tropic humid. Figure 1 also shows the climatic regions within Texas. Table 1 gives other details of the validation stations.

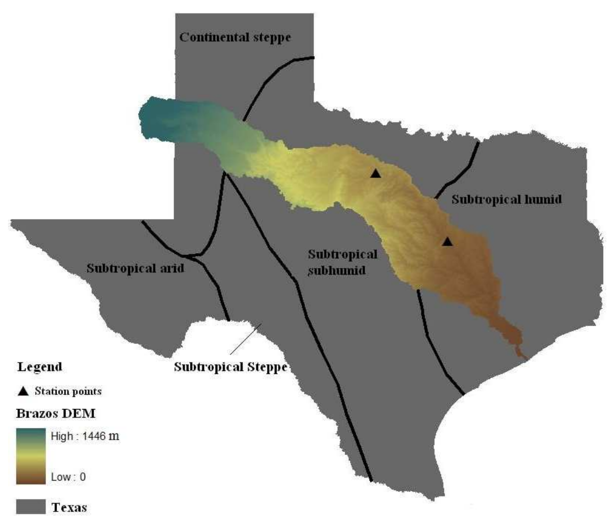


Figure 1. Location of validation stations and climate regions within Texas

Table 1. Location of Validation stations

Station name	Latitude	Longitude	Drainage area (sq miles)
Brazos rv nr Waco	31.535	-97.073	19993
Brazos rv nr Southbend	33.024	-98.643	22673

3. METHODOLOGY

3.1 Model description

The VIC-3L is a large scale land surface model and is used for simulating land-atmosphere fluxes by solving water and energy balance at a daily or sub-daily temporal scale. It was first developed at University of Washington by Liang et al., 1994. The land surface is essentially divided into grids of specified resolution. Each of these cells is simulated independent of each other. Land surface is divided into different vegetation covers in such a way that multiple vegetation classes can exist within a cell. The vegetation parameters considered by the model include root depth, root fraction, LAI, stomatal resistance, albedo, etc. The soil is partitioned into three layers vertically, and the main soil parameters include hydraulic conductivity, thickness of each soil layer, soil moisture diffusion parameters, initial soil moisture, bulk density and particle density. The model has been widely used, particularly for streamflow and soil moisture simulations. Abdulla et al. (1996), Nijssen et al. (1997), Lohmann et al. (1998), and Nijssen et al. (2001), primarily used VIC for streamflow simulation. Sheffield et al. (2004), Andreadis and Lettenmaeir (2006), Sheffield and Wood (2008), and Shukla and Wood (2008) demonstrated the use of VIC simulated soil moisture and runoff in the context of drought .

In addition to VIC, a river routing model was used to route streamflow to the desired station location. The routing model was developed by Lohmann et al. (1996, 1998). In this routing scheme, the surface runoff simulated by VIC in each grid cell is transported to the outlet of the grid cell using a unit hydrograph approach. Then, runoff from each grid cell is routed through the channel using a linearized Saint-Venant equation.

3.2 Drought classification using standardized streamflow index

In this study, the drought characteristic focussed on was severity. Theory of runs was used to derive severity from VIC simulated streamflow. A run is defined as a portion of time series of drought variable X_t in which all values are either above or below a threshold level X_0. Accordingly it can be called a positive or a negative run. The threshold level may be constant or it may vary with time. For this study, the drought variable X_t chosen was standardised streamflow index (SSFI). The concept of SSFI is based on the standardised precipitation index (SPI) by Mckee et al. (1993) and has been applied by Modarres (2007). It is statistically similar to SPI. SSFI for a given period can be defined as:

$$SSFI = \frac{F_i - \overline{F}}{\sigma}$$

Where F_i is the flow rate in time interval i, $\overline{F}$ is the mean of the series and σ is the standard deviation of the series. The drought classification based on SSFI is similar to that based on SPI. Table 2 gives the details of SSFI classification. SSFI values less than -1 were considered for calculating the drought severity. The cumulative deficit gives the severity value. Figure 2 explains drought characterisation using the theory of runs. All the shaded portions indicate drought events. The annual severity value for all the VIC girds within the basin over the period 1949-2000 was calculated using this method.

Table 2. SSFI classification

SSFI value	Classification
2.0 or more	Extremely wet
1.5 to 1.99	Very wet
1.0 to 1.49	Moderately wet
-0.99 to 0.99	Near normal
-1.0 to -1.49	Moderately dry
-1.5 to -1.99	Severely dry
-2.0 or less	Extremely dry

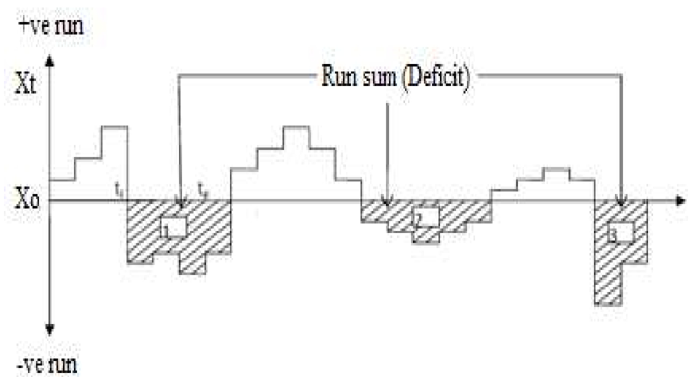


Figure 2. Theory of runs

3.3 Regionalization based on severity

In this study, an entropy based index known as directional information transfer (DIT) was used for the grouping of grids into homogenous regions. The concept of entropy was first used in the context of communication theory. In communication theory, entropy measures the uncertainty of a random event, or rather the information contained in it through the observations of it (Yang and Burn, 1994). Thus it can be inferred that the observations from an uncertain event will contain more information than the observations from a certain event. Since there will be some kind of information transfer between different sites, the observations made at one site infer information about other sites too to some extent. This information transfer among the stations is termed as 'mutual information'. Thus entropy can be used to measure the information content of observations and mutual information can be used to measure the information transfer. Thus entropy and mutual information provide a threefold measure of information at a station, information transfer and loss between stations, and description of relationships among stations according to the information transfer between them (Yang and Burn,1994). This makes it unique from other conventional similarity measures.

In addition to the use of DIT, the conventional k-means clustering has also been used to identify homogenous regions, and the results obtained from both these methods were compared.

3.3.1 Entropy concepts

Starting with the basic concept of entropy, Shannon entropy for a random variable X is defined as (Lathi,1968):

$$H(X) = \sum_{i=1}^{m} P_i \log_2 P_i$$

where P_i's are the probabilities, and m denotes the total number of class intervals. H(X) is the marginal entropy of X, which means the measure of information contained in X. If two random variables (X,Y) are considered, the mutual information or the measure of information transfer between them can be computed as (Lathi,1968):

$$T(X,Y) = H(X) - H(X/Y)$$

where H(X/Y) represents the information lost during transmission. It can be estimated as:

$$H(X/Y) = \sum_{i,j} p(X_i, Y_j) \log_2 \frac{p(X_i, Y_j)}{p(Y_j)}$$

There are studies in which mutual information has been used as a similarity measure for clustering purposes (Kraskov and Grassberger, 2009) and as a distance measure (Cover and Thomas, 1991). It was found that mutual information as a similarity measure works much better than the Pearson correlation or Euclidean distance (Priness et al., 2007).

3.3.2 Directional information transfer

This is a standardized version of mutual information. Directional information transfer is the fraction of the information transferred from one site to another. The concept of DIT was introduced by Coombs et al. (1970) in the field of mathematical psychology as a coefficient of constraint (Fass, 2006). It is a normalized version of mutual information between two gauges to obtain the fraction of information transferred from one site to another as a value between 0 and 1. DIT is a much better index than mutual information because the upper bound of mutual information can vary from site to site, depending on the marginal entropy value at the respective station which makes the mutual information, a not so good index of dependence. DIT can thus be expressed as:

$$DIT_{xy} = \frac{T(X,Y)}{H(X)};\ DIT_{yx} = \frac{T(X,Y)}{H(Y)}$$

where DIT_{xy} describes the fractional information inferred by station X about Y, and DIT_{yx} is the fractional information inferred by station Y about X, T(X,Y) is the mutual information between X and Y, H(X) and H(Y) are the marginal entropy values for X and Y respectively. The marginal entropy values are calculated using the formula for Shannon entropy, the mutual information between X and Y can be calculated as T(X,Y) = H(X)-H(X/Y) where H(X/Y) is equivalent to the loss of information H_{lost}.

$$DIT = (H - H_{lost}) / H = 1 - (H_{lost} / H)$$

It should also be noted that while the mutual information term is symmetric, DIT is no longer symmetrical. The concept of using entropy for the purpose of regionalization in hydrology was introduced by Yang and Burn (1994). Not many applications of DIT in hydrology exist. Alfonso et al. (2010) used DIT as a criteria to determine the independency of water level monitoring stations which helped in designing an optimum network for the same.

3.3.3 Regionalization using DIT

While using DIT for regionalization, those stations for which both DIT_{xy} and DIT_{yx} are high can be considered to be strongly dependent since information can be mutually inferred between them. If neither DIT is high, then the two stations should remain in separate groups. If only one DIT is high, say DIT_{xy}, then station Y, whose information can be predicted by X, can join station X if station Y does not belong to any other group; otherwise it stays in its own group. But, by no means can X enter station Y's group (Yang and Burn, 1994). DIT can be distinguished from traditional similarity measures like correlation coefficient, since it is based on the information connection between stations.

The number of groups formed is controlled by the threshold value of DIT. A higher threshold value will lead to a larger number of groups. However, the size of each group will be small. A lower threshold value will result in the formation of a small number of groups, but the size of each group will be larger.

In this study, the series of annual drought severity values for the years 1949-2000 at each grid was considered as variable, X. The probability estimation was done using the histogram method. A bin size of 7 was chosen. Using the expression for Shannon entropy, the marginal entropies were estimated. Having obtained the marginal entropies, the next step is the estimation of joint and conditional probabilities and then mutual information. Then the DIT matrix for Brazos basin was estimated. Since the basin of interest has a total of 719 grids, the size of the DIT matrix will be 719X719. The DIT values ranged between 1 and 0.059. A threshold value for DIT was chosen for grouping. Since there is no guideline for choosing the value, the decision rather depends upon whether it produces a reasonable number of groups. In this study after trials with different threshold values, a value of 0.4 was chosen, since values higher or lower than that produced several or too few groups after regionalization. To demonstrate how the choice of threshold value affects the number of groups formed, a small sample of DIT matrix with only 8 stations is given in Table 3.

It can be seen that the maximum DIT value corresponds to station pair 1 and 2 (0.54 and 0.45) and the smallest is for station pair 1 and 5 (0.12 and 0.14) respectively.

Table 3. Sample DIT matrix for 8 stations

Station	1	2	3	4	5	6	7	8
1	1	0.54	0.20	0.13	0.12	0.20	0.21	0.47
2	0.45	1	0.52	0.17	0.15	0.32	0.25	0.32
3	0.25	0.50	1	0.28	0.15	0.49	0.19	0.29

4	0.18	0.15	0.21	1	0.42	0.25	0.48	0.22
5	0.14	0.15	0.17	0.40	1	0.21	0.29	0.15
6	0.22	0.30	0.50	0.22	0.19	1	0.19	0.23
7	0.26	0.27	0.22	0.50	0.31	0.23	1	0.20
8	0.41	0.28	0.26	0.16	0.14	0.20	0.21	1

Consider a threshold of 0.35. The groups formed based on the grouping principles explained earlier will comprise: 1,2,3,6 and 8 in group 1 and 4,5 and 7 in group 2. Figure 3 shows the grouping when the threshold is 0.35. If instead of 0.35, we choose a lower threshold, say 0.2 all 8 stations will fall under one group. This proves the statement that lower the threshold, smaller the group numbers, and larger the group size.

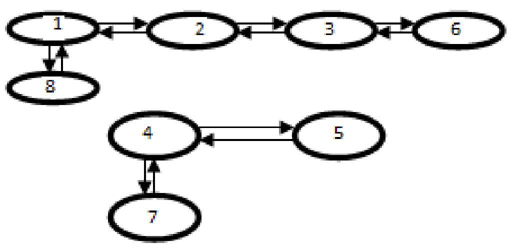


Figure 3. Grouping for threshold more than 0.35

If we choose a much higher threshold, say 0.45 then it can be seen that initially, stations 1,2,3 fall in one group and 4,7 falls in another group. For stations 5 and 8, there is no combination for which both DIT_{xy} and DIT_{yx} are higher than the threshold. Next, check whether any one value of DIT_{xy} or DIT_{yx} is higher than the threshold. It can be seen that DIT_{18} is 0.47 which is higher than the threshold. Since 8 does not belong to any group, it can be put into the group of station 1. For station 5, since none of the DIT_{xy} or DIT_{yx} values are above the threshold, it does not fall in either group 1 or 2.

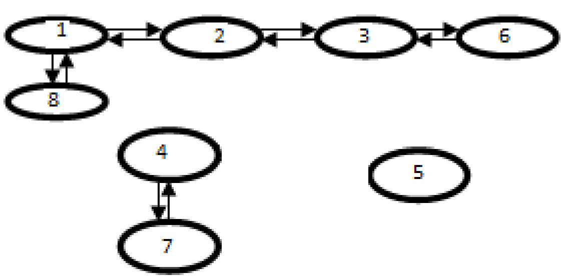


Figure 4. Grouping for threshold more than 0.45

3.3.4 Comparison with k-means clustering

To understand how well the entropy based index of DIT succeeds in grouping similar stations, the results were compared with the conventional k-means clustering method. K-means is a hard clustering algorithm in which a collection of N vectors will be classified into K groups. The aim of the algorithm is to find the center of clusters (also known as centroids) for each group. The algorithm minimizes the objective function which is essentially a dissimilarity function.

The steps in the algorithm includes: (1) Initialise the centroids k_i = 1,2,...k by randomly selecting k points from among all data points. (2) Determine the membership matrix U by equation: $u_{ij} = 1$ if $\|Xj - ki\|^2 \le \|Xj - kk\|^2$, for all $k \neq i$, $u_{ij} = 0$ otherwise

(3) Compute the dissimilarity function

$F = \sum_{i=1}^{k} [\sum_{k, x_k \in G_i} \|x_k - k_i\|^2]$. Stop if its improvement over the previous iteration is below a threshold.

(4) Compute new centroids using:

$$k_i = \frac{1}{|G_i|} \sum_{k, x_k \in G_i} x_k \text{and go to step } (2).$$

The performance of the algorithm depends on the initial position of centroids. Since we do not know the value of k apriori, cluster validity indices were employed to determine an estimate of k to be used. The Calinski-Harabasz index was used as the cluster validity index in this study. The optimum number of clusters corresponded to the highest value of the index.

In hydrology, K-means algorithm and its variants have been used, primarily as part of regionalisation of watersheds by Bhaskar and O'Connor (1989), Burn and Goel (2000), Rao and Srinivas (2005), and Isik and Singh (2008), to name a few.

4. DATA

For this study, the VIC model for streamflow simulation was run at 1/8th degree resolution and hence all the input files including forcing files, soil and vegetation parameters have this resolution. The model needs climatic forcing data at a daily temporal scale, and the forcing variables commonly used are daily precipitation, wind speed and air temperature extremes. The time period of data used was for the latter half of 20th century: 1949-2000. The gridded forcing data at 1/8th degree resolution required for driving the model was obtained from Maurer et al. (2002) who has provided a data base for 15 delineated basins over United States, Canada and Mexico. From this, a subset for Neches basin was derived for this study. Apart from forcing data, soil and land cover data is also required by VIC model. The soil parameters which were not be used for calibration were obtained from LDAS (Land Data Assimilation System). Vegetation parameters needed were also obtained from LDAS. The leaf area index (LAI) needed was obtained from Moderate Resolution Imaging Spectroradiometer (MODIS) data.

The data needed in the routing scheme includes a fraction file, flow direction file, Xmask file, flow velocity and diffusion files, and unit hydrograph file. ArcMap was used for the preparation of the files, and the DEM files needed for creating the required files were obtained from USGS hydro 1k datasets.

5. RESULTS AND DISCUSSIONS

5.1 Setting of VIC model

Since VIC model involves a lot of parameters, calibration of the same can become quite tedious. The recommended parameters, and the plausible range of values for each of them are given in Table 4. In this study, six soil parameters were considered for calibration purposes. As far as the calibration of the routing model is concerned, the suggested parameters for adjustment include velocity and diffusivity. If only monthly streamflows are required, velocity and diffusivity values of 800 m^2/s and 1.5 m/s are deemed acceptable.

Table 4. Details of calibration parameters

Soil parameter	Unit	Range of values
Infiltration shape parameter (b_{inf})	None	0-0.4
Maximum sub-surface flow rate (Ds_{max})	mm/day	0-30
Fraction of Ds_{max} when non linear flow starts (Ds)	None	0-1
Depth of second soil layer (D_2)	meter	0.1-1.5
Depth of third soil layer (D_3)	meter	0.1-1.5
Fraction of maximum soil moisture when non linear flow starts (W_s)	None	0-1

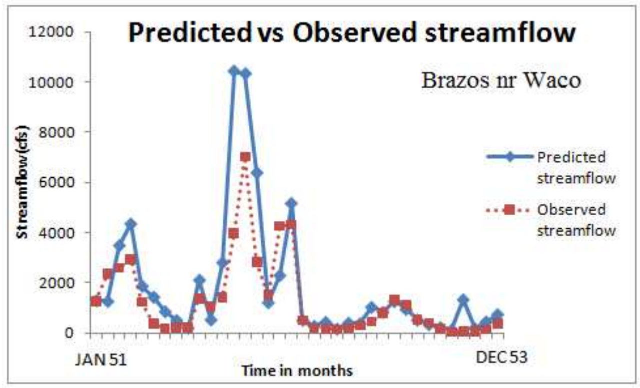
5.2 Validation of the model

The streamflow obtained after calibrating the model parameters were validated using the USGS streamflow data. For this purpose, the routing model was used to route the flow to the selected station locations. The results from the routing model were aggregated to a monthly scale (in cfs) and compared with the observed gauge data (in cfs). The three performance criteria selected were correlation coefficient, Nash-Sutcliffe efficiency and mean flow ratio. A higher value of correlation coefficient and Nash efficiency indicates good performance of the model. The closer the value is to 1, the more accurate the model is. Validation of the results obtained from the calibrated model with respect to the observed streamflow values at the respective gauges are shown in Figure 5. Table 5 gives a summary of performance measures at each of these stations. The time period of validation was from 1951-1953.

From the values obtained it can be seen that the model performance is acceptably good. It can also be seen from the table that since the mean flow values at both stations are more than 1, the model shows a tendency to overpredict the streamflow values.

Table 5. Results of validation

Station	Correlation coefficient	MF ratio	NSE
Brazos nr Waco	0.9	1.276	0.619
Brazos nr Southbend	0.87	1.584	0.514



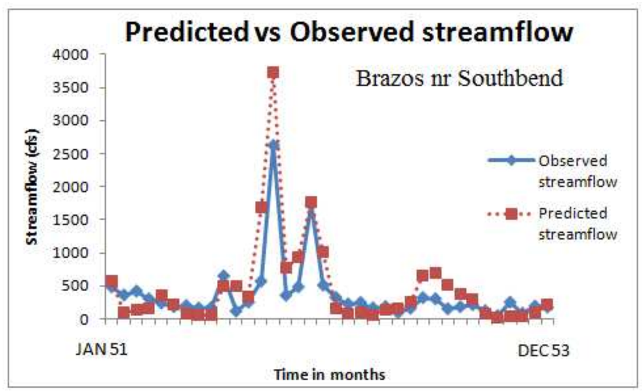


Figure 5. Validation results at the stations

5.3 Results of grouping

Overall, there were 719 grids to be grouped. While grouping, attention was also paid to two additional factors apart from the suggested grouping principles: (1) Weightage was given to the distance between station pairs. (2) Since Brazos river basin extends across three different climate regions, care was taken to not group grids belonging to different climate regions. For the selected threshold of 0.4, a total of eight regions were formed. The groups formed along Brazos basin are shown in Figure 7. Totally 8 regions were formed. Table 6 gives details of the groups formed.

Table 6. Details of the regions formed using DIT

Region	Number of grids	Annual average severity	Climatic region
1	67	5.631	Subtropical humid

2	37	7.239	Subtropical humid
3	169	10.677	Subtropical humid
4	82	9.127	Subtropical subhumid
5	73	12.388	Subtropical subhumid
6	130	16.502	Subtropical subhumid
7	48	9.178	Continental steppe
8	91	6.838	Continental steppe

It can be seen that region 3 is the largest group formed covering about 24% of the total basin area and region 2 is the smallest, covering just about 6.88% of the total basin area. The annual average severity value is the highest for region 6 and comes up to about 16.502 and is lowest for the region 1 and comes up to about 5.631. Figure 6 is a graph showing the variation of average severity and percentage area across the regions.

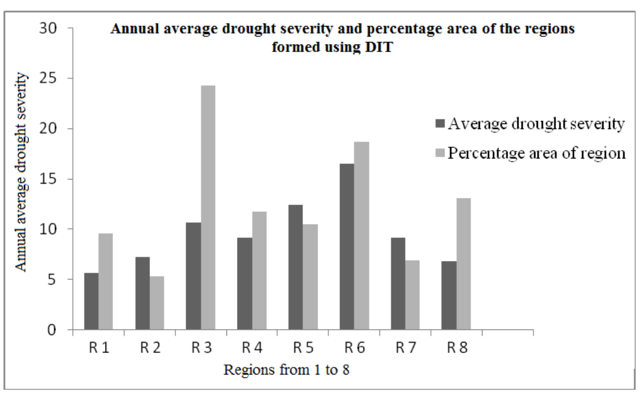


Figure 6. Average annual severity and percentage area for regions found using DIT

In the case of k-means clustering, according to the Calinski-Harabasz validity index, the number of clusters selected were 7 for which the lowest index value was obtained. Figure 8 shows the variation of Calinski-Harabasz index with number of clusters. Table 7 gives the details of the groups formed.

It can be seen that region 5 is the largest group formed and covers about 34.7% of the total basin area and region 1 is the smallest with 5.88% of the total basin area. The annual average severity is highest for region 5 and corresponds to 14.77. Region 1 has the least annual average severity of 4.519. Figure 9 shows the variation of average severity and percentage area across the regions. Figure 10 shows the regions formed while k-means clustering was used.

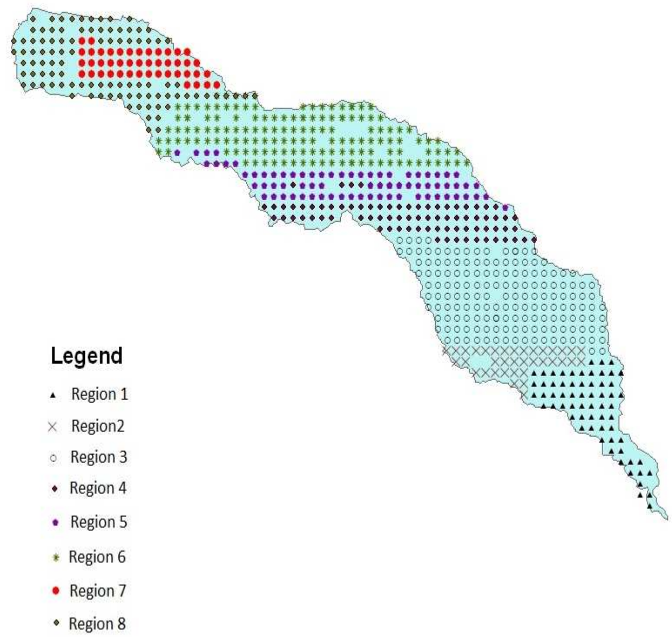


Figure 7. Homogenous regions within Brazos basin using DIT

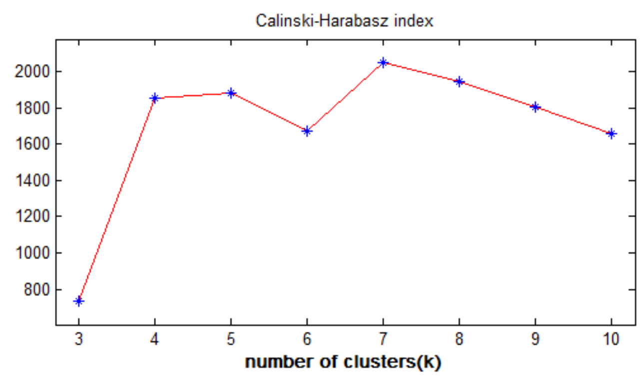


Figure 8. Variation of Calinski-Harabasz index with cluster numbers

Table 7. Details of the regions formed using K-means

Region	**Number of grids**	**Annual average severity**	**Climatic region**
1	41	4.519	Subtropical humid
2	76	8.012	Subtropical humid
3	86	9.770	Subtropical humid
4	58	9.617	Subtropical humid
5	242	14.770	Subtropical subhumid

6	43	10.296	Subtropical subhumid
7	151	9.255	Continental steppe

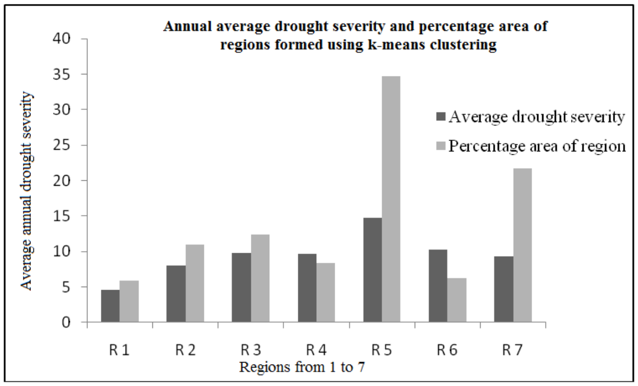


Figure 9. Average annual severity and percentage area for regions found using K-means

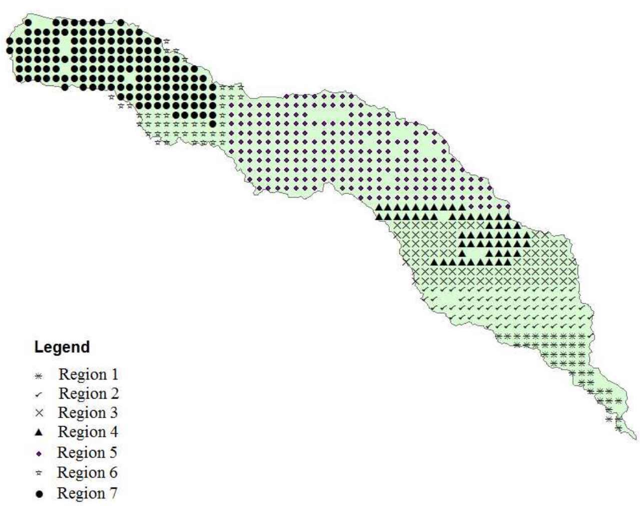


Figure 10. Homogenous regions within Brazos basin using k-means clustering

It can be seen from Figures 9 and 10 that the results obtained using entropy based index and a conventional clustering method are quite similar.

Figure 11 shows the precipitation pattern within the Brazos river basin. The average precipitation rate within each region formed using DIT and k-means clustering, is given in Table 8. It can be seen that the drought severity pattern closely follows the precipitation pattern within the basin, with the exception being the upstream portion of the basin, which showed slightly lower severity despite getting low rainfall.

Table 8. Average daily precipitation rate for the regions

Average daily precipitation rate for the region under consideration (mm/day)								
Regions	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
DIT	2.87	2.61	2.18	2.21	1.92	1.19	1.25	2.7
k-means	2.93	2.52	2.33	2.27	1.85	1.36	1.23	-

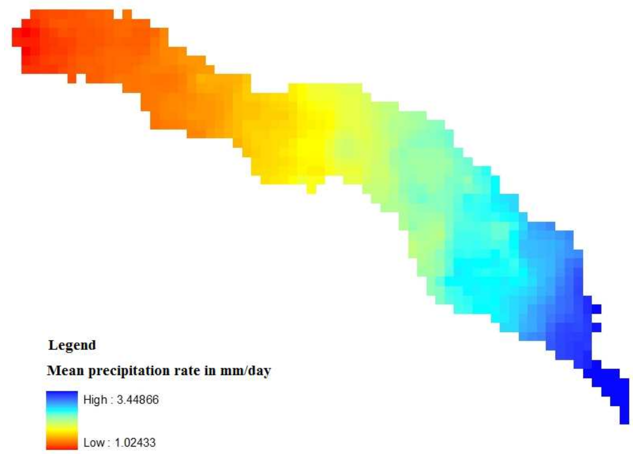


Figure 11.Average daily precipitation rate over Brazos basin

6. CONCLUSIONS

The study applies entropy for purposes of regionalization of Brazos River basin using directional information transfer. DIT groups stations based on the information connection between them, and can be considered as a better index than mutual information due to its standardized form, and is also superior to other statistical similarity measures, like correlation coefficient. Results of regionalization are also compared with the k-means clustering.

Out of the eight regions formed using DIT, the lower Brazos basin showed relatively lower severity, which could be attributed to the fact that rainfall is relatively higher over the part. The middle Brazos basin showed higher severity levels. In the case of regions formed using k-means clustering, the middle and upper regions of Brazos basin showed higher severity than the lower basin. The pattern of severity is similar for both the methods and in general, follows the precipitation pattern within the basin. Further investigation as to the changes in water demand, land use and land cover, and its subsequent effects need to be conducted to ascertain its influence on streamflow drought .

7. REFERENCES

[1] Abdulla, F.A., Lettenmaier, D.P.,Wood, E.F and Smith, J.A 1996. "Application of a macroscale hydrologic model to estimate the water balance of the Arkansas-Red river basin", *J. Geophys. Res.*, 101(D3), 7449-7459.13pp.

[2] Alfonso, L., Lobbrecht. A., and Price,R. 2010. “Optimization of water level monitoring network in polder systems using information theory”, *Water Res. Research,* 46,

[3] Andreadis, K.M., and Lettenmaier,D.P. 2006. “Trends in 20th century drought over the continental United States”, *Geophysical Research Letters* 33**,** p. L10403 10410.11029/12006GL025711.

[4] Bhaskar, N.R. and O'Connor, C.A. 1989. “Comparison of method of residuals and cluster analysis for flood regionalization”, *J. Water Res. Pl. and Mgmt.* 115(6),793–808.

[5] Burn, D.H. and Goel, N.K. 2000. “The formation of groups for regional flood frequency analysis”, *Hydro. Sci. J.,* 45(1), 97–112.

[6] Byzedi, M. and Saghafian, B. 2009. Regional analysis of streamflow drought: A case study for Southwestern Iran, *W. Acad. Sci.*, 57, 447-451.

[7] Cover, T.M. and Thomas, J.A. 1991. “Elements of Information Theory”, John Wiley and Sons, New York.

[8] Fass,D.M. 2006. Human sensitivity to mutual information.PHD dissertation, Rutgers state university of N.J., New Brunswick.

[9] Kraskov, A. and Grassberger, P. 2009. “MIC: Mutual Information based Hierarchial Clustering*” Inform. Theory Stat. Learn.* 101-123.

[10] Lathi, B.P. 1968. “An introduction to Random Signals and Information Theory”, International Textbook Company, Scanton, Pennsylvania.

[11] Liang, X., Lettenmaier,D.P.,Wood,E.F. and Burges,S.J. 1994. A Simple hydrologically Based Model of Land Surface Water and Energy Fluxes for GSMs, *J. Geophys. Res.*, 99(D7), 14,415-14,428.

[12] Lohmann, D., Nolte-Holube,R. and Raschke, E. 1996. A large-scale horizontal routing model to be coupled to land surface parametrization schemes, *Tellus*, 48(A), 708-721.

[13] Lohmann, D., Raschke, E., Nijssen, B. and Lettenmaier, D.P. 1998. Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model, *Hydrol. Sci. J.*, 43(1), 131-141.

[14] Maurer, E.P., Wood, A.W., Adam, J.C., Lettenmaier, D.P. and Nijssen, B. 2002. A Long-Term Hydrologically-Based Data Set of Land Surface Fluxes and States for the Conterminous United States, *J. Clim.,* 15(22), 3237-3251.

[15] McKee, T.B., Doesken, N.J., Kleist, J.1993. “The relationship of drought frequency and duration to time scales”, *Eight conf. App. Clim.*, Anaheim, CA.

[16] Mishra, A.K. and Singh, V.P. 2010. A review of drought concepts, *J.Hydrol*., 391, 202-216.

[17] Modarres, R. 2007. Streamflow drought time series forecasting, *Stoch. Environ. Res. Risk Assess.,* 22, 223-233.

[18] Nathan, R.T. and McMahon, T.A. 1990. Identification of homogenous regions for the purpose of regionalisation, *J. Hydrol*., 121, 217-238.

[19] Nijssen, B.N., Lettenmaier, D.P., Liang, X., Wetzel, S.W. and Wood, E.F.1997. “Streamflow simulation for continental-scale river basins”, *Water Resour. Res.*, 33(4), 711-724.

[20] Nijssen, B.N., O'Donnell, G.M., Lettenmaier, D.P. and Wood, E.F. 2001. “Predicting the discharge of global rivers”, *J. Clim.*, 14, 3307-3323.

[21] Sheffield, J., Goteti, G., Wen, F. and Wood, E.F.2004. “A simulated soil moisture based drought analysis for the United States”, *J. Geophys. Res.,* 109, p. D24108 24110.21029/22004JD005182.

[22] Sheffield, J., and Wood, E.F.2008. “Global Trends and Variability in Soil Moisture and Drought Characteristics, 1950–2000, from Observation-Driven Simulations of the Terrestrial Hydrologic Cycle”, *J. Clim.*, 21, 432–458.

[23] Shukla, S., and Wood, A.W.2008.“Use of a standardized runoff index for characterizing hydrologic drought”, *Geophys.ResLetters*35,L02405,02410.01029/02007GL032487.

[24] Tallaksen, L.M. and Hisdal, H. 1997. Regional analysis of extreme streamflow drought duration and deficit volume, *FRIEND'97 - Regional Hydrology: Concepts and Models for Sustainable Water Resource Management.*

[25] Wilhite, D.A. and Glantz, M.H.1985. “Understanding the drought phenomenon: the role of definitions”, *Water Int.,* 10, 111-120.

[26] Yang,Y. and Burn, D.H. 1994. An entropy approach to data collection network design, *J.Hydrol*., 157,307-324.