Pareto-optimality and a search for robustness: choosing solutions with desired properties in objective space and parameter space

Gift Dumedah, Aaron A. Berg and Mark Wineberg

ABSTRACT

Multi-objective genetic algorithms are increasingly being applied to calibrate hydrological models by generating several competitive solutions usually referred to as a Pareto-optimal set. The Pareto-optimal set comprises non-dominated solutions at the calibration phase but it is usually unknown whether all or only a subset of non-dominated solutions at the calibration phase remains non-dominated at the validation phase. In practice, users would like to know solutions (and their associated properties) which remain non-dominated at both the calibration and validation phases. This study investigates robustness of the Pareto-optimal set by developing a model characterization framework (MCF). The MCF uses cluster analysis to examine the distribution of solutions in parameter space and objective space, and conditional probability to combine linkages between the distributions of solutions in both spaces. The MCF has been illustrated for calibration output generated from application of the Non-dominated Sorting Genetic Algorithm-II to calibrate the Soil and Water Assessment Tool for streamflow in the Fairchild Creek watershed in southern Ontario. Our results show that not all non-dominated solutions found at the calibration phase perform the same for different validation periods. The MCF illustrates that robust solutions – non-dominated solutions which cluster in similar locations in parameter space and objective space – performed consistently well for several validation periods.

Key words | multi-objective evolutionary algorithms, non-dominance, parameter estimation, Pareto-optimality, robustness

Gift Dumedah (corresponding author)
School of Geography and Earth Sciences,
McMaster University,
Hamilton, ON, L8S4L8,
Canada
E-mail: *dgiftman@hotmail.com*

Aaron A. Berg
Department of Geography,
University of Guelph,
Guelph, ON, N1G2W1,
Canada

Mark Wineberg
Department of Computing, and Information Science,
University of Guelph,
Guelph, ON, N1G2W1,
Canada

INTRODUCTION

Calibration of hydrological models using multi-objective genetic algorithms (MOGAs) usually generate competitive solutions which are not dominated by any other solution when compared using model evaluation objectives. These non-dominated solutions, usually called a Pareto-optimal set, allow users several decision scenarios which represent alternative trade-offs between model evaluation objectives. From a decision-making standpoint, users are interested to know whether the Pareto set has a subset of solutions which are sensitive or less sensitive to small changes in parameter values, a necessary condition in practice. Additionally, it is usually unknown whether a subset of the Pareto-optimal set remains non-dominated at both calibration and validation time periods. In such cases, users would prefer solutions that are less sensitive to model parameter perturbations, and have a high level of performance in evaluation objective space at both calibration and future time steps.

The evaluation of performance of solutions in combination with their parametric variations are generally addressed using robust optimization methods. Robust optimization methods are usually categorized into deterministic methods (Gunawan & Azarm 2004, 2005; Li *et al.* 2005) and stochastic (or probabilistic) methods (Bertsimas &

doi: 10.2166/hydro.2011.120

Thiele 2006; Chen *et al*. 2007; Goh & Sim 2010; Goh *et al*. 2010). Deterministic methods use pre-determined intervals for parameter values to optimize the objective function, whereas probabilistic techniques use statistical measures (e.g. mean and variance) to evaluate sensitivity of parametric variations before they are incorporated into optimizing the objective function. Several studies (Tsutsui & Ghosh 1997; Ray 2002; Jin & Sendhoff 2003; Gunawan & Azarm 2004, 2005; Li *et al*. 2005) have linked the performance of solutions to their parametric variations (or robustness), and others (Chen *et al*. 2007; Goh & Sim 2010) have applied decision rules to address robust optimization in linear programs.

Furthermore, the evaluation of the Pareto-optimal set to select compromise solution(s) has been recommended in the literature (Coello Coello *et al*. 2002; Marler & Arora 2004). As a result, various studies (Khu & Madsen 2005; Taboada & Coit 2006; Tang *et al*. 2006; Bekele & Nicklow 2007; Ferreira *et al*. 2007; Grierson 2008; Hejazi *et al*. 2008; Crispim & de Sousa 2009; Dumedah *et al*. 2010) have evaluated the Pareto-optimal set to select a subset of solutions using different decision criteria. But very few studies have actually investigated the sensitivity of the Pareto-optimal set to small changes in model parameter values and its performance for future time steps.

The Pareto-optimal set, like most calibration outputs, is generated by evaluating the performance of model parameter sets in objective (or decision) space. While the distribution of parameter values can be used to test the performance of solutions, the evaluation of solutions in parameter space is limited in the model calibration literature. In other words, evaluating the distribution of solutions by examining the linkages between parameter space and objective space is not a typical approach in most calibration methods.

As a result, a framework is needed to evaluate the distribution of solutions in objective space and parameter space, and to examine the linkages between the two spaces. In MOGA literature, several studies (Jin & Sendhoff 2003; Deb & Gupta 2005, 2006; Gunawan & Azarm 2005; Li *et al*. 2005; Nazemi *et al*. 2006; Goh & Sim 2010) have examined linkages between objective space and parameter space. The linkage between the distribution of solutions in parameter space and objective space has been shown to provide robust Pareto-optimal solutions (Gunawan & Azarm 2005; Li *et al*. 2005; Nazemi *et al*. 2006). The study by Nazemi *et al*. (2006) determined robust solutions by finding the overlap of solutions between a Pareto-optimal set and a set of solutions with higher posterior probability.

Deb & Gupta (2005, 2006) provided two methods of finding robust multi-objective solutions. The first method optimizes the mean effective objective value which is determined by averaging a finite set of neighboring solutions. The second method optimizes individual evaluation objectives but adds a constraint to restrict a pre-defined limiting change in objective values. This method is synonymous with the interactive approach as the acceptable change in objective values is defined using user input to guide the search.

Our study expands on these linkages by finding robust solutions and the assessment of these solutions at both calibration and validation phases. We designed a model characterization framework (MCF) to define robustness by using conditional probability to combine the distribution of solutions in parameter space and objective space. The MCF includes a specially defined indicator, a choice index for finding solutions which have small perturbations in objective space. The MCF has been illustrated for calibration output which was generated by calibrating the Soil and Water Assessment Tool (SWAT) for simulations of streamflow in the Fairchild Creek watershed in southern Ontario, Canada. SWAT was calibrated using the Non-dominated Sorting Genetic Algorithm-II (NSGA-II).

The remaining part of this paper is organized as follows. The next section describes the MCF to evaluate the distribution of non-dominated solutions in objective space and parameter space. This section also provides a description of the NSGA-II method used to calibrate SWAT for streamflow. The results and discussion section presents the NSGA-II/SWAT calibration output – a Pareto-optimal set evaluated across different simulation periods. This section also implements the MCF to choose solutions with the desired behavior from the Pareto set, which are generated using NSGA-II. The paper concludes with a discussion about the utility of the MCF to select robust solutions. The conclusion emphasizes the importance of robustness and the choice index, and the significance of evaluating the distribution of solutions in two spaces: objective space and parameter space.

METHODS

The following subsections describe the model characterization framework. Using the concept of Pareto-optimality we provide a framework to search for robust solutions by evaluating the clustering of solutions in both objective space and parameter space. The section also describes the NSGA-II method, the study area and the SWAT model used to simulate streamflow.

Model characterization framework

Here we outline the model characterization framework, shown in Figure 1, to simultaneously assess the distribution of solutions in both parameter space and objective space. The evaluation of Pareto-optimality is usually limited to objective space but we can look at commonalities that the various solutions have to get an understanding of the properties of parameter space as well.

Note that in this study 'parameter set' is synonymous with 'parameter pathway'. A parameter pathway can be visualized as parameter values in a vector linked together by an imaginary line connecting contiguous parameters in the vector. A parameter pathway emphasizes the interconnectedness of model parameter values for evaluating the quality of a solution in objective space.

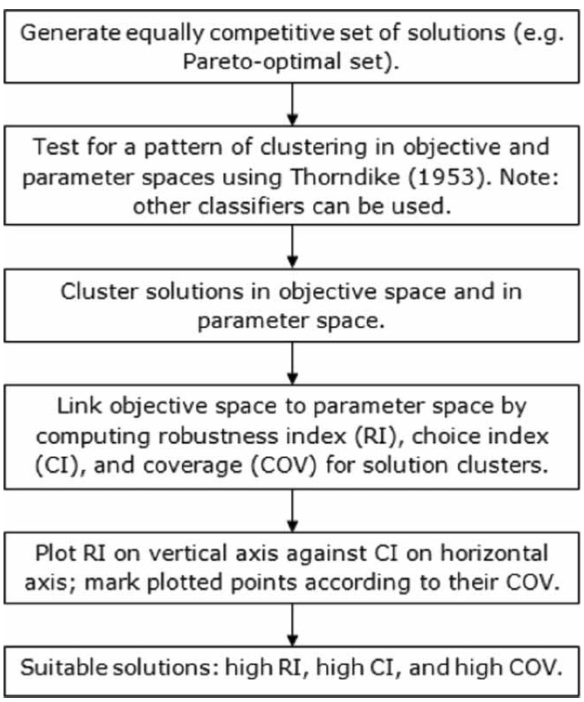

Figure 1 | Key procedures of the MCF to select a subset of equally competitive solutions using robustness, choice index and coverage.

Clustering in parameter space and objective space

We measure the distribution of solutions in parameter space by clustering model parameter values of the solutions found on the Pareto frontier. An application of a specific distance function is used to compare parameter vectors, which in future applications can be weighted to emphasize one parameter over another. As we have no bias towards one parameter over another, we normalize each parameter and take the Euclidean distance between the normalized vectors. We call the clusters thus formed 'parameter clusters' (PC) and index them so that the ith parameter cluster is denoted PC_i.

Similarly we cluster the distribution of solutions in objective space using a distance measure that combines the model evaluation functions (bias and RMSE in this study, that is, solutions discovered on the Pareto frontier). Specifically, a weighted Euclidean distance was used to combine bias and RMSE, although any distance measure tailored to the system under investigation could be used. These clusters are called 'fitness clusters' (FC) and the jth fitness cluster is denoted FC_j. The term 'fitness' is a terminology used in a genetic algorithm to refer to the quality of a solution or the performance of a candidate solution in a population.

Other classifiers different from cluster analysis could be applied to analyze the distribution of solutions, so clustering should not be thought of as the only approach. One suggestion is that cluster analysis should be used to analyze the distribution of solutions if and only if a pattern of clustering is observed following the test of clustering outlined in Thorndike (1953). The procedure shown in Thorndike (1953) also illustrates the 'knee' approach that shall be used in this paper to determine the appropriate number of clusters to be used in clustering a specific dataset based on the relative size of the reduction of variance introduced by increasing the number of clusters.

Once the clusters have been found, the distribution of solutions in parameter space and objective space can be combined using conditional probability as a means to find linkages between patterns observed in the two spaces. For example, given that a solution is found among a cluster of solutions in objective space, we can examine if these solutions can be generated from similar parameter pathways or perhaps there are other very different pathways that can produce the same behavior. To achieve this we outline the model characterization framework to define model scenarios based on the clustering of solutions in parameter space and objective space.

In this paper we refer to the clustering of solutions in parameter space as locality. Parameter sets that cluster closely can be described as local solutions and those solutions scattered and located far apart are non-local. Also solutions can be classified based on their clustering or arrangement in objective space. The arrangement of solutions in objective space describes a *choice* property such that local choice solutions are clustered closely in objective space while non-local choice solutions are scattered when evaluated in objective space.

Linking parameter space and objective space: categorizing Pareto-optimal solutions

Using the clusters in parameter space and objective space we can find comparisons between solutions by evaluating linkages between the two spaces. Specifically, we define four categories: *Robust*, *Sensitive*, *Local Choice* and *Non-local Choice* to illustrate these linkages. Note that these categories are tied to the clustering of solutions in both parameter space and objective space.

Robustness is the ability for a set of parameters to remain unchanged (i.e. insensitive) regardless of small changes or perturbations in the internal and external structure of a system (Jin & Sendhoff 2003; Deb & Gupta 2005; Ross *et al.* 2008; Willink 2008). Using this definition, robust parameter sets are defined as solutions that have similar values (i.e. belong to similar clusters) in both parameter space and objective space. Robustness links parameter space to objective space in a way that similarities in parameter space imply similarities in objective space.

We now determine whether a cluster of solutions is robust or sensitive through the use of a robustness index. The robustness index $\mathrm{RI}_{i,j}$ for the jth cluster in parameter space and the ith cluster in objective space is equivalent to the frequency-based estimate of the conditional probability $P(\mathrm{FC}_i|\mathrm{PC}_j)$ and can be computed as shown in Equation (1). $\|\mathrm{FC}_i\|$ represents the number of members (i.e. magnitude) of the cluster i in objective space (or fitness space) and $\|\mathrm{PC}_j\|$ is the number of members in cluster j in parameter space:

$$\mathrm{RI}_{i,j} = \frac{||\mathrm{FC}_i \cap \mathrm{PC}_j||}{||\mathrm{PC}_j||} \tag{1}$$

The robustness index $\mathrm{RI}_{i,j}$ shall also be denoted as $\mathrm{RI}(\mathrm{FC}_i, \mathrm{PC}_j)$. The $\mathrm{RI}_{i,j}$ varies between a value close to unity, representing a very robust solution, and a value close to zero, indicating a very sensitive solution. Robust solutions satisfy the condition: $\mathrm{RI}_{i,j} \approx 1.0$. *Sensitive* solutions, on the other hand, are parameter sets that have very similar pathways in parameter space but produce very different results in objective space. A set of solutions is categorized as sensitive if their robustness index is close to zero (i.e. $\mathrm{RI}_{i,j} \approx 0.0$).

Similarly, the choice index $\mathrm{CI}_{i,j}$ for the ith cluster in objective space and the jth cluster in parameter space is analogous to the frequency-based estimate of the conditional probability $P(\mathrm{PC}_j|\mathrm{FC}_i)$ and can be computed as shown in Equation (2). Given a solution in objective space cluster, the $\mathrm{CI}_{i,j}$ indicates whether the specified cluster has more or fewer alternatives (i.e. choices) in parameter space to generate a similar response in objective space. That is, the $\mathrm{CI}_{i,j}$ operates as a selection index:

$$\mathrm{CI}_{i,j} = 1 - \frac{||\mathrm{FC}_i \cap \mathrm{PC}_j||}{||\mathrm{FC}_i||} \tag{2}$$

where $\|\mathrm{FC}_i\|$ represents the number of members in cluster i in objective space and $\|\mathrm{PC}_j\|$ is the number of members in cluster j in parameter space. The choice index $CI_{i,j}$ shall also be denoted as $\mathrm{CI}(\mathrm{FC}_i, \mathrm{PC}_j)$. $\mathrm{CI}_{i,j}$ varies between a value close to zero, indicating a local choice, and a value close to one for non-local choice. This special purposed index, $\mathrm{CI}_{i,j}$, is used to emphasize whether there are several or a few choices in parameter space to produce a specified solution in objective space.

Clusters of solutions that are *local choice* have similar pathways in parameter space as well as similar fitness values, such that no other solutions exist that have similar fitness values with different parameter pathways. A strong local choice solution satisfies the condition: $CI_{i,j} \approx 0.0$. On the other hand, solutions possessing the *non-local choice* property have different solutions in parameter space but result in similar scenarios in objective space. Specifically, non-local choice solutions have several alternative parameter values to generate a specified watershed response. A strong non-local choice solution satisfies the condition $CI_{i,j} \approx 1.0$.

Note that solutions used in this type of analysis are equally competitive but they possess unique properties. Solutions do not need to be determined in a multi-objective fashion so the framework can apply to single-objective problems.

Using linkages between robustness and choice index: choosing a solution with desired behavior

The set of solutions categorized based on a robustness index ($RI_{i,j}$) and choice index ($CI_{i,j}$) can be combined by linking the two indices. The rationale for this analysis is that, by combining the two indices, a decision can be made as to whether a particular solution is robust enough and, if not, whether other solutions with similar performance but with different parameter pathways exist that may be in a robust region. Additionally, the analysis allows selection of parameter sets with the desired behavior from robust or sensitive solutions. This analysis is illustrated in Figure 2 where robustness index is plotted on the vertical axis against choice index on the horizontal axis for a hypothetical set of

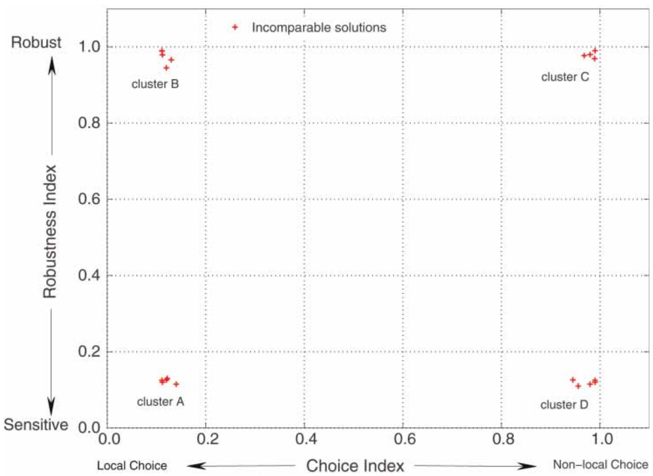

Figure 2 | Different clusters of Pareto-optimal solutions categorized based on robustness index and choice index.

solutions. The solution clusters A, B, C, and D combine two sets of properties, one from robustness index and the other from choice index. As a result, robust/sensitive solution clusters on the robustness index are either non-local choice or local choice.

Using Figure 2 as a decision tool, a user can further discern between two robust clusters (clusters B and C). Cluster C is robust but this cluster has a subset of solutions which can be generated from several other parameter pathways (non-local choice). In contrast, cluster B is robust but only a few similar parameter pathways can generate similar objective values (local choice). Hence a combined evaluation of robustness and choice index can be useful to distinguish between two similar robust clusters. Non-local choice solutions are equally relevant for evaluation of the physical meaning of model parameters and their sensitivity. Because non-local choice solutions have several parameter pathways and produce similar watershed responses, these pathways could be examined for their physical meaning by choosing the most descriptive.

Furthermore, a user could be interested in a parameter pathway that falls in one of the two sensitive clusters (clusters A and D). Cluster D is sensitive and non-local choice, that is, it comprises a set of solutions which belong to other parameter pathways as well as many other clusters in objective space. This cluster of solutions has alternatives in both parameter space and objective space. Because there are other alternatives to solutions in this region, those choices should be explored for decision-making, for example, on the basis of the physical meaning of parameter values. Indeed, a cluster that has similar performance but different parameter pathways might be found in a robust region and therefore a desirable solution. Cluster A is sensitive and local choice; solutions in this region also belong to several other parameter pathways but they are unified in objective space. A set of solutions in this region will be ideal for assessing the sensitivity and the physical meaning of parameter pathways. Although Figure 2 is illustrated for a hypothetical set of solutions, an analysis for real results is shown in the results section.

Study area

To investigate the concept of Pareto-optimality and validate the model characterization framework we apply the framework in a case study. The case study uses NSGA-II to calibrate SWAT for streamflow in the Fairchild Creek (FCr) watershed in southern Ontario. The FCr watershed shown in Figure 3 is a sub-catchment of the Grand River located in the south-eastern portion of the Grand River Basin. The FCr has a drainage area of about 400 km^2. The topography of the FCr watershed is gently sloping with an elevation interval of about 325 m at the upstream to 184 m at the downstream. The soils in the upstream portion of the watershed are predominately loam. The middle and lower sections have fine textured soils that are a mix of silt, clay, loam and some sand. A large portion of the watershed is cleared for agricultural activity but specific land cover distributions are approximately agriculture (64%), forestry (21%), pasture (9%), urban (5%) and open water and wetlands (1%). Streamflow in the FCr watershed collected from 1990 to 2003 varies widely from year to year, with the maximum and minimum daily discharges ranging between 75.9 m^3/s on January 17, 1995 and 0.145 m^3/s on July 16, 1999, respectively. The average annual daily discharge is 3.49 m^3/s with a standard deviation of 0.88 m^3/s. Extreme discharges are occasionally observed in January due to sudden high temperatures which result in snowmelt floods. But overall maximum discharges occurred in March and April due to accumulated snow.

SWAT and its calibration using NSGA-II

The Soil and Water Assessment Tool (SWAT) is a semi-distributed hydrological model that was developed by the United States Department of Agriculture in the early 1990s and has been continuously updated (Arnold et al. 1998; Arnold & Fohrer 2005). SWAT runs continuously at a daily time step and is capable of simulating long-term impacts of land management and climate on water and sediment in large river basins with varying soils and land cover (Arnold & Fohrer 2005). As a physically based model, SWAT inputs include specific information on topography, drainage networks, land use/cover, weather and land management practices. SWAT uses a digital elevation model (DEM) to delineate the watershed into sub-basins which are further divided into hydrologic response units (HRUs). HRUs are

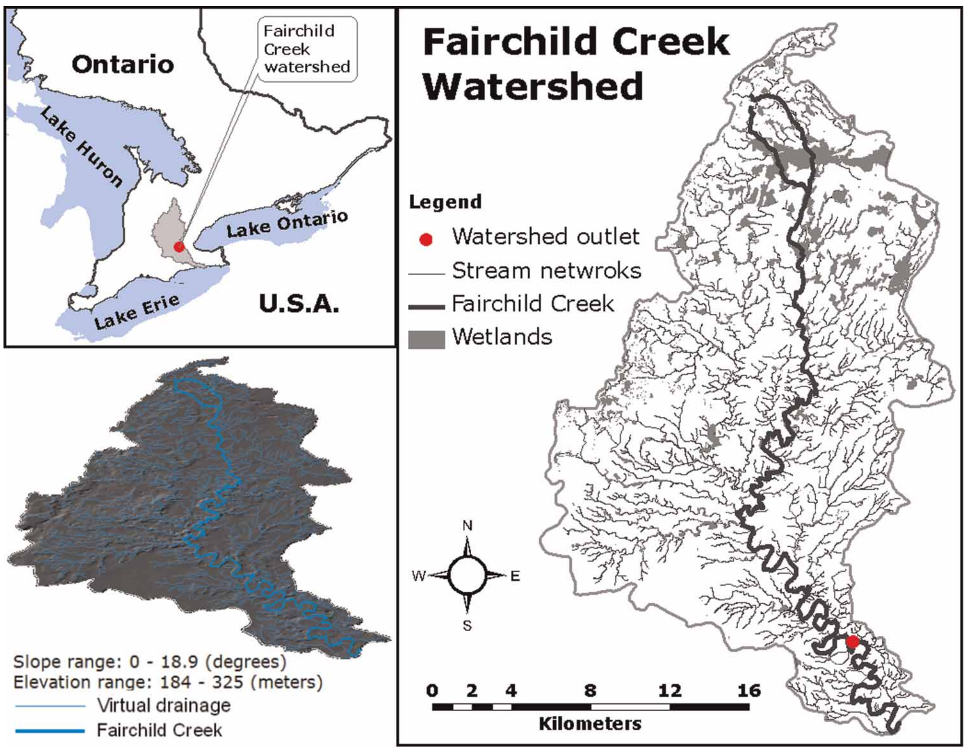

Figure 3 | Study area: Fairchild Creek watershed in southern Ontario (data source: Natural Resources Canada).

homogenous land areas within sub-basins which are categorized by specification of the unique land use and soil information. SWAT computes runoff volume using either the runoff curve number approach or the Green and Ampt infiltration equation (Green & Ampt 1911). Further description of SWAT can be found in various sources (Arnold et al. 1998; Neitsch et al. 2001; Arnold & Fohrer 2005).

The following procedures were undertaken to set up SWAT for the FCr watershed. The FCr watershed was delineated into sub-basins using a 10 m resolution DEM and a drainage network. The delineation was based on the distribution of drainage networks, and soil and land-cover information in the FCr watershed. The sub-basins were further subdivided into HRUs using both land use/cover and soil distribution. The generated HRUs are homogeneous land areas within sub-basins based on a unique combination of land use and soil distribution within a sub-basin. Next, weather information based on precipitation, minimum and maximum temperatures, solar radiation, humidity and wind are generated for the sub-basins from weather stations within and surrounding the FCr watershed. SWAT assembles all the above information to generate input files and initializes parameter values ready for calibration. Table 1 shows 24 SWAT model parameters, their descriptions and

Table 1 | SWAT model parameters, their lower and upper bounds, and their descriptions

Model parameter	Change type	Lower bound	Upper bound	Description
TIMP	Absolute	0.01	1.00	Snow pack temperature lag factor
SURLAG	Absolute	0.00	24.00	Surface runoff lag time (d)
SFTMP	Absolute	−5.00	5.00	Snowfall temperature (°C)
SMTMP	Absolute	−5.00	5.00	Snowmelt base temperature (°C)
SMFMX	Absolute	0.00	10.00	Melt factor for snow on Jun. 21 (mm/ °C d)
SMFMN	Absolute	0.00	10.00	Melt factor for snow on Dec. 21 (mm/ °C d)
MSK-CO1	Absolute	0.00	10.00	Calibration coefficient used to control impact of the storage time constant for normal flow (km)
MSK-CO2	Absolute	0.00	10.00	Calibration coefficient used to control impact of the storage time constant for low flow (km)
MSK-X	Absolute	0.00	0.30	Weighting factor controlling relative importance of inflow rate and outflow rate in determining water storage in reach segment
CH-K1	Absolute	0.00	150.00	Effective hydraulic conductivity in tributary channel alluvium (mm/h)
CN2	Relative	−0.35	0.35	Initial SCS runoff curve number for moisture condition II
CH-N2	Relative	−0.50	0.50	Manning's n value for main channel
CH-K2	Absolute	0.01	150.00	Effective hydraulic conductivity in main channel alluvium (mm/h)
ALPHA-BF	Absolute	0.00	1.00	Baseflow alpha factor (d)
GWQMN	Absolute	0.00	5000.00	Threshold depth of water in the shallow aquifer required for return flow to occur (mm)
GW-REVAP	Absolute	0.02	0.20	Groundwater 'revap' (transfer of groundwater to upper soil layers) coefficient
GW-DELAY	Absolute	0.00	500.00	Groundwater delay (d)
RCHRG-DP	Absolute	0.00	1.00	Deep aquifer percolation fraction
REVAPMN	Absolute	0.00	500.00	Threshold depth of water in the shallow aquifer for 'revap' to occur (mm)
CANMX	Absolute	0.00	100.00	Maximum canopy storage (mm H_2O)
ESCO	Absolute	0.00	1.00	Soil evaporation compensation factor
EPCO	Absolute	0.00	1.00	Plant uptake compensation factor
SOL-AWC	Relative	−0.50	0.50	Available water capacity of the soil layer (mm H_2O/mm soil)
SOL-K	Relative	−0.80	0.80	Saturated hydraulic conductivity (mm/h)

types of modification used during the parameter estimation procedure.

Next, the NSGA-II is applied to calibrate SWAT for simulations of streamflow. The NSGA-II/SWAT framework was run through 400 generations with a population of 200 parameter sets which are evaluated based on model bias in Equation (3) and RMSE in Equation (4). Though model bias is bounded by RMSE (i.e. bias cannot be greater than RMSE), within those bounds the two objective functions are independent as they assess different aspects of the calibration process (Gupta *et al.* 1998). The default crossover probability of 0.8 and a mutation probability of $1/l$ (where l is the number of variables) were used to guide the search. The modeling period was 12 years, starting from 1992 to 2003. The model calibration period was set to the first six years (i.e. 1992–1997) and the remaining six years (i.e. 1998–2003) were used to validate model outputs:

$$\text{Bias} = \left(\sum_{i=1}^{k} (y_{s,i} - y_{o,i}) \right) \Big/ k \quad (3)$$

$$\text{RMSE} = \sqrt{\left(\sum_{i=1}^{k} (y_{s,i} - y_{o,i})^2 \right) \Big/ k} \quad (4)$$

where $y_{o,i}$ is the observed streamflow for ith day, $y_{s,i}$ the simulated streamflow for ith day, and k the period of simulation in days.

RESULTS AND DISCUSSION

This section presents the NSGA-II results for calibrating SWAT for streamflow and analyzes the resulting Pareto set by implementing the model characterization framework. The solutions were evaluated for two linkages: one based on linkages between objective space and parameter space, and the other based on linkages between robustness and the choice index. This combined evaluation of distribution of solutions allows the identification of solutions which remain competitive across several validation periods. That is, the evaluation is capable to identify high-performing streamflow solutions under changing meteorological conditions in the watershed.

The solutions in the Pareto-optimal set describe a variety of acceptable simulations; finding an objective rationale to select one solution amongst numerous possible solutions can be a daunting task. The Pareto frontier is a trade-off surface that is represented by a range of parameter sets which have equally competitive RMSE and model bias values. Two different types of trade-off are apparent on the Pareto frontier: the first is the trade-off between the parameter sets based on their model parameter values (parameter space) and the second trade-off is between parameter sets based on their predicted streamflows (objective space). As a result, when choosing one solution over another from the Pareto frontier, some compromise between optimizing either RMSE or model bias inherently occurs. In the discussion below, we evaluate the solutions on the Pareto frontier for properties of robustness and choice index on two fronts: parameter space and objective space.

Evaluation of robustness and choice properties for parameter sets

There are solutions on the Pareto frontier that possess unique properties either from the parameter values themselves or on the Pareto frontier but it is unknown if these solutions possess robustness or choice properties. By evaluation of robustness and choice properties, we intend to examine the distribution of parameter values and the spread of RMSE and bias values on the Pareto frontier. Specifically, the following is an implementation of the model characterization framework described in later subsections.

First, using cluster analysis on the Pareto frontier, six unique clusters are readily identified in objective space. A k-means clustering was applied using a Euclidean distance measure, where the six clusters and their cluster sizes are $\|FC_1\| = 60$, $\|FC_2\| = 7$, $\|FC_3\| = 18$, $\|FC_4\| = 13$, $\|FC_5\| = 10$ and $\|FC_6\| = 7$. The decision to use six clusters was based on Figure 4 where the percent of variance is explained by different numbers of clusters k, and $k = 6$ provided the most suitable number of clusters as it falls at the knee.

Second, on the distribution of parameter values, eight parameter clusters (denoted PC_1, PC_2, PC_3, PC_4, PC_5, PC_6,

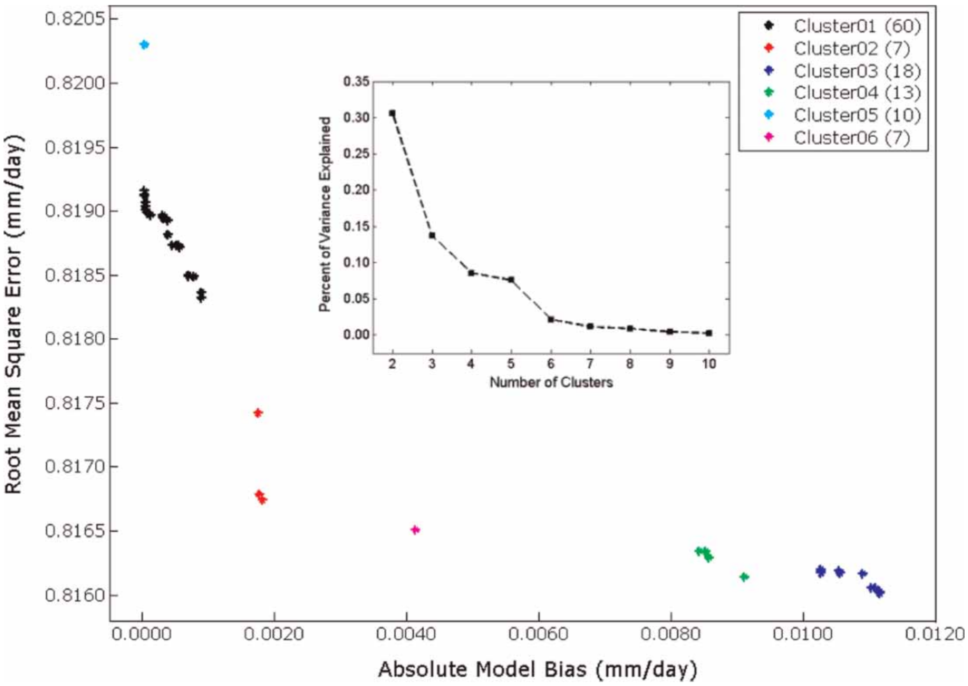

Figure 4 | Cluster of solutions found on the Pareto frontier for calibration output. Inset plot is a number of clusters and their corresponding variances based on the *knee* method (Thorndike 1953).

PC_7 and PC_8) were identified in parameter space. Again, the *k*-means clustering technique was applied to all parameter values across the 115 parameter sets using a Euclidean distance measure. The eight clusters are obtained using the same procedure as was done for objective space clustering; the plot for parameter space is shown in Figure 5. For the 115 solutions, PC_1 has 14 solutions, PC_2 has 33 solutions, PC_3 has 20 solutions, PC_4 has 14 solutions, PC_5 has 5 solutions, PC_6 has 10 solutions, PC_7 has 12 solutions and PC_8 has 7 solutions.

In summary, the number of clusters in parameter space is eight and there are six clusters in objective space. The number of clusters is problem-dependent and a function of the Pareto frontier and its corresponding distribution of solutions in both spaces.

Based on clustering in both spaces, we evaluate the level of partitioning, robustness and choice properties for solutions. In association with partitioning of the Pareto frontier, we use coverage to indicate the proportion of solutions in a specified partition. Coverage is a measure of

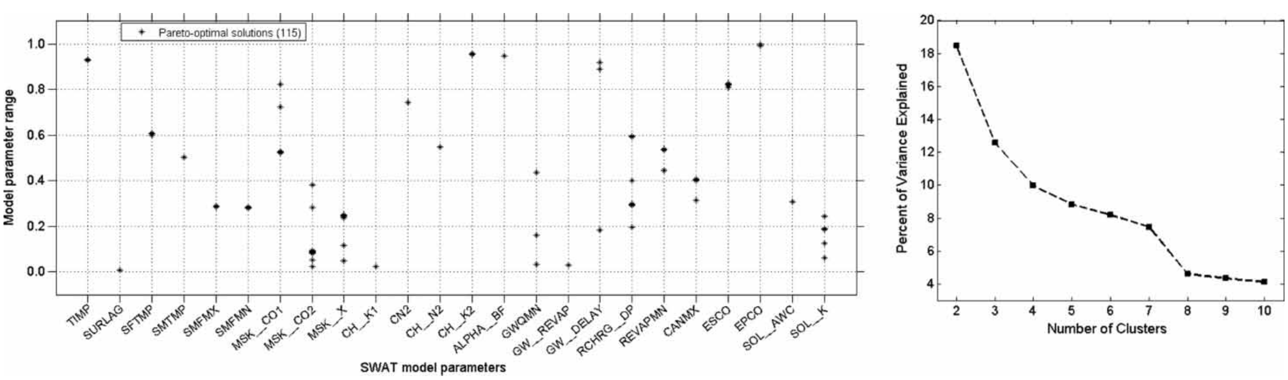

Figure 5 | Left-hand side graph shows plotted points for SWAT model parameters and their clustering in parameter space. Right-hand side plot is a number of clusters and their corresponding variances based on the *knee* method (Thorndike 1953).

how well a specific cluster in both parameter and objective spaces is represented on the Pareto frontier. The level of coverage gives information about regions of concentration (or similarity) of solutions on the Pareto frontier. To partition the Pareto frontier, the number of solutions in the overlap region between a cluster in objective space and a cluster in parameter space are divided by the number of all solutions on the Pareto frontier. Table 2 illustrates the partitioning of the Pareto frontier into different clusters (e.g. FC_1).

Overall, PC_2 represents the most concentrated region in parameter space and FC_1 has the most coverage in objective space. The most concentrated clusters, particularly, in parameter space describe a representative pathway for clustering in parameter space. These results are important because there are solutions which represent locations of maximum concentrations for parameter-by-parameter clustering but do not describe a representative pathway for clustering in parameter space. That is, information about individual parameter clusters is important but they are not adequate to describe a representative pathway to generate information about parameter relationships.

Next, the evaluation of robustness determines clusters (e.g. PC_1) in parameter space which overlap with clusters (e.g. FC_1) in objective space. Table 2 illustrates the computation of robustness index for different clusters (e.g. FC_1) on the Pareto frontier with respect to the clusters (e.g. PC_2) in parameter space. The most robust combination of parameter clusters are RI(PC_1, FC_1), RI(PC_3, FC_1), RI(PC_5, FC_1), RI(PC_6, FC_5) and RI(PC_8, FC_1).

In addition, there are sensitive solutions in Table 3, that is, solutions with very similar pathways in parameter space which produce different responses at the watershed outlet. The cluster PC_4 has 14 solutions located close together in parameter space but they fall into clusters FC_1 (with 10 solutions) and FC_2 (with four solutions) in objective space. A similar pattern is observed for 12 solutions in cluster PC_7 which are split into clusters FC_3 (with seven solutions) and FC_6 (with five solutions) in objective space.

The non-local choice describes solutions that have similar RMSE and bias values on the Pareto frontier but have different values in parameter space. Table 2 shows computation of the choice index for different clusters (e.g. PC_1) in parameter space with respect to clusters (e.g. FC_1) on the Pareto frontier. In Table 2, the 60 solutions in FC_1 are split into six clusters in parameter space; PC_1 has 14 solutions, PC_2 has four solutions, PC_3 has 20 solutions, PC_4 has 10 solutions, PC_5 has five solutions and PC_8 has seven solutions. This example shows how a set of solutions in FC_1 with very similar RMSE and bias values in objective space are split into six different clusters in parameter space. That is, there are several choices in parameter pathways to generate similar solutions in FC_1.

In contrast, strong local choice solutions are exemplified in the overlapped region between FC_5 and PC_6 where all the solutions in the objective cluster FC_5 are located in the parameter cluster PC_6. Similarly, local choice solutions are obtained in CI(FC_4, PC_2). Again, local choice solutions are different from robust solutions as conditionality for robustness is in parameter space while conditionality for the choice index is in objective space. As we have seen in Table 2, although solutions in the overlapped region between FC_4 and PC_2 are located closely in objective space and parameter space these solutions are not robust. They are sensitive on the robustness scale as constituent solutions in PC_2 are all not located in FC_3 but they are split into other objective clusters: FC_1, FC_2, FC_3 and FC_6.

Linkages between robustness and choice index: solutions with desired properties

Now we illustrate linkages between the level of coverage and the set of solutions categorized as robust, sensitive, local choice and non-local choice. The evaluation of robustness and choice properties has selected a category of solutions in the overlapped region between clusters in parameter space and objective space. These categories of solutions represent isolated regions across the clusters in both parameter space and objective space.

As illustrated in Table 2, there are five categories of solutions representing robust regions, and a similar pattern is demonstrated for sensitive, local choice and non-local choice solutions. To decide which category of solutions from which to select a robust solution, the categories should be evaluated using their level of coverage which is computed in Table 2. An expanded coverage signifies a concentrated region on the Pareto frontier either in parameter space or objective space. Although the NSGA-II properties do not allow for full sensitivity analysis in these

Table 2 | The partitioning of solutions (coverage), robustness and choice indices for 1992–1997 calibration output

	FC_1	FC_2	FC_3	FC_4	FC_5	FC_6	Sum
Number of solutions in overlapped cluster							
PC_1	14	0	0	0	0	0	14
PC_2	4	3	11	13	0	2	33
PC_3	20	0	0	0	0	0	20
PC_4	10	4	0	0	0	0	14
PC_5	5	0	0	0	0	0	5
PC_6	0	0	0	0	10	0	10
PC_7	0	0	7	0	0	5	12
PC_8	7	0	0	0	0	0	7
Sum	60	7	18	13	10	7	115
Coverage							
PC_1	0.122	0.000	0.000	0.000	0.000	0.000	0.122
PC_2	0.035	0.026	0.096	0.113	0.000	0.017	0.287
PC_3	0.174	0.000	0.000	0.000	0.000	0.000	0.174
PC_4	0.087	0.035	0.000	0.000	0.000	0.000	0.122
PC_5	0.043	0.000	0.000	0.000	0.000	0.000	0.043
PC_6	0.000	0.000	0.000	0.000	0.087	0.000	0.087
PC_7	0.000	0.000	0.061	0.000	0.000	0.043	0.104
PC_8	0.061	0.000	0.000	0.000	0.000	0.000	0.061
Sum	0.522	0.061	0.157	0.113	0.087	0.061	1.000
Robustness index							
PC_1	1.000	0.000	0.000	0.000	0.000	0.000	1.000
PC_2	0.121	0.091	0.333	0.394	0.000	0.061	1.000
PC_3	1.000	0.000	0.000	0.000	0.000	0.000	1.000
PC_4	0.714	0.286	0.000	0.000	0.000	0.000	1.000
PC_5	1.000	0.000	0.000	0.000	0.000	0.000	1.000
PC_6	0.000	0.000	0.000	0.000	1.000	0.000	1.000
PC_7	0.000	0.000	0.583	0.000	0.000	0.417	1.000
PC_8	1.000	0.000	0.000	0.000	0.000	0.000	1.000
Sum	4.835	0.377	0.917	0.394	1.000	0.477	8.000
Choice index							
PC_1	0.767	1.000	1.000	1.000	1.000	1.000	5.767
PC_2	0.933	0.571	0.389	0.000	1.000	0.714	3.608
PC_3	0.667	1.000	1.000	1.000	1.000	1.000	5.667
PC_4	0.833	0.429	1.000	1.000	1.000	1.000	5.262
PC_5	0.917	1.000	1.000	1.000	1.000	1.000	5.917
PC_6	1.000	1.000	1.000	1.000	0.000	1.000	5.000
PC_7	1.000	1.000	0.611	1.000	1.000	0.286	4.897
PC_8	0.883	1.000	1.000	1.000	1.000	1.000	5.883
Sum	7.000	7.000	7.000	7.000	7.000	7.000	42.000

Table 3 | Partitioning of solutions (coverage), robustness and choice indexes for calibration output

Parameter cluster	Objective cluster	Solutions	CI	RI
Robustness and choice indexes for 1992–1997 calibration output				
PC_1	FC_1	16,17,18,19,20,23,49,51,61,65,73,79,99,108	0.767	1.000
PC_2	FC_1	7,21,37,102	0.933	0.121
PC_2	FC_2	84,103,105	0.571	0.091
PC_2	FC_3	12,28,46,57,70,74,78,80,83,89,91	0.389	0.333
PC_2	FC_4	1,14,35,38,44,47,53,67,69,81,93,94,101	0.000	0.394
PC_2	FC_6	52,92	0.714	0.061
PC_3	FC_1	5,9,13,15,27,29,30,31,34,39,43,45,48,50,62,63,64,71,82,114	0.667	1.000
PC_4	FC_1	8,10,11,22,24,25,26,32,36,54	0.833	0.714
PC_4	FC_2	107,110,112,113	0.429	0.286
PC_5	FC_1	41,42,72,97,106	0.917	1.000
PC_6	FC_5	2,3,4,76,77,85,87,88,90,100	0.000	1.000
PC_7	FC_3	33,56,75,86,98,109,111	0.611	0.583
PC_7	FC_6	59,60,66,68,104	0.286	0.417
PC_8	FC_1	6,40,55,58,95,96,115	0.883	1.000

concentrated regions, these locations still indicate a unified space in either parameter space or objective space to explore further. Among the five robust regions in Table 2, the category of solutions in $P(FC_1, PC_3)$ has the highest coverage (0.17), indicating an overlapped region from which to select a strong robust solution. Similar analysis could be applied to relate coverage for the selection of sensitive, local choice and non-local choice solutions.

Furthermore, following the analysis method in an earlier subsection we determine linkages between robustness and choice index for solution clusters. The relationship between robustness and choice index for our solutions is shown in Figure 6 where we illustrate different pairs of properties for solution clusters. The plotted points are sized based on their level of coverage; a higher level of coverage is identified with a larger maker. As a result, Figure 6 is using three properties: robustness, choice index and coverage to discern between clusters of solutions. As seen from Figure 6, four groups of clusters are identified; robust and non-local choice, robust and local choice, sensitive and non-local choice, and sensitive and local choice. The pair of properties for different solution clusters divides the Pareto-optimal set into specific categories from which a decision-maker can select a single solution based on the properties of the problem under consideration. Again, this categorization is a function of the set of solutions obtained from the Pareto frontier.

To select a robust solution, a decision-maker can choose from five strongly robust clusters: $P(PC_1, FC_1)$, $P(PC_3, FC_1)$, $P(PC_5, FC_1)$, $P(PC_6, FC_5)$ and $P(PC_8, FC_1)$. The cluster $P(PC_5, FC_1)$ is strongly robust but it has solutions whose objective values could be determined from several other parameter pathways. But we can reduce the diversity in parameter pathways by selecting $P(PC_8, FC_1)$ or $P(PC_1, FC_1)$ as they both have fewer parameter pathways than $P(PC_5, FC_1)$. As seen from Figure 6, the cluster $P(PC_6, FC_5)$ has the most unified parameter pathway for producing similar solutions in objective space. But the cluster $P(PC_6, FC_5)$ does not have the highest coverage among the five robust solutions as it trails both $P(PC_1, FC_1)$ and $P(PC_3, FC_1)$ on the coverage scale. Overall, we have two distinct clusters of robust solutions: one with a unified parameter pathway and a moderate coverage $P(PC_6, FC_5)$ and the other with several choices in parameter space and the largest coverage $P(PC_3, FC_1)$.

Evaluation of Pareto-optimal solutions for several validation periods

Solution clusters in the Pareto-optimal set were evaluated across 15 different validation periods to determine their

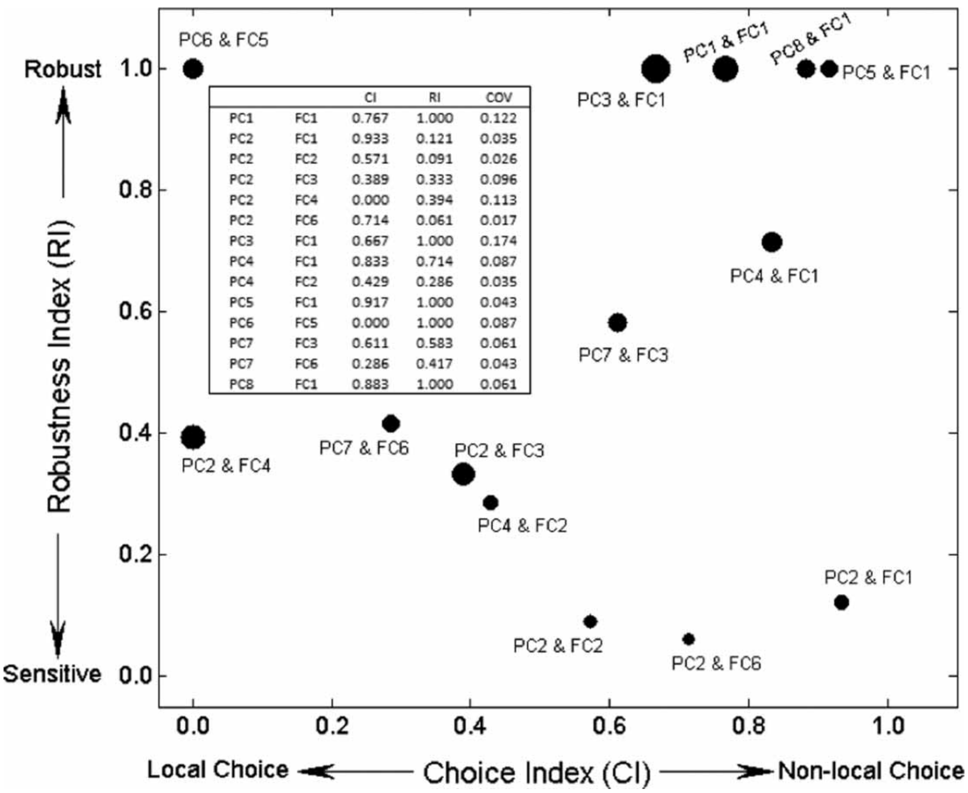

Figure 6 | Pareto-optimal solution clusters evaluated based on robustness and choice index. Note that plotted dots are sized based on their corresponding level of coverage.

performance levels. To facilitate this analysis, Pareto-optimal solutions in different clusters are listed in Table 3 with their associated robustness and choice indexes. Solutions which remained non-dominated for various validation periods are shown in Table 4.

The results show that solutions which remained non-dominated are variable for different validation periods. Some validation periods (e.g. 2000–2003) have a large set of non-dominated solutions while others (e.g. 1999–2002) have small memberships. Notwithstanding this variability, a subset of solutions has consistently remained non-dominated for several validation periods. For example, solutions 51,73,99 from $PC_1 \cap FC_1$, solutions 6,55,58,96,115 from $PC_8 \cap FC_1$ and solutions 48,50,71 from $PC_3 \cap FC_1$. These solutions are associated with high robustness index and their clusters performed well across several validation periods. Additionally, these solutions are generally associated with clusters which have more choices in objective space.

Notes on the utility of the MCF

Results in the above sections have demonstrated the utility of the MCF to select a subset of Pareto-optimal solutions by exploring relationships in parameter space and objective space when conducting parameter estimation. The MCF procedure, more importantly, determined solutions which remain competitive when evaluated for future simulation periods. In contrast to traditional methods, the MCF has linked robustness to other properties including coverage and choice index. The MCF has illustrated that robustness is not enough information alone to evaluate the performance solutions across future simulation periods, and that choice index and coverage provide important supplementary information to evaluate solution clusters. These indexes were shown in a unit interval to illustrate the relative robustness or sensitivity for different solution clusters.

The capability to determine solutions with desired properties has a practical appeal to both modelers and

Table 4 | Performance of Pareto-optimal solutions and solution clusters across several validation periods

Validation period	Non-dominated solutions
1998–1999	2,3,4,6,48,50,55,67,71,73,75,76,77,85,87,88,90,95,98,99,100,102,109,111
1999–2000	6,55,58,67,95,96,99,101,102
2000–2001	6,40,58,67,70,71,73,83,96,99,102,115
2001–2002	2,6,16,18,19,20,23,40,48,50,51,54,55,58,61,65,67,68,71,73,77,85,87,88,95,96,99,102,112,115
2002–2003	1,5,12,14,20,28,33,35,37,38,44,47,48,50,56,72,74,81,84,89,91,97,99,102,103,105,106
1998–2000	6,55,58,67,71,73,75,95,96,98,99,102,106,109,111
1999–2001	6,58,67,96,99,102,106
2000–2002	6,40,58,67,96,99,102,106,115
2001–2003	1,6,14,20,35,40,44,47,48,50,51,55,58,67,71,73,92,95,96,99,102,106,115
1998–2001	6,10,55,67,95,99,102
1999–2002	6,58,67,96,102
2000–2003	1,6,14,20,35,40,44,47,48,50,51,58,67,71,73,92,96,99,102,106,115
1998–2002	6,55,67,95,102,106,115
1999–2003	6,40,58,67,96,99,102,106,115
1998–2003	6,40,48,50,51,55,58,67,71,73,92,95,96,98,99,102,106,109,115

decision-makers. That is, users can select robust solutions with different behaviors: robust solutions with few choices in model performance or robust solutions with several alternative performances in objective space. This is useful information for modelers and users alike who would use both types of solutions under different conditions. For example, robust solutions with several choices in model performance are important in data assimilation operations where a stable model state is crucial to estimate the ensemble of model forecasts under different simulation periods.

Additionally, the MCF illustrates regions in parameter space which could be used to specify parameter limits and to speed up future model parameterizations. Sensitive parameter pathways would also be avoided in these operations. Note that the MCF applied two pieces of information on the sensitivity of model parameters: sensitivity on a parameter-by-parameter basis and sensitive parameter pathways which actually use parameter relationships. While traditional sensitivity analysis usually focuses on the former, the MCF incorporates the latter to emphasize parameter relationships. Clustered regions were observed on a parameter-by-parameter basis, but these individual clustered regions play a small role when finding the dominant pathway across all model parameters.

CONCLUSIONS

This study has explored the concept of Pareto-optimality to search for robust solutions by investigating the distribution of solutions in objective space and parameter space. Our method applied an NSGA-II framework to calibrate SWAT for simulations of streamflow based on two objectives: RMSE and model bias. The NSGA-II output comprises a set of solutions, each with a trade-off between RMSE and model bias.

Using Pareto-optimal solutions from the NSGA-II method, we developed and applied MCF to identify robust, sensitive, local choice and non-local choice solutions by evaluating the relationships between patterns observed in objective space and those in parameter space. The model characterization framework used cluster analysis to examine the distribution of solutions in objective space and parameter space, and conditional probability to combine the distribution of solutions in both spaces. Using this framework, we were able to compute different indices that were used to separate robust solutions from sensitive, local choice and non-local choice solutions.

Furthermore, our study evaluated Pareto-optimal solutions across 15 different validation periods. As illustrated, not all non-dominated solutions found at the calibration phase perform the same for different validation periods. The solutions that perform consistently well for several validation periods have been shown to come from clustered parameter space; demonstrating the importance of the distribution of solutions in parameter space. Additionally, the evaluation reveals the influence of robustness and coverage on the performance level of solution clusters for different validation periods. The framework outlines an evaluation criteria for robust solutions and emphasizes the important connection between robustness and choice index to further differentiate a set of robust solutions into two distinct subcategories.

Overall, a key attribute of our method is the integration of robustness, coverage and choice index properties. Our results from the MCF illustrate robustness in a way that similarities in parameter pathways imply similarities in objective space, and choice index by showing how similar watershed responses can be generated from several parameter pathways. There are also sensitive solutions on the Pareto frontier which have similar parameter pathways but produce different responses in objective space. The properties of robustness, coverage and choice index are important for decision-making purposes and are critical for accurate description of the watershed under different conditions/scenarios. More importantly, robustness has been shown to be strongly associated with the performance of solutions across several validation periods.

ACKNOWLEDGEMENTS

This work is supported by the Natural Sciences and Engineering Research Council of Canada and the Canadian Foundation for Climate and Atmospheric Sciences. We thank Robert Collier for his help with NSGA-II. We are grateful to Drs Wanhong Yang and Yongbo Liu, and Adam Bonnycastle for providing data and suggestions on model set-up. We thank Dr Bryan Tolson for his comments on the experimental set-up for validating the model characterization framework. We thank the anonymous reviewers for their comments and efforts.

REFERENCES

Arnold, J. G. & Fohrer, N. 2005 SWAT2000: current capabilities and research opportunities in applied watershed modelling. *Hydrol. Process.* **19**, 563–572.

Arnold, J. G., Srinivasan, R., Muttiah, R. S. & Williams, J. R. 1998 Large area hydrologic modeling and assessment part 1: model development. *J. AWRA* **34**, 73–89.

Bekele, E. G. & Nicklow, J. W. 2007 Multi-objective automatic calibration of SWAT using NSGA-II. *J. Hydrol.* **341**, 165–176.

Bertsimas, D. & Thiele, A. 2006 A robust optimization approach to inventory theory. *Oper. Res.* **54** (1), 150–168.

Chen, X., Sim, M. & Sun, P. 2007 A robust optimization perspective on stochastic programming. *Oper. Res.* **55** (6), 1058–1071.

Coello Coello, C. A., Van Veldhuizen, D. A. & Lamont, G. B. 2002 *Evolutionary Algorithms for Solving Multi-Objective Problems*. Kluwer, Amsterdam.

Crispim, J. A. & de Sousa, J. P. 2009 Partner selection in virtual enterprises: a multi-criteria decision support approach. *Int. J. Prod. Res.* **47** (17), 4791–4812.

Deb, K. & Gupta, H. 2005 Searching for robust Pareto-optimal solutions in multi-objective optimization. *LNCS* **3410**, 150–164.

Deb, K. & Gupta, H. 2006 Introducing robustness in multi-objective optimization. *Evolut. Comput.* **14** (4), 463–494.

Dumedah, G., Berg, A. A., Wineberg, M. & Collier, R. 2010 Selecting model parameter sets from a trade-off surface generated from the Non-Dominated Sorting Genetic Algorithm-II. *Wat. Res. Mngmnt.* **24** (15), 4469–4489.

Ferreira, J. C., Fonseca, C. M. & Gaspar-Cunha, A. 2007 Methodology to select solutions from the pareto-optimal set: a comparative study. In *Genetic and Evolutionary Computation Conference. Proc 9th Annual Conference on Genetic and Evolutionary Computation, London, July*. ACM, New York, pp. 789–796.

Goh, J., Hall, N. G. & Sim, M. 2010 *Robust Optimization Strategies for Total Cost Control in Project Management*. Graduate School of Business, Stanford University, California, USA.

Goh, J. & Sim, M. 2010 Distributionally robust optimization and its tractable approximations. *Oper. Res.* **58** (4), 902–917.

Green, W. H. & Ampt, G. A. 1911 Studies on soil physics, 1: The flow of air and water through soils. *J. Agric. Sci.* **4**, 1–24.

Grierson, D. E. 2008 Pareto multi-criteria decision making. *Adv. Engng. Inf.* **22** (3), 371–384.

Gunawan, S. & Azarm, S. 2004 Non-gradient based parameter sensitivity estimation for single objective robust design optimization. *Trans. ASME J. Mech. Design* **126** (3), 395–402.

Gunawan, S. & Azarm, S. 2005 Multi-objective robust optimization using a sensitivity region concept. *Struct. Multidisc. Optim.* **29** (1), 50–60.

Gupta, H., Sorooshian, S. & Yapo, P. O. 1998. Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. *Water Res. Res.* **34**, 751–763.

Hejazi, M. I., Cai, X. & Borah, D. K. 2008 Calibrating a watershed simulation model involving human interference: an application of multi-objective genetic algorithms. *J. Hydroinf.* **10** (1), 97–111.

Jin, Y. & Sendhoff, B. 2003 Trade-off between performance and robustness: an evolutionary multiobjective approach. *Evolut. Multi-Criterion Optim. (EMO)* **2632** (68), 237–251.

Khu, S. T. & Madsen, H. 2005 Multiobjective calibration with Pareto preference ordering: an application to rainfall-runoff model calibration. *Water Res. Res.* **41** (3), W03004.

Li, M., Azarm, S. & Aute, V. 2005 A multi-objective genetic algorithm for robust design optimization. In *GECCO '05. Proc. 2005 Conference on Genetic and Evolutionary Computation*. ACM, New York.

Marler, R. T. & Arora, J. S. 2004 Survey of multi-objective optimization methods for engineering. *Struct. Multidisc. Optim.* **26** (6), 369–395.

Nazemi, A., Xin, Y. & Chan, A. H. 2006 Extracting a set of robust pareto-optimal parameters for hydrologic models using NSGA-II and SCEM. In *2006 IEEE Congress on Evolutionary Computation*. pp. 1901–1908.

Neitsch, S. L., Arnold, J. G., Kiniry, J. R. & Williams, J. R. 2001 *Soil and Water Assessment Tool: Theoretical Documentation, version 2000*. Soil and Water Research Service, Temple, TX.

Ray, T. 2002 Constrained robust optimal design using a multiobjective evolutionary algorithm. In *Proc. Congress on Evolutionary Computation (CEC'2002)*. IEEE Service Center, Piscataway, NJ, pp. 419–424.

Ross, A. M., Rhodes, D. H. & Hastings, D. E. 2008 Defining changeability: reconciling flexibility, adaptability, scalability, modifiability, and robustness for maintaining system lifecycle value. *Syst. Engng.* **11** (3), 246–262.

Taboada, H. & Coit, D. 2006 Data mining techniques to facilitate the analysis of the Pareto-optimal set for multiple objective problems. In *Proc. Industrial Engineering Research Conference*, Orlando, FL, May. IERC, New York.

Tang, Y., Reed, P. & Wagener, T. 2006 How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration? *Hydrol. Earth Syst. Sci.* **10**, 289–307.

Thorndike, R. L. 1953 Who belongs in the family? *Psychometrika* **18**, 267–276.

Tsutsui, S. & Ghosh, A. 1997 Genetic algorithms with a robust solution searching scheme. *IEEE Trans. Evolut. Comput.* **1** (3), 201–219.

Willink, R. 2008 What is robustness in data analysis? *Metrologia* **45**, 442–447.

First received 22 September 2010; accepted in revised form 11 March 2011. Available online 22 June 2011