DEVELOPMENT OF A PYTHON ALGORITHM FOR AUTOMATED DATA ORGANIZATION FROM GROUNDWATER MONITORING WELLS OF THE GUARANI AQUIFER SYSTEM IN BRAZIL

GUSTAVO SCHMIDT CABRAL[1]; PAULA JOSYANE DOS SANTOS FRANCISCO[2]; GUILHERME PAZZAGLIA[2]; JULIANA PERTILLE[3]

[1]Universidade Federal de Pelotas – gsc.geologia@gmail.com
[2]Universidade Federal de Pelotas – paula.josyane.eng@gmail.com / pazzagfx@gmail.com
[3]Universidade Federal de Pelotas – juliana.pertill@gmail.com

1. INTRODUCTION

The efficient management of water resources is a critical issue on a global scale, especially in regions where the availability of groundwater plays a fundamental role in providing drinking water and ensuring environmental sustainability. The Guarani Aquifer System (GAS), located in South America, is one of the largest groundwater reserves in the world, spanning parts of Brazil, Argentina, Paraguay, and Uruguay. In Brazil, this aquifer plays a vital role in supplying water to various communities, industries, and agricultural activities.

Monitoring groundwater is essential to understand variations in water levels and quality over time. To this end, networks of monitoring wells were established throughout Brazil, collecting a massive amount of hydrogeological data. However, analyzing this data often becomes a challenging task due to its heterogeneity and the lack of an efficient system for organization and analysis.

The main challenge of this work is to propose the development of an innovative algorithm that automates the organization of data collected in groundwater monitoring wells in the GAS. The algorithm aims to transform this data into a structured table, facilitating time series analysis.

The Integrated Groundwater Monitoring Network of the Geological Survey of Brazil (SGB-CPRM) is responsible for collecting and maintaining hydrogeological data across the country. However, this data is often available in diverse and non-standardized formats, making it difficult to use it efficiently for long-term analysis and informed decision-making. The lack of an automated system for organizing and standardizing this data is a significant obstacle for researchers, water resource managers and decision makers.

The central problem addressed by this work is the absence of an automated tool that can transform the raw data from the monitoring wells of the GAS into a structured table, which is easily accessible and ready for time series analysis. Currently, manually organizing this data is time-consuming and error-prone, which limits the ability to harness the full potential of available hydrogeological information.

The main goal of this work is to develop an algorithm that allows the automated organization of data from groundwater monitoring wells in the GAS, provided by the SGB-CPRM Integrated Groundwater Monitoring Network, in the format of a structured and standardized table. Specifically, we seek to achieve the following objectives: (1) Collect and integrate raw data from different sources and formats, ensuring data integrity and quality; (2) Develop an algorithm that is capable of structuring the data into a table with relevant fields, such as collection date and water levels; (3) Validate the algorithm through tests and comparisons with manual data organization methods; (4) Demonstrate the usefulness of the

organized table for time series analysis and decision making related to groundwater management.

2. METHODOLOGY

The methodology adopted in this study involved the extraction and organization of data from water level monitoring wells in the GAS, located in Brazil. To this end, a sample of 39 wells was used, and their data were made available by the Integrated Groundwater Monitoring Network (RIMAS), maintained by the Geological Survey of Brazil (SGB-CPRM) (Figure 1).

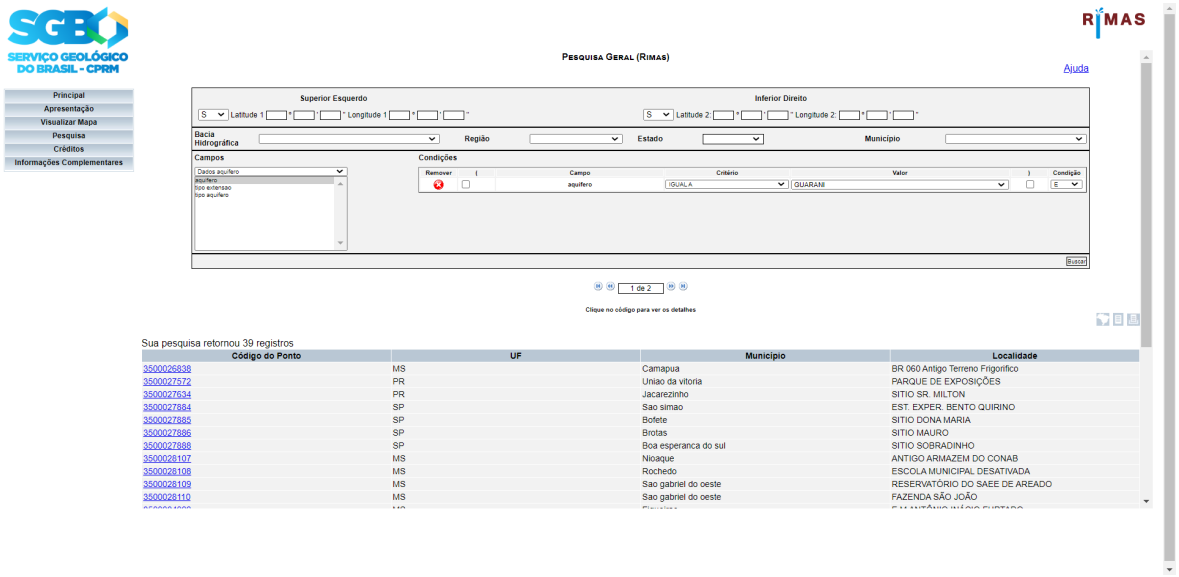

Figure 1. RIMAS interface web.

The data provided by RIMAS were in .csv file format, which posed an initial challenge due to the need to transform this raw data into an organized structure suitable for water level time series analysis. In this regard, a programming-based approach was employed, utilizing Python libraries, including pandas, datetime, and numpy (Figure 2).

The first step of the process involved using the pandas library to import the data in .csv format, enabling efficient data manipulation. Subsequently, the datetime library was used to standardize and convert the date and time information present in the data, ensuring temporal consistency throughout the dataset.

The creation of a personalized algorithm was the subsequent step, with the primary objective of automating the organization and restructuring of raw well data. This algorithm performed tasks such as ordering information, correcting potential errors, and standardizing date and time formats. The numpy library played a crucial role in specific mathematical operations applied to the data whenever necessary to enhance its quality and cohesion.

The final result of this process was a table organized in the appropriate format for analyzing time series of water levels in monitoring wells. This table now represents a solid and structured foundation, ready to be used in investigating water level trends and behaviors in the GAS.

In summary, the methodology employed in this study combined data collection from reliable sources, the application of programming techniques in Python, and the utilization of specialized libraries to automate the organization of

raw data. This approach facilitated the creation of a well-structured database suitable for time series analysis, providing a robust foundation for subsequent investigations related to water levels in the GAS.

Figure 2. Python libraries used in the development of the algorithm.

3. RESULTS AND DISCUSSION

An algorithm was developed in Python using the Pycharm development environment with the aim of processing, organizing, and optimizing data related to water levels in monitoring wells (Figure 3).

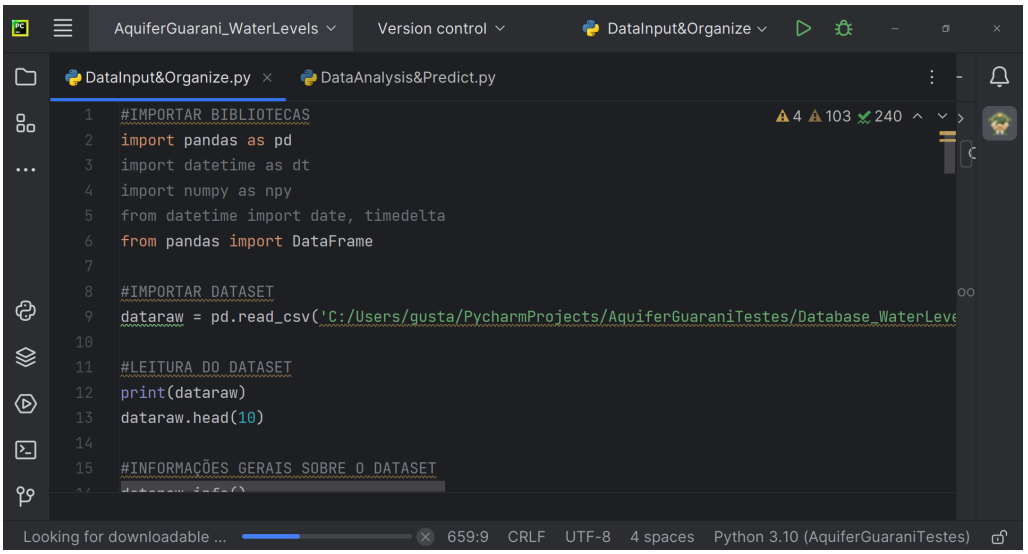

Figure 3. Pycharm interface presenting the algorithm that was developed.

The main results can be summarized as follows: (1) Data Reading and Transformation: The algorithm was able to read the raw data from the database, understanding its structure and data format. During this process, important data processing was carried out, converting the decimal separator from comma (",") to point (.) and adjusting the data types of the columns, making the water level information compatible with the format float64 and dates in string format; (2) Data Grouping: Based on predefined criteria, the algorithm created separate dataframes for each well and its respective water level. This allowed for a more efficient organization of information, facilitating subsequent analyses; (3) Data Consolidation: One of the most significant results was the consolidation of all data into a single database. This meant that water level data from different wells were related to a single date column, optimizing the database for creating time series

analysis graphs; (5) Identification of Empty Spaces: The algorithm also played an important role in identifying 'No data' records. These empty spaces represent situations where the water level gauge has failed to measure due to operational errors or equipment problems. This is essential to ensure the quality of the data and the reliability of subsequent analyses; (4) Comparison of Results: Finally, the study allowed for a comparison between the raw database imported from the website (Table 1) and the database format developed by the algorithm (Table 2). This comparison revealed the substantial improvements achieved by the algorithm, demonstrating its effectiveness in organizing and optimizing data for subsequent analysis.

Id Ponto	Latitude Decimal	Longitude Decimal	Numero de sequencia	Data da medicao	Hora da medicao	Nivel da agua
3500026838	-19.637.778	-54.082.222	1	03/12/2010	12:00	36,43
…	…	…	…	…	…	…

Table 1. Structure of the raw database imported from RIMAS.

Index	Date	Well1_WaterLevel	…	Well39_WaterLevel
1	01/01/2011	36,45	…	NA
…	…	…	…	…

Table 2. Database structure for reading in time series analysis, result obtained from the developed algorithm (Figure 2).

4. CONCLUSIONS

Based on the methodology applied, it is possible to conclude that: (1) The algorithm developed in this work plays a fundamental role in the preparation, organization, and simplification of data related to water levels in GAS monitoring wells; (2) Furthermore, it facilitates time series analysis, operational error agreements, and significantly improves data quality, making it a valuable application for future studies in the field of hydrogeological data analysis and water resources management. These results reinforce the importance of continuous development of tools and algorithms to enhance the monitoring and management of water resources, thus contributing to environmental sustainability and water security.

5. BIBLIOGRAPHIC REFERENCES

McKinney, W. 2010. Data structures for statistical computing in python. In Proceedings of the 9th Python in Science Conference (Vol. 445, pp. 51–56).