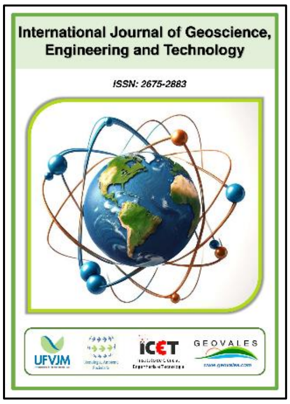
International Journal of Geoscience, Engineering and Technology
ISSN: 2675-2883
DOI:
Volume 10 – Nº 1 – Year: 2024 – pp.40-45



Received:
October 18, 2024

Accepted:
October 29, 2024

Published:
October 31, 2024

Automating Hydroclimatic Data Processing: A VBA Module for Efficient Handling of Large Datasets in Microsoft Access

Gláucio Ferreira Loureiro[1], Carlos Rogério de Mello[2]

[1] Universidade Federal dos Vales do Jequitinhonha e Mucuri, Teófilo Otoini-MG, Brasil
[2] Universidade Federal de Lavras, Lavras-MG, Brasil

Email address
glaucio.loureiro@ufvjm.edu.br (Gláucio F. Loureiro) – Corresponding author.
crmello@ufla.br (Carlos R. Mello)

Abstract

Hydroclimatic data plays a crucial role in the study of water resources and the surrounding environment. These data are typically obtained from the National Water Agency and the state agencies. However, limited data series are available, and the data may contain undesirable characteristics. Data processing is usually performed using modules developed in spreadsheets; however, this method is impractical when dealing with large volumes of data. This led to the development of a Visual Basic for Applications (VBA) module in Microsoft Office Access to handle the extensive datasets. The application was designed to automatically process empty fields, clean invalid characters, and group data when large gaps appeared at the end of some tables. Finally, it calculated the monthly, daily, and annual averages. The application has proven to be highly intuitive because users only need to follow the sequence of buttons displayed on the screen.

Keywords: Database, Programming module, Hydrological, calculations.

1. Introduction

Hydrological data play a fundamental role in studies related to water resources and surrounding environmental changes. As Aragão et al. (2013) explained, for integrated watershed management, it is essential to understand the generated surface runoff. This understanding must include an analysis of climatic factors, such as rainfall and evaporation, among other variables. Tucci (2001) emphasized that hydrological parameters can be divided into three categories: climatic parameters, runoff parameters, and characteristics of the receiving environment. Therefore, hydrological data must be observed over time in order to obtain more detailed information. The collection of time-series data is essential for building a hydrological model because it involves large volumes of information. Currently, a significant portion of these data is provided by the National Water Agency (ANA) or state regulatory agencies, with a smaller portion managed by small reading stations operated by companies or individuals. However, owing to the recent nature of much of the monitoring, or the difficulty in reading the devices, the data often presents failures, errors, or configurations that are not ideal for researchers, requiring considerable time for tabulation and data correction. Kalteh and Hjorth (2009) pointed out that one of the biggest obstacles in time-series studies is related to data issues, especially missing data. This is critical, because models involving surface runoff can be severely compromised. Paiva and Paiva (2001) stressed that reliable hydrological time series are crucial for all techniques that involve hydrological calculations. However, Tucci (2001) highlighted the difficulty in obtaining long-error-free precipitation datasets. Various studies have focused on developing resources for the processing and handling of hydrological data. Billah et al. (2016) developed web services that automate the workflow for pre-processing grid data for the preparation of inputs to the Variable Infiltration Capacity (VIC) model. According to Gan et al. (2020), this workflow includes information that allows for independent reproduction of the model results while also documenting the steps used to create the input files. The need for consistent longterm datasets has made the use of databases an essential tool for managing these data. As Silberschatz et al. (2006) explained, a database management system is a collection of programs that access a set of data, which is referred to as a database. A database is essentially a collection of data that requires access to a system. Microsoft Access is an example of a Relational Database Management System (RDBMS) that includes macros, forms, queries, and reports, providing tools for manipulating, defining, and controlling large volumes of data. Despite this, a significant number of hydroclimatological studies have been conducted using spreadsheets and word processors, which are not ideal for managing large datasets. The improvement and wider distribution of hydroclimatic monitoring stations across the country highlight the need for appropriate techniques and tools to manage the vast amount of data collected. With advancements in computing, the analysis of hydroclimatic processes has become faster and more accurate, allowing for the exploration of a greater number of alternatives in a shorter time frame (Cirilo et al., 1997). As water resource management institutions grow, so does the demand for technologies such as software that supports decision-making regarding the use of water resources (Almeida, 2006). In light of this, the objective of this work was to develop a tool using Access as a platform, with a module programmed in Visual Basic for Applications, to process hydro climatological data. This tool aims to assist hydrology professionals in applying mathematical models and supporting decision-making processes.

2. Methodology

The application was developed at the Water and Soil Engineering Department of the Federal University of Lavras. It utilizes Microsoft Office©, in which data are imported from a Microsoft Office Excel spreadsheet into Microsoft Office Access. The modules used in Access were programmed in VBA, and responsible for manipulating externally obtained data and checking for unfilled gaps or inconsistent data. The software is divided into several modules.

The first module, named "Apaga Vazios” (Remove Empty Fields) uses a set of queries to delete empty spaces in the spreadsheet and prevent blank values from being considered zeros, which could affect the average calculations and other computations that depend on the accuracy of the dataset.

Another key module is the "Processar Dados" (Data Processing) module, which is structured in VBA. This module generates a complete series of dates, from the first to the last date of the year, and checks for unwanted characteristics. This feature helps hydrological data analysts visualize any gaps in the date sequence, as in many cases, the sequential order of dates in the table does not reflect the actual time sequence. This makes it easier to identify hidden or missing dates and to correct flaws in the series.

The "Agrupar Dados" (Group Data) feature allows fields that are completely empty in the spreadsheet to be removed, leaving only fields with valid values. This is particularly important for calculations involving averages because errors can occur when empty fields are included. Through VBA programming, valid data are extracted from the original fields in the spreadsheet and transferred to a predetermined table in Access.

The "Processar Dados" (Process Data) module, also programmed in VBA, handles daily average calculation since most data series contain multiple readings per day. These calculations were performed using loop structures and conditional statements, allowing a daily average to be calculated for each hydroclimatic variable, forming a daily data series. Based on the daily data, it is also possible to calculate "Média Mensal" (monthly averages) and "Média Anual" (annual averages). These calculations are performed using queries that group the data by month or year and compute the respective averages. The ‘Delete Data’ button runs a macro that executes multiple queries, which are resources available in Access, to clear all data from the tables and queries in the application. This feature is essential for preventing the mixing of hydrological data from different series. In addition, there is a text field in the application that stores the file path where the program is saved. This text field is read each time the VBA modules in the application are executed.

3. Results and discussion

The application features a final screen with four buttons on the left, three buttons on the right, and an additional button to delete data (Figure 1).

The application operates by importing data to be processed into a database. This import is performed into a pre-existing table created within the application for this purpose (Figure 2). The import process requires specific steps to ensure that the table structure within the application remains constant. Therefore, the Excel or Access spreadsheet to be imported must have variable headers that exactly match the headers in the application table.

After clicking the 'Import' button, an assistant provided by Access will guide the import process. The first screen displays a box requesting the file path of the document to be imported. This is illustrated in Figure (3).

After starting the assistant, it will guide the user through the final steps of the import process. At the end, the name 'Dados_Importados' must be entered to replace the existing table in the database. However, as shown in Figure (4), care must be taken to ensure that the imported spreadsheet headers exactly match the headers expected by the application.

After completing the import process, data were available for manipulation within the application. The next step begins by clicking the first button, labeled 'Apaga Vazios' (Delete Empty Cells), which removes any empty cells at the beginning or end of the imported spreadsheet or any data entries without dates, as these may cause errors when the application performs calculations. Figure (5) shows a message indicating how many rows that are completely devoid of data need to be deleted. In this example, no rows need to be deleted because the data have already been pre-processed manually.

Subsequently, the user clicks the 'Tratar Dados' (Process Data) button, which checks for any letters in place of numbers or other undesirable patterns that could interfere with the calculations. Next, the 'Agrupar Dados' (Group Data) button was executed, which aggregates the data originally separated by hours into a daily scale. This is the first calculation performed using this application. Finally, the user clicks on 'Processar Dados' (Process Data), where the application calculates the averages on a monthly and annual scale.

The processed data is presented in a table format and can be exported to an Excel spreadsheet for use in other applications. Once the data have been worked on, the application provides the option to export the processed data to an Excel spreadsheet. The export is simple: the user just needs to select the spreadsheet displayed by the application, execute the computer's 'copy' command, and then paste the data into any Excel worksheet. In addition, the software features a display showing the generated dates and a field where the user can input the file path where the application is saved, allowing the user to store the application in any folder on their computer. Carleton et al. (2005), in their research, developed a database designed to facilitate the storage, retrieval, and analysis of hydrological data. The authors concluded that, among the various advantages of data processing, this type of tool allows for data sharing over the Internet, or database replication, which can be used within a network to facilitate collaboration among multiple users, making it a powerful tool for water managers and researchers.

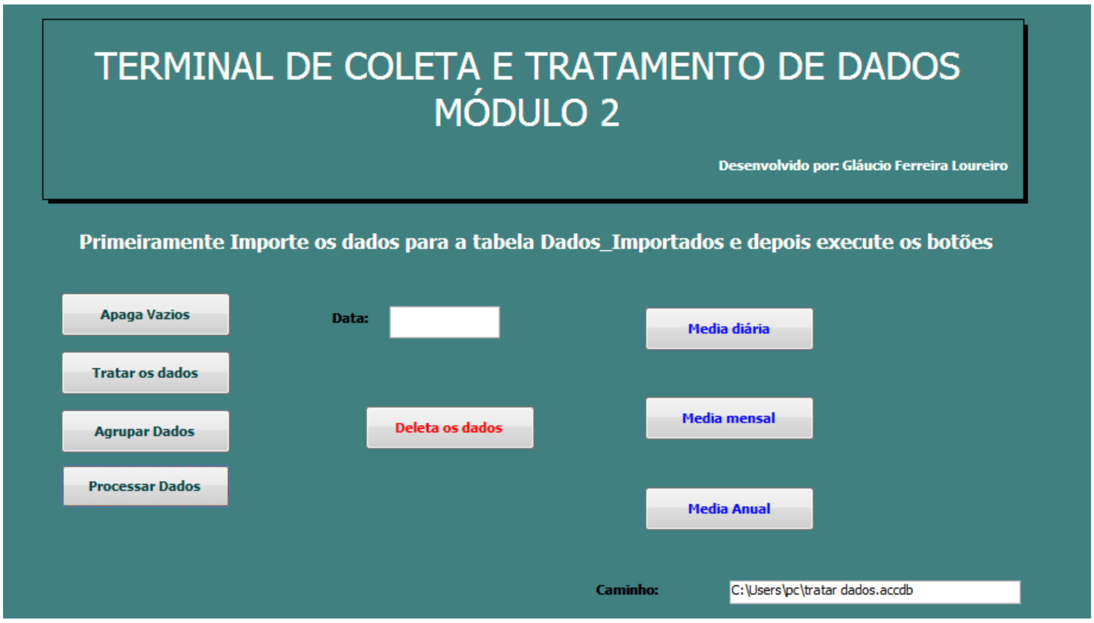


Figure 1 – Main application screen.

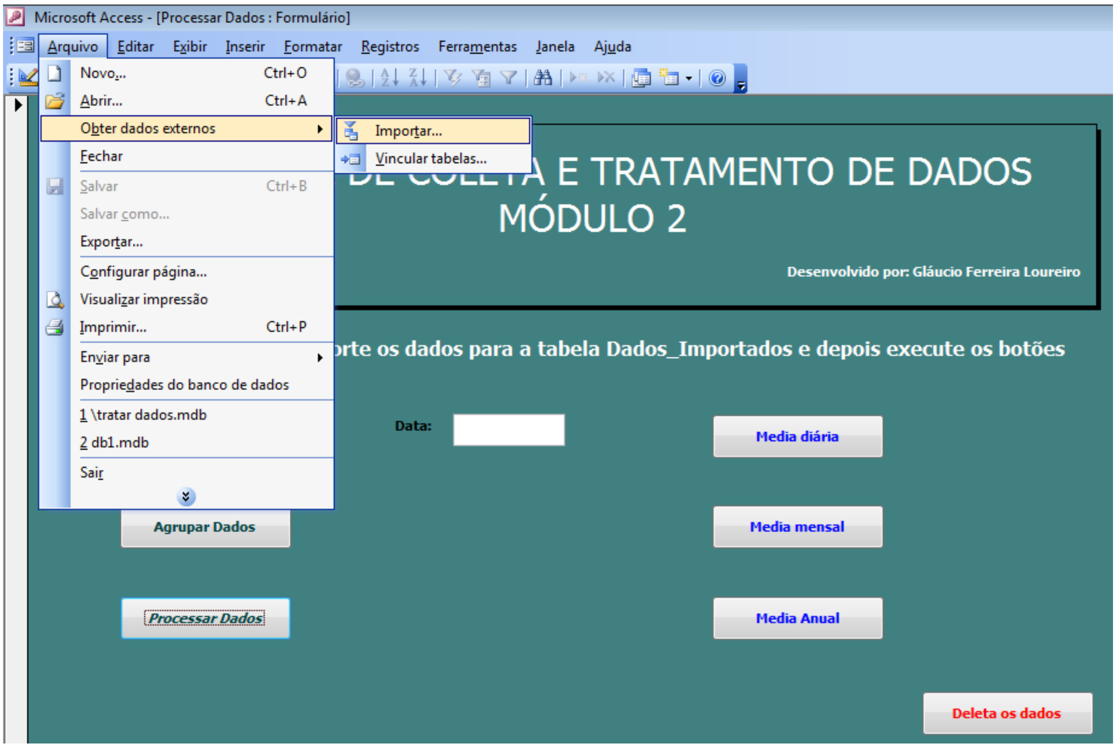


Figure 2 – Data Import Screen.

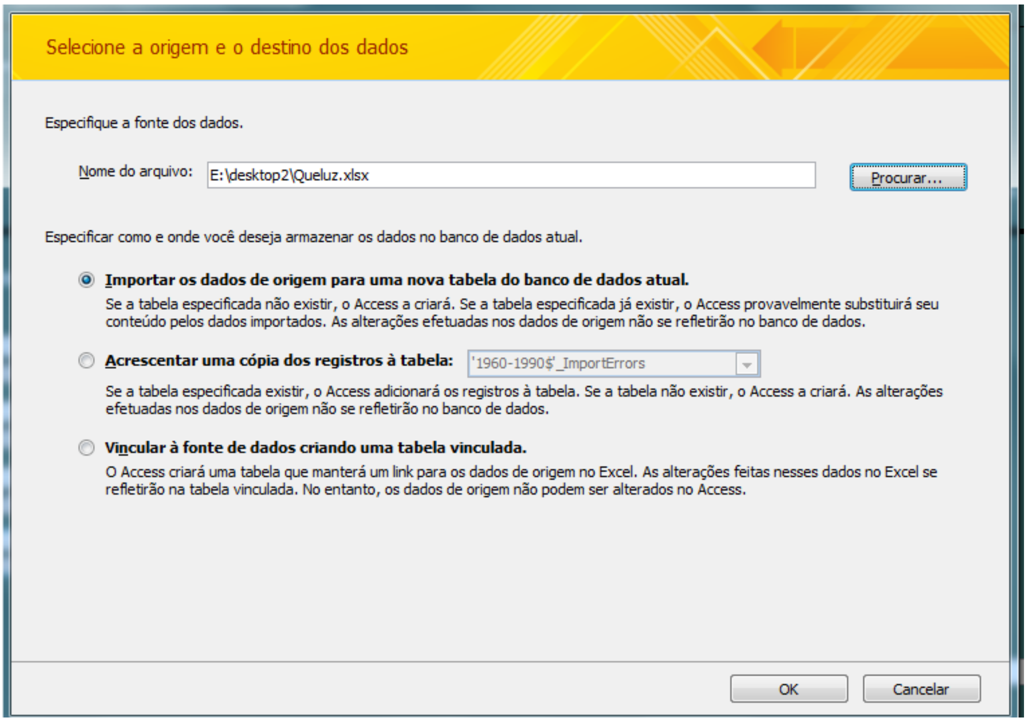


Figure 3 – Import Assistant Screen.

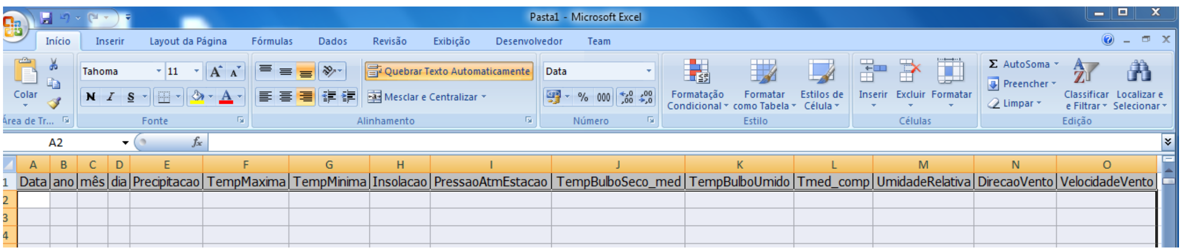


Figure 4 – Screen Showing the Spreadsheet Template Header.

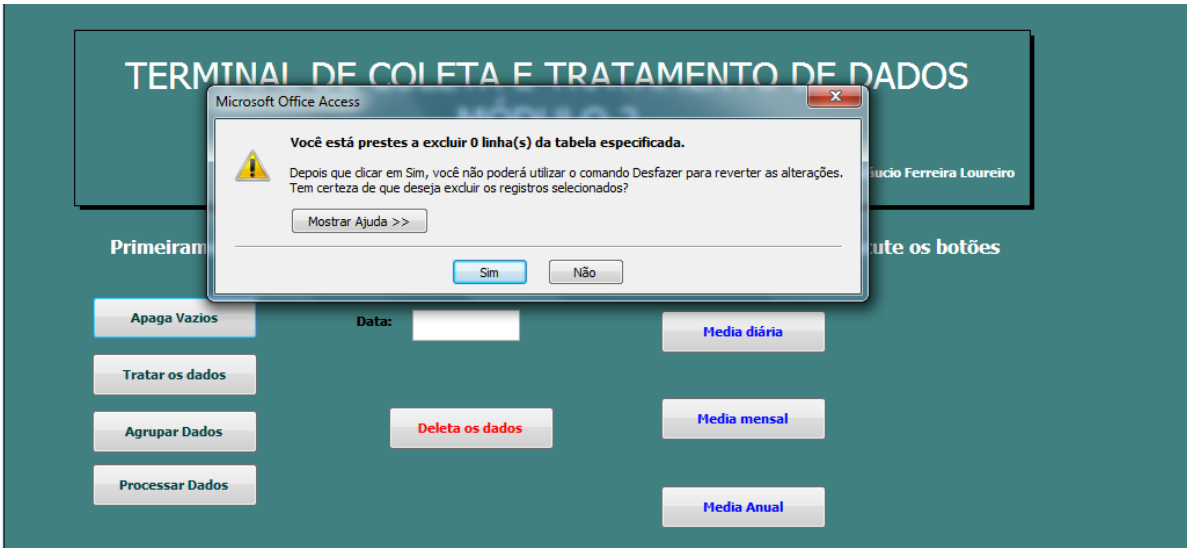


Figure 5 – Screen Showing the Number of Rows to Be Deleted from the Database Table.

4. Conclusion

The application featured a highly user-friendly interface that facilitated its use by study groups involving professors and undergraduate research students. Additionally, the time required for data manipulation and tabulation was significantly reduced compared with previous manual methods, demonstrating its effectiveness as a tool for database management and manipulation. The final results calculated by the application showed strong consistency with a precision of up to 16 decimal places. Therefore, this tool is a valuable asset in handling, processing, and organizing databases, combined with its ease of use and simplicity in performing calculations. Future developments of the application will include calculations of monthly averages over a 30-year period, which could be used, for example, to derive a hydroclimatological normal.

References

Almeida, C.N., 2006. Personalização de Planilhas Excel® para Programação de Modelos – O Caso da Simulação da Operação de Reservatórios. In: *Simpósio de Recursos Hídricos Do Nordeste*, 7. João Pessoa: Universidade Federal da Paraíba.

Aragão, R., Cruz, M.A.S., Amorim, J.R.A., Mendonça, L.C., Figueiredo, E.E. and Srinivasan, V.S., 2013. *Análise de sensibilidade dos parâmetros do modelo SWAT e simulação dos processos hidrossedimentológicos em uma bacia no agreste nordestino*. Revista Brasileira de Ciência do Solo, 37(4), pp. 1091-1102. https://doi.org/10.1590/S0100-06832013000400026

Billah, M.M., Goodall, J.L., Narayan, U., Essawy, B.T., Lakshmi, V., Rajasekar, A. and Moore, R.W., 2016. *Using a data grid to automate data preparation pipelines required for regional-scale hydrologic modeling*. Environmental Modelling & Software, 78, pp. 31-39. https://doi.org/10.1016/j.envsoft.2015.12.010

Carleton, C.J., Dahlgren, R.A. and Tate, K.W., 2005. *A relational database for the monitoringand analysis of watershed hydrologic functions: I. Database design and pertinent queries*. Computers & Geosciences, 31(4), pp. 393-402. https://doi.org/10.1016/j.cageo.2004.10.007

Cirilo, J.A, Baltar, A.M., Rolim Junior, A.L., Torres Filho, C.O. and Azevedo, J.R.G., 1997. *Processamento integrado de dados para análise hidrológica*. Revista Brasileira de Recursos Hídricos, 2(2), pp. 21-44. http://dx.doi.org/10.21168/rbrh.v2n1.p15-28

Gan, T. Tarboton, D.G., Dash, P., Gichamo, T.Z. and Horsburgh, J.S., 2020. *Integrating hydrologic modeling web services with online data sharing to prepare, store, and execute hydrologic models*. Environmental Modelling & Software, 130, pp. 104731. https://doi.org/10.1016/j.envsoft.2020.104731

Kalteh, A.M. and Hjorth, P., 2009. *Imputation of Missing values in precipitation-runoff process database*. Journal of Hydrology Research, 40(4), pp. 420-432. https://doi.org/10.2166/nh.2009.001

Paiva, J.B.D. and Paiva, E.C.D. *Hidrologia aplicada à gestão de pequenas bacias* – Porto Alegre: ABRH, 2001.

Silberschatz, A., Korth, H.F. and Sudarshan, S., 2006. *Sistema de Banco de Dados*. Rio de Janeiro: Elsevier, 2006.

Tucci, C.E.M. *Hidrologia*: *ciência e aplicação.* Porto Alegre: Ed. Universidade/UFRGS: ABRH, 2001.