## HydRoVars: An R Tool to collect hydrological variables.

Soft data collection for realistic hydrological modelling

Alejandro Sánchez-Gómez\*^, Katrin Bieger\*\*, Christoph Schürz\*\*\*, Silvia Martínez-Pérez\*, Hendrik Rathjens\*\*\*\*, Eugenio Molina-Navarro\*.

\* Department of Geology, Geography and Environment. University of Alcalá. (E-mail\*: [alejandro.sanchezg@uah.es](mailto:alejandro.sanchezg@uah.es), [alejandro.sanchezgomez97@gmail.com](mailto:alejandro.sanchezgomez97@gmail.com))

\*\*Department of Ecoscience. Aarhus University.

\*\*\*Department of Computational Landscape Ecology. Helmholtz Centre for Environmental Research.

\*\*\*\*Stone Environmental, Inc.

**Supplementary Materials**

### Appendix A: Output files generated

Output file 1: Studied basins data: properties and gauging data availability

| Basin data | Area (km²) | Range and number of years with observed data | Data availability for the studied period |
| --- | --- | --- | --- |
| 1, Navaluenga (3231) | 699.31 | 1974-2019 (45) | Complete |
| 2, Matallana (3049) | 252.26 | 2010-2019 (9) | Uncomplete Years = 2010, days left = 273 |
| 3, Villarejo de Montalban (3211) | 136.21 | 1969-2019 (50) | Complete |
| 4, Peralejo de las Truchas (3001) | 408.56 | 1945-2019 (74) | Complete |
| 5, Priego Escabas (3045) | 329.24 | 1912-2019 (107) | Complete |
| 6, Santa Maria del Val (3040) | 117.79 | 2010-2019 (9) | Uncomplete Years = 2010, days left = 273 |
| 7, Jabalera (3249) | 85.16 | 1977-2019 (42) | Complete |
| 8, Huete (3172) | 359.43 | 1965-2019 (54) | Complete |
| 9, Torote (3193) | 254.67 | 1972-2019 (47) | Complete |
| 10, La Pueblanueva (3251) | 222.35 | 1978-2019 (41) | Complete |
| 11, Ventosa (3030) | 942.43 | 1945-2019 (74) | Complete |
| 12, La Peraleja (3173) | 260.79 | 1966-2019 (53) | Complete |
| 13, Villasequilla de Yepes (3164) | 1320.77 | 1971-2019 (48) | Complete |
| 14, Valverde de los Arroyos (3165) | 279.32 | 2011-2019 (8) | Years without records = 2010 , Uncomplete Years = 2011, days left = 181 |
| 15, Malpica (3212) | 412.12 | 1972-2019 (47) | Complete |
| 16, Taravillas (3268) | 184.27 | 1982-2019 (37) | Complete |
| 17, Romanones (3237) | 319.21 | 1975-2019 (44) | Complete |
| 18, Priego Trabaque (3186) | 389.17 | 1969-2019 (50) | Complete |
| 19, Bujalaro (3060) | 1027.91 | 1961-2019 (58) | Complete |

Output file 2: Weather and runoff coefficient data: Average temperature and precipitation for the studied period, average runoff rate, and maximum and minimum annual values

| Basin ID | Basin | Region | Mean Temperature (ºC) | Mean Precipitation (mm) | Mean Runoff rate | Minimum Runoff rate | Maximum Runoff rate |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | Navaluenga | IMP | 10.30 | 897 | 0.374 | 0.264 | 0.565 |
| 2 | Matallana | IMP | 9.75 | 907 | 0.473 | 0.250 | 0.573 |
| 3 | Villarejo de Montalban | IMP | 14.68 | 592 | 0.136 | 0.033 | 0.292 |
| 4 | Peralejo de las Truchas | CRB | 9.35 | 770 | 0.380 | 0.215 | 0.488 |
| 5 | Priego Escabas | CRB | 11.40 | 733 | 0.374 | 0.241 | 0.455 |
| 6 | Santa Maria del Val | CRB | 9.92 | 783 | 0.450 | 0.191 | 0.741 |
| 7 | Jabalera | DTH | 13.69 | 469 | 0.086 | 0.034 | 0.242 |
| 8 | Huete | DTH | 13.26 | 534 | 0.052 | 0.025 | 0.102 |
| 9 | Torote | DTH | 13.95 | 486 | 0.056 | 0.030 | 0.103 |
| 10 | La Pueblanueva | DTH | 15.32 | 522 | 0.064 | 0.013 | 0.152 |
| 11 | Ventosa | DTL | 10.62 | 553 | 0.079 | 0.045 | 0.144 |
| 12 | La Peraleja | DTL | 13.19 | 541 | 0.021 | 0.006 | 0.059 |
| 13 | Villasequilla de Yepes | DTL | 15.20 | 356 | 0.019 | 0.009 | 0.025 |
| 14 | Valverde de los Arroyos | MIX | 9.72 | 873 | 0.364 | 0.201 | 0.479 |
| 15 | Malpica | MIX | 15.06 | 555 | 0.168 | 0.046 | 0.307 |
| 16 | Taravillas | MIX | 10.10 | 695 | 0.188 | 0.086 | 0.296 |
| 17 | Romanones | MIX | 12.57 | 534 | 0.063 | 0.042 | 0.095 |
| 18 | Priego Trabaque | MIX | 12.56 | 626 | 0.034 | 0.009 | 0.081 |
| 19 | Bujalaro | MIX | 11.34 | 508 | 0.106 | 0.065 | 0.140 |

Output file 3: Groundwater recession data: Duration of the recession curve, average recession constant and alpha values, and average coefficient of determination obtained

| Basin ID | Basin | Region | Average recession duration | Average coefficient of determination | Average groundwater recession constant | Average alpha value |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | Navaluenga | IMP | 74 | 0.908 | -0.024 | 0.976 |
| 2 | Matallana | IMP | 42 | 0.989 | -0.045 | 0.956 |
| 3 | Villarejo de Montalban | IMP | 26 | 0.970 | -0.052 | 0.950 |
| 4 | Peralejo de las Truchas | CRB | 163 | 0.944 | -0.007 | 0.993 |
| 5 | Priego Escabas | CRB | 140 | 0.884 | -0.006 | 0.994 |
| 6 | Santa Maria del Val | CRB | 127 | 0.954 | -0.013 | 0.987 |
| 7 | Jabalera | DTH | 53 | 0.935 | -0.035 | 0.966 |
| 8 | Huete | DTH | 65 | 0.891 | -0.022 | 0.978 |
| 9 | Torote | DTH | 46 | 0.535 | -0.039 | 0.962 |
| 10 | La Pueblanueva | DTH | 31 | 0.928 | -0.053 | 0.949 |
| 11 | Ventosa | DTL | 79 | 0.903 | -0.008 | 0.992 |
| 12 | La Peraleja | DTL | 48 | 0.888 | -0.055 | 0.947 |
| 13 | Villasequilla de Yepes | DTL | 100 | 0.876 | -0.012 | 0.988 |
| 14 | Valverde de los Arroyos | MIX | 44 | 0.978 | -0.056 | 0.945 |
| 15 | Malpica | MIX | 33 | 0.956 | -0.066 | 0.937 |
| 16 | Taravillas | MIX | 85 | 0.915 | -0.011 | 0.989 |
| 17 | Romanones | MIX | 98 | 0.916 | -0.008 | 0.992 |
| 18 | Priego Trabaque | MIX | 90 | 0.892 | -0.014 | 0.986 |
| 19 | Bujalaro | MIX | 67 | 0.924 | -0.012 | 0.988 |

Output file 4: Groundwater contribution data: Values used for the baseflow separation and groundwater contribution estimated

| Basin | Basin ID | Region | Alpha value used | BFImax value used | Estimated groundwater contribution (fraction) |
| --- | --- | --- | --- | --- | --- |
| Navaluenga | 1 | IMP | 0.990 | 0.40 | 0.308 |
| Matallana | 2 | IMP | 0.989 | 0.30 | 0.239 |
| Villarejo de Montalban | 3 | IMP | 0.985 | 0.35 | 0.231 |
| Peralejo de las Truchas | 4 | CRB | 0.995 | 0.60 | 0.513 |
| Priego Escabas | 5 | CRB | 0.997 | 0.65 | 0.549 |
| Santa Maria del Val | 6 | CRB | 0.997 | 0.70 | 0.417 |
| Jabalera | 7 | DTH | 0.985 | 0.45 | 0.394 |
| Huete | 8 | DTH | 0.993 | 0.55 | 0.476 |
| Torote | 9 | DTH | 0.985 | 0.55 | 0.423 |
| La Pueblanueva | 10 | DTH | 0.980 | 0.55 | 0.384 |
| Ventosa | 11 | DTL | 0.996 | 0.60 | 0.555 |
| La Peraleja | 12 | DTL | 0.995 | 0.60 | 0.360 |
| Villasequilla de Yepes | 13 | DTL | 0.995 | 0.45 | 0.370 |
| Valverde de los Arroyos | 14 | MIX | 0.980 | 0.20 | 0.182 |
| Malpica | 15 | MIX | 0.975 | 0.35 | 0.278 |
| Taravillas | 16 | MIX | 0.993 | 0.50 | 0.460 |
| Romanones | 17 | MIX | 0.996 | 0.50 | 0.456 |
| Priego Trabaque | 18 | MIX | 0.990 | 0.45 | 0.395 |
| Bujalaro | 19 | MIX | 0.993 | 0.50 | 0.478 |

### Appendix B: Code snippets

The following lines present the main sections of the R code created to apply the methodology presented in this manuscript. The full code can be found in the GitHub repository included in the Data availability section.

Snippet 1: Creating input files

# FILE 1.- Basins csv file  
 # Input data: Shapefile with the delineated basins.   
 basins <- read\_sf("1\_Used\_files/GIS/Shapefiles/basins\_studied.shp") %>% arrange(., id)  
 basins\_csv <- basins %>% rename(Basin\_ID = id) %>%   
 mutate(area = st\_area(.)) %>%   
 # Introducing the gauging stations codes (Manually) # User action: Introduce the gauging stations code  
 mutate(gauging\_code = c(3231, 3049, 3211, 3001, 3045, 3040,  
 3249, 3172, 3193, 3251, 3030, 3173,  
 3164, 3165, 3212, 3268, 3237, 3186 , 3060)) %>%   
 st\_drop\_geometry(.) %>%   
 write.csv(x = basins\_csv, file = "1\_Used\_files/Created\_csv/1\_basins\_file.csv", row.names = F)  
   
 # FILE 2. Gauging points csv file  
 #Input data: weather grid and delineated basins.  
 # Grid points

pcp\_points <- read\_sf("1\_Used\_files/GIS/Shapefiles/weather\_grid\_UTM.shp")  
 basins <- read\_sf("1\_Used\_files/GIS/Shapefiles/basins\_studied.shp") %>% arrange(., id)  
   
 # 2.1. Buffer created for basins (1 km distance)  
 basins\_buffer <- st\_buffer(basins, dist = 1000) # User action: define buffer (m)  
   
 # 2.2.Clipping grid points with the basins buffer (region column is not necessary)  
 grid\_points\_clip <- st\_intersection(pcp\_points, basins\_buffer[, c("id", "Basin", "geometry")])  
   
 write.csv(x = grid\_points\_clip\_csv, file = "1\_Used\_files/Created\_csv/2\_ids\_stations\_file.csv", row.names = F)

Snippet 2: Runoff aggregation

study\_period <- c(2010:2018) # User action: Define study period  
 areas <- basins\_file$area # Drainage areas for converting flow to milimeters  
   
 obs\_anual <- list() # Empty list   
 for(i in 1:length(cods)){  
 gaug\_st <- filter(gauging\_data\_tagus, cod == cods[i]) %>% filter(year(date) %in% study\_period)

caud\_anual <- gaug\_st[,c("date", "obs\_flow")] %>% group\_by(year = year(date)) %>% summarise(., obs\_m3 = mean(obs\_flow)) %>% # Average annual flow (m3/s)  
 mutate(obs\_mm = (obs\_m3\*86400\*365\*1000)/ (areas[i])) %>% # Annual contribution (mm/year)  
 cbind(bas = i) %>% .[,c("bas", "year", "obs\_mm")] # Final table with Basin\_ID, Year and Annual contribution  
 obs\_anual[[i]] <- caud\_anual # List with the gauged data for all the basins  
 }

Snippet 3: Precipitation aggregation

path <- "1\_Used\_files/Data/weather\_data/pcp\_spain/" # Directory with weather data  
 # User action: Define the starting and ending dates for the weather data  
 init\_date <- as.Date("1951-01-01")  
 end\_date <- as.Date("2019-12-31")  
 dates <- seq(init\_date, end\_date, 1) # A sequence of dates for the entire period with data is created  
   
 pcp\_grid\_points <- read.csv("1\_Used\_files/Created\_csv/2\_ids\_stations\_file.csv") %>% arrange(., Basin\_ID) # File with IDs, names, and location of the grid points, and basins data   
   
 # Loop for calculating the annual precipitation of each basin trough the average of the annual precipitation for each station within the basin  
 pcp\_bas\_list <- list() #empty list  
 for(i in 1:length(unique(pcp\_grid\_points$Basin\_ID))){ # i --> Basin ID  
 filt\_st <- filter(pcp\_grid\_points, Basin\_ID == i) # Basin data and precipitations points inside  
 stations <- filt\_st[,1] #Precipitations points inside each basin  
 pcps\_sts <- c()  
 for(n in 1:length(stations)){ # n --> Weather stations identifier within each basin  
 st\_dat <- read\_table(paste(path, stations[n], "\_PCP.txt", sep = ""), skip = 1, col\_names = F) %>% #read the precipitation file for each point  
 mutate(date = ymd(dates), pcp = X1) %>% .[,c("date", "pcp")] %>% group\_by(year(date)) %>%   
 summarise(pcp\_year = sum(pcp)) # calculate the total precipitation for each year  
 colnames(st\_dat) <- c("year", "pcp")  
   
 pcp\_st <- filter(st\_dat, year %in% study\_period) %>% .[,"pcp"] # Filtering with the study period  
 pcps\_sts <- tibble(pcps\_sts, pcp\_st, .name\_repair = "unique") # Table with annual precipitation data for all the points of a basin   
 }  
 pcp\_bas <- pcps\_sts %>% apply(., 1, mean) %>% cbind(year = c(study\_period)) %>%   
 tibble(year = .[,"year"], pcp\_y = .[,"."]) %>% .[,c("year", "pcp\_y")] # Calculate for each basin the average precipitation of all the precipitation points within  
 pcp\_bas\_list[[i]] <- pcp\_bas[, "pcp\_y"] %>% cbind(year = c(study\_period), bas = i) %>% .[,c("bas", "year", "pcp\_y")] }

Snippet 4: Runoff rate calculation

anual\_runoff\_rate <- list() #empty list   
 basin\_runoff\_rate <- list() #empty list  
 basin\_runoff\_rates <- c() #empty vector

# loop for calculating runoff coefficient  
 for(i in 1:length(pcp\_bas\_list)){  
 anual\_runoff\_rate[[i]] <- obs\_anual[[i]] %>% left\_join(pcp\_bas\_list[[i]], by = "year") %>%   
 mutate(Basin\_ID = bas.x, Year = year, Pcp = pcp\_y, Runoff = obs\_mm, Runoff\_rt = Runoff/Pcp) %>%   
 .[,c("Basin\_ID", "Year", "Pcp", "Runoff", "Runoff\_rt")] # List with the annual values

basin\_runoff\_rate[[i]] <- anual\_runoff\_rate[[i]] %>% summarise(Basin\_ID = mean(Basin\_ID), Mean\_pcp = mean(Pcp), Mean\_runoff = mean(Runoff), Runoff\_rate = mean(Runoff\_rt), Max\_runoff\_rate = max(Runoff\_rt), min\_runoff\_rate = min(Runoff\_rt), Runoff\_rate\_sd = sd(Runoff\_rt)) %>% unlist(.) # List with the average precipitation, runoff and runoff rate values for the entire period; and maximum and minimum runoff rate  
 basin\_runoff\_rates <- basin\_runoff\_rates %>% rbind(basin\_runoff\_rate[[i]]) # Merge the average list values  
 }

Snippet 5: Recession curve linear regression

# Peak 1  
  
reg\_pk1 <- lm(log(peak\_1\_data$obs\_flow[peak\_1\_recesion])~ seq(1, length(peak\_1\_data$date[peak\_1\_recesion]), 1)) # Linear regression  
sum\_reg\_pk1 <- summary(reg\_pk1) # Results of the regression  
  
rec\_const\_pk1 <- sum\_reg\_pk1$coefficients[2,1] # Slope of the curve  
adj\_r2\_pk1 <- sum\_reg\_pk1$adj.r.squared # Coefficient of determination  
alpha\_value <- 2.71828182846^(rec\_const\_pk1) # Alpha value calculation  
  
peak\_1\_recession\_data <- tibble(Basin\_ID = basin\_ID,  
 peak = 1,  
 recess\_days = length(peak\_1\_recesion), # Recession curve duration  
 det\_coef = adj\_r2\_pk1,   
 gw\_rec\_const = rec\_const\_pk1,  
 alpha\_value = alpha\_value,  
 peak\_range = paste(peak\_1[1], peak\_1[length(peak\_1)], sep= ":"),   
 recess\_range = paste(peak\_1\_recesion[1], peak\_1\_recesion[length(peak\_1\_recesion)], sep= ":"),  
 basin\_info = basin\_information)

Snippet 6: Baseflow filter application

#Filter parameters  
 alpha <- 0.99 # User action: define the alpha parameter value  
 bfi\_max <- 0.4 # User action: define the BFImax parameter value

# Apply the filter to estimate baseflow  
 bfsep <- baseflow\_sep(df = basin\_data,   
 Q = "obs\_flow",   
 alpha = alpha,   
 BFIma =bfi\_max,   
 method = "two\_param") # Run baseflow filter with the selected parameters  
   
 bf\_sep\_data <- basin\_data %>% # Precipitation and streamflow data  
 mutate(baseflow = bfsep$B, runoff = bfsep$R) # Adding the baseflow filter output

# Calculate the baseflow index and create basin output data  
  
 bf\_index <- sum(bfsep$B) / (sum(bfsep$B)+sum(bfsep$R)) #Calculating the baseflow contribution  
 bf\_index  
   
 basin\_gwc\_data <- tibble(basins\_file[basins\_file$Basin\_ID == basin\_ID, c("Basin", "Basin\_ID", "region")]) %>%   
 mutate(alpha = round(alpha,3), bfi\_max\_used = round(bfi\_max,3), BF\_Rate = round(bf\_index,3))

### Appendix C: Lithology of the study case basins

Table C1: Basins total area and relative area of each geologic substrate (%)

| ID and basin | Total area (sq. km) | IMP % | CRB % | DTH % | DTL % |
| --- | --- | --- | --- | --- | --- |
| 1, Navaluenga | 699 | 98 | 0 | 2 | 0 |
| 2, Matallana | 252 | 92 | 0 | 8 | 0 |
| 3, Villarejo de Montalban | 136 | 79 | 4 | 17 | 0 |
| 4, Peralejo de las Truchas | 409 | 2 | 83 | 8 | 7 |
| 5, Priego Escabas | 329 | 0 | 74 | 24 | 1 |
| 6, Santa Maria del Val | 118 | 0 | 86 | 9 | 5 |
| 7, Jabalera | 85 | 0 | 14 | 85 | 1 |
| 8, Huete | 359 | 0 | 9 | 74 | 17 |
| 9, Torote | 255 | 0 | 0 | 100 | 0 |
| 10, La Pueblanueva | 222 | 23 | 0 | 77 | 0 |
| 11, Priego Trabaque | 942 | 0 | 27 | 57 | 16 |
| 12, La Peraleja | 261 | 0 | 1 | 73 | 26 |
| 13, Villasequilla de Yepes | 1321 | 0 | 35 | 18 | 48 |
| 14, Valverde de los Arroyos | 279 | 62 | 18 | 8 | 12 |
| 15, Malpica | 412 | 52 | 5 | 43 | 0 |
| 16, Taravillas | 184 | 19 | 59 | 4 | 18 |
| 17, Romanones | 319 | 0 | 57 | 29 | 14 |
| 18, Ventosa | 389 | 5 | 50 | 17 | 27 |
| 19, Bujalaro | 1028 | 3 | 44 | 23 | 29 |

### Appendix D: Examples of hydrographs of the study case subbasins

|  |
| --- |
| Figure D1: Hydrographs comparison of subbasins located in the different geological regions. |