Geostatistical Analysis of Groundwater Data

Reshma H S, P N Chandramouli

***Abstract**: Water resources are stressed because of the country's increasing population and increased water requirements. Even though a good understanding of both surface and groundwater hydrological systems make it possible to manage these resources properly.*

To study the main characteristics of formation of clusters of groundwater levels, statistical analysis has been used. Geostatistics is a class of statistics used to analyze and predict the values associated with spatial or spatiotemporal phenomena. It incorporates the spatial (and in some cases temporal) coordinates of the data within the analyses.

The Statistical analysis is applied to monthly groundwater levels fluctuation data over a period of 2004-2017 in Mysuru, Mandya, Chamarajanagara and Hassan districts of Southern Karnataka in India. The groundwater levels data is collected from 197 Observation Wells from the districts.

The Statistical methods like K-Means Clustering and Agglomerative Hierarchical Cluster Analysis is used to perform the datasets. Grouping is made using AHC method, during this process results are obtained by graph called Dendrogram. The obtained results are compared with the LULC maps of all 4 districts. Different grouping (cluster) is made for groundwater level fluctuations for proper conclusion to arrive.

***Keywords:** Agglomerative hierarchical clustering, K-means cluster, Statistical method.*

I. INTRODUCTION

Groundwater is the most favored source of water in different user segments in India on record of its close all universal accessibility, dependability and low capital expense. The expanding reliance on ground water as a solid source of water has brought about aimless extraction in different pieces of the nation without due respect to the energizing limits of aquifers and other natural elements. Then again, there are territories in the nation, where ground water advancement is imperfect disregarding the accessibility of adequate assets, and channel order regions experiencing issues of water logging and soil saltiness because of the steady ascent in ground water levels. Groundwater levels are observed to be straightly subject to the effective amount of accumulated rainfall. An impact of climatic change, there will be increasing evapotranspiration and decreasing in rainfall which will resulting in the depletion of groundwater levels, due to this some wells become dry and also other wells become less production due to loss of available drawdown. The effect of climatic changes on groundwater levels can be analyzed and estimated by the relationship between the groundwater level fluctuations and historical climate records.

To investigate the groundwater system several techniques are used, here Statistical method and Modelling methods have been reported to study the sensitivity of the groundwater system and also to understand the cluster analysis methods applied to the groundwater level.

1.1 OBJECTIVES:

- The main objective of this work is to define and apply a framework to study the variation in the groundwater levels over a long period of time (12 years) and large area (Four districts of Southern Karnataka). Using Cluster Analysis to determine the true groups of data according to their similarities to each other.
- To obtain difference in variation of groundwater in terms of similar groups from Cluster Analysis.
- Understand the fluctuations in the groundwater levels behaviour of the groundwater system by comparing the results obtained from the Cluster Analysis.

II. STUDYAREA & DATA SETS

The present study attempts to develop a cluster analysis of groundwater level fluctuations for four districts taken separately and all the districts combined together (i.e. Mandya, Mysore, Chamarajanagara, Hassan) of Southern Karnataka in India with a clear focus on availability of sparse datasets.

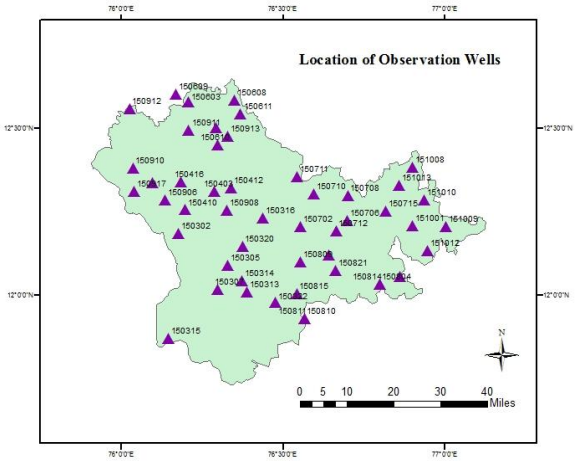

Figure-1 Location of Observation Wells in Mysuru District

Revised Manuscript Received on October 15, 2019.

Reshma H S, M.tech in Hydraulics, Department of Civil Engineering, National Institute of Engineering, Mysore, India

P. N. Chandramouli, Professor, Department of Civil Engineering, National Institute of Engineering, Mysore, India.

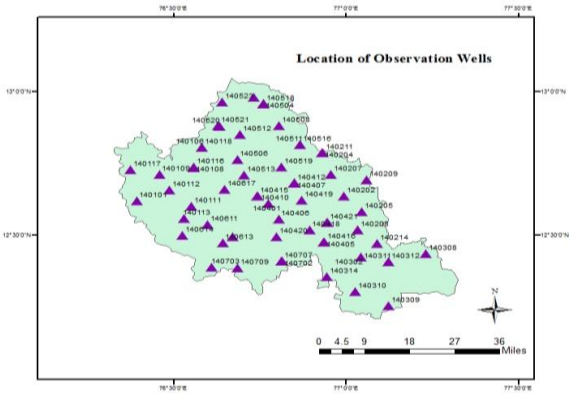

Figure-2 Location of Observation Wells in Mandya District

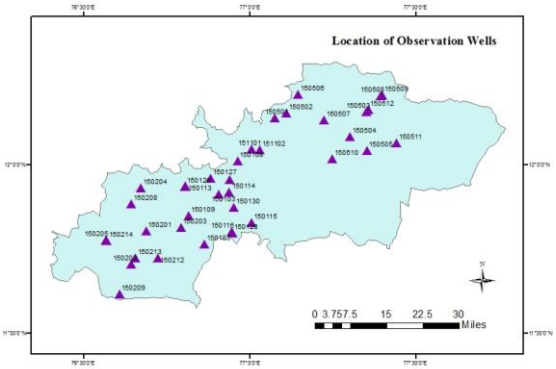

Figure-3 Location of Observation Wells in Chamarajanagara District

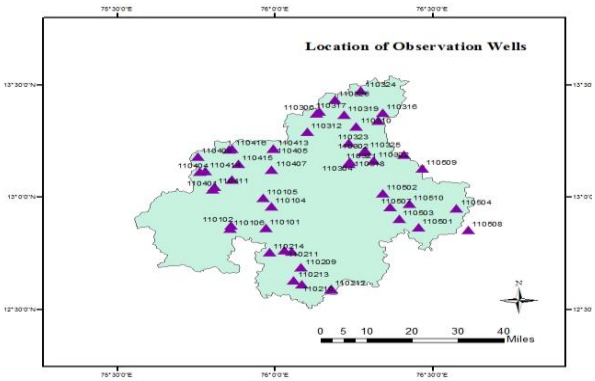

Figure-4 Location of Observation Wells in Hassan District

III. PROCEDURE FOR PAPER SUBMISSION

3.1 Statistical Analysis

Statistical analysis is used to carry out the interpretation of the data. It is one of the numerical measures which describe the characteristics of a sample and a single numbered output can be displayed to study the nature of a data set. All mathematical and statistical calculations were influenced utilizing to Excel 2007 (Microsoft office) and XLSTAT.

3.2 K-MEANS CLUSTERING

K-means clustering used for distance measure, the distance measure will determine the similarity between two elements and it will influence the shape of the clusters.

K- Means clustering method is not suitable for the sample size under 400. Only it suits when sample size is exceeding 1000, hence agglomerative hierarchical clustering method is adopted for the process of clustering the GWLs series.

3.3 AGGLOMERATIVE HIERARCHICAL CLUSTERING

In statistics, hierarchical clustering (also known as hierarchical cluster analysis or HCA) is a cluster analysis method that seeks to construct a cluster hierarchy. There are generally two types of hierarchical clustering strategies

Agglomerative: This is a "bottom-up" approach: each observation starts in its own cluster and as one moves up the hierarchy, pairs of clusters are merged.

Divisive: This is a "top-down" approach: all observations begin in one cluster and splits occur recursively as one moves the hierarchy downwards.

The fusions and splits are generally determined in a greedy way. Hierarchical clustering results are usually shown in a Dendrogram.

3.3.1 Process of AHC method

- The process begins by calculating the N object dissimilarity.
- Then two objects are clustered together to minimize a given criterion of agglomeration, thus creating a class consisting of these two objects.
- Then the dissimilarity is calculated using the agglomeration criterion between this class and the other N-2 objects. The two objects or classes of objects are then clustered together, the clustering of which minimizes the agglomeration criterion.

The Figure-5 (Dendrogram) represents formation of number of clusters. The complete dendrogram shows the progressive object clustering. If truncation is requested, a broken line will mark the level of truncation. After truncation, the truncated dendrogram shows the classes.

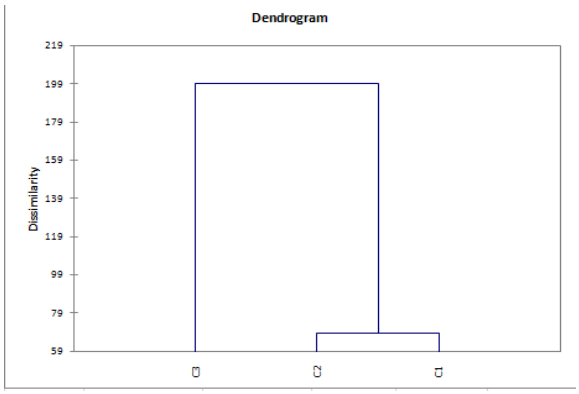

Figure-5 cluster indicating graph

3.4 HOW CLUSTER IS FORM?

Cluster is formed based on the centroid method. Automatically centre point is formed in between the number of points in the group/cluster. Where the point is nearest to the centroid point, formed a cluster. This process is continued throughout the steps until the required number of cluster is formed. (For example: K= 1, 2, 3...).

IV. RESULTS AND DISCUSSION

All the necessary analysis for Cluster Analysis in the groundwater system is carried out on the basis of the methodology and input requirements. The analysis is performed by Statistical Analysis of the groundwater level data series. For four districts taken separately and all districts combined Mandya, Mysuru, Chamarajanagara & Hassan (i.e. MMCH), the results and discussions will be presented. This chapter presents all the results and discussions necessary for the research objective.

The monthly GWLs data applicable to the 60 observation wells for Mandya district, 49 observation wells for Mysuru district, 37 observation wells for Chamarajanagara district and 50 observation wells for Hassan district totalling to 196 groundwater level observation wells. The period considered for data from the GWLs series is 12 years, i.e. January 2004 to January 2017.

In Agglomerative Hierarchical Clustering (AHC) method analysis, dendrogram graph is appeared for each month of the year (i.e. 12 graphs). By this graph number of cluster formation is obtained. And also by the result appeared in the last table called result by object displays the each observation well belongs to which cluster (i.e. 1, 2 or 3). Here 1, 2 and 3 indicates High, Medium & Low. That can be form groups.

4.1 RESULTS OBTAINED DURING THE AHC METHOD IN THE FORM OF DENDROGRAM GRAPH

The below graphs represents the formation of number of clusters (i.e. 3). And also represents the Observation Well ID's belongs to which group.

Graph 1: Month of January

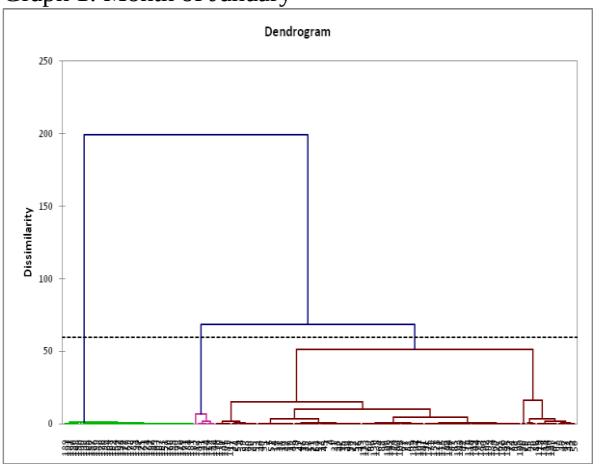

Figure-6

Graph 2: Month of February

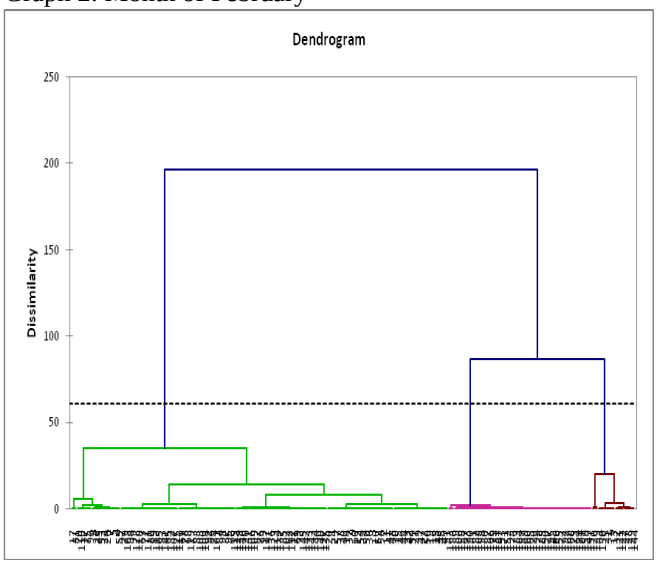

Figure-7

Graph 3: Month of March

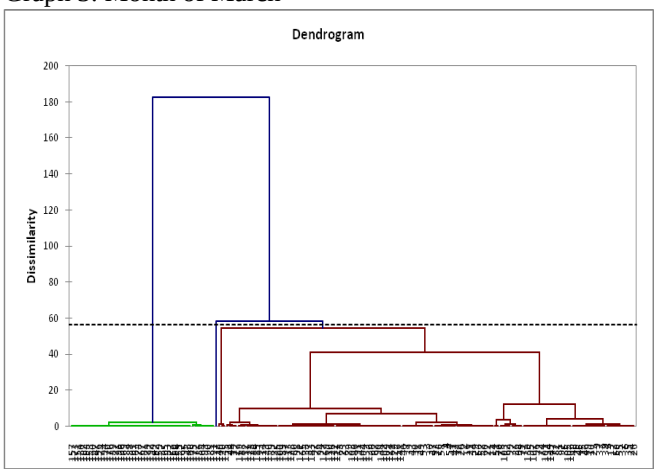

Figure-8

Graph 4: Month of April

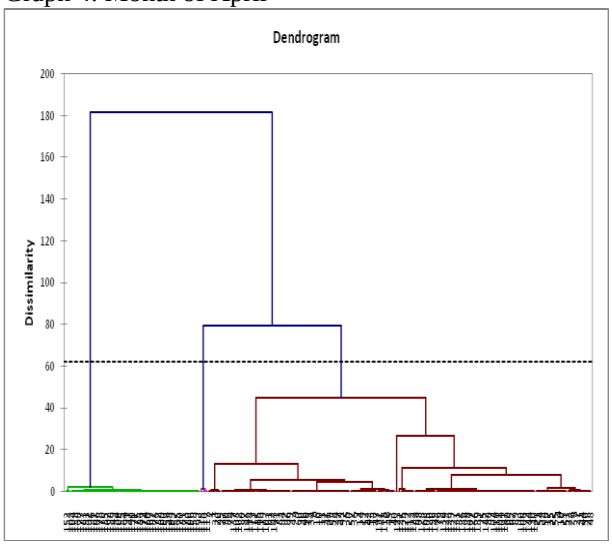

Figure-9

Graph 5: Month of May

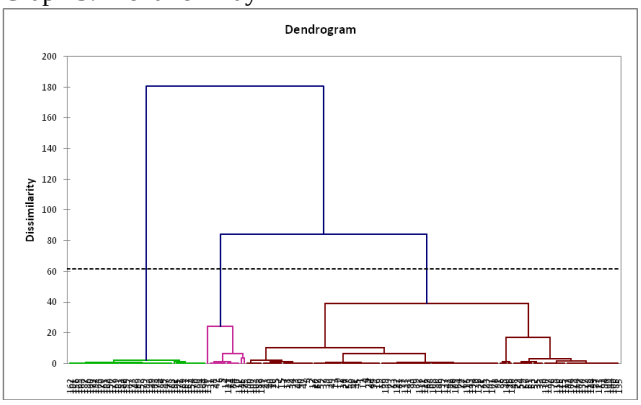

Figure-10

Graph 6: Month of June

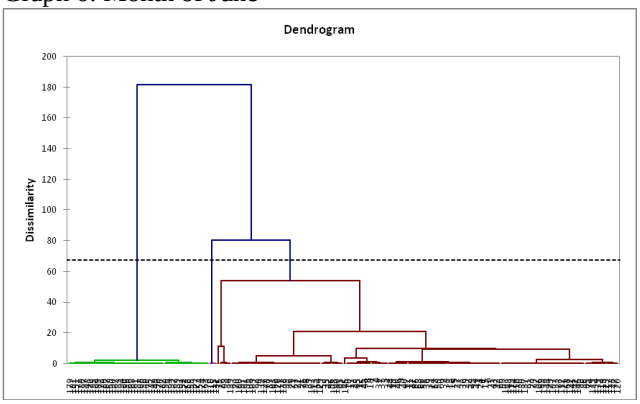

Figure-11

Graph 7: Month of July

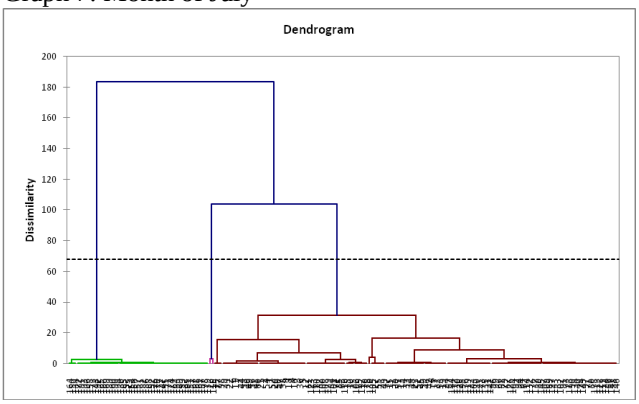

Figure-12

Graph 8: Month of August

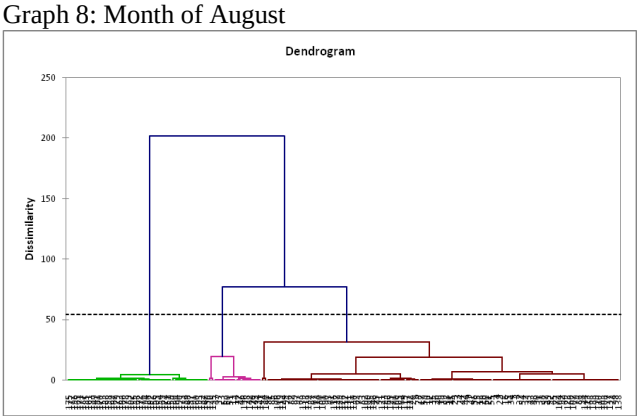

Figure-13

Graph 9: Month of September

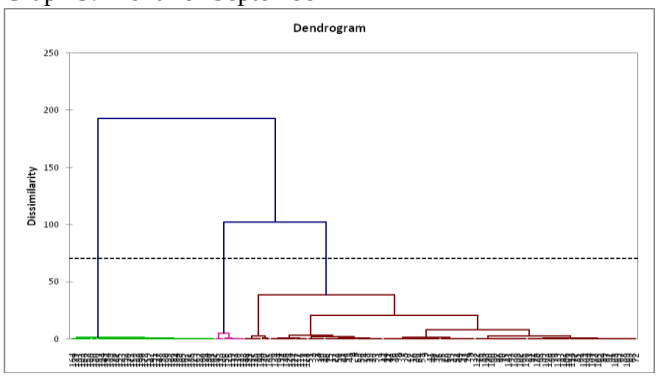

Figure-14

Graph 10: Month of October

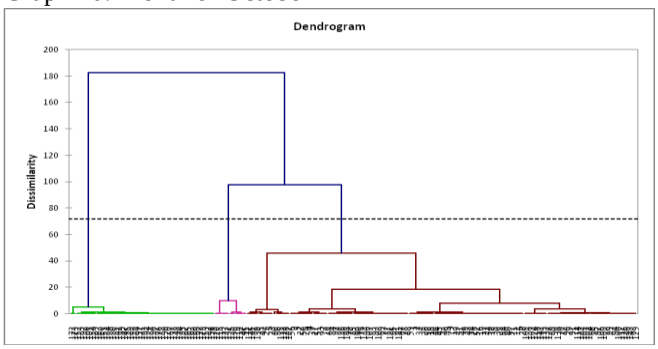

Figure-15

Graph 11: Month of November

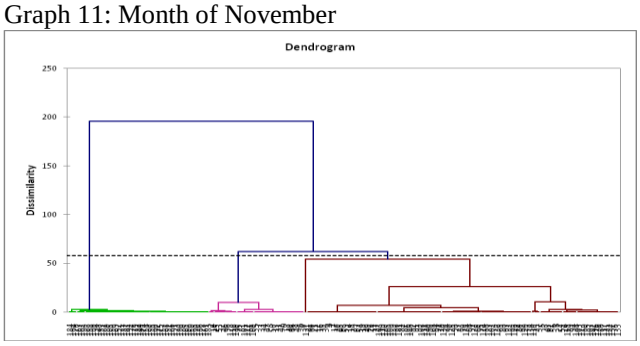

Figure-16

Graph 12: Month of December

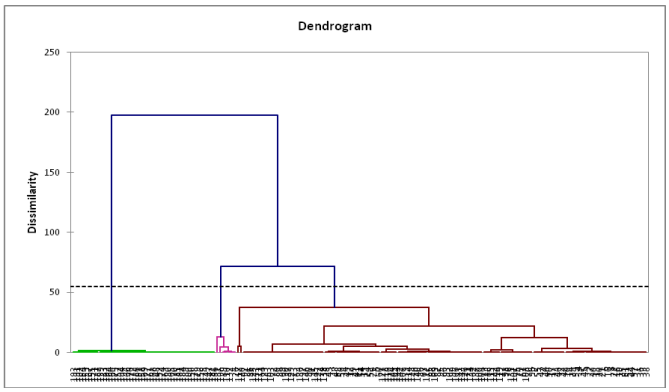

Figure-17

- The above all graphs (**figure-6 to 17**) formed 3 clusters and obtained results are compared with the LULC map, to identify the groundwater level is Low, Average and High, and to discuss the reason for the fluctuations in groundwater levels.

4.2 COMPARISION OF CLUSTER WITH LULC MAP

- The formed clusters are plotted on the LULC map. Clusters are represents in different colors (dots) on map are: Yellow- High, Red- Low and Black- Average.
- Legends: LULC map

1- Water, **2**-Wetlands-mixed, **3**- Snow, **5**-Range-Brush, **6**- Urban/Rural area, **7**- Orchard , **8**- Agricultural land-generic, **10**- Irrigated, surface water double crop, **11**- Irrigated, surface water continues crop, **13**- Irrigated, groundwater, double crop, **14**- Irrigated, conjunctive use, single crop, **16**- Irrigated, conjunctive use, double crop.

4.2.1 Description for the Following Cases

Formation of clusters obtained by the AHC analysis of GWLs (2004-2017) of 197 observation wells for all 4 districts (MMCH) is given in below figures. It is observed from the figures, 197 observation wells are made 3 groups. Yellow and Black colour represents High and Average in GWLs, because the observation wells are dugged/ lies in Wetlands-Mixed, Nearby Water Bodies, Snow, Irrigated Conjunctive Double Crop, Range Brush, Irrigated Surface Water Double Crop, Irrigated surface water double crop, Irrigated GW double crop. But GWL may vary seasonally or if proper recharge is not done or it may be depends on the rainfall also. Red colour represents Low in GWL because; the observation wells are lies in Urban/ Rural areas.

Case 1: Month of January

In the month of January are few observation wells are low in GWLs (Red), remaining are sufficient in GWLs (Black and Red). Reason is mentioned above paragraph.

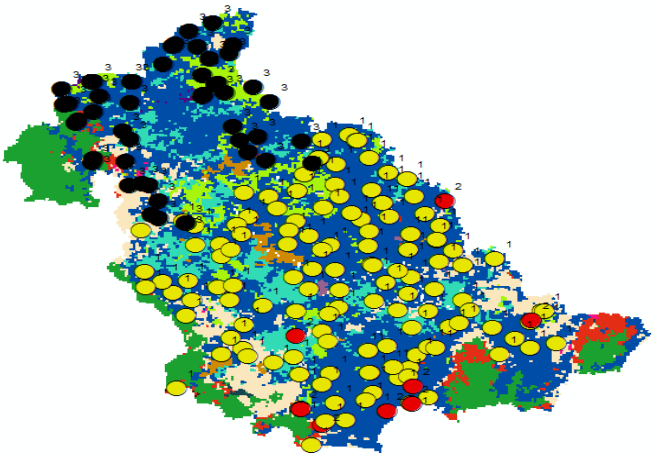

Figure-18

Representation of no. of clusters on LULC map of 4 districts for the month of January

Case 2: Month of February

In the month of February, sufficient GWL is available for some observation wells (black and yellow), but in this case red color observation wells are distributed, because of depletion of GWLs, it may due to dry land, improper recharge, low rainfall, or it may be depends on the no. of users.

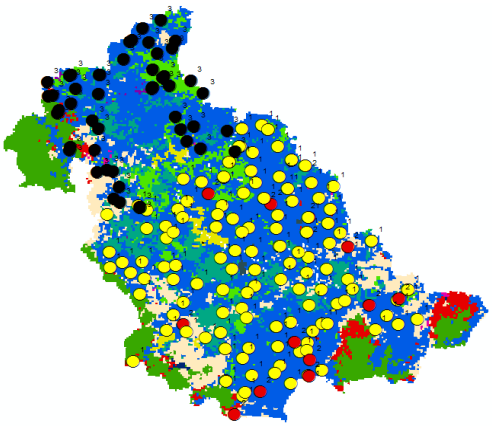

Figure-19

Representation of no. of clusters on LULC map of 4 districts for the month of February

Case 3: Month of March

In the case of March, it can observed that only one observation wells are shown in red color that means it's very low in GWLs. Because it lies near by urban / rural areas. And also March is summer season.

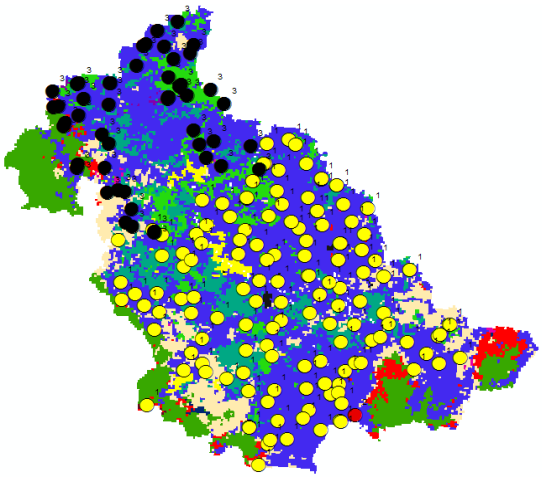

Figure-20

Representation of no. of clusters on LULC map of 4 districts for the month of March

Case 4: Month of April

In the case of April, only few observation wells are low in GWLs usually in urban areas, because it is summer season. Remaining observation wells are sufficient in GWLs because that are surrounded by wetlands, irrigated lands.

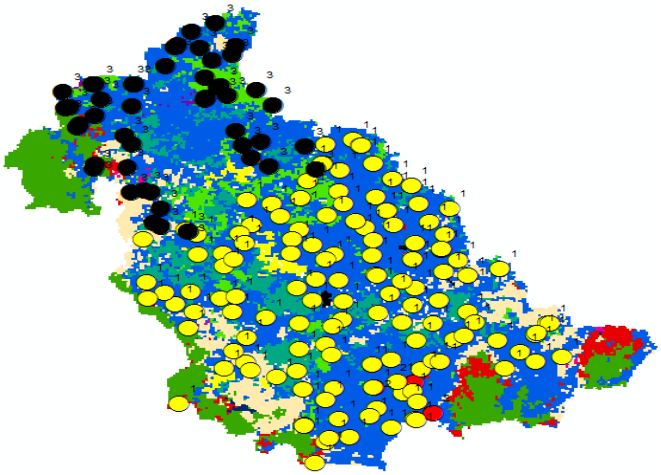

Figure-20

Representation of no. of clusters on LULC map of 4 districts for the month of April

Case 5: Month of May

In the case of May, due to summer season in some areas GWLs are get depleted. And it is indicated by red color.

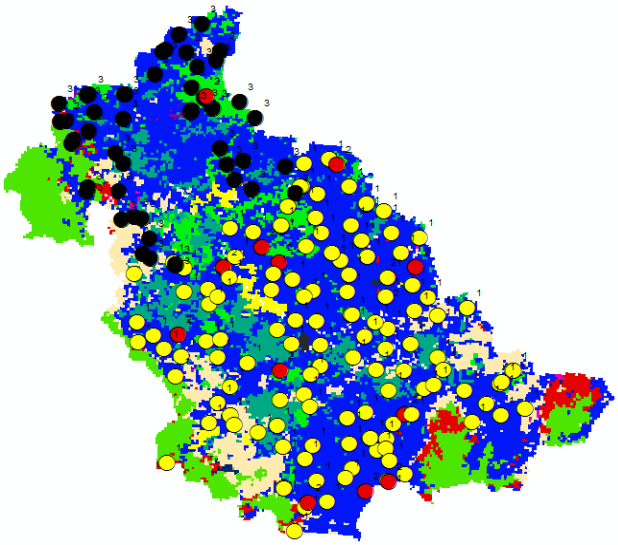

Figure-21

Representation of no. of clusters on LULC map of 4 districts for the month of May

Case 6: Month of June

In the case of June, usually all the observation wells are sufficient in GWLs but it may vary in water levels. The water levels are in rich, because due to monsoon rainy season. Hence, low level observation wells are less in number. The above mentioned same reason is carried out upto the month of September. Because from the month of June to September monsoon rainy season.

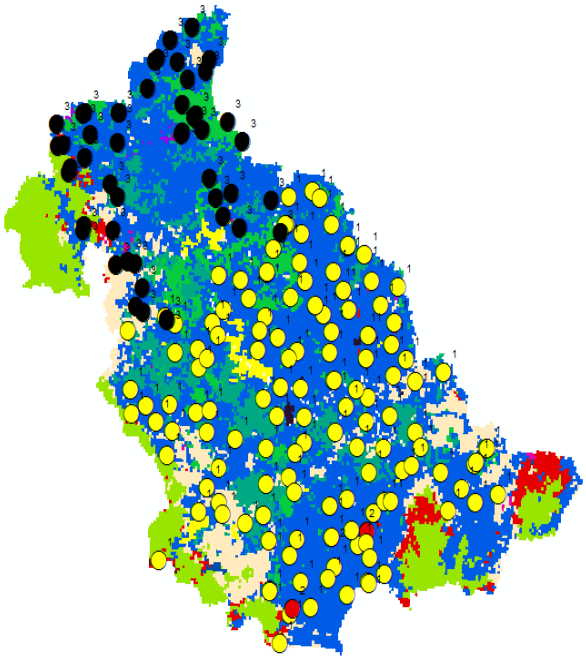

Figure-22

Representation of no. of clusters on LULC map of 4 districts for the month of June

Case 7: Month of July

Here also low GWLs observation wells are only one in number. Because it lies in Range-Brush.

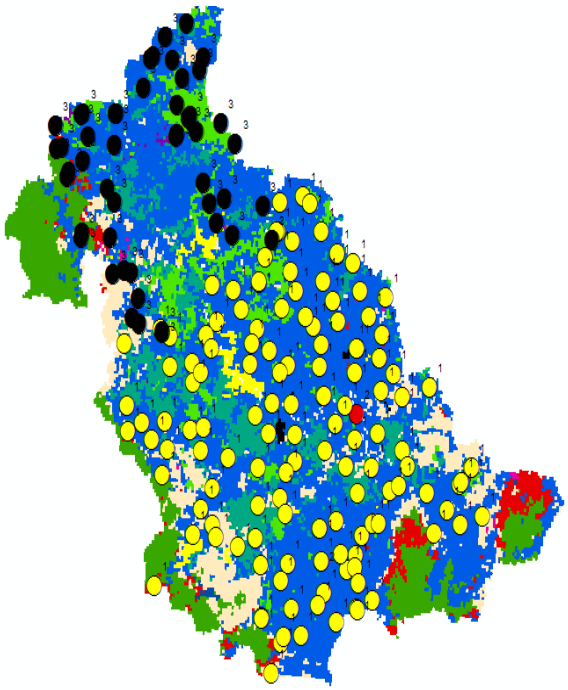

Figure-23

Representation of no. of clusters on LULC map of 4 districts for the month of July

Case 8: Month of August

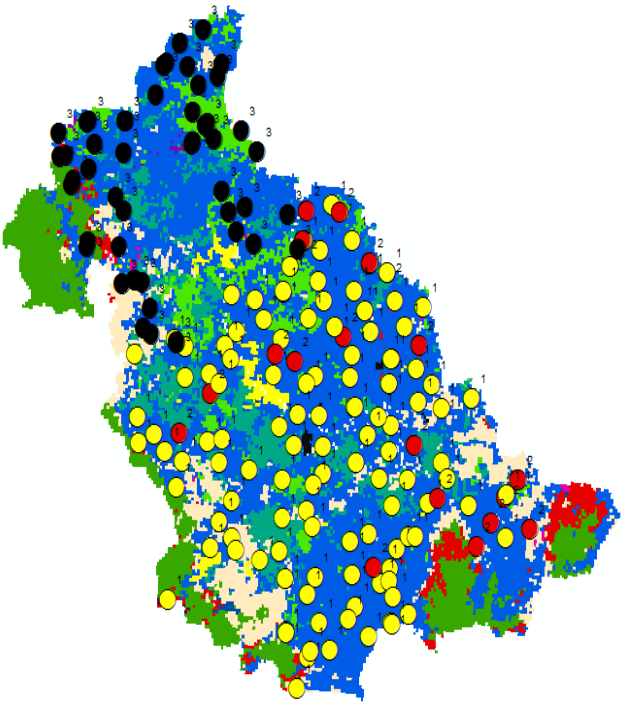

Figure-24

Representation of no. of clusters on LULC map of 4 districts for the month of August

Case 9: Month of September

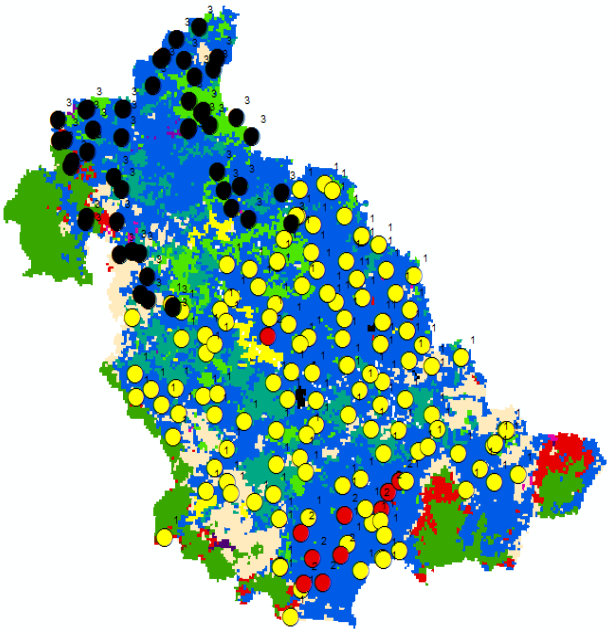

Figure-25

Representation of no. of clusters on LULC map of 4 districts for the month of September

Case 10: Month of October

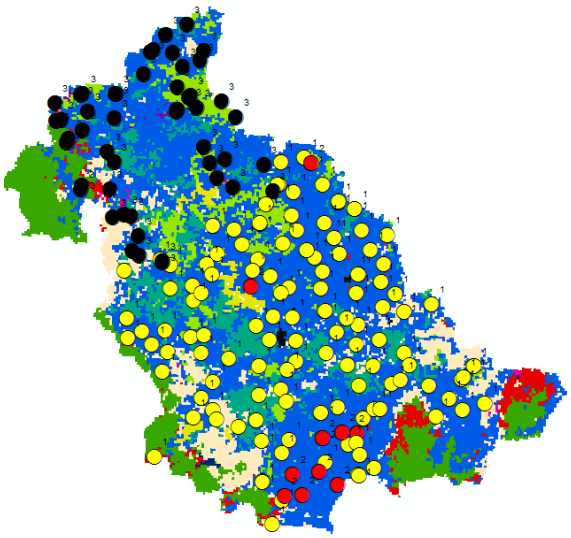

Figure-26

Representation of no. of clusters on LULC map of 4 districts for the month of October

Case 11: Month of November

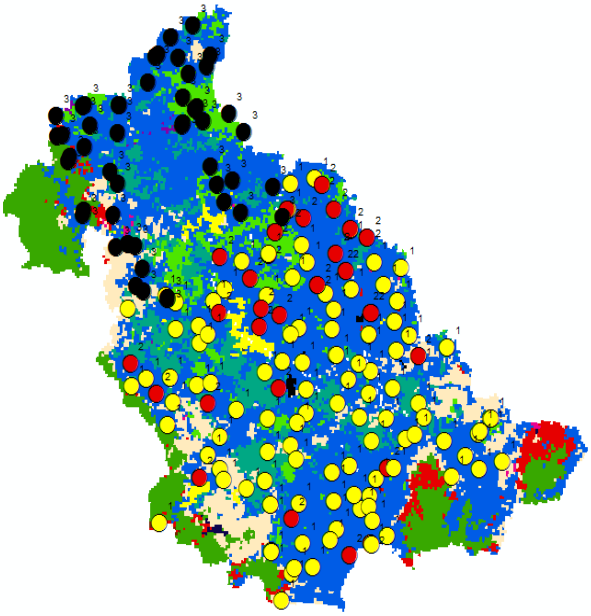

Figure-27

Representation of no. of clusters on LULC map of 4 districts for the month of November

Case 12: Month of December

In the case of December, in this month there would be a winter season. Here GWLs in low level are more, because of less rainfall, or it may be less agricultural activity.

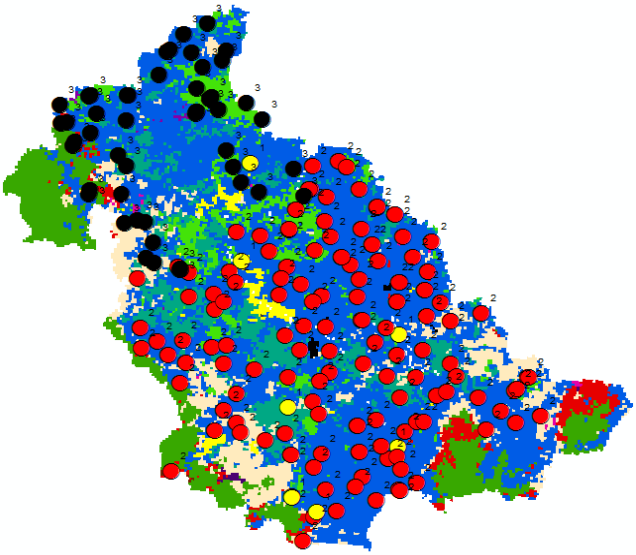

Figure-28
Representation of no. of clusters on LULC map of 4 districts for the month of December

V. CONCLUSION

In this study, Statistical Analysis is carried out for groundwater level series data sets in Mysuru, Mandya, Chamarajanagara and Hassan.

Generally, from this work, following cases are observed:

- High of GWLs
- Medium of GWLs
- Low of GWLs

The rising of GWLs may be due to the recharge exceeding the discharge over the years, and the falling of GWLs shows that discharges are more than recharges. The GWL fluctuations may also depends on the agricultural activities, dryness of water bodies, rainfall etc.

From this work, it is known that rising and falling of groundwater depends on the Landuse Landcover map, agricultural activities and also depends on the climatic season. By using Statistical methods clusters can be made for any type of data sets. When compare to K-Means clustering method, AHC method is best method when the sample size is less in numbers (i.e. under 300 to 400).

REFERENCES

1. Akira Kawamura ,Shigeyuki Ishihara., 2017. Clustering Fluctuation Patterns of Groundwater Levels in Tokyo Caused by the Great East Japan Earthquake Using Self-Organizing Maps: International Water Resource Association.
2. Belkhiri L, Boudoukha, and Mouni L., 2011. A multivariate Statistical Analysis of Groundwater Chemistry Data: Int. J. Environ. Res., 5(2):537-544.
3. Deeptha Thattai, Sathyanathan rangarajan., et al., 2017. Geostatistical analysis of groundwater parameters in Kerala State, India: jr. Of industrial pollution control 33(s3) (2017) pp 1496-1499
4. Gurhan Yalcin, M ,Ali Tumuklu., et al., 2009. Application of multivariate statistical approach to identify heavy metal sources in bottom soil of the Seyhan River (Adana), Turkey: Environ Monit Assess (2010) 164:311–322.
5. Ji-Hoon Kim,1 Rak-Hyeon Kim., et al., 2005. Multivariate statistical analysis to identify the major factors governing groundwater quality in the coastal area of Kimje, South Korea: *Hydrol. Process.* 19, 1261–1276.
6. Kouping Chen, Jiu J. Jiao., et al., 2007. Multivariate statistical evaluation of trace elements in Groundwater in a coastal area in Shenzhen, China: Environmental Pollution 147, 771-780.
7. Matthias Templ, Peter Filzmoser , Clemens Reimann., 2008. Cluster analysis applied to regional geochemical data: Problems and possibilities: Applied Geochemistry 23 (2008) 2198–2213.
8. Ranjan Sarukkalige., 2012. Geostatistical Analysis of Groundwater Quality in Western Australia: a review source: An International Journal (ESTIJ), ISSN: 2250-3498, Vol.2, No. 4.
9. Sana'a Odat., 2015. Cluster and Factor Analysis of Groundwater in Mafraq Area, Jordan., Current World Environment: Vol. 10(2), 422-431 (2015).
10. Sajil kumar, P J, Jegathambal, P, James, E J, 2011. Water Institute, Karunya University, Coimbatore. Multivariate and Geostatistical Analysis of Groundwater Quality in Palar River Basin: a review source: International journal of geology, (issue 4, volume 5).
11. Salah, Hamad., 2009. Geostatistical analysis of groundwater levels in the south Al Jabal Al Akhdar area using GIS. Water resource department, General Water Authority, 32718, Benghazi, Libya, 25. - 28. 1. (2009)
12. Samson S and Elangovan k., 2017. multivariate Statistical Analysis to assess Groundwater quality in Namakkal district, Tamil Nadu, India.
13. Sandow Mark Yidana., 2010. Groundwater classification using multivariate statistical methods: Southern Ghana: Journal of African Earth Sciences 57, 455–469.
14. Vivienne H. McNeil, Malcolm E. Cox, Micaela Preda., 2005. Assessment of chemical water types and their spatial variation using multi-stage cluster analysis, Queensland, Australia: Journal of Hydrology 310 (2005) 181–200.
15. Yurdanur Unal, Tayfun Kindap and Mehmet Karaca., 2003. Redefining the Climate Zones of Turkey Using Cluster Analysis: International Journal of Climatology. **23**: 1045–1055.
16. Yuxiao Jianga,B , Huaming Guo.,et al., 2015. Principal component analysis and hierarchical cluster analyses ofarsenic groundwater geochemistry in the Hetao basin, Inner Mongolia: Chemie der Erde 75 , 197–205.

AUTHORS PROFILE

Reshma H.S, completed Bachelor Of Engineering in the stream of Civil Engineering from G.Madegowda Institute of Technology, Mandya and Masters in the field of Hydrualics and Water Resource Engineering from National Institute of Engineering, Mysuru Karnataka, 570 008, India
reshmahs1995@gmail.com.

Dr. P.N.Chandramouli is Professor in the Department of Civil Engineering and Head, Centre for Research & Consultancy (CRC). He has received his Ph.D from IIT-Roorkee, M.E. from IISc-Bangalore and B.E. from University of Mysore. His teaching and research interests are in the field of Hydraulics and Water Resource Engineering.Karnataka, 570008, India.
chandramouli1963@yahoo.co.in