12-2016

An examination of geographic patterns of soil climate and its classification in the U.S. system of soil taxonomy

Hans Edwin Winzeler
Purdue University

Graduate School Form
30 Updated 12/26/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Hans Edwin Winzeler

Entitled

AN EXAMINATION OF GEOGRAPHIC PATTERNS OF SOIL CLIMATE AND ITS CLASSIFICATION IN THE U.S. SYSTEM OF SOIL TAXONOMY

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Phillip R. Owens
Co-chair

Brad Joern
Co-chair

Venkatesh Merwade

Darrel Schulze

Zamir Libohova

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Phillip R. Owens and Brad Joern

Approved by: Joe M. Anderson 12/7/2016

Head of the Departmental Graduate Program Date

AN EXAMINATION OF GEOGRAPHIC PATTERNS OF SOIL CLIMATE AND ITS CLASSIFICATION IN THE U.S. SYSTEM OF SOIL TAXONOMY

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Hans Edwin Winzeler

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2016

Purdue University

West Lafayette, Indiana

ACKNOWLEDGEMENTS

My dissertation advisor Phillip R. Owens provided support for me during the various phases of this project and I wish to thank him and acknowledge his indispensable help. I also wish to thank the Graduate School at Purdue University for their support. I would like to also thank Professor James R. Schupp.

TABLE OF CONTENTS

Page

ABSTRACT

Winzeler, Hans E. Ph.D., Purdue University, December 2016. An Examination of Geographic Patterns of Soil Climate and its classification in the U.S. System of Soil Taxonomy. Major Professors: Brad Joern and Phillip R. Owens.

Soil climate, the record of temporal patterns of soil moisture and temperature, is an important component of the structure of U.S. Soil Taxonomy. The U.S. Soil Survey has used the Newhall Simulation Model (NSM) for estimating soil climate from atmospheric climate records at weather stations since the 1970s. The current soil climate map of the U.S. was published in 1994 by using NSM runs from selected weather stations along with knowledge-based hand-drawn mapping procedures. We developed a revised soil climate mapping methodology using the NSM and digital soil mapping techniques.

The new methodology is called Grid Element Newhall Simulation Model (GEN), where a coordinate system is used to divide geographic space into a grid and each element or grid-cell serves as a reference area for querying and organizing model input, and for organizing and displaying model output. The GEN was used to make a soil moisture map of the conterminous U.S. (GEN-CONUS). GEN-CONUS and the 1994 map were compared to each other and to two sets of weather station data from years 1961

to 1990 and years 1971 to 2000 (National climate data center, NCDC). Agreement between GEN-CONUS and the 1994 map was 75.6%. GEN-CONUS had higher agreement than the 1994 map with NSM
output from NCDC data for 1961-1990 and 1971-2000 (kappa = 0.845 and 0.777). The GEN methodology was also used to generate a map of projected soil climate in the year 2080 for part of the Southern Rocky Mountains, predicting expansion of the Ustic and contraction of the Udic moisture regimes.

Soil climate in the conterminous US is expected to change in response to global climate change. Soil moisture and temperature are strongly influenced by atmospheric climate variables. The Grid Element Newhall Simulation Model (GEN), an updated NSM for geographic raster data, was developed and applied in this project to future climate simulations available from International atmospheric climate prediction projects. These included a simulation of 1) current climate conditions, 2) climate in year 2070 under a radiative forcing increase scenario of 2.6 W m^{-2} above pre-industrial levels (a low estimate) and 3) climate in the year 2070 under a radiative forcing scenario increase of 8.5 W m^{-2} (higher estimate). Soil climate classification was analyzed to determine the extent and character of soil climate reclassification that might be necessary in coming decades. Results indicate that 18% of the land area of the conterminous US would be reclassified into a new temperature regime in the low-radiative forcing scenario and 37% would be reclassified in the high forcing scenario. In general, soil moisture decreased in future climate change scenarios, leading to increased water deficits for many geographic areas due to greater evapotranspiration and warmer soil temperature during the growing season. The dominant temperature regime change was that from the Mesic temperature regime to the Thermic and Hyperthermic regimes under both the high and low radiative forcing increases. Changes from the colder temperature regimes in Northern states to more Mesic regimes was also noted. The geographic pattern expected for changes in moisture regime shows far more change in the western part of CONUS than in the east, with changes from moist conditions to more arid conditions predominating. Some limited areas in the arid Southwest are expected to become wetter, particularly under the high radiative forcing estimate. Orographic changes in moisture and temperature follow the general trend of increasing temperature and decreasing moisture in future climate change scenarios.

As a driver of soil development and a key factor of soil formation, climate influences physical and chemical properties of soils as they form from geological and biological material. In this study we examine soil climate as simulated by the NSM and its relation to georeferenced point observations of soil properties measured and recorded over many decades by the National Cooperative Soil Survey. The goal is to determine the strength and direction of relationships between geographic observations of soil properties that may have been influenced by climate and the simulations of soil climate for the same locations. An additional goal is to determine whether the NSM as a process model contributes substantially to an accounting of the interaction between atmospheric climate and any resulting soil properties, or whether a simpler observational model that does not include simulation of soil moisture and temperature interactions might be sufficient or superior to this simulation approach. The observational model includes the same input directly taken from atmospheric climate datasets as that used to populate the NSM, but does not include simulation of how the atmospheric climate would translate into soil climate through simulation of moisture and temperature dynamics in the soil.

We find that the NSM may have some value as a tool to explain a few relationships between climate and soil properties observed in the NCSS dataset, but that direct observation without simulation also shows promise. Severe limitations in the NCSS data include unknown sampling biases, ambiguous geographical precision of observation, inconsistent sampling and analysis protocols, incomplete data records, etc. Limitations of the usefulness of the NSM include high levels of multicollinearity among model output parameters, adherence to moisture modelling behavior that does not account for the complexities of preferential flow, the assumption of free-drainage in all soils modelled, the lack of a ponding routine or a realistic accounting of snow melt dynamics, as well as other limitations. These limitations may restrict the results of this study from providing firm conclusions, but exploratory analysis does indicate some positive correlations between atmospheric climate and soil properties, particularly after atmospheric datasets are applied to simulation of soil climate through the NSM.

CHAPTER 1. A METHODOLOGY FOR EXAMINING CHANGES IN SOIL CLIMATE GEOGRAPHY THROUGH TIME

1.1 Introduction

Soil climate may be defined as the long-term record of seasonal and diurnal patterns of perhaps the most dynamic properties of soils, those of moisture and temperature. Historical maps of soil climate in the U.S. have largely been conceived of and displayed as thematic maps, with polygons delineating approximate geographic boundaries between taxonomic groupings. The traditional methodology for the production of such maps in the U.S. Soil Survey has relied on expert knowledge and delineation of areas by hand. This reliance on experts to create soil climate maps through manual delineation has several limitations (Zhu et al., 2001). These include limitations to the size of the soil body that can be delineated, limited ability to update maps rapidly and efficiently, and the inevitability of errors when maps are drawn with visual examination of environmental covariates (Zhu et al., 2001). In addition, hand-delineated expert maps require experts for every map iteration, making them inefficient in cases when iterations are desirable, such as when maps of soils of multiple time periods are desired. Also, knowledge that facilitated the production of a map made with expert knowledge most often remains tacit within the mapping product (Hudson, 1992). When this happens discussion with the maker of the map may be the only way to determine why certain delineations were made. This is particularly problematic in fields in which soil change is under study because the timescale for such change can be much longer than the working lives of individual investigators. Expert knowledge, if not systematically applied, can be inconsistent, with multiple experts providing conflicting or differing opinions. Some opinions may change given further evidence or consideration. In contrast to methodologies of mapping using hand delineation through expert knowledge, digital soil mapping techniques based on geographical information systems data layers use environmental covariates and models to produce map output in systematic repeatable ways (McBratney et al., 2003). In this paper we test whether map production through direct application of a soil moisture model to geospatial data layers can lead to more consistent model output than a historical hand-delineated map made using expert knowledge.

Broad scale soil climate maps are useful in harmonization of local soil surveys, including the effort currently underway by the USDA/NRCS Soil Survey Division, the Soil Data Joining and Re-correlation initiative, to correct abrupt changes in soil maps at political boundaries (Dobos et al., 2010; Scheffe et al., 2012). Continental scale soil climate maps offer versions of soil properties that can be analyzed and used without artifacts caused by variations in analysis at local political boundary lines. As such, they can be correlated to soil taxonomic properties and integrated with historical soil maps to provide greater consistency. Broad-scale soil climate maps can also be generated for multiple iterations of climate data to assess climate change. Future biotic conditions accompanying climate change, particularly with respect to soil climate as it affects agricultural and forest productivity are a central interest of the USDA Climate Change Science Plan (USDA, 2012).

Soil climate has been an important component of the taxonomic structure of U.S. *Soil Taxonomy* since the release of the 7th Approximation and the first publication of Soil Taxonomy (Soil Survey Staff, 1960; Soil Survey Staff, 1975). Soil climate characteristics have been used to differentiate taxa from the Order (the highest level) down to the suborder, great group, subgroup, family, and soil series levels. Soil climate was originally envisioned in *Soil Taxonomy* to follow the older concepts of zonality and intrazonality (USDA, 1938), which were considered untenable as a natural classification because the concepts were not based on discernible soil properties (Smith, 1986). It was reasoned that soil climate properties, while often dynamic within daily, seasonal, and annual patterns, were nevertheless measurable quantities that could be observed and recorded. Early adoption of soil climate concepts led to a recognition that though the markers of soil climate could be immediately observed (in terms of soil moisture content and soil temperature at the time of observation), the actual long-term climate that would determine a soil's moisture or temperature regime would require extrapolation from records of atmospheric climate until appropriate data sets of long-term soil climate could be populated (Smith, 1986).

The process of soil climate simulation modeling based on atmospheric climate station data for various periods of record has been operational at the Order and Suborder level in mapping applications internationally (Van Wambeke, 1982), and within the U.S. (Smith, 1986; Soil Survey Staff, 1975; USDA-SCS, 1994) for the past four decades. Recently, Bonfante et al. (2011) addressed gaps and resolution conflicts between physically-based models, USDA soil moisture classes, and climate-driven approaches such as the Newhall Simulation Model (NSM). They explored several strategies for improving soil climate estimates within soil taxonomy schemes and recommended simulation modeling as one viable approach. Other recommendations included a greater reliance on physical measurement and possible modifications of taxonomic definitions based directly on matric potential measurements taken over time rather than soil moisture control sections.

The NSM is a software tool designed to integrate monthly atmospheric climate data into information relevant to soil classification categories by simulating soil moisture and temperature data for calendar days (Van Wambeke, 1982, 1986; Smith, 1986; Newhall and Berdanier, 1996; Jeutong et al., 2000; Yamoah et al., 2003). The NSM was originally developed by Guy Smith and Franklin Newhall in 1972 (Newhall and Berdanier, 1996) and has been used by the U.S. National Cooperative Soil Survey to simulate soil climate for weather stations in soil survey areas (Smith, 1986; USDA-SCS, 1994; Van Wambeke, 1986). The most recent map of soil climate covering the continental U.S., *Soil Climate Regimes of the United States* (USDA-SCS, 1994), relied on the NSM to support mapping of soil moisture and temperature regimes. The NSM has been used in the U.S. and internationally in studies of soil taxonomy, responses of crops to weather, and yield predictions (Bonfante et al., 2011; Van Wambeke, 1982; Jeutong et al., 2000; Yamoah et al., 2003; Costantini et al., 2002; Waltman et al., 2011).

The NSM is considered a mesoscale model. Because NSM assumes precipitation excess exits the soil as runoff or as deep percolation, resulting soil moisture estimates are only valid for well-drained soils associated with relatively level landscapes. The model lacks a runoff/ponding subroutine and functions on a calendar year rather than hydrological year with no carryover from the previous year. It does not account for snowmelt and also lacks a mechanism for accounting for antecedent moisture conditions. In spite of these limitations, it is widely believed that in most cases the NSM provides a reasonable approximation of soil moisture (number of days moist, days dry) and temperature (number of days $<5^oC$ to $>8^oC$) on a monthly time-step. NSM does not require intensive, serially complete daily weather data, but rather monthly summary data of atmospheric precipitation and air temperature. Such input data is readily available for remote areas of the U.S. and many parts of the world. By contrast, field scale models are often more computationally complex and generally require additional measurements of wind speed, solar radiation, relative humidity, cropping, and other parameters in their evapotranspiration subroutines that are not easily acquired across broad geographic regions or remote mountainous landscape settings, and over long-term temporal records (Costantini et al., 2002; Williams et al., 1989). The NSM generates a mesoscale approximation of soil climate that is applicable to soil survey and taxonomic classification (Smith, 1986).

The NSM can be compared to similar process models, but it retains features uniquely suited to taxonomic classification of soil climate. While other models generate inferences of soil moisture and temperature parameters from climate records, such as the field scale models EPIC and CENTURY (Costantini et al., 2002; Williams et al., 1989), the NSM couples water balance calculations more directly with available water-holding capacity and gives output of predicted soil taxonomic classes. The soil moisture calendar output from NSM defines the days that a soil's moisture control section is moist, partly moist and partly dry, or dry within the context of soil temperature thresholds at 5°C and 8°C. These thresholds are needed for taxonomic classification of soil moisture and temperature regimes. The soil moisture and temperature calendars given by the NSM are used to assign taxonomic classes according to US Soil Taxonomy (USDA, 1999).

The scope of this paper is to apply the NSM to gridded raster atmospheric climate data sets using a Grid Element Newhall Simulation (GEN) methodology. Previous applications of the NSM have been limited to point observations of climate station data for a select number of climate data stations rather than gridded data sets covering complete geographic areas (USDA-SCS, 1994). The GEN methodology involves application of NSM to raster datasets for mapping soil moisture and offers several advantages. It gives complete geographic coverage of model output, rather than output for individual point observations, so that mapped patterns of NSM results can be visualized and analyzed geographically for a given period of climatic record. It operates independently of expert knowledge used in the soil climate mapping process and creates a more quantitative output that is not influenced by individual bias due to experiences limited to particular soil regions. While some regional knowledge is likely lost in the broad application of a simple model, greater efficiency and transparency is gained. Finally, multiple iterations of soil moisture output given differing input values of atmospheric climate are possible. Rich data sets of atmospheric climate change can be run through the model, making the tool useful for studying soil change.

The objective in this study is to introduce the GEN methodology and to determine whether spatially gridded geographic modeling of soil moisture regimes with the NSM using raster climate data better predicts NSM model moisture output for weather stations compared to a map made with more traditional expert knowledge methods. Two maps will be compared to each other and to weather station output. The two maps are: *Soil Climate Regimes of the United States* (USDA-SCS, 1994), and the digital output presented here (GEN-CONUS).

1.2 Methods and Materials

1.2.1 Software and Areal Estimates

Mapping tasks were carried out using System for Automated Geoscientific Analysis (SAGA) software version 2.0.7 (SAGA, 2012) and ArcGIS 10 software (ESRI, 2012). All areal estimates were made using Albers Equal Area projection parameters. Higher resolution map layers were resampled to the common ½ arc minute of geographic degree (approximately 800 m resolution in the projected condition) for the full extent of the Conterminous United States. All vector (polygons and point location) map products were projected and rasterized to the common target 800 m resolution in Albers Equal Area projection.

1.2.2 The Newhall Simulation Model (NSM)

The NSM, Java version 1.5.1 is an updated version of the original Newhall Simulation Model developed by Franklin Newhall and Guy Smith in 1972 (Newhall and

Berdanier, 1996; USDA-NRCS, 2011; Waltman et al., 2011; Waltman et al., 2012). The mechanics of the model were not changed in the Java version, but internal calculations and software architecture were updated and made more efficient. The NSM was used to simulate seasonal water balance patterns and calendars for soil moisture in the calculated soil moisture control section in three categories (moist, partly moist and partly dry, and dry) and temperature (number of days <5^{o}C, 5 to 8^{o}C, and >8^{o}C), defining taxonomic windows of soil climate regimes.

The mechanics of the NSM are briefly described here, but more detail can be found in Van Wambeke (1986) and Newhall and Berdanier (1996). In the NSM, the soil is assumed to behave as a reservoir with a fixed capacity determined by its water holding capacity. Water was added by precipitation (Newhall and Berdanier, 1996). Water in excess of retention capacity was assumed to exit the soil as runoff or deep leaching. Stored water was removed by evapotranspiration using Thornthwaite's formula (1948). The soil was divided into segments of 25 mm of water retention difference to the depth of the available water holding capacity. It was then divided into 8 segments, each representing 3.13 mm of water retention difference. The moisture retention was assumed to range from 33 kPa, when all segments are filled, to 1500 kPa or dryer, when all segments are empty. The time step for the model was 360 days per year, with each month given equal influence of 30 days. Monthly precipitation was simulated in light precipitation events and heavy precipitation events. Light precipitation was assumed to account for half of the monthly precipitation in the first half of the month. Total monthly potential evapotranspiration using Thornthwaite's formula (Thornthwaite,

1948) was subtracted from light precipitation to give net moisture activity (NMA). If the resulting value was positive the depth increments were filled, starting at the top of the soil column with half of the NMA. If negative, half of the NMA was applied to the soil column to exhaust the filled segments by diagonal removals called slants, starting with the lowest slant number. Slants were conceptualized as zones of moisture removal oriented diagonally at 45 degree angles from lower soil horizons toward surface horizons. Lower slants are closer to the lower soil horizons and higher slants are closer to the surface. Removal by consecutive slants, starting with lower slants and continuing to higher slants, required greater amounts of potential evapotranspiration units to remove water as the soil became dryer. Next, heavy precipitation was assumed to account for half of the monthly precipitation in the second half of the month. Heavy precipitation was applied to fill available segments by depth increments, and was not subject to evapotranspiration before being absorbed by the soil. The moisture control section was defined in Soil Taxonomy as having an upper boundary the depth to which a dry soil is moistened by 2.5 cm of water moving downward from the surface in 24 hours and a lower boundary as the depth to which a dry soil will be moistened by 7.5 cm of water within 48 hours (USDA, 1999). In the NSM this zone was approximated by the depths of the cumulative water retention difference of 25 and 75 mm (Newhall and Berdanier, 1996). For each moisture state generated (number of segments either wet or dry), the NSM classified the moisture control section either dry in all parts, dry in some parts and moist in other parts, or moist in all parts, for each day of the yearly analysis. An annual calendar of days moist, moist/dry, and dry was generated to make the final determination of the soil moisture classification. This process was iterated for each of the approximately 12 million grid cells for the PRISM-STATSGO2 data set described below to create the output map (GEN-CONUS) examined in this paper. It was also used to classify soil climate from historical records from weather stations, as described below.

While the scope of this paper was limited to examination of moisture regimes, the NSM also estimates temperature regimes and bioclimatic indicators.

1.2.3 Weather Input Data

The 30-year Normals of monthly precipitation and air temperature data were extracted from the weather stations of the National Weather Service(NWS) Cooperative Network (NCDC, 2012a; NCDC, 2012b) that were serially complete for the periods of 1961-1990 and 1971-2000. (The term “normal” refers to a year in which the value for precipitation or temperature is plus or minus one standard deviation of the long-term mean annual value (USDA, 1999).) The total number of weather stations was 4,221 and 5,032 for 1961-1990 and 1971-2000 periods, respectively and comprised the first two data sets for which the NSM was run. The weather stations are distributed throughout the U.S. and are denser in the eastern part of the country compared to the west part (Figure 1.3) for the two periods of normals that were used in the validation process. The raster data set used in the GEN methodology was from the Parameter Regression on Independent Slopes model (PRISM, PRISM Climate Group, 2012; Di Luzio, et al., 2008) and included 30-year monthly precipitation and air temperature estimates provided in a raster format at a resolution of ½ arc minute of degree for the climate record of 1971 to 2000.

1.2.4 Other Model Inputs

Other input data consisted of root-zone Available Water Holding Capacity (AWHC) from the USDA-NRCS digital general soil map of the U.S. (STATSGO2) soil database in a raster format (1:250,000 scale; 250 m resolution grid) (USDA-NRCS, 2011; USDA-NRCS, 2007) and elevation from the shuttle radar topography mission (SRTM) data set (CGIAR, 2011).

Elevation data were resampled from the native SRTM resolution to match the resolution of PRISM climate data inputs using bilinear convolution so that one elevation estimate was available for each climate data raster cell. Elevation was not used to calculate soil moisture in the GEN-CONUS output, but was collected as metadata in the operational database for future runs if temperature lapse rates due to elevation are required. In the PRISM datasets lapse rates were already accounted for, so calculating them again in the GEN-CONUS output would be redundant.

The AWHC data layer was derived from effective rooting depth AWHC of the whole soil adjusted for rock fragments. The calculation of AWHC reflects particle size distribution, organic matter, depth to root restricting layer, salt content, and bulk density. Miscellaneous land types and areas with zero values for AWHC were assumed to be non-soil in the model runs and were excluded from geographic analysis. This occurred in areas with water bodies, rock outcrop and badlands, urban lands, and other non-soil areas.

Each model run used model default values for the offset between mean annual air temperature and mean annual soil temperature of 2.5° C.

1.2.5 The GEN Methodology

In the Grid Element Newhall (GEN) Methodology a geographic area was segmented by a regular coordinate grid system. Each cell of the grid system provided a reference area in which data from all the input GIS layers were queried. The dataset for each grid cell element was then populated with values for each of the inputs required by the NSM for which spatially referenced data exist. These include monthly temperature and precipitation values, AWHC, latitude, longitude, and elevation (Figure 1.1). The NSM was then run individually for each grid cell and model outputs were then aggregated and classed for a thematic map.

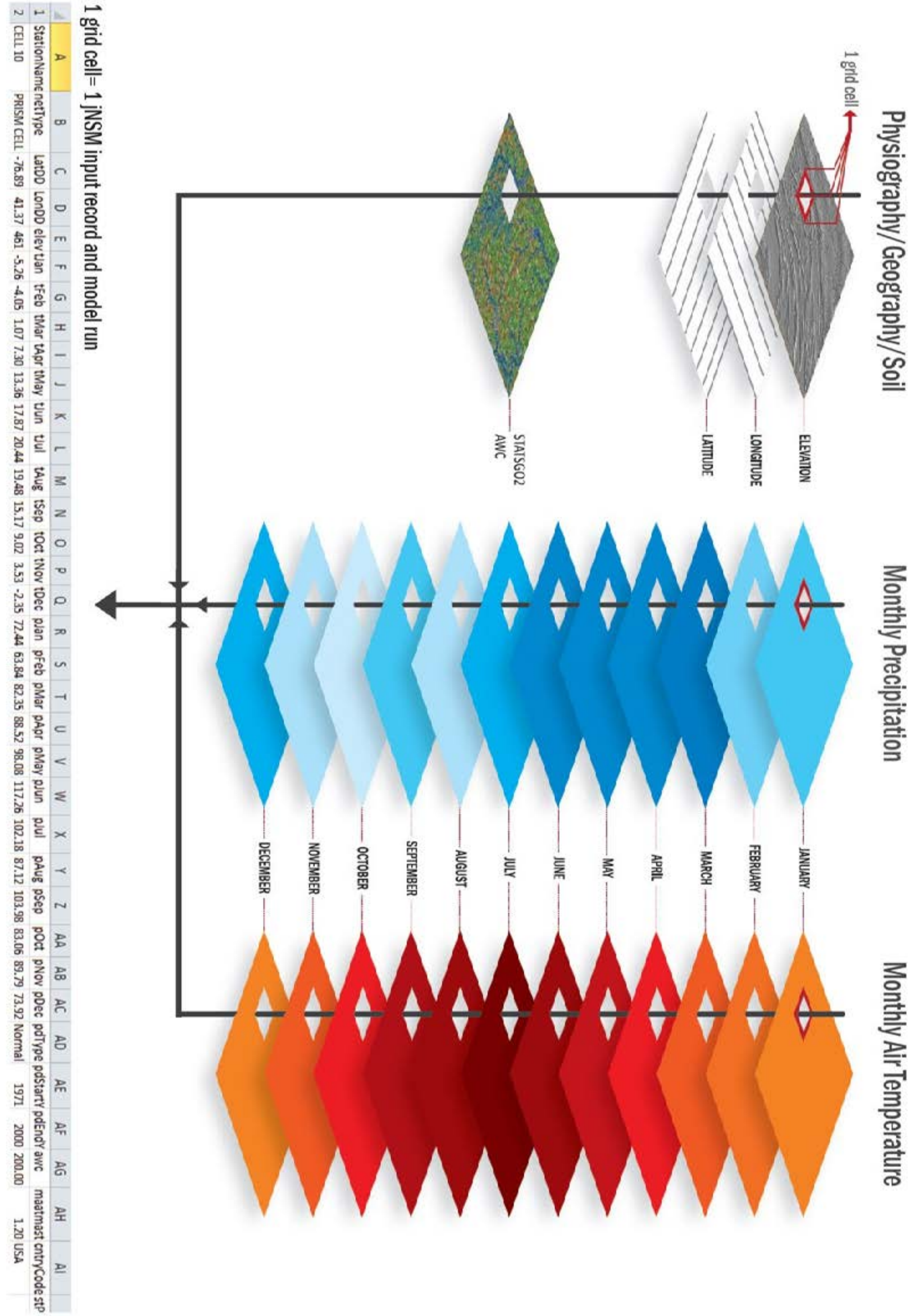

Figure 1.1. The Grid Element Newhall Simulation Model (GEN) methodology for creating maps of soil moisture regimes using the Newhall Simulation Model (NSM).

Monthly precipitation and mean air temperature normals are taken from the input raster data sets. Each grid cell of the ½ arc minute conterminous U.S. represents one model run. The model is run on consecutive grid cells until the geographic area of interest is covered.

1.2.6 The Grid Element Newhall Conterminous U.S. Moisture (GEN-CONUS) Map

The GEN methodology was run on each geographic ½ arc minute of degree of the conterminous U.S. with monthly total precipitation and mean monthly air temperature derived from PRISM data, a total of 12,114,036 model runs. Each model run consisted of the following inputs: 12 monthly air temperature rasters, 12 monthly precipitation rasters, AWHC estimates, elevation, latitude, and longitude. Output analyzed consisted of soil moisture regime classes for each of the 12,114,036 map pixels. The monthly air temperature and precipitation rasters (PRISM, 2011) represented 30-year normal values for the climate period of 1971 to 2000. The minimum map delineation was limited by the pixel resolution of ½ minute of arc, or about 800 m. The output is referred to as the Grid Element Newhall Conterminous U.S. Moisture (GEN-CONUS) map.

1.2.7 The 1994 Map of Soil Moisture Regimes

To test the hypothesis that the GEN methodology was useful for mapping soil moisture regimes for the U.S., the GEN-CONUS moisture regime map was compared to an existing analog soil climate map. The soil moisture regime portion of the Soil Climate map of the Conterminous U.S. (USDA-SCS, 1994) served as our reference data layer for geospatial analysis (Figure 1.2). This map is referred to as the 1994 map or analog map.

The USDA-SCS (1994) methodology for production of the 1994 map included manual delineations of the interpolated soil climate regimes on 1:1,000,000 topographic base maps from a collection of sources, such as hardcopy soil surveys, STATSGO maps, and a small sampling of Newhall Simulation Model runs for selected weather stations for an unspecified period of record. The documentation for the mapping product does not indicate how many weather stations or what period of atmospheric climate data were run through the NSM. Lines placed by the expert map makers were aided by visual inspection of assumed climate covariates (USDA-SCS, 1994). Procedures varied by state, with personnel in some states placing boundaries according to dominant vegetation maps, some according to previous maps of soil climate, and some based on soil temperature studies (USDA-SCS, 1994). The minimum delineation of the map is reported as 2,266 km^2.

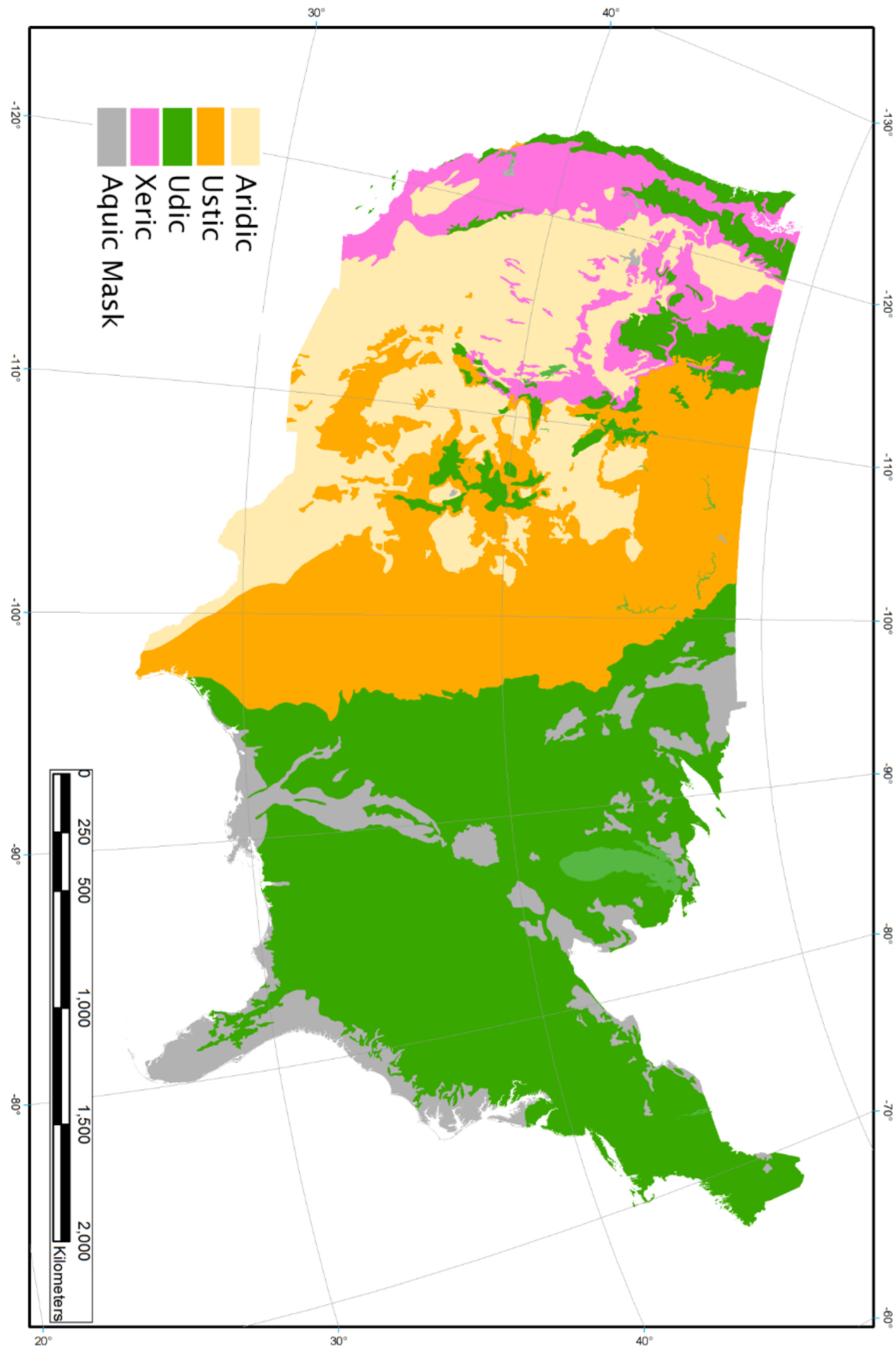

Figure 1.2. The *Soil Climate Regimes of the U.S.* (USDA-SCS, 1994) map represents a

traditional approach, aggregating from county and STATSGO2-level information, with a small sample of individual weather station runs of the Newhall Simulation Model, and combining these outputs with *ad hoc* regional expert knowledge or assumed rules. This visualization of the 1994 map shows only the moisture regimes, not moisture subgroups or temperature regimes.

1.2.8 NSM Runs for Weather Stations for 1961 to 1990 and 1971 to 2000 Periods

The NSM was run for all weather stations with complete data for two periods of interest, 1961-1990, and 1971-2000 (NCDC, 2012a; NCDC, 2012b). The first 30-year period represents the climate normal data that presumably would have been most influential in the development of the 1994 map (USDA-SCS, 1994). The second climate normal data set, 1971-2000, was used in the development of the PRISM data set (PRISM, 2011) that served as the input data for the GEN-CONUS map. The climate station locations, their ecological regions (USEPA, 2016), and the output of the soil moisture regimes in the GEN-CONUS map relative to ecological regions are shown in figure 1.3.

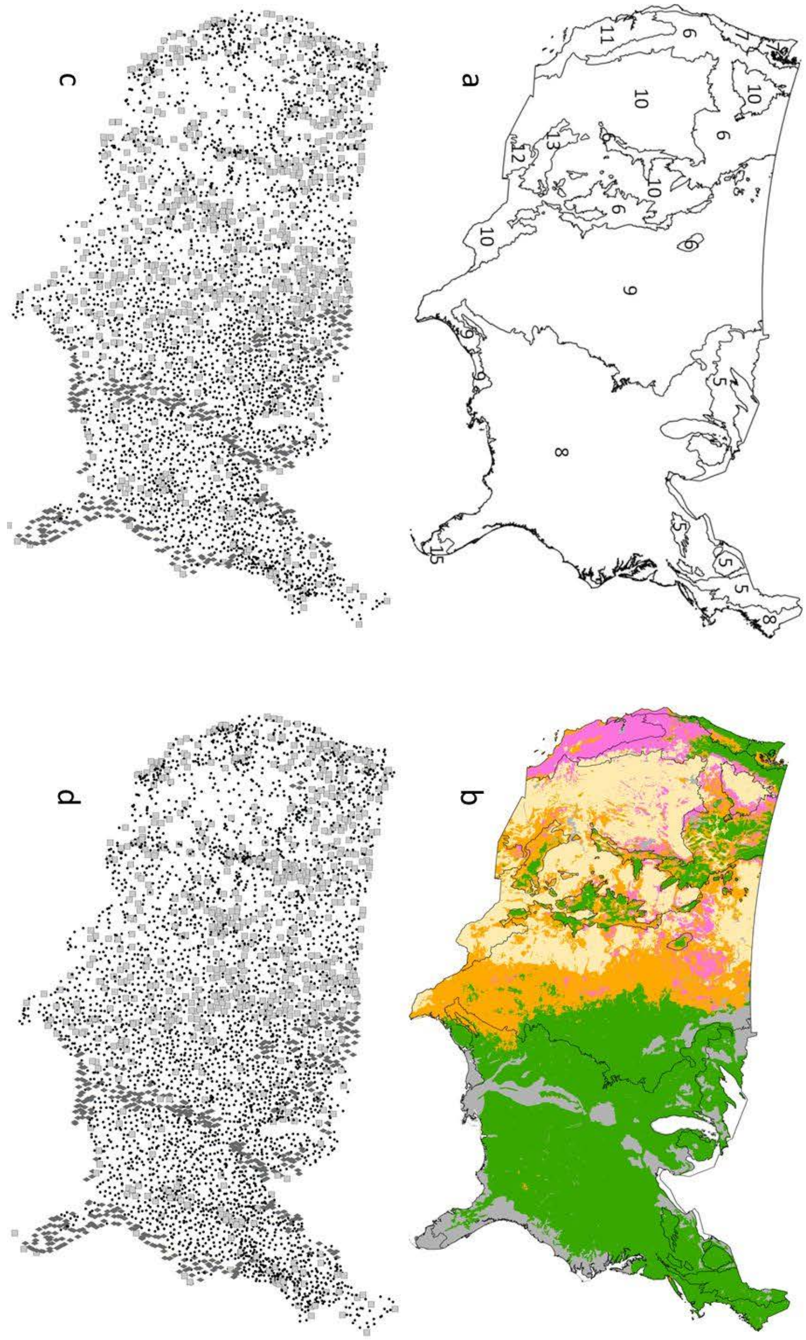

Figure 1.3. a) The ecological regions given in USEPA (2016): 5 – Northern Forests, 6 – Northwestern Forested Mountains, 7 – Marine West Coast, 8 – Eastern Temperate Forests, 9 – Great Plains, 10 – North American Deserts, 11 – Mediterranean California, 12 – Southern Semiarid Highlands, 13 – Temperate Sierras, 15 – Tropical Wet Forests; b) The GEN-CONUS output of soil moisture regimes (with the same color key as figure 1.2) with ecological zones; c) NCDC climate station data locations, 1961- 1990 ; and d) NCDC climate station data locations 1971-2000. In c) and d) black dots indicate climate stations with model output that matched the GEN-CONUS map; grey squares had output that did not match; and gray diamonds are stations on soils that were classed as Aquic in the SCS 1994 map.

1.2.9 Map Comparisons

To highlight differences in soil moisture regimes between the map produced by the GEN-CONUS methodology and 1994 map, both maps were overlain and areas of agreement were tabulated and analyzed using contingency table analysis techniques (Cohen, 1968). Cross tabulations of land area geographic extents of major taxonomic moisture regimes between the 1994 map and the GEN-CONUS output map were stratified by level 1 North American Ecoregions (USEPA, 2016) in order to facilitate understanding of areas where the two maps may differ to greater or lesser degrees. These ecoregions for North America divide the conterminous U.S. into 10 major ecological zones. This was done to highlight geographic patterns of major differences in the interpretation of soil moisture regimes.

Cross tabulation of map areas was conducted after conversion of both maps to Albers equal area projection. Cross tabulation consists of examining two maps in the following way. The area of the first map classified with the first classification category is compared to the classification found in the second map for the same area. Total area of each classification category in the second map that falls within the area of the first classification category of the first map is summed. This is done for each classification category in the first map, until a list is populated with the total area of each classification category of the first map and all the classifications categories from the second map that geographically intersect the first map.

Total land area for each polygon of each soil moisture regime was then calculated for the two maps. Land area classified into the 4 major soil moisture regimes for the two maps was cross-tabulated. Because output from the NSM is largely constrained to well-drained upland soils without perched or permanent water tables, areas that were delineated as Aquic in the 1994 map were assumed to be Aquic, and were excluded from the comparison with GEN-CONUS. This exclusion ensured that soil landscapes that were dominantly driven by groundwater flow were not included in the spatial comparisons.

1.2.10 Climate Station NSM Model Run Comparisons

Part of the difference between the two maps may result from the differences in atmospheric climate for the time period immediately preceding the publication of the 1994 map and the 1971-2000 data set used to create GEN-CONUS. To test the extent to which such temporal climate change may have influenced map output, we analyzed two sets of climate station NSM output. Presuming that the best available data for the production of the 1994 map would have been the weather station data encompassing 1961-1990, we used this set of climate normal as a proxy for a ground-truth for atmospheric climate for the period. Likewise we used the 1971-2000 climate normal as a proxy for a ground-truth relative to the GEN-CONUS map. We used contingency table analysis to compare the two sets of normals to determine whether climate differences may account for a large portion of the difference between the maps.

1.2.11 Climate Change Illustration

To illustrate the utility of the GEN methodology for soil climate forecasting, we ran a soil climate change simulation using the A1B scenario for climate change in the year 2080 from the Special Report on Emissions Scenarios (IPCC, 2000), for a portion of the Rocky Mountains. The A1B scenario is characterized by rapid industrial and population growth and a balanced emphasis on multiple energy sources. Of the multiple climate change scenarios published in the report this scenario predicted the median amount of change for 2080. We used the Hadley Centre Coupled Model, Version 3 (Gordon et al., 2000; Ramirez-Villegas and Jarvis, 2010; CIAT, 2012) with the A1B, because it represents a median amount of change compared to other scenarios.

These data are available in raster format from CIAT (2012) at the same spatial resolution as the PRISM data described in the GEN-CONUS described above. Twenty-four rasters representing 12 monthly air temperature values and 12 monthly precipitation values expected in 2080 were obtained from CIAT (2012) and used for the climate change scenario. The GEN methodology was applied assuming that soil properties and elevation values used in NSM would be consistent with the STATSGO2 values. The GEN methodology for the 2080 scenario is identical to the methodology described for the GEN-CONUS project, except that the future climate change scenario from CIAT (2012) was used as input rather than the raster dataset from PRISM. All other inputs were the same. The climate change scenario is referred to as the GEN2080

We applied the GEN methodology to Major Land Resource Areas 34A, 34B, 48A, 48B, and 51 in the Rocky Mountain Range and Forest and the Western Range and Irrigated Region Land Resource Regions (USDA-NRCS, 2002). This area was chosen because of its high elevation contrast (1,200 to 4,300 m) and diversity of soil climate regimes over complex relief. This illustration was undertaken to show spatial patterns between elevation and the GEN-CONUS output, the 1994, and the GEN2080.

1.2.12 Statistical Analysis

Contingency table analysis is a technique for finding dependence structures among multivariate categorical variables within two populations. First, the frequency distributions of all variables are displayed in tabular form, with the first population displayed along the x direction and the second along the y direction. Then, for each category in the first population, the frequency distributions of occurrence of all other populations from the second population are summed in sequence. The result is a table that summarizes all combinations of categorical cross tabulation. The tabular results are then analyzed using measures of association, such as Cohen’s kappa and percentage agreement.

Statistical analysis included the application of Cohen's kappa coefficient for determining the degree of model agreement between map predictions of soil moisture regime class for unit of land area, and between map predictions and NSM runs using climate station data for specific point locations of climate stations. Cohen's kappa is given as

$$K = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \tag{1}$$

where Pr(a) is the agreement among maps where the number of instances (here, map pixels that are categorized identically) of agreement between the two maps is divided by the total number of instances of observations (the total number of map pixels), and Pr(e) is the probability of random agreement given as

$$\Pr(e) = (\sum \frac{n_i}{n} + \frac{m_i}{n})$$

[2]

where $\Pr(e)$ is the probability of chance agreement assuming random selection, n_i is the number of instances of the second population matching the i th category in the first population and m_i is the number of instances of the first population matching the i th category of the second population, and n is the land area of the total population of map pixels (Cohen, 1968). Because kappa includes the probability of agreement occurring by chance, it is considered more robust than simple percent agreement.

1.3 Results and Discussion

General patterns in the GEN-CONUS map are roughly similar to those of the 1994 map, but there are important differences (Figure 1.4). The overall agreement between the 1994 map and the GEN-CONUS map is 75.6% of land area, with a kappa agreement of 0.642 (Table 1). Differences between the two maps are strongest in the Southern Semiarid Highlands (12), with only 45.6 % agreement between the two maps (Table 2). The 1994 map predicts only Aridic and Ustic regimes for this ecological region, while the GEN-CONUS map predicts greater diversity including Aridic, Ustic, Udic, and Xeric moisture regimes. In fact, 51% of the land area in the Aridic region in the 1994 map is predicted to be Ustic in the GEN-CONUS map (Table 2). The ecological region with the highest kappa agreement is the Eastern Temperate Forests (5), with 98.8% agreement and a kappa coefficient of 0.730. This high level of agreement is due in part to the fact that 97% of the non-Aquic land area is classed as Udic in both maps. A sizable area

(16,800 km^2) classed as Ustic in the GEN-CONUS map is classed as Udic in the 1994 map. Much of this area is found in Eastern Texas and may be due to negative summer water balance associated with higher evapotranspiration and relatively lower summer precipitation in the period of record. Differences in the Great Plains (9) ecological region include a higher prevalence of the Aridic moisture regime in the GEN-CONUS output (22% of the land area) than in the 1994 map (11% of the land area), and an overall agreement of 60.7% and kappa of 0.361. Because land areas categorized as Aquic in the 1994 map were excluded from analysis, no analysis was done of the Tropical Wet Forests (15) region in the Southern tip of Florida, as it was entirely covered by the Aquic regime in the 1994 map. The Northern Forests (5) region showed 100% agreement between the two maps with the Aquic moisture regime excluded from analysis, with the Udic moisture regime comprising all 301,796 km^2 of the map area.

Table 1.1. Contingency table showing overall level of agreement between maps of soil moisture regimes for the conterminous United States in land area (km^2) meeting cross class groupings. Soil moisture regime maps compared are the 1994 map and the current Newhall Simulation Model output (GEN-CONUS) given in this paper. Locations mapped Aquic in the 1994 map were excluded from analysis.

	1994 map					
GEN-CONUS map	Aridic	Ustic	Udic	Xeric	Agreement (%)	Kappa
		--- *km*^{2} ---				
Aridic	1,170,456	391,160	23,820	98,587	75.6	0.642
Ustic	290,696	1,058,982	158,521	136,017		
Udic	2,100	349,365	2,989,198	51,941		
Xeric	114,190	129,984	36,278	305,986		

Table 1.2. Contingency table showing level of agreement between maps of soil moisture regimes for each of the 10 major ecological regions of the Conterminous United States in land area (km^2) meeting cross class groupings. Soil moisture maps compared are the SCS 1994 map and the current Newhall Simulation Model output given in this paper. Locations mapped Aquic in the 1994 map were excluded from analysis.

Eastern Temperate Forests (8)

	1994 map					
GEN-CONUS map	Aridic	Ustic	Udic	Xeric	Agreement (%)	Kappa
		--- km^2 ---				
Aridic	-	-	-	-		
Ustic	-	36,381	16,800	-	98.8	0.730
Udic	-	9,312	2,029,531	-		
Xeric	-	0	0	-		

Great Plains (9)

	1994 map					
GEN-	Aridic	Ustic	Udic	Xeric		
		--- km^2 ---				
Aridic	133,879	334,371	4,428	-		
Ustic	80,234	802,498	62,293	-	60.7	0.361
Udic	26	232,591	363,347	-		
Xeric	15,745	106,415	6,527	-		

Marine West Coast (7)

	1994 map					
GEN-	Aridic	Ustic	Udic	Xeric		
		--- km^2 ---				
Aridic	-	-	-	-		
Ustic	-	169	5,116	451	61.5	0.033
Udic	-	-	25,070	973		
Xeric	-	0	9,576	525		

Mediterranean California (11)

	1994 map					
GEN-	Aridic	Ustic	Udic	Xeric		
		--- km^2 ---				
Aridic	1,496	-	-	3,595		
Ustic	8,057	381	854	9,984	75.9	0.151
Udic	-	92	339	2,428		
Xeric	19,063	322	382	138,455		

North American Deserts (10)

	1994 map					
GEN-	Aridic	Ustic	Udic	Xeric		
		--- km^2 ---				
Aridic	998,006	27,461	5,075	84,407		
Ustic	163,483	25,696	13,883	22,882	73.0	0.220
Udic	581	887	821	2,543		
Xeric	66,595	5,076	3,249	44,211		

Northern Forests (5)

	1994 map			
GEN-	Aridic	Ustic	Udic	Xeric

			--- *km²* ---			
Aridic	-	-	-	-		
Ustic	-	-	-	-	100	-
Udic	-	-	301,796	-		
Xeric	-	-	-	-		

Northwestern Forested Mountains (6)

			1994 map			
GEN-	Aridic	Ustic	Udic	Xeric		
			--- *km²* ---			
Aridic	13,363	7,708	627	17,249		
Ustic	13,804	107,529	71,366	78,738	56.8	0.378
Udic	780	84,255	201,965	32,303		
Xeric	12,643	11,405	19,012	137,308		

Southern Semiarid Highlands (12)

			1994 map			
GEN-	Aridic	Ustic	Udic	Xeric		
			--- *km²* ---			
Aridic	14,898	113	-	-		
Ustic	19,860	5,732	-	-	45.6	0.142
Udic	361	539	-	-		
Xeric	3,520	211	-	-		

Temperate Sierras (13)

			1994 map			
GEN-CONUS	Aridic	Ustic	Udic	Xeric		
			--- *km²* ---			
Aridic	4,254	8,816	-	-		
Ustic	11,447	59,539	-	-	55.8	0.032
Udic	561	25,278	-	-		
Xeric	1,298	3,123	-	-		

Tropical Wet Forests (15) (not analyzed, entirely Aquic)

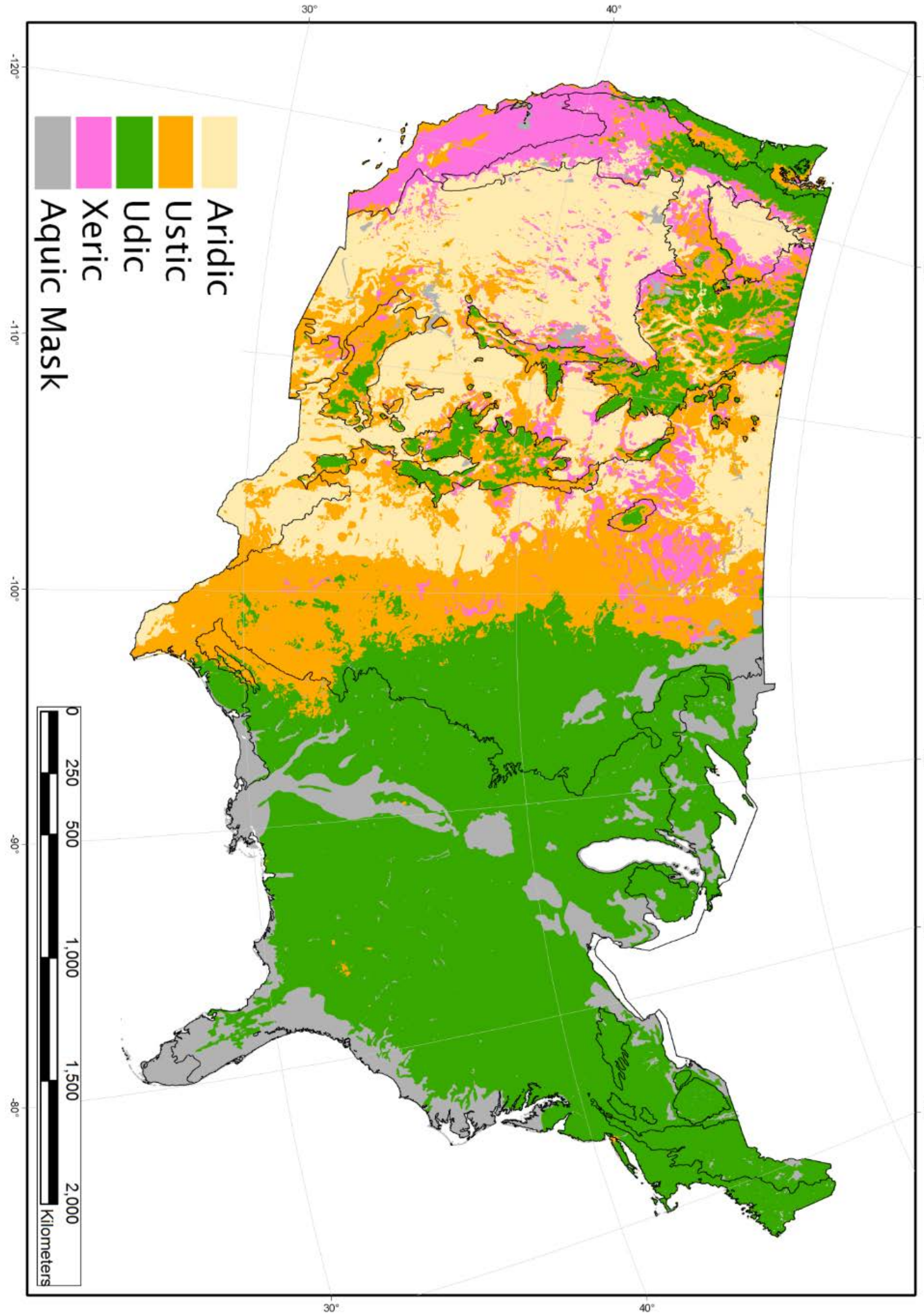

Figure 1.4. Grid Element Newhall Simulation Model (GEN-CONUS) map of soil moisture regimes made from gridded output from PRISM data, STATSGO2 data, and elevation data run through Newhall Simulation Model. The Aquic mask indicates areas that were classified as Aquic in the 1994 map and were excluded from analysis because the NSM does not model the Aquic moisture regime.

Comparisons between NSM output from the climate station locations for the periods 1961-1990 and 1971-2000 showed strong agreement (92%), indicating that differences in modeled soil climate for the two periods were likely relatively minor and that comparisons between the two maps (the 1994 map and the GEN-CONUS map) would not be overly influenced by differences in soil climate when the maps were produced (Table 1.3). The GEN-CONUS map showed agreement of 90.1%, kappa 0.845, with the climate station output for the period 1971-2000. The 1994 map had less agreement with the climate stations, 75.6%, kappa 0.623, for the period immediately preceding map production, 1961-1990. Interestingly, the GEN-CONUS map had higher agreement with the climate stations in 1961-1990 than the 1994 map had, kappa 0.777, even though GEN-CONUS was produced using the PRISM data set that was based on climate from the 1971-2000 period. This indicates that the GEN-CONUS map is a more consistent representation of NSM output than the 1994 map for both periods that were analyzed and that climate differences between the two periods of record were probably not a major factor in the differences between the two maps.

Table 1.3. Contingency table showing overall level of agreement between Newhall Simulation output for climate station data and the 1994 map, and for the station data and the Grid Element NSM (GEN-CONUS) output map. The table shows the number of climate stations in each moisture regime by calculation period meeting cross class groupings.

		NCDC Climate Normals 1961-1990				
NCDC Climate Normals 1971-2000	Aridic	Ustic	Udic	Xeric	Agreement (%)	Kappa
		--- *# stations* ---				
Aridic	816	8	0	5	91.9	0.869
Ustic	99	487	33	24		
Udic	0	52	2161	4		
Xeric	65	39	0	245		
		GEN-CONUS map				
NCDC Climate Normals 1971-2000	Aridic	Ustic	Udic	Xeric		
		--- *# stations* ---				
Aridic	816	51	19	29	90.1	0.845
Ustic	36	631	48	20		
Udic	8	30	2,206	1		
Xeric	156	31	0	234		
		1994 map				
NCDC Climate Normals 1961-1990	Aridic	Ustic	Udic	Xeric		
		--- *# stations* ---				
Aridic	486	309	72	79	75.6	0.623
Ustic	38	375	83	79		
Udic	2	143	1,787	25		
Xeric	25	41	21	190		
		GEN-CONUS map				
NCDC Climate Normals 1961-1990	Aridic	Ustic	Udic	Xeric		
		--- *# stations* ---				
Aridic	688	114	7	71	85.1	0.777
Ustic	19	535	21	47		
Udic	5	230	1,681	0		
Xeric	7	29	1	238		

Figure 1.5a gives a view of output for the Rocky Mountain Range and Forest and the Western Range and Irrigated Region Land Resource Regions SRTM elevation. GEN-CONUS output (Figure 1.5c) shows greater adherence to topographical effects in the region than the 1994 map (Figure 1.5b). In the GEN-CONUS output for the region, 94.7% of the land area above elevation 3,300 m (16,048 km^2) is classified as Udic, while 89% of the land area above elevation 3,300 m (15,135 km^2) is classified as Udic in the 1994 map. Due to topographical orographic effects, these high-elevation areas are in fact more likely Udic than Ustic.

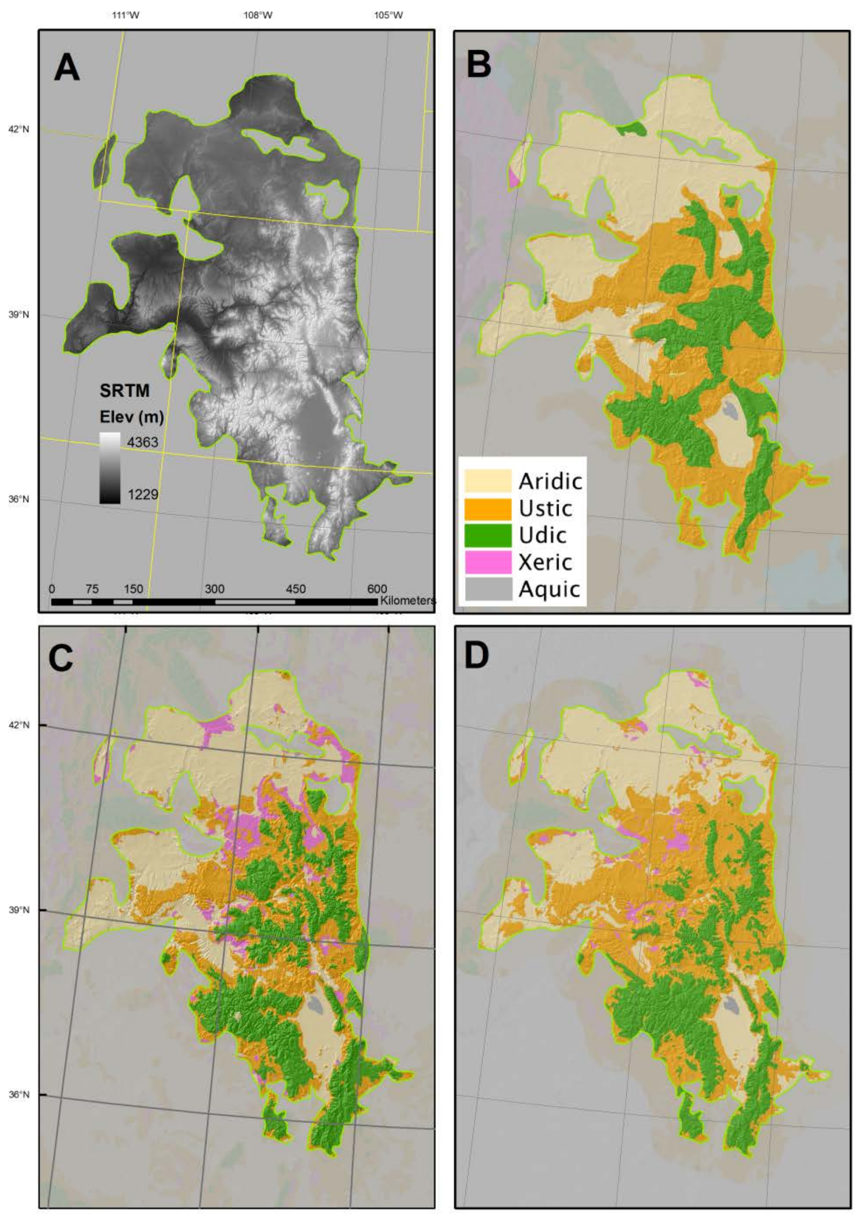

Figure 1.5. Rocky Mountain Range and Forest and the Western Range and Irrigated Region (USDA-NRCS,2002). Upper left: a digital elevation model for the area of interest; upper right, the soil moisture regimes for the region given in the 1994 SCS map; lower left: GEN-CONUS output for the 1971-2000 climate data; lower right: GEN output for the A1B climate change scenario (IPCC, 2000; CIAT, 2012) in the year 2080. The yellow lines in A indicate state boundaries of Wyoming, Colorado, Utah, Arizona, and New Mexico.

Figure 1.5d shows the climate change scenario in 2080 using the GEN methodology. While a direct comparison between the GEN-CONUS map and GEN2080 is not straightforward because the methodologies used to create the input data from the two data sources discussed above differ, the two output maps indicate an increase in the Ustic moisture regime of about 40% of land area and a decrease in the Udic moisture regime of around 12%. With the GEN methodology applied to future climate simulations soil scientists can visualize changes in soil climate. Understanding the change in soil climate may provide better planning for climate change scenarios.

1.4 Conclusion

The most important difference between the 1994 moisture regime map and the GEN-CONUS map is that of methodology. The 1994 map is a static thematic analog map developed by experts working with limited data to convey knowledge about soil climate. The GEN-CONUS map is a systematic application of a soil climate model to an atmospheric data set with no expert intervention. As such, its primary purpose is to clearly display the results of model output. The GEN methodology is repeatable and can be run in different iterations with differing data sets, as the GEN2080 scenario illustrates. The 1994 analog map is a snapshot of what a particular group of experts assembled for the purpose of creating a map accomplished at a given time given the best information available to them. As such, it is not repeatable and no iterations can be run.

Hudson (1992) argued that soil maps created through expert knowledge are often based on understandings of relationships between soil forming factors and soil properties that are not verbally or numerically expressed. He referred to this as the problem of tacit knowledge and claimed that it creates serious inefficiencies in soil survey operations. In mapping projects based on tacit knowledge, the output map itself is often the only published expression of the understanding of the mapping soil scientist. In order to be repeated by new investigators, a large body of understanding has to be built in the mind of each investigator. The GEN methodology, by contrast, is a completely transparent methodology for transforming data of atmospheric climate into soil climate classifications. Such a transparent methodology allows for systematic examination of processes that influence the reliability of eventual output. The task for improving a map becomes not one of increasing the expertise of the map makers who can be relied upon to make improved maps, but one of either improving the assumptions that drive the model that makes the map, or of improving the accuracy of the underlying data.

The GEN methodology offers greater flexibility than the methodology of the 1994 map because it allows for changes in input or model function, such as different scenarios of atmospheric climate inputs, improved estimates of atmospheric climate inputs, and improvements in subroutines of the model. As such it is infinitely iterable. The GEN-CONUS map can be thought of as a visualization of an iteration of the GEN methodology. Such an approach lends itself to multiple iterations with incremental improvement, each of which can result in an output map expressing a current state of a mapping effort. Because expert knowledge is not used in map production, model runs can be easily made whenever improvements might facilitate higher accuracy of mapping. Expert knowledge remains an important component of the process by its assessment of model output and by its application to model improvement.

Because the GEN-CONUS map had a higher level of agreement with the NSM output for climate station normals for both 1961-1990 and 1971-2000, it is a better representation of the results of the Newhall Simulation Model on the specific geographic application of the conterminous U.S. than the map produced by the Soil Conservation Service in 1994. This is not necessarily to say that it is a better map. While adherence to a model that the U.S. soil mapping community has embraced and widely used for 40 years is remarkable, extensive testing of the NSM model against real long-term measures of soil moisture and temperature is needed.

CHAPTER 2. MAPPING SOIL CLIMATE CHANGE WITH DIGITAL SOIL MAPPING USING A GEOGRAPHICAL SOIL CLIMATE SIMULATION MODEL

2.1 Introduction

Soil climate is the long-term record of seasonal and diurnal patterns of moisture and temperature in soil (Brady and Weil, 2001). Soil moisture is a key variable that constrains plant transpiration and photosynthesis and can impact water, energy biogeochemical cycles (Seneviratne et al., 2010). Soil temperature influences evapotranspiration rates, biomass production, chemical and physical weathering of parent materials, biotic activity, and soil organic matter dynamics (Brady and Weil, 2001; Jobbagy and Jackson, 2000). Accelerated climate change driven by increased radiative forcing ongoing in this century, is expected to have profound influence on natural and agricultural systems in decades to come, with long-term repercussions in subsequent centuries (IPCC, 2014; Rosenzweig et al., 2008). These changes will influence plant transpiration and photosynthesis, water, energy and biogeochemical cycles, and land-use. Estimates may be important to assess future seasonal and diurnal patterns of soil moisture and temperature that can be expected under scenarios of climate change (Seneviratne et al., 2010). Soil climate is always changing, but its patterns are predictable to some degree within the context of atmospheric climate and particular soil properties relevant to the status of soil moisture and temperature fluctuations. In this paper we examine future and current soil climate through the Grid Element Newhall Simulation Model (GEN) developed to allow for geographical application of the Newhall Simulation Model to facilitate mapping of soil change (Winzeler et al., 2013). The GEN can be considered a geographical tool used to examine soil climate in scenarios of climate change within a digital soil mapping (DSM) context.

DSM is a set of techniques oriented within geographical information systems (GIS) in which environmental covariates, legacy soil data, and models are used to produce soil map output in systematic and repeatable ways (McBratney et al., 2003). Soil covariates are incorporated into DSM because it has been widely demonstrated that soils vary across geographical space as influenced by spatially distributed soil forming factors (Jenny, 1941). Many of the so-called 5 state factors of soil formation can be represented through environmental covariate datasets. Datasets representing geologic age and composition of parent materials, topography, climate, and organisms are available to be integrated within GIS in a DSM context. One of the goals of DSM is to produce a soil spatial prediction function with spatially autocorrelated errors (McBratney et al., 2003). Tools within DSM include georeferenced global datasets, uniform raster formats, and software tools for integrating disparate datasets into a cohesive whole (Waltman, S.W., 2011; Global workshop on Digital Soil Mapping, 2015).

Soils are currently changing and anticipated to continue to change due to direct effects from climate change and indirectly through their various responses to climate change (Davidson and Janssens, 2006). Climate influences soil properties by governing patterns of moisture and temperature fluctuation in soils as well as influencing soil-forming factors that can drive processes of soil change (Davidson and Janssens, 2006; Cowell and Urban, 2010). Fluctuations of soil moisture and temperature affect soil carbon, primary weathering rates, mineralization and nutrient cycling rates, and oxidation rates (Brady and Weil, 2001). Temperature influences soil carbon decomposition in complex ways, potentially creating positive or negative feedback loops by stimulating both primary productivity as well as microbial decomposition rates of soil carbon sources (Davidson and Janssens, 2006; Lal, 2004). Climate change, due to internal and external forcing mechanisms, is predicted to cause a rise of global surface temperature over the 21st Century (IPCC, 2014). Temperatures in soils are expected to rise globally, with soils undergoing higher rates of potential evapotranspiration and consequently greater water deficits during the growing season in many areas (Cowell and Urban, 2010).

The climate simulation model used by the National Cooperative Soil Survey to support soil mapping efforts was developed in 1972 by Guy Smith and Franklin Newhall and is commonly referred to as the Newhall Simulation Model (NSM) (Newhall and Berdanier, 1996). It was developed as a way to simulate the ways in which atmospheric climate influences soil moisture and temperature conditions. The distinction between atmospheric climate and soil climate is important because soil moisture and temperature are influenced by variables distinct from atmospheric climate such as aspect, topography, snowmelt dynamics, insolation, and soil properties such as organic matter content, particle size, texture, moisture holding capacity, and others. (Smith, 1986). Soil climate classes in the U.S. were developed to accord with observations of natural vegetation and cropping patterns (Smith, 1986). Temperature regime as estimated for the soil in the NSM is valid for the main root zone, estimated to be between a depth of 5 to 100 cm (Soil Survey Staff, 1999). The moisture regime is estimated from the moisture control section (MCS), which is defined as having an upper boundary of the depth to which a dry soil (tension of more than 1500 kPA, but not air-dry) will be wetted by 2.5 cm of precipitation in a 24-hour period and a lower boundary the depth to which the same soil will be wetted by 7.5 cm of precipitation in a 48 hour period (Soil Survey Staff, 1999). The purpose of the development of the MCS was to permit calculation of moisture regimes from the climate record with the NSM (Smith, 1986). The upper limit was chosen such that periods of measured dryness would not be influenced by brief light showers during the dry season in dry landscapes, and the lower limit was arbitrarily selected so as to limit the depth when calculating the moisture status from the soil climate (Smith, 1986)

The NSM is considered a mesoscale model. Because NSM assumes precipitation excess exits the soil as runoff or as deep percolation, resulting soil moisture estimates are valid for well-drained soils associated with relatively level landscapes. The model lacks a runoff/ponding subroutine and functions on a calendar year rather than hydrological year with no carryover from the previous year. It does not account for snowmelt and also lacks a mechanism for accounting for antecedent moisture conditions. In spite of these limitations, it is widely believed that in most cases the NSM provides a reasonable approximation of soil moisture (number of days moist, days dry) and temperature (number of days $<5^oC$ to $>8^oC$) on a monthly time-step. NSM does not require intensive, serially complete daily weather data, but rather monthly summary data of atmospheric precipitation and air temperature. Such input data is readily available for remote areas of the U.S. and many parts of the world, and is useful on raster datasets with complete grid-cell coverage of geographic areas. By contrast, field scale models are often more computationally complex and generally require additional measurements of wind speed, solar radiation, relative humidity, cropping, and other parameters in their evapotranspiration subroutines that are not easily acquired across broad geographic regions or remote mountainous landscape settings, and over long-term temporal records (Costantini et al., 2002; Williams et al., 1989). The NSM generates a mesoscale approximation of soil climate that is applicable to soil survey and taxonomic classification (Smith, 1986). The NSM has been used in the U.S. and internationally in studies of soil taxonomy, soil mapping, responses of crops to weather, and yield predictions (Bonfante et al., 2011; Emadi et al., 2016; Van Wambeke, 1982; Jeutong et al., 2000; Yamoah et al., 2003; Costantini et al., 2002; Waltman et al., 2011).

The NSM was recently updated to Java version 1.6.0, allowing for greater cross-platform versatility of the model (Waltman, 2012). The Grid Element Newhall (GEN) methodology was then developed to allow for updates to soil climate maps coincident with updated and newly available atmospheric climate datasets (Winzeler et al., 2013). GEN represents a continuous coverage pixel-by-pixel application of NSM for continental-scale rasters of soil climate. The GEN methodology is used in this study to obtain soil climate classifications for soil temperature and moisture regimes for different scenarios of climate change. This methodology will allow for analysis of soil change in terms of classification differences in soil moisture and temperature regimes from today's climate to climate after 60 years of climate change under different predicted radiative forcing scenarios.

Atmospheric climate model output from general circulation models is available for download to researchers studying climate change. A leading model for researchers in North America is ModelE2 from the Goddard Institute for Space Studies of the National Aeronautics and Space Administration (Nazarenko et al., 2015). It represents an institutional branch of the international effort to model climate in the Coupled Model Intercomparison Project Phase 5 (CMIP5) with simulations of atmospheric radiative forcing scenarios outlined in the 5th Assessment Report of the IPCC. The CMIP5 datasets for this study are available from the Worldclim.org public interface (Hijmans, et al., 2005; Nazarenko et al., 2015; IPCC, 2014).

GEN can be used to give estimates of soil change by integrating geographically referenced climate prediction data with georeferenced soil information (Winzeler et al., 2013). We use it here in this study on the datasets describing future climate scenarios obtained from NASA in order to examine the ways in which soils can be expected to change with changing climate in the entire Conterminous U.S.A. (CONUS).

2.2 Methods and materials

Mapping tasks were carried out using System for Automated Geoscientific Analysis (SAGA) Software version 2.0.7 (SAGA, 2012) and ArcGIS 10 software (ESRI, 2012). All areal estimates were made using Albers Equal Area projection parameters. Higher resolution map layers were resampled to the common 2.5 arc minute of geographic degree (approximately 4.5 km resolution in the projected condition) for the full extent of the Conterminous United States. All vector (polygons and point location) map products were projected and rasterized to the common target 2.5 km resolution in Albers Equal Area projection.

2.2.1 Grid Element Newhall Simulation Model (GEN)

The GEN methodology is a geographic application of the Newhall Simulation Model. NSM was originally designed to operate on inputs of monthly temperature and precipitation summaries available from discreet weather stations. GEN takes advantage of the availability of datasets representing climate variability across geographic space in regularly spaced grid cells by applying the NSM sequentially to grid cells to represent uninterrupted geographic space. In GEN each grid cell of a continuous raster dataset representing monthly precipitation and temperature data for a given geographic region is coupled with soil information and run through the model to generate output. Inputs to the model include monthly temperature and precipitation values for current and future climate scenarios, available water holding capacity of the soil (AWHC), latitude, longitude, and elevation. Elevation for each grid cell was obtained from the Shuttle Radar Topography Mission (SRTM) data set (CGIAR, 2015). The AWHC data layer was derived from effective rooting depth AWHC of the whole soil adjusted for rock fragments (Waltman, 2011; USDA-NRCS, 2007). The calculation of AWHC reflects particle size distribution, organic matter, depth to root restricting layer, salt content, and bulk density. Miscellaneous land types and areas with zero values for AWHC were assumed to be non-soil in the model runs and were excluded from geographic analysis.

This occurred in areas with water bodies, rock outcrop and badlands, urban lands, and other non-soil areas. The GEN was run for each grid cell and model outputs were aggregated and classed to make thematic maps. The technique allows for model runs to be updated when new model inputs become available.

In the NSM, the soil is assumed to behave as a reservoir with a fixed capacity determined by its water holding capacity. Water was added by precipitation (Newhall and Berdanier, 1996). Water in excess of retention capacity was assumed to exit the soil as runoff or deep leaching. Stored water was removed by evapotranspiration using Thornthwaite's formula (1948). The soil was divided into segments of 25 mm of water retention difference to the depth of the available water holding capacity. It was then divided into 8 segments, each representing 3.13 mm of water retention difference. The moisture retention was assumed to range from 33 kPa, when all segments are filled, to 1500 kPa or dryer, when all segments are empty. The time step for the model was 360 days per year, with each month given equal influence of 30 days. Monthly precipitation was simulated in light precipitation events and heavy precipitation events. Light precipitation was assumed to account for half of the monthly precipitation in the first half of the month. Total monthly potential evapotranspiration was subtracted from light precipitation to give net moisture activity (NMA). If the resulting value was positive the depth increments were filled, starting at the top of the soil column with half of the NMA. If negative, half of the NMA was applied to the soil column to exhaust the filled segments by diagonal removals called slants, starting with the lowest slant number. Slants were conceptualized as zones of moisture removal oriented diagonally at 45 degree angles from lower soil horizons toward surface horizons. Lower slants are closer to the lower soil horizons and higher slants are closer to the surface. Removal by consecutive slants, starting with lower slants and continuing to higher slants, required greater amounts of potential evapotranspiration units to remove water as the soil became dryer. Next, heavy precipitation was assumed to account for half of the monthly precipitation in the second half of the month. Heavy precipitation was applied to fill available segments by depth increments, and was not subject to evapotranspiration before being absorbed by the soil. The moisture control section was defined in Soil Taxonomy as having an upper boundary the depth to which a dry soil is moistened by 2.5 cm of water moving downward from the surface in 24 hours and a lower boundary as the depth to which a dry soil will be moistened by 7.5 cm of water within 48 hours (USDA, 1999). In the NSM this zone was approximated by the depths of the cumulative water retention difference of 25 and 75 mm (Newhall and Berdanier, 1996). For each moisture state generated (number of segments either wet or dry), the NSM classified the moisture control section either dry in all parts, dry in some parts and moist in other parts, or moist in all parts, for each day of the yearly analysis. An annual calendar of days moist, moist/dry, and dry was generated to make the final determination of the soil moisture classification. This process was iterated for each of the approximately 480,000 grid cells for each climate layer available from Worldclim at 2.5 ArcMinutes of spatial resolution (Hijmans et al., 2005).

The categorical delineations among concepts of soil moisture and temperature regimes were developed in the US system of *Soil Taxonomy*, in part, to match observed geographic patterns of cropping and land use management. Smith mentions the use of temperature isotherms 22°, 15°, and 8°C to separate areas suited to the production of citrus, cotton, winter wheat, spring wheat, corn, and small grains (Smith, 1986). Output from the NSM relies on fine temperature and moisture delineations relevant to cropping.

2.2.2 Climate data

GEN methodology was applied to three climate datasets for different climate change scenarios to obtain soil climate classification within the context of US Soil Taxonomy for the entire CONUS.

1) Current conditions (approximate radiative forcing of approximately 2.9 W m^{-2} radiative forcing relative to pre-industrial levels); and
2) Conditions predicted in 2070 under the influence of an increase of 2.6 W m^{-2} radiative forcing relative to pre-industrial levels; and
3) Conditions predicted in 2070 under the influence of an increase of 8.5 W m^{-2} radiative forcing.

Radiative forcing is defined as the rate of energy change at the top of the atmosphere relative to pre-industrial energy levels considered to be the year 1750.

Output obtained included soil temperature and moisture regimes for these three different climate scenarios. Soil changes were analyzed in terms of classification differences in soil moisture and temperature regimes from today's climate to climate after 50+ years of global climate change predicted under the two different radiative forcing scenarios.

Data simulated from Global Circulation Model E for representative concentration pathways from climate projections used in the Fifth Assessment IPCC Report Coupled Model Intercomparison Project 5 (CMIP5) was obtained from WorldClim.org (Hijmans et al., 2005). Model output chosen was the GISS-E2 model from NASA Goddard Institute for Space Studies (Nazarenko et al., 2015) under scenarios representing representative concentration pathways 2.6 W m^{-2} and 8.5 W m^{-2}. These two scenarios were chosen as they represent the higher and lower radiative forcings considered by the IPCC in the 5th assessment document, and can be thought of as lower and higher scenarios for climate change respectively (IPCC, 2014). Year 2070 was chosen as it was considered close enough in the recent future to be relevant to today's land-use decisions and distant enough to show significant effects of climate change.

2.2.3 Analysis

Analysis of output was conducted first by dividing the CONUS region into 20 dominant North American Ecoregions (NAE) representing 20 areas of general similarity in type, quality, and quantity of environmental resources (USEPA, 2016). Total land area falling in each classification category for soil moisture and temperature regime was summed for each NAE in each of the three climate scenarios. Comparisons among radiative forcing scenarios were used to indicate the extent of soil climate change predicted for each NAE. Contrasts for the change in land area extent under each soil and moisture class for each NAE were summarized. The area of change from one moisture regime to another and from one temperature regime to another was summed and characterized.

2.3 Results and Discussion

Soil climate for 2070 was, in general, warmer and less moist than current conditions, with land area classed as Udic declining between 4% - 8% and that classed as Thermic increasing between 6% - 15% (Figures 2.1 -2.3). Land area classed with Ustic, Aridic, and Xeric moisture regimes increased, reflecting decreasing amounts of seasonal soil moisture in summers in future scenarios with higher evapotranspiration rates expected with higher temperatures. Land area of the Cryic, Frigid, and Pergelic moisture regimes decreased markedly in both 2070 scenarios relative to the current conditions with increasing temperature in simulations in 2070. Because the Aquic moisture regime is not handled by the NSM and because it is not known how long a duration of saturation leads to formation of an Aquic Moisture Regime, there is no clear way to determine changes in the large areas of aquic moisture regime in future climate change scenarios. Presumably it could be greater along costs with sea level rise and could increase in inland areas experiencing increased rainfall and decrease in areas predicted to have less rainfall. It is worth noting, however, that the aquic moisture regime is not used as a formative element or even a criterion in taxa in S*oil Taxonomy* (Soil Survey Staff, 1999). Because the NSM relies on an assumption of free drainage and is not able to predict aquic conditions, all output examined is valid only for well-drained soils. The NSM only predicts soil moisture regimes that are used above the series level. (“The formative term “aqua” [when used in suborder designations] refers to aquic conditions, not an aquic moisture regime” – Soil Survey Staff, 1999.)

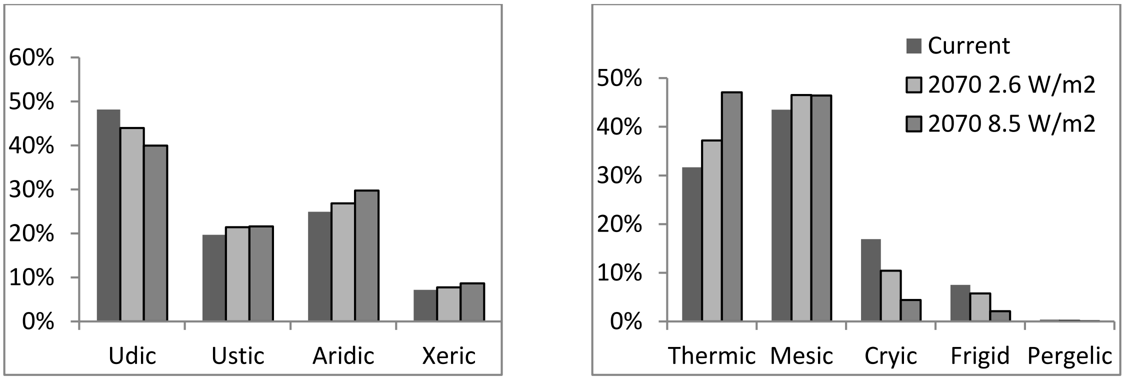

Figure 2.1. Land area for Conterminous U.S.A. moisture regimes (left) and temperature regimes (right) under current conditions, and conditions (Thermic class here also includes hyperthermic) predicted in 2070 under 2.6 W m^{-2} and 8.5 W m^{-2} radiative forcing scenarios from GEN model applied to GISS-E2 model.

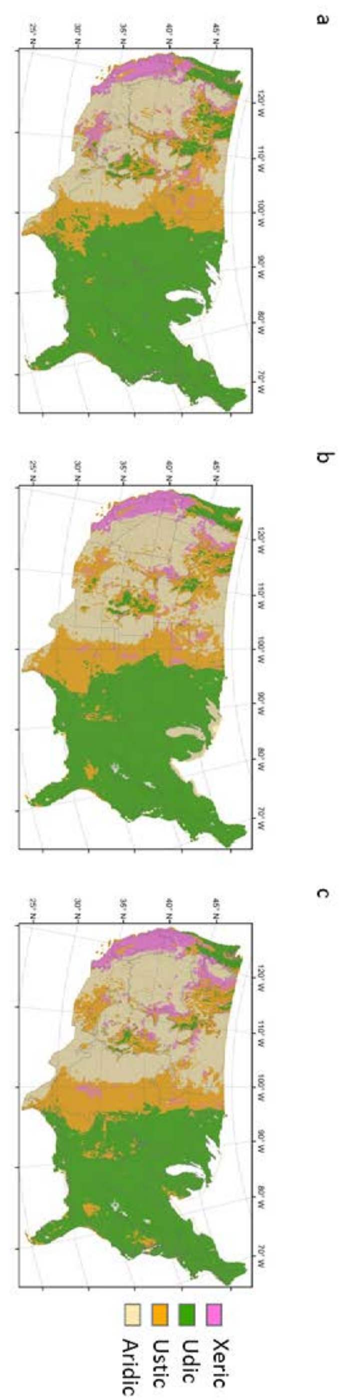

Figure 2.2. Moisture regimes predicted for a) current conditions and for conditions predicted in 2070 under b) 2.6 W m^{-2} and c) 8.5 W m^{-2} radiative forcing scenarios from GEN model applied to GISS-E2 model.

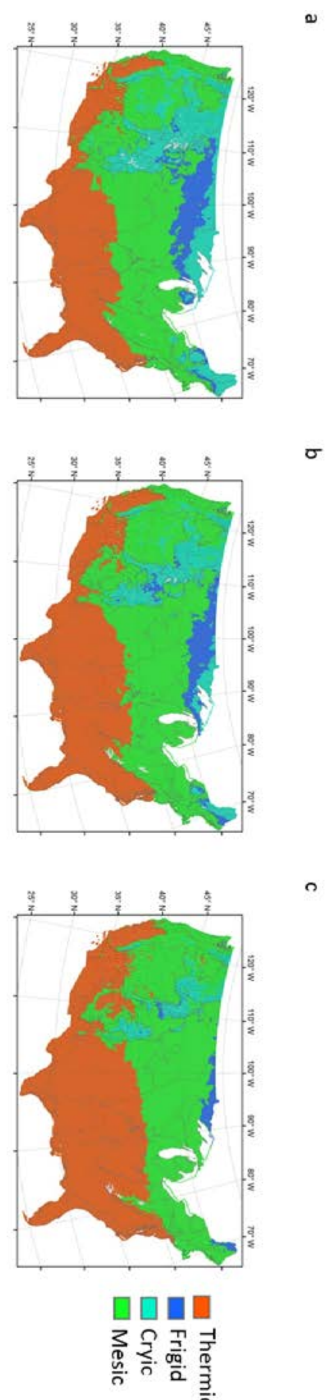

Figure 2.3. Temperature regimes predicted for a) current conditions and for conditions predicted in 2070 under b) 2.6 W m^{-2} and c) 8.5 W m^{-2} radiative forcing scenarios from GEN model applied to GISS-E2 model.

Summary data for moisture and temperature regimes in individual ecological regions indicates decreases in soil moisture associated with greater evapotranspiration in 2070 scenarios relative to current conditions as well as greater soil temperature (Figure 2.4 and Tables 2.1 and 2.2). Mountainous and highland regions such as the Western Cordillera are expected to become warmer and drier, with many elevation-influenced cold regions moving from Frigid and Cryic regimes to Udic or Ustic. Pergelic regimes in this ecological zone are expected to become severely reduced in area as many become Cryic or Frigid, or even Mesic.

Table 2.1. Summary data for soil moisture regimes in individual ecological regions under current conditions, conditions in 2070 under low radiative forcing, and in 2070 under high radiative forcing.

		Current Conditions				Year 2070, Low Radiative Forcing, 2.6 W m^{-2}				Year 2070, High Radiative Forcing, 8.5 W m^{-2}			
Map Code	North American Ecological Zone	Udic	Ustic	Aridic	Xeric	Udic	Ustic	Aridic	Xeric	Udic	Ustic	Aridic	Xeric
		Land area %				Land area %				Land area %			
1	Atlantic highlands	100	0	0	0	100	0	0	0	100	0	0	0
2	Central USA plains	100	0	0	0	100	0	1	0	94	5	1	0
3	Cold Deserts	1	9	87	4	0	10	87	3	0	5	93	1
4	Everglades	88	11	0	0	84	15	0	0	76	23	0	0
5	Marine west coast forest	78	22	0	0	70	28	0	2	65	27	0	7
6	Mediterranean California	0	8	5	87	0	11	3	87	0	7	2	91
7	Mississippi alluvial and Southeast USA coastal plains	96	4	0	0	90	10	0	0	87	13	0	0
8	Mixed wood plains	100	0	0	0	100	0	0	0	99	1	0	0

9	Mixed wood shield	100	0	0	0	100	0	0	0	99	1	0	0
10	Ozark-Ouachina-Appalachian forests	100	0	0	0	99	1	0	0	95	5	0	0
11	South cental semiarid prairies	24	42	34	0	11	49	37	3	3	50	44	4
12	Southeastern USA plains	93	7	0	0	89	11	0	0	85	15	0	0
13	Tamailipas-Texas semiarid plain	0	66	33	1	0	63	37	0	0	34	66	0
14	Temperate prairies	76	21	2	0	76	20	3	1	62	33	5	0
15	Texas-Louisiana coastal plain	61	29	10	0	50	32	18	0	59	22	18	0
16	Upper Gila Mountains	11	43	32	14	8	65	23	4	2	61	31	6
17	Warm Deserts	0	8	83	9	0	17	79	4	0	18	78	4
18	West-Central semiarid prairies	2	66	27	5	0	48	49	2	0	34	66	0
19	Western Cordillera	36	41	15	8	27	45	20	8	19	49	25	8
20	Western Sierra Madre piedmont	0	30	50	20	1	50	38	12	0	60	32	8

Table 2.2. Summary data for soil temperature regimes in individual ecological regions under current conditions, conditions in 2070 under low radiative forcing, and in 2070 under high radiative forcing.

		Current Conditions					Year 2070, Low Radiative Forcing					Year 2070, High Radiative Forcing				
Map Code	NAE	Thermic/ HyperTh.	Mesic	Cryic	Frigid	Pergelic	Thermic/ HyperTh.	Mesic	Cryic	Frigid	Pergelic	Thermic/ HyperTh.	Mesic	Cryic	Frigid	Pergelic
		-- % land area in moisture regime --					-- % land area in moisture regime --					-- % land area in moisture regime --				
1	Atlantic highlands	0	29	59	13	0	0	54	31	15	0	0	88	1	12	0
2	Central USA plains	0	99	0	1	0	0	100	0	0	0	25	75	0	0	0
3	Cold Deserts	1	68	26	5	0	4	77	12	6	0	20	76	2	2	0
4	Everglades	100	0	0	0	0	100	0	0	0	0	100	0	0	0	0
5	Marine west coast forest	0	97	3	0	0	2	96	1	0	0	7	92	0	0	0
6	Mediterranean California	95	5	0	0	0	97	3	0	0	0	99	1	0	0	0
7	Mississippi alluvial and Southeast USA coastal plains	91	9	0	0	0	94	6	0	0	0	99	1	0	0	0
8	Mixed wood plains	0	59	14	27	0	0	87	2	10	0	0	99	0	1	0
9	Mixed wood shield	0	5	80	15	0	0	22	39	40	0	0	77	0	23	0
10	Ozark-Ouachina-Appalachian forests	28	72	0	0	0	48	52	0	0	0	75	25	0	0	0
11	South cental semiarid prairies	51	49	0	0	0	61	39	0	0	0	77	23	0	0	0

12	Southeastern USA plains	75	25	0	0	0	87	13	0	0	0	98	2	0	0	0
13	Tamailipas-Texas semiarid plain	100	0	0	0	0	100	0	0	0	0	100	0	0	0	0
14	Temperate prairies	4	61	12	24	0	8	70	2	20	0	23	72	0	5	0
15	Texas-Louisiana coastal plain	100	0	0	0	0	100	0	0	0	0	100	0	0	0	0
16	Upper Gila Mountains	18	71	12	0	0	23	75	2	0	0	50	50	0	0	0
17	Warm Deserts	93	7	0	0	0	97	3	0	0	0	99	1	0	0	0
18	West-Central semiarid prairies	0	42	14	45	0	0	71	4	24	0	0	98	0	2	0
19	Western Cordillera	0	16	78	1	4	1	25	72	2	1	3	43	48	6	0
20	Western Sierra Madre piedmont	91	9	0	0	0	96	4	0	0	0	99	1	0	0	0

Figure 2.4. North American Ecoregions from USEPA dataset (USEPA, 2016). Key to ecoregions: 1 - Atlantic highlands, 2 - Central USA plains, 3 - Cold Deserts, 4 – Everglades, 5 - Marine west coast forest, 6 - Mediterranean California, 7 - Mississippi alluvial and Southeast USA coastal plains, 8 - Mixed wood plains, 9 - Mixed wood shield, 10 - Ozark-Ouachina-Appalachian forests, 11 - South cental semiarid prairies, 12 - Southeastern USA plains, 13 - Tamailipas-Texas semiarid plain, 14 - Temperate prairies, 15 - Texas-Louisiana coastal plain, 16 - Upper Gila Mountains, 17 - Warm Deserts, 18 - West-Central semiarid prairies, 19 - Western Cordillera, 20 - Western Sierra Madre piedmont. (Hillshade basemap was derived from SRTM data (Jarvis et al., 2008.)

Increasing temperatures during the growing season can be expected to drive higher rates of evapotranspiration, leading to greater water deficits in the 2070 scenarios when compared to the scenarios of current conditions. This accords with findings of other researchers (Seneviratne, et al., 2010), and the need to document soil change and update soil mapping products in the coming decades. The change from the Mesic regime to Thermic and Hyperthermic regimes is the dominant temperature regime change predicted in 2070 under the high radiative forcing estimate (Table 2.3). Under the lower radiative forcing estimate equal changes from the Frigid to the Mesic and from the Mesic to the Thermic/Hyperthermic can be expected (Table 2.3). In both scenarios, land changes from the Cryic to Mesic, Cryic to Frigid, and Pergelic to Cryic are important. Changes in moisture regime in 2070 under both radiative forcing estimates affect Udic regimes changing to Ustic more than other changes, reflecting expected drier conditions during the growing seasons. Changes to Aridic regimes from Udic and from Ustic are also predicted.

Table 2.3. Changes in temperature regimes from current conditions to those in 2070 under low radiative forcing and high radiative forcing scenarios

Change to temperature regime	Climate condition in 2070 2.6 W m^{-2}	8.5 W m^{-2}
	------------ 10,000 km^2	-------------
No change	659	510
Mesic to Thermic/Hyperthermic	43	122
Cryic to Mesic	25	90
Cryic to Frigid	30	18
Frigid to Mesic	44	61
Pergelic to Cryic	2	3
Total fraction of land area changing class	18%	37%

The geographic pattern expected for changes in moisture regime shows far more change in the western part of the CONUS, with changes from moist conditions to more arid conditions in many cases (Figure 2.5). Some limited areas in the arid Southwest are expected to become wetter, particularly under the high radiative forcing estimate when some areas are expected to change from Aridic to Xeric or Ustic (Figure 2.5B).

The geographic changes expected for temperature regimes are perhaps less complex and follow broadly latitudinal patterns. The temperature regime fronts for the Thermic/Udic, the Mesic/Frigid, and the Frigid/Cryic are expected to proceed northerly in latitude in both scenarios. With respect to orography, many mountainous regions are expected to change temperature regime from Pergelic to Cryic or Frigid, from Cryic to Frigid or Mesic, and from Frigid to Mesic.

CHAPTER 3. AN EXPLORATION OF THE APPARENT INFLUENCE OF SOIL CLIMATE ON SOIL PROPERTIES IN GEOGRAPHIC ANALYSIS OF THE NATIONAL COOPERATIVE SOIL SURVEY LABORATORY CHARACTERIZATION DATABASE

3.1 Introduction

Soil climate, the long-term record of seasonal and diurnal patterns of moisture and temperature in soil, is considered a driver of soil development and a key soil-forming factor influencing the variations in soil properties within soil profiles and over geographic spaces (Brady and Weil, 2001; Smith, 1986; Jenny, 1941). Atmospheric climate as it influences soil climate determines the strength and nature of flux factors that induce physical and chemical changes in soils (Buol et al., 1989). These flux factors include evapotranspiration, radiant and atmospheric temperature, winds, precipitation, photosynthesis and carbon cycling, and water flow (Buol et al., 1989). The soil can further be considered as a system on which energy inputs and outputs play an active role in chemical and physical transformations (Runge, 1973; Smeck et al., 1983). These energy inputs from climate, or flux factors, can be quantified through observation of the climatic record with relevant mathematical models (Rasmussen, 2005).

Direct influences of climate on soils include precipitation and temperature. Precipitation influences soils as moisture constrains plant transpiration and photosynthesis and determines water and energy biogeochemical cycles (Seneviratne et al., 2010), increases the amount of the hydrogen ion in soils (Buol et al., 1989), and increases clay content through weathering and clay translocation through the soil profile (Jenny, 1980; Levine and Ciolkosz, 1983). Precipitation can also influence erosion and deposition rates through removals and deposits of surface materials. Finally, precipitation can provide direct inputs into the soil system such as deposition of rain-borne particles, nitrogen, and carbonic acid. Temperature influences plant growth rates and subsequent productivity (Rasmussen et al., 2005), it drives evapotranspiration, and it increases the rate of chemical reactions important for neoformation of clays, transformations of compounds in soils, and decomposition rates of organic matter (Brady and Weil, 2001). Seasonal patterns of precipitation and temperature are important as they determine niches supportive of particular vegetation and cropping regimes and they influence carbon stability. Two geographic areas with the same annual precipitation and temperature, for instance, may have widely different vegetation and soil carbon content due to temporally different patterns and degrees of fluctuation of temperature and moisture throughout the months of the year.

Classic climosequences are studies of soil variability in which attempts to isolate climate as a soil forming factor by selection of sites in which other soil forming factors are relatively constant and climate alone is observed to account for variability have provided much information about the influence of climate on soil properties (Schaetzl and Anderson, 2005). In contrast to the climosequence approach where data are carefully selected to minimize the confounding influence of soil forming factors other than climate, recent investigations have emphasized the use of new technologies to examine vast datasets at a continental scale to explore the relationships between soil properties and climate to include the full range of all soil forming factors along with ample data to tease out relevant relationships through large datasets (Scull, 2009). This new “top-down” approach requires sufficient and extensive sampling across the entire range of soil forming factors to determine the effect of one factor in the face of the full range of variability of the others (Scull, 2009). To inform continental scale studies of soil and climate with this approach, extensive datasets representing measured soil properties and climate are necessary.

One particularly useful dataset for the top-down approach to studying the influence of climate on soil properties is the laboratory data available from the soil characterization database of the National Cooperative Soil Survey (NCSS, 2015). The database contains detailed geographic, pedometric, classification, and laboratory data for 48,586 pedons within the geographical boundaries of the conterminous states of the US (CONUS) and more pedons outside this area. Each pedon has been characterized with standard laboratory testing at samples representing each named horizon down to various depths. The data on these pedons have been collected over decades of research representing soil observations on hundreds of millions of acres of land by the US National Cooperative Soil Survey (NCSS, 2015).

Another important dataset is the detailed record of atmospheric climate of the CONUS available from the PRISM Climate Group (PRISM, 2015). We used the monthly precipitation and temperature estimates for the years 1971 – 2000 to give a picture of the long-term climate. In this dataset the CONUS is divided into a grid with approximately 12 million pixels, each representing a geographic unit of ½ arc minute of degree. Each georeferenced grid cell is assigned an average monthly temperature and precipitation value for the period of interest using the Parameter Regression on Independent Slopes model (PRISM, 2015).

Models have been used to simulate soil moisture and temperature patterns for given sets of atmospheric climate data throughout the history of the NCSS, but more validation and verification of the model outputs is needed (Smith, 1986). A recent advance has been the use of the Newhall Simulation Model (NSM), the most dominant soil climate model used by the NCSS, on georeferenced grid cells representing continuous coverage of climate observations at regularly spaced intervals across the CONUS (Winzeler et al., 2013). This has allowed for simulations of soil climate change in response to radiative forcing scenarios released by the Intergovernmental Panel on Climate Change (IPCC, 2014).

Attempting to evaluate the usefulness of model output of the NSM using direct measurements of soil moisture and temperature over very short episodes of geologic time may not lend sufficient perspective relevant to pedogenesis due to the short-term nature of the observations. In this paper we propose to use soil variables themselves as indicators of long-term climate influences. Several soil properties are influenced by climate. By examining the relationships between these soil properties and the climate variables as given by atmospheric climate records, soil climate simulations, and estimates of net primary productivity derived from atmospheric climate, we seek to examine the utility of output from the NSM.

Relevant indicators of the signature of climate on the soil used in this study are soil pH, clay, carbon, and cation exchange capacity. Soil pH at a depth of 50 cm in the soil profile, we believe, is deep enough to dampen some of the effects of anthropic surface treatments such as lime and other soil amendments, but shallow enough to be fully in the solum and susceptible to the chronic influence of climate. The clay accumulation index (CAI) designed to quantify the degree of argillic horizon development in the B horizon is one indicator of the extent and intensity of climate's influence on pedogenesis (Levine and Ciolkosz, 1983). Likewise the observation of the maximum clay percentage within the pedon to a depth of 2 m may relate to weathering of primary and secondary minerals as well as neoformation of clays relevant to the activity of local conditions of soil climate. Soil carbon and CEC are both influenced by climate through its influence on plant production rates of biomass and microbial decomposition rates. Many other factors such as land management, topography, and parent material play a role in soil carbon dynamics and may confound any simple causal relationships.

We examine soil properties with depth in this study because some soil properties with depth are expected to be more relevant to soil variability as influenced by climate than the properties found on the surface alone. Soils are more easily modified at the surface by tillage, vegetation, liming, or other local influences and more profoundly influenced over time at greater depth by chronic influences such as climate. Some soil properties are best looked at with respect to losses from surface horizons and accumulations at deeper horizons, so looking at changes in properties with depth accomplishes two things. First, it normalizes the effects of disparate parent materials as accumulations and losses in the profile are expressed relative to amounts initially present in the unmodified parent material. Second, it shows the influence of climate drivers as they facilitate horizonation, or change in soil properties with depth.

The Newhall Simulation Model is a convenient tool for examining soil variability with respect to climate variability. It is used to integrate atmospheric climate data into simulations of soil climate using a monthly time step. More detailed descriptions of the model can be found elsewhere (Van Wambeke, 1986; Winzeler et al., 2013). The NSM is considered a mesoscale model. Because NSM assumes precipitation excess exits the soil as runoff or as deep percolation, resulting soil moisture estimates are valid for well-drained soils associated with relatively level landscapes. The model lacks a runoff/ponding subroutine and functions on a calendar year rather than hydrological year with no carryover from the previous year. It does not account for snowmelt and also lacks a mechanism for accounting for antecedent moisture conditions. In spite of these limitations, it is widely believed that in most cases the NSM provides a reasonable approximation of soil moisture (number of days moist, days dry) and temperature (number of days $<5^oC$ to $>8^oC$) on a monthly time-step. NSM does not require intensive, serially complete daily weather data, but rather monthly summary data of atmospheric precipitation and air temperature. Such input data are readily available for remote areas of the U.S. and many parts of the world, and is useful on raster datasets with complete grid-cell coverage of geographic areas. By contrast, field scale models are often more computationally complex and generally require additional measurements of wind speed, solar radiation, relative humidity, cropping, and other parameters in their evapotranspiration subroutines that are not easily acquired across broad geographic regions or remote mountainous landscape settings, and over long-term temporal records (Costantini et al., 2002; Williams et al., 1989). The NSM generates a mesoscale approximation of soil climate that is applicable to soil survey and taxonomic classification (Smith, 1986). The NSM has been used in the U.S. and internationally in studies of soil taxonomy, soil mapping, responses of crops to weather, and yield predictions (Bonfante et al., 2011; Emadi et al., 2016; Van Wambeke, 1982; Jeutong et al., 2000; Yamoah et al., 2003; Costantini et al., 2002; Waltman et al., 2011).

The NSM was recently updated to Java version 1.6.0, allowing for greater cross-platform versatility of the model (Waltman, 2012). The Grid Element Newhall (GEN) methodology was then developed to make allow for updates to soil climate maps coincident with updated and newly available atmospheric climate datasets (Winzeler et al., 2013). GEN represents a continuous coverage pixel-by-pixel application of NSM for continental-scale rasters of soil climate. The GEN methodology is used in this study to obtain soil climate estimates and to assess their relevance in the prediction of soil properties within ordinary least squares and geographically weighted regression models.

The NSM is largely untested with respect to how well the outputs of the model relate to actual soil variability observed over geographic space in widely different geologic regimes of different parent materials. In this paper we test the outputs of the NSM according to their correspondence to soil properties known to vary with climate, and relative to each other with respect to information redundancy and collinearity. It is important to note that testing the relationship between today's climate and today's soil properties necessarily ignores the influence of paleoclimate, which in some cases may be significant. Climates are not constant over pedogenic time, and some soil properties may show influences of paleoclimates that are not accounted for in the record of current climate. Nevertheless, in most cases the current climate has profound influence on pedogenic processes such that the signature of current climate can be traced in the soil properties observed. In addition, datasets of paleoclimate lack the degree of spatial and temporal resolution necessary for NSM model runs and are not easily commensurate to contemporary climate models.

Other model variables representing soil forming factors other than climate were used in this study to eliminate as many confounding effects as possible. These included variables useful for integrating atmospheric climate data to provide energy estimates and estimates of primary plant productivity, and models accounting for the local influence of terrain on soils. The energy estimates and primary productivity estimates were those developed by Rasmussen in his energy accounting approach to quantitative pedogenesis (Rasmussen, 2005). The three terrain variables chosen were determined in other studies to be sufficient for creating a geometric signature needed to accurately classify terrain-unit maps within an automatic terrain classification system (Iwahashi and Pike, 2007). We used the three terrain variables recommended by Iwahashi and Pike to increase the efficiency of our analysis and most fully account for the effects of terrain with the minimum number of variables. By removing effects of terrain, plant productivity, and energy inputs in the regression models, the effects of soil climate on soil properties were clarified and confounding influences were minimized.

The hypothesis for this study is three fold. First, we test whether the influence of climate on soils can be observed by noting changes in soil properties over geographic space coincident with changes in atmospheric climate over the same geographic space. Second, we gauge the validity of the NSM by observing its ability to provide reasonable predictor variables in a multiple linear regression approach used to model soil properties as influenced by climate. Finally, we gauge the strength of the NSM by comparing its ability to provide reasonable predictor variables for strong regression models relating NSM output to soil variables across geographic space to a climate predictor model (CLIM) that uses only atmospheric climate inputs and no moisture and temperature simulations over time.

3.2 Methods and Materials

To test the validity and strength of the NSM as a predictive tool for explaining soil variability across geographic space relative to an observational climate predictor, we obtained data from disparate sources and integrated them into a single GIS database. Data from PRISM provided atmospheric climate estimates of precipitation and temperature in a monthly time step from 1970 – 2000 (PRISM, 2015). Data from the shuttle radar topography mission provided elevation estimates needed for terrain analysis (Jarvis et al., 2008). Shape files representing the physiographic divisions of soil and land use areas, the Major Land Resource Areas (MLRAs), were obtained from the USDA (USDA, 2006). The database containing soil laboratory data and pedon descriptions was downloaded from the National Cooperative Soil Survey (NCSS, 2015). The database contains 48,586 pedons within the geographical boundaries of this study, each analyzed for physical and chemical characteristics at multiple depths according to soil horizon delineations. One potential limitation of this database includes lack of transparency about the choices made regarding soil observations. In most cases it is unclear, for instance, if a detailed soil observation was undertaken to characterize an area, provide a contrasting soil from more typical soil samples for an area, to satisfy technical criteria, or for any other reason. As such, treating the database as a mine of random observations may introduce unknown biases. Nevertheless, the database is uniquely powerful in terms of the extent of the information available, the detailed information for each pedon, its combination of pedological observations and laboratory measurements, and for its overall comprehensive and cohesive characteristics.

For this study Entisols were excluded as they are by definition made up of unaltered parent material below the A horizon; as such they do not reflect a climatic signature with depth in the soil. Histosols were excluded as well, as they do not reflect a climatic signature on a mineral matrix, but rather are made of a build-up of organic material often within a closed-depression context; they are different enough from mineral soils that their inclusion in the models would represent a confounding of some of the relationships sought. Only pedons within the CONUS boundaries, with sufficient depth for all analyses, populated data for the response variables, and developed in non-contrasting parent materials without obvious lithologic discontinuities were chosen for inclusion in this study (Figure 3.1).

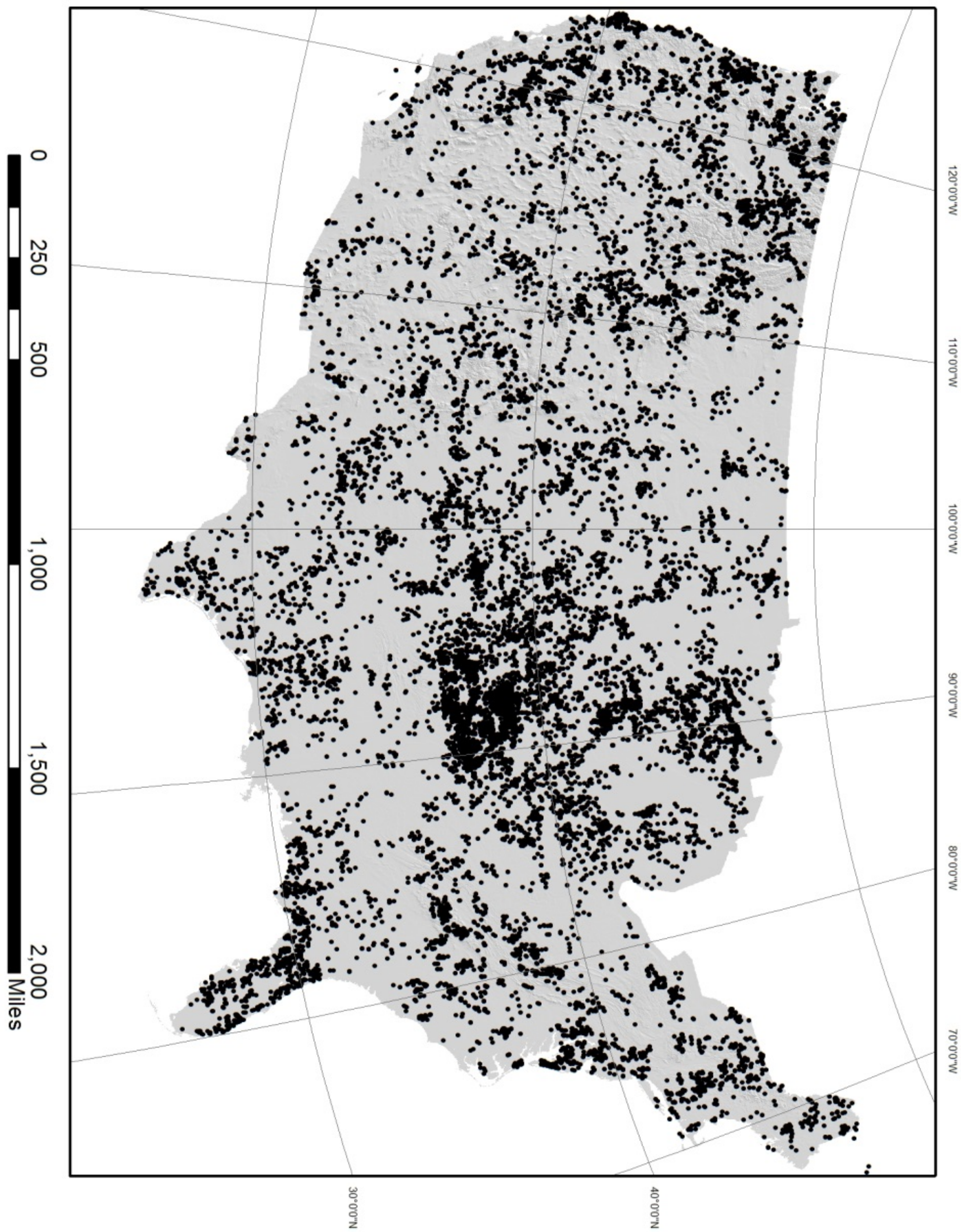

Figure 3.1. Point locations with sufficient data for analyzing pH at 50 cm depth, approximately 22,000 pedons.

Variables analyzed included soil pH at 50 cm depth, the soil clay accumulation index (CAI), soil clay estimate (kg m^{-2} to a depth of 2 m), soil carbon in the upper 25 cm and upper 200 cm (kg m^{-2}), clay maximum (percentage of the mineral fraction by dry weight in the upper 2 m), and cumulative cation exchange capacity (NH_4OAc extraction in the upper 50 cm, 100 cm and 150cm, expressed in mol m^{-2}). The soil CAI represents the degree of argillic horizon development and is calculated with the formula

$$\sum[(B_c - C_c) \times (T)]$$

where B_c is the percentage clay by weight of the mineral fraction < 2mm of the soil in the B horizon(s), C_c is the percentage clay by weight of the mineral fraction < 2mm of the C horizons, and T is the thickness in cm of the B horizons (Levine and Ciolkosz, 1983). Where the CAI is less than 0 it was assumed that a lithologic discontinuity between parent materials of the B and C horizons accounted for the difference. Pedons with reported lithologic discontinuities or with negative CAI values were eliminated from the analysis of CAI.

The NSM was run for each grid cell of the CONUS using the input data described in Winzeler et al. (2013). Output from the NSM includes soil climate classes and the following variables.

- Bio5drycum – the cumulative number of days in the year that the soil moisture control section (SMCS) is simulated to be warmer than 5°C and fully dry (1500 kPa, or greater, of moisture retention)

- Bio5mdcum – the cumulative number of days in the year that the SMCS is simulated to be warmer than 5°C and partially dry and partially moist
- Bio5mstcum – the cumulative number of days in the year that the SMCS is simulated to be warmer than 5°C and fully moist
- Bio8mstcon – the consecutive number of days in the year that the SMCS is simulated to be warmer than 8°C and fully moist
- Smrdrycons – the consecutive number of days in the summer that the SMCS is simulated to be fully dry
- Wtrmscns– the consecutive number of days in the winter that the SMCS is simulated to be fully moist
- Yrdrycum – the cumulative number of days in the year that the SMCS is simulated to be fully dry
- Yrmdcum – the cumulative number of days in the year that the SMCS is simulated to be partially moist and partially dry
- Yrmstcons – the consecutive number of days in the year that the SMCS is simulated to be fully moist
- Yrmstcum – the cumulative number of days in the year that the SMCS is simulated to be fully moist

Rasmussen's energy estimates were processed using the PRISM data to determine energy input into the soil system and to estimate net primary production of vegetation (Rasmussen et al., 2005). These included the following variables.

- $\sum E_{ppti} = (c)(P_{effi})(MAT_i)$, where *i* is the month of P_{eff} (the amount of precipitation greater than potential evapotranspiration), *c* is the specific heat of water, P_{eff} is the mass of water and MAT is the change in temperature from 0°C to the mean monthly air temperature; 12 monthly estimates are summed to give an annual total (Rasmussen et al., 2005)
- $NPP_i = \{3000/[1 + e^{(1.315 - 0.119)(MATi)}]\}$ $(days_i/365\ d\ yr^{-1})$, with the NPP variable representing the amount of plant production during the year (Rasmussen et al., 2005)

For all models three terrain variables were used. These were slope, curvature, and surface roughness (Iwahashi and Pike, 2007). They were calculated from SRTM data at 1 km resolution (Jarvis et al., 2008.)

The experimental unit used for all regression models was the Major Land Resource Areas, 226 total units for the CONUS. Each MLRA represents a unit of similarity with respect to soils, land use, and physiography (USDA, 2006). These were chosen because the underlying data structure of the individual pedons examined precluded the use of un-aggregated data due to the high degree of spatial autocorrelation and clustering among observations. MLRAs are large enough to accommodate large bodies of available pedons, but small enough to represent unique geographic zones of similarity. To avoid unrealistic values, if the MLRA had fewer than 4 soil observations it was omitted from regression models. All explanatory variables within each MLRA were also averaged to obtain a representative value for each variable. These average values were then used in the regression models.

NSM and terrain variables were used in ordinary least squares and geographically weighted regressions to explain variability of soil properties related to soil climate (NSM model). The complete model run included all variables. Many of these variables are correlated, leading to multicollinearity problems in the models. To remove multicollinearity, each model was run in a stepwise progression after removing any term with a variance inflation factor >7.5 (ArcGIS, 2015). After variance inflation was eliminated, terms with p-values > 0.05 were removed in the same stepwise manner. Final models were checked with Akaike's Information Criterion (AIC) to determine strongest models. The multiple R^2 value was reported for the final model (ArcGIS, 2015).

To assess the relative worth of the NSM a non-process model (CLIM model) was developed and used to create similar regression models for comparison. For this model, atmospheric climate, Rasmussen energy and NPP estimates, and terrain variables were used to create ordinary least squares and geographically weighted regressions to explain variability of soil properties related to climate in the same way as the NSM model. This CLIM model uses annual averages of climatic variables without any simulation of how these variables might influence seasonal fluctuations of soil moisture and temperature. It is believed that this CLIM model will provide a useful contrast to the simulations provided by the NSM. The comparison between the CLIM and the NSM thus represents the marginal value of including detailed pedon-based simulation of soil moisture and temperature with terms accounting for seasonal fluctuations. If the NSM performs better than the CLIM model, it can be assumed that detailed simulation is worthwhile. If it does not improve upon the CLIM model, then this suggests that the NSM model might be overly complicated and inefficient. The performance of CLIM and NSM regression models were compared by assessing differences in R^2 and AIC. Significant model terms were reported for each model set.

Geographically weighted regression (GWR) was used in models displaying a high degree of possible nonstationarity, indicated by a significant Jarque-Bera statistic (Scull, 2009; Miller et al., 2007). In most cases the GWR model performed only slightly better than the OLS model as it allowed regression coefficients to vary locally, based on greater weighting of neighboring data regions. The improvement may not be enough to warrant separate regression models for each region in all cases, as the satisfaction of the stationarity assumption leads to more universally applicable models. The two strongest models, those for soil pH at 50 cm depth and the soil accumulation index, were only marginally improved with a GWR approach. The OLS models accounted for 76% and 31% respectively and the GWR models accounted for 77% and 33% of the variability in the NSM models (Table 3.1). The improvement may not be enough to justify the use of GWR, which is a more complex regression approach as it gives separate linear coefficients for each spatial unit.

Table 3.1. Tabular output for regression models explaining soil variability by Newhall Simulation Model variables. Goodness of fit is reported with the adjusted R2 statistic for both ordinary least squares (OLS) and geographically weighted regression (GWR) models. Values for predictor variables are reported p-values in the final model.

NSM Models					Predictor variables used in the best models and their significance*																				
Response Variable	OLS Adj. R^2	GWR Adj. R^2		**	Yrmstcons		Yrmdcum		Yrdrycum		Wtrmscns		Smrdrycons		Bio8mstcon		Bio5mstcum		Bio5mdcum		Bio5drycum		Slope		Roughness
Soil pH at 50 cm	0.76	0.77				+	0.000			+	0.000					-	0.000			+	0.000	-	0.000	-	0.011
Soil Clay Accum Index (200 cm)	0.31	0.33		-	0.000	+	0.000			+	0.002					+	0.000					-	0.016		
Soil Clay KG 2 m x 1 m x 1 m (includes shallower pedons)	0.27	0.35														+	0.000	+	0.000			-	0.000		
Soil C upper 25 cm	0.23	0.41						-	0.000					-	0.014							+	0.000		

Soil Clay Max Single PM (200 cm)	0.22	0.33				+	0.000									+	0.000					-	0.000	+	0.008
Cumulative pedon CEC, upper 100 cm, NH4 extraction	0.18	0.22				+	0.000			+	0.000	+	0.014												
Soil C upper 200 cm	0.18	0.18				+	0.063					-	0.000	-	0.000									-	0.006
Cumulative pedon CEC, upper 150 cm, NH4 extraction	0.16	0.17				+	0.000			+	0.000	+	0.064												
Cumulative pedon CEC, upper 50 cm, NH4 extraction	0.16	0.21				+	0.000			+	0.000	+	0.013												

*Curvature and Yrmstcum variables not significant and not included in any model

** Sign to the left of predictor variable indicates direction of influence for the predictor variable.

Regression-driven visualizations were developed for visual assessment for successful models, in which variance inflation has been removed, the adjusted R^2 accounts for a good proportion of the variability (>0.30), all terms are significant ($p < 0.05$), and the AIC is observed to be lowest for the suite of variables used.

3.3 Results

In general, NSM models were stronger and more complex than CLIM models, with more predictor variables and higher goodness of fit. Ordinary least squares regression results ranged in goodness of fit from 0.16 to 0.76 for models with the NSM terms and from 0.08 to 0.73 for the CLIM models (Tables 3.1 and 3.2). In general, the NSM models explained slightly more of the variability of the response variables than the CLIM models.

Table 3.2. Tabular output for regression models explaining soil variability by naïve climate predictor (CLIM) model variables. Goodness of fit is reported with the adjusted R2 statistic for both ordinary least squares (OLS) and geographically weighted regression (GWR) models. Values for predictor variables are reported p-values in the final model.

CLIM Models					Predictor variables used in the best models and their significance* (p-values)										
Response Variable	OLS Adj. R^2	GWR Adj. R^2		**	Mean Annual Air Temperature		Mean Annual Precipitation		Slope		Roughness		Annual E_{ppt}		Annual NPP
Soil pH at 50 cm	0.73	0.82		+	0.000	-	0.000			-	0.000			-	0.016
Soil Clay Accum Index (200 cm)	0.28	0.31		+	0.000	-	0.000							+	0.018
Soil Clay KG 2 m x 1 m x 1 m (includes shallower pedons)	0.25	0.35		+	0.000			-	0.000			-	0.005	+	0.000
Soil C upper 25 cm	0.35	0.45		-	0.000	+	0.003	+	0.068					-	0.000
Soil Clay Max Single PM (200 cm)	0.20	0.32		+	0.001	-	0.000	-	0.023	+	0.003	+	0.000		
Cumulative pedon CEC, upper 100 cm, NH4 extraction	0.12	0.28		+	0.000							-	0.000	+	0.001
Soil C upper 200 cm	0.16	0.16								-	0.017			-	0.001

Cumulative pedon CEC, upper 150 cm, NH4 extraction	0.08	0.25		+	0.001							-	0.000	+	0.001
Cumulative pedon CEC, upper 50 cm, NH4 extraction	0.10	0.28		+	0.005							-	0.000	+	0.009

*Curvature not significant

** Sign to the left of predictor variable indicates direction of influence for the predictor variable.

Table 3.3. Regression equations used to create regression-based visualization figures. AIC values (lower is better) indicate model quality. R^2 values indicate model goodness of fit.

Model terms and Response variable	AIC	OLS Adj. R^2	Equation
NSM: Soil pH at 50 cm	366	0.76	6.27 - 0.0520 Slope - 1.24 Roughness + 0.00566 Bio5drycum - 0.00585 Bio5mstcum + 0.0118 Wtrmscns + 0.00998 Yrmdcum
CLIM: Soil pH at 50 cm	391	0.73	8.68 - 0.00187 Precipitation + 0.0320 Temperature - 1.40 Roughness - 0.000334 NPP
NSM: CAI	2073	0.31	260 - 6.12 Slope + 1.81 Bio5mstcum + 2.95 Wtrmscns + 2.34 Yrmdcum - 1.94 Yrmstcons
CLIM: CAI	2078	0.28	159 - 0.137 Precipitation + 19.2 Temperature + 0.0734 NPP

NSM: 25 cm cumulative carbon (kg m^{-2})	1080	0.27	11.7 + 0.529 Slope - 7.15 Roughness - 0.0150 Bio8mstcon - 0.0319 Yrdrycum
CLIM: 25 cm cumulative carbon (kg m^{-2})	1047	0.35	7.57 + 0.00982 Precipitation - 0.266 Temperature + 0.151 Slope - 0.00393 NPP

Multicollinearity was a serious concern with both sets of models, but especially with NSM models. Final models for NSM included between 23 – 46% of the initial variables, the majority of them having been removed due to multicollinearity and non-significance (Table 3.1). Final models for CLIM were more efficient, with between 29 – 71% of the initial variables used in the final models, with a smaller fraction of the initial terms removed due to multicollinearity and non-significance (Table 3.2). The CLIM models were therefore more efficient, with fewer redundant terms.

The terms in the most NSM models and with the greatest significance were Yrmdcum, Wtrmscons, Smrdrycons, and Bio5mstcum. The direction of influence with the term Yrmdcum was positive in all cases, indicating that the cumulative number of days the SMCS is partially moist and partially dry in the year leads to an increase in pH at 50 cm depth, greater clay accumulation, greater soil clay maximum in the upper 200 cm, greater cumulative pedon CEC in the upper 100 cm and upper 150 cm and upper 50 cm, and greater soil carbon in the upper 200 cm (Table 3.1). This accords with expectations regarding soil carbon stability in soils that are moist enough to support plant growth, but dry enough to restrict microbial breakdown of carbon at least during parts of the year. Likewise, for clay accumulation, it is expected that periods of moisture followed by periods of dryness are important for mobilizing clay movement in a downward direction (lessivage) during the development of argillic horizons. CEC is dominated by organic matter and clays, so it is expected that it should show similar patterns to both carbon and clay.

For the CLIM models, the most effective predictor variable was annual air temperature, with positive influence on soil pH, soil clay accumulation index, soil clay, soil CEC at three depths, but with a negative influence on soil carbon in the upper 25 cm (Table 3.2). Annual net primary productivity was also highly significant in all models but one. Greater NPP resulted in lower pH, lower soil carbon in the upper 25 cm and 200 cm, and in increased soil clay accumulation, soil clay, and cumulative CEC.

In OLS regressions, CLIM models accounted for less variability than NSM models for all models except the one made for soil carbon in the upper 25 cm (Tables 3.1 and 3.2). The model for soil carbon in the upper 25 cm was highly influenced by atmospheric temperature, probably reflecting the dominant influence of temperature on soil carbon metabolism rates by microorganisms, particularly at shallow depths where atmospheric temperature fluctuations have greater influence than at deeper zones. For all other models, NSM explained more soil variability than CLIM. This implies that simulation of the activity of soil moisture and temperature in the soil profile over a monthly time step can enrich understanding of the aspects of soil climate that influence measured soil properties.

The only OLS models with $R^2 > 0.30$ were soil pH at 50 cm, CAI, and C in the upper 25 cm (only for the CLIM model) (Figures 3.2 -3.7). NSM output regression models had lower AIC values than those for the CLIM models for most models, indicating better model quality and implying that the NSM has useful variables that are not present in the simpler climate model CLIM (Table 3.3).

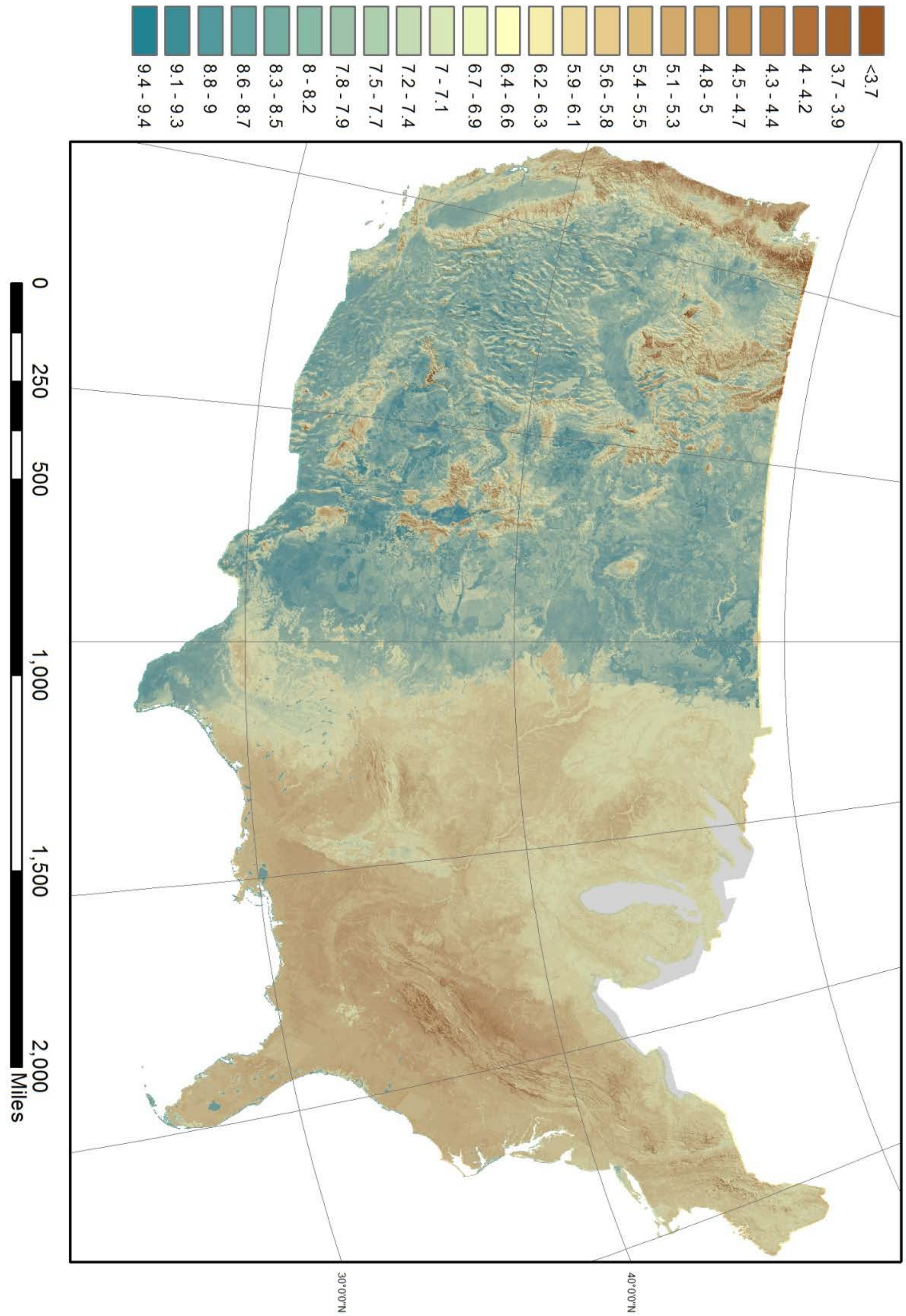

Figure 3.2. Regression output visualization for pH at depth 50 cm with NSM predictor variables.

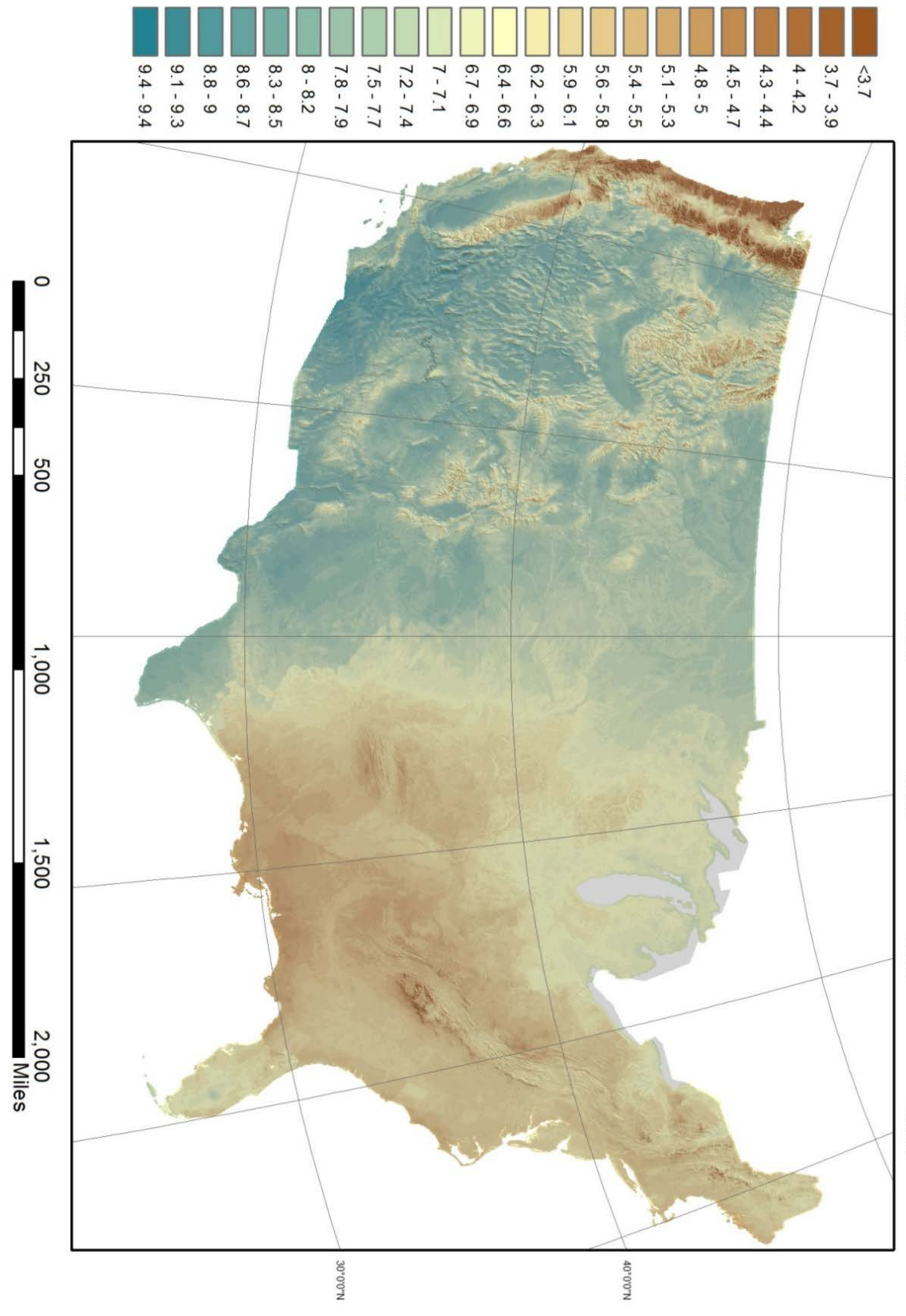

Figure 3.3. Regression output visualization for pH at depth 50 cm with CLIM predictor variables.

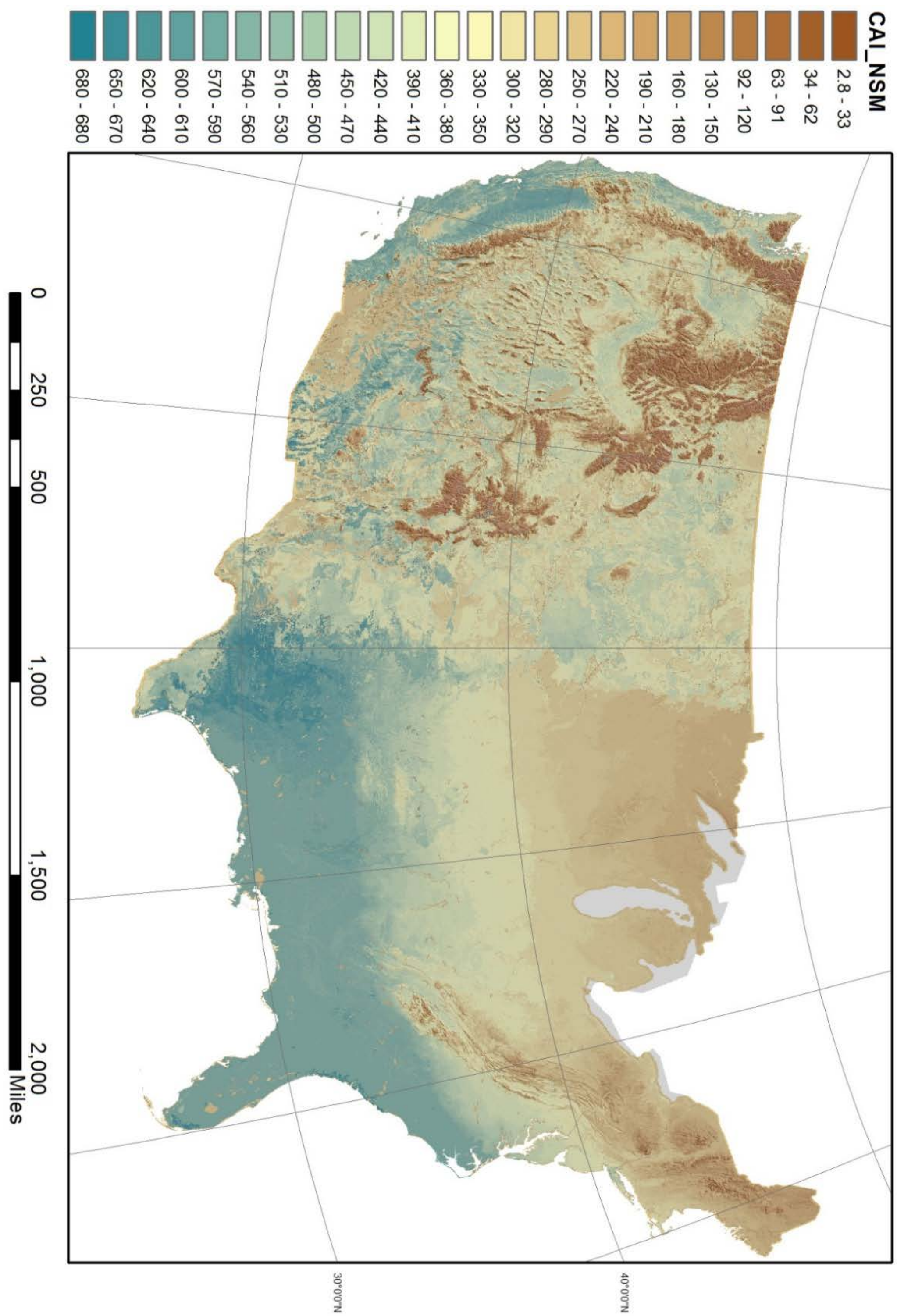

Figure 3.4. Regression output visualization for clay accumulation index (CAI) with NSM predictor variables.

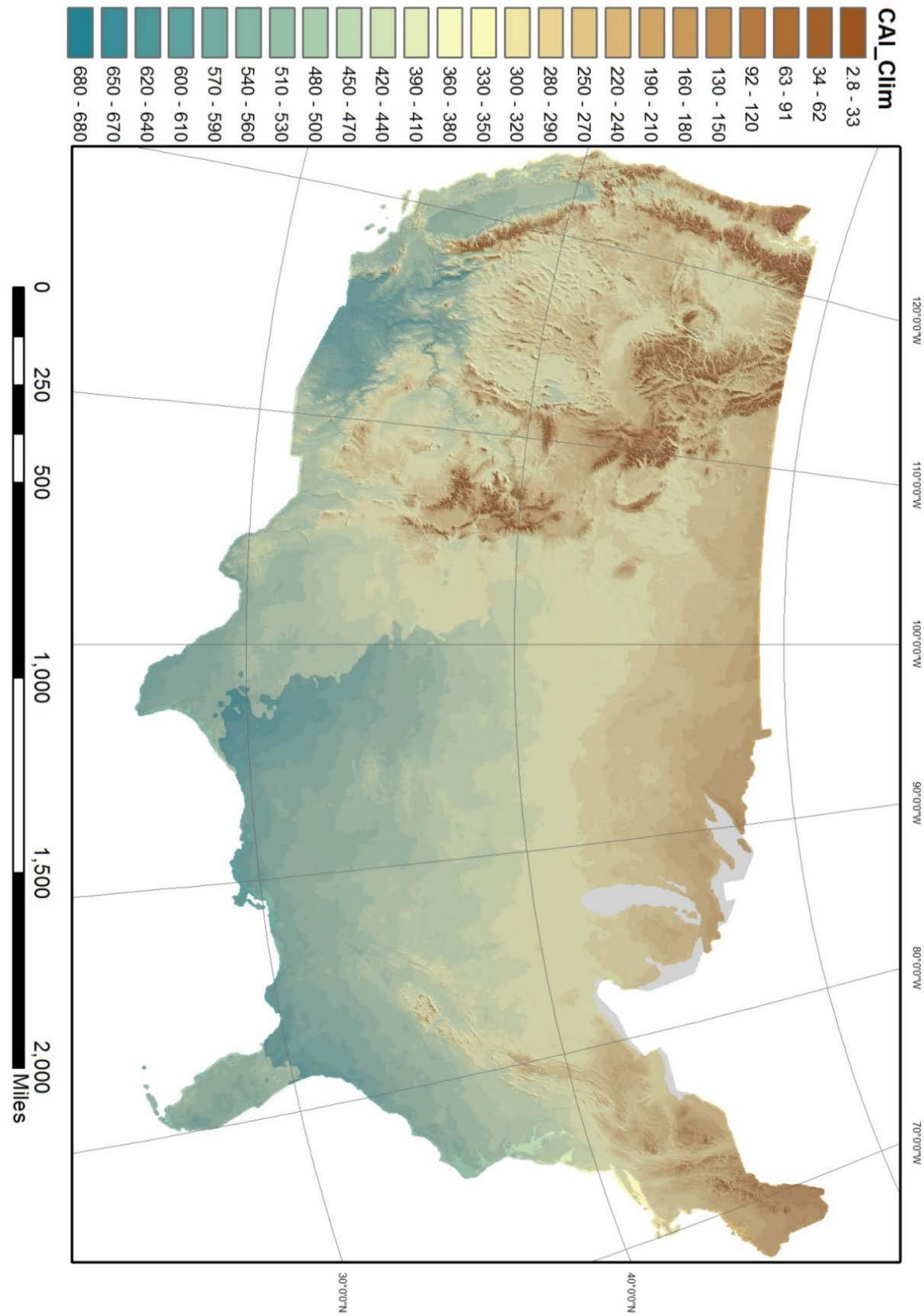

Figure 3.5. Regression output visualization for clay accumulation index (CAI) with CLIM predictor variables.

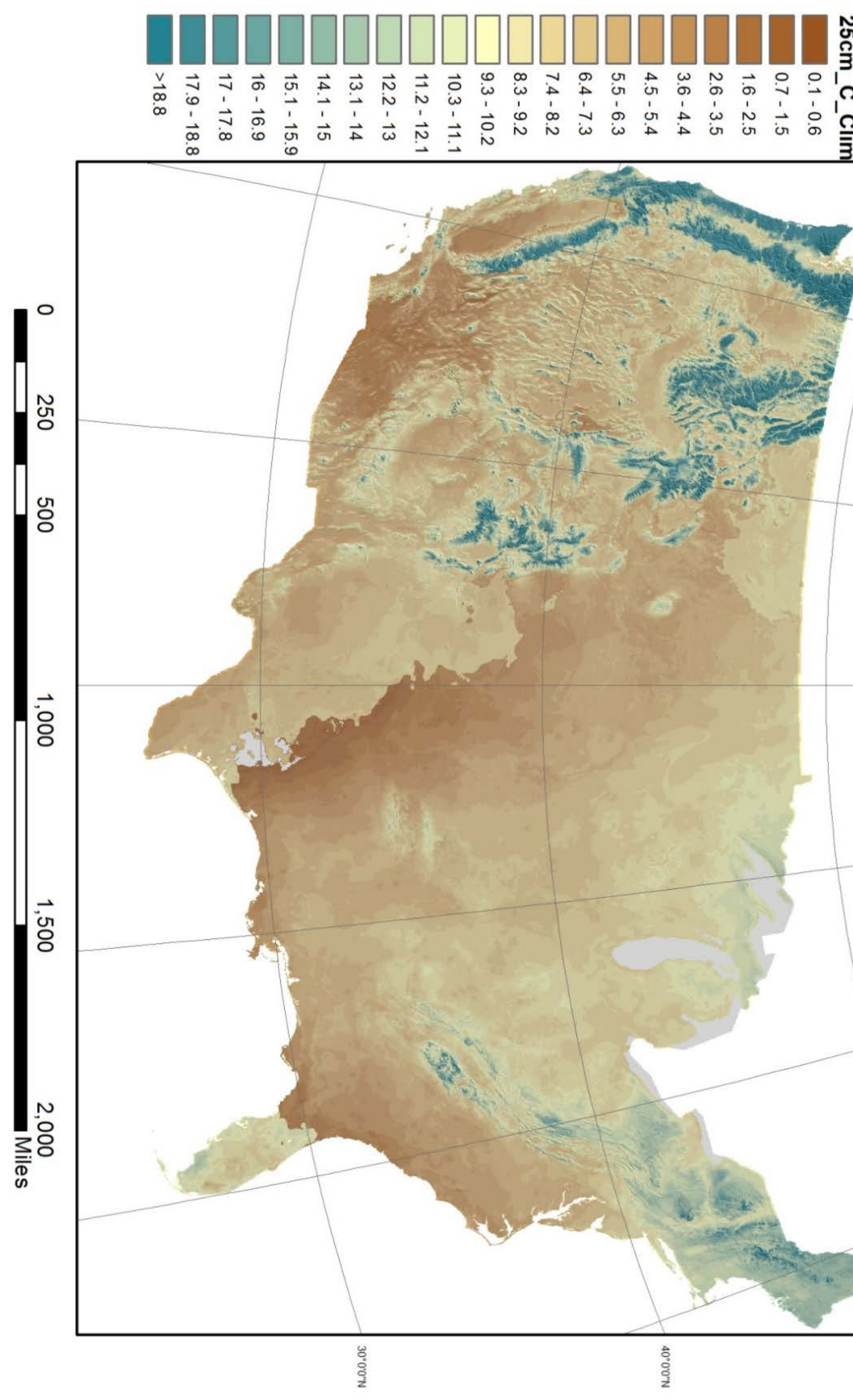

Figure 3.6. Regression output visualization for 25 cm depth carbon content (kg m^{-2}) with CLIM predictor variables.

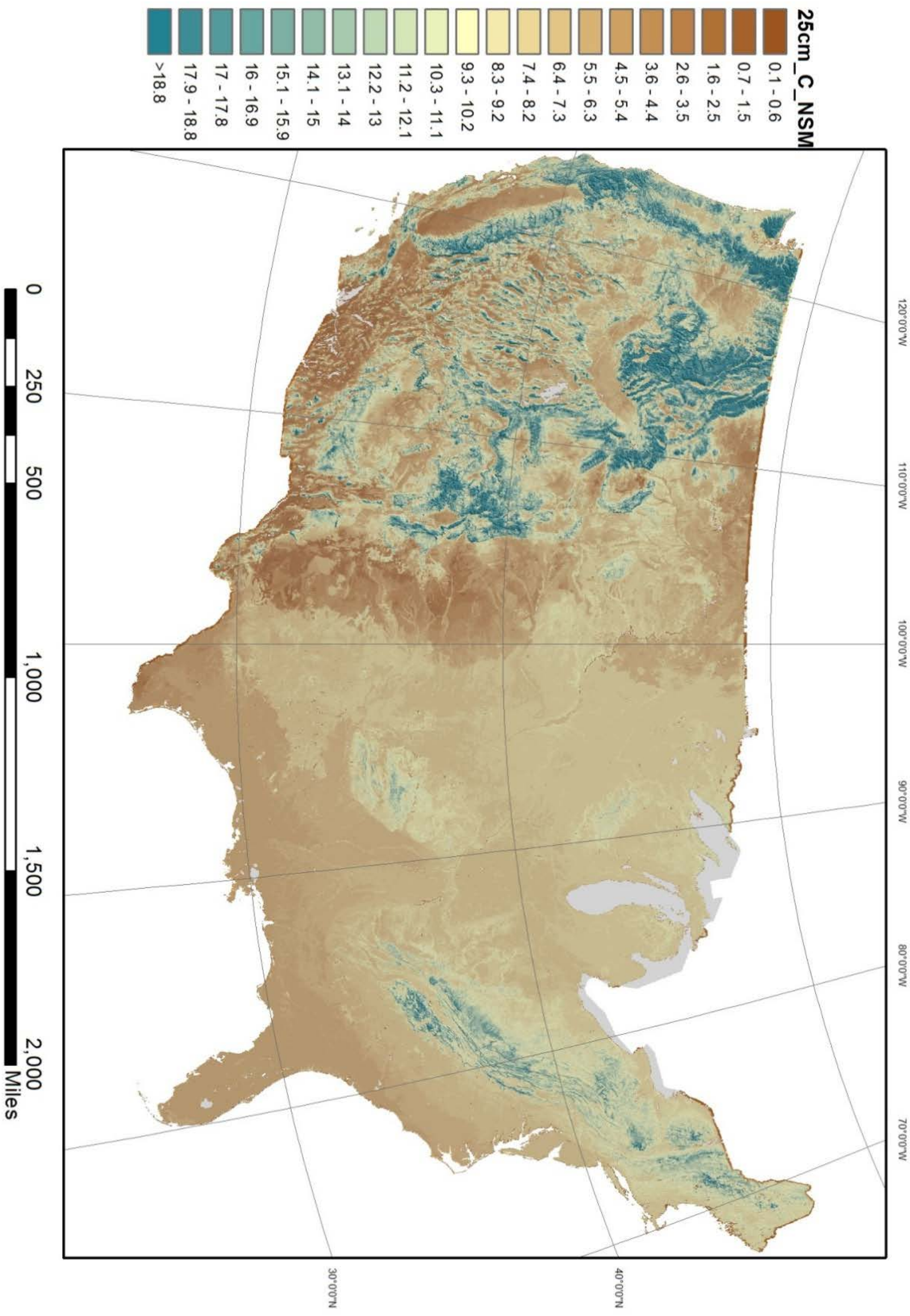

Figure 3.7. Regression output visualization for 25 cm depth carbon content (kg m^{-2}) with NSM predictor variables.

3.4 Discussion

While the NSM provides an integration of atmospheric climate variables and can provide estimates of a soil's wetting/drying cycles and temperature fluctuations, much of the output variables from the model are highly correlated, limiting their usefulness in regression models. With careful removal of terms, however, some of this multicollinearity can be controlled. When correlated variables are removed the NSM's performance as a predictive tool shows some strength and its terms generally improve upon model performance from regressions involving atmospheric climate predictor variables alone (CLIM model). Removal of redundant terms in the NSM makes the model more efficient and strengthens its performance as a predictive tool. The CLIM model also performed well and provided an efficient set of predictor variables for exploring soil variability related to climate. Its terms were more efficient, but had slightly less power than those from NSM (Table 3.3).

The NCSS database used in this project was not populated with samples that were specifically chosen for the needs of the paper. Nor is it known whether the sampling was done in a random fashion, or using other sampling techniques. As such, unknown biases may be included in the sample set. Some MLRA regions are well represented with many, seemingly randomly chosen samples, while some MLRAs are poorly represented with sparse samples that may have been taken for specific reasons not germane to the kind of data characterization that would have been best for this project. Limitations in the input data may be made up for by data extensiveness, particularly for data of the pH at 50 cm depth, giving good average representations of geographic variability. Observations of other variables, however, may have been too sparsely taken or may include unknown biases. Conclusions about relationships between climate variables and soil properties should be treated with some amount of caution.

Soil pH at a depth of 50 cm was shown to be positively correlated to climate variables related to weathering intensity. Increased rainfall and temperature lead to decreased soil pH through well-known pedogenic processes (Buol et al., 1989). Leaching of carbonates and bases from the soil profile occurs in climates where precipitation exceeds evapotranspiration. Exchangeable aluminum and hydrogen dominate very low pH soils as they are less soluble than bases, while free calcium and sodium may dominate soils with very high pH in conditions marked by high evapotranspiration. In conditions when rainfall exceeds evapotranspiration, calcium and sodium ions can be removed from the soil and replaced by exchange acidity. The dissolution of carbonic acid in soil water decreases soil pH and can result from respiration of microorganisms in processes influenced by soil moisture and temperature (Brady and Weil, 2001). In arid regions accumulation of soluble salts in the soil solution contributes to high pH values. Other processes influencing soil pH related to soil climate include the accumulation of organic matter, acids from biological metabolism, oxidation of nitrogen and sulfur, and plant uptake of cations (Brady and Weil, 2001). These processes all tend to promote greater acidity when rainfall and temperature levels support sufficient plant growth and microbial processes.

Clay content of soils is influenced by weathering of primary minerals to progressively smaller particle sizes, by conditions favoring neoformation of clays, and by underlying parent material in which the soil develops. Climate influences clay through the degree of weathering and the degree of lessivage driven primarily by precipitation, and through temperature by driving chemical transformations. Increased soil moisture generally leads to greater degrees of weathering of primary minerals, greater translocation of clays to deeper horizons, and to conditions more favorable to neoformation of clays. Higher temperatures with adequate rainfall promote greater degrees of chemical transformation of clays on a continuum from more smectitic to more kaolinitic (Brady and Weil, 2001).

The cation exchange capacity of most soils increases with pH because higher pH is associated with higher amounts of salts, higher base saturation, and higher levels of exchangeable bases (Brady and Weil, 2001). As soils weather, particularly under the influence of high rainfall and high temperature, CEC decreases as several processes take hold. Clay mineralogy changes from the smectitic to the more kaolinitic spectrum as positive ions are removed from the soil. Soil organic matter, a major contributor to soil CEC, may be degraded by microbial respiration under conditions of high temperature and moisture. Working independently and confounding expectations with respect to the influence of climate in these processes is the influence of underlying parent material. The models in this paper are insufficient to explain the influence of soil climate on CEC due, in part, to the influence of other soil forming factors not accounted for in the models such as parent material, geologic age, human management, or organisms.

Another variable that was insufficiently explained by the CLIM and the NSM models was soil C. Soil C is influenced by climate through weathering of carbon-containing primary minerals, through the climate's influence on net primary productivity, soil C decomposition, and through soil organic C dynamics. Soil organic C mineralization and decomposition follow temperature and moisture conditions and are governed by the stability of organic matter, the availability of substrate, the physiology of soil microflora, and physiochemical controls such as pH and the availability of oxygen, moisture and other conditions (Lützow and Kögel-Knabner, 2009).

The degree to which soil properties are influenced by climate has been observed for some time (Jenny, 1941). Testing the relationships between soils and climate with continental-scale datasets is a new task that shows some promise and can refine and develop the fundamental understandings (Scull, 2009). We have shown in this paper that a top-down approach, as recommended by Scull, that accounts for soil variability with depth in the profile may provide a particularly rich environment for further study. The influence of climate can be traced in the soil properties observed at a continental scale using the NSM. As it offers pertinent variables that relate to weathering effects in soils the NSM model is a useful tool for exploring continental-scale soil variability as influenced by long-term climate.

LIST OF REFERENCES

ArcGIS. 2015. Modeling spatial relationships toolset concepts. http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/interpreting-ols-results.htm Accessed Sunday, May 3, 2015.

Bonfante, A., A. Basile, P. Manna, and F. Terribile. 2011. Use of physically based models to evaluate USDA soil moisture classes. Soil Sci. Soc. Am. J. 75:181-191.

Brady, N.C., R.R. Weil. 2001. The nature and properties of soils, 13th edition. Prentice Hall, NJ.

Buol, S.W., F.D Hole, R.J. McCracken. 1989. Soil genesis and classification. Iowa State University Press, Ames, IA.

CIAT. 2012. Delta method, SRES A1B, HadCM3, 2080s (Precipitation and Mean Temperature) [dataset]. URL: http://www.ccafs-climate.org/data/.

CGIAR. 2011. SRTM 90m Digital Elevation Data. Accessed November, 2011. URL: http://srtm.csi.cgiar.org/.

Cohen, J. 1968. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol. Bull. 70:213-220.

Costantini, E.A.C., F. Castelli, S. Raimondi, and P. Lorenzoni. 2002. Assessing soil moisture regimes with traditional and new methods. Soil Sci. Soc. Am. J. 66:1889-1896

Cowell, C.M., and M.A. Urban. 2010. The changing geography of the U.S. water budget: twentieth-century patterns and twenty-first-century projections. Annals of the Associations of American Geographers. 100:740-754.

Davidson, E.A., I.A. Janssens. 2006. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. Nature. 440:165-173.

Di Luzio, M., G.L. Johnson, C. Daly, J.K. Eischeid, and J.G. Arnold. 2008. Constructing retrospective gridded daily precipitation and temperature datasets for the Conterminous United States. Bull. Am. . Meteorol. Soc. 47:475-497.

Dobos, E., T. Bailkó, E. Micheli, and J. Kobza. 2010. Legacy soil data harmonization and database development. J.L. Boettinger, D.W. Howell, A.C. Moore, A.E. Hartemink, and S. Kienast-Brown (Eds), In: Boettinger, Howell, Moore, Hartemink, Kienast-Brown, *Digital Soil Mapping: Progress in Soil Science*, volume 2, part IV. Springer: New York. P309-323.

Emadi, M., A.R. Shahriari, F. Sadegh-Zadeh, et al. 2016. Geostatistics-based spatial distribution of soil moisture and temperature regime classes in Mazandaran province, northern Iran. Archives of Agronomy and Soil Science. 62:502-522.

Gordon, C., C. Cooper, C.A. Senior, H. Banks, J.M. Gregory, T.C. Johns, J.F.B. Mitchell, and R.A. Wood. 2000. The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. Clim. Dyn., 16:147-168.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis, 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978.

Hudson, B.D. 1992. The soil survey as paradigm-based science. Soil Sci. Soc. Am. J. 56:836-841.

IPCC. 2014. Climate change 2014: Synthesis report. Contribution of Working Groups I, II, and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Geneva, Switzerland.

Iwahashi, J., R.J. Pike. 2007. Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. Geomorphology. 86:409-440.

Jarvis, A., H.I. Reuter, A. Nelson, E. Guevara. 2008. Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database (http://srtm.csi.cgiar.org).

Jenny, H. 1941. Factors of soil formation: A system of quantitative pedology. New York: McGraw-Hill.

Jenny, H. 1961. Derivation of state factor equations of soils and ecosystems. Soil Sci. Soc. Am. Proc. 25:385-388.

Jeutong, F., K.M. Eskridge, W.J. Waltman, and O.S. Smith. 2000. Comparison of bioclimatic indices for prediction of maize yields. Crop Sci. 40:1612-1617.

Lal, R. 2004. Soil carbon sequestration impacts on global climate change and food security. Science. 304:1623-1627.

Levine and Ciolkosz. 1983. Soil development in till of various ages in northeastern Pennsylvania. Quat. Res. 19: 85-99.

Lützow, M.V., I. Kögel-Knabner, K. Ekschmitt, E. Matzner, G. Guggenberger, B. Marschner, H. Flessa. 2006. Stabilization of organic matter in temperate soils: mechanisms and their relevance under different soil conditions – a review. European Journal of Soil Science. 57:426-445.

McBratney, A.B., M.L. Mendoça Santos, and B. Minasny. 2003. On digital soil mapping. Geoderma. 117:3-52.

Miller, J., J. Franklin, and R. Aspinall. 2007. Incorporating spatial dependence in predictive vegetation models. Ecological Modelling. 202 (3-4): 225-42.

National Cooperative Soil Survey. 2015. National cooperative soil survey characterization database. http://ncsslabdatamart.sc.egov.usda.gov/. Accessed Monday, December 7, 2015.

Nazarenko, L., GA Schmidt, RL Miller, N. Tausnev, M. Kelley, R. Ruedy, GL Russell, I Aleinov, M. Bauer, S. Bauer, R Bleck, V Canuto, Y Cheng, TL Clune, AD Del Genio, G Faluvegi, JE Hansen, RJ Healy, NY Kiang, D Koch, AA Lacis, AN LeGrande, J Lerner, KK Lo, S Menon, V Oinas, J Perlwitz, MJ Puma, D Rind, A Romanou, M Sato, DT Shindell, S Sun, K Tsigaridis, N Unger, A Voulgarakis, MS Yao, JL Zhang. 2015. Future climate change under RCP emission scenarios with GISS ModelE2. Journal of Advances in Modeling Earth Systems. 7:244-267.

NCDC. 2012a. Maps of annual 1961 – 90 normal temperature, precipitation, and degree days. U.S. Department of Commerce. National Oceanic and Atmospheric Administration. Accessed November, 2011. URL: http://www.ncdc.noaa.gov/oa/documentlibrary/clim81supp3/clim81.html.

NCDC. 2012b. Monthly station normals of temperature, precipitation, and heating and cooling degree days 1971-2000. Climatography of the United States, No. 81, National Oceanic and Atmospheric Administration, Asheville, NC. Accessed November, 2011. URL: http://www.ncdc.noaa.gov/oa/climate/normals/usnormalsprods.html.

Newhall, F. and C.R. Berdanier. 1996. Calculation of soil moisture regimes from the climatic record. Soil Surv. Invest. Rep. No. 46. Natural Resources Conservation Service: Lincoln, NE

PRISM Climate Group. 2015. Oregon State University. http://prism.oregonstate.edu Accessed Sunday, May 3, 2015.

Ramirez-Villegas, J. and A. Jarvis. 2010. Downscaling global circulation model outputs: the delta method decision and policy analysis working paper No. 1. CIAT. Accessed July, 2012. URL: http://www.ccafs-climate.org/media/ccafs_climate/docs/Downscaling-WP-01.pdf

Rasmussen, C., R.J. Southard, and W.R. Horwath. 2005. Modeling energy inputs to predict pedogenic environments using regional environmental databases. Soil Sci. Soc. Am. J. 69:1266-1274.

Rosenzweig, C.D., M. Karoly, M. vicarelli, P. Nefotis, Q. Wu, G. Casassa, A. Menzel, et al. 2008. Attributing physical and biological impacts to anthropogenic climate change. Nature 453:353-57.

Runge, E.C.A. 1973. Soil development sequences and energy models. Soil Science. 115:183-193.

SAGA GIS Development Team. 2012. System for Automated Geoscientific Analysis. Open Source Geospatial Foundation Project. http://sourceforge.net/projects/saga-gis/

Schaetzl, R.J., and S. Anderson. 2005. Soils: Genesis and Geomorphology. Cambridge University Press. New York, NY.

Scheffe, K. F., C. Loerch, and P. Finnell. 2012. Implementing soil data joining and recorrelation in MLRA soil survey projects. SSSA Intern. Meetings, 2012.

Scull, P. 2009. A top-down approach to the state factor paradigm for use in macroscale soil analysis. Annals of the Association of American Geographers. 100:1-12.

Seneviratne, S.I., T. Corti, E. Davin, et al. 2010. Investigating soil moisture-climate interactions in a changing climate: A review. Earth-Science Reviews. 99:125-161.

Smeck, N.E., E.C.A. Runge, E.E. Mackintosh. 1983. Dynamics and genetic modeling of soil systems. In L.P. Wilding et al. (ed.) Pedogenesis and soil taxonomy. Elsevier, NY.

Smith, G.D. 1986. The Guy Smith interviews: rationale for concepts on soil taxonomy. SMSS Technical Monograph No. 11. Cornell University, NY.

Soil Survey Staff. 1960. Soil classification, a comprehensive system, 7th approximation. USGPO Washington, D.C.

Soil Survey Staff. 1975. Soil taxonomy – a basic system of soil classification for making and interpreting soil survey. Handb. No. 436, SCS-USDA. USGPO Washington, D.C.

Soil Survey Staff. 1999. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. 2nd edition. Natural Resources Conservation Service. U.S. Department of Agriculture Handbook 436.

Thornthwaite, C.W. 1948. An approach toward a rational classification of climate. Geogr. Rev. 38(1): 55-94.

USDA. 1938. Soils and Men. Yearbook of Agriculture. USGPO. Washington, D.C.

USDA. 1999. Soil Taxonomy: a basic system of soil classification for making and interpreting soil surveys. USGPO Washington, D.C.

USDA. 2006. MLRA Geographic Database, version 4.2. http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_052440.zip Accessed Wednesday, March 30, 2016.

USDA-NRCS. 2006. U.S. General Soil Map (STATSGO2). Accessed November, 2011. URL: http://soils.usda.gov/survey/geography/statsgo/.

USDA-NRCS. 2007. National Soil Information System (NASIS) query to estimate root zone available water holding capacity using the Digital General Soil Map of the United States (STATSGO2). National Soil Survey Center, Lincoln, NE

USDA-NRCS. 2011. JAVA Newhall Simulation Model – Update to a traditional soil climate simulation model. Abstract Soil Sci. Soc. Am. Int. Annual Meeting. San Antonio, TX.

USDA. 2012. USDA Climate change science plan. Accessed July, 2012. URL: http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1047955.pdf

USDA-NRCS. 2002. US48MLRA [dataset]. Accessed July, 2012. URL: http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs143_012674.zip

USDA-SCS. 1994. Soil Climate Regimes of the United States. Soil Conservation Service: Lincoln, NE.

USEPA. 2016. Ecoregions of North America. Accessed November, 2016. URL: http://www.fs.fed.us/rm/ecoregions/products/map-ecoregions-north-america/.

Van Wambeke, A. 1982. Calculated moisture and temperature regimes of Africa. Soil Management Support Services. SCS, USDA. USGPO Washington DC.

Van Wambeke, A. 1986. Newhall Simulation Model, a BASIC program for the IBM PC (DOS 2.0 or later). Cornell University: Ithaca, NY.

Von Lützow, M., I. Kögel-Knabner. 2009. Temperature sensitivity of soil organic matter decomposition – what do we know? Biol. Fertil. Soils. 46:1-15.

Waltman, D. 2012. Newhall Simulation Model v 1.5.1 (JAVA Core Model). Accessed November, 2011. URL: http://code.google.com/p/newhall/.

Waltman, S.W., D.A. Miller, B. Bills, and W.J. Waltman. 2011. JAVA Newhall Simulation Model – Update to a traditional soil climate simulation model. Abstract Soil Sci. Soc. Am. Int. Annual Meeting. San Antonio, TX.

Waltman, S.W. 2011. Use of Gridded Soil Survey Geographic Data (gSSURGO and gSTATSGO2). Accessed February 2, 2016. URL: http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_051790.pdf

Williams, J.R., C.A. Jones, J.R. Kiniry, and D.A. Spanel. 1989. The EPIC crop growth model. Trans. ASAE 32:497-511.

Winzeler, H.E., P.R. Owens, S.W. Waltman, W.J. Waltman, Z. Libohova, D. Beaudette. 2013. A methodology for examining changes in soil climate geography through time: U.S. soil moisture regimes for the period 1971-2000. Soil Sci. Soc. Am. J. 77:213-225.

Yamoah, C.F., A. Bationo, T.J. Wyatt, B. Shapiro, and S. Koala. 2003. Simulated weather variables effects on millet fertilized with phosphate rock sand in the Sahel. Nutr. Cycling Agroecosyst.s. 67: 167-176.

Zhu, A.X., B. Hudson, J. Burt, K. Lubich, and D. Simonson. 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. *Soil Sci. Soc. Am. J.* 65:1463-1472.

VITA

Hans Edwin Winzeler
hewinzeler@gmail.com
Gettysburg, Pennsylvania
717.337.1561

B.A. English – Ohio University
B.S. Environmental Sciences – University of Illinois
M.S. Agronomy – Purdue University
Ph.D. Agronomy – Purdue University

Current employment: Pennsylvania State University Fruit Research and Extension Center – Technician. Ploughman Cider at P&Q – Cider and Wine Maker.

Previous employment: Material Matters, Inc. – Project Manager