Asian Journal of Water, Environment and Pollution, Vol. 17, No. 4 (2020), pp. 11-21.
DOI 10.3233/AJW200046

Clustering of Groundwater Wells and Spatial Variation of Groundwater Recharge in Sina Basin, India

Thendiyath Roshni*, Jesu V. Nayahi[1], Madan K. Jha[2], Mandal Nehar, Choudhary Sourav and Pawan S. Wable[3]

Department of Civil Engineering, National Institute of Technology Patna, Bihar - 800013, India
[1]Department of Computer Science, Anna University, Chennai - 627007, India
[2]Department of Agricultural and Food Engineering, Indian Institute of Technology Kharagpur - 721302, India
[3]Visiting Scientist (Hydrology), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India
✉ roshni@nitp.ac.in

Received October 25, 2018; revised and accepted July 18, 2020

Abstract: A spatial and temporal analysis of groundwater levels, topography, and precipitation is required to properly manage the groundwater resource. The present paper explains it in two parts: (1) spatial analysis of groundwater levels and selection of suitable clustering approach for selection of representative wells and (2) spatial and temporal variations of groundwater recharge calculated by three numerical models: Chaturvedi model, Amritsar model and ERAS model. Four clustering techniques including K-Means clustering algorithm, Hierarchical clustering technique, canopy and expectation maximisation (EM) were used for the clustering of groundwater levels. Among these, the canopy technique presents more reliable results compared to the other techniques for the spatial analysis of groundwater levels and the formation of representative wells in the Sina basin. For the groundwater recharge estimation, Chaturvedi model and ERAS model values were found very close. The recharge values show consistency with the precipitation data and found that 15% of precipitation contributes to annual groundwater recharge. Spatio-temporal variation of groundwater recharge correlated with precipitation is also carried out for the selected basin. The results show a drastic decline in the groundwater recharge from the year 1990 to 2008. An empirical expression is also developed for groundwater recharge estimation in terms of groundwater level. This provides regional scale information about the basin and helps to understand the groundwater exploitation scenario for instance.

Keywords: Clustering, groundwater, recharge, ERAS model, spatio-temporal analysis.

Introduction

About 34% of the annual water supply contributed by groundwater is an important freshwater resource to mankind. It is essential to meet the needs of all aspects of life, which varies from industrial to cultural requirements. The surface water available for drinking, industrialisation and irrigation purposes cannot fulfil the requirement; hence, the high demand for groundwater. Due to the dense population, random distribution of water resources, change of climate and economic development there is increased demand for surface water (Krishnamurthy et al., 1996; Murthy, 2000; Selvam et al., 2014).The hike in demand and dependency is increasing exploitation of ground water resources. Therefore, quantifying the rate of natural groundwater recharge is a basic prerequisite for the efficient management of scarce groundwater resources.

*Corresponding Author

The natural groundwater recharge is a very complex and dynamic phenomenon. Its determination involves several unresolved problems that require additional research, thereby making the direct measurement of natural recharge extremely difficult (Bouwer, 2002; Roshni et al., 2019). Researchers are more interested in the systematic approach and remediations because the availability of water resource is dependent on various climatic factors. Hence, it is important to understand and study the rainfall patterns and their impact on groundwater recharge. Due to large datasets, sometimes it is difficult for the spatio-temporal analysis. Hence, clustering is carried out for convenience in the analysis (Sahoo and Jha, 2016). For the recharge estimation of spatially varied groundwater level fluctuations data, a representative sample for each cluster is convenient for the estimation and further analysis. Under cluster analysis, a data analysis tool for clustering data with similar characteristics forming a meaningful group are referred (Sahoo and Jha, 2016).

Thus, the motivation for this study was three-fold: (1) to compare and select a suitable clustering technique for grouping spatially varied groundwater levels in the Sina basin, Maharashtra. This will ease the temporal analysis of groundwater levels when there is huge number of spatial data available. (2) To compare and quantify groundwater recharge using the Amritsar model (Sehgal, 1973), Chaturvedi model (Chaturvedi, 1973) and ERAS model (Bougharìou et al., 2014). (3) To relate the groundwater recharge with groundwater level. This provides a clear picture of the groundwater recharge scenario in the Sina basin. The results are analysed and correlated with the variation of groundwater level and precipitation

Study Area

The Sina river basin is a drought-prone area, located in Maharashtra state, India characterised under a semiarid region. Figure 1 shows the study area. It has nine

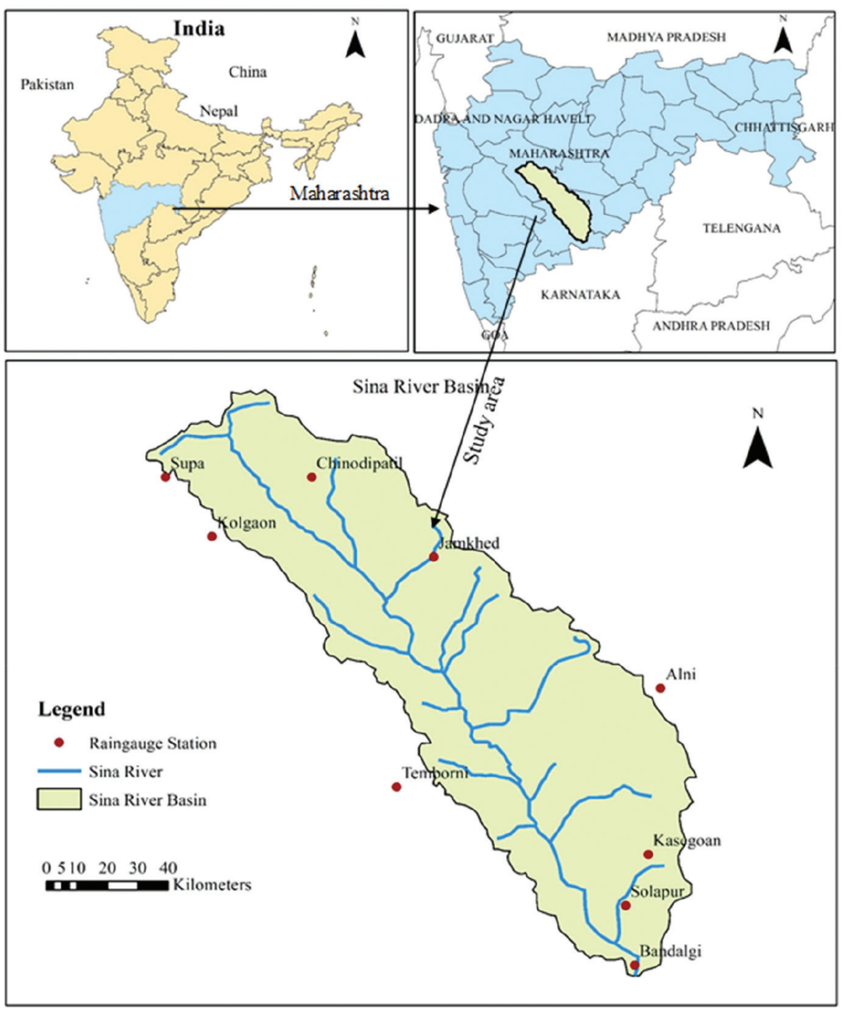

Figure 1: Study area map of the Sina basin.

raingauge stations: Temborni, Kasegoan, Chinodipatil, Bandalgi, Solapur, Alni, Jamkhed, Kolgaon, and Supa. The details of the basic basin data are tabulated in Table 1. The groundwater level data is collected for 133 wells for the pre-monsoon season for the period 1985-2008 (Wable et al., 2017). The monthly precipitation data for nine rain gauge stations were collected from IMD Pune from 1985 through 2008 and the groundwater level data from Centre for Groundwater Board (CGWB).

Table 1: Information about the study area

Latitude	17°28′N–19°16′N
Longitude	74°28′E–76°7′E
Geographical area	12,444 km^2
Average annual rainfall	644 mm
Topography	420–964 m MSL
Climate	Semiarid
Average maximum temperature	40.5° C (May)
Average minimum temperature	10.5 °C (December)

Methodology

Clustering Process

Clustering (Han and Kamber, 2006; Written and Frank, 2005) is the process of grouping the input data based on similarity. The types of clustering includes K-Means, Hierarchical, Expectation maximisation and Canopy.

The K-means (Arthur and Vassilvitskii, 2007; Han and Kamber, 2006; Written and Frank, 2005) is a partitioning-based clustering technique used to group the given data based on similarity estimated using distance metrics. The Euclidean distance metric is used to compute the distance between two points or objects in this algorithm. Hierarchical clustering (Han and Kamber, 2006; Written and Frank, 2005) are of two types mainly agglomerative and divisive. The agglomerative hierarchical clustering technique starts clustering each object being considered as a single element cluster. Then based on the similarity the most similar objects are grouped together. The divisive hierarchical clustering algorithm is the reverse process in which the algorithm starts with all objects forming a single cluster. Expectation maximization (EM) (Han and Kamber, 2006; Written and Frank, 2005) algorithm is a model-based clustering technique used to fit the given data to a mathematical model. EM clustering algorithm is used to estimate the probability distribution of each cluster. Likewise, Canopy clustering (McCallum, 2000) is used as a pre-clustering technique to handle large data sets. Canopy is formed by a random point chosen from the given data set and it is removed from the set of input data.

Selection of Representative Well

Cluster analysis was carried out using K-Means, EM, hierarchical and canopy methods. The statistical method was employed to get the representative wells of selected clusters (Figure 2).

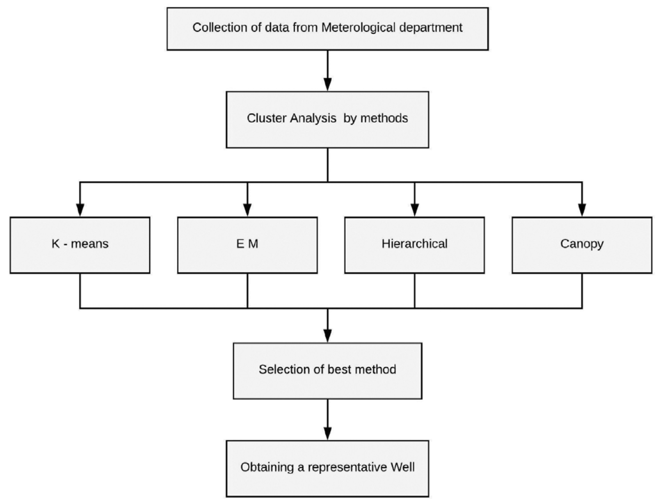

Figure 2: Flow chart of Cluster analysis for the selection of representative well.

Equivalent Rainfall Computed using Thiessen Polygons

For the calculation of average rainfall, Thiessen polygon using area-weighted method was employed due to the smoothness of calculation of average rainfall when the station is located beyond the boundary of drainage (Bartoletti et al., 2018). It has its superiority over the arithmetic method as the weightage of the rainfall station is considered.

Groundwater Recharge Estimation

The groundwater recharge was calculated using the well-known Amritsar model (Sehgal, 1973), Chaturvedi model (Chaturvedi, 1973) and ERAS model (Boughariou et al., 2015). The Amritsar formula was by Sehgal in 1973 using regression analysis for certain inter-streams alluvial tracts in Punjab. Hence, the formula met a good agreement with the actual observation having precipitation between 580 and 710 mm. The Amritsar model (Sehgal, 1973) is written as

$$R = 2.5\ (P - 16)^{0.5} \tag{1}$$

where R is recharge in inch and P is precipitation in inch.

Chaturvedi in 1943 developed an empirical equation, which calculates the amount of rainfall that infiltrates in the ground when it exceeds 380 mm (15 inches) and the expression is based on the water table fluctuation method. The major constraint lies in its narrower region of precipitation values beyond which the empirical equation gets changed by considering the other factors. The expression (Chaturvedi, 1973) is

$$R = 1.35\ (P - 14)^{0.5} \tag{2}$$

where R is recharge in inch and P is precipitation in inch.

Likewise, by studying the Alomartes spring in Granda, an ERAS code was derived in mathematical terms from the MEDA model. The recharge of the aquifer under ERAS is considered to be a fraction of effective rainfall worked out as the difference between effective precipitation and the actual evapotranspiration. The expression is given as (Boughariou et al., 2015)

$$R = M\ (P - T^{\beta})^{N} \tag{3}$$

where P is the annual precipitation (mm), T is the average air temperature (°C) and β is the dimensionless calibration parameter for the conversion of temperature into potential evapotranspiration. The value of β ranges from 1.3 (cold zones) to 1.6 in warmer zones. The value of M lies in the range of 0.01 and 0.06 and N in the range of 1.22 and 0.9. Suitable combinations of constants are selected based on the characteristics of the area for the recharge estimation. The values of constants are assumed for the arid regions (Boughariou et al., 2015).

However, one of the major constraints related to the ERAS model ($R \leq P$) is its model satisfying equation:

$$M \times (P - T^{\beta})^{N} \leq (P - T^{\beta}) \tag{4}$$

$$M \times (P - T^{\beta})^{N-1} \leq 1 \tag{5}$$

The output of the model (equation 3) is correct only when it satisfies the equations 4 and 5.

Finally, the groundwater recharge using different models are calculated and compared in nine regions delineated by Thiessen polygons.

Results and Discussion

Clustering Analysis

The experiments on clustering techniques were conducted using Weka 3.8.1 tool (Hall et al., 2009). The groundwater level collected in 133 wells in the Sina basin during 1990, 1995, 2000, 2005 and 2008 is clustered using four different clustering techniques such as K-means, hierarchical, expectation maximisation (EM) and canopy. The clustering algorithms were executed to produce five clusters.

Figure 3 shows the box plot, which represents the average groundwater depth determined by the different clustering algorithms with the number of clusters being set as 5 (Cluster 0, Cluster 1, Cluster 2, Cluster 3 and Cluster 4).

A review on the pre-clustering approach to K-Means Clustering (Kumar et al., 2014) explains that canopy clustering method effectively divides the data into overlapping subsets by approximately measuring the distance between the data within the cluster.

Likewise, if more than four clustering algorithm is considered, then analysis from other models includes clustering by density-based models such as density-based spatial clustering of application with noise (DBSCAN), ordering points to identify the clustering structure (OPTICS) and Subspace models such as biclustering (co-clustering of two or more clustering components) (Wikipedia.org).

Hence, in the present study, the canopy method of clustering showed very less variation as its groundwater level was more staggered and the spatial variation of data was very close to the average groundwater depth for that cluster. Hence, the canopy method is selected (due to less variation of data) (Table 2) for the clustering of groundwater levels in the Sina basin and the variation of groundwater levels for five clusters is graphically represented in Figure 4. Figure 4a is the scatter plot of the groundwater level at various wells grouped by a canopy and Figure 4b is the histogram showing the average groundwater depth over all the 5 years taken in the study fitted using the normal distribution.

Selection of Representative Well

The representative well in the current study was mainly calculated as the arithmetic average of all the wells present inside the cluster. Arithmetic average proved to be a good option due to increasing staggering behaviour of all the wells clubbed inside each cluster (Figure 5).

Equivalent Rainfall Computed by Thiessen Polygon

For the calculation of average rainfall for a given area, the Thiessen polygon method was used. It has its superiority over the arithmetic method as the weightage of the rainfall station is considered. The Thiessen polygon for different rain gauge stations shown in Figure 6 helps to calculate the average precipitation.

Computation of Groundwater Recharge

Groundwater recharge was calculated for each Thiessen polygons to estimate the recharge potential zones. Of

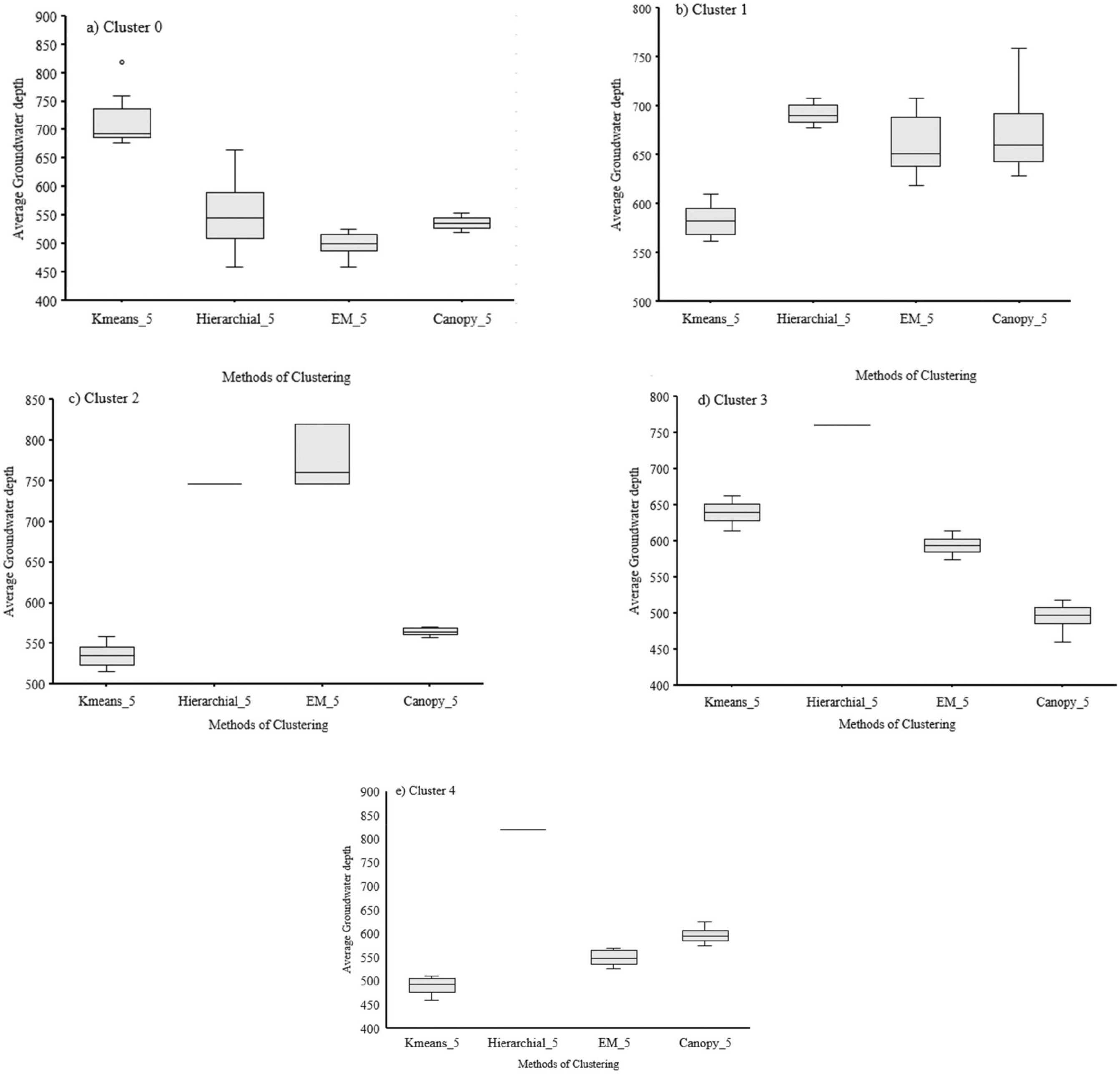

Figure 3: Box plot of the average groundwater levels by different clustering methods (a) Cluster 0, (b) Cluster 1, (c) Cluster 2, (d) Cluster 3, (e) Cluster 4.

Table 2: Standard deviation values for different clusters by four selected methods

Methods	*Cluster 0*	*Cluster 1*	*Cluster 2*	*Cluster 3*	*Cluster 4*
Kmeans	41.783	15.040559	13.187479	14.493442	15.081789
Hierarchical	51.562	10.470436	-	-	-
EM	18.388	27.077941	38.972534	11.77517	14.066875
Canopy	**10.083**	**44.962293**	**4.2654849**	16.898623	3.6228277

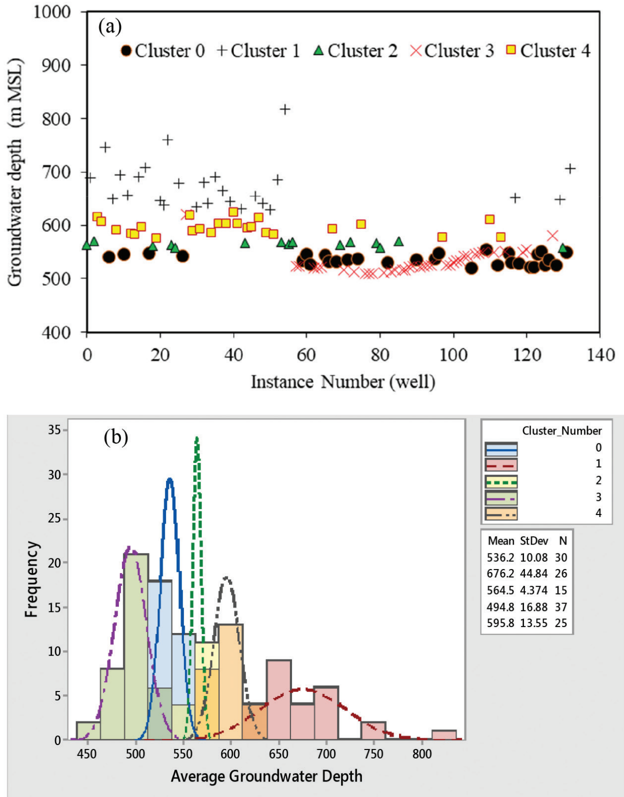

Figure 4: (a) A plot of groundwater wells variation for different clusters in canopy clustering and (b) Histogram plot of average groundwater depth.

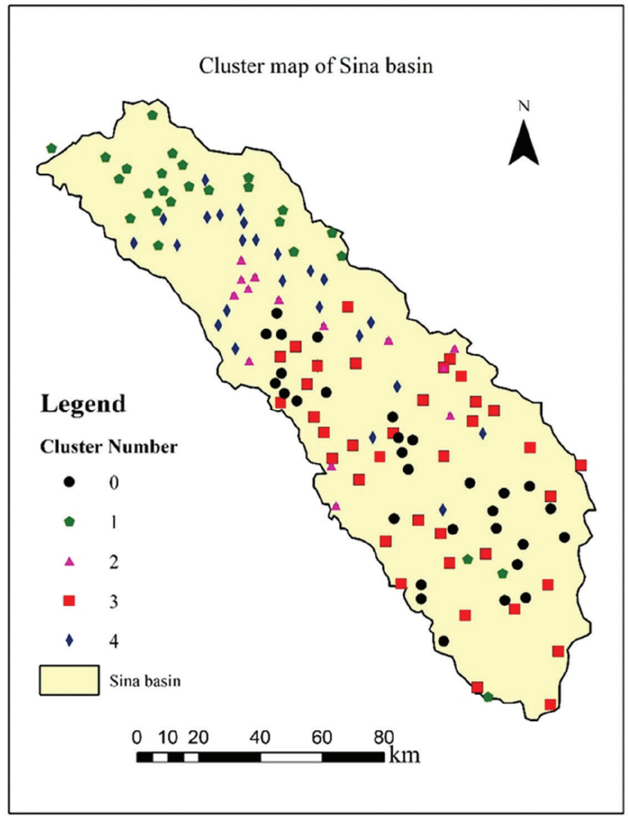

Figure 5: Cluster Map of Sina basin.

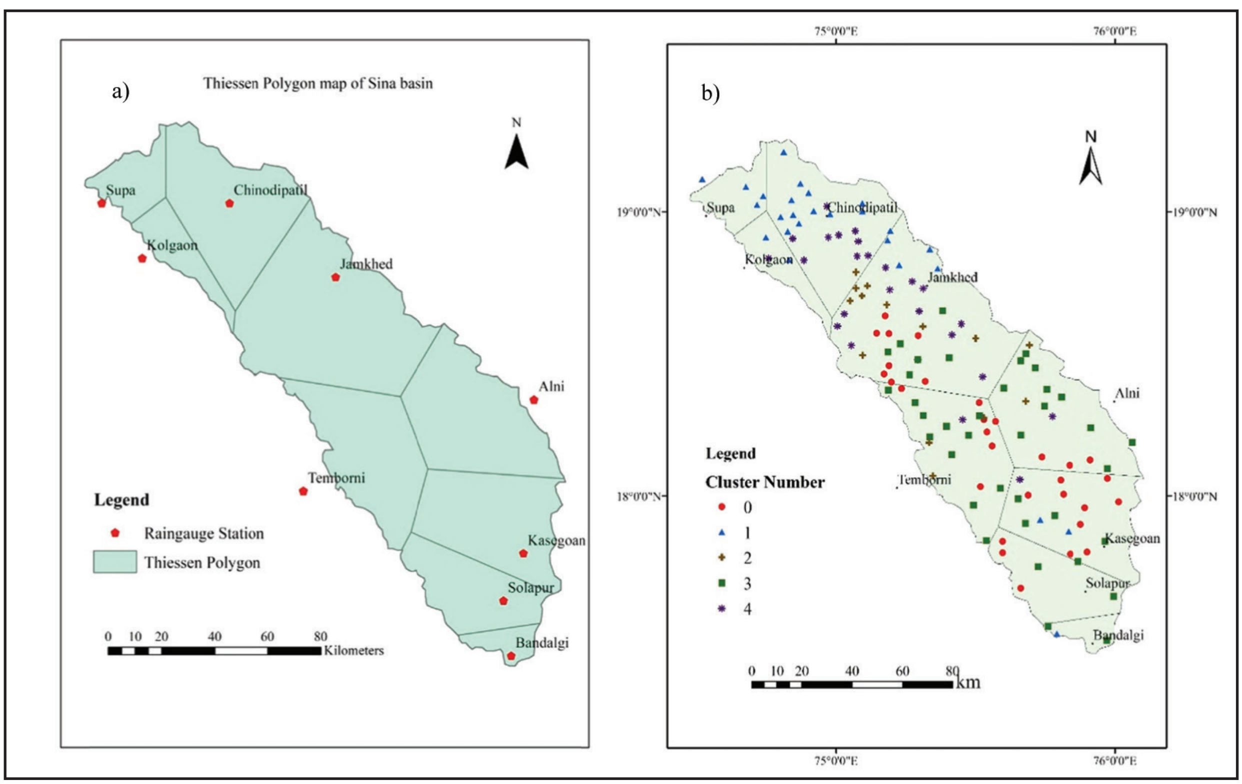

Figure 6: A plot of Thiessen polygon (a) for different rain gauge stations in the basin (b) along with five selected groundwater level clusters.

these three models, it was found that Chaturvedi model and ERAS model showed almost similar values of average annual groundwater recharge (Table 3) for nine Thiessen polygons (Figure 7a). Figure 7b shows the percentage contribution of precipitation in the annual recharge of Sina basin for the period 1990-2005. The distribution of groundwater recharge is similar to that of precipitation while the percentage of precipitation that contributes to recharge varies from 9% to 22% leaving an average of 15%. Spatial variations of groundwater recharge calculated by ERAS model are shown in Figure 8. To identify the recharge potential zones, a plot of topography map and spatial analysis of groundwater level map is analysed (Figure 9a, b). The figure showed higher elevation at Supa, Chinodipatil and Kolgoan rain gauge stations and a lower slope towards the Kasegoan,

Table 3: Recharge calculation for the rain gauge stations

Polygon	*Stations*	*Chaturvedi Model (mm/year)*	*Amritsar Model (mm/year)*	*ERAS Model (mm/year)*
1.	Temborni	103.14	168.56	101.72
2.	Kasegoan	116.44	196.05	119.65
3.	Chinodipatil	76.28	109.04	73.40
4.	Bandalgi	127.67	218.70	136.91
5.	Solapur	129.64	222.64	140.15
6.	Alni	151.00	264.81	179.43
7.	Jamkhed	124.77	212.88	132.26
8.	Kolgaon	56.58	53.98	58.77
9.	Supa	56.36	53.19	58.63

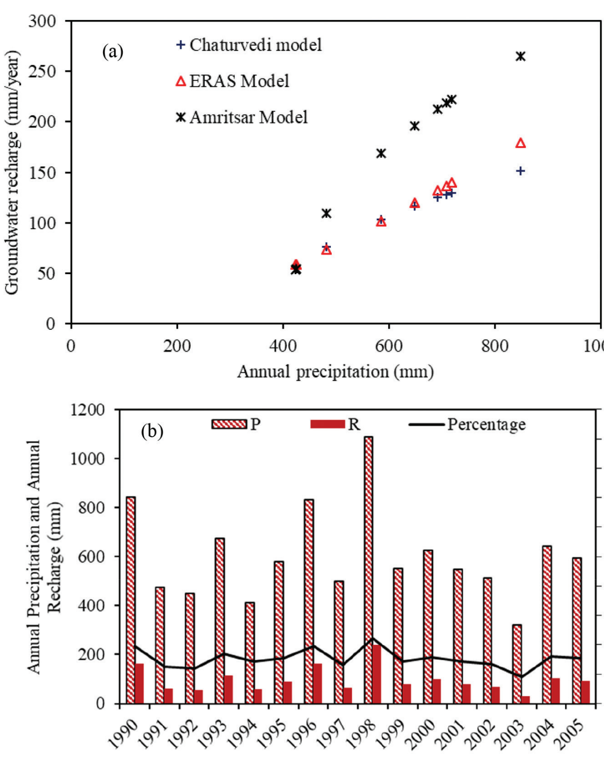

Figure 7: (a) Comparison of annual groundwater recharge by Chaturvedi model, ERAS model and Amritsar model (b) Percentage of annual precipitation (P) to groundwater recharge (R).

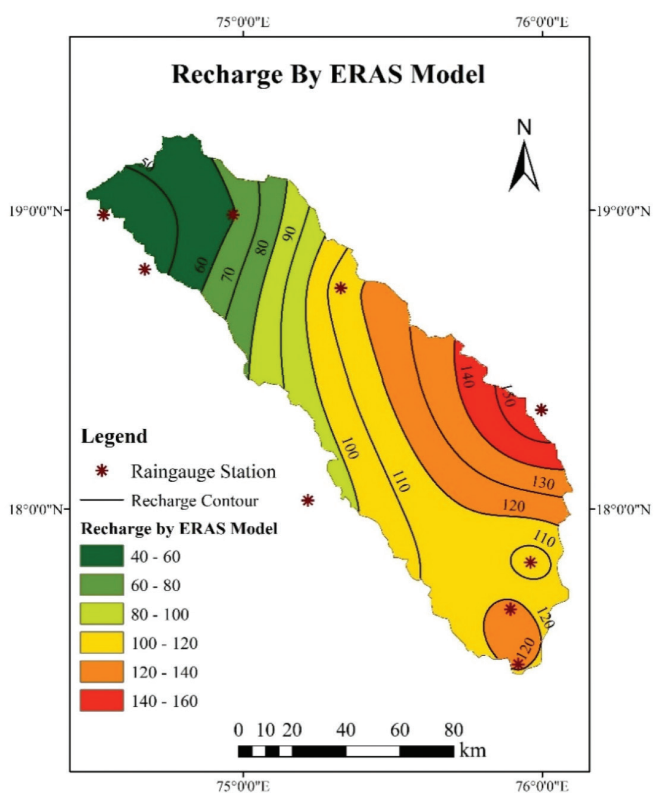

Figure 8: Recharge potential map of Sina basin using ERAS model.

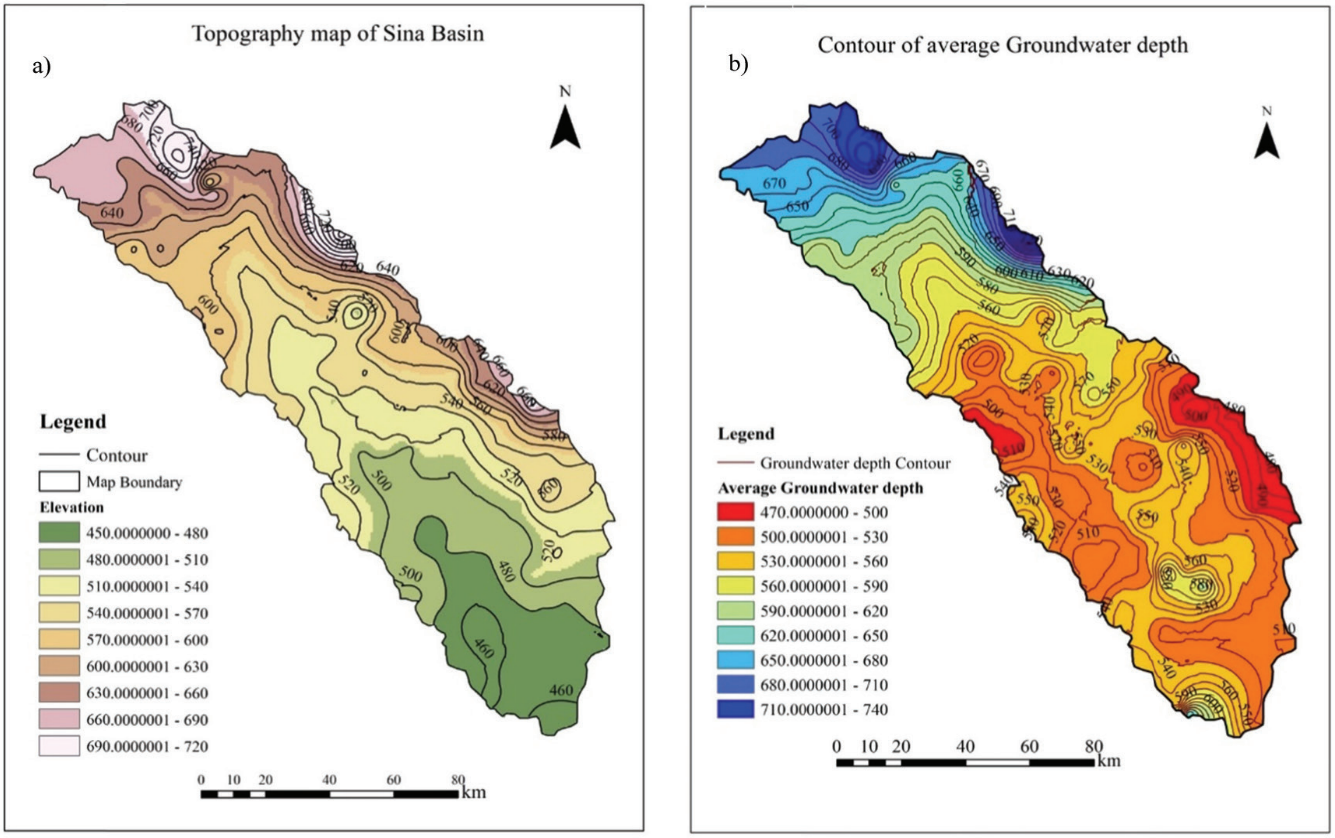

Figure 9: (a) Topography map of Sina basin and (b) average groundwater level by Canopy method.

Solapur and Bandalgi stations. These regions situated in the south of the basin are the expected targets for the best recharge zones. Alni region was examined as the best region of higher recharge potential zone by Chaturvedi, Amritsar and ERAS models.

Hence, the groundwater recharge was calculated by the above mentioned three models, namely the Chaturvedi model, ERAS model and Amritsar model. However, for the determination of recharge other groundwater recharge models include land surface models such as variable infiltration capacity (VIC), MODIS dataset, Grace models, Wetspass model, Noah model and Mosaic model (Niraula et al., 2017).

Spatio-Temporal Variation of Groundwater Recharge with Rainfall

The spatio-temporal variation of groundwater recharge along with rainfall is calculated for the years 1990 and 2008. The percentage variation of groundwater recharge and rainfall with respect to the base period of 1990 is calculated and plotted in Figure 10a, b. It is evident from the figures that there is a considerable decrease in the precipitation and groundwater recharge for all the regions in the Sina basin. However, a maximum of 20% improvement in groundwater recharge with respect to 1990 is seen in Supa, Jamkhed and Alni locations of the study area.

Relation between Groundwater Recharge and Groundwater Level

The higher is the water level, the greater is the amount of storage water in the aquifer. Thus, a proportionality exists between the groundwater recharge and groundwater level. As the groundwater table declines, the groundwater recharge reduces. Furthermore, a reverse relationship is seen with the groundwater level (below the ground) with that of the recharge. The observed groundwater level and the estimated groundwater recharge plot firms this assumption (shown in Figure 11). The expression is given below:

$$R = 300e^{-0.23d} \tag{6}$$

Where R is the groundwater recharge (mm), d is the groundwater level (bgl) in m.

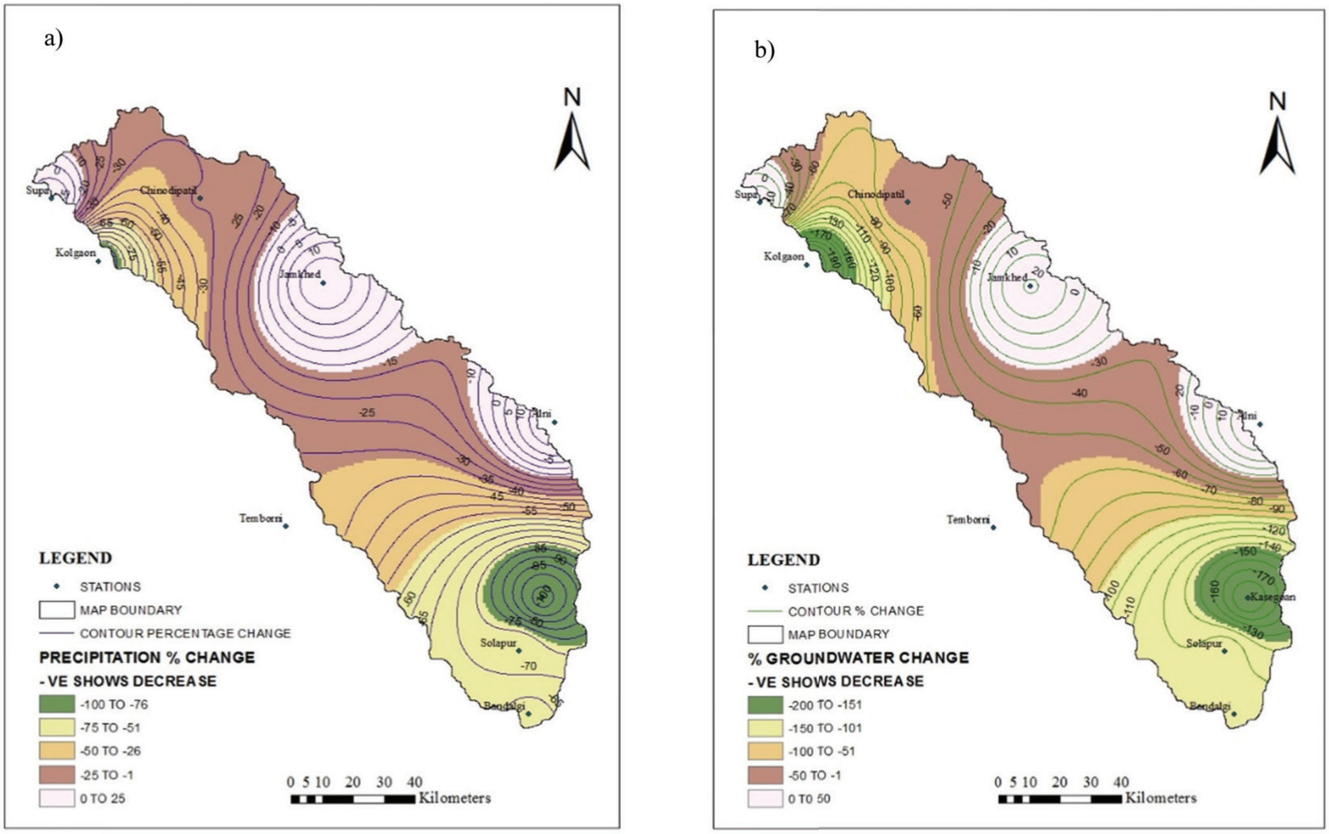

Figure 10: Percentage variation in (a) precipitation and (b) groundwater recharge for year 2008 with respect to base period 1990.

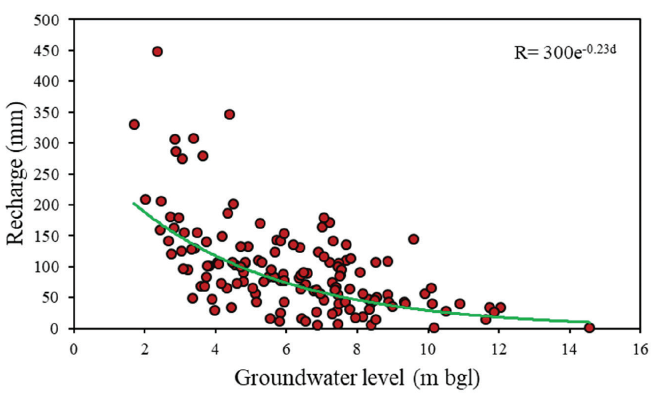

Figure 11: Relationship between groundwater recharge and groundwater level for data in the Sina basin.

Hence, the trend line equation obtained for different polygons is used to predict groundwater recharge for different water table depth and getting a zone of maximum recharge.

A study by Gemitzi et al. (2017) has put forward an empirical groundwater recharge equation which depends on weather, geological and hydrological properties. These same authors have developed an empirical groundwater recharge equation based on Regression analysis. The recharge equation uses SWAT AET values which is given as:

$$GR = 0.4999(P - \mathrm{AET}_{\mathrm{SWAT}}) + 0.041 \quad (7)$$

Whereas, the monthly groundwater recharge equation based on MODIS AET data is

$$GR = 0.5174(P - \mathrm{AET}_{\mathrm{MODIS}}) + 0.2145 \quad (8)$$

where GR corresponds to monthly groundwater recharge (mm), P is monthly precipitation (mm) and AETSWAT and AETMODIS are monthly actual evapotranspiration values (mm).

However, the validation of these models is done by comparing it with the locally observed groundwater fluctuation data obtained from the state and central agencies. The results can also be validated by studying the water inflow and outflow from the water balance approach.

Conclusions

A spatial analysis of different wells of the Sina basin was performed for the assessment of groundwater levels. Spatial data of groundwater levels were put to cluster analysis for the determination of the best suitable method for the determination of representative well for each cluster. Groundwater recharge for each Thiessen polygons was calculated using three different numerical models: Amritsar Model, Chaturvedi Model and ERAS Model.

For the spatial analysis of groundwater levels, four clustering techniques: K-means, hierarchical, expectation maximisation (EM) and canopy were selected and the clustering algorithms were executed for five clusters. By statistical analysis and graphical indicators, the canopy technique was found to be the most suitable form of clustering the formation of representative well for each cluster. Thiessen polygons were plotted for the nine rain gauge stations in the study area. Groundwater recharge assessment was done using Amritsar Model, Chaturvedi Model and ERAS Model for nine regions. The result shows that Chaturvedi Model and ERAS Model share the credit of similar recharge values in the selected basin. It was found that about 15% of the annual precipitation contributes to the annual recharge in the Sina basin. Later, ERAS model results were used for the spatio-temporal analysis of groundwater recharge. Groundwater potential zones were delineated by a combined view of visual comparison, contouring the flow direction of groundwater levels and topography of the basin. Spatio-temporal variation of groundwater recharge and precipitation for the years 1990 and 2008 were compared and plotted. The results indicate a decline in groundwater recharge and precipitation in most of the regions of the Sina basin compared to the base period year 1990. Hence, proper management measures should be taken to conserve waters by limiting the excessive pumping of groundwater. Finally, an empirical relation is developed for finding groundwater recharge in terms of groundwater level (measured below ground level). This information traces the local resources and helps to understand only the regional water system.

Acknowledgement and Conflicts of Interest

The authors would like to thank IMD Pune and Centre for Groundwater Board (CGWB) for providing data to complete the analysis and extends thanks to the editor and reviewers' comments in improving the quality of the paper. The authors also declare no conflict of interest in this article.

References

Arthur, D. and S. Vassilvitskii (2007). K-means^{++}: The advantages of careful seeding. *In:* 2007 ACM-SIAM symposium on discrete algorithms (SODA'07).

Bartoletti, N., Casagli, F., Marsili-Libelli, S., Nardi, A. and L. Palandri (2018). Data-driven rainfall/runoff modelling based on a neuro-fuzzy inference system. *Environmental Modelling & Software*, **106:** 35-47.

Boughariou, E., Saidi, S., Barkaoui, A.E., Khanfir, H., Zarehloul, Y. and S. Bouri (2014). Mapping recharge potential zones and natural recharge calculation: Study case in Sfax region. *Arabian Journal of Geosciences*, **8:** 5203-5221.

Bouwer, H. (2002). Artificial recharge of groundwater: Hydrogeology and engineering. *Hydrogeology Journal*, **10:** 121-142.

Chaturvedi, R.S. (1973). A note on the investigation of ground water resources in western districts of Uttar Pradesh. *In:* Annual Report. U.P. Irrigation Research Institute, 86-122.

Gemitzi, A., Ajami, H. and H.H. Richnow (2017). Developing empirical monthly groundwater recharge equations based on modeling and remote sensing data – Modeling future groundwater recharge to predict potential climate change impacts. *Journal of Hydrology*, **546:** 1-13.

Hall, M., Frank, E., Holmes, G.P., Fahringer, B., Reutemann, P. and I.H. Witten (2009). The WEKA Data Mining Software: An Update. SIGKDD Explorations 11.

Han, J. and M. Kamber (2006). Data Mining Concepts and Techniques. Morgan Kaufmann Publishers.

Krishnamurthy, J., Venkatesesa, Kumar, N., Jayraman, V. and M. Manivel (1996). An approach to demarcate ground water potential zones through remote sensing and a geographical information system. *International Journal of Remote Sensing*, **7:** 1867-1884.

McCallum, A., Nigam, K. and L.H. Ungar (2000). Efficient clustering of high dimensional data sets with application to reference matching. *In:* Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining ACM-SIAM Symposium on Discrete Algorithms. 169-178.

Murthy, K.S.R. (2000). Groundwater potential in a semi-arid region of Andhra Pradesh—A geographical information system approach. *International Journal of Remote Sensing*, **21:** 1867-1884.

Niraula, R., Meixner, T., Ajami, H., Rodell, M., Gochis, D. and C.L. Castro (2017). Comparing potential recharge estimates from three Land Surface Models across the western US. *Journal of Hydrology*, **545:** 410-423.

Roshni, T., Jha, M.K., Deo, R.C. and A. Vandana (2019). Development and evaluation of hybrid artificial neural network architectures for modeling spatio-temporal groundwater fluctuations in a complex aquifer system. *Water Resources Management*, **33:** 2381-2397.

Sahoo, S. and M.K. Jha (2016). Pattern recognition in lithology classification: Modeling using neural networks, self-organizing maps and genetic algorithms. *Hydrogeology Journal*, **25:** 311-330.

Sehgal, S.R. (1973). Ground Water Resources of Punjab State – A Recent Study. Third Annual Research Session. Central Board of Irrigation and Power, New Delhi.

Selvam, S., Magesh, N.S., Chidambaram, S., Rajamanickam, M. and M.C. Sashikkuma (2014). A GIS-based identification of groundwater recharge potential zones using RS and IF technique: A case study in Ottapidaram taluk, Tuticorin district, Tamil Nadu. *Environmental Earth Science*, **73:** 3785-3799.

Wable, P.S., Jha, M.K. and S. Murasingh (2017). Evaluation of groundwater resources for sustainable groundwater development in a semiarid river basin of India. *Environmental Earth Science*, **76:** 601. DOI 10.1007/s12665-017-6912-2

Written, I.H. and E. Frank (2005). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann Publication.