Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization

Zhenxing Zhang, Thorsten Wagener, Patrick Reed, and Rashi Bhushan

Received 14 January 2008; revised 2 April 2008; accepted 25 June 2008; published 18 September 2008.

[1] Approaches to predictions in ungauged basins have so far mainly focused on a priori parameter estimates from physical watershed characteristics or on the regionalization of model parameters. Recent studies suggest that the regionalization of hydrologic indices (e.g., streamflow characteristics) provides an additional way to extrapolate information about the expected watershed response to ungauged locations for use in continuous watershed modeling. This study contributes a novel multiobjective framework for identifying behavioral parameter ensembles for ungauged basins using suites of regionalized hydrologic indices. The new formulation enables the use of multiobjective optimization algorithms for the identification of model ensembles for predictions in ungauged basins for the first time. Application of the new formulation to 30 watersheds located in England and Wales and comparison of the results with a Monte Carlo approach demonstrate that the new formulation will significantly advance our ability to reduce the uncertainty of predictions in ungauged basins.


1. Introduction

[2] Hydrologic models are important tools to simulate the functional behavior of watersheds in terms of how they partition (into different pathways), store and eventually release precipitation for both scientific and operational studies. The degree of spatial conceptualization of processes necessary to simplify reality means that usually at least some of the model’s parameters are not directly observable from physical watershed characteristics, but have to be estimated through a process of calibration [e.g., Beven, 1989, 2006]. Significant progress has been made over the last 15 years in developing powerful calibration algorithms that can reliably find globally optimum solutions in the often highly complex parameter space of typical watershed models, if observations of the watershed response of interest are available (see discussions by Wagener and Gupta [2005] and Gupta et al. [2008]).

[3] While this approach has enhanced our predictive capability at gauged locations, sufficiently long and high-quality observations of streamflow are often not available; that is, we face the problem of predictions in ungauged basins [Sivapalan et al., 2003]. This problem could arise either because of the fact that no gauging station exists or that we want to predict the impact of future environmental (e.g., land use) change on hydrologic variables [Wagener, 2007]. Alternative approaches to calibration are required to estimate parameters in these circumstances. Thus far, solutions to the problem of providing continuous flow predictions at ungauged locations have focused on developing better a priori parameter estimates derived (directly or indirectly) from physical watershed characteristics. Some studies have directly used observable characteristics such as soil hydraulic and vegetation properties to derive a priori estimates [e.g., Atkinson et al., 2003; Koren et al., 2003], assuming that the correlation between these properties and parameters is sufficiently strong to at least provide reasonable initial estimates. Others have used the indirect approach of model parameter regionalization in which a model is calibrated to a large number of gauged watersheds and regression equations between its parameters and physical watershed characteristics are derived [e.g., Weeks and Ashkanasy, 1985; Jakeman et al., 1992; Post et al., 1998; Sefion and Howarth, 1998; Abdulla and Lettenmaier, 1997; Seibert, 1999; Fernandez et al., 2000; Merz and Blöschl, 2004; Wagener et al., 2004; McIntyre et al., 2005; Wagener and Wheater, 2006; Blöschl, 2005; Parajka et al., 2007]. These regression equations can then be used to derive a priori parameter estimates at ungauged locations. Other studies investigated the value of physical watershed similarity or spatial proximity to guide a priori parameter selection, but showed mixed results [e.g., Vandewiele and Elias, 1995; Merz and Blöschl, 2004; McIntyre et al., 2005]. The drawbacks of both direct and indirect approaches to ungauged basin parameter estimation have been discussed in detail elsewhere and we will not review this discussion here [Beven, 2001, 2006; Wagener and Wheater, 2006]. In general, many of these studies concluded that a priori
estimates are often not sufficiently reliable and should be treated as very uncertain.

[5] An alternative to the derivation of a priori model parameter estimates from physical watershed characteristics lies in the regionalization of streamflow characteristics. Prior studies have found that the correlations between streamflow characteristics and physical watershed characteristics are often significantly higher than between model parameters and watershed characteristics [e.g., Poff et al., 2006a, 2006b; Yadav et al., 2007]. Although the correlations between streamflow and watershed characteristics do not directly provide predictions of the continuous streamflow hydrograph, regionalized flow characteristics provide dynamic aspects of the watershed system that can be used to constrain hydrologic model predictions. Yadav et al. [2007] revised the regionalization of flow characteristics by including estimates of uncertainty in the regression equations. Uncertainty ranges, in addition to the expected values of flow characteristics, are valuable sources of information for constraining a model and thus for providing ensemble predictions at ungauged locations. Regionalizing flow characteristics has several advantages compared to model parameter regionalization.

[5] 1. The relationships between hydrologic indices and physical watershed characteristics are not obscured by model structural error or an ill-defined calibration problem as in the case of model parameter regionalization.

[6] 2. The regionalized indices are independent of any watershed model. The information can therefore be used to constrain any model.

[7] 3. The approach also allows for the inclusion of commonly available information on hydrologic flow characteristics, e.g., regionalized flood or low-flow frequency characteristics.

[8] In this paper, we contribute a flexible problem formulation strategy that advances our ability to consider increasing numbers of hydrologic indices and their associated uncertainty, as well as more complex models. The problem formulation advances, for the first time, our ability to apply multiobjective global optimization algorithms to enhance ensemble predictions of continuous streamflow within ungauged basins. We test this approach on 30 watersheds located in the UK using a typical lumped watershed model. Regionalized uncertainties are derived using the approach suggested by Yadav et al. [2007]. In the following sections, we present the regionalization approach, the novel problem formulation strategy and its multiobjective solution using evolutionary optimization.

2. Methods

2.1 Regionalization of Hydrologic Indices

[9] The regionalization of hydrologic indices as constraints has been introduced by Yadav et al. [2007] using the following procedural steps.

[10] 1. Extract hydrologic indices of interest from historical streamflow observations for many watersheds. Hydrologic indices can include flow characteristics such as runoff ratio, time to peak, mean flow, etc.

[11] 2. Regionalize hydrologic indices through multiple linear regressions against available physical watershed and climatic characteristics.

[12] 3. Calculate standard (90%) prediction limits for regression equations on the basis of assumptions of normality of residuals.

[13] 4. At the location without streamflow observation, run a selected hydrologic watershed model in a Monte Carlo framework in which parameters are sampled from independent uniform distributions. These distributions require definition of upper and lower bounds for each parameter, which could be either generally feasible ranges or a priori feasible ranges for the specific location.

[14] 5. Calculate the earlier chosen hydrologic indices for each simulation produced by the sampled parameter sets.

[15] 6. The prediction limits of the regression equations provide the expected range for individual hydrologic indices. Compare the locally simulated indices with those ranges. Parameter sets producing indices that fall inside the prediction limits are considered behavioral, all other parameter sets are considered nonbehavioral and rejected as possible representations of the watershed at hand.

[16] 7. Use all behavioral simulations to create ensemble predictions in the ungauged basin.

[17] Yadav et al. [2007] found that three hydrologic indices were needed (simultaneously) to attain sufficient ensemble spread for reliable predictions for the thirty UK watersheds. The three indices were high pulse count, defined as the number of annual occurrences during which flow exceeds 3 times the median daily flow, runoff ratio, defined as average annual runoff divided by average annual precipitation, and slope of the flow duration curve (FDC), defined as slope of the FDC between 33% and 66% flow exceedance values of streamflow. Topographic watershed slope and wetness ratio (long-term average precipitation over long-term average potential evapotranspiration) were used to predict runoff ratio. BFHOST (a measure based on soil type and hydrogeology) was the predictor variable for the flow duration curve slope. The combination of BFHOST, topographic slope and wetness ratio (the latter two predictors were not always strongly correlated) is needed to predict the high pulse count.

[18] While the study by Yadav et al. [2007] provides a proof of concept, their use of a Monte Carlo sampling approach limits the flexibility and scalability of the methodology to parsimonious hydrologic models and three hydrologic indices and is thus similar to the state of hydrologic model calibration before the introduction of powerful global optimization algorithms. Larger numbers of hydrologic indices pose increasingly more severe behavioral constraints and consequently increase the computational demands for Monte Carlo sampling. This problem is only exacerbated with increasing model complexity, which would result in a much higher dimensionality of the parameter space to be searched and also mean longer run times per model evaluation. We believe that the number of constraints ultimately will remain limited since it will otherwise lead to an overconditioning of the parameter space and might even include a potential conflict between constraints, which cannot be resolved. Thus leading to the problem of potentially finding no behavioral solution, which would be equivalent to a rejection of all models as sometimes done by others in gauged basins [Beven, 2006]. A new problem formulation and a new – more efficient – solution strategy are required to address these issues.
2.2. Multiobjective Calibration Formulation

[19] The main contribution of our research is a new formulation of the model identification problem for ungauged (with respect to the system response of interest) basins. The reformulation creates a multiobjective search problem that can be solved by multiobjective optimization tools to identify an ensemble of models consistent with a larger number of regionalized constraints. Consistent (or behavioral) means that the simulated hydrologic indices of these models fall within the regionalized ranges. The intent of this methodology is to provide a new approach to utilizing regional information for model identification and in this way enhance the size and complexity of ungauged prediction problems that can be solved reliably. Mathematically, evolutionary multiobjective optimization seeks to quantify optimal trade-offs between conflicting optimization objectives. The term “optimal trade-offs” refers to Pareto optimal or nondominated solution sets where one of a solution’s objective values cannot be improved without degrading its performance in one or more of the remaining objectives. Equation (1) reformulates the search for models satisfying regional constraints as an optimization of multiple objectives.

[20] All of the hydrologic indices are normalized to range between −1 and 1 to enhance the search by enforcing scaling invariance across the objectives (i.e., one objective with a large range or variability does not dominate search). Figure 1 illustrates the use of 3 hydrologic indices building on the results of Yadav et al. [2007]. The transformation of N hydrologic constraints to a multiobjective search formulation requires N+1 objective functions. The first N objectives minimize the distance to the centers of the prediction limits for the indices and are defined in equation (1) for \( j = 1, 2, \ldots, N \),

\[
\text{Minimize } f_j = \left| I_j - I_{cj} \right| / \delta_j \tag{1}
\]

where \( I_j \) is the jth hydrologic index calculated on the basis of the simulation; \( I_{cj} \) is the center of the prediction limit of the jth hydrologic index, (i.e., the expected value in the regression), and \( \delta_j \) is the spread of the prediction limit of the jth hydrologic index, (i.e., the range from its minimum to its maximum). The normalization of the indices means their expected values have been transformed to be equal to zero. Geometrically, when three indices are used (see Figure 1b) equation (1) seeks to minimize each of the model predicted indices’ distances from their expected values that form the (0,0,0) origin of the unit sphere. Yadav et al. [2007] assume a Gaussian distribution for the regression-based predictions, even though they only consider lower and upper limits. Equation (1) biases multiobjective search toward the mean of the regionalized indices.

[21] The objective shown in equation (2) is formulated to explicitly conflict with the N objectives of equation (1) and to yield a Pareto set. Equation (2) forces the multiobjective search to identify behavioral parameter groups at the outer extremes of the indices’ ranges (i.e., maximally distant from origin of sphere in Figure 1). In equation (2) the objective minimizes the distance from a candidate behavioral parameter set’s model simulated hydrologic index value (\( I_j \)) to the

---

**Figure 1.** Visualization of the problem formulation. (a) For a single hydrologic index, \( I_{cj} \) is the center of the prediction limit, and \( \delta_j \) is the symmetric distance from the center of the prediction limit to either the minimum or the maximum index value for the prediction range. After standardization, the center of the prediction limit is 0, and the prediction limit ranges from −1 to 1. (b) For three hydrologic indices, all behavioral simulations reside inside of a standardized sphere with the center of the prediction limit at \([0, 0, 0]\).
outer edges of the regionalized indices ranges as shown below:

\[
\text{Minimize } f_A = 1 - \frac{1}{N} \sum_{j=1}^{N} \left( \frac{I_j - I_{\text{low}}}{I_{\text{high}} - I_{\text{low}}} \right)
\]

In combination, equations (1) and (2) present a novel formulation for the prediction in ungauged basins problem and lead to the identification of behavioral parameter sets over the full extent of the indices-space sphere as shown in Figure 1 for a three-indices problem. This formulation creates a spherical behavioral subspace within the feasible three-dimensional index space (Figure 1b). The spherical solution space excludes some of the feasible area if the full ranges for each index would be considered; that is, the full space would be a cube, thus implicitly including the assumption that the “edges” for the expected regionalized indices are less likely to contain good parameter sets. The decision space for the multiobjective problem formulation shown in equations (1) and (2) is the parameter space of any hydrologic models employed.

2.3. Multiobjective Evolutionary Search

[22] Recent studies have shown that evolutionary multiobjective optimization algorithms can facilitate the solution of water resources problems with several conflicting objectives [Bekele and Nicklow, 2005; Frugt et al., 2003; Kollat and Reed, 2006; Tang et al., 2007]. These methods use population-based search operators that mimic Darwinian natural selection to search for and approximate the full Pareto optimal set. The Pareto optimal set is composed of solutions in which an improvement in a single objective requires a corresponding reduction of performance in other objectives (i.e., trade-offs). Plotting the Pareto optimal set yields the Pareto frontier, which is maximally a (M-1) dimensional surface given M objectives. Evolutionary multiobjective methods exploit their population-based search to provide full approximations to the Pareto frontier in a single algorithm run. In this study, the four-objective calibration problem shown in equations (1) and (2) is solved using the epsilon dominance archiving Nondominated Sorted Genetic Algorithm II (\(\varepsilon\)-NSGAII). The \(\varepsilon\)-NSGAII has been shown to be an efficient and effective multiobjective solution tool for water resources applications [Tang et al., 2007].

[23] Kollat and Reed [2006] developed the \(\varepsilon\)-NSGAII algorithm as an extension of the original NSGA-II algorithm [Deb et al., 2002] by adding \(\varepsilon\)-dominance archiving [Laumanns et al., 2002] and automatic parameterization to enhance the solution of high-order Pareto optimization problems (i.e., problems with three or more objectives). For this study, the \(\varepsilon\)-NSGAII provides (1) an excellent ability to maintain solution diversity for the full subspace of interest, (2) reliable search dynamics based on its adaptive population sizing and epsilon dominance archiving, and (3) minimum user inputs for its search parameters Epsilon dominance archiving requires that users specify \(\varepsilon\) values that represent their precision or resolution requirements for optimizing each objective. The \(\varepsilon\) values serve to discretize the objective space into user specified \(\varepsilon\) blocks. In terms of the problem formulation given in equations (1) and (2) these blocks provide hypercubes across the sphere defining the behavioral indices space (Figure 1). In this case the \(\varepsilon\)-NSGAII’s \(\varepsilon\)-dominance archiving stores a single Pareto optimal solution for each \(\varepsilon\) block (for details see Kollat and Reed [2007]). The epsilon dominance archiving grid bounds the Pareto optimal set size, which is important for this application since the search problem for ensemble streamflow predictions can theoretically have an infinite number of Pareto optimal solutions. Moreover, Kollat and Reed [2007] have demonstrated that epsilon dominance archiving can be used to directly control the resolution of large Pareto optimal sets (i.e., behavioral parameter ensembles in this study) as well as the computational demands associated with their identification.

3. Case Study

3.1. Data Overview and Regionalization of Indices

[24] A set of 30 small to medium-sized (~50–1100 km²) watersheds distributed throughout England and Wales are employed in this study. These watersheds cover a range of watershed characteristics including soil type, topography and land use, as well as climate (Table 1). Time series of daily precipitation, potential evapotranspiration and streamflow are available for a 10-year period. All results shown later are based on running the model for the full 10-year period, though only shorter periods are plotted for visual evaluation. Further details of the watersheds can be found in the work by Yadav et al. [2007] and are not repeated here.

[25] All 30 watersheds were employed to develop the regional regression relations between hydrologic indices and watershed characteristics using a sixfold cross validation experiment by Yadav et al. [2007]. The 30 watersheds were broken into six groups of five watersheds. Twenty five watersheds were used in turn to estimate regression relationships for the three regionalized hydrologic indices in the five “ungauged” watersheds. No locally observed streamflow information was therefore considered to estimate the indices ranges. The ranges for the 3 hydrologic indices are defined as the prediction limits of the regional regression equations for high pulse count, runoff ratio, and slope of flow duration. Figures 2a–2c show the regionalized ranges as well as the observed values for the three indices. Figure 2 shows that most of the observed indices fall within the regionalized ranges (more than two thirds for each of the three cases), and only a few observed indices deviate considerably from these expected intervals.

[26] From the 30 watersheds, 3 watersheds were selected as initial test cases as they represent three different scenarios for the quality of the hydrologic indices regionalization, i.e., good hydrologic indices regionalization (Kirkbymills watershed, watershed 13 in Figure 2), medium hydrologic indices regionalization (Harrowdown Old Mill watershed, watershed 26 in Figure 2), and poor hydrologic indices regionalization (Shalford watershed, watershed 30 in Figure 2). For the Kirkbymills watershed, all three hydrologic indices (high pulse count, runoff ratio, and slope of flow duration curve) are within in the prediction limits provided by hydrologic indices regionalization. For the Harrowdown Old Mill watershed, two hydrologic indices (runoff ratio and slope of flow duration curve) are captured by the prediction limits. For the Shalford watershed, only one hydrologic index (high pulse count) falls within its
prediction limits, while the prediction limits for high pulse count are very narrow. The narrow prediction limits for high pulse count are the result of a major part of the prediction interval being below 0 and therefore producing infeasible values for high pulse count. The negative portion results from the purely empirical regression without physical constraints. The proposed multiobjective search formulation for behavioral parameter sets is demonstrated in detail for the three selected watersheds. Results for the remaining 27 watersheds are also reported for completeness. Of course, in the truly ungauged case outside a scientific study, one would not a priori know whether a regionalization worked well or not, which makes it even more important to have a robust procedure for model identification.

3.2. Hydrologic Model

[27] To enable comparison with results by Yadav et al. [2007], this study uses the same lumped watershed model, HYMOD, with five parameters. HYMOD consists of a soil moisture accounting component which represents storage using a Pareto probability distribution for storage capacities. Runoff is produced by overflow from these storages and subsequently split into a parallel routing component. The routing component consists of two series of linear reservoirs representing quick flow and slow flow response, respectively. Details of the HYMOD model parameters are summarized in Table 2. Parameter ranges were set on the basis of an initial sensitivity analysis. Input data required by the model are mean areal precipitation and evapotranspiration. The model has been widely used for hydrologic research, readers interested in further details can see Boyle et al. [2000] or Wagener et al. [2001]. The model requires the optimization of 5 conceptual parameters.

3.3. Overview of Computational Experiments

[28] The decision variables for the multiobjective calibration formulation shown in equations (1) and (2) are the five parameters of HYMOD. The search for behavioral simulations is transferred to a four-objective, five-decision variable multiobjective calibration problem. The model is run for a 10-year period for the analyses shown. However, a 1-year warm-up period is excluded when evaluating the model performance.

[29] The Monte Carlo uniform random sampling used by Yadav et al. [2007] is the benchmark used to test the new multiobjective formulation for behavioral sets (and its solution) presented here. The comparative results were developed by first analyzing the Kirkby mills, Harrowdown Old Mill, and Shalford watersheds, which represent a range from good to poor performance for regionalized hydrologic indices. In order to compare the $\varepsilon$-NSGAII and Monte Carlo approaches for these three watersheds thoroughly, 50 random seed trial runs are performed and the statistics based on the 50 random runs were analyzed. For each seed run, the total number of function evaluations (NFE) is set to 100,000 for both $\varepsilon$-NSGAII and Monte Carlo sampling. For the $\varepsilon$-NSGAII, the most important algorithm parameter is then $\varepsilon$ value for each of the objective functions. As each objective is based on a standardized hydrologic index, the impacts of unit and magnitude of different indices are minimal. The $\varepsilon$ values for all objectives are set to be 0.01 here. Changing the value of $\varepsilon$ would alter the density with which the index

<table>
<thead>
<tr>
<th>Watershed Number</th>
<th>Station Number</th>
<th>River</th>
<th>Watershed</th>
<th>Area (km²)</th>
<th>BFI</th>
<th>P/PE</th>
<th>P (mm a⁻¹)</th>
<th>Q (mm a⁻¹)</th>
<th>PE (mm a⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27035</td>
<td>Aire</td>
<td>Kildwick Bridge</td>
<td>282.3</td>
<td>0.369</td>
<td>2.021</td>
<td>1191.7</td>
<td>769.0</td>
<td>589.8</td>
</tr>
<tr>
<td>2</td>
<td>25006</td>
<td>Greta</td>
<td>Rutherford Bridge</td>
<td>86.1</td>
<td>0.209</td>
<td>2.090</td>
<td>1279.7</td>
<td>864.3</td>
<td>612.4</td>
</tr>
<tr>
<td>3</td>
<td>52010</td>
<td>Brue</td>
<td>Lovingston</td>
<td>155.2</td>
<td>0.473</td>
<td>1.466</td>
<td>845.6</td>
<td>421.0</td>
<td>576.6</td>
</tr>
<tr>
<td>4</td>
<td>50001</td>
<td>Taw</td>
<td>Umberleigh</td>
<td>826.2</td>
<td>0.424</td>
<td>2.052</td>
<td>1245.9</td>
<td>725.0</td>
<td>607.1</td>
</tr>
<tr>
<td>5</td>
<td>24004</td>
<td>Bedburn Beck</td>
<td>Bedburn</td>
<td>74.9</td>
<td>0.464</td>
<td>1.452</td>
<td>886.5</td>
<td>525.8</td>
<td>610.5</td>
</tr>
<tr>
<td>6</td>
<td>47008</td>
<td>Thrushel</td>
<td>Tinhay</td>
<td>112.7</td>
<td>0.385</td>
<td>1.917</td>
<td>1235.7</td>
<td>868.1</td>
<td>644.7</td>
</tr>
<tr>
<td>7</td>
<td>67018</td>
<td>Dee</td>
<td>New Inn</td>
<td>53.9</td>
<td>0.273</td>
<td>3.470</td>
<td>2208.5</td>
<td>1868.2</td>
<td>636.4</td>
</tr>
<tr>
<td>8</td>
<td>60010</td>
<td>Tywi</td>
<td>Nantgarredig</td>
<td>1090.4</td>
<td>0.478</td>
<td>2.970</td>
<td>1911.9</td>
<td>1146.3</td>
<td>643.7</td>
</tr>
<tr>
<td>9</td>
<td>23006</td>
<td>South Tyne</td>
<td>Featherstone</td>
<td>321.9</td>
<td>0.329</td>
<td>2.396</td>
<td>1528.4</td>
<td>1102.4</td>
<td>637.9</td>
</tr>
<tr>
<td>10</td>
<td>22001</td>
<td>Coquet</td>
<td>Morwic</td>
<td>569.8</td>
<td>0.487</td>
<td>1.964</td>
<td>878.7</td>
<td>460.6</td>
<td>447.5</td>
</tr>
<tr>
<td>11</td>
<td>62001</td>
<td>Teifi</td>
<td>Glen Teifi</td>
<td>893.6</td>
<td>0.532</td>
<td>2.218</td>
<td>1405.9</td>
<td>1017.8</td>
<td>633.8</td>
</tr>
<tr>
<td>12</td>
<td>28046</td>
<td>Dove</td>
<td>Izac Walton</td>
<td>83.0</td>
<td>0.783</td>
<td>1.782</td>
<td>1143.1</td>
<td>787.2</td>
<td>641.7</td>
</tr>
<tr>
<td>13</td>
<td>27042</td>
<td>Dove</td>
<td>Kirkby mills</td>
<td>59.2</td>
<td>0.595</td>
<td>1.867</td>
<td>959.8</td>
<td>578.6</td>
<td>514.0</td>
</tr>
<tr>
<td>14</td>
<td>45005</td>
<td>Otter</td>
<td>Doton</td>
<td>202.5</td>
<td>0.538</td>
<td>1.511</td>
<td>975.3</td>
<td>463.9</td>
<td>645.5</td>
</tr>
<tr>
<td>15</td>
<td>25005</td>
<td>Leven</td>
<td>Leven Bridge</td>
<td>196.3</td>
<td>0.432</td>
<td>1.127</td>
<td>708.5</td>
<td>292.4</td>
<td>628.4</td>
</tr>
<tr>
<td>16</td>
<td>53017</td>
<td>Boyd</td>
<td>Bitton</td>
<td>47.9</td>
<td>0.459</td>
<td>1.182</td>
<td>809.7</td>
<td>380.3</td>
<td>648.7</td>
</tr>
<tr>
<td>17</td>
<td>54016</td>
<td>Roden</td>
<td>Rodington</td>
<td>259.0</td>
<td>0.609</td>
<td>1.304</td>
<td>705.4</td>
<td>235.5</td>
<td>541.1</td>
</tr>
<tr>
<td>18</td>
<td>55014</td>
<td>Lugg</td>
<td>Byton</td>
<td>203.3</td>
<td>0.668</td>
<td>1.544</td>
<td>1035.3</td>
<td>623.8</td>
<td>670.7</td>
</tr>
<tr>
<td>19</td>
<td>22006</td>
<td>Blyth</td>
<td>Hartford Bridge</td>
<td>269.4</td>
<td>0.460</td>
<td>1.241</td>
<td>710.7</td>
<td>245.0</td>
<td>572.5</td>
</tr>
<tr>
<td>20</td>
<td>37005</td>
<td>Cole</td>
<td>Lexden</td>
<td>238.2</td>
<td>0.526</td>
<td>0.773</td>
<td>585.6</td>
<td>146.0</td>
<td>757.3</td>
</tr>
<tr>
<td>21</td>
<td>42004</td>
<td>Test</td>
<td>Broadlands</td>
<td>1035.9</td>
<td>0.898</td>
<td>1.255</td>
<td>807.9</td>
<td>319.9</td>
<td>643.9</td>
</tr>
<tr>
<td>22</td>
<td>31010</td>
<td>Chater</td>
<td>Fosters Bridge</td>
<td>68.9</td>
<td>0.511</td>
<td>0.871</td>
<td>641.0</td>
<td>238.4</td>
<td>726.0</td>
</tr>
<tr>
<td>23</td>
<td>38003</td>
<td>Mimram</td>
<td>Panshanger Park</td>
<td>133.9</td>
<td>0.937</td>
<td>0.858</td>
<td>638.4</td>
<td>127.7</td>
<td>744.0</td>
</tr>
<tr>
<td>24</td>
<td>33029</td>
<td>Stringside</td>
<td>Whitebridge</td>
<td>98.8</td>
<td>0.857</td>
<td>0.914</td>
<td>632.7</td>
<td>151.4</td>
<td>692.2</td>
</tr>
<tr>
<td>25</td>
<td>43006</td>
<td>Nadder</td>
<td>Wilton</td>
<td>220.6</td>
<td>0.813</td>
<td>1.188</td>
<td>858.2</td>
<td>397.3</td>
<td>722.6</td>
</tr>
<tr>
<td>26</td>
<td>32004</td>
<td>Ise brook</td>
<td>Harrowdown Old Mill</td>
<td>194.0</td>
<td>0.551</td>
<td>0.889</td>
<td>645.1</td>
<td>221.3</td>
<td>725.2</td>
</tr>
<tr>
<td>27</td>
<td>39020</td>
<td>Colin</td>
<td>Bibury</td>
<td>106.7</td>
<td>0.938</td>
<td>1.087</td>
<td>808.8</td>
<td>405.2</td>
<td>744.0</td>
</tr>
<tr>
<td>28</td>
<td>29003</td>
<td>Ludd</td>
<td>Louth</td>
<td>55.2</td>
<td>0.899</td>
<td>1.052</td>
<td>674.8</td>
<td>282.6</td>
<td>641.7</td>
</tr>
<tr>
<td>29</td>
<td>48003</td>
<td>Fal</td>
<td>Tregony</td>
<td>87.0</td>
<td>0.694</td>
<td>2.136</td>
<td>1267.9</td>
<td>715.0</td>
<td>593.7</td>
</tr>
<tr>
<td>30</td>
<td>39029</td>
<td>Tillingbourne</td>
<td>Shalford</td>
<td>59.0</td>
<td>0.888</td>
<td>1.094</td>
<td>803.7</td>
<td>294.9</td>
<td>734.4</td>
</tr>
</tbody>
</table>

*P, Precipitation; PE, potential evapotranspiration; Q, streamflow; BFI, base flow index.
Table 2. Description of Parameters for the Lumped Watershed Model Used

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Unit</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>C&lt;sub&gt;max&lt;/sub&gt;</td>
<td>Maximum storage capacity</td>
<td>mm</td>
<td>1</td>
<td>500</td>
</tr>
<tr>
<td>b</td>
<td>Index describing distribution of storage capacity</td>
<td></td>
<td>0.1</td>
<td>2</td>
</tr>
<tr>
<td>α</td>
<td>Flow split coefficient (fraction quick flow)</td>
<td></td>
<td>0.1</td>
<td>0.99</td>
</tr>
<tr>
<td>K&lt;sub&gt;q&lt;/sub&gt;</td>
<td>Time constant of linear quick flow reservoir</td>
<td>d&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>0.1</td>
<td>0.99</td>
</tr>
<tr>
<td>K&lt;sub&gt;s&lt;/sub&gt;</td>
<td>Time constant of linear slow flow reservoir</td>
<td>d&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>0</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Figure 2. Prediction limits of regional regression equations and observed hydrologic indices for each of the 30 watersheds (the shaded bar is the prediction limit and the asterisk is the corresponding observed hydrologic index): (a) high pulse count, (b) runoff ratio, and (c) slope of flow duration curve. Watersheds 13, 26, and 30 are the Kirkey mills, Harrowdown Old Mill, and Shalford watersheds, respectively, which are the selected watersheds for detailed analysis in the study. The order of watersheds is the same as the order in Table 2 and Figure 6. The watersheds are sorted in descending order by the number of behavioral simulation found through Monte Carlo sampling, thus expressing the difficulty of finding behavioral simulations. The search is becoming more and more difficult from watershed 1 to watershed 30.
space is filled with behavioral solutions. The other $\varepsilon$-NSGAII parameters, including the probability of mating, the probability of mutation, and the initial population size, are set on the basis of the recommendations of Tang et al. [2006].

[30] On the basis of the initially detailed analysis of the Kirkbymills, Harrowdown Old Mill, and Shalford watersheds, the second set of $\varepsilon$-NSGAII and Monte Carlo sampling experiments were refined for the remaining 27 watersheds. The total number of search simulations was reduced to 30,000 and both approaches are run with a single random seed only, since little variability was observed for the 3 selected watersheds’ results evaluated with 50 seeds.

3.4. Evaluation Metrics

[31] The primary evaluative metric used in this study quantifies the number of behavioral simulations found through $\varepsilon$-NSGAII and Monte Carlo sampling as a function of the overall number of model runs executed. The reliability and sharpness of the behavioral ensembles, as introduced by Yadav et al. [2007], are the hydrologic prediction metrics for evaluating the behavioral ensembles. By definition, reliability is the fraction of time during which the observed streamflow lies within the prediction range of the ensemble. At the same time, the reduction of uncertainty of prediction range is measured by sharpness. The ratio of the prediction range of the streamflow ensemble produced by sampling from the behavioral predictions to the range produced by the a priori feasible parameters is calculated, and sharpness is defined as 1 minus the ratio. A perfect prediction with a single parameter set (and therefore a single hydrograph) would result in a value of 100% for both reliability and sharpness.

4. Results

[32] The results are based on the detailed analysis of three test watersheds (Kirkbymills, Harrowdown and Shalford) and the less detailed analysis of all 30 watersheds. We compared Monte Carlo and multiobjective formulations to demonstrate the efficiency and value of the new approach, while discussing differences in the resulting reliability and sharpness as defined below. The results presented for the three test watersheds are based on the average performance calculated for 50 random trial runs for both multiobjective search and Monte Carlo analysis. A sensitivity analysis has been included to enable evaluation of why the difficulty in identifying behavioral parameter sets varies between watersheds.

4.1. Kirkbymills Watershed: Good Regionalization Case

[33] The results of both $\varepsilon$-NSGAII and Monte Carlo analyses for the Kirkbymills watershed are plotted in Figure 3. Figure 3a shows the number of behavioral simulations found through $\varepsilon$-NSGAII and Monte Carlo uniform random sampling versus the total number of function (model) evaluations (NFE) executed. The Monte Carlo approach finds on average about 8,000 behavioral simulations after a total of 100,000 simulations, while the $\varepsilon$-NSGAII approach finds approximately 17,000 behavioral simulations after a total of 100,000 simulations.

[34] The Monte Carlo approach shows an approximately linear rate in identifying behavioral models due to the uniform random sampling applied (remember that the plot is on log scale), though the slope of that linear rate will differ between watersheds. Figure 3a shows that $\varepsilon$-NSGAII identifies behavioral models at a higher than linear rate. For the first 10,000 simulations evaluated, the $\varepsilon$-NSGAII and Monte Carlo find 20% and 8% of the evaluated simulations to be behavioral, respectively. The Kirkbymills watershed test case suggests that the multiobjective search methodology proposed in this study is likely to be more effective than a Monte Carlo analysis even if the search problem should be relatively easy as judged by the good regionalization result, which of course would not be known if the watershed was truly ungauged.

[35] Figures 3b and 3c show the average reliability (fraction of observations captured by ensemble) and sharpness (fraction of uncertainty reduction of constrained versus unconstrained ensemble) for high, medium, and low flows for $\varepsilon$-NSGAII for the Kirkbymills watershed. High, medium and low flows are defined on the basis of the 25 and 75 flow percentiles. It can be seen that both reliability and sharpness start to converge approximately at an NFE of 30,000. The reliabilities for medium and low flows are close to 1 (meaning almost 100% of the observations are captured) and the reliability for high flows is about 88%. The sharpness values for high, medium, and low flows are centered around 0.5, which represents a reduction of predictive uncertainty by 50%. Consider that while reliability is valuable for assessing the prediction ensemble performance, it is currently only a binary performance metric, which means the reliability is zero for a time step when the observation falls outside the ensemble range, even if it might only fall outside by a small margin. The ensemble hydrograph ranges and the observed hydrograph at the Kirkbymills watershed are shown in Figure 3d to allow visual examination of the quality of the result. One can see a significant uncertainty reduction while the observations are still captured.

4.2. Harrowdown Old Mill Watershed: Medium Regionalization Case

[36] For the Harrowdown Old Mill watershed the observed runoff ratio and the observed slope of the flow duration curve hydrologic indices both fall within the prediction limits provided by the regionalization (see Figures 2a–2c). The comparison of the $\varepsilon$-NSGAII and Monte Carlo sampling’s dynamic rates for finding behavioral simulations is shown in Figure 4a. After evaluating 100,000 simulations, the average number of behavioral simulations found through Monte Carlo sampling and through the $\varepsilon$-NSGAII are 2,700 and 17,000, respectively. The primary difference between the Harrowdown and Kirkbymills case studies lies in the relative rates in which Monte Carlo analysis and $\varepsilon$-NSGAII identify behavioral parameter sets. After 10,000 model evaluations Figure 3a shows that 1.5% of the Monte Carlo simulations were behavioral versus 15% of the $\varepsilon$-NSGAII’s simulations, thus indicating that this case is more difficult than the Kirkbymills study.
The average reliability and sharpness for high, medium, and low flows for the \( \varepsilon \)-NSGAII generated behavioral ensemble are shown in Figures 4b and 4c. It can be seen that the reliabilities for medium flows and low flows (close to 100%), and the sharpness values for medium flows and low flows are high. On the other hand, both reliability and sharpness for high flows are relatively low compared to the corresponding reliability and sharpness of the other flow regimes. This is likely a consequence of the observed high pulse count (constraining high flows) falling outside of the prediction limit provided by the regionalization. During visual examination of simulated and observed time series in Figure 4d it can be seen that the ensemble represents the dynamics of the observed hydrograph especially for medium and low flow periods.

4.3. Shalford Watershed: Poor Regionalization Case

In the Shalford watershed, only the observed high pulse count falls within its regionalized prediction limits, while the other two observed hydrologic indices plot outside (see Figures 2a–2c). Figure 5a clearly shows that the Shalford test case is far more challenging to solve in terms
of the relative rates for identifying behavioral parameter sets for both Monte Carlo and \( \varepsilon \)-NSGAII analyses. The Monte Carlo sampling approach only identified 30 behavioral simulations out of the 100,000 simulations evaluated. Despite this, \( \varepsilon \)-NSGAII was still able to identify approximately 10,000 behavioral simulations. For this test case \( \varepsilon \)-NSGAII maintained an average rate of identifying 1 behavioral set for every 10 simulations evaluated (10%).

[19] Reliability and sharpness values are shown in Figures 5b and 5c, respectively. It is not surprising that all of the reliabilities are lower than before. The high-flow predictions have the overall highest reliability of 0.7, which is expected since high pulse count was the only hydrologic index that fell within the predictions limits attained from the regional regressions. It is also noted that all of the sharpness values are very high. This is because the prediction limits for high pulse count are very narrow, though they do capture the observed high pulse count in the Shalford watershed. It can be seen from Figure 5d that the ensemble predictions represent the dynamics of the observed hydrograph. The deviation of the ensemble predictions from the observed hydrograph is not signifi-
cant, which is not well reflected in the binary reliability measure though. With a difficult case like the Shalford watershed, the \( \varepsilon \)-NSGAII approach is considerably more efficient than Monte Carlo sampling.

### 4.4. Reasons for Differences in Identification Efficiency

Figure 6 provides plots of the cumulative distributions of the behavioral parameter populations to better illustrate potential reasons for the differences in the difficulty of identifying behavioral models. These plots are similar to the Regional Sensitivity Analysis [Hornberger and Spear, 1981] approach as used by Freer et al. [1996] or Wagener et al. [2001]. The plots of Figure 6 visualize how the three different hydrologic indices constrain the parameter space individually and together. A steep cumulative distribution indicates that the behavioral parameter sets only cover a small part of the feasible parameter space and consequently that the parameter identification search problem will be more difficult.

The top row of Figure 6 shows how the runoff ratio (RR) constrains the water balance parameters of the HYMOD model. The amount of evapotranspiration is
defined by the soil moisture storage size, which in turn is defined by parameters \( C_{\text{max}} \) and \( b \). The three routing parameters are insensitive and their distributions are uniform across the feasible parameter ranges. In the second row of Figure 6 the parameter distributions are constrained by the slope of the flow duration curve (SFDC) solely. One can see that the resulting cumulative distributions are considerably different from the top row. Parameters \( b \) and \( k_q \) are insensitive, while parameter \( k_s \) seems severely constrained for the Shalford and Harrowdown watersheds, which both represent very difficult search cases. The constraint seems to mainly act on the parameters that distribute the flow volume. The high pulse count (HPC) impacts all parameters, but in particular alpha and \( k_s \) for the Shalford and Harrowdown watersheds. Again the more difficult watersheds have much narrower solution spaces for several of the parameters. The bottom row results in Figure 6 show the combined impact of using all of the indices and nicely visualizes that the behavioral cumulative parameter distributions are much steeper for several parameters for the Shalford and Harrowdown watersheds. KirkbyMills remains closer to uniform distributions, thus suggesting that identifying behavioral parameter sets for this watershed should be a much easier task.

4.5. Overall Result of the Remaining Watersheds

[42] For the remaining 27 watersheds a total of 30,000 simulations and a single random seed were used to compare \( \varepsilon \)-NSGAII and Monte Carlo sampling. On the basis of the prior results, 30,000 NFEs are sufficient to show differences between optimization-based and random sampling. A single seed is appropriate since little variability was observed in the results for the three watersheds analyzed in the previous sections. The number of behavioral simulations found using \( \varepsilon \)-NSGAII and Monte Carlo sampling are shown in Figure 7. The watersheds on the \( x \) axis are ranked according to the number of behavioral simulations found through Monte Carlo sampling since it provides an indicator of search
problem difficulty (the same order is used in Figure 2). For watersheds 1 and 2, the number of behavioral simulations found by both methods is close to 10,000, which reflects a 30% success rate in identifying behavioral parameter sets (i.e., 10,000 out of a total of 30,000 evaluated). For those two watersheds, the \(\varepsilon\)-NSGAII appears to find fewer behavioral parameter sets than the Monte Carlo analysis, though this result is a consequence of how the \(\varepsilon\)-NSGAII “thins” the final solutions set to be equal to or less than the number of cells in the user-specified epsilon-dominance grid (i.e., the algorithm finds a very consistent number of solutions).

[43] The \(\varepsilon\)-NSGAII’s overall performance is fairly consistent for all 30 watersheds despite the great variability in search problem difficulty. The algorithm’s success rate for identifying behavioral parameter sets ranges from 30% for watershed 1 (the easiest problem) to 10% for watershed 30 (a difficult search problem). Figure 7 clearly shows that the uniform random sampling is severely impacted by problem difficulty leading to nearly complete failure to identify behavioral simulations for watershed 30 when limited to 30,000 model executions.

5. Discussion and Conclusions

[44] In this study we present a multiobjective framework to advance prediction in ungauged basins problems that generalizes our ability to use powerful optimization algorithms for model identification. The results shown in the previous sections demonstrate that the formulation using the \(\varepsilon\)-NSGAII optimization algorithm is more efficient than a Monte Carlo–based approach, especially for difficult cases. For example, while only about three model evaluations (executions) are needed for both \(\varepsilon\)-NSGAII and Monte Carlo approaches to identify one behavioral simulation for watershed 1 (Figure 7), eight are needed for \(\varepsilon\)-NSGAII in watershed 27, whereas the Monte Carlo approach needs to run the model 1,000 times to achieve the same result. For watershed 30 (Shalford), \(\varepsilon\)-NSGAII needs about 10 simulations to identify a behavioral simulation while the Monte Carlo sampling requires approximately 3,000 simulations.

The cumulative behavioral parameter distribution plots of Figure 6 visualize why certain search problems are harder than others. The larger the search behavioral solution space, and thus the more uniform the behavioral parameter distribution, the easier the search.

[45] The high efficiency and robustness of the \(\varepsilon\)-NSGAII in finding behavioral parameter sets indicates that (1) more complex (incl. distributed) hydrologic model decision spaces can be searched using the multiobjective prediction in ungauged basins (PUB) framework presented in this study, and (2) an increasing number of hydrologic indices can be tested for constraining ensemble predictions for ungauged basins. Reducing the uncertainty on predictions in ungauged basins is a current focal point of the hydrologic community [Sivapalan et al., 2003]. Ensemble predictions in ungauged basins derived from a priori feasible model parameter ranges often provide very uncertain streamflow estimates unless local historical observations of streamflow are available. The regionalization of hydrologic indices can be a successful way to extrapolate characteristics of the watershed response behavior to ungauged locations as constraints on any model. In this paper, we demonstrate that PUB can be viewed as a new class of multiobjective search problems. Reformulating the PUB problem as a multiobjective search for behavioral model parameter sets, generalizes our ability to test a wide range and number of hydrologic indices to maximize our use of the information available in streamflow data sets. The PUB problem formulation contributed in this study can be easily adapted to solve larger, more computationally demanding model identification problems. This will have particular value for future studies seeking to develop high fidelity, probabilistic scenarios for environmental change under uncertainty [Wagener, 2007].

[46] Acknowledgments. The first and third authors of this work were partially supported by the National Science Foundation under grants EAR-0418798, EAR-0609791, EAR-0609741, and CBET-0640443. Partial support for the second author was provided by SAHRA under NSF-STC grant EAR-064-9876800, the National Weather Service Office of Hydrology under grant numbers NOAA/NA04NWS4620012 and NOAA/DG 133W-03-666 SE-0916, and the USDI under grant USDI/USGS 432–41(69AR).
Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the writers and do not necessarily reflect the views of the National Science Foundation.

References


