Multi-period and multi-variable calibration of SWAT+ using gridded input datasets and a novel R package

International SWAT Conference (Aarhus, Denmark)

June 30th, 2023

Mauricio Zambrano-Bigiarini[1,2], Rodrigo Marinao Rivas[2,3]

[1] Department of Civil Engineering, Universidad de La Frontera, Temuco, Chile.
[2] Center for Climate and Resilience Research (CR2), Santiago, Chile.
[3] MSc(c) in Engineering Sciences, Universidad de La Frontera, Temuco, Chile.

Motivation

- Several gridded datasets have become available in the last decades on a global scale, with increasing spatial and temporal resolution and low latency times (Lettenmaier et al., 2015; Sheffield et al., 2018; Chawla et al., 2020)

- These products have opened new opportunities to advance Earth Sciences modelling

Motivation

- Different multi-objective optimisation (MOO) algorithms and calibration approaches have been used over decades to obtain consistent parameters for hydrological model applications.

- MOO generates a set of solutions that represent the Pareto optimal front, which is a set of solutions that cannot be improved in one objective without sacrificing performance in another objective. This provides a more diverse set of parameter sets than SOO algorithms, which typically generate a single "pseudo-optimal" solution.

- MOO can help identify trade-offs between different objectives, allowing decision-makers to make informed choices on the modelling processes based on their priorities.

- Multi-objective calibration is able to exploit all of the information about the physical system contained in observed time series obtained from diverse data sources (Yapo et al., 1998) (e.g., Q, SM, ET, GPP, NO_3).

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).

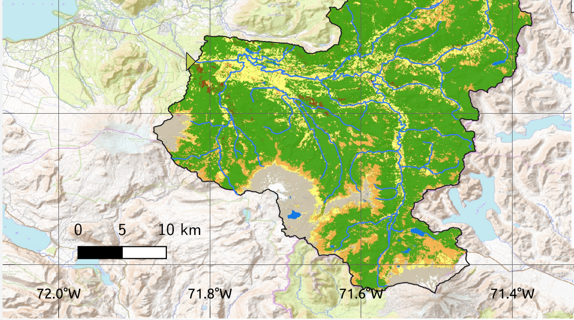
Main features:

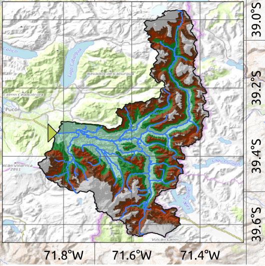
- Model-independent: can be used to calibrate **R-based models** (e.g., TUWmodel, GR-models) and **R-external models** (e.g., SWAT+, SWAT, Raven, WEAP, MODFLOW).
- Platform-independent: It can be run in **GNU/Linux**, **MacOS** and **Windows** machines.
- Computationally efficient: It takes advantage 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).

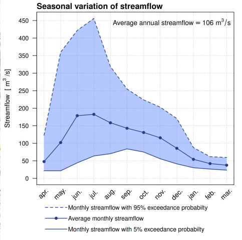
A first version of the package is available on CRAN.

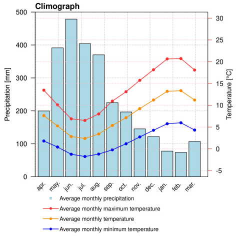
Study area: Trancura antes de Llafenco catchment

- **Andean sub-basin**
- **Lat: 39.33°S**
- **Pluvio-nival regime**
- **Area: 1380 km2**
- **P_{annual}: 2 700 mm**
- **T_{annual}: 3-14°C**









Goodness-of-fit metrics

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

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

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

2) Pearson correlation coefficient: r

It measures whether the shape of two ts are similar or not, ignoring bias.

$$r = \frac{\text{cov}(sim, obs)}{\sigma_{sim}\sigma_{obs}}$$

Case study 1: Multi-period calibration (1992-2014)

Objective 1: $KGE_{lf}(Q1)$

- Period: Wet (1992-2009)
- Variable: Streamflows
- Time step: Daily

Objective 2: $KGE_{lf}(Q2)$

- Period: Dry (2010-2014)
- Variable: Streamflows
- Time step: Daily

- Streamflow gauge station: (*Trancura antes de Llafenco*, BNA code: 9414001, Lat: -39.33°, Lon: -71.77°).
- GoF: Kling-Gupta efficiency for low flows (Garcia et al., 2017).

Case study 1 (multi-period): calibrated parameters (12)

ID	Short description	Units	Archive	Type of change	Range
`latq_co`	Lateral soil flow coefficient - linear adjustment	–	`hydrology.hyd`	repl	0-0.05
`perco`	Percolation coefficient	–	`hydrology.hyd`	repl	0-0.95
`epco`	Plant water uptake factor	–	`hydrology.hyd`	repl	0-1
`revap_min`	Threshold depth of water for revap to occur	m	`aquifer.aqu`	repl	0-10
`flo_min`	Minimum aquifer storage to allow return flow	m	`aquifer.aqu`	repl	0-10
`alpha_bf`	Baseflow alpha factor	$1/day$	`aquifer.aqu`	repl	0-1
`revap`	Groundwater revap coefficient	–	`aquifer.aqu`	repl	0.02-0.2
`rchg_dp`	Deep aquifer percolation fraction	–	`aquifer.aqu`	repl	0-1
`fall_tmp`	Snowfall temperature	$\circ C$	`snow.sno`	repl	-5-5
`melt_tmp`	Snow melt base temperature	$\circ C$	`snow.sno`	repl	-5-5
`melt_max`	Maximum melt rate for snow during year	$mm/^{\circ}C/day$	`snow.sno`	repl	0-10
`melt_min`	Minimum melt rate for snow during year	$mm/^{\circ}C/day$	`snow.sno`	repl	0-10

Case study 2 (multi-variable): Multi-variable calibration (1992-2014)

Objective N° 1: $KGE_{lf}(Q)$

- Variable: Streamflow (Q)
- GoF: Modified Kling-Gupta efficiency proposed by (Garcia et al., 2017). For **low flows**.
- Time step: Daily
- Source of observed data: River gauge (*Trancura antes de Llafenco*, BNA code: 9414001, Lat: -39.33°, Lon: -71.77°).

Objective N° 2: $rPearson(ET)$

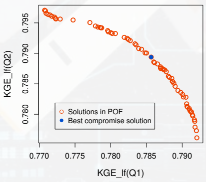
- Variable: Actual evapotranspiration (ET)
- GoF: Pearson correlation coefficient (*rPearson*)
- Time step: Monthly
- "Observed" data source: Gridded ET products
  - Case study B.1: ERA5-Land ($\sim$ 10 km)
  - Case study B.2: SSEbop ($\sim$ 1 km)

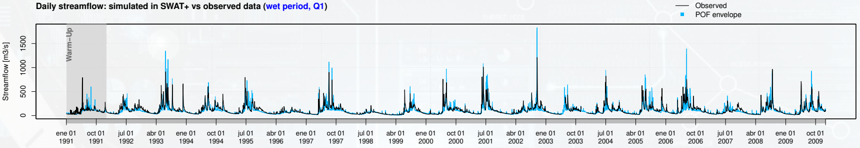
Case study 2 (multi-variable): Calibrated parameters (12)

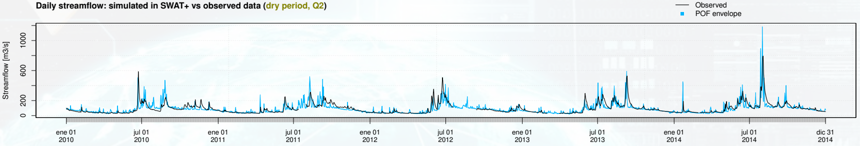
ID	Short description	Units	Archive	Type of change	Range
latq_co	Lateral soil flow coefficient - linear adjustment	—	hydrology.hyd	repl	0-0.05
perco	Percolation coefficient	—	hydrology.hyd	repl	0-0.95
epco	Plant water uptake factor	—	hydrology.hyd	repl	0-1
revap_min	Threshold depth of water for revap to occur	m	aquifer.aqu	repl	0-10
flo_min	Minimum aquifer storage to allow return flow	m	aquifer.aqu	repl	0-10
alpha_bf	Baseflow alpha factor	$1/day$	aquifer.aqu	repl	0-1
revap	Groundwater revap coefficient	—	aquifer.aqu	repl	0.02-0.2
rchg_dp	Deep aquifer percolation fraction	—	aquifer.aqu	repl	0-1
fall_tmp	Snowfall temperature	$\circ C$	snow.sno	repl	-5-5
melt_tmp	Snow melt base temperature	$\circ C$	snow.sno	repl	-5-5
tmp_base	Minimum (base) temperature for plant growth	$\circ C$	plants.plt	addi	-5-5
tmp_opt	Optimal temperature for plant growth	$\circ C$	plants.plt	addi	-5-5

Case study 1: Multi-period calibration, wet and dry

Pareto optimal front envelope

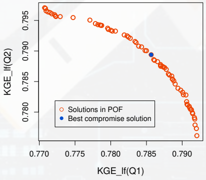


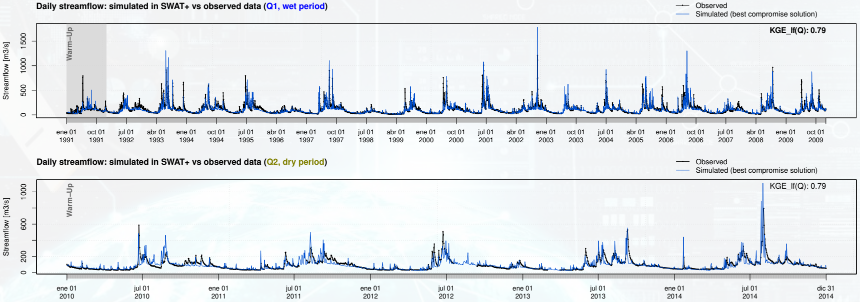




Case study 1: Multi-period calibration, wet and dry

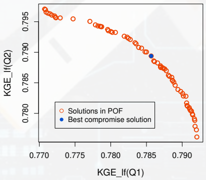
Best compromise solution

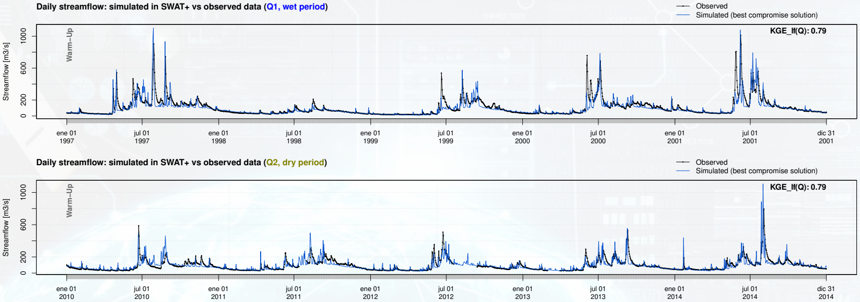




Case study 1: Multi-period calibration, wet and dry (zoom)

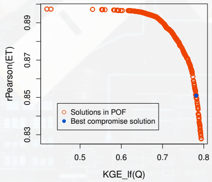
Best compromise solution

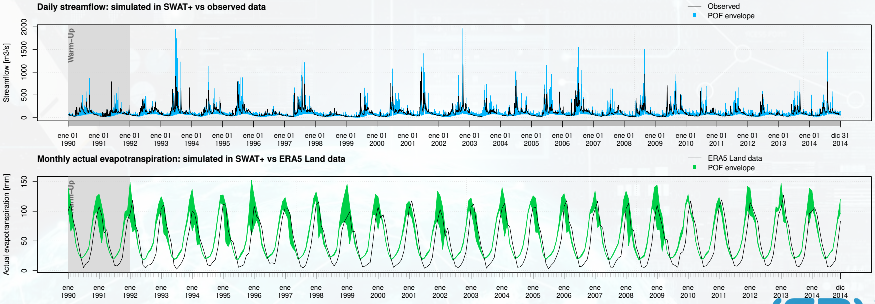




Case study 2.1: Multi-variable calibration (daily Q and monthly ETa from ERA5-Land)

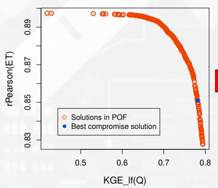
Pareto optimal front envelope

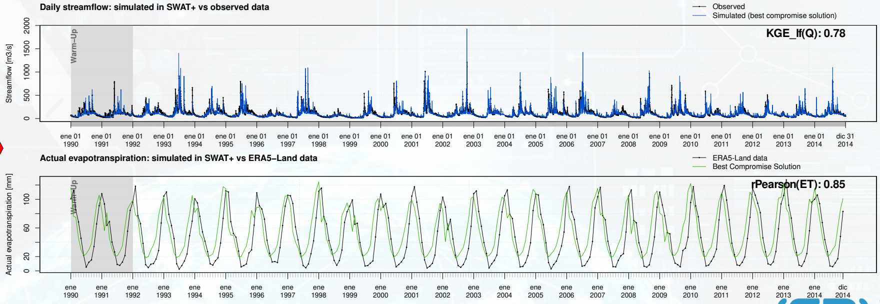




Case study 2.1: Multi-variable calibration (daily Q and monthly ETa from ERA5-Land)

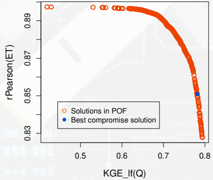
Best compromise solution

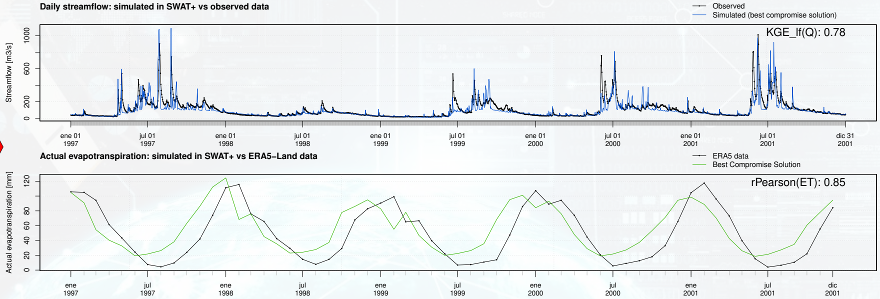




Case study 2.1: Multi-variable calibration (zoom)

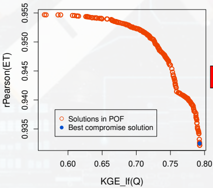
Best compromise solution

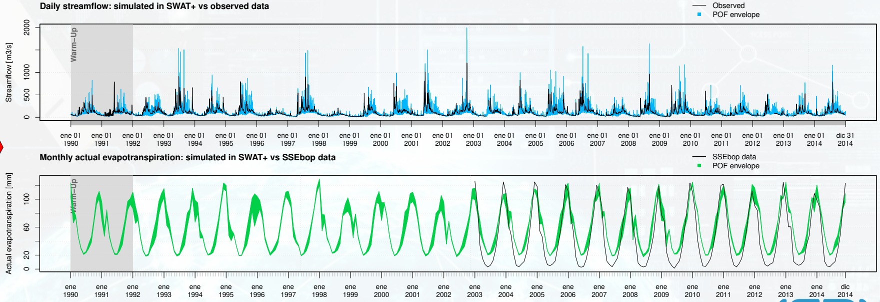




Case study 2.2: Multi-variable calibration (daily Q and monthly ETa from SSEBoP

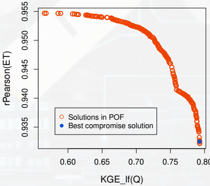
Pareto optimal front envelope

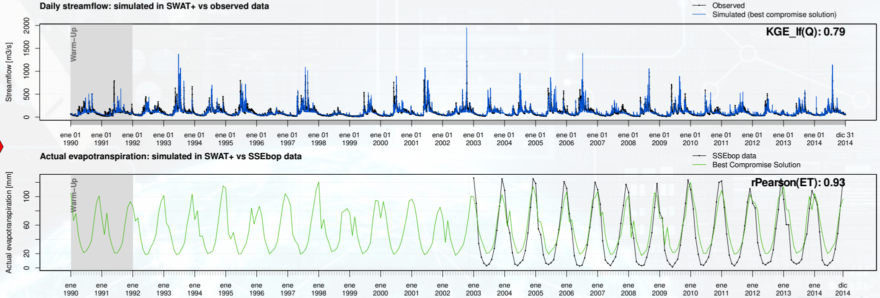




Case study 2.2: Multi-variable calibration (daily Q and monthly ETa from SSEBoP

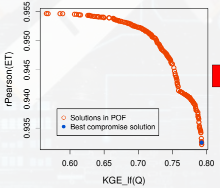
Best compromise solution

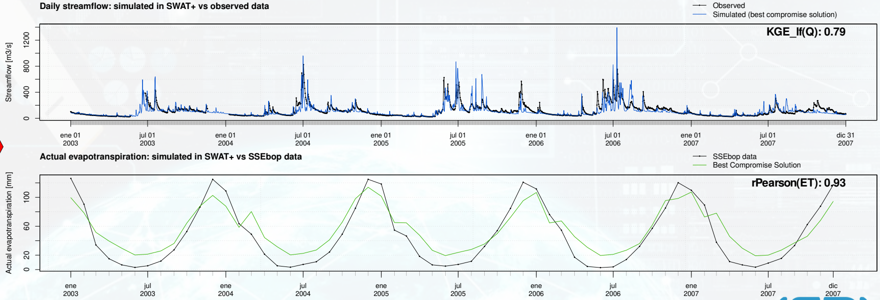




Case study 2.2: Multi-variable calibration (daily Q and monthly ETa from SSEBoP (zoom)

Best compromise solution





Conclusions

1. In the multi-period calibration problem, the **best compromise solution** obtained with hydroMOPSO was able to simulataneously provide a good representation of **two contrasting temporal periods**: wet (1992-2009, $KGE_{lf}(Q1) = 0.79$) and dry (2010-2014, $KGE_{lf}(Q2) = 0.79$).
2. In the multi-variable calibration problem with ERA5-Land ETa data, the **best compromise solution** obtained with hydroMOPSO provided a good representation of daily streamflows ($KGE_{lf}(Q) = 0.78$) and monthly actual evapotranspiration ($r(ETa) = 0.85$).
3. In the multi-variable calibration problem with SSEBoP data, the **best compromise solution** obtained with hydroMOPSO provided a good representation of daily streamflows ($KGE_{lf}(Q) = 0.79$) and monthly actual evapotranspiration ($r(ETa) = 0.93$).
4. Given the model-independent nature of the package and the versatility of the R environment, we hope the broader community of hydrologists and Earth scientists will use `hydroMOPSO` for the multi-objective calibration of a wide class of hydrological and environmental models.

< mauricio.zambrano [at] ufrontera.cl >

References I

Chawla, I., Karthikeyan, L., Mishra, A.K., 2020. A review of remote sensing applications for water security: Quantity, quality, and extremes. Journal of Hydrology 585, 124826. doi:10.1016/j.jhydrol.2020.124826.

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.

Lettenmaier, D.P., Alsdorf, D., Dozier, J., Huffman, G.J., Pan, M., Wood, E.F., 2015. Inroads of remote sensing into hydrologic science during the WRR era. Water Resources Research 51, 7309–7342. doi:10.1002/2015WR017616.

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.

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.

References II

Netzel, P., Slopek, J., 2021. Comparison of different implementations of a raster map calculator. Computers & Geosciences 154, 104824. doi:10.1016/j.cageo.2021.104824.

Sheffield, J., Wood, E.F., Pan, M., Beck, H., Coccia, G., Serrat-Capdevila, A., Verbist, K., 2018. Satellite remote sensing for water resources management: Potential for supporting sustainable development in data-poor regions. Water Resources Research 54, 9724–9758. doi:10.1029/2017WR022437.

Silva-Coira, F., Paramá, J.R., Ladra, S., López, J.R., Gutiérrez, G., 2020. Efficient processing of raster and vector data. PLoS One 15, e0226943. doi:10.1371/journal.pone.0226943.

Yapo, P.O., Gupta, H.V., Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. Journal of Hydrology 204, 83–97. doi:10.1016/S0022-1694(97)00107-8.

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.

? 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:

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

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

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