THE EFFECTS OF DIGITAL ELEVATION MODEL RESOLUTION ON THE CALCULATION AND PREDICTIONS OF TOPOGRAPHIC WETNESS INDICES

by

DAMION RYAN DROVER

(Under the Direction of C. Rhett Jackson)

ABSTRACT

Variable source areas of pollution can be mapped using topographic wetness indices (TWI) derived from digital elevation models. These maps come in an infinite number of resolutions, and affect the calculation of terrain attributes depending on the resolution used. The purpose of this project was to study the effects of DEM resolution on the calculation of TWI maps on a small Upper Coastal Plain watershed. The maps were then used to make predictions of soil attributes across the watershed. It was found that the optimal TWI resolution depended on the soil attribute. For example, the mineral soil horizons correlated best to the 30m TWI, while the organic soil horizon correlated best to the 5m TWI. Prediction maps were made of the soil attributes by themselves, and also cokriged with TWI. These maps and methods are potentially useful to forest managers and to researchers for future projects.

INDEX WORDS: Digital Elevation Models, Topographic Wetness Index, Downslope Index, Variable Source Areas, Soils, Biofuels, Resolution, Geographic Information System, Prediction Map, Basin Delineation

THE EFFECTS OF DIGITAL ELEVATION MODEL RESOLUTION ON THE CALCULATION AND PREDICTIONS OF TOPOGRAPHIC WETNESS INDICES

by

DAMION RYAN DROVER

BS, University of Wisconsin-Oshkosh, 2008

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2011

THE EFFECTS OF DIGITAL ELEVATION MODEL RESOLUTION ON THE CALCULATION AND PREDICTIONS OF TOPOGRAPHIC WETNESS INDICES

by

DAMION RYAN DROVER

Major Professor:	C. Rhett Jackson
Committee:	Daniel Markewitz
	Lan Mu

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2011

DEDICATION

To my grandfather, Dr. Thomas Kelley, for inspiration and support all these years.

ACKNOWLEDGEMENTS

I would like to thank my committee: Rhett Jackson, Dan Markewitz, and Lan Mu, for advice and support on this project.

I want to give a special thanks to Enhao Du, at the Savannah River Site. He was always eager to help me in any way he could, and there are parts of my thesis that I would still be fighting with if it wasn't for him.

In the same vein, Dr. John Dowd spent hours writing my Downslope Index program, and put up with my lack of understanding of virtually anything he said to me. I could not have included that index if it wasn't for him. I'm very grateful for his help.

Finally, thanks to all the others who helped me on my project: Louise Jacques, Patrick Bussell, Ben Morris, Bob Bahn, Erin Harris, and Meg Williamson.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

One of the largest exports in the Southeast U.S. is forest products. Interest in biofuels using forest biomass has increased recently, leading to more research into better forest management BMPs. The USDA Forest Service, along with the Oak Ridge National Laboratory, University of Georgia and Oregon State University are researching the impacts of intensive forest management for biofuels on water quality and quantity at the Savannah River Site in South Carolina. Surface runoff of saturated areas, transporting excess nutrients and contaminants, is a potential water quality issue under investigation. Detailed maps of variable source areas and soil characteristics would therefore be helpful prior to treatment.

The availability of remotely sensed and computed digital elevation models (DEMs) and spatial analysis tools make it easy to calculate terrain attributes. These terrain attributes can be used in models to predict saturated areas or other attributes in the landscape. With laser altimetry, an area can be flown to produce very high resolution data, and the resulting data can be resampled into any resolution of DEM desired. Additionally, there exist many maps that are in various resolutions of DEM, such as those acquired from the U.S. Geological Survey.

Problems arise when using maps derived from different resolution DEMs. For example, saturated areas can be under or overestimated depending on the resolution used. The purpose of this study was to examine the effects of DEM resolution on the calculation of topographic wetness indices used to predict variable source areas of saturation, and to find the best resolutions to produce prediction maps of soil attributes like nitrogen, carbon, bulk density and soil texture for low-relief, humid-temperate forested hillslopes.

Topographic wetness indices were calculated based on the derived terrain attributes, slope and specific catchment area, from five different DEM resolutions. The DEMs were resampled from LiDAR, which is a laser altimetry remote sensing method, obtained from the USDA Forest Service at Savannah River Site. The specific DEM resolutions were chosen because they are common grid cell sizes (10m, 30m, and 50m) used in mapping for management applications and in research. The finer resolutions (2m and 5m) were chosen for the purpose of determining how finer resolutions performed compared with coarser resolutions at predicting wetness and related soil attributes.

The wetness indices were compared across DEMs and with each other in terms of quantile and distribution differences, then in terms of how well they each correlated with measured soil attributes. Spatial and non-spatial analyses were performed, and predictions using regression and geostatistics were examined for efficacy relative to each DEM resolution. Trends in the raw data and analysis results were also revealed.

1.2 Literature Review

Terrestrial surface run-off can transport excess nutrients, sediment and contaminants to streams, rivers and lakes (Walter 2000). Run-off is produced primarily by two different processes. Hortonian overland flow occurs when the infiltration capacity of the soil is exceeded by rainfall intensity (Horton 1933, Betson 1964). This occurs usually on semi-arid and Midwestern agricultural soils. In humid, well-vegetated areas, infiltration capacity is rarely exceeded, so the Horton flow model doesn't work (Dunne and Black 1970, Dunne et al. 1975, Ward 1984, Walter 2000). Instead, the soil is saturated until its storage capacity is exceeded. This is called saturation excess, and can be responsible for subsurface flow, return flow and direct precipitation onto saturated areas (Dunne et al. 1975). When rain falls on an area that has exceeded its storage capacity, it will produce run-off in much the same way as Hortonian flow. As the soil gets saturated, the saturated area will tend to get larger. The area expands during storm and subsurface flow and contracts during drier periods (Hewlett and Hibbert 1967). This can be on the scale of storms or seasons (Dunne and Black 1970, Dunne et al. 1975, Frankenberger 1999). The saturated area

is thus called a variable source area (VSA) (or variable contributing area, as in Beven and Kirkby 1979), because it can be the source of runoff and it varies through space and time.

VSAs can be caused by shallow soils overlying relatively impermeable material, converging hillslopes, concave plan or profile curvature, or shallow slopes (Walter et al. 2000). The last three cases can be predicted from maps derived from a digital elevation model (DEM), making it possible to map where VSAs are most likely to occur on the landscape. The locations and extent of shallow soils overlying impermeable material cannot be derived from a DEM. These include perched water tables caused by shallow lenses of relatively impermeable material (O'Loughlin 1981). The existence of shallow soils and perched water tables confound and underestimate predictions of the spatial distribution of VSAs (Beven 1997). Also, the existence of deep permeable soils in areas with low slope or concave curvature cannot be derived from DEMs either (Quinn et al. 1991), which may cause an overestimation of VSA distribution (Beven 1997). Despite the shortcomings of using topographic attributes to discern where VSAs are most probable, the models used can be considerably accurate (Quinn et al. 1991, Western et al. 1999, Agnew et al. 2006, Grabs et al. 2009), and the inputs can be simple, inexpensive and readily available.

One model used extensively for its simplicity is the TOPMODEL (Beven and Kirkby, 1979) topographic wetness index (TWI):

$$\mathrm{Ln}(a/\tan\beta)$$

where "a" is the specific catchment area and tanβ is the local slope. The specific catchment area is conceptually an integration of the surface and subsurface drainage from an upslope area per unit width of contour. (Speight 1980, Moore et al. 1991). The local slope is considered to be, in terms of gridded DEMs, the maximum difference in elevation from one cell to an adjacent cell divided by the distance between the centers of the two cells.

Each parameter has its own assumptions (Beven 1997): 1) that subsurface flow can be approximated by a succession of steady-state flow rates, and that the input to the saturated zone is uniform across the catchment, affecting a. 2) The slope is a good representation of hydraulic gradient,

affecting tanβ. These assumptions, like most assumptions in hydrological models, are not met everywhere. Beven (1997) gives more detail about the validity and seriousness of the assumptions underlying TOPMODEL, and the crudeness of such an approach. However, he notes that "the approach offers simplicity, inexpensive calculations and a minimum number of effective parameters". In addition, TWI represents a theoretical distribution of flow that can be calculated for any catchment, and can be used to predict VSAs, saturation excess overland flow and subsurface flow (Quinn 1995).

The idea behind the TWI is to provide an index of relative saturation across a basin, where it assumes hydrological similarity among areas with the same index value (Beven and Kirkby 1979, Wolock 1995, Frankenberger 1999). TOPMODEL uses the natural links between hydrology and geomorphology of a catchment to make predictions about one process (hydrology) from an index of measures from another dataset (topography) (Quinn and Beven 1993).

There have been modifications and hybrids to TWI. Some have used TWI with the Soil Conservation Service (SCS) curve number method (Lyon et al. 2004) for example. Schneiderman et al. (2007) altered the Generalized Watershed Loading Function (Haith and Shoemaker, 1987) into the Variable Source Loading Function model using the SCS–CN equation, and spatially distributed the runoff response according to ln(a/tanβ) to simulate saturation excess runoff from VSAs. Agnew et al. (2006) used TWI in conjunction with monthly saturation probability maps and were able to predict actual saturation with relative differences consistently less than ten percent between observed and simulated soil moisture.

Specific catchment area (SCA) is calculated using the total area draining to a cell, divided by the length of one of the sides of the cell (contour length). The direction of tanβ (flow direction) is also considered when calculating the total draining area. There are a number of ways to calculate the flow direction, and several papers have discussed the relative efficacy and applications of the choice of method. One common method is to use the single flow algorithm (SFA). To calculate SCA using SFA for a target cell in a regular raster grid, the adjacent cell with the maximum drop in elevation is used to determine flow direction. All the cells upslope, plus the target cell, are counted and multiplied by their

areas (which is the contour length squared) (O'Callaghan and Mark, 1984). This method is fairly accurate at finer DEM resolutions, but as grids become coarser than 50m errors tend to increase, especially on divergent hillslopes (Quinn 1991). Instead of routing all flow to one downslope cell, the multiple flow algorithm partitions the flow to all downslope cells, based on a weighting scheme. That weighting scheme is in proportion to the tanβ local slope of each downslope cell (Quinn 1991). Tarboton (1997) criticized this method as being too dispersive to be used effectively to calculate the upslope area. He instead proposed a different method that uses eight triangular facets centered on the center cell of a 3x3 grid. This method minimizes the grid bias that occurs as a result of the orientation of a regular grid, and enables the program to fit planes to the cells. Flow direction is calculated using a proportion of a facet based on the steepest downslope cells, and in this way only one or two downslope cells gets all the flow (Figure 1.1). Upslope area is then calculated as the target cell's area plus any neighbors that have some fraction draining to the target cell (Tarboton 1997). Tarboton named this method "D-inf" because it uses continuous direction instead of the discrete cardinal and diagonal directions used in other methods.

Since the local slope (tanβ) only describes the slope over what is essentially a point, it does not take into account what happens downslope to influence flow. To get around this limitation, Hjerdt et al. (2004) developed what is known as the downslope index, where the tanβ parameter is replaced with $\tan\alpha_d$ (Figure 1.2). $\tan\alpha_d$ is based on findings from Speight (1980) who found that both upslope and downslope conditions affect drainage, and from Crave and Gascuel-Odoux (1997) that calculated head difference from the stream to each point and used that information as a predictor of soil class. Instead of using the stream as an exit point, Hjerdt et al. (2004) used an arbitrary loss of head (d) based on relief, DEM resolution, and local soil transmissivity. $\tan\alpha_d$ was calculated using d/L_d, where L_d was the horizontal length one would have to travel to drop d units of elevation (Figure 1.2). They simulated a hillslope from a modified sine wave. They then simulated groundwater tables calculated from tanβ and $\tan\alpha_d$ to illustrate groundwater behavior in convergent and divergent parts of the hillslope. They found the $\tan\alpha_d$ to be more realistic in terms of representing a groundwater table. It was also found to smooth out

downslope areas as d increased, due to its integrative nature, instead of mimicking every detail of the topography, as happens with tanβ. As d approached zero, $\tan\alpha_d$ approached tanβ.

The downslope wetness index (DWI = $\ln(a/\tan\alpha_d)$ has been used with success in several studies, such as Seibert et al. (2007) and Inamdar and Mitchell (2007). However, Guntner et al. (2004) found that, in steeper terrain, local slope (tanβ) outperformed $\tan\alpha_d$ at delineating saturated areas. Sorensen et al. (2006) found that the success of the parameter depended largely on the variable being measured or predicted, but in general the local slope outperformed $\tan\alpha_d$. Grabs et al. (2009) found that DWI was more accurate than TWI for predicting the location and structure of wetlands.

Differences among DEM resolutions can have impacts on the calculation of terrain attributes. Specific contributing area is directly affected by resolution, because with equal area larger grid cells (and thus longer contour lengths) will produce smaller SCA values. Vieux (1993), Thompson et al. (2001), and others found that lower mean and maximum gradient result from lower resolution DEMs. Quinn et al. (1995) concluded that larger grid cells are unrepresentative and difficult to validate in small catchments, the shape and size of a catchment may be altered by excessively large cells, and lower resolution results in bias toward large TWI values. Straumann et al. (2007) found that gentler slopes were less susceptible to decreased resolution than steeper slopes. They found loss of representation of channelized flow at 40m resolution, and lower overall SCA. Both gradient and SCA decreased in range with decreased resolution (Straumann 2007).

Few studies have compared DEM resolutions of less than 10m. Zhang and Montgomery (1994) used DEMs at 2m, 4m, 10m, 30m and 90m resolutions on 30-40° slopes. They found that the 10m resolution improved the modeling of run-off generation processes from 30m resolution, but that no significant improvement was found at resolutions finer than 10m. They suggested that the resolution should be chosen at the scale of the process being modeled. Lane et al. (2004) used 2m, 4m, 8m, 16m, 32m and 64m DEMs and showed that areas of saturation tended to increase, while channel definition decreased with decreasing resolution. Kienzle (2004) used grid sizes ranging from 5m – 100m on flat, moderately sloped and steep relief areas, and discovered that the optimal resolution depended on the

terrain attribute, but was generally within the range of 5m – 20m. As with other studies, mean slope (%) decreased with decreasing resolution.

Soil moisture is related to the mobility and transformation of soil physical and chemical attributes. Whelan and Gandolfi (2002) used TWI, along with Monte Carlo simulations, to predict the spatial distribution of soil organic carbon (SOC) and denitrification. They used the relationship between oxidizable SOC and denitrification spatial variation, and the assumption of correlation between SOC and soil wetness as reasoning for using TWI. Also, the simple parameters in TWI allowed the iterations necessary for Monte Carlo simulations.

Moore et al. (1993) found fairly strong significant correlations between TWI and several soil attributes (A horizon thickness, sand and silt concentration, SOM concentration, pH, and extractable P) in a small Colorado catchment. In the Catskills, however, Johnson et al. (2000) found poor correlations between TWI and the same soil attributes. They proposed that the topographic dependence of the soil chemical properties was at a smaller scale than the size of the DEMs (5m) used in their study as a possible reason for the poor correlations. Gessler et al. (1995) showed strong correlations between TWI and several soil attributes. Zinko et al. (2005) found a positive linear relationship between TWI and plant species richness in boreal Sweden. They also found that depth to groundwater correlations with TWI were strong. These and other studies are good evidence that topographic attributes can be used to predict soil attributes, and that scale and DEM resolution are factors to consider.

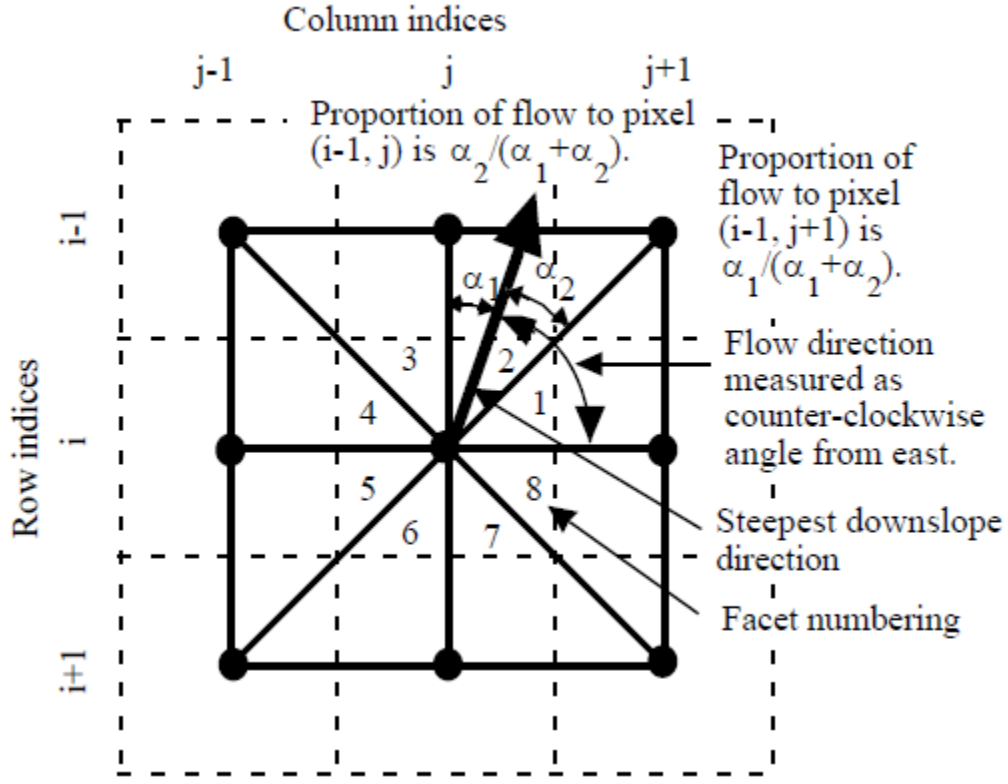

Figure 1.1. Tarboton's facets and method for calculating "D-inf". From: Tarboton (1997)

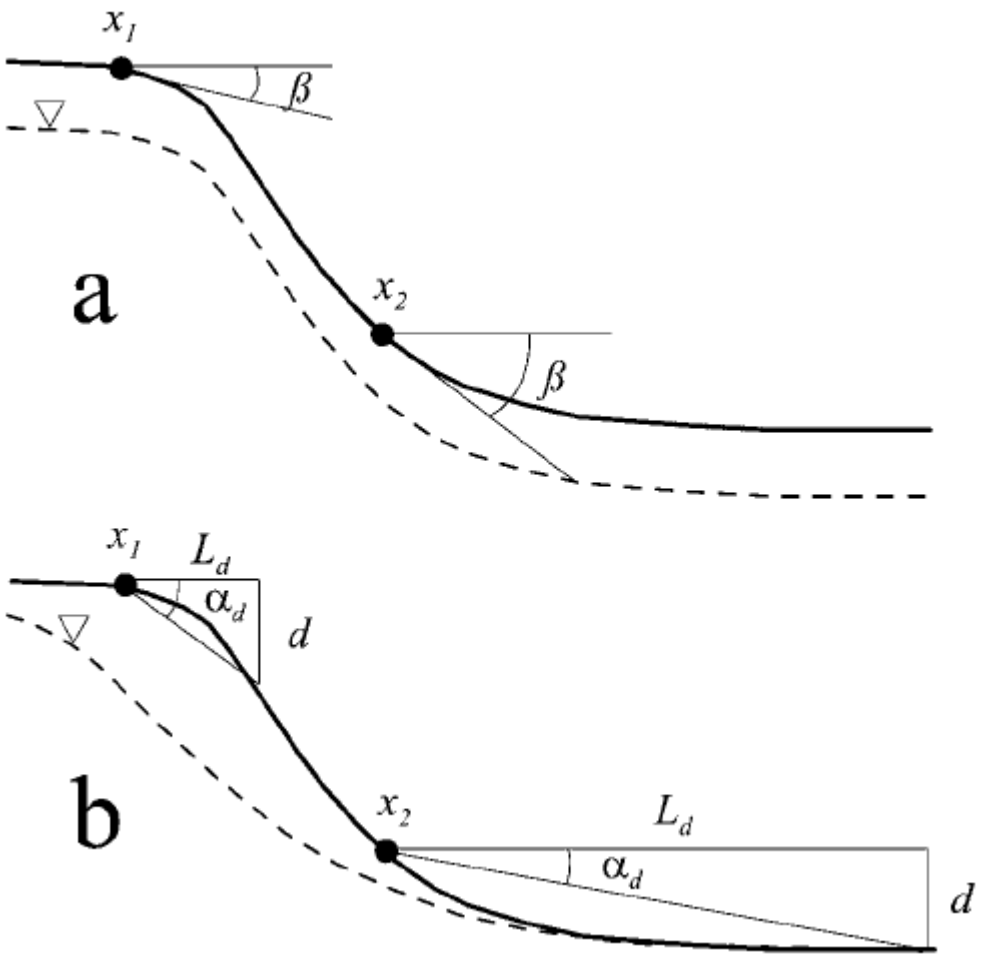

Figure 1.2. Illustration of differences between (a) the local slope and (b) the downslope index. The dotted line is the hydraulic gradient estimated using each method. From: Hjerdt (2004).

1.3 Thesis Outline

Chapter 2 is a supporting chapter that describes the study areas. It highlights problems encountered while attempting to choose the correct DEM for basin delineation. It also contains the soils and vegetation description, and includes maps of each of the study areas.

Chapter 3 was written to stand alone, so it includes some redundant information from the first two chapters. Its purpose was to examine the effects of DEM resolution on the calculation of topographic wetness indices, and the subsequent abilities of different resolution indices to predict soil attributes. Chapter 3 takes a global non-spatial approach to the data, while Chapter 4 considers the locations of

sample points as explanatory variables to help improve the index predictions of soil attributes. Chapter 5 summarizes the results of the study and provides recommendations for future research.

1.4 Literature Cited

Agnew, L.J., S. Lyon, P. Gérard-Marchant, V.B. Collins, A.J. Lembo, T.S. Steenhuis, and M.T. Walter. 2006. Identifying hydrologically sensitive areas: Bridging the gap between science and application. *Journal of Environmental Management* 78: 63-76.

Betson, R.P. 1964. What is watershed runoff? *Journal of Geophysical Research* 69(8): 1541-1552.

Beven K.J., M.J. Kirkby. 1979. A physically-based, variable contributing area model of basin hydrology. *Hydrological Science Bulletin* 24: 43-69.

Beven, K., 1997. Topmodel: a critique. *Hydrological Processes* 11: 1069–1085.

Crave, A., and C. Gascuel-Odoux. 1997. The influence of topography on time and space distribution of soil surface water content. *Hydrological Processes* 11: 203–210.

Dunne T., R.D. Black. 1970. Partial area contributions to storm runoff in a small New England watershed. *Water Resources Research* 6: 1296–1311.

Dunne, T., T. R. Moore, and C. H. Taylor. 1975. Recognition and prediction of runoff-producing zones in humid regions. *Hydrological Sciences Bulletin* 20: 305-327.

Frankenberger J.R., E.S. Brooks, M.T. Walter, M.F. Walter, T.S. Steenhuis. 1999. A GIS-based variable source area model. *Hydrological Processes* 13(6): 804–822.

Gessler, P.E., I.D. Moore, N.J. McKenzie, P.J. Ryan. 1995. Soil–landscape modeling and spatial prediction of soil attributes. *International Journal of Geographical Information Systems* 4: 421–432.

Grabs, T., J. Seibert, K. Bishop, and H. Laudon. 2009. Modeling spatial patterns of saturated areas: a comparison of the topographic wetness index and a dynamic distributed model. *Journal of Hydrology* 373: 15-23.

Guntner, A., J. Seibert, and S. Uhlenbrook. 2004. Modeling spatial patterns of saturated areas: an evaluation of different terrain indices. *Water Resources Research* 40: W05114, doi:10.1029/2003WR002864.

Haith, D.A. and L.L. Shoemaker. 1987. Generalized Watershed Loading Functions for stream flow nutrients. *Water Resources Bulletin* 23(3): 471–478.

Hewlett, J.D. and A.R. Hibbert. 1967. Factors affecting the response of small watersheds to precipitation in humid regions. In: *Forest Hydrology* (eds. W.E. Sopper and H.W. Lull). Pergamon Press, Oxford. 275-290.

Hjerdt, K.N., J.J. McDonnell, J. Seibert, and A. Rodhe. 2004. A new topographic index to quantify downslope controls on local drainage. *Water Resources Research* 40: W05602, doi:10.1029/2004WR003130.

Horton, R.E. 1933. The role of infiltration in the hydrological cycle. *Transactions of the American Geophysical Union* 14: 446-460.

Inamdar, S.P. and M.J. Mitchell. 2007b. Contributions of riparian and hillslope waters to storm runoff across multiple catchments and storm events in a glaciated forested watershed. *Journal of Hydrology* 341: 116–130.

Johnson, C., J.J. Ruiz-Mendez, and G.B. Lawrence. 2000. Forest soil chemistry and terrain attributes in a Catskills watershed. *Soil Science Society of America Journal* 64: 1804-1814.

Kienzle, S.W. 2004. The effect of DEM raster resolution on first order, second order and compound terrain derivatives. *Transactions in GIS* 8: 83-111.

Lane S. N., C. J. Brookes, M. J. Kirkby, and J. Holden. 2004. A network-index-based version of TOPMODEL for use with high-resolution digital topographic data. *Hydrological Processes* 18: 191-201.

Lyon, S.W., M. T. Walter, P. Gerard-Marchant and T.S. Steenhuis. 2004. Using a topographic index to distribute variable source area runoff predicted with the SCS curve-number equation. *Hydrological Processes* 18: 2757–2771.

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

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

O'Callaghan, J.F., and D.M. Mark. 1984. The extraction of drainage networks from digital elevation data. *Computer Vision Graphics and Image Processing* 28: 323-344.

O'Loughlin, E.M. 1981. Saturation regions in catchments and their relations to soil and topographic properties. *Journal of Hydrology* 53: 229-246**.**

Quinn, P.F., K.J. Beven, P. Chevallier, and O. Planchon. 1991. The prediction of hillslope flowpaths for distributed modeling using digital terrain models. *Hydrological Processes* 5: 59-80.

Quinn, P. F., and K.J. Beven. 1993. Spatial and temporal predictions of soil moisture dynamics, runoff, variable source areas and evapotranspiration for Plynlimon, Mid-Wales. *Hydrological Processes* 7: 425-448.

Quinn, P. F., K. J. Beven, and R. Lamb. 1995. The ln(a/tan b) index: how to calculate it and how to use it in the TOPMODEL framework. *Hydrological Processes* 9: 161-182.

Schneiderman, E.M., T. S. Steenhuis, D. J. Thongs, Z. M. Easton, M. S. Zion, A. L. Neal, G. F. Mendoza, and M. T. Walter. 2007. Incorporating variable source area hydrology into a curve number based watershed model. *Hydrological Processes* 21(25): 3420-3430.

Seibert, J., J. Stendahl, R. Sorensen. 2007. Topographical influences on soil properties in boreal forests. *Geoderma* 141: 139-148.

Sørensen, R., U. Zinko, and J. Seibert. 2006. On the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrology and Earth System Sciences* 10: 101–112.

Speight, J.G. 1980. The role of topography in controlling throughflow generation: a discussion. *Earth Surface Processes* 5: 187-219.

Straumann, R.K., and R.S. Purves. 2007. Resolution sensitivity of a compound terrain derivative as computed from LiDAR-based elevation data. Fabrikant S.I. and Wachowicz M. (eds.) *Lecture Notes in Geoinformation and Cartography*. The European Information Society – Leading the Way with Geo- Information, Proceedings of AGILE 2007, 8.–11 May 2007, Aalborg, Denmark, 87-109.

Tarboton, D. G. 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resources Research* 33(2): 309–319.

Thompson, J.A., J.C. Bell, and C.A. Butler. 2001. Digital elevation model resolution: effects on terrain attribute calculation and quantitative soil-landscape modeling. *Geoderma* 100: 67–89.

Vieux, B.E. 1993. DEM aggregation and smoothing effects on surface runoff modeling. *Journal of Computing in Civil Engineering* 7: 310-338.

Walter, M.T., M.F. Walter, E.S. Brooks, T.S. Steenhuis, J. Boll, K. Weiler. 2000. Hydrologically sensitive areas: variable source area hydrology implications for water quality risk assessment. *Journal of Soil and Water Conservation* 3: 277-284.

Ward, R.C. 1984. On the response to precipitation of headwater streams in humid areas. *Journal of Hydrology* 74: 171–189.

Western, A.W., R.B. Grayson, G. Bloschl, G.R. Willgoose, T.A. McMahon. 1999a. Observed spatial organization of soil moisture and its relation to terrain indices. *Water Resources Research* 35: 797–810.

Whelan, M.J. and C. Gandolfi. 2002. Modeling of spatial controls on denitrification at the landscape scale. *Hydrological Processes* 16: 1437-1450.

Wolock, D.M. and G.J. McCabe. 1995. Comparison of single and multiple flow algorithms for computing topographic parameters in TOPMODEL. *Water Resources Research* 31: 1315-1324.

Zhang W., D.R. Montgomery. 1994. Digital elevation model grid size, landscape representation, and hydrologic simulations. *Water Resources Research* 30: 1019-1028.

Zinko, U., J. Seibert, M. Dynesius, and C. Nilsson. 2005. Plant species numbers predicted by a topography based groundwater flow index. *Ecosystems* 8: 430-441.

CHAPTER 2

SITE CHARACTERISTICS

2.1 Basin delineation

Basin delineation was impacted by DEM resolution, making a definitive delineation for the basins problematic. Figure 2.1.1 shows basin C delineation drawn using the watershed tool in the Spatial Analyst extension of ArcGIS 9.3 (ESRI Inc.).

The procedure for delineating basins is outlined in the flowchart in Figure 2.1.2. The flow direction was calculated directly from the DEM, and was one of two inputs for the watershed tool. The pour points were georeferenced flumes that were installed prior to the study. Since the locations of the flumes were constant, the only other input that influenced basin delineation was the flow direction. The key to understanding how elevation influenced basin delineation was to understand how elevation influenced flow direction.

Flow direction was calculated using an algorithm that uses the steepest drop among eight neighbors from a cell. Note: it was not calculated using multiple flow directions, as were the other derived terrain attributes in Chapters 3 and 4. The watershed tool does not accept the Taudem 4.6 multiple flow direction raster as a valid input. Depending on the cell size, the flow directions over the basin will differ (Figures 2.1.3 and 2.1.4). This is because the elevation values are extended across the area as the cell size increases, resulting in neighbors that vary with resolution.

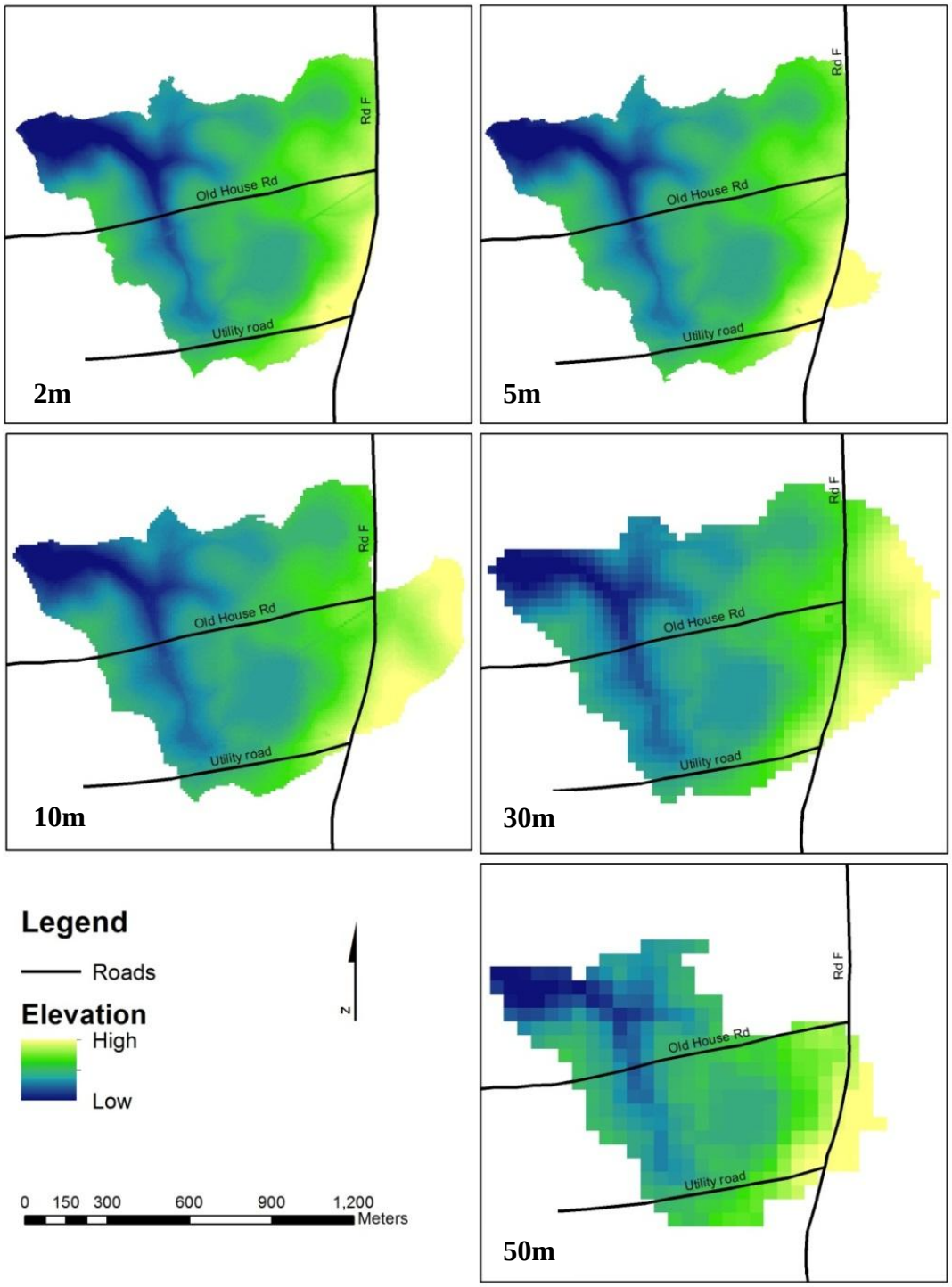

Figure 2.1.1. Basin C delineated from different DEM resolutions. The roads are included for context.

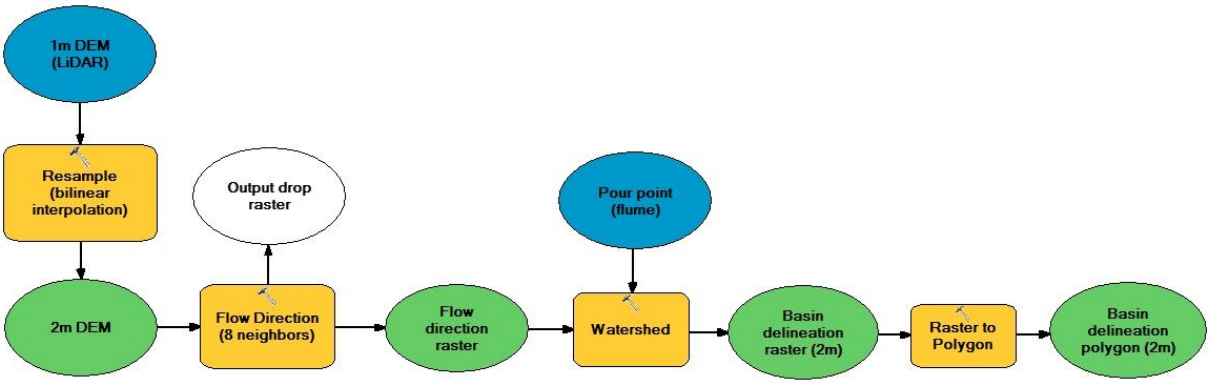

Figure 2.1.2. Flow chart showing the steps required by ArcGIS 9.3 to draw basin boundaries. These steps were repeated for each resolution.

In the 5m resolution basin C, the eastern boundary runs along Road F. In this case, the road acts as a ridge because the flow directions divide down its center (Figure 2.1.3, inset). Instead of the flow running from the east toward the pour point for basin C, it runs into a ditch and flows north. In the 30m resolution basin C, the road doesn't appear as a definite feature (Figure 2.1.4), and the flow runs over where the road should appear on the map, toward the basin C pour point.

The 50m basin C had a different problem. Like the 30m raster, the cells were too large to show any definition in the roads, so again the road did not act as a ridge. In this case the ridge was simply an artifact of how the elevation cells lined up with each other.

The watershed tool starts at the pour point and follows the flow direction until it reaches a cell where there is no other cell that "flows" into it. Figure 2.1.5 illustrates this with a pour point (flume) designated by a green dot. The basin as delineated by the program is in yellow. The arrows point in the direction of the flow out of its cell. None of the cells outside of the yellow area have an arrow that points toward the yellow area. The program interprets this as the boundary of the watershed.

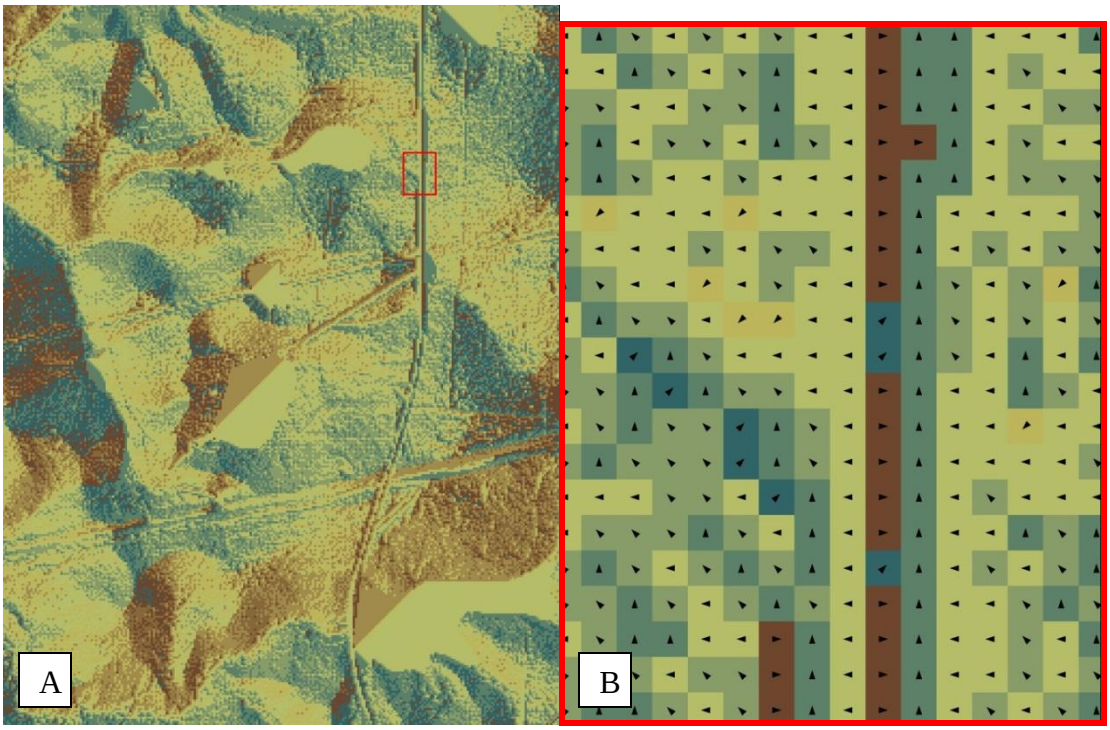

Figure 2.1.3. (A) Flow direction raster derived from 5m DEM. Note the road (Rd F) that goes through the red box. In the inset (B) on the right, the flow directions are pointed away from the road on both sides. The ditch appears as the long bluish vertical line running alongside the road.

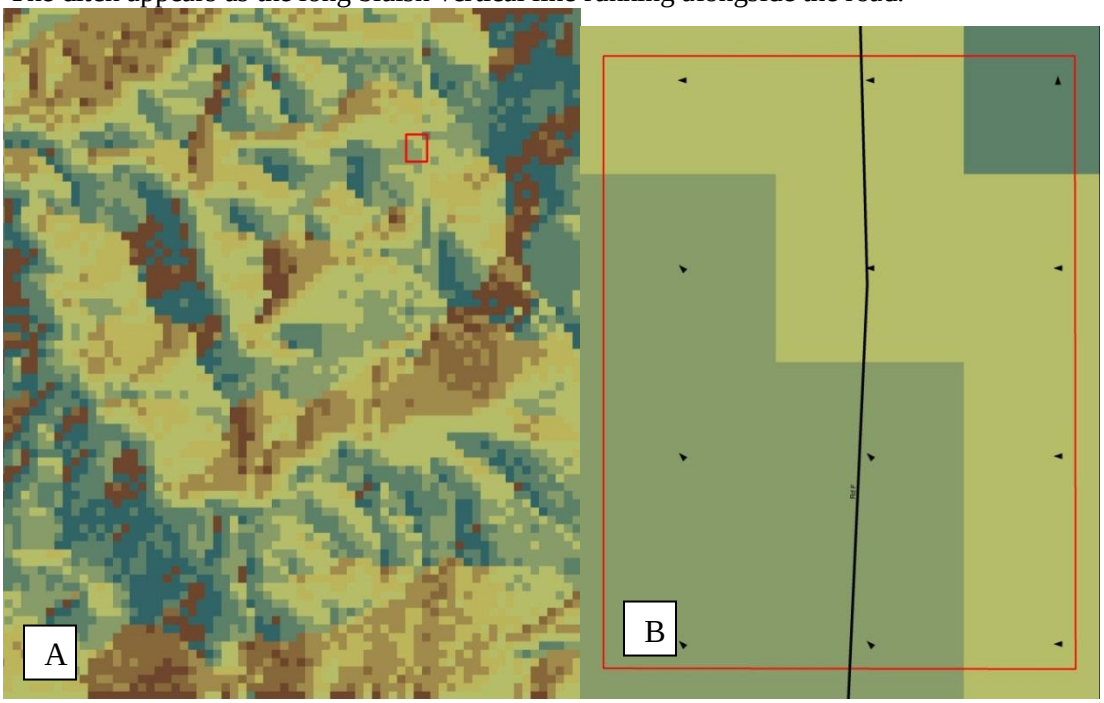

Figure 2.1.4. (A) Flow direction raster derived from 30m DEM. The red rectangle is in the exact same position in the landscape as in Figure 2.1.3. Note how the road that appeared in the 5m flow direction raster is not apparent in the 30m raster. The black line in the inset (B) on the right was added to show where the road should be. The flow direction arrows do not change when they encounter the "road".

The decision of which resolution creates the correct delineation is not an easy one. According to topography, the finest resolution should provide the most accurate flow paths. The road is a real feature that would, in reality, divert flow the way it is represented in the map. However, flow throughout the watershed includes more than simply surface flow. If flow were allowed to travel beneath the road, then perhaps the 30m delineation would be the most accurate. But without ground truth, even high definition maps produced using LiDAR do not provide adequate information, and any predictions of flow direction is merely speculation. This example illustrates the need for site-specific ground truth to supplement the information gained from topography.

As the examples of the different basin delineations based on resolution illustrate, the choice of resolution has consequences on calculation of area draining through a point in a watershed. The areas of the C basin as calculated at the five different resolutions in this study are as follows: 2m = 0.970km^2, 5m = 0.997km^2, 10m = 1.116km^2, 30m = 1.167km^2, and 50m = 0.803km^2.

2.2 Soils and vegetation

Historically the site was used as cattle and agricultural land. The topography is moderately sloping (2-5%), and the soils have a sandy surface layer underlain by a shallow loamy-clay argillic layer.

The topsoils in the study area (Basins C, B and R) were primarily sandy to sandy loam in texture. The upland areas of the basins were Dothan series sand (Plinthic Kandiudults), and the lowest areas were Pickney series sand (Cumulic Humaquepts). Additionally, the soil in the upper stream reaches of the B basin was Ogeechee series (Typic Endoaquults) with sandy loam topsoil. The upper riparian zones were mostly a complex of Vaucluse (Fragic Kanhapludults) and Ailey (Arenic Kanhapludults) series with loamy sand topsoil. Some of the samples were also taken from the topsoil of Fuquay series sand (Arenic Plinthic Kandiudults). Figure 2.2.1 shows the spatial distribution of the soil types (NRCS Soil Survey Map of SRS. Hand delineated from 1990 SCS Soil Survey onto 1:15840 mylar quad sheets then digitized by NRCS, Columbia SC on contract with the Savannah River Forest Station).

The vegetation in the study area was mixed hardwood in the lowland parts of the basins, and graduated to pines in the upper elevations of the basins (Figure 2.2.2).

The average temperature, according to a record from 1948-2005 by the South Carolina Department of Natural Resources, ranges from 34.5°F in January to 92.4°F in July. The average annual precipitation is 48.95 inches.

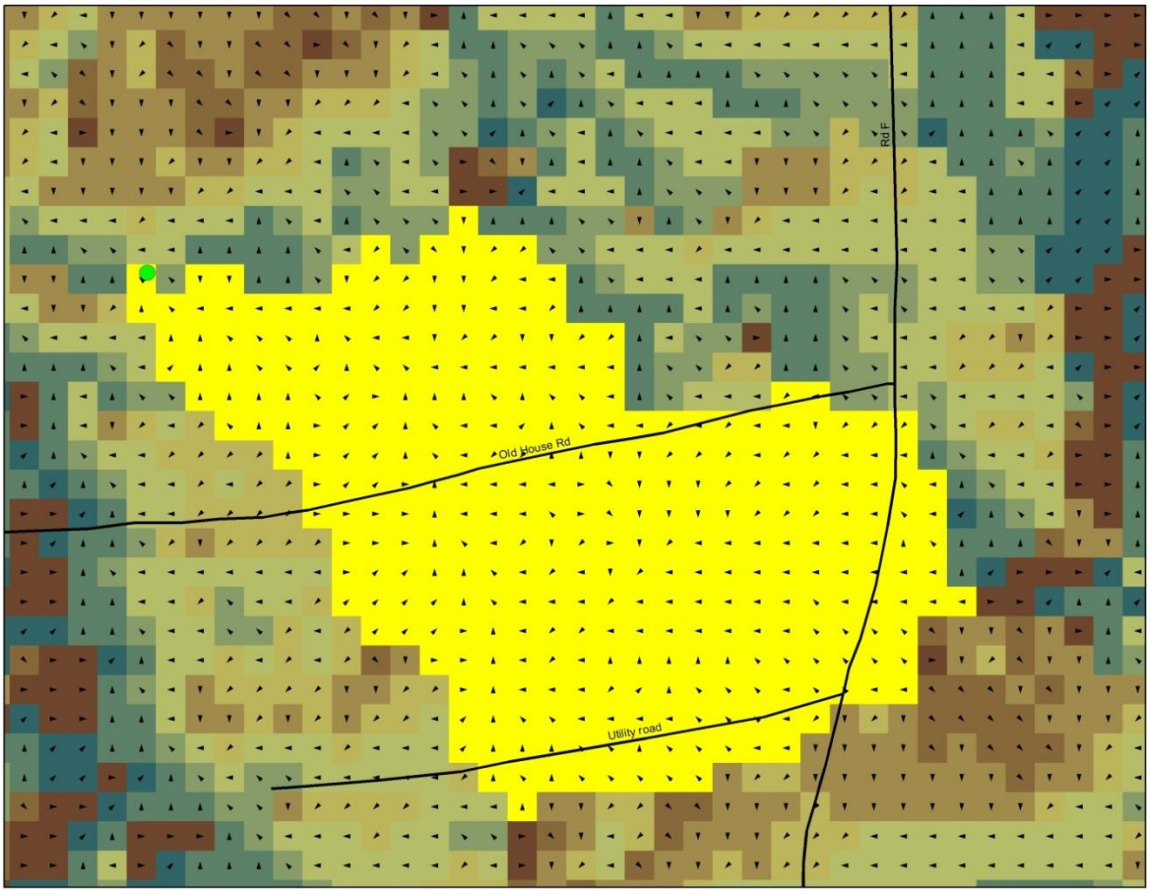

Figure 2.1.5. C basin as calculated by the watershed tool from a 50m DEM (yellow area). The green dot is the location of the flume that was used as a pour point. The arrows indicate the direction of flow from its respective cell.

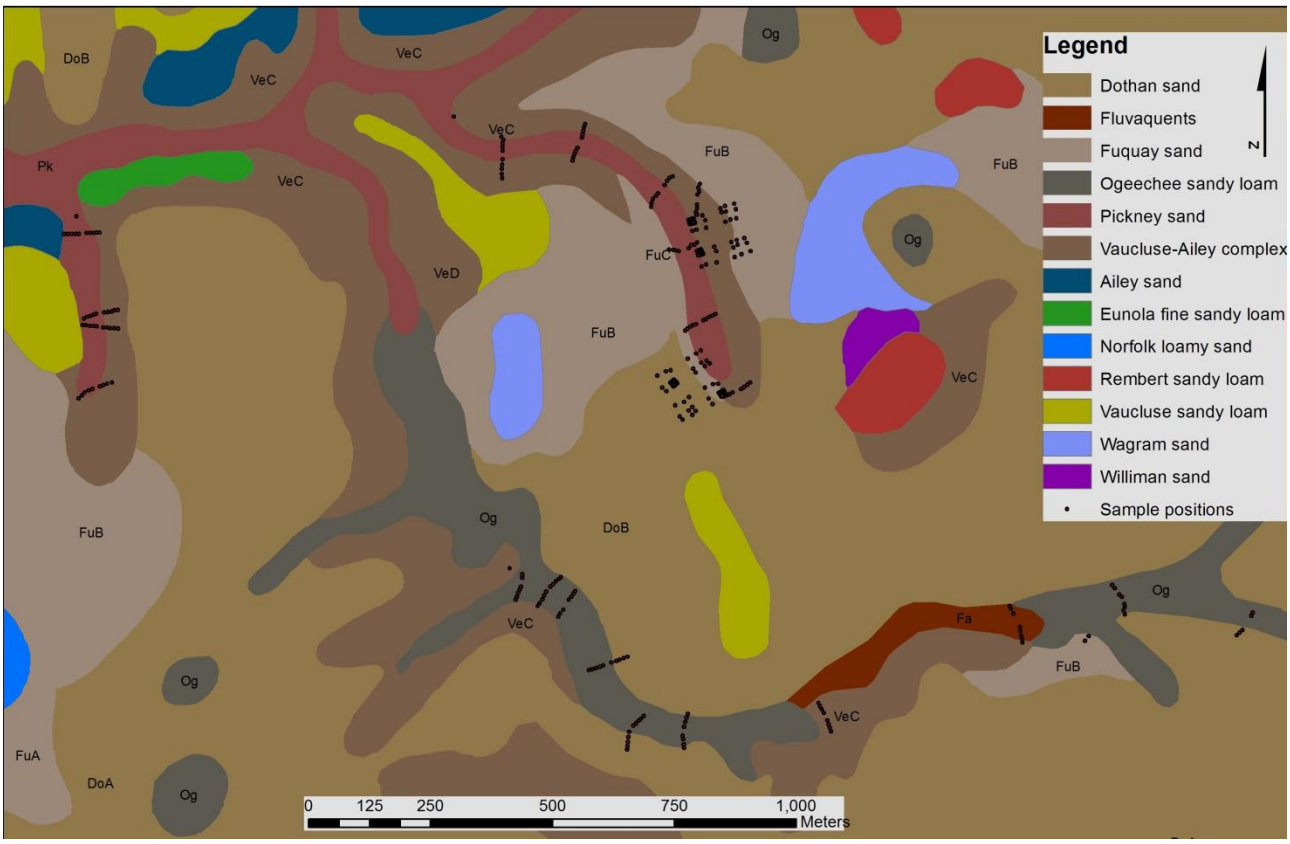

Figure 2.2.1. Spatial distribution of soil types in the study area. Source: USDA Forest Service, 2007.

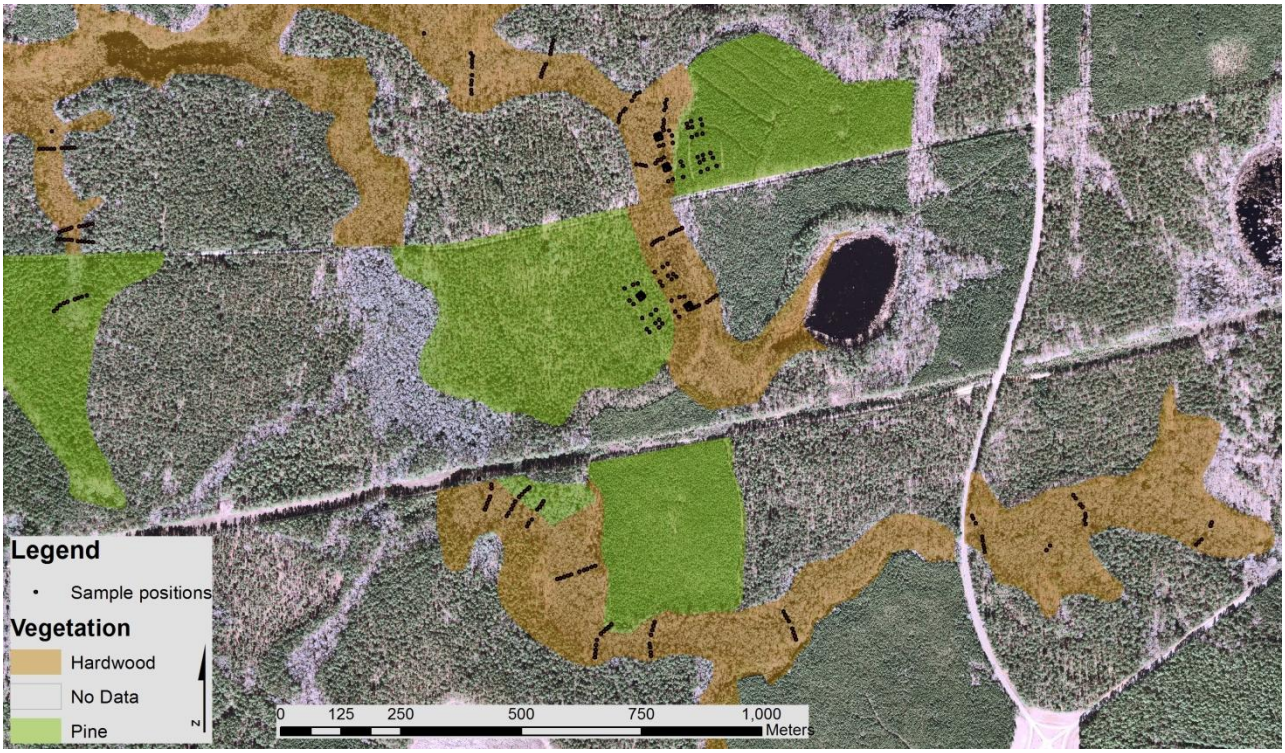

Figure 2.2.2. Spatial distribution of the vegetation surrounding the sample positions in the study area. Source: USDA Forest Service, 2007.

Chapter 3

DEM RESOLUTION EFFECTS ON NON-SPATIAL RELATIONSHIPS AND PREDICTIONS

3.1 Introduction

Characterizing the soil attributes of a large area can be expensive if it is done by collecting soil samples and interpolating the results. For this reason, there is a demand for relatively cheaper alternatives, such as remote sensing or soil-landscape modeling using terrain attributes (Park et al. 2001). Soil attributes have been correlated to terrain attributes such as elevation, slope, aspect, and curvature (Moore et al. 1993, Gessler et al. 1995, Park et al 2001). It has been shown in numerous studies that hydrology is related to the catenary development of soils throughout a landscape (Moore et al. 1993, Gessler et al. 1995, Thompson et al. 2001). Topographic Wetness Index (TWI), and Downslope Wetness Index (DWI) have been associated with process-based predictions of soil attributes such as texture, organic matter content, and horizon thickness, based on the wetness index predictions of variable source areas of soil moisture (Moore et al. 1993, Thompson et al. 2001). The TWI has also been used to enhance soil survey maps (Moore et al. 1993). It is from these studies that hypotheses can be made regarding the magnitude and direction of correlations among terrain and soil attributes.

TWI has also been called a "Compound Topographic Index" in a number of studies (Gessler et al. 1995, Fraise et al. 2001, Gobin et al. 2001) because it is constructed from two topographic attributes, slope and specific catchment area, that are derived from a DEM. One of the components of TWI is the local slope (i.e. the slope derived from the elevation difference of one cell to an adjacent cell in a DEM). Hjerdt et al. (2004) made the case that the downslope topography is an important factor when considering drainage, and that local slope does not take this into account. The DWI is calculated a bit differently. Its slope is determined by using a somewhat arbitrarily chosen elevation drop (based on overall relief of the study area) and finding the horizontal distance to travel from a cell to reach that drop. The resulting

tangent of the angle is the slope of the cell. The DWI can in this way integrate changes in the terrain over a longer distance than the local slope, and therefore better account for downslope influence. Additionally, Hjerdt et al. (2004) have found that the slope calculated by the DWI method was more robust among varying DEM resolutions than local slope.

A popular and relatively simple way to compute maps of TWI and DWI is by using digital elevation models (DEMs) in a raster grid format. In a geographical information system such as ArcGIS (ESRI Inc.), a user can input a DEM available from the US Geological Survey and create a TWI map to predict soil moisture variability across a landscape. One of the potential issues involved when using this method is that DEMs are available at different resolutions (i.e. grid cell sizes). This can result in large error in predictions made using TWI or DWI at resolutions that either do not capture enough detail in the landscape, or capture too much detail to accurately represent the processes within the landscape.

The purpose of this study was to determine which resolution DEMs would be optimal for producing wetness indices that correlated highly with soil attributes, and could thus be used predictively. It was expected that the 5m and 10m grid sizes would produce higher correlations of topographic indices with soil characteristics than TWIs and DWIs than the larger grid sizes, such as 30m and 50m, or the smallest (2m) grid size.

Distributions of the TWIs and DWIs were compared across DEM resolutions to determine the effects of DEM resolution on the calculations of both indices. The quantile and moment information were compared across resolutions, between wetness indices, and between the entire C basin and sample positions to see if the sample positions were representative of the entire basin. These comparisons were also done to see between which resolutions the effect of choosing a resolution would have the largest and smallest effect.

Because slope has been shown to be influenced by changes in resolution (Zhang and Montgomery 1994, Hjerdt 2004), the distributions of slope calculated as local slope and for the downslope index were also compared. It was expected that the ranges of both the TWIs and DWIs would decrease and that the means and minimum values would increase as resolution decreased. This was based on previous studies

(Zhang and Montgomery 1994, Hjerdt 2004) that found that the slope would decrease both in range and mean values as resolution decreased. Because the minimum values of slope are not affected much, the maximum values of TWI or DWI should not be affected much by a decrease in slope with decrease in resolution, but the minimum values would instead increase.

Recognizing that TWI and DWI are measures of relative saturation probability, it should also be considered that the spectrum of dry to saturated soils have effects that vary differently for different soil attributes. For example, decomposition speeds up as soils move from dry to moist, but then slows down or stops when soils are saturated. Thus, if wetness index is an accurate indicator of saturation, there should be a non-linear relationship between wetness index and soil attributes that is related to decomposition rate. For this reason, non-linear relationships between the wetness indices and each soil attribute were also explored.

3.2 Study site

The study was conducted in three sub-catchments (R, B, and C) of the Fourmile Branch catchment within the Savannah River Site, in Barnwell County, South Carolina, USA (Figure 3.1.1). Historically the site was used as cattle and agricultural land. The topography is moderately sloping (2-5%), and the soils have a sandy surface layer underlain by a shallow loamy-clay argillic layer. These conditions suggest that interflow occurs commonly, but vertical percolation and groundwater flow are the dominant flow paths on this watershed. The vegetation at the study site is primarily mixed hardwood and loblolly pine (*Pinus taeda*).

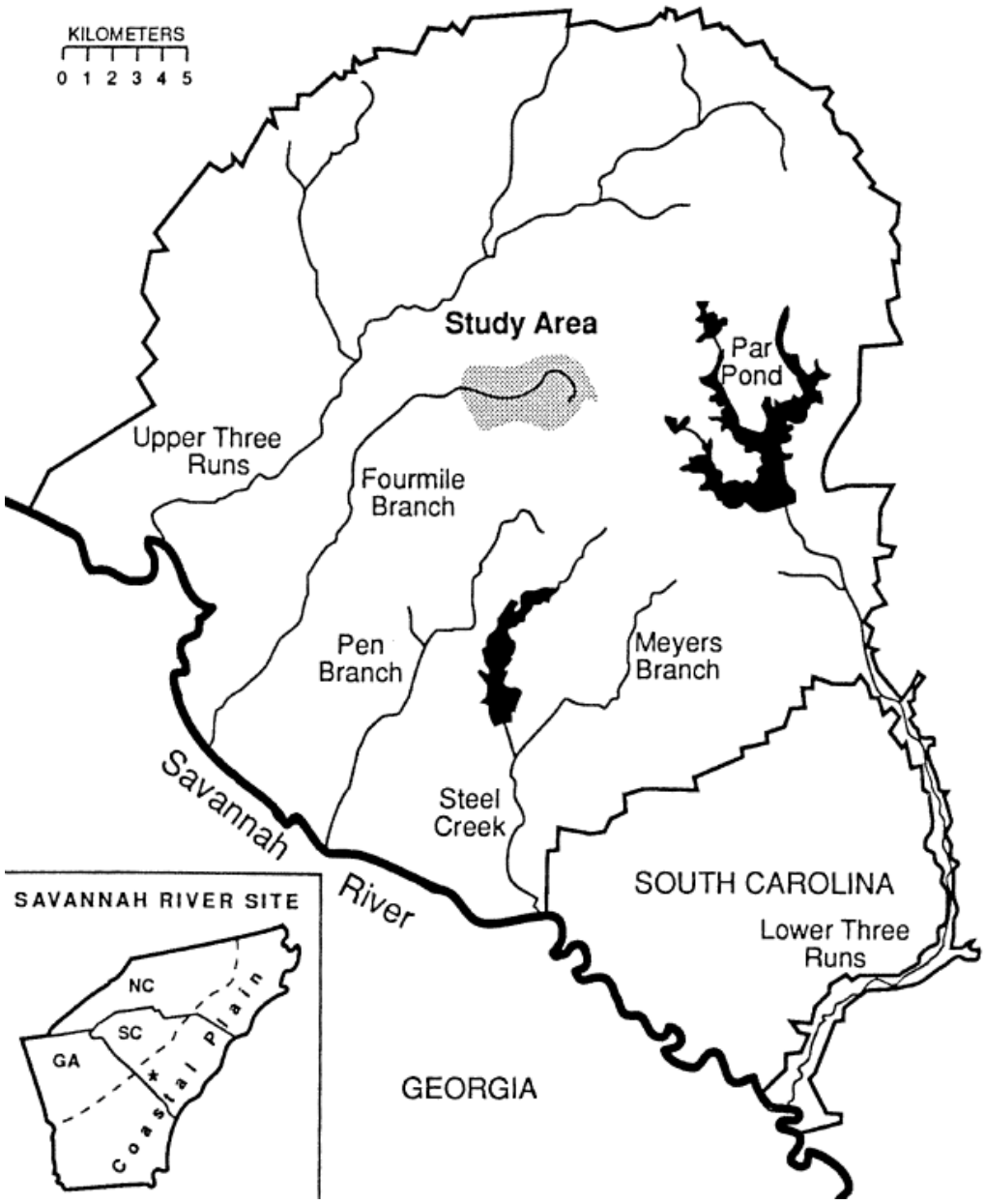

Figure 3.2.1. The Savannah River Site in South Carolina. Source: Dosskey and Bertsch, 1994. The study area in this depiction is the same as was used in this study.

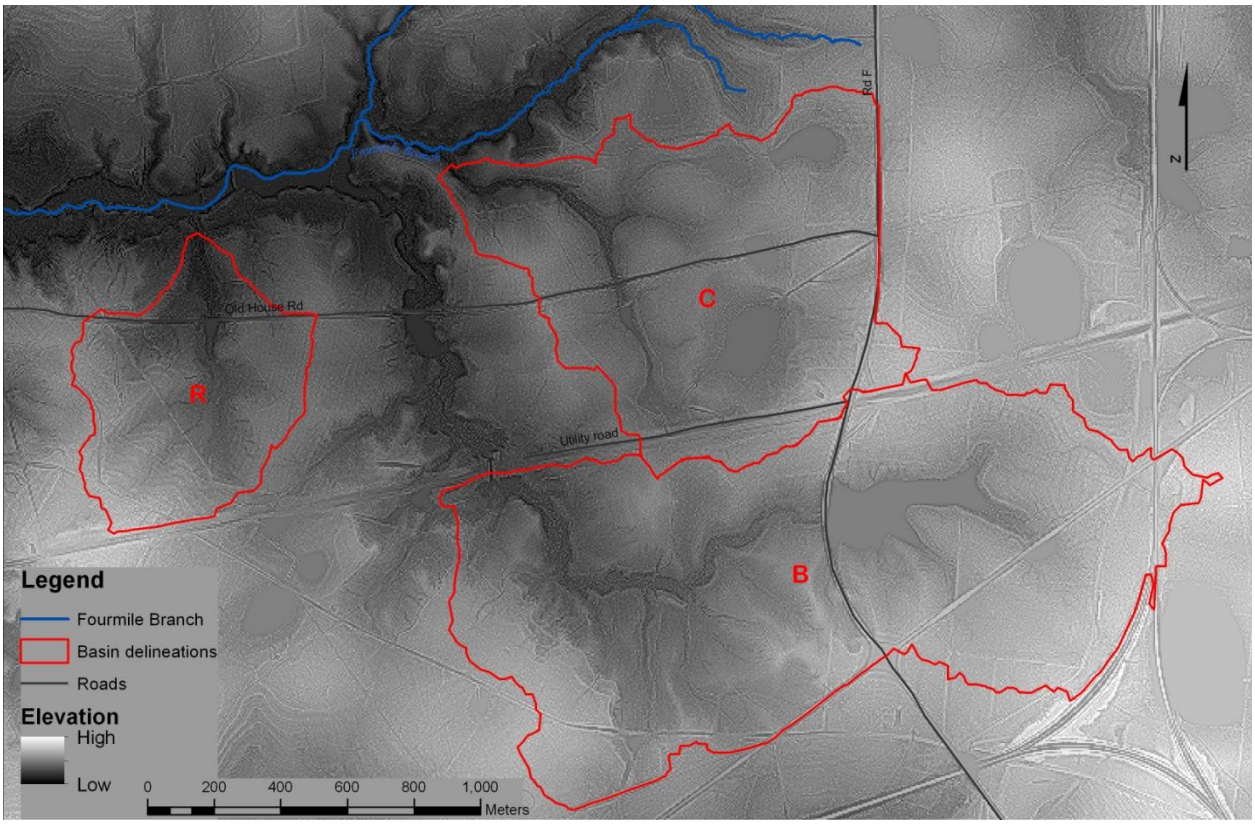

Figure 3.2.2. Sub-catchments within Fourmile Branch catchment at the Savannah River Site.

3.3 Methods

3.3.1 Calculation of TWI and DWI

A 1m (raster cell size 1m x 1m) LiDAR-derived digital elevation model (DEM) was resampled using ArcGIS 9.3 (ESRI, hereafter referred to as ArcGIS) resample tool (ArcGIS >Data Management > Raster > Raster Processing > Resample). The resulting raster grid sizes (resolutions) were 2m, 5m, 10m, 30m and 50m. LiDAR data were obtained from the USDA Forest Service in 2009. The metadata showed that the LiDAR data were collected in 2009.

Basin delineation was performed, using ArcGIS, on three watersheds (R, B and C) draining into Fourmile Branch at the Savannah River Site in South Carolina (Figure 3.1). Each watershed was delineated at each resolution, and area was calculated for each resulting map. The pour point for each basin was a flume installed previous to this study. The flume was georeferenced using an Archer (Juniper

Systems Inc.) handheld GPS unit (Hemisphere XF101 DGPS). The inputs for the basin delineation tool (ArcGIS >Spatial Analyst > Hydrology > Watershed) were flow direction and the pour point. Flow direction was calculated using the ArcGIS flow direction tool (ArcGIS >Spatial Analyst > Hydrology > Flow direction). After the watersheds were delineated, they were converted to shapefiles and were used to clip (ArcGIS >Data Management > Raster > Raster Processing > Clip) the DEMs at each resolution to the watershed boundaries.

Lateral transects were placed orthogonally to the streams (Figure 3.3.1). Four to five positions were marked along each transect on each side of the stream, spaced approximately 8m apart. Two 100x100m grids (Figure 3.3.2) were placed on hillslopes within the same catchment. The "N" grid was placed on a hillslope with a predominantly northwest aspect. The "S" grid was placed on a hillslope with a predominantly northeast aspect. The grids were arranged in a 6/10 cyclic sampling pattern (Burrows 2002) with two smaller grids arranged in a (scaled down but otherwise identical) cyclic sampling pattern nested within each larger grid. If the pattern was regular, there would be 10 points at 10m intervals to have every possible distance in 10m increments between 10 – 100m (10/10 pattern). In a 6/10 cyclic pattern, there are 6 points out of a possible 10 to account for the same increments of intervals as in a 10/10 with fewer sample points. The points were laid out as in Figure 3.3.3. The reason for this pattern was to maximize the number of pairs in each lag for the purpose of constructing semivariograms, while minimizing the number of samples needed to be collected. Each position was georeferenced using an Archer (Juniper Systems Inc.) handheld GPS unit (Hemisphere XF101 DGPS). For better accuracy, the positions in the grids were also measured using a tape measure and compass.

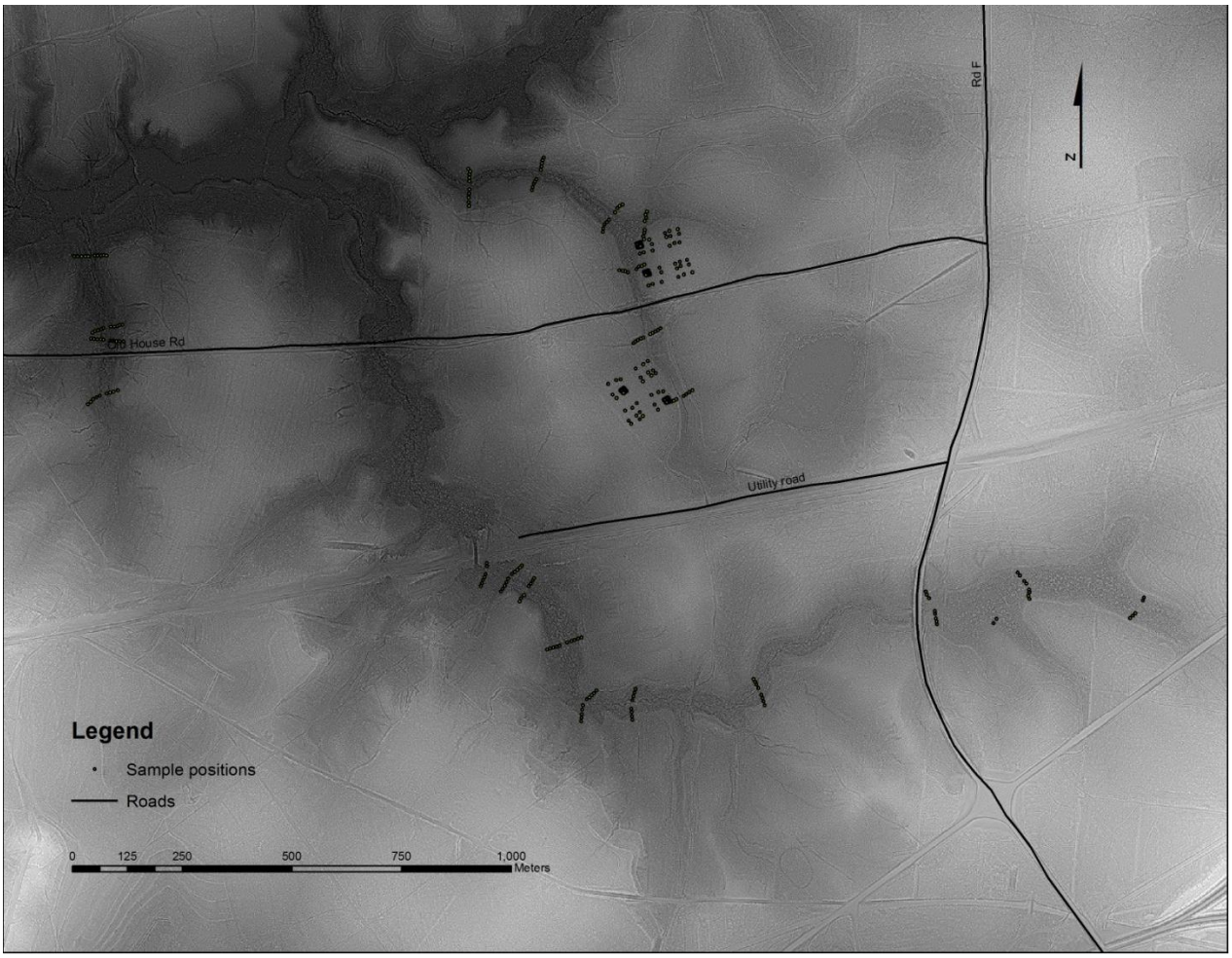

Figure 3.3.1. Map of sample positions on all three basins.

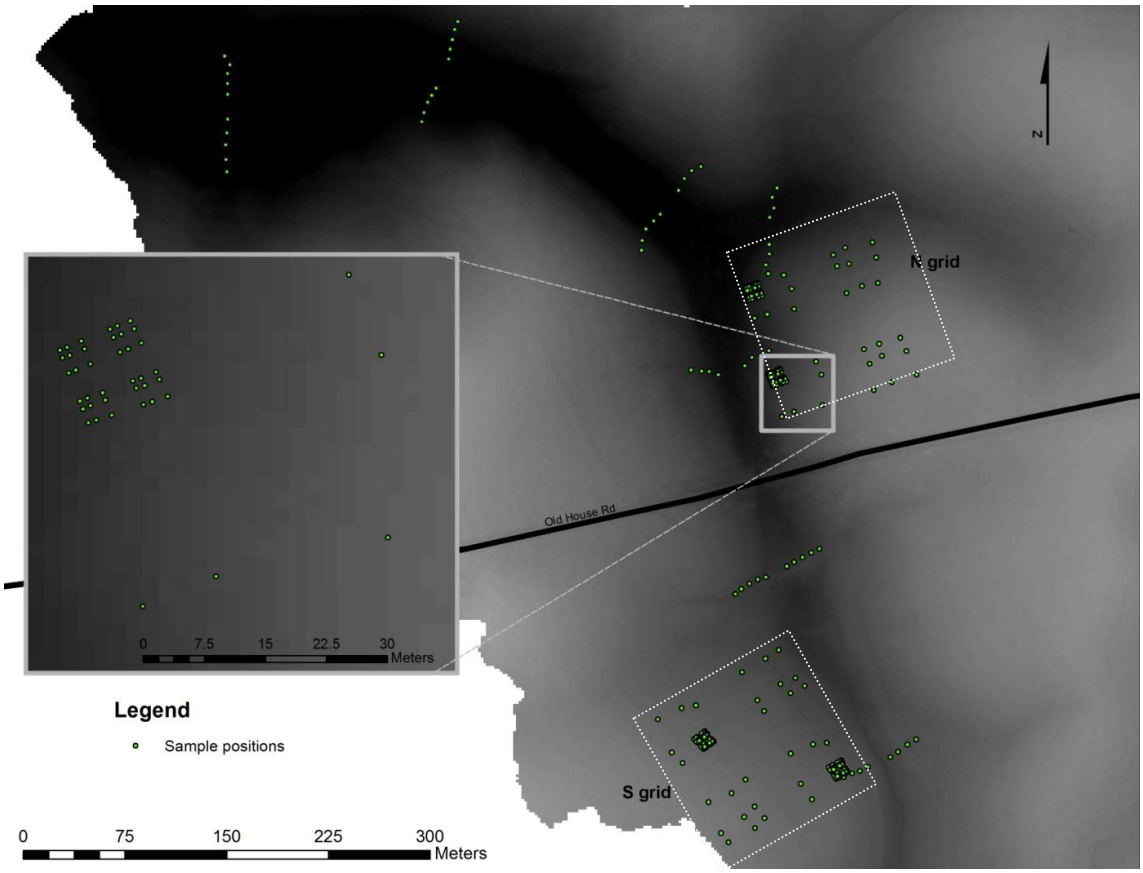

Figure 3.3.2. Map of sample positions on the C basin. The N and S grids (within dotted lines) each contain two smaller grids that are similar in layout, but 1/100th of the size (inset).

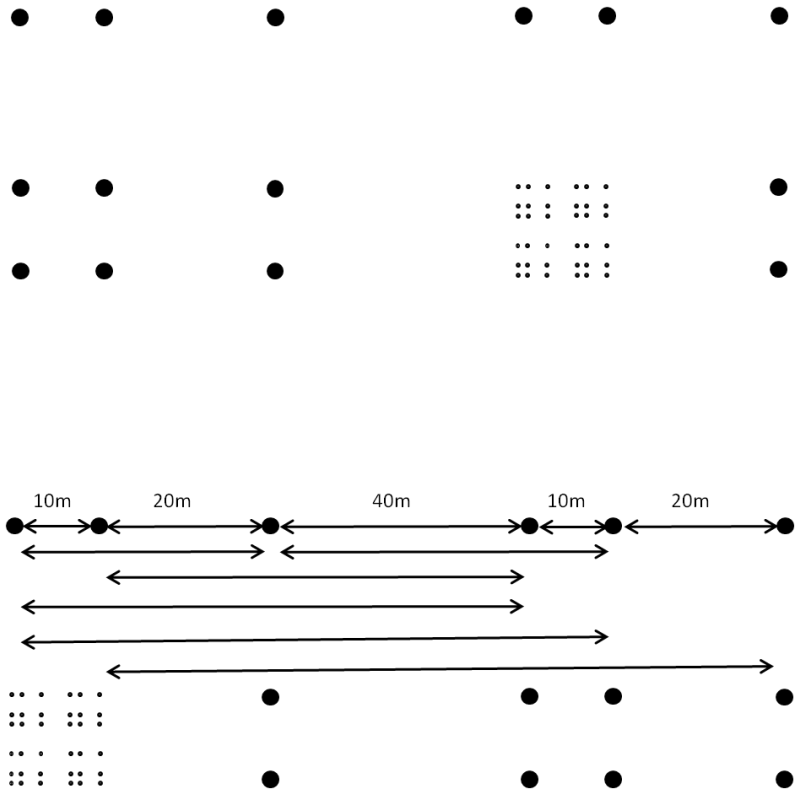

Figure 3.3.3. Sampling design for grids. The intervals between points were arranged so that every interval (as indicated by the arrows) in increments of 10m from 0-100m for the larger grids, and increments of 1m from 0-10m for the smaller grids were taken into account. The same pattern for each row is repeated for the columns.

The components of the Topographic Wetness Index (TWI) were calculated using TauDEM v. 4.6 (http://hydrology.usu.edu/taudem/taudem4.0/taudem40.html). Specific Contributing Area (SCA) and slope (tan β) were calculated using the "D-inf" flow direction/slope tool in TauDEM v. 4.6. TWI was then calculated in raster calculator (ArcGIS >Map Algebra > Raster Calculator), using the following equation: Ln(SCA/tanβ). TWI was calculated in this way at each DEM resolution, and designated by its respective resolution (i.e. 2m TWI, 5m TWI, 10m TWI, 30m TWI and 50m TWI). The TWI values were extracted to the sample positions (ArcGIS > Spatial Analyst > Extraction > Extract Values to Points) using bilinear interpolation.

Downslope Wetness Index (DWI) was calculated using a program written by Dr. John Dowd in the Geology Department at the University of Georgia. For a complete description of how this program

works, see Appendix A. The program constructed a matrix from the coordinate and elevation values of each DEM. At each elevation value it scanned a ring of adjacent cell values for an elevation difference of d=-0.5m. If that difference criterion was met or exceeded, the length to that cell was calculated. Otherwise, it scanned the next ring. The difference in elevation value between d and the found cell's elevation value was used as a proportion from which to calculate the adjusted length from the center cell to the found cell. The slope ($\tan\alpha_{0.5}$) for the center cell was calculated by dividing d by the adjusted length. DWI was calculated in Raster Calculator using the following equation: $\ln(SCA/\tan\alpha_{0.5})$ DWI was calculated in this way for each DEM resolution, and designated by its respective resolution (i.e. 2m DWI, 5m DWI, 10m DWI, 30m DWI and 50m DWI). The DWI values were extracted to the sample positions (ArcGIS > Spatial Analyst > Extraction > Extract Values to Points) using bilinear interpolation.

3.3.2 Soil data collection

Soil attributes were used to assess the predictive capability of each index-resolution combination, and to calibrate the best model. Terrain and soil attribute data were collected from three separate small watersheds. One watershed was sampled heavier than the other two, and was used to create the models. The other two served as validation sets. Soil attributes considered "cheaper" such as bulk density and soil texture were used as regressors in the models to improve the predictions. Terrain attributes such as the elevation and distance to streams were also used to improve the models. The wetness indices were also replaced by their components (slope and specific catchment area) to see how well they explained the variance of the soil attributes.

Soil samples and profile information were collected from each watershed. At each position, a hole was augered to the depth of the argillic layer. The soil profile was examined for horizonation and redoximorphic features. The depth to each horizon change was recorded, along with the presence and type of redoximorphic feature.

In July 2010, the organic horizon (including forest floor litter) was collected using a 26cm diameter ring from each position. Everything within the ring was collected and placed in a paper bag.

Roots were excluded from all samples. Two mineral soil samples were collected at a depth of 7.5cm, using a 20lb. slide hammer and sleeves with an inner diameter of 7.5cm and 7.5cm length. The sample intended for bulk density measurement was collected first.

The organic horizon sample was oven dried at 65°C, weighed, ground in a Wiley mill and then ball-milled for carbon and nitrogen (CN) analysis. The bulk density sample was oven dried at 105°C, weighed, and stored for texture analysis. The remaining sample was air dried for several days, sieved in a 2mm sieve to separate coarse and fine fragments, and ball-milled for CN analysis. No oven-dried correction was applied.

Well-mixed, pulverized subsamples (1mg) were analyzed for carbon and nitrogen concentrations using a CHN elemental analyzer (CE Instruments – NC 2100, CE Elantech Inc., Lakewood, NJ).

Soil particle size distribution was determined using the hydrometer method as described in Gee and Bauder (1979). Soil samples weighing 50g were soaked for 20 hours in a solution of sodium hexametaphosphate for chemical dispersion, and then mechanically dispersed using a soil mixer as in Bouyoucos (1962). The fractions of interest were sand (0.53 – 2.00mm), silt (0.002 – 0.53mm) and clay (<0.002mm). The sand fraction was estimated by reading the density measurement at 30s, doubling the reading and subtracting from 100, as in Bouyoucos (1951). The clay fraction was calculated using the Day method as described in Gee and Bauder (1979). The silt fraction was calculated by difference. A blank solution of deionized water was measured at the same time as other measurements, and subtracted from the other measurements to account for changes in density due to water temperature.

3.3.3 Electrical conductivity

An electromagnetic induction (EMI) probe transmits a primary magnetic field that induces electrical current in the earth. The electrical current generates a secondary magnetic field that is measured by the sensor of the probe. The information from the secondary magnetic field is converted to a measure of electrical conductivity (mS/cm). In December 2010, and then again in January and March 2011, a Dualem-2 EMI probe (Dualem, Milton, ON, Canada) was used in combination with an Archer

(Juniper Systems Inc.) handheld GPS unit (Hemisphere XF101 DGPS) to measure the electrical conductivity (EC) of the stream and adjacent hillslopes in Basin C. The probe was carried approximately 50cm above ground, longitudinally along the stream. The unit recorded a data point consisting of X and Y coordinates, and horizontal and vertical conductivity every meter of travel. The probe was carried in passes parallel to the stream, spaced approximately 10m apart, covering an area of approximately 195,000 square meters. The data were kriged at thirty-five 10m lags using ArcGIS to produce raster surfaces of electrical conductivity, which were then extracted to the sample positions (ArcGIS > Spatial Analyst > Extraction > Extract Values to Points) using bilinear interpolation.

3.3.4 Data analysis

Hillslope and catchment scales were compared by using the positions from only one grid to represent hillslope, and both grids and C basin transects to represent catchment in the C basin.

3.3.4.1 Data preparation

Carbon and nitrogen values were measured as concentrations, but in order to relate to common literature reports, it was desirable that the results reflect their values as carbon and nitrogen content. Therefore, bulk density (for the 0-7.5cm depth A horizon samples), and weight of organic matter (for the organic horizon samples) were needed to convert the values. The conversion formulas for each are below:

Content (kg/ha) = bulk density (g/cm^3) * N or C % * (10mg/g) * ($kg/10^6 mg$) * ($10^8 cm^2/ha$) * 7.5cm

Content (kg/ha) = π $(13cm)^2$ * OM wt. (g) * N or C % * (10mg/g) * ($kg/10^6 mg$) * ($10^8 cm^2/ha$)

All data were transformed where appropriate prior to analyses. All indices were skewed to the right, so they were all log-transformed for normality. Nitrogen and carbon in both organic and mineral layers, and silt were also log-transformed. Bulk density was transformed using Box-Cox with a lambda of 2.5. The means and standard deviations were calculated before being back transformed for reporting.

Only data used in regressions and model predictions were transformed. Distribution data were left raw because no assumptions of normality were necessary.

3.3.4.2 Additional terrain attributes

Slope as calculated using Taudem 4.6 (above), and as calculated using the Dowd program were extracted to sample positions in ArcGIS using bilinear interpolation.

Distance to nearest stream was calculated for each sample position in ArcGIS. The streams were generated using the ArcHydro Tools extension (ArcHydro>Terrain preprocessing>Stream definition) and the number of flow accumulation cells needed to define a stream. The number of cells chosen depends on the grid size of the flow accumulation raster. In this case, the stream was chosen arbitrarily as ephemeral and intermittent streams where water was observed on site during the winter wet period. The number of cells (10,000) was chosen that would generate a stream to closely approximate those areas.

3.3.4.3 TWI and DWI comparisons

Distributions of the TWI and DWI values for both the C basin and the sample positions were computed using JMP 9.0.0 (SAS Institute Inc.). Quantiles and moments were used to quantify key differences among the distributions. There were over 200,000 cells in the C basin at the 2m resolution, while there were only approximately 300 cells in the basin at the 50m resolution. There were 236 cells used to calculate the distributions for the sample positions at all resolutions. Minimum, maximum, interquartile range, and median differences between each resolution were calculated to quantify the effect of DEM resolution on the calculation of TWI and DWI. The effect of wetness index type was also quantified in this way. The indices were then each binned into discrete integer classes. For example, index class 1 was all values x_i where $1 \leq x_i < 2$. Cumulative distribution plots were constructed, and the areas for each index class were calculated as a percentage of the entire area of the basin and compared across resolutions.

3.3.4.4 Regression analyses

Unless otherwise noted, all statistics were run in JMP 9.0.0 (SAS Institute Inc.). Linear regression was performed on combinations of the wetness indices and soil (BD, C and N, EC, and textural classes) attributes. After examination of the regression plots a non-linear relationship was found. A quadratic fit was then applied to the data, and examined in place of linear regression. The reasoning behind exploration of the non-linear fit was that conditions for decomposition and several chemical mechanisms perform best under moist but not saturated conditions. These chemical transformations are complex, not enough data or parameters were collected, and cause-effect relationships were beyond the scope of this thesis. Nonetheless, non-linear relationships could possibly be used to understand the resolution effects across wetness index classes.

Multiple non-linear regressions were performed on combinations of the wetness indices and soil attributes. Any variables that were likely to display multicollinearity were not put in the same model together. For example, slope and the topographic indices showed a strong relationship (R^2>0.5), so they were never in the same model. Elevation, distance from streams and weights of the organic horizon samples were added as regressors in multiple regressions to improve predictions of bulk density, as well as C and N in both the organic and mineral (0-7.5cm depth) horizons.

Variables were put into multiple regression models based on their relationships as mentioned above. If they were not significant at the alpha level =0.10, they were removed. When models were optimized for the calibration data in watershed C, they were applied to the data available from the B and R watersheds. The best model was chosen by using the RMSE between the predicted and observed values of the variable. In order to keep distinction among the resolutions, each model only used the data from its respective resolution. For example, for the 5m resolution, only 5m elevation and 5m TWI could be used together, not 10m elevation and 5m TWI. Some of the variables (e.g. EC and organic layer attributes) had significant model effects, but couldn't be used in the models to be validated by the B and R transects because they were not measured on those watersheds. They were instead used in models with the data points north of Old House Road and validated using the data points south of Old House Road.

3.4 Results and discussion

3.4.1 Terrain attribute comparisons

The effect of resolution on each wetness index was quantified by absolute difference of quantile and moment distribution characteristics between succeeding changes in resolution (Tables 3.4.1 and 3.4.2). The largest resolution effect on the minimum index value was between the 2 and 5m TWIs and the 5 and 10m DWIs. There was an almost six index value difference in minimums between the 2 and 50m TWIs and DWIs. The smallest resolution effect on the minimum index value was between the 30 and 50m resolutions for both the TWIs and DWIs. There was a clear trend of increasing minimum index values for both indices with decreasing resolution. This was expected as mentioned in the introduction, based on the reasoning that as slope decreases, the minimum index values increase. The maximums showed no obvious trend across resolutions, and most of the resolution effects were considerably smaller than those for the minimums. The largest resolution effect on maximums for both indices was between the 30 and 50m resolutions. The smallest effect was between 5 and 10m TWIs, and between 2 and 5m DWIs. There was a span of three index values across all resolutions for the maximum TWIs, but less than one index value difference across all resolutions for the maximum DWIs. As a result of increasing minimums and small differences among maximums, the ranges followed a trend where they decreased with decreasing resolution.

The interquartile range (IQR) did not show nearly as much variation across resolutions as did the minimum-maximum range, for either wetness index. The largest resolution effect for IQR was between 5 and 10m resolutions for both indices, and this effect only varied by 0.541 for TWI and 0.409 for DWI. By comparison, the smallest resolution effect for the minimum-maximum range was 0.946 between the 10 and 30m DWI, more than twice as high as the IQR. The largest effect for the minimum-maximum range was 2.754 between the 10 and 30m TWI. So in this case, choosing to use DWI would mean that the effect of the choice of resolution between 10 and 30m would affect the total range of values a third as much as choosing to use TWI.

In fact, using DWI instead of TWI was more robust to differences between 10 and 30m resolutions for all distribution characteristics except variance. The opposite was true for the instance where the choice was between the 5 and 10m resolution. In that case, the TWI was more robust to resolution effects than the DWI in all distribution characteristics except IQR. Between the 2 and 5m resolutions, DWI proved to be more robust to resolution effects across all distribution characteristics.

The median values showed a clear trend of increasing index values with decreasing resolution. The largest effect for the median was between the 2 and 5m resolutions of both indices, and the smallest effect was between the 30 and 50m resolutions. What this means is that the choice of using either a 30 or 50m DEM does not impact the median of TWI or DWI as much as the choice of using either a 2 or 5m DEM. In terms of range, the impact of the choices between 2m and 5m is less than the impact of the choice between 30m and 50m resolutions. The choice of using either TWI or DWI impacts their distribution characteristics more when using a 10m DEM than with other resolutions.

Table 3.4.1. Distribution characteristics across resolutions for each wetness index in the C basin. Sample sizes for each resolution were: 2m = 211,133; 5m = 36,146; 10m = 10,259; 30m = 1,214; and 50m = 313.

	2m TWI	5m TWI	10m TWI	30m TWI	50m TWI	2m DWI	5m DWI	10m DWI	30m DWI	50m DWI
Min	1.194	2.693	4.047	5.341	5.974	1.155	2.382	4.121	5.341	5.974
Max	18.659	18.758	18.838	17.378	15.435	17.855	17.861	17.613	17.887	16.937
Range	17.465	16.07	14.791	12.037	9.461	16.7	15.479	13.49	12.55	10.96
Median	4.583	6.125	6.875	7.629	7.920	4.988	6.292	7.294	7.810	8.126
IQR	2.779	2.480	1.940	1.531	1.673	2.713	2.469	2.060	1.767	2.022

Table 3.4.2. Resolution and wetness index effects on distribution characteristics in the C basin. Lower numbers mean that the characteristic is more robust to choice of either resolution or wetness index.

TWI	Min	Max	Range	Median	IQR
2 vs. 5	1.499	0.099	1.400	1.542	0.299
5 vs.10	1.354	0.080	1.274	0.750	0.541
10 vs. 30	1.294	1.460	2.754	0.754	0.409
30 vs. 50	0.633	1.943	2.576	0.291	0.143
DWI					
2 vs. 5	1.227	0.006	1.221	1.304	0.244
5 vs.10	1.739	0.248	1.987	1.002	0.409
10 vs. 30	1.220	0.274	0.946	0.516	0.292
30 vs. 50	0.633	0.950	1.583	0.316	0.255
TWI vs. DWI					
2m	0.039	0.804	0.765	0.405	0.066
5m	0.311	0.897	0.586	0.167	0.011
10m	0.074	1.225	1.299	0.419	0.120
30m	0.000	0.509	0.509	0.181	0.237
50m	0.000	1.502	1.502	0.206	0.349

The cumulative distribution functions (CDF) for both wetness indices for the sample points are shown in Figure 3.4.1. The sample points did not represent the range of TWI values relative to the entire basin very well. The loss in the maximum end of the range from basin to positions was more severe (a difference of 2-8 index values) than the loss in the minimum end (1-2). Furthermore, the maximums for the basins did not vary across resolutions as much as with the sample points. With the exception of the 2m and 50m TWIs the sample positions represented the IQR of the basin well. The distributions of the 5 and 10m TWIs were similar for the positions. With that exception, there was an obvious decrease in range with a decrease in TWI for both the sample points and the entire basin. The same patterns held in general between the DWIs for basin and sample positions (Figure 3.4.2).

Figures 3.4.3 – 3.4.12 show the spatial distribution of TWI and DWI. As would be expected, the figures show that the finer resolutions have much greater detail in separation of potential variable source areas. The wettest areas (blue) are shown to include small tributaries in the finer resolutions, but appear to shift the bulk of the wetness to widen the ephemeral streams in the coarser resolutions.

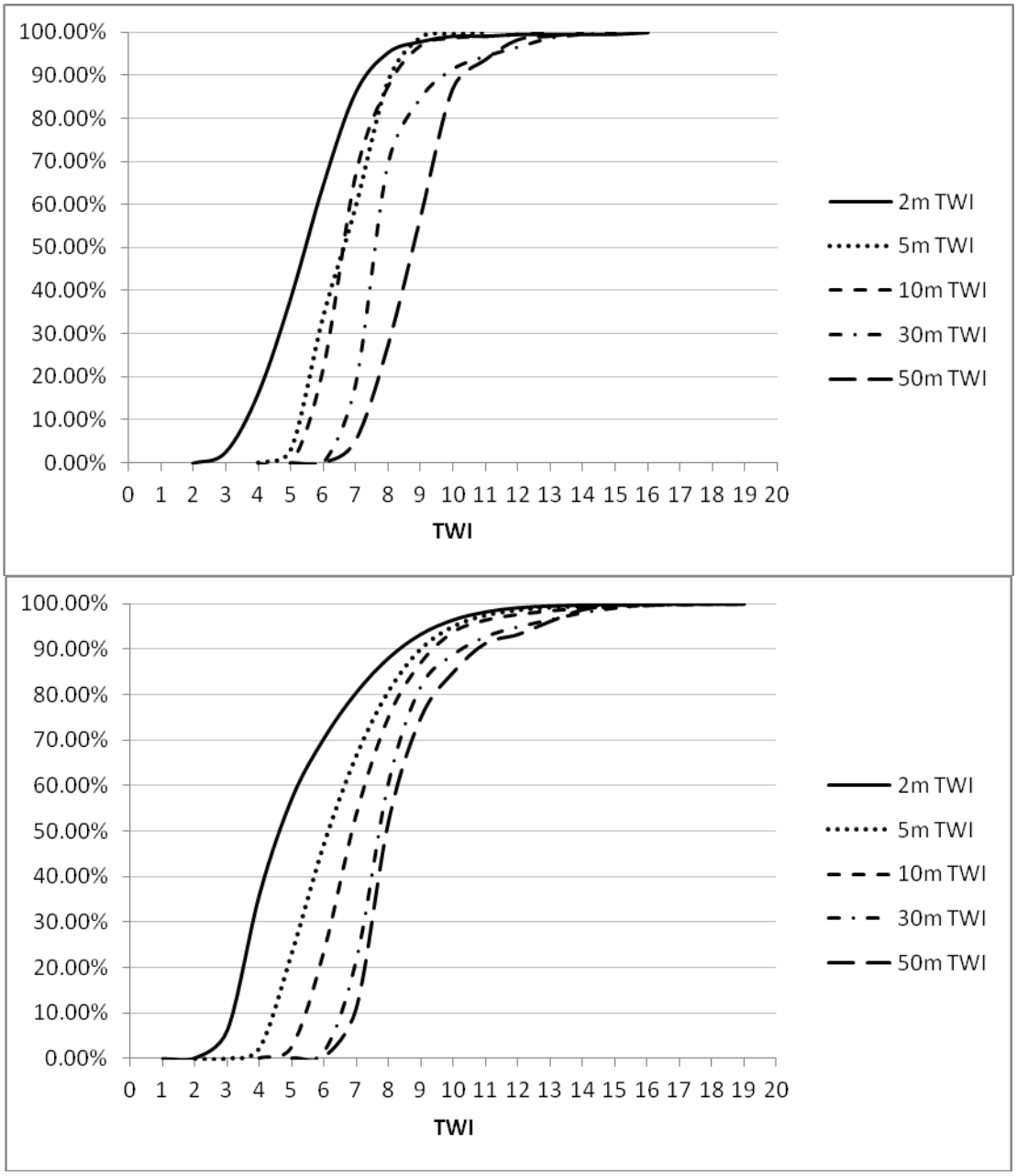

Figure 3.4.1. Cumulative distribution function plots for binned TWI at sample points (top) and the entire C basin (bottom).

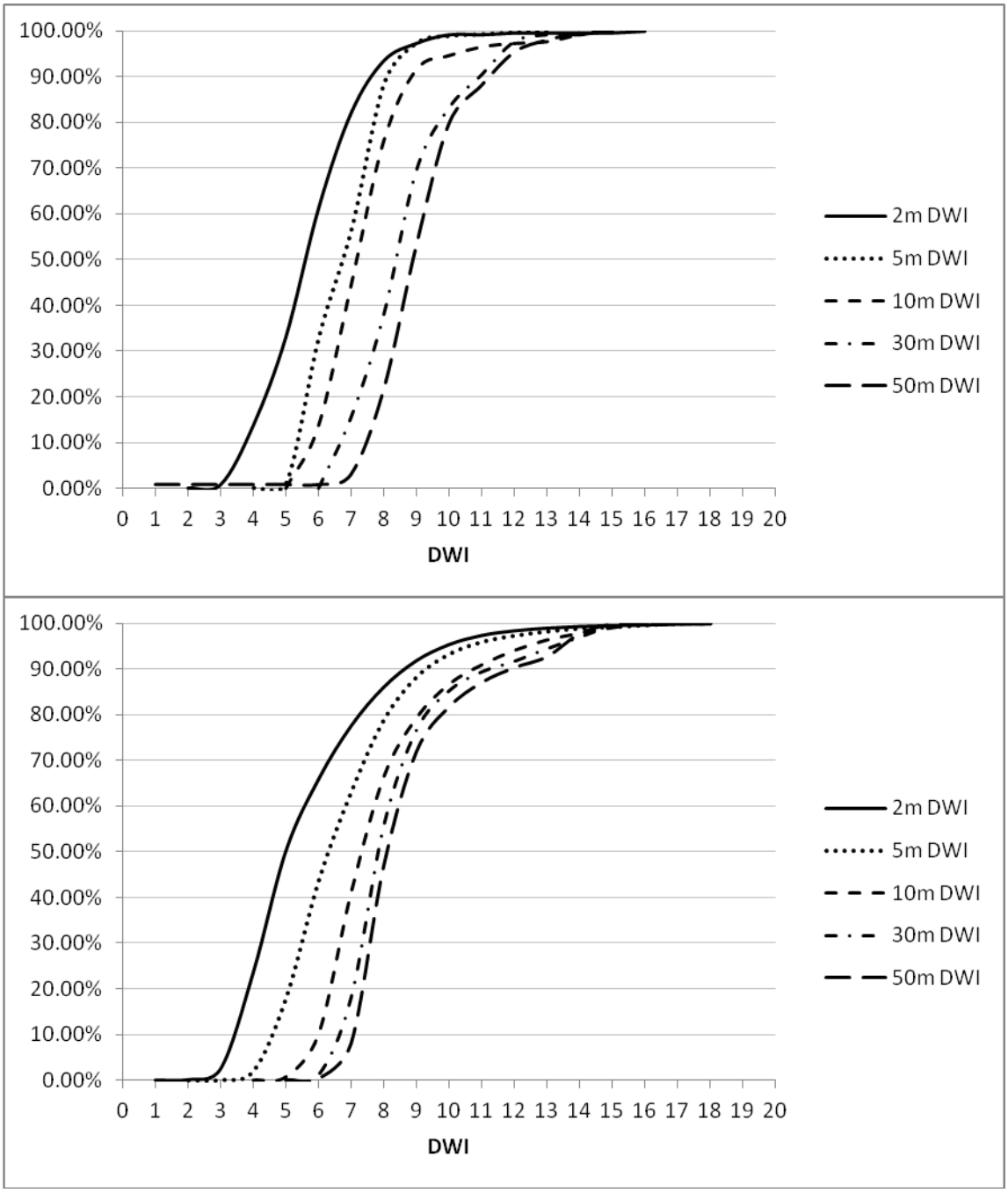

Figure 3.4.2. Cumulative distribution function plots for binned DWI at sample points (top) and the entire C basin (bottom).

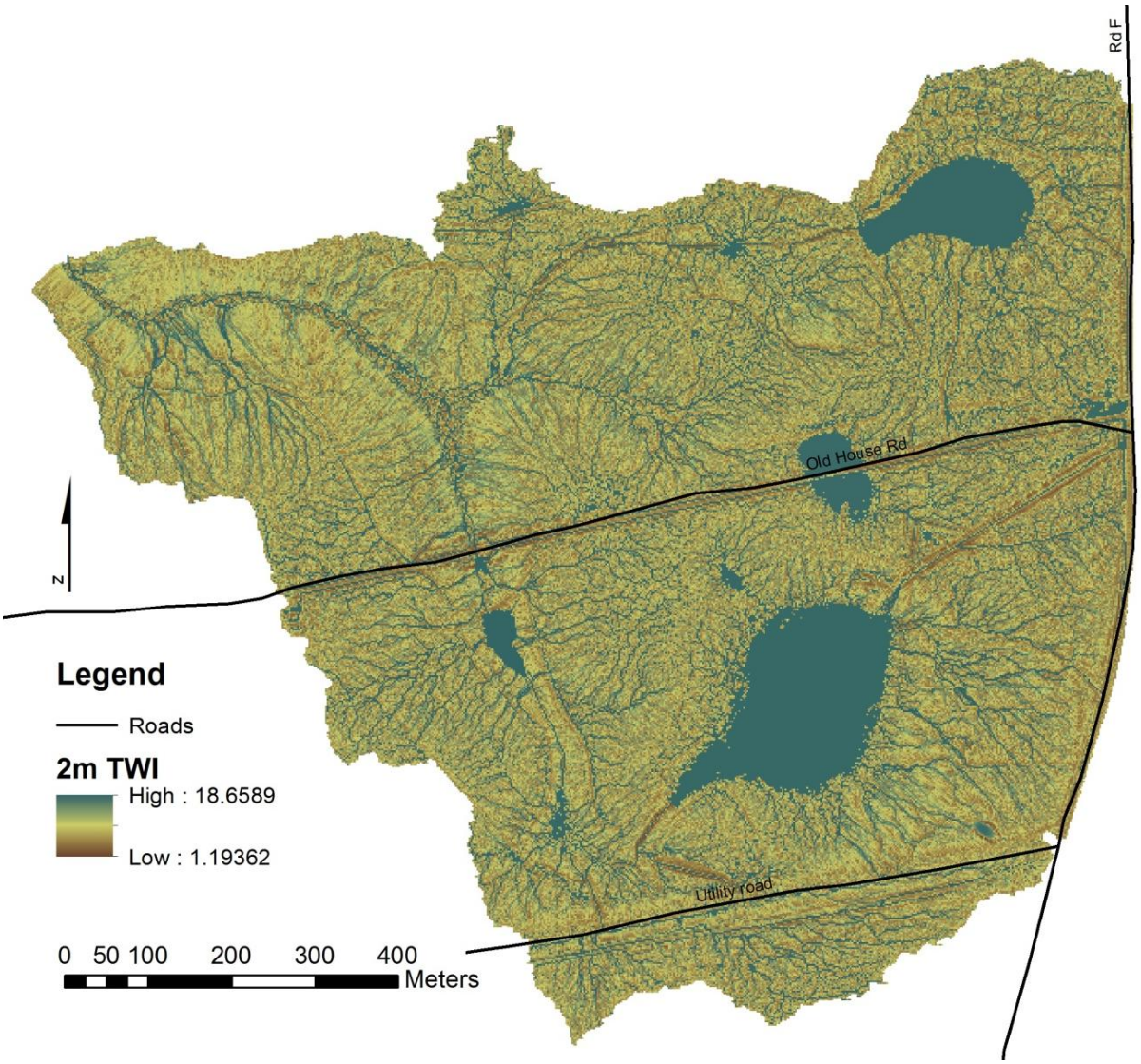

Figure 3.4.3. Map of 2m TWI in basin C.

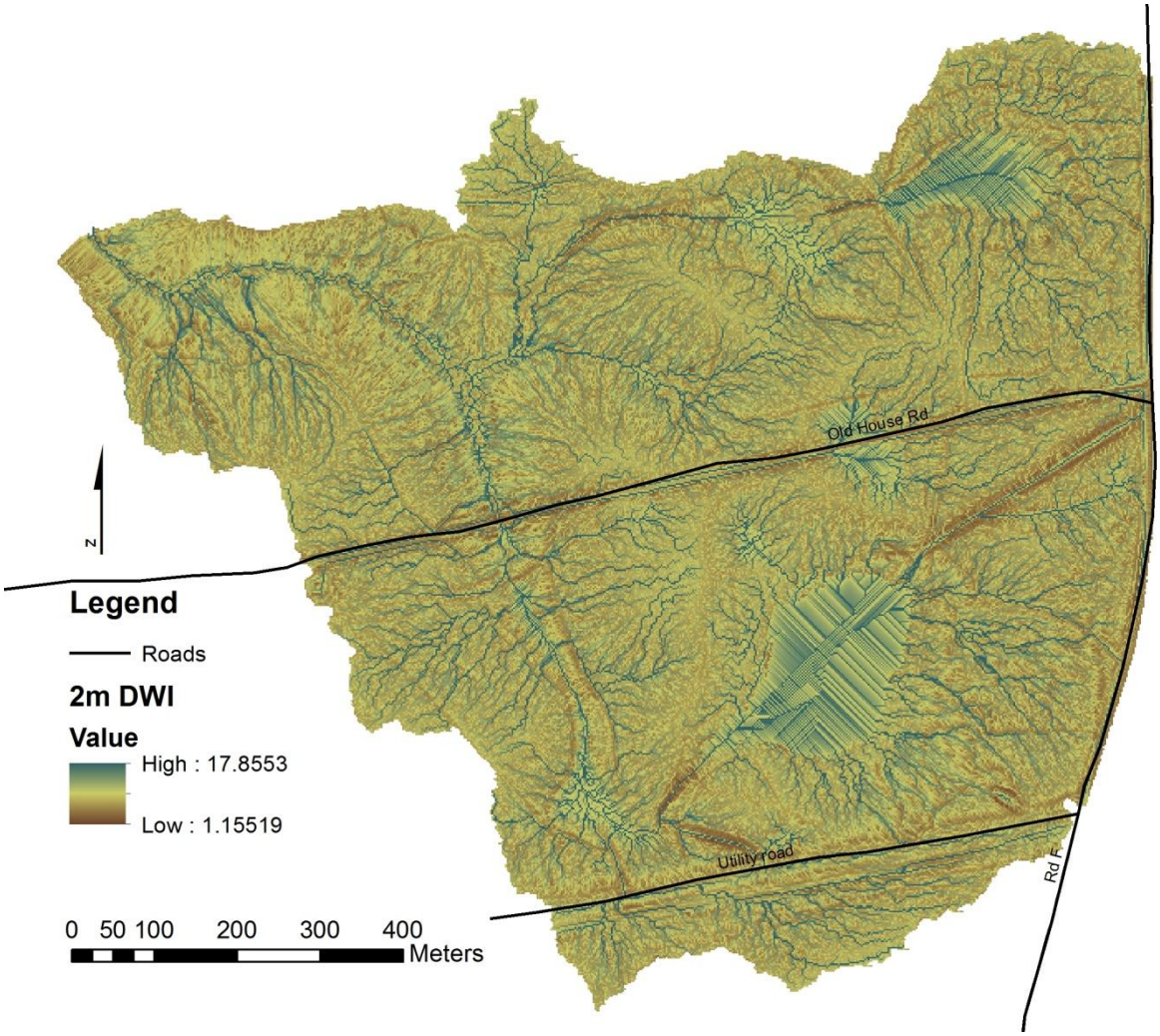

Figure 3.4.4. Map of 2m DWI in basin C.

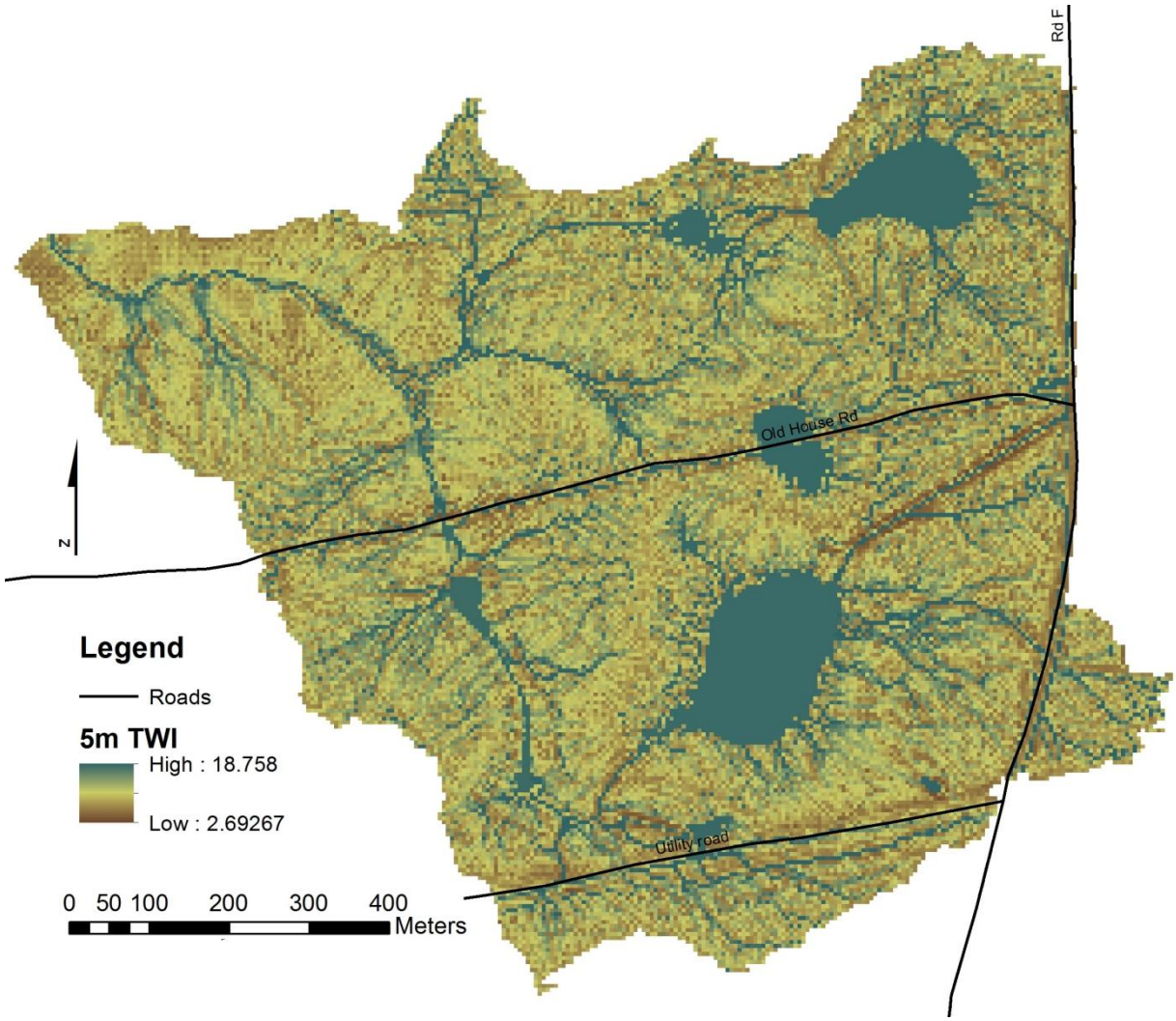

Figure 3.4.5. Map of 5m TWI in basin C.

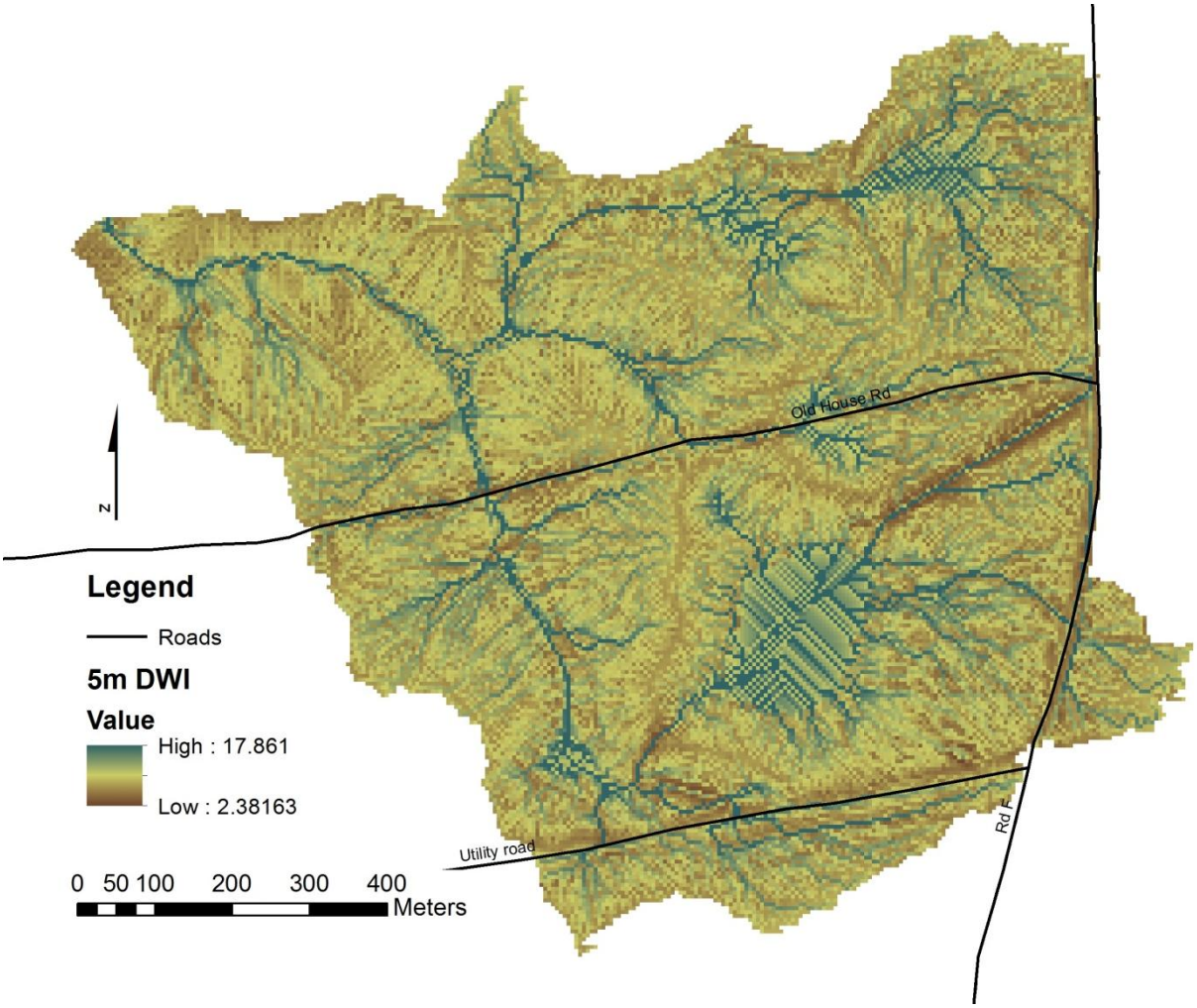

Figure 3.4.6. Map of 5m DWI in basin C.

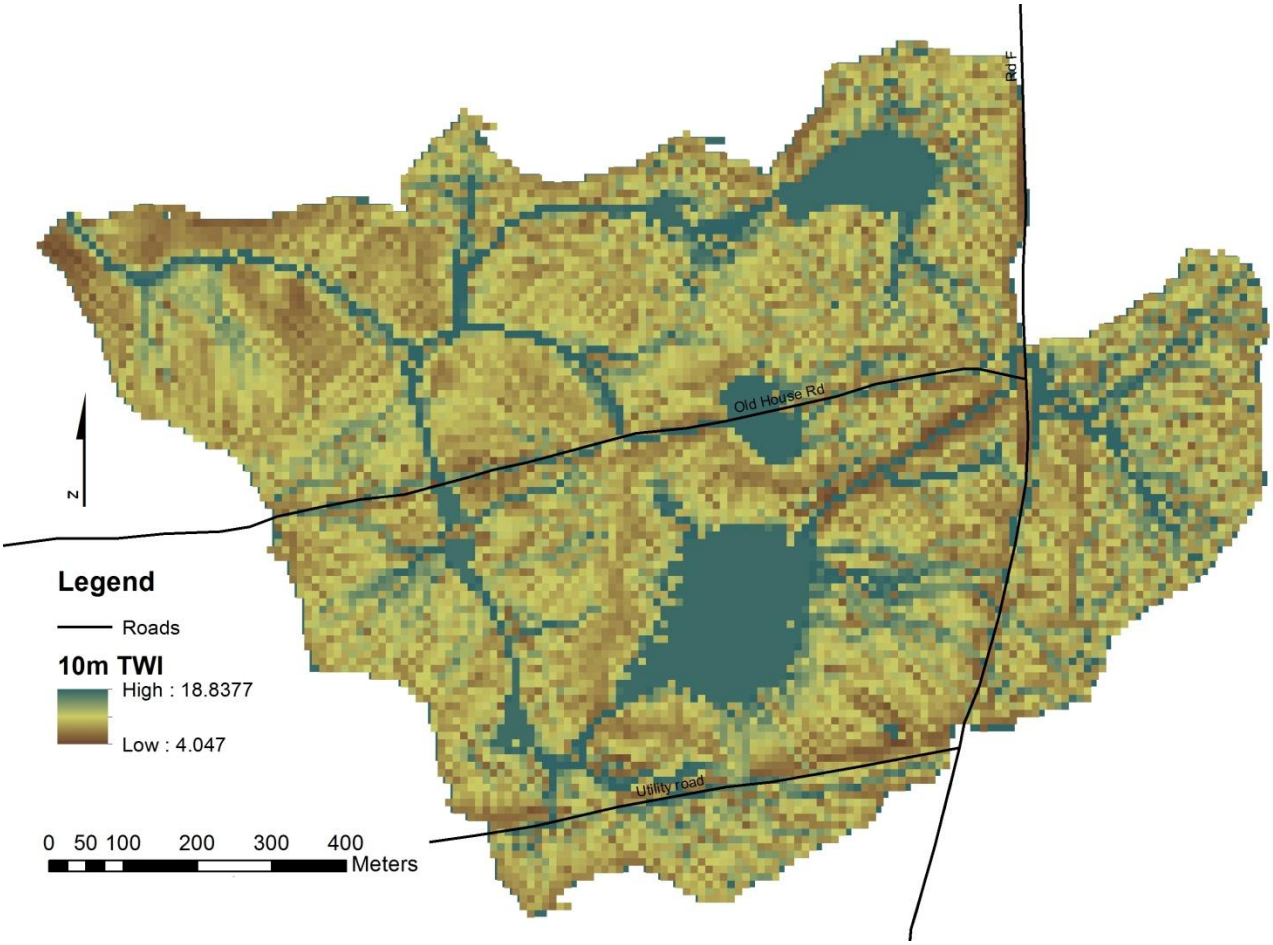

Figure 3.4.7. Map of 10m TWI in basin C.

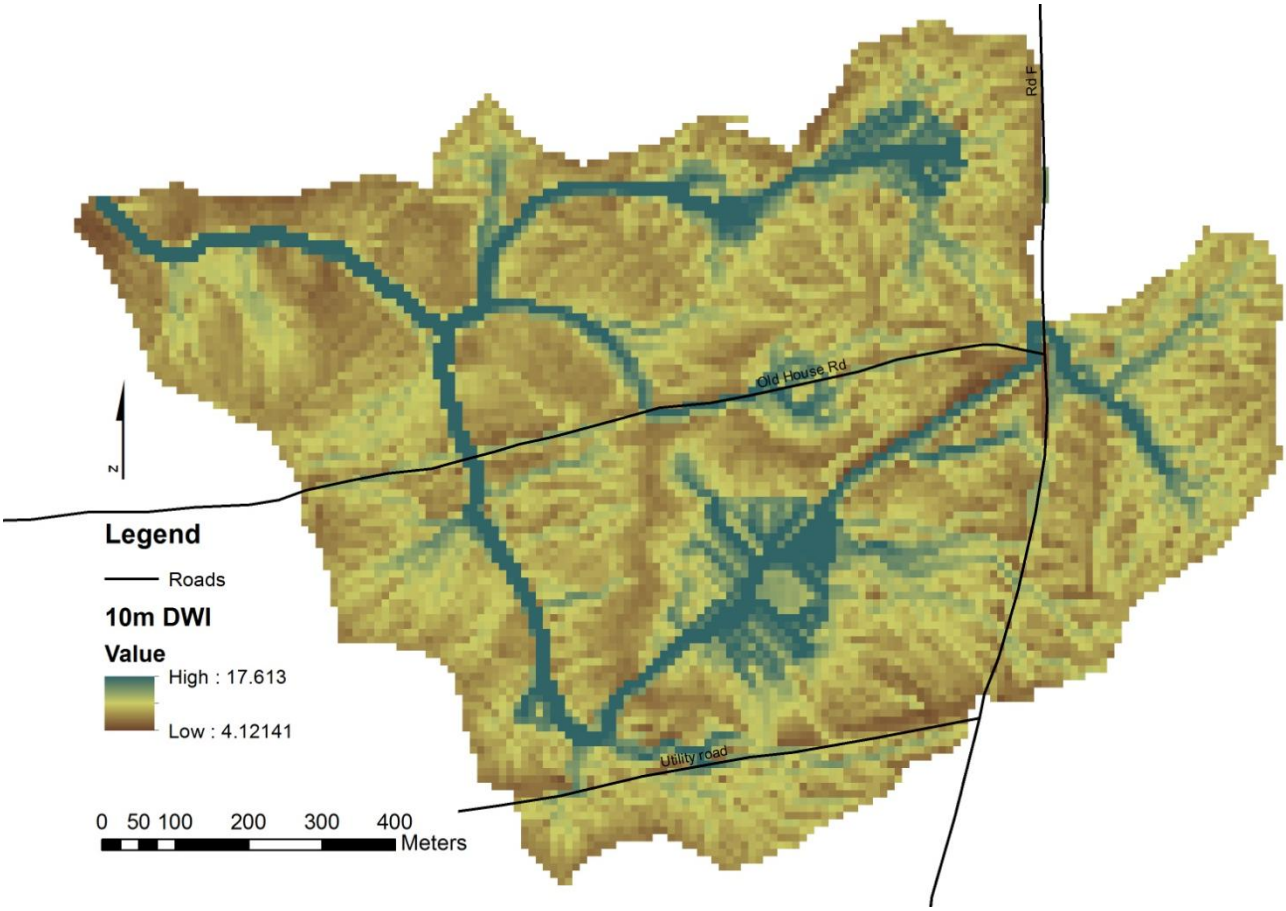

Figure 3.4.8. Map of 10m DWI in basin C.

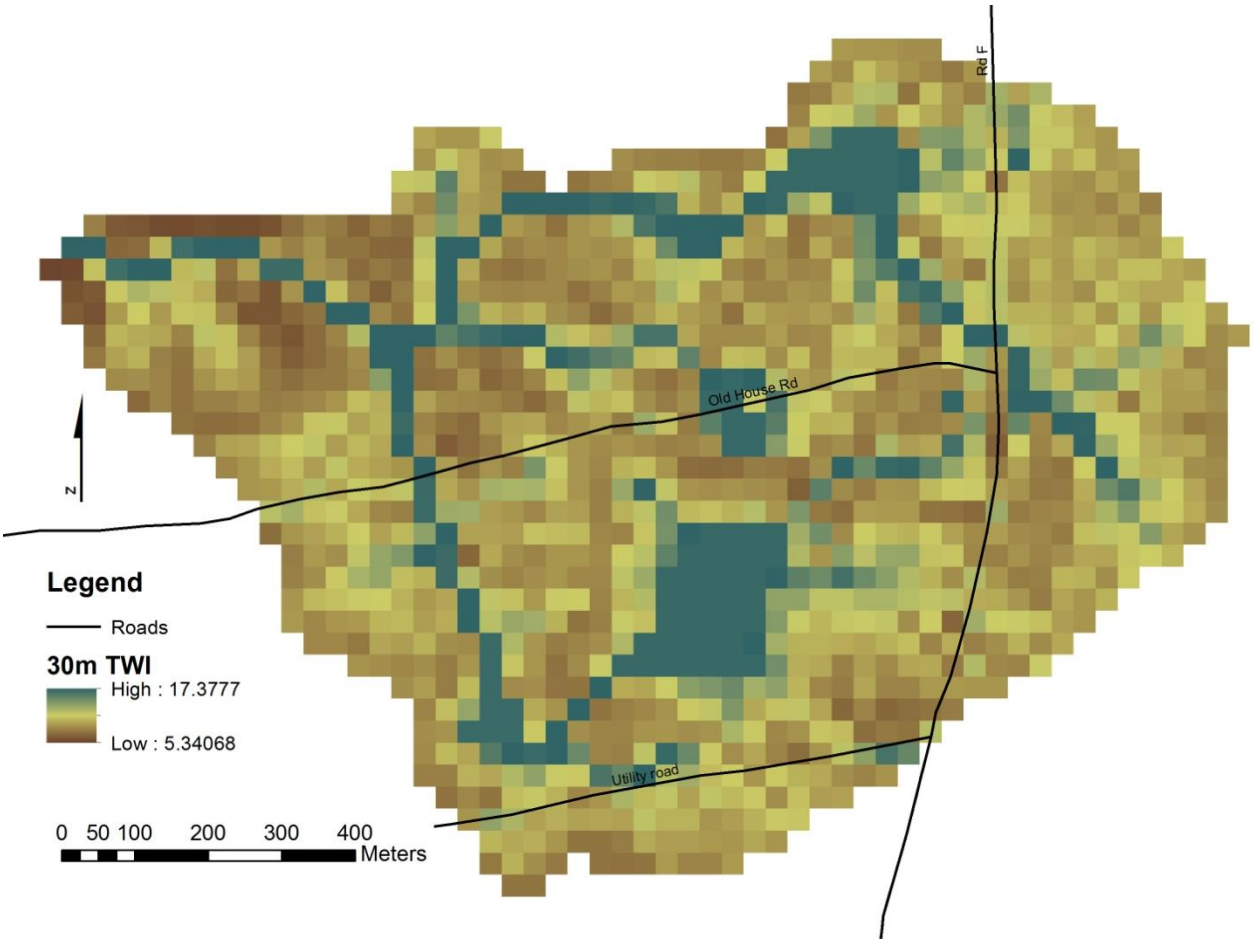

Figure 3.4.9. Map of 30m TWI in basin C.

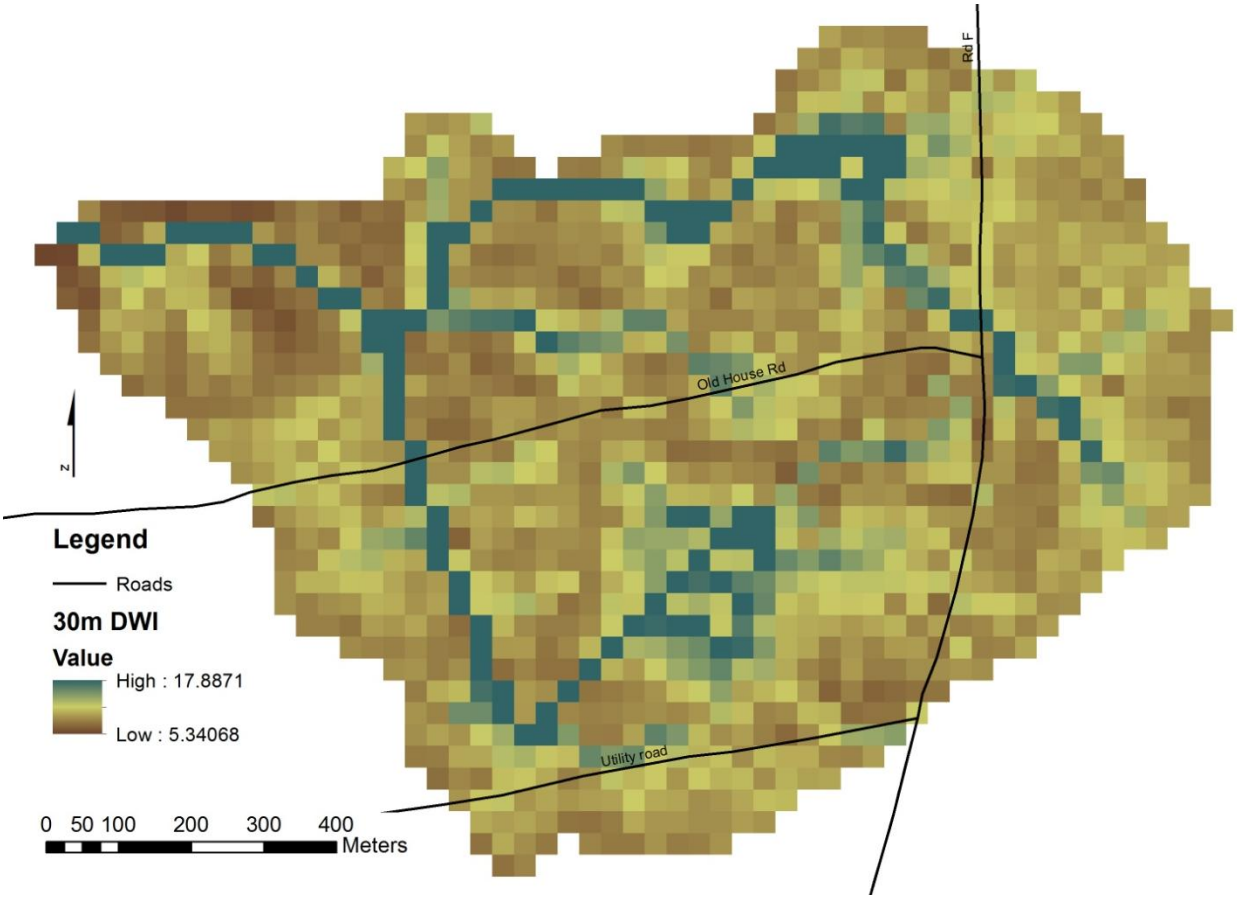

Figure 3.4.10. Map of 30m DWI in basin C.

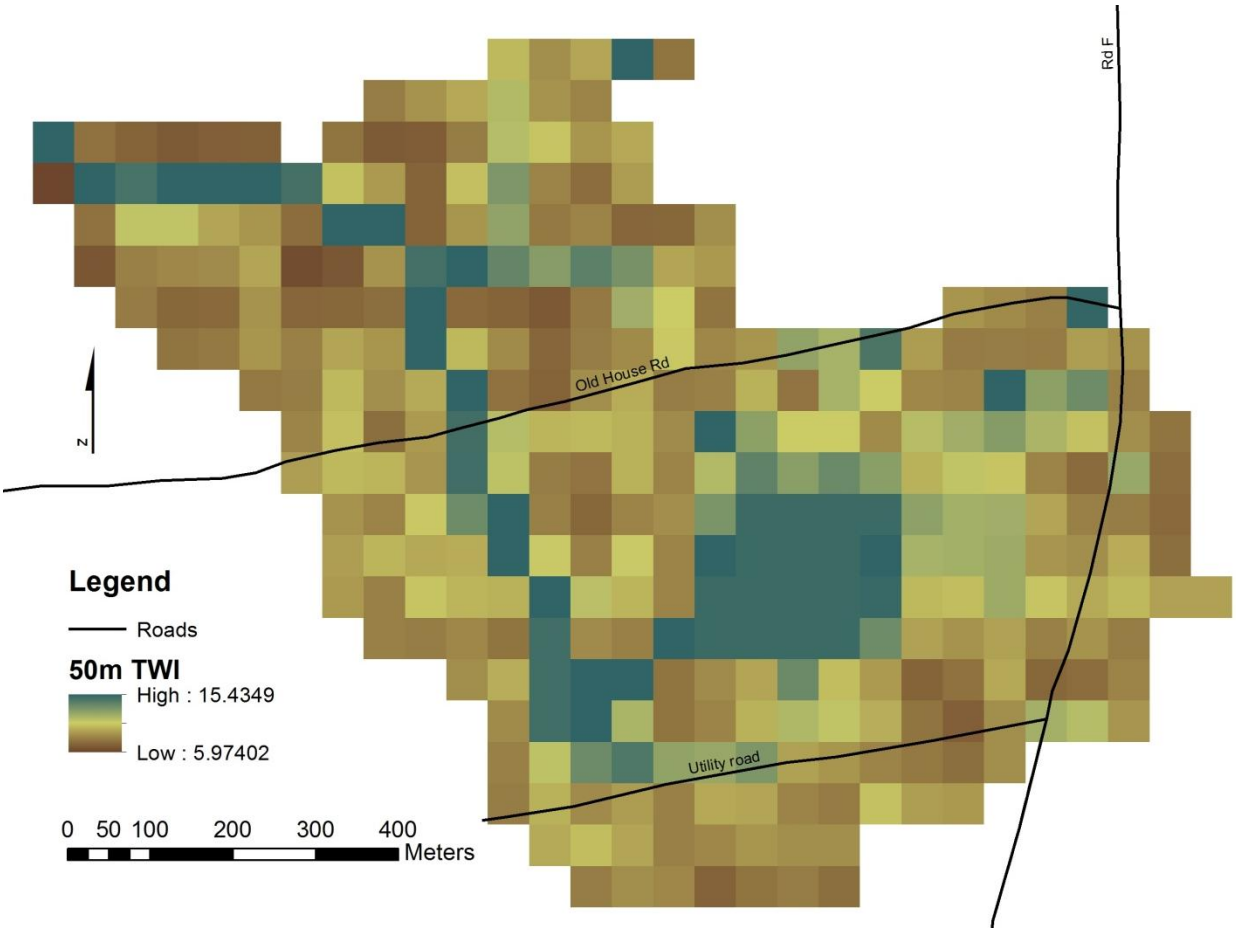

Figure 3.4.11. Map of 50m TWI in basin C.

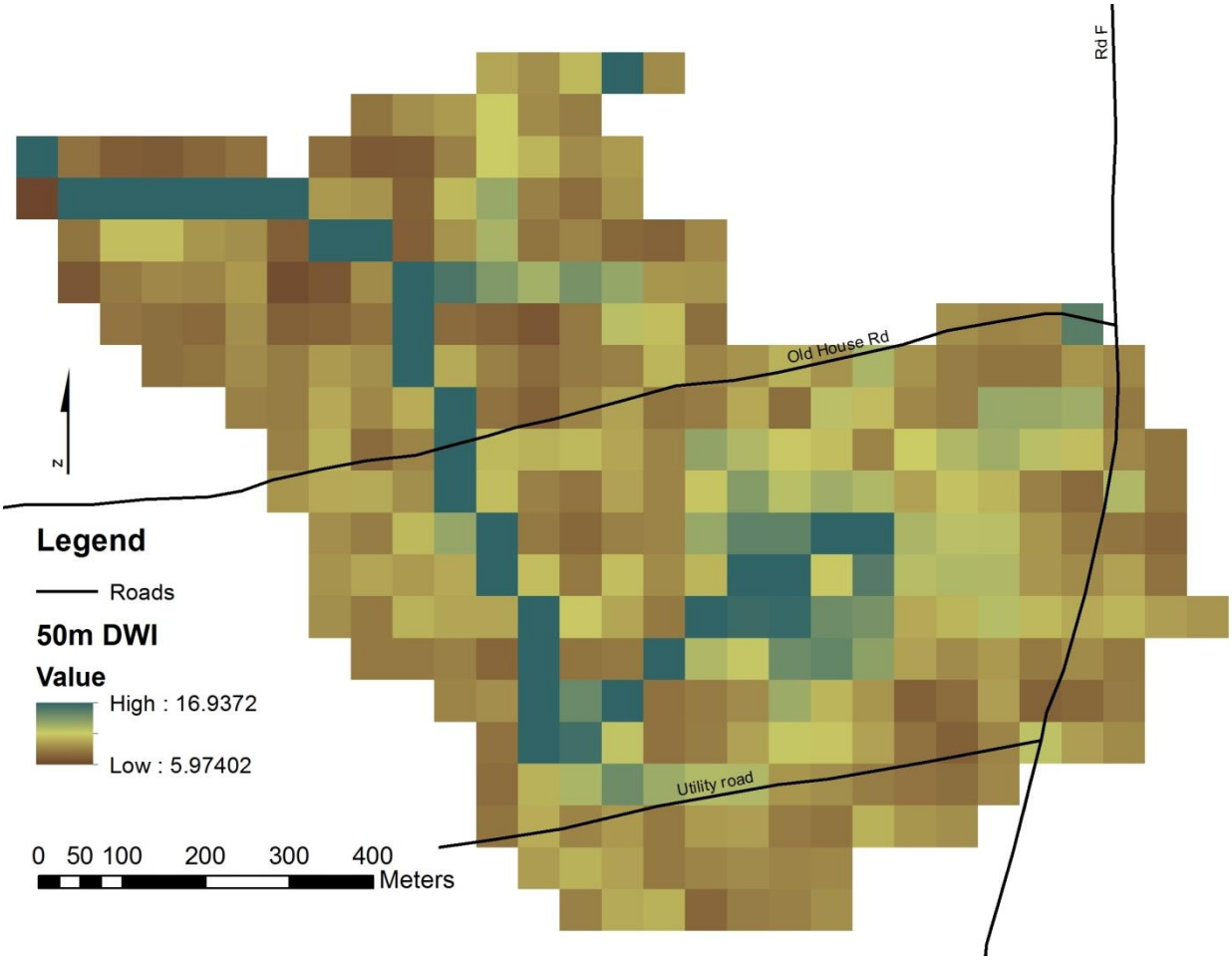

Figure 3.4.12. Map of 50m DWI in basin C.

3.4.2 Soil data

Prior to running non-linear regressions, data transformations were performed on each soil attribute where necessary. Figures 3.4.13 – 3.4.15 show the distributions of N and C in the mineral soil and organic layers before and after log-transformation. Silt was also log-transformed. Bulk density was transformed using a Box-Cox transformation with $\gamma = 2.5$. Electrical conductivity data for all three months were normal, so no transformation was necessary (Figure 3.4.16).

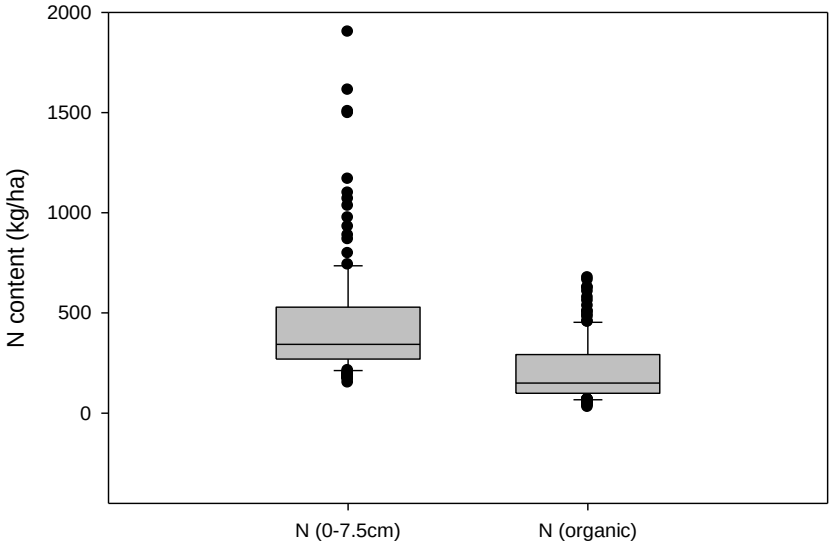

Figure 3.4.13. Distributions of nitrogen content for the mineral and organic soils at sample positions in the C basin, prior to log-transformation (n = 142).

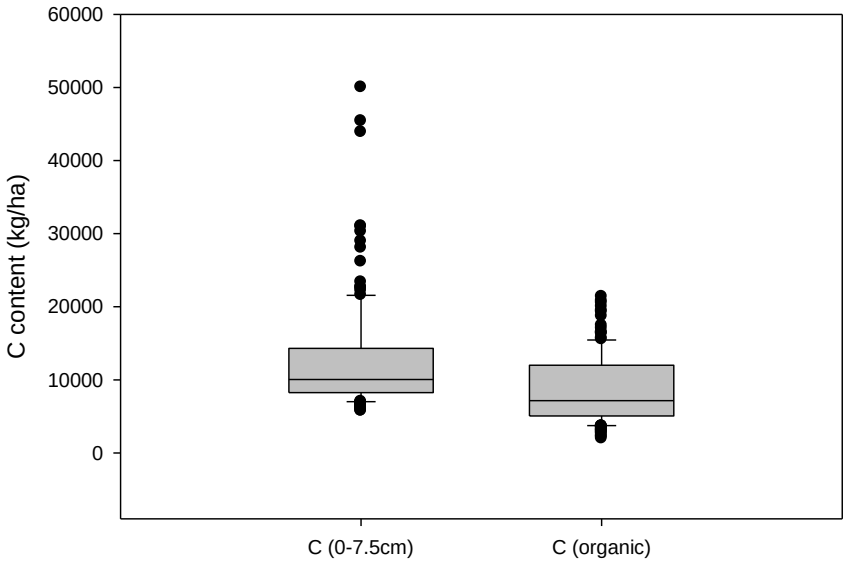

Figure 3.4.14. Distributions of carbon content for the mineral and organic soils at sample positions in the C basin, prior to log-transformation (n=142).

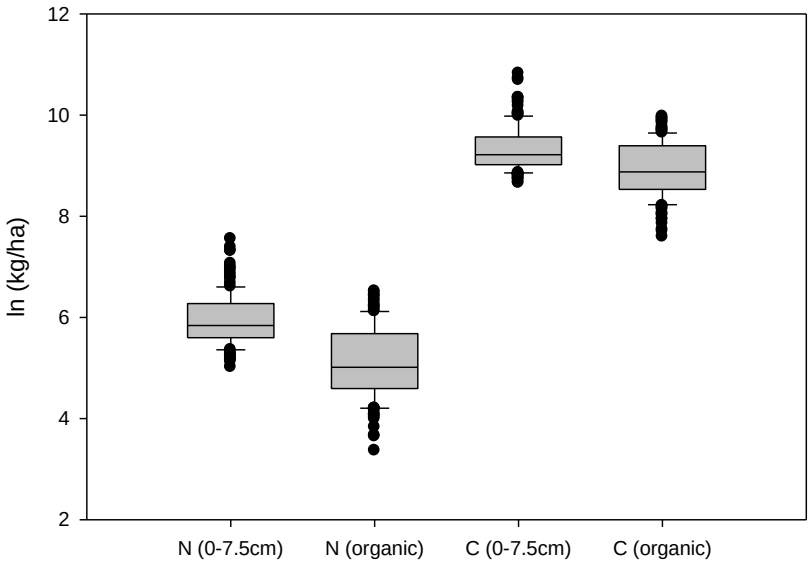

Figure 3.4.15. Distributions of nitrogen and carbon content for the mineral and organic soils at sample positions in the C basin, following log-transformation (n=142).

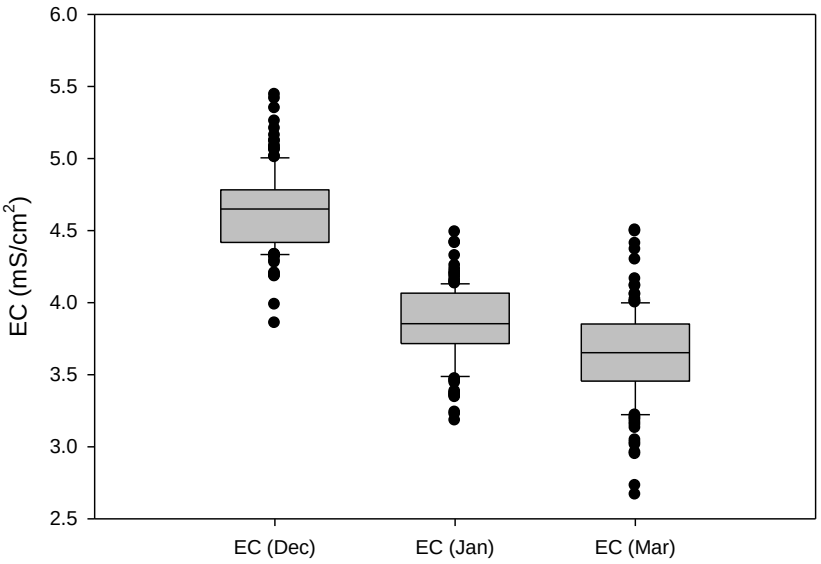

Figure 3.4.16. Distributions of electrical conductivity (EC) at sample positions in the C basin. No transformations were necessary (n = 236).

3.4.2 Non-linear relationships at the catchment scale

Linear regressions did not describe the relationships between soil attributes and indices as well as non-linear regressions (Figure 3.4.17). Also, the linear regressions violated the assumption of homoscedasticity, whereas the non-linear regressions showed reasonable homoscedasticity. Figures 3.4.17- 3.4.21 illustrate the differences among non-linear relationships between the soil attributes and wetness indices at various resolutions. The general pattern, as expected, for both N and C in both mineral and organic soil is that as moisture increases the N and C content decreases until a certain point and then increases again. The noise in the plots might have been caused by the sampling method. The samples were collected as points, and were not composited. Had they been composited, there might have been better agreement with the indices, especially as the cells increased in size.

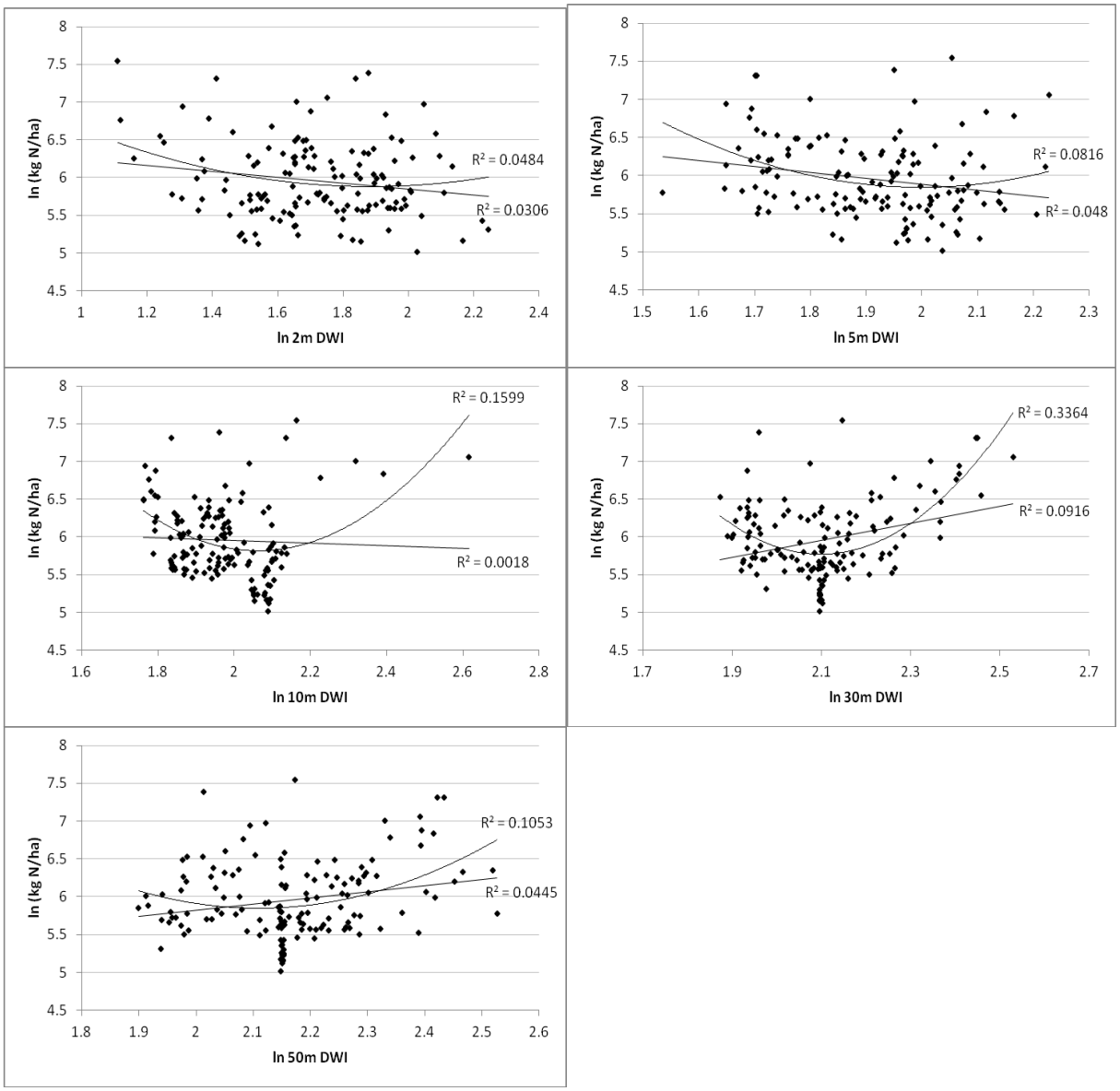

Figure 3.4.17. Linear and non-linear fits for DWI vs. N (0-7.5cm) at 2m, 5m, 10m, 30m, and 50m resolutions in the C basin. Note that the non-linear fit shows higher coefficient of determination and improves homoscedasticity (n=142).

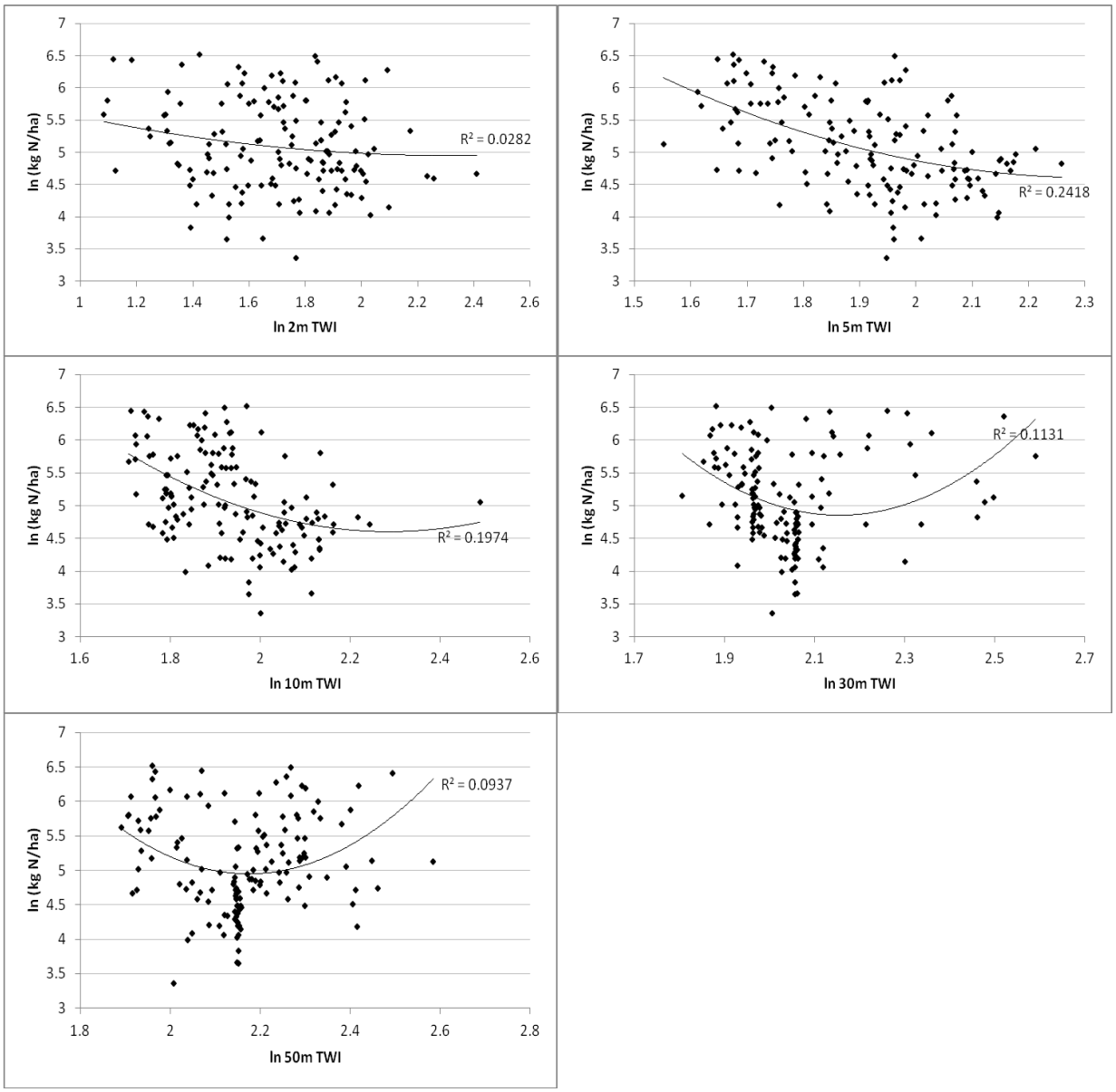

Figure 3.4.18. Non-linear fits for TWI vs. N (organic) at 2m, 5m, 10m, 30m, and 50m resolutions in the C basin (n=142).

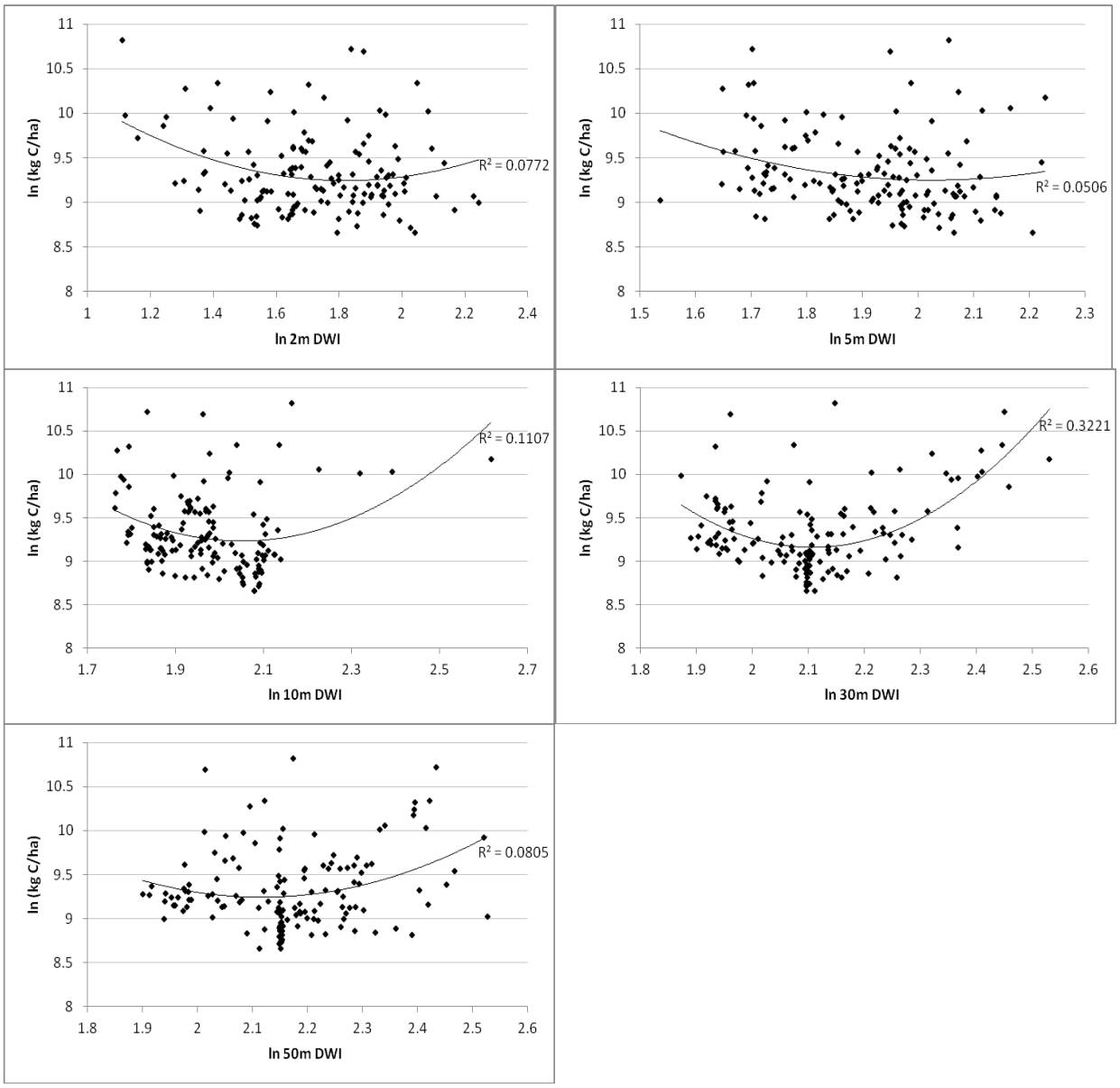

Figure 3.4.19. Non-linear fits for DWI vs. C (0-7.5cm) at 2m, 5m, 10m, 30m, and 50m resolutions in the C basin (n=142).

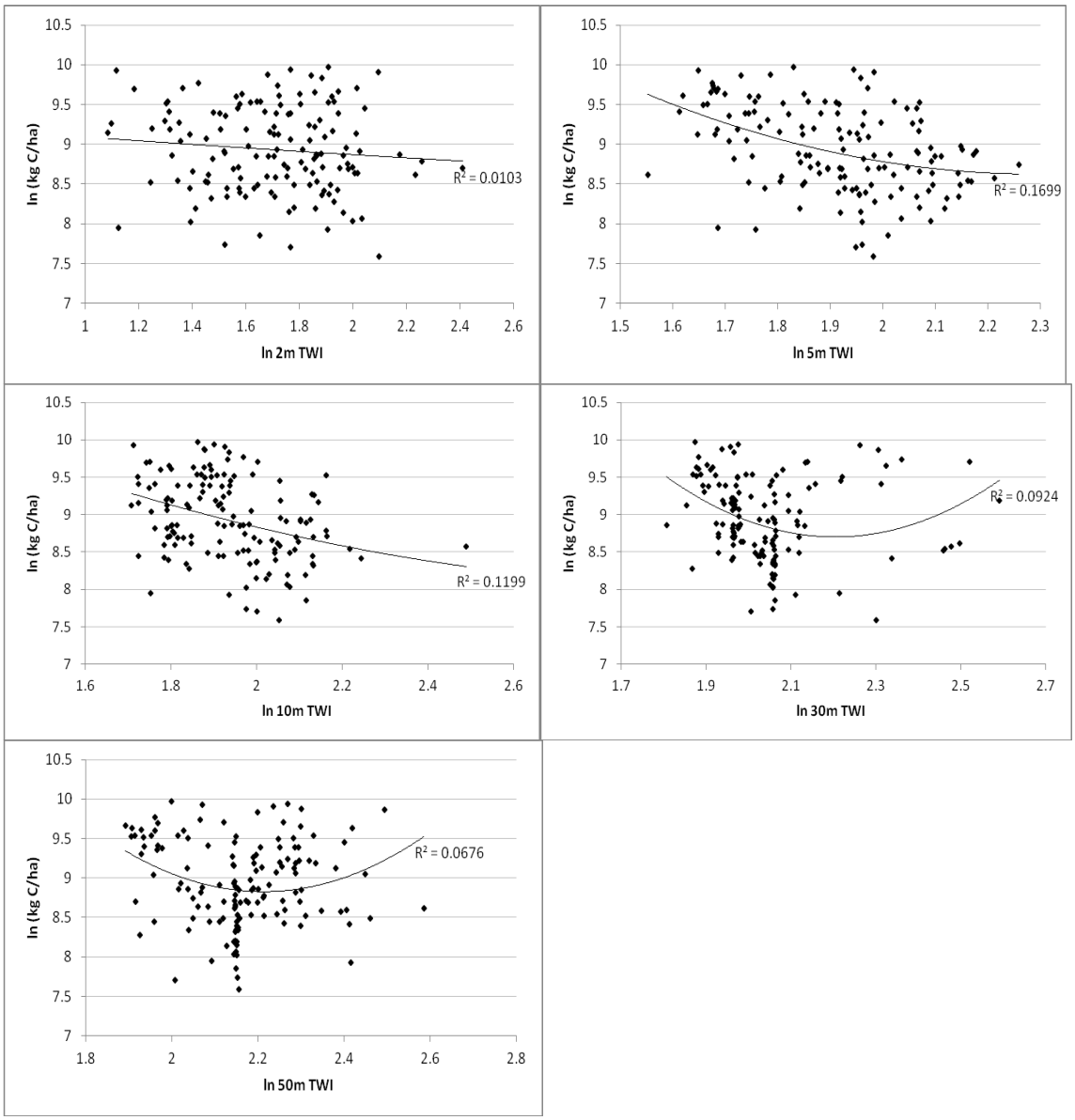

Figure 3.4.20. Non-linear fits for TWI vs. C (organic) at 2m, 5m, 10m, 30m, and 50m resolutions in the C basin (n=142).

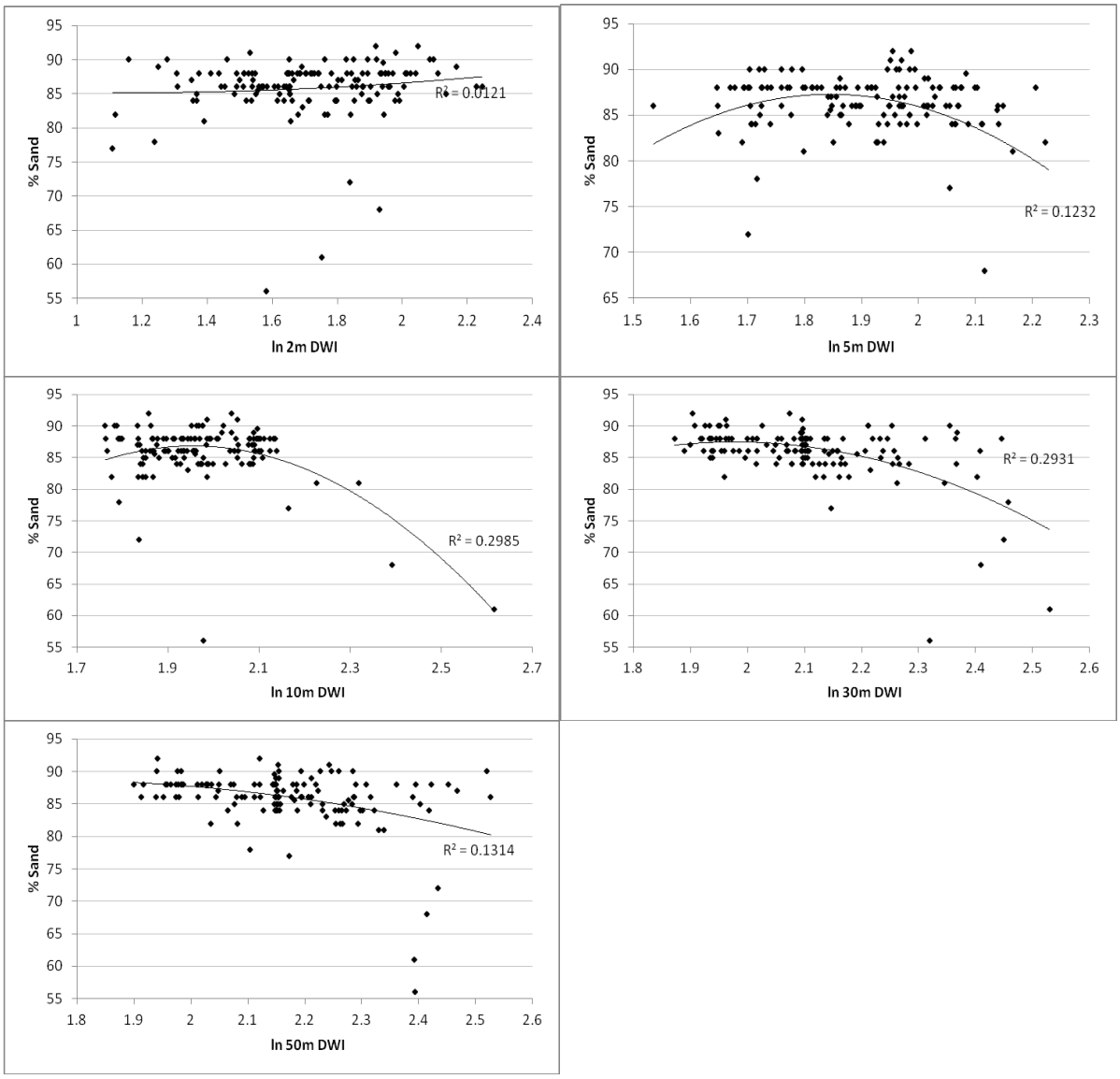

Figure 3.4.21. Non-linear fits for DWI vs. % sand at 2m, 5m, 10m, 30m, and 50m resolutions in the C basin (n=142).

After running non-linear regressions for each wetness index vs. each soil attribute within the C basin, it was found that DWI performed better overall for mineral soil attributes and TWI performed better overall for organic soil attributes (Table 3.4.3). In general, both indices calculated at the 30m resolution had the highest correlations with mineral soil attributes. The indices calculated at the 5m resolution had the highest correlations with organic soil attributes. Twenty-five and thirty-five percent of the variation of bulk density (D_b) was explained by the 30m TWI and DWI respectively. Within the 30m

resolution DWI explained more variation for all soil attributes, except December electrical conductivity (EC) and silt, than TWI. TWI showed a higher correlation with silt than DWI at the 30m resolution, but DWI showed a higher correlation with silt than TWI at the 10m resolution.

Variations in C and N content in the organic horizon were best explained by the 5m resolution of both TWI and DWI. They were the only samples that were taken above the mineral soil layer. Variations in C and N content in the mineral horizon were best explained by the 30m resolution of both TWI and DWI.

Electrical conductivity (mS/cm^2) showed the best relationships for December at the 10m DWI, and 30m TWI. The best relationships for January were at the 30m TWI and DWI. March EC showed very weak to no relationships with any resolution of either wetness index.

None of the soil profile characteristics measured (A horizon thickness, depth to argillic layer, and depth to redoximorphic features) were significantly correlated with either wetness index. Both indices at the 2m resolution showed very weak to no relationships with all of the soil attributes.

Table 3.4.3. Non-linear regression results (R^2) for wetness indices vs. soil attributes at the catchment scale. NS denotes $p>0.05$. Sample sizes for N and C in both mineral and organic horizons were 142, otherwise other attributes were n=236.

TWI	D_b	N (0-7.5cm)	C (0-7.5cm)	N (organic)	C (organic)	Dec EC	Jan EC	Mar EC	Silt
2m	0.05*	NS	NS	NS	NS	NS	NS	NS	NS
5m	0.17**	0.10**	0.08*	0.24**	0.17**	0.07**	0.09**	NS	0.10**
10m	0.20**	0.15**	0.11**	0.20**	0.12**	0.11**	0.08**	0.04*	0.18**
30m	0.25**	0.23**	0.20**	0.11**	0.09*	0.15**	0.12**	NS	0.23**
50m	0.13**	0.05	NS	0.09*	0.07*	0.05*	NS	NS	0.13**
DWI									
2m	0.06**	0.05	0.08*	0.05	NS	NS	NS	0.03	NS
5m	0.12**	0.08*	0.05	0.20**	0.15**	0.07**	0.07**	0.05*	0.13**
10m	0.28**	0.16**	0.11**	0.16**	0.09*	0.15**	0.06**	NS	0.29**
30m	0.35**	0.34**	0.32**	0.14**	0.11**	0.09**	0.26**	NS	0.22**
50m	0.17**	0.11**	0.08*	0.10**	0.08*	0.14**	0.05*	NS	0.22**

*p<0.01, **p<0.001.

3.4.3 Non-linear relationships at the hillslope scale

Within the N grid (Table 3.4.4), Bulk density, C and N in the mineral soil, and January EC all showed the best relationships with the 30m resolution of both indices. The organic horizon C and N had no relationships with either index at any resolution. The best relationship for silt was with the 30m TWI and the 10m DWI. The 2m resolution for both indices showed very weak to no relationships with any soil attributes.

Within the S grid (Table 3.4.5), there were very weak to no relationships between the soil attributes and the 2m and 5m resolutions of either wetness index. Bulk density also showed very weak to no relationships with either index at any resolution. The 30m and 50m resolutions for both indices showed the best relationships with N and C in the mineral and organic horizons. There was not one particular index or resolution that associated well with EC for all months. Silt showed the best relationship with the 30m TWI, with an R^2=0.52, p<0.0001. It is unclear why the organic horizon attributes had different relationships with the indices from catchment to hillslope scale.

Table 3.4.4. Non-linear regression results (R^2) for wetness indices vs. soil attributes at the hillslope scale (N grid). NS denotes p>0.05. Sample sizes for N and C in both mineral and organic horizons were 43, and all other attributes were n=100.

TWI	D_b	N (0-7.5cm)	C (0-7.5cm)	N (organic)	C (organic)	Dec EC	Jan EC	Mar EC	Silt
2m	0.05*	NS	NS	NS	NS	NS	NS	0.03	NS
5m	0.17**	NS	NS	NS	NS	0.07**	0.09**	NS	0.10**
10m	0.20**	0.26*	0.14	NS	NS	0.11**	0.08**	0.04*	0.18**
30m	0.25**	0.35**	0.26*	NS	NS	0.15**	0.12**	NS	0.23**
50m	0.13**	0.28*	0.26*	NS	NS	0.05*	NS	NS	0.13**
DWI									
2m	0.06**	NS	NS	NS	NS	NS	NS	0.03	NS
5m	0.12**	NS	NS	NS	NS	0.07**	0.07**	0.05*	0.13**
10m	0.28**	0.17	NS	NS	NS	0.15**	0.06**	NS	0.29**
30m	0.35**	0.29**	0.38**	NS	NS	0.09**	0.26**	NS	0.22**
50m	0.17**	0.26*	0.24*	NS	NS	0.14**	0.05*	NS	0.22**

*p<0.01, **p<0.001.

Table 3.4.5. Non-linear regression results (R^2) for wetness indices vs. soil attributes at the hillslope scale (S grid). NS denotes p>0.05. Sample sizes for N and C in both mineral and organic horizons were 83, and all other attributes were n=100.

TWI	D_b	N (0-7.5cm)	C (0-7.5cm)	N (organic)	C (organic)	Dec EC	Jan EC	Mar EC	Silt
2m	NS	NS	NS	NS	NS	NS	NS	NS	NS
5m	NS	NS	NS	NS	NS	NS	NS	0.10*	NS
10m	NS	NS	NS	0.14*	NS	0.19**	0.24**	0.19**	0.13*
30m	0.09	0.18**	0.17**	0.30**	0.16**	0.16**	0.36**	0.13*	0.52**
50m	0.08	0.22**	0.16**	0.24**	NS	NS	0.29**	0.11*	0.21**
DWI									
2m	NS	NS	NS	NS	NS	NS	NS	NS	NS
5m	NS	NS	NS	NS	NS	0.08	NS	0.10	NS
10m	0.07	0.13*	NS	NS	NS	0.20**	0.15**	0.30**	0.14*
30m	NS	0.19**	0.16**	0.22**	0.11	NS	0.25**	0.14*	0.14**
50m	0.08	0.20**	0.14*	0.22**	NS	NS	0.29**	0.11*	0.21**

*p<0.01, **p<0.001.

3.4.5 Predictions of soil attributes using terrain attributes

3.4.5.1 Catchment scale

Both N and C in the B and R basins were best predicted by the 30m TWI and bulk density (R^2 =0.20, p <0.0001 and R^2 = 0.16, p < 0.0001, respectively). The addition of other variables did not improve the relationships. The mean of measured A horizon N content in the test area (B and R basins, data combined) was 1005.76 ± 554.05 kg/ha and the mean of the predicted values was 512.05 ± 257.81 kg/ha. The RMSE for the predicted values was 626.27 kg/ha. The ranges of the measured and predicted values for the test area were included to place the RMSE in context. The range of the predicted N content for the test area was 229.40-1033.66 kg/ha, and the range of measured N content for the test area was 422.37-2004.20 kg/ha. The mean of measured A horizon C content in the test area was 23929.32 ± 8223.61 kg/ha and the mean of the predicted values was 13856.68 ± 5024.16 kg/ha. The RMSE for the predicted values was 11637.76 kg/ha. The range of predicted C content for the test area was 7497.95-23495.12 kg/ha, and the range of measured C content for the test area was 10807.85-34743.76 kg/ha.

The best combination of variables to predict bulk density were 5m TWI and silt ($R^2 = 0.13$, $p<0.0001$). The mean of measured bulk density values in the test area was 1.092 ± 0.283 g/cm^3 and the mean of the predicted values was 1.102 ± 0.129 g/cm^3. The RMSE for the predicted values was 0.246 g/cm^3. The range of predicted bulk density for the test area was 0.859-1.318 g/cm^3, and the range of measured bulk density for the test area was 0.633-1.566 g/cm^3.

3.4.5.2 Hillslope scale

The best combination of variables to predict A horizon N content were 30m TWI, bulk density, organic horizon sample mass, and December EC ($R^2 = 0.54$, $p <0.0001$). The mean of A horizon N content in the test area (south of Old House Road) was 314.08 ± 159.17 kg/ha and the mean of the predicted values was 379.44 ± 84.65 kg/ha. The RMSE for the predicted values was 213.12 kg/ha. To put the RMSE in context, the range of the test variable was 151.30-1611.81 kg/ha and the range of the predicted variable was 227.21-755.54 kg/ha.

The best combination of variables to predict A horizon C content were 30m TWI, bulk density, organic horizon sample mass, and December EC ($R^2 = 0.49$, $p <0.0001$). The mean of A horizon C content in the test area was 9392.29 ± 3913.01 kg/ha and the mean of the predicted values was 10398.23 ± 2172.10 kg/ha. The RMSE for the predicted values was 5427.88 kg/ha. The range of the test variable was 5762.81-43889.51 kg/ha and the range for the predicted variable was 6838.39-20753.96 kg/ha.

The best combination of variables to predict organic horizon N content were 5m TWI and bulk density ($R^2 = 0.22$, $p <0.0001$). The mean of organic horizon N content in the test area was 116.57 ± 82.10 kg/ha and the mean of the predicted values was 255.50 ± 72.14 kg/ha. The RMSE for the predicted values was 160.79 kg/ha. The range of the test variable was 28.87-575.90 kg/ha and the range for the predicted variable was 157.04-489.61 kg/ha.

The best combination of variables to predict organic horizon C content were 5m TWI and bulk density ($R^2 = 0.31$, $p <0.0001$). The mean of organic horizon C content in the test area was 6206.39 ± 3550.92 kg/ha and the mean of the predicted values was 11099.18 ± 2917.48 kg/ha. The RMSE of the

predicted values was 6017.08 kg/ha. The range of the test variable was 2212.10-20773.22 kg/ha and the range for the predicted variable was 6989.49-21075.67 kg/ha.

The best combination of variables to predict bulk density (0-7.5cm depth) were 30m DWI, organic matter weight, and silt ($R^2 = 0.37$, $p < 0.0001$). The mean of bulk density in the test area was 1.303 ± 0.167 g/cm^3 and the mean of the predicted values was 1.073 ± 0.048 g/cm^3. The RMSE for the predicted values was 0.245 g/cm^3. The range of the test variable was 1.076-1.569 g/cm^{3}and the range for the predicted variable was 0.896-1.182 g/cm^3.

3.4.6 Discussion of regressions and model predictions

In general, stronger relationships existed between soil attributes and the indices that were calculated from coarser resolutions (i.e. 10 – 50m) than with those calculated from finer resolutions. This might be the result of clustering of soil attribute values, reflecting the scale of the processes behind the values. The larger cells would accommodate clusters (see Figure 3.4.18), or would match the scale of the processes better than the smaller cells. One exception to this general trend was the relationships between the 5m indices and the organic layer C and N. One possibility for this exception could be that the organic layer processes are much more local and affected by smaller changes in surface topography than the mineral layer. The spatial relationships among terrain and soil attributes will be explored further in Chapter 4.

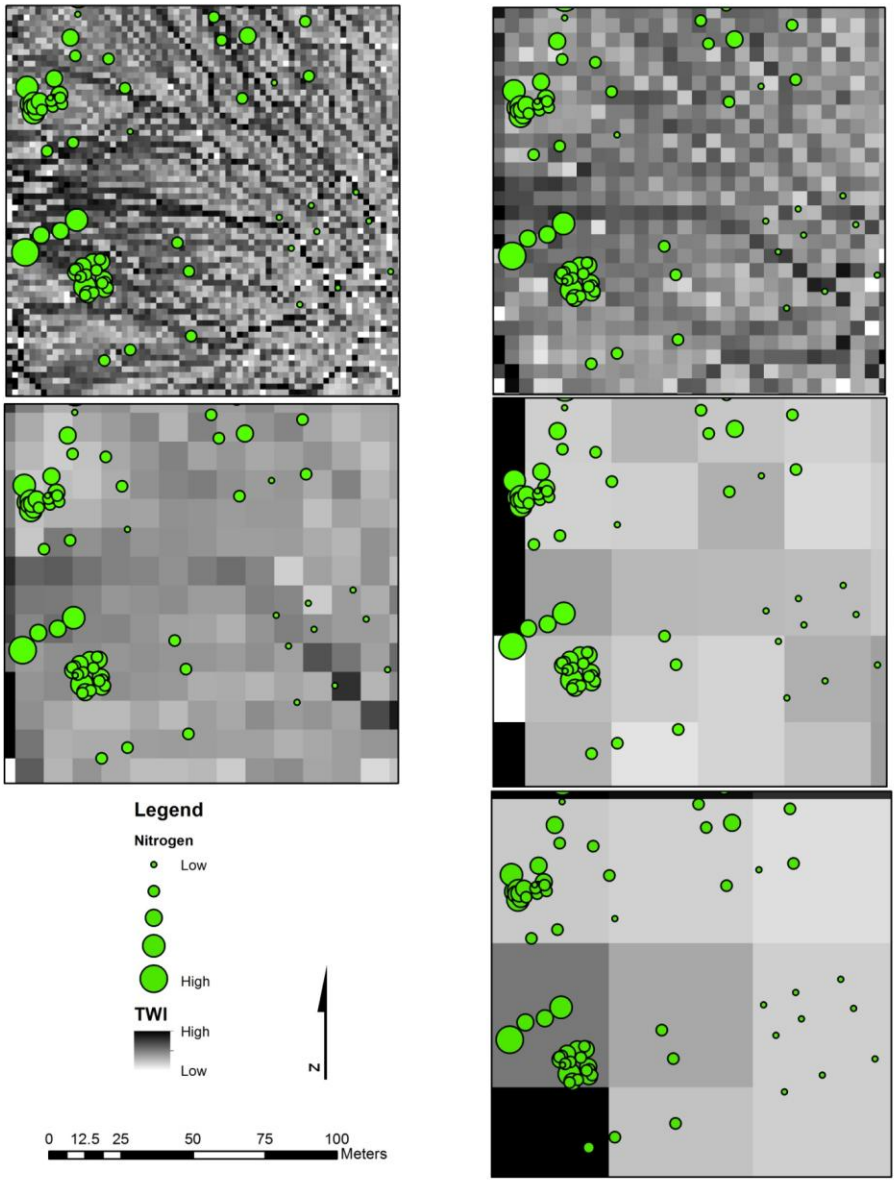

Figure 3.4.22. TWI maps at different resolutions. The nitrogen values are relatively clustered and the larger cells accommodate this clustering better than the smaller cells.

3.5 Conclusion

Comparisons among the indices calculated at different resolutions revealed that as resolution decreased, the ranges of index values decreased. This has the effect of underestimating wet areas and overestimating dry areas as resolutions become increasingly coarse. The choice of resolution had the smallest effect on the medians of wetness indices when choosing between 30m and 50m resolutions, and the smallest effect on the ranges of wetness indices when choosing between 2m and 5m resolutions.

Finer resolution wetness indices showed better detail and separation of distinct variable source areas. As resolutions decreased, the wettest areas shifted from the variable source areas to the ephemeral streams.

Non-linear regressions showed that at the catchment scale C and N in mineral soil correlated better with 30m resolution indices, and C and N in the organic horizon correlated better with 5m resolution indices. At the hillslope scale, C and N in mineral soil correlated best with 30m resolution indices. However, C and N in the organic horizon only showed significant correlation with indices in the S grid, and the best correlations were with 30m resolution indices.

As suggested by the non-linear regressions at the catchment scale, the models that best predicted the mineral soil attributes included indices at the 30m resolution, and the models that best predicted the organic horizon attributes included indices at the 5m resolution. This pattern might be related to the scales of processes affecting soil change in each horizon. For example, the distribution of vegetation and leaf litter might change at a smaller scale than the hydrologically driven distribution of N and C in the mineral soil.

3.6 Literature Cited

Bouyoucos, G. J. 1951. A recalibration of the hydrometer method for making mechanical analysis of soils. *Agronomy Journal 43*: 434-488.

Bouyoucos, G. J. 1962. Hydrometer method improved for making particle size analysis of soils. *Agronomy Journal 54*: 464-465.

Burrows S.N., S.T. Gower, M.K. Clayton, D.S. Mackay, D.E. Ahl. 2002. Application of geostatistics to characterize leaf area index (LAI) from flux tower to landscape scales using a cyclic sampling design. *Ecosystems 5*: 667- 79.

Fraisse, C.W., K.A. Sudduth, N.R. Kitchen. 2001. Delineation of site-specific management zones by unsupervised classification of topographic attributes and soil electrical conductivity. *Transactions of the ASAE 44*(1): 155–166.

Gee, G.W., J.W. Bauder. 1979. Particle size analysis by hydrometer: a simplified method for routine textural analysis and a sensitivity test of measurement parameters. *Soil Science Society of America Journal 43*: 1004-1007.

Gessler, P.E., I.D. Moore, N.J. McKenzie, P.J. Ryan. 1995. Soil–landscape modelling and spatial prediction of soil attributes. *International Journal of Geographical Information Systems 9*: 421–432.

Gobin, A., P. Campling, J. Feyen. 2001. Soil-landscape modeling to quantify spatial variability of soil texture. *Physics and Chemistry of the Earth. Part B: Hydrology, Oceans and Atmosphere 26* (1): 41– 45.

Hjerdt, K. N., J. J. McDonnell, J. Seibert, and A. Rodhe. 2004. A new topographic index to quantify downslope controls on local drainage, *Water Resources Research 40*, W05602, doi:10.1029/2004WR003130.

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

Park, S.J., K. McSweeney, B. Lowery. 2001. Identification of the spatial distribution of soils using a process-based terrain characterization. *Geoderma 103*: 249–272.

Quinn, P., K. Beven, P. Chevallier, O. Planchon. 1991. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrological Processes 5*: 59–79.

Tarboton, D. G. 1997. A new method for determination of flow directions and upslope area in grid digital elevation models. *Water Resources Research 33*(2): 309–319.

Thompson, J.A., J.C. Bell, C.A. Butler. 2001. Digital elevation model resolution: effects on terrain attribute calculation and quantitative soil– landscape modeling. *Geoderma 100*: 67–89.

Wolock, D. M. and C. V. Price. 1994. Effects of digital elevation model map scale and data resolution on a topography-based watershed model. *Water Resources Research 30*: 3041-3052.

Zhang, W. and D. R. Montgomery. 1994. Digital elevation model grid size, landscape representation, and hydrologic simulations. *Water Resources Research 30*(4): 1019-1028.

Chapter 4

DEM RESOLUTION EFFECTS ON SPATIAL RELATIONSHIPS AND PREDICTION

4.1 Introduction

Chapter 3 discussed the aspatial relationships between soil and terrain attributes calculated from different DEM resolutions. Ordinary Least Squares (OLS) regression and associated descriptive statistics are global models that provide an appropriate starting point to understanding the relationships among variables. However, they lack information on variation with geographical location. When using global models like OLS, the independence of data is assumed, but when using autocorrelation the assumption is implicitly that data are not independent. This chapter will focus on the variation of soil and terrain attributes across space. The question remains the same: Is there an optimal DEM resolution that will predict soil attributes using a topographic wetness index? Spatial analysis will be used to identify clusters of high and low attribute values, and will help to visualize the spatial nature of the relationship between terrain and soil attributes. Additionally, end-products of this study will be continuous prediction surfaces of several soil attributes produced using geostatistics. The ability to produce accurate prediction maps from sparse soil sampling and readily available terrain attributes is valuable to inform sampling and land use planning, because it reduces the cost and increases the efficiency of those activities (Moore et al. 1993).

Regionalized Variable Theory (RVT) uses the probability distribution of a variable at a given point in space to estimate values at unrecorded places and give errors for those estimates (Webster and Oliver 2001). RVT is based on Tobler's First Law of Geography, which states that everything is related to everything else, but near things are more related than distant things (Tobler 1970). While OLS regression and other aspatial methods assume independence of residuals (e.g. absence of spatial autocorrelation), geostatistics uses spatial autocorrelation of residuals to provide more information for use

in prediction. A semivariogram is a tool used for measuring spatial dependence by using the distance between pairs of variables as a parameter.

The semivariogram is half the average squared difference of the values of a variable at two positions, and is plotted against the distance between those two positions. The pairs are placed in lags, much like the bins used in histograms, to make use of the empirical semivariogram more manageable. The correct choice of lag distance and number of lags is important for defining the shape and extent of the semivariogram. According to Tobler's First Law, it is expected that as distance increases, the semivariogram will increase. Pairs that are closer together should have a smaller difference than those that are spaced farther apart. After a certain distance, known as the range, the semivariogram will reach an asymptote (i.e sill) where spatial correlation ceases. The nugget (y-intercept) is the random unknown error among the data, and the partial sill is the difference between the nugget and sill. The nugget, range and sill are used to describe the model fit to the semivariogram, which is used to perform geostatistical methods such as kriging.

Kriging can be used to make predictions across a continuous surface, based on discrete measured points. It also provides a measure of the error of those predictions both locally and globally, by using a cross-validation technique. Universal kriging with anisotropy makes use of the assumption that not only distance but also direction between pairs of sample points can be used to explain variation in the landscape (ESRI Inc.). There are different types of kriging, and many publications (Webster and Oliver 2001, Wackernagel 2006, Diggle and Ribeiro 2007) have addressed these differences, so not much space will be devoted to them here. Choosing the right type of kriging depends primarily on the knowledge of the trend of the data. If the mean is assumed constant but unknown, then ordinary kriging is appropriate. If the mean is instead a function of the locations across space, such as a second-order polynomial, then universal kriging is probably a better choice. The function for universal kriging at each location is essentially a linear or non-linear regression equation: $Z(s) = \mu(s) + \varepsilon(s)$, where $Z(s)$ is the value of the variable at a certain location s, $\mu(s)$ is the function of the trend, and $\varepsilon(s)$ is the random error at location s. Kriging uses location as one of the explanatory variables in the regression.

Cokriging uses cross-covariance of residuals between two or more variables in addition to spatial autocorrelation to help describe the model on which the prediction is based. Cokriging was the method that was used in this study, because the goal was to improve predictions of soil attributes using terrain attributes. The terrain attributes were derived from DEMs, so they could be sampled more heavily than the soil attributes.

Data are not always suitable for analysis using geostatistics. Before using kriging, it is important to use exploratory spatial data analysis to discover trends, distribution, autocorrelation, and outliers. For this purpose, the data for each attribute were examined using histograms, Voronoi maps, trend analysis, global and local Moran's I, semivariograms and cluster analysis before any kriging or cokriging was attempted.

4.2 Methods

4.2.1 Suitability analysis for geostatistics

The global Moran's I (Spatial Statistics Tools) was used to test for spatial autocorrelation for each soil attribute to see if the data were suitable for use in geostatistics. If the Z score, index, and p values indicated that the data were significantly clustered, Anselin local Moran's I (Spatial Statistics Tools) was used to see where the clustering occurred, and if there was a trend in the clustering across the sample positions. The local Moran's I produced a map of local Z scores, indices, and p values for each sample position. It also indicated if the position was part of a cluster of high or low values. This was valuable for visualization of patterns across the study area.

To test for trends in the data, a trend analysis was used (Geostatistical Analyst). This analysis shows the type of trend and the direction of the trend. By rotating the locations with respect to the grids, the trend changed so that the direction of the trend could be seen. This analysis views the data from a horizontal position, in contrast to the other approaches used here that look at the data from above.

Finally, Voronoi diagrams (Geostatistical Analyst) were used to visually inspect patterns of attribute values across the study area. This was useful for comparing the soil attributes with terrain attributes (e.g. elevation and wetness indices), and with other soil attributes (e.g. texture and EC).

4.2.2 Geostatistics

If the data were significantly spatially autocorrelated, and there was a trend, the next step was to krige the data to create a prediction surface. Each variable was inspected for normality using a histogram, and data were transformed if necessary. Because there were trends found, universal kriging was chosen.

The semivariogram for the soil attribute was inspected as a preliminary step to prediction kriging. The search direction tool was used to detect anisotropy among the data. Anisotropy direction was aligned with the direction where the semivariogram slope was the greatest, indicating a more rapid change in differences between pairs in that direction. The lags were adjusted to the optimal search distance. The number of neighbors chosen was a minimum of 10 and a maximum of 30. This seemed like a reasonable number since the number of positions was between 142 and 263, depending on the dataset used.

The data were cokriged with wetness indices, and the resulting surfaces were compared using RMSE to see which wetness index was able to improve the predictions the best. The wetness indices were sampled over the entire watershed, but only those within the area of the soil attribute sample points were included. A standard study area polygon was drawn around the outermost points (Figure 4.2.1) and filled with 5,000 random points to which were extracted the values of TWI, DWI, and elevation rasters at all resolutions. After accounting for missing data among the random points, the actual count was 3,917. This was done to reduce bias due to the different number of points and point locations in the basin at each resolution. After the soil attribute was kriged, the same major and minor range, anisotropy direction, lag size and number of lags were retained for all cokriging with terrain attributes. This was done to constrain for the effect of resolution on the cokriging.

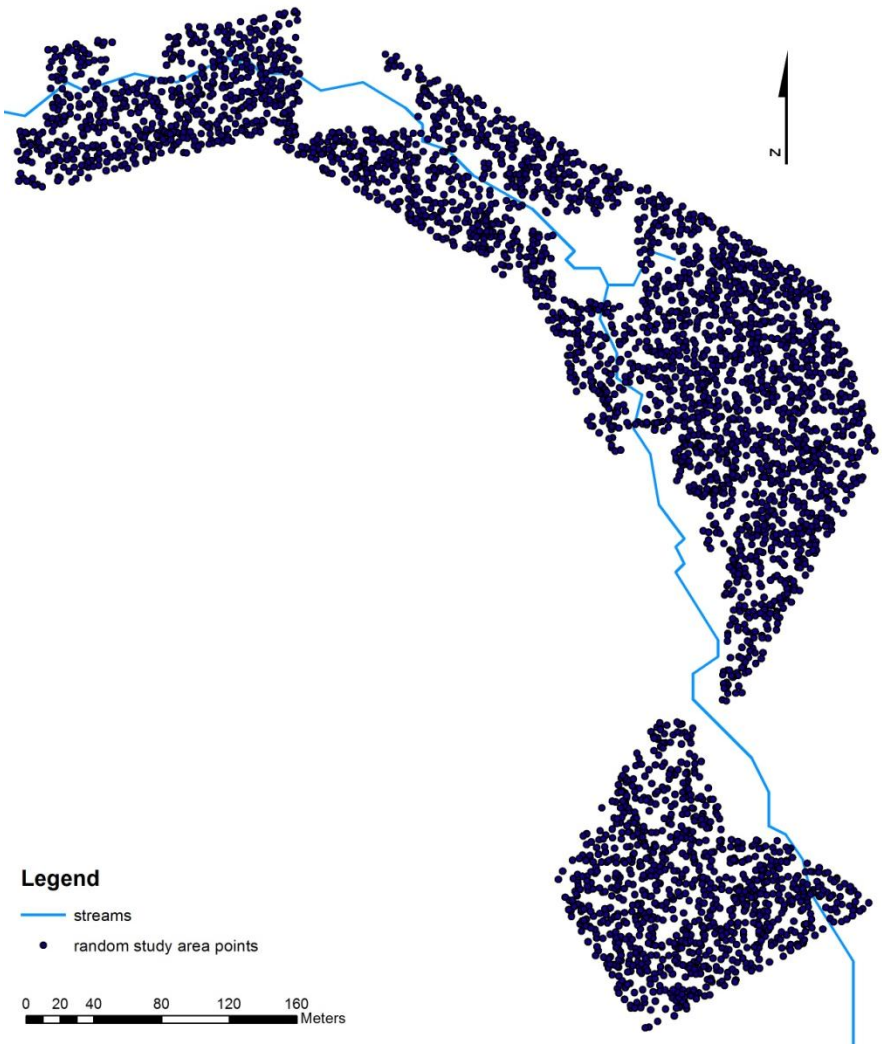

Figure 4.2.1. Random points generated for the study area. Each point was given a TWI, DWI and elevation value extracted from each resolution of its respective raster grid (total of 15 values per point). These were used for cokriging with soil attributes to improve predictions.

4.2.3 Statistical tools

All analyses were performed in ArcGIS 10 (ESRI Inc.). Tools referenced in parentheses were extensions available within ArcGIS 10.

4.3 Results and discussion

To save space in this section, N (0-7.5cm) was used as an example of the results. All soil attributes shared similar types of patterns (e.g. clustering and variable trends), though the exact nature of those patterns varied depending on the attribute under examination. Appendix B contains the results of the tests and associated figures for all of the soil attributes. Relevant comparisons among soil and terrain attributes are represented in tables below.

The global Moran's I test showed that all soil attributes were significantly spatially autocorrelated ($p < 0.0001$). The local Moran's I test (Figure 4.3.1) showed that N (0-7.5cm) was clustered with high values (red) near the stream north of Old House Rd, and low values (blue) in the S grid. The trend analysis (Figure 4.3.2) showed that in one direction the N values followed a linear trend, while in another direction they followed a quadratic polynomial trend. The Voronoi diagram (Figure 4.3.3) showed that in general the N values seemed to be higher near the stream than they were uphill from the stream. The stream was included in the figure to provide context for the values. The frequency distribution (Figure 4.3.4) of N values indicated that the data should be log-transformed before kriging.

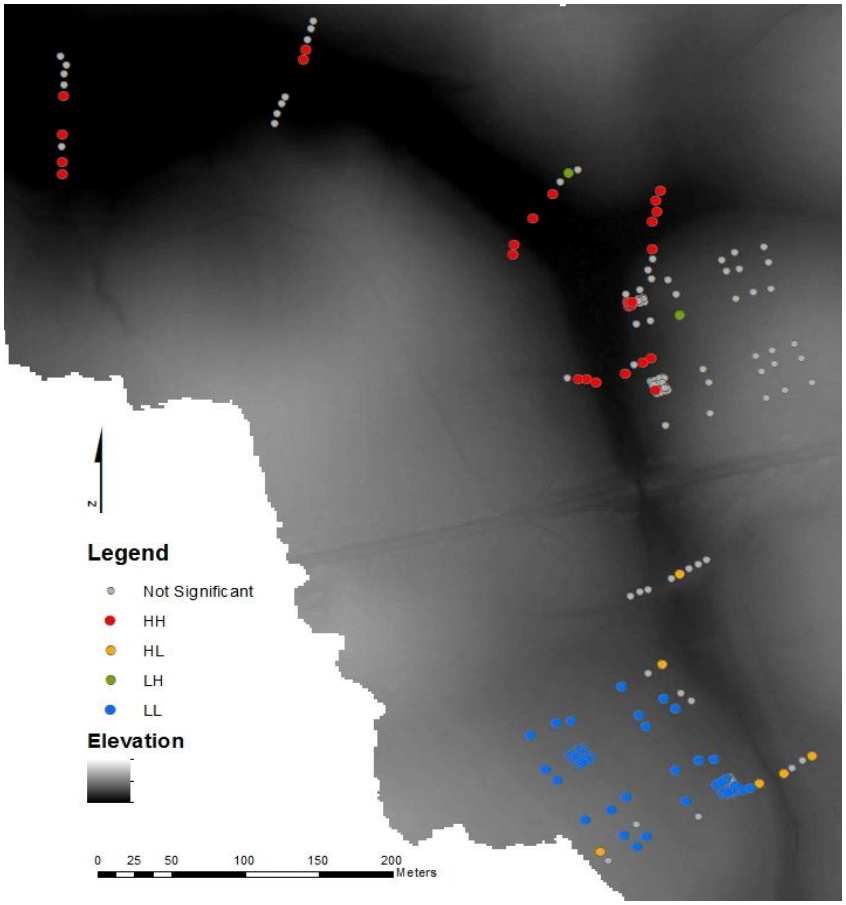

Figure 4.3.1. Local Moran's I map, showing significant ($p<0.05$) clustering of high and low values of N. HH and LL represent high and low clustering, respectively. HL and LH represent high values surrounded by low values and low values surrounded by high values, respectively.

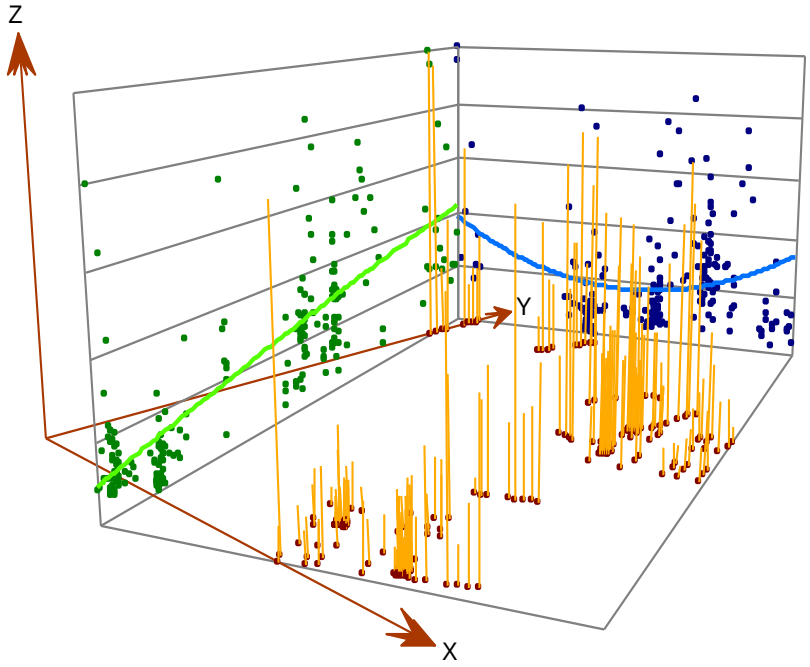

Figure 4.3.2. Trend analysis of N (0-7.5cm). The X-axis runs west-east and the y axis runs south-north. The Z axis is the values of the variables. Note that the trend that runs approximately NNW is close to linear, while the trend that runs ENE is a 2nd order polynomial

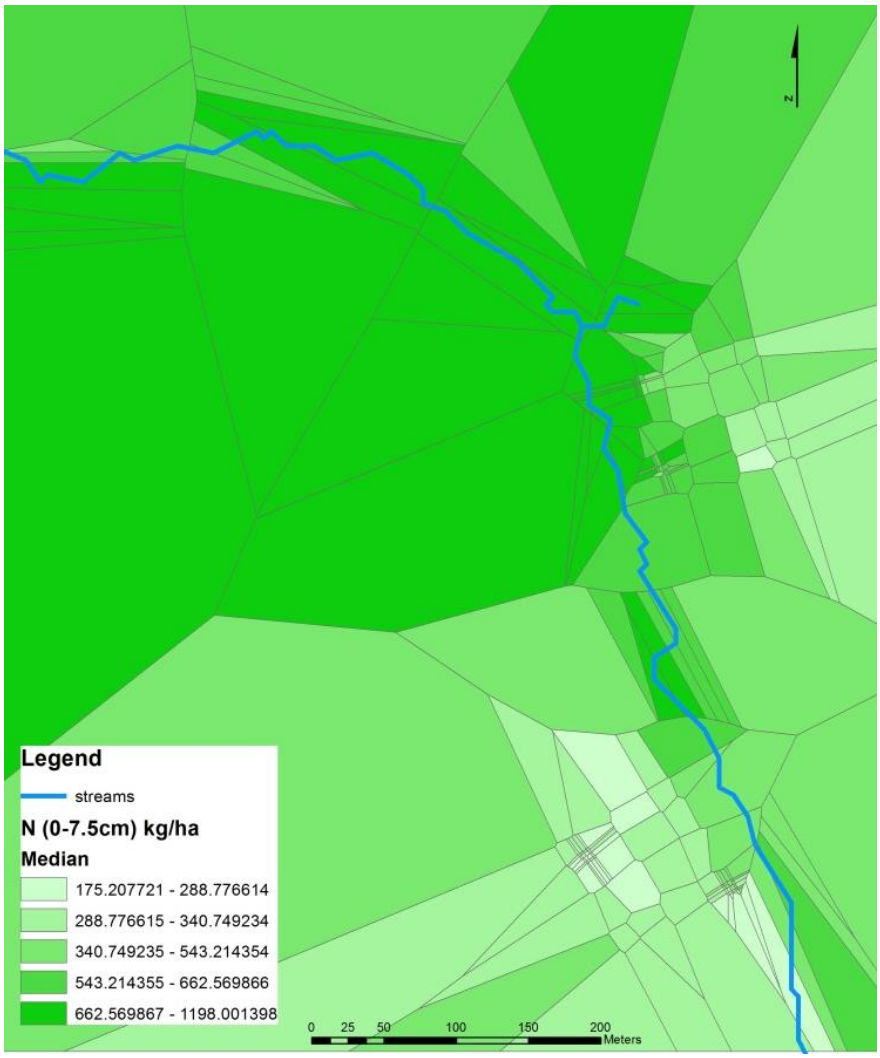

Figure 4.3.3. Voronoi map of N (0-7.5cm) values on the C basin. The high values appear to be associated with the ephemeral stream.

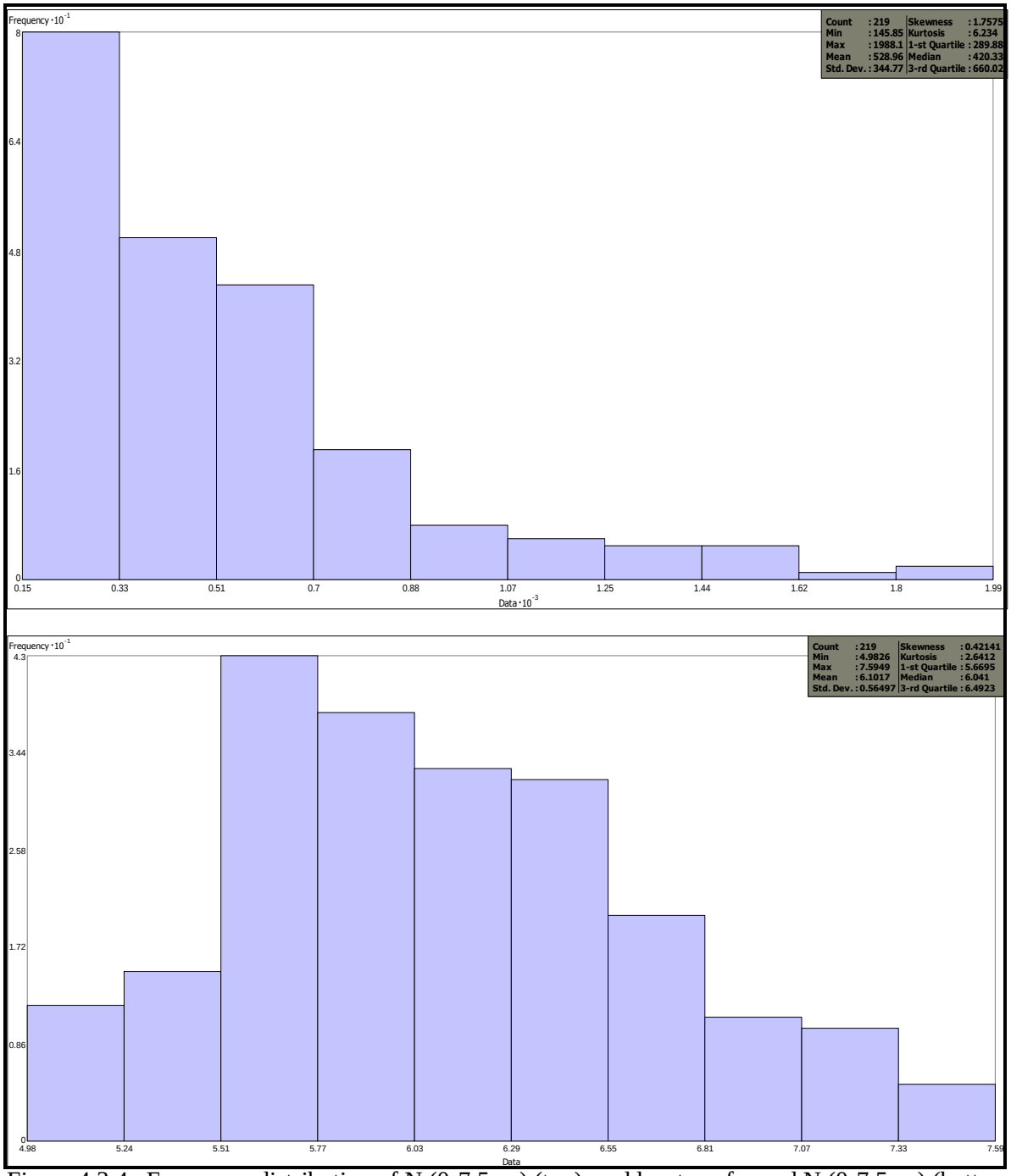

Figure 4.3.4. Frequency distribution of N (0-7.5cm) (top), and log-transformed N (0-7.5cm) (bottom).

After analyzing the semivariogram (Figure 4.3.5), anisotropy in the SW-NE direction was found. The root mean square error (RMSE) for N (0-7.5cm) was 257.01 kg/ha, and the range of values was 145.9 – 1988.1 kg/ha. Comparisons between prediction error results from kriging of N alone and cokriging of N with TWI, DWI and elevation are shown in Table 4.3.1. The 30m resolution showed the best improvement in RMSE for each wetness index among the other resolutions for those indices. Cokriging with elevation actually raised the RMSE score, so was not a good choice for cokriging with N.

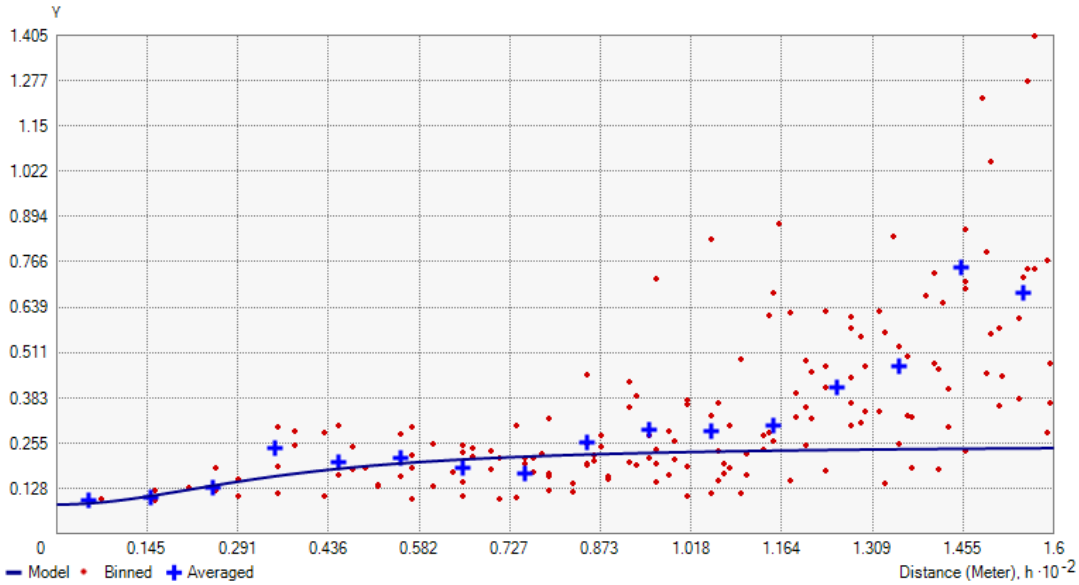

Figure 4.3.5. Semivariogram for N (0-7.5cm).

Similarly, the 30m resolution was also the best choice for cokriging with C (0-7.5cm) with both TWI and DWI. Again, elevation was not a good choice. Table 4.3.2 shows the resolutions for each wetness index that made the best improvements for each soil attribute RMSE, and Figure 4.3.6 illustrates the spatial differences among cokriged C and N in both the organic and mineral horizons. The range of values for each attribute was included to provide scale for the RMSE. The 5m resolution TWI and DWI improved the organic horizon N and C predictions the best relative to other resolutions. This was similar to Chapter 3, in terms of the finer resolutions better predicting the organic horizon C and N, and the coarser resolutions better predicting the mineral soil C and N. One reason, as posed in Chapter 3, might

be because the scale of processes affecting the spatial variation in the organic horizon is smaller than the scale of the processes affecting the spatial variation of the mineral horizon. Figure 4.3.7 illustrates the spatial differences between the organic and mineral horizon N. The smaller "islands" of values in the organic horizon map indicate more variation over a smaller area.

Assuming correlation with the wetness indices, these values should be correlated better given more variation of wetness index over the same area, which was what happened when a 5m wetness index was used instead of a 30m wetness index.

Bulk density was not improved by any cokriging except with 10m DWI, and it only improved from a RMSE of 0.1374 to 0.1368.

Table 4.3.1. Prediction error for kriged maps with N alone, and N cokriged with each resolution of index and elevation.

N (0-7.5cm)	RMSE
Alone	257.01
2m TWI	257.60
5m TWI	256.01
10m TWI	257.55
30m TWI	255.02
50m TWI	256.46
2m DWI	257.83
5m DWI	256.56
10m DWI	257.50
30m DWI	253.50
50m DWI	256.44
2m elevation	263.28
5m elevation	263.15
10m elevation	262.43
30m elevation	260.55
50m elevation	265.12

Table 4.3.2. Prediction error improvement. The "best index" is the index that best improved the RMSE from that of the attribute alone. The 5th column is the RMSE with EC added as a covariate to the "best index" model. The range of values was provided for scale of error.

Attribute	RMSE alone	Best index	RMSE w/ index	RMSE w/ index and EC	Range
N (0-7.5cm)	257.01	30m DWI	253.50	253.42	145.9 - 1988.1
C (0-7.5cm)	6773.48	30m DWI	6703.60	6733.93	5762.8 - 65276.0
N (organic)	111.60	5m TWI	110.55	109.37	28.9 - 804.4
C (organic)	3949.35	5m TWI	3886.25	3895.53	1983.9 - 23518.0
Bulk density	0.1374	10m DWI	0.1368	0.1372	0.2258 - 1.5690

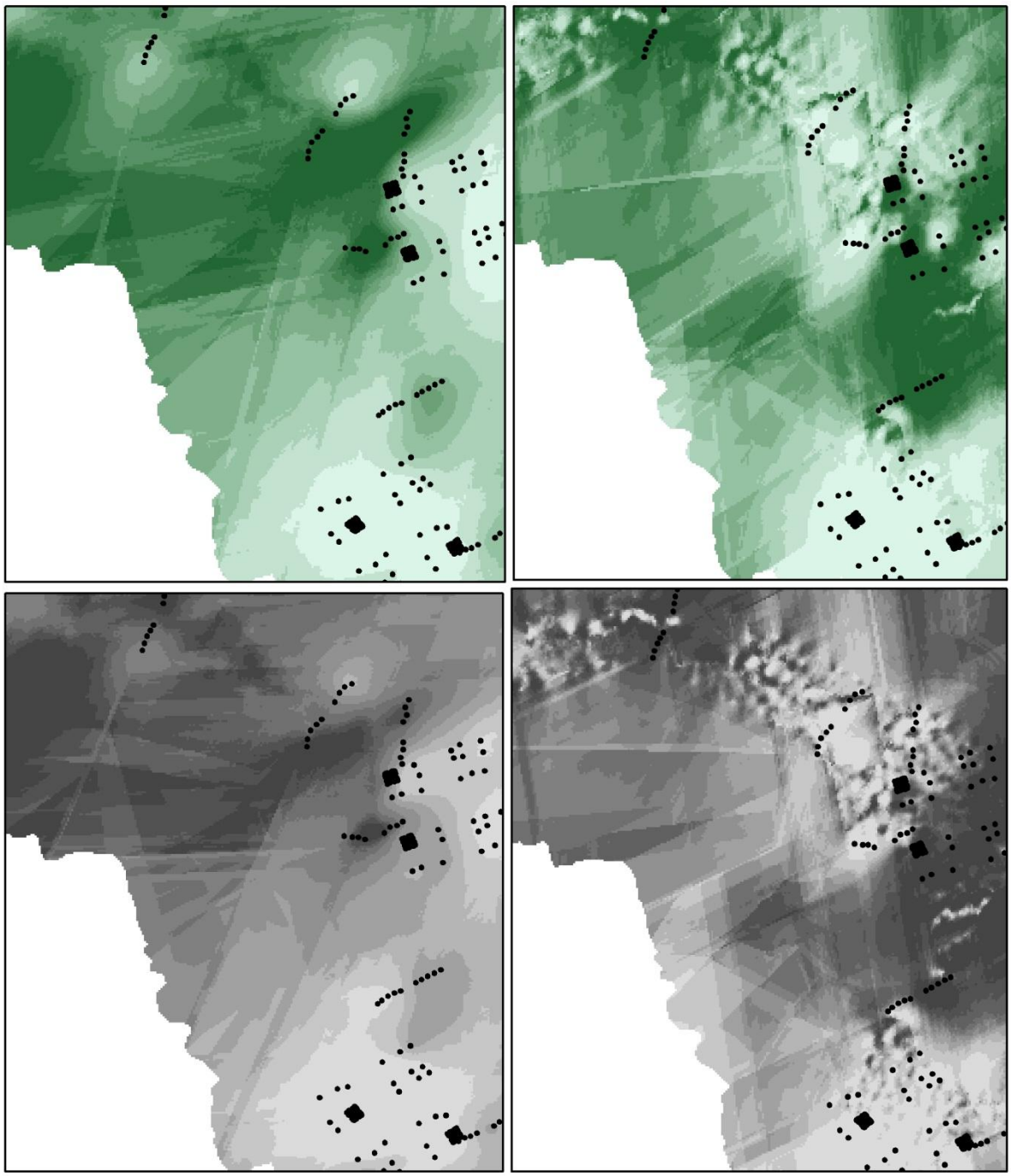

Figure 4.3.6. N (0-7.5cm) cokriged with 30m DWI (top left), N (organic horizon) cokriged with 5m TWI (top right), C (0-7.5cm) cokriged with 30m DWI (bottom left), and C (organic horizon) cokriged with 5m TWI (bottom right).

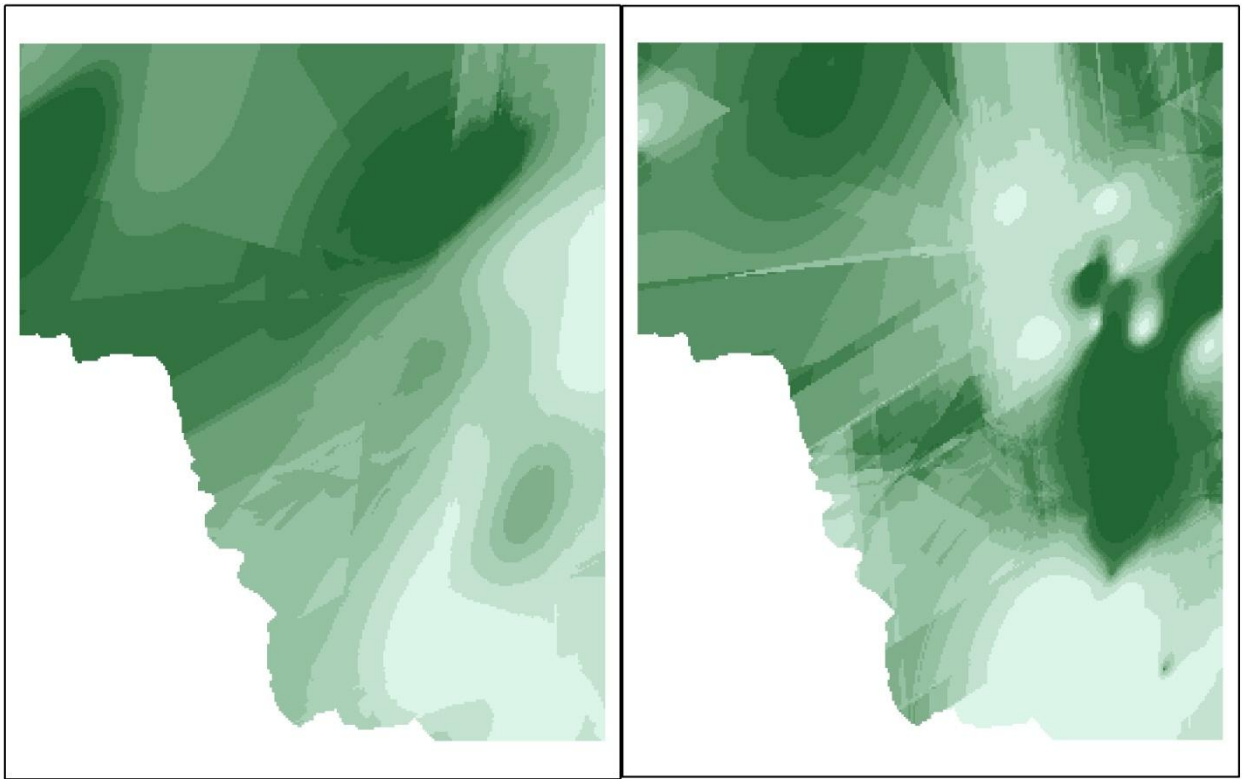

Figure 4.3.7. N (0-7.5cm) (left) and N (organic) (right). The smaller "islands" suggest smaller scale differences among values in the organic horizon map.

Soil electrical conductivity from March 2011 was also added as a third covariate in the cokriging models, for two reasons. First, the kriged map of EC values (Figure 4.3.8) appeared to resemble the study area in terms of stream definition. Second, EMI is a relatively inexpensive sampling method, and information about how it works to improve sampling estimates would be valuable. Only the attribute-index combinations from Table 4.3.2 were used. While all cokriging with the added EC parameter improved the prediction error relative to the attributes alone, it only improved the error relative to that from cokriging with the best index in mineral and organic horizon N (Table 4.3.2). The resulting maps were compared next to the maps with only the attribute and those cokriged with an index (Figure 4.3.9).

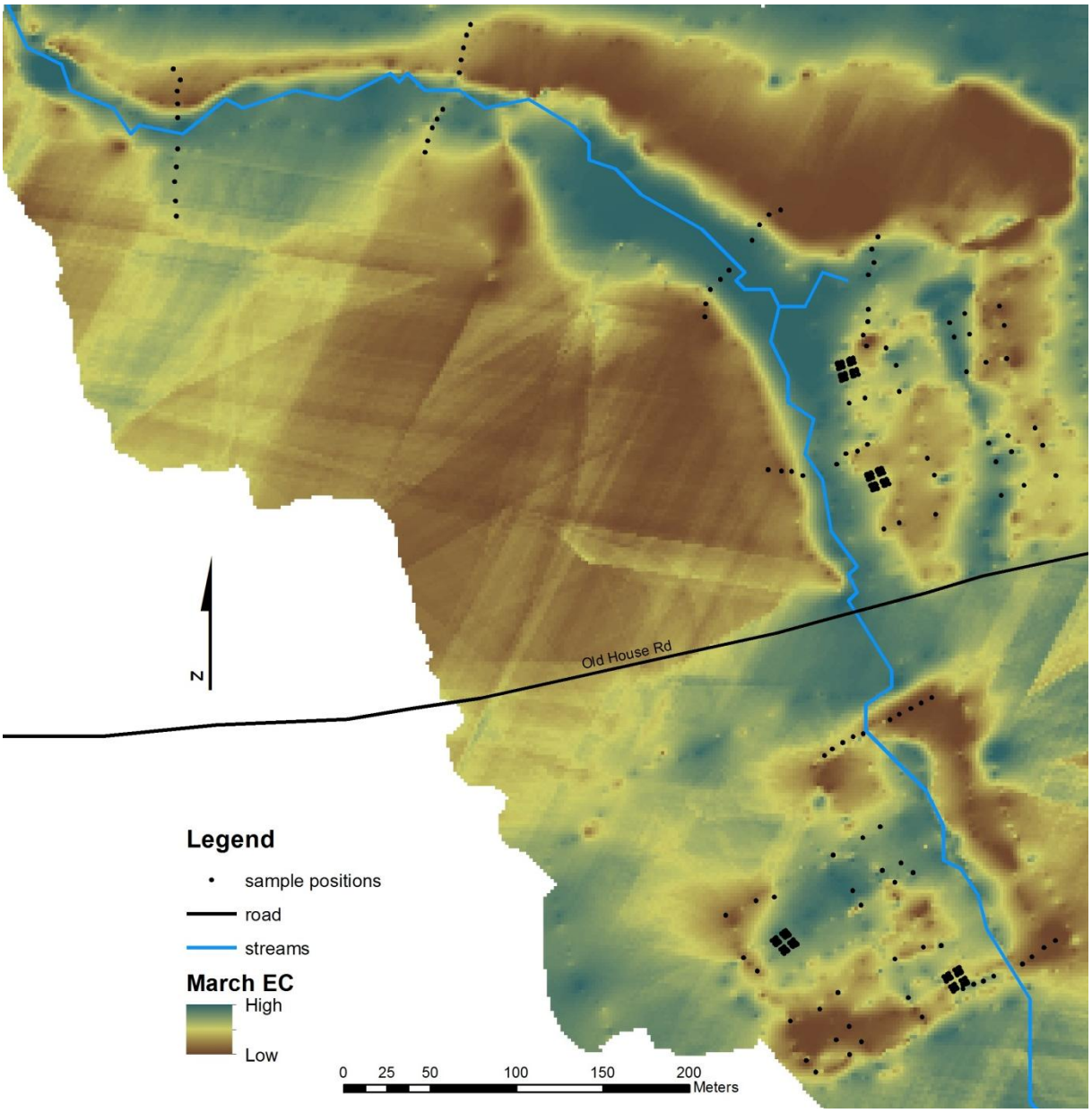

Figure 4.3.8. Kriged prediction map of EC values across the study area. The stream and road were added for context. The greatest density of sampling was done near the transects and grids, so the accuracy of this map decreases with distance from these features.

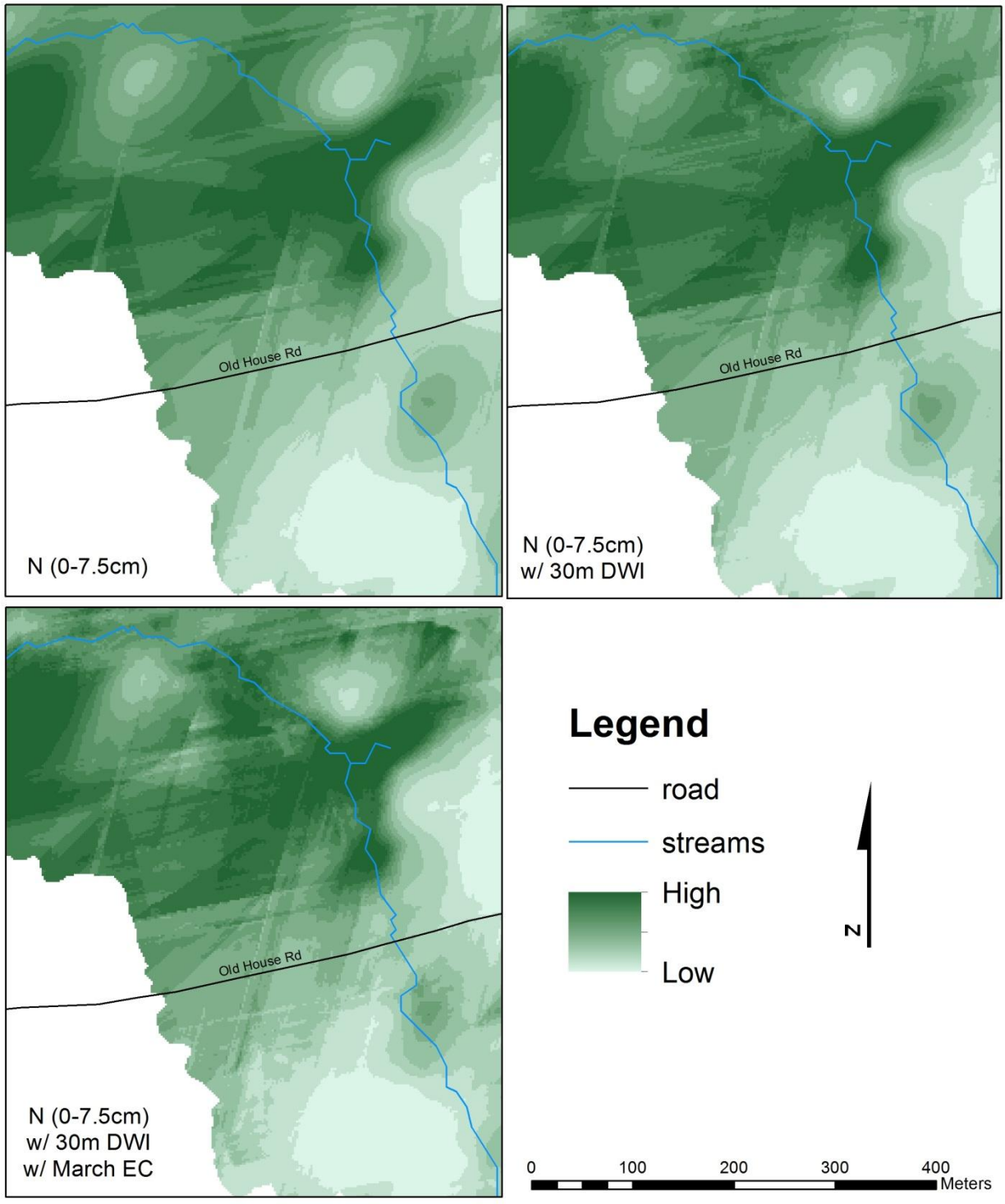

Figure 4.3.9. Spatial distribution of N (0-7.5cm) values alone (top left), cokriged with 30m DWI (top right), and cokriged with 30m DWI and March EC (bottom left).

4.4 Conclusion

The spatial analyses shown here indicate that the choice of DEM resolution should consider the scale effects of processes on the variable being measured. In this case, the organic horizon C and N appeared to vary more over smaller areas than the mineral horizon. The prediction maps produced using finer resolution wetness indices performed better than those from coarser resolution indices. In contrast, all of the attributes from the mineral soil (0-7.5cm depth) had lower prediction errors from coarser resolution indices. These results agreed with the results from Chapter 3.

4.5 Literature cited

Diggle P.J. and P.J. Ribeiro. *Model-based Geostatistics*. New York: Springer; 2007.

ESRI Inc. How kriging works. *ArcGIS 10 Desktop Help*. Accessed 18 October 2011. http://help.arcgis.com/en/arcgisdesktop/10.0/help

ESRI Inc. Modeling spatial relationships. *ArcGIS 10 Desktop Help*. Accessed 19 October 2011. http://help.arcgis.com/en/arcgisdesktop/10.0/help

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

Tobler WR. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography 46*: 234–240.

Wackernagel H. *Multivariate Geostatistics: an Introduction with Applications*. New York: Springer; 2006.

Webster R. and M. Oliver. *Geostatistics for Environmental Scientists*. Chichester, UK: Wiley; 2001.

Chapter 5

CONCLUSION

This study showed that there is no optimal DEM resolution for every application when predicting soil attributes from terrain attributes. The choice depends on the particular combination of terrain and soil attributes under examination. For example, the 5m resolution performed better, both using spatial and non-spatial statistics, when predicting organic horizon C and N, while the 30m resolution performed better when predicting mineral soil C and N. This was probably due to differences in the scale of the processes governing the status of C and N in their respective horizons. A mechanistic study might reveal exactly what are those processes, but that was beyond the scope of this study. It was found that the correlations between certain soil attributes and topographic wetness indices had non-linear trends both spatially and non-spatially. This might serve as a starting point for such investigations into soil chemical and physical processes.

All quantile measures of wetness index were sensitive to DEM resolution. The differences between index values calculated at different DEM resolutions were not trivial. Consider that the index is on a log-scale, so the actual differences were very large. The ranges of the wetness indices decreased with decreased resolution, with the minimums affected the most. What this means is that the wetness index calculated from a 50m DEM will overestimate the saturation of an area relative to a 2m DEM by several orders of magnitude. Or vice versa, the 2m DEM might be vastly underestimating the saturation of that area. The indices should be calibrated to the site in order to decide which resolution is more accurate. The larger grid cell sizes might actually be more accurate if they describe below-surface flow in areas where below-surface flow is dominant. Again, calibration to the particular site would be recommended for future studies and applications.

The basin area was shown in Chapter 2 to be sensitive to DEM resolution, due to the effects on flow direction. This could influence the results of stream flow calculations that take basin area into account. Choosing the correct resolution is here again problematic because the actual flow paths may or may not follow topography.

Using slope as a predictive terrain attribute resulted in better correlations overall than with wetness indices. The slope decreased with decreasing resolution, and those effects were seen as decreased correlations with soil attributes. A decrease in correlation does not imply that the slopes were not correct, it just shows that the use of larger grid cell sizes results in lower predictive ability of slope with certain soil attributes.

Finally, this study produced maps that illustrate the spatial patterns of soil attributes such as C and N in the organic and mineral horizons, bulk density, soil texture, and electrical conductivity for the purpose of informing future research and management applications.

Appendix A

DOWNSLOPE INDEX PROGRAM

This program was created by Dr. John Dowd of the Geology Department at the University of Georgia. It reads elevation information and calculates slope length. The slope length can then be converted to slope by dividing by the elevation drop between the target cell and the cell chosen by this program (i.e. the d criterion). It reads data from either an Excel spreadsheet or an .mtx file. It will automatically convert an Excel into an .mtx file for faster processing. It is assumed the user knows how to extract shapefile or raster information into an Excel spreadsheet prior to use.

Figure A.1 shows a map of the elevation values, with each pixel colored according to elevation value. Using a map of cells as shown in Figure A.2, a specific cell can be targeted and used to test the program. The values for the cell's location and elevation appear in windows above the map.

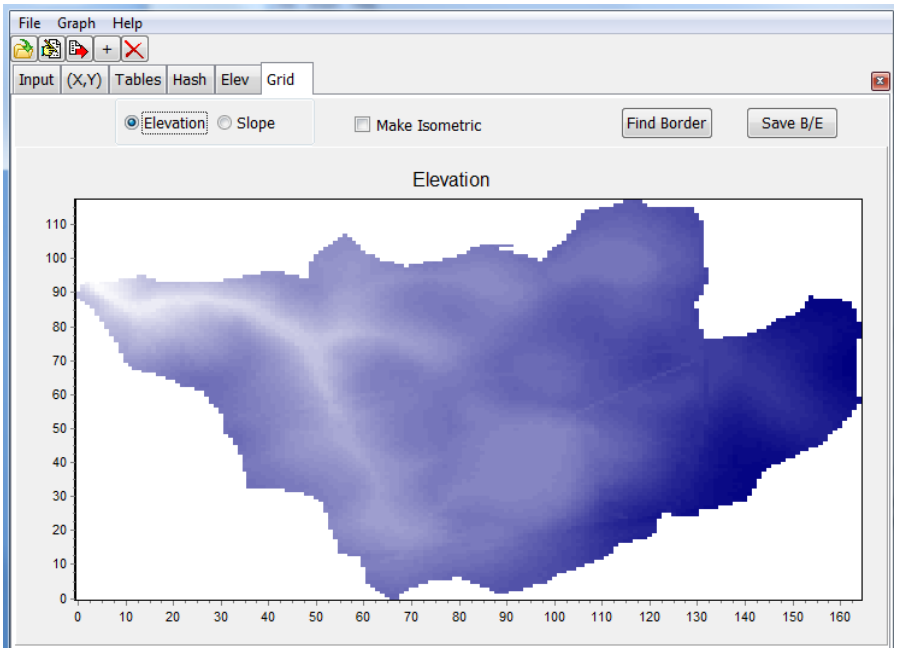

Figure A.1. Graph of elevation values.

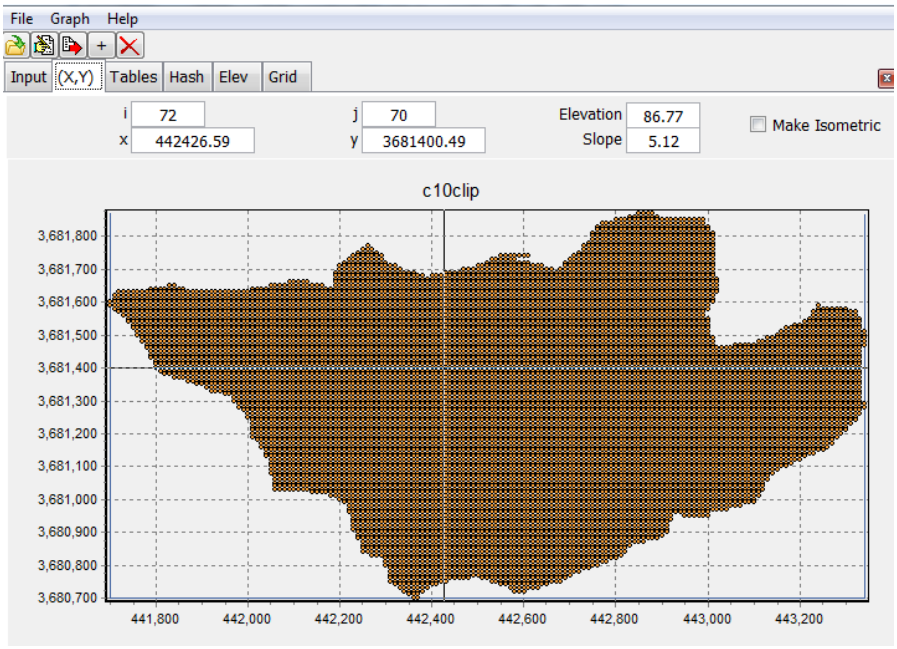

Figure A.2. Using a cursor, specific cells can be selected for testing purposes.

Under the "Hash" tab (Figure A.3) the user can select a certain cell and see either the elevation values for the target cell and all adjacent cells or the difference in elevation between the target cell and the adjacent cells. The adjacent cells are in the first ring of cells (n=1) surrounding the target cell. This is an important concept, as this program uses the number of rings in the calculation of slope length.

The d value is the specified drop in elevation that is used to calculate the downslope index. The program works by checking each adjacent cell to the target cell to see if the "d criterion" is met. If it is met, the program calculates the number of cells multiplied by the cell length. For example, if the cell size is 10m and the d criterion is met within the first ring of cells in a cardinal direction, then the grid length is 10. If the cell chosen is in a diagonal direction the grid length is multiplied by 1.414. This grid length is then adjusted based on the proportion between the actual difference in elevation between the cells and the value of d. For example, the target cell has an elevation value of 85.3m and the cell with the minimum elevation in the first ring is 84.52m (see Figure A.3). The elevation difference is -0.78m (the negative sign is there to signify that it is a drop in elevation). If the value of d is -0.5m, then the program can stop at this ring and calculate the grid length between the two cells (14.14m): Because the actual elevation

drop is more than the required drop, the grid length is adjusted by the proportion 0.5/0.78. This new value is given to the target cell as the adjusted slope length. IT IS NOT THE SLOPE. The d value can then be divided by the slope length to get the actual slope value for that cell. This calculation can be carried out easily using the output from this program in an Excel spreadsheet before converting back to a shapefile.

If no cells in the adjacent ring meet the d criterion, then the program continues to the next ring (n=2), and so on until either the "max n" is reached (Figure A.4) or the d criterion is met. If max n is reached, the cell is given an arbitrary "flag" value (Figure A.5) to show that it never met the d criterion. This max n is adjustable, and is in place for cases where the d criterion will not be met in a realistic fashion, such as from the bottom of a sink hole. It also makes the program run faster.

The user has the option of either running the program for one target cell or for the entire grid. The "Run 1" option is used for testing the program, and will give information (if "Verbose" is checked) on which cell it chose, in which ring, the elevation, grid length, and adjusted length. The user can then manually check using the grid shown in Figure A.5. If "Multiple n" is selected, the program runs as described in the previous paragraphs. If it is not selected, the program runs only for the ring specified (as in Figure A.6).

Figure A.7 shows the output of the slope length values, and Figure A.8 shows the output in a spreadsheet that can be copied and pasted into an Excel spreadsheet and converted back into a shapefile.

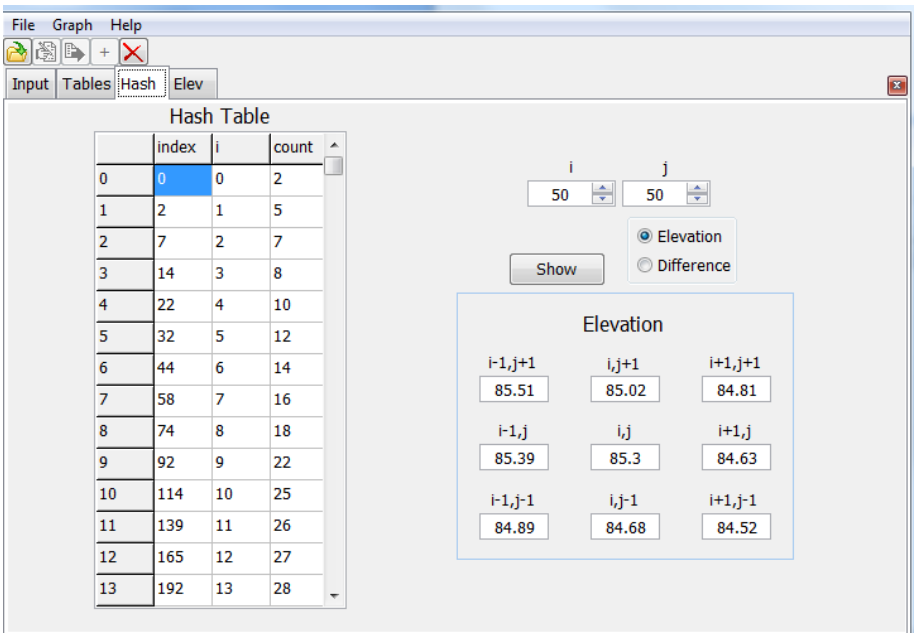

Figure A.3. Option to show the elevation values for a target cell (i,j) and adjacent cells (n=1 ring), or the difference in elevation between the target cell and adjacent cells.

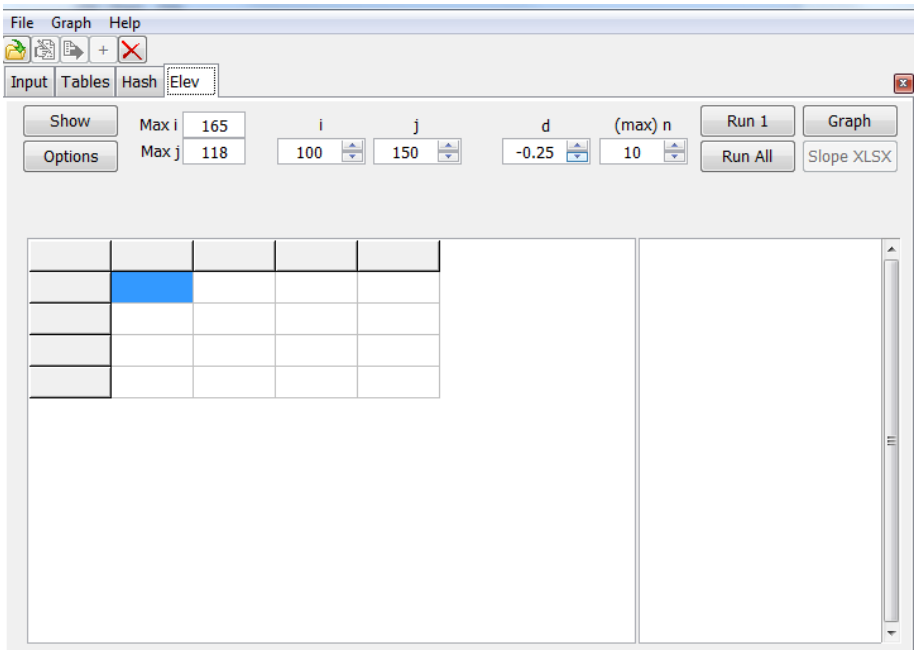

Figure A.4. Options for how to calculate the downslope length. "Run 1" only runs the program for the cell at the i and j location selected in the boxes. The d value is the elevation drop that the program uses to decide when to stop and measure length. The max n is the maximum number of rings the program will search before it stops. "Run all" runs the program for every cell in the grid.

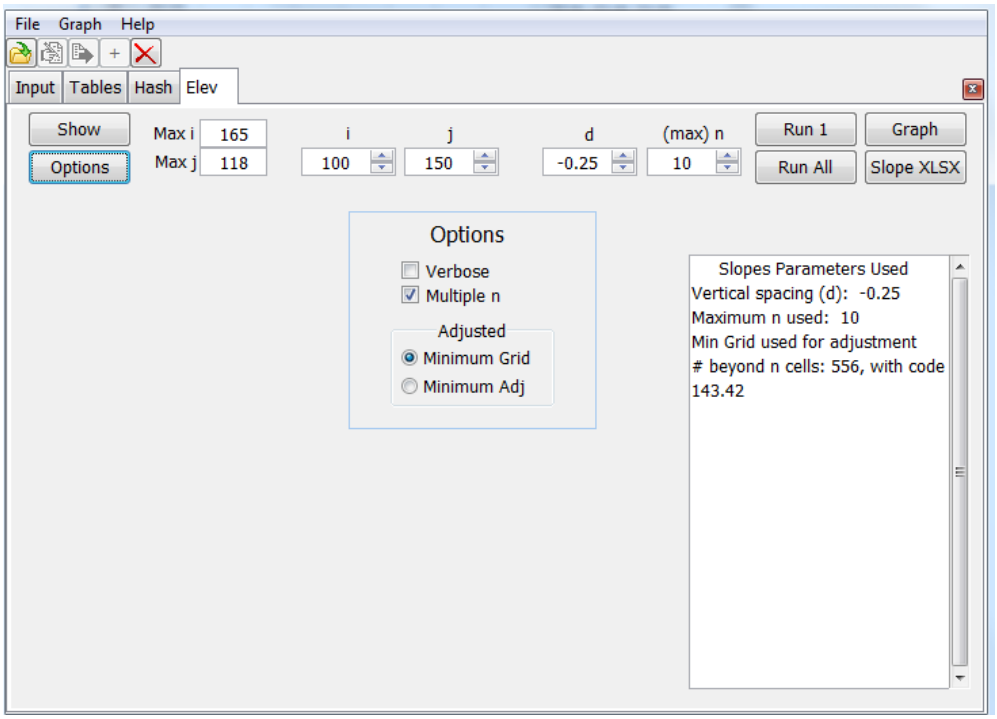

Figure A.5. By running with shown option settings, the program runs for all rings until either the d criterion is met or the max n is reached. Instead of all the information shown when "Verbose" is selected, this display shows the number of cells beyond max n and the slope length value they are given.

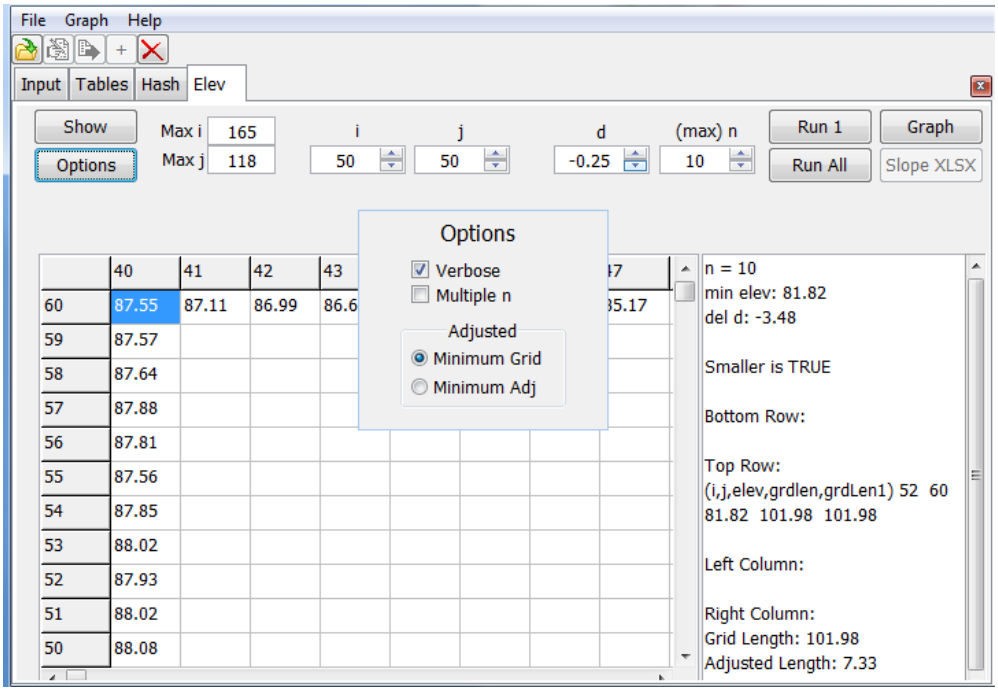

Figure A.6. By running with shown option settings, the program only runs for the ring specified, and shows all information regarding the cell with the minimum elevation found, and the adjusted length to that cell.

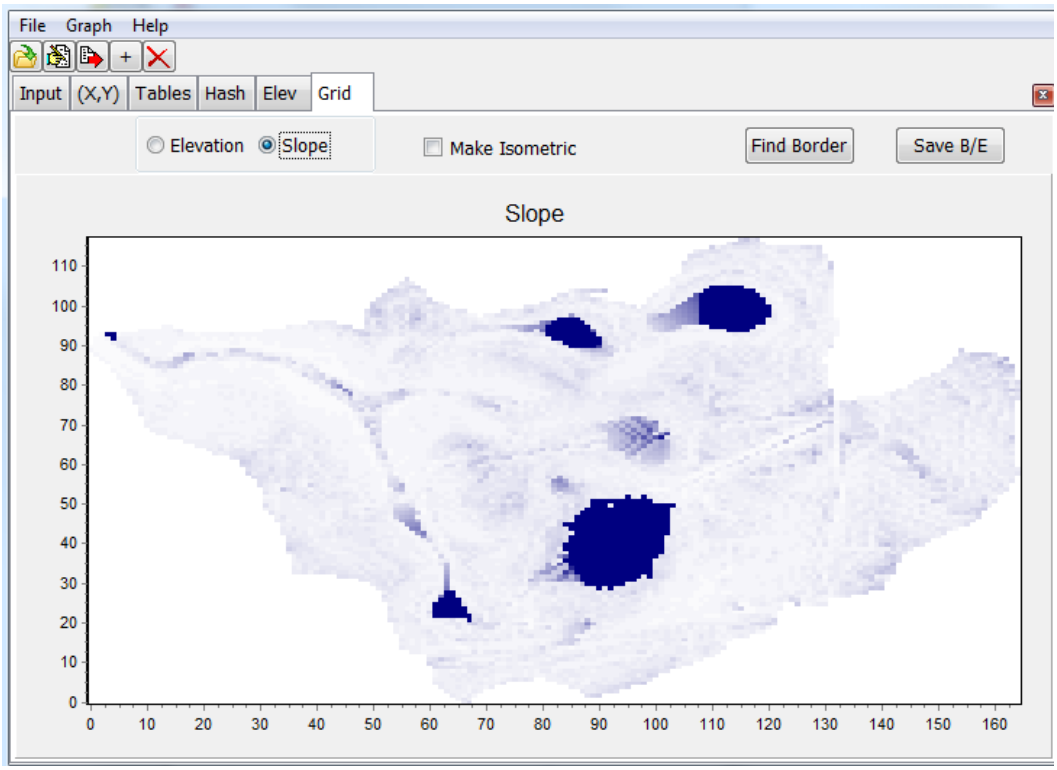

Figure A.7. Graph of slope length values.

File Edit Options Chart

11167 x 4 Real

	0	1	2	3
0	442851.0000	681870.0000	92.6100	24.6665
1	442861.0000	681870.0000	92.5900	31.1962
2	442871.0000	681870.0000	92.8200	16.0705
3	442881.0000	681870.0000	92.9100	20.7970
4	442831.0000	681860.0000	92.5700	12.6270
5	442841.0000	681860.0000	92.4600	25.2539
6	442851.0000	681860.0000	92.5700	22.0971
7	442861.0000	681860.0000	92.5900	19.9650
8	442871.0000	681860.0000	92.7800	15.0448
9	442881.0000	681860.0000	92.9600	11.7851
10	442891.0000	681860.0000	93.2200	9.3039
11	442801.0000	681850.0000	92.2800	24.3829
12	442811.0000	681850.0000	92.3800	8.0646
13	442821.0000	681850.0000	92.2900	12.1526

Figure A.8. Output from program. Columns 0 and 1 are the original x,y coordinates of each cell, column 2 is the elevation of each cell and column 3 is the calculated slope length.

Appendix B

SPATIAL DATA

B.1 N (0-7.5cm)

n = 219

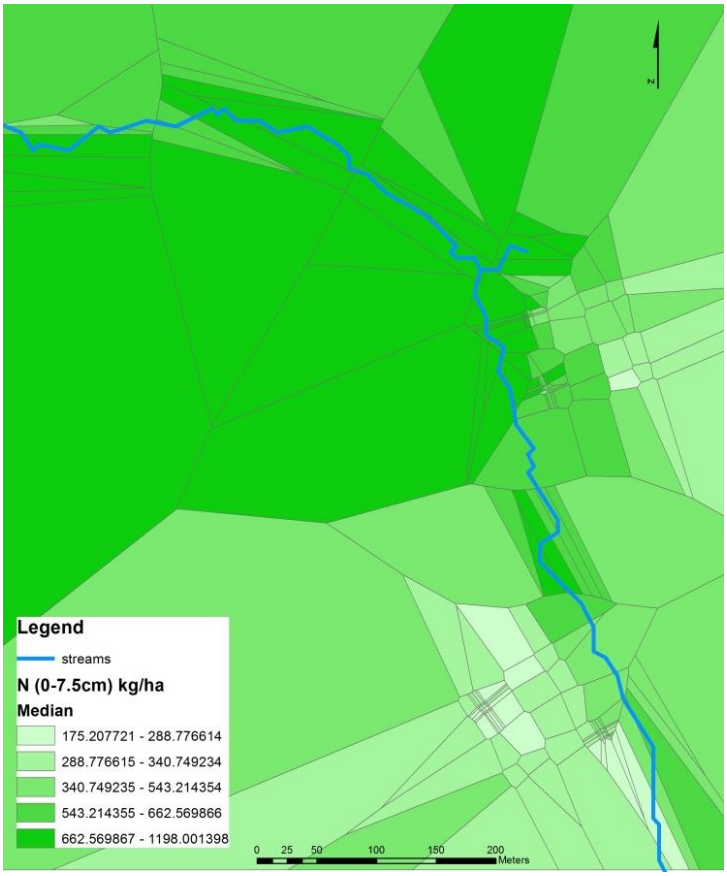

Figure B.1.1. Voronoi map of N (0-7.5cm) values on the C basin. The high values appear to be associated with the ephemeral stream.

Global Moran's I statistics:

Index: 0.353
Z score:15.993
P < 0.0001

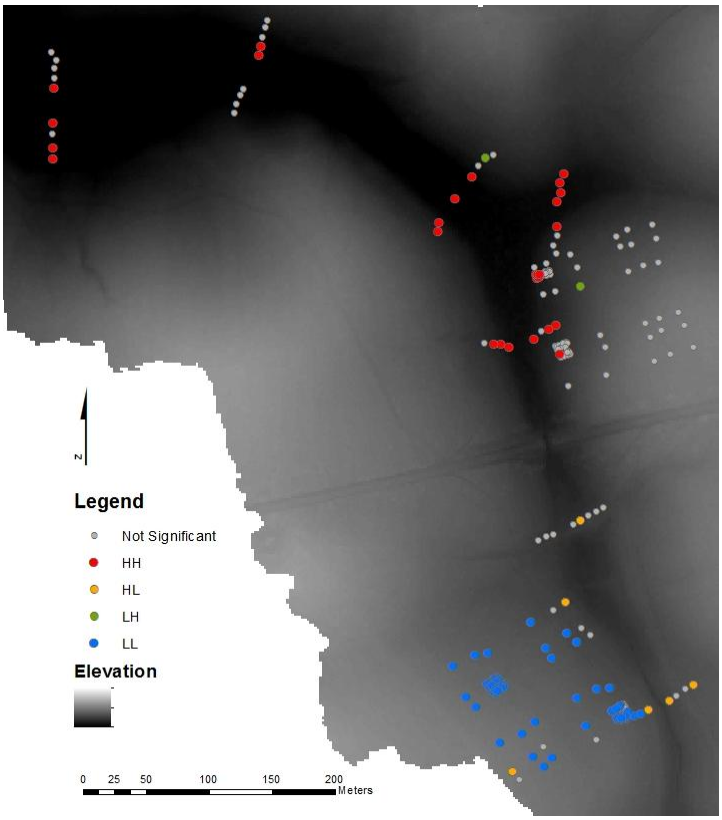

Figure B.1.2. Local Moran's I map, showing significant ($p<0.05$) clustering of high and low values of N. HH and LL stand for high and low clustering, respectively. HL and LH stand for high values surrounded by low values and low values surrounded by high values, respectively.

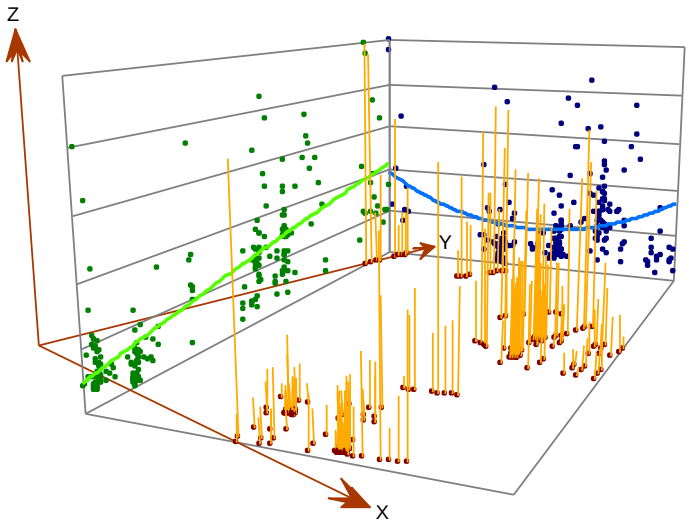

Figure B.1.3. Trend analysis of N (0-7.5cm). The X-axis runs west-east and the y axis runs south-north. Note that the trend that runs approximately NNW is close to linear, while the trend that runs ENE is a 2nd order polynomial.

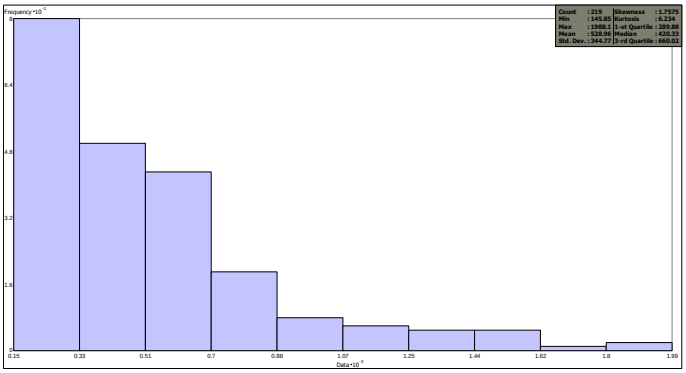

Figure B.1.4. N (0-7.5cm) frequency distribution. The axes and statistics box are illegible, but the point of this figure is the shape of the histogram.

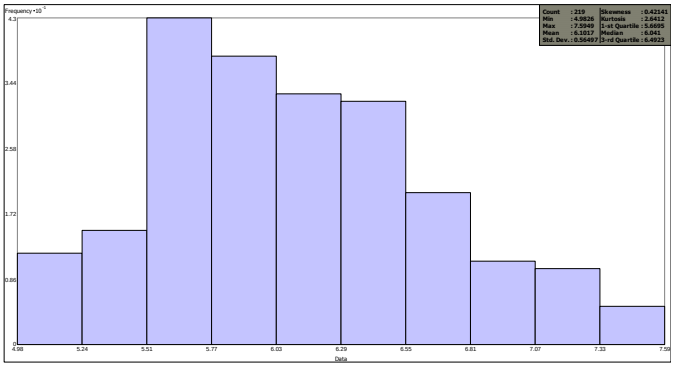

Figure B.1.5. Log-transformed N (0-7.5cm) frequency distribution.

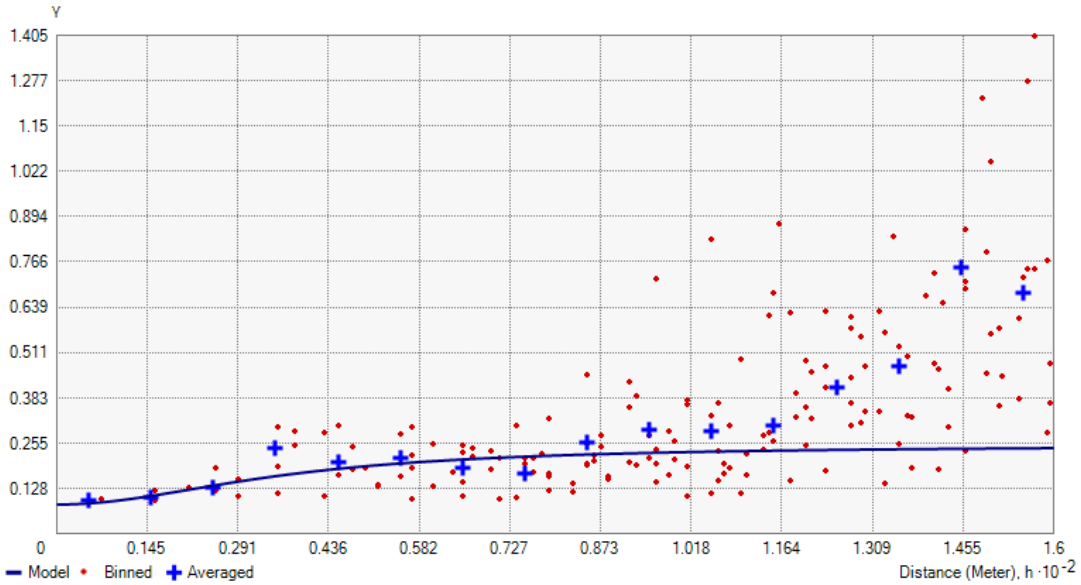

Figure B.1.6. Semivariogram for N (0-7.5cm). The major axis is 160m, minor axis is 100m and the direction is 47°

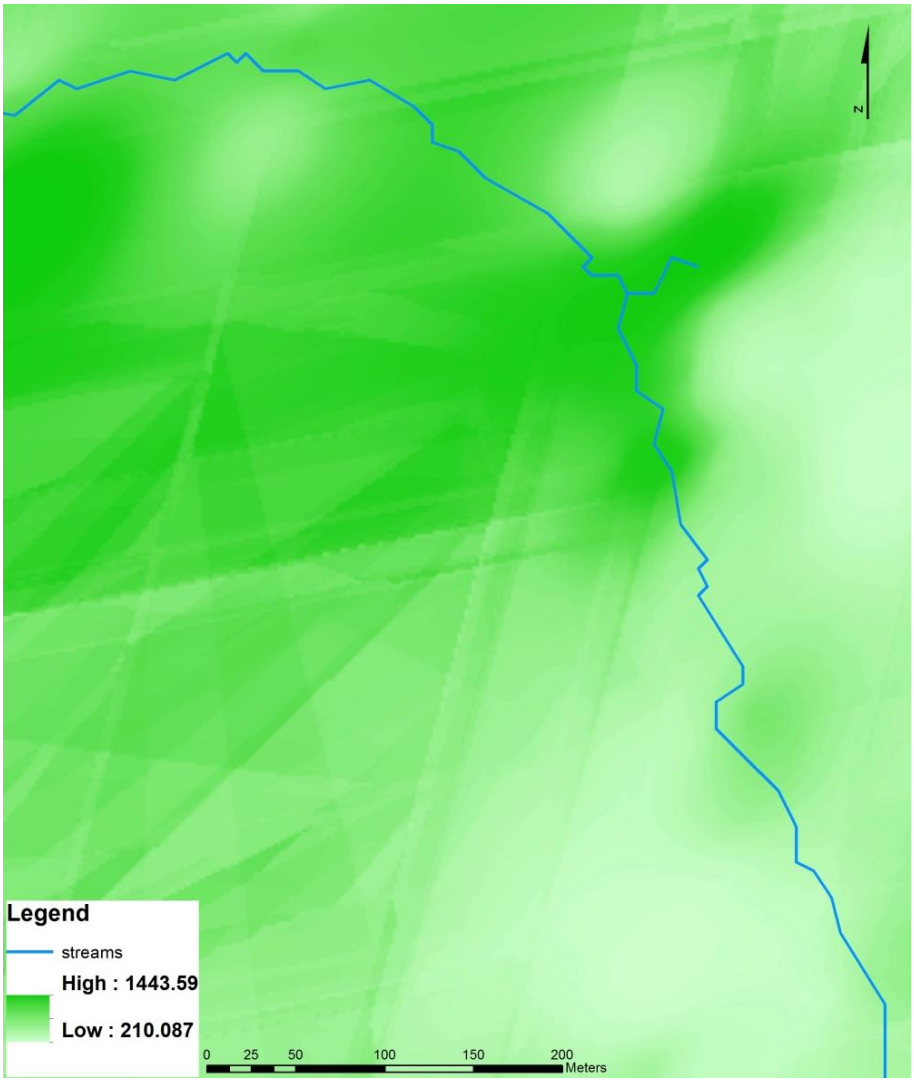

Figure B.1.7. Kriged map of N (0-7.5cm) (kg/ha), using above semivariogram. RMSE = 257.01. Avg. Standard Error = 0.4248.

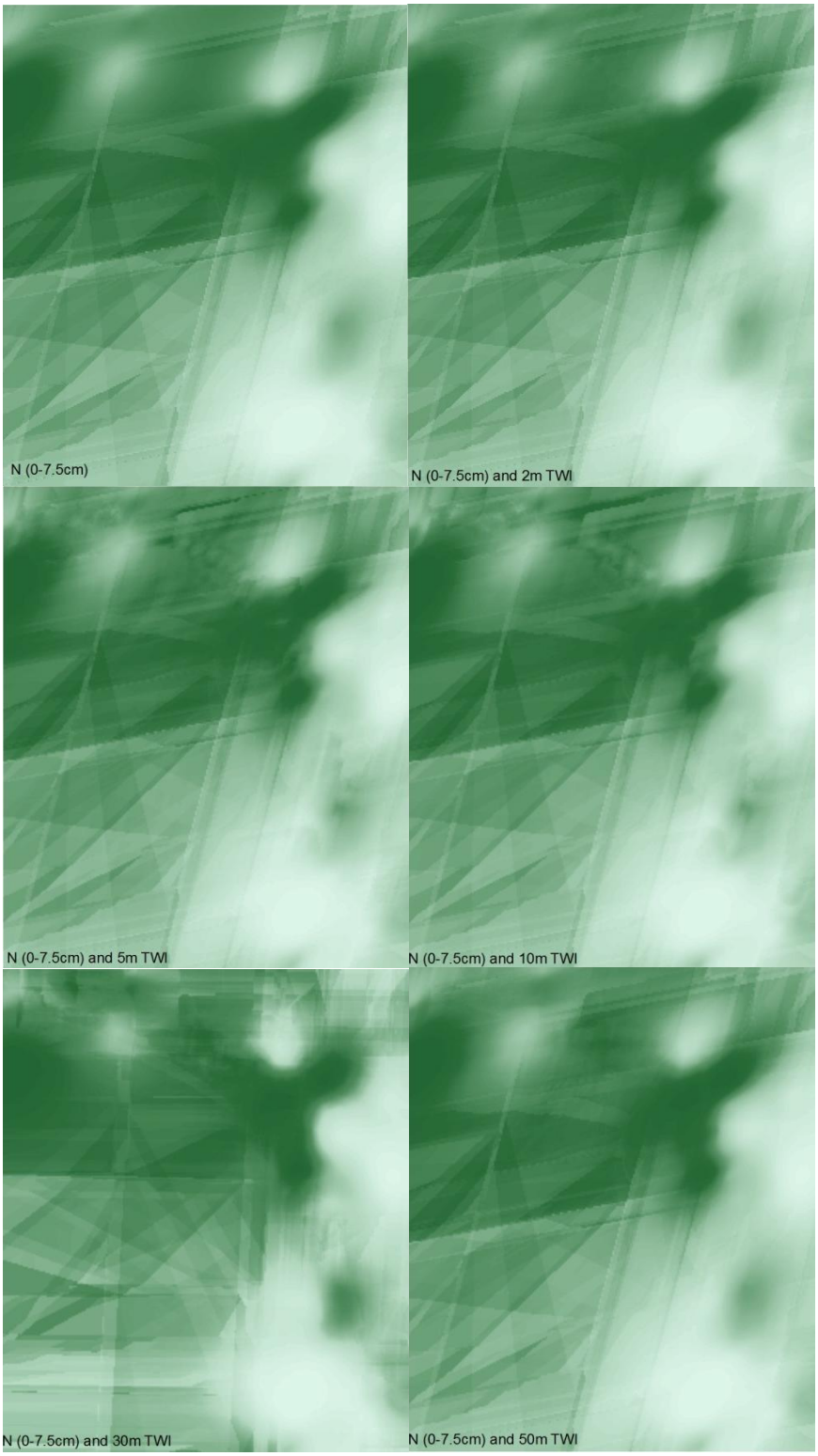

Figure B.1.8. Effects of cokriging with different TWI resolutions on the prediction of N.

B.2 C (0-7.5cm)

n=219

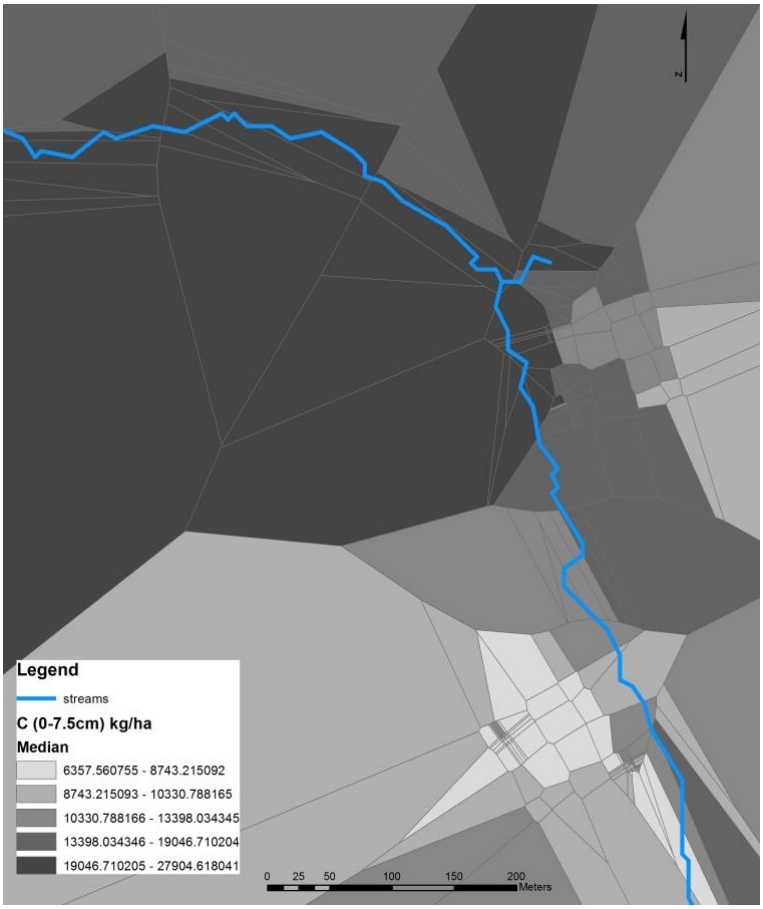

Figure B.2.1. Voronoi map of C (0-7.5cm) values on the C basin. The high values appear to be associated with the ephemeral stream.

Global Moran's I statistics:

Index: 0.289
Z score: 13.234
P < 0.0001

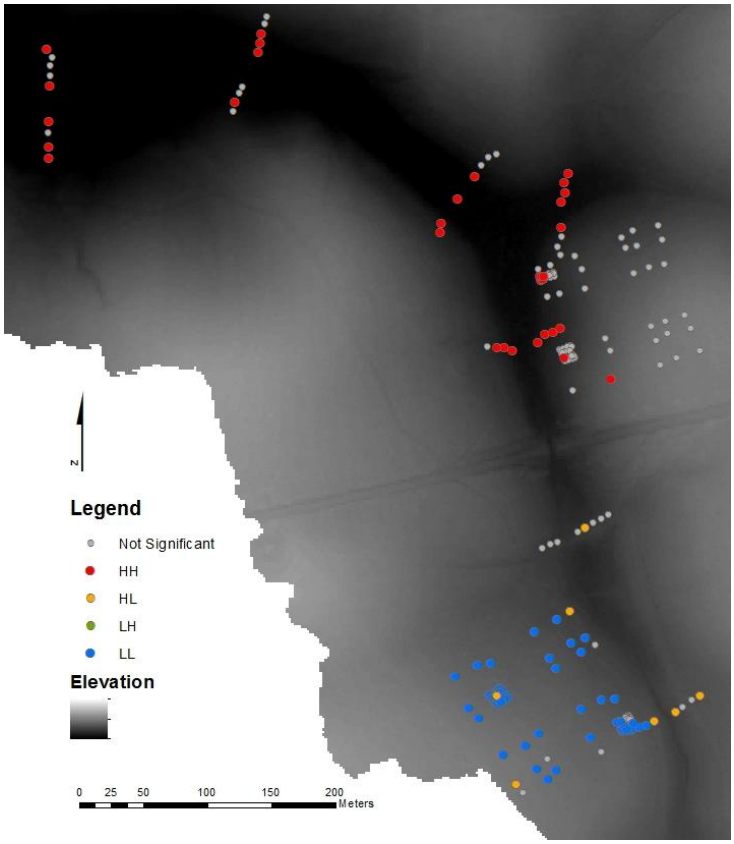

Figure B.2.2. Local Moran's I map, showing significant ($p<0.05$) clustering of high and low values of N. HH and LL stand for high and low clustering, respectively. HL and LH stand for high values surrounded by low values and low values surrounded by high values, respectively.

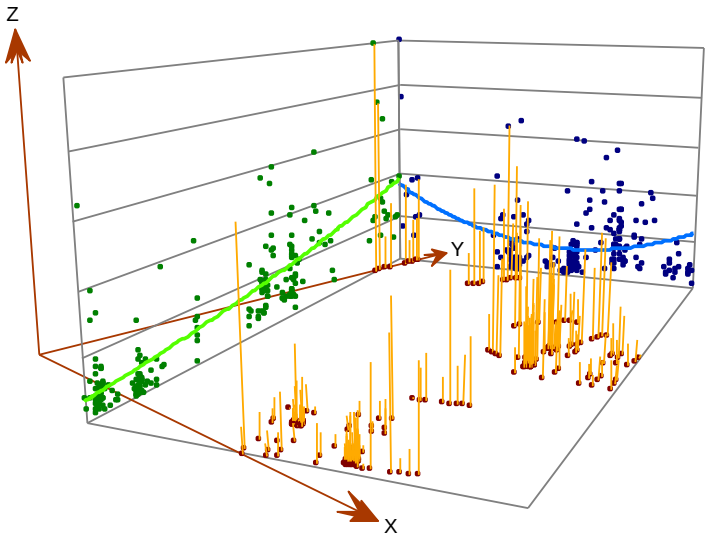

Figure B.2.3. Trend analysis of C (0-7.5cm). The X-axis runs west-east and the y axis runs south-north. Note that the trend that runs approximately NNW is close to linear, while the trend that runs ENE is a 2nd order polynomial.

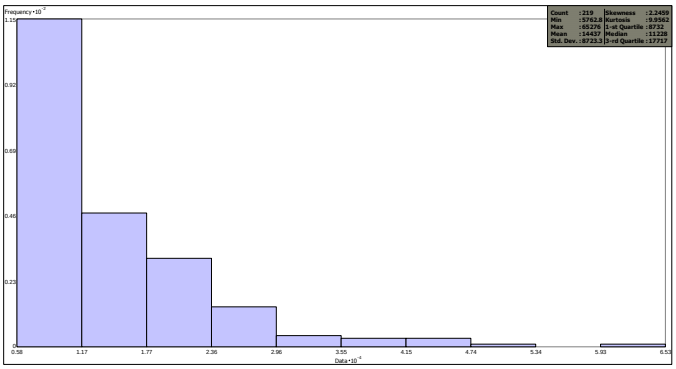

Figure B.1.4. C (0-7.5cm) frequency distribution.

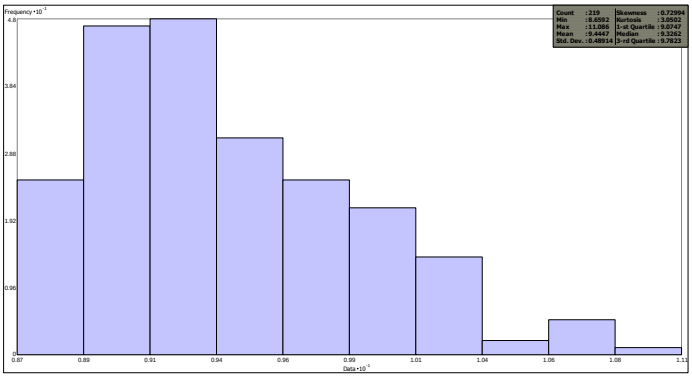

Figure B.2.5. Log-transformed C (0-7.5cm) frequency distribution.

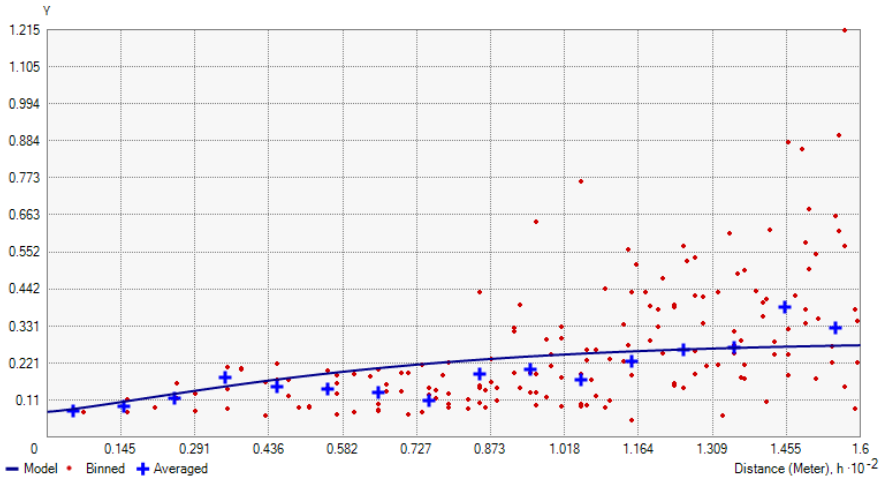

Figure B.2.6. Semivariogram for C (0-7.5cm). The major axis is 160m, minor axis is 100m and the direction is 55.6°.

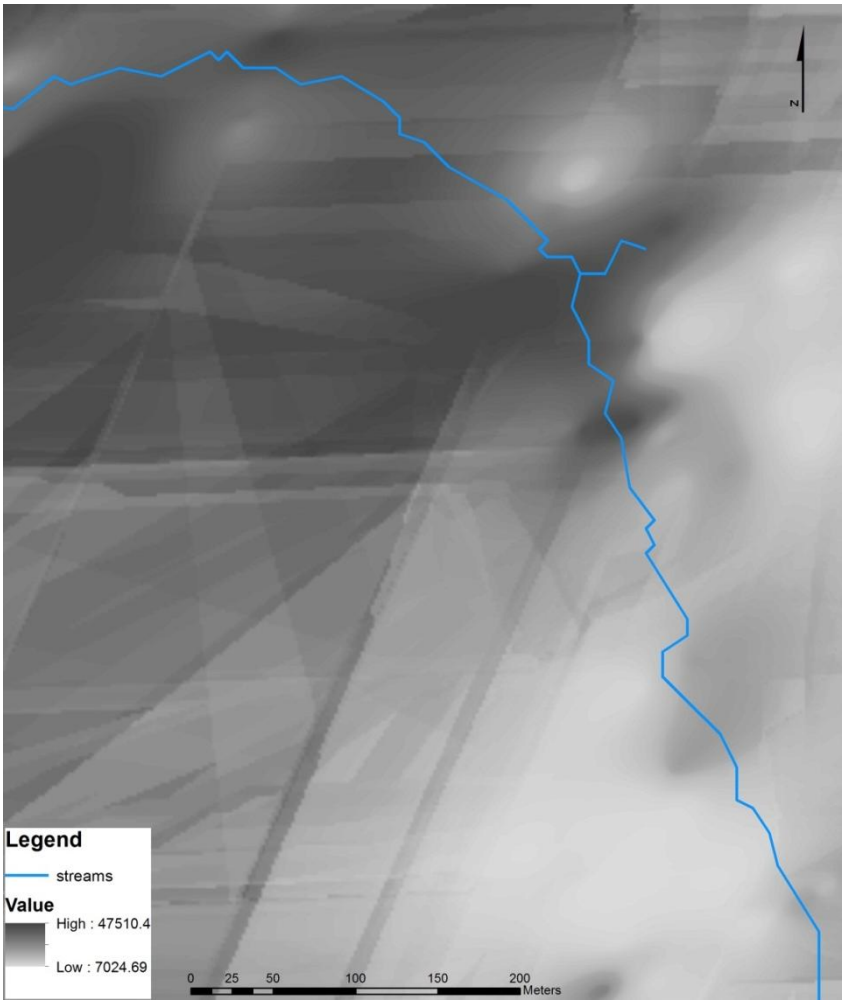

Figure B.2.7. Kriged map of C (0-7.5cm) (kg/ha), using above semivariogram. RMSE = 6773.48. Avg. Standard Error = 0.4246.

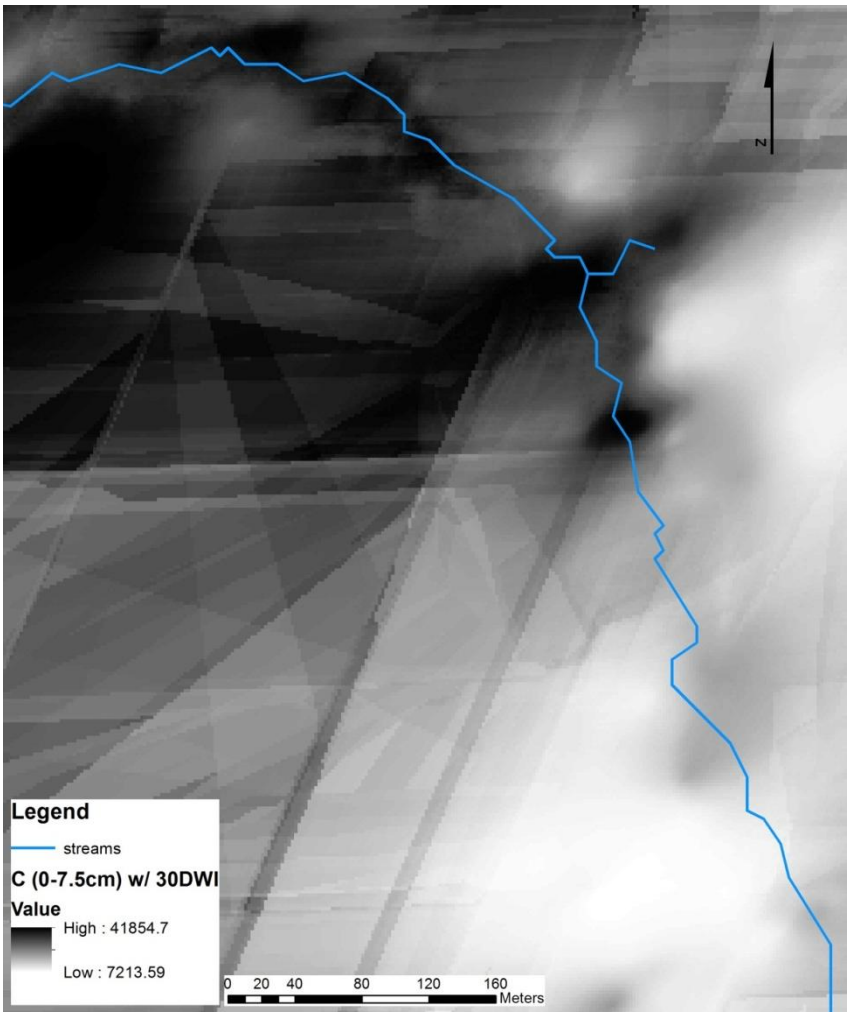

Figure B.2.8. Map of C (0-7.5cm) cokriged with 30m DWI. RMSE = 6703.60. Avg. Standard Error = 0.3509.

B.3 N (organic horizon)

n= 156

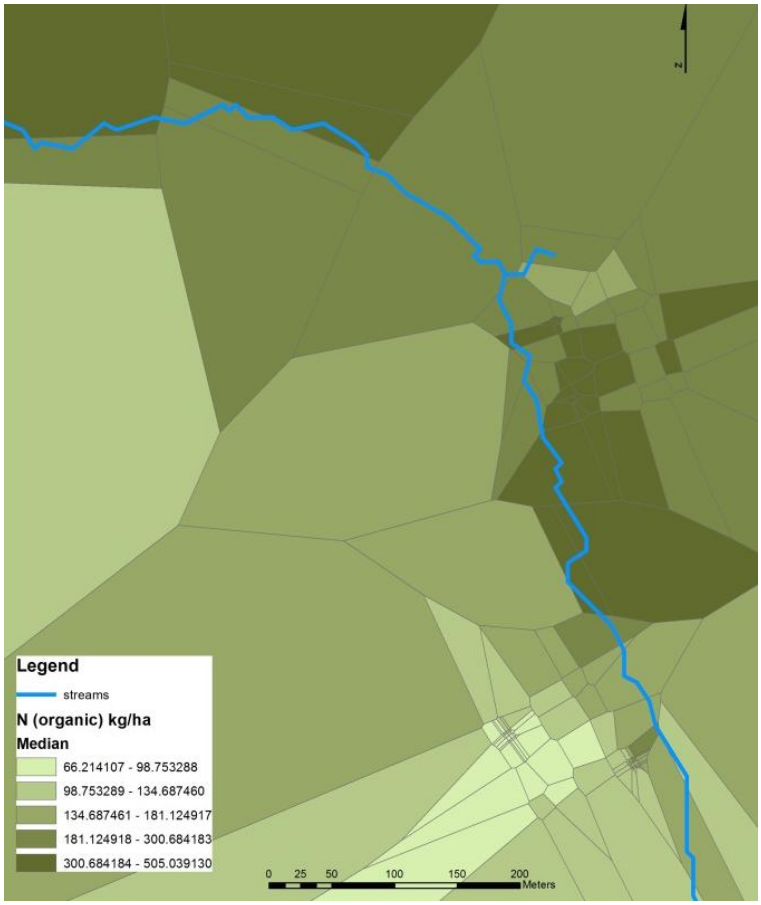

Figure B.3.1. Voronoi map of N (organic) values on the C basin.

Global Moran's I statistics:

Index: 0.407
Z score: 14.883
P < 0.0001

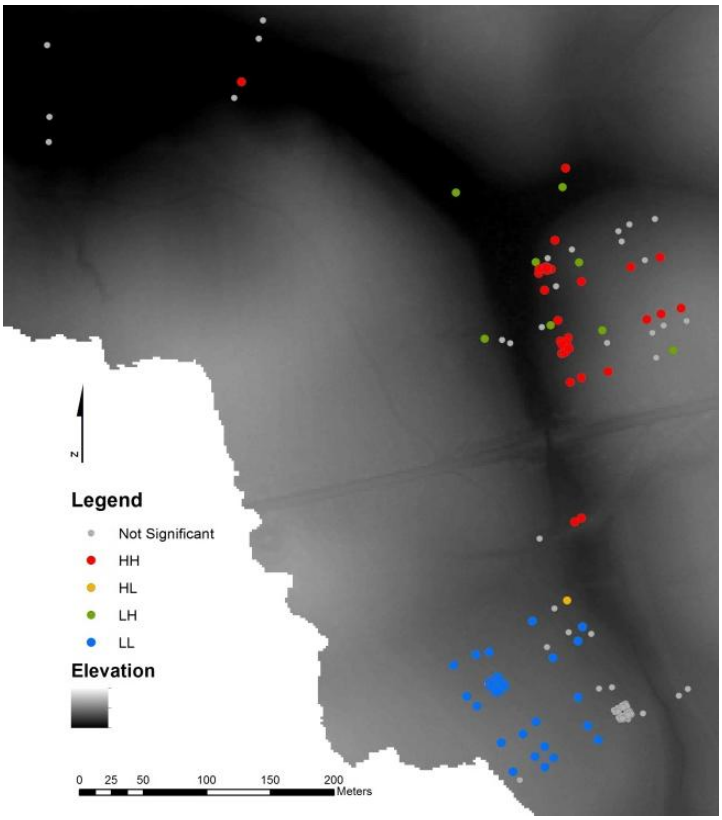

Figure B.3.2. Local Moran's I map, showing significant ($p<0.05$) clustering of high and low values of N. HH and LL stand for high and low clustering, respectively. HL and LH stand for high values surrounded by low values and low values surrounded by high values, respectively.

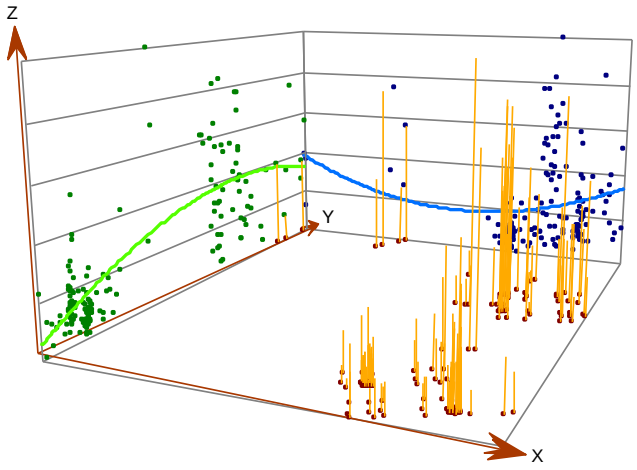

Figure B.3.3. Trend analysis of N (organic). There are 2nd degree polynomial trends in both a N-S direction and an E-W direction.

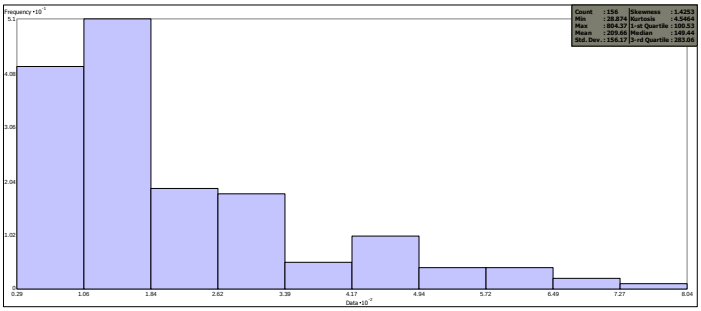

Figure B.3.4. N (organic) frequency distribution.

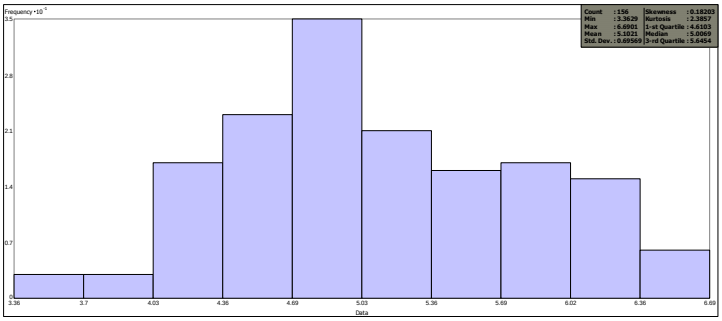

Figure B.3.5. Log-transformed N (organic) frequency distribution.

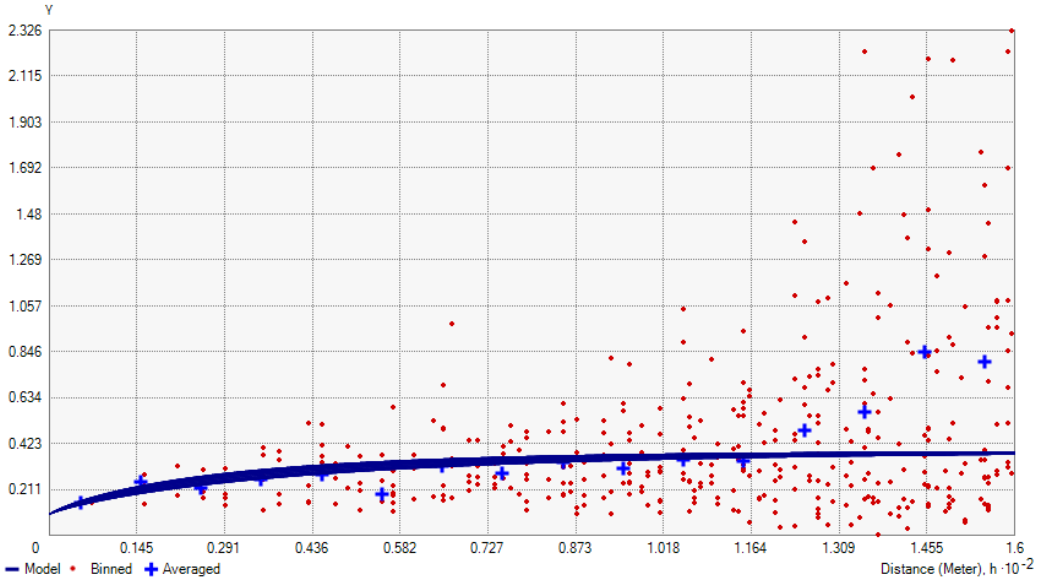

Figure B.3.6. Semivariogram for N (organic). The major axis is 160m, minor axis is 100m and the direction is 25.7°

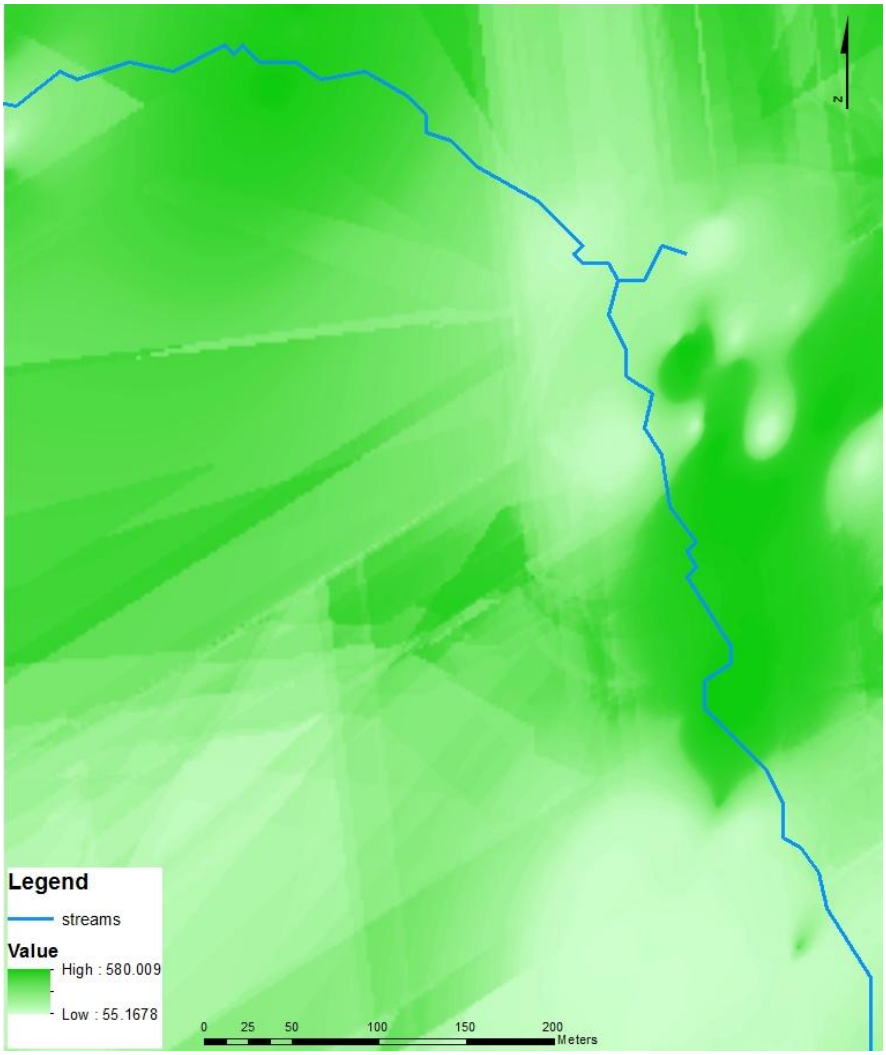

Figure B.3.7. Kriged map of N (organic) (kg/ha), using semivariogram from figure B.3.6. RMSE = 111.60. Avg. Standard Error = 0.6217.

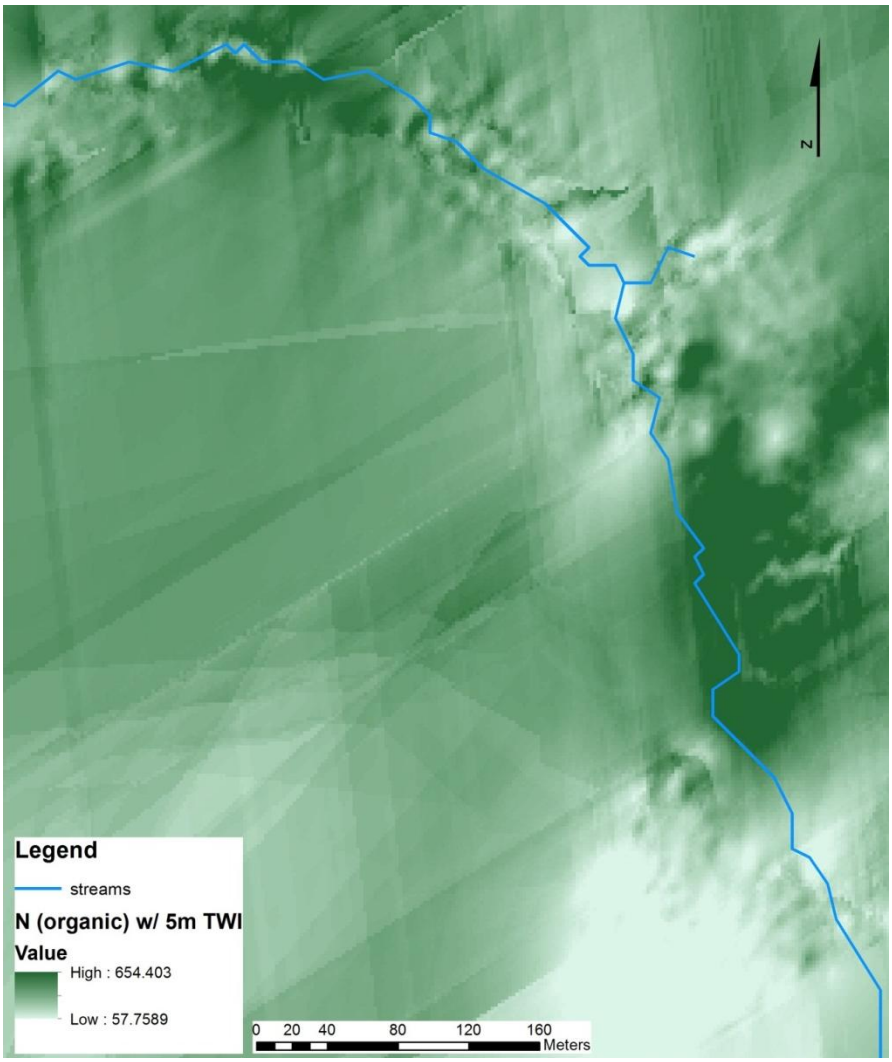

Figure B.3.8. Map of N (organic) cokriged with 5m TWI. RMSE = 110.23. Avg. Standard Error = 0.6141.

B.4 C (organic horizon)

n = 156

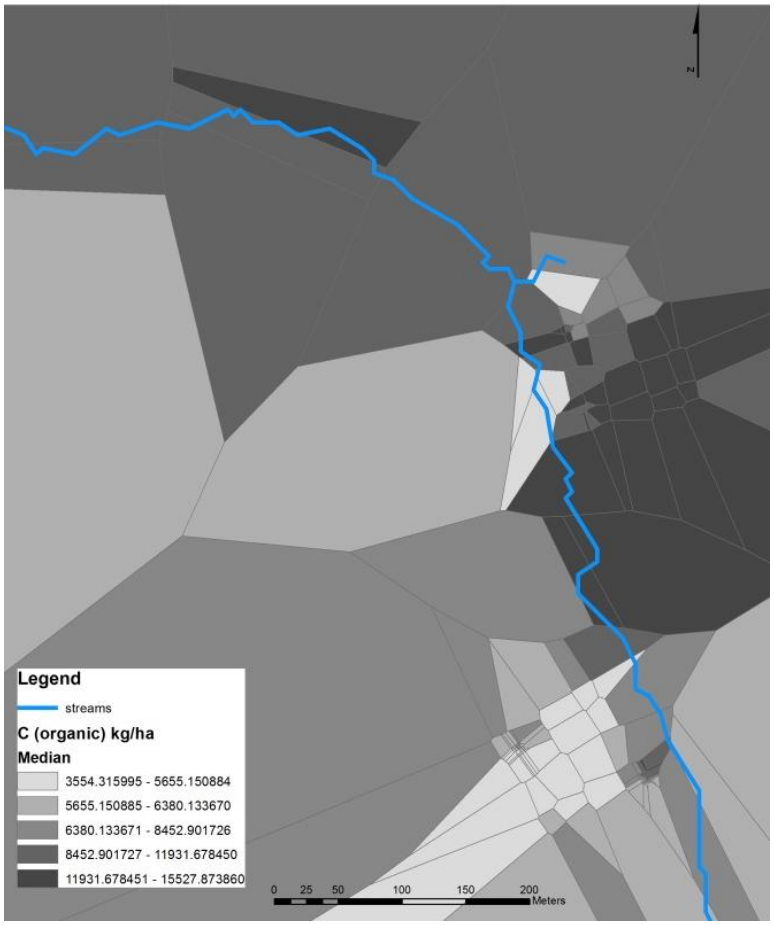

Figure B.4.1. Voronoi map of C (organic) values on the C basin.

Global Moran's I statistics:

Index: 0.244
Z score: 8.967
P < 0.0001

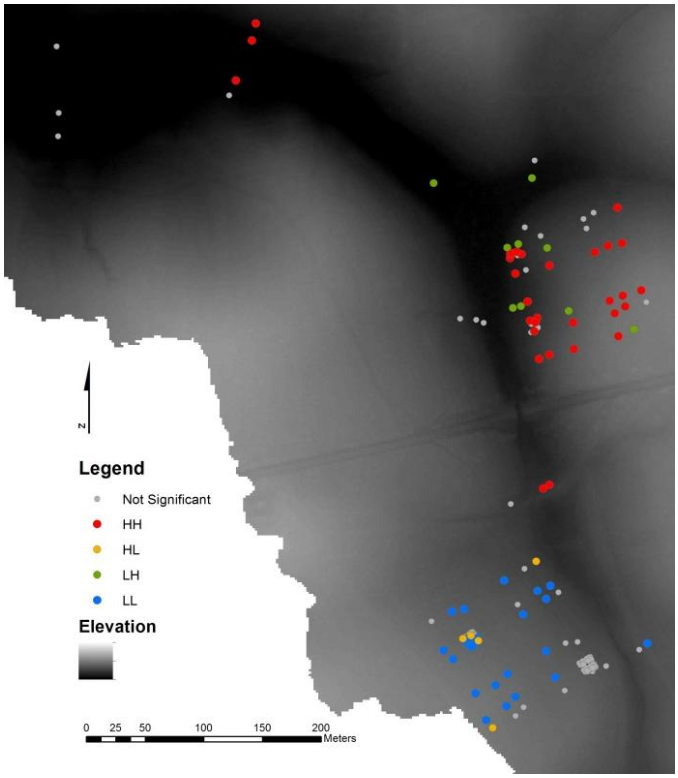

Figure B.4.2. Local Moran's I map, showing significant ($p<0.05$) clustering of high and low values of C. HH and LL stand for high and low clustering, respectively. HL and LH stand for high values surrounded by low values and low values surrounded by high values, respectively.

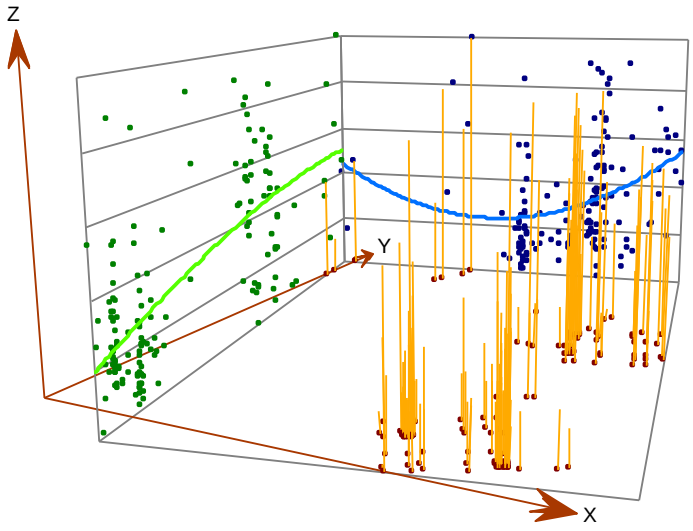

Figure B.4.3. Trend analysis of C (organic).

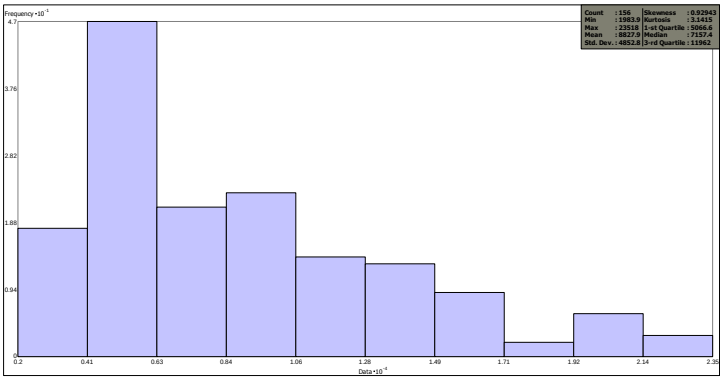

Figure B.4.4. C (organic) frequency distribution.

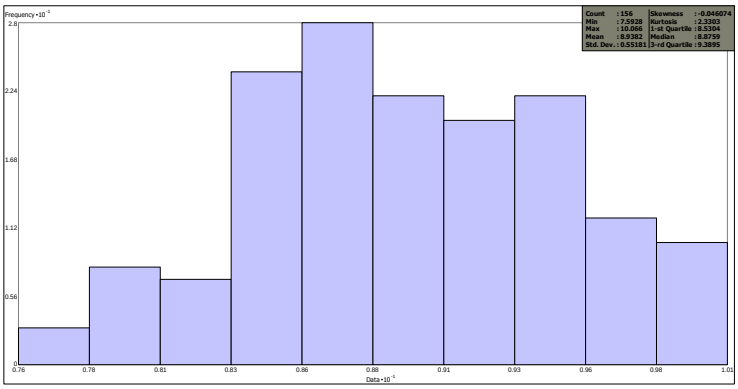

Figure B.4.5. Log-transformed C (organic) frequency distribution.

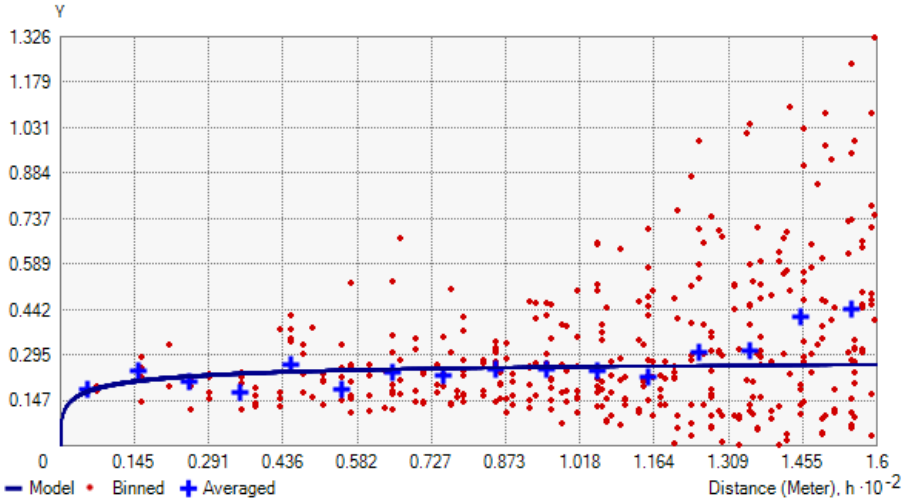

Figure B.3.6. Semivariogram for C (organic). The major axis is 160m, minor axis is 120m and the direction is 27.6°

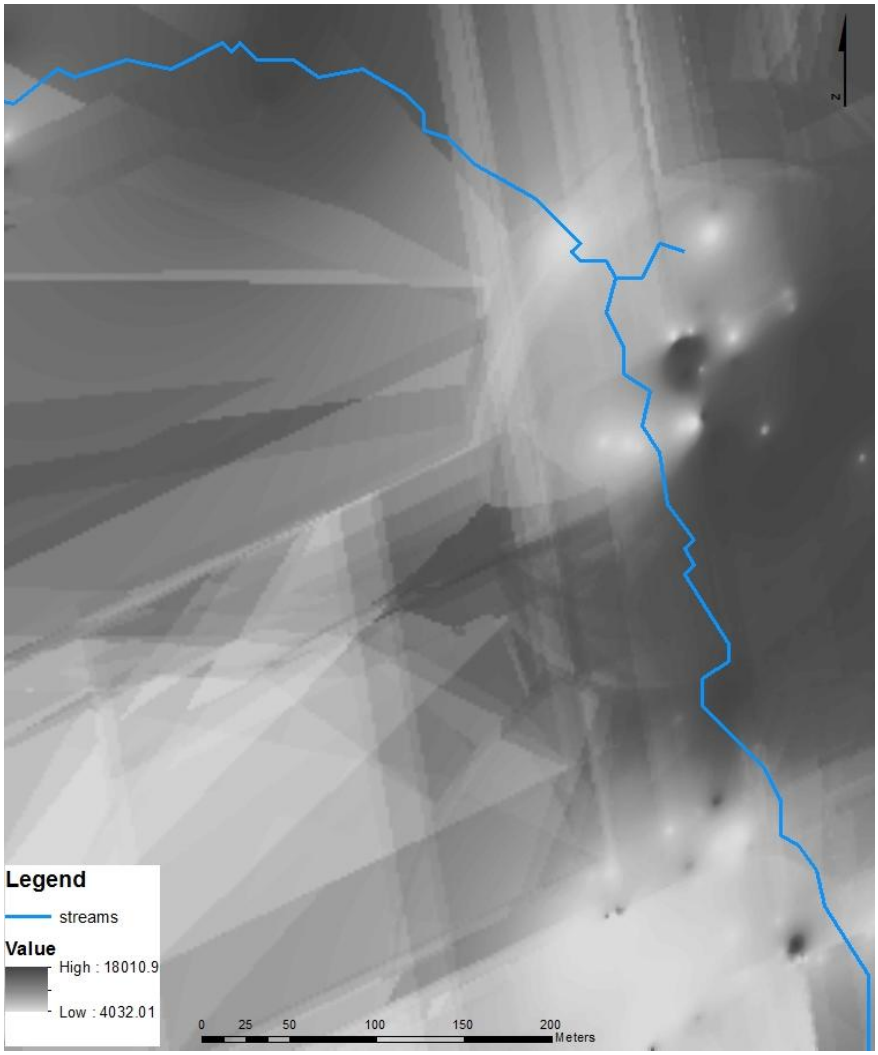

Figure B.4.7. Kriged map of C (organic) (kg/ha), using semivariogram from figure B.4.6. RMSE = 3949.35. Avg. Standard Error = 0.5600.

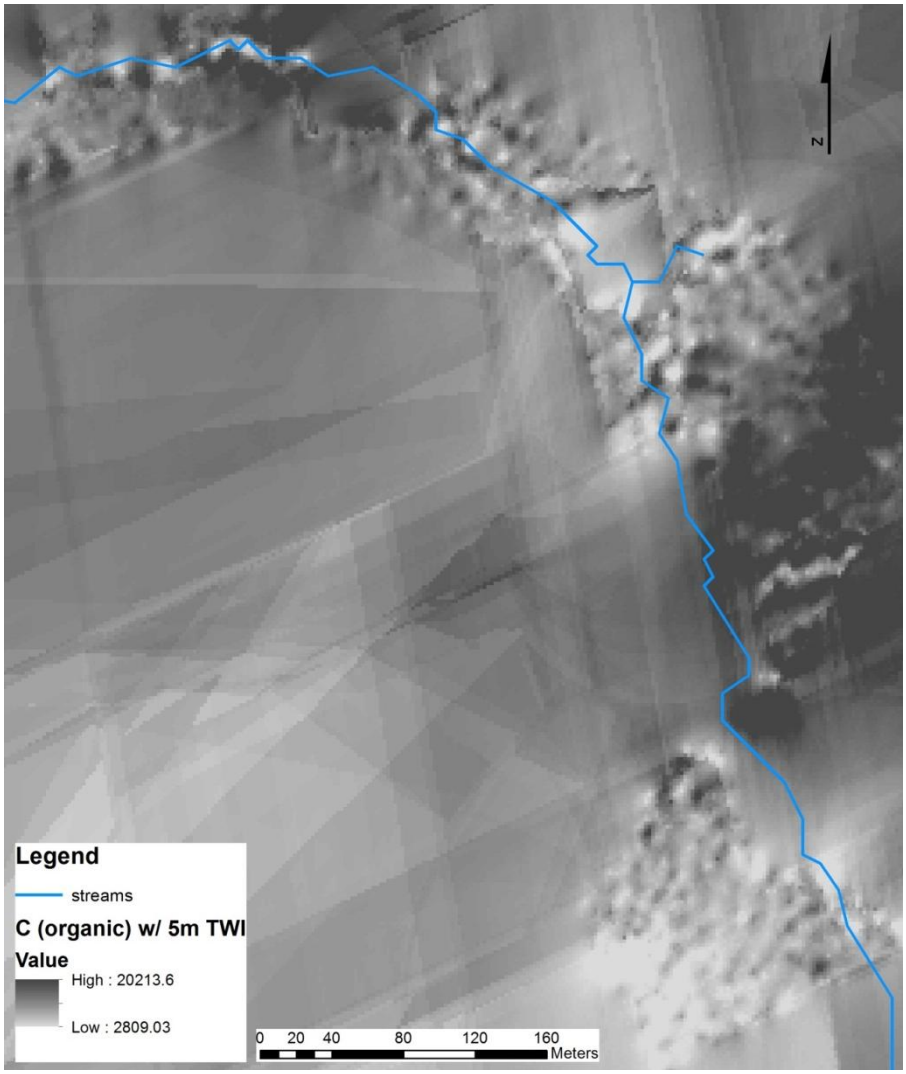

Figure B.4.8. Map of C (organic) cokriged with 5m TWI. RMSE = 3886.25. Avg. Standard Error = 0.5531.

B.5 Sand

n = 263

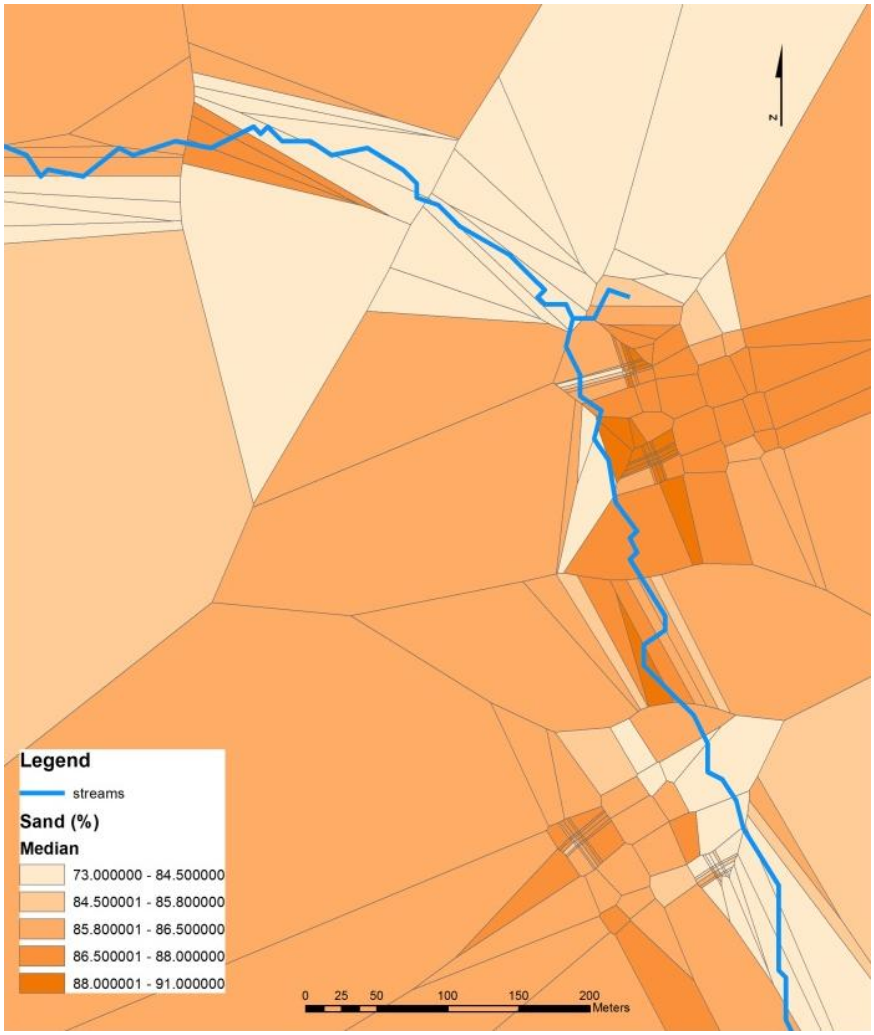

Figure B.5.1. Voronoi map of sand (%) values on the C basin.

Global Moran's I statistics:

Index: 0.176
Z score: 7.996
P < 0.0001

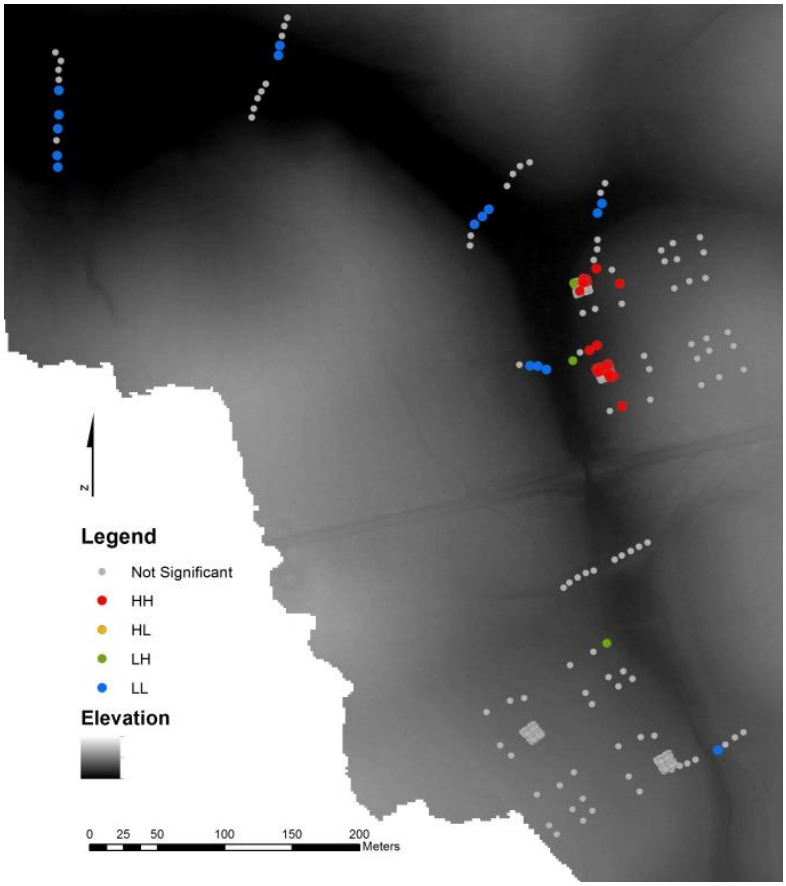

Figure B.5.2. Local Moran's I map, showing significant ($p<0.05$) clustering of high and low values of sand. HH and LL stand for high and low clustering, respectively. HL and LH stand for high values surrounded by low values and low values surrounded by high values, respectively.

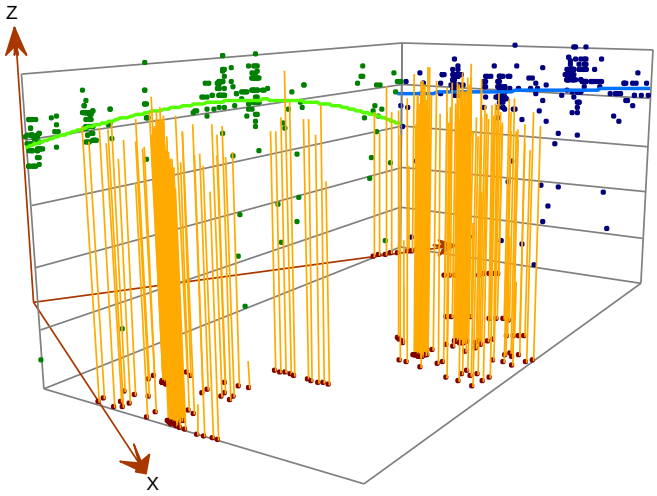

Figure B.5.3. Trend analysis of sand.

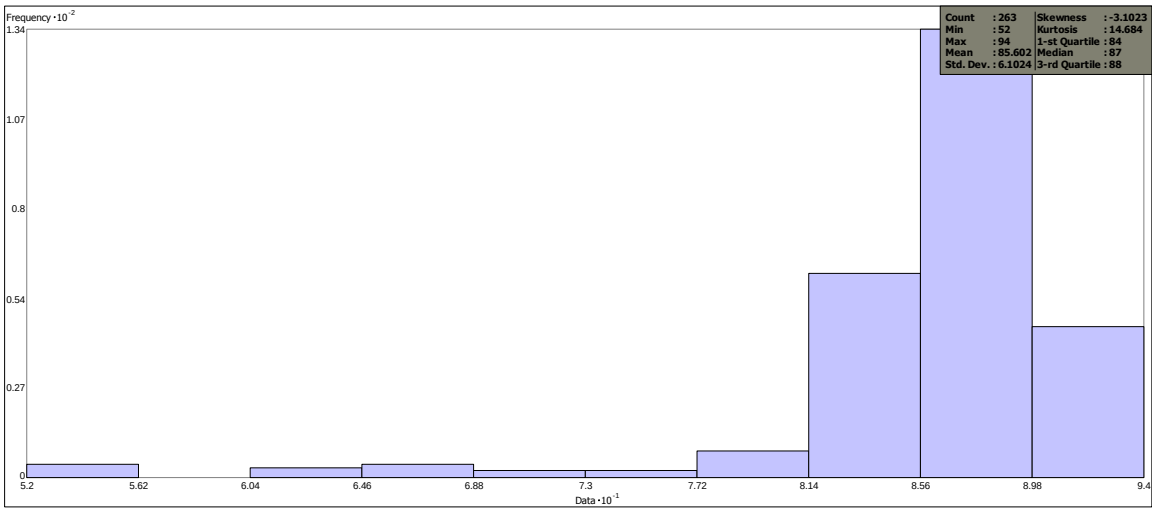

Figure B.5.4. Sand frequency distribution.

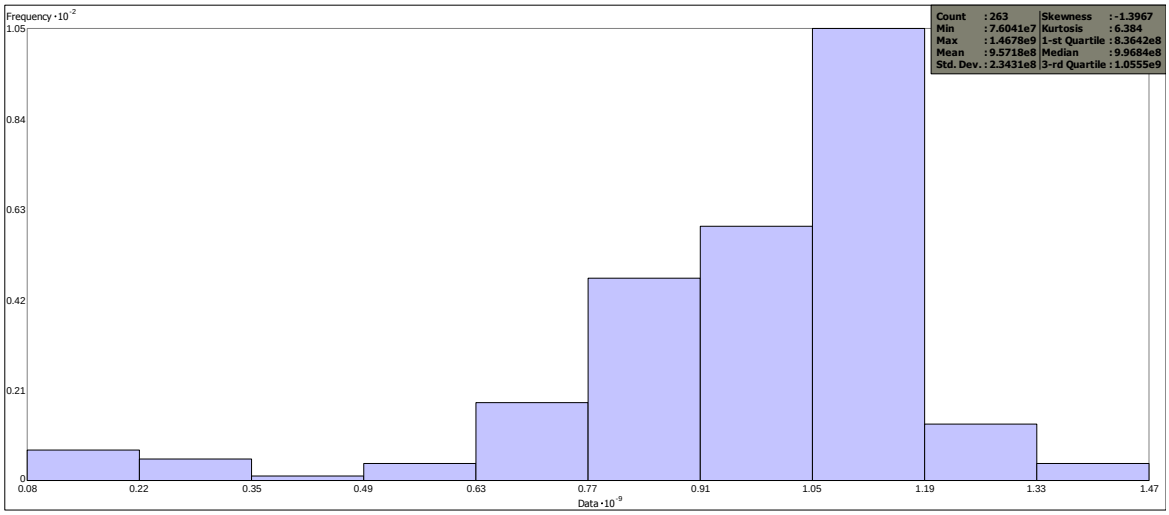

Figure B.5.5. Box-Cox-transformed (λ=5) sand frequency distribution.

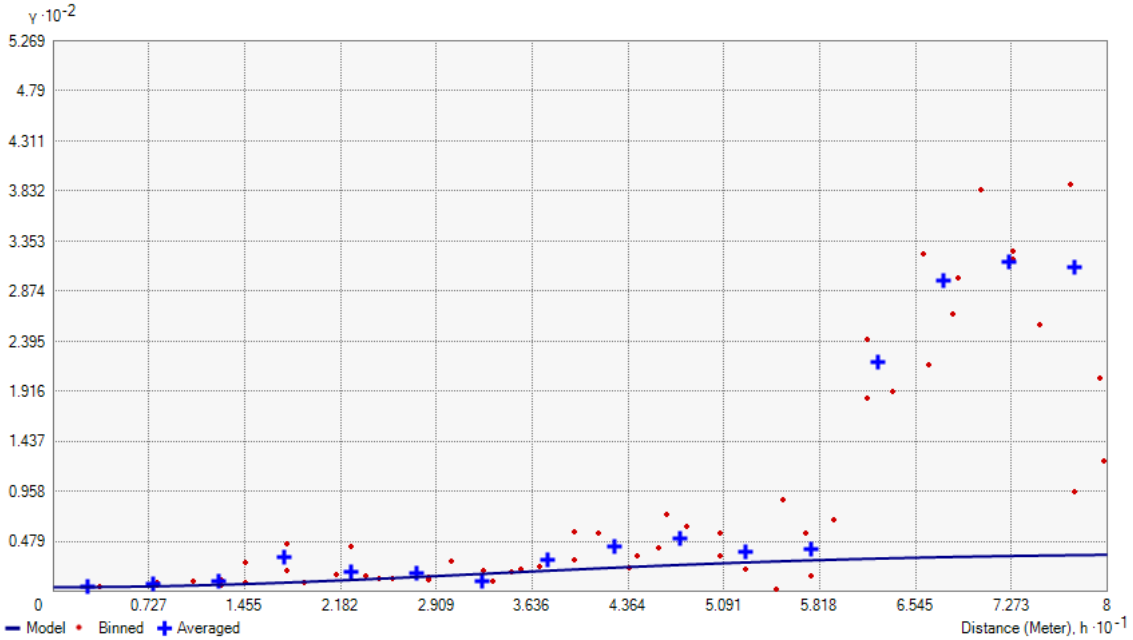

Figure B.5.6. Semivariogram for sand. The major axis is 80m, minor axis is 20m and the direction is 25.4°

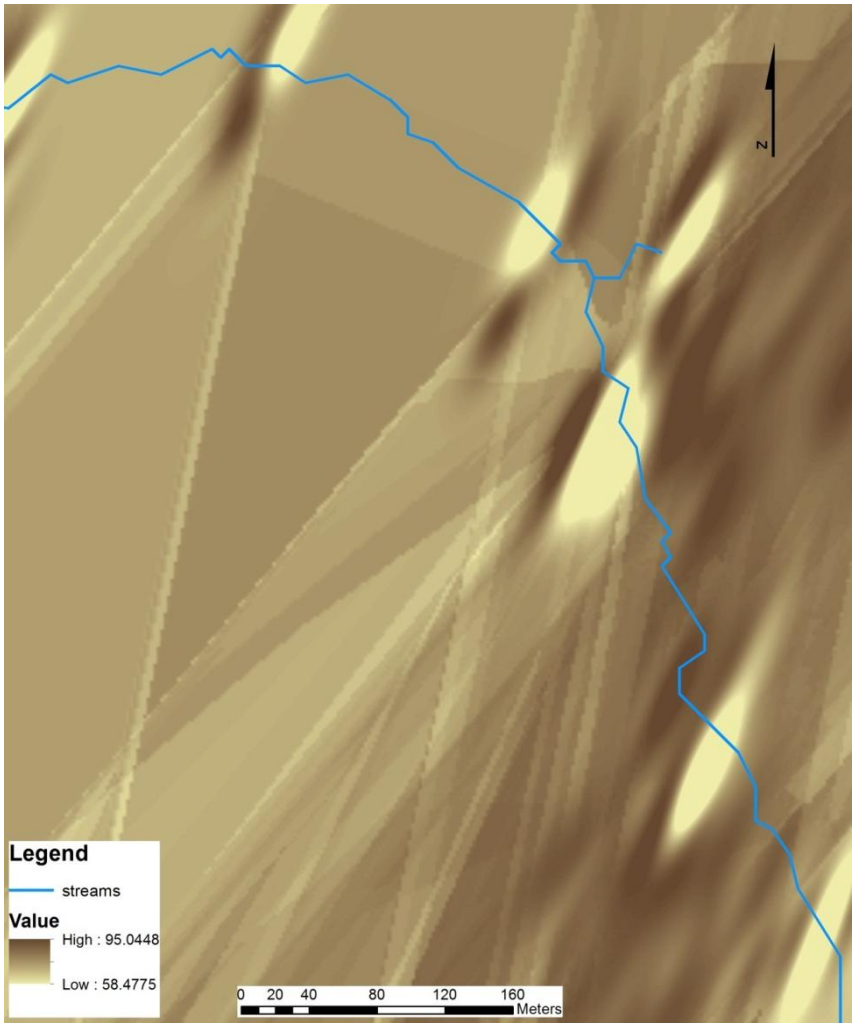

Figure B.5.7. Kriged map of sand (%), using semivariogram from figure B.5.6. RMSE = 5.414.

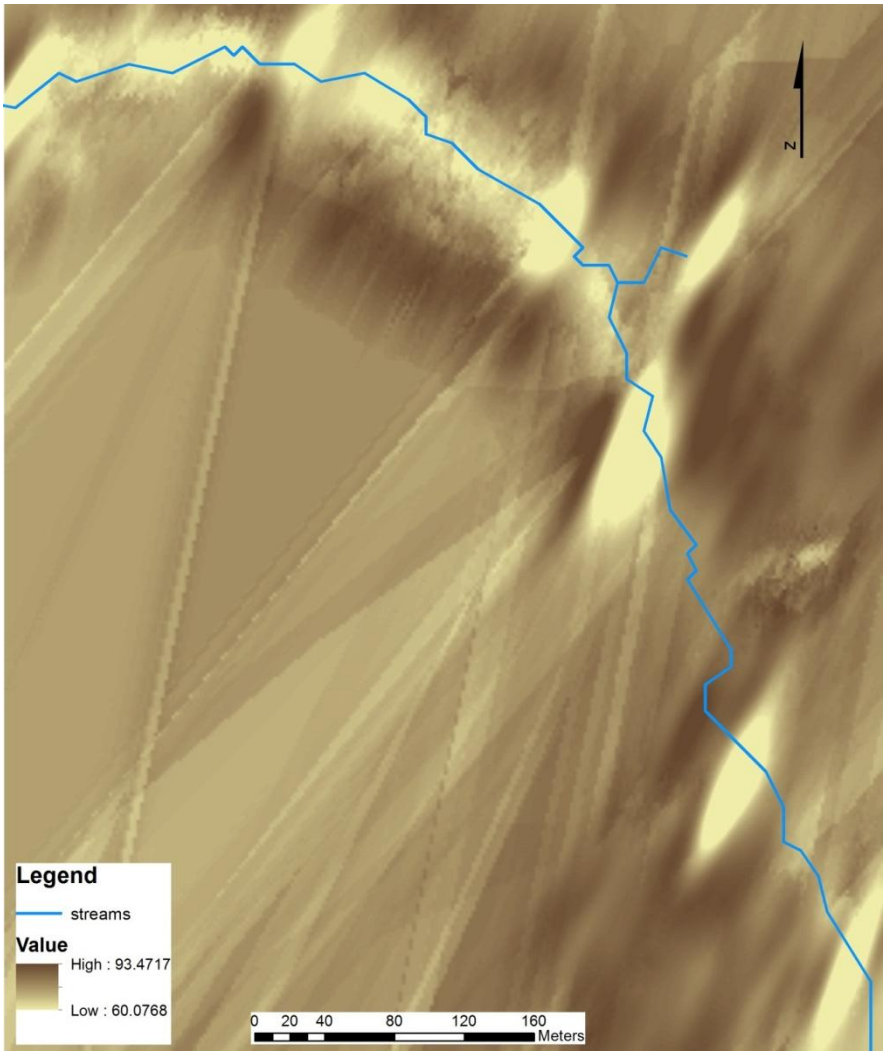

Figure B.5.8. Map of sand (%) cokriged with 10m DEM. RMSE = 5.190.

B.6 Silt

n=263

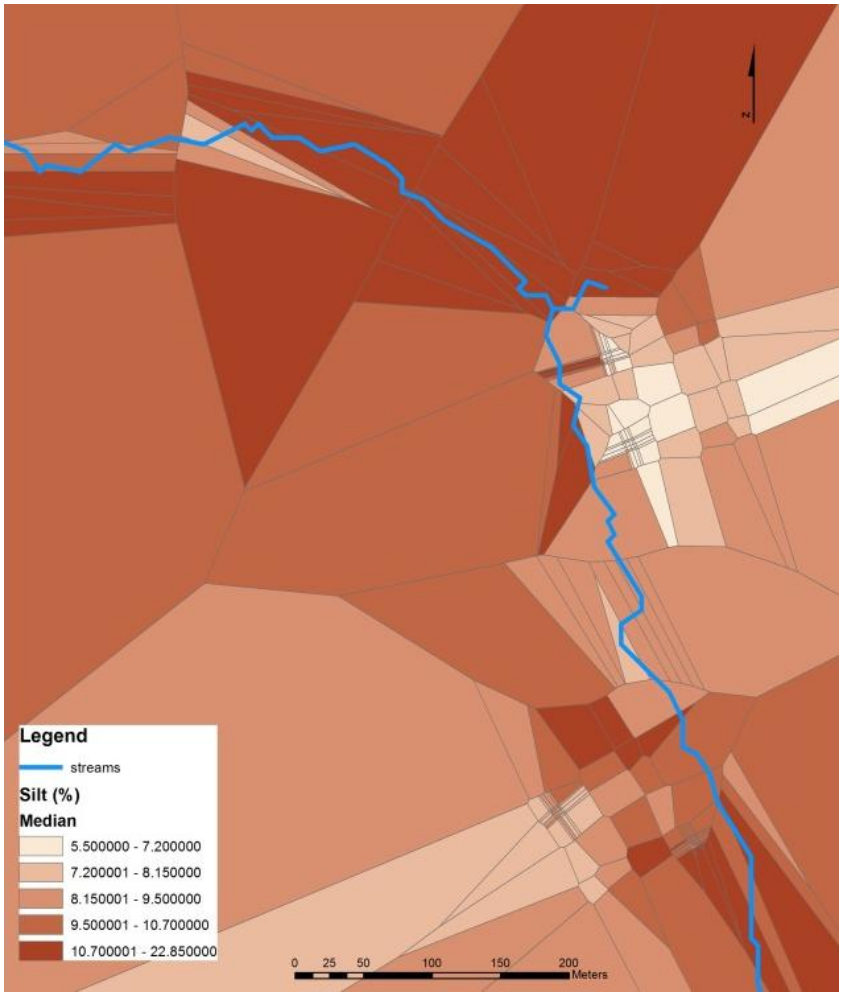

Figure B.6.1. Voronoi map of silt (%) values on the C basin.

Global Moran's I statistics:

Index: 0.165
Z score: 7.508
P < 0.0001

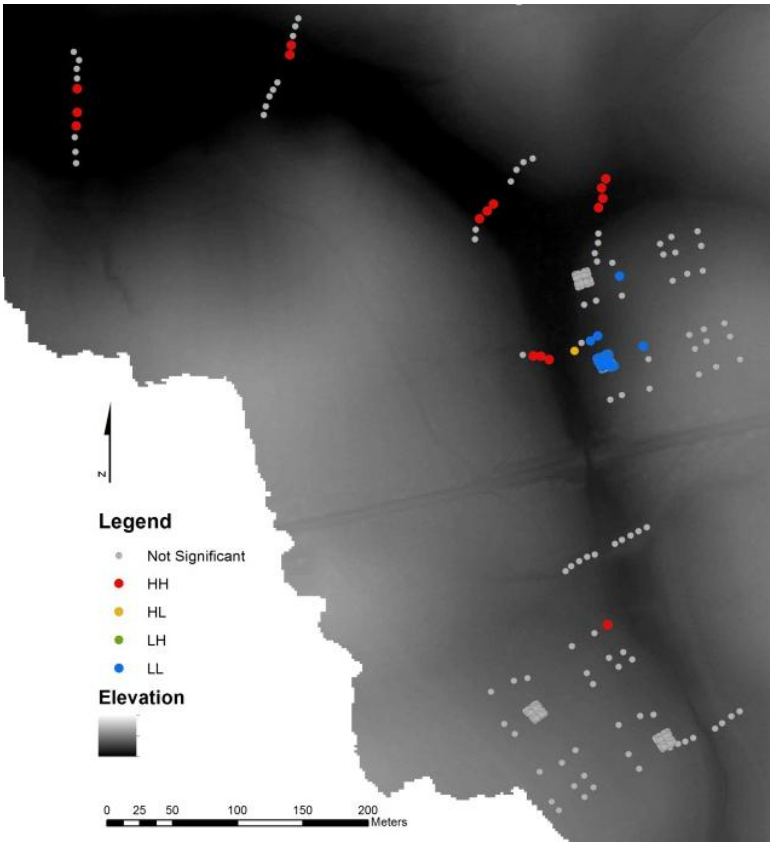

Figure B.6.2. Local Moran's I map, showing significant ($p<0.05$) clustering of high and low values of silt. HH and LL stand for high and low clustering, respectively. HL and LH stand for high values surrounded by low values and low values surrounded by high values, respectively.

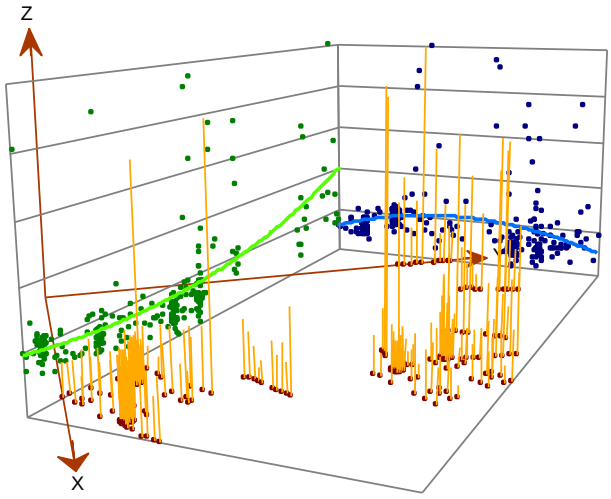

Figure B.6.3. Trend analysis of silt.

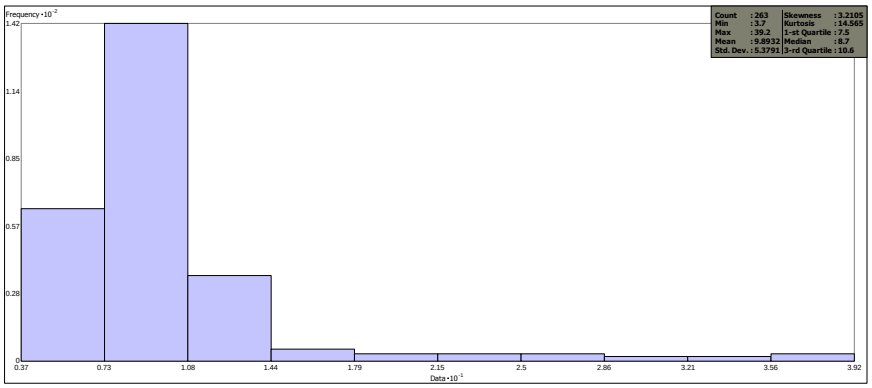

Figure B.6.4. Silt frequency distribution.

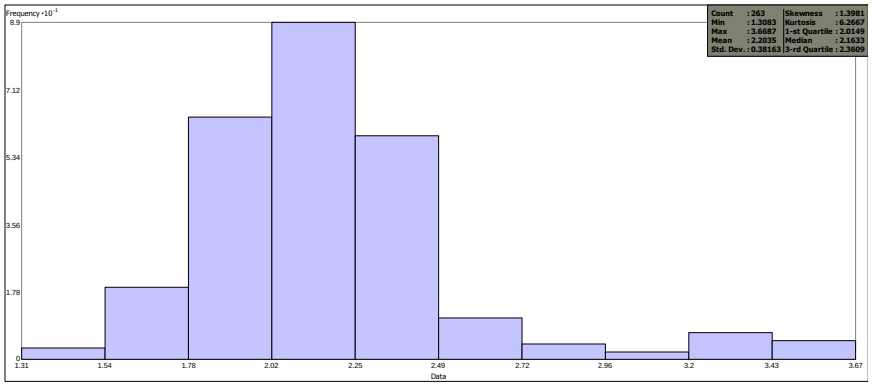

Figure B.6.5. Log-transformed silt frequency distribution.

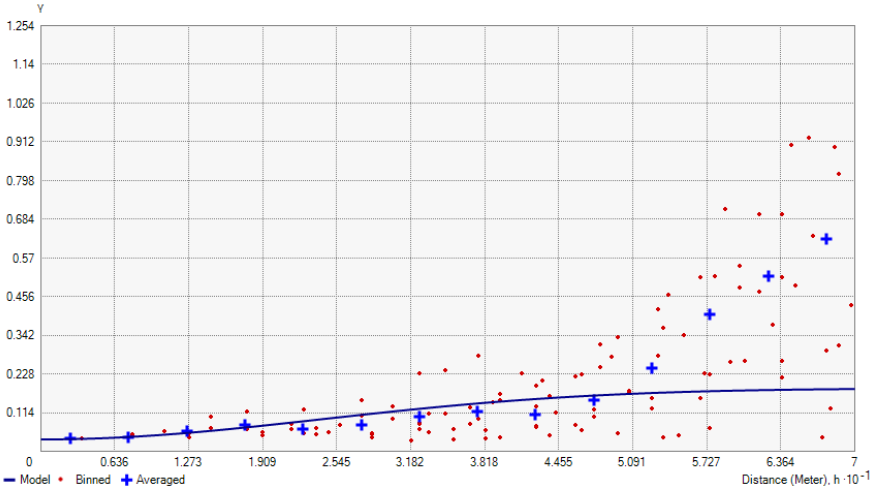

Figure B.6.6. Semivariogram for sand. The major axis is 70m, minor axis is 20m and the direction is 26.8°

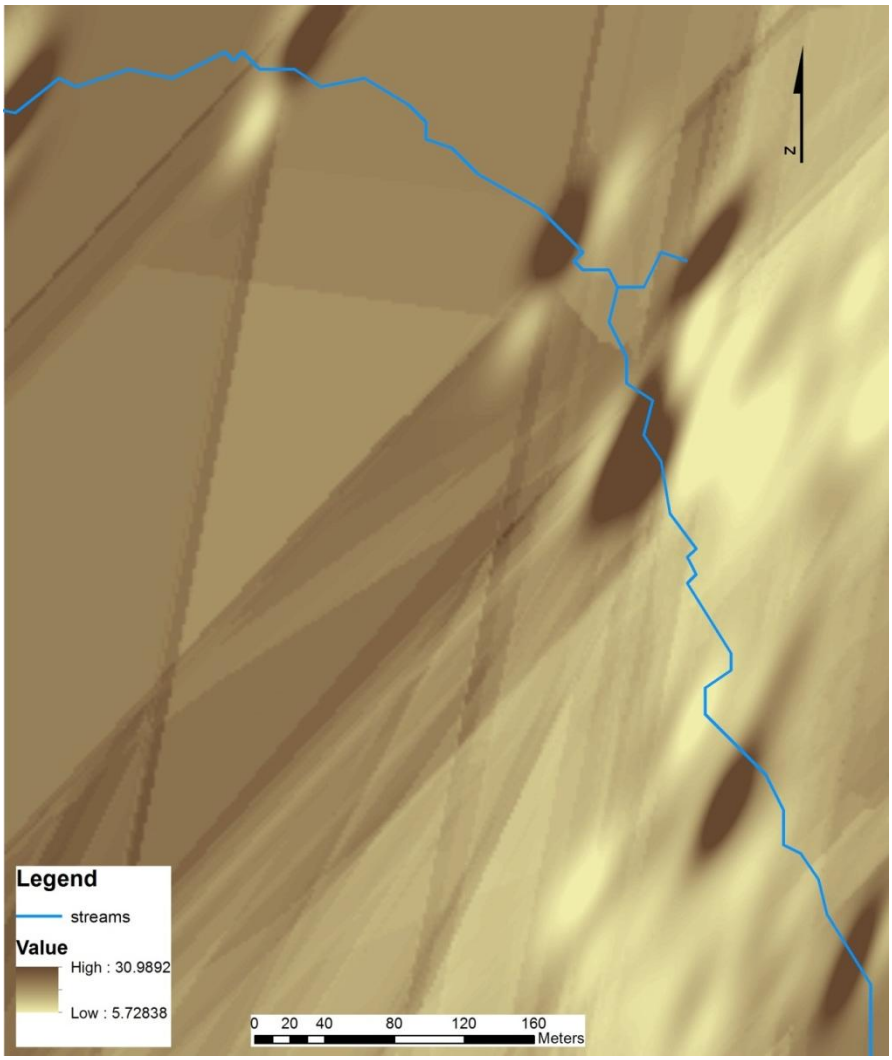

Figure B.6.7. Kriged map of silt (%), using semivariogram from figure B.6.6. RMSE = 4.214.

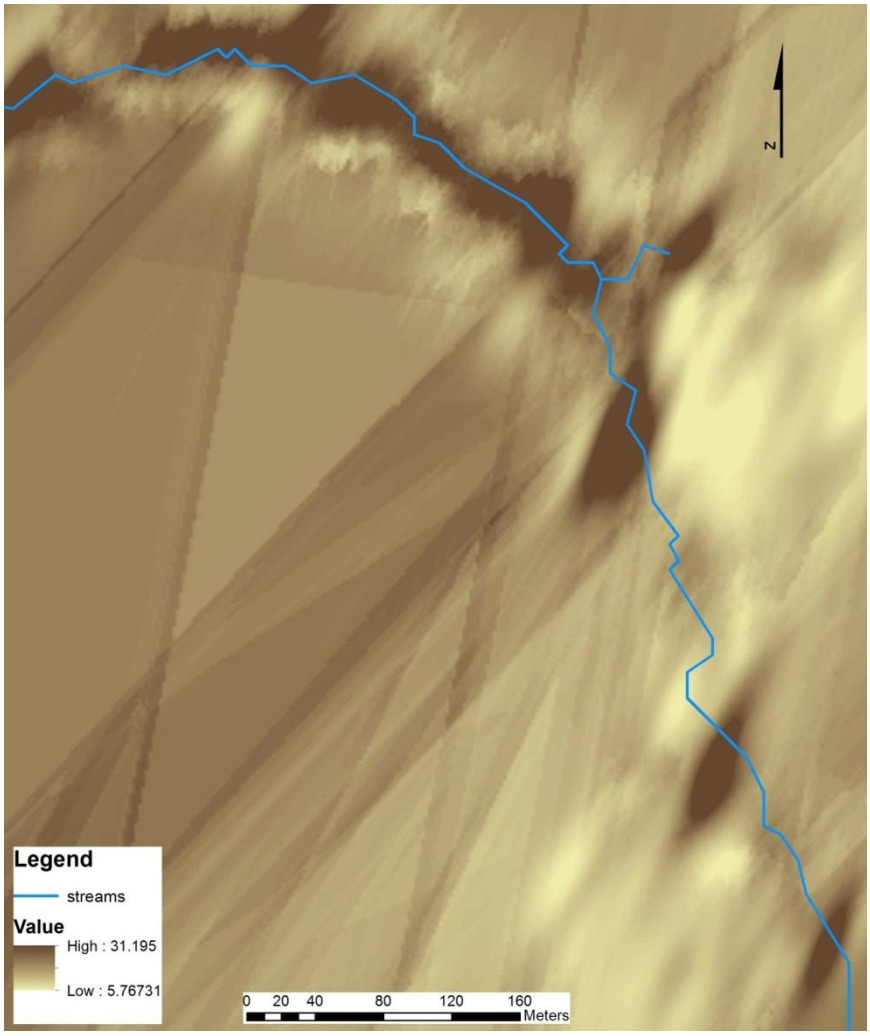

Figure B.6.8. Map of silt (%) cokriged with 30m DWI. RMSE = 3.978.

B.7 Bulk density

n = 261

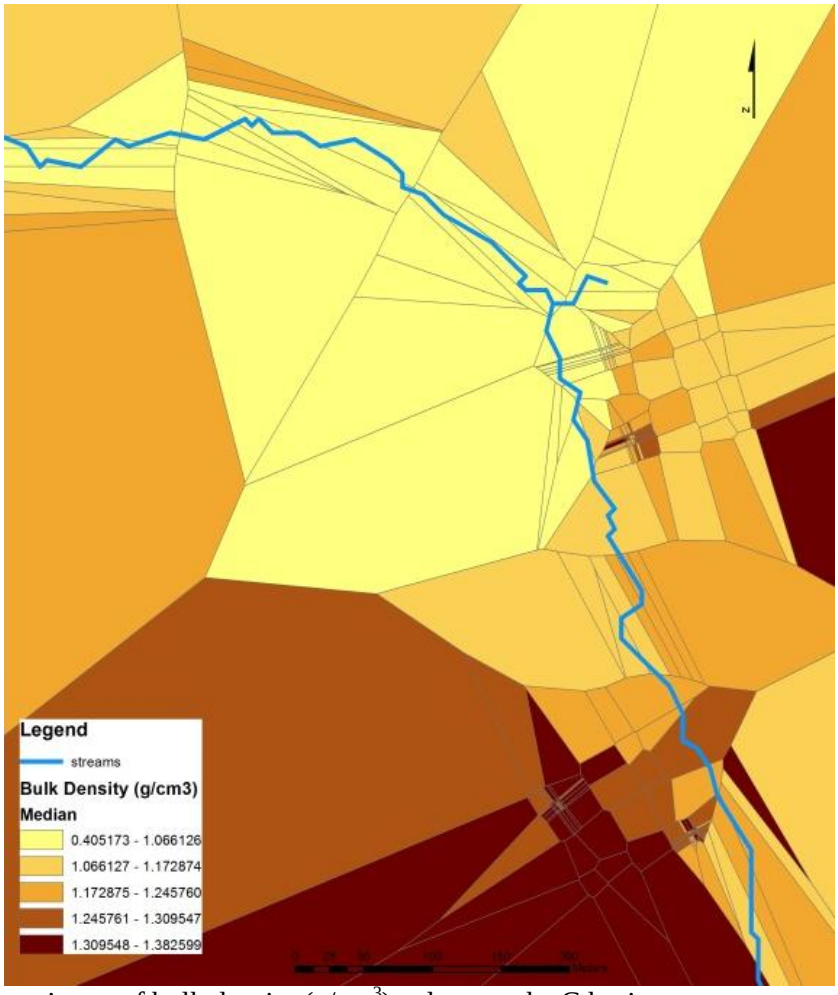

Figure B.7.1. Voronoi map of bulk density (g/cm^3) values on the C basin.

Global Moran's I statistics:

Index: 0.370
Z score: 21.036
P < 0.0001

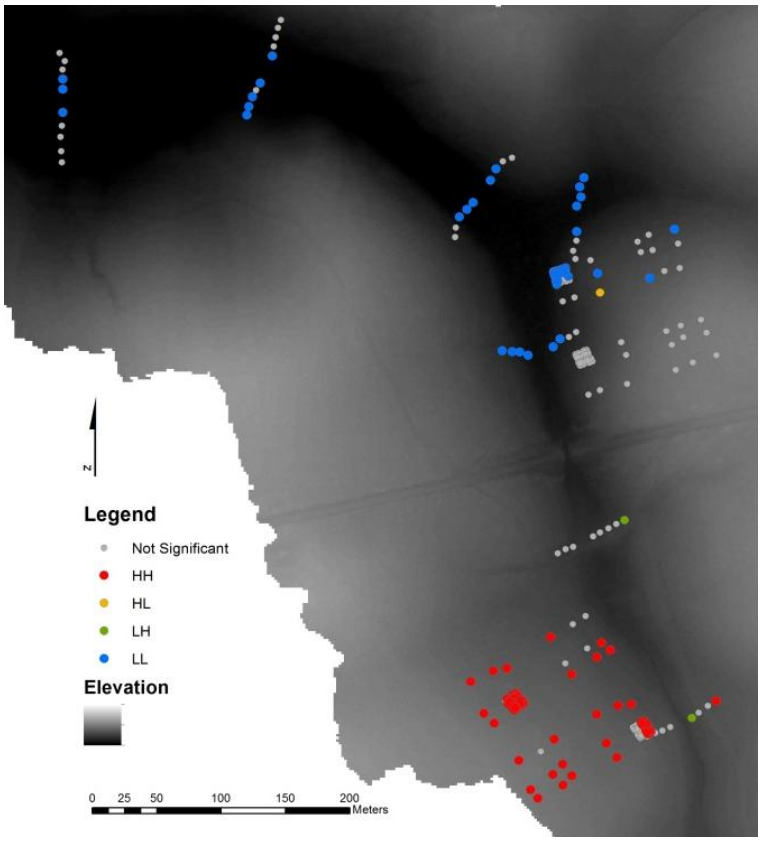

Figure B.7.2. Local Moran's I map, showing significant ($p<0.05$) clustering of high and low values of bulk density. HH and LL stand for high and low clustering, respectively. HL and LH stand for high values surrounded by low values and low values surrounded by high values, respectively.

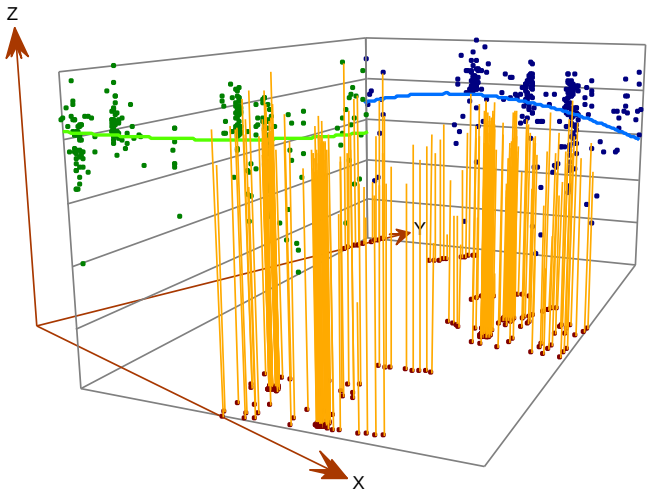

Figure B.7.3. Trend analysis of bulk density.

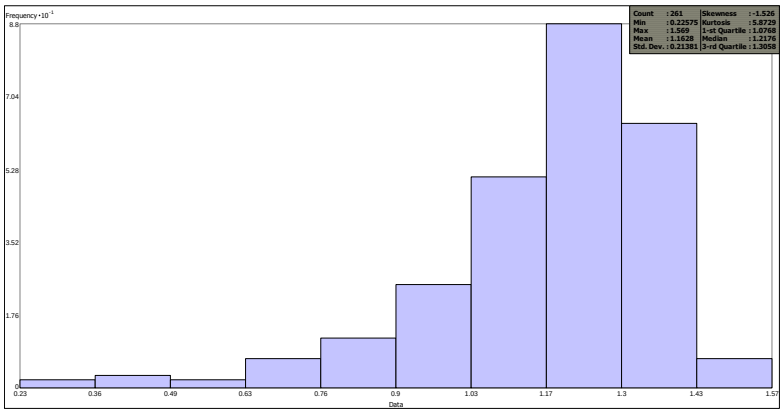

Figure B.7.4. Bulk density frequency distribution.

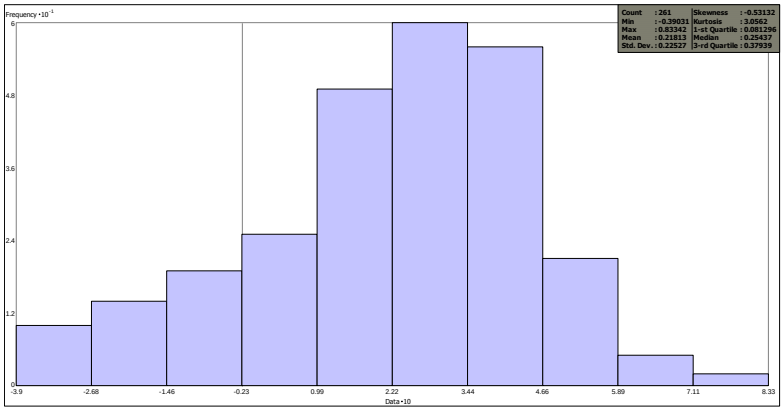

Figure B.7.5. Box-Cox-transformed (λ=2.5) bulk density frequency distribution.

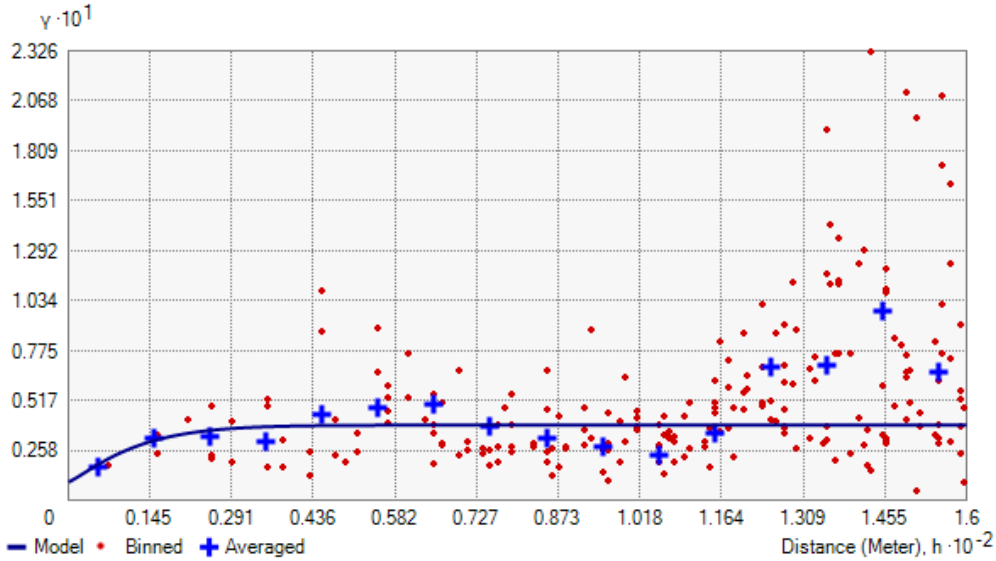

Figure B.7.6. Semivariogram for bulk density. The major axis is 160m, minor axis is 120m and the direction is 122.5°

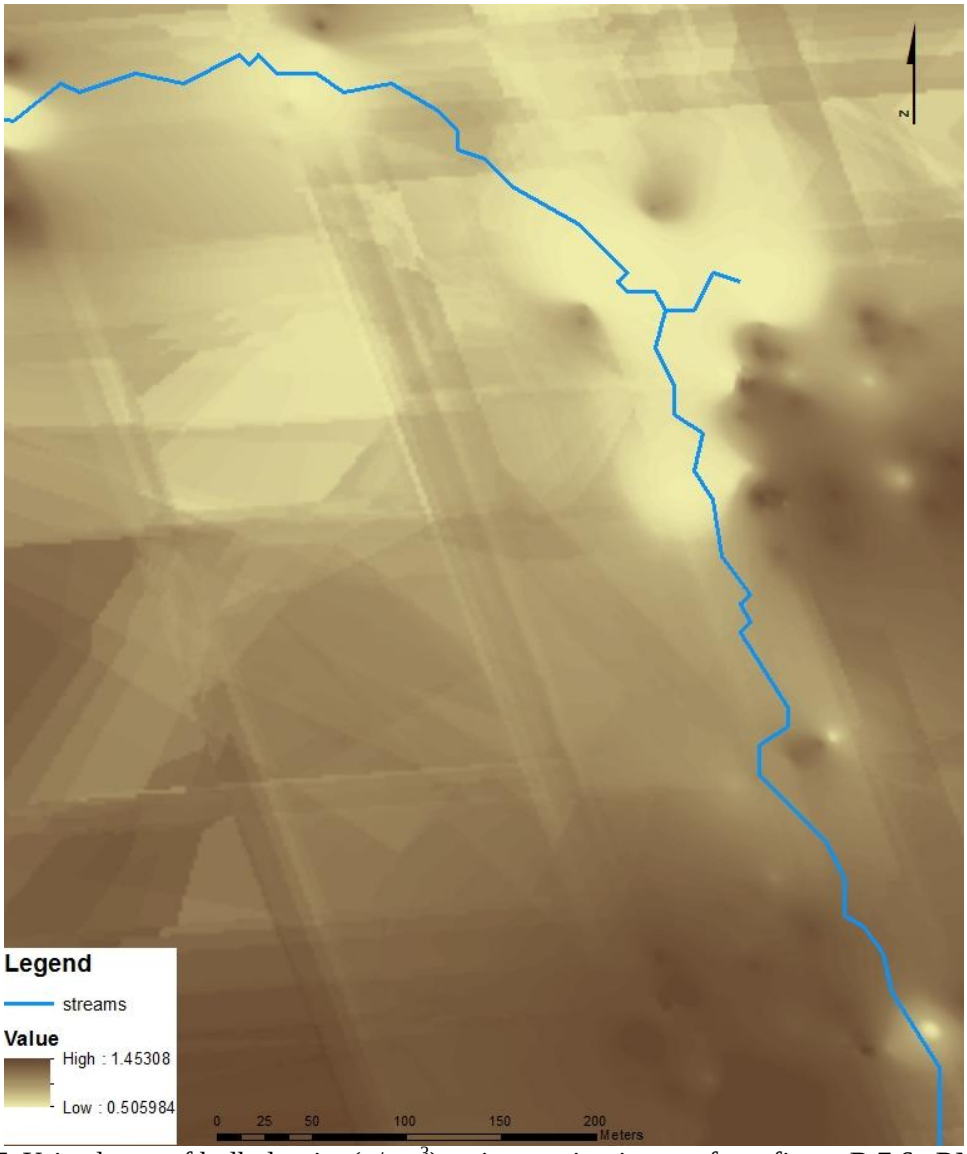

Figure B.7.7. Kriged map of bulk density (g/cm^3), using semivariogram from figure B.7.6. RMSE = 0.1374.

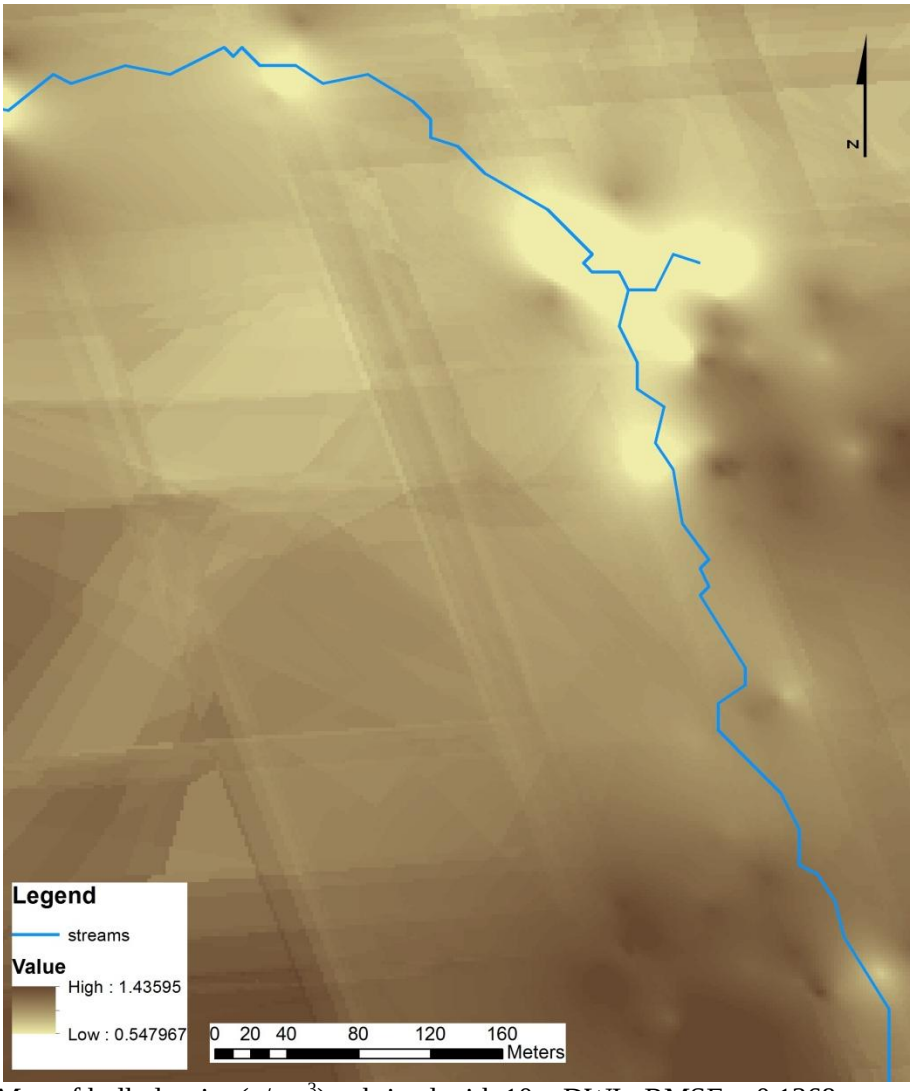

Figure B.7.8. Map of bulk density (g/cm^3) cokriged with 10m DWI. RMSE = 0.1368.

B.8 A horizon thickness

n=142

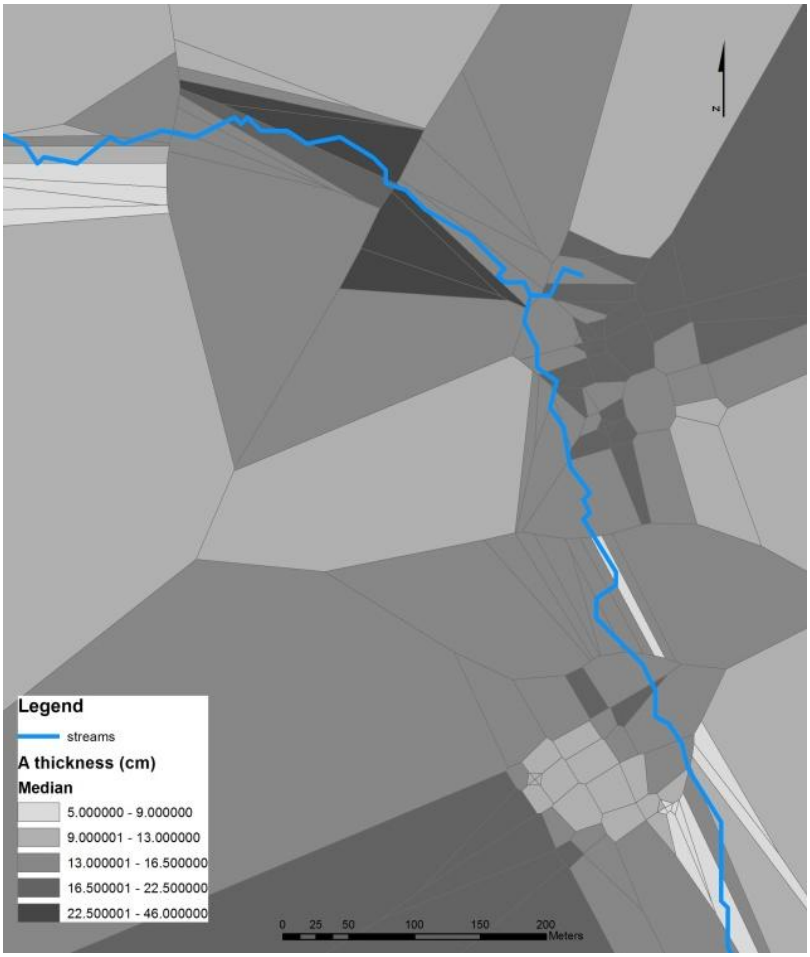

Figure B.8.1. Voronoi map of A horizon thickness (cm) values on the C basin.

Global Moran's I statistics:

Index: 0.176
Z score: 6.953
P < 0.0001

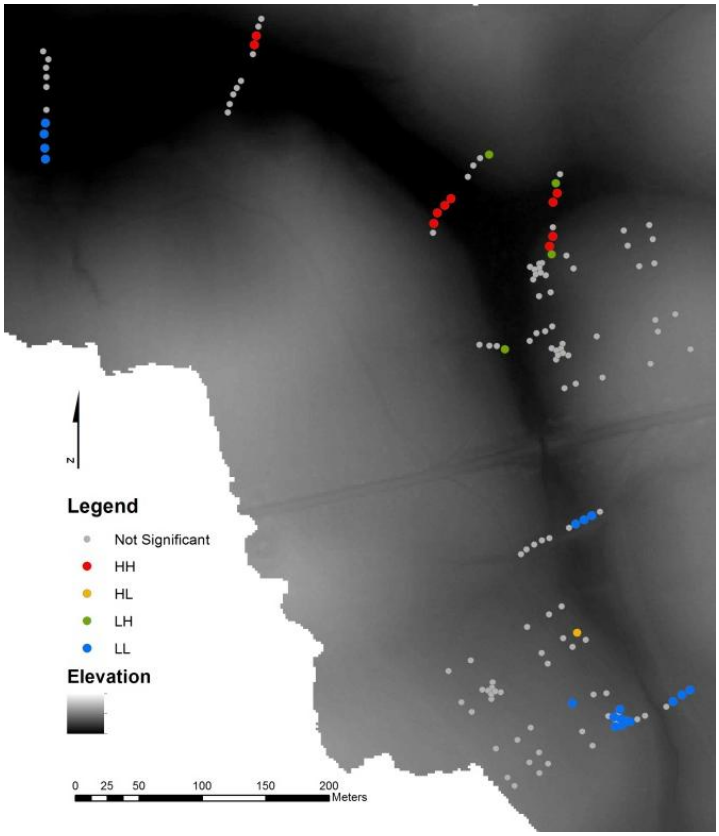

Figure B.8.2. Local Moran's I map, showing significant ($p<0.05$) clustering of high and low values of A thickness. HH and LL stand for high and low clustering, respectively. HL and LH stand for high values surrounded by low values and low values surrounded by high values, respectively.

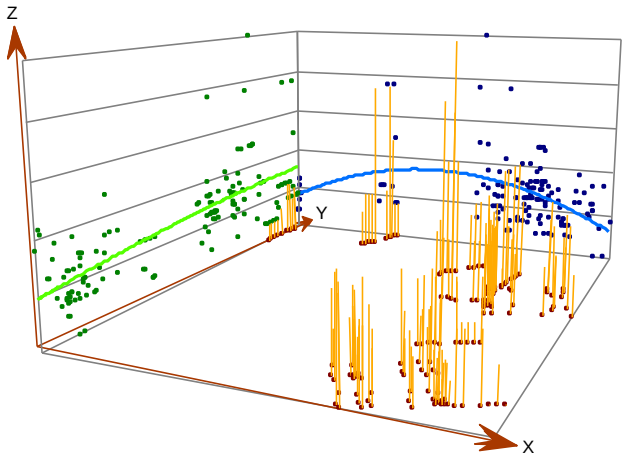

Figure B.8.3. Trend analysis for A horizon thickness.

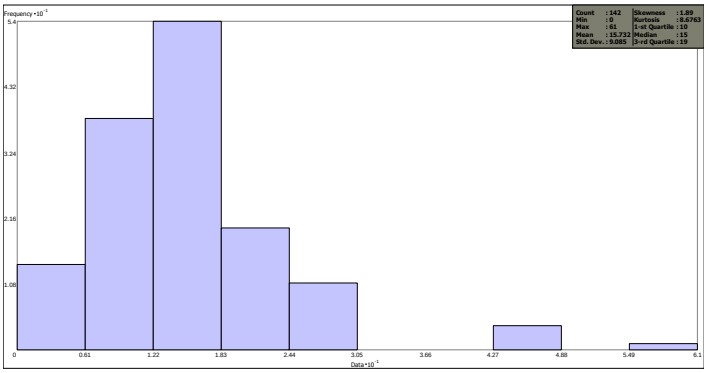

Figure B.8.4. A horizon thickness frequency distribution. Note: Since some of the values are 0, the data couldn't be log-transformed.

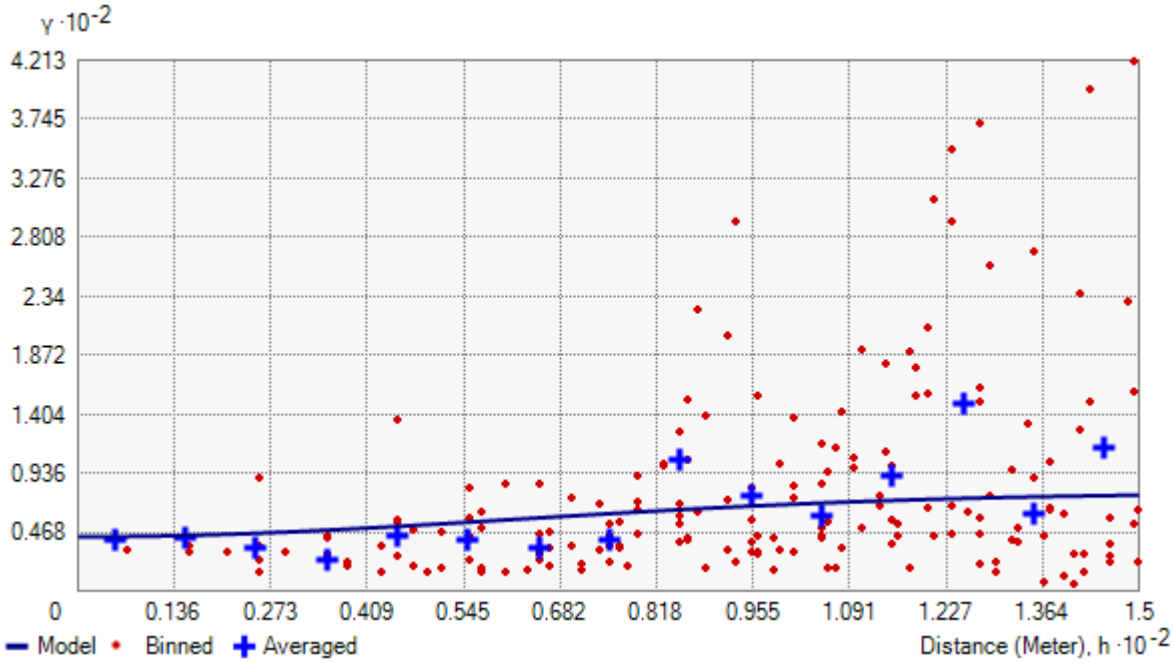

Figure B.8.5. Semivariogram for A thickness. The major axis is 150m, minor axis is 90m and the direction is 136.7°

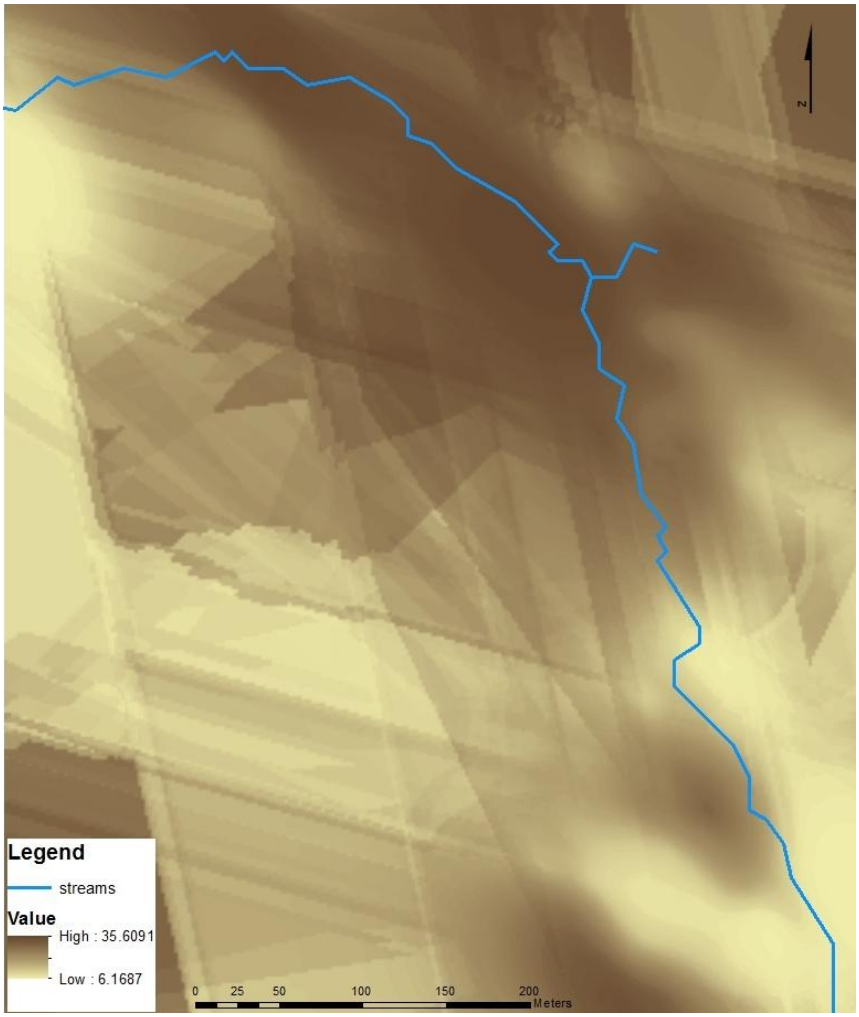

Figure B.8.6. Kriged map of A horizon thickness (cm), using semivariogram from figure B.8.5. RMSE = 8.158.

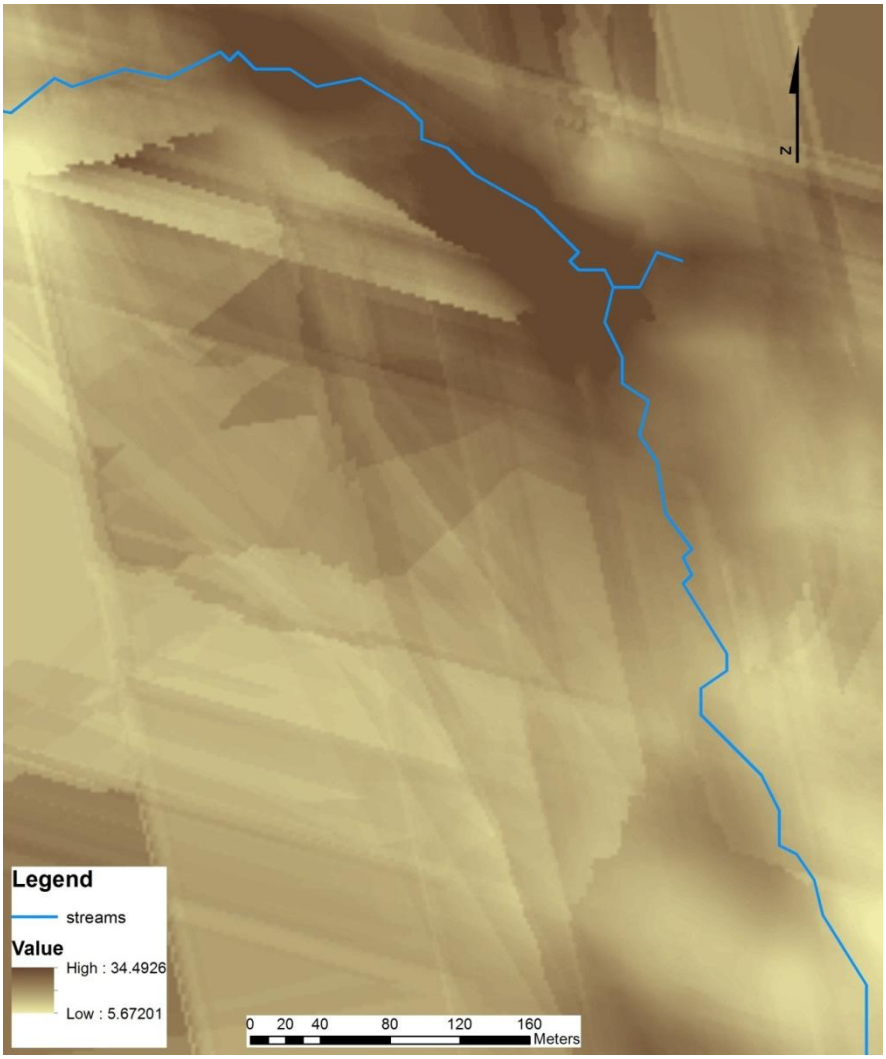

Figure B.8.7. Map of A horizon thickness (cm) cokriged with 50m DEM. RMSE = 7.984.