Water Resour Manage (2016) 30:4399–4413
DOI 10.1007/s11269-016-1428-1

Wavelet Spectrum and Self-Organizing Maps-Based Approach for Hydrologic Regionalization -a Case Study in the Western United States

A. Agarwal[1,2,3] · **R. Maheswaran**[4,5] · **J Kurths**[1,2] · **R. Khosa**[3]

Received: 13 February 2016 / Accepted: 4 July 2016 /
Published online: 12 July 2016

Abstract Hydrologic regionalization deals with the investigation of homogeneity in watersheds and provides a classification of watersheds for regional analysis. The classification thus obtained can be used as a basis for mapping data from gauged to ungauged sites and can improve extreme event prediction. This paper proposes a wavelet power spectrum (WPS) coupled with the self-organizing map method for clustering hydrologic catchments. The application of this technique is implemented for gauged catchments. As a test case study, monthly streamflow records observed at 117 selected catchments throughout the western United States from 1951 through 2002. Further, based on WPS of each station, catchments are classified into homogeneous clusters, which provides a representative WPS pattern for the streamflow stations in each cluster.

Keywords Wavelet power spectrum · Regionalization · Ungauged catchments · K-means technique · Self-organizing map

✉ R. Maheswaran
maheswaran27@yahoo.co.in

[1] Institute of Earth and Environmental Science, University of Potsdam, 14476 Potsdam, Germany

[2] Postdam Institute for Climate Impact Research, Research, P.O. Box 601203, 14412 Potsdam, Germany

[3] Water Resource Engineering, Indian Institute of Technology, Delhi, India

[4] MVGR College of Engineering, Vizianagaram, India

[5] Saint Anthony Falls Laboratory, University of Minnesota, Minneapolis, MN, USA

1 Introduction

In many situations, robust estimation of streamflow at the site of interest is an essential component as it is inevitably required for the design of hydraulic structure, flood studies, engineering developments (dam, diversion structures, power plants) and, more importantly, environmental studies (land use planning and management, stream habitat assessment, extreme events and climate impact studies). However, in many of the developing countries, the length of the records are not long enough to get reliable estimates of extreme events. One basic approach when encountering such situations is to pool the information from other hydrologically similar catchments, a method traditionally termed regionalization (Bloschl and Sivapalan (1995); Saf (2009).

In the past, there have been several attempts (Franchini and Suppo 1996; Allende et al. 2009; Cutore et al. 2007; Coelho et al. 2012; Sang 2013),to develop a suitable framework for regionalization using different approaches. The criteria range all the way vary from similarity in the streamflow signature (Atiem and Harmancioğlu 2006; Goyal and Gupta 2014; Latt et al. 2015), geographical location and catchment characteristics (Chen et al. 2011; Rao and Srinivas 2008) to catchment river network complexity, model parameters and uncertainty (Vandewiele et al. 1991; Cutore et al. 2007; Bock et al. 2015). Razavi and Coulibaly (2013) provide a detailed review of several methods for hydrologic regionalization. Some of the very recent additions to the list include Bock et al. (2015), where the authors applied a parameter regionalization scheme to transfer parameter values and model uncertainty information from gauged to ungauged areas. Sivakumar et al. (2013) proposed a comprehensive account of catchment classification and concludes that there are no established criteria for regionalization due to the i) scarcity of data and the ii) subjectivity involved in the selecting of attributes, weights, threshold value, and distance measure.

Even though there are a plethora of methods available, there are some important bottlenecks which need to be addressed. One important issue, addressed in this paper, is the temporal heterogeneity present in streamflow signatures which may complicate efforts to develop a regionalization framework. Devito et al. (2005) emphasized the need to focus on spatiotemporal similarities and differences between streamflow regimes to create spatially contiguous homogeneous hydrologic clusters. These spatially contiguous regions may be identified by climate variability, physiological characteristics, and statistical similarity of streamflow regimes. In order to capture the underlying temporal characteristics, we will use the concept of the wavelet spectrum combined with clustering algorithms.

In the last decade, wavelet analysis has gained significant attention from researchers from various fields due to its characteristic ability to capture the temporal variability at multiple scales. In the past, wavelets have been used in hydrology for several applications, such as forecasting (Sahay and Srivastava 2014; Kisi 2011; Sehgal et al. 2014a, b),downscaling (Lakhanpal 2015). The power spectrum estimated from the wavelet spectrum provides information about the signature of the energy distribution across scales and can be used for understanding the scale of the dominant processes. In the past, several studies have shown that the wavelet spectrum (Labat 2005) can capture the temporal variability in a streamflow. In the field of regionalization, Saco and Kumar (2000) and Zoppou et al. (2002) used wavelet based stream flow signatures for establishing similarity in catchments. The important issue with the study by Zoppou et al. (2002) is that they have not considered the application of the proposed approach to real-time observed streamflow and have used k-means clustering algorithm. Two key issues with this work are (1) intrinsic disadvantages of k-means, such as being sensitive to initial distribution (initialization procedure), demanding the number of cluster beforehand (Chen et al. 2010), being prone to

getting stuck in local minima, and displaying a result that is biased towards the number and location of the initial codebooks (Shahapurkar and Sundareshan 2004). (2) Regionalization based on model simulation may not fully reflect the catchment characteristics which are essential in these types of studies. As an extension of these works, we intend in the present paper to develop a regionalization framework using wavelet spectrum and self-organizing map (SOM). The SOM has been shown to be a more robust clustering method than k-means in terms of the dependency upon initialization. Several studies have shown that the SOM is a powerful tool for clustering and performs well in the detection of noisy signals, preserves topology and, more importantly, performs better where initialization remains an issue. (Chen et al. 2010).

Therefore, in this paper, attempts are made:

1. To develop a wavelet-coupled SOM-based approach for a regionalization framework, and compare the results with those obtained from the wavelet-based k-means approach. For this purpose, we have utilized the streamflow observed at 117 stations within the western United States.
2. To show that wavelet analysis is a powerful tool for capturing multiscale variability and can be effectively used to characterize the underlying hydrologic system.

Additionally, the temporal evolution of the stations clustering behavior were studied in order to understand the effect of non-stationarity in the hydrologic classification of catchments.

2 Study Area and Dataset

In this study, streamflow stations from western United States (US) are used to test the efficacy of the technique in identifying homogeneous hydrological clusters. Monthly streamflow data over an extensive network of 117 gauging stations are included. Fig. 1 shows the geographical location of these streamflow stations spreading over 11 states. Streamflow data at a monthly scale were obtained from the USGS Geological Survey database (http://nwis.waterdata.usgs.gov/nwis). In the recent past, several similar studies have used this dataset (Sivakumar and Woldemeskel 2015; Agarwal et al. 2016).

Observed streamflow data for the period from October 1951 to September 2002 (i.e. water year) for the total 52 years and all 117 stations were analyzed. It is observed that the records cover a variety of stations and from different geographical locations, climatic influences, drainage characteristics, and land use (for details Refer to Sivakumar and Singh (2012).

3 Methodology

Before embarking on the proposed methodology, we introduce the basic concepts of wavelets and clustering methods, such as the k-means and SOM.

3.1 Wavelet Transform

In the present study, continuous wavelet transform-based power spectra are employed to yield a better representation of the energy distribution of the streamflow signal at different scales. It

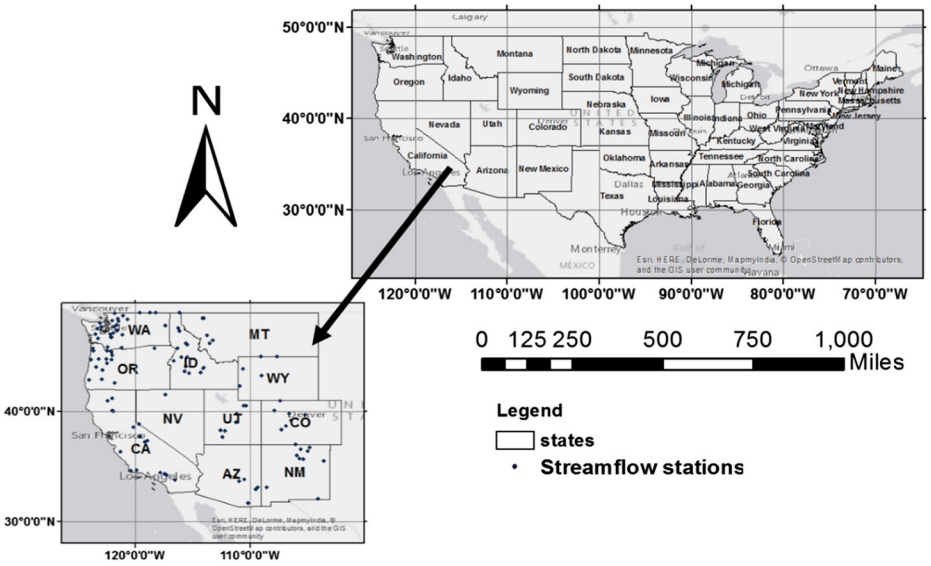

Fig. 1 Geographical location of 117 streamflow stations selected from the western United States where AZ – Arizona; CA – California; CO – Colorado; ID – Idaho; MT – Montana; NM – New Mexico; NV – Nevada; OR – Oregon; UT – Utah; WA – Washington; WY – Wyoming

has been shown in several studies that this is a very useful tool for capturing the variability and information hidden in the raw signal. (Agarwal 2015; Sehgal et al., 2014a, b).

Daubechies (1992) defined the continuous wavelet transform (CWT) for a squared integral function $f(t)$ of time t as:

$$W_{a,b} = \int_{-\infty}^{\infty} f(t)\Psi^*_{a,b}(t).dt \quad a, b \in R, \quad a \neq 0 \tag{1a}$$

with

$$\Psi^*_{a,b}(t) = \frac{1}{\sqrt{|a|}}\Psi\left(\frac{t-b}{a}\right) \tag{1b}$$

where Ψ represents a family of functions called wavelets and * represent the conjugate function. The parameters a and b denote the scale and location, respectively. $W_{a,b}$is the wavelet coefficient at scale a and location b; varying the values of a creates dilation and contraction effects depending on$a > 1$ *or* $a < 1$ respectively. Similarly, by varying b, we can analyze the function at different temporal locations. The factor $\frac{1}{\sqrt{|a|}}$is used to normalize the energy for different values of *a and b*. (Torrence and Compo 1998; Nourani et al. 2009).

The wavelet power at given scale can be estimated as the absolute value squared of the wavelet transform and is given by

$$W_{wp}(a) = \sum_b \left|W_{a,b}\right|^2 \tag{2}$$

Using Equation (2), wavelet power across all values of *'a'* can be estimated; the plot between *'a'* and wavelet power is called global wavelet power spectrum, or wavelet spectral signature, of the given time series. The wavelet power spectrum is a very convenient description of the fluctuation of variance at different frequencies (*'a'*). In the present study,

the wavelet power spectrum of each of the streamflow time series was obtained using above Equation (1)-(2) above.

3.2 K-Means Clustering

The k-means technique (Morissette and Chartier 2013) is a partition-based clustering technique widely used for its easy implementation and fast convergence characteristics. However, the technique suffers from the trap in local minima, along with initialization issues (Rao and Srinivas 2008). For, more information the readers are referred to Rao and Srinivas (2008).

3.3 Self-Organizing Map (SOM)

The SOM, also known as Kohnen's neural network, is an architecture suggested for artificial neural networks which have been used extensively in various fields, such as brain maps (Kohonen 2012), clustering (Chen et al. 2010), competitive learning (Mehta and Jain 2009) etc.

The basic SOM algorithm is iterative, which implies that the SOM is trained using unsupervised learning. The algorithm assumes that all data vectors are located in a d-dimensional space and begins with random weight initialization, which is an attempt to achieve optimum values in the subsequent iterative steps. The mathematical details related are clearly given in Haykin and Lippmann (1994).

Due to its unique ability to locate a pattern even in complicated structures, the SOM has been used extensively in hydrological applications, such as assessing water quality and managing ecosystems (Céréghino and Park 2009), hydrologic modelling and analysis (Zhou et al. 2008), and identification of hydrologic homogeneous regions (Lin and Chen 2006).

3.4 Wavelet Coupled SOM-Based Method for Regionalization

The streamflow data, which is considered as the signature representation of the catchment characteristics, from all stations is standardized by subtracting the mean and dividing by the standard deviation to bring the data to a comparable platform. The standardized streamflow series is transformed into wavelet coefficients at different time and frequency scales using the Morlet wavelet(Giri et al. 2014). Depending on the data span of the streamflow time series (Maheswaran and Khosa 2012), the wavelet coefficients are estimated at up to 7 levels of decomposition. Then, the wavelet power spectrum is estimated for each station (for a total of 117 stations) and is used as the basis to form homogeneous clusters using k-means and self-organizing map techniques. The number of clusters is decided based on measurements such as silhouette values and homogeneity measures. The entire algorithm was coded and implemented using the functions available in MATLAB R2015a.

4 Model Application

As explained above, wavelet coefficients were estimated at different scales using the Morlet wavelet for each of the streamflow series from the 117 stations considered in this study. Fig. 2(a) shows, for example, the normalized streamflow time series observed from Rio Hondo near Valdez, NM (USGS Station #08,276,500) and the results obtained from wavelet analysis.

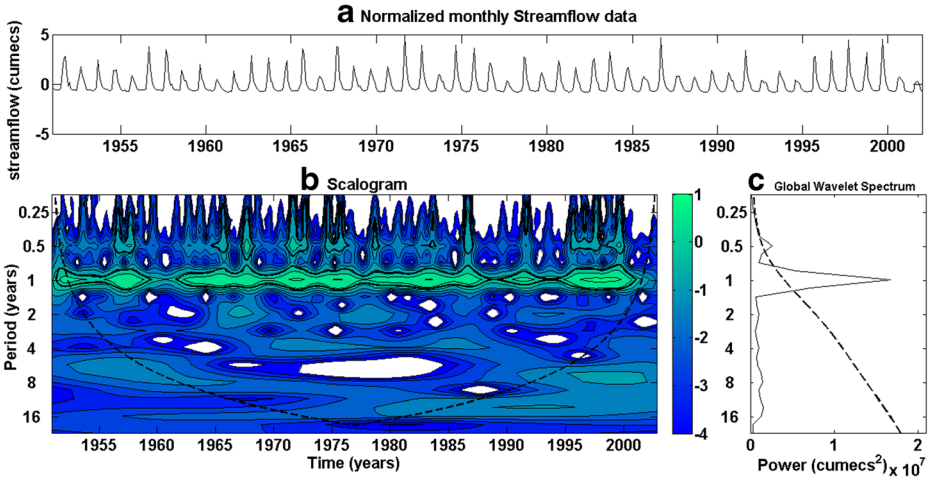

Fig. 2 **a** Normalized monthly streamflow time series from Rio Hondo near Valdez, NM (USGS Station #08276500)); **b** scalogram; **c** global power spectrum. The dashed lines in the scalogram represent the region beyond, which is affected by the boundary effects. The thick contour (dashed lines) indicates a 5% significance level

The normalized wavelet scalogram, i.e. $|W_n(a)|\,\frac{2}{\sigma^2}$ is shown in Fig. 2(b). Fig. 2(c), showing the plot of the power vs. scale, clearly indicates the relative power of features at different scales.

Following a similar procedure, the wavelet spectrum was estimated for all the 117 streamflow stations. The next step was to investigate the spatial organization of these wavelet spectra from all 117 streamflow stations by employing the two clustering techniques (k-means & SOM). The optimal number of clusters was decided using the Dunn and silhouette validation indices described briefly in Agarwal et al. 2016. Interested readers can find more details in (Dunn 1973).

Figure 3 shows the values of the Dunn and silhouette validation indices plotted against the number of clusters for both types of clustering techniques. It is seen that the probable number of clusters are 7, 4 and 3 based on the Dunn index and silhouette value for both k-means and SOM. All three possibilities for the optimal number of clusters were tested and based on proper discretion; one of them was selected.

When the clusters were formed based on an optimal number of 3, it was observed that the clusters were unstableand thereby signaling the need for an additional cluster. Similarly, for an optimal number of 7, most of the stations were present in only four clusters and the remaining three clusters were found to be almost empty. Rao and Srinivas (2008)suggests ignoring these empty clusters and working with the number of highly populated clusters which works out to be 4. Hence, 4 is selected as an optimal number of clusters for further analysis.

Further, observing the higher Dunn index values at cluster 3 (Fig. 3a) obtained using the k-means technique, one might argue that the k-means technique shows good cluster quality in comparison to the SOM technique, but it should be noted that this is an internal validation index (i.e. the clustering result is evaluated based on the data that was clustered itself), and Manning et al. (2008) cautions that high scores on internal criteria in cluster evaluation does not necessarily result in effective information retrieval applications.

Based on these observations, the 117 western United States streamflow stations are segregated into four clusters using the wavelet power spectrum method coupled with clustering

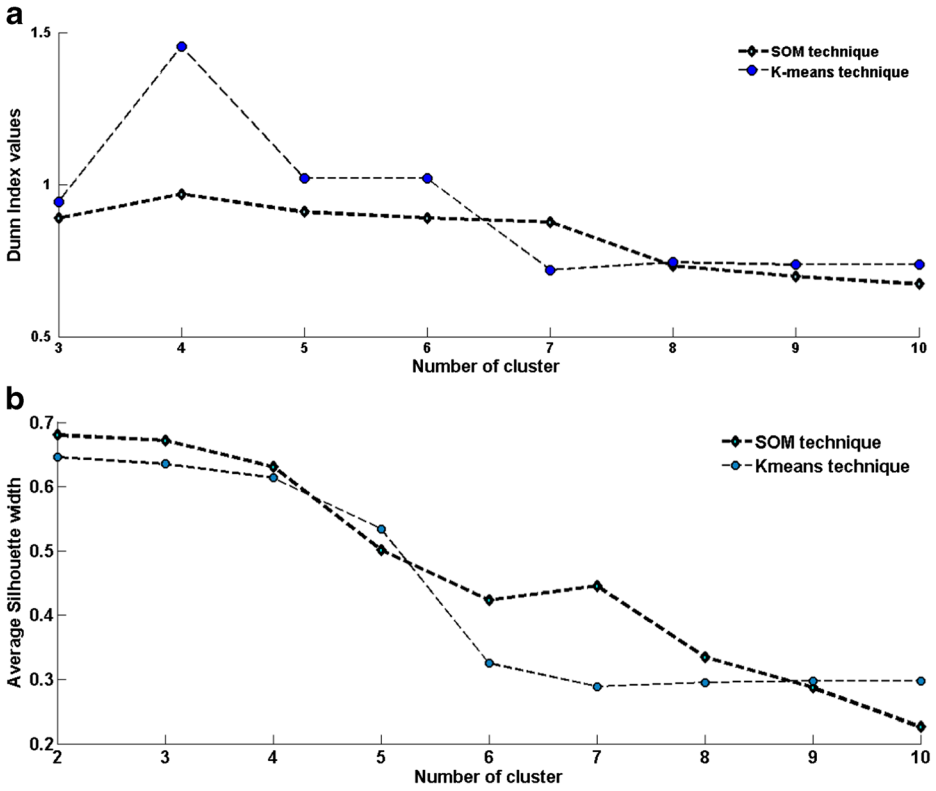

Fig. 3 Plot of cluster indices against the number of clusters: **a** Dunn index, **b** average silhouette width

(k-means & SOM techniques) as explained above. Table 1 shows the number of stations that fall into each cluster category using k-means and SOM techniques. Clusters 1 and 4 from the k-means (SOM) analysis are the largest, and together they contain 84 % (80 %) of the stations, which is similar to the results obtained by Sivakumar and Singh (2012) on the same set of data in which the authors use the correlation dimension method.

However, it is important to note that there were certain differences (total number of stations in the cluster, the position of stations in the cluster) in our result when compared to the one obtained by Sivakumar and Singh (2012). This may be explained by certain limitations of the correlation dimension method, such as (1) underestimating the dimensionality for small data sizes, i.e. not being able to deal with the limitations of the data, (2) the fact that the

Table 1 Number of stations in each cluster

Cluster Number	Number of stations	
	K-means technique	SOM technique
1	70	60
2	2	7
3	16	16
4	29	34

dimensionality and complexity of streamflow could change with respect to the temporal scale and (3) its vulnerability to the effects of data size, noise, temporal correlation, etc.

5 Results and Discussion

The geographical distribution of the resultant clusters is mapped in Fig. 4 for both clusterings (k-means and SOM) techniques. The most prominent feature of the result is that the spatial extent of the streamflow stations in the clusters prominently shows the ability of the WPS to capture the underlying driving forces rather than just forming clusters based on geographical proximity. For example, cluster 1 (in both cases) consists of stations that are uniformly distributed throughout the study area, but the majority of stations belong to WA, OR, MD, and ID, i.e. showing geographical proximity. On the other hand, cluster 4 has its stations equally distributed throughout WA, AZ and NM. Similarly, clusters 2 and 3 show a disperse distribution of stations. These results demonstrate that, apart from the geographical similarity, there is another governing mechanism viz. physiographic features, climatic patterns, similarity at different temporal scales and statistical similarity of streamflow regimes which might influence the streamflow pattern.

It was observed that the average silhouette value was found to be 0.60 and 0.65 for k-means and SOM technique respectively. The general interpretationfollows that high value of silhouette width indicates that the stations are well matched to their clusters and poorly matched to other clusters (an appropriate solution). In the present context, it can be seen that the SOM performs marginally better than the k-means in terms of silhouette value.

Further, in order to explore the similarities in the station characteristics of each cluster, the individual wavelet scalogram and spectrum were examined. Table 2 provides a summary of the distinct, prominent features that are characteristic for the given clusters, and this information can be used for further application in hydrology.

To understand the homogeneity of each cluster (resulting from the k-means and SOM techniques), the normalized global wavelet spectrum (NGWS) is plotted for each station in a given cluster. Fig. 5 clearly shows the NGWS pattern for any station belonging to a given

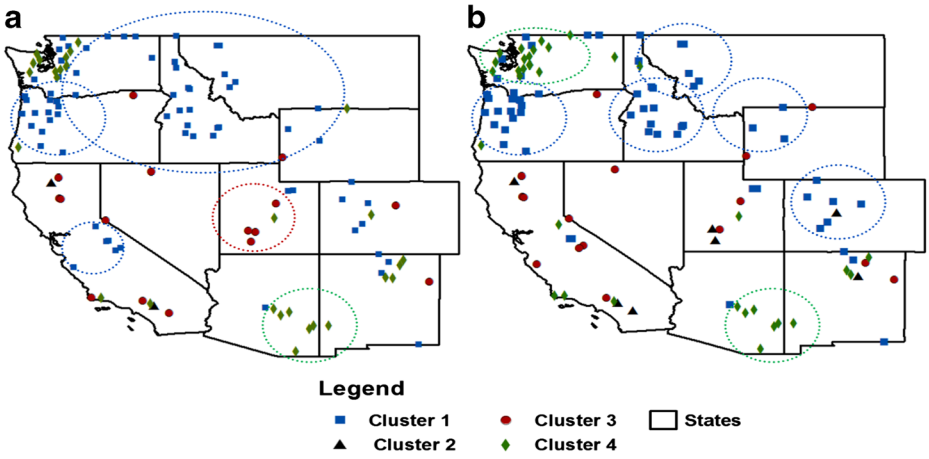

Fig. 4 Cluster-wise geographical distribution of streamflow stations for the (**a**) k-means technique and (**b**) SOM technique

Table 2 Characteristic features identified from the representative wavelet scalogram and spectrum for all 4 clusters

Cluster number	Annual Variation	Decadal Variability	Presence of dominating scale	Geographical spread	Drainage area (square mile)	Elevation (miles)
1	Yes (high)	No (Nil)	Yes (Nil)	High spatial spread	Mean area = 896.73 Maximum area=13550 Minimum area= 22.30	Mean elevation = 3192.991 Maximum elevation =8006.29 Minimum elevation = 32.6
2	No (Nil)	Yes (high)	Yes (Low)	Low spatial spread	Mean area = 152.21 Maximum area=425.00 Minimum area= 8.80	Mean elevation = 4734.286 Maximum elevation =9830.00 Minimum elevation = 700.00
3	Yes (Low)	Yes (High)	No (Medium)	Low spatial spread	Mean area = 335.29 Maximum area= 2060.00 Minimum area= 22.90	Mean elevation = 4318.209 Maximum elevation = 7670.00 Minimum elevation = 339.2
4	Yes (High)	Yes (Medium)	No (High)	High spatial spread	Mean area = 915.11 Maximum area= 7896 Minimum area= 13.40	Mean elevation = 2402.708 Maximum elevation = 7502.94 Minimum elevation = 21.00

cluster. It can be seen that the pattern is unique for a given cluster and differs strongly when compared to any other cluster.

Although it is clear from Fig. 5a and 5b that in both clustering techniques the pattern of normalized global wavelet spectra (NGWS) is almost similar, but under closer inspection of the clusters, it becomes evident that the results from the SOM technique present a clearer pattern.

Recalling Fig. 4, we had observed that some stations move from one cluster to another when we used the SOM as a clustering technique. Close analysis of Fig. 5a reveals that the NGWS of certain stations in clusters 3 and 4 do not have similar spectra when compared with

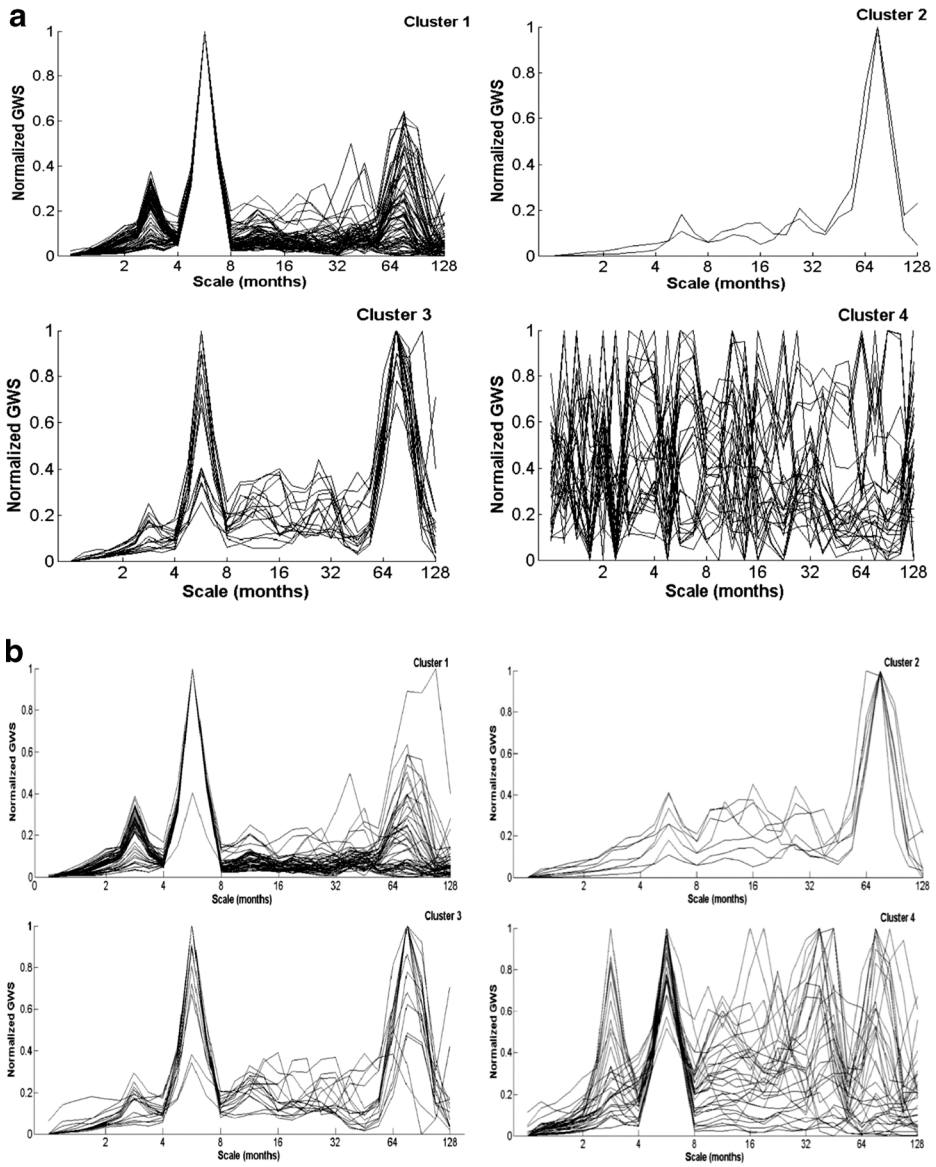

Fig. 5 Normalized global wavelet spectrum (NGWS) for each streamflow station resulting from the (**a**) k-means technique and (**b**) SOM technique

the remaining stations of the respective cluster. For example, it is quite evident from Fig. 5a that some of the streamflow stations from clusters 3 and 4 show a similar pattern to cluster 2 and none of the stations from cluster 1. Moreover, we have already seen in Fig. 4 that a total of five stations move to cluster 2 from clusters 3 and 4 (no stations from cluster 1) upon changing the clustering approach from k-means to SOM. Interestingly, all these five stations have NGWS patterns similar to the one which is characteristic of cluster 2 (Figure 5b). This very clearly shows the superiority of SOM over k-means. This might be because the k-means fails to allocate these stations into the appropriate clusters due to initialization problems associated with it. On the other hand, the SOM technique is quite efficient and can capture the underlying variability. Further, a comparison of the results of both clustering techniques also shows that k-means is sensitive to initial distribution, due to which some of the stations are located in the wrong cluster, whereas the overall clustering performance of the SOM is better than that of k-means. Thus, we propose the use of self-organizing maps as possible substitutes for the classical k-means clustering technique.

The above analysis also revealed that the traditional measure of clustering strength, such as Silhouette values, is not a robust measurement as the values obtained from SOM and k-means were nearly the same (please refer section 3).

Physical Interpretation of the Clusters In this section, we describe the possible physical mechanism underlying the observed pattern of multiscale variability for each of the clusters.

a) Cluster 1

In the case of cluster 1, the variability across scales is associated with wavelet spectra having higher energy values between six-month and annual time scales. Also, there is a high energy mode or oscillation having a scale of >2–3 years. The catchments in cluster 1 are mainly present in states like Idaho, Oregon and Montana, and some in states like Colorado and Wyoming. Most of the precipitation in this region occurs during winter storms. In general, in these regions, rainfall and snow contributions produce a single period of above-normal flow conditions. Moreover, during the summer and fall seasons, dry conditions exist and the streamflow is very low. This kind of seasonal cycle may be the reason for the higher energy at six-month and one-year scales. The high energy on the inter-annual (> 2 years) scale signifies the presence of the low-frequency mode which might represent the baseflow contribution during the summer and fall seasons. During such periods, the majority of the flow of the streams may be from the groundwater contribution. It is also possible that the low-frequency oscillation may be due to the influence of teleconnections with climatic indices driving the precipitation in these regions. However, in order to assert that possibility, the precipitation characteristics across scales must be investigated, which is beyond the scope of the present study.

b) Cluster 2

For the stations which fall into cluster 2, the dominant features are the inter-annual oscillations which have higher energy than that of the intra-annual and annual features. This cluster represents the regions which receive rainfall from November to late April. Further, it is characterized by the presence of the most extensive and productive aquifer, capable of storing a large volume of rainfall which later maintains the baseflow during the dry conditions (Saco and Kumar 2000). The high variability of the rainfall must be canceled out by the great amount of storage facilitated by the large aquifer and regulating dams.

c) Cluster 3

The regions in this cluster are characterized by a high energy distribution at strong annual and inter-annual scales, whereas there is less or no energy at intra-annual scales. These regions have

characteristics similar to the regions that fall into cluster 1, except that the regions in cluster 3 have lower energy at smaller scales. This kind of behavior can be attributed to the fact that streamflow contributions in these regions are mainly through seasonal rainfall, and the baseflow and may not receive much contribution from the snowmelt.

d) Cluster 4

The stations that fall into cluster 4 are generally characterized by an equal distribution of energy across intra-annual, annual and inter-annual scales. The streamflow in such catchments may be associated with annual peaks due to snowmelt, rainfall and baseflow contribution resulting in inter-annual cycles. However, these regions may also have short-term, snowmelt-driven, high flows resulting in flashy type flows which are characterized by high energy at the 3–4-month scale.

Further, in order to study the physical characteristics of the catchments associated with each of the clusters, features such as the area and elevation of the catchment outlet are related to the cluster characteristics resulting from the SOM technique. Figs. 6 (a) and (b) show the box plot of drainage areas and elevation of streamflow stations for each cluster. The figures show that the resultant clusters are, to some degree, stratified by both drainage area and elevation.

Effects of Non-Stationarity on the Clustering To study the ability of the proposed method to capture the non-stationarity of the association of stations with each other, the entire data set was divided into three segments (the first two spanning 20 years and the third one the remaining period of 22 years) with an overlapping period of 5 years. The clustering analysis was done using the wavelet-SOM method for each segment of the data, and the results from each of the data segments were compared to the others to understand the evolution of the association or relationship between the stations in a cluster with time. Fig. 7 shows the results of the analysis where it can be seen that there is a movement of stations across the clusters, and this clearly indicates the ability of the proposed approach to capture the temporal evolution of the station characteristics, possibly due to land use change, change in climatic influences, etc. It can be seen that there is remarkable evidence of non-stationarity in the streamflow behavior expressed, in terms of varying intensity/energy over time. For example, the stations in cluster 2 show intermittent features at the annual scale. However, the reasons for the non-stationarity are beyond the scope of the present study.

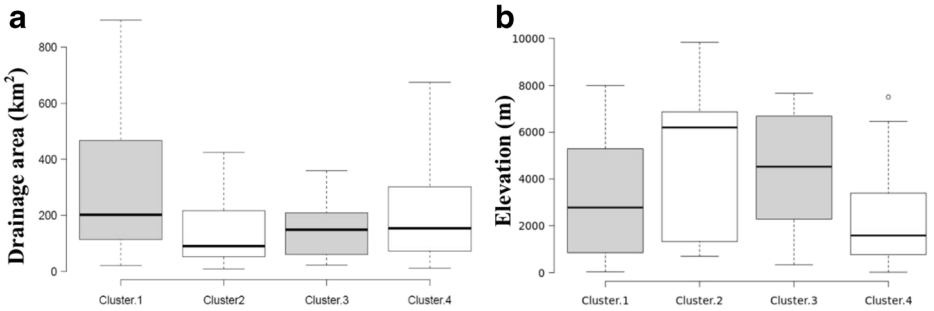

Fig. 6 Box plot of streamflow stations in each cluster using wavelet SOM technique (**a**) Drainage area, (**b**) Elevation

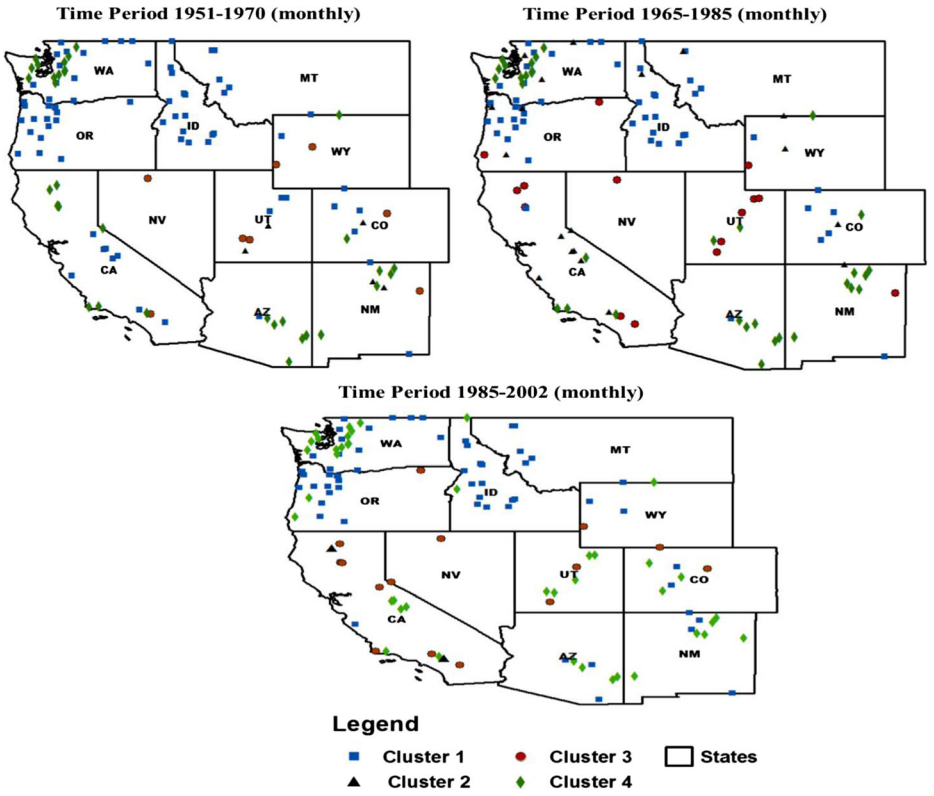

Fig. 7 Cluster-wise geographical distribution of streamflow stations with different time period (months)

6 Conclusion

This study has attempted to develop a wavelet-based SOM method for catchment regionalization. Application of the method to streamflow data from 117 monitoring stations in the contiguous western United States offered promising results for regionalization. The results lead to the following concluding remarks:

a) The wavelet power spectrum appears to be an important statistic in capturing the catchment characteristics. The 117 stations studied are categorized into 4 clusters, each having a distinct wavelet power spectrum pattern across the different scales considered.
b) The proposed clustering method coupled wavelet-based approach for regionalization of hydrologic catchments overcomes some of the limitations(such as data limitation, temporal scale variation, dimensionality, etc.) of the existing approaches.
c) The analysis of the clusters revealed the physical mechanism underlying a homogeneous region. Based on the previous studies, an attempt was made to provide a physical justification for the set of clusters obtained. It was observed that the timing and intensity of rainfall, snow, and contributions from groundwater influenced the streamflow signature of the catchments. However, further investigation is necessary to understand how the catchment characteristics influence the streamflow signature.

d) The comparison of the results of the two clustering techniques shows that the k-means is sensitive to initiative distribution, whereas the overall clustering performance of SOMs is better than that of k-means. Thus, we propose the use of self-organizing maps as a possible substitute for the most classical k-means clustering technique.

Acknowledgments This research was funded by Deutsche Forschungsgemeinschaft (DFG) (GRK 2043/1) within the graduate research training group "Natural risk in a changing world (NatRiskChange) at the University of Potsdam and the Department of Science and Technology, India, through the INSPIRE Faculty Fellowship held by MaheswaranRathinasamy.

References

Agarwal A (2015) Hydrologic regionalization using wavelet-based multiscale entropy technique. Dissertation, Indian Institute of Technology Delhi

Agarwal A, Maheswaran R, Sehgal V, Khosa R, Sivakumar B, Bernhofer C (2016) Hydrologic regionalization using wavelet-based multiscale entropy method. J Hydrol 538:22–32

Allende TC, Mendoza ME, and Lopez GE, Morales-Manilla L (2009) Hydrogeographical regionalization: an approach for evaluating the effects of land cover change in watersheds. A case study in the Cuitzeo Lake Watershed, Central Mexico. Water Resour Manag 23(12):2587–2603

Atiem IA, Harmancioğlu NB (2006) Assessment of Regional Floods Using L-Moments Approach: The Case of the River Nile. Water Resour Manag 20(5):723–747

Bloschl G, Sivapalan M (1995) Scale issues in hydrological modeling: a review. Hydrol Process 9(3–4):251–290

Bock AR, Hay LE, McCabe GJ, Markstrom SL, Atkinson RD (2015) Parameter regionalization of a monthly water balance model for the conterminous United States. Hydrol Earth SystSc 12:10023–10066

Céréghino R, Park YS (2009) Review of the self-organizing map approach in water resources: a commentary. Environ Modell & Softw 24(8):945–947

Chen Y, Qin B, Liu T, Liu Y, Li S (2010) The Comparison of SOM and K-means for Text Clustering. Comput Inform Sci 3(2):268

Chen LH, Lin GF, Hsu CW (2011) Development of Design Hyetographs for Ungauged Sites Using an Approach Combining PCA, SOM and Kriging Methods. Water Resour Manag 25(8):1995–2013

Coelho AC, Labadie JW, Fontane DG (2012) Multicriteria decision support system for regionalization of integrated water resources management. Water Resour Manag 26(5):1325–1346

Cutore P, Cristaudo G, Campisano A, Modica C, Cancelliere A, Rossi G (2007) Regional Models for the Estimation of Streamflow Series in Ungauged Basins. Water Resour Manag 21(5):789–800

Daubechies I (1992) Ten lectures on wavelets, Philadelphia: Society for industrial and applied mathematics (Vol. 61):198–202

Devito K, Creed I, Gan T, Mendoza C, Petrone R, Silins U, Smerdon B (2005) A framework for broad-scale classification of hydrologic response units on the Boreal Plain: is topography the last thing to consider? Hydrol Process (19):1705–1714

Dunn JC (1973) A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters (73):32–57

Franchini M, Suppo M (1996) Regional analysis of flow duration curves for a limestone region. Water Resour Manag 10(3): 199–218

Giri BK, Mitra C, Panigrahi PK, Iyengar AS (2014) Multi-scale dynamics of glow discharge plasma through wavelets: Self-similar behavior to neutral turbulence and dissipation. Chaos 24(4):0431–0435

Goyal MK, Gupta V (2014) Identification of homogeneous rainfall regimes in Northeast Region of India using fuzzy cluster analysis. Water Resour Manag 28(13):4491–4511

Haykin S, Lippmann R (1994) Neural networks, a comprehensive foundation. Int J Neural Syst 5(4):363–364

Kisi O (2011) Wavelet regression model as an alternative to neural networks for river stage forecasting. Water Resour Manag 25(2):579–600

Kohonen T (2012) Self-organization and associative memory (Vol. 8). Springer-Verlag New York Inc, New York

Labat D (2005) Recent advances in wavelet analyses: part 1. A review of concepts. J Hydrol 314(1):275–288

Lakhanpal A (2015) Statistical downscaling of GCM outputs using wavelet based model. Dissertation, Indian Institute of Technology Delhi

Latt ZZ, Wittenberg H, Urban B (2015) Clustering hydrological homogeneous regions and neural network based index flood estimation for ungauged catchments: an Example of the Chindwin River in Myanmar. Water Resour Manag 29(3):913–928

Lin GF, Chen LH (2006) Identification of homogeneous regions for regional frequency analysis using the self-organizing map. J Hydrol 324(1):1–9

Maheswaran R, Khosa R (2012) Wavelet–Volterra coupled model for monthly stream flow forecasting. J Hydrol (450):320–335

Manning CD, Raghavan P, Schutze H (2008) Introduction to Information Retrieval. In: Cambridge University press, Cambridge, England pp 450–416

Mehta R, Jain SK (2009) optimal operation of a multi-purpose reservoir using neuro-fuzzy technique. Water Resour Manag 23:509–529

Morissette L, Chartier S (2013) The k-means clustering technique: General considerations and implementation in Mathematica. Tutor Quant Methods Psychol 9(1):15–24

Nourani V, Komasi M, Mano A (2009) A multivariate ANN-wavelet approach for rainfall–runoff modeling. Water ResourManag 23(14):2877–2894

Rao AR, Srinivas V (2008) Regionalization of watersheds: an approach based on cluster analysis. Springer Netherlands. doi:10.1007/978-1-4020-6852-2

Razavi T, Coulibaly P (2013) Streamflow prediction in ungauged basins: Review of regionalization methods. J Hydrol Eng 18(8):958–975

Saco P, Kumar P (2000) Coherent modes in multiscale variability of streamflow over the United States. Water Resour Res 36(4):1049–1067

Saf B (2009) Regional flood frequency analysis using L-moments for the West Mediterranean region of Turkey. Water ResourManag 23(3):531–551

Sahay RR, Srivastava A (2014) Predicting monsoon floods in rivers embedding wavelet transform, genetic algorithm and neural network. Water Resour Manag 28(2):301–317

Sang YF (2013) Improved wavelet modeling framework for hydrologic time series forecasting. Water Resour Manag 27(8):2807–2821

Sehgal V, Sahay RR, Chatterjee C (2014a) Effect of utilization of discrete wavelet components on flood forecasting performance of wavelet based ANFIS models. Water Resour Manag 28(6):1733–1749

Sehgal V, Tiwari MK, Chatterjee C (2014b) Wavelet bootstrap multiple linear regression based hybrid modeling for daily River discharge forecasting. Water ResourManag 28(10): 2793–2811

Shahapurkar SS, Sundareshan MK (2004) Comparison of self-organizing map with k-means hierarchical clustering for bioinformatics applications. Neural Netw, Int Joint Conference 2:1221–1226

Sivakumar B, Singh VP (2012) Hydrologic system complexity and nonlinear dynamic concepts for a catchment classification framework. Hydrol Earth Syst Sc 16(11):4119–4131

Sivakumar B, Woldemeskel FM (2015) A network-based analysis of spatial rainfall connections. Enviro Modell & Soft 69:55–62

Sivakumar B, Singh VP, Berndtsson R, Khan SK (2013) Catchment classification framework in hydrology: challenges and directions. J Hydrol Eng 20(1):A4014002

Torrence C, Compo GP (1998) A practical guide to wavelet analysis. B Am Meteorol Soc 79(1):61–78

Vandewiele GL, CY X, Huybrechts W (1991) Regionalisation of physically-based water balance models in Belgium. Application to ungauged catchments. Water Resour Manag 5(3):199–208

Zhou HC, Peng Y, Liang GH (2008) The research of monthly discharge predictor-corrector model based on wavelet decomposition. Water ResourManag 22(1):217–227

Zoppou C, Neilsen O, Zhang L (2002) Regionalization of daily stream flow in Australia using wavelets and k-means analysis Tech. Rep., Australian National University. Available from http://www.maths.anu.edu.au/research.reports/mrr/mrr02.003/abs.html