Chapter 3
A Generalized Additive Soil Depth Model for a Mountainous Semi-Arid Watershed Based Upon Topographic and Land Cover Attributes

T.K. Tesfa, D.G. Tarboton, D.G. Chandler, and J.P. McNamara

Abstract Soil depth is an important input parameter in hydrological and ecological modeling. Presently, the soil depth data available in national soil databases (STATSGO, SSURGO) is provided as averages within generalized map units. Spatial uncertainty within these units limits their applicability for spatially distributed modeling. This work reports a statistical model for prediction of soil depth in a semi-arid mountainous watershed that is based upon topographic and other landscape attributes. Soil depth was surveyed by driving a rod into the ground until refusal at geo-referenced locations selected to represent the range of topographic and land cover variations in Dry Creek Experimental Watershed, Boise, Idaho, USA. The soil depth survey consisted of a model calibration set, measured at 819 locations over 8 sub-watersheds, and a model testing set, measured at 130 locations randomly distributed over the remainder of the watershed. Topographic attributes were derived from a Digital Elevation Model. Land cover attributes were derived from Landsat TM remote sensing images and high resolution aerial photographs. A Generalized Additive Model was developed to predict soil depth over the watershed from these attributes. This model explained about 50% of the soil depth spatial variation and is an important improvement towards solving the need in distributed modeling for distributed soil depth input data.

Keywords Generalized additive models · Explanatory variables · Land cover attributes · Soil depth · Topographic attributes

3.1 Introduction

Soil depth is one of the most important input parameters for hydrological and ecological models. Its spatial pattern, significantly affects soil moisture, runoff generation, and subsurface and groundwater flow (Freer et al., 2002; McNamara

T.K. Tesfa (✉)
Pacific Northwest National Laboratory, PO Box 999 Richland, WA 99352, USA
e-mail: Teklu.Tesfa@pnl.gov

J.L. Boettinger et al. (eds.), *Digital Soil Mapping*, Progress in Soil Science 2,
DOI 10.1007/978-90-481-8863-5_3,

et al., 2005; Stieglitz et al., 2003). Consequently, its accurate representation is becoming increasingly important. It is highly variable spatially, and laborious, time-consuming and difficult to practically measure even for a modestly sized watershed (Dietrich et al., 1995). There is thus a need for models that can predict the spatial pattern of soil depth.

The national soil databases (SSURGO & STATSGO) have been the main sources of soil depth information used in hydrological and ecological modeling in the United States. In these soil databases, soils are spatially represented as discrete map units with sharp boundaries. A map unit may be comprised of more than one soil component but these components are not represented spatially within the map unit. As a result, soil attributes are spatially represented at map unit level as a mean or some other representative value of the components. Such a representation limits quantification of the variability of soil attributes within each class, and class boundaries generalize the spatial pattern of the soil properties, absorbing small scale variability into larger class units (Moore et al., 1993; Zhu, 1997). There is a need in spatially distributed modeling for fine scale models of soil depth that do not have these limitations. Past efforts to develop fine scale models include fuzzy logic, statistical and physically based approaches (Dietrich et al., 1995; Moore et al., 1993; Zhu, 1997).

In this chapter, we develop a statistical model for prediction of the spatial pattern of soil depth over complex terrain from topographic and land cover attributes in a mountainous semi arid watershed. Topographic and land cover attributes intended to have explanatory capability for soil depth were derived from a digital elevation model (DEM) and Landsat TM remote sensing images. A Generalized Additive Model (GAM) (Hastie and Tibshirani, 1990) was applied to predict soil depth based on these topographic and land cover attributes using soil depth data measured at 819 points at 8 sub-watersheds within Dry Creek Experimental Watershed (DCEW). This calibration data set was randomly divided into a training subset consisting of 75% of the data and a validation subset consisting of the remaining 25% that was used to estimate the prediction error for variable and model complexity selection (see Chapter 7). Soil depth data measured at an additional 130 random points within DCEW was used as an out of sample data set to test the model results. The Nash-Sutcliffe efficiency coefficient, which is widely used to assess the predictive accuracy of models, was used to evaluate the efficiency of the soil depth model.

3.2 Study Area

This study was carried out in the Dry Creek Experimental Watershed (DCEW), about 28 km^2 in area, located in the semi-arid southwestern region of Idaho, USA (Fig. 3.1). The area is composed of mountainous and foothills topography with elevations that range from 1,000 to 2,100 m (Williams et al., 2008). The landscape is typified by moderately steep slopes with average slope of about 25%, with steeper north facing slopes than south facing slopes, and is strongly dissected by streams.

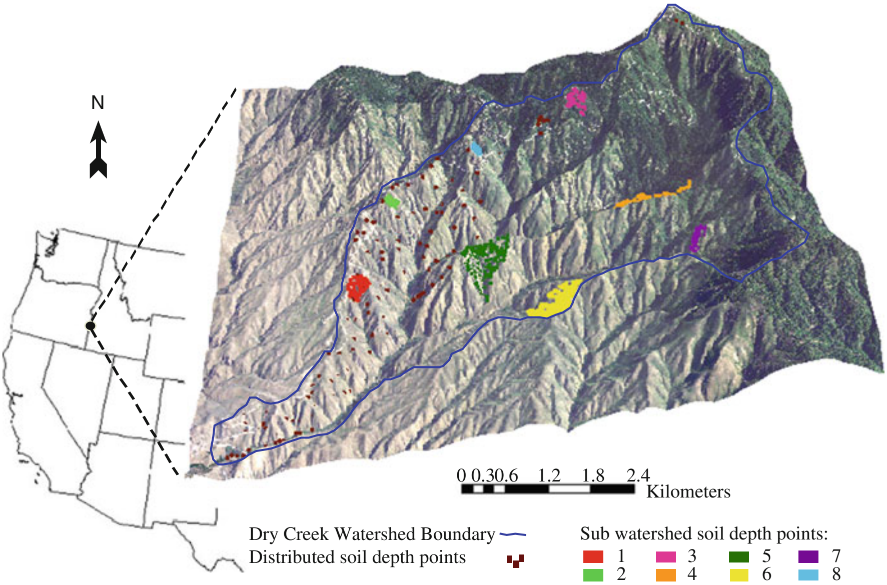
N
0 0.3 0.6 1.2 1.8 2.4
Kilometers
Dry Creek Watershed Boundary
Distributed soil depth points
Sub watershed soil depth points:
1
2
3
4
5
6
7
8

Fig. 3.1 Dry Creek Experimental Watershed (DCEW) near Boise, ID, in the Western USA. Points show locations where soil depth was sampled

The climate is a steppe summer dry climate at low elevation and moist continental climate with dry summers at high elevation (McNamara et al., 2005). Precipitation is highest in winter, as snow in the highlands and rain in the lowlands, and in spring in the form of rain. There are occasional summer thunderstorms. The average annual precipitation ranges from 37 cm at lower elevations to 57 cm at higher elevations (Williams, 2005). The average monthly temperatures are highest in July and lowest in January. Streamflow typically remains low in the early and mid winter and peaks in the early to mid spring due to snowmelt (McNamara et al., 2005).

Vegetation varies with elevation and landscape aspect (McNamara et al., 2005; Williams, 2005). Grass (south facing aspects) and sagebrush (north facing aspects) are dominant at lower elevations. Upper elevations are dominated by ponderosa pine (*Pinus ponderosa*) and Douglas-fir (*Pseudotsuga menziesii*) forest with patches of lodgepole pine (*Pinus contorta*) and aspen (*Populus tremuloides*). Middle elevations range from grass and shrublands to open forest of ponderosa pine and Douglas-fir.

Soils in this area are formed from weathering of the underlying Idaho Batholith, which is a granite intrusion ranging in age from 75 to 85 million years (Lewis et al., 1987; USDA, 1997). The dominant rock type is biotite granodiorite which consists of medium to coarse-grained rocks composed of plagioclase, quartz, potassium feldspar, and biotite (Johnson et al., 1988). The soils are classified into three general great groups according to US Soil Taxonomy: Argixerolls, Haploxerolls, and Haplocambids. These soils range from loam to sandy loam in texture and are generally well drained with high surface erosion potential (USDA, 1997). The Natural Resource Conservation's soil survey of the Boise Front (SSURGO soil database for survey area symbol ID903 obtained from Idaho NRCS office) provides a more detailed description of the soils underlying the watershed.

3.3 Methodology

3.3.1 Field and Digital Data

Eight sub-watersheds were selected to represent the elevation, slope, aspect and land cover variability present within the DCEW. Soil depth was surveyed at a total of 819 points within these sub-watersheds. Survey locations were chosen to represent the range of topographic and land cover variation in the sub-watersheds. At each survey location three depth replicates two to three meters apart were collected by driving a 220 cm long 1.27 cm diameter sharpened copper coated steel rod graduated at 5 cm interval into the ground using a fence post pounder until refusal. The survey was carried out in the early springs of 2005 and 2006, when the ground was relatively wet so that the rod penetrated more easily. The first author carried out this survey for 761 of the points in seven sub-watersheds, while soil depth data for 58 points in the eighth sub-watershed, had been previously collected using the same methods (Williams et al., 2008). The data from these 819 points are designated as the calibration dataset. A further 130 soil depth observations were collected using the same method at randomly distributed locations, at least 50 meters away from the selected sub-watersheds, over the remainder of the watershed. These are designated as the testing dataset (Fig. 3.1).

A wide range of topographic and land cover attributes were chosen as potential regression explanatory variables for the prediction of soil depth. Fifty five topographic variables (Table 3.1) were derived from the 1/3 arc second DEM obtained from the USGS seamless data server, which was projected to a 5 m resolution grid for the derivation of the topographic attributes. Of these, 36 were new topographic attributes that we derived following the approach described in Tarboton and Baker (2008). Ten land cover variables (Table 3.2) were derived from the Landsat TM imagery (path 41 row 30 obtained from the USGS) and an aerial photograph (obtained from NRCS Idaho State Office). Details on the derivation of these geospatial input variables are given in Tesfa et al. (2009).

3.3.2 Statistical Analysis

3.3.2.1 Normalization

Box Cox transformations (Equation (3.1)) were used to transform the measured soil depth (sd) and each explanatory variable so that their distribution was near normal.

Table 3.1 Topographic attributes derived from DEM; derivations and equations are in Tesfa et al. (2009)

Symbol	Description
elv**	Elevation above sea level
sca**	Specific catchment area from the D∞ method. This is contributing area divided by the grid cell size (from TauDEM[a] specific catchment area function)
plncurv**	Plan curvature is the curvature of the surface perpendicular to the direction of the maximum slope (From ArcGIS spatial analysis tools curvature function). A positive value indicates upwardly convex surface; a negative value indicates upwardly concave surface; and zero indicates flat surface
prfcurv	Profile curvature is the curvature of the surface in the direction of maximum slope (From ArcGIS spatial analyst tools curvature function) (Moore et al., 1993, 1991). A negative value indicates upwardly convex surface; a positive value indicates upwardly concave surface and zero indicates flat surface. See Table 29.1)
gncurv	The second derivative of the surface computed by fitting a fourth order polynomial equation to a 3×3 grid cell window (From ArcGIS spatial analyst tools curvature function) (Moore et al., 1993, 1991).
aspg	The direction that a topographic slope faces expressed in terms of degrees from the north (From ArcGIS spatial analyst tools aspect function).
slpg**	Magnitude of topographic slope computed using finite differences on a 3×3 grid cell window (From ArcGIS spatial analyst tools slope function).
ang**	The D∞ flow direction: This is the direction of the steepest outwards slope from the triangular facets centered on each grid cell and is reported as the angle in radians counter-clockwise from east (TauDEM Dinf Flow Directions function).
ad8	D8 Contributing Area: The number of grid cells draining through each grid cell using the single flow direction model (TauDEM D8 Contributing Area function)
sd8	The D8 slope: The steepest outwards slope from a grid cell to one of its eight neighbors reported as drop/distance, i.e. tan of the angle (TauDEM D8 Flow Directions function).
stdist	D8 Distance to Stream: Horizontal distance from each grid cell to a stream grid cell traced along D8 flow directions by moving until a stream grid cell as defined by the Stream Raster grid is encountered (TauDEM Flow Distance to Streams function).
Slpt	D∞ slope (Tarboton, 1997): The steepest outwards slope from the triangular facets centerd on each grid cell reported as drop/distance, i.e. tan of the slope angle (TauDEM Dinf Flow Directions function)
plen	D8 Longest Upslope Length: The length of the flow path from the furthest cell that drains to each cell along D8 flow directions. (TauDEM Grid Network Order and Flow Path Lengths function)
tlen	D8 Total Upslope Length: The total length of flow paths draining to each grid cell along D8 flow directions (TauDEM Grid Network Order and Flow Path Lengths function)
sd8a**	Slope averaged over a 100 m path traced downslope along D8 flow directions (from GRAIP[b], D8 slope with downslope averaging function)
p	The D8 flow direction grid representing the flow direction from each grid cell to one of its adjacent or diagonal neighbors, encoded as 1–8 counter-clockwise starting at east (TauDEM D8 Flow Directions function)
sar	Wetness index inverse: an index calculated as slope/specific catchment area (TauDEM wetness index inverse function)

Table 3.1 (continued)

Symbol	Description
sph8	D8 horizontal slope position
modcurv**	Curvature modeled based on field observed curvature using a regression equation on plan curvature, D8 horizontal slope position, wetness index inverse and general curvature, see Tesfa et al. (2009) for details
lhr*	Longest D∞ horizontal distance to ridge, see Tesfa et al. (2009) for details
shr*	Shortest D∞ horizontal distance to ridge, see Tesfa et al. (2009) for details
ahr*	Average D∞ horizontal distance to ridge, see Tesfa et al. (2009) for details
lhs*	Longest D∞ horizontal distance to stream, see Tesfa et al. (2009) for details
shs*	Shortest D∞ horizontal distance to stream, see Tesfa et al. (2009) for details
ahs*	Average D∞ horizontal distance to stream, see Tesfa et al. (2009) for details
lvr*	Longest D∞ vertical rise to ridge, see Tesfa et al. (2009) for details
svr*	Shortest D∞ vertical rise to ridge, see Tesfa et al. (2009) for details
avr**	Average vertical rise to ridge computed over multiple (D∞) paths from ridge to each point, see Tesfa et al. (2009) for details
lvs**	Longest vertical drop to stream computed over multiple (D∞) paths from point to stream, see Tesfa et al. (2009) for details
svs*	Shortest D∞ vertical drop to stream, see Tesfa et al. (2009) for details
avs*	Average D∞ vertical drop to stream, see Tesfa et al. (2009) for details
lsr*	Longest surface distance to ridge, see Tesfa et al. (2009) for details
ssr*	Shortest surface distance to ridge, see Tesfa et al. (2009) for details
asr*	Average surface distance to ridge, see Tesfa et al. (2009) for details
lss*	Longest surface distance to stream, see Tesfa et al. (2009) for details
sss*	Shortest surface distance to stream, see Tesfa et al. (2009) for details
ass*	Average surface distance to stream, see Tesfa et al. (2009) for details
lps*	Longest Pythagoras distance to stream, see Tesfa et al. (2009) for details
sps*	Shortest Pythagoras distance to stream, see Tesfa et al. (2009) for details
aps*	Average Pythagoras distance to stream, see Tesfa et al. (2009) for details
lpr*	Longest Pythagoras distance to ridge, see Tesfa et al. (2009) for details
spr*	Shortest Pythagoras distance to ridge, see Tesfa et al. (2009) for details
apr*	Average Pythagoras distance to ridge, see Tesfa et al. (2009) for details
lsph∞*	D∞ Longest horizontal slope position, see Tesfa et al. (2009) for details
ssph∞*	D∞ Shortest horizontal slope position, see Tesfa et al. (2009) for details
asph∞*	D∞ Average horizontal slope position, see Tesfa et al. (2009) for details
lspv**	Longest vertical slope position computed as longest vertical drop divided by the longest vertical drop plus longest vertical rise to ridge, see Tesfa et al. (2009) for details
sspv*	Shortest vertical slope position, see Tesfa et al. (2009) for details
aspv*	Average vertical slope position, see Tesfa et al. (2009) for details
lspp*	Longest Pythagoras slope position, see Tesfa et al. (2009) for details
sspp*	Shortest Pythagoras slope position, see Tesfa et al. (2009) for details
asp*	Average Pythagoras slope position, see Tesfa et al. (2009) for details
lspr*	Longest slope position ratio, see Tesfa et al. (2009) for details
sspr*	Shortest slope position ratio, see Tesfa et al. (2009) for details
aspr*	Average slope position ratio, see Tesfa et al. (2009) for details

* New topographic variables derived using enhanced terrain analysis

**Topographic variables selected for modeling soil depth

[a] TauDEM is the Terrain Analysis Using Digital Elevation Models software (http://www. engineering.usu.edu/dtarb/taudem)

[b] GRAIP is the Geomorphologic Road Analysis Inventory Package software (http://www. engineering.usu.edu/dtarb/graip)

Table 3.2 Landsat remote sensing image based data and their descriptions; equations are in Tesfa et al. (2009)

Symbol	Description
lc	Land cover map derived from Landsat TM image using supervised classification (this method is described in Section 10.2.4) in ERDAS IMAGINE. Land cover is represented as a numerical value encoded as follows: 1 Road, rock outcrop and bare, 2 Grass, 3 Mixed grass and shrub, 4 Shrub, riparian and deciduous forest, 5 Coniferous forest
*pc1***	First principal component from ERDAS IMAGINE principal component analysis of Landsat Thematic Mapper bands 1, 2, 3, 4, 5, and 7
pc2	Second principal component derived from principal component transformation of Landsat TM image in ERDAS IMAGINE (Jensen, 1996)
pc3	Third principal component derived from principal component transformation of Landsat TM image in ERDAS IMAGINE (Jensen, 1996)
tc1	First tasseled cap component derived from tasseled cap transformation of Landsat TM image in ERDAS IMAGINE (represents brightness)
tc2	Second tasseled cap component derived from tasseled cap transformation of Landsat TM image in ERDAS IMAGINE (represents greenness)
tc3	Third tasseled cap component derived from tasseled cap transformation of Landsat TM image in ERDAS IMAGINE (represents wetness)
ndvi	Normalized difference vegetation index calculated in ERDAS IMAGINE (Jensen, 1996) (see Table 29.1 and Section 20.2.3)
vi	Vegetation index calculated in ERDAS IMAGINE (Jensen, 1996)
cc	Canopy cover index calculated in ERDAS IMAGINE (Zhu and Band, 1994)

** Land cover variables selected for modeling soil depth

$$t(x) = \frac{(x^{\lambda} - 1)}{\lambda} \tag{3.1}$$

Here, $t(x)$ denotes the transform of variable x with transformation parameter λ. λ was selected to maximize the Shapiro-Wilks Normality Test W-statistic as implemented in R (R Development Core Team, 2007).

3.3.2.2 Model

We applied Generalized Additive Models (GAM) (Hastie and Tibshirani, 1990) to predict soil depth using the explanatory variables. GAM is a statistical approach that generalizes multiple regression by replacing linear combinations of the explanatory variables with combinations of nonparamtertic smoothing or fitting functions, estimated through a backfitting algorithm. The GAM model is:

$$E(sd|x_1, x_2, \ldots, x_p) = \alpha + f_1(x_1) + f_2(x_2) + \cdots + f_p(x_p) \tag{3.2}$$

where, $x_1, x_2, \ldots, x_p$ are explanatory variables (predictors), sd is soil depth (response variable) and f_i are non-parametric smoothing splines that relate sd to the $x_1, x_2, \ldots, x_p$. The model assumes that the mean of sd is an additive combination of nonlinear functions of the explanatory variables $x_1, x_2, \ldots, x_p$. We used the GAM package as implemented in R (R Development Core Team, 2007).

3.3.2.3 Variable Selection and Model Complexity

Questions in developing a predictive regression model include which potential explanatory variables to use and what to do about interdependent explanatory variables. Many of the explanatory variables that we derived from the DEM (Table 3.1) were variants on similar quantities, so we were specifically concerned about the effect of explanatory variable correlation on model prediction error. A correlation matrix giving the cross correlation between all 65 explanatory variables was computed using all 819 data points in the calibration dataset. Random Forest (Breiman, 2001), a classification and regression package (this is described in Section 15.2.3) in R (R Development Core Team, 2007), was used to calculate a measure of explanatory variable importance (see Section 29.2.3.2) for the prediction of soil depth. Due to randomness in the Random Forest method the variable importance varies slightly each time it is run. We therefore ran Random Forest 50 times using all 819 data points in the calibration dataset with all 65 potential explanatory variables with soil depth as the response variable and averaged variable importance across these runs. Explanatory variables were then ordered based upon their importance measures.

The number of explanatory variables in a model is a measure of model complexity. We used the correlation matrix, together with the Random Forest importance values to develop sets of explanatory variables representing models of differing complexity by eliminating the variable of lesser importance from pairs of variables with correlation above a designated threshold (from 0.15 to 0.9 in increments of 0.05). Variables were filtered out working sequentially from high to low correlation until no pairs with correlation greater than the threshold remained. Lower thresholds result in fewer variables, so a range of models with differing complexity were developed. This approach reduced the correlation between variables selected for inclusion in a model. Models of differing complexity were also constructed using explanatory variables directly from the variable list ordered by importance. Figure 3.2 shows the explanatory variables with importance values greater than or equal to 0.009, ordered based on their average importance values from 50 RF runs with all 819 calibration data points and all 65 explanatory variables.

To evaluate appropriate model complexity, we randomly split the calibration sample of 819 data points into two parts, designated as the training and validation sets. The separate testing dataset of 130 points randomly distributed across the watershed was withheld from this process, so that it could be used for evaluation of the final model. GAM was applied, using the training data set of 614 data points to fit the models. Prediction error was computed for both the training and validation data set. The validation data set prediction error provides an out of sample estimate appropriate for trading off variance due to complexity with bias due to too few explanatory variables (see e.g. Hastie et al., 2001). The results from this analysis allowed us to select the explanatory variables and degree of model complexity.

3.3.2.4 Calibration and Testing

Once the explanatory variables and model with appropriate complexity had been selected, GAM was applied using the full calibration data set as input. It was used

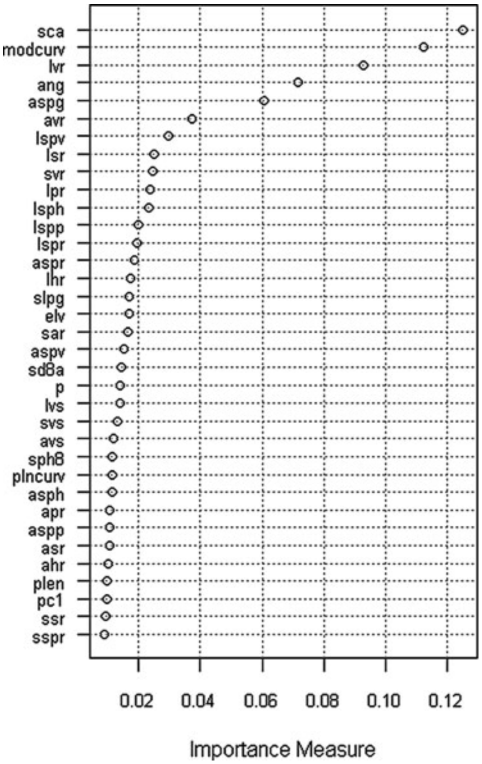

Fig. 3.2 Variable importance measure of the Box Cox transformed explanatory variables averaged from 50 RF model runs

to predict soil depth for the entire watershed. We then compared the testing dataset with the GAM soil depth values at testing locations using the Nash-Sutcliffe efficiency coefficient (NSE), which is a measure widely used to quantitatively assess the predictive accuracy of a model.

$$\mathrm{NSE} = 1 - \frac{\sum(\mathrm{SD_o} - \mathrm{SD_p})^2}{\sum(\mathrm{SD_o} - \mathrm{SD_m})^2} \tag{3.3}$$

where; $\mathrm{SD_o}$, $\mathrm{SD_p}$, and $\mathrm{SD_m}$ are observed (measured), predicted, and mean of observed (measured) soil depths respectively.

3.4 Results and Discussion

3.4.1 Variable Selection and Model Complexity

Figure 3.3 shows the variation of mean square prediction error for training and validation datasets versus model complexity in terms of the number of input variables. The continuous lines in this figure are from models developed using explanatory variables selected based on Random Forest importance directly. There is a new GAM model for each additional input variable. The symbols in this figure are from

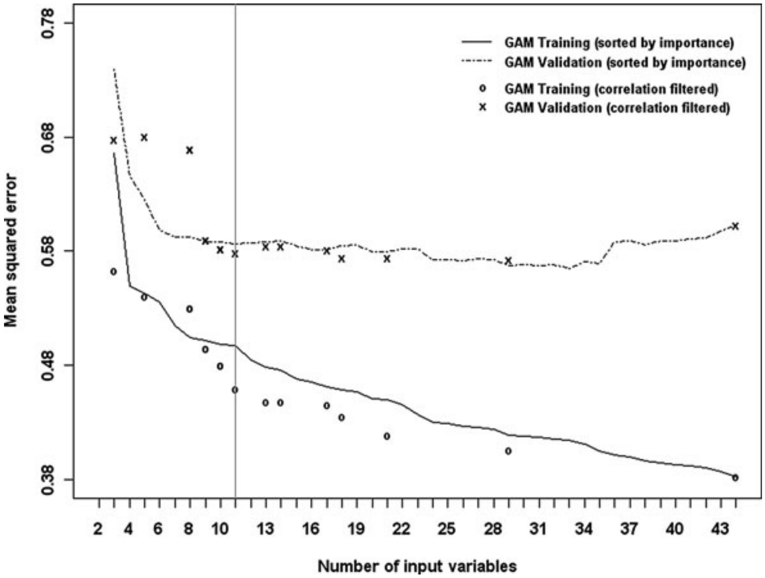

Fig. 3.3 Number of input variables (Model complexity) vs. mean squared error. Explanatory variables selected directly using importance (*continuous*) and filtered by correlation (*symbols*)

models developed using cross correlation as a filter to reduce inter-dependence among explanatory variables. There is a new GAM model with different number of input variables for each correlation threshold. Figure 3.3 reports training and validation errors separately.

For both the importance-selected and correlation-filtered models, the training error decreases progressively as additional input variables are added while the validation error decreases initially and then flattens out and starts to increase. The use of correlation-filtered explanatory variables resulted in lower error. The validation error starts to increase for complexity more than 11 correlation-filtered variables (Fig. 3.3). Although there are fluctuations on validation MSE that go slightly below the 11 variable complexity, for 18 and 21 input variables, in our judgment the point of diminishing returns has been reached at 11 input variables. Consequently we selected 11 correlation-filtered explanatory variables as representing the optimum GAM complexity for this dataset. Tables 3.1 and 3.2 list all the topographic and land cover explanatory variables derived for modeling soil depth. Variables derived using new DEM analysis methods are identified with single asterisk (*) and variables selected by this variable selection procedure are identified by double asterisks (**). Ten of the 11 selected explanatory variables are topographic variables, with three (avr, lspv, lvs), variables derived using the new DEM analysis methods.

3.4.2 Model Evaluation

Based on the selection of 11 correlation-filtered explanatory variables above, GAM was applied to the full calibration set of 819 data points. Figure 3.4 shows the scatter

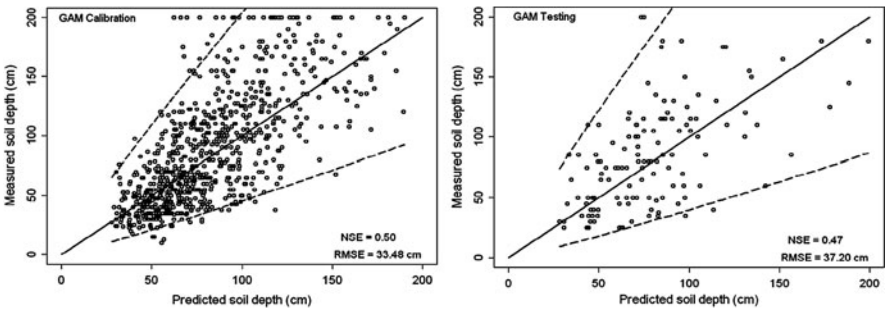

Fig. 3.4 Predicted soil depth vs. measured soil depth with plus and minus two standard error for calibration (*left*) and testing (*right*) data

plots of predicted versus measured soil depth for the calibration (left) and testing (right) data and their Nash-Sutcliffe Efficiency (NSE) and root mean squared errors (RMSE) after transforming back into space of soil depth. The testing data was not used at all in model development. In this figure, the diagonal (central) lines represent the 1:1 line (predicted = observed). The two diverging dash lines, above and below the 1:1 line, show the predicted soil depth plus and minus two standard errors representing 95% confidence intervals. These lines diverge as a result of the Box-Cox back transformation (Figs. 3.4 and 3.5).

Figure 3.5 shows the soil depth map created using GAM at 5 meter grid scale which improves the scale of soil depth representation as compared to the map unit based soil depth maps that can be created using conventional soil survey approach

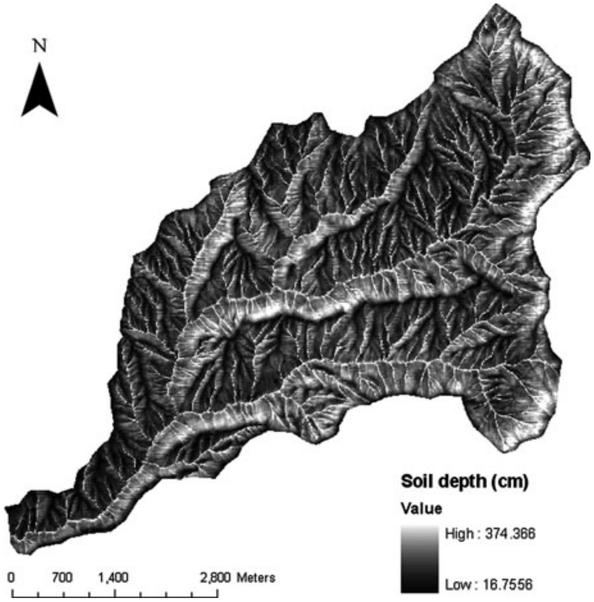

Fig. 3.5 Soil depth map predicted using GAM model

(see Sections 29.2.3.1 and 29.3.2). This models the ridges (convex areas) and south facing slopes as having shallower soils compared to the valleys (concave areas) and the north facing slopes respectively. This agrees with existing literature (e.g. Dietrich et al., 1995). As compared to soil depth maps created using conventional.

3.5 Conclusions

A statistical model has been developed that predicts soil depth using topographic and land cover attributes. The topographic attributes were found to be more important than the land cover attributes in predicting the soil depth. The model was able to explain about 50% of the measured soil depth variability in an out of sample test. New topographic variables derived from the DEM played an important role in this model. Considering the uncontrolled uncertainties due to the complex local variation of soil depth, DEM errors and GPS reading errors, this is considered an important improvement towards solving the need for distributed soil depth information in distributed hydrological and ecological modeling.

Acknowledgments The data collection part of this research was funded by the Utah Drought Management project, a project of USDA-CSREES Special Research Grant no. 2005-34552-15828 from the USDA Cooperative State Research, Education, and Extension Service. The modeling part was funded by the Inland Northwest Research Alliance (INRA). Field assistance was provided by Pam Aishlin and other students at Boise State University.

References

Breiman, L., 2001. Random forests. Machine Learning, 45:5–32.

Dietrich, W. E., Reiss, R., Hsu, M.-L., and Montgomery, D. R., 1995. A process-based model for colluvial soil depth and shallow landsliding using digital elevation data. Hydrological Processes, 9:383–400.

Freer, J., McDonnell, J. J., Beven, K. J., Peters, N. E., Burns, D. A., Hooper, R. P., Aulenbach, B., and Kendall, C., 2002. The role of bedrock topography on subsurface storm flow. Water Resources Research, 38(12):1269–1285.

Hastie, T., and Tibshirani, R., 1990. Generalized Additive Models. Chapman and Hall, London.

Hastie, T., Tibshirani, R., and Friedman, J., 2001. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 533 pp., Springer, New York.

Jensen, J. R., 1996. Introductory to Digital Image Processing: A Remote Sensing Perspective, 2 ed., Prentice-Hall, Englewood Cliffs, NJ.

Johnson, K.M., Lewis, R.S., Bennet, E.H., and Kiilsgaard, T.H., 1988. Cretaceous and tertiary intrusive rocks of south-central Idaho, pp. 55–86. In: Link, P.K., and Hackett, W.R. (eds.), Guidebook to the Geology of Central and Southern Idaho: Idaho Geological Survey, Bulletin 27, University of Idaho, Moscow, Idaho.

Lewis, R.S., Kiilsgaard, T.H., Bennet, E.H., and Hall, W.H., 1987. Lithologic and chemical characteristics of the central and southeastern part of the southern lobe of the Idaho Batholith, pp. 171–196. In: Vallier, T.L., Brooks, H.C. (eds), Geology of the Blue Mountains region of Oregon, Idaho, and Washington – the Idaho Batholith and Its Border Zone: U.S. Geological Survey Professional Paper 1436.

McNamara, J.P., Chandler, D., Seyfried, M., and Achet, S., 2005. Soil moisture states, lateral flow, and streamflow generation in a semi-arid, snowmelt-driven catchment. Hydrological Processes 19(20):4023–4038.

Moore, I.D., Gessler, P.E., Nielsen, G.A., and Peterson, G.A., 1993. Soil attribute prediction using terrain analysis. Soil Science Society of America Journal 57(2):443–452.

Moore, I.D., Grayson, R.B., and Ladson, A.R., 1991. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrological Processes 5(1):3–30.

R Development Core Team, 2007. R: A Language and Environment for Statistical Computing, edited, R Foundation for Statistical Computing, Vienna, Austria.

Stieglitz, M., Shaman, J., McNamara, J., Engel, V., Shanley, J., and Kling, G.W., 2003. An approach to understanding hydrologic connectivity on the hillslope and the implications for nutrient transport. Global Biogeochemical Cycles 17(4):1105–1120.

Tarboton, D.G., and Baker, M.E., 2008. Towards an algebra for terrain-based flow analysis, pp. 167–194. In: Mount, N.J., Harvey, G.L., Aplin, P., and Priestnall, G., (eds.), Representing, Modeling and Visualizing the Natural Environment: Innovations in GIS 13. CRC Press, Boca Raton, FL.

Tesfa, T.K., Tarboton, D.G., Chandler, D.G., and McNamara, J.P., 2009. Modeling soil depth from topographic and land cover attributes, Water Resources Research 45:W10438, doi: 10.1029/2008WR007474.

USDA, 1997. Soil survey of the Boise Front Project, Idaho: Interim and supplemental report, Boise, Idaho.

Williams, C.J., 2005, Characterization of the spatial and temporal controls on soil moisture and streamflow generation in a semi-arid headwater catchment, Masters Thesis, Boise State University.

Williams, C.J., McNamara, J.P., and Chandler, D.G., 2008. Controls on the temporal and spatial variability of soil moisture in a mountainous landscape: the signatures of snow and complex terrain. Hydrology and Earth System Sciences 5(4):1927–1966.

Zhu, A.X., 1997. A similarity model for representing soil spatial information. Geoderma 77(2–4):217–242.

Zhu, A.X., and Band, L.E., 1994. A knowledge-based approach to data integration for soil mapping. Canadian Journal of Remote Sensing 20:408–418.