HYDROLOGICAL PROCESSES
Hydrol. Process. **28**, 5220–5235 (2014)
Published online 5 September 2013 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/hyp.10001

Rainfall–runoff model parameter estimation and uncertainty evaluation on small plots

Keewook Kim,[1,2]* Gene Whelan,[1] S. Thomas Purucker,[1] Thomas F. Bohrmann,[1,3] Michael J. Cyterski,[1] Marirosa Molina,[1] Yin Gu,[1] Yakov Pachepsky,[4] Andrey Guber[5] and Dorcas H. Franklin[6]

[1] *National Exposure Research Laboratory, Ecosystem Research Division, US Environmental Protection Agency, Athens, GA, 30605, USA*
[2] *Oak Ridge Institute for Science and Education, US Department of Energy, Oak Ridge, TN, 37830, USA*
[3] *Cardno ENTRIX, Raleigh, NC, 27612, USA*
[4] *Agricultural Research Service, Environmental Microbial and Food Safety Lab, US Department of Agriculture, Beltsville, MD, 20705, USA*
[5] *Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lancing, MI, 48824, USA*
[6] *Agricultural Research Service, US Department of Agricultural, and Department of Crop & Soil Sciences, University of Georgia, Athens, GA, 30602, USA*

Abstract:

Four seasonal rainfall simulations in 2009 and 2010 were applied to a field containing 36 plots (0.75 × 2 m each), resulting in 144 runoff events. In all simulations, a constant rate of rainfall was applied then halted 60 min after initiation of runoff, with plot-scale monitoring of runoff every 5 min during that period. Runoff was simulated with the Kinematic Runoff and Erosion/Simulator of Transport with Infiltration and Runoff (KINEROS2/STWIR) field-scale model, whose hydrodynamics are based on the kinematic wave equation. Because of the non-linear nature of the model and a highly parameterized model with respect to the available data, several approaches were investigated to upscale nine runoff-related parameters from a series of small monitored plots to the field scale. Inverse modeling was performed using the model-independent Parameter ESTimation (PEST) algorithm to individually calibrate the nine KINEROS2/STWIR parameters on 36 plots. The parameters were averaged, and bootstrapping was used to assess uncertainty of the parameters via estimation of confidence intervals (CI). A Monte Carlo simulation using the bootstrap results showed reasonable field-scale representation of flow rates. Median values of calibrated parameters were within the 95% CI obtained with bootstrapping. The simulated results for the median values associated with the 90% CI flow rates produced similar trends as those exhibited with the observed data, suggesting that median values of the calibrated parameters from the PEST inverse modeling could be used to represent the field scale. Copyright © 2013 John Wiley & Sons, Ltd.

KEY WORDS watershed modeling; inverse modeling; bootstrap; PEST; KINEROS2; STWIR

Received 11 April 2013; Accepted 6 August 2013

INTRODUCTION

Quantifying rainfall–runoff relationships is a precondition for estimating the hydrologic responses of catchments, evaluating ecological functioning of landscapes and assessing management practices with respect to contaminant transport and environmental protection. The complexity of runoff generation and flow is easy to perceive but difficult to express mathematically without substantial simplifications. Various simplifications have been suggested, and as a result, many rainfall–runoff models have been proposed and used (Beven, 2012).

Scale dependence in parameter values has been noted as the inherent challenge in rainfall–runoff modeling (Brasington and Richards, 1998; Vivoni *et al*., 2007). Information about rainfall–runoff relationships has been mostly collected on a relatively small spatial scale (e.g. laboratory/plot), whereas it is needed for larger spatial units (e.g. field). This is a typical situation for environmental studies, where research-based models generally must be applied to larger areas (extrapolation) with incomplete data coverage (interpolation) (Bierkens *et al*., 2000; Finke *et al*., 2002). Because of spatial heterogeneity, field-scale parameter values can be remarkably different from laboratory-scale measurements, and upscaling bridges the gap (Cushman *et al*., 2002; Neuman and Di Federico, 2003; Deng, 2009). Bierkens *et al*. (2000) divided upscaling methods into four classes: averaging model inputs or outputs, finding representative parameters, averaging model equations and simplifying models. Of these, averaging

*Correspondence to: Keewook Kim, National Exposure Research Laboratory, Ecosystem Research Division, US Environmental Protection Agency, Athens, GA 30605, USA.
E-mail: kim.keewook@epa.gov, keewookkim79@gmail.com

model inputs or outputs is relatively simple, as demonstrated by recent studies (Tegnander and Gimse, 1998; Zhu and Mohanty, 2002; Crow *et al.*, 2005; Kunstmann, 2005; Yoo *et al.*, 2007; Ahrens and Beck, 2008). Averaging can be performed on calibrated input parameters, followed by inspecting the output, or the input remains unchanged, followed by combining (averaging) the output (Bierkens *et al.*, 2000).

Efficient calibration methods have been implemented in several software packages [e. g. PEST (Doherty, 2005; Tonkin and Doherty, 2009), UCODE (Poeter and Hill, 1998) and OSTRICH (Matott, 2005)]. These methods have been widely employed over the years using various hydrological models (Al-Abed and Whiteley, 2002; Durelle *et al.*, 2003; Lin and Anderson, 2003; Rabideau *et al.*, 2005; Gallagher and Doherty, 2007; Immerzeel and Droogers, 2008; Maneta *et al.*, 2008; Guber *et al.*, 2009, 2010, 2011). Model calibration is shown to provide more information (e.g. measurement of accuracy to sample estimates) when combined with the bootstrapping statistical technique (Efron and Tibshirani, 1993). Bootstrapping allows estimation of the sample parameter distribution using simple methods (Varian, 2005), infers variability in an unknown distribution from which data are drawn by resampling (Felsenstein, 1985) and has been used widely to assess variable uncertainty by estimating confidence intervals (CIs) (Felsenstein, 1985; Tegnander and Gimse, 1998; Yoo *et al.*, 2007; Ahrens and Beck, 2008).

The objectives were to develop field-scale values for rainfall–runoff parameters associated with the Kinematic Runoff and Erosion/Simulator of Transport with Infiltration and Runoff (KINEROS2/STWIR) model from small plots distributed throughout a field, and establish an appropriate and reasonable procedure for parameter calibration on a highly parameterized model. One hundred forty four runoff experiments were implemented on 36 small plots, and the relationships between rainfall and runoff were investigated. Using the resulting experimental data set, overland flow was modeled at the plot scale to upscale results to the field scale. Inherent difficulties in the process included reliance on a non-linear runoff model, modeling parameters that are difficult to estimate, a highly parameterized model with respect to available data and a lack of field-scale monitoring. A combination of statistical approaches provides estimated medians and distributions for critical input runoff parameters, evaluates uncertainty of model parameters and performance with the bootstrapping technique and provides field-scale model performance when coupled with a Monte Carlo assessment.

METHODS AND PROCEDURES

Field method

During 2009 through 2010, four seasonal rainfall simulations (Events) were applied to a field containing 36 plots (0.75 × 2 m each), resulting in 144 runoff events. The experimental design for the field is presented in Figure 1, where the coloured plots represent those that were used in the study. Butler *et al.* (2006) describe many characteristics of the study site, which is located in Oconee County, Georgia, USA (33°47′N, 83°23′W, elevation 225 m). The experimental set-up was established on a Cecil coarse sandy loam soil (fine, kaolinitic, thermic and Typic Kanhapludult) with mixed tall fescue-Bermuda grass vegetation on 3–10% slopes (mean of 6%). Experimental plots were delineated with galvanized sheet metal (23-cm width) placed in the ground to a depth of 18 cm, and a runoff collection system was installed at the down-slope end of each plot; each plot contained five rain gauges. This is similar to the runoff collection design reported by Franklin *et al.* (2006). Prior to all rainfall simulations, the forage

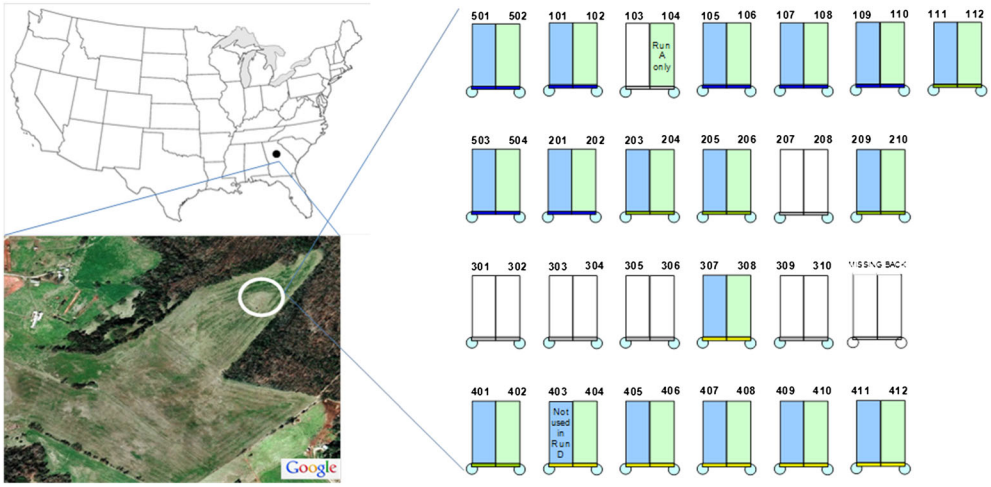

Figure 1. Experimental design for the field study in Watkinsville, Georgia, USA, where the coloured plots represent those that were used in the study

canopy height was standardized by trimming the vegetation to a 10-cm stubble height.

Two rainfall simulators (Tlaloc 3000, Joern's Inc., West Lafayette, IN) were used to apply rain, with each simulator applying water simultaneously on two adjacent plots. In all simulations, a constant rate of rainfall was applied at a pressure of 43.1 kPa and was continued for an additional 60 min after initiation of runoff. Rainfall rates from plot to plot ranged from 44 to 85 mm/h, with an overall mean and median of 67 and 66 mm/h, respectively. Variation in rainfall rates and patterns for different plots was due to the use of different nozzles and hydraulic lines and nonuniform spray. The duration of rainfall, which depended on the time until initiation of runoff, was a function of season and plot and varied from a mean of 79 min during the wet spring to 188 min during the dry summer. The amount of rainfall varied from a mean of 93 mm in the wet spring to 196 mm in the dry summer. For reference, the 2-year and 100-year 60-min precipitations for this area are about 45 and 97 mm, respectively (Frederic *et al.*, 1977). Runoff volume was measured in 5-min increments over the 60-min runoff period.

Table I summarizes field conditions during the Events. The highest air temperatures were during Event C, followed by Event D, with more moderate temperatures during Events A and B. Event A consistently had the highest antecedent rainfalls, whereas Event C had the lowest. When comparing Events B and D, Event D had higher 10-day and 30-day antecedent rainfalls, but Event B had higher 50-day and 90-day antecedent rainfalls. The wettest and driest initial soil moisture conditions were seen during Events A and C, respectively. When coupling the lower 50-day and 90-day antecedent rainfall with the higher temperatures during Event D, Event D had a lower initial soil moisture which resulted in longer delays prior to initiation of runoff.

Table I. Summary of meteorological statistics on experimental plots and field conditions for each Event

Event	Date	Field simulation[c]			Antecedent rainfall[d] (mm)				Average daily temperature[d] (°C)			Degree of saturation of soil	
		Average rainfall amount (mm)	Average rainfall duration (min)	Average rainfall intensity (mm/hr)	10-day	30-day	50-day	90-day	Max	Min	Mean	NOAA[d]	Plot[e]
A	09/29/2009–10/16/2009 ($n = 37^a$)	94 (11.4[b])	84 (10.8[b])	67 (4.7[b])	134.0	266.6	385.9	451.3	22.4	11.3	17.0	0.173	—
B	03/02/2010–03/18/2010 ($n = 36^a$)	93 (13.8[b])	79 (9.3[b])	70 (7.2[b])	10.4	104.1	260.6	490.4	13.7	2.8	8.9	0.166	0.03–1
C	06/22/2010–07/08/2010 ($n = 36^a$)	196 (36.4[b])	187 (32.1[b])	63 (6.0[b])	30.3	43.9	116.1	158.1	29.3	22.7	26.3	0.095	0.02–0.55
D	10/05/2010–10/21/2010 ($n = 35^a$)	142 (44.5[b])	130 (39.6[b])	66 (5.5[b])	117.7	129.4	159.3	270.8	20.8	13.3	17.2	0.147	0.16–0.87

[a] '*n*' refers to the number of plots.
[b] Standard deviation.
[c] Observed directly from the field.
[d] Observed at NOAA weather station (GA Watkinsville 5 SSE located at 33°47′N, 83°23′W, elevation 225.9 m). Period of antecedent rainfall was counted from the first date of the each Run. Average daily temperature was estimated during each Event. Degree of saturation is depth-averaged value (0–50 cm depth) at AM12:00 on the start date of each Event.
[e] Range of degree of saturation directly measured on each plot prior to the simulation. There is no soil moisture observation in Event A.

Rainfall–runoff modeling

Using a modified version of the KINEROS2 model (Woolhiser *et al.*, 1990; Semmens *et al.*, 2008; http://www.tucson.ars.ag.gov/kineros/) coupled with the STWIR model (Guber *et al.*, 2009), overland runoff was simulated for the Watkinsville plots. KINEROS2/STWIR is an event-oriented, physically based model that describes the processes of interception, infiltration, surface runoff and erosion from small agricultural and urban watersheds. Although only a single plane was necessary in these simulations, KINEROS2/STWIR can describe a watershed by a cascade of planes and channels. Overland flow, channel flow, erosion and sediment transport are solved with partial differential equations using finite-difference techniques; spatial variation of rainfall, infiltration, runoff and erosion parameters can be accommodated. Using the method of characteristics, overland water flow is described by the

kinematic wave equation (Eagleson, 1970; Woolhiser *et al.*, 1990; Singh, 1997; Guber *et al.*, 2009):

$$\frac{\partial y_0}{\partial t} + \frac{\partial q_0}{\partial x} = i - f \tag{1}$$

where y_0 is average depth of overland flow (cm), t is time (h), q_0 is discharge per unit width of overland flow ($cm^2\,h^{-1}$), x is flow distance (cm), i is rainfall intensity ($cm\,h^{-1}$) and f is infiltration rate ($cm\,h^{-1}$). Discharge per unit width of overland flow can be expressed as a function of flow depth:

$$q_0 = \alpha_0 y_0^{m_0} \tag{2}$$

where parameters α_0 and m_0 are defined according to Manning's [Equation (3)] hydraulic resistance laws, respectively, as

$$\alpha_0 = n^{-1} S_0^{1/2}, m_0 = 5/3 \tag{3}$$

where n is Manning's roughness coefficient for overland flow, and S_0 is the average gradient of the overland flow element (i.e. surface land slope; $cm\,cm^{-1}$). The Parlange equation (Parlange *et al.*, 1982) is employed to estimate infiltration rate through the soil surface:

$$f = K_s \left[1 + \frac{\sigma}{exp(\sigma I/B)}\right] \tag{4}$$

$$B = (G_n + y_0)(\theta_s + \theta_i) \tag{5}$$

where K_s is saturated hydraulic conductivity ($cm\,h^{-1}$), σ is a dimensionless parameter that represents soil texture, I is infiltrated depth (cm), G_n is net capillary drive (cm) and θ_s and θ_i are soil volumetric water contents at saturation and initial conditions, respectively ($cm^3\,cm^{-3}$). KINEROS2/STWIR also considers rainfall interception, soil moisture redistribution and surface unevenness. To describe these features, seven additional parameters are needed: interception depth and fraction of canopy cover for rainfall interception; pore size distribution index; coefficient of variance of saturated hydraulic conductivity; volumetric rock fraction for soil moisture redistribution; average microtopographic relief and spacing for surface unevenness.

Model parameterization

This study employed the model-independent Parameter ESTimation (PEST) software (Doherty, 2005) and bootstrap techniques (Efron and Tibshirani, 1993) to calibrate KINEROS2/STWIR parameters and evaluate model performance for the 36 Watkinsville plots. Parameter calibration included two approaches: a three-step sequential procedure, using the inverse model PEST, in which parameters varied by plot and a PEST/Bootstrap procedure in which parameter values were averaged across plots.

The sequential three-step calibration procedure was implemented because there were more calibration parameters than necessary data, so each calibration procedure used parameter values from the preceding step as initial values. First, the parameters were calibrated by plot by Event. Common parameters that are independent of Event were then calibrated by plot across Events while holding the other parameters constant. Finally, the common parameters were held constant, and other parameters were recalibrated by plot by Event.

The PEST/Bootstrap calibration procedure was performed to reduce over-parameterization and determine parameter values at a field scale (i.e. across plots). It also evaluated parameter uncertainty and model performance. Details on PEST and parameter calibration procedures are presented in the following sections.

Parameter estimation. In the PEST algorithm, an objective function related to the difference between model-calculated parameters and field observations is minimized iteratively. The relationship between the set of model parameters (x) and field observations (y) can be written as:

$$y = M(x) \tag{6}$$

where x and y are both vectors with n (number of parameters) and m (number of observations) entries, respectively, and M is a non-linear function that relates model parameters to field observations. The linearization of Equation (6) by Taylor's theorem is

$$\hat{y} = y + H(\hat{x} - x) \tag{7}$$

where $\hat{x}$ and $\hat{y}$ are the vectors with the estimated parameters and model-calculated observations, respectively, and H is a Jacobian matrix of partial derivatives with m rows and n columns that is calculated as

$$H[i,j] = \frac{\partial M(\hat{x})[i]}{\partial x[j]} \tag{8}$$

During PEST iterations, the parameter vector, $\hat{x}$, is updated by adding an upgrade vector u:

$$u_k = \left(H_k^T O H_k\right)^{-1} H_k^T O\left(y - \hat{y}_k^-\right) \tag{9}$$

$$\hat{x}_k^+ = \hat{x}_k^- + u_k \tag{10}$$

where k is the iteration level, and T indicates a transpose in the matrix operation. The superscripts negative and positive indicate prior and posterior estimates, respectively. O is the observation weight matrix. The PEST iterations are repeated until the following objective function converges to the minimum value:

$$\Phi = [\hat{y} - y - H(\hat{x} - x)]^T O[\hat{y} - y - H(\hat{x} - x)] \quad (11)$$

A detailed description of the PEST algorithm can be found in Doherty (2005). PEST has been widely used for a number of inverse applications including rainfall–runoff (Gallagher and Doherty, 2007; Immerzeel and Droogers, 2008), chemical and microbial transport (Baginska *et al*., 2003) and those applications with the KINEROS2/STWIR model (Guber *et al*., 2009, 2010, 2011).

Initialization of model parameters. Thirteen key runoff parameters are used in KINEROS2/STWIR to describe overland flow from a plane. Four were measured in the field: slope angle of a plane (Slope), initial degree of saturation (θ_i or Sat), fraction of canopy cover (Canopy) and total porosity (θ_s or Por). These were fixed at their measured values in all calibration procedures, except for the initial degree of saturation, which had measured values for only three of the four Events. Nine sample parameters required calibration: Manning's roughness coefficient (*n* or Manning); interception depth (In); average micro-topographic relief (Relief); average micro-topographic spacing (Spacing); saturated hydraulic conductivity (K_s); coefficient of variance of saturated hydraulic conductivity (C_v); mean capillary drive (G); volumetric fraction of rock (Rock); and pore size distribution index (Dist). Table II presents upper and lower bounds for the nine 'calibrated' parameters which were determined to be within physically sufficient ranges based on values suggested or used in previous studies and obtained from preliminary simulations. As part of the sequential three-step calibration procedure, six parameters were considered 'common' and maintained at the same value across seasons (i.e. Events): Slope, Relief, Spacing, Por, Rock and Dist. Table III summarizes measured, calibrated, and common parameters.

Sequential three-step calibration procedure. Central to the sequential three-step calibration procedure was an iterative approach, as outlined in Figure 2(a), to assign values to the nine calibration parameters to avoid them being assigned values equivalent to their lower or upper bounds. Initial values were assigned, and PEST was used to calibrate the parameters simultaneously. Updated calibrated values were assigned as starting values for the next iteration unless those values were at the bounds

Table II. Upper and lower bounds of calibrated parameters used in this study based on the values suggested or used in references and obtained from preliminary simulations

Parameter*	Bound		Values suggested or used in the reference	Reference
	Lower	Upper		
Manning	0.05	0.8	0.1–0.63	Woolhiser *et al*. (1990)
			0.053–0.8	MacArthur and DeVries (1993)
C_v	0.01	50.0	0.1–2.	http://www.tucson.ars.ag.gov/kineros/
			0.1	Guber *et al*. (2011)
			0.02–27.3	Preliminary simulations
Relief (mm)	0.01	400.0	0, 0.5	Guber *et al*. (2011)
			0.65–19.0	Preliminary simulations
Spacing (m)	0.001	2.0	0.3	Guber *et al*. (2011)
			0.002–1.75	preliminary simulations
In (mm)	0.01	300.0	0.5–4.1	Woolhiser *et al*. (1990)
			4.77–101.3	Preliminary simulations
K_s (mm/hr)	0.2199	266.3	0.6–210.	Woolhiser *et al*. (1990)
			0.2199–266.3	Meyer *et al*. (1997)
			0.3–73.3	Guber *et al*. (2009)
			17.2, 48.3	Guber *et al*. (2011)
G (mm)	0.1	500.0	50.0–410.	http://www.tucson.ars.ag.gov/kineros/
			46.0–407.	Woolhiser *et al*. (1990)
			1.0–263.	Guber *et al*. (2009)
			100.0, 306.	Guber *et al*. (2011)
			16.6–133.3	Preliminary simulations
Rock	0.0	0.2	0.0–0.11	Preliminary simulations
Dist	0.14	1.43	0.15–0.694	Rawls *et al*. (1982)
			0.14–1.43	Meyer *et al*. (1997)

*Manning: Manning's roughness coefficient; C_v: coefficient of variance of saturated hydraulic conductivity; Relief: average micro-topographic relief; Spacing: average micro-topographic spacing; In: interception depth; K_s: saturated hydraulic conductivity; G: mean capillary drive; Rock: volumetric fraction of rock; Dist: pore size distribution index.

Table III. Summary of sequential calibration procedures for the 13 KINEROS2/STWIR parameters, indicating those that varied by plot by Event, by plot across Events and across plots by Event

Parameters	Parameter		Sequential Calibration Part 1 (Fig. 1a)									PEST/Bootstrap Calibration Part 2 (Fig. 1b)		
				First (1st)		Second (2nd)		Third		Median value determined from the Third Calibration				
	Measured	Calibrated	Common Parameters across Events	Varies by Plot by Event	Varies by Plot across Events	Varies by Plot by Event	Varies by Plot across Events	Varies by Plot by Event	Varies by Plot across Events	Varies across Plots by Event	Varies across Plots across Events	Value Varies by Plot by Event	Value varies by Plot across Events	Values Averaged across Plots by Event
Slope	√		√		A		A		A		A		A	
Por	√		√	A			C		2nd		2nd		2nd	
Sat	√			A		A		A		A		A		
Canopy	√				A		A		A		A		A	
Manning		√		C		1st		C		M				C
C_v		√		C		1st		C		M				C
Relief (mm)		√	√	C			C		2nd		M			C
Spacing (m)		√	√	C			C		2nd		M			C
In (mm)		√		C		1st		C		M				C
K_s (mm/hr)		√		C		1st		C		M				C
G (mm)		√		C		1st		C		M				C
Rock		√	√	C			C		2nd		M			C
Dist		√	√	C			C		2nd		M			C

A = assigned the measured value from field sampling; C = calibrated; M = computed median; 1st = value fixed to the first calibration results; 2nd = value fixed to the second calibration results.

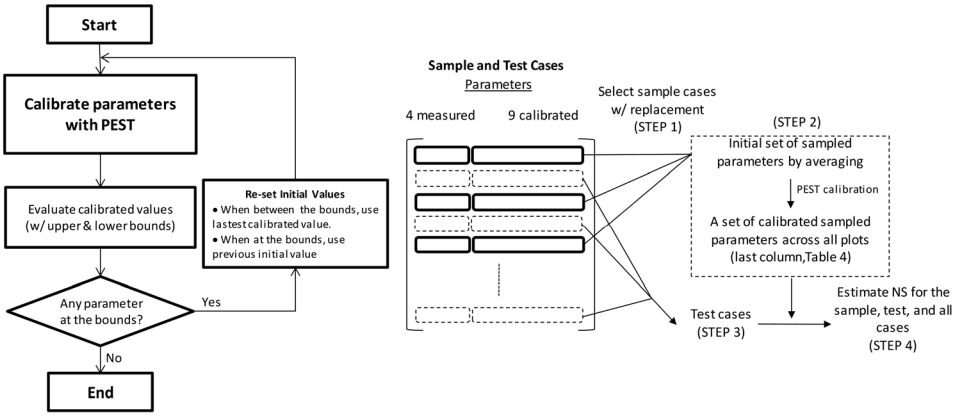

Figure 2. (a) Flow chart of iterative parameter calibration procedure by plot and (b) PEST/bootstrap calibration by Event, coupled with single PEST calibration, for each iteration. Solid and dashed rectangles refer to sample and test cases, respectively

of their initial ranges (Table II). Estimates at the bounds were rejected, and values reverted to the prior estimate. The calibration was repeated plot by plot, for each of the 144 plots, until all parameters were within the bounds outlined in Table II.

The sequential three-step calibration procedure associated with the 13 KINEROS2/STWIR parameters is presented in Table III using the process outlined in Figure 2(a). The first calibration determined an initial estimate for the nine calibrated parameters by plot by Event, fixing the measured parameters. In the second calibration, five of the six common parameters (i.e. excluding Slope, which was constant) were calibrated simultaneously for all Events by PEST, where the initial values for the common parameters were set to their averaged values from the first calibration across Events, and the other parameters (Table III) were fixed to their calibrated values. Upper and lower bounds were assumed as $1.5\times$(largest calibrated value) and $0.5\times$(smallest calibrated value), respectively. In the third calibration, common or measured parameters were fixed at values from the second calibration, and the remaining parameters were recalibrated by plot by Event. At the end of the first and the third calibrations, Nash–Sutcliffe (NS) model efficiency between observed and calculated cumulative flows was estimated for each plot as follows:

$$NS = 1 - \frac{\sum_{t=1}^{T}(O^t - M^t)^2}{\sum_{t=1}^{T}\left(O^t - \overline{O}\right)^2} \tag{12}$$

where O^t, $\overline{O}$ and M^t are observed cumulative flows at time step t, average cumulative value, and modeled cumulative flows at time step t, respectively. T is the number of the simulation time steps on each plot.

Given the large number of calibration parameters and the non-linear nature of the KINEROS2/STWIR model, inverse modeling using PEST was sensitive to initial values. Although not a requirement, results from the third calibration were used as initial conditions for the sensitivity analysis and PEST/Bootstrapping calibration procedure.

Sensitivity analysis. Parameter values that drive models are subject to change and error, and a sensitivity analysis is the investigation of these potential changes and errors and their impacts on conclusions drawn from the model (Pannell, 1997). A sensitivity analysis was performed on the KINEROS2/STWIR calibrated parameters using Event B plots. With a baseline assumption of NS equaling unity (i.e. modeled output is the same as the observations), the variation of error in NS (i.e. 1 – NS) was determined to be a function of the percent change in each parameter within the range identified in Table II, holding the other parameters to values obtained from the third calibration.

PEST/Bootstrap calibration. The purpose of the bootstrap procedure is to estimate CIs for calibrated parameters and uncertainty through sample distributions. Statistics describing model performance were obtained by calibrating a set of parameters to the KINEROS2/STWIR model throughout all plots using PEST and estimating errors between observations and model output. Using the third calibration results as a starting point, a single application of the bootstrap procedure is described in Figure 2(b), where plots for each iteration are divided into sample (i.e. randomly selected with replacement) and test (unselected) plots, differentiated by Event (A, B, C or D), where the sample and test plots are rolled into sample and test cases, respectively. Plots that are not selected as sample plots are automatically assigned to test plots. Each row in Figure 2(b) corresponds to a plot associated with a specific Event, which is described by four measured and nine calibrated parameters; values of which were determined during the calibration procedure outlined in Figure 2(a).

In Step 1 of the procedure outlined in Figure 2(b), the sample plots were randomly selected with replacement from the total plots by Event. For Step 2, each of the nine calibrated parameters [solid boxes in Figure 2(b)] was averaged across the sample plots selected in Step 1, representing initial values for a simultaneous recalibration using PEST that resulted in a new set of values that are the same and applicable across all plots. In this step, only a single PEST calibration was performed without regard to whether the calibrated parameters are at their bounds. Under Step 3, while maintaining the four field-measured parameters (Slope, Sat, Canopy, and Por) as constant, intrinsic parameters for each plot, the KINEROS2/STWIR model was executed on the test plots as single, deterministic applications. From these applications, a fourth step determined the NS between observed and calculated flow rates using Equation (13), where the mean model covers the entire simulation period, providing cross-validation between sample and test plots:

$$NS = 1 - \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}\left(Q_{o,i}^t - Q_{m,i}^t\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T_i}\left(Q_{o,i}^t - \overline{Q_{o,i}}\right)^2} \tag{13}$$

$$\overline{Q_{o,i}} = \frac{1}{T_i}\left(\frac{1}{N}\sum_{i=1}^{N} O_i\right) \tag{14}$$

where N is number of plots (i.e. number of sample, test, or all plots), and T_i is the number of total simulation time steps on ith plot. $Q_{o,i}^t$ and $Q_{m,i}^t$ are observed and modeled flow rates per unit area of depth (cm h^{-1}) on the ith plot at time step t (h), respectively. O_i is the observed total flow depth (cm) on the ith plot. Steps 1 through 4 were repeated 10 000 times.

Parameter uncertainty and model performance

Evaluation of parameter uncertainty. A principal goal of bootstrap theory is to produce a well-behaved CI (Efron and Tibshirani, 1993). Because of the bounds specified on the KINEROS2/STWIR parameters, the distributions of parameters obtained by bootstrap could be skewed (Efron, 1987), which will likely be a common problem in physical models with bounded parameter spaces. If this is not accounted for properly, skewed bootstrap distributions will lead to biased bootstrap CIs and poor coverage properties.

Efron (1987) introduced bias-correlated, accelerated bootstrap CIs (BCa intervals) as a remedy. Typically, a CI can be estimated by the percentile method:

$$\left(\hat{\theta}_{lo}, \hat{\theta}_{up}\right) = \left(\hat{\theta}^{*(\alpha)}, \hat{\theta}^{*(1-\alpha)}\right) \tag{15}$$

where $\hat{\theta}_{lo}$ and $\hat{\theta}_{up}$ are lower and upper CIs, respectively, and $\hat{\theta}^{*(\alpha)}$ indicates the αth percentile of bootstrap replications. Instead of Equation (15), the BCa method uses Equation (16) for CI.

$$\left(\hat{\theta}_{lo}, \hat{\theta}_{up}\right) = \left(\hat{\theta}^{*(\alpha_1)}, \hat{\theta}^{*(\alpha_2)}\right) \tag{16}$$

where

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha)}}{1 - \hat{a}\left(\hat{z}_0 + z^{(\alpha)}\right)}\right)$$

$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\alpha)}}{1 - \hat{a}\left(\hat{z}_0 + z^{(1-\alpha)}\right)}\right)$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function, and $z^{(\alpha)}$ is the αth percentile point of a standard normal distribution. $\hat{z}_0$ and $\hat{a}$ can be calculated using Equations (17) and (18).

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#\left\{\hat{\theta}^*(b) < \hat{\theta}\right\}}{B}\right) \tag{17}$$

$$\hat{a} = \frac{\sum_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^3}{6\left\{\sum_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^2\right\}^{3/2}} \tag{18}$$

where $\hat{\theta}^*(b)$ is the bootstrap replication, $\hat{\theta}$ is the original estimate, $\hat{\theta}_{(i)}$ is the leave-one-out jackknife estimate, $\hat{\theta}_{(\cdot)}$ is the average value and B is the number of bootstrap replications.

Evaluating field-scale model performance with Monte Carlo simulation. The PEST/Bootstrap procedure estimates a set of parameters common to all selected sample cases (i.e. all plots within an Event), whose results may overestimate flows on some plots and underestimate flows on others. Thus, individual model performance (i.e. NS) on a plot-by-plot basis, using error statistics estimated by comparing individual plot observations and calculations, may not adequately reflect overall field-scale model performance which could be better served by considering flows from all plots. Based on distributions of each parameter obtained by PEST/Bootstrap calibration, a Monte Carlo analysis was performed to quantify uncertainty at the field scale. Each parameter value was randomly and individually selected from its distribution obtained by PEST/Bootstrap calibration, and a set of parameters was then applied to the KINEROS2/STWIR model for all plots. This procedure was repeated through 10 000 iterations, with results being compared with observed flows for each Event.

RESULTS AND DISCUSSION

Parameter calibration on plot scale

Sequential three-step parameter calibration with PEST. The nine KINEROS2/STWIR calibrated parameters were initially calibrated by plot by Event using PEST ['First (1st)' in Table III]; therefore, each calibrated parameter had a different value for different Events on each plot. For most plots, parameters were calibrated within three iterations [Figure 2(a)], with a maximum number of 30 iterations. Over-parameterization and model nonlinearity contributed to excellent model performance (i.e. NS ranges from 0.0996 to 1.0, with a mean of 0.9765, throughout all plots), as there are 12 runoff data points and nine calibrated parameters per plot. Results of a typical first calibration are presented in Figure 3, as illustrated by Plot B407.

Based on the first PEST calibration [i.e. 'First (1st)' in Table III], the common parameters were simultaneously recalibrated through the seasons [i.e. 'Second (2nd)' in Table III]. Following this, the remaining calibrated parameters were individually recalibrated by plot for each Event (i.e. 'Third' in Table III). The calibrated results matched well with the observed values, where NS ranged from 0.8449 to 1.0, with a mean of 0.9841 throughout all plots, excluding three that had negative NS values. Results of a typical third calibration are illustrated by Plot A108 (Figure 3).

Sensitivity analysis. Figure 4 presents results of the sensitivity analysis on the nine calibrated parameters for

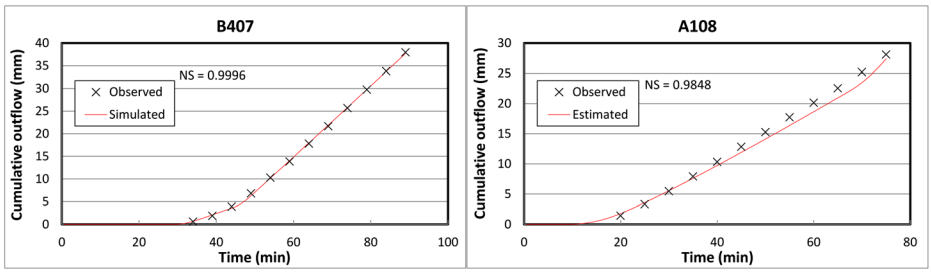

Figure 3. Typical observed and simulated cumulative outflow and NS for Plots B407 and A108, using the calibration procedure outlined in Figure 2(a), corresponding to the 'first' and 'third' calibrations, respectively, summarized in Table III. NS values were estimated using cumulative flow

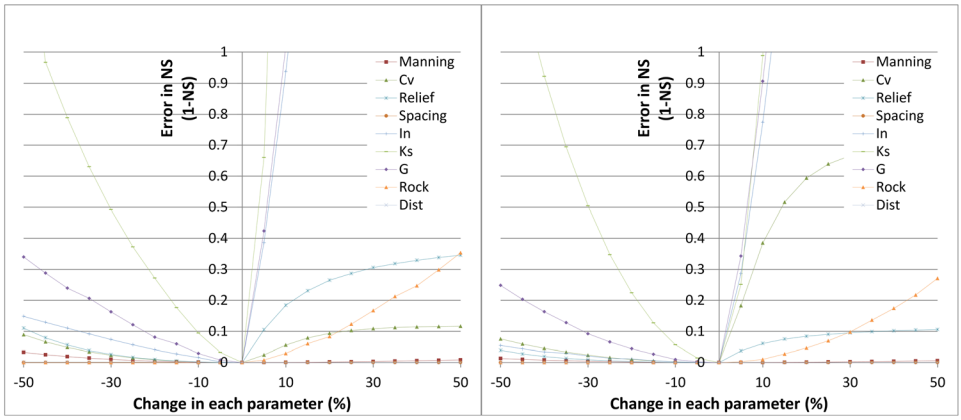

Figure 4. Sensitivity analysis for plots B107 and B408, illustrating the variation of error in NS as a function of the percent change in parameters (corresponds to 'third calibration' in Table III). X-axis represents the % change in each parameter within its range, holding other parameters fixed

plots B107 and B408. As these two illustrate, there can be significant variability in parameter sensitivity from plot to plot because of nonlinearity in the KINEROS2/STWIR model. The number of plots by parameter that exhibits a variation of error in NS larger than 0.3 is summarized in Table IV. The parameters naturally grouped themselves into two categories – more sensitive (Spacing, In, Ks, G, C_v and Relief) and less sensitive (Manning, Rock and Dist). For example, In and K_s exhibited variations of error in NS larger than 0.3 for all 36 Event B plots and Rock could have been in the more sensitive category.

Evaluation of parameter uncertainty and field-scale model performance

Parameter calibration with PEST/Bootstrap. Using the procedure outlined in Figure 2(b), PEST/Bootstrap parameter calibrations were performed for each Event; distributions of the nine calibrated parameters are presented as box plots in Figure 5, with median values ('◦') estimated from the distributions for Events A through D. Overlaid on the box plots are median values ('×') from the third calibration. The PEST/Bootstrap ('◦') and third calibration's median values ('×') were generally located near each other or within the first and third quartiles of the distribution. It should be noted that several parameters were calibrated at their lower/upper bounds; for example, the average micro-topographic spacing was calibrated at its lower bound (=0.001), which is near zero, in 2248 iterations. Because its physical lower bound is zero, the value can be considered reasonable.

On the other hand, when parameters have little influence on output variability, their values, as determined from the inverse calculations, can vary within the specified range

Table IV. Number of plots by parameter with a variation of error in NS larger than 0.3

	Dist	Manning	Rock	C_v	Relief	Spacing	G	In	K_s
Number of plots in sensitive bin	3	9	19	26	28	33	33	36	36
Sensitivity	less	less	less	more	more	more	more	more	more

The values in the table represent the number of plots, out of 36, in Event B where the output was sensitive to change by the corresponding parameter.

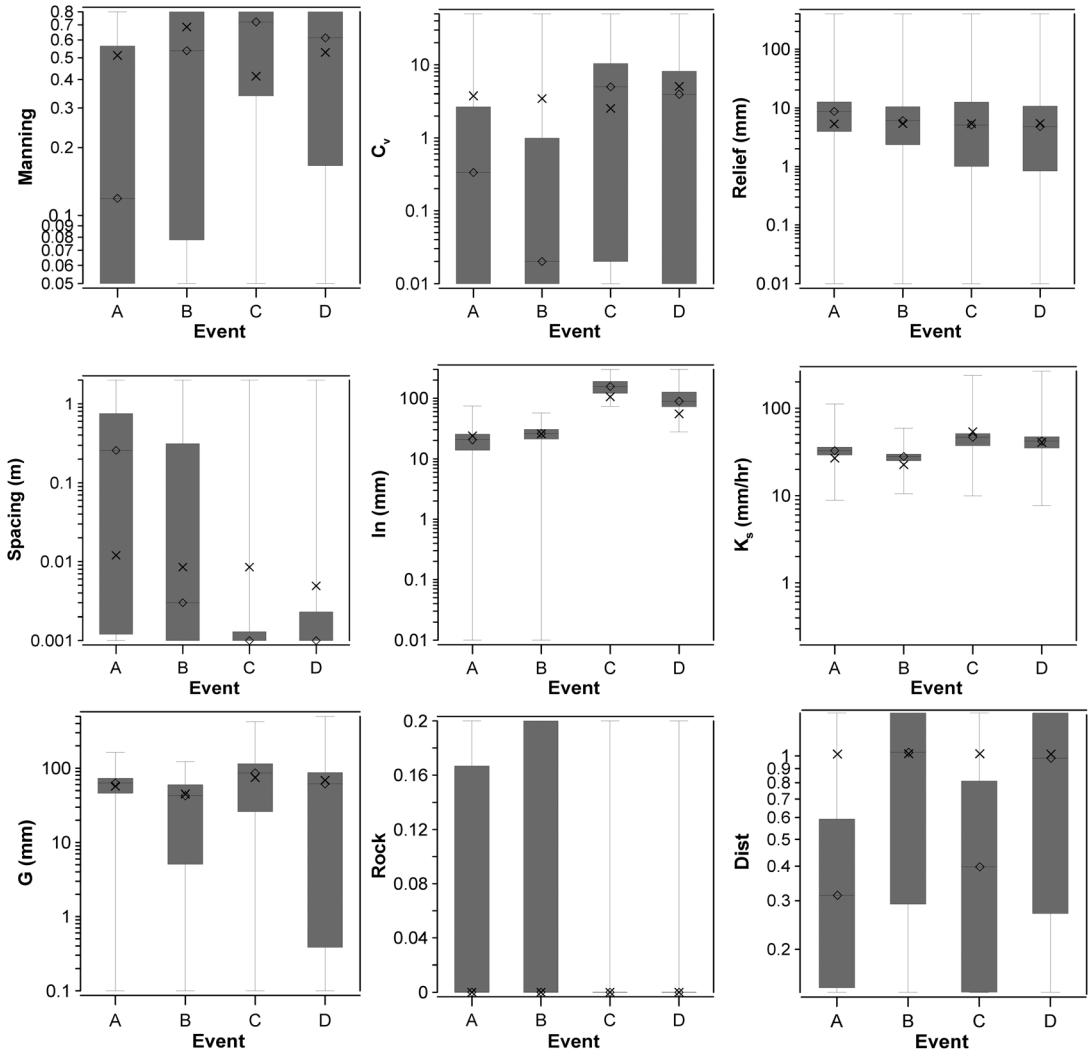

Figure 5. Box-plots and median values ('◦') of calibrated parameters by PEST/bootstrap throughout Events. Y-axis represents the lower and upper bounds listed in Table II. '×' represents median of the third calibration result in Table III

without severely impacting the calibration process, which would include accumulations at a boundary, as illustrated by the least sensitive parameters (i.e. Manning, Rock and Dist). For example, in Event A, Manning's roughness coefficient was calibrated at its lower bound (=0.05) in 2812 iterations; Manning, Rock and Dist were calibrated at the bounds from 2.5 to 36.7% with mean of 17.5%. The more sensitive parameters were calibrated at the bounds (0–7.6% with mean of 0.7%) in a relatively small number of iterations.

Figure 6 presents the distributions of NS on sample, test and all plots for each Event, where the mean model covers the entire simulation period, which includes nonzero values prior to initiation of runoff. Although a positive NS occurred in most iterations for Events A and B (98.6 and 99.9%, respectively) in sample plots, Events C and D rarely exhibited a positive NS (9.60 and 5.79%, respectively). These results could have been caused by uncertainty in parameters, influence of parameters, correlations between parameters that are unmeasured or unaccounted for in this analysis and/or relatively poor performance of KINEROS2/STWIR model under drier conditions.

Evaluation of parameter uncertainty. Figure 7 presents 95% CIs by Event of the six more sensitive KINEROS2/STWIR calibrated parameters, using BCa method; a similar analysis producing similar results was conducted using the percentile method. The results are based on the PEST/Bootstrap calibration procedure [i.e. see Figure 2(b)] and illustrate similar ranges in parameter CIs throughout the Events except for interception depth (In). Because of drier soil moisture conditions in Events C and D (Table I), interception depth (In), which accounts for rainfall being held on foliage by surface tension, appears to calibrate to relatively larger values for Events C and D than Events A and B.

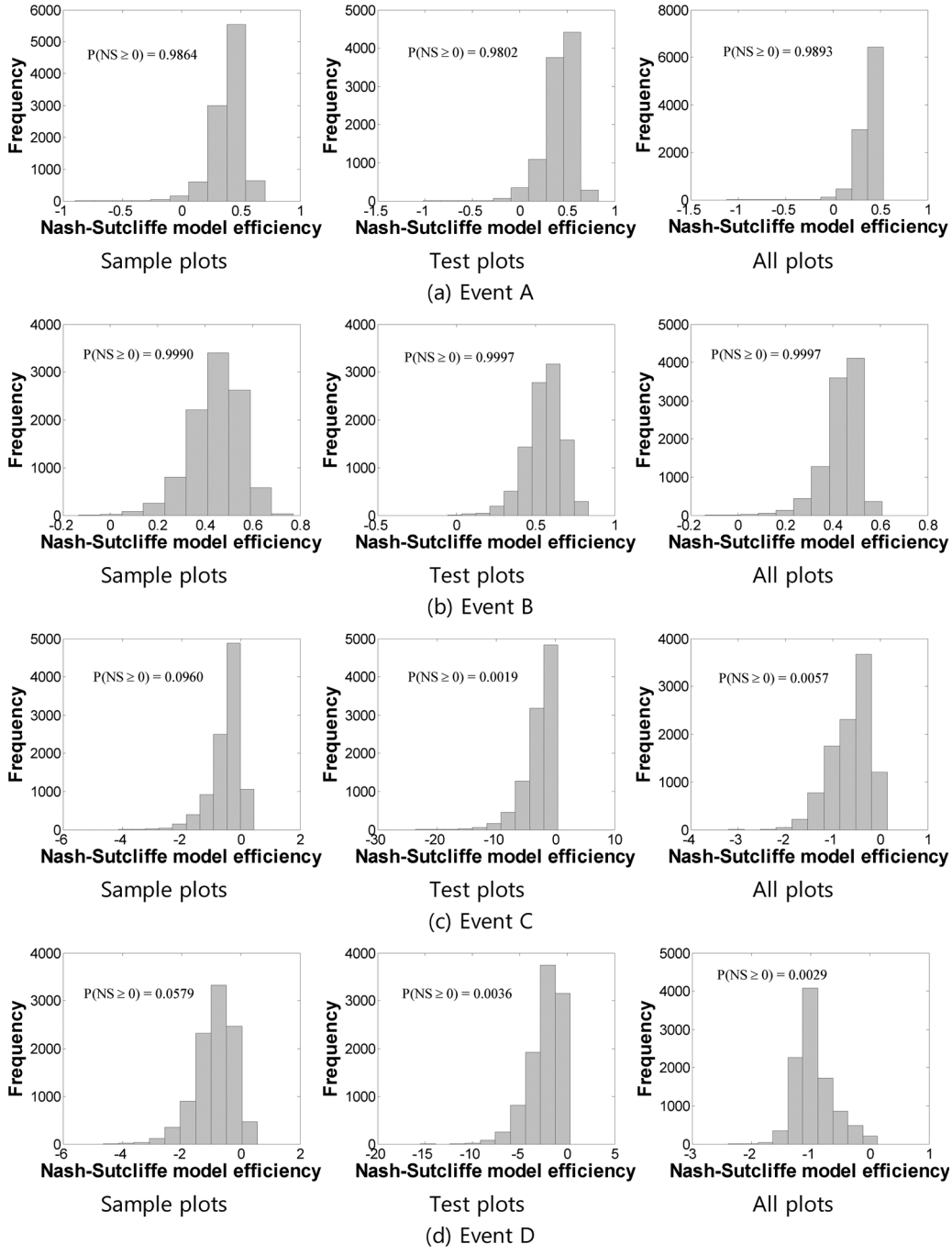

Figure 6. Distribution of NS on sample, test and all plots for each Event. NS values were estimated using time-varying flow rates

Figure 7 also accounts for commonality of the parameters across plots by Events (i.e. seasons). The impact of not considering commonality across plots is reflected in the median (i.e. 50th percentile) parameter values ('×' in figures) that were obtained after the third calibration. Although a number of the median values comfortably occur within the 95% CIs, 37.5% are outside. The relationship between the third calibration and the follow-on PEST/Bootstrap calibration is presented in Table III. The third calibration accounts for consistency across Events through the common parameters; hence, the PEST/Bootstrap procedure inherently reflects considerations across Events through its initial conditions.

Evaluation of field-scale model performance and its uncertainty. Figures 6 and 8 present overall field-scale model performance. Figure 6 summarizes the distribution of the NS model efficiency which compares every flow

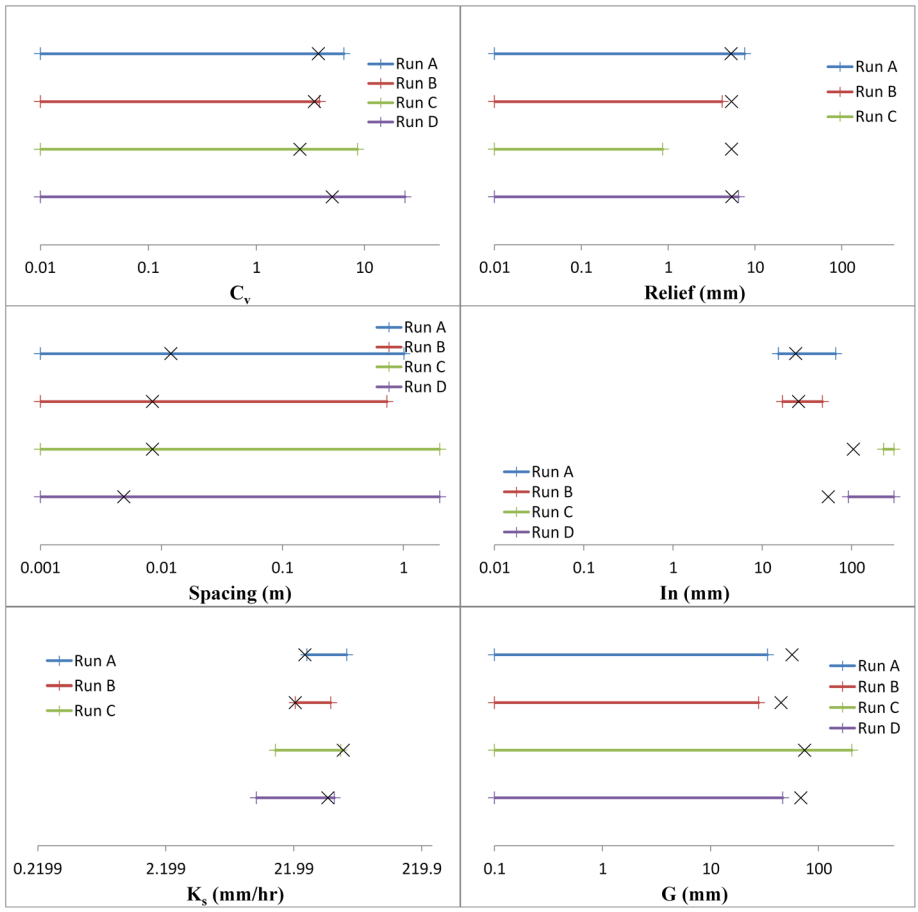

Figure 7. 95% confidence interval of parameters obtained by BCa method (step 2 results outlined in Figure 2). '×' represents the median of the third calibration results in Table III. The range of the x-axis represents the lower and upper bounds in Table II

simulation and observation on a plot-by-plot basis, using the procedure outlined in Figure 2(b). Figure 8 summarizes the results of the Monte Carlo analysis by comparing variability between simulated and observed flow rates by Event. The left-most plots in Figure 8 present the 95% (red) and 90% (blue) CIs for the observed flow rates, as well as the median (black) curve. The centre plots present the 90% CI of the flow rates, as they vary in time. The grey region indicates the 95% CIs about the lower and upper observed 90% CI curve. The solid black curve indicates simulated flows from the Monte Carlo analysis (i.e. variability across plots), and the dashed-red curve indicates the 90% CI for the simulated flow rates using median values from the third calibration (i.e. variability across Events). The right-most plots present median flow rates, as they vary in time, and the grey region indicates the 95% CIs about the observed median curve, as they vary in time. The solid black curve indicates the simulated median flow rates from the Monte Carlo analysis, and the dashed-red curve indicates the simulated flow rates using median values from of the third calibration. There is no simulated median flow curve (solid black line) for Event C because more than 50% of the values were zero.

The vertical red lines in Figure 8 indicate the end time associated with the plot that has the shortest simulation duration in the Event. Because each plot in the field experiments was stopped 60 min after initiation of runoff, information after the vertical red line does not fully represent the characteristics of all plots at the field scale, but it was presented for completeness because any conclusions based on these data should be considered carefully. For example, when all plots in the field are considered in the analysis (i.e. left of the vertical red line), the observed and simulated (i.e. Monte Carlo and median) results associated with the 90% CI flow rates (i.e. centre figures in Figure 8) exhibit similar trends and tend to fall within the upper and lower 95% CIs of the observed data. The same can be said for medians of the flow rates (i.e. right-most figures in Figure 8) for Events A and B. Because of the disparity in runoff start times for Events C and D, it is difficult to draw conclusions that lend insight into why Events C and D do not behave as well as Events A and B in the calibration process.

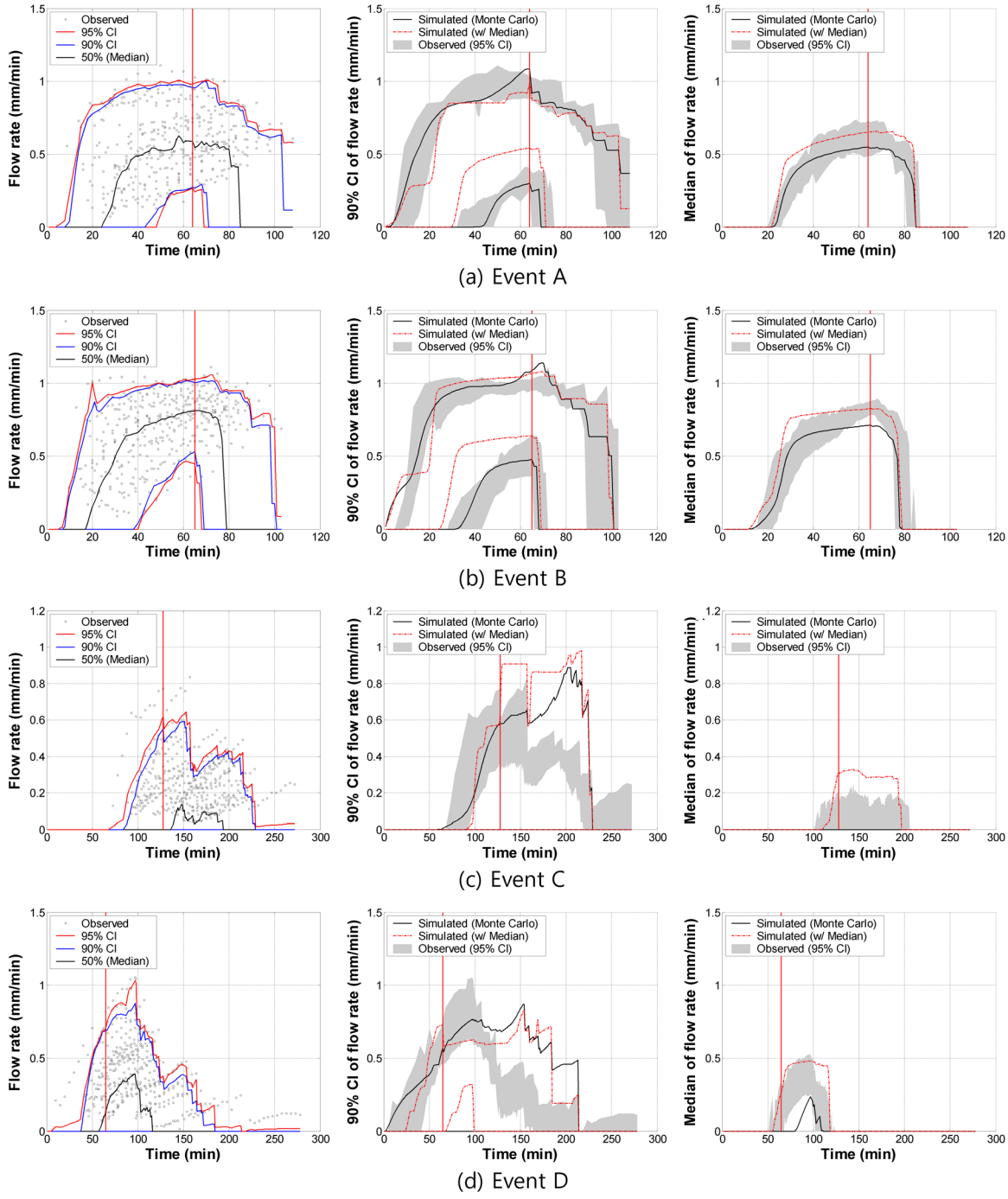

Figure 8. Observed flow rate from each plot and its confidence interval in time (left), and comparison of 90% confidence interval (middle) and median (right) to observations, Monte Carlo simulations, and simulations with median values from the third calibration (Table III). Vertical red line represents end time of the plot with the shortest total simulation time. (a) Event A, (b) Event B, (c) Event C and (d) Event D.

As noted earlier, field experiments associated with Events C and D were performed under drier conditions when the soil contained low antecedent moisture (Table I). As a result, there were large differences in rainfall duration (runoff start time) throughout the plots (Table I). Figure 9 presents distributions of runoff start times from observations, Monte Carlo (i.e. across plots) simulations and median values determined from the third calibration (i.e. by plot). The runoff start times for Events C and D were as high as 218 min, with one plot in Event D (i.e. D403) never exhibiting runoff. On the other hand, no plots for Events A and B had runoff start times greater than 60 min, with most (85%) starting before 30 min. Although the total number of observations is relatively small (37, 36, 36 and 35 for

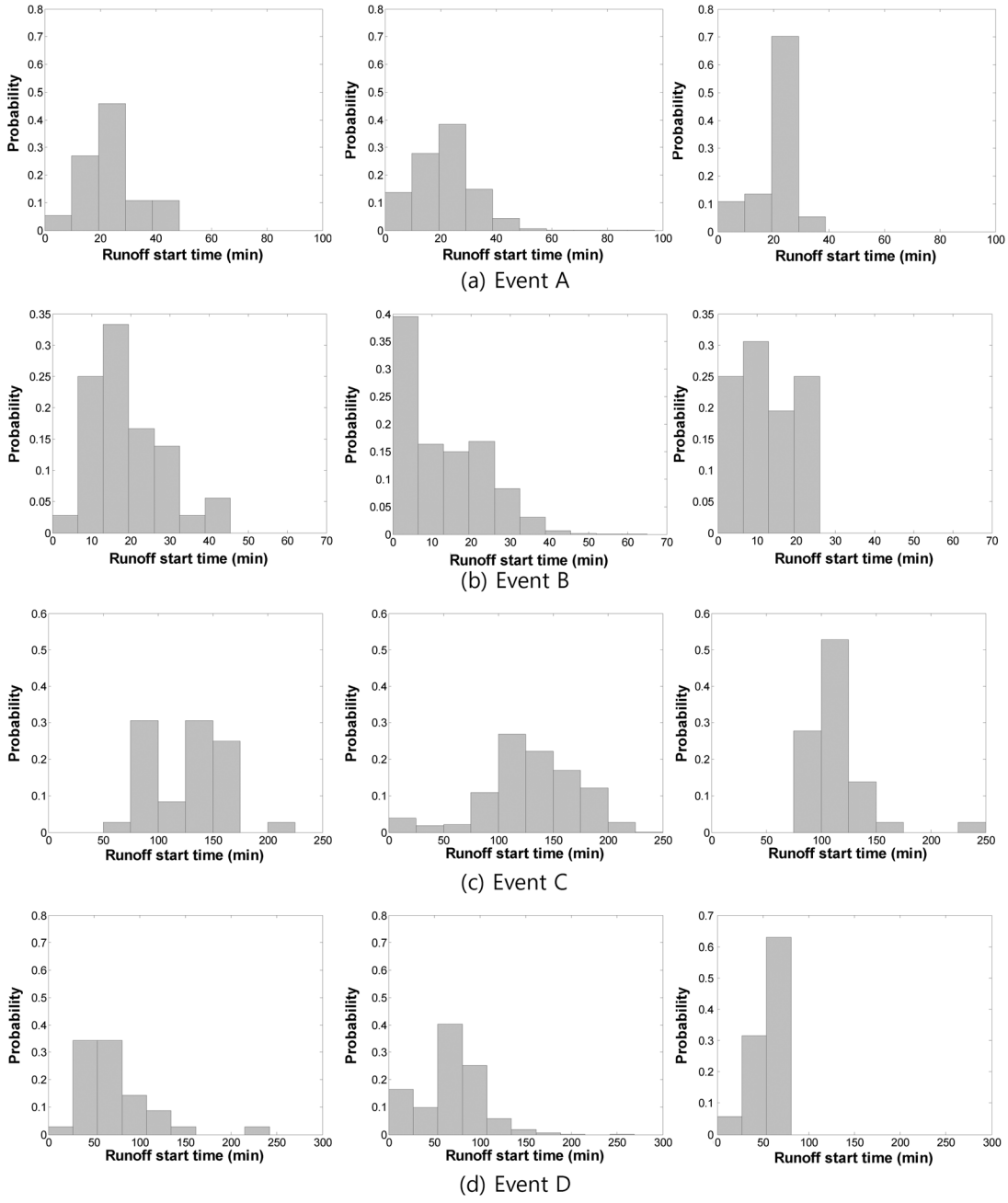

Figure 9. Distribution of runoff start times for each Event of observed data (left), Monte Carlo simulations (middle) and simulations with median values (right) from the third calibration (Table III). Statistics exclude plot with no runoff. (a) Event A, (b) Event B, (c) Event C and (d) Event D.

Events A through D, respectively), the ranges in runoff start times and overall distribution shapes and ranges in runoff start times are similar between observations and both simulations. For example, average and median start times with statistics for Events A through D, for simulations and observations, match fairly well, whereas statistics obtained using the medians are slightly underestimated (Table V). Also, peaks of the simulation distributions reasonably matched the observations, except in Event B which was largely underestimated – 36% and 37% of the average and median Event B runoff simulated start times, respectively, compared with the observations. The last column in Table V also identifies the probability of beginning the simulated runoff prior to the last observed time associated with runoff (i.e. observed runoff start time + 60 min). For example, the simulated runoff start times for Events C and D exceeded the longest observed time associated with runoff in 23 and 17% of the Monte Carlo iterations, respectively. The corresponding probabilities for Events A and B were essentially zero.

Table V. Statistics of observed and simulated runoff start time on each Event. Statistics exclude plots with no runoff

Events		Average (min)	Median (min)	Standard deviation (min)	Probability of no runoff
A	Observed	23.8	24.0	10.8	—
	Simulated (Monte Carlo)	21.3	21.0	10.6	0.0003
	Simulated (with median)	20.3	21.0	7.1	0.0000
B	Observed	19.4	17.5	9.3	—
	Simulated (Monte Carlo)	12.4	11.0	11.0	0.0000
	Simulated (with median)	12.1	11.0	8.0	0.0000
C	Observed	127.5	134.5	32.1	—
	Simulated (Monte Carlo)	129.6	130.0	43.7	0.2270
	Simulated (with median)	111.4	108.0	15.7	0.0278
D	Observed	69.7	57.0	39.6	—
	Simulated (Monte Carlo)	66.3	70.0	34.4	0.1742
	Simulated (with median)	52.6	55.0	12.8	0.0000

CONCLUSIONS

A series of monitored rainfall–runoff experiments were conducted over four seasons in 2009 and 2010 on a field containing 36 plots (0.75×2 m each), resulting in 144 runoff events. Runoff was simulated using KINEROS2/STWIR field-scale model, whose hydrodynamics are based on the kinematic wave and Parlange equations. Because of the (1) nonlinearity of the model, (2) modeling parameters that are difficult to estimate, (3) over-parameterization of the model with respect to the available data and (4) no field-scale monitoring, a combination of statistical approaches was investigated to evaluate field-scale model performance using input parameters based on plot-scale.

Of the four upscaling methods identified by Bierkens *et al*. (2000), parameter upscaling was performed and compared using two statistical approaches. The first approach averaged model outputs with a sequential three-step parameter calibration using the inverse model PEST. An over-parameterization problem occurred when calibrating by plot by Event, as there were equally good alternatives for multiple optimizations, each resulting in excellent model performance. A sequential three-step procedure improved the problem by maintaining consistency across Events by plot, addressing the over-parameterization problem by reducing the number of calibration parameters and by finding appropriate initial parameter values through the sequential steps. The second approach averaged model inputs and found representative parameters with a combination of PEST and bootstrap coupled with a Monte Carlo assessment. This approach assessed the uncertainty of the parameters through estimating CIs and determining parameter distributions.

As a result of these two statistical approaches, the median values of the more sensitive parameters, obtained by the sequential three-step procedure, were within the 95% CI obtained by bootstrapping, suggesting that median values from the PEST inverse modeling (i.e. first approach) can represent the study area. Monte Carlo simulations, using bootstrap results, showed reasonable field-scale representation of flow rates by the KINEROS2/STWIR model throughout the seasons. Although the KINEROS2/STWIR model executed admirably, this study suggests that care should be taken when using the model under dry conditions although additional studies are necessary to draw more definitive conclusions. Given the non-linear nature of the model, significant potential for a highly parameterized model and multiple datasets within plots without field-scale monitoring, our results suggest that the median value from PEST inverse modeling can be used to represent the unmonitored field-scale values, where an initial range of hard-to-define runoff parameters can be bracketed by the 95% CI.

ACKNOWLEDGEMENTS

This research was supported in part by an appointment to the Research Participation Program at the US Environmental Protection Agency, Office of Research and Development, administered by the Oak Ridge Institute for Science and Education through Interagency Agreement No. DW8992298301 between the US Department of Energy and the US Environmental Protection Agency. The Richard B. Russell Agricultural Research Center, Agricultural Research Service, US Department of Agriculture, in Athens, GA provided technical assistance on the field work. Data collection and collation services were provided by students under contract with the Student Services Contracting Authority. The US Environmental Protection Agency through its Office of Research and Development funded and managed the research described here. It has been subjected to Agency review and approved for publication.

REFERENCES

Ahrens B, Beck A. 2008. On upscaling of rain-gauge data for evaluating numerical weather forecasts. *Meteorology and Atmospheric Physics* **99**(3-4): 155–167. DOI: 10.1007/s00703-007-0261-8.

Al-Abed NA, Whiteley HR. 2002. Calibration of the Hydrological Simulation Program Fortran (HSPF) model using automatic calibration

and geographical information systems. *Hydrological Processes* **16**: 3169–3188. DOI: 10.1002/hyp.1094.

Baginska B, Milne-Home W, Cornish PS. 2003. Modelling nutrient transport in Currency Creek, NSW with AnnAGNPS and PEST. *The Modelling of Hydrologic Systems* **18**(8-9): 801–808. DOI: 10.1016/S1364-8152(03)00079-3.

Beven KJ. 2012. *Rainfall-Runoff Modelling: the Primer*, 2nd edn. Wiley-Blackwell: Hoboken, NJ.

Bierkens MFP, Finke PA, de Willigen P. 2000. *Upscaling and Downscaling Methods for Environmental Research*. Kluwer Academic Publishers: Dordrecht, The Netherlands.

Brasington J, Richards K. 1998. Interactions between model predictions, parameters and DTM scales for TOPMODEL. *Computers & Geosciences* **24**(4): 299–314. DOI: 10.1016/S0098-3004(97)00081-2.

Butler DM, Franklin DH, Cabrera ML, Tasistro AS, Xia K, West LT. 2006. Evaluating aeration techniques for decreasing phosphorus export from grasslands receiving manure. *Journal of Environmental Quality* **37**: 1279–1287. DOI: 10.2134/jeq2007.0289.

Crow WT, Ryu D, Famiglietti JS. 2005. Upscaling of field-scale soil moisture measurements using distributed land surface modeling. *Advances in Water Resources* **28**: 1–14. DOI: 10.1016/j.advwatres.2004.10.004.

Cushman JH, Bennethus LS, Hu BX. 2002. A primer on upscaling tools for porous media. *Advances in Water Resources* **25**(8-12): 1043–1067. DOI: 10.1016/S0309-1708(02)00047-7.

Deng H. 2009. Upscaling Reactive Transport Parameters for Porous and Fractured Porous Media. Ph.D Dissertation, Florida State University.

Doherty J. 2005. *PEST: Model-Independent Parameter Estimation User Manual*, 5th edn. Watermark Numerical Computing: Australia.

Durelle TS, Gooseff MN, Bencala KE, Runkel RL. 2003. Automated calibration of a stream solute transport model: implications for interpretation of biogeochemical parameters. *Journal of the North American Benthological Society* **22**(4): 492–510.

Eagleson PS. 1970. *Dynamic Hydrology*. McGraw-Hill: New York.

Efron B. 1987. Better bootstrap confidence intervals. *Journal of the American Statistical Association* **82**: 171–200.

Efron B, Tibshirani R. 1993. *An Introduction to the Bootstrap*. Chapman & Hall/CRC: Boca Raton, FL.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**(4): 783–791.

Finke PA, Bierkens MFP, de Willigen P. 2002. Choosing appropriate upscaling and downscaling methods for environmental research. *IAHS Red Book Series* **273**: 405–409.

Franklin DH, Cabrera ML, Calvert VH. 2006. Fertilizer source and soil aeration effects on runoff volume and quality. *Soil Science Society of America Journal* **70**: 84–89. DOI: 10.2136/sssaj2003.0114.

Frederic RH, Myers VA, Auciello EP. 1977. Five- to 60-Minute Precipitation Frequency for the Eastern and Central United States. NOAA Technical Memorandum NWS HYDRO-35, NOAA, National Weather Service: Silver Springs, MD.

Gallagher M, Doherty J. 2007. Parameter estimation and uncertainty analysis for a watershed model. *Environmental Modelling & Software* **22**(7): 1000–1020. DOI: 10.1016/j.envsoft.2006.06.007.

Guber AK, Yakirevich AM, Sadeghi AM, Pachepsky YA, Shelton DR. 2009. Uncertainty evaluation of coliform bacteria removal from vegetated filter strip under overland flow condition. *Journal of Environmental Quality* **38**: 1636–1644. DOI: 10.2134/jeq2008.0328.

Guber AK, Pachepsky YA, Yakirevich AM, Shelton DR, Sadeghi AM, Goodrich DC, Unkrich CL. 2010. STWIR, a Microorganism Transport with Infiltration and Runoff add-on Module for the KINEROS2 Runoff and Erosion Model Documentation and User Manual. U.S. Department of Agriculture, ARS.

Guber AK, Pachepsky YA, Yakirevich AM, Shelton DR, Sadeghi AM, Goodrich DC, Unkrich CL. 2011. Uncertainty in modelling of faecal coliform overland transport associated with manure application in Maryland. *Hydrological Processes* **25**: 2393–2404. DOI: 10.1002/hyp.8003.

Immerzeel WW, Droogers P. 2008. Calibration of a distributed hydrological model based on satellite evapotranspiration. *Journal of Hydrology* **349**(3-4): 411–424. DOI: 10.1016/j.jhydrol.2007.11.017.

Kunstmann H. 2005. Upscaling of land-surface parameters through direct moment propagation. *Advances in Geosciences* **5**: 127–131.

Lin Y-F, Anderson MP. 2003. A digital procedure for ground water recharge and discharge pattern recognition and rate estimation. *Ground Water* **41**: 306–315. DOI: 10.1111/j.1745-6584.2003.tb02599.x.

MacArthur R, DeVries JJ. 1993. *Introduction and Application of Kinematic Wave Routing Techniques using HEC-1*. Hydrologic Engineering Center, US Army Corps of Engineers: Davis, CA.

Maneta M, Schnabel S, Jetten V. 2008. Continuous spatially distributed simulation of surface and subsurface hydrological processes in a small semiarid catchment. *Hydrological Processes* **22**: 2196–2214. DOI: 10.1002/hyp.6817

Matott LS. 2005. OSTRICH: an Optimization Software Tool: Documentation and User's Guide, Version 1.6. State University of New York at Buffalo, Department of Civil, Structural and Environmental Engineering.

Meyer PD, Rockhold ML, Gee GW. 1997. Uncertainty Analyses of Infiltration and Subsurface Flow and Transport for SDMP Sites. NUREG/CR-6565, U.S.NRC: Washington, DC.

Neuman SP, Di Federico V. 2003. Multifacetad nature of hydrologic scaling and its interpretation. *Reviews of Geophysics* **41**(3): 1014. DOI: 10.1029/2003RG000130.

Pannell DJ. 1997. Sensitivity analysis of normative economic models: theoretical framework and practical strategies. *Agricultural Economics* **16**: 139–152. DOI: 10.1016/S0169-5150(96)01217-0.

Parlange J-Y, Lisle I, Braddock RD, Smith RE. 1982. The three-parameter infiltration equation. *Soil Science* **133**: 337–341.

Poeter EP, Hill MC. 1998. Documentation of UCODE: a Computer Code for Universal Inverse Modeling. U.S. Geological Survey Water-Resources Investigations Report 98-408: Denver, Colorado.

Rabideau AJ, Matott LS, Jankovic I, Craig JR, Becker MW. 2005. Influence of numerical precision on the calibration of AEM-based groundwater flow models. *Environmental Geology* **48**(1): 57–67. DOI: 10.1007/s00254-005-1258-6.

Rawls WJ, Brakensiek DL, Saxton KE. 1982. Estimation of soil water properties. *Transactions of the American Society of Agricultural Engineers* **25**: 1316–1320.

Semmens DJ, Goodrich DC, Unkrich CL, Smith RE, Woolhiser DA, Miller SN. 2008. KINEROS2 and the AGWA modeling framework. In *Hydrological Modelling in Arid and Semi-Arid Areas*, Wheater H, Sorroshian S, Sharma KD (eds). Cambridge University Press: London; 49–68.

Singh VP. 1997. *Kinematic Wave Modeling in Water Resources: Environmental Hydrology*. John Wiley & Sons: New York.

Tegnander C, Gimse T. 1998. Flow simulations to evaluate upscaling permeability. *Mathematical Geology* **30**(6): 717–731. DOI: 10.1023/A:1022347403212.

Tonkin MJ, Doherty J. 2009. Calibration-constrained Monte Carlo analysis of highly-parameterized models using subspace techniques. *Water Resources Research* **45**: W00B10. DOI: 10.1029/2007WR006678.

Varian H. 2005. Bootstrap tutorial. *The Mathematica Journal* **9**: 768–775.

Vivoni ER, Entekhabi D, Bras RL, Ivanov VY. 2007. Controls on runoff generation and scale-dependence in a distributed hydrologic model. *Hydrology and Earth System Sciences* **4**(3): 983–1029. DOI: 10.5194/hess-11-1683-2007.

Woolhiser DA, Smith RE, Goodrich DC. 1990. *KINEROS, a Kinematic Runoff and Erosion Model: Documentation and User Manual*. ARS-77. USDA, ARS: Washington, DC.

Yoo C, Lee JH, Kim K. 2007. Evaluation of the Clark unit hydrograph parameters depending on basin and meteorological condition: 2. Estimation of parameter variability. *Journal of Korea Water Resources Association* **40**(2): 171–182. (in Korean)

Zhu J, Mohanty BP. 2002. Upscaling of soil hydraulic properties for steady state evaporation and infiltration. *Water Resources Research* **38**(9): 1178. DOI: 10.1029/2001WR000704.

