



**The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:**

**Document Title:** Identification of Deceased Border Crossers:  
Investigating Spatial and Skeletal  
Attributes of Migrant Deaths

**Author(s):** Caitlin Claire Madigan Vogelsberg

**Document Number:** 306325

**Date Received:** April 2023

**Award Number:** 2017-IJ-CX-0002

**This resource has not been published by the U.S. Department of Justice. This resource is being made publicly available through the Office of Justice Programs' National Criminal Justice Reference Service.**

**Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.**

IDENTIFICATION OF DECEASED BORDER CROSSERS: INVESTIGATING SPATIAL  
AND SKELETAL ATTRIBUTES OF MIGRANT DEATHS

By

Caitlin Claire Madigan Vogelsberg

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Anthropology – Doctor of Philosophy

2018

## ABSTRACT

### IDENTIFICATION OF DECEASED BORDER CROSSERS: INVESTIGATING SPATIAL AND SKELETAL ATTRIBUTES OF MIGRANT DEATHS

By

Caitlin Claire Madigan Vogelsberg

International migration research has primarily focused on cultural, sociological, and economic components. Understanding geographic mortality patterns and the skeletal attributes among the deceased using a mixed-model framework can provide insight into route selection for irregular entrants via the Mexican border. This research investigates the spatial and skeletal properties of deceased undocumented border crossers (UBCs) recovered in the southern Arizona desert and the relationship between recovery location and country of origin.

Previous research investigating spatial patterns in the distribution of identified UBCs recovered in the jurisdiction of the Pima County Office of the Medical Examiner (PCOME) in Tucson, Arizona demonstrated positive spatial autocorrelation between individualizing attributes (such as biological sex and country of origin) and their recovery location (Vogelsberg 2018). Those results indicated several influencing factors, such as country of origin, on final recovery location. This research project combines spatial patterns and cranial skeletal indicators of geographic origin to improve country of origin prediction during the identification process.

An optimized global linear model developed using craniometric and macromorphoscopic factors for a sample ( $n = 25$ ) of identified Mexican and Guatemalan individuals analyzed at the PCOME was incorporated into several geographically weighted regression (GWR) platforms to predict country of origin. The best performing GWR analysis accounted for just over half of the variation in the data ( $R^2 = 0.540$ ). This is an increase from the global model ( $R^2 = 0.432$ ) which did not incorporate recovery location and attributes of other individuals found nearby. Other

indicators of model goodness-of-fit show more accurate country of origin predictions using the GWR method.

Model testing on individuals with presumptive Mexican identifications (n = 8) resulted in the correct allocation for country of origin for two individuals and provided promising results for future application. Although sample sizes were small, the potential for applying mixed-model methods is clearly demonstrated. As more individuals are identified and added to the model reference sample, the utility of this predictive method will improve. Furthermore, the application of these techniques to situations in which the physical location of an individual might correlate with their personal attributes is demonstrated.

This research provides the forensic and humanitarian community with supplemental information to aid in the investigation of undocumented border crossers recovered from the southern Arizona desert. Enhancements to the identification process by better directing missing persons searches may increase the number of identified individuals and the return of their remains to their families.

Copyright by  
CAITLIN CLAIRE MADIGAN VOGELSBERG  
2018

To my grandparents; whose cheers of “Don’t Settle!” still motivate me.

## ACKNOWLEDGEMENTS

## TABLE OF CONTENTS

LIST OF TABLES .....	x
LIST OF FIGURES .....	xiii
CHAPTER 1. INTRODUCTION .....	1
Project Goals .....	4
Research Questions .....	4
Biological Data Analysis Questions .....	4
Geospatial Data Analysis Questions .....	5
Combined Data Analysis Research Questions .....	5
Organization of Chapters .....	6
REFERENCES .....	8
CHAPTER 2. HISTORICAL RELATIONS BETWEEN THE UNITED STATES AND MEXICO .....	11
Introduction .....	11
Creation of the Borderlands .....	11
Border Concept .....	12
European Occupation .....	13
The Spanish Frontier .....	14
The Mexican Frontier .....	15
American Developments .....	16
United States and Mexico Immigration and Policy Relations .....	18
Mexican Politico-economic and Immigration Trends .....	18
First Wave .....	18
Second Wave .....	20
Third Wave .....	21
Current Political-Economic Developments .....	22
Immigration Policy Trends .....	24
First Phase .....	25
Second Phase .....	26
Third Phase .....	27
Illegalization of Immigration .....	28
Militarization of the Border .....	29
Conclusions .....	30
APPENDIX .....	32
REFERENCES .....	34
CHAPTER 3. MIGRATION THEORIES IN THE US-MEXICO CONTEXT .....	39
Introduction .....	39
Immigration Initiation .....	39
“Push-Pull” .....	40
World Systems Theory .....	41



New Economics of Migration Theory .....	42
Immigration Perpetuation Theories .....	43
Dual-market Labor System .....	43
Cumulative Causation and Migrant Networks.....	44
Conclusions.....	46
REFERENCES .....	48
CHAPTER 4. BIODISTANCE BACKGROUND.....	51
Introduction.....	51
Utility and Theory.....	52
Theories.....	53
Nonmetric Approaches .....	55
Metric Approaches.....	59
Contemporary Research.....	61
Regional Investigations.....	63
Conclusions.....	66
REFERENCES .....	67
CHAPTER 5. METHODS & MATERIALS .....	73
Methods Introduction.....	73
Biological Methods.....	74
Cranio-metric Data.....	74
Macromorphoscopic Data.....	79
Random Forest Models .....	81
Geospatial Methods .....	82
Exploratory spatial data analysis (ESDA) .....	83
Geographically Weighted Regression.....	86
Assumptions and Limitations .....	87
Materials Introduction.....	88
Biological Data .....	88
Cranio-metric Dataset .....	89
Macromorphoscopic Dataset .....	90
Random Forest Model Dataset.....	91
Geospatial Data.....	92
Geobiological Data .....	98
Conclusions.....	100
REFERENCES .....	101
CHAPTER 6. BIOLOGICAL DATA RESULTS.....	105
Introduction.....	105
Cranio-metric Data Analysis.....	105
Macromorphoscopic Data Analysis.....	112
Random Forest Models .....	114
Conclusions.....	121
REFERENCES .....	122

CHAPTER 7. GEOSPATIAL RESULTS .....	124
Introduction.....	124
Exploratory Spatial Data Analysis (ESDA).....	124
Identified Individuals .....	131
Join Count .....	136
Geographically Weighted Regression Analysis .....	140
General Linear Models .....	140
OLS Model Choice .....	142
Geographically Weighted Regression Analyses.....	143
Model 4.2 .....	144
spgwr.....	145
GWmodel.....	150
Model 4.3 .....	153
spgwr.....	154
GWmodel.....	157
ArcMap .....	160
Discrepancy Investigation.....	164
Model Testing.....	166
Conclusions.....	169
REFERENCES .....	170
CHAPTER 8. DISCUSSION AND CONCLUSIONS .....	172
Introduction.....	172
Biological Data .....	173
Cranio-metric And Macromorphoscopic Data Analysis .....	173
Random Forest Modeling Analyses.....	173
Future Biological Studies.....	174
Geographic Data Analysis .....	175
Exploratory Spatial Data Analysis (ESDA).....	175
Year of Recovery .....	175
Identification Status .....	178
Sex and Country of Origin.....	179
Future Geographic Studies.....	180
Geobiological Model .....	180
Concluding Remarks.....	182
REFERENCES .....	184

## LIST OF TABLES

Table 5.1. Howells cranial measurements and abbreviations (Spradley 2013).....	75
Table 5.2. Cranial measurements used in initial principal components analyses.....	77
Table 5.3. Group properties of craniometrics principal components analyses.....	78
Table 5.4. Measurements used in final principal components analyses.....	78
Table 5.5. Macromorphoscopic trait names and abbreviations.....	80
Table 5.6. Population group distribution of craniometric dataset.....	89
Table 5.7. Country of origin distribution of craniometric dataset.....	90
Table 5.8. Ancestral group distribution of the macromorphoscopic dataset.....	91
Table 5.9. Geo-origin distribution of the macromorphoscopic dataset by data source.....	91
Table 5.10. Distribution of dataset used in random forest modeling.....	92
Table 5.11. Distribution of identified UBC remains by sex and country of origin.....	97
Table 5.12. Demographic distribution of identified individuals in ESDA.....	98
Table 5.13. Demographic information of identified individuals with geo- biological data.....	99
Table 6.1. Measurement variables used in PCAs.....	106
Table 6.2. Factor scores from PCA results.....	107
Table 6.3. Variable loadings of first 3 factors.....	108
Table 6.4. First 2 factor loadings for Hispanic samples.....	110
Table 6.5. Classification rates of the three-way CAP analysis.....	112
Table 6.6. Out-of-box error rates for RFMs run for ancestry and country of origin.....	115
Table 6.7. Confusion matrix for RFM for ancestry prediction.....	116
Table 6.8. Confusion matrix for RFM for country of origin prediction.....	117

Table 7.1. Demographic distribution of identified individuals used in ESDA methods .....	127
Table 7.2. Global Moran's I statistics for all individuals.....	128
Table 7.3. Number of identified individuals by year.....	130
Table 7.4. LISA cluster map and Moran's scatter plot for identification status variable for all individuals.....	132
Table 7.5. Global and local Moran's I values obtained in ESDA.....	139
Table 7.6. Country of origin and sex of individuals used in GWR analysis .....	140
Table 7.7. Summary results of global linear models .....	141
Table 7.8. AICc indexes of global linear models .....	141
Table 7.9. Diagnostics for Model 4.2 using spgwr.....	145
Table 7.10. Prediction results for Model 4.2 using spgwr.....	149
Table 7.11. Diagnostics for Model 4.2 using GWmodel.....	150
Table 7.12. Coefficient ranges calculated using GWmodel and spgwr on Model 4.2 .....	150
Table 7.13. Prediction results for Model 4.2 using GWmodel.....	152
Table 7.14. GWR accuracy comparisons for Model 4.2 .....	153
Table 7.15. Diagnostics for Model 4.3 using spgwr.....	154
Table 7.16. Prediction results for Model 4.3 using spgwr.....	156
Table 7.17. Diagnostics for Model 4.3 using GWmodel.....	157
Table 7.18. Prediction results for Model 4.3 using GWmodel.....	159
Table 7.19. Coefficient ranges calculated using GWmodel and spgwr on Model 4.3 .....	160
Table 7.20. GWR accuracy comparisons for Model 4.3 .....	160
Table 7.21. Diagnostics for Model 4.3 with 50km hexagonal grid .....	161
Table 7.22. Prediction results for Model 4.3 using ArcMap 50km hexagonal grid .....	163
Table 7.23. GWR accuracy comparisons for Model 4.3 using ArcMap hexagonal grid .....	163

Table 7.24. Prediction results for test samples using spgwr.....	168
Table 7.25. Prediction accuracy comparison for test model and Model 4.3 .....	168

## LIST OF FIGURES

Figure 5.1. Flow-chart of project materials and methods. ....	73
Figure 5.2. VIMGUI visualization of missing craniometric data, sorted from most to least. ....	76
Figure 5.3. Corridors designations used in geospatial analyses.....	83
Figure 5.4. Exact and approximate locations of all recovered UBC remains from 2001 January 2017 (N = 2,616). ....	94
Figure 5.5. Locations of all recovered UBCs with accurate data used in project (N = 1,723). ....	95
Figure 5.6. Locations of UBC remains with accurate data coded by identification status (ID'd = 1,051) and (UID = 680). ....	95
Figure 5.7. Number of recoveries per year for all recovered remains (N = 2,616). ....	96
Figure 5.8. Identified individuals with geobiological data available.....	99
Figure 6.1. Selected measurements for final PCA methods. ....	106
Figure 6.2. Scree plot for 12 variable PCA run on standardized data. ....	107
Figure 6.3. Group centroids of first 2 factors pooled by ancestry. ....	109
Figure 6.4. Group centroids of first 2 factors pooled by country of origin.....	111
Figure 6.5. Ordination plots using (a) Mahalanobis (m=10) and (b) alternate Gower (m = 10) distance measures. ....	113
Figure 6.6. The ancestry out-of-bag error estimate, across 500 trees. ....	116
Figure 6.7. The country of origin out-of-bag error estimate, across 500 trees. ....	117
Figure 6.8. Importance plot of variables used in ancestry prediction RFM. ....	119
Figure 6.9. Importance plot of variables used in country of origin prediction RFM.....	120
Figure 7.1. Number of recoveries per year for all recovered remains (N = 2,616). ....	125
Figure 7.2. Number of recoveries per year for identified individuals with accurate GPS data (n = 1,051).....	125
Figure 7.3. Global Moran's I plot for year of recovery, using kNN (k = 41). ....	128

Figure 7.4. LISA Cluster map and Moran's scatter plot for year of recovery variable for all individuals. ....	129
Figure 7.5. LISA cluster map and Moran's scatter plot for identification status variable for all individuals. ....	131
Figure 7.6. LISA cluster map and Moran's scatter plot for year of recovery of identified individuals. ....	132
Figure 7.7. LISA cluster map and Moran's scatter plot for sex of identified individuals. ....	133
Figure 7.8. LISA cluster map and Moran's scatter plot for country of origin of identified individuals. ....	134
Figure 7.9. LISA cluster map and Moran's scatter plot for country of origin of individuals not from Mexico. ....	135
Figure 7.10. LISA cluster map and Moran's scatter plot for sex of individuals not from Mexico. ....	135
Figure 7.11. Grid cell transformation. ....	136
Figure 7.12. LISA cluster map and Moran's scatter plot for identification status variable for all individuals using grid transformation. ....	137
Figure 7.13. LISA cluster map and Moran's scatter plot for sex for identified individuals using grid transformation. ....	138
Figure 7.14. LISA cluster map and Moran's scatter plot for country of origin for identified individuals using grid transformation. ....	138
Figure 7.15. Data Set 2 (n = 22) labeled by country of origin. ....	144
Figure 7.16. Data Set 2 labeled by sex. ....	144
Figure 7.17. Coefficient values for Model 4.2 variables. ....	147
Figure 7.18. Model 4.2 predicted (a) and actual country of origin values (b) using spgwr. ....	148
Figure 7.19. Model 4.2 predicted (a) and actual country of origin values (b) using GWmodel. ....	151
Figure 7.20. Data Set 3 (n = 25) labeled by country of origin. ....	154
Figure 7.21. Data Set 3 labeled by sex. ....	154
Figure 7.22. Model 4.3 predicted (a) and actual country of origin values (b) using spgwr. ....	155

Figure 7.23. Model 4.3 predicted (a) and actual country of origin values (b) using GWmodel. 158

Figure 7.24. Model 4.3 predicted (a) and actual country of origin values (b) using ArcMap 50km hexagonal grid. .... 162

Figure 7.25. Plot of first 2 factors labeled by country of origin. .... 165

Figure 7.26. Individual # 21 PCA and PCOa scores compared to all country means. .... 165

Figure 7.27. Geographic position of individual # 21 (star)..... 166

Figure 7.28. Predicted country of origin values of test dataset (a) and actual countries (b) using spgwr. .... 167

Figure 8.1. LISA Cluster map and Moran's scatter plot for year of recovery variable for all individuals. .... 176

Figure 8.2. LISA cluster map and Moran's scatter plot for identification status variable for all individuals. .... 178



## CHAPTER 1. INTRODUCTION

Undocumented immigration into the United States continues to be at the forefront of political, economic, and social arenas at both the state and national levels. These discussions often involve the unauthorized entry of international migrants into the United States via the southern border, with debates ranging from worries over American job security to humanitarian groups raising awareness of the plight of undocumented border crossers (UBCs) during and after their journey (Kelly 2017; Martínez et al. 2013; Reineke & Martínez 2014). Yet, this region is characterized by a history of constant migration.

Although there are several ways for a person to reside irregularly in the United States, such as overstaying a work or tourist visa, a steady number of people continue to clandestinely cross the Mexican border, and sometimes more than once in their lifetime (Massey 1987; Jones-Correa & de Graauw 2013). The number of undocumented migrants crossing into the US, and subsequently dying, has dominated the caseload of forensic anthropologists working in the region over the years. As a response to this continuing problem, the Pima County Office of the Medical Examiner (PCOME) in Tucson, Arizona, began keeping official records on the number of recovered suspected UBC remains in 2001, even though border crossing deaths were already part of the regular casework. Averaging over 170 UBC cases a year, the PCOME investigates the deaths of all unknown individuals from 11 of the 15 Arizona counties, including the four that border northern Mexico. The total geographic area of the PCOME's jurisdiction is ca. 27,000mi<sup>2</sup>, with deceased migrants found primarily within the Tucson Sector— one of 20 regional divisions created by the United States Border Patrol (USBP). Most cases are concentrated in Pima and Santa Cruz counties, although recoveries from as far north as La Paz have occurred. Current forensic anthropology UBC cases total over 2,600, with ~36% still unidentified (PCOME 2017).

In line with the Border Patrol apprehension rates, the PCOME has seen a remarkable increase in the number of undocumented border crosser deaths over the last 30 years. Regarding the Tucson Sector, Anderson (2008) reports that the number of cases averaged 19 per year prior to 1998, steadily climbing to 75 in 2001, and doubling by 2002. In fact, 2010 saw the highest rate of recovered remains (n = 222) (PCOME 2017). While apprehension rates have steadily declined over the last 12 fiscal years, the number of recovered individuals remains steady and high (CBP 2015; Colibrí 2018). The dramatic increase in the number of deaths along the border in both Arizona and the Greater Southwest is in part due to the dangerous routes many irregular entrants are forced to take (Rubio-Goldsmith et al. 2006). These paths are generally established via socially and culturally-driven processes such as *coyotes* (paid travel guides) or prior travel route knowledge (Massey 1988; Massey & Espinosa 1997; Singer & Massey 1998; Orrenius 1999; Tellez 2007). The deceased are often ill-equipped or otherwise unprepared for the journey through harsh terrain and perish, and their remains are later recovered by local or federal law enforcement for identification.

The Colibrí Center for Human Rights is a nonprofit organization working with the PCOME and other organizations to assist the families of missing migrants by taking missing persons reports, collecting family reference samples for DNA identification, and providing information to the public about the current humanitarian crisis. Colibrí currently has over 3,000 missing persons reports in their database. This means that the number of missing individuals far exceeds the number yet to be identified in Arizona (Colibrí Center 2018).

Efforts are ongoing for the PCOME and Colibrí to identify all individuals and return them to their families. This process involves researching the demographics of the deceased (age, sex, etc.) through the analysis of recovered remains and any associated cultural effects

(Anderson 2008). To date, these demographic and cultural studies dominate the literature concerning migrant deaths in southern Arizona. Studies on the analysis of the geographic location of these deaths in southern Arizona have been conducted but have been limited to exploring temporal or topographical trends (Chamblee 2006; Lawrence & Wildgen 2012; Giordano & Spradley 2017). This is in part due to the relatively recent availability of geographic data to researchers outside of Arizona law enforcement and the PCOME through the online open geographic information systems (OGIS) provided at [www.humaneborders.info](http://www.humaneborders.info). This type of centralized database for information on migrant deaths is unique to the PCOME and southern Arizona and should be implemented by other states and jurisdictions handling these types of cases. A directed analysis focusing on the location and distribution of international migrant deaths within using this southern Arizona data is warranted. If successful, this research should encourage other jurisdictions undergoing a similar crisis to collect this type of geographic and demographic information for more applicable research involving patterns specific to their own region.

The purpose of this research is to examine the biological and geospatial properties of human remains recovered in southern Arizona by investigating how individuals are represented over this wide geographic area. Previous research has demonstrated that migration is a patterned process, including the jobs immigrants take, where they settle, and the routes they use to enter the country (Portes & Rumbaut 2014; Téllez 2007). I posit that similar patterns should emerge concerning where people die as they clandestinely travel through the Sonoran Desert, such as between recovery location and the individual's country of origin.

## **Project Goals**

Two goals drove this research: First, to create a model for predicting the country of origin of deceased migrant remains recovered from the southern Arizona desert. This was completed by combining open geographic information system (OGIS) data of the recovery location of deceased undocumented border crossers and the well-established methods correlating the geographic or ancestral origins of an individual with their skeletal attributes. Second, to expand the body of knowledge regarding Hispanic individuals through continued skeletal data collection and population variation documentation. Novel data collected at the PCOME was added to the Forensic Anthropology Databank (FDB) and Macromorphoscopic Databank (MaMD) to aid in future research projects on these populations. The addition of data from new and existing populations within these databanks is necessary to improve reference samples and thus the empirical power of analytical methods in forensic anthropology.

## **Research Questions**

Several research questions directed these project goals. They are grouped based on the project's three main areas of investigation: biological and geospatial analyses and the interaction between the two.

### *Biological Data Analysis Questions*

The research question related to the biological analysis centers on whether there are correlations between an individual's craniometric and macromorphoscopic features and their country of origin. This question will be investigated using cranial data of identified individuals from Hispanic populations, including individuals analyzed by the PCOME. A comparative dataset from identified American Black and White populations will also be analyzed to

maximize the cranial differences (both metric and nonmetric) that distinguish Hispanic populations from other ancestral groups. This will improve the discriminating powers of these analyses and the subsequent predictive abilities of the project-developed models.

### *Geospatial Data Analysis Questions*

The research question related to the geospatial data involves whether there are spatial correlations between an individual's demographic information, specially year of recovery, sex, and country of origin, and the location where their remains were recovered in southern Arizona. Pilot investigations into the general spatial properties of these data have demonstrated evidence of positive autocorrelation for sex and year of recovery. Due to the patterned nature of international migration between the United States and Mexico, there should also be statistically significant spatial patterning of individuals recovered in southern Arizona. This will indicate a geographic relationship between individuals based on certain aspects of their migration experience, including their place of origin.

### *Combined Data Analysis Research Questions*

If there are positive correlations and promising predictive capabilities following the investigation of the first two research questions, the last question focuses on whether the predictive powers of the biological (metric and macromorphoscopic data) and geographic (the location of where people from certain countries are recovered) data can be combined to predict where an unknown individual originates. While this third research question will only be testable using a small subset of the data, its results may be applied to the PCOME operating procedures wherein its true functionality in facilitating a more efficient identification can be tested.

Through this research a more efficient use of local and national resources directed to the identification and repatriation of migrant remains in southern Arizona can be established. Additionally, the potential for applying these methods to other border regions undergoing similar circumstances will be exhibited. Many agencies (including the PCOME) undertake the responsibility of identifying any individual who dies in their jurisdiction, regardless of nationality. Therefore, means to quickly associate missing persons reports with unidentified remains should be investigated.

### **Organization of Chapters**

To understand the nature of the geospatial properties of migrant deaths in the southern Arizona desert, this dissertation takes the following format. Chapter Two discusses the regional history of the Greater Southwest as it pertains to the political and economic ties between the United States and Mexico. It also looks at the pertinent political developments related to Mexican and Latin American immigration to the United States and how its negative connotations have helped create the current humanitarian crisis.

Chapter Three reviews the theoretical approaches to international migration which best correspond to the current state of migration of Mexican and Latin American individuals into the United States. This discussion places the presumed geospatial patterns of migrant deaths into a broader historical and theoretical context of immigration occurring region.

Chapter Four provides a theoretical and methodological introduction to the topic of biological distance (or biodistance) analysis in physical anthropology. As the cranial data in this study is being used as a proxy for the geographic origin of an individual, understanding the biological mechanisms behind it is pertinent. This chapter also briefly discusses current research

into the identification of Hispanic populations in the region in efforts to identifying unknown migrant individuals.

The methods and associated materials used to investigate the research questions are introduced in Chapter Five. These include both biological and geographic data from several sources. The biological data are a combination of craniometric and macromorphoscopic data from Hispanic individuals with known demographic information. The biological analyses include separate techniques to handle the craniometric and macromorphoscopic data, as well as random forest models combining the two. The geographic data were first evaluated using exploratory spatial data analysis (ESDA) techniques to understand the extent of spatial patterning in the locations of migrant death recoveries. The geographic data derives from the undocumented migrant cases analyzed at the Pima County Office of the Medical Examiner provided by the nonprofit group Humane Borders. These include both identified and unidentified individuals, but the identified will be of primary focus to understand the geographic patterns of migrant deaths in the southern Arizona desert. Then, the two datasets (biological and geographic) were analyzed using methods of geographically weighted regression (GWR) to predict the country of origin of unknown skeletal remains.

The results of the biological analyses are presented and discussed in Chapter Six. These include the craniometric, macromorphoscopic, and random forest model results from the cranial data. The ESDA results from the geospatial analyses and the GWR modeling and prediction results are given in Chapter Seven. A discussion of the implications of the biological and geographic results and concluding thoughts are provided last in Chapter Eight. Future research directions are also presented as this is unfortunately an ongoing crisis, with a presumed need for refined techniques and further inquiries.

## REFERENCES



## REFERENCES

- Anderson BE. 2008. Identifying the dead: Methods utilized by the Pima County (Arizona) office of the medical examiner for undocumented border crossers: 2001-2006. *Journal of Forensic Sciences*, 53(1): 8–15.
- CBP.gov. 2015. (United States Border Patrol) Total Illegal Alien Apprehensions by Fiscal Year (Oct. 1st through Sept. 30th) Washington DC: United States Border Patrol, Available from: <https://www.cbp.gov/sites/default/files/documents/BP%20Total%20Apps%2C%20Mexico%2C%20OTM%20FY2000-FY2015.pdf>.
- CBP.gov. 2016. (United States Border Patrol) Southwest Border Deaths by Fiscal Year (Oct. 1st through Sept. 30th) Washington, DC: U.S. Customs and Border Protection. Available from: <https://www.cbp.gov/sites/default/files/documents/BP%20Southwest%20Border%20Sector%20Deaths%20FY1998%20-%20FY2015.pdf>.
- Chamblee JF, Christopherson GL, Townley M, DeBorde D, Hoover R. 2006. Mapping migrant deaths in southern Arizona: The humane borders GIS. Unpublished report.
- Colibrí Center for Human Rights. 2018. Press Kit; <http://www.colibricenter.org/wp-content/uploads/2018/04/Press-Kit-English-PDF.pdf>.
- Grannis W. 2012. Detecting and Assessing Aggregate Human Pedestrian Migration Patterns Using High Resolution Multispectral Imagery and Geographic Information Systems Texas State University: Doctoral dissertation.
- Giordano A, Spradley MK. 2017. Migrant deaths at the Arizona–Mexico border: Spatial trends of a mass disaster. *Forensic Science International*, 280: 200-212.
- Jones-Correa M, de Graauw E. 2013. The Illegality Trap: The Politics of Immigration & the Lens of Illegality. *Daedalus*, 142(3): 185-198.
- Kelly A. 2017. FACT CHECK: Have Immigrants Lowered Wages for Blue-Collar American Workers? August 4; Available from: <https://www.npr.org/2017/08/04/541321716/fact-check-have-low-skilled-immigrants-taken-american-jobs>.
- Lawrence S, Wildgen J. 2012. Manifold Destiny: Migrant deaths and destinations in the Arizona desert. *Growth and Change*, 43(3): 482-504.
- National Academies of Sciences, Engineering, and Medicine. 2017. The economic and fiscal consequences of immigration. National Academies Press.

- Martínez D, Reineke R, Rubio-Goldsmith R, Anderson B, Hess G, Parks B. 2013. A continued humanitarian crisis at the border: Undocumented border crosser deaths recorded by the Pima County Office of the Medical Examiner, 1990-2012. Tucson: Binational Migration Institute, University of Arizona.
- Massey DS. 1987. Understanding Mexican migration to the United States. *American Journal of Sociology*, 92(6): 1372-1403.
- Massey DS. 1988. Economic development and international migration in comparative perspective. *The Population and Development Review*, 383-413.
- Massey D, Espinosa K. 1997. What's Driving Mexico-U.S. Migration? A Theoretical, Empirical, and Policy Analysis. *American Journal of Sociology*, 102(4): 939-999.
- Orrenius PM. 1999. The role of family networks, coyote prices and the rural economy in migration from Western Mexico: 1965-1994. Dallas: Federal Reserve Bank of Dallas, Working Paper, 9910.
- Pima County Office of the Medical Examiner (PCOME). 2017. Annual Report. Tucson, Arizona.
- Portes A, Rumbaut R. 2014. *Immigrant America : A Portrait*. University of California Press.
- Reineke R, Martínez DE. 2014. Migrant Deaths in the Americas (United States and Mexico). *Fatal Journeys*. pp. 45 - 83.
- Rubio-Goldsmith R, McCormick MM, Martinez D, Duarte MI. 2006. The “Funnel Effect” and Recovered Bodies of Unauthorized Migrants Processed by the Pima County Office of the Medical Examiner, 1990-2005. Tucson: Binational Migration Institute, Mexican American Studies and Research Center, University of Arizona, Report.
- Rubio-Goldsmith, RM, O’Leary AO, Soto A. 2014. Protocol Development for the Standardization of Identification and Examination of UBC Bodies Along the U.S.-Mexico Border: A Best Practices Manual. Tucson: Binational Migration Institute, University of Arizona.
- Singer A, Massey D. 1998. The Social Process of Undocumented Border Crossing among Mexican Migrants. *The International Migration Review*, 32(3): 561-592.
- Téllez MEA. 2007. Detour through the Sonora-Arizona Desert: New routes of Mexican emigration to the United States. *The Journal of Latino-Latin American Studies*, 2 (4): 4-19.

## **CHAPTER 2. HISTORICAL RELATIONS BETWEEN THE UNITED STATES AND MEXICO**

### **Introduction**

The current geopolitical climate concerning undocumented migrants perishing while crossing the United States-Mexico border is the culmination of centuries of unrest between several populations vying for land and its control. This makes the border region a place of dynamic processes involving the exchange of culture, commodities, and people. Understanding how each side of the border developed separately, yet integrated, can illuminate the current state of the border region. This chapter will explore both the cultural and political-economic history of the Greater Southwest of North America in order to place the present United States and Mexico border relations into the proper context. Only then can studies that involve the current state of political affairs involving undocumented immigration be understood. In particular, this context helps highlight how the present trends of migrant border crossings in southern Arizona and its high rate of deaths were created through these historical processes.

### **Creation of the Borderlands**

The area known as the Greater Southwest of North America has a long history of human occupation that has shaped the current landscape both physically and culturally. From the Spanish arrival in the mid-1500s, to the American expansion in the 1800s, the region had already been inhabited for over 12,000 years, with dramatic changes occurring in a relatively short amount of time (Spicer 1962). In this context, the lands designated here as the “Greater Southwest” include the southern portions of California, Arizona, New Mexico, and Texas and the northern Mexican states of Sonora, Chihuahua, Coahuila, Nuevo León, and Tamaulipas (Sheridan & Parezo 1996).

Beginning as an unknown region aside from small groups of indigenous peoples, and considered a terra nullius (empty land) to most Europeans, starting in the 1500s the Greater Southwest became a place of hope, conquest, and eventual division. A historical trend of ‘outsiders’ entering the region to control and profit from its inhabitants and resources was established; only to be repeated by the next group (Spicer 1962). Physical markers of nation-states and their political control were constructed where once only verbal or culturally-understood distinctions between land ownership were necessary (St. John 2011). Time and again, a new group entered the region and attempted to control the people already living there, who themselves wish to live as they had been; or at least since the adjustments to the preceding intrusion (Spicer 1962). As such, a composite social and cultural reality developed. Today, this can be readily seen in both the lived experiences of those within the United States and Mexico border region, and those attempting to traverse it.

### *Border Concept*

The Greater Southwest region has transformed from a frontier, to an area defined by a border (or zone between two polities), to one surrounded by a boundary, or line of segregation (Nevins 2010). The physical and ideological differences between these definitions have impacted not only the physical landscape, but also the experience of people traversing them. While these definitions are the typification of ideals:

the terms describe ongoing processes of construction of different types of geographically and historically specific divisions and zones of contact between territories, ones that reflect power relations between groups and that have very real impacts on people’s lives (Nevins 2010: 229).

Borders “shape and are shaped by what they contain, and what crosses or is prevented from crossing them. The ‘container’ and the ‘contents’ are mutually formative” (Anderson & O’Dowd

1999: 594). The border, however, must be understood outside two-dimensional terms as it merely “represents an arbitration, and simplification, of complex geo-political, political, and social struggles” (Anderson & O’Dowd 1999: 595) becoming much more than a necessary line to cross in the movement between nations.

Often the literature switches between the terms borders and boundaries as though they are synonymous, but a distinction in the rigidity and implications behind each term has been made. While there are several types of boundaries, all go beyond just being political zones (as with borders) and are distinct, enforceable barriers when deemed necessary (Prescott 1987; 2008). The political transformation of the US-Mexico border from a place lacking definition to one of boundaries patrolled by man and machine only took roughly three decades; a relatively short amount of time.

Despite current political rhetoric, the boundary between the United States and Mexico may be better described as selectively permeable; with a degree of openness dependent on who or what is trying to cross it, rather than an inflexible division between the two countries. Anderson and O’Dowd (1999) point out how the differential porosity of the US-Mexico border is evident in the ease with which some peoples and goods are able to pass as opposed to others. However, the government did not “set out to close the border, but rather to improve its ability to manipulate spatial controls to reflect state priorities” (St. John 2011: 6). In the case of the movement of people, it has either been a negligible line, or a barrier depending on the time in history, as will be discussed shortly.

## **European Occupation**

When Spanish explorers began entering the region in the 1530s, the Greater Southwest was not a barren landscape but one inhabited by diverse indigenous peoples with their own

connections. Indigenous groups including the Pimas, Navajos, Apaches, Hohokam, and Pueblos peoples lived and thrived in the region for centuries before Europeans began their explorations west (Spicer 1972; Sheridan 1965). This section provides a brief discussion of European campaigns in the region and their actions taken against Native populations. This historical background provides reference for the present-day cultures living in the Greater Southwest and highlights how the current political and social powerplays occurring within (and between) these communities is not a new phenomenon. The turbulent, yet entwined, environment of the region stems from its unnatural division, as seen through its deep historical ties.

### *The Spanish Frontier*

As the first of several groups, Spanish explorers began entering and claiming the lands of central and northern Mexico in the early 1500s, continuing well into the 1800s throughout the modern states of Texas, New Mexico, Arizona, and California. With such a large area claimed, settlers slowly made their way into the areas beyond New Spain's northern capital of Mexico City (Spicer 1962). Through a mutual relationship between the indigenous Native Americans and Spanish which provided protection and labor for each, the frontier ecology of the new Spanish holdings was growing (Radding 1998). Yet, social unrest became the prevailing theme as the region vied for control and autonomy; while still hoping to benefit from the other's presence.

State-run *presidios*, or fortified military towns, were developed to not only protect the Spanish holdings but also as an attempt to directly incorporate Indians into Spanish politics and culture (Spicer 1962). The process of inclusion helped rectify what the Spanish saw as a "cultural void" that needed to be filled in the indigenous peoples (Spicer 1962). The indigenous groups were lacking the physical symbols of what the Spanish considered culture. Therefore, it was up to the Spanish to provide them the means. The lasting effects of *presidios* and missions, both

physically and culturally, are still witnessed today in the region in the forms of modern urban towns and cities such as the Presidio San Agustín del Tucsón, the founding structure of the modern multi-cultural city of Tucson, Arizona (Spicer 1962).

### *The Mexican Frontier*

In 1821, following the 11-year Mexican War of Independence, Mexico broke free from Spanish rule with the relatively peaceful transition under the Plan of Iguala, also known as The Plan of the Three Guarantees (Weber 1982). Part of the Plan was that all inhabitants of the new state would become Mexican citizens, including the indigenous populations. Therefore, it was no longer the job of the Spanish crown to deal with these foreign lands and peoples, but the new polity. Word of independence spread slowly through the northern frontier and Weber (1982) describes the Mexican frontier as an exceptionally transitional area due to its distance from the center of power in Mexico City. Furthermore, many military operations were removed from the frontier lands to assist in the War of Independence leaving armed holdings scarce and political control up for debate (Radding 1997).

The new Mexican Republic found itself struggling to find political standing while the indigenous peoples still tried to maintain their traditional way of life (Radding 1997). Unlike the Spanish who were more interested in pacification and superficial acculturation, the Mexican agenda was focused on embedding society-level change through forced assimilation (Spicer 1962). The Mexican frontier wavered under the added effects of a government attempting to quickly transition from a monarchy to a constitutional order along with its own power struggles with local tribes. The assimilation program did not go as planned and the Mexican government resorted to military colonization and deportation within northern Mexico for some native groups (Spicer 1962). Questions over loyalty and the resilience of the new government were quiet, but

present, and have long underlined its political endeavors in the region (Weber 1982). Thus, Mexico began on unsteady political and economic terms that did not go unnoticed even on the fringes, which were some of the first areas to cast doubt on the new authority. This laid the foundations for later American encroachment on the land and its resources, continuing the often-contentious efforts for control of the people and goods of the region.

### **American Developments**

Without the separation of the Greater Southwest into two nations, their hotly debated border and the movement of people and goods over it would not be an issue. Therefore, the creation of the United States and Mexico border must be reviewed. Again, the dynamic nature of the region, including the struggle between the indigenous peoples and foreign frontiersman to control the land and its peoples, is highlighted.

Starting in the mid-1820s, Mexico successfully began inviting American and foreign colonists to Texas to increase the population and revenue from the country's periphery (Weber 1982). Thus, began another power struggle between Mexican and indigenous peoples and the new American arrivals. In 1846, Mexico refused an offer made by the United States to purchase California and New Mexico and US President James Polk declared war on Mexico (Weber 1982). The Mexican War lasted approximately two years and ended with the Treaty of Guadalupe Hidalgo, confirming previous US claims to Texas and the relinquishment of California and parts of New Mexico and Arizona from Mexico to the United States (Cutter 1978). Also included in the treaty were all of present-day Nevada, Utah, parts of Colorado, and the northern two thirds of Arizona. This once again threw the area of the Greater Southwest into



political and cultural conflict with the introduction of yet another new government and set of cultural norms (Sheridan 1995).

Finally, in 1854 with the Gadsden Purchase, the continental United States took its current form (Schmidt 1961). While a relatively small parcel compared to that of most other US land deals, the Purchase claimed a portion of New Mexico and the remaining land south of the Gila River in Arizona. According to Schmidt (1961) the Purchase was “the culmination of the widespread and swelling popular conviction that it was the unmistakable preordained mission of the American republic to expand over the whole continent” (pg. 245). It was more than just a simple money-making endeavor; it was one filled with conviction to preside over the land and everything within it. It was also a fairly easy transaction in part due to the need for funds by Mexican President Santa Anna to remain in power (Schmidt 1961).

Drafting the Mexican-American border was not an easy task as it was a physical and ideological power struggle, which still continues (St. John 2011). While the eastern-most portion of the border followed the natural boundary of the Rio Grande, lines were simply drawn across the map connecting significant places on the western edge (El Paso, San Diego Bay, etc.) and even then only after several heated meetings and negotiations (St. John 2011). The new southern Arizona boundary was based on a seemingly arbitrary line through the region with disregard for its current inhabitants, but still with significant considerations made by its crafters (Schmidt 1961). Disputes over land and its indigenous inhabitants, especially the Tohono O’odham, were still an issue for the United States. However, the government’s sole focus was no longer on controlling the peoples of their new territories, but rather to also colonize the land with its own citizens and reap its rewards (Spicer 1962). The creation of the US-Mexico border became not

only a symbol for the control over the region, but also a physical barrier which affects the lives, and deaths, of many people.

## **United States and Mexico Immigration and Policy Relations**

Policies regarding the movement of goods and people between Mexico and the United States have shifted from mere formalities to highly politicized issues over the course of several decades (Appendix 2-1). Seeking a balance between minimizing perceived threats and maximizing monetary benefits, the US-Mexico border has been under intense scrutiny by both sides. The previously discussed tradition of pacification laid the groundwork for the current marginalization of “outsiders” and the justification for international migration policies. This section provides another temporal review of the Greater Southwest since the final separation between the two sovereignties in the 1850s, highlighting the politico-economic trends surrounding trans-border movement and related political agenda. Three phases of economic and legislative actions, with Mexican immigration often a central focus, have become apparent. The creation of the illegal immigrant and the increase in border enforcement and militarization are also considered here as they often fall in line with these economic and political events. Through these lenses, the current state of trade and immigration affairs between these connected national systems can be understood.

### *Mexican Politico-economic and Immigration Trends*

#### First Wave

The first wave of Mexican emigration to the United States can be attributed to a combination of development and industrialization in the Southwestern US and the commercialization of Mexican agriculture in the 1870s, which ended with the US market crash

in the 1920s. Political action responding to the travel of goods and people over the newly created border soon followed.

The economic hardships of Mexico, such as the Mexican Revolution (1910 - 1920), “pushed” many people north across the border where employment prospects were more fruitful, in part from foreign-based opportunities (Mora-Torres 2001). “The annual number of Mexican entries grew from just 10,000 in 1913, on the eve of World War I, to 68,000 in 1920, and peaked at 106,000 in 1924. According to official U.S. statistics, some 621,000 Mexicans entered the United States between 1920 and 1929 (Cardoso 1980), a figure not reached again for decades” (Durand et al. 2001: 109).

Cooperative policing of the US-Mexico border began in the 1870s with both sides sending military units to suppress a new rise in violence (Mora-Torres 2001). However, the border itself was not highly restricted and people were able to cross freely to work and acquire goods (St. John 2011). The transformation of an easy trip across the border to one labeled as “international immigration” occurred via restrictive policies at the beginning of the 20<sup>th</sup> century (Appendix 2-1).

Following the United States market crash in 1929, the number of Mexican immigrants dropped dramatically due to the decline of available work. Immigrants were also scapegoated as a cause for the collapse, leading to the deportation of roughly 453,000 individuals between 1929 and 1937 (Durand et al. 2001). During this same period, the modernization of the Mexican state was in progress and included massive agricultural investments in both the private and public sectors. “Millions of hectares were distributed to peasants in an effort to recreate the communal land system of the past and reverse the enclosures of the Porfirian era” (Massey 1988: 403). This also led many Mexican nationals to voluntarily return home in hopes of capitalizing on the

successful economic growth and development (Massey 1988).

### Second Wave

The second wave of Mexican immigration is primarily defined by the United States' efforts at remedying World War II labor shortages with the largest campaign thus far to attract migrant workers to the United States: the Bracero Program. Not surprisingly, this wave ended with the Immigration and Nationality Act (INA) of 1965, one of the most restrictive immigration policies to date.

Beginning in 1942, the Bracero Program provided temporary work visas to Mexican agricultural migrants around the Southwest. The program was meant to be a provisional fix to labor shortages (Calavita 1992). The success of the program on the productivity of the US agricultural market, as well as the appeal of low-wage foreign workers, resulted in the Bracero Program lasting until 1964. "In all, some 4.6 million braceros and 565,000 legal emigrants entered the United States during the period 1940-1964" (Massey 1988: 404). Part of the program's success was that as opposed to European immigrants who were more permanent in their relocation, Mexican migration tended to be cyclical in nature. Workers tended to return home at the end of the harvesting season and their visas expired. Therefore, the fear of long-term labor loss in Mexico and jobs for Americans was suppressed with each nation benefiting from the deal.

The first official end of the Program in 1947 was welcomed by both the Braceros and American farmers since now they could continue working or employing foreign laborers (albeit illegally) without having to deal with the 'red tape' from either government (Martin 1999). Only under smaller visa programs such as the H-2 section of the Immigration and Nationality Act of 1952 were farmers legally allowed to continue to apply for foreign labor, and only then after first recruiting in the United States (Martin 1999). The termination of the Bracero Program and the

passing of the INA of 1965, which reduced annual guest worker visas, led to a new era of Mexican and Latin American immigration which continues today (Calavita 1992).

### Third Wave

The third, and current, wave of Mexican immigration led to the ongoing humanitarian crisis in the greater Southwest underlying this research project. It began in the late 1960s and has been the most proliferative thus far. Legislation such as the 1965 INA, followed by the cessation of the Bracero Program, resulted in more people needed, and wanting, to work in the United States than opportunities to do so legally (Calavita 1992). This imbalance led many to seek alternative measures to enter and work as they once had.

This current wave has also resulted in Mexico becoming the preponderant country of origin of all foreign-born individuals in the United States with many traveling and living unauthorized. Since 1980, Mexican-born individuals have been the largest immigrant group in the United States and comprise approximately 28% of the 41.3 million foreign-born living here (~11.6M) (Zong & Batalova 2015). Additionally, the number of immigrants from Central America has more than tripled since the 1980s, with approximately 3.4 million currently living in the United States and approximately 85% of them (2.9M) from the Northern Triangle (El Salvador, Guatemala, and Honduras) (Lesser & Batalova 2017). Several reasons are attributed to this dramatic increase and include social and economic factors in both the United States and Mexico, as well as changes in US immigration policy to be discussed more in-depth in the next section.

Although many individuals continued their traditional patterns of work migration, albeit now deemed 'illegal', the establishment of the Border Industrialization Program (BIP) in 1965 provided much-needed opportunities (Urquidi & Villarreal 1978). This program began one of the largest United States investments into the region with the building of *maquiladoras*, or assembly

plants, in Mexican border cities. Traditionally, *maquiladora* workers have been young women who stay within the country to work, unlike the Bracero Program which was comprised mainly of men who went elsewhere for employment. In 1975, the number of *maquiladoras* increased by 288% from 1970, and in 1975 at least a third of regional income and employment stemmed from BIP projects (Urquidi & Villarreal 1978).

Even though the BIP and other neoliberal development strategies, such as the privatization of agricultural land and the lifting of foreign investor restrictions (McMichael 2016), resulted in brief periods of economic growth, the Mexican economy was still struggling. A rising national debt, inflation, and a dramatically decreasing value of the peso led to Mexico declaring a debt moratorium in 1982 (Greenberg et al. 2012). What followed were several loans and reform packages by the International Monetary Fund, World Bank, United States, and others (McMichael 2016).

Several systemic issues within Mexico also contributed to a continuing influx of migrant workers to the United States during the 1980s and 90s. These include a growing Mexican population, disparities between United States and Mexican earnings, and agricultural modernization tied to a shift from subsistence farming to cash crops in Mexico reduced the need for local workers (Massey 2002; Vernez & Ronfeldt 1991). Combined with the US reliance on cheap labor and restrictive and dysfunctional policies, these external issues left a large and important role in the American economy in need, with limited means to properly fill it (Calavita 1992; Rosenblum & Brick 2011).

## **Current Political-Economic Developments**

The North American Free Trade Agreement (NAFTA) created the largest political-

economic policy thus-far regarding United States and Mexico commodity exchanges. The Agreement was signed in 1994 and removed tariffs and other trade barriers between the United States, Mexico, and Canada. Meant to ease trade more so between Mexico and the United States, there was also hope of “fast tracking” Mexico towards modernization. Additional goals included foreign capital investments to help stabilize exchange rates, as well as provide funds for Mexico’s debts. Therefore, NAFTA would provide the (relative) stability of the US market for investors as well as access to it (Castañeda 1993).

The long-term effects of NAFTA are many and depending on the vantage point, both positive and negative. Retrospective studies have shown that despite some fears of losing US jobs to the southern market, larger positive economic trends overshadowed any loss of employment (Burfisher 2001). Unsurprisingly NAFTA appears to have had a much larger positive effect on the Mexican economy than the United States regarding both economic growth and trade balance (Burfisher 2001). All Mexicans did not react positively to NAFTA and different sectors of the populations were impacted differently by the neoliberal reforms that surrounded the treaty. The implementation of NAFTA came on the heels of privatization measures on the agricultural and public sectors set forward by Harvard economist-turned President Carlos Salinas de Gortari (1988-1994) to solve Mexico’s debt problems (McMichael 2016). Many indigenous groups including the Zapatista of Chiapas vocally opposed the deal (McMichael 2016). They took active resistance to the Agreement as it would directly affect their ability to continue their subsistence farming-based livelihoods with American and Canadian farmers now able to export crops to Mexico (Greenberg et al. 2012). On the northern edge, an influx of trade and “the force of globalization hits borderlands quite hard: the transitions making borders into economic bridges rather than barriers are hardly nice or easy. Extant transnational

communities in border regions thus are among the first to grapple with the positive and negative effects of trade liberalization” (McCrossen 2009: 51).

Today, in terms of relative gross domestic product (GDP), the Mexico-America disparity is outside the top 10 worldwide (at 17<sup>th</sup> in 2004) behind many other neighbor-countries with similar economic gaps in Africa and the Middle East. However, cumulative differentials of wealth, commerce, and border length seen between the US and Mexico are still worth noting (Moré 2011). Their intertwining economic relationship only exacerbates the political and social operations of the two countries in terms of trans-border movement and the restriction of certain aspects, such as people.

### *Immigration Policy Trends*

Following in line with, and often in response to, the aforementioned immigration waves, several phases of international immigration policy reform and border enforcement measures were instituted (Appendix 2-1). While the United States has made international immigration a key issue in state and federal politics, the regulation of Mexican and Latin American individuals was not a primary focus until the early to mid-20<sup>th</sup> century. Most instances of immigration-related policies focused on a specific country or ethnic group which at that time was deemed in need of regulation, while being less restrictive to others. This inconsistency in who is and is not allowed into the United States reflects the nation’s insecurity and duplicity regarding immigrants (Calavita 1994). While some major developments in the history of international immigration legislation must be addressed, those that deal with Mexican immigrants will be the primary focus here.



### First Phase

The first phase of immigration policy changes began not long after the final shaping of the present-day United States. Despite American rail company's active recruitment of East Asian labor, the Chinese Exclusion Act of 1882 barred the immigration of Chinese workers for 10 years and denied their eligibility for naturalization. In fact, some of the first duties of patrolmen along the US-Mexico border were to find Chinese workers attempting to illegally cross into the US. In 1894, the Bureau of Immigration Enforcement was created under the Treasury Department and expanded the state's control over official ports of entry and in 1914 they were outfitted with boundary inspectors patrolling with boats and automobiles (Nevins 2010). Yet by 1919 there were only 151 immigrant inspectors for the entirety of the southern border and they were expected to remain at one of the 20 official ports of entry, while mobile patrolmen numbered around 60 (Nevis 2010).

More rigorous border-crossing requirements were put into place in the 1910s and even then, it was still fairly easy to cross without the proper documents (St. John 2011). These policies included the Immigration Act of 1917 which attempted to further exclude those deemed "unfit" for American citizenship. The Act called for a literacy test and entry fee of \$8 a person but was still mostly focused on Asian immigrants (Nevins 2010). As many immigrants were illiterate even in their native language or were unable to pay the tax, the 1917 Act drove many entrants to find other, and more circumferential, means to enter the country. Mexican nationals were initially included in the 1917 Act, but those working in the agricultural, mining, and railroad sectors were quickly made exempt. Already, the importance of seasonal migration from the south was being recognized and slowly making its way from the border region into the country's interior (Ettinger 2009).

The 1920s saw both a massive wave of immigration into the United States as well as more aggressive attempts by the US government to mitigate them. While the Quota Act of 1921 put the first numerical limits on international immigrants (at 3% of the 1910 census), the Immigration Act of 1924 made them permanent, at 2% of the 1890 census. It also excluded those ineligible to citizenship (in part via the imbedded Oriental Exclusion Act) but did not restrict countries in the Western Hemisphere, further highlighting the reliance of the US agricultural market on Mexican and Latin American immigrant labor (Ngai 2004; Nevins 2010). In addition, this was the first law ever to eliminate “the statute of limitations on deportations for nearly all forms of unlawful entry and provided for the deportation at any time of a person.... without a valid visa or without inspection” (Ngai 2004: 60).

The US Border Patrol was established shortly after as part of the Labor Appropriation Act of 1924 and made a division of the Department of Labor with 450 agents. The number of officers increased to 916 by 1939. In 1940, the USBP was expanded and moved to the Department of Justice with personnel totaling approximately 1,531 officers (Hernandez 2010).

### Second Phase

The 1920s saw a rise in deportations and the number continued to increase following the immigrant scapegoating for the 1929 market crash. An estimated “...415,000 Mexicans were forcibly expelled from the United States between 1929 and 1935, and ... 85,000 more left ‘voluntarily,’ but usually under intense pressure from local officials” (Cornelius 1978: 16). Immigration policy went fairly unchanged for the next decade, but the Second World War highlighted the need for supplementing the American workforce. This led to the relatively immigrant-friendly Bracero Program implemented in 1942. Again, immigration policy went mostly untouched during the post-war rebuilding period and throughout the Cold War. The exception was the McCarran-Walter Act of 1952 which focused more on blocking social

“undesirables” from entering the country (Hing & Romero 2004). The next, and historically largest, piece of legislation did not come to fruition until 1965.

The Immigration and Nationality Act (INA) of 1965 officially ended the Bracero program and brought an overhaul to the immigration system. While the INA removed the old national-origins quotas, it replaced residence visas with uniform caps at 20,000 for all countries outside the Western Hemisphere (Ngai 2004). Of *those* visas, 75% were reserved for “specified ‘preference’ relatives of citizens and lawful permanent residents, and an unlimited number was available to immediate relatives... of U.S. citizens” (Hing & Romero 2004: 95). INA also reduced the number of annual guest worker visas from 450,000 to zero (Jones-Correa & de Graauw 2013). The preference of family reunification over short-term migrants reflected the increasing frustration with government oversight in the agricultural labor market, as well as the decrease in demand for foreign labor due to improved farm mechanization (Calavita 1992; Ngai 2004; Massey et al. 2002). The final severance of the Bracero program via the INA, however, left many without proper means to enter the country to work and, as a result, many of them resorted to undocumented and “illegal” means to continue to support themselves and their families (Alarcón 2011).

### Third Phase

Beginning in the 1980s with the Immigration Reform and Control Act (IRCA) of 1986, the current phase of immigration policy creation focuses on the illegalization of immigration and militarization of the Mexican border. IRCA was meant to “regain control of the border” (Calavita 1994) by increasing border control funding, imposing civil and criminal penalties against companies employing undocumented immigrants, as well as providing amnesty for immigrants who could prove US residency for eight or more years. Approximately 2.7 million immigrants, including 2.3 million from Mexico, were able to legalize their status under the

IRCA; but it failed to address why individuals were continuing to migrate irregularly and thus was no more than temporary patch to a flawed system (Calavita 1989; Bean et al. 1990; Jones-Correa & de Graauw 2013). Further contradictions arose between the inclusive economic policies of NAFTA and General Agreement on Tariffs and Trade (GATT) reforms, and the concurrent restrictive immigration policies such as the IRCA and 1990 Immigration Act. Massey et al. explain that at this time, “U.S. policies transformed what had been a relatively open and benign labor process with few negative consequences into an exploitative underground system of labor coercion that put downward pressure on the wages and working conditions not only of undocumented migrants but of legal immigrants and citizens alike” (2002: 5).

### **Illegalization of Immigration**

After the IRCA, which increased Border Patrol staff by 50 percent, most attempts at immigration reform still focused on fixing the “problem” of immigration, while only perpetuating it as such. The historical production of what it means to be an “illegal” has primarily affected Mexican individuals and “continuously re-stages the US-Mexico border in particular as the theatre of an enforcement ‘crisis’ and thus constantly re-renders “Mexican” as the distinctive national name for migrant ‘illegality’” (De Genova 2004: 171). Immigrants only became “illegal” after the creation of legislation which removes their “legality” and is continually manifested in both name and ideology through policy creation and police state strategies (De Genova 2004). For example, the 1996 Illegal Immigration Reform and Immigration Responsibility Act (IIRIRA) increased border fortifications and the number of felonies associated with immigration (including unauthorized re-entry), blurring the lines of criminality and immigration which was traditionally a civil matter (Jones-Correa & de Graauw 2013; Martinez & Slack 2013).

Through the criminalization of immigration, both social and legal ramifications now affect individuals and exploit an already vulnerable workforce, potentially exposing them to actual illicit activities if incarcerated (Martinez & Slack 2013). Further control is developed through the pacifying concept of “deportability” as migrants are less likely to raise objections to workplace or social injustices under the fear of being deported (De Genova 2002). Therefore, with a system working against them both socially and politically, many immigrants are left with few options other than entering and working in the United States without proper documentation.

### **Militarization of the Border**

The relationship between border enforcement and immigration legislation has become a centerpiece of political discourse, in part due to the aforementioned discrepancies between US policy and practice (Calavita 1989). Despite NAFTA’s increase in commodity exchanges between the United States and Mexico, it restricted worker mobility both across the border and within the US (Alvarez 2012). Around this same time, the Border Patrol’s “prevention through deterrence” program was initiated which included some 4,000 new agents deployed to the Mexican border (Boyce 2015). Today, there are over 59,000 Customs and Border Patrol employees, with just over 22,900 officers and 19,800 Border Patrol Agents (CBP 2017). Other measures included high-intensity lighting, 10-foot high steel fencing, and remote surveillance systems and sensors (Cornelius 2001; Boyce 2015). Whether this was in preparation for the feared influx of Mexican immigrants or in response to previous immigration trends, it began the deadliest period for undocumented migration into the Greater Southwest (Castañeda 1993).

The first round, “Operation Hold-the-Line” was implemented in 1993 within the El Paso sector of the USBP. This was soon followed in 1994 by “Operation Gatekeeper” and “Safeguard” in the San Diego and Tucson sectors, respectively (Nevins 2010). Closer

collaboration between the United States military and USBP has now influenced the Border Patrol's role as the main enforcement taskforce for the Immigration and Naturalization Service (INS) (Dunn 2001). The justification for military support has been to aid in drug enforcement, yet much of the assistance has been towards immigration control. Dunn (2001) finds issue in this dichotomy between the military's goal to suppress enemy threats and law enforcement's supposed focus on the application of the legal system (through the Border Patrol), yet the military influence on USBP practice along the border is apparent.

With the implementation of several USBP programs, including "Operation Gatekeeper," it was hoped that increased enforcement at traditional areas of entry for 'illegal immigrants' would deter crossings as it would leave crossing routes through harsher terrain as the only alternative (Cornelius 2001). This was unfortunately not the case. Enforcement of traditional areas of entry did not always deter people from crossing; instead, many were left with no option other than crossing the border through the rugged desert. There have been over 2,800 cases of undocumented border crosser deaths (UBCs) in the Tucson Sector alone since the data started being collected in the mid-1990s (PCOME 2017). The militarization of the border and corresponding illegalization of immigration has made it not only deadly to enter the United States, but also dangerous to stay.

## **Conclusions**

The rich history of the Greater Southwest is rooted in cultural, political, and economic relations, creating a dynamic environment which still reflects its past. Beginning in the mid-1800's, the struggle over land repeatedly led to the imbalance of power and resources, particularly for indigenous or native Mexicans. The opportunities presented by the United States

for work have been mired by the lack of opportunities of legitimacy along with them (Calavita 1994). Rather than accepting the United States reliance on cheap, foreign labor (as created by past events) and providing resources to this necessary section of the United States population, harmful laws and policies have been established instead (Calavita 1994). In as such, individuals looking for work have continued their journey through the southern border to disastrous results. The present intense program at keeping “illegal” immigrants out of the United States has not quelled the urge of many to continue to try and enter. The historical events of the 19th and 20th centuries provide a solid framework for understanding the US and Mexican borderlands as well as the actors within it.

While it may be ill-advised to examine the migrant experience in fragments, or to study it so holistically as to lose the important details of the process, it is useful to approach the subject from an academic perspective. Thus, the initiation and perpetuation of Mexican immigration to the United States from a theoretical standpoint will be explored in the next chapter.

## APPENDIX



## APPENDIX

### Appendix 2-1. Important Policy and Legislation Timeline.

<b>Date</b>	<b>Policy or Legislation</b>
1882	Chinese Exclusion Act
1894	Bureau of Immigration formed
1917	Immigration Act- established literacy test and entrance fee
1924	Immigration Act- created immigrant quota system
1924	Creation of the US Border Patrol (USBP) under Dept of Labor
1940	USBP moves to under Dept of Justice
1942	Beginning of Bracero Program
1947	First end to Bracero Program
1952	McCarren-Walter Act
1952	Immigration and Nationality Act (INA)– creation of H2 agricultural working visas
1964	Final end of Bracero Program
1965	Immigration and Nationality Act (INA) – reduced guest worker visas
1965	Border Industrialization Program (BIP)
1986	Immigration Reform and Control Act (IRCA)
1990	Immigration Act of 1990
1993	USBP “Operation Hold-the-Line”
1994	North American Free Trade Agreement (NAFTA) signed
1994	USBP “Operation Gatekeeper” & “Operation Safeguard”

## REFERENCES

## REFERENCES

- Alarcón R. 2011. US Immigration policy and the mobility of Mexicans (1882-2005). *Migraciones Internacionales*, 6(20): 185-218.
- Alvarez Jr RR. 2012. Reconceptualizing the space of the Mexico-US borderline, In: Wilson, TM, Donnan, H, editors. *A Companion to Border Studies*. Wiley-Blackwell, pp. 538–556.
- Anderson J, O’Dowd L. 1999. Borders, border regions and territoriality: Contradictory meanings, changing significance. *Regional Studies*, 33(7): 593–604.
- Bean FD, Bean FD, Edmonston B, Passel JS. 1990. *Undocumented Migration to the United States: IRCA and the Experience of the 1980s*. The Urban Institute.
- Boyce GA. 2015. The rugged border: Surveillance, policing and the dynamic materiality of the US/Mexico frontier. *Environment and Planning D: Society and Space*, 34(2): 245-262.
- Burfisher ME, Robinson S, Thierfelder K. 2001. The impact of NAFTA on the United States. *The Journal of Economic Perspectives*, 15(1): 125-144.
- Calavita K. 1989. The contradictions of immigration lawmaking: The immigration reform and control act of 1986. *Law & Policy*, 11(1): 17-47.
- Calavita K. 1992. *Inside the state: the Bracero program, immigration, and the I.N.S.* New York, Routledge.
- Calavita K. 1994. U.S. immigration policy: Contradictions and projections for the future. *Indiana Journal of Global Legal Studies*, 2(1): 143-152.
- Cardoso LA. 1980. *Mexican Emigration to the United States 1897-1931: Socio-economic Patterns*, Tucson: University of Arizona Press.
- Castañeda JG. 1993. Can NAFTA Change Mexico? *Foreign Affairs*, 72(4): 66-80.
- CBP.gov. 2017. (United States Border Patrol) CBP Snapshot. Washington DC: United States Border Patrol, Available from: <https://www.cbp.gov/sites/default/files/assets/documents/2017-Apr/cbpsnapshot-april2017.pdf>
- Cornelius WA. 1978. Mexican migration to the United States: causes, consequences, and U.S. responses. Working Paper. Center for International Studies, MIT.
- Cornelius WA. 2001. Death at the border: Efficacy and unintended consequences of US immigration control policy. *Population and Development Review*, 27(4): 661-685.

- Cutter DC. 1978. The legacy of the Treaty of Guadalupe Hidalgo. *New Mexico Historical Review*, 53(5), 305–315.
- De Genova NP. 2002. Migrant “illegality” and deportability in everyday life. *Annual Review of Anthropology*, 31(1): 419-447.
- De Genova NP. 2004. The legal production of Mexican/migrant “illegality”. *Latino Studies*, 2(2): 160-185.
- Dunn TJ. 2001. Border militarization via drug and immigration enforcement: human rights implications. *Social Justice*, 28(2): 7-30.
- Durand J, Massey DS, Zenteno RM. 2001. Mexican Immigration to the United States: Continuities and Changes. *Latin American Research Review*, 107-127.
- Ettinger P. 2009. *Imaginary Lines: Border Enforcement and the Origins of Undocumented Immigration, 1882-1930*. University of Texas Press.
- Greenberg JB, Weaver T, Browning-Aiken A, Alexander WL. 2012. The Neoliberal Transformation of Mexico. In: *Neoliberalism and Commodity Production in Mexico*, JB. Greenberg, T Weaver, A Browning-Aiken, WL Alexander (eds), pp.1-32.
- Hernandez, K. 2010. *Migra!: A History of the U. S. Border Patrol*. California: University of California Press.
- Hing B, Romero A. 2004. 1965 to 1990: From Discriminatory Quotas to Discriminatory Diversity Visas. In *Defining America: Through Immigration Policy*, Temple University Press.
- Jones-Correa M, de Graauw E. 2013. The illegality trap: The politics of immigration & the lens of illegality. *Daedalus*, 142(3): 185-198.
- Lesser G, Batalova J. 2017. *Central American Immigrants in the United States*. Washington, DC: Migration Policy Institute.
- Martin PL. 1999. Emigration Dynamics in Mexico. The case of agriculture. In: *Emigration Dynamics in Developing Countries*. Vol. 3. Mexico, Central America and the Caribbean, R Appleyard (ed.). Brookfield: Ashgate, pp. 117 – 177.
- Martinez D, Slack J. 2013. What part of “illegal” don’t you understand? The social consequences of criminalizing unauthorized Mexican migrants in the United States. *Social & Legal Studies*, 22(4): 535-551.
- Massey DS. 1988. Economic development and international migration in comparative perspective. *The Population and Development Review*, 4(3): 383-413.

- Massey DS, Durand J, Malone NJ. 2002. *Beyond Smoke and Mirrors: Mexican Immigration in an Era of Economic Integration*. New York: Russell Sage Foundation.
- McCrossen A. 2009. *Land of necessity: Consumer culture in the United States-Mexico borderlands*. Durham: Duke University Press.
- McMichael P. 2016. *Development and social change: A global perspective, 6<sup>th</sup> Edition*. Sage Publications.
- Mora-Torres J. 2001. *The Making of the Mexican Border: The State, Capitalism, and Society in Nuevo León, 1848-1910*. Austin: University of Texas Press.
- Nevins J. 2010. *Operation Gatekeeper and Beyond: The War On “Illegals” and the Remaking of the U.S.- Mexico Boundary*. New York: Routledge.
- Ngai M. 2004. *Impossible Subjects: Illegal Aliens and The Making of Modern America*. Princeton, NJ: Princeton University Press.
- Pima County Office of the Medical Examiner (PCOME). 2017. *Annual Report*. Tucson, Arizona.
- Prescott V, Triggs GD. 2008. *International Frontiers and Boundaries: Law, Politics and Geography*. Leiden: Martinus Nijhoff.
- Prescott V. 1987. *Political Boundaries and Frontiers*. London: Allen and Unwin.
- Radding C. 1997. *Wandering Peoples: Colonialism, Ethnic Spaces, and Ecological Frontiers in Northwestern Mexico*. Duke University Press, pp.1700-1850.
- Radding C. 1998. *The Colonial Pact and Changing Ethnic Frontiers in Highland Sonora, 1740-1840*. In: *Contested Ground: Comparative Frontiers on the Northern and Southern Edges of the Spanish Empire*. Tucson: University of Arizona Press, pp. 52-66.
- Rosenblum MR, Brick K. 2011. *US Immigration Policy and Mexican/Central American Migration Flows: Then and Now*. Washington DC: Migration Policy Institute.
- Ross SR. 1978. *Views across the border: the United States and Mexico*. Albuquerque: University of New Mexico Press.
- Schmidt LB. 1961. *Manifest Opportunity and the Gadsden Purchase*. *Arizona and the West* 3(3): 245–264.
- Sheridan TE. 1995. *Arizona: A History*. Tucson: University of Arizona Press.
- Sheridan TE, Parezo NJ. 1996. *Paths of life: American Indians of the Southwest and northern Mexico*. Tucson: University of Arizona Press.

- Spicer ES. 1962. *Cycles of Conquest: The Impact of Spain, Mexico, and the United States on the Indians of the Southwest, 1533-1960*. Tucson: University of Arizona Press.
- St. John, R. 2011. *Line in the Sand: A History of the Western US-Mexico Border*. Princeton: Princeton University Press.
- Truett LJ, Truett DB. 2007. NAFTA and the maquiladoras: boon or bane? *Contemporary Economic Policy*, 25(3): 374-386.
- Urquidi V, Villareal SM. 1978. Economic importance of Mexico's northern border region. In: SR Ross (ed.). *Views across the Border: The United States and Mexico*. Albuquerque: University of New Mexico Press, pp.141-162.
- Vernez G, Ronfeldt D. 1991. The Current Situation in Mexican Immigration. *Science*, 251(4998): 1189-1193.
- Weber DJ. 1982. *The Mexican Frontier, 1821-1846: The American Southwest Under Mexico*. Albuquerque: University of New Mexico Press.
- Zong J, Batalova J. 2015. *Frequently Requested Statistics on Immigrants and Immigration in the United States*. Washington, DC: Migration Policy Institute.

## CHAPTER 3. MIGRATION THEORIES IN THE US-MEXICO CONTEXT

### Introduction

The migration of peoples between the United States and Mexico has connected the two countries for centuries as highlighted in Chapter 2. Many of the attempts to explain the initiation and continuation of migration use various models, but no one theory will suffice in explaining the motives of such a large and diverse group. A combination of ideas is necessary to understand the past and present patterns of US-Mexico immigration considering its long and complex history. This chapter will discuss the pertinent economic, historical, and socioeconomic factors behind United States and Mexico migration and place them into a larger theoretical framework.

Castles et al. (2014) rightly point out how migration is not a “problem to be solved” but rather a collective **process** and the result of societal transformations based on many social, economic, and political factors. While they argue against too much scholarly deconstruction, it is difficult to not try and provide specific explanations of this patterned human behavior such as money, political turmoil, etc. Theories tend to include economic, historical, or socioeconomic factors (or a combination thereof) for which a brief review of just a few of the core models which apply to the United States and Mexico case study follows.

### Immigration Initiation

Several approaches help to understand the general history of US-Mexico migration at the preliminary emigration stages. These methods highlight the stimulus for initial out-trips from Mexico but are less-equipped to explain the conditions behind sustaining the movements of people. For those, a separate reflection on the pertinent theories behind the persistence of US-Mexico migration will be conducted.

### *“Push-Pull”*

As early as the 1880s, the combined forces behind many migration events were recognized. Ravenstein developed several “laws” of migration which revolved around the internal movement of peoples from rural English settings to more opportunistic towns (Griggs 1977). Perhaps the most well-known migration “law” incorporates the economic decision-making factors behind an individual’s move. These factors act as motivators which both “push” them from their homes, like poor wages and social unrest, and another set of complimentary “pull” factors attracting them to a particular town or region.

In Mexico, “push” factors such as the land-ownership reform by President Porfirio Diaz in the 1880s and a rocky post-Revolutionary economy drove many to look for work in the western United States (Cornelius 1978; Henderson 2011). Meanwhile, rapid development in the mining, ranching, and railway sectors demanded more laborers, “pulling” many to the United States for more lucrative work than could be found at home. This theory has also been used to explain the success of the Bracero Program wherein the “pull” of guaranteed work and wages and the “push” of limited at-home prospects for rural Mexicans led to high number of men emigrating throughout the course of the program, and well after (Massey et al. 2002).

The active recruitment of laborers, such as those of railway and agricultural companies, may have been artificially creating “pull” factors and thus this theory can only be used “post-factum” to explain existing flows, rather than be used for understanding their creation (Portes & Rumbaut 2014). Therefore, intervening factors like the migrant’s age, prior knowledge about where they are going, and available jobs also affects their decision and process of migration (Lee 1966). A Mexican laborer directly recruited to work on a railway line will have a much different experience than one blindly leaving home in search of any type of employment.



While the push-pull theory may work well for understanding the motivations for some first generations of immigrants, it falls short for explaining sustained patterns of migration like that seen between Mexico and the United States. Circular migration, as was common for many Mexican immigrants in the mid and late 20<sup>th</sup> century, included multiple trips between the two locations.

### *World Systems Theory*

World systems theory does well at describing how the global political ecology has played a part in Mexican migration patterns. It describes countries as either part of the core or peripheral zones of the world economy, wherein the surplus of the production activities in the (often poorer) periphery are transferred to those in the (richer) core, with little reciprocation (Wallerstein 1984). This theory explains that the development of capital works within the peripheral zones upsets the traditional social and labor systems and thus initiates out migration (Wallerstein 1984). In this context, the “core” is the United States, which has set its stake into the “peripheral” (Mexico) to exploit its resources: both natural and human (Massey et al. 2002). This seems to be the case with the US recruitment of labor at both the turn of the 20<sup>th</sup> century and in the 1940s with the Bracero program. It can also be applied to the development of industrial zones along the US-Mexico border in both 1800s with the *Zona Libre* and the *maquiladoras* of the BIP in the 1960s. The ease at which American companies were able to develop and recruit throughout the Greater Southwest, as well as federal policies facilitating this supported the “economic and spatial advantages of Mexican migrant labor for U.S. employers” (Rodriguez 2004: 458). This reliance on cheap Mexican labor and lucrative business opportunities around the border only continued the tradition of the United States profiting from Mexico established in the mid-1800’s (Massey et al. 2002).

Massey & Espinosa (1997) argue that in the context of repeat migration “growth in direct foreign investment is associated with lower probabilities of repeat migration among both legal and illegal migrants” (1997: 974). Contrary to this, American investments may have played a larger role in creating underemployment in certain sectors of the labor market. Coupled with Mexican land reform, many workers within the rural sector were forced to look elsewhere, including more stable jobs in America (Cornelius 1978). As the tradition of migration to the United States was not built on the idea of permanent relocation north, the concept of a complete upset of the peripheral zones in evoking outmigration are not best highlighted in this Mexican case study. But rather, this economic development is better understood as a short-term catalyst for leaving, as described better in the new economics of migration theory.

### *New Economics of Migration Theory*

The New Economics of Migration (Stark & Bloom 1985) perspective helps contextualize Mexican migration to the United States as a familial strategy to address economic uncertainties created by “imperfectly developed markets” in areas that have been affected by capitalist penetration (Castles et al. 2014). Herein, a member from one family is sent to work for remittances which motivates other families to send their own members when they see the original’s rise in material wealth. In the 1930s and 1970s attempts at rebuilding Mexico’s agricultural market were made by giving small parcels of land to rural farmers, but without the necessary capital to run them. As a familial coping strategy, young men were sent to work in the United States to earn enough money to develop and sustain the family farm, and hopefully return to it later (Cornelius 1978; Heyman 1991; Massey et al. 2002). As explained by Massey et al. “most migrants are not interested in relocating north of the border permanently. This assertion is backed up by the very high rates at which money is remitted back to families at home and the

extensive repatriation of U.S. savings” (2002: 62). Higher wages available in the United States “offer Mexicans an incentive to migrate not only because they yield higher expected lifetime earnings, but also because they offer poor families a way to loosen liquidity constraints and manage risks” (Massey et al. 1994: 713). While accounting for migrant agency in the initial decision-making process, the new economics of migration only really works regarding the initial motivators for leaving, and other than increasing investment opportunities, lacks the explanatory power for repeat migration.

### **Immigration Perpetuation Theories**

So far there has been only a limited focus on the cyclical reality of US-Mexico immigration in this discussion. Many thoughts have been published to contextualize the perpetuation of immigration between the two countries and those which best frame this process are addressed here.

#### *Dual-market Labor System*

As jobs available for migrants in the United States were primarily for low-skilled workers willing to come in large numbers under less than ideal circumstances, the idea of the dual-market labor system as an explanation was invoked (Reich et al. 1973; Piore 1979). Falling within the historical-structuralist paradigm, this theory divides the labor market into two sectors: with domestic, native individuals supplying the higher-skilled and paying primary job sector, and immigrants supplying the secondary, or low-skilled, low-paying, and temporary sector (Morawska 2012). The continued and easy access to migrants willing to fulfill the secondary market perpetuates its subservient level as employers know they do not need to improve conditions which may be necessary to attract native workers. It holds that “international

migration [is] caused by the structural demand for cheap and dispensable labour rather than decisions made by the individuals” (Morawska 2012: 58). Furthermore, immigration policies tend to also reinforce these dual labor systems as they create barriers for unauthorized migrations to achieve citizenship and become more lucrative in the labor market (Massey et al. 1994). In America, immigrants from Mexico and other Latin American countries (documented and undocumented alike) are usually relegated to lower tier agricultural, industry, or hospitality jobs with difficulty in achieving upward mobility (Cornelius 1978).

This line of thought, however, removes worker agency and presents migrants as mostly confined to low-skilled jobs; which is certainly not always the case. In fact, the sociological effects of immigrant typification may play a large role in the perpetuation of the dual markets wherein certain communities are seen as “smarter” or “harder working” and thus able to succeed in while “widespread discrimination may hold that certain groups are able only to perform low-wage menial labor” (Portes & Rumbaut 2014: 140). This theory is thus based more on labor demand than supply and the type of workers deemed appropriate as fulfilling the demand such as those from Mexico (Massey et al. 1993).

### *Cumulative Causation and Migrant Networks*

In attempts at incorporating the decision-making process, cumulative causation builds upon the concept of migrant networks and transnational communities (Massey et al. 1994). Migrant network theories look at how social ties in both the sending and receiving countries facilitate migrant choices. Networks become forms of social capital which lowers both the social and economic costs of migration and tends to encourage new migrants. Transnational communities are then created within the receiving country as a shared space for immigrants; connecting them to their home nations via both communication and remittances. Although still

fairly economically-based, cumulative causation theories suggest migration induces changes within societies and self-perpetuates future migration events at both the individual and community level. Migration begets more migration through social networks and their association with an individual (and then, community) earning more money, being more successful, or gaining new skill sets, amongst other social, economic, and cultural factors.

...migration as a social process suggests that, while structural factors may initiate migration, once it has begun, they fade into the background. As the migrant career develops and trips are repeated, aspects of the migrant experience itself come increasingly to dominate the decision to make another trip (Massey 1987: 1385).

These labor recruitment and cyclical investment trips set in motion the cumulative process built on social networks and the reliance on US work for overcoming financial hardships or for improving one's socioeconomic status.

The long economic and cultural relationship between the United States and Mexico have created migrant networks with deep ties. Following division of the two nations, the ebb and flow of work and jobs needed by both sides has driven individuals to seek out work away from home. Using survey data from rural communities within the Mexican Migration Project Massey and Garcia España (1987) empirically show that individuals were more likely to emigrate if they lived in a household with a prior migrant or in a community with a large number of people who had worked in the United States. Furthermore, an individual is more likely to make several return trips as "being confident of their abilities to come and go with ease, and to gain access to US employment...migrants with strong network ties tended to take shorter and more frequent trips" (Kossoudji 1992 in Massey et al. 1994: 730). Additionally, sociologists Curran and Rivero-Fuentes (2003) found in their survey of the 1999 Mexican Migrant Project Data that the percentage of households which had adults who had emigrated to the United States ranged from

2% - 66%, and that individuals with access to networks of their own gender increased their chances for international migration (Mullan 1989).

Migrant networks also established the transnational communities seen in large urban centers like Chicago and Detroit (Majka & Mullan 2002) and transformed the demographic of many areas of the Southwestern United States to having majority foreign-born residents (Portes & Rumbaut 2014). Border communities become the places of cultural and monetary exchange, and –as opposed to places of division- they are reconceptualized as places of bridging, acting as “connectors of the diverse and disparate, as well as of the history and meanings of people and places” (Alvarez 2012: 552). While historical-structural factors may cause a person to migrate, it is the social process and dynamics which sustains migration and affects the choices of actions in their communities (Massey 1987).

## **Conclusions**

Given the prolonged and complex politico-economic and social history between the United States and Mexico, it is apparent that there are many variables which contribute to the current state of immigration relations. Some factors which drove the first generation to migrate may no longer be applicable, yet the sustaining power of their actions and the knowledge they gained is still evident in current migration practices. The consequences of recent immigrant enforcement measures have led to a new era of Mexican immigration defined by familial separations and increased border militarization (Massey 2018). Understanding the foundations and practical motives behind international migration is necessary for contextualizing the current issues regarding immigration policy and reform. It also helps explain why so many would risk their lives to try and enter the United States hoping for better opportunities. The established reliance on cheap labor in the United States along with restrictive and dysfunctional policies

leaves a large and important role in the American economy wanting, with limited means for people to do so with proper authorization (Calavita 1994; Rosenblum & Brick 2011).

The effects of this prolonged and complex relationship of the Mexican immigrant life experience should thus be recognizable in death. Presumably, these will become apparent in the subsequent investigation of the recovery location of identified individuals who have perished while trying to traverse the southern Arizona border.

## REFERENCES



## REFERENCES

- Alvarez Jr. RR. 2012. Reconceptualizing the space of the Mexico-US borderline, In: A Companion to Border Studies. Wilson TM, Donnan H (eds). Wiley-Blackwell, pp. 538–556.
- Calavita K. 1994. U.S. immigration policy: Contradictions and projections for the future. *Indiana Journal of Global Legal Studies*, 2(1): 143-152.
- Castles S, De Haas H, Miller MJ. 2014. *The Age of Migration: International Population Movements in the Modern World (5th Edition)*. New York: The Guilford Press.
- Cornelius WA. 1978. Mexican migration to the United States: causes, consequences, and U.S. responses. Working Paper. Center for International Studies, MIT.
- Curran SR, Rivero-Fuentes E. 2003. Engendering migrant networks: The case of Mexican migration. *Demography*, 40(2): 289-307.
- Grigg DB. 1977. EG Ravenstein and the “laws of migration”. *Journal of Historical Geography*, 3(1): 41-54.
- Henderson TJ. 2011. *Beyond Borders: A History of Mexican Migration to the United States (Vol. 13)*. John Wiley & Sons.
- Heyman, JM. 1991. *Life and Labor on the Border: Working People of Northeastern Sonora, Mexico, 1886-1986*. Tucson: University of Arizona Press.
- Lee ES. 1966. A theory of migration. *Demography*, 3(1): 47-57.
- Majka L, Mullan, B. 2002. Ethnic Communities and Ethnic Organizations Reconsidered: South-East Asians and Eastern Europeans in Chicago. *International Migration*, 40(2): 71-92.
- Massey DS. 1987. Understanding Mexican migration to the United States. *American Journal of Sociology*, 92(6): 1372-1403.
- Massey DS. 2018. *The Gendered Consequences of Immigration Enforcement, Feb 6. The Gendered Policy Report*. University of Minnesota. Available at: <http://genderpolicyreport.umn.edu/the-gendered-consequences-of-immigration-enforcement>.
- Massey DS, Arango J, Hugo G, Kouaouci A, Pellegrino A, Taylor JE. 1993. Theories of international migration: A review and appraisal. *Population and Development Review*, 19(3): 431-466.

- Massey DS, Arango J, Hugo G, Kouaouci A, Pellegrino A, Taylor JE. 1994. An evaluation of international migration theory: The North American case. *Population and Development Review*, 20 (4): 699-751.
- Massey DS, Durand J, Malone NJ. 2002. *Beyond smoke and mirrors: Mexican immigration in an era of economic integration*. New York: Russell Sage Foundation.
- Massey DS, España FG. 1987. The social process of international migration. *Science*, 237(4816): 733-738.
- Morawska E. 2012. Historical-Structural Models of International Migration In: Martiniello, M and Rath J, (eds), *An Introduction to International Migration Studies*, Amsterdam: Amsterdam University Press, pp. 57-77.
- Mullan BP. 1989. The impact of social networks on the occupational status of migrants. *International Migration*, 27(1): 69-86.
- Piore M. 1979. *Birds of Passage: Migrant Labor and Industrial Societies*. Cambridge: Cambridge University Press.
- Portes A, Rumbaut RG. 2014. The Three Phases of U.S.-Bound Immigration. In: *Immigrant America: A Portrait*. Oakland: University of California Press.
- Reich M, Gordon DM, Edwards RC. 1973. Dual labor markets: A theory of labor market segmentation. *American Economic Review*, 63(2): 359-365.
- Rosenblum MR, Brick K. 2011. *US Immigration Policy and Mexican/Central American Migration Flows: Then and Now*. Washington DC: Migration Policy Institute.
- Stark O, Bloom D. 1985. The New Economics of Labor Migration. *The American Economic Review*, 75(2): 173-178.
- Wallerstein I. 1984. *The politics of the world-economy: The states, the movements, and the civilizations*. Cambridge University Press.

## CHAPTER 4. BIODISTANCE BACKGROUND

### Introduction

Biological distance, or biodistance, analyses incorporate methods and theories from the fields of geography and human biology to investigate the relationships between individuals as expressed in their physical remains (Larsen 1997). Expanding upon Tobler’s “First Law of Geography” wherein, “everything is related to everything else, but near things are more related than distant things” (1970: 234), biodistance studies are based on the logic that individuals who live close will be more similar than individuals who live further away. These similarities can be observed in the prevalence of both phenotypic and genotypic traits in populations, such as vestigial tails, which result from the transference of genetic information via gene flow (Scott & Turner 2000). The degree of expression and the corresponding power of discrimination vary depending on the spatial scale under investigation (Relethford & Lees 1982). More population-based traits such as incisal winging may be better suited for regional studies, while allelic changes may be best for discerning familial groups. The traits used, and the spatial scale applied are dependent on the research questions asked and must be project-appropriate.

Biodistance studies were some of the earliest scholarly endeavors in physical anthropology, although they were not called such until the 1970s (Hefner et al. 2016). Early works investigated individual human “types”, but soon shifted towards identifying racial, or geographic, groups (Scott & Turner 2000). These groups were based on phenotypic attributes such as skin color and facial structure from the premise that people from one area of the world looked more similar than others. Several field-founding researchers established both metric and nonmetric standards during this time (including those from Morton and later Martin)—with updated versions of many of these methods are still in use today (Hefner et al. 2016). Current

research has shifted to documenting local and temporal variation in populations using biodistance studies driven by genetic models (Konigsberg 2006). These broad categories of quantitative and qualitative analyses each have their own applications and advantages to understanding human biological variation.

Following is a review of the theoretical and historical developments of biodistance analysis, to better understand its significance in the field of physical and forensic anthropology. These methods also form the base of the biological data analysis discussed in Chapter 5 and allow for the prediction of ancestral origins of an individual based on their skeletal remains.

### **Utility and Theory**

Larsen (1997) identifies three motivators for biological distance studies in anthropology: 1) understanding the evolutionary history of a population, 2) addressing archaeological or bioarcheological issues such as kin groups, and 3) providing context for population variability studies. More recently, biodistance analyses have been applied to studies of intra- and inter-group variation for understanding the relationship between group affinity and geographic location(s). Scales ranging from small regional areas to larger geographic zones—like continents—have been investigated to assess which biological traits may help identify an individual's geographic heritage. Ultimately, biodistance studies are either looking for levels of group similarity or dissimilarity, and the results used as proxies for genetic information when that data is incomplete or unknown. These biological patterns can then be applied to the associated covariates of location and time for use in predictive modeling and classifications.

Prompted by Washburn's (1951) push for more rigorous hypothesis testing, research into the biological and evolutionary basis behind morphological traits began in the mid-20<sup>th</sup> century. These included investigations into the correlation between genotype and phenotype to support

the use of physical traits as appropriate substitutes for the underlying genetic structure of a population. Grüneberg (1952) described “quasi-continuous” skeletal traits in mice, wherein multiple expressions were possible, and were interpreted as the byproducts of inherited continuous traits. Berry and Berry (1967) applied these methods to human samples to see the extent of variation in the human skeleton, as well as the applicability of multivariate statistical methods on subjects other than mice. Relethford (1994; 2001; 2002) has provided a number of studies highlighting the correlation between skeletal data and population genetic structure. Dental traits have also proven useful in organizing closely related populations, as thoroughly outlined by Scott and Turner (2000) in their justification for using tooth morphology to study ancient human remains. Similar reviews and studies have been conducted using craniometric data, as well, and more information on the validity of these approaches will be provided in their dedicated sections. While some argue for at least more samples as there are variables measured (Corruccini 1975), Howells (1973) argued that  $n > 50$ , and the appropriate model chosen, would be sufficient to represent the underlying genetic relationships observed in the physical observations (Pietrusewsky 2000).

### *Theories*

Model-free and model-bound theories underlay biodistance studies. Model-free studies indirectly apply population structure models to assess biological differences, while model-bound approaches directly incorporate measures of population affinity to estimate the genetic parameters under investigation (Relethford & Lees 1982). Understanding the biological correlates of the presumed genetic relationship are at the foundation for both methods. What varies is the initial investigatory approach. Either approach could theoretically confound or misrepresent any actual relationships as they may be masked by the research design itself, or

erroneously assumed and resulting in false-positive associations.

Many studies incorporate the model-bound approach of isolation by distance for measuring similarity and dissimilarity between groups. These models assume a positive correlation between genetic relatedness and spatial distance, and a negative correlation between relatedness and temporal distance when controlling for the other variable (time or space, respectively) (Konigsberg 1990). One model is the migration matrix which looks at the effects of gene flow on localized populations and calculates the probability that an individual from one population came from another subpopulation (Konigsberg 1990). This model is useful in studies investigating patterns of human migration within a relatively-closed region where there is assumed to be a finite number of core populations contributing to the observed biological variation; such as in Latin America where indigenous groups had been in relative isolation from the rest of North and South America prior to European conquest. Alternatively, the unidimensional stepping-stone model assumes an infinite number of populations in a linear environment who symmetrically exchange migrants (and genetic information) between the adjacent populations at a known rate (Konigsberg 1990). This results in a proportional increase in biological distance as the geographic distance increases between populations.

Model-free approaches are more common in physical anthropology, with the primary focus being the investigation of relative levels of group similarity or differences to understand the effects of population structure on inter-group variation (Pietrusewsky 2000). Model-free studies are “investigated using pattern comparison and correlational techniques, seeking to assess the general degree of relationships, but not their exact form” (Relethford & Lees 1982: 116). Where comparative studies attempt to identify patterns of similarity and relate them to other covariates such as biological or historical factors, differentiation studies measure the amount of

variation without regard to a pattern (Relethford & Lees 1982). Therefore, differentiation methods make it possible to assess the effect of various processes, like migration, on quantitative variants using analysis of variance procedures. Such as the case in Lees and Relethford's 1982 assessments of anthropomorphic measurements across Ireland which highlighted local levels of isolation and diffusion (Cabana & Clark 2011). Conversely, comparative methods investigate the degree of correlation within a population as derived from metric data, usually in the form of a correlation coefficient; the most common method in physical anthropology being the Mahalanobis distance statistic ( $D^2$ ). These comparisons are then indirectly applied to population structure models. No matter the approach, however, biological distance methods are relativistic and comparable only when appropriate (Scott & Turner 2000).

### *Nonmetric Approaches*

Early scholarly pursuits used morphological, or nonmetric, traits to describe the physically-observed differences between 'races. These traits were chosen not only to record (and support) the assumed biological and social hierarchy between populations, but also from a drive to collect and classify the natural world (Wolpoff & Caspari 1997). Nonmetric traits, as the name implies, are those which are described and recorded qualitatively, usually in terms of their presence, absence, or degree of expression. Studies are traditionally divided between epigenetic (or discontinuous) and morphological (i.e. macromorphoscopic) trait investigations; each with their own appropriate applications.

Nonmetric traits were the focus of some of the first physical anthropological studies such as those by Russell (1863-1903) and LeDouble (1848-1913) with a large sample size and scope. But, they were representative of the time in their focus on documenting anomalous features of the crania rather than contextualizing their results (Hefner et al. 2016). Laughlin and Jørgensen

(1956) set the precedent for contemporary studies by looking at regional differences within the closed 'system' of Greenland. The authors used continuous and discontinuous morphological data to understand the degree of divergence for the various indigenous groups in conjunction with the known historical context. This study corroborated the historical record for Eskimo population interaction within Greenland and established the usefulness of nonmetric cranial traits as a proxy for known genetic relatedness (Laughlin & Jørgensen 1956).

Epigenetic traits are the result of interactions between environmental and biological factors and were not investigated in human studies until Berry and Berry in 1967. This term, however, is outdated as it is now used to reference the effects of environmental factors on DNA and development, such as methylation (Pink et al. 2016), rather than skeletal variants. Therefore, the term "discontinuous" is preferred to describe these traits and avoid confusion. Even the term "nonmetric trait" now tends to refer to those defined below, while "macromorphoscopic" refers to those used in forensic anthropologic research (Pink et al. 2016).

The developmental features associated with discontinuous traits manifest in a number of skeletal variants, usually recorded as either present or absent. Traits are classified primarily as the result of hyper- or hypostotic processes, or as foramina and canals for nerves and blood vessels (Saunders 1989). Traits from the skull, teeth, and postcranial skeleton have been documented and applied to ancient and modern skeletal remains to understand population variation and relatedness. These traits are commonly used in biological distance analyses of past populations for understanding group relationships. Studies regarding the heritability of discontinuous traits routinely report high correlations between phenotype and genetic kinship, which supports their use as proxies for genetic data when investigating biological affinity of past populations (Laughlin & Jørgensen 1967; Berry & Berry 1967; Ossenbergs 1976; Hanihara



2008). Work by Ossenberg (1976) and others (Berry & Berry 1967; Cheverud & Buikstra 1981) highlighted several advantages of epigenetic traits over metric methods—primarily their low environmental plasticity—setting the course for further detailed reports on nonmetric traits such as trait associations and expression symmetry. This includes Hauser and DeStefano’s (1989) ubiquitous report documenting 84 epigenetic [sic] variants of the skull, now routinely cited due to its standardization of trait identification and data recording.

Morphological, or macromorphoscopic, traits are described as “quasicontinuous variables of the cranium reflected as soft-tissue differences in the living” (Hefner 2016: 302) with expressions divided into five class types: 1) bone shape, 2) feature morphology, 3) suture shape, 4) feature presence or absence, and 5) feature prominence or protrusion (Hefner 2016). These traits are also assumed to be the result of genetic and environmental factors which affect cranial development but are more associated with geographic variation as opposed to inter-population relatedness. Morphological traits have been the study of focus primarily in forensic anthropology as they assist in group classification procedures used to assess the ancestry of unknown remains. Research into the applicability of macromorphoscopic traits can be traced back to Hooton’s work in the 1920s and 30s (1925; 1936) with statistical investigations beginning in the 1970s with Finnegan and McGuire (1979). However, it has not been until the last decade that further standardization and testing of the effectiveness, and appropriateness, of macromorphoscopic traits for assessing ancestry has been more meticulously studied (Hefner 2009; Hefner & Ousley 2014). Due to the inherent biases surrounding traditional morphological trait observation and application (Brues 1990), the need for protocol to further standardize and objectify traits, like that for craniometric data, was deemed necessary.

Current morphological trait studies attempt to remove these biases by focusing on

standardizing a limited number of traits which have proven most useful in ancestral group classification and performing necessary statistical analyses (Hefner 2009; Klales & Kenyhercz 2015; Plemons & Hefner 2016). Most traits involve features of the midface (such as nasal aperture width and shape) as well as palatal and vault morphologies. While the traits themselves may not necessarily be new, their objective recording is. Data collection procedures such as the computer program *Macromorphoscopic Traits 1.61* (MMS) (Plemons & Hefner 2016), answered the call to reduce interobserver error and calculate method accuracy. Again, traditional cranial morphology research focused on classifying unknown individuals into one of the Black, White, Native American/Asian population groups. But, as the global population becomes more mobile, these traditional, and limited, ancestral groups are no longer as applicable or appropriate. The question over the reliability of the nonmetric trait list method has long been debated but only more recently addressed (Brues 1990; Rhine 1990; Hefner & Ousley; Hefner 2007; Hefner 2009; Hughes et al. 2011).

Additionally, the growing Hispanic population, or individuals of admixed Latin American descent, in the United States has become an area of increased research in forensic anthropology. At present, it is just as likely for an American forensic anthropologist to analyze an individual of Mexican heritage as Swedish or West African; due in part to the approximately 56.5 million Hispanic individuals currently living in the United States, a group comprising the second largest behind American Whites (Flores 2017). Researchers have begun to test the applicability of macromorphoscopic analysis on Hispanic populations in addition to other groups from Asia and South Africa (Birkby et al. 2008; L'Abbé et al. 2011; Hughes et al. 2011; Hurst 2012; Ratliff 2014; Hefner 2015). In order to refine trait studies and improve the current methods used to identify unknown remains, a wide scope of human variation must be incorporated. A

more in-depth discussion of this work will continue later in the chapter.

### *Metric Approaches*

Metric analyses in biodistance studies followed a similar path of development to nonmetric analyses. The transition from purely descriptive reports on populations to statistically-based and theoretically-grounded investigations was gradual and reflected the overall change within the field of physical anthropology. Metric analyses include both cranial and postcranial elements with standardized recording methods. Modern methods for each are presented in Buikstra and Ubelaker's (1994) *Standards for Data Collection from Human Skeletal Remains* and the updated Langley et al.'s 2016 *Data Collection Procedures for Forensic Skeletal Material 2.0*.

Blumenbach, considered by some to be the father of physical anthropology, was interested in skull morphology and its application to geographically-based racial classes he termed *Caucasoids, Mongolians, Ethiopians, Malayans, and Americans*. Although he saw the other races as derivatives of the ideal European type, he still recognized regional variation and gradation with his "transitional races" of the Australian Aborigines and Malays (Wolpoff & Caspari 1997). Later, Samuel Morton, an American typologist, investigated regional variation between Amerindian populations under the assumption that skeletal morphology was the result of biological relatedness and could be reflected metrically (Buikstra 2006). Morton's (1839) *Crania Americana* was still one of the first works to note geography as a covariate; although he did not account for environmental effects on skeletal plasticity, and thus missed some of the factors acting on cranial morphology. In the 1900s, Hooton statistically investigated regional variation using natural breaks within the metric and indicial data of Pueblo Indians and Canary Islanders. His work incorporated temporal data to answer a variety of archaeological questions,

in particular those regarding the peopling of America (Buikstra 2006).

Giles and Elliott (1962) were the first to apply discriminant analyses to understanding cranial differences between American subpopulations after what was once purely a descriptive approach. Using eight cranial measurements from the Terry and Todd collections, they created discriminant functions to classify crania into the three broad “racial” categories of *American Negro*, *White*, and *Indian* (their *terminology*) with a high degree of accuracy when separated by sex (Giles & Elliott 1962). The measurements primarily highlighted overall group size and shape differences, which may explain why they were useful for sex discrimination as well. However, all measurements must be available for the method to work appropriately, therefore limiting its applicability in cases of incomplete remains. Birkby (1966) used this approach with a more robust American Indian subsample and emphasized the importance of using functions representative of the investigatory populations and that inappropriate applications may lead to erroneous results.

Howells’ 1973 *Cranial Variation in Man*, however, is most often referenced in craniometric literature because of his refinement of landmarks (and measurements) first standardized by the works of Martin, Vallois, Broca and others, along with the assignment of three-letter codes to cranial measurements (Hefner et al. 2016). Howell presents 16 multivariate discriminate functions for 20 un-pooled populations developed using 70 cranial measurements. The measurements reflect overall cranial size and shape differences and were analyzed at both multivariate and principal factor levels. These methods were (and continue to be) successful at distinguishing geographic clusters, predicting group affiliation, and understanding the structure of prehistoric populations (Zegura 1971; Howells 1973; Droessler 1981; Key 1983).

Cranial measurements have been the traditional source of biological data used by

physical anthropologists to analyze the geographic origin of individuals. This is in part due to the relatively early, and easy, implementation of classification statistics. Today, the primary method used for metric ancestry analysis is the computer program FORDISC 3.1 (Jantz & Ousley 2005). Using a combination of modern and historic, community-supplied data from populations around the world, analysts can input their own cranial and postcranial measurements to aid in the assessment of an unknown individual's ancestry via made-to-order discriminant functions (Ousley & Jantz 2012; Spradley & Jantz 2011); reporting Fisher's linear discriminant analysis as discriminant scores for two groups and the canonical variate scores for three or more groups.

Unlike Giles and Elliott's equations, any available measurements can be used to generate these custom classifications with the further option of program-selected best fit models. While a powerful tool, it is only as good as the data entered— both by the user and within the database— and therefore must not be viewed as an infallible magical program, but rather another tool in the skeletal analyst's repertoire (Ousley & Jantz 2012; Jantz & Ousley 2013; Dudzik & Kolarowicz 2016; Dudzik & Jantz 2016).

### *Contemporary Research*

Biodistance studies shifted from categorizing populations to investigating genetic-relatedness with more recent advances and applications in bioarcheological and group admixture research. One such advancement was the application of multivariate statistics, such as Mahalanobis' distance ( $D^2$ ) and Smith's Mean Measure of Divergence (MMD) for metric and nonmetric analyses, respectively, to investigate the relationship of populations using several variables at once. These statistics expanded the depth of information gained from skeletal remains to include inter- and intrapopulation differences on a variety of scales (Konigsberg 2006). Multivariate approaches permit the simultaneous analysis of the relationship of multiple

variables from one or more groups and provide a good proxy to measure overall group relatedness (Larsen 1997) and are “exceptionally well suited for investigating interrelationships among the variables, examining group differences, and making other inferences of the variables and groups selected” (Pietrusewsky 2000: 490).

Advances in human population genetics have also presented new avenues for comparing craniometric classification results. Relethford (1994) noted similar patterns between genetic and craniometric data and comparable levels of inheritance supporting the use of phenotypic data as proxies for genetic affiliation. Saunders (1989) provided a thorough review of postcranial nonmetric skeletal variation studies including discussions on issues of data collection methods and the appropriateness of two theoretical frameworks—genetic hypotheses and biological population models— for understanding biodistance study results. Her discussions expand this burgeoning field of research to include the biological and environmental causes behind traits (as opposed to just being the result of them) and their applicability to physical anthropological research (Saunders 1989).

Geographic spatial analyses are now commonly combined with skeletal biodistance studies to help visualize group-relatedness, provided the data and study scale are appropriate. Trait distribution and continuity levels as assessed by spatial autocorrelation techniques can be addressed numerically and graphically in several ways depending on the type of questions asked of the data. Quantitative methods such as Moran’s *I* measure the level of spatial autocorrelation and is a variation on Pearson’s correlation coefficient, a measure of the relatedness between two variables often used in physical anthropology (Legendre & Legendre 1998). Tests of both Global Moran’s and Local Indicators of Spatial Association (LISA) can be conducted to explore the intrinsic variation of a sample set and later addressed in subsequent analyses (Anselin 1995).

Relethford (2008) provides the quintessential marriage of craniometric and geographic data by performing a geostatistical analysis on body and craniofacial size for almost 200 populations throughout Ireland. The methods include a combination of variogram analyses and kriging to assess the pattern of spatial variation of skeletal attributes across the island nation. While the results show an apparent cline in craniofacial size across the country, the results must be understood under historical and biological frameworks (in this case the effects of the Viking conquest). Smaller scale studies can also benefit from these methods as Hefner (2012) showed in his spatial analysis of burials excavated from the historic Alameda-Stone cemetery in Tucson, AZ. This analysis detected both spatial and biological grouping of individuals buried in the cemetery as visualized through variograms and cluster analyses, respectively. The applicability of spatial analyses at several scales bodes well for further application to physical and forensic anthropology research questions.

### **Regional Investigations**

Several studies have used biodistance data in conjunction with biological and historical evidence to answer broad-anthropological questions regarding populations associated with the Greater Southwest. One such was Salzano and Callegari-Jacque's (1988) case study on the peopling of South America. Therein, archaeological and ethnographic records were combined with biological variability data (such as cranial morphology and craniometrics) from numerous populations across the continent. Using data from sources over 50 years, the genetic variability of the continent was analyzed using geographic and adaptive lenses. In sum, the variation of indigenous South Americans should be viewed as gradual clines rather than distinct categories with much more research necessary to fully understand the biological relationships of these groups (Salzano & Callegari-Jacque 1988).

A more contemporary study by Willermet et al. (2013) used dental traits to detect levels of biological affinity in ancient Central American remains and corroborated the relationship between observed morphological variations and group relatedness as seen in the historical and biological record. The authors recognized the inefficiency of applying just one theoretical model to the data, lest they lost sight of the interaction of several variables behind the observed skeletal morphology. Therefore, by framing the results (Mahalanobis distance values) in three models—cultural group, geographic model, and temporal period— a fuller depth of information about the populations could be observed. Then, the data can be applied to modern Mexican and Central American populations to understand levels of admixture and regional variation more holistically.

Hispanics, or individuals with admixed Latin American, European, and African ancestry, have been the focus of many recent biodistance studies in forensic anthropology. This is in part due to their growing majority in United States demographics and the ongoing migrant death crisis along the United States and Mexico border. Several researchers have undertaken the necessary steps to build a working knowledge base for skeletal attributes of these populations. Both metric-based (Spradley 2014) and morphological (Hefner et al. 2015) approaches have been developed to classify known and unknown Hispanic individuals, even with the potential problem of multiple countries of origin included when using the broad demographic term of Hispanic (e.g. Guatemala, Mexico, etc.).

By expanding upon the known power of craniometric analyses to differentiate between ‘traditional’ ancestry groups, research comparing craniometric data from undocumented border crossers (UBCs) and Hispanic subpopulations has been conducted to look for differences between groups found throughout Latin America (Spradley et al. 2008; Spradley 2013; Spradley 2014). Using cranial and postcranial metric data from UBCs examined at both the PCOME and



Texas State University-San Marcos, the researchers set out to develop sectioning points and classification functions for Hispanic populations, and to investigate the variation within and between samples of American Whites, Blacks, and Hispanics using both traditional measurements and geometric morphometric methods. Separation of test populations consisting of individuals recovered in Arizona and Texas and known Mexican and Guatemalan nationals was detectable using Mahalanobis distance measures, in addition to the possibility of predicting group affiliation for the UBC samples. For example, the Arizona UBC individuals were most similar to the identified Mexican sample, which likely reflects the known preponderance of individuals entering Arizona being from Mexico (Spradley 2014).

Nonmetric traits have also proven useful for discriminating Hispanic populations. Birkby et al. (2008) presented a suite of cranial nonmetric traits (both epigenetic and morphological) observed more frequently in individuals of Southwest Hispanic ancestry examined at the Pima County Office of the Medical Examiner (PCOME). While more informative than investigatory, the report provided the foundation for future trait studies by providing the most useful traits for identifying Southwest Hispanics. In addition, the authors discuss the concept of the cultural profile, or the combination of geographic, cultural, and physical characteristics associated with a set of remains that distinguishes it as belonging to that of a UBC; the most common subgroup of Hispanic individuals examined at their office. Hurst (2012) took a closer examination of the eight Birkby et al. (2008) traits and found that several were useful for distinguishing European and African Americans from Hispanics analyzed at the PCOME, thus supporting their continued use during forensic analysis. Exclusively-macromorphoscopic trait studies have also begun to incorporate Hispanic populations into their datasets. While large-scale separation between American Black, White, and pooled-Hispanic sample populations proved most powerful, Hefner

et al. (2015) were able to correctly classify the Guatemalan and SW Hispanic subgroups using macromorphoscopic traits, albeit at a lesser rate than craniometric analyses. Further investigation into Random Forest Modeling (RFM), however, has shown the strength of combining both morphoscopic and metric analyses of ancestry classification which result in higher classification rates than when the methods are used separately via discriminant function analysis (Hefner et al. 2014). This technique accounts for within and between group variation, “reducing misclassification rates, and capturing aspects of cranial shape, size, and morphology” (Hefner et al. 2014: 583).

In all, the need for increased sample sizes (in both overall numbers and sub-groups) must drive future studies so that more areas of Latin America are available from which to compare unknown individuals. Multiple lines of evidence, including data types and analytic methods, must be investigated as to better understand human variation and the factors behind geographically associated skeletal markers of ancestry.

## **Conclusions**

The appropriateness of incorporating skeletal analysis, of both metric and nonmetric traits, to assess the ancestral origin of an individual from a biodistance approach has been established. It is within this theoretical and practical framework that the biological analysis portion of this study will be conducted using the methods outlined in the next chapter.

## REFERENCES

## REFERENCES

- Anselin L. 1995. Local indicators of spatial association—LISA. *Geographical Analysis*, 27(2): 93-115.
- Berry RJ, Berry AC. 1967. Epigenetic variation in the human cranium. *Journal of Anatomy*, 101: 361-379.
- Birkby WH. 1966. An evaluation of race and sex identification from cranial measurements. *American Journal of Physical Anthropology*, 24(1): 21-27.
- Birkby WH, Fenton TW, Anderson BE. 2008. Identifying southwest Hispanics using nonmetric traits and the cultural profile. *Journal of Forensic Sciences*, 53(1): 29–33.
- Brues AM. 1990. The once and future diagnosis of race. In: *Skeletal Attribution of Race: Methods for Forensic Anthropology*, Gill GW and Rhine S (eds.). Albuquerque: Maxwell Museum of Anthropology, pp. 1-9.
- Buikstra JE. 2006. History of research in skeletal biology. In: *Handbook of North American Indians: Environment, Origins, and Population*, Vol. 3, DH Ubelaker and WC Sturtevant, (eds.). Washington DC: Smithsonian Institution, pp. 504-523.
- Buikstra JE, Ubelaker DH. 1994. Standards for data collection from human skeletal remains: proceedings of a seminar at the Field Museum of Natural History, organized by Jonathan Haas. Fayetteville: Arkansas Archeological Survey.
- Cabana GS, Clark JJ. 2011. *Rethinking Anthropological Perspectives on Migration*. Gainesville: University Press of Florida.
- Cheverud JM, Buikstra JE. 1981. Quantitative genetics of skeletal nonmetric traits in the rhesus macaques on Cayo Santiago. II. Phenotypic, genetic, and environmental correlations between traits. *American Journal of Physical Anthropology*, 54(1): 51-58.
- Corruccini RS. 1975. Multivariate analysis in biological anthropology: some considerations. *Journal of Human Evolution*, 4: 1–19.
- Droessler J. 1981. *Craniometry and Biological Distance: Biocultural Continuity and Change at the Late-Woodland Mississippian Interface*. Center for American Archaeology, Evanston.
- Dudzik B, Jantz, RL. 2016. Misclassifications of hispanics using Fordisc 3.1: Comparing cranial morphology in Asian and hispanic populations. *Journal of Forensic Sciences*, 61(5): 1311-1318.

- Dudzik B, Kolatorowicz A. 2016. Craniometric Data Analysis and Estimation of Biodistance. In: Biological Distance Analysis, Pilloud MA, Hefner JT (eds.). San Diego: Academic Press. pp. 35-60.
- Finnegan M, McGuire SA. 1979. Classification systems for discrete variables used in forensic anthropology. *American Journal of Physical Anthropology*, 51(4):5 47-53.
- Flores A. 2017. Statistical portrait of Hispanics in the United States. Pew Research Center Hispanic Trends, 18 September; <http://www.pewhispanic.org/2017/09/18/facts-on-u-s-latinos/#hispanic-rising-share>
- Giles E, Elliot O. 1962. Race identification from cranial measurement. *Journal of Forensic Sciences*, 7: 147-151.
- Hanihara T. 2008. Morphological variation of major human populations based on nonmetric dental traits. *American Journal of Physical Anthropology*, 136(2): 169-182.
- Hauser, G, De Stefano GF. 1989. Epigenetic variants of the human skull. Stuttgart, Germany: E. Schweizerbart'sche Verlagsbuchhandlung.
- Hefner JT. 2007. The statistical determination of ancestry using cranial nonmetric traits. Doctoral Dissertation, Gainesville: The University of Florida.
- Hefner JT. 2009. Cranial nonmetric variation and estimating ancestry. *Journal of Forensic Sciences*, 54(5): 985-995.
- Hefner JT. 2012. Biological distance and geospatial analysis. In: *Skeletal Biology of the Alameda-Stone Cemetery, Vol. II*. Hefner JT, Keur MA, Heilen MP (eds.). Statistical Research, Inc. Technical Series, No. 10-04, pp. 4.1-4.23.
- Hefner JT, Ousley SD. 2006. Morphoscopic traits and the statistical determination of ancestry II. Proceedings of the 58th Annual Meeting of the American Academy of Forensic Sciences; Feb 20–25; Seattle, WA. Colorado Springs, CO: American Academy of Forensic Sciences.
- Hefner JT, Pilloud MA, Black CJ, Anderson BE. 2015. Morphoscopic trait expression in “hispanic” populations. *Journal of Forensic Sciences*, 60(5): 1135-1139.
- Hefner JT, Pilloud MA, Buikstra JE, Vogelsberg CCM. 2016. A brief history of biological distance analysis. In: *Biological Distance Analysis*, Pilloud MA, Hefner JT (eds.). San Diego: Academic Press. pp. 3-22.
- Hefner JT, Spradley MK, Anderson B. 2014. Ancestry assessment using random forest modeling. *Journal of Forensic Sciences*, 59(3): 583-589.

- Howells WW. 1973. Cranial variation in man. Papers of the Peabody Museum of Archaeology and Ethnology. Vol. 67, Cambridge: Harvard University.
- Hughes CE, Juarez CA, Hughes TL, Galloway A, Fowler G, Chacon S. 2011. A simulation for exploring the effects of the “trait list” method’s subjectivity on consistency and accuracy of ancestry estimations. *Journal of Forensic Sciences*, 56(5): 1094-1106.
- Hurst CV. 2012. Morphoscopic trait expressions used to identify southwest Hispanics. *Journal of Forensic Sciences*, 57(4): 859–865.
- Jantz RL, Ousley SD. 2013. Introduction to Fordisc 3. In: *Forensic Anthropology: An Introduction*, Tersigni Tarrant MA, Shirley NR (eds.). Boca Raton: CRC Press. pp.253-269.
- Key PJ.1983. *Craniometric Relationships Among Plains Indians: Culture-Historical and Evolutionary Implications*. Report of Investigations. University of Tennessee
- Klales AR, Kenyhercz MW. 2015. Morphological assessment of ancestry using cranial macromorphoscopsics. *Journal of Forensic Sciences*, 60(1): 13-20.
- Konigsberg LW. 1990. Analysis of prehistoric biological variation under a model of isolation by geographic and temporal distance. *Human Biology*, 62(1): 49-70.
- Konigsberg LW. 2006. A post-Neumann history of biological and genetic distance studies in bioarchaeology. In: *Bioarchaeology: The Contextual Analysis of Human Remains*, Buikstra JE, Beck LA (eds.). Amsterdam: Academic Press Elsevier, pp. 263-280.
- L’Abbé, EN, Van Rooyen C, Nawrocki SP, Becker PJ. 2011. An evaluation of non-metric cranial traits used to estimate ancestry in a South African sample. *Forensic Sciences International*, 209(1-3): 195.e1-7.
- Langley NR, Jantz LM, Ousley SD, Jantz RL, Milner G. 2016. *Data Collection Procedures for Forensic Skeletal Material 2.0*. University of Tennessee and Lincoln Memorial University.
- Larsen CS. 1997. Biological distance and historical dimensions of skeletal variation. In: *Bioarchaeology: Interpreting Behavior from the Human Skeleton*. Larsen CS (ed). Cambridge: Cambridge University Press, pp. 357-401.
- Laughlin WS, Jørgensen JB. 1956. Isolate variation in Greenlandic Eskimo crania. *Human Heredity*, 6(1): 3-12.
- Legendre P, Legendre L. 1998. Spatial analysis. In: *Numerical ecology: Developments in environmental modeling* Vol. 20. Elsevier, pp. 707-785.

- Ousley SD, Jantz RL. 2012. "Fordisc 3.0 and Statistical Methods for Estimating Sex and Ancestry." In: Dirkmaat DC (ed.), *A Companion to Forensic Anthropology*. Chichester: Wiley Blackwell. pp. 311-329.
- Ossenberg NS. 1976. Within and between race distances in population studies based on discrete traits of the human skull. *American Journal of Physical Anthropology*, 45: 701-716.
- Pietrusewsky MI. 2000. Metric analysis of skeletal remains: methods and applications. In: *Biological Anthropology of the Human Skeleton*, Katzenberg MA and Saunders SR, (eds.). New York: Wiley-Liss, pp. 375-415.
- Pink CM, Maier C, Pilloud MA, Hefner JT. 2016. Cranial nonmetric and morphoscopic data sets. In: *Biological Distance Analysis*. Pilloud MA, Hefner JT (eds.). San Diego: Academic Press. pp. 91-107.
- Plemons AM, Hefner JT. 2016. Ancestry estimation using macromorphoscopic traits. *Academic Forensic Pathology*, 6: 400-412.
- Ratliff MD. 2014. Evaluating morphoscopic trait frequencies of Southeast Asians and Pacific Islanders. Doctoral dissertation. University of Montana.
- Relethford JH. 1994. Craniometric variation among modern human populations. *American Journal of Physical Anthropology*, 95(1): 53-62.
- Relethford JH. 2001. Global analysis of regional differences in craniometric diversity and population substructure. *Human Biology*, 73: 629-636.
- Relethford JH. 2002. Apportionment of global human genetic diversity based on craniometrics and skin color. *American Journal of Physical Anthropology*, 118: 393-398.
- Relethford JH. 2008. Geostatistics and spatial analysis in biological anthropology. *American Journal of Physical Anthropology*, 136(1): 1-10.
- Relethford JH, Blangero J. 1990. Detection of differential gene flow from patterns of quantitative variation. *Human Biology*, 62(1): 5-26.
- Relethford JH, Lees F. 1982. The use of quantitative traits in the study of human population structure. *Yearbook of Physical Anthropology*, 25:113-132.
- Salzano FM, Callegari-Jacques SM. 1988. South American Indians: A case study in evolution. *Research Monographs of Human Population Biology*, No. 6. Oxford: Clarendon.
- Saunders SR. 1989. Nonmetric skeletal variation. In: *Reconstruction of Life from the Skeleton*, İşcan MY, Kennedy KAR (eds.). New York: Liss. pp. 95-108.

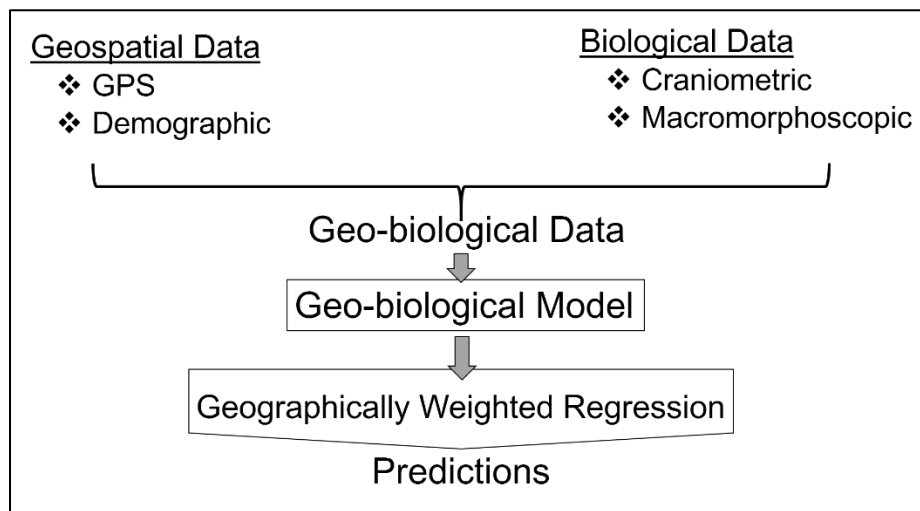
- Scott GR, Turner CG. 2000. *The Anthropology of Modern Human Teeth: Dental Morphology and Its Variation in Recent Human Populations*. Vol. 20. Cambridge University Press, pp. 254-259.
- Spradley MK, Jantz RL, Robinson A, Peccerelli F. 2008. Demographic change and forensic identification: problems in metric identification of Hispanic skeletons. *Journal of Forensic Sciences*, 53(1): 21-28.
- Spradley MK. 2013. *Project IDENTIFICATION: Developing Accurate Identification Criteria for Hispanics*. US Department of Justice.
- Spradley MK. 2014. Toward Estimating Geographic Origin of Migrant Remains Along the United States-Mexico Border. *Annals of Anthropological Practice*, 38(1): 101–110.
- Spradley MK, Jantz RL. 2011. Sex estimation in forensic anthropology: skull versus postcranial elements. *Journal of Forensic Sciences*, 56: 289–296.
- Tobler WR. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(1): 234-240.
- Washburn SL. 1951. Section of anthropology: The new physical anthropology. *Transactions of the New York Academy of Sciences*, 13(7): 298-304.
- Willermet C, Edgar HJ, Ragsdale C, Aubry BS. 2013. Biodistances among Mexican, Maya, Toltec, and Totonac groups of central and coastal Mexico. *Chungara: Revista de Antropología Chilena*, 45(3): 447-459.
- Wolpoff MH, Caspari R. 1997. *Race and human evolution*. New York: Simon and Schuster.
- Zegura, SL. 1971. *A Multivariate Analysis of the Inter- and Intra-Population Variation Exhibited by Eskimo Crania*. Doctoral dissertation, University of Wisconsin, Madison.



## CHAPTER 5. METHODS & MATERIALS

### Methods Introduction

The present research project followed the methods outlined below to investigate the relationship between a person's demographic information (primarily, country of origin) and the recovery location of their remains in the southern Arizona desert using geospatial and biological data sets apply it to new recoveries. The geospatial data includes demographic information and the GPS coordinates of an individual's recovery location; and the biological data includes both cranial measurements and macromorphoscopic indicators of ancestry. Together, these data are now referred to as geobiological data. These geobiological data were combined to form the associated geobiological model and run through a geographically weighted regression analysis (GWR) to predict the country of origin of newly recovered remains (Figure 5.1).



**Figure 5.1. Flow-chart of project materials and methods.**

Random Forest models were also explored to find more information on both the metric and nonmetric traits which most differentiate Hispanic individuals from other populations

commonly investigated in American forensic laboratories. The data derived from biological and geospatial sources and a description of each type are presented later in this chapter.

## **Biological Methods**

Cranial measurements and macromorphoscopic indicators of ancestry were collected from identified and unidentified individuals investigated by the PCOME. In addition, data collected by other researchers from populations situated near the U.S.–Mexico border (e.g. Mexican nationals and indigenous groups) were also compiled (Spradley 2013). These data were used to assess inter-and-intra regional population differences in cranial morphology for the identified individuals, and for the application to newly developed models for assessing the ancestry of unknown individuals using probabilistic methods.

Biological data collection and analysis was subdivided by data type (craniometric vs. macromorphoscopic) using their respective methods and the results combined for more robust analyses which will be explained in greater detail as follows. The craniometric and macromorphoscopic data analysis results from the Hispanic individuals possessing both data types were applied to the subsequent geobiological model.

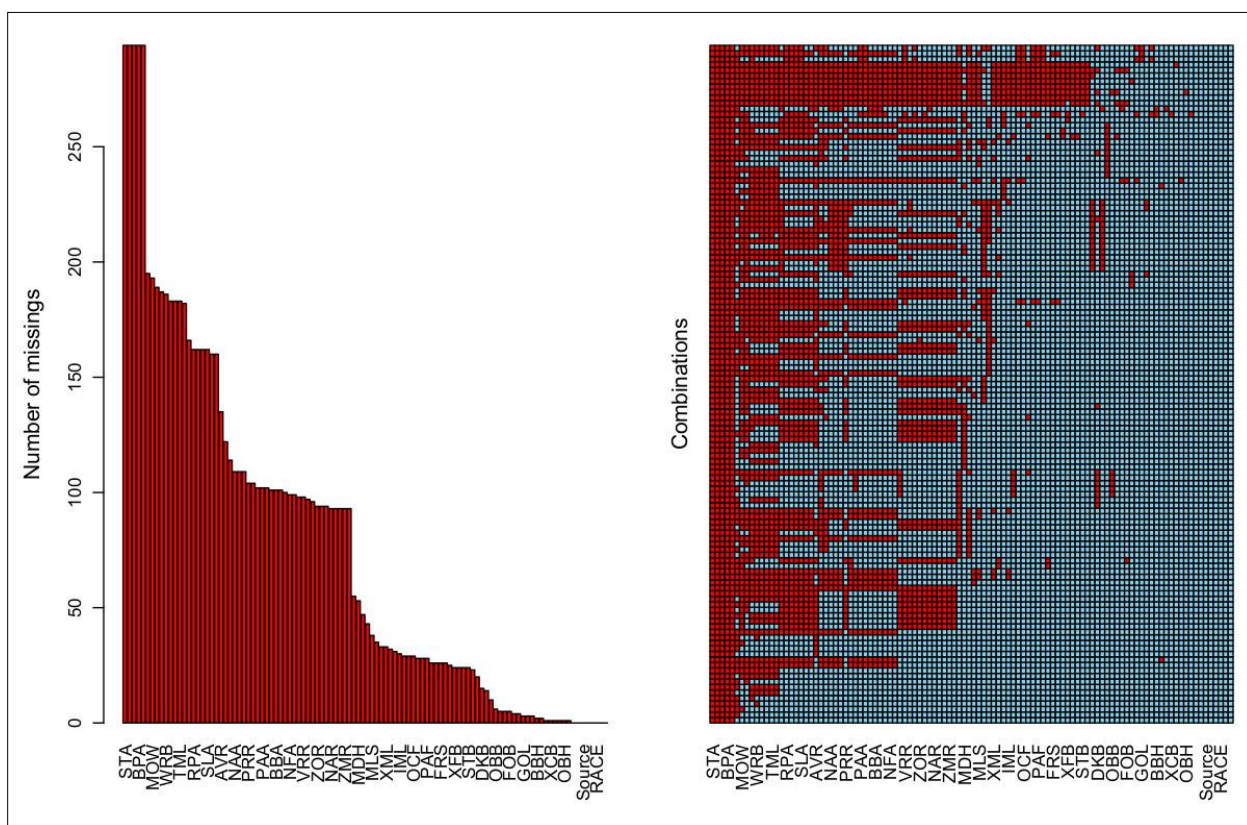
### *Craniometric Data*

All novel cranial data collection followed standard protocols established by previous work on these populations to maintain consistency and allow for data pooling. This included the collection of eighty standard cranial landmarks using a Microscribe® G2X digitizer and processed using the program 3Skull (Ousley 2004) to calculate inter-landmark distances and to store the data (linear and coordinate). These data included both standard and novel craniometric data defined by Howells (1973) (See Table 5.1).

**Table 5.1. Howells cranial measurements and abbreviations (Spradley 2013)**

<b>Abbreviation</b>	<b>Measurement</b>	<b>Abbreviation</b>	<b>Measurement</b>
GOL	Glabella-Occipital	PAF	Parietal Fraction
NOL	Nasion-Occipital	OCC	Occipital Chord
BNL	Basion-Nasion	OCS	Occipital Subtense
BBH	Basion-Bregma	OCF	Occipital Fraction
WFB	Minimum Frontal Breadth	FOL	Foramen Magnum Length
XCB	Max Cran Breadth	FOB	Foramen Magnum Breadth
XFB	Max Frontal Breadth	NAR	Nasion Radius
ZYB	Bizygomatic Breadth	SSR	Subspinale Radius
AUB	Biauricular Breadth	PRR	Prosthion Radius
ASB	Biasterionic Breadth	DKR	Dacryon Radius
BPL	Basion-Prosthion Length	ZOR	Zygoorbitale Radius
NPH	Nasion-Prosthion Height	FMR	Frontomolare Radius
NLH	Nasal Height	EKR	Ectoconchion Radius
JUB	Bijugal Breadth	ZMR	Zygomaxillare Radius
NLB	Nasal Breadth	AVR	M1 Alveolar Radius
MAB	External Palate Breadth	BRR	Bregma Radius
MAL	External Palate Length	VRR	Vertex Radius
MDH	Mastoid Height	LAR	Lambda Radius
OBH	Orbital Height	OSR	Opisthion Radius
OBB	Orbital Breadth	BAR	Basion Radius
DKB	Interorbital Breadth	NAA	Nasion Angle
NDS	Nasion-Dacryon Subtense	PRA	Prosthion Angle
WNB	Simotic Chord	BAA	Basion Angle, Nasion-Prosthion
SIS	Simotic Subtense	NBA	Nasion Angle
ZMB	Bimaxillary Breadth	BBA	Basion Angle, Nasion-Bregma
SSS	Zygo-Maxillary Subtense	BRA	Bregma Angle
FMB	Bifrontal Breadth	SSA	Zygomaxillary Angle
NAS	Nasio-Frontal Subtense	NFA	Nasio-Frontal Angle
EKB	Bi-Orbital Breadth	DKA	Dacryal Angle
DKS	Dacryon Subtense	NDA	O-Dacryal Angle
IML	Inferior Malar Length	SIA	Simiotic Angle
XML	Maximum Malar Length	FRA	Frontal Angle
MLS	Malar Subtense	PAA	Parietal Angle
WMH	Minimum Malar Height	OCA	Occipital Angle
UFBR	Upper Facial Breadth	GLS	Glabella Projection
PAS	Parietal Subtense	RPA	Radio-Parietal Angle
STB	Bistephanic Breadth	ROA	Radio-Occipital Angle,
FRC	Frontal Chord	BSA	Basal Angle,
FRS	Frontal Subtense	SBA	Sub-Bregma Angle
FRF	Frontal Fraction	SLA	Sub-Lambda Angle
PAC	Parietal Chord		
RFA	Radio-Frontal Angle		

Variations in data collection methods (digital versus manual) and incomplete remains have led to missing data within the data sets. The R package VIMGUI was used to visualize and impute missing data using the k-Nearest Neighbor (kNN) algorithm (as in Figure 5.2). Previous research into the imputation of craniometric data has supported the use of  $k = 5$  neighbors in cases of 50% missing craniometric data (Kenyhercz & Passalacqua 2016). Prior to imputation, all variables missing 100 or more values ( $\sim 50\%$  of the sample set) were removed from further analysis. This left 73 craniometric variables for analysis, presented in Table 5.2.



**Figure 5.2. VIMGUI visualization of missing craniometric data, sorted from most to least.**

**Table 5.2. Cranial measurements used in initial principal components analyses**

Measurements			
GOL	FRS	OBB	LAR
NOL	FRF	DKB	OSR
BNL	PAC	NDS	BAR
BBH	PAS	WNB	NAA
XCB	PAF	SIS	PRA
XFB	OCC	ZMB	BAA
WFB	OCS	SSS	NBA
ZYB	OCF	FMB	BBA
AUB	FOL	NAS	BRA
ASB	FOB	EKB	SSA
BPL	NAR	DKS	NFA
NPH	SSR	IML	DKA
NLH	PRR	XML	FRA
JUB	DKR	MLS	PAA
NLB	ZOR	WMH	OCA
MAB	FMR	GLS	UFBR
MAL	EKR	STB	
MDH	ZMR	FRC	
OBH	BRR	VRR	

To minimize any effect of size on population differences, the data were analyzed in both standardized and unstandardized formats. Standardization and all other statistical analyses were conducted using the program STATISTICA (TIBCO 2017). The data were then checked for outliers and data from countries with fewer than 5 individuals were removed from analysis, including: Costa Rica, Honduras, Jamaica, Peru, and the United States.

Principal components analyses (PCAs) were conducted on both standardized and unstandardized data, as well as pooled and un-pooled country-of-origin groups at multiple levels to find the best model. Pooled groups included American Blacks, American Whites, and Hispanics, and un-pooled included Mexico, Guatemala, El Salvador, American Blacks, and American Whites. First, a factor analysis of the 26 FORDISC 3.1 (Ousley & Jantz 2012)

measurements were conducted. The 15 variables with the highest factor loadings were selected and a principal components analysis was completed. From those 15, another factor analysis was conducted to identify the next 12 variables with the highest factor loadings and a PCA was run with just those selected. The same process was repeated to find the 4 highest loading variables for the last model-selection PCA analysis. In total, 18 PCAs were conducted: 9 on the standardized data groups and 9 on the unstandardized data. A table outlining the analyses is presented in Table 5.3. A model using 12 variables on standardized data was chosen for the final analyses based on Eigenvalues and variable factor loadings of the aforementioned PCAs. These variables are presented in Table 5.4.

**Table 5.3. Group properties of craniometrics principal components analyses**

<b>Unstandardized Data</b>	<b>Pooling</b>	<b>Grouping</b>	<b>Standardized Data</b>	<b>Pooling</b>	<b>Grouping</b>
4 Variables	Un-pooled	Hispanics only	4 Variables	Un-pooled	Hispanics only
		All groups			All groups
	Pooled	All groups		Pooled	All groups
12 Variables	Un-pooled	Hispanics only	12 Variables	Un-pooled	Hispanics only
		All groups			All groups
	Pooled	All groups		Pooled	All groups
15 Variables	Un-pooled	Hispanics only	15 Variables	Un-pooled	Hispanics only
		All groups			All groups
	Pooled	All groups		Pooled	All groups

**Table 5.4. Measurements used in final principal components analyses**

<b>Abbreviation</b>	<b>Measurement</b>
GOL	glabella-occipital
NOL	nasion-occipital
BNL	basion-nasion
WFB	minimum frontal breadth
ZYB	byzygomatic breadth
AUB	biauricular breadth

<b>Abbreviation</b>	<b>Measurement</b>
BPL	basion-prosthion length
JUB	bijugal breadth
MAB	external palate breadth
ZMB	bimaxillary breadth
EKB	bi-orbital breadth
UFBR	upper facial breadth

The resulting PCA values maximize group differences and were used as a data reduction technique for the application of the craniometric data to the geobiological model. Only ancestry and country of origin were investigated since data standardization removes size differences generally associated with sex. Finally, a basic factor analysis by ancestral group and country of origin for each case was performed. This method explores the variables which have the greatest effect on differentiating between the groups. The resulting principal component scores for Hispanic samples analyzed at the PCOME were used in the subsequent geobiological model creation as explained below.

#### *Macromorphoscopic Data*

Collecting macromorphoscopic data is another aspect of best practices for the assessment of ancestry, so these techniques were also incorporated into the present project (SWGANTH 2013). Macromorphoscopic data were collected using the software program MMS (v1.6.1) (Hefner 2016) with reference data from the Macromorphoscopic Databank (MaMD, Hefner 2018). MMS provides detailed descriptions and trait expression illustrations for 17 macromorphoscopic traits of the cranium (Table 5.5). Due to their high correlations and low inter-observer error rates, 6 traits were used in the subsequent analyses and are designated by an asterisk (\*) (Hefner 2009; Kamnikar et al. 2018).

**Table 5.5. Macromorphoscopic trait names and abbreviations**

<b>Trait Name</b>	<b>Abbreviation</b>
Anterior Nasal Spine*	ANS*
Inferior Nasal Aperture*	INA*
Interorbital Breadth*	IOB*
Malar Tubercle	MT
Nasal Aperture Shape	NAS
Nasal Aperture Width*	NAW*
Nasal Bone Contour*	NBC*
Nasal Bone Shape	NBS
Nasal Overgrowth	NO
Nasofrontal Suture	NFS
Orbital Shape	OS
Palate Shape	PS
Postbregmatic Depression*	PBD*
Posterior Zygomatic Tubercle	PZT
Supranasal Suture	SS
Transverse Palatine Suture	TPS
Zygomaxillary Suture	ZS

Methods for the analysis of the macromorphoscopic data involved performing canonical analyses of the principal coordinates (CAP) with the R package BiodiversityR (Kindt & Coe 2005). The CAP method transforms categorical variables into continuous variables using an unconstrained ordination procedure, while also calculating principal coordinate (PCOa) values using one of several measures of similarity/dissimilarity. These values are used in a traditional canonical analysis (either regression or classification) to visualize group differences (Hefner 2016). The CAP method can also assign unknown individuals to a group through discriminant function analyses, as the CAP transformations relieves the traditional violations committed when using DFAs on nonmetric traits (Hefner 2016).

In the BiodiversityR package, the user can specify the  $m$  values, distance measures, and ultimately plot the discriminant function results following the ordination procedures. The  $m$



value is the number of axes analyzed by the discriminant analysis and the R program can choose the appropriate number to maximize the classification rate (Kindt & Coe 2005; Hefner 2007). The distance measures are indices of dissimilarity in which “the ultimate goal is to arrive at clusters of objects which display *small within-cluster* variation, but *large between-cluster* variation” (Kachigan 1991: 262). Several distance measures are made available in BiodiversityR and for this project the Mahalanobis and Alternate Gower distances were tested.

The Mahalanobis distance (or  $D^2$ ) is commonly used in ancestry analyses using craniometric data. It considers the intercorrelation of traits in the interpretation of group similarities, with smaller distances indicated more group similarity (Spradley 2006). The alternate Gower distance (Anderson et al. 2006) is a modification to the traditional Gower (1971) measure which can handle nominal, ordinal, and binary data but in its modified form “omits double-zeros and divides by the number of pairs with at least one above-zero value, and does not scale columns” (Kindt & Coe 2005). This makes it is useful for the current MMS dataset with both binary and categorical scores.

Previous research using CAP procedures indicate the Chi-square Distance as the most effective for correctly classifying individuals into ancestral groups using macromorphoscopic data (Hefner 2007); however, this measure was unavailable in the present R package. Therefore, this project also tests the effectiveness of these alternative distance methods on a similar data set. The resulting principal coordinate (PCOa) scores for Hispanic samples analyzed at the PCOME were applied to the geobiological model creation as explained later.

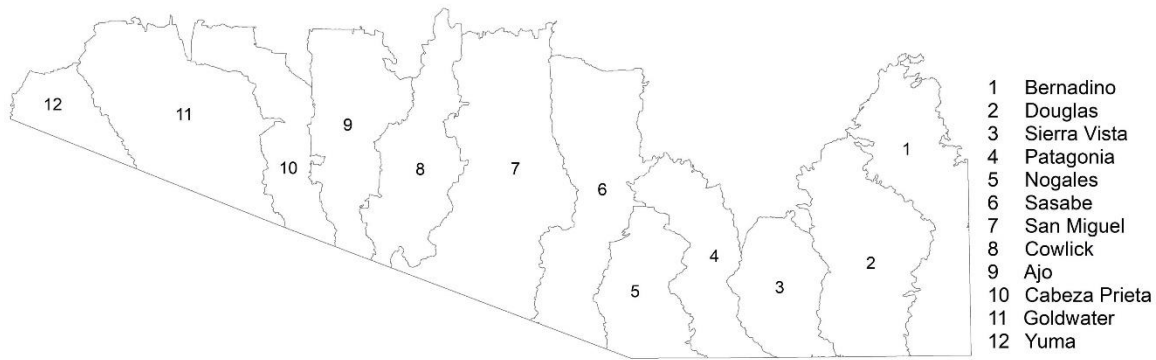
### *Random Forest Models*

Finally, since improved classification rates were reported using combined craniometric and macromorphoscopic data, random forest models (RFMs) using both datasets were created to

classify the sample by ancestry and country of origin using the R package Rattle (Williams 2011). RFMs also do not require data normalization and imputation of missing data is built into the package (Williams 2011). Additionally, the RFMs can identify which variables (both metric and nonmetric) are contributing the most to each model (Hefner et al. 2014). Partitioning of the data set is not necessary in the model creation process and it is not advised for small sample sizes, so the groups were pooled into broad ancestry groups (Williams 2011). However, due to the fundamentals of forest creation and the implementation of out-of-bag (OBB) prediction, this is not worrisome. OBB estimates are a type of bootstrapping which applies trees to the data that were not used in building that tree for the creation of accurate error rates (Williams 2011).

### **Geospatial Methods**

The geospatial and demographic data collected from the Humane Borders website and PCOME casefiles were analyzed using several methods with the programs ArcGIS®, GeoDa (Anselin et al. 2006), and RStudio (RStudio Team 2016). The programs produce similar but not exact results, as although they perform the same equations, the methods for computation and presentation differ slightly (Kekez 2015). The appropriate platform was chosen depending on the task performed and the visual and analytical capabilities needed for the analysis. In all analyses, the UTM X and Y values were used. The covariables of year of recovery, sex, country of origin, and recovery corridor were coded as necessary for investigation in this research project. Recovery corridors were designed and labeled based on watersheds and mountain ranges between the border region and northern urban areas (Chamblee 2018, personal communication). A map identifying these corridors is presented in Figure 5.3.



**Figure 5.3. Corridors designations used in geospatial analyses.**

Caution is used when describing the physical location of cases. While it is presumed this is where the individual died, both anecdotal and analytical data show that remains, especially skeletonized, can be moved from their original death location; due to several taphonomic processes including scattering by animals and humans, as well as environmental-related transport like floods and storms. Additionally, one individual may be represented by several case numbers due to postmortem scattering, and only reconciled when either genomic or geospatial investigations are conducted. When multiple cases are determined to be from the same individual, the initial case number (and thus oldest) is used. This explains why the number of cases reported may change from one reporting period to the next, as cases from more recent years become absorbed into older cases. This information is included in the quarterly updates performed on the PCOME and Humane Borders mapping data. The phrase “recovery location” is preferred and used over “location of death.”

*Exploratory spatial data analysis (ESDA)*

Methods of exploratory spatial data analysis (ESDA) were employed to begin the investigation into the spatial relationship of recovered individuals. These methods closely follow traditional exploratory analyses, while also incorporating factors of spatial dependency when

looking at the distributions and patterns in the data. In essence, exploratory analyses are data-driven by approaching the process with few preconceived notions, theories, or hypotheses about the relationship of the data.

Pilot studies have indicated positive spatial autocorrelation in these data, so more nuanced spatial autocorrelation analyses were deemed appropriate. These include the Global Moran's I and Local Indicator of Spatial Association (LISA) cluster maps. The Global Moran's I statistic ranges from -1 to 1 and measures the spatial autocorrelation of a single variable. Herein, the covariance is divided by the maximum-likelihood estimator of the variance with values greater than 0 indicating positive autocorrelation and those less than 0 reflecting negative autocorrelation. Since Moran's I summarizes the entire area using a single statistic, spatial homogeneity is assumed (Legendre & Legendre 1998). As this is presumably not the case in our current sample set spanning thousands of miles, LISA cluster maps were also created using GeoDa (Anselin et al. 2006). A LISA statistic "gives an indication of the extent of significant spatial clustering of similar values around that observation [and] the sum of LISAs for all observations is proportional to a global indicator spatial association" (Anselin 1995: 94). By using this statistic and its associated cluster maps, it is possible to visually assess and statistically test for local levels of variation including spatial clusters, or "hotspots", and areas of local instability (Anselin 1995). The resulting maps color-code the data by the type of spatial autocorrelation present (i.e. positive, negative, or none).

Prior to analysis, a spatial weights matrix ( $W$ ) was created to measure the relative location of all the data points to each other, and provide a template for further diagnostic and analytical testing. These matrixes "express the neighbor structure between observations as a  $n \times n$  matrix,  $W$ , in which the elements of the matrix are the spatial weights" (Anselin & Rey 2014:

35). The spatial weights “ $\{w_{ij}\}$ , are any set of constants that specify which  $j$  subareas in the study area have variate values directly spatially related with  $X_i$ ” (Cliff & Ord 1981: 8). The weights are then used to compute the spatial lag for an observation via the weighted average of neighboring locations (Anselin 2005). Nearest neighbor algorithms are used in regression and predictive models such as these where the points around a given data point effect its predicted group classification; as those that are closer to it contribute more than those that are further away. Conventional nearest neighbor values ( $k$ ) are calculated by taking the square root of the sample size (Diggle 1979; Duda et al. 2001). In this project, a distance-based spatial weights matrix was created using the  $k$ -nearest neighbor of 41.

Although the use of  $k$ -nearest neighbor calculations for the spatial weights matrix rely on neighbor relations as opposed to distance, it is possible that the relatively large area under investigation in this study is still minimizing the actual amounts of spatial autocorrelation present. Therefore, a spatial weight using a grid-based joins count was also investigated with the minimum nearest neighbor distance was calculated (approximately 13 km) to estimate the size of the grid cells. Then, a theoretical grid was placed over the study area, with its cells being used to define new point relations via counting both the number of points per cell and assessing their proximity to other points (both within their own cell and between cells). Large areas of empty cells (those without cases) were dissolved into larger cells to minimize the effects of distance, and the relative amount of each case variable found within the cell was calculated. These include the ratio of unidentified individuals to identified, females to males, and individuals from Non-Mexican countries to those from Mexico. This transformation minimizes the heterogeneity of the study area as seen in regions of high and low recovery densities and allows for consistent analysis. These ratios were the basis of the grid cell autocorrelation analyses.

## *Geographically Weighted Regression*

Following data exploration, geographically weighted regression (GWR) analyses were conducted. GWR performs like standard regression methods with the added capabilities of describing spatial variations in the input data (Grose et al. 2008). It also allows for the prediction of new locations of data or the dependent attribute of a data point with a known location. It is the latter application of GWR that is important in this project to predict the country of origin of an unknown deceased migrant based on their skeletal attributes and where they were recovered.

The process first requires the development of global linear models and several were investigated. The best performing model for predicting country of origin was created by combining the first two factor scores of the craniometric PCAs and the two PCOa scores derived from the macromorphoscopic CAP; now termed the geobiological model.

Geographically weighted regression (GWR) analyses were run on the chosen global geobiological model using the R packages *spgwr* (Williams 2011) and *GWmodel* (Gollini et al. 2015). For each analysis, the optimal bandwidth value was calculated, and an adaptive Gaussian kernel was applied. An adaptive kernel means the bandwidth distances will change according to the spatial density of the points, and a nearest neighbors' approach was applied as opposed to a fixed distance. This is useful in cases such as this where there are areas of more and less dense data points (Fotheringham et al. 2002). Each program contains functions for applying a geographically weight regression to a previously-developed global model and for predicting unknown samples. Gollini et al. 2015 provides an in-depth discussion at the functional differences between the two packages in their description of *GWmodel*.

Lastly, GWR testing was completed on the geobiological model in ArcMap (ESRI 2016) by placing a 50 km hexagonal grid over the study area. This approach may then be preferable for

situations with less accurate GPS data as there is a 50 km “buffer” built into the model.

Diagnostic tests were run using this platform to compare to the results generated from the R packages.

### **Assumptions and Limitations**

Several assumptions must be addressed regarding the nature of the geospatial analyses and their application to this research project. Issues of collinearity of the data must be taken into account in the geographically weighted regression analyses. Collinearity occurs when the independent variables “have a high degree of correlation, or are close to exhibiting a deterministic linear relationship” (Brunsdon et al. 2012: 1). The PCA and PCOA values applied to the OLS model are correlated in that together they express the variation seen in the crania of individuals from different ancestral groups, and therefore may pose a problem. However, most issues of collinearity involve the precision of the coefficient outputs, and as the ultimate predictive (or dependent variable) output is of interest here, limiting the effects of it may not be necessary (Brunsdon et al. 2012). Tests for collinearity will still be conducted, as well as spatial patterning of the model’s residuals, to understand the appropriateness of the final model.

Finally, statistics that test for spatial autocorrelation “assume stationarity, meaning that the underlying process should have at least roughly the same parameter values (mean and variance) for the entire study area... and assume that the spatial covariance structure of the variable (i.e., the values of spatial autocorrelation at different spatial distances or lags) is similar over the entire study area” (Wagner & Fortin 2005: 1979 -80). Due to the large geographic area under investigation (roughly 27,000mi<sup>2</sup>) and long time-period (approximately 16 years), spatial and temporal heterogeneity is assumed. Yet, the fact that the identified dataset consists of

predominately young Mexican males may produce an issue. Diagnostic tests of normality and the removal of outliers will help address the distribution of the dataset but may also diminish the ability to look at a wide variety of individuals. Although a large proportion of the migrant dataset are Mexican males, the need to predict the origin of individuals from other countries is high. The removal of outliers, presumably non-Mexican males, will limit the applicability of the model; at least until a larger sample set of individuals from other countries is available. Currently, the limited number of countries and small overall sample size presents a cautionary scenario for the application of the geobiological model. Individuals entering the United States come from many different countries and regions for which more robust sample sizes may not currently be available, but identifications are ongoing and thus reference samples continue to grow. At the moment regional specificity may not be possible, but general country or region is sufficient for directing identification efforts.

## **Materials Introduction**

The present research uses three data types to investigate the relationship between the demographic information and the recovery location of undocumented border crosser (UBC) remains in the southern Arizona desert: 1) biological, 2) geospatial, and 3) geobiological. While complementary, each data type is derived from several sources and the basis of several related research questions.

## **Biological Data**

The biological data used in this project consist of cranial measurements and macromorphoscopic indicators of ancestry from both identified and unidentified individuals investigated at the PCOME. In addition, biological data from other populations situated near the



US–Mexico border (e.g. regional indigenous groups) were obtained as reference samples for ancestral groups similar to future individuals potentially recovered by the PCOME.

### *Craniometric Dataset*

Dr. Spradley and the NIH-funded Project IDENTIFICATION began the skeletal reference database of Hispanic individuals processed at the PCOME, the Forensic Anthropology Center at Texas State (FACTS), as well as two documented cemeteries from Mexico. These data are made available for research via the Forensic Databank (FDB) and through direct data-requests (Ousley & Jantz 1998; Spradley 2013). The Forensic Databank is a resource compiled by forensic anthropologists in part for the “continuous testing and revision of standards for identification to reflect biological changes in recent populations” (Jantz & Moore-Jansen 1988: 5-6). These Hispanic data were the craniometric reference samples used in this project and acted as a template for the newly collected craniometric data taken at the PCOME. Additionally, a modern American Black and American White dataset from the Forensic Databank was also used to test the discriminating power of the craniometric data within and between the groups. The demographic distribution of the craniometric sample set is presented in Tables 5.6 and 5.7.

**Table 5.6. Population group distribution of craniometric dataset**

<b>Population</b>	<b>Female</b>	<b>Male</b>	<b>Total</b>
<b>Hispanic</b>	40	245	284
<b>American Black</b>	126	191	317
<b>American White</b>	289	452	741
<b>Grand Total</b>	455	888	1343

**Table 5.7. Country of origin distribution of craniometric dataset**

<b>Country of origin</b>	<b>Female</b>	<b>Male</b>	<b>Total</b>
<b>United States</b>	415	643	1058
<b>Mexico</b>	19	214	232
<b>Guatemala</b>	16	24	40
<b>El Salvador</b>	5	7	12
<b>Grand Total</b>	455	888	1343

*Macromorphoscopic Dataset*

Macromorphoscopic indicators of ancestry are also routinely collected during case analyses at the PCOME using MMS v.1.6.1 (Plemons & Hefner 2016). The scores from identified individuals analyzed between 2007 and 2017 were obtained upon request from the Macromorphoscopic Databank and the PCOME (MaMD, Plemons & Hefner 2016). As with the craniometric dataset, macromorphoscopic data from contemporary American Black and White individuals were also used to explore the within and between group differences of the Hispanic population.

The MaMD categorizes the data by ancestral background in a 4-scheme hierarchical method. This includes a broad ‘ancestry’ classification, and a more specific ‘geo-origin’ group (Hefner 2018). The ancestry group is typographical in nature with grouping based on broad geographical regions. This is further refined to geo-origin which “corresponds to a specific geographical location of birth (i.e., country) ...referring only to place of birth (when known) or sample origin when region of birth is not provided” (Hefner 2018: 4).

The sex and ancestry for the identified cases with MMS data used in this project is presented in Table 5.8 while the geo-origin distribution of each ancestral group by sex is broken down in Table 5.9. Individuals with missing data (primarily scores for post-bregmatic depression) were removed.

**Table 5.8. Ancestral group distribution of the macromorphoscopic dataset**

<b>Ancestry</b>	<b>Female</b>	<b>Male</b>	<b>Total</b>
African	116	176	292
European	157	128	285
Hispanic	103	265	368
<b>Grand Total</b>	<b>376</b>	<b>569</b>	<b>945</b>

**Table 5.9. Geo-origin distribution of the macromorphoscopic dataset by data source**

<b>Ancestry</b>	<b>Data source</b>	<b>Geo-Origin</b>	<b>Female</b>	<b>Male</b>	<b>Total</b>
<b>African</b>	Terry Collection	American Black	116	176	292
<b>European</b>	Terry Collection	American White	157	128	285
<b>Hispanic</b>	Multiple	Colombia	66	168	234
		Guatemala	28	65	93
		SW Hispanic	3	8	11
	PCOME	Mexico	3	20	23
		Guatemala	3	1	4
		Honduras	0	2	2
		El Salvador	0	1	1
<b>Grand Total</b>			<b>376</b>	<b>569</b>	<b>945</b>

*Random Forest Model Dataset*

Since improved classification rates were reported using a combination of craniometric and macromorphoscopic data (Hefner et al. 2014), random forest models (RFMs) will be created. The data included in these methods derived from the aforementioned biological datasets, as well as an outside source (Hefner et al. 2014). Both the broad ancestral grouping and country of origin of the individuals were investigated. A breakdown of the demographics of the individuals used in the RFM section is presented in Table 5.10.

**Table 5.10. Distribution of dataset used in random forest modeling**

<b>Ancestry</b>	<b>Country of Origin</b>	<b>Female</b>	<b>Male</b>	<b>Total</b>
<b>American Black</b>	United States	6	31	37
<b>American White</b>	United States	23	45	68
<b>Hispanic</b>	Mexico	3	28	31
	Guatemala	3	2	5
	El Salvador	0	2	2
	Honduras	0	1	1
	Unknown	2	19	21
<b>Grand Total</b>		37	128	165

### **Geospatial Data**

The geospatial data for this project came from the Arizona OpenGIS Initiative for Deceased Migrants ([www.humaneborders.info](http://www.humaneborders.info)) sponsored by the nonprofit group Humane Borders (Chamblee et al. 2006). The website provides GIS-based tools to grant the public access to free, high-quality, downloadable spatial data on all migrant deaths analyzed by the Pima and Maricopa County Office of the Medical Examiner since 2001 using interactive mapping tools. Only data from cases analyzed by the PCOME was used in the current project.

The website provides three mapping tools using the free Google Map platform in both English and Spanish. The tool most pertinent to this project, the “Map of Migrant Mortality,” enables the user to perform a variety of searches, including that for single or groups of cases with fine-tuning parameters for the variables of sex, year of death, and cause of death. In cases of incomplete or skeletonized remains, cause of death was reported as either “Skeletal Remains” or “Undetermined.” Otherwise, the PCOME-determined cause of death is reported, if known.

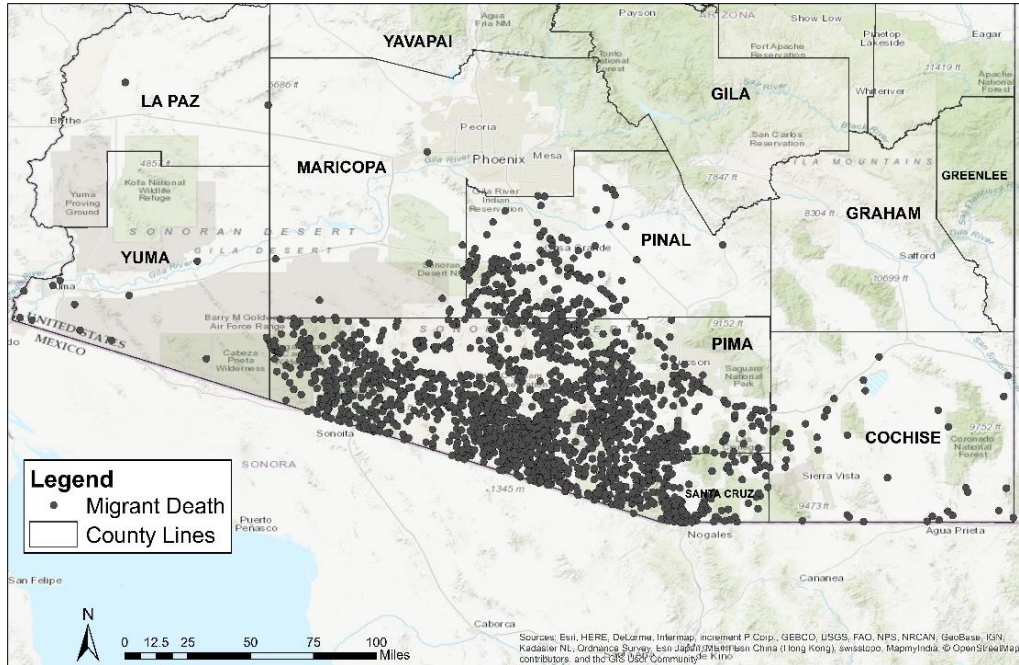
For each search, cases are plotted with the relevant data for each tabulated below the map. The user is also able to obtain basic case information when individuals are selected on the map. These data include, but are not limited to: case number, sex, age, reporting date (when found), and recovery location data in GPS coordinates (UTMs), latitude and longitude, and

descriptive formats. UTM coordinates are the primary geographic data used in the subsequent geospatial analyses.

Many cases have imprecise GPS data, including many completed prior to the implementation of standardized collection and record-keeping procedures. Therefore, it was advised to only use cases with known UTM coordinates from GPS data or addresses; those with only vague physical descriptions or otherwise non-exact data for recovery location were dropped. Map data are updated from PCOME case records regularly, so data for this project were limited to January 2001 to January 2nd, 2017; totaling 2,616 PCOME cases (1,723 identified and 893 unidentified). However, after the removal of cases with unverified UTMs, the final project sample totaled  $N = 1,731$ , with  $n = 1,051$  identified and 680 unidentified.

Due to the public nature of the OGIS platform, some case information is withheld, including country of origin and the condition of the body when it was recovered. As these data may be useful in understanding where and when people crossed, they were obtained from PCOME records when available.

Figure 5.4 displays the location of all 2,616 deceased migrants recovered during this project period (including those with estimated geolocations).



**Figure 5.4. Exact and approximate locations of all recovered UBC remains from 2001 January 2017 (N = 2,616).**

Figure 5.5 displays the location of all 1,731 deceased migrants with verified UTM coordinates used in this study; while Figure 5.6 differentiates between identified and unidentified (UID) individuals.

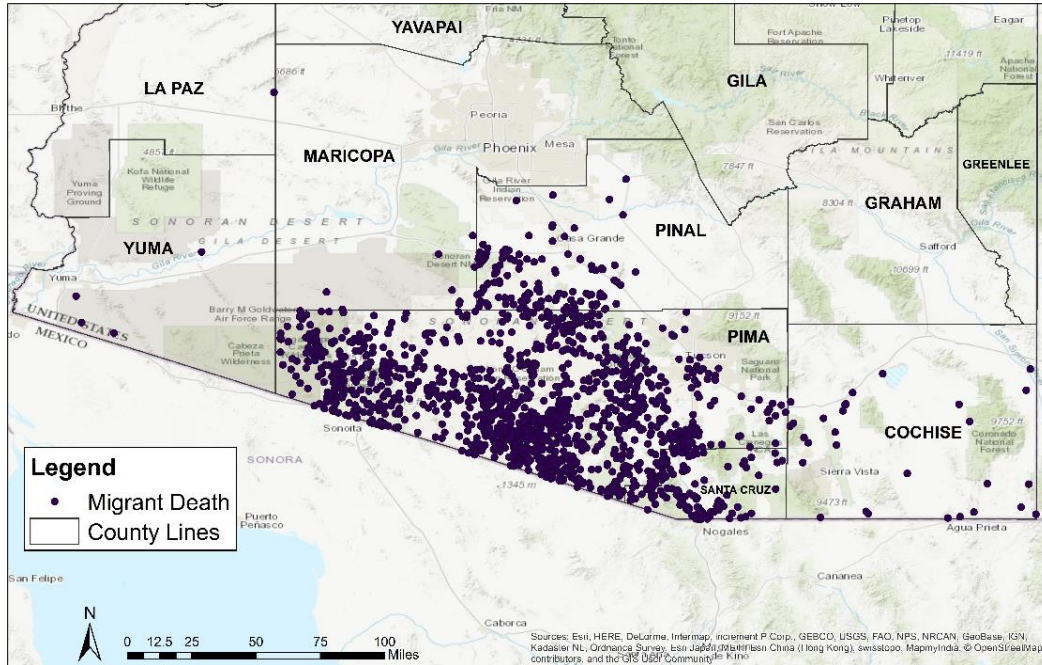


Figure 5.5. Locations of all recovered UBCs with accurate data used in project (N = 1,723).

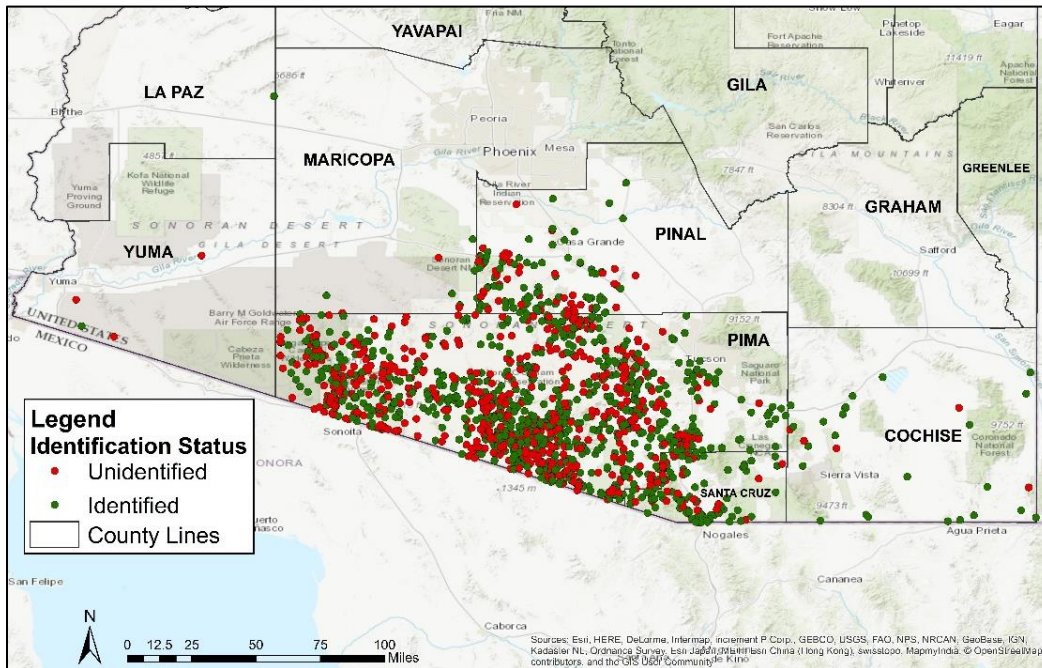
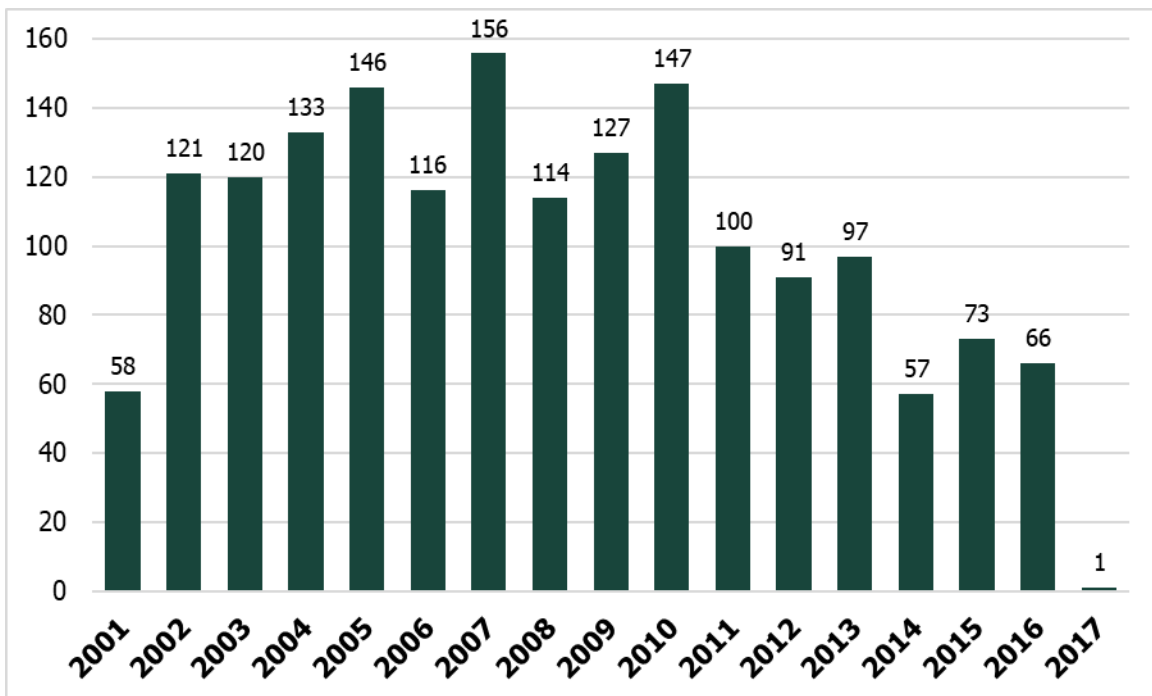


Figure 5.6. Locations of UBC remains with accurate data coded by identification status (ID'd = 1,051) and (UID = 680).

The yearly distributions of deceased UBCs are presented in Figure 5.7. A note should be made that the year of recovery (as presented here) and the year of death are not necessarily the same. Often, individuals die in remote regions or are left behind by other members of their parties, not to be found for days, weeks, or even years. Therefore, the date of recovery may be some time following the individual’s actual death. Cases also often consist of only partial skeletal remains, with complete recovery impossible or drawn out over time, which may lead to subsequent remains coded as additional cases. As individuals are re-associated (usually via DNA or geographic-proximity analyses), their reports are aggregated and assigned the earliest case number. This explains any discrepancies between total and yearly number of cases presented by the PCOME in their annual reports from year to year (PCOME 2016).



**Figure 5.7. Number of recoveries per year for all recovered remains (N = 2,616).**



Table 5.11 presents the demographic distribution of the sex and country of origin for all identified UBC cases analyzed by the PCOME during this project’s timeframe (N= 1,723), and Table 5.12 presents the demographics of the individuals used in the exploratory spatial data analysis of this research project (n = 1,051).

**Table 5.11. Distribution of identified UBC remains by sex and country of origin**

<b>Country of Origin</b>	<b>Female</b>	<b>Male</b>	<b>Total</b>
<b>Mexico</b>	227	1207	1434
<b>Guatemala</b>	46	120	165
<b>El Salvador</b>	12	37	49
<b>Honduras</b>	3	34	37
<b>Ecuador</b>	5	6	11
<b>Peru</b>	4	3	7
<b>Brazil</b>	1	3	4
<b>Colombia</b>	1	2	3
<b>Costa Rica</b>	0	3	3
<b>Dominican Republic</b>	2	0	2
<b>Chile</b>	0	1	1
<b>Jamaica</b>	0	1	1
<b>Nicaragua</b>	1	0	1
<b>Venezuela</b>	0	1	1
<b>Unknown*</b>	0	3	3
<b>Grand Total</b>	302	1421	1723

**\*Fetus**

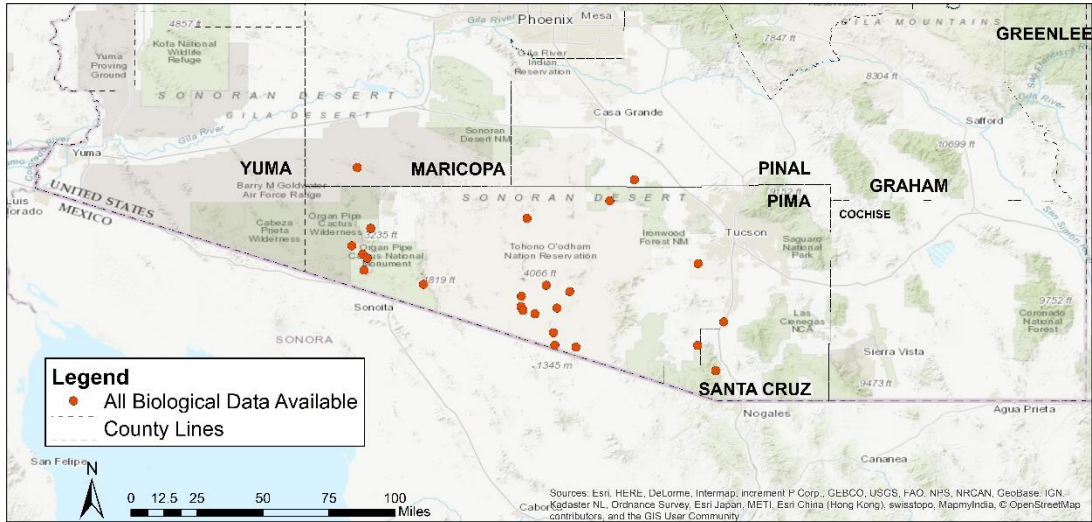
**Table 5.12. Demographic distribution of identified individuals in ESDA**

<b>Country of Origin</b>	<b>Female</b>	<b>Male</b>	<b>Total</b>
<b>Mexico</b>	130	746	876
<b>Guatemala</b>	28	73	101
<b>El Salvador</b>	9	24	33
<b>Honduras</b>	1	23	24
<b>Ecuador</b>	3	4	7
<b>Peru</b>	1	2	3
<b>Colombia</b>	1	1	2
<b>Costa Rica</b>	0	1	1
<b>Dominican Republic</b>	1	0	1
<b>Jamaica</b>	0	1	1
<b>Nicaragua</b>	1	0	1
<b>Unknown*</b>	0	1	1
<b>Grand Total</b>	175	876	1051

**\*Fetus**

### **Geobiological Data**

Preliminary analyses of the spatial properties of the geospatial data have shown significant positive autocorrelation between sex, year of recovery, and the location remains were recovered (Vogelsberg 2018). The development of a model to predict the country of origin of unknown remains requires the confluence of both the geographic and biological data derived from these separate investigations. Therefore, a third, study-derived set of data designated as geobiological was investigated to establish the relationship between location of death and ancestral background. These data are the products of the biological analyses combined with the geospatial data and were used for the model building discussed in previously. Figure 5.8 presents a map of the identified PCOME individuals who had biological and geospatial data available, regardless of the accuracy of the GPS data.



**Figure 5.8. Identified individuals with geobiological data available.**

The sex and country of origin distribution for the individuals with geobiological data (n = 27) is provided in Table 5.13.

**Table 5.13. Demographic information of identified individuals with geobiological data**

Country of Origin	Female	Male	Total
Mexico	3	18	21
Guatemala	3	1	4
El Salvador	0	2	2
<b>Grand Total</b>	<b>6</b>	<b>21</b>	<b>27</b>

Of these, only 24 had accurate GPS data required for geographically weighted regression analyses. The individuals were then pooled by sex, since measurement standardization during the biological data analyses accounted for any possible sexual dimorphism present.

## **Conclusions**

The methods and associated datasets described here will assist in the project goal of predicting the country of origin of newly recovered individuals and aid in the investigatory process for identifying undocumented border crossers. By exploring the geospatial and biological patterns of identified individuals, more about the unidentified can be understood.

## REFERENCES

## REFERENCES

- Anderson MJ, Ellingsen KE, McArdle BH. 2006. Multivariate dispersion as a measure of beta diversity. *Ecology Letters*, 9: 683–693.
- Anselin L. 1995. Local indicators of spatial association—LISA. *Geographical Analysis*, 27(2): 93-115.
- Anselin L, 2005. Spatial statistical modeling in a GIS environment. In: *GIS, Spatial Analysis, and Modeling*, Maguire DJ, Batty M, Goodchild MF (eds). ESR Press, pp.93-111.
- Anselin L, Syabri I. Kho Y. 2006. GeoDa: An introduction to spatial data analysis. *Geographical Analysis*, 38(1): 5-22.
- Anselin L, Rey SJ. 2014. *Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL*. GeoDa Press LLC.
- Brunsdon C Charlton M, Harris P. 2012. Living with Collinearity in Local Regression Models. In: *Proceedings of the 10th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Brazil*.
- Chamblee JF, Christopherson GL, Townley M, DeBorde D, Hoover R. 2006. Mapping migrant deaths in southern Arizona: The humane borders GIS. Unpublished report.
- Cliff AD, Ord JK. 1981. *Spatial autocorrelation*. London: Pion.
- Diggle PJ. 1979. Statistical methods for spatial point patterns in ecology. In: *Spatial and Temporal Analysis in Ecology*, RM Cormack and JK Ord (eds). Fairfield, MD: International Cooperative Pub. House. pp. 95-150.
- Duda RO, Hart PE, Stork DG. 2001. *Pattern classification*. 2nd Edition. New York: John Wiley & Sons.
- ESRI Inc. 2016. ArcMap 10.5.1 [Computer software]. Redlands, CA: Esri Inc.
- Gollini I, Lu B, Charlton M, Brunsdon B, Harris P. 2015. GWmodel: An R package for exploring spatial heterogeneity using Geographically Weighted Models. *Journal of Statistical Software*, 63(17): 1-50.
- Gower JC. 1971. A general coefficient of similarity and some of its properties. *Biometrics*, 1:857-871.
- Grose D, Brunsdon C, Harris R. 2008. *Introduction to Geographically Weighted Regression (GWR) and to Grid Enabled GWR*. Lancaster University.

- Hefner JT. 2007. The Statistical Determination of Ancestry Using Cranial Nonmetric Traits. Doctoral dissertation, University of Florida.
- Hefner JT. 2009. Cranial nonmetric variation and estimating ancestry. *Journal of Forensic Sciences*, 54(5): 985-995.
- Hefner JT, Spradley MK, Anderson B. 2014. Ancestry assessment using random forest modeling. *Journal of Forensic Sciences*, 59(3): 583-589.
- Hefner JT. 2016. Biological Distance Analysis, Cranial Morphoscopic Traits, and Ancestry Assessment in Forensic Anthropology. In: *Biological Distance Analysis*, Pilloud MA, Hefner JT (eds). San Diego: Academic Press. pp. 301-315.
- Hefner JT. 2018. The macromorphoscopic databank. *American Journal of Physical Anthropology*, 00: 1-11.
- Hefner JT, Spradley MK, Anderson B. 2014. Ancestry assessment using random forest modeling. *Journal of Forensic Sciences*, 59(3): 583-589.
- Howells WW. 1973. Cranial variation in man. *Papers of the Peabody Museum of Archaeology and Ethnology*. Vol 67. Cambridge, Mass.: Harvard University Press.
- Jantz RL, Moore-Jansen PH. 1988. A data base for forensic anthropology. Washington, DC USA: Final Report to the National Institute of Justice.
- Kachigan SK. 1991. *Multivariate statistical analysis: A conceptual introduction*. Radius Press.
- Kamnikar KR, Plemons AM, Hefner JT. 2018. Intraobserver Error in Macromorphoscopic Trait Data. *Journal of Forensic Sciences*, 63(2): 361-370.
- Kekez F. 2015. Clustering of Immigration Population in Helsinki Metropolitan Area, Finland: A Comparative Study of Exploratory Spatial Data Analysis Methods. Master's Thesis. Geography and Geoinformatics. University of Helsinki. Department of Geosciences and Geography. Division of Geography.
- Kenyhercz MW, Passalacqua NV. 2016. Missing data imputation methods and their performance with biodistance analyses. In: *Biological Distance Analysis*, Pilloud MA, Hefner JT (eds). San Diego: Academic Press. pp. 181-194.
- Kindt R, Coe R. 2005. *Tree diversity analysis. A manual and software for common statistical methods for ecological and biodiversity studies*. World Agroforestry Centre (ICRAF), Nairobi, Kenya.
- Legendre P, Legendre L. 1998. Spatial analysis. In: *Numerical Ecology: Developments in Environmental Modeling Vol. 20*. Elsevier, pp. 707-785.

- Ousley SD. 2004. 3skull [computer program]. Windows version 2.0.77;  
[http://www.mercyhurst.edu/departments/applied\\_forensic\\_sciences/faculty\\_staff.html](http://www.mercyhurst.edu/departments/applied_forensic_sciences/faculty_staff.html)
- Ousley SD, Jantz RL. 1998. The forensic data bank: documenting skeletal trends in the United States. In: *Forensic Osteology: Advances in the Identification of Human Remains*. Reichs KJ (ed), Springfield: Charles C. Thomas. pp. 441–458.
- Ousley SD, Jantz RL. 2012. Fordisc 3.0 and Statistical Methods for Estimating Sex and Ancestry. In: *A Companion to Forensic Anthropology*. Dirkmaat DC (ed), Chichester: Wiley Blackwell. pp. 311-329.
- Pima County Office of the Medical Examiner (PCOME). 2016. Annual Report. Tucson, Arizona.
- Plemons AM, Hefner JT. 2016. Ancestry estimation using macromorphoscopic traits. *Academic Forensic Pathology*, 6: 400-412.
- RStudio Team. 2016. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA.
- Scientific Working Group for Forensic Anthropology (SWGANTH). 2013. Ancestry Analysis. Published Document Accessed at  
[https://www.nist.gov/sites/default/files/documents/2018/03/13/swganth\\_ancestry\\_assessment.pdf](https://www.nist.gov/sites/default/files/documents/2018/03/13/swganth_ancestry_assessment.pdf)
- Spradley MK. 2006. Biological anthropological aspects of the African diaspora. Geographic Origins, Secular Trends, and Plastic Versus Genetic Influences Utilizing Craniometric Data Doctoral dissertation, University of Tennessee.
- Spradley MK. 2013. Project IDENTIFICATION: Developing Accurate Identification Criteria for Hispanics. US Department of Justice.
- TIBCO Software Inc. 2017. Statistica (data analysis software system), version 13.  
<http://statistica.io>.
- Vogelsberg CCM. 2018. Assessing the Spatial Patterns of Undocumented Border Crosser (UBC) Deaths in the Southern Arizona Desert. Proceedings of the 70<sup>th</sup> Annual Meeting of the American Academy of Forensic Sciences; Feb 19-24, Seattle, WA
- Wagner HH, Fortin MJ. 2005. Spatial analysis of landscapes: concepts and statistics. *Ecology*, 86(8): 1975-1987.
- Williams GJ. 2011. *Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery, Use R!*, New York: Springer.



## **CHAPTER 6. BIOLOGICAL DATA RESULTS**

### **Introduction**

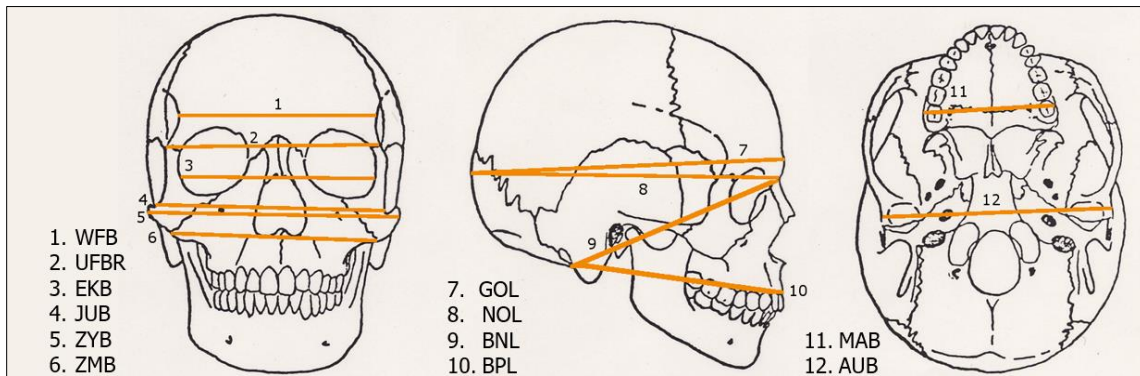
The biological data included craniometric measurements and macromorphoscopic traits from individuals from several populations including; identified undocumented border crossers, other groups with Hispanic origins, and populations commonly analyzed in American forensic anthropological casework. The purpose of the biological data analyses were to prepare the data for the application to the geobiological model. The results of these analyses will be presented here as they pertain the datatype (craniometric, macromorphoscopic, or combined) and interpreted as appropriate.

### **Craniometric Data Analysis**

Following the 18 principal component analyses (PCAs) (9 on the standardized data groups and 9 on the unstandardized data), a subset of 12 standardized measurement variables was chosen for the final analyses. Table 6.1 presents the abbreviations and names of these 12 variables while Figure 6.1 highlights the measurements on a skull line drawing.

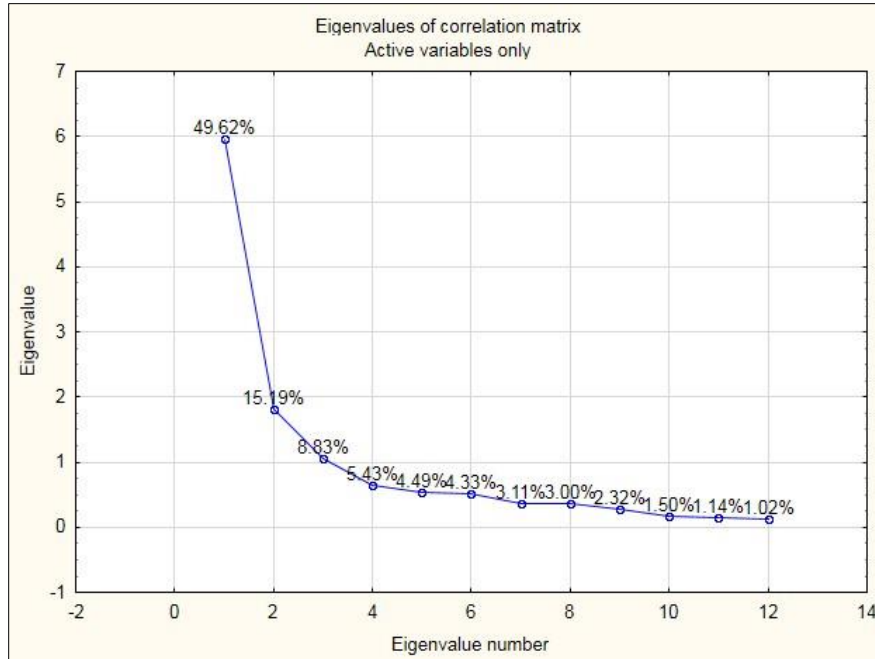
**Table 6.1. Measurement variables used in PCAs**

Abbreviation	Measurement
GOL	Glabella-Occipital
NOL	Nasion-Occipital
BNL	Basion-Nasion
WFB	Minimum Frontal Breadth
ZYB	Bizygomatic Breadth
AUB	Biauricular Breadth
BPL	Basion-Prosthion Length
JUB	Bijugal Breadth
MAB	External Palate Breadth
ZMB	Bimaxillary Breadth
EKB	Bi-Orbital Breadth
UFBR	Upper Facial Breadth



**Figure 6.1. Selected measurements for final PCA methods.**

The scree plot derived from the final PCA, as well as a table of the 12 associated factor eigenvalues is presented in Figure 6.2 and Table 6.2, respectively.



**Figure 6.2. Scree plot for 12 variable PCA run on standardized data.**

**Table 6.2. Factor scores from PCA results**

Factor	Eigenvalue	% Total	Cumulative %
1	5.954416	49.62013	49.6201
2	1.822715	15.18929	64.8094
3	1.059989	8.83324	73.6427
4	0.651974	5.43311	79.0758
5	0.539288	4.49407	83.5698
6	0.520142	4.33452	87.9044
7	0.373363	3.11136	91.0157
8	0.359835	2.99863	94.0143
9	0.278824	2.32354	96.3379
10	0.179586	1.49655	97.8344
11	0.137308	1.14423	98.9787
12	0.12256	1.02134	100

The scree plot graphically presents the eigenvalues of the derived factors documented in Table 6.2, and each eigenvalue “corresponds to the equivalent number of variables which the factor represents” (Kachigan 1991: 246). Here, the first factor accounts for as much variance as

nearly 6 of the original 12 variables (nearly 50% of the variation), while the second accounts for almost two variables (1.82, or 15% of the variation) (Table 6.2).

The number of factors used to evaluate a model is chosen by the point where difference in the amount of variance explained by each subsequent factor becomes less and less relative. Traditionally, this is visually assessed as the factor at which the crook or “elbow” in the scree plot occurs, or when the eigenvalues are lower than one. In this case, the first 3 factors explain roughly 73.64% of the total variance as described by these 12 cranial measurements (Figure 6.2).

The first 3 factor loadings of each variable are presented in Table 6.3. These values highlight which variables are most important in the variability between the three groups tested (Hispanics, American Whites, and American Blacks).

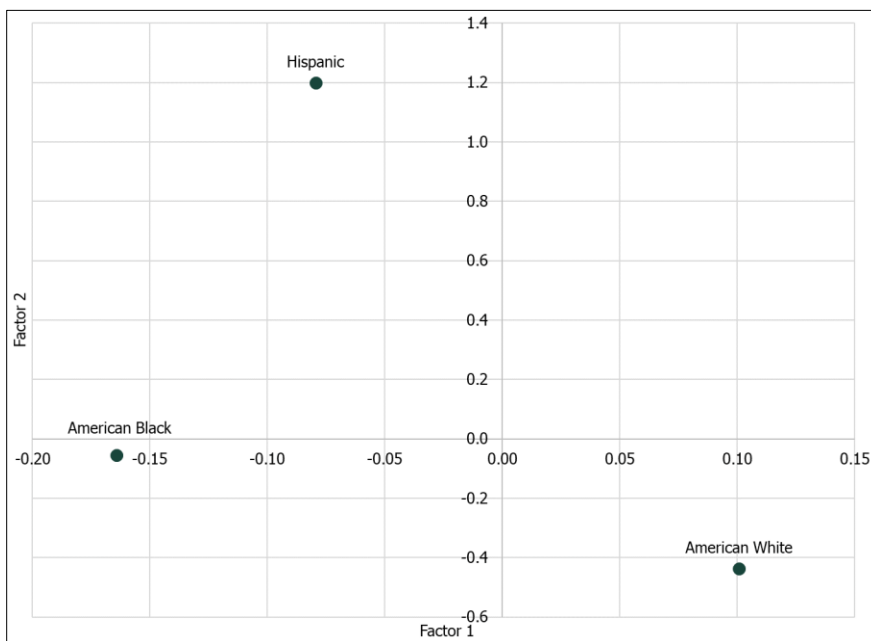
**Table 6.3. Variable loadings of first 3 factors**

<b>Variable</b>	<b>Factor 1</b>	<b>Factor 2</b>	<b>Factor 3</b>
<b>GOL</b>	-0.634	<b>-0.666</b>	0.008
<b>NOL</b>	-0.621	<b>-0.674</b>	-0.016
<b>BNL</b>	-0.653	<b>-0.480</b>	0.201
<b>WFB</b>	-0.680	-0.123	<b>-0.424</b>
<b>ZYB</b>	<b>-0.837</b>	0.273	-0.146
<b>AUB</b>	<b>-0.728</b>	0.293	-0.261
<b>BPL</b>	-0.605	-0.103	<b>0.656</b>
<b>JUB</b>	-0.652	0.347	-0.080
<b>MAB</b>	-0.606	0.379	<b>0.483</b>
<b>ZMB</b>	-0.653	<b>0.487</b>	0.166
<b>EKB</b>	<b>-0.852</b>	0.085	-0.125
<b>UFBR</b>	<b>-0.858</b>	0.012	-0.193
<b>Variance Explained</b>	5.954	1.823	1.060
<b>% Total</b>	49.62%	15.19%	8.83%

**Red denotes highest loadings**

In these analyses, the measurements of upper facial breadth (UFBR), biorbital breadth (EKB), bizygomatic breadth (ZYB), and biauricular breadth (AUB) contribute most to factor 1, in that order. The large negative loadings of these variables indicate the measurements have a high contribution to the differences seen between the three groups. When referring to Figure 6.1, it is evident that these measurements heavily involve midfacial width and length; including interorbital breadth (IOB), facial breadth (UFBR), etc. The second factor loading is most affected by the length of the cranium (GOL, NOL, and BNL) as well as some midfacial width (ZMB). Lastly, while the measurements associated with factor 3 explain the least amount of the variance (only about 9%), those corresponding to facial width and length still play a part (WFB, BPL, and MAB).

A plot of the mean of the first two factor values for the three ancestral groups is presented in Figure 6.3 and the three groups separate well.



**Figure 6.3. Group centroids of first 2 factors pooled by ancestry.**

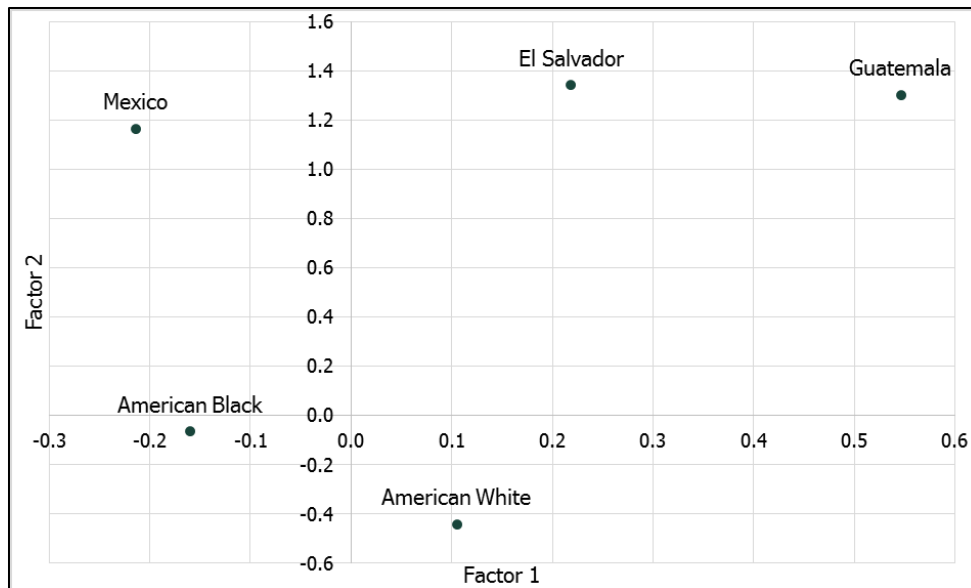
A factor analysis was also performed to see which measurements had the highest loadings for the top three PCA factors. The factors which best distinguished the Hispanic population from the other American groups were of primary focus and are presented in Table 6.4. These factor loadings mostly highlight measurements which involve the midface.

**Table 6.4. First 2 factor loadings for Hispanic samples**

<b>Variable</b>	<b>Factor 1</b>	<b>Factor2</b>
<b>GOL</b>	-0.648	<b>-0.666</b>
<b>NOL</b>	-0.623	<b>-0.634</b>
<b>BNL</b>	<b>-0.730</b>	-0.384
<b>WFB</b>	<b>-0.718</b>	0.023
<b>ZYB</b>	<b>-0.879</b>	0.254
<b>AUB</b>	<b>-0.749</b>	0.366
<b>BPL</b>	-0.608	-0.318
<b>JUB</b>	<b>-0.891</b>	0.265
<b>MAB</b>	<b>-0.701</b>	0.171
<b>ZMB</b>	<b>-0.707</b>	0.318
<b>EKB</b>	<b>-0.858</b>	0.160
<b>UFBR</b>	<b>-0.881</b>	0.078
<b>Variance Explained</b>	6.858	1.525
<b>% Total</b>	57.1%	12.7%

**Red denotes highest loadings**

A PCA grouping by country of origin was also run to see if the craniometric data could further differentiate the Hispanic subpopulations. The group centroids for the individuals from Mexico, Guatemala, El Salvador, and the United States (still separated as American Blacks and Whites) is plotted in Figure 6.4.



**Figure 6.4. Group centroids of first 2 factors pooled by country of origin.**

Similar to the pooled ancestral groups (Figure 6.3), the individuals from countries that predominately make up the Hispanic cohort separate well from the American Blacks and Whites. All have positive means for factor 2 and samples from El Salvador and Guatemala tend to differentiate themselves from those from Mexico with positive factor 1 values, as well. This may indicate the possibility of more nuanced country of origin predictions in the future for unknown Hispanic individuals.

More nuanced interpretations of the PCA results is not appropriate as this type of analysis is used as a technique for pre-processing the data before further biological distance analysis methods. In this case, the principal components analysis was used as a data-reduction technique to make the craniometric data applicable to the geobiological model. It was the first two PCA scores for the Hispanic individuals from the more conservative ancestral analysis that were chosen for application to the geobiological model process.

## Macromorphoscopic Data Analysis

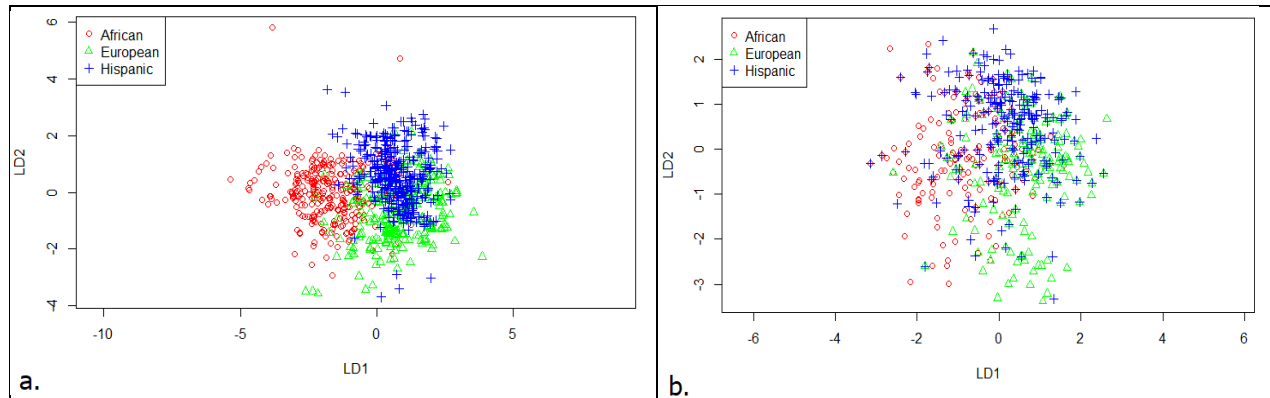
The examination of the macromorphoscopic data set began with evaluating the effectiveness of several distance measures within the canonical analyses of the principal coordinates (CAP) using the R package BiodiversityR (Kindt & Coe 2005). The measures were chosen based on the dataset and the appropriateness of the method for analysis. For example, the Jaccard measure was not implemented as it is based on presence/absence data, which the MMS data are not exclusively such. However, as CAP transforms categorical variables into continuous data using an unconstrained ordination procedure it overcomes the inappropriateness of using MMS data with some of the available distance measures, such as the Mahalanobis. Ultimately, the Mahalanobis and alternate Gower distance measures were applied.

The R package also provides both the correctly classified percentage for each group as well as the appropriate number of axes (m) to use in the CAP analyses. The results of the two 3-group CAP discrimination analyses using the Mahalanobis and alternate Gower distance measures are presented in Table 6.5, and the respective ordination plots are presented in Figure 6.5.

**Table 6.5. Classification rates of the three-way CAP analysis**

<b>Distance Measure</b>	<b>m</b>	<b>Group</b>	<b>% Correct</b>
Mahalanobis	10	African	71.3
		European	84.9
		Hispanic	80.4
		<b>Overall</b>	<b>78.9</b>
Alternate Gower	10	African	62.1
		European	61.8
		Hispanic	67.8
		<b>Overall</b>	<b>64.2</b>





**Figure 6.5. Ordination plots using (a) Mahalanobis ( $m=10$ ) and (b) alternate Gower ( $m=10$ ) distance measures.**

In both cases, an axes number of  $m = 10$  was calculated for between group distinction; but the Mahalanobis distance measure had a higher overall classification rate of 78.9% compared to the alternate Gower's of 64.2%. It also had a higher classification rate for the Hispanic population at 80.4%, which is of ultimate interest in this study as this is the predominant population of undocumented border crossers. The lowest classification rate using the Mahalanobis distance was for individuals of African ancestry at 71.3%, but this is still better than chance. The alternate Gower index was the most consistent for each of the three groups with a range of roughly 62 - 64% correct classification, but again, only slightly better than classifying the individual correctly than by chance. The ordination plots also presents more group separation using the alternate Gower method than with the Mahalanobis as is apparent in the point densities of each group (Figure 6.5).

Previous research using a 3-group method and the Chi-square distance measure achieved similar results to those obtained here (80.3% correct and 78.9% overall, Table 6.5) with an overall correct classification rate of 75.46% and 74.81% of the Native American subsample; the closest subgroup to the current Hispanic sample (Hefner 2007). While the present research had a

poorer correct classification rate for American Black individuals than Hefner's (2007) results, it had a higher correct classification rate for the American White individuals (72.22%). Due to the absence of the Chi-square distance measure in the BiodiversityR package, and the convention of using the Mahalanobis distance for other methods of biodistance analyses this measure was chosen for the final CAP analysis (Anderson & Willis 2003). Again, while this measure is calculated using a variance-covariance matrix, the transformation of the MMS data into continuous values during the CAP analysis overcomes this issue. The comparability of these results to previous research (Hefner 2007) support the application of the two PCOa scores generated with the BiodiversityR package to the geobiological model and the use of this method for future macromorphoscopic data analyses.

### **Random Forest Models**

Several iterations of a Random Forest Model (RFM) using the R program Rattle (Williams 2011) were run using a combination of variables for predicting both the ancestry and country of origin of unknown individuals. These include: all the data available for all of the individuals, the 16 variables from the previously described PCA and CAP models for all of the individuals, all data available on just those with known demographic information, and just the 16 model variables for the positively identified data. A table summarizing the out-of-bag (OBB) error rates for each of the RFMs grouped by prediction level is presented in Table 6.6. OBB predictions measure the performance of the model by “using the observations that are not included in the “bag”—... the subset of the training dataset used for building the decision tree” (Williams 2001: 251) to develop each predictive tree. The OBB predictions are then aggregated and the error rate is calculated on this value (Liaw & Wiener 2002). An unbiased estimate of

error, the OBB rate ensures overfitting of the model does not occur, as long as enough decision trees have been created (Williams 2001).

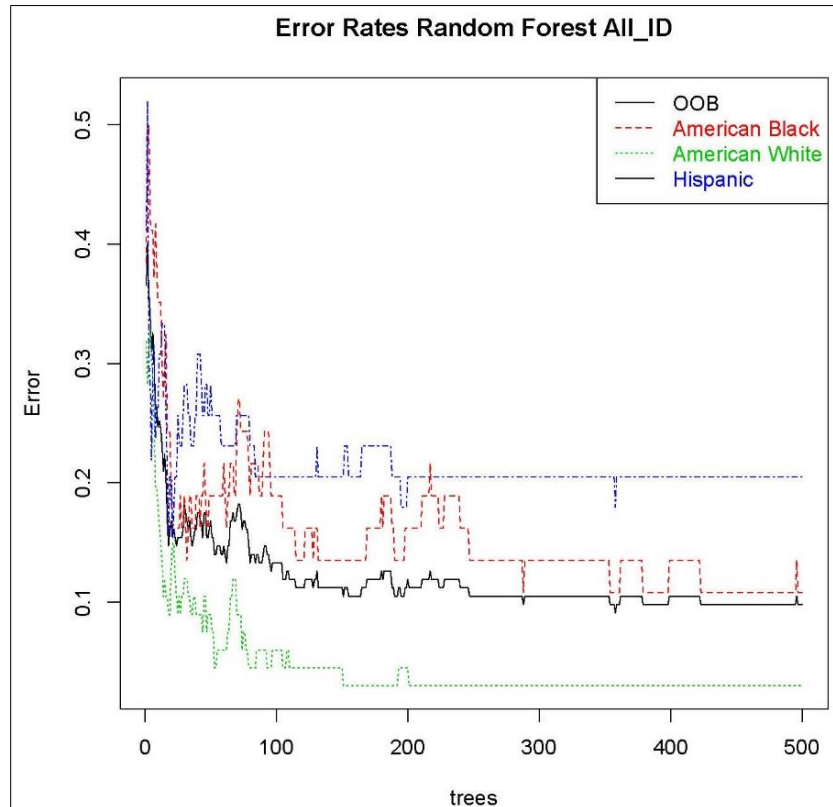
**Table 6.6. Out-of-box error rates for RFMs run for ancestry and country of origin**

<b>ANCESTRY</b>	<b>All</b>	<b>Identified</b>	<b>COUNTRY OF ORIGIN</b>	<b>All</b>	<b>Identified</b>
<b>All Data</b>	13.33%	9.79%	<b>All Data</b>	27.88%	16.08%
<b>Model</b>	16.36%	11.89%	<b>Model</b>	23.64%	13.99%

The RFM for ancestry prediction using all data with known demographic information performed best with an accuracy of just over 90% (Table 6.6). For predicting country of origin, however, the model using the 16 variables from the craniometric and macromorphoscopic analyses performed best with an accuracy rate of approximately 86%. The confusion matrix for both models are presented in Tables 6.7 and 6.8, while the OBB error plots are shown in Figures 6.8 and 6.9. These matrixes present the number of correctly and mis-classified samples for each ancestral group (Table 6.7) and country of origin (Table 6.8), and the associated overall classification error for that category. The error plot “reports the accuracy of the forest of trees (in terms of error rate on the y-axis) against the number of trees that been included in the forest (the x-axis)” (Williams 2011: 255-256). The OBB line is the model’s overall error rate while the other lines show the error rates of the dependent variables with the addition each additional tree to the forest (Williams 2011).

**Table 6.7. Confusion matrix for RFM for ancestry prediction**

	<b>Black (n=37)</b>	<b>White (n=67)</b>	<b>Hispanic (n=39)</b>	<b>Error (%)</b>
<b>Black</b>	33	3	1	10.8
<b>White</b>	1	65	1	2.9
<b>Hispanic</b>	1	7	31	20.5



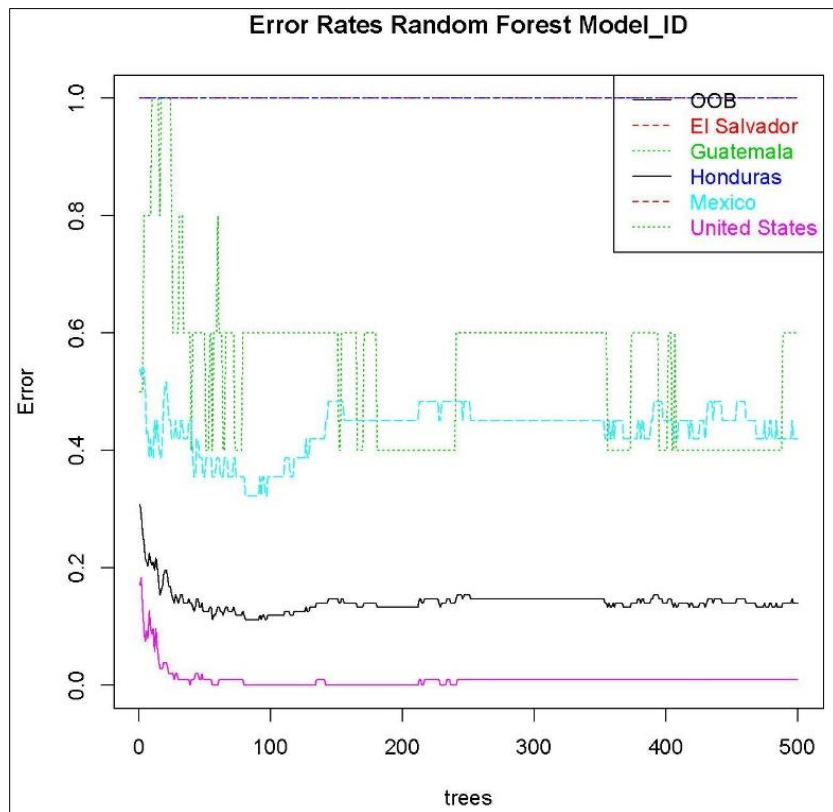
**Figure 6.6. The ancestry out-of-bag error estimate, across 500 trees.**

In the ancestry model, the American White group had the lowest error rate (2.9%) with 65 correctly classified and 1 each incorrectly classified as American Black and Hispanic. The American Black group had an overall classification error rate of 10.8%, while the Hispanic group had the highest average error rate at 20.5% (Table 6.7). Most of the mis-classification of Hispanic individuals are for American Whites, which is likely a reflection of the genetic heritage

of this group involving the admixture of European and Indigenous American populations (Wang et al. 2008). The error plot (Figure 6.6) indicates that the model performed best for all three groups after roughly 150 trees were added and did not significantly improve after 250 for either the American White or Hispanic group classifications (as this is when the error line levels off).

**Table 6.8. Confusion matrix for RFM for country of origin prediction**

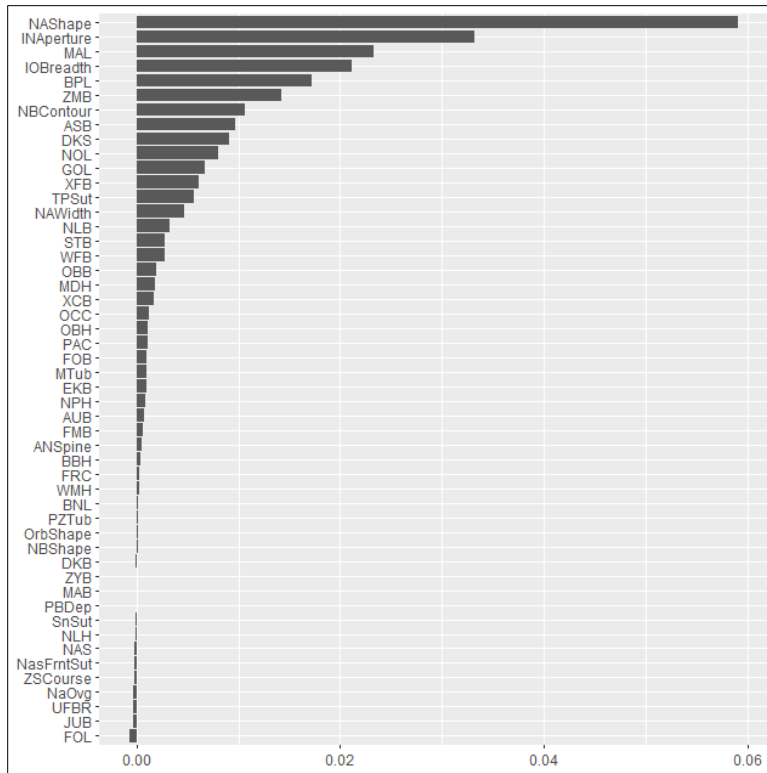
	El Salvador (n=2)	Guatemala (n=5)	Honduras (n=1)	Mexico (n=31)	United States (n=104)	Error (%)
El Salvador	0	0	0	1	1	100
Guatemala	0	2	0	2	1	60
Honduras	0	0	0	1	0	100
Mexico	0	1	0	18	12	41.9
United States	0	0	0	1	103	0.96



**Figure 6.7. The country of origin out-of-bag error estimate, across 500 trees.**

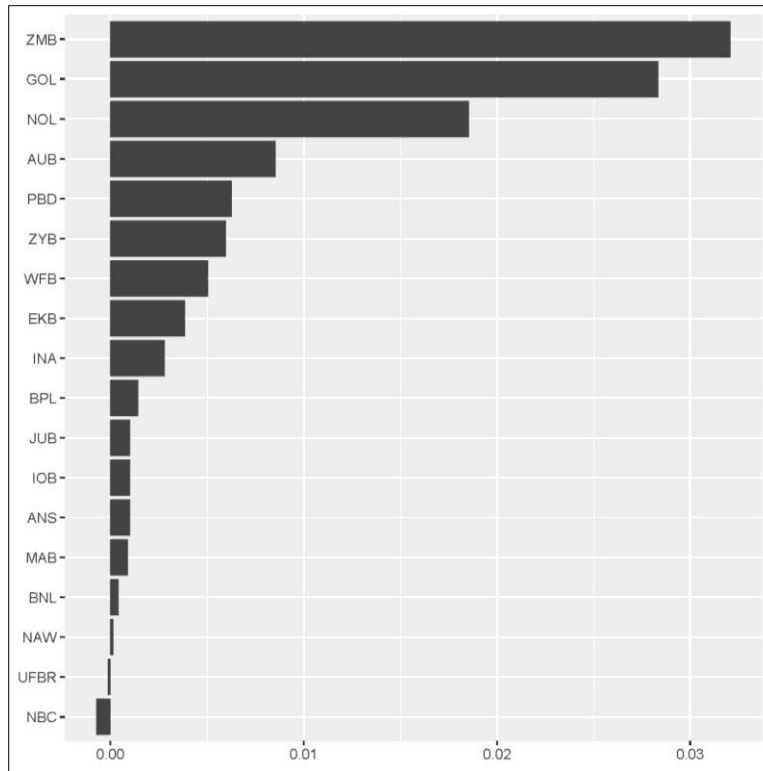
The error rates for correctly predicting the country of origin of an individual are much higher, even though the overall rate is roughly 86% (Table 6.6). For example, neither of the two El Salvadoran individuals were correctly predicted regardless of how many trees were added; mostly due to the small sample sizes for it and several of the Latin American countries present in the dataset (Table 6.8). However, the model did relatively well for predicting which individuals were from the United States; which for now corroborates the usefulness these types of models for predicting ancestry until more individuals from these other countries can be collected.

The importance of each variable used in the respective ancestry and country of origin forest model was weighed and visual displays of the mean decrease in accuracy are presented in Figures 6.8 and 6.9. As each variable is removed, the impact of the variable on the accuracy of the model is calculated; with more important variables having larger impacts. The variables are plotted in descending order of importance and contain a mix of both craniometric and macromorphoscopic variables (y axis). The horizontal (x) axis is the mean decrease in accuracy when the variable is removed (Figure 6.8 and 6.9).



**Figure 6.8. Importance plot of variables used in ancestry prediction RFM.**

In the ancestry analysis, the first 6 variables are the most important to the predictive powers of the model, with less than a 10% loss of accuracy after nasal bone contour (NBC) (Figure 6.8). These six variables include the nasal aperture shape (NAS), the infranasal aperture border (INA), external palate length (MAL), the interorbital breadth (IOB), basion-prosthion length (BPL), and zygomaxillary breadth (ZMB). As with the craniometric PCA analyses, these measurements and features are mostly related to the width and length of the midface which have been proven as useful for explaining ancestral group variance (Spradley 2013; Hefner 2016).



**Figure 6.9. Importance plot of variables used in country of origin prediction RFM.**

In the country of origin prediction forest model, the first 3 variables of zygomaxillary breadth (ZMB), maximum cranial length (GOL), and the nasion-opisthion length (NOL) account for roughly 80% of the model (Figure 6.9). This indicates that overall length and width measurements may be more important for distinguishing between country of origin groups rather than midface features like those highlighted in the ancestry model.

In fact, the first of the macromorphoscopic variables is post-bregmatic depression at fifth most important and infranasal aperture shape as the ninth most important variable for this forest. But since this model was mostly accurate at distinguishing individuals from the United States versus Latin American countries, this must also be taken into consideration. The inclusion of both American White and Black individuals in the US sample suggests this forest model is probably only extracting the overall cranial size differences between Hispanics and non-



Hispanics. A more nuanced country of origin models may be more beneficial by not including non-Hispanic individuals from the United States. Unfortunately, this requires much larger sample sizes than are currently available.

## **Conclusions**

Correlations between craniometric and macromorphoscopic indicators of ancestry, and an individual's ancestral group and country of origin have been highlighted using several methods. The craniometric PCA and the macromorphoscopic PCOa results from the more conservative ancestral group analyses of Hispanic individuals will be applied with the geospatial data in the forthcoming geobiological model discussed in Chapter Seven. The results of the random forest model testing will be used in future investigations involving refinement of the geobiological model.

## REFERENCES

## REFERENCES

- Anderson MJ, Willis TJ. 2003. Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology*, 84(2): 511-525.
- Hefner JT. 2016. Biological distance analysis, cranial morphoscopic traits, and ancestry assessment in forensic anthropology. In: *Biological Distance Analysis*. Pilloud MA, Hefner JT (eds). San Diego: Academic Press. p. 301-315.
- Kachigan SK. 1991. *Multivariate Statistical Analysis: A Conceptual Introduction*. New York: Radius Press.
- Kindt R, Coe R. 2005. *Tree diversity analysis. A manual and software for common statistical methods for ecological and biodiversity studies*. World Agroforestry Centre (ICRAF).
- Spradley MK. 2013. *Project IDENTIFICATION: Developing Accurate Identification Criteria for Hispanics*. US Department of Justice.
- Spradley MK. 2014. Toward estimating geographic origin of migrant remains along the United States-Mexico border. *Annals of Anthropological Practice*, 38(1): 101–110.
- Wang S, Ray N, Rojas W, Parra MV, Bedoya G, et al. 2008. Geographic patterns of genome admixture in Latin American Mestizos. *PLoS Genetics*, 4(3): e1000037.
- Williams GJ. 2011. *Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery, Use R!*, New York: Springer.

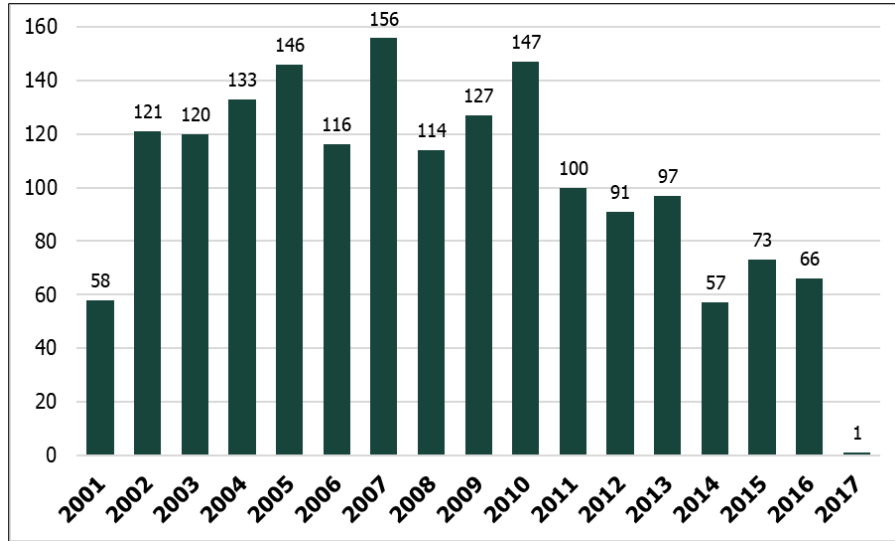
## **CHAPTER 7. GEOSPATIAL RESULTS**

### **Introduction**

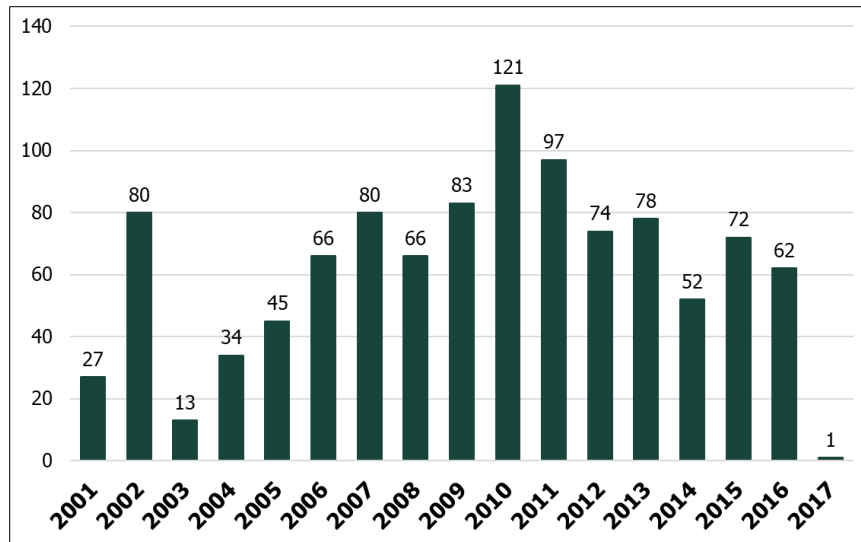
The geospatial analysis portion of this research project included both exploratory analyses and applications of geographically weighted regression (GWR) procedures. The exploratory process was necessary to confirm the presence of spatial patterns amongst the recovered undocumented border crossers in southern Arizona. Only upon the presence of positive spatial autocorrelation can the use of GWR methods be appropriate. This chapter first presents the results of the Exploratory Spatial Data Analysis (ESDA) followed by the investigations into the geospatial model and subsequent GWR analysis.

### **Exploratory Spatial Data Analysis (ESDA)**

To begin the exploratory analysis, descriptive statistics of the cases analyzed at the PCOME during the study's timeframe were calculated ( $N = 2,616$ ). The yearly distribution of recoveries is presented in Figure 7.1; while the distribution of identified individuals with accurate GPS data used in the ESDA is presented in Figure 7.2 ( $n = 1,051$ ).



**Figure 7.1. Number of recoveries per year for all recovered remains (N = 2,616).**



**Figure 7.2. Number of recoveries per year for identified individuals with accurate GPS data (n = 1,051).**

The number of recovered cases increased during the beginning half of the decade, peaking in 2007 at 156 (Figure 7.1). However, only 80 of those had accurate GPS data (Figure 7.2). The year 2010 had the most recoveries with accurate GPS at n = 121 (Figure 7.2). The ratio

between the number of recoveries and those with accurate GPS data rises over several years and therefore is biased towards later dates and should be considered in the subsequent analyses.

While the number of recoveries has declined slightly over the last 10 years, death rates appear to be increasing compared to the number of border patrol apprehensions. “Although the number of UBCs investigated by the PCOME decreased [between 2009 and 2011], the number of apprehensions in the Tucson sector decreased at a much faster rate during the same period, from 241,673 to 123,285. This suggests that the number of unauthorized crossers traversing the area also decreased substantially” (Martinez et al. 2013: 13).

The distribution of identified cases by sex and country of origin with accurate GPS data is presented in Table 7.1. The majority are from Mexico, making up roughly 74% of all females and 85% of males, with the next highest group from Guatemala at 16% and 8%, respectively. This is consistent with the previously documented demographic trends of immigration within the Tucson sector of both living and deceased immigrants (Martinez et al. 2013). While people from several Central American and Caribbean countries are also present, they have been pooled accordingly in the ESDA to increase group sample sizes.

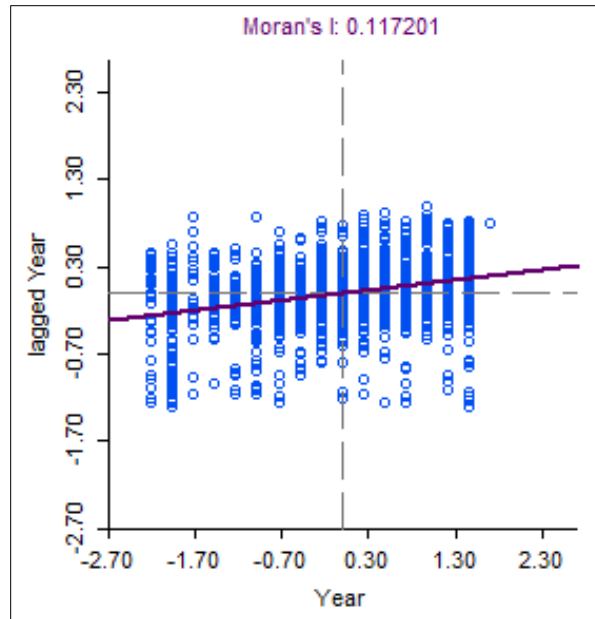
**Table 7.1. Demographic distribution of identified individuals used in ESDA methods**

<b>Country of origin</b>	<b>Female</b>	<b>Male</b>	<b>Total</b>
<b>Mexico</b>	130	746	876
<b>Guatemala</b>	28	73	101
<b>El Salvador</b>	9	24	33
<b>Honduras</b>	1	23	24
<b>Ecuador</b>	3	4	7
<b>Peru</b>	1	2	3
<b>Colombia</b>	1	1	2
<b>Costa Rica</b>	0	1	1
<b>Dominican Republic</b>	1	0	1
<b>Jamaica</b>	0	1	1
<b>Nicaragua</b>	1	0	1
<b>Unknown*</b>	0	1	1
<b>Grand Total</b>	175	876	1051

**\*Fetus**

Pilot studies indicate positive spatial autocorrelation in these data, so more nuanced explorations of spatial autocorrelation analyses were deemed appropriate (Vogelsberg 2018). These include the Global Moran’s I and Local Indicator of Spatial Association (LISA) cluster maps with associated Moran’s I values.

The results of the Global Moran’s I test shows evidence of slight positive spatial autocorrelation ( $I = 0.1172$ ) for year of recovery using a spatial weights matrix with a nearest neighbor of ( $k = 41$ ) for all individuals. An example of this Global Moran’s plot is presented in Figure 7.3.



**Figure 7.3. Global Moran's I plot for year of recovery, using kNN ( $k = 41$ ).**

Evidence of positive spatial autocorrelation is also present for the variables of sex, country of origin, and identification status as expressed in the positive Global Moran's I statistic (Table 7.2).

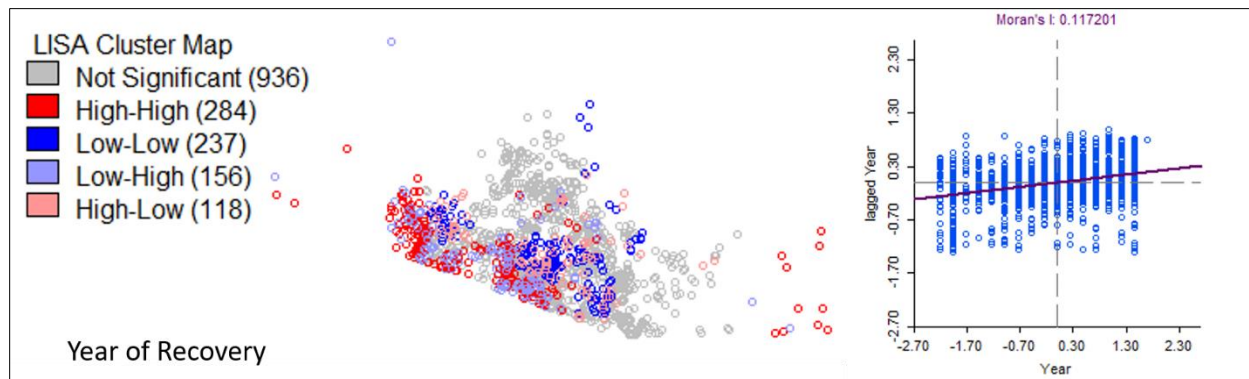
**Table 7.2. Global Moran's I statistics for all individuals**

<b>Variable</b>	<b>Global Moran's I</b>
<b>Year of Recovery</b>	0.117
<b>Country of Origin</b>	0.025
<b>Sex</b>	0.036
<b>Identified vs. Unidentified</b>	0.029

However, this statistic only suggests the presence of spatial autocorrelation within the sample, not necessarily where it exists. LISA analyses can assess the extent of significant spatial clustering around an observation relative to the overall global indicators of spatial association



(Lloyd 2007). Maps highlighting the clustering of attributes can also be created using the LISA values, which is useful for understanding the sample’s geospatial properties. The Moran’s scatter plot and LISA cluster maps for all individuals for the variable of year of recovery are presented in Figures 7.4.



**Figure 7.4. LISA Cluster map and Moran's scatter plot for year of recovery variable for all individuals.**

For year of recovery (Figure 7.4), the points on the LISA cluster maps shaded in red indicate “high-high” clusters, or individuals recovered in later (or “higher”) years that are located near others also recovered in later years; while the blue indicates the “low-low” clusters, or individuals recovered in earlier years near others recovered around the same time. The “low-low” cluster represents the years 2001-2010 and the “high-high” from 2011-2017. The remaining categories indicate individuals recovered in earlier years surrounded by later individuals (“low-high”) and individuals recovered in later years surrounded by earlier individuals (“high-low”), as well as non-significant clusters (grey).

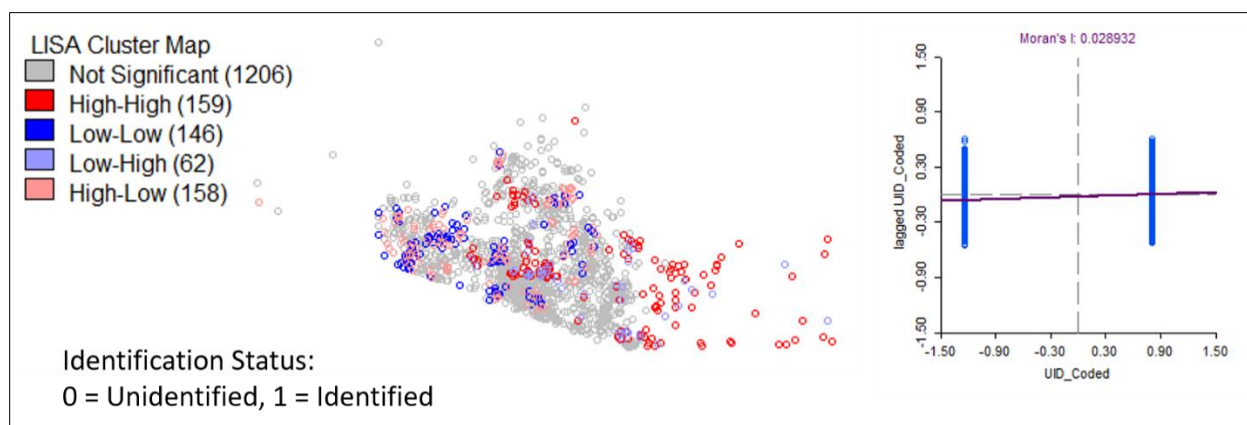
Both the “high-high” and “low-low” clusters represent a positive spatial association wherein an individual is surrounded by individuals of similar years. “Low-high” and “high-low”

values indicate a negative association as there is a cooccurrence of the year groups. Additionally, there are several non-significance clusters. As the frequency counts on the legend indicates, there are more of the “high-high” (n = 284) and “low-low” clusters (n = 237) than the negative spatial autocorrelation options (n = 158 and n = 118, respectively).

The blue, or older, cluster of cases are more apparent in the center of the state while the red, or clusters of cases from more recent years, appear more in the western half of the state. Table 7.3 presents the number of individuals recovered each year, with those recovered in the “low” or earlier years (2001 – 2010) totaling 615, and 436 from the “high” or later years of 2011 – 2017.

**Table 7.3. Number of identified individuals by year**

<b>Year</b>	<b>Identified</b>
<b>2001</b>	27
<b>2002</b>	80
<b>2003</b>	13
<b>2004</b>	34
<b>2005</b>	45
<b>2006</b>	66
<b>2007</b>	80
<b>2008</b>	66
<b>2009</b>	83
<b>2010</b>	121
<b>2011</b>	97
<b>2012</b>	74
<b>2013</b>	78
<b>2014</b>	52
<b>2015</b>	71
<b>2016</b>	63
<b>2017</b>	1
<b>Grand Total</b>	<b>1051</b>



**Figure 7.5. LISA cluster map and Moran's scatter plot for identification status variable for all individuals.**

Figure 7.5 presents the LISA cluster map and associated Moran's I plots for all individuals coded as identified (1) or unidentified (0). There is slight positive spatial autocorrelation between identified (red, or "high-high") and unidentified (blue, or "low-low") cases with  $I = 0.029$ . The identified cluster around the eastern and center portions of the state, while the unidentified are more frequent in the western regions. Several reasons may be involved in this spatial distribution which will be discussed in Chapter 8.

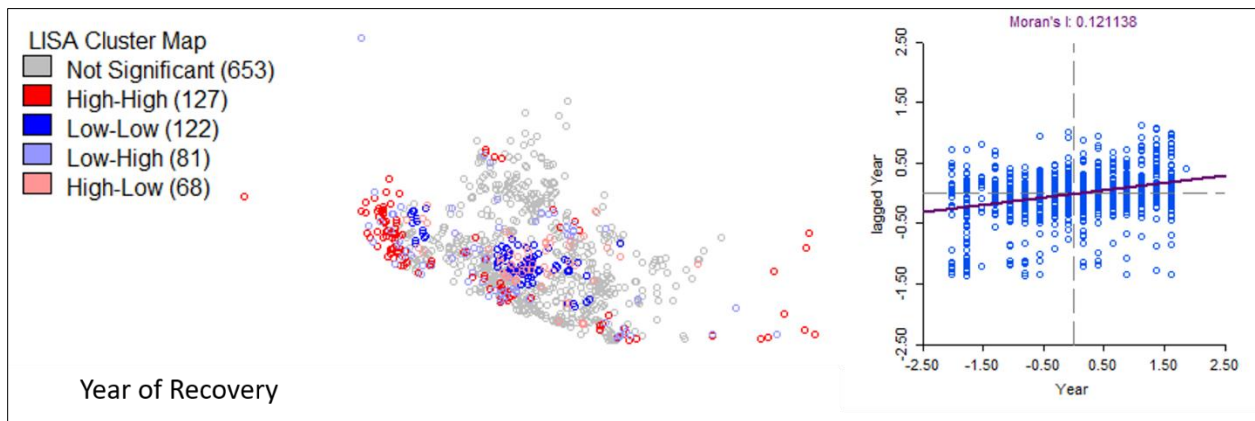
### *Identified Individuals*

To continue the investigation into the emerging spatial patterns in the recovery locations of migrant remains in southern Arizona, the next step was to focus only on the identified individuals. The results of the Global Moran's I statistic using a spatial weights matrix and  $k = 33$  was created for the variables of year of recovery, sex, and country of origin are presented in Table 7.4. The positive values (albeit small) reflect the presence of positive spatial autocorrelation for all three variables.

**Table 7.4. LISA cluster map and Moran's scatter plot for identification status variable for all individuals**

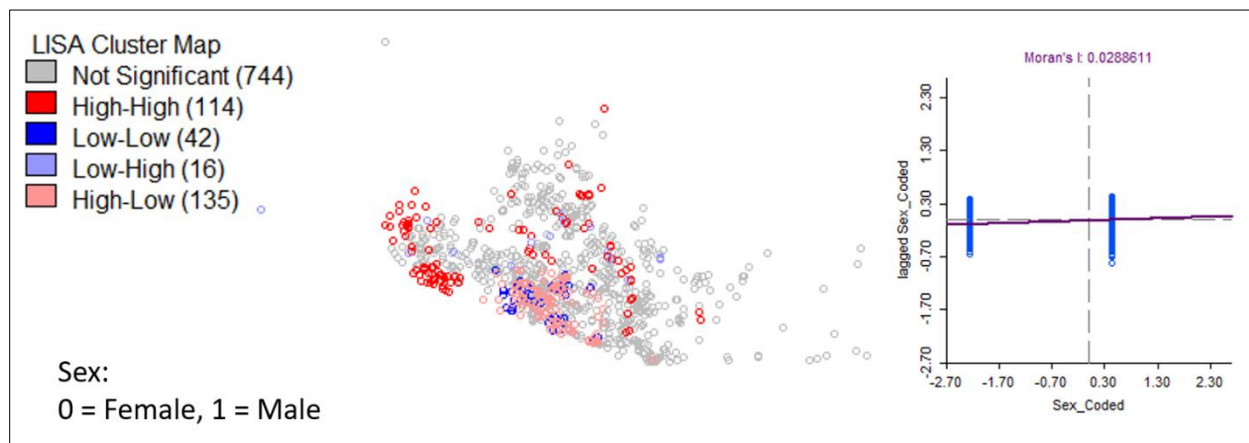
<b>Variable</b>	<b>Global Moran's I</b>
<b>Year of Recovery</b>	0.121
<b>Sex</b>	0.029
<b>Country of Origin</b>	0.025

Compared to the Global Moran's I for all individuals (Table 7.2), the *I* values of identified individuals are similar; if only slightly improved in year of recovery and country of origin. Again, LISA statistical analyses for these variables are warranted to account for the large geographic area of the study. The Moran's scatter plot and LISA cluster maps of identified individuals for the variables of year of recovery, sex, and country of origin are presented in Figures 7.6 through 7.8.



**Figure 7.6. LISA cluster map and Moran's scatter plot for year of recovery of identified individuals.**

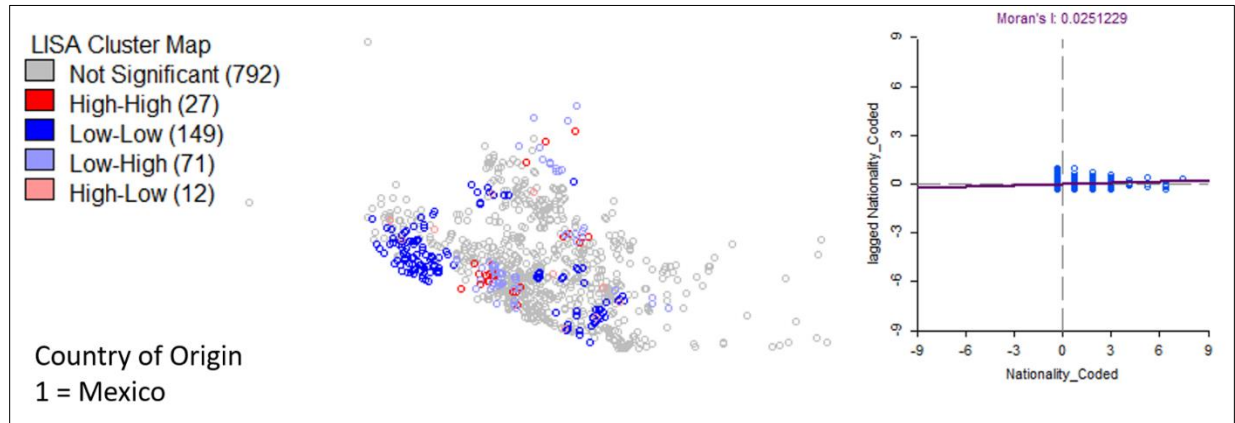
Year of recovery presents a similar pattern as the map for all individuals and a slightly higher *I* value (Figure 7.4). This is not surprising as these individuals were included in the previous analysis and therefore the same results are being highlighted, just with fewer cases.



**Figure 7.7. LISA cluster map and Moran's scatter plot for sex of identified individuals.**

When tests of spatial autocorrelation for sex are performed on the subsample of identified individuals, patterns in the recovery location of males and females emerge (Figure 7.7). Females are clustering within the south-central portion of the study area (as indicated by the blue “low-low” and pink “high-low” clusters) while males tend to have a larger spread across the state, with a higher proportion in the western section. This may indicate that females are more likely to travel within the central zone, while males travel throughout the state. This seems to follow a similar pattern to year of recovery and when the number of females from the two halves of the study period are compared; the proportion of females found in the earlier years (2001-2010) is almost 3.5 times more than the later (2010-2017), at  $n = 136$  and  $n = 39$ , respectively. That seems to indicate women were more crossing through these areas in the early 2000s in these areas; while men have continued to cross throughout the state. Indeed, the clusters of men at the

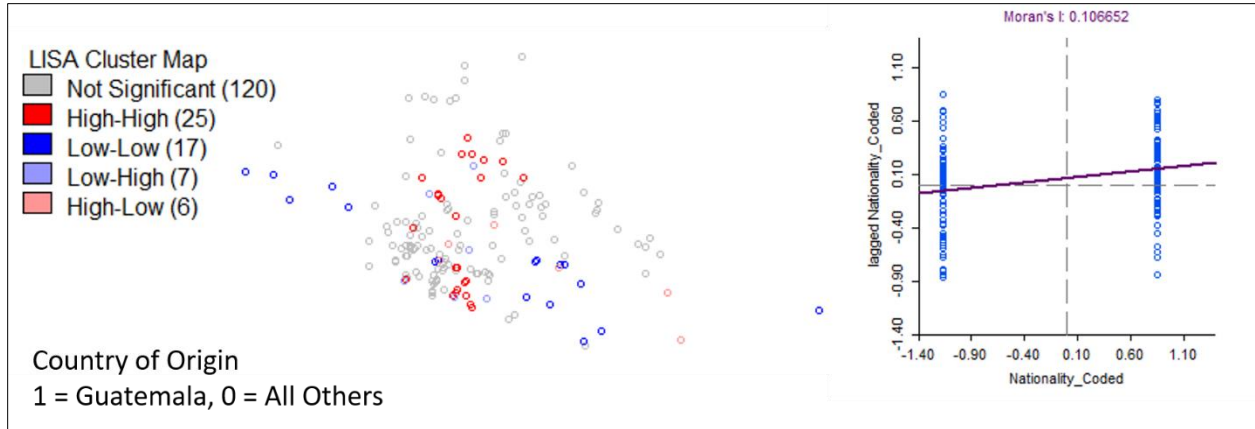
most western portion of the study area from the entire study period (2001-2017) when investigated further.



**Figure 7.8. LISA cluster map and Moran's scatter plot for country of origin of identified individuals.**

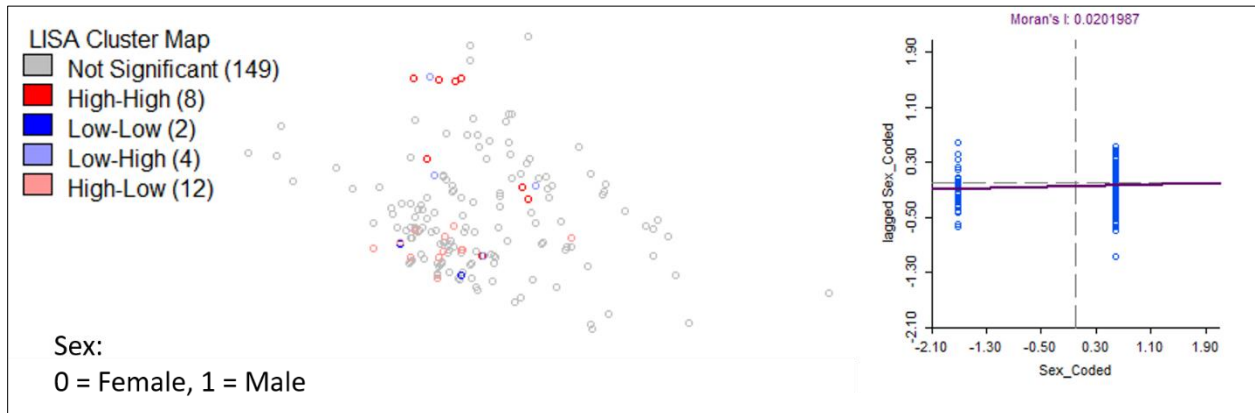
The LISA cluster map for the country of origin of identified individuals (Figure 7.8) and its respective Moran's I also indicate positive spatial autocorrelation. In this analysis, each country was coded 1-8 in descending order of overall counts (Table 5.12) and less represented South American and Caribbean nations pooled, respectively. Mexico was coded as 1, as seen in the high number of “low-low” (149) and “low-high” (71) clusters throughout the state and higher concentrations in the western and central regions.

The high proportion of Mexican males in the sample may have potentially biased the results, so these were removed to see if there was any evidence of spatial autocorrelation amongst individuals from other countries. Figure 7.9 presents the LISA results for the country of origin for individuals excluding Mexico, while Figure 7.10 presents the LISA results for the sex of individuals not from Mexico.



**Figure 7.9. LISA cluster map and Moran's scatter plot for country of origin of individuals not from Mexico.**

In this analysis, Guatemala was coded as 1, while all other countries were pooled and coded as 0. While this greatly diminished the sample (from  $n = 1,051$  to 175), there is still positive spatial autocorrelation of the individuals from countries other than Mexico.



**Figure 7.10. LISA cluster map and Moran's scatter plot for sex of individuals not from Mexico.**

The analysis of spatial autocorrelation for sex within the subsample of identified individuals not from Mexico (Figure 7.10) also presents a small, but positive Moran's I; though

there are very few numbers of significant LISA clusters. Discussions as to why this pattern may be present are available in Chapter 8.

### *Join Count*

Placing a 40 km<sup>2</sup> grid over the study area resulted in the shapefile in Figure 7.11. In each cell, the ratio of unidentified individuals to identified, females to males, and individuals from non-Mexican countries to those from Mexico were calculated and formed the basis of the following ESDA results. Instead of the *k* nearest neighbors-derived spatial weights matrix, a queen's contiguity method was employed to calculate the relationship between the polygons for all subsequent analyses.

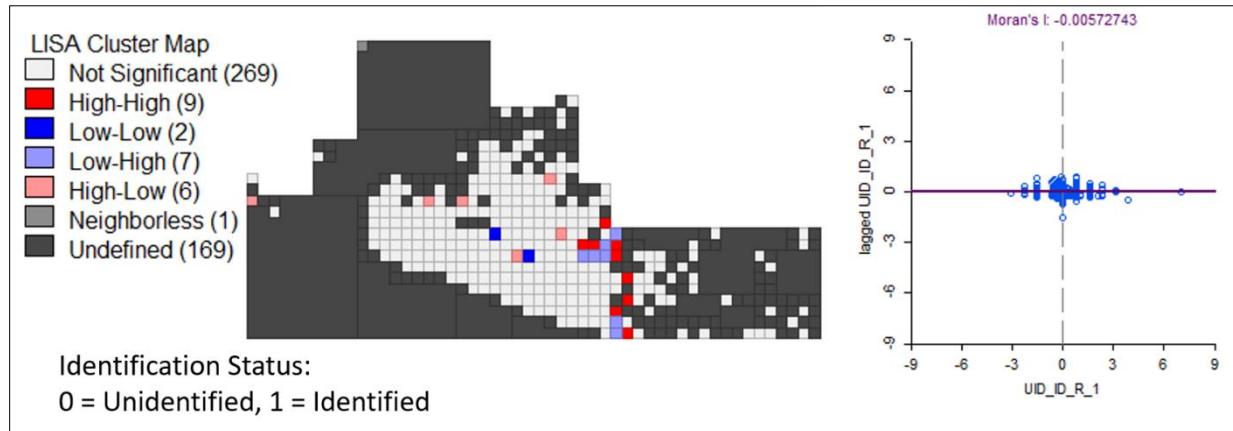


**Figure 7.11. Grid cell transformation.**

Figure 7.12 presents the LISA cluster map and Moran's scatter plot for the ratio of unidentified to identified individuals. The cells without data (i.e. no cases) are shaded a dark

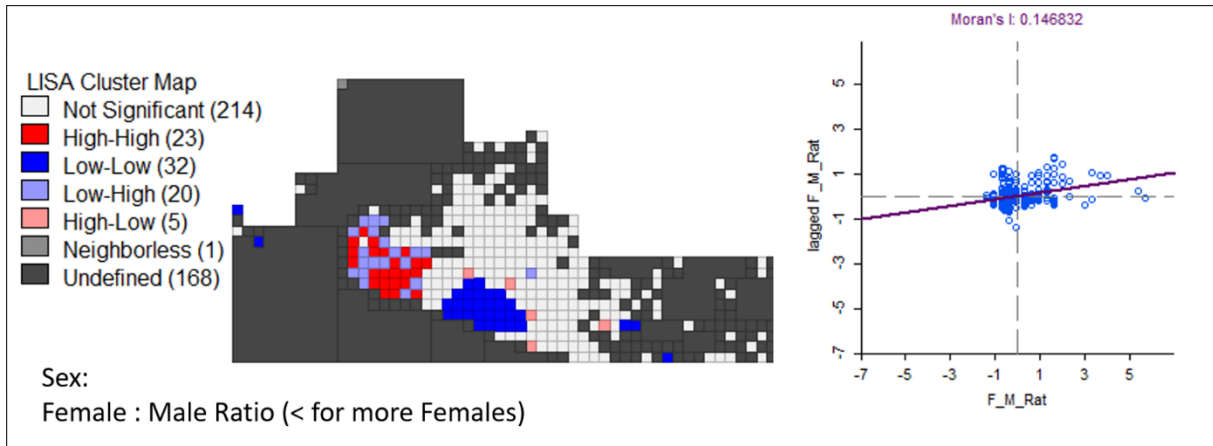


grey, while the red and blue clustering symbology is the same as before for high and low values, respectively.



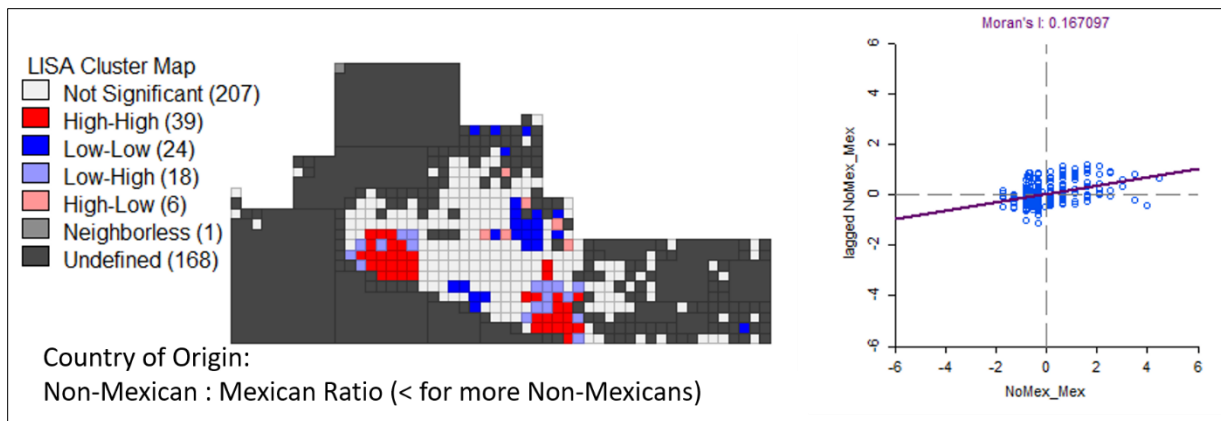
**Figure 7.12. LISA cluster map and Moran's scatter plot for identification status variable for all individuals using grid transformation.**

Compared to the point data analysis presented in Figure 7.5, a different pattern is emerging for identification status (Figure 7.12) with a decrease in the Moran's I value indicating (very slight) negative spatial autocorrelation. This lower value, however, does not necessarily mean there is no significant spatial autocorrelation. Rather, this transformation may be masking the case-by-case correlations and thus is not the best representation of the data. But the clustering of identified individuals within the center of the state and unidentified towards the west is still present.



**Figure 7.13. LISA cluster map and Moran's scatter plot for sex for identified individuals using grid transformation.**

The LISA analysis for the sex of the identified individuals (females : males) is presented in Figure 7.13; showing similar patterns to the point data analysis seen in Figure 7.7. However, the calculation of the ratio of females to males (wherein lower values indicated more females than males within each cell), overcomes the male bias which may have originally been skewing the case-by-case analyses. This results in a higher *I* value and clusters of regions with more females than males recovered are emerging, but still within the same central region as before.



**Figure 7.14. LISA cluster map and Moran's scatter plot for country of origin for identified individuals using grid transformation.**

The last LISA cluster map was the country of origin of identified individuals, calculated using the ratio of persons not from Mexico to those from Mexico (wherein higher ratios mean more individuals from Mexico) is presented in Figure 7.14. Once again, there is a positive Moran's I value indicating positive spatial autocorrelation of cases, along with the same pattern of more migrants from Mexico recovered in the western and central regions of the state (Figure 7.6). The clusters of areas with more individuals not from Mexico are also highlighted in this transformation.

Table 7.5 summarizes all the Moran's I values calculated throughout this exploratory analysis, with the general trend of more positive spatial autocorrelation extracted at more localized levels of analysis.

**Table 7.5. Global and local Moran's I values obtained in ESDA**

<b>GLOBAL MORAN'S I</b>		<b>LISA MORAN'S I</b>	
<b>Variable</b>	<b>I value</b>	<b>Variable</b>	<b>I value</b>
<b>All individuals</b>		<b>All individuals</b>	
<b>Year of Recovery</b>	0.11720	<b>Identified vs. Unidentified</b>	0.02893
<b>Country of Origin</b>	0.02507	<b>Identified individuals</b>	
<b>Sex</b>	0.03469	<b>Country of Origin</b>	0.02512
<b>Identified vs. Unidentified</b>	0.02893	<b>Sex</b>	0.02886
<b>Identified individuals</b>		<b>Country of Origin, sans Mexico</b>	0.10665
<b>Year of Recovery</b>	0.12114	<b>Sex, sans Mexico</b>	0.02019
<b>Country of Origin</b>	0.02512	<b>Grid: all individuals</b>	
<b>Sex</b>	0.02886	<b>UID : ID</b>	-0.00573
		<b>Female : Male</b>	0.14683
		<b>Not Mexico : Mexico Origin</b>	0.16709

## Geographically Weighted Regression Analysis

### *General Linear Models*

Given the presence of positive spatial autocorrelation for several variables, specifically the sex and country of origin of identified individuals, the development of models to potentially predict these attributes in future remains was reasoned possible. Due to the large amount of data necessary, including both types of biological data as well as accurate GPS data, this left  $n = 24$  individuals for analysis. Data were standardized prior to biological data analysis so males and females were pooled. A demographic breakdown of these individuals is provided in Table 7.6.

**Table 7.6. Country of origin and sex of individuals used in GWR analysis**

<b>Country of Origin</b>	<b>Total</b>
<b>Mexico</b>	18
<b>Guatemala</b>	4
<b>El Salvador</b>	2
<b>Grand Total</b>	24

First, ordinary least squares (OLS) models using three separate data and variable combinations were investigated: sex, corridor of recovery, and all PCA and PCOA values. These data combinations included: 1) those presented in Table 7.6, 2) the removal of the two El Salvadoran individuals, and 3) the inclusion of three individuals Mexican individuals (2 males, 1 female) with GPS data labeled as “vague physical description (precise to within 10 mi/15 km)” in the OGIS database. This last inclusion increased sample sizes and tested the effectiveness of adding individuals with less precise GPS data to the model.

Four OLS models were created for potential application to the GWR process:

Model 1: Country of origin = PCA1 + PCA2 + PCOa1 + PCOa2 + Corridor

Model 2: Country of origin = PCA1 + PCA2 + PCOa1 + PCOa2 + Sex

Model 3: Country of origin = PCA1 + PCA2 + PCOa1 + PCOa2 + Corridor + Sex

Model 4: Country of origin = PCA1 + PCA2 + PCOa1 + PCOa2

The models were compared by both their overall significance ( $p$ ) values, adjusted  $R^2$  values, and corrected Akaike Information Criterion (AICc) indexes. Table 7.7 presents the  $p$  and adjusted  $R^2$  values for all 4 models in all 3 data combinations while the AICc indexes are presented in Table 7.8. In general, higher  $R^2$  values are desired as it means more variance is explained by the model than other untested variables, while lower AICc indexes are favored. By convention, a difference of  $\geq 3$  in AICc values demonstrates an improvement in model fit (Fotheringham et al. 2002). A significance threshold of  $p < 0.05$  was used.

**Table 7.7. Summary results of global linear models**

<b>Data Set</b>	<b>Model</b>	<b>P Value</b>	<b>Adjusted <math>R^2</math></b>	<b>Residual Squares</b>
<b>(1) All Data</b>	1	0.407	0.016	1.591
	2	0.220	0.109	1.553
	3	0.257	0.103	1.553
	4	0.396	0.013	7.609
<b>(2) No Outliers</b>	1	0.025	0.373	1.553
	2	0.023	0.382	1.540
	3	0.049	0.341	1.553
	4	0.010	0.410	1.564
<b>(3) Samples Added</b>	1	0.009	0.402	1.591
	2	0.008	0.416	1.553
	3	0.018	0.384	1.553
	4	0.004	0.432	1.591

**Table 7.8. AICc indexes of global linear models**

<b>Data Set</b>	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>	<b>Model 4</b>
<b>1</b>	60.175	57.783	61.181	57.482
<b>2</b>	26.260	25.930	31.006	21.863
<b>3</b>	22.676	22.069	26.478	18.758

Ultimately, the small number of El Salvadoran individuals in Data Set 1 did not permit a usable linear model (Table 7.7). The removal of these outliers resulted in the better performance of the models but with a small sample size ( $n = 22$ , Data Set 2). To increase sample numbers, three individuals of Mexican origin with less precise GPS data were added, for a sample set of  $n = 25$  (designated as Data Set 3). Their GPS data was designated as precise to within 10 miles; while not appropriate for geospatial analyses, these data's imprecision may be negligible in the GWR analysis.

### *OLS Model Choice*

The best performing model for predicting country of origin was Model 4: the combination of just the first two factor scores of the craniometric PCA scores and the macromorphoscopic PCOa scores. This is not surprising as these variables are the result of the earlier ancestry and country of origin differentiation methods using the biological data. The fact that adding the sex of the individual or the corridor in which they were found (as conducted in Models 1-3) also produces statistically significant results highlights the importance these variables may have on the routes in which people take and die.

The performance values presented in Tables 7.7 and 7.8, indicate Model 4 performs only slightly better than Model 2 (which combined the biological data and sex of the individual) and only marginally better than Model 1, which only incorporated recovery corridor with the biological data.

Models 1 and 3 were ultimately not selected due to the variable redundancy regarding corridor location and the inherent nature of geographic weighted regression. Local spatial relationships are incorporated into the regression equations and thus the "dummy variable" designating the corridor of recovery is not necessary (Williams 2011). But, the statistical

significance of this global model still demonstrates the importance of location in the prediction of an individual's country of origin. Model 2 was also not chosen due to application issues involving the necessary independent variable of sex. In the case of incomplete or ambiguous skeletal remains, the sex of the individual may not be certain and therefore the model would not always be applicable.

Model 4 was thus chosen for further analysis. It had the lowest AICc values at a margin greater than 3 which suggests any differences in goodness-of-fit is probably more than error. Two datasets were applied to this Model: Data Set 2 (n = 22), which removed the El Salvadoran outliers and Data Set 3 (n = 25) which added the 3 Mexican individuals with less precise GPS data. In subsequent discussions these tests are designated as Model 4.2 and 4.3, respectively, with the number after the decimal point indicating the data set.

To be sure that a spatial analysis is appropriate, a Global Moran's I was performed on the model's residuals to test for spatial autocorrelation. An  $I$  value of -0.0536 and a non-significant p-value of 0.973 (when  $\alpha < 0.05$ ) for Model 4.2 and  $I = -0.0395$  and  $p = 0.491$  for Model 4.3 indicated no spatial autocorrelation in the residuals. This verifies spatial interdependence of the samples (in that the country of origin of one individual does not affect the biological values of the nearby individuals) so unbiased and efficient parameter estimates can be assumed. This bodes well for the subsequent geographically weighted regression tests.

### *Geographically Weighted Regression Analyses*

Geographically weighted regression (GWR) analyses on the two data sets (Model 4.2 and 4.3) using the R packages *spgwr* (Williams 2011) and *GWmodel* (Gollini et al. 2015) were conducted. Further investigatory approaches were taken using ArcMap (ESRI 2016) and will be also be presented.

Model 4.2

For reference, maps of the individuals in Data Set 2 as coded by sex and country of origin are presented in Figures 7.15 and 7.16.

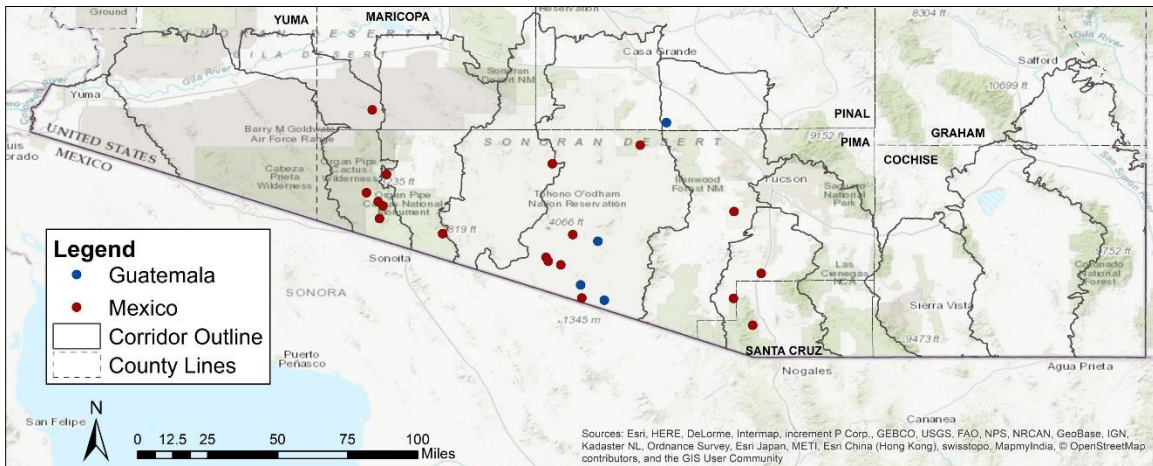


Figure 7.15. Data Set 2 (n = 22) labeled by country of origin.

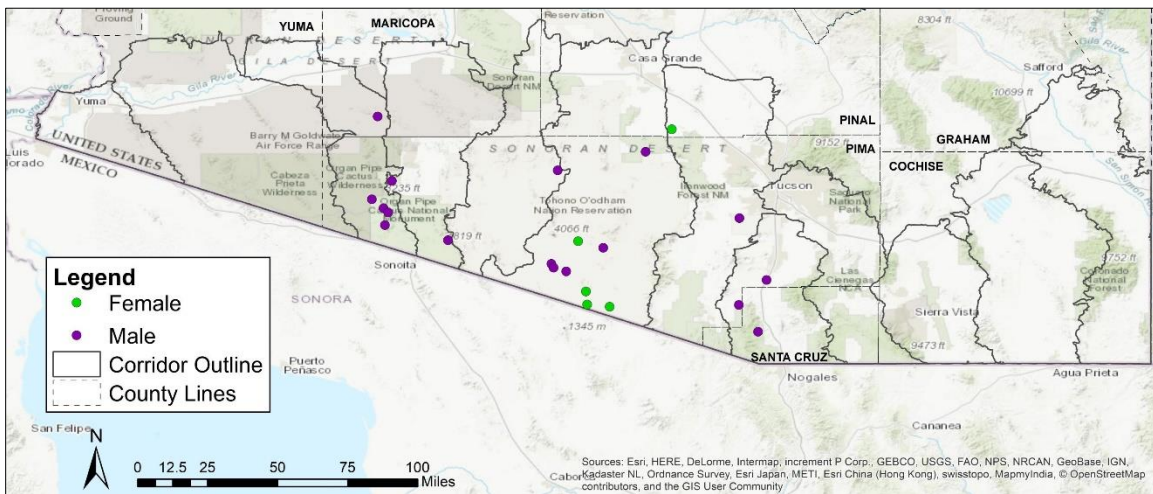


Figure 7.16. Data Set 2 labeled by sex.



## spgwr

First, the R package *spgwr* (Williams 2011) was used to explore the geographically weighted regressive properties of the individuals as they vary across the southern Arizona desert. The diagnostic criteria of the GWR analysis using Data Set 2 is presented in Table 7.9.

**Table 7.9. Diagnostics for Model 4.2 using spgwr**

<b>Output</b>	<b>Value</b>
<b>Residual Squares</b>	1.455
<b>AICc</b>	25.566
<b>R<sup>2</sup></b>	0.555

Compared to the global model, the AICc increased from 21.863 (Table 7.8) to 25.566 (Table 7.9). This suggests the global model is a better fit as the difference is greater than 3; however, other measures are available for understanding model performance. The residual squares equals the difference between an observed value and the predicted value provided by the model. Smaller values indicate a closer fit between the model and the observed data. The residual squares decreased from 1.564 in the global model (Table 7.7) to 1.455 when using GWR methods (Table 7.9). This indicates less difference between the actual country of origin score and that predicted by Model 4.2 using spgwr.

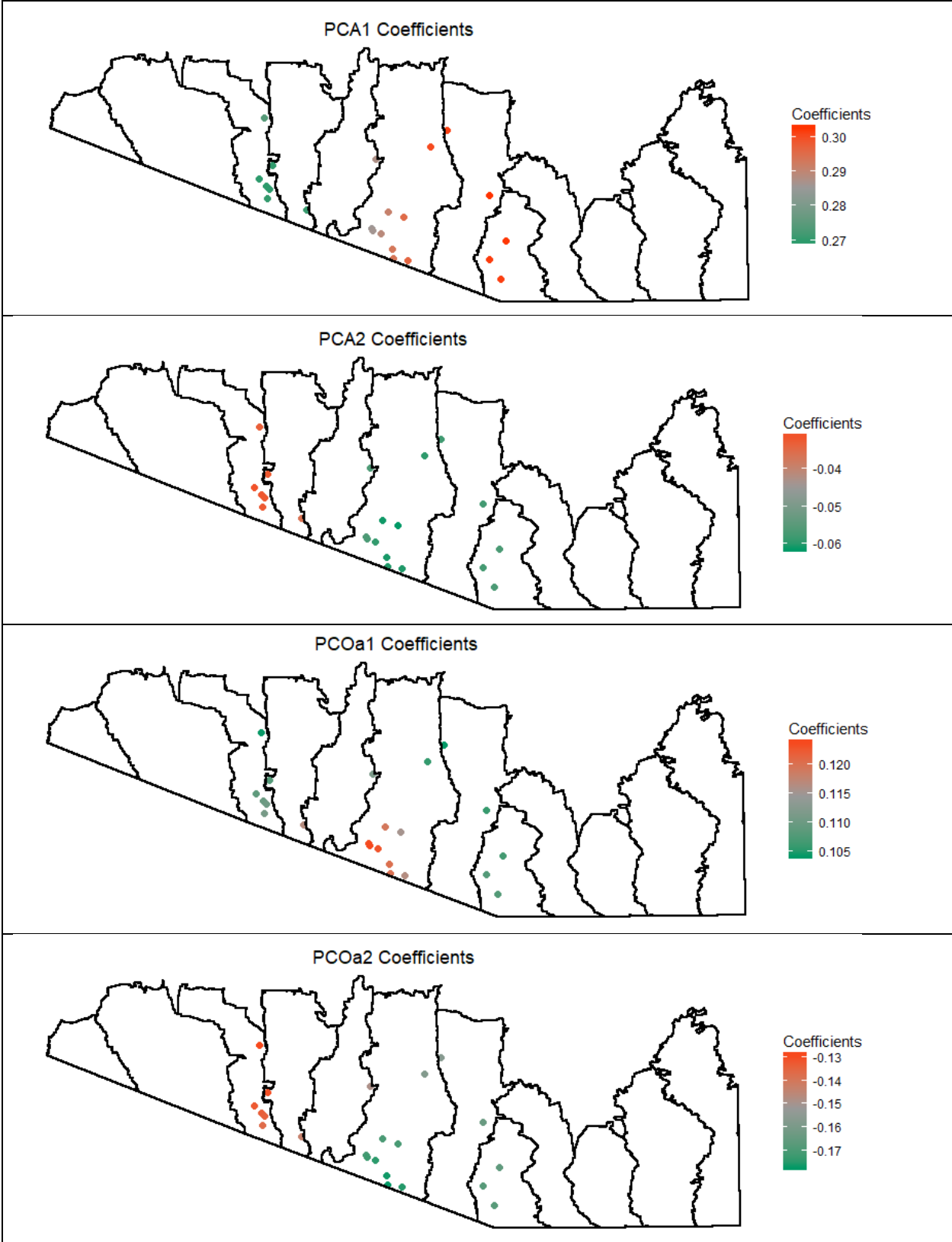
The R<sup>2</sup> value is another goodness-of-fit measure and expresses the amount of variation explained by the model. Usually the adjusted R<sup>2</sup> is reported to take "...into account the number of independent variables in the model and [reflect]... model parsimony" (Charlton et al. 2009: 2). However, in GWR the adjusted R<sup>2</sup> accounts for the effective degrees of freedom of the independent variables, which are a product of the bandwidth. This adjustment removes some of

the interpretive ability of the value for reflecting the proportion of the variance explained by the model (ESRI 2016). So, the more meaningful unadjusted value is reported here.

Using spgwr to analyze Model 4.2, the  $R^2 = 0.555$  (Table 7.9) indicates that the model is explaining just over half of the variation seen in the variables. This is an increase in the global  $R^2$  value of 0.401, but a large portion is still unexplained, and the result of variables not currently being taken into consideration.

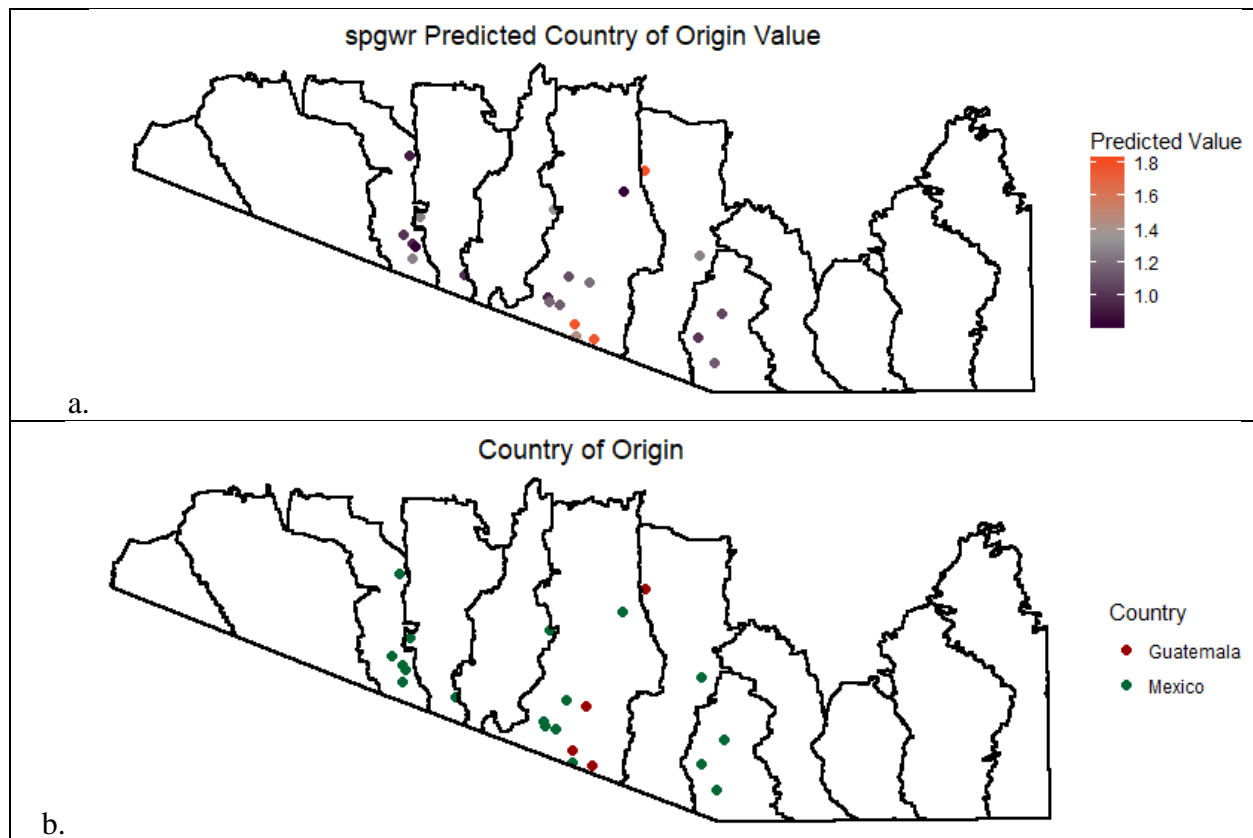
The coefficients for each of the independent variables from the Model 4.2 spgwr analysis is presented in Figure 7.17. The coefficient values indicate the relationship between the variable and the overall predicted value, in this case country of origin. A rise in coefficient values means as the independent variable (i.e. values of PCA1) increases so does the dependent variable (the country of origin coded: Mexico = 1 and Guatemala = 2). Similarly, as the coefficient decreases, so does the country of origin code. Higher coefficients would be more related to individuals from Guatemala and lower coefficients to those from Mexico.

The coefficients of Model 4.2 express a gradient across the landscape with a slight differentiation between the individuals in the western portion of the state (specifically the western-most Cabeza Prieta corridor) to those in the central regions. The exception is with PCOa1, which there appears to be a pocket of higher values in the central San Miguel corridor compared to the rest of the state. While slight, this may mean these individuals have a somewhat more positive relationship with this variable to their country of origin. It is worth investigating the country of origin of these individuals to see if this is the corresponding factor.



**Figure 7.17. Coefficient values for Model 4.2 variables.**

Although coefficient data is useful in understanding the relationship of the biological variables to the country of origin, the predictive power of the GWR model is of most interest here for the future application to the forensic anthropological identification process at the Pima County Office of the Medical Examiner. To evaluate this, the predicted country of origin values can be plotted and compared to the actual values, as in Figure 7.18.



**Figure 7.18. Model 4.2 predicted (a) and actual country of origin values (b) using spgwr.**

There is good overall visual correspondence between the two plots with at least 3 of the 4 Guatemalan individuals correctly classified (coded [a] orange and [b] red). A comparison of the

calculated values is presented in Table 7.10 and can be used to evaluate how well the model is performing. As before, Mexico is coded as 1 and Guatemala is 2.

**Table 7.10. Prediction results for Model 4.2 using spgwr**

<b>Individual</b>	<b>Country of Origin Coded</b>	<b>Predicted Value</b>
1	1	1.145
2	1	0.830
3	1	1.092
4	2	1.768
5	1	1.001
6	1	1.147
7	1	0.961
8	1	0.976
9	1	1.065
10	1	0.947
11	2	1.805
12	1	1.257
13	1	1.264
<b>14</b>	<b>1</b>	<b>1.443</b>
15	1	1.269
16	1	0.995
17	1	1.169
18	1	1.322
19	1	0.905
20	2	1.758
<b>21</b>	<b>2</b>	<b>1.201</b>
22	1	0.823

**Bold denotes greatest inaccuracies**

The largest divergence between the two values seems to be with # 21, where the individual was predicted as 1.2 which is closer to Mexico (1) than their actual country of origin code of 2 for Guatemala. The second largest divergence is individual # 14, whose predicted value was 1.44 and was from Mexico (1). Tests of model accuracy were also conducted and will be presented after the results of the GWmodel analysis of Model 4.2.

## GWmodel

Data Set 2 was then analyzed using the R package GWmodel (Gollini et al. 2015). The two packages are mostly analogous, with similar diagnostics to allow for result comparisons.

The summarized results of the GWR analysis using GWmodel on Model 4.2 are presented in Table 7.11.

**Table 7.11. Diagnostics for Model 4.2 using GWmodel**

<b>Output</b>	<b>Value</b>
<b>Residual Squares</b>	1.485
<b>AICc</b>	24.403
<b>R<sup>2</sup></b>	0.546

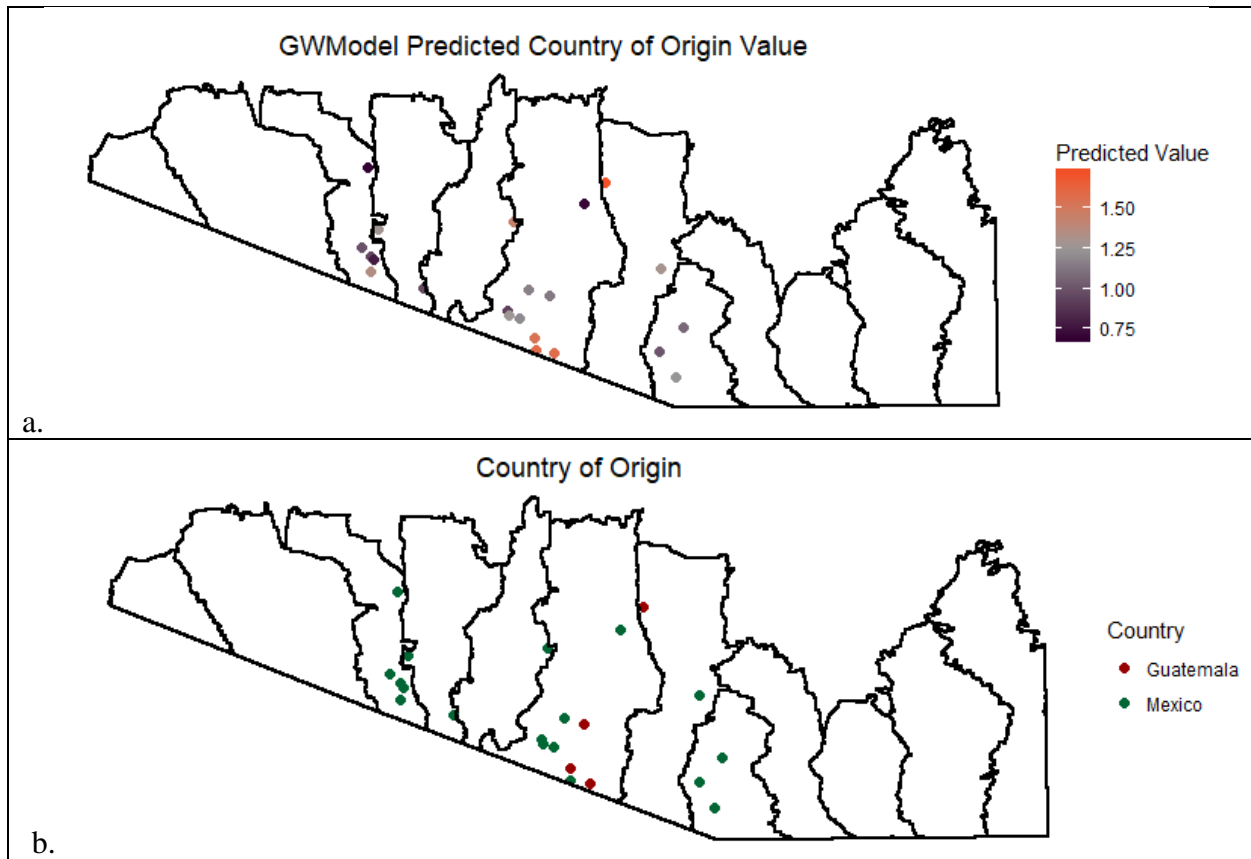
Compared to the spgwr results, the R<sup>2</sup> has decreased slightly while the residual squares has slightly increased. These values indicate a poorer fitting model but being so slight it may just be calculation differences rather than actual differences in model performance.

The coefficient values for each of the independent variables of Model 4.2 were calculated and compared to those derived from the spgwr package. Overall, the coefficient ranges are similar and are presented in Table 7.12.

**Table 7.12. Coefficient ranges calculated using GWmodel and spgwr on Model 4.2**

<b>GWmodel</b>				<b>spgwr</b>		
	<b>Min.</b>	<b>Median</b>	<b>Max.</b>	<b>Min.</b>	<b>Median</b>	<b>Max.</b>
<b>X Intercept</b>	1.005	1.011	1.026	1.004	1.015	1.038
<b>PCA1</b>	0.279	0.290	0.301	0.270	0.290	0.302
<b>PCA2</b>	-0.059	-0.055	-0.037	-0.062	-0.057	-0.032
<b>PCOa1</b>	0.104	0.109	0.120	0.104	0.111	0.124
<b>PCOa2</b>	-0.173	-0.164	-0.143	-0.178	-0.165	-0.130

Since the potential to predict the country of origin of future unidentified recovered migrants is most important for this project, the differences in these values were mapped in Figure 7.19 and the values are presented in Table 7.13.



**Figure 7.19. Model 4.2 predicted (a) and actual country of origin values (b) using GWmodel.**

**Table 7.13. Prediction results for Model 4.2 using GWmodel**

<b>Individual</b>	<b>Country of Origin Coded</b>	<b>Predicted Value</b>
1	1	1.196
2	1	0.699
3	1	1.162
<b>4</b>	<b>2</b>	<b>1.542</b>
5	1	1.004
6	1	1.238
7	1	0.947
8	1	0.954
9	1	1.072
10	1	0.926
11	2	1.714
12	1	1.279
13	1	1.295
<b>14</b>	<b>1</b>	<b>1.584</b>
15	1	1.344
16	1	0.988
17	1	1.267
18	1	1.434
19	1	0.685
<b>20</b>	<b>2</b>	<b>1.577</b>
<b>21</b>	<b>2</b>	<b>1.130</b>
22	1	0.776

**Bold denotes greatest inaccuracies**

The results of the GWmodel are visually very close (Figure 7.19) but numerically less so (Table 7.13), especially when compared to the spgwr results (Table 7.10). One way to test the accuracy of the model is to look at the prediction and prediction uncertainty accuracies.

Prediction accuracy is measured by the root mean squared prediction error (RMSPE) and the mean absolute prediction error (MAPE), both of which should tend to zero. Prediction uncertainty accuracy is measured by the mean and standard deviation (SD) of the prediction z-score data (mean ZS and SD ZS, respectively) ... where for unbiased prediction standard errors, the mean and SD of the z-scores should tend to zero and unity, respectively (Gollini et al. 2015: 40).



The results of these three tests on the prediction values derived from both the GWmodel and spgwr packages is presented in Table 7.14.

**Table 7.14. GWR accuracy comparisons for Model 4.2**

	<b>RMSPE</b>	<b>MAPE</b>	<b>Mean ZS</b>	<b>SD ZS</b>
<b>GWmodel</b>	0.336	0.270	0.003	0.235
<b>spgwr</b>	0.257	0.191	-0.040	0.235

These accuracy tests confirm the superficial differences in the performance of the two R packages when comparing the predicted values to the actual country of origin codes. While not substantially different, the spgwr package performed better than the GWmodel at predicting the country of origin of individuals based on their recovery location and skeletal attributes. The spgwr package also had higher  $R^2$  values, but the corrected AICc values indicate a better fit under the parameters set by the GWmodel. To further investigate the properties of GWR in these test conditions, Data Set 3, with individuals with less accurate GPS data, was evaluated using both R packages.

#### Model 4.3

GWR analyses using both R packages were performed with Data Set 3 to test the effectiveness of the model with samples containing less than precise GPS data; in this case within 10 miles or 15 kilometers. The location of the remains of the individuals comprising Data Set 3 are presented in Figure 7.20 labeled by country of origin and Figure 7.21 by sex.

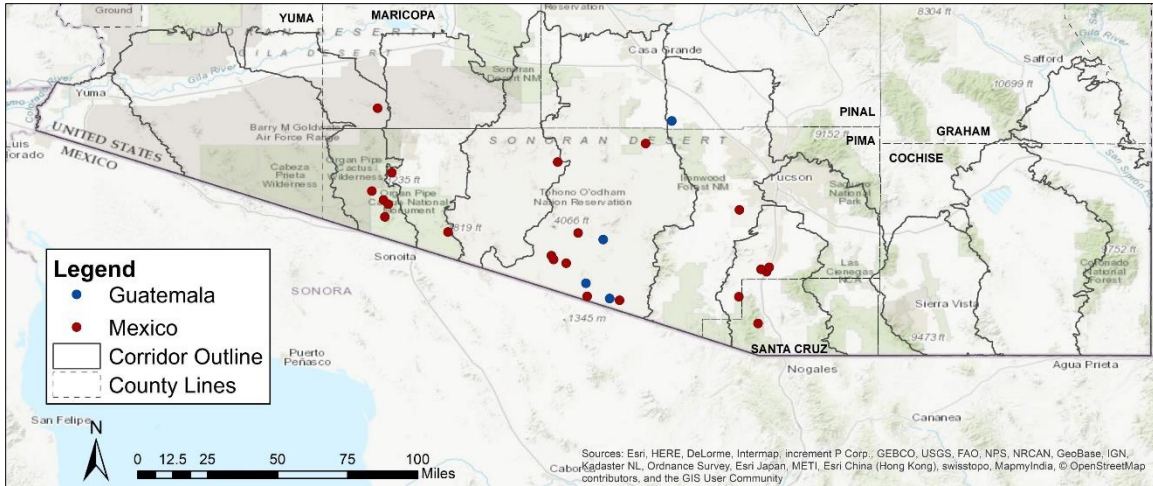


Figure 7.20. Data Set 3 (n = 25) labeled by country of origin.

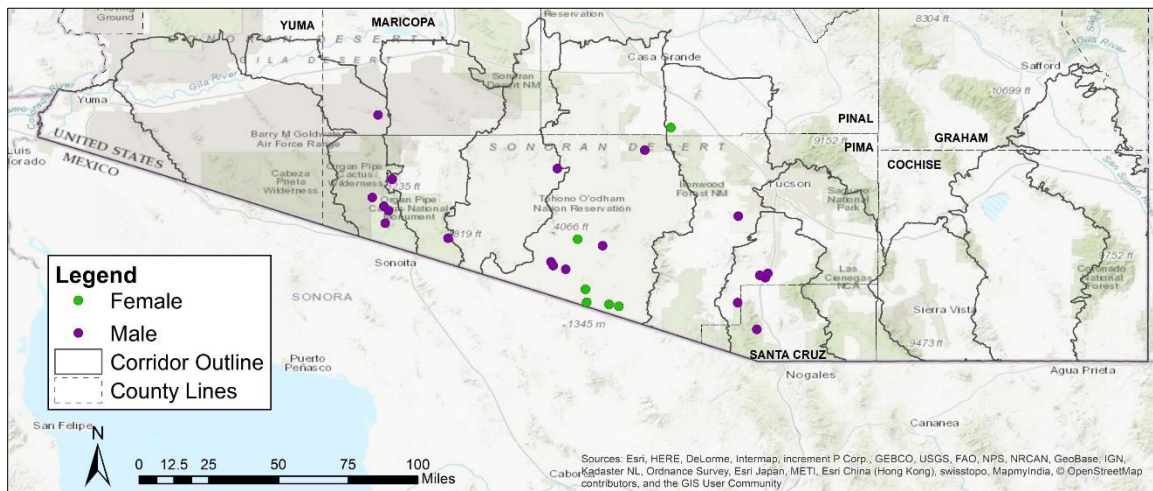


Figure 7.21. Data Set 3 labeled by sex.

spgwr

The output of the analysis using spgwr on Model 4.3 is presented in Table 7.15.

Table 7.15. Diagnostics for Model 4.3 using spgwr

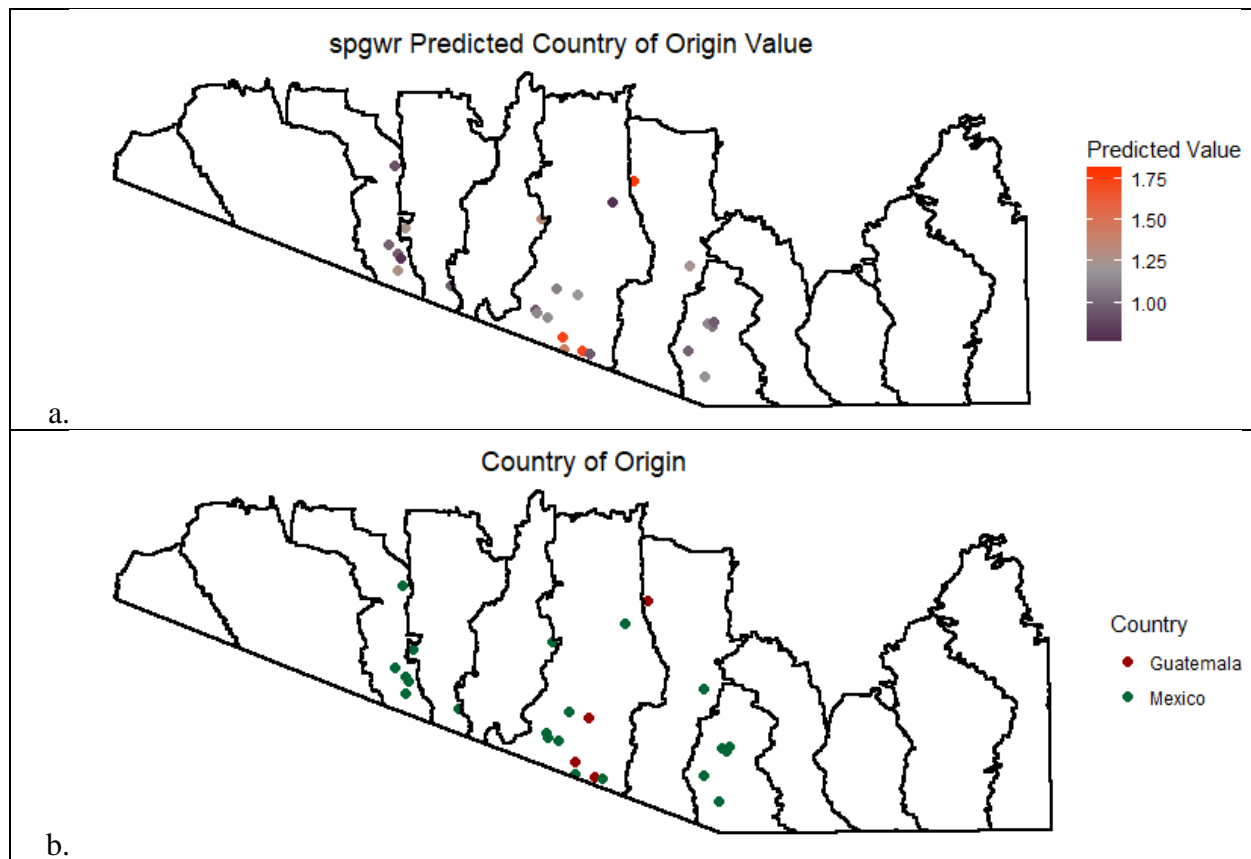
Output	Value
Residual Squares	1.545
AICc	20.854

$R^2$

0.540

Compared to Model 4.2 (Table 7.9), the results of the GWR analysis using spgwr indicate the models are performing similarly. The AICc decreased from 25.566 to 20.854, indicating a better fitting model as it is  $> 3$ . The residual squares and  $R^2$  values are also similar which suggests the addition of less precise data does not decrease the relative power of the analysis, at least at this stage.

The maps of the predicted versus actual countries of origin values are presented in Figure 7.22, while the corresponding numerical output is presented in Table 7.16.



**Figure 7.22. Model 4.3 predicted (a) and actual country of origin values (b) using spgwr.**

**Table 7.16. Prediction results for Model 4.3 using spgwr**

<b>Individual</b>	<b>Country of Origin Coded</b>	<b>Predicted Value</b>
1	1	1.154
2	1	0.792
3	1	1.109
4	2	1.750
5	1	0.988
6	1	1.173
7	1	0.963
8	1	0.964
9	1	1.066
10	1	0.939
11	2	1.792
12	1	1.261
13	1	1.251
<b>14</b>	<b>1</b>	<b>1.448</b>
15	1	1.279
16	1	0.986
17	1	1.137
18	1	1.343
19	1	0.916
20	2	1.721
<b>21</b>	<b>2</b>	<b>1.206</b>
22	1	0.806
23	1	1.133
24	1	0.952
25	1	0.959

**Bold denotes greatest inaccuracies**

Again, there is good correspondence between the actual and predicted values for country of origin using this model, especially for samples # 23-25 (the three added Mexican individuals) (Table 7.16). The relative accuracy of the spgwr results on Model 4.3 will be assessed following the results of the GWmodel analysis.

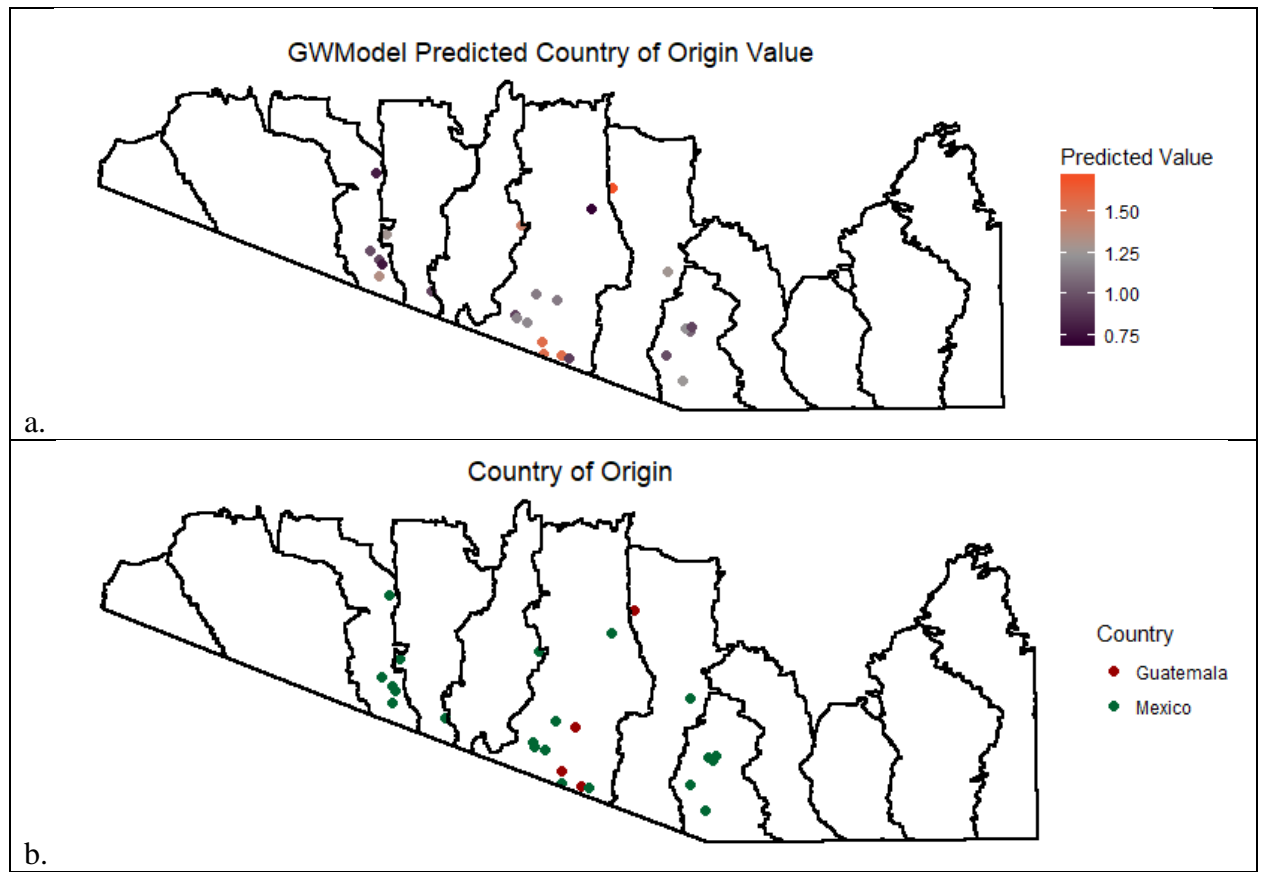
## GWmodel

Data Set 3 was then analyzed using the R package GWmodel. The diagnostic results (Table 7.17) are similar to those derived from the spgwr package, with only slight differences in each diagnostic value. Again, the model is explaining just over half of the variance seen in the data ( $R^2$ ), which is more than the global model at 0.432 (Table 7.7).

**Table 7.17. Diagnostics for Model 4.3 using GWmodel**

<b>Output</b>	<b>Value</b>
<b>Residual Squares</b>	1.531
<b>AICc</b>	21.250
<b><math>R^2</math></b>	0.544

The maps and tables for the comparison between predicted and actual values are presented in Figure 7.23 and Table 7.18, respectively.



**Figure 7.23. Model 4.3 predicted (a) and actual country of origin values (b) using GWmodel.**

**Table 7.18. Prediction results for Model 4.3 using GWmodel**

<b>Individual</b>	<b>Country of Origin Coded</b>	<b>Predicted Value</b>
1	1	1.193
2	1	0.704
3	1	1.142
4	2	1.567
5	1	0.981
6	1	1.258
7	1	0.958
8	1	0.956
9	1	1.073
10	1	0.931
11	2	1.703
12	1	1.280
13	1	1.276
<b>14</b>	<b>1</b>	<b>1.559</b>
15	1	1.339
16	1	0.984
17	1	1.183
18	1	1.438
19	1	0.799
20	2	1.568
<b>21</b>	<b>2</b>	<b>1.144</b>
<b>22</b>	<b>1</b>	<b>0.775</b>
23	1	1.218
24	1	0.926
25	1	0.947

**Bold denotes greatest inaccuracies**

The coefficient ranges for both geographically weighted regression platforms using Data Set 3 are presented in Table 7.19. The values are very similar for the minimum, median, and maximum values and in some cases identical. But, overall differences in each package’s predictive powers is evident. The comparative descriptive statistics from testing the accuracy of the model predictions is presented in Table 7.20.

**Table 7.19. Coefficient ranges calculated using GWmodel and spgwr on Model 4.3**

GWmodel	GWmodel			spgwr		
	Min.	Median	Max.	Min.	Median	Max.
<b>X Intercept</b>	0.953	0.976	0.987	0.954	0.972	0.981
<b>PCA1</b>	0.267	0.278	0.281	0.268	0.277	0.281
<b>PCA2</b>	-0.036	-0.025	-0.020	-0.031	-0.025	-0.020
<b>PCOa1</b>	0.124	0.132	0.136	0.124	0.131	0.133
<b>PCOa2</b>	-0.182	-0.175	-0.156	-0.180	-0.173	-0.158

**Table 7.20. GWR accuracy comparisons for Model 4.3**

	RMSPE	MAPE	Mean ZS	SD ZS
<b>GWmodel</b>	0.308	0.241	0.005	0.227
<b>spgwr</b>	0.249	0.185	-0.001	0.227

The lower root mean squared prediction error (RMSPE) and mean absolute prediction error (MAPE) values indicate more accurate country of origin predictions using the spgwr package (Table 7.20). The lower mean and standard deviations of the prediction z-scores signify fewer differences in the average overall predicted and actual country of origin code values. As with Model 4.3, although the coefficient values were similar between the two packages, the prediction accuracy diagnostics are better with spgwr. Gollini et al. (2015) provide an in-depth comparison of the two packages which may explain the computational differences seen between GWmodel and spgwr.

### ArcMap

Model 4.3 was also tested using ArcMap (ESRI 2016) by placing a 50 km<sup>2</sup> hexagonal grid over the study area and correlating the data points to the hexagon in which they fell, similar to that performed in the ESDA. Then, geographically weighed regression analyses were run



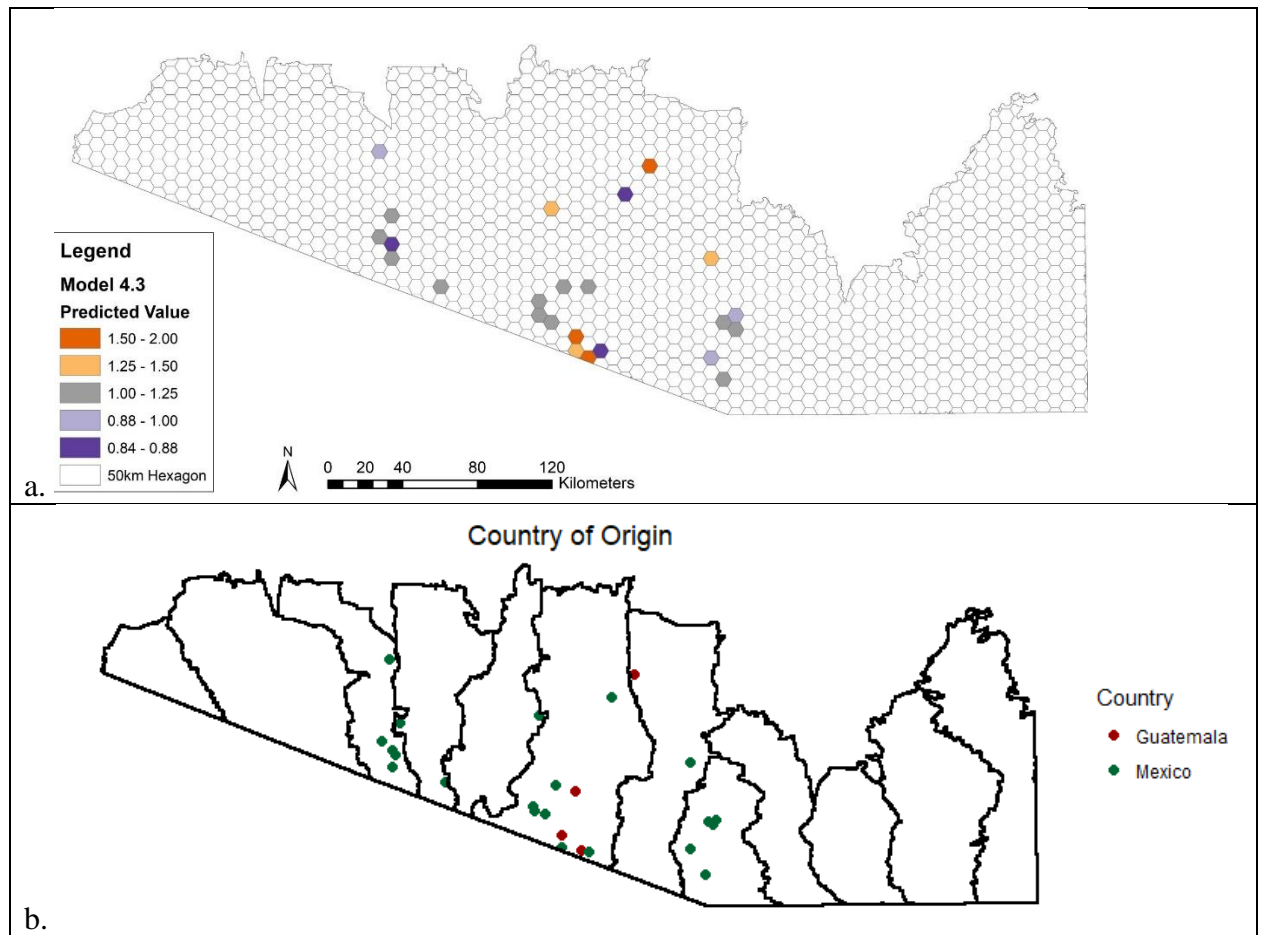
using the grid's intercept value with the point data. The resulting output table for Model 4.3 is presented below (Table 7.21).

**Table 7.21. Diagnostics for Model 4.3 with 50km hexagonal grid**

<b>Output</b>	<b>Value</b>
<b>Residual Squares</b>	1.360
<b>AICc</b>	33.728
<b>R<sup>2</sup></b>	0.595

While the R<sup>2</sup> value has increased from 0.54 from the R package results (Tables 7.15 and 7.17) to just under 60% of the variance explained (R<sup>2</sup> = 0.595), the corrected AIC value has also risen from 20.854 to over 33. This suggests a poorer fitting model in general but, it is explaining more of the variance than before.

The predictive value map is presented in Figure 7.24 and the numerical results are available in Table 7.22.



**Figure 7.24. Model 4.3 predicted (a) and actual country of origin values (b) using ArcMap 50km hexagonal grid.**

**Table 7.22. Prediction results for Model 4.3 using ArcMap 50km hexagonal grid**

<b>Individual</b>	<b>Country of Origin Coded</b>	<b>Predicted Value</b>
1	1	1.016
2	2	1.827
3	1	1.005
4	1	1.180
5	1	0.947
6	1	1.428
7	1	0.843
8	2	1.872
9	1	0.950
10	1	1.005
11	1	1.277
12	1	1.034
13	1	1.133
14	1	1.064
15	1	0.849
16	1	0.849
17	1	1.239
18	1	1.186
19	1	1.142
20	1	1.319
<b>21</b>	<b>2</b>	<b>1.212</b>
22	1	1.232
23	1	0.950
24	1	0.885
25	2	1.797

**Bold denotes greatest inaccuracies**

As with the previous two models, the main outlier is individual # 21, with a prediction value closer to 1 (for Mexico) than their actual code of 2 for Guatemala. The results of the accuracy of this hexagonal grid transformation is presented in Table 7.23.

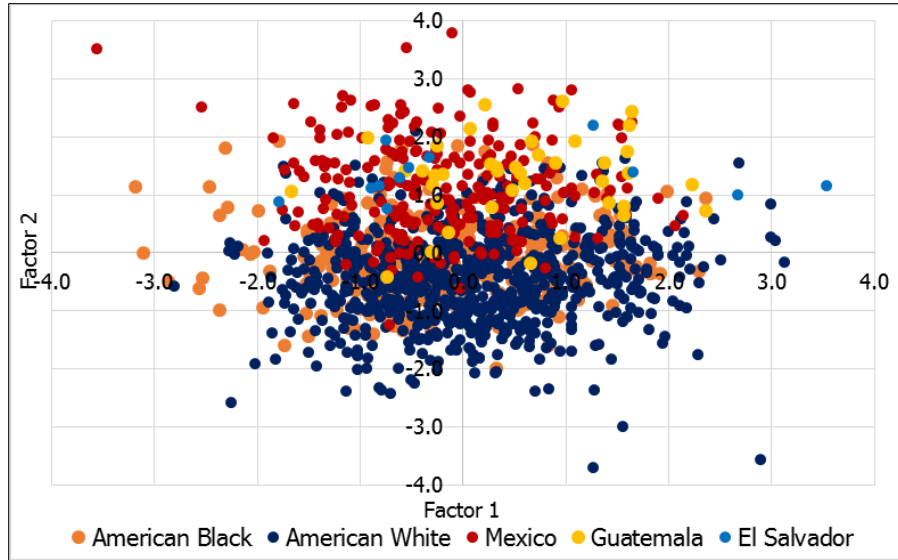
**Table 7.23. GWR accuracy comparisons for Model 4.3 using ArcMap hexagonal grid**

	<b>RMSPE</b>	<b>MAPE</b>	<b>Mean ZS</b>	<b>SD ZS</b>
<b>ArcMap</b>	0.235	0.171	-0.007	0.211

While these results are an improvement over the R packages, the subsequent application of this platform becomes difficult with the addition of more data. As each hexagon must take on the attributes of either a single point or a combination of the points (usually the mean), the addition of more samples means multiple points within a single hexagon can occur. These results may be of interest in understanding spatial patterning of migrant deaths but are less desirable when attempting to predict new individual data points. The spgwr platform is preferred for the purposes of this project as it provides more accurate prediction results when analyzing the country of origin of individuals using cranial data.

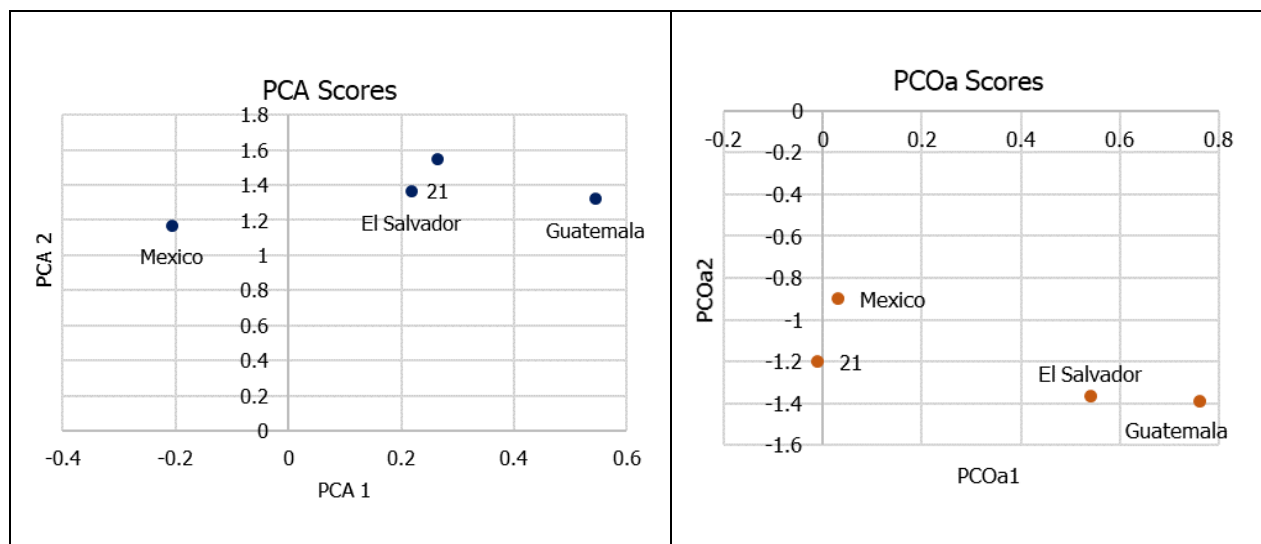
### **Discrepancy Investigation**

Two individuals (# 14 and # 21) were consistently over- and under-predicted, respectively, when using the R packages; while # 21 was underpredicted using all three platforms. Several elements may be responsible for these discrepancies, including the nature of the current sample set. When revisiting the plot of the samples by the first two PCA scores labeled by their country of origin (Figure 7.25) the amount of overlap amongst the individuals from different countries is apparent. The fact that the group means separate well (Figure 6.3) means that when applied to the geobiological model, the first two PCA scores may do well at predicting the country of origin for individuals whose values are close to the group means, but less so for outliers.



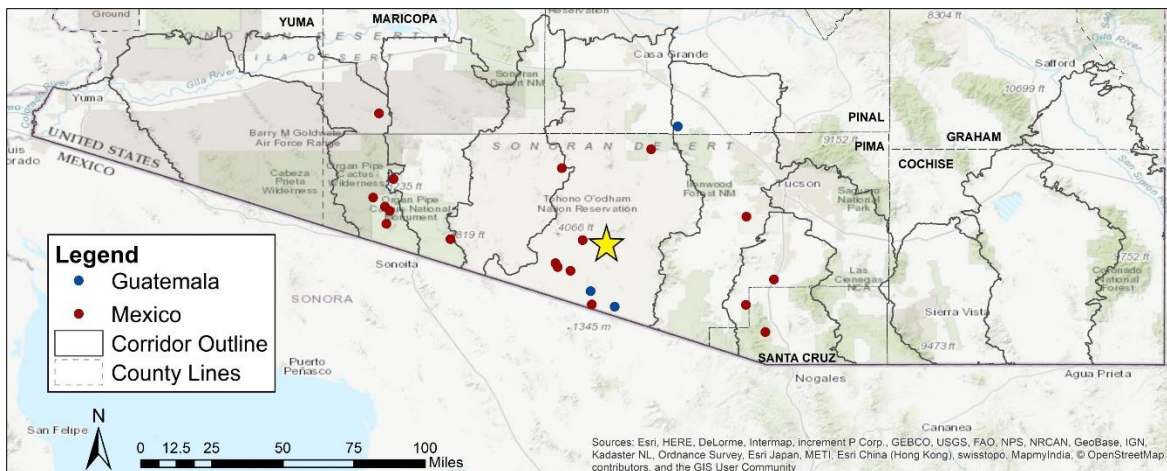
**Figure 7.25. Plot of first 2 factors labeled by country of origin.**

This overlap could explain the consistently poor prediction for individual # 21. Although Guatemalan, their prediction scores were closer to 1 (Mexican). When their PCA and PCOa scores were compared to the respective country means (Figure 7.26) the PCA scores are closer to the dropped El Salvadorans and the PCOa scores to that of Mexico.



**Figure 7.26. Individual # 21 PCA and PCOa scores compared to all country means.**

Revisiting the geographic position of sample # 21 (Figure 7.27, starred), it is closest to a Mexican individual, but still in center of the state like the other Guatemalans. This shows the importance the biological aspects of the model play into the prediction, and thus the need for increased sample sizes.



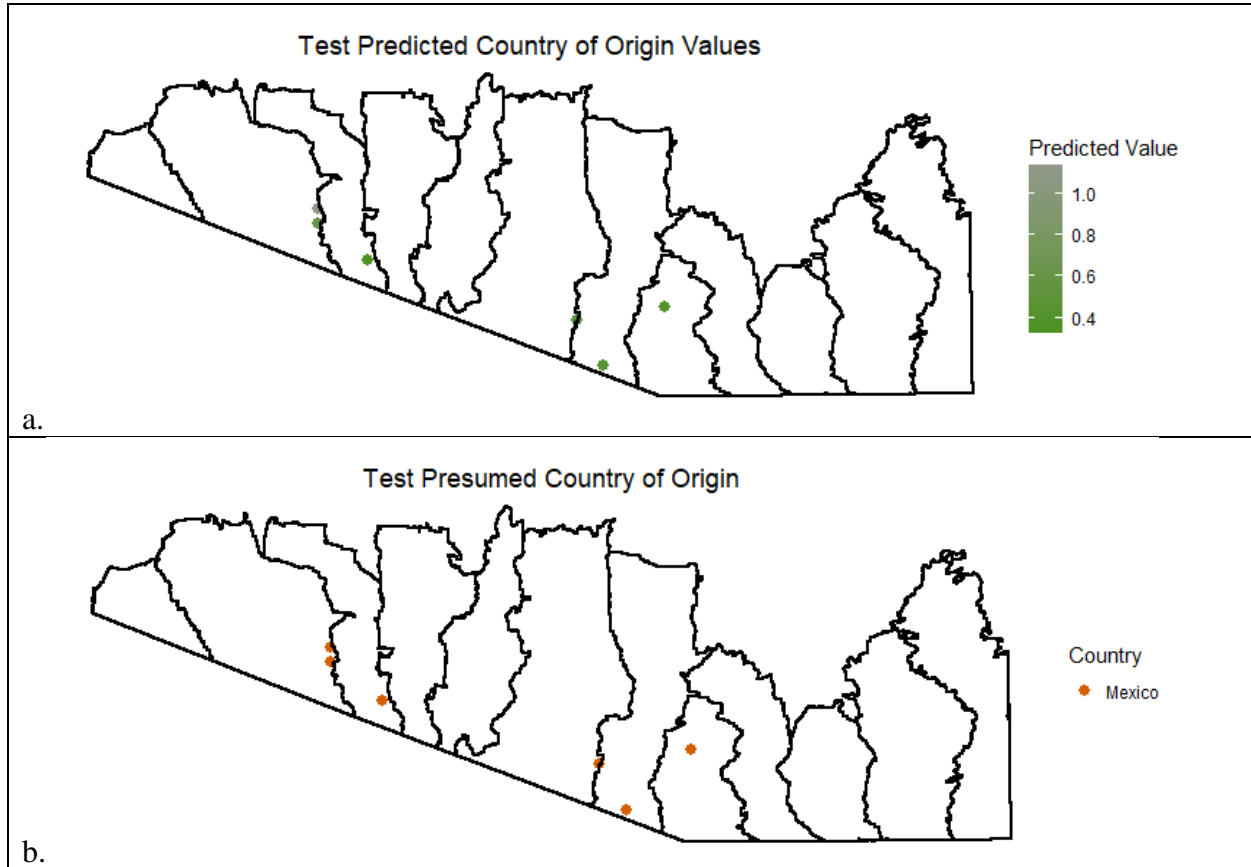
**Figure 7.27. Geographic position of individual # 21 (star).**

## Model Testing

Applying the geobiological model is different than traditional regression, as a single equation is not generated. Rather, GWR “allows to explore and analyse the spatial distribution of variables by fitting a regression equation to every point in space and thus study spatial heterogeneity and spatial relationships” (Blachowski 2016; 1002). The R package *spgwr* provides a user-friendly application of predicting unknown dependent variables of a data set via user-developed parameters (such as those created in the previous sections).

To test the final model, a subsample of eight individuals with strong presumptive identifications was used. These individuals are currently pending identification and therefore

were not applied to the original model. All eight are presumed to be Mexican nationals. The results of the prediction are presented in Figure 7.28 and Table 7.24.



**Figure 7.28. Predicted country of origin values of test dataset (a) and actual countries (b) using spgwr.**

**Table 7.24. Prediction results for test samples using spgwr**

<b>Individual</b>	<b>Presumed Country of Origin Code</b>	<b>Predicted Value</b>
1	1	0.514
2	1	1.119
3	1	0.540
4	1	0.343
5	1	0.399
6	1	0.774
7	1	0.586
8	1	0.445

The result of the predicted country of origin of the eight presumed Mexican individuals has a much wider range than produced in the final GWR model. All the individuals have country of origin codes of 1 (for Mexican) and while only samples # 2 and # 6 are relatively close to this value, the others are at least closer to 1 than 2 (the coded value for Guatemalans). This would likely result in the prediction of these individuals as Mexicans as opposed to another country of origin. One reason for this may be the spatial placement of these individuals, including two within a new corridor (the western Goldwater corridor) and one on the southern edges of the Sasabe Corridor (Figure 5.3). These results highlight the need for a large and diverse sample set to develop threshold values for multiple countries of origins.

The diagnostic values of the test sample versus the final GWR model are presented in Table 7.25.

**Table 7.25. Prediction accuracy comparison for test model and Model 4.3**

	<b>RMSPE</b>	<b>MAPE</b>	<b>Mean ZS</b>	<b>SD ZS</b>
<b>spgwr Test</b>	0.472	0.439	0.611	0.393
<b>Model 4.3</b>	0.249	0.185	-0.001	0.227



The increase in error values is likely the product of the location of the test samples in areas where there were none in the original model. Also, biological variation not seen in the reference data may also be affecting the prediction values of the test individuals. As cases recovered from different locations are added to the model database, the accuracy should increase in these new areas. Therefore, the model must remain a dynamic resource to incorporate more of the PCOME's jurisdiction and cases of individuals from other countries.

## **Conclusions**

The presence of positive spatial autocorrelation is shown in the exploratory spatial data analyses and provides a promising foundation for the employment of predictive modeling of the data. The global model which best explains an individual's country of origin includes both craniometric and macromorphoscopic data reduction values, and further supports their use in the assessment of ancestry from cranial remains. The combination of this model with the recovery location of the individual using methods of geographically weighted regression produces a useful method to aid in the identification process at the Pima County Office of the Medical Examiner in Tucson, Arizona in the near future.

## REFERENCES

## REFERENCES

- Charlton M, Fotheringham S, Brunson, C. 2009. Geographically weighted regression. White paper. National Centre for Geocomputation. National University of Ireland Maynooth.
- Esri Inc. 2016. ArcMap 10.5.1 [Computer software]. Redlands, CA: Esri Inc.
- Fotheringham AS, Brundson C, Charlton M. 2002. Geographically weighted regression: The analysis of spatially varying relationships. Chichester, West Sussex, England: Wiley.
- Gollini I, Lu B, Charlton M, Brunson B, Harris P. 2015. GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models. *Journal of Statistical Software*, 63(17), 1-50.
- Lloyd CD. 2007. Local models for spatial analysis. Boca Raton: CRC/Taylor & Francis.
- Martinez DE, Reineke R, Rubio-Goldsmith R, Anderson BE, Hess GL, Parks BO. 2013. A Continued Humanitarian Crisis at the Border: Undocumented Border Crosser Deaths Recorded by the Pima County Office of the Medical Examiner, 1990–2012. Tucson: Binational Migration Institute, University of Arizona.
- Williams GJ. 2011. *Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery, Use R!*, New York: Springer.
- Vogelsberg CCM. 2018. Assessing the Spatial Patterns of Undocumented Border Crosser (UBC) Deaths in the Southern Arizona Desert. Proceedings of the 70<sup>th</sup> Annual Meeting of the American Academy of Forensic Sciences; Feb 19-24, Seattle, WA.

## CHAPTER 8. DISCUSSION AND CONCLUSIONS

### Introduction

Over the past two decades, more than 3,000 individuals have died while trying to cross the United States-Mexico border. This ongoing humanitarian crisis has led to the partnership of several organizations in efforts to understand and mitigate the situation. Included are collaborations between nonprofit groups such as Humane Borders and the Colibrí Center for Human Rights, and the Pima County Office of the Medical Examiner (PCOME) in Tucson, Arizona. Official counts and recovery locations of undocumented border crosser recoveries have been made since 2000, and as of the end of 2017, 992 (or 36%) of the 2,816 cases analyzed by the PCOME remain unidentified (PCOME 2017).

Discrepancies are apparent between the number of recoveries made in southern Arizona and the number of missing persons reported (Colibrí 2018). The purpose of this project was to explore the geospatial and biological patterns of identified migrants to help understand more about the unidentified. By combining the skeletal information of identified deceased undocumented border crossers and the recovery location of their remains in the southern Arizona desert, the emerging patterns can aid in the prediction of country of origin of newly recovered individuals. This will hopefully help close the gap between the number of missing individuals and the number of unidentified remains. This application of forensic anthropological and geographic methods also highlights such sociological theories of international migration as migrant networks and cumulative causation. The results of this project provide new ways to understand the social, political, and biological connections involving undocumented border crossers. It also provides a new avenue for identifying future unknown remains and returning them to their families.

## **Biological Data**

### *Craniometric And Macromorphoscopic Data Analysis*

The biological data analyses conducted in this project focused on the question of whether there were any correlations between an individual's craniometric and macromorphoscopic features and their country of origin. The results of analyses using both types of biological data (separately and combined in random forest models) correspond well with previous research using a similar dataset to assess the use of skeletal data for classifying individuals by their geographic origins (Hefner 2007; Spradley et al. 2008; Spradley 2013; Spradley 2014; Ross et al. 2014; Tise 2014; Spradley & Jantz 2016). These includes the results of both the pooled and un-pooled group analyses using the craniometric and macromorphoscopic data.

The current biological dataset of individuals from countries other than Mexico is quite small (in the case of El Salvadorans and Guatemalans) or absent (in the case of Hondurans or Ecuadorans). As the PCOME has identified individuals from 14 countries which would be classified as Hispanic, the need for representatives from these populations is evident. Although many of the individuals recovered in the southern Arizona desert are from the countries used in this study as opposed to South American and Caribbean nations, an adequate representation of the possible variation is still necessary for more accurate predictions.

### *Random Forest Modeling Analyses*

The results of the RFM further highlight the contribution of both craniometric and macromorphoscopic traits to the biological variation observed in individuals from several ancestral backgrounds. Although the RFM analyses had the highest classification rates for Hispanic individuals using a 3-group system (Hispanics, American Whites and American Blacks), its application to predicting a more specific country of origin produced considerably

high error rates. In its current state, RFMs may be useful in assessing if an individual is Hispanic or not, after which the craniometric PCA and macromorphoscopic CAP methods would be run separately. Once more individuals from the Hispanic subgroups are obtained, the classification error rates would presumably decrease and assist in the prediction of an individual's country of origin using random forest models. As of now, these RFM results support the geobiological model's incorporation of both metric and nonmetric measures of ancestry for predicting group classification.

### Future Biological Studies

As more identifications are made and the number of individuals from Hispanic countries increases, the variables used in the biological data analyses should be reassessed. In that way the biological variables applied to the geobiological model will continue to reflect the features of the cranium which best differentiate the subgroups. It is assumed that as new countries are added to the sample of identified remains, more refined group differentiation will be possible and needs to be reflected in the geobiological analysis.

It would also be beneficial to investigate how the results of the random forest modelling can be combined with the geographically weighted regression analyses to predict the country of origin of unknown remains. As both methods combine the classification powers of metric and nonmetric cranial traits, it would be more efficient to directly apply the GWR to the RFM procedures rather than taking the extra steps of performing biological data analyses separately and then combining the results. Investigations into the R code capable of combining these two methods would be advantageous to this project.

The application of the top performing variables from the RFM results to the geobiological model should be explored first since the use of PCA analyses on truly unidentified

individuals may be inappropriate. This method refinement could also potentially increase the number of individuals available to create the geobiological model as fewer variables are required. Several individuals had to be excluded from the original PCA and CAP procedures due to missing data and this could be circumvented if other variables were designated as having higher importance via forest building.

## **Geographic Data Analysis**

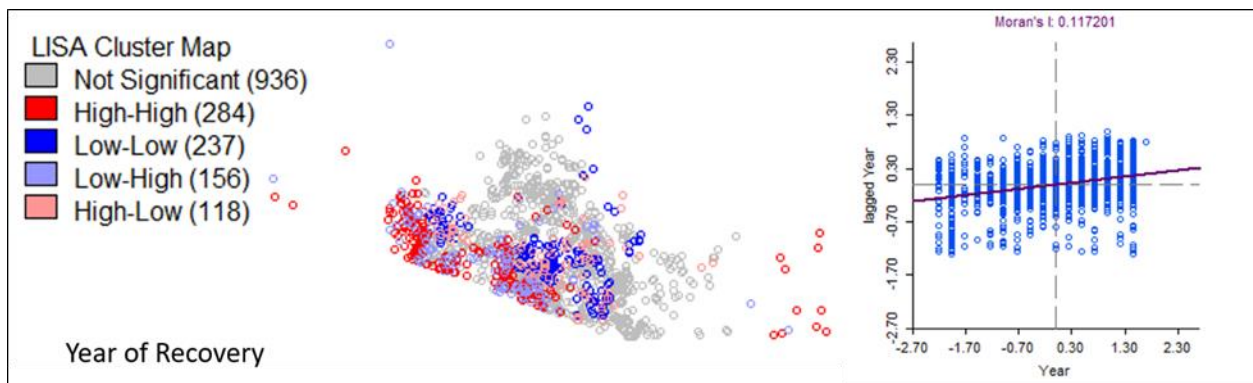
### *Exploratory Spatial Data Analysis (ESDA)*

The results of the ESDA procedures highlight several variables with positive spatial autocorrelation at both the global and local levels. The implications of the results for year of recovery, identification status, and the sex and country of origin of identified individuals are worth more consideration as they illustrate previously-discussed political and theoretical themes involving international migration in the Greater Southwest. These include the effects of United States Border Patrol enforcement measures and evidence of migrant social networks as illustrated in the recovery location of individuals who died while trying to clandestinely cross the United States-Mexico border.

### Year of Recovery

The apparent shift in recovery locations throughout the study timeframe seems to reflect a shift in the routes individuals took, and ultimately died along, from the center of the state to the more remote western areas (Figure 8.1, below for reference). This shift correlates well with the well-established phenomenon in the Greater Southwest immigration literature deemed the “funnel effect.” This was the result of the previously-discussed US Border Patrol Operations “Safeguard” and “Gatekeeper” in the mid-1990’s which made traditional routes of migration

riskier, and migrants were “funneled” to more remote desert areas (Rubio-Goldsmith 2006). While migrants crossed (and died) in the more central crossing points at the turn of the 21<sup>st</sup> century, they had to find alternative routes in the western and eastern portions of the state as these traditional entry points became more militarized (Télez 2007). More remote regions have had an increase in recoveries in later years as this shift in travel amongst migrants occurs.



**Figure 8.1. LISA Cluster map and Moran's scatter plot for year of recovery variable for all individuals.**

Another interpretation of the emerging year of recovery local cluster patterns is that it is actually capturing relative recovery efforts by location over time. As migrant routes shift so do enforcement and humanitarian effort. This means that more people would be in the area to come upon migrant remains. Migrant travel areas within the western corridors may only be highlighted in later years due to more intense search and recovery of individuals in this region, while really having been in use throughout the last 17 years by migrants. Other researchers have hypothesized this possibility but dismissed it due a consistent number of remains being found over the decade despite an increase in the numbers USBP agents (Gordano & Spradley 2017). However, they also reported an increase over the years in the proportion of recoveries being

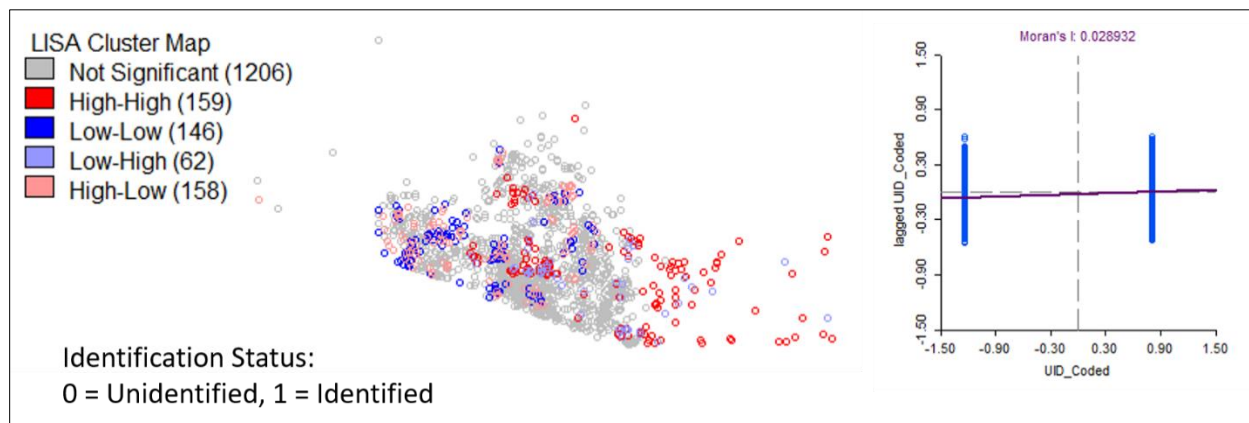


skeletal remains (Gordano & Spradley 2017). This could indicate that these are older routes which are actually less used now, as indicated in the relative number of skeletal remains being recovered.

A more nuanced examination of the condition of the recovered remains per cluster in relation to their year of recovery may be useful. During times of high enforcement, or more intense search and recovery efforts for UBCs, it would be expected that deceased individuals are found closer to their time of death or within a shorter time frame. During times of low enforcement, or less intense recovery efforts, a longer time would elapse between when an individual died and when their remains were recovered. When patrolling increases again in the area, the amount of skeletal or mummified remains would presumably increase in these areas. Further investigation into the post-mortem interval (PMI) of the remains would be necessary to properly assess how long the area was in use by migrants, and if this interpretation is correct. If the skeletonized remains were mostly given PMIs of greater than 5 years, this would support the line of reasoning that they are more established corridors that are just now being patrolled more heavily. Yet, if the remains had shorter PMIs of 6 months to one year (very possible in the hot, dry Sonoran Desert) this would indicate the establishment of new routes coinciding with increased recovery efforts in areas. A combination of the two may indicate more established routes which have been used throughout the years and are just now being investigated more. A comparison of the date of border fence erection and positive autocorrelation temporal clusters may also shed light on the relationship between the effects of physical barriers on migratory routes. Whatever the reason for these clusters, they appear to be areas of migrant activity that should be investigated more closely.

## Identification Status

The spatial autocorrelation present in identified versus unidentified remains (as shown in Figure 8.2 again for reference) may reflect either differences in investigatory jurisdictions (county Sheriff's versus US Border Patrol, etc.) or a correlation with year of recovery.



**Figure 8.2. LISA cluster map and Moran's scatter plot for identification status variable for all individuals.**

As more identifications have been made in the eastern and central portions of the state, the cluster maps may be exposing relative efforts put into identifying UBC remains by these jurisdictions. More likely, the year of recovery may influence the identification of an individual; with older cases potentially having higher rates of identification as they have had more time to be investigated. Following the breaks associated with the high and low the year of recovery clusters, the total number of identified cases from “older” years (2001-2010) totals 615 individuals while 436 are from 2011 – 2017 (Table 7.3). Older cases seem to have been identified more than newer cases and this may line up well with the year of recovery clusters.

### Sex and Country of Origin

The LISA analyses on the sex and country of origin of identified individuals also showed areas of positive spatial autocorrelation. The likely source of these results is the fact that the majority of the PCOME cases are Mexican males (PCOME 2017). However, when all Mexican individuals (both males and females) were removed from analysis, clustering still occurred. While this group removal greatly diminished the sample set (from  $n = 1051$  to 175), there is still (if not more) overall positive spatial autocorrelation amongst the individuals from countries other than Mexico.

This supports previously discussed sociological and cultural theories of network migration (Massey & España 1987; Mullan 1989). These theories propose that people from the same country or area tend to share information and thus should travel along similar routes; whether it is through coyote choice or a collective knowledge about which routes had been successful for others. This has been documented in both social and ethnographic research and now in the locations in which they are recovered when their trips were unsuccessful (Massey & España 1987; Massey & Espinosa 1997; Curran & Rivero-Fuentes 2003; Téllez 2007).

Compared to the country of origin analysis, it appears that the sex of a migrant is less important regarding the route on which they died. When all individuals were analyzed for correlations amongst men and women, the number of significant clusters totaled around 15%. When the Mexican individuals were removed, only about 6% of the samples were significant. The country an individual came from seemed to be more important than their sex when considering where their remains were recovered. In each analysis, the Moran's  $I$  is only slightly positive, but marginally more so when all individuals are included. In the context of migrant network theories, this makes sense as knowledge is usually passed through cultural groups, in this case home country. But the presence of some autocorrelation amongst the sexes also seems

to support the engendered studies of international migration and the role sex plays in route choice (Curran & Rivero-Fuentes 2003; Téllez 2007).

The research question prompting the geographic data investigations into spatial correlations between an individual's demographic information and their recovery location can also be answered affirmatively. The attributes of a migrant, including their sex and where they are from does influence their travel route and subsequent recovery location if they unfortunately die while clandestinely crossing the United States-Mexico border.

### Future Geographic Studies

Future studies should make more nuanced investigations into the positively significant clusters regarding when the remains were recovered. This would include the relative body condition of the identified and unidentified remains to see if individuals found closer to their time of death have higher rates of identification. The PCOME has reported that identification rates have decreased over the last decade, while the postmortem interval (PMI) of remains has increased (PCOME 2017). This may support the hypothesis that long-used migration routes in the western portion of the state are only being found more recently. If the cases are a mix of skeletonized and fleshed remains, then it is possible that these are long established routes. The PMI of the skeletal remains would provide further insight into the amount of time between death and recovery and could be correlated to the length of time the area was being used by migrant travelers.

### **Geobiological Model**

The current success of the geobiological model at predicting the country of origin of Mexican and Guatemalan individuals is promising for its direct application to the forensic

identification process at the PCOME in the near future. There is potential for combining the regressive powers of the biological and geographic data to predict where an unknown individual is from to facilitate a more efficient identification, at least in this southern Arizona example.

Several considerations must be made, though, as the model is currently in an initial stage. At present, only two countries are included (Mexico and Guatemala) and though tests using similar cases have proven successful, the addition of more data to the model's structure is necessary. The limitations of the geobiological model must be considered when applied by PCOME researchers. Additionally, cut-off points for differentiating between predicted country codes must be decided. At present, it seems the value of 1.5 (in-between the country codes) is sufficient, but the addition of more individuals, and more variation, may require further refinement. As more samples are added the prediction groups may become more robust and require a less subjective prediction decision. Until then this must be taken into consideration.

How to deal with more than two country groups must also be investigated as well as corresponding cut-off value decisions. The geobiological model parameters are meant to be dynamic and grow with the relative reference data. The results have yet to be tested in a real-world setting, but once the presumptive identifications presented in the testing phase of the geobiological model are confirmed, the usefulness of this model will be seen. It is hoped that the model can be applied soon to the PCOME procedures.

Further studies into how to increase the number of individuals available to create the geobiological model should also be explored. These could include reducing the biological data necessary or incorporating cases with less precise GPS data. While it is ill-advised to rely on the results of studies using inaccurate geospatial data, the level at which this may be for this study has not been determined. Additionally, if an appropriate data transformation can be created, like

that of the hexagonal grid placement, the threshold for GPS accuracy may be lowered.

Discussions with geographers and other experts in the field of geographic information systems may shed light on the suitability of these explorations.

### **Concluding Remarks**

The results of this research support known associations described in both biodistance and international immigration literature. Included are the correlations between an individual's skeletal morphology and their geographic origins, as well as migrant network theories. The positive spatial autocorrelation exhibited in the attributes of the undocumented border crossers recovered in the southern Arizona desert presents a new set of data to corroborate previously known patterns of population substructure and migrant decision-making processes. Existing network theories of international migration are supported by the presence of clusters of individuals from the same country and sex throughout the landscape. As these networks lower both the social and economic costs of migration, they tend to encourage new migrants along existing routes (Mullan 1989; Massey & Espinosa 1997; Orrenius 1999). These networks are highlighted in the clustering of identified individuals by their country of origin and sex, and further investigations may improve our understanding of how social ties might facilitate migrant choices.

The geobiological model process also highlights the importance of the sex and recovery location of deceased migrants on understanding who they were and their migration choices. Although the biological data analysis removed any of the variation caused by the sex of individual, it was still a significant variable in the global model exploration. The significance of the corridor of recovery in the global model investigations also highlights the importance of spatial correlation on migrant death and recovery locations as presented here and in other

research (Chamblee et al. 2006; Grannis 2012; Lawrence & Wildgen 2012; Giordano & Spradley 2017). The methods of spatial exploration and geographically weighted regression may also prove useful in related humanitarian crises and broad geographic events where the location of a person is a meaningful variable in understanding or predicting who they were in life. The collection of geographic and demographic of deceased undocumented border crossers should therefore be undertaken by other law enforcement agencies along the US border to enhance their own identification procedures.

Lastly, this research improves the body of knowledge regarding the geographic locations of undocumented border crosser deaths by adding a variable unavailable before: the country from where they came. At present, the results of the exploratory spatial data analysis might be most applicable to the PCOME's investigative process as the geobiological model sample set is still small. New unidentified cases can be mapped in relation to those previously-identified to see if they fall within the present sex or country of origin clusters. This may provide an insight as to which international authorities or humanitarian groups would be most helpful in identifying the individual. Then, the remains can ultimately be returned to their loved ones to aid in the healing process caused by the current state of international immigration in the United States.

## REFERENCES



## REFERENCES

- Chamblee JF, Christopherson GL, Townley M, DeBorde D, Hoover R. 2006. Mapping migrant deaths in southern Arizona: The humane borders GIS. Unpublished report.
- Colibrí Center for Human Rights. 2018. Press Kit; <http://www.colibricenter.org/wp-content/uploads/2018/04/Press-Kit-English-PDF.pdf>.
- Grannis W. 2012. Detecting and Assessing Aggregate Human Pedestrian Migration Patterns Using High Resolution Multispectral Imagery and Geographic Information Systems Texas State University: Doctoral dissertation.
- Giordano A, Spradley MK. 2017. Migrant deaths at the Arizona–Mexico border: Spatial trends of a mass disaster. *Forensic Science International*, 280:200-212.
- Lawrence S, Wildgen J. 2012. Manifold Destiny: Migrant deaths and destinations in the Arizona desert. *Growth and Change*, 43(3):482-504.
- Massey DS, España FG. 1987. The social process of international migration. *Science*, 237(4816): 733-8.
- Mullan BP. 1989. The Impact of Social Networks on the Occupational Status of Migrants. *International Migration*, 27: 69-86.
- Orrenius PM. 1999. The role of family networks, coyote prices and the rural economy in migration from Western Mexico: 1965-1994. Dallas: Federal Reserve Bank of Dallas, (Working Paper, 9910).
- Pima County Office of the Medical Examiner (PCOME). 2017. Annual Report. Tucson, Arizona.
- Ross, A., Slice, D. E., & Ubelaker, D. H. 2014. Population Affinities of Hispanic Crania: Implications for Forensic Identification. In *Biological Affinity in Forensic Identification of Human Skeletal Remains: Beyond Black and White* (pp. 155–164). Boca Raton, FL: CRC Press.
- Rubio-Goldsmith R, McCormick M, Martinez DE, Duarte IM., 2006. The 'Funnel Effect' & Recovered Bodies of Unauthorized Migrants Processed by the Pima County Office of the Medical Examiner, 1990-2005.
- Spradley MK. 2013. Project IDENTIFICATION: Developing Accurate Identification Criteria for Hispanics. US Department of Justice.
- Spradley MK. 2014. Toward Estimating Geographic Origin of Migrant Remains Along the United States-Mexico Border. *Annals of Anthropological Practice*, 38(1): 101–110.

- Spradley MK, Jantz RL, Robinson A, Peccerelli F. 2008. Demographic change and forensic identification: problems in metric identification of Hispanic skeletons. *Journal of Forensic Sciences*, 53(1): 21-28.
- Spradley MK, Jantz RL. 2016. Ancestry estimation in forensic anthropology: geometric morphometric versus standard and nonstandard interlandmark distances. *Journal of Forensic Sciences*, 61(4):892-897.
- Téllez M. 2007 Detour through the Sonora-Arizona Desert: New routes of Mexican emigration to the United States. *Journal of Latino/Latin American Studies*, 2(4):4-19.
- Tise, ML. 2014. *Craniometric Ancestry Proportions among Groups Considered Hispanic: Genetic Biological Variation, Sex-Biased Asymmetry, and Forensic Applications*. University of South Florida.