


RESEARCH ARTICLE

# Application of the patient rule induction method to detect hydrologic model behavioural parameters and quantify uncertainty

Ashkan Shokri<sup>1</sup>  | Jeffrey P. Walker<sup>1</sup> | Albert I. J. M. van Dijk<sup>2</sup> | Ashley J. Wright<sup>1</sup> | Valentijn R. N. Pauwels<sup>1</sup>

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

<sup>2</sup>Fenner School of Environment and Society, Australian National University, Canberra, Australia

## Correspondence

Ashkan Shokri, Department of Civil Engineering, Monash University, Melbourne, Australia.

Email: ashkan.shokri@monash.edu

## Funding information

Monash International Postgraduate Research Scholarship (MIPRS); Australian Research Council (ARC) Future Fellowship, Grant/Award Number: FT130100545; Australian Research Council (ARC), Grant/Award Number: DP140103679

## Abstract

Finding an operational parameter vector is always challenging in the application of hydrologic models, with over-parameterization and limited information from observations leading to uncertainty about the best parameter vectors. Thus, it is beneficial to find every possible behavioural parameter vector. This paper presents a new methodology, called the patient rule induction method for parameter estimation (PRIM-PE), to define where the behavioural parameter vectors are located in the parameter space. The PRIM-PE was used to discover all regions of the parameter space containing an acceptable model behaviour. This algorithm consists of an initial sampling procedure to generate a parameter sample that sufficiently represents the response surface with a uniform distribution within the “good-enough” region (i.e., performance better than a predefined threshold) and a rule induction component (PRIM), which is then used to define regions in the parameter space in which the acceptable parameter vectors are located. To investigate its ability in different situations, the methodology is evaluated using four test problems. The PRIM-PE sampling procedure was also compared against a Markov chain Monte Carlo sampler known as the differential evolution adaptive Metropolis (DREAM<sub>zS</sub>) algorithm. Finally, a spatially distributed hydrological model calibration problem with two settings (a three-parameter calibration problem and a 23-parameter calibration problem) was solved using the PRIM-PE algorithm. The results show that the PRIM-PE method captured the good-enough region in the parameter space successfully using 8 and 107 boxes for the three-parameter and 23-parameter problems, respectively. This good-enough region can be used in a global sensitivity analysis to provide a broad range of parameter vectors that produce acceptable model performance. Moreover, for a specific objective function and model structure, the size of the boxes can be used as a measure of equifinality.

## KEYWORDS

equifinality, hydrological model, parameter estimation, PRIM-PE, uncertainty quantification

## 1 | INTRODUCTION

A spatially distributed hydrological model is a numerical simplification of water cycle processes in a spatially explicit way, often based on a grid. Many such models have been developed over the last decades (e.g., Arnold, Srinivasan, Muttiah, & Williams, 1998; Borah et al. 2002; Famiglietti & Wood, 1994; Liang, Lettenmaier, Wood, & Burges, 1994; Refsgaard et al. 1995; Wigmosta, Vail, & Lettenmaier, 1994; van Dijk 2010). To optimally apply these models, an efficient and effective

calibration is needed. Two types of approaches have been developed for calibration: (a) manual approaches, which rely mostly on expert judgement and (b) automatic approaches, which employ computer-based routines to find the best parameter vector(s). Although manual calibration still has the potential to provide a good estimation of the parameters, it needs a significant amount of expert labour to obtain an acceptable result (Gupta, Sorooshian, Hogue, & Boyle, 2013). So the development of robust automatic calibration methods that also provide parameter uncertainties is required.

One of the main difficulties in automated calibration (hereafter simply referred to as calibration) is equifinality (Beven, 1993). This issue can be approached from two points of view. First, equifinality can be interpreted as a problem that arises from an inadequate goodness-of-fit measure (or objective function). This means that, due to the overall averaging that is inherent to single value objective functions, different simulation outputs can result in similar objective function values. This reasoning logically led to the development of multi-objective calibration methods (e.g., Yapo, Gupta, & Sorooshian, 1998). Efstratiadis and Koutsoyiannis (2010) reviewed the application of multi-objective calibration studies from the previous decade, concluding that in complex problems with uncertainty, a multi-objective approach can augment identifiability of parameters.

Second, the mismatch between model complexity and information delivered to the model from observations can be a reason for equifinality. To solve this problem, two different types of solutions have been proposed: solutions that reduce the model complexity to match the available information of the observation data (e.g., Arkesteijn and Pande, 2013; Bardsley, Vetrova, & Liu, 2015; Diodato, Brocca, Bellocchi, Fiorillo, & Guadagno, 2014; Dooge, 1997; Fenicia, Kavetski, & Savenije, 2011; Hill, 2006; Schoups, van de Giesen, & Savenije, 2008; Sivakumar, 2008; Sivapalan, Zhang, Vertessy, & Blöschl, 2003; Tonkin & Doherty, 2005) and solutions that accept insufficiency of the observation information while retaining all plausible parameter vectors until a new observation becomes available. In the latter case, the inability to find a behavioural parameter vector is understood to represent uncertainty in the results. Consequently, different simulations using different plausible parameter vectors are shown to be a set of plausible results (Franks & Beven, 1997; Freer, Beven, & Ambrose, 1996). It is noted that uncertainty quantified in this way does not include the model structural uncertainty caused by the model inadequacy itself. Previous studies have attempted to quantify structural uncertainty using different model structures (e.g., Clark et al., 2015; Viney et al., 2009). However, the assessment of model structural uncertainty is beyond the scope of this paper.

To quantify parameter uncertainty, statistical methods such as Bayesian inference have been routinely used in hydrological model parameter estimation. Bayesian methods estimate the posterior parameter distribution, conditional on observations by merging a prior knowledge about the parameters and information from observations. Research on Bayesian parameter estimation, initiated in the early 1980s (e.g., Kuczera, 1983; Sorooshian & Dracup, 1980) is a field of continued interest (e.g., Ajami, Duan, & Sorooshian 2007; Bates & Campbell, 2001; Duan, Ajami, Gao, & Sorooshian, 2007; Hsu, Moradkhani, & Sorooshian, 2009; Keating, Doherty, Vrugt, & Kang, 2010; Kuczera & Parent, 1998; Laloy, Rogiers, Vrugt, Mallants, & Jacques, 2013; Malama, Kuhlman, & James, 2013; Marshall, Nott, & Sharma, 2004; Samanta, Clayton, Mackay, Kruger, & Ewers, 2008; Schoups & Vrugt, 2010; Shafii, Tolson, & Matott, 2014; Smith & Marshall, 2008; Smith, Marshall, & Sharma, 2015; Smith, Sharma, Marshall, Mehrotra, & Sisson, 2010; Thyer et al., 2009; Vrugt, Ter Braak, Clark, Hyman, & Robinson, 2008; Vrugt, Ter Braak, Diks, et al., 2009; Wöhling & Vrugt, 2011). However, the complexity and non-linearity of models causes difficulties in summarizing posterior distributions and is a major obstacle that Bayesian inference for hydrological calibration faces. (Marshall

et al., 2004; Vrugt, 2016). To overcome these difficulties, a number of Monte Carlo based methods were developed to assist modellers to generate samples from posterior distributions and estimate parameter distributions based on these.

One of the best known methods for uncertainty analysis based on Bayesian inference is the Generalized Likelihood Uncertainty Estimation (GLUE; Beven & Binley, 1992) method, which uses the Monte Carlo approach to generate samples from the posterior distribution. GLUE has been widely used in hydrological model calibration. The application of GLUE has been widely criticized for two reasons. First, the informal likelihood function that is often used and, second, the usage of a simple uniform sampling procedure. This sampling approach may work well for low dimensional problems, but it is computationally problematic for a high dimensional parameter space. Samples generated using this assumption are highly likely to be inefficiently scattered over the entire parameter domain (Blasone et al., 2008; Iorgulescu, Beven, & Musy, 2005; Jia & Culver, 2008; Kuczera & Parent, 1998; Stedinger, Vogel, Lee, & Batchelder, 2008; Vrugt, Ter Braak, Gupta, et al., 2009). This sample dispersion in high-dimensional problems causes a significant reduction in the chance of generating near optimal parameter vectors. When calibrating a 13 parameter lumped model at the watershed scale, Jia and Culver (2008) found just 381 near-optimal parameter vectors out of 50,000 trials. However, it should be noted that there is nothing in the GLUE framework that prevents a formal likelihood function and/or a more efficient sampling method being applied (e.g., Blasone et al., 2008).

In addition to the GLUE framework, there are a number of other approaches of sampling that have been developed to overcome the computational difficulties. The basis of the majority of these approaches are Markov chain Monte Carlo (MCMC) approaches. To effectively produce a more focused sample around the regions with a higher likelihood and avoid sampling too much in less important regions, the MCMC approach uses a Markov chain that generates a random walk through the parameter space. Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953) developed the random walk Metropolis algorithm. A number of variants of the random walk Metropolis were developed including Metropolis–Hastings (Hastings, 1970), adaptive Metropolis (Haario, Saksman, & Tamminen, 2001), delayed rejection adaptive Metropolis (Haario, Laine, Mira, & Saksman, 2006), shuffled complex evolution Metropolis algorithm (Vrugt, Gupta, Bouten, & Sorooshian, (2003)), differential evolution Markov chain (Ter Braak, 2006), and differential evolution adaptive Metropolis (DREAM; Vrugt et al., 2008; Vrugt, ter Braak, Gupta, et al., 2009).

The outputs of these methods are samples from posterior parameter distribution. They can be used to calculate the posterior distribution mean, variance, skewness, kurtosis, the higher moments of the distribution, the marginal distributions of individual parameters, the joint probability distribution between two parameters, or a fitted parametric density function. Unfortunately, in high dimensional and complex problems, these statistics cannot provide an interpretation of the posterior distribution shape. Having such an interpretation assists modellers in understanding the global sensitivity of the model performance to its parameters and provides a broad understanding of the range of parameter vectors that produce acceptable model performance, which is very valuable in hydrological modelling. For instance, it can be used as a prior

to a statistical inference of model parameters. Having a continuous space of a good-enough region of the parameter space is also beneficial when the reproduction of a new parameter vector ensemble within the good-enough region is of interest, especially when the model is computationally expensive to run. In this situation, the continuous definition is useful to produce a new good-enough sample effortlessly.

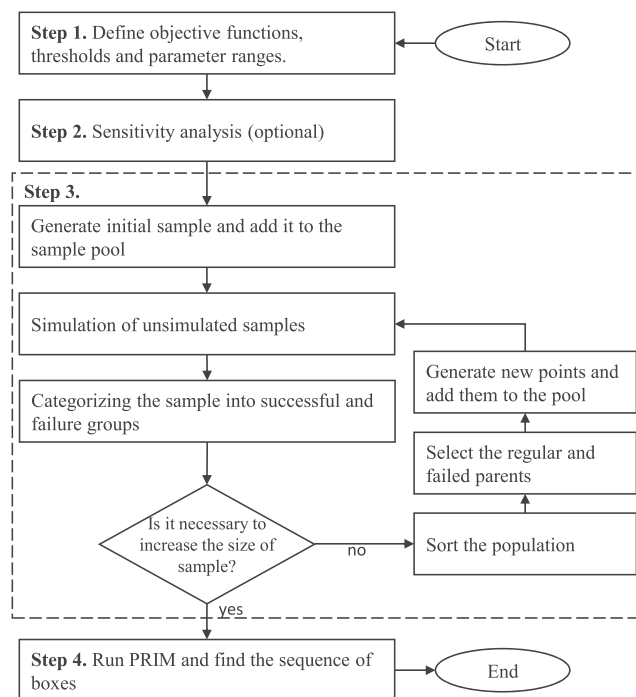
The patient rule induction method (PRIM; Friedman & Fisher, 1999) is a tool that is used to find a region of interest using a discrete sample and then define it in an interpretable way using a set of hypercubes (or boxes). PRIM is a well-known method in the field of decision-making under deep uncertainty conditions and is mostly known as a part of robust decision-making (Bryant & Lempert, 2010; Lempert, 2002; Lempert et al., 2016; Lempert, Popper, & Bankes, 2003; Shortridge & Guikema, 2016) framework. However, despite the ability of PRIM to find a region of interest using a discrete sample, this tool has not been tested for parameter uncertainty analysis of hydrological models.

The aim of this study is to propose a methodology based on PRIM, called here the PRIM for parameter estimation (PRIM-PE), which determines all the plausible model parameterizations that produce good-enough model performances as an interpretable region in the form of a set of hypercubes (boxes). This region is defined as an area that contains all of the parameters with an objective function value better than a predefined threshold. The resulting region represents a border that separates the good-enough region from the rest of the parameter space. This border is technically an interpretable approximation of a cross-section of the objective function. The main innovation of the PRIM-PE is the form of its final results. In this algorithm, an easier to interpret definition of the good-enough region helps the modellers to know the model and its reaction to the parameters in an alternative way. Moreover, the low theoretical complexity facilitates the PRIM-PE implementation while it provides an accurate understanding of the model. The PRIM needs a sample of evaluated parameter vectors. Every sample of parameter vectors will lead to reliable box edges as long as the sample is well spread in the vicinity of the border. On the basis of this, the characteristics of an efficiently spread sample for the current methodology differ from an efficiently spread sample for other Bayesian methods, which try to generate the sample uniformly (uniform sampler in some of the GLUE applications) or with higher density at the region with a higher likelihood. Therefore, a new sampler is also proposed to generate an appropriate and efficient sample. The PRIM-PE methodology was first tested with a number of mathematical test problems. Then, it was benchmarked against the DREAM<sub>25</sub> (Laloy & Vrugt, 2012; Vrugt, 2016) method by solving a 5-parameter hydrological model, and finally, it was used to calibrate a spatially distributed hydrological model.

## 2 | METHODOLOGY

### 2.1 | The PRIM-PE approach

Given the uncertainties and presence of equifinality inherent in calibration, each parameter vector performing better than a specific threshold is valuable. Therefore, to assist in finding regions containing these parameter vectors, a new parameter estimation approach (PRIM-PE)



**FIGURE 1** Schematic of the patient rule induction method for parameter estimation (PRIM-PE) methodology

is proposed. This algorithm starts with generating a parameter sample that includes several different parameter vectors, which are then used to generate simulations. On the basis of a predefined set of rules, the generated parameter vectors are then categorized as a “success” or “failure.” The success rules should be able to separate sets with good-enough performance from the unacceptable. A good choice for such rules could be a number of objective functions and thresholds to test the plausible sets. Finally, the PRIM routine is used to define a sequence of “hypercubes” (referred to from herein as boxes) in the parameter space that cover just the successful points. These boxes are in fact a set of boundaries for the parameter values that form a hypercube in parameter space. If the sample size is large enough, modeller can be sure that any random parameter vector generated within these boxes will result in good-enough simulations and hence represent a good-enough parameter vector. Thus, the region that these boxes cover are referred to as the good-enough region. Figure 1 shows a schematic of the algorithm.

The first step in this method is to define a number of rules to categorize different parameter vectors into the failure or success category. The easiest way to do this is to choose one or more objective function(s) with a cut-off threshold for each. A parameter vector can then be considered as a successful solution if all (or a subset) of its objective functions are better than the corresponding thresholds. Many different formulations for measuring the goodness-of-fit are proposed in the literature. Legates and McCabe (1999) reviewed different types of objective functions and concluded that correlation-based objective functions are not suitable to assess the performance of hydrological models and that the Nash–Sutcliffe efficiency (NSE), root mean square error, and mean absolute error are preferable. In this study, the NSE is used as the objective function. The formulation of the NSE is

$$NSE = 1 - \sum_{t=1}^T \frac{(Q_o^t - Q_m^t)^2}{(Q_o^t - \bar{Q}_o)^2}, \quad (1)$$

where  $Q_o^t$  and  $Q_m^t$  are the observed and modelled value at the  $t$ th time step,  $T$  is the number of time steps, and  $\bar{Q}_o$  is the averaged observed value. It is noted that despite the popularity of the NSE, there are some disadvantages in using it as an objective function. The most important is its high sensitivity to extreme values (Legates & McCabe, 1999). After defining the objective function, a cut-off threshold is determined on the basis of expert judgement and the objectives of the problem. However, it can also be chosen as a small amount worse than the global optimum objective function value: Even though the equifinality issue causes difficulty to find the optimal solution, finding an approximation of the best objective function is easily possible. In this study, the best objective function value was first approximated using a genetic algorithm (Goldberg, 1989) optimization procedure. In order to define a good-enough region around the best objective function value, the threshold is then defined slightly lower than the optimal objective function value by a subjectively chosen fraction. After defining the objective functions and thresholds, a set of physically meaningful ranges for parameter values are defined by the modeller.

The second step involves an optional sensitivity analysis. It provides a better overall understanding of the simulation system, and it also helps to pay more attention to highly sensitive parameters, discarding the less sensitive parameters, and consequently reduce the dimensionality of the problem. In this study, the sensitivity analysis was performed by changing one parameter at a time. Although the optimum solution calculated in the previous step is only one of the possible optimum solutions, it can be used as the centre point in the sensitivity analysis. Furthermore, the results of the sensitivity analysis can be verified after finding the good-enough region using random points from that region.

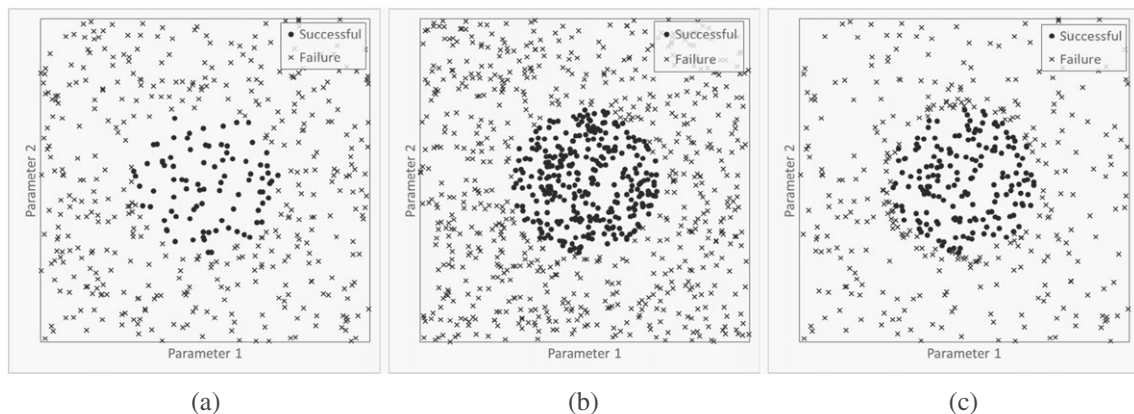
After the first two steps, a sample of parameter vectors, which is representative of the model response surface, is generated. To obtain a reliable result, the density of produced samples within the parameter space should be sufficiently high. In principle, the parameter estimation process can be undertaken using a completely random sample, but obtaining such a dense sample in a high-dimensional problem could easily be computationally problematic. Therefore, a sequential sampling procedure is proposed to effectively reduce the sample size, while preserving the density of points in the desired region in the parameter space (step 3).

To be more specific, a high density of points in and near the successful region is necessary, but the density of points in the surrounding failure region is not important. For instance, consider a two-parameter problem (Figure 2), which shows two randomly generated samples with a low (500 points) and a high (2,000 points) density. As can be seen, the denser sample is more illustrative of the successful area. The sequential procedure that is proposed here produces a sample with higher density around the successful area and lower density elsewhere. Figure 2c is a varying density sample with 500 points, which illustrates the boundary of good-enough region more accurately than Figure 2a with the same sample size.

The sampling approach starts with generating a completely random initial sample and simulating its members. After that, the points are sorted on the basis of their success. Then, the successful points are sorted on the basis of their local densities, and the failure points are sorted on the basis of their objective function values. Figure 3 shows an example of the sorting procedure. The local density is defined as the distance to the  $k$ th nearest point ( $k$  is a small integer). After sorting, the first  $n_1$  points (regular parents) in addition to  $n_2$  points from the failed population (failed parents) are selected to produce the new generation. The selection of the failed parents is based on a tournament selection (Miller & Goldberg, 1995) so that a number of the failed points ( $t$  points) are selected randomly and among them the points with the lowest local density chosen as a failure parent. This selection strategy improves the chance of searching in the low density regions to spot the other potentially good-enough regions.

After selecting the parents,  $m$  points are generated randomly around each selected point in a distance less than the average local density of successful points ( $n_1$ ,  $n_2$ , and  $m$  should be set by the modeller). After generating the new points, they are added to the sample, and the process is again repeated. The stop criterion for this iterative process is defined on the basis of the final sample size or average local density of successful points. This process assists in producing a sample of points with a good density in the successful region, and in problems with multiple optima, it can produce a sample with a sufficient density in all of the optima regions.

It should be noted that generating new points becomes challenging when the parent is located near the edges of the parameter space. In this situation, there is possibility of producing infeasible children (points out of the parameter space). Three main strategies can be



**FIGURE 2** Examples of different sampling approaches: (a) 500 points; (b) 2,000 points; and (c) 500 points using the proposed sampling method

	Density function	Objective function	Successful?
Point 1	0.52	9.33	1
Point 2	0.07	9.57	1
Point 8	0.07	7.79	1
Point 7	0.04	7.99	1
Point 20	0.88	6.53	0
Point 13	0.85	6.38	0
Point 3	0.21	6.37	0
Point 4	0.86	5.66	0
Point 14	0.79	5.49	0
⋮	⋮	⋮	⋮
Point 9	0.69	0.96	0

Selected points  
to produce the  
next generation  
( $n=6$ )

**FIGURE 3** An example of sorting in the sampling procedure. Consider a problem with an objective function (that should be maximized) in the range of  $[0, 10]$  and a predefined threshold equal to 7. In the first generation, 4 points (out of 20 points) perform better than the threshold (the grey area); they are sorted on the basis of the density functions, whereas the rest of the points are sorted on the basis of the objective functions

adopted for such situations: (a) to discard those children and produce new ones, (b) to map the infeasible children into the feasible space, and (c) to artificially expand the domain and assign a low value as the objective function of those children. The first and second strategy lead to a higher density at the locations that the infeasible points are mapped into. Conversely, by expanding the parameter space, the uniformity of the sample is retained, but a number of infeasible points emerge at the final sample. In this study, the later strategy was adopted.

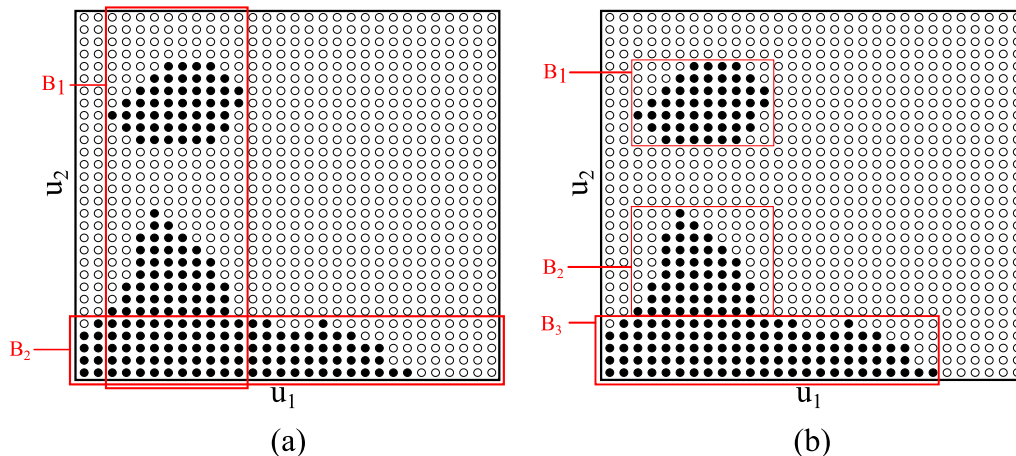
The fourth and final step is to extract the sequence of boxes that define the good-enough region using the PRIM method. Here, a brief description of the PRIM methodology is presented. Full details are available in Friedman and Fisher (1999).

PRIM is a clustering algorithm that finds one or more boxes in the parameter space using three criteria: density, coverage, and interpretability. Density is defined as the total number of successful points divided by the total number of points inside the box. Coverage is the total number of successful points inside the box divided by the total number of successful points, and interpretability is approximated by the number of dimensions defining each box. Figure 4a shows a box sequence with lower density than Figure 4b but higher interpretability.

As can be seen, each box in the result of PRIM can be expressed in the form of a subset of parameters which are constrained by upper and lower bounds, such that

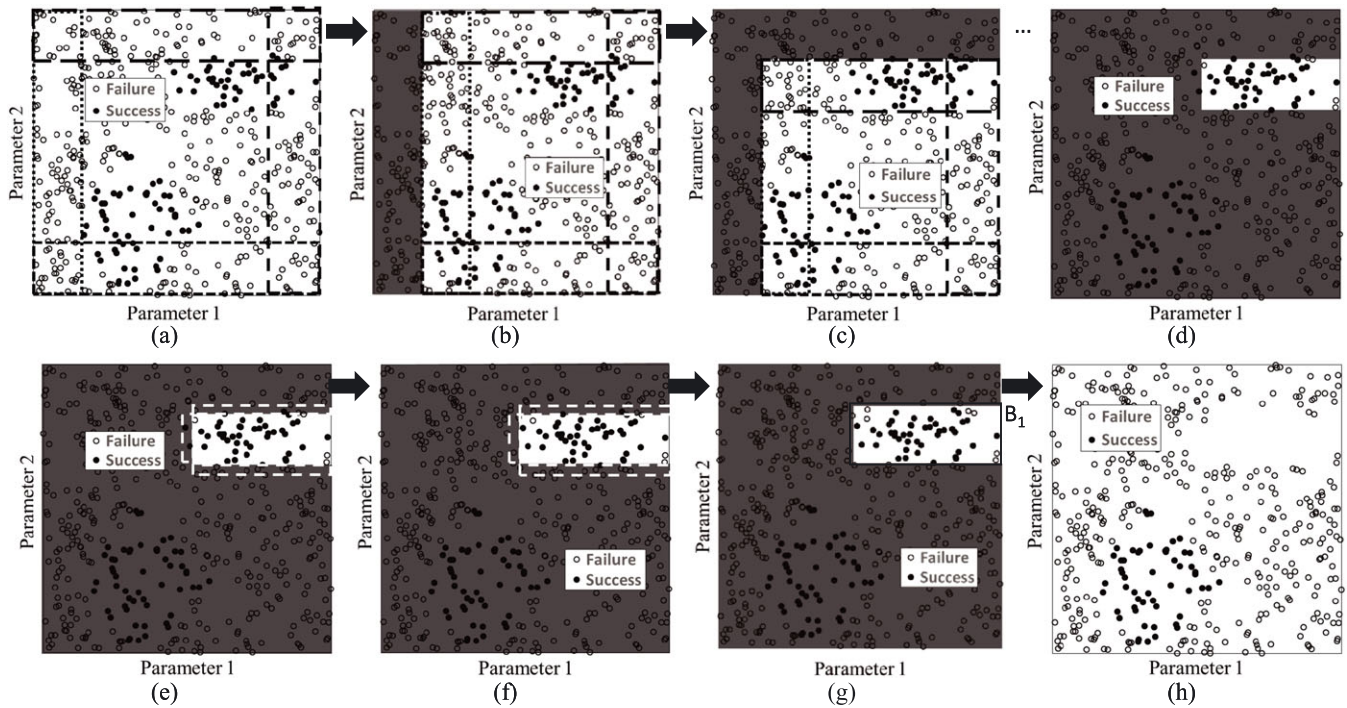
$$B = \{a_j < x_j < b_j, j \in L\}, \quad (2)$$

where  $B$  is the box boundary set,  $x_j$  is the  $j$ th parameter,  $a_j$  and  $b_j$  are the lower and upper bounds, and  $L$  is the number of parameters. In order to find these subsets, the PRIM uses a two-step procedure: peeling and pasting. In the peeling step, a box ( $B$ ) covering the whole parameter space is considered as an initial box; then, in an iterative procedure among available peeling alternatives, the one that increases the density of successful points inside the box by removing the box is removed. (Figure 5a). The peeling alternatives are narrow boxes on each side of the main box. This iterative process is continued until the density or coverage criteria exceed a predefined value. Consequently, various alternative boxes can be defined on the basis of a different density or coverage criteria. Generally, by increasing the value of the acceptable density, the size of the boxes, and as a result their coverage, reduces, while the number of boxes needed to represent the good-enough region increases.



**FIGURE 4** Example of sequence of boxes with different criteria: (a) low density and high interpretability and (b) high density and low interpretability





**FIGURE 5** Top panel: the row peeling procedure. Bottom panel: the pasting procedure. Black dashed lines represent the options to be removed in the peeling procedure, white dashed lines represent the options to be added in the pasting procedure, and the grey area is the removed parts. (a) The initial box that contains all points and four options to be peeled. (b and c) The illustrations of the remaining box after removing the first and second parts. (d) The final box at the end of the peeling procedure. (e) The options of the pasting in the first and second pasting stages. (f) The remaining box after removing the first and second parts. (g) The final box (B1). After finding the final box, the successful points are removed from the sample and (h) used for the next step to find the next box

Consequently, it is harder to interpret a larger sequence of boxes. Thus, there is an inherent trade-off between the acceptable density, the coverage, and the interpretability of the resulting box sequence. The optimal density should be selected on the basis of the type of problem. For instance, if one is looking for a rough estimate of a “region of interest” with a high interpretability, a lower acceptable density should be chosen. Alternatively, if an accurate representation of the “region of interest” is required, a higher density and greater number of smaller boxes should be considered. In the parameter estimation applications of PRIM, high density boxes are preferred because accurate boxes are sought to perfectly partition the parameter space.

At the end of the peeling procedure, the pasting procedure starts. The pasting process is an inverse version of peeling in which narrow boxes are added to the sides of the main box in each iteration. Figure 5b shows the pasting process. After finding the final box, the successful points inside the box are removed from the sample pool, and the process to find the next box starts. Finally, after removing all successful points, the final sequence of boxes is considered the good-enough region. Bryant (2009) developed a tool to perform a PRIM analysis called The Scenario Discovery Tool to Support Robust Decision Making. In this study, this tool, with minor modifications to automatize the procedure, is used for the PRIM analysis.

## 2.2 | MCMC sampling with DREAM<sub>ZS</sub>

To benchmark the proposed method, the results were validated against the results of the DREAM<sub>ZS</sub> algorithm (Laloy & Vrugt, 2012; Vrugt, 2016), which is an elegant approach to quantify parameter uncertainty

and posterior distributions. This algorithm is an adaptive multiple-chain MCMC simulation, being an adaptation of an MCMC algorithm entitled the shuffled complex evolution Metropolis algorithm (Vrugt et al., 2003). The DREAM<sub>ZS</sub> methodology runs multiple chains simultaneously, to explore the entire parameter space. This method uses a snooker sampling approach (Ter Braak & Vrugt, 2008) on the basis of an archive of the past states of the chains rather than their current states. A detailed description of the DREAM<sub>ZS</sub> algorithm was presented in Laloy and Vrugt (2012). In this paper, the test case of Vrugt et al. (2008) is used to validate the PRIM methodology.

## 3 | PRIM-PE METHOD EVALUATION

In order to test the performance of the PRIM-PE in different situations, a number of test problems were solved including (a) modelling a cosine function using a linear equation to show general characteristics of the proposed method, (b) a banana shape distribution fitting problem to test its ability in the identification of a curved good-enough region, (c) a multiple optima problem to test the applicability of multiple good-enough regions, and (d) a lumped rainfall-runoff hydrological model calibration problem. As was mentioned in the previous sections, the PRIM-PE method consists of a sampling and a rule induction component. In the current section, to evaluate the method performance, the characteristics of the sampling procedure were first explored and compared against an existing MCMC method (DREAM<sub>ZS</sub>). The rule induction component (PRIM) was then employed to detect the good-enough region as a sequence of interpretable boxes, and the accuracy of those boxes was finally evaluated. It should be emphasized that

any sample can be used as the input to the rule induction component (including the sample from the MCMC methods). For instance, among the resulting points from DREAM<sub>ZS</sub>, those that make up a given percentage of the probability mass of the target cumulative distribution function can be considered as successful points. However, the sampling procedure in the PRIM-PE method was designed to optimize the efficiency and accuracy of the rule induction component (PRIM) by producing a uniformly distributed sample over the good-enough region.

### 3.1 | A two-dimensional banana-shaped distribution fitting test problem

Because the PRIM defines the good-enough region as a sequence of boxes, a concern arises regarding its ability to face highly non-linear and/or concave good-enough regions. To overcome this concern, a two-dimensional non-linear banana-shaped distribution was used to test the PRIM-PE capability to cope with such problems. This test problem has previously been used to test different parameter uncertainty quantification problems (e.g., Vrugt et al., 2003). This distribution is a twisted bivariate normal distribution. The density function is

$$f(\theta_1, \theta_2) = \exp\left(-\frac{1}{2}\left(\frac{\theta_1^2}{100} + \frac{(\theta_2 + b\theta_1^2 - 100b)^2}{1}\right)\right), \quad (3)$$

in which  $\theta_1$  and  $\theta_2$  are the parameters and  $b$  is a constant, which determines the non-linearity. In this study, a value of 0.1 was assigned to  $b$ , and the logarithm of the density function was used as the objective function. Figure 6a shows the response surface of this problem.

#### 3.1.1 | Sampling efficiency and accuracy

To evaluate the efficiency and accuracy of the PRIM-PE sampling procedure, three samples with different sizes (i.e., 1,000, 5,000, and 10,000) were generated. As can be seen in Figure 6a, the best value of Figure 3 is slightly higher than  $-4.2$ , so the threshold for the PRIM-PE was selected as  $-5$ . The top panel in Figure 7(a–c) shows these samples. In these figures, the blue and red points represent the successful (behavioural) and failure (non-behavioural) points, and the black line is the good-enough region. According to these figures, it is observed that PRIM-PE managed to produce samples that are uniformly distributed over the good-enough region and a narrow buffer around it. This

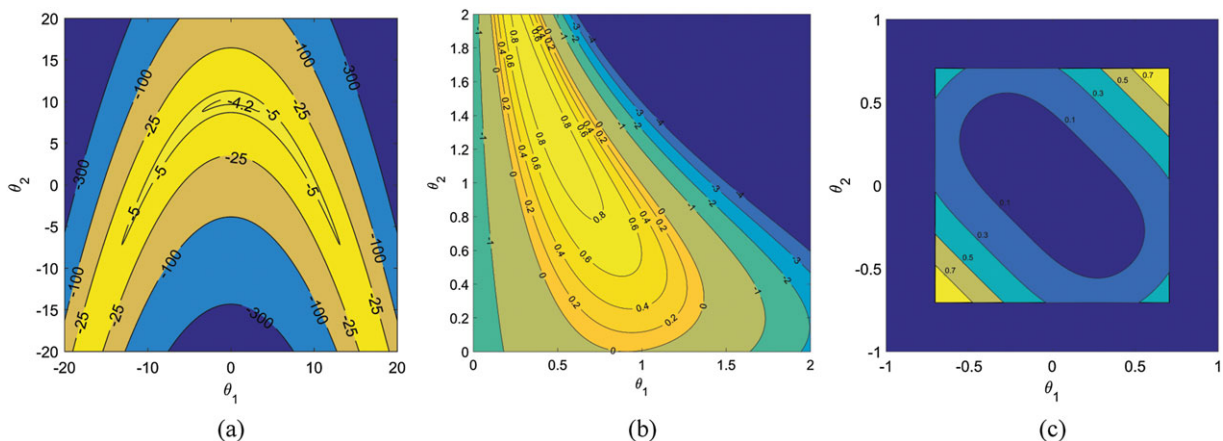
uniformity reveals the border efficiently. As shown in the figures, even the sample with the lowest size (1,000 points) had a uniform distribution and an acceptable coverage.

In the next step, to have a better understanding about the characteristics of the samples, three samples with the same sizes as the previous ones (i.e., 1,000, 5,000, and 10,000) were generated by the DREAM<sub>ZS</sub> method. For this purpose, prior PDFs were considered to be uniform distributions, and the likelihood function was Equation 3. However, the concentration of the points in the good-enough region can potentially be improved by applying informed prior Probability Density Functions (PDF). The second panel of Figure 7 (d–f) shows these samples. The red points represent all parameter vectors that were evaluated, and the blue points are the parameters that were selected as part of the chain to express the target distribution. The resulting samples had higher concentrations at the central parts (the region with higher likelihood), and they perfectly captured the target probability distributions. By comparing the first two panels, it can be concluded that, although the resulting samples of both methods reveal the good-enough boundaries, the PRIM-PE, due to its sample uniform distributions, was more efficient for this purpose, and it provides a better opportunity to define the good-enough boundary accurately. Conversely, the PRIM-PE was not as accurate as the MCMC methods (more specifically, the DREAM<sub>ZS</sub>) at generating a sample with the target distribution, which is not a PRIM-PE objective.

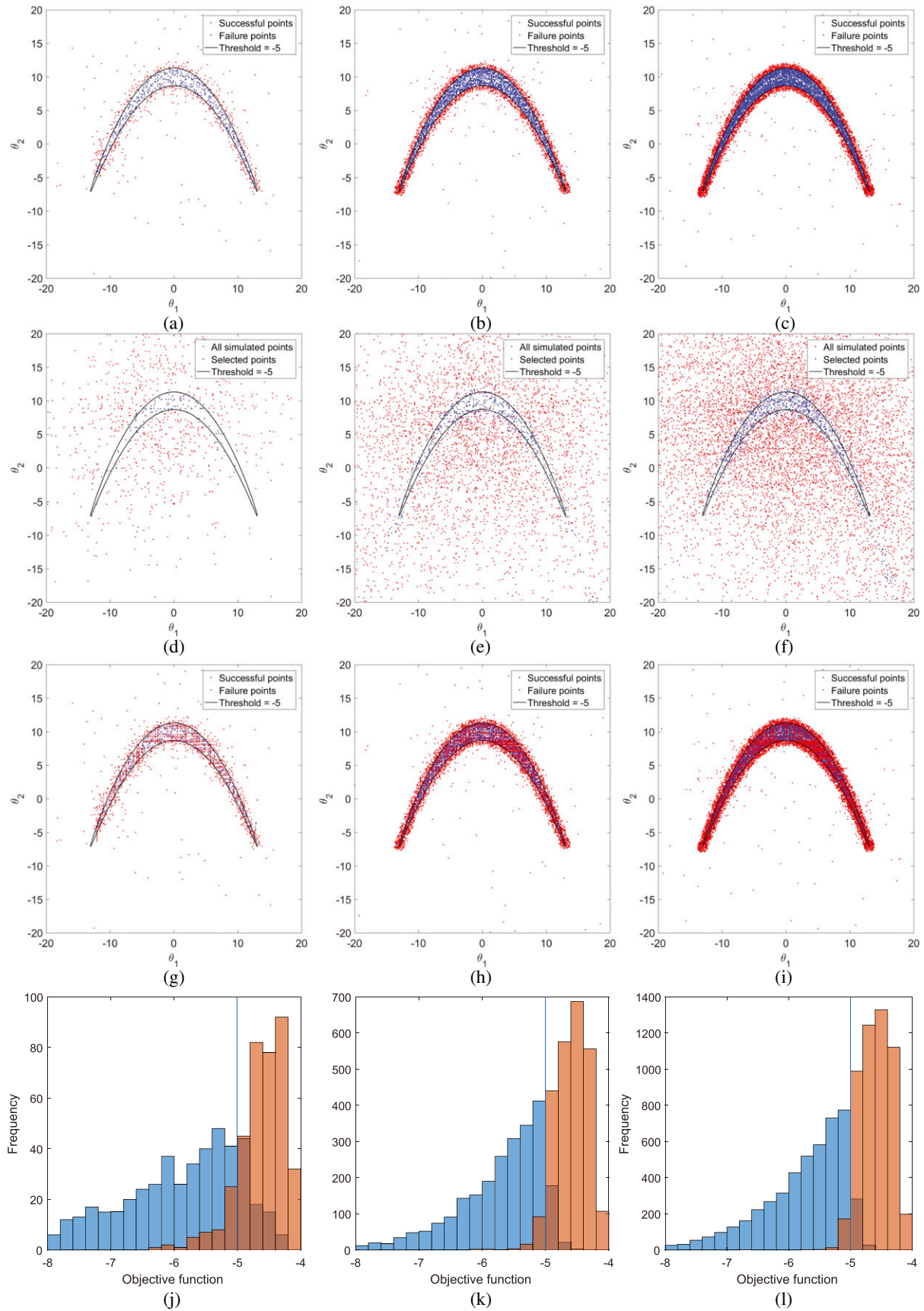
It should be mentioned that a step likelihood function (just similar to the goodness measure of the PRIM-PE) was also tested for the DREAM<sub>ZS</sub> algorithm. The resulting sample was poorly distributed and has not captured the good-enough region properly (result is not presented here). The main reason is, if an appropriate density function is not applied as the likelihood function, due to a poorly defined Metropolis ratio, the DREAM<sub>ZS</sub> algorithm cannot properly search the domain. Consequently, it is unlikely that desirable results will be achieved.

#### 3.1.2 | Accuracy of PRIM

After generating a sample, the rule induction component (PRIM) of the PRIM-PE could be employed to define the good-enough region as a set of boxes. Figure 7 (g–i) illustrate these boxes for different samples. According to these figures, by increasing the sample size, the accuracy of the boxes in terms of correctly capturing the good-enough region



**FIGURE 6** Response surfaces of the test problems: (a) banana-shaped distribution fitting; (b) cosine; and (c) multi-optima



**FIGURE 7** Comparison of the sampling processes for the two-dimensional banana-shaped distribution fitting test problem. The bottom row represents the histograms of the points being categorized as non-behavioural (blue) and behavioural (orange). (a) 1,000 points (PRIM-PE); (b) 5,000 points (PRIM-PE); (c) 10,000 points (PRIM-PE); (d) 1,000 points (DREAM<sub>ZS</sub>); (e) 5,000 points (DREAM<sub>ZS</sub>); (f) 10,000 points (DREAM<sub>ZS</sub>); (g) 1,000 points (Boxes); (h) 5,000 points (Boxes); (i) 10,000 points (Boxes); (j) Boxes accuracy (1,000 points); (k) Boxes accuracy (5,000 points); (l) Boxes accuracy (10,000 points)



increased. It should be noted that the PRIM algorithm in this example was forced to produce the boxes with 100% density (meaning that all the points inside the boxes should be behavioural). However, it can be seen that, in the examples with a lower sample sizes, there were some overestimations and underestimations, which were caused by low density in those specific points. For instance, in places with lower number of failure points at the edge, the algorithm wrongly detected the boundary towards outside the good-enough region. This means that, using a high density sample at edges can help the PRIM to avoid this errors.

### 3.1.3 | Validation

To validate the resulting boxes, their skill in representing the good-enough region should be evaluated. To do so, a number of independent samples (with the same size of the training samples) were generated, and the resulting box sequences were employed to distinguish the behavioural and non-behavioural points. Figure 7(j-l) represents the histogram of the points being categorized as non-behavioural (blue) and behavioural (orange), and the vertical line shows the threshold (which was selected as -5). The accuracy of the produced boxes is measured as the ratio of the number of correctly recognized behavioural and non-behavioural points to their total number. According to these figures, the resulting boxes could successfully recognize more than 85% of the behavioural and non-behavioural points in all experiments, and these ratios increased by increasing the sample size. The detailed values of accuracies are represented in Table 1.

## 3.2 | Cosine test problem

The second problem is a cosine test problem. In this problem, a cosine function as "truth" was modelled by a simple equation. The truth relation between  $X$  and  $Y$  is given by

$$Y = \cos(X), \quad (4)$$

with the model equation

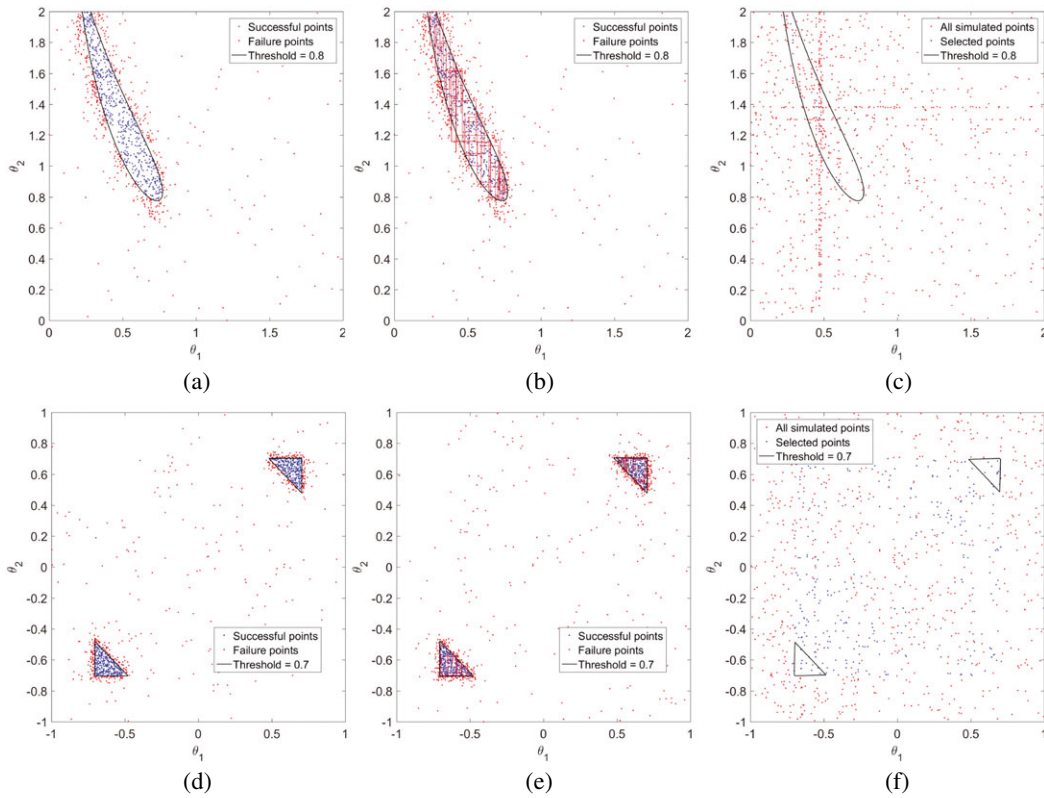
$$Y = 1 - a(X + \epsilon)^n \quad (5)$$

used to simulate Equation 4 and  $a$  and  $n$  the parameters to be estimated. Accordingly, 1,000 samples were randomly generated in the range of  $[-\pi, \pi]$  with the true value of the corresponding  $Y$  calculated using Equation 4. For each parameter vector, the NSE was calculated for the results of Equation 5. In this equation,  $\epsilon$  represents a normal Gaussian random noise with 0.1 and 0.2 mean and standard deviation. Figure 6b shows the response surface of this test problem. The highest possible value of NSE for this problem is approximately 0.9. A slightly lower value of 0.8 was considered as the threshold. Similar to previous test problem, first, the PRIM-PE sampling procedure and the DREAM<sub>ZS</sub> were used to generate samples with different sizes (1,000, 5,000, and 10,000). After this, the good-enough region was detected by the PRIM. For this problem, a likelihood function consistent with previous studies (e.g., Vrugt (2016)) was adopted for DREAM<sub>ZS</sub>. This likelihood function transforms the residual between the truth and simulated  $Y$  values in a Gaussian likelihood function. Moreover, all prior PDFs were considered to be uniform distributions. Similar to the first case study, by using informed prior PDFs and likelihood function, the DREAM<sub>ZS</sub> is expected to generate a sample with higher concentration in the good-enough region. The top panel of Figure 8 shows the sample and box sequence generated using the PRIM-PE and the sample generated using the DREAM<sub>ZS</sub> with size of 1,000. The details of the validation for all the experiments with different sample sizes are presented in Table 1. A same pattern as the first test problem was obtained. By increasing the sample size, more accurate box sequences were achieved. Another considerable point is the better accuracy of this test problem using a lower number of boxes. The main reason for this difference is the simpler (less concave) shape of the good-enough region so that the boxes could better represent the region.

**TABLE 1** Accuracy of the PRIM-PE resulting boxes in detecting the good-enough region

Problem	No. pars	Threshold	MAD (%)	Sample size	No. of boxes	$\alpha$ (%)	$\beta$ (%)
Banana-shaped	2	-5	100	1,000	70	86.7	86.4
		-5	100	5,000	213	95.4	91.8
		-5	100	10,000	348	96.3	93.6
Cosine	2	0.8	100	1,000	31	94.0	92.6
		0.8	100	5,000	102	97.5	94.8
		0.8	100	10,000	147	98.3	94.4
Multi-optima	2	0.7	100	1,000	23	96.5	87.5
		0.7	100	5,000	75	98.7	96.8
		0.7	100	10,000	109	99.1	97.6
HYMOD	5	16.1	100	50,000	924	80.7	99.5
		16.2	100	50,000	1,101	81.0	99.0
		16.5	100	50,000	1,152	88.0	98.5
		1.7	100	50,000	1,392	88.6	97.8
HYMOD15	15	14.5	80	1,000,000	5,350	78.9	96.5
		14.5	90	1,000,000	7,932	70.1	98.4
		14.5	100	1,000,000	9,778	65.2	97.7

Note.  $\alpha$  = the percentage of behavioural points that were detected correctly;  $\beta$  = the percentage of non-behavioural points that were detected correctly; HYMOD = hydrological model; MAD = minimum accepted density; PRIM-PE = patient rule induction method for parameter estimation.



**FIGURE 8** Comparison of the sampling processes for the cosine and multi-optima test problems with 1,000 sample size: (a) PRIM-PE sample (cosine); (b) box sequence (cosine); (c) DREAM<sub>ZS</sub> sample (cosine); (d) PRIM-PE sample (multi-optima); (e) box sequence (multi-optima); and (f) DREAM<sub>ZS</sub> sample (multi-optima)

### 3.3 | Multi-optima test problem

The third test problem is designed to test the ability of the PRIM-PE in a problem with multiple optima and also with optimum points located at the edges of the parameter space. The formulation of its objective function is

$$f = \begin{cases} \frac{(\theta_1 + \theta_2)^2}{2} + \frac{(\theta_1 - \theta_2)^4}{8}, & |\theta_1| \text{ and } |\theta_2| < \sqrt{2} \\ 0, & |\theta_1| \text{ or } |\theta_2| \geq \sqrt{2} \end{cases} \quad (6)$$

Figure 6c shows the response surface of this test problem. Two optimum points and consequently two good-enough regions are located at the top right and the bottom left in the parameter space. The optimum value of the objective function (which is occurring in two points) is approximately 0.8, and a slightly lower value of 0.7 was considered as the threshold. Similar to the previous test problems, the sampling process of the PRIM-PE method and the DREAM<sub>ZS</sub> were used to generate samples with different sizes (1,000, 5,000, and 10,000) after which the good-enough region was detected using the PRIM-PE. In this problem, Equation 6 was used as the likelihood function, and all prior PDFs were considered to be uniform distributions. Similar as in the previous case studies, if informed prior PDFs and likelihood were implemented, the DREAM<sub>ZS</sub> would potentially produce more samples in the good-enough regions. The bottom panel of Figure 8 shows the sample and the box sequence generated using the PRIM-PE and the sample generated by DREAM<sub>ZS</sub> with size of 1,000. The details of the validation of all the experiments with different sample sizes are presented in Table 1. As can be seen in the figures, the method successfully detected both optimum (good-enough) regions. A notable point in this problem

is the fact that the optimum points are located exactly at the edge of the parameter space ( $[\sqrt{2}, \sqrt{2}]$  and  $[-\sqrt{2}, -\sqrt{2}]$ ). In this situation, the resulting children in the sampling procedure could be generated out of the boundaries. As is mentioned in the methodology section, three main strategies can be adopted for this kind of situations: (a) to discard those children and produce new ones, (b) to map the infeasible children (points out of the parameter space) into the search space, and (c) to artificially expand the domain and assign a low value as the objective function of those children. Using the first and second strategy leads to a sample with higher density at the edge or the location that the infeasible points are mapped into. Conversely, by expanding the parameter space, the uniformity of the sample is retained, but a number of infeasible points emerge at the final sample. Because these infeasible points can be distinguished before simulation (just by checking the boundaries), they do not impose a strong computational effort. For this reason, in this problem, the expanding strategy was adopted to cope with infeasible children.

### 3.4 | HYMOD

The last test problem is the parameter estimation of a parsimonious lumped rainfall-run-off model called HYDrological MODEL (HYMOD; Boyle, Gupta, & Sorooshian, 2000). The HYMOD model consists of two series of reservoirs including a set of three reservoirs to model surface run-off and a single reservoir to represent baseflow. This model has five parameters including the maximum capacity in the catchment ( $C_{\max}$ ), the degree of spatial variability of soil moisture capacity within the catchment ( $b_{\exp}$ ), the coefficient that divides the flow into two parts of slow and quick run-off ( $\alpha$ ), and two parameters for routing system

**TABLE 2** HYMOD parameter boundaries and descriptions

Parameter	Description	Minimum	Maximum	Unit
$C_{\max}$	Maximum storage in watershed	1	500	mm
$b_{\exp}$	Spatial variability of soil moisture storage	0.1	2.0	—
$\alpha$	Distribution factor between two reservoirs	0.10	0.99	—
$R_s$	Residence time slow flow reservoir	0.001	0.100	days
$R_q$	Residence time quick flow reservoir	0.10	0.99	days

Note. Adapted from Vrugt et al. (2008). HYMOD = hydrological model.

that describe the residence times of the reservoirs ( $R_s$  and  $R_q$ ). For more details about this model, refer to previous papers (e.g., Boyle et al., 2000; Vrugt et al., 2008).

Because of its simple structure, the limited number of parameters and consequently the ease of analysing the results, HYMOD was used as a part of the validation process for the PRIM-PE. Similar to the previous test problems, the sampling procedure of the PRIM-PE and DREAM<sub>z5</sub> was used to generate a number of samples and then the PRIM employed to induct a sequence of boxes that covering the good-enough region. In order to maintain consistency with other previous studies, the publicly available DREAM<sub>z5</sub> package and the built-in HYMOD function with the data set for the Leaf River catchment (Vrugt, 2016) were used. A 2-year time period from September 30, 1952 to September 30, 1954 was selected for this study. The parameter boundaries were also selected similar to those defined in the previous application of the DREAM<sub>z5</sub> algorithm (Table 2).

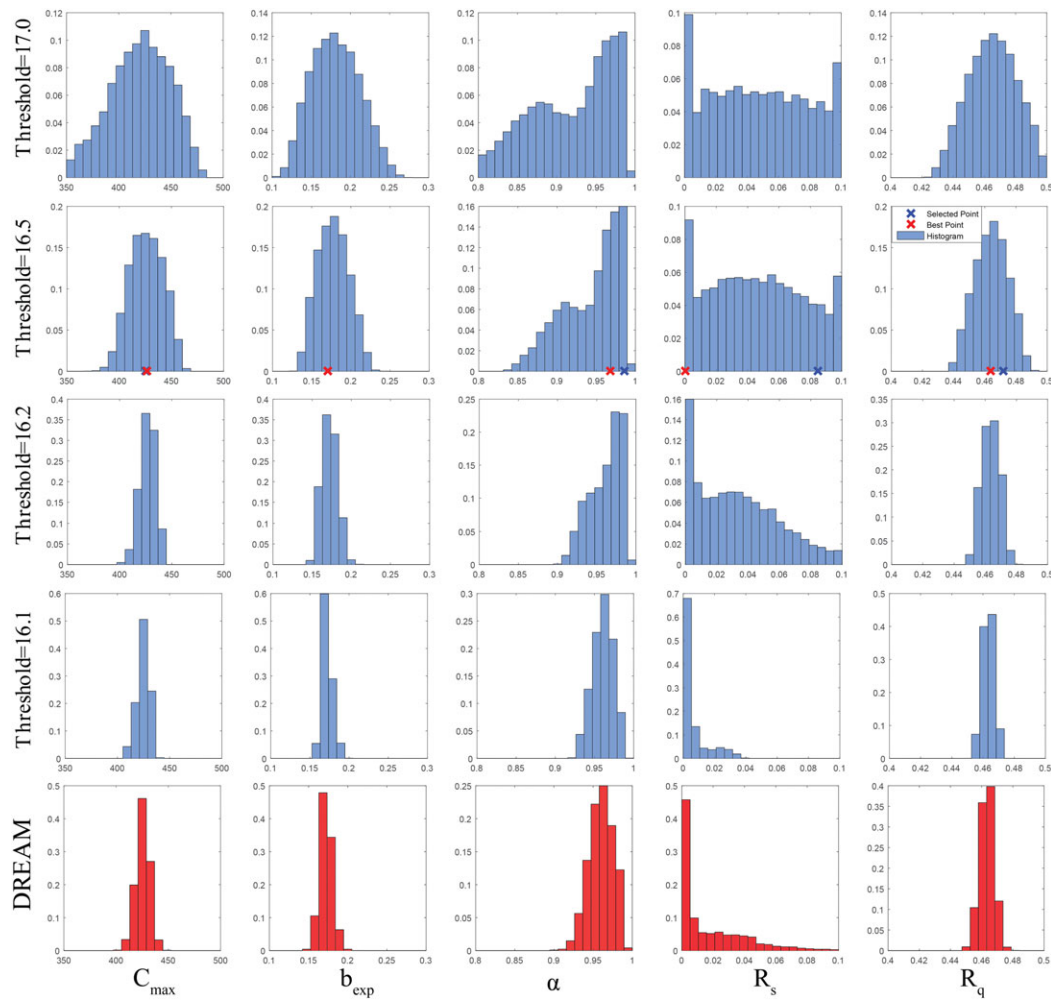
To quantify the model parameter uncertainty, a sample size of 50,000 parameter vectors was generated using the DREAM<sub>z5</sub> with  $N = 5$  Markov chains using a Gaussian likelihood (consistent with previous studies, e.g., Vrugt 2016). The convergence of the algorithm was checked using the  $\hat{R}$  convergence diagnostic (introduced by Gelman & Rubin, 1992). The results showed that the  $\hat{R}$  values for all the parameters were less than the suggested threshold (1.2) after 20% of trials. In the PRIM-PE algorithm application, a similar sample size was used, in which 100 points were initially generated randomly. Then, the rest of the population was generated in 2,495 generations. The root mean square error was considered as the objective function. To show the changes in shape of the good-enough region by changing the cut-off threshold, the PRIM-PE was run using a set of different thresholds. As the global optimum objective function was 16.03, thresholds of 16.10, 16.20, 16.50, and 17.00 were chosen. Although the objective of the PRIM-PE was not finding the target distribution, a sample with a uniform distribution within the good-enough region was of interest. To have a better understanding about the characteristics of the samples, the marginal distributions of the resulting sample were investigated. Figure 9 shows the histogram of the successful points in the resulting samples generated by the proposed method (four upper panels) and the marginal distribution of last 20% of parameters found by the DREAM<sub>z5</sub> algorithm (the bottom panel). It should be noted that the ratio of 20% was chosen to be consistent with the previous DREAM applications (e.g., Vrugt et al., 2008). Higher ratios (up to 80%) resulted in practically identical marginal distributions. To facilitate the analysis, the range of the horizontal axis for each parameter was fixed.

As can be seen, the resulting histograms from the PRIM-PE using a strict threshold (very near to the global optimum objective function, 16.1) were quite similar to the marginal distribution of the DREAM<sub>z5</sub>. Both histogram sets show small dispersions around the global optimum point. However, the characteristics of the resulting histograms of the proposed method are affected by changing the threshold. Some of the parameters including  $C_{\max}$ ,  $b_{\exp}$ , and  $R_q$  simply widened, indicating that by moving away from the optimal point, the objective function reduces relatively monotonically. Conversely, the histograms of  $\alpha$  and  $R_s$  show major changes in the general shapes. This shape changing suggests that for a particular threshold around 16.5, a new region of good-enough parameter vectors emerges. This fact is clearly observable in the histogram of  $\alpha$  that was estimated using a threshold of 16.5, where a new peak was detected. Also, investigating the evolution of the  $R_s$  histogram, it can be seen that for thresholds greater than 16.5 that the histogram evolves to a relatively uniform distribution.

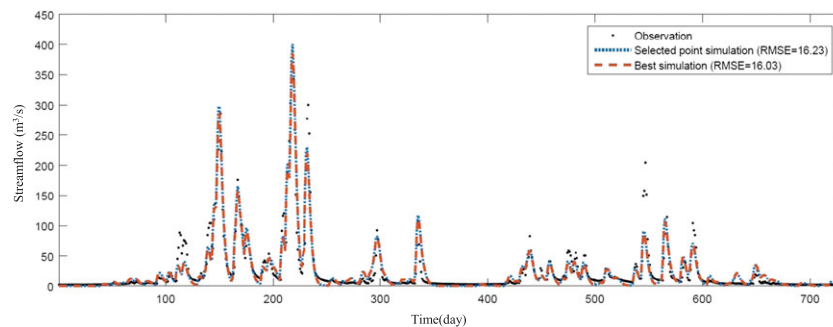
In order to make a better perception about the difference among thresholds, two individual simulations were carried out (Figure 10). One simulation using the optimum parameter vector (red) and another using a parameter vector randomly selected within the 16.5 region but out of the 16.1 (blue). The parameter values are shown as red and blue in at Figure 9. Figure 10 shows that the difference between the simulations is small. So perhaps the blue cross can be considered as an acceptable simulation. However, given the location of the points, it can be seen that despite the presence of a distinct similarity between performance of the blue (selected) and the red (optimum) point, it has limited probability in the posterior probability density calculated by the DREAM<sub>z5</sub> algorithm.

In the PRIM-PE, the next step after sample generation was to use the PRIM to define the box sequence. Table 3 shows the boundaries of the five largest boxes calculated by the PRIM for the thresholds of 16.1 and 16.5. These boxes can now be used for generating as many good-enough parameter vectors as the modeller needs with a low computational cost. Also, these boxes provide valuable information about the shape of the good-enough region. For example, by comparing the boundaries of  $R_s$  for the region with thresholds 16.1 and 16.5, it can be seen in the latter case that most of the boxes boundaries are equal or very near to their initial ranges, indicating that the model sensitivity to  $R_s$  reduces significantly if the acceptable threshold is more than 16.5.

In order to test the proposed model in a more complex problem, 10 rainfall multipliers (ranging between 0.5 and 1.5) for the major events were added to the model (similar to the previous studies from, e.g., Vrugt et al., 2008; Vrugt, Ter Braak, Diks, et al., 2009). These multipliers



**FIGURE 9** The first four upper panels show the histogram of the successful point in the resulting samples generated by the PRIM-PE and the last panel is the marginal distribution of the parameters from the DREAM<sub>ZS</sub> algorithm. Each column represents a particular parameter. The red crosses on the second panel show the global optimum point position and the blue cross (in first two parameters are overlapped with the red ones) are the location of a random parameter vector within the boundary of 16.5 but outside the 16.1 region



**FIGURE 10** Simulations of the best parameter vector (red) and a selected parameter vector (blue)

were then treated as new model parameters to increase the number of dimensions. Table 1 shows the performances of the resulting boxes.

One of the opportunities that is provided by the PRIM is the interpretability of the results. Two measures of interpretability could be defined for the box sequences, more specifically the number of boxes and the number of dimensions that are used in definition of each box. These measures can be controlled by setting the minimum acceptable density (MAD). A lower MAD leads to a higher interpretability by producing fewer boxes with higher volumes and perhaps a lower number

of dimensions. In this way, the boxes would be easier to interpret, but they would also provide a more inaccurate estimate of the boundaries. It means these boxes contain a part of the surrounding area of the good-enough region.

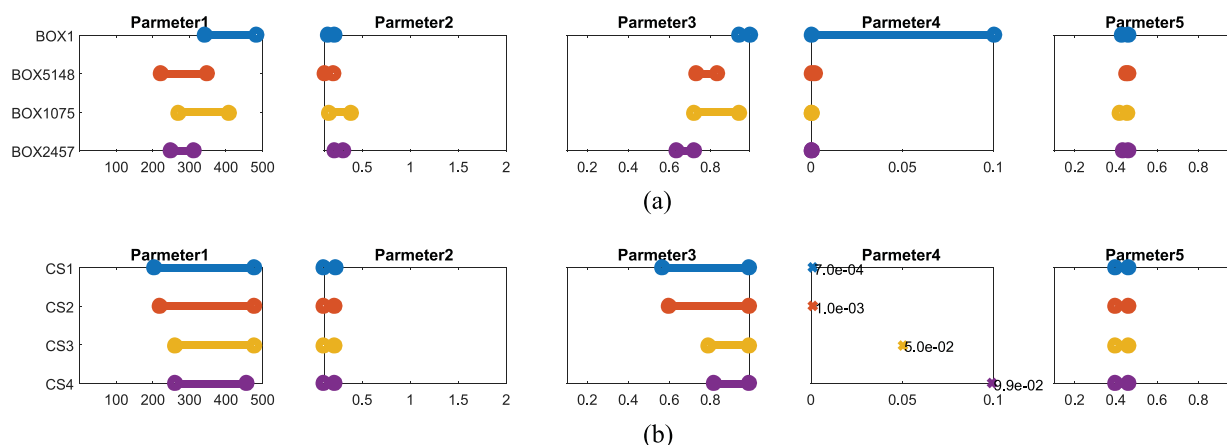
To have a better understanding about the interpretable results, four boxes of the HYMOD15 problem with  $MAD = 80\%$  are shown in Figure 11a (in order to improve readability, just five main parameters were employed). In this figure, because the region is defined as a number of boxes, each box could be shown by a number of ranges. Using



**TABLE 3** The boundaries of the five largest resulting boxes that are calculated using thresholds 16.1 and 16.5

$C_{\max}$		$b_{\exp}$		$\alpha$		$R_s$		$R_q$	
LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
Threshold = 16.1									
420	432	0.164	0.180	0.947	0.976	0.000	0.003	0.459	0.469
418	427	0.164	0.181	0.942	0.964	0.000	0.004	0.458	0.469
421	430	0.163	0.182	0.954	0.971	0.000	0.004	0.457	0.470
420	429	0.161	0.185	0.944	0.968	0.000	0.002	0.456	0.471
422	431	0.162	0.176	0.949	0.976	0.000	0.006	0.460	0.467
Threshold = 16.5									
426	500	0.150	0.205	0.959	0.990	0.000	0.100	0.449	0.479
417	500	0.143	0.214	0.942	0.990	0.000	0.100	0.457	0.468
417	450	0.148	0.195	0.903	0.990	0.006	0.083	0.453	0.480
423	500	0.141	0.209	0.969	0.990	0.014	0.100	0.450	0.473
403	435	0.154	0.205	0.941	0.990	0.000	0.100	0.444	0.468

Note. LB and UB indicate the lower bound and upper bound.

**FIGURE 11** Examples of the estimated boxes and estimation cross-section of the good-enough region from the box sequence of HYMOD15 problem with MAD = 80%. (a) Four samples of boxes from the HYMOD15 problem; (b) four cross-section of the box sequence from the HYMOD15 problem

these boxes, a clearer understanding about the good-enough parameter region and consequently about the model can be achieved. For example, it can be observed that most of the boxes had  $[0, 0.1]$  as the boundaries of Parameter4 (e.g., the Box1), which was the entire feasible domain. In these boxes, the boundaries of other parameters were limited to a particular ranges (i.e., Parameter1,  $[300, 500]$ ; Parameter2,  $[0.10, 0.22]$ ; Parameter3,  $[0.90, 1.00]$ ; Parameter5,  $[0.42, 0.48]$ ). Conversely, there were a number of other boxes with a narrow boundaries for Parameter4,  $[0, 0.001]$  (e.g., Box5148, Box1075, and Box2457), which had a wider range of other parameters (i.e., Parameter1,  $[200, 500]$ ; Parameter2,  $[0.10, 0.30]$ ; Parameter3,  $[0.55, 1.00]$ ; and Parameter5,  $[0.42, 0.48]$ ). From this analysis, it can be concluded that a low value of Parameter4 ( $R_s$ ) provides an opportunity to have wider ranges of other parameters, but for values more than 0.001, the model did not show any considerable sensitivity.

These boxes also are useful to interpret the good-enough region by providing an opportunity to make cross-sections. In other words, for a specific value of one or more parameters, the possible range of the other parameters can be determined. For example, the reaction of the model to Parameter4 (which was discussed previously) can be

explained with the analysis of the cross-sections as well. Figure 11b shows a number of cross-sections, which were determined by considering Parameter4 as constant. As can be seen, for low values of Parameter4 (less than 0.001), a wider range of the other parameter is observed, and by increasing the value of Parameter4, the ranges of other parameters decrease until a particular range after which they remain constant.

In summary, the results of these comparisons show the ability of the PRIM-PE method in the detection of the good-enough region. Producing a sample with the posterior distribution was not the aim of this study. Instead, the PRIM-PE method provides a broader understanding about the sensitivity of the model to its parameters by providing an opportunity to simplify the resulting good-enough region by automatically reducing the dimension of the region. Furthermore, the MCMC methods do not usually use a threshold in the sampling procedure, and the density of the points in the sample is proportional to their target distribution. It means that the density on the border of the good-enough region is often lower than in the central parts. By increasing the threshold, this rapidly decreases. Therefore, although the resulting sample from the MCMC methods can find the posterior distribution efficiently,

they are probably not efficient at good-enough region border analysis. The border analysis can provide an area in which most of the points there have at least a minimum performance, and if that region is interpretable, it provides a good insight into the parameter space and consequently the model.

## 4 | SPATIALLY DISTRIBUTED HYDROLOGICALLY DISTRIBUTED MODEL DEMONSTRATION

### 4.1 | Introduction

The spatially distributed hydrological model used in this study was the Australian Water Resources Assessment-Land (AWRA-L), which is a part of the AWRA system (Van Dijk, 2010). This model was jointly developed by the Australian Commonwealth Scientific and Industrial

Research Organisation and the Bureau of Meteorology. AWRA-L (version 0.5) is a grid-distributed model that simulates water flows in the surface, shallow and deep soil, groundwater, and vegetation. AWRA-L has relatively few parameters compared with fully process-based spatial hydrological models (which can have dozens of parameters per grid cell) and has been proven to work well in Australia (Renzullo et al., 2014; van Dijk, Peñma-Arancibia, & Bruijnzeel, 2012; van Dijk, Renzullo, & Rodell, 2011).

This model consists of a number of grids, with each grid including two (adjustable to any integer number) hydrological response units (HRUs), representing the deep- and shallow-rooted vegetation areas. Soil and vegetation water and heat fluxes are simulated separately for each HRU. The HRUs have the same ground water storage, but each HRU has its own parameter vectors and state variables. Table 4 shows the main parameter vectors and their proposed values. In this model, the soil water content is partitioned into three stores: (a) the surface

**TABLE 4** Simulation model parameters list, default, minimum, maximum value, and the sensitivity of the parameters

Name	Default		Max		Min		Sensitivity	
	HRU1	HRU2	HRU1	HRU2	HRU1	HRU2	Mean score	Rank
alb_dry	0.26	0.26	0.52	0.52	0.13	0.13	0	24
alb_wet	0.16	0.16	0.32	0.32	0.08	0.08	0	24
beta	4.5	4.5	8.4	8.4	0.7	0.7	3	1
cGsmax	0.03	0.03	0.06	0.06	0.015	0.015	0.29	17
ER_frac_ref	0.2	0.05	0.25	0.125	0.05	0.025	0.05	21
Fw_conn	1	1	1	1	1	1	0	24
FsoilE <sub>max</sub>	0.2	0.5	0.9	0.9	0.2	0.2	0.67	12
fvegref_G	0.15	0.15	0.3	0.3	0.05	0.05	0	24
FwaterE	0.7	0.7	0.8	0.8	0.6	0.6	0	24
Gfrac_max	0.3	0.3	0.5	0.5	0.2	0.2	0	24
hveg	10	0.5	15	1	5	0.2	0	24
InitLoss	5	5	10	10	0	0	1.05	9
LAl <sub>max</sub>	8	8	10	10	7	7	0	24
LAl <sub>ref</sub>	2.5	1.4	3	2	2	1	0	24
PrefR	150	150	2,000	2,000	100	100	1.29	7
S_sls	0.1	0.1	0.6	0.6	0.05	0.05	0.43	14
S0FC	30	30	60	60	15	15	0.05	21
SdFC	1,000	1,000	2,000	2,000	500	500	2.43	3
SsFC	200	200	400	400	100	100	0	24
SLA	3	10	9	30	1.5	5	2.86	17
Throw	1,000	150	2,000	30	500	7.5	1.43	19
Tsenc	60	10	120	20	30	5	0	24
Ud0	4	0	8	0	2	0	0.43	14
Us0	6	6	10	10	1	1	1.14	8
Vc	0.35	0.65	0.45	1	0.3	0.5	0.14	19
w0limE	0.85	0.85	0.9	0.9	0.6	0.6	0	24
w0ref_alb	0.3	0.3	0.35	0.35	0.19	0.19	0	24
wlimU	0.3	0.3	0.5	0.5	0.15	0.15	1.81	6
wslimU	0.3	0.3	0.5	0.5	0.15	0.15	0.05	21
Sgref-scale	8.15	8.15	16.3	16.3	-1.63	-1.63	2.76	2
Sgref_shape	2.34	2.34	4.68	4.68	-0.468	-0.468	2.05	5
FdrainFC_scale	0.0685	0.0685	0.137	0.137	0.01	0.01	1.05	9
FdrainFC_shape	3.179	3.179	4	4	-7	-7	2.29	4
K_gw_scale	0.047	0.047	0.094	0.094	-0.047	-0.047	0.52	13
K_gw_shape	-0.0508	-0.0508	-0.1016	-0.1016	0.01016	0.01016	0	24
K_rout_scale	0.141	0.141	0.282	0.282	0.01	0.01	0.76	11
K_rout_int	0.284	0.284	0.568	0.568	0.01	0.01	0.43	14
Three-parameter problem			23-parameter problem			Insensitive parameters		

Note. HRU = hydrological response unit.

soil layer, (b) a shallow layer that both shallow-rooted and deep-rooted vegetation have access to, and (c) a deep layer where just deep-rooted vegetation extracts water. There is also a ground water storage, which is modelled using a simple linear storage modelled without any lateral flow. Table 4 shows the list of AWRA-L parameters with their default values and the absolute ranges, selected based on recommended values by Van Dijk (2010) and a trial and error process.

## 4.2 | Test area and data set

The Murrumbidgee Basin covers 8% of the Murray–Darling Basin (87,283 km<sup>2</sup>) and is located in southern New South Wales, Australia.

Because there are many gauged catchments in the eastern half of the basin and the stream flow is considerably less regulated because of irrigation requirements (unlike the western half), the eastern half (21 catchments) was studied, and the results of the calibration for a sample catchment (Adjungbilly Dbalara) were presented here (Figure 12). The yearly average rainfall is 530 mm with a fairly uniform temporal distribution throughout the year, but a strong east–west gradient from 2,000 to less than 300 mm. This rainfall generates an average of 53 mm of run-off per year over the basin. The water availability of the basin is 4,270 GL/year of which 10% is an inter-basin transfer from the Snowy Mountains Hydro-electric Scheme (CSIRO, 2008).

The model input includes climate data (minimum and maximum temperature, rainfall, and radiation) from the Terrestrial Ecosystem Research Network ecosystem modelling and scaling infrastructure from 2007 to 2011 at 0.01-degree resolution; a land cover map (Lymburner, Geoscience Australia, & Australian Bureau of Agricultural and Resource Economics, 2011), which was originally at 250 m resolution and upscaled to a 0.01 degree resolution; and observed stream-flow data for validation and calibration of the model and final results (available online at <http://realtimedata.water.nsw.gov.au/water.stm>).

The data from 2007 to 2009 were used to initialize the model, whereas data from 2009 and 2010 were used for parameter estimation and from 2010 to 2011 were used for validation. While comparatively short, the validation results suggest that the length of the calibration period was adequate.

## 5 | RESULTS

### 5.1 | Sensitivity analysis

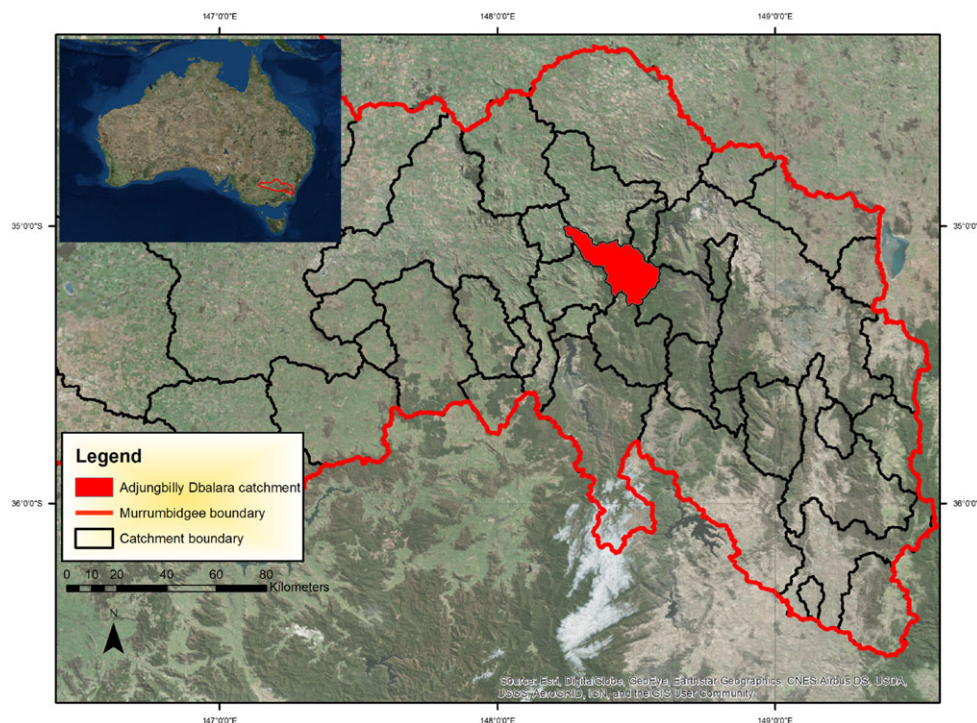
A sensitivity analysis was conducted for each catchment (21 catchments in the upper Murrumbidgee) and a total of 37 unknown parameters. For each catchment, parameters were divided into four categories on the basis of the objective function degradation that they caused; after which, a score of 0 to 3 was assigned to each parameter. These sensitivity scores were assigned on the basis of Table 5.

A sample of the sensitivity analysis for a single catchment (Adjungbilly Dbalara) is presented in Figure 13. The average of the scores for all catchments shows the overall sensitivity of each parameter with the parameter sensitivity rank presented in Table 4.

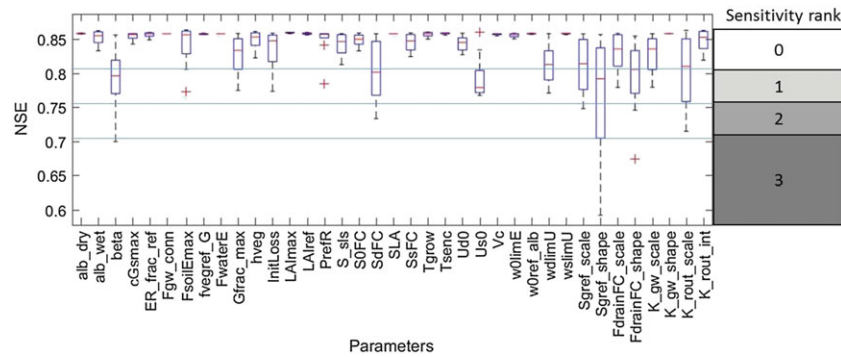
After the sensitivity analysis, two calibration problem set-ups were solved. First, a three-parameter calibration problem for one of the

**TABLE 5** Sensitivity analysis score assigning rules

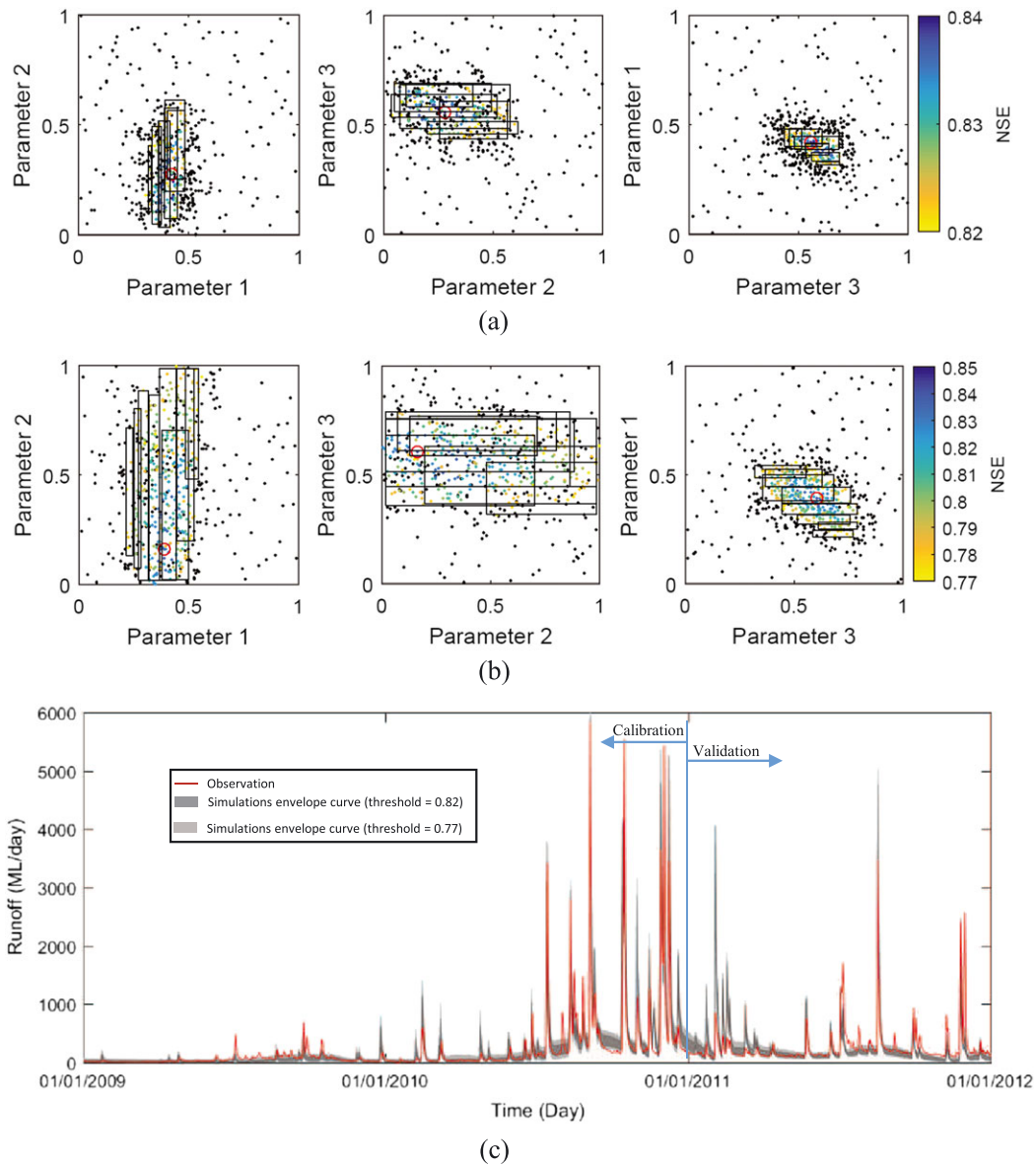
Amount of degradation in objective function	Corresponding sensitivity score
0.00 to 0.05	0
0.05 to 0.10	1
0.10 to 0.15	2
More than 0.15	3



**FIGURE 12** Murrumbidgee Basin location and its eastern catchments



**FIGURE 13** Sensitivity analysis of parameters for the Adjungbilly Dbalara catchment. NSE = Nash-Sutcliffe efficiency



**FIGURE 14** The resulting sequence of boxes for the three-parameter problem (threshold = 0.82) and simulation of a randomly generated sample in the boxes. (a) Boxes sequence (threshold = 0.82); (b) boxes sequence (threshold = 0.77); (c) simulation of random generated point inside boxes (30 points)

catchments was solved to provide a visualization opportunity to show different aspects of the results. Then, a 23-parameter problem with the sensitive parameters (the parameters with average sensitivity score more than zero) was undertaken.

## 5.2 | Three-parameter calibration

According to Table 4, the three most sensitive parameters are beta, SdFC and Sgref-scale (in this section, they are referred to as Parameters



1 to 3, respectively). Results were obtained using a total of 600 simulations including 100 initial samples and 500 generated samples in 10 iterations. Two threshold values, 5% and 10% less than the best obtained objective function, were used to solve the problem. The maximum value obtained from genetic algorithm optimization was 0.86, so the thresholds of 0.82 (5% less than 0.86) and 0.77 (10% less than 0.86) were selected. As a first step, points were scattered uniformly in the domain, and gradually, the density around the successful region increased. In the first case (threshold = 0.82), due to the smaller successful region, no successful point was generated in the initial step, and even the best point failed to perform better than the threshold, and successful points were produced in the next generations. Conversely, in the second case (threshold = 0.77), a number of successful points were detected in the first iteration, and their number increased during the next generation. Figure 14a,b show the final populations of both cases. By comparing them, it can be seen that the density of points in the first case was higher than in the second case because the same amount of points were scattered in a smaller region. So the stricter threshold led to a denser successful region.

The best solution (the red circle) had a relatively better performance at each step. The successful region can also be considered as a bound that every point inside that is likely to be a behavioural parameter vector. Also, the size of this region can be used as a relative measure of equifinality. This means that a bigger size of the successful region can be used as an indicator of the relative degree of equifinality.

Tables 6 and 7 summarize the boundaries of the sequences of boxes for a threshold equal to 0.82 (6 boxes) and 0.77 (8 boxes), respectively. Figure 14a,b presents the boxes visually. These boxes cover all successful points and no failure point (according to the defined criteria of density = 100%). Figure 14c shows the envelope curves of the

run-off simulations for a number of samples (five samples in each box), which were randomly generated. It can be seen that all simulations had good-enough model performances. By comparing the results from the first and second case, it can be seen that reducing the threshold caused a wider simulation bound. It should be mentioned that, because the size of the boxes was not equal and there were overlaps between the boxes, the sampling can be done in a way that the probability of selecting each point in the good-enough region becomes uniform.

### 5.3 | Twenty-three-parameter calibration

A similar process to that conducted for the three-parameter problem was carried out for the 23-parameter problem. A 10,000 sample size was chosen for this problem on the basis of examination and preliminary calibration results. The sample was generated using an initial population of 5,000 points and 100 iterations, with 50 new points being generated in each iteration. The threshold used for this problem was 5% less than the optimum value (0.82). In this case, the criteria for generating the boxes was a density equal to 100% (similar to the three-parameter problem). The resulting sequence consisted of 107 boxes. Figure 15 shows an envelope curve of 214 simulations for different parameter vectors randomly generated inside the boxes (2 points in each box).

After finding the good-enough region, the validity of sensitivity analysis result was tested using a parameter ensemble that was randomly generated from that region. The results showed that the overall order of the sensitive parameters was similar to the order calculated using an optimal parameter vector as the centre point. Every sensitive parameter from the initial analysis remained sensitive and only a few insensitive parameters (alb-wet, fvegreg-G, Gfrac-max, hveg, LAlmax, Tsenc, and wOlimE) showed limited sensitivity (average score less than 0.45).

**TABLE 6** Sequence of boxes for the three-parameter problem and threshold = 0.82

Box	Parameter 1		Parameter 2		Parameter 3	
	Min	Max	Min	Max	Min	Max
1	0.403	0.481	0.255	0.566	0.435	0.608
2	0.393	0.483	0.199	0.613	0.458	0.515
3	0.390	0.447	0.074	0.577	0.482	0.684
4	0.364	0.413	0.035	0.518	0.533	0.639
5	0.335	0.374	0.048	0.493	0.560	0.689
6	0.318	0.333	0.104	0.408	0.583	0.687

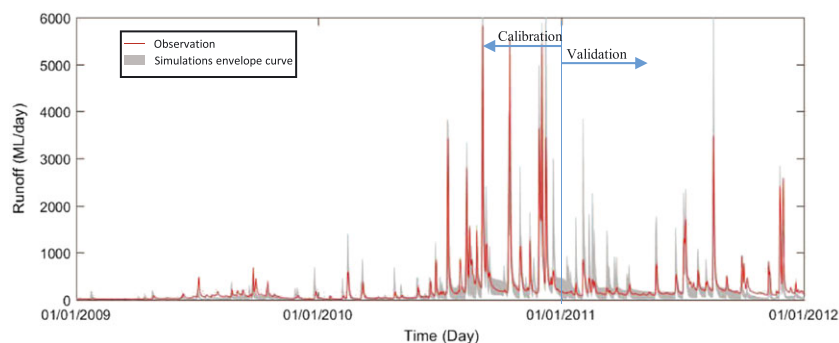
**TABLE 7** Sequence of boxes for the three-parameter problem and threshold = 0.77

Box	Parameter 1		Parameter 2		Parameter 3	
	Min	Max	Min	Max	Min	Max
1	0.446	0.525	0.198	0.985	0.367	0.630
2	0.484	0.544	0.481	0.985	0.318	0.556
3	0.379	0.499	0.019	0.703	0.357	0.683
4	0.368	0.444	0.018	0.985	0.445	0.758
5	0.318	0.366	0.018	0.864	0.445	0.790
6	0.272	0.316	0.018	0.884	0.514	0.757
7	0.250	0.281	0.072	0.801	0.611	0.790
8	0.213	0.245	0.129	0.714	0.587	0.769

## 6 | DISCUSSION

The goal of this paper was to develop a methodology (PRIM-PE) to provide an interpretable understanding of the model performance in parameter space to help modellers understand its global sensitivity and provide a broad understanding about the parameter vectors that have an acceptable performance. The PRIM-PE methodology uses a sequence of steps to find a good-enough region in the parameter space to provide a better perspective about the parameter space. The first step of the procedure is a sampling routine, which makes an estimate about the response surface using a limited sample size that is focused on the good-enough region. After the sampling, the resulting sample is used as input to PRIM to evaluate the sequences of boxes as representative of the good-enough region.

There are different components in the sampling procedure designed to improve its performance. Two of the most important are as follows: (a) the sorting strategy, which forces the sampling method to generate a sample uniformly within the good-enough region and a narrow buffer to reveal the boundaries as clear as possible and provide the opportunity for the PRIM to detect the good-enough region accurately and (b) the failed parent search component, which helps the sampling method to not be trapped in a local optimum, prevent premature convergences, and provides the capability to capture multiple optima.



**FIGURE 15** Simulation of a randomly generated sample in the boxes for the 23-parameter problem

To test the capability of the PRIM-PE method, four test problems were used. The first three problems were used to test the performance of the PRIM-PE in different situations (e.g., curved and multiple optima good-enough regions), and the sampling procedure was benchmarked against the DREAM<sub>ZS</sub>. The results for all three problems showed that the PRIM-PE method effectively generated samples with a higher density in the good-enough regions and the surrounding area to provide a good definition of the border of the good-enough region. The achieved focus of points saves a considerable amount of computation effort that would be needed for simulating unnecessary points far from the intended region. The PRIM results also demonstrate that it can define the good-enough region using a set of boxes. Moreover, benchmarking of the PRIM-PE against the DREAM<sub>ZS</sub> algorithm using the fourth test problem (HYMOD parameter estimation) showed that with a threshold close to the optimum objective function, the generated samples were similar, but by increasing the value of the threshold, the good-enough region (estimated by the PRIM-PE method) widened. In all the DREAM<sub>ZS</sub> runs, prior PDFs were considered to be uniform distributions. Thus, it can be expected that informed prior PDFs would improve the comparison. However, to define a proper prior and likelihood requires a good understanding of the system as well as rich technical knowledge, which complicate the use of the MCMC methods.

After sampling, the PRIM was successfully used to define the box sequence as a representative of the good-enough region. The benefit of finding such a region instead of the conventional results demonstrates that this region is a continuous area potentially containing just every possible combination of good-enough points. This method does not require predefined parameter ranges in addition to a set of rules to separate the good-enough points from the other points normally defined as measures of the objective function and their thresholds (similar to the GLUE). By finding the entire good-enough region, if a particular set of rules was used, adding a new rule is computationally much cheaper than executing the entire process again, because the previously calculated good-enough region can be used as the new searching domain for the new rule. Moreover, the collection of parameter vectors produced by other methods could be refined using additional rules. In this case, if the number of successful points within the collection is limited, an additional procedure of generating samples is needed. However, given the fact that this study has been performed using just a single rule, future work could address the effects of different objective functions and more rules.

The good-enough region can be considered as a confidence boundary of the optimal parameter vector, with each point inside being a

possible effective solution. Moreover, the total size of the resulting sequence of boxes from the PRIM-based method can be used as a measure of equifinality. As all parameter vectors located in the sequence of boxes share a minimum performance, they can be useful for various problems. One application of these boxes is for estimating the parameter vector in an ungauged catchment. If a parameter vector has been categorized as acceptable in a number of surrounding catchments, it can possibly be a good estimation of the parameter in the ungauged catchment. Having every possible parameter vector for a number of nearby catchments allows using a parameter from the intersection of the other catchments' good-enough region as a better choice for the ungauged catchment. In this way, the knowledge from a number of catchments can be combined to estimate a more reliable parameter vector for the ungauged catchment. Overlapping of the good-enough regions is also useful in other problems, for example, when for a specific location, two series of observations are available. This may occur when a river has two gauging stations. For the upper catchment, two sequences of boxes could be obtained. By overlapping the sequences of boxes for the doubly gauged area, a smaller (less uncertain) good-enough region would be found. Furthermore, these boxes can be used as an input for data assimilation methods such as the Ensemble Kalman Filter.

The PRIM-PE could also be used as a prior estimate of the search region for other MCMC methods. Because the PRIM-PE can work with a limited sample size, but with a commensurate decrease in accuracy, it can roughly estimate the good-enough region boundaries. These rough estimates can assist other MCMC methods to search the parameter region more efficiently.

It should be acknowledged that the PRIM-PE has a number of limitations. First, the process of finding a suitable cut-off threshold demands a prior knowledge about the optimum objective function value. Second, the PRIM-PE is not designed to work with unconstrained search spaces. Third, the PRIM-PE is designed to solve hydrological parameter estimation problems with a limited number of parameters (maximum around 20–25 parameters). Therefore, its performance was tested for this type of problems. The performance of the PRIM-PE in higher dimensional problems will be the subject of future studies.

It should be emphasized that the main objective of the PRIM-PE is not to find the target distribution. The PRIM provides an opportunity to simplify the resulting good-enough region by defining a number of boxes and automatically reducing the dimension of the region as much as possible. Furthermore, it generates an area that one can be almost certain that every point in that area has at least a minimum perfor-

mance, and if that region is interpretable, it can provide a good insight into the parameter area and consequently the model. Conversely, as the MCMC methods do not usually use a threshold in the sampling procedure, the density of the points in the sample is proportional to their target distribution. This means that the density on the border is often less than central parts and by increasing the threshold it rapidly decreases. Therefore, even though the resulting sample can address the posterior distribution efficiently, it is not expected to be efficient at border analysis.

## 7 | CONCLUSIONS

A new parameter estimation and uncertainty quantification method based on the PRIM, called PRIM-PE, has been proposed. The aim of this method is to find a range in parameter space that covers every acceptable parameter vector. The PRIM-PE method consists of two main parts. First, an elegant sampling procedure is used to explore the parameter space efficiently to generate samples focused on the good-enough region and the surrounding area. The proposed sampling method starts with an initially completely random generation, after which the new generations are produced mostly in the regions with good-enough performances. Then, the PRIM is used to define the good-enough region in the parameter space as a sequence of boxes. The method was tested using four different test problems and its sampling procedure compared with the DREAM<sub>25</sub>. After that a hydrological model (AWRA-L) was calibrated using two settings (the three-parameter and 23-parameter calibration problems). The results showed that the PRIM-PE method can efficiently capture the good-enough region.

## ACKNOWLEDGMENTS

Ashkan Shokri acknowledges the support from the Monash University in form of a Monash Graduate Scholarship (MGS) and Monash International Postgraduate Research Scholarship (MIPRS). Valentijn Pauwels is supported through ARC Future Fellowship Grant FT130100545. The research is supported through ARC Grant DP140103679. This research is also supported in part by the Monash eResearch Centre and eSolutions-Research Support Services through the use of the Monash Campus HPC Cluster.

## ORCID

Ashkan Shokri  <http://orcid.org/0000-0002-4925-7937>

## REFERENCES

- Ajami, N. K., Duan, Q., & Sorooshian, S. (2007). An integrated hydrologic bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resources Research*, 43(1), W01403. <https://doi.org/10.1029/2005WR004745>.
- Arkesteijn, L., & Pande, S. (2013). On hydrological model complexity, its geometrical interpretations and prediction uncertainty. *Water Resources Research*, 49(10), 7048–7063.
- Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area hydrologic modeling and assessment part i: Model development.
- Bardsley, W. E., Vetrova, V., & Liu, S. (2015). Toward creating simpler hydrological models: A lasso subset selection approach. *Environmental Modelling & Software*, 72, 33–43.
- Bates, B. C., & Campbell, E. P. (2001). A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall–runoff modeling. *Water Resources Research*, 37, 937–947. <https://doi.org/10.1029/2000WR900363>.
- Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, 16(1), 41–51.
- Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, 6(3), 279–298.
- Blasone, R.-S., Vrugt, J. A., Madsen, H., Rosbjerg, D., Robinson, B. A., & Zyvoloski, G. A. (2008). Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov chain Monte Carlo sampling. *Advances in Water Resources*, 31(4), 630–648.
- Borah, D. K., Xia, R., & Bera, M. (2002). DWSM - A dynamic watershed simulation model. In V. P. Singh & D. K. Freyert (Eds.), *Mathematical model for small watershed hydrology* (pp. 113–166). Highlands Ranch, Colorado: Water Resources Publications, LLC.
- Boyle, D. P., Gupta, H. V., & Sorooshian, S. (2000). Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resources Research*, 36(12), 3663–3674.
- Bryant, B. (2009). sdtoolkit: Scenario discovery tools to support robust decision making.
- Bryant, B. P., & Lempert, R. J. (2010). Thinking inside the box: A participatory, computer-assisted approach to scenario discovery. *Technological Forecasting and Social Change*, 77(1), 34–49.
- Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., ... Rasmussen, R. M. (2015). A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resources Research*, 51(4), 2498–2514.
- Diodato, N., Brocca, L., Bellocchi, G., Fiorillo, F., & Guadagno, F. M. (2014). Complexity-reduction modelling for assessing the macro-scale patterns of historical soil moisture in the euro-mediterranean region. *Hydrological Processes*, 28(11), 3752–3760.
- Dooge, J. C. (1997). Searching for simplicity in hydrology. *Surveys in Geophysics*, 18(5), 511–534.
- Duan, Q., Ajami, N. K., Gao, X., & Sorooshian, S. (2007). Multi-model ensemble hydrologic prediction using bayesian model averaging. *Advances in Water Resources*, 30(5), 1371–1386.
- Efstratiadis, A., & Koutsoyiannis, D. (2010). One decade of multi-objective calibration approaches in hydrological modelling: A review. *Hydrological Sciences Journal*, 55(1), 58–78.
- Famiglietti, J. S., & Wood, E. F. (1994). Multiscale modeling of spatially variable water and energy balance processes. *Water Resources Research*, 30(11), 3061–3078.
- Fenicia, F., Kavetski, D., & Savenije, H. H. G. (2011). Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development. *Water Resources Research*, 47(11), W11510. <https://doi.org/10.1029/2010WR010174>.
- Franks, S. W., & Beven, K. J. (1997). Bayesian estimation of uncertainty in land surface–atmosphere flux predictions. *Journal of Geophysical Research: Atmospheres*, 102(D20), 23991–23999.
- Freer, J., Beven, K., & Ambrose, B. (1996). Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the glue approach. *Water Resources Research*, 32(7), 2161–2173.
- Friedman, J. H., & Fisher, N. I. (1999). Bump hunting in high-dimensional data. *Statistics and Computing*, 9(2), 123–143.
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472.
- Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. New York, NY: Addison-Wesley Pub. Co, Reading, Mass.
- Gupta, H. V., Sorooshian, S., Hogue, T. S., & Boyle, D. P. (2013). Advances in automatic calibration of watershed models.
- Haario, H., Saksman, E., & Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, 7(2), 223–242.

- Haario, H., Laine, M., Mira, A., & Saksman, E. (2006). DRAM: Efficient adaptive MCMC. *Statistics and Computing*, 16(4), 339–354.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97–109.
- Hill, M. C. (2006). The practical use of simplicity in developing ground water models. *Ground Water*, 44(6), 775–781.
- Hsu, K.-I., Moradkhani, H., & Sorooshian, S. (2009). A sequential Bayesian approach for hydrologic model selection and prediction. *Water Resources Research*, 45(12), W00B12. <https://doi.org/10.1029/2008WR006824>.
- Iorgulescu, I., Beven, K. J., & Musy, A. (2005). Data-based modelling of runoff and chemical tracer concentrations in the haute-montue research catchment (switzerland). *Hydrological Processes*, 19(13), 2557–2573.
- Jia, Y., & Culver, T. B. (2008). Uncertainty analysis for watershed modeling using generalized likelihood uncertainty estimation with multiple calibration measures. *Journal of Water Resources Planning and Management*, 134(2), 97–106.
- Keating, E. H., Doherty, J., Vrugt, J. A., & Kang, Q. (2010). Optimization and uncertainty assessment of strongly nonlinear groundwater models with high parameter dimensionality. *Water Resources Research*, 46(10), W10517. <https://doi.org/10.1029/2009WR008584>.
- Kuczera, G. (1983). Improved parameter inference in catchment models: 1. Evaluating parameter uncertainty. *Water Resources Research*, 19(5), 1151–1162.
- Kuczera, G., & Parent, E. (1998). Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm. *Journal of Hydrology*, 211(1–4), 69–85.
- Laloy, E., & Vrugt, J. A. (2012). High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(zs) and high-performance computing. *Water Resources Research*, 48(1), W01526. <https://doi.org/10.1029/2011WR010608>.
- Laloy, E., Rogiers, B., Vrugt, J. A., Mallants, D., & Jacques, D. (2013). Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov chain Monte Carlo simulation and polynomial chaos expansion. *Water Resources Research*, 49(5), 2664–2682.
- Legates, D. R., & McCabe, G. J. (1999). Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, 35(1), 233–241.
- Lempert, R. J. (2002). A new decision sciences for complex systems. *Proceedings of the National Academy of Sciences*, 99(suppl 3), 7309–7313.
- Lempert, R. J., Popper, S. W., & Banks, S. C. (2003). Shaping the next one hundred years: New methods for quantitative, long-term policy analysis / Robert J. Lempert, Steven W. Popper, Steven C. Banks. RAND and [Lancaster: Gazelle] [distributor], Santa Monica.
- Lempert, R. J., Warren, R., Henry, R., Button, R. W., Klenk, J., & Giglio, K. (2016). *Defense resource planning under uncertainty: An application of robust decision making to munitions mix planning*. Santa Monica, California: Rand Corporation.
- Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research: Atmospheres*, 99(D7), 14415–14428.
- Lymburner, L., Geoscience Australia, & Australian Bureau of Agricultural and Resources Economics (2011). *The national dynamic land cover dataset*. Geoscience Australia: Symonston, A.C.T.
- Malama, B., Kuhlman, K. L., & James, S. C. (2013). Core-scale solute transport model selection using Monte Carlo analysis. *Water Resources Research*, 49(6), 3133–3147.
- Marshall, L., Nott, D., & Sharma, A. (2004). A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling. *Water Resources Research*, 40(2), W02501. <https://doi.org/10.1029/2003WR002378>.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6), 1087–1092.
- Miller, B. L., & Goldberg, D. E. (1995). Genetic algorithms, tournament selection, and the effects of noise. *Complex systems*, 9(3), 193–212.
- Refsgaard, J. C., Storm, B. (1995). MIKE SHE. In V.P. Singh (Ed.), *Computer models of watershed hydrology* (pp. 809–846). Highlands Ranch, Colorado.
- Renzullo, L. J., van Dijk, A. I. J. M., Perraud, J. M., Collins, D., Henderson, B., Jin, H., ... McJannet, D. L. (2014). Continental satellite soil moisture data assimilation improves root-zone moisture analysis for water resources assessment. *Journal of Hydrology*, 519, Part D(0), 2747–2762.
- Samanta, S., Clayton, M. K., Mackay, D. S., Kruger, E. L., & Ewers, B. E. (2008). Quantitative comparison of canopy conductance models using a Bayesian approach. *Water Resources Research*, 44(9), W09431. <https://doi.org/10.1029/2007WR006761>.
- Schoups, G., & Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resources Research*, 46(10), W10531. <https://doi.org/10.1029/2009WR008933>.
- Schoups, G., van de Giesen, N. C., & Savenije, H. H. G. (2008). Model complexity control for hydrologic prediction. *Water Resources Research*, 44(12), W00B03. <https://doi.org/10.1029/2008WR006836>.
- Shafii, M., Tolson, B., & Matott, L. S. (2014). Uncertainty-based multi-criteria calibration of rainfall-runoff models: A comparative study. *Stochastic environmental research and risk assessment*, 28(6), 1493–1510.
- Shortridge, J. E., & Guikema, S. D. (2016). Scenario discovery with multiple criteria: An evaluation of the robust decision-making framework for climate change adaptation. *Risk Analysis*, 36, 2298–2312. <https://doi.org/10.1111/risa.12582>.
- Sivakumar, B. (2008). Dominant processes concept, model simplification and classification framework in catchment hydrology. *Stochastic Environmental Research and Risk Assessment*, 22(6), 737–748.
- Sivapalan, M., Zhang, L., Vertessy, R., & Blöschl, G. (2003). Downward approach to hydrological prediction. *Hydrological Processes*, 17(11), 2099. <https://doi.org/10.1002/hyp.1426>.
- Smith, T., Sharma, A., Marshall, L., Mehrotra, R., & Sisson, S. (2010). Development of a formal likelihood function for improved Bayesian inference of ephemeral catchments. *Water Resources Research*, 46(12), W12551. <https://doi.org/10.1029/2010WR009514>.
- Smith, T., Marshall, L., & Sharma, A. (2015). Modeling residual hydrologic errors with Bayesian inference. *Journal of Hydrology*, 528, 29–37.
- Smith, T. J., & Marshall, L. A. (2008). Bayesian methods in hydrologic modeling: A study of recent advancements in Markov chain Monte Carlo techniques. *Water Resources Research*, 44(12), W00B05. <https://doi.org/10.1029/2007WR006705>.
- Sorooshian, S., & Dracup, J. A. (1980). Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Correlated and heteroscedastic error cases. *Water Resources Research*, 16, 430–442. <https://doi.org/10.1029/WR016i002p00430>.
- Stedinger, J. R., Vogel, R. M., Lee, S. U., & Batchelder, R. (2008). Appraisal of the generalized likelihood uncertainty estimation (glue) method. *Water Resources Research*, 44(12), W00B06. <https://doi.org/10.1029/2008WR006822>.
- Ter Braak, C. J. F. (2006). A Markov chain Monte Carlo version of the genetic algorithm differential evolution: Easy Bayesian computing for real parameter spaces. *Statistics and Computing*, 16(3), 239–249.
- Ter Braak, C. J. F., & Vrugt, J. A. (2008). Differential evolution Markov chain with snooker updater and fewer chains. *Statistics and Computing*, 18(4), 435–446.
- Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S. W., & Srikanthan, S. (2009). Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis. *Water Resources Research*, 45(12), W00B14. <https://doi.org/10.1029/2008WR006825>.
- Tonkin, M. J., & Doherty, J. (2005). A hybrid regularized inversion methodology for highly parameterized environmental models. *Water Resources Research*, 41(10), W10412. <https://doi.org/10.1029/2005WR003995>.
- Van Dijk, A. I. J. M. (2010). Landscape model (version 0.5) technical description, AWRA Tech. Rep. 3, WIRADA/CSIRO Water for a Healthy Country Flagship, Canberra.



- Van Dijk, A. I. J. M., Peñma-Arancibia, J. L., & Bruijnzeel, L. A. (2012). Land cover and water yield: Inference problems when comparing catchments with mixed land cover. *Hydrology and Earth System Sciences*, 16(9), 3461–3473.
- Van Dijk, A. I. J. M., Renzullo, L. J., & Rodell, M. (2011). Use of gravity recovery and climate experiment terrestrial water storage retrievals to evaluate model estimates by the Australian water resources assessment system. *Water Resources Research*, 47(11), W11524. <https://doi.org/10.1029/2011WR010714>.
- Viney, N. R., Bormann, H., Breuer, L., Bronstert, A., Croke, B., Frede, H., ... Willems, P. (2009). Assessing the impact of land use change on hydrology by ensemble modelling (luchem) ii: Ensemble combinations and predictions. *Advances in Water Resources*, 32(2), 147–158.
- Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, 75, 273–316.
- Vrugt, J. A., Gupta, H. V., Bouten, W., & Sorooshian, S. (2003). A shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, 39(8), 1–14. <https://doi.org/10.1029/2002WR001642>.
- Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., & Robinson, B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research*, 44(12), W00B09. <https://doi.org/10.1029/2007WR006720>.
- Vrugt, J. A., Ter Braak, C., Diks, C., Robinson, B. A., Hyman, J. M., & Higdon, D. (2009). Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, 10(3), 273–290.
- Vrugt, J. A., ter Braak, C. J. F., Gupta, H. V., & Robinson, B. A. (2009). Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? *Stochastic Environmental Research and Risk Assessment*, 23(7), 1011–1026.
- Wöhling, T., & Vrugt, J. A. (2011). Multiresponse multilayer vadose zone model calibration using Markov chain Monte Carlo simulation and field water retention data. *Water Resources Research*, 47(4), W04510. <https://doi.org/10.1029/ignorespaces2010WR009265>.
- Wigmosta, M. S., Vail, L. W., & Lettenmaier, D. P. (1994). A distributed hydrology-vegetation model for complex terrain. *Water Resources Research*, 30(6), 1665–1679.
- Yapo, P. O., Gupta, H. V., & Sorooshian, S. (1998). Multi-objective global optimization for hydrologic models. *Journal of Hydrology*, 204(1–4), 83–97.

**How to cite this article:** Shokri A, Walker JP, van Dijk AIJM, Wright AJ, Pauwels VRN. Application of the patient rule induction method to detect hydrologic model behavioural parameters and quantify uncertainty. *Hydrological Processes*. 2018;1–21. <https://doi.org/10.1002/hyp.11464>