

Improving the spatial representation of soil properties and hydrology using topographically derived initialization processes in the SWAT model

Daniel R. Fuka,¹ Amy S. Collick,^{2†} Peter J.A. Kleinman,² Daniel A. Auerbach,^{3*‡}
R. Daren Harmel⁴ and Zachary M. Easton¹

¹ Department of Biological Systems Engineering, Virginia Tech, Blacksburg, VA, 24061, USA

² Pasture Systems and Watershed Management Research Unit, USDA-ARS, University Park, PA, USA

³ Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, 14853, USA

⁴ Grassland Soil and Water Research Laboratory, USDA-ARS, Temple, TX, 76502, USA

Abstract:

Topography exerts critical controls on many hydrologic, geomorphologic and biophysical processes. However, many watershed modelling systems use topographic data only to define basin boundaries and stream channels, neglecting opportunities to account for topographic controls on processes such as soil genesis, soil moisture distributions and hydrological response. Here, we demonstrate a method that uses topographic data to adjust spatial soil morphologic and hydrologic attributes: texture, depth to the C-horizon, saturated conductivity, bulk density, porosity and the water capacities at field (33 kpa) and wilting point (1500 kpa) tensions. As a proof of concept and initial performance test, the values of the topographically adjusted soil parameters and those from the Soil Survey Geographic Database (SSURGO; available at 1 : 20 000 scale) were compared with measured soil pedon pit data in the Grasslands Soil and Water Research Lab watershed in Riesel, TX. The topographically adjusted soils were better correlated with the pit measurements than were the SSURGO values. We then incorporated the topographically adjusted soils into an initialization of the Soil and Water Assessment Tool model for 15 Riesel research watersheds to investigate how changes in soil properties influence modelled hydrological responses at the field scale. The results showed that the topographically adjusted soils produced better runoff predictions in 50% of the fields, with the SSURGO soils performing better in the remainder. In addition, the *a priori* adjusted soils result in fewer calibrated model parameters. These results indicate that adjusting soil properties based on topography can result in more accurate soil characterization and, in some cases, improve model performance. Copyright © 2016 John Wiley & Sons, Ltd.

KEY WORDS SWAT model; infiltration excess runoff; soil characteristics; topography; surface characterization; Hortonian runoff; non-point source

Received 4 June 2015; Accepted 21 April 2016

INTRODUCTION

Researchers, planners and government agencies often rely on watershed models to provide insight into the consequences of land management on nutrient and sediment loss to water bodies (Gassman *et al.*, 2007; Van Liew *et al.*, 2012; Bosch *et al.*, 2013; Crossman *et al.*, 2013; Niraula *et al.*, 2013). In many of these models [e.g. Soil and Water Assessment Tool (SWAT), Generalized Watershed Loading Function, Hydrological Simulation Program—Fortran],

the representation of runoff generation and other hydrologic processes is highly dependent upon the soil and topographic data that are used to initialize the model. Therefore, accurate characterization of the spatial variation in soils and topography is essential to meaningfully initialize watershed models, and this parameterization exhibits a strong influence on the quality of information provided by the model.

Topography is one of the key controls of soil development (pedogenesis), and the catena, or sequence of soils along a topographic gradient, provides fundamental insight into how water moves in landscapes (Jenny, 1941; Foth, 1943; Mitsch and Gosselink, 2000; Lin, 2012). Watershed models that directly address the interaction of topography and soils, such as TOPMODEL (Beven *et al.*, 1995), Distributed Hydrology Soil Vegetation Model (Wigmosta *et al.*, 1994) and the Soil Moisture Routing model (Zollweg

*Correspondence to: Dan Auerbach, Department of Ecology and Evolutionary Biology, Cornell University, Ithaca NY 14853, USA.
E-mail: auerbach.dan@gmail.com

† Current Address: Department of Agriculture, Food, and Resource Sciences, University of Maryland Eastern Shore, Princess Anne, MD 21853, USA

‡ Current Address: Office of Wetlands, Oceans & Watersheds, USEPA, Washington, DC 20460, USA

et al., 1996; Frankenberger *et al.*, 1999; Easton *et al.*, 2007a, b) have performed well in landscapes dominated by variable source area hydrology where saturation excess runoff occurs (Dunne and Black, 1970). However, watershed models that are structured around infiltration excess or Hortonian flow (runoff generation as a function of soil infiltration capacity rather than saturation state) generally incorporate topographic features only to define average slopes and to delineate the basin area that contributes runoff to a reach. Nevertheless, soil characteristics are critically important to Hortonian flow-dominated systems (Horton, 1933), and field-scale models can benefit from extending the influence of topography on these soil characteristics. Adapting watershed models, such as the SWAT, to better represent spatially explicit variation in soil properties as a function of topographic position promises to improve predictions of where and when runoff occurs, thereby improving a model's ability to inform landscape management decisions.

Unfortunately, direct measurement of many important soil properties is time-consuming and difficult (Hadzick *et al.*, 2011; McBratney *et al.*, 2006; Schaap *et al.*, 2001; Vereecken *et al.*, 2010; Wösten *et al.*, 1995, Wösten *et al.*, 2001), but pedotransfer function (PTF) models can relate easily obtained *in situ* measurements of soil hydraulic properties such as bulk density, porosity or available water content to more difficult to characterize properties, such as saturated hydraulic conductivity and infiltration rate (Tseng and Jury, 1993; Levi *et al.*, 2015). At least since Jenny (1941) suggested that parent material, climate, biota, relief and time were factors controlling soil formation, PTFs have offered a means to extend limited measurements of soil properties to other physical locations according to indices of similarity (Bouma, 1989; McBratney *et al.*, 2003; Levi *et al.*, 2015). This spatial extrapolation has been accomplished through various methods including elevation-derived topographic indices, surface reflectance obtained from remote sensing platforms or some combination of these and other data that quantify energy inputs to the system (Moore *et al.*, 1993; Hengl *et al.*, 2007; Nield *et al.*, 2007), especially as digital elevation models have increased in spatial resolution through time. The spatial transfer of hydrologic properties through the landscape is often accomplished through statistical methods ranging from simple regression to more complex multivariate regression kriging methods and to a lesser extent deterministic methods, utilizing the knowledge of soil scientists to develop models of soil prediction (Minasny *et al.*, 2008; Saadat *et al.*, 2008; Levi *et al.*, 2015). While advances in PTF research allow estimates of such properties as soil water characteristics across spatial scales (Smettem *et al.*, 2004; Sharma *et al.*, 2006; Jana and Mohanty, 2011), the applications of digital soil mapping tools for vadose zone modelling are underutilized (Minasny *et al.*, 2013) and the development of spatial

PTFs has been minimal (McBratney *et al.*, 2003). Thus, incorporating such approaches into watershed and field-scale modelling initializations offers unique opportunities to improve soil characterization and model performance (Collick *et al.*, 2014).

Such opportunities include improving the simulation of runoff generation and shallow transient water tables, which can be simulated through techniques that infer soil properties from topographic gradients (Easton *et al.*, 2008, 2011). For example, Collick *et al.* (2014) recently demonstrated an approach to distribute soil properties such as texture and depth via a PTF. This work combined the spatial distribution of hydrological characteristics of Moore *et al.* (1993) with the hydrological parameter relationships of Saxton and Rawls (2006). They showed that terrain metrics correlated well with particle size distribution, organic matter and depth, as well as with soil hydraulic properties including hydraulic conductivity (Ksat) and available water content (AWC). These relationships are significant because available soil databases, such as the Soil Survey Geographic Database (SSURGO), are subject to various uncertainties including the boundaries of landscape units (i.e. the soil-type polygons), the extrapolation of pedon point observations to soil-type polygons and the variable intensity of sampling. Despite the critical need for soil surveys and the importance of the resulting maps to natural resource management (Simonson, 1991), available soil data often fail to provide relevant hydrologic information for increasingly detailed studies, often conducted at field scales but over large spatial extents (Petersen, 1991; Moore *et al.*, 1993).

To address this challenge, we present a new approach to distribute soil properties according to topography within the SWAT model. The method seamlessly integrates the calculation of a topographic index (TI) into SWAT initialization. The TI is then used to distribute soil depth, texture, saturated conductivity and AWC that influence field-level hydrology and thereby nutrient and sediment loss. To evaluate the approach, we compared calculated TI values to pedon data collected from soil pits in the Grassland, Soil and Water Research Laboratory (GSWRL) Riesel Watersheds near Riesel, Texas, and assessed the uncalibrated model performance of the topographically distributed initialization relative to the current SWAT implementation that incorporates standard SSURGO data. Further comparison to an unconstrained calibration of a SWAT initialization for each field illustrates the strengths and weaknesses of each approach.

MATERIALS AND METHODS

Overview

Initialization of SWAT commonly proceeds via ArcSWAT, a GIS interface that facilitates model set-up

and parameterization from soils, land use and digital elevation datasets (Veith *et al.*, 2008; Nietsch *et al.*, 2009; Arnold *et al.*, 2010; Winchell *et al.*, 2013). We developed an ArcMap toolbox for the ArcSWAT interface (ArcMap versions 10.0 and later) that introduces topographic attributes without disrupting the standard initialization process (i.e. no changes to source code or other supporting files are necessary). This TopoSWAT toolbox (Fuka and Easton, 2015) acquires and processes the requisite data layers needed for general SWAT model watershed delineation, updates the SWAT databases and creates the necessary SWAT lookup tables to link spatial data to the databases. To permit global application, this method uses the FAO-UNESCO Digital Soil Map of the World (FAO, 2007) as a starting point from which to distribute parameter values, although any digital soil map with comparable feature data and supporting databases could be incorporated. In addition to a standard ArcSWAT initialization using SSURGO data, we tested the TopoSWAT approach by comparing model parameter values and simulation outputs to soil pit data and runoff observations at the Riesel watersheds.

Watershed description

The Riesel experimental watersheds were established by the USDA Soil Conservation Service (now NRCS) in 1938 to better understand hydrologic processes in agricultural watersheds that affect soil erosion, nutrient

cycling, flood events and the agricultural economy (USDA, 1942; Harmel *et al.*, 2007). The Riesel watersheds contains 340 ha of federally owned and operated land in the 2372 ha Brushy Creek watershed within the 45 000 km² Texas Blackland Prairie (Figure 1). The long-term agricultural management records and abundance of measurements (Harmel *et al.*, 2014) make this site ideal for rigorous model comparison.

Surface drainage in the Riesel watersheds is moderate to rapid after storm events. The highly expansive Houston Black clay vertisol soils dominate the watershed, with typical particle size distributions of 17% sand, 28% silt and 55% clay (Harmel *et al.*, 2007). These deep, well-drained soils formed from weakly consolidated calcareous clays with 1–3% slopes in upland areas. This soil has low saturated conductivity when wet (approximate saturated hydraulic conductivity of 1.5 mm/h), although preferential crack flow causes high infiltration rates when the soil is dry (Allen *et al.*, 2005; Arnold *et al.*, 2005). Land use in the region is divided evenly between grassland and cropland, with the latter mostly cotton, grain sorghums, corn, wheat, oats and hay. The grassland is mostly improved pasture, with native range on the shallower and steeper soils. Soil erosion is considered a major management problem in the region (Texas Almanac, 2014). Climate in the region consists of long, hot summers and short, mild winters (average July high temperatures are ~35 °C; January low average ~2 °C). The growing season extends from mid-March to mid-

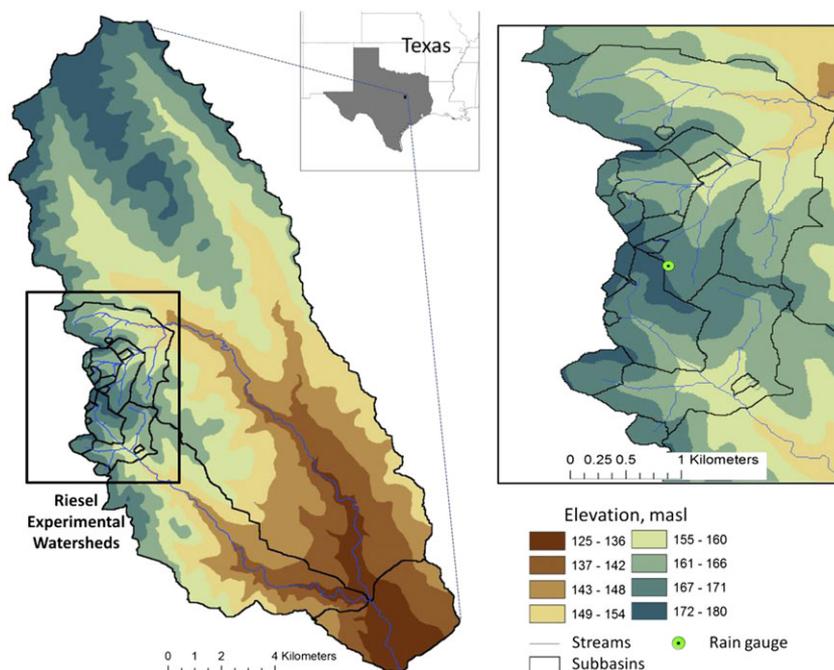


Figure 1. Location of the Grassland, Soil and Water Research Laboratory experimental watersheds near Riesel, Texas, digital elevation model, agricultural fields and precipitation gauge in central Texas, USA

November with most of the ~90 cm of annual precipitation delivered during convective thunderstorms in the summer (Harmel *et al.*, 2006).

Soil pedon data

Data from Holtan *et al.* (1968) were used to identify, digitize and geolocate eight soil pits in the Riesel watersheds. Variables digitized from the report include the pit ID, soil name, soil texture, depth to the C-horizon, drainage class, saturated conductivity, bulk density, porosity and the field capacities at 33 kpa (~field capacity) and 1500 kpa (~wilting point) tensions.

Field runoff data

The Riesel watersheds have an extensive field-level runoff collection network with two major monitored catchments, which together include 10 experimental fields and three mixed land uses currently monitored, for a total of 15 flow gages¹ (Figure 1). Approximately 1300 gage years of daily runoff and 700 gage years of soil loss data are available (Harmel *et al.*, 2014). There is also an extensive continuously monitored precipitation gauge network. All data are available at <http://www.ars.usda.gov/spa/hydro-data>. Discharge measurements are made by continuously recording flow levels in a stilling well located in each calibrated flume or weir structure. ISCO 6700 samplers are installed at each site to automatically collect water quality samples in each runoff event. Data were aggregated to daily values for use in this study.

SWAT description

SWAT is a watershed-scale hydrologic and non-point source pollution water quality model that combines soil, land use and management information with weather data to simulate surface and subsurface hydrological processes and the associated chemical and sediment fluxes. The meteorological forcing data required to run the model include precipitation, temperature, relative humidity, wind speed and solar radiation, although SWAT has the ability to estimate missing values using a weather generator. The core SWAT executable is compiled from FORTRAN and requires a number of carefully formatted text input files.

The ArcSWAT interface (for ArcGIS 10.1) facilitates basin delineation, the overlay of spatial data for soils, land use, elevation and management and the subsequent parameterization of SWAT from standard databases. ArcSWAT lumps unique combinations of land use, soil types and slope classes into hydrologic response units (HRUs). All HRUs of a given soil type, land use and

slope class have identical physical representation regardless of physical location within a sub-basin. This aggregation reduces computational complexity while simultaneously reducing the influence of the underlying spatial configuration (Easton *et al.*, 2008).

Forcing data. Despite the availability of numerous precipitation gauges within the research station, we selected a single, centrally located station (RG89) to minimize the complexity of performance biases across sub-basins. While it seems intuitive that the closest precipitation gauge to each catchment would best represent the individual catchments, this has been demonstrated to frequently not be the case both in high-density (Auerbach *et al.*, 2016) as well as low-density (Fuka *et al.*, 2014a) precipitation gauge networks. In addition, unconstraining the study to use the best fitting precipitation gauge for each specific catchment by definition would increase degrees of freedom with the effect of a net increase in total system performance, so the intention of this study is to use a single central precipitation gauge as a control allowing the differences in performances between catchments to be more easily represented. Daily maximum and minimum temperatures are more spatially uniform than precipitation and have less overall influence on the daily hydrologic mass balance. Consequently, we derived temperature values from the closest Global Historical Climate Network (GHCN) weather station, 15 km south of the research station at Marlin, TX (TX MARLIN 3 NE, GHCN: USC00415611). Relative humidity, solar radiation and wind speed were simulated using SWAT's weather generator (WGEN_US_First Order).

Spatial data. Elevation data, spatial and tabular attribute land use data and spatial and tabular soil data are required data for model initialization. A digital elevation model with a resolution of 1/3 arcsecond (roughly 10 m) was obtained for the watershed area from the National Elevation Dataset from the US Geological Survey National Map service (<http://www.nationalmap.gov>). The land use layer for the models was created by converting shapefiles of the GSWRL field runoff plots to a raster layer at the resolution of the digital elevation model (DEM) and then overlaying this map onto the National Land Cover Database 2011 dataset from the US Geological Survey National Map service. The default ArcSWAT values were used for the management parameters of National Land Cover Database classes. In order to correctly represent management that occurred on the Riesel watersheds, a SWAT-specific database was developed that included dates and operation types for crop planting, tillage, fertilization and harvest. These data are available for every runoff plot represented in the study

¹For consistency with NOAA and USGS terminology, 'gauge' will refer to weather sensors, and gage will refer to flow sensors.

(<http://www.ars.usda.gov/spa/hydro-data>), with over 1000 specific management operations included in the SWAT management tables. For the standard SWAT initialization, SSURGO spatial layers were downloaded from the USDA WebSoilSurvey service, and soil attributes were obtained from the SSURGO database established for SWAT (Sheshukov *et al.*, 2012).

TopoSWAT. The TopoSWAT initialization replaces the SSURGO soil layer with soil parameters that are dynamically distributed from a spatial combination of the FAO-UNESCO Digital Soil Map of the World (FAO, 2007) and TI classes based on the base digital elevation data (Fuka and Easton, 2015: Available at: 10.6084/m9.figshare.1342823). The TI [Equation 1] represents the propensity of a landscape unit to soil saturation and subsequent runoff generation by combining two important controls on these processes: the upslope contributing area (α) that drains through any given point and the local slope gradient ($\tan \beta$) (Beven and Kirkby, 1979; Easton *et al.*, 2008).

$$TI = \ln(\alpha/\tan\beta) \quad (1)$$

Our approach requires that two or more TI classes be defined for the smallest SWAT sub-basin area in the initialization. For this study, a raster of TI values was resampled to create a TI class layer by division into 10 equal area classes, ranging from TI class 1 (the 10% of the watershed with the highest TI) to a TI class 10 (the 10% of the watershed with the lowest TI). The method could accommodate other numbers and modes of class division, but the 10 equal area increments addressed upslope, midslope and downslope differences and allowed alignment with the smallest area on which management could be performed. The default TopoSWAT protocol used for this study varied each soil parameter by $\pm 15\%$ of the mean relative to 10 TI classes. Users familiar with the watershed of interest could adjust these percentages across a wider or narrower range (although we are performing this study as if the basin was ungauged and no expert knowledge of the basin is available). Field capacity, wilting point soil moisture capacity and AWC were calculated using the soil water characteristic relationships from Saxton and Rawls (2006) and applied to the TI-based soils in the model. In order to examine the potential benefits of the methodology for modelling ungauged basins and data-scarce areas, soil parameters were not further adjusted to fit pedon measurements in the watershed.

A single slope class is used during HRU delineation to avoid accounting for slope twice within the TopoSWAT initialization [note slope is explicit in Equation 1]. Two slope classes were used for the

SSURGO-based initialization, given the relatively flat topography and to ensure consistency with the TopoSWAT-distributed soil initialization. This yielded similar number of HRUS (Table I, $n=918$ and $n=900$ for the SSURGO-based and TopoSWAT respectively), making the new method computationally similar to the traditional initialization methods.

Analysis

We first compared the topographically distributed soil data and SSURGO-derived soil data to the measured pit data at eight locations in the Riesel watersheds that were well distributed across the TI classes.

Within SWAT, model performance, measured as goodness of fit to observed runoff, was employed to evaluate the SSURGO-based and distributed TopoSWAT initializations without calibration. Uncalibrated performance provided an indication of how these approaches might perform in ungauged locations and prevented inter-parameter compensation (e.g. storage contributions from adjusting AWC may be compensated for by soil depth), which often occurs during calibration in multiparameter watershed models and complicates inferences regarding model results. In addition, we assessed the default (uncalibrated) performance of these initializations against simple SWAT model initializations (three sub-basins, 3 HRUs per sub-basin) that were individually calibrated to each runoff station. Calibration of each runoff station was performed using the Differential Evolution Optimization package (Ardia and Mullen, 2009) coupled to the SWATmodel package (Fuka *et al.*, 2014b) in R (R Core Team, 2013). The calibration was performed for years 2000–2008 with a 2-year warm-up period. Calibrated parameters are shown in Table II.

RESULTS

Relationships between TI classes and measured soil properties

TI classes and soil pedon data exhibited significant correlations, as predicted by Moore *et al.* (1993). This result fit with the expectation that downslope soil

Table I. Differences in total HRUs from resulting change in slope classes (no. of slope class) and soil database resolution (no. of distinct soils)

	SSURGO soils	Distributed soils
No. of distinct soils	33	20
No. of slope classes	2	1
Total HRUs	918	900

Table II. Calibrated parameters used for Differential Evolution Optimization, within an initial range published in the literature

Variable	Definition	Range
GW_DELAY	Groundwater delay (days)	1–180 days
ALPHA_BF	Baseflow alpha factor (1/days)	0–1
SURLAG	Surface runoff lag time (days)	1–180 days
GWQMN	Threshold depth of water in the shallow aquifer (m)	1–200 mm
CN_{II}	Average SCS CN_{II} value	40–85
RCHRG_DP	Deep aquifer percolation fraction	0–1.0
REVAPMN	Depth of water in the aquifer for revap (mm)	0–500 mm
GW_REVAP	Groundwater revap coefficient	0–0.2
ESCO (HRU)	Soil evaporation compensation factor	0.2–0.99
EPCO (HRU)	Plant uptake compensation factor	0.2–0.99

movement into areas with higher TI classes generates deeper soils with higher clay content and lower saturated conductivity than lower TI classes (Figure 2a–f). The strongest correlations with TI class were evident for field capacity at 33 kpa, porosity and AWC (Figure 2c, d, f, respectively). Soil depth, saturated conductivity (Ksat) and bulk density were also correlated although not as strongly (Figure 2a, b, d, respectively).

Soil parameter values at initialization

Parameter values in the standard SSURGO-based initialization poorly represented the measured soil

features at the locations corresponding to sampled pedons (Figure 3). The SSURGO data show an unrealistically deep maximum soil depth value across the landscape (red '+' in Figure 3a) and consequently over-predicted AWC (Figure 3c). In contrast, the dynamically distributed soil initialization via TopoSWAT produced parameter values for soil depth and AWC that more accurately reflected measurements and the varied soil depths of this region (Figure 3).

Neither initialization estimated Ksat accurately, but variability in Ksat was better captured by the topographically distributed soils. An outlier Ksat value (measured at >60 mm/h) was significantly higher than the characteristic regional values and especially poorly estimated.

Clear differences in the accuracy of initial estimates were apparent for AWC values, with the SSURGO initialization substantially over-predicting measurements (especially at low values). The TopoSWAT values fell much closer to the 1:1 line, with the tendency to over-predict low AWC and under-predict high measurements.

Hydrologic model performance comparison

Both SWAT model initializations resulted in similar uncalibrated performance, measured in terms of daily NSE for field runoff. The topographically distributed soils outperformed SSURGO in 7 of 15 fields, while SSURGO was superior in the remaining 8 (Figure 4; Table III). Performance differences were small, however, with the maximum absolute difference=0.07 (Figure 4f). Model

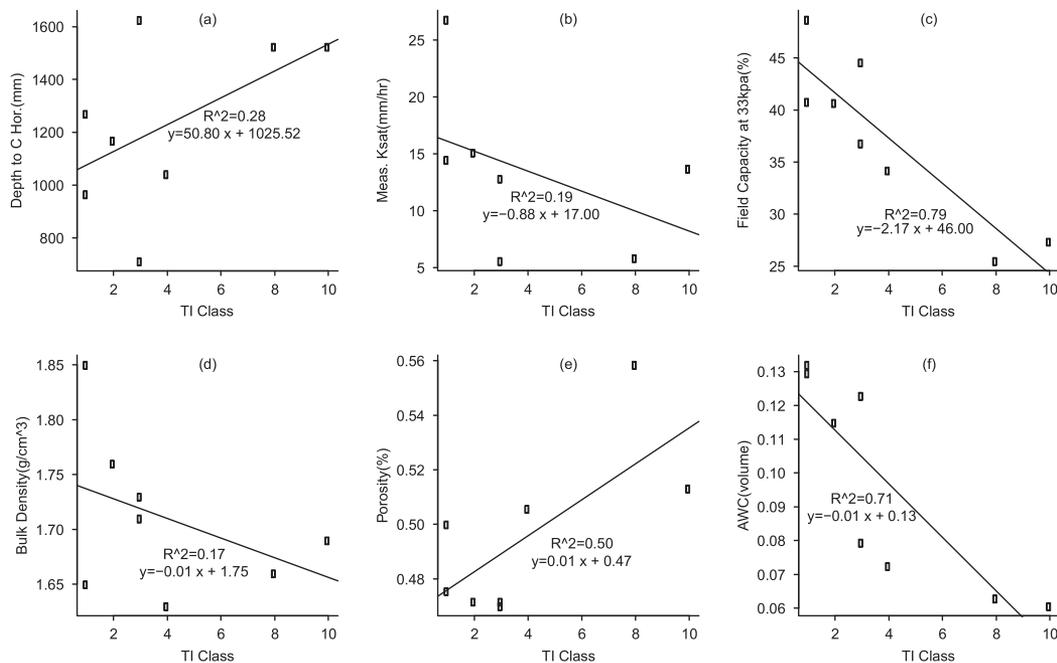


Figure 2. Relationship between TI class and measured soil depth (a), Ksat (b), field capacity (c), bulk density (d), porosity (e) and AWC (f) for soil pedons at the Riesel watersheds

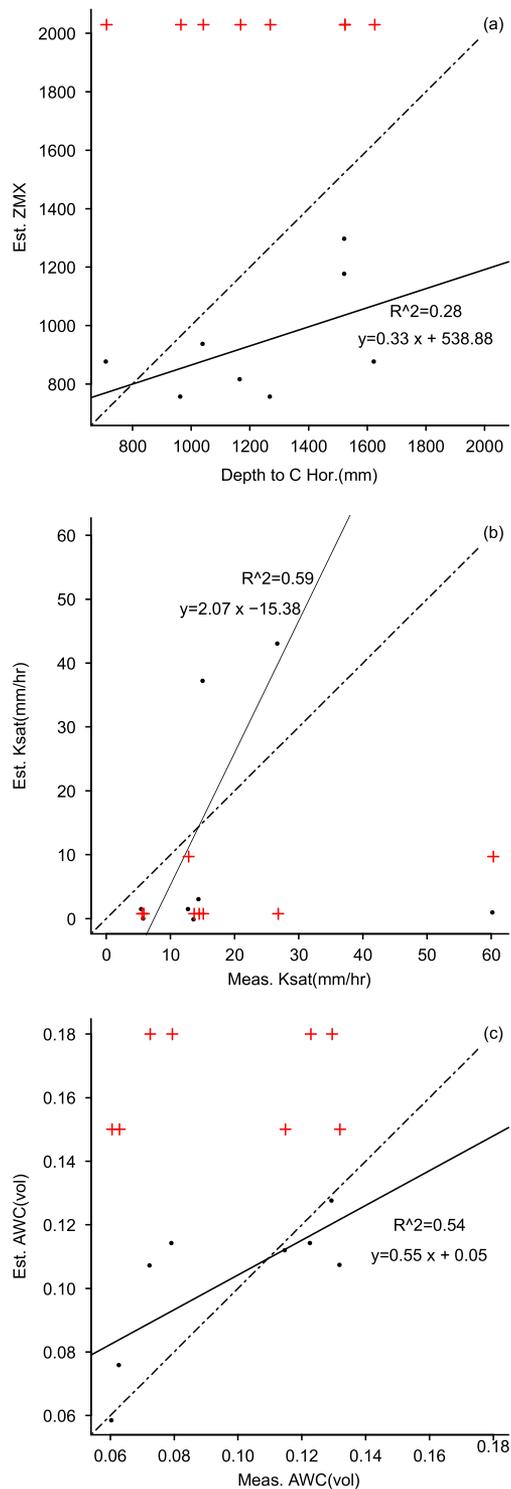


Figure 3. Comparison of soil pedon values at the Riesel watersheds against the default parameter initializations from ArcSWAT using the standard SSURGO initialization (red +s) and the default TI-distributed soils (black dots). Deviation from the 1:1 line of estimated (y -axis) versus measured (x -axis) reflects the accuracy of the initialized values. The regression line and Pearson's R^2 correlation coefficient for the dynamically distributed soils are shown in panels for (a) soil depth (ZMX), (b) Ksat and (c) AWC. The outlier in Ksat, which is uncharacteristic of regional soils, was excluded from the regression in (b)

performance was modestly better in fields with larger catchment area (Figure 4 and Table III), but dynamically distributed soils performed better than SSURGO in the smaller catchment areas (Figure 4f; Table III).

After calibrating the models, the mean performance improvement in each basin was an NSE increase of 0.12 (Table III). However, the magnitude of increased NSE varied across basins (standard deviation/range, Table III). Field Y14, a poorly performing field in both initializations, had an NSE increase of 0.28 with calibration (from NSE of 0.41 to 0.69). Furthermore, the uniformly poor performance of the smallest fields (P1–P4, four runoff plots aggregated as field PS) demonstrated the influence of modelling and or input data limitations rather than the initialization or calibration approach (Table III).

DISCUSSION

In the context of known relationships between TI and soil characteristics in saturation excess-dominated soils (Ciolkosz and Waltham, 2000; Buchanan *et al.*, 2014), these results support the use of topographic data and analyses to create spatially distributed soil maps for hydrological modelling purposes in infiltration excess-dominated soils, especially where high-resolution soil data are not available. Soil scientists have historically focused on the vertical relationships of soil horizons and soil-forming processes (Buol *et al.*, 1989), rather than the horizontal relationships that are important in describing water movement across and through a landscape (Easton *et al.*, 2008; Collick *et al.*, 2014). Spatial soil patterns are typically represented as choropleth maps with discrete lines representing the boundaries between map units (Moore *et al.*, 1993). Nevertheless, these divisions may reflect political, land cover or some other arbitrary boundary and may poorly capture the heterogeneity of vadose zone properties and processes. Moreover, the sampled pedon, which forms basis for map unit values, may be many kilometres from a point of interest, and the range of values for attributes describing hydraulic properties can vary by an order of magnitude or more as shown in Figure 3. Thus, the use of topographic data to infer spatial variability in soil properties can help to meet the demand for increasingly high-resolution modelling when resources are unavailable for conventional sampling. Furthermore, this approach allows hydrologic modellers to take full advantage of the steadily increasing resolution of DEMs to enhance soil characterization via the best available datasets (e.g. with Light Detection and Ranging and drone imagery rather than satellite sensing).

The model performance without calibration elucidates the benefits of the topographically distributed soil parameters, as calibration risks decoupling parameter

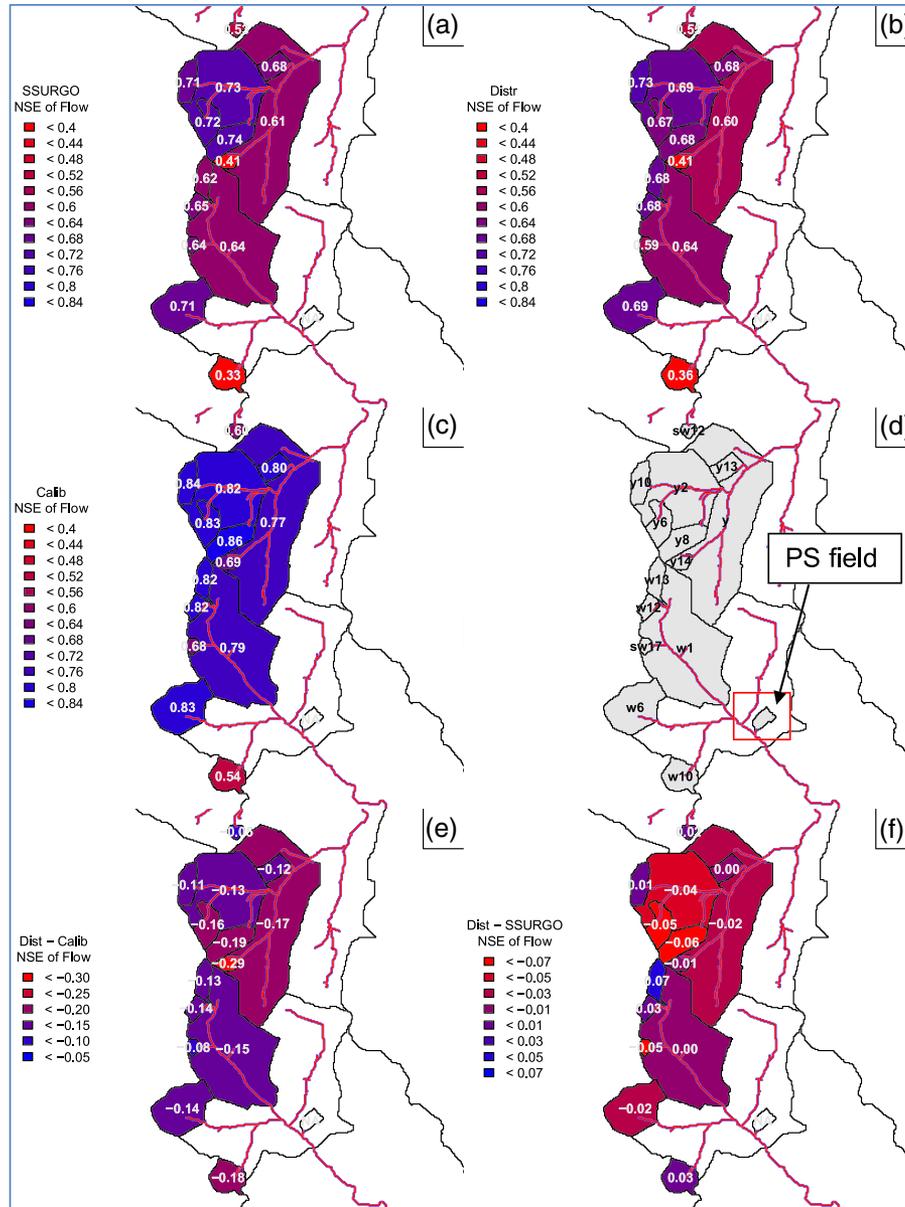


Figure 4. Performance for each runoff station ArcSWAT initialization using SSURGO (a), ArcSWAT initialization using dynamically distributed soils (b), for individually optimally calibrated 9HRU SWAT initializations (c), the difference in performance between the dynamically distributed soils initializations and optimal individual station calibration (e) and the difference in performance between the dynamically distributed soils and SSURGO soils initializations (f). A map of station names (d) is included for cross reference with the results in Table III

values from the inferred hydrologic processes. For example, the parameters AWC, soil depth and saturated conductivity can compensate for each other in influencing basin outflow, despite their unique effects on soil chemistry and hydraulics. While it is possible to numerically coerce models of both soil representations to match the outlet runoff, the act of calibration permits inter-parameter compensation and can reduce the mechanistic meaning of the model. This can result in non-repeatable and unrealistic parameter sets (Vitek and Kalibera, 2012) and may hinder evaluation of potential management scenarios that are modelled using parameter adjustments.

The default parameter set from the dynamically distributed soil initialization therefore represents the combination of independent physical data regarding soil attributes with the known variation in hydrologic processes that are a result of topography, a detail that the current best available soil maps often miss. This can permit a better understanding of the allocation of model uncertainty among structural, data and observation components. For example, both implementations tended to misrepresent the storage but did so in different manners. The SSURGO-based initialization captured baseflow fairly poorly, sometimes under-predicting and sometimes over-

Table III. Performance for each runoff station for an optimal calibration at the runoff station, an uncalibrated ArcSWAT initialization using SSURGO data, an uncalibrated ArcSWAT initialization using dynamically distributed soils and the difference in performance between the dynamically distributed soils and SSURGO soil initializations

Runoff station	Area (ha)	Calibrated single basin/field	NSEs		
			SSURGO	TopoSWAT	TopoSWAT-SSURGO
Y	125x.0	0.77	0.61 (−0.16)	0.60 (−0.17)	−0.02
W1	71.2	0.79	0.64 (−0.15)	0.64 (−0.15)	0.00
Y2	53.4	0.82	0.73 (−0.09)	0.69 (−0.13)	−0.04
W6	17.1	0.83	0.71 (−0.12)	0.69 (−0.14)	−0.02
Y10	8.5	0.84	0.71 (−0.12)	0.73 (−0.11)	0.01
Y6	8.5	0.83	0.72 (−0.11)	0.67 (−0.16)	−0.05
Y8	8.4	0.87	0.74 (−0.13)	0.68 (−0.19)	−0.06
W10	8.0	0.54	0.33 (−0.21)	0.36 (−0.18)	0.03
Y13	4.6	0.80	0.68 (−0.13)	0.68 (−0.12)	0.00
W13	4.6	0.82	0.62 (−0.20)	0.68 (−0.13)	0.07
W12	4.0	0.82	0.65 (−0.16)	0.68 (−0.14)	0.03
Y14	2.3	0.69	0.41 (−0.28)	0.41 (−0.29)	−0.01
SW17	1.2	0.68	0.65 (−0.03)	0.59 (−0.08)	−0.05
SW12	1.2	0.60	0.53 (−0.07)	0.55 (−0.05)	0.02
PS ^a	0.4	0.23	0.29 (0.06)	0.22 (−0.01)	−0.07

Runoff stations are listed in order of size of contributing area.

^a Fields P1–P4 were aggregated into one area for model simulation, called PS.

predicting, in part because of the single, unduly large soil depth value and the higher than measured AWC. This artificially high soil storage resulted in an increase in base flow as an absorbed moisture front in the soils seeping into the storage below the soil profile.

While this study was performed to clarify the confidence with which this methodology could be applied to ungaged basins and data-scarce areas, the relationship between TI class and pedon measurements suggests that additional pedon data could help to establish a better characterization of soil parameters in a basin by refining the TI divisions. The age and limited number of pedons in this area hindered a cross-validation study (i.e. leave-one-out replicates) to further examine how the quantity of available measurements might affect the soil parameter distribution.

Model computational time, as well as the computational input and output requirements, for a simulation increases linearly with the number of HRUs that are used in a SWAT model, and it is common that the bulk of the computational time for a SWAT modelling project is used in the calibration of the model. An additional benefit to the distributed soil method is that a better representation of the soil parameters decreases the need to calibrate several of the most sensitive soil parameters (Van Griensven *et al.*, 2006), which can significantly reduce the amount of calibration required; a portion of this increase is compensated for by the removal of the requirement to add slope classes, traditionally starting at 3, as well as the ability to decrease the number of base soil types when using lower resolution soils maps like FAO.

Initial perception for this methodology could consider the resulting parameterizations more complex, but the method for initialization is in fact simpler. The more

complex parameterization comes from lower density soil classification data that is currently distributed with the TopoSWAT initialization system for the entire globe, and parameters distributed with TI that is calculated from the DEM, already required. The traditional watershed parameterization requires a user to locate and download an additional soil map and download, add to the project database and perform quality checks on additional corresponding soil reference tables, steps not required for the methodology presented here.

In this study, the ArcSWAT initialization using dynamically distributed soils consistently produced greater flows relative to the SSURGO-based soils. This suggests the applied relevance of the proposed approach, whereby inaccuracy because of spatial homogeneity of soil parameters risks inaccurate evaluation of management alternatives dependent on these soil parameters. Underestimating high flow events could lead to less conservative predictions regarding nutrient, sediment and pesticide transport in a management planning context.

CONCLUSION

Ultimately, the goal of this research is to identify the source areas of nutrient and sediment loss more accurately and precisely, although in this article, we focus on ways to better represent the soil system within the model. Refining the soil properties for SWAT model initialization can elucidate where high runoff potential and high nutrient availability intersect, thereby better informing spatially targeted agricultural and water resource management decisions.

While this method only improved the model performance in 7 out of 15 basins, better representation of the parameters controlling the processes within the field is important, even if the benefit may not be fully realized at the outlet. The net effect of the method has the added benefit of reducing the calibration needs of the model, as removing soil depth, Ksat field capacity, bulk density, porosity and AWC from the calibration process decreases the number of iterations the calibration process will require before optimum performance will be achieved. In addition, decreasing the number of calibrated parameters increases the probability that a parsimonious model will result.

This study demonstrated that in an ungaged location, a generalized soil distribution may not represent the soil system as well as is possible, and indeed, further knowledge can be used to better develop a relationship describing soil genesis and ultimately water quality impacts. Topographically distributed soils can better represent the spatial heterogeneity of soil properties, and further work should explore how the increasing availability of high-resolution elevation data can be best coupled with existing soil datasets.

ACKNOWLEDGEMENTS

We would like to acknowledge the funding support from NSF [award no. 1343802, EarthCube Building Blocks: A Broker Framework for Next Generation Geoscience (BCube)], USDA NRCS Conservation Innovation Grants Program (award no. 69-3A75-13-232), a NatureNet postdoctoral fellowship from The Nature Conservancy, and computational support from National Center for Atmospheric Research (Computational and Information Systems Laboratory. 2015. Yellowstone: IBM iDataPlex System (NCAR Community Computing). Boulder, CO: National Center for Atmospheric Research. <http://n2t.net/ark:/85065/d7wd3xhc>).

REFERENCES

- Allen PM, Harmel RD, Arnold JG, Plant B, Yelderma J, King KW. 2005. Field data and flow system response in clay (Vertisol) shale terrain, north central Texas. *USA Hydrologic Processes* **19**: 2719–2736.
- Ardia, D, Mullen, K. 2009. DEoptim: differential evolution optimization in R. *R package version 2.2–2*. <http://CRAN.R-project.org/package=DEoptim> (accessed October 2013).
- Arnold JG, Allen PM, Volk M, Williams JR, Bosch DD. 2010. Assessment of different representations of spatial variability on SWAT model performance. *Transactions of the ASABE* **53**(5): 1433–1443.
- Arnold JG, Potter KN, King KW, Allen PM. 2005. Estimation of soil cracking and the effect on surface runoff in a Texas Blackland Prairie watershed. *Hydrologic Processes* **19**: 589–603.
- Auerbach DA, Easton ZM, Walter MT, Flecker AS, Fuka DR. 2016. Evaluating weather observations and the climate forecast system reanalysis as inputs for hydrologic modeling in the tropics. *Hydrological Processes* DOI:10.1002/hyp.10860
- Beven KJ, Lamb R, Quinn P, Romanowicz R, Freer J. 1995. TOPMODEL. In *Computer Models of Watershed Hydrology*, Sing VP (ed). Water Resources Publications: Colorado; 627–668.
- Beven KJ, Kirkby MJ. 1979. Towards a simple physically based variable contributing model of catchment hydrology. *Hydrological Sciences Bulletin* **24**(1): 43–69.
- Bosch NS, Allan JD, Selegean JP, Scavia D. 2013. Scenario-testing of agricultural best management practices in Lake Erie watersheds. *Journal Of Great Lakes Research* **39**(3): 429–436. DOI:10.1016/j.jglr.2013.06.004
- Bouma, J. 1989. Advances in soil science. *Using soil survey data for quantitative land evaluation*. Springer-Verlag New York Inc.: New York, NY, USA; (pp. 177–213).
- Buchanan BP, Fleming M, Schneider RL, Richards BK, Archibald J, Qiu Z, Walter MT. 2014. Evaluating topographic wetness indices across central New York agricultural landscapes. *Hydrology and Earth System Science* **18**: 3279–3299. DOI:10.5194/hess-18-3279-2014
- Buol SW, Hole FD, McCracken RJ. 1989. *Soil genesis and classification*. Iowa State University Press: Ames, Iowa; 456.
- Cioikosz EJ, Waltham WJ. 2000. *Pennsylvania's fragipans*. Agronomy Department, the Pennsylvania State University: State College, PA, USA.
- Collick AS, Fuka DR, Kleinman PJA, Buda AR, Weld JL, White MJ, Veith TL, Bryant RB, Bolster CH, Easton ZM. 2014. Predicting phosphorus dynamics in complex terrains using a variable source area hydrology model. *Hydrological Process Early View* DOI:10.1002/hyp.10178
- Crossman J, Futter MN, Oni SK, Whitehead PG, Jin L, Butterfield D, Baulch H, Dillon PJ. 2013. Impacts of climate change on hydrology and water quality: future proofing management strategies in the Lake Simcoe watershed, Canada. *Journal of Great Lakes Research* **39**(1): 19–32. DOI:10.1016/j.jglr.2012.11.003
- Dunne T, Black RD. 1970. Partial area contributions to storm runoff in a small New England watershed. *Water Resources Research* **6**(5): 1296–1311.
- Easton ZM, Gérard-Marchant P, Walter MT, Petrovic AM, Steenhuis TS. 2007a. Hydrologic assessment of an urban variable source watershed in the Northeast United States. *Water Resources Research* **43**: W03413. DOI:10.1029/2006WR005076
- Easton ZM, Gérard-Marchant P, Walter MT, Petrovic AM, Steenhuis TS. 2007b. Identifying dissolved phosphorus source areas and predicting transport from an urban watershed using distributed hydrologic modeling. *Water Resources Research* **43**: W11414. DOI:10.1029/2006WR005697
- Easton ZM, Fuka DR, Walter MT, Cowan DM, Schneiderman EM, Steenhuis TS. 2008. Re-conceptualizing the Soil and Water Assessment Tool (SWAT) model to predict runoff from variable source areas. *Journal of Hydrology* **348**: 279–291. DOI:10.1016/j.jhydrol.2007.10.008
- Easton ZM, Walter MT, Fuka DR, White ED, Steenhuis TS. 2011. A simple concept for calibrating runoff thresholds in quasi-distributed variable source area watershed models. *Hydrological Processes* **25**(20): 3131–3143.
- FAO, 2007. <http://www.fao.org/geonetwork/srv/en/metadata.show?id=14116>. Accessed 05/01/2014.
- Foth HD. 1943. Fundamentals of Soil Science. In *8th Edition (1990)*. John Wiley & Sons: Hoboken, NJ, USA.
- Frankenberger JR, Brooks ES, Walter MT, Walter MF, Steenhuis TS. 1999. A GIS-based variable source area hydrology model. *Hydrological Processes* **13**(6): 805–822.
- Fuka, D.R., Easton, Z.M. 2015. TopoSWAT Source. figshare. DOI: 10.6084/m9.figshare.1342823
- Fuka DR, Walter MT, MacAlister C, Steenhuis TS, Easton ZM. 2014a. SWATmodel: a multi-operating system, multi-platform SWAT model package in R. *Journal of the American Water Resources Association* **50** (5): 1349–1353.
- Fuka DR, Walter MT, MacAlister C, Degaetano AT, Steenhuis TS, Easton ZM. 2014b. Using the Climate Forecast System Reanalysis as weather input data for watershed models. *Hydrological Processes* **28**(22): 5613–5623. DOI:10.1002/hyp.10073
- Gassman PW, Reyes MR, Green CH, Arnold JG. 2007. The Soil and Water Assessment Tool: historical development, applications, and future research directions. *Transactions of the ASABE* **50**: 1211–1250.

- Hadzick ZZ, Guber AK, Pachepsky YA, Hill RL. 2011. Pedotransfer functions in soil electrical resistivity estimation. *Geoderma* **164**: 195–202. DOI:10.1016/j.geoderma.2011.06.004
- Harmel RD, Haney RL, Smith DR, White MW, King KW. 2014. USDA-ARS Riesel Watersheds, Riesel, TX, USA: water quality research database. *Water Resources Research* **50**: 8374–8382. DOI:10.1002/2013WR015191
- Harmel RD, Bonta JV, Richardson CW. 2007. The Original USDA-ARS Experimental Watersheds in Texas and Ohio: contributions from the past and visions for the future. *Transactions of the ASABE* **50**(5): 1669–1675.
- Harmel RD, Richardson CW, King KW, Allen PM. 2006. Runoff and soil loss relationships for the Texas Blackland Prairies ecoregion. *Journal of Hydrology* **331**(3–4): 471–483.
- Hengl T, Toomanian N, Reuter H, Malakouti MJ. 2007. Methods to interpolate soil categorical variables from profile observations: lessons from Iran. *Geoderma* **140**(4): 417–427. DOI:10.1016/j.geoderma.2003.08.018
- Holtan, H.N., England, C.B., Lawless, G.P., and Schumaker, G.A., 1968. Moisture tension data for selected soils on experimental watersheds. USDA-ARS publication RS-41-144 October 1968. (archive available at: <https://archive.org/details/moisturetensiond14holt>)
- Horton RE. 1933. The role of infiltration in the hydrologic cycle. *Transactions, American Geophysical Union* **14**: 446–460.
- Jana RB, Mohanty BP. 2011. Enhancing PTFs with remotely sensed data for multiscale soil water retention estimation. *Journal of Hydrology* **399** (3–4): 201–211.
- Jenny H. 1941. *Factors of soil formation: a system of quantitative pedology*. McGraw-Hill: NY.
- Levi MR, Schaap MG, Rasmussen C. 2015. Application of spatial pedotransfer functions to understand soil modulation of vegetation response to climate. *Vadose Zone Journal* **14**(9). DOI: 10.2136/vzj2014.09.0126
- Lin, H. 2012. Hydrogeology: addressing fundamentals and building bridges to understand complex pedologic and hydrologic interactions. Chapter 1. Henry Lin (Ed.) 2012. *Hydrogeology: Synergistic Integration of Soil Science and Hydrology*. Academic Press/Elsevier: Cambridge, Massachusetts, USA. DOI: 10.1016/B978-0-12-386941-8.00001-0.
- McBratney AB, Mendonça Santos ML, Minasny B. 2003. On digital soil mapping. *Geoderma* **117**: 3–52. DOI:10.1016/S0016-7061(03)00223-4
- McBratney AB, Minasny B, Viscarra Rossel R. 2006. Spectral soil analysis and inference systems: a powerful combination for solving the soil data crisis. *Geoderma* **136**: 272–278.
- Minasny B, McBratney AB, Malone BP, Wheeler I. 2013. Digital Mapping of Soil Carbon. In *Advances in Agronomy*, **118**. Academic Press: Newark, NJ.
- Minasny B, McBratney AB, Tranter G, Murphy BW. 2008. Using soil knowledge for the evaluation of mid-infrared diffuse reflectance spectroscopy for predicting soil physical and mechanical properties. *European Journal of Soil Science* **59**: 960–971. DOI:10.1111/j.1365-2389.2008.01058.x
- Mitsch WJ, Gosselink JG. 2000. *Wetlands*, 3rd edition. John Wiley and Sons: NY.
- Moore ID, Gessler PE, Nielsen GA, Peterson GA. 1993. Soil attribute prediction using terrain analysis. *Soil Science Society of America Journal* **57**(2): 443–452.
- Nield SJ, Boettinger JL, Ramsey RD. 2007. Digitally mapping gypsic and natric soil areas using Landsat ETM data. *Journal of Soil Science Society of America* **71**: 245–252.
- Nietsch SL, Arnold JG, Kiniry JR, Williams JR, King KW. 2009. *Soil and Water Assessment Tool, Theoretical Documentation, Version 2009, Technical Report TR-406*. Texas Water Resources Institute: College Station, TX.
- Niraula R, Kalin L, Srivastava P, Anderson C. 2013. Identifying critical source areas of nonpoint source pollution with SWAT and GWLF. *Ecological Modelling* **268**: 123–133. DOI:10.1016/j.ecolmodel.2013.08.007
- Petersen C. 1991. Precision GPS navigation for improving agricultural productivity. *GPS World* **2**(1): 38–44.
- Saadat H, Bonnell R, Sharifi F, Mehuis G, Namdar M, Ale-Ebrahim S. 2008. Landform classification from a digital elevation model and satellite imagery. *Geomorphology* **00**: 453–464. DOI:10.1016/j.geomorph.2008.01.011
- Saxton KE, Rawls WJ. 2006. Soil water characteristic estimates by texture and organic matter for hydrologic solutions. *Soil Science Society of America Journal* **70**(5): 1569–1578. DOI:10.2136/sssaj2005.0117
- Schaap MG, Leij FJ, van Genuchten MT. 2001. Rosetta: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology* **251**: 163–176.
- Sharma SK, Mohanty BP, Zhu J. 2006. Including topography and vegetation attributes for developing pedotransfer functions. *Soil Science Society of America Journal* **70**(5): 1430–1440.
- Sheshukov AY, Daggupati P, Douglas-Mankin KR, Lee MC. 2012. High spatial resolution soil data for watershed modeling: 1. Development of a SSURGO-ArcSWAT utility. *Journal of Natural and Environmental Sciences* **2**(2): 15–24.
- Simonson RW. 1991. The US soil survey—contributions to soil science and its application. *Geoderma* **48**(1): 1–16.
- Smettem K, Pracilio G, Oliver Y, Harper R. 2004. Data availability and scale in hydrologic applications. In *Development of pedotransfer functions in soil hydrology*, Pachepsky Y, Rawls WJ (eds). Developments in soil science. Elsevier: Amsterdam; 253–272.
- Texas Almanac. 2014. <http://www.texasalmanac.com/topics/environment/soils-texas>. Accessed 05/01/2014
- Tseng P-H, Jury WA. 1993. Simulation of field measurement of hydraulic conductivity in unsaturated heterogeneous soil. *Water Resources Research* **29**(7): 2087–2099. DOI:10.1029/93WR00578
- USDA. 1942. *The agriculture, soils, geology, and topography of the Blacklands experimental watershed, Waco, Texas. Hydrologic Bulletin No. 5*. USDA-SCS: Washington, D.C..
- Van Griensven A, Meixner T, Grunwald S, Bishop T, Diluzio M, Srinivasan R. 2006. A global sensitivity analysis tool for the parameters of multi-variable catchment models. *Journal of Hydrology* **324**(1): 10–23.
- Van Liew MW, Feng S, Pathak TB. 2012. Climate change impacts on streamflow, water quality, and best management practices for the Shell and Logan Creek watersheds in Nebraska. *International Journal of Agriculture and Biological Engineering* **5**(1): 13–34. DOI:10.3965/j.ijabe.20120501.003
- Veith TL, Sharpley AN, Arnold JG. 2008. Modeling a small, northeastern watershed with detailed field-level data. *Transactions of the ASABE* **51** (2): 471–483.
- Verecken H, Weynants M, Javaux M, Pachepsky Y, Schaap MG, Genuchten MT. 2010. Using pedotransfer functions to estimate the van Genuchten-Mualem soil hydraulic properties: a review. *Vadose Zone Journal* **9**: 795Y820.
- Vitek J, Kalibera T. 2012. R3: repeatability, reproducibility and rigor. *ACM SIGPLAN Notices* **47**(4a): 30–36.
- Wigmosta MS, Vail LW, Lettenmaier DP. 1994. A distributed hydrology vegetation model for complex terrain. *Water Resources Research* **30**(6): 1665–1680.
- Winchell M, Srinivasan R, Di Luzio M, Arnold J. 2013. *ArcSWAT interface for SWAT user's guide*. Blackland Research Center: Texas Agricultural Experiment station and USDA Agricultural Research Service.
- Wösten JHM, Pachepsky YA, Rawls WJ. 2001. Pedo-transfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. *Journal of Hydrology* **251**: 123–150.
- Wösten HAB, Ruardy TG, van der Mei HC, Busscher HJ, Wessels JGH. 1995. Interfacial self-assembly of a *Schizophyllum commune* hydrophobin into an insoluble amphipathic membrane depends on surface hydrophobicity. *Colloids and Surfaces. B, Biointerfaces* **5**: 189–195.
- Zollweg JA, Gburek WJ, Steenhuis TS. 1996. SMORMod - A GIS-integrated rainfall-runoff model. *Transactions of the ASAE* **39**(4): 1299–1307.