

Effects of detailed soil spatial information on watershed modeling across different model scales

Trevor Quinn^{a,1}, A.-Xing Zhu^{b,c,*}, James E. Burt^{c,2}

^a Michael Baker Corporation, PA, USA

^b State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Building 917, Datun Road, An Wai, Beijing 100101, China

^c Department of Geography, University of Wisconsin-Madison, 550 North Park Street, Madison, WI, USA

Received 18 October 2004; accepted 6 June 2005

Abstract

Hydro-ecological modelers often use spatial variation of soil information derived from conventional soil surveys in simulation of hydro-ecological processes over watersheds at mesoscale (10–100 km²). Conventional soil surveys are not designed to provide the same level of spatial detail as terrain and vegetation inputs derived from digital terrain analysis and remote sensing techniques. Soil property layers derived from conventional soil surveys are often incompatible with detailed terrain and remotely sensed data due to their difference in scales. The objective of this research is to examine the effect of scale incompatibility between soil information and the detailed digital terrain data and remotely sensed information by comparing simulations of watershed processes based on the conventional soil map and those simulations based on detailed soil information across different simulation scales. The detailed soil spatial information was derived using a GIS (geographical information system), expert knowledge, and fuzzy logic based predictive mapping approach (Soil Land Inference Model, SoLIM). The Regional Hydro-Ecological Simulation System (RHESys) is used to simulate two watershed processes: net photosynthesis and stream flow. The difference between simulation based on the conventional soil map and that based on the detailed predictive soil map at a given simulation scale is perceived to be the effect of scale incompatibility between conventional soil data and the rest of the (more detailed) data layers at that scale. Two modeling approaches were taken in this study: the lumped parameter approach and the distributed parameter approach. The results over two small watersheds indicate that the effect does not necessarily always increase or decrease as the simulation scale becomes finer or coarser. For a given watershed there seems to be a fixed scale at which the effect is consistently low for the simulated processes with both the lumped parameter approach and the distributed parameter approach.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Geographic information systems; Remote sensing; Scale; Soils; Hydrology; Environmental modeling; Model bias; SoLIM

1. Introduction

Increasingly geographic information systems (GIS) are used to parameterize the landscape for hydro-ecological models operating at the mesoscale level (10–100 km²). One of the emergent stumbling blocks in the integration of GIS and watershed models is the problem of combining data sets of varying levels of spatial details (Ehleringer and Field, 1993; Blöschl, 1998;

* Corresponding author. Tel.: +86 10 6488 9681/+1 608 262 0272; fax: +1 608 265 3991.

E-mail addresses: tquinn@mbakercorp.com (T. Quinn), axing@lreis.ac.cn, axing@geography.wisc.edu (A.-X. Zhu), jburt@geography.wisc.edu (J.E. Burt).

¹ Fax: +1 703 960 9125.

² Tel.: +1 608 263 6640; fax: +1 608 265 3991.

Thieken et al., 1999; Western and Blöschl, 1999; Zhu, 2005). In other words, spatial data needed for hydro-ecological models at the mesoscale level are not always available at that level of spatial detail. For example, information on the spatial variation of soils are often derived from the *conventional* 1:24,000 scale soil maps (which often have a minimum mapping unit between 2.5 and 5 ha), while information on the spatial variation of vegetation are often derived from remotely sensed imagery at 30 m resolution (or finer). Consequently, the level of spatial detail in conventional soil maps are often much coarser than the level of spatial detail in the corresponding vegetation map. Thus the levels of spatial details between the two maps are incompatible. This incompatibility is referred to as scale incompatibility because it is a result of the difference in map scale and/or spatial resolution among the data layers.

When using data sets of varying scales, researchers are often faced with the question of ‘... what is the appropriate scale at which to simulate hydro-ecological processes over mesoscale watersheds?’ (Band, 1993; Wolock and Price, 1994; Zhang and Montgomery, 1994; Van Gardingen et al., 1997; Koren et al., 1999; Georges and Chen, 2002; Haddeland et al., 2002; Ranjan and Wurbs, 2002). Since GIS allows rapid processing and parameterization of spatial data, it is tempting to operate the model at the scale of the most detailed data layers involved in the modeling effort, even if other data layers do not match that scale.

It is important to note that this scale incompatibility can cause the spatial co-variation of model parameters to be characterized incorrectly (Zhu, 2000), and thus lead to incorrect model output and result interpretation. One common example of scale incompatibility results from the use of soil data in hydro-ecological models, which require variables about local soil water storage capacity and transmissivity. Soil information are typically derived from conventional polygon-based soil maps, with a scale likely to be substantially lower than that of other data used by the model, such as terrain data derived from standard digital elevation models (DEMs). Modelers often overlay (spatially combine) high resolution (10–30 m) topographic and vegetation data with generalized soil information derived from the conventional soil survey (1:24,000) to estimate the co-variation of terrain, vegetation, and soil conditions over space. This overlay can result in poor local correspondence between key soil variables such as available moisture and other model parameters such as leaf area index or solar insolation. In such cases, the scale of the original soil survey prevents the parameterization of small areas where soil properties deviate from those of

the larger, surrounding soil body. Band and Moore (1995) identified scale incompatibility as a potential problem in extending hillslope hydrologic models to regional scales.

Some of the effect of scale incompatibility among data layers may be dependent on the scale at which the model is run. In this paper, we refer to this as the *simulation scale*. Data sets are not often available at a resolution that will permit realistic process simulations at very fine simulation scales (meters or less). As a result, watershed models that simulate these processes over large spatial extents must find a way to describe environmental conditions using effective parameters rather than directly observed values. For some models, these parameters are produced by partitioning the landscape into *hillslope* units and aggregating the spatial data within each hillslope unit. The degree to which landscape heterogeneity is generalized and aggregated by the model can be thought of as the size of these hillslope units which in turn can be thought of as the model simulation scale.

Hillslope partitioning is one of the more flexible methods of varying simulation scale. Hillslopes – the areas in a watershed that drain to each stream link on either bank side – capture much of the spatial variation of incident short-wave radiation and seem particularly suitable for landscape parameterization in mountainous terrain (Band et al., 1991; Moore et al., 1991) (Fig. 1). Varying the extent of the stream network can change the number and size of the hillslopes in any given watershed.

This research examines how watershed modeling responds to the scale incompatibility between the generalized soil property information and the other but more detailed environmental information at different model simulation scales as approximated by different levels of hillslope partitioning. The Regional Hydro-Ecological Simulation System (RHESSys) (Running et al., 1989; Band et al., 1991, 1993) is used in this research for the simulation of two watershed processes: net photosynthesis and stream flow. Two commonly used general modeling frameworks (the lumped parameter approach and the distributed approach (Maidment, 1993)) are used to simulate these processes and to examine the effect of scale incompatibility.

Two versions of spatial soil information each at different level of spatial details are used for comparison in examining the effects of scale incompatibility on the simulated processes. The first version is a conventional soil map and the second is soil information derived from a soil-land inference approach (Soil Land Inference Model, SoLIM) (Zhu, 1997, 1999; Zhu et al., 2001).

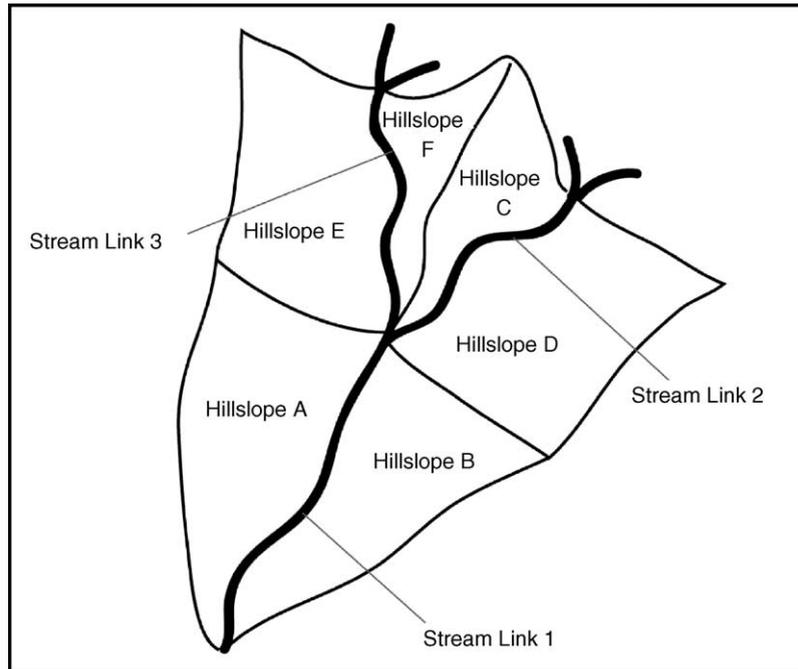


Fig. 1. Hillslope partitions. Hillslopes A and B drain to stream link 1. Hillslopes C and D drain to stream link 2. Hillslopes E and F drain to stream link 3.

Zhu (2000) established that the co-variation of soil properties and other landscape parameters is better represented through the SoLIM approach than with conventional soil maps. This allows us to assume that the model runs – based on the detailed soil spatial information from SoLIM – provide the most realistic output available at each particular model simulation scale, given that the scale of the other environmental parameters permits. Under this assumption the difference in a modeled process between the simulation based on the conventional soil map and that based on the soil information from SoLIM at a given scale can be perceived as a measure reflecting the effect due to the scale incompatibility between the conventional soil map and other detailed environmental parameters at that scale. Thus, in this study the examination of how this effect changes across scale becomes the examination of the difference in model results between the two simulations over different levels of hillslope partitioning.

In the next section we briefly describe the RHESSys model, which is used for the simulation of the hydro-ecological processes. In Section 3, this is followed by a discussion of the two ways (conventional soil survey and the SoLIM approach) for characterizing the soil spatial variation. Section 4 describes the research design and study area. Section 5 presents the results from the two watersheds which are discussed in Section 6.

Conclusion and future efforts are then presented in Section 7.

2. RHESSys

RHESSys is described in a series of journal articles in the late 1980s and 1990s (Running et al., 1987, 1989; Band et al., 1991, 1993; Nemani et al., 1993; Mackay and Band, 1997). Band et al. (1991, 1993) developed RHESSys by integrating TOPMODEL, a hillslope hydrology model (Beven and Kirkby, 1979), with FOREST-BGC, an ecosystem simulation that models water, nitrogen, and carbon cycles in forests (Running and Coughlan, 1988; Running and Gower, 1991). A geographic data processing component (a set of GIS and remote sensing techniques) organizes vegetation, topographic, and soil variables into a multi-tiered hierarchy of processing units that capture major spatial variation in model parameters.

2.1. Landscape partitioning in RHESSys

RHESSys organizes landscape parameters derived from GIS or remote sensing techniques based on hillslopes. Hillslopes are defined using an automated hillslope partitioning method described by Band (1989). The method first calculates *upslope drainage area*

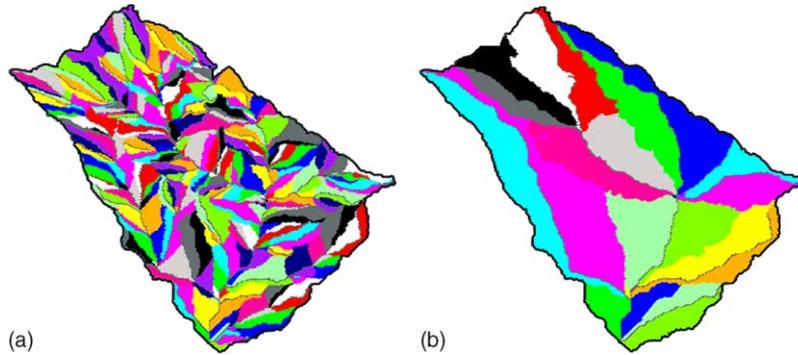


Fig. 2. Hillslope partitioning: (a) at a 100-pixel minimum UDA threshold; (b) at a 1200-pixel minimum UDA threshold.

(UDA) from a DEM for each pixel in the watershed. The UDA is the total land surface area that drains to the pixel and it increases geometrically down slope from ridge tops at the periphery of the watershed to the basin outlet. A minimum UDA threshold is defined for pixels to be classified as streams. Once defined, streams are then divided into *links* (unbranched segments of stream), each of which contains exactly two hillslopes, one on either bank, which contribute water to that stream link and no other. These hillslopes become a basic unit of landscape parameterization in RHESSys. The UDA threshold controls the level of landscape generalization: lower UDA thresholds generate more extensive stream networks, which in turn generate smaller and more numerous hillslopes (i.e., more detailed model simulation scale) (Fig. 2).

There are two approaches to representing RHESSys model parameters (e.g., leaf area index, saturated hydraulic conductivity, solum depth) within the hillslope units: the lumped parameter and distributed

parameter approach, as defined in Maidment (1993). In the lumped parameter approach, mean values for model parameters are estimated and used to represent the entire hillslope unit. The distributed parameter approach allows model parameters to co-vary within the hillslope unit. Hillslopes are divided into multiple elevation bands, which are in turn divided into multiple wetness intervals, as defined by Beven and Kirkby's (1979) hydrologic similarity index (Fig. 3). Each wetness interval contains its own set of model parameters.

2.2. Main hydro-ecological processes and soil variables in RHESSys

Band et al. (1993) integrated TOPMODEL (Beven and Kirkby, 1979), a hillslope hydrology model, into RHESSys, providing automated parameterization and simulation of hydro-ecological fluxes and storage. Among the many processes and fluxes simulated by

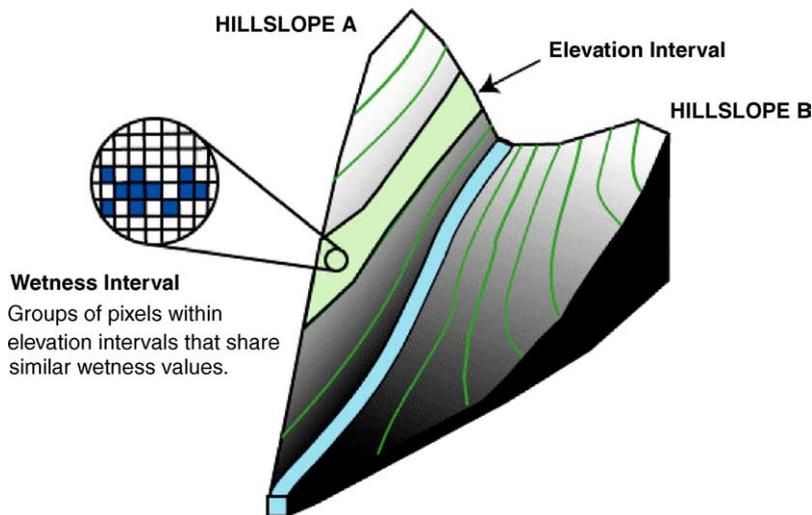


Fig. 3. Elevation and wetness intervals. These hillslope subdivisions allow representation of within hillslope variation of model parameters.

RHESSys, two main hydro-ecological fluxes (*stream flow* and *net photosynthesis* (PSN)) are used in this study to examine the effect of scale incompatibility. Stream flow consists of return flow, runoff from saturated portions of the hillslope, and base flow (see Sivapalan et al. (1987), Famiglietti and Wood (1990) and Band et al. (1993) for details). PSN is calculated using gross canopy photosynthesis and total canopy maintenance respiration. *Gross canopy photosynthesis* ($\text{kg C m}^{-2} \text{day}^{-1}$) is calculated using a CO_2 diffusion gradient, radiation and temperature-controlled mesophyll CO_2 conductance, canopy water vapor conductance, leaf area index, and day length (Lohammar et al., 1980). *Night canopy respiration* is computed from night average temperature and foliar biomass. Maintenance respiration for stems and roots is calculated using biomass and daily average air and soil temperature. Total maintenance respiration, subtracted from gross canopy photosynthesis, provides *net canopy photosynthesis*.

Two key soil variables, *rooting zone depth* and *soil saturated hydraulic conductivity*, are required for running RHESSys. Rooting zone depth affects the local soil profile saturation deficit. Hydraulic conductivity is used to calculate soil transmissivity, which is a component of the wetness index calculation. Data on the spatial variation of these two variables are derived in two different ways with each corresponding to one of the two ways in characterizing soil spatial variation. The methods used in this paper for deriving spatial data on these two variables are exactly the same as these described in Zhu and Mackay (2001) and are briefly described in Section 4.2.

3. Soil spatial models

3.1. Conventional soil mapping

Soil mapping involves two major processes: (1) conceptually organizing soil properties or parameters into taxonomic classes and (2) representing the variation of these taxonomic classes over space (Zhu, 1997). The *soil series*, a level of U.S. soil taxonomy, is the most basic description of soil variation in the parameter domain (USDA Soil Survey Division, 1993). It is designed to describe groups of soils that share similar horizon arrangements and properties (such as thickness, color, texture, organic content, rock fragment content, and mineral composition). The soil series is defined to provide relatively narrow ranges of values for each property.

Polygon delineations represent the variation of soil in the spatial domain. At the 1:24,000 scale, they

typically describe a dominant soil series or a group of series that cannot be mapped separately at that scale (USDA Soil Survey Division, 1993). The groupings of taxonomic classes that are mapped separately are called “soil map units”, which are defined to be mutually exclusive across the survey area.

The polygons and their map unit labels in the conventional soil survey are not designed to provide a level of spatial detail that approaches the scale of watershed hydro-ecological processes simulated at the mesoscale level, although such surveys are frequently used in modeling such processes (Band and Moore, 1995). One of the most widely available soil datasets for watershed modeling is the 2nd order soil survey of the U.S. Natural Resource Conservation Service at the 1:24,000 scale. This type of survey is designed for general agricultural and urban planning purposes: the assessment of soil suitability for farming, urban development, timber harvesting, or some other land use.

For modeling purposes, soil properties are assumed to be homogenous throughout the soil survey polygon, and the typical soil property value of the dominant soil class in the mapping unit is assigned to the polygon. Although there is normally some information in the survey legend about the degree of class “mixing” within each map unit and information about the range of property values within each class, this information is rarely used in models (Band and Moore, 1995). This is because the distribution of multiple minor soil classes and deviations away from typical property values within each class are not recorded in the conventional soil survey in a spatially explicit manner. Although it may be known that a particular map unit contains two to three secondary soil classes, the modeler does not know *where* these soil classes occur within the polygon. Likewise it may be known that the depth of the primary soil class ranges from 80 to 120 cm, but the modeler does not know the spatial distribution of this property, and thus is forced to assign a “typical” value of 100 cm to the entire polygon. In this way, the unrealistic representation of soil properties inherent in the conventional soil mapping is transferred to the environmental model that uses the information.

3.2. Detailed soil spatial model

Zhu and Band (1994), Zhu et al. (1996) and Zhu (1997, 1999, 2000) developed SoLIM to address the limitations of the conventional soil survey. While a full discussion of the SoLIM model is outside the scope of this article, in general terms, SoLIM consists of three components: a similarity model that portrays soil as a

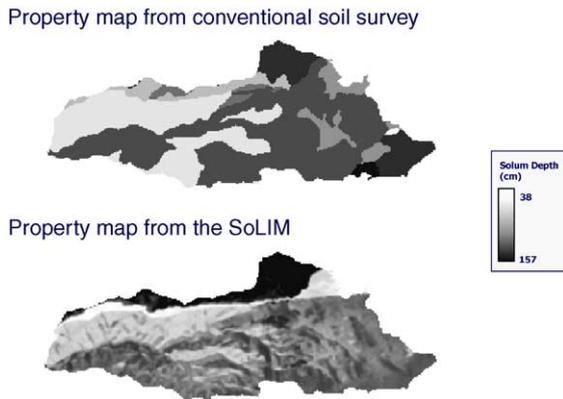


Fig. 4. Solum depth map generated from a conventional soil survey (top) and from the SoLIM model for the North Fork Elk Creek watershed (bottom).

continuum in both the spatial and attribute domains, an inference procedure for deriving soil similarity information, and soil information products such as soil series maps and spatially continuous soil property maps. The similarity model recognizes that soil types are continuous spatial features. At the boundaries of soil polygons there is considerable uncertainty about which category the soil falls into, and in fact such locations may display characteristics of two or more soil types in varying degrees. The inference procedure provides a method for generating these similarity vectors based on expert knowledge, GIS-derived data, and fuzzy logic. The output from the inference procedure – a similarity vector for each pixel in the study area – can be used to produce spatially continuous soil property data (Zhu et al., 1997; Zhu, 1997). The property values that result from a fuzzy system of representation are more detailed and more accurate than those resulting from a Boolean system of representation, as in conventional soil maps (Fig. 4) (Zhu et al., 1997).

4. Study areas and research design

The effect on model output due to scale incompatibility may be compared across several model simulation scales. The examination of this effect across a range of model simulation scales is conducted over two watersheds in western Montana. First, two versions of soil property maps are derived: one from the conventional soil survey and the other from the output of the SoLIM model (Zhu and Mackay, 2001). A series of hillslope partitioning schemes are then generated for each of the watersheds, ranging from “coarse” hillslope schemes with fewer, larger hillslopes to “fine” hillslopes schemes with more, smaller hillslopes

(Fig. 2). Terrain, leaf area index, and soil property data are organized for each of these hillslope schemes. The model is then run for both the conventional and detailed soil data for each partitioning scheme and the output from the model based on the conventional soil map is then compared to that from the model based on the detailed soil information.

4.1. Study area

The study area is a portion of the University of Montana’s Experimental Forest in western Montana, the Lubrecht watershed that is characterized by moderate to strong relief (Fig. 5). The climate is semi-arid to semi-humid, with mean annual precipitation between 50 and 76 cm (Ross and Hunter, 1976). Moisture conditions vary by slope aspect and elevation with low-elevation, south-facing slopes generally drier than higher-elevation, north-facing slopes (Zhu and Mackay, 2001).

Mountain slopes in the study area are covered primarily with second growth Douglas-fir forests along with smaller amounts of western larch and ponderosa pine. Ponderosa pine forest dominates at lower elevations, while Douglas-fir continue to elevations of roughly 1650 m, beyond which subalpine fir and Engelmann-spruce are predominant (Zhu and Mackay, 2001).

There are three major geology types in the study area: belt rocks, granite, and limestone, which have weathered to form twelve soil series (Table 1) (Zhu and Mackay, 2001). Approximately 90% of the soils in the study area are Inceptisols—poorly developed soils with minimal organic content.

The two watersheds under study (Fig. 5b), Cap Wallace Gulch and North Fork Elk Creek, are adjacent tributaries to Elk Creek, which is a tributary to the Blackfoot River. The Cap Wallace Gulch watershed is 6.2 km² and lies just to north of North Fork. North Fork is 18.6 km². The DEM and remotely sensed data are both at 30 m pixel resolution.

4.2. Data preparation

Two sets of environmental parameters were assembled for the model, one for each method of soil parameterization. RHESSys requires elevation, gradient, aspect, UDA, leaf area index, and station climate data, along with soils information. The only difference in the two sets was the representation of soil properties (conventional or detailed). All other data layers remained the same and the pixel size is 30 m by 30 m.

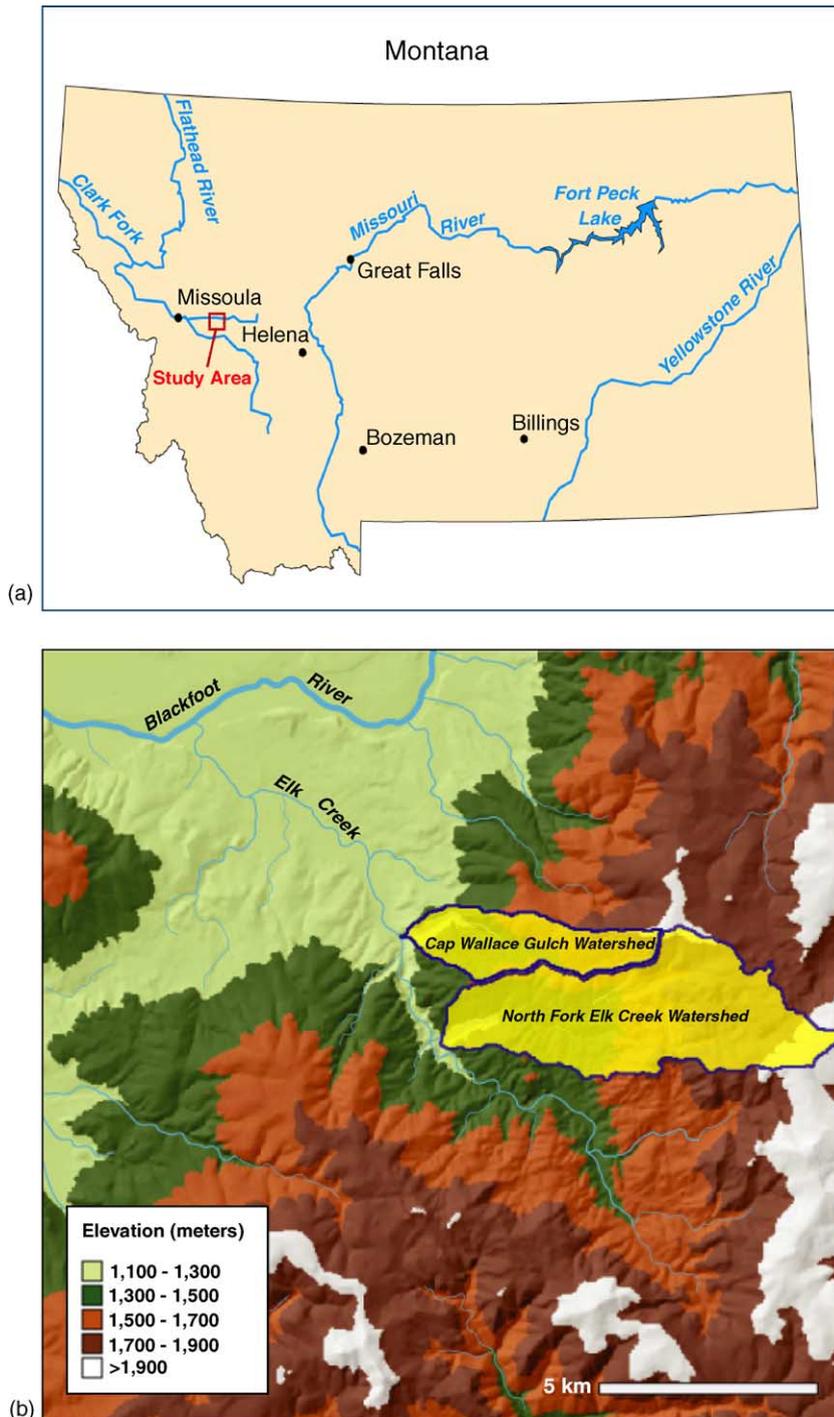


Fig. 5. Location of study area: (a) general location; (b) the two study watersheds and surrounding elevations.

RHESSys requires two soil properties: saturated hydraulic conductivity and rooting zone depth. Saturated hydraulic conductivity for each of the soil series in the study area was approximated by soil texture using Clapp and Hornberger (1978). The spatial variation of

rooting zone depth is very difficult to characterize across a landscape, so solum depth (depth to the bottom on the B-horizon) was used as a surrogate.

Hydraulic conductivity and solum depth values were assigned to conventional survey polygons using the

Table 1
Soil series in the Lubrecht study area

Parent materials	Soil series	Soil subgroup	Soil order
Granite	Ambrant	Udic Ustochrepts	Inceptisol
	Elkner	Typic Cryochrepts	Inceptisol
	Ovando	Typic Cryorthents	Entisol
	Rochester	Typic Ustorthents	Entisol
Belt (shale)	Evaro	Typic Cryochrepts	Inceptisol
	Sharrott	Lithic Ustochrepts	Inceptisol
	Tevis	Dystric Eutrochrepts	Inceptisol
	Winkler	Udic Ustochrepts	Inceptisol
	Winkler (Cool)	Udic Ustochrepts	Inceptisol
Limestone	Repp	Typic Ustochrepts	Inceptisol
	Trapps	Typic Eutroboralfs	Alfisol
	Whitore	Typic Cryochrepts	Inceptisol

dominant soil series of each map unit. A conventional survey was acquired for the study area in the form of SSURGO (Soil Survey Geographic Database) from the U.S. Natural Resource Conservation Service. Each soil polygon was related to a map unit in the SSURGO database, which consisted of one or more components (soil series). The soil series with the highest representative fraction in the map unit was used. Soil textures (used to approximate hydraulic conductivity) and solum depth for each of the soil series in the area were determined through field research. These field values were applied to all the polygons in the study area. The resulting vector maps of soil property polygons were then converted into raster format, using grid cell size (30 m by 30 m) that matched that of the terrain and leaf area index data layers.

Detailed hydraulic conductivity and solum depth maps were generated using the SoLIM-derived similarity vector and a linear additive weighting calculation as described in [Zhu and Mackay \(2001\)](#). Fuzzy membership maps were generated for each of the soil series in the study area using the knowledge acquisition and inference techniques described in [Zhu and Band \(1994\)](#), [Zhu et al. \(1996, 1997\)](#) and [Zhu \(1999, 2000\)](#). Six major environmental conditions were used to drive the inference: elevation, aspect, profile curvature, gradient, canopy coverage, and parent material. The same typical property values that were applied to the dominant soil series in the SSURGO polygons were used in the linear additive weighting equation. The resulting detailed soil property maps matched the spatial extent and resolution of the terrain data and leaf area index, which was a requirement for processing the data layers into RHESSys input files.

Table 2
Hillslope partitioning schemes for Cap Wallace Gulch

Threshold (pixels)	Number of partitions	Mean partition size (pixels)
50	77	59
100	56	113
200	26	251
300	18	365
400	14	471
500	10	663
600	6	1111
700	6	1111
800	6	1111

4.3. Parameterization and data organization

RHESSys receives environmental data in the form of aggregate values for each hillslope in a particular partitioning scheme. This research aimed to vary the partitioning schemes to produce a wide range of model simulation scales. Model simulation scale was assessed by calculating the mean size of the hillslope partitions ([Tables 2 and 3](#)), which were created by stream networks defined by progressively increasing UDA thresholds. Stream networks defined by large UDA thresholds are highly simplified (un-branched), and require large increases in the threshold to generate ever more simplified stream networks. This explains why multiple threshold settings at the high end of the threshold size range produce the same number of hillslope partitions ([Tables 2 and 3](#)).

Once hillslope partitions were developed, the terrain, leaf area index, and soils data layers were processed into “cartridge files”, which stores environmental parameters for each hillslope in the partitioning scheme. Two cartridge files were assembled for each partitioning scheme in each of the two study watersheds: one cartridge file for the environmental dataset that included

Table 3
Hillslope partitioning schemes for North Fork

Threshold (pixels)	Number of partitions	Mean partition size (pixels)
100	198	104
200	118	175
300	78	264
400	54	382
500	38	542
600	26	793
700	22	938
800	18	1145
900	18	1145
1000	18	1145

conventional soil data and the other for the dataset that included detailed soil spatial information. For simulations using the lumped parameter approach, the cartridge file is the only input file needed because the model calculates a hillslope mean value for each of the environmental variables, which is used to determine the daily response of that hillslope to the precipitation and temperature conditions specified in the climate file.

For the distributed parameter approach, a distributed parameterization scheme is used which allows environmental variables to co-vary at a more detailed level than that provided by the hillslopes. In addition to the cartridge file described above, a frequency file is used to capture the co-variation of parameters within each slope. The co-variation is captured by dividing each hillslope into regular elevation intervals, then within each elevation interval, assembling pixels into groups based on the hydrologic similarity index (the topographic wetness intervals) (Mackay and Band, 1997). These groups are then used as the aggregation units for the set of environmental variables (including the soil properties derived from detailed and conventional maps). This portion of the experiment was designed to examine whether any relationships between model simulation scale and soil data scale also occurred in the distributed scheme.

4.4. Model output and comparison

The model was operated for a 365-day simulation (after a 1 year initialization run) for the cartridge file containing the conventional soil information and the cartridge file containing detailed soil information. This process was repeated for each model scale within each watershed. Differences were examined for two

model outputs: stream flow and net photosynthesis (Fig. 6).

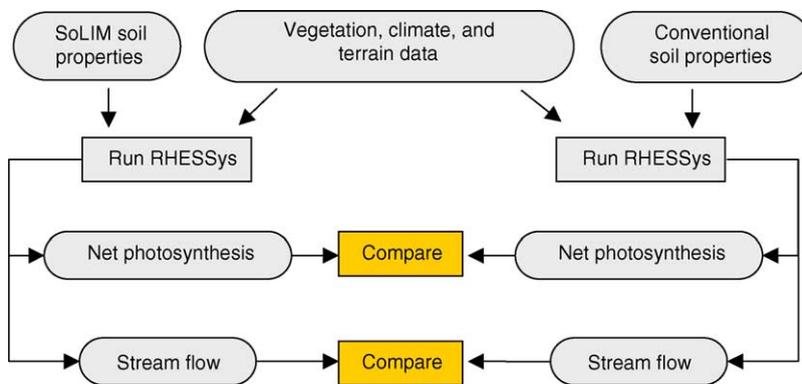
Root mean squared difference (RMSD) in daily output values and total annual output difference were used to characterize the performance of conventional soil data models runs relative to detailed soil data model runs over the 365-day study period. RMSD values for stream flow and net photosynthesis were determined by calculating the squares of the difference between the two runs for each of 365 daily values, then taking the square root of the mean of those values. These two root mean squared difference calculations are expressed as

$$RMSD_{streamflow} = \sqrt{\frac{\sum_{t=1}^n (S_{conv}^t - S_{det}^t)^2}{n}}$$

$$RMSD_{PSN} = \sqrt{\frac{\sum_{t=1}^n (PSN_{conv}^t - PSN_{det}^t)^2}{n}}$$

where S_{conv} and S_{det} are stream flows for the conventional and detailed soil data runs, respectively, PSN_{conv} and PSN_{det} are net photosynthesis for the conventional and detailed soil data runs, t is the day number of the study period, ranging from 1 to n , the final day of the study period (in this study, day 365). Because stream flow and PSN are relatively low for large portions of the study period, differences were relatively low for large portions of the study period.

The relative performance of the two methods of soil parameterization must be judged in terms of the model’s response to relatively infrequent heavy rainfall events, as these generate most of the annual stream flow and constitute a key driver of net photosynthesis. The root mean squared difference (RMSD) was chosen as a statistical measure because it captures both positive and



At one particular model scale

Fig. 6. RHESSys output comparison at a given scale.

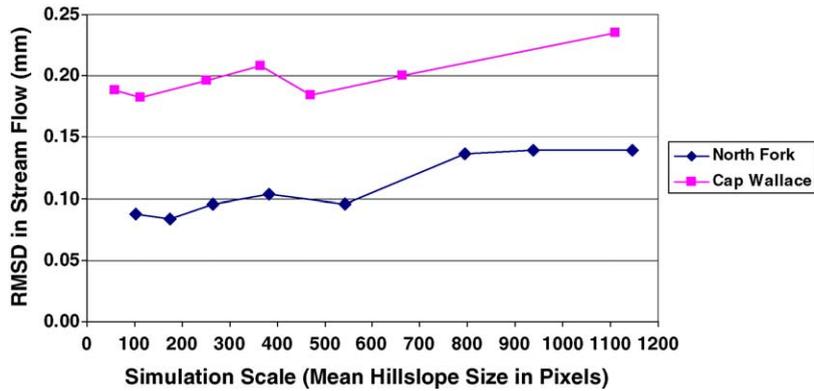


Fig. 7. Root mean squared difference in stream flow by simulation scale for Cap Wallace Gulch and North Fork Elk Creek (lumped approach).

negative differences and accentuates these relatively infrequent, large differences between the conventional and detailed model runs. Assuming that the detailed parameterization scheme results in the most realistic model output at a given level of hillslope partitioning, a small RMSD means that the conventional soil data is generating relatively realistic output at that scale.

5. Results

5.1. Lumped parameter approach

Figs. 7 and 8 display RMSD results for the lumped parameter approach for Cap Wallace Gulch and North Fork. The graphs show trends in the conventional soil data model run performance relative to detailed data runs at a range of model simulation scales. The conventional soil model run output for stream flow seems to approach the detailed model run benchmark at two model simulation scales [150 pixels at 30 m × 30 m pixel (13,500 m²) and around 500 pixels

(45,000 m²)] (Fig. 7), with some declines in performance at scales intermediate to these two points. In contrast, net photosynthesis outputs show consistent improvements in the conventional soil run performance as hillslope size increases until a point near 500 pixels (Fig. 8). This is the point of minimum difference between model outputs of each of the two soil parameterization techniques. For both stream flow and net photosynthesis in both watersheds, the difference between the conventional and detailed soil model runs increases at model scales beyond a mean hillslope size of roughly 500 pixels.

5.2. Distributed parameter approach

Results for the distributed parameter approach (Figs. 9 and 10) show much smaller differences between the conventional and detailed model runs, but it may be possible to describe some trends in the conventional soil data performance across the range of model simulation scales. Stream flow outputs for Cap

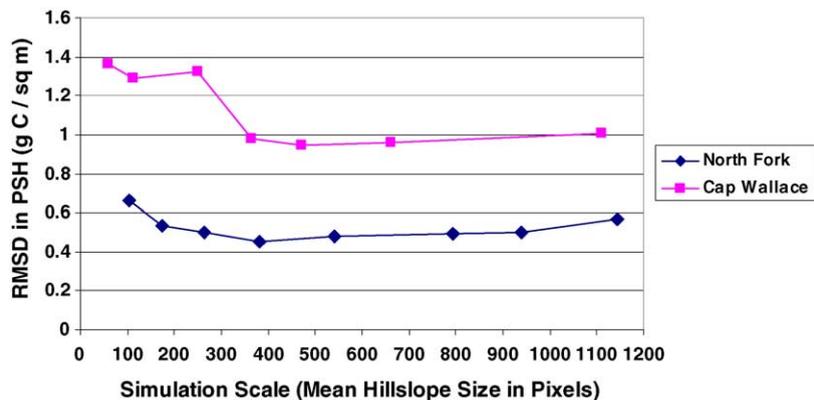


Fig. 8. Root mean squared difference in net photosynthesis by simulation scale for Cap Wallace Gulch and North Fork Elk Creek (lumped approach).

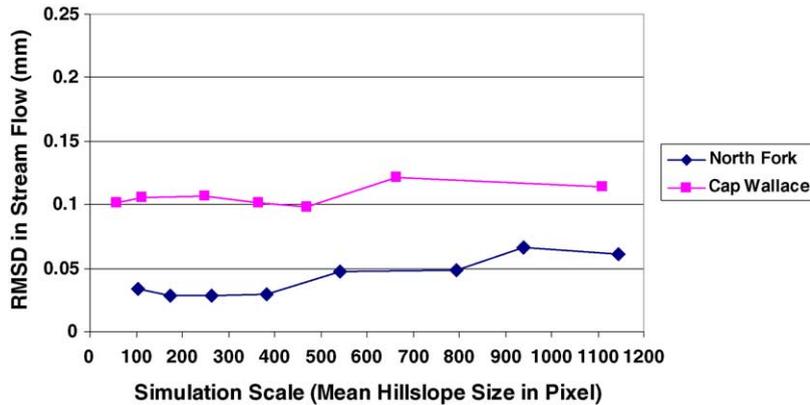


Fig. 9. Root mean squared difference in stream flow by simulation scale for Cap Wallace Gulch and North Fork Elk Creek (distributed approach).

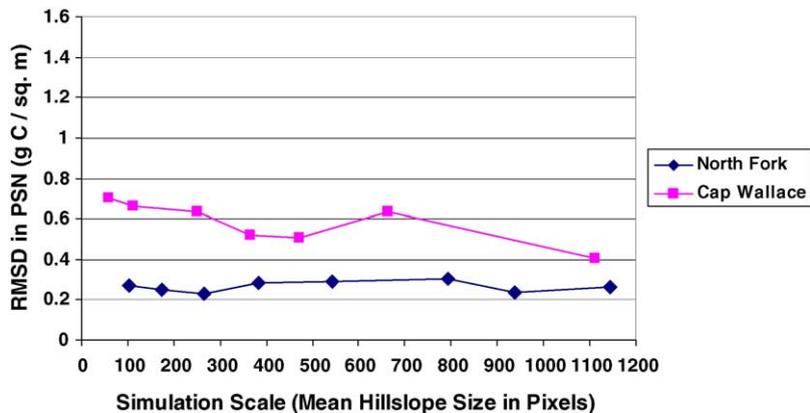


Fig. 10. Root mean squared difference in net photosynthesis by simulation scale for Cap Wallace Gulch and North Fork Elk Creek (distributed approach).

Wallace Gulch and North Fork show relatively consistent performance by the conventional soil data at scales with mean hillslope sizes between 400 and 500 pixels. As with the lumped parameterization scheme, net photosynthesis output shows improvement in conventional run performance up to a point of minimal difference at a model scale near 500 pixel mean

hillslope size. With the exception of net photosynthesis for North Fork Elk Creek, the distributed approach tended to produce relatively degraded performance (larger differences) by the conventional soil data at hillslope scales between 400 and 800 pixels.

Tables 4 and 5 show RMSD values as a percentage of mean daily stream flow and net photosynthesis over the

Table 4
Differences in stream flow and net photosynthesis (PSN) for Cap Wallace Gulch (lumped approach)

Mean hillslope size (pixels)	Root mean squared difference (RMSD) in stream flow (mm)	RMSD as % of mean daily stream flow ^a	Root mean squared difference (RMSD) in PSN (mm)	RMSD as % of mean daily PSN ^a
59	0.19	45	1.36	46
113	0.18	45	1.29	43
251	0.20	49	1.32	43
365	0.21	55	0.98	31
471	0.18	49	0.95	30
663	0.20	55	0.96	30
1111	0.24	65	1.01	31

^a From simulation based on information from the conventional soil map.

Table 5
Differences in stream flow and net photosynthesis (PSN) for Cap Wallace Gulch (distributed approach)

Mean hillslope size (pixels)	Root mean squared difference (RMSD) in stream flow (mm)	RMSD as % of mean daily stream flow ^a	Root mean squared difference RMSD in PSN (mm)	RMSD as % of mean daily PSN ^a
59	0.10	19	0.70	29
113	0.11	19	0.66	28
251	0.11	19	0.63	27
365	0.10	18	0.52	23
471	0.10	17	0.51	22
663	0.12	21	0.63	29
1111	0.14	24	0.36	16

^a From simulation based on information from the conventional soil map.

Cap Wallace Gulch watershed for the lumped parameterization and the distributed parameterization schemes, respectively. These values indicate that the influence of model simulation scale on scale incompatibility effects under the distributed approach is somewhat smaller than under the lumped parameterization.

6. Discussion

It seems likely that certain model simulation scales “reduce” the effect that results from overgeneralization of soil properties. In this experiment, this model simulation scale effect is demonstrated by the improvement in the conventional model run performance in the 50–500-pixel range. At the low end of this range, hillslope schemes appear to dissect soil property patches enough to allow the inclusions represented by detailed soil data to be parameterized and incorporated into model output. This exposes the scale incompatibility of the conventional soil information relative to the other data layers. Detailed soil data covaries with other environmental parameters to produce noticeably different output from conventional data, which explains some of the larger RMSD values, especially for net photosynthesis (Figs. 8 and 10). As model scale approaches 500 pixels, more generalized hillslope schemes tend to “average out” the spatial variation of the detailed soil data such that hillslope-aggregated soil parameters for conventional and detailed soil data reach a better level of agreement. Here the incompatibility effect is minimized because all environmental data layers are effectively “re-sampled” to a more generalized resolution that matches that of the conventional soil data.

The appearance of relative agreement in RMSD values at some model scales (Figs. 7–10) appears to coincide with modal soil patch sizes. Although the mode in the frequency distributions of soil patch size

(Figs. 11 and 12) is quite weak, these figures do show a somewhat higher number of soil patches at or near 500 pixels relative to sizes just above and below for both soil properties. The explanation for this effect lies in the “design scale” of the conventional soil survey. Soil survey map units were designed to capture, at a certain level of generalization, the variation of soil properties, which are a function of the other environmental parameters being modeled. Thus the patches and the other environmental parameters reach optimal co-variation only at particular scales, beyond which the co-variance with other environmental variables increases model sensitivity to soil parameterization method.

At model scales that roughly match the scale of the conventional soil data, the generalization in both the conventional and the detailed soil property maps approaches the same level. This research assumes that the detailed soil property maps provide the most realistic model output at a given scale, which then implies that the conventional maps produce the most realistic output at a scale that matches the original scale of the soil survey. When the model simulation scale is set at a finer resolution than the modal soil property patch size, the soil property map produced from the conventional soil survey is inadequate to provide the

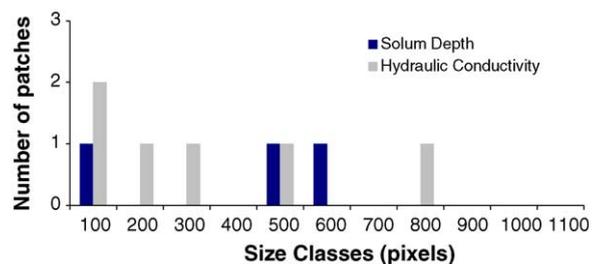


Fig. 11. Distribution of soil patch sizes for solum depth and hydraulic conductivity properties (Cap Wallace Gulch).

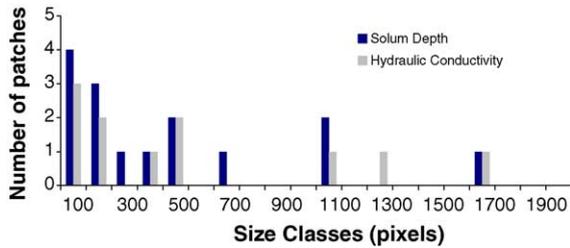


Fig. 12. Distribution of soil patch sizes for solum depth and hydraulic conductivity properties (North Fork).

level of spatial detail needed. Conversely, when the model scale is set at a coarser resolution than the modal soil property patch size, the conventional soil property map is generalized to a level that creates another kind of unrealistic representation of soil property variation, and model outputs based on the conventional maps again deviate from these based on the detailed soil information.

The sensitivity to soil representation method are obviously reduced by parameterization schemes that allow within-hillslope co-variation of environmental variables (as in the distributed approach) (Figs. 9 and 10), but not so much that they eliminate the impact of model simulation scale. Reduced effect is again found at model simulation scales that approximate the scale of the conventional soil data. The effect due to soil data scale incompatibility in the distributed scheme, however, may not be large enough to justify tuning model simulation scale to meet the scale of the original soil representation scheme.

The preceding discussion implies that there is an “optimal” model scale that minimizes the effect on model outputs due to scale incompatibility. Two problems arise in the pursuit of an optimal model simulation scale. The first is the result of the hillslope partitioning scheme, which is most relevant to the lumped approach and the second is the errors in assigning typical property values to soil polygons.

It is highly unlikely that hillslope partitioning at any scale will exactly match the configuration of soil property variation defined by the conventional map-derived soil patches. Certain spatial soil patterns, especially those involving changes in bedrock geology, will not follow hillslope patterns. Although soil patches might express the mean value of the detailed soil variation within the patch area, the partition might split or merge soils patches. This could generate mean values for conventional soil information that deviate from the mean values of the same partition in the detailed soil layer.

Assigning typical property values to soil polygons involves errors. Even it were possible to get the hillslope partitioning scheme to match the soil patch configuration, the property values of the soil patches would not always match mean soil property values derived from detailed soil data. There is significant uncertainty in the assignment of a typical soil property value to a given map unit, since there is no information about the spatial distribution of soil class and property deviations within the polygons belonging to the map unit. Model runs could only completely eliminate scale incompatibility effects when hillslope configuration exactly matched patch configuration, and the mean property value of patches accurately described the real spatial variation of that property within the patch area. These limitations make a certain amount of difference between the conventional and detailed soil parameterization techniques presented here unavoidable.

7. Conclusions and future research

The difference in model output due to scale incompatibility between soil information from conventional soil maps and detail information about other environmental variables does change with change of model simulation scale. The results of this research suggest that, when using conventional soil data, operating the watershed models at scales approaching those of the soil data layers reduces bias in model predictions. The scale of the soil property data, which is often converted from polygon-model soil surveys, can be assessed using a frequency distribution of soil patches—contiguous areas of pixels sharing the same property value. Modeling at scales that are much smaller or much larger than modal soil patch sizes could introduce errors due to the scale incompatibility between soil data and other detailed environmental data (such as digital terrain data and remotely sensed data).

This study also shows that the distributed parameterization scheme may reduce the degree of effect on model results due to scale incompatibility between conventional soil map and other detailed environmental data layers. However, the change in pattern of the effect across model simulation scale is very similar to that under the lumped parameterization scheme.

This research has examined patterns of mean daily difference between conventional and detailed soil data model runs using a 365-day data set. It is important to note, however, that the most significant gaps between model runs using the two soil parameterization schemes emerge during the period of moisture stress in the late

summer as reported by Zhu and MacKay (2001). During this period, available moisture becomes a crucial limiting factor in net photosynthesis and other ecological indicators, making the model especially sensitive to soil transmissivity and solum depth values. It may be more appropriate to divide the study year into three periods, and focus analysis on the hot and dry portion of the year. This may reveal clearer patterns of difference between model runs over a range of scales.

Due to the limitation on the availability of detailed soil information the current study was conducted over two small watersheds over which the numbers of soil polygons are rather small. As a result the modal size of soil patches in the conventional soil map over each watershed is not clearly shown. A more extensive study area is currently under development and a repeat of this exercise is underway to confirm the results found in this paper. Future research efforts should include more watersheds in different environmental settings to further test and confirm the hypothesis.

Acknowledgements

This research is supported by a grant from Natural Resources Conservation Service, United States Department of Agriculture under Agreement No. 69-5F48-9-00186 and the support provided to A.-Xing Zhu by Chinese Academy of Sciences via its “One-Hundred Talents Program”.

References

- Band, L.E., 1989. A terrain-based watershed information system. *Hydrol. Process.* 3, 151–162.
- Band, L.E., 1993. Effect of land surface representation on forest water and carbon budgets. *J. Hydrol.* 150, 749–772.
- Band, L.E., Moore, I.D., 1995. Scale: landscape attributes and geographical information systems. *Hydrol. Process.* 9, 401–422.
- Band, L.E., Peterson, D.L., Running, S.W., Coughlan, J., Lammers, R., Dungan, J., Nemani, R., 1991. Forest ecosystem processes at the watershed scale: basis for distributed simulation. *Ecol. Model.* 56, 171–196.
- Band, L.E., Patterson, P., Nemani, R., Running, S.W., 1993. Forest ecosystem processes at the watershed scale: incorporating hillslope hydrology. *Agric. For. Meteorol.* 63, 93–126.
- Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.* 23, 43–69.
- Blöschl, G., 1998. Scale and Scaling in Hydrology—A Framework for Thinking and Analysis. John Wiley, Chichester, pp. 68–110.
- Clapp, R.B., Hornberger, G.M., 1978. Empirical equations for some soil hydraulic properties. *Water Resour. Res.* 14, 601–604.
- Ehleringer, J.R., Field, C.B. (Eds.), 1993. *Scaling Physiological Processes: Leaf to Globe*. Academic Press, San Diego, p. 310.
- Famiglietti, J., Wood, E.F., 1990. Evapotranspiration and runoff from large land areas: land surface hydrology for global circulation models. *Surv. Geophys.* 12, 179–204.
- Georges, V., Chen, T., 2002. Sensitivity of a large-scale hydrologic model to quality of input data obtained at different scales; distributed versus stochastic non-distributed modelling. *J. Hydrol.* 264, 101–112.
- Haddeland, I., Matheussen, B.V., Lettenmaier, D.P., 2002. Influence of spatial resolution on simulated streamflow in a macroscale hydrologic model. *Water Resour. Res.* 38, 291–2910.
- Koren, V., Finnerty, B.D., Schaake J.C., 1999. Scale dependencies of hydrologic models to spatial variability of precipitation. *J. Hydrol.* 217, 285–302.
- Lohammar, T., Larsson, S., Linder, S., Falk, S.O., 1980. FAST—simulation models of gaseous exchange in Scots pine. *Ecol. Bull.* 32, 505–523.
- Mackay, D.S., Band, L.E., 1997. Forest ecosystem processes at the watershed scale: dynamic coupling of distributed hydrology and canopy growth. *Hydrol. Process.* 11, 1197–1217.
- Maidment, D.R., 1993. GIS and hydrologic modeling. In: Goodchild, M.F., Parks, B.O., Steyaert, L.T. (Eds.), *Environmental Modeling with GIS*. Oxford University Press, New York, pp. 147–167.
- Moore, I.D., Grayson, R.B., Ladson, A.R., 1991. Digital terrain modeling: a review of hydrological, geomorphological and biological applications. *Hydrol. Process.* 5, 3–30.
- Nemani, R., Pierce, L., Running, S., Band, L.E., 1993. Forest ecosystem processes at the watershed scale: sensitivity to remotely sensed leaf area index estimates. *Int. J. Rem. Sens.* 14, 2519–2534.
- Ranjan, S.M., Wurbs, R.A., 2002. Scale-dependent soil and climate variability effects on watershed water balance of the SWAT model. *J. Hydrol.* vol. 256, 264–285.
- Ross, R.L., Hunter, H.E., 1976. *Climax Vegetation of Montana Based on Soils and Climate*. USDA Soil Conservation Service, Bozeman, Montana, p. 36.
- Running, S.W., Coughlan, J.C., 1988. A general model of forest ecosystem processes for regional applications: I. Hydrologic balance, canopy gas exchange and primary production processes. *Ecol. Model.* 42, 125–154.
- Running, S.W., Gower, S.T., 1991. FOREST-BGC, a general model of forest ecosystem processes for regional applications. II. Dynamic carbon allocation and nitrogen budgets. *Tree Physiol.* 9, 147–160.
- Running, S.W., Nemani, R., Hungerford, R.D., 1987. Extrapolation of synoptic meteorological data in mountainous terrain, and its use for simulating forest evapotranspiration and photosynthesis. *Can. J. For. Res.* 17, 472–483.
- Running, S.W., Nemani, R.R., Peterson, D.L., Band, L.E., Potts, D.F., Pierce, L.L., 1989. Mapping regional forest evapotranspiration and photosynthesis by coupling satellite data with ecosystems simulation. *Ecology* 70, 1090–1101.
- Sivapalan, M., Wood, E.F., Beven, K.J., 1987. On hydrologic similarity: 2. A scaled model of storm runoff production. *Water Resour. Res.* 23, 2266–2278.
- Thielen, A.H., Lücke, A., Diekkrüger, B., Richter, O., 1999. Scaling input data by GIS for hydrological modeling. *Hydrol. Process.* 13, 611–630.
- USDA (United States Department of Agriculture) Soil Survey Division, 1993. *Soil Survey Manual*. Washington DC: United States Department of Agriculture, p. 238.

- Van Gardingen, P.R., Foody, G.M., Curran, P.J. (Eds.), 1997. Scaling-up: From Cell to Landscape. Cambridge University Press, New York, pp. 113–136.
- Western, A.W., Blöschl, G., 1999. On the spatial scaling of soil moisture. *J. Hydrol.* 217, 203–224.
- Wolock, D.M., Price, C.V., 1994. Effects of digital elevation model map scale and data resolution on a topography-based watershed model. *Water Resour. Res.* 30, 3041–3052.
- Zhang, W., Montgomery, D.R., 1994. Digital elevation model grid size, landscape representation, and hydrologic simulation. *J. Hydrol.* 102, 1019–1028.
- Zhu, A.X., 1997. A similarity model for representing soil spatial information. *Geoderma* 77, 217–242.
- Zhu, A.X., 1999. A personal construct-based knowledge acquisition process for natural resource mapping. *Int. J. Geogr. Informat. Sci.* 13, 119–141.
- Zhu, A.X., 2000. Mapping soil landscape as spatial continua: the neural network approach. *Water Resour. Res.* 36, 663–677.
- Zhu, A.X., 2005. Research issues on uncertainty in geographic data and GIS-based analysis. In: McMaster, R., Ursery, L. (Eds.), *A Research Agenda For Geographic Information Science*. CRC Press, New York, pp. 197–223.
- Zhu, A.X., Band, L.E., 1994. A knowledge-based approach to data integration for soil mapping. *Can. J. Rem. Sens.* 20, 408–418.
- Zhu, A.X., Mackay, D.S., 2001. Effects of spatial detail of soil information on watershed modeling. *J. Hydrol.* 248, 54–77.
- Zhu, A.X., Band, L.E., Dutton, B., Nimlos, T.J., 1996. Automated soil inference under fuzzy logic. *Ecol. Model.* 90, 123–145.
- Zhu, A.X., Band, L.E., Vertessy, R., Dutton, B., 1997. Deriving soil property using a soil land inference model (SoLIM). *Soil Sci. Soc. Am. J.* 61, 523–533.
- Zhu, A.X., Hudson, B., Burt, J., Lubich, K., Simonson, D., 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. *Soil Sci. Soc. Am. J.* 65, 1463–1472.