ABSTRACT

Janna R Caspersen. MEASURING GEOSPATIALLY EXPLICIT PERCEPTIONS OF SUDANESE ETHNIC GROUP LOCATIONS: A COMPARISON OF SUBJECT-MATTER-EXPERTS AND ONLINE DATA (Under the direction of Dr. Tracy Van Holt) Department of Geography, May 2013.

The geospatial perceptions of Sudan and South Sudan subject-matter-experts (SMEs) regarding the location of Sudanese ethnic groups were collected and analyzed, in order to determine how to synthesize and gain meaning from multiple geospatially explicit responses. This study utilizes and attempts to build on methodologies based in the geographic sub-fields of participatory mapping and participatory geographic information systems, as well as, anthropology. To determine how well the previously examined SME perceptions of ethnic group location compare to online news articles, methods based in network analysis, were applied in an attempt to verify the SME perceptions.

Geospatially explicit response maps were collected at the Sudan Studies Conference (April 2012) depicting SMEs' perceptions of ethnic group location. From the hand drawn response maps, digital raster layers were created, aggregated and displayed using graduated colors. The Geospatial Similarity Analysis found that areas where seven or more SMEs agreed an ethnic group was located, was the minimum amount of overlap necessary to indicate agreement. A geospatial consensus analysis was applied using the cultural consensus model (CCM) to determine if the respondents' percentage of overlap with one another indicated culturally shared knowledge, concerning the location of ethnic groups. A geospatial cultural consensus or shared knowledge was found for two of the four ethnic groups analyzed.

To verify the locations indicated by SMEs, an independent data source was used in a geospatially linked semi-automated network text analysis, paired with a content analysis. These analyses were used to identify co-occurrences of location names and ethnic group names within

articles from the Sudan Tribune. The Euclidian distance between expert-indicated location and locations cited in the Sudan Tribune were determined in order to characterize their agreement geospatially. To gain a more comprehensive understanding of the co-occurrences of ethnic groups and locations, each one was examined and contextually classified. The co-occurrences contextually classified as indigenous lands proved to have greater geospatial agreement with the location indicated by SMEs, more so than any other contextual classification (political conflict, ethnic conflict, resources, oil and history).

Measuring consensus geospatially among multiple informants can be applied to any type of geospatial knowledge, in this case the representation of expert perceptions, making it a valuable addition to participatory mapping methods. Semi-automated network text analysis that codes for contextually indicative terms could be further developed and used to examine and model the distribution of ethnic groups in a more comprehensive manner.

MEASURING GEOSPATIALLY EXPLICIT PERCEPTIONS

OF SUDANESE ETHNIC GROUP LOCATIONS:

A COMPARISON OF SUBJECT-MATTER-EXPERTS AND ONLINE DATA

A Thesis

Presented to

The Faculty of the Department of Geography

East Carolina University

In Partial Fulfillment

of the requirements for the Degree

Masters of Arts in Geography

by

Janna Ruth Caspersen

May 2013

MEASURING GEOSPATIALLY EXPLICIT PERCEPTIONS
OF SUDANESE ETHNIC GROUP LOCATIONS:
A COMPARISON OF SUBJECT-MATTER-EXPERTS AND ONLINE DATA

by

Janna R Caspersen

APPROVED BY:

DIRECTOR OF THESIS: ______________________________

Tracy Van Holt, PhD

COMMITTEE MEMBER ______________________________

Thomas W. Crawford, PhD

COMMITTEE MEMBER: ______________________________

Jeffery C. Johnson, PhD

COMMITTEE MEMBER: ______________________________

Burrell E. Montz, PhD

CHAIR OF THE DEPARTMENT OF GEOGRAPHY: ______________________________

Burrell E. Montz, PhD

DEAN OF THE GRADUATE SCHOOL: ______________________________

Paul J. Gemperline, PhD

ACKNOWLEDGMENTS

I would like to thank some of the people that have made this research possible. Firstly, my advisor, Dr. Van Holt for all the hours she has put into guiding me and helping me learn what academic research is all about. Secondly, I would like to thank my committee members: Dr. Johnson, Dr. Crawford, and Dr. Montz; for all the guidance and support they provided to me over the course of the thesis process. My research team, including: Dr. Lobban, Dr. Fluehr-Lobban, Anthony France, Christine Gares, Jamie Brinkley, and Christopher Cooper, also deserves a big thank you, for their willingness to share their expertise and time with me. I would also like to thank Dr. Mulcahy and Dr. Reynolds for their assistance in designing my map figures and explaining various statistical measures. Finally, my work would not be possible without the excellent advice and support I have received from my friends and family. Thanks Mom and Dad!

TABLE OF CONTENTS

LIST OF TABLES vii

LIST OF FIGURES viii

CHAPTER 1: AN INTRODUCTION 1

CHAPTER 2: MEASURING AGREEMENT USING GEOSPATIALLY EXPLICIT PERCEPTIONS OF SUDANESE ETHNIC GROUP LOCATIONS BY SUBJECT-MATTER-EXPERTS AND THE CULTURAL CONSENSUS MODEL 5

- Introduction 5
- Background 7
  - Assessing Ethnic Group Boundaries 7
  - Mapping Ethnic Groups 9
- Methods 10
  - Reference & Response Maps 10
  - Ethnic Group Elicitation 12
  - Participatory Mapping 12
  - Geospatial Similarity Analysis 14
  - Geospatially Applying the Cultural Consensus Model 18
- Results 22
  - What constitutes an area of agreement? 22
  - Is there a consensus among the SMEs' concerning their perceived locations of ethnic groups? 25
- Discussion 27
  - Size Matters 27

Orientation and Livelihood 28

Relaxing Cultural Consensus Model for Geospatial Studies 29

Recommendations for Future Research 30

Conclusion 31

CHAPTER 3: COMPARING SUBJECT-MATTER-EXPERTS' PERCEPTIONS OF SUDANESE ETHNIC GROUP LOCATIONS TO AN ONLINE NEWS SOURCE 32

Introduction 32

Research Objectives 34

Methods 35

Data Sources 36

Participatory Data Collection & Analysis 36

Network Analysis to Extract Desired Text from the Sudan Tribune 37

Content Analysis to Identify the Co-occurrences of Locations & Ethnic Groups 38

Scale Classification of *ST Locations* 39

ST/SME Distance to Measure Geospatial Agreement 39

Ethnic Group Effect on ST/SME Distance 40

Contextual Classification of *ST Occurrences* 40

Correspondence Analysis 43

Livelihood Classification of Ethnic Groups 44

Results 45

Is there agreement between SME perceptions of ethnic group location and the reported location of ethnic groups through the Sudan Tribune? 45

Do any of the contextual classifications have a higher level of geospatial overlap with the ethnic group locations indicated by SMEs? 47
Is there greater agreement between SME location and ST locations (less distance) for ethnic groups with sedentary livelihoods rather than with nomadic livelihoods? 55
Discussion 56
The Influence of Context on Agreement between the Sudan Tribune and the SMEs 56
Distant Outliers Illuminate Displacement 59
Nomadic/ Sedentary Tradition v. Reality 61
Lessons Learned & Future Research Opportunities 62
Conclusion 64
CHAPTER 4: CONCLUSION 65
REFERENCES 68
APPENDIX A: IRB NOTIFICATION OF EXEMPT CERTIFICATION 74
APPENDIX B: RESPONSE MAP 75
APPENDIX C: REFERENCE MAP 76
APPENDIX D: LIST OF ETHNIC GROUPS INCLUDED IN THE INITAL DATA COLLECTION 77
APPENDIX E: DEMOGRAPHIC INFORMATION COLLECTED FROM RESPONDENTS 78
APPENDIX F: EARL MAPS FOR ALL THE ETHNIC GROUPS INITIALLY INVESTIGATED 79

APPENDIX G: RESPONDENT AREA OVERLAP WITH EACH AGREEMENT THRESHOLDS FOR THE DANAGLA 86
APPENDIX H: RESPONDENT PERCENT AREA OVERLAP WITH EACH AGREEMENT THRESHOLDS FOR THE DANAGLA 87
APPENDIX I: RESPONDENT AREA OVERLAP WITH EACH AGREEMENT THRESHOLDS FOR THE SHILLUK....................88
APPENDIX J: RESPONDENT PERCENT AREA OVERLAP WITH EACH AGREEMENT THRESHOLDS FOR THE SHILLUK....................89
APPENDIX K: RESPONDENT AREA OVERLAP WITH EACH AGREEMENT THRESHOLDS FOR THE JA ALIYIN....................90
APPENDIX L: RESPONDENT PERCENT AREA OVERLAP WITH EACH AGREEMENT THRESHOLDS FOR THE JA ALIYIN....................91
APPENDIX M: RESPONDENT AREA OVERLAP WITH EACH AGREEMENT THRESHOLDS FOR THE NGOK 92
APPENDIX N: RESPONDENT PERCENT AREA OVERLAP WITH EACH AGREEMENT THRESHOLDS FOR THE NGOK 93
APPENDIX O: DANAGLA RESPONDENT-BY-RESPONDENT PERCENT AREA OVERLAP 94
APPENDIX P: JA ALIYIA RESPONDENT-BY-RESPONDENT PERCENT AREA OVERLAP 95
APPENDIX Q: NJOK RESPONDENT-BY-RESPONDENT PERCENT AREA OVERLAP ...96
APPENDIX R: SHILLUK RESPONDENT-BY-RESPONDENT PERCENT AREA OVERLAP 97

LIST OF TABLES

Table 1: Tukey HSD Post-Hoc Test on the average percent area of overlap of all respondent's polygons with each agreement threshold..24

Table 2: Statistical Results for Consensus Analysis...25

Table 3: Respondent Competency Scores for Consensuses Analysis 26

Table 4: *ST/SME Distance* Statistics by Ethnic Group..46

Table 5 Tukey's Honestly-Significant-Difference Test of the variation in between the average *ST/SME Distance* of each Ethnic Group..46

Table 6: *ST/SME Distance* Statistics by Context Classification..47

Table 7: Contextual Classification by Ethnic Group Cross-Tabular Table 48

Table 8: Two Sample t-test of the variation between the average *ST/SME Distance* of nomadic and sedentary ethnic groups..55

LIST OF FIGURES

Figure 1: Ethnic Agreement Raster Layer (EARL) Maps for the four ethnic groups analyzed....15

Figure 2: Example of respondent-by-agreement threshold overlap calculation used for the Geospatial Similarity Analysis 17

Figure 3: Example of the respondent-by-respondent overlap calculation used for the CCM 21

Figure 4: Mean percent area of overlap of all respondent's polygons with each agreement threshold....23

Figure 5: Mean area of overlap of all respondent's polygons with each agreement threshold 23

Figure 6: Correspondence Analysis....49

Figure 7: Geospatial Agreement between the Sudan Tribune and SME Perceptions of the Bari Ethnic Group....51

Figure 8: Geospatial Agreement between the Sudan Tribune and SME Perceptions of the Fur Ethnic Group....53

Figure 9: Geospatial Agreement between the Sudan Tribune and SME Perceptions of the Beja Ethnic Group....54

Figure 10: Geospatial Agreement between the Sudan Tribune and SME Perceptions of the Ngok Ethnic Group....57

Figure 11: Geospatial Agreement between the Sudan Tribune and SME Perceptions of the Shilluk Ethnic Group 60

CHAPTER 1:

AN INTRODUCTION

To understand the drivers of ethnic conflict in our ever shifting political and environmental landscape, a better way to characterize the distribution of ethnic groups, at varying temporal and spatial scales, is needed. To document the distribution of ethnic groups in areas where there is a lack of on-the-ground data and the distribution of populations is volatile due to conflict, we can consult expert respondents. However, it is challenging to assess areas of agreement with geospatial data and identify which pixels stay in and which ones are eliminated. Traditionally, geographers have used participatory mapping and focus groups to delineate areas, but this remains a challenge since individual responses can be influenced by group dynamics (Krueger and Casey, 2009). The cultural consensus model (CCM), which assesses shared knowledge/agreement between respondents, offers potential. However, to date, this approach has not been applied to geospatially explicit data. Still even if a consensus among experts is found, there still is the possibility that expert knowledge is distinct (Boster and Johnson 1989) and verification by independent sources can help substantiate the data. Network text analysis has the power to identify connections within textual data, that can be linked geospatially (Van Holt et al. 2013), and used to verify the aggregated experts' perceptions.

The field of geography has served a pivotal role in the use of geographic information systems and science (Openshaw 1991). It is from this substantial tradition that this research applies basic functions such as area and distance measurement tools, as well as more specialized spatial aggregation and cartographic visualization capabilities. Participatory mapping is a method that combines local knowledge and professional mapping techniques to produce a map that shows things such as land division and ownership or local natural and unnatural resources. Often it is understood to be driven by local people, with the assistance of GPS and GIS technicians (Chapin, Lamb and Threlkeld 2005). Participatory GIS or PGIS originated when GPS and GIS technologies were integration with Participatory Learning and Action (PLA) (Rambaldi et al. 2006). The primary goal of PGIS is to empower underrepresented populations through the formal production of their own knowledge using geographic information systems and technologies (Rambaldi et al. 2006).

The production of dynamic maps through methodologies grown from participatory mapping and PGIS, has been used by Borjorquez-Tapia to integrate multiple geospatially explicit data layers with expert knowledge regarding suitable overwintering-habitat characteristics for butterflies (Bojorquez-Tapia et al. 2003). Their analysis required integrating multiple independent-group developed models, for suitable habitat criteria with weighting, which proved to be challenging because there was not a consensus between the groups (Bojorquez-Tapia et al. 2003). The expansion of geospatially explicit participatory research methodologies has made participatory mapping into an avenue for comparing and aggregating a wide variety of geospatial knowledge. A method to systematically integrate multiple perceptions and measure their agreement geospatially would be a valuable addition to participatory mapping methods because it allows for the collection and aggregation of group knowledge.

CCM is based in anthropology and was designed to systematically measure culturally shared knowledge (Romney et al. 1986). The CCM has been applied previously to unconstrained pile-sort data, which is a method sometimes used in structured interviewing where the respondent sorts a stack of notecard, each with some related term on it, into as many or as few piles as they see fit, to measure and compare the competence (or shared knowledge) of experts and novices (Boster and Johnson 1989). In this study the CCM was used to determine if the respondents’ geospatially agreed with one another, on the area of ethnic groups, due to shared knowledge. The extensive applicability of participatory mapping methods makes apply the CCM to geospatially explicit data a valuable addition to the sub-field of participatory research within geographic field methods.

Once the experts’ responses have been analyzed and a geospatially explicit ethnic group distribution is defined, those locations could be checked for accuracy using an independent data source and network analysis. GIS and network analysis both have a strong basis in topology (spatial relationships between features) and are inherently spatial, making the two highly compatible for a wide range of research applications (Curtin 2007). The field of network analysis has recently been applied to geographic studies by geospatially linking and visualizing locations within network text-data (Giuliani 2007; Maggioni and Uberti 2011; Broekel and Boschma 2012). Social network analysis programs, such as ORA (Carley et al. 2011b) have the ability to geospatially visualize networks generated from text sources and are able to generate KML file for use in other GIS interfaces, such as ArcGIS (ESRI 2011).

This thesis is broken into four chapters: an introduction (Chapter 1), two independent academic research articles (Chapters 2 and 3), and a conclusion (Chapter 4). Chapter 2 addresses two research questions, concerning how to synthesize and gain meaning from multiple, geospatially explicit, drawn map responses. The two questions addressed in this chapter are:

- What constitutes an area of agreement?
- Is there a consensus among the SMEs’ concerning their perceived locations of ethnic groups?

Chapter 3 addresses three research questions, verifying the aggregated responses (from Chapter 2) using geospatially linked network text analysis and context analysis of articles from a Sudanese news source. The three questions addressed in this chapter are:

- Is there agreement between SME perceptions of ethnic group location and the reported location of ethnic groups through the Sudan Tribune?
- Do any of the contextual classifications have a higher level of geospatial overlap with the ethnic group locations indicated by SMEs?
- Is there greater agreement between SME location and ST locations (less distance) for ethnic groups with sedentary livelihoods rather than with nomadic livelihoods?

Chapter four serves as a conclusion, in which key findings are summarized and the implications of those findings are discussed.

CHAPTER 2

MEASURING AGREEMENT USING GEOSPATIALLY EXPLICIT PERCEPTIONS OF SUDANESE ETHNIC GROUP LOCATIONS BY SUBJECT-MATTER-EXPERTS AND THE CULTURAL CONSENSUS MODEL

INTRODUCTION

To understand the drivers of ethnic conflict in our ever shifting political and environmental landscape, we need a better way to characterize the location of ethnic groups dynamically. One of the greatest obstacles in analyzing civil conflict and geography are the limitations within the available data (Buhaug and Lujala 2005). Since the number of civil wars has been steadily increasing since 1945 (Fearon 2004), conflict research has evolved from pairwise comparisons between countries (Maoz and Abdolali 1989; Bremer 1992) to analyses within countries (Buhaug and Gates 2002; Buhaug and Rod 2006; Gleditsch 2007). Conflict appears to be tied to environmental resources although researchers do not agree on whether scarcity (Homer-Dixon 1994; Ross 2004a; Theisen 2008), abundance (Gilmore et al. 2005), or uneven distribution (Ross 2004a, 2004b) is most relevant. Ethnic heterogeneity has also been shown to be correlated with ethnic group conflict (Sambanis 2001; Van Holt et al. 2012).

Sudan and South Sudan are countries whose histories are profoundly affected by chronic ethnic conflict beginning at independence in 1955 and continuing through their separation on July 9th 2011 (Lobban, Kramer and Fluehr-Lobban 2002; Themner and Wallensteen 2012; Lobban and Fluehr-Lobban 2012). Even after the recent split of Sudan, into two independent countries (Sudan and South Sudan), conflict persists over Abyei, a parcel of land located on either side of the new border (Themner and Wallensteen 2012). Over the past four decades the chronic conflict has led many Sudanese ethnic groups into a series of emigration and immigration process both domestically and/or internationally (Ayers 2010). Thus the location of many Sudanese ethnic groups has likely changed, partially or wholly, from the traditional/historical location.

The existing maps of Sudanese ethnic groups that are available today are helpful in conceptualizing the amount of diversity present in Sudan. However, they are static maps that are in many cases created by one or two individuals who exhaust historical and anthropological data, such as the ethnic group maps created by Izady (2012) who synthesized data sources dating back to 1954. These maps are among the few geospatially explicit maps that show ethnic group patterns of distribution in Sudan and South Sudan. One of the most widely recognized ethnic group maps of Africa was created in 1959 by George Peter Murdoch (Murdoch 1959). Other more recent maps of Sudan and South Sudan have been published anonymously online (Muturzikin 2007); the sources cited include historical atlases as well as many online sources such as Ethnologue (Lewis, Simons, and Fennig 2013) and The World Atlas of Language Structures Online (Forkel 2011).

The most ideal way to map changes in ethnic group location would be to use on the ground data collection and verification, with the ethnic groups leading the mission. Countries such as Sudan and South Sudan are so large in area, (2,505,810 km^2 the size of Alaska, Texas, and Maine combined) and population (60,602,000 projected for 2025) (Lobban et al. 2002) that to map ethnic groups on the ground would take an enormous amount of resources and time.

An alternative that would make the task of integrating these dynamic changes in ethnic group location more manageable would be to consult multiple Sudanese and non-Sudanese subject-matter-experts (SMEs) in a geospatially explicit mapping exercise of Sudanese ethnic groups. Expert perspectives are easily solicited and often consist of current knowledge and experience regarding the geospatial nature of ethnic group population flows that have occurred both historically and in recent years. Expert knowledge combined with hand-drawn participatory mapping is an inexpensive and relatively simple way to collect geospatial data. When using hand-drawn responses, it can be assumed that each response will be unique in size, shape, and location. This variation in data makes choosing which boundaries to use and how to aggregate the responses challenging.

Before multiple experts' geospatially explicit perceptions are able to be converted into a descriptive and accurate map, research is needed on how to synthesize and gain meaning from multiple responses. In this study, Sudan SMEs will draw their perceptions of ethnic group locations on a georeferenced map. To determine how many overlapping responses are necessary to constitute an area of agreement, the hand-drawn responses will be aggregated and compared to one another using a Geospatial Similarity Analysis.

In order to attach statistical meaning to the overlap in responses, the cultural consensus model (CCM) will be applied to the geospatial data. The CCM will help determine if the respondents' mapped geospatial overlap is based on shared knowledge. Given this background this paper addresses two research questions: first, what constitutes an area of agreement; and second, is there a consensus among the SMEs' concerning their perceived locations of ethnic groups?

BACKGROUND

Assessing Ethnic Group Boundaries

One of the challenges in participatory mapping is how to synthesize many individuals' responses into a single map. A systematic method that will allow researchers to measure the extent to which multiple geospatially explicit responses agree with one another is needed. One way to meet this challenge is to compare the geospatial similarity or overlap of individual response boundaries. To determine if the geospatial similarity (or overlap) of respondents' drawn boundaries indeed indicates agreement and is a product of culturally shared knowledge, the CCM will be applied to geospatial data for the first time.

The cultural consensus theory states that culturally shared knowledge can be quantified by assessing agreement among respondents (Romney, Weller and Batchelder 1986). With any culture, there is some amount of knowledge that is shared by some groups and individuals that flows through a subsystem of knowledge patterns (Roberts 1964; Kroeber 1948). The CCM reasons that if respondents have a high level of consensus or agreement in their answers then their answers symbolize culturally shared knowledge and thus represent culturally correct responses (Romney et al. 1986). The culturally correct responses are calculated differently in the formal model than in the informal model. The formal model applies Bayesian weighting to all the responses in order to estimate the answer key whereas the informal model uses the factor scores (average) of all the responses to estimate the culturally correct answers (Weller 2007).The formal CCM is used for multiple-choice and true/false data whereas the informal model can process ordinal, interval, and ratio-scaled data (Weller 2007).

To measure shared knowledge, respondents are given a series of multiple choice or true/false questions (Romney et al. 1986). This information is then transformed into an agreement matrix. The CCM evaluates each respondent and measures the individual competency of cultural knowledge based on the agreement between respondents' answers (Weller 2007) and the culturally correct answer. Cultural consensus theory assumes that there is one culturally correct response to each question (Romney et al. 1986). The theory of cultural consensus makes three assumptions that in practice are used as guidelines for applying the model. One, there exists a common single truth that respondents from the same “cultural reality” will respond the same to questioning (Romney et al. 1986, 317). Two, an individual respondent’s responses are independent of other respondent’s influence (Romney et al. 1986). Three, the questions asked of respondents are equally difficult to answer (Romney et al. 1986).

The CCM is neutral and thus does not make any assumption about what a ‘proper’ answer is; instead the correct answer is gathered from the pairwise comparison of all the responses (Romney, Batchelder and Weller 1987, 164). The downside of applying the CCM to geospatial data is that although it will describe the extent to which the responses agree with one another (based on culturally shared knowledge), the model will not produce a single culturally correct geospatially explicit ethnic group boundary.

Mapping Ethnic Groups

Mapping of ethnic groups’ indigenous lands has expanded significantly over the past five decades (Chapin, Lamb and Threlkeld 2005) due to its potential to assist in documenting land use, designing resource management plans, as well as the preserving the historical and cultural knowledge of underrepresented people (Herlihy and Knapp 2003). Many methods have been used for this type of mapping, from community sketch mapping to participatory geographic information systems (PGIS); most emphasize participation with the community(s) that is being documented (Chapin et al. 2005). PGIS is the combination of participant’s spatial knowledge and geospatial tools (Rambaldi et al. 2006).

PGIS has been used in studies of local land use such as mapping the intensity of grazing in pastoral lands by using local expert’s knowledge of grazing and digitizing their drawn boundaries using hardcopy transparencies and GIS software (Bemigisha et al. 2009). PGIS can be undertaken using any level of technology, from making maps in the dirt with natural materials to utilizing a mixture of GIS technologies and systems (Rambaldi et al. 2006).

Scale mapping is a participatory methodology that is primarily concerned with producing maps that are georeferenced and use symbology and scale in a way that accurately orients its respondents' drawn responses (Rambaldi et al. 2006, 5). *Scaled map* responses can be transformed and digitally manipulated using *PGIS spatial analysis*. *PGIS spatial analysis* makes use of current technologies to analyze geospatial questions concerning simple measures such as the time and cost (Rambaldi et al.2006), as well as more rigorous measures concerning patterns of language (Luo et al. 2007) or conservation and management of resources on national reserves (Bernard, Barbosa and Carvalho 2011).

METHODS

To begin the research, maps were created as data gathering tools, designed to record hand-drawn expert perceptions concerning the location of ethnic groups. The ethnic groups included in this research were selected by experts who have a deep knowledge of Sudan and South Sudan's ethnic history and landscape. Interviews were conducted using participatory mapping techniques and a convenience sample. To determine the number of overlapping responses that are necessary to constitute an area of agreement between respondents' perceptions of ethnic group locations, a Geospatial Similarity Analysis was conducted. The CCM was the applied to geospatial in order to determine if the overlap of respondents' drawn boundaries actually indicates agreement based on shared knowledge.

Reference & Response Maps

In order to facilitate the collection and analysis of geospatial data solicited from multiple respondents, two maps were created of Sudan and South Sudan: a response map and a reference map. The response map was created and used to record hand-drawn perceptions of ethnic group location boundaries (Appendix B). The reference map was created in case a respondent felt they needed additional land-cover information to orient their response.

The preliminary response map was evaluated by Dr. Richard Lobban and Dr. Carolyn Fluehr-Lobban both of whom have worked in Sudan for the past thirty years (Lobban and Fluehr-Lobban 2012). They identified features (cities, roads, and other reference points) to include to orient respondents. Selecting appropriate features is important to the legibility of any map, but the cartography for maps used in scaled mapping must allow respondents to relate to the map and orient themselves geospatially (Rambaldi et al.2006).

On the final response map, state borders (UNDP 2010), expert selected cities (UNDP 2010), roads (UNDP 2010), and rivers (FAO Southern Sudan 2004) were included as were mountain ranges that were defined as areas above 800 meters, from a digital-elevation-model (DEM) (Lehner, Verdin and Jarvis 2006). The expert selected features were designated because they are easily recognized and are thus able to help orient many respondents. The mountain ranges were included because other orienting features were absent in those areas and a few of the ethnic groups have historically inhabited mountainous regions. The reference map included all the same features as the response map except, instead of the DEM, a Landsat satellite image of Sudan and South Sudan was used (Appendix C).

Each response map (Appendix B) had an ethnic group labeled at the top. At the bottom was a confidence scale (1-5) where the respondents could report their level of confidence in the accuracy of their responses, five being very confident and one being not confident (Appendix B). For the analysis, any response map that had a confidence value of two or lower was not included.

Ethnic Group Elicitation

As requested by the research team, the Lobbans provided a list of 25 ethnic groups to use for the SME's participatory mapping, with the guideline that the ethnic groups listed are approximately the same size. Ethnic groups in Sudan and South Sudan are generally divided into three levels of ethnic identity. For example the Dinka ethnic group (1^{st}) is very large and is made up of smaller unique ethnic branches such as the Ngok, Malwal, and Bor (2^{nd}); from these there are smaller sub-groups known as family groups (3^{rd}) that are based unique lineages (Lobban and Fluehr-Lobban 2012). We chose to map at the second level of identity, ethnic branch, due to each groups size, scale, and media coverage. The final list includes the following ethnic groups: Azande, Bari, Beja, Beni Amer, Bor, Danagla, East Jikany, Hadendowa, Humr, Ja Aliyin, Kakwa, Lou, Madi, Malwal, Masalit, Messirya, Murle, Ngok, Rizaygat, Shaygiya, Shilluk, Taisha, Talodi, West Jikany, and Zaghawa (Appendix D).

The response maps were organized into respondent packets that contained one map per ethnic group, the reference map, and an extra response map. After the first four interviews, we added an additional ethnic group, the Fur, who are a rather large group and unrepresented in our initial list. We used the extra drawing map in each respondent packet to map the Fur.

Participatory Mapping

We interviewed seventeen experts attending *The Sudanese Studies Conference* in Tempe, Arizona (May 18-20, 2012). To identify the experts, we asked the conference organizers to identify which people would be most informed on our subject of study. We arranged some interviews prior to the conference via email, then many more interviews were scheduled at the conference. In order to attract experts to our study we distributed flyers, invited people at our roundtable *Participatory Understanding of Sudanese Ethnic Groups*, and used conference events to network.

In the interview SME *scaled mapping* was conducted. The respondents were asked to draw on a map where they perceive each ethnic group to be located. They were given one ethnic group per map and one map at a time. All responses were voluntary, thus some respondents chose not to respond to certain ethnic groups, in most cases this was because the respondent did not recognize the ethnic group and thus could not define a location. This process took anywhere from five to fifteen minutes for all twenty-five ethnic groups. Additional demographic data were collected from the respondents including age, ethnicity (group, family, and branch), languages spoken, birth place, current residence, education level, and current employment (Appendix E).

The hand-drawn response maps were scanned at 200 dpi resolution and stored in Tagged Image file format. Then each of the response maps was georeferenced using four control points. The average remote sensing (RMS) error for all 398 maps was 0.6924 km, ranging from .02 km to 10.96 km, with a standard deviation of 0.75 km. Response polygons were created using ArcScan (ESRI 2011) to heads-up digitize (generate a poly-line and a polygon) each drawn response; these were used for applying geospatial data to the CCM. The response polygon was then converted into a raster layer, using a cell size of 1.28 km, and assigned a value of one, while the remaining cells were assigned zeros. The raster layers were used in the Geospatial Similarity Analysis. The raster of that was rectified to *Africa Albers Equal Area Conic* projection and the *WGS 1984* datum. These two references were selected in order to keep the distortion of area at a minimum (Krygier and Wood 2005).

Geospatial Similarity Analysis

To determine how many overlapping expert responses are necessary to constitute an area of agreement, a Geospatial Similarity Analysis was used. To integrate all 17 of the respondents' raster layers for each of the 26 ethnic groups, the cell statistics function was used to add the raster layers together to create an *Ethnic Agreement Raster Layer* (EARL) (Appendix F). The EARLs reflect the number of respondents that agree that an ethnic group is present at each pixel.

Each raster layer is made of many cells, each containing a value of 1 or 0. A value of 1 indicates that the expert respondent included that pixel in the area where he or she perceived that ethnic group to be located. A value of 0 indicates that that pixel was not included by the expert respondent. Thus when the cells statistics function adds the raster layers of all the respondents together for each ethnic group, each unique pixel value is aggregated. For example, if three experts all indicated the same pixel has a particular ethnic group residing in it, after summing that pixel would have a value of 3. To perform the Geospatial Similarity Analysis, ethnic groups with at least ten overlapping responses were selected, because they provide the widest range of agreement values for analysis (EARLs ranged from 0 to 13). Four of the twenty-six EARLs met or exceeded the ten overlap minimum: the Ngok, Danagla, Ja Aliyin and Shilluk (Figure 1).

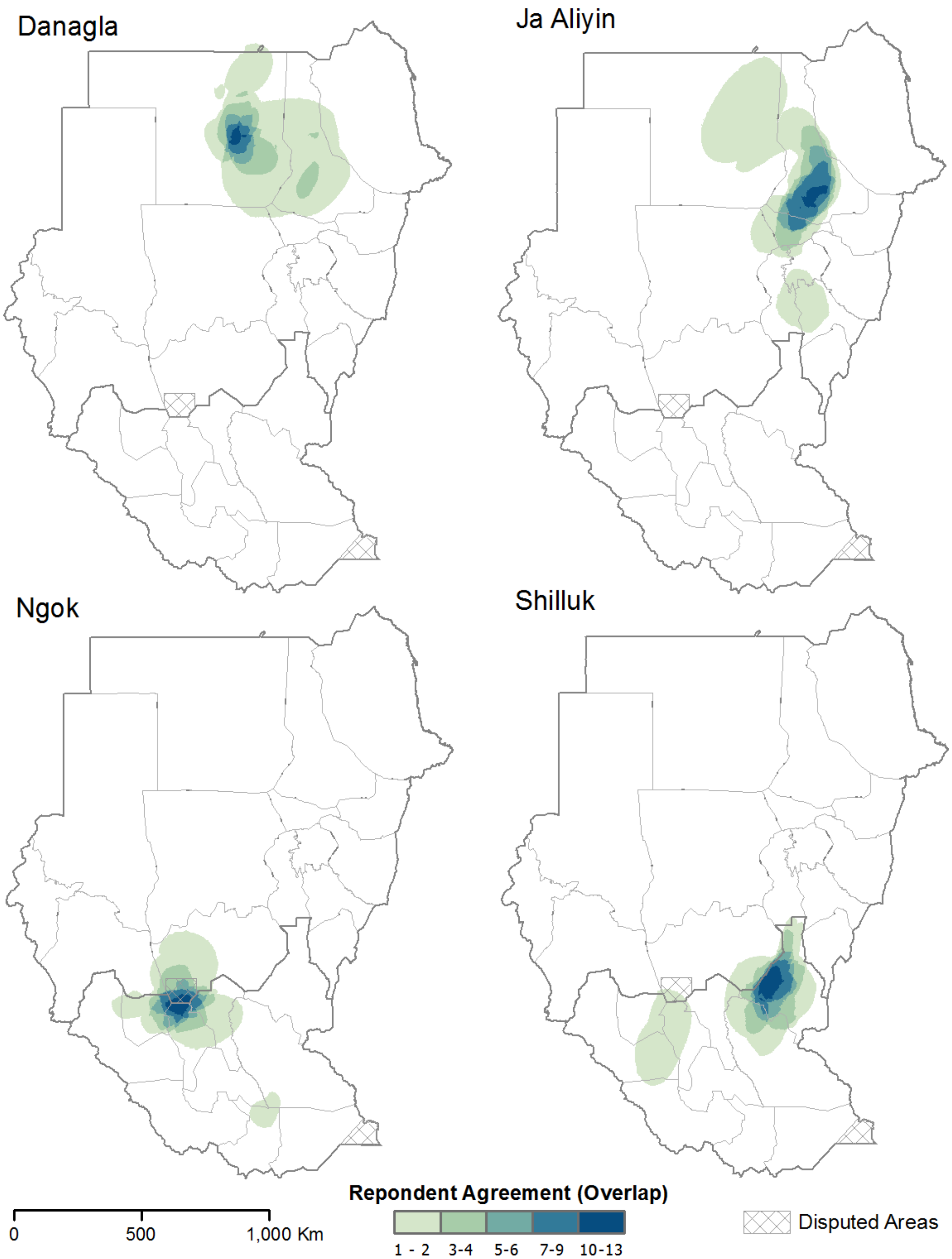


Figure 1: Ethnic Agreement Raster Layer (EARL) maps for the four ethnic groups analyzed (reflecting the number of overlapping respondents)

To calculate the geospatial similarity, the area (km^2) and percent area of overlap each respondent had with an agreement area was assessed, using the previously generated response polygons. Agreement areas were defined by dividing the full range of agreement values into four thresholds of agreement (≥3, ≥5, ≥7, ≥10). The amount of area each threshold covered was calculated. Then, using the intersect tool in ArcGIS (ESRI 2011), the area of intersection each respondent had with each agreement threshold was determined (Appendices G-N).

To produce the percent area that each respondent shared with each agreement threshold, the individual response area of intersection was divided by the area of the designated agreement threshold. For example (Figure 2), Ja Aliyin, respondent 23, overlapped with three or more other respondents for 45,389 km^2, and with ten or more respondents for 4,402 km^2. These area measurements were then divided by the total area where there are ≥3 (61,884 km^2) and ≥10 (4403 km^2) responses present (Figure 2). The result is then the percentage of the total area the respondent covers (Figure 2). Respondent 23's response covers 73% of the area where more than three respondents have indicated the Ja Aliyin ethnic group is located and 100% of the area where more than ten respondents agree (Figure 2).

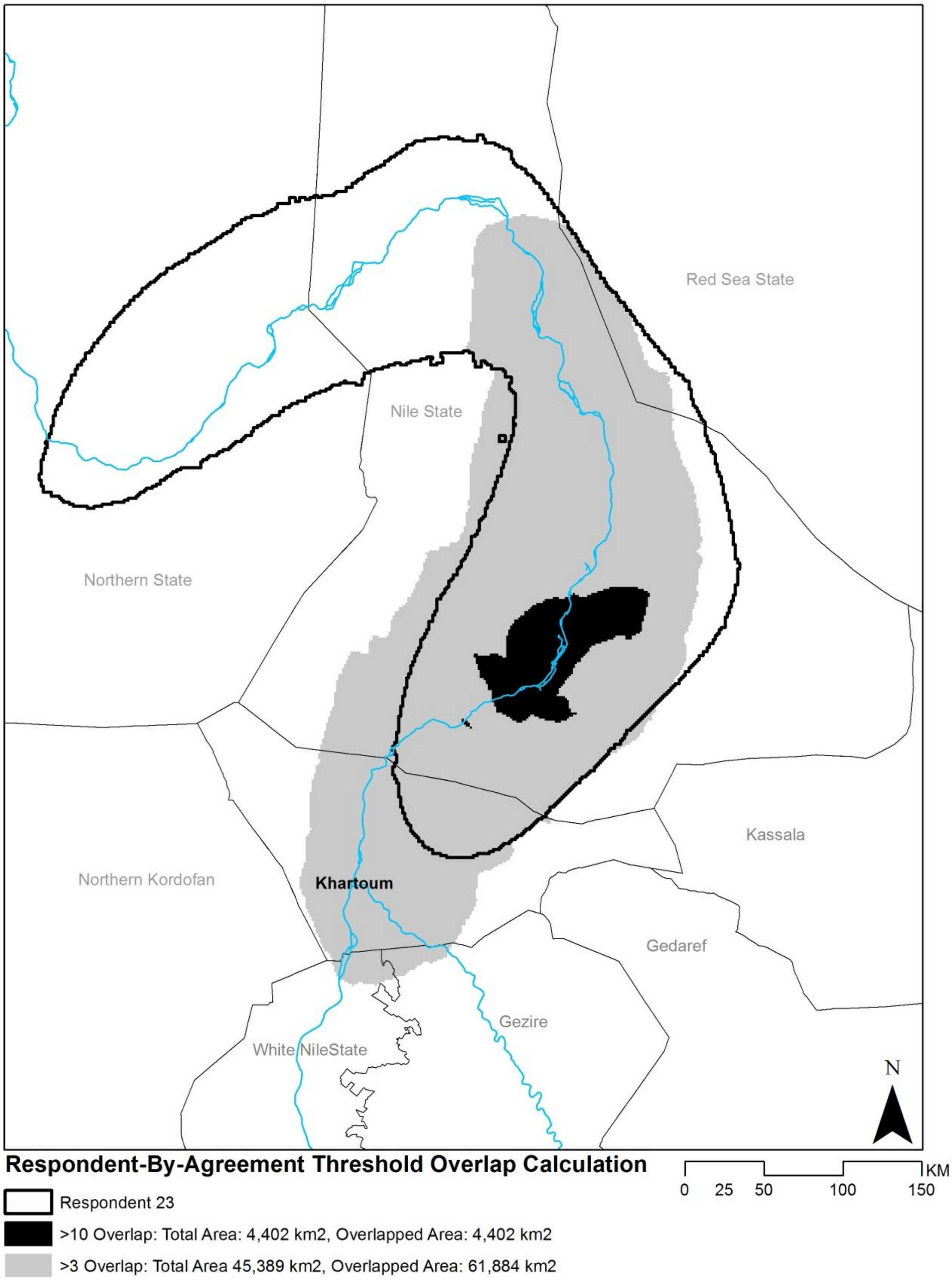


Figure 2: Example of respondent-by-agreement threshold overlap calculation used for the Geospatial Similarity Analysis

The mean area and mean percent area of overlap for all respondents were calculated for each ethnic group and each threshold. Using SPSS, the average area of overlap and average percent area overlap of response polygons were then analyzed with an analysis of variance (ANOVA) to determine if a significant difference existed between the thresholds. This was followed by a Tukey HSD Post-Hoc Test to see how the thresholds of agreement were different across all four ethnic groups.

Geospatially Applying the Cultural Consensus Model

To determine if the respondents' geospatial overlap indicates shared knowledge, the informal CCM was applied. The informal CCM is being applied, rather than the formal model, because the data describing the averaged percent area of overlap between each pair of respondents is ratio data which cannot be used in the formal model. The averaged percent of overlap between each pair of respondents represents the extent to which each pair agrees with one another that a certain ethnic groups is present in a certain area.

Before the model can be properly applied, the eigenvalue ratios must be examined to determine if the data is a good fit and conforms to the first assumption of the informal CCM, that there is a single pattern of responses present in the data (Weller 2007). There must be at least a three-to-one ratio between the first and the second eigenvalues to indicate a cultural consensus and a single pattern of responses (Weller 2007). If the ratio between the first and second eigenvalues is less than three-to-one, there are either multiple factors or no factors present in the data that would explain the first factor (Romney et al. 1986, Weller 2007).

The first eigenvalue (factor loadings) calculated by the informal CCM represent the respondents' competence scores which are used to measure variation in respondents' knowledge (Weller 2007). The competency scores are essentially a pairwise comparison of overlap between each pair of respondents, representing the extent to which each respondent geospatially agreed with all the other respondents. Thus, if one respondent consistently has greater overlap with a greater number of respondents, their competency score will be higher than the respondents with less overlap with fewer other respondents. The competency scores range from 0 to 1; the average competency score to indicate that a respondent possesses shared knowledge is greater than 0.50 (Weller 2007, 363). The average of these competency scores across all respondents provides a measure of agreement in the data (Weller 2007; Weller 1987), indicating agreement among the respondents, and therefore shared knowledge. To test if the variation in response polygons influences the CCM competency scores, a bivariate, Pearson's two-tailed correlation analysis was run on the respondent's individual polygon size and their CCM competency scores for each ethnic group.

The second eigenvalue (set of factor scores), outputted by the informal CCM, is hypothetically the culturally correct answer, which in practice is the averaging of respondents' individual responses by their competency scores and aggregating the responses (Weller 2007). However, in this case, due to use of geospatial data, the second eigenvalue actually represents the average overlap between each pair of respondents.

Following the informal CCM, a minimum residuals factor analysis was run on the transposed agreement matrix, using no factor rotation in UCINET 6 (Borgatti 2002). The agreement matrix contained the averaged percent of area overlap between each pair of response polygons (Appendices O-R). When applying the CCM, the matrices used must be transposed to make the questions themselves, the mapped ethnic group locations, the unit of analysis (rows), or in this case to make the percent overlap between each pair of respondents the unit of analysis, and to make the respondents the variables (columns) (Weller 2007). To create the transposed respondent-by-respondent agreement matrix, the area of each response and the area of intersection between each pair of respondents were determined. To produce the percent area of intersection between each pair of respondents, the area of each intersection was then divided by the total area of each corresponding response (Figure 3). The result is two different percentages due to the size variation between response polygons, resulting in an asymmetrical matrix. The CCM processing design within UCINET does not accommodate such data so the matrices were symmetrized. The percentage of overlap between each pair of respondents was averaged so that the table could be symmetrical and the factor analysis could be run.

For example, the first and third respondents had an intersection area of 27,023 km^2 and individual areas of 31,564 km^2 (Respondent 1) and 40,391 km^2 (Respondent 3) (Figure 3). To produce the percent overlap, the area of intersection by each respondent was divided by the total area of the respondent’s polygon. The first respondent's polygon overlaps with 67% of the third respondent's polygon whereas the third respondent's polygon overlaps with 86% of the first respondent's polygon (Figure 3). The variation in size of response affected the percentage: the first respondent’s polygon had a smaller area, and thus had a higher percentage of overlap. The first and third respondents’ responses were then averaged (77%), to symmetrize the matrix.

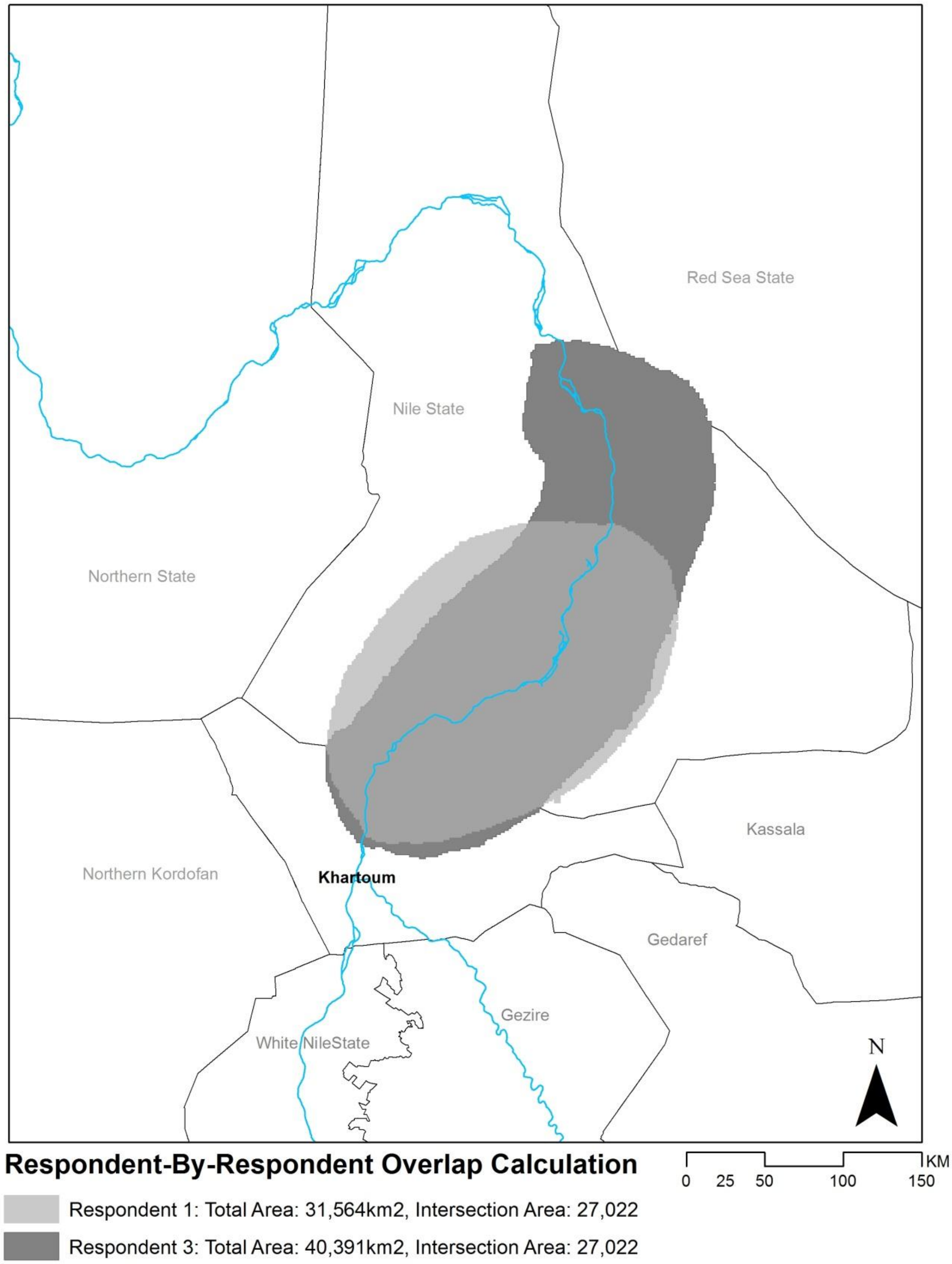


Figure 3: Example of the respondent-by-respondent overlap calculation used for the CCM.

RESULTS

What constitutes an area of agreement?

The *Ethnic Agreement Raster Layers* (EARLs) appeared visually to be in general agreement where five or more of the respondents' polygons overlapped. For example, the Ja Aliyian EARL shows an agreement area forming when overlapping responses reach five or more with a maximum respondent overlap of eleven (Figure 1). At the agreement threshold of ≥5, the average responses for all four ethnic groups exceed 40% overlap with the aggregate agreement area (Figure 4). The average percent area of overlap a respondent's polygon shares with the ≥7 threshold is greater than 50% for all four ethnic groups (Figure 4). The average area a respondent shares with a threshold agreement area decreases as the threshold for overlapping increases (Figure 4 and 5). This is because the area which respondents overlap gets smaller as the level of agreement increases. For example, the Ja Aliyian respondents had the highest average overlap area with all four thresholds in spite of the shrinking size of the threshold area (Figure 5).

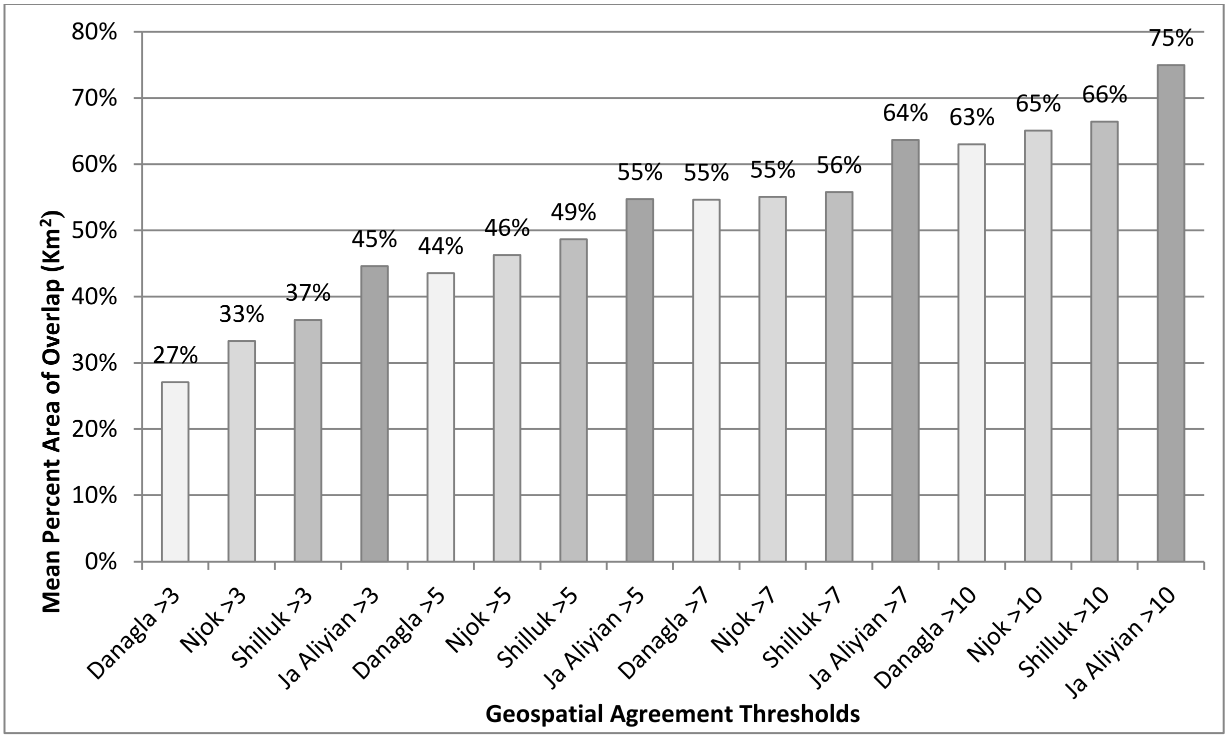


Figure 4: Mean percent area of overlap of all respondent's polygons with each agreement threshold

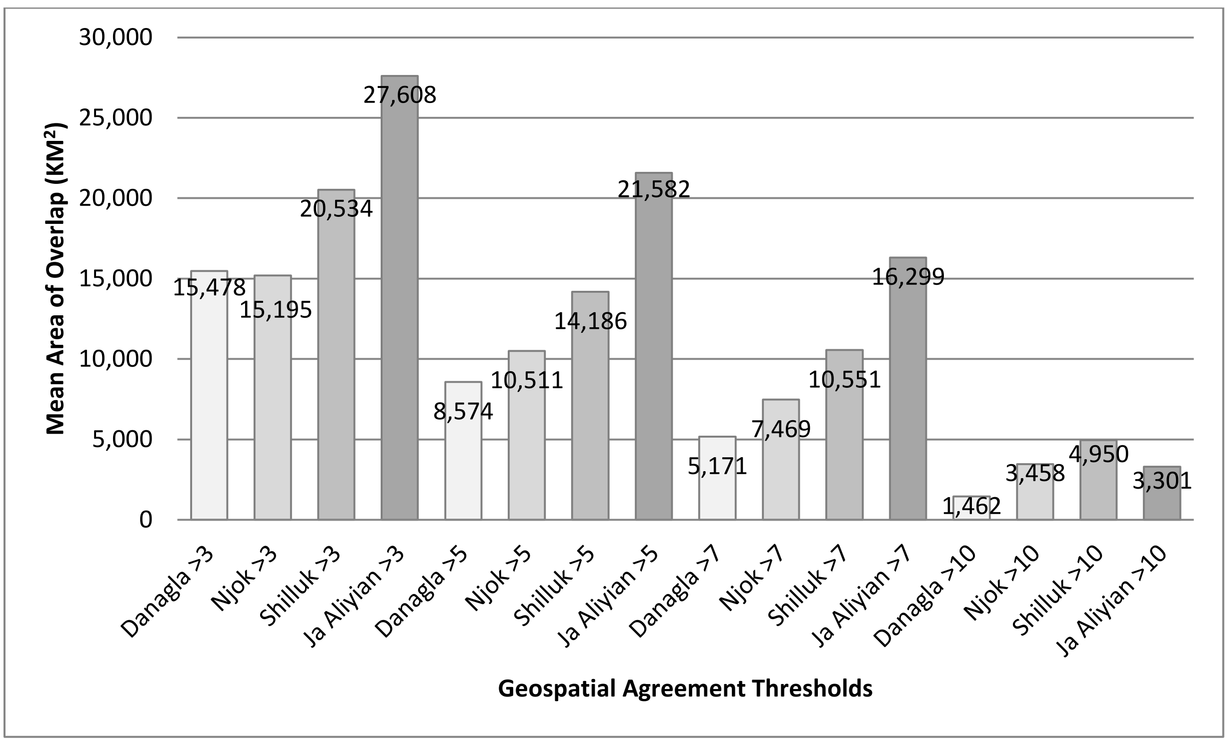


Figure 5: Mean area of overlap of all respondent's polygons with each agreement threshold

The ANOVA showed that there was a highly significant difference between the mean percent area overlap (p=.000), of all respondents collectively between all the different thresholds (≥3, ≥5, ≥7, ≥10). The Tukey HSD Post-Hoc comparison found that there was a highly significant difference between the ≥3 and ≥7 thresholds (p=0.001) and between the ≥5 and ≥10 thresholds (p=0.008) (Table 1). The mean difference between the ≥3 and ≥7 thresholds (22%) was greater than the difference between the ≥5 and ≥10 thresholds (20%) (Table 1). The difference in the mean percent area of overlap of for all respondents between the ≥5 and ≥7 thresholds, ≥7 and ≥10 thresholds, and the ≥3 and ≥5 threshold were all found to not be significantly different (p=0.429, p=0.322, p=0.126) (Table 1). This suggests that although there is not a significant difference between neighboring thresholds, the ≥7 threshold has the most significant variation, indicating agreement.

Table 1: Tukey HSD Post-Hoc Test on the average percent area of overlap of all respondent's polygons with each agreement threshold

Thresholds of Agreement	Mean Difference	Significance
≥3 & ≥5	0.13	0.126
≥3 & ≥7	0.22	0.001
≥5 & ≥7	0.09	0.429
≥5 & ≥10	0.2	0.008
≥7 & ≥10	0.1	0.322

The minimum number of overlaps that is sufficient to constitute an area of agreement is seven or more. The ≥7 agreement threshold provides a high level of average overlap with the agreement area (over 50%). The difference between the thresholds of ≥7 and ≥10 is not significant and since the size of the aggregate agreement area for the threshold of ≥7 is larger than the area for the ≥10 threshold, the ≥7 threshold appears to be empirically more satisfying.

The analysis showed that the thresholds of ≥ 3 and ≥ 5 have on average less than 50% overlap with the aggregate agreement area.

Is there a consensus among the SMEs' concerning their perceived locations of ethnic groups?

Typically when applying the CCM, there must be at least a three-to-one ratio between the first and the second eigenvalues in a minimum residuals factor analysis to indicate consensus (Weller 2007). This is because this ratio indicates that there is one underlying fact that explains the data, shared knowledge, which in this case concerns the boundaries of specific ethnic groups' locations. In this analysis, Ja Aliyia had a ratio between the first and second eigenvalues of 4.89 and the Shilluk had a ratio of 3.18, both indicating a consensus between respondents (Table 2). The Dangala had a slightly higher ratio of 2.89 than the Ngok, which had the lowest ratio of 2.48 (Table 2). These two although they do not make the formal three-to-one ratio, have other indicators that an area of agreement exists between responses.

Table 2: Statistical Results from Consensus Analysis

Ethnic Group	Ja Aliyia	Shilluk	Njok	Danagla
# negative competencies	0	0	0	0
largest eigenvalue	7.38	6.64	5.86	5.37
second largest eigenvalue	1.51	2.08	2.36	1.86
ratio of largest to next	4.89	3.18	2.48	2.89

The average competence scores for individual respondents range from 0.35 (Respondent 26) to 0.77 (Respondent 1), with no negative values. Thirteen of the seventeen respondents (Table 3) met or exceeded the average competency threshold of 0.50 to indicate shared knowledge (Weller 2007, 363). The average across all respondents' competence scores for each ethnic group ranged from 0.71 for the Ja Aliyia responses to 0.52 for the Danagla respondents (Table 3) indicating agreement within the data (Weller 2007; Weller 1987).

Table 3: Respondent Competency Scores from Consensuses Analysis

Respondent	Ja Aliyia	Shilluk	Danagla	Njok	Individual Respondent Average
1	0.85	0.69	0.76	0.79	0.77
3	0.91	0.62	0.83	0.55	0.73
4	0.55	0.56	0.71	-	0.61
5	-	0.79	-	0.60	0.70
6	0.25	0.82	0.46	0.43	0.49
7	0.77	0.60	0.84	0.04	0.56
16	0.73	0.63	0.06	0.57	0.50
17	0.71	0.26	0.73	0.52	0.56
18	-	0.39	-	0.36	0.38
19	0.79	0.75	0.02	0.63	0.55
20	-	-	0.06	0.69	0.38
21	0.80	0.24	0.63	0.79	0.62
23	0.75	0.83	0.55	0.63	0.69
24	0.68	0.84	0.47	0.64	0.66
25	0.87	0.62	0.81	0.75	0.76
26	0.50	0.31	0.06	0.51	0.35
27	0.74	0.83	0.74	0.74	0.76
Average	0.71	0.61	0.52	0.58	

The responses formed a strong consensus for the locations of the Ja Aliyia and the Shilluk ethnic groups. The responses for the Ngok and Danagla ethnic groups, though they did not meet the conventional three-to-one ratio to reach a consensus, showed strong signs of agreement. For example, the Ngok response polygons had the greatest number of overlapping responses (13). The Ngok respondents also had the lowest amount of agreement area lost between threshold ≥ 3 and ≥ 10 and the response polygons on average overlapped 50% with the ≥ 7 and ≥ 10 threshold areas. This shows that there is a precise area where 13 experts agree the Ngok are located. In the consensus analysis; however, the Ngok had the lowest eigenvalue ratio of 2.48.

DISCUSSION

The Geospatial Similarity Analysis of the ethnic groups showed that there were indeed areas of agreement between SMEs, and the geospatial consensus analysis verified that there was shared spatial knowledge between the respondents. Respondents' drawn polygon size, livelihood, and mapping orientation affects agreement.

Size Matters

The respondents who made larger polygons were more likely to overlap with other respondents and thus receive a higher level of agreement using the EARL data. Respondent 23 was found to have one of the largest individual polygons when compared to the polygon size of all other respondents; Respondent 23 had the largest response for Danagla, the second largest response for the Ngok, and the third largest for Ja Aliyian. Out of 17 Shilluk respondents, the smallest response polygon covered 2,651 km^2 (respondent 18) and the largest response polygon covered 34 times that (90,661 km^2) (respondent 23), an 88,000 km^2 difference. Out of the four ethnic groups, the average size of the responses was 32,305 km^2 and the average size of Respondent 23's responses were 101,438 km^2. The CCM showed that Respondent 23 had an average competency score of 0.69, which is the sixth highest out of 17 respondents, ranging from 0.35 to 0.77 (Table 3). Although the size of individual response polygons affects the Geospatial Similarity Analysis, the size does not directly determine the CCM eigenvalues. A bivariate Pearson two-tailed correlation analysis was run on the respondent's individual polygon size and their CCM competency scores for each ethnic group. The correlation analysis showed that the p values for each ethnic group were less than 0.05, meaning that the size of respondents' polygons did not affect the CCM competency scores.

Orientation and Livelihood

The Ja Aliyia and the Shilluk had high agreement according to the geospatial consensus analysis (Table 2). This could be because both are located near a large city and along a river both of which were clearly mapped. The Ja Aliyin were perceived to be just north of Khartoum, the capital (Appendix F) and along the Nile River. The Shilluk were perceived to be located in and north of Malakal, a large city in the north of South Sudan (Appendix F) and along the White Nile River.

On the other hand, the Ngok and Danagla had the lowest agreement for the geospatial consensus analysis (Table 2). The Danagla were perceived to be located along the Nile River but there is no city center to help orient respondents (Appendix F). All of the Danagla respondents placed them along the Nile, north of Khartoum, and all but three placed them in the Northern State (Ash Shimaliyya) near its capital of Dunqulah, which was not marked on the response map. The Ngok were perceived to be located in the southern half of a disputed area along the Sudan/South Sudan border, in the state of Southern Kordofan and extending into South Sudan states of Warab, Unity, and Northern Bahar (Appendix F). This relatively lower consensus could be due to the unrest in the area. However, the fact that the Ngok had the highest number of overlaps shows that there is a high level of consensus about their general location.

The variation in CCM scores could also be due to livelihood as it relates to production activities. The Ja Aliyin and the Shilluk are both mostly settled and not migratory. The Ja Aliyin in the nineteenth century became merchants and traders of goods and slaves, but mostly live a settled agriculturist's life herding camels and cattle (Lobban et al. 2002). The Shilluk are also settled farmers who fish and own livestock and dairy cows (Lobban et al. 2002). Since both lead predominantly sedentary lifestyles, their relatively consistent location helps explain their higher CCM scores. The Danagla, on the other hand, are known for being migrant merchants who traded all over Sudan and South Sudan (Lobban et al. 2002). The Ngok has been settled for over two centuries in Abyei (Johnson 2008); however, the area is disputed between Sudan and South Sudan (Themner and Wallensteen 2012). The unsettled nature of the Shilluk and the Ngok likely explains their lower CCM scores.

Relaxing the Cultural Consensus Model for Geospatial Studies

Due to the high degree of individuality, regarding SMEs' ethnic group location responses (the data used to produce the correlation matrices), the informal CCM may need to have a more flexible eigenvalue ratio (traditionally three-to-one) to indicate shared knowledge. The formal CCM uses true/false, multiple-choice, or ranked data and thus has a finite number of possible responses where only one response is culturally correct (Romney et al. 1986). The informal CCM used here is able to evaluate any correlation matrix such as one created from pile-sort data. Boster and Johnson (1989) conducted free (unrestricted) pile sort interviews and asked expert and non-expert respondents to sort types of fish into as many or as few piles, as they see fit, based on their perceptions. This method may lead to some respondents making many piles while others make few. Parallel to the variation seen in pile sort data is the variation observed in geospatial data, where each response varies in size (some respondents draw relatively large polygons, while others draw small polygons, a few respondents may even draw multiple polygons while most draw one). This variation within possible responses is problematic because it makes comparing individual responses difficult; this is known as the lumper-splitter problem (Weller and Romney 1988).

The varied nature of data used in the informal CCM presents limitations in regards to the resulting eigenvalues. The CCM results for the Ngok and Danagla ethnic groups did not meet the necessary eigenvalue ratio (three-to-one) to indicate shared knowledge. The results from the Geospatial Similarity Analysis show that these two ethnic groups are perceived to be in a single, generally agreed upon location by multiple experts. This ratio may need to be more flexible when utilizing the informal CCM because the data used to produce the correlation matrices have a higher degree of individuality.

Recommendations for Future Research

Several adjustments are recommended for future research. Orienting the respondents proved problematic within states that did not have major features such as large rivers or main roads. Even though it may add what looks like clutter, the state capitals would be helpful orienting features in all of the states.

The open-endedness of our drawing instructions led to great variation in the size and position of respondents' drawn polygons, possibly interfering with and/or stunting the amount of overlap/agreement between respondents. Using more specific drawing instructions, such as, "designate the center of an ethnic group location and then draw a polygon for the full extent of that same ethnic group," could help standardize the geospatial responses and possibly increase the amount of overlap/agreement. While conducting this type of analysis it may also be interesting to quantify and analyze the area where respondents fail to overlap with each other.

Soliciting interviews from a greater number of experts would also be helpful in future research, because it would allow for a greater understanding of agreement thresholds. By incorporating more people with diverse educational backgrounds, the respondent overlap/agreement could be greater for all the ethnic groups. With the presumed increase in overlap/agreement, a researcher could use a scree plot of additional unique polygons by area, to help determine when adding additional people no longer improves the map. It would also be interesting to include people with varying levels of expertise; this may help illuminate any education-based bias. This study focused on how to aggregate and analyze multiple geospatial perceptions; however, the accuracy of these perceptions is not addressed. Using outside, independent sources to verify the geospatial perceptions would aid in understanding and measuring their accuracy.

CONCLUSION

This research focused on how to synthesize and gain meaning from multiple geospatially explicit responses. Through the Geospatial Similarity Analysis, it was determined that to identify the location of an ethnic group there needs to be a minimum of seven expert responses that overlap. Geospatially applying the CCM attached meaning to the geospatial response overlap, indicating that for two of the four ethnic groups analyzed, the respondents' overlap was based on shared knowledge and formed a consensus. Being able to aggregate and add meaning to geospatially explicit perceptions of multiple SMEs enables researchers to produce geospatial data that is current, and incorporates multiple perspectives at a relatively rapid pace. Though there is still much work to be done, maps created using this method have the potential to be powerful avenues of communication between researchers and decision makers.

CHAPTER 3

COMPARING SUBJECT-MATTER-EXPERTS' PERCEPTIONS OF SUDANESE ETHNIC GROUP LOCATIONS TO AN ONLINE NEWS SOURCE

INTRODUCTION

The movement of ethnic group populations across Sudan and South Sudan has been in flux for decades making the precise location of ethnic groups difficult to pin-point. The ethnic group conflict in Sudan and South Sudan began with unified Sudan's independence in 1955 and has persisted until their division in July 2011 (Lobban, Kramer and Fluehr-Lobban 2002; Themner and Wallensteen 2012; Lobban and Fluehr-Lobban 2012). Conflict has continued since the separation in 2011 over the Abeyi region, along the new border (Themner and Wallensteen 2012). This near constant ethnic group conflict fosters an environment of mass movements and shifts in the locations of various populations, displacing millions from their homelands and livelihoods (Ayers 2010). Characterizing these changes within countries that are as large (2,505,801 km^2, larger than Alaska and Texas combined) and as populated (60,602,000 by 2025) (Lobban et al. 2002, 326) as Sudan and South Sudan is challenging.

To better understand ethnic group location, the dynamic distribution of ethnic groups needs to be better documented, so that distribution due to factors such as resources or livelihood (sedentary/nomadic) are not confused with the distribution induced by conflict. A way to make the process of documenting the locations of multiple ethnic populations more feasible is to solicit and analyze Sudan and South Sudan subject-matter-expert (SME) perceptions regarding the location of ethnic groups, using participatory mapping and participatory geographic information systems (PGIS) methodologies (Chapter 2). The analysis of the aggregated SME perceptions through the use of geospatial similarity and the CCM showed that areas where seven or more SMEs indicated an ethnic group to be located was sufficient for agreement and in some cases indicated a consensus among the SMEs (Chapter 2).

Now that multiple geospatially explicit SME perceptions of ethnic group location have been aggregated and analyzed, those locations need to be examined for accuracy. One way to verify the SME responses is to geospatially link news articles from an independent data source that refers to ethnic groups and their location and compares their relative locations. If the aggregated SME responses provide the same locations indicated by the geospatially linked network text analysis, then such a method could be used to update the locations of ethnic groups and other under-represented populations that are written about in the news.

Van Holt et al. (2012) showed that incidents of severe conflict that were reported in the Sudan Tribune were associated with the presence of multiple ethnic groups reported at the same location. Van Holt used the network analysis software ORA (Carley et al. 2011b) to apply the Data-to-Model (D2M) approach (Carley et al. 2011a) to network text analysis (Carley 1997). The D2M approach allows researchers to recognize relationships between words (actors, organizations, etc.) and the core concepts of the text (Popping 2000), as a means to depict social networks (Diesner and Carley 2005). When the text data refers to a specific group or region, researchers are able to generate a prompt ethnographic assessment, which serves as a socio-cultural profile (Carley, Bigrigg and Diallo 2012). In Van Holt et al. (2012), ethnic groups, severe-conflict terms, and locations were the terms analyzed, and the co-occurrences of ethnic group by location and severe conflict by location were mapped.

Compared to conventional scientific approaches of ethnographic assessment, D2M (Carley et al. 2011a; Carley et al. 2011b) combines vast amounts of data from key sources and is an expedient process that can be done in a matter of weeks (Pfeffer and Carley 2012). In light of the previously stated advantages associated with D2M, it has the potential to function as a verification method for any type of data that refers to a specific group or region. In this case it will be used to verify geospatially explicit location data that focuses on ethnic groups in Sudan and South Sudan. If decision makers and researchers are able to conceptualize and visualize the dynamic flow of ethnic groups in regions and states where politics and populations are volatile, it will greatly aid in understanding conflict.

RESEARCH OBJECTIVES

To understand the extent of agreement regarding ethnic group location, we ask the following research question: Is there agreement between SME perceptions of ethnic group location and the reported location of ethnic groups through the Sudan Tribune? Once this question is answered, it is important to verify and determine how well SME's perceptions of ethnic group location compare to co-occurrences of locations and ethnic groups within an online source, the Sudan Tribune (www.sudantribune.com, 2004-2008), a content analysis was used to identify locations within the Sudan Tribune news articles that appeared within seven words of an ethnic group name, referred to as *ST locations*. The *ST locations* are then geospatially linked and compared to the respective *SME location*.

To understand the context in which Sudanese ethnic group names appear, we asked: Do any of the contextual classifications have a higher level of geospatial overlap with the ethnic group locations indicated by SMEs? An *ST occurrence* is defined as the co-occurrence of an ethnic group name and an *ST location* within seven words of each other in the text. In order to determine the meanings of the *ST occurrences,* the context for each was classified. The *ST occurrences* that are contextually classified as indigenous land are hypothesized to have greater agreement (shorter geographic distance) between the *SME locations* and the *ST locations,* than *ST occurrences* that are contextually classified into one of the other seven contextual classifications (ethnic conflict, political conflict, historical, resource, oil, and other).

To determine if the distribution and recognition of ethnic groups is influenced by their livelihood classification (nomadic/sedentary), we asked a third research question: Is there greater agreement between SME location and ST locations (less distance) for ethnic groups with sedentary livelihoods rather than with nomadic livelihoods? The agreement between *SME location* and *ST location*s for sedentary ethnic groups is hypothesized to be greater (shorter distance) than the agreement between *ST location* and *SME location*s depicting nomadic ethnic groups (longer distance).

METHODS

To identify where SMEs perceive ethnic groups to be located, the centroid of their agreement area was used (Chapter 2). To identify locations that are connected to specific ethnic groups in the Sudan Tribune, a network text analysis and a manual content analysis were conducted. To quantify the geospatial agreement between the *SME locations* and the *ST locations* the straight-line distance was determined. A Tukey HSD Test was applied to determine if the ethnic groups themselves had an effect on the distance between *SME locations* and *ST locations*. To better characterize the co-occurrences ethnic group names had with location names in the original source text of the Sudan Tribune, the *ST Occurrences* were classified into one of seven contextual classifications. In order to better understand the relationships between the seven contextual classifications and five ethnic groups, a correspondence analysis was used to generate a cross-tabular table, effectively synthesizing the *ST occurrences*. To determine if the livelihood classification of sedentary rather than nomadic ethnic groups has greater agreement between

SMEs and the Sudan Tribune, a two sample t-test was applied to determine if a significant difference exists between the average *ST/SME distance* for nomadic and sedentary ethnic groups. Then two ANOVAs were conducted to test if a significant difference between nomadic and sedentary ethnic groups average *ST/SME distance* (found in the t-test) was influenced by the livelihood classification itself.

Data Sources

To perform the network text analysis, the text published in the Sudan Tribune from 2004 to 2008 was used, this data source was chosen because it is an English language, online news source (www.sudantribune.com), with an easily accessible archive. To conduct a network text analysis, a thesaurus must be developed that includes terms and words of interest. The ethnic groups that were included in the thesaurus were compiled by SME Richard Lobban, who has worked in Sudan for the past thirty years (Lobban and Fluehr-Lobban 2012). Locations were coded and linked geospatially, using the National Geospatial Intelligence Agency dataset for Sudan and South Sudan (50,954 recoded locations) (NGIA 2011).

Participatory Data Collection & Analysis

For the SME data collection, two maps were created using ArcMap (ESRI 2011): a response map used by respondents to record (draw) ethnic group locations (Appendix B) and the reference map used to orient respondents (Appendix C). We interviewed SMEs attending *The Sudanese Studies Conference* in Tempe, Arizona (May 18-20, 2012) and asked them to indicate where ethnic groups were located on the map. The hand-drawn locations were digitized and aggregated. Then, through the Geospatial Similarity Analysis and by applying the CCM to geospatial data, it was determined that when seven or more respondents were in geospatial agreement the individual responses overlapped more than 50% with that area (see Chapter 2 for details). Five ethnic groups were chosen for analysis because they met the SME agreement threshold of ≥ 7 (See Chapter 2) and were coded in the semi-automated network analysis. The five ethnic groups include the Ngok, Shilluk, Fur, Beja and Bari.

The centroid of the area where seven or more SMEs agreed is the *SME location* for the comparative analysis between the locations cited in the Sudan Tribune and the locations indicated by SMEs. There were some instances where seven or more SMEs agreed but the agreement area did not form a single polygon. In these instances, a single centroid was created at the center of the multiple polygons.

Network Analysis to Extract Desired Text from the Sudan Tribune

AutoMap (Carley et al. 2011a) was used to generate a two-mode network depicting ethnic groups by locations. AutoMap searches for the words provided in a user-defined thesaurus (in this case ethnic group names and location names) and generates links between terms that are found within seven words of each other. A window size (word distance) of seven was chosen to preserve the relational meanings (Van Holt et al. 2012). If the word distance was greater than seven words, the network would pick up on many connections that are contextually disconnected. If the word distance was less than seven words, there would be unrecognized contextual connections. For example, in the text, "The SPLA army has currently positioned itself as an occupation force in **South Sudan**[1] which is grabbing lands of the **Bari** tribe in **Central Equatoria State** (CES) without any regard to the rule of law (Odiong, 2007)," Automap would recognize words coded for in the thesaurus and would generate a link between the Bari, South Sudan, and CES. However, if the text read, "The SPLA army has currently positioned itself in **South Sudan** as an occupation force which is grabbing **Bari** tribe lands in **Central Equatoria State** (CES) without any regard to the rule of law," only CES and Bari would be linked.

[1] Bold indicates terms included in the user defined thesaurus.

The thesaurus was developed to code the text using an a priori and inductive approach. The a priori approach codes for words that are defined by the research team, based on previous knowledge of the theory and related research. Words coded using the inductive approach were high frequency words that appeared more than twenty times throughout all five years of data and that did not get coded for initially in the a priori thesaurus. To gather the relevant source text, the search tool in ORA, a network analysis program (Carley et al. 2011b) was used to extract all the articles that had one of the five ethnic group names in it, a total of 361 articles.

Content Analysis to Identify the Co-occurrences of Locations & Ethnic Groups

Examining the 361 Sudan Tribune news articles first requires the removal of duplicate articles. There were sixty-nine articles that were duplicates of text or short clips of other full articles (for example, documents that the Sudan Tribune used as news alerts). The removal of duplicates brought the number of articles down to 292. Within the remaining articles, there were three that did not contain the ethnic groups of interest, leaving 289 articles to examine.

Each of the 289 articles was manually searched for the ethnic group name, and all location names within seven words were recorded, per article, as an *ST occurrence*. Each *ST occurrence*, is a dyadic relationship between an ethnic group name and location in a single article. For example, if Beja and Khartoum appeared within seven words of each other more than once in a single article, only one occurrence would be recorded; if Beja and Khartoum appear within seven words of each other in multiple articles, each of those dyads is an occurrence.

The original source texts, from the Sudan Tribune, were examined to verify the co-occurrences and understand the context. Seven location names were found in the Sudan Tribune that were not included in the NGIA datasets of Sudan and South Sudan and thus were not coded in the thesaurus. To create a more detailed model of the distribution of ethnic groups across

Sudan and South Sudan, the seven locations (four domestic, three international), were geospatially linked. To accomplish this, Google Earth was used and the geographic coordinates for available locations were recorded, classified, and mapped.

Scale Classification of ST Locations

The *ST location*s were classified for scale into one of six scale classes: villages, cities, states, mountain regions (Nuba Mountains, Red Sea Hills), sub-country regions (West Sudan, East Sudan, Southern Blue Nile), and countries. Many of the villages were not geospatially linked because we were unable to locate them in the NGIA dataset or on Google Earth. The regional location references such as Eastern Sudan or Southern Blue Nile were also not geospatially linked because the boundaries are not clearly defined. With the unlinked locations and regional references excluded, there were 228 *ST occurrences* mapped out of 288 total *ST occurrences*.

ST/SME Distance to Measure Geospatial Agreement

To quantify the geospatial agreement between the SMEs' perceptions of ethnic group location and the locations indicated in the Sudan Tribune, the Euclidean distance (km) between each *ST location* and the respective *SME location* (*ST/SME distance*) was determined using the near tool in ArcGIS (ESRI 2011). To characterize the *ST location*s that were originally polygons (locations such as states or countries), the centroid of the polygon was generated. All the data were projected into the Africa Equidistant Conic projection to obtain the most accurate distance measures.

Ethnic Group Effect on ST/SME Distance

To determine if the ethnic group itself has an effect on the distance between geospatially linked *ST occurrences* and the location indicated by SMEs, a Tukey Honestly Significant Difference (Tukey HSD) test was used. To quantify the significant differences between ethnic groups' average *ST/SME distance,* the Tukey HSD Test was conducted using the individual *ST/SME distance* measured for each geospatially linked *ST occurrence* (N=228), classified by ethnic group.

Contextual Classification of ST Occurrences

Analyzing the *ST/SME distance* helps us to geospatially conceptualize how well the *ST locations* agree with the *SME locations*. However to genuinely understand how the locations in the Sudan Tribune are connected to the ethnic group, the original Sudan Tribune articles were consulted and characterized by manually conducting a content analysis. A manual search within seven words on either side of the ethnic group name was conducted and any location reference was recorded as an *ST location* and geospatially linked. The word distance of seven was applied consistently in order to preserve continuity of analysis from the semi-automated network text analysis to the manual content analysis. All the scale classes were included in the content analysis (villages, cities, states, mountain regions, sub-country regions and countries).

There were 114 articles that did not have a location name within seven words of the ethnic group name. The 175 remaining articles (Bari 5, Beja 77, Fur 6, Ngok 25, Shilluk 62) that contained *ST location*s were then classified for context, creating an *ST occurrence*. If the same location appeared more than once as an *ST location* within a single article, only one *ST occurrence* was recorded and the classification was a synthesis of all the connections that location had with that ethnic group name. Each article ranged from having one to eight *ST occurrences*. The *ST occurrences* were text examined for context and classified into one of seven classifications: indigenous land, ethnic conflict, political conflict, historical, resource, oil, and other.

The dyadic relationships classified as indigenous land were situations in the text describing ownership of land or text referring to an ethnic group as original/historical residents. For example a dyad that reads,

Augustburger said aid agencies had until recently failed to recognize the importance of what was happening in the **Shilluk**[2] Kingdom of northern **Upper Nile** because they were preoccupied with the huge humanitarian crisis in Darfur, where some one-million people have fled their homes (ST 2004).

would be identified and classified as indigenous land because it is describing the ethnic group as being originally from there, people of this land.

Dyads that were classified as ethnic conflict had a context that conveyed the ethnic group in question being in conflict with another specified ethnic group. For example, "'Heavy fighting' broke out in the town of **Malakal** on Thursday morning between Nuer and **Shilluk** ethnic groups, according to local sources" (IRIN 2004). The relationships classified as political conflict involved context that conveyed the ethnic group or member of the specified ethnic group as being in conflict with a political entity including police, military, government officials, or the government in general with references to Khartoum. For example,

Chronic poverty and neglect by the authorities prompted the region's largest ethnic group, the **Beja**, to take up arms against **Khartoum** in 1996, eventually forging an alliance with the much larger, southern-based Sudan People's Liberation Movement (Reeves 2006a).

[2] Bold indicates the ethnic group name or the location name identified in the Sudan Tribune for the contextual classification of ST occurrences.

The dyadic relationships that were classified as historic were any happenings before Sudan's united independence in 1955. For example, "According to the CPA Abyei is defined as the area of the nine **Ngok** Dinka chiefdoms transferred to **Kordofan** (in the North) in 1905" (Lupai 2007).

Oil and resources were classified separately due to the known importance of oil in Sudan and South Sudan (Patey 2007). Those that were classified as oil were exclusively referring to the breakdown of oil revenue between the Sudan, South Sudan, and the regions within Abyeia. For example,

Net oil revenues to be divided six ways during the interim period: 50 percent for the national government, 42 percent for the government of South Sudan, and two percent each for (southern) **Bahr el Ghazal** region, (northern) **Western Kordofan**, **Ngok** Dinka people and Misseriya people (Factbox 2004).

When more than one location was present within seven words of the ethnic group name, both were classified independently, but within the same classification, in this case oil. The resource classification was assigned when an ethnic group was being described within the context of resource scarcity or abundance. For example, "But in the slums of **Port Sudan**, the **Beja** and other pastoralists displaced from the surrounding rural areas by drought, the seizure of prime land and food shortages live on less than $1 a day" (Reeves 2006b).

The classification of "other" was used for connections that were contextually absent between the location and the ethnic group name. This included instances where the article was written in a location but the article was about somewhere or something else or when the ethnic group name and location were referred to independently. For example, It also has targeted African Muslims in the Nuba Mountains and attacked mosques and religious schools of the rebellious **Beja** people in the east. In **Darfur**, as many as 50,000 have died and 1.2 million have been driven from their villages in a conflict that is more racially based than religious (Raghavan 2004).

Also included in the "other" classification were instances in the text where a location name was referring to the ethnic group, such as Masalite which is both the name of a place and the name of an ethnic group and was inappropriately connected as a location three times to the Fur ethnic group. To determine if any particular class of context is in greater agreement with the SMEs' perceptions than any other contextual class, the total *ST occurrences* for each contextual classification was determined and the average *ST/SME distance* by contextual classification was calculated.

Correspondence Analysis

To aggregate and analyze the contextual classifications assigned to each *ST occurrence* the relationships between the five ethnic groups and the seven contextual classifications were synthesized into a cross-tabular table using UCINET (Borgatti, Everett and Freeman 2002). To determine if the relationship between the contextual classifications (rows) and ethnic groups (columns) is significant and not produced by chance, StatXact8 (Mehta and Patel 2001) was used to calculate the Chi-Square statistic, because it estimates the exact p-value by applying the Monte Carlo test. To visualize the correspondence analysis UCINET's visualization tool, Netdraw, was used to represent the cell values of the table as links in a network reflecting the row and column relationships.

Livelihood Classification of Ethnic Groups

To understand the variation between nomadic and sedentary ethnic groups, the five ethnic groups were coded as nomadic or sedentary by SME, Richard Lobban (Lobban 2012). Two out of the five ethnic groups being analyzed the Beja and the Ngok, are normally nomadic people (Lobban 2012, Cutler 1991). The Beja are nomadic pastoralists and the Ngok are cattle herders (Lobban, Kramer and Fluehr-Lobban 2002). The Fur, Shilluk and Bari ethnic groups all live sedentary lifestyles (Lobban 2012). The Fur ethnic group consists of hill famers with livestock that do not require migration, the Shilluk are farmers and fisherman, and the Bari are slash and burn farmers (Lobban 2012). The lifestyles described above are how these ethnic groups have classically been characterized; however, the reality can be quite different due to displacement events, such as drought and conflict (Lobban 2012).

A two sample t-test was used to determine if there is a significant difference between the average *ST/SME distance* for the livelihood classifications (nomadic/sedentary). The two sample t-test used the *ST/SME distance* as the dependent variable and livelihood classification as the independent-grouping variable.

Two ANOVAs were then conducted to determine if the difference between nomadic and sedentary ethnic groups with respect to how well the *SME locations* agreed with the *ST locations*, was influenced by the livelihood classification itself. The international locations, those outside of Sudan, were excluded from these analyses (34 *ST occurrences*). The first ANOVA was conducted to evaluate the variation between ethnic groups with regard to their average *ST/SME distance*. The ethnic group ANOVA used the *ST/SME distance* as the dependent variable and ethnic group as the independent variable, meaning that the ANOVA generated the mean *ST/SME distance* for each ethnic group and then used that mean to analyze the total variation explained by ethnic group. The second ANOVA was then used to evaluate if any more of the variation between ethnic groups regarding *ST/SME distance* would be explained by the livelihood classification. The livelihood ANOVA also used *ST/SME distance* as the dependent variable and added livelihood classification as a second independent variable (nomadic/sedentary).

RESULTS

Is there agreement between SME perceptions of ethnic group location and the reported location of ethnic groups through the Sudan Tribune?

None of the *ST locations* were precisely the same as the *SME locations* since the centroid of the *SME location* was used for analysis, instead of a polygon. Of the five ethnic groups analyzed, three had *ST occurrences* located in the SME agreement area, defined as an area where seven or more SMEs agree an ethnic group is located. All five of the analyzed ethnic groups that had *ST occurrences* that were not located in this area as well. The average distance between the *ST locations* and the *SME locations* was 357 km, with a maximum of 2,091 km and a minimum of 51 km (Table 4). The Bari, who are slash and burn farmers, had the lowest average distance (179 km) between the locations taken from the Sudan Tribune and where SMEs perceived them to be (Table 4). The Beja, a traditionally nomadic ethnic group, had the largest average distance (722 km) suggesting that the Beja are more geospatially spread out than the other ethnic groups analyzed (Table 4).

Table 4: *ST/SME Distance* Statistics by Ethnic Group

Ethnic Group	Total *ST Occurrences*	Geospatially Linked *ST Occurrences*	Average *ST/SME Distance* (km)	Range
Bari	7	6	179	53-304
Beja	135	97	722	72-2,091
Fur	14	3	420	218-823
Ngok	37	32	197	51-449
Shilluk	95	90	269	62-873
Total	288	228	357	51-2,091

The Tukey HSD test showed that there are significant differences between the average *ST/SME distances* for the Beja and Ngok (p=0.000), the Beja and Bari (p=0.032), and the Beja and Shilluk (p=0.000) (Table 5). The Beja were found to be significantly different from the three ethnic groups with the lowest average *ST/SME distance* (Table 5). The significant difference found by the Tukey HSD (Table 5) test between the two nomadic ethnic groups (Beja and Ngok) indicates that the livelihood classification does not explain variation in *ST/SME distance* by ethnic group.

Table 5: Tukey's Honestly-Significant-Difference Test of the variation between the average *ST/SME Distance* of each Ethnic Group.

Ethnic Group	Ethnic Group	Difference in Average ST/SME Distance (km)	p-value	95% Confidence Interval	
				Lower	Upper
Bari	Beja	-627.05	0.032	-1220.71	-33.39
Bari	Fur	-366.33	0.742	-1187.44	454.78
Bari	Ngok	-143.48	0.968	-750.70	463.74
Bari	Shilluk	-150.08	0.959	-743.36	443.20
Beja	Fur	260.72	0.753	-332.94	854.38
Beja	Ngok	483.57	0.000	266.94	700.19
Beja	Shilluk	476.96	0.000	303.20	650.73
Fur	Ngok	222.85	0.855	-384.37	830.07
Fur	Shilluk	216.25	0.858	-377.03	809.53
Ngok	Shilluk	-6.60	1.000	-222.19	208.98

Do any of the contextual classifications have a higher level of geospatial overlap with the ethnic group locations indicated by SMEs?

The content analysis of 175 Sudan Tribune articles resulted in a total of 288 *ST occurrences* that were categorized into one of six classifications. The geospatially linked co-occurrences contextually classified as indigenous land had the shortest average *ST/SME distance* (179 km) suggesting that when ethnic groups are mentioned in news articles concerning their indigenous land, the location names associated with the ethnic group name agreed with the locations indicated by SMEs (Table 6).

Table 6: *ST/SME Distance* Statistics by Context Classification

Contextual Classification	Total *ST Occurrences*	Geospatially Linked *ST Occurrences*	Average *ST/SME Distance* (km)	Range
Indigenous Land	116	83	206	51-1,299
Political Conflict	110	97	576	62-1,299
Ethnic Conflict	15	12	722	62-1,299
Historical	9	9	328	51-449
Resources	7	6	631	51-1299
Oil	4	4	260	151-297
Other	27	17	793	75-2091
Total	288	228		51-2,091

A correspondence analysis of these occurrences showed that the association between the ethnic groups and the contextual classifications are significant, with a Chi-square p-value of less than 0.000 (Table 7). The three ethnic groups with the lowest average *ST/SME distance* (the Bari, Fur, and Ngok) had more than 50% of their occurrences classified as indigenous land (Table 7). Most *ST occurrences* were classified as indigenous land (40%), followed by political conflict (38%) (Table 7).

Table 7: Contextual Classification by Ethnic Group Cross-Tabular Table with a Chi-squared p-value of 0.000

Classification	Bari	Beja	Fur	Ngok	Shilluk	Row Total
Indigenous Land	4	39	2	20	51	116
	57%	29%	14%	54%	54%	40%
Political Conflict	2	69	4	2	33	110
	29%	51%	29%	5%	35%	38%
Ethnic Conflict	1	6	3	0	5	15
	14%	4%	21%	0%	5%	5%
Historical	0	0	0	9	0	9
	-	-	-	24%	-	3%
Resources	0	5	0	2	0	7
	-	4%	-	5%	-	2%
Oil	0	0	0	4	0	4
	-	-	-	11%	-	1%
Other	0	16	5	0	6	27
	-	12%	36%	-	6%	9%
Column Total	7	135	14	37	95	288
	100%	100%	100%	100%	100%	100%

In the network visualization of the correspondence analysis (Figure 6) the red circles represent the contextual classifications and the blue squares represent the ethnic groups, the links connecting the red circles to the blue squares represent *ST Occurrences.* All five ethnic groups were connected to the contextual classifications of political conflict and indigenous land (Figure 6).

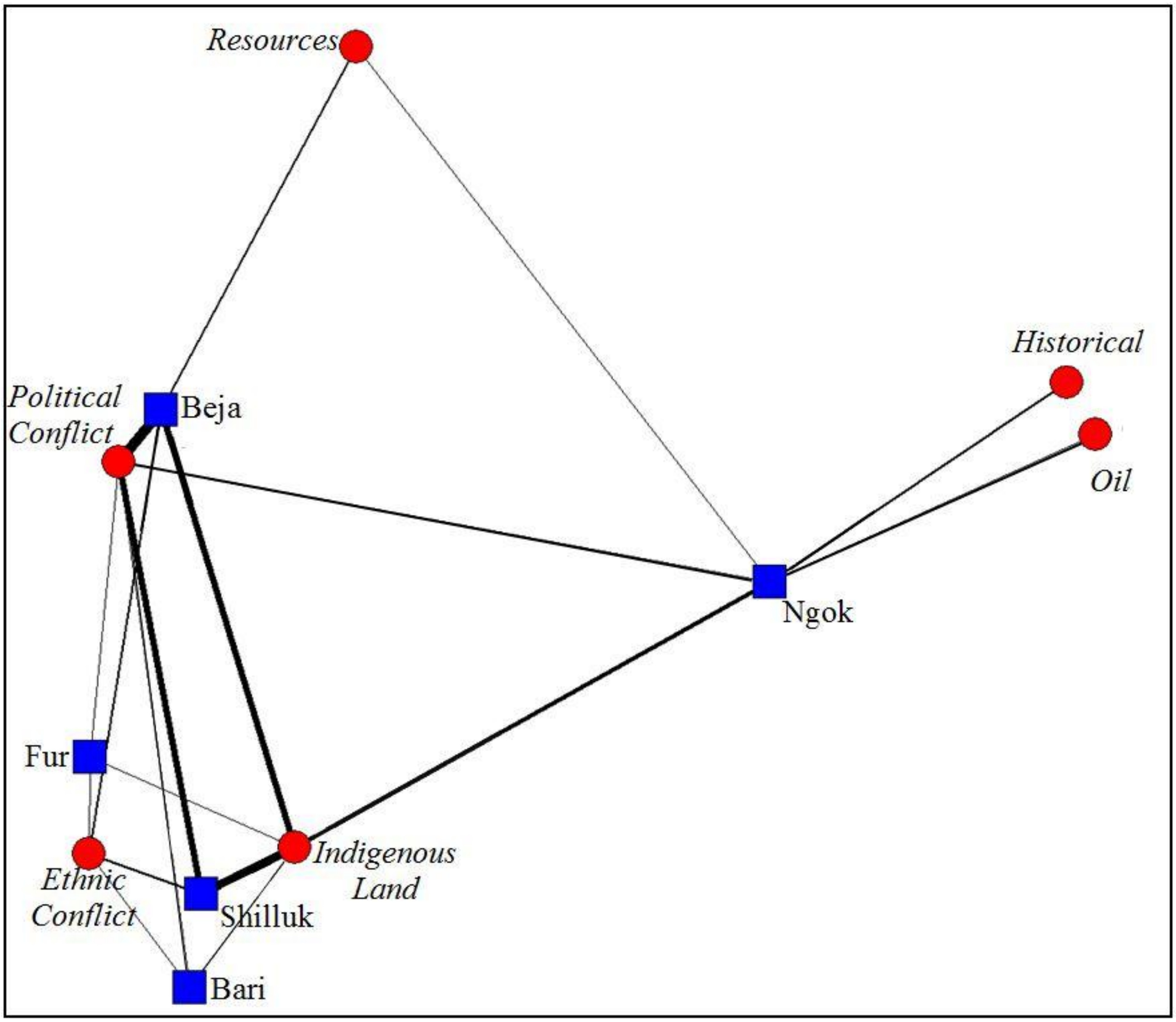


Figure 6: Correspondence Analysis (Network visualization of cross-tabular table depicting contextual classification by ethnic group)

The Bari had the highest amount of agreement between the *ST location*s and the *SME location* (Figure 7); they had the highest percentage of *ST occurrences* classified as indigenous land (57%) (Table 7) and the shortest average *ST/SME distance* (Table 4).

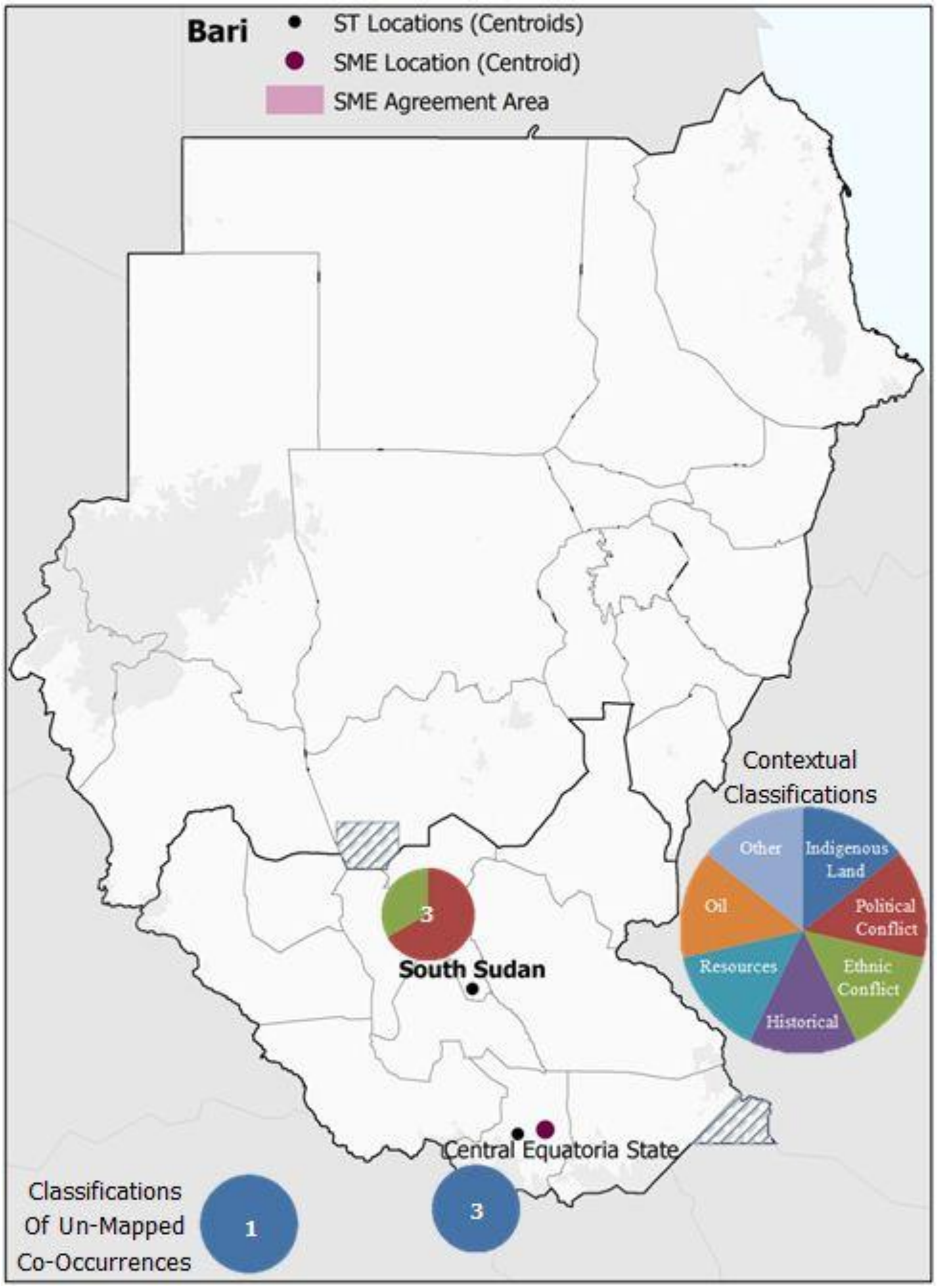


Figure 7: Geospatial Agreement between the Sudan Tribune and SME Perceptions of the Bari Ethnic Group and Contextual Classification by Location (The SME Bari Agreement Area is too small to see under the *SME location*)

The *ST occurrences* classified as “other” were done so because contextually the location did not connect to the ethnic group; this is evident in the fact that this contextual classification has the highest average *ST/SME distance* (793 km) (Table 6). The Fur had the largest portion of *ST occurrences* classified as other (36%) and ethnic conflict (21%) out of all five ethnic groups (Table 7). The ST occurrences classified as ethnic conflict have the second highest average *ST/SME distance* (722 km) (Table 6). This could be due to the fact that in some cases ethnic conflict occurs when one ethnic group leaves its homeland and forces other to leave theirs causing a chain reaction of displacement and ethnic groups to move further and further away. The Fur also had the second highest average distance (420 km) (Table 4) between *ST location* and *SME location*s or the second lowest level of agreement (Figure 8).

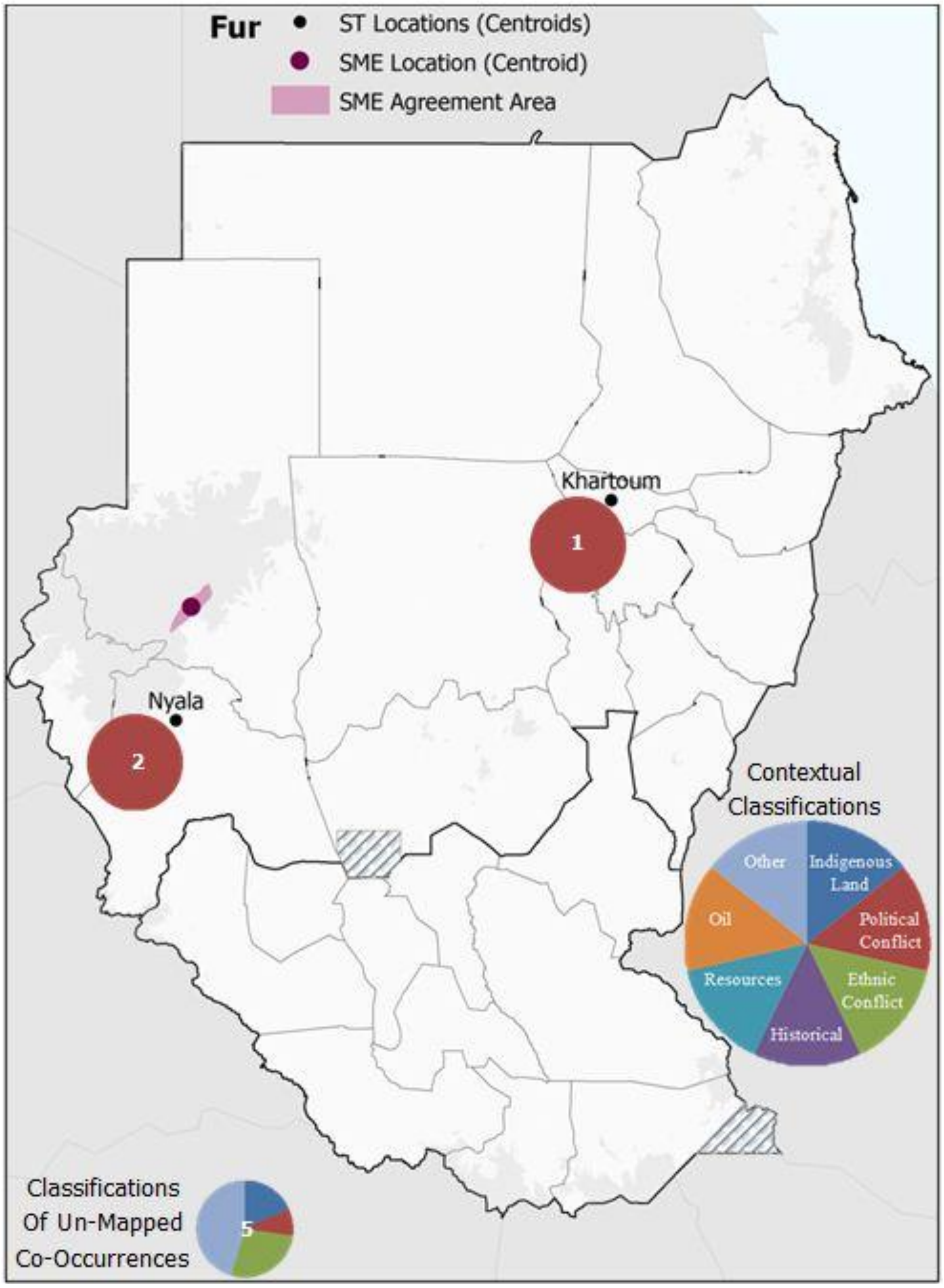


Figure 8: Geospatial Agreement between the Sudan Tribune and SME Perceptions of the Fur Ethnic Group and Contextual Classification by Location

The Beja accounted for 46% of the total occurrences and 42% of the geospatially linked occurrences. The Beja had the longest average *ST/SME distance* (722 km) (Table 4) meaning the least amount of agreement between the ST and SMEs (Figure 9). Of the Beja *ST occurrences* 51% were classified as political conflict, a larger portion than any other ethnic group (Table 7).

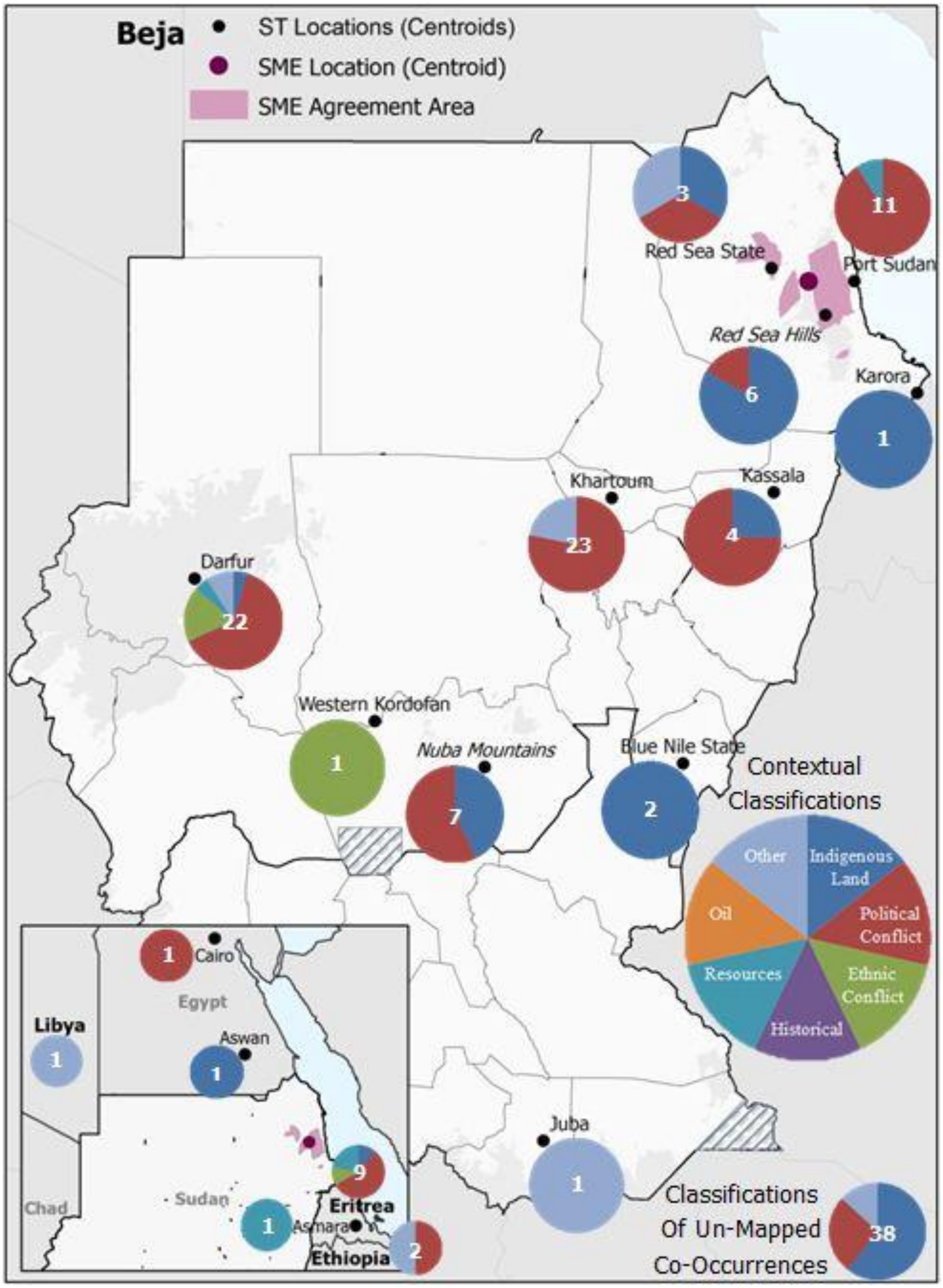


Figure 9: Geospatial Agreement between the Sudan Tribune and SME Perceptions of the Beja Ethnic Group and Contextual Classification by Location

Is there greater agreement between SME location and ST locations (less distance) for ethnic groups with sedentary livelihoods rather than with nomadic livelihoods?

There is a significant difference (p=0.000) in the average *ST/SME distance* between ethnic groups classified by livelihood (nomadic or sedentary) (Table 8). The two sample t-test showed that the *ST/SME distance* for sedentary ethnic groups was on average 206 km (Table 8). For nomadic ethnic groups, the *ST/SME distance* was on average 523 km (Table 8). The nomadic ethnic group names were connected to locations in the Sudan Tribune that were almost twice as far away from where SMEs perceived them to be than sedentary ethnic group names.

Table 8: Two Sample t-test of the variation between the average *ST/SME Distance* of nomadic and sedentary ethnic groups

Presence of Country Classification	Nomadic			Sedentary			Pooled Variance t	Significance of t (p-Value)
	N	Average *ST/SME Distance* (km)	SD	N	Average *ST/SME Distance* (km)	SD		
Included	106	560.34	477.76	96	236.51	254.18	-5.92	0.000
Excluded	98	522.51	477.42	74	206.13	282.73	-5.06	0.000

The ethnic group ANOVA that only incorporated ethnic groups, found that 32% (multiple R squared value=0.324, p=0.000) of the variation in *ST/SME distance* was explained by ethnic group classification. The livelihood ANOVA, that included the additional independent variable livelihood classification, did not increase the amount of variation explained (Multiple R Squared Value=0.324, p=0.000). Therefore, according to these ANOVA results, livelihood classification does not help to explain the variation in *ST/SME distance*.

DISCUSSION

The Influence of Context on Agreement between the Sudan Tribune and the SMEs

The ethnic group with the second lowest average *ST/SME distance* is the Ngok (197 km) (Table 4), meaning that relative to other groups, the Ngok had the highest agreement between the location indicated by SMEs and the locations found in the Sudan Tribune. Many of the Ngok have been displaced from Abyei and found refuge in Khartoum (Assal 2006; Cohen 2008). However, through this content analysis, no *ST occurrences* were found between the Ngok and Khartoum (Figure 10).

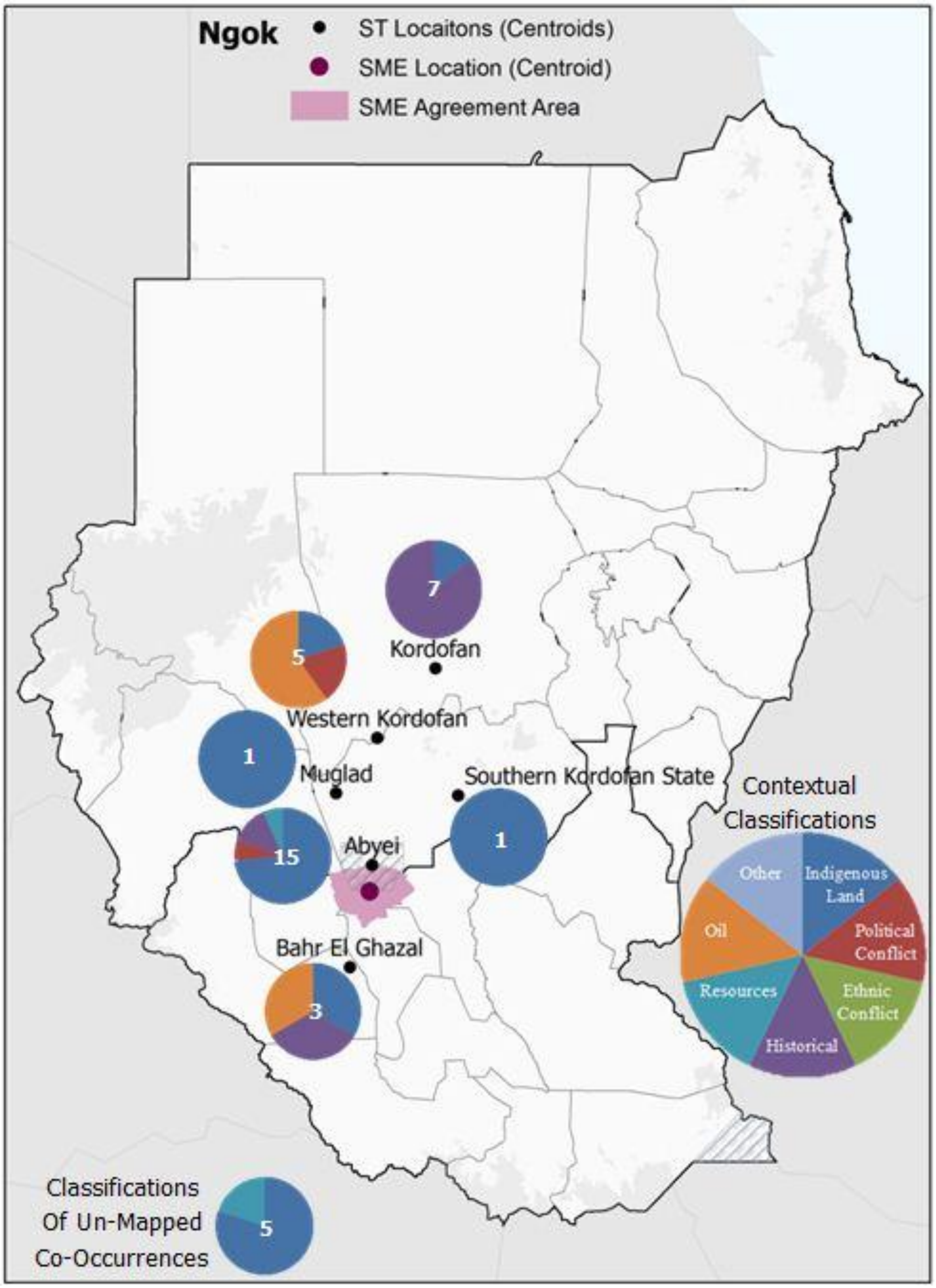


Figure 10: Geospatial Agreement between the Sudan Tribune and SME Perceptions of the Ngok Ethnic Group and Contextual Classification by Location

The Ngok is the only ethnic group analyzed that had *ST occurrences* classified as either oil (260 km) or historical (328 km) (Table 6). Articles published by the Sudan Tribune from 2004-2008 that contained the ethnic group name Ngok were more focused on the politics of their homeland rather than their displaced whereabouts. This suggests that retrieving contextual information from the original source text is necessary; otherwise false assumptions about the location of an ethnic group, are possible.

The classification of ethnic conflict has the second largest average distance (722 km) second only to the classification of other (793 km) (Table 6). The Fur had the highest percentage (21%) of *ST occurrences* classified as ethnic conflict (Table 7). However, none of the *ST occurrences* with the Fur were able to be spatially linked because of the fine scale at which their locations were cited (Figure 8). The *SME location* for the Fur is also located in the North Darfur State.

The recoded history of ethnic conflict involving the Fur shows that many of the actual conflicts took place on their indigenous lands, located in the Northern Darfur State. The Northern Darfur State experienced several years of drought in the mid-1980s which forced the nomadic Arab ethnic groups to move south in order to care for their camels, into lands traditionally inhabited by pastoralist ethnic groups (King and Osman 2004). The Fur (the largest ethnic group in the Darfur region) and other sedentary pastoralists in the area began having conflicts over water-wells with the nomads from the north, who had obtained guns from neighboring countries, Chad and Libya (King and Osman 2004). The conflict worsened when the Khartoum Prime Minister supplied guns to the pastoralists to help them defend their wells and homes (King and Osman 2004).

If the Fur *ST occurrences* classified as ethnic conflict had been geospatially linked, the average *ST/SME distance* may have been much shorter for the Fur and for ethnic conflict classification because presumably the geospatial linkages would be in one of the Darfur states. The Fur had a total of 14 *ST occurrences;* however, only three were able to be linked geospatially. The inability to link the vast majority of the *ST occurrences* may have decreased the levels of geospatial agreement (increased average *ST/SME distance*) exhibited by the Fur ethnic group.

Distant Outliers Illuminate Displacement

The Shilluk had one *ST occurrence* that was a noticeable outliers connected to Darfur, which is 873 km from the *SME location* (Figure 11).

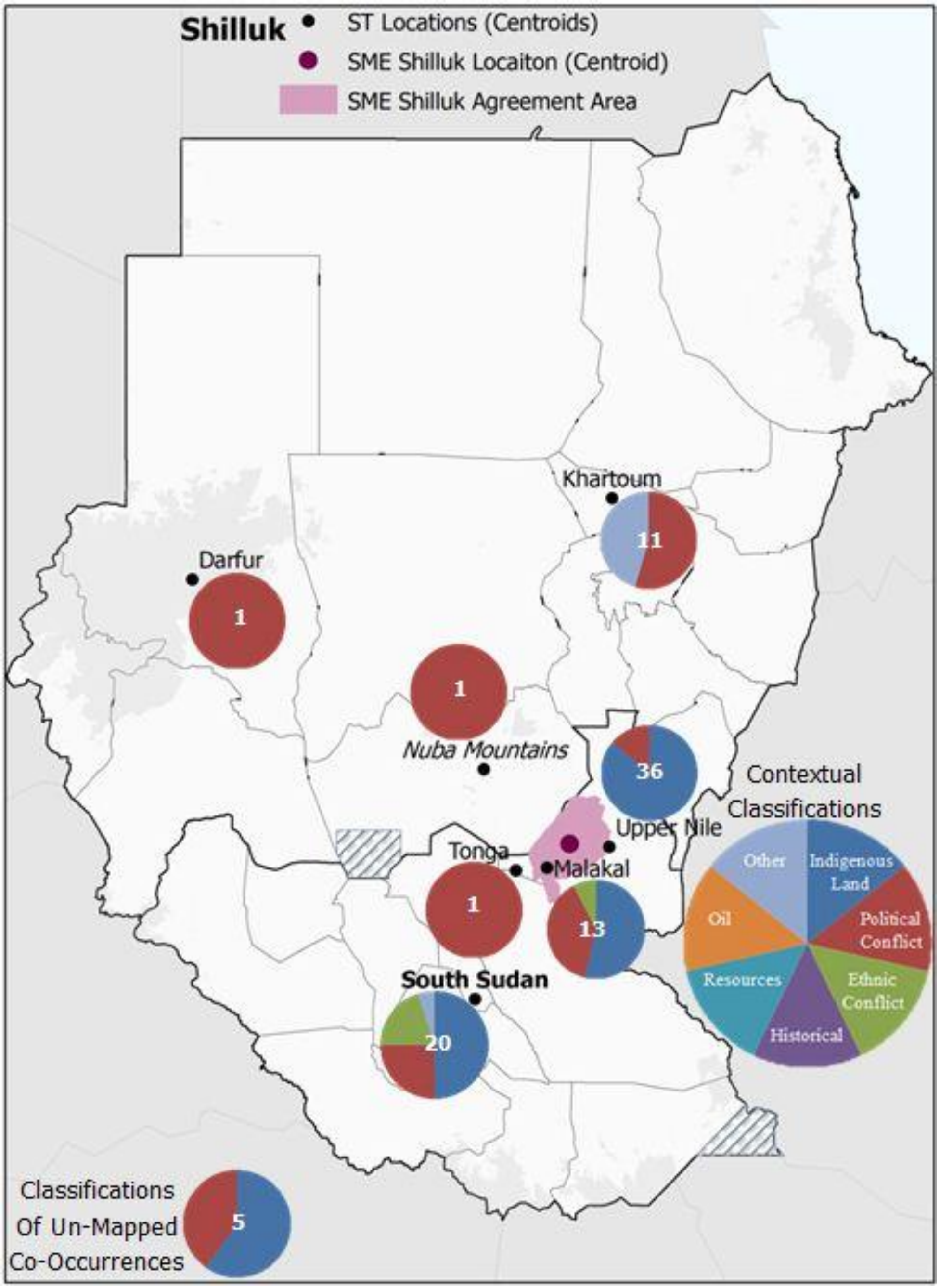


Figure 11: Geospatial Agreement between the Sudan Tribune and SME Perceptions of the Shilluk Ethnic Group and Contextual Classification by Location

Upon examining the original article that produced the link, it was clear that this connection was due to perceived similar experiences. This is not because the Shilluk inhabit Darfur, but rather they are referred to simultaneously because the people of Darfur and the Shilluk people share in similar conflicts. For example the article read, "…while Khartoum continues its campaigns of human destruction in **Darfur** and in the **Shilluk** Kingdom of Upper Nile Province in southern Sudan…(Reeves 2004)" Knowing this helps illuminate the common struggles faced by ethnic groups during domestic conflict.

It is necessary to code international locations even when modeling individual countries or ethnic groups, such as Sudan and/or the Beja. The Beja are the only ethnic group connected internationally by *ST occurrences* (Figure 9). The co-occurrences between the Beja and countries such as Eritrea illustrate the displacement of portions of the Beja population fleeing conflict (Dahl 1991; Young 2011). Thousands of refugees and internally displaced persons (IDPs) have been generated by the war in Eastern Sudan, where the Beja population has traditionally lived (Young 2011). The Beja have fled to the northern side of the Eritrean border and into the northwest region of Ethiopia, recently due to conflict (Young 2011) but historically due to drought induced famine (Dahl 1991; Cutler 1991). The co-occurrences that connected Eritrea and Ethiopia to the Beja were contextually classified in the content analysis as resources, since there was not a class specifically for displacement (Figure 9). This type of international displacement may have played a part in making the average *ST/SME distance* exhibited by the Beja larger, because the centroid of the entire country is used as the geospatial link (Figure 9).

Nomadic/ Sedentary Tradition v. Reality

The Shilluk who are traditionally farmers and fishermen live a sedentary lifestyle (Lobban 2012). The Shilluk had a longer average distance than the Ngok, a seasonally nomadic ethnic group (Lobban 2012) (Table 4). However, this distribution is based on news articles from 2004-2008, so the analyzed ethnic group's traditional ways of life and historical homeland may be less represented than the reality of their current lifestyle and location. A study conducted by

Assal (2006) from the University of Khartoum showed that the largest IDP camp in Khartoum (the capital city with the largest number of IDPs) was inhabited by a variety of ethnic groups including the Dinka (Ngok), which made up a quarter of the entire camp (25.4%), the Fur (13.1%), the Shilluk (4.1%), and the Bari (4%). This shows that the traditional lands and ways of life for some Sudanese ethnic groups may not characterize their recent reality.

The Ngok had a much smaller range (51-449 km) of *ST/SME distances* than the Beja (72-2,091 km) (Table 4), who are traditionally nomadic (Lobban 2012). However, the constant conflict that spread throughout Sudan and South Sudan since their united independence in 1955 has shifted the reality of many ethnic groups' lifestyles (Lobban 2012). The Ngok in the years leading up to the Comprehensive Peace Agreement in 2005, concerning the disputed area of Abyei, were internally displaced to the north (Cohen 2008). If words specifically associated with displacement were coded and a semi-automated network text analysis searching for those types of co-occurrences with ethnic groups had been conducted, displacement of ethnic groups could be modeled.

Lessons Learned & Future Research Opportunities

A few modifications are recommended for future research. By using the centroid of the *SME location* instead of the edge of the polygon to measure the Euclidean distance between *ST locations* and *SME locations*, this study may have overestimated the *ST/SME distance*. In the future it may be wise to use the edge of agreement polygon itself instead of its centroid.

In this study we were unable to observe possible temporal variation because all five years of data were synthesized into one analysis. This type of data modeling could be done in the future using single years to show if/how the co-occurrences shift geospatially on an annual basis.

This research used the Sudan Tribune, an online, national, English language newspaper. The Sudan Tribune was chosen because it provided consistent national coverage; however, the large audience could have influenced some reporting bias towards larger, more politically established ethnic groups. In the future, utilizing more localized news sources may prove to better characterize the location of ethnic groups, particularly ethic groups with smaller populations or ethnic groups that live in underrepresented regions. The greatest challenges faced when attempting to use local news sources are the language barrier and the access to paper archives if websites do not exist.

During the contextual classification of *ST occurrences*, a few of them had contexts referring specifically to displacement. Since displacement was not one of the classification options, we classified them as either ethnic conflict or resources, whichever seemed most appropriate. In future analysis of Sudan related data, a class for displacement should be included; in hindsight this may have better illuminated the *ST occurrences* classified as ethnic conflict or resource.

To avoid the time consuming task of manually classifying the context of each co-occurrence it may be helpful to use automated techniques to code for words that are associated with relevant research themes (Van Holt et al. 2013). For example, the *ST occurrences* classified as indigenous land could be further analyzed for common words associated specifically with indigenous land, such as native, home, or local. By including these words and testing to see if they identify context correctly, future semi-automated network text analysis could be structured to pick up on specific types of textual connections.

CONCLUSION

Location analysis that integrates multiple SMEs knowledge and network text analysis of independent online data bring us another step closer to being able to visualize and verify the dynamic geospatial flow of ethnic groups'. However, there is still more work that need to be done before this method can be confidently used to map the locations of ethnic groups. By evaluating the distance between the *SME locations* and *ST locations* and using that to characterize their agreement, it was determined that only a few of the *ST occurrences* overlapped with the SME area of agreement. Through the content analysis and contextual classification, it was discovered that there is greater agreement between *ST occurrences* classified as indigenous land than any other contextual classification. Also, according to the ANOVAs the livelihood classification (sedentary or nomadic) of an ethnic group does not help to explain the variation in *ST/SME distances*. With refinement and further research, this method of location analysis could be a reasonable and powerful tool for researchers and decision makers to conceptualize and document the dispersion of ethnic groups in regions and states where politics and populations are volatile.

CHAPTER 4

CONCLUSION

Participatory mapping and PGIS methodologies were utilized to geospatially represent the dynamic distribution of 26 selected Sudanese ethnic groups, as perceived by SMEs. To account for the lack of current, on the ground, geospatially explicit data regarding Sudanese ethnic groups, Sudan subject-matter-experts (SMEs) were interviewed and their perceptions were recorded using scaled and georeferenced maps. To analyze the geospatial agreement between expert's perceptions, a Geospatial Similarity Analysis and geospatial consensus analysis were developed and applied.

The CCM has been widely applied in the field of anthropology to assess shared knowledge (Romney, Weller and Batchelder 1986), but never before has it been used on geospatially explicit data. By applying the CCM in this manner, this research has developed a method to systematically measure geospatial consensus among multiple informants, expanding the applicability of participatory-mapping methods. Maps created using this method have the potential to be powerful avenues of communication between researchers and decision makers because of its ability to incorporate, visualize, and gain meaning from multiple viewpoints. However, there is more work that needs to be done in order to refine the manner in which the geospatially explicit data is collected. In order to better understand thresholds of agreement between respondents, soliciting a greater number of people from varying levels of expertise may be helpful. The presumably increased number of overlapping responses would illuminate a threshold that defines how much overlap is possible before additional perspectives cease to improve the consensus.

Semi-automated network text analysis has the ability to create geospatially linked data from independent non-geospatial text data sources. To determine if the SMEs' aggregated perceptions of ethnic group locations agree geospatially with an independent data source, articles from the Sudan Tribune online news source were examined for co-occurrences within the text of locations and ethnic groups. The co-occurrences were identified using a geospatially linked semi-automated network text analysis and a manual content analysis. The distances between the location indicated by multiple SMEs and locations cited in the Sudan Tribune were determined in order to geospatially characterize agreement. To investigate the textual connection further, each co-occurrence was contextually classified.

Through the contextual classification of geospatially linked co-occurrences (identified by the semi-automated network text analysis and content analysis), it has been shown that certain types of context relate to locations differently than other types of context. For example, locations cited in the context of indigenous land related more closely to where Sudan SMEs perceived an ethnic group to be than contextual classifications such as political-conflict. Therefore, in future research using semi-automated network text analysis, additional coding for contextually indicative terms could help illuminate findings that would otherwise be lost, due to the variety of contexts within which co-occurrences can be cited. There is still a need for a closer examination of contextually revealing terms and the words that commonly occur in various contexts need to be identified. Once the contextually indicative terms are identified, this type of data modeling (using semi-automated network analysis) could use multiple individual years to show how the distribution of specified populations shifts geospatially on an annual basis, possibly depicting the displacement patterns of specific ethnic groups.

In conclusion, this research developed and applied a method for utilizing the CCM to determine whether or not geospatial overlap of responses is a product of shared knowledge. This research also found that the context of geospatially linked co-occurrence can be helpful in examining, explaining, and visualizing certain research themes geospatially. This type of locational analysis, that integrates SME knowledge and independent data sources, advances the ability of researchers and decision makers to model the dynamic distribution on ethnic group populations across varying temporal and spatial scales.

REFERENCES

Assal, M. (2006). Whose Rights Count? National and International Responses to the Rights of IDPs in the Sudan. Brighton, United Kingdom: Development Research Centre on Migration, Globalization and Poverty - University of Sussex.

Ayers, A. (2010). Sudan's uncivil war: the global-historical constitution of political violence. *Review of African Political Economy, 37*(124), 153-171. doi: 10.1080/03056244.2010.483888

Bemigisha, J., Carranza, J., Skidmore, A. K., McCall, M., Polce, C., & Prins, H. H. T. (2009). Representation of Uncertainty and Integration of PGIS-based Grazing Intensity Maps Using Evidential Belief Functions. *Transactions in GIS, 13*(3), 273-293. doi: 10.1111/j.1467-9671.2009.01162.x

Bernard, E., Barbosa, L., & Carvalho, R. (2011). Participatory GIS in a sustainable use reserve in Brazilian Amazonia: Implications for management and conservation. *Applied Geography, 31*(2), 564-572. doi: 10.1016/j.apgeog.2010.11.014

Bojorquez-Tapia, L. A., Brower, L. P., Castilleja, G., Sanchez-Colon, S., Hernandez, M., Calvert, W., . . . Juarez, M. D. (2003). Mapping expert knowledge: Redesigning the Monarch Butterfly Biosphere Reserve. *Conservation Biology, 17*(2), 367-379. doi: 10.1046/j.1523-1739.2003.01309.x

Borgatti, S., Everett, M., & Freeman, L. (2002). UCINET for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technology.

Boster, J. S., & Johnson, J. C. (1989). Form or Function - A Comparison Of Expert And Novice Judgments of Similarity Among Fish. *American Anthropologist, 91*(4), 866-889. doi: 10.1525/aa.1989.91.4.02a00040

Bremer, S. (1992). Dangerous Dyads - Conditions Affecting the Likelihood of Interstate War, 1816-1965. *Journal of Conflict Resolution, 36*(2), 309-341. doi: 10.1177/0022002792036002005

Broekel, T., & Boschma, R. (2012). Knowledge networks in the Dutch aviation industry: the proximity paradox. *Journal of Economic Geography, 12*(2), 409-433. doi: 10.1093/jeg/lbr010

Buhaug, H., & Gates, S. (2002). The geography of civil war. *Journal of Peace Research, 39*(4), 417-433. doi: 10.1177/0022343302039004003

Buhaug, H., & Lujala, P. (2005). Accounting for scale: Measuring geography in quantitative studies of civil war. *Political Geography, 24*(4), 399-418. doi: 10.1016/j.polgeo.2005.01.006

Buhaug, H., & Rod, J. (2006). Local determinants of African civil wars, 1970-2001. *Political Geography, 25*(3), 315-335. doi: 10.1016/j.polgeo.2006.02.005

Carley, K. (1997). Network Text Analysis: The Network Position of Concepts. In C. Roberts (Ed.), *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences From Texts and Transcripts* (pp. 79-100). Hillsdale: Lawrence Erlbaum Associates.

Carley, K., Bigrigg, M., & Diallo, B. (2012). Data-to-model: a mixed initiative approach for rapid ethnographic assessment. *Computational and Mathematical Organization Theory, 18*(3), 300-327. doi: 10.1007/s10588-012-9125-y

Carley, K., Columbus D., Bigrigg M., & Kunkel F. (2011a). AutoMap User's Guide: Carnegie Mellon University, School of Computer Science, Institute for Software Research, CMU-ISR-11-107R.

Carley, K., Reminga, J., Storrick, J., & Columbus, D. (2011b). ORA User's Guide: Carnegie Mellon University, School of Computer Science, Institute for Software Research, CMU-ISR-11-107R.

Chapin, M., Lamb, Z., & Threlkeld, B. (2005). Mapping Indigenous Lands. *Annual Review of Anthropology, 34*, 619-638. doi: 10.1146/annurev.anthro.34.081804.120429

Cohen, R. (2008). Listening to the Voices of the Displaced: Lessons Learned. Washington, DC: Brookings-Bern Project on Internal Displacement.

Curtin, K. (2007). Network Analysis in Geographic Information Science: Review, Assessment and Projections (Vol. 34, pp. 103-111): Cartography and Geographic Information Society.

Cutler, P. (1991). The Political-Economy of Famine In Ethiopia And Sudan. *Ambio, 20*(5), 176-178.

Dahl, G. (1991). The Beja of Sudan and The Famine Of 1984-1986. *Ambio, 20*(5), 189-191.

Diesner, J., & Carley, K. (2005). Revealing Social Structure From Texts: Meta-Matrix Text Analysis as a Novel Method for Network Text Analysis In V. Narayanan & D. Armstrong (Eds.), *Causal Mapping for Information Systems and Technology Research: Approaches, Advances, and Illustrations* (Vol. 4, pp. 81-108). Harrisburg: Idea Group Publishing.

ESRI. (2011). ArcGIS Desktop: Release 10. Redlands: Environmental Systems Research Institute.

FACTBOX. (28 May 2004). Key terms of Pacts to End Sudan's 21-Year War, *Sudan Tribune*. Roubaix - France: OVH.

FAO South Sudan (2004). Hydrology. United Nations Institute for Training and Research & United Nations Sudan Information Gateway.

Fearon, J. D. (2004). Why do some civil wars last so much longer than others? *Journal of Peace Research, 41*(3), 275-301. doi: 10.1177/0022343304043770

Forkel, R. (2011). World Atlas of Language Structures Online. 2013, from http://wals.info/

Gilmore, E., Gleditsch, N. P., Lujala, P., & Rod, J. K. (2005). Conflict diamonds: A new dataset. *Conflict Management and Peace Science, 22*(3), 257-272. doi: 10.1080/07388940500201003

Giuliani, E. (2007). The selective nature of knowledge networks in clusters: evidence from the wine industry. *Journal of Economic Geography, 7*(2), 139-168. doi: 10.1093/jeg/lbl014

Gleditsch, K. (2007). Transnational dimensions of civil war. *Journal of Peace Research, 44*(3), 293-309. doi: 10.1177/0022343307076637

Herlihy, P., & Knapp, G. (2003). Maps of, by, and for the peoples of Latin America. *Human Organization, 62*(4), 303-314.

Homerdixon, T. F. (1994). ENVIRONMENTAL SCARCITIES AND VIOLENT CONFLICT - EVIDENCE FROM CASES. *International Security, 19*(1), 5-40. doi: 10.2307/2539147

IRIN. (23 July 2004). Heavy Fighting in Southern Sudan between Nuer and Shilluk, *Sudan Tribune Online (45)*.

Izady, M. Ethnic Groups in Former (United) Sudan. Retrieved 9-28-2012, from http://gulf2000.columbia.edu/

Johnson, D. H. (2008). Why abyei matters - The breaking point of Sudan's comprehensive peace agreement? *African Affairs, 107*(426), 1-19. doi: 10.1093/afraf/adm070

King, M., & Osman, M. (2004). *ENVIRONMENTAL DEGRADATION AS A CAUSE OF CONFLICT IN DARFUR.* Paper presented at the Environmental Degradation and the Conflict in Darfur, University of Khartoum.

KM, C., D, C., MW, B., & F, K. (2011). AutoMap User’s Guide: Carnegie Mellon University, School of Computer Science, Institute for Software Research.

Kroeber, A. (1948). *Anthropology*. New York: Harcourt, Brace.

Krueger, R., & Casey, M. A. (2009). *Focus Groups: A Practical Guild for Applied Research* (4th ed.). Thousand Oaks, California: SAGE Publications, Inc.

Krygier, J., & Wood, D. (2005). *Making Maps*. New York: The Guilford Press.

Lehner, B., Verdin, K., & Jarvis, A. (2006). HydroSHEDS.

Lewis, P., Simons, G., & Fennig, C. (2013). Ethnologue Languages of the World. from http://www.ethnologue.com

Lobban, R., & Fluehr-Lobban, C. (2012). [Conflict Period Email].

Lobban, R., Kramer, R., & Fluehr-Lobban, C. (2002). *Historical Dictionary of the Sudan* (Third ed.). Lanham, Maryland: The Scarecrow Press, Inc.

Luo, W., Hartmann, J., Liu, J. X., & Huang, P. W. (2007). Geographic patterns of Zhuang (Tai) kinship terms in Guangxi and border areas: a GIS analysis of language and culture change. *Social & Cultural Geography, 8*(4), 575-596. doi: 10.1080/14649360701529840

Lupai, J. (27 November 2007). Jihad for Abyei Oil is Disintegration of Sudan, *Sudan Tribune Online*.

M., R. J. (1964). The Self-Management of Cultures. In W. Goodenough (Ed.), *Explorations in Cultural Anthropology* (pp. 433-454). New York: McGraw-Hill.

Maggioni, M. A., & Uberti, T. E. (2011). Networks and geography in the economics of knowledge flows. *Quality & Quantity, 45*(5), 1031-1051. doi: 10.1007/s11135-011-9488-z

MAOZ, Z., & ABDOLALI, N. (1989). REGIME TYPES AND INTERNATIONAL CONFLICT, 1816-1976. *Journal of Conflict Resolution, 33*(1), 3-35. doi: 10.1177/0022002789033001001

Mehta, C., & Patel, N. (2007). StatXact 8 User Manual. Cambridge, MA: CYTEL Software Corporation.

Murdoch, G. P. (1959). *Africa: Its Peoples and their Culture History*. New York: McGraw-Hill Book Company.

NGIA. (2011). National Geospatial Intelligence Agency Sudan Country File.

nickname), M. a. (2007). Distribution of African Languages Retrieved 9-28-2012, from http://www.muturzikin.com/en.html

Odiong, L. (Dec 18 2008). The SPLA is an Army of Occupation and Land Grabbing, *Sudan Tribune*.

Openshaw, S. (1991). A View on The GIS Crisis In Geography, Or, Using GIS to Put Humpty-Dumpty Back Together Again. *Environment and Planning A, 23*(5), 621-628.

Patey, L. A. (2007). State rules: oil companies and armed conflict in Sudan. *Third World Quarterly, 28*(5), 997-1016. doi: 10.1080/01436590701371728

Pfeffer, J., & Carley, K. M. (2012). Rapid modeling and analyzing networks extracted from pre-structured news articles. *Computational and Mathematical Organization Theory, 18*(3), 280-299. doi: 10.1007/s10588-012-9122-1

Popping, R. (2000). *Computer-Assisted Text Analysis*. Thousand Oaks: Sage.

UNDP (United Nation Development Program). 2010. Roads, States, Settlements. United Nations Institute for Training and Research & UN Sudan Information Management Working Group (IMWG).

Raghavan, S. (21 September 2004). Policies, Economics Turned Muslims Against Muslims in Sudan's Darfur, *Sudan Tribune Online*.

Rambaldi, G., Kwaku Kyem, P., McCall, M., & Weiner, D. (2006). Participatory Spatial Information Management and Communication in Developing Countries. *The Electronic Journal of Information Systems in Developing Countries, 25*(1), 1-9.

Reeves, E. (April 6, 2004). Khartoum Continues Campaigns of Human Destruction in Southern Sudan and Darfur, *Sudan Tribune Online*.

Reeves, E. (Feb 14, 2006). As the International Community Postures, Sudan s Many Crises Deepen: Darfur, Chad, Eastern Sudan, Southern Sudan---All Remain Victims of National Islamic Front Tyranny and Brutality, *Sudan Tribune Online*.

Reeves, E. (June 28, 2006). The Regime has a Long and Deadly History of Such Suspensions, both in Southern Sudan and Currently in Eastern Sudan, *Sudan Tribune Online*.

Romney, A. K., Batchelder, W. H., & Weller, S. C. (1987). Recent Applications of Cultural Consensus Theory. *American Behavioral Scientist, 31*(2), 163-177. doi: 10.1177/000276487031002003

Romney, A. K., Weller, S. C., & Batchelder, W. H. (1986). Culture as Consensus - A Theory Of Culture And Informant Accuracy. *American Anthropologist, 88*(2), 313-338. doi: 10.1525/aa.1986.88.2.02a00020

Ross, M. L. (2004a). How do natural resources influence civil war? Evidence from thirteen cases. *International Organization, 58*(1), 35-67. doi: 10.1017/s02081830458102x

Ross, M. L. (2004b). What do we know about natural resources and civil war? *Journal of Peace Research, 41*(3), 337-356. doi: 10.1177/0022343304043773

Sambanis, N. (2001). Do ethnic and nonethnic civil wars have the same causes? A theoretical and empirical inquiry (part 1). *Journal of Conflict Resolution, 45*(3), 259-282. doi: 10.1177/0022002701045003001

Sudan Tribune: Plural News and Views on Sudan. (2003-2012). *Sudan Tribune Online*.

Sudan Tribune (May 12, 2004). U.N. Sees New Humanitarian Crisis in Sudan's Upper Nile Region, *Sudan Tribune Online*.

Theisen, O. M. (2008). Blood and Soil? Resource Scarcity and Internal Armed Conflict Revisited. *Journal of Peace Research*, *45*(6), 801-818. doi: 10.1177/0022343308096157

Themner, L., & Wallensteen, P. (2012). Armed Conflicts, 1946-2011. *Journal of Peace Research, 49*(4), 565-575. doi: 10.1177/0022343312452421

Van Holt, T., Johnson, J., Brinkley, J., Carley, K., & Caspersen, J. (2012). Structure of ethnic violence in Sudan: a semi-automated network analysis of online news (2003-2010). *Computational and Mathematical Organization Theory*, *18*(3), 340-355. doi: 10.1007/s10588-012-9124-z

Van Holt, T., Johnson, J., Carley, K., Brinkley, J., & Diesner, J. (2013). *Rapid Ethnographic Assessment for Cultural Mapping*. To be published in: Journal of Poetics.

Weller, S. C. (1987). Shared Knowledge, Intracultural Variation, and Knowledge Aggregation. *American Behavioral Scientist, 31*(2), 178-193. doi: 10.1177/000276487031002004

Weller, S. C. (2007). Cultural consensus theory: Applications and frequently asked questions. *Field Methods, 19*(4), 339-368. doi: 10.1177/1525822xo7303502

Weller, S. C., & Romney, A. K. (1988). Structured Interviewing. Newbury Park, California: Sage.

Young, J. (2011). Beja: Local Conflict, Marginalization, and the Threat to Regional Security. Pretoria, South Africa: Institute for Security Studies.

APPENDIX A: IRB NOTIFICATION OF EXEMPT CERTIFICATION

EAST CAROLINA UNIVERSITY
University & Medical Center Institutional Review Board Office
4N-70 Brody Medical Sciences Building· Mail Stop 682
600 Moye Boulevard · Greenville, NC 27834
Office **252-744-2914** · Fax **252-744-2284** · **www.ecu.edu/irb**

Notification of Exempt Certification

From: Social/Behavioral IRB
To: Janna Caspersen
CC:
Tracy Van Holt
Date: 5/11/2012
Re: UMCIRB 12-000807
Participatory Understanding of Sudanese Ethnic Groups

I am pleased to inform you that your research submission has been certified as exempt on 5/11/2012. This study is eligible for Exempt Certification under category #2.

It is your responsibility to ensure that this research is conducted in the manner reported in your application and/or protocol, as well as being consistent with the ethical principles of the Belmont Report and your profession.

This research study does not require any additional interaction with the UMCIRB unless there are proposed changes to this study. Any change, prior to implementing that change, must be submitted to the UMCIRB for review and approval. The UMCIRB will determine if the change impacts the eligibility of the research for exempt status. If more substantive review is required, you will be notified within five business days.

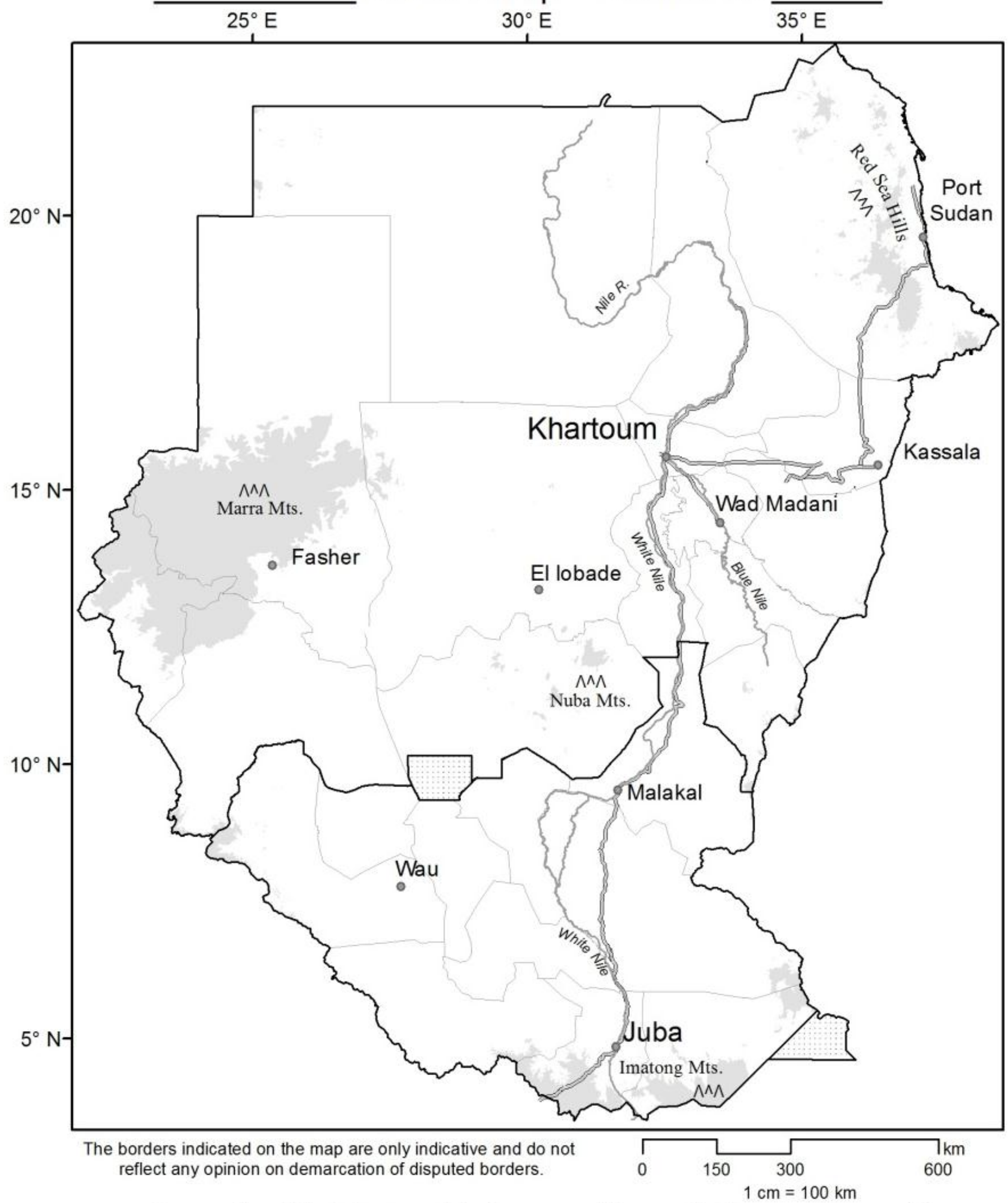
The UMCIRB office will hold your exemption application for a period of five years from the date of this letter. If you wish to continue this protocol beyond this period, you will need to submit an Exemption Certification request at least 30 days before the end of the five year period.

The Chairperson (or designee) does not have a potential for conflict of interest on this study.

APPENDIX B: RESPONSE MAP

(USED TO RECORD SUBJECT-MATTER-EXPERT PERCEPTIONS)

____________ Ethnic Group Informant # ______



On a scale of 1 to 5, how confident are you with your ethnic group placement?

1	2	3	4	5
Not Confident		**Neutral**		**Very Confident**

APPENDIX C: REFERENCE MAP (AVAILABLE TO RESPONDENTS DURING THE GEOSPATIAL DATA COLLECTION)

APPENDIX D: LIST OF ETHNIC GROUPS INCLUDED IN THE INITAL DATA COLLECTION

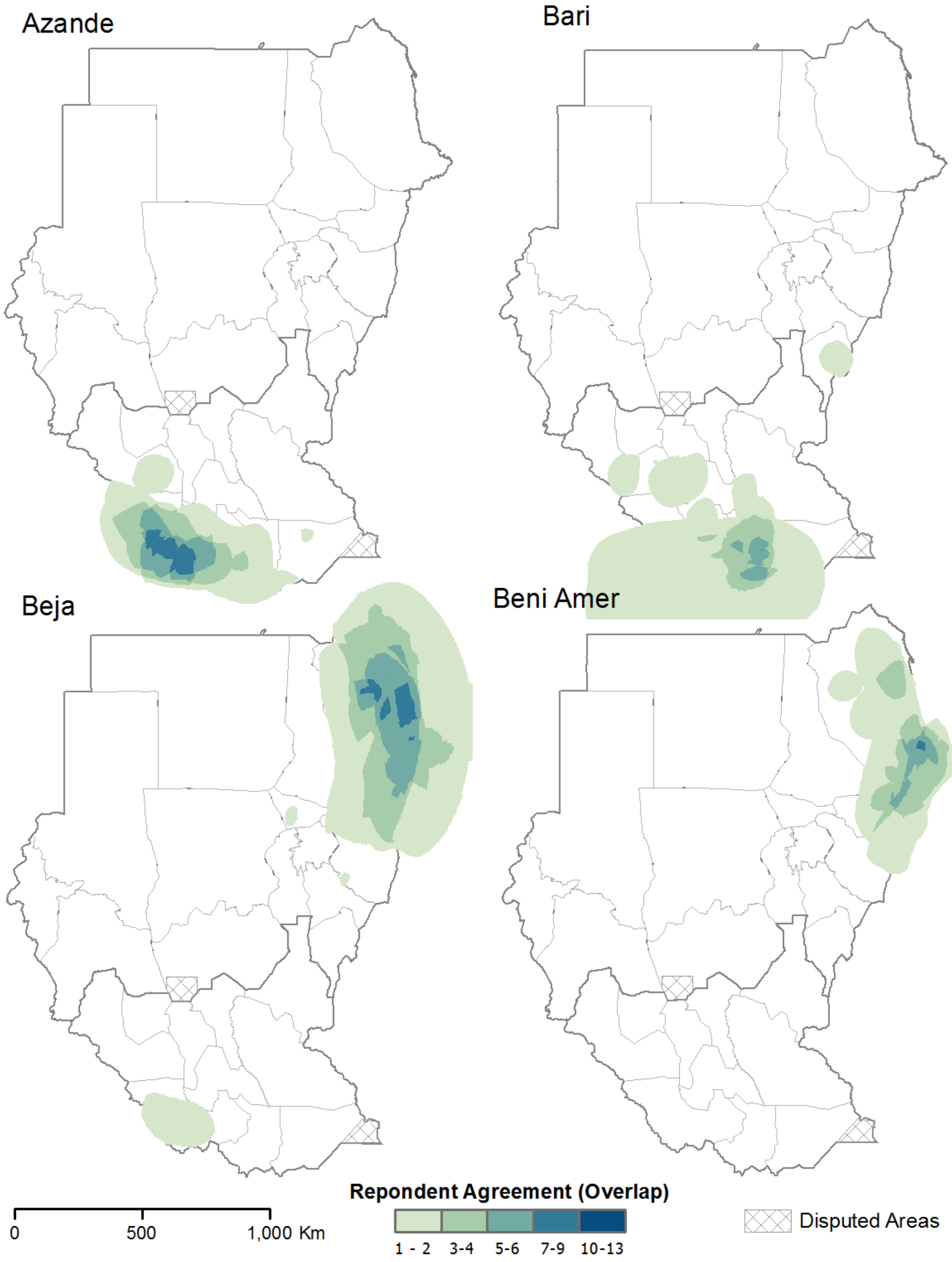
Ethnic Groups Used in Expert Mapping Prescribed by experts Dr. Lobban & Dr. Fluehr Lobban
Azande
Bari
Beja
Beni Amer
Bor
Danagla
East Jikany
Hadendowa
Humr
Ja Aliyin
Kakwa
Lou
Madi
Malwal
Masalit
Messirya
Murle
Ngok
Rizaygat
Shaygiya
Shilluk
Taisha
Talodi
West Jikany
Zaghawa
Fur *Added at the conference by research team, due to lack of representation*

APPENDIX E: DEMOGRAPHIC INFORMATION COLLECTED FROM RESPONDENTS

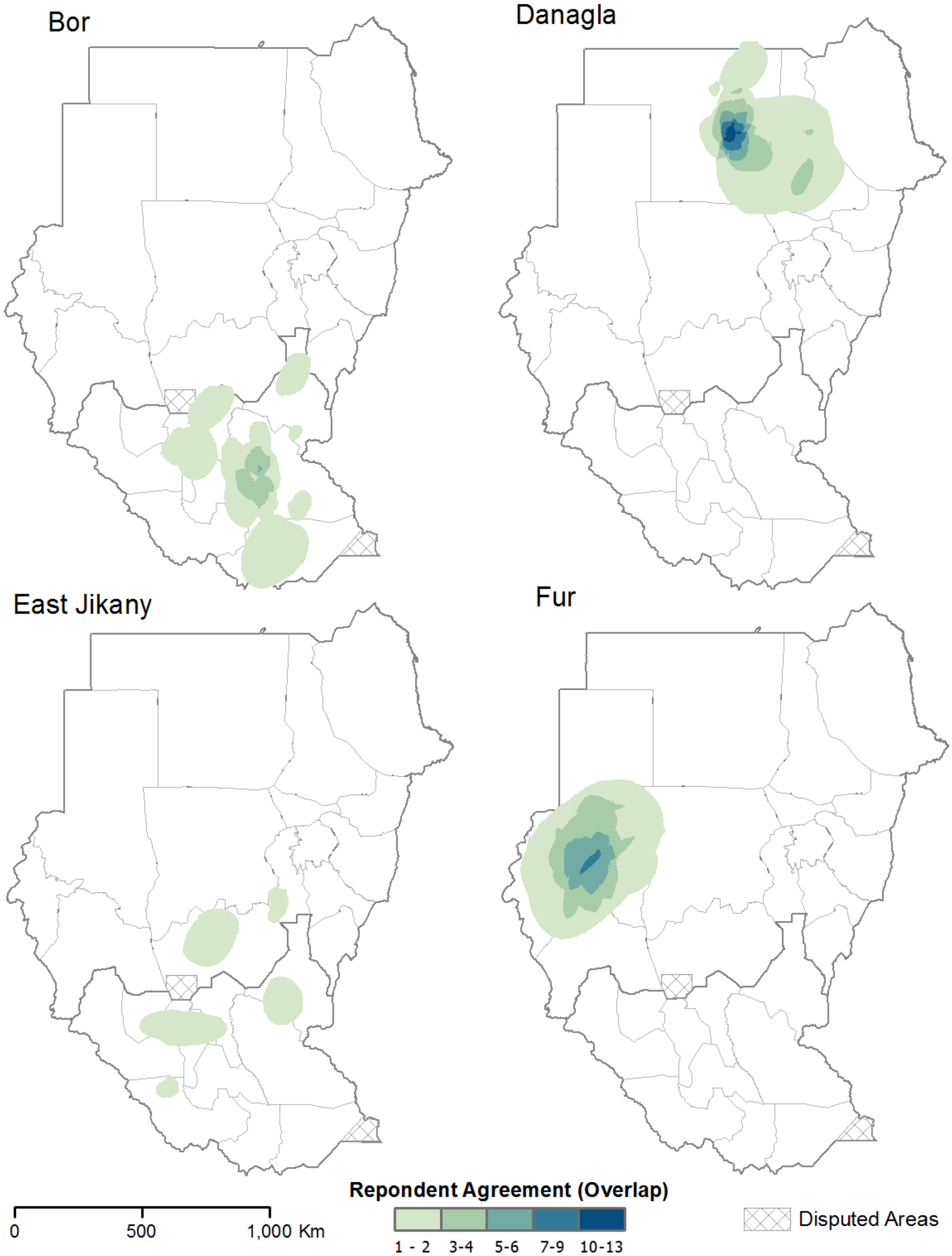
(OMITTED FOR PRIVACY: CURRENT RESIDENCE AND BIRTH PLACE)

Expert Interview Demographic Data								
ID#	Age	Sex	Ethnic Identities			Languages	Edu.	Employment
			Group	Branch	Family			
1	61	M	KUKU			KUKU (BARI), ENGLISH	PHD	PROFESSOR
2	32	F	WHITE, AMERICAN			ENGLISH, SWAHIL	MA	CIVIL-MILITARY
3	67	M	USA			ENGLISH, DUTCH, SUDANESE, ARABIC	PHD	PROFESSOR
4	73	M	ARAKIIN	HAUIMAB	BABIKERS	ARABIC, ENGLISH, POLISH	MA	RETIRED
6	47	M	KUKU			KUKU, ENGLISH, ACHLOI, ARABIC, SWAHLI	PHD	
7	55	M	PENSYLVANIA, GERMAN			ENGLISH, SUDANESE ARABIC, BARI, DANGALA, KINYARWANDA	PHD	ASSOCIATE PROFESSOR
8	60	F	SUDANESE			ARABIC, ENGLISH, POLISH	PHD	SELF EMPLOYED
16	63	M	SUDANESE	GALI	ABDALLAH	ARABIC, ENGLISH, HEUSA	PHD	ASSOCIATE PROFESSOR
17	75	M	HAMAR			ARABIC	PHD	RETIRED
18	37	M	AWLAD HAMID			ARABIC	MS	ENGINEER
19	67	F	UKRAINE/ GERMAN			ENGLISH, ARABIC, SPANISH	PHD	PROFESSOR
20		M				ARABIC, ENGLISH	MS	SELF EMPLOYED
21	57	M	SUDANESE/ JAALIYIN	RUBATAB		ARABIC, ENGLISH	PHD	SEMI-CONDUCTOR
23	45	M	AFRICAN AMERICAN	BARTTI		ARABIC, ENGLISH, HAWSA	BA	ARTIST
24	68	M	BRITISH			ENGLISH, FRENCH, SOME ARABIC	PHD	RETIRED
26	65	M	BARI			BARI	PHD	TEACHER
27	68	M	AMERICAN			ARABIC, ENGLISH, FRENCH	PHD	PROFESSOR, NAVAL OFFICER

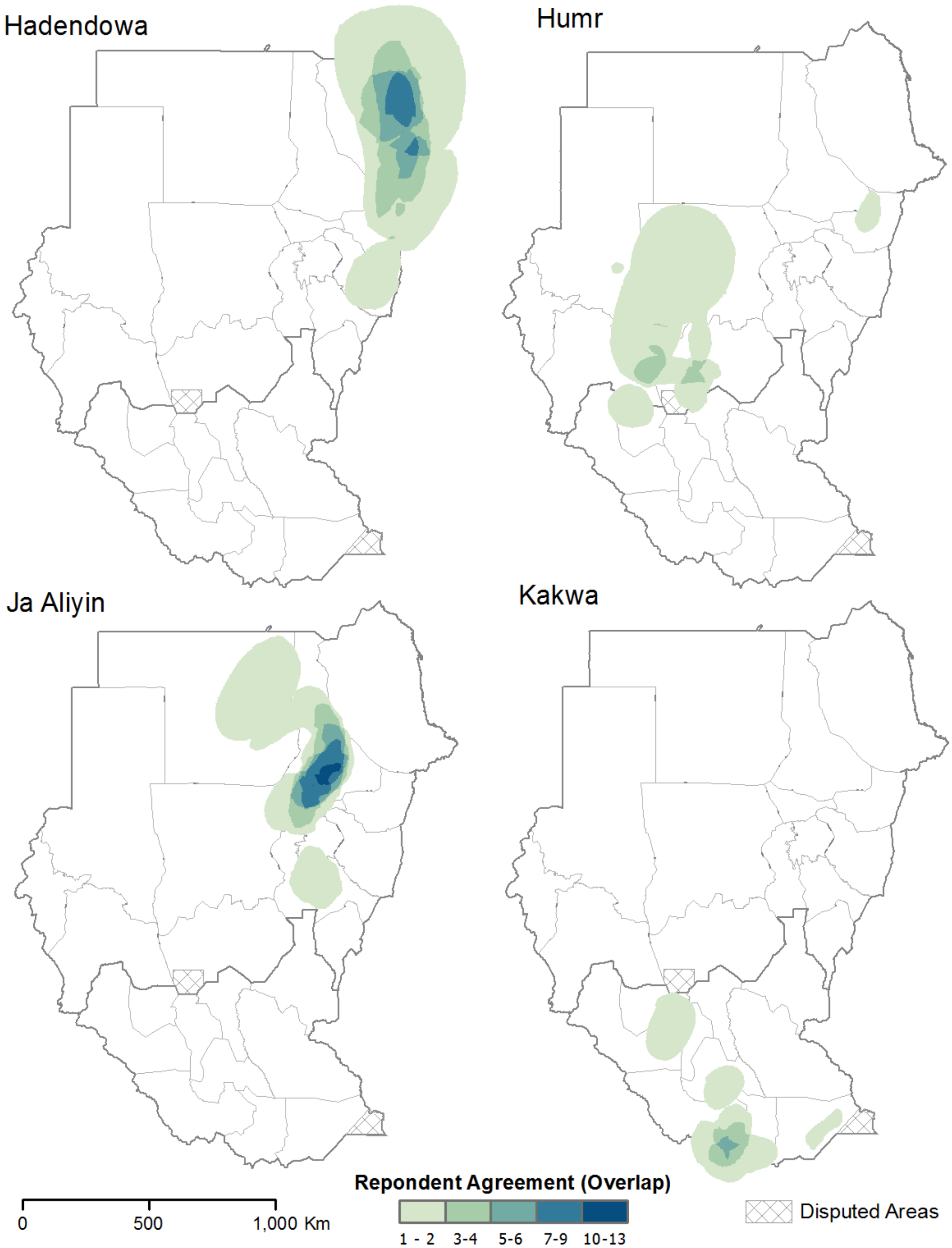
APPENDIX F: EARL MAPS FOR ALL THE ETHNIC GROUPS INITIALLY INVESTIGATED



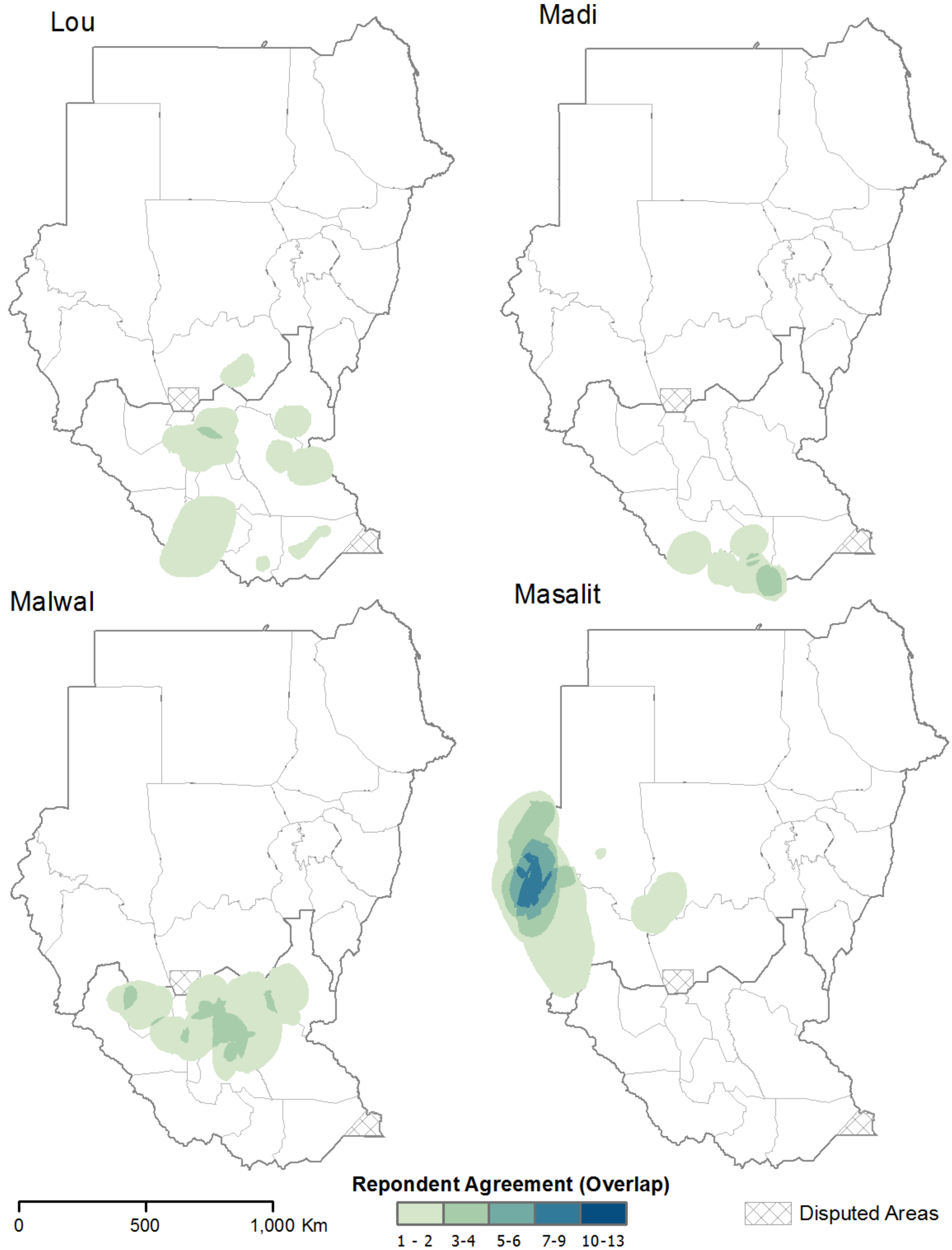
APPENDIX F (CONTINUED): EARL MAPS FOR ALL THE ETHNIC GROUPS INITIALLY INVESTIGATED



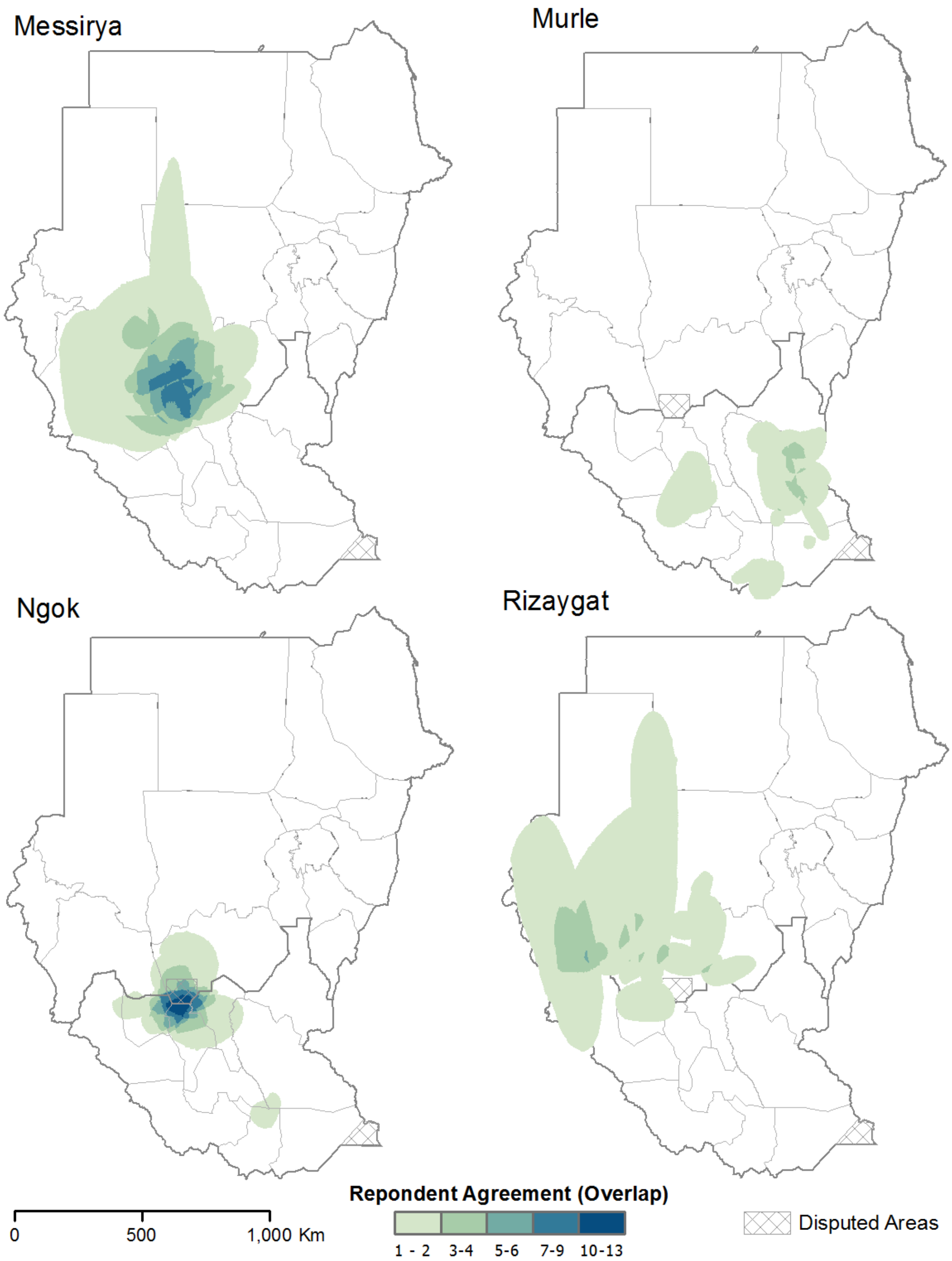
APPENDIX F (CONTINUED): EARL MAPS FOR ALL THE ETHNIC GROUPS INITIALLY INVESTIGATED



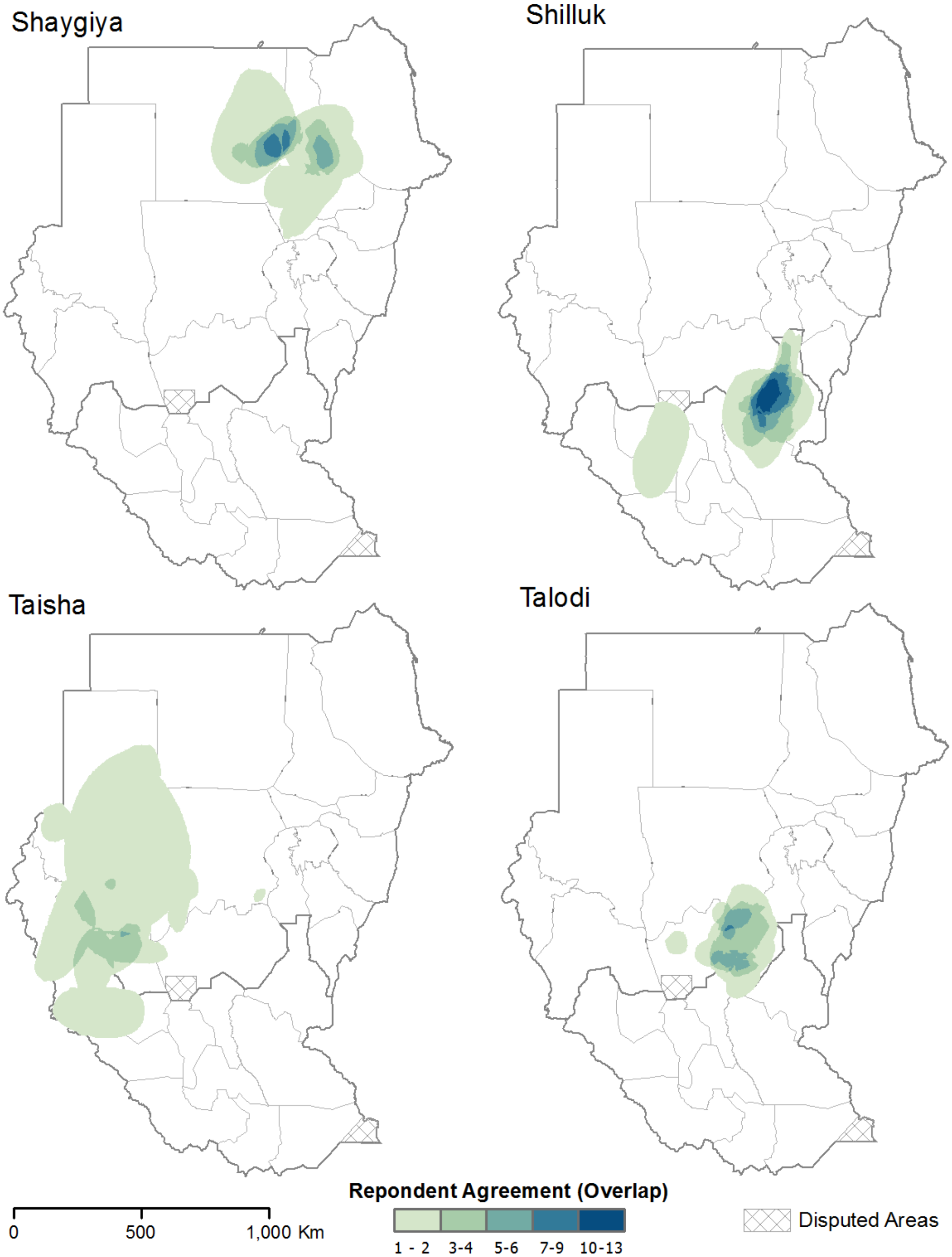
APPENDIX F (CONTINUED): EARL MAPS FOR ALL THE ETHNIC GROUPS INITIALLY INVESTIGATED



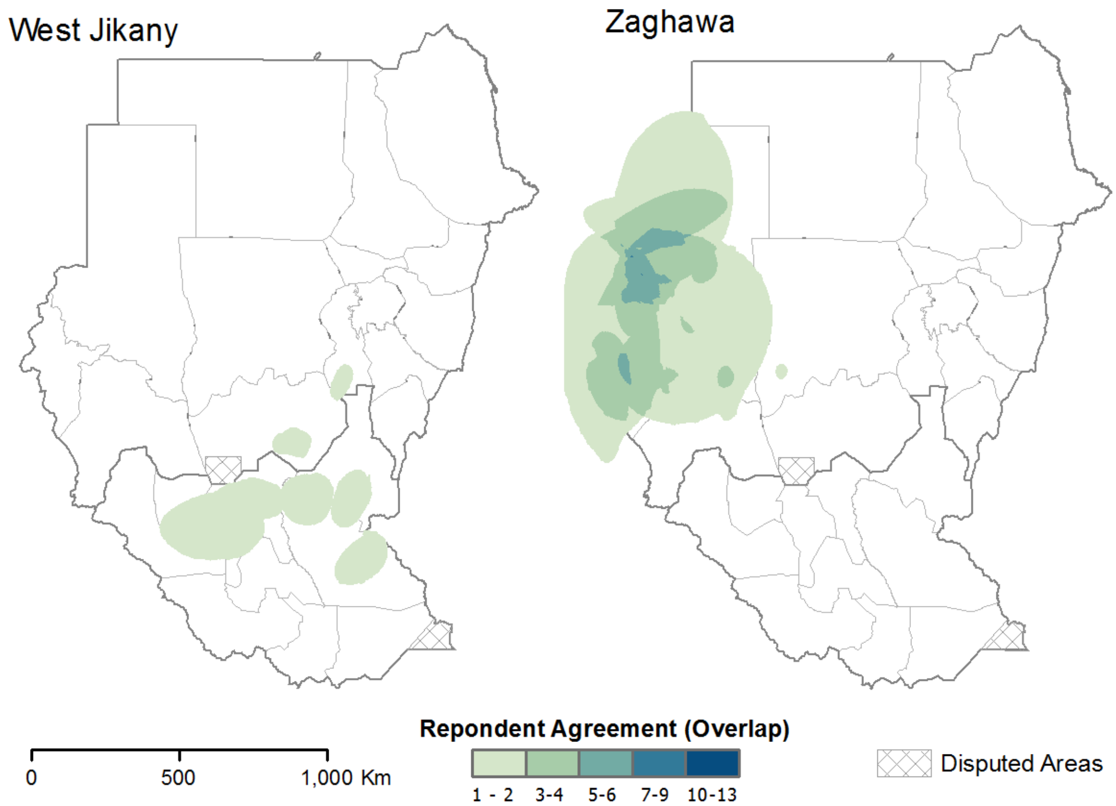
APPENDIX F (CONTINUED): EARL MAPS FOR ALL THE ETHNIC GROUPS INITIALLY INVESTIGATED



APPENDIX F (CONTINUED): EARL MAPS FOR ALL THE ETHNIC GROUPS INITIALLY INVESTIGATED



APPENDIX F (CONTINUED): EARL MAPS FOR ALL THE ETHNIC GROUPS INITIALLY INVESTIGATED



APPENDIX G: RESPONDENT AREA OVERLAP WITH EACH AGREEMENT THRESHOLDS FOR THE DANAGLA (KILOMETERS)

Danagla Respondents	3 ≥ Overlap (KM)	5 ≥ Overlap (KM)	7 ≥ Overlap (KM)	10 ≥ Overlap (KM)
1	27.5	15.4	8.8	2.2
3	17.0	15.6	8.6	2.2
4	20.3	14.5	8.6	2.2
5				
6	25.1	9.5	4.9	1.2
7	8.5	8.4	7.8	2.2
16	0.6	0.0	0.0	0.0
17	3.4	3.4	3.4	2.0
18				
19	0.6	0.0	0.0	0.0
20	0.0	0.0	0.0	0.0
21	28.3	12.8	7.7	2.2
23	41.9	14.9	7.6	2.2
24	14.4	8.2	4.0	1.3
25	14.1	12.8	8.3	2.2
26	7.5	0.0	0.0	0.0
27	23.1	13.0	7.9	2.2

APPENDIX H: RESPONDENT PERCENT AREA OVERLAP WITH EACH AGREEMENT THRESHOLDS FOR THE DANAGLA (KILOMETERS)

Danagla Respondents	% Area Shared 3 ≥ Overlap	% Area Shared 5 ≥ Overlap	% Area Shared 7 ≥ Overlap	% Area Shared 10 ≥ Overlap	Average %
1	51%	83%	100%	100%	84%
3	32%	84%	97%	100%	78%
4	38%	79%	97%	100%	78%
5					
6	47%	52%	55%	54%	52%
7	16%	46%	88%	100%	62%
16	1%	0%	0%	0%	0%
17	6%	19%	38%	93%	39%
18					
19	1%	0%	0%	0%	0%
20	0%	0%	0%	0%	0%
21	53%	70%	87%	100%	77%
23	78%	81%	86%	100%	86%
24	27%	44%	45%	61%	44%
25	26%	69%	93%	100%	72%
26	14%	0%	0%	0%	3%
27	43%	70%	89%	100%	76%

APPENDIX I: RESPONDENT AREA OVERLAP WITH EACH AGREEMENT THRESHOLDS FOR THE SHILLUK (KILOMETERS)

Shilluk Respondents	3 ≥ Overlap (KM)	5 ≥ Overlap (KM)	7 ≥ Overlap (KM)	10 ≥ Overlap (KM)
1	21.9	16.1	13.3	7.4
3	25.0	16.1	11.8	7.1
4	19.5	14.7	8.8	3.8
5	32.7	24.1	17.1	7.4
6	27.8	21.1	17.5	7.4
7	12.6	9.5	8.5	6.4
16	19.9	15.3	12.5	6.9
17	10.5	4.2	1.8	0.1
18	2.7	2.4	1.6	0.0
19	23.9	19.1	15.0	7.1
20				
21	18.5	6.3	1.9	0.1
23	50.3	27.3	18.8	7.4
24	25.4	23.7	17.9	7.4
25	34.3	19.4	12.8	5.8
26	7.1	4.7	3.9	2.3
27	17.1	17.0	16.1	7.4

APPENDIX J: RESPONDENT PERCENT AREA OVERLAP WITH EACH AGREEMENT THRESHOLDS FOR THE SHILLUK (KILOMETERS)

Shilluk Respondents	% Area Shared 3 ≥ Overlap	% Area Shared 5 ≥ Overlap	% Area Shared 7 ≥ Overlap	% Area Shared 10 ≥ Overlap	Average %
1	**39%**	**55%**	**70%**	**99%**	**66%**
3	**45%**	55%	62%	96%	64%
4	**35%**	51%	47%	50%	46%
5	**58%**	83%	91%	100%	83%
6	**49%**	72%	93%	100%	79%
7	**22%**	33%	45%	86%	47%
16	**35%**	53%	66%	93%	62%
17	**19%**	14%	9%	1%	11%
18	**5%**	8%	9%	0%	5%
19	**42%**	65%	79%	95%	71%
20					
21	**33%**	22%	10%	1%	16%
23	**89%**	94%	99%	100%	96%
24	**45%**	81%	95%	100%	80%
25	**61%**	67%	68%	78%	68%
26	**13%**	16%	21%	30%	20%
27	**30%**	58%	85%	100%	69%

APPENDIX K: RESPONDENT AREA OVERLAP WITH EACH AGREEMENT THRESHOLDS FOR THE JA ALIYIN (KILOMETERS)

Ja Aliyin Respondents	3 ≥ Overlap (KM)	5 ≥ Overlap (KM)	7 ≥ Overlap (KM)	10 ≥ Overlap (KM)
1	31.1	28.6	22.0	4.4
3	38.9	37.4	25.4	4.4
4	16.9	15.9	11.5	1.2
5				
6	18.3	9.9	6.2	0.0
7	55.4	39.4	25.6	4.4
16	18.1	18.1	14.7	4.2
17	12.4	12.4	11.8	3.8
18				
19	26.5	23.8	20.1	4.4
20				
21	56.1	36.3	24.8	4.4
23	45.4	34.3	24.4	4.4
24	19.4	17.6	14.9	4.4
25	52.5	28.6	25.6	4.4
26	10.6	8.7	4.7	0.7
27	12.8	12.8	12.7	4.4

APPENDIX L: RESPONDENT PERCENT AREA OVERLAP WITH EACH AGREEMENT THRESHOLDS FOR THE JA ALIYIN (KILOMETERS)

Ja Aliyin Respondents	% Area Shared 3 ≥ Overlap	% Area Shared 5 ≥ Overlap	% Area Shared 7 ≥ Overlap	% Area Shared 10 ≥ Overlap	Average %
1	50%	72%	86%	100%	77%
3	63%	95%	99%	100%	89%
4	27%	40%	45%	28%	35%
5					
6	30%	25%	24%	0%	20%
7	90%	100%	100%	100%	97%
16	29%	46%	57%	95%	57%
17	20%	32%	46%	87%	46%
18					
19	43%	60%	79%	100%	70%
20					
21	91%	92%	97%	100%	95%
23	73%	87%	95%	100%	89%
24	31%	45%	58%	99%	58%
25	85%	73%	100%	100%	89%
26	17%	22%	18%	16%	18%
27	21%	32%	50%	100%	51%

APPENDIX M: RESPONDENT AREA OVERLAP WITH EACH AGREEMENT THRESHOLDS FOR THE NGOK (KILOMETERS)

Ngok Respondents	3 ≥ Overlap (KM)	5 ≥ Overlap (KM)	7 ≥ Overlap (KM)	10 ≥ Overlap (KM)
1	15.1	12.6	10.0	5.3
3	15.7	9.8	7.6	3.8
4				
5	20.4	11.0	8.5	4.8
6	11.4	8.3	4.7	2.3
7	0.3	0.0	0.0	0.0
16	19.1	14.7	11.3	5.0
17	3.1	3.1	2.5	1.1
18	2.1	1.9	0.3	0.0
19	16.5	13.0	8.7	4.1
20	28.9	18.0	13.4	5.3
21	19.2	17.0	13.0	5.3
23	27.4	17.6	11.7	5.3
24	25.5	15.6	10.2	5.2
25	26.1	15.3	11.8	5.3
26	5.4	4.2	2.3	0.8
27	22.2	16.6	11.2	5.3

APPENDIX N: RESPONDENT PERCENT AREA OVERLAP WITH EACH AGREEMENT THRESHOLDS FOR THE NGOK (KILOMETERS)

Ngok Respondents	% Area Shared $3 \geq$ Overlap	% Area Shared $5 \geq$ Overlap	% Area Shared $7 \geq$ Overlap	% Area Shared $10 \geq$ Overlap	Respondent Average %
1	33%	56%	73%	99%	65%
3	34%	43%	56%	71%	51%
4					
5	45%	49%	63%	91%	62%
6	25%	37%	35%	42%	35%
7	1%	0%	0%	0%	0%
16	42%	65%	83%	93%	71%
17	7%	13%	18%	21%	15%
18	5%	9%	2%	0%	4%
19	36%	57%	64%	77%	59%
20	63%	79%	99%	100%	85%
21	42%	75%	96%	100%	78%
23	60%	78%	86%	99%	81%
24	56%	69%	76%	98%	74%
25	57%	68%	87%	100%	78%
26	12%	18%	17%	15%	15%
27	49%	73%	82%	100%	76%

APPENDIX O: DANAGLA RESPONDENT-BY-RESPONDENT PERCENT AREA OVERLAP

Informant	1	3	4	5	6	7	16	17	18	19	20	21	23	24	25	26	27
1	100%	37%	57%		21%	24%	0%	10%		3%	0%	32%	52%	48%	31%	0%	68%
3	76%	100%	69%		58%	48%	0%	20%		0%	0%	76%	79%	40%	70%	0%	63%
4	69%	41%	100%		24%	28%	0%	12%		0%	0%	32%	47%	32%	36%	0%	65%
5																	
6	20%	29%	20%		100%	13%	0%	3%		0%	0%	69%	99%	1%	27%	0%	12%
7	99%	97%	95%		53%	100%	0%	41%		0%	0%	80%	79%	46%	94%	0%	89%
16	0%	0%	0%		0%	0%	100%	0%		0%	6%	1%	100%	0%	0%	0%	0%
17	100%	99%	100%		32%	100%	0%	100%		0%	0%	79%	69%	61%	100%	0%	100%
18																	
19	3%	0%	0%		0%	0%	0%	0%		100%	0%	0%	0%	11%	0%	0%	0%
20	0%	0%	0%		0%	0%	1%	0%		0%	100%	0%	76%	0%	0%	13%	0%
21	37%	43%	30%		80%	22%	1%	9%		0%	0%	100%	99%	11%	38%	0%	25%
23	10%	8%	8%		20%	4%	6%	1%		0%	27%	17%	100%	3%	7%	5%	7%
24	81%	34%	45%		2%	19%	0%	10%		15%	0%	16%	28%	100%	18%	0%	69%
25	75%	85%	73%		68%	56%	0%	24%		0%	0%	81%	84%	26%	100%	0%	60%
26	0%	0%	0%		0%	0%	0%	0%		0%	100%	0%	100%	0%	0%	100%	0%
27	95%	44%	75%		16%	30%	0%	14%		0%	0%	31%	49%	57%	34%	0%	100%

APPENDIX P: JA ALIYIA RESPONDENT-BY-RESPONDENT PERCENT AREA OVERLAP

Informant	1	3	4	5	6	7	16	17	18	19	20	21	23	24	25	26	27
1	100%	86%	50%		30%	99%	39%	28%		77%		84%	79%	56%	94%	15%	36%
3	67%	100%	37%		23%	95%	43%	31%		55%		89%	89%	43%	92%	24%	31%
4	90%	87%	100%		63%	97%	16%	4%		53%		84%	63%	60%	92%	0%	8%
5																	
6	24%	23%	28%		100%	58%	0%	0%		8%		44%	13%	16%	30%	0%	0%
7	30%	37%	16%		22%	100%	17%	12%		26%		48%	38%	19%	45%	10%	12%
16	68%	96%	15%		0%	100%	100%	55%		71%		100%	100%	30%	100%	24%	61%
17	71%	100%	6%		0%	100%	81%	100%		77%		100%	100%	42%	100%	38%	80%
18																	
19	85%	78%	32%		12%	92%	45%	33%		100%		80%	86%	50%	86%	22%	41%
20																	
21	35%	47%	19%		23%	66%	24%	16%		30%		100%	71%	20%	67%	11%	17%
23	30%	43%	13%		6%	47%	22%	15%		30%		66%	100%	22%	52%	13%	15%
24	87%	86%	51%		31%	100%	27%	26%		71%		74%	89%	100%	89%	19%	30%
25	51%	64%	28%		20%	80%	31%	21%		43%		88%	74%	31%	100%	14%	22%
26	44%	89%	0%		0%	96%	40%	43%		58%		80%	99%	35%	74%	100%	28%
27	88%	98%	11%		0%	100%	87%	78%		93%		100%	100%	48%	100%	24%	100%

APPENDIX Q: NJOK RESPONDENT-BY-RESPONDENT PERCENT AREA OVERLAP

Informant	1	3	4	5	6	7	16	17	18	19	20	21	23	24	25	26	27
1	100%	38%		84%	17%	0%	67%	8%	14%	42%	100%	67%	69%	48%	100%	35%	54%
3	32%	100%		22%	10%	3%	37%	16%	0%	21%	47%	52%	100%	87%	37%	5%	50%
4																	
5	44%	14%		100%	8%	0%	30%	2%	7%	23%	85%	30%	28%	18%	72%	19%	29%
6	15%	11%		14%	100%	0%	40%	11%	0%	80%	30%	39%	37%	55%	20%	0%	68%
7	0%	1%		0%	0%	100%	0%	0%	0%	0%	0%	0%	38%	4%	0%	0%	0%
16	31%	21%		27%	20%	0%	100%	2%	1%	29%	44%	48%	44%	36%	41%	6%	39%
17	38%	92%		23%	58%	0%	25%	100%	0%	74%	78%	66%	100%	100%	46%	0%	100%
18	100%	0%		100%	0%	0%	15%	0%	100%	0%	100%	9%	35%	0%	100%	97%	0%
19	34%	20%		35%	68%	0%	49%	12%	0%	100%	49%	56%	53%	63%	41%	9%	82%
20	24%	13%		39%	8%	0%	22%	4%	3%	14%	100%	23%	26%	19%	46%	8%	20%
21	53%	48%		45%	33%	0%	81%	10%	1%	55%	77%	100%	74%	69%	66%	12%	78%
23	18%	30%		14%	10%	27%	24%	5%	1%	17%	28%	24%	100%	51%	23%	5%	25%
24	22%	47%		16%	27%	5%	35%	9%	0%	36%	36%	40%	91%	100%	28%	2%	48%
25	51%	23%		70%	11%	0%	44%	5%	7%	26%	99%	43%	45%	31%	100%	18%	35%
26	98%	18%		100%	0%	0%	45%	0%	37%	30%	100%	43%	53%	14%	100%	100%	30%
27	35%	39%		36%	48%	0%	54%	13%	0%	67%	54%	64%	63%	69%	45%	7%	100%

APPENDIX R: SHILLUK RESPONDENT-BY-RESPONDENT PERCENT AREA OVERLAP

Informant	1	3	4	5	6	7	16	17	18	19	20	21	23	24	25	26	27
1	100%	68%	36%	51%	60%	42%	70%	3%	0%	58%		7%	70%	66%	39%	29%	51%
3	54%	100%	29%	59%	37%	22%	54%	7%	0%	37%		34%	69%	53%	53%	9%	33%
4	41%	41%	100%	70%	38%	27%	19%	31%	0%	23%		50%	100%	52%	82%	39%	31%
5	34%	51%	42%	100%	54%	25%	31%	12%	7%	43%		31%	100%	56%	77%	12%	43%
6	50%	39%	28%	65%	100%	40%	52%	1%	9%	79%		2%	96%	66%	46%	14%	60%
7	71%	47%	41%	62%	81%	100%	51%	0%	0%	58%		2%	100%	54%	48%	35%	57%
16	65%	63%	16%	42%	59%	28%	100%	0%	2%	66%		0%	55%	65%	27%	7%	49%
17	7%	19%	60%	40%	4%	0%	0%	100%	0%	0%		100%	100%	24%	97%	5%	1%
18	2%	0%	0%	95%	100%	0%	23%	0%	100%	100%		0%	100%	62%	40%	0%	65%
19	51%	42%	18%	56%	85%	31%	64%	0%	10%	100%		0%	84%	68%	36%	8%	58%
20																	
21	3%	16%	17%	17%	1%	1%	0%	18%	0%	0%		100%	49%	6%	48%	4%	0%
23	18%	23%	23%	38%	30%	15%	15%	12%	3%	25%		33%	100%	24%	43%	15%	19%
24	61%	62%	43%	77%	74%	30%	65%	10%	6%	71%		14%	87%	100%	60%	13%	64%
25	20%	34%	37%	58%	29%	15%	15%	23%	2%	20%		64%	85%	33%	100%	11%	23%
26	39%	15%	47%	25%	24%	28%	10%	3%	0%	12%		15%	78%	19%	28%	100%	19%
27	70%	58%	38%	88%	100%	47%	74%	1%	10%	90%		1%	99%	94%	62%	19%	100%