

Performance of the dynamically dimensioned search algorithm: influence of parameter initialization strategy when calibrating a physically based hydrological model

Franck Lespinas, Ashu Dastoor and Vincent Fortin

ABSTRACT

This study presents an evaluation of the performance of the dynamically dimensioned search (DDS) algorithm when calibrating the hydrological component of the Visualizing Ecosystems for Land Management Assessments (VELMA) ecohydrological model. Two calibration strategies were tested for the initial parameter values: (1) a 'high-cost strategy', where 100 sets of initial parameter values were randomly chosen within the overall parameter space, and (2) a 'low-cost strategy', where a unique set of initial parameter values was derived from the available field data. Both strategies were tested for six different values of the maximum number of model evaluations ranging between 100 and 10,000. Results revealed that DDS is able to converge rapidly to a good parameter calibration solution of the VELMA hydrological component regardless of the parameter initialization strategy used. The accuracy and convergence efficiency of the DDS algorithm were, however, slightly better for the low-cost strategy. This study suggests that initializing the parameter values of complex physically based models using information on the watershed characteristics can increase the efficiency of the automatic calibration procedures.

Key words | automatic calibration, hydrological model, optimization algorithm, VELMA

Franck Lespinas (corresponding author)
Ashu Dastoor
Air Quality Research Division,
Environment and Climate Change Canada,
2121 Trans-Canada Highway, Dorval H9P 1J3,
Canada
E-mail: francklespinas@yahoo.fr

Vincent Fortin
Meteorological Research Division,
Environment and Climate Change Canada,
2121 Trans-Canada Highway, Dorval H9P 1J3,
Canada

INTRODUCTION

Most environmental watershed models have parameters that must be adjusted in order to reproduce the observed processes with good accuracy in the investigated watersheds. This process is called parameter optimization or model calibration, and can be implemented following two different/complementary approaches: manual and automatic. Manual calibration consists of adjusting the model parameters to capture trends and patterns in observed data through a visual trial and error procedure using knowledge of the watershed, site-specific observations, literature values as well as experience (e.g., [Knights et al. 2014](#)). Manual calibration is a useful learning exercise to understand the model characteristics and the mechanism of processes occurring in the studied watershed. However, manual calibration is also

subjective, and labor- and time-intensive ([Zhang et al. 2008](#)). Automatic calibration consists of using an optimization algorithm that searches for an optimal parameter set that minimizes the model output errors relative to the available measured data (e.g., [Tolson & Shoemaker 2007](#); [Lespinas et al. 2014](#)). These methods require defining an objective function that measures the error in model outputs with respect to the observation data, and selecting an optimization algorithm that optimizes (minimizes, most often) the selected objective function.

A global optimization algorithm, dynamically dimensioned search (DDS), was recently introduced by [Tolson & Shoemaker \(2007\)](#) for automatic calibration of complex watershed simulation models with a large number of

doi: 10.2166/nh.2017.139

parameters. Compared to most global optimization methods, DDS starts the parameter search from a unique initial parameter value set, requires essentially no parameter tuning and automatically scales the search to find the best attainable parameter combination within the user-specified maximum number of objective function evaluations (Tolson & Shoemaker 2007). Using the Soil Water Assessment Tool (SWAT), a popular hydrological model, Tolson & Shoemaker (2007) showed that DDS is able to converge rapidly to a good calibration solution (as opposed to a global optimal solution) and easily avoids poor local optima. As a result, DDS is ideally suited for computationally expensive optimization problems such as distributed watershed model calibration. DDS has been successfully applied for calibrating distributed watershed models (e.g., Haghnegahdar et al. 2014).

Tolson & Shoemaker (2007) compared the performance of DDS with the more commonly used shuffled complex evolution (SCE) algorithm for calibrating SWAT in the Cannonsville Reservoir watershed (New York, USA). The authors clearly showed DDS to be more efficient and effective than SCE in calibrating SWAT. Matott et al. (2011) investigated the performance of various MATLAB and Python optimizers applied to two simulation-based optimization case studies involving groundwater flow and contaminant transport. Their results provided empirical evidence that direct search algorithms and heuristic variants, such as DDS, are good choices for application to simulation-based optimization problems involving groundwater management. Wallner et al. (2012) investigated the impact of several optimization algorithms on the performance and robustness of the Hydrologic Modeling System from the Hydrological Engineering Center (HEC-HMS) model applied on the Aller-Leine River Basin (Germany). These authors found that DDS performed better than the other tested optimization methods in terms of the objective function values obtained after 1,000 model iterations. Yen et al. (2015) developed and demonstrated use of a computational procedure for evaluating efficiency of optimization methods in: (1) finding the optimal (best) parameter solutions within a reasonable number of model evaluations, (2) identifying the parameters that adequately allow reproduction of several hydrologic and water-quality responses (e.g., flow, sediment, and nutrients) simultaneously at

multiple locations in the watershed, and (3) delineating the regions associated with good calibration solutions within the whole parameter space. The authors applied their proposed procedure to evaluate the performance of six optimization methods including DDS for calibrating SWAT in the Eagle Creek watershed (Indiana, USA). Their results showed that DDS surpassed all other methods in terms of convergence speed and behavioral statistics. Chu et al. (2015) recently proposed a modified version of DDS, Heuristically Dynamically Dimensioned Search with Sensitivity Information (HDDS-S), which assigns a non-uniform probability to each model parameter to be modified by DDS on the basis of a sensitivity analysis. The above studies clearly reveal the advantages of using DDS in calibrating complex physically based hydrological models. However, despite the increasing interest in using DDS to calibrate complex models, the influence of the procedure used for initializing the model parameter values on the performance of DDS has not been investigated and which is the subject of this study.

Many of the above-cited studies used SWAT for various hydrological applications. SWAT is a spatially semi-distributed river basin model developed by the Agricultural Research Service of the United States Department of Agriculture. It was developed to evaluate the effects of alternative management decisions on water resources and pollution in mesoscale and large river basins. The model simulates long-term runoff, and export of sediments, nutrients, and contaminants from rural watersheds, especially those dominated by agriculture (Arnold et al. 1998). Despite its high popularity, SWAT is difficult to implement into the watersheds where forcing and/or geographical soil data are scarce because the model requires significant knowledge of watershed characteristics.

The Visualizing Ecosystems for Land Management Assessments (VELMA) model, developed by the US Environment Protection Agency (EPA), is a relatively simple ecohydrological model that links hydrological and biogeochemical processes within watersheds to simulate water cycling, carbon and nitrogen cycling in plants and soil, and the transport of dissolved forms of carbon and nitrogen from the terrestrial landscape to streams (Abdelnour et al. 2011). VELMA has been applied to simulate hydrological, carbon and nitrogen dynamics in a forested

watershed of the Pacific Northwest (Abdelnour *et al.* 2011, 2013). VELMA was extended to include biogeochemical processes of mercury coupled to carbon cycle and was used to simulate stream mercury and methylmercury concentrations in a forested watershed of South Carolina (Golden *et al.* 2012; Knightes *et al.* 2014). The VELMA model was recently adapted at Environment and Climate Change Canada to simulate and study mercury cycling in diverse Canadian forested watersheds under the Canadian Mercury Science Program framework. In all the above-cited studies, site-specific observations in combination with literature values were used as initial parameter values of the model; the parameter values were then manually adjusted through visual analysis to capture trends and patterns in observed data. While manual calibration was made easier in the above studies because of the availability of large field data, it would be more difficult to calibrate the model in watersheds where availability of field data is scarce. In order to apply VELMA to a large number of watersheds across Canada, an automatic calibration of the model is required, particularly where observed field data are limited.

The objective of this study is to propose a parameter optimization framework to efficiently and accurately calibrate the hydrological component of VELMA using DDS for two wetland-dominated watersheds which are typical of the forested Precambrian Shield in south-central Ontario, Canada. The model calibration is performed to adjust the hydrological parameters in order to reproduce the observed daily stream flow for the recent two decades at the outlet of these watersheds, and to capture the observed subsurface hydrological dynamics. A sensitivity analysis of the DDS algorithm is performed with respect to the choice of initial values of the calibrated parameters and with respect to the choice of maximum number of model evaluations. Finally, recommendations are provided to efficiently calibrate the hydrological component of VELMA using DDS algorithm. The investigated watersheds are representative of large portions of the Precambrian Shields; therefore, the results found could potentially be extended across much of south-central Ontario. More generally, this study provides modelers with valuable information for calibrating VELMA with the DDS algorithm for small forested watersheds of northern latitudes.

MATERIALS

The hydrological model

The eco-hydrological model VELMA, developed at the United States Environmental Protection Agency (EPA), is a spatially distributed differential mass balance model that simulates the lateral and vertical movement of water, heat, and nutrients within the soil, between the soil and the vegetation, and from the soil surface and vegetation to the atmosphere (Abdelnour *et al.* 2011, 2013). The modeling domain of VELMA is a three-dimensional matrix covering the topographical surface (x-y) and four soil layers (z). It consists of three coupled components: (1) a hydrological model that simulates soil water infiltration, evapotranspiration, drainage, surface and subsurface runoff, (2) a soil temperature model that simulates daily temperatures of soil layers, and (3) a plant-soil model that simulates ecosystem carbon and nitrogen storage and the cycling of carbon and nitrogen between a plant biomass layer and the active soil pools. For the purposes of this paper only the hydrological component is used. A complete description of the hydrological component of VELMA as well as the associated equations can be found in Abdelnour *et al.* (2011). Required input data include air temperature, precipitation, soil texture, and digital elevation model (DEM).

The original 'Processing' version of VELMA is not suitable for coupling the model with an automatic calibration technique such as the DDS algorithm because the simulation computational time for this version is too high to allow for a large number of model simulations required by the calibration algorithm. VELMA was thus re-written in Fortran in order to facilitate an efficient calibration of VELMA with DDS. Test simulations from the Fortran version of VELMA against the original 'Processing' version of VELMA confirmed that the output variables computed from both versions of the model were consistent.

The model calibration algorithm

The DDS algorithm is a simple stochastic single-solution-based heuristic global search algorithm that was developed for the purpose of finding the best attainable parameter combination (the lowest value of an objective function) within a

specified maximum number of model evaluations (Behrangi et al. 2008). The algorithm initially searches globally and becomes increasingly local as the number of model iterations (or model evaluations) approaches the maximum allowable number of iterations. DDS is, therefore, not designed to converge to the precise global optimum but to converge to the region of the global optimum in the best case or the region of a good local optimum in the worst case.

The DDS inputs are the maximum number of model evaluations allowed, m ; the vectors of lower, \mathbf{X}^{min} , and upper, \mathbf{X}^{max} , bounds for all D calibration parameters, and the initial parameter values $\mathbf{X}^0 = [X_1, \dots, X_D]$. The maximum number of model evaluations (m) is set according to the desired computational time required to be spent on the optimization problem. The value of m therefore depends on the time to run the model and the available computational resources. The initial parameter values can be either specified by the user or randomly sampled within the lower and upper bounds of the parameters. Note that DDS starts with initializing a unique set of parameter values while the other global optimization methods generally start with initializing a population of parameter values (population-based algorithms); therefore, choice of the initial values may impact optimization results.

For the first iteration, the objective function F (see the section 'Performance evaluation') is evaluated for the initial parameter values solution, $F(\mathbf{X}^0)$; this value becomes the best objective function value ($F_{best} = F(\mathbf{X}^0)$) corresponding to the best solution $\mathbf{X}^{best} = \mathbf{X}^0$. For the successive iterations, a set of new parameter values is selected by randomly perturbing the values from a normal distribution. The objective function F is then evaluated for the new parameter set (or candidate solution) \mathbf{X}^{new} . The objective function values $F(\mathbf{X}^{new})$ and F_{best} are then compared and the current best solution \mathbf{X}^{best} is then updated to the candidate solution \mathbf{X}^{new} only if $F(\mathbf{X}^{new})$ is better than F_{best} . This process is repeated until the number of iterations reaches the user-defined maximum number of model evaluations. The number of parameters being perturbed decreases with the increasing number of iterations, so that the solver searches more globally at the beginning and more locally at the end of the optimization procedure (Tolson & Shoemaker 2007). DDS has no other criteria to terminate the algorithm.

The only control parameter of the DDS algorithm is the scalar neighborhood size perturbation parameter, r , which defines the random perturbation size standard deviation as a fraction of the model parameter range (Tolson & Shoemaker 2007). The default value of 0.2 recommended by Tolson & Shoemaker (2007) is used in this study since this value allows the algorithm to escape regions around poor local minima. Theoretically, this value allows a sampling range that practically spans the normalized parameter range for a current parameter value halfway between the lower and the upper bounds.

STUDY SITES

The Harp (45°23'N, 79°08'W) and Dickie (45°09'N, 79°05'W) basins are located in the District of Mukoska County, Ontario, and are typical of the shallow till-rock ridge physiographic region of Central Ontario. They lie on a southern extension of the Precambrian Shield near the southern limit of the Boreal ecozone, which has a humid continental climate with long cool summers (Köppen class Dfb) (Buttle & Eimers 2009). The watersheds are part of the Great Lakes St Lawrence forest region and are predominantly covered by mixed deciduous coniferous forests (Dillon et al. 1991). The area underwent selective logging in the 1800s and early 1900s, primarily for eastern white pine and red pine, but the forests have regrown after farming was abandoned in the early 1900s (Dillon et al. 1991, 2003). Very little forest harvesting has occurred during the past several decades, leaving a predominance of secondary growth forests with stand age of more than 100 years (Dillon et al. 1991). The economy of the region is controlled largely by tourism activities, and extensive shoreline development associated with a number of cottage dwellings surround the Harp and Dickie lakes (Dillon et al. 2003).

Meteorology, stream flow, and water chemistry of the watersheds have been monitored by the Ontario Ministry of Environment (OMOE), Dorset Environmental Science Centre's (DESC) long-term monitoring program, which has operated since the late 1970s. Mean annual surface air temperature ranged from 3.5°C (1992) to 6.5°C (1998) with an average of 5°C for the period 1984–2008. Average winter and summer temperatures were –8.3°C and 17.2°C,

respectively. Annual precipitation ranged from 754 mm (1998) to 1,273 mm (2008) with an average of 980 mm. Precipitation distribution is relatively homogeneous throughout the year (Devito *et al.* 1996). About 30% of precipitation falls as snow (Devito *et al.* 1996).

Soils in the DESC area are poorly developed and contain very low proportions of silt and clay-sized particles (Jeffries & Snyder 1983). Infiltration is rapid through the sandy tills of the watersheds, which produces a saturated layer at the till–bedrock contact in areas of thin till, and runoff occurs as subsurface lateral flow through this layer (Peters *et al.* 1995). The upland–wetland hydrologic connection appears to change from a transient to a continuous link when the watershed is covered predominantly ($\geq 50\%$) with surficial deposits classified as minor till plain (Devito *et al.* 1996). Due to the shallowness of the surficial deposits associated with the effective impermeability of the bedrock, only local groundwater aquifers develop and the hydrology varies considerably over a year (Devito *et al.* 1996). Detailed descriptions of the physiography, geology, and geochemistry of the studied watersheds can be found in Jeffries & Snyder (1983).

For this study, HP5 and DE10 watersheds (Figure 1) in Harp and Dickie basins, respectively, were selected because they are the largest watersheds of the gauged streams in these basins and by themselves they supply a large amount of the runoff water to the Harp and Dickie lakes (Buttle & Eimers 2009). HP5 and DE10 have a watershed area of 190.5 and 78.9 ha and a mean slope of 3 and 1%, respectively (Buttle & Eimers 2009). Elevations range from 316/344 m (HP5/DE10) at the stream gauging stations to 422/382 m (HP5/DE10) at the northeastern/northwestern (HP5/DE10) ridgeline (Figure 1). Harp watershed is underlain by gneiss and other Precambrian metasedimentary bedrock including amphibolite and schist while Dickie watershed is underlain by hornblende migmatite (LaZerte & Dillon 1984). Surficial geology of the HP5 ranges from minor till plain (38.1%), composed of sandy loam and sand deposits between 1 and 10 m thick, to thin deposits with thickness less than 1 m and exposed bedrocks (48.6%), and peatland (13.3%) (Buttle & Eimers 2009). The DE10 catchment has lower variable surficial geology, with predominantly thin till (82.9%) and peat (17.1%). Thin surficial deposits in both watersheds

are composed of sand, silt, and gravel of granitic composition (Dillon *et al.* 1991). Measured annual runoff for HP5 ranges from 318 mm (2000) to 862 mm (1985) with an average of 582 mm for the period 1984–2001; while measured annual runoff for DE10 ranges from 303 mm (1987) to 622 mm (1985) with an average of 483 mm for the period 1984–2007. Most of the runoff in these streams is delivered during spring snowmelt, which generally occurs from mid-March to early May (Buttle & Eimers 2009).

METHODOLOGY

Data

Daily air temperature and precipitation data are used to force the hydrological model; whereas, daily observed stream flow data are used to calibrate the hydrological model parameters. Regular meteorological monitoring at the study sites has been conducted since 1979 by DESC, OMOE (Yao *et al.* 2009). Stream flow measurements have also been continuously conducted since 1979 to 2002 and 2007 at the mouth of the HP5 and DE10 watersheds, respectively (Yao *et al.* 2009). For computational efficiency, the model calibration was performed on a shorter period than the available data period. The VELMA model was calibrated over the periods 1984–2002 and 1984–2007 for the watersheds HP5 and DE10, respectively. These periods cover a dry period that lasted from the mid-1980s until the beginning of the 1990s and was followed by a wetter period until the mid-1990s (Dillon & Molot 2005). Meteorological data were obtained from the HPP2 and HYP2 meteorological stations located within 1 km of the watersheds HP5 and DE10, respectively (Yao *et al.* 2009). Daily observed stream flow measurements were obtained from gauging weirs/flumes located at the mouth of the HP5 and DE10 watersheds (Yao *et al.* 2009).

A 20 m resolution DEM of the HP5 and DE10 watersheds (Natural Resources Canada 2013) was used to compute flow direction, flow contribution area, delineate geographical boundaries of the watersheds, and generate a channel network. Each 20 m \times 20 m soil column was divided

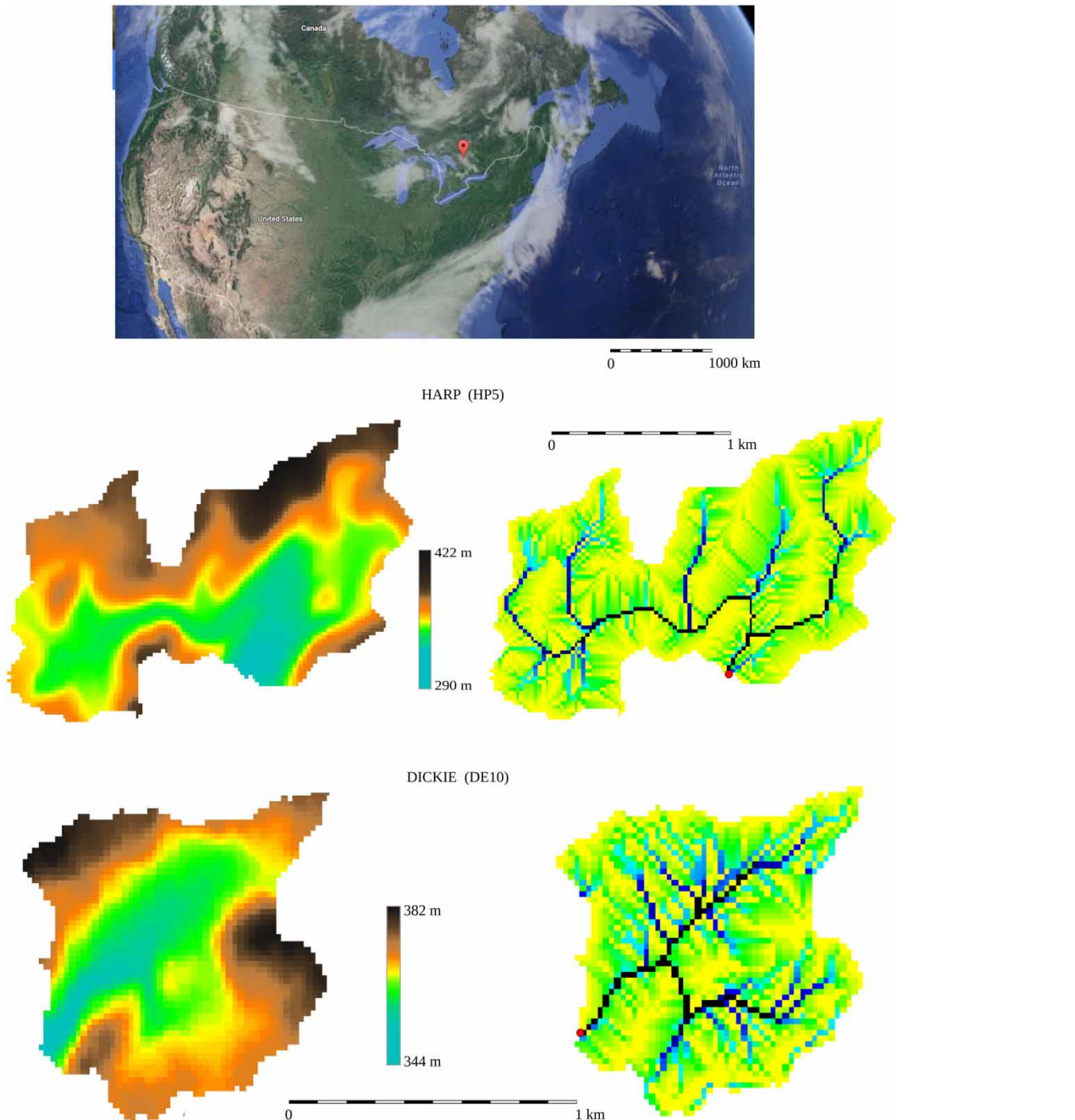


Figure 1 | Location and geomorphology of the studied watersheds. Upper panel: locations of the watersheds. Middle and lower panels: digital elevation model (left panels) and flow river network (right panels) of the HP5 and DE10 watersheds. Dots indicate the location of the gauging stations.

into four layers of equal thickness: a surface layer, two intermediate layers, and a deep layer. In this study, total depth to bedrock is a parameter to be calibrated and is considered

homogeneous in the entire watershed (see the section ‘Optimization strategies’). The dominant soil texture is loamy sand in the study watersheds (Lozano 1987).

Optimization strategies

The complexity of the optimization problem of any watershed model is dependent on the number of parameters that need to be adjusted and the wideness of the ranges of parameter values (Zhang *et al.* 2008). The hydrological component of VELMA has 13 parameters which require calibration for each investigated basin. These parameters are related to soil properties, evapotranspiration, and snow-related processes (Table 1). Preliminary sensitivity analyses revealed that surface hydraulic conductivity and both lateral and vertical decay of the hydraulic conductivity with depth have the largest influence on the model outputs. This is in agreement with the results found by Abdelnour *et al.* (2013).

Two sets of parameter optimization experiments were conducted using different parameter initialization methodologies to calibrate hydrological components of VELMA with the DDS algorithm. For the first set of experiments (named hereafter ‘high-cost strategy’; ‘cost’ refers to the computational resources requirements), 100 initial parameter values sets were created by randomly sampling

from the parameter ranges shown in Table 1. For the second set of experiments (named hereafter ‘low-cost strategy’), a unique set of initial parameter values was developed either from literature values or from previous calibrations of VELMA to different watersheds (Table 1, column ‘Observations’).

The parameter value ranges were determined either from literature values (Dingman 1994; Buttle & Eimers 2009) or from previous calibrations of VELMA to different watersheds (e.g., Abdelnour *et al.* 2011; Golden *et al.* 2012). For some parameters, the value ranges were extended to allow the model to be representative of a large number of watersheds (Table 1, ‘Fixed a priori’ in the ‘References’ column). The observed values for field capacity, porosity, wilting point, and surface saturation hydraulic conductivity were obtained following Dingman (1994). Field observations indicate that the average soil depth to bedrock in the HP5 and DE10 watersheds is about 1 m (LaZerte & Dillon 1984; Devito *et al.* 1989; Schiff *et al.* 2002; Dillon *et al.* 2003; Dillon & Molot 2005); therefore, soil depth to bedrock was set to 1 m for the second experiment. In the absence of field data, the remaining parameter values for the second

Table 1 | Parameters and parameter value ranges used in the automatic calibration of the hydrological model of VELMA

Name	Definition	Range	References	Observations
<i>Soil properties</i>				
θ_i^c	Field capacity in layer i	0.091–0.396	Dingman (1994)	0.125 ^a
ϕ_i	Porosity in layer i	0.398–0.501	Dingman (1994)	0.437 ^a
θ_i^w	Wilting point in layer i	0.033–0.272	Dingman (1994)	0.055 ^a
$sdtb$	Soil depth to bedrock (m)	0.1–10.0	Fixed a priori	1.0 ^a
K_s	Surface saturated hydraulic conductivity (mm day ⁻¹)	14–5,040	Dingman (1994)	1,466 ^a
f_v	Vertical decay rate of K_s (l m ⁻¹)	0.0–10.0	Fixed a priori	5.0
f_l	Lateral decay of K_s (l m ⁻¹)	0.0–10.0	Fixed a priori	5.0
<i>Snow</i>				
T_s	Threshold temperature (°C)	–5.0 –5.0	Fixed a priori	0.0
T_m	Melting temperature (°C)	1.0–10.0 (+ T_s)	Fixed a priori	5.0
dd	Degree-day factor for melt (mm °C ⁻¹ day ⁻¹)	1.0–10.0	Fixed a priori	5.0
ros	Rain on snow parameter	0.1–1.0	Fixed a priori	0.5
<i>Evotranspiration</i>				
K_{PET}	PET calibration parameter	0.1–1.0	Fixed a priori	0.5
c_{ET}	ET shape factor	1.0–10.0	Fixed a priori	5.0

^aRefers to the available information on the watershed characteristics found in LaZerte & Dillon (1984), Lozano (1987), Devito *et al.* (1989), Schiff *et al.* (2002), Dillon *et al.* (2003) and Dillon & Molot (2005).

experiment were fixed to the halfway values between the lower and upper bounds of these parameters.

The automatic model parameter calibration procedure using DDS algorithm was repeated six times with each initial parameter value set (i.e., 100 high-cost strategy randomly selected initial value sets and one observationally defined low-cost strategy initial value set) using 100, 500, 1,000, 2,500, 5,000, and 10,000 as the maximum number of model objective function evaluations (m). In addition, VELMA-DDS calibration experiments using the low-cost strategy were repeated ten times (named hereafter 'trials') for each value of m using identical setup. Thus, a total of 660 model parameter calibration experiments were performed using the VELMA-DDS framework (i.e., 600 experiments for the high-cost strategy and 60 experiments for the low-cost strategy). The high- and low-cost strategies investigate the influence of the choice of initial parameter values on the optimization results; whereas calibration experiments using different values of m evaluate the efficiency of the algorithm and its convergence characteristics. The performance differences between the ten trial experiments allow an investigation of the sensitivity of the DDS optimization technique to the random perturbation method utilized by DDS to generate candidate parameter value solutions during the calibration procedure.

Optimization Software Toolkit for Research Involving Computational Heuristics (OSTRICH; Matott 2005) framework was used to set up the VELMA-DDS calibration procedure. OSTRICH is a model-independent calibration and optimization tool consisting of a number of popular optimization algorithms including the DDS.

Performance evaluation

In this study, the optimization objective function describes how well the daily stream flow values are simulated by the model. The Nash–Sutcliffe (NS) coefficient (Nash & Sutcliffe 1970), a normalized form of the sum of squares of residuals of square root of daily discharges, was used as an objective function to calibrate the model.

$$NS(\sqrt{Q}) = 1 - \frac{\sum (\sqrt{Q_{obs,i}} - \sqrt{Q_{mod,i}})^2}{\sum (\sqrt{Q_{obs,i}} - \sqrt{Q_{obs,i}})^2} \quad (1)$$

where $\sqrt{Q_{obs,i}}$ and $\sqrt{Q_{mod,i}}$ are the square roots of observed and modeled daily (i) discharges, respectively; $\sqrt{Q_{obs,i}}$ is the average $\sqrt{Q_{obs,i}}$ over the calibration period, and $i = 1, 2, \dots, N$, where N is the total number of observed and modeled data pairs. NS determines how well the plot of the observed values versus the modeled values fits the 1:1 line, and it ranges from $-\infty$ to 1, where $NS = 1$ indicates a perfect fit between the simulated and observed hydrograph. NS quantifies the ability of the model to explain variance of the variable of interest, i.e., the improvement achieved by the model in simulating the variable of interest compared to a basic reference model simulating a constant value equal to the average observed value of the variable over the time period of calibration (Lespinas et al. 2014). NS calculated on square root values of the variable was chosen instead of the classical NS calculated on raw values because the latter tends to emphasize large errors associated with flood events.

The water content of soil layers at the beginning of the calibration period can affect the model results and, consequently, the computation of the objective function. By default, VELMA sets the initial water content of all the cells in all four layers of the watershed to the field capacity value specified for each soil type. Special care was therefore taken to minimize the uncertainties due to the initialization of water content by discarding the first year of the simulation results in the computation of the objective function. Setting up an appropriate spin-up period allows initial soil moisture levels to adjust to the site's biophysical conditions, and is commonly used in hydrological modeling studies (e.g., Lespinas et al. 2014).

The DDS algorithm performance was assessed by computing an average objective function value across 100 high-cost strategy experiments at the end of each additional objective function evaluation step for a given value of m . In the case of the low-cost strategy experiments, an average objective function value was calculated across ten trial experiments after each additional objective function evaluation step for a given value of m . Average objective function values were then plotted against the number of objective function evaluation steps completed for each value of m ; thus, comparative results were obtained in calibrating VELMA hydrology using $m = 100$ to 10,000. Such plots, named convergence plots, are a popular method in the studies comparing performance of optimization

algorithms (e.g., Tolson & Shoemaker 2007; Zhang et al. 2008; Huang et al. 2014). Averaging of the objective function values over multiple optimization experiments allows minimization of the statistical variability and the influence of the selection of initial parameter values (Behrangi et al. 2008). Given that the average algorithm performance does not provide a complete picture of results, the distributions of the best (i.e., final) objective function values of 100 high-cost initial value experiments and ten low-cost trial experiments for each value of m were also graphically assessed using box-and-whisker diagrams.

RESULTS

Figure 2 shows, for each value of the maximum number of model evaluations m , the evolution of average best NS

values as a function of the number of model evaluations completed by the DDS algorithm for both high-cost and low-cost strategies. For both strategies, the objective function values change relatively quickly over the initial 300 model evaluations, and then change relatively slowly regardless of the maximum number of model evaluations. Also, the differences in the average NS values between the various maximum number of model evaluation (m) calibration experiments are generally larger for the initial 300 model evaluations than for the remaining model evaluations. The behavior of the DDS performance, however, varies according to the watershed and the strategy. For HP5, in the case of the high-cost strategy, the efficiency of DDS to improve the objective function values over the initial 300 model evaluations is inversely related to the maximum number of model evaluations (Figure 2(a)); whereas for DE10, the DDS performance is similar enough whatever the maximum

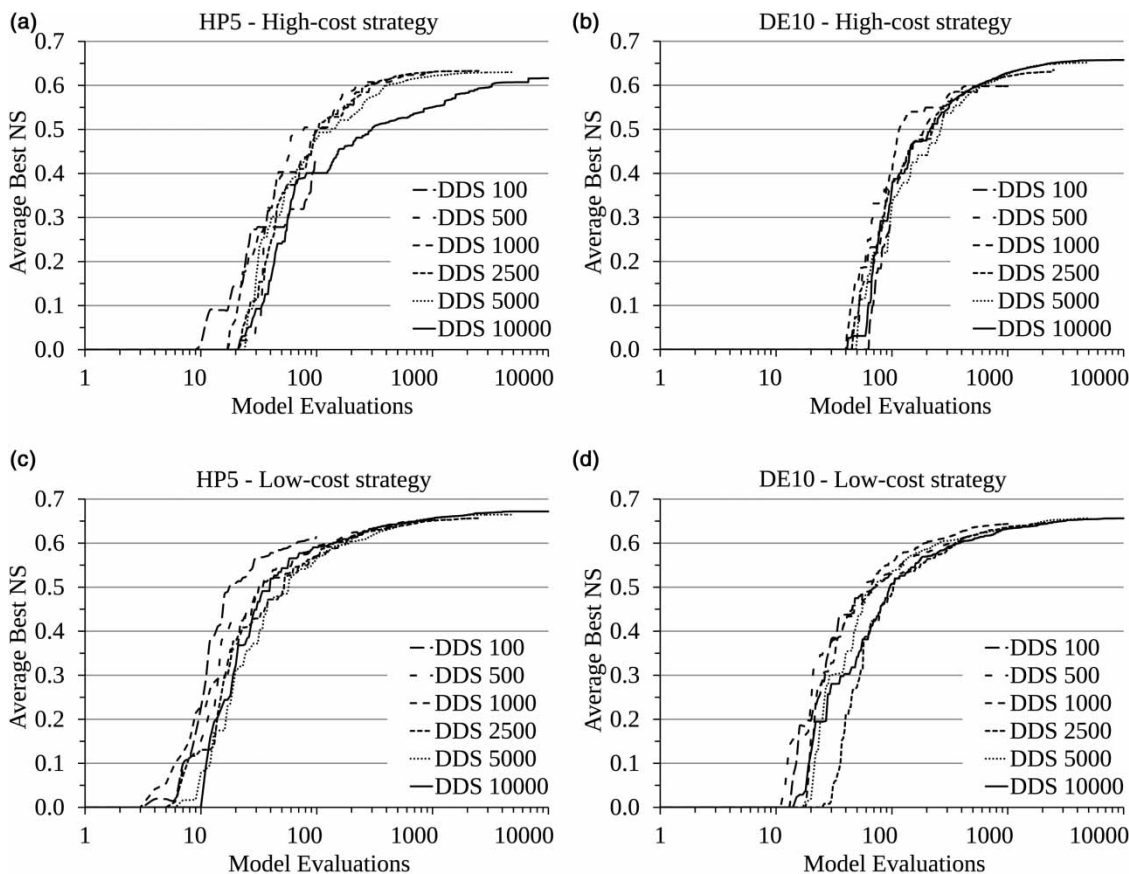


Figure 2 | Evolution of average best NS value across 100 high-cost strategy experiments (a) and (b) and across ten low-cost strategy experiments (c) and (d) as a function of the number of model evaluations for each value of m .

number of model evaluations chosen (Figure 2(b)). The DDS algorithm associated with the low-cost strategy converges more efficiently towards a better solution than DDS associated with the high-cost strategy. For example, the average best NS value obtained for HP5 after 100 model evaluations is 0.447(0.614), 0.505(0.594), 0.503(0.570), 0.497(0.572), 0.479(0.566), and 0.401(0.590) for the high (low)-cost

strategy for DDS with $m = 100, 500, 1,000, 2,500, 5,000,$ and $10,000,$ respectively. Also, the final average NS values obtained from the low-cost strategy are higher than those obtained from the high-cost strategy (Table 2).

Figure 3 shows box-and-whisker diagrams of the final objective function values for the 100 high-cost and the ten low-cost strategy experiments for each value of m . Generally, the quartile values of the final objective functions strongly increase between DDS with $m = 100$ and 500 , and increase relatively slowly thereafter. The distribution of the final objective function values, however, varies according to the watershed, the strategy, and the maximum number of model evaluations. For example, the interquartile range of DDS solutions for the high-cost strategy is higher for the HP5 watershed than for the DE10 watershed when the maximum number of model evaluations is more than 500 (Figure 3(a) and 3(b)). The quartile values obtained from the low-cost strategy are mostly higher than those obtained from the high-cost strategy irrespective of the

Table 2 | Final average best NS values

m	HP5		DE10	
	High-cost	Low-cost	High-cost	Low-cost
100	0.447	0.614	0.290	0.530
500	0.612	0.638	0.594	0.616
1,000	0.630	0.652	0.598	0.644
2,500	0.633	0.657	0.635	0.646
5,000	0.631	0.665	0.651	0.656
10,000	0.616	0.672	0.657	0.656

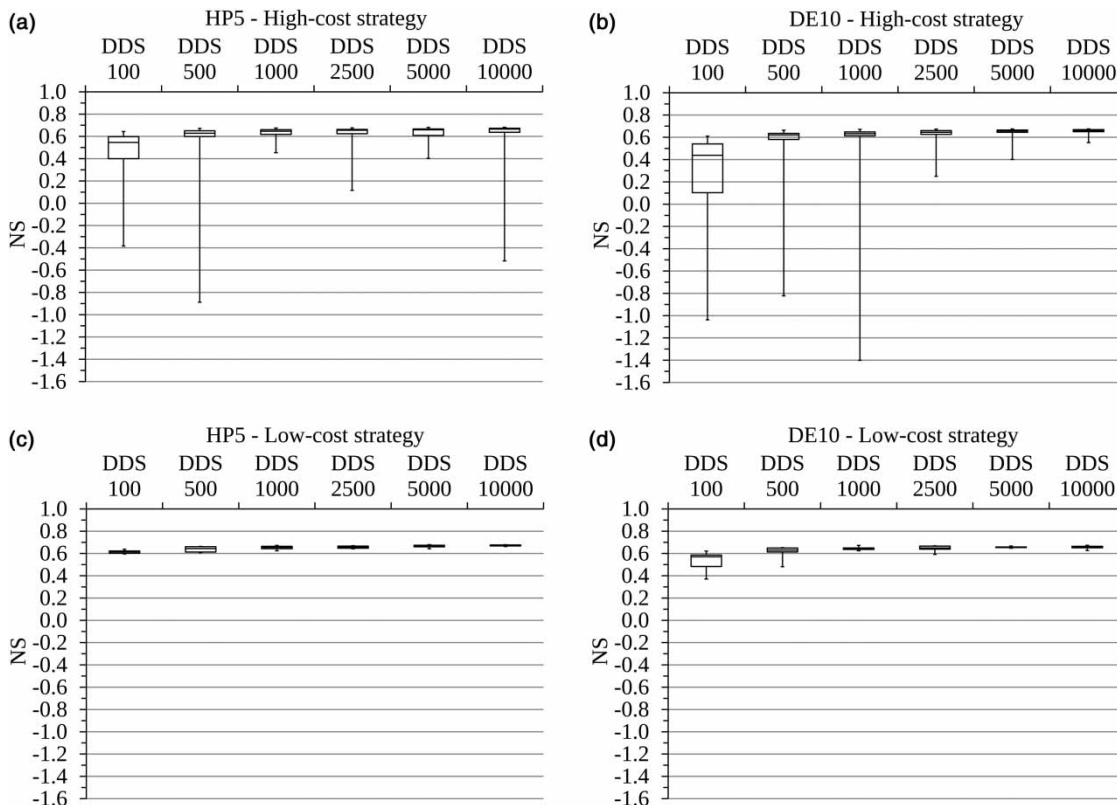


Figure 3 | Box-and-whisker diagram of the best NS values across 100 high-cost strategy (a) and (b) and across ten low-cost strategy experiments (c) and (d) for each value of m .

watershed considered. For example, the median final NS values obtained for HP5 are 0.547(0.614), 0.628(0.644), 0.646(0.655), 0.655(0.656), 0.659(0.666), and 0.665(0.673) for the high (low)-cost strategy for DDS with $m = 100, 500, 1,000, 2,500, 5,000,$ and $10,000,$ respectively. The minimum final NS values are all positive for the low-cost strategy while they can be negative for the high-cost strategy, indicating that in some cases DDS cannot avoid poor local optima even after 10,000 model evaluations. However, the maximum NS values obtained by the high-cost strategy experiments are mostly larger than those obtained from the low-cost strategy for all values of m . For each value of m and for each watershed, the differences of the maximum NS values between the high-cost strategy and the low-cost strategy are, however, equal to or less than 0.01. Also, for both watersheds and both strategies, the maximum NS values increased by less than 0.01 after $m = 1,000$ model evaluations.

Figures 4 and 5 show the distribution of the optimized parameter values obtained from both strategies for the HP5 and DE10 watersheds, respectively. The optimized parameter values are shown for the 30 best optimization results following a decreasing order of the ranks of the NS values. The minimum and maximum values of the Y axis are equal to the range of the parameter values defined in Table 1. Most of the optimized parameter values of the best simulations are very close to each other, thus defining a convergence region in the parameter space where most local optima can be found as well as potentially the global optimum of the objective function. The optimized parameter values are very close between both investigated watersheds, which is not surprising given that their geological characteristics and land cover are very similar. The optimized parameter values are consistent with the available field observations. For example, the optimized soil depth to bedrock for the 30 best simulations varies from 0.103(0.981) to 1.124(1.444) m with an average of 0.732(1.204) m for the HP5 (DE10) watershed. These values are consistent with the soil depth observations in the investigated watersheds (e.g., LaZerte & Dillon 1984; Devito et al. 1989; Dillon et al. 2003). Also the optimized surface saturated hydraulic conductivity values are within the ranges of the hydraulic conductivities measured in soils of the Harp watershed (Hinton et al. 1993; Devito et al. 1996). Only the optimized

values of the field capacity, porosity, and wilting point strongly vary between the best simulations. Although the results obtained from the low-cost strategy are very similar to those obtained from the high-cost strategy, the dispersion of the optimized parameter values is somewhat higher than for the high-cost strategy.

Figure 6 illustrates the simulated and observed hydrographs over the period, January 1994–December 1995, obtained from the best model calibration simulations (i.e., NS = 0.681 and 0.675 for the HP5 and DE10 watersheds, respectively). Overall, the watershed model captures the daily runoff dynamics over the entire calibration period, although it tends to overestimate low flows and underestimate peak flows. The model reproduces the snow melting in spring, the low flows in summer, and the rapid runoff response due to storm events in fall with a good accuracy. Table 3 shows the domain average annual values of runoff, rainfall, and evaporation computed from the best simulations and averaged over all the soil columns located within the watershed delineations. The partition of runoff between the different soil layers indicates that the main flow pathway is the subsurface runoff in the two first soil layers in the investigated watersheds; thus, the main runoff pathway is located within 500 cm depth below the surface when the soil depth to bedrock is about 1 m. Moreover, both surface runoff and deep subsurface runoff are second-order contributors to the total water losses at the watershed-scale, compared to the evaporation which is responsible for about 60% of the water losses for both watersheds. Seasonally, the respective contributions of the surface runoff and the subsurface runoff to the total runoff are similar (not shown).

DISCUSSION

Overall, the results presented above indicate that DDS is able to calibrate efficiently the hydrological model of VELMA irrespective of the initial parameter values and the calibration strategy used. Quality of the model optimization increases very rapidly over the first 300 model evaluations, and increases slowly thereafter. This behavior is in agreement with other studies examining the efficiency of DDS in improving the calibration results of complex

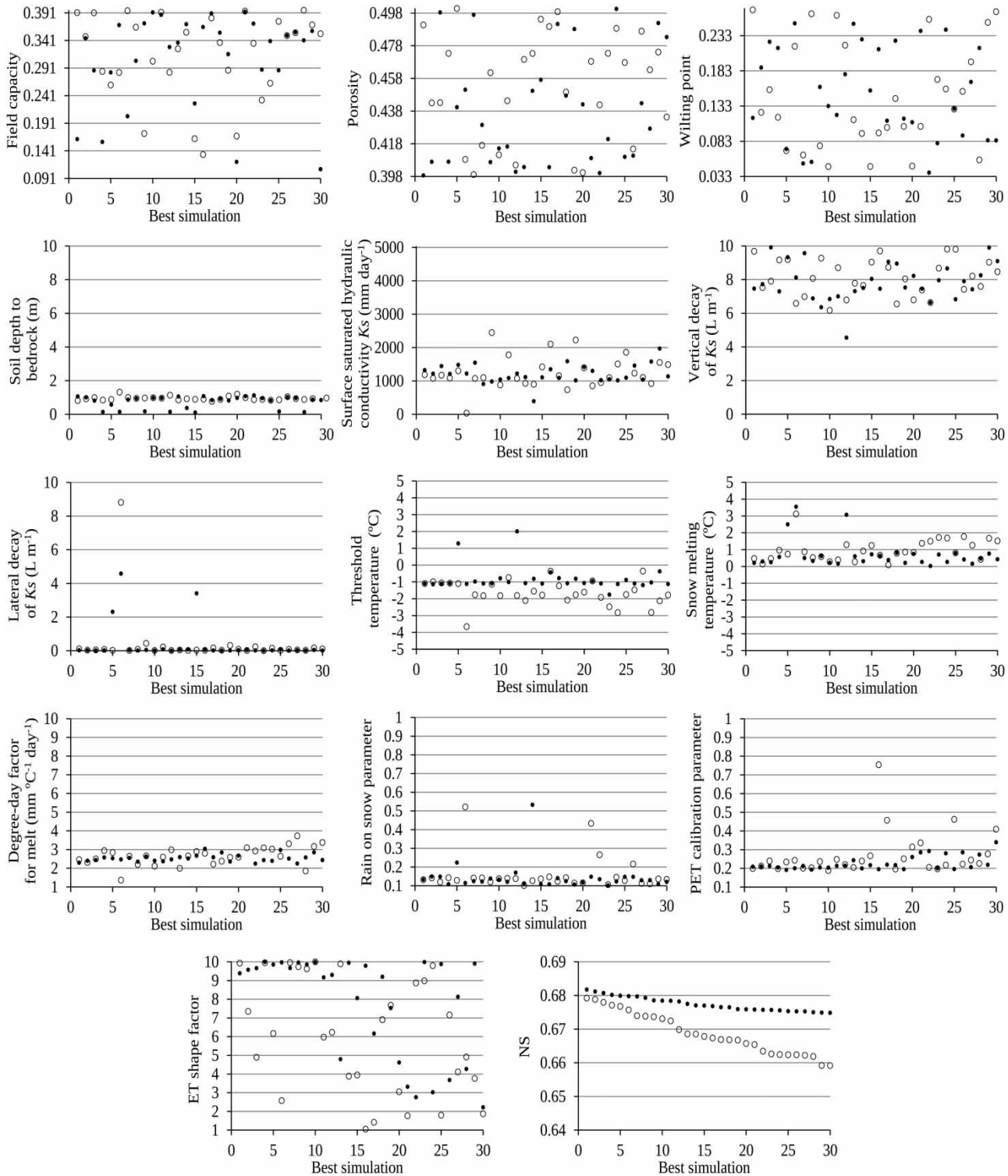


Figure 4 | Optimized parameter values obtained from the high-cost (black dots) and the low-cost (white dots) strategies for the 30 best optimization results for the HP5 watershed. The minimum and maximum values of Y axis are equal to the range of the parameter values defined in Table 1.

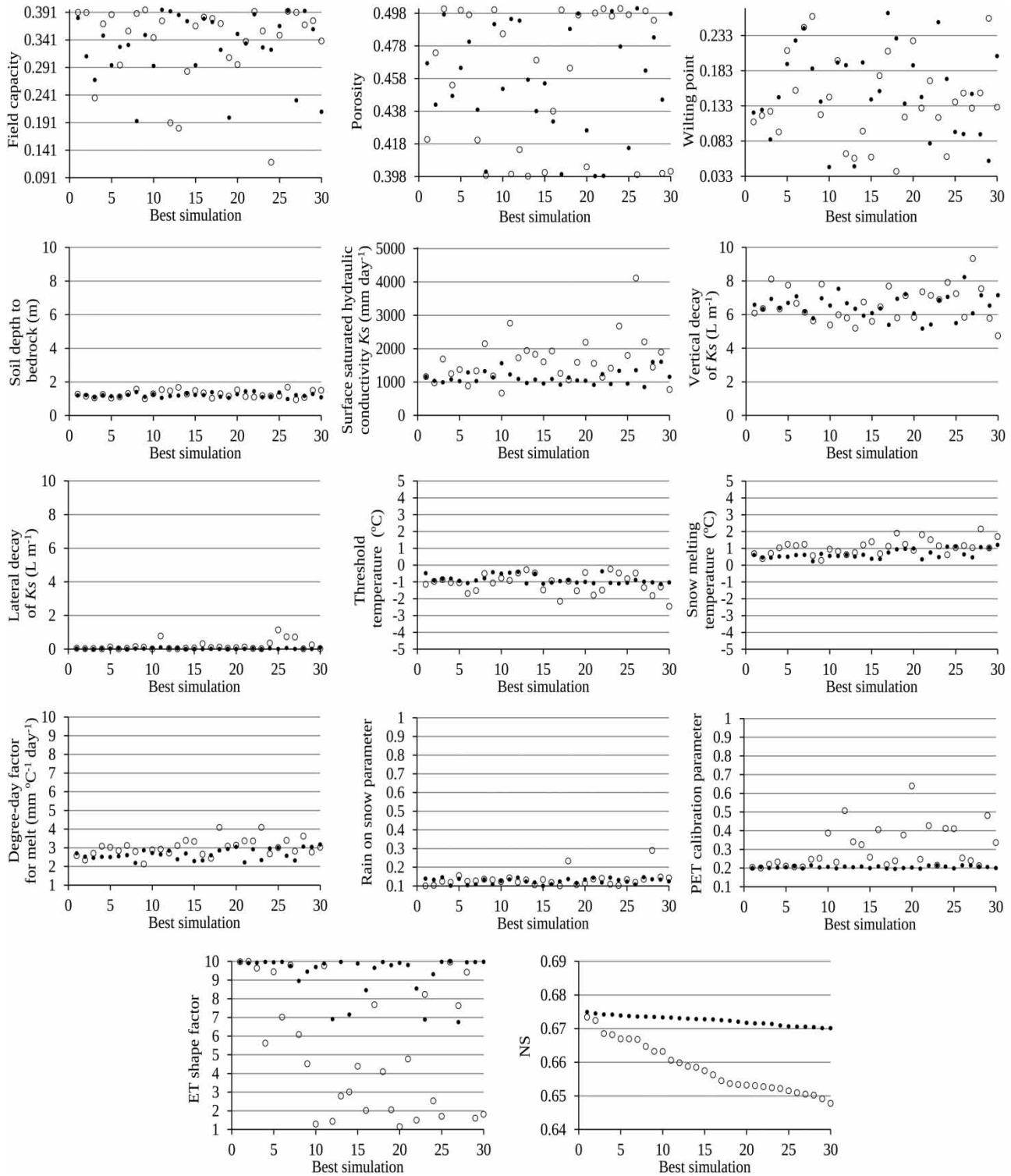


Figure 5 | Optimized parameter values obtained from the high-cost (black dots) and the low-cost (white dots) strategies for the 30 best optimization results for the DE10 watershed. The minimum and maximum values of Y axis are equal to the range of the parameter values defined in Table 1.

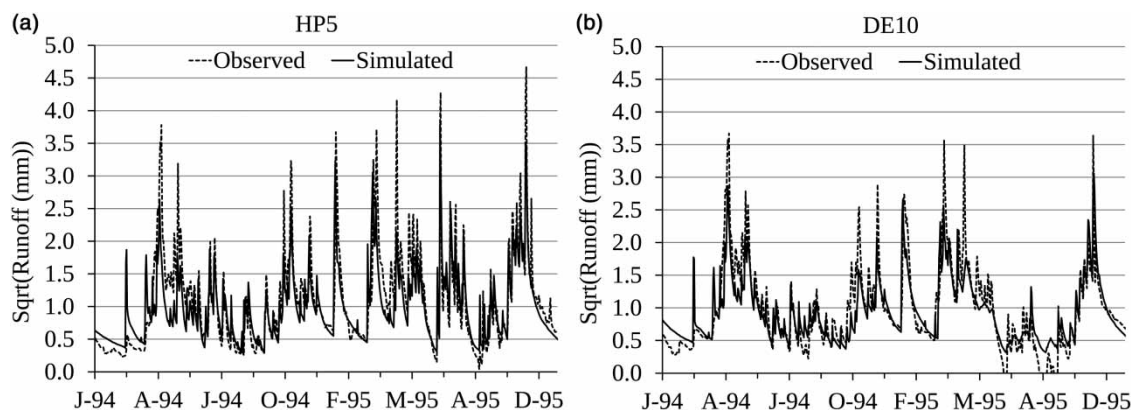


Figure 6 | Time series of the observed and simulated discharge for the watershed HP5 (a) and DE10 (b) over the period January 1994–December 1995. Simulated discharge refers to the best calibration results obtained from all the optimization experiments.

Table 3 | Average annual hydrological fluxes (in mm year⁻¹) simulated over the period 1985–2002

	HP5	DE10
Surface runoff	36	34
Subsurface runoff layer 1	148	105
Subsurface runoff layer 2	189	224
Subsurface runoff layer 3	22	20
Subsurface runoff layer 4	13	12
Rainfall	971	971
Total evaporation	566	598

watershed models (Tolson & Shoemaker 2007; Matott *et al.* 2011; Huang *et al.* 2014).

The DDS performance in calibrating the hydrological model of VELMA using the low-cost strategy is slightly better than that obtained using the high-cost strategy, both in terms of the algorithm convergence efficiency and accuracy. Results of this study also reveal that DDS initialized with some parameter values may fail to reach good calibration solutions even after 10,000 model evaluations (Figure 3). This indicates that initialization of the parameters is a crucial step in the calibration of a watershed model. Some studies using DDS to calibrate watershed models take care of the initialization of the parameter values before implementing DDS (e.g., Haghnegahdar *et al.* 2014; Huang *et al.* 2014), while others fix them randomly (e.g., Tolson & Shoemaker 2007). This is mainly because the field information is not always available directly or the high-degree of conceptuality of watershed models does not

allow to easily initialize the parameters from the watershed physical characteristics. As an alternative to the field information, the initial parameter values might also be fixed according to the past calibrations of the model for watersheds with similar characteristics, as performed by Haghnegahdar *et al.* (2014).

Beyond NS statistics, the convergence of most of the optimized parameters into a region well delimited in the parameter space is evidence of the robustness of the DDS algorithm to calibrate the hydrological model of VELMA. Most of the optimized parameter values are close to the available observations, which validate, at least partially, the representativeness of the physical processes in the model. The exceptions are the field capacity, porosity, and the wilting point, for which the large dispersion of the optimized parameters reveals equifinality issues (Duan *et al.* 1993). Note that the preliminary sensitivity analyses indicate that these parameters have a very low impact on model results in terms of NS values (not shown), which is in agreement with the results found by Abdelnour *et al.* (2013). Although the ranges of the parameter values were mostly fixed a priori (Table 1), they were chosen to be sufficiently large to allow for the parameter values to be adjusted by the calibration procedure with a sufficient degree of flexibility. The optimized values for the rain-on-snow parameter and the ET shape factor are very close to the minimum and maximum values, respectively (Figures 4 and 5). This indicates that the rain-on-snow effect is likely very limited in the investigated watersheds, and the minimum boundary value of this parameter might therefore be

fixed to 0 instead of 0.1. Conversely, the maximum boundary value of the ET shape factor might be increased in order to improve the optimization results.

Other optimization methods could have been used to calibrate the hydrological model of VELMA. Zhang *et al.* (2008) compared performance of multiple optimization algorithms to calibrate the popular watershed model SWAT for four watersheds with different climatic and hydrologic conditions. The authors found that the relative performance of the algorithms is dependent on the selected watershed, and that no one optimization algorithm consistently performs better than other algorithms for all watersheds. It should be noted that control parameters of optimization algorithms can be tuned to improve their performance for specific calibration problems (Behrangi *et al.* 2008). Based on these considerations, it seems more useful to choose an optimization algorithm and adjust its control parameters in order to make it more adapted for the calibration problem of interest than using a batch of optimization algorithms.

In this study, performance of the DDS algorithm is evaluated from the perspective of a distributed watershed model calibration with a limited number of model evaluations and, thus, obtaining the global optimum solution is not a reasonable expectation. It is well-known that most optimization algorithms have difficulty in locating the global optimum because the response surfaces often contain multiple local optima with regions of attraction of differing size, discontinuities, and long ridges and valleys (Thyer *et al.* 1999). Furthermore, even in the case where the global optimum is found, the values of calibrated model parameters may be degraded because of potential errors in observed data used in the model calibration process or the representation of processes in the model. Finally, the calibration problem formulated in this study could be further extended to include alternative low-flow weighting schemes for the hydrological model calibration, and alternative objective functions (including multi-objective functions) for model evaluations.

It is possible to improve the performance of the DDS algorithm in reaching good calibration solutions more efficiently. For example, Huang *et al.* (2014) proposed a modified version of the DDS algorithm – MDDS (M for modified) – that makes full use of sensitivity information in the optimization procedure. The authors show that

MDDS outperforms the DDS algorithm when the maximum number of model evaluations is chosen to be less than 2,500, and that the advantages of the MDDS algorithm are more obvious for high-dimensional distributed hydrological models. Tolson & Shoemaker (2007) and Yen *et al.* (2016) have demonstrated that the performance of DDS could vary with respect to the perturbation factor (r) in terms of convergence speed. Because the calibration problem proposed here has one main region of attraction, DDS might be fine-tuned by reducing r to 0.1 when the optimized parameter sets are close to the region of the global optimum in order to reach higher NS values more rapidly. The optimization results found in this study are also dependent on the modeling choices presented in the section ‘Optimization strategies’. For example, the model parameters that have been fixed might also be calibrated in order to increase the performance of VELMA hydrology. The soil depth to bedrock was calibrated then divided equally between the four soil layers of the model to reduce the dimensionality of the optimization problem and also because detailed information on the soil profile was not available for the investigated watersheds. This choice could be questioned since equal soil layer thickness along the soil profile favors deep flow and slow runoff response to rainfall (Abdelnour *et al.* 2013). Note, however, that the large decrease of the optimized vertical hydraulic conductivities in the soil profile (Figures 5 and 6) leads to rapid vertical flow and subsequent lateral flow in the top two subsurface layers (Table 3), which likely allows compensation for errors in the soil layer thicknesses. A calibration of the individual soil layer thicknesses, therefore, does not guarantee improvement of the model results. Investigation of these options is, however, beyond the scope of this study; the aim here is to provide a sensitivity analysis of the DDS algorithm performance with respect to the choice of the initial values of the model parameters to be calibrated and the choice of maximum number of model evaluations for arriving at the calibration solution.

CONCLUSION

This paper investigates the performance of the DDS algorithm in calibrating the spatially distributed hydrological

model of VELMA for reproducing the hydrological functioning of two main tributaries of the Harp and Dickie lakes (Ontario, Canada) over the recent two decades. Both investigated watersheds are typical of the shallow-till-rock ridge physiographic region of Central Ontario, and therefore present similar characteristics. The DDS performance was evaluated against the choices of the initial parameter values and the maximum number of model evaluations employed in the calibration procedure. Two model parameter value initialization strategies were implemented in the VELMA-DDS optimization framework. The first strategy consisted of randomly sampled initial values from the feasible parameter space (high-cost strategy), while the second strategy consisted of using a unique initial set of parameter values which were partly derived from field observations (low-cost strategy) and partly fixed to the mid-values of the parameter ranges. Furthermore, the DDS calibration algorithm was tested for six different values of the maximum number of model evaluations ranging from 100 to 10,000.

The results show that the DDS algorithm is able to converge efficiently to a good calibration solution of the hydrological model of VELMA irrespective of the initialization strategy chosen for the parameter values. The objective function of the VELMA-DDS framework which tests the performance of the model in simulating daily water discharges improves very rapidly over the first 300 model evaluations then increases relatively slowly over the remaining model evaluations. The optimization results are, however, slightly better for the low-cost strategy than for the high-cost strategy, both in terms of the algorithm convergence efficiency and accuracy. The probability that DDS fails to avoid poor local optima is larger for the high-cost strategy than for the low-cost strategy even after 10,000 model evaluations. The main exceptions are for the maximum NS values, where DDS performance obtained from the high-cost strategy is slightly larger than that obtained from the low-cost strategy. For both watersheds and both strategies, the maximum NS values increased by less than 0.01 after 1,000 model evaluations.

The distribution of the optimized parameter values for the best simulations indicates that most of the parameters converge to a well-defined region in the parameter space where many local optima, as well as likely the global

optimum, can be found. The optimized parameter values are consistent with field observations. Comparison of simulated hydrological fluxes with measurements indicates that the hydrological model of VELMA is able to capture the observed subsurface dynamics in HP5 and DE10. The results of this study are evidence of the robustness of the DDS algorithm to converge to a region close to the global optimum, and the results also validate the representativeness of hydrological processes in the model.

Based on the above results, it is recommended that field measurements or any other relevant information related to watershed characteristics be considered to initialize the hydrological parameter values of VELMA for calibrating the model with DDS. The maximum number of model evaluations should be chosen to be at least 1,000 in order to obtain a good calibration solution of the hydrological model of VELMA. Beyond 1,000 model evaluations, since DDS does not improve significantly the optimization results, it would be preferable to use a local optimization method initialized from the final DDS solution in order to identify a more precise estimate of the local optimum close to where DDS converged, as recommended by Tolson & Shoemaker (2007). In case the initial parameter values are randomly chosen, an ensemble of calibration trials initialized from different parameter value sets spanning the range of allowable parameter values should be performed in order to avoid poor local optima.

This study demonstrates that the use of available information on the watershed characteristics to initialize the parameter values can aid in improving the efficiency of the DDS algorithm to reach good calibration results in a limited number of model evaluations. This result is particularly important for the calibration of computationally intensive watershed models; however, further studies should be conducted for other watersheds as well as other hydrological models. More generally, this study reveals the need for modelers to have access to field information on the watershed characteristics for calibrating and validating complex physically based distributed models. In this context, use of detailed databases on land cover types and soil properties appear potentially helpful in calibrating such complex hydrological models as VELMA at regional-scale.

The next phase of this project is to use the developed VELMA-DDS modeling framework to calibrate the nutrient

and mercury components of VELMA with the aim to efficiently simulate nutrient and mercury fluxes in the two investigated watersheds which will be the subject of a forthcoming study. In future, the VELMA-DDS modeling framework is planned to be used in other boreal forested watersheds for which the available data are scarce in order to better understand transport and transformation of mercury from atmospheric deposition to aquatic ecosystems.

REFERENCES

- Abdelnour, A., Stieglitz, M., Pan, F. & McKane, R. 2011 [Catchment hydrological responses to forest harvest amount and spatial pattern](#). *Water Resources Research* **47** (9), W09521. doi:10.1029/2010WR010165.
- Abdelnour, A., McKane, R., Stieglitz, M., Pan, F. & Cheng, Y. 2013 [Effects of harvest on carbon and nitrogen dynamics in a pacific Northwest forest catchment](#). *Water Resources Research* **49** (3), 1292–1313. doi:10.1029/2012WR012994.
- Arnold, J. G., Srinivasan, R., Muttiah, R. R. & Williams, J. R. 1998 [Large area hydrologic modeling and assessment part I: model development](#). *Journal of the American Water Resources Association* **39**, 73–89.
- Behrangi, A., Khakbaz, B., Vrugt, J. A., Duan, Q. & Sorooshian, S. 2008 [Comment on ‘Dynamically dimensioned search algorithm for computationally efficient watershed model calibration’ by Brian A. Tolson and Christine A. Shoemaker](#). *Water Resources Research* **44** (12), W12603. doi:10.1029/2007WR006429.
- Buttle, J. M. & Eimers, M. C. 2009 [Scaling and physiographic controls on streamflow behaviour on the precambrian shield, south-central Ontario](#). *Journal of Hydrology* **374** (3–4), 360–372. doi: 10.1016/j.jhydrol.2009.06.036.
- Chu, J., Peng, Y. & Ding, W. D. 2015 [A heuristically dynamically dimensioned search with sensitivity information \(HDDS-S\) and application to river basin management](#). *Water* **7** (5), 2214–2238.
- Devito, K. J., Dillon, P. J. & LaZerte, B. D. 1989 [Phosphorus and nitrogen retention in five precambrian shield wetlands](#). *Biogeochemistry* **8** (3), 185–204. doi:10.1007/BF00002888.
- Devito, K. J., Hill, A. R. & Roulet, N. 1996 [Groundwater-surface water interactions in headwater forested wetlands of the Canadian shield](#). *Journal of Hydrology* **181** (1–4), 127–147. doi:10.1016/0022-1694(95)02912-5.
- Dillon, P. J. & Molot, L. A. 2005 [Long-term trends in catchment export and lake retention of dissolved organic carbon, dissolved organic nitrogen, total iron, and total phosphorus: the Dorset, Ontario, study, 1978–1998](#). *Journal of Geophysical Research* **110**, G01002. doi:10.1029/2004JG000003.
- Dillon, P. J., Molot, L. A. & Scheider, W. A. 1991 [Phosphorus and nitrogen export from forested stream catchments in central Ontario](#). *Journal of Environmental Quality* **20** (4), 857–864. doi:10.2134/jeq1991.00472425002000040025x.
- Dillon, P. J., Somers, K. M., Findeis, J. & Eimers, M. C. 2003 [Coherent response of lakes in Ontario, Canada to reductions in sulphur deposition: the effects of climate on sulphate concentrations](#). *Hydrology and Earth System Sciences* **7** (4), 583–595.
- Dingman, S. 1994 *Physical Hydrology*. Prentice Hall, Upper Saddle River, NJ.
- Duan, Q., Gupta, V. K. & Sorooshian, S. 1993 [Shuffled complex evolution approach for effective and efficient global minimization](#). *Journal of Optimization Theory and Applications* **76** (3), 501–521.
- Golden, H. E., Knightes, C. D., Conrads, P. A., Davis, G. M., Feaster, T. D., Journey, C. A., Benedict, S. T., Brigham, M. E. & Bradley, P. M. 2012 [Characterizing mercury concentrations and fluxes in a coastal plain watershed: insights from dynamic modeling and data](#). *Journal of Geophysical Research* **117**, G01006. doi:10.1029/2011JG001806.
- Haghnegahdar, A., Tolson, B. A., Davison, B., Seglenieks, F. R., Klyszejko, E., Soulis, E. D., Fortin, V. & Matott, L. S. 2014 [Calibrating environment Canada’s MESH modelling system over the Great Lakes basin](#). *Atmosphere-Ocean* **52** (4), 281–293. doi:10.1080/07055900.2014.939131.
- Hinton, M. J., Schiff, S. L. & English, M. C. 1993 [Physical properties governing groundwater flow in a glacial till catchment](#). *Journal of Hydrology* **142** (1–4), 229–249. doi:10.1016/0022-1694(93)90012-X.
- Huang, X., Liao, W., Lei, X., Jia, Y., Wang, Y., Wang, X., Jiang, Y. & Wang, H. 2014 [Parameter optimization of distributed hydrological model with a modified dynamically dimensioned search algorithm](#). *Environmental Modelling & Software* **52**, 98–110. doi:10.1016/j.envsoft.2013.09.028.
- Jeffries, D. S. & Snyder, W. R. 1983 *Geology and Geochemistry of the Muskoka-Haliburton Study Area. Data Report DR 83/3*. Ontario Ministry of the Environment, Toronto, Ontario, Canada.
- Knightes, C. D., Golden, H. E., Journey, C. A., Davis, G. M., Conrads, P. A., Marvin-DiPasquale, M., Brigham, M. E. & Bradley, P. M. 2014 [Mercury and methylmercury stream concentrations in a coastal plain watershed: a multi-scale simulation analysis](#). *Environmental Pollution* **187**, 182–192. doi:10.1016/j.envpol.2013.12.026.
- LaZerte, B. D. & Dillon, P. J. 1984 [Relative importance of anthropogenic versus natural sources of acidity in lakes and streams of central Ontario](#). *Canadian Journal of Fisheries and Aquatic Sciences* **41** (11), 1664–1677. doi:10.1139/f84-205.
- Lespinas, F., Ludwig, W. & Heussner, S. 2014 [Hydrological and climatic uncertainties associated with modeling the impact of climate change on water resources of small Mediterranean coastal rivers](#). *Journal of Hydrology* **511**, 403–422. doi:10.1016/j.jhydrol.2014.01.033.
- Lozano, F. C. 1987 *Physical and Chemical Properties of the Soils at the Southern Biogeochemical Study Site. Report BGC-018*.

- Dorset Research Center, Ontario Ministry of the Environment, Dorset, Ontario, Canada.
- Matott, L. S. 2005 *An Optimization Software Tool: Documentation and User's Guide*. Department of Civil, Structural and Environmental Engineering, State University of New York at Buffalo, New York.
- Matott, L. S., Leung, K. & Sim, J. 2011 Application of MATLAB and Python optimizers to two case studies involving groundwater flow and contaminant transport modeling. *Computers & Geosciences* **37** (11), 1894–1899. doi:10.1016/j.cageo.2011.03.017.
- Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models, part 1 – a discussion of principles. *Journal of Hydrology* **10** (3), 282–290. [http://dx.doi.org/10.1016/0022-1694\(70\)90255-6](http://dx.doi.org/10.1016/0022-1694(70)90255-6).
- Natural Resources Canada 2013 Canadian Digital Elevation Model, Product Specifications. Map Information Branch, GeoGratis Client Services. Available at http://ftp.maps.canada.ca/pub/elevation/dem_mne/highresolution_hauteresolution/HRDEM_Product_Specification.pdf.
- Peters, D. L., Buttle, J. M., Taylor, C. H. & LaZerte, B. D. 1995 Runoff production in a forested, shallow soil Canadian Shield basin. *Water Resources Research* **31** (5), 1291–1304. doi:10.1029/94WR03286.
- Schiff, S. L., Devito, K. J., Elgood, R. J., McCrindle, P. M., Spoelstra, J. & Dillon, P. 2002 Two adjacent forested catchments: dramatically different NO₃⁻ export. *Water Resources Research* **38** (12). doi:10.1029/2000WR000170.
- Thyer, M., Kuczera, G. & Bates, B. C. 1999 Probabilistic optimization for conceptual rainfall-runoff models: a comparison of the shuffled complex evolution and simulated annealing algorithms. *Water Resources Research* **35** (3), 767–773. doi:10.1029/1998WR900058.
- Tolson, B. A. & Shoemaker, C. A. 2007 Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research* **43** (1), W01413. doi:10.1029/2005WR004723.
- Wallner, M., Haberlandt, U. & Dietrich, J. 2012 Evaluation of different calibration strategies for large scale continuous hydrological modelling. *Advances in Geosciences* **31**, 67–74. doi:10.5194/adgeo-31-67-2012.
- Yao, H., Deveau, M. & Scott, L. 2009 *Hydrological Data for Lakes and Catchments in Muskoka/Haliburton (1978–2007)*. Data Report 09/01, Ontario Ministry of the Environment, Toronto, Ontario, Canada.
- Yen, H., Jeong, J., Tseng, W.-H., Kim, M.-K., Records, R. M. & Arabi, M. 2015 Computational procedure for evaluating sampling techniques on watershed model calibration. *Journal of Hydrologic Engineering* **20** (7). doi:10.1061/(ASCE)HE.1943-5584.0001095.
- Yen, H., Jeong, J. & Smith, D. R. 2016 Evaluation of dynamically dimensioned search algorithm for optimizing SWAT by altering sampling distributions and searching range. *Journal of the American Water Resources Association* **52** (2), 443–455.
- Zhang, X., Srinivasan, R., Zhao, K. & Liew, M. V. 2008 Evaluation of global optimization algorithms for parameter calibration of a computationally intensive hydrologic model. *Hydrological Processes* **23** (3), 430–441. doi:10.1002/hyp.7152.

First received 3 March 2017; accepted in revised form 16 October 2017. Available online 16 November 2017

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.