

The remarks made by the Reviewers and the Associate Editor are written in italics, and the replies in normal font and blue color. In the citation of the remarks, the original page, paragraph, table, figure and line numbers have been used. However, in the replies these numbers refer to the numbers in the revised manuscript unless otherwise specified.

## **Reviewer #2**

*The authors introduce the OGW (optimization of gauge weights) method and compare their results with the IDW (inverse distance weighting) method, based on data from seven catchments in Australia. Furthermore, results of a case study are presented. This is an interesting study, which deserves to be published – but the presentation quality should be improved. I have just noticed that quite some of my potential comments have already been stated by Reviewer # 1, so I repeat just those, which I find most urgent.*

### **General comments:**

- (1) *I found the usage units for “streamflow” quite confusing: In Fig.3 it is  $m^3/s$ , in Fig. 9 it is mm/h. Is this obtained by dividing the flow in  $[m^3/s]$  through the catchment area? Which units should be used in Equation 5 (for  $q$  and  $t$ )? There is a square root of [streamflow squared, divided by time]. The result can be hardly [mm], as given in Table 2.*

This is now addressed in Figure 4 and table 3. Further on P7L205 the following phrase has been added

“Lastly, to provide a comparison with rainfall observations, streamflow is converted to units of mm/h. This calculation involves dividing the streamflow by the catchment area”

- (2) *Please check the references throughout: The “scopus” information is hardly readable, like: “<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84884921481&doi=10.1016%2Fj.jhydrol.2013.09.004&partnerID=40&md5=3554bc7f56537b1b95a3a51160ebe40a>, 2013.” I would regards the DOI information as sufficient.*

The references have been updated to show only the URL where applicable and only the DOI for most references.

### **Specific comments:**

- (1) *Abstract: “For a selection of 7 Australian catchments this methodology was able to yield improvements of 15.3% and 7.1% ..” This is a bit misleading, since you have excluded 2 catchments to get these numbers – right?*

This was not intended to be misleading. It is now addressed on P1L6

“For a selection of 5 Australian catchments”

- (3) *Page 2, Line 35: “.. gauge density of 0.4 gauges per  $km^2$  “. Please check the numbers. According to Xu et al. (2013) this should be “per 1000  $km^2$  ”*

This is now addressed on P2L36

“until a gauge density of 0.4 gauges per 1000  $km^2$ ”

- (4) *P 3, L 68: How did you select the catchments? I understand that all are susceptible to flooding, but are these the catchments, which are most affected?*

The phrase on P3L86 now includes mention of “major flooding”. The phrase now reads;

“This study uses seven Australian catchments that are susceptible to major flooding, and having different rainfall gauge densities, climatology, area, response times, flow characteristics, and seasonality of rainfall. Further, data availability was also an important factor in choosing the catchments.”

(5) *A map with the locations of the catchments would be very helpful.*

A catchment locality map is now shown in Figure 1

(6) *You should provide some information on average precipitation in the different catchments. This would help to interpret the results in Fig. 2. Values for Tully look intimidating, but I have just found out that Tully (town) is one of the wettest places in Australia.*

Table 1 and its caption has been updated to include the requested information on average annual rainfall.

(7) P 4, L 116: *“taken to be 2 for this study” – why?*

This is now addressed on P5L137

“is an arbitrary power parameter taken to be 2 for this study. This is a commonly applied value , used operationally by the BoM, and consistent with Liu et al (2018)”

(8) P 6, L 158: *Please explain “UH” 2*

The acronym UH is described to be unit hydrograph on P6L162

(9) *Figure 1: Values for Hurdle and Warwick seem to be concentrated around ~ 75 %. Do you have an explanation for that? In general I would have expected at least some values closer to 100 %.*

Values of 75% are considered to be quite high. Due to quality control, maintenance, and equipment malfunction it is unlikely that 100% is ever reached.

(10) *Table 4: You give 3 digits after the decimal point, are those digits significant? With such an uneven distribution, the standard deviation (which is much larger than the mean) is not very meaningful.*

Table 5 now uses 1 digit after the decimal and the standard deviation has been removed.

(11) *Table 4: The GR4H values for Tully is 5525 mm – can this be true?*

It is oddly large. But yes it is true. This is now addressed on P9L246

“According to the IDW rainfall interpolation the Tully catchment does not observe enough rainfall to explain the observed streamflow. Consequently, it is likely that the OGW method compensates for this by weighting gauges that observe the highest rainfall for the Tully.”

(12) *Equation 6: Shouldn't there be a parenthesis? So you would just be summing up the r values – and the symbol r needs to be explained.*

Equation 6 has been updated. An explanation for  $r$  is given on P10L299

(13) *Figure 8: One split sample for GR4H IDW shows a totally different behavior. Do you know why?*

A specific reason is not known. A generic reason is given on P10L302

“Each split sample produced different behavior according to the data set it was calibrated against. It is expected that results would become more homogenous when longer data sets are used for optimization.”

*(13) Figure 4: One split sample for GR4H IDW shows also a different behavior. Do you know why? And is it maybe the same as in Fig. 8?*

A specific reason is not known. A generic reason is given on P10L302

“Again each split sample produced different behavior according to the data set it was calibrated against and it is expected that results would become more homogenous when longer data sets are used for optimization.”

*(14) Figure 9: The PDM model seems to saturate at ~ 2.5 mm/h. Do you have an explanation for this behavior?*

This is now addressed on P11L325

“It is hypothesized that this occurs as result of insufficient rainfall being observed to simulate the observed streamflow.”

*Further tests on the PDM indicate it can simulate larger events but does not due to poor performance in other times of the streamflow record.*

**(Some) technical corrections:**

*I did not specifically search for typos, but found quite some – so there are probably more. Some examples:*

*P 1 L 1: hydrolgical Same line: applictions*

Corrected

*P 2 L 55: The nature .... imply*

Corrected

*P 3 L 77: Köppen-Gieger (should be “-Geiger”).*

Corrected

*P 4, L 95: Queenslad*

Corrected

*P 4, L 106: operator*

Corrected

*P 4, L 127: “to potentially have” should be “do ...” right?*

No, they potentially have larger weights. It is not definite.

*P 5, L 146: “Thus the water balance is not closed and potential biases in observation data corrected for.” seems to be incomplete.*

Corrected

*P 6, L 171: “The skill ... were ..”*

Corrected

*P 10, L 281: “weigh” should be “weight” – right?*

Corrected

# Insights from a new methodology to optimize rain gauge weighting for rainfall-runoff models

Ashley J. Wright<sup>1</sup>, David E. Robertson<sup>2</sup>, Jeffrey P. Walker<sup>1</sup>, and Valentijn R.N. Pauwels<sup>1</sup>

<sup>1</sup>Department of Civil Engineering, Monash University, Clayton, Victoria, Australia

<sup>2</sup>Commonwealth Scientific and Industrial Research Organisation, Clayton, Victoria, Australia

**Correspondence:** Ashley Wright

**Abstract.** The simulation of streamflow continues to be of significant importance for societies and their economies. Hydrologic modelling of the rainfall-runoff process is essential for the simulation of streamflow. Typically rainfall runoff models use areal rainfall estimates that are based on geographical features. This paper introduces a new methodology to optimally blend the weighting of gauges for the purpose of obtaining superior streamflow simulation. For a selection of 5 Australian catchments this methodology was able to yield improvements of 15.3% and 7.1% in optimization and evaluation periods respectively. Catchments with a low gauge density, or an overwhelming majority of gauges with a low proportion of observations available, are not well suited to this new methodology. Models which close the water balance and demonstrate internal model dynamics that are consistent with a conceptual understanding of the rainfall-runoff process yielded consistent improvement in streamflow simulation skill.

## 1 Introduction

Rainfall-runoff models form the basis of hydrologic simulations for a wide range of applications, extending from forecasting floods through to water resources assessments and the design of water management structures. Hydrologic models generate streamflow simulations from forcing data including rainfall. The quality of any hydrologic simulation is dependent on the availability of high quality estimates of areal rainfall for model optimization and simulation.

For rainfall-runoff models the concept of *garbage in, garbage out* is inescapable. An improper treatment of forcing data leads to a cascade of errors which renders all other modeling techniques ineffective (Kuczera et al., 2010). Rainfall is the primary forcing for models which simulate streamflow and can be derived from in-situ gauges, satellite and ground mounted RADARs, and re-analysis products. The quality of observations, their quality control, and the methods to aggregate rainfall observations all play a key role in the effective identification of hydrologic model parameters (Te Linde et al., 2008) and the assimilation (Vanden-Eijnden et al., 2013) of observed states and outputs. Due to relative accuracy, precision, and the near immediate availability of quality controlled observations, methods that utilize gauge data to derive areal rainfall incident upon a catchment are essential for operational flood forecasts (Li et al., 2016) and are the focus of this study.

The quality of gauge based areal rainfall estimates hinges upon systematic errors, quality control of observations, the design of rain gauge networks, and methodologies to interpolate areal rainfall from the gauge network. In a review of WMO gauge

inter-comparison studies, Sevruk et al. (2009) stressed the need to correct for systematic error sources such as wind, evaporation, and shielding. Rodda and Dixon (2012) highlight that systematic errors from wind contribute most significantly towards rainfall undercatch while Pollock et al. (2018) demonstrated that a conventional cylinder shaped rain gauge mounted at 0.5 m can underestimate rainfall measured at the ground by a pit gauge by more than 23% on average. Quality control of gauged data sets is often performed via comparison with re-analysis (Robertson et al., 2015) or RADAR based products (Qi et al., 2016). Rain gauge networks are used at different temporal and spatial scales for meteorological, climatological, and hydrologic purposes. Few rain gauge networks are designed to optimize the ability of hydrologic models to simulate streamflow and flood events (Chacon-Hurtado et al., 2017). Dong et al. (2005) demonstrated that local climatic and geographic conditions play a role in the ability of each gauge to improve hydrologic simulation and that some gauges hinder the successful simulation of streamflow. For their study catchment Xu et al. (2013) found that hydrologic simulation skill improved significantly until a gauge density of 0.4 gauges per 1000 km<sup>2</sup> was reached and that simulation skill degraded for rain gauge densities larger than 1.4 gauges per km<sup>2</sup>. Through random sampling Anctil et al. (2006) explored the impact that different rainfall gauge weightings had on streamflow forecasts and determined that the best performing combination of rain gauges came from a sub-sample of the available rain gauges. These studies either indicate that systematic errors have not been sufficiently quality controlled or that the methodology to interpolate areal rainfall from the gauge network does not adequately weight important gauges. Therefore this study focuses on developing an interpolation method that adequately weights important rainfall gauges.

There are a variety of interpolation methods used to obtain an areal rainfall estimate from a gauged network ranging from simple methods such as the Thiessen's polygons and inverse distance weighting (IDW) methods, to more complex methods such as kriging. In a review of different methods for the spatial interpolation of rainfall data Ly et al. (2013) found no optimal spatial interpolation method for operational hydrology, and that interpolation performance often depended on the density of the gauge network, the spatial and temporal resolution of the data, and parameters such as those used by the semi-variogram in kriging. Mair and Fares (2011); Chen et al. (2017), and Liu et al. (2018) compared interpolation methods based on their ability to simulate streamflow, with each study using a different hydrologic model and catchment. In respective order they found that ordinary kriging, principal component regression with residual correction, and IDW interpolation of gauged rainfall data yielded superior streamflow simulations. However, none of the studies included all three methodologies. The use of a single but different catchment and single but different hydrologic model for each study makes it difficult to determine the robustness of each methodology. It is hypothesized that the best performing interpolation methods are not optimized for hydrologic simulations and that their success is likely to be greatly influenced by the choice of hydrologic model, chosen catchment, and simulation metric.

The search for an optimal interpolation of rainfall gauges for hydrologic simulation is essentially an inverse modeling question that asks, how do hydrologists extract optimal rainfall forcings from streamflow simulations? Methods to extract optimal rainfall from streamflow range from those that do not need rainfall forcing such as a cascading linear inversion of runoff (Hino, 1986) to those that require rainfall forcing such as linear inversion (Kirchner, 2009), optimization of storm multipliers (Kavetski et al., 2006; Vrugt et al., 2008), reduction of forcing data (Wright et al., 2017), and the inclusion of ancillary data (Wright et al., 2018). Further Schaeffli et al. (2007) and Duethmann et al. (2013) use precipitation correction

factors that are suitable for mountainous or data sparse regions. The nature of these studies implies that rainfall obtained from interpolation methods is not *true* rainfall. However, no example is found in the literature which presents a methodology to optimally interpolate rainfall gauge data for hydrologic simulations and provide insights derived therein. To develop an algorithm for the optimization of gauge weights (OGW) that allows the estimation of rainfall which optimizes hydrologic simulation skill it is necessary to include additional degrees of freedom in the parameter optimization process. Comparisons of streamflow forecast skill when forced with varying quantitative precipitation estimates (Heistermann and Kneis, 2011), and rain gauge, radar, and satellite based precipitation estimates (Yilmaz et al., 2005), demonstrate that calibration may obscure error and bias in rainfall estimates, making it essential that rainfall estimates and their resulting calibrated parameters be thoroughly evaluated outside of the optimization period. Many hydrologic purists are strong advocates of *Occam's Razor* and are, rightfully so, concerned that adding hydrologic parameters will result in over-fitting of hydrologic models. Keeping this in mind the value of additional parameters should be determined through the ability of those parameters to provide improved solutions for hydrologic simulations in either an evaluation or forecast period.

For the development of a robust methodology to optimally interpolate rainfall gauges it is essential to consider a variety of; catchments, rainfall-runoff models, rainfall interpolation schemes, simulation metrics, and quality of data. This study develops a methodology to optimally interpolate rainfall gauges for streamflow simulation using a variety of catchments, rainfall runoff models and quality of data. A review of different methods for the spatial interpolation of rainfall data by Ly et al. (2013) found that no spatial interpolation methodology consistently provided the best streamflow simulations. Consequently streamflow simulations forced by OGW rainfall are benchmarked against those obtained with IDW rainfall and are deemed suitable if there is a consistent improvement noticed. Situations in which the OGW is not appropriate are highlighted. Lastly, the OGW methodology is part of the optimization process. As such using an alternate simulation metric would still yield the OGW for that metric.

This paper contributes to the hydrologic sciences by developing a novel OGW method which is used in the estimation of areal rainfall for hydrologic simulations and providing insights on the potential influences that catchment characteristics and the choice of hydrologic model pose for areal rainfall estimates based on gauge interpolation.

## 2 Data set

With locations shown in Figure 1 this study uses seven Australian catchments that are susceptible to major flooding, and having different rainfall gauge densities, climatology, area, response times, flow characteristics, and seasonality of rainfall. Further, data availability was also an important factor in choosing the catchments. Data are collected for the 7 year period beginning 1<sup>st</sup> of January 2007 and ending 31<sup>st</sup> of December 2013. With the exception being that rainfall in the Onkaparinga and South Esk catchments was first recorded on 1<sup>st</sup> of November 2007. Rainfall and potential evapotranspiration (PET) data are used to force the models. Monthly PET data from the Australian Water Availability Project (AWAP) (Raupach et al., 2009, 2012) were used at and disaggregated to a uniform hourly rate for each month. This disaggregation strategy is appropriate since streamflow simulations are relatively insensitive to the temporal resolution of PET (Samain and Pauwels, 2013). Both hourly rainfall and

streamflow data (Bureau of Meteorology, 2019b) were obtained from the Bureau of Meteorology (BoM) with the catchments  
95 being delineated using the Australian geofabric data set (Bureau of Meteorology, 2019a). All rainfall gauges within 10 km of  
the catchment boundary were considered. A summary of the catchment properties is given in Table 1, whilst an indication of  
the available observations for each rain gauge in each catchment is given in Figure 2. When no gauge within the catchment had  
an observation, rainfall is assumed to be zero. Ephemeral streams were defined as those having zero flows in > 4% of their  
records (Bennett et al., 2016), and Köppen Geiger climate classifications were taken from Peel et al. (2007). Brief summaries  
100 of each catchments main characteristics are as follows:

- *Hurdle Creek at Bobinawarrah* has a small catchment that drains to an ephemeral tributary stream of the frequently  
flooded Ovens River in north-eastern Victoria. It has a temperate climate and a near uniform seasonal rainfall distribution.
- *Onkaparinga River upstream of Hahndorf dissipator* lies near the head of the flood prone Onkaparinga River catchment.  
The ephemeral stream is located within a temperate climate in south-eastern South Australia and has winter rains with  
105 drier summers.
- *South Esk River at Llewellyn* is part of a river system with a long history of flooding. The intermittent stream is located  
in north-eastern Tasmania and belongs to a temperate climate that has a near uniform seasonal rainfall distribution.
- *Clarence River at Paddys Flat* is an intermittent stream at the headwaters of a flashy catchment that observed its worst  
flood in January of 2013. The catchment lies in the eastern part of New South Wales just south of the Queensland border  
110 in a temperate climate that has winter rains and drier summers.
- *Tully River at Euramo* is a perennial stream that has been consistently subject to moderate floods over the last two  
decades. It lies within a tropical monsoon climate in far north Queensland that has the bulk of its rainfall over summer.
- *Condamine River at Warwick* is an ephemeral stream that contributes to the slow moving Condamine River. The Con-  
damine River has seen some of the countries worst flooding, having had 3 majors floods in the last decade alone. The  
115 catchment lies north of the New South Wales border in the eastern part of Queensland in a temperate climate that has  
winter rains and drier summers.
- *Isaac River at Yatton* is an ephemeral stream that lies within a hot semi-arid climate in north Queensland that is subject  
to large summer rainfall events. The Isaac River is a tributary of the Mackenzie and Fitzroy Rivers and consequently  
plays a vital role in flood warning. The catchment of Yatton has been subject to 3 major floods in the last decade.

### 120 3 Methods

In this study, the effects of two different methods of estimating areal rainfall on streamflow simulation performance are com-  
pared using three different hydrologic models. Here the methods of estimating areal rainfall are introduced, followed by a  
description of the hydrologic modelling approach and techniques for evaluating streamflow simulation performance.



### 3.1 Areal rainfall estimation

125 In essence the estimation of areal rainfall  $r_i$ , at time step  $i$ , from rainfall gauges involves adequately estimating the weighting of gauges and is determined by

$$r_i = \sum_{j=1}^n (r_{i,j} \cdot w_{i,j}), \quad (1)$$

where  $r_{i,j}$  and  $w_{i,j}$  are the rainfall volume and gauge weighting for each timestep  $i$  and gauge  $j$  respectively. The  $\times$  operator indicates element by element multiplication. For each timestep where a gauge had an observation, the gauge weights  $w_{i,j}$  are automatically scaled to unity by

$$w_{i,j} = f_{i,j} \cdot w_j / (f_{i,j} \times w_j), \quad (2)$$

where  $w_j$  is the gauge weight for each location and  $f_{i,j}$  is the binary filter array which describes if a gauge had an observation at each time step.

#### 3.1.1 Inverse distance weighting

135 In contrast to the OGW method, both gauge weights and areal rainfall estimation are calculated prior to the estimation of model parameters. Gauge weights were calculated by

$$w_j = 1/d_j^p, \quad (3)$$

where  $d_j$  is the distance from gauge  $j$  to the catchment centroid and  $p$  is an arbitrary power parameter taken to be 2 for this study. This is a commonly applied value, used operationally by the BoM, and consistent with Liu et al. (2018).

#### 140 3.1.2 Optimization of gauge weights

Methods such as IDW do not consider the hydrologic model performance in estimating areal rainfall. To consider the effects of areal rainfall estimation on streamflow simulations an approach to optimize gauge weights as a part of the hydrologic model optimization process is introduced. To allow for the identification of optimal gauge weights via a parameter estimation algorithm, maximum and minimum parameter bounds need to be identified. During the optimization process, if weighting for a gauge were able to become zero then data for that gauge is effectively discarded. If this were able to occur it is possible that the chosen likelihood or objective function becomes maximized by discarding multiple gauges and assigning an unrealistic weight to one gauge. In this situation the hydrologic model parameters could be over-fitted. Over-fitting is defined to occur when the addition of parameters leads to improved performance in the optimization periods and decreased performance in the evaluation periods. To avoid this gauge weights,  $w_j^{\min}$ , were assigned a minimum un-scaled value of one. It is then logical that gauges which have observations available for a larger proportion of time to potentially have larger gauge weightings. To ensure that the gauges with less observations contribute less to areal rainfall estimates, this maximum un-scaled gauge weighting reduced

according to an appropriate negative exponential. The maximum allowable un-scaled gauge weightings  $w_j^{\max}$  that could be explored were calculated as

$$w_j^{\max} = \max(1.01, (n - 1) \times 0.55 / 0.45 \times (1 - 2.03 \times e^{-4.74 \times p_j})), \quad (4)$$

155 where  $n$  is the total number of gauges used and  $p_j$  is the proportion of observations available for gauge  $j$ . The constants in equation 4 can be modified to allow different parameter spaces to be searched for each gauge, that has a rainfall time series which is more or less complete than the surrounding gauges. Further,  $w_j^{\max}$  were restricted to values that ensured a maximum scaled weighting of 0.55 if that gauge had all observations available. Similar to the IDW method, any alterations to forcing data via re-analysis products or inclusion of additional gauges will accordingly require the optimization process to be revised.

## 160 **3.2 Hydrologic models**

The hydrologic models used in this study were chosen based on their proven performance, contrasting routing schemes, and complexity. A description of the parameters and their ranges can be found in Table 2. Both the modèle de Génie Rural à 4 paramètres Horaire (GR4H) and Hydroliska Byråns Vattenbalansavdelning (HBV) model use unit hydrograph (UH) routing whilst the Probability Distributed Model (PDM) uses a dynamic storage based routing approach. Both HBV and PDM close  
165 the water balance whilst GR4H does not (Perrin et al., 2003). Each catchment is modelled in a lumped fashion.

### **3.2.1 GR4H**

The GR4H is an hourly application of the modèle de Génie Rural à 4 paramètres Journalier (GR4J) model (Perrin et al., 2003) which utilizes two state variables, soil water storage and routing water storage. The model is mathematically parsimonious and consists of 4 parameters which, along with the state variables, describe evapotranspiration, percolation, and both slow and fast  
170 runoff. A potential problem with the GR4H model is that the catchment exchange parameter X2 allows for both the import or export of water into the model. Thus the water balance is not closed and potential biases in observation data are not able to be corrected.

### **3.2.2 PDM**

The PDM as described by Moore (2007) consists of a set of functions used to describe various hydrologic systems. The model  
175 consists of four states, a cascade of two linear stores which are used to describe surface runoff, a linear store used to describe subsurface flow and a catchment based soil moisture store that is considered to consist of soil moisture stores with varying capacities that are able to be represented by a Pareto distribution. The PDM uses 9 model parameters and routing is a dynamic storage based process.

### **3.2.3 HBV**

180 Of the many HBV model (Lindström et al., 1997) variants, this paper uses the version developed by Matgen et al. (2006) and used by Pauwels and De Lannoy (2015). The model consists of three state variables, a soil reservoir, a slow reservoir,

and a fast reservoir. The 11 parameters, 3 state variables, and subsequent governing equations are used to calculate the actual evapotranspiration, infiltration, effective rainfall, percolation, and proportion of effective rainfall that enters the fast and slow reservoirs, outflow from each reservoir, and UH based routing.

### 185 3.3 Parameter estimation

The estimation of model and OGW parameters was performed simultaneously using the Differential Evolution Adaptive Metropolis algorithm with past state sampling and snooker updating (DREAM<sub>ZS</sub>) of Vrugt (2016). The default settings outlined in Vrugt (2016) were used. DREAM<sub>ZS</sub> finds the posterior parameter distribution which maximizes a chosen likelihood function.

#### 190 3.3.1 Likelihood function

The optimal weighting of rainfall gauges will produce rainfall specific to the chosen likelihood function and will change if a likelihood function which places priority on low flows is chosen instead of a likelihood function that places priority on high flows. The same can be said if objective functions are maximized instead of likelihood functions. For this study a Gaussian likelihood function (Thiemann et al., 2001) is used.

### 195 3.4 Evaluation strategy

The evaluation of optimized parameters and gauge weightings was carried out via split sample testing. A warm-up period of one year beginning 1<sup>st</sup> of January 2007 and ending 31<sup>st</sup> of December 2008 was used. In the Onkaparinga and South Esk catchments a warm-up period of two months was used. The skill of streamflow simulations was then evaluated for the six year period beginning 1<sup>st</sup> of January 2008 and ending 31<sup>st</sup> of December 2013. The optimization and evaluation periods were 5 and  
200 1 years long respectively. The optimization and evaluation periods were alternated until all 6 years of evaluation data had been used for evaluation. This resulted in a total of 6 split samples for each catchment.

The Root Mean Square Error (RMSE) was chosen to evaluate model simulations for widespread applicability to explain simulation trends such as bias, standard deviation, and correlation. RMSE is given by

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^t (q_i^s - q_i^o)^2}{t}}, \quad (5)$$

205 for which  $t$  represents the length of the optimized simulation period, and  $q_i^s$ , and  $q_i^o$ , represent the simulated, and observed streamflow at time step  $i$  respectively. Relative improvements in simulation skill are determined by calculating the ratio of RMSE between the stated cases. Lastly, to provide a comparison with rainfall observations, streamflow is converted to units of mm/h. This calculation involves dividing the streamflow by the catchment area.

## 4 Results and discussion

210 The results of the split sample testing are presented as an RMSE average of the samples for each catchment and model for the optimization and evaluation periods in Tables 3 and 4 respectively. The HBV model consistently yields lower RMSE values for the optimization period using both the IDW and OGW rainfall products. As expected, the addition of extra parameters for the optimized rainfall product consistently led to notable improvements throughout the optimization period. There are however a few split samples in which the additional parameters used in the OGW methodology led to a marginal increase in RMSE.

215 This can happen when the scaling of gauge weightings, as described in 3.1.2, does not allow the optimization process to attain the same gauge weights as the IDW method. A possible scenario in which this could happen occurs when the IDW method inappropriately applies a weighting larger than 55% to a single gauge.

Whilst the OGW rainfall estimates generally led to improvements in the evaluation period, there was much less consistency in performance between models. For IDW rainfall the GR4H model had the lowest RMSE for 2 out of the 7 catchments. The

220 best simulations were obtained for 5 out of the 7 catchments when OGW rainfall was used. On average the observed relative improvements in the evaluation period were  $-7.9\%$ ,  $6.2\%$ , and  $8.3\%$  for the GR4H, HBV, and PDM models, respectively. If the best IDW simulation for all models is compared to the best OGW for all models simulation an average improvement of  $4.3\%$  was observed.

To uncover the reasons for the apparent mismatch in performance between the optimization and evaluation periods it is

225 necessary to reflect on the methodology and available data. The two catchments in which the IDW simulations performed better than the OGW simulations both have deficiencies in data availability. Table 1 and Figure 2 show that the Onkaparinga catchment has 21 gauges, only one of which has data available for more than 55% of the time. Since they are more able to contribute to an increase in simulation skill in the optimization period, gauges with a larger proportion of observations available in the optimization period are more likely to obtain higher weights. If these gauges have an overall low proportion

230 of observations available they are more likely to not have observations available in the evaluation period. This is then likely to lead to an increase in simulation skill for the optimization period but not for the evaluation period. This improvement in skill in the optimization period and decrease in skill in the evaluation period was observed in the Onkaparinga catchment, a catchment with the overwhelming majority of gauges having  $< 50\%$  data availability. Consequently, this technique is not recommended for catchments that have a overwhelming majority of rainfall gauges with poor ( $< 50\%$ ) data availability. Aside from the

235 Onkaparinga catchment the catchment of Yatton is the only other catchment in which the IDW rainfall simulations perform better than the OGW simulations in evaluation periods. The Yatton catchment has one gauge per  $2806 \text{ km}^2$  whereas every other catchment has at least one gauge per  $200 \text{ km}^2$ . Consequently, a minimum gauge density of at least one gauge per  $200 \text{ km}^2$  is recommend. Zeng et al. (2018) demonstrated that identification of model parameters became increasingly difficult as the density of the rain gauge network decreased. This effect is made worse by the OGWs. Consequently, a poor gauge density

240 does not allow for adequate estimation of hydrologic model parameters, gauge weightings, and areal rainfall. If the results of these two catchments are not considered, comparing the best OGW simulation for all models to the best IDW simulation for all models led to improvements of  $15.3\%$  and  $7.1\%$  in the optimization and evaluation periods, respectively.

Applying all three models and selecting the best performing model resulted in the OGW leading to considerable improvements. It is however concerning to see that the simulation skill in the evaluation periods consistently decreased for the GR4H model. If some models are not well suited to be used with OGW rainfall estimates then it is imperative to know which elements of the model structure contribute to this decrease in skill. The average annual change in rainfall and net forcing, where net forcing are defined as rainfall minus evapotranspiration, between IDW and OGW simulations, are shown in Figure 3 and Table 5 respectively. According to the IDW rainfall interpolation the Tully catchment does not observe enough rainfall to explain the observed streamflow. Consequently, it is likely that the OGW method compensates for this by weighting gauges that observe the highest rainfall for the Tully. OGW rainfall is larger than IDW rainfall for all hydrologic models and catchments, except Yatton. It was also common for both HBV and PDM to obtain larger OGW rainfall estimates than GR4H. This trend is in stark contrast to the pattern in the change of net forcing. The mean change in net forcing across catchments was significantly lower for HBV and PDM than it was for GR4H, indicating that the internal dynamics of the HBV and PDM adapt to a change in rainfall by altering evapotranspiration while GR4H does not.

It is common for hydrologists to link the improved simulation performance and increase in the number of parameters with over-fitting. The consistent improvement in simulation skill in the evaluation periods for the PDM and HBV model demonstrates that over-fitting does not occur for these models. However, the consistent decrease in simulation skill in the evaluation periods for the GR4H model demonstrates that over-fitting did occur for this model. By testing different gauge weightings the OGW methodology takes greater advantage of the available rainfall time series when compared to the IDW methodology. As such it is hypothesized that the GR4H model is not able to take advantage of the additional data and is subject to over-fitting.

#### 4.1 Case Study - Paddys Flat

This case study was designed to develop a deeper understanding of the impact that the OGW and subsequent areal rainfall estimates had on the hydrologic model simulations and internal dynamics. Paddys Flat was chosen for the case study due to the improvement in streamflow simulation skill in the evaluation period being similar to the mean improvement of streamflow simulation skill in the evaluation period across all 7 catchments. Further, results observed in the Paddys Flat case study are considered to be representative of the OGW methodology. Results and their associated discussion are presented sequentially to describe the impacts on rainfall, model parameters, the water balance, and streamflow. The observed streamflow and IDW areal rainfall estimate for Paddys Flat can be seen in Figure 4.

The comparison of the cumulative rainfall volumes obtained for the split sample simulations for each model at Paddys Flat is shown in Figure 5 demonstrating that the OGW tends to lead to an increase in cumulative rainfall volumes. Each split sample obtained through model evaluation is represented by a different line. The cumulative rainfall volumes estimated by GR4H are less than those estimated by HBV and PDM. Each split sample produced different behavior according to the data set it was calibrated against. It is expected that results would become more homogenous when longer data sets are used for optimization. However, it is necessary to determine if there is a tendency for low or high magnitude rainfall observations to be estimated differently by the OGW or IDW interpolation methods.

Plots of the OGW estimation of areal rainfall obtained by each model versus the IDW estimation of areal rainfall in Figure 6 reveal that, when compared to the IDW interpolation method, the OGW methodology did not have a tendency to predict greater rainfall volumes for small or large magnitude rainfall observations. The larger cumulative rainfall volumes observed by the OGW method were therefore a result of a slight tendency to estimate greater rainfall volumes for both small and large magnitude rainfall observations. The impact each gauge has on the estimation of areal rainfall was then explored.

Hydrologic simulation skill based OGWs allows gauges which add value to evaluation periods or forecasts to be identified. The gauge weightings determined by the IDW interpolation method and OGW method for each split sample and model are shown in Figure 7. Each model selected similar gauges and weightings for the majority of gauges. There was however one or two gauges which only one model assigned weight to. Therefore the results from the analysis of rainfall estimates obtained through the OGW method do not indicate that GR4H, HBV, and PDM require model specific rainfall forcings, and nor do they give sufficient reasoning as to why the OGW simulations decrease in simulation skill for GR4H and improve in simulation skill for HBV and PDM. It is therefore likely that some difference in model parameters and subsequent impact on internal model dynamics caused the increase/decrease in skill.

To demonstrate the change or lack of change in model parameters that were observed for each rainfall estimation method, the Maximum a Posteriori (MAP) parameters obtained for each split sample simulation are shown for the OGW and IDW simulations for Paddys Flat in Figure 8. Despite the increase in the amount of parameters needed for OGW streamflow simulations, the relatively small change in parameter spread observed for each model and rainfall interpolation method suggests that there is little increase in the hydrologic model parameter uncertainty. GR4H parameters did not noticeably alter whilst parameters that influence soil moisture content, evapotranspiration, and percolation change significantly for both HBV and PDM. This lack of change in parameters for GR4H indicates that the model is not taking advantage of the additional data. Further, parameters influencing fast flow and baseflow change for the HBV model. The change in these specific parameters indicate that internal model dynamics are sensitive to changes in forcing data for both HBV and PDM but not for GR4H. This presents a basis as to why rainfall minus evapotranspiration change much less for both HBV and PDM than GR4H.

The water balance is given as

$$s_i = \sum_{m=1}^i (r_m - q_m - e_m), \quad (6)$$

where  $s_i$ ,  $r_m$ , and  $e_m$  are the catchment storage, rainfall forcing, and evapotranspiration at time steps  $i$  and  $m$  respectively. It should be noted that for GR4H a catchment exchange process that either abstracts or provides additional water may occur. Figure 9 shows the water balance for the split sample simulations for the OGW and IDW methods for each model at Paddys Flat. Again each split sample produced different behavior according to the data set it was calibrated against and it is expected that results would become more homogenous when longer data sets are used for optimization. After the warm-up period of one year, simulations for both the HBV and PDM models reached a point of equilibrium in which a rise or decline in  $s_i$  was eventually countered by a subsequent and opposing rise or decline in  $s_i$ . This trend is the cornerstone of a model that is able to close the water balance. A non-zero positive or negative catchment exchange parameter in GR4H led to simulations which do

not close the water balance. This is the reason why  $s_i$  gradually increased for GR4H. Referring back to Figure 8 it can be seen  
310 that the catchment exchange parameter was marginally above the mean of the minimum and parameter bounds. This mean is  
zero and explains why there is only a gradual departure from zero.

The IDW storage profiles and the OGW storage profiles obtained using GR4H remained inconsistent with the current conceptual understanding of the rainfall-runoff process. The storage profiles for GR4H have different trajectories for each split sample which became more dispersed with the inclusion of OGW rainfall estimates. In contrast, both HBV and PDM demonstrated storage profiles that tend towards an equilibrium. The storage profiles of HBV and PDM for each split sample are more  
315 similar to each other than when the IDW rainfall estimates were used. Further, the storage profiles obtained using the HBV and PDM were remarkably similar for all split samples regardless of whether or not the IDW or OGW rainfall estimates were used. As such both, HBV and PDM demonstrate improved internal model dynamics which are consistent with a conceptual understanding of the rainfall-runoff process. Lastly, the ability of each model to represent different streamflow events was  
320 analyzed.

The OGW estimation of areal rainfall enabled both HBV and PDM to simulate internal dynamics with improved consistency. Models that have more consistency in their representation of internal dynamics are better positioned to benefit from the inclusion of soil moisture data for calibration and/or assimilation purposes (Li et al., 2016). Similar benefits are likely to be observed when updating internal states through the assimilation of observed streamflow. The relative impact of the OGW methodology  
325 on streamflow, when compared to the IDW methodology, can be observed in Figure 10. Events can be distinguished by lines of hysteresis and the observed streamflow. For all three models there are a number of flood events that are not adequately simulated. It is hypothesized that this occurs as a result of insufficient rainfall being observed to simulate the observed streamflow. It is clear that for larger magnitude events the OGW methodology brought the streamflow simulations closer to observations for the HBV and PDM throughout the evaluation period. Conversely, the use of OGW rainfall areal rainfall estimates with GR4H  
330 led to marginally improved streamflow simulations for one event and significantly worse streamflow simulations for another event.

## 5 Conclusions

This study developed a methodology to optimally weight rainfall gauges such that improved streamflow simulation skill could be obtained for three different conceptual hydrologic models. The OGW methodology developed was tested on seven Australian catchments and a comparison of streamflow simulations obtained using the IDW and OGW methods demonstrated an  
335 improvement in streamflow RMSE of 15.3% and 7.1% in the optimization and evaluation periods respectively, for catchments that have multiple rainfall gauges with observations available  $> 50\%$  of the time and more than one rainfall gauge every 200  $\text{km}^2$ . The methodology did not work equally well for the three hydrologic models chosen. Improvements in evaluation periods were only noticed for the PDM and HBV hydrologic models, and not for the GR4H hydrologic model. The most likely  
340 explanation for this is the inability of the GR4H model to represent internal dynamics that are consistent with a conceptual understanding of the rainfall-runoff process. Further research could explore the impact that OGW has on parameter uncertainty

and possible interactions between OGW and the inclusion of catchment exchange or bias correction parameters. Catchment characteristics were not found to influence the applicability of the OGW methodology. This methodology opens new possibilities for model evaluation, understanding of forcing uncertainty, and data assimilation studies in hydrology.

345 *Author contributions.* AW conducted the experimental work, contributed towards the theory and wrote the paper. JW assisted in the writing process. VP and DR contributed towards the theory and assisted in the writing process.

*Acknowledgements.* The authors would like to extend their gratitude to the anonymous reviewers for their comments and recommendations. This work was supported by the Multi-modal Australian Sciences Imaging and Visualisation Environment (MASSIVE) (<http://www.massive.org.au>).



## References

- 350 Ancil, F., Lauzon, N., Andréassian, V., Oudin, L., and Perrin, C.: Improvement of rainfall-runoff forecasts through mean areal rainfall optimization, *J. Hydrol.*, 328, 717–725, <https://doi.org/10.1016/j.jhydrol.2006.01.016>, 2006.
- Bennett, J. C., Robertson, D. E., Ward, P. G., Hapuarachchi, H. P., and Wang, Q.: Calibrating hourly rainfall-runoff models with daily forcings for streamflow forecasting applications in meso-scale catchments, *Environ. Model. Softw.*, 76, 20–36, <https://doi.org/10.1016/j.envsoft.2015.11.006>, 2016.
- 355 Bureau of Meteorology: Australian Hydrological Geospatial Fabric (Geofabric), <http://www.bom.gov.au/water/geofabric/index.shtml>, 2019a.
- Bureau of Meteorology: Water Data Online, <http://www.bom.gov.au/waterdata/>, 2019b.
- Chacon-Hurtado, J. C., Alfonso, L., and Solomatine, D. P.: Rainfall and streamflow sensor network design: A review of applications, classification, and a proposed framework, *Hydrol. Earth Syst. Sci.*, 21, 3071–3091, <https://doi.org/10.5194/hess-21-3071-2017>, 2017.
- Chen, T., Ren, L., Yuan, F., Yang, X., Jiang, S., Tang, T., Liu, Y., Zhao, C., Zhang, L., Chen, T., Ren, L., Yuan, F., Yang, X., Jiang, S., Tang,
- 360 T., Liu, Y., Zhao, C., and Zhang, L.: Comparison of Spatial Interpolation Schemes for Rainfall Data and Application in Hydrological Modeling, *Water*, 9, 342, <https://doi.org/10.3390/w9050342>, 2017.
- Dong, X., Dohmen-Janssen, C. M., and Booij, M. J.: Appropriate spatial sampling of rainfall for flow simulation, *Hydrol. Sci. J.*, 50, 279–298, <https://doi.org/10.1623/hysj.50.2.279.60656>, 2005.
- Duethmann, D., Zimmer, J., Gafurov, A., Güntner, A., Kriegel, D., Merz, B., and Vorogushyn, S.: Evaluation of areal precipitation
- 365 estimates based on downscaled reanalysis and station data by hydrological modelling, *Hydrol. Earth Syst. Sci.*, 17, 2415–2434, <https://doi.org/10.5194/hess-17-2415-2013>, 2013.
- Heistermann, M. and Kneis, D.: Benchmarking quantitative precipitation estimation by conceptual rainfall-runoff modeling, *Water Resour. Res.*, 47, <https://doi.org/10.1029/2010WR009153>, 2011.
- Hino, M.: Improvements in the inverse estimation method of effective rainfall from runoff, *J. Hydrol.*, 83, 137–147,
- 370 [https://doi.org/10.1016/0022-1694\(86\)90188-5](https://doi.org/10.1016/0022-1694(86)90188-5), 1986.
- Kavetski, D., Kuczera, G., and Franks, S. W.: Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, *Water Resour. Res.*, 42, <https://doi.org/10.1029/2005WR004376>, 2006.
- Kirchner, J. W.: Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, *Water Resour. Res.*, 45, <https://doi.org/10.1029/2008WR006912>, 2009.
- 375 Kuczera, G., Renard, B., Thyer, M., and Kavetski, D.: There are no hydrological monsters, just models and observations with large uncertainties! , *Hydrol. Sci. J.*, 55, 980–991, <https://doi.org/10.1080/02626667.2010.504677>, 2010.
- Li, Y., Grimaldi, S., Walker, J., and Pauwels, V.: Application of Remote Sensing Data to Constrain Operational Rainfall-Driven Flood Forecasting: A Review, *Remote Sens.*, 8, 456, <https://doi.org/10.3390/rs8060456>, 2016.
- Lindström, G., Johansson, B., Persson, M., Gardelin, M., and Bergström, S.: Development and test of the distributed HBV-96 hydrological
- 380 model, *J. Hydrol.*, 201, 272–288, [https://doi.org/10.1016/S0022-1694\(97\)00041-3](https://doi.org/10.1016/S0022-1694(97)00041-3), 1997.
- Liu, S., Li, Y., Pauwels, V. R. N., and Walker, J. P.: Impact of Rain Gauge Quality Control and Interpolation on Streamflow Simulation: An Application to the Warwick Catchment, Australia, *Front. Earth Sci.*, 5, <https://doi.org/10.3389/feart.2017.00114>, 2018.
- Ly, S., Charles, C., and Degré, A.: Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale. A review , *Biotechnol. Agron. Soc. Environ.*, 17, 392–406, 2013.

- 385 Mair, A. and Fares, A.: Comparison of rainfall interpolation methods in a mountainous region of a tropical island, *J. Hydrol. Eng.*, 16, 371–383, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000330](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000330), 2011.
- Matgen, P., Henry, J.-B., Hoffmann, L., and Pfister, L.: Assimilation of remotely sensed soil saturation levels in conceptual rainfall-runoff models, in: *IAHS-AISH Publ.*, 303, pp. 226–234, 2006.
- Moore, R. J.: The PDM rainfall-runoff model, *Hydrol. Earth Syst. Sci.*, 11, 483–499, <https://doi.org/10.5194/hess-11-483-2007>, 2007.
- 390 Pauwels, V. R. N. and De Lannoy, G. J. M.: Error covariance calculation for forecast bias estimation in hydrologic data assimilation, *Adv. Water Resour.*, 86, 284–296, <https://doi.org/10.1016/j.advwatres.2015.05.013>, 2015.
- Peel, M. C., Finlayson, B. L., and McMahon, T. A.: Updated world map of the Köppen-Geiger climate classification, *Hydrol. Earth Syst. Sci.*, 11, 1633–1644, <https://doi.org/10.5194/hess-11-1633-2007>, 2007.
- Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, *J. Hydrol.*, 279, 275–289, [https://doi.org/10.1016/S0022-1694\(03\)00225-7](https://doi.org/10.1016/S0022-1694(03)00225-7), 2003.
- 395 Pollock, M. D., O'Donnell, G., Quinn, P., Dutton, M., Black, A., Wilkinson, M. E., Colli, M., Stagnaro, M., Lanza, L. G., Lewis, E., Kilsby, C. G., and O'Connell, P. E.: Quantifying and Mitigating Wind-Induced Undercatch in Rainfall Measurements, *Water Resour. Res.*, 54, 3863–3875, <https://doi.org/10.1029/2017WR022421>, 2018.
- Qi, Y., Martinaitis, S., Zhang, J., and Cocks, S.: A Real-Time Automated Quality Control of Hourly Rain Gauge Data Based on Multiple Sensors in MRMS System, *J. Hydrometeorol.*, 17, 1675–1691, <https://doi.org/10.1175/jhm-d-15-0188.1>, 2016.
- 400 Raupach, M., Briggs, P., Haverd, V., King, E., Paget, M., and Trudinger, C.: Australian Water Availability Project (AWAP): CSIRO Marine and Atmospheric Research Component: Final Report for Phase 3, [http://www.csiro.au/awap/doc/CTR\[\\_\]013\[\\_\]online\[\\_\]FINAL.pdf](http://www.csiro.au/awap/doc/CTR[_]013[_]online[_]FINAL.pdf), 2009.
- Raupach, M., Briggs, P., Haverd, V., King, E., Paget, M., and Trudinger, C.: Australian Water Availability Project, <http://www.csiro.au/awap>, 2012.
- 405 Robertson, D. E., Bennett, J. C., and Wang, Q. J.: A strategy for quality controlling hourly rainfall observations and its impact on hourly streamflow simulations, [www.mssanz.org.au/modsim2015](http://www.mssanz.org.au/modsim2015), 2015.
- Rodda, J. C. and Dixon, H.: Rainfall measurement revisited, 67, <https://doi.org/10.1002/wea.1884>, 2012.
- Samain, B. and Pauwels, V. R. N.: Impact of potential and (scintillometer-based) actual evapotranspiration estimates on the performance of a lumped rainfall–runoff model, *Hydrol. Earth Syst. Sci.*, 17, 4525–4540, <https://doi.org/10.5194/hess-17-4525-2013>, 2013.
- 410 Schaeffli, B., Talamba, D. B., and Musy, A.: Quantifying hydrological modeling errors through a mixture of normal distributions, *J. Hydrol.*, 332, 303–315, <https://doi.org/10.1016/j.jhydrol.2006.07.005>, 2007.
- Sevruk, B., Ondrás, M., and Chvíla, B.: The WMO precipitation measurement intercomparisons, *Atmos. Res.*, 92, 376–380, <https://doi.org/10.1016/j.atmosres.2009.01.016>, 2009.
- 415 Te Linde, A. H., Aerts, J., Hurkmans, R., and Eberle, M.: Comparing model performance of two rainfall-runoff models in the Rhine basin using different atmospheric forcing data sets, *Hydrol. Earth Syst. Sci.*, 12, 943–957, <https://doi.org/10.5194/hess-12-943-2008>, 2008.
- Thiemann, M., Trosset, M., Gupta, H., and Sorooshian, S.: Bayesian recursive parameter estimation for hydrologic models, *Water Resour. Res.*, 37, 2521–2535, <https://doi.org/10.1029/2000WR900405>, 2001.
- Vanden-Eijnden, E., Weare, J., Vanden-Eijnden, E., and Weare, J.: Data Assimilation in the Low Noise Regime with Application to the Kuroshio, *Mon. Weather Rev.*, 141, 1822–1841, <https://doi.org/10.1175/MWR-D-12-00060.1>, 2013.
- 420 Vrugt, J. A.: Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation, *Environ. Model. Softw.*, 75, 273–316, <https://doi.org/10.1016/j.envsoft.2015.08.013>, 2016.

**Table 1.** Catchment area, number of rainfall gauges, the % of time that no gauge had an observation for each of the study catchments, average annual IDW rainfall over the simulation period, and minimum and maximum elevation. It should be noted that IDW rainfall is not regarded as the truth. Statistics shown are for the collected rainfall record beginning 1<sup>st</sup> of January 2007 and ending 31<sup>st</sup> of December 2013, except for in the Onkaparinga and South Esk catchments where rainfall was first recorded on 1<sup>st</sup> of November 2007.

Catchment	Number of gauges	% of time no gauge works	Catchment area km <sup>2</sup>	Annual IDW rain-fall [mm]	Minimum elevation [m]	Maximum elevation [m]
Hurdle Creek	12	14.6	156	582.6	165.3	1062.4
Onkaparinga	21	12.5	208	485.4	179.0	616.6
Paddys Flat	27	2.6	3112	765.1	168.6	1300.6
South Esk	13	6.902	2289	616.2	177.5	1527.3
Tully	7	4.1	1390	2133.2	6.7	1198.6
Warwick	21	5.3	1379	502.1	446.2	1361.7
Yatton	7	3.6	19639.5	1015.8	99.0	1084.9

Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., and Robinson, B. A.: Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resour. Res.*, 44, <https://doi.org/10.1029/2007WR006720>, 2008.

Wright, A., Walker, J., and Pauwels, V.: Estimating rainfall time series and model parameter distributions using model data reduction and inversion techniques, *Water Resour. Res.*, 53, <https://doi.org/10.1002/2017WR020442>, 2017.

Wright, A. J., Walker, J. P., and Pauwels, V. R. N.: Identification of Hydrologic Models, Optimized Parameters, and Rainfall Inputs Consistent with In Situ Streamflow and Rainfall and Remotely Sensed Soil Moisture, *J. Hydrometeorol.*, 19, 1305–1320, <https://doi.org/10.1175/JHM-D-17-0240.1>, 2018.

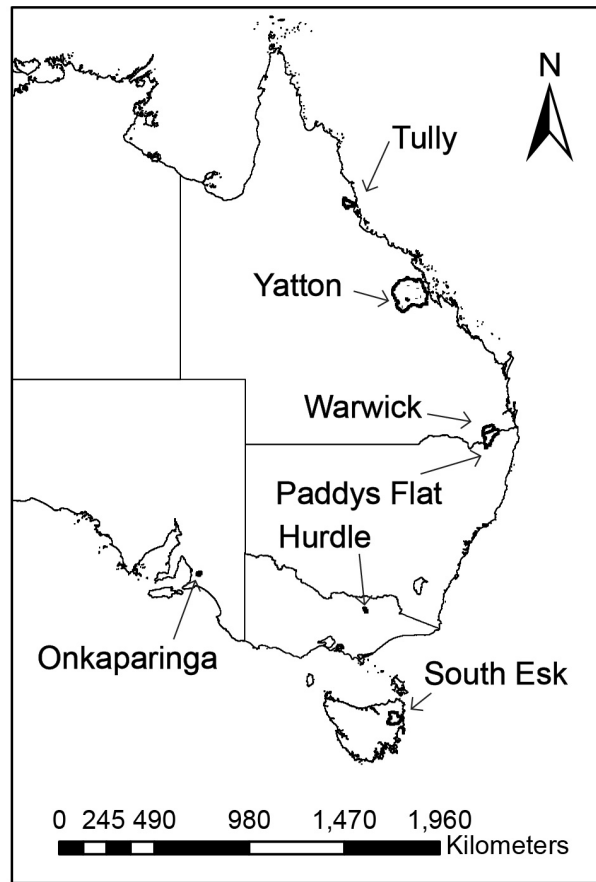
Xu, H., Xu, C.-Y., Chen, H., Zhang, Z., and Li, L.: Assessing the influence of rain gauge density and distribution on hydrological model performance in a humid region of China, *J. Hydrol.*, 505, 1–12, <https://doi.org/10.1016/j.jhydrol.2013.09.004>, 2013.

Yilmaz, K. K., Hogue, T. S., Hsu, K.-L., Sorooshian, S., Gupta, H. V., and Wagener, T.: Intercomparison of rain gauge, radar, and satellite-based precipitation estimates with emphasis on hydrologic forecasting, *J. Hydrometeorol.*, 6, 497–517, <https://doi.org/10.1175/JHM431.1>, 2005.

Zeng, Q., Chen, H., Xu, C.-Y., Jie, M.-X., Chen, J., Guo, S.-L., and Liu, J.: The effect of rain gauge density and distribution on runoff simulation using a lumped hydrological modelling approach, *J. Hydrol.*, 563, 106–122, <https://doi.org/10.1016/j.jhydrol.2018.05.058>, 2018.

**Table 2.** Parameters and ranges used for the optimization of hydrologic models .

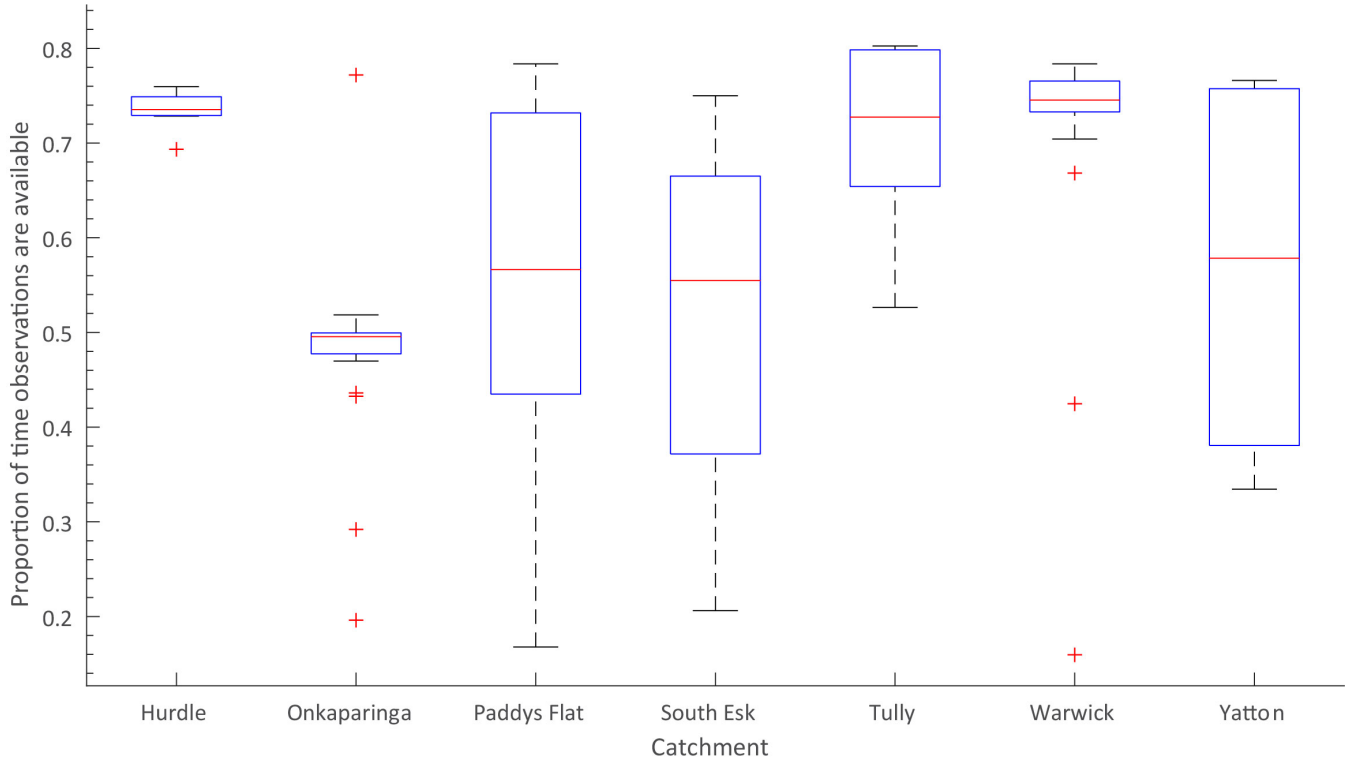
Model	Parameter	Description	Range	Unit
GR4H	X1	Production store capacity	1 - 3000	mm
	X2	Catchment exchange	-27 - 27	mm
	X3	Routing store capacity	1 - 660	mm
	X4	length of unit hydrograph	1 - 240	hr
HBV	$\lambda$	Parameter for evapotranspiration	$10^{-2}$ - 3	-
	$S_{\max}$	Capacity of soil reservoir	$1 - 10^3$	mm
	b	Parameter for infiltration	$10^{-2}$ - 3	-
	Pe	Percolation parameter	$10^{-2} - 50$	mm/h
	$\alpha$	Fast reservoir parameter	$10^{-2}$ - 3	-
	$\beta$	Percolation parameter	$10^{-2}$ - 3	-
	$S2_{\max}$	Capacity of fast reservoir	1 - 2000	mm
	$\kappa_2$	Parameter for outflow from fast reservoir	$10^{-1} - 10^2$	mm <sup>3</sup> /h
	$\psi$	Parameter for outflow from fast reservoir	$10^{-1}$ - 3	-
	$\kappa_1$	Parameter for outflow from slow reservoir	$10^{-3}$ - 1	mm <sup>2</sup> /h
	UH	length of unit hydrograph	1 - 240	h
PDM	$C_{\max}$	Maximum store capacity	1 - 500	mm
	b	Parameter that controls spatial variability of $C_{\max}$	$10^{-3}$ - 1.8	-
	$b_e$	Actual evaporation exponent	0.1 - 5	-
	$b_g$	Recharge function exponent	0.2 - 6.7	-
	$k_b$	Baseflow constant	1 - 2000	h/mm <sup>2</sup>
	$C_{\min}$	Minimum store capacity	$(0 - 1) \times C_{\max}$	mm
	$S_t$	Soil tension storage capacity	$(0 - 1) \times C_{\max}$	mm
	$\kappa_1$	Time constant for linear reservoir	1 - 300	h
	$\kappa_2/$	Time constant for linear reservoir	$(10^{-6} - 100) \times k_1$	h



**Figure 1.** The location of catchments used in this study.

**Table 3.** Average RMSE [mm/h] for simulated streamflow in the optimization periods for each catchment using IDW and OGW rainfall. Values in bold indicate the lowest obtained RMSE.

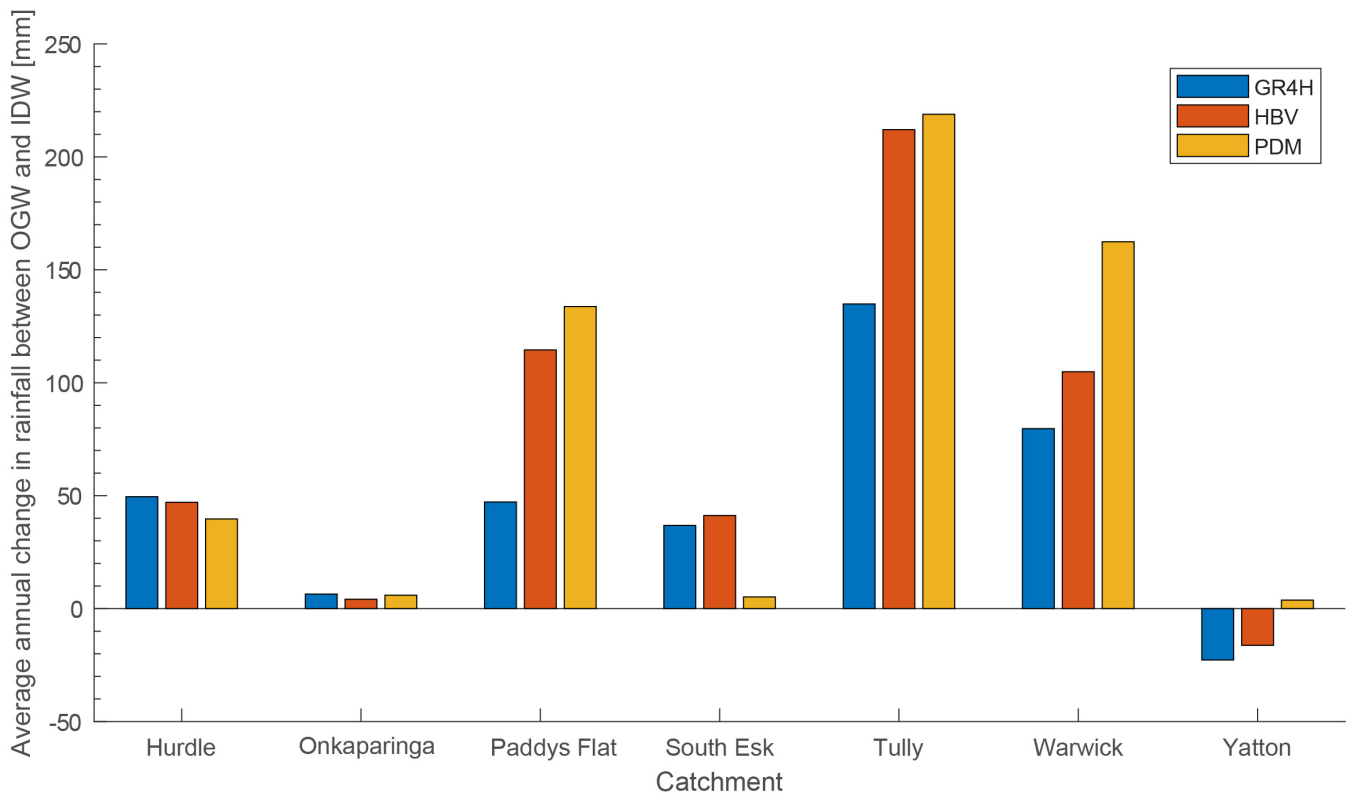
Catchment	GR4H		HBV		PDM	
	IDW	OGW	IDW	OGW	IDW	OGW
Hurdle Creek	0.033	0.031	0.032	<b>0.028</b>	0.033	0.028
Onkaparinga	0.040	0.039	0.037	<b>0.037</b>	0.039	0.039
Paddys Flat	0.068	0.053	0.061	<b>0.048</b>	0.084	0.069
South Esk	0.037	0.033	0.035	<b>0.032</b>	0.043	0.039
Tully	0.175	0.177	0.165	<b>0.151</b>	0.170	0.157
Warwick	0.032	0.023	0.032	<b>0.023</b>	0.047	0.038
Yatton	0.062	0.056	0.053	<b>0.048</b>	0.061	0.057



**Figure 2.** Box plots showing the median, 25<sup>th</sup> and 75<sup>th</sup> percentile, end points and outliers that describe the proportion of observations available for gauges within each catchment.

**Table 4.** Average RMSE [mm/h] for simulated streamflow in evaluation periods for each catchment using IDW and OGW rainfall. Values in bold indicate the lowest obtained RMSE.

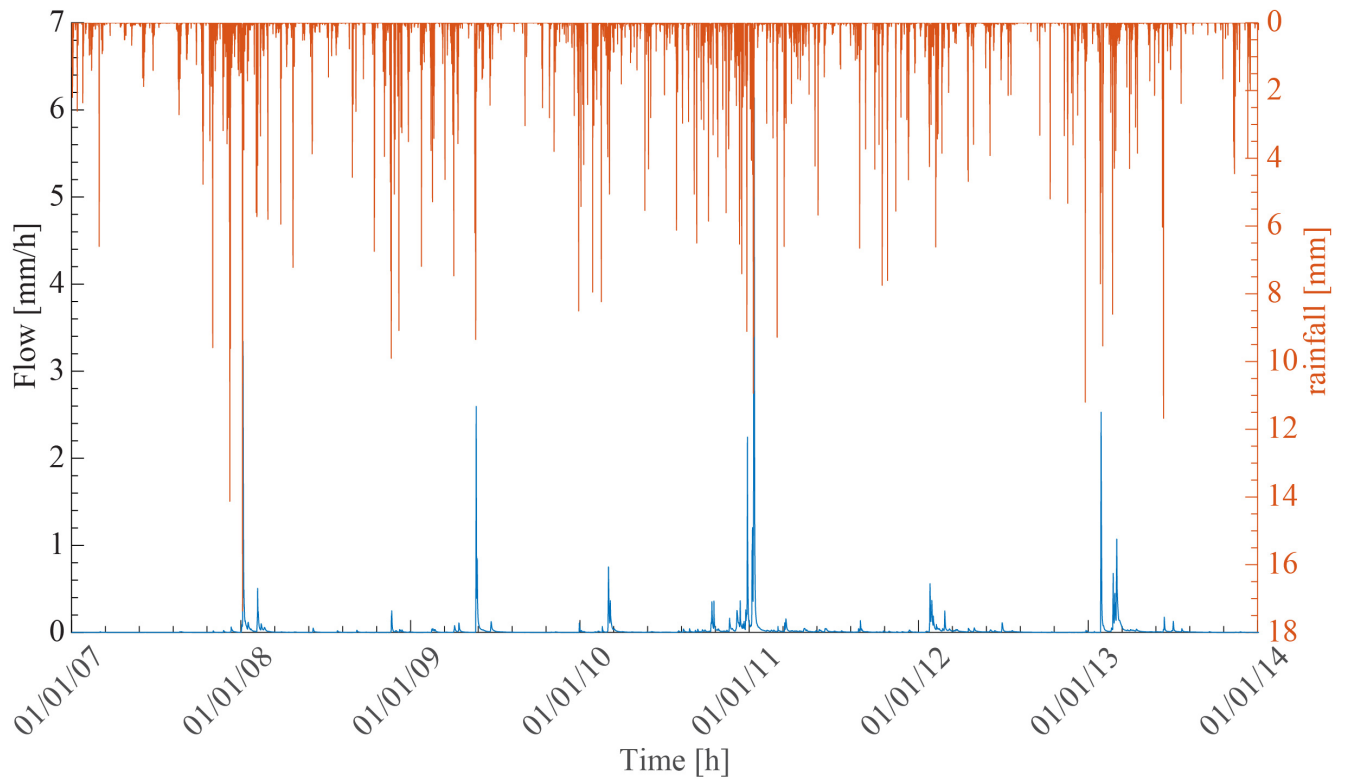
Catchment	GR4H		HBV		PDM	
	IDW	OGW	IDW	OGW	IDW	OGW
Hurdle Creek	0.033	0.033	0.041	0.036	0.037	<b>0.029</b>
Onkaparinga	<b>0.035</b>	0.036	0.038	0.039	0.038	0.039
Paddys Flat	0.083	0.103	0.082	<b>0.077</b>	0.093	0.080
South Esk	0.037	<b>0.037</b>	0.044	0.043	0.053	0.044
Tully	0.176	0.184	0.161	<b>0.151</b>	0.167	0.156
Warwick	0.035	0.038	0.042	<b>0.031</b>	0.050	0.041
Yatton	<b>0.068</b>	0.075	0.069	0.072	0.068	0.078



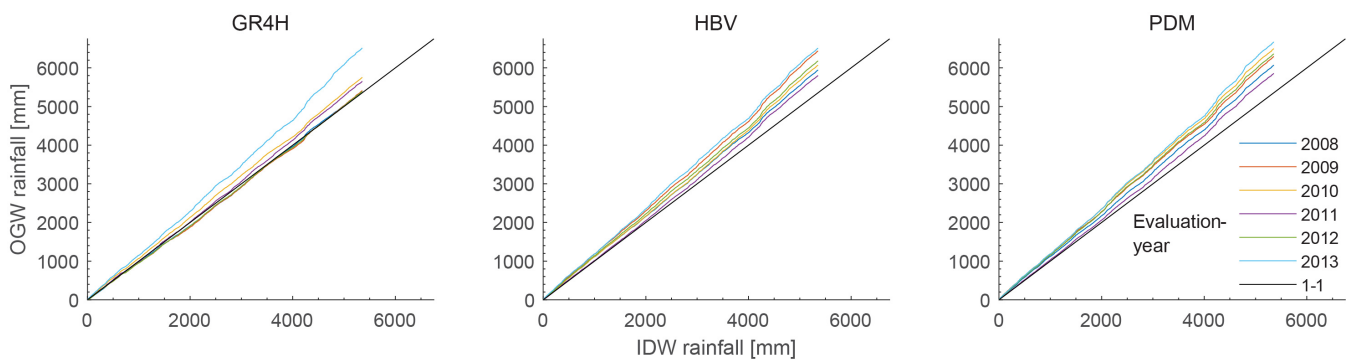
**Figure 3.** The average annual change in estimated areal rainfall for the simulation period between the OGW and IDW areal rainfall estimation methods.

**Table 5.** Average annual change [mm] in rainfall minus evapotranspiration between the IDW and OGW rainfall simulations. A positive change indicates that net model forcings were higher for the OGW simulations.

	GR4H	HBV	PDM
Hurdle Creek	40.0	22.9	19.8
Onkaparinga	5.1	0.57	6.3
Paddys Flat	-13.0	3.5	-33.3
South Esk	54.5	-5.3	16.6
Tully	789.3	201.8	189.2
Warwick	9.4	2.6	14.0
Yatton	115.5	-11.6	22.5
Mean	143.0	60.1	33.6

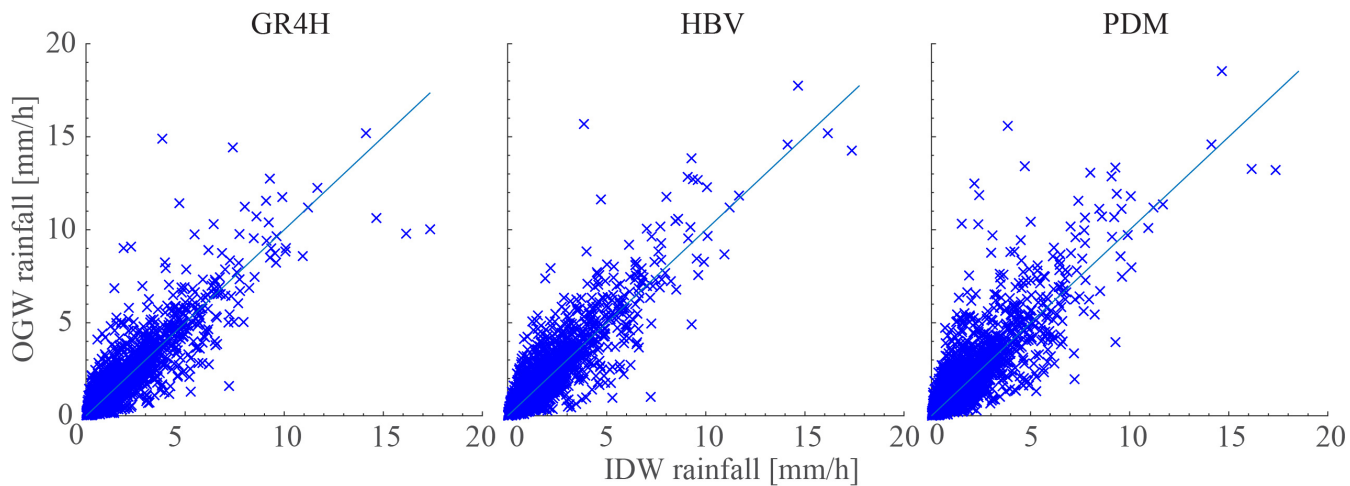


**Figure 4.** Observed streamflow and IDW areal rainfall estimates for Paddys Flat.

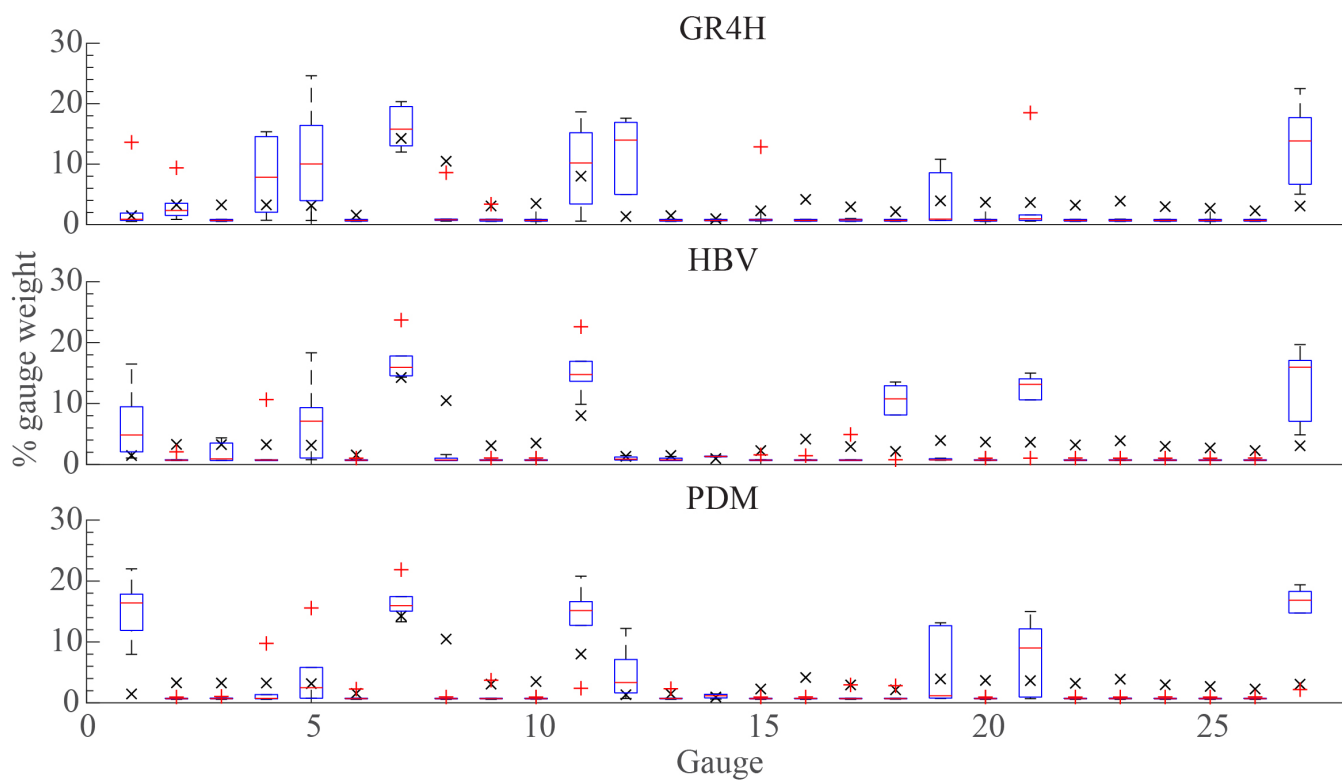


**Figure 5.** Cumulative areal rainfall estimates obtained for each OGW simulation when compared to the cumulative IDW rainfall areal rainfall estimate for Paddy's Flat. 1-1 reference lines are shown in black. Each colour represents a different split sample.

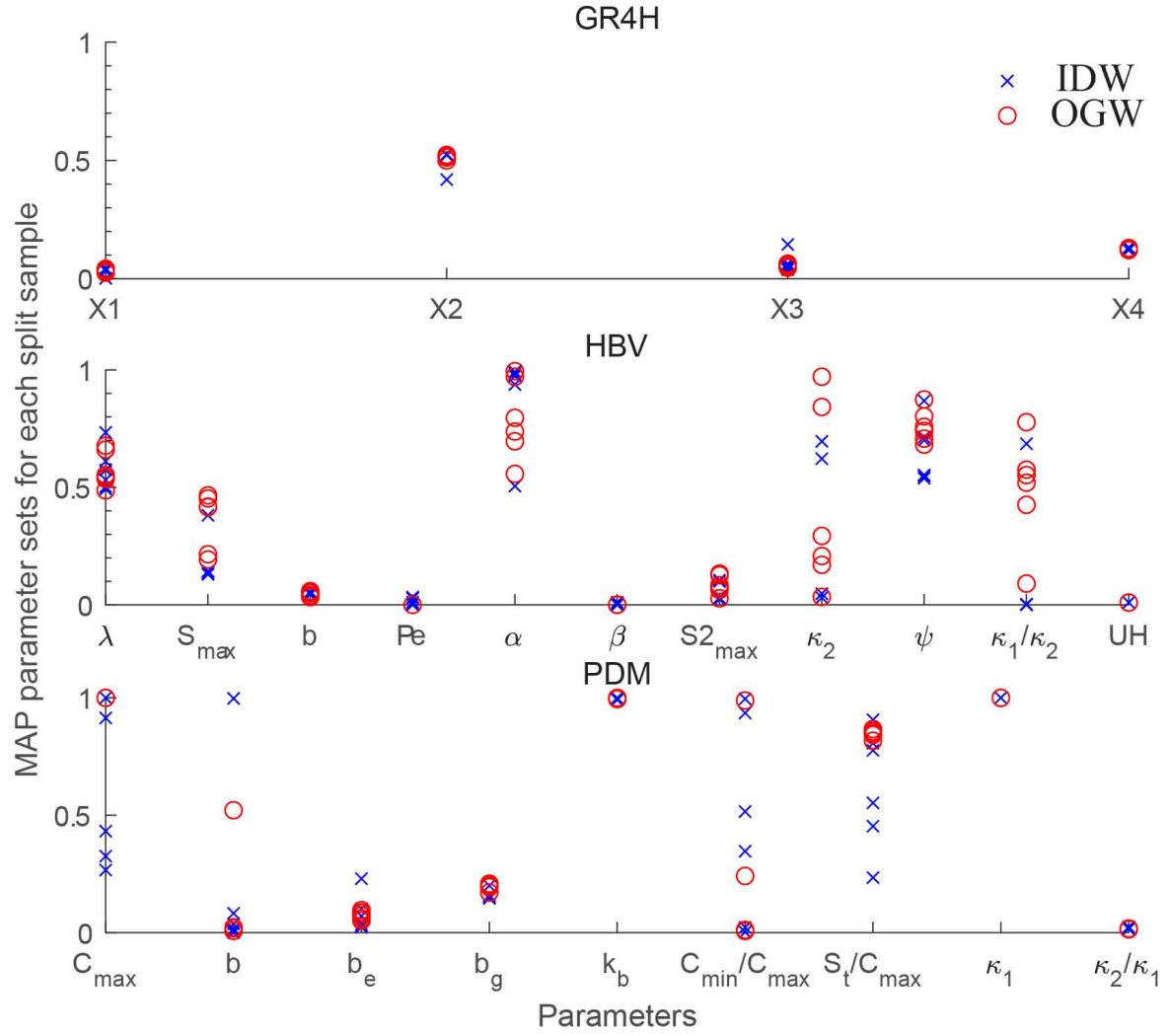




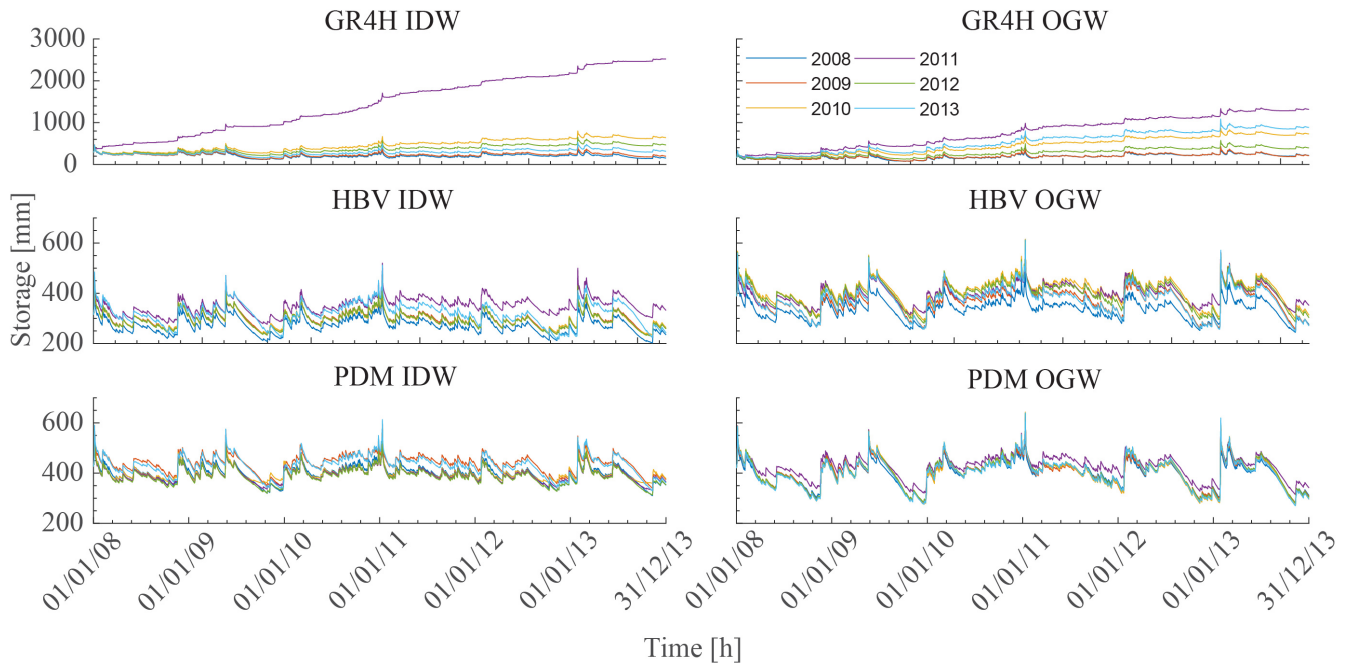
**Figure 6.** A 1-1 comparison of the average OGW areal rainfall estimates with the IDW areal rainfall estimates for each model for Paddy's Flat.



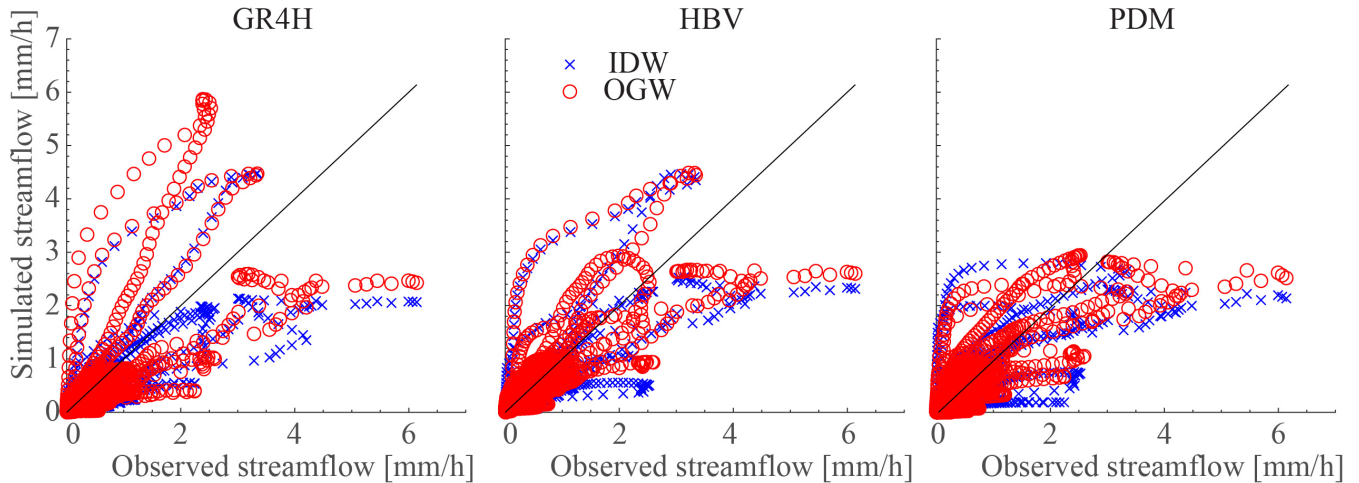
**Figure 7.** A comparison of the IDW gauge weights, as indicated by a  $\times$ , with OGW gauge weights for each split sample simulation for Paddy's Flat. A  $+$  indicates a potential outlier.



**Figure 8.** Relative optimized parameters for each rainfall-runoff model using OGW and IDW rainfall estimates for Paddy's Flat. Values of 0 and 1 indicate the lower and upper bounds for each parameter respectively.



**Figure 9.** A representation of each models ability to resolve storage for each split sample and areal rainfall estimation method respectively for Paddy's Flat. Each time series starts at the end of the warm-up period and is adjusted to have the same storage as other samples from the model and rainfall estimation method pair. Each colour represents a different split sample.



**Figure 10.** A comparison of the streamflow simulations for the evaluation periods for each model and areal rainfall estimation method for Paddy's Flat. Events can be distinguished by lines of hysteresis and the observed streamflow. 1-1 reference lines are shown in black.