

University of Denver

Digital Commons @ DU

Electronic Theses and Dissertations

Graduate Studies

1-1-2016

The Spatial Distribution of Terrestrial Stable Carbon Isotopes in North America, and the Impacts of Spatial and Temporal Resolution on Static Ecological Models

Sydney M. Firmin
University of Denver

Follow this and additional works at: <https://digitalcommons.du.edu/etd>



Part of the [Ecology and Evolutionary Biology Commons](#), and the [Geography Commons](#)

Recommended Citation

Firmin, Sydney M., "The Spatial Distribution of Terrestrial Stable Carbon Isotopes in North America, and the Impacts of Spatial and Temporal Resolution on Static Ecological Models" (2016). *Electronic Theses and Dissertations*. 1201.

<https://digitalcommons.du.edu/etd/1201>

This Thesis is brought to you for free and open access by the Graduate Studies at Digital Commons @ DU. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons @ DU. For more information, please contact jennifer.cox@du.edu, dig-commons@du.edu.

The Spatial Distribution of Terrestrial Stable Carbon Isotopes in North America, and the
Impacts of Spatial and Temporal Resolution on Static Ecological Models

A Thesis
Presented to
The Faculty of Natural Sciences and Mathematics
University of Denver

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts

by
Sydney M. Firmin
August 2016
Advisor: Rebecca L. Powell

Author: Sydney M. Firmin

Title: The Spatial Distribution of Terrestrial Stable Carbon Isotopes in North America, and the Impacts of Spatial and Temporal Resolution on Static Ecological Models

Advisor: Rebecca L. Powell

Degree Date: August 2016

ABSTRACT

Due to the unique spatial and temporal characteristics of ecological phenomena, the extent and grain size of spatial data sets essentially filter the observations. This thesis examines the impacts of temporal and spatial resolution on the modeling of terrestrial stable carbon isotopic landscapes (isoscapes). I model the distribution of leaf stable carbon isotope composition ($\delta^{13}\text{C}$) for the continent of North America at multiple temporal and spatial resolutions. I generate each $\delta^{13}\text{C}$ isoscape variation by first predicting the relative abundance of C_3/C_4 vegetation cover using monthly climate grids, crop distribution/type grids, and remote sensing data of plant life form, and then applying the respective leaf $\delta^{13}\text{C}$ endmembers to each pixel.

One application of isoscapes is predicting the geographic origin of migratory animals by relating the isotopic signature of animal tissue to environmental isotope values. I conduct multiple exercises in geographic origin assignment using known-origin feather isotope data of mountain plover (*Charadrius montanus*) chicks as an indirect means of testing the impact of resolution on $\delta^{13}\text{C}$ isoscapes. Results indicate that temporal resolution does have a significant impact on predicted isoscape layers, and in turn, geographic origin assignment efficacy. Temporal periods that did not correspond to tissue growth exhibited a mismatch in the range of predicted vegetation $\delta^{13}\text{C}$ values

relative to the range of measured feather $\delta^{13}\text{C}$ values and therefore were not useful in generating geographic origin assignments. The spatial resolution of modeled $\delta^{13}\text{C}$ minimally impacted assignment accuracy and precision compared to temporal resolution; however, the current analysis was limited by the spatial resolution of the input data set. These results should be further explored to better characterize spatiotemporal ecological characteristics of migratory animals and to improve modeling of the isotopic landscape itself.

ACKNOWLEDGEMENTS

I would like to start by thanking my advisor Dr. Rebecca Powell. She has been an invaluable source of guidance and rationality throughout my time at the University of Denver, and I cannot begin to express my gratitude for the doors she has opened for me.

I am grateful to my advisory committee, Dr. Jing Li and Dr. Donald Sullivan, for their feedback and help in completing my thesis; and Dr. Shannon Murphy for taking time to serve as outside chair for my Thesis Defense. I would also like to thank Dr. Michael Wunder of the University of Colorado, Denver for sharing his Mountain Plover chick data, as well as and his guidance and feedback in learning the methods to geographic origin assignment.

I would like to acknowledge Dr. Cathy Durso, Dr. Michael Minn and Dr. Hannah Vander Zanden for helping me learn the programming language R; my peers in the geography department for their feedback and support; and my many mentors and peers at the 2015 SPATIAL short course at the University of Utah. I would additionally like to acknowledge the NSF for making my attendance to the SPATIAL short course possible.

And of course, I am grateful for all of the support and love I have received from my family and friends; without them I would not have been able to start, let alone finish my thesis.

TABLE OF CONTENTS

CHAPTER 1: Introduction.....	1
1.1 Background and Motivation	1
1.2 Research Questions	5
1.3 Intellectual Merit: Geographic Representation	6
1.4 Broader Impacts: Connecting Geography and Ecology	7
1.5 Summary of Chapters	7
 CHAPTER 2: Background	10
2.1 The Hazards and Opportunities of Spatial Analysis	10
2.2 Stable Terrestrial Carbon Isoscapes	19
2.3 Using Isoscapes for Geographic Origin Assignment	27
 CHAPTER 3: A Terrestrial Stable Carbon Isoscape of North American	35
3.1 Introduction	35
3.2 Methods	37
3.3 Results	47
3.4 Discussion	55
3.5 Conclusion	58
 CHAPTER 4: The Impact of Temporal Resolution on Isoscapes in Mixed Grasslands ...	60
4.1 Introduction	60
4.2 Methods	62
4.3 Results	71
4.4 Discussion	80
4.5 Conclusion	87
 CHAPTER 5: The Impact of Spatial Resolution on Stable Carbon Isoscapes	89
5.1 Introduction	89
5.2 Methods	92
5.3 Results	93
5.4 Discussion	96
5.5 Conclusion	98
 CHAPTER 6: Summary and Conclusions	100
6.1 Summary of Research	100
6.2 Limitations and Suggested Improvements	103
6.2 Future Research	105

REFERENCES	108
APPENDICES	115
Appendix 1: Soil Organic Matter (SOM) Validation Data Set	115
Appendix 2: Rescaling Equation and Calibration Data Plots	117
Appendix 3: Geographic Origin Assignment Code	125

LIST OF FIGURES

Figure 1. Two aspects of the MAUP. Source: Jelinski & Wu, (1996)	12
Figure 2. Frequency and ranges of leaf $\delta^{13}\text{C}$ for C_3 and C_4 plants. Source: Tipple & Pagani, (2007)	21
Figure 3. C_3/C_4 dominance in relation to temperature and atmospheric CO_2 . Source: Edwards <i>et al.</i> , (2010)	23
Figure 4. Collapsed MODIS IGBP land-cover classification for the year 2001	40
Figure 5. Number of months over a year each 5-minute raster pixel meets C_4 climate Collatz-crossover model conditions for two parameters	43
Figure 6. Individual output layers from the intermediate isoscape modeling steps	45
Figure 7. The four permutations of the annual, 5-min, $\delta^{13}\text{C}$ isoscape	49
Figure 8. Difference maps between the herbaceous land cover rules (Fig. 8-A) and the temperature crossover parameters (Fig. 8-B)	51
Figure 9. Locations of SOM validation sites	52
Figure 10. Seasonal C_3/C_4 land cover	63
Figure 11. Original Mountain Plover chick sample sites within breeding range. Source: Wunder <i>et al.</i> , (2005)	66
Figure 12. Four seasonal $\delta^{13}\text{C}$ isoscapes	71
Figure 13. Twelve monthly $\delta^{13}\text{C}$ isoscapes	73
Figure 14. Plots of population-level assignment accuracy	75
Figure 15. Plots of population-level assignment precision	78
Figure 16. Calibration points and rescaling equations for May isoscape	83
Figure 17. Calibration points and rescaling equations for June isoscape	84
Figure 18. Example of varying spatial resolutions of raster data model	91
Figure 19. Annual $\delta^{13}\text{C}$ isoscape and the Mountain Plover breeding range	94

Figure 20. Three variations of $\delta^{13}\text{C}$ isoscapes with differing spatial resolutions 95

Figure 21. Population-level assignment accuracy and precision 97

CHAPTER 1: Introduction

“An ecosystem is the interacting system made up of all the living and nonliving objects in a specified volume of space” -Weathers et al., 2013

“Space and time frame all aspects of the discipline of geography.” – Goodchild, 2013

1.1 Background and Motivation

Because ecology is the study of connections and interactions between living and non-living entities sharing space, the complexities of modeling ecological patterns and processes are directly connected to fundamental geographic questions concerning spatial resolution and extent. The spatial properties of a model often determine the range of patterns and processes that can be observed (Dark & Bram, 2007). Because both biotic and abiotic processes important to ecology occur across a range of spatial and temporal resolutions, when developing an ecological model, an essential question becomes, “what is the best way to represent these processes?” (Goodchild, 2011).

Spatial resolution and extent are essential components all spatial models. Models using the same data aggregated at different spatial and temporal resolutions can yield different, and occasionally contradictory, results (Openshaw & Taylor, 1977; Jelinski & Wu, 1996; Dark & Bram, 2007). These spatial modeling considerations have been defined as properties of the *modifiable areal unit problem* (MAUP). The MAUP is made

up of two separate but related phenomena: the *scale effect* and the *zonation effect* (Jelinski & Wu, 1996; Dark & Bram, 2007). The scale effect occurs when data are aggregated into fewer, larger units, resulting in decreased variance, while the mean values remain the same (Bram & Dark, 2007). The zonation effect occurs when the data set is recombined into different areal units (i.e., same area but different shapes and/or locations), which results in different data values (impacting both the mean and the variance of the data) (Jelinski & Wu, 1996; Dark & Bram, 2007).

The MAUP is an issue for nearly any quantitative study using spatial data (Openshaw & Taylor, 1979; Dark & Bram, 2007). Due to the proliferation of freely available spatial data over the past few decades, researchers in a wide range of fields commonly exploit data sets created without the specific purpose of their research, meaning they do not have direct control over the spatial or temporal resolution of input data (Dark & Bram, 2007). As a result, selection of spatial grain size and spatial extent is often arbitrary (Jelinski & Wu, 1996; Bark & Bram, 2007). However, the MAUP does not necessarily need to be viewed as a problem; rather, it reflects the characteristics of real-world phenomena and how they are structured (Jelinski & Wu, 1996). If the MAUP is recognized and dealt with intentionally, it can be leveraged to characterize the structure, function and dynamics of the processes the model is attempting to capture. Geography provides the conceptual framework for the handling of spatial data across diverse research communities (Goodchild, 2013).

Ecological models often seek to represent a myriad of complex environmental processes in a single model. However, a fair amount of uncertainty exists in modeling

individual ecological processes and how they might interact with one another. Due to these factors, effective ecological models are often created by modeling indirect measures to characterize ecosystem processes. This is a particularly useful strategy when the study area spans a large spatial extent, such as a region or continent (Turner, 2005; Wu *et al.*, 2013; Heffernan *et al.*, 2014; Fei *et al.*, 2016).

One example of a *macrosystem* ecological model is the spatio-temporal distribution of stable isotopes in the environment, which can be represented as an isoscape (isotopic landscape). Isoscapes are used in a wide variety of research, ranging from forensic anthropology to food-web ecology (Bowen & West, 2008; Bowen 2010; West *et al.*, 2010). In this research, I generate a terrestrial stable carbon isoscape of North America by predicting general vegetation distributions and relating this prediction to expected leaf carbon isotope ratio ($\delta^{13}\text{C}$) content. Spatial representations of stable carbon isotopes are used in a range of ecological and environmental science research. For example, because ratios remain relatively constant during the assimilation that occurs between an animal's diet and their body tissues, the patterns of $\delta^{13}\text{C}$ that exist in a landscape can be used to constrain the geographic origin of animals based on the $\delta^{13}\text{C}$ (generally in combination with other stable isotopes) within their body tissue (Wunder *et al.*, 2005; Hobson *et al.*, 2012). Mapping $\delta^{13}\text{C}$ can also be used to characterize physiological controls on, as well as sources of biosphere-atmosphere gas exchange (Bowling *et al.*, 2002; Suites *et al.*, 2005).

Stable carbon isoscapes have been created and published at global extents (Lloyd and Farquhar, 1994; Still *et al.* 2003; Suits *et al.*, 2005) and continental extents for Africa

and South America (Still and Powell, 2010; Powell *et al.*, 2012). However, stable carbon isoscapes have yet to be published for continents spanning mid- to high- latitudes such as North America or Eurasia. The spatial distribution of terrestrial stable carbon ($\delta^{13}\text{C}$) isotopes is determined primarily by plant functional type distribution (i.e., the relative composition of C_3/C_4 plants). Because North America and Eurasia have temporally offset but geographically co-dominant C_3/C_4 grasslands, these continents present a modeling challenge of both spatially and temporally mixed pixels. The modeling of mid-latitude grasslands to generate a stable terrestrial carbon isoscape therefore offers an excellent opportunity to explore the impacts of spatial and temporal resolution. Generating a stable carbon isoscape for North America contributes to a gap in the current data sets available to the stable isotope research community, as well as develops best practices for the representation of dynamic landscapes in static ecological models.

Assessing the impact of spatial and temporal resolution on ecological models is quite challenging because directly comparing the same data set at different spatial or temporal resolutions essentially only tests the resampling or interpolation method, not the impact of resolution on the accuracy of the data set or the resulting model itself. To work around this, I propose creating an ecological model at a series of different spatial and temporal resolutions, and then comparing the performance of the models when applied to an ecological analysis, specifically, a geographic origin assignment for a migratory species.

Geographic origin of an individual organism can be predicted by relating the isotopic signature of animal tissue to spatially explicit modeled values of stable isotopes

entering the food web (Kelly & Finch, 1998; Hobson *et al.*, 1999; Wunder, 2010; Bowen *et al.*, 2014; Vander Zanden *et al.*, 2014). This is possible because organisms assimilate the stable isotope composition of their environment (with some modification) into their body tissue, meaning that “you are what you eat, isotopically” (DeNiro & Epstein, 1978). Because an “isotopic fingerprint” is left on the inert tissues of an animal by their local environment, isotopes can serve as a forensic tool for ecology (Ehleringer & Monson, 1993; Bowen & West, 2008; Wunder & Norris 2008; Hobson *et al.*, 2010; Wunder, 2010; Bowen *et al.*, 2014).

In order to test the impact of varying spatial and temporal resolutions on the stable carbon isoscape model, I conduct multiple exercises in geographic origin assignment, using known-origin feather isotope data of mountain plovers (*Charadrius montanus*) (Wunder *et al.*, 2005; Wunder, 2010; Bowen *et al.*, 2014; Vander Zanden *et al.*, 2014). By using known-origin feather isotope data, I am able to compare the assignment predictions of each temporal and spatial variation of the $\delta^{13}\text{C}$ isoscape model against a validation data set and thereby indirectly assess impact.

1.2 Research Questions

This research explicitly links the fields of geography and ecology, explores best practices for representing variability in a static models, and guides new questions about ecological process by specifically considering the following guiding research questions:

1. *What is the spatial distribution of vegetation (i.e., leaf) stable carbon isotopes on the continent of North America?*

2. *How can seasonal variation of vegetation cover be represented in a static model? What is the impact of temporally aggregating the seasonal variation of mid-latitude grasslands impact confidence in the final model?*
3. *What is the impact of spatial resolution on overall confidence in the model? What is the impact of aggregating fine-grained spatial data versus resampling coarse-grained spatial data on the final data product?*

1.3 Intellectual Merit: Geographic Representation

Several previous efforts to characterize mid-latitude grasslands have harnessed the phenological differences in C₃ and C₄ grasses as a mechanism for distinguishing plant functional types (Wang *et al.*, 2011; Wang *et al.*, 2013; Zhong *et al.*, 2015). Although these methods have produced promising results, the end products are ultimately a static output, (i.e., the average C₃/C₄ abundance over a season). Static models assume temporal consistency (Bowen *et al.*, 2005; Vander Zanden *et al.*, 2015). This assumption often creates temporally averaged models that do not represent a “real” environment at any one given point in time. This is a particularly important consideration for migratory animals, who only inhabit a region at a specific time of the year. Animals integrate environmental signals via their food web at varying spatial and temporal resolutions (Hobson & Norris, 2008). My research will explore best practices for representing temporally dynamic processes in a static data product.

Representing spatial and temporally dynamic phenomena at an appropriate resolution is challenging. It is an additional challenge to quantify the impact of model

spatial and temporal resolutions on analysis results. Directly comparing data products of differing spatial and temporal resolutions is difficult, particularly when they are derived from the same input data. A direct comparison can identify disagreements that are potentially an artifact of aggregation and resampling methods, rather than assessing the impact on model output (Gotway & Young, 2002; Pontius *et al.*, 2007). My research develops and explores a novel technique to assess impacts of spatio-temporal resolution on an ecological model.

1.4 Broader Impacts: Connecting Geography and Ecology

Recently, due to environmental threats to biomes globally, as well as wide-spread availability of large extent, spatially explicit data, *macrosystems ecology* has emerged as a rapidly growing framework within ecology which facilitates investigations that emphasize spatial and temporal patterns, heterogeneity, and interactions of processes across multiple scales (Turner, 2005; Wu *et al.*, 2013; Heffernan *et al.*, 2014; Fei *et al.*, 2016). These emphases are all core tenants of studies within the field of geography. In my research, I will apply a geographic perspective to challenges commonly faced in ecology, and thereby contribute to building explicit connections between the two disciplines.

1.5 Summary of Chapters

Chapter 2: Background provides a literature review and theoretical context to my thesis research. Traditional geographic and geostatistical problems such as the *modifiable areal unit problem* (MAUP) and the *change of support problem* (COSP) are defined and

discussed in the context of ecological modeling and research. The potential of leveraging these “problems” in ecological models is considered. Next, literature regarding the modeling of plant functional types and carbon isoscapes is summarized. Finally, the use of stable carbon isoscapes for geographic origin assignments is reviewed, detailing principles of geographic origin assignments using the probability surfaces model.

Chapter 3: Research Question 1 details the steps to generate an annual terrestrial $\delta^{13}\text{C}$ isoscape for the continent of North America that is complimentary to the stable carbon isoscapes generated for Africa (Still & Powell, 2010) and South America (Powell *et al.*, 2012). I identify “optimal” parameters for modeling the spatial distribution of stable carbon isotopes in North America at an annual temporal and 5-min spatial resolution by comparing four variations of the $\delta^{13}\text{C}$ models to a reference data set of soil organic matter (SOM) values spanning the Great Plains, USA. This chapter informs the parameters used to generate the spatial and temporal variations of $\delta^{13}\text{C}$ isoscapes for research questions 2 and 3 (Chapters 4 and 5, respectively).

Chapter 4: Research Question 2 explores the impact of temporal resolution on stable carbon isoscape models in mixed grassland regions. This chapter describes methods for assessing and comparing the efficacy of different isoscape models by using geographic origin assignment with known-origin tissue samples as a method for model “validation”. Seasonal and monthly terrestrial $\delta^{13}\text{C}$ isoscapes are developed using the methods outlined in Chapter 3, and compared to the annual temporal resolution product. The results of this analysis are discussed, with special attention paid to the impact of temporal resolution on $\delta^{13}\text{C}$ isoscape models in regards to temporally varying grasslands.

Chapter 5: Research Question 3 expands the methods and techniques developed in Chapter 4 by exploring the impacts of spatial resolution on modeling stable terrestrial carbon isoscapes. Two additional annual stable carbon isoscapes are generated at different spatial resolutions (1-kilometer and 10-minute) using the methods outlined in Chapter 3. These two isoscapes as well as the annual 5-minute isoscape are assessed using the geographic origin assignment validation method detailed in Chapter 4 to evaluate the impact that spatial resolution has on representing terrestrial $\delta^{13}\text{C}$ isoscapes and the efficacy of assigning the geographic origin of mountain plovers.

Chapter 6: Summary and Conclusions summarizes the findings of my thesis research and draws general conclusions on limitations and best practices for approaching fundamental geographic problems such as the MAUP or the COSP when creating isoscapes and using geographic origin assignment methods. Recommendations for future avenues of research are presented.

CHAPTER 2: Background

2.1 The Hazards and Opportunities of Spatial Analysis

2.1.1 Describing the Spatiotemporal Characteristics of Data

Spatially-explicit models created in a Geographic Information System (GIS) are based off of digital spatial data sets. The fundamental spatial and temporal characteristics of a spatial data set are described by *resolution* and *extent*. *Extent* refers to the total geographic area that the data set encompasses. *Extent* can also refer to the temporal duration or range of data (Turner *et al.*, 1989; Fassnacht *et al.*, 2006; Goodchild, 2011). *Resolution* refers to the degree of detail, and applies both to the spatial and temporal dimensions (Turner *et al.*, 1989; Hay *et al.*, 2001). In the case of raster (i.e. gridded) data, spatial resolution is explicitly defined as the ground dimensions of the raster cells, or pixels (Jelinski & Wu, 1996; Fassnacht *et al.*, 2006; Goodchild, 2011). Although the term scale has been inconsistently defined across disciplines, in the context of geographic information systems (GIS) -and the context of this research- *scale* refers to the combination of resolution and extent (Fassnacht *et al.*, 2006; Goodchild, 2011).

2.1.2 The Modifiable Areal Unit Problem

All quantitative studies that utilize spatial data are subject to the modifiable unit areal unit problem (MAUP). An iconic study on the impacts of the MAUP by Openshaw

& Taylor (1979) illustrated how modifying the basic areal units (spatial grain size and zone) of an input data set can radically change the outcome of analysis. The study analyzed data from 99 counties in Iowa and found that the correlation between the percentage of elderly voters and percentage of Republican voters ranged from +0.979 to -0.811, depending on the basic areal unit of analysis. This effect has been repeatedly demonstrated in studies in both human and physical geography (Openshaw & Taylor, 1979; Jelinski & Wu, 1996; Hamil *et al.*, 2016).

The MAUP has two major components: the *scale effect* and the *zonation effect*. The *scale effect* is primarily a result of data aggregation, which causes variations in numerical results strictly due to the number of areal units used to characterize a given area (Openshaw & Taylor, 1979; Dark & Bram, 2007). The *scale effect* occurs when data are combined into fewer, larger units. As a result, the variance of the data are reduced while the mean of the data set remains the same, as illustrated in *Fig. 1 (a-c)*. This effectively smoothes the data and results in information loss (Dark & Bram, 2007). The *zonation effect* describes differences in data values as a result of the way in which smaller areal units are grouped into larger areal units (Openshaw & Taylor, 1979; Dark & Bram, 2007), and impacts both the variance and mean of the data set. The impact of zonation on mean and variance is less predictable than the impact of linear aggregation, as illustrated in *Fig. 1 (d-f)* (Jelinski & Wu, 1996; Dark & Bram, 2007).

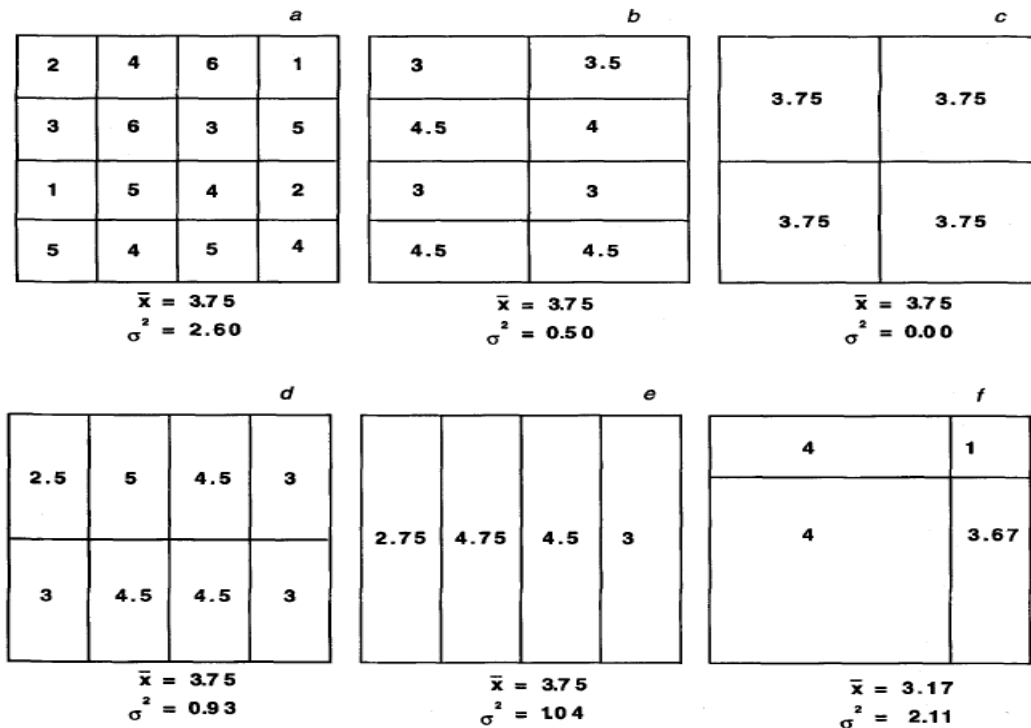


Figure 1 Example showing the two aspects of MAUP; (a-c) represent the scale effect, showing how aggregation reduces variation and spatial heterogeneity, while (d-f) depict the zonation effect, showing how zones can greatly impact variance, and even the mean of a data set, even when maintaining the same number of areal units Source: Jelinski & Wu, (1996).

The MAUP is also associated with the concept of *ecological fallacy*, which occurs when aggregate data values are applied to an individual within the data set (Gotway & Young, 2002; Dark & Bram, 2007). This can occur in studies using spatial data that do not differentiate between the spatial associations caused by aggregation of the data versus the actual associations between the data (Openshaw, 1984; Dark & Bram, 2007). Ecological fallacy is an extreme condition of cross-scale inference or downscaling (Goodchild, 2011). Making inferences at a finer resolution than the input data (i.e., downscaling) can be just as hazardous as aggregation (Hamil *et al.*, 2016), as attempting to make inferences at finer (or coarser) resolution than the original data set may produce

invalid results. Although the MAUP is well documented and researched, no formal solution exists (Dark & Bram, 2007).

2.1.3 The Modifiable Temporal Unit Problem

All of the hazards associated with the MAUP also apply to selecting and modifying the resolution and extent of temporal data (Dark & Bram, 2007). Although the MAUP addresses the intertwined nature of space and time (Cressie, 1996), the particular complications involved in using temporal data have recently been specifically identified as the Modifiable Temporal Unit Problem (MTUP) (Coltekin *et al.*, 2011; De Jong & Bruin, 2012; Cheng & Adepeju, 2014).

Like the MAUP, the MTUP identifies the hazards associated with arbitrary selection or modification of sampling units. MTUP is composed of three components: the aggregation effect (related to temporal resolution or interval), the segmentation effect (how intervals are defined in time), and the boundary effect (duration or extent of the temporal window) (Coltekin *et al.*, 2011; Cheng & Adepeju, 2014). The aggregation effect is analogous to the MAUP scale effect, in which the aggregation of data results in loss of data heterogeneity. Temporal data are often aggregated for ease of processing or in order to assimilate the data to a regular sampling unit (Cheng & Adepeju, 2014). The temporal segmentation effect is related to the zonation effect in the MAUP. The temporal segmentation effect identifies that the way the data are divided into smaller units impacts the results of analysis. For example a week-long segment could be defined as spanning Sunday through the following Saturday, versus to starting on Mondays and ending on

Sundays. Although the temporal grain is equivalent in both definitions, results of analysis could be different (Cheng & Adepeju, 2014).

The impacts of the MTUP have been documented in multiple studies. De Jong & de Bruin (2012) found that temporal aggregation schema directly impact model results. In particular they found that a large portion of the variation was a result of changing aggregation bins (de Jong and de Bruin, 2012). In a case study using crime data in central London, Cheng & Adepeju (2014) found that temporal unit had a significant effect on the temporal duration, spatial extent and statistical significance of analysis results, and suggested that aggregation could actually be harnessed to more rapidly identify significant clusters of crime than data at finer temporal resolutions. The findings of Cheng & Adepeju (2014) suggest that the segmentation effect could assist in characterizing the cyclic patterns of crime.

Often, decisions regarding temporal scale are rooted in data availability, or based on empirical analysis rather than theory (Coltekin *et al.*, 2011). Much like the MAUP, there is not necessarily a solution to the MTUP. Rather, it is a potential source of error that must be explicitly accounted for.

2.1.4 Combining Mismatched Spatial Data (The Change of Support Problem)

The MAUP, which has been well documented and researched within the field of geography, is directly related to the change of support problem (COSP). The COSP is defined in geostatistics as the problems associated with integrating spatial data that exist in different forms or dimensions (e.g., point and raster data) (Cressie, 1996; Gotway &

Young, 2002). In this context, support refers to the geometrical size, shape and spatial orientation of the regions associated with the data measurements (Cressie, 1996). To integrate spatially mismatched data sets, data are often averaged or aggregated to a matching form for analysis, resulting in a new variable. The new variable is related to the original, but due to the transformation may possess different statistical and spatial properties (Gotway & Young, 2002). Many statistical methods have been developed for combining spatially mismatched data, (e.g., kriging, Bayesian areal regression models); however, each introduces its own error and limitations (Gotway & Young, 2002).

2.1.5 The MAUP and Raster Data

Raster (i.e., gridded) data are often used when creating spatially explicit ecological models, as the data represent a continuous response variable. Raster data represent a particular instance of the MAUP, because a regular grid is arbitrarily imposed over a study area, defining the grain and extent of the data (Hay *et al.*, 2001; Lechner *et al.*, 2011). For raster data sets, the modifiable units are the individual pixels, and spatial resolution is determined by the area of the pixels (Jelinski & Wu, 1996). Users of raster data sets must be aware that the areal units are often preset and arbitrary, or limited by the physical design of the instrument system, and may not properly represent the phenomenon they are trying to capture (Soranno *et al.*, 2014). For example, the spatial resolution of remotely sensed data is determined by the mechanics of the sensor, rather than a grain size significant to the research to which it is applied (Dark & Bram, 2007; Lecher *et al.*, 2011). Results of spatial analysis may therefore represent the associations between the original units, rather than the phenomenon itself (Fassnacht *et al.*, 2006;

Lechner *et al.*, 2011; Soranno *et al.*, 2014). Remotely sensed data can neither provide more detail than this minimum mapping unit, nor accurately portray spatial patterns at a coarser resolution than the resolution at which it was captured (Hay *et al.*, 2001; Fassnacht *et al.*, 2005).

The risk associated with both MAUP and COSP when creating spatially explicit models is increased when using multiple spatial data sets (Lechner *et al.*, 2011). Selecting the most appropriate scale for analysis while assimilating disparate data sets is important because the spatial dynamics of a process at one resolution may be unimportant at another resolution (Gotway & Young, 2002). Often, data are spatially aggregated to make them compatible with other data sets. However, this is statistically problematic because variability of the original data values may be reduced, and because there is an inherent statistical instability in the transition between spatial resolutions (Dark & Bram, 2007). Methods and best practices for combining incompatible spatial data continues to be an active area of statistical research (Gotway & Young, 2002).

2.1.6 The MAUP as an Opportunity

“Thus the researcher whose model fits reality at a level that is less than perfect, as all models must, is left not knowing whether the misfit is due to the effects of spatial resolution, or due to an imperfection in the model, or both” –Goodchild, 2011

Both the MAUP and the COSP have been researched as “problems.” However, there is a well-accepted paradigm within ecology that physical and ecological processes naturally operate within different spatial and temporal scales (resolution and extent)

(Turner *et al.*, 1989; Dark & Bram, 2007). Within the framework of *macrosystems ecology*, researchers have begun to harness the MAUP or COSP “problems” to characterize the spatial characteristics of ecological processes. The resolution of a model is a property of observation and analysis, and only becomes fixed once the researcher sets spatial parameters (Hay *et al.*, 2001; Lechner *et al.*, 2011; Soranno *et al.*, 2014). Due to the unique spatial characteristics of ecological entities, the extent and grain size of a data set essentially filter the phenomena that are captured (Dark & Bram, 2007). Therefore, changing the resolution or extent of observation may better capture the phenomena being studied. Modifying areal units has the potential to characterize important information about the structure, function and dynamics of the phenomena being modeled (Jelinski & Wu, 1996; Hay *et al.*, 2001; Soranno *et al.*, 2014).

For example, a recent study by Griffith *et al.* (2015) found that at a 100 km spatial grain, the modeled spatial distribution of C₄ vegetation cover (which is directly correlated with $\delta^{13}\text{C}$) was more strongly correlated to the $\delta^{13}\text{C}$ isotopic ratios of bison tissue than to soil $\delta^{13}\text{C}$ isotopic ratios. Such correlations provide information about the spatial properties of the underlying ecological processes; i.e., the bison assimilate $\delta^{13}\text{C}$ of their environment at a spatial grain on the order of 100 km, while the processes driving assimilation of plant $\delta^{13}\text{C}$ in soil operate at much finer spatial resolutions (Griffith *et al.*, 2015). When the process being studied is influenced by factors that operate at a finer spatial resolution represented by the data set, the results of data analysis may ultimately be misleading (Goodchild, 2011). Conversely, including too much detail may lead to overwhelming complexity in the representation of the system being studied (Jelinski &

Wu, 1996), and aggregation may be necessary in order to create more meaningful units of analysis (Gotway & Young, 2002).

To mitigate the challenge of selecting an appropriate spatial grain for a model, there is an emerging trend in *macrosystems* and *landscape* ecology is to conduct multi-scale studies, such that spatial resolution (and/or extent) is the independent variable being tested (Turner *et al.*, 1989; Fei *et al.*, 2016). Multi-scale studies specifically sample data at multiple grain sizes, to characterize ecosystems at multiple spatial resolutions, while mitigating the effects of the MAUP. For spatial models that rely on data of a fixed sampling unit, creating models at multiple scales is a way of ensuring the best possible resolution is captured given the input data (Marceau *et al.*, 1994; Fassnacht *et al.*, 2006).

Additionally, researchers are commonly limited to using indirect data sets; however, when dealt with explicitly and appropriately, resolution has the potential to characterize the structure, function and dynamics of complex systems embedded in time and space (Jelinski & Wu, 1996; Hay *et al.*, 2001; Soranno *et al.*, 2014). Due to the paradoxical risk and potential presented by the MAUP, best practices for using spatial data in ecological studies should continue to be researched. In particular, I am interested in exploring the impacts of traditional geographic modeling “problems” such as the MAUP, the MTUP and the COSP on static, spatially-explicit ecological models by exploring the impact of spatial and temporal resolution on a terrestrial stable carbon isoscape for the continent of North America.

2.2 Stable Terrestrial Carbon Isoscapes

2.2.1 Isotopes and Isoscapes

Most elements found on a periodic table exist in several forms, containing the same number of protons and electrons, and possessing the same chemical properties, but with different numbers of neutrons and therefore unique atomic masses (Wassenaar, 2008; West *et al.*, 2010; Weathers *et al.*, 2013). The different number of neutrons cause isotopes to behave slightly differently from one another, resulting in differing isotopic concentrations (through a process called fractionation) as a function of space and time (Bowen, 2010). Fractionation, or the separation of isotopes into different concentrations, occurs in any system where physical or chemical processes produce an isotopic concentration different from the concentration of the input source (Bowen 2010). Isotopic values are expressed as δ , where $\delta = \frac{(R_{sample} - R_{standard})}{R_{standard}} \times 1000\text{‰}$, and compares the abundance of the heavy isotopic abundance to the light isotope (e.g., $^{13}\text{C}/^{12}\text{C}$) (Bowen, 2010; West *et al.*, 2010).

Because isotopic fractionation results in distinctive spatial and temporal patterns at landscape to global scales, and biotic entities tend to assimilate their environmental isotopic signatures, isotopic analysis has been used for a wide variety of spatially explicit applications, such as determining the source of a city's water supply, food and trade regulation, or deducing the origin of illicit drugs (West *et al.*, 2007; Ehleringer, 2000; Bowen, 2010; West *et al.*, 2010). The spatio-temporal distribution of isotopes can be represented as an "isotope" (a combination of the words isotopic landscape) (Bowen,

2010; West *et al.*, 2010). Isoscapes are spatially explicit, grid-based (i.e., raster) models, created either by combining data sets to model fractionating processes or by interpolating point data into a continuous grid surface (Bowen, 2010; West *et al.*, 2010). Global isoscapes of varying qualities have been created for stable H, C, N, and O (Bowen, 2010).

An important consideration that studies using stable isotopes often ignore is that isoscapes are not intensively sampled data. Rather, isoscapes are predictions modeled from a set of expectations (Wunder, 2010). Isoscapes describe general patterns of variation by filling in gaps found in observational data with predictions, but should not be treated as observational data themselves (Bowen, 2010; Wunder, 2010). Additionally, as spatially explicit models, isoscapes are subject to the standard geographic “problems” such as the MAUP, the MTUP, and the COSP. With this in mind, isoscapes have proven to be a powerful tool for identifying and characterizing spatial biogeochemical processes (Bowen, 2010).

2.2.2 Terrestrial Stable Carbon Isoscapes

In the case of stable carbon isotopes, the most abundant isotope, ^{12}C , has an average relative abundance of 98.90%, and the less common ^{13}C has a relative abundance of 1.10% (West *et al.*, 2010). All plants discriminate against the heavier carbon isotope ^{13}C during photosynthesis and integration of CO_2 into their structure. However, photosynthetic pathway directly effects the magnitude of ^{13}C discrimination (Δ) in carbon fixation (Lloyd & Farquhar, 1994; Still & Powell, 2010; Powell *et al.*, 2012). Relative to plants utilizing the C_3 photosynthetic pathway, C_4 plants are enriched in $\delta^{13}\text{C}$ (Still & Powell, 2010). In fact, the difference in isotopic discrimination between C_3 and C_4

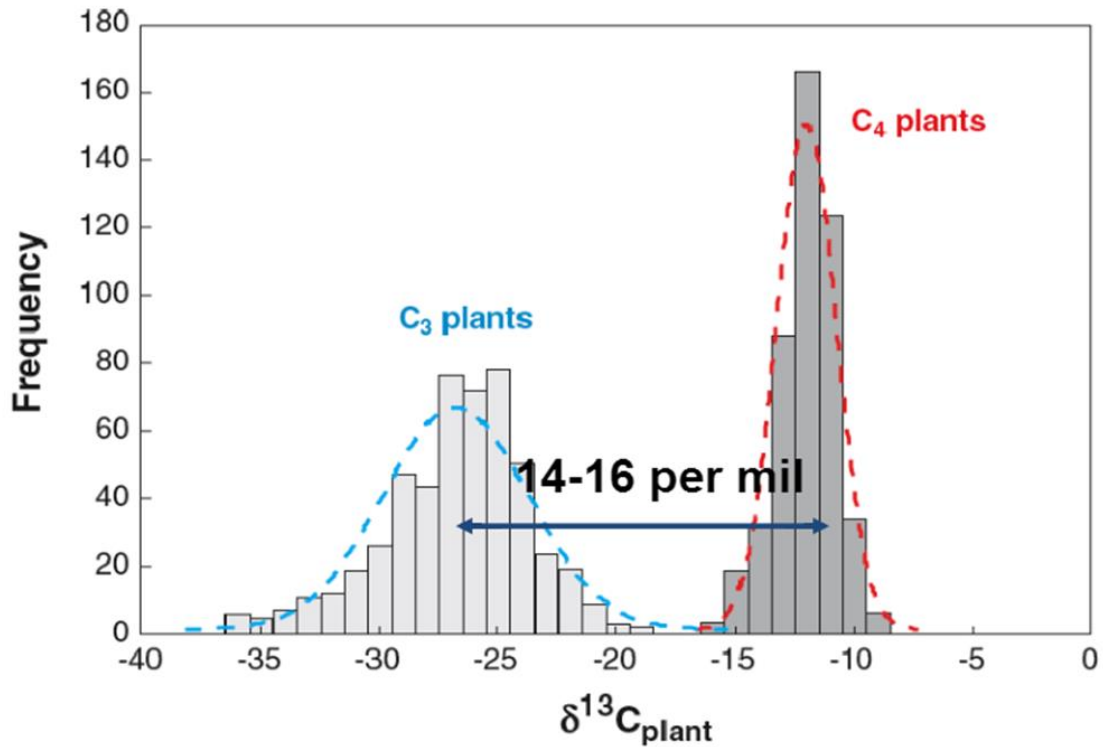


Figure 2 Frequency and ranges of leaf $\delta^{13}C$ ratios for globally sampled C_3 and C_4 plants (Cerling & Harris, 1999; Tiplle & Pagani, 2007).

photosynthesis is so large that the carbon isotope distributions of each photosynthetic pathway are almost entirely non-overlapping, as illustrated in *Fig. 2* (Ehleringer *et al.*, 1986; Tiplle & Pagani, 2007; Bowen & West, 2008). Because photosynthesis is the primary terrestrial carbon fractionating process, C_3 and C_4 biogeography is the primary determinant of terrestrial $\delta^{13}C$ spatial and temporal distribution (Cerling *et al.*, 1997; Bowen, 2010). Plant functional type distributions at macro-ecological (i.e., continental) spatial extents can be predicted by combining remote sensing data on vegetation life form, because C_4 plants are almost exclusively herbaceous, and climate data to model conditions in which C_4 grasses out compete C_3 grasses (Still *et al.*, 2003; Still & Powell, 2010; Powell *et al.*, 2012). Continental-extent terrestrial carbon isoscapes have been

generated at relatively high spatial resolutions (~10 km) for both Africa (Still & Powell, 2010) and South America (Powell *et al.*, 2012).

2.2.3 Modeling the Biogeography of Plant Functional Types

The first-order fractionation process for terrestrial stable carbon isotopes is photosynthesis, meaning spatial patterns of stable carbon isotopes on a landscape are determined by the spatial distribution of vegetation cover. In addition to being isotopically distinct, C₃ and C₄ plants are also favored under different climatic conditions (Farquhar *et al.*, 1989; Ehleringer & Monson, 1993; Sage *et al.*, 1999; Still *et al.*, 2003). The C₄ photosynthetic pathway likely evolved as an efficient response to the negative effects of photorespiration in hot climates (Ehleringer *et al.*, 1997). This adaptation causes C₃ and C₄ species to thrive in different environmental conditions (Teeri & Stowe, 1976; Murphy & Bowman, 2007; Von Fischer *et al.*, 2008). At higher temperatures, increased photorespiration causes C₃ plants to be less competitive, while at lower temperatures, the additional energy demands of the C₄ pathway to concentrate carbon and reduce photorespiration become a disadvantage (Ehleringer, 1978). This dynamic results in a *crossover temperature*, at which the two photosynthetic pathway have equal quantum yields; temperatures higher than this favor C₄ photosynthesis and temperatures lower than the crossover point favor C₃, as depicted in *Fig. 3* (Ehleringer, 1978; Ehleringer *et al.*, 1997; Collatz *et al.*, 1998; Still *et al.*, 2003; Powell *et al.*, 2012; Griffith *et al.*, 2015). Using climate data, the crossover temperature framework (referred to as the Collatz-crossover model) can be employed to model C₃ or C₄ dominance at a given location on a landscape (Collatz *et al.*, 1998; Still *et al.*, 2003; Griffith *et al.*, 2015). A major limitation

of the Collatz-crossover model is that does not predict the percentage belonging to each plant functional type, only which photosynthetic pathway (plant functional type) should be dominant given the climate (Still & Powell, 2010).

The climate metric used in previous terrestrial carbon isoscape models as a crossover temperature is a mean monthly temperature (22°C) with a minimum rainfall constraint (25 mm) to exclude desert and Mediterranean climates (Ehleringer *et al.*, 1997; Collatz *et al.*, 1998; Clark, 1998; Still *et al.*, 2003). More recently, research has indicated that a better climate metric for a crossover temperature might be a monthly maximum temperature (27°C) with the same minimum rainfall constraint (25 mm), as a monthly maximum temperature does not average in overnight low temperatures.

Most C₄ plants are herbaceous, and about 60% are grasses (Teeri & Stowe, 1976; Teeri *et al.*, 1980; Edwards *et al.*, 2010). Therefore, it is important to consider vegetation growth forms in a plant functional type model (Still & Powell, 2010). This can be accounted for with a remotely sensed vegetation continuous fields (VCF) product, which is a MODIS (Moderate Resolution Imaging Spectroradiometer) satellite-based collection

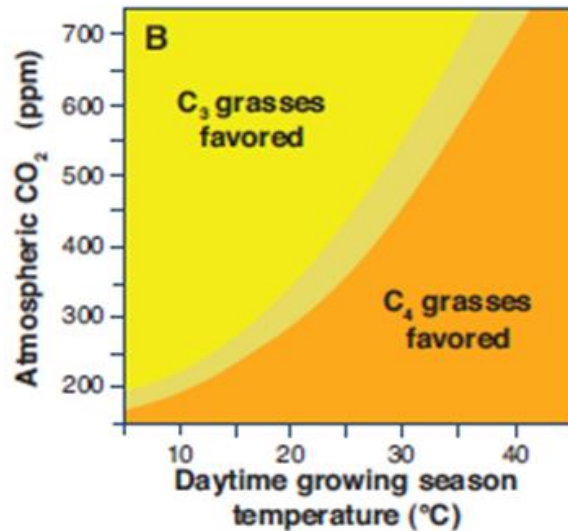


Figure 3 Threshold for C₃/C₄ dominance in relation to environmental temperature and atmospheric CO₂ concentration (Edwards *et al.*, 2010).

of data layers containing the proportional estimates for vegetation cover types: woody vegetation, herbaceous vegetation, and bare ground (e.g., DeFries *et al.*, 1995) (Still & Powell, 2010; Powell *et al.*, 2012).

Finally, due to the heavy impact humans have on landscapes, it is important to consider the spatial distribution of croplands in a plant functional type model (Still & Powell, 2010), as crop types do not always match the climatic conditions that control the distribution of natural herbaceous plant functional types. For example, the Great Plains biome once covered the middle third of the North American continent; however it has been almost entirely converted into row-crop agriculture (Sage *et al.*, 1999). Where corn (C₄) is often planted in C₃ climates, the converse is also true, where soy bean (C₃) is found to be planted broadly in formerly C₄-dominant grasslands, such as the savannas of Brazil (Leff *et al.*, 2004). Croplands can be accounted for by data sets that combine remotely sensed data with agricultural statistics to depict the spatial distribution of crop types (e.g., Ramankutty *et al.*, 2008; Monfreda *et al.*, 2008).

2.2.4 Challenges of Modeling Mid-Latitude Grasslands

Modeling the vegetation cover of North America presents a challenge because vegetation cover varies both inter- and intra- annually. Because vegetation cover is the first order determinant of terrestrial stable carbon isotope fractionation over a landscape, this challenge is pertinent for creating a $\delta^{13}\text{C}$ isoscape. In tropical regions, climate conditions are relatively static year round, and the spatial distribution of C₃ and C₄ dominant grasses is also temporally consistent. The seasonal variation of mid-latitudes results in intra-annual variability of relative C₃/C₄ abundance, and mid-latitude grasslands

are often also temporally variable at an intra-annual extent. This section describes some vegetation patterns found on the continent of North America that may be difficult to capture in a static model.

At latitudes where the Collatz *cross-over temperature* is reached for only part of the year, C₃ and C₄ grass abundances are often temporally offset (Ehleringer & Monson, 1993). This is observed in the Great Plains grasslands, where a mild spring followed by hot summer creates conditions where C₃ and C₄ grass floras geographically co-occur, but relative abundancies shift though the course of the season (Sage *et al.*, 1999), as C₃ grasses emerge earlier in the growing season and then are replaced by C₄ grasses during the hot summer months (Ode *et al.*, 1980, Monson *et al.*, 1983). This offsets direct competition because each plant functional type effectively has a unique growing season, and therefore captures environmental resources at different points in the year.

Deserts in North America can also have shifting seasonal compositions of plant functional type based on climatic conditions. For example, in the Chihuahuan Desert, winter rain promotes C₃ grasses and summer rain favors C₄ grasses (Kemp, 1983). Precipitation is equally likely for each season, and severe seasonal droughts occur in the spring and autumn, punctuating both the winter and summer growing seasons. This weather pattern prevents one community from persisting longer than a single growing season due to a lack of available moisture, and therefore prevents one community from becoming annually dominant over another.

In some parts of Texas, temporal offsetting does not occur as it does in the Great Plains. The long and hot summers favor C₄ plants to such an extent that a dense turf is

created, capturing space and nutrients during the winter and excluding C₃ grasses entirely even though the temperatures are mild enough in the winter for C₃ plants to be favored. Due to this phenomenon, some C₄-dominated grasslands in the southernmost portions of the Great Plains become active at the same time as C₃ dominated grasslands in the northern regions (Tieszen *et al.*, 1997).

Because C₃/C₄ grasses are sensitive to when precipitation occurs, there is also inter-annual elasticity in the relative composition of C₃/C₄ grasses in the Great Plains (Sage *et al.*, 1999; Winslow *et al.*, 2003; Murphy & Bowman, 2007). When the summer growing season is dry, C₃ grasses have an advantage and flourish during the mild spring; however when the spring is dry and the summer wet, C₄ species have an advantage, shifting overall composition towards C₄ dominance (Monson *et al.*, 1983; Winslow *et al.*, 2003; Murphy & Bowman, 2007).

Recent studies suggest that capturing plant phenology via remote sensing can be an effective tool for determining the relative abundance and extents of C₃ and C₄ communities with asynchronous growing seasons (Tieszen *et al.*, 1997; Davidson & Csillag, 2003; Wang *et al.*, 2011; Wang *et al.*, 2013; Zhong *et al.*, 2015). However, most phenological plant functional type models ultimately produce a single static output, and do not account for seasonal variability in the final representation.

Generating a stable carbon isoscape for the continent of North America presents an opportunity to explore methods for representing temporally varying phenomena (e.g. seasonal vegetation cover) in a static model. In this work, I will test the impacts of both temporal and spatial resolution of a terrestrial stable carbon isoscape for North America. I

intend to apply the models in a geographic origin assignment exercise to test the performance of each isoscape variation's assignment prediction against a validation data set of known-origin tissue locations.

2.3 Using Isoscapes for Geographic Origin Assignment

2.3.1 Isotopes and Animal Migration

“The movement of organisms in space and time defines their interaction with their environments and, therefore, comprises a fundamental aspect of their ecology and evolutionary history” -Hobson & Norris, 2008

Because animals directly assimilate the isotopic composition of their diet into their tissues, a common application of isoscapes is wildlife or human forensics. This includes characterizing long-distance migration by relating tissue isotopes to the isotopic patterns across a landscape, and predicting (assigning) a geographic origin to the tissue (Wunder, 2010). Geographic origin assignment can be leveraged as validation method for isoscapes by comparing assignment predictions to a tissue's known origin (Vander Zanden et al., 2014; Vander Zanden et al. 2015). This is the strategy I will employ to evaluate the impacts of spatial and temporal resolution of stable terrestrial carbon isoscapes of North America.

Characterizing the ranges or environmental connectivity of animals that migrate long distances can be challenging due to costs and/or physical limitations (Hobson & Norris, 2008; Wunder, 2010; Hobson *et al.*, 2010). Using stable isotopes as intrinsic spatial markers has many advantages over the use of extrinsic tools such as tags, rings or

transmitters (Hobson & Norris, 2008). Isotopic analysis does not require initial capture and marking of an organism, reducing the statistical bias that is introduced in extrinsic methods in which animals must be sampled at one site, then later recaptured at a second location after migration (Wassenaar, 2008; Wunder, 2010). Many migratory animals such as insects or songbirds are simply too small to carry extrinsic markers such as rings or tags. Additionally, current isotopic sampling methods on metabolically inactive tissues are typically non-invasive (Vander Zanden *et al.*, 2014).

Using stable isotopes to analyze migratory origin assignment has been applied to organisms ranging from insects to mammals in both aquatic and terrestrial environments. A large number of migration studies using stable isotopes has focused on birds because they are relatively easy to capture and sample non-destructively (Rubenstein & Hobson, 2004). For stable isotopes to function as a tracer for animal migration, the tissues of the animal must contain one or more of the isotopes of interest, and the organism must migrate between isotopically distinctive landscapes, while retaining a record of the isotopic signatures they travel through (Wassenaar, 2008).

2.3.2 Isotopes in Animal Tissues

Different tissues represent different time periods in an animal's life. Animals have both metabolically active and metabolically inert tissues. Inert tissues (i.e. keratinous tissues) are tissues that do not change isotopically and therefore represent well-defined growth periods and are isotopically stable once fully formed (Rubenstein & Hobson, 2004; Wassenaar, 2008). Inert tissues are typically used for long-distance migration studies. In order to accurately connect tissues with a temporal period, the ecology and

physiology of the organism must be well documented. The growth period of a tissue is key for characterizing the physical space that tissue represents. The temporal aspect of tissue growth defines the physical space integrated by an animal's tissue (Wassenaar, 2008). For example, slow growing feathers or claws may integrate the stable isotope signature of larger areas (depending on the migratory angle of the organism) than more rapidly growing tissues.

2.3.3 Statistical Assumptions of Geographic Origin Assignments

Geographic origin assignment models are founded on two basic statistical assumptions. First is the assumption of statistical independence; i.e., that samples are drawn randomly from the target population of interest. Second is the assumption of process homogeneity; i.e., that all individuals within a population in the same location are subject to the same processes that generate isotopic variance (Wunder & Norris, 2008). The assumption of process homogeneity is almost never met in reality, as measured isotopic variability is introduced from a number of different sources, e.g., differences between individuals, isotopic heterogeneity within an individual, and variance in isotope-ratio mass spectrometry (IRMS) measurement error (Wunder & Norris, 2008). In instances with a clustered sampling design, the sample site is a better sampling unit than individuals because stable isotope values are related to geographic variables, and groups of individuals from the same location should share the same response variable (Wunder & Norris, 2008).

2.3.4 Probability Surfaces Framework

Although multiple analytical frameworks have been developed for geographic origin assignment using isoscapes, a frequently used method in recent studies is the probability surfaces model (Wunder *et al.*, 2005; Wunder & Norris, 2008; Wunder, 2010; Bowen *et al.*, 2014; Vander Zanden *et al.*, 2014; Vander Zanden *et al.* 2015). The probability surfaces model is a structure for isotope-based geographic assignment using a semi-parametric Bayesian framework, such that the output probability surface depicts the relative probability that any point in space is the true origin of a given tissue sample (Wunder & Norris, 2008).

The first step in performing a geographic origin assignment is to obtain a continuous surface representation of environmental stable isotope values (i.e., one or more isoscapes), which provides the spatial distribution of environmental stable isotope ratios across a landscape that individuals are related to (Wunder & Norris, 2008; Bowen *et al.*, 2014). Applying Bayes' theorem to determine the relative probability (A_i) that site i is the true location of origin for a tissue sample with a measured isotopic composition of δ_s yields *Equation 1*, where $P(A_i)$ is the prior probability associated with location i , and the denominator is the total sum of the probabilities across all possible locations (Wunder *et al.*, 2005; Bowen *et al.*, 2014).

$$P(A_i|\delta_s) = \frac{P(A_i)P(\delta_s|A_i)}{\sum P(A_i)P(\delta_s|A_i)} \quad \text{Equation 1}$$

To account for isotopic fractionation that occurs through the food web and assimilation into the fully formed tissue, a rescaling function is generated by relating known-origin sample values to the isoscape surface (Wunder, 2010; Bowen *et al.*, 2014). The general form of the isotopic rescaling function is provided in *Equation 2*, where the value of a sampled individual at a given geographic location (δ_s) is related to the modeled isoscape value δ_i , and ε is the difference between the function's predicted value for the location and the sampled feather (Hobson *et al.*, 2012; Bowen *et al.*, 2014; Wunder, 2010). The rescaling function is both organism and isoscape specific and accounts for the fractionation that occurs through the food web during the growth of the sampled tissue, as well as local variance of stable isotope values at the sampling site. The rescaling function generates a second isoscape that is calibrated (or rescaled) to the sampled tissues being assigned (a tissue isoscape) (Wunder & Norris, 2008).

$$\delta_s = f(\delta_p) + \varepsilon \tag{Equation 2}$$

The final step is to create a probability of origin surface for each individual being assigned. Per-pixel relative probabilities are calculated by relating the calibrated tissue isoscape to the individual sample (δ_s) (*Equation 3*). The output is a probability of origin surface, in which each raster cell is assigned a value between 0 and 1, corresponding to the likelihood that the individual sampled originated from that location. The probability of origin surface is rescaled so that all probability values included in the surface sum to a

value of one (Wunder & Norris, 2008). Essentially, geographic assignment is an evaluation of the probability that the sampled isotopic composition of tissue δ_s would be observed at any location i (Wunder *et al.*, 2005, Wunder *et al.*, 2010; Bowen *et al.*, 2014).

$$P(A_i|\delta_s) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{\left(\frac{-(\delta_s - (\beta_0 + \beta_1\hat{\delta}_{p,i}))^2}{2\sigma_i^2}\right)} \quad \text{Equation 3}$$

The probability surface model accounts for variance from multiple sources (e.g., per-pixel variance of isoscape estimates, between individuals at a given location, and analytical error). The probability surface model can also be applied to multiple isotopes and provides a range of potential assignments for a given individual (Wunder & Norris 2008). The biggest disadvantage to the probability surface model is that it is computationally intensive (Wunder & Norris, 2008). However, the probability surface model is the method included for geographic assignment analysis in the online cyber-GIS system IsoMAP (Isoscapes Modelling, Analysis and Prediction; <http://isomap.org>) (Bowen *et al.*, 2014). Additionally, R code for geographic origin assignments was recently published by Vander Zanden *et al.* (2014).

2.3.5 Using the Spatial Distribution of Carbon Isotopes for Geographic Origin Assignment Research

Stable carbon isotope ratios vary in time and space and are primarily determined by plant functional type biogeography. When isotopes are used as tracers for migration

studies, organisms are assumed to be consistent isotopic integrators of the environment. However, this assumption may not always hold for $\delta^{13}\text{C}$ if the species selectively consumes a specific plant functional type. For herbivores, tissue $\delta^{13}\text{C}$ concentrations reflect the C_3 and C_4 ratio present in their diet. $\delta^{13}\text{C}$ in long-term inert tissues (mammal teeth), have been used to characterize evolutionary relationships of mammalian grazers and grasslands (Cerling *et al.*, 1997; MacFadden, 2000; Cerling *et al.*, 2003; West *et al.*, 2006). Stable carbon isotopes therefore can be used for geographic origin assignment if the animal migrates between distinctive C_3 and C_4 habitats and does not integrate plant functional types selectively (Rubenstein & Hobson, 2004; West *et al.*, 2006).

Although the majority of isotopic migration studies using $\delta^{13}\text{C}$ have been applied in marine environments, stable carbon has been employed as an isotopic tracer in a few bird migration studies (Rubenstein & Hobson, 2004; Bowen, 2010). In a study on Ecuadorian hummingbirds, Hobson *et al.* (2002) measured a $\delta^{13}\text{C}$ response in tail feathers that was consistent with plant functional type distributions in relation to altitude (Hobson *et al.*, 2002). A multi-isotope origin assignment conducted by Hobson *et al.* (2012) on Afro-tropical migrating birds identified some limitations in using $\delta^{13}\text{C}$. The grain of the $\delta^{13}\text{C}$ isoscape model used was spatially coarse, which reduced the representation of habitats that are rare or highly localized. Additionally, the authors proposed that improved understanding of the composition and origins of species-specific diets would greatly improve origin assignment. For example, knowing if the birds are integrating a diet of local insects from specific habitats versus consuming insects from many different habitats over a large spatial extent would assist in selecting the appropriate grain for

migratory origin assignment (Hobson *et al.*, 2012). This relates to the MAUP, as selecting an optimal spatial resolution (i.e., one corresponding to the spatial area integrated by the organism through diet) could potentially improve geographic assignment models.

2.3.6 Geographic Origin Assignment and Inter-Annual Temporal Variation

Often, the isoscapes used in geographic origin assignments are long-term averages of multiple growing seasons, which do not necessarily reflect a biologically relevant time period when dealing with animals that assimilate isotopes from their food web within a single growing season (Vander Zanden *et al.*, 2014; Tonra *et al.*, 2015; Vander Zanden *et al.*, 2015). More recently, studies have explored the impacts of inter-annual variation on geographic origin assignments using $\delta^2\text{H}$ isoscapes. Results of these recent studies by Vander Zanden *et al.* (2014) and Tonra *et al.* (2015) suggest that temporal extent does not have a significant impact on the accuracy and precision of origin assignments (Vander Zanden *et al.*, 2014, Tonra *et al.*, 2015, Vander Zanden *et al.* 2015). However, this is still a topic of active research, and each of these recent studies have explored inter-annual variation only. The impacts of intra-annual variation, as well as spatial grain, have not yet been examined in this context and present an opportunity for further research.

CHAPTER 3: A Terrestrial Stable Carbon Isoscape of North America

3.1 Introduction

What is the spatial distribution of vegetation stable carbon isotopes on the continent of North America?

The natural variability of stable carbon isotope composition ($\delta^{13}\text{C}$) in the environment has been studied and harnessed for a wide variety of applications ranging from characterizing sources and sinks for biosphere-atmosphere CO_2 exchange to characterizing animal migration (Suites *et al.*, 2005; Wunder *et al.*, 2005; Bowen, 2010; Hobson *et al.*, 2012; Powell *et al.*, 2012). The spatio-temporal patterns of $\delta^{13}\text{C}$, represented as an isoscape (isotopic landscape), provide opportunities for spatially explicit ecological research. Measuring vegetation $\delta^{13}\text{C}$ can determine the relative contribution of plant functional types (i.e., C_3 vs. C_4) to biomass at a given location, or measuring tissue $\delta^{13}\text{C}$ can determine the relative contribution of plant functional types to an animal's diet (Cerling *et al.*, 2003; Still *et al.*, 2003; Suites *et al.*, 2005; Bowen, 2010; West *et al.*, 2010). In fact, because $\delta^{13}\text{C}$ is integrated to tissue through diet with little modification, in addition to identifying diet it can be used to determine trophic niches (Hobson, 2011; Cerling *et al.*, 2003), or even approximate the location the tissue was grown (Wunder *et al.*, 2005; Hobson *et al.*, 2012). Spatially explicit soil $\delta^{13}\text{C}$ has been

used to reconstruct prehistoric shifts in woody and herbaceous vegetation on a landscape and determine the environmental conditions hominids evolved in (Cerling *et al.*, 2011). Although several models of C₃/C₄ composition have been generated at a global scale (Lloyd and Farquhar, 1994; Still *et al.* 2003; Suits *et al.*, 2005), as well as $\delta^{13}\text{C}$ isoscapes at finer resolutions (~10 km) for the continents of Africa and South America (Still & Powell, 2010; Powell *et al.*, 2012), a fine spatial resolution terrestrial $\delta^{13}\text{C}$ isoscape has not yet been modeled for North America. In this section, I generate an annual $\delta^{13}\text{C}$ isoscape for North America at 5-minute spatial resolution.

The methods used to generate the Africa and South America isoscape predict $\delta^{13}\text{C}$ by modeling environmental conditions rather than interpolating spatially explicit observations. Generating a $\delta^{13}\text{C}$ for the continent of North America using the same methods provides an opportunity to further test the model assumptions with reference data due to the availability of observational data spanning the continent. Additionally, because North America spans mid-to-high latitudes, C₃/C₄ composition of grasslands varies seasonally. Thus, the impact of averaging seasonally varying communities can be explored and compared against different reference data sets, such as soil organic matter (SOM, which represents a long term average of isotopes in an area), or data collected from vegetation surveys. The North American $\delta^{13}\text{C}$ isoscape will provide a new research tool to the stable isotope community that compliments the continental scale isoscapes that have been created for Africa (Still & Powell, 2010) and South America (Powell *et al.*, 2012).

The initial goal of my thesis research is to generate a moderate spatial resolution terrestrial $\delta^{13}\text{C}$ isoscape following the general methods established by Still & Powell, 2010; and Powell *et al.*, 2012. The results of this chapter inform the methods for research questions 2 and 3, (Chapters 4 and 5, respectively) by determining the optimal parameters to use for modeling the spatial distribution of stable carbon isotopes in North America.

3.2 Methods

The first-order determinant of the spatio-temporal distribution of $\delta^{13}\text{C}$ on a landscape is the relative composition of plant functional types (i.e., photosynthetic pathway) contributing to total vegetation. Therefore, in order to generate a $\delta^{13}\text{C}$ isoscape I first model the relative C_3/C_4 proportion per 5-min pixel for the extent of the North American, and then apply $\delta^{13}\text{C}$ endmembers to the modeled per-pixel relative distribution of plant functional types (Still & Powell, 2010; Powell *et al.*, 2012). To predict the spatial distribution of plant functional types, I make three major assumptions. (1) All C_4 vegetation is herbaceous; therefore, the first step separate woody cover from herbaceous cover. (2) The distribution of crop plant functional types do not follow climate rules, so the distribution of agricultural crops (herbaceous) must be accounted for using an agricultural lands data set. (3) The dominant photosynthetic pathway of natural herbaceous cover is determined by climate rules (i.e., the crossover temperature), and therefore natural herbaceous cover is partitioned based on these rules.

The input data sets used to predict the relative abundance of vegetation life forms are the same data sets used by Powell *et al.* (2012), and are listed in *Table 1*. The Moderate Resolution Imaging Spectroradiometer (MODIS) Vegetation Continuous Fields (VCF), Climatic Research Unit (CRU) 2.0 climatology, and agricultural lands layers are all modeled or interpolated gridded data sets, and introduce error and uncertainty into the final isoscape. The $\delta^{13}\text{C}$ composition of vegetation is predicted using endmember values reported in the literature for trees, shrubs, C_3 grasses and C_4 grasses. The endmembers are applied to the modeled relative abundance of vegetation woody vegetation, C_3 herbaceous and C_4 herbaceous to generate a stable terrestrial carbon isoscape.

Table 1 Input data sets used to model terrestrial stable carbon isoscape with original spatial resolutions. Adapted from Powell *et al.*, 2012.

<i>Dataset</i>	<i>Year</i>	<i>Spatial Resolution</i>	<i>Citation</i>
MODIS Vegetation Continuous Fields Yearly L3 Global Collection 3 (MOD44B)	2001	500 m	Hansen <i>et al.</i> , 2003
MODIS Land Cover Type Yearly L3 Global Collection 5 (MOD12Q1)	2001	500 m	Friedl <i>et al.</i> , 2010
CRU 2.0 Climatology 1961-1990		10 min	New <i>et al.</i> , 2002
Agricultural Lands in the Year 2000 (M3-Cropland Data)	2000	5 min	Ramankutty <i>et al.</i> , 2008
Harvested Area and Yields of 175 Crops (MB-Crops Data)	2000	5 min	Monfreda <i>et al.</i> , 2008
MODIS Vegetation Indices Monthly L3 Global Collection 5 (MOD13A3)	2000	1 km	Solano <i>et al.</i> , 2010

I assess variations of two modeling parameters: the crossover temperature metric (mean monthly temperature versus monthly maximum temperature) used to discriminate between areas of C_3 and C_4 dominance, and the decision rules used to partition woody and herbaceous cover based on the MODIS VCF data set (rules that assign minimum vs. maximum herbaceous cover in mixed-classification regions). Testing each combination of these two model parameters results in four isoscape iterations with different relative C_3/C_4 herbaceous cover in pixels that contain natural grasslands.

3.2.1 Study Area

The study area is constrained to the conterminous landmass of North America. The area extends as far south as the Panama Canal, as that is the northern cutoff for the South America carbon isoscape generated by Powell *et al.* (2012), and as far north as the Arctic Circle. Under the current climatic conditions, latitudes higher than 60°N do not have C₄ grasses, excluding some select species that are found in Alaska and northwestern Canada (Sage, 1999). However, creating a hard cutoff at 60°N might limit future analyses or modeling change. Therefore, the northern bounding latitude is 66.5°N, which includes the majority of North American landmass.

3.2.2 Estimating %woody and %herbaceous Vegetation Cover

The first step to generating a $\delta^{13}\text{C}$ isoscape is to determine the per-pixel percent of woody and herbaceous vegetation cover. The MODIS Vegetation Continuous Fields (VCF) Yearly L3 Global Collection 3 for the year 2001 includes layers that represent per-pixel percent tree cover, percent herbaceous cover, and percent bare surface at 500-m spatial resolution (Hansen *et al.*, 2003). However, the VCF data product was designed to map percent tree cover, and a major limitation of the VCF “herbaceous” layer is that it does not discriminate between woody shrubs and grasses, which is important for discriminating between C₃ and C₄ vegetation (Hansen *et al.*, 2003; Still *et al.*, 2003; Powell *et al.*, 2012). Additionally, the VCF layers represent the percent of canopy cover per grid cell, which results in an overestimation of the herbaceous layer, as maximum crown cover is fixed at 80% (Hansen *et al.*, 2003). To address these shortcomings, I convert the VCF percent tree-cover (i.e., canopy cover) layer to percent tree-crown

($\%tree-crown = \%tree-cover/0.8$), and create a percent non-tree layer ($\%non-tree$) by subtracting the adjusted percent tree-crown layer and the VCF percent bare layer from 100, assuming everything that is not tree crown or bare is non-tree ($\%non-tree = 100 - \%tree-crown + \%bare$) (Powell et al. 2012). This step results in a final $\%tree-crown$ layer, and an intermediate $\%non-tree$ vegetation layer, to be further classified into herbaceous, shrub and crop layers.

Next, I divide the intermediate $\%non-tree$ layer into shrub ($\%shrub$) and herbaceous ($\%herbaceous$) layers by using the land-cover descriptions for the 17-class International Geosphere-Biosphere Programme (IGBP) land-cover classification,

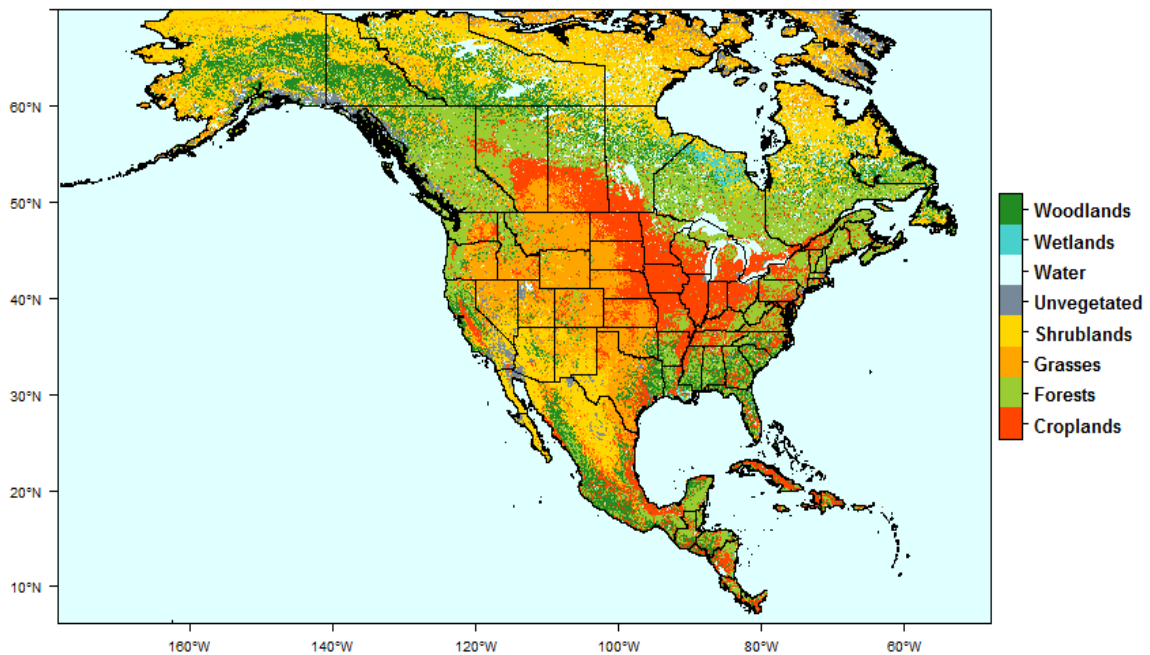


Figure 4 Collapsed MODIS IGBP land-cover classification for the year 2001. The “Shrublands” and “Woodlands” classifications include the three IGBP classes that are identified as a mixed classification containing both shrubs and grasses (i.e., “open shrubland”, “woody savanna” and “savanna”). In North America, areas falling into these classifications are constrained to the Chihuahuan Desert region and Northern Canada/Alaska

included in the MODIS 500-m Global Land Cover Type product for 2001 (Loveland & Belward 1997). A generalized IGBP classification schema for North America (2001) is depicted in *Fig. 4*. The *%non-tree* layer pixels that correspond to the land-cover classes described as predominantly shrub and/or tree by the IGBP classification system are assigned to the *%shrub* layer, and for pixels that correspond to classes composed predominantly of herbaceous cover, the *%non-tree* layer is assigned as *%herbaceous*. For pixels that correspond to IGBP classes composed of a mixture of shrubs and grasses (i.e., “open shrub-land”, “woody savanna”, and “savanna”) *%non-tree* layer is partitioned into *%shrub* and *%herbaceous* following rules described by Powell *et al.* (2012). The maximum-herbaceous and minimum-herbaceous rules result in two sets of *%shrub* and *%herbaceous* layers, one assuming maximum possible herbaceous cover and one assuming minimum herbaceous cover. I refer to the maximum herbaceous layers as “max-herbaceous” and the minimum herbaceous as “min-herbaceous” layers using each respective set of mixed classification rules. Each of these layers, as well as the *%tree-crown* layer are spatially aggregated and resampled in R to a 5-min spatial resolution, compatible with the crop type data layer (Powell *et al.*, 2012).

3.2.3 Accounting for Managed Agro-Ecosystems

To account for the managed agro-ecosystems, (i.e., C₄ crops grown in C₃ climates, and vice-versa), the *%herbaceous* layer is first partitioned into *%natural-herbaceous* and *%crop* layers using the Cropland 2000 data set, which provides per-pixel cropland percentage at a 5-min resolution (Ramankutty *et al.*, 2008). Following the methods established by Powell *et al.* (2012), I make two simplifying assumptions: that

cropland only contains herbaceous vegetation, and that the *%tree-crown* layer is more accurate than the *%shrub* and *%herbaceous* layers. Therefore, incorporating the cropland layer does not change the *%tree-crown* layer. Instead, the *%herbaceous*, *%shrub* and *%crop* values are adjusted to change on a per-pixel basis. The integration of the Cropland 2000 data results in new *%crop* and *%natural-herbaceous* data layers, as well as a final *%shrub* data layer. I then apply a second global data set delineating crop type (Monfreda *et al.*, 2008) to separate the *%crop* layer into either C₃ or C₄ vegetation (Powell *et al.*, 2012). The final outputs from this step are *%C₃-crop* and *%C₄-crop* layers, and a *%natural-herbaceous-adjusted* layer (Powell *et al.*, 2012).

3.2.4 Partitioning the Natural Herbaceous Layer into C₃ and C₄ Layers

The relative C₃/C₄ composition of the *%natural-herbaceous-adjusted* layer is determined by climate. Regions are classified as favorable to C₄ vegetation based on a cross-over temperature metric applied to a global climate data set (Daly *et al.*, 2002; New *et al.*, 2002). Two cross-over temperature metrics were tested: mean monthly temperature (>22°C; Collatz *et al.*, 1998) and maximum monthly temperature (>27°C; Griffith *et al.*, 2015). I refer to each of these parameters as “mean-temperature” and “max-temperature”, respectively. The number of months each pixel meets each constraint is presented in *Fig. 5*.

In terms of an annual classification, pixels that are classified by the crossover-temperature metric as C₄-favorable for 6 or more months out of the year are classified as C₄-dominant. Pixels that have 0 months classified as being C₄-favorable are classified as entirely C₃-dominant. Pixels that classified as having at least one month, but less than six,

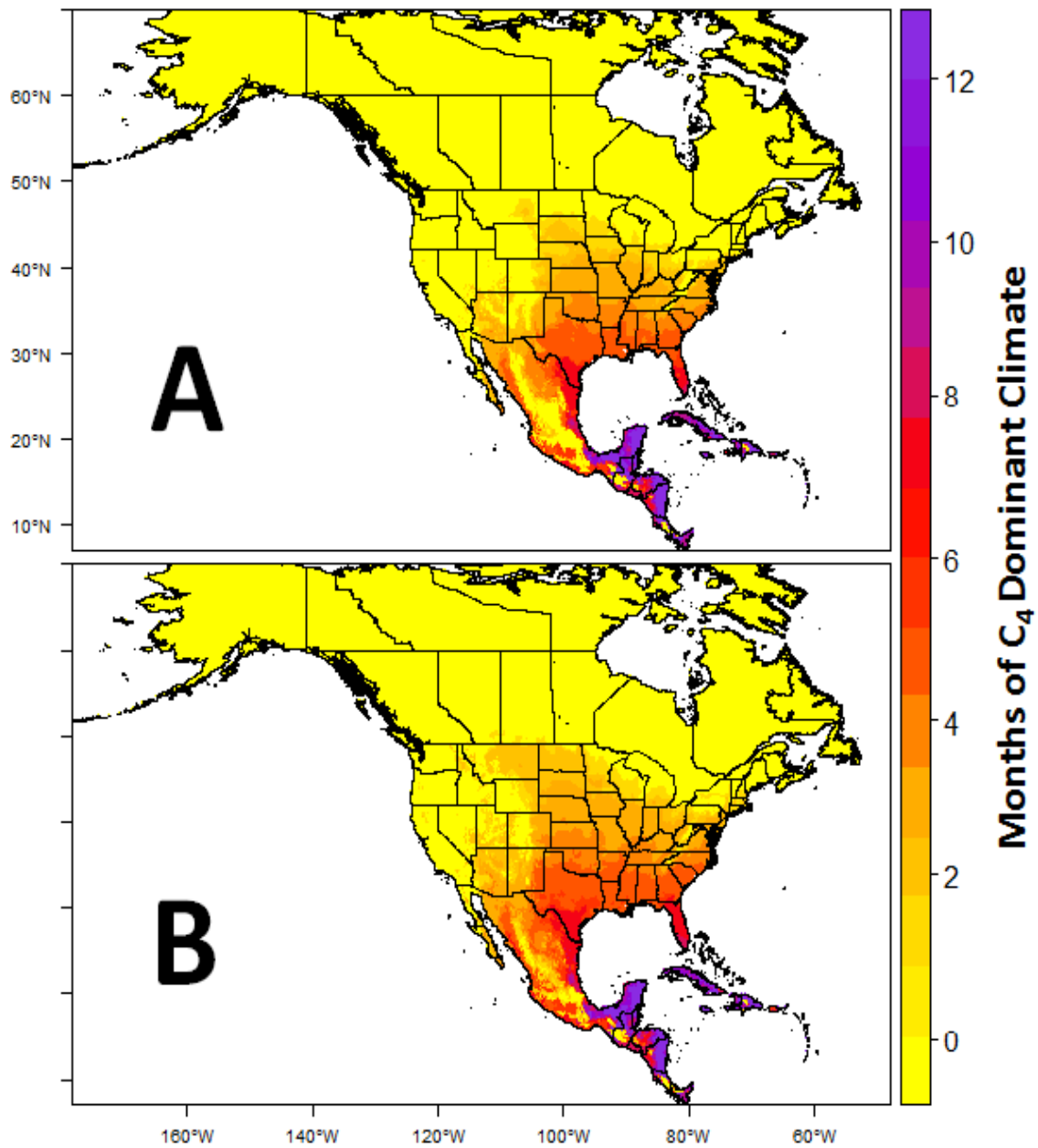


Figure 5 Total number of months each 5-minute raster pixel meets the C₄-dominant conditions. These conditions are modeled as either a mean-temperature $\geq 22^{\circ}\text{C}$ and mean monthly precipitation $> 25\text{mm}$ (Collatz *et al.*, 1998) (A) or a monthly maximum-temperature $\geq 27^{\circ}\text{C}$ and mean monthly precipitation $> 25\text{mm}$ (Griffith *et al.*, 2015) (B).

that are C₄-favorable are expected to be areas where C₃ and C₄ grasses co-exist (i.e., mixed grasslands) and are classified as mixed.

For pixels classified as “mixed”, I create a ratio of C₄ plants to total vegetation abundance. Following the methods of Still *et al.* (2003) and Powell *et al.* (2012), I assume that C₄ percent cover of a pixel is proportional to the ratio of vegetation productivity (approximated by the MODIS Normalized Difference Vegetation Index, or NDVI) in C₄-climate months to the total growing season vegetation productivity. The per-pixel growing season is defined as those months with mean temperature >25°C (or maximum temperature >27°C) and mean precipitation > 25mm, because most grasslands require at least this much precipitation to grow (Collatz *et al.* 1998; Powell *et al.*, 2012). This process results in two sets of C₃ and C₄ mask layers, one set corresponding to each cross-over temperature metric. Each cross-over temperature metric is applied to each version of the land cover rules of the *%natural-herbaceous-adjusted* layer (max-herbaceous and min-herbaceous), resulting in four variations of *%C₃-natural-herbaceous* and *%C₄-natural-herbaceous* layers. The sum of the *%C₃-natural-herbaceous*, *%C₃-crop*, *%C₄-natural-herbaceous*, *%C₄-crop*, and *%shrub* layers is equal to the adjusted *%non-tree* vegetation layer generated from the VCF data set.

Finally, I generate layers that represent vegetation life form by photosynthetic pathway. The final *%C₄-herbaceous* layer is generated by summing the *%C₄-natural-herbaceous* and the *%C₄-crop* layer. Likewise, the final *%C₃* layer is generated by summing the *%C₃-natural-herbaceous* and *%C₃-crop*. The final *%woody-cover* layer is generated by combining the *%tree-crown* and *%shrub* layers. The layers for the

maximum-herbaceous land-cover rule and mean monthly crossover temperature constraint are presented in *Fig. 6*.

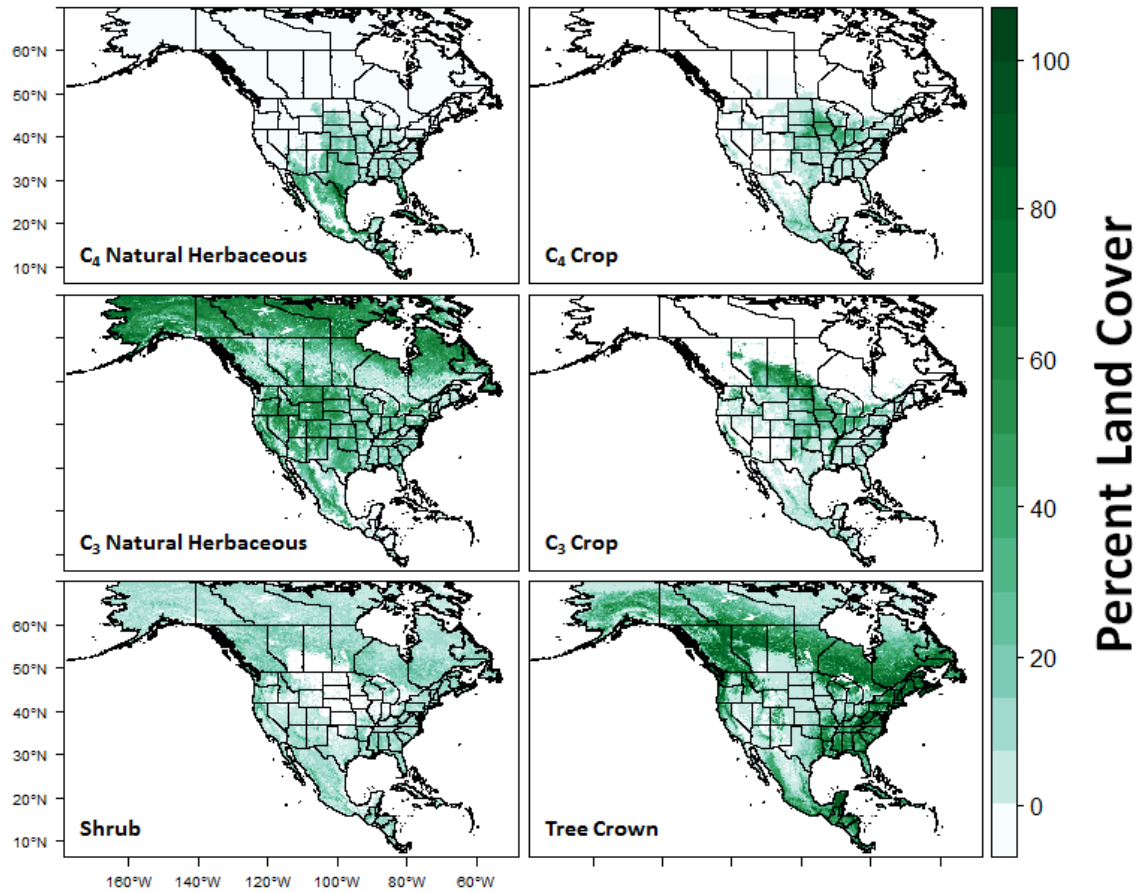


Figure 6 Individual output layers representing the maximum herbaceous land-cover rule and mean-temperature crossover climate parameter. Each map layer depicts the percent area within each 5-min pixel covered by the specific vegetation type.

3.2.5 Estimating Leaf $\delta^{13}\text{C}$

Leaf $\delta^{13}\text{C}$ is calculated based on the vegetation life form layers generated in the previous steps, and corresponding $\delta^{13}\text{C}$ endmember values from the literature (Powell *et al.*, 2012; *Table 2*). To apply $\delta^{13}\text{C}$ endmembers, the combined percent-cover layers are converted to percent-vegetation layers and then multiplied by the isotopic endmember values associated with the respective photosynthetic pathways. A standard deviation raster for each model is created by applying the literature-based standard deviation values (*Table 2*) to the final vegetation output. This process is repeated for each variation of the vegetation cover model (mean-temperature/max-herbaceous, mean-temperature/min-herbaceous, max-temperature/max-herbaceous and max-temperature/min-herbaceous).

Table 2 Values based on literature review used as endmembers to estimate leaf $\delta^{13}\text{C}$ values for each grid cell based on relative vegetation cover type.

Vegetation type	Leaf $\delta^{13}\text{C}$ (‰) ($\mu \pm \sigma$)	Source	No. samples	Location
C ₄ herbaceous (grasses)	-12.5 \pm 1.1	Cerling et al. 1997	†	global
C ₃ herbaceous (grasses)	-26.7 \pm 2.3	Cerling et al. 1997	†	global
Woody (Trees & Shrubs)	-27.2 \pm 2.5	Diefendorf et al. 2010	166	North America

† $\delta^{13}\text{C}$ values were estimated from 825 modern plants; the number of C₃ versus C₄ grasses is not specified.

3.2.6 Validating the Stable Terrestrial Carbon Isoscapes

Validating a product at a continental scale is inherently difficult, particularly for data with relatively coarse spatial resolution. While recognizing that a point-to-raster comparison introduces the change of support problem (COSP), I compare the four permutations of the final isoscape outputs to an independent data product of plot-level soil organic matter (SOM) $\delta^{13}\text{C}$ values (Von Fischer *et al.*, 2008). This data set sampled

two to four soil cores of 55 native prairie systems widely distributed across the Great Plains, reporting the mean $\delta^{13}\text{C}$ value for A and B soil horizons as well as plant roots. Although SOM and fine root material $\delta^{13}\text{C}$ values are both available in the Von Fischer data set, SOM is selected because it is a temporally-averaged $\delta^{13}\text{C}$ representation of multiple years of vegetation, as opposed to what is currently growing. I convert my leaf $\delta^{13}\text{C}$ predictions to SOM predictions by applying a constant offset of +2‰, identified by Bowling *et al.* (2008) and used to estimate $\delta^{13}\text{C}$ values of SOM in Powell *et al.* (2012). This offset accounts for fractionation that occurs between plant matter and SOM. In order to evaluate each combination of parameters on the $\delta^{13}\text{C}$ isoscape, I calculate the residuals between the predicted $\delta^{13}\text{C}$ value and the observed $\delta^{13}\text{C}$ value. Locations and values of the measured SOM $\delta^{13}\text{C}$ values reported by Von Fischer *et al.* (2008) are provided in a table in *Appendix 1*.

3.3 Results

3.3.1 Four Models of Geographic Distribution of leaf $\delta^{13}\text{C}$

Four vegetation $\delta^{13}\text{C}$ isoscapes were generated to test the impact of two sets of modeling parameters. These include a mean-temperature /max-herbaceous (Fig. 7-A), max-temperature /max-herbaceous (Fig. 7-B), mean-temperature/min-herbaceous (Fig. 7-C), and max-temperature /min-herbaceous (Fig. 7-D).

The spatial patterns of mean leaf $\delta^{13}\text{C}$ are directly determined by the relative distributions of plant functional types, and many patterns were consistent in all four permutations, as only the C₃/C₄ composition of natural grasslands was varied between

each version. The most isotopically depleted pixels occurred in the boreal forests in Canada, where there is the highest percentage of tree cover relative to grasslands or other vegetation types. The most isotopically enriched pixels occurred along the border between southeastern Texas and Mexico, an area which was categorized predominantly as shrub and grassland under the IGBP classification system, and modeled as having ≥ 6 months of climate conditions favoring C_4 grasses. The Great Plains region of North America was modeled as annually mixed C_3/C_4 grasslands, resulting in a wide range of $\delta^{13}C$ values depending on the relative composition of vegetation predicted in each pixel. The percent crop cover is a fixed value in all four isoscape permutations. Therefore, in all permutations, a significant amount of C_4 crops were modeled in areas classified as C_3 dominant in Iowa, Wisconsin, Minnesota, and Illinois (Fig. 6), resulting in enriched mean leaf $\delta^{13}C$ values relative to what would have been modeled without taking crops into account. Similarly, the southeastern forests of the United States maintain consistent leaf $\delta^{13}C$ values across all permutation because the tree-crown layer was assumed to have high fidelity and also remained fixed in all permutations.

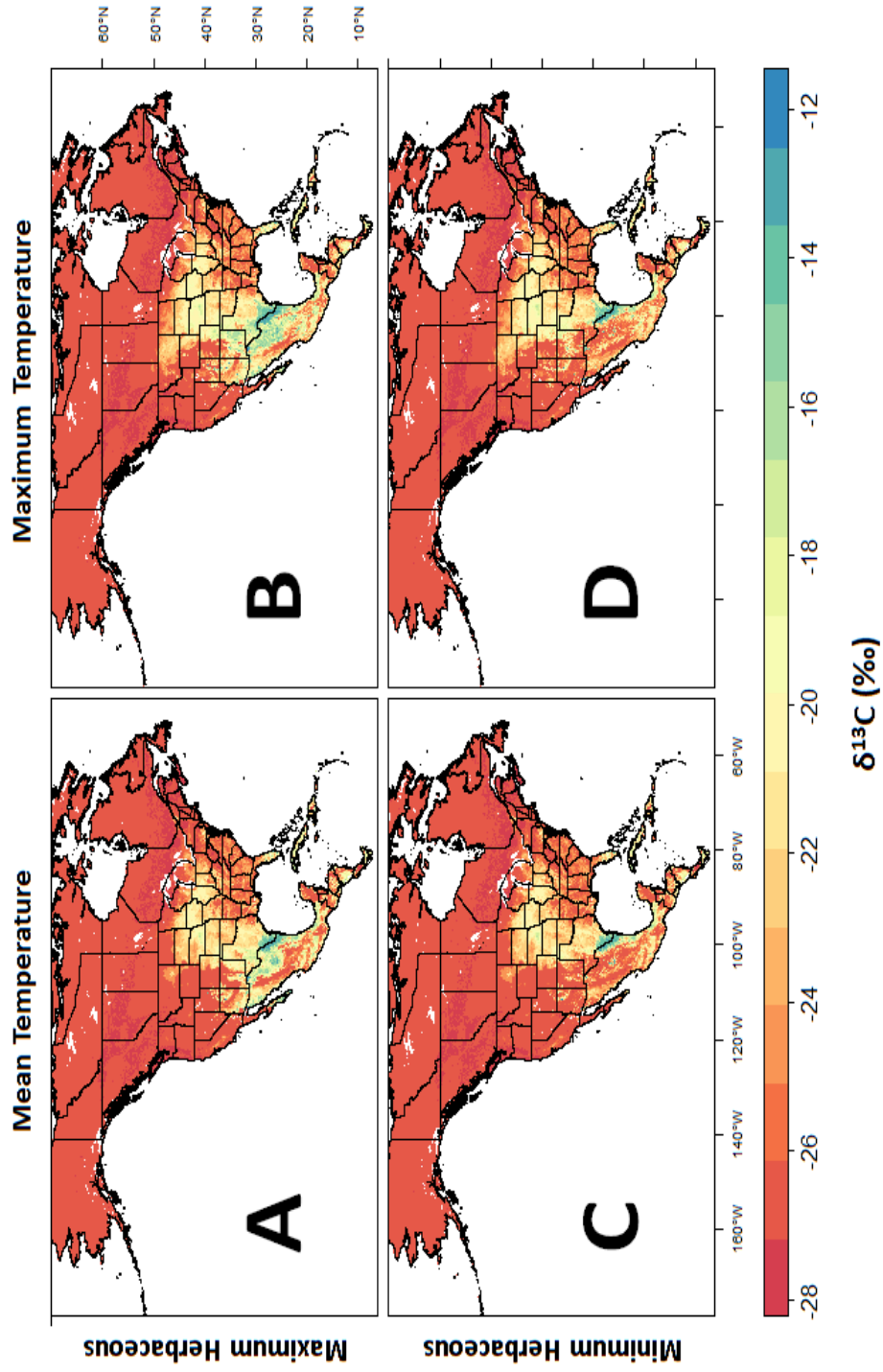


Figure 7 The four permutations of the annual 5-min resolution stable terrestrial carbon isoscape. The left column contains the monthly mean-temperature iterations (Collatz *et al.*, 1998) with maximum (A) and minimum (C) herbaceous land-cover rules, and the right column contains the monthly maximum-temperature iterations (Griffith *et al.* 2015) with maximum (B) and minimum (D) herbaceous land-cover rules.

There were, however, notable differences between each isoscape permutation. The areas of greatest difference due to land-cover rules directly corresponded to the IGBP shrubland classification (Fig. 4). The land-cover rules resulted in the biggest differences in the Chihuahuan desert region, as highlighted in the land-cover rules difference map (Fig. 8-A), created by subtracting the mean-temperature/min-herbaceous isoscape (Fig. 7-B) from the mean-temperature/max-herbaceous isoscape layer (Fig. 7-A). I also performed this calculation on the max-temperature permutations; however, the output was the same.

To directly compare the differences that result from changing the crossover temperature parameter, I created a difference map between the number of C₄-dominant months modeled by the max-temperature crossover parameter and the mean-temperature crossover parameter (Fig. 9). The climate parameters showed no difference in number of C₄ months north of the U.S.-Canadian border, south of the Mexico-Guatemala border and along the West Coast of the United States. Changing the crossover-climate parameter shifts the boundary of C₄ presence at higher latitudes and high elevations; specifically, the max-temperature parameter predicts greater C₄ presence at higher latitudes and high elevations than the mean-temperature parameters.

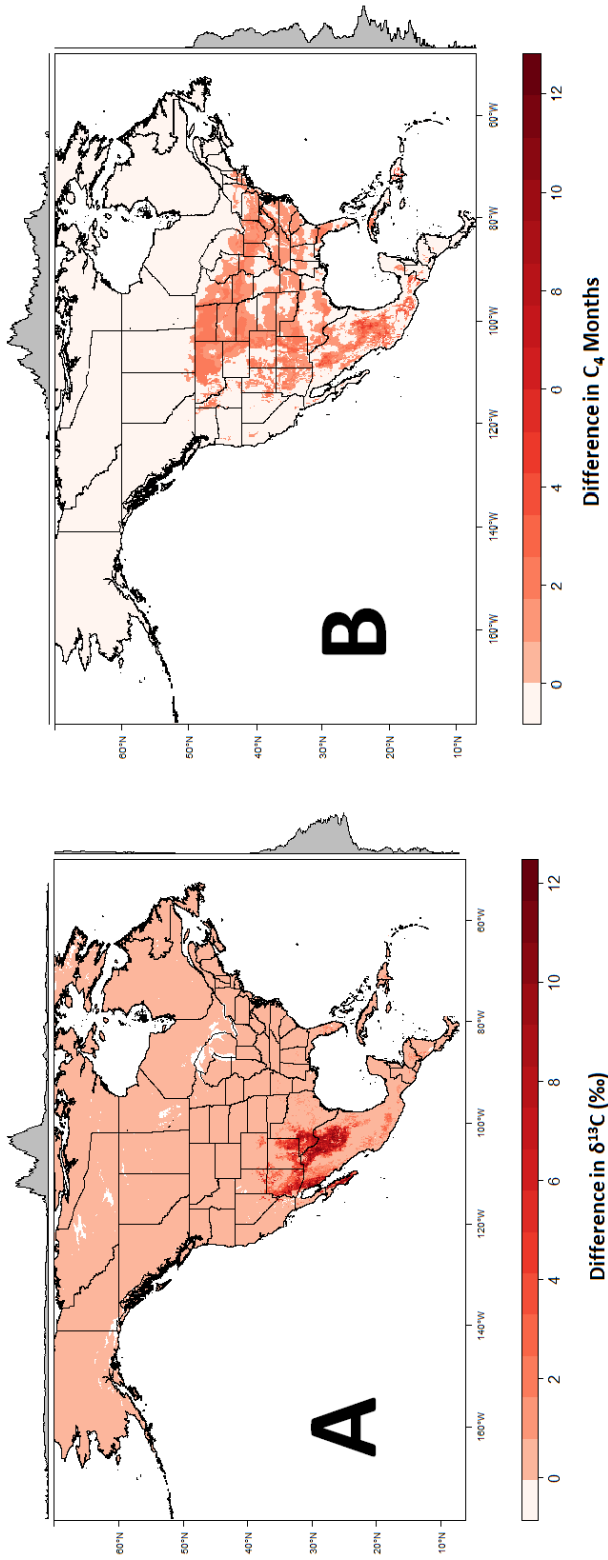


Figure 8-A The difference between maximum-herbaceous and minimum-herbaceous isoscapes (max-min). Higher values correspond to areas of higher uncertainty based on land-cover assignment rules. **Fig 8-B** depicts the difference in C_4 – dominant months between the monthly mean-temperature crossover parameters (Collatz *et al.*, 1998) and the monthly maximum-temperature crossover parameters (Griffith *et al.*, 2015). The maximum-monthly temperature parameters include a greater area of C_4 -dominant months, as well as a greater number of month. The plot on the top of each figure depicts the magnitude of difference as a function of longitude, and the plots to the right of each figure chart difference as a function of longitude.

3.3.2 Evaluation of Models against an Independent Data Set

Each permutation of the annual $\delta^{13}\text{C}$ isoscape was compared to an observed soil organic matter (SOM) data set of 55 sample sites spanning the Great Plains, USA (Von Fischer *et al.*, 2008), depicted in Fig. 9. The residuals for all four isoscape permutations had a positive bias. This means that given the +2‰ offset, the isoscape models typically predicted a more depleted $\delta^{13}\text{C}$ value than the observed SOM $\delta^{13}\text{C}$. Residuals for the isoscape permutations generated with the mean monthly temperature parameter were similar in magnitude. The mean of residuals for the mean-temperature/max-herbaceous permutation was 3.26‰, and the mean of residuals for the mean-temperature/min-

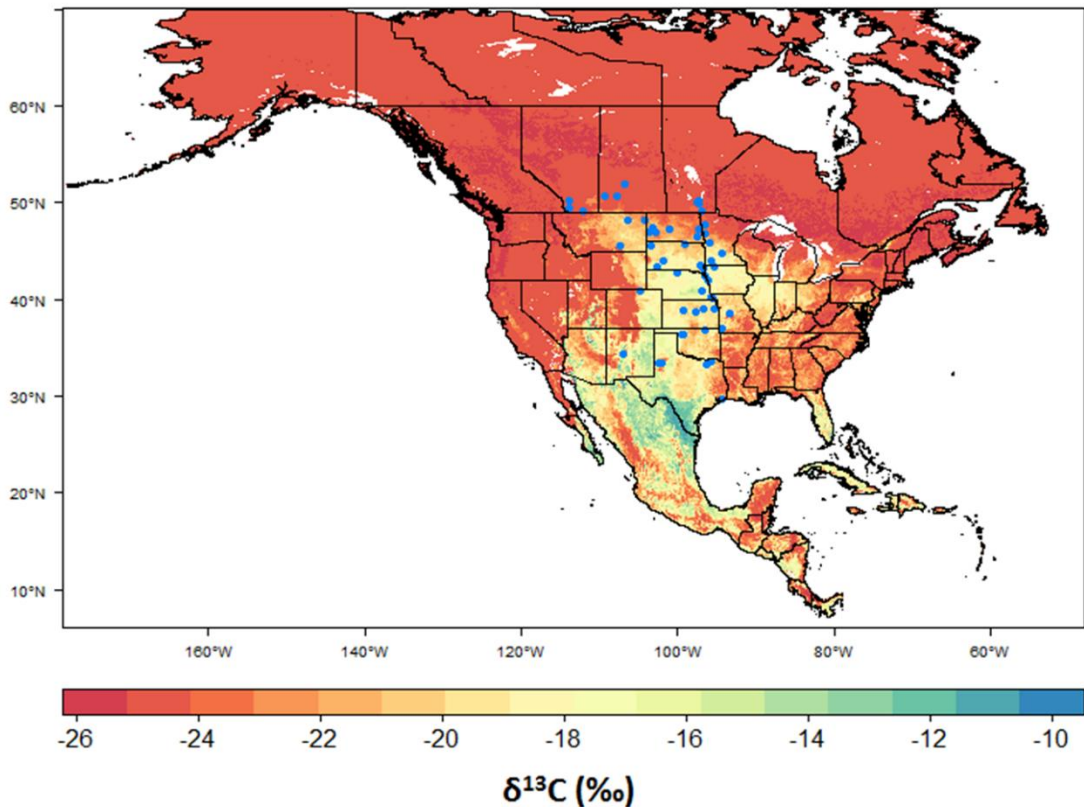


Figure 9 Distribution of SOM observation sites, depicted in blue, over the max-herbaceous /max-temperature isoscape permutation, converted from leaf to SOM predictions.

herbaceous permutation was 3.34‰. Residuals for isoscapes generated with the maximum-monthly temperature parameter were slightly lower than the mean-monthly temperature permutations. The mean of residuals for the max-temperature/max-herbaceous permutation was 2.17‰ and the max-temperature/min-herbaceous model was 2.28‰. The residuals for permutations generated with the minimum-herbaceous land-cover rules were slightly larger than the permutations generated with the maximum-herbaceous land-cover rules.

The land-cover rules resulted in different $\delta^{13}\text{C}$ predictions between permutations (different residuals) at only five of the SOM sample sites, all of which are located within the extent of the 2001 IGBP shrubland classification. At these five sites, the max-herbaceous permutations had lower residuals than their paired min-herbaceous permutations. With the exception of the Seivelleta, NM site, all four permutations predicted more isotopically enriched values than the observed SOM, meaning that C_4 abundance was over-predicted relative to soil values.

The ten sample sites located north of the Canadian border, as well as seven sample sites within the Great Plains, USA, had identical residuals across all four isoscape permutations. Both crossover-temperature parameters predicted 0 months of C_4 dominance north of the U.S.-Canadian border, meaning the region is predicted as being entirely C_3 dominant regardless of the crossover climate parameter used. The Great Plains sites included in this group are clustered in Northern Kansas, Missouri and Southern Nebraska, an area that is modeled as having no difference between the mean-temperature and max-temperature crossover parameter permutations (Fig. 8).

The max-temperature isoscape permutations had smaller residuals than the mean-temperature isoscape permutations at most sites. However, at three of the 55 observed SOM sites, the mean-temperature permutations had smaller residuals than the max-temperature permutations. These sites were Lange-Ferguson in South Dakota, Schaefer Prairie in Minnesota and Custer Battlefield in Montana. At all three of the sites, the max-temperature permutations predicted a more isotopically enriched value (overestimating C₄ presence) relative to the mean-temperature. These three sites are all located on the northern and western edges of the modeled extent of C₄ presence in the mean-temperature permutations.

The largest difference between the residuals of the mean- and max-temperature permutations occurred at a site in the northern Colorado Front Range, at an elevation > 1650 m, where the residual of the mean-temperature crossover parameters $\delta^{13}\text{C}$ predictions was approximately 6‰ more than the max-temperature permutations.

Overall, the crossover climate parameter had a greater impact on the ability to more accurately predict the observed SOM $\delta^{13}\text{C}$ values than the land-cover rules. All permutations typically predicted more depleted $\delta^{13}\text{C}$ values than what was observed; however, the mean-temperature permutations underestimated C₄ abundance more dramatically than the max-temperature permutations. Based on these validation data, the permutation of best fit was the max-temperature/max-herbaceous.

3.4 Discussion

Prior work has estimated the crossover latitude from C₄- to C₃- dominated communities between 43° and 45°N (Ehleringer *et al.*, 1997; Tieszen *et al.*, 1997). This is more similar to what is modeled by the mean-temperature crossover parameter than the max-temperature crossover parameter, as the latter predicts the transition from mixed to C₃ dominant at approximately 48°N. The three soil organic matter (SOM) sites at which the mean-temperature permutations had smaller residuals than the max-temperature permutations all occurred along the edge of the extent of C₄ presence predicted by the mean-temperature permutations, and the max-temperature permutations predicted more isotopically enriched values than what was observed. This could potentially indicate a need for different crossover parameters, where the max-temperature parameter appears to perform better at lower latitudes and higher elevations, the mean-temperature may perform better at the higher latitudes where C₄ dominance transitions into C₃.

One advantage of the max-temperature crossover parameter is that, by using a monthly maximum temperature as opposed to the mean, daytime temperatures are not smoothed out by overnight lows. Using monthly maximum temperature had the most impact at higher elevations and deserts, where the max-temperature permutations included more of the foothills of the Rocky Mountain range, higher latitudes, and more of the Chihuahuan and Sonoran deserts as either mixed or C₄ dominant relative to the mean-temperature permutations.

Von Fischer *et al.* (2008) has similar results and also found weak agreement between their SOM data and the general Collatz *et al.* (1998) crossover model not

accounting for crops. The sites sampled by Von Fischer *et al.* (2008) were all located on intact native prairies and grasslands, meaning that the impacts of cropland and different vegetation life forms (i.e., trees and shrubs) were not captured by the sample. However, this was an appropriate data set for testing the land-cover rules and climate crossover parameters, as these variations in model parameters most directly impact the modeling of natural herbaceous vegetation. Additionally, SOM was an appropriate validation data set for the annual isoscapes, as SOM captures a long-term average of vegetation cover, as opposed to the vegetation that is present at a particular point in time. The predictions of $\delta^{13}\text{C}$ generated from the isotope models are highly generalized, static averages of the leaf $\delta^{13}\text{C}$ for each 5-min pixel.

One limitation of the SOM data set is that the SOM values are effectively point observations being compared to a 5-minute (approximately 10x10 km) pixels. Von Fischer *et al.* (2008) also proposed that the disagreement they found between their observed SOM data and the Collatz-crossover predictions was potentially due to spatial context. Finer-scale prediction is not necessarily the intended use of the Collatz-crossover model, so when it was applied to the spatially complex Great Plains region, it presented weaknesses in discriminating mixed grasslands. Because the range of global climate conditions vary much more than climate conditions within the Great Plains, the parameters used for the global crossover model did not have the same explanatory power when applied to a more constrained, regional extent (Von Fischer *et al.*, 2008). At finer spatial resolutions, subtler differences between environmental conditions become more important for plant functional type discrimination.

Some of the disagreements between the $\delta^{13}\text{C}$ isoscape predictions and SOM data may also be related to the change of support problem (COSP). Comparing the plots sampled by Von Fischer *et al.* (2008) to 5-min spatial resolution raster pixels is essentially a point-to-raster conversion, where the raster pixels represent an average of values over a much larger area than what is captured in the 1m^2 quadrant plots. When values extracted from the raster cell are compared to the soil plot, the extracted raster value is implied to be equivalent to the point; however, the pixel integrates values over a much larger spatial grain than that.

In comparing the four permutations to observed SOM, only five of the 55 observed SOM points were impacted by the land-cover parameter, and both of the max-herbaceous permutations performed better than the min-herbaceous. The disagreement between land-cover rules can be interpreted as uncertainty in land cover, indicating the greatest uncertainty due to the land-cover classification rules overlaps the IGBP shrubland classification. This region may be most impacted by modeled herbaceous land-cover parameters for a few different reasons. First, the IGBP shrubland classification is a mixed class comprised of both herbaceous and shrub; therefore, pixels in this class are directly impacted by the varying land cover rules. The land-cover rules are a somewhat arbitrary step function that split mixed classes into shrub and herbaceous layers based on the woody vegetation threshold ranges included in the IGBP land-cover class descriptions. The larger the value of the step function, the greater the difference is between the max- and min- herbaceous permutations. Second, there is a large percentage of barren cover in deserts. This model calculates isotopic values based on relative

percentage vegetation cover. This means that shifts in the shrub/herbaceous composition in areas with a high barren percentage will result in a more dramatic difference in plant functional type percentage because the total vegetation cover is very low.

The positive bias observed when comparing modeled values to a reference SOM data set may indicate the need for a different offset parameter to convert leaf vegetation $\delta^{13}\text{C}$ to soil $\delta^{13}\text{C}$. In an ideal model, the mean of the residuals should be close to 0. The offset I used was derived from multiple data sets of soil organic matter (SOM) in forest environments (Bowling *et al.*, 2008). The +2‰ offset may therefore not be appropriate for grasslands or shrubland environments. The leaf-to-soil offset could also be latitude dependent, requiring different offsets for varying latitude thresholds.

3.5 Conclusion

Based on the comparison of all four variations of the leaf $\delta^{13}\text{C}$ isoscape to an observed soil organic matter (SOM) data set, the max-temperature/max-herbaceous model is selected as the best-fit model iteration to use for the research questions that follow. The greatest error of the max-temperature crossover parameter is overestimation of C_4 presence at “edge” dominance sites.

To better understand strengths and limitations of the methods used to generate stable terrestrial carbon isoscapes at mid-latitudes, the four models should be validated against additional data that span larger spatial extent than the U.S. Great Plains, and that include sites other than preserved prairies. The poor fit between the SOM observations was likely in part due to the change of support problem (COSP). To mitigate the impact

of COSP on my subsequent research questions concerning spatial and temporal resolution, I elect to use a different method for model assessment, harnessing geographic origin assignment as a mechanism for comparison and validation.

CHAPTER 4: The Impact of Temporal Resolution on Isoscapes in Mixed Grasslands

4.1 Introduction

How can seasonal variation of vegetation cover be represented in a static model? How does temporally aggregating the seasonal variation of mid-latitude grasslands impact confidence in the final model?

Not only do environments change with time, many animals occupy different spaces as a function of time, making both space and time fundamental aspects of ecology (Hobson & Norris, 2008). Stable isotopes have been identified as a natural recorder that can be harnessed to trace ecological processes and activities such as migration. Because environmental isotopes are directly integrated into animal tissues through diet, the environment where an animal's tissue was grown can be inferred by relating tissue isotopic values to an environmental isoscape (West *et al.*, 2006; Bowen, 2010). In the case of migratory animals, the temporal component of tissue growth is also spatial (Wassenaar 2008).

Commonly, geographic origin assignment exercises assume that the environmental isotopic values (environmental isoscapes) are assumed to be static over time (Hobson *et al.*, 2003; Wunder *et al.*, 2008; Hobson *et al.*, 2012). Recently, some

research has begun to explore the impact of inter-annual temporal variability on stable hydrogen isoscapes and geographic origin assignment exercises (Vander Zanden et al., 2014; Vander Zanden et al., 2015; Tonra et al., 2015). I intend to extend this research by exploring the impact of *intra*-annual variation on $\delta^{13}\text{C}$ isoscapes and geographic origin assignments.

Unlike the continents of South America and Africa, North America spans high enough latitudes that seasonality plays a major role in grassland composition (Ode *et al.*, 1980, Monson *et al.*, 1983; Ehleringer & Monson, 1993; Sage, 1999). Modeling the composition of mid-latitude grasslands requires a number of important decisions on how to represent the geographically co-dominant, but temporally offset, communities of plant functional types. Temporally offset co-dominance means that the C_3/C_4 composition of grasslands varies within a season, resulting in pixels that are both spatially mixed (when considered as individual snapshots within the season) as well as temporally mixed (different relative compositions over the course of the season at one geographic location). Several recent studies have analyzed remote sensing data to discriminate plant functional type communities using time-series data and vegetation phenology (Guan *et al.*, 2012; Wang *et al.*, 2011; Wang *et al.*, 2013; Zhong *et al.*, 2015). However, instead of examining how to best harness phenology for plant functional type discrimination, I explore how intra-annual variability of vegetation composition impacts a static representation of the stable carbon isoscape. This research explores the importance of matching the temporal resolution of tissue growth to the stable isoscape used for geographic origin assignment.

To accomplish this aim, I compare three representations of the $\delta^{13}\text{C}$ isoscape model with different temporal resolutions. The annual $\delta^{13}\text{C}$ isoscape generated in Research Question 1 (Chapter 3) with temporally mixed pixels was compared to a set of seasonally aggregated $\delta^{13}\text{C}$ isoscapes (four data products, representing early-, mid-, late-growing season and winter) and monthly $\delta^{13}\text{C}$ isoscapes; the latter correspond to the finest temporal resolution of the input climate data. I use the max-temperature parameter and max-herbaceous land-cover rule for all temporal variations applied to the isoscape, based on the evaluation with a reference data set (Section 3.3.2).

4.2 Methods

4.2.1 Generating Seasonal and Monthly Resolution Isoscapes

To generate the seasonal and monthly $\delta^{13}\text{C}$ isoscape data products, the same process described in Chapter 3.2 is followed, the only difference being that the natural herbaceous layer is partitioned into C_3/C_4 components at different temporal resolutions. For monthly representations of the $\delta^{13}\text{C}$ isoscape, pixels are categorized as either C_3 - or C_4 -dominant for each month based on the maximum monthly temperature (“max-temperature”) crossover climate parameters (Griffith *et al.*, 2015). Although the pixels may be spatially mixed in terms of the type of vegetation cover (e.g., woody or crop), the natural herbaceous vegetation itself is categorized as entirely C_3 or entirely C_4 . For the seasonal isoscapes, the number of C_4 -dominant months per pixel are summed into four seasonal periods (early, mid, late, and winter). Seasons are defined as follows: early- (March, April and May), mid-(June, July and August), late-(September, October and

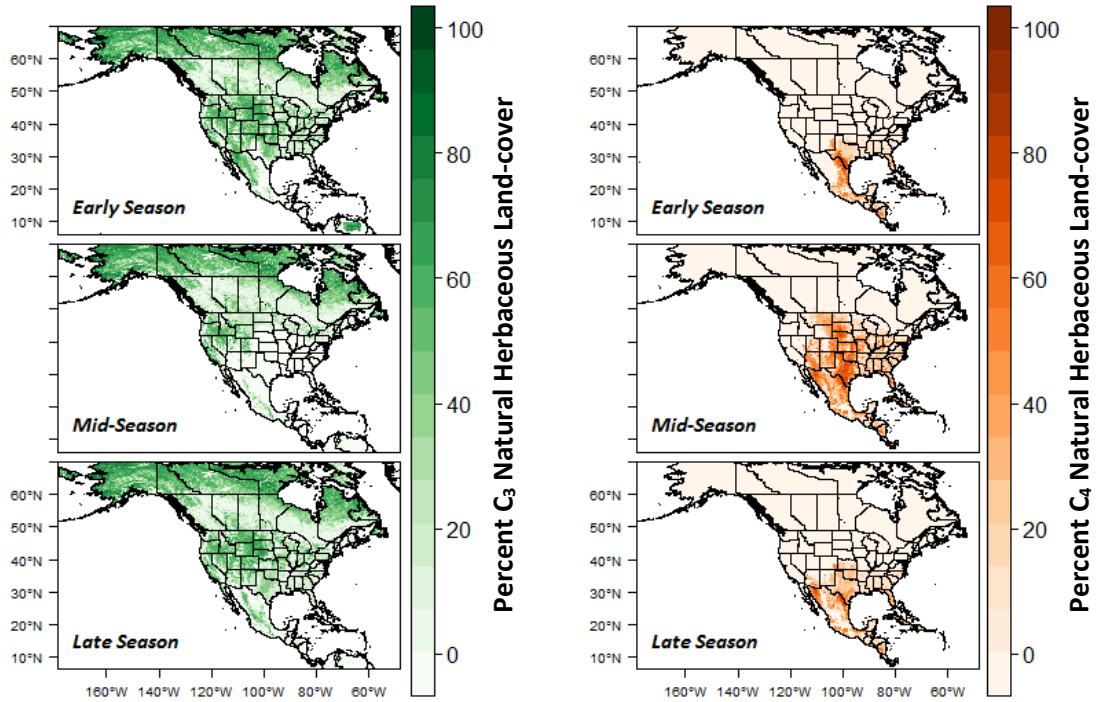


Figure 10 Seasonal C_3/C_4 land cover for the maximum-herbaceous land-cover rule and max-temperature parameter permutation. The figures in green depict the percent of C_3 natural herbaceous land cover. The orange figures depict the seasonal C_4 natural herbaceous land-cover.

November) growing seasons, and winter (December, January and February). Within each seasonal period, the proportion of C_4 to C_3 vegetation is estimated by generating an NDVI-weighted ratio based on the NDVI during C_4 -favorable months to NDVI across the season (Section 3.2.4) to generate temporally mixed representations for the season. The percent cover of C_3 and C_4 natural herbaceous vegetation by season is depicted in *Fig. 10*. The result is four seasonal isoscapes and twelve monthly isoscapes. Each temporal set corresponds to the same annual representation (Section 3), with differing temporal resolutions. All of the isoscapes are generated at a 5-min spatial resolution.

4.2.2 Geographic Origin Assignments

A potential application of the $\delta^{13}\text{C}$ isoscape for North America is predicting the unknown origin of individual migratory animals using the stable isotope values of their tissue. Here, I assess the impact of the temporal resolution of the $\delta^{13}\text{C}$ isoscape through a geographic origin assignment exercise (e.g., Bowen *et al.*, 2014; Vander Zanden *et al.*, 2014). Geographic origin assignment is the process of predicting the origin of an animal tissue by relating the isotopic content of the tissue to an environmental isoscape. Using known-origin tissues allows me to assess the impacts of the temporal resolution of an isoscape on the precision and accuracy of geographic assignment prediction.

I use a data set of feather isotope values from 118 known-origin mountain plover (*Charadrius montanus*) chicks, collected and analyzed for $\delta^{13}\text{C}$ in 2001 and 2002 (Wunder *et al.*, 2005). The sample sites are depicted along with the mountain plover breeding range in *Fig. 11*, adapted from Wunder *et al.* (2005). Mountain plover chicks forage independently for invertebrates in the first few hours after they hatch; contour and flight feathers in sheath begin to show at about 7 days old (Wunder *et al.*, 2005). The chicks cannot fly prior to their first feathering, and therefore the isotopic content of their first contour and flight feathers are derived from a spatially constrained environment (limited to ~56 ha) (Knopf & Rupert, 1996; Wunder *et al.*, 2005), and the feather isotopes of the plover chicks are assumed to be known-origin. The plover chicks would have grown their first contour feathers over a period between mid-May and June of each respective year; feathers were sampled between June 12th and August 18th (Wunder *et al.*, 2005).

To test each of the isoscape variations, each of the individual mountain plover chicks (n=118) are treated as if they were from an unknown origin, and assigned an origin based on the calibrated feather-isoscape. Following the methods described in the following section (*Section 4.2.3*), probability surfaces are generated for each individual (Wunder, 2010; Vander Zanden *et al.*, 2014; Vander Zanden *et al.*, 2015). Probability surfaces are a surface in which each pixel is assigned a probability of true origin. For any given assignment probability raster, the per-pixel probability of origin are normalized so that all pixel probabilities within an assignment raster sum to 1 across the study area extent. High probability values correspond to relatively high likelihood that the pixel is the true origin.

Geographic origin assignments are performed for each individual and for each isoscape variation. This method of model assessment allows for me to explore the impact of temporal resolution in the context of geographic origin assignment, a common application isoscapes.

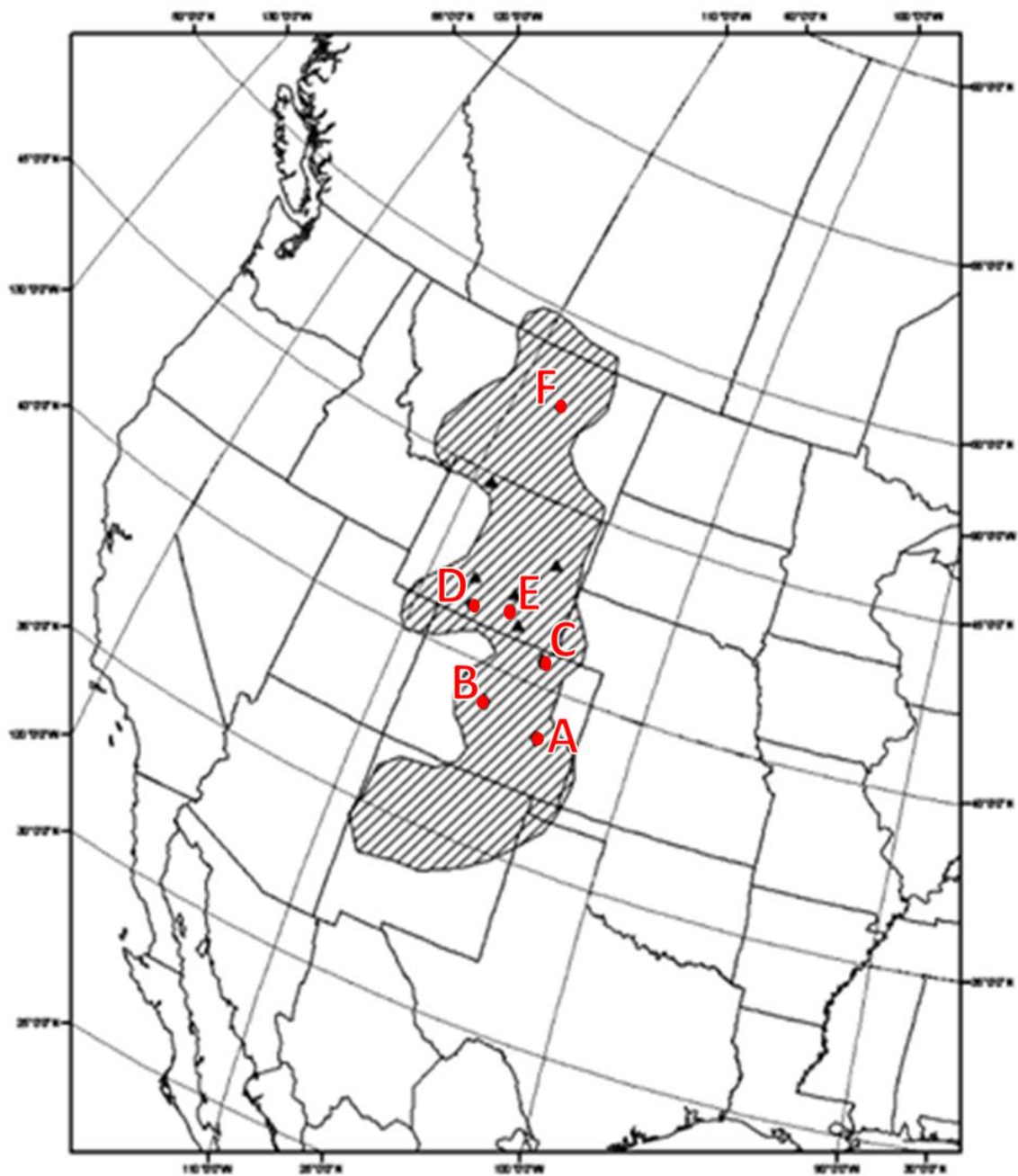


Figure 11 Map depicting the plover chick sample sites within the mountain plover breeding range, adapted from Wunder *et al.* (2005). The sampling sites of data included in the current analysis are labeled with red letters.

4.2.3 Geographic Origin Assignment Methods

The three basic steps of geographic origin assignment are as follows: (1) create a rescaling equation to calibrate the environmental isoscape to tissue isotope values, (2) convert the environmental isoscape to an animal tissue isoscape, and (3) relate measured isotope values from individual animals to the tissue-calibrated isoscape to generate a probability of origin surface (Wunder & Norris, 2008; Wunder, 2010).

In this study, the rescaling equation step is performed by bootstrapping all of the original known-origin data to create a series of rescaling functions for each seasonal and monthly isoscape. Bootstrapping is a statistical technique by which population data are simulated by randomly sampling the original sample with replacement repeatedly (Efron & Tibshirani, 1993). One thousand bootstrapped re-samples of the data are generated, and each is used to generate a rescaling equation, resulting in 1000 sets of slopes and intercepts. To convert the environmental isoscape to a tissue isoscape, each bootstrapped rescaling equation is applied to the $\delta^{13}\text{C}$ isoscape, resulting in 1000 $\delta^{13}\text{C}$ tissue-calibrated isoscapes. Each of the calibrated rasters is then collected into a raster stack. This geographic origin assignment process estimates variance and error from three sources (Wunder & Norris, 2008; Wunder, 2010).

To generate per-pixel probability of origin estimates, a final feather-calibrated isoscape must be generated and the associated per-pixel variance must be calculated. The final rescaled feather isoscape is calculated as the mean of the 1000 stacked rasters. The variance estimate consists of three components: (1) average variance between individuals at a site, (2) rescaling variance, and (3) isoscape variance. The measure of the average

variance between individuals at a single sample site is calculated by squaring and fitting the vector of standard deviations of regression residuals from each rescaling equation (1000 standard deviation values) to a gamma distribution, then finding the average. Variance related to the rescaling process is calculated as the variance of each pixel from the stack of 1000 calibrated rasters. Finally, isoscape variance is estimated by squaring the per-pixel values of the standard deviation surface that correspond to each of the $\delta^{13}\text{C}$ isoscapes based on literature values (Section 3.2.5). Each of the per-pixel estimates of variance are summed, and the square root of this value is used as the standard deviation for the assignment process. The bootstrapping, as well as capturing variance by per-pixel calibrations and fitting the mean gamma distribution of the rescaling equation residuals are all novel methods for creating geographic origin assignment probability surfaces. Because calibration functions are both tissue and isoscape dependent, the entire calibration process is repeated for each isoscape raster surface, corresponding to each temporal subset isoscape (Wunder & Norris, 2008; Wunder, 2010; Vander Zanden *et al.*, 2014; Vander Zanden *et al.*, 2015).

Finally, geographic origin assignments are performed for individuals by relating the measured isotope value to each pixel in the calibrated isoscape with a probability density function that accounts for the three sources of variance and error detailed in the previous paragraph. The R code used to generate the geographic origin assignments adapted from Vander Zanden *et al.* (2014) is available in *Appendix 3*.

4.2.4 Assessing Model Efficacy through Accuracy, Precision and Similarity

To explore the impact of temporal resolution on the isoscape models and geographic origin assignment, the accuracy, precision and similarity of the generated geographic origin assignments of each individual raster were assessed following methods described in Vander Zanden *et al.* (2014). The winter-seasonal and monthly (i.e., January, February and December) assignments were not included in the efficacy evaluations, as the mountain plover breeding range has very little active vegetation during these periods and is not well represented by the isoscapes.

At the population-level, both accuracy and precision are assessed across a sequence of relative probability values, ranging from 0.01 (low probability) to 0.99 (high probability), at 0.01 increments. Relative probability for each assignment raster is calculated by dividing each pixel in the entire raster surface by the maximum probability for that particular individual assignment (i.e., maximum value across the entire spatial extent of the data set). Population-level assignment *accuracy* is determined by the proportion of the known-origin validation locations included in the assignment region at each probability interval. This measure evaluates the capacity of the assignment model to accurately identify a region of origin that includes the known-origin sites, given the temporal window the assignments were based on (Vander Zanden *et al.*, 2014). Population-level assignment *precision* is assessed as the median proportion of total surface area included in the assignment region (across all individuals) for each relative probability interval. Smaller proportions of the spatial domain indicate higher precision in assignment surface (Vander Zanden *et al.*, 2014).

Individual-level assessments of accuracy and precision are quantified as the difference between geographic origin assignments generated for the same individual (e.g., mountain plover 6645) with different environmental isoscapes (e.g., the May isoscape versus the June isoscape). The individual-level accuracy is assessed as the difference in relative probability between assignments at the known-origin location of individual birds. This metric determines if the relative likelihood of origin at the actual, known location of origin increased or decreased as the result of changing the temporal window used for assignment. Individual assignment precision is assessed as the difference in area of the posterior probability surface at the relative probability density value equal to or greater than the relative probability at the known origin. Differences in individual-level assessments were tested for statistical significance with two-tailed paired t-tests (Vander Zanden *et al.*, 2014).

The similarity index is used to determine if differing isoscapes result in significantly different geographic origin assignments in a per-pixel comparison. Similarity is only evaluated at an individual-level and is assessed by directly comparing two origin assignments for an individual bird generated using differing isoscapes. The index was adopted from the Expected fraction of Shared Presences (ESP) metric (Godsoe, 2014; Vander Zanden *et al.*, 2014). The index calculates the ratio of the number of shared cell values between two assignment rasters to the total number of possible cells. More simply, the similarity index represents the percentage overlap between the two assignments, over the entire range of relative probabilities (Vander Zanden *et al.*, 2014).

4.3 Results

4.3.1 Seasonal Isoscapes

Four seasonal isoscapes together represent the temporal extent of a full year (Fig. 12). The cropland signature is the same in all four models because there is no temporally coded data for the croplands. This results in isotopic enrichment in the Great Lakes, USA, region due to the dominance of C₄ crops, especially corn. North of the U.S.-Canada border there is no seasonal variation. Grasslands in the region are modeled as C₃-dominant year round. The West Coast of the United States is also C₃-dominant year round, due to a Mediterranean climate.

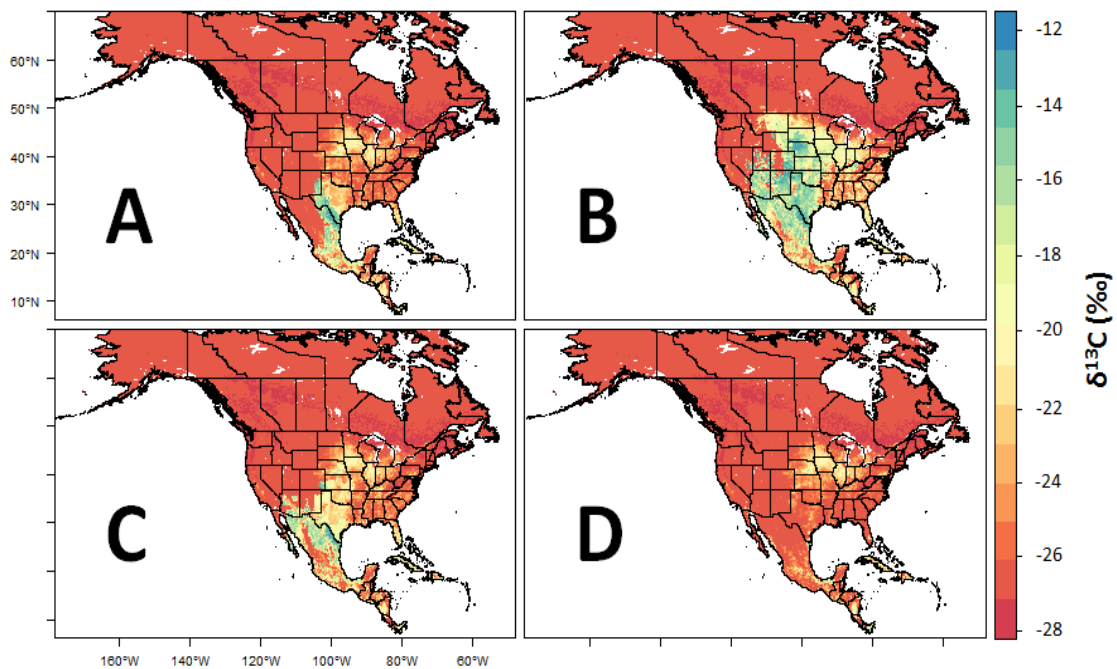


Figure 12 Seasonal $\delta^{13}\text{C}$ isoscapes. The “early season” represents March, April, and May (A). The “mid-season” includes June, July, and August (B). The “late season” includes September, October, and November (C). Winter includes December, January, and February (D).

Significant differences occur between each of the seasonal isoscapes. The largest extent of C₄ presence occurs in the mid-season isoscape (*Fig. 12-B*), and smallest extent of C₄ presence is modeled in the winter isoscape (*Fig. 12-D*). The late-season isoscape has a larger proportion of pixels classified as having C₄ presence relative to the early-season isoscape. The most dramatic seasonal changes occur in the Great Plains, the Sonoran Desert and Chihuahuan Desert regions. Additionally, there is an isotopically enriched signature in the Appalachian region of the United States in the mid-season isoscape, relative to the early- and late-season isoscapes. In the annual isoscape, seasonal variations are effectively smoothed out by averaging.

4.3.2 Monthly Isoscapes

Monthly isoscapes were the finest temporal resolution products generated, given the temporal resolution of the CRU climate data (*Fig. 13*). Similar to the seasonal isoscapes, the same temporally static cropland data set is apparent in each monthly model. The months between November and March resemble one another relatively closely because very few regions are classified as C₄-dominant during this time period (except for the pixels where the C₄ crop signal is present). At the monthly temporal resolutions, natural herbaceous grasslands are classified as either C₃ or entirely C₄. The monthly isoscapes also show differences from one another relative to seasonality. The May and June isoscapes show increasing C₄ presence in the Southern Great Plains and in Mexico. June, July and August months all predict C₄-dominant grasslands for large areas

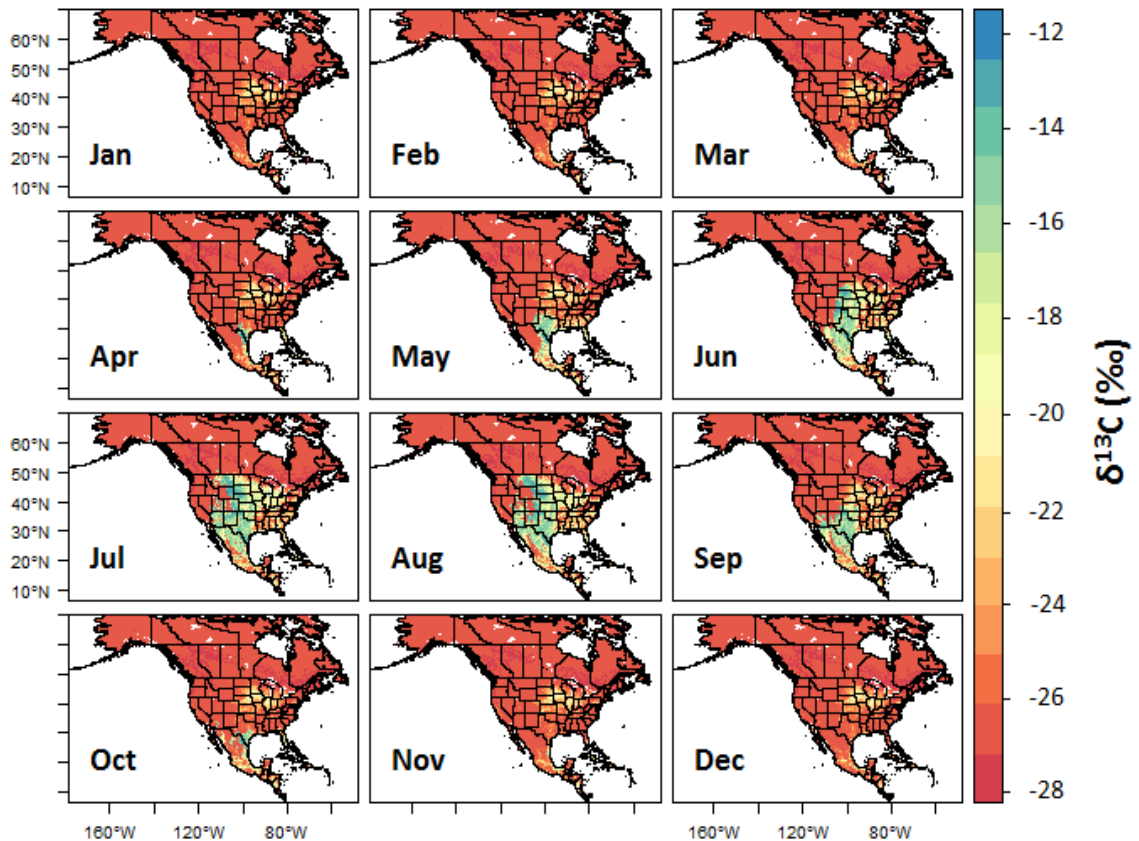


Figure 13 Monthly $\delta^{13}\text{C}$ isoscapes. The natural herbaceous component of each pixel is classified as either C_3 - or C_4 -dominant based on the climate mask for that month.

of the Great Plains, USA. C_4 dominance is modeled as first occurring around the Gulf of Mexico in April, expanding to the north and west in May and June, and then occupying the majority of the Great Plains region in July and August before retreating south in September and essentially dissipating in October. The May and September isoscapes are broadly similar, although the September isoscape predicts a greater extent of C_4 presence across East Texas and along the U.S.-Mexico border.

4.3.3 Geographic Origin Assignment Evaluation

Geographic origin probability surfaces were created for the 118 known-origin mountain plover chicks using each of the seasonal and monthly isoscapes, as well as the annual isoscape. For each isoscape, 1000 rescaled (i.e., feather-calibrated) isoscapes were generated from bootstrap sampled linear regressions that relate feather and environmental $\delta^{13}\text{C}$ values. Geographic origin assignments for the individual samples were made using the mean of the 1000 rescaled tissue isoscapes, while accounting for variance due to random sampling and other sources of uncertainty. Finally, the accuracy, precision, and similarity of the temporal-isoscape variations were assessed and compared.

Population Level Accuracy. Population-level accuracy was assessed as the proportion of known-origin sample locations included in the area defined by each relative probability threshold. This assessment resulted in relatively uniform accuracies for each isoscape variation (*Fig. 14*). The proportion of validation individuals included in each relative probability threshold decreases as the relative probability threshold increases, as higher probabilities are more selective and therefore include less area in general. Overall, the accuracies associated with assignments to the annual and each of the seasonal isoscapes were relatively high. Mid-season isoscape assignments had relatively higher population-level accuracy than annual isoscape assignments. The proportion of individuals included at the 0.99 relative probability interval was 0.75 for the mid-season assignments and 0.65 for the annual assignments. Early- and late-season isoscape assignments had similar population-level accuracy to one another, both of which were lower than the population accuracy of annual isoscape assignments. The proportion of

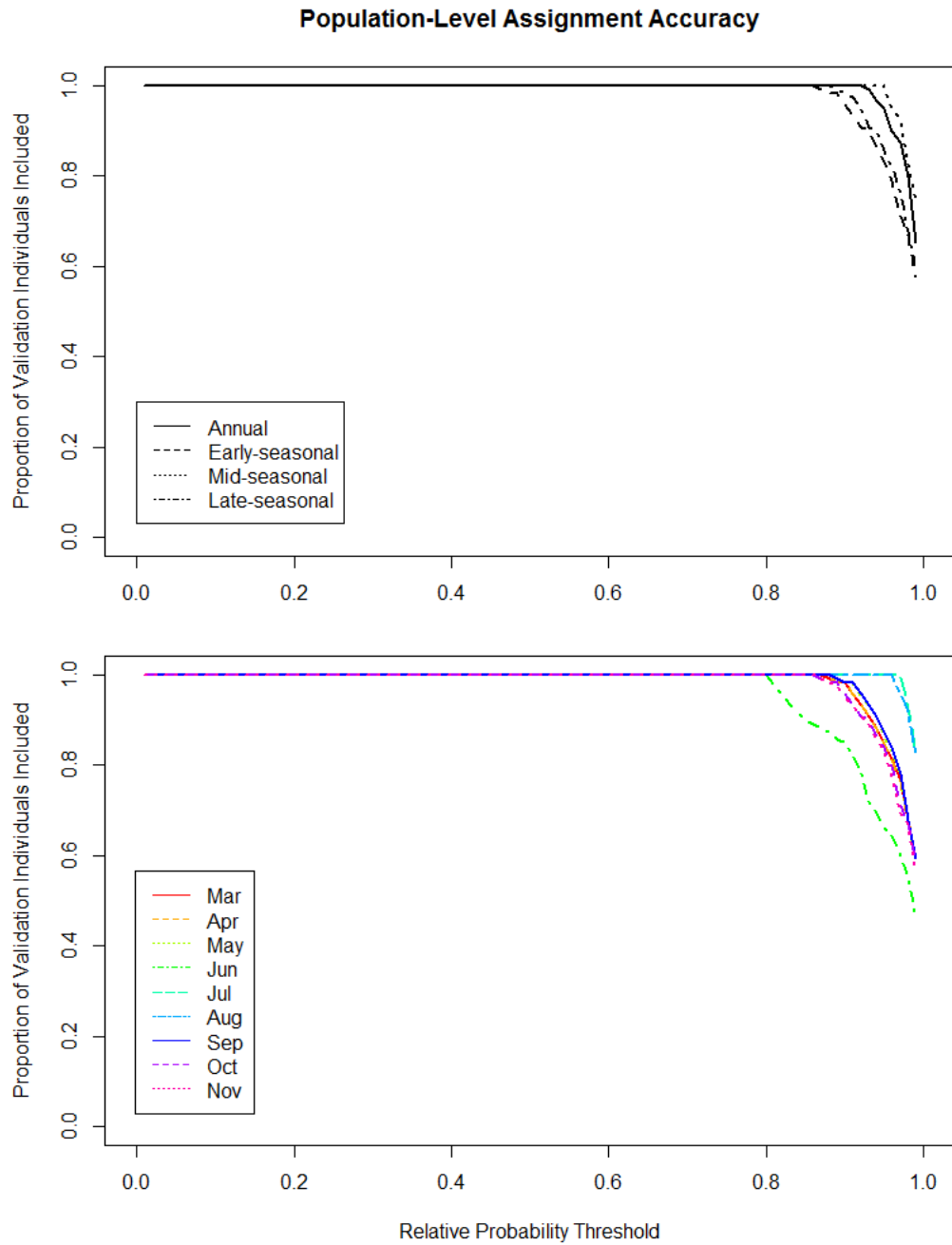


Figure 14 Population-level assignment accuracy, measured as the proportion of validation individuals for which the known origin was included at each relative probability interval.

individuals included at the 0.99 relative probability interval was 0.58 for early-season assignments and 0.59 for late-season assignments.

The monthly-isoscape assignments performed similarly to each of their respective seasonal assignment, with the exception of the June isoscape-based assignments, which had the lowest population level-accuracy, with only 0.47 validation individuals included at 0.99 relative probability interval. July and August isoscape assignments both included >0.80 of the known-origin locations at the 0.99 relative probability interval, which was higher than any of the seasonal or annual assignments.

Changes in Individual-Level Accuracy. Individual-level accuracy was measured as the difference in relative probability between different isoscape assignments (e.g., early-season – mid-season) at the known origin of each individual (Vander Zanden *et al.*, 2014). This metric indicates whether relative probability of origin at the true location of origin increased or decreased when the assignment was based on a different temporal window. The early-, mid- and late-season isoscape assignment probabilities were all significantly different from one another (p -value < 0.001 for all paired differences). Mid-season isoscape assignments had the highest accuracy compared to both the early- and late-season isoscape assignments. Early-season isoscape assignments had significantly lower accuracy than late-season assignments.

For monthly isoscape assignments, I compared the biologically relevant months (i.e., the months that correspond to feather growth), June and July, to one another. Between the two sets of assignments, July isoscape based assignments had significantly higher accuracy than June assignments (p < 0.001). I then compared the July isoscape

assignments to the mid-season and annual assignments. The July isoscape assignments had significantly higher accuracy than both the mid-season and annual assignments ($p < 0.001$).

Population-Level Precision. Population-level precision was calculated as the as the median proportion of total surface area included in an assignment at each relative probability interval (Vander Zanden *et al.*, 2014; *Fig. 15*). The early- and late-season isoscape assignments included a smaller proportion of total area at each given probability interval, indicating higher precision, than the mid-season and annual isoscape assignments. For the annual assignments, the median portion was equal to 1.0 (i.e., included the entire surface) until a relative probability threshold 0.94. The population level mid-seasonal assignments showed even lower precision. For mid-seasonal assignments, the median proportion of the surface included is equal to 1.0 until the 0.99 relative probability threshold, where the median proportion of included surface area drops to 0.97. The high population accuracies correspond to very low precision assignments.

A similar pattern was also found for the associated monthly assignments. Monthly assignment precision was highest for the late-spring (i.e., April and May) and early-autumn (i.e., September and October) isoscape assignments. The July and August monthly isoscape assignments both had very low precision; the entire surface was included at the 0.99 relative probability interval, as indicated by the horizontal lines in the precision plots (*Fig. 15-B*).

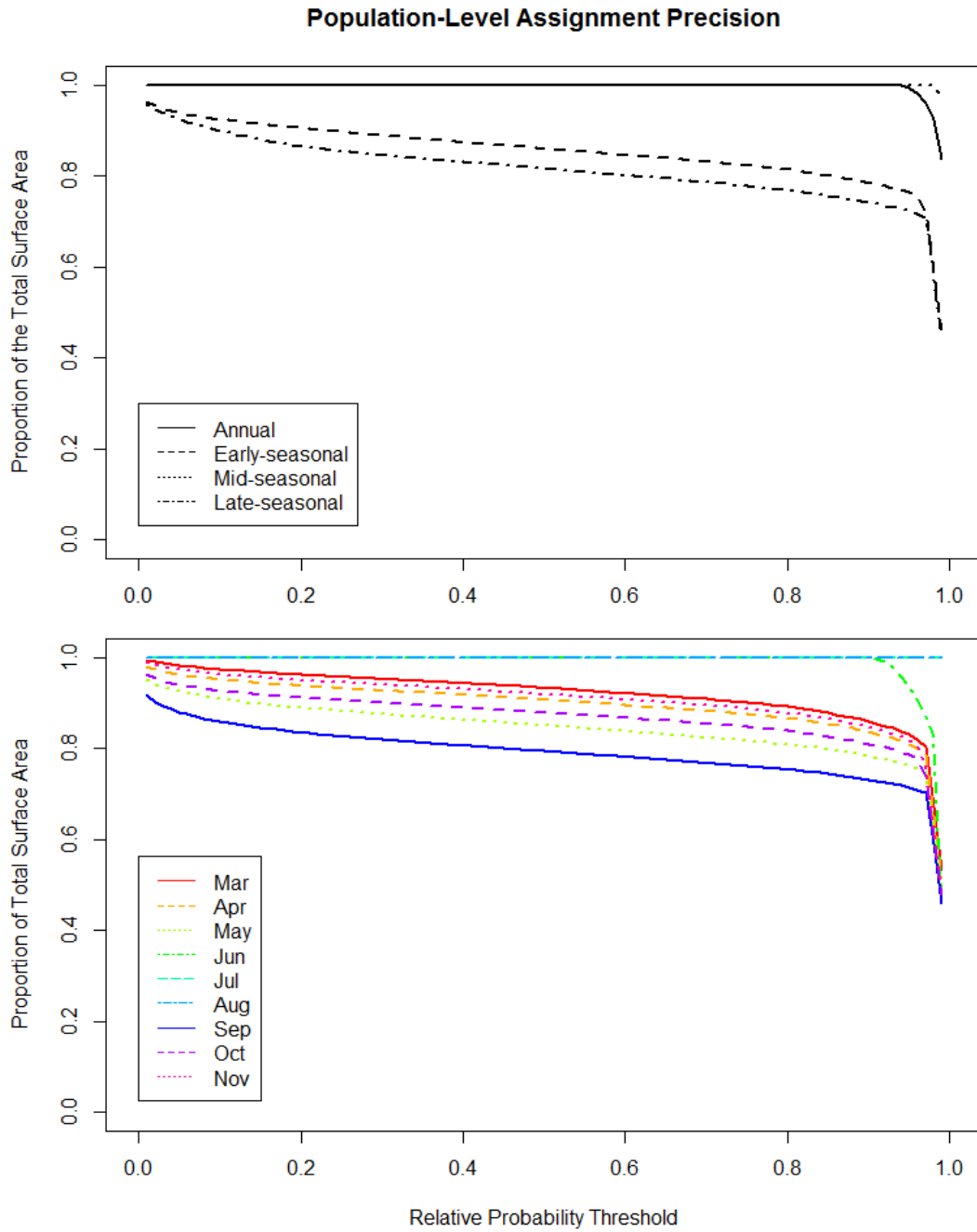


Figure 15 Population-level precision for the annual and three growing season isoscapes, measured as the median proportion of total surface area included in an assignment at each given relative probability interval.

Individual-Level Precision. Change in individual-level precision was measured as the change in area (between two isoscapes) of the posterior probability surface area defined by all pixels with values within the probability interval corresponding to that of the pixel of known origin. The differences in individual-level precision between the early- and mid-season isoscape assignments were significant, with early-season assignments having improved precision ($p = 0.005$). The differences between the early- and late-season isoscape assignments and the mid- and late-season isoscape assignments were all significant as well ($p < 0.001$, $p < 0.001$). Of the three seasonal isoscape assignments, assignments based of the late-season isoscape had the highest precision, and the mid-season isoscape assignments had the lowest precision based on the means of pair-wise differences.

Assignments based on the June isoscape had higher precision than the assignments based on the July isoscapes; however, the differences were not statistically significant as indicated by a paired two-tailed t-test ($p = 0.699$). The difference in precision between the annual and mid-season isoscape assignments was also not statistically significant ($p = 0.244$). In the pairwise comparisons of June assignments to mid-seasonal and June to annual assignments, there were no statistically significant differences.

Similarity. The similarity index represents the percentage overlap between the two assignments, over the entire range of relative probabilities (Vander Zanden *et al.*, 2014). The mean similarity indices between all comparisons combinations of the annual, seasonal and monthly assignments were relatively high (mean < 0.899). The annual

isoscape assignments were slightly more similar to the mid-season isoscape assignments (mean similarity = 0.9772) than to the early-season isoscape assignments (mean similarity = 0.9131). The early- and late-season isoscapes were more similar to one another than to the mid-season isoscape assignments, but the difference between similarity index means was not greater than 0.05 in any combination. The mean similarity between monthly-isoscape geographic origin assignments were consistent, with May and July monthly isoscape assignments having the lowest average similarity index (mean = 0.9118), indicating they are the least similar assignments, and April and May having the highest (mean = 0.9532), indicating they are the most similar.

4.4 Discussion

4.4.1 Seasonal and Monthly Isoscapes

Vegetation of the North American landscape is difficult to characterize in a single annual model due to the temporally offset, but geographically co-dominant communities that occur due to seasonality. For example, the Chihuahuan desert has two distinctive communities each year. In the winter, the landscape is C₃-dominant; however, in the summer the area is entirely C₄-dominant. Both seasonal communities are punctuated by intense drought, so neither is able to persist beyond a very limited time period (Kemp, 1983; Sage *et al.*, 1999). This shift in C₃/C₄ dominance is effectively captured by the seasonal isoscapes, whereas the annual isoscape averages out the two vegetation communities and predicts a landscape that never actually exists. Overall, the plant functional type dynamics predicted by the seasonal and monthly isoscapes appear to better match the descriptions of previous floristic survey studies (Ode *et al.*, 1980;

Monson *et al.*, 1983), where the spring months are C₃-dominant, the summer months are dominated by C₄ grasses, and there is a resurgence of C₃ grasses in the fall.

In the seasonal and monthly isoscapes, only “natural” herbaceous vegetation was modeled as being temporally variable. This is a major limitation in the current modeling methods, as the majority of plants do not actively persist at mid- to high-latitudes throughout an entire year. Another limitation of only modeling temporal shifts in natural herbaceous C₃ and C₄ dominance is made most apparent by the persistent C₄ crop signal in the Great Lakes region, USA, in every isoscape permutation. Being able to model the temporal variability of crops would better account for multi-cropping (seasonal crops), as well as fallow fields, and being able to “mask out” areas where no active vegetation is present would better constrain the extent of possible migratory animal origins.

I did not include the winter and winter-month isoscape assignments in the assessments for accuracy, precision and similarity because there were many limitations in the modeling methods that caused poor representation of the winter months. The model does not account for active versus dormant vegetation. During the winter months, the entire mountain plover breeding range is subject to snow cover (mean monthly temperature < 0°C), and therefore does not likely have much active vegetation. However, the winter isoscape, with modifications, has the potential provide information about migratory animals that spend the winter at lower latitudes and summer at higher latitudes on the North American continent.

4.3.2 Calibration Points as an Indicator of Temporal Fit

Mountain plover chicks grow their first set of feathers from mid-May to July (Wunder et al. 2008). These months are therefore biologically relevant in terms of time frame, and other periods over the year, are biologically irrelevant time periods. The mid-season, as well as the summer month (e.g. July and August) isoscape assignments all had very high accuracy at the population level. July and August isoscape-based assignments had the highest levels of accuracy. However, the high accuracy was coupled with having relatively low precision in comparison to the other isoscape assignments based on biologically irrelevant temporal windows. This dynamic might be explained by the calibration equations used to convert the environmental isoscape predictions into a feather isoscape.

The mountain plover chicks come from both C₃ and C₄ dominant environments, as indicated by the $\delta^{13}\text{C}$ values ranging from -21.9‰ to -12.4‰. This is well matched by the biologically relevant isoscape predictions of $\delta^{13}\text{C}$ values at the known-origin sites (-26.7‰ to -12.9‰). However, the isoscapes generated for irrelevant months predict only C₃ dominant values, constrained to a range of -26.7‰ to -25.5‰. This difference in range predicted by the environmental isoscape in comparison to the feather values indicates that the irrelevant time periods are not an appropriate fits for generating the mountain plover geographic origin assignments, and will not generate meaningful calibration equations. The gain in assignment precision seen in assignments based on irrelevant time periods is artificially gained from extremely steep calibration equations generated by trying to fit a wide range of feather $\delta^{13}\text{C}$ values to a narrow range of $\delta^{13}\text{C}$ isoscape predictions (*Fig.*

16). All biologically irrelevant temporal windows have very similar $\delta^{13}\text{C}$ isotope predictions within the mountain plover breeding range, and the resulting feather isotope calibration equations all look similar to the calibration points displayed in *Fig. 16*. Calibration point and rescaling equation plots for all isoscapes are included in *Appendix 2*.

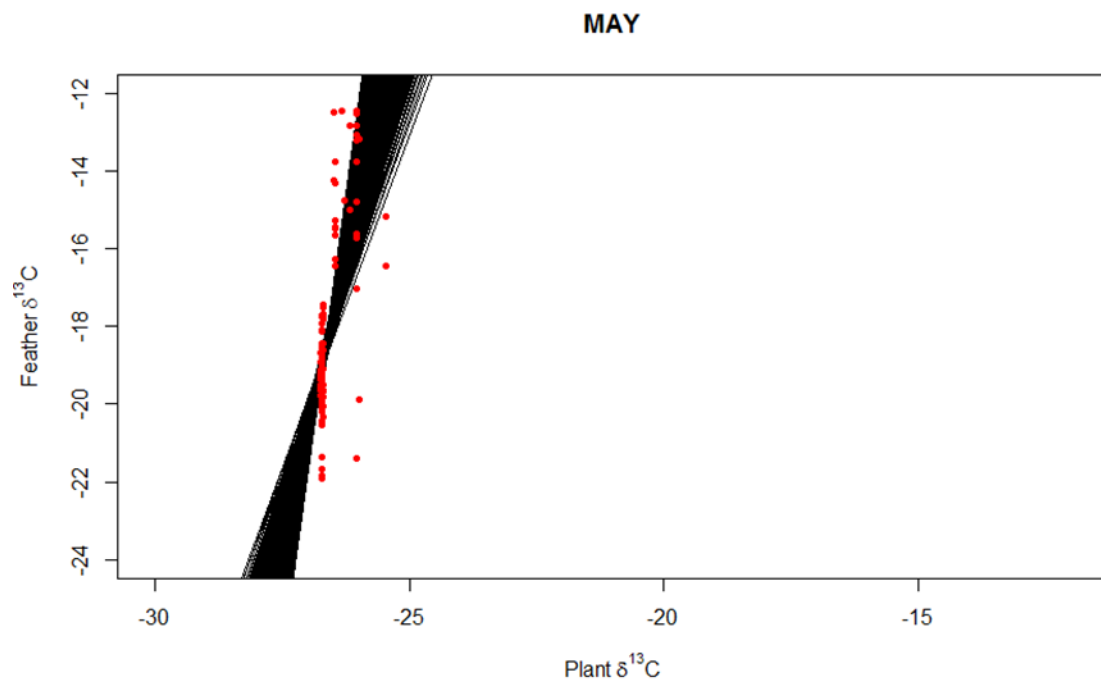


Figure 16 May isoscape environmental to feather tissue isoscape calibration points and equations. All other isoscape calibration points and rescaling equations are included in *Appendix 2*.

June, the most biologically relevant month to the sampled population of mountain plover chicks, was expected to yield high accuracy and precision assignments; however, this was not the case. June assignments typically had the lowest relative accuracy as well as relatively low precision. One factor causing this may be due to the influence of a cluster of mountain plover chicks sampled in a poorly modeled environment. As

presented in *Fig. 17*. The points highlighted by the purple circle are all mountain plover feather samples that did not correspond to the values predicted by the isoscape. For these points, feather $\delta^{13}\text{C}$ values indicate C_4 -dominant diet, while predicted leaf $\delta^{13}\text{C}$ values indicate a C_3 -dominant environment. These feather samples all come from the same sampling location (*Site C in Fig. 11*), classified as a fallow field (Wunder *et al.*, 2005). This mismatch indicates that this site was poorly modeled by the isoscape. This could potentially be due to the coarse spatial grain of the isoscape. The isotopic signature of each pixel is an average of all the estimated vegetation included in the 5-min area. Therefore, it is possible that the $\delta^{13}\text{C}$ values reflected in the tissues were assimilated from

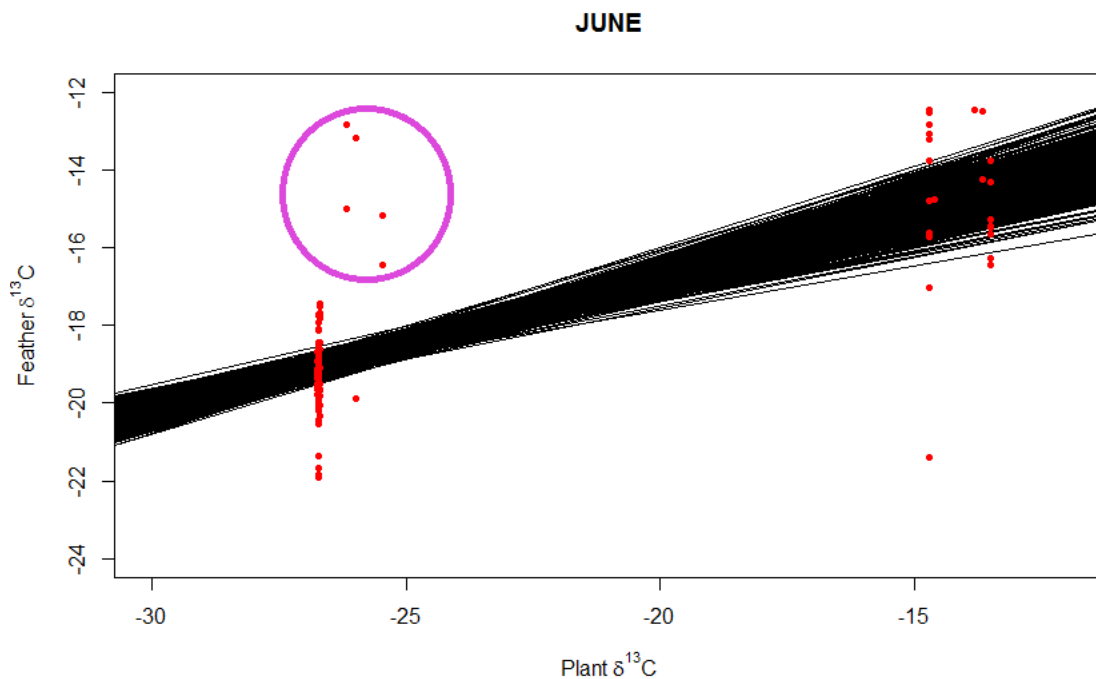


Figure 17 Known-origin feather calibration points and isoscape rescaling equations for the June isoscape. The x-axis are original $\delta^{13}\text{C}$ environmental isoscape values extracted to the known-origin feather sampling locations. The y-axis depicts sampled feather $\delta^{13}\text{C}$ values. The lines depict the 1000 bootstrapped rescaling equations. The points highlighted in the purple circle are feather samples that do not match the modeled environment.

a site, such as a fallow field, that was not spatially dominant, and the associated $\delta^{13}\text{C}$ signal was muted by the $\delta^{13}\text{C}$ signal of the more dominant vegetation cover. Another possibility is that the long-term agricultural and climate data sets used to generate the isoscape is not representative of the environment this particular group of mountain plover chicks were interacting with. In other words, “statistical June” is not equivalent to June, 2001.

Additionally, the monthly isoscapes are limited in that they are binary representations of C_3 vs. C_4 dominance in mixed grasslands areas. The methods used to model plant functional types require a temporal factor to predict mixed grasslands. Mixed grassland pixels are classified as entirely C_3 or entirely C_4 at the monthly temporal resolution. For example, the early summer months such as June may have C_4 dominant environmental conditions; however, there is still likely a considerable C_3 presence due to the environmental conditions of the previous month. Also, it is important to note that the monthly data is an arbitrary temporal block, as are the seasonal groupings, and both are subject to the modifiable temporal unit problem (MTUP). For example, the results may be different if the early season was defined as February, March, April instead of March, April, May. Much like the first law of geography, which states “all things are related, but things that are close together are more closely related,” the vegetation communities that occur in close temporal sequence are related. The binary modeling of the grasslands at a monthly temporal resolution effectively reduces within grassland heterogeneity.

The annual isoscape assignments were most similar to the mid-season and summer-month (i.e., July and August) isoscapes. The annual representation of leaf $\delta^{13}\text{C}$ is

artificial in the sense that annual average vegetation is an environment that an animal will never actually interact with at a particular point in time. This is particularly important when working with migratory animals, which interact with different environments across a landscape as a function of time.

4.3.3 Assessing Geographic Assignment Efficacy

This analysis has highlighted some limitations of using relative probability intervals to assess the efficacy of geographic origin assignments. For assignments with very low precision, corresponding accuracy metrics are generally high, because there is almost equal probability that the individual came from any location on the continent. Relative probability is calculated by normalizing posterior probability by the maximum value of the data set; therefore, in cases of low precision, the resulting relative probability of every pixel is quite high across large spatial extents.

The similarity metric was also problematic metric because isoscapes are not necessarily separate (i.e., independent) models, but rather permutations of the same model. The only differences between isoscapes occur in natural herbaceous areas; therefore, there was inherently a high similarity between assignments, particularly in the cropland- and forest-dominated areas. Additionally, there is a potential problem in assessing the similarity between isoscape assignments that were generated across temporal scales (e.g., annual isoscape assignments versus a seasonal or monthly isoscape assignment), again because the coarser temporal resolution model is not independent of the data included in the finer temporal resolution model. Assignments between temporal

resolutions corresponding to the same dates will inherently be more similar than two isoscape assignments being compared for different dates at the same temporal resolution.

4.5 Conclusion

In order to effectively use high temporal resolution models, it is critical to have a basic understanding of the ecological properties of the migratory animals being assigned. For example, comparing tissues to an environment the organism did not interact with (e.g., biologically irrelevant temporal windows) would yield meaningless results as there is not a connection between the organism and modeled environment. For example, the assignment of mountain plover chick tissues to biologically irrelevant time periods appeared to result in improved precision; however, this gain in precision was artificial and caused by forcing an isoscape calibration equation between a wide range of $\delta^{13}\text{C}$ feather values and a narrow range of environmental $\delta^{13}\text{C}$ predictions. Care should be taken to not assign tissue to a habitat- either in time or space- that the tissue was not grown in.

Isoscapes corresponding to most of the biologically relevant time frames (i.e., annual, mid-season, July and August) resulted in assignments with high accuracy, but relatively low precision. The low precision is likely a result of the assignments being bimodal, identifying birds as either growing tissue in a C_3 or C_4 environment. Using $\delta^{13}\text{C}$ isotopes for assignment allows for tissues to be identified as either coming from C_3 - or C_4 -dominant environments. Further constraining assignments by limiting the assignment region to the known breeding range or employing other isotopes would likely improve prediction.

One biologically relevant month, June, resulted in assignments with relatively low precision and accuracy. This is potentially due to a cluster of mismatched points all coming from the sample site, where the feathers of the birds had a distinct C₄ signal, but the isoscape predicted the environment as C₃. This mismatch could be caused by large pixels muting the influence of local sites, such as crop fields, or mixed grasslands at the local scale, where natural grasslands were classified as either C₃ or C₃. By examining calibration points and rescaling equations, clusters can be identified and characterize potential sources of uncertainty within the isoscape. Further refinement of the current methods to generate the isoscapes of differing temporal resolutions, such as considering active vs. dormant vegetation, or modeling mixed grasslands at monthly resolution could improve both the accuracy and precision of the models. Additionally, the 5-min spatial resolution of each of these models is much larger than the 56 hectares that a mountain plover would assimilate while feathering. This is the basis of my third research question: how does spatial resolution impact the efficacy of using $\delta^{13}\text{C}$ models used for geographic origin assignment?

CHAPTER 5: The Impact of Spatial Resolution on Modeling Stable Terrestrial Carbon Isoscapes

5.1 Introduction

What is the impact of spatial resolution on overall confidence in the model? What is the impact of aggregating fine-grained spatial data versus resampling coarse-grained spatial data on the final data product?

Spatial resolution of raster data sets, or spatial “scale,” can be defined as the size of the two-dimensional raster cells, or pixel dimensions (Goodchild, 2011). Isoscapes are represented as a gridded, continuous response variable. The pixels of an isoscape estimate the approximate isotope value for the area of the landscape represented by the pixel. The annual 5-min isoscape generated in Chapter 3 represents the annual and spatial average of each 5-min pixel’s area. Animal tissues for geographic origin assignment are typically spatially coded as coordinates (i.e., a point) which are compared to gridded isoscape values. The change in support problem (COSP) and the modifiable areal unit problem (MAUP) are fundamental spatial considerations often ignored in geographic origin assignments. However, the spatial area over which an organism integrates an environmental isotopic signal varies. By matching the spatial resolution of an isoscape to the spatial grain and extent that the organism integrates, there is potential to improve

origin assignment accuracy, as well as better characterize the processes by which organisms integrate the isotopic signature of their environment.

A common standard protocol for integrating data sets of different spatial resolutions into a common modeling framework (e.g., the IsoMap cyber-GIS system) is to aggregate all of the input data products to the spatial resolution of the data set with the coarsest grain size (Bowen *et al.*, 2014). However, this reduces variance of data values, as well as spatial heterogeneity. This research question seeks to explore the impact of varying spatial resolutions on confidence in the final model predictions. Spatial resolution is an important consideration in environmental modeling prediction, because both aggregating and disaggregating data introduce error into the final product.

Spatial resolution defines the level of detail and content of a spatial model (Goodchild, 2010). With coarser spatial data, detail is lost. However, coarser spatial resolution generally corresponds to an increase in accuracy. In the geographic assignment example, the “correct” assignment location is more likely to be identified when larger grid cells are used. *Fig. 18* depicts a theoretical example of the potential impact of spatial resolution on a geographic assignment prediction. At a coarse spatial resolution, more area is included, increasing accuracy (there is a higher probability that the “true origin” is included in the predicted area) but reducing precision. The finer resolution model has higher precision, but potentially lower accuracy.

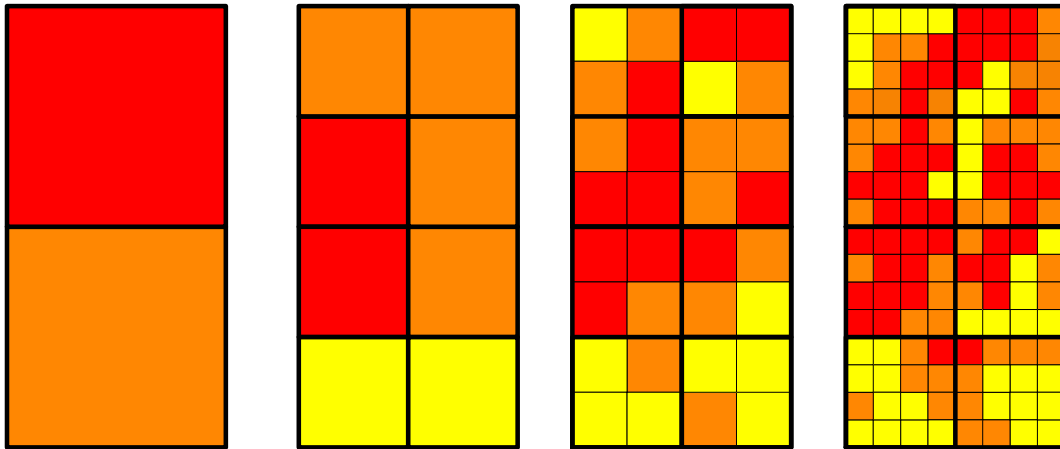


Figure 18 Example of varying spatial resolution of raster data model, and potential impact on geographic assignment. Red represents high probability of origin, and yellow represents lower probability.

Another consideration is how well the spatial resolution of an ecological model “matches” the phenomena being represented. For example, a recent study by Griffith *et al.*, (2015) has shown that the spatial resolution of a model is an important consideration and should be matched to the spatial resolution the processes occur at (e.g., the area of environment that an organism assimilates into its tissues). In this way, modifying the spatial resolution of an ecological model may improve accuracy. By paying attention to how spatial resolution impacts the representation of a phenomena or performance of a spatial model, traditional geographic “problems” may be leveraged to characterize the processes being studied (*Section 2.1.6*). This chapter explores the impact of spatial resolution on the $\delta^{13}\text{C}$ isoscape and the geographic origin assignment of mountain plover chicks.

5.2 Methods

To explore the impact of spatial resolution on modeling outcomes, two additional annual isoscape models are generated following the methods described in Chapter 3.2, resampling all input data to two different input resolutions. The spatial resolutions tested were determined and constrained by the spatial resolutions of the input data sets. The global input data sets that were used to generate the North America product (*Table 2, Chapter 3*) range in spatial resolution from 500 meter to 10 minute (~18.5 km at 40° latitude). For the fine-grained data product, all input data are disaggregated to 1-km spatial resolution. For the spatially coarse data product, all input data are aggregated to 10-min spatial resolution.

To make the raster grids of each data product compatible, the resample function in the R raster package is used to snap each data product to perfectly matching raster grids. For continuous data, the bilinear resampling method is used; for categorical data (e.g., climate masks, land-cover classes), the nearest neighbor resampling method is used. The resampling of data to coerce matching raster grids is an instance of the MAUP zonation effect, and results in uncertainty and error in the final model.

The impact of spatial resolution is tested using the geographic origin assignment methods detailed in *Section 4.2.2*. The resulting products are also compared to the annual 5-min spatial resolution model generated in Chapter 3. Due to the memory intensive demands of processing 1-km resolution data, all three spatial resolution rasters (1-km, 5-min and 10-min) are cropped to the spatial extent corresponding to the approximate known mountain plover breeding range (Wunder *et al.*, 2008; *Fig. 18*). I assess the

accuracy and precision of each model using the methods, described in *Section 4.2.3* (Vander Zanden *et al.*, 2014). The population-level accuracy was assessed as the proportion of known-origin sites (of the assigned individuals) included across a sequence of relative probability thresholds, and precision is measured as the median surface area included at each relative probability interval. The individual-level accuracy metric is assessed as the difference between the relative assignment probabilities for two different isoscape assignments at the known-origin site, and precision is assessed as the difference in surface area included at the relative probability threshold equal to that predicted at the known-origin location.

However, in contrast to Vander Zanden *et al.* (2014), similarity is not assessed because this metric requires the pixels to be matching in size during assignment. Because all of the pixel probabilities within the extent are rescaled to sum to 1, the fine-grained resolutions, having more pixels, will inherently have lower relative probabilities.

5.3 Results

5.3.1 The Impact of Spatial Resolution on Modeled $\delta^{13}\text{C}$ in the Mountain Plover Breeding Range

The general spatial patterns of $\delta^{13}\text{C}$ are very similar across all three model variations generated at differing target spatial resolutions (*Fig. 19*). Although the 1-km model output does have more spatial variation than the two coarser permutations, there is some detectible blockiness due to the coarser cropland and climate data used in the model. Most notably, the range of isotopic values mapped decreases as spatial resolution

decreases. For the full continental extent of the isoscape, the range of pixel values predicted for both the 1-km and 5-min products was -27.2 to -12.5. For the 10-min product, the range of values predicted was -27.2 to -12.8, indicating that no pixels were modeled as entirely C₄ dominant at a 10-min spatial resolution. This smoothing is most apparent in the southwest corner of the mapped extent. In the 1-km permutation, there are isotopically enriched values depicted in dark blue, as well as no-value pixels (white). This area is a desert near Flagstaff, AZ. The white pixels are classified as 100% barren (0% vegetation) by the vegetation continuous fields (VCF) data set, and therefore have no leaf $\delta^{13}\text{C}$ signature.

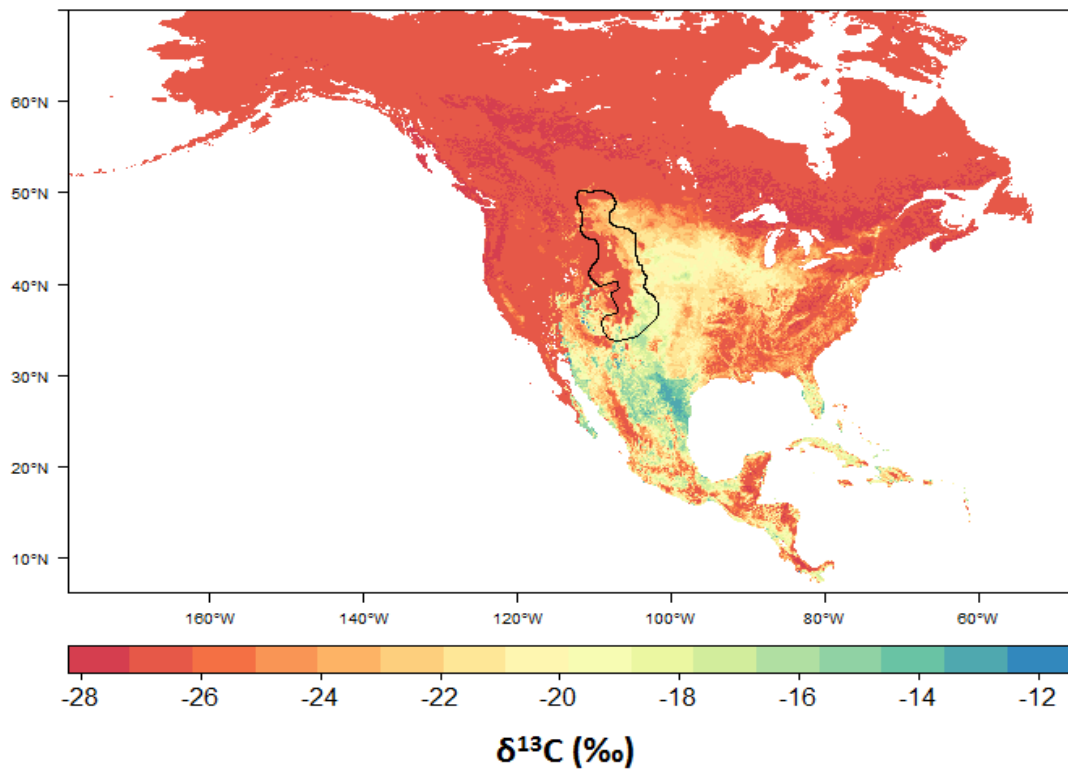


Figure 18 The annual $\delta^{13}\text{C}$ isoscape at five-minute spatial resolution. The mountain plover breeding range is outlined in black (Wunder *et al.*, 2005).

5.3.2 Geographic Origin Assignment Evaluation on Spatial Resolution

There was essentially no difference in population-level accuracy for the three isoscapes of different spatial resolutions, nor for population-level precision (Fig. 20). Individual-level accuracy was compared as the relative probability predicted at the known-origin location by each geographic origin assignment model. However, the comparison of individual-level between spatial resolution assignments did result in significant differences between the 10-min and 1-km models ($t = -4.573$, $p\text{-value} < 0.001$)

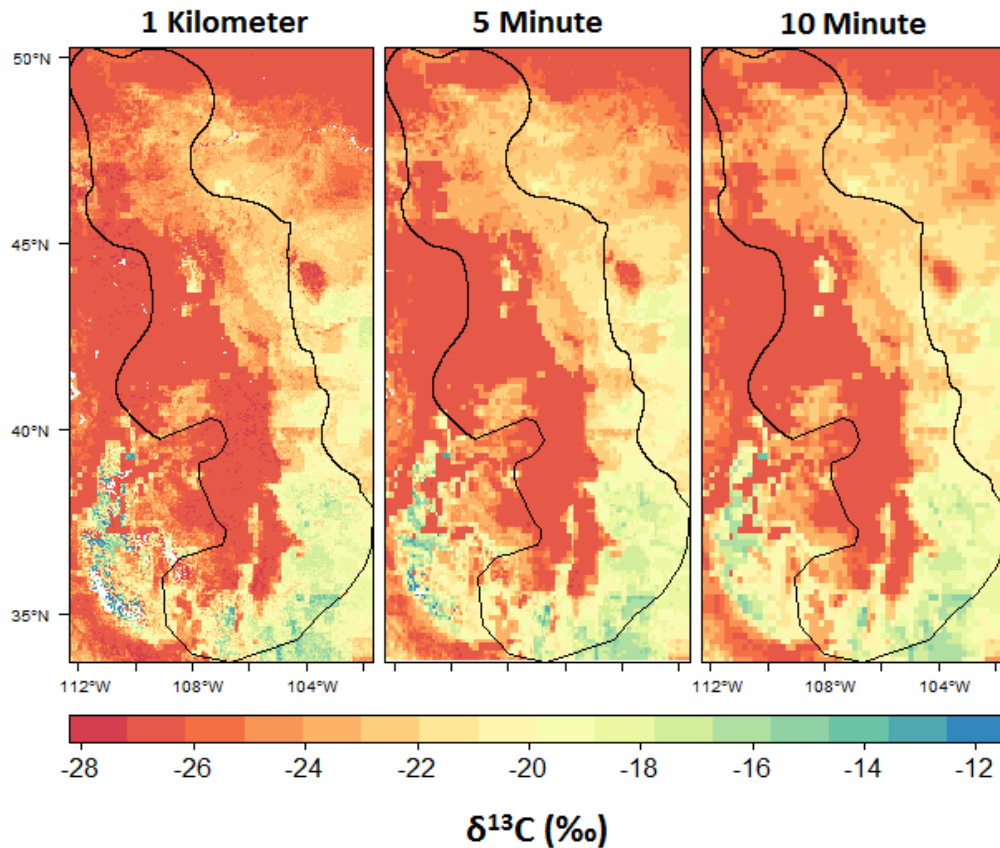


Figure 20 Three $\delta^{13}\text{C}$ isoscape products in the known mountain plover breeding range. The leftmost isoscape is 1-kilometer spatial resolution, the middle isoscape is 5-minute spatial resolution, and the rightmost isoscape is 10-minute resolution.

with the 10-min isoscape assignments having higher accuracy. Using relative probability accounts for the differences in the number of pixels included in each raster at different spatial resolutions. The differences in accuracy between the 5-min and 1-km ($t = 2.782$, $p\text{-value} = 0.006$) and the 10-min and 5-min ($t = -2.689$, $p\text{-value} = 0.008$) were also statistically significant, with increased spatial resolution (i.e., finer grain size) resulting in decreased accuracy.

The individual-level precision assessments had similar results as the individual-level accuracy. The 10-mi and 1-km isoscape assignments had the largest differences in assignment precision, with the 1-km assignments having higher precision. The differences in precision between 5-min and 1-km assignments and the 10-min and 5-min assignments were both significantly different from one another. As expected, the higher resolution assignments had higher precision relative to coarser resolution assignments.

5.4 Discussion

Given the three spatial resolutions being tested, there was not a significant change in geographic origin assignment efficacy at the population level, although there were statistically significant differences when assignments were compared at the individual-level. This tradeoff in accuracy and precision at the individual level is not surprising in itself, as increased model precision is often at the expense of accuracy, and vice-versa.

The reason the individual level accuracy and precision metrics resulted in statistically significant differences, whereas the population-level metrics did not. With a greater number of pixels, there is likely a greater number of distinct relative probability

values. However, the population-level accuracy metric divides (bins) the relative probabilities into 99 sequence intervals; therefore, the population-level accuracies are near equivalent for all three isoscapes (*Fig. 21*). The same type of phenomena occurs with the precision metrics. The binning at the population-level causes differences between the isoscape permutation assignments to be very small, but at the individual level they are statistically significant.

These results should be interpreted through the perspective of improving geographic origin assignment utility using isoscapes. First, it should be noted that geographic origins surfaces do not identify the specific location (pixel) that a migratory animal comes from, as there are likely many pixels that have similar or the same isotopic values within a given extent; rather, the goal is to constrain the possible environment. Therefore, the small differences in relative probabilities between pixels gained from finer spatial resolution and improved precision may not be meaningful in the context of geographic origin assignments. Additionally, the processing of high-resolution data is

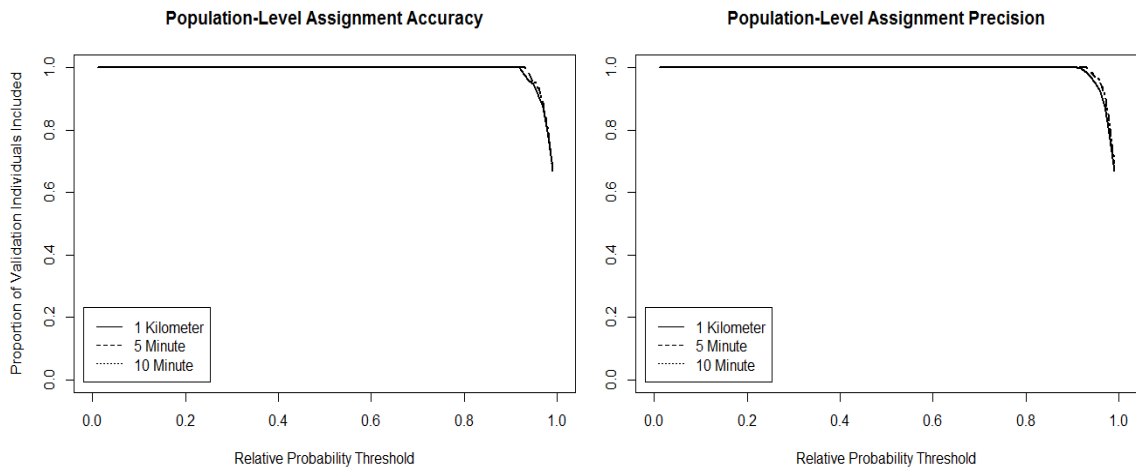


Figure 21 Population-level assignment accuracy and population-level precision.

substantially more memory intensive, which can quickly become prohibitive in terms of time and processing power.

Additional limitations to this analysis should be considered. First, each of the grid sizes tested are all still too coarse to represent a biologically relevant spatial scale for the mountain plover chicks (~56 hectares). Thus, the mountain plover chicks are effectively assimilating a point-sized area relative to the smallest pixel area. In terms of the change of support problem (COSP), the comparison of bird feather to the isoscape raster surface is a point (individual origin) to raster conversion. Finally, the input data did not change between models. If all input data had been collected at the finest grid size, the results of the assignments at each resolution may have been more different.

5.5 Conclusion

Although there were statistically significant differences identified between each isoscape permutation of differing spatial resolutions, I found there was not a meaningful impact of spatial resolution on geographic origin assignments in the context of interpretation. The improvements in precision gained from assignments using finer spatial resolution isoscapes were always at the cost of accuracy. Because models with higher spatial resolution are often proportionately more memory intensive, I propose that it is best to use isoscapes generated at the coarsest-resolution input data set for initial geographic origin assignments.

The concept of harnessing areal units to characterize the way an animal like mountain plover chicks interact with an environment may be more useful at smaller

spatial extents, where data may be collected at specific spatial resolutions, as many ecological studies examining intrinsic scale have begun to explore. The remotely sensed data used to generate the $\delta^{13}\text{C}$ isoscapes all occur at a spatial resolution considerably coarser than what would be relevant to flightless chicks. Starting with data at a finer spatial grain than the estimated area of assimilation for the animal being studied, and then aggregating up to find a point of improved model fitness would be a better strategy to test this concept, and should be further explored.

CHAPTER 6: Summary and Conclusions

6.1 Summary of Research

Overview

This research was driven by an interest in leveraging the spatial and temporal characteristics of ecological data and models to characterize the ecological phenomena being researched. I explored these concepts in the context of modeling the spatial patterns of stable carbon isotopes for the continent of North America. I experimented with model rules and parameters, temporal resolution, and spatial resolution as factors that impact the efficacy of creating isoscapes, and performing geographic origin assignments. The findings presented in this thesis serve as initial investigations of the representation of dynamic ecological processes in static models, as well as possible best practices for using a terrestrial stable carbon isoscape for geographic origin assignments.

Research Question 1: What is the spatial distribution of vegetation (i.e., leaf) stable carbon isotopes on the continent of North America?

Through a comparison to soil organic matter observations in the Great Plains, USA, the results from my study demonstrated the potential limitations of using mean-monthly temperature as the climate metric for the Collatz-crossover model in North

America, as well as the potential need for multiple “cross-over temperatures” in relation to latitude. I used a $\delta^{13}\text{C}$ SOM data set for validation data because unlike vegetation surveys, SOM represents a long-term average of plant functional type abundancies, and therefore is appropriate data to test an annual isoscape. However, there was not strong agreement between the observed data set and any of the isoscape permutations, indicating general model uncertainty. Uncertainty was greatest in deserts and higher elevations where the land-cover rules and crossover metrics had the greatest impact. The max-temperature/max-herbaceous isoscape model permutation had the lowest residuals in comparison to the validation data set and therefore was used in the following research chapters.

Research Question 2: How can seasonal variation of vegetation cover be represented in a static model? In what way does temporally aggregating the seasonal variation of mid-latitude grasslands impact confidence in the final model?

The known-origin geographic origins assignments of mountain plover chicks indicated that temporal resolution did have a significant impact on isoscape modeling and geographic origin assignment efficacy. In general, Accuracy and precision improved using biologically relevant finer resolution temporal windows (e.g., mid-seasonal assignments were more accurate and precise than annual assignments, because mid-seasonal isoscapes do not average in periods of time that the birds do not interact with their environment). Most geographic origin assignments based on biologically relevant temporal windows (i.e., mid-season, July and August) had relatively high accuracy and lower precision. However, the relatively high precision found in assignments based off

biologically irrelevant time periods was artificial and only due to the extremely steep tissue calibration equations applied to the environmental isoscapes. Geographic origin assignments should not be created with isoscapes that represent an environment the organism could not have grown tissue in, spatially or temporally.

June, which is a biologically relevant time period created assignments with low accuracy and precision relative to other months. This is likely a result of mismatched feather $\delta^{13}\text{C}$ values to environmental isoscape $\delta^{13}\text{C}$ predictions resulting in a cluster of points skewing the calibration equations, thereby reducing both accuracy and precision. This mismatch indicates limitations in the isoscape model, perhaps resulting from coarse spatial resolution or long-term statistical data not appropriately representing the environment experienced by the mountain plover chicks. The major implication of these results is that it is very important to review the observational data, isoscape predictions, and calibration curves prior to generating geographic origin assignments.

Research Question 3: What is the impact of spatial resolution of the final product on overall confidence in the model? What is the impact of aggregating fine-grained spatial data versus resampling coarse-grained spatial data on the final data product?

The purpose of an isoscape-based geographic origin assignment is to generally constrain the environment from which an animal originates, rather than to identify a specific point of origin. For this reason, using coarser spatial resolution models might be more ideal, because my results showed that coarser resolution isoscape improved accuracy and reduced computer processing times. There was little difference between geographic origin assignments between the 1-km, 5-min and 10-min isoscapes at the

population level in terms of accuracy and precision. However, at the individual-level, the efficacy assessment metrics indicated there was a significant difference in accuracy and precision between each set of assignments. For example, finer-resolution assignments had higher precision but lower accuracy, and at coarser resolutions assignments had higher accuracy and lower precision. There was always a tradeoff in accuracy and precision with changes to the spatial resolution of the isoscape that was used.

It is possible that the different spatial resolutions did not show obvious improvement over one another because none came close to a biologically relevant spatial scale. Even the 1-km spatial resolution model is still much coarser than the 56-ha patches that mountain plover chicks inhabit prior to their first feathering. Additionally, the input data did not change between isoscape variations, which meant spatial heterogeneity was only preserved for data that had a spatial resolution equal to or finer than the target resolution.

6.2 Limitations and Suggested Improvements

Many of the model limitations of the terrestrial carbon isoscapes were exacerbated in the seasonal and monthly isoscape permutations by the methods used to vary temporal resolution. Because many of the input data layers represented annual averages (e.g., croplands, vegetation cover) there was no way to account for active versus dormant vegetation, or multi-cropping cycles on a landscape. Lacking temporally variable data on active vegetation is problematic for the seasonal and monthly isoscapes because they are modeled at fine enough temporal resolutions where whole pixels may not contain active vegetation for that entire time period.

Even the natural herbaceous grasslands, which were modeled as temporally variable, were limited by the modifiable temporal unit problem (MTUP). The monthly temporal resolution is an arbitrary temporal unit, and may not be appropriately capturing temporal variability. Additionally, the monthly isoscapes do not model mixed grasslands, and therefore lose some model heterogeneity as the grasslands are binary. This also does not take into account that the growing season is a progression, and grassland communities are directly impacted by the vegetation that grows before it (i.e., there is not an instantaneous turnover from C₃ dominance to C₄ dominance).

Results for Research Questions 2 and 3 are also limited in that geographic origin assignments are both isoscape and species specific, meaning that thus far these results only have been applied to mountain plovers and a specific set of $\delta^{13}\text{C}$ isoscapes. The mountain plover breeding range is spatially constrained to the mountains, foothills and immediate grasslands east of the Rockies. This means that large portions of the continent were not sampled for calibration between isoscape and feather $\delta^{13}\text{C}$ values. Due to the limited range of the mountain plovers, I was unable to leverage the dramatic temporal variation that occurs in the Chihuahuan desert and the majority of the Great Plains to explore isoscape properties or geographic origin assignment methods. The geographic origin assignment validations would potentially be improved by using a more widely distributed species for known-origin calibrations and assignments. Also, the measures of variance from calibration and rescaling equation residuals need to be further tested to assure they are appropriate and effective.

I found little variation in the population-level accuracy and precision curves relative to results reported for previous studies. The accuracy and precision metrics were originally developed and used in the context of precipitation-based $\delta^2\text{H}$ isoscape assignments (Vander Zanden *et al.* 2014). My results suggest some limitations in these assessment metrics in the context of assignments based on $\delta^{13}\text{C}$ isoscapes. The population-level metrics may have lacked texture because geographic origin assignments based on $\delta^{13}\text{C}$ isoscapes are somewhat bi-modal. This is because as birds are in-essence classified as belonging to either a C_3 -dominant environment or a C_4 -dominant environment. The lack of differentiation between the population-level metrics may be related to binary individual assignments, resulting in a muted population-level impact.

6.2 Future Research

The initial motivation behind my research was to characterize the biogeography of both plant and animal species by leveraging the spatial and temporal qualities of ecological models. The results of my research provide a foundation for many different avenues of further exploration.

Initial isoscape generation could be explored by comparing the isoscapes generated using the Collatz-crossover model, which differentiates C_3 and C_4 dominance by using a crossover temperature, to the research using plant functional type phenology as a mechanism for differentiating between C_3 and C_4 vegetation (*Section 2.2.5*). Using phenology instead of climate conditions may account for any “lag time” that may occur between environmental conditions and present vegetation. Previous work has modeled phenology using MODIS bi-weekly NDVI data (Wang *et al.*, 2011; Wang *et al.*, 2013;

Zhong *et al.*, 2015), and these models have finer temporal resolution than the monthly climate metrics.

Further research on seasonal variation is necessary for two reasons. First, finer-temporal resolution isoscapes may improve accuracy and precision by narrowing the time frame to a biologically relevant time period. Second, future research could examine whether finer-temporal resolution isoscapes can be leveraged to identify both where and *when* a migratory animal occupies a habitat. This could be tested by applying known-origin geographic assignment validations on individuals that breed or molt during a different time period.

The examination of temporal and spatial resolution in the context of geographic origin assignments and isoscapes should also be applied to stable hydrogen isoscapes. It would be relevant to understand how spatial and temporal variation impact hydrogen isotope representation across a landscape. Additionally, multi-isotope geographic origin assignments have the potential to increase accuracy and precision of predictions. Combining stable hydrogen and stable carbon isoscapes may result in new opportunities.

Finally, testing the impact of modifying areal units at a much smaller spatial extent may lead to more opportunities for leveraging model properties to characterize ecological phenomena. In a perfect world, animals are “ideal integrators” of the environment, where all sources of environmental isotopes are equally integrated into animal tissue. Animals selectively consume food within an ecosystem, meaning they assimilate their environments in different ways, as well as different spatial extents. It may be possible that the MAUP could be harnessed to characterize the ways in which an

animal “uses” its environment, both in terms of scale, as well as food-web partitioning. For this reason, the MAUP, the MTUP and the COSP should continue to be explored in the context of environmental stable isotopes and geographic origin assignments. Ultimately, considering the parameters used to model environmental properties, allows further explorations concerning the interactions of animals and their landscapes.

REFERENCES

- Bowen, G.J. (2010) Isoscapes: Spatial pattern in isotopic biogeochemistry. *Annual Review of Earth and Planetary Sciences*, **38**, 161-187.
- Bowen, G. J. & West, J.B. (2008). Isotope Landscapes for Terrestrial Migration Research. *Tracking Animal Migration with Stable Isotopes* (ed. by K.A. Hobson & L.I. Wassenaar), pp.79-106. Elsevier, USA.
- Bowen, G.J., Liu Z., Vander Zanden, H.B., Zhao L., & Takashi, G. (2014) Geographic assignment with stable isotopes in IsoMAP. *Methods in Ecology and Evolution*, **5**, 201-206.
- Bowling, D.R., Pataki, D.E., & Randerson, J.T. (2008) Carbon isotopes in terrestrial ecosystem pools and CO₂ fluxes. *New Phytologist*, **178**, 24-40.
- Cerling T.E. & Harris, J.M. (1999) Carbon isotope fractionation between diet and bioapatite in ungulate mammals and implications for ecological and paleoecological studies. *Oecologia*. **120**, 347–63
- Cerling, T.E., Harris, J.M., MacFadden, B.J., Leakey, M.G., Quade J., Eisenmann, V., & Eleringer, J.R. (1997) Global vegetation change through the Miocene/ Pliocene boundary. *Nature*. **389**, 153–158.
- Cerling, T.E., Harris, J.M., & Passey, B.H. (2003). Diets of east African Bovidae based on stable isotope analysis. *Journal of Mammalogy*, **84**, 456-470.
- Cerling, T.E., Wynn, J.G., Andanje, S.A., Bird, M.I., Korir, D.K., Levin, N.E., Mace, W., Macharia, A.N., Quade, J., & Remien, C.H. (2011) Woody cover and the hominin environments in the past 6 million years. *Nature*, **476**:51-56.
- Cheng, T. & Adepeju, M. (2014) Modifiable Temporal Unit Problem (MTUP) and its effect on space-time cluster detection. *PLoS One*, **9**, 1-10.
- Collatz, G.J., Berry, J.A., & Clark, J.S. (1998) Effects of climate and atmospheric CO₂ partial pressure on the global distribution of C₄ grasses: present, past, and future. *Oecologia*, **114**, 441–454.
- Cressie, N. (1996). Change of support and the modifiable areal unit problem. *Geographical Systems*, **3**, 159-180.

- Daly, C., Gibson, W.P., Taylor, G.H., Johnson G.L., & Pasteris P. (2002) A knowledge-based approach to the statistical mapping of climate. *Climate Research*, **22**, 99-123.
- Davidson, A., & Csillag, F. (2003). A comparison of three approaches for predicting C₄ species cover of northern mixed grass prairie. *Remote Sensing of Environment*, **86**, 70–82.
- DeFries, R.S., Field, C.B., & Fung, I. (1995) Mapping the land surface for global atmosphere-biosphere models: toward continuous distributions of vegetation's functional properties. *Journal of Geophysical Research: Atmospheres*, **100**, 2867-2882.
- DeNiro, M. J. & Epstein, S. (1978) Influence of diet on the distribution of carbon isotopes in animals. *Geochimica et Cosmochimica Acta*, **42**, 495-506.
- Diefendorf, A.F., Mueller, K.E., Wing, S.L., Koch, P.L., & Freeman, K.H. (2010) Global patterns in leaf ¹³C discrimination and implications for studies of past and future climate. *Proceedings of the National Academy of Science*. **107**, 5738-5743.
- Edwards E.J., Osborne, C.P., Strömberg, C.A.E., Smith S.A., & Consortium, C.G. (2010) The origins of C₄ grasslands: integrating evolutionary and ecosystem science. *Science*, **328**, 587–591.
- Efron, B. & Tibshirani, R.J. (1993). The Bootstrap Estimate of Standard Error. *An Introduction to the Bootstrap*. pp 45-57. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, USA.
- Ehleringer, J.R., Sage, R. F., Flanagan, L.B., & Pearcy, R.W. (1991) Climate change and the evolution of C₄ photosynthesis. *Trends in Ecology & Evolution*, **6**, 95–99.
- Ehleringer, J.R., & Monson, R.K. (1993) Evolutionary and ecological aspects of photosynthetic pathway variation. *Annual Review of Ecology and Systematics*, **24**, 411–439.
- Ehleringer, J.R., Cerling, T.E., & Helliker, B.R. (1997) C₄ photosynthesis, atmospheric CO₂, and climate. *Oecologia*, **112**, 285–299.
- Ehleringer, J.R., Casale, J.F., Lott M.J., & Ford, V.L. (2000) Tracing the geographical origin of cocaine. *Nature*, **408**, 311-312.
- Faber-Langendoen, D. & Tester, J. R. (1993) Oak Mortality in Sand Savannas Following Drought in East-Central Minnesota. *Bulletin of the Torrey Botanical Club*, **120**, 248-256.
- Fassnacht, K.S., Cohen, W.B., & Spies, T.A. (2006) Key issues in making and using satellite-based maps in ecology: A primer. *Forest Ecology and Management*, **222**, 167-181.

- Fei, S., Guo, Q., & Potter, K. (2016) Macrosystems ecology: novel methods and new understanding of multi-scale patterns and processes. *Landscape Ecology*, **31**, 1-6
10.1007/s10980-015-0315-0
- Foody, G.M., & Dash, J. (2007) Discriminating and mapping the C₃ and C₄ composition of grasslands in the northern Great Plains, USA. *Ecological Informatics*, **2**, 89-93.
- Godsoe, W. (2014) Inferring the similarity of species distributions using Species' Distribution Models. *Ecography*, **37**, 130-136.
- Goodchild, M.F. (2013) Prospects for a Space- Time GIS. *Annals of the Association of American Geographers*, **103**, 1072-1077
- Goodchild, M.F. (2011). Scale in GIS: An overview. *Geomorphology*, 139, 5-9.
- Goodin, D.G., & Henebry, G.M. (1997). A technique for monitoring ecological disturbance in tallgrass prairie using seasonal NDVI trajectories and a discriminant function mixture model. *Remote Sensing of the Environment*, **61**, 270-278.
- Gotway, C.A., & Young, L.J. (2002) Combining Incompatible Spatial Data. *Journal of the American Statistical Association*, **97**, 632-648.
- Griffith, D.M, Anderson, T.M., Osborne, C.P., Stromberg, C.A.E, Forrestel, E.J., & Still, C.J. (2015) Biogeographically distinct controls on C₃ and C₄ grass distributions: merging community and physiological ecology. *Global Ecology and Biogeography*, **24**, 304-313.
- Griffith, D.M., Cotton, J.M., Powell R.L., & Still, C.J. (2015) Comparing soil and bison $\delta^{13}\text{C}$ to field estimates of C₄ plant abundances in North America. Poster Presentation. AGU Annual Meeting. B53D-0590.
- Guan, L., Liangyun, L., Peng, D., Hu, Y., Jiao, Q. & Liu L. (2012) Monitoring the distribution of C₃ and C₄ grasses in a temperate grassland in northern China using moderate resolution imaging spectroradiometer normalized difference vegetation index trajectories. *Journal of Applied Remote Sensing*, **6**, 1-13.
- Hamil, K.A.D., Iannone III, B.V., Huang, W.K., Fei, S., & Zhang, H. (2016) Cross-scale contradictions in ecological relationships. *Landscape Ecology*, **31**, 7-18.
- Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Carroll, M., Cimiceli, C., & Sohlberd, R.A. (2003) Global percent tree cover at a spatial resolution of 500 meters: first results of the MODIS vegetation continuous fields' algorithm. *Earth Interactions*, **7**, 10.
- Hay, G.J., Marceau, D.J., Dube, P., & Bouchard, A. (2001) A multiscale framework for landscape analysis: Object-specific analysis and upscaling. *Landscape Ecology*, **16**, 471-490.

- Heffernan, J.B., Soranno, P.A., Angilletta Jr., M.J., Buckley, L.B., Gruner, D.S., Keitt, T.H., Kellner, J.R., Kominoski, J.S., Rocha, A.V., Xiao, J., Harms, T.K., Goring, S.J., Koenig, L.E., McDowell, W.H., Powell, H., Richardson, A.D., Stow, C.A., Vargas, R., & Weathers, K.C. (2014) Macrosystems ecology: understanding ecological patterns and processes at continental scales. *Frontiers in Ecology and Environment*, doi: 5-14. 10.1890/130017
- Hobson, K.A., Wassenaar, L.I., Mila, B., Lovette, I., Dingle, C., & Smith, T.B. (2002). Stable isotopes as indicators of altitudinal distributions and movements in an Ecuadorean hummingbird community. *Oecologia*. **136**, 302-308.
- Hobson K.A. & Norris, D.R. (2008) Animal Migration: A context for using new techniques and approaches. *Tracking Animal Migration with Stable Isotopes* (ed. by K.A. Hobson & L.I. Wassenaar), pp.1-19. Elsevier, USA.
- Hobson, K.A. (2011) Isotopic ornithology: a perspective. *Journal of Ornithology*, **152**, S49-S66.
- Hobson, K.A., Van Wilgenburg, S.L. Wassenaar, L.I. & Larson K. (2012) Linking hydrogen ($\delta^2\text{H}$) isotopes in feathers and precipitation: sources of variance and consequences for assignment isoscapes. *PLoS ONE*, **7**, e35137.
- Hobson, K.A., Van Wilgenburg, S.L., Wassenaar, L.I., Powell, R.L., Still, C.J., & Craine, J.M. (2012) A multi-isotope ($\delta^{13}\text{C}$, $\delta^{15}\text{N}$, $\delta^2\text{H}$) feather isoscape to assign Afrotropical migrant birds to origins. *Ecosphere*, **3**, 1-20.
- Jelinski, D.E., & Wu, J. (1996) The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecology*, **11**, 129-140.
- Kemp, P.R. (1983) Phenological Patterns of Chihuahuan Desert plants in relation to the timing of water availability. *Journal of Ecology*, **71**, 427-436.
- Knopf, F.L. & Rupert, J.R. (1996) Productivity and movements of mountain plovers breeding in Colorado. *Wilson Bull*, **108**, 28-35.
- Leff, B., Ramankutty, N., & Foley, J.A. (2004) Geographic distribution of major crops across the world. *Global Biogeochemical Cycles*, doi: 10.1029/2003GB002108.
- Loveland, T.R., & Belward, A.S. (1997) The IGBP-DIS global 1 km land cover data set, DISCover: First results. *International Journal of Remote Sensing*, **18**, 3291-3295.
- MacFadden, B.J. (2000) Cenozoic mammalian herbivores from the Americas: reconstructing ancient direct and terrestrial communities. *Annual Review of Ecological Systems*, **31**, 33-59.

- Marceau, D.J., Howarth, P. J. & Gratton, D.J. (1994) Remote sensing and the measurement of geographical entities in a forested environment. 1. The scale and spatial aggregation problem. *Remote Sensing of the Environment*, **49**, 93-104.
- Monfreda, C., Ramankutty, N. & Foley, J. A. (2008) Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochemical Cycles*, doi: 22:10.1029/2007GB002947.
- Monson, R.K., Littlejon, R.O., Jr & Williams, G.J., III. (1983) Photosynthetic adaptation to temperature in four species from the Colorado shortgrass steppe: A physiological model for coexistence. *Oecologia*, **58**, 43-51.
- Murphy, B. P., & Bowman, D. M. J. S. (2007) Seasonal water availability predicts the relative abundance of C₃ and C₄ grasses in Australia. *Global Ecology and Biogeography*, **16**, 160-169.
- New, M., Lister, D., Hulme, M., & Maken, I. (2002) A high-resolution data set of surface climate over global land areas. *Climate Research*, **21**, 1-25.
- Ode, D. J., Tieszen, L.L., & Lerman, J.C. (1980) The seasonal contribution of C₃ and C₄ plant species to primary production in a mixed prairie. *Ecology*, **61**, 1304-1311.
- Openshaw, S. (1984) Ecological fallacies and the analysis of areal census data. *Environment and Planning*, **16**, 17-31.
- Openshaw, S. & Taylor, P. (1979) A million or so correlation coefficients: three experiments on the modifiable areal unit problem. *Statistical Applications in the Spatial Sciences* (ed. by N. Wrigley). pp. 127-144. Pion, London.
- Pearcy, R.W. (1990) Sunflecks and Photosynthesis in Plant Canopies. *Annual Review of Plant Physiology and Plant Molecular Biology*, **41**, 421-453.
- Pontius, R.G. Jr., Thontteh, O. & Chen, H. (2008) Components of information for multiple resolution comparison between maps that share a real variable. *Environmental Ecological Statistics*, **15**, 111-142.
- Powell, R.L., Yoo, E-H., & Still, C.J. (2012) Vegetation and soil carbon-13 isoscapes for South America: integrating remote sensing and ecosystem isotope measurements. *Ecosphere*, **3**, 109, doi: 10.1980/ES12-00162.1.
- Ramankutty, N., Evan, A.T., Monfreda, C., & Foley, J.A. (2008) Farming the Planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochemical Cycles*, doi: 22:10.1029/2007GB002952.
- Rubenstein, D.R. & Hobson, K.A. (2004) From birds to butterflies: animal movement patterns and stable isotopes. *TRENDS in Ecology and Evolution*, **19**, 5, doi:10.1016/j.tree.2004.03.017.

- Sage R.F. (1999) Why C₄ photosynthesis? *C₄ Plant Biology* (ed. by R.F. Sage & R.K. Monson), pp. 3-16. Academic Press, San Diego.
- Sage, R.F., Wedin, D.A., & Li, M. (1999) The Biogeography of C₄ Photosynthesis: Patterns and Controlling Factors. *C₄ Plant Biology* (ed. by R.F. Sage & R.K. Monson), pp. 313-317. Academic Press, San Diego.
- Still, C.J., Berry, J.A., Collatz, G.J. & DeFries, R.S. (2003) Global distribution of C₃ and C₄ vegetation: carbon cycle implications. *Global Biogeochemical Cycles*, **17**, 1006, doi: 10.1029/2001GB001807.
- Still, C. J. & Powell, R. L. (2010) Continental-scale distributions of plant stable carbon isotopes. *Isoscapes: Understanding movement, pattern, and process on Earth through isotope mapping* (ed. by J.B. West, G.J. Bowen, T.E. Dawson and K. Tu), pp. 179-193. Springer, Dordrecht, the Netherlands.
- Suits, N.S., Denning, A.S., Berry, J.A., Still, C.J., Kaduk, J., Miller, J., & Baker, I.T. (2005) Simulation of carbon isotope discrimination of the terrestrial biosphere. *Global Biogeochemical Cycles*, doi: 19.10.1029/2003GB002141.
- Teeri, J.A. & Stowe, L.G. (1976) Climatic patterns and the distribution of C₄ grasses in North America. *Oecologia*, **23**, 1-12.
- Teeri J.A., Stowe, L.G., & Livingstone, D.A. (1980) The distribution of C₄ species of the Cyperaceae in North America in relation to climate. *Oecologia*, **47**, 307-310.
- Tieszen, L. L., Reed, B. C., Bliss, N. B., Wylie, B. K., & DeJong, D. D. (1997) NDVI, C₃ and C₄ production, and distributions in Great Plains grassland land cover classes. *Ecological Applications*, **7**, 59– 78.
- Tipple, B.J. & Pagani, M. (2007) The Early Origins of Terrestrial C₄ Photosynthesis. *Earth and Planetary Sciences*. **35**, 435-461.
- Tonra, C.M., Both, C., & Marra, P.P. (2015) Incorporating site and year-specific deuterium ratios ($\delta^2\text{H}$) from precipitation into geographic assignments of a migratory bird. *Journal of Avian Biology*, **46**, 266-274.
- Turner, M.G., Dale, V.H., & Gardener, R.H. (1989) Predicting across scales: Theory development and testing. *Landscape Ecology*, **3**, 245-252.
- Turner, M.G. (2005) Landscape ecology: what is the state of the science? *Annual Review of Ecology, Evolution and Systematics*, **36**, 319-344.
- Vander Zanden, H.B., Wunder, M.B., Hobson, K.A., Van Wilgenburg, S.L., Wassenaar, L.I., Welker, J.M & Bowen, G.J. (2014) Contrasting assignment of migratory organisms to geographic origins using long-term versus year-specific precipitation isotope maps. *Methods in Ecology and Evolution*, **5**, 891-900.

Vander Zanden, H.B., Wunder, M.B., Hobson, K.A., Van Wilgenburg, S.L., Wassenaar, L.I., Welker, J.M & Bowen, G.J. (2015) Space-time tradeoffs in the development of precipitation-based isoscape models for determining migratory origin. *Journal of Avian Biology*, **46**, 658-667.

Von Fischer, J.C., Tieszen, L.L., & Schimel, D.S. (2008) Climate controls on C₃ vs. C₄ productivity in North American grasslands from carbon isotope composition of soil organic matter. *Global Change Biology*, **14**, 1-5, doi:10.1111/j.1365-2486.2008.01552.x

Wang, C., Fritschi, F.B., Stacey, G., & Yang, Z. (2011) Phenology-Based Assessment of Perennial Energy Crops in North American Tallgrass Prairie. *Annals of the Association of American Geographers*, **101**, 742-751.

Wang, C. Hunt, E.R. Jr., Zhang, L. & Guo, H. (2013) Phenology-assisted classification of C₃ and C₄ grasses in the U.S. Great Plains and their climate dependency with MODIS time series. *Remote Sensing of the Environment*, **138**, 90-101

Wassenaar, L.I. (2008) An introduction to light stable isotopes for use in terrestrial animal migration studies. *Tracking Animal Migration with Stable Isotopes* (ed. by K.A. Hobson & L.I. Wassenaar) pp. 21-43. Elsevier, USA

West, J. B., Bowen, G. J., Dawson, T. E., & Tu, K. P. (2010). Context and Background for the Topic and Book, *Isoscapes: Understanding movement, pattern, and process on Earth through isotope mapping* (ed. by J.B. West, G.J. Bowen, T.E. Dawson & K. Tu). Springer, Dordrecht, The Netherlands.

Winslow, J. C., Hunt, E. R. Jr., & Piper, S. C. (2003) The influence of seasonal water availability on global C₃ and C₄ grassland biomass and its implications for climate change research. *Ecological Modelling*, **163**, 153–173.

Wu, J., Jones, B., Li, H., & Loucks, O.L. (2013) Key concepts and research topics in landscape ecology revisited: 30 years after the Allerton Park workshop. *Landscape Ecology*, **28**, 1-11.

Wunder, M.B. (2010) Using isoscapes to model probability surfaces for determining geographic origins. *Isoscapes: Understanding Movement, Pattern, and Process on Earth Through Isotope Mapping* (ed. by J.B. West, G.J. Bowen, T.E. Dawson & K. P. Tu) pp. 251-270. Springer, Netherlands.

Wunder, M.B., Kester, C.L., Knopf, F.L., & Rye, R.O. (2005) A test of geographic assignment using isotope tracers in feathers of known origin. *Oecologia*, **144**, 906-617

Wunder, M.B. & Norris, R.D. (2008) Analysis and Design for Isotope-Based Studies of Migratory Animals. *Tracking Animal Migration with Stable Isotopes*. (ed. by K.A. Hobson & L.I. Wassenaar) pp.107-128. Elsevier, USA.

Zhong, C., Wang, C., & Wu, C. (2015) MODIS-Based Fractional Crop Mapping in the U.S. Midwest with Spatially Constrained Phenological Mixture Analysis. *Remote Sensing*, **7**, 512-529.

APPENDICIES

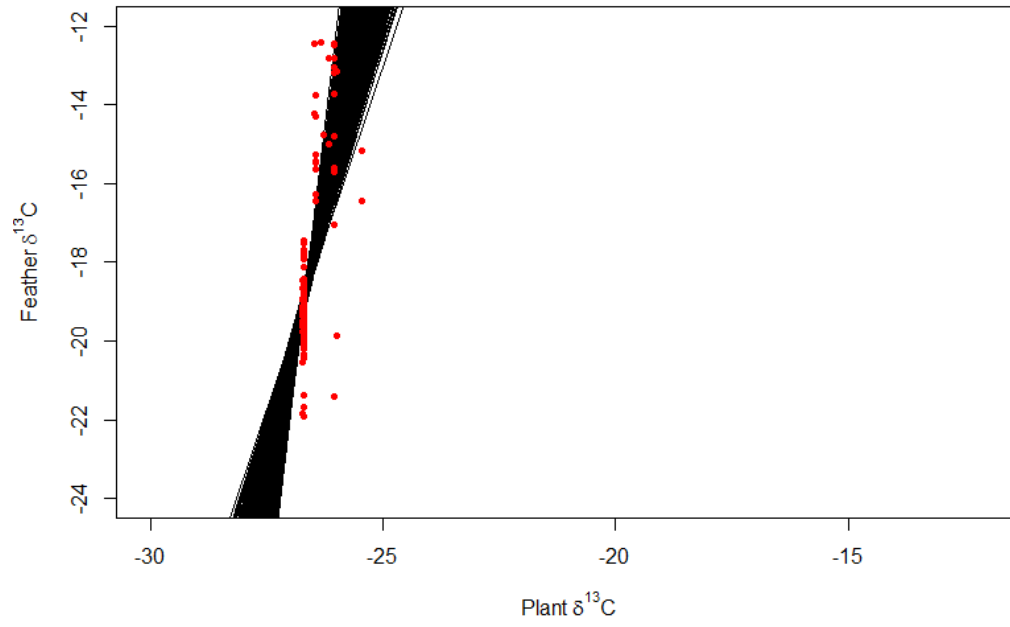
APPENDIX 1: Soil Organic Matter (SOM) Validation Data Set

Site	Latitude	Longitude	SOM $\delta^{13}\text{C}$
Anahuac Wildlife Refuge, TX	29.67	-94.4	-15.0
Clymer's Prairie, TX	33.32	-96.2	-14.4
Lubbock, TX	33.41	-102.1	-15.5
Muleshoe, TX	33.5	-102.4	-14.2
Tridens Prairie, TX	33.64	-95.7	-14.4
Seviellita, NM	34.35	-106.9	-16.7
Woodward, OK	36.42	-99.3	-18.6
Freedom, OK	36.45	-99.4	-14.1
Tallgrass Prairie, OK	36.88	-96.5	-16.3
Diamond Grove, MO	37.03	-94.3	-15.6
Drover's Prairie, MO	38.53	-93.3	-19.3
Land Institute, KS	38.73	-97.6	-15.3
Fort Hays, KS	38.86	-99.3	-15.6
Fall Leaf Prairie, KS	39	-95.2	-18.3
Konza Prairie, KS	39.09	-96.6	-14.4
Squaw Creek Wildlife Refuge, MO	40.08	-95.4	-16.8
Indian Cave State Park, NE	40.26	-95.6	-16.0
CO State/LTER, CO	40.84	-104.7	-15.9
Nine Mile Prairie, NE	40.87	-96.8	-15.5
Loess Hills Wildlife Refuge, IA	42.05	-96.1	-15.7
Stone State Park, IA	42.52	-96.5	-14.0
Niobrara Nature Preserve, NE	42.77	-100	-17.8
Second Niobrara Site	42.77	-100	-18.4
Newton Hills State Park, SD	43.26	-96.6	-18.3
Lange-Furgeson Site, SD	43.33	-102.6	-18.3
Cayler Prairie, IA	43.4	-95.2	-17.7
Makoce Washte, SD	43.55	-97	-16.3
Lundblad, MN	43.94	-95.7	-18.7

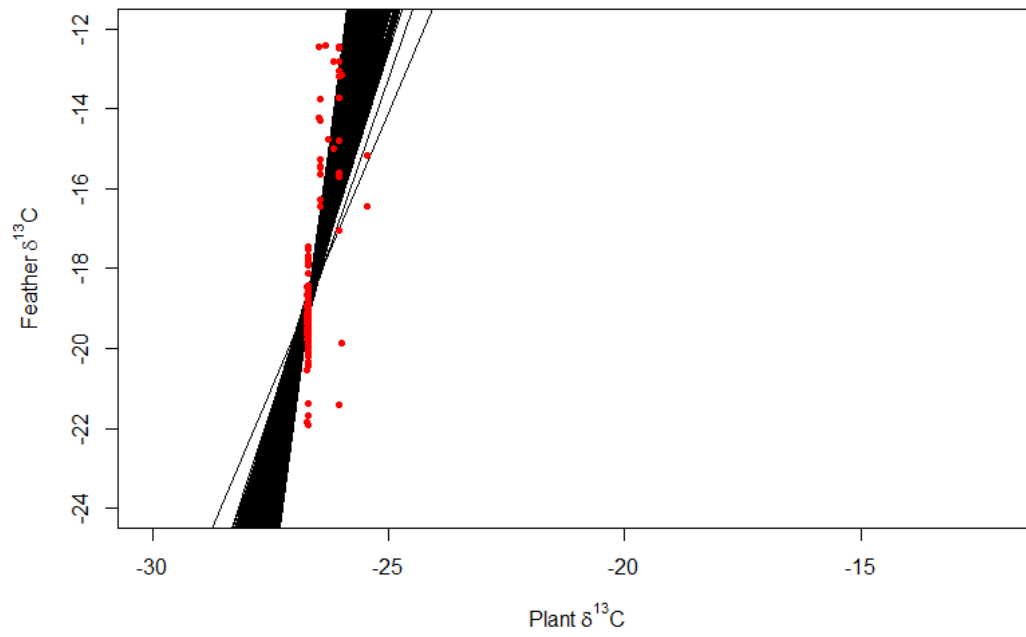
Cottonwood, SD	43.96	-101.9	-18.1
Schaefer Prairie, MN	44.72	-94.3	-19.8
Antelope Prairie, SD	45.51	-103.3	-20.4
Custer Battlefield, MT	45.54	-107.4	-25.0
Ordway Prairie, SD	45.72	-99.1	-19.0
Staffanson, MN	45.82	-95.8	-17.6
Eastern ND Tallgrass Prairie, ND	46.42	-97.5	-18.2
Bluestem Prairie, MN	46.84	-96.5	-19.5
Dickinson, ND	46.89	-102.8	-18.9
Sheyenne Grassland, ND	46.5	-97.5	-21.1
Western ND Mixed Prairie, ND	47	-103.5	-20.1
Oakville, ND	47.2	-97.3	-20.5
Cross Ranch, ND	47.25	-101	-19.7
Teddy Roosevelt N.P., ND	47.45	-103.2	-21.9
Pembina Prairie, MN	47.69	-96.4	-17.9
Glasgow, MT	48.12	-106.4	-20.3
Bainville, MT	48.14	-104.2	-20.5
Milk River, Alberta	49.08	-112.1	-23.4
Tolstoi Prairie, Manitoba	49.08	-96.8	-21.0
Living Prairie, Manitoba	49.88	-97.3	-21.4
Head Smashed In, Alberta	49.5	-113.8	-24.1
Grosse Isle, Manitoba	50.07	-97.5	-20.6
Oak Hammock, Manitoba	50.2	-97.2	-19.1
Stavely, Alberta	50.22	-113.9	-25.2
Matador, Saskatchewan	50.67	-109.3	-24.1
Biddulph, Saskatchewan	50.68	-107.7	-22.9
Kernan Prairie, Saskatchewan	51.9	-106.7	-25.1

APPENDIX 2: Rescaling Equation and Calibration Data Plots

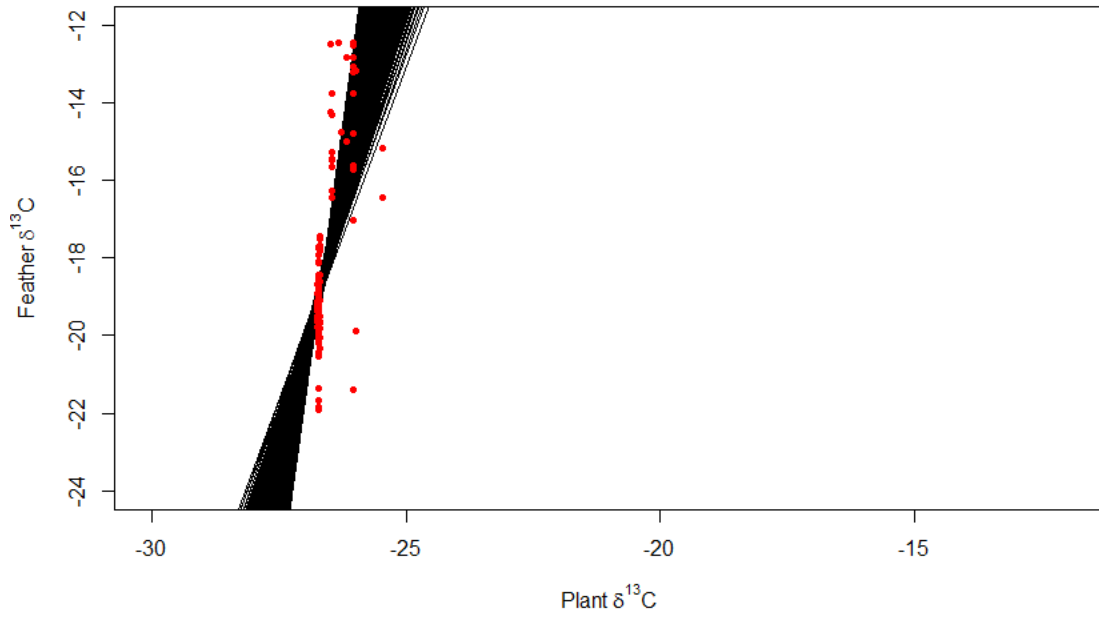
MARCH



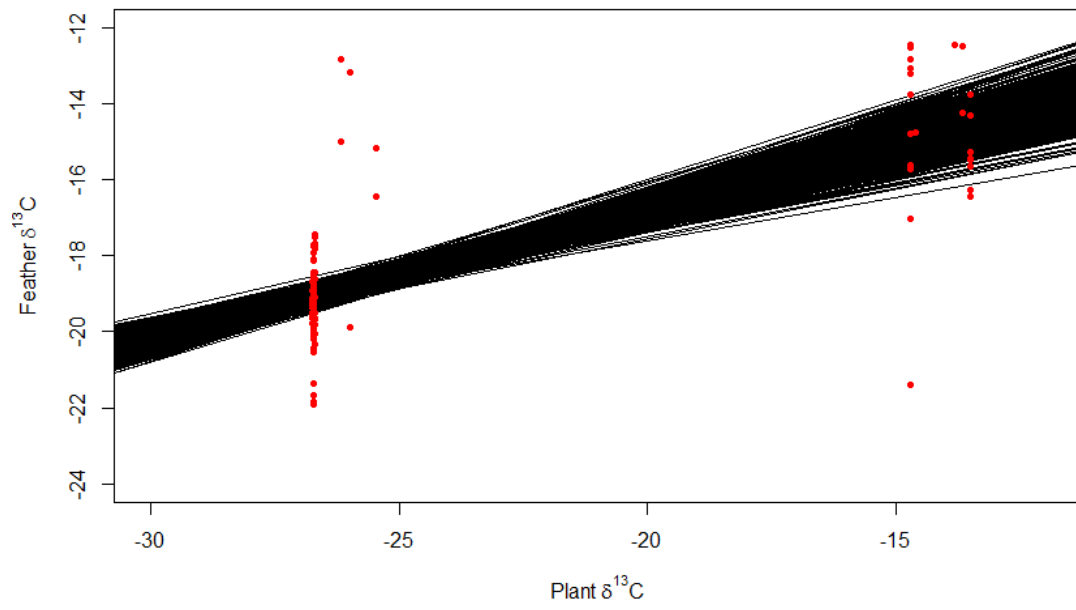
APRIL



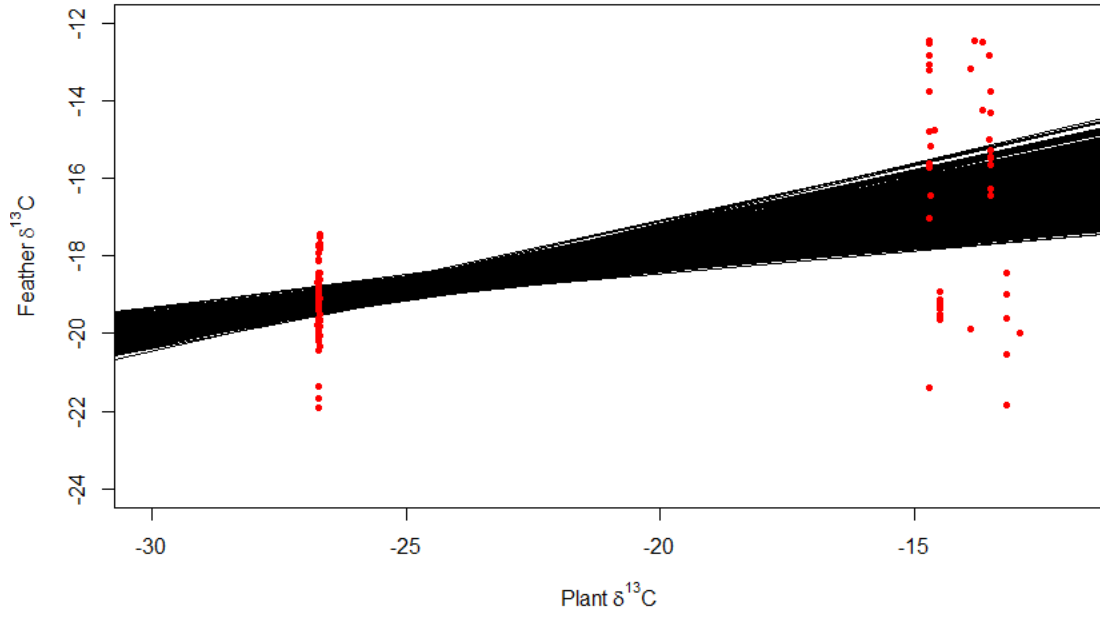
MAY



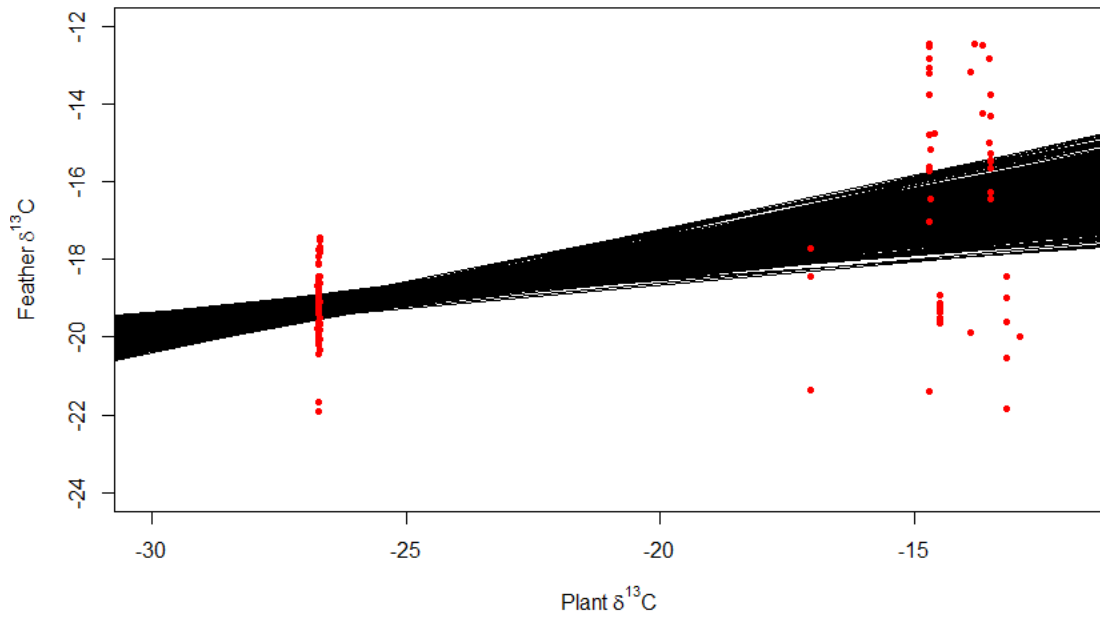
JUNE



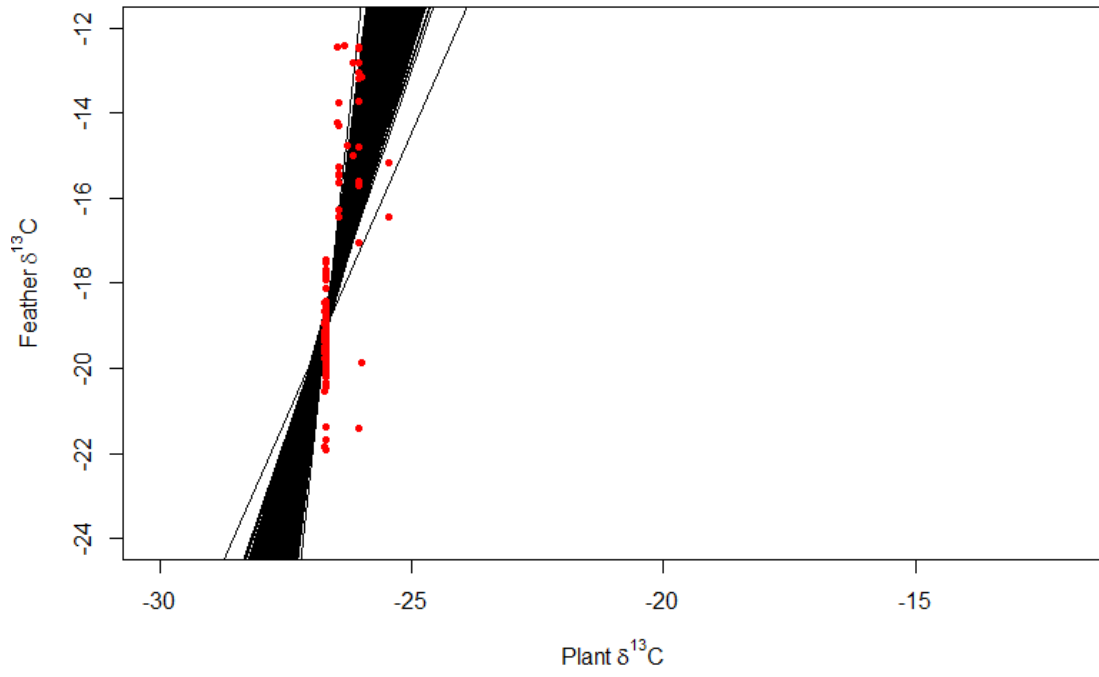
JULY



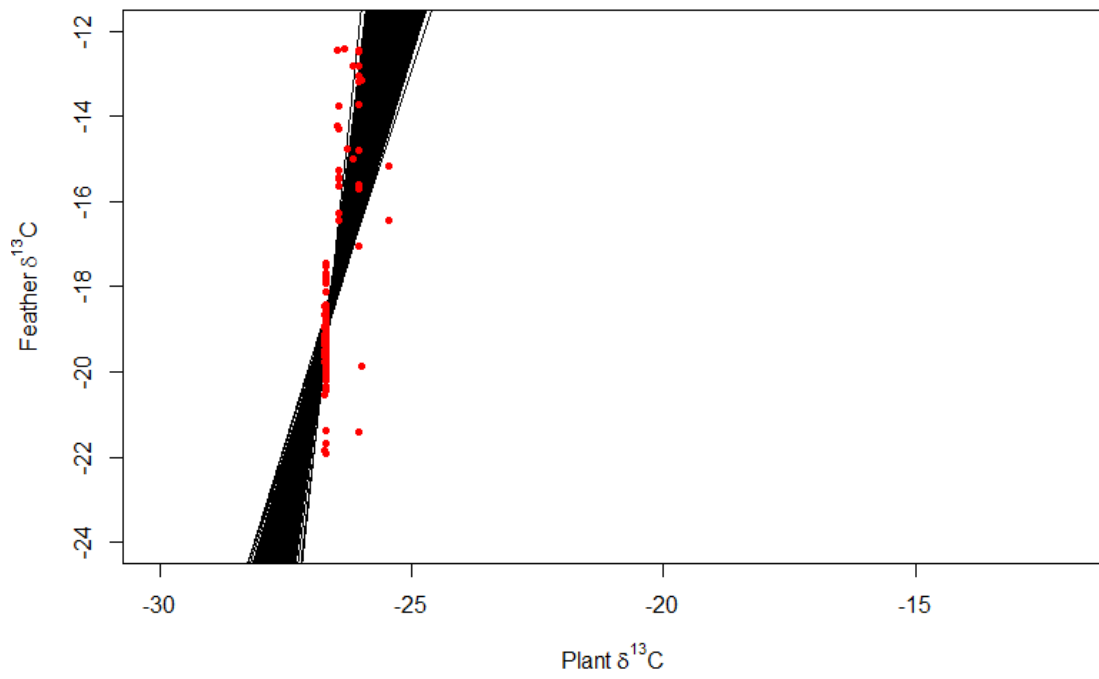
AUGUST



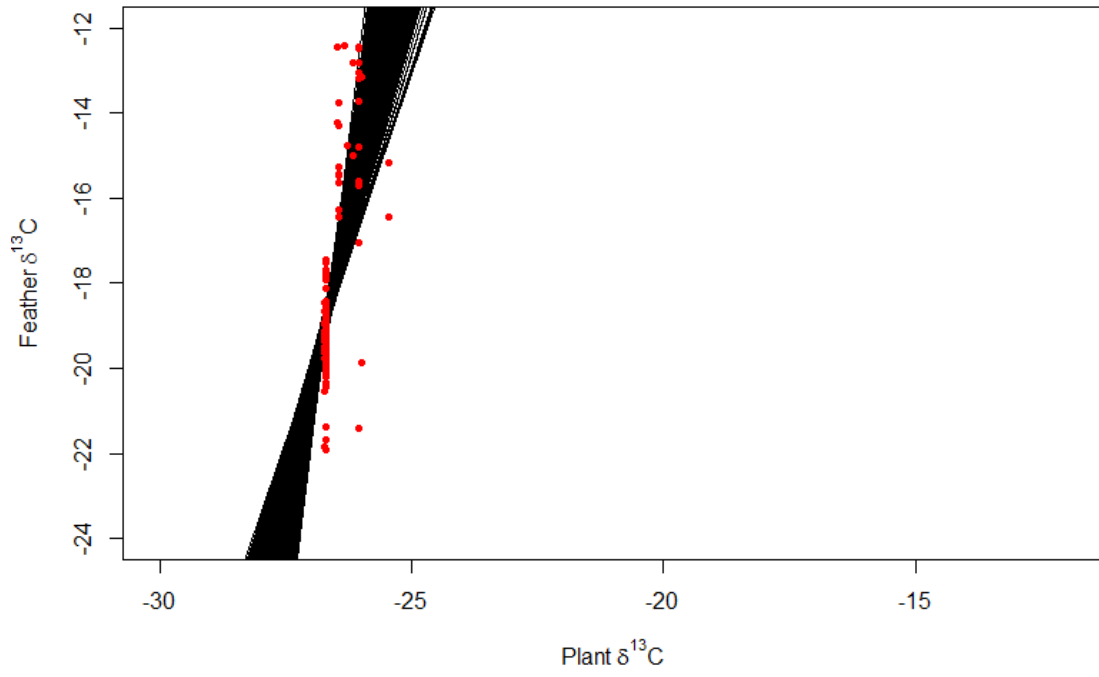
SEPTEMBER



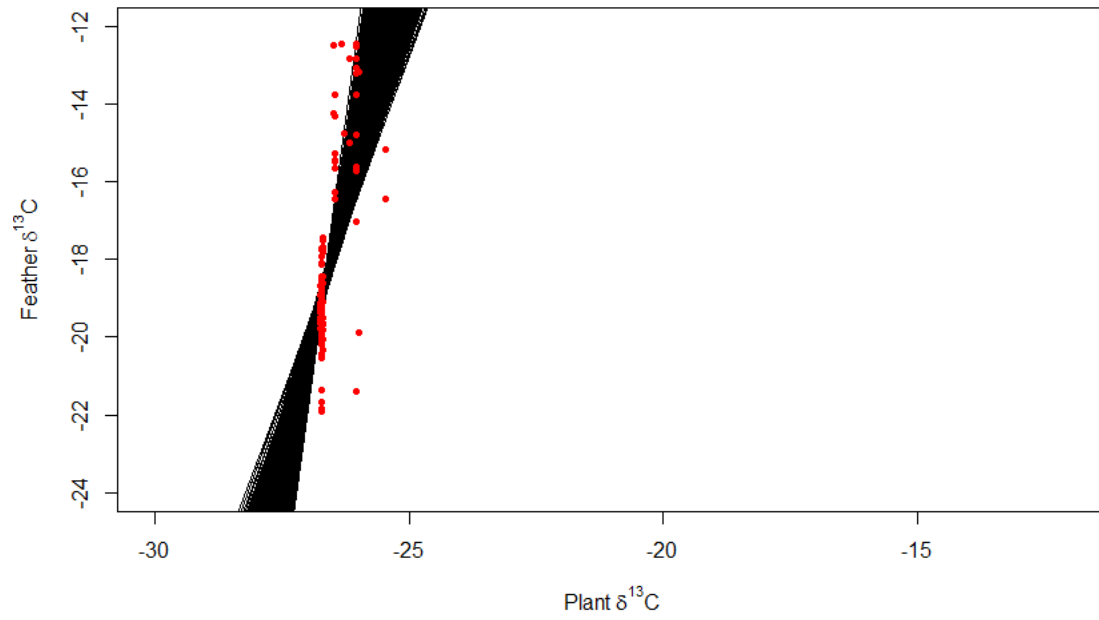
OCTOBER



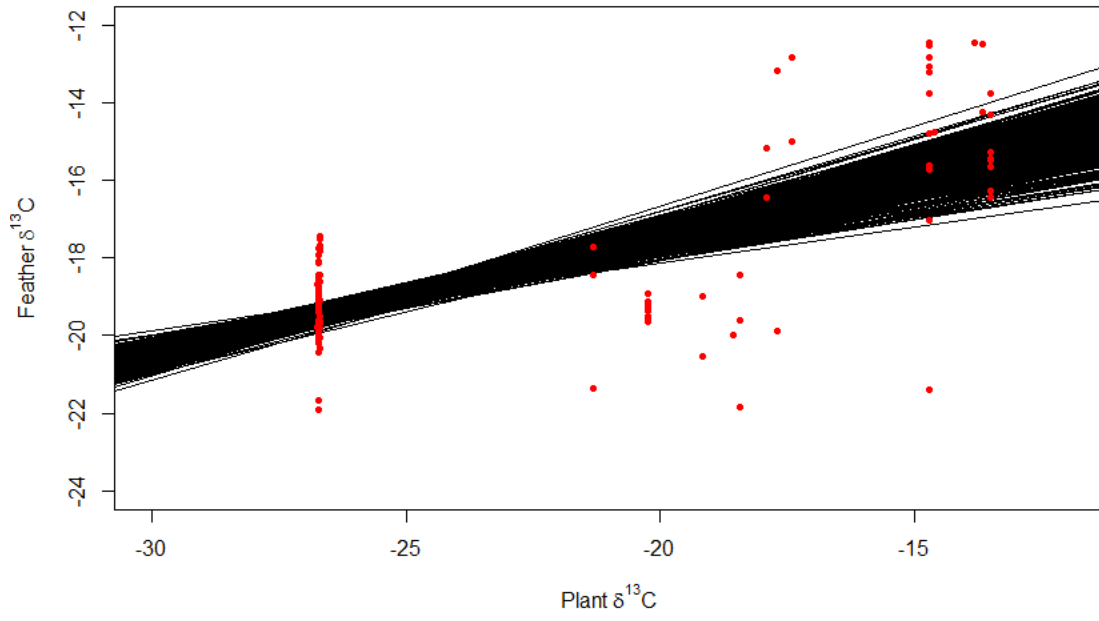
NOVEMBER



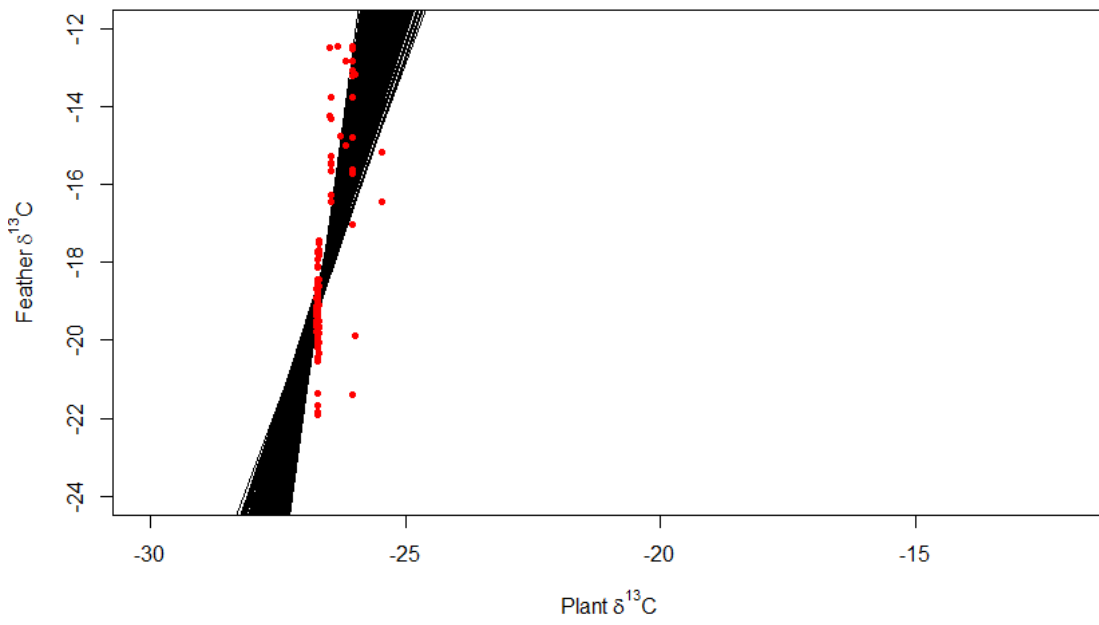
EARLY SEASON

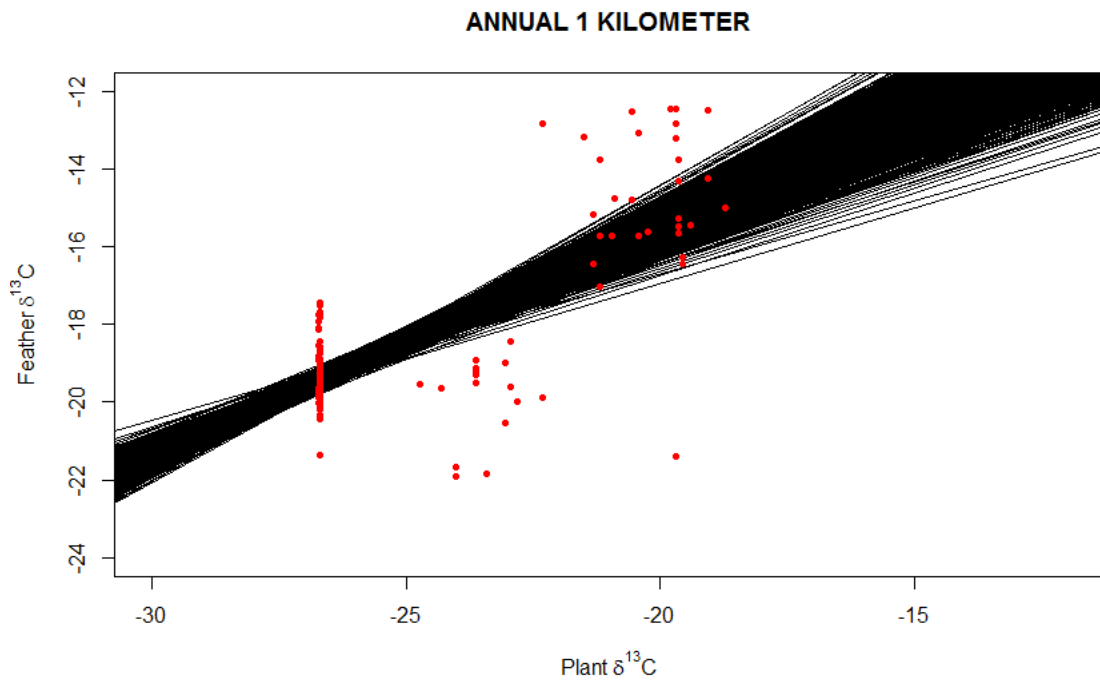
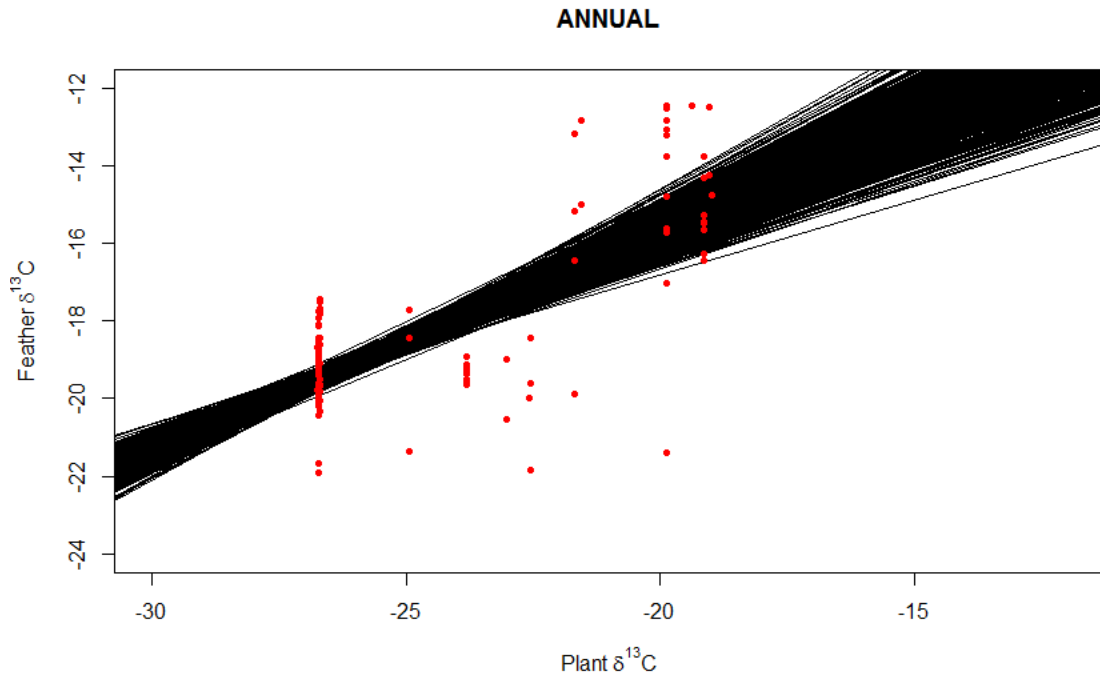


MID-SEASON

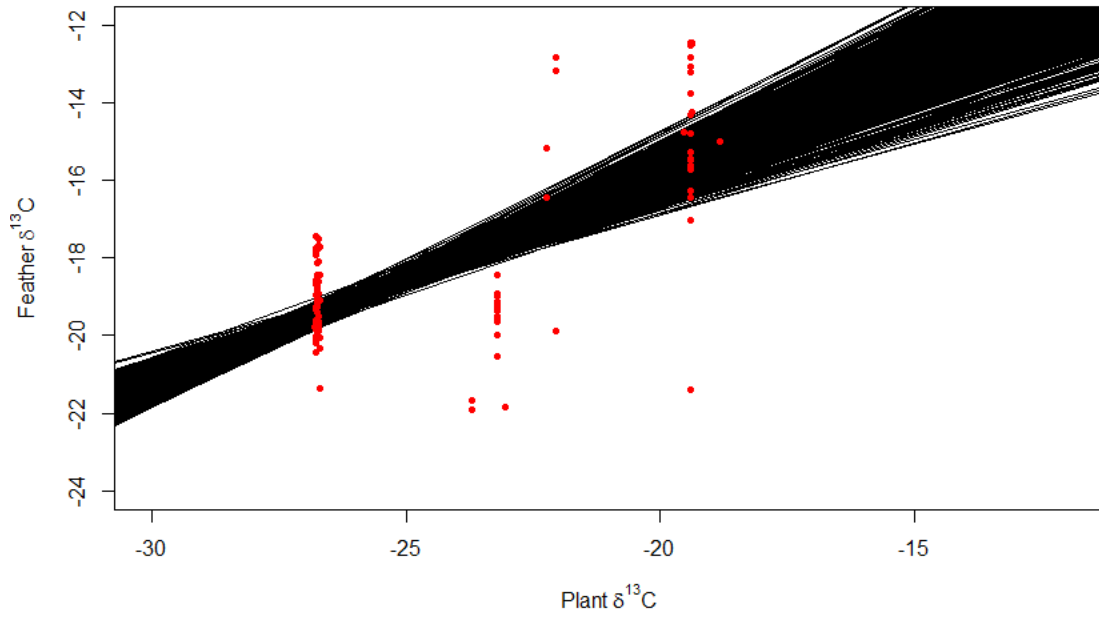


LATE SEASON





ANNUAL 10MIN



APPENDIX 3: Geographic Origin Assignment Code

```
##necessary libraries

library(raster)
library(rgdal)
library(maps)
library(mapdata)
library(fields)
library(spam)
library(pracma)
library(usdm)
library(SDMTools)
library(MASS)

##----- RESCALING FUNCTION -----##
##The function conducts 1000 random samples of data with replacement and generates a
regression line for each
##The output contains the slopes and intercepts of each of the 1000 regression lines, as
well as the standard deviation of equation residuals
##Input is a .csv table that has the calibration data tissue d13C, and the corresponding
plant mean and SD extracted from the vegetation isoscape

##Function requires:
  #table = the filename (with directory, if applicable) from which to load the data
  #tissue.mean = column # containing d13C tissue values of individuals sampled
  #plant.mean = column # containing d13C plant values
  #plant.SD = column # containing of d13C plant values

rescale <- function(table.ex, tissue.mean, plant.mean, plant.SD) {
##create empty vectors to save values to, change length value to number of iterations.
  slopes <- vector('numeric', length=1000)
  intercepts <- vector('numeric', length=1000)
  sds <- vector('numeric', length=1000)

##bootstrapping function, repeats resampling 1000 times, creating a rescaling equation
with a slope, intercept and residuals
  for (k in 1:1000){
    table <- read.table(table.ex, header=TRUE,
                        sep=",", na.strings="NA")
    boot <- sample(seq_len(nrow(table)), nrow(table), replace = TRUE)
    calibration <- table[boot,]
```

```

tissue.d13c <- vector('numeric', length(calibration)) ##creates vector of bootstapped-
sample feather d13C
tissue.d13c <- calibration[, tissue.mean]
plant.d13c <- vector('numeric', length(calibration)) ##creates vector of bootstapped-
sample extracted (from isoscape) plant d13C
plant.d13c <- calibration[, plant.mean]

lmResult.k <- lm(tissue.d13c~plant.d13c) ##linear regression model
intercepts[k] <- coef(lmResult.k)[1] ##saves intercepts to vectors
slopes[k] <- coef(lmResult.k)[2] ##saves slopes to vector
sds[k] <- sd(lmResult.k$residuals) ##standard deviation of residuals

}
##returns data frame containing 1000 slopes, intercepts and standard deviations of
residuals
return(data.frame(slopes, intercepts, sds))
}

##----- RASTER CONVERSION -----##
##Conversion function uses the output from the rescaling function to rescale plant raster
and rescale SD raster (used in the pooled error)

##Function requires:
#original.raster = the filename (and directory) of the original vegetation raster
#reg.par = the output from the function above
#scratch.dir = the directory of a scratch folder to store the rasters temporarily

raster.conversion <- function (original.raster, reg.par, scratch.dir) {

for (i in 1:length(reg.par[,1])) {
reg.par.i <- reg.par[i,]
raster.i <- original.raster*reg.par.i$slopes + reg.par.i$intercepts ##creates
bootstrapped (1000) rescaled rasters
name <- paste(scratch.dir, i, ".grd", sep="")
writeRaster(raster.i, filename=name, overwrite=TRUE)
}
setwd(scratch.dir)
all.files <- dir(pattern=".grd")
n <- length(all.files)
all.rasters <- stack(all.files) ##stacks all rescaled rasters from folder

```



```

mean.raster <- stackApply(all.rasters, fun=mean, indices=c(rep(1,n))) ##calculates the
mean of rescaled rasters, pixel by pixel
var.raster <- stackApply(all.rasters, fun=var, indices=c(rep(1,n))) ##calculates the per
pixel variance and creates a variance raster
return(list(mean.raster=mean.raster, var.raster=var.raster)) ##returns mean raster and
variance raster
}

```

```
##----- ASSIGNMENT FUNCTION -----##
```

```
##Assignment function uses the likelihood term to determine the probability that an
individual sample was from a particular geographic location and writes an ascii file to a
chosen directory
```

```
##Function requires:
```

```

#rescaled_raster = the tissue-specific d13C raster created in the conversion
function (output or raster.conversion)
#rescaled_SD_raster = the variance raster created in the conversion function.
This the component of the error term related to the rescaling process.(output of
raster.conversion)
#precip_SD_raster = this is the SD raster associated with the original vegetation
isoscape. This is the plant component of the variance term.
#SDS = the individual component of the variance term. Uses the sds from rescale
function= gamma mean mean=alpha*beta featherraster[[3]] (output of rescale)
#assign_table = this a csv filename (and directory, if applicable) containing the
tissue d2H values of the individuals for which the assignments will be made
#d13Ctissue = column number in the assign_table with the d13c tissue values
#ID = column number with individual identifiers
#save_dir is where the output assignments should be saved as an ascii, but could
be changed

```

```

assignment <- function(rescaled_raster, rescaled_var_raster, plant_SD_raster, SDS,
assign_table, d13Ctissue, ID, save_dir){
  out <- fitdistr((SDS)^2, "gamma", lower=0) ##calculates gamma fit distribution of the
standard deviation of the residuals
  SD_indv <- out$estimate[1]*out$estimate[2] ##calculates average of fitted gamma
distribution to account for average variance among individuals "at a site"
  error <- sqrt((rescaled_var_raster) + (plant_SD_raster)^2 + SD_indv) ##sum of the
error results, includes variance raster, mean gamma of sds, and plant isoscape standard
deviation
  data <- read.table(assign_table, sep=",", header=T) ##read in table of feather d13C
values
  data <- data[1:118,] ##change length value to number of samples being assigned a
geographic origin
  n <- length(data[,d13Ctissue])

```

```

for (i in 1:n){
  indiv.data <- data[i,] ##feather isotope value
  indiv.id <- indiv.data[1, ID] ##feather ID
  assign <- (1/(2*pi*error^2))*exp(-1*(indv.data[1, d13Ctissue]-
rescaled_raster)^2/(2*error^2)) ##assignment function accounting for error term
  assign_norm <- assign/cellStats(assign, "sum") ##normalize so all pixels sum to 1
  filename <- paste(save_dir, indiv.id, ".like", ".asc", sep="")
  writeRaster(assign_norm, file=filename, format="ascii", overwrite=TRUE)
}
}

```