Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Hydrology 295 (2004) 211–224
Journal of Hydrology
www.elsevier.com/locate/jhydrol

Effects of distribution-based parameter aggregation on a spatially distributed agricultural nonpoint source pollution model

Eileen Chen[a,1], D. Scott Mackay[b,*]

[a]Environmental Remote Sensing Center, University of Wisconsin—Madison, 1225 West Dayton St., Madison, WI 53715, USA
[b]Department of Geography, University of Buffalo, The State University of New York, 105 Wilkeson Quadrangle, Buffalo, NY 14261, USA

Received 9 April 2003; revised 12 March 2004; accepted 19 March 2004

Abstract

Accurate predictions of sediment yield from distributed models of runoff and sediment yield depends in part of how well matched the model structure is to input data spatial representation. This study investigated how model structure and input data representation affect sediment predictions made using the Soil and Water Assessment Tool (SWAT). The study focused on the integration of two specific components of SWAT: the Modified Universal Soil Loss Equation (MUSLE) and the hydrologic response unit (HRU). MUSLE, a watershed erosion model, was applied to different levels of watershed partitioning and alternative HRU schemes for a watershed and its two subwatersheds over a 4-year period of measured stream flow and sediment yield. The results show that across different levels of watershed partitioning, HRUs do not conserve sediment. Instead, HRUs introduce almost half of the variability in sediment generation, which has previously been attributed solely to input data aggregation. This occurs for two reasons. First, MUSLE defines a nonlinear relationship between sediment generation and HRU area, but the sediment load is scaled linearly from the HRU level to the subwatershed level. Second, HRUs aggregate land areas without regard for the surface connectivity assumptions, which are implicit in MUSLE. These conflicts caused by the integration of HRU and MUSLE makes it difficult to determine the effect of different land use on soil erosion. This study indicates that greater attention should be made to structuring the data inputs to match the underlying assumptions of sub-models within SWAT.

Keywords: Aggregation error; Geographic information systems; Hydrologic modeling; Nonpoint source pollution; Sediment transport

1. Introduction

Integrated hydrological computer models are increasingly used to facilitate watershed planning and management (Joao and Walsh, 1992; Wilson, 1996; Lohani et al., 2002). One way to build hydrological models capable of modeling complex natural systems is to integrate a number of simpler process-based sub-models. For example, the Soil and Water Assessment Tool (SWAT, Arnold et al., 1998) integrates sub-models for sediment delivery, soil erosion, and rainfall–runoff. Each sub-model is an abstraction of some component of a larger system, which carries with it assumptions related to the scale

* Corresponding author. Tel.: +1-716-645-2722; fax: +1-716-645-2329.
E-mail addresses: dsmackay@buffalo.edu (D.S. Mackay); echen@uwalumni.com (E. Chen).
[1] Present address: Alameda County Water District, 43885 South Grimmer Boulevard, Fremont, CA 94538, USA. Tel.: +1-510-668-4473.

0022-1694/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2004.03.029

at which the sub-model was derived, the availability of measured data, and the understanding of the modeled processes when the sub-model was conceived. These assumptions reflect a specific model purpose and its intended scope of capabilities. When sub-models are integrated to explain or predict hydrological responses, it may not be clear that all of the sub-model assumptions are being met. Empirical and conceptual sub-models, which are common elements of integrated watershed models used for management, including SWAT, are most prone to this issue as not all of their assumptions can be explicitly defined (Oreskes et al., 1994). Conflicting assumptions resulting from sub-model integration can lead to unreliable simulation outputs that can easily go undetected (Mackay and Robinson, 2000). Undesirable model behavior from conflicting assumptions can also increase model predictive uncertainty, which emerges during model parameterization (Beven, 1995).

One way to potentially minimize conflicts between model components is to ensure that they work with a common data model and have supporting tools for processing input data and preparing parameters. Distributed hydrologic models are now routinely coupled with geographical information systems (GIS), which offer an unprecedented flexibility in the representation and organization of spatial data. Watershed information can be represented in as simple a form as a linked-lumped approach in which the watershed is divided into lumped subwatersheds connected by streams, or as detailed as a fully distributed scheme where each point on a digital terrain surface is explicitly described (Maidment, 1993). The success of the combined GIS and hydrologic model depends in part on how well the aggregation of spatial information in the GIS matches the needs of the modeling components. The effects of data aggregation on model input parameters, such as slope, are well studied. Less is known about conceptual mismatches between the representation of spatial information and the underlying assumptions within model components. For example, hydrologic model components usually require assumptions about the presence or absence of flux and no-flux boundaries, which may not be enforced by the GIS. This mismatch is a model structure issue rather than a data aggregation issue, but the two effects may co-occur.

Some researchers have examined the effects of spatial aggregation of watershed information on model performance. Mamillapalli et al. (1996) show that the accuracy of SWAT improves with increasing detail of watershed representation, within the limits imposed by the available data. Bingner et al. (1997) show that annual fine sediment yield is highly sensitive to the number of sub-basins used to represent a watershed, but runoff is not. FitzHugh and Mackay (2000, 2001) demonstrate that sediment generation prediction is sensitive not only to landcover and soil attributes but also the area of the hydrological response unit (HRU) used to capture sub-basin variability of soil and land cover. They show a nonlinear relationship between sediment generation and the size of the conceptual unit used to characterize a portion of a watershed. These results do not adequately explain why runoff and sediment yield respond differently to spatial data aggregation. However, one emerging hypothesis is that the aggregation effect observed is partly due to model structure, and hence the sub-models for runoff and sediment yield respond differently to the same change in spatial representation. The goal of this paper is to test this hypothesis by focusing on the effect of model structure in relation to watershed discretization. The Soil Water and Assessment Tool (SWAT) Version 2000 (Neitsch et al., 2000, 2001) is used for this study and is described in more detail in the following sections. SWAT predicts sediment and pollutant loadings leaving a watershed and the spatial distribution of soil loss and pollution contributions within a watershed. SWAT is one of the models currently available for use as an aid in implementing the Environmental Protection Agency's Total Maximum Daily Loads.

2. Methodology

2.1. Model description

SWAT is a long-term distributed-parameter model, designed to predict the impact of land management practice in an agricultural watershed (Arnold et al., 1998). It is a public domain model supported by the US Department of Agriculture, Agricultural Research Service at the Grassland, Soil and Water Research

Laboratory in Temple, TX, USA. SWAT version 2000 used in this research is also currently part of the Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) version 3 modeling package, which was developed by the US Environmental Protection Agency to facilitate state and local water quality assessments. In this study, the SWAT Arcview interface (DiLuzio et al., 2001) was used to write SWAT input files from GIS data layers. The topographic parameterization (TOPAZ) digital landscape analysis package (Garbrecht and Martz, 1995) was used to partition the watershed into subwatersheds and derive the stream network.

SWAT estimates surface runoff with the SCS curve number method. Sediment generation from each HRU is calculated by the MUSLE:

$$\text{sed} = 11.8(Q_{\text{surf}} q_{\text{peak}} \times \text{Area})^{0.56} \times K \times C \times P \times \text{LS} \times \text{CFRG} \quad (1)$$

where sed is the sediment generation (metric ton), Q_{surf} is the surface runoff (mm), q_{peak} is the peak runoff rate (m^3/s), Area is the HRU area (km^2), K is the USLE soil erodibility factor, C is the USLE cover and management factor, P is the USLE support practice factor, LS is the USLE topographic factor, and CFRG is the coarse fragment factor. The K, C, P, LS and CFRG factors will be discussed together as the sediment detachment factors and the Q_{surf} and q_{peak} as the runoff energy terms. The peak runoff rate is calculated with the modified rational formula:

$$q_{\text{peak}} = \frac{\alpha Q_{\text{surf}} \times \text{Area}}{3.6 \times t_{\text{conc}}} \quad (2)$$

where q_{peak} is the peak runoff rate (m^3/s), Q_{surf} is the surface runoff (mm), Area is the HRU area (km^2), t_{conc} is the time of concentration (h), and α is the fraction of daily rainfall that occurs during the time of concentration. The value of α is calculated as:

$$\alpha = 1 - \exp[2t_{\text{conc}} \times \ln(1 - \alpha_{0.5})] \quad (3)$$

where t_{conc} is the time of concentration and $\alpha_{0.5}$ is the fraction of daily rain falling in the half-hour highest intensity rainfall. The time of concentration is the sum of overland flow time and channel flow time. Overland flow time is defined as the time it takes for water to travel from the furthest point in the sub-basin to a stream channel and it is computed as:

$$t_{\text{ov}} = \frac{L_{\text{slp}}^{0.6} n^{0.6}}{18\text{slp}^{0.3}} \quad (4)$$

where L_{slp} is the average sub-basin slope length (m), slp is the average sub-basin slope (m/m), and n is the Manning's roughness coefficient. Channel flow time is computed as:

$$t_{\text{ch}} = \frac{0.62 L n^{0.75}}{\text{Area}^{0.125} \text{slp}_{ch}^{0.375}} \quad (5)$$

where L is the channel length (km), n is the Manning's roughness coefficient for the channel, Area is the HRU area (km^2) and slp_{ch} is the channel slope (m/m).

The maximum amount of sediment that can be transported from a channel segment is determined by:

$$\text{conc}_{\text{sed,ch,mx}} = c_{\text{sp}} v_{\text{ch,pk}}^{\text{sp exp}} \quad (6)$$

where $\text{conc}_{\text{sed,ch,mx}}$ is the maximum concentration of sediment that can be transported or the channel carrying capacity (t/m^3), $v_{\text{ch,pk}}$ is the peak channel velocity (m/s), c_{sp} and sp exp are constants defined by the user. The peak channel velocity is defined as:

$$v_{\text{ch,pk}} = \frac{q_{\text{ch,pk}}}{A_{\text{ch}}} \quad (7)$$

where $q_{\text{ch,pk}}$ is the peak flow rate (m^3/s) and A_{ch} is the cross-sectional area of flow in the channel. The peak flow rate is calculated as:

$$q_{\text{ch,pk}} = \text{prf} q_{\text{ch}} \quad (8)$$

where prf is the peak rate adjustment factor, and q_{ch} is the average rate of flow (m^3/s).

2.2. SWAT spatial representation

SWAT operates effectively at two scales: (1) whole watersheds consisting of sub-basins interconnected by a stream network, and (2) sub-basins divided into HRUs consisting of unique combinations of land cover and soils (Srinivasan et al., 2000). HRUs are calculated by treating land cover and soil attributes as statistical distributions, and as such they transform spatial information in the form of soil and land cover maps into nonspatial distributional information. Implicit in the concept of the HRU is the assumption that there is no interaction between HRUs, that is there is no flux from one HRU to another. Sediment and nutrient generation are calculated separately for each HRU and then summed to determine total sub-basin fluxes. The benefits of HRUs are (1) efficient use of nonspatial data sets such as the Agricultural Census and National Resources Inventory (Srinivasan et al., 2000) and (2) minimal computational costs of simulations by lumping similar soil and land use areas into a single unit (Neitsch et al., 2000).

SWAT calculates soil erosion in each HRU with the Modified Universal Soil Loss Equation (MUSLE) (Williams, 1975). MUSLE augments the rainfall energy factor from the Universal Soil Loss Equation (USLE) (Wischmeier and Smith, 1965) with a runoff energy factor. Imbedded in the runoff energy factor are watershed characteristics such as drainage area, stream slope and watershed shape, as these are implicit in the use of unit hydrograph theory (Williams, 1975). The runoff factor represents the energy needed to detach and transport sediment out of a basin. The original MUSLE equation was developed using data from 18 small watersheds and it explained the majority of the variation observed in sediment yield (Williams, 1975). Williams and Berndt (1977) combined MUSLE with the soil conservation service (SCS) runoff curve number technique (USDA-SCS, 1972) and tested the output on 26 watersheds. The combined models explained most of the variation in monthly and annual sediment loadings. It is important to note that both the original and modified MUSLE equations were developed and validated on whole watersheds. The key property of the watershed for MUSLE is that all component sediment fluxes are integrated at the outlet, as assumed by unit hydrograph theory. As such MUSLE assumes a closed simulation unit with a single absorbing flux boundary at the watershed outlet.

This integrative property of watersheds does not strictly hold for the HRU, which is defined on land use and soil attributes rather than flux boundaries. Furthermore, nonadjacent areas can be lumped and modeled as a single HRU, which imposes artificial connectivity that may ignore true no-flux boundaries. For instance, two similar land use areas, occurring in a single sub-basin but separated by a topographic divide, may be merged into a single HRU. This effectively connects these two areas and inflates the effective runoff (and energy) generating area for sediment transport. Parameters in MUSLE, such as the runoff energy factor, assume a topographically based spatial connectivity, which the HRUs clearly do not strictly follow. In addition, the sediment generated from each HRU is summed to estimate the total sub-basin sediment load. The summation is only valid if the sediment generation scales linearly with the HRU area. The violation of sub-model assumptions may have unexpected effects on model output, although these effects may be masked by calibration or other interdependent model processes. The question becomes: to what extent are sediment generation and sediment yield predictions altered by using HRUs and does this matter for watershed scale sediment predictions? To properly address these questions we must consider the mechanisms by which sediment is moved from sub-basins to the streams, and how sediment is routed through the stream network.

2.3. A closer look at MUSLE

Following Eqs. (1) and (2) FitzHugh and Mackay (2000) show that sediment yield varies nonlinearly with HRU area. Specifically, sediment yield is proportional to HRU area raised to the 1.12 power:

$$\text{sed} \propto (\text{Area} \times \text{Area})^{0.56} \tag{9}$$

and is insignificantly offset with a change in time of concentration (FitzHugh and Mackay, 2000). The nonlinear relationship between sediment yield and HRU area means that sediment prediction for each sub-basin cannot be accurately captured with an area-weighted summation over all its HRUs. Consider, for example, the application of MUSLE to a 100 km^2 watershed with spatially uniform soil and land cover. Using one HRU, Eq. (9) yields $(100^{1.12}) = 174$. The same watershed using 10 equal-area HRUs will yield $(10^{1.12}) \times 10 = 132$. Without altering the physical properties of the watershed the sediment yield with 10 HRUs would be about 25% less than with only 1 HRU. In effect, the choice in the number of HRUs alters sediment generation even if the USLE source factors remain the same. The only way to compensate for this aggregation effect when calibrating the model is to adjust USLE factors. However, there is no clear justification for doing this.

To more clearly separate the data aggregation effect associated with sub-basin size from this model structural effect, the MUSLE equation was linearized with respect to HRU area as:

$$\mathrm{sed}' = \frac{(Q_{\mathrm{surf}})^{0.56}(q'_{\mathrm{peak}})^{0.56}(\mathrm{Area})^{1.12}}{\mathrm{Area}^{0.12}} \times K \times C \times P \times \mathrm{LS} \times \mathrm{CFRG} \quad (10)$$

where sed′ is the linearized sediment load from each HRU, and q' is the peak flow rate without the HRU area term. Again, based on FitzHugh and Mackay (2000), time of concentration can be neglected. It is important to note that Eq. (10) should not be considered an alternative to the MUSLE equation, and should not be used in an application setting to predict sediment yield. Instead, this linearization is simply a tool used here to help distinguish the effect of data aggregation from the nonlinear scaling of the Area term on sediment generation.

2.4. Study site and data

The study site was the Pheasant Branch watershed in Dane County, Wisconsin (Fig. 1). The watershed outlet is located at US Geological Survey (USGS) gauging station 05427948, which drains an area of 4730 ha. Pheasant Branch is comprised of two major branches-the 2871 ha North Fork and the 1389 ha South Fork. The North Fork has an average slope of 6% and consists largely of dairy farms with a predominance of corn and alfalfa fields. The South Fork shares similar topography, with an average slope of 5%, but it is more developed, with approximately 15% of the land use classified as residential or commercial, compared to the North Fork which has only about 6%. The major soil types of the study area are silt loams. We chose a simulation period from 1978 to 1981 when the outlet gauge and two nested USGS gauges, no. 05427943, which monitored the North Fork, and no. 05427945, which monitored the South Fork, were simultaneously active. According to the Dane County Regional Planning Commission, during 1976–1977, the watershed outlet gauge (no. 05427948) recorded the highest suspended sediment load per unit area of all rural streams monitored in Dane county (Krug and Gerald, 1986). The sediment and associated nutrient and agricultural chemical

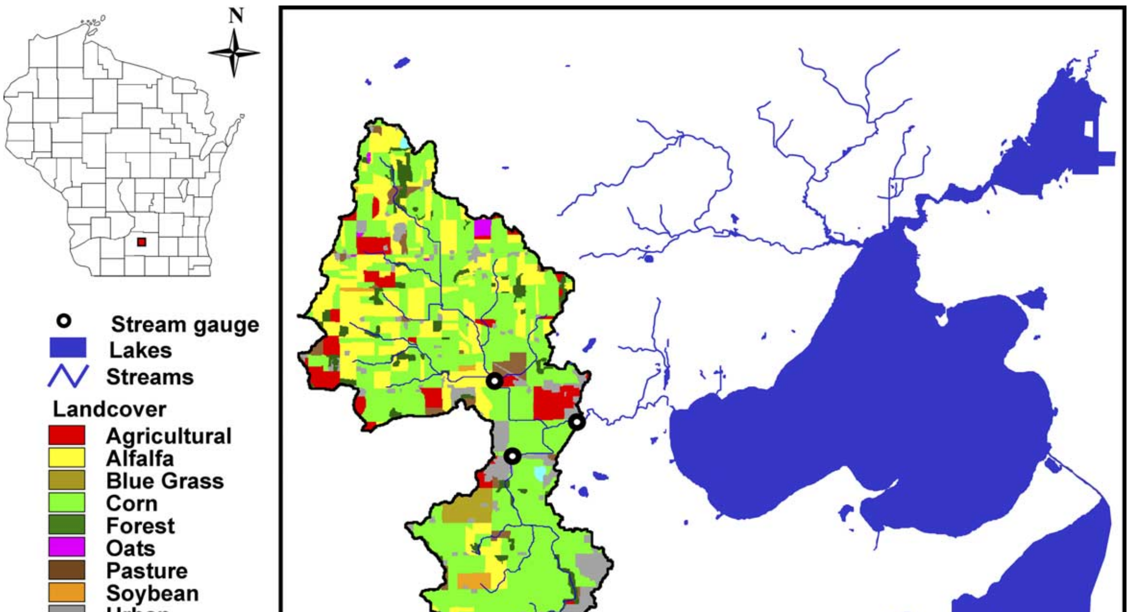


Fig. 1. Study site—Pheasant Branch Watershed, WI.

loading have significant impact on the water quality of Lake Mendota down stream (Wisconsin DNR, 1997).

Daily stream flow data were available for all three gauges. Daily sediment yield were available for the South Fork and watershed outlet gauges. Sediment yield were only measured during storm periods for the North Fork gauging station, and so monthly values were extrapolated using stream flow data. The total annual sediment yield was checked against published values in the USGS Water-Resources Investigations Report (85-4068), which were estimated from the instantaneous stream discharge. Daily precipitation and minimum and maximum temperature data were downloaded from the Charmany Farm National Weather Service Cooperation station (no. 471416), located about 1 km southeast of the watershed. Three GIS input layers were used to parameterize SWAT: a 1:15,840 digital soil data derived from the Dane County soil survey; a Dane County digital elevation model (DEM), gridded at 11.5 m; and a land cover map derived for the present study from 1:660 color aerial photographs, taken in July of 1979, obtained from the Farm Service Agency. Part of the main channel network had been straightened and dredged for agricultural drainage purposes (North Fork Pheasant Branch Watershed Committee, 1999). Because the engineered channels are too narrow to be resolved in the DEM, a 1:24,000 stream network map, produced by the Department of Natural Resources, was merged with the DEM to ensure the proper representation of the main stream channels.

2.5. Watershed configuration and model simulation

The Pheasant Branch watershed was partitioned into four different watershed delineations of 5, 25, 95, and 179 sub-basins, respectively, following procedures detailed in FitzHugh and Mackay (2000). Each watershed delineation was further divided into different numbers of HRUs by setting the landcover and soil thresholds as shown in Fig. 2. The thresholds were measured in terms of percent area of each attribute within a sub-basin. Minor land cover types below the landcover threshold were eliminated and within the selected land cover types, minor soil types were eliminated according to the soil threshold. Unlike the threshold option, which selects the dominant soil type within the selected land cover, the dominant land cover and soil option chooses the two attributes independently, so that the dominant soil for the subwatershed is not necessarily the dominant soil associated with the dominant land cover (DiLuzio et al., 2001; Mamillapalli et al., 1996). Simulations were performed with the original SWAT 2000 source code and with the modified code containing the linearized MUSLE (Eq. (10)), for

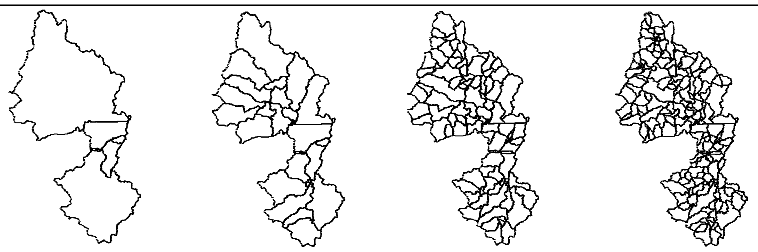
		5 subbasin	**25 subbasin**	**95 subbasin**	**179 subbasin**
Subbasins					
Threshold:					
Land Cover	**Soil**	**Number of HRUs**	**Number of HRUs**	**Number of HRUs**	**Number of HRUs**
Dominant		**5**	**25**	**95**	**179**
20%	**15%**	**13**	**75**	**350**	**675**
10%	**10%**	**23**	**181**	**626**	**1095**
5%	**5%**	**79**	**769**	**998**	**1569**

Fig. 2. Watershed configurations. The Pheasant Branch watershed was partitioned into four different watershed delineations with 5, 25, 95, and 179 sub-basins. Each watershed delineation was further divided into different number of HRUs by setting the landcover:soil threshold as: dominant:dominant, 20:15, 10:10 and 5:5%.

the 16 watershed configurations, using the same input files.

The original SWAT 2000 model was calibrated at the watershed outlet using the watershed configuration with 95 sub-basins and 626 HRUs. The calibrated model parameters were used for all simulations. Fig. 3 shows the calibrated monthly stream flow and sediment yield for the simulated period. The monthly stream flow, on average, is over predicted. Most of the over prediction occurred during July 1978, March 1979, and September 1981. The modified coefficients of efficiency (MCOE) (Legates and McCabe, 1999) for the watershed outlet, North Forth and South Fork were 0.368, 0.104 and −0.834, respectively. The MCOE for the average monthly sediment yield were 0.391, −0.825, and 0.164. Details on calibrating SWAT for both runoff and sediment yield can be found in FitzHugh and Mackay (2000). Although accurate predictions were not the goal of this study, the calibration served to establish a realistic base line for making relative comparisons of the model output. Sections 3 and 4 will focus mainly on the sediment yield and sediment generation predictions, since the annual stream flow is not greatly affected by watershed partitioning (Fig. 4). This is consistent with results reported in Bingner et al. (1997) in a different study area and FitzHugh and Mackay (2000) in the same study area for a different set of years.

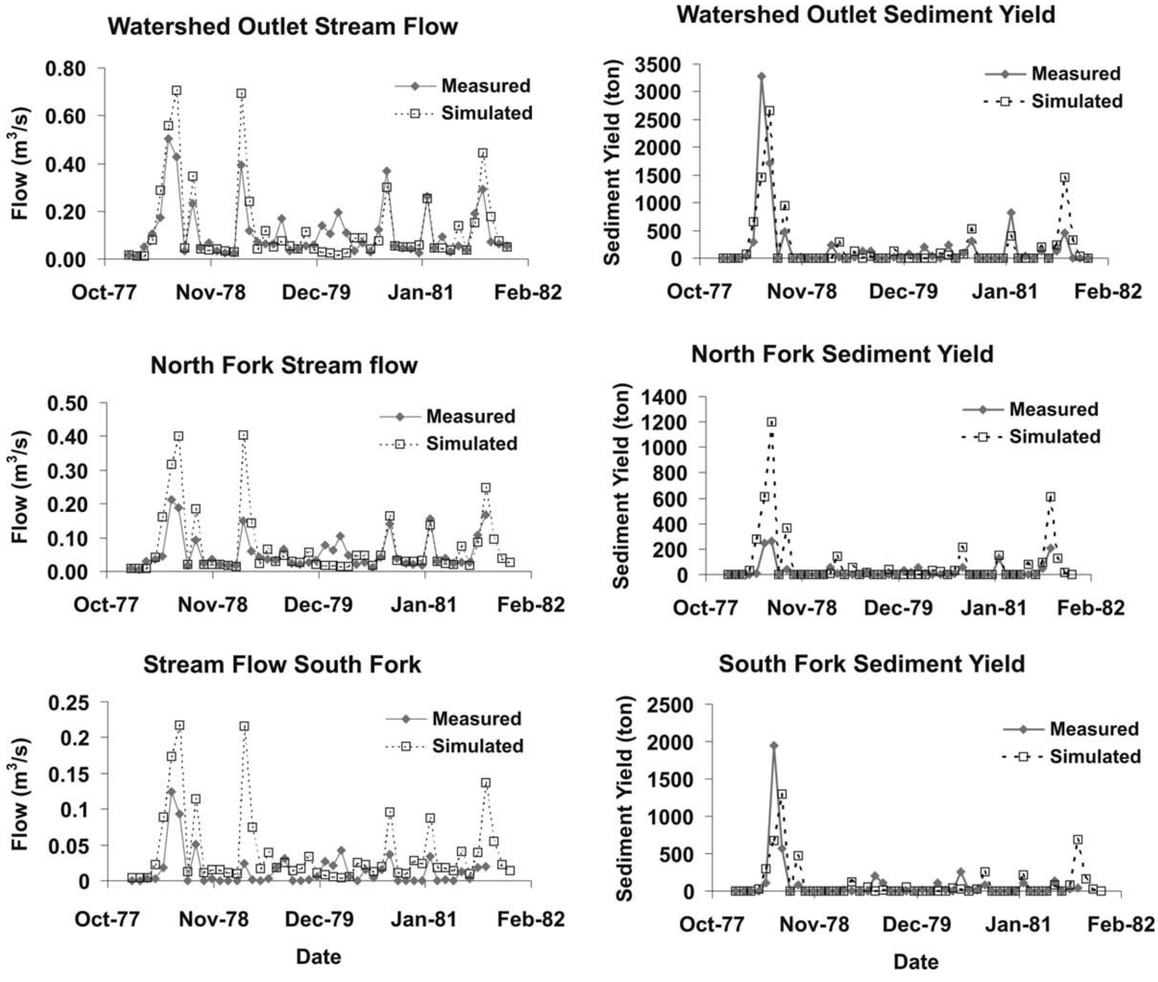


Fig. 3. Monthly average measured and calibrated stream flow and sediment yield for 1979–1981.

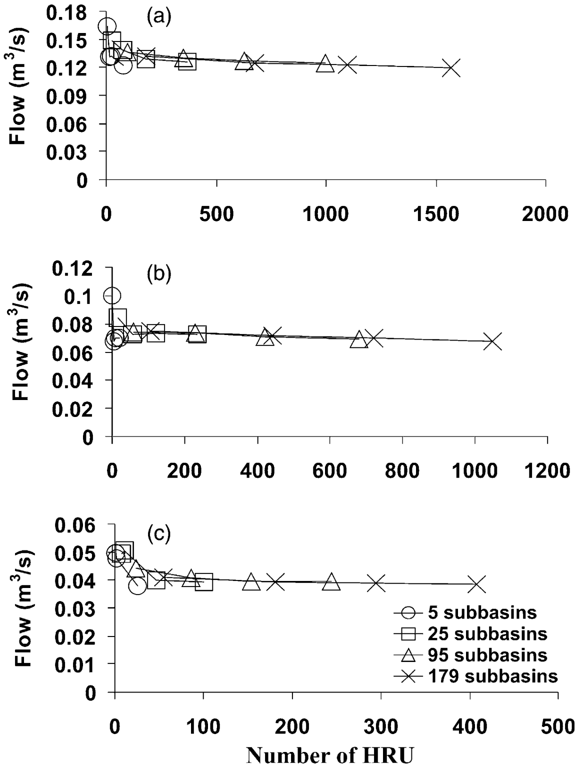


Fig. 4. Average annual stream flow at: (a) Pheasant Branch outlet gauge, (b) North Fork gauge, and (c) South Fork gauge.

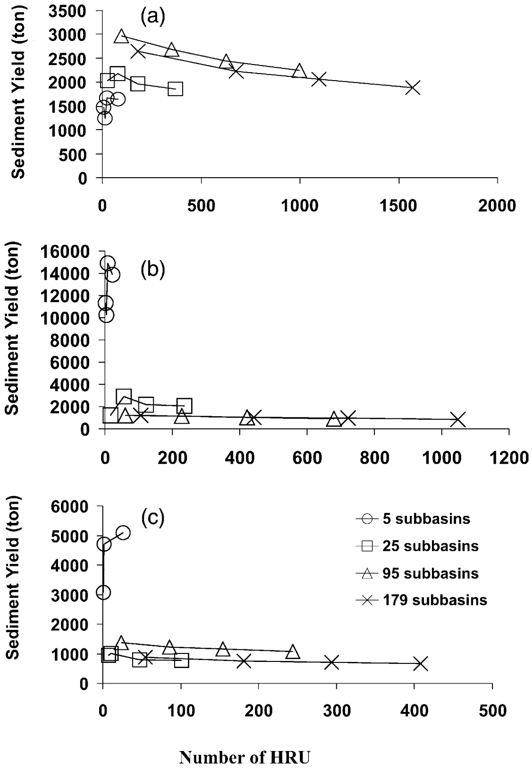


Fig. 5. Average annual sediment yield for: (a) Pheasant Branch, (b) North Fork, and (c) South Fork.

3. Results

3.1. Sediment yield

Sediment yield is the amount of sediment transported out of a watershed or subwatershed. This value is often used for model evaluation and calibration because it can be compared against readily available measured data sets, such as those published by the USGS. Sediment yield reflects the integrated response of sediment generation processes (at sub-basin scale) and stream processes (at watershed scale). Fig. 5 shows the average annual sediment yield for the 16 watershed configurations at the three gauges. The North Fork and South Fork exhibit similar trends across levels of partitioning. For the North Fork, the five sub-basin configurations produce sediment yields ranging from 11,400 to 14,900 t, but as the sub-basin number increases to 25, the sediment yield drops sharply to between 2000 and 2900 t. Increasing the number of sub-basins to 95 and 179 varies sediment yield by less than 400 t. Similarly, for the South Fork, the five sub-basin configurations deliver sediment ranging from 3100 to 5100 t, which drops to between 1400 and 700 t when the number of sub-basins increases to 179. At the watershed outlet, sediment yields increase as the number of sub-basins increases from 5 to 95 and then decreases slightly at 179 sub-basins. With the exception of the three coarsest watershed configurations, sediment yield decreases as the number of HRUs increases within a given watershed delineation.

3.2. Sediment generation

Sediment generation is the amount of overland soil loss due to water erosion. Based on Eq. (10) we expected sediment generation calculated from the SWAT to exhibit greater variance than the linearized model, since the original model output is influence by

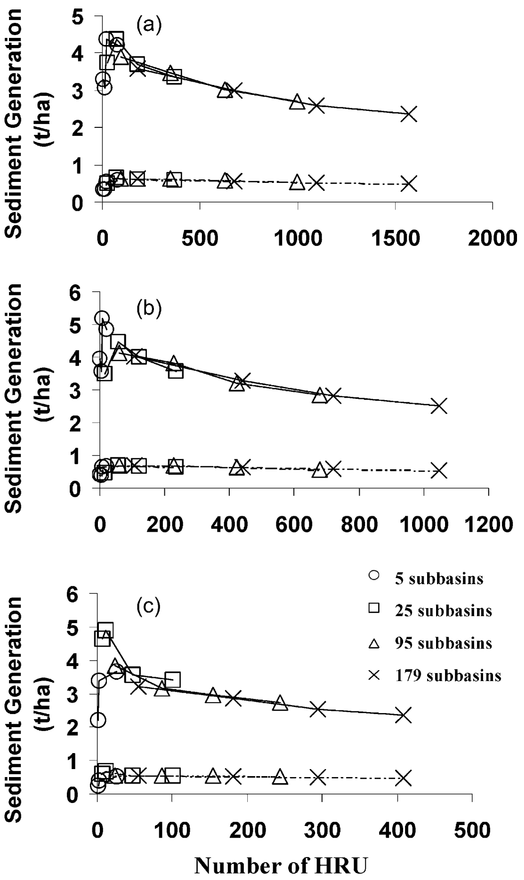


Fig. 6. Average annual sediment generation for: (a) Pheasant Branch, (b) North Fork, and (c) South Fork. The linearized model simulations are in dashed lines and the original model simulations are in solid lines.

both data aggregation effect and area-energy effect while the linearized model is only affected by the data aggregation.

Fig. 6 shows the change in annual average sediment generation versus change in the number of HRUs and sub-basins. All simulations show an initial increase in sediment generation at the coarse watershed configurations, and then a decrease as the number of HRUs increase to 1569. For the entire Pheasant Branch watershed, from the watershed configuration that generated the maximum sediment output to the most detailed watershed configuration, the sediment generation predicted by the original model decreases by 46%, while the sediment generation predicted by the linearized model decreases by only 25%. Similarly, for the North Fork, the sediment generation decreases by 45% with the original model and 23% with the linearized model. For the South Fork sediment generation decreases by 51 and 29% for the original and linearized models, respectively.

4. Discussion

4.1. Sediment yield

Channel transport processes are modeled differently for the internal and external stream links. This explains the large decrease in yield observed between the five sub-basin and the 25 sub-basin delineations, at the North Fork and South Fork gauges. At the five sub-basin level, the two nested gauges are each located at the outlet of an external channel link, in which there is no modeled sediment deposition. All the sediment generated overland is transported out of these reaches and into the next higher order stream, where its fate depends on the channel carrying capacity of that stream. As the number of sub-basins increases from 5 to 25, the nested gauges become associated with internal links. Consequently, the sediment routing algorithm allows excess sediment to be deposited up-stream of the gauges and a large decrease in sediment yield is observed. The watershed outlet does not experience large changes in sediment yield, because the outlet is located on an internal channel link in all watershed delineations. The different treatments among internal and external stream channels are due to the fact that channel transport processes, such as deposition and re-suspension are implicit in MUSLE. Subwatersheds with external channel links satisfy the watershed assumption of MUSLE and thus do not require additional routing. For higher order stream channels routing is needed to make connections between the sub-basins.

However, the application of multiple HRUs within sub-basins ignores the routing processes needed to link each simulated area. MUSLE is calculated at the HRU level and not directly at the sub-basin level. The five sub-basin configurations over predict the average sediment yield by approximately 9800–14,400 t for the North Fork, and approximately 1900–3900 t for the South Fork. The overestimation of sediment yield is probably the result of the following two factors or a combination of: (1) when one HRU is used for each sub-basin, the aggregated parameter input used in the MUSLE does not adequately capture the effects of the various land-use and soils in the sub-basins; (2) when multiple HRUs are applied in each sub-basin to compensate for the data aggregation effect, the sum of the runoff energy from the individual HRUs does not adequately describe the transport processes for the entire sub-basin. During model calibration, the USLE *P*-factor was adjusted to lower the sediment generation, but within a realistic range of *P*-values, the sediment generation remained over predicted.

The subtler variations in sediment yield can be attributed to the source-limited subwatersheds and the geomorphological representation of the subwatersheds in which the gauges are located. The sub-basin size and shape influence channel characteristics, which are used to calculate the carrying capacity of each reach. A watershed is described here as transport-limited, if more sediment is being generated than the streams can transport; it is source-limited if the streams can transport more sediment than is generated (Keller et al., 1997; FitzHugh and Mackay, 2000, 2001). Fig. 7 shows sub-basins that have net sediment deposition, or, in other words, were transport-limited during the study period. Deposition occurs mostly during large storm events when larger amounts of soil are eroded. Sub-basins that are transport-limited act effectively as bottlenecks, so that sediment yield down stream is more sensitive to the carrying capacity of the upstream reaches and less sensitive to the actual amount of sediment generated in each sub-basin. For source-limited sub-basins, sediment yield output is affected more by sediment generation and hence by the HRU parameters.

For the North Fork, sediment yield is slightly higher for the 25 sub-basin configurations than the 95 and 179. The decrease is due to the internalization of a single reach draining into the outlet channel. Sediment yield is stable for the 95 and 179 sub-basin configurations, because the shape of the outlet sub-basin remains fixed and it drains numerous transport-limited sub-basins. The South Fork also drains transport-limited sub-basins, but the shape of the outlet sub-basin also changes with watershed partitioning, and hence there is a slight decrease in output stability. For the Pheasant Branch watershed outlet, the sediment yield increases slightly as the number of sub-basins increases from 5 to 95. This occurs because there is a decrease in sediment deposition and an increase in the number of source-limited sub-basins surrounding the outlet. The increase in the number of source-limited sub-basins means that all sediment generated in these areas is contributing to the outlet.

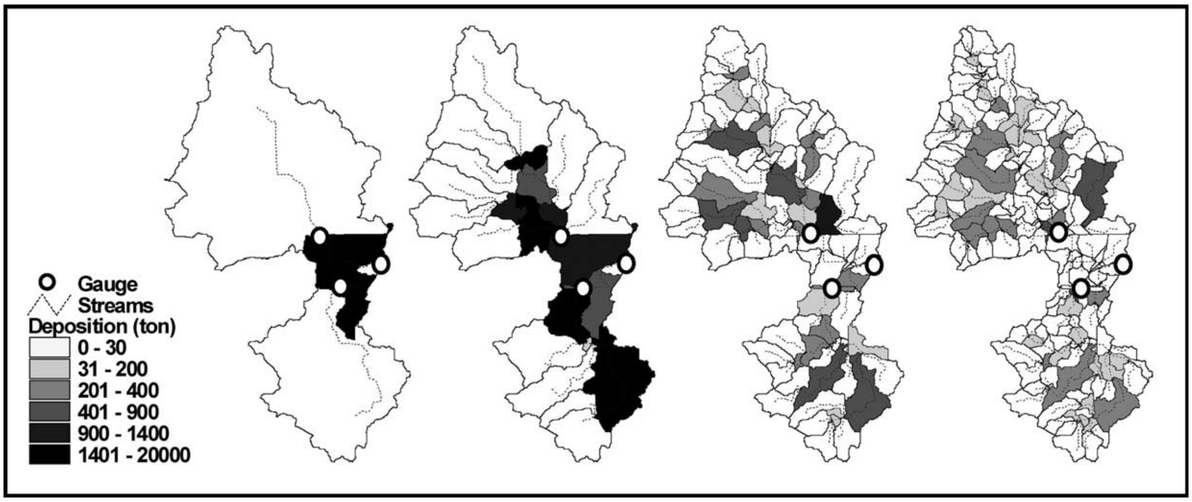


Fig. 7. Net average annual sediment deposition for the four watershed delineations with the landcover: soil threshold at 5:5%. There is no deposition in the external channel links and the spatial distribution of the transport-limited reaches is affected by the watershed delineation.

As the number of sub-basins increases to 179, there is a small decrease in yield due to a decrease in sediment generation and flow.

The behavior of sediment yield predictions is also very sensitive to calibration and model settings. The degree to which a watershed is transport-limited or source-limited depends partially on the user-defined parameters; c_{sp} and sp exp, in the channel carrying capacity equation, and the peak rate adjustment factor in Eq. (8). Sediment yield can be adjusted further by changing the channel erosion parameters, such as channel erodibility factor and channel cover factor. The default value for channel erodibility factor (0.0) was used for this study because no measurements were available to estimate the parameter. The default value represents nonerosive channels. If the channels are allowed to erode, then SWAT re-entrains sediment when the sediment concentration is less than the channel carrying capacity; sediment accretion occurs otherwise. The channel processes have the overall effect of simplifying the calibration process, but they also mask problems in modeled overland erosion processes.

4.2. Sediment generation

There are three major interacting components controlling the behavior of sediment generation in SWAT: MUSLE runoff energy factor, detachment factors, and the HRU area. The effect of the HRU area is shown by the comparison between the original MUSLE and the linearized model as summarized in Fig. 6. Both models were run using the same input files. The observed difference in the output is caused by the nonconservative summation of the MUSLE output from the HRUs. In the coarse watershed configurations (5–75 total HRUs), both models exhibit a general increase in sediment generation with an increase in HRUs. As the number of HRUs is further increased, the original model output continues to show a marked decrease in sediment generation while the linearized model shows relative stability. Beyond 626 HRUs, the rate of decrease in sediment yield for the original model is reduced, and the two models again exhibit a similar gentle decreasing trend. As to be expected, the power function (i.e. HRU area raised to a power of 1.12) in MUSLE has greater impact when the area term is large. In the coarse watershed configurations, which contain the largest HRUs, sediment generation is dominated by the time of concentration. This is not entirely in agreement with the conclusions of FitzHugh and Mackay (2000), but it does narrow the range of basin sizes where time of concentration is a critical factor. This will be discussed further below. Between 75 and 626 HRUs the HRU area plays a more important role. In this range, a significant part of the observed variance in the model output can be attributed to model structure and does not reflect the changes in watershed information content brought about through data aggregation.

Fig. 8 shows the effect of the MUSLE runoff energy and detachment factors on sediment generation. Fig. 8a shows the average area-weighted sediment detachment factors for each watershed

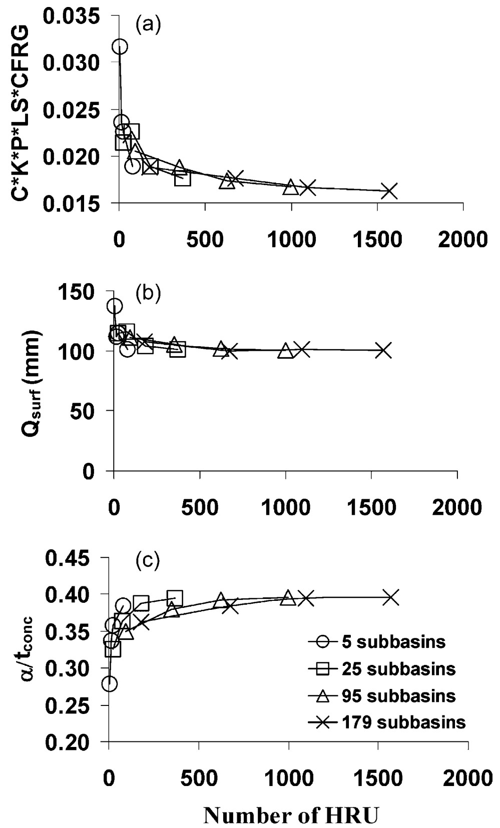


Fig. 8. Average MUSEL run-off energy and detachment factors, plotted against number of HRUs: (a) sediment detachment factors, which contains the USLE K, C, P, LS factors plus CFRG factor; (b) average annual surface run-off; and (c) average α/t_{conc} from the peak run-off rate equation.

configuration. The change in the distribution of the detachment factors is the direct result of data aggregation associated with watershed partitioning. Fig. 8b shows the change in average surface runoff across the different watershed configurations. Surface runoff decreases slightly from 5 to 23 HRUs but remains fairly stable as the number of HRUs is increased to 1569. This reflects the small changes in the average runoff CN, which decreases from 73 to 69, and then to 66, respectively. Fig. 8c shows the estimated ratio between α and t_{conc} in the peak run-off equation. The ratio is estimated using basin average t_{conc}. Since the same weather data was used for all simulations, a constant is used for $\alpha_{0.5}$. Because α also contains t_{conc}, Fig. 8c effectively expresses the relationship between the time of concentration in MUSLE and the level of watershed partitioning. The ratio between α and t_{conc} is the only factor in MUSLE that exhibits the same initial increasing trend observed in the sediment generation.

Time of concentration is affected by data aggregation because it uses sub-basin average topographical parameters, but it is also influenced by the application of spatial attributes associated with the HRUs. Within t_{conc}, the channel time of concentration is also nonlinearly related to the HRU area. As the HRU area increases, the channel time of concentration decreases, and sediment generation increases. The effect of this area term on sediment generation cannot be shown as easily as the other area terms in MUSLE. The problem it creates is similar, in that there is a conceptual violation of the model assumptions and sediment generation is forced to scale linearly with area within sub-basins. In addition to responding to HRU area, sediment generation is also influenced by the channel length parameter inside t_{conc}. The channel length for each HRU is calculated by multiplying the channel length from the most distant point in the sub-basin to the sub-basin outlet by the fraction of the respective HRU area in the sub-basin. As the channel lengths increase, t_{conc} increases, and the sediment generation decreases. The use of HRU channel length raises two issues: (1) because HRUs are nonspatial, the channel length parameter does not have a direct physical interpretation; and (2) MUSLE takes the channel length to a power factor, which means the sum of the HRU channel length is not equal to the sub-basin channel length.

Model structural effects such as those associated with the integration of MUSLE and HRU are difficult to fully quantify due to a lack of overland soil loss measurements. Inferences made with available data, such as sediment yield, are unreliable because sediment generation behavior can easily be obscured under a combination of model calibration and channel processes. It may be argued that the accuracy of internal values, like sediment generation, may not be critical if the ultimate goal is to provide satisfactory estimates of watershed response. Nevertheless, resolving the physics in the model structure will help expand its applicability to producing distributed outputs for targeting site-specific management practice, and for extending its use to ungauged watersheds (Grayson et al., 1992).

There exist a number of alternative water erosion models, for example, Agricultural Policy/Environmental eXtender (APEX) (Williams et al., 1998), Water Erosion Prediction Project (WEPP) (Laflen et al., 1997), and Revised Universal Soil Loss Equation (RUSLE) (Renard et al., 1991), which could provide alternative sub-basin sediment generation for SWAT. However, these models require higher resolution data sets that are not practical for large watersheds. Based on the results of this paper, we suggest that, for SWAT, the maximum possible number of sub-basins supported by the resolution of the DEM be used for the simulations, and that one HRU be applied per sub-basin. This approach can easily be implemented with the GIS interface available with SWAT, and it avoids the problem of having to sum up nonlinear terms used in the MUSLE, but retains the ability to represent soil and land cover variability. In addition, over time, if the landscape becomes more fragmented from land use changes, the same set of sub-basins can be used so that consistent comparisons can be made. For example, if corn farmers in a watershed decide to diversify their crops to include cabbage, lima beans, sugar beet, and potato, the change in land use will increase the number of HRUs used to represent the watershed. If they retain the same farming practices as before, then in theory, there should be no change in the sediment generation predictions from MUSLE, since all of the different crops have the same detachment factors and the CNs remain the same as before. But because there is an increase in the number of HRUs, an artificial decrease in overland sediment erosion will be predicted. This means that output comparisons made before and after the crop change will be misleading. The same issue can also occur with inter-watershed comparison studies, where two watersheds that are relatively similar in terms of MUSLE input parameters could potentially produce very different sediment outputs, solely due to the difference in HRU configurations.

5. Conclusions

The simulated results for the Pheasant Branch watershed show that model structure plays an important role in the model response to spatial data aggregation. The integration of MUSLE and HRU are conceptually incompatible because MUSLE, a watershed response model, is being applied to HRUs, which are not watersheds but nonspatial subunits of a watershed. Our original question pertaining to the extent of this effect on sediment generation and sediment yield predictions cannot be fully answered without overland erosion measurements. However, this paper has shown that one direct consequence of the violation of sub-model assumption is the inability of SWAT to conserve sediment generation among different levels of watershed partitioning. Variations observed in sediment generation outputs at different watershed discretizations are due to the averaging of the topographic attributes, changes in the statistical distribution of the sediment detachment factors, and the linear scaling of the MUSLE outputs from the HRUs to the sub-basin level. The latter neglects the nonlinear relationship between the MUSLE runoff energy and sediment generation, thus causing an artificial decrease in sediment generation when the HRU area is decreased. Furthermore, the treatment of sediment transport in first order streams may not be appropriate when multiple HRUs are used. There is no physical connection between the sediment generated at each of the HRUs and the sub-basin total. These findings have direct implications for the interpretation of model output for both modelers attempting to optimize watershed representation and those making comparison between watersheds of contrasting levels of heterogeneity. These issues are inevitable with a highly flexible tool such as SWAT, but it is possible to rely on sub-basins as the means of representing spatial variability of soil and land cover information and thereby avoiding the spatial mismatches associated with HRUs.

Acknowledgements

Primary funding support for this research was by USDA CSREES Hatch (Project no. WIS04453), and an EPA Nutrient Science STAR grant (R830669). Further funding by NASA (NAG5-6435) for the Integrated Remote Sensing Resource Center at UW-Madison is also acknowledged. Susan Neitsch at the Blackland Research Center, Texas Agricultural Experiment Station in Temple, TX provided assistance with the UNIX version of SWAT 2000, in which we made code modifications.

References

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

Beven, K., 1995. Linking parameters across scales: subgrid parameterizations and scale dependent hydrological models. Hydrological Processes 9, 509–525.

Bingner, R.L., Garbrecht, J., Arnold, J.G., Srinivasan, R., 1997. Effect of watershed subdivision on simulation runoff and fine sediment yield. Transactions of the ASAE 40(5), 1329–13335.

DiLuzio, M., Srinivansan, R., Arnold, J., 2001. Arcview Interface for SWAT2000 User's Guide, Blackland Research Center, Texas Agricultural Experiment Station, Temple, TX.

FitzHugh, T.W., Mackay, D.S., 2000. Impacts of input parameter spatial aggregation on an agricultural nonpoint source pollution model. Journal of Hydrology 236(1–2), 35–53.

FizHugh, T.W., Mackay, D.S., 2001. Impact of subwatershed partitioning on modeled source- and transport-limited agricultural nonpoint source pollution model. Journal of Soil and Water Conservation 56(2), 137–143.

Garbrecht, J., Martz, L.W., 1995. TOPAZ: An automated landscape analysis tool for topographic evaluations, drainage identification, watershed segmentation and subcatchment parameterization: Overview. ARS Pub. No. NAWQL 95-1. US Department of Agriculture, Agricultural Research Service, Durant, OK, USA.

Grayson, R.B., Moore, I.D., Thomas, McMahon, A., 1992. Physically based hydrologic modeling 2. Is the concept realistic? Water Resources Research 26(10), 2659–2666.

Joao, E.M., Walsh, S.J., 1992. GIS implications for hydrologic modeling: simulation of nonpoint pollution generated as a consequence of watershed development scenarios. Computer Environment and Urban Systems 16, 43–63.

Keller, E.A., Valentine, D.W., Gibbs, D.R., 1997. Hydrological response of small watersheds following the Southern California painted cave fire of June 1990. Hydrological Processes 11(4), 401–414.

Krug, W.R., Gerald, L.G., 1986. Effects of urbanization on streamflow, sediment loads, and channel morphology in Pheasant Branch Basin near Middleton, Wisconsin, U.S. Geological Survey. Water-Resources Investigations Report, 85–4068.

Laflen, J.M., Elliot, W.J., Flanagan, D.C., Meyer, C.R., Nearing, M.A., 1997. WEPP-prediction water erosion using a process-based model. Journal of Soil and Water Conservation 52(2), 96–102.

Legates, D.R., McCabe, G.J. Jr., 1999. Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model validation. Water Resources Research 35(1), 233–241.

Lohani, V., Kibler, D.F., Chanat, J., 2002. Constructing a problem solving environment tool for hydrologic assessment of land use change. Journal of the American water Resources Association 38(2), 439–452.

Mackay, D.S., Robinson, V.B., 2000. A multiple criteria decision support system for testing integrated environmental models. Fuzzy Set and Systems 113, 53–67.

Maidment, D.R., 1993. In: Goodchild, M.F., Parks, B.O., Steyaert, L.T. (Eds.), Environmental Modeling with GIS, Oxford University Press, Oxford, pp. 147–167.

Mamillapalli, S., Srinivasan, R., Arnold, J.G., Engel, B.A., 1996. Effect of Spatial Variability on Basin Scale Modeling, In Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, New Mexico, January 1996, National Center for Geographic Information and Analysis, Santa Barbara, CA.

Neitsch, S.L., Arnold, J., Kiniry, J.R., Williams, J.R., 2000. Soil and Water Assessment Tool User's Manual Version 2000, Blackland Research Center, Texas Agricultural Experiment Station, Temple, TX.

Neitsch, S.L., Arnold, J., Kiniry, J.R., Williams, J.R., 2001. Soil and Water Assessment Tool Theoretical Documentation Version 2000, Blackland Research Center, Texas Agricultural Experiment Station, Temple, TX.

North Fork Pheasant Branch Watershed Committee, March 1999. http://www.pheasantbranch.org

Oreskes, N., Shraderfrechette, K., Belitz, K., 1994. Verification, validation, and confirmation of numerical-models in the earth-sciences. Science 263(5147), 641–646.

Renard, K.G., Foster, G.R., Weesies, G.A., Porter, J.P., 1991. RUSLE: revised universal soil loss equation. Journal of Soil and Water Conservation 46(1), 30–33.

Srinivasan, R., Muttiah, R.S., Dyke, P.T., Walker, C., Arnold, J., 2000. Hydrologic Unit Model for the United States (HUMUS), Texas Agricultural Experiment Station, Blackland Research Center, Temple, TX, http://srph.brc.tamus.edu/humus/html/paper.html.

USDA-SCS, 1972. National Engineering Handbook, Hydrology Section 4, Chapter 4–10, US Department of Agriculture, Soil Conservation Service, Washington, DC.

Williams, J.R., 1975. Sediment-yield prediction with universal equation using runoff energy factor. In Present and prospective technology for prediction sediment yield and sources: Proceedings of the sediment yield workshop, USDA Sedimentation Lab, Oxford, MS, November 28–30, 1972. ARS-S-40, pp. 244–252.

Williams, J.R., Berndt, H.D., 1977. Sediment yield prediction based on watershed hydrology. Transactions of ASAE 20(6), 1100–1104.

Williams, J.R., Arnold, J.G., Srivinasan, R., Ramanarayanan, T.S., 1998. APEX: a new tool for predicting the effects of climate and CO2 changes on erosion and water quality. In: Boardman, J., Favis-Mortlock, D. (Eds.), Modeling Soil Erosion by Water, NATO ASI Series I: Global Environmental Change, NATO ASI Series I: Global Environmental Change, vol. 55. Springer, Berlin, pp. 441–449.

Wilson, J.P., 1996. GIS-based Land Surface/Subsurface Modeling: New Potential for New Models?, NCGIA Third International Conference/Workshop on Integrating GIS and Environmental Modeling, January 21–25, 1996, Santa Fe, New Mexico.

Wischmeier, W.H., Smith, D.D., 1965. Predicting Rainfall-Erosion Losses from Cropland East of the Rocky Mountains, Agriculture Handbook 282. USDA_ARS.

Wisconsin, D.N.R., 1997. Nonpoint Source Control Plan for the Lake Mendota Priority Watershed Project: Project Summary. Publication WT-481-97, Wisconsin Department of Natural Resources, Madison, WI, http://www.dnr.state.wi.us/org/water/wm/nps/plans/mensum/mendota.htm.