Evaluation of soil clustering techniques to characterise hydrological soil processes at the catchment scale with SWAT+

International SWAT Conference (Strasbourg, France)

July 12th, 2024

Mauricio Zambrano-Bigiarini[1,2], Fernando Gimeno[3], Mauricio Galleguillos[4,2], Rodrigo Marinao[5]

[1] Department of Civil Engineering, Universidad de La Frontera, Temuco, Chile.
[2] Center for Climate and Resilience Research (CR2), Santiago, Chile.
[3] PhD(c) in Natural Resources Sciences, Universidad de La Frontera, Temuco, Chile.
[4] Faculty of Engineering and Sciences, Universidad Adolfo Ibañez, Santiago, Chile.
[5] MSc(c) in Engineering Sciences, Universidad de La Frontera, Temuco, Chile.

Motivation -I

- Soil properties not only affect the slow response of the catchment but also the water available for plants and recharge rates into the groundwater system.
- Therefore, an accurate characterisation of the spatial heterogeneity of soil properties is crucial for a reliable representation of soil hydrological processes in any hydrological model.
- In particular, in SWAT and SWAT+ the dynamic partitioning of P into Q and ET depends critially of the **soil map used as input data**.
- Null Hypothesis: the use of coarse-resolution global soil maps should limit our ability to represent soil hydrological processes

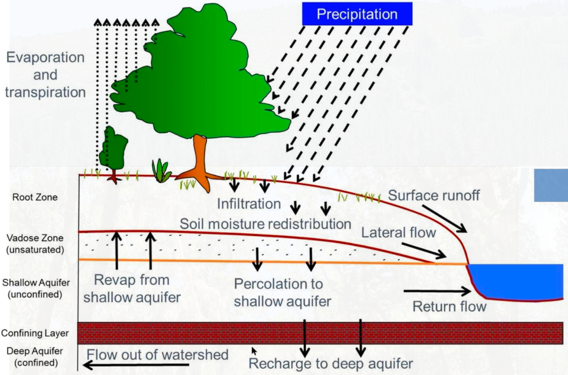
However, **is this really so?**

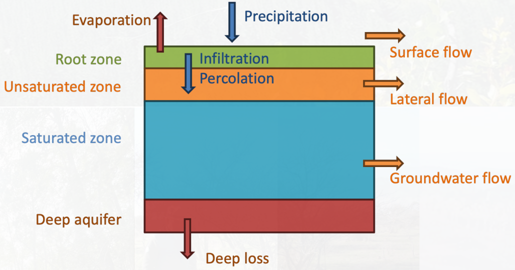
Motivation - II

- Unexpectedly, studies using different soil maps have been focused in the total streamflow response (e.g., NSE(Q), KGE(Q)) rather on soil hydrological processes affected by the soil properties (e.g., low flows).
- In particular, in 2021 we had a SWAT+ model with acceptable values of KGE and NSE during the validation period.
- However, we realise that the ET was not correctly represented for **native Chilean forest** → we computed local vegetation parameters.
- Unexpectedly: KGE and NSE become worse than in the previous case → we were right for the wrong reasons !!

Therefore, we decided to study soil properties with more detail in SWAT+.

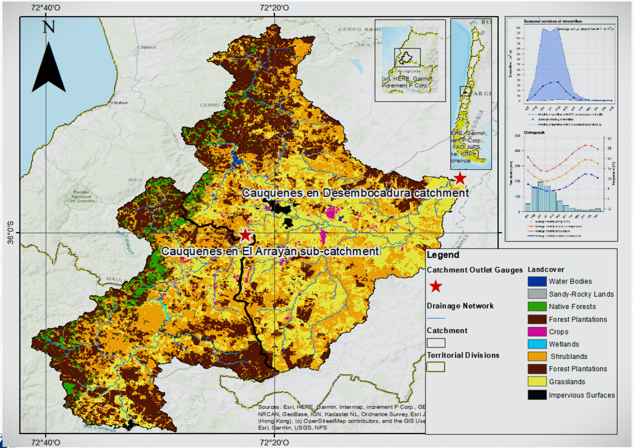
Hydrologcial processes in SWAT+ (Land Phase)





Source: Čerkasova et al. (2024). (Workshop 2024).

Study area: general description



Drainage area	1750 km^2
Elevations	108-736 m a.s.l.
Annual P	934 mm
Dominant cover	Exotic plantations (41%)

In situ soil data at "Playa Blanca" site

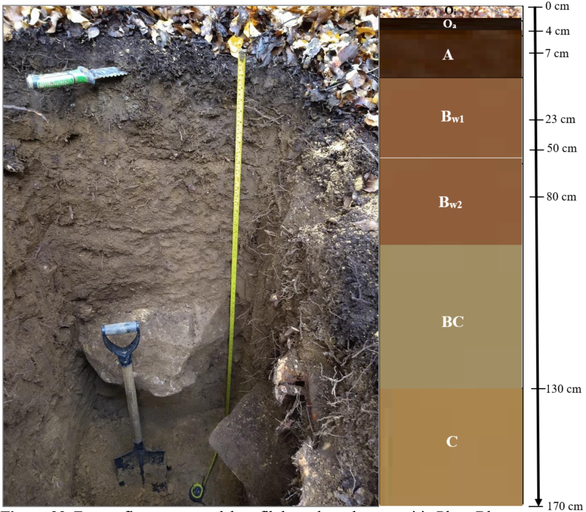


Figura 33. Fotografía y esquema del perfil de suelo en bosque, sitio Playa Blanca.

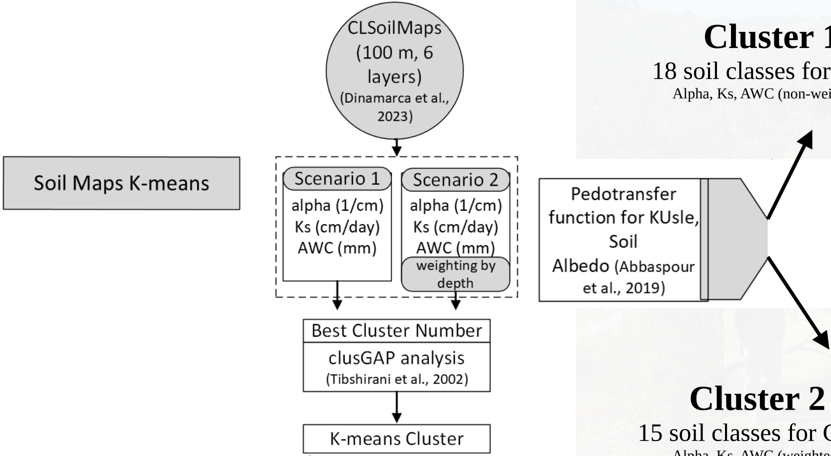
Soil Layer	Depth, [cm]	Bulk Density, [Mg/m3]	True Density, [Mg/m3]	Porosity, [%]	Clay, [%]	Silt, [%]	Sand, [%]
Oi-Oa	0-4	s/c	-	-	-	-	-
A	7-23	0,83	2,43	65,8	6,5	18,4	75,1
Bw1	23-50	0,81	2,49	67,5	6,4	20,4	73,1
Bw2	50-80	1,29	2,63	51,1	23,8	16,6	59,6
BC	80-130	1,40	2,63	46,9	21,5	15,3	63,2
C	130-170	1,53	2,66	42,2	26,8	18,6	54,6

Soil Layer	Depth, [cm]	Soil Water Content, [%]			
		2 kPa	60 kPa	330 kPa	15000 kPa
Oi-Oa	0-4	-	-	-	-
A	7-23	0,552	0,313	0,278	0,214
Bw1	23-50	0,559	0,342	0,273	0,135
Bw2	50-80	0,465	0,320	0,284	0,167
BC	80-130	0,366	0,259	0,229	0,120
C	130-170	0,389	0,310	0,288	0,152

CLSoilMaps: a new gridded soil map for Chile

- CLSoilMaps (Dinamarca et al., 2023) is a **new gridded soil map** with **physical properties** and **hydraulic parameters** at **100 m** spatial resolution for **six standardized depths**.
- It is publicly available for **continental Chile** and **binational basins** shared with Argentina at `https://doi.org/10.5281/zenodo.7464210`.
- It is based on digital soil mapping (DSM) and pedotransfer functions (Rosetta V3), following **GlobalSoilMap** project standards.
- Thousands of in situ soil profiles were collected for different land use conditions (e.g. agricultural, forest, peatland, shrubland, and Andean grassland), and tehn combined wuith several environmental covariates based on the SCORPAN soil forming factors.
- In particular, field capacity, permanent wilting point, total available water capacity, and Van Genuchten's soil hydraulic parameters were derived with Rosetta V3 algorithm.

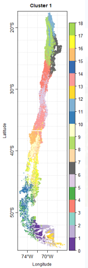
CLSoilmaps4SWAT: CLSoilmaps → SWAT+



Cluster 1

18 soil classes for Chile

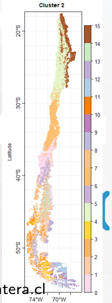
Alpha, Ks, AWC (non-weighted)



Cluster 2

15 soil classes for Chile

Alpha, Ks, AWC (weighted)



All gridded soil data used with SWAT+

N°	Gridded soil data	Spatial resolution	Spatial extent	N° Layers	Layer depths [cm]	Description
1	HWSD v1.2	1000 m	Global	2	30, 100	Harmonized World Soil Database (FAO, 2012).
2	DSOLMap	250 m	Global	6	5, 15, 30, 60, 100, 200	Derived from world soil observations with ML (López-Ballesteros et al., 2023)
3	Cluster 1	100 m	Chile	6	5, 15, 30, 60, 100, 200	Based on CLSoilMaps (Dinamarca et al., 2023), clustered using unweighted **Alpha, Ks, AWC**.
4	Cluster 2	100 m	Chile	6	5, 15, 30, 60, 100, 200	Based on CLSoilMaps (Dinamarca et al., 2023), clustered using weighted **Alpha, Ks, AWC**.

SWAT+ model setup - I

Static maps:

- Watershed delineation: 50 Ha threshold for stream identification + local drainage network.
- Slops: Three classes: 0-8%, 8-30%, >30%
- Land Use: Dynamic (1990, 1999, 2004, 2009, 2013, 2018)
- Vegetation: Phenological cycle of local vegetation reconstructed based on LAI.

HRUs:

- HWSD v1.2: 1107 HRUs
- DSOLMap: 1478 HRUs
- Cluster 1: 1224 HRUs
- Cluster 2: 1162 HRUs

SWAT+ model setup - II

Climate and hydrology:

- Climate data: CR2met v2.5 (P, Tmx, Tmn, PET-HS), 0.1°, daily (Boisier, 2023).
- Routing: Variable storage.
- P partitioning: SCS-CN.

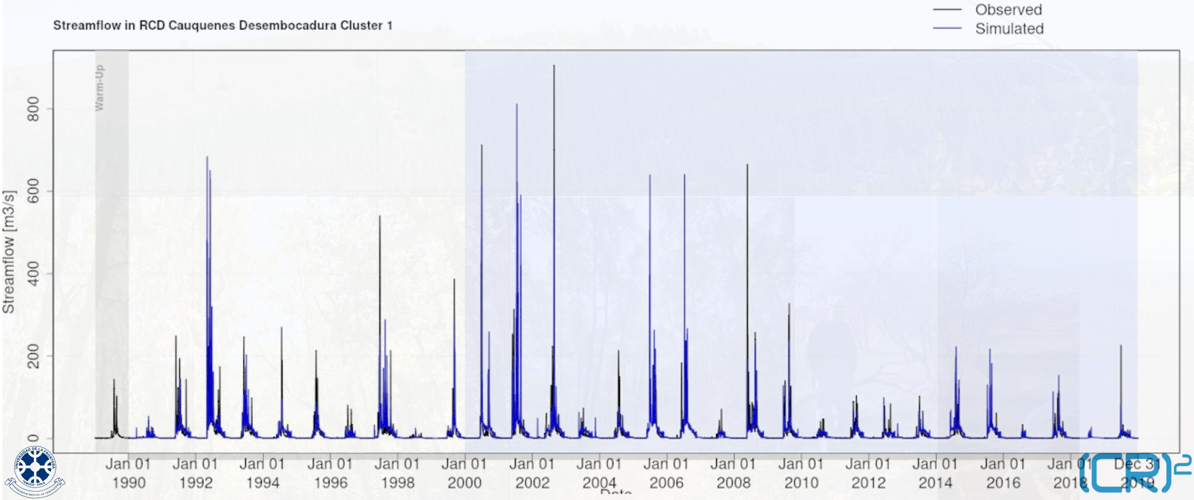
Simulation:

- Streamflow station: Cauquenes en Desembocadura (ID: 7339001).
- Warm up: 1990-1991 (2 years).
- Calibration: 2000-2019 (20 years).
- Validation: 1992-1999 (8 years).
- Calibration type: MOO with `hydroMOPSO` (Marinao-Rivas and Zambrano-Bigiarini, 2023, 2021).
- Objective 1: KGE_{lf} for low flows (Garcia et al., 2017).
- Objective 2: $APFB$ for high flows (Mizukami et al., 2019).

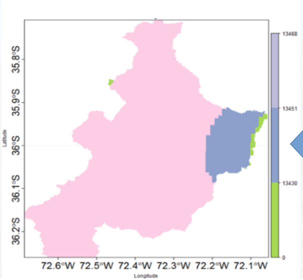
Calibration parameters (8)

N°	ID	Short description	Units	Input file	Type of change	Range
1	CN	Curve number for condition II	–	hydrology.hyd	addi	[-25, 25]
2	PERCO	Percolation coefficient	–	hydrology.hyd	repl	[0.6, 0.95]
3	LATQ_CO	Lateral soil flow coefficient - linear adjustment	–	hydrology.hyd	repl	[0, 1]
4	LAT_TIME	Lateral flow travel time	*days*	hydrology.hyd	repl	[0, 180]
5	ALPHA_BF	Baseflow alpha factor	$1/day$	aquifer.aqu	repl	[0, 1]
6	REVAP	Groundwater revap coefficient	–	aquifer.aqu	repl	[0.02, 0.2]
7	FLO_MIN	Minimum aquifer storage to allow return flow	m	aquifer.aqu	repl	[0, 10]
8	REVAP_MIN	Threshold depth of water for revap to occur	m	aquifer.aqu	repl	[0, 10]

Results: Daily streamflows for Cluster 1

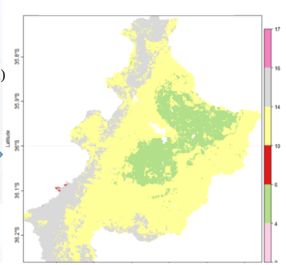


Results: High and low flows in CAL and VER



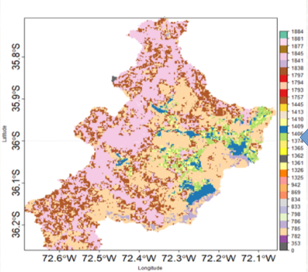
HWSD v1.2
3 soil classes

Cluster 1
6 soil classes
Alpha, Ks, AWC (non-weighted)



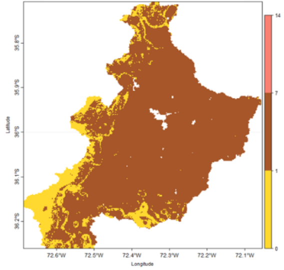
Daily CAL (2000-2020)		
Gridded soil data	**KGElf**	**APFB**
HWSD	0,52	**0,07 %**
DSOLmap	0,52	0,22 %
Cluster1	**0,57**	0,60 %
Cluster2	0,53	2,46 %

Daily VER (1990-1999)		
Gridded soil data	**KGElf**	**APFB**
HWSD	0,6	15,00 %
DSOLmap	0,57	**7,59 %**
Cluster1	**0,61**	28,76 %
Cluster2	0,55	12,93 %

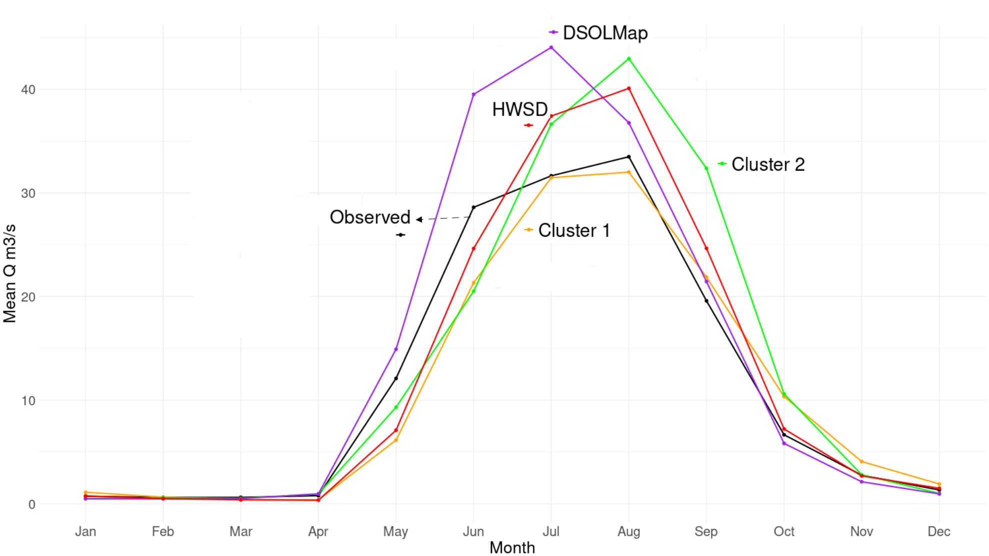


DSOLMap
32 soil classes

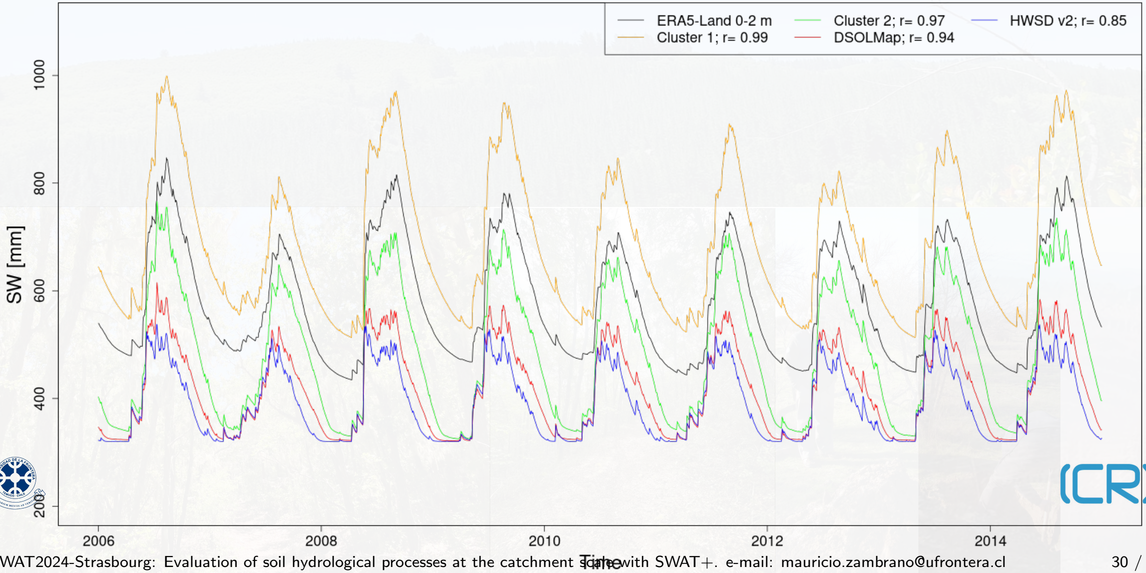
Cluster 2
3 soil classes
Alpha, Ks, AWC (weighted)



Results: Mean monthly streamflows



Results: Soil Water Content



Conclusions and Outlook

Conclusions

1. The performance of all the soil maps tested here (i.e., global, local) was good at daily time scale and not very different among them when using NSE and KGE for total streamflow (not shown here).
2. However, the simulated seasonal streamflows were closer to the observed ones when using the local soil map (Cluster 1).
3. The use of a local soil property maps (Cluster 1 and Cluster 2) resulted in a slight improvement in the representation of low flows (CAL, VAL), but high flows were not well captured during the verification period.
4. The use of a local soil property maps (Cluster 1) resulted in the best reproduction of the daily soil moisture over the whole soil profile of the catchment.

Outlook

1. To review the numerical ranges used to modify some model parameters (e.g., CN2, ALPHA_BF).
2. To add hydrological signatures (McMillan, 2020, 2021; Westerberg and McMillan, 2015) to the verification phase.
3. To implement gwflow to improve the representation of low flows..

< mauricio.zambrano [at] ufrontera.cl >

References I

Boisier, J., 2023. CR2MET: A high-resolution precipitation and temperature dataset for the period 1960–2021 in continental chile (v2.5), zenodo [data set]. doi:10.5281/zenodo.7529682.

Dinamarca, D.I., Galleguillos, M., Seguel, O., Faúndez Urbina, C., 2023. Clsoilmaps: a national soil gridded database of physical and hydraulic soil properties for chile. Scientific Data 10, 630. doi:10.5281/zenodo.7464209.

Garcia, F., Folton, N., Oudin, L., 2017. Which objective function to calibrate rainfall–runoff models for low-flow index simulations? Hydrological Sciences Journal 62, 1149–1166. doi:10.1080/02626667.2017.1308511.

Lin, Q., Li, J., Du, Z., Chen, J., Ming, Z., 2015. A novel multi-objective particle swarm optimization with multiple search strategies. European Journal of Operational Research 247, 732–744. doi:10.1016/J.EJOR.2015.06.071.

Lin, Q., Liu, S., Zhu, Q., Tang, C., Song, R., Chen, J., Coello, C.A.C., Wong, K.C., Zhang, J., 2016. Particle swarm optimization with a balanceable fitness estimation for many-objective optimization problems. IEEE Transactions on Evolutionary Computation 22, 32–46. doi:10.1109/TEVC.2016.2631279.

López-Ballesteros, A., Nielsen, A., Castellanos-Osorio, G., Trolle, D., Senent-Aparicio, J., 2023. DSOLMap, a novel high-resolution global digital soil property map for the SWAT+ model: Development and hydrological evaluation. Catena 231, 107339. doi:10.1016/j.catena.2023.107339.

References II

Marinao-Rivas, R., Zambrano-Bigiarini, M., 2021. Towards best default configuration settings for nmpso in multi-objective optimization. 2021 IEEE Latin American Conference on Computational Intelligence, LA-CCI 2021 doi:10.1109/LA-CCI48322.2021.9769844.

Marinao-Rivas, R., Zambrano-Bigiarini, M., 2023. hydroMOPSO: Multi-Objective Calibration of Hydrological Models using MOPSO. URL: https://CRAN.R-project.org/package=hydroMOPSO. r package version 0.1-3.

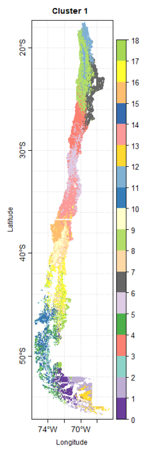
McMillan, H., 2020. Linking hydrologic signatures to hydrologic processes: A review. Hydrological Processes 34, 1393–1409. doi:10.1002/hyp.13632.

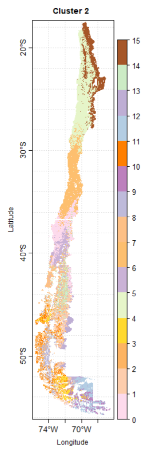
McMillan, H.K., 2021. A review of hydrologic signatures and their applications. Wiley Interdisciplinary Reviews: Water 8, e1499. doi:10.1002/wat2.1499.

Mizukami, N., Rakovec, O., Newman, A.J., Clark, M.P., Wood, A.W., Gupta, H.V., Kumar, R., 2019. On the choice of calibration metrics for "high-flow" estimation using hydrologic models. Hydrology and Earth System Sciences 23, 2601–2614. doi:10.5194/hess-23-2601-2019.

Westerberg, I., McMillan, H.K., 2015. Uncertainty in hydrological signatures. Hydrology and Earth System Sciences 19, 3951–3968. doi:10.5194/hess-19-3951-2015.

Chilean soil maps





Calibration software: hydroMOPSO

hydroMOPSO (Marinao-Rivas and Zambrano-Bigiarini, 2023, 2021) is an R package for multi-objective optimisation/calibration of hydrological models. Is based on the NMPSO algorithm (Lin et al., 2016, 2015), which combines two search mechanism (PSO and genetic operators).

Main features:

- Model-independent: can be used to calibrate **R-based models** (e.g., TUWmodel, GR-models) and **R-external models** (e.g., SWAT+, SWAT, Raven, WEAP, MODFLOW).
- Platform-independent: It can be run in **GNU/Linux**, **MacOS** and **Windows** machines.
- Computationally efficient: It takes advantage of **multi-core machines** and **network clusters** → important reduction of execution time.
- Highly configurable: It has several **fine-tuning options** and an effective default configuration. (Marinao-Rivas and Zambrano-Bigiarini, 2021).

A first version of the package is available on CRAN.

Goodness-of-fit metrics - I

1) Kling-Gupta efficiency for low flows: KGE_{lf}

Specially formulated for low streamflow by Garcia et al. (2017):

$$KGE_{lf} = \frac{KGE(Q) + KGE\left(\frac{1}{Q+\epsilon}\right)}{2}$$

where:

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad and \quad \epsilon = \mu_{obs}/100$$

and:

- $r = rPearson =$ Pearson correlation coefficient between observed and simulated values;
- $\alpha = \sigma_{sim}/\sigma_{obs}$; and
- $\beta = \mu_{sim}/\mu_{obs}$.

Goodness-of-fit metrics - II

2) Anual Peak Flow Bias (APFB), [%]

Proposed by Mizukami et al. (2019) to identify differences in high streamflow values.

$$APFB = \sqrt{\left(\frac{\overline{Qmax_{sim}}}{\overline{Qmax_{obs}}} - 1\right)^2}$$

where:

- $\overline{Qmax_{sim}}$: mean of the simulated annual maximum streamflow series.
- $\overline{Qmax_{obs}}$: mean of the observed annual maximum streamflow series.

hydroMOPSO default configuration

`hydroMOPSO` implements NMPSO (Lin et al., 2016), a novel multi-objective algorithm that combines two search mechanisms to maintain the diversity of the population and accelerate its convergence towards the Pareto-optimal front (POF). The two mechanisms are based on PSO and genetic operations. A balanceable fitness estimation (BFE) method is used to rank particles in an external archive, in order to provide an effective guidance to the true POF, while keeping diversity among particles.

Marinao-Rivas and Zambrano-Bigiarini (2021) defined a default configuration for the NMPSO algorithm, based on different tests. Sixteen different combinations were tested, made from: i) the swarm size (N), ii) the maximum number of particles in the external archive (N_e), and iii) the maximum amount of genetic operations in the external archive (max_{go})

The default configuration established in this study was:

$$N = 10 \;\; particles$$

$$N_e = 100 \;\; particles$$

$$max_{go} = 50 \;\; crossovers/mutations$$