Robustness of Conceptual Rainfall-Runoff Models: How this Varies across Australian Catchments?

Danlu Guo[1,2], Feifei Zheng[2], Hoshin Gupta[3], Holger Maier[2,4]

[1] Department of Infrastructure Engineering, The University of Melbourne, Parkville, VIC Australia.
Danlu.guo@unimelb.edu.au

[2] College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, Zhejiang. China.
feifeizheng@zju.edu.cn

[3] Department of Hydrology and Atmospheric Sciences, The University of Arizona, Tucson, AZ, USA.
hoshin@email.arizona.edu

[4] School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, SA Australia.
holger.maier@adelaide.edu.au

Full story: https://doi.org/10.1029/2019WR026752

What motivated this study?

Robustness of ANN rainfall-runoff models (Zheng et al., 2017)

AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2017WR021470

Key Points:
- Impact of calibration and evaluation data allocation on model performance is examined using a large number of catchments
- Robustness of model performance can be very poor if statistical properties of the data are ignored during data allocation
- While obtained using ANN-type models, the results are broadly relevant to all classes of hydrological models

On Lack of Robustness in Hydrological Model Development Due to Absence of Guidelines for Selecting Calibration and Evaluation Data: Demonstration for Data-Driven Models

Feifei Zheng[1], Holger R. Maier[1,2], Wenyan Wu[2,3], Graeme C. Dandy[2], Hoshin V. Gupta[4], and Tuqiao Zhang[1]

[1]College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, Zhejiang, China, [2]School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, SA, Australia, [3]Department of Infrastructure Engineering, Melbourne School of Engineering, University of Melbourne, Australia, [4]Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ, USA

Robustness of conceptual rainfall-runoff models (Guo et al., 2020)

Water Resources Research

RESEARCH ARTICLE
10.1029/2019WR026752

Key Points:
- We investigate the robustness of CRR models across calibration/evaluation data splits with a large-sample approach
- Different data splits markedly impact CRR model performance, particularly for accurately reproducing the mean and variability of runoff
- Low performance robustness is related to high runoff skewness and aridity, variable baseflow contribution, and low rainfall-runoff ratio

Supporting Information:
- Supporting Information S1

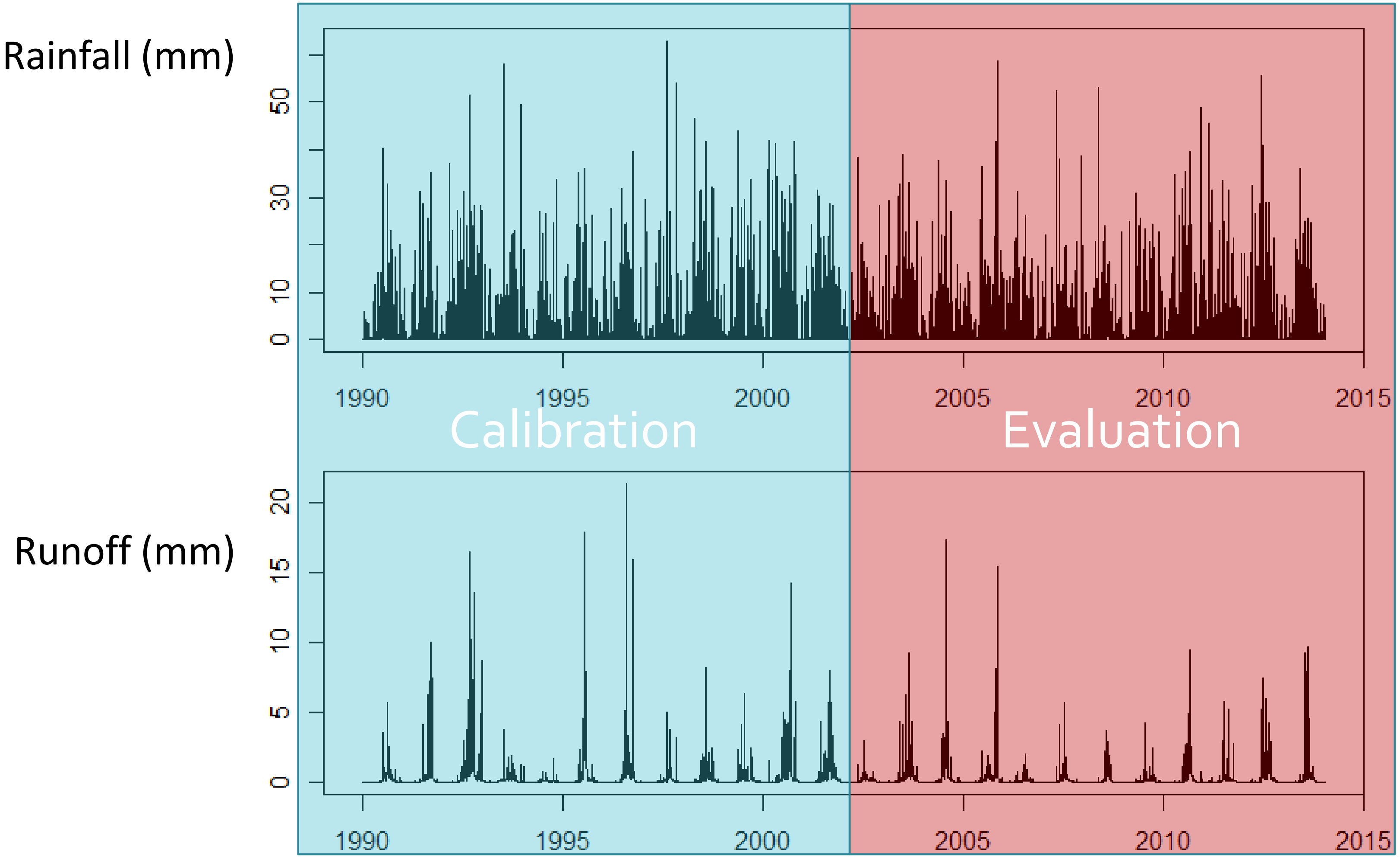
On the Robustness of Conceptual Rainfall-Runoff Models to Calibration and Evaluation Data Set Splits Selection: A Large Sample Investigation

Danlu Guo[1,2], Feifei Zheng[2], Hoshin Gupta[3], and Holger R. Maier[2,4]

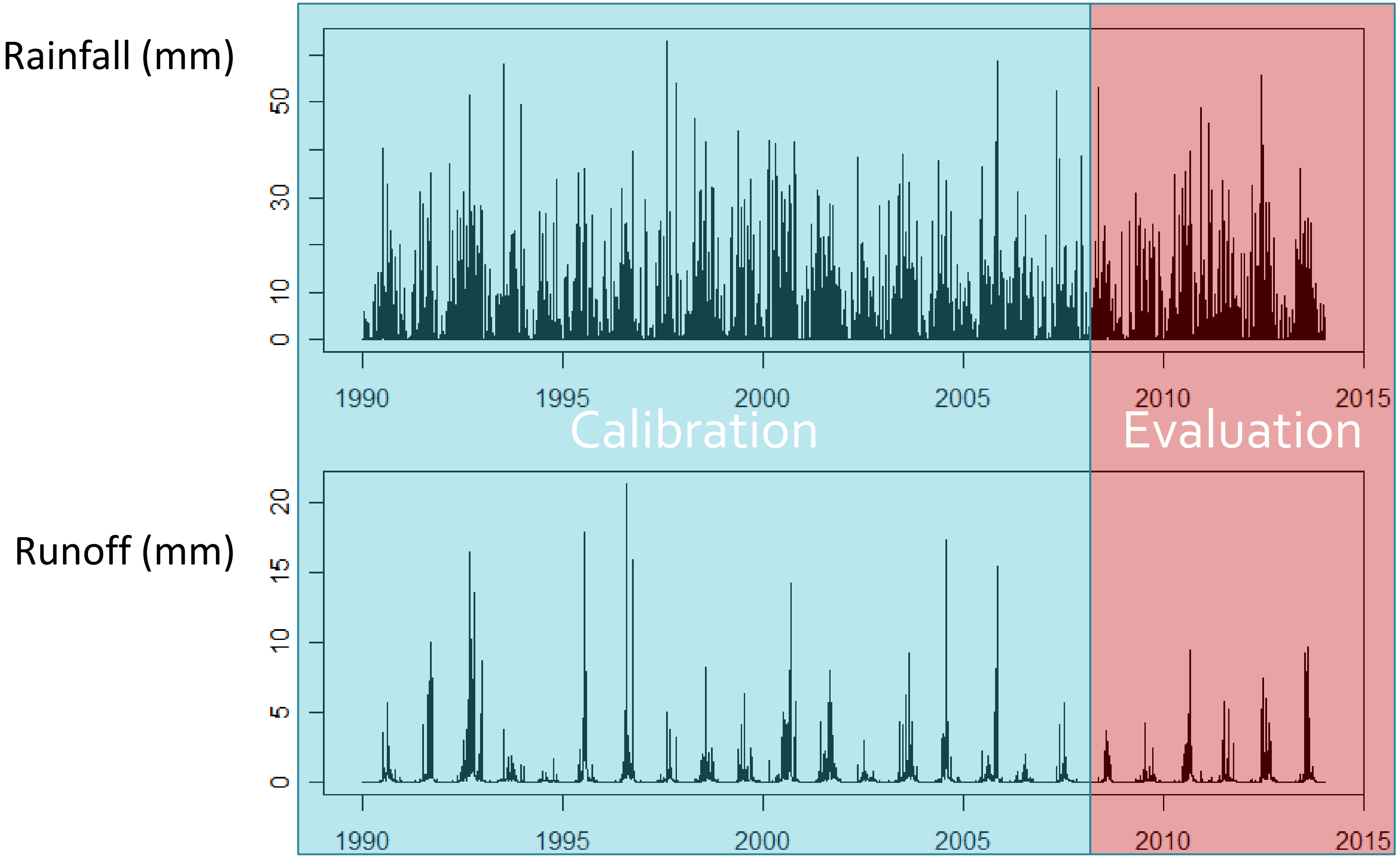
[1]Department of Infrastructure Engineering, The University of Melbourne, Parkville, Victoria, Australia, [2]College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, China, [3]Department of Hydrology and Atmospheric Sciences, The University of Arizona, Tucson, AZ, USA, [4]School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, South Australia, Australia

Abstract Conceptual rainfall-runoff (CRR) models are widely used for runoff simulation and for prediction under a changing climate. The models are often calibrated with only a portion of all available data at a location and then evaluated independently with another part of the data for reliability assessment. Previous studies report a persistent decrease in CRR model performance when applying the calibrated model to the evaluation data. However, there remains a lack of comprehensive understanding about the nature of this "low transferability" problem and why it occurs. In this study we employ a large sample

CRR model structure is determined by the data within calibration period

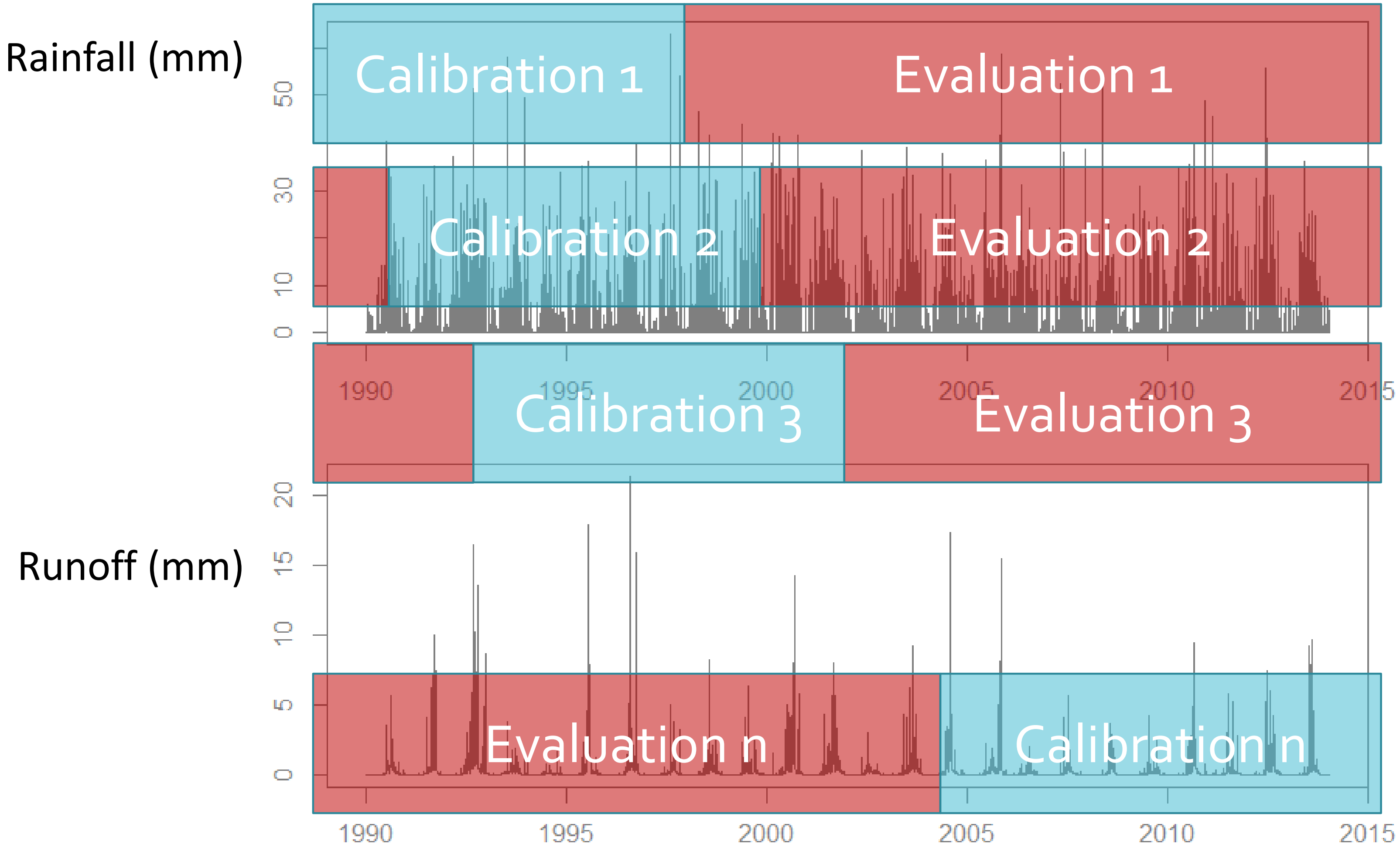


So, we have trouble if the hydro-climatic conditions differ too much between calibration & evaluation periods



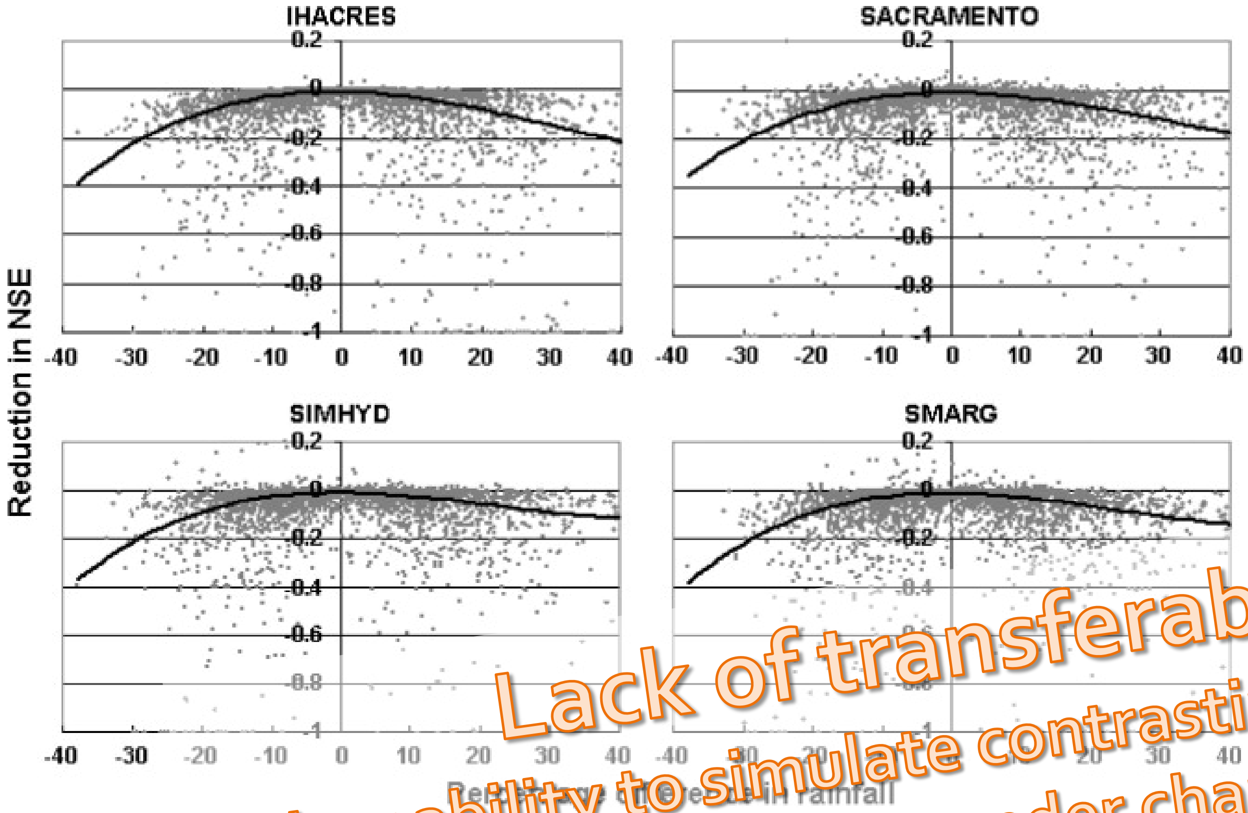
Split-sample test
has been used to understand

how performance differs between calibration and evaluation data



A consistent finding:

(CRR) models perform worse under conditions that differ from calibration period



e.g. Vaze et al. (2010)

and Bastola et al., 2011;
Merz et al., 2011
Coron et al., 2012;
Coron et al., 2014;
Broderick et al., 2016
…

Lack of transferability
(low ability to simulate contrasting conditions)
Indicating suitability under changing climate

The remaining question…

We know that CRR model performance decrease at a catchment, when calibration & evaluation conditions differ, but …

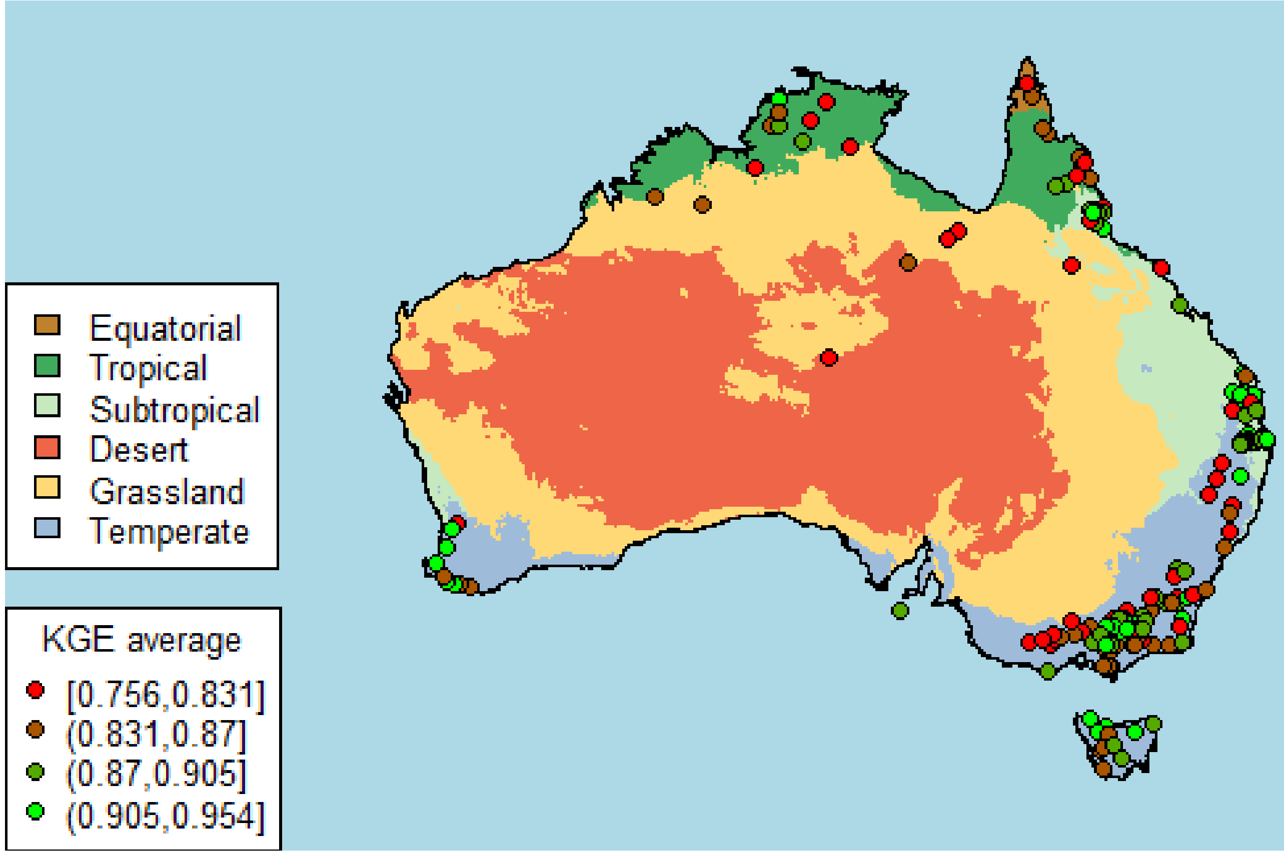
Do the extents of performance variation (robustness) differ across catchments?

How do they change across catchment characteristics?

↓

large-sample hydrology!

Included 163 HRS catchments (large-sample hydrology)

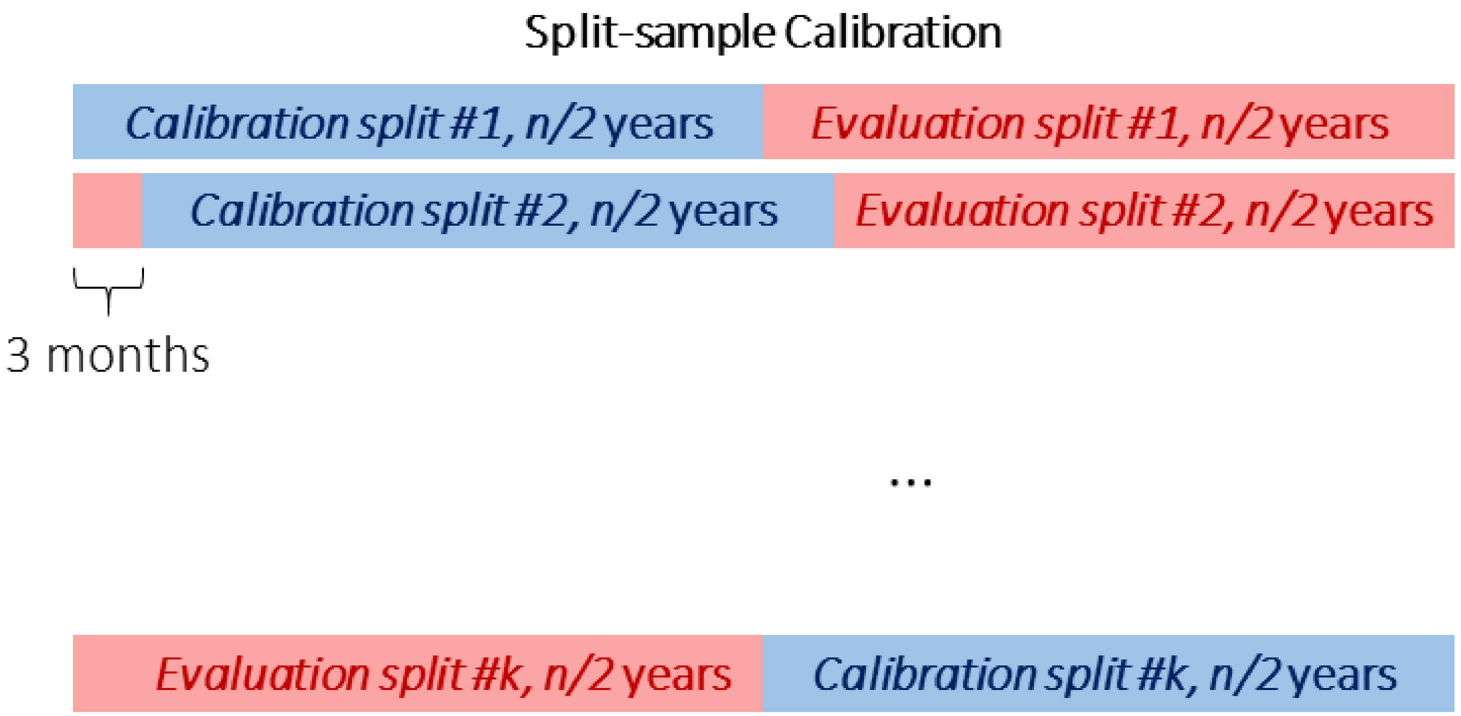


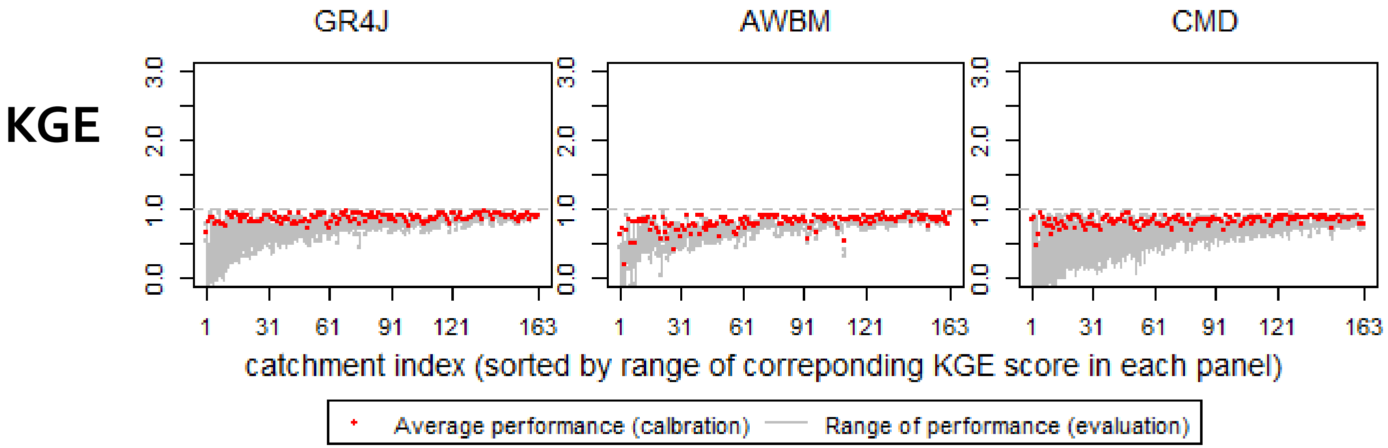
- Ran GR4J on all with SCE algorithm for parameter optimization
- 100 replicates on each catchment starting from different random seeds
- Filtered out catchments with:

1) <0.75 mean KGE;
2) high KGE variability (95 CI >3% mean KGE)

Large number of split-sample tests on each catchment

with 50:50 calibration: evaluation data split





How does CRR model robustness differ across catchments?

Overall KGE robustness

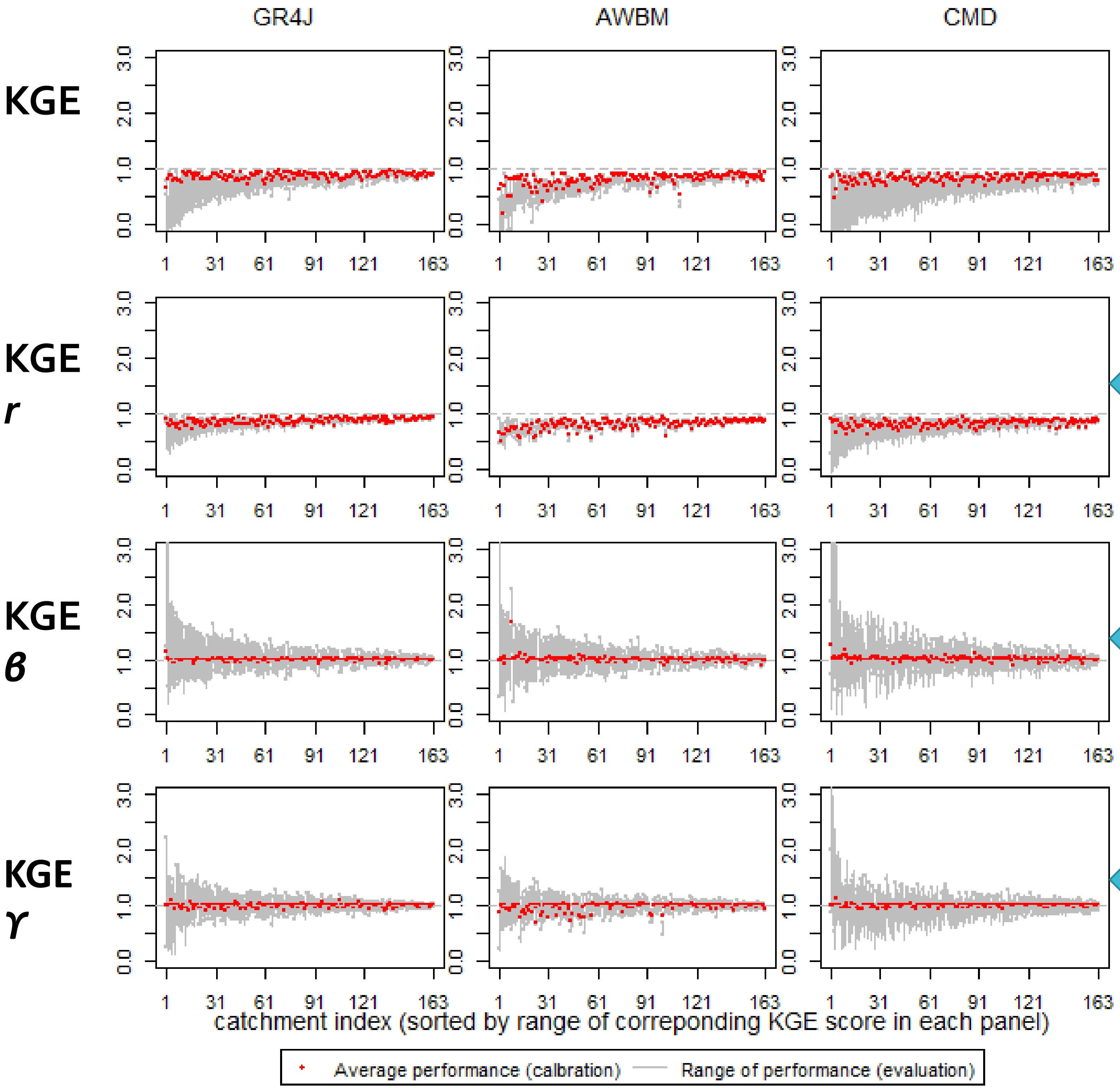
- Focusing on three common CRR models GR4J, AWBM and CMD
- We used **the range of evaluation performance from all data splits** to assess model robustness. Evaluation performance is based on KGE, a weighted average of how well the model simulates the observed data in:

1) serial correlation;
2) mean;
3) Variability

- Each grey bar summarizes the variation in evaluation performance from all the calibration/evaluation splits at each catchment along the x-axis (163 in total).
- The longer a grey bar is, the higher variability so low robustness the catchment has.

Across all KGE components, CRR models are more robust in simulating the serial correlation instead of the mean and variability.

This is likely due to the high correlation between runoff and rainfall (a model input), so that it is easier for a CRR model to obtain the serial correlation structure from input data

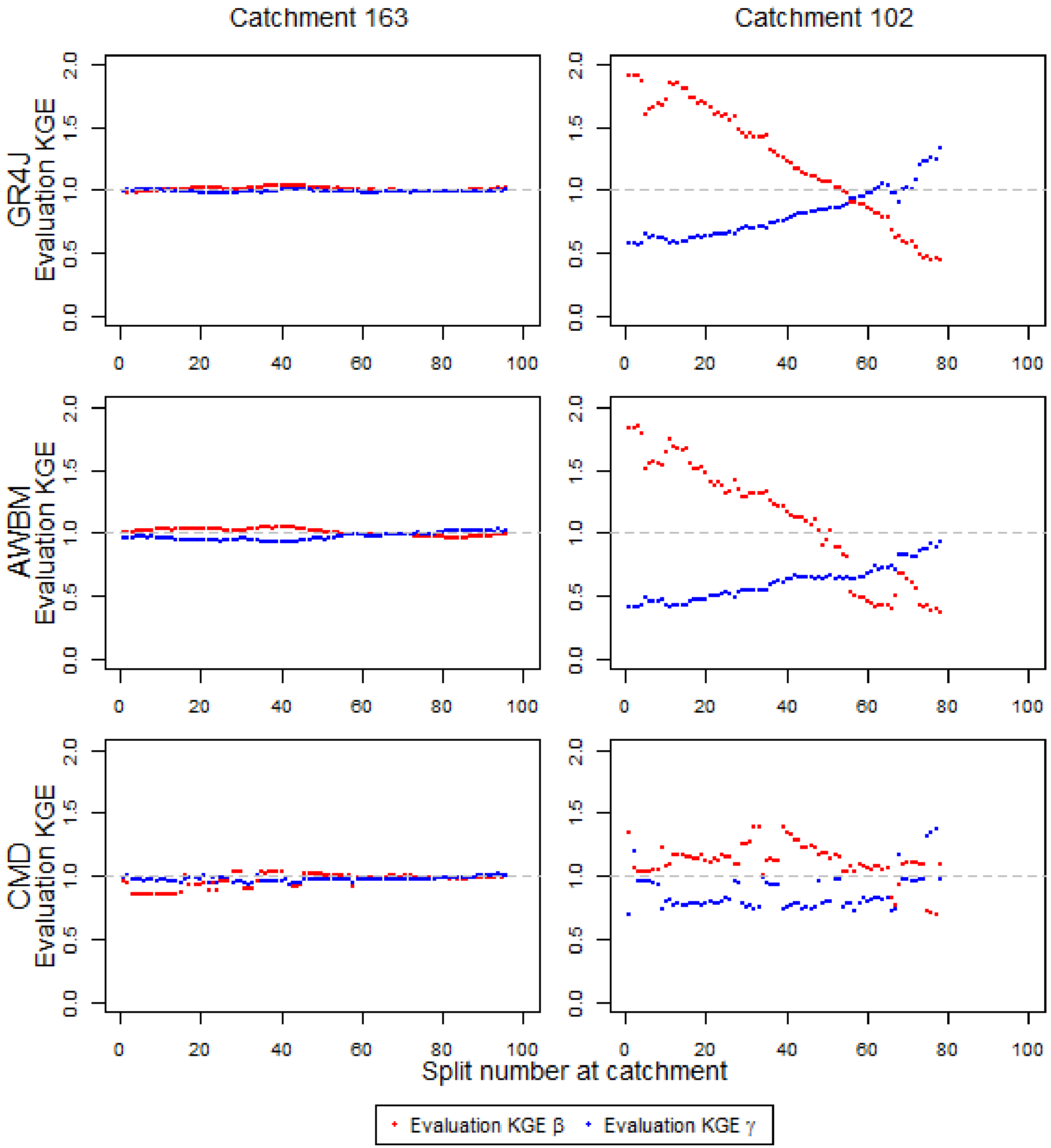


Sim. vs. Obs. correlation

Mean

Variability

Catchments show varying consistency of evaluation performance across different splits

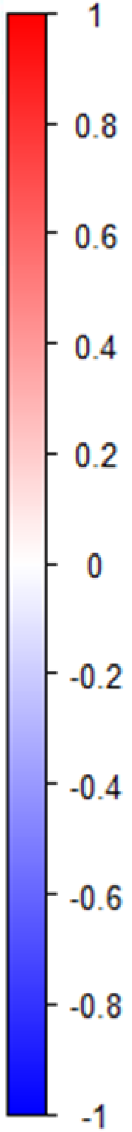


- This is illustrated with two extreme catchments using the GR4J results. Catchment 163 shows very consistent evaluation performance over time.
- **Catchment 102 indicates a 'compensation effect'** between **the evaluation *β* and *ϒ* (mean and variability)** – which negatively correlated with each other.
- Specifically, **overestimation of the mean (*β* > 1)** tends to be **associated with an underestimation of variability (*ϒ* <1)**, and vice versa.
- ~20-30 such catchments – suggesting possible time-varying change of the rainfall-runoff relationships that need to be further investigated.

What are the key factors for varying model robustness across catchments?

low robustness tends to occur at catchments with:
Low RR ratio,
high variability of baseflow contribution,
high runoff skewness

	Mean_runoff	Mean_rainfall	Mean_PET	Mean_BFI	Mean_aridity	Mean_RR_ratio	SD_BFI	SD_aridity	SD_RR_ratio	SD_runoff	Skew_runoff	Area
SD(KGE_GR4J)	-0.68	-0.67	0.18	-0.29	0.61	-0.65	0.55	0.48	-0.13	-0.33	0.52	0.07
SD(KGE_AWBM)	-0.68	-0.67	0.1	-0.27	0.57	-0.67	0.51	0.43	-0.2	-0.39	0.52	0.1
SD(KGE_CMD)	-0.46	-0.45	-0.28	-0.07	0.19	-0.43	0.13	0.12	-0.44	-0.42	0.12	0.1



- We now summarize model robustness with the SD of KGE at each catchment
- The SD values are correlation with a wide range of catchment characteristics including the long-term conditions and the mid- and short-term variability of rainfall, PET and runoff, together with the catchment topography.

- The key characteristics affecting CRR model robustness are:

1) long-term average Q & P
2) RR ratio and aridity
3) variability of annual BFI and Q skewness

Both studies found ***runoff skewness*** as a key factor related to ***low robustness***, but via different pathways!

Robustness of ANN rainfall-runoff models (Zheng et al., 2017)

AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE

10.1002/2017WR021470

Key Points:

- Impact of calibration and evaluation data allocation on model performance is examined using a large number of catchments
- Robustness of model performance can be very poor if statistical properties of the data are ignored during data allocation
- While obtained using ANN-type models, the results are broadly relevant to all classes of hydrological models

On Lack of Robustness in Hydrological Model Development Due to Absence of Guidelines for Selecting Calibration and Evaluation Data: Demonstration for Data-Driven Models

Feifei Zheng[1], Holger R. Maier[1,2], Wenyan Wu[2,3], Graeme C. Dandy[2], Hoshin V. Gupta[4], and Tuqiao Zhang[1]

ANN models require a data-split process before calibration, for which it is more difficult to allocate data of similar statistical properties across calibration/evaluation datasets – if runoff skewness is high.

Robustness of conceptual rainfall-runoff models (Guo et al., 2020)

Water Resources Research

RESEARCH ARTICLE

10.1029/2019WR026752

Key Points:

- We investigate the robustness of CRR models across calibration/evaluation data splits with a large-sample approach
- Different data splits markedly impact CRR model performance, particularly for accurately reproducing the mean and variability of runoff
- Low performance robustness is related to high runoff skewness and aridity, variable baseflow contribution, and low rainfall-runoff ratio

Supporting Information:

- Supporting Information S1

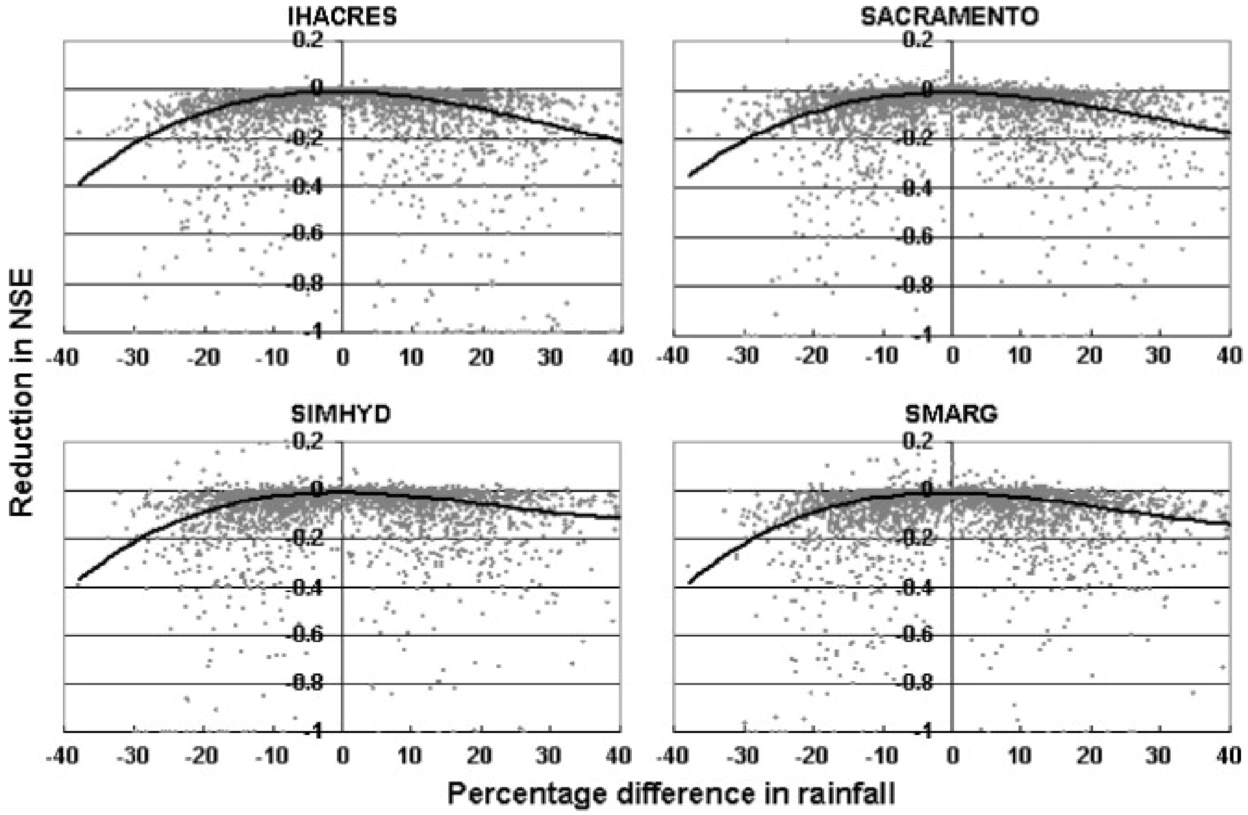
On the Robustness of Conceptual Rainfall-Runoff Models to Calibration and Evaluation Data Set Splits Selection: A Large Sample Investigation

Danlu Guo[1,2], Feifei Zheng[2], Hoshin Gupta[3], and Holger R. Maier[2,4]

[1]Department of Infrastructure Engineering, The University of Melbourne, Parkville, Victoria, Australia, [2]College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, China, [3]Department of Hydrology and Atmospheric Sciences, The University of Arizona, Tucson, AZ, USA, [4]School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, South Australia, Australia

For CRR models, high skewness tends to lead to smaller store capacities in the calibrated models (Appendix A2), which makes them 'less flexible' to deal with different hydro-climatic conditions.

Difference between ***robustness*** and ***transferability*** analyses



e.g. Vaze et al. (2010)

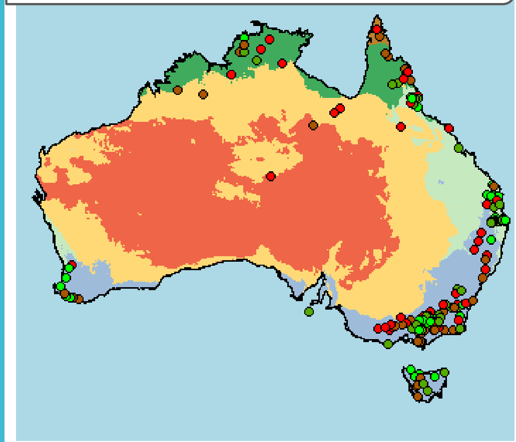
and Bastola et al., 2011;
Merz et al., 2011
Coron et al., 2012;
Coron et al., 2014;
Broderick et al., 2016
…

Transferability: relative difference of catchment conditions and model performance (calibration vs. evaluation) *-> how similar should calibration and evaluation conditions be to warrant unchanged model performance?*

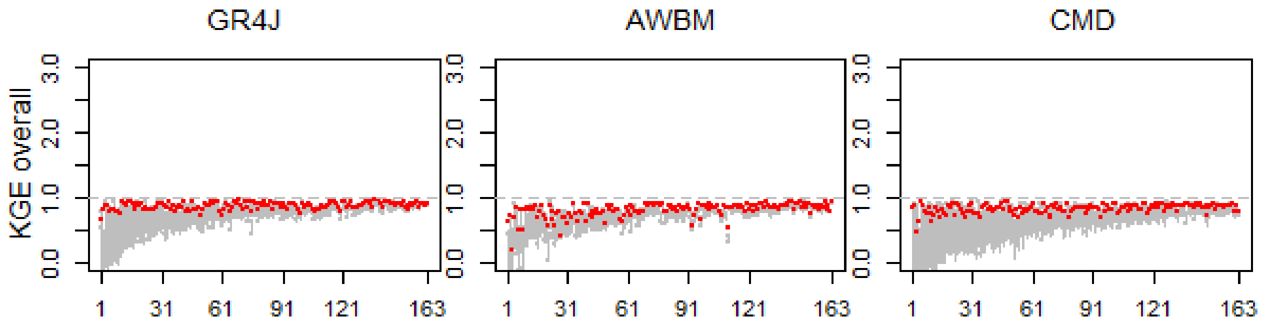
Robustness: variation in absolute model performance *-> under what catchment condition is a model more likely to have stable/unstable performance?*

A re-cap of the whole story…

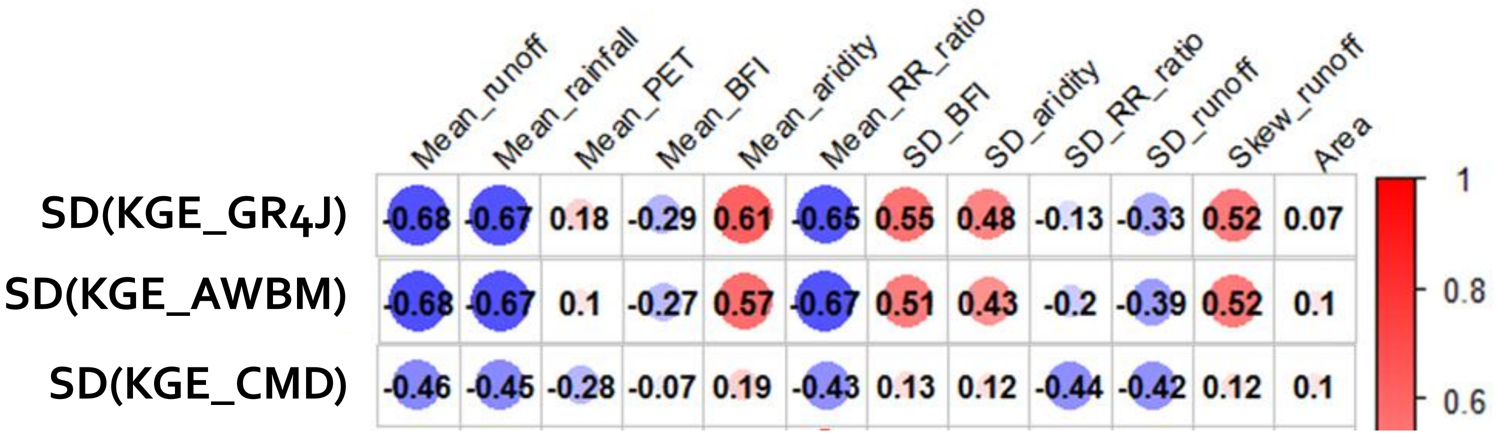
163 HRS catchments



Model performance (KGE) robustness from split-sample calibration



Low robustness tends to come with:
Low RR ratio, high variability of baseflow contribution, high runoff skewness (more arid?)

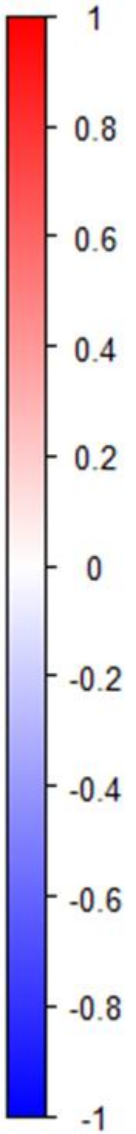


Full story: https://doi.org/10.1029/2019WR026752

Appendix

What are the key factors for varying model robustness across catchments? (Same as P13 but including cross-correlations)

	Mean_runoff	Mean_rainfall	Mean_PET	Mean_BFI	Mean_aridity	Mean_RR_ratio	SD_BFI	SD_aridity	SD_RR_ratio	SD_runoff	Skew_runoff	Area
SD_KGE_GR4J	-0.68	-0.67	0.18	-0.29	0.61	-0.65	0.55	0.48	-0.13	-0.33	0.52	0.07
SD_KGE_AWBM	-0.68	-0.67	0.1	-0.27	0.57	-0.67	0.51	0.43	-0.2	-0.39	0.52	0.1
SD_KGE_CMD	-0.46	-0.45	-0.28	-0.07	0.19	-0.43	0.13	0.12	-0.44	-0.42	0.12	0.1
Mean_runoff		0.91	-0.11	0.37	-0.79	0.98	-0.6	-0.65	0.55	0.75	-0.46	-0.37
Mean_rainfall			-0.01	0.31	-0.81	0.83	-0.52	-0.69	0.41	0.67	-0.48	-0.37
Mean_PET				-0.26	0.54	-0.18	0.47	0.45	0.43	0.29	0.42	0.17
Mean_BFI					-0.46	0.4	-0.73	-0.33	0.05	-0.05	-0.39	-0.11
Mean_aridity						-0.77	0.68	0.84	-0.12	-0.36	0.61	0.44
Mean_RR_ratio							-0.64	-0.64	0.54	0.71	-0.48	-0.35
SD_BFI								0.6	0.02	-0.14	0.6	0.05
SD_aridity									-0.02	-0.28	0.61	0.33
SD_RR_ratio										0.8	0.25	-0.31
SD_runoff											0.13	-0.33
Skew_runoff												0



- **Rows 1-3:** Spearman's rank correlations between model robustness (SD of KGE) and 12 catchment hydro-climatic characteristics, across all 163 catchments.
- **Rows 4-14:** Correlations amongst the 12 hydro-climatic catchment characteristics.
- The plot highlights **key factors related to low model robustness (solid blue and red dots in Rows 1-3) and their cross-correlation.**

Lower modelled store capacity are found at catchments with higher runoff skewness

- Relationships between Mean_RR_ratio, Mean_aridity, Skew_runoff and SD_BFI, and the calibrated store capacity of each catchment from GR4J, AWBM and CMD (all in mm, averaged across all calibration subperiod, so each dot represents a catchment.
- Shaded cells highlight correlations and scatter plots between each pair of catchment characteristic and calibrated store capacity. The pairwise Spearman's correlations are shown in the top-right triangle. **The plot illustrates the relationship between low store capacity and high runoff skewness.**