Water Resources Research

Supporting Information for

Simulating runoff under changing climatic conditions: revisiting an apparent deficiency of conceptual rainfall-runoff models

Keirnan J. A. Fowler[1], Murray C. Peel[1], Andrew W. Western[1], Lu Zhang[2] and Tim J. Peterson[1]

[1]Department of Infrastructure Engineering, University of Melbourne, Victoria, Australia

[2]CSIRO Land and Water Flagship, Canberra, ACT, Australia

Contents of this file

Introduction

This supporting information contains the following:

1. **Pareto results for every case study examined in this study (Figures S1 and S2).** Specifically, we provide a plot for each catchment showing the Pareto curve obtained for each of the five model structures, for a total of 430 curves.

2. **Comparison table for Kling Gupta Efficiency versus Nash Sutcliffe Efficiency (Figure S3)** highlighting problems noted by Gupta et al. (2009).

3. **Plotted comparison of calibration results between AMALGAM and the single-objective optimizer CMA-ES (Figure S4)**. This comparison was done in ten catchments, for five model structures, for each of the two objectives.

4. **Four additional case studies pursuant to Section 3.2 of the article (Figure S5).** Section 3.2 showed an example (Figure 11) where the endpoints of two Pareto curves (those for GR4J and GR4JMOD) were similar despite divergence mid-curve. As the article explains, "…use of a single-objective DSST…would lead to the erroneous conclusion that the alterations to GR4JMOD by Hughes et al. (2013) made negligible difference to the model's capabilities. In contrast, Figure 11 shows that this is not the case by the divergence of the purple GR4JMOD curve from the orange GR4J curve." To demonstrate that Figure 11 is not an isolated case, in Figure S4 we provide examples from four other catchments.

5. **Catchment-by-catchment information for the 12 catchments in which no model structure was able to meet modelling standard 1 (Table S1).** We provide additional information for the interested reader who wishes to know the names and physical properties of the catchments that were most difficult to model.

6. **Catchment location map (Figure S6)** showing location of catchments where no model structure could meet a given modelling standard, referenced in Section 3.4 of the article.

7. **Text discussing categorization of model failure based on the shape of the Pareto Curve (Text S1).** This text expands upon the summary provided in Section 3.4.

8. **Diagram in support of Text S1 (Figure S7)** providing a visual summary of model failure categorization for each model structure.

9. **Text regarding bushfires (Text S2)** outlining efforts to investigate whether the hydrological effects of bushfires might be associated with catchments where no model structure could meet a given modelling standard. Referenced in Section ## of the article.

10. **Diagrams of simulation bias for selected parameter sets (Figures S8 to S11)**, following a format suggested by Coron et al. (2014). These plots are the same as Figure 13 of the article, but are shown for more case studies.

11. **References.**

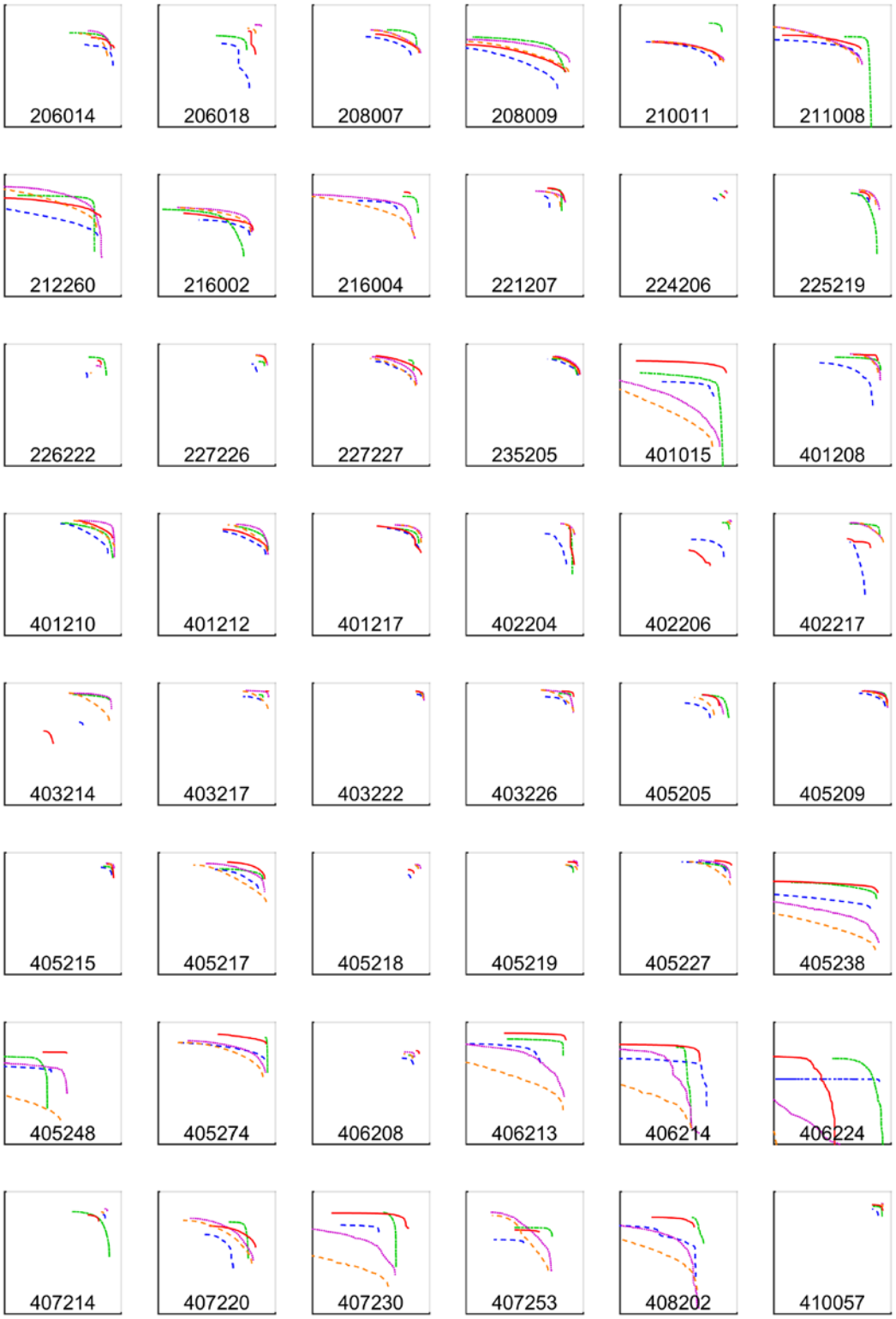


Figure S1. Pareto curves, for each of the five model structures, in 48 of the 86 study catchments. The y axis is KGE over non-dry period (axis limits: 0,1) and the x axis is KGE over dry period (axis limits: 0,1). For plot legend, refer next figure.

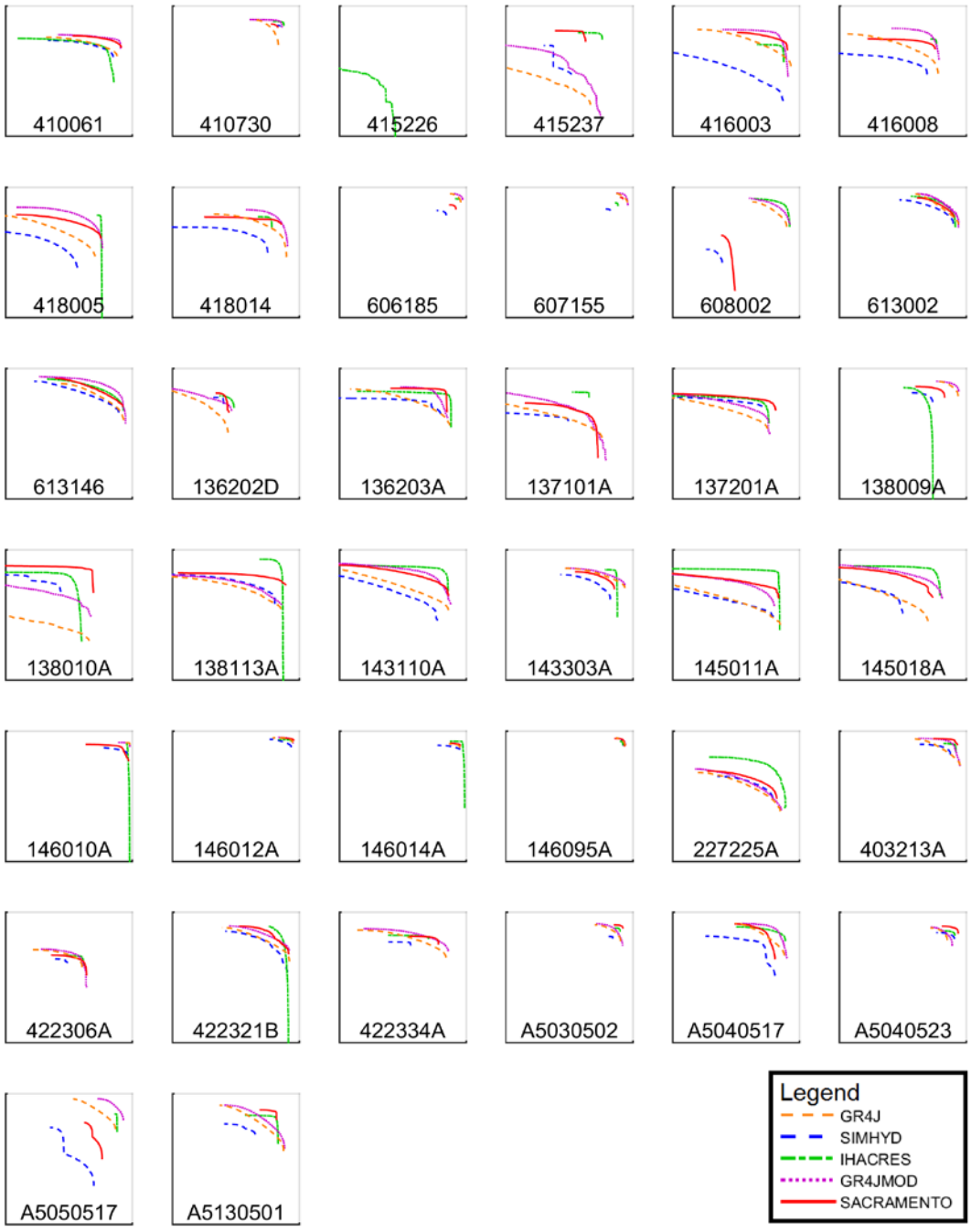
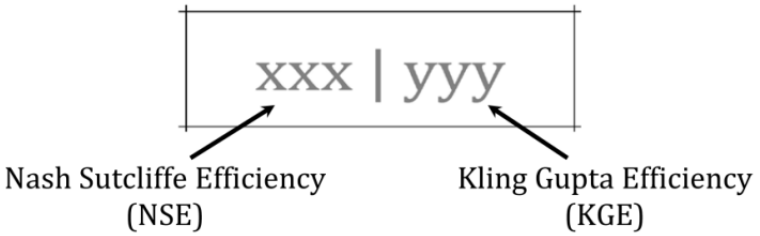


Figure S2. Pareto curves, for each of the five model structures, in 38 of the 86 study catchments. The y axis is KGE over non-dry period (axis limits: 0,1) and the x axis is KGE over dry period (axis limits: 0,1). For a plot legend, refer next figure.



Low variability of flows (ie. σ_{daily} = 0.7mm/d)

NSE and KGE values if Linear Correlation = 0.90

Relative Bias \ Relative error in standard deviation	0%	-5%	-10%	-20%	-30%
0%	**0.80** | 0.90	**0.81** | 0.89	**0.81** | 0.86	**0.80** | 0.78	**0.77** | 0.68
-5%	0.80 | 0.89	0.81 | 0.88	0.81 | 0.85	0.80 | 0.77	0.77 | 0.68
-10%	0.79 | 0.86	0.80 | 0.85	0.80 | 0.83	0.79 | 0.76	0.76 | 0.67
-20%	0.76 | 0.78	0.77 | 0.77	0.77 | 0.76	0.76 | 0.70	0.73 | 0.63
-30%	0.71 | 0.68	0.72 | 0.68	0.72 | 0.67	0.71 | 0.63	0.68 | 0.56

High variability of flows (ie. σ_{daily} = 4.0mm/d)

NSE and KGE values if Linear Correlation = 0.90

Relative Bias \ Relative error in standard deviation	0%	-5%	-10%	-20%	-30%
0%	**0.80** | 0.90	0.81 | 0.89	0.81 | 0.86	0.80 | 0.78	0.77 | 0.68
-5%	**0.80** | 0.89	0.81 | 0.88	0.81 | 0.85	0.80 | 0.77	0.77 | 0.68
-10%	**0.80** | 0.86	0.81 | 0.85	0.81 | 0.83	0.80 | 0.76	0.77 | 0.67
-20%	**0.80** | 0.78	0.81 | 0.77	0.81 | 0.76	0.80 | 0.70	0.77 | 0.63
-30%	**0.80** | 0.68	0.80 | 0.68	0.81 | 0.67	0.80 | 0.63	0.77 | 0.56

Figure S3. Values for Nash Sutcliffe Efficiency (NSE) and Kling Gupta Efficiency (KGE) for different combinations of standard deviation σ and bias β. Linear correlation r is constant at 0.9. The analysis is done for two values of standard deviation: σ = 0.7 mm/d (10th percentile – 9 catchments of 86 had a lower value); and σ = 4.0 mm/d (90th percentile – 9 catchments of 86 had a higher value). Note that Equation 4 from Gupta et al. (2009) requires the average flow to be defined in absolute terms; a value of 0.7 mm/d or 255 mm/yr was assumed, which was close to the median value for the study catchments. Orange and purple values mark issues with the NSE, as follows: *Orange:* Gupta et al. (2009) noted that the bias term is normalised by the observed standard deviation, which means that in catchments with high flow variability (as in this study) the magnitude of the bias can be high without penalising the NSE score. *Purple*: Gupta et al. (2009) noted a further problem with the treatment of the standard deviation σ in the NSE, regarding the ratio $\sigma_{simulated}$ / $\sigma_{observed}$. Although this ratio should ideally have a value of unity, the optimum value for NSE occurs when the ratio is equal to the linear correlation.

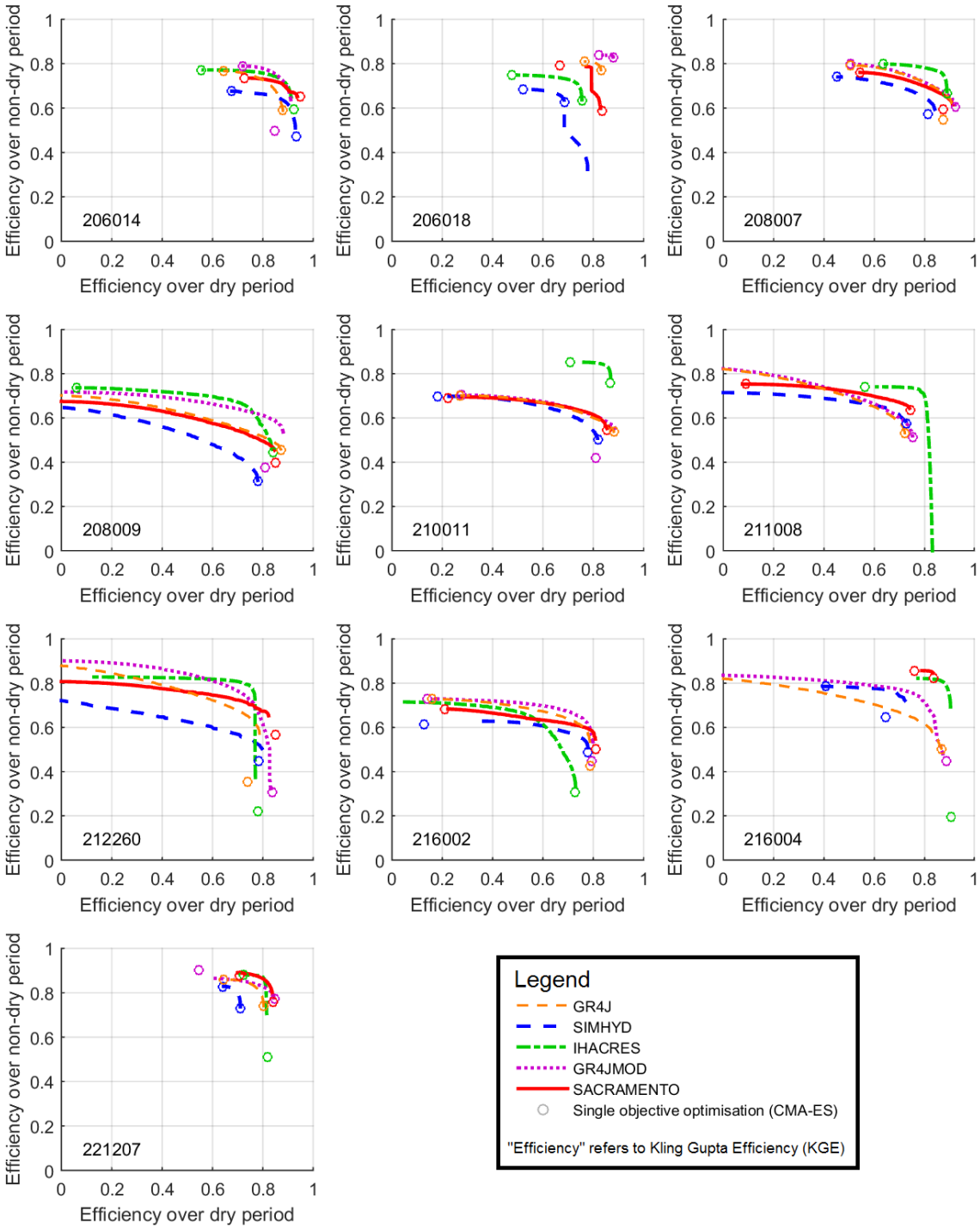


Figure S4. Pareto curves, for each of the five model structures, in 10 selected study catchments, along with single-objective optimization results using CMA-ES.

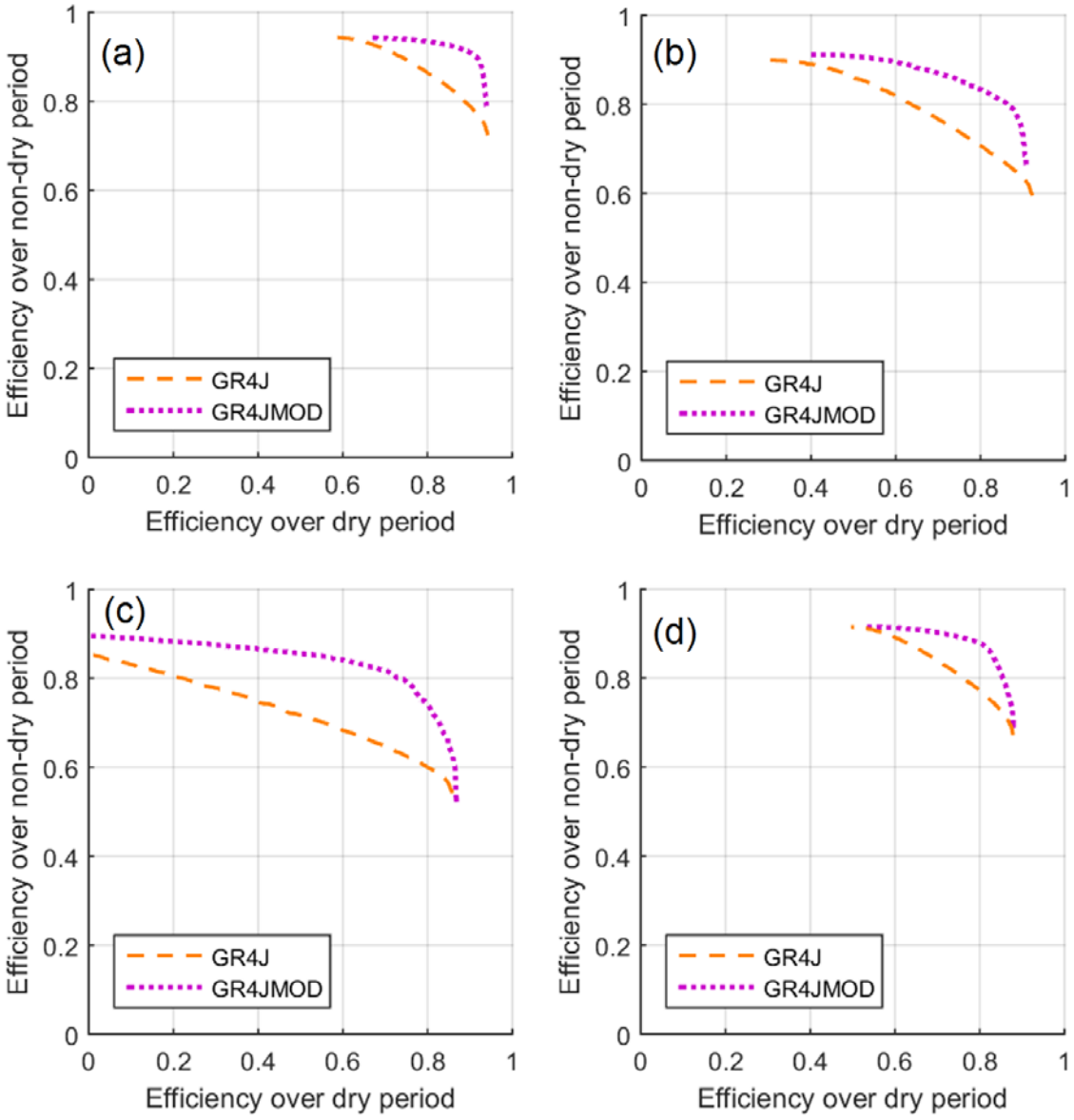


Figure S5. Four cases where the GR4J and GR4JMOD curves have similar endpoints. In each case the single-objective Differential Split Sample Test results would lead to the erroneous conclusion that the improvements introduced into GR4JMOD by Hughes et al. (2013) made negligible difference to the model's capabilities. (a) 401210 Snowy Creek below Granite Flat, Victoria; (b) 405217 Yea River at Devlins Bridge, Victoria; (c) 143110A Bremer River at Adams Bridge, Queensland; (d) A5040517 First Creek at Waterfall Gully, South Australia.

Catchment number & name	Catch-ment area *(km²)*	Mean annual rainfall *(mm/yr)*	Dry period flow ratio *(-)*	Average slope *(%)*	Forest Cover *(%)*	Farm Dam Develop-ment *(ML/km²)*
208009 - Barnard River at Barry	152	958	0.766	23.1	98.4	n/a
216002 - Clyde River at Brooman	862	1066	0.521	22.7	97.4	n/a
405248 - Major Creek at Graytown	292	608	0.061	4.3	72.3	4.13
406214 - Axe Creek at Longlea	236	600	0.068	5.0	71.4	32.06
406224 - Mount Pleasant Creek at Runnymede	243	518	0.053	2.4	21.4	6.93
407253 - Piccaninny Creek at Minto	681	491	0.441	1.5	59.3	22.54
408202 - Avoca River at Amphitheatre	77	616	0.125	13.3	49.5	5.35
415226 - Richardson River at Carrs Plains	130	504	0.002	1.3	17.9	2.58
136202D - Barambah Creek at Litzows	652	876	0.158	9.1	96.1	n/a
137101A - Gregory River at Burrum Highway	454	918	0.154	4.1	98.3	n/a
138010A - Wide Bay Creek at Kilkivan	352	819	0.056	10.4	97.7	n/a
422306A - Swan Creek at Swanfels	82	1012	0.345	24.2	96.4	n/a

Table S1. Characteristics for each of the 12 study catchments in group FF; that is, for which no model structure was able to meet modelling standard 1 ($KGE_{non\text{-}dry}$ = 0.7; KGE_{dry} = 0.7). Farm dam development data is only available for catchments in the state of Victoria.

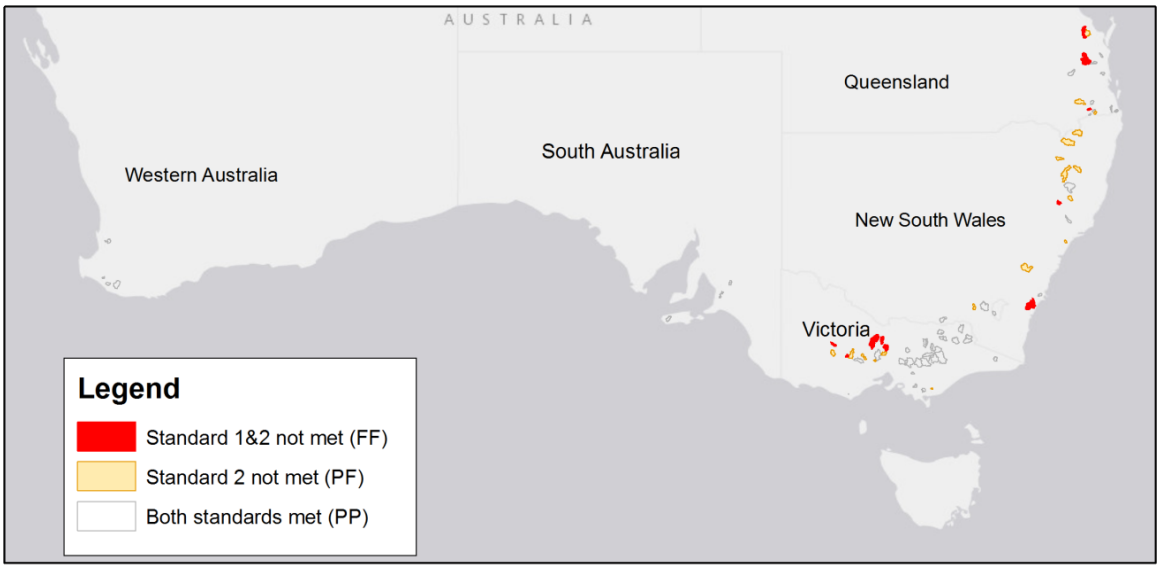


Figure S6. Map of catchments where no model structure could meet a given modelling standard.

Text S1 – Categorisation of model failure based on shape of the Pareto Curve

In this section we examine those catchments where the model structures failed to meet Standard 1 and/or Standard 2. We analyse the Pareto Curves and consider what the form of these curves may indicate about the type of model failure.

Figure S7 divides the instances where model structures failed to meet the standards (ie. Case C in Figure 14) into sub-categories according the type of failure, which we deduce from the form of their Pareto Curves. Having failed to meet the standard, every instance listed in these tables is one where no single parameter set can simulate flows satisfactorily in both wet and dry periods. The endpoints of the curves indicate whether the model structure is able to meet the standard in a given objective when optimized to it in isolation. If it can meet one or both standards in isolation, we classify this as a different type of failure compared to instances where the standard can be met in neither objective.

The results are relatively well spread between the failure types, making generalisation difficult (Figure S7). GR4J and GR4JMOD were exceptionally good at meeting the lower standard (KGE = 0.7) in a given objective provided that they were calibrated to it in isolation; hence zero catchments in Failure Type 1 (see also Figure 7). IHACRES had a much lower failure rate (hence lower totals) but among failure instances, Failure Type 1 was relatively common; that is, IHACRES failures tended to belong to the “appears to be deficient in this catchment, regardless of climate” category, particularly for the more stringent of the two standards (Figure S7). This category was also common for the model structure SIMHYD. It was more common for model structures to fail to meet the KGE=0.7 standard in the dry period (Failure types 3 and 1) than in the wet (Failure types 2 and 1). This distinction was not evident for KGE = 0.8. In general, the lack of a dominating category, particularly for the higher standard, speaks against the generalisation that model structures are generally poor at simulating dry conditions.

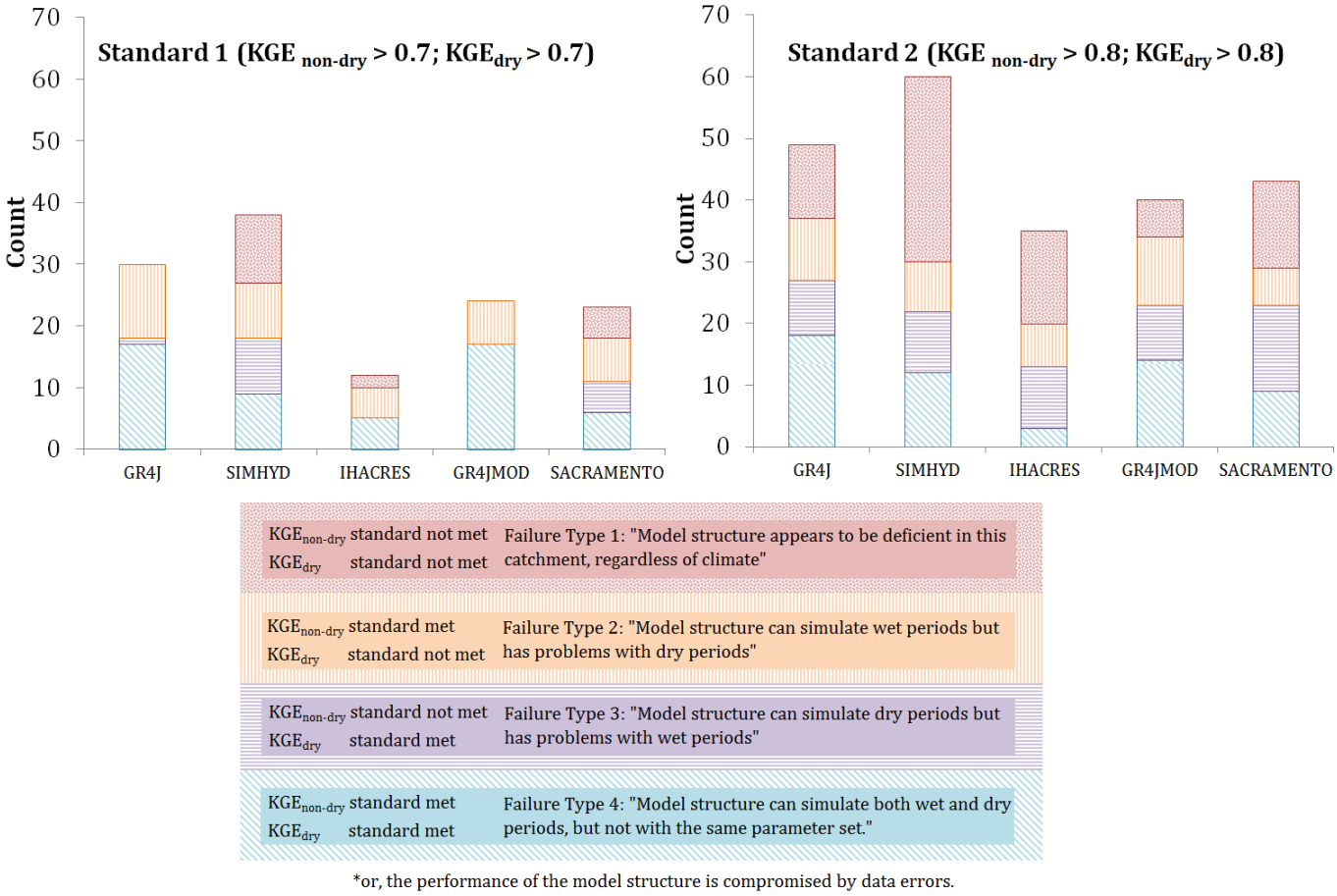


Figure S7. Details of model failure for those catchments where a given model structure failed.

Text S2 – Discussion of bushfires

We considered the possible effect of bushfires on the hydrology of the test catchments. As with farm dams, data on historic bushfire severity was only available for the State of Victoria (Department of Environment, Land, Water and Planning, 2015). There have been a number of significant bushfires in Victoria over the last 15 years, notably in 2003, 2007 and 2009. Also relevant to this study was a significant fire in 1983. Bushfires in Australia have been shown to cause changes in catchment hydrology, including a decrease in plant water use immediately following a fire, followed by a period of higher water use as saplings recolonise (eg. Kuczera 1987; Cornish and Vertessy, 2001). These changes may cause temporary degradation of performance of a calibrated model. In the current study, bushfires could confound the results if the 7-year dry period immediately followed a bushfire. If this was the case, bushfires could be a reason for model failure in the dry period. Given that the modelling period adopted in this study runs to 2010, the 2007 and 2009 fires occurred too late in the record (ie. are too recent) for this confounding to occur. To examine the 2003 and 1983 fires, we looked for catchments where the 7-year period was immediately after the fire. Although many catchments had dry periods starting around 2003 (Figure 5b), only three of these catchments were within the extent of the 2003 bushfires (401217, 403222 and 402204) and none of these three failed Modelling Standards 1 or 2. Very few catchments had dry periods starting around 1983 (Figure 5b); only one of these was in Victoria (407214) and this catchment was not burned in the 1983 fires. Based on the above discussion, it is concluded that there is no evidence that poor model performance observed in Groups PF and FF is the result of bushfires.

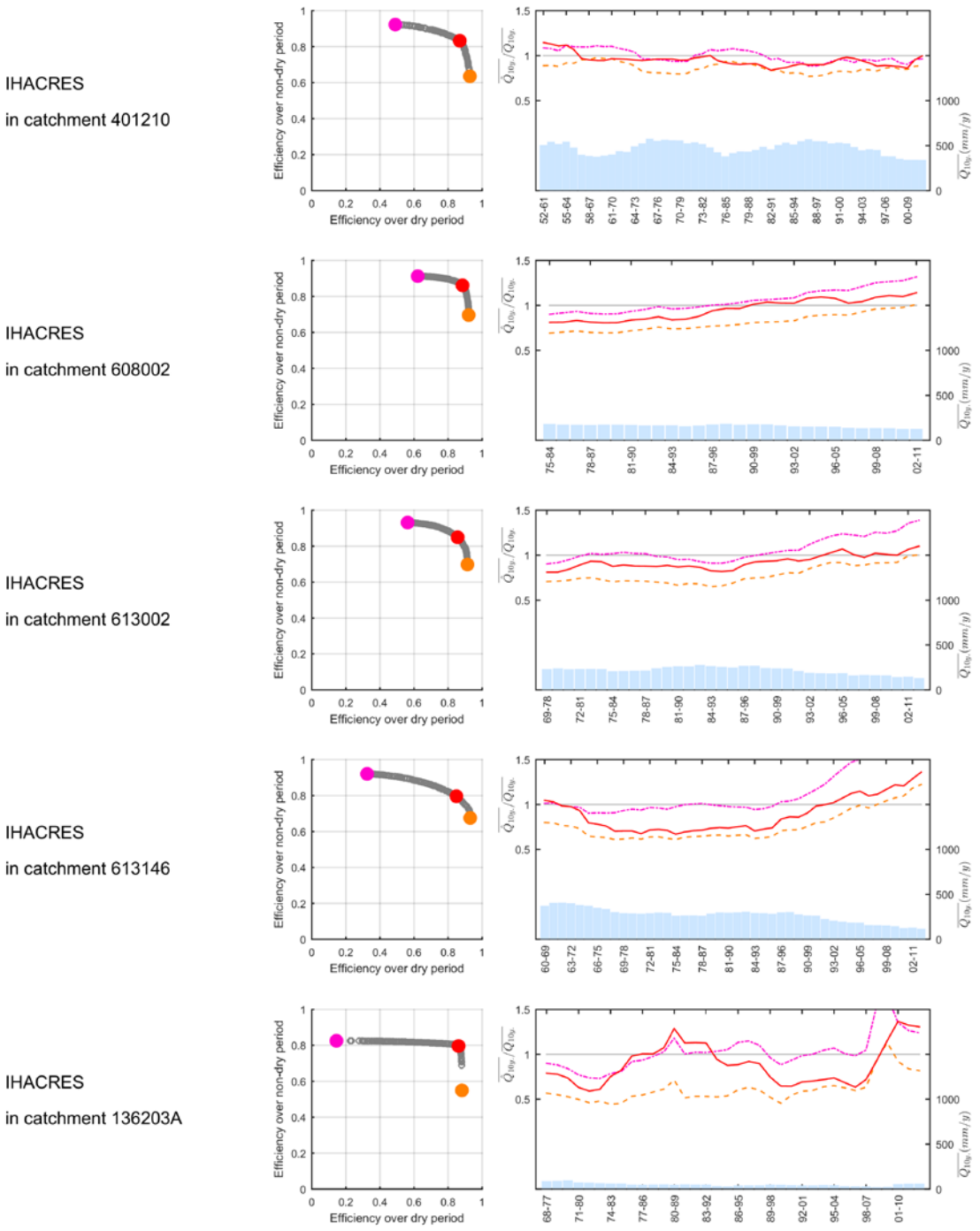


Figure S8. Long-term simulation bias for three selected parameter sets, after Coron et al. (2015, Figure 5), 1 of 4. Simulation bias is plotted as a ten-year moving average for three selected parameter sets from the Pareto curve. The ten-year average streamflows are also plotted for reference, in blue. This figure is the first of a series of four, showing a total of 21 case studies selected based on the shape of their Pareto Fronts. Specifically, three criteria were used to select these 21 case studies: (1) At least 1 parameter set within [0.8+, 0.8+] - that is, the model structure meets Standard 2 in this catchment (refer paper for definitions); (2) Calibrating to the non-dry period in isolation results in reductions in KGE_{dry} by at least 0.10 (ie. to 0.70) relative to the parameter set in (1); and (3) Calibrating to the dry period in isolation results in reductions in $KGE_{non\text{-}dry}$ by at least 0.10 (ie. to 0.70) relative to the parameter set in (1).

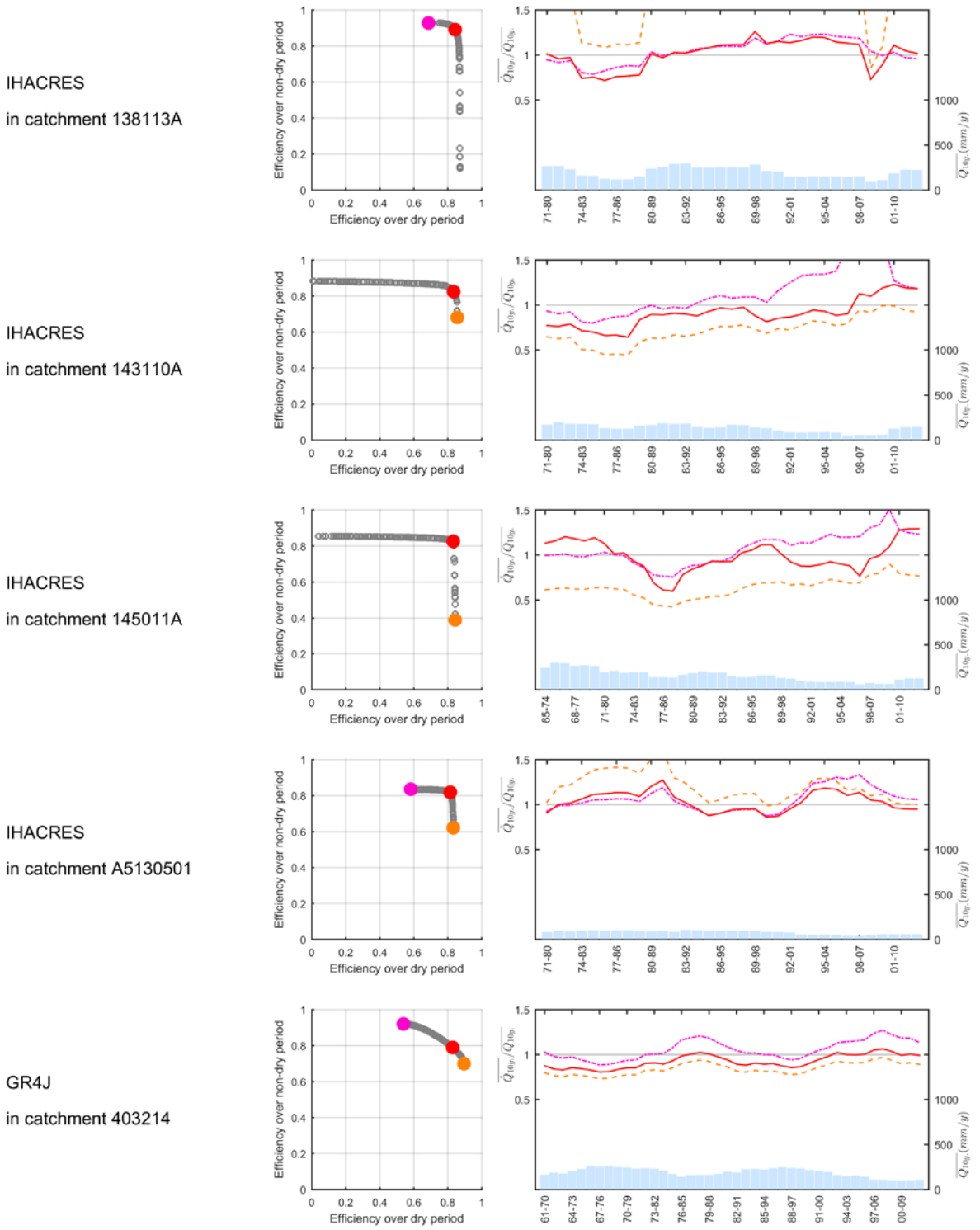


Figure S9: Long-term simulation bias for three selected parameter sets, after Coron et al. (2015, Figure 5), 2 of 4.

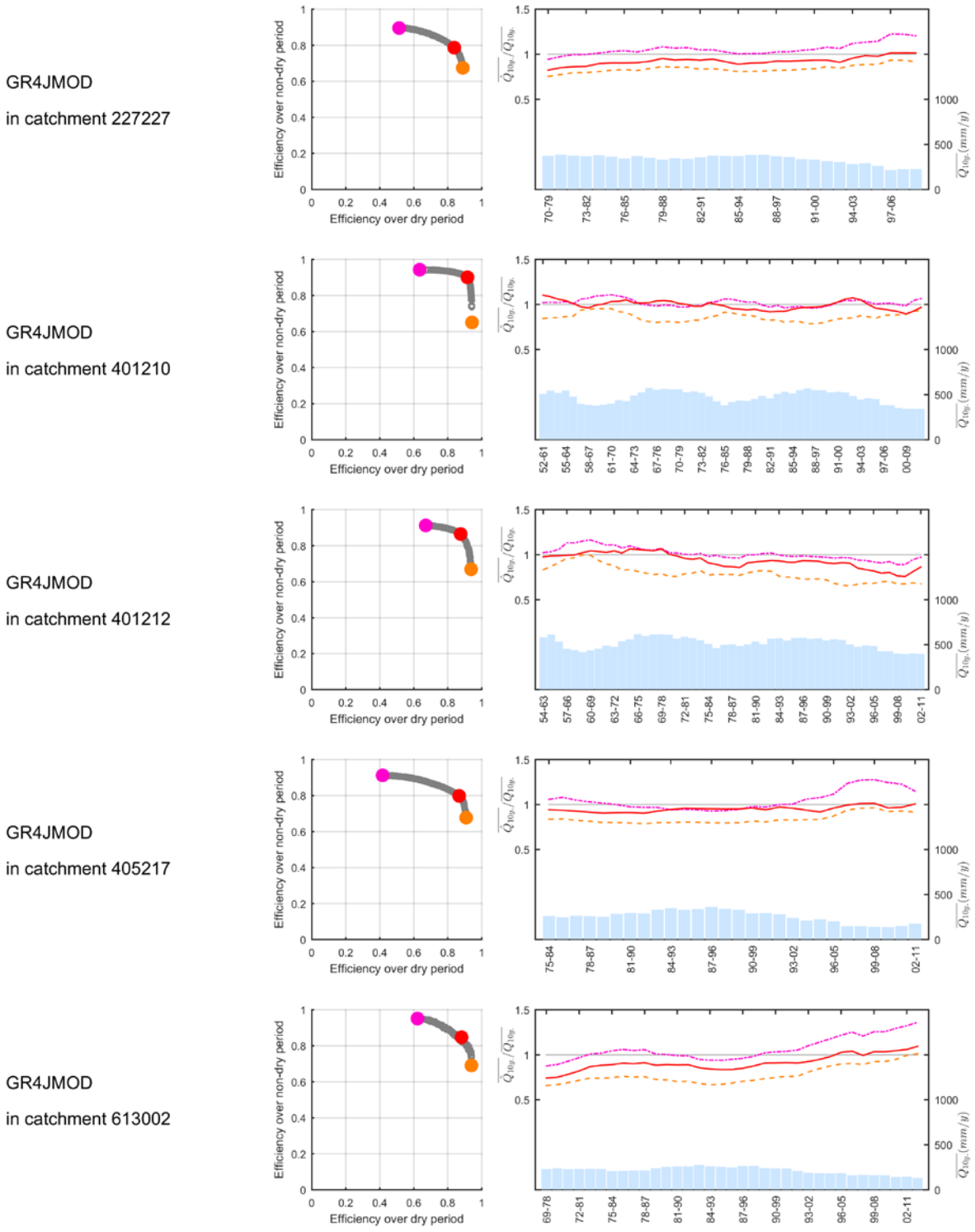


Figure S10: Long-term simulation bias for three selected parameter sets, after Coron et al. (2015, Figure 5), 3 of 4.

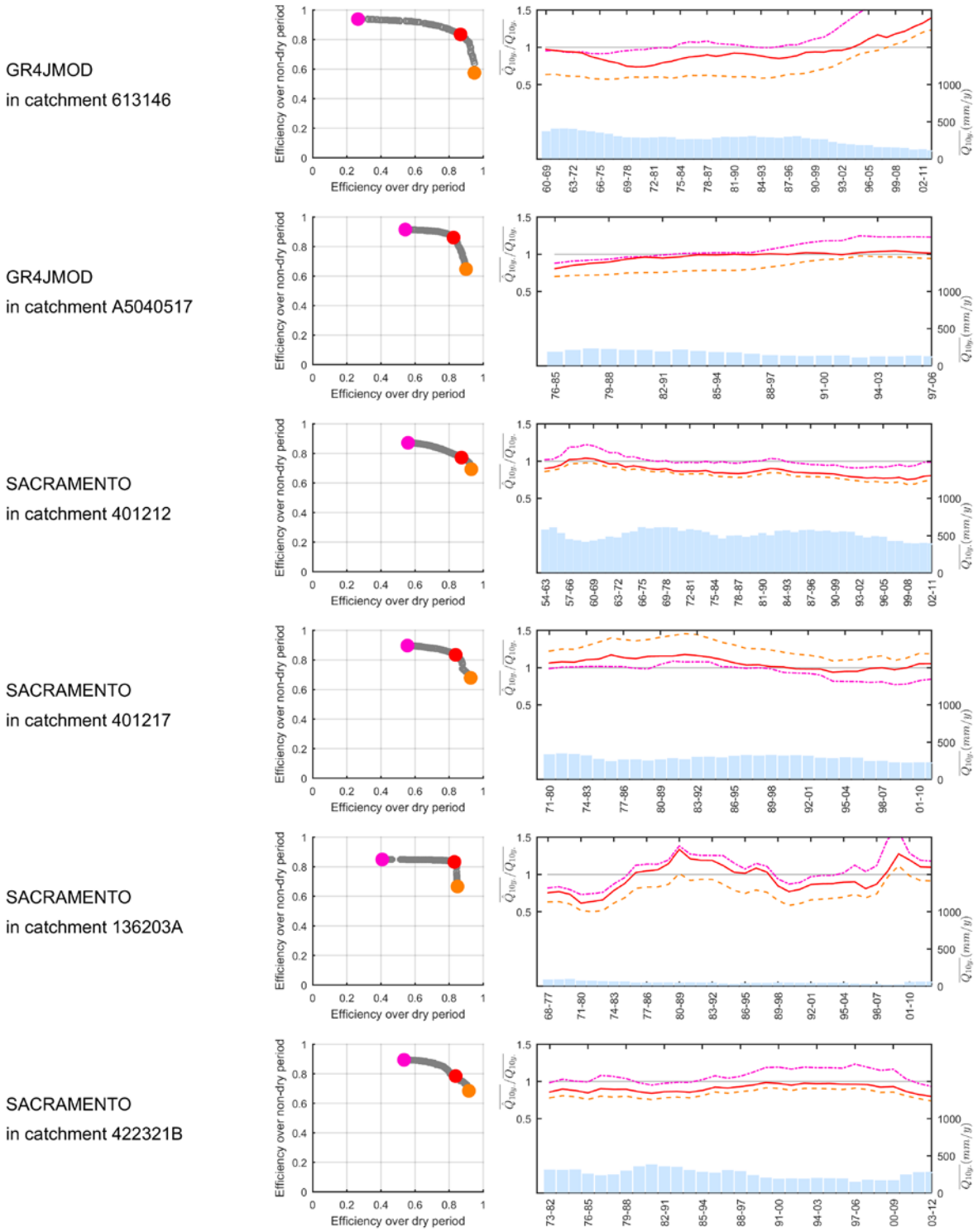


Figure S11: Long-term simulation bias for three selected parameter sets, after Coron et al. (2015, Figure 5), 4 of 4.

References

Cornish, P. M., & R. A. Vertessy (2001). Forest age-induced changes in evapotranspiration and water yield in a eucalypt forest. Journal of Hydrology, 242(1), 43-63.

Coron, L., V. Andréassian, C. Perrin, M. Bourqui, and F. Hendrickx (2014), On the lack of robustness of hydrologic models regarding water balance simulation: a diagnostic approach applied to three models of increasing complexity on 20 mountainous catchments. Hydrol. Earth Syst. Sci., 18(2), 727-746.

Department of Environment, Land, Water and Planning (2015), Historic Bushfire severity. Publically available GIS dataset "BLD_FIRE_HISEV" available at http://services.land.vic.gov.au/SpatialDatamart/search.html?action=search, accessed 04/08/2015.

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez (2009), Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol., 377(1), 80-91.

Hughes, J., R. Silberstein, and A. Grigg (2013), Extending rainfall-runoff models for use in environments with long–term catchment storage and forest cover changes. In Piantadosi, J., R.S. Anderssen, and J. Boland (eds) MODSIM2013, 20th International Congress on Modelling and Simulation.

Kuczera, G. (1987). Prediction of water yield reductions following a bushfire in ash-mixed species eucalypt forest. Journal of Hydrology, 94(3), 215-236.