Soft data collection for realistic hydrological modelling: a reproducible methodology developed in R for estimating the runoff coefficient and baseflow index.

Alejandro Sánchez Gómez[1], Katrin Bieger[2], Christoph Schürz[3], Silvia Martínez Pérez[1], José Manuel Rodríguez Castellanos[1], Hendrick Rathjens[4], Eugenio Molina Navarro[1]

[1]*Department of Geology, Geography and Environment, University of Alcalá (UAH), Spain*

[2] *Department of Ecoscience, Aarhus University, Denmark*

[3]*Department of Computational Landscape Ecology, Helmholtz Centre for Environmental Research (UFZ), Germany*

[4]*Stone Environmental, Inc.*

SWAT Conference 2023

Aarhus, Denmark

Universidad de Alcalá ACMA Agua, Clima y Medio Ambiente SWAT Soil & Water Assessment Tool AARHUS UNIVERSITY

Index

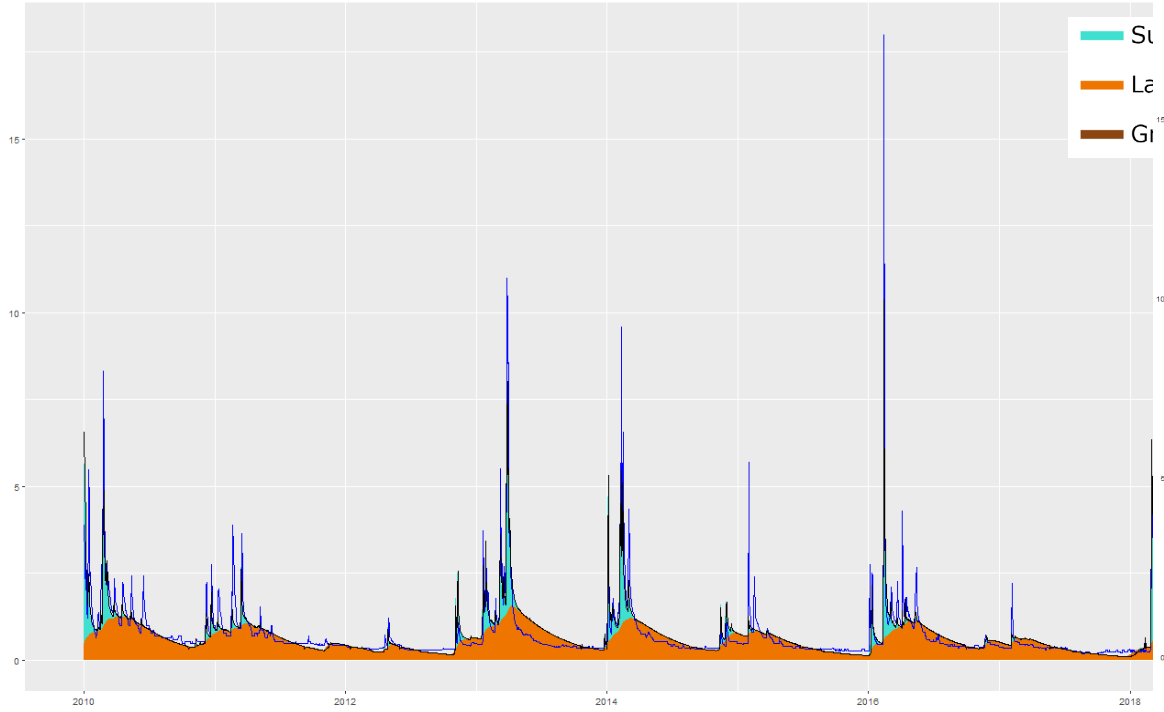
Introduction. Role of soft calibration in hydrological modelling works

Soft calibration → Ensure that a model reproduce variable (crop yields, nutrient loads, sediments, streamflow contribution) accordingly to a expected value.

Determine **suitable ranges for sensitive parameters on the water balance** in order to represent its variables realistically **before starting Hard calibration** (evapotranspiration rate, groundwater, etc.).

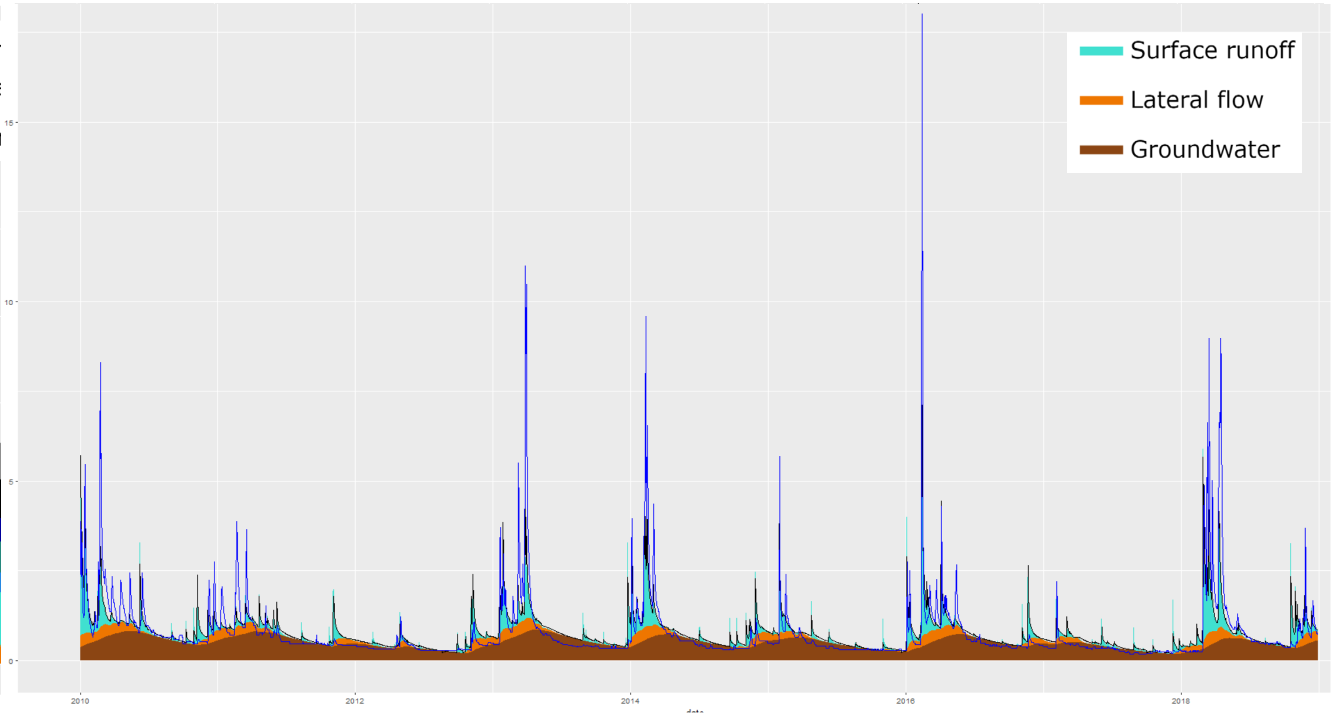
Example 1: Hard calibration

NSE	R^2	PBIAS
0.748	0.754	-0.8



Example 2: Soft calibration + Hard calibration

NSE	R^2	PBIAS
0.57	0.61	3.5



Introducción. *Soft data*

Soft data: *expert knowledge, estimations, measured data.*

- *Maximum historical streamflow record.*
- *Suitable range of values for a parameter (e.g., soil depth).*
- *Estimated amount of sediment.*

Soft data: *Key hydrological soft variable*

How can these variables be estimated?

- Runoff coefficient (Water yield / Precipitation).

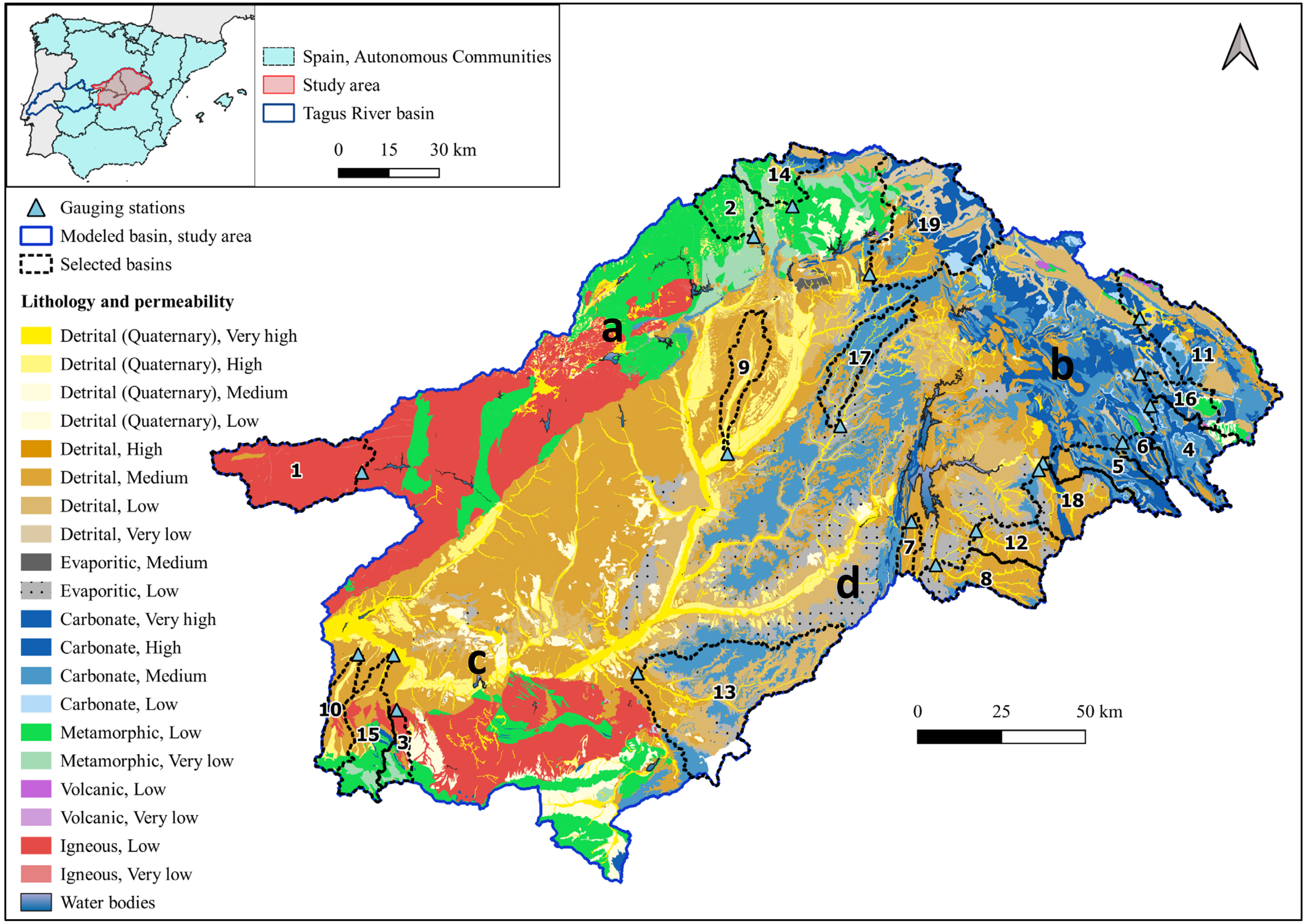
- What fraction of the precipitation volume becomes streamflow?
- What is evapotranspirated?

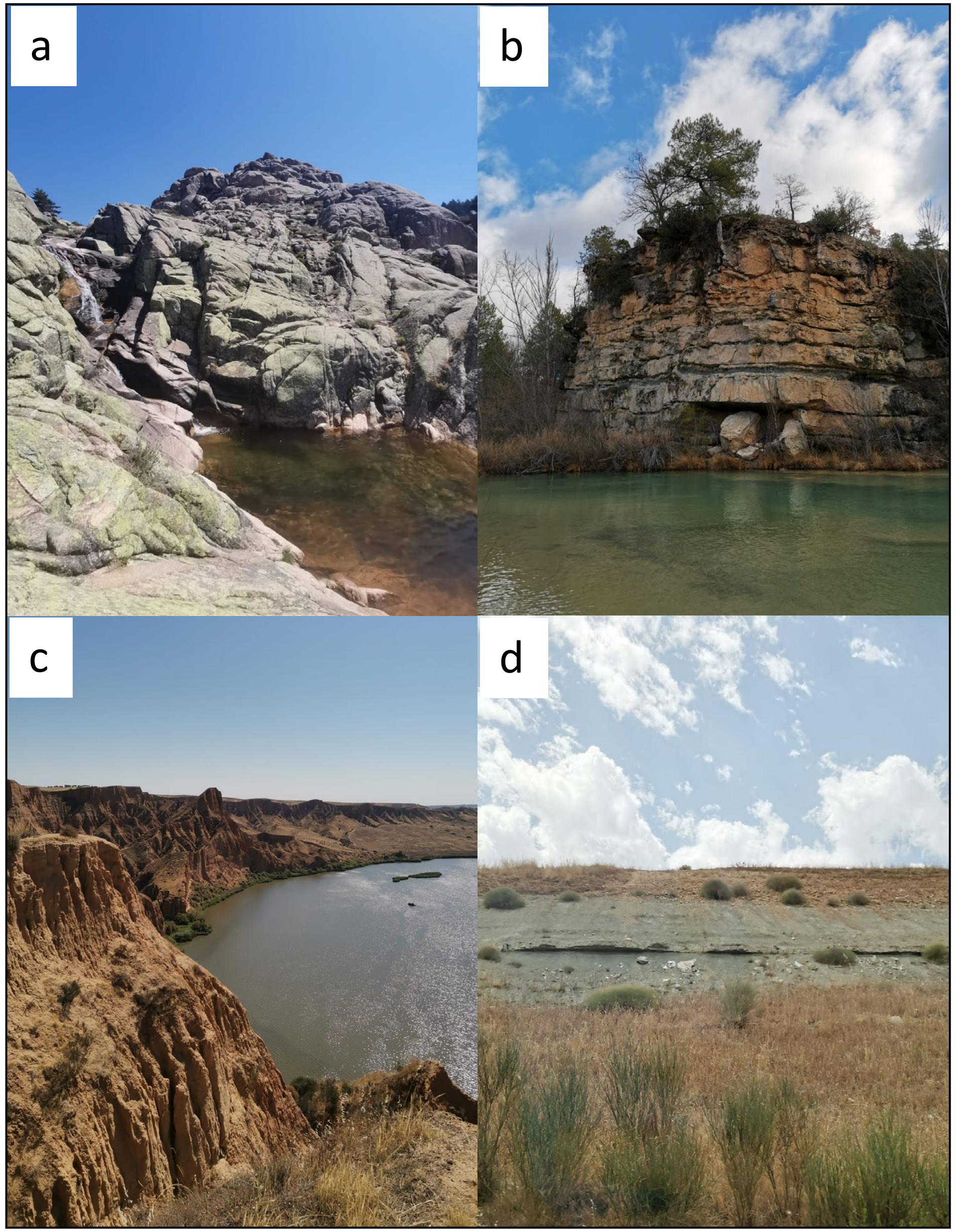
- Groundwater contribution to the streamflow (Groundwater Flow / Water yield).

What fraction of the streamflow comes from aquifers?

Introduction. Geological heterogeneity in the study area

Tagus River basin → Geological heterogeneity → Differences in hydrological processes → Soft data needed for **different geological regions.**





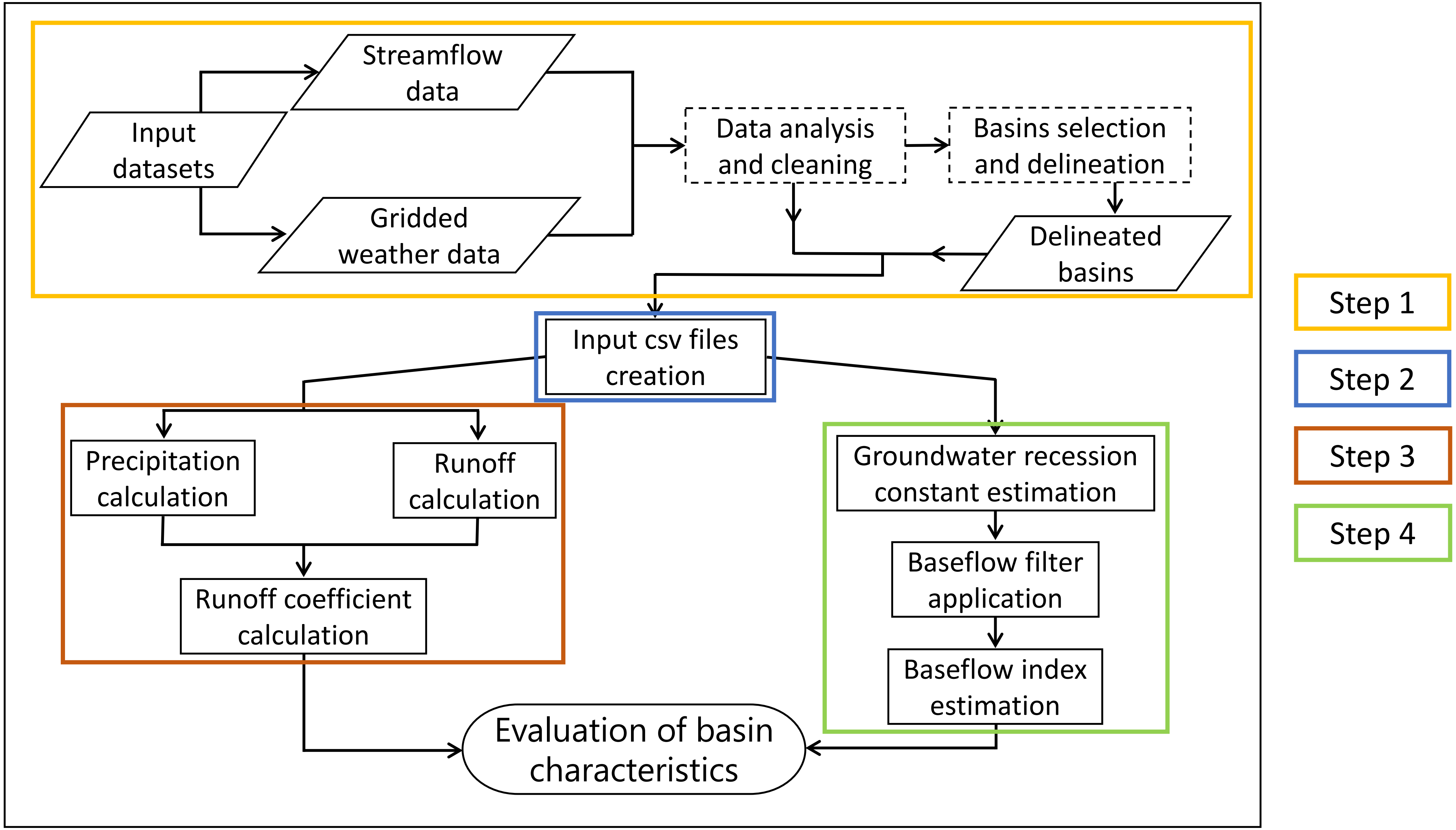
Objectives

- To develop and make available for SWAT users a methodology to estimate two key variables for a soft calibration process: runoff rate and groundwater contribution to streamflow.

- To present the Tagus River basin study case to highligth the importance of analysing these two indices.

Repository link
https://github.com/alejandrosgz/Soft_data_collection_methodology

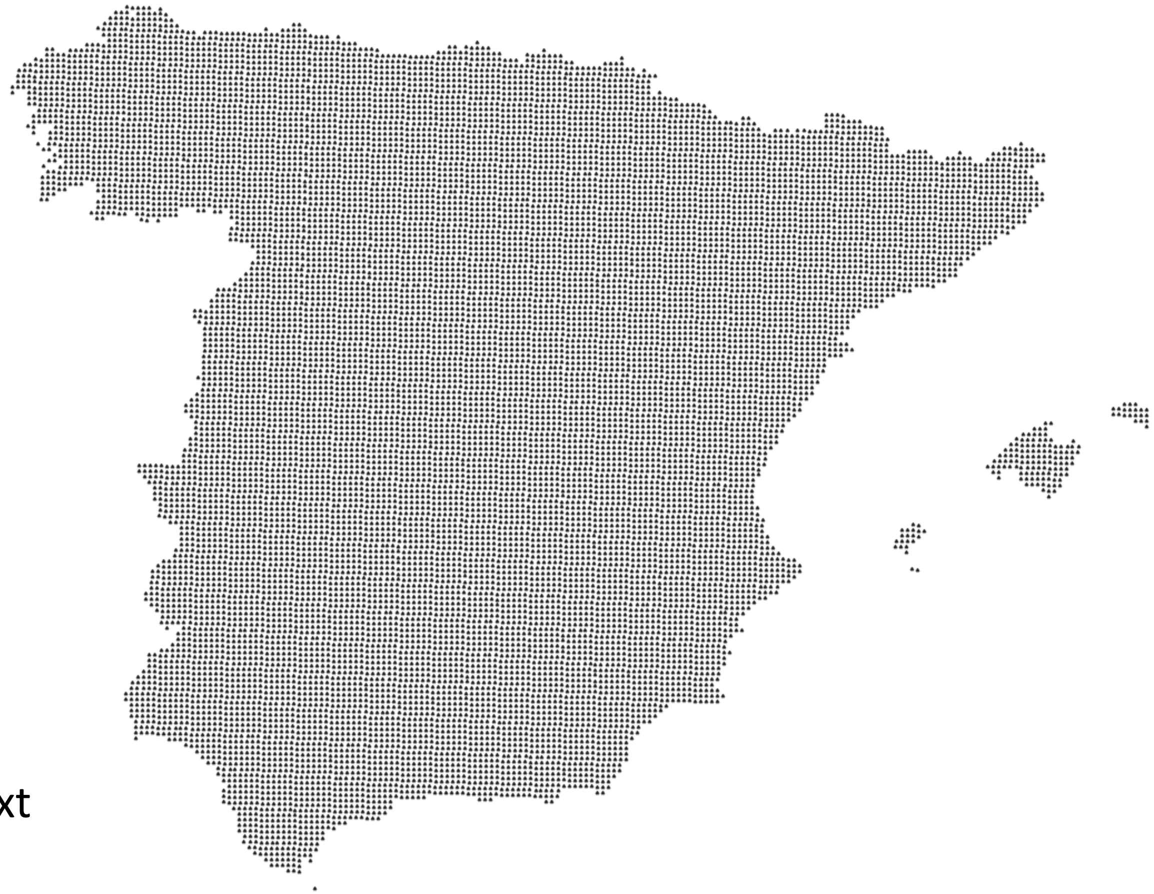
Methodology´s workflow



Step 1: Input data preparation

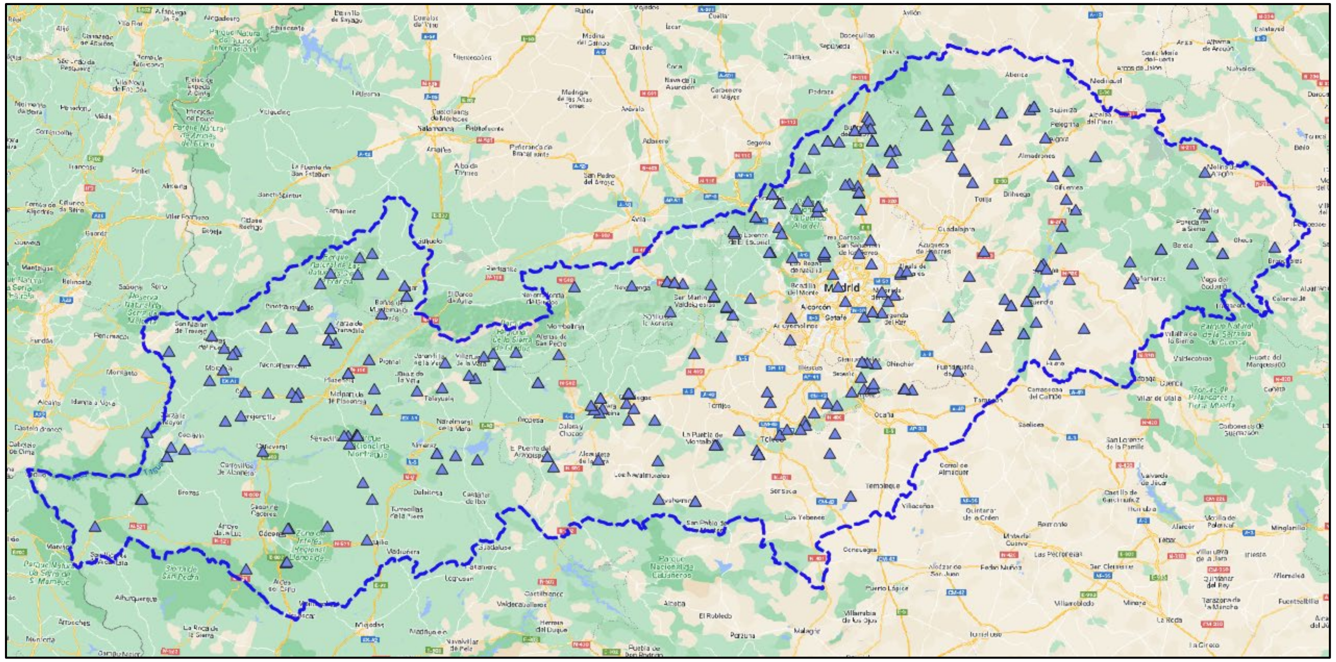
Gridded weather data:

- Gridded data with high resolution → Highly recommended (simplify interpolation).
- AEMET (Spanish meteorological agency) grid. 5 km resolution. Daily temperature (minimum and maximum) and precipitation data interpolated from real weather stations. Available serie: 1951-2021.
- Converted to SWAT 2012 format (Senent et al., 2021) and available at https://swat.tamu.edu/data/spain/. Txt files, one for each point of the grid.



Senent-Aparicio, J., Jimeno-Sáez, P., López-Ballesteros, A., Giménez, J. G., Pérez-Sánchez, J., Cecilia, J. M., & Srinivasan, R. (2021). Impacts of swat weather generator statistics from high-resolution datasets on monthly streamflow simulation over Peninsular Spain. *Journal of Hydrology: Regional Studies*, *35*, 100826.

Step 1: Input data preparation



Streamflow gauged data:

- Streamflow daily records (m^3/s).
- File with the data for all the gauging stations with an identifier for each gauging station.
- Used datasets: Downloaded from the gauged streamflow yearly report for the Tagus River basin https://ceh.cedex.es/anuarioaforos/demarcaciones.asp.
- Data availability must be checked (only the years with complete data must be used).

```
# A tibble: 2,399,462 x 4
   indroea fecha      altura caudal
     <int> <chr>       <dbl>  <dbl>
 1    3001 01/10/1945   0.46   1
 2    3001 02/10/1945   0.47   1.15
 3    3001 03/10/1945   0.46   1
 4    3001 04/10/1945   0.48   1.31
 5    3001 05/10/1945   0.48   1.31
 6    3001 06/10/1945   0.47   1.15
 7    3001 07/10/1945   0.48   1.31
 8    3001 08/10/1945   0.49   1.46
 9    3001 09/10/1945   0.49   1.46
10    3001 10/10/1945   0.5    1.62
# i 2,399,452 more rows
```

Basin ID	Basin	Period with available data	Complete data (2010-2018)
1	Navaluenga	1974-2019	Yes
2	Matallana	2010-2019	No, 273 days missing
3	Villarejo de Montalban	1969-2019	Yes
4	Peralejo de las Truchas	1945-2019	Yes
5	Priego Escabas	1912-2019	Yes
6	Santa Maria del Val	2010-2019	No, 273 days missing
7	Jabalera	1977-2019	Yes
8	Huete	1965-2019	Yes
9	Torote	1972-2019	Yes

Step 1: Input data preparation

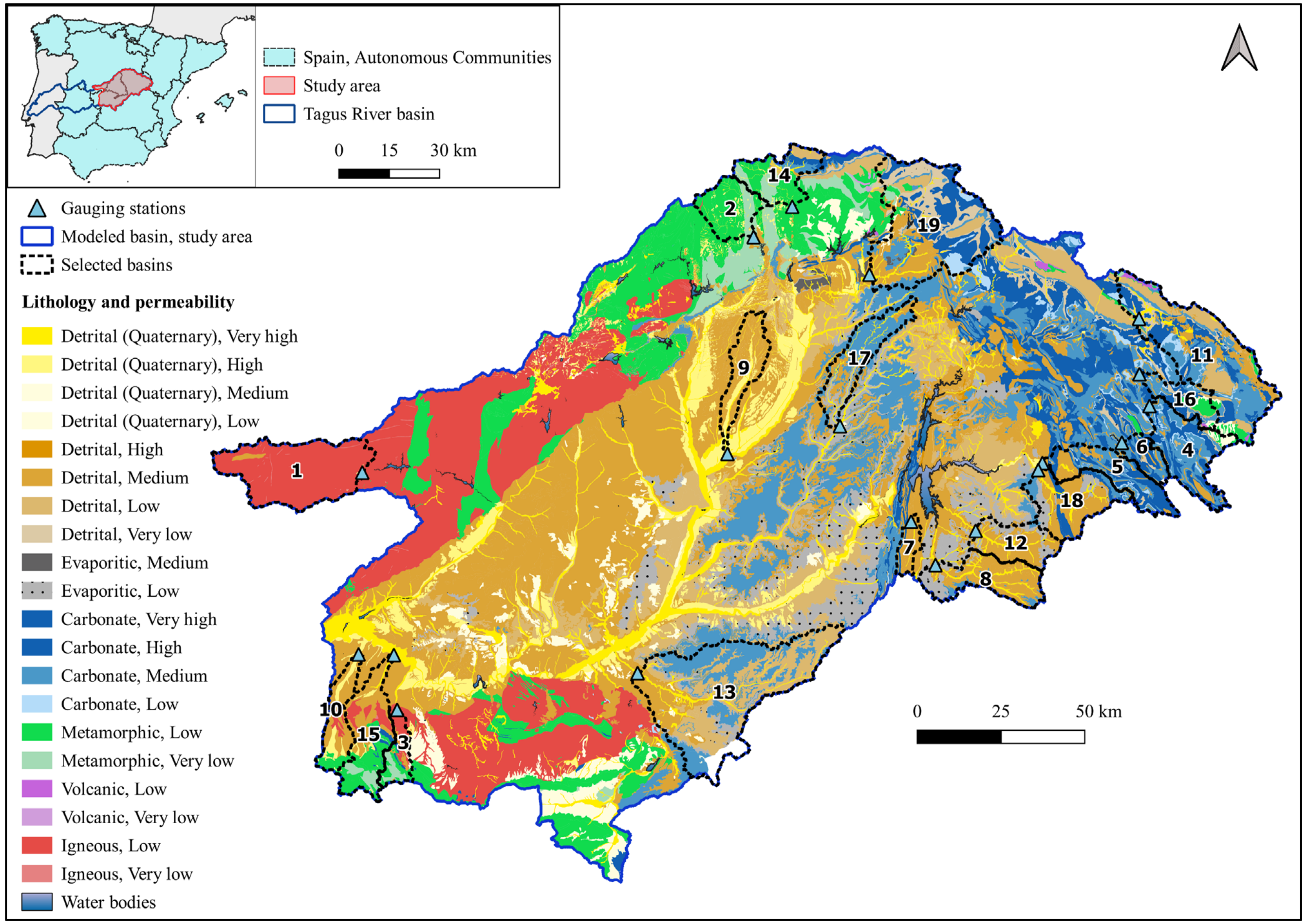
Subbasins´ selection and delineation

Selection criteria:

- Presence of a gauging station with enough available data.
- Underlying lithology (representative of the different geological regions).
- Undisturbed regime (absence of reservoirs or relevant water withdrawals).

Delineation:

- Performed using the GRASS module in QGIS (r.watershed and r.water.outlet tools). DEM required.



Step 2: Input files creation

Created from:

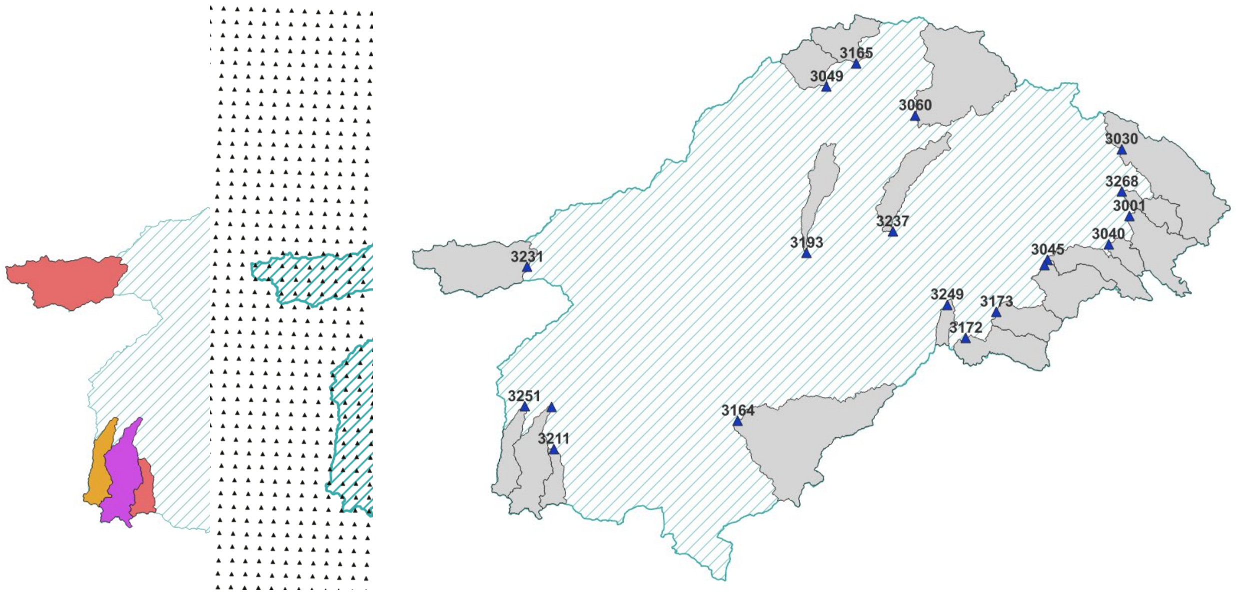
- Vector layer with all the delineated subbasins polygons.
- Vector layer with the location of all the weather points.
- Identifiers for the streamflow gauging stations to be used.

File 1: Subbasins data

- ID, Name, Area, Gauging station code and optional additional variable (geology groups in this case).

File 2: Weather data points to be used

- Identifier for each point of the weather grid that is located within a subbasin and have to be used to calculate precipitation or temperature.



	A	B	C	D	E
1	Basin	Basin_ID	area	gauging_code	region
2	Navaluenga	1	699307071.8	3231	IMP
3	Matallana	2	252264276.4	3049	IMP
4	Villarejo de N	3	136208984.8	3211	IMP
5	Peralejo de la	4	408557741.2	3001	CRB
6	Priego Escaba	5	329242825.5	3045	CRB
7	Santa Maria d	6	117788153	3040	CRB
8	Jabalera	7	85160052		
9	Huete	8	35942651		
10	Torote	9	25466816		
11	La Pueblanue	10	22235068		

	A	B	C	D	E	F	G
1	ID	NAME	LAT	LONG	ELEVATION	Basin_ID	Basin
2	7603	7603_PCP	40.346	-5.082	1520	1	Navaluenga
3	7604	7604_PCP	40.346	-5.016	1433	1	Navaluenga
4	7605	7605_PCP	40.345	-4.95	1402	1	Navaluenga
5	7606	7606_PCP	40.345	-4.885	1342	1	Navaluenga
6	7607	7607_PCP	40.344	-4.819	1185	1	Navaluenga
7	7608	7608_PCP	40.343	-4.754	1045	1	Navaluenga
8	7712	7712_PCP	40.397	-5.212	1636	1	Navaluenga
9	7713	7713_PCP	40.397	-5.146	1601	1	Navaluenga
10	7714	7714_PCP	40.396	-5.081	1557	1	Navaluenga
11	7715	7715_PCP	40.396	-5.015	1505	1	Navaluenga
12	7716	7716_PCP	40.395	-4.95	1423	1	Navaluenga
13	7717	7717_PCP	40.395	-4.884	1295	1	Navaluenga
14	7718	7718_PCP	40.394	-4.818	1149	1	Navaluenga

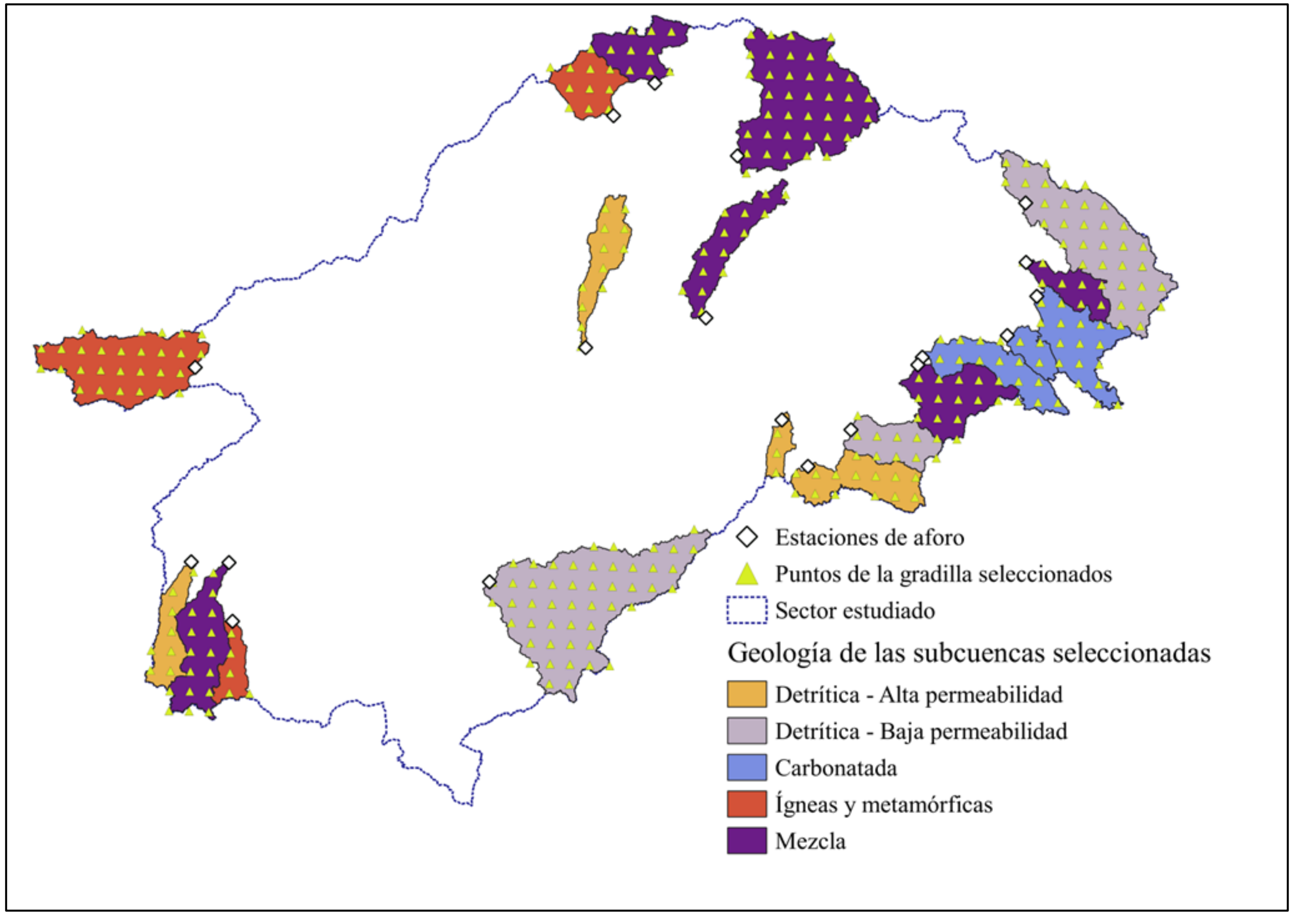
Step 3: Runoff coefficient calculation

Annual precipitation calculation (mm)

- Daily basin precipitation calculated from the mean of the weather points within.
- Sum of daily precipitation to calculate anual precipitation.

Annual water yield calculation (mm)

- Average streamflow in the year calculation (m^3/s) from daily records.
- Converted to anual water yield using the subbasin area and number of seconds in a year (mm/year).



Step 3: Runoff coefficient calculation

Runoff coefficient calculation

Rc = Water yield (mm) / Precipitation (mm)

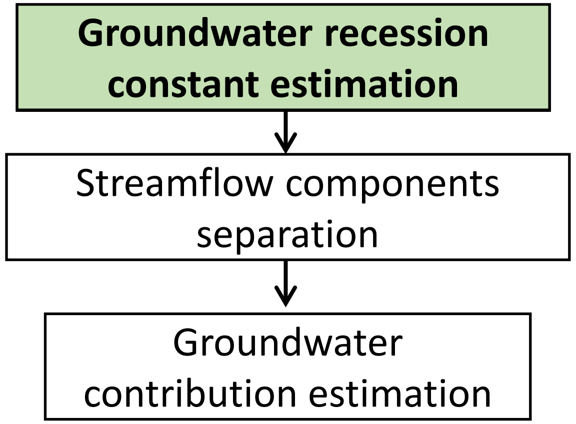
Annual average, minimum and maximum values for the selected period.

Values obtained for each subbasins and also aggregated to geological regions.

Table III: Subbasins runoff coefficients. Average, minimum and maximum value for the entire period.

ID and Subbasin	Region	Mean Temperature	Mean Precipitation	Mean Runoff coefficient	Min Runoff coefficient	Max Runoff coefficient
1, Navaluenga	IMP	11.53	897	0.37	0.26	0.56
2, Matallana	IMP	9.08	907	0.47	0.25	0.57
3, Villarejo de Montalban	IMP	15.46	592	0.14	0.03	0.29
4, Peralejo de las Truchas	CRB	9.74	770	0.38	0.22	0.49
5, Priego Escabas	CRB	10.85	733	0.37	0.24	0.46
6, Santa Maria del Val	CRB	9.78	783	0.45	0.19	0.74
7, Jabalera	DTH	13.75	469	0.09	0.03	0.24
8, Huete	DTH	13.29	534	0.05	0.02	0.10
9, Torote	DTH	13.23	486	0.06	0.03	0.10
10, La Pueblanueva	DTH	15.89	522	0.06	0.01	0.15
11, Ventosa	DTL	11.03	553	0.08	0.05	0.14
12, La Peraleja	DTL	13.32	541	0.02	0.01	0.06
13, Villasequilla de Yepes	DTL	14.39	356	0.02	0.01	0.03
14, Valverde de los Arroyos	MIX	9.85	873	0.36	0.20	0.48
15, Malpica	MIX	15.99	555	0.17	0.05	0.31
16, Taravillas	MIX	10.46	695	0.19	0.09	0.30
17, Romanones	MIX	12.01	534	0.06	0.04	0.10
18, Priego Trabraque	MIX	11.29	626	0.03	0.01	0.08
19, Bujalaro	MIX	10.75	508	0.11	0.07	0.14

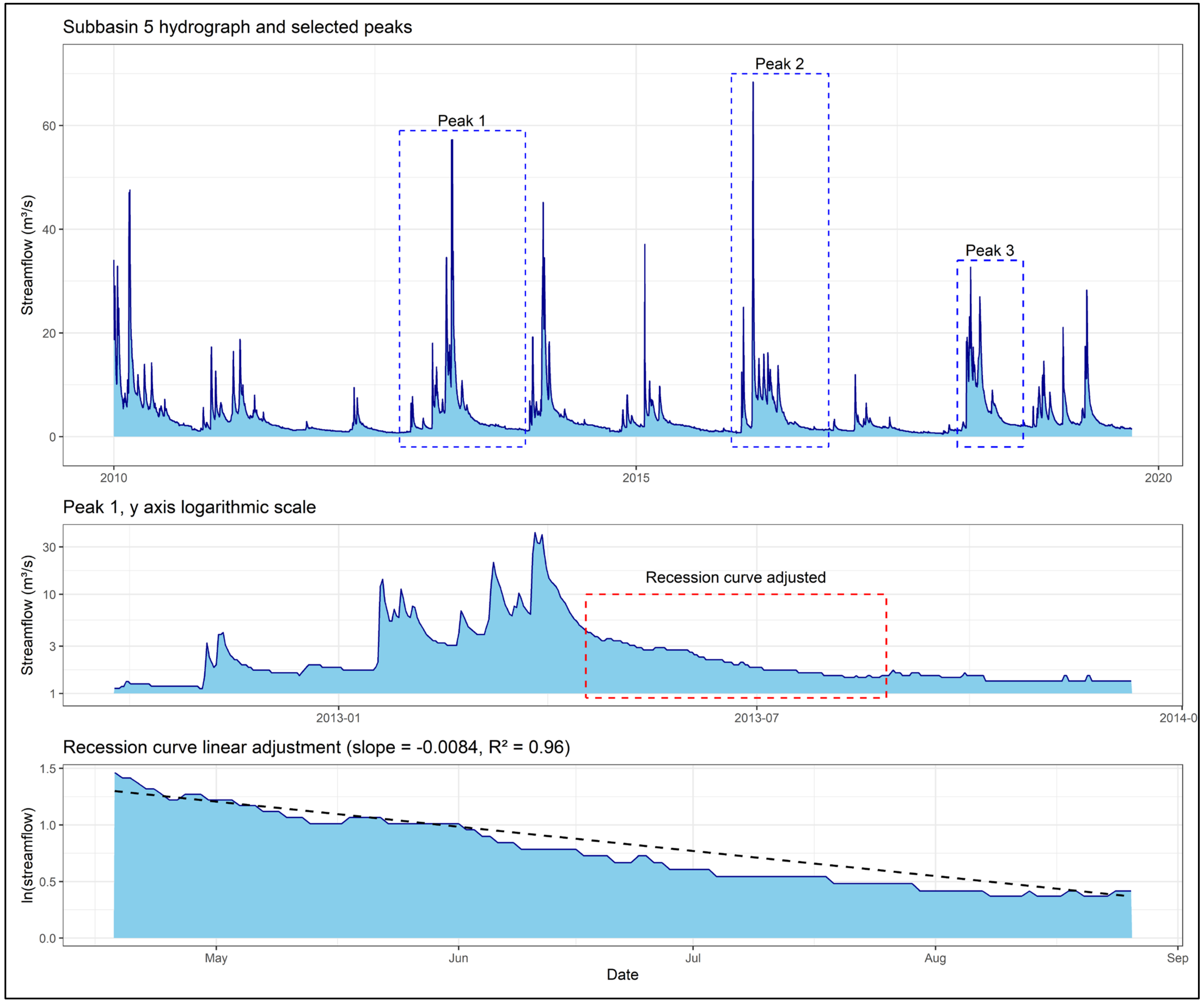
Step 4: Groundwater contribution estimation



- Using the Maillet (1905) equation, α can be estimated.

$$Q_t = Q_0 * e^{-\alpha * t}$$

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



Step 4: Groundwater contribution estimation

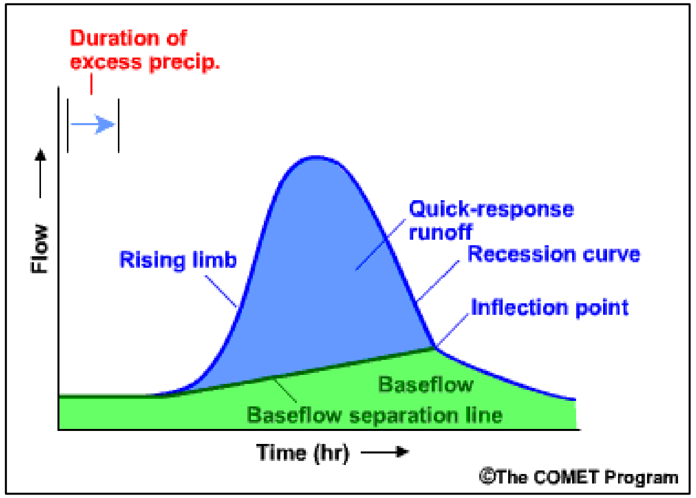
Groundwater recession constant estimation

↓

Streamflow components separation

↓

Groundwater contribution estimation

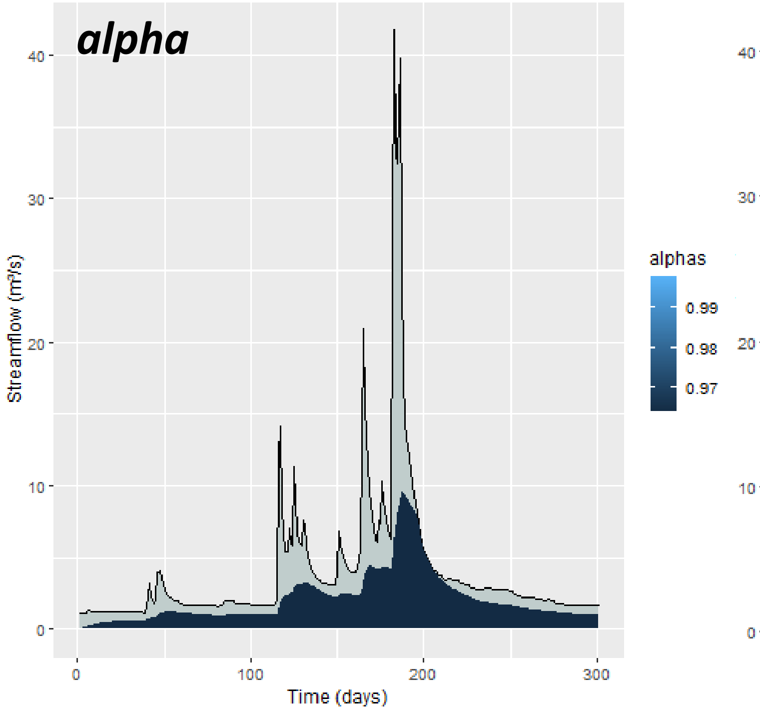


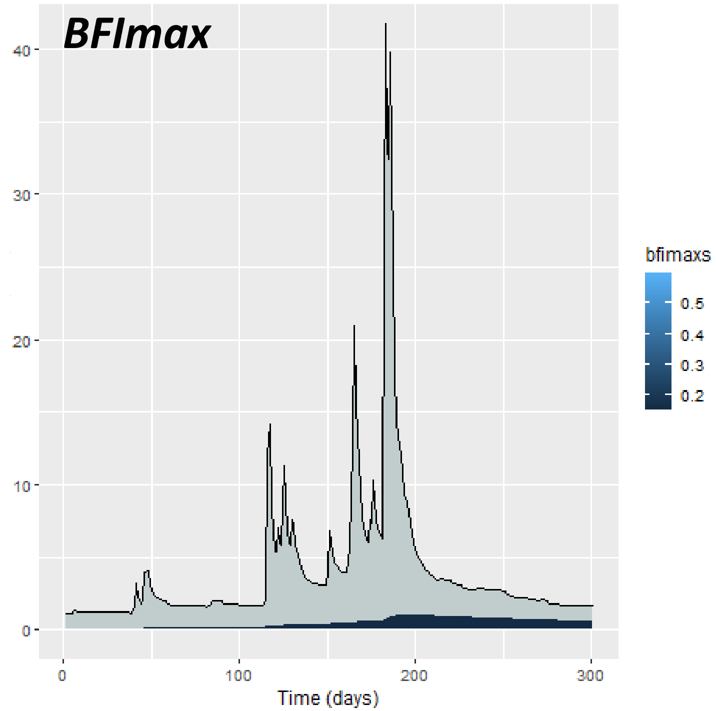
Eckhardt, K. (2005). How to construct recursive digital filters for baseflow separation. Hydrological Processes: An International Journal, 19(2), 507-515.

- Performed using a digital baseflow filter, based on Eckhardt (2005).

$$b_k = \frac{(1 - BFImax) * alpha * b_{k-1} + (1 - alpha) * BFImax * y_k}{1 - alpha * BFImax}$$

- This filter uses two parameters:
  - ➢ *alpha*: Controls the immediacy of the baseflow response to precipitation or recession processes. $alpha = e^{-\alpha}$
  - ➢ *BFImax*: Controls the maximum baseflow value for each day.





Step 4: Groundwater contribution estimation

Groundwater recession constant estimation

↓

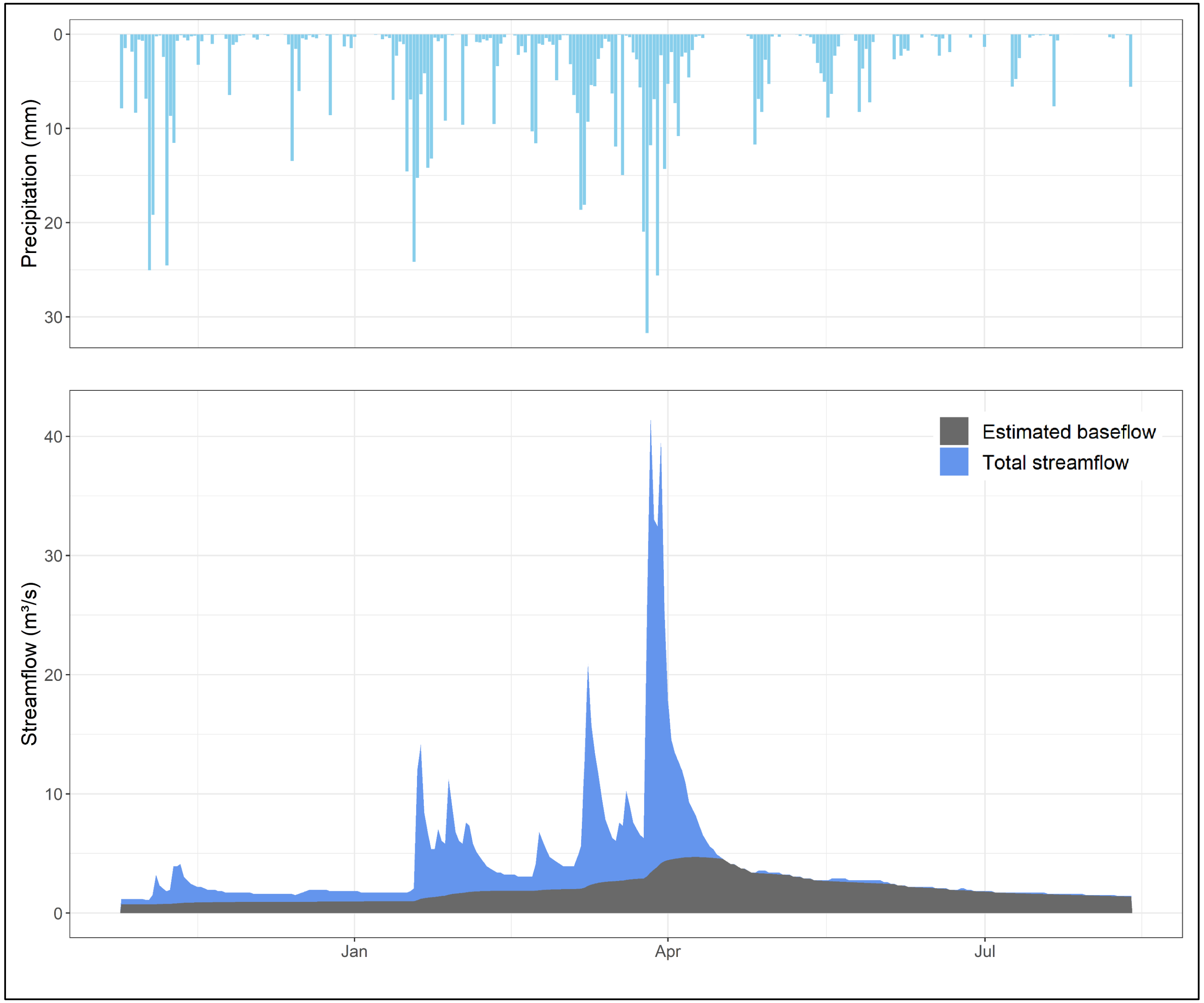
Streamflow components separation

↓

Groundwater contribution estimation

- Once a realistic separation has been performed, the baseflow contribution can be calculated.

$$Bf_c = \frac{\sum_1^n bf_k}{\sum_1^n (bf_k + rn_k)}$$



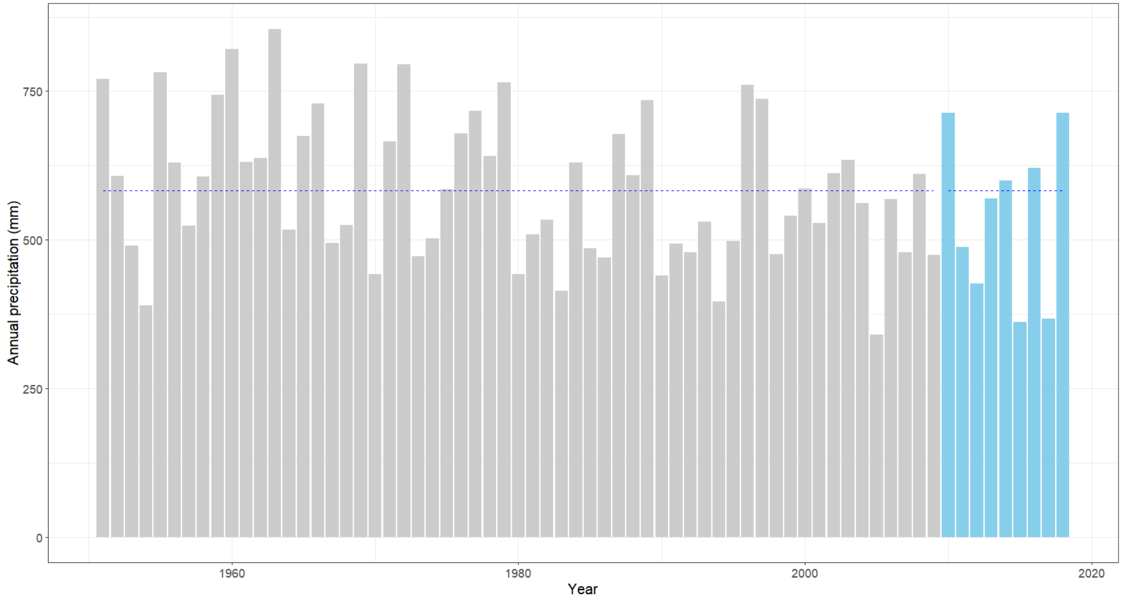
Methodology considerations

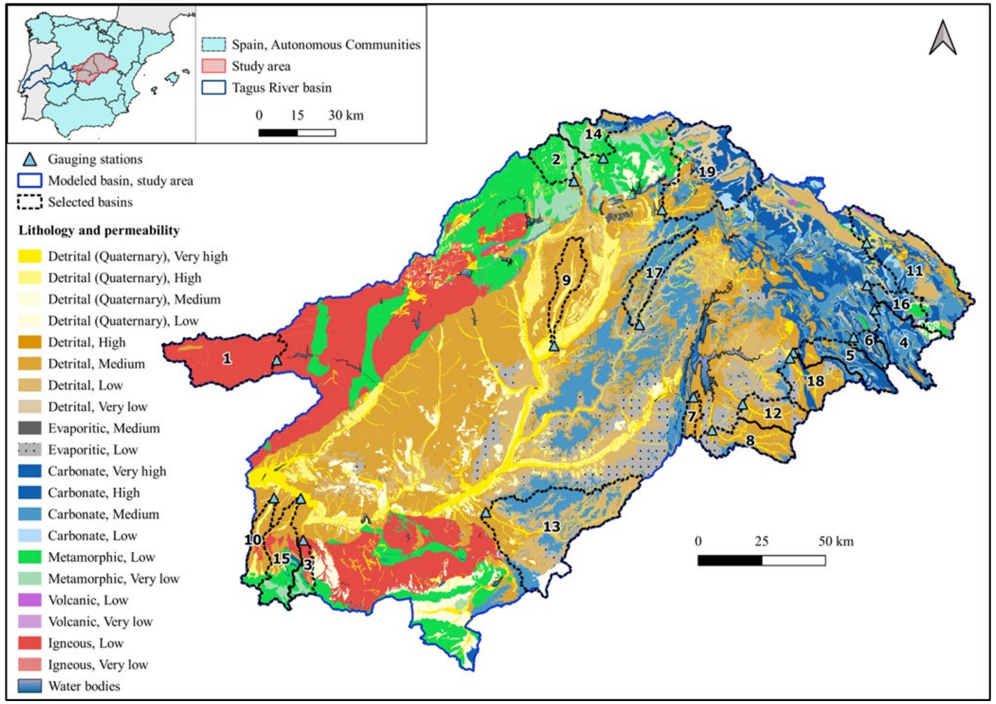
Selected period and subbasins:

For climatic/hydrological characterizations → **Time period should be long and heterogeneous enough** to be representative (~ 30 years).

However, for modelling purposes → Target values for the modelling period should be estimated.

In this study case, due to the subbasins selection criteria, **all the subbasins are located in the headwaters** → Steeper slopes and more humid and colder conditions.





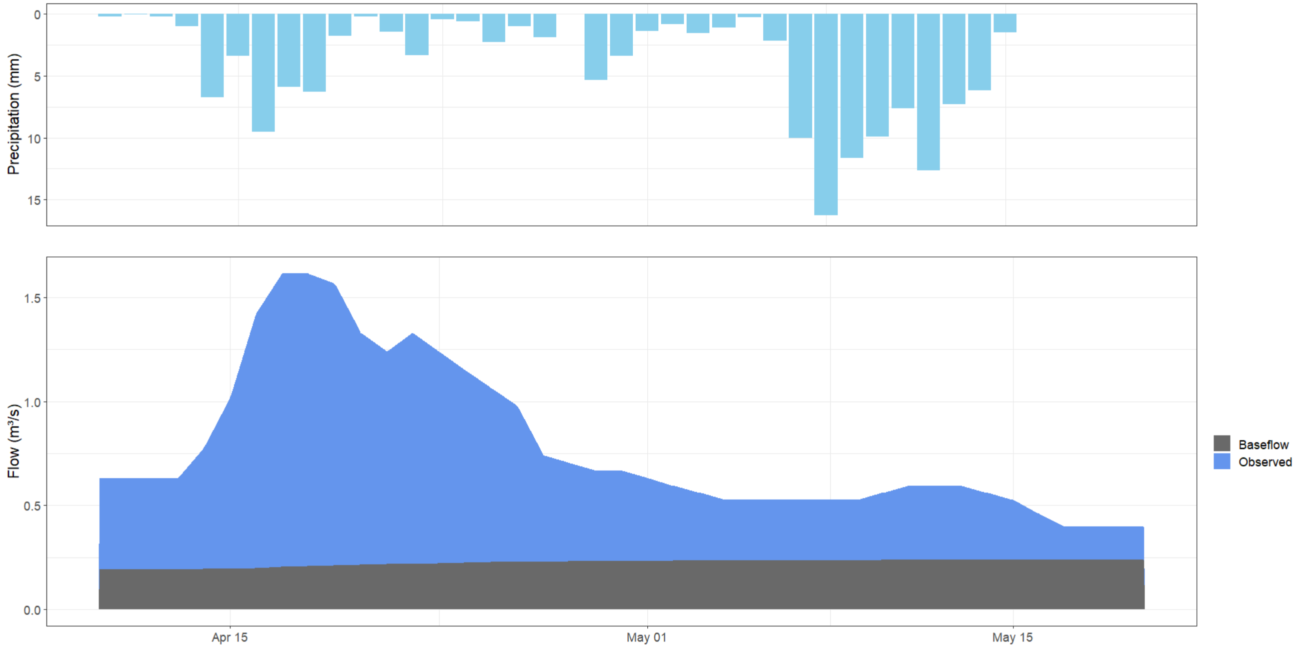
Methodology considerations

Input data reliability:

Datasets, especially streamflow, should be analysed to avoid potential mistakes.

- Runoff rate calculation → Lack of data understimate this variable → Only years with complete data should be used.
- Baseflow contribution estimation → Anomalous record → Wrong estimation if not replaced.

Basin ID	Basin	Period with available data	Complete data (2010-2018)
1	Navaluenga	1974-2019	Yes
2	Matallana	2010-2019	No, 273 days missing
3	Villarejo de Montalban	1969-2019	Yes
4	Peralejo de las Truchas	1945-2019	Yes
5	Priego Escabas	1912-2019	Yes
6	Santa Maria del Val	2010-2019	No, 273 days missing
7	Jabalera	1977-2019	Yes
8	Huete	1965-2019	Yes
9	Torote	1972-2019	Yes



Methodology considerations

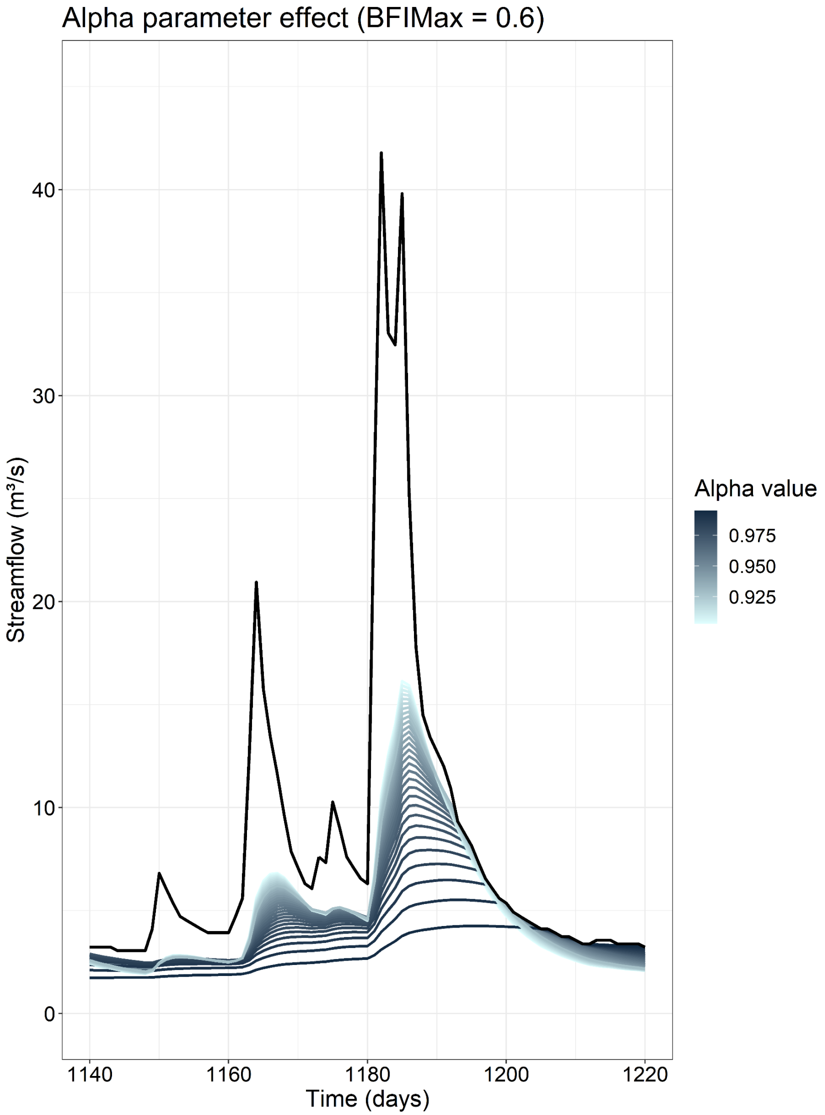
Baseflow filter parameters selection:

The estimated α values were not always used in the filter → A realistic separation should be preferred than using the estimated values.

Values of alpha lower than 0.95, in our opinion, should be avoided.

Similarly, BFImax should have an appropriate value considering the basin geology.

Basin	Region	Estimated alpha	Used alpha
Navaluenga	IMP	0.976	0.975
Matallana	IMP	0.956	0.965
Villarejo de Montalban	IMP	0.950	0.951
Peralejo de las Truchas	CRB	0.993	0.993
Priego Escabas	CRB	0.994	0.995
Santa Maria del Val	CRB	0.987	0.995
Jabalera	DTH	0.966	0.987
Huete	DTH	0.978	0.992
Torote	DTH	0.962	0.978
La Pueblanueva	DTH	0.949	0.978
Ventosa	DTL	0.992	0.998
La Peraleja	DTL	0.947	0.990
Villasequilla de Yepes	DTL	0.988	0.998
Valverde de los Arroyos	MIX	0.945	0.980
Malpica	MIX	0.937	0.985
Taravillas	MIX	0.989	0.995
Romanones	MIX	0.992	0.998
Priego Trabraque	MIX	0.986	0.990
Bujalaro	MIX	0.988	0.995

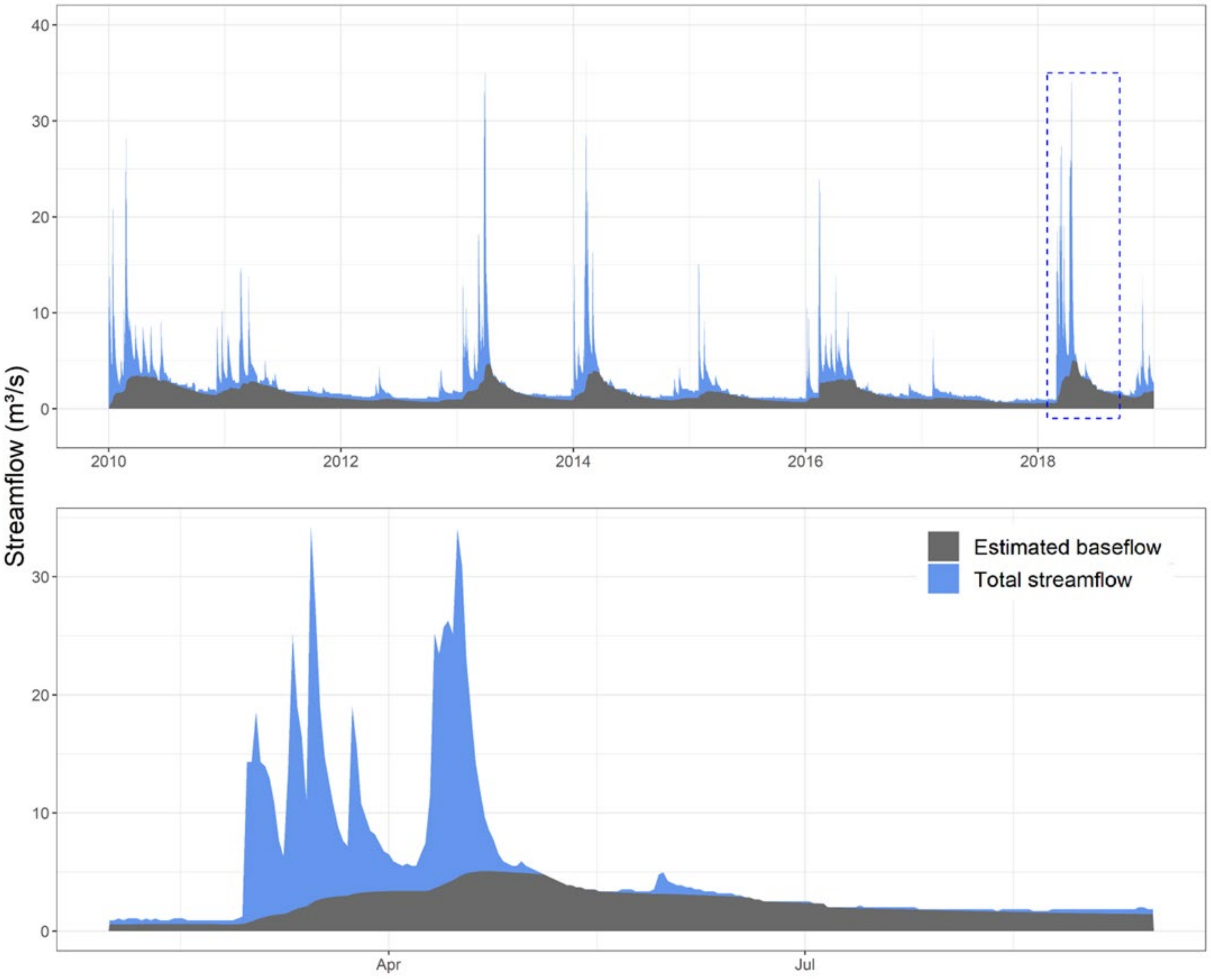


Methodology considerations

Time scale when performing the streamflow components separation

When working with series with more tan one year, the streamflow components separation should be **evaluated for individual events**, reaching a reasonable separation for some of them.

Longer time series including several peaks → Visual perception of a larger and unrealistic amount of groundwater contribution.



Results and conclusions

- Runoff coefficient and baseflow contribution estimations were obtained for 19 subbasins of the Tagus River basin, representative of four geological regions.
- Average values of these variables might guide a soft calibration, ensuring that the water balance is simulated realistically.
- A reproducible methodology has been developed for the entire Spanish territory and is aplicable to other regions with similar datasets.
- Expert knowledge and the explained considerations should guide the use of the methodology presented.

Table 1: Average, minimum and maximum runoff coefficients for the geological regions.

Region	Mean Temperature	Mean Precipitation	Mean Runoff coefficient	Min Runoff coefficient	Max Runoff coefficient
IMP	12.02	799	0.327	0.14	0.47
CRB	10.12	762	0.400	0.37	0.45
DTH	14.04	503	0.065	0.05	0.09
DTL	12.91	483	0.040	0.02	0.08
MIX	11.72	632	0.153	0.03	0.36

Table 3: Average parameter values used in the baseflow filter and groundwater indexes estimated at region scale.

Region	Mean alpha used	Mean BFImax used	Estimated baseflow index	Baseflow index standard deviation
IMP	0.964	0.23	0.22	0.03
CRB	0.994	0.55	0.46	0.10
DTH	0.984	0.39	0.34	0.03
DTL	0.995	0.38	0.30	0.11
MIX	0.990	0.36	0.31	0.13

Publications:

Sánchez-Gómez, A., Bieger, K., Schürz, C., Martínez-Pérez, S., Rathjens, H., Molina-Navarro, E., Forthcoming. Soft data collection for realistic hydrological modelling: A reproducible methodology developed in R for estimating the runoff coefficient and the baseflow index. (Under review). Environmental Modelling & Software.

Sánchez-Gómez, A., Schürz, C., Bieger, K., Martínez-Pérez, S., Molina-Navarro, E., Forthcoming. Using sensitivity analysis and soft calibration of geological regions to improve the representation of hydrological processes in a SWAT+ model. (Under review). Journal of Hydrology.



Thanks for your attention!

alejandro.sanchezg@uah.es




Universidad de Alcalá
MINISTERIO DE CIENCIA E INNOVACIÓN
Castilla-La Mancha
Cofinanciado por la Unión Europea
PARQUE CIENTÍFICO Y TECNOLÓGICO DE CASTILLA - LA MANCHA
CEAGU