HYDROLOGICAL PROCESSES
Hydrol. Process. (2010)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/hyp.7820

Delineation of hydrologically similar units in a watershed based on fuzzy classification of soil hydraulic properties

Priyabrata Santra,[1] Bhabani Sankar Das[2]* and Debashish Chakravarty[3]

[1] *Central Arid Zone Research Institute, Regional Research Station, Jaisalmer, Rajasthan 345 001, India*
[2] *Agricultural and Food Engineering Department, Indian Institute of Technology, Kharagpur, West Bengal 721302, India*
[3] *Mining Engineering Department, Indian Institute of Technology, Kharagpur, West Bengal 721302, India*

Abstract:

Evaluation of flow and transport processes in a watershed-scale requires that the watershed be divided into homogenous spatial units referred to as hydrologically similar units (HSUs). Although a few discretization schemes are already in use, a universally acceptable method of obtaining HSUs is yet to emerge. In this study, we developed a fuzzy inference system (FIS) to classify the saturated hydraulic conductivity (K_s) and two water-retention parameters α and n into fuzzy logic-based soil hydrologic classes (FSHCs). Analysis of these classes showed that soil properties within an FSHC have less variability and those between two FSHCs have large variability. This result suggested that soils belonging to a specific FSHC may be more similar than those across different FSHCs and may be grouped together to represent an HSU. Soils within a specific hydrologic class were aggregated to delineate HSUs within the watershed. For the Dengei Pahad micro-watershed (DPW), this approach showed five distinct regions representing a discretized zone having similar soil hydraulic properties. Application of this approach on a larger international database of soil hydraulic properties revealed that the developed hydrologic classes are quite comparable across different databases. The delineated HSUs based on these FSHCs were also better than the soil series map of the watershed in maintaining the soil heterogeneity of the watershed. Moreover, this new discretization scheme using the SWAT modelling environment showed better performance than the soil series-based discretization approach. Copyright © 2010 John Wiley & Sons, Ltd.

KEY WORDS fuzzy inference system; saturated hydraulic conductivity; water-retention curve; hydrologically similar unit; hydrological response unit; discretization

Received 2 August 2009; Accepted 14 June 2010

INTRODUCTION

Over the last few decades, several distributed, physically based hydrological models such as Soil and Water Assessment Tool (SWAT), Système Hydrologique Européen (SHE), Agricultural Non-Point Source (AGNPS), etc. have been developed with the capability of generating hydrologic process-wise outputs over a watershed. Most of these models differ in the way a catchment is divided into homogeneous spatial units over which transport equations representing hydrologic processes are solved (Beven, 2002). The processing of dividing a catchment into hydrologically similar units (HSUs) (Karvonen *et al.*, 1999) is often referred to as discretization (Dehotin and Braud, 2008). An HSU may be viewed as a distributed landscape unit in a watershed having unique land-use, soil type, geology, and topography. The flow and transport domains are assumed to be more uniform within an HSU than between HSUs allowing simpler mathematical treatment of hydrologic processes in a watershed. Although a few promising methods are available, discretization of a watershed into HSUs remains a challenging task requiring extensive manual and subjective intervention (MacMillan *et al.*, 2004; Francke *et al.*, 2007).

Discretization generally involves multiple and opposing considerations (Dehotin and Braud, 2008). Specifically, the choice of a discretization scheme depends on both modelling and management objectives and the availability of relevant data at the desired scale. Traditionally, HSUs are identified from field measurements and field mapping of landscape features such as soil, geology, land-use, slope, and hydrologic processes (Karvonen *et al.*, 1999; Bull *et al.*, 2003), collectively referred to as hydro-landscape features (Dehotin and Braud, 2008). Several approaches exist for discretizing watersheds based on hydro-landscape features. For example, Reggiani *et al.* (1999) used the Strahler order of topographic stream segments to divide a watershed into representative elementary watersheds. Although conservation equations for mass, momentum, and energy transport at the representative elementary watershed scale were derived in their study, parameterization of vadose zone continues to be a challenging task. Kite and Kouwen (1992) discretized a watershed into hydrotopes based on land-use characteristics. The hydrotope concept, however, ignores the natural heterogeneity of hydrological parameters (Band and Moore, 1995; Bonta, 1998). More recently, Flügel (1995) introduced the concept of hydrological response

* Correspondence to: Bhabani Sankar Das, Agricultural and Food Engineering Department, Indian Institute of Technology, Kharagpur, West Bengal 721302, India. E-mail: bsdas@iitkgp.ac.in

units (HRU) to discretize watersheds based on the slope, river network, and land-use characteristics. Similar to HSUs, hydrologic processes are assumed to be more similar within an HRU than between HRUs (Blöschl and Sivapalan, 1995). The HRU approach yields a large number of virtual spatial units (Mamillapalli *et al.*, 1996), each of which is required to be parameterized. Several researchers have delineated HRUs in geographical information system (GIS) platforms by overlaying land-use, soil, and slope data (Flügel, 1995; Karnoven *et al.*, 1999), resulting into large permutation of HRUs with the formidable task of parameterizing each of these landscape units. Despite a large number of studies, no standard set of hydro-landscape features or a universal approach exists for discretization of a watershed (Wooldridge and Kalma, 2001).

In this study, we hypothesized that soil hydraulic properties of water-retention characteristics (relationship between volumetric water content, θ, and matric potential head, h) and the saturated hydraulic conductivity (K_s) may be effectively used to discretize a watershed because of their following desirable characteristics: (1) hydraulic properties influence hydrologic processes of infiltration, percolation, runoff, and sediment loss; (2) hydraulic properties may be expressed in terms of basic soil properties, land surface, and vegetation characteristics through pedotransfer functions (PTFs); and (3) soil hydraulic properties are traditionally used for defining similar soils (Miller and Miller, 1956; Das *et al.*, 2005). The hypothesis was examined by classifying soils based on soil hydraulic properties measured in a micro-watershed.

Over the last two decades, advanced computing methods have been developed for pattern classification (Li *et al.*, 2008; Zuo *et al.*, 2008). Specifically, k-means or fuzzy rule-based classifications are efficient tools to classify domains having multiple parameters and parameter range while providing expert knowledge-based inferences about the system (Burrough *et al.*, 2000; Triantafilis *et al.*, 2003; Cifarelli *et al.*, 2007). Classifications through k-means clustering is easy and simple over a spatial scale, but have the disadvantage of using the concept of abrupt change of boundary between two classes. Such disadvantages of a linear classifier may be easily overcome using a fuzzy rule-based classification system. Theoretical aspects of fuzzy classification in soil science are described by McBratney and Odeh (1997). Bardossy and Disse (1993) suggested that the fuzzy rule-based model may be an alternative to the complex model for the description of infiltration and percolation through soil. Metternicht (2003) reported that fuzzy-based classification of salt-affected soil is better than crisp classification. Goktepe *et al.* (2005) have also reported the superior performance of fuzzy rule-based classification than k-means classification from their study on soil clustering based on the plasticity index and shear strength. The superior performance of fuzzy logic over the USLE (universal soil loss equation) approach has also been reported for delineating soil erosion hazard map in watershed scales (Mitra *et al.*, 1998; Ahamed *et al.*, 2000; Tran *et al.*, 2002). Tayfur *et al.* (2003) have shown that a simple fuzzy logic algorithm based on rainfall intensity and slope data of a watershed predicts sediment load from bare soil surface better than the physically based model. Bende-Michl (2005) successfully used fuzzy-based rules for delineating chemical hydrological response units (CHRUs) within river catchments for modelling water and nutrient fluxes. Recently, Yan *et al.* (2007) delineated site-specific management zones based on the fuzzy c-mean clustering approach for precision agriculture.

In this study, we developed a fuzzy logic-based fuzzy inference system (FIS) to classify soil hydraulic properties. Soils having similar class of hydraulic properties were then aggregated to represent an HSU, leading to the discretization of the selected micro-watershed. This newly developed discretization approach was then tested using standard statistical approaches and a distributed, physically based hydrological model.

MATERIALS AND METHODS

Description of study area and soil sampling

This study was carried out at the Dengei Pahad micro-watershed (DPW) ($\sim$42 km^2), which is a part of the Western Catchment of the Chilika Lake in Orissa, India. The DPW has a steep topography with elevations ranging from 5 to 451 m. Details of soil sampling and analytical procedures are described in Santra and Das (2008). Briefly, basic soil properties and hydraulic properties were analysed in 100 surface (0–10 cm) soil samples collected on a nested grid from the DPW (Santra, 2009). Basic soil properties included bulk density (ρ_b), soil organic carbon (OC) content, particle size distribution, and pH. Soil hydraulic properties included K_s and two van Genuchten water-retention parameters, α and n (van Genuchten, 1980). The database containing all these properties is referred to as the Chilika database, which serves as a local database. An international soil database, the UNSODA, ver. 2.0 (Leij *et al.*, 1996; Nemes *et al.*, 2001) was also used to compare the results obtained from the Chilika database. A subset ($N = 315$) of soils from the UNSODA database containing data on K_s, α, and n was extracted for this purpose. This database covers a wide range of soils and is referred to as the UNSODA-HYD database in this study.

Soil hydraulic properties

A power function relationship is commonly used to describe the θ–h relationship (van Genuchten, 1980):

$$S_e = \left(\frac{\theta - \theta_r}{\theta_s - \theta_r}\right) = \left[\frac{1}{1 + (\alpha|h|)^n}\right]^{1-\frac{1}{n}} \quad (1)$$

where S_e is the relative saturation, θ_r is the residual soil water content (cm^3 cm^{-3}), θ_s is the saturated soil water content (cm^3 cm^{-3}), and α (cm^{-1}) and n are shape parameters of the water-retention curve. The parameter

α changes with different soil types for the same value of n and vice versa. Thus, the ordered pair α and n uniquely characterizes the water-retentive capacity of soil. Similarly, the ordered pair α and n is also used for describing the unsaturated hydraulic conductivity, $K(\theta)$ (van Genuchten, 1980):

$$K(S_e) = K_s S_e^{0\cdot5}\left[1 - \left(1 - S_e^{n/(n-1)}\right)^{1-1/n}\right]^2 \qquad (2)$$

While a $\theta(h)$ function describes the soil's ability to store water, a $K(\theta)$ function describes the soil's capacity to allow water to flow through soil. Thus, the parameters of these two functions (K_s, α, and n) comprehensively describe the soil water regime and may be ideal for identifying HSUs in a watershed.

Delineation of HSUs based on fuzzy classification of soil hydraulic properties

k-Means and fuzzy c-mean clustering approaches. Before the fuzzy logics were developed to classify soil hydraulic parameters, the k-means and fuzzy c-mean clustering techniques were employed to cluster the training datasets on $\ln(K_s)$, $\ln(\alpha)$, and n both in the Chilika ($N = 100$) and the UNSODA-HYD ($N = 315$) databases. In the k-means clustering approach, clusters were derived using the Euclidean distance as the distance measure and the overall sum of the distances between objects and the respective cluster centres as the objective function J:

$$J = \sum_{j=1}^{k}\sum_{i=1}^{n}\left\|x_i^{(j)} - c_j\right\|^2 \qquad (3)$$

where i represents data points, j represents cluster, $\left\|x_i^{(j)} - c_j\right\|$ is the chosen distance measure between a data point $x_i^{(j)}$ and cluster centre c_j. The k-means clustering was repeated for three to nine clusters to identify the optimum number of clusters based on the mean silhouette values. The silhouette values range from $+1$ for points that are very distant from neighbouring clusters, through zero for points that do not distinctly belong to a single cluster, to -1 for points that are probably assigned to the wrong cluster.

Once the k-means clusters were identified, fuzzy c-mean clustering (Bezdek, 1981) was carried out to refine the cluster definitions. In the fuzzy c-mean clustering, members of a particular cluster are defined by their membership grade or membership functions (MFs) instead of a crisp boundary between clusters. Clusters found during k-means clustering are used as inputs for the fuzzy c-mean clustering and the clustering procedure is carried out by minimizing:

$$J_m = \sum_{i=1}^{n}\sum_{j=1}^{k} u_{ij}^{\,m}\left\|x_i^{(j)} - c_j\right\|^2 \qquad 1 \le m < \infty \qquad (4)$$

where m is the fuzzy exponent and a value of $m = 2$ was used, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the ith value of d-dimensional measured data, c_j is the d-dimension centre of the cluster, and $||*||$ is any measure expressing the similarity between any measured data and the cluster centre. The default Euclidean distance option in *fcm* function of Matlab 7·1 was used as a measure of similarity between data points and cluster centre. Fuzzy partitioning was carried out iteratively by updating the membership u_{ij} and the cluster centres c_j as follows:

$$u_{i_j} = \frac{1}{\displaystyle\sum_{k=1}^{c}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}} \text{ and } c_j = \frac{\displaystyle\sum_{i=1}^{n} u_{i_j} x_i}{\displaystyle\sum_{i=1}^{n} u_{i_j}^{\,m}} \qquad (5)$$

This iteration stops when $\max_{ij}\left\{\left|u_{ij}^{(k+1)} - u_{ij}^{(k)}\right|\right\}$ is less than ε, where ε is the predetermined termination criterion and k is the iteration number. This procedure converges to a local minimum or a saddle point of J_m.

Development of a FIS. FIS is the process of formulating the mapping from a given input to an output using fuzzy logic. The FIS comprises five steps: fuzzification of the input variables, application of the fuzzy operator in the antecedent, implication from the antecedent to the consequent, aggregation of the consequents across the rules, and defuzzification. The FIS was developed by defining the MFs for each soil hydraulic property. Box plots of data points belonging to a particular cluster were used as a guide to define MFs. Overlapping MFs for a given hydraulic property were merged to form a new MF as a part of supervised classification step. Thus, each hydraulic property was grouped into several classes. Fuzzy logics or rules were formulated on the basis of the associated class for each hydraulic property, leading to a particular soil cluster. For each of the fuzzy rules, three parts were defined as the antecedents corresponding to each hydraulic property and one as consequent corresponding to the soil hydrological class. Multiple parts in the antecedent of a rule were joined together through *min* fuzzy operator. The output fuzzy set of the rule was truncated to the membership degree obtained from antecedent of the rule using *min* fuzzy operator as the implication function. Both the Mamdani- (Zadeh, 1973; Mamdani, 1975) and Sugeno type (Sugeno, 1985) of FISs were investigated. The output of the FIS was treated as discrete soil hydrological classes, which was fuzzified through triangular MFs for the Mamdani type of FIS. Singleton output MFs were followed for the Sugeno type of FIS. Thus, the FIS was developed using $\ln(K_s)$, $\ln(\alpha)$, and n to derive soil hydrological classes as outputs. Output fuzzy sets of soil hydrological classes obtained from multiple rules were aggregated together through the *max* fuzzy operator, and finally the aggregated fuzzy set of soil hydraulic class was defuzzified using the centroid method.

Once the FIS was developed, it was evaluated on each pixel of spatial map of $\ln(K_s)$, $\ln(\alpha)$, and n. The output map of soil hydrological classes obtained from the FIS was converted to a raster format in ArcGIS 9·1. As we have defined the non-overlapping triangular MFs for the soil hydrological classes, the fuzzy-classified map was reclassified using the same class boundary as defined in the triangular MF for each class. For testing the accuracy of the newly developed FIS, a synthetic soil hydraulic property database was simulated in such a way that each data points lies within 0·5 times of the standard deviation from the cluster centre. The reference soil hydrological class for each synthetic data set was determined through application of a single major rule applicable for that data set. The accuracy of the FIS was tested on the basis of these synthetic data, which were considered as reference.

Evaluation of fuzzy-classified soil maps

The hydrologic classes derived from the above FIS are referred to as the fuzzy-based soil hydrological classes (FSHCs). Soils within an FSHC are considered to have similar triplets of $\ln(K_s)$, $\ln(\alpha)$, and n. To examine this similarity, both intra- and inter-class variations of selected soil and hydrological (hydro-landscape) attributes were examined. We also compared the resulting soil hydrological classes with the taxonomically classified soil series map (Scale: 1 : 250 000) for the DPW available from the National Bureau of Soil Survey and Land-Use Planning (NBSS&LUP), Nagpur (NBSS&LUP, 2005). Random locations ($N = 260$) within the watershed were generated for this purpose and corresponding class values were extracted from both the maps using spatial analysis tool of ArcGIS 9·1. The variability of hydro-landscape attributes was assessed by estimating the coefficient of variation (CV) for these locations. Additionally, we also derived similar classes for $\ln(K_s)$, $\ln(\alpha)$, and n obtained from the UNSODA-HYD database. The resulting FSHCs were compared with the FSHCs from the Chilika database. The FSHCs for the Chilika database and the UNSODA-HYD database are referred to as FSHCC and FSHCU, respectively.

Evaluation of FSHC-based HSUs

Soils within a specific FSHC may be treated to be similar in terms of producing a similar hydrologic response and may be grouped together (aggregated) for delineating HSUs. But for delineating the hydraulic property-based HSUs within a watershed, surface maps of hydraulic properties were required. The surface maps of hydraulic properties are difficult and expensive to prepare and hence are not commonly available. However, the surface map of basic soil properties, e.g. bulk density, OC content, sand content, silt content, clay content, etc. are relatively easier to prepare and, therefore, such maps are mostly available in soil survey reports. Therefore, PTFs, which have the capability to translate basic soil properties to soil hydraulic properties, were used here to generate the required maps of soil hydraulic properties of the watershed. To achieve this, surface maps of bulk density, OC content, silt content, clay content, and pH were first prepared using the regression-kriging (RK) approach. Because the digital elevation model (DEM) data was derived from the 90 m × 90 m resolution Shuttle Radar Topographic Mission (SRTM) data, we also adopted this grid size in the RK approach. A significant trend was observed in the watershed due to elevation on spatial data of basic soil properties and hence the data on elevation was used as a covariate during RK of basic soil properties. The kriged map of basic soil properties were cross validated through root mean squared residual (RMSR) and mean squared deviation ratio (MSDR) (Minasny and McBratney, 2007). The RMSR estimates the accuracy of prediction (e.g. smaller RMSR values indicate more accuracy of prediction). The MSDR measures the goodness of fit of the theoretical estimate of error (Bishop and Lark, 2008). We used the RK method of the type C (Odeh *et al*., 1995), in which the trend function is modelled using the ordinary least squares approach and ordinary kriging (OK) is performed on the residuals of the trend function. The final prediction of RK is obtained through summing the predicted values from regression analysis and residuals from kriging analysis. Prepared maps of the basic soil property were then combined through regression-based PTFs (Santra and Das, 2008) using raster calculator option of ArcGIS 9·1 to generate surface maps for $\ln(K_s)$, $\ln(\alpha)$, and n. The accuracy of the raster map of each soil hydraulic property was checked through the validation approach using 100 observed sampling locations. The surface map of the three soil hydraulic parameters was then passed through developed FIS to delineate the watershed into HSUs. The discretization approach was finally tested by examining the stream flow data for 2004–2006 using the SWAT simulation environment (Arnold *et al*., 1998; Neitsch *et al*., 2002). The GIS-enabled ArcSWAT (ver. 1.0.7) was used for data handling for flow simulations.

Input specification in the SWAT model. Two major inputs such as the DEM and the land-use classes for the SWAT were obtained from the remote-sensing data. The DEM was obtained from the SRTM image (http://srtm.csi.cgiar.org/) and projected to the Universal Transverse Mercator (UTM) with zone number 45 (Figure 1a). The SRTM data has a spatial resolution of 90 m (Jarvis *et al*., 2006). A total of 17 sub-basins were delineated in the DPW using a threshold limit for originating a stream as 109 ha in the SWAT model. Similarly, the land-use grid was prepared from the AWiFs image acquired by IRS-P6 satellite. The raw image was classified into land-use classes (Figure 1b) with the help of ground truth data using the ERDAS IMAGINE (ver. 8.0) software. The land-use classes were 42% agricultural land, 22% forested area with evergreen and deciduous trees, and 36% wetlands with natural shrubs for the DPW.

Two simulation set-ups were prepared using soil data representing (1) the FSHC-based HSUs and (2) soil series data for the DPW. Simulations corresponding to

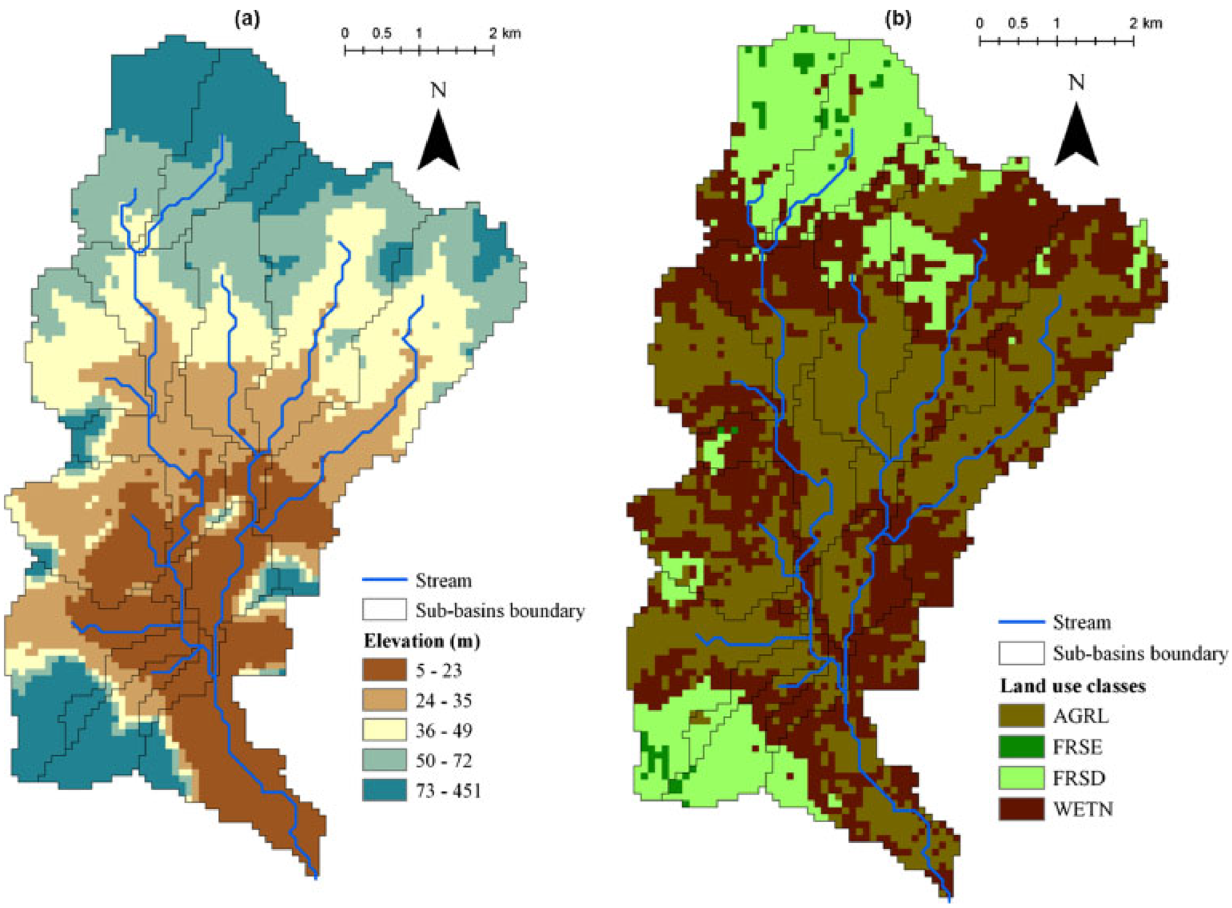

Figure 1. GIS-based grids of the watershed used in SWAT simulation: (a) digital elevation model and (b) land-use class grid of the watershed. AGRL, agricultural land; FRSE, evergreen forest; FRSD, deciduous forest; WETN, non-forested wetlands

Table I. Basic soil properties of the dominant soil series of the study area obtained from the series-level soil data of Orissa published by NBSS&LUP (2005)

Soil series	Depth (cm)	Soil properties						
		Organic carbon (%)	Sand (%)	Silt (%)	Clay (%)	Coarse fragment vol. (%)	pH	EC (dS m^{-1})
Singarazu	0–13	0·36	66·9	14·4	18·7	—	5·3	0·08
	13–28	0·28	50	23·5	26·5	—	5·6	0·08
	28–60	0·17	50·5	19·1	30·4	—	5·8	0·04
	60–90	0·20	50·1	18·2	31·7	4	5·8	0·03
	90–127	0·12	55·6	17·3	27·1	7	5·8	0·03
Tarlakota	0–9	0·26	64·4	20·7	14·9	4	5·4	0·10
	9–41	0·20	60·3	21·1	18·6	12	5·8	0·08
	41–83	0·20	58·2	19·3	22·5	18	6	0·09
	83–102	0·16	57·1	18·1	24·8	25	6·3	0·05
Jamguda	0–18	1·82	33	37·7	29·3	—	6	0·95
	18–42	1·07	27	31·3	41·7	—	6	0·47
	42–68	0·67	26·8	30	45·2	—	6·1	0·38
	68–96	0·58	43·1	22·8	34·1	—	6·1	0·35
	96–124	0·47	45·3	23·1	31·6	—	6·2	0·27
	124–152	0·33	22·2	32	45·8	—	6	0·21
Nuagarh	0–21	0·6	14·4	48·5	37·1	—	7·3	0·42
	21–48	0·54	13·6	43·5	42·9	—	7·6	0·27
	48–82	0·35	35·6	15·4	49	—	7·9	0·13
	82–105	0·36	38·7	15·2	46·1	—	8	0·95
	105–155	0·2	40·2	14·8	45	—	8·2	0·76
Bandhadwar	0–14	0·80	47·6	15·6	36·8	50	5	0·05
	14–31	0·60	42·5	12·6	44·9	65	5	0·05
	31–69	0·40	44·5	10·0	45·5	75	4·9	0·05

these two discretization approaches are referred to as SWAT-FSHC and SWAT-Series, respectively. Soil properties for different soil series are shown in Table I. Soil properties [ρ_b, OC, sand content, silt content, clay content, pH, electrical conductivity (EC), θ_{FC}, θ_{PWP}, K_s, albedo, soil erodibility] used in the SWAT-FSHC were estimated by averaging each measured soil property within a given HSU. Parameters such as θ_{FC} and θ_{PWP} were estimated using the fuzzy-classified van Genuchten (VG) parameters α and n for respective FSHCs. The albedo was estimated from the averaged spectral reflectance over 350–2500 nm wavelength, which was measured using a handheld spectroradiometer. The erodibility was estimated using the algorithm built in the SWAT modelling environment. Other input data remained the same in both the simulation set-ups. In the ArcSWAT environment, HSUs are delineated by overlaying sub-basin, soil, and land-use grids to obtain unique combinations of these three landscape attributes. These unique combinations were further reduced by using a threshold value of 10% for both soil and land-use grid to finally obtain 85 and 60 HRUs for the SWAT-FSHC and SWAT-series, respectively. Each delineated landscape unit is called the HRU in the ArcSWAT. Thus, each FSHC-based HSU or the area representing a soil series contained several HRUs. In this fashion, an HSU may be viewed as consisting of several HRUs as per the SWAT modelling approach.

Daily weather data on rainfall, maximum temperature, minimum temperature, relative humidity, and wind speed were collected for 11 years (1996–2006) from nearby weather station of Indian Meteorological Department, Bhubaneswar, Orissa. The observed solar radiation data was not available for the station and, therefore, was calculated from maximum and minimum temperature (Hargreaves and Samani, 1985). Monthly averages and other statistical weather parameters were calculated for this weather station and included in the weather generator database.

Calibration and validation of SWAT. The daily measured outflow data from the watershed were collected for 3 years from 2004 to 2006. We conducted total SWAT simulations in three steps. In the first step, sensitive parameters were identified through SWAT simulation for the year 2004–2006 using the sum of square on residuals (SSR) as the objective function. In the second step, the sensitive parameters were auto-calibrated through SWAT simulation during the year 1996–2005. In the auto-calibration process, simulations for the first 8 years (1996–2003) were performed without any measured outflow data, which stabilized the SWAT model set-up. Then SWAT simulation during the next 2 years (2004–2005) was done for the real model calibration using the available measured daily outflow data. In the third step, the calibrated SWAT model was validated with the measured outflow data for the year 2006. The modelling efficiency during calibration and validation was evaluated using the Nash–Suthcliffe efficiency (NSC) criterion (Nash and Suthcliffe, 1970):

$$\mathrm{NSC} = 1 - \frac{\sum_{i=1}^{N}(M_i - O_i)^2}{\sum_{i=1}^{N}(O_i - \overline{O}_i)^2} \tag{6}$$

where M_i is the modelled or simulated value of flow, O_i is the observed value of flow, $\overline{O}_i$ is the mean of observed values of flow, and N is the number of observations. The NSC value ranges between $-\infty$ to 1, and the higher the value, the more efficient is the calibration. A negative NSC value indicates that the mean of the observed value would have been a better predictor than the simulated values with the SWAT model. Because we intended to compare the simulation performance for different soil data sources, none of the soil-related parameters were selected for calibration although soil depth, available water capacity, and K_s were found to be sensitive for both the SWAT simulation set-ups. These three parameters were directly provided as model inputs.

RESULTS AND DISCUSSIONS

Discretization of a watershed is a key step in implementing distributed physically based hydrological models. Inasmuch as a fixed guideline or standard suite of hydro-landscape features are not available for the discretization of a watershed, we hypothesized that soil hydraulic properties of water-retention characteristics and saturated hydraulic conductivity may be treated as comprehensive hydro-landscape attributes for discretization. Hydraulic properties are known to depend on soil, terrain, and vegetation characteristics. Moreover, because hydraulic properties dominantly control the partition of rainfall into infiltration and runoff excess, the discretization of a watershed based on these properties should inherently satisfy hydrologic modelling and management goals. We discretized a micro-watershed by grouping areas having similar soil hydraulic properties using the fuzzy-based classification approach and combined the results with the distributed physically based model SWAT for modelling runoff.

Classification of soil hydraulic properties

Analysis of estimated silhouette values (Figure 2a) in the k-means clustering shows that the Chilika dataset has seven broad clusters or classes for $\ln(K_s)$, $\ln(\alpha)$, and n. Of these, clusters numbered 3, 4, and 7 are well separated with positive and distinct silhouette values (Figure 2b). Clusters with the negative silhouette values (clusters numbered 1, 2, 5, and 6) are less reliable and, hence, need to be reclassified. Although supervised classification is possible in the k-means algorithm, we used seven clusters as inputs for the fuzzy c-mean algorithm to reclassify the dataset on $\ln(K_s)$, $\ln(\alpha)$, and n. Cluster centres obtained from fuzzy c-mean clustering are shown in the three-dimensional $\ln(K_s)$, $\ln(\alpha)$ and n feature space (Figure 3)

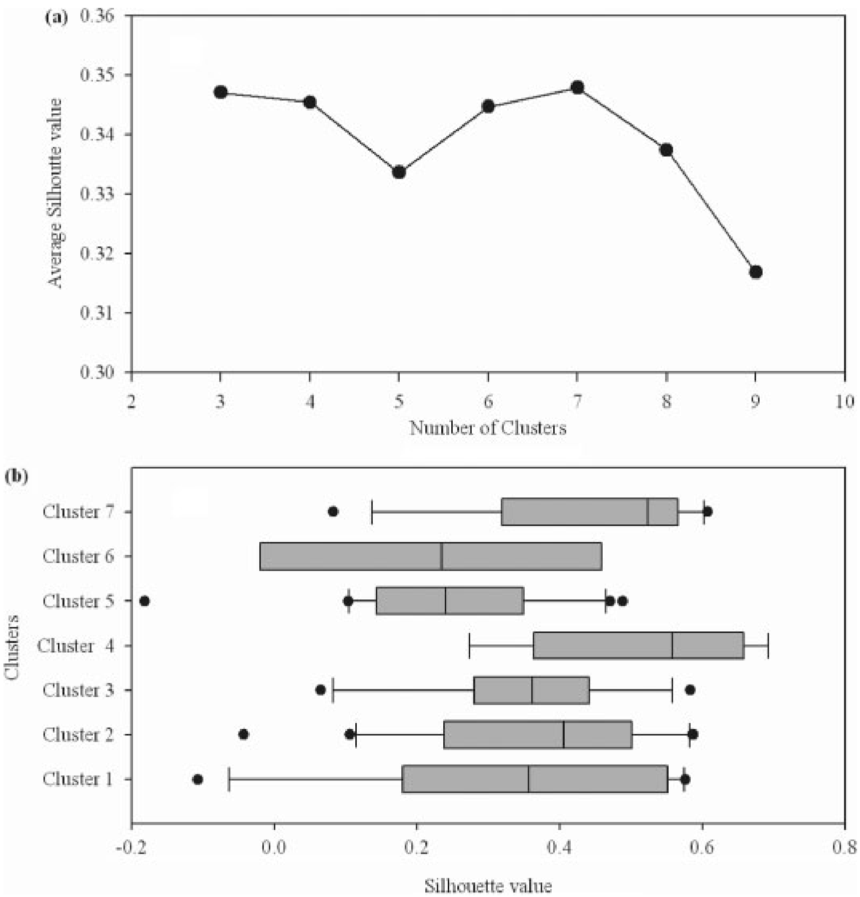

Figure 2. Average silhouette value for different number of clusters (a) and silhouette value for seven different clusters (b) for the Chilika database

and their magnitudes are listed in Table II. Cluster centres with close proximity may be merged together. For example, cluster centres 1 and 4 for $\ln(K_s)$ are similar and may be merged to a single centre. Similarly, cluster centres 2 and 3 for $\ln(K_s)$ are also merged. For $\ln(\alpha)$, cluster centres 1 and 3 are merged; cluster centres 2, 5, and 6 are merged; and cluster centres 4 and 7 are merged to single centres. Similarly, for n, cluster centres 2, 4, 5, and 7 are merged and cluster centres 3 and 6 are merged to single centres. These supervised cluster centres suggest that there exist broadly five classes of $\ln(K_s)$, three classes of $\ln(\alpha)$, and three classes of n (Table II), which yields seven distinct clusters for the triplets of $\ln(K_s)$, $\ln(\alpha)$, and n.

Five supervised centres of $\ln(K_s)$ at 0·663, 1·643, 2·774, 3·509, and 5·157 cm day^{-1} in the Chilka database may be referred to as very low, low, medium, high and very high class respectively. These five centres of $\ln(K_s)$ correspond well with very slow, slow, moderate, moderately rapid, and rapid permeability classes of O'Neal (1952) cited in soil survey reports, respectively. Three supervised centres of $\ln(\alpha)$ at $-4·137$, $-2·220$, and $-0·319$ may be referred to as low, medium, and high classes respectively. Three supervised centres of n at 1·130, 1·185, and 1·259 may be referred to as low, medium, and high class respectively. Table III shows the fuzzy-based soil clusters based on the above supervised centres of $\ln(K_s)$, $\ln(\alpha)$, and n. Hereinafter,

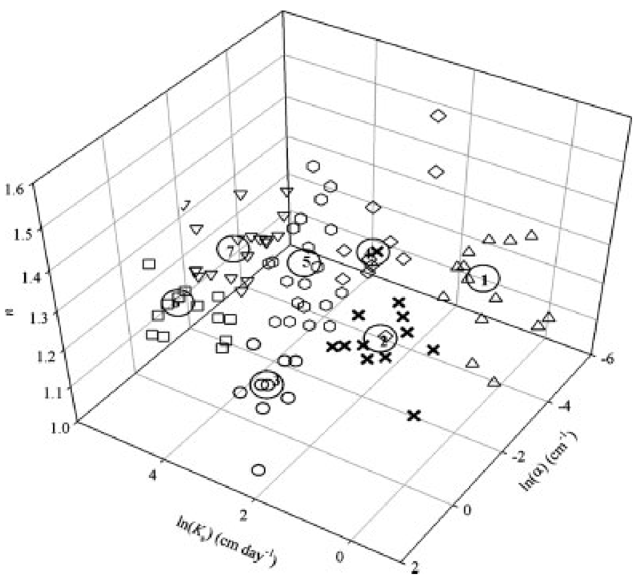

Figure 3. Observed values of hydraulic parameters in three-dimensional space with cluster centres (represented by big circles) after fuzzy c-mean clustering of the Chilika database with seven clusters. Numbers in circles are codes for the soil hydrological classes FSHCC1-FSHCC7

these seven clusters of hydraulic properties are referred to as the FSHC for the Chilika (FSHCC) database or simply classes FSHCC 1–7. The Gaussian MFs for the individual centre of each hydraulic property were developed after assuming the cluster centre as the mean

Table II. Cluster centres for each hydraulic parameter with unsupervised and supervised fuzzy c-mean clustering for the Chilika database

Soil cluster	Cluster centres with unsupervised fuzzy c-mean clustering			Cluster centres after supervision		
	$\ln(K_s)$	$\ln(\alpha)$	n	$\ln(K_s)$	$\ln(\alpha)$	n
Cluster 1	2·886	0·044	1·130	2·774	−0·319	1·130
Cluster 2	4·948	−2·217	1·253	5·157	−2·220	1·259
Cluster 3	5·226	−0·682	1·189	5·157	−0·319	1·185
Cluster 4	2·662	−3·555	1·266	2·774	−4·137	1·259
Cluster 5	3·509	−2·428	1·267	3·509	−2·220	1·259
Cluster 6	1·643	−2·014	1·180	1·643	−2·220	1·185
Cluster 7	0·663	−4·179	1·249	0·663	−4·137	1·259

$\ln(K_s)$, logarithm of saturated hydraulic conductivity (cm day^{-1}); $\ln(\alpha)$, logarithm of van Genuchten water-retention parameter, α (cm^{-1}); n, van Genuchten water-retention parameter.

Table III. Rules for classification used to develop the fuzzy inference system

Fuzzy rules	$\ln(K_s)$	$\ln(\alpha)$	n	Fuzzy-based soil hydrological class for Chilika database
Rule 1	Medium	High	Low	FSHCC 4
Rule 2	Very high	Medium	High	FSHCC 6
Rule 3	Very high	High	Medium	FSHCC 7
Rule 4	Medium	Low	High	FSHCC 3
Rule 5	High	Medium	High	FSHCC 5
Rule 6	Low	Medium	Medium	FSHCC 2
Rule 7	Very low	Low	High	FSHCC 1

$\ln(K_s)$, logarithm of saturated hydraulic conductivity (cm day^{-1}); $\ln(\alpha)$, logarithm of van Genuchten water-retention parameter α (cm^{-1}); n, van Genuchten water-retention parameter; FSHCC, Fuzzy-based soil hydrological classes.

(μ). The standard deviation (σ) for this Gaussian MF was estimated assuming that the inter-quartile range for the observed data is the same as the 3σ around mean ($\mu \pm 3\sigma$). The final MFs for each hydraulic property and soil hydrological class are presented in Figure 4. Fuzzy rules and MFs were further used to develop the FIS for fuzzy classification.

To compare the distinctness of FSHCs derived for the Chilika database, the above procedure was repeated for the UNSODA-HYD. A total of five FSHCs (FSHCU 1–5) were found for the UNSODA-HYD database. Out of these five clusters, FSHCU 1, 2, 3, and 4 matched closely with the cluster FSHCC 1, 4, 6, and 7, respectively. Ranges of soil hydraulic parameters corresponding to FSHCU 5 were not present in the Chilika database. In general, hydraulic parameter ranges for a particular class within the UNSODA-HYD database are higher than those for the Chilika database. Therefore, a single hydrological class of the UNSODA-HYD database may spread over two or three classes of the Chilika database, indicating that the local databases are more detailed but can be more fragmented compared to global databases such as the UNSODA-HYD. Hydrologic classes from the Chilika database may be more appropriate for discretizing small watersheds. Alternatively, hydrological classes derived from the global UNSODA-HYD may lose small-scale details that may be more important for representing hydrologic processes at regional scales. Testing of the FSHCC with the FIS developed in this study revealed that five FSHCCs are classified with >95% accuracy, whereas FSHCC 3 and 5 are classified with 75% accuracy. Low accuracy for FSHCC 3 and 5 may be due to overlap in MF for the medium and high classes of $\ln(K_s)$.

The fuzzy logic-based classification of soil hydraulic properties from the Chilika and UNSODA-HYD databases appears to provide a convenient way to classify soils based on $\ln(K_s)$, $\ln(\alpha)$, and n of the van Genuchten (1980) water retention and hydraulic conductivity models. The FSHCs appear to have several desirable properties: (1) fuzzy logic-based clustering and classification method has a sound statistical basis with a step-by-step methodology, (2) soil properties within an FSHC have less variability and between two FSHCs have large variability, (3) class boundaries are comparable to those observed for a relatively large international database of soil hydraulic properties, and (4) soil hydraulic properties are themselves unique parameters that control the flow processes in soil and are thereby linked to two primary hydrologic responses of runoff and sediment loss. With these desirable properties, soils belonging to a single FSHC may be grouped together to represent a HSU. This study, for the first time, shows that, in addition to saturated hydraulic conductivity, the hydraulic properties of unsaturated soil may also be rationally classified into distinct classes.

Error involved in the final raster map of soil hydraulic properties

The RMSR and MSDR values of estimated basic soil properties through RK are presented in Table IV. The errors associated with basic soil properties are less and quite comparable with the reported RMSR values by several studies (Robinson and Metternicht, 2006; Kerry and Oliver, 2007; Minasny and McBratney, 2007; Santra *et al.*, 2008). The MSDR values are close to one for each of the soil properties, which indicates less error variation. Therefore, the map of basic soil properties prepared through kriging may be considered to be reliable for its

Table IV. Root mean squared residual (RMSR) and mean squared deviation ratio (MSDR) of predicted values of basic soil properties with RK approach

Soil properties	Root mean squared residual (RMSR)	Mean squared deviation ratio (MSDR)
Organic carbon (%)	0·38	1·205
Sand (%)	15·07	0·991
Silt (%)	5·50	0·966
Clay (%)	13·43	1·009
pH	1·11	1·173

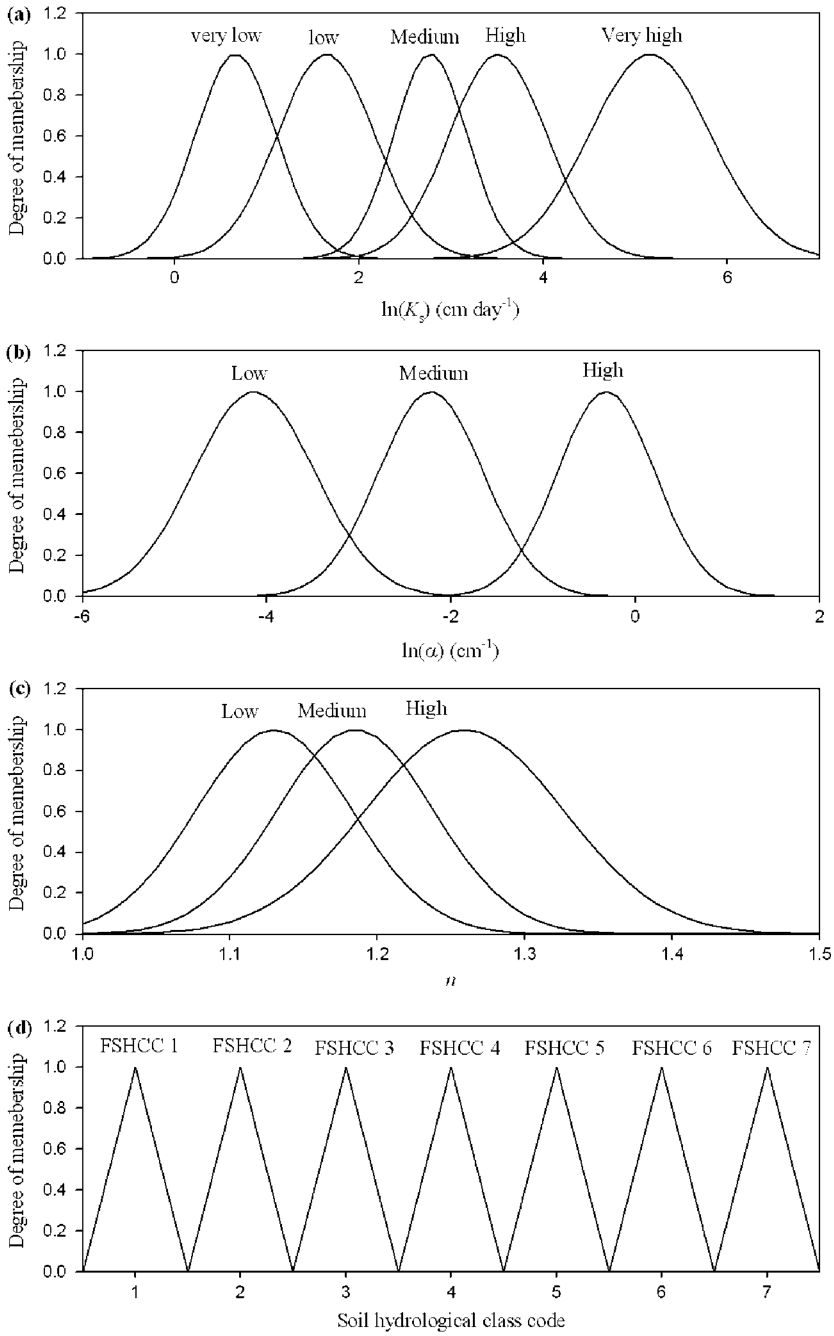

Figure 4. Membership functions for (a) logarithmically transformed saturated hydraulic conductivity, (b) logarithmically transformed van Genuchten parameter α, (c) van Genuchten parameter n, and (d) soil hydrological classes

further use. The krigged maps of basic soil properties were combined through PTFs to obtain the raster map of hydraulic properties. Therefore, the error associated with the surface map of each hydraulic property was contributed from two sources: the error from kriging and the error from the application of PTFs. We have tried to quantify the propagation of these errors in the final map of hydraulic properties (Table V). Kriging has magnified the error by an average of 20% over point estimation of hydraulic properties. To avoid such error propagation, it was also possible to prepare a soil hydraulic property map through direct kriging of their point-based measurements but resulted with more RMSR values than from the approach adopted in this study. This was due to large variability of soil hydraulic properties in two-dimensional space, and, therefore, the estimation of reliable spatial correlation structure was difficult.

Inter- and intra-class variability within HSUs

Application of FSHC-based FIS on the surface map of $\ln(K_s)$, $\ln(\alpha)$, and n for the whole DPW area resulted into five distinct HSUs. The HSUs corresponding to FSHCC 1 and FSHCC 4 were not found in the watershed. This may be due to application of the FIS on interpolated soil hydraulic property maps, whereas the developed FSHCCs were based on the measured data points ($N=$

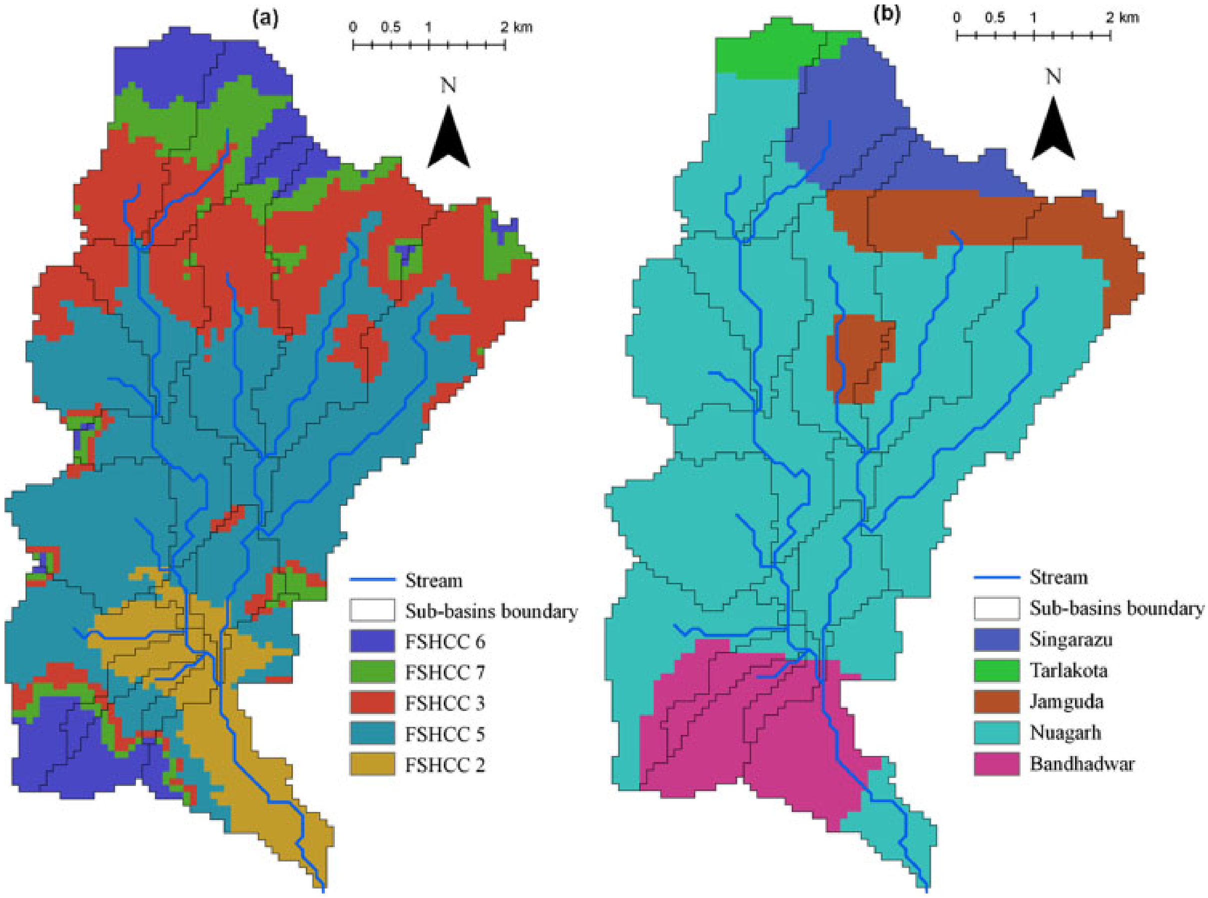

Figure 5. Soil map of the watershed showing (a) FSHCs based on soil hydraulic properties and (b) taxonomic classes based on soil series

Table V. Validation of the raster map of each hydraulic property and comparison of the resulted root mean squared residuals (RMSR) with those obtained from simple kriging of hydraulic properties and point estimation of hydraulic properties through PTFs

Soil hydraulic properties	Root mean squared residual (RMSR)		
	Raster map of hydraulic properties after kriging of basic soil properties and then application of PTFs	Raster map of hydraulic properties prepared directly through kriging of point-based measurements	Point estimate of hydraulic properties through PTFs
$\ln(K_s)$ (cm day^{-1})	1·596	1·540	1·358
$\ln(\alpha)$ (cm^{-1})	1·288	1·400	1·145
n (−)	0·079	0·076	0·062

Table VI. Variation of soil hydraulic properties both within and between the mapping units of taxonomically classified and fuzzy-based soil map of the DPW watershed

Soil properties[a]	Variation within mapping unit				Variation between mapping units			
	Taxonomically classified soil map		Fuzzy-based soil map		Taxonomically classified soil map		Fuzzy-based soil map	
	μ_{CV}[b]	σ_{CV}[c]	μ_{CV}	σ_{CV}	μ_{CV}	σ_{CV}	μ_{CV}	σ_{CV}
$\ln(K_s)$	26·42	103·14	6·19	61·86	38·14	46·33	20·79	63·75
$\ln(\alpha)$	5·48	81·16	5·59	36·32	7·94	47·02	6·26	27·53
n	1·00	86·44	1·78	38·16	2·32	58·48	4·18	21·10

[a] $\ln(K_s)$, logarithm of saturated hydraulic conductivity (cm day^{-1}); $\ln(\alpha)$, logarithm of van Genuchten water-retention parameter α (cm^{-1}); n, van Genuchten water-retention parameter.
[b] μ_{CV}, mean of coefficient of variation.
[c] σ_{CV}, standard deviation of coefficient of variation.

Table VII. Average values of soil properties for each soil hydrological class of fuzzy-classified map of the watershed

Soil properties[a]	Mean values for each soil hydrological class of fuzzy-classified map				
	FSHCC 6[b]	FSHCC 7	FSHCC 3	FSHCC 5	FSHCC 2
Soil depth (cm)	20	40	100	125	125
Number of horizons	2	2	4	5	5
ρ_p	2·88	2·74	2·65	2·54	2·38
ρ_b	1·50	1·50	1·50	1·53	1·56
OC	0·90	0·91	0·96	0·94	0·92
Sand	87·11	73·07	63·52	50·10	24·86
Silt	7·27	11·68	12·82	15·11	23·34
Clay	5·62	15·26	23·66	34·79	51·80
pH	4·5	4·5	4·7	6·2	7·1
ln(EC)	3·119	3·721	4·054	4·759	5·674
$\ln(K_s)$	5·150	4·456	3·953	3·257	2·015
$\ln(\alpha)$	−2·402	−2·571	−2·529	−2·273	−2·413
n	1·346	1·333	1·305	1·240	1·165

[a] ρ_p, particle density (Mg m^{-3}); ρ_b, bulk density (Mg m^{-3}); OC, organic carbon content (%); sand, sand content (%); silt, silt content (%); clay, clay content (%); ln(EC), logarithm of electrical conductivity (dS m^{-1}); $\ln(K_s)$, logarithm of saturated hydraulic conductivity (cm day^{-1}); $\ln(\alpha)$, logarithm of van Genuchten water-retention parameter α (cm^{-1}); n, van Genuchten water-retention parameter.
[b] FSHCC represents fuzzy-based soil hydrological class from the Chilika database.

100). Figure 5a shows distinct HSUs as dictated by the FSHCs. Each of these regions is a discretized zone in the watershed having similar soil hydraulic properties. Soil taxonomy also allows classification of soils into similar pedogenic units called soil series, which is generally used along with other landscape attributes to delineate HRUs. The soil series-based classes showed that soils in the DPW belong to five different soil series (Figure 5b). About 69·2% of the total watershed area falls under the Nuagarh soil series. In fact, all the five FSHCC classes may be seen in the Nuagarh series, although a large part (45·3%) is represented by FSHCC 5. The CV values for $\ln(K_s)$, $\ln(\alpha)$, and n within a class and between the classes calculated from 260 randomly selected locations in the DPW for both these soil maps are listed in Table VI. The mean and standard deviation of CV (μ_{CV} and σ_{CV}) reveal that the five FSHCC maps are generally more similar in $\ln(K_s)$, $\ln(\alpha)$, and n within a class and more dissimilar between the classes than those for the soil series-based map. Thus, fuzzy-classified maps appear to retain the inherent heterogeneity of the watershed while maintaining the similarity within an HSU more consistently than the taxonomically classified maps. The utility of the soil hydraulic property-based fuzzy delineation of HSUs is further examined using the SWAT model in the following section.

Runoff simulation using the newly developed discretization scheme

The fuzzy logic-based classification of soil hydraulic properties appears to provide a step-by-step approach to discretize a watershed into HSUs. Moreover, the delineated HSUs based on these FSHCs appeared to be more similar within classes than those derived from taxonomically classified maps. Thus, this new discretization approach may be a unique and robust method to delineate HSUs and simulate hydrologic response in a watershed. The modelling environment of detailed process-based hydrological model may be the best option to test this discretization approach but was unavailable during this study. We used the public-domain SWAT model for testing this new discretization approach. The SWAT model delineated 85 HRUs based on FSHCCs and 60 HRUs based on soil series maps in the DPW. Because the DPW is also classified into five distinct FSHCs (and, therefore, five HSUs), each HSU contains several HRUs. Thus, an HSU may be viewed as a superset of HRUs and each HRU within a given HSU may be assigned the same area-averaged soil properties. Basic soil properties for five FSHCCs are listed in Table VII. Similarly, the taxonomic classification of soils in the DPW also had five distinct soil series (Table I). Each series also contained several HRUs for all of which the corresponding soil series data were assigned. These two approaches of delineating HSUs were tested by creating two simulation set-ups, which were similar in all required inputs except for soil properties. Both these set-ups were first calibrated and then validated to simulate runoff from the DPW.

Calibration of the SWAT model. On the basis of the sensitivity analysis of 27 flow parameters of the SWAT model, top-ranked sensitive parameters for both SWAT-FSHC and SWAT-series set-up are listed in Table VIII. From the table, it is observed that the effective hydraulic conductivity in main channel alluvium (CH_K2), base flow alpha factor (Alfa_Bf), initial soil conservation service (SCS) runoff curve number for moisture condition II (CN2), Manning's roughness coefficient for the main channel (CH_N), surface runoff lag coefficient (Surlag), and soil evaporation compensation factor (ESCO) were most sensitive in descending order with the same rank for both the SWAT set-ups. Besides these, two groundwater parameters—groundwater delay time (GWdelay) and threshold depth of water in the shallow aquifer required

Table VIII. Sensitive parameters for simulation of daily out flow in the watershed with their calibrated values

Parameters[a]	Rank of parameters according to their sensitivity		Calibrated parameters		
	SWAT-FSHC	SWAT-series	Replacement method[b]	SWAT-FSHC[c]	SWAT-series[c]
Alpha_Bf	2	2	1	0·1178	0·1269
CH_K2	1	1	1	342·13	491·10
CH_N	4	4	1	0·2982	0·3000
CN2	3	3	3	+19·78	+22·48
ESCO	6	6	1	0·2202	0·0226
GWdelay	7	13	1	287·07	480·69
GWqmn	8	17	1	3546·9	5000
Surlag	5	5	1	3·039	18·563

[a] Alpha_Bf, base flow alpha factor (days); CH_K2, effective hydraulic conductivity in main channel alluvium (mm/hr); CH_N, manning's roughness coefficient for the main channel; CN2, initial SCS runoff curve number for moisture condition II; ESCO, soil evaporation compensation factor; GWdelay, groundwater delay time (days); GWqmn, threshold depth of water in the shallow aquifer required for return flow to occur (mm H_2O); Surlag, surface runoff lag coefficient.
[b] If replacement method is 1, then parameters were replaced by value and if it is 3, then parameters were multiplied by value (%) during calibration.
[c] SWAT-FSHC, SWAT simulation set-up with fuzzy-based soil hydrological class; SWAT-series, SWAT simulation set-up with soil series map.

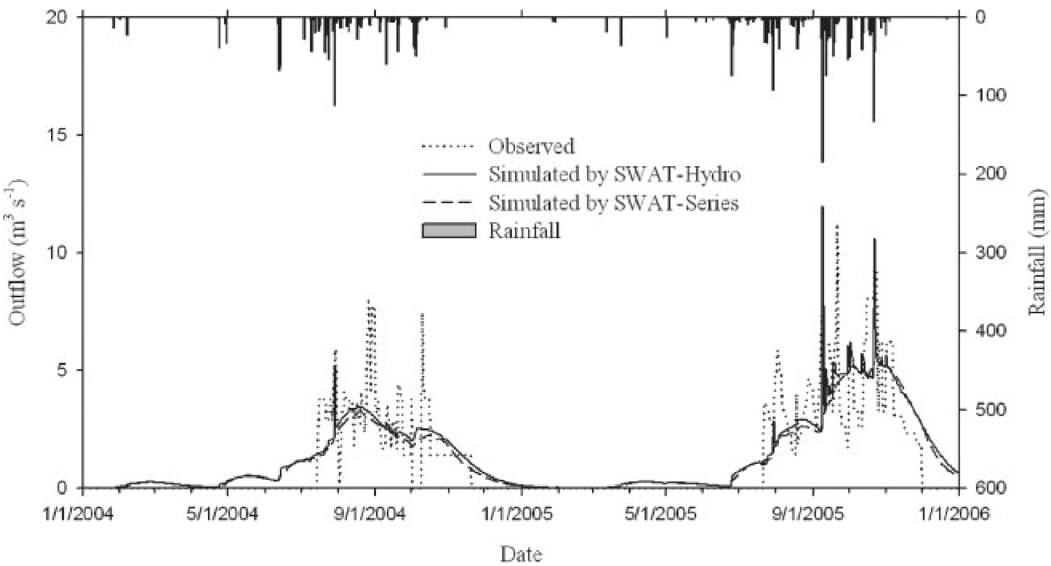

Figure 6. Daily observed and simulated outflow from the watershed during the calibration period using SWAT simulation set-up with FSHC (SWAT-FSHC) and SWAT simulation set-up with soil series (SWAT-series)

for return flow to occur (GWqmn)—were also observed to be sensitive, although their ranks were different in the two SWAT set-ups. These eight sensitive parameters were then auto-calibrated and the corresponding calibrated values for both SWAT-FSHC and SWAT-series simulation set-ups are listed in Table VIII. During the calibration process, only the CN2 value was changed by the multiplication method, which implies that initial old value for this parameter was multiplied by a number (%) and thus maintained the dissimilarity between HRUs. The remaining seven parameters were changed by replacing the old value by a new number and thus a single value was assigned in the whole watershed. Alpha_Bf values for both the SWAT set-ups ranged from 0·1 to 0·3, which indicates a slow response of groundwater flow to recharge. High values for hydraulic conductivity (CH_K2) and roughness coefficient (CH_N) for the main channel may be due to the presence of sand and gravels and hilly terrain of this watershed. The ESCO value is low for the SWAT-series simulation set-up, which indicates that most of the evaporative demand is extracted from deeper soil. The lag time to move water from the bottom of the soil profile to shallow aquifer, i.e. GW_delay is more for the SWAT-series than for SWAT-FSHC. Therefore, deeper groundwater table and low contribution of groundwater to stream flow is expected in the SWAT-series set-up. This is evident from the high GWqmn value for the SWAT-series (Table VIII). The value of Surlag is high for the SWAT-series simulation set-up, which indicates less water storage on the catchment and consequently more runoff volume for the streams. The SCS curve number was increased in both the SWAT simulation set-ups after calibration.

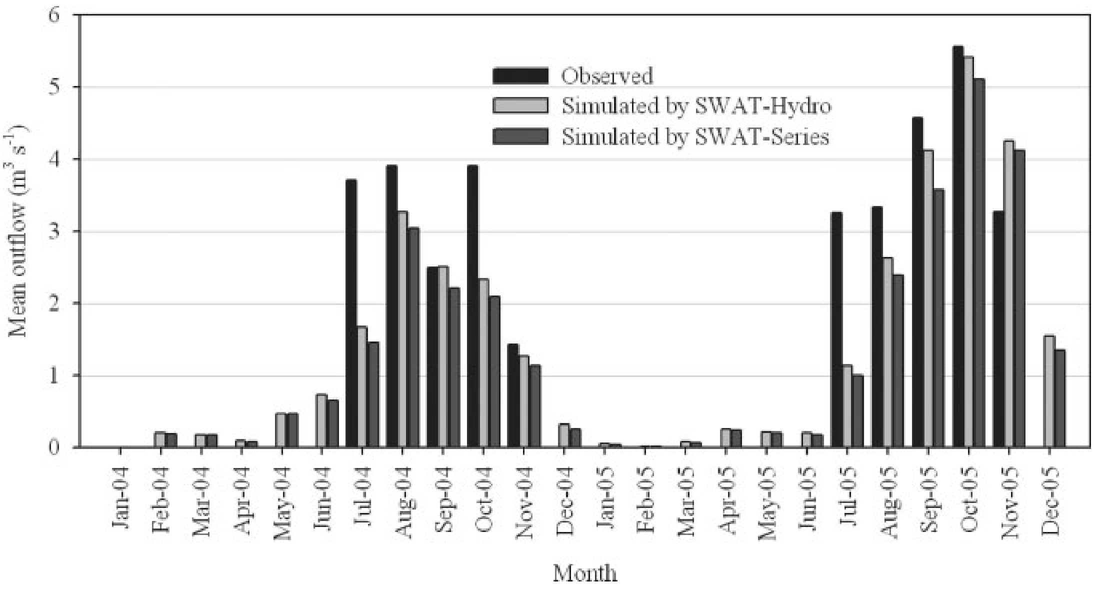

Figure 7. Monthly mean of the daily observed and simulated outflow from the watershed during the calibration period using SWAT simulation set-up with FSHC (SWAT-FSHC) and SWAT simulation set-up with soil series (SWAT-series)

The observed and simulated daily outflow (m^3 s^{-1}) from the watershed during the calibration period is shown in Figure 6. Although the simulated daily outflows for both the SWAT set-ups are close to each other, the SWAT-FSHC appears to perform slightly better in the monsoon season than the SWAT-series. The SWAT-series fails to capture the peak value of flow. In case of SWAT-FSHC, the simulated peak flow corresponding to high rainfall events matches well with the observed stream flow peaks. The monthly average of daily outflow (m^3 s^{-1}) during the calibration period for both the SWAT set-ups is given in Figure 7. The predicted value of the monthly average of the daily outflow was lower than the observed monthly average of the daily outflow in most cases in both the SWAT set-ups. The calibration of both the SWAT set-ups improved the simulation performance as seen from high NSC values in Table IX. The SWAT-FSHC had a slightly better performance with increased from $-3{\cdot}31$ before calibration to $0{\cdot}67$ after calibration; the corresponding increase for SWAT-series was from $-2{\cdot}62$ to $0{\cdot}66$. The difference in NSC values between these two simulation set-ups was more in the monsoon season (June–November). In terms of prediction of monthly average of daily outflow, the performance of SWAT-FSHC was also better than that of SWAT-series.

Validation of the SWAT model. The observed and simulated value of the daily and monthly average of daily outflow from the watershed during monsoon season in 2006 is presented in Figures 8 and 9 respectively. These figures show that the SWAT-FSHC simulated observed peaks are similar to its performance during calibration. In contrast, the SWAT-series predicted daily outflow as an average value throughout the season with none of the simulated values matching the observed peak values. Figure 9 shows that the average daily outflow rates during July and August were 6·60 and 9·44 m^3 s^{-1} respectively. Both simulation set-ups under-predicted outflow volumes for these 2 months, although slightly higher outflow volumes were observed in SWAT-FSHC than those in SWAT-series simulations.

The NSC values for the simulation of the daily outflow and monthly average of daily outflow from the watershed (Table IX) show that SWAT-FSHC simulation results are slightly better than those of SWAT-series. Although similar observations were observed during the calibration period, the difference in the NSC value between these two simulation set-ups was more for the validation period, indicating the slightly superior performance of SWAT-FSHC compared to the performance of SWAT-series in simulating stream flow outside the calibration period. In general, the NSC value is lower for the validation period than for the calibration period for both set-ups. Both of these two SWAT set-ups were calibrated in a period with low annual rainfall and, therefore, when validated in a period with high rainfall, the simulated stream flow did not match all the observed peak flows. A comparison between the observed and simulated runoff and the corresponding NSC value suggested that SWAT-FSHC performed slightly better than SWAT-series in simulating outflow from the watershed. The performance of SWAT-FSHC may perhaps be improved through different ways. For instance, the inclusion of management practices for land-use classes in simulations may improve the performance. We have also used the rainfall data from the weather station, which is not located within the watershed. The inclusion of local rainfall data of the watershed may further improve the performance of the FSHCC scheme. Finally, direct inclusion of soil hydraulic properties in a more detailed physically based model may fully demonstrate the utility of the current HSU-delineation approach.

Table IX. Nash–Sutcliffe coefficient for simulating the daily outflow and monthly mean of daily outflow from the watershed

Time period	Season	Daily outflow		Monthly mean of daily outflow	
		SWAT-FSHC[a]	SWAT-series[a]	SWAT-FSHC	SWAT-series
Calibration					
2004	Total year	0·62	0·60	0·76	0·69
	Monsoon	0·45	0·42	0·41	0·24
2005	Total year	0·68	0·67	0·83	0·81
	Monsoon	0·54	0·50	0·65	0·55
2004–2005 (pooled)	Total year	0·67	0·66	0·80	0·76
	Monsoon	0·53	0·51	0·57	0·45
Validation					
2006	Total year	0·45	0·42	0·92	0·88
	Monsoon	0·30	0·25	0·86	0·76

[a] SWAT-FSHC, SWAT simulation set-up with fuzzy-based soil hydrological class; SWAT-series, SWAT simulation set-up with soil series map.

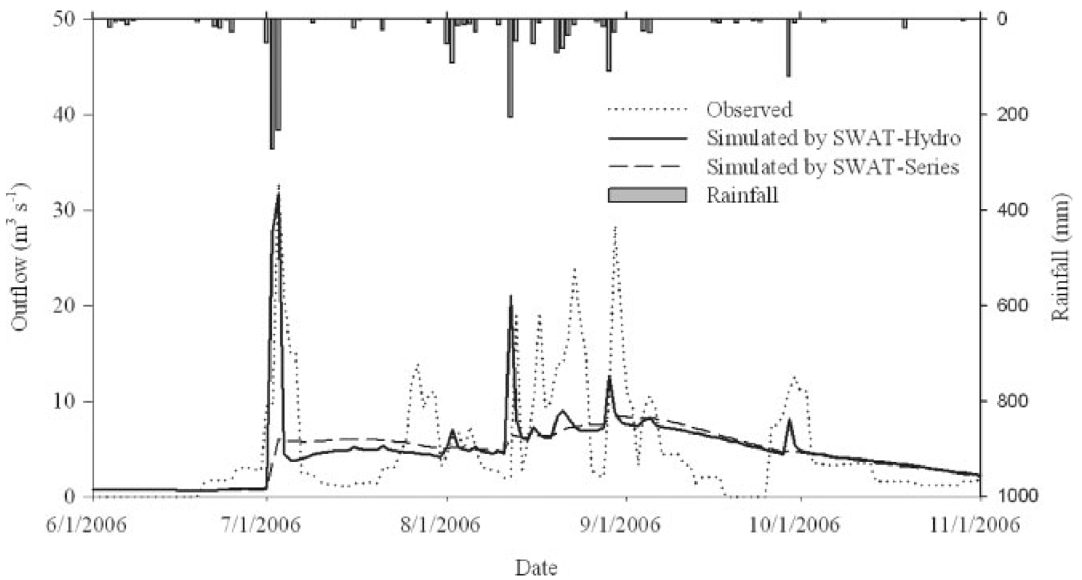

Figure 8. The daily observed and simulated outflow from the watershed during the validation period using SWAT simulation set-up with FSHC (SWAT-FSHC) and SWAT simulation set-up with soil series (SWAT-series)

CONCLUSIONS

In this study, we have shown that soil hydraulic properties of water-retention characteristics and saturated hydraulic conductivity may be treated as comprehensive hydro-landscape attributes for discretization. We developed a FIS to classify K_s and two water-retention parameters α and n into distinct soil hydrologic classes. Analysis of these classes showed that soil properties within an FSHC have less variability and between two FSHCs have large variability. This result suggested that soils belonging to a specific FSHC may be more similar than those across different FSHCs and may be grouped together to represent a HSU. Soils within a specific hydrologic class were aggregated to delineate HSUs within the watershed. For the DPW, this approach showed five distinct regions representing a discretized zone having similar soil hydraulic properties. Application of this approach on a larger international database of soil hydraulic properties revealed that the developed hydrologic classes are quite comparable across different databases. The delineated HSUs based on these FSHCs were also better than the soil series map of the watershed in maintaining the soil heterogeneity of the watershed. Soil hydraulic properties within a mapping unit were also more similar for the FSHC-based HSU map than for the soil series map. Thus, the fuzzy logic-based classification of soil hydraulic properties appears to provide a step-by-step way to discretize a watershed into HSUs. In the present study, we have used only surface soil properties to delineate the FSHC-based HSUs due to unavailability of soil data for deeper layers, inclusion of which may result in more accurate delineation of HSUs within the watershed. This is because this layering of soil will result in different hydrologic responses of the total soil profile than those of surface soil. However, the hydrological behaviour of surface soil may largely influences the hydrological response of the total soil profile. Therefore, the present approach shows the possibility of classifying landscapes into similar zones based on soil hydraulic properties, which may be further improved.

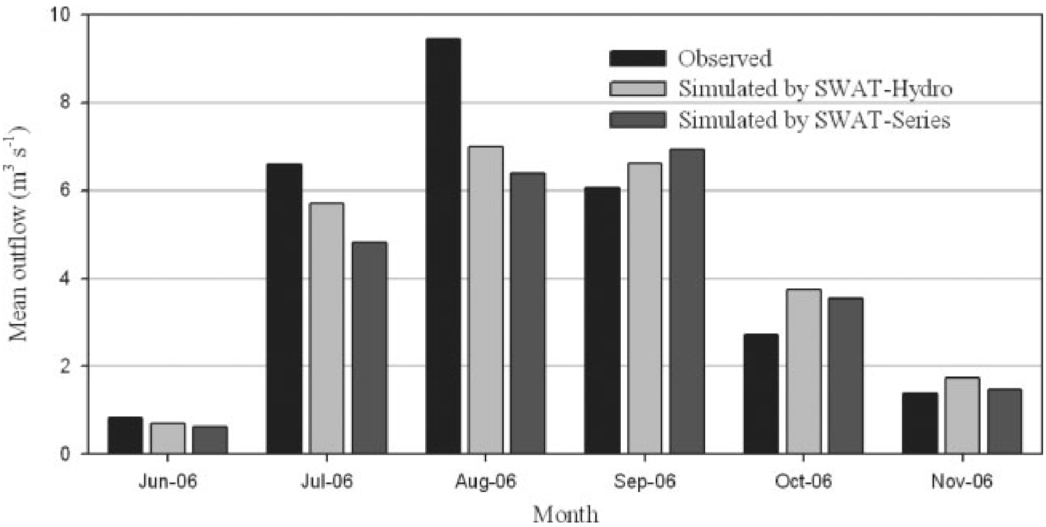

Figure 9. Monthly mean of the daily observed and simulated stream outflow from the watershed during the validation period using SWAT simulation set-up with FSHC (SWAT-FSHC) and SWAT simulation set-up with soil series (SWAT-series)

We examined the utility of this new discretization scheme for describing the hydrologic response using the process-based SWAT modelling environment. The SWAT-based simulation results suggested that the newly developed discretization scheme provided better performance than the soil series-based discretization approach. Therefore, the FSHC approach may be used to delineate HSUs in a watershed. In particular, for a small-scale watershed or an agricultural watershed, where there is large possibility that the total watershed is covered by a single mapping unit, the HSUs map will be more useful in simulating hydrological response. The proposed discretization scheme is based on soil hydraulic parameters, which dominantly control hydrological processes and, therefore, may offer a sound method for discretization of a watershed for use with a process-based hydrological model.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support provided by the Chilika Development Authority, Bhubaneswar, India, for conducting a part of the study and for providing the secondary data on runoff. The senior author also acknowledges the Central Arid Zone Research Institute, Jodhpur, for granting permission for study leave at IIT, Kharagpur.

REFERENCES

Ahamed TRN, Gopal Rao K, Murty JSR. 2000. Fuzzy class membership approach to soil erosion modeling. *Agricultural Systems* **63**: 97–110.

Arnold JG, Srinivasan R, Muttiah RS, Williams JR. 1998. Large area hydrologic modeling and assessment part I: model development. *Journal of the American Water Resources Association* **34**(1): 73–89.

Band LE, Moore ID. 1995. Scale: landscape attributes and geographical information systems. In *Scale Issues in Hydrological Modelling*, Kalma JD, Sivapalan M (eds). Wiley: Chichester, UK; 159–180.

Bardossy A, Disse M. 1993. Fuzzy rule-based models for infiltration. *Water Resources Research* **29**(2): 373–382.

Bende-Michl U. 2005. Development of a rule-based fuzzy set classification model to determine process-based chemical hydrological response units (CHRUs) within the ArcView GIS ®. In *MODSIM 2005 International Congress on Modelling and Simulation*, Zerger A, Argent RM (eds). December. Modelling and Simulation Society of Australia and New Zealand: University of Melbourne; 1346–1353. ISBN: 0-9758400-2-9.

Beven K. 2002. Towards an alternative blueprint for a physically-based digitally simulated hydrological modelling system. *Hydrological Processes* **16**: 189–206.

Bezdek JC. 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press: New York.

Bishop TFA, Lark RM. 2008. Reply to "standardized vs. customary ordinary cokriging..." by A. Papritz. *Geoderma* **146**(1–2): 397–399.

Blöschl G, Sivapalan M. 1995. Scale issues in hydrological modeling. *Hydrological Processes* **9**: 251–290.

Bonta JV. 1998. Spatial variability of runoff and soil properties on small watersheds in similar soil-map units. *Transactions of the ASAE* **41**: 575–858.

Bull LJ, Kirkby MJ, Shannon J, Dunsford H. 2003. Predicting hydrological similar surfaces (HYSS) in semi-arid environments. *Advances in Environmental Monitoring and Modelling* **1**(2): 1–26.

Burrough PA, van Gaans PFM, MacMillan RA. 2000. High-resolution landform classification using fuzzy k-means. *Fuzzy Sets and Systems* **113**: 37–52.

Cifarelli C, Nieddu L, Seref O, Pardalos PM. 2007. K-T.R.A.C.E: a kernel *k*-means procedure for classification. *Computers and Operations Research* **34**: 3154–3161.

Das BS, Haws NW, Rao PSC. 2005. Defining geometric similarity in soils. *Vadose Zone Journal* **4**: 264–270.

Dehotin J, Braud I. 2008. Which spatial discretization for distributed hydrological models? Proposition of a methodology and illustration from medium to large-scale catchments. *Hydrology and Earth System Sciences* **12**: 769–796.

Flügel WA. 1995. Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Bröl, Germany. *Hydrological Processes* **9**: 423–436.

Francke T, Güntner A, Mamede G, Müller EN, Bronstert A. 2007. Automated catena-based discretization of landscapes for the derivation of hydrological modelling units. *International Journal of Geographical Information Science* **22**(2): 111–132.

Goktepe AB, Altun S, Sezer A. 2005. Soil clustering by fuzzy c-mean algorithm. *Advanced Engineering Software* **36**: 691–698.

Hargreaves GH, Samani Z. 1985. A reference crop evapotranspiration from temperature. *Transactions of the ASAE* **1**(2): 96–99.

Jarvis A, Reuter HI, Nelson A, Guevara E. 2006. *Hole-filled Seamless SRTM Data V3*, International Centre for Tropical Agriculture (CIAT): Colombia USA.

Karnoven T, Koivusalo H, Jauhianinen M, Palko J, Weppling K. 1999. A hydrological model for predicting runoff from different land-use areas. *Journal of Hydrology* **217**: 253–265.

Kerry R, Oliver MA. 2007. Comparing sampling needs for semivariograms of soil properties computed by the method of moments and residual maximum likelihood. *Geoderma* **140**: 383–396.

Kite GW, Kouwen N. 1992. Watershed modelling using land classification. *Water Resources Research* **28**: 3193–3200.

Leij FJ, Alves WJ, van Genuchten MTh, Williams JR. 1996. *Unsaturated soil hydraulic database, THE UNSODA 1·0 User's Manual*. US Environmental Protection Agency: Ada, Oklahoma, Report EPA/600/R-96/095.

Li L, Wang J, Cao Z, Zhong E. 2008. An information-fusion method to identify pattern of spatial heterogeneity for improving the accuracy of estimation. *Stochastic Environmental Research and Risk Assessment* **22**: 689–704.

MacMillan RA, Jones RK, McNabb DH. 2004. Defining a hierarchy of spatial entities for environmental analysis and modeling using digital elevation models (DEMs). *Computer, Environment and Urban Systems* **28**(3): 175–200.

MacQueen JB. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press: Berkeley; 281–297.

Mamdani EH, Assilian S. 1975. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies* **7**(1): 1–13.

Mamillapalli S, Srinivasan R, Arnold JG, Engel BA. 1996. Effect of spatial variability on basin scale modeling. In *Proceedings of the 3rd International Conference Integrating GIS and Environmental Modeling*, Santa Fe: New Mexico; USA. Maine, USA: NCGIA (National Center for Geographic Information and Analysis). 10–15.

McBratney AB, Odeh IOA. 1997. Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements, and fuzzy decisions. *Geoderma* **77**: 85–113.

Metternicht GI. 2003. Categorical fuzziness: a comparison between crisp and fuzzy class boundary modeling for mapping salt-affected soils using Landsat TM data and a classification based on anion ratios. *Ecological Modelling* **168**: 371–389.

Miller EE, Miller RD. 1956. Physical theory for capillary flow phenomena. *Journal of Applied Physics* **27**: 324–332.

Minasny B, McBratney AB. 2007. Spatial prediction of soil properties using EBLUP with the Matérn covariance function. *Geoderma* **140**: 324–336.

Mitra B, Scott HD, Dixon JC, McKimmey JM. 1998. Application of fuzzy logic to the prediction of soil erosion in a large watershed. *Geoderma* **86**: 183–209.

Nash JE, Suthcliffe JE. 1970. River flow forecasting through conceptual model. *Journal of Hydrology* **10**: 282–290.

NBSS&LUP. 2005. *Soil Series of Orissa*, National Bureau of Soil Survey and Land-use Planning: Nagpur, India.

Neitsch SL, Arnold JG, Kiniry JR, Srinivasan R, Williams JR. 2002. *Soil and water assessment tool (SWAT). SWAT user manual, version 2000*. Water Resources Institute: College Station, Texas, TWRI Report TR-192.

Nemes A, Schaap MG, Leij FJ, Wösten JHM. 2001. Description of the unsaturated soil hydraulic database THE UNSODA version 2.0. *Journal of Hydrology* **251**: 151–162.

Odeh IOA, McBratney AB, Chittleborough DJ. 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. *Geoderma* **63**: 197–214.

O'Neil AM. 1952. A key for evaluating soil permeability by means of certain field clues. *Soil Science Society of America Proceedings* **16**: 312–315.

Reggiani P, Hassanizadeh SM, Sivapalan M, Gray WG. 1999. A unifying framework for watershed thermodynamics: constitutive relationship. *Advanced Water Resources Research* **22**(4): 367–398.

Robinson TP, Metternicht G. 2006. Testing the performance of spatial interpolation techniques for mapping soil properties. *Computers and Electronics in Agriculture* **50**: 97–108.

Santra P. 2009. Delineation and characterization of hydrologically-similar zones based on spatial distribution of soil hydraulic properties. PhD thesis, Indian Institute of Technology: Kharagpur, India.

Santra P, Das BS. 2008. Pedotransfer functions for soil hydraulic properties developed from a hilly watershed of Eastern India. *Geoderma* **146**: 439–448.

Santra P, Chopra UK, Chakraborty D. 2008. Spatial variability of soil properties and its application in predicting surface map of hydraulic properties in an agricultural faro. *Current Science* **95**(7): 937–945.

Sugeno M. 1985. *Industrial Applications of Fuzzy Control*, Elsevier Science Inc.: New York, NY, USA; 269.

Tayfur G, Ozdemir S, Singh VP. 2003. Fuzzy logic algorithm for runoff-induced sediment transport from bare soil surfaces. *Advanced Water Research* **26**: 1249–1256.

Tran LT, Ridgley MA, Duckstein L, Sutherland R. 2002. Application of fuzzy logic- based modeling to improve the performance of Revised Universal Soil Loss Equation. *Catena* **47**: 203–226.

Triantafilis J, Odeh IOA, Minasny B, McBratney AB. 2003. Elucidation of physiographic and hydrogeological features of the lower Namoi valley using fuzzy k-means classification of EM34 data. *Environmental Modelling and Software* **18**: 667–680.

van Genuchten MTh. 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal* **44**: 892–898.

Wooldridge SA, Kalma JD. 2001. Regional-scale hydrological modelling using multiple-parameter landscape zones and a quasi-distributed water balance model. *Hydrology and Earth System Sciences* **5**(1): 59–74.

Yan Li, Zhou Shi, Feng Li, Hong-Yi Li. 2007. Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. *Computers and Electronics in Agriculture* **56**: 174–186.

Zadeh LA. 1973. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man, and Cybernetics* **3**(1): 28–44.

Zuo W, Zhang D, Wang K. 2008. On kernel difference-weighted *k*-nearest neighbor classification. *Pattern Analysis and Applications* **11**: 247–257.

