Proceedings of the 2022 Pre-ICIS SIGDSA Symposium

Special Interest Group on Decision Support and Analytics (SIGDSA)

12-12-2022

Optimizing the Frequency of Data Collection through Citizen Science - An Application of Analysis of Variance and Data Visualization

Alexandra Mroz

Lakshmi Iyer

Basant Maheshwari

Akramul Bhuiyan

Optimizing the Frequency of Data Collection through Citizen Science - An Application of Analysis of Variance and Data Visualization

An Extended Abstract

Alexandra Mroz
Appalachian State University
mrozaw@appstate.edu
Basant Maheshwari
Western Sydney University
B.Maheshwari@westernsydney.edu.au

Lakshmi S. Iyer
Appalachian State University
iyerLs@appstate.edu
Akramul Bhuiyan
Appalachian State University
bhuiyanm@appstate.edu

Introduction

India is the largest user of groundwater in the world with an estimated usage of 230 km3 per year (World Bank, 2010; Saha et al., 2018; Goswami and Ghosal, 2022). Globally, areas under groundwater irrigation are the highest in India (39 million ha), followed by China (19 million ha) and the USA (17 million ha), and at present 204 km3 y-1 of groundwater is pumped annually in India (Siebert et al., 2010). Recognizing the problem and the importance of groundwater in rural areas of the country, The Managing Aquifer Recharge and sustaining Groundwater Use through Village-Level Intervention (MARVI) project commenced in 2011 to improve water security for rural communities. The project aims to teach rural communities to become more self-efficient and better manage their precious resources. MARVI Project aims to address the following problems:

1. What are the patterns of groundwater changes in the wells over time, across villages? How are these patterns related to parameters such as rainfall?
2. What is the minimum number of wells that should be targeted for data collection to get a sense of an aquifer's behavior? In other words, what is the minimum number of wells that will give famers a good understanding of the water level fluctuations around their village? How does the water level in their wells change based on rainfall amounts during the monsoon season?
3. How often should well data be collected to get a sense of an aquifer's behavior?

This research in progress addresses the last research question. MARVI project aims to help five villages in the northwestern part of India. In this study, a sample of 10 wells from each village was selected and analyzed to better optimize the frequency of data collection. This study uses ANOVA to analyze the variation in monthly groundwater levels and determine if the data could be collected bi-weekly, instead of continuing to collect data weekly.

Literature Review

There are many factors that influence the fluctuation of groundwater levels. The most important ones include: rainfall, evaporation, temperature and pumping patterns (Yadav et al. 2019). When it comes to groundwater, data collection it serves two main purposes: (Szucs & Jordan, 1994)

1. Find out natural processes and changes that are taking place in the aquifer
2. Detect pollution and serve as an alarm system

To find the optimal sampling frequency it is best to collect the data on a regular basis over an extended period of time. With a sizable sample reading from each aquifer analysts can then move on to study the data and begin drawing conclusions.

Data collection

As mentioned earlier, MARVI project started by collecting data from 250 wells in five different villages. The number of monitored wells was later changed to 30 wells per village totaling 120 wells. The villages that are a part of the project are Varni, Sunderpura, Hinta, Dharta, and Badgaon. The MARVI project used the citizen science approach and trained volunteers, called Bhujal Jankaars (Groundwater Informed), to collect the well data every week. The data is being recorded using a mobile app "MyWell". In the app BJs record the data on the water level in each well, water quality, rainfall amount, and geographic data. The data is later transferred into an Excel spreadsheet and cleaned.

Data analysis

This study seeks to address the question of how often the data should be collected to maintain a good understanding of the aquifers' behavior and make the data collection process less labor-intensive. The analysis of variance (ANOVA) focused on comparing weekly differences of the sample for each month over the span of 10 years. ANOVA was used to determine if the variance for each month over the years was statistically significant.

	Dharta	Varni	Sunderpura	Badgaon	Hinta
January	0.00711	0.922831	2.24E-11	4.62E-13	0.000356
February	0.140747	0.049555	0.000007	3.72E-10	0.449142
March	0.177929	0.170356	0.08031	0.000009	0.140502
April	0.21376	3.76E-07	0.383838	0.000252	0.522487
May	0.190299	4.95E-07	0.003277	0.019701	0.023671
June	0.000964	0.485313	0.485329	0.200951	0.000044
July	0.002032	0.377376	0.000116	0.018528	0.433483
August	1.42E-19	2.58E-23	2.17E-08	4.47E-35	0.000844
September	0.808491	0.006401	0.673147	2.12E-09	0.000003
October	0.004549	0.848997	5.84E-21	2.03E-08	0.008434
November	1.06E-07	2.83E-07	0.01231	3.07E-10	0.530533
December	0.920153	0.002803	0.003118	1.46E-11	0.30005

Table 1. ANOVA results

Table 1 shows the results of the ANOVA test for each month for each village. P-values above 0.05 indicate that the variance between the weekly groundwater changes over the years is not significant. Based on this analysis several months in which the changes in the groundwater levels over the years don't exhibit a lot of fluctuation have been found. The next step of the analysis was to determine for which of those months the average weekly fluctuations in the well water levels don't exceed a certain threshold. For the purpose of this analysis, a threshold of (-)2% was chosen.

The last step of the analysis was to pick the months for which there is little fluctuation in groundwater levels over the years and the average weekly percent change in groundwater levels does not exceed 2% (Tables 2 and 3). The data for those months was then transformed into bi-weekly data by removing every other record, and average bi-weekly percent changes were calculated. This was done to further confirm that the biweekly data collection would yield similar results in terms of understanding and maintaining the aquifers' water levels.

	Dharta	**Varni**	**Sunderpura**	**Badgaon**	**Hinta**
January	-	-1.22	-	-	-
February	-1.47	-	-	-	-
March	-0.90	-1.67	0.06	-	-0.96
April	-0.01	-	-0.60	-	-0.82
May	-0.74	-	-	-	-
June	-	-0.26	-0.11	0.33	-
July	-	1.67	-	-	1.19
August	-	-	-	-	-
September	1.24	-	2.63	-	-
October	-10.62	-2.36	-	-	-
November	-	-	-	-	-3.72
December	-3.09	-	-	-	-2.55

Table 2. Average weekly percent changes in water levels for months with p-value more than 0.05

	Dharta	**Varni**	**Sunderpura**	**Badgaon**	**Hinta**
January	-	-4.47	-	-	-
February	**-1.63**	-	-	-	**-2.00**
March	**-1.66**	-6.73	**0.19**	-	-2.50
April	**0.14**	-	**-1.42**	-	**-1.32**
May	**-1.20**	-	-	-	-
June	-	**0.25**	**-0.28**	**0.32**	-
July	-	9.91	-	-	**1.50**
August	-	-	-	-	-
September	5.22	-	-	-	-
October	-	-	-	-	-
November	-	-	-	-	-
December	-	-	-	-	-

Table 3. Average bi-weekly percent changes in water levels for months with p-value more than 0.05 and average weekly percent changes below 2%

Conclusions

The findings from this study can now indicate in which months the data could be collected less frequently, i.e. biweekly rather than weekly, without compromising the data quality or losing track of the groundwater levels. Although it might have seemed that the results would be similar for all the villages because of their proximity to each other, the results seem to be different for each village. This means it may be necessary to use different measurement frequencies depending upon the local geology and other factors.

References (complete list available upon request)