Analysing the Capability of a Catchment's Spectral Signature to Regionalize Hydrological Parameters

Laura Fragoso-Campón [1], Pablo Durán-Barroso [2] and Elia Quirós [3]

[1] Department of Graphic Expression, Universidad de Extremadura, Cáceres, 10003, Spain

[2] Department of Construction, Universidad de Extremadura, Cáceres, 10003, Spain

[3] Department of Graphic Expression, Universidad de Extremadura, Cáceres, 10003, Spain

Corresponding Author: Laura Fragoso-Campón. Universidad de Extremadura, Escuela Politécnica, Avda. de la Universidad s/n, Cáceres, 10003, Spain. E-mail: laurafragoso@unex.es

Contenido

LIST OF ACRONYMS

AEMET	Spanish Meteorological Agency
Best-Par	Best combination of the calibrated parameters in each catchment
BETA	Parameter that determines the relative contribution to runoff from rain or snowmelt
BI	Brightness index
CFMAX	Degree-Δt factor for snow melting that starts if temperatures are above TT.
CFR	Refreezing coefficient
CI	Colour index
CR	Climatic regionalization
CWH	Water holding capacity of melted water that refreezes again when temperature decrease below TT
DEM	Digital elevation models
ESDC	European Soil Data Centre
FC	Maximum soil moisture storage
GAP	Genetic algorithm and Powell optimization
GGR	Global ground regionalization.
GLCM	Grey-level cooccurrence matrix
GNDVI	Green normalized difference vegetation index
GoF	Goodness-of-fit
GR	Ground regionalization
GW	Groundwater boxes
IGN	Spanish National Geographic Institute
IWMI	International Water Management Institute
K_0	Recession coefficient of STZ
K_1	Recession coefficient of SUZ
K_2	Recession coefficient of SLZ
KGE	Kling-Gupta Efficiency
LogReff	Model efficiency for log (Q)
LP	Soil moisture value above which actual evapotranspiration (AET) reaches PET.
MAXBAX	Length of triangular weighting function
MeanDiff	Mean difference
MSAVI2	Second modified soil-adjusted vegetation index
Mtry	Number of variables to be selected when growing the trees
NDI45	Normalized difference index
NDVI	Mean normalized difference vegetation index

NDWI	Normalized difference water index
NIR	Near-infrared spectral bands
NSE	Nash-Sutcliffe efficiency
Ntree	Number of decision trees
PCALT	Increase of precipitation with height increment
PERC	Maximum percolation from the SUZ to the SLZ
PGR	Physical ground regionalization
R^2	Coefficient of determination
Reff	Model efficiency
ReffPeak	Efficiency for peak flows
Reg-Par	Regionalized parameters
RF	Random forest algorithm
S1	Sentinel-1
S2	Sentinel-2
SAIH	Automatic Hydrological Information Systems
SAR	Synthetic Aperture Radar
SAVI	Soil-adjusted vegetation index
SFCF	Snowfall correction factor
SGR	Spectral ground regionalization
SLZ	Storage in the soil lower zone
SNAP	Sentinel Application Platform
SP	Seasonal variability in degree-Δt factor
STZ	Storage in the soil top zone
SUZ	Storage in the soil upper zone
SWIR	Shortwave infrared spectral bands
TCALT	Decrease of temperature with height increment
TT	Threshold temperature at which the accumulation of precipitation is in the form of snow below it
UZL	Maximum percolation from the STZ to the SUZ
VIS	Visible spectral bands
VolumeError	Volume error

TABLES SI

TABLE SI.1 Studies Performing Model Parameter Regionalization

Study	Hydrological Method	Spatial analysis	Regionalization model	Type	Scale†	Study Area	GoF‡ function	Validation	Regionalization
Merz and Blöschl (2004)	HBV	Lumped	Similarity-based	Global and local regression, nearest neighbour, Kringing	R	Austria	NSE	0.63	0.56
Hundecha and Bárdossy (2004)	HBV	Semi-distributed	Regression	Transfer equations	R	Germany	R^2	0.79-0.90	0.76-0.92
Booij (2005)	HBV	Semi-distributed	Regression	Transfer equations	R	The Netherlands	R^2	0.87	0.92
Götzinger and Bárdossy (2007)	HBV	Distributed	Regression	Transfer equations, Lipschitz and monotony condition.	R	Germany	NSE	0.53	0.5
Parajka et al. (2007)	HBV	Semi-distributed	Similarity-based	Spatial proximity	R	Austria	NSE	0.66-0.69	0.62-0.66
Oudin et al. (2008)	GR4J	Lumped	Regression and Similarity-based	Regression, physical similarity and spatial proximity	R	France	NSE	0.78	0.71-0.74
Jin et al. (2009)	HBV	Lumped	Similarity-based	Spatial proximity		China	NSE	0.78	0.72
Masih et al. (2010)	HBV	Semi-distributed	Hydrological similarity	-	R	Iran	NSE	0.62	0.47[§1]
Arsenault and Brissette (2016)	HSAMI	Lumped	Similarity-based	Physical similarity and spatial proximity	R	Canada	NSE	0.8	0.7
Beck et al. (2016)	HBV	Lumped & distributed	Similarity-based	Physical similarity	G	Global	Own aggregate objective fuction	0.64	0.49
						Arid		0.6	0.48
Jillo et al. (2017)	HBV	Lumped	Regression	Transfer equations	R	Ethiopia	KGE	0.63	0.5

Study	Hydrological Method	Spatial analysis	Regionalization model	Type	Scale†	Study Area	GoF‡ function	Validation	Regionalization
Jayathilake and Smith (2019)	Probability Distributed Model (PDM)	Distributed	Hydrological similarity	-	R	USA	NSE	0.57	NA
Zamoum and Souag-Gamane (2019)	GR2M	Lumped	Similarity-based	Clustering PCA and SOM	R	Algeria	NSE	> 0.7	0.52[§3]
Beck et al. (2020)	HBV	Distributed	Regression	Transfer equations	G	Global	KGE	0.69	0.46
						Arid	KGE	0.30[§4]	-0.05[§5]

†Regional (R) or Global (G) Scale
‡Goodness of fit function

§1 Estimated from Table 7 in Masih et al. (2010) using the 7th best ranked
§2 Estimated from Table 4 in Tegegne and Kim (2018)
§3 Estimated from Table 5 in Zamoum and Souag-Gamane (2019)
§4 Visual interpretation of Figure 5 in Beck et al. (2020)
§5 Visual interpretation of Figure 2 in Beck et al. (2020)

TABLE SI.2 List of predictors used in the regionalization of the HBV parameters

Thematic description	Predictor	Description
Climatic Regionalization		
Topography	DEM25_mean	Mean value of the catchment zonal statistic on the DEM25 (digital elevation model of the terrain at 25 m spatial resolution)
	gosDEM_mean	Mean value of the catchment zonal statistic on the gosDEM (AEMET ground observation stations digital elevation model)
	D_elevation	Value representing the diference between DEM25_mean and gosDEM_mean
	Delta_mean	Mean value of the catchment zonal statistic on an auxiliary raster of differeces between DEM25 and gosDEM)
	Delta_sd	Standar deviation value of the catchment zonal statistic on raster (DEM25-gosDEM)
	DE_mean	Mean value of the catchment zonal statistic on the DE (Euclideam distance to the nearest AEMET ground observation station)
	DE_sd	Standar deviation value of the catchment zonal statistic on the DE (Euclideam distance to the nearest AEMET ground observation station)
Climatic	T_mean	Annual mean daily temperature of the long-term series (2008-2019)
	P_mean	Annual mean total precipitation of the long-term series (2008-2019)
Ground Regionalization		
Vegetation coverage	TCCe	Percent of tree canopy cover of evergreen forest (needleleaf and broadleaf)
	TCCb	Percent of tree canopy cover of deciduous broadleaf forest
	TCCd	Percent of tree canopy cover of *Dehesas* in the watershed (*Dehesas* are evergreen broadleaf forest with 20<TCC<60%)
	CCs	Percent of evergreen shrubs cover
	CCp	Percent of herbaceous vegetation cover
	CCagr	Percent of agricultural land cover
	CCi	Percent of impervious surfaces
Morphology	Rf	Horton form factor
	Rc	Gravellius compactness coefficient
	Re	Elongation ratio
	Slope_m	Mean value of the catchment zonal statistic on the slope derived of the DEM25 (digital elevation model of the terrain at 25 m spatial resolution)
	Slope_sd	Standar deviation value of the catchment zonal statistic on the slope derived of the DEM25 (digital elevation model of the terrain at 25 m spatial resolution)
Topsoil properties	Silt_m	Mean value of the catchment zonal statistic on silt content in topsoil (0-20 cm)
	Silt_sd	Standard deviation value of the catchment zonal statistic on silt content in topsoil (0-20 cm)

Thematic description	Predictor	Description
	Sand_m	Mean value of the catchment zonal statistic on sand content in topsoil (0-20 cm)
	Sand_sd	Standard deviation value of the catchment zonal statistic on sand content in topsoil (0-20 cm)
	Clay_m	Mean value of the catchment zonal statistic on clay content in topsoil (0-20 cm)
	Clay_sd	Standard deviation value of the catchment zonal statistic on clay content in topsoil (0-20 cm)
	Coarse_m	Mean value of the catchment zonal statistic on coarse fragements content in topsoil (0-20 cm)
	Coarse_sd	Standard deviation value of the catchment zonal statistic on coarse fragements content in topsoil (0-20 cm)
	Bulk_m	Mean value of the catchment zonal statistic on Bulk density in topsoil (0-20 cm)
	Bulk_sd	Standard deviation value of the catchment zonal statistic on Bulk density in topsoil (0-20 cm)
	awc_m	Mean value of the catchment zonal statistic on the available water capacity (AWC) for the topsoil fine earth fraction (0-20 cm)
	awc_sd	Standard deviation value of the catchment zonal statistic on the available water capacity (AWC) for the topsoil fine earth fraction (0-20 cm)
Subsoil properties	WRB_RG	Percent of Regosols (RG)
	WRB_LP	Percent of Leptisols (LP)
	WRB_CM	Percent of Cambisols (CM)
	WRB_FL	Percent of Fluvisols (FL)
	WRB_AC	Percent of Acrisols (AC)
	WRB_LV	Percent of Luvisols (LV)
	HG_1	Percent of hydro-geological class type HG1
	HG_2	Percent of hydro-geological class type HG2
	HG_4	Percent of hydro-geological class type HG4
	DIMP_D	Percent of deep depth (> 80 cm) to the impermeable layer
	DIMP_S	Percent of shallow depth (< 80 cm) to the impermeable layer
	AWCSUB_#	Percent of NaN subsoil available water capacity
	AWCSUB_VL	Percent of very low subsoil available water capacity
	AWCSUB_L	Percent of low subsoil available water capacity
	AWCSUB_M	Percent of medium subsoil available water capacity
	AWCSUB_	Percent of high subsoil available water capacity

Thematic description	Predictor	Description
	H	
Spectral Sentinel- 1 (SAR)	VH_i_j	j value of the backscattering coefficient of VH polarization in i image date adquision (i= s for summer; i=w for winter) (j=m for mean; j=sd for standard deviation; j=q25 for 0.25 quantil; j=q50 for 0.50 quantil;j=q75 for 0.75 quantil)
	VV_i_j	j value of the backscattering coefficient of VV polarization in i image date adquision (i= s for summer; i=w for winter) (j=m for mean; j=sd for standard deviation; j=q25 for 0.25 quantil; j=q50 for 0.50 quantil;j=q75 for 0.75 quantil)
	VH_i_g_j	j value of the t texture metric of VH polarization in i image date adquision (i= s for summer; i=w for winter) (j=m for mean; j=sd for standard deviation; j=q25 for 0.25 quantil; j=q50 for 0.50 quantil;j=q75 for 0.75 quantil) (g=Con for contrast; g=Dis for Dissimilarity; g=GLCMm for GLCMMean; g=GLCMv for GLCMVariance)
	VV_i_g_j	j value of the t texture metric of VV polarization in i image date adquision (i= s for summer; i=w for winter) (j=m for mean; j=sd for standard deviation; j=q25 for 0.25 quantil; j=q50 for 0.50 quantil;j=q75 for 0.75 quantil) (g=Con for contrast; g=Dis for Dissimilarity; g=GLCMm for GLCMMean; g=GLCMv for GLCMVariance)
Spectral Sentinel- 2 (Optical)	Bk_j	j value of the k Sentinel- 2 band (j=m for mean; j=sd for standard deviation; j=q25 for 0.25 quantil; j=q50 for 0.50 quantil;j=q75 for 0.75 quantil) (k= 1, 2, 3, 4, 5, 6, 7, 8, 8A, 11 and 12)
	Bk_g_j	j value of the t texture metric of the k Sentinel- 2 band (j=m for mean; j=sd for standard deviation; j=q25 for 0.25 quantil; j=q50 for 0.50 quantil;j=q75 for 0.75 quantil) (g=Con for contrast; g=Dis for Dissimilarity; g=GLCMm for GLCMMean; g=GLCMv for GLCMVariance)(k= 1, 2, 3, 4, 5, 6, 7, 8, 8A, 11 and 12)
	SI_m_j	j value of the m Soil Index (j=m for mean; j=sd for standard deviation; j=q25 for 0.25 quantil; j=q50 for 0.50 quantil;j=q75 for 0.75 quantil) (m = BI for brightness index; m=CI for colour index)
	VI_n_j	j value of the n Vegetation Index (j=m for mean; j=sd for standard deviation; j=q25 for 0.25 quantil; j=q50 for 0.50 quantil;j=q75 for 0.75 quantil) (n=GNDVI for the green normalized difference vegetation index; n=MSAVI2 for

Thematic description	Predictor	Description
		the second modified soil-adjusted vegetation index); n=NDI45 for the normalized difference index; n=NDVI for the normalized difference vegetation index; n= SAVI for the soil-adjusted vegetation index)
	WI_p_j	j value of the p Water Index (j=m for mean; j=sd for standard deviation; j=q25 for 0.25 quantil; j=q50 for 0.50 quantil;j=q75 for 0.75 quantil) (p=NDWI for the normalized difference water index)

TABLE SI.3 Goodness of fit functions considered in this study

Function name	Description	Equation	Perfect fit value
Reff	Model efficiency	$1-\frac{\sum(Q_{osb}-Q_{sim})^2}{\sum(Q_{osb}-\overline{Q_{obs}})^2}$	1
KGE	Kling-Gupta Efficiency	$1-\sqrt{((r-1)^2+(\alpha-1)^2+(\beta-1)^2)}$ r: Pearson correlation coefficient between Q_{sim} and Q_{osb} α: ratio between standard deviation of Q_{sim} and Q_{osb} β: ratio between $\overline{Q_{sim}}$ and $\overline{Q_{osb}}$	1
R^2	Coefficient of determination	$\frac{\left(\sum(Q_{osb}-\overline{Q_{obs}})\cdot(Q_{sim}-\overline{Q_{sim}})\right)^2}{\sum(Q_{osb}-\overline{Q_{obs}})^2\cdot\sum(Q_{sim}-\overline{Q_{sim}})^2}$	1
LogReff	Model efficiency for log (Q)	$1-\frac{\sum(\ln Q_{osb}-\ln Q_{sim})^2}{\sum(\ln Q_{osb}-\overline{\ln Q_{obs}})^2}$	1
MeanDiff	Mean difference	$\frac{\sum(Q_{osb}-Q_{sim})}{n}\cdot 365$ n: number of times steps	0
VolumeError	Volume error	$1-\frac{\lvert\sum(Q_{osb}-Q_{sim})\rvert}{\sum(Q_{osb})}$	1
ReffPeak	Efficiency for peak flows	$1-\frac{\sum(peakQ_{osb}-peakQ_{sim})^2}{\sum(peakQ_{osb}-\overline{peakQ_{obs}})^2}$	1
Objective	Goodness of fit measure self-defined in this study	0.2·Reff + 0.6·KGE + 0.2·Reff Peak	1

TABLE SI.4 Best-Par Cluster Statistics (median and standard deviation) of HBV parameters

	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5	
variable	median	sd	median	sd	median	sd	median	sd	median	sd
PCALT	10.25	0.48	23.77	0.70	10.00	2.28	17.00	2.34	10.31	1.43
TCALT	0.60	0.01	0.60	0.01	0.64	0.02	0.60	0.03	0.63	0.02
TT	0.28	0.14	-2.00	0.09	-1.32	1.07	0.34	0.05	0.34	0.52
CFMAX	3.04	0.48	2.93	0.70	1.73	0.65	2.39	0.56	2.02	1.46
CFR	0.05	0.00	0.05	0.00	0.05	0.00	0.05	0.00	0.05	0.00
CWH	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.10	0.00
SFCF	0.76	0.15	0.52	0.17	0.50	0.00	0.90	0.17	0.74	0.20
FC	211.95	108.28	243.79	38.94	504.77	164.03	63.79	125.31	182.40	124.85
LP	0.45	0.09	0.36	0.21	0.52	0.36	0.90	0.23	0.63	0.25
BETA	4.75	0.42	4.00	0.68	2.92	1.92	5.00	0.30	4.98	1.99
PERC	2.17	1.38	1.50	0.67	3.00	0.80	1.73	0.88	1.62	1.89
UZL	0.50	1.63	21.54	2.05	24.89	2.83	11.47	8.16	0.87	1.47
K0	0.50	0.00	0.50	0.03	0.50	0.01	0.50	0.00	0.50	0.00
K1	0.11	0.07	0.14	0.03	0.08	0.01	0.14	0.07	0.14	0.07
K2	0.04	0.03	0.05	0.00	0.01	0.00	0.02	0.01	0.07	0.04
MAXBAS	2.30	0.17	1.30	0.07	2.05	0.23	2.02	0.03	2.20	0.62
Cet	0.07	0.11	0.05	0.05	0.06	0.16	0.25	0.11	0.04	0.14

TABLE SI.5 RMSE of the regionalized parameters compared to the calibrated parameters

	RMSE			
Parameter	CR	SGR	PGR	GGR
PCALT	3.59	N/A	N/A	N/A
TCALT	0.02	N/A	N/A	N/A
TT	0.56	N/A	N/A	N/A
CFMAX	0.94	N/A	N/A	N/A
SFCF	0.18	N/A	N/A	N/A
FC_1	N/A	168.02	157.84	162.63
FC_2	N/A	101.43	94.59	97.88
FC_3	N/A	24.02	22.44	23.06
LP	N/A	0.26	0.26	0.26
BETA	N/A	1.39	1.34	1.36
PERC	N/A	1.24	1.17	1.18
UZL	N/A	6.66	6.16	6.62
K0	N/A	0.03	0.03	0.03
K1	N/A	0.06	0.06	0.06
K2	N/A	0.03	0.03	0.03
MAXBAS	N/A	0.36	0.39	0.36
Cet	N/A	0.13	0.10	0.12

FIGURES SI

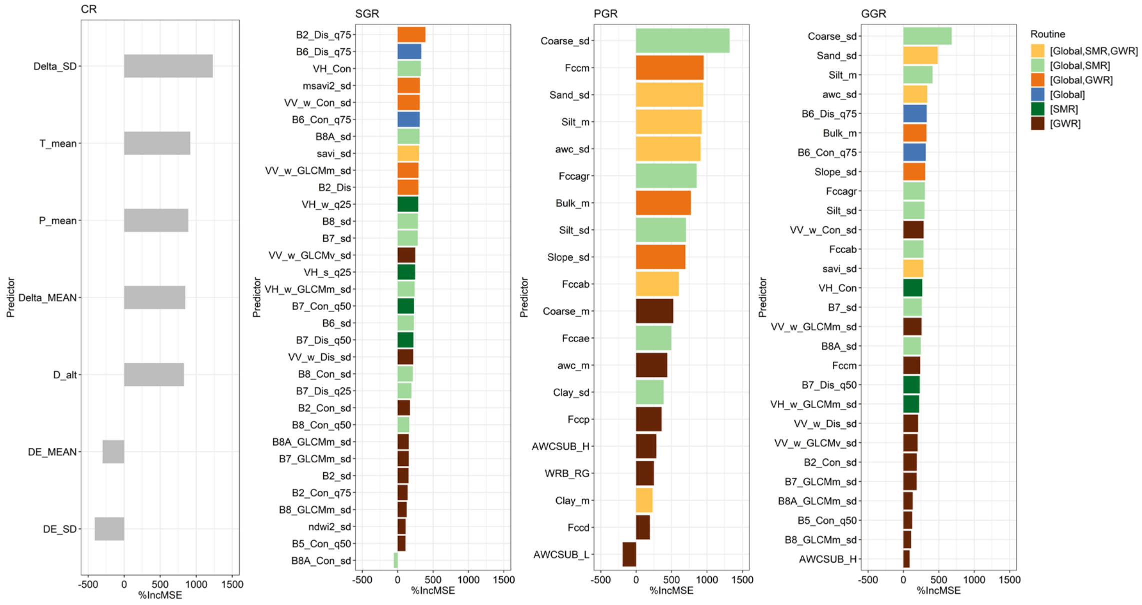


FIGURE SI.1 Importance of the predictors in the regressions. %IncMSE represents the increase of the mean squared error when a predictor is randomly permuted. The colour in the legend shows the level of importance of the prediction of each routine: the global importance computation, or specifically for the soil moisture routine (SMR) or the groundwater routine (GWR)

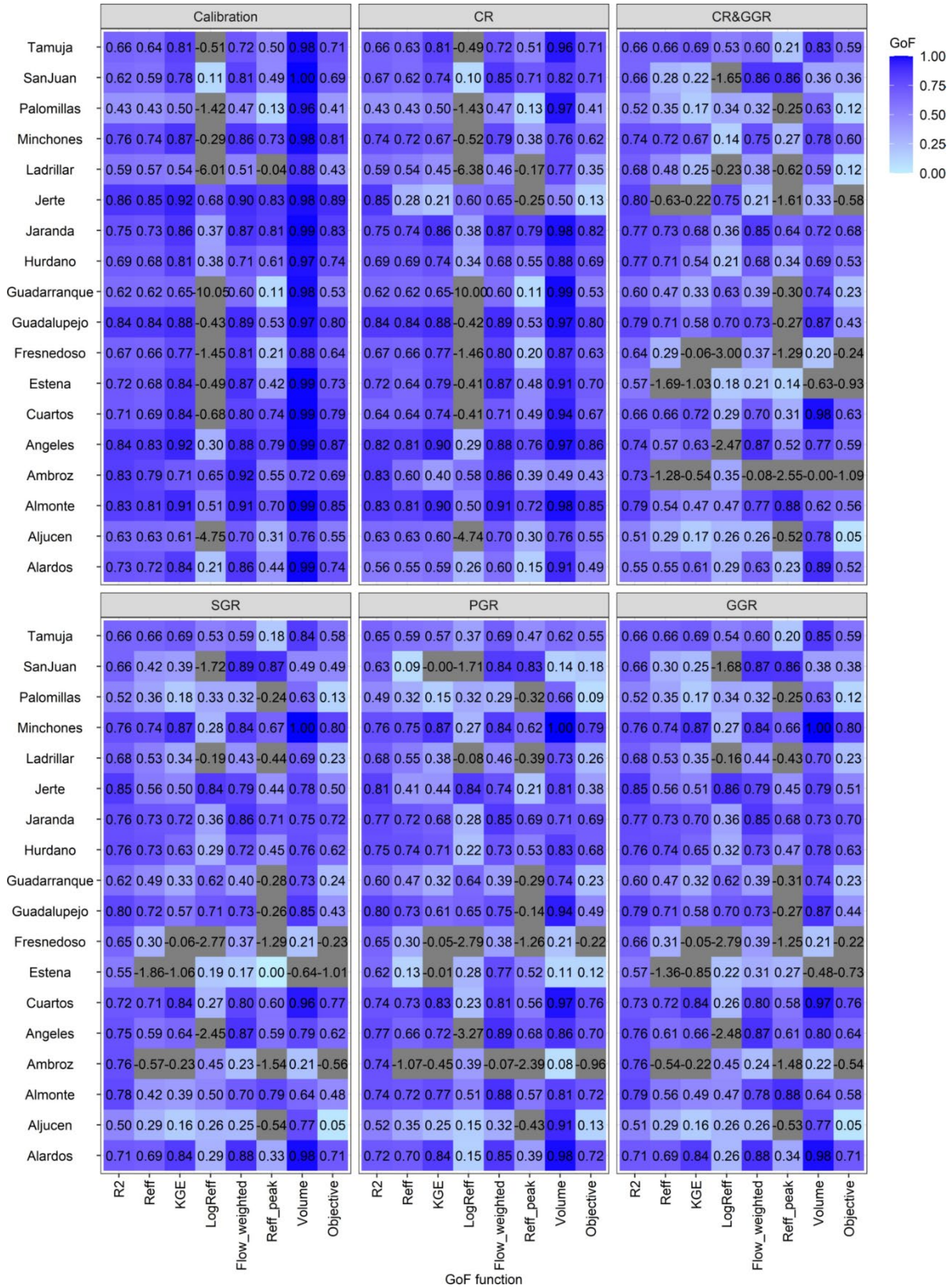


FIGURE SI.2 Goodness of fit values in the catchments for the different scenarios in the regionalization of the parameters, where the value for a perfect fit is 1 and values below 0 indicate poor fit (represented in grey colour)

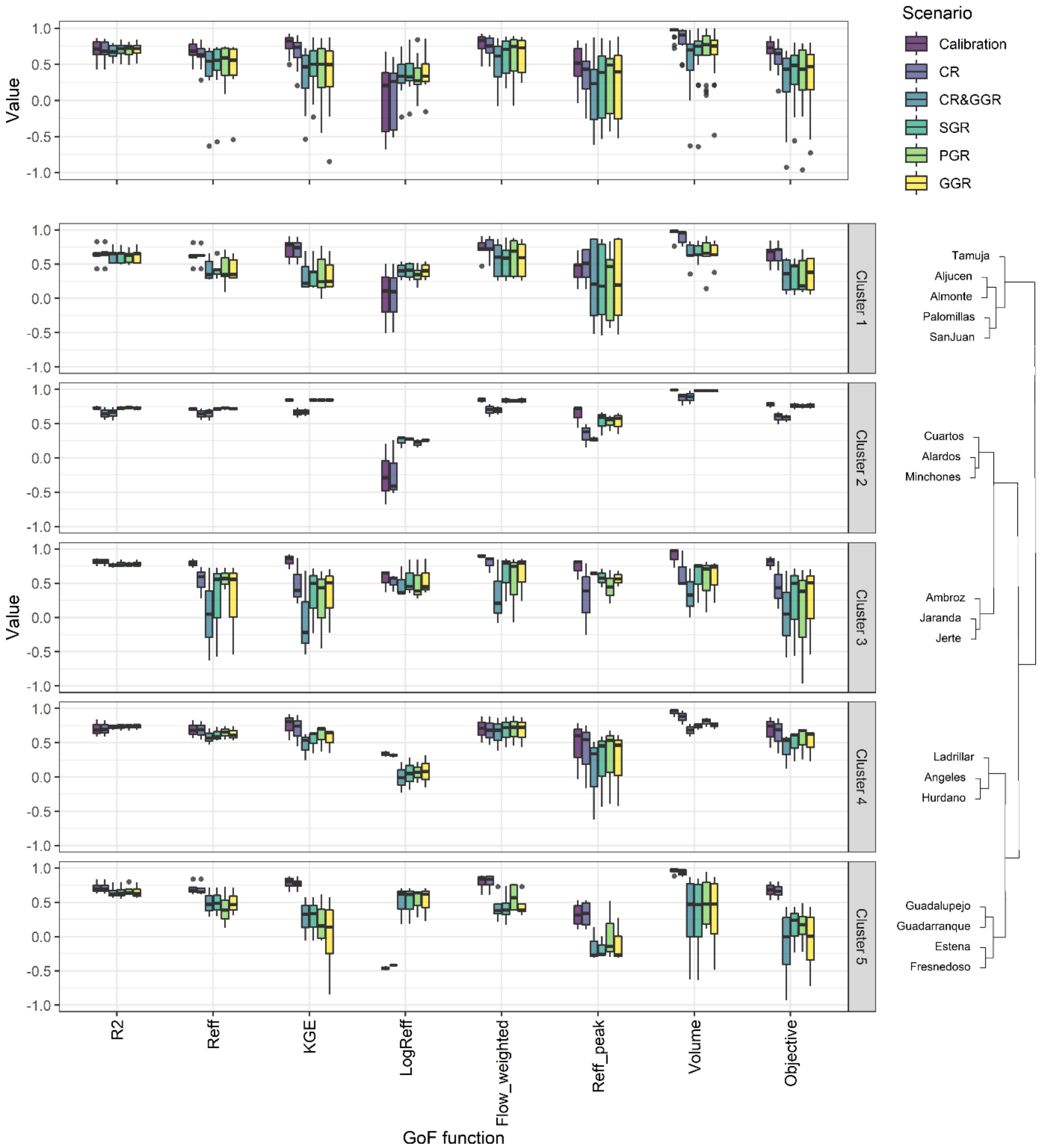
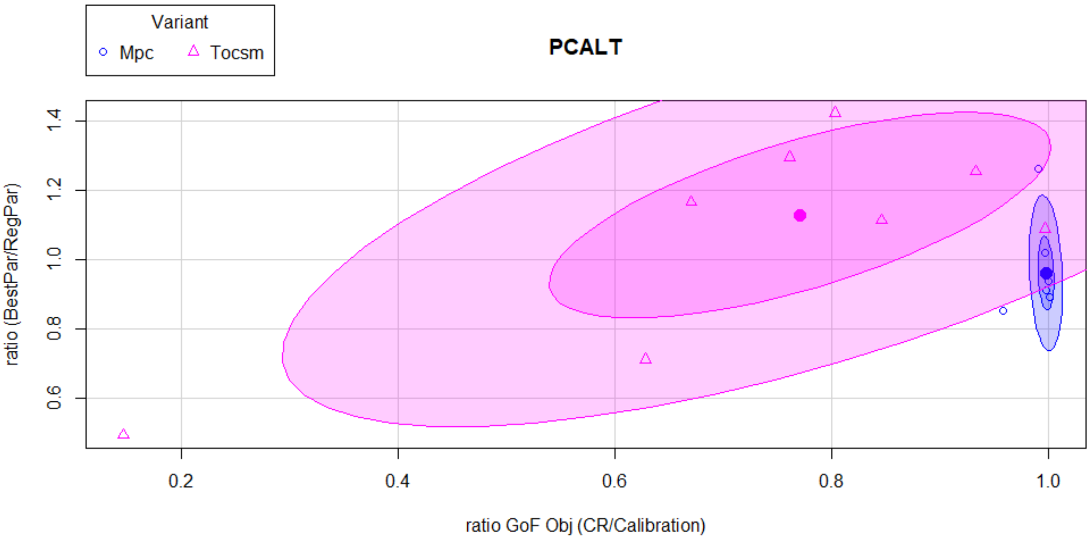
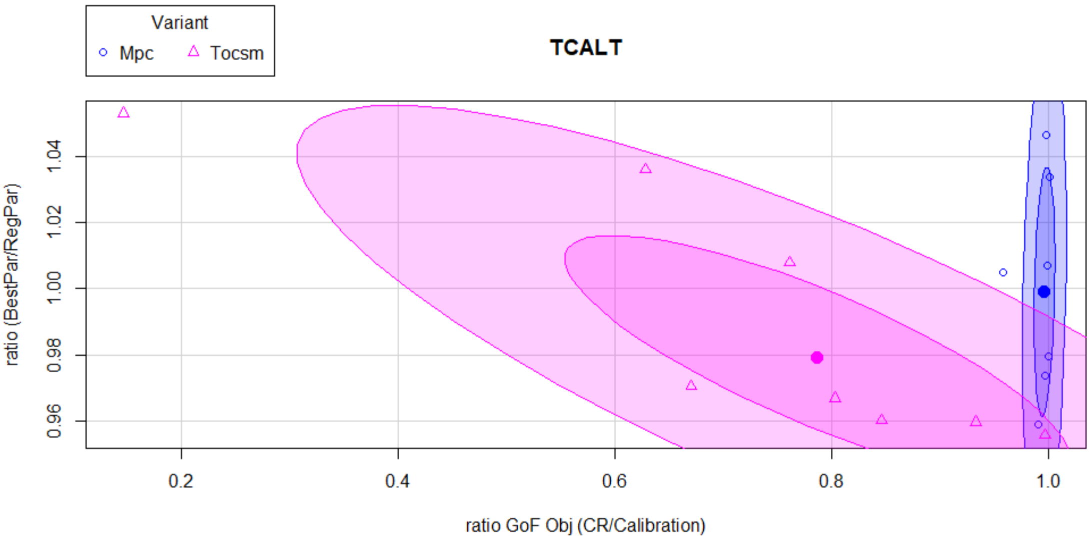


FIGURE SI.3 Boxplot graphic of the goodness of fit function values in the different scenarios considered for the regionalization of the parameters (please note that values below -1 in Figure 7 are not represented in this graphic)





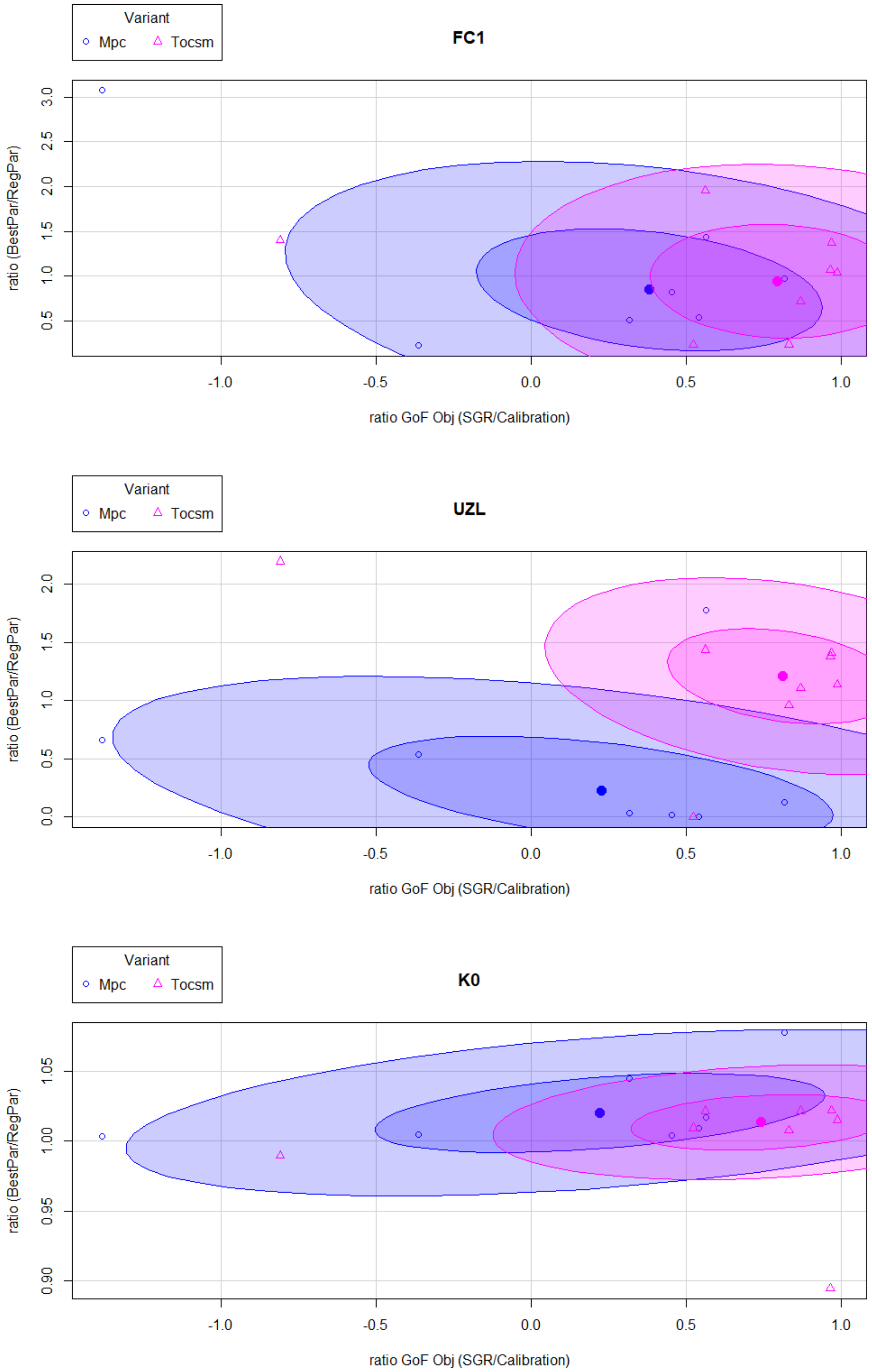
(a)

FIGURE SI.4 Scatterplot displaying the relationship between the ratio of the regionalized GoF objective function to the calibration values and the ratio of BestPar to Regpar. The value for a perfect fit of both ratios is 1 and the ellipses draw a 95% and 50% confidence levels. In all the examples are represented the most sensitive parameters of the hydrological model in both bioclimatic variants in the study area: the Mediterranean pluviseasonal-continental (Mpc) which is the driest variant, and temperate oceanic sub-Mediterranean (Tocsm) which is the most humid variant..

(a) For the Climatic Regionalization (CR) scenario are represented PCALT and TCALT

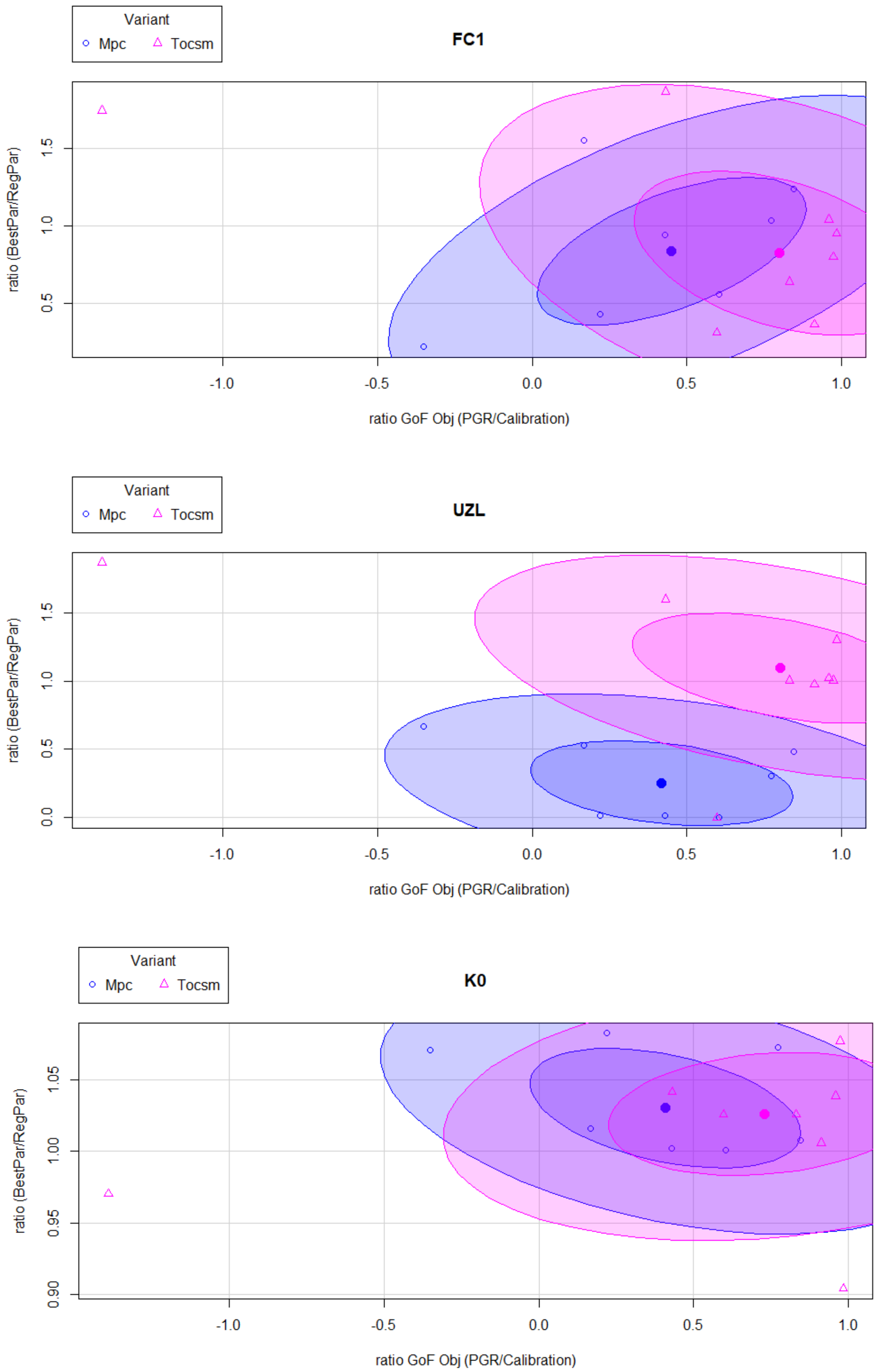
(b) For the Spectral Ground Regionalization (SGR) scenario are represented FC, UZL and K0.

(c) For the Physicall Ground Regionalization (PGR) scenario are represented FC, UZL and K0.



(b)

FIGURE SI.4 (continued)



(c)

FIGURE SI.4 (continued)

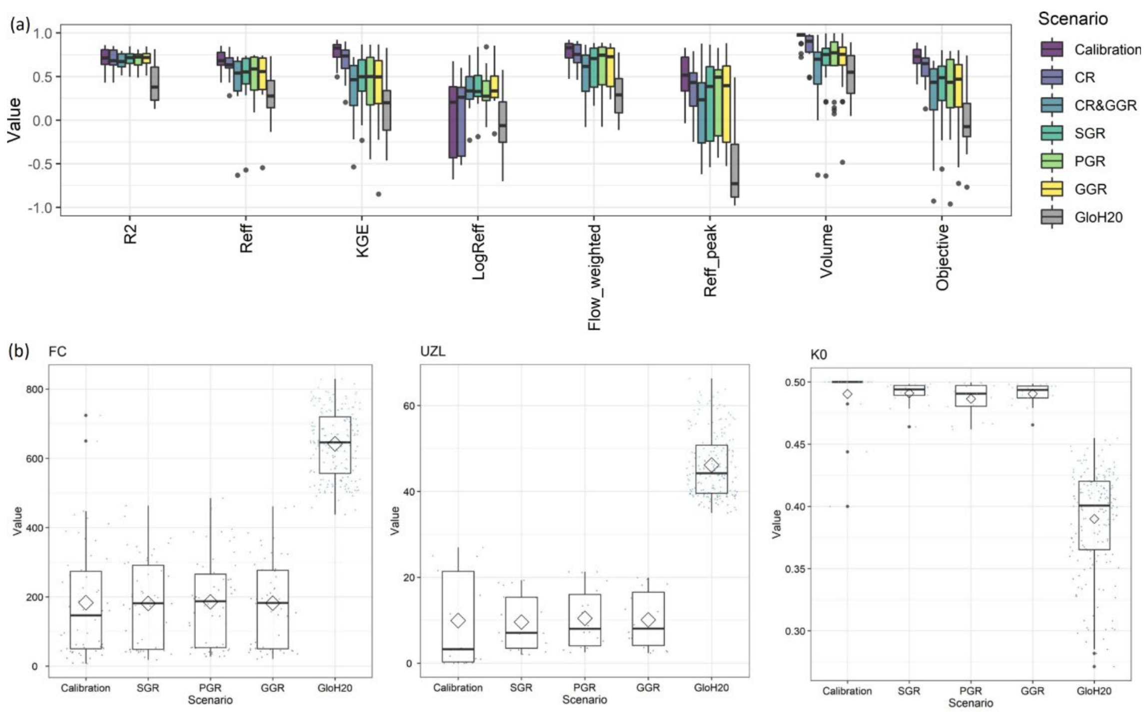


FIGURE SI.5 Performance of the GloH2O regionalized parameter dataset. (a) Boxplot graphic of the goodness of fit function values in the scenarios proposed in our work. (b) Boxplot graphic of the averaged ensemble-mean GloH2O regionalized parameters compared to the calibrated and scenario-regionalized parameters proposed in our work