Supplement of Hydrol. Earth Syst. Sci., 24, 827–847, 2020
https://doi.org/10.5194/hess-24-827-2020-supplement

Hydrology and Earth System Sciences
Open Access
EGU

Supplement of

A data-based predictive model for spatiotemporal variability in stream water quality

Danlu Guo et al.

Correspondence to: Danlu Guo (danlu.guo@unimelb.edu.au)

Supplementary Materials

Table S1. Data sources of the potential spatial predictors for water quality (i.e. catchment characteristics). See Lintern et al. (2018b) for details.

	Catchment characteristic	Data Source
Climate	Average annual radiation (MJ $m^{-2}day^{-1}$)	(Geoscience Australia, 2011)
	Average temperature (°C)	(Geoscience Australia, 2011)
	Average temperature of warmest quarter (°C)	(Geoscience Australia, 2011)
	Average temperature of coldest quarter (°C)	(Geoscience Australia, 2011)
	Maximum temperature of hottest month (°C)	(Geoscience Australia, 2011)
	Minimum temperature of coldest month (°C)	(Geoscience Australia, 2011)
	Annual average rainfall (mm)	(Geoscience Australia, 2011)
	Average rainfall of the wettest quarter (mm)	(Geoscience Australia, 2011)
	Average rainfall of the driest quarter (mm)	(Geoscience Australia, 2011)
	Average rainfall of the coldest quarter (mm)	(Geoscience Australia, 2011)
	Average rainfall of the warmest quarter (mm)	(Geoscience Australia, 2011)
	Annual average catchment rainfall erosivity (MJ $mm^{-1}ha^{-1}hr^{-1}yr^{-1}$)	(Geoscience Australia, 2011)
Hydrology	Average annual runoff (mm)	(Geoscience Australia, 2011)
	Average of average daily flow (ML d^{-1})	Calculated using instantaneous flows from DELWP (2016)
	Standard deviation of average daily flow (ML d^{-1})	Calculated using instantaneous flows from DELWP (2016)
	Pereniality of runoff (%) (proportion of "contribution to mean annual discharge by the driest six months of the year" (Geoscience Australia, 2011))	(Geoscience Australia, 2011)
	Mean number of days where there is no flow annually (days $year^{-1}$)	Calculated using daily flows from DELWP (2016)
	Mean 7-day low flow (ML d^{-1})	Calculated using instantaneous flows from DELWP (2016)
	Mean Base Flow Index	Calculated using method outlined in Grayson et al. (1996)
	Maximum distance upstream to dam wall or reservoir (km)	(Geoscience Australia, 2011)
	Area of catchment comprised of farm dams (%)	(Department of Environment Land Water and Planning Victoria, 2016)
	Total storage capacity of dams in catchment normalized to average daily flow (ML $ML^{-1}d^{-1}$)	(Geoscience Australia, 2004)
Land use	Area of catchment urbanized (%)	(Bureau of Rural Sciences, 2010)
	Area of catchment made up of roads (%)	(Bureau of Rural Sciences, 2010)
	Area of catchment used for horticulture (%)	(Bureau of Rural Sciences, 2010)
	Area of catchment used for agriculture (%)[1]	(Bureau of Rural Sciences, 2010)
	Area of catchment used for pastures (grazing) (%)	(Bureau of Rural Sciences, 2010)
	Area of catchment used for cropping (%)[2]	(Bureau of Rural Sciences, 2010)
Land cover	Mean width of vegetated riparian zone (m)	(Department of Environment Land Water and Planning, 2014)
	Average fragmentation of riparian zone (%)	(Department of Environment Land Water and Planning, 2014)
	Area of catchment covered with grass (%)[3]	(Geoscience Australia, 2011)
	Area of catchment covered with forest (%)[4]	(Geoscience Australia, 2011)
	Area of catchment covered with shrubs (%)[5]	(Geoscience Australia, 2011)
	Area of catchment covered with woodland (%)[6]	(Geoscience Australia, 2011)
	Area of catchment bare (%)	(Geoscience Australia, 2011)

Soil type and geology	Area of catchment underlain by unconsolidated bedrock (%)	(Geoscience Australia, 2011)
	Area of catchment underlain by igneous bedrock (%)	(Geoscience Australia, 2011)
	Area of catchment underlain by sedimentary bedrock (%)	(Geoscience Australia, 2011)
	Area of catchment underlain by mixed igneous and sedimentary bedrock (%)	(Geoscience Australia, 2011)
	Average soil TP content (mg kg^{-1})	(Terrestrial Ecosystem Research Network, 2016)
	Average soil TN content (mg kg^{-1})	(Terrestrial Ecosystem Research Network, 2016)
	Average soil clay content (%)	(Terrestrial Ecosystem Research Network, 2016)
	Area of catchment with saline aquifers (%)	(Department of Agriculture and Water Resources, 2013)
Topography	Catchment area ($km^{2)}$	(Geoscience Australia, 2011)
	Mean catchment elevation (m)	(Geoscience Australia, 2011)
	Maximum catchment elevation (m)	(Geoscience Australia, 2011)
	Area of catchment made up of valley bottoms (%)	(Geoscience Australia, 2011)
	Total catchment length (km)	(Geoscience Australia, 2011)
	Mean catchment slope (%)	(Geoscience Australia, 2011)
	Mean channel slope (%)	Calculated using BOM (2012)

1. Agricultural activities include all primary production activities including plantation forests, grazing pastures, cropping and horticulture. This includes both dryland and irrigation agricultural activities.

2. Cropping refers to the production of commodities such as cereals, beverage and spice crops, hay, oilseeds, sugar, cotton, alkaloid poppies and pulses.

3. Grass refers to grasslands with tussock, hummock, reeds/rushes.

4. Forest refers to rainforests, Eucalypt forests, mangroves and low closed forests (e.g., Acacia, Melaleuca or Banksia species). Areas with high density of vegetation (>30% cover) and tall trees (>10 m).

5. Shrubs refers to open and dry woodlands and shrublands with hummock or tussock grass, Melaleuca shrublands, lignum shrublands, saltbush and chenopods. Areas with vegetation <2 m tall.

6. Woodlands refer to areas with medium trees (<10 m) at medium density (<30% cover).

Table S2. Data sources of the potential temporal predictors for water quality. See Guo et al. (2019) for details.

Data		**Source**
Daily rainfall (mm)		Australian Water Availability Project (AWAP) (Raupach et al., 2009, 2012) Available from: http://www.csiro.au/awap; http://www.bom.gov.au/jsp/awap/index.jsp
Daily average temperature (°C)		
Daily actual ET (mm)		Australian Water Resources Assessment (Frost et al., 2016) Available from: http://www.bom.gov.au/water/landscape
Daily average root zone soil moisture		
Daily average deep soil moisture		
Monthly NDVI	January 1994 – December 1999	Advanced Very High Resolution Radiometer product (AVHRR) (Eidenshink, 1992) Available from: https://earthdata.nasa.gov/
	January 2000 – December 2013	Moderate Resolution Imaging Spectroradiometer (MODIS); MOD13A3 (NASA LP DAAC, 2017) Available from: https://earthdata.nasa.gov/

Table S3. Log-sinh transformation parameter (a and b) values for 50 potential spatial predictors for stream water quality (i.e. catchment characteristics).

Catchment characteristics	***a***	***b***
Annual radiation (MJ $m^{-2}day^{-1}$)	3.458	2.052
Annual temperature (°C)	2.425	3.133
Annual rainfall (mm)	0.008	0.001
Erosivity (MJ $mm^{-1}ha^{-1}hr^{-1}yr^{-1}$)	0.030	0.000
Driest quarter rain (mm)	0.099	0.003
Wettest quarter rain (mm)	0.002	0.003
Warmest quarter rainfall (mm)	0.039	0.005
Coldest quarter rainfall (mm)	0.001	0.001
Coldest month minimum temperature (°C)	4.999	0.000
Hottest month maximum temperature (°C)	0.000	0.002
Coldest quarter mean temperature (°C)	4.986	4.996
Warmest quarter mean temperature (°C)	3.805	2.193
Average of average daily flow (ML d^{-1})	0.002	0.001
Average of average daily flow (ML d^{-1})	0.034	0.002
Standard deviation of average daily flow (ML d^{-1})	0.012	0.430
Pereniality of runoff (%) (proportion of 'contribution to mean annual discharge by the driest six months of the year'	0.106	0.152
Mean number of days where there is no flow annually (days $year^{-1}$)	0.000	0.066
Mean 7-day low flow (ML d^{-1})	0.045	3.319
Mean Base Flow Index	4.896	0.000
Maximum distance upstream to dam wall or reservoir (km)	0.034	0.006
Area of catchment comprised of farm dams (%)	0.000	5.000
Total storage capacity of dams in catchment normalized to average daily flow (ML $ML^{-1}d^{-1}$)	0.003	0.002
Area of catchment urbanized (%)	0.000	0.135
Area of catchment made up of roads (%)	0.055	0.729
Area of catchment used for agriculture (%)	4.998	4.995
Area of catchment used for pastures (grazing) (%)	0.174	0.114
Area of catchment used for cropping (%)	0.000	0.079
Area of catchment used for horticulture (%)	0.000	0.373
Mean width of vegetated riparian zone (m)	0.293	0.013
Average fragmentation of riparian zone (%)	0.174	0.132
Area of catchment covered with grass (%)	0.000	0.158
Area of catchment covered with forest (%)	0.238	0.020
Area of catchment covered with shrubs (%)	0.000	0.403
Area of catchment covered with woodland (%)	0.002	0.108
Area of catchment bare (%)	0.000	5.000
Area of catchment underlain by unconsolidated bedrock (%)	0.024	0.050
Area of catchment underlain by igneous bedrock (%)	0.034	0.068
Area of catchment underlain by sedimentary bedrock (%)	4.998	4.995
Area of catchment underlain by mixed igneous and sedimentary bedrock (%)	0.000	0.032
Average soil TP content (mg kg^{-1})	0.044	4.744
Average soil TN content (mg kg^{-1})	0.213	1.733

Average soil clay content (%)	0.000	0.021
Area of catchment with saline aquifers (%)	0.001	0.000
Catchment area (km^2)	0.177	0.001
Mean catchment elevation (m)	0.044	0.001
Area of catchment made up of valley bottoms (%)	0.002	0.074
Total catchment length (km)	0.003	0.001
Mean catchment slope (%)	0.078	0.068
Mean channel slope (%)	0.029	4.899
Average soil clay content (%)	0.103	0.040

Table S4. Box-Cox transformation parameter (λ) values for the six water quality constituents and the nineteen potential temporal predictors. Values in bracket show the standard deviation of individual site-level λ.

Water Quality Constituent	**λ**
TSS	-0.249 (0.287)
TP	-0.058 (0.181)
FRP	-0.836 (1.056)
TKN	0.141 (0.342)
NO_x	0.107 (0.305)
EC	-0.024 (0.921)
Temporal predictors	**λ**
Rainfall (mm)	0.106 (0.041)
Rainfall on previous day (mm)	0.108 (0.028)
Averaged rainfall over previous 3 days (mm)	0.157 (0.022)
Averaged rainfall over previous 7 days (mm)	0.220 (0.025)
Averaged rainfall over previous 14 days (mm)	0.192 (0.046)
Averaged rainfall over previous 30 days (mm)	0.116 (0.075)
Streamflow (mm d^{-1})	-0.015 (0.225)
Streamflow on previous day (mm d^{-1})	-0.027 (0.207)
Averaged Streamflow over previous 3 days (mm d^{-1})	-0.032 (0.207)
Averaged Streamflow over previous 7 days (mm d^{-1})	-0.030 (0.2)
Averaged Streamflow over previous 14 days (mm d^{-1})	-0.021 (0.198)
Averaged Streamflow over previous 30 days (mm d^{-1})	-0.004 (0.195)
Dry spell length in the past 14 days (days)	0.257 (0.089)
NDVI for the month	3.715 (1.998)
Water temperature (°C)	0.357 (0.269)
Air temperature (°C)	0.231 (0.244)
Evaporation (mm)	0.019 (0.13)
Root zone soil moisture (%)	0.913 (0.648)
Deep soil moisture (%)	0.357 (0.269)

Table S5. The key temporal predictor for each water quality constituent, and the two key factors that are mostly closely related to the spatial variation of each temporal predictor (see Section 2.3 in the main text for detailed selection process). The corresponding Spearman's correlation coefficients (R) are also shown in the last column.

Constituent	**Key factors that affect temporal variability**	**Key factors that affect spatial variability in temporal effects**	**Spearman's R**
TSS	Same-day streamflow	Annual rainfall	0.722
		Hottest month maximum temperature	-0.575
	7-day antecedent streamflow	Annual runoff	-0.536

		Mean elevation	-0.465
	Water temperature	Daily flow standard deviation	0.204
		Total catchment length	0.177
	Soil moisture root	Percentage area with saline aquifers	0.507
		Hottest month maximum temperature	0.495
	Soil moisture deep	Maximum distance upstream to dam wall or reservoir	-0.275
		Percentage area covered by grassland	-0.24
TP	Same-day streamflow	Annual rainfall	0.695
		Hottest month maximum temperature	-0.556
	30-day antecedent streamflow	Erosivity	-0.675
		Percentage cropping area	0.626
	Water temperature	Percentage agricultural area	0.382
		Percentage area used for roads	0.274
	Soil moisture root	Percentage pasture area	0.564
		Hottest month maximum temperature	0.557
	Soil moisture deep	Percentage area underlain by mixed igneous bedrock	-0.23
		Maximum distance upstream to dam wall or reservoir	-0.21
FRP	Same-day streamflow	Percentage agriculture area	0.392
		Percentage area underlain by mixed igneous bedrock	0.314
	Water temperature	Total catchment length	-0.28
		Coldest quarter mean temperature	0.232
	Soil moisture deep	Percentage area used for roads	-0.21
		Percentage aea covered by woodland	0.204
TKN	Same-day streamflow	Annual rainfall	0.713
		Hottest month maximum temperature	-0.618
	30-day antecedent streamflow	Erosivity	-0.823
		Percentage cropping area	0.694
	NDVI	Mean_7daylowflow	0.42
		Maximum distance upstream to dam wall or reservoir	-0.366
	Water temperature	Coldest quarter rainfall	-0.386
		Maximum distance upstream to dam wall or reservoir	0.374
	Soil moisture root	Warmest quarter mean temperature	0.6
		Percentage pasture area	0.588
	Soil moisture deep	Hottest month maximum temperature	-0.274
		Warmest quarter mean temperature	-0.269
NOx	Same-day streamflow	Total storage capacity of dams in catchment	-0.493
		Mean soil TN content	0.458
	30-day antecedent streamflow	Coldest quarter rainfall	-0.413
		Hottest month maximum temperature	0.396
	NDVI	Percentage area covered by woodland	-0.442

		Maximum elevation	-0.428
	Water temperature	Percentage area underlain by mixed igneous bedrock	0.266
		Percentage urbanized area	-0.2
	Soil moisture root	Annual temperature	0.44
		Warmest quarter average temperature	0.338
	Soil moisture deep	Percentage horticulture area	0.341
		Wettest quarter rainfall	-0.334
EC	Same-day streamflow	Percentage area covered by grassland	-0.347
		Percentage area covered by woodland	-0.317
	14-day antecedent streamflow	Percentage area covered by forest	0.324
		PerForest_Ext	0.276
	Water temperature	Coldest month minimum temperature	-0.328
		Mean catchment slope	0.28
	Soil moisture root	Mean 7-day low flow	0.33
		Average soil TN content	0.303
	Soil moisture deep	Maximum elevation	0.366
		Percentage area covered by woodland	0.312

30

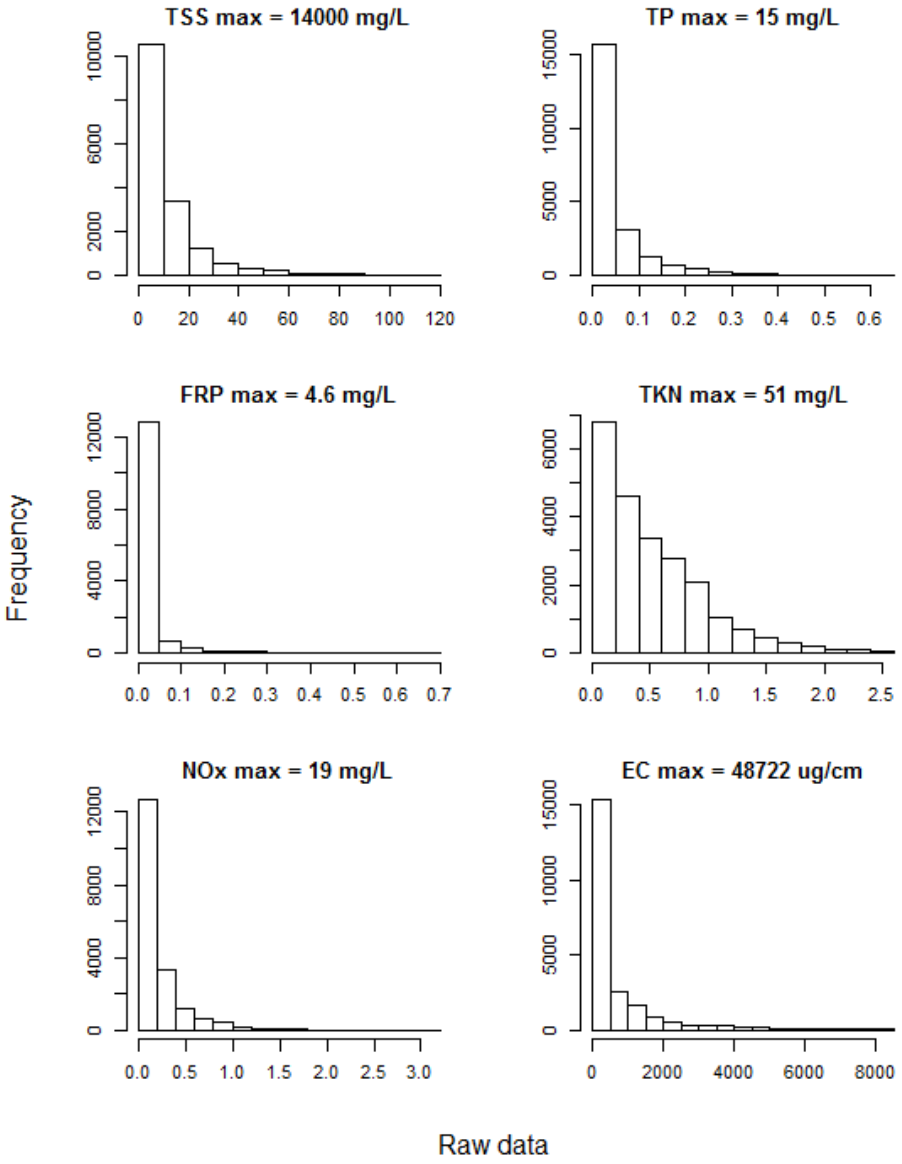


Figure S1. Distribution of the raw water quality data across all catchments. Each panel shows one constituent with only the above-DL data. To help visualizing the highly skewed data, the top percentile of data for each constituent were not plotted, while the maximum value was shown in the corresponding panel title.

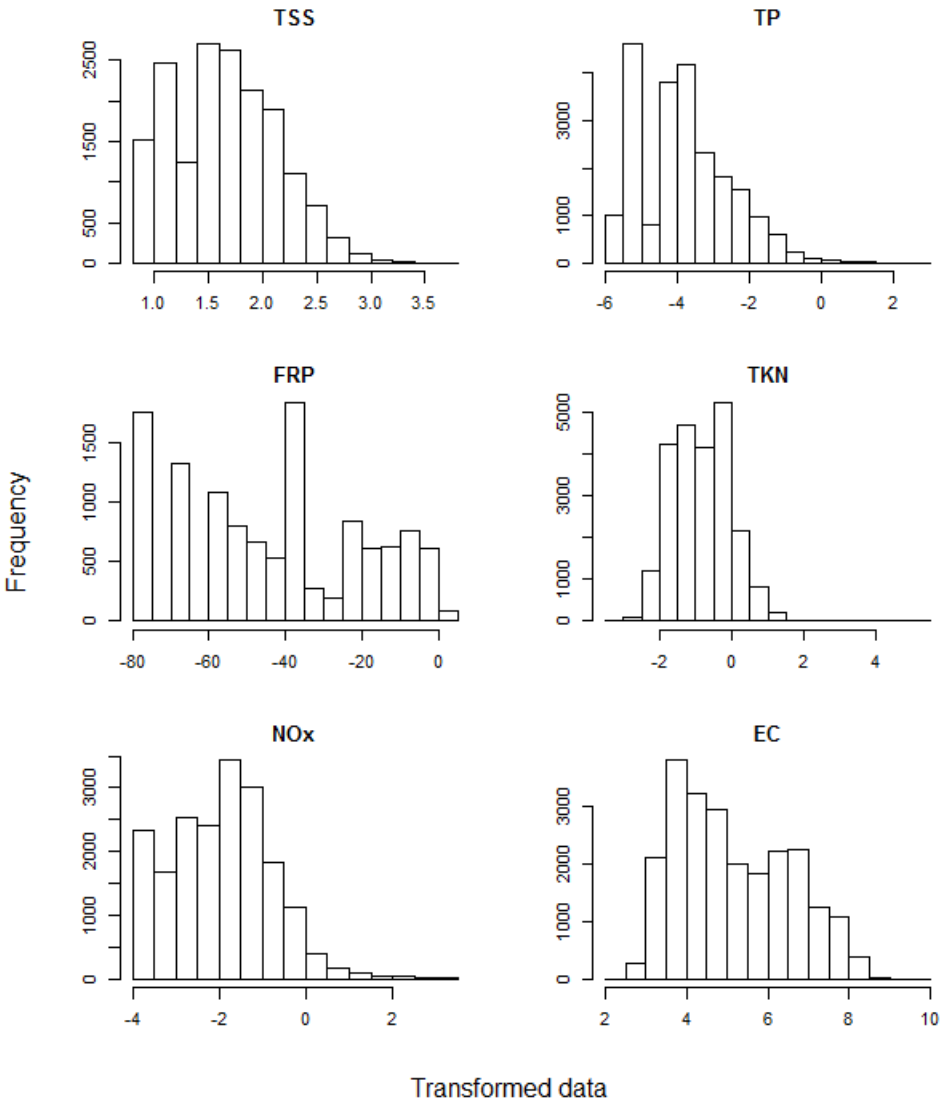
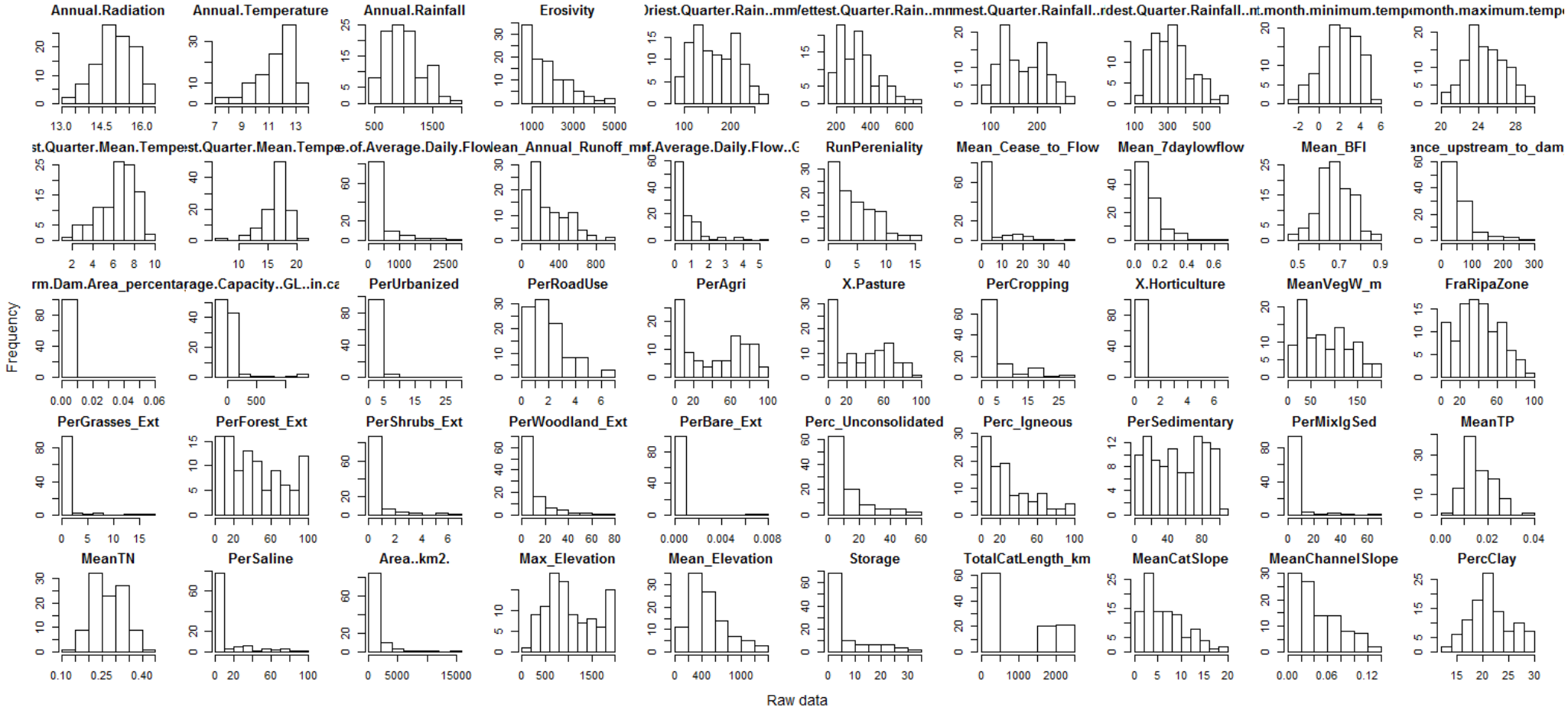


Figure S2. Distribution of the transformed water quality data across all catchments. Each panel shows one constituent with only the above-DL data.



38

39 **Figure S3. Distribution of the raw data for catchment characteristics included as potential spatial predictors in the model.**

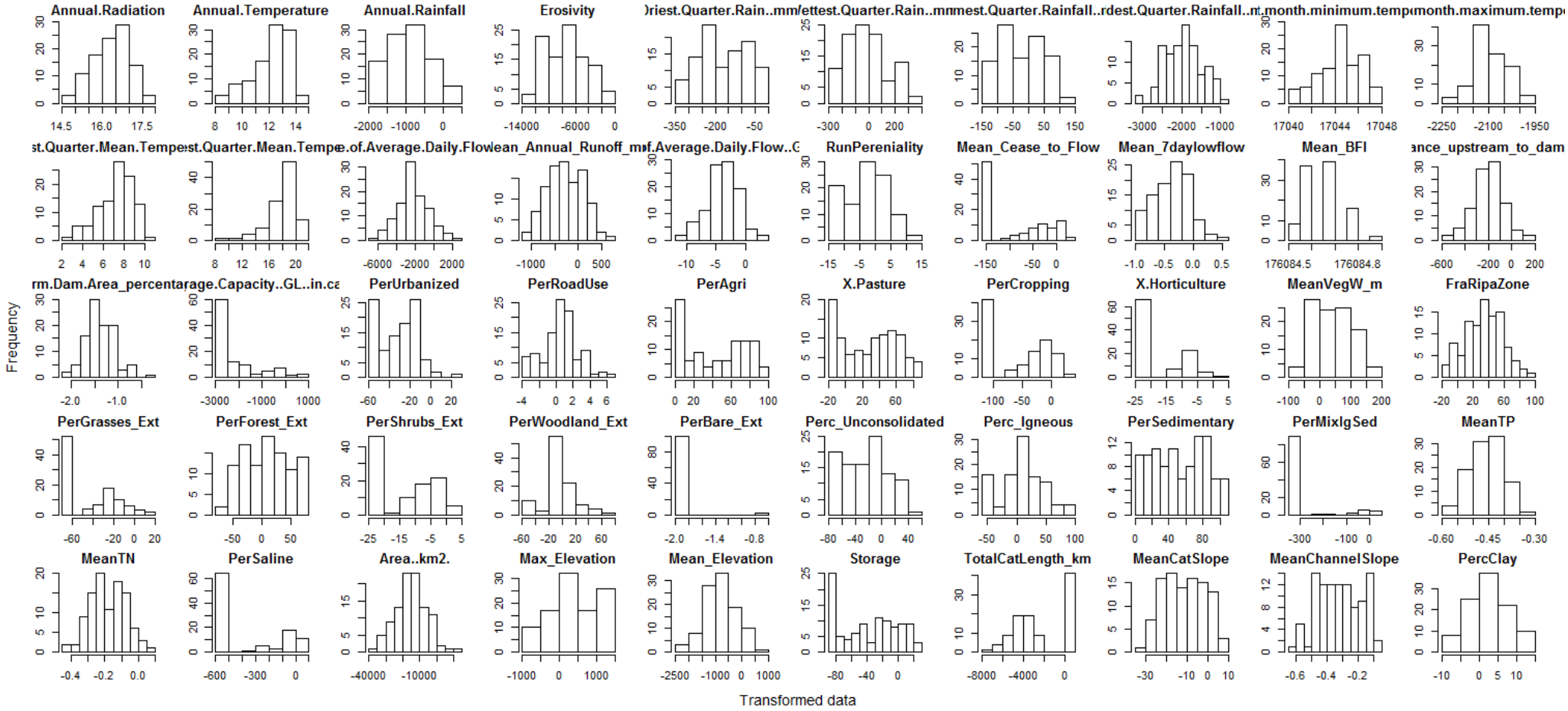


Figure S4. Distribution of the transformed data for catchment characteristics included as potential spatial predictors in the model.

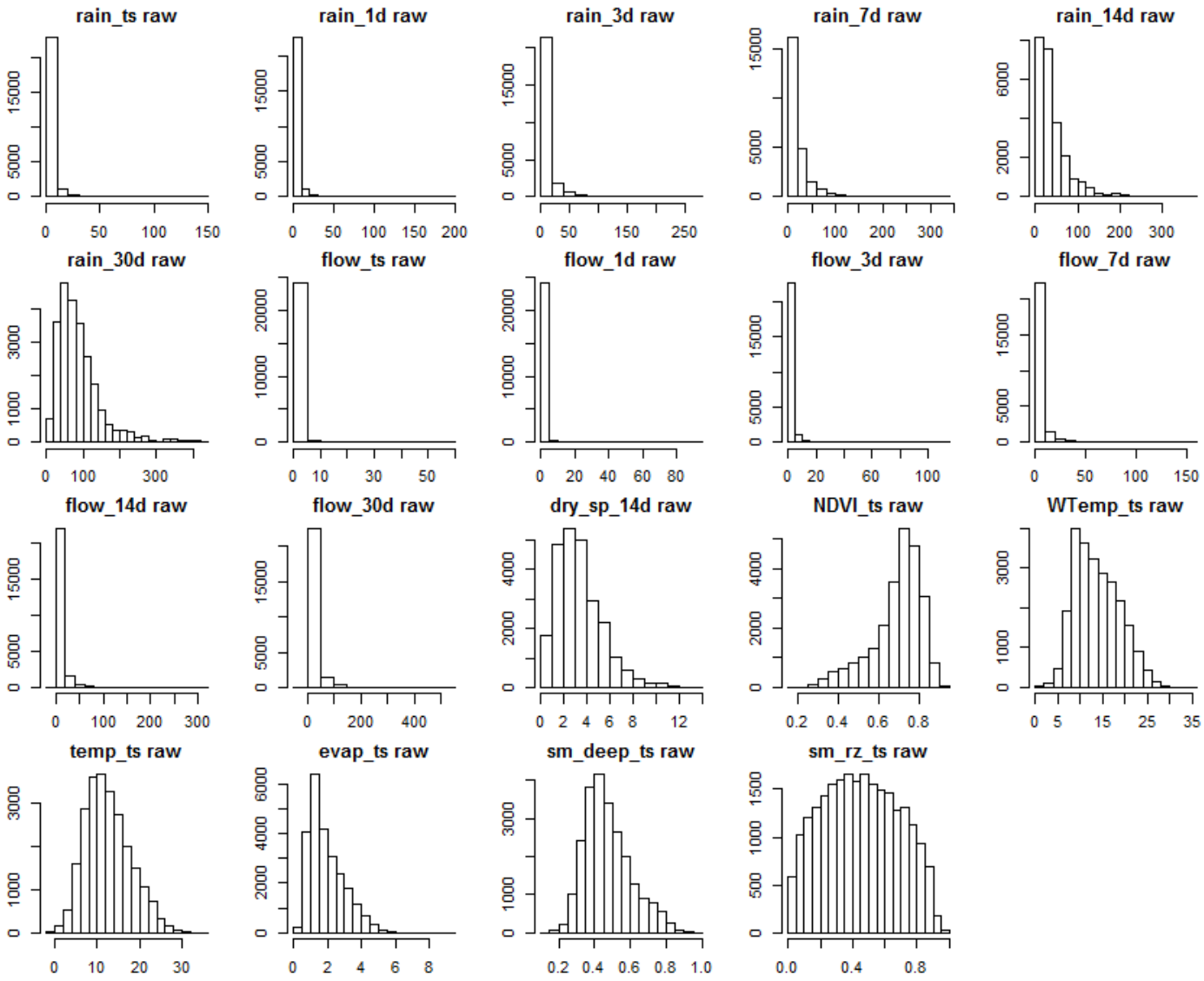
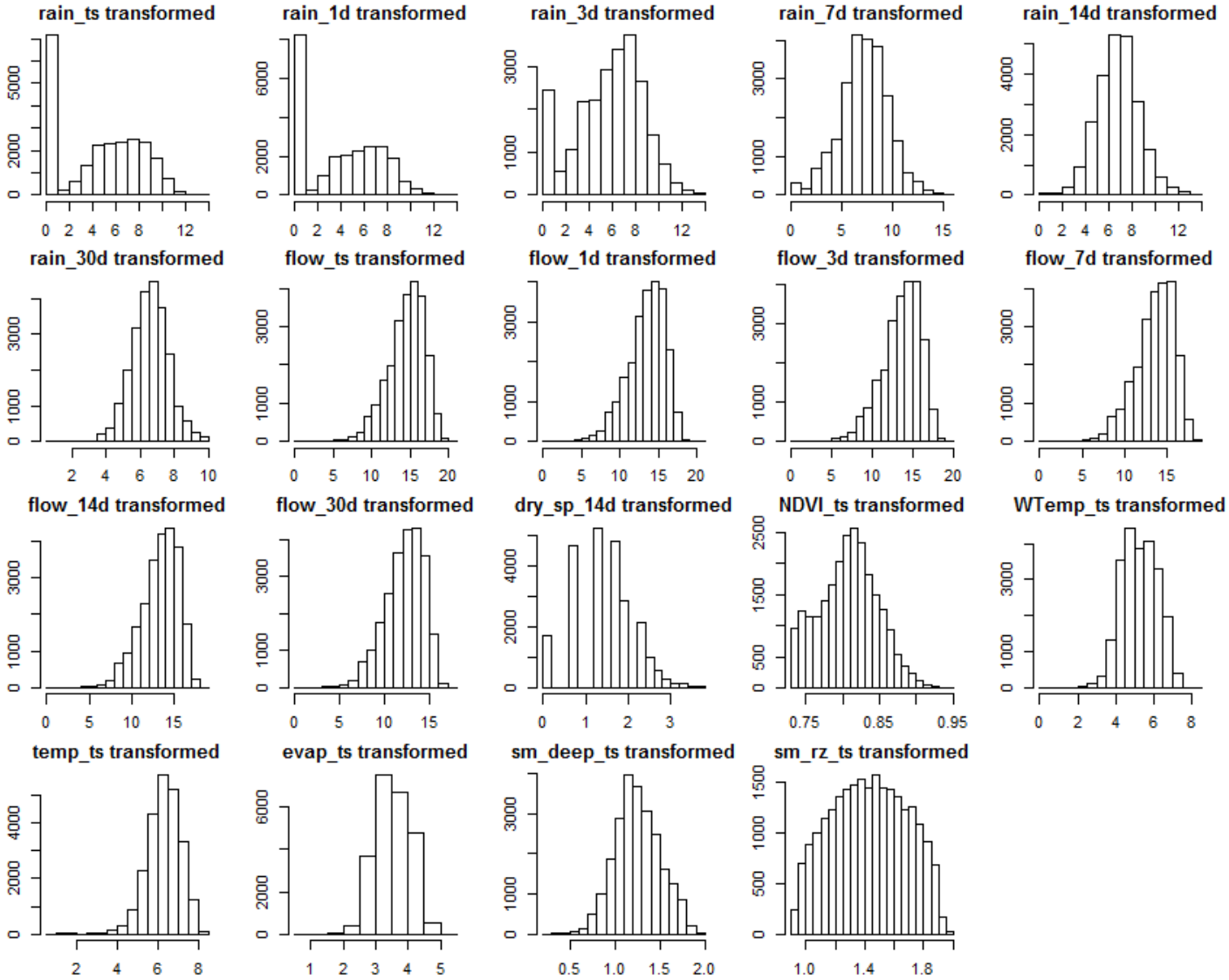


Figure S5. Distribution of the raw data for hydro-climatic and vegetation variables included as potential temporal predictors in the model.



47

48 **Figure S6. Distribution of the transformed data for transformed (Box-Cox) hydro-climatic and vegetation variables included as potential temporal predictors in the model.**

Figure S4. The two key factors that are mostly closely related to the spatial variation of each temporal predictor of each water quality constituents, as highlighted in the coloured cells (see Section 2.3 in the main text for detailed selection of the two key factors). Colours indicate the corresponding Spearman's correlation coefficients (R) from -1 (red) to 1 (blue).

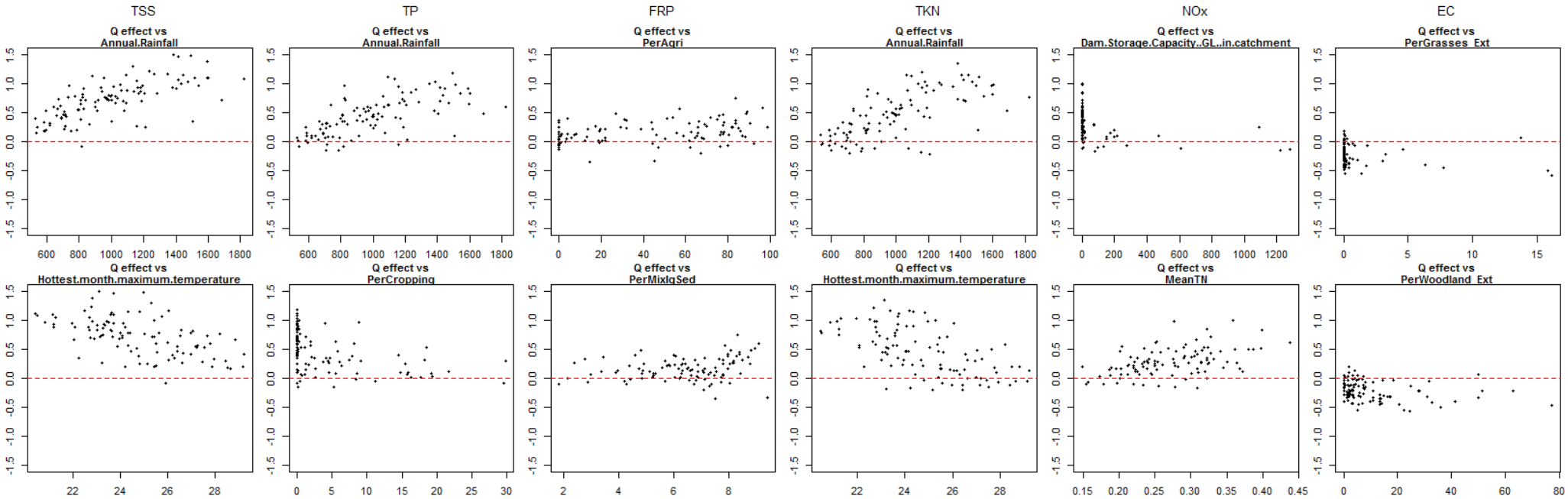


Figure S5. Effects of streamflow across catchments against the two most important catchment landscape characteristics, for each constituent (see Section 2.3 in the main text for detailed selection of the two key factors). Red dash lines indicate the zero levels, and thus differentiate positive and negative streamflow effects

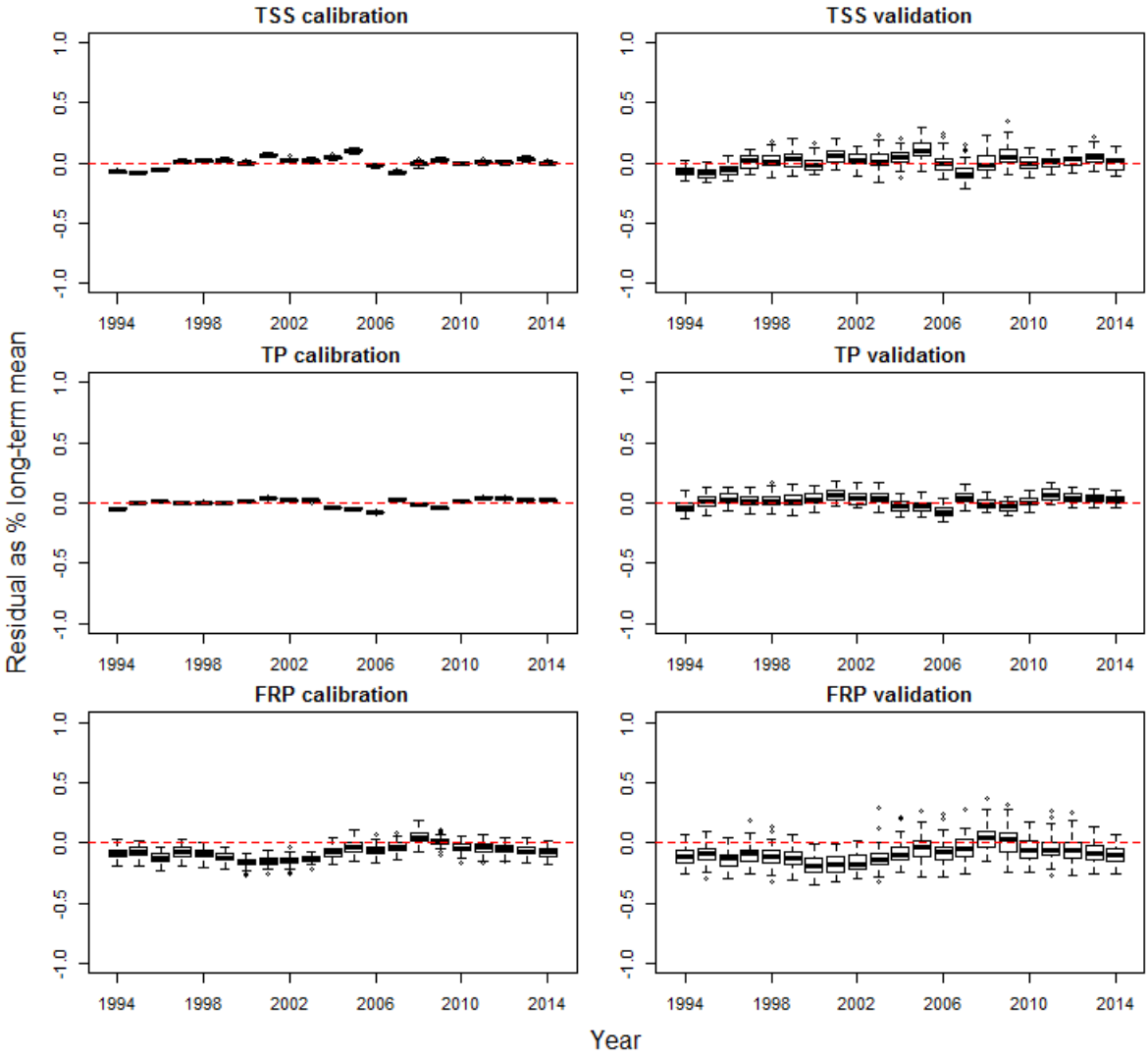


Figure S6. Annual average residuals of the models for TSS, TP and FRP, as % of long-term average. All values are presented in a Box-Cox transformed scale.

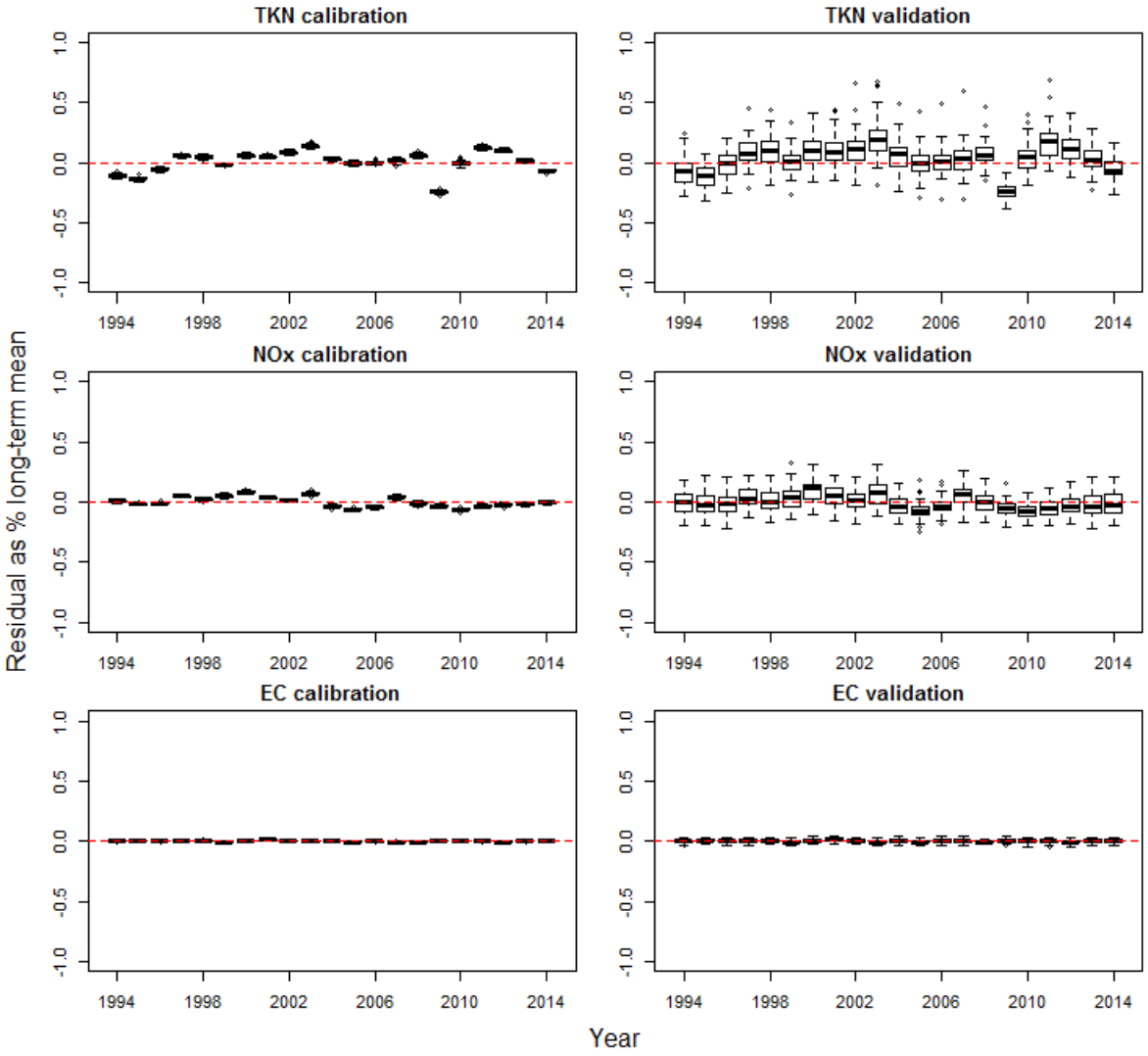


Figure S7. Annual average residuals of the models for TKN, NOx and EC, as % of long-term average. All values are presented in a Box-Cox transformed scale.

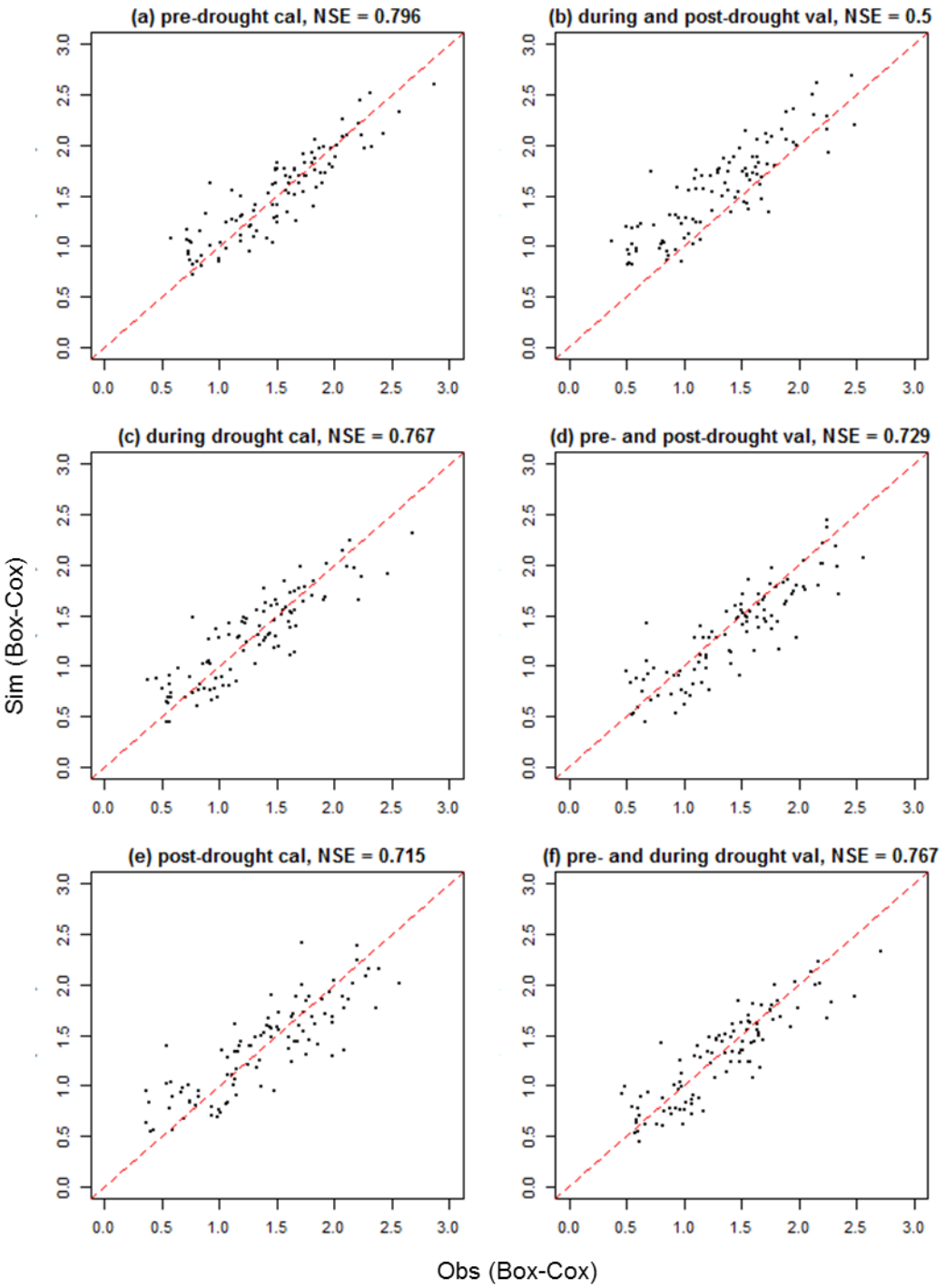


Figure S8. Comparison of the TSS model performance, as the simulated against observed site-level mean concentrations across three different calibration/validation periods for calibrations on the pre-drought (1994-1996), drought (1997-2009) and the post-drought (2010-2014) periods, respectively, see Section 2.4 for details of the calibration and validation approach.

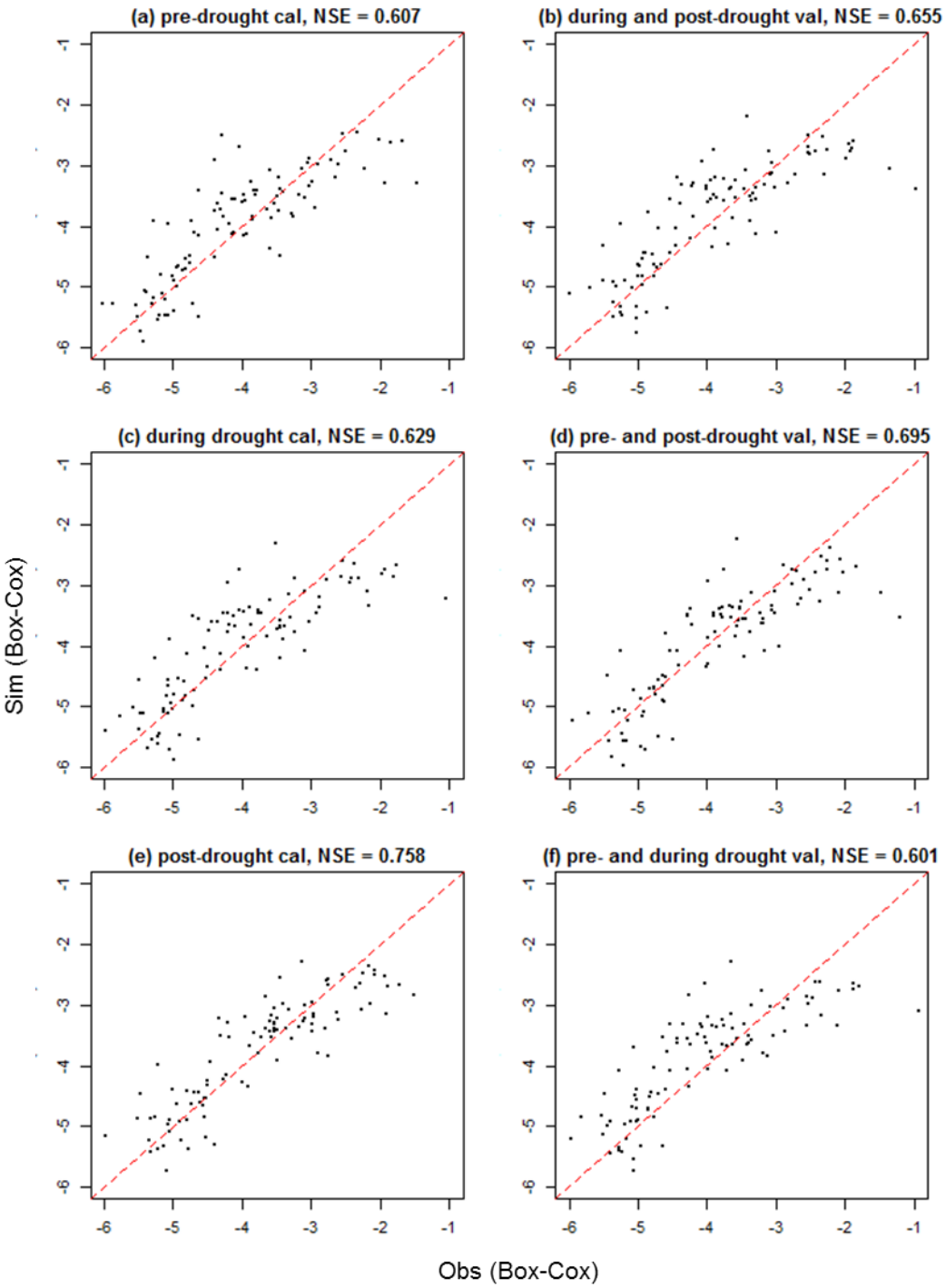


Figure S9. Comparison of the TP model performance, as the simulated against observed site-level mean concentrations across three different calibration/validation periods for calibrations on the pre-drought (1994-1996), drought (1997-2009) and the post-drought (2010-2014) periods, respectively, see Section 2.4 for details of the calibration and validation approach.

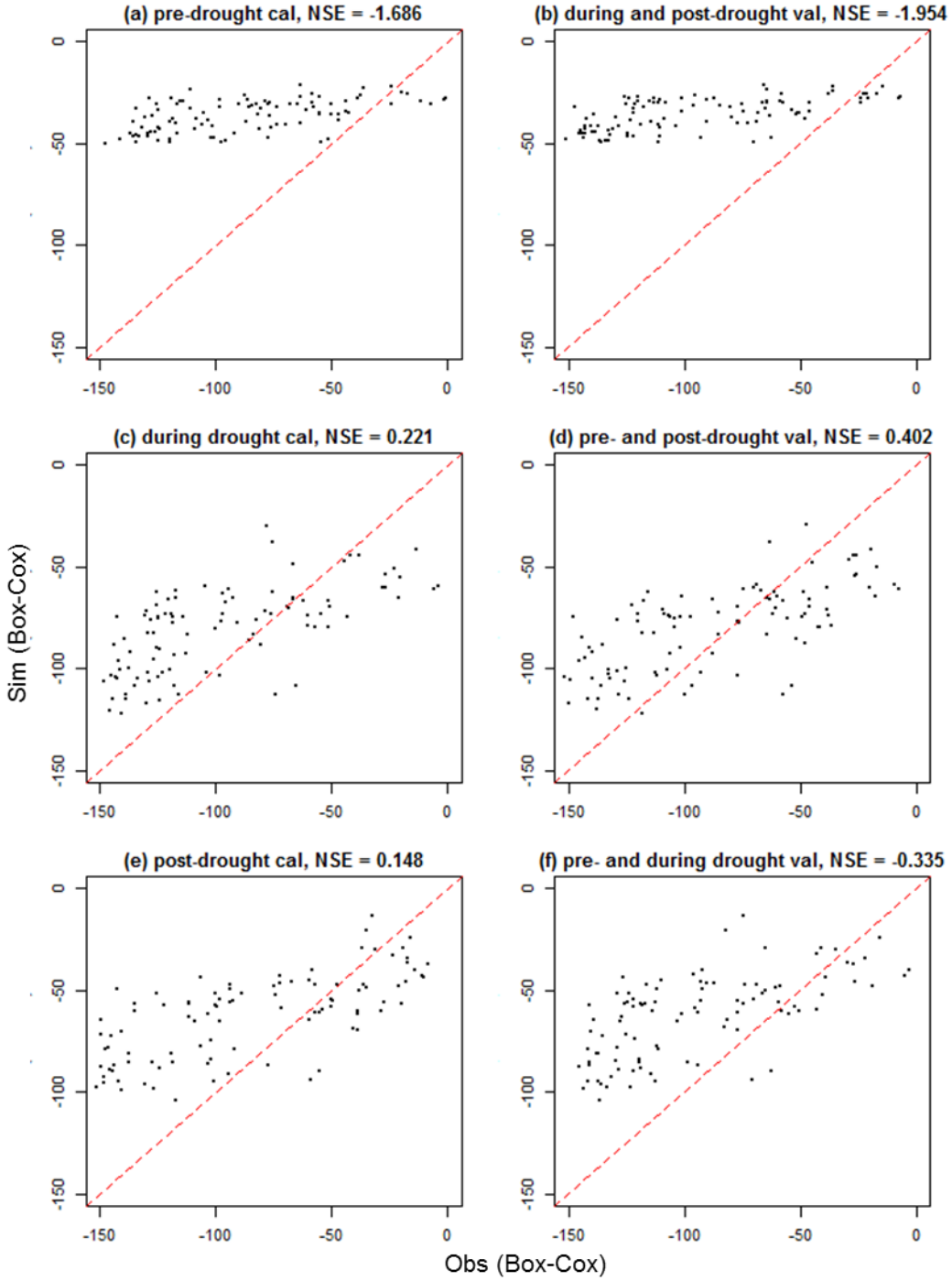


Figure S10. Comparison of the FRP model performance, as the simulated against observed site-level mean concentrations across three different calibration/validation periods for calibrations on the pre-drought (1994-1996), drought (1997-2009) and the post-drought (2010-2014) periods, respectively, see Section 2.4 for details of the calibration and validation approach. Note that the unstable performance can be resulted by the poor performance for the full model, see Section 3.1.

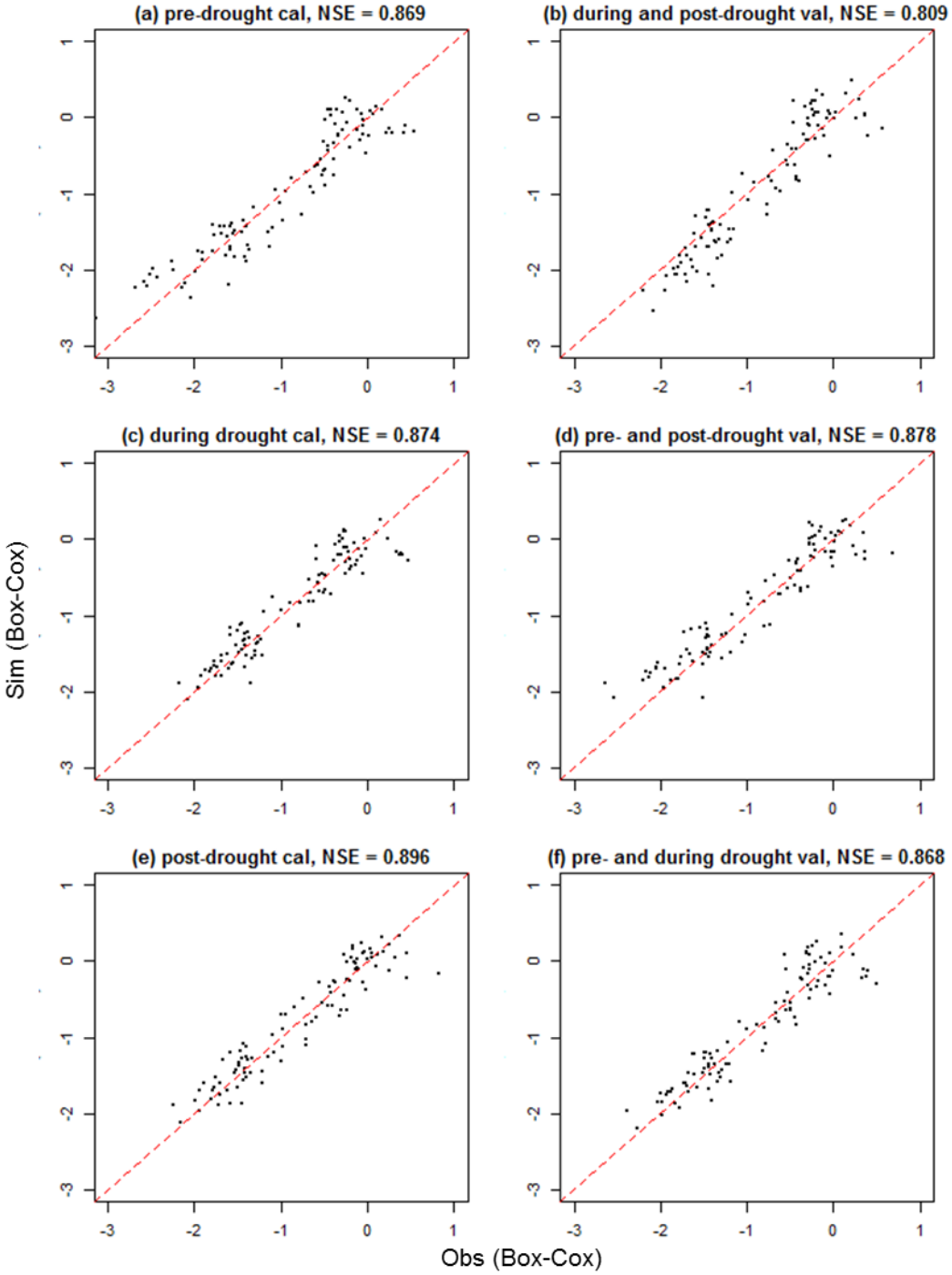


Figure S11. Comparison of the TKN model performance, as the simulated against observed site-level mean concentrations across three different calibration/validation periods for calibrations on the pre-drought (1994-1996), drought (1997-2009) and the post-drought (2010-2014) periods, respectively, see Section 2.4 for details of the calibration and validation approach.

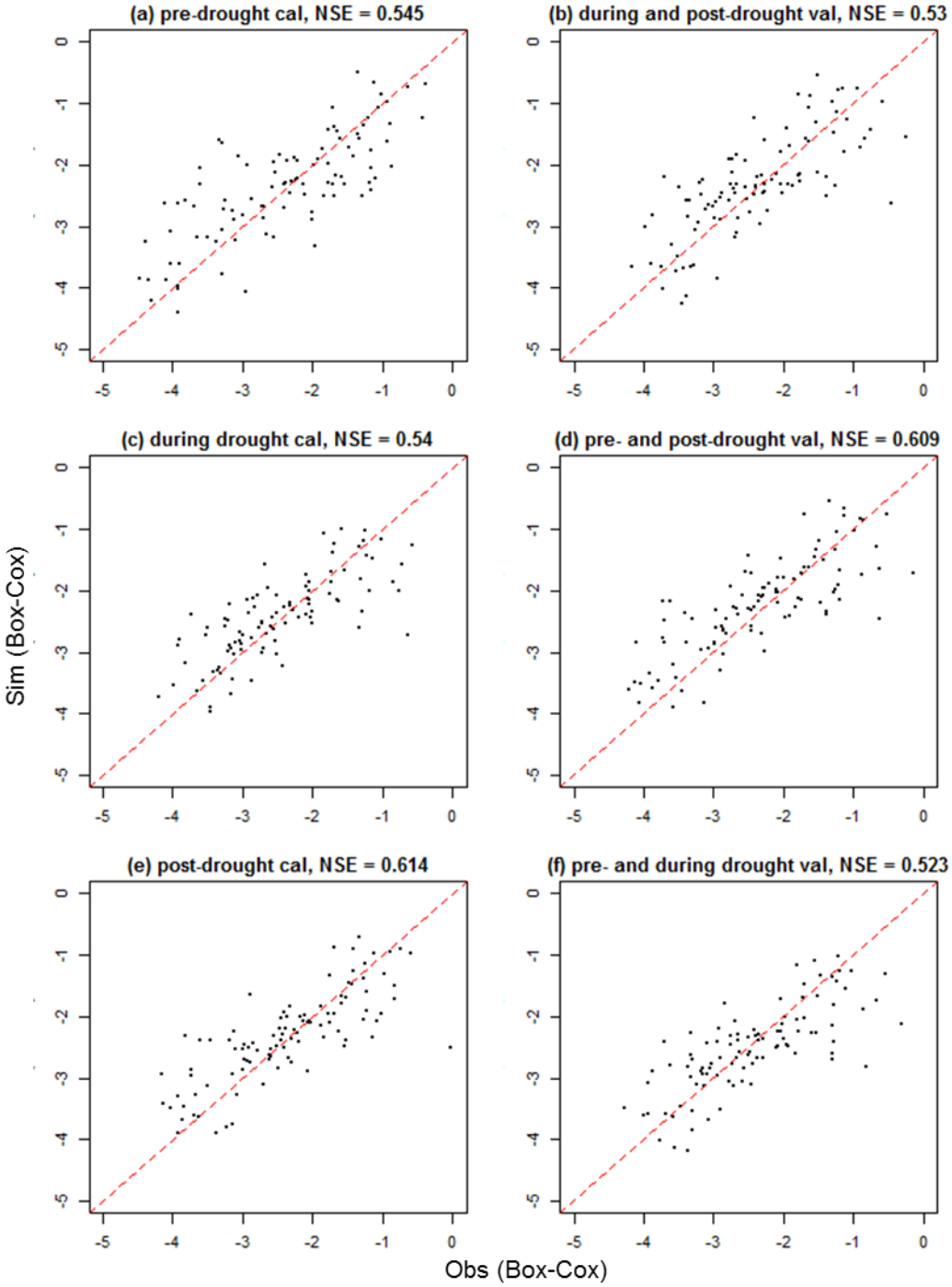


Figure S12. Comparison of the NO_x model performance, as the simulated against observed site-level mean concentrations across three different calibration/validation periods for calibrations on the pre-drought (1994-1996), drought (1997-2009) and the post-drought (2010-2014) periods, respectively, see Section 2.4 for details of the calibration and validation approach.

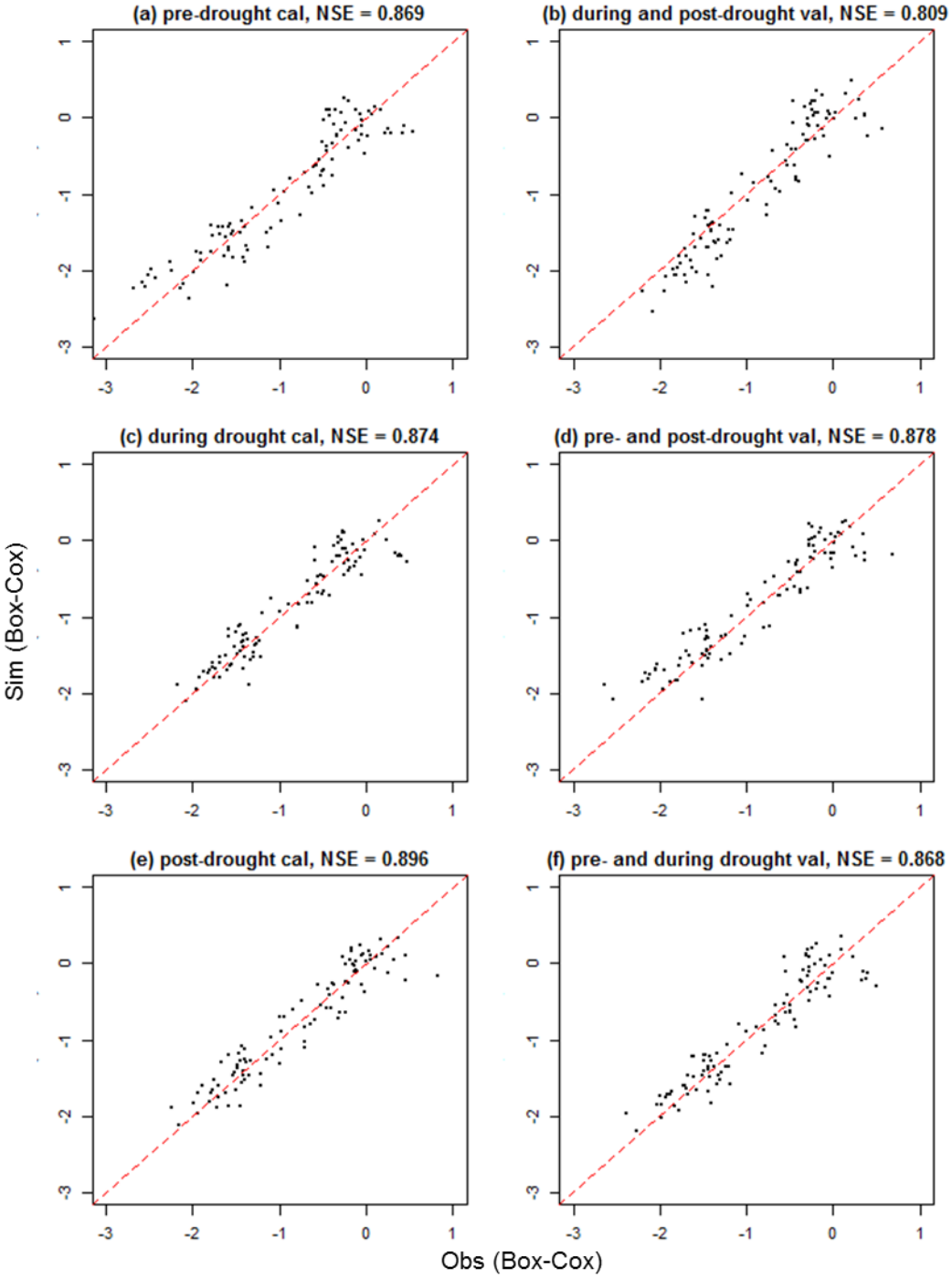


Figure S13. Comparison of the EC model performance, as the simulated against observed site-level mean concentrations across three different calibration/validation periods for calibrations on the pre-drought (1994-1996), drought (1997-2009) and the post-drought (2010-2014) periods, respectively, see Section 2.4 for details of the calibration and validation approach.

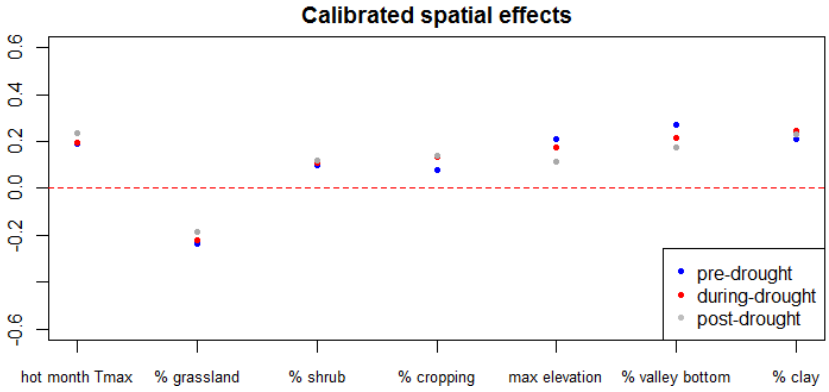


Figure S14. Effects of the seven key predictors for the spatial variability in TSS across 102 sites, summarized by the posterior mean of the calibrated parameter values for each predictor, to the pre-, during- and post-drought periods (differentiated by colour). The seven key predictors are, from left: hottest month maximum temperature, percentage catchment area as grassland, percentage catchment area as shrub, percentage catchment area as cropping land, maximum catchment elevation, percentage catchment area made up of valley bottoms, and average soil clay content.