

Hydrovars: an R tool to collect hydrological variables

Alejandro Sánchez-Gómez^{a,*}, Katrin Bieger^b, Christoph Schürz^c, Silvia Martínez-Pérez^a, Hendrik Rathjens^d and Eugenio Molina-Navarro^a

^a Department of Geology, Geography and Environment, University of Alcalá, Madrid, Spain

^b Department of Ecoscience, Aarhus University, Aarhus, Denmark

^c Department of Computational Landscape Ecology, Helmholtz Centre for Environmental Research, Leipzig, Germany

^d Stone Environmental, Inc., Montpelier, USA

*Corresponding author. E-mail: alejandro.sanchezgomez97@gmail.com; alejandro.sanchezg@uah.es

 AS, 0000-0003-1085-0941

ABSTRACT

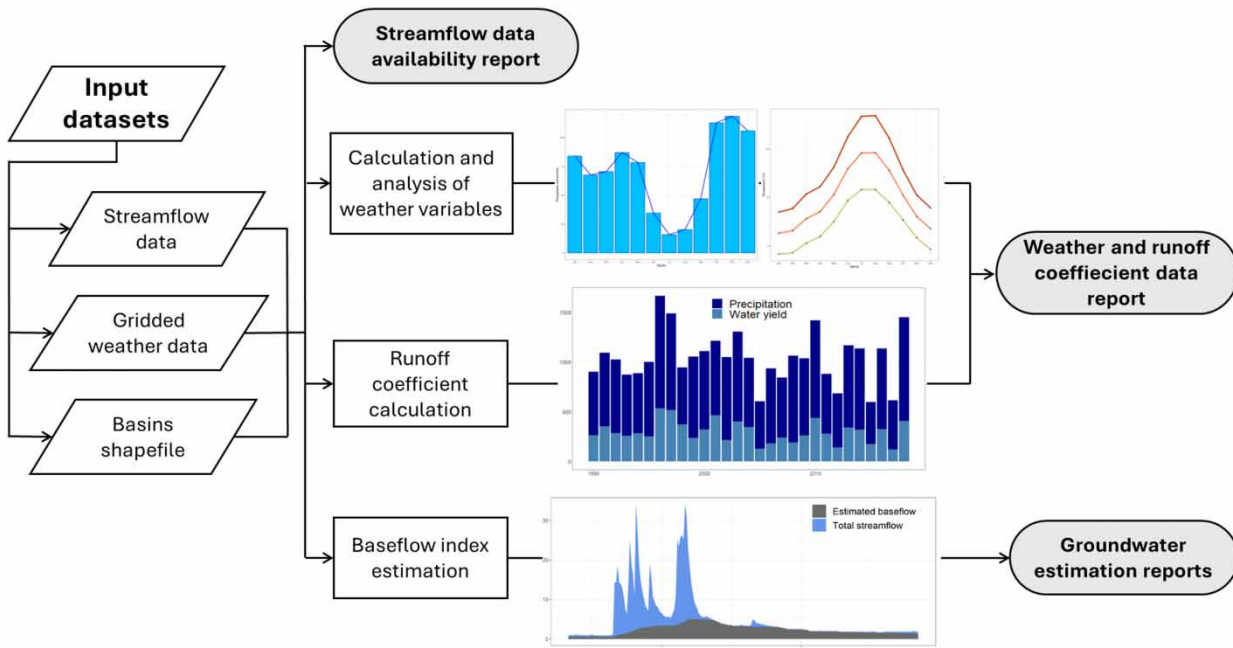
Hydrological models can benefit from soft calibration, a process by which the proper simulation of hydrological variables is proved while or before addressing hard calibration. Soft calibration reduces the probability of obtaining a statistically accurate but unrealistic model. However, it requires soft data, which is often hard to acquire or unavailable. This work presents HydRoVars, an R tool developed to facilitate the estimation of data which can be implemented in a soft calibration procedure. It allows us to estimate two key hydrological indices (the runoff coefficient and baseflow index) and weather-related variables at the catchment scale for one or numerous basins. The runoff coefficient is calculated automatically from precipitation and streamflow datasets. Groundwater contribution is estimated through a semi-automatic process based on a baseflow filter which considers hydrogeological properties. Modellers would benefit from incorporating soft calibration in their calibration procedures, and this tool might help to estimate these relevant hydrological variables in their modelled area. The tool has been tested in 19 subbasins of the Tagus River basin (Spain) located in different geological regions. In the test cases, we demonstrate the usefulness of this tool to improve the model representation and gain an understanding of the catchments' hydrology.

Key words: groundwater assessment, hydrological modelling, hydrological processes, R, soft calibration, soft data

HIGHLIGHTS

- An R tool has been developed to collect weather and hydrological variables.
- It allows estimating the runoff coefficient, and the baseflow index at the catchment scale.
- Runoff coefficient can be automatically obtained for multiple basins.
- Guidance is provided for a realistic groundwater contribution estimation.
- Its usefulness has been tested in the Upper Tagus River basin (Spain).

GRAPHICAL ABSTRACT

HydRoVars: Easy collection of weather and hydrological data**1. INTRODUCTION**

Hydrological models are extremely useful for addressing challenges in managing water resources, and the availability of user-friendly interfaces facilitates their use (Gassman *et al.* 2014). However, hydrological modelling applications sometimes present results which are solely based on analysing the performance of a certain metric (or metrics) for the calibrated variable (e.g., streamflow, evapotranspiration, etc.), without further discussing if the simulated hydrological processes in the model resemble the reality of the study area (Bahremand 2016; Acero Triana *et al.* 2019).

To avoid equifinality or obtaining an accurate (in terms of statistical performance) but unrealistic (in terms of hydrological processes representation) model (Muleta 2012; Molina-Navarro *et al.* 2017), modellers should attempt to ensure that at least the main components in the water balance and the streamflow components are properly represented considering the characteristics of the modelled area (e.g., Guse *et al.* 2014; Sánchez-Gómez *et al.* 2023). Some researchers discussed this issue and proposed the use of soft calibration techniques to achieve more realistic hydrological models (e.g., Seibert & McDonnell 2002; Arnold *et al.* 2015), i.e., assessing model results against soft data, such as annual averages of the water balance components or the expected relative contribution of the different components of the streamflow. Soft calibration can be followed by hard calibration, which would involve the comparison of simulated and observed time series of a certain variable (e.g., streamflow). The works of Pfannerstill *et al.* (2017) and Chawanda *et al.* (2020) are examples of successfully using soft data for the calibration of hydrological models.

Soft calibration requires soft data, a concept that encompasses different types of information; from a range of expected values for a model parameter to a maximum water level value based on historical records (Efstratiadis & Koutsoyiannis 2010). Two key soft hydrological variables are the runoff coefficient and the baseflow index (Blume *et al.* 2007). The runoff coefficient is the fraction of precipitation volume that becomes streamflow. It is not constant and depends on factors such as soil type, lithology, topography, or the characteristics of precipitation events. The runoff coefficient can also be used to estimate the evapotranspiration ratio, which is the main loss of the water balance in Mediterranean regions (García-Ruiz *et al.* 2011). In contrast to precipitation and streamflow, measurements of actual evapotranspiration are rare at the catchment scale. Thus, despite having some limitations, the runoff coefficient helps to understand the water balance of a basin. The runoff coefficient can be estimated for any area where precipitation and streamflow data are available. Even though both

streamflow and precipitation data are available, studies in Spain estimating this variable are limited to specific rainfall events, focusing on the factors that influence runoff generation (e.g., Kirkby *et al.* 2002; Rodríguez-Blanco *et al.* 2012). To the best of our knowledge, there is a lack of studies analysing long time series that could be used for calibration purposes.

Baseflow is the fraction of the streamflow that is released from natural water storage systems such as glaciers or aquifers, and therefore does not respond immediately to precipitation events. Groundwater is the main source of baseflow in Mediterranean areas, and it is responsible for maintaining streamflow during dry periods, which is essential for water supply, water quality, and ecosystem functioning (López-Vera 2012). Groundwater flow depends on factors such as lithology, topography, climate, or soil type, among others. The baseflow index indicates the fraction of the total streamflow that originates from baseflow and can be estimated through different methods: through hydrological models (Samper *et al.* 2015), mass-balance methods with chemical tracers (Ortega *et al.* 2015), or applying numerical or graphical methods to streamflow records to perform a separation of its components (baseflow and direct runoff) (Lyne & Hsollick 1979; Custodio & Llamas 1983). Various digital filters for this separation have been developed, allowing us to estimate the baseflow index using streamflow data (Arnold & Allen 1999; Eckhardt 2005; Kang *et al.* 2022).

Values of runoff coefficient or baseflow index for a certain catchment can be found in the literature if previous hydrological studies for that area exist. Some studies might also provide expected values for certain regions, depending on their climate and/or their hydrogeological characteristics (e.g., Custodio & Llamas 1983). However, a reliable soft calibration procedure would involve obtaining runoff coefficient and baseflow index estimates specifically for the catchment and time period of interest.

Obtaining values for these two indices in a particular catchment often requires tedious work of data collection and processing. Regarding the runoff coefficient, it would involve the collection of precipitation data and its interpolation to have a catchment average, as well as gathering streamflow gauging records for the same period and converting them to the appropriate units to calculate the desired coefficient. Regarding the baseflow index, the work would involve collecting hydrographs of several streamflow peaks to then apply hydrograph separation techniques (Custodio & Llamas 1983; WMO 2008), thus being able to calculate the groundwater contribution for the period of interest. This sometimes has been done manually (e.g., Molina-Navarro *et al.* 2016), while other authors have used automatic computerized techniques (Kang *et al.* 2022). Among those, Arnold & Allen (1999) developed a software that incorporates a digital filter for baseflow estimation that has been used in numerous modelling applications (e.g., Meaurio *et al.* 2015; Senent-Aparicio *et al.* 2021a). Koffler *et al.* (2022) developed an R package which estimates the baseflow index through a linear interpolation of minima turning points in a streamflow time series, which could also lead to uncertainties depending on the hydrological regime (WMO 2008). A more reliable hydrograph separation, however, would entail an active user involvement to guide the separation process.

Considering the above, there is a need to develop tools to derive realistic values of these two soft data variables for one or several catchments, for a certain period, and using the same working environment. This would ultimately encourage modelers to subsequently perform soft calibration procedures, thus advancing towards a better hydrological modelling. Besides, in those basins that present heterogeneous hydrological properties, deriving these variables for several sub-catchments could allow carrying out zonal calibrations, which might result in more robust models (e.g., Sánchez-Gómez *et al.* 2022).

This work presents HydRoVars, a semi-automatic R tool that allows us to collect two soft data variables at the catchment scale, namely the runoff coefficient and the baseflow index, for a certain area and time period. It also allows us to gather and process weather variables (precipitation and temperature) at different time scales. Some advantages of this tool are that it uses weather and streamflow datasets which are readily available, that it entirely works in the R environment, and that it allows us to estimate easily weather and hydrological variables for numerous basins, which will save time for users. In comparison with other tools that perform hydrological analyses, this one is semi-automatic, allowing the user to control certain processes such as the streamflow components separation, where the properties of the studied area should be considered.

To prove the applicability of this tool, it has been applied to 19 subbasins of the Upper Tagus River Basin (UTRB) in Spain. The UTRB is geologically heterogeneous and modelling its complex hydrological system can greatly benefit from spatially distributed soft information on the dominant hydrological processes.

2. METHODOLOGY

2.1. General workflow

The developed tool can be downloaded from the repository provided in the supplementary material (Sánchez Gómez 2024) (see Data availability section) and requires R and RStudio (R Core Team 2022). It uses the following input datasets: (i) a file

with daily streamflow data from the selected gauging stations with an identifier; (ii) a directory with gridded meteorological data (precipitation and temperature) and a vector layer (point type) with the location and identifier of the grid points; and (iii) a single vector layer (polygon type) with the delineation of all the basins to be analysed (i.e. the drainage areas of the streamflow stations selected, that should be prepared beforehand). Although temperature data is not required to calculate the selected indices, it is worthy to include it as a variable to extract, since it directly affects evapotranspiration and therefore the runoff coefficient, so it might be useful for analysis and discussion of results.

Once the input datasets are ready, the first step consists in creating two csv files that will be used as inputs in posterior processes. The first file, created from the basins vector file, contains information about the analysed basins. The second file defines the weather grid points that should be used to calculate the weather variables in each basin. Once the input files are created and read, the R code allows estimating the runoff and baseflow coefficients, being the user guided both by the comments provided in the code and by the recommendations given in this manuscript, particularly towards the baseflow index calculation. Two additional files which contain the results of the baseflow index calculation are created too. Finally, some output files which summarize the more relevant information are automatically created with the code; examples can be found in Appendix A (see appendices in Supplementary Material). Appendix B shows the main lines of the code to address these processes, but the full code explained in greater detail can be obtained from the GitHub repository included in the Data availability section. Figure 1 presents the general workflow of the proposed methodology, and the different steps are fully described in the next subsections.

The data sources used in this manuscript allow us to directly reproduce the methodology proposed anywhere in Spain, but as long as the data requirements abovementioned are available, it can also be applied anywhere in the world with minor code adaptations to read different input data formats.

2.2. Study area

The usefulness of this methodology has been tested in the UTRB; a catchment highly relevant for water management in Spain. The Tagus River (Figure 1) is the longest in the Iberian Peninsula, and with approximately 11 million inhabitants, its basin is the most populated. Water resources in this basin are scarce considering the high water demand for irrigation and human consumption, and it is expected that climate change will exacerbate this situation (Bednar-Friedl *et al.* 2022). In addition, a water transfer system diverts water from the headwaters of the Tagus River to the southeast of Spain to satisfy irrigation

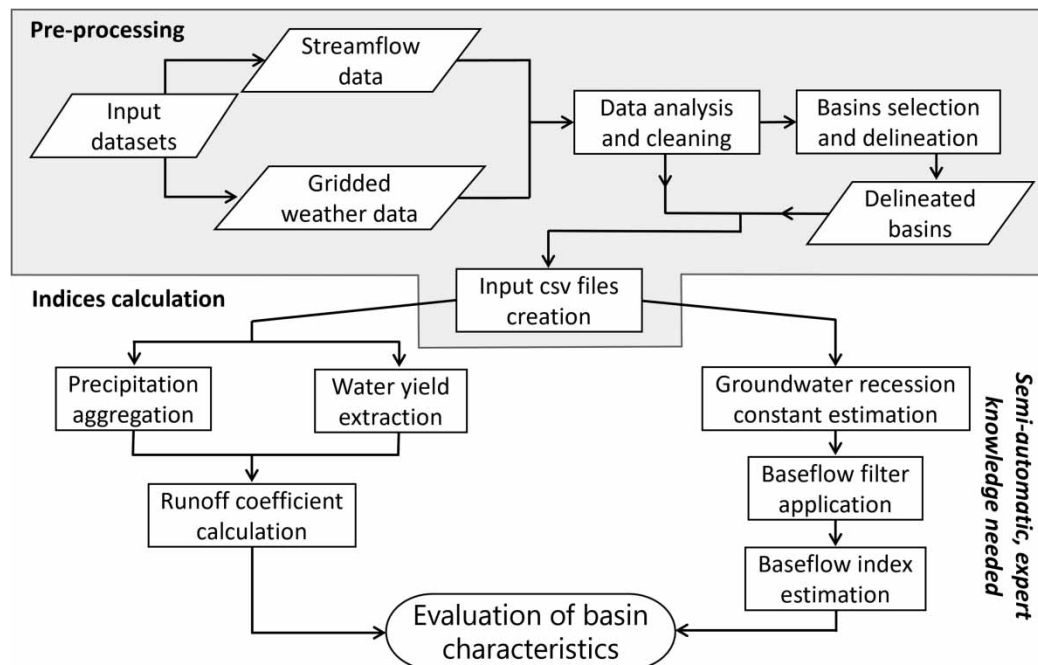


Figure 1 | Workflow chart of the presented methodology.

and human consumption demands. Thus, realistic hydrological modelling of this basin is of utmost importance in order to make reliable future assessments of, for example, climate change impacts.

The study area for this work is the upper part of the Tagus River basin (UTRB), which has a size of 33,880 km², i.e., approximately 40% of the entire Tagus River basin (Figure 2). It is delimited by three mountain ranges: the Central System in the north, the Iberian System in the east, and the Toledo Mountains in the south. The UTRB altitude varies between 2,400 m.a.s.l. in the Central System mountain range and 360 m.a.s.l. at its outlet in Talavera de la Reina (Toledo).

The study area is characterized by a continental Mediterranean climate, which varies depending on factors such as altitude, proximity to the ocean and latitude. Chazarra Bernabé *et al.* (2018) classified it as temperate (C) and dry (B) following the Köppen–Geiger climate classification for the period 1980–2010. The mean precipitation for that period was 550 mm, with values higher than 1,200 mm in the Central System and lower than 400 mm in the driest areas of the southern part of the UTRB. The mean temperature is 12 °C, ranging from 8 to 10 °C in the Central System to 13 °C in the east and 17 °C in the west (Peral García *et al.* 2017). The average precipitation and water yield of the basin decreased over the last decades, which can be attributed to the impacts of climate change.

Natural vegetation and agricultural lands are the main land covers in the UTRB, occupying around 48 and 47% of the total area, respectively. The dominant crops are cereals, followed by other crops such as sunflower or legumes. Urban land covers around 3% of the total area.

Due to mountain systems with different bedrocks, the UTRB is characterized by a heterogeneous geology, and the dominant hydrological processes vary within the different geological zones (Custodio & Llamas 1983; Sánchez-Gómez *et al.* 2022). Direct runoff is expected to be predominant in the Central System and in the Toledo mountains, which are composed of igneous and metamorphic rocks with low permeability. In contrast, the carbonate rocks of the Iberian System are important aquifers and baseflow is expected to be more relevant there. The sedimentary deposits of the basin also have varying hydrological properties, including Tertiary rocks and Quaternary alluvial sediments with high permeability that are relevant aquifers, and detrital and evaporitic materials with low permeability.

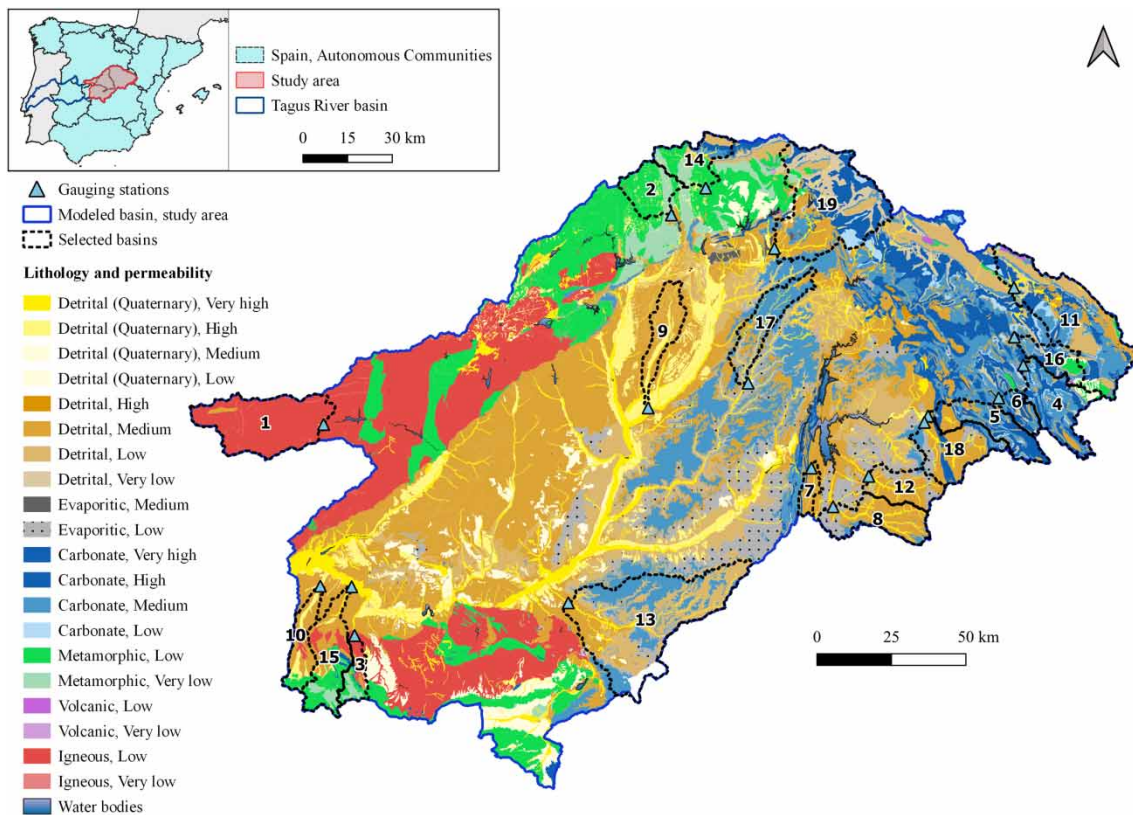


Figure 2 | Study area location, lithology, permeability, and selected basins (see section 2.3). Lithology and permeability information taken from IGME (del Pozo Gómez 2009).

2.3. Input data preparation

Daily precipitation and temperature data for the UTRB were obtained from the Spanish Meteorological Agency (AEMET) grid (5 km resolution, [Peral García et al. 2017](#)), which includes daily data for all the Spanish territory from 1951 to 2019. This database was transformed into SWAT readable format by [Senent-Aparicio et al. \(2021b\)](#) and was the one used in this work due to its convenient format. A point vector layer containing the location of the centroids of the weather grid cells was created and each point was assigned a unique identifier. Streamflow data were collected from the CEDEX streamflow gauging stations annual report ([CEDEX 2021](#)), which contains daily records of all gauging stations in Spain.

Truthful and complete input data are necessary to perform a reliable analysis of hydrological and weather variables. Thus, for the studied period, the used datasets were graphically and statistically analysed (range of values, absence of outliers or null values, etc.). In the case of streamflow data, verification of data availability for the entire working period should be done, since missing data can lead to incorrect results. The methodology allows us to create a report on the availability of streamflow data (see example in Appendix A1).

The last required input is a vector file containing the polygons of the basins to be analysed. This step was intentionally taken out of the R workflow, since they might already be delineated, or the user might prefer to delineate them using a GIS software to guarantee a reliable delineation (often the delineation steps need to be visually inspected at various zoom levels and/or by overlaying them with aerial imagery or other spatial data).

The selection of basins has to be based on two criteria: the availability of streamflow records for the desired working period, and having a mostly undisturbed regime, without any major reservoirs in the catchment or significant water withdrawals. Users should consider that the lack of streamflow records or the presence of elements which alter the hydrological processes might affect the reliability of the results obtained through the methodology. An additional third criterion was used in the study case presented in this manuscript: the geological region in which each basin is located. Grouping the subbasins in regions (geological or of any other desired type) might be useful for deriving average indices values per region for a subsequent soft calibration process. Finally, 19 subbasins were identified as suitable for the calculation of runoff and baseflow coefficients for the different geologic zones ([Figure 2](#)).

For the study case presented, four different geological regions or substrates were defined according to the lithology and permeability of the UTRB ([del Pozo Gómez 2009](#)). The defined regions are: (i) an impervious region (IMP), with low and very low permeability metamorphic and igneous rocks; (ii) a carbonate region (CRB), composed of carbonate rocks with high and medium permeability mainly; (iii) a high and medium permeability detrital region (DTH), conformed by high and medium permeability Tertiary sedimentary rocks and Quaternary sediments; and (iv) a low permeability detrital region (DTL), conformed by detrital and evaporitic materials with low permeability (i.e., marls, gypsum). As most of the subbasins had more than one type of substrate within, the proportion of each material in each subbasin was calculated ([Appendix C](#)). Subbasins where at least 70% of the area was composed of one substrate were considered representative for the IMP, CRB, and DTH regions. However, DTL substrates were not dominant in any gauged subbasin, and therefore subbasins with at least 25% of this substrate were considered representative for this region. Three to four representative subbasins were selected for each geological zone ([Figure 1](#), [Appendix C](#)), and six basins with a mixed lithology were chosen additionally, in order to characterize the hydrologic behaviour in heterogeneous regions (MIX).

The working period selected to obtain the runoff and baseflow coefficient was 2010–2018. This is because the main goal of this manuscript is to present a methodology to obtain these two variables for a certain modelling period so they can afterwards guide a subsequent soft calibration process. This period seemed convenient since it contains the latest available data and enough years for a future model evaluation in which the observed indices can be compared with simulated ones. Besides, this period covers a large interannual climate variability, with dry, wet, and average years, which is desirable for future modelling purposes so the calculation of runoff and baseflow coefficients takes into account different climatic conditions. If the purpose of using the tool is to characterize the climate and hydrological behaviour in a region, longer periods would be more appropriate (i.e., at least 20–30 years). Some of the subbasins selected for the study case (subbasins 2, 6 and 14) did not have streamflow records for the entire time period ([Appendix A 1](#)), but were considered important for ensuring that all of the geological regions in the UTRB were adequately represented. In these cases, only the years with complete data were used for calculating the runoff and baseflow coefficients. More information about the selected subbasins and their available streamflow data can be found in [Appendix A](#).

2.4. Creating workflow input files

The first step performed within the proposed R workflow creates two files that identify the relevant input data for the individual basins of interest. After loading the basins vector file, the *sf* and *dplyr* packages (Pebesma 2018; Wickham 2022) are used to calculate the area of each basin, and a field with the identifier of the gauging station for each basin must be added, finally creating the file *1_basins_file.csv*. Next, all the weather points located within a basin or a 1 km buffer outside its boundary are assigned to the respective basin using the weather grid vector file and the basin shapefile. Buffer extension is defined by the user, and depends on weather data resolution. A csv file containing the identifier of each grid point located in the basins and which basin it belongs to is then created and saved as *2_ids_stations_file.csv*. Note that in the weather datasets used in the case study presented, the identifiers and location of the precipitation and temperature grid points are the same, so only one file is needed for both variables. If that is not the case, two different files have to be created (following the same methodology) and used for the calculation of precipitation and temperature (see section 2.5). The code used to create both files can be found in Appendix B (Snippet 1). The two csv files created are the input files required for the calculation of the runoff and baseflow coefficients (see sections 2.5 and 2.6).

2.5. Runoff coefficient calculation

For the calculation of the runoff coefficient, input precipitation and streamflow datasets (section 2.3) and the input csv files created in the previous working step (section 2.4) are used. The code (Appendix B, Snippet 2 and 3) automatically aggregates precipitation and runoff data and calculates the average runoff coefficient (Appendix B, Snippet 4) both for the entire period and annual values. The observed streamflow data are aggregated from daily streamflow records (m^3/s) to annual contribution values (mm/year) using the basin area. Daily precipitation data (mm) is aggregated to calculate annual precipitation (mm/year) for each grid point and then the average precipitation for each basin is derived. In addition, the average maximum, minimum, and mean temperature is calculated if temperature data are included. Then, the average runoff coefficient for the selected period in each basin is calculated, also allowing us to extract annual values to analyse the interannual variation. If the area of study contains several regions, like the geological regions of the study case presented, the code allows the user to calculate the average runoff coefficient per region too.

2.6. Baseflow contribution estimation

The methodology proposed calculates the baseflow contribution using a two-parameter baseflow filter function following the equations in Eckhardt (2005), which are in turn based on the Lyne & Hollick (1979) algorithm and subsequent modifications by Chapman & Maxwell (1996) and Chapman (1999). This algorithm has already been used in other automated methods for estimating baseflow (e.g., Arnold & Allen 1999), but in this manuscript, rather than applying a fully automatic baseflow separation, we provide recommendations for a semi-automatic application guided by expert knowledge, and thus ensuring an accurate and realistic calculation of baseflow contribution for the different basins to be analysed.

In the following lines, the theoretical basis for the baseflow separation and the baseflow index calculation are explained, since a detailed explanation helps to understand how to obtain the required parameters for the baseflow filter application. The baseflow filter uses the following equation:

$$b_k = \frac{(1 - \text{BFI}_{\max}) * \alpha * b_{k-1} + (1 - \alpha) * \text{BFI}_{\max} * y_k}{1 - \alpha * \text{BFI}_{\max}} \quad (1)$$

where the baseflow for a day (b_k) (same units than the observed data) is calculated from the total streamflow of this day (y_k) and the baseflow in the previous day (b_{k-1}) using the parameters BFI_{\max} and α . BFI_{\max} is the maximum value of baseflow contribution expected for 1 day (fraction), and α is defined in Equation (2) by Eckhardt (2005) as a recession constant in the form:

$$b_k = \alpha * b_{k-1} \quad (2)$$

where b_k is derived from b_{k-1} multiplied by α (notice that in the referred manuscript this parameter is defined as 'a', but in the R function is computed as ' α ', nomenclature that we decided to keep in the manuscript). This might lead to confusion, since the recession constant in hydrogeological literature is usually presented as α , and defined as in the following

equation (Maillet 1905; Zhu *et al.* 2010):

$$Q_t = Q_0 * e^{-\alpha * t} \quad (3)$$

where Q_0 is the streamflow at the beginning of the recession and Q_t is the streamflow on day t . However, Equation (3) can be converted to a linear equation using logarithms:

$$\ln(Q_t) = \ln(Q_0) - \alpha * t \quad (4)$$

where the intercept is the streamflow at the beginning of the recession (Q_0) and the slope is the classic groundwater recession constant (α). Considering Equations (2) and (3), when $t = 1$, and making the assumption that during the recession all the flow is baseflow, it can be inferred that:

$$alpha = e^{-\alpha} \quad (5)$$

where $alpha$ is the parameter used in the filter and α is the classic groundwater recession constant.

Considering the above, the proposed procedure to estimate the groundwater contribution consists in the following steps: (i) estimate the value of $alpha$, (ii) apply the baseflow filter, changing the values of $alpha$ and BFI_{max} until a realistic baseflow separation is achieved, and (iii) calculate the groundwater contribution. These steps are guided by the tool.

For the estimation of $alpha$ as incorporated in the filter, the groundwater recession constant can be calculated in its classic way (α), using Equation (3). If this equation is converted to a linear one using logarithms (Equation (4)), α becomes the slope of this linear equation and can be estimated. To reduce uncertainty, we recommend identifying the groundwater recession curve for three representative streamflow peaks in each basin. These peaks can be selected by considering aspects such as their magnitude, duration, and shape (Figure 3). Then, the linear regression of the natural logarithm of the streamflow on the recession time (Figure 3) is performed (code in Appendix B, Snippet 5), obtaining a linear equation similar to Equation (4). The initial point of the recession curve has to be selected when the direct runoff ceases and the baseflow takes maximum values.

Despite the ambiguity in determining the beginning and the end of the recession curve (Blume *et al.* 2007), representing the streamflow with a semi-logarithmic axis can help to determine these points, since the recession curve conforms approximately to a straight line (Figure 3). A minimum of 10 days of recession curve, which ensures that the linearity of the equation is a good approximation (Chapman 1999), and a minimum determination coefficient value of 0.8 are recommended as criteria for calculating the groundwater recession constant. Even though a recession curve without any precipitation is desirable for adjusting the linear regression, it is not always possible to comply with this condition. A longer recession curve might be preferred despite including some precipitation events, since the recession curve tends to be steeper at the beginning of the recession.

Once the slope of the linear regression equation (α) is calculated, the recession constant as included in the baseflow filter ($alpha$) can be derived following Equation (5) (Appendix B, Snippet 6), thus having three values for this parameter (from the recession of three hydrograph peaks) that can be used as a reference to apply the baseflow filter. At this point, the R code creates the file `3_alpha_estimation.csv`, which stores the estimated $alpha$ values and the information collected to address this working step (e.g., recession duration, peak range, etc.). Figure 3 resumes the process to estimate the groundwater recession constant.

Once an estimated value of $alpha$ has been obtained, the baseflow filter can be applied. Different values of $alpha$ and BFI_{max} should be tried until reaching a realistic baseflow separation (code in Appendix B, Snippet 6). In order to know how these parameters work, their effects on the baseflow estimation have been assessed (Figure 4, animated versions of this plot can be found in the GitHub repository). Both parameters are noticeably sensitive to the baseflow component separation. BFI_{max} directly affects the amount of baseflow that the filter estimates, as it is the maximum baseflow ratio expected (Figure 4, left). $alpha$ controls the response of baseflow to episodes of flood or recession. Higher values of $alpha$ result in a lower sensitivity of baseflow to those changes (smoother fluctuations), having also an impact on the total amount of baseflow.

Both parameters are closely related to the lithology: basins with aquifer systems will have a higher contribution of groundwater to streamflow than impervious basins, and therefore should have higher BFI_{max} values. Permeable basins with higher aquifer-river connectivity (i.e., sandy aquifers) will be more sensitive to recharge-discharge episodes than basins with less

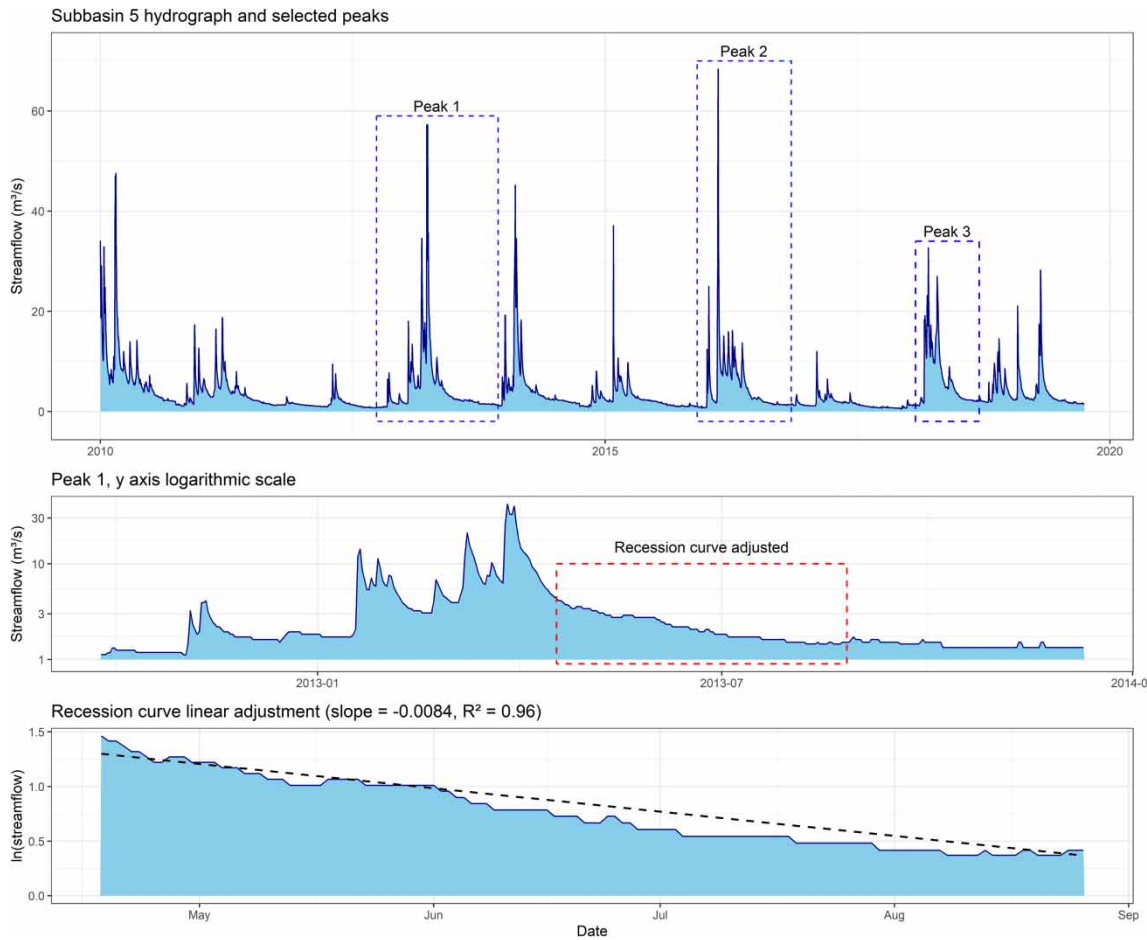


Figure 3 | Peaks selection, groundwater recession curve identification, and linear adjustment performed for subbasin 5 of the study case presented.

permeable substrates (e.g., clayey aquifers) and therefore the α value, which is related to the transmissivity, will be lower. Accordingly, the properties of a basin must be considered before setting the parameters of the baseflow filter, as for example reported by Kang *et al.* (2022).

The baseflow separation is computed for the entire observed data period (Appendix B, Snippet 6), but it is evaluated individually for each peak defined during the α estimation. A realistic separation should be achieved for the three peaks, ensuring the suitability of the selected parameters. Although it would vary according to the basin characteristics, an appropriate baseflow separation may be similar to the one presented in Figure 5, where baseflow fraction regarding total streamflow is maximum during dry periods and minimum when there are precipitation events. Precipitation is plotted above the baseflow separation, which helps to perform and evaluate it.

Once a realistic baseflow separation has been achieved, the contribution of groundwater to the streamflow (Bf_c) for the evaluated period is calculated by the code (Appendix B, Snippet 6) using the following equation:

$$Bf_c = \frac{\sum_1^n bf_k}{\sum_1^n (bf_k + m_k)} \quad (6)$$

where n is the number of days for which the filter is applied, bf_k is the baseflow contribution for the day k and m_k is the runoff contribution for the day k (units equal to the observed data). Then, the R code creates the `4_groundwater_results.csv` file, which stores the values of the parameters used in the filter and the calculated baseflow contribution.

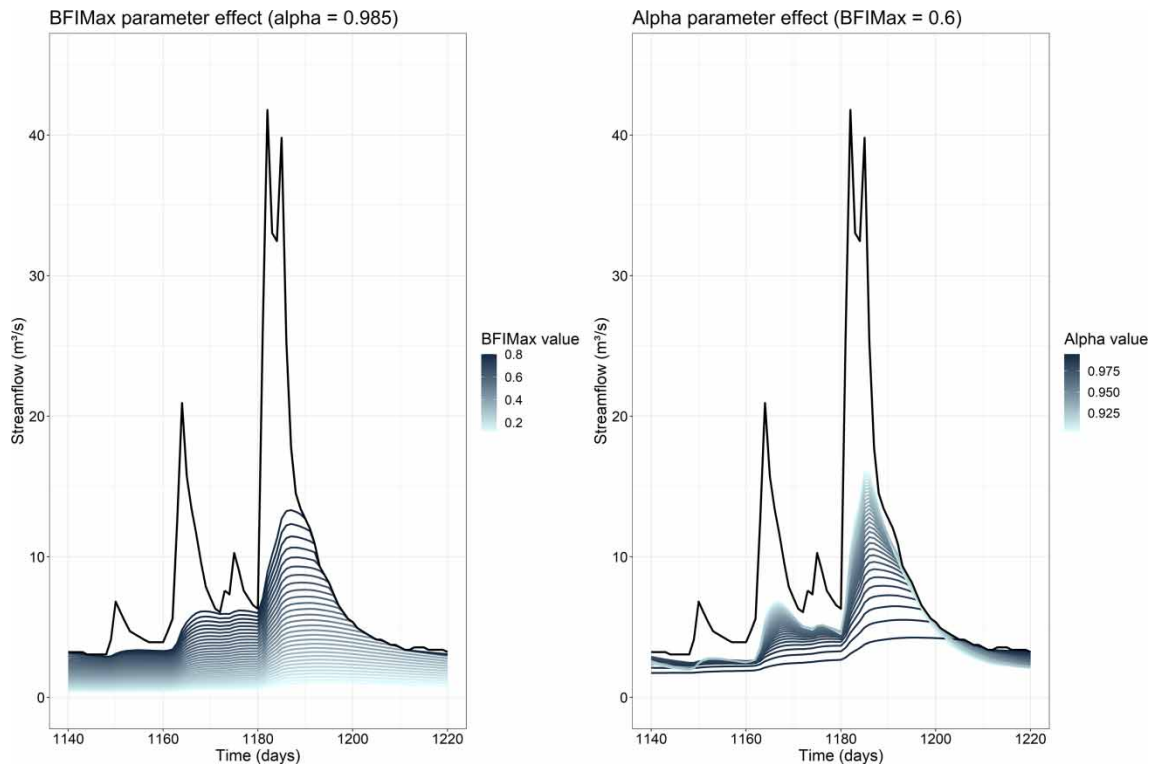


Figure 4 | Effect of the filter parameters on the baseflow separation.

3. RESULTS AND DISCUSSION

This section will first present and briefly discuss the results of applying the methodology developed in R to the case study selected, in order to prove its usefulness and soundness. Then, the novelty and contribution to the hydrological modelling sphere of this work are discussed.

3.1. Study case application

3.1.1. Runoff coefficients

Runoff coefficients were calculated for 19 subbasins of the UTRB for the period 2010–2018 (with some exceptions for subbasins 2, 6, and 14), and both average and annual values were evaluated considering the precipitation, temperature, and geological region. Appendix A (Output file 2) contains the mean, maximum and minimum runoff coefficient values obtained for each subbasin, and the mean precipitation and temperature for the time series.

The variability obtained in the runoff coefficient values highlights the importance of obtaining and analysing soft data before calibrating a hydrological model. The values for the mean runoff coefficient for the different subbasins of the UTRB ranged from 47% in subbasin 2 to 2% in subbasin 13. Within the subbasins, the runoff coefficient also varies greatly from year to year (e.g., the maximum value for subbasin 10 is 15 times higher than the minimum value, Appendix A 2), being thus necessary to evaluate a period representative of the weather conditions. As expected, the runoff coefficient is related to the mean precipitation (less precipitation led to lower values of the runoff coefficient), to the mean temperature (higher temperatures result in higher evapotranspiration rates and lower runoff coefficients), and also to the geological region (Appendix A 2). However, in some years and subbasins, the runoff coefficient seems to be also influenced by other factors. Table 1 presents the average runoff coefficient per geological region calculated from the full results (Appendix A2).

The influence of lithology becomes evident when aggregating to the region scale. The runoff coefficients are highest in the IMP and CRB regions, where not only precipitation and temperature are more favourable for generating runoff, but also the lithology and the mountainous topography. It should be noted that the subbasins selected in the CRB region are closely located (Figure 1) and thus have a similar climate. Therefore, the runoff coefficient varies less than within the IMP region,

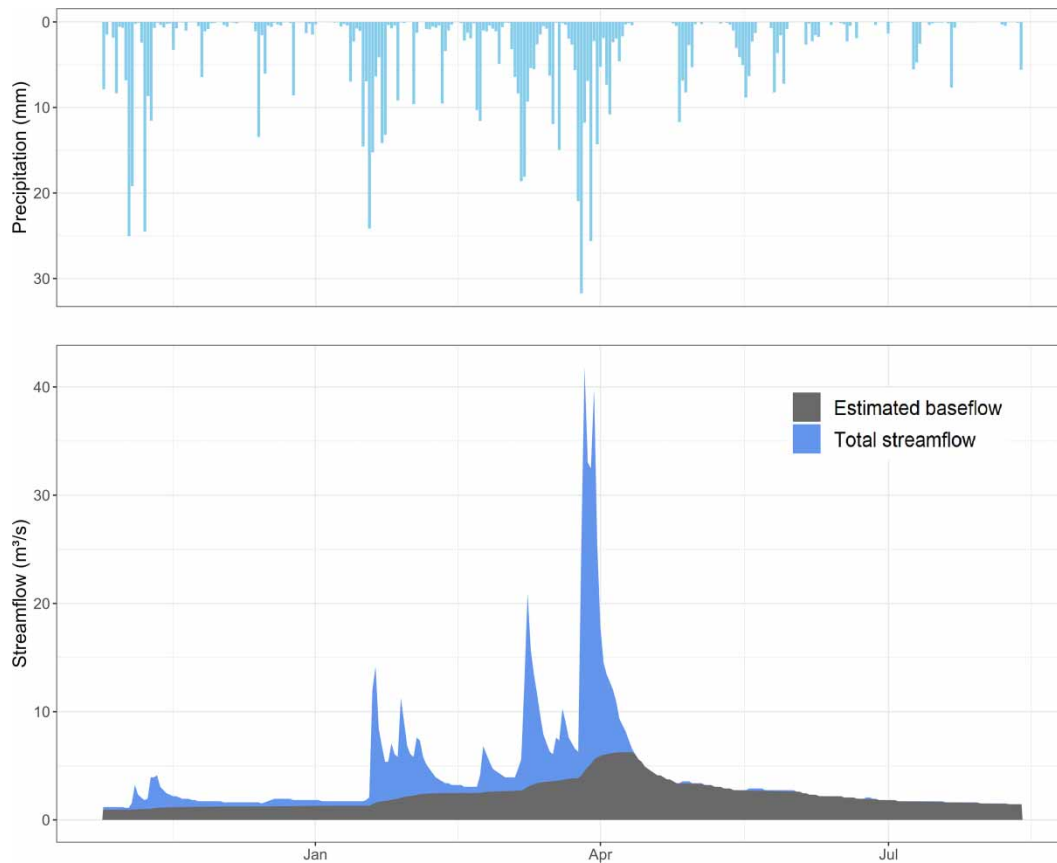


Figure 5 | Baseflow separation example performed for the peak 1 of the subbasin 5.

Table 1 | Average temperature, precipitation, and average, minimum and maximum runoff coefficients for the geological regions

Region	Mean temperature (°C)	Mean precipitation (mm)	Mean runoff coefficient	Min runoff coefficient	Max runoff coefficient
IMP	11.6	799	0.33	0.14	0.47
CRB	10.2	762	0.40	0.37	0.45
DTH	14.1	503	0.07	0.05	0.09
DTL	13.0	483	0.04	0.02	0.08
MIX	11.9	632	0.15	0.03	0.36

where subbasin 3 is located in a more arid and warmer region than subbasins 1 and 2 (Appendix A2). The DTH and DTL regions are flatter and warmer than the CRB region, and evapotranspiration is favoured under these conditions (Custodio & Llamas 1983), resulting in runoff coefficients of less than 10%. Even though some areas of the DTL region are composed of medium permeability carbonate and detrital materials (Appendix C), the runoff coefficient in this region is lower than in the DTH region, indicating a high influence of the low permeability materials, which in combination with a flat topography, low precipitation, and warmer temperatures, leads to very high evapotranspiration. Due to the flat topography, it would be possible that a small amount of the recharged water is released to streams downstream of the gauging stations, but it is not expected to be a significant amount.

As expected, the subbasins used to characterize regions with a mixed lithology (MIX) yielded the widest range of runoff coefficient values. In line with the results discussed earlier, mixed lithology subbasins consisting of igneous, metamorphic, or carbonated materials had higher runoff coefficient values, while lower values were obtained for the mixed subbasins where detrital materials with low permeability are present (e.g., subbasins 17 or 18, Appendix A 2).

3.1.2. Groundwater contribution analysis

Table 2 summarizes the data and results of the process for estimating the *alpha* filter parameter in the different geological regions. This includes the average duration of the recession curve, the average coefficient of determination, the average α , and the average *alpha* value and standard deviation. Results for the individual subbasins can be found in Appendix A (Output file 3).

The lithology had a noticeable influence on the recession curves seen in the hydrographs for the different geological regions (Appendix D). The recession curve duration is directly proportional to the aquifer properties of the subbasin. Subbasins with a carbonate lithology had an average recession curve duration that was three times longer than the recession curve duration of the IMP or DTH regions, which can also be observed when comparing the hydrographs (Appendix D). Since some of the subbasins in the DTL region include areas with a carbonate lithology, the duration of the recession curve was also longer (due to the CRB carbonate lithology areas). In the CRB subbasins, the end point of the recession curves was chosen when a precipitation event caused a relevant increase in the streamflow, while in the IMP or DTL region, this point was generally chosen when the streamflow ceased. Due to the higher transmissivity of the DTH aquifers and its smaller storage capacity compared to the CRB subbasins, groundwater flow is lower in these subbasins, and they are characterized by a smaller and faster recession process that does not always maintain streamflow throughout the entire dry season. In the IMP region, groundwater was not expected to be as relevant as in the other substrates, and accordingly, the expected recession time was shorter. However, other hydrological processes (snowmelt, precipitation frequency) in subbasins 1 and 2 of this region might significantly impact the recession curve.

The average values for the coefficient of determination of the recession curves selected were in general higher than 0.9, which indicates that this selection was appropriate. The subbasins with the highest determination coefficients are characterized by more homogeneous recession processes, and the recession curve selection was easier (e.g., in the CRB or IMP regions).

The obtained *alpha* values are related to the recession curve duration and magnitude. They generally matched our expectations: the CRB region yielded the highest *alpha* (and therefore the longest recession period), while the lowest value was obtained for the IMP region. As discussed earlier, some other processes occurring in the IMP region might be extending the recession period, and therefore the obtained *alpha* values are similar to the DTH values. The standard deviation of *alpha* (Table 2) reveals the variability in the recession of the subbasins within a region. It was the largest in the heterogeneous MIX and DTL regions, and smallest in the CRB region, which indicates that the recession process is similar in its subbasins.

The BFI_{max} values were initially estimated based on the recommended values by Eckhardt (2005), together with the knowledge of the research team on the study area. However, these recommended values led in some cases to an overestimation of baseflow (although they could be appropriate for other regions), and were then tuned by using the filter towards achieving a realistic hydrograph separation by visual inspection following the recommendations in the literature (e.g., Custodio & Llamas 1983). The final selected values for *alpha* (Table 3) slightly differed in some cases from the estimated ones (Table 2), as a realistic streamflow components separation for the three evaluated peaks was preferred than using the exact estimated values.

Table 3 shows the average values of the parameters used for the different regions and the value of the baseflow index estimated with the baseflow filter. Appendix A3 contains this information for each of the subbasins.

Subbasins located in geological regions with low permeability were expected to have a lower baseflow index, while those located in regions with relevant aquifers were expected to have higher. A baseflow index of 26% was estimated for the IMP region (Table 3). This value might be considered high for impervious basins, but based on the knowledge about the subbasins and the assessment performed, it seems reasonable. Some indicators, such as the presence of flow during dry periods or

Table 2 | Results at region scale of the groundwater recession curves linear adjustment

Region	Average recession curve time (days)	Mean coefficient of determination	Mean groundwater recession constant (α)	Mean <i>alpha</i> value	<i>Alpha</i> standard deviation
IMP	48	0.956	-0.040	0.961	0.014
CRB	144	0.927	-0.009	0.991	0.004
DTH	49	0.910	-0.037	0.964	0.012
DTL	76	0.897	-0.025	0.976	0.025
MIX	70	0.930	-0.028	0.973	0.025

Table 3 | Average parameter values used in the baseflow filter and groundwater indices estimated at region scale

Region	Mean α used	Mean BFI_{max} used	Estimated baseflow index	Baseflow index standard deviation
IMP	0.988	0.35	0.26	0.04
CRB	0.996	0.65	0.49	0.07
DTH	0.986	0.52	0.42	0.04
DTL	0.995	0.55	0.43	0.11
MIX	0.988	0.42	0.37	0.12

phreatophyte vegetation suggest a relevant groundwater contribution, and these indicators have been observed in some of the subbasins characterized as impervious in this work (Martín-Loeches *et al.* 2020).

As expected, the CRB region had the highest baseflow index (Table 3): it is dominated by a carbonate geology and its climate is colder and more humid. Around 50% of the streamflow was estimated as groundwater contribution, which is in agreement with previous studies (Sánchez-Gómez *et al.* 2022) and considered as a realistic value taking into account the properties of the region. The baseflow index is higher in some of the studied subbasins, e.g., subbasin 6, where a groundwater contribution of 55% was estimated (Appendix A 3).

Despite their different permeability, the groundwater contribution obtained in the two detrital regions was similar (Table 3). As mentioned in section 2.3, this could be because the subbasins considered as representative of the DTL region are actually composed of mixed materials (with relevant areas of carbonate and detrital materials with medium and high permeability). Runoff coefficients obtained for the DTL region are noticeably lower (Table 2), pointing to a very high evapotranspiration in the low permeability areas (percolation is reduced and more water is therefore available for evapotranspiration). In contrast, in the permeable areas of the DTL subbasins, a larger fraction of water infiltrates, recharges the groundwater and becomes streamflow. In consequence, groundwater contribution in the DTL region is higher than could be expected considering its lithology, since the permeable areas might contribute more to the streamflow than the low permeability ones. The hydrographs of the subbasins within the DTL region reveal this behaviour, since baseflow is maintained throughout the year (Appendix D).

In the subbasins located in mixed lithology regions, the average baseflow index was 37%. Four of the subbasins representative of this region (subbasins 16–19, Appendix C) are mostly composed of carbonate materials from medium to high permeability (Figure 1), while only two of the subbasins have impervious igneous and metamorphic materials.

The standard deviation of the estimated baseflow rate among subbasins located in the same geological region was highest in the MIX region (0.12), followed by the DTL region (0.12), where more different subbasins are grouped, while it is smaller in CRB (0.07) and in IMP and DTH regions (0.04).

Applying a baseflow filter such as the one used in this workflow requires some considerations that are worth mentioning. As explained earlier, although α values were estimated in the methodology, the used values slightly differed in some cases, as a realistic baseflow separation was prioritized. An α value higher than 0.97, regardless of the estimated one, is recommended to obtain a realistic baseflow separation. The influence on the α estimation of factors such as snowmelt, precipitation events, reservoirs, or withdrawals or releases related to human activities, should be analysed and corrected if possible. Regarding data reliability, when applying the filter, the amount of streamflow in 1 day affects the baseflow calculation in the following days. Therefore, if an anomalous value of streamflow is recorded (i.e., a very low or zero value in between higher values), we recommend correcting this value to achieve realistic estimates of baseflow. The data used for this work had only very few anomalous values that were corrected before applying the filter. Lastly, the time scale used to evaluate the separation of the hydrograph components can also have a significant impact on the baseflow index estimation. Figure 6 compares how a peak is represented when looking at it within an entire time series and when looking at it individually. Longer time series including several peaks might lead to a distorted visual perception of the baseflow separation that might make the user think that the groundwater contribution is larger than it actually is. Thus, the proposed separation procedure based on evaluating several hydrograph peaks individually is strongly recommended.

3.2. Contribution of the methodology presented to hydrological modelling

The methodology presented in this manuscript provides an R tool to estimate both the runoff coefficient and the baseflow index for one or several basins, indices that can be used in a subsequent soft calibration process to guarantee achieving

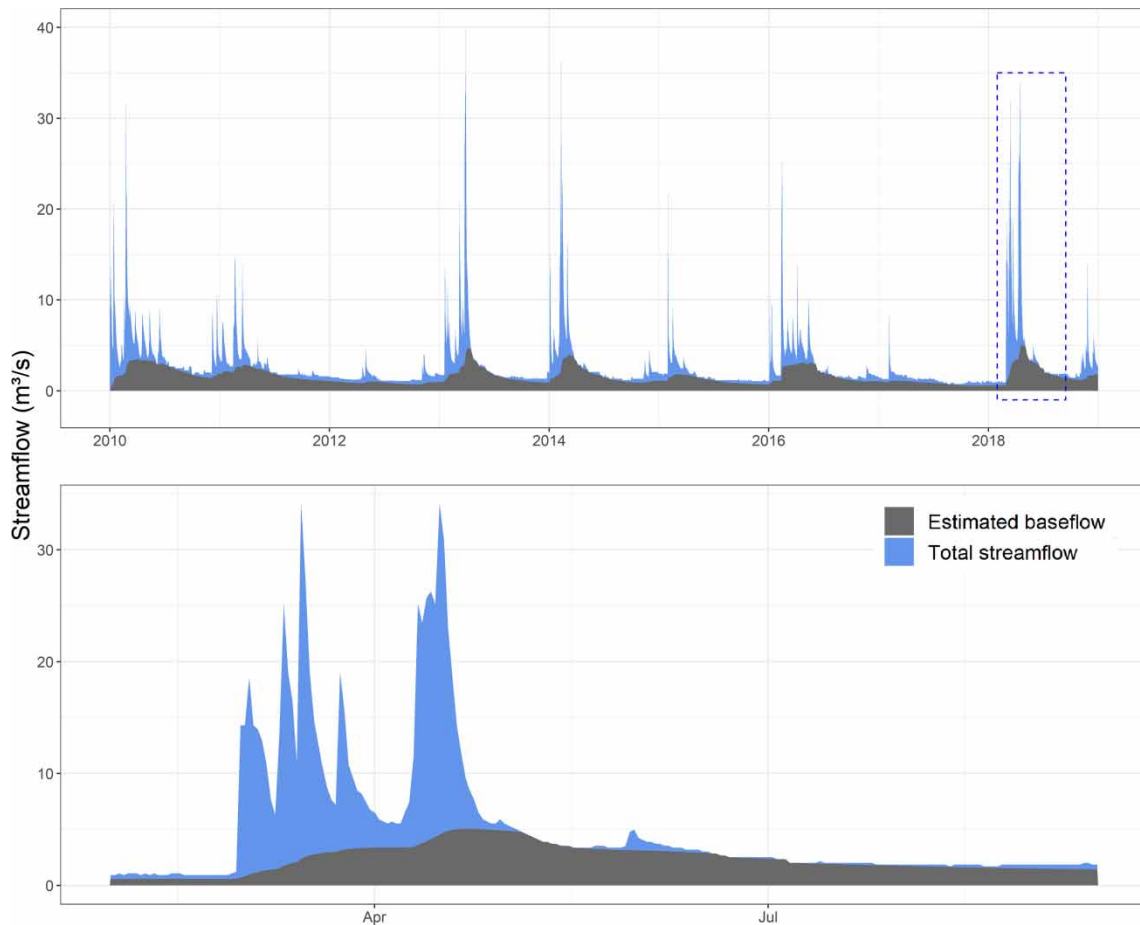


Figure 6 | Time-scale effect on the visual perception of the baseflow separation performed with the filter.

realistic simulations (Pfannerstill *et al.* 2017). Besides, it has been tested in 19 subbasins of a Spanish catchment, revealing the variability in those indices that might exist within a similar catchment, and the importance to consider that variability if accurate modelling wants to be achieved.

First, the methodology allows to automatically calculate the runoff coefficient for a number of basins using gridded precipitation data and streamflow records. This facilitates modellers' work, since the runoff coefficient calculation can be tedious, involving data interpolation, aggregation, and units' conversion. The data sources used for the study case presented as an example allow us to fully reproduce the methodology in Spain. However, with some code modifications, it can be applied to any other place with gridded precipitation data, which is often available (e.g., Molina-Navarro *et al.* 2017; Blankenau *et al.* 2020).

Regarding the baseflow contribution to the streamflow, several baseflow filters have already been developed by other authors (Arnold & Allen 1999; Koffler *et al.* 2022). These tools are fully automated, easy to apply, but this involves some simplifications that can lead to a separation of the hydrograph yielding a baseflow contribution that might not fully match with the real one. This could make modellers guide a subsequent calibration procedure based on inaccurate values, particularly if they do not have a strong hydrological background and/or they do not have good knowledge of the study area modelled. With the methodology presented in this manuscript, we propose a semi-automatic, supervised application of a baseflow digital filter, with a previous step of hydrograph recession curves analysis to get a realistic value of the groundwater recession constant (a parameter needed by the filter) before its application.

We provide guidance on how to address this workflow in R, the same working environment as for the runoff coefficient calculation, without the need for installing new software or converting the data. Although the proposed methodology addresses the hydrograph separation after analysing three hydrograph peaks (and their respective recession curves), it still

might not guarantee a fully accurate baseflow separation (e.g., baseflow underestimated towards the end of long low flows periods or slightly overestimated during peaks, Figure 5) and for some cases, a compromise is needed to deal with these inaccuracies. These limitations have also been reported by other authors (Kang *et al.* 2022) and should be kept in mind when applying digital filters. In addition to the methodology applicability, the theoretical description of the filter presented in the manuscript shed some light on the understanding of the equation governing those filters, particularly towards the different definitions of the groundwater recession constant used in the literature. Collaterally, the guided calculation of this constant provides values that can be used in a subsequent calibration of models such as SWAT (notice that 'alpha' in SWAT is equivalent to α is the classic groundwater recession constant, see Equation (5)).

Finally, the large differences between subbasins and geological regions obtained for the two soft indices analysed in the study case (Tables 1 and 3) reveal the need to take into account the regional differences in these indices before addressing a hydrological model calibration if an accurate, realistic, and robust model is to be achieved. However, unfortunately, very often a same parameter set is used of an entire catchment, despite being heterogeneous (Efstratiadis & Koutsoyiannis 2010). The use of the methodology proposed in this manuscript can help to obtain values of these indices at a regional level, thus facilitating a zonal parameterization of a subsequent calibration procedure and contributing to a more realistic and reliable hydrological modelling.

4. CONCLUSIONS

This work presents HydRoVars, an R coded tool which allows us to collect and analyse weather and hydrological variables, specifically the runoff coefficient and the baseflow index, for a single basin or a group of them. Thus, it facilitates the step of data collection prior to a soft calibration, a recommended procedure for reducing equifinality and thus the probability of obtaining a statistically satisfactory but unrealistic model.

The tool automatically calculates basin-specific runoff coefficients (annual and average values) based on rainfall and streamflow data. The baseflow index is estimated through a supervised, semi-automatic application of a digital baseflow filter. The code includes a previous step to estimate one of the filter parameters (*alpha*) from the hydrograph recession curve, and the manuscript provides guidance to address it, thus ensuring a realistic and successful application of the filter to finally obtain a baseflow index value. Collection of weather data (precipitation and temperature) at varied temporal scales is also included, which allows us to characterize the climate conditions of the studied area.

To demonstrate the usefulness of this methodology, these two indices were estimated in 19 subbasins of the UTRB (Spain), located in four different geological regions. Besides testing the methodology, the work resulted in new insights about the hydrological properties of the basin, revealing large differences in both indices within the subbasins analysed. This in turn proves the need of taking into account the regional differences in these indices before addressing a hydrological model calibration if a realistic model is to be achieved.

The datasets used in the case study provided allow us to replicate the methodology in the entire Spanish territory, but with small code adjustments, it can be applied to any other region with available streamflow and precipitation data (preferably gridded). Besides the code, all the data of the case study are available in the repository for readers to reproduce the work presented. This methodology, computed in R environment for both indices, might contribute to towards realistic studies in the hydrological modelling community, particularly in areas with large climatic or geological heterogeneity.

ACKNOWLEDGEMENTS

Alejandro Sánchez-Gómez received support from the University of Alcalá (UAH) PhD Fellowships Program. This study was also supported by the TwinTagus Project, PID2021-128126OA-I00, funded by MCIN/AEI/10.13039/501100011033 and FEDER (EU), the Department of Education, Culture and Sports of Castilla-La Mancha Government and the European Union (IMPACT Project, SBPLY/21/180225/000092) and the Department of Economy, Business and Employment of Castilla-La Mancha Government (CEAGU, 2023/00029/001). The authors want to thank AEMET for the gridded weather datasets and the Water Resources Management and Planning Research Group (particularly Dr Javier Senent Aparicio) from the Catholic University of Murcia for adapting the weather datasets format and make it available (<https://swat.tamu.edu/data/spain/>). The authors want to thank CEDEX for making available the gauging datasets. Authors also want to acknowledge Lupe, Germán, Antonio and Carlos (UAH Physics Department), for their help understanding the mathematics behind the baseflow filter equations.

DATA AVAILABILITY STATEMENT

All the data and code used for this work is available in the following GitHub repository: <https://github.com/alejandrosz/HydRoVars>, where detailed explanations about its installation and functioning are included. Weather datasets used in this work can be downloaded here: <https://swat.tamu.edu/data/spain/>. Gauging datasets used in this work can be downloaded here: <https://ceh.cedex.es/anuarioaforos/demarcaciones.asp>.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

- Acero Triana, J. S., Chu, M. L., Guzman, J. A., Moriasi, D. N. & Steiner, J. L. 2019 Beyond model metrics: The perils of calibrating hydrologic models. *Journal of Hydrology* **578**, 124032.
- Arnold, J. G. & Allen, P. M. 1999 Automated methods for estimating baseflow and ground water recharge from streamflow records. *Journal of the American Water Resources Association* **35** (2), 411–424.
- Arnold, J. G., Youssef, M. A., Yen, H., White, M. J., Sheshukov, A. Y., Sadeghi, A. M., Moriasi, D. N., Steiner, J. L., Amatya, D. M., Skaggs, R. W., Haney, E. B., Jeong, J., Arabi, M. & Gowda, P. H. 2015 Hydrological processes and model representation: Impact of soft data on calibration. *Transactions of the ASABE* **58** (6), 1637–1660.
- Bahreman, A. 2016 HESS opinions: Advocating process modeling and de-emphasizing parameter estimation. *Hydrology and Earth System Sciences* **20** (4), 1433–1445.
- Bednar-Friedl, B., Biesbroek, R., Schmidt, D. N., Alexander, P., Børsheim, K. Y., Carnicer, J., Georgopoulou, E., Haasnoot, M., Le Cozannet, G., Lionello, P., Lipka, O., Möllmann, C., Muccione, V., Mustonen, T., Piepenburg, D. & Whitmarsh, L. 2022 Europe. In: *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Pörtner, H. O., ed.). Cambridge University Press, Cambridge, UK. New York, USA, pp. 1817–1927.
- Blankenau, P. A., Kilic, A. & Allen, R. 2020 An evaluation of gridded weather data sets for the purpose of estimating reference evapotranspiration in the United States. *Agricultural Water Management* **242**, 106376.
- Blume, T., Zehe, E. & Bronstert, A. 2007 Rainfall-Runoff response, event-based runoff coefficients and hydrograph separation. *Hydrological Sciences Journal* **52** (5), 843–862.
- CEDEX 2021 Anuario de Aforos. Centro de Estudios Hidrográficos. Available from: <https://ceh.cedex.es/anuarioaforos/default.asp> (accessed 15 December 2023).
- Chapman, T. 1999 A comparison of algorithms for stream flow recession and baseflow separation. *Hydrological Processes* **13** (5), 701–714.
- Chapman, T. & Maxwell, A. I. 1996 Baseflow separation-comparison of numerical methods with tracer experiments. In: *Hydrology and Water Resources Symposium 1996: Water and the Environment, Institution of Engineers*, Barton, Australia, pp. 539–545.
- Chawanda, C. J., Arnold, J., Thiery, W. & van Griensven, A. 2020 Mass balance calibration and reservoir representations for large-scale hydrological impact studies using SWAT+. *Climatic Change* **163** (3), 1307–1327.
- Chazarra Bernabé, A., Flórez García, E., Peraza Sánchez, B., Tohá Rebull, T., Lorenzo Mariño, B., Criado Pinto, E., Vicente Moreno, J. V., Romero Fresneda, R. & Botey Fullat, R. 2018 *Mapas Climáticos de España (1981–2010) y ETo (1996–2016)*. Agencia Estatal de Meteorología, Madrid, Spain.
- Custodio, E. & Llamas, M. R. 1983 *Hidrología Subterránea*. Omega, Barcelona, Spain.
- del Pozo Gómez, M. 2009 *Mapa Litoestratigráfico, de Permeabilidades e Hidrogeológico de España A Escala 1:200.000*. IGME, Madrid, Spain.
- Eckhardt, K. 2005 How to construct recursive digital filters for baseflow separation. *Hydrological Processes* **19** (2), 507–515.
- Efstratiadis, A. & Koutsoyiannis, D. 2010 One decade of multi-objective calibration approaches in hydrological modelling: A review. *Hydrological Sciences Journal* **55** (1), 58–78.
- García-Ruiz, J. M., López-Moreno, J. I., Vicente-Serrano, S. M., Lasanta-Martínez, T. & Beguería, S. 2011 Mediterranean water resources in a global change scenario. *Earth-Science Reviews* **105**, 121–139.
- Gassman, P. W., Sadeghi, A. M. & Srinivasan, R. 2014 Applications of the SWAT model special section: Overview and insights. *Journal of Environmental Quality* **43** (1), 1–8.
- Guse, B., Reusser, D. E. & Fohrer, N. 2014 How to improve the representation of hydrological processes in SWAT for a lowland catchment-temporal analysis of parameter sensitivity and model performance. *Hydrological Processes* **28** (4), 2651–2670.
- Kang, T., Lee, S., Lee, N. & Jin, Y. 2022 Baseflow separation using the digital filter method: Review and sensitivity analysis. *Water* **14** (3), 485.
- Kirkby, M., Bracken, L. & Reaney, S. 2002 The influence of land use, soils and topography on the delivery of hillslope runoff to channels in SE Spain. *Earth Surface Processes and Landforms* **27** (13), 1459–1473.
- Koffler, D., Gauster, T. & Laaha, G. 2022 *Lfstat: Calculation of Low Flow Statistics for Daily Stream Flow Data*. Available from: <https://cran.r-project.org/web/packages/lfstat/lfstat.pdf> (accessed 15 December 2023).
- López-Vera, F. 2012 Groundwater: The invisible resource. *International Journal of Water Resources Development* **28** (1), 141–150.

- Lyne, V. & Hollick, M. 1979 Stochastic time-variable rainfall-runoff modelling. In *Institute of Engineers Australia National Conference*, Barton, Australia, pp. 89–93.
- Maillet, E. T. 1905 *Essais d'hydraulique souterraine et fluviale*. Librairie Sci., A. Hermann, Paris, France.
- Martín-Loeches, M., Pavón-García, J., Molina-Navarro, E., Martínez-Santos, P., Almeida, C., Reyes-López, J., Cienfuegos-Hevia, I. & Sastre-Merlín, A. 2020 Hydrogeochemistry of granitic mountain zones and the influence of adjacent sedimentary basins at their tectonic borders: The case of the Spanish Central System Batholith. *Hydrogeology Journal* **28** (7), 2477–2500.
- Meaurio, M., Zabaleta, A., Uriarte, J. A., Srinivasan, R. & Antigüedad, I. 2015 Evaluation of SWAT models performance to simulate streamflow spatial origin. The case of a small forested watershed. *Journal of Hydrology* **525**, 326–334.
- Molina-Navarro, E., Hallack-Alegría, M., Martínez-Pérez, S., Ramírez-Hernández, J., Mungaray-Moctezuma, A. & Sastre-Merlín, A. 2016 Hydrological modeling and climate change impacts in an agricultural semiarid region. Case study: Guadalupe River Basin, Mexico. *Agricultural Water Management* **175**, 29–42.
- Molina-Navarro, E., Andersen, H. E., Nielsen, A., Thodsen, H. & Trolle, D. 2017 The impact of the objective function in multi-site and multi-variable calibration of the SWAT model. *Environmental Modelling & Software* **93**, 255–267.
- Muleta, M. K. 2012 Model performance sensitivity to objective function during automated calibrations. *Journal of Hydrologic Engineering* **17** (6), 756–767.
- Ortega, L., Manzano, M., Custodio, E., Hornero, J. & Rodríguez-Arévalo, J. 2015 Using ^{222}Rn to identify and quantify groundwater inflows to the Mundo River (SE Spain). *Chemical Geology* **395**, 67–79.
- Pebesma, E. 2018 Simple features for r: Standardized support for spatial vector data. *The R Journal* **10** (1), 439–446.
- Peral García, C., Navascués Fernández-Victorio, B. & Ramos Calzado, P. 2017 *Serie de Precipitación Diaria En Rejilla Con Fines Climáticos*. Agencia Estatal de Meteorología, Madrid, Spain.
- Pfannerstill, M., Bieger, K., Guse, B., Bosch, D. D., Fohrer, N. & Arnold, J. G. 2017 How to constrain multi-objective calibrations of the SWAT model using water balance components. *JAWRA Journal of the American Water Resources Association* **53** (3), 532–546.
- R Core Team 2022 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available from: <https://www.R-project.org/>.
- Rodríguez-Blanco, M. L., Taboada-Castro, M. M. & Taboada-Castro, M. T. 2012 Rainfall–runoff response and event-based runoff coefficients in a humid area (Northwest Spain). *Hydrological Sciences Journal* **57** (3), 445–459.
- Samper, J., Pisani, B. & Espinha Marques, J. 2015 Hydrological models of interflow in Three Iberian Mountain Basins. *Environmental Earth Sciences* **73** (6), 2645–2656.
- Sánchez Gómez, A. 2024 *HydRoVars: An R tool to collect hydrological variables*. GitHub Repository. Available from: <https://github.com/alejandrosz/HydRoVars>.
- Sánchez-Gómez, A., Martínez-Pérez, S., Pérez-Chavero, F. M. & Molina-Navarro, E. 2022 Optimization of a SWAT model by incorporating geological information through calibration strategies. *Optimization and Engineering* **23** (4), 2203–2233.
- Sánchez-Gómez, A., Martínez-Pérez, S., Leduc, S., Sastre-Merlín, A. & Molina-Navarro, E. 2023 Streamflow components and climate change: Lessons learnt and energy implications after hydrological modeling experiences in catchments with a Mediterranean climate. *Energy Reports* **9**, 277–291.
- Seibert, J. & McDonnell, J. J. 2002 On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. *Water Resources Research* **38** (11), 23–1–23–14.
- Senent-Aparicio, J., George, C. & Srinivasan, R. 2021a Introducing a new post-processing tool for the SWAT+ model to evaluate environmental flows. *Environmental Modelling & Software* **136**, 104944.
- Senent-Aparicio, J., Jimeno-Sáez, P., Cecilia, J. M. & Srinivasan, R. 2021b Impacts of swat weather generator statistics from high-resolution datasets on monthly streamflow simulation over Peninsular Spain. *Journal of Hydrology: Regional Studies* **35**, 100826.
- Wickham, H. 2022 *Tidyverse: Easily Install and Load the 'Tidyverse'* (version 1.3.2). Available from: <https://CRAN.R-project.org/package=tidyverse> (accessed 15 December 2023).
- WMO 2008 *Manual on Low-Flow Estimation and Prediction*. Operational Hydrology Report, no. 50. World Meteorological Organization, Geneva, Switzerland.
- Zhu, R., Zheng, H. & Liu, C. 2010 Estimation of groundwater residence time and recession rate in watershed of the loess plateau. *Journal of Geographical Sciences* **20** (2), 273–282.

First received 15 December 2023; accepted in revised form 1 April 2024. Available online 9 April 2024