A LOGISTIC REGRESSION AND MARKOV CHAIN MODEL FOR THE PREDICTION OF NATION-STATE VIOLENT CONFLICTS AND TRANSITIONS

THESIS

Nicholas J. Shallcross, MAJ, USA

AFIT-ENS-MS-16-M-128

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-MS-16-M-128

A LOGISTIC REGRESSION AND MARKOV MODEL FOR THE PREDICTION OF NATION-STATE VIOLENT CONFLICTS AND TRANSITIONS

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

Nicholas J. Shallcross, BS

MAJ, USA

March 2016

AFIT-ENS-MS-16-M-128

A LOGISTIC REGRESSION AND MARKOV CHAIN MODEL FOR THE PREDICTION OF NATION-STATE VIOLENT CONFLICTS AND TRANSITIONS

THESIS

Nicholas J. Shallcross, BS

MAJ, USA

Committee Membership:

Dr. D. K. Ahner
Chair

Lieutenant Colonel Brian J. Lunday, Ph.D.
Member

AFIT-ENS-MS-16-M-128

Abstract

The prediction and forecasting of violent conflict, is of vital importance to formulate coherent national strategies effecting regional and worldwide stability and security. Using open source data, this research formulates and constructs a suite of statistical models that predict future transitions into and out of violent conflict and forecasts the regional and global incidences of violent conflict over a ten-year time horizon. A total of thirty predictor variables are tested and evaluated for inclusion in twelve conditional logistic regression models, which calculate the probability that a nation will transition from its current conflict state, either "In Conflict" or "Not in Conflict", to a new state in the following year. These probabilities are then used to construct a series of nation-specific Markov chain models that forecast violent conflict, as well as yield insights into regional conflict trends out to year 2024 and beyond. The logistic regression models proposed in this study achieve training dataset accuracies of 88.76%, and validation dataset accuracies of 84.67%. Additionally, the Markov models achieve three year forecast accuracies of 85.16% during model validation. Given the current state of included predictor variables, this study predicts that global violent conflict rates remain constant through year 2024, but are projected to increase beyond that timeframe with 95 of the 182 considered nations projected to be in a state of violent conflict from the current 84 nations in conflict.

KEYWORDS: Conflict Transitions, Logistic Regression, Markov Models

This work is dedicated to my wife for her tireless devotion, support, and love over the course of my career.

Acknowledgments

I would like to express my sincere appreciation to my advisor, Dr. Darryl Ahner, for his direction and support throughout the course of this research. I also wish to thank Lieutenant Colonel Brian Lunday, Dr. Richard Deckro, and Major Russell Walter for their insights and contributions.

Nicholas J. Shallcross

Table of Contents

Page

Abstract iv
Table of Contents vii
List of Figures ix
List of Tables xi
I. Introduction 1
1.1 General Issue 1
1.2 Problem Statement 2
1.3 Research Objective and Focus 3
1.4 Research Questions 3
1.5 Methodology 4
1.6 Study Assumptions and Limitations 5
1.7 Implications 6
1.8 Overview of Remaining Chapters 7
II. Literature Review 8
2.1 Overview 8
2.2 Nation-State Conflict Prediction: Relevant Research 8
2.3 Markov Models and the Prediction and Spread of Disease Epidemics 17
2.4 Relevant Variables 20
2.5 Summary 24
III. Methodology 26
3.1 Chapter Overview 26
3.2 Conditional Conflict Database Development 26
3.3 Logistic Regression 46

3.4 Markov Models 70

3.5 Summary..78

IV. Analysis and Results..80

4.1 Chapter Overview..80

4.2 Analysis of Region Specific Conditional Logistic Regression Models 80

4.3 Analysis of Significant Conflict Transition Variables 100

4.4 Analysis of Nation Specific Markov Models 115

4.5 Forecasting Global Conflict trends...130

V. Conclusions and Recommendations 145

5.1 Chapter Overview..145

5.2 Conclusions of Research 145

5.3 Significance of Research 150

5.4 Recommendations for Future Research...150

Appendix A: Regional Assignments of Nations..154

Appendix B: Region Specific Conditional Logistic Regression Models.......................155

Appendix C: Markov Model Outputs 158

Appendix D: Regional Conflict Forecasts for 2016, 2019, and 2024............................164

Appendix E: Transience Scores by Region 171

Appendix F: Story Board 178

Bibliography 179

List of Figures

Page

Figure 1: National Level Violent Conflict in 2014 2

Figure 2: Study Methodology 4

Figure 3: In Conflict Database 45

Figure 4: Plots of the Logit π(x) and Logit Transformation g(x) 47

Figure 5: Hypothesis Test for Model Significance 52

Figure 6: Hypothesis Test for Covariate Significance 53

Figure 7: Hosmer-Lemeshow Hypothesis Test 58

Figure 8: Receiver Operating Characteristic Curve 60

Figure 9: Relative Risk Relation to Classification Table 61

Figure 10: JMP Whole Model Test 63

Figure 11: JMP Parameter Estimates 64

Figure 12: SMOTE Pseudo-code 66

Figure 13: Nation Specific Conflict Transition Markov Model 71

Figure 14: Conflict Transition Probability Markov Tool 77

Figure 15: Graph of Training and Validation ROC Curves. 84

Figure 16: Analysis of Cut-Point Effects on Classification Accuracy and False-Negative Rates 94

Figure 17: Covariate Correlation to Dependent Variable 101

Figure 18: Arab & North African States (In Conflict) Covariate Effects 102

Figure 19: Arab & North African States (Not in Conflict) Covariate Effects 104

Figure 20: Eastern Europe & Central Asia (In Conflict) Covariate Effects 105

Figure 21: Eastern Europe & Central Asia (Not in Conflict) Covariate Effects 106

Figure 22: Latin America (In Conflict) Covariate Effects 107

Figure 23: Latin America (Not in Conflict) Covariate Effects 108

Figure 24: OECD (In Conflict) Covariate Effects 109

Figure 25: OECD (Not in Conflict) Covariate Effects 110

Figure 26: South & East Asia (In Conflict) Covariate Effects 111

Figure 27: South & East Asia (Not in Conflict) Covariate Effects................................ 112

Figure 28: Sub-Saharan Africa (In Conflict) Covariate Effects 113

Figure 29: Sub-Saharan Africa (Not in Conflict) Covariate Effects.............................. 114

Figure 30: Average HIIK Conflict Intensity Levels by Bin 120

Figure 31: Identification of Significant Outliers by HIIK Bin 121

Figure 32: Average HIIK Conflict Intensity Levels with Outliers Removed 123

Figure 33: Example of First and Second Sojourn Times 125

Figure 34: Example of Significantly Long Sojourn Times 126

Figure 35: Long Run Conflict Probabilities... 127

Figure 36: Example of Mean Recurrence Times 129

Figure 37: Transience Score Histogram 144

List of Tables

Page

Table 1: Transition Probability Matrix: Conflict at *t* vs. *t-1*, 1970-2009 13

Table 2: Region Specific Relevant Variables 22

Table 3: Conflict Items 29

Table 4: HIIK Proxy Measures 30

Table 5: Mapping of Conditional Dependent Variable.. 31

Table 6: Independent Variables 35

Table 7: Border Conflict Score Example 36

Table 8: Mapping of Regime Type 37

Table 9: Number of Variables per Nation-year Instance; Worst Data............................. 43

Table 10: Variables Requiring Data Imputation 44

Table 11: Purposeful selection of Covariates Methodology 54

Table 12: Classification Table 59

Table 13: Discrimination Measures 61

Table 14: Summary Statistics of Regional Model Data... 67

Table 15: Nodal Model Training and Validation Year Sets 68

Table 16: Sub-Saharan, Given Non-Conflict Logistic Regression Model 81

Table 17: Sub-Saharan Africa "Not in Conflict" Conditional Model Comparison 82

Table 18: AUC Values by Region and Model 85

Table 19: Overall Classification Accuracies Given Fixed Cut-point of 0.50 87

Table 20: Comparison of Model Accuracies with the Boekestein Model 89

Table 21: Validation Data Set Accuracies below Accuracy Benchmark of 80%............. 90

Table 22: Model Rare Event Accuracies Given Fixed Cut-point of 0.50........................ 92

Table 23: Effects of Adjusted Cut-points on Model Accuracy and False Negative Rates 95

Table 24: Hosmer-Lemeshow Goodness of Fit Test Results... 96

Table 25: Hosmer-Lemeshow Test Significant Outliers 97

Table 26: Audit of Significant Outliers 99

Table 27: Overall Assessment of Conditional Logistic Regression Models 100

Table 28: Ranking of Variables in terms of Model Statistical Significance.................. 115

Table 29: Comparison of HIIK Observed and Expected Incidences of Conflict using a 0.50 Cut Point for 2014. 116

Table 30: Markov Model Validation Statistics 119

Table 31: HIIK Intensity Bin Assignments 120

Table 32: Significant Outliers, 2014 122

Table 33: Forecasting Assessment Matrix... 131

Table 34: Significant Changes in Conflict Probability Over 10 Year Period................ 132

Table 35: Summary of Conflict Forecasts: World View 133

Table 36: Arab & North African States Conflict Forecast Summary............................ 134

Table 37: Eastern Europe & Central Asia Conflict Forecast Summary 135

Table 38: Latin America Conflict Forecast Summary... 136

Table 39: OECD Conflict Forecast Summary 137

Table 40: South & East Asia Conflict Forecast Summary 138

Table 41: Sub-Saharan Africa Conflict Forecast Summary 139

Table 42: Top 25 Most Transient Nations... 142

Table 43: Top 25 Least Transient Nations... 143

Table 44: Regional Assignments of Nations 154
Table 45: Summary of Arab and North African State Models 155
Table 46: Summary of Eastern Europe & Central Asian Models 155
Table 47: Summary of Latina American Models 156
Table 48: Summary of OECD Models 156
Table 49: Summary of South & East Asian Models 156
Table 50: Summary of Sun-Saharan African Models 157
Table 51: Markov Model results by Nation 158
Table 52: Arab & North African States 2, 5, and 10 Year Forecasts 164
Table 53: Eastern Europe & Central Asia 2, 5, and 10 Year Forecasts 165
Table 54: Latin America 2, 5, and 10 Year Forecasts 166
Table 55: OECD 2, 5, and 10 Year Forecasts 167
Table 56: South & East Asia 2, 5, and 10 Year Forecasts 168
Table 57: Sub-Saharan Africa 2, 5, and 10 Year Forecasts 169
Table 58: Arab & North African States Transience Scores 171
Table 59: Eastern Europe & Central Asia Transience Scores 172
Table 60: Latin American Transience Scores 173
Table 61: OECD Transience Scores 174
Table 62: South & East Asia Transience Scores 175
Table 63: Sub-Saharan Africa Transience Scores 176

A LOGISTIC REGRESSION AND MARKOV MODEL FOR THE PREDICTION OF NATION-STATE VIOLENT CONFLICTS AND TRANSITIONS

I. Introduction

"It makes no difference what men think of war, said the judge. War endures. As well ask men what they think of stone. War was always here. Before man was, war waited for him. The ultimate trade awaiting its ultimate practitioner."

Cormac McCarthy, Blood Meridian

1.1 General Issue

Violent conflict between competing groups has been a pervasive and driving force for all of human history. It has evolved from small skirmishes between unarmed groups, wielding rudimentary weapons, to industrialized global conflagrations. Global incidences of violent conflict are at historically high levels, with 223 individual ongoing violent conflicts occurring throughout the globe, as shown in Figure 1 (Heidelberg Institute for International Conflict Research, 2014). While some of these conflicts are new, many have been ongoing for a decade or more, with no potential resolution in sight. Many recent studies have focused, with much success, on identifying the factors relevant for the accurate prediction armed conflict in nations. However, these studies have mainly focused on predicting conflict in the following year or two. While there is much to be gained from these analyses, a more operationally relevant question is: where and when will conflict transitions occur? A conflict transition is an event in which a nation transitions into or out of a state of violent conflict.

Conflict transitions by their very definition are rare events and. while some conflicts, are brought about by the unforeseen "Black Swan" events, many times there are overt but subtle indicators that a conflict is becoming more likely. Moreover, research in support of this study has identified a trend that, once a nation enters a certain conflict state, it tends to remain in such a state until some new event or events occur to disrupt this "conflict inertia". To answer the question concerning when and where conflict transitions will occur, this study develops a collection of conditional logistic regression and Markov chain models to predict when and where these conflict transitions are likely to occur and subsequently forecast global conflict incidences using open source data.

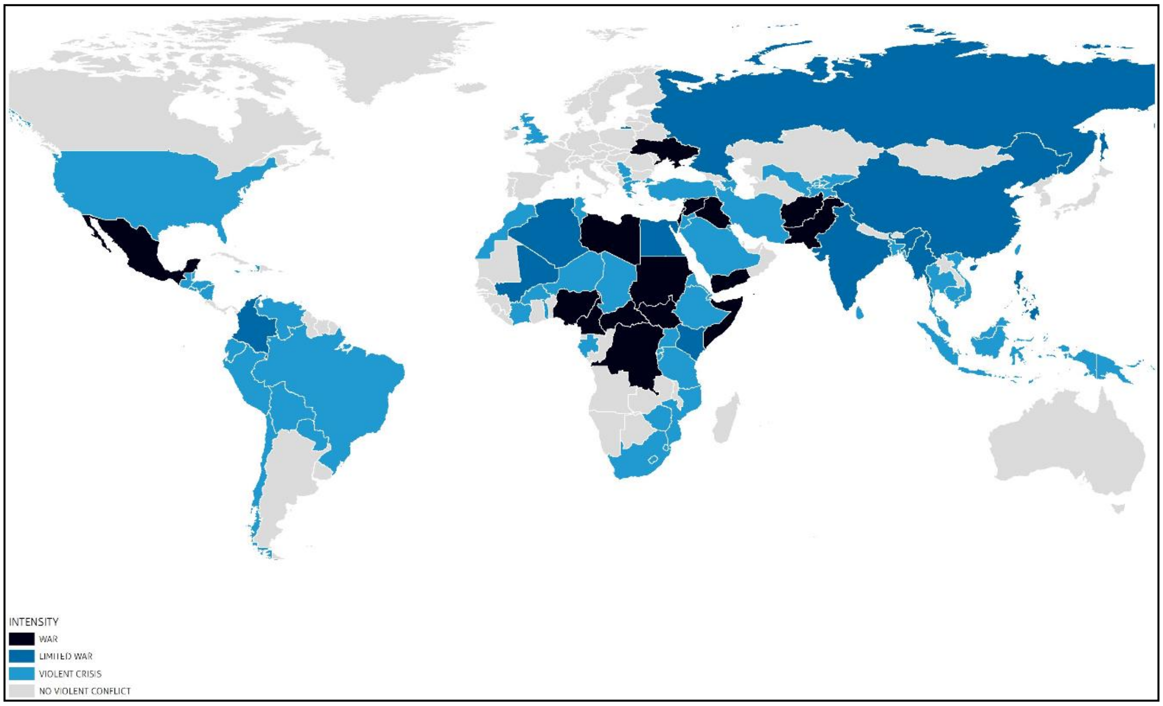


Figure 1: National Level Violent Conflict in 2014
(Heidelberg Institute for International Conflict Research, 2014)

1.2 Problem Statement

Use open source data to develop statistical models that predict and lend insight concerning when and where the world's nations transition into or out of violent conflict.

1.3 Research Objective and Focus

The objective of this study is to predict future worldwide and long-term conflict trends, national conflict transience indices, and to identify the exacerbating and/or enabling factors that lead to increased or decreased probabilities of conflict transitions. This study utilizes Markov modeling methods supported by conditional logistic regression models to predict transitions into and out of violent conflict, and the subsequent forecasting of global incidences of violent conflict. This study analyzed global conflict and its contributing factors for the years 2004 through 2014.

1.4 Research Questions

This study seeks to answer the five following research questions pertaining to the prediction of conflict transitions and forecasting of global conflict.

Question 1

How accurately can statistical models predict conflict transitions for individual nations?

Question 2

What factors are the significant predictors of conflict transitions?

Question 3

How is the number of global conflicts predicted to change by 2024 and beyond?

Question 4

What nations are susceptible to conflict transitions; which nations appear invulnerable to conflict transitions?

Question 5

Which nations, currently not in conflict, are identified as near-term risks for transitions into violent conflict?

1.5 Methodology

This study compiles and formats over 30 disparate databases into a single conditional conflict database (CCD). This effort is required for the development of twelve region specific conditional logistic regression models, one set for nations classified as "in conflict" and the other set for nations classified as "not in conflict". Using these conditional logistic regression models, we develop nation-specific Markov models to forecast conflict status transitions and future conflict trends for 182 of the world's nations. The complete study methodology is presented in Figure 2.

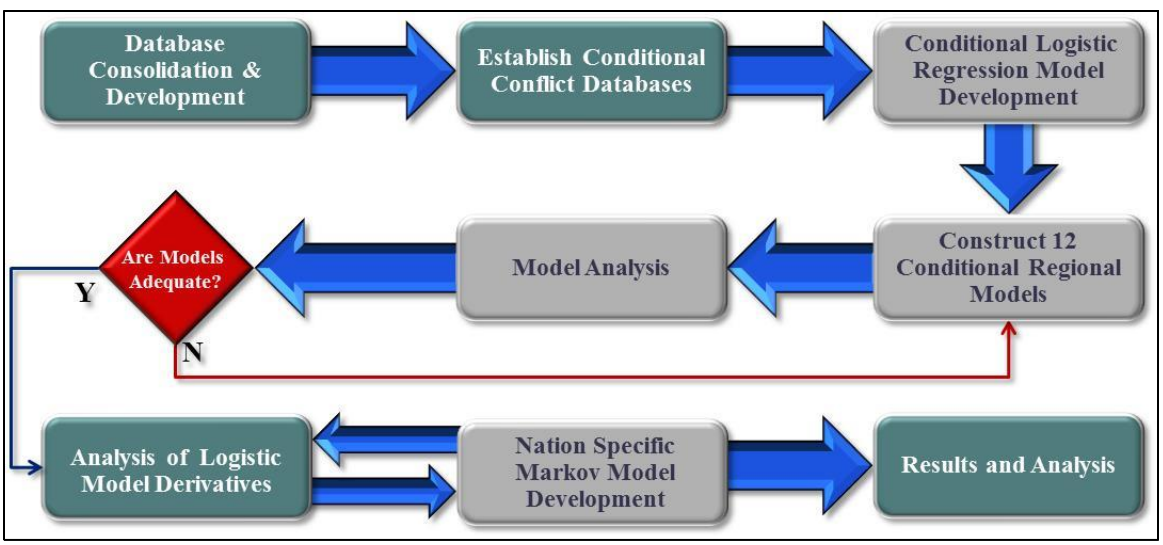


Figure 2: Study Methodology

1.6 Study Assumptions and Limitations

Assumptions

Four underlying assumptions were required to proceed with the methodology and analysis in this study. Similar to previous conflict prediction studies, this research first assumes the existence of statistical and trend variables that are accurate and viable predictors of violent conflict. Second, this study assumes that any variable identified as significant within the model will remain relevant from year to year, and for the duration of the conflict forecasting period. Next, this study assumes that the six geographic regions utilized for the development of the conditional logistic regression models provide suitable commonality in terms of economy, geography, ethnic, and religious demographics to facilitate the modeling effort. Finally, to support the forecasting of global conflict, this study assumes that regional factors relevant to conflict remain unchanged throughout the forecasting period.

Limitations

Data availability is this study's single greatest limitation, mandating a combination of data lag prediction and data imputation for the development of the conditional logistic regression models. Data lag prediction refers to the requirement of using data sets that may be one-to-three years behind the dependent variable, a suboptimal but proven method for the prediction of violent conflict in nations (Boekestein, 2015). Missing data further exacerbates the lag in the data sets, and it must be accounted for using statistical data imputation methods available in commercial software. In addition to these limitations, this study requires an expanded conflict data set, spanning the years 2004 through 2014, in order to capture enough instances of conflict transitions to effectively build the regional logistic regression models. While an expanded data set is not in itself a limitation, the dynamic conditions of the contemporary operating environment do not wholly resemble the conditions present a decade earlier, which may result in a loss of fidelity in some of the final recommended models constructed using these older data sets. In addition to the independent variables, the dependent variable "transitions into conflict" is limited by the availability of data provided by the Heidelberg Institute for Conflict Research (HIIK). This variable is derived from the HIIK's annually published Conflict Barometer; the 2014 Conflict Barometer is the most current available publication, and thus year 2014 sets the benchmark for all forecasting analyses conducted in this study.

1.7 Implications

Dr. George Box once remarked that "Essentially, all models are wrong, but some are useful" (Box, 1979). While predictive accuracy is an important aspect of this study, it was never the goal to develop a model with perfect accuracy. Instead, it is the goal of this study to gain relevant and actionable insights from the suite of models developed herein. These insights include identifying the regional factors relevant to conflict transitions, nations susceptible or "immune" to these transitions, and regional conflict trends that may impact future policy decisions. It is the expectation of this research to provide commanders and national level leadership an accurate and tractable analysis to aid the development and execution of future foreign policy and security strategies.

1.8 Overview of Remaining Chapters

Including the introduction, this thesis is comprised of five chapters. Chapter 2 reviews previous studies and literature pertaining to conflict prediction, logistic regression, and Markov models. The detailed literature review is vital to narrow the research scope of this study and provide insights into viable methods for the modeling and prediction of violent conflict transitions. Chapter 3 presents an in-depth discussion of the data base design methodology, mathematics, notation, modeling approach, and software required to answer the study questions. Chapter 4 provides a validation of both the conditional logistic regression and Markov models, and presents a comprehensive analysis of the results obtained from said models. Finally, Chapter 5 summarizes the methodology, results, conclusions, and limitations of this research, and finally proposes operationally relevant studies that may capitalize on the methodology and results developed in this thesis.

II. Literature Review

"Therefore, just as water retains no constant shape, so in warfare there are no constant conditions."

Sun Tzu, The Art of War

2.1 Overview

This research is an effort to define a methodology for the use of multi-state Markov chain models (MCM) for the prediction of nation-state transitions into and out of violent conflict. With this objective in mind, this chapter is broken down into five sections, beginning with this overview. The second section of this chapter is a survey of previous nation-state conflict prediction studies, with a focus on models predicting conflicts post 2001, their methodologies and subsequent predictive success rates. The third section reviews non-conflict oriented prediction studies utilizing multi-state Markov models, with an emphasis on viral epidemiology and spatial relations. The final section provides a synopsis and definitions for the different levels of conflict, which may be modeled as states within the MCM, and examines common prediction variables used in previous studies as well as additional variables that may be relevant to this analysis. This review is not exhaustive; instead it examines the variables and regional dynamics, highlighting the nature and factors unique to modern violent conflict.

2.2 Nation-State Conflict Prediction: Relevant Research

For the purpose of this research, we are primarily concerned with analytical and predictive studies conducted during the Era of Persistent Conflict: the period following the terrorist attacks of September 11th, 2001, influenced by the dynamic and unique challenges posed by the modern international political landscape. The seminal work, *A Global Forecasting Model of Political Instability*, a Central Intelligence Agency (CIA) funded study led by Dr. Jack A. Goldstone, part of the CIA's Political Instability Task Force, derived a series of models predicting political instability two years prior to event onset (Goldstone, et al., 2005). Utilizing a set of global, open-source data, spanning the time frame of 1955 – 2003, the CIA study compiled an exhaustive list of instability events, with the final problem set including nearly 300 "Adverse Regime Changes", "Ethnic Wars", "Revolutionary Wars", and "Genocides/Politicides" (Goldstone, et al., 2005). The study's dependent variable was the onset of political instability brought about by the occurrence of one or more of the problem set events. Multiple methodologies, including event history models, logistic regression, neural networks, and Markov processes were employed to identify factors associated with political instability, the onset of which is considered a rare event, given definitions laid out in their study. As a result, the case control method, common in epidemiological analysis of rare occurrences, became their primary methodology (Goldstone, et al., 2005). The study initially tested hundreds of variables under the assumption that the complexity associated with the onset of political instability would require an equally complex model or set of models, each specific to regime type and problem set event (Goldstone, et al., 2005). In actuality, the CIA-funded study determined these initial assumptions to be incorrect, noting that a small subset of the original variables and a relatively simple model were sufficient to model political instability across various regime types.

What separates the CIA-funded study from past conflict and political instability prediction studies was the ability to significantly reduce the unexplained variance in the model. Previous quantitative studies, sought only to find statistically significant variables, but paid little attention to the variables ability to explain variance within the overall model (Ward, Greenhill, & Bakke, 2010). However, as Dr. Michael Ward so adroitly points out, previous studies spent significant effort in the pursuit of finding statistically significant variables but little effort in determining what variables actually improve the predictive ability of the models. As a result, most of these models fail to achieve predictive accuracy rates in excess of 50%, and often times are convoluted and difficult to interpret. For nearly three years, the Political Instability Task Force struggled to develop a model having an accuracy greater than 60-70%. However, their methodology combined with an internally developed four-part regime categorization yields postdictive accuracy rates of 80% or greater (Goldstone, et al., 2005). However, the CIA funded study can only achieve these postdictive rates on a subset of randomly sampled, politically vulnerable nation-states, and thus cannot achieve "whole world" accuracy (Boekestein, 2015).

In 2007, the Center for Army Analysis (CAA) initiated the Forecast and Analysis of Complex Threats (FACT) study, which eventually became a series of four studies (FACT I-IV), each refining the data and methodology of the previous study. The original study directors Shearer and Marvin sought to develop a methodology to "predict the future conflict of select nation-states, but in a manner that facilitated explanation;" in essence a relatively simple model that was still relevant to the Army Staff (Shearer & Marvin, 2010). Conflict data used in the FACT studies was collected from the Heidelberg Institute of International Conflict Research (HIIK), which at the time classified conflict intensity levels into six categories: No Conflict, Dispute, Latent

Conflict, Crisis, Severe Crisis, and War; these categories have since been updated to: 0 – No Conflict, 1 – Dispute, 2 – Non-Violent Crisis, 3 – Violent Crisis, 4 – Limited War, and 5 – War (Heidelberg Institute for International Conflict Research, 2014). As part of the methodology, the FACT study maps the four highest HIIK intensity levels to two categories: Conflict and Peace (Shearer & Marvin, 2010). In addition to the HIIK data, the FACT studies utilized a variety of open source governmental and non-governmental databases, such as the *World Bank*, *Food and Agricultural Organization of the United Nations*, and the *Polity IV Project* to gather feature (macro-structural indicators) data. The methodologies employed in FACT I-III used a common weighted moving average forecasting model combined with a factor analysis algorithm to classify their specific future feature vectors, known as the K-Nearest Neighbor. Using principal component analysis (PCA) was used to create the multiple features employed in the study, ultimately maximizing the explained variance within the data. The K-Nearest Neighbor algorithm then classifies each of these feature vectors as a function of the n-closest past feature vectors, with decision rules requiring either a simple or super-majority for a classification of Conflict or Peace, with best results occurring when K = 7 (Shearer & Marvin, 2010). The FACT studies yielded accuracies in excess of 85% when the predicted nation scores were classified as conflict, peace, or uncertain. However, this high postdictive accuracy is due to the fact that 25% of the 157 considered nations are categorized as "uncertain", reducing the overall confidence in the predictive ability of the model.

In his 2011 paper, *Predicting Armed Conflict, 2010-2050*, Hegre employs dynamic multinomial logit model estimation techniques to develop a three-state transition probability matrix capable of predicting changes in global and regional incidences of armed conflict out to year 2050. The Hegre study created and used the Uppsala/PRIO conflict data set, consolidating relevant data for 169 countries from 1970 to 2009. The Uppsala/PRIO data reports three conflict levels: "No Conflict" or less that 25 combat-related deaths per year, "Minor Conflict" or between 25 and 999 combat related deaths per year, and "Major Conflict" when greater than 1000 combat related deaths are reported in a year (Hegre et al., 2011). The primary predictive methodology employed by Dr. Hegre, was a C++ based simulation based upon a statistical model of conflict onset, escalation, and termination dependent on a set of both endogenous and exogenous variables (Hegre et al., 2011). The methodology employs a nine step process of (1) Estimating the underlying statistical model through dynamic multinomial estimation; (2) Developing assumptions about the distribution of the exogenous variables; (3) Simulating conflicts for the current year; (4) Drawing a realization of the coefficients of the multinomial logit model (Equation 8); (5) Calculating the nine probabilities of transition between states shown in Table 1; (6) Randomly drawing whether a country experiences conflict, based on estimated probabilities; (7) Updating the values of the explanatory variables; (8) Repeat steps (4) – (7) for each year of the forecast; and finally (9) Repeating step (3) – (8) a number of times to even out the impact of individual realizations of the multinomial logit coefficients and the individual values of the probability distributions (Hegre et al., 2011).

A dynamic, multinomial logit model was used to estimate the probability transition matrix with the outcome at time t, based on a $t\text{-}1$ set as the indicator variables. The model is identified by setting the baseline outcome to $j = 0$, "No Conflict", resulting in the estimates β_1 and β_2 being interpreted as the impact of the explanatory variable, x, on the probability of being in “Minor Conflict” and “Major Conflict” relative to “No Conflict” (Hegre et al., 2011). Essentially, this model shows which variables increase the risk of conflict onset; however, the predicted duration of the conflict is calculated through the use of interaction terms between the states at *t-1* and the predictor variables, producing the transition probability matrix shown in Table 1.

Table 1: Transition Probability Matrix: Conflict at *t* vs. *t-1*, 1970-2009
(Hegre, Karlsen, Nygard, Strand, & Urdal, 2011)

Conflict at *t-1*	No Conflict	Minor Conflict	Major Conflict	Total
No Conflict	5116 (0.966)	156 (0.029)	23 (0.004)	5295 (1.000)
Minor Conflict	145 (0.207)	481 (0.689)	72 (0.103)	698 (1.000)
Major Conflict	24 (0.070)	70 (0.205)	247 (0.724)	341 (1.000)
Observations	5285	707	342	6334

Row proportions in parentheses

The Hegre Model divides the world into nine regions, based upon the observation that conflict tends to cluster in a few geographical regions, sharing similar rates associated with risk factors such as infant mortality rates or poverty levels. These regions are: South and Central America and the Caribbean; Western Europe, North America, and Oceania; Eastern Europe; Western Asia and North Africa; West Africa; East and Central Africa; Southern Africa; South and Central Asia; Eastern and South East Asia. However, the methodology further investigates the “neighborhoods” associated with each nation. The neighborhood of country A is defined as all n countries $[B_1 \ldots B_N]$ that share a border with A; where country A shares a border with country B_i if there is less than 100km distance between any points of their territories (Hegre et al., 2011). This was an important factor in their methodology, as it allowed them to model the cross border effects of conflict on near neighbors, creating a measure of neighborhood effects, relevant to each country.

The Hegre model is unique from previous conflict prediction studies, as it does not restrict its predictions to solely the onset of conflict, thereby excluding ongoing conflicts. Additionally, the Hegre Model simultaneously predicts conflict onset, escalation, and termination, allowing for the prediction of both the global and regional incidence of armed conflict. The prediction horizon is also unique to the Hegre model due to its length, 7-9 years, with an average postdictive accuracy (across all regions) of 79%, and a false positive rate of 8.5% given a probability threshold of $p > 0.3$, for the state of interest (Hegre et al., 2011). As the title of Hegre's paper indicates, his objective was to predict conflicts out to year 2050, which he accomplished through the use of projections of predictor variables, as provided by the UN World Population Prospects and the International Institute of Applied Systems Analysis (Hegre et al., 2011). Using this data, Hegre predicts an overall decline in the global incidence level of violent conflict; a decline attributed to improvements in variables associated with infant mortality, education and youth bulges (Hegre et al., 2011). However, since these long term predictions are based on projections as opposed to actual data, the Hegre Model estimates should be interpreted as long-term global, and to a lesser extent regional, conflict trends, given projected conditions as opposed to specific national level predictions.

Boekestein conducted the most recent analysis concerning the prediction of future nation-state conflict, in his study *A Predictive Logistic Regression Model of World Conflict Using Open Source Data* (Boekestein, 2015). As the name implies, the Boekestein model uses logistic regression similar to the CIA-funded and FACT studies to produces a parsimonious model that is tailored to each of the six geographical regions identified in his study: Sub-Saharan Africa; South and East Asia; Eastern Europe and Central Asia; Arab Nations; Organization for Economic Cooperation and Development (OECD); and Latin America, comprising 180 of the 193 United Nations member nations, with the states of Palestine and Kosovo also included for consideration. As in previous studies, conflict intensity or level of violence was chosen as the dependent variable for this study and is based off the levels calculated by the HIIK. The HIIK levels of violence are calculated using the five metrics of: Weapons – light or heavy, Personnel – number engaged per month; Casualties – number per month, destruction – infrastructure, accommodation, economy, and culture; and Cross Border Refugees and Internally Displaced Persons (IDP) – number per month (Heidelberg Institute for International Conflict Research, 2014). Using these metrics, the HIIK assigns one of the six aforementioned intensity levels to every identified political conflict. The Boekestein model subsequently maps these six levels of conflict to two dependent variables: "Not Violent Conflicts": Levels 0 – 2, and "Violent Conflicts": Levels 3 – 5.

Twenty-two statistic and four trend variables were considered for this study, thirteen of which are common to the CIA funded and the FACT studies (Boekestein, 2015). The data supporting these variables is gathered from multiple sources to include the World Bank, HIIK, and the CIA World Fact Book, with some sources maintaining data sets from 1970. As Boekestein points out, many of these data sets are not complete or available for the current year of the study, requiring a two or three year lag in the model to predict current year nation-state conflict levels. Additionally some variables had significant gaps in the data requiring imputation to complete the data set. For example the data set supporting the variable "Conflict in Bordering States", whose calculation took into account the number of bordering nations, and the percent border shared with nation *i*, required the imputation of data for 29 island nations (Boekestein, 2015). A rigorous variable screening process, to check for collinearity among the set of 26 variables was implemented prior to model development using three separate analysis methods. Despite the rigorous testing, the initial Boekestein models failed to achieve postdictive accuracy rates in excess of 76%. To improve the model, several factor analysis and noise reduction techniques were used to reduce the initial set of 23 variables to a set of six factors, with highly correlated variables represented by a single factor (Ahner, Boekestein, & Deckro, 2015). Given the nature of the study, it was also desirable to minimize the number of false negative reports by the model, i.e., the number of times the model predicts "Not in Violent Conflict" when in actuality the nation in question is in "Violent Conflict" (Boekestein, 2015). This objective was accomplished by adjusting the logistic regression cutoff level, for which the default setting was 0.5, through extensive sensitivity testing. The testing determined a potential need for an additional variable to explain a nation's region, due to the nature of the particular nations consistently reporting as either false positives or false negatives. This insight led to the construction of separate model for each of the six previously identified regions. Each model employs a specific subset of variables from the original 26 statistic and trend variables that best describe the conflict risk factors unique to each region. This methodology resulted in a reduction of false negative predictions in the range of 2–7%, and a combined postdictive accuracy for both the model and validation sets of 80.22%, given a logistic regression cutoff of 0.28 (Ahner, Boekestein, & Deckro, 2015).

2.3 Markov Models and the Prediction and Spread of Disease Epidemics

The prediction of the outbreak and spread of disease epidemics in many ways is analogous to the study and prediction of violent conflict and its antecedents. Additionally given the various states of disease a host, outbreak, or epidemic can exist in, such as no signs of disease, susceptible, infected, and cured, the prediction methodology lends itself to the use of Markov models. The 2007 paper, *Bayesian Markov switching models for the early detection of influenza epidemics*, explores a methodology for the early detection of influenza outbreaks, using a two-state Markovian process. The methodology created by Martinez-Beneito and his team, employs a two-state, or binary, hidden Markov process in which the population is in a non-epidemic or epidemic phase, states 0 and 1 respectively. The underlying concept of the model is to associate the variable $Y_{i,j}$, the difference in disease rates between weeks i and $i+1$ in year j, with $Z_{i,j}$, the unobserved random variable that indicates the state of the system (Martinez-Beneito et al., 2007). The model for the $Y_{i,j}$ variable is specific to the state and season of the system, and is either an Gaussian white-noise process (non-epidemic) or an autoregressive process of order 1 (epidemic). Upon determination of the model, the parameters $P_{0,0}$ and $P_{1,1}$ were estimated using the Bayesian paradigm requiring the specification of prior distributions.

To validate the model's predictive accuracy, Martinez-Beneito compared its performance using a near term partial and complete data set. The model was constructed using a dataset covering a nine-year period, allowing the team to develop robust estimates of the various parameters used in the model. However, given the nature of disease outbreaks, time horizons are measured in weeks as opposed to months or years, requiring that the model be tested using limited subset of the near-term preceding weeks. The results showed that, even with the reduced data set, the model predicted the same incidence of epidemic in 93% of the scenarios of the model using the full data set, given a $p > 0.30$ (Martinez-Beneito et al., 2007).

In his 2004 paper, *The analysis of hospital infection data using hidden Markov models*, Cooper proposes a new process to analyze infections that are generally considered endemic to hospitals, and are carried asymptomatically before infections begin to appear in proportions of the patient population. The data associated with hospital-acquired infections generally consists of short time series with low number counts (Cooper & Lipsitch, 2004). For his analysis, Cooper stresses the importance of patient-to-patient transmission, which shares many similarities to conflict spillover from one state to the next. The transmission chain is modeled using a structured hidden continuous time Markov chain over a short time increment h. Germane to this discussion are the parameters β/N, the transmission rate to each susceptible patient in population N given an infected host; v, the probability of being a pathogen host; μ, patient discharge rate; and $C_t \in \{0, 1, 2, \ldots N\}$, the state of the system given as the number of infected hosts at time t (Cooper & Lipsitch, 2004).

In this model new infections arise due to cross-infection, at a rate proportional to the product of the number of infected hosts, C_t, and the number of susceptible patients, ($N - C_t$). New infections can also occur in the newly discharge susceptible population (Cooper & Lipsitch, 2004). In the modeling of cross border conflict spill-over, the parameter β/N, can be interpreted as the proportion of a nation's border that shares a mutual border with a state currently in violent conflict; where v and μ are the respective probabilities of entering and terminating a conflict given a neighboring state is in conflict.

In his concluding remarks, Cooper identifies several limitations associated with his methodology, specifically that it may not be appropriate for large systems. He specifically states: “A further limitation is that while such a model may be appropriate for a single ward or unit, for larger hospital populations made up of several interacting units its value is not so clear” (Cooper & Lipsitch, 2004). He ties the reason for this limitation to the model’s use of short time series with limited data, increasing the collinearity between multiple variables. Therefore, the overall methodology is likely not appropriate for the prediction of nation-state conflict, but the modeling of disease transmission gives insight on how to possibly model cross-border conflict spillover.

The final methodology we will explore is the *Modeling of Viral Epidemiology in Connected Networks*, discussed by Spears of the Naval Research Laboratory. In this instance, Spears adapts methodology for the prediction and spread of disease epidemics and applies them to the spread of computer viruses in a network. Given the level of interconnectedness shared by most nations, a result of globalization, it is easy visualize the current geo-political topology as a vast network, where conflict in one state sends shockwaves through the network, eventually affecting numerous other nations. The methodology for this research employs very general discrete-time Markov chains and continuous-time differential equations to model the propagation of viral attacks in a network. The network envisioned in this in this study consists of N nodes that exist in one of four medical conditions or states: S, susceptible; E, exposed; I, infected; and C, cured (Spears, 2001). The discussion of the methodology builds upon two- and three-state Markov chains, but for the purposes of this discussion we will focus on his four condition *S-E-I-C* model. In this model a susceptible node must be exposed to the pathogen before it becomes infected, infected before cured, cured before susceptible.

The transition associated with this Markov model requires that $I' - I$ more nodes become infected at time j, where I' can be less than, equal to, or greater than I, with similar requirements for the other three medical conditions (Spears, 2001). The transition probabilities for this model take on binomial characteristics that either a node exists in a specific medical condition, or it does not. Four variables are employed in this model: α, the probability a susceptible patient become exposed to a pathogen; μ, the probability an exposed patient is infect; δ, the probability an infected patient is cured; and δ', the probability a cured patient become susceptible. In the end, the methodology employed by Spears may permit the modeling of the spread of violent conflict as a function of bordering states, or geographic nearest neighbors in the case of island nations. Additionally, through the depiction of strongly and weekly connected nodes, we have a methodology that may simulate secure and porous international borders.

2.4 Relevant Variables

As stated previously, many conflict prediction studies have expended substantial effort and resources in the pursuit of statistically significant variables while failing to understand how those variables improve the predictive qualities of their respective models. When analyzing conflict predictor variables used in previous studies, one must ask: “will these variables still remain significant in future conflicts?” Furthermore, will variables currently identified as insignificant in current conflicts become significant as the nature of violent conflict evolves? When analyzing and studying different conflict predictor variables, the analyst must avoid falling into a common trap propagated through dystopian visions of future conflict that are so common in this day and age (Johnson, 2014). The trap is the belief that future global incidences of violent conflict will only increase, eventually becoming unmanageable by most national governments. However, as Hegre noted, if UN projections prove reliable, his model actually predicts the opposite outcome, with the global incidence of conflict decreasing by 2050. All this being said, it is imperative the analyst understands the effects of historical predictor variables on current conflicts while staying abreast of emerging trends, predictors, and their effects that will frame the nature of future conflicts.

The recent Boekestein study created a model using 27 total variables, achieving accuracy rates in excess of 80% by region. Given these results, one can assume the set of 27 statistic and trend variables represent a set of available predictors that offer excellent predictive accuracy of modern violent conflicts, if properly tailored for different world regions. As is seen in Table 2, each region in the Boekestein model has a particular subset of relevant variables, with some regions requiring as few as two and other regions as many as nine. Additionally, the importance of the variables, referenced by the index corresponding to the variable-region intersection in Table 2, is also region specific. For example, individual freedom statistics were shown to be the most significant variable in three regions: Sub-Sahara Africa, Easter Europe and Central Asia, and the OECD, but is the fifth most significant variable for Arab Nations and Latin America. However, this table is not all-inclusive due to the absence of variables: border conflict, religious diversity, ethnic diversity, and the HIIK trend, which were removed during final model construction (Boekestein, 2015).

Table 2: Region Specific Relevant Variables
(Boekestein, 2015)

Variable/Region	Sub-Sahara Africa	South and East Asia	Arab Nations	Easterm Europe and Central Asia	OECD	Latin America
Freedom	1		5	1	1	5
2 Yr Freedom Trend						9
3 Yr Freedom Trend	6					
5 Yr Freedom Trend	8					
Regime Type (Central)	7				4	
Regime Type(Democratic)	9					
Polity IV					3	2
GDP Per Cpita						6
Refugees Asyulm	3		3			7
Refugee Origin		4				
Unemployment	5					8
Rural Population						3
Infant Mortality					2	4
Caloric Intake		1			5	
Death Rate		3	1			1
Arable Land			2			
Population Growth Rate					7	
Improved Water	2					
Trade	4	2	4	2	6	

A review of previous studies reveals that variables such as political statistics, conflict history, infant mortality rates (IMR), population statistics, civil liberties, and ethnic/religious dominance or diversity are frequently employed as significant predictors of violent conflict. In addition to these variables, the Hegre model introduces the variables related to current conflict intensity, education, youth bulges, international treaties, neighborhood characteristics (a conglomeration of growth rates, per capita GDP, education levels, IMR, and other political considerations), and oil (Hegre et al., 2011). The oil variable is of particular interest due to the hypothesis that nation-states whose GDP is dependent upon primary commodities through export revenue, such as oil, tend towards weaker governmental institutions putting them at greater risk for violent conflict (Hegre et al., 2011).

Economic factors such as gross domestic product (GDP), primary commodity exports, and income growth have also been demonstrated as significant predictors of violent conflict. As noted in their 2002 paper, *On the Incidence of Civil War in Africa*, Collier and Hoeffler employ an econometric model to predict incidences of conflict in Africa (Collier & Hoeffler, 2002). While their study primarily focused on African nations, the authors note that patterns of conflict in Africa largely resemble other developing regions throughout the globe, indicating that economic variables may be useful conflict predictors in other regions. Similarly, other studies have shown that population variables, specifically ethnic and religious oriented statistics, are powerful predictors of and historical contributors to violent conflict. In the 2001 paper titled *Ethnicity, Insurgency, and Civil War*, Fearon and Laitin aurgue that post-Cold War civil wars and insurgency were driven and exacerbated through numerous ethnic and religious factors, not the least of which was ethnic nationalism (Fearon & Laitin, 2001).

In his book *Out of the Mountains: The Coming age of the Urban Guerilla*, Kilcullen discusses several drivers of violence and instability that may compliment the current set of common conflict predictors. The basic premise of his work is that future conflict is likely to occur in the urban sprawl of coastal mega-cities, and in the peri-urban settlements that exist in Africa, the Middle East, Latin America, and Asia (Kilcullen, 2013). He discusses the growth of criminal violence networks combined with a simultaneous decay or complete lack of basic infrastructure (such as sanitation) as two drivers of instability. The respective rates of growth and decay of these two predictors may serve viable and significant variables in a predictive model. The same variables, along with national inflation rates, changes to military expenditure/manning levels, and the application of international sanctions are echoed in the paper, *Statistical Approaches to Developing Indicators of Armed Conflict*, whose purpose is to "explore the feasibility of developing a meaningful system of indicators of armed violence" (Kisielewski, Rosa, & Asher, 2010).

2.5 Summary

In this chapter, we surveyed multiple sources to identify useful methods, theories, and variables to enable the development of a methodology for the prediction of violent nation-state conflict using Markov Chain Models. Previous conflict prediction studies employ multiple methodologies to include logistic regression and simulation with the best models achieving accuracy rates in excess of 80%. While methodologies and variables differed between studies, the common trend was to construct region-specific models to better estimate the global incidence of violence. This methodology allows for the use of region-specific significant variables, whose value when applied to a global model may be insignificant or even detrimental to the predictive accuracy of the model. Next, we survey a group of studies using Markov models for the prediction and analysis of the spread of disease epidemics, which share common traits with the spread of conflict. The studies featured in this chapter use multi-state hidden Markov models, emphasizing patient-to-patient transfer of pathogens. Notable studies combine Markov models with strong and weakly connected patient networks, which may provide a suitable methodology to the modeling of the nation-states within their various regions. Finally, we review commonly used and emerging predictor variables that are relevant to modern conflict. Such variables include recent conflict history, infant mortality rates, various political and economic statistics, and region type. Subject matter experts also identify predictors such as crime rates, population migration, and changes to military spending and force levels as possible drivers of instability in susceptible nations.

III. Methodology

"This is no formula of war. No one dares to arrogantly claim to have the perfect method in the sphere of war. No one has ever been able to use one method to win all wars. But it does not mean that there are no rules regarding war."

Qiao Liang, Unrestricted Warfare

3.1 Chapter Overview

This research examines the methods germane to the prediction and forecasting of nation-state violent conflict transitions. Section 3.2 describes the methodology guiding the development of the conditional conflict database, to include variable selection, database design, and data imputation. Next section 3.3 discusses the mathematical principals and development of the conditional logistic regression models to include the Synthetic Minority Over-sampling Technique. Finally, Section 3.4 describes the theory and development of the nation specific Markov models used to determine the near- and long-term conflict trends of the nation-states examined in this research.

3.2 Conditional Conflict Database Development

Nation-State Case Selection

This study examines the incidences of violent conflict for 181 of the 193 member states of the United Nations, as well as Palestine, which is referred throughout this study as the West Bank (United Nations, 2015). The 182 nations states examined in this research are consistent with those surveyed in the Boekestein model (Boekestein, 2015). The 12 member states not considered in this study are Andorra, Dominica, Liechtenstein, The Marshall Islands, Monaco, Nauru, Palau, Saint Kitts and Nevis, Saint Lucia, Saint

Vincent and the Grenadines, San Marino, and Tuvalu. These nations were omitted due their relatively small populations, combined with inadequate or incomplete data. Similar to previous studies, disputed territories and regions such as Nagorno-Karabakh, South Ossetia, Western Sahara, Somaliland, and Taiwan are also omitted from consideration.

Case selection, the specific years of interest for each nation-state status observation, was predicated on the availability of adequate data, combined with the requirement to capture sufficient amounts of data relevant to the current operational environment. Cases for all 182 nation states are drawn from the years 2004 to 2014, the 11-year period immediately following the United States led invasion of Iraq in March 2003; a total of 2,002 individual nation-year cases. This time period was ultimately selected based on the findings of several recent operational environment assessments that emphasized the importance of current visible trends for meaningful conflict prediction (Johnson, 2014).

Description of the Dependent Variable

This study utilizes the conditional dependent variable "Conflict Transition given Previous Year Status", which is derived from the Heidelberg Institute for International Conflict Research (HIIK) conflict intensity levels for each nation. To understand how the HIIK derives a nation's conflict intensity score, one must first define a set of conflict measures and conflict items which constitute the key elements of the score. The HIIK definitions for *Conflict Measures* and *Conflict Items* are provided.

Conflict Measures

> *Conflict measures are actions and communications carried out by a conflict actor in the context of a political conflict. They are constitutive*

for an identifiable conflict if they lie outside established procedures of conflict regulations and – possibly in conjunction with other conflict measures – if they threaten the international order or a core function of the state. Core state functions encompass providing security of a population, integrity of territory and of a specific political, socioeconomic or cultural order (Heidelberg Institute for International Conflict Research, 2014).

Conflict Items

Conflict items are material or immaterial goods pursued by conflict actors via conflict measures. Due to the character of conflict measures, conflict items attain relevance for the society as a whole – either for coexistence within a given state or between states. This aspect constitutes the genuinely political dimension of political conflicts (Heidelberg Institute for International Conflict Research, 2014).

The 2014 Conflict Barometer developed by the HIIK utilizes the 10 conflict items described in Table 3.

Table 3: Conflict Items
(Heidelberg Institute for International Conflict Research, 2014)

The Heidelberg Institute for International Conflict Research Conflict Items	
Conflcit Items	**Description**
System / Ideology	Conflict actor aspires to change the ideological, religious, socioeconomic, or judicial orientation of the political system or regime.
National Power	The power to govern a state.
Autonomy	Attaining or extending political self-rule of a population within a state of a dependent territory without striving for independence.
Secession	The aspired separation of a part of a territory aiming to establish a new state or to merge with another state.
Decolonization	The desired independence of a dependent territory from foreign rule.
Subnational Predominance	The attainment of de-facto control by a government, a non-state organization, or a population over a territory or a population.
Resources	The pursuit of the possession of natural resources or raw materials, or the profits gained thereof.
Territory	The desire to change the course or alter an international border.
International Power	The change aspired in the power constellation in the international system or regional system therein, especially by changing military capabilities or the political or economic influence of a state.
Other Items	A residual category.

To determine a conflict's intensity level, the HIIK utilizes five proxy measures to assess the means and consequences of the given conflict. The means of conflict include the weapons and personnel involved therein, while the conflict consequences includes the casualties, refugees and internally displaced persons (IDP), and destruction sustained by said conflict (Heidelberg Institute for International Conflict Research, 2014). The parameters and assigned values are provided in Table 4.

Table 4: HIIK Proxy Measures

(Heidelberg Institute for International Conflict Research, 2014)

		Conflcit Means		
		0 Points	1 Point	2 Points
Conflict Consequences	0 Point	Violent Crisis	Violent Crisis	Limited War
	1 Point	Violent Crisis	Limited War	War
	2 Ponts	Limited War	War	War

Weapons			
		Employment	
		Light	Heavy
Type	Light	0 Points	
	Heavy	1 Point	2 Points

Personnel		
Low	Medium	High
Pax $\leq$ 50	50 < Pax $\leq$ 400	Pax > 400
0 Points	1 Point	2 Points

Casulaties		
Low	Medium	High
Cas $\leq$ 20	20 < Cas $\leq$ 60	Cas > 60
0 Points	1 Point	2 Points

Refugees & IDPs		
Low	Medium	High
Ref $\leq$ 1000	1000 < Ref $\leq$ 20,000	Ref > 20,000
0 Points	1 Point	2 Points

Destruction		
Low	Medium	High
Within 0 Dimensions	Within 1 - 2 Dimensions	Within 3 - 4 Dimensions
0 Points	1 Point	2 Points

The intensity levels of a particular conflict are an attribute sum of the conflict measures for a given geographic area and time period. The HIIK employs a six-level model with the following intensity levels: 0 – No Conflict, 1 – Dispute, 2 – Non-violent Crisis, 3 – Violent Crisis, 4 – Limited War, 5 – War (Heidelberg Institute for International Conflict Research, 2014). Nations that were or are currently experiencing multiple conflicts are assigned an overall HIIK intensity level equating to the highest level assigned to any of the ongoing conflicts for a particular year. These levels were subsequently mapped to a binary variable called "Level of Violence", with levels 0 through 2 mapped to "Non-Violent Conflicts" and levels 3 through 5 mapped to "Violent Conflicts" (Boekestein, 2015).

The conditional dependent variable "Conflict Transition given Previous Year Status" is mapped to the level of violence variable for the preceding year ($y - 1$) and the level of violence variable for the following year. A transition is said to occur if the status changes over the course of the year. The mapping of the second order dependent variable from the HIIK conflict intensity levels is provided in Table 5.

Table 5: Mapping of Conditional Dependent Variable

HIIK Intesity Level	HIIK Terminology	Level of Violence	Status Year (y - 1)	Conflict Transtion year (y), given status year (y - 1)
0	No Conflict	Non-Violent Conflicts	Not in Conflict	Transition into Non-conflict (0)
1	Dispute			
2	Non-violent Crisis			Transition Into Conflict (1)
3	Violent Crisis	Violent Conflicts	In Conflict	Transition into Non-conflict (0)
4	Limited War			
5	War			Transition Into Conflict (1)

Independent Variable Selection

This study incorporates 26 nation specific statistic variables and four trend variables obtained from six data repositories: the Heidelberg Institute for International Conflict Research, The World Bank, Central Intelligence Agency (CIA) World Fact Book, Freedom House, the Center for Systemic Peace, and the Food & Agriculture Organization of the United Nations (UN FAO). Variable selection was heavily influenced by the Center for Army Analysis FACT studies, 12 variables in common (Reed, 2013); the CIA – Goldstone study, three variables in common (Goldstone, et al., 2005), and the Boekestein study, 25 variables in common (Boekestein, 2015). These and similar studies have repeatedly demonstrated the significance of theses variables as conflict predictors, hence their consideration in this study. Recent studies, such as those conducted by Hegre at the University of Oslo, have expounded the necessity of including statistical variables representing emerging trends within the operational environment such as military spending, urbanization of populations, loss of natural sources of fresh water, and burgeoning youth populations that are seen as future drives of instability (Hegre et al., 2011). Therefore, the following five variables are also included for consideration within this study and are defined as:

Military expenditure (Percent of central government expenditure):

Military expenditures data from the Stockholm International Peace Research Institute (SIPRI) are derived from the North Atlantic Treaty organization (NATO) definition, which includes all current and capital expenditures on the armed forces, including peacekeeping forces; defense ministries and other government agencies engaged in defense projects; paramilitary forces, if these are judged to be trained and equipped for military operations; and military space activities. Such expenditures include military and civil personnel, including retirement pensions of military personnel and social services for personnel; operation and maintenance; procurement; military research and development; and military aid (World Bank, 2015).

Military expenditure (Percent of gross domestic product):

This variable is defined in the same fashion as above, but takes into account the relative defense expenditure as it relates to the total national output.

Population ages 0 – 14 (percent of total):

This variable is based on a nation’s population between the ages 0 to 14 as a percentage of the total population. Population is based on the *de facto* definition of population (World Bank, 2015). This variable is referred to as "Youth Bulge" throughout this study.

Renewable internal freshwater resources per capita (cubic meters):

Renewable internal freshwater resources flows refer to internal renewable resources (internal river flows and groundwater from rainfall) in the country. Renewable internal freshwater resources per capita are calculated using the World Bank's population estimates (World Bank, 2015). Due to the limitations of this data set, this variable is the average of the 2007, 2012, and 2013 statistics for each nation, and it is subsequently fixed as a stationary variable.

Government Type:

This is a six-level indicator variable derived from the Polity IV scores for each nation. Polity is defined as a political or governmental organization; a society or institution with an organized government (Marshall, Gurr, & Jaggers, 2014). Polity IV scores a nations political body on a 21 point scale of -10 (fully autocratic) to 10 (fully democratic), with additional identifiers -66 (indicating foreign interruption), -77 (indicating anarchy), and -88 (indicating a transitional government). From these scores the six levels are defined as: Level 0: *Autocratic Government* (Polity IV: -10 to -6); Level 1: *Emerging Democratic Government* (Polity IV: -5 to +5); Level 2: *Democratic Government* (+6 to +10); Level 3: *Foreign Interruption* (Polity IV: -66); Level 4: *Anarchy* (Polity IV: -77); Level 5: *Transitional Government* (Polity IV: -88). This variable was included to provide greater fidelity when modeling political instability within a nation. The Center for Systemic Peace provides polity scores for 166 of the 182 nations considered in this study. To account for this data gap, government type was correlated to the "Regime Type" variable that is discussed later in this chapter.

Overview of Independent Variables

Table 6 provides a synopsis of the 30 statistical and trend variables. Several near-year data sets (for years 2012, 2013, and 2014) are missing. These occurrences result in a "data lag" ranging between 1 and 2 years, based upon the year 2014 (forecast year 0), for 14 of the 30 variables. In cases involving variables with a data lag, the variable i at year j will be used to predict conflict at year $j + lag(i)$. For example, the variable arable land has two-year lag in the data set requiring that the 2012 data model 2014 conflicts. There are two serious implications when constructing a predictive model using "lagged" data. The first implication is that we are attempting to develop a predictive tool using less current data that may not capture or completely disregards current trends that lead to conflict transitions, thus reducing the accuracy of the model. The second implication is that such data ultimately increases the overall variance in the model due to increased forecasting time horizons. In addition to the data lag, incomplete data sets (i.e., variable-year instances with less than 182 entries) are also pervasive. Imputation methods employed to replace missing data are covered later in this chapter.

Table 6: Independent Variables

Year of First Data Set	Lag (years)	Variable	Number of Entries per Year 2012	2013	2014
World Bank Variable					
1961	2	Arable Land (hectares per person)	181		
1961	1	Birth Rate (per 1,000 people)	182	182	
1961	1	Death Rate (per 1,000 people)	182	182	
1961	1	Fertility Rate (births per woman)	182	182	
1960	0	GDP Per Capita (current USD)	179	179	164
1990	0	Improved Water Source (% population with access)	178	175	175
1960	1	Life Expectancy (years)	182	182	
1990	2	Military Expend (% Gov Spending)	100		
1988	0	Military Expend (% GDP)	179	141	131
1961	0	Infant Mortality rate (per 1,000 live births)	182	182	182
1961	0	Population ages 0 - 14 (% of total population)	182	182	182
1961	0	Population density (people per square kilometer)	181	181	181
1961	0	Population Growth (annual %)	181	182	182
1990	1	Refugee Population by county of asylum (% population)	159	160	
1990	1	Refugee population by country of origin (% population)	180	181	
1962	Locked	Renewable Fresh Water per Capita (cubic meters, average of 2004 - 2014 data)	174	174	174
1960	0	Trade (% GDP)	168	161	130
1991	1	Unemployment (total % of labor force)	171	171	
CIA World Fact Book Variables					
2010	0	Border Conflict Score	182	182	182
	Locked	Regime Type (3 level indicator variable)	182	182	182
	Locked	Ethnic Diversity (% of Dominant Ethnic Group)	182	182	182
	Locked	Religious Diversity (% of Dominant Ethnic Group)	174	174	174
Freedom House, The Center for Systemic peace, and Food & Agriculture Organization of the United Nations Variables					
1972	0	Freedom Score (Average of Civil Liberties and Political Rights (scores 0 to 1))	180	180	180
1960	0	Polity IV (Political behavior score -10 to 10, and -66, -77, -88)	166	166	166
1960	0	Government Type (6 level indicator variable derived directly from Polity IV scores)	166	166	166
1961	1	Caloric Intake (average caloric intake from all sources per person)	39	39	
Trend Variable					
1996	1	2 Yr Conflict Intensity Trend (Derived from HIIK intensity levels)	182	182	182
	1	2 Yr Freedom Trend (Derived from Freedom Score)	179	180	180
	1	3 Yr Freedom Trend (Derived from Freedom Score)	179	180	180
	1	5 Yr Freedom Trend (Derived from Freedom Score)	180	181	181

A majority of the independent variables are self-explanatory in both origin and function; however, the derivation of several key statistical and trend variables requires further discussion.

Border Conflict Score

Conflicts in bordering states are cited as a variable of interest in both the CIA-Goldstone (as a binary indicator variable) and the Boekestein studies. The developed border conflict score seeks to model the external pressures applied to a nation as a function of HIIK intensity level of a nation's bordering neighbors for a given year, and the relative proportion of the international border attributed to each of those nations. The international border data was obtained from the CIA World Fact Book. The equation for calculating the Border Conflict Score is defined in Equation 1.

$$Cb_{ij} = \sum_{i=1}^{n} x_{ij} p_i \text{ for } \forall\, j$$

Equation 1: Border Conflict Score (Boekestein, 2015)

Where:

Cb_{ij} = Conflict in border states statistic
n = number of bordering nations
x_{ij} = HIIK intensity level for nation i for year j
p_i = percent of border shared with nation i
i = Country $\in$ {1, 2, ..., 182}
j = Years $\in$ {1996, 1997,..., 2014}

The border conflict score for Afghanistan in 2014 is provided as an example of the variable calculation in Table 7.

Table 7: Border Conflict Score Example

Afghanistan Boder Conflict Score 2014			
Bordering State	Border (km)	p_i	x_{ij}
China	91	0.015	4
Iran	921	0.154	3
Pakistan	2670	0.446	5
Tajikistan	1357	0.227	3
Turkmenistan	804	0.134	1
Uzbekistan	144	0.024	2
Border Conflict Score (Cb_{ij})		**3.61**	

Regime Type

Regime Type is a three level indicator variable that was first cited in the CIA-Goldstone study as a significant predictor of political instability. The Boekestein study was the first to employ the variable in its current simplified form after mapping the 57 government descriptions provided in the CIA World Fact Book to 10 then subsequently three nominative variables as shown in Table 8 (Boekestein, 2015).

Table 8: Mapping of Regime Type
(Boekestein, 2015)

Expanded Regime Type	
Class	**Total**
Communist	4
Dictatorship	2
Military Junta	1
Monarchy	24
Theocracy	2
Democracy	39
Republic	107
Transitional Government	2
Disputed	1
Gand Total	**182**

Reduced Regime Type	
New Class	**Total**
Central/Ruling Party	36
Democratic	137
Emerging, Transitional, recent change, disputed	9
Grand Total	**182**

The three levels of this variable are mapped as: Level 0: *Central rule / ruling party*; Level 1: *Emerging, transitional, or disputed*; Level 2: *Democratic government*. Unlike the *Government Type* indicator variable, *Regime Type* is locked, meaning that it cannot change from year to year. *Regime Type* is correlated with the new dynamic indicator variable *Government Type*, which is envisioned as the primary means for model political institutions. The continued use of *Regime Type* within this study is as a modeling alternative to the new variable.

Freedom Score

The statistical variable *Freedom Score* was first identified as a significant variable during the Boekestein study which sought to develop a variable that incorporated the highly correlated aspects of the *Civil Liberties* and *Political Rights* variables aggregated by the Freedom House data base (Boekestein, 2015). Freedom house has compiled this data set since 1972, and it currently covers 195 nations and 15 disputed territories (Freedom House, 2015). For 2015, Freedom House adopted a new scheme for its two variables which they believe provided more nuanced information than the older 7-point scoring system; Freedom House now scores *Political Rights* on a 40-point scale, and *Civil Liberties* on a 60-point scale (Freedom House, 2015).

As in the Boekestein study, this analysis combines *Political Rights* and *Civil Liberties* to create the variable *Freedom Score* by taking the average of the normalized scores for each nation-year instance. Scores were normalized to remove bias attributed to having an uneven dual scoring system utilized by Freedom House. The derivation of the Freedom Score is provided in Equations 2, 3, and 4.

$$nPr_{ij} = \frac{Pr_{ij}}{40}$$

Equation 2: Normalized Political Rights

$$nCl_{ij} = \frac{Cl_{ij}}{60}$$

Equation 3: Normalized Civil Liberties

$$FS_{ij} = \frac{nPr_{ij} + nCl_{ij}}{2}$$

Equation 4: Freedom Score

Where:

FS_{ij} = Freedom score for country i in year j
Pr_{ij} = Political rights score for country i in year j
$n\mathrm{Pr}_{ij}$ = Normalized political rights score for country i in year j
Cl_{ij} = Civil liberties score for country i in year j
nCl_{ij} = Normalized civil liberties score for country i in year j
i = Country $\in$ {1, 2, ..., 182}
j = Years $\in$ {1996, 1997,..., 2014}

Conflict and Freedom Trend Variables

Trend variables seek to predict conflict transitions through modeling the change in trajectory of a specific nation's conflict intensity levels and freedom scores. Previous conflict prediction studies have successfully employed trend variables as indicators of instability. Due to the nature of their calculations, all trend variables experience a one year lag in the model.

Change in HIIK conflict intensity is modeled as a two-year trend variable dividing the change in HIIK intensity levels for the years in question by the number of intensity levels, as shown in Equation 5. The objective of this variable is the improvement of conflict transition forecasting through the forecasting of increased or decreased levels of violence.

$$2YCIT_{i,j} = \frac{HIL_{i,j-1} - HIL_{i,j-2}}{6}$$

Equation 5: Two Year HIIK Trend Variable

Where:

$2YCIT_{i,j}$ = Two year conflict intensity trend for country i in year j
$HIL_{i,j}$ = HIIK intensity level for country i in year j
i = Country $\in$ {1, 2, ..., 182}
j = Years $\in$ {1996, 1997,..., 2014}

Like the HIIK conflict intensity trend variable, the two-, three-, and five-year freedom trends also seek to forecast conflict transitions through the modeling of a nation's Polity functions. In addition to the two-year trend variable, three- and five-year variables are also included to improve forecasting over longer time horizons as shown in Equation 6.

$$2YFT_{i,j} = FS_{i,j-2} - FS_{i,j-1}$$
$$3YFT_{i,j} = FS_{i,j-3} - FS_{i,j-1}$$
$$5YFT_{i,j} = FS_{i,j-5} - FS_{i,j-1}$$

Equation 6: Two-, Three-, and Five-Year Freedom Trend Variables

Where:

$2YFT_{i,j}$ = Two-year freedom trend for country i in year j
$3YFT_{i,j}$ = Three-year freedom trend for country i in year j
$5YFT_{i,j}$ = Five-year freedom trend for country i in year j
FS_{ij} = Freedom score for country i in year j
i = Country $\in$ {1, 2, ..., 182}
j = Years $\in$ {1996, 1997,..., 2014}

Database Design and Construction

Data Base Criteria

The design of the Conditional Conflict Database (CCD) facilitates the eventual construction of the conditional logistic regression and Markov models and consists of two sub-databases: the "In Conflict" and "Not in Conflict". The "In Conflict" database includes all instances of nations transitioning from a state of conflict (either remaining in conflict or transitioning out of conflict), while the "Not in Conflict" database includes all instances of transitioning from a state of non-conflict. The CCD meets three design criteria essential for the development of studies using logistic regression and Markov models: 1 – Common nomenclature and time frame across all datasets; 2 – Automated raw-data refreshment; and 3 – Easily searchable/sortable by nation, year-group, region, etc. The objective of the database design is the creation of six region specific databases which are used to develop the conditional logistic regression models.

Issues

The primary obstacle in the creation of the master database was the sorting, cataloguing, and formatting of the over *30* disparate databases that are loaded into the CCD. Between all datasets there exist *338* separate entries for nations, regions, and territories (NRT), of which only *182* are considered in this study. Additionally, a transliteration system was developed to ensure a common naming convention for all 338 NRTs, in addition to the unique catalogue numbers (*1* through *338*) assigned to each entity. A uniform database structure based on that used by the World Bank is employed to format the raw databases and segregate the "top" *182* nations-of-interest, creating the usable structures which are loaded into the CCD.

The master database requires a total *78,078* separate entries to properly catalogue the *2,002* separate nation-year instances included in this study. Manual database updates are cumbersome, time-consuming, and prone to human error. To overcome this obstacle, a Microsoft Office visual basic (VBA) based consolidated database tool was developed to compile the *39* separate identifying-information and data spreadsheets into one consolidated file, which is subsequently time-stamped with the most recent compile date. This tool enables timely and error free data updates of the CCD for any dataset conforming to the World Bank format.

Data Imputation

As shown in Table 6, the raw datasets employed for this study had numerous instances of missing data. Since the study considers 182 of the world's nations, data-year sets containing less than 182 data points require the data imputation prior to final consolidation in the CCD. In general, nations with fledgling or unstable governments lack the ability to track and consolidate the large amounts of statistical data required for this study. For the data considered in this study, a total of *1,602* or *80%* of the nation-year instances had between *28* and *30* of the *30* possible variables, with the average of 28.5 variables per nation-year instance. However, within the considered dataset, there exist *32* nation-year instances that have less than *23* of the *30* possible variables; the complete list provided in Table 9.

Table 9: Number of Variables per Nation-year Instance; Worst Data

Index	Country	Year	Total Data Sets
149	South Sudan	2009	13
149	South Sudan	2010	13
149	South Sudan	2011	13
149	South Sudan	2012	13
149	South Sudan	2006	16
149	South Sudan	2007	16
149	South Sudan	2008	16
109	Montenegro	2005	18
109	Montenegro	2006	18
149	South Sudan	2005	18
107	Micronesia, Federated States of	2012	20
109	Montenegro	2004	20
107	Micronesia, Federated States of	2011	21
149	South Sudan	2004	21
107	Micronesia, Federated States of	2014	22
107	Micronesia, Federated States of	2013	22
107	Micronesia, Federated States of	2004	22
107	Micronesia, Federated States of	2005	22
107	Micronesia, Federated States of	2006	22
136	Samoa	2014	22
176	Vanuatu	2014	22
179	West Bank	2004	22
179	West Bank	2005	22
179	West Bank	2006	22
179	West Bank	2007	22
179	West Bank	2008	22
179	West Bank	2009	22
179	West Bank	2010	22
179	West Bank	2011	22
179	West Bank	2012	22
179	West Bank	2013	22
179	West Bank	2014	22

A total of 2,903 of the 62,062 statistical data points required imputation prior to final consolidation in the CCD. The JMP statistical software package was employed to impute the missing data. JMP imputes missing data points by analyzing values in other columns and rows, developing an estimate of the missing value(s) (Hinrichs & Boiler, 2010). Imputed values are expectations conditioned on the non-missing values of each row in the data set (SAS Institute, 2015). Two separate data imputation methods, isolated variable and holistic imputation (using entire data set), were conducted and compared to identify the optimal variables to import into the master CCD. The final imputation method selection was based on the statistical similarity (average) of the imputed data to the raw data for nation-year instances within the same region. Additionally, the imputation of the "Polity IV" and "Government Type" data was based off of the "Regime Type" variable. Table 10 provides the list of variables requiring data imputation as well as the method of imputation employed.

Table 10: Variables Requiring Data Imputation

Variable Name	Imputation Method
Arable Land	Holistic Imputation
GDP Per Capitia	Isolated Variable
Improved Water	Holistic Imputation
Military Expend (% Gov Spending)	Holistic Imputation
Military Expend (% GDP)	Holistic Imputation
Population density	Isolated Variable
Population Growth	Holistic Imputation
Refugee (Asylum)	Isolated Variable
Refugee (Origin)	Isolated Variable
Fresh Water per Capita	Holistic Imputation
Trade (% GDP)	Holistic Imputation
Unemployment	Holistic Imputation
Polity IV	Based off regime type
Government Type	Based off regime type
Caloric Intake	Holistic Imputation
Freedom Score	Holistic Imputation
2 Yr Freedom Trend	Holistic Imputation
3 Yr Freedom Trend	Holistic Imputation
5 Yr Freedom Trend	Holistic Imputation
Religious Diversity	Holistic Imputation

Conditional Conflict Database Structure

The "In Conflict" and "Not in Conflict" CCDs share a common database structure that includes the catalogue number, standard name and code, the base year, transition year-pair, the year code, supporting HIIK data, region, and all statistical data from *2004* to *2014*. The database also provides summary statistics concerning the total instances and transitions of interest included within the dataset. An example of the CCD structure is provided in Figure 3.

Left Node: Conflict - Conflict, Conlict - No Conflict
Total Instances: 731
Total Transitions (out of conflict): 111
Date Created: 11/16/2015 13:05

Index	Country	Country Code	Year	Year Pair	Year Code	States In Conflict	HIIK Conflict Intensity	Region	Geographic COCOM	Full Category	Transition to Conflict (Y)	Arable Land	Birth Rate
1	Afghanistan	AFG	2004	2004-2005	AFG2004	1	4	Eastern Europe and Central Asia	CENTCOM	ConCon	1	0.345812	47.403
1	Afghanistan	AFG	2005	2005-2006	AFG2005	1	4	Eastern Europe and Central Asia	CENTCOM	ConCon	1	0.336215	46.291
1	Afghanistan	AFG	2006	2006-2007	AFG2006	1	5	Eastern Europe and Central Asia	CENTCOM	ConCon	1	0.325413	45.078
1	Afghanistan	AFG	2007	2007-2008	AFG2007	1	5	Eastern Europe and Central Asia	CENTCOM	ConCon	1	0.313947	43.763
1	Afghanistan	AFG	2008	2008-2009	AFG2008	1	5	Eastern Europe and Central Asia	CENTCOM	ConCon	1	0.304082	42.361
1	Afghanistan	AFG	2009	2009-2010	AFG2009	1	5	Eastern Europe and Central Asia	CENTCOM	ConCon	1	0.295796	40.899
1	Afghanistan	AFG	2010	2010-2011	AFG2010	1	5	Eastern Europe and Central Asia	CENTCOM	ConCon	1	0.288323	39.414
1	Afghanistan	AFG	2011	2011-2012	AFG2011	1	5	Eastern Europe and Central Asia	CENTCOM	ConCon	1	0.281253	37.952
1	Afghanistan	AFG	2012	2012-2013	AFG2012	1	5	Eastern Europe and Central Asia	CENTCOM	ConCon	1	0.274387	36.556
1	Afghanistan	AFG	2013	2013-2014	AFG2013	1	5	Eastern Europe and Central Asia	CENTCOM	ConCon	1	0.267682	35.254

Figure 3: In Conflict Database

Creation of the CCD requires that all instances of conflict transition, from one year to another, are identified and catalogued according to whether a transition from their current state occurred for the preceding year. This formulation results in the possibility that data specific nation-year transition instances may be included in both the "In Conflict" and "Not in Conflict" databases.

Regional Assignments

The practice of creating region-specific conflict prediction models has been employed in several previous studies. These studies have shown a relationship between the duration and scope of violent conflict and the significance of regional commonalities such as the incidence of natural resources, physical geography, adjacent border conflicts, and population demographics (Buhag, 2005). Additionally, it has been shown that conflict risk factors such as poverty, famine, and despotism tend to cluster in so-called

“Bad Neighborhoods”, with an observable cross-border effect (Hegre, Karlsen, Nygard, Strand, & Urdal, 2011). The regional assignments utilized in this study are based on the six-region world model developed in the Boekestein model, and are comprised of: Sub-Sahara Africa, South and East Asia, Eastern Europe and Central Asia, Arab and North African States, Latin America, and the Organization of Economic Cooperation and Development (OECD) nations (Boekestein, 2015). The number of nations assigned to a specific region ranges from 17 (Arab & North African states) to 49 (Sub-Sahara Africa). The regional assignments for the 182 nations considered in this study are provided in Appendix A.

3.3 Logistic Regression

Overview of Logistic Regression Concepts and Theory

Logistic regression was employed as the regression method for this study due to the binary response of the conditional dependent variable, where a nation given its current status either “Transitions / Remains in Conflict” or “Transitions / Remains out of Conflict”. As in any regression model, the goal of this analysis is to construct the best fitting, parsimonious, and operationally interpretable model to describe the relationship between the dependent and independent variables (Hosmer, Lemeshow, & Sturdivant, 2013). Linear regression is not used since the dichotomous nature of the data used in this study violates many of the assumptions required for linear regression specifically those of **measurement** (dependent variable is continuous and unbounded), **homoscedasticity** (constant residual variance over regressor hull), and **normality** (residuals are normally distributed) (Hosmer, Lemeshow, & Sturdivant, 2013). The measurement assumption is violated through the use of the dichotomous variable that is constrained to 0 or 1. This in turn violates the normality assumption of the distribution of errors, which themselves can only assume values of 0 or 1. Finally, the homoscedasticity assumption is violated due to non-constant variance of the error terms associated with each instance.

In logistic regression, the conditional mean of the dichotomous response is bounded between 0 and 1, or simply $0 \leq E(Y|x) \leq 1$. This results in the non-constant variance discussed previously as the response approaches 0 or 1 producing the “S-curve” shown in Figure 4. The curve itself resembles the plot of a continuous distribution of a random variable, leading to the use of the logistic distribution to model the conditional mean for a dichotomous response (Hosmer, Lemeshow, & Sturdivant, 2013).

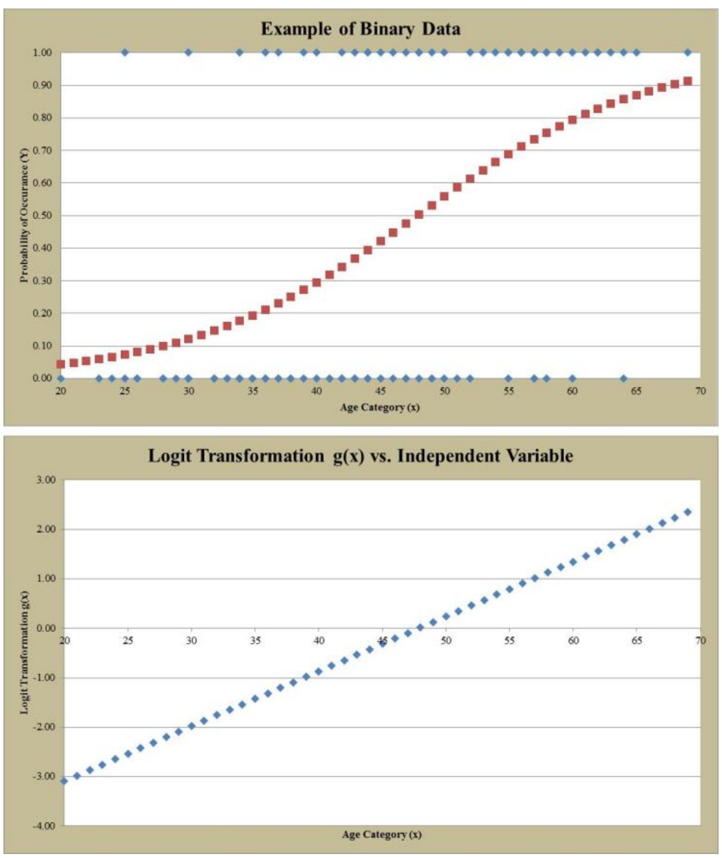


Figure 4: Plots of the Logit π(x) and Logit Transformation g(x)

The Logit and Logit Transformation

While other continuous distributions can adequately model dichotomous data, the logistic distribution provides superior mathematical flexibility in conjunction with operationally meaningful estimates of the covariate effects. For the purposes of this study the conditional mean will be represented as $\pi(x) = E(Y|x)$, where the logit $\pi(x)$ represents the probability of the response is equal to 1, or for the purposes of this study a "Transition into Conflict," given the covariate(s). The specific form of the logistic model is given in Equation 7.

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x+\cdots+\beta_n x}}{1+e^{\beta_0+\beta_1 x+\cdots+\beta_n x}} = \frac{e^{g(x)}}{1+e^{g(x)}}$$

Equation 7: The General Logistic Regression Model

The general logistic regression model effectively ensures that the probability estimate of conflict transition is bounded between 0 and 1. The error associated with the model assumes a binomial distribution with an expected value given by $E(\varepsilon|Y=1) = 1-\pi(x)$ with a probability of $\pi(x)$, or $E(\varepsilon|Y=0) = -\pi(x)$ with a probability of $1-\pi(x)$. These properties of the error term result in the binomial distribution with the properties of $E(\varepsilon|Y) = 0$ and $Var(\varepsilon|Y) = \pi(x)[1-\pi(x)]$ (Hosmer, Lemeshow, & Sturdivant, 2013).

Central to the development of this study's logistic regression models is the concept of the logit transformation *g(x)*. The logit encompasses many of the desirable properties of the linear regression model such as a continuous, unbounded response that is linear within its parameters as shown in Figure 4 (Hosmer, Lemeshow, & Sturdivant, 2013). The logit transformation is calculated by taking the natural logarithm of the odds ratio: $\pi(x)/[1-\pi(x)]$ presented in Equation 8.

$$g(x) = ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x + \cdots + \beta_n x$$

Equation 8: The Logit Transformation

The covariate parameters β_i are estimated through the method of maximum likelihood which seeks to determine the estimates of the covariate parameters that agree most closely with the observed data of the response (Hosmer, Lemeshow, & Sturdivant, 2013). As with the error terms, each sample observation follows a binomial distribution with the likelihood function given by Equation 9.

$$l(\vec{\beta}) = \prod_{i=1}^{n} \pi(x_i)^{y_i}[1-\pi(x_i)]^{1-y_i}$$

Equation 9: Likelihood Function

Where:

$\vec{\beta}$ = $(\beta_0, \beta_1, \ldots, \beta_n)$
$\pi(x_i)$ = ith response probability
y_i = ith response obsetvation

The principal of maximum likelihood simply seeks to maximize the expression provided in Equation 10. However, the use of the Log-likelihood function provided in Equation 11 provides a simpler means of estimating the covariate parameters (Hosmer, Lemeshow, & Sturdivant, 2013).

$$L(\vec{\beta}) = \sum_{i=1}^{n} \{y_i ln[\pi(x_i)] + (1 - y_i) ln[1 - \pi(x_i)]\}$$

Equation 10: Log-likelihood Function

Testing for Model and Coefficient Significance

As in linear regression, the basic premise for determining the significance of any logistic regression model is comparing the model containing the covariates of interest to the model without those parameters via hypothesis testing. The comparison method in logistic regression is the likelihood ratio test, which assumes a Chi-square (χ^2) distribution. In order to conduct the hypothesis tests using the likelihood ratios, we must calculate the deviance in the likelihood values of the saturated and fitted models. The deviance (*D*) statistic is shown in Equation 11.

$$D = -2 \sum_{i=1}^{n} \left\{ y_i ln \left[\frac{\pi(x_i)}{y_i} \right] + (1 - y_i) ln \left[\frac{1 - \pi(x_i)}{1 - y_i} \right] \right\}$$

Equation 11: Deviance of the Saturated and Fitted Models

Given that the likelihood $l(\vec{\beta})$ of the saturated model (i.e. the model containing the entire set of variables) is equal to 1.0, it follows that the deviance is equal to $D = -2ln[likelihood\ of\ the\ fitted\ model]$. It should be noted that the deviance statistic has the same function in logistic regression as the residual sum-of-squares (SSE) does in linear regression (Hosmer, Lemeshow, & Sturdivant, 2013).

To assess the significance of the covariate in question, the statistic *G*, the negative two log ratio of the deviance statistics, with and without the variable in question, is calculated. The statistic *G* has the same function in logistic regression as the numerator of the partial F-test does in linear regression (Hosmer, Lemeshow, & Sturdivant, 2013). The statistic G, which assumes a Chi-square distribution, can be calculated as either the ratio of likelihoods between the different models as shown in Equation 12, or as the differences between the deviances of the two models as shown in Equation 13.

$$G = -2ln\left[\frac{Likelihood\ without\ variable}{Likelihood\ with\ variable}\right]$$

Equation 12: Likelihood Ratio Method

$$G = D(Model\ without\ variable) - D(Model\ with\ variable)$$

Equation 13: Difference in deviances method

Figure 5 presents a likelihood ratio hypothesis test using JMP software output. Model significance for a given confidence level (1-α)% is determined through a standard hypothesis test wherein the null hypothesis (H_0), the intercept only model is sufficient is tested against the alternate hypothesis (H_a), the reduced model is equivalent to the full model. The *G* statistic is compared against the Chi-Square test statistic $\chi^2_{(1-\alpha,\ n)}$, for a given confidence level and *n* degrees of freedom, the difference in the number of variables between the two models. In this example the null hypothesis is rejected if $G > 7.815$.

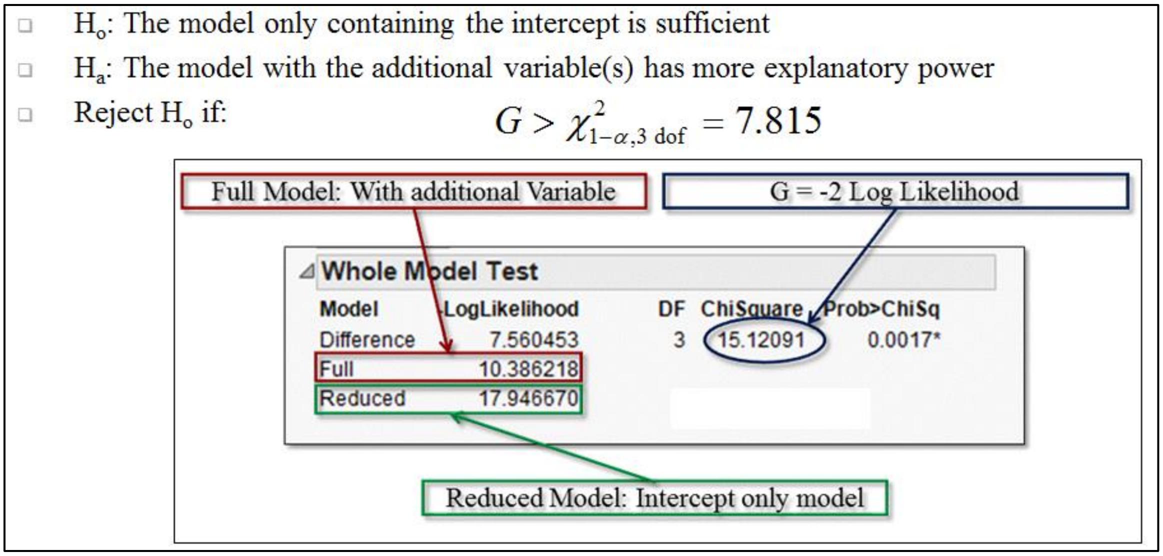


Figure 5: Hypothesis Test for Model Significance

Variable significance for a given confidence level (1-α)% is determined through a standard hypothesis test similar to that discussed previously. In this case, the Wald statistic (*W*), which follows a Chi-square distribution with one degree of freedom, is used as the test statistic. The null hypothesis (H_0), the variable does not significantly contribute to the model, is tested against the alternate hypothesis (H_1), the variable significantly contributes to the model. The *W* statistic is compared against the Chi-Square test statistic $\chi^2_{(1-\alpha,\ 1)}$, for a given confidence level and one degree of freedom. In this test the null hypothesis is rejected if $G > \chi^2_{(1-\alpha,1)}$. Figure 6 provides an example of such a test. In this case, "Refugee Asylum" is identified as a significant variable, while the intercept, "Trade (% GDP)", and "Religious Diversity" fail the test for a 0.05 level of significance.

- H_o: The variable does not significantly contribute to the model
- H_a: The variable significantly contributes to the model
- Reject H_o if: $W > \chi^2_{1-\alpha,1\ \text{dof}} = 3.841$

W = Wald Test Statistic

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-25.24666	13.925276	3.29	0.0698
Refugee (Asylum)	2.13222e-5	1.0112e-5	4.45	0.0350*
Trade (% GDP)	0.08733891	0.0489827	3.18	0.0746
Religious Diversity	21.5581165	13.449431	2.57	0.1090

Figure 6: Hypothesis Test for Covariate Significance

Model Building Strategies

The Purposeful Selection of Covariates method is the model building strategy employed throughout this study. The strategy entails a seven step, iterative process that individually analyzes each of the independent variables, fits and analyzes a preliminary effects model, assesses covariate interaction, and assesses the fit and adequacy of the main effects model (Hosmer, Lemeshow, & Sturdivant, 2013). The methodology guiding the purposeful selection of covariates is provided in Table 11.

Table 11: Purposeful selection of Covariates Methodology

Step 1	**Univariate assessment of all candidate variables**	Candidate variables for the first multivariable model are selected based on the univariate test p-value. Include if $p \leq 0.25$
Step 2	**Fit a multivariable model containing all covariates identified in step 1**	Assess the importance of each covariate using the its p-value analyzed at traditional levels. Eliminate all vaiariables (one at time), that do not significantly contribute to the model.
Step 3	**Assessment of initial covariate estimates**	Compare the values of the estimated coefficients in the reduced model, built during step 2, to the original model identified in step 1. Identify any variable whose $\Delta\beta \geq 20\%$, as this indicated one or more excluded variables are important in providing adjustment to effect of the variable in question, and should be added back into the model.
Step 4	**Add each variable not selected in step 1 to model obtained at the conclusion of step 3.**	Variables are added one at a time, checking for variable significance using the p-value. The final model produced in step 4 is referred to as the preliminary main effects model.
Step 5	**Construct the Main Effects Model.**	For each continuous variable, check the assumption of logit linearity as a function of the covariate.
Step 6	**Check for covariate interaction within the Main Effects Model.**	Create a list of possible pairs of variables that have a realistic possibility of interacting. This can include the various levels of categorical variables. Interaction terms are added and tested one at a time for significance in univariable model. Significant interaction terms are added to the Main Effects Model.
Step 7	**Assess model Adequacy and Fit.**	Assess model adequacy using the Hosmer-Lemsow Goodness of Fit test, analysis of classification tables, and the receiver operating charactersitic curve.

The first step entails fitting separate univariate logistic regression models for each variable. The significance of each variable is assessed based on the standard Chi-square test. Candidate variables for the initial multivariate model are screened and selected based on p-values less or equal to 0.25. This relaxed selection criteria allows for the inclusion of possibly significant variables that may not have been included in the model otherwise (Hosmer, Lemeshow, & Sturdivant, 2013).

The second and third steps involve the fitting of the initial multivariate model containing all the variables identified in the first step, assessing model and covariate significance, followed by a systematic removal, one variable at a time, and analysis of variables based upon the studies desired significance level ($p = 0.05$, was the standard significance level employed throughout this study). As part of the systematic analysis of variables in the second and third steps, a comparison of the coefficient values of the variables remaining in the model, prior to and following the removal of a variable, was conducted. If the change in coefficient value ($\Delta\beta_i$) for any variable was greater than ±20%, it indicated the possible importance of the removed variable within the model; the variable is subsequently added back into the model on the next iteration.

During the fourth step, variables initially excluded from consideration, are systematically added back into the model and tested for significance creating the preliminary effects model in the fourth step. In the fifth step, each covariate within the preliminary effects model is checked for logit linearity. If a covariate is found to behave in a nonlinear fashion, appropriate transformations are applied and tested. During the sixth step, covariate is tested for significance within the final model. Interaction between two variables implies that the effect of each variable is not constant over the levels of the other variable (Hosmer, Lemeshow, & Sturdivant, 2013). Ultimately, the final decision to include interaction terms in the main effects model must be based on statistical significance of the interaction term, and practical considerations such as whether the interaction term improves the model and whether it operationally relevant. Following the addition of significant interaction terms to the preliminary effects model, the systematic model reduction of variables described in the second step is repeated with the coefficients of the main effects locked. The model constructed at the end of the sixth step is known as the main effects model (Hosmer, Lemeshow, & Sturdivant, 2013).

Assessing Model Fit and Adequacy

Assessing the fit and adequacy of model fit is the seventh and final step of the purposeful selection of covariates method. However, further discussion of the various methods employed in this steps warrant a separate section within this chapter. Three methods are employed in concert to provide a holistic assessment of the fit and adequacy of the conditional logistic regression models in this study, those methods were: The Hosmer-Lemeshow Goodness of Fit Test, classification tables, and the area under the curve for model-specific receiver operating characteristic curves.

The Hosmer-Lemeshow goodness of fit assesses the overall fit of probability $\pi(x_i)$ based population sub-groups, through the use of the Hosmer-Lemeshow statistic $\hat{C}$. Two grouping strategies are generally employed; the first is based on the percentiles of the estimated probabilities, and the second is based on the actual fixed values of the same probabilities (Hosmer, Lemeshow, & Sturdivant, 2013). In general the population is broken into 10 sub-groups (g), but more or fewer can be used depending on the data set. The squared differences between the expected and observed observations, for both success “1” and failure “0” responses for each sub-group are calculated added. The summation of the sub-group specific statistics is known as the Hosmer-Lemeshow goodness of fit statistic ($\hat{C}$) and is presented in its entirety in Equation 14 (Hosmer, Lemeshow, & Sturdivant, 2013).

$$\hat{C} = \sum_{k=1}^{g} \left[\frac{\left(o_{1k} - \hat{e}_{1k}\right)^2}{\hat{e}_{1k}} + \frac{\left(o_{0k} - \hat{e}_{0k}\right)^2}{\hat{e}_{0k}} \right]$$

$$o_{1k} = \sum_{j=1}^{c_k} y_j$$

$$o_{0k} = \sum_{j=1}^{c_k} \left(m_j - y_j\right)$$

$$e_{1k} = \sum_{j=1}^{c_k} m_j \hat{\pi}_j$$

$$e_{0k} = \sum_{j=1}^{c_k} m_j \left(1 - \hat{\pi}_j\right)$$

$$\bar{\pi}_k = \frac{1}{n_k} \sum_{j=1}^{c_k} m_j \hat{\pi}_j$$

Equation 14: Hosmer-Lemeshow Test Statistic

Where:

g = Number of sub-groups
o_{1k} = Number of "1" or success observations within the kth sub-group
$\hat{e}_{1k}$ = Number of "1" or expected successes within the kth sub-group
o_{0k} = Number of "0" or failure observations within the kth sub-group
$\hat{e}_{0k}$ = Number of "0" or expected failures within the kth sub-group
$\bar{\pi}_k$ = The average estimated probability in the kth sub-group

Like other logistic regression test statistics, $\hat{C}$ follows a Chi-square distribution with given significance level (α) and $g - 2$ degrees of freedom, where g is the number of sub groups employed with within the goodness of fit test. Model fit is also assessed through a standard hypothesis test where the null hypothesis (H_0), there is evidence of model fit, is tested against the alternate hypothesis (H_1), there is little evidence of model fit. The $\hat{C}$ statistic is compared against the Chi-Square test statistic $\chi^2_{(1-\alpha,\ g-2)}$, for a given confidence level and one degree of freedom (the difference in the number of variables between the two models). In this test the null hypothesis is accepted if $\chi^2_{(1-\alpha,\ g-2)} > \hat{C}$.

An example of such the Hosmer-Lemeshow hypothesis test for $\alpha = 0.05$ is shown in Figure 7.

H_o:	**Model appears fit data**	
H_a:	**There is little evidence of model fit**	
Decision Rule: Reject if T.S. > C		
C =	**0.236**	**Test Result**
T.S. =	**3.841**	**Fail to Reject: Model Apears to fit the data well.**
P{T.S. > C}	**0.627**	

Figure 7: Hosmer-Lemeshow Hypothesis Test

Classifications tables, or confusion matrices, gauge the adequacy of logistic regression models through the depiction of the total number of true-positive, false-positive, true-negative, and false-negative responses as they relate to the total population. True-positive and true-negative are referenced as model sensitivity and model specificity respectively. These tables are the result of cross-classifying the dichotomous response variable, with the value of the outcome variable $\pi(x_i)$ (Hosmer, Lemeshow, & Sturdivant, 2013). The cross classification is dependent on a probability cut-point which assigns values that fall below the cut-point to the "0" or failure response, and values greater that the cut-point to the "1" or success response; in general the cut-point is initially set at 0.5. An example of a standard classification table is presented in Table 12.

Table 12: Classification Table

Standard Classification Table			
	Observed		
Classified	Transition to Conflict = 1	Remain/Transition out of Conflict = 0	Total
Transition to Conflict = 1	5	1	6
Remain/Transition out of Conflict = 0	4	116	120
Total	9	117	126

Med Cut Point:	0.50
Model Acuracy:	0.960

As can be observed in Table 12, the model sensitivity or true-positive ($y_i = 1|Y_i = 1$) rate is given by the five correctly predicted "1" responses out of a total of nine occurrences. Additionally, the model specificity or true-negative ($y_i = 0|Y_i = 0$) is given by the 116 correctly classified "0" responses out of a total of 117 occurrences, given a cut-point equal to 0.50. The overall model accuracy is gauged by the overall proportion of "true" responses to the total number of observations, and is provided in Equation 15.

$$Model\ Accuracy = \frac{\sum True\ Positive\ Obs + \sum True\ Negative\ Obs}{\sum Observations}$$

Equation 15: Logistic Regression Model Accuracy

The final method used to gauge the overall adequacy of the logistic regression model is the total area under the curve (AUC) for receiver operating characteristic (ROC) curves. Unlike classification tables which depend on a single cut-point, ROC curves provide a better and more comprehensive description of model adequacy over the entire range of model responses (Hosmer, Lemeshow, & Sturdivant, 2013). The ROC curve, whose use originates from signal theory, provides a means to measure the model's (receiver) ability to detect true responses (signal) in the presence of noise. The graph of the ROC curve plots the model's probability of detecting a true signal (sensitivity) as a function of the probability of detecting a false signal (1 – specificity) over the entire range of cut-point values as shown in Figure 8 (Hosmer, Lemeshow, & Sturdivant, 2013).

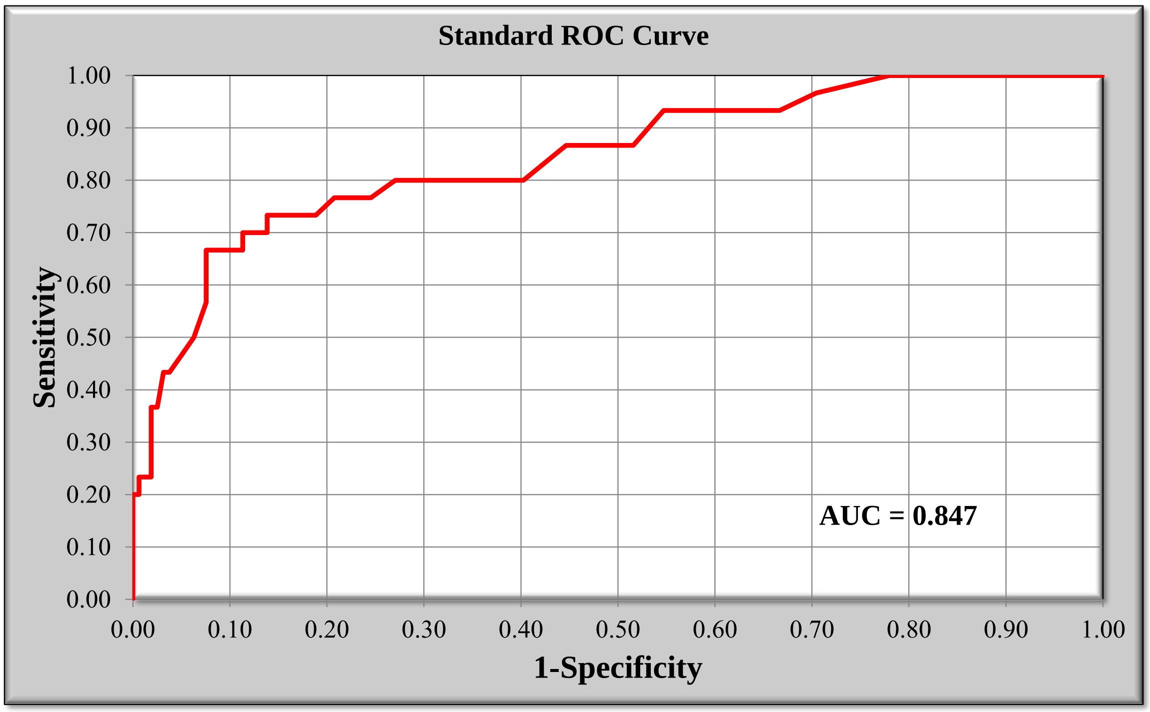


Figure 8: Receiver Operating Characteristic Curve

The ROC area under the curve ranges from 0.5 to 1.0 and provides a measure of the model's ability to effectively discriminate between observations experiencing the outcome of interest versus those who do not (Hosmer, Lemeshow, & Sturdivant, 2013). Model's with low AUC values nearing 0.50, are said to have little to no discrimination capacity, or that the model provides little predictive benefit over that of a coin toss. While there is not set standard for gauging the adequacy of model discrimination, the criteria provided in Table 13 set the guidelines for logistic regression model analysis employed in this study.

Table 13: Discrimination Measures
(Hosmer, Lemeshow, & Sturdivant, 2013)

AUC	General Guidelines
AUC = 0.5	No Model Discrimination
$0.5 < AUC \leq 0.7$	Poor Discrimination
$0.7 < AUC \leq 0.8$	Acceptable Discrimination
$0.8 < AUC \leq 0.9$	Excellent Discrimination
$AUC > 0.9$	Superior Discrimination

Interpretation of the Logistic Regression Model

The odds ratio can be used to approximate another measure known as the relative risk, which is the ratio of outcome probabilities (Hosmer, Lemeshow, & Sturdivant, 2013). The concept of relative risk can be related to the classification table, and model sensitivity/specificity as shown in Figure 9.

Standard Classification Table		
	Observed	
Classified	**Transition to Conflict = 1**	**Remain/Transition out of Conflict = 0**
Transition to Conflict = 1	$e^{\beta_0+\beta_1}/_{1+e^{\beta_0+\beta_1}}$	$e^{\beta_0}/_{1+e^{\beta_0}}$
Remain/Transition out of Conflict = 0	$\left(1/_{1+e^{\beta_0+\beta_1}}\right)$	$\left(1/_{1+e^{\beta_0}}\right)$
Total	**1**	**1**

Figure 9: Relative Risk Relation to Classification Table

Interpretation of the effects of significant covariates, as they relate to conflict transition is central to the research questions and operational relevancy of this study. For this purpose, logistic regression analysis employs the concept of the odds ratio (OR), which is a measure of association that approximates the likelihood for the dichotomous response given a certain covariate remains in the model. Equation 16 provides the derivation of the univariate odds ratio.

$$OR = \frac{\left[\dfrac{\left(e^{\beta_0+\beta_1} / 1 + e^{\beta_0+\beta_1} \right)}{\left(1 / 1 + e^{\beta_0+\beta_1} \right)}\right]}{\left[\dfrac{\left(e^{\beta_0} / 1 + e^{\beta_0} \right)}{\left(1 / 1 + e^{\beta_0} \right)}\right]} = e^{\beta_1}$$

Equation 16: Univariate Odds Ratio

The basic interpretation of the odds ratio is illustrated in the following example. If a certain model has an odds ratio of 2, it can be said that the odds of experiencing the outcome of interest, given the certain covariate effect is present, is 2 to 1. Conversely if the odds ratio is 0.5, it can be said that odds of experiencing the outcome of interest is half of that when a certain covariate effect is present. The mechanics and interpretation of multivariate model odds ratios is very similar to that of the univariate method, which is presented to demonstrate the basic premises of the concept.

Overview of JMP Software and Output

JMP is a statistical software package developed by the SAS Corporation that is employed to construct and analyze the logistic regression models for this study. For this reason, a brief discussion of the JMP model output is warranted. Figure 10 displays the JMP whole model test for significance. The JMP output displays the log-likelihood values for both the Full and Reduced models, as well as the Chi-Square distributed *G* statistic. This interface enables the analyst to quickly ascertain the significance of the overall model. This is done through visual inspection of the p-value given as "Prob>ChiSq" in the JMP interface. For the purposes of this study the threshold for model significance was set at for p-values $\leq$ 0.05; the JMP default threshold of significance indicated by "*" (Hinrichs & Boiler, 2010).

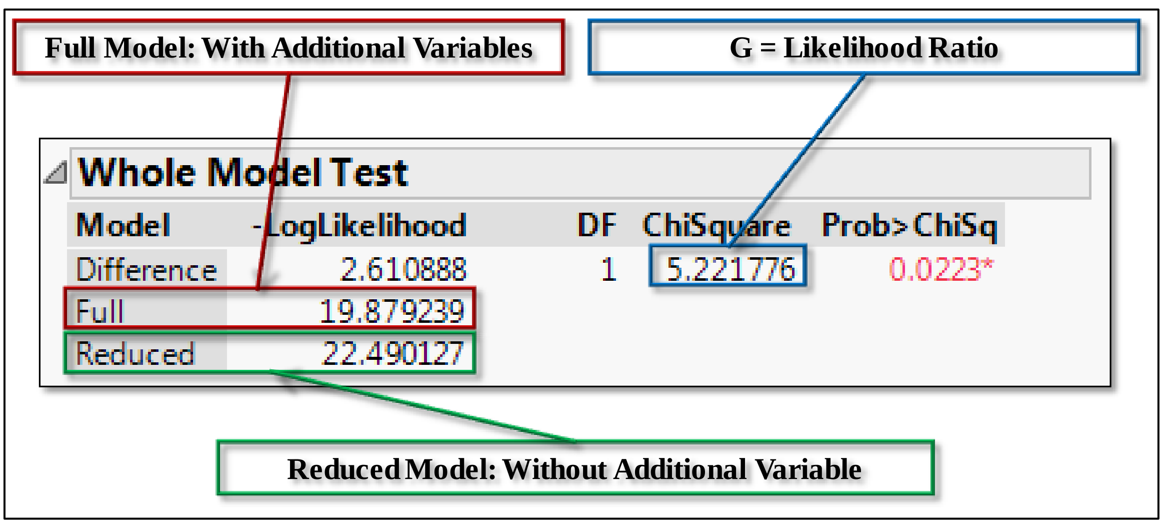


Figure 10: JMP Whole Model Test

The JMP environment also provides estimates of coefficients combined with the overall significance of the variables included in the model. As with the whole model test, the threshold for covariate significance was set at p-values $\leq$ 0.05. The JMP estimates of the covariate coefficients (β_i) are provided in the first column of Figure 11. It should be observed that these estimates have the opposite sign when included in the final model. For example, the logit transformation for the Figure 11 estimates is given as: $g(x) = 25.247 - 2.12 \times 10^{-5}x_1 - 0.087x_2 - 21.558x_3$.

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-25.24666	13.925276	3.29	0.0698
Refugee (Asylum)	2.13222e-5	1.0112e-5	4.45	0.0350*
Trade (% GDP)	0.08733891	0.0489827	3.18	0.0746
Religious Diversity	21.5581165	13.449431	2.57	0.1090

Figure 11: JMP Parameter Estimates

Synthetic Minority Over-sampling Technique

The nature of predicting conflict transitions, which are decidedly rare events, results in significantly unbalanced conditional conflict data sets; a data set is said to be unbalanced if the classification categories are not approximately equally represented (Chawla et al., 2002). Due to the method of maximum likelihood, which is employed to estimate the covariate coefficients, the imbalance of the data set will favor the observation response, success or failure that forms the majority of the population responses. This results in the tendency to misclassify the observations of interests, i.e., conflict transitions, in favor of the majority response, no transition from current status. To compensate for this phenomenon, the Synthetic Minority Over-sampling Technique (SMOTE) was utilized to enable development of the conditional logistic regression model for a single specific region, which experienced significant issues with misclassification.

The SMOTE methodology conducts an over-sampling of the minority class through the creation of "synthetic" observations (Chawla et al., 2002). The generation of synthetic observations is conducted in the feature space of the observations, through the creation of segments joining the k (in this study k = 5), nearest neighbors. The synthetic examples added to the original data set result in the creation of larger and less specific data regions, which allows better training of the minority dataset (Chawla et al., 2002). Synthetic data points utilized in this study were generated using a MATLAB sub-routine based on the pseudo-code provided in Figure 12.

```
Algorithm SMOTE(T, N, k)
Input: Number of minority class samples T; Amount of SMOTE N%; Number of nearest
    neighbors k
Output: (N/100) * T synthetic minority class samples
    (* If N is less than 100%, randomize the minority class samples as only a random
    percent of them will be SMOTEd. *)
    if N < 100
      then Randomize the T minority class samples
            T = (N/100) * T
            N = 100
    endif
    N = (int)(N/100) (* The amount of SMOTE is assumed to be in integral multiples of
    100. *)
    k = Number of nearest neighbors
    numattrs = Number of attributes
    Sample[ ][ ]: array for original minority class samples
    newindex: keeps a count of number of synthetic samples generated, initialized to 0
    Synthetic[ ][ ]: array for synthetic samples
    (* Compute k nearest neighbors for each minority class sample only. *)
    for i ← 1 to T
            Compute k nearest neighbors for i, and save the indices in the nnarray
            Populate(N, i, nnarray)
    endfor

    Populate(N, i, nnarray) (* Function to generate the synthetic samples. *)
    while N ≠ 0
            Choose a random number between 1 and k, call it nn. This step chooses one of
            the k nearest neighbors of i.
            for attr ← 1 to numattrs
                    Compute: dif = Sample[nnarray[nn]][attr] − Sample[i][attr]
                    Compute: gap = random number between 0 and 1
                    Synthetic[newindex][attr] = Sample[i][attr] + gap * dif
            endfor
            newindex++
            N = N − 1
    endwhile
    return (* End of Populate. *)
    End of Pseudo-Code.
```

Figure 12: SMOTE Pseudo-code

(Chawla et al., 2002)

Construction of Regional Logistic Regression Models

Model Dataset Overview

Six regional logistic regression models, consisting of two sub-models, conditioned on a nation's conflict status prior of the year of transition were developed for this study. The use of "In Conflict" and "Not in Conflict" conditional models is a requirement for the subsequent development of the nation specific Markov conflict transition models. The "In Conflict" models include all instances of nations in conflict for year $i-1$, that either remain in conflict or transition out of conflict in year i. Similarly, the "Not in Conflict" models include all instances of nations not in conflict for year $i-1$, that remain out of conflict or transition into conflict in year i. The data set utilized for the training and validation models covered the years *2004* to *2013*; data for year *2014* was reserved for Markov model development due to the lack of HIIK conflict data for year *2015*. The summary statistics for the regional model data is provided in Table 14.

Table 14: Summary Statistics of Regional Model Data

Regional Models	Sub-Saharan Africa		South and East Asia		Eastern Europe and Central Asia		Arab & North African States	
Statisitcs	In Conflict	Not In Conflict	In Conflict	Not In Conflict	In Conflict	Not In Conflict	In Conflict	Not In Conflict
Number of Cases	228	262	123	157	117	166	95	75
Number of Transitions	37	42	19	19	19	23	6	14
Transition rate (%)	16.2%	16.0%	15.4%	12.1%	16.2%	13.9%	6.3%	18.7%

Regional Models	Latin America		OECD		World View (Totals of Regions)	
Statisitcs	In Conflict	Not In Conflict	In Conflict	Not In Conflict	In Conflict	Not In Conflict
Number of Cases	95	174	75	255	733	1089
Number of Transitions	19	26	11	13	111	137
Transition rate (%)	20.0%	14.9%	14.7%	5.1%	15.1%	12.6%

On average, the "In Conflict" models experience transitions of interest (i.e. transitions out of conflict) in 15.1% of all cases, while the "Not in Conflict" models experience transitions into conflict in 12.6% of all cases. These average transition rates were instrumental in the identification of the training and validation data sets, which sought to maintain these rates for model development.

Design of Training and Validation Data Sets

The overarching concept guiding the selection of the training and validation data sets was to identify the data subsets, for each conditional model, that provided transition rates comparable to the “World View” averages shown in Table 14. In general, this principal was adhered to for every model except the Arab and North African “In Conflict” model and the OECD “Not in Conflict” model which experienced below average out-of-state transition rates of 6.3% and 5.1% respectively. Training and validation data sets were initially standardized across all models with year sets *2004* to *2010* specified for the training models, and year sets *2011* to *2013* specified for model validation. However, during the construction of the twelve conditional models, it became clear that a standardized year set across regions resulted in the development of sub-optimal conditional models. Continuous analysis and model refinement the construction of the conditional models resulted in the selection of the model specific data year sets provided in Table 15. The final selection of data year-sets is predicated on balancing the competing requirements of maintaining individual model transition rates on par with world averages and constructing models that adequately predict the rare events of interest.

Table 15: Nodal Model Training and Validation Year Sets

Regional Models	Sub-Saharan Africa		South and East Asia		Eastern Europe and Central Asia	
Year Sets	In Conflict	Not In Conflict	In Conflict	Not In Conflict	In Conflict	Not In Conflict
Training Year Set	2004-2010	2004-2010	2004-2010	2008-2011	2006-2011	2004-2010
Validation Year Set	2011-2013	2011-2013	2011-2013	2012-2013	2012-2013	2011-2013
Markov Year Set	2014	2014	2014	2014	2014	2014

Regional Models	Arab & North African States		Latin America		OECD	
Year Sets	In Conflict	Not In Conflict	In Conflict	Not In Conflict	In Conflict	Not In Conflict
Training Year Set	2004-2010	2004-2009	2008-2011	2004-2010	2004-2010	2005-2009
Validation Year Set	2011-2013	2010-2013	2012-2013	2011-2013	2011-2013	2010-2013
Markov Year Set	2014	2014	2014	2014	2014	2014

Issues Encountered with Rare Event Prediction

As stated previously, the method of maximum likelihood estimation tends to favor the majority of occurrences in unbalanced data sets, resulting in the misclassification of rare events. All twelve nodal models experienced minor-to-moderate detrimental impacts to model prediction accuracy as a result of this phenomenon. The careful selection of training and validation year-sets enabled the mitigation of misclassification issues in 11 of the 12 nodal models. However, all initial Eastern Europe & Central Asia "Not in Conflict" models failed to properly classify a single conflict transition in any of the validation models. To correct this deficiency, the Synthetic Minority Over-sampling Technique was utilized to produce 48 additional minority instances (transitions from no-conflict into conflict) for the training model data set. The additional data points were generated using a SMOTE algorithm developed for the MATLAB modeling environment using the 16 conflict transitions from year sets 2004 to 2010 as the primary input (MathWorks, 2015). Following generation, the 48 instances were analyzed to ensure completeness and similarity to the original variables. It was noted that design variables such as "Government Type" and "Regime Type" were approximated as continuous variables with values ranging from 0.20 to 0.70. To correct this issue, values less than 0.50 were rounded down to 0, where those greater than or equal to 0.50 were rounded up to 1, while ensuring only 1 level for each variable assigned to a particular instance. Five separate models were developed using the SMOTE training-set, all of which predicted at least one of the seven observed conflict transitions from year-set 2011-2013. Discussion and analysis of the final nodal models for each region is discussed in Chapter IV.

3.4 Markov Models

Overview

The use of nation-specific Markov models as a forecasting tool for conflict and conflict transitions is envisioned to provide operationally relevant and tractable analysis of future conflict trends. This study utilizes the two-state Markov Model depicted in Figure 13, providing the probabilities of conflict transition for the following year, given the current conflict status of the nation in question. The Markov model base year (year 0) is for this study is affixed at 2014, corresponding to the data provided in the most recent HIIK conflict barometer. Subsequently, the transition probabilities for the Markov base year are calculated using the conditional logistic regression models and applied to all 182 nations in the 2014 data set. The use of Markov models provides insights into expected transition times, mean recurrence, as well as the long-run proportions that a nation will remain in a particular status. Ultimately the use of Markov models provides operationally relevant global conflict forecasting with prediction horizons greater than one year.

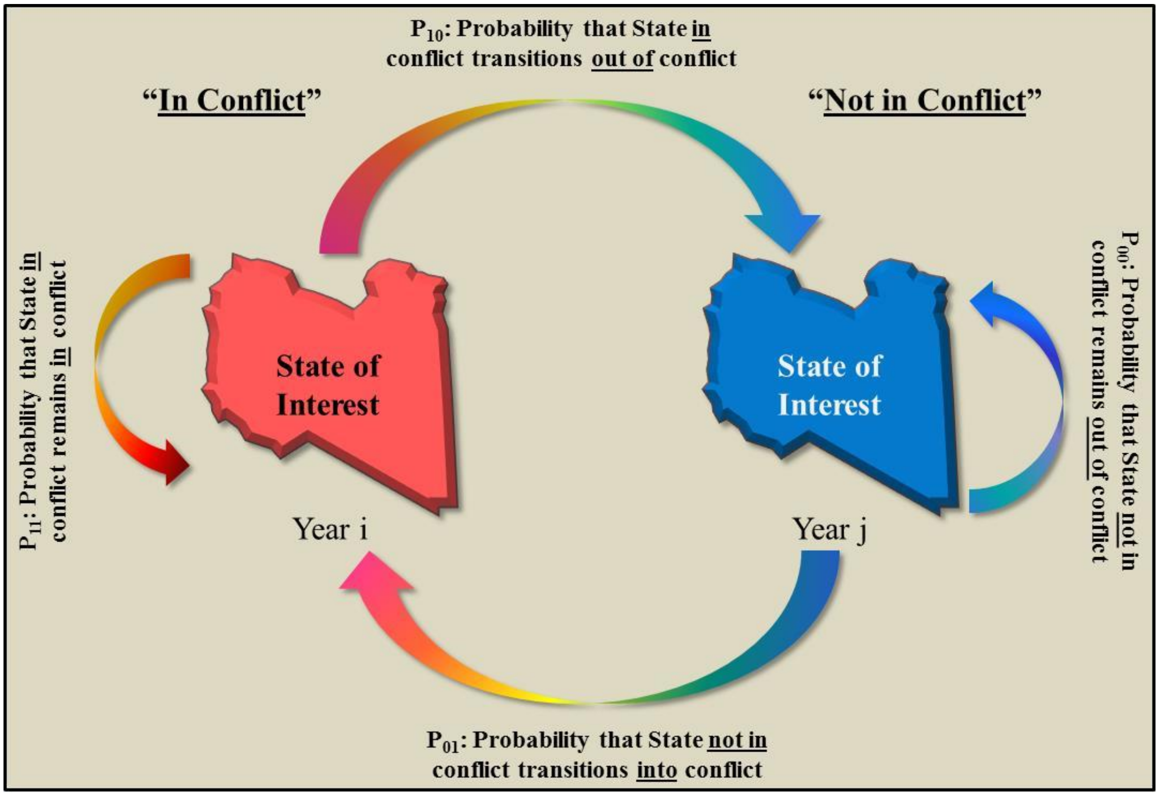


Figure 13: Nation Specific Conflict Transition Markov Model

The Markov Process

A discrete time Markov process is a stochastic process given $\{X_n, n = 1, 2, ...,\}$ that takes on a finite number of possible values, which for the purposes of this study will include the entire set of non-negative integers. If $X_n = i$, the process is said to be in state i at time *n* (Ross, 2014). Given that the current system in in state i at time *n*, there exists a fixed probability P_{ij} that the system will transition to state j, at time $n + 1$, as shown in Equation 17.

$$P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0\} = P\{X_{n+1} = j | X_n = i\} = P_{ij}$$

Equation 17: Markov Chain (Ross, 2014)

This equation may be interpreted as the conditional probability of any future state X_{n+1}, given past states and the present state X_n, is independent of all previous states, and is conditioned only on the current state (Ross, 2014). This is known as the one-step transition probability, and can also be represented as: $P_{ij} = P_{ij}^0 = P_{ij}^1$, which will provide a useful form of notation when dealing with *n*-step transition probabilities as shown in Equation 18. It follows then that the sum of the probabilities of all transition options given a current state, is equal to *1*.

$$\sum_{j=0}^{\infty} P_{ij} = 1, \forall\, i = 0, 1, 2, \ldots$$

Equation 18: Summation of Transition Probabilities for a Current State

This feature of the Markov state probabilities is illustrated in the generic two-state Markov chain (***P***) depicted in Equation 19.

$$\boldsymbol{P} = \begin{Vmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{Vmatrix}$$

Equation 19: Generic Two-state Markov Chain

Generic Two-state Markov Chain In this example the probability of remaining in state "0", given you are currently in "0" is given by $P_{00} = \alpha$, while the probability of transitioning from state "0" to state "1" is given by $P_{01} = 1 - \alpha$. Similarly, the probability of transitioning to state "*0*" from state "*1*" is $P_{10} = 1 - \beta$, and remaining in state "*1*" is $P_{11} = \beta$. For the purposes of the study, the following state transition probabilities are defined.

P_{00} = Probability that nation not in conflict remains out of conflict
P_{01} = Probability that nation not in conflict transitions into of conflict
P_{10} = Probability that nation in conflict transitions out of conflict
P_{11} = Probability that nation in conflict remains in conflict

A brief discussion of the accessibility of states within a Markov model is warranted before we proceed further. State j is said to be accessible from state i if $P_{ij}^{n} > 0$ (Ross, 2014). Additionally, the states of the models employed in this study are said to communicate, since they are always accessible from each other. This is germane to this study, as all included nations have the potential to transition from one state to another (i.e., there exist no probabilities such that $P_{ij}^{n} = 0$). However, as will be discussed later, there are numerous states that have transition probabilities approaching 0. This condition results in some very interesting phenomena when analyzing the stability, recurrence and long-run proportions of nations and is discussed in Chapter IV.

Chapman-Kolmogorov Equations

The Chapman-Kolmogorov equations provide a method for computing the probability that a system currently in state i will transition to state j after n additional transitions (Ross, 2014). The concept of the n-step transition probability is easily relatable to the 1-step transition discussed previously and is shown in Equation 20.

$$P_{ij}^{n} = P\{X_{n+k} = j | X_k = i\}$$

Equation 20: Markov n-step probability

Computation of the n-step transition probabilities occurs via sum-product of the transition probabilities for periods k and n, as shown in Equation 21.

$$P_{ij}^{n+m} = \sum_{k=0}^{\infty} P_{ik}^{n} P_{kj}^{m}$$

Equation 21: Chapman-Kolmogorov Equation

The multiplication of transition probabilities $P_{ik}^{n} P_{kj}^{m}$, represent the probability that, starting in state i the process will transition to state j in $n + m$ transitions, through a path that will move through state k at the n^{th} transition (Ross, 2014). This concept is subsequently adapted to the calculation of the n-step transition matrix probabilities $\boldsymbol{P}^{(n)}$, through the use of matrix multiplication as shown in Equation 22.

$$\boldsymbol{P}^{(n+m)} = \boldsymbol{P}^{(n)} \cdot \boldsymbol{P}^{(m)}$$

Equation 22: Transition probabilities for the n-step matrix

This extremely powerful concept of using matrix multiplication to simultaneously determine the transition probabilities of each state for a given time period, forms the basis of the conflict forecasting and analysis tool developed for this study.

Sojourn Times and Variance

Relevant to the forecasting of conflict transitions is the expected time to the first conflict transition, or simply given that a nation is currently in state i, what is the expected time R_j until is it is in state j? The time to first transition, from a designated time 0, is simply calculated by taking the inverse of the probability given the system is currently in state i, the system will transition into state j. The expected time to the first transition and its variance are calculated using Equation 23.

$$
\begin{aligned}
&E\left[R_j\right]=\frac{1}{p_{ij}^0} \\
&\therefore \\
&E\left[R_1\right]=\frac{1}{p_{01}^0};\ \mathrm{E}\left[R_0\right]=\frac{1}{p_{10}^0} \\
&VAR\left[R_j\right]=\frac{p_{jj}}{\left(1-p_{jj}\right)^2}
\end{aligned}
$$

Equation 23: Expected Time to First Transition and Variance

Where:

R_1 = Time to 1st transition (non-conflict to conflcit)
R_0 = Time to 1st transition (conflict to non-conflcit)
p_{01}^0 = Transition probability at time zero (non-conflict to conflcit)
p_{10}^0 = Transition probability at time zero (conflict to non-conflcit)

It should be remembered that, due to the memoryless properties of the Markov model, time 0 is relative and can be designated at any point in time.

Recurrence and Long-Run Proportions

As stated previously, the Markov models employed in this study have states that are accessible from every other state; creating a condition known as positive recurrence. A state j is said to be positive recurrent if the number of expected transitions it takes to start and then return to state j is less than infinity (i.e., $m_j < \infty$) (Ross, 2014). The mean recurrence for any state is given by Equation 24.

$$m_j = \frac{1}{\pi_j}$$

Equation 24: Mean recurrence time for state *j*

Where:

$\pi_j =$ Long run proportion of time spent in state j

The concept of recurrence leads directly to the idea of Markov model long run proportions π_j, the expected percentage of time a system will be in state j. The long run proportions of a Markov chain are closely associated with the eigenvalues of the transition matrix ***P*** (Ross, 2014). Additionally, like the state specific transition probabilities p_{ij}, the long run proportions must also sum to 1. The derivation of the two-state long run proportions is provided in Equation 25.

$$\begin{aligned}
&\pi_0 = p_{00} \cdot \pi_0 + p_{10} \cdot \pi_1 \\
&\pi_1 = p_{01} \cdot \pi_0 + p_{11} \cdot \pi_1 \\
&\pi_0 + \pi_1 = 1 \\
&\pi_0 = \frac{p_{10}}{1 + p_{10} - p_{00}}, \ \pi_1 = \frac{p_{01}}{1 + p_{10} - p_{00}}
\end{aligned}$$

Equation 25: Two-State Long Run Proportions

Where:

$\pi_0 =$ Long run proportion of time spent not in conflict
$\pi_1 =$ Long run proportion of time spent in conflict

It should be noted, that the long run proportions π_j can be approximated by raising the transition probability matrix P, to a significantly high power, as demonstrated in the Equation 26.

$$\begin{aligned}
&P^{\infty} \approx P^{n} = \begin{Vmatrix} \pi_0 & \pi_1 \\ \pi_0 & \pi_1 \end{Vmatrix} \\
&\textit{where} \\
&n >> 50 \\
&n = \text{number of periods into the future}
\end{aligned}$$

Equation 26: Long Run Proportion Approximation

Development of the Nation Specific Markov Models

The construction of the two-state transition probability matrix required that the region-specific conditional models calculate the transition probabilities for the year 2014 data set. The process involved applying both conditional models, for a specific region, to the same data set, resulting in four distinct transition probabilities for each of the 182 nations considered. Specifically, the transition probabilities p_{00} and p_{01} were calculated from the "Not in Conflict" models, while p_{10} and p_{11} were calculated from the "In Conflict" models. These probabilities are subsequently compiled into a VBA enabled, Microsoft Excel workbook known as the "Conflict Transition Probability Markov Chain Tool." The tool enables the automated calculation of the n-step transition probabilities for a specified time-period, first and second sojourn times and variances, the mean recurrence times, as well as the long-run proportions for each state. An example of the tool is shown in Figure 14.

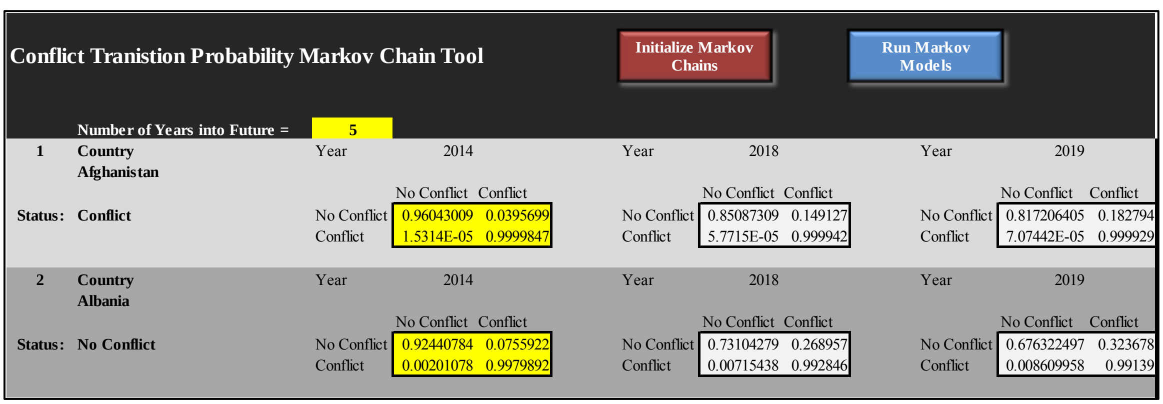


Figure 14: Conflict Transition Probability Markov Tool

3.5 Summary

This chapter described the methodology employed to construct the conditional conflict database, as well as the theory and methodology guiding the development of the logistic regression models that are used to calculate the transition probabilities required for the Markov Models. The creation methodology enables repeatability of this study's results as well a means to evaluate additional alternatives associated with variable selection and model development.

This methodology examined 30 statistical variables acquired from several open sources for the development of both the dependent variable and the conditional logistic regression models. The data resources employed in this methodology are similar to, or updates of, the previous analytical efforts discussed in Chapter II. The data sets utilized in this study are professionally created and maintained by reputable organizations, that strive to maintain the most current and accurate data. However, the nature of data collection in less than fully permissive environments results in incomplete and often time lagged data sets that form the basis of this study. Despite less than timely and perfect in data, there exist methods and techniques that enable the construction and relevant analysis of robust conflict prediction models.

The strengths of the models developed for this study lie in their ability to "operationalize" complex regional conflict environments to key underlying factors that influence conflict transitions. Additionally, the combination of logistic-regression and Markov models enables long range forecasting of world-wide conflict trends that is not possible with logistic-regression models alone. Moreover, the models developed for this study are surprisingly not limited in their predictive power by the quantity, quality, and in some instances the timeframe of the information currently available. As will be discussed in later on in this work, seminal events such as the Arab Spring or the rise of the Islamic State signal a possible paradigm shift of relevant predictors of violent conflict, a shift that may take several years of data collection to fully realize the precursors and impacts of these events. The implementation of the methodologies previously described, and the relevant analysis is presented in Chapter IV.

IV. Analysis and Results

"However, the pulse of the God of War is hard to take. If you want to discuss war, particularly the war that will break out tomorrow evening or the morning of the day after tomorrow, there is only one way, and that is to determine its nature with bated breath, carefully feeling the pulse of the God of War today."

Qiao Liang, Unrestricted Warfare

4.1 Chapter Overview

The purpose of this chapter is to describe and analyze the results of the methodology discussed in Chapter III. First, in Section 4.2, we discuss the construction and validation of the six regional conditional logistic regression models. Next, Section 4.3 provides an in depth analysis of the significant variables by region and conditional model. Subsequently, Section 4.4 examines the construction, validation, and results for the nation specific Markov models. Finally, in Section 4.5 we provide an analysis of future global conflict trends developed from the Markov models.

4.2 Analysis of Region Specific Conditional Logistic Regression Models

Development of the Regional Conditional Logistic Regression Models

The Purposeful Selection of Covariates method was employed in the construction of all twelve conditional logistic regression models (two per region) used in this study. The method provides a systematic means to efficiently construct meaningful and operationally relevant models that achieve suitable classification accuracies in both the training and validation data sets. Initial analysis of the logistic regression models focused on maximizing the area under the curve (AUC) for the specific ROC curves while simultaneously ensuring appropriate model and variable significance. It is the goal of this study to ensure an AUC greater than 0.80 for models with p-values less than or equal to 0.05. As seen in Table 16, the conditional model developed for Sub-Saharan African States classified as not in conflict the previous year, is significant with p-value of 0.00001 with a complimentary AUC of 0.874, indicating the model is both highly significant and an excellent discriminator. Additionally all seven variables have p-values considerably less than 0.05, indicating high levels of significance, and further reinforcing the overall suitability of this model for conflict transition prediction.

Table 16: Sub-Saharan, Given Non-Conflict Logistic Regression Model

Sub- Saharan Africa (Given Non-Conflict) Model

Variable	Coefficient	G	p
Arable Land	7.801	13.640	0.000
Birth Rate	-0.474	8.740	0.003
Infant Mortality rate	0.053	8.240	0.004
Youth Bulge	0.346	6.470	0.011
Refugee (Asylum)	5.91E-06	5.890	0.015
Trade (% GDP)	-0.052	7.830	0.005
Freedom Score	-5.637	12.760	0.000
Log-Likelihood =	43.446		
G =	40.754		
P =	**0.00001**		
AUC =	**0.874**		

Given the multitude of potential variable combinations (equivalent to 30!) for each model, there exist multiple potential significant conditional models for each region. Consequently, multiple distinct models were developed, analyzed and compared to identify the optimal conditional models for each region. If the initial analysis indicated the models experienced satisfactory significance and discrimination, an in-depth analysis of overall model performance was conducted; an example of such a comparison is presented in Table 17.

Table 17: Sub-Saharan Africa "Not in Conflict" Conditional Model Comparison

Sub- Saharan Africa (Given Non-Conflict) Model 1

Variable	Coefficient	G	p
Arable Land	7.801	13.640	0.000
Birth Rate	-0.474	8.740	0.003
Infant Mortality rate	0.053	8.240	0.004
Youth Bulge	0.346	6.470	0.011
Refugee (Asylum)	5.91E-06	5.890	0.015
Trade (% GDP)	-0.052	7.830	0.005
Freedom Score	-5.637	12.760	0.000

Log-Likelihood =	43.446
G =	40.754
P =	0.000
AUC =	0.874

Training Data Set Model

Sub-Saharan Africa (Right Node): 2004-2010			
	Observed		
Classified	Transition to Conflict = 1	Remain/Transition out of Conflict = 0	Total
Transition to Conflict = 1	6	4	10
Remain/Transition out of Conflict = 0	24	155	179
Total	30	159	189

Med Cut Point:	0.50
Model Acuracy:	0.852

Validation Data Se Model

Sub-Saharan Africa (Right Node): 2011-2013			
	Observed		
Classified	Transition to Conflict = 1	Remain/Transition out of Conflict = 0	Total
Transition to Conflict = 1	2	0	2
Remain/Transition out of Conflict = 0	10	61	71
Total	12	61	73

Med Cut Point:	0.50
Model Acuracy:	0.863

Sub- Saharan Africa (Given Non-Conflict) Model 2

Variable	Coefficient	G	p
Arable Land	6.574	11.300	0.001
Population Growth	-1.991	11.820	0.001
Infant Mortality rate	0.028	3.690	**0.055**
Caloric Intake	1.70E-03	6.170	0.013
Refugee (Asylum)	0.000	5.690	0.017
Trade (% GDP)	-0.052	8.890	0.003
Freedom Score	-6.575	14.590	0.000

Log-Likelihood =	41.656
G =	44.334
P =	0.000
AUC =	0.873

Training Data Set Model

Sub-Saharan Africa (Right Node): 2004-2010			
	Observed		
Classified	Transition to Conflict = 1	Remain/Transition out of Conflict = 0	Total
Transition to Conflict = 1	5	1	6
Remain/Transition out of Conflict = 0	25	158	183
Total	30	159	189

Med Cut Point:	0.50
Model Acuracy:	0.862

Validation Data Se Model

Sub-Saharan Africa (Right Node): 2011-2013			
	Observed		
Classified	Transition to Conflict = 1	Remain/Transition out of Conflict = 0	Total
Transition to Conflict = 1	0	0	0
Remain/Transition out of Conflict = 0	12	61	73
Total	12	61	73

Med Cut Point:	0.50
Model Acuracy:	0.836

In the example illustrated in Table 17, two distinct conditional models were developed for Sub-Saharan African states classified as not in conflict. In this example, Model 1 is the same model shown in Table 16, while Model 2 has replaced the variables "Birth Rate" and "Youth Bulge" with "Population Growth" and "Caloric Intake" respectively. Initial analysis indicates satisfactory significance for both models and their respective variables and equivalent discriminatory powers as indicated by their AUC values. Model performance was then compared using the Training and Validation datasets as described previously, with a classification cut-point fixed at 0.50 for all comparisons. The analysis focused on four criteria in descending order: (1) Overall Predictive Accuracy of the model on the validation data set, (2) Overall Predictive Accuracy of the model on the training data set, (3) Overall ability to properly classify rare-events (transitions from current state) in both data sets, and (4) Minimum number of "False Negatives" in the validation data set, shown in the bottom left-hand quadrant of the classification table.

In this example, a rare-event is considered a "Transition to Conflict" which occurs in 42 of the 262 total instances across both data sets. Analysis of the results shows that Model 1 outperforms Model 2 in three of the four criteria: higher validation model predictive accuracy (0.863), classification of rare events (8 of 42 transitions), and 10 total false negatives as opposed to 12 in Model 2. Despite having a slightly lower training data set accuracy (0.852), attributed to classifying 155 of the 159 true-negative instances, using our criteria, Model 1 is considered to be the superior of the two prospective "non-conflict" conditional models for the Sub-Saharan Africa region. Similar analyses were conducted on the conditional models for all regions, ultimately identifying the 12 conditional models used in this study.

The objective of this model building strategy was the construction of parsimonious conditional logistic regression models that achieve prediction accuracies in excess of 80% for both the training and validation data sets. A summary of the final conditional models for each region is provided in Appendix B.

Model Validation and Analysis

Three methods were employed in the validation and analysis of the 12 conditional logistic regression models: (1) Receiver Operating Characteristic area under the curve values, (2) Classification Accuracy, and (3) Hosmer-Lemeshow Goodness of Fit Tests. These analyses assess the suitability of the conditional logistic regression models in terms of overall discriminative power, model accuracy (with an emphasis on rare-events), and the model's approximation of the data.

Receiver Operating Characteristic Area Under the Curve Analysis

As stated earlier, the Receiver Operating Characteristic (ROC) curve graphically depicts a models ability to detect a signal in the presence of noise across the entire range of possible cut-points. ROC Curves are developed for both the training and validation data sets as means to assess the overall model performance. An example of a typical set of training and validation ROC curves for a conditional model is provided in Figure 15.

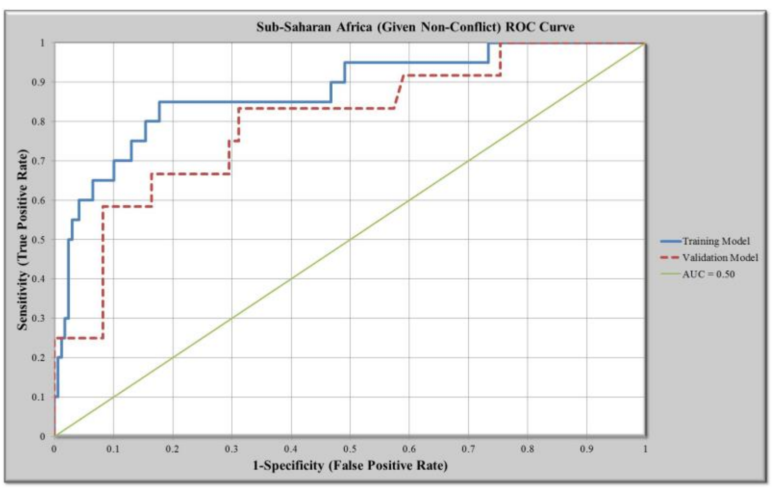


Figure 15: Graph of Training and Validation ROC Curves.

Visual inspection of the two ROC curves shows that the model in this example provides better discrimination than would be obtained by simple binary guess (e.g. a fair coin toss) represented by the AUC = 0.50 diagonal line that bisects the graph. It is also readily apparent that the discriminative power of the training model exceeds that of the validation model, a logical result, which is a product of our model building strategy. To understand the model's performance more precisely, we employ AUC analysis using the criteria described in Chapter III. Table 18 summarizes the AUC values for both the training and validation data sets for all conditional models across the six geographic regions and as a combined world model.

Table 18: AUC Values by Region and Model

Receiver Operating Characterist AUC Scores by Region					
Region	Model	Training Data Set		Validation Data Set	
		AUC	Assessment	AUC	Assessment
Arab & North African States	In Conflict	0.962	Superior	0.500	No Model Discrimination
	Not in Conflict	0.930	Superior	0.520	Poor
Eastern Europe & Central Asia	In Conflict	0.972	Superior	0.659	Poor
	Not in Conflict	0.946	Superior	0.651	Poor
Latin America	In Conflict	0.878	Excellent	0.750	Acceptable
	Not in Conflict	0.952	Superior	0.776	Acceptable
OECD	In Conflict	0.914	Superior	0.561	Poor
	Not in Conflict	0.974	Superior	0.735	Acceptable
South & East Asia	In Conflict	0.938	Superior	0.689	Poor
	Not in Conflict	0.932	Superior	0.696	Poor
Sub-Saharan Africa	In Conflict	0.889	Excellent	0.704	Acceptable
	Not in Conflict	0.874	Excellent	0.796	Acceptable
Combined World Model	In Conflict	0.887	Excellent	0.655	Poor
	Not in Conflict	0.922	Superior	0.743	Acceptable

The assessed performances of all the models using the training data set ranges from excellent discrimination ($0.80 \leq AUC < 0.90$) to superior discrimination ($AUC \geq 0.90$); with 8 of the 12 models assessed as superior discriminators on the training data set. However, there is a noticeable degradation in model performance on the validation data sets, with performance assessments generally ranging from poor discrimination ($0.50 \leq AUC < 0.70$) to acceptable discrimination ($0.70 \leq AUC < 0.80$). Of interest is the significant decline in model performance for the Arab and North African conditional models, each of which experiences performances losses in excess of 40% from the training to the validation data sets. Initially, it was theorized that the relative rarity of conflict transitions was the root cause of the degradation in performance. However, analysis of validation sets for other region models shows that this theory is not highly correlated with validation model performance. Another possible explanation in the degradation of the validation model performance may be linked to the "Arab Spring". The Arab Spring and its resulting conflicts have continued to engulf Southwest Asia and North Africa since the Tunisian revolution. This date is significant to the Arab & North African models, due to resulting conflicts in otherwise stable regimes that occur only in the validation data sets.

Another interesting occurrence that is observed in the AUC scores is the general trend for "Not in Conflict" models to experience better performance at both regional and combined world levels than their "In Conflict" counterparts. This trend is observed in both the training and validation data sets, and it occurs in 19 of the 24 instances presented in Table 18. A possible explanation of this phenomenon may be related to the inability to accurately collect data from nations experiencing conflict. The results suggest the data associated with nations that transition or remain out of conflict provides improved predictive performance over nations that tend to be in conflict.

Analysis of Model Validity Based on Classification Accuracy

Analysis of classification tables provides a second method used to assess logistic regression model validity and suitability. Classification table analysis, as it pertains to this study, focuses on three areas: (1) Overall model accuracy, (2) Percentage of rare-events properly classified, and (3) Model false negative rate. Initial analyses fixed the classification table cut point at 0.50, thereby classifying all instances with $\pi_i < 0.50$ as transitioning/remaining out of conflict, and all instances with $\pi_i \geq 0.50$ as transition/remaining in conflict. As stated previously, the objective of our model building strategy is to construct models that achieve classification accuracies in excess of 80% for both the training and validation data sets. A summary of the overall model strategies using the fixed cut-point of 0.50 is presented in Table 19 which details the accuracies and total instances per data set for each model.

Table 19: Overall Classification Accuracies Given Fixed Cut-point of 0.50

Model Accuracies Using 0.50 Classification Cut Point							
Region	Model	Cut Point	Training Data Set		Validation Data Set		Traning and Validation
			Accuracy	No. Instances	Accuracy	No. Instances	
Arab & North African States	In Conflict	0.50	94.2%	52	74.4%	43	85.3%
	Not in Conflict	0.50	93.3%	60	60.0%	15	86.7%
Eastern Europe & Central Asia	In Conflict	0.50	92.5%	67	82.8%	29	89.6%
	Not in Conflict	0.50	86.0%	171	76.7%	43	84.1%
Latin America	In Conflict	0.50	81.1%	37	83.3%	30	82.1%
	Not in Conflict	0.50	90.9%	132	88.1%	42	90.2%
OECD	In Conflict	0.50	88.7%	53	86.4%	22	88.0%
	Not in Conflict	0.50	96.0%	126	95.0%	101	95.6%
South & East Asia	In Conflict	0.50	87.3%	79	84.1%	44	86.2%
	Not in Conflict	0.50	87.9%	66	88.0%	25	87.9%
Sub-Saharan Africa	In Conflict	0.50	86.4%	154	82.4%	74	85.1%
	Not in Conflict	0.50	85.2%	189	86.3%	73	85.5%
Combined World Results	In Conflict	0.50	88.2%	442	81.8%	242	86.0%
	Not in Conflict	0.50	89.1%	744	87.0%	299	88.5%

Model accuracies exceeded the 80% classification accuracy benchmark in all 12 training data sets, and in 9 of the validation data sets. Training data set accuracies averaged 88.2% for “In Conflict” conditional models, and 89.1% for “Not in Conflict” models, with 5 of the 12 training data sets yielding accuracies above 90%. As expected, the models experience some degradation in their classification accuracies when applied to the validation data set, but they still achieve average accuracies of 81.8% and 87.0% for the “In Conflict” and “Not in Conflict” models respectively. The overall classification accuracies for both the training and validation data sets exceed the 80% benchmark for all regions and are considered suitable for the purposes of this study. Similar to the AUC analysis, the “Not in Conflict” models generally experience greater predictive accuracies than the “In Conflict” counterparts, with the phenomenon observed in 19 of the 24 instances provided in Table 19. The exception to this trend seems to occur more frequently in the Arab & North African, and the Eastern Europe & Central Asian models than in the rest of the regions.

These results compare favorably with historical studies which have struggled to achieve prediction accuracies greater than 80%. Studies such as the CAA-led Forecast and Analysis of Complex Threats (Reed, 2013) or the Political Instability Task Force’s global forecasting model (Goldstone, et al., 2005) only achieve accuracies greater than 80% on limited and very specific data sets. On the other hand, the Boekestein model achieved accuracies approaching 80% without implementing special conditions to enable prediction accuracy; these model accuracies were subsequently compared to those developed by this study (Boekestein, 2015). To enable a one-to-one model comparison by region, we have developed weighted regional accuracies for both the training and validation data sets; Table 20 provides a comparison of this study's results with the recent Boekestein study. As can be seen, both models perform very well at the regional level, with all training data sets yielding accuracies in excess of 80%. Both models perform similarly at the regional level, however the conditional logistic regression / Markov chain (C-LR/MC) model developed for this study achieves higher overall prediction accuracies at the combined world level. Comparison of the respective model performance on the validation data sets reveals that the C-LR/MC model realizes a significant improvement in prediction accuracy over the Boekestein model. The C-LR/MC model attains higher prediction accuracies for each of the six regions for the validation data set, and a 84.67% weighted prediction accuracy at the combined world level.

Table 20: Comparison of Model Accuracies with the Boekestein Model

Comparison of Boekestien Model Accuracies with Conditional Logistic Regression/Markov Chain Weighted Accracies by Region				
Region	Training Data Set Accuracies		Validation Data Set Accuracies	
	Boekestein Model	Conditional LR/MC Weighted Accuracies	Boekestein Model	Conditional LR/MC Weighted Accuracies
Arab & North African States	84.31%	**93.72%**	70.59%	**70.68%**
Eastern Europe & Central Asia	77.38%	**87.83%**	75.00%	**79.16%**
Latin America	90.12%	**88.75%**	77.78%	**86.10%**
OECD	95.96%	**93.84%**	92.42%	**93.46%**
South & East Asia	90.48%	**87.57%**	76.79%	**85.51%**
Sub-Saharan Africa	82.31%	**85.74%**	74.49%	**84.34%**
Combined World Results	86.63%	**88.76%**	78.30%	**84.67%**

Three of the validation data sets, both Arab & North African conditional models, and the Eastern Europe & Central Asia "Not in Conflict" model, fail to achieve accuracies greater than 80%. The classification tables for these three validation data sets are shown in Table 21.

Table 21: Validation Data Set Accuracies below Accuracy Benchmark of 80%

Arab States (Given Conflict): 2011-2013			
	Observed		
Classified	Transition/Remain in Conflict = 1	Transition out of Conflict = 0	Total
Transition/Remain in Conflict = 1	32	1	33
Transition out of Conflict = 0	10	0	10
Total	42	1	43

Med Cut Point:	0.50
Model Acuracy:	0.744

Arab States (Given Non-Conflict): 2010 - 2013			
	Observed		
Classified	Transition to Conflict = 1	Remain/Transition out of Conflict = 0	Total
Transition to Conflict = 1	1	2	3
Remain/Transition out of Conflict = 0	4	8	12
Total	5	10	15

Med Cut Point:	0.50
Model Acuracy:	0.600

E. Europe & Central Asia (Given Non-Conflict): 2011 - 2013			
	Observed		
Classified	Transition to Conflict = 1	Remain/Transition out of Conflict = 0	Total
Transition to Conflict = 1	2	5	7
Remain/Transition out of Conflict = 0	5	31	36
Total	7	36	43

Med Cut Point:	0.50
Model Acuracy:	0.767

The effects of the Arab Spring on model accuracy become apparent in the Arab and North African models, specifically in the "In Conflict" model which misclassifies 11 of the 43 instances. The model, developed using data that completely pre-dates the Arab Spring, achieves an accuracy of 74.4% and classifies nearly a quarter (10) of the total instances as transitioning out of conflict, when in reality only one such transition occurs during the 2011 to 2013 time period (i.e., Oman in 2011 – 2012). The Arab & North African "Not in Conflict" model experienced even greater misclassification rates (40% of all instances misclassified), resulting in an overall classification accuracy of 60% for the validation model. However, three misclassified transitions: Libya (2010 – 2011), Syria

(2010 – 2011), and Tunisia (2010 – 2011), all of which transition into conflict in 2011, are directly related to the Arab Spring, and it is likely these nations would have remained out of conflict had this event not occurred.

As noted previously, the initial models developed for the Eastern Europe and Central Asia “Not in Conflict” data experienced numerous classification issues, often failing to properly classify any transitions into conflict. As a result, the Synthetic Minority Oversampling Technique (SMOTE) was employed to aid development of a model that achieved satisfactory classification accuracy in both the training and validation data sets. These initial models maximized the likelihood of these nations transitioning or remaining out of conflict resulting in significant false-negative rates (in excess of 20% of all instances), the complete failure to classify any nation as transitioning into conflict, and model accuracies in the 70% range. Despite failing to generate classification accuracies above 80%, the final “Not in Conflict” model is a significant improvement over the earlier versions, providing better overall classification accuracy with reduce false-negative rates.

Accurate model building challenges for Eastern Europe and Central Asia may be the result of an ethnically diverse and widespread geographic region that straddles the both Eastern and Western civilization. The conflicts within this region generally take on two forms; in the east conflicts are generally the result of long standing tribal conflicts and foreign intervention, while in the west financial crises, immigration, and political turmoil (notably in the former Soviet states) exacerbate political and societal instability. Ultimately, future studies may wish explore a realignment of the nations within this geographic region in order to improve model performance.

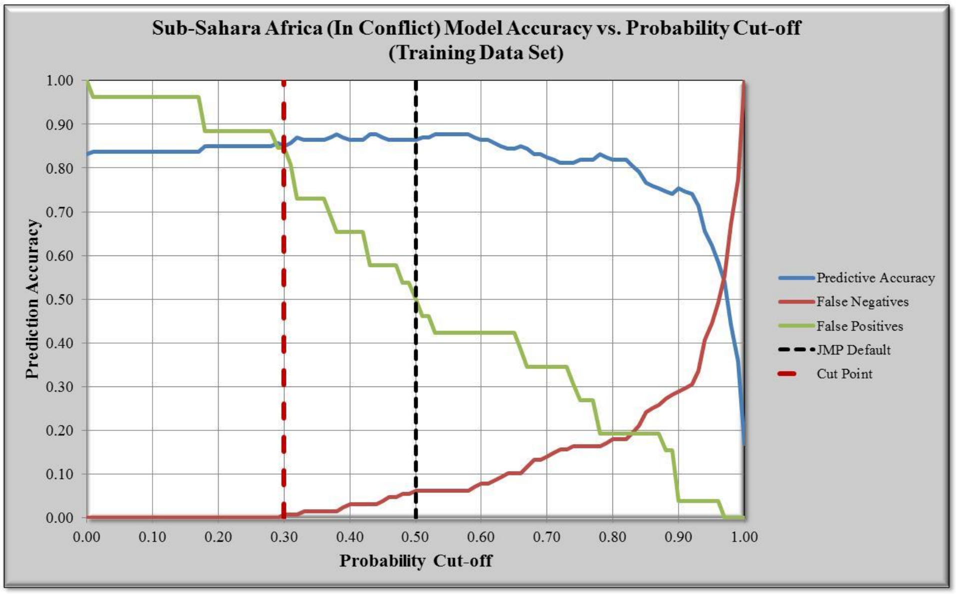
While overall model accuracies are considered to meet or exceed expectations, further analysis is required to ascertain model performance concerning rare events (i.e., transitions into or out of conflict). The principal of maximum likelihood will favor the majority population in any data set, at the expense of the minority. With transition rates ranging from 5–20% across all data sets it is possible achieve benchmark classification accuracies simply by only properly classifying the majority population of the conditional model. As part of the overall model assessment, rare event accuracies must be taken into account.

Table 22: Model Rare Event Accuracies Given Fixed Cut-point of 0.50

Model Rare Event Accuracies Using 0.50 Classification Cut Point							
Region	Model	Cut Point	Training Data Set		Validation Data Set		Traning and Validation
			Accuracy	No. Instances	Accuracy	No. Instances	
Arab & North African States	In Conflict	0.50	60.0%	5	0.0%	1	50.0%
	Not in Conflict	0.50	66.7%	9	20.0%	5	50.0%
Eastern Europe & Central Asia	In Conflict	0.50	76.9%	13	50.0%	6	68.4%
	Not in Conflict	0.50	79.7%	64	28.6%	7	74.6%
Latin America	In Conflict	0.50	42.9%	7	0.0%	2	33.3%
	Not in Conflict	0.50	61.1%	18	37.5%	8	53.8%
OECD	In Conflict	0.50	37.5%	8	0.0%	3	27.2%
	Not in Conflict	0.50	55.6%	9	50.0%	4	53.8%
South & East Asia	In Conflict	0.50	23.1%	13	33.3%	6	26.3%
	Not in Conflict	0.50	41.7%	12	0.0%	2	35.7%
Sub-Saharan Africa	In Conflict	0.50	50.0%	26	36.4%	11	45.9%
	Not in Conflict	0.50	20.0%	30	16.7%	12	19.1%
Combined World Results	In Conflict	0.50	48.6%	72	31.0%	29	43.6%
	Not in Conflict	0.50	59.2%	142	26.3%	38	52.2%

Rare event classification accuracies by region and model are provided in Table 22. Across all regions, the “In Conflict” models correctly classified 35 of the 72 (48.6%) transitions out of conflict, and the “Not in Conflict” models correctly classified 84 of the 142 (59.2%) transitions into conflict for all twelve training set models. Expectedly, validation rare-event classification accuracies are generally lower than their training counterparts at the regional level, with 9 of 29 (31.0%) transitions out of conflict, and 10 of 38 (26.3%) transitions into conflict properly classified as an aggregate model. It is observed that 4 of the 12 validation models failed to properly classify a single transition instance, represented by an assigned accuracy of 0.00%. However, in each of these cases the total number of observed transitions is less than or equal to 3, resulting in below average transition rates for the four regions. It is therefore assessed that overall model suitability is not affected by this singular result. Ultimately, model rare event classification accuracies, for the aggregate data sets, average 43.6% for "In Conflict" models, and 52.2% for "Not in Conflict" models, which is considered acceptable given the above overall predictive accuracies of the logistic regression models combined with the relative rarity of conflict transitions within the data set.

The final classification table analysis involves adjusting the cut-point in order to limit the number of false-negative classifications while maintaining suitable model accuracy. In this study, a false-negative is defined as a nation classified as transitioning/remaining out of conflict, when in fact the nation remains/transitions into conflict. Given the operational implications of misclassifying a potential transition into conflict, it is arguably better to reduce the model's false positive rate, which is achieved by adjusting the cut-point for each conditional model, than to misclassify a nation as being "Not in Conflict". A typical cut point analysis is presented in Figure 16, which graphs the conditional model accuracy, false-negative rate, and false-positive rate as function of the probability cut-point. As is the case for all models, the false-negative rate declines as the cut-point approaches zero. The vertical dashed lines represent the JMP-default cut points and adjusted default cut-points.



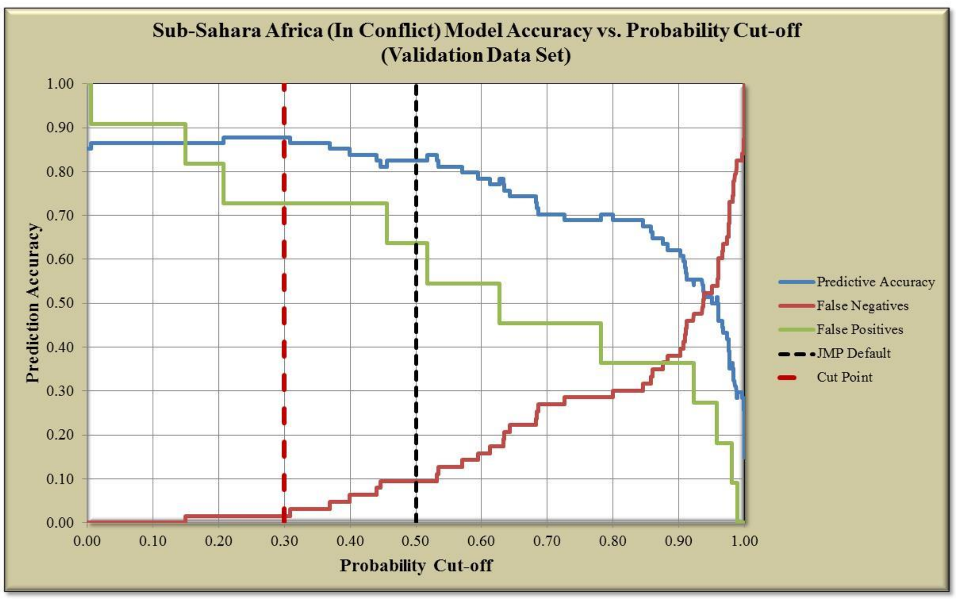


Figure 16: Analysis of Cut-Point Effects on Classification Accuracy and False-Negative Rates

Adjustment of the classification table cut-point seeks to balance three objectives: minimize false-negative rate, maintain model accuracy, and minimize the deviation from the JMP default cut-point of 0.50 for both the training and validation models. Minimization of the deviation in the adjusted cut-point from the JMP-default is desired due to its effects on limiting the model's false-positive rate, which increases as the cut-point approaches zero. The adjusted cut-points were set to values less than the JMP-Default in 10. However, in the remaining two cases, the default cut point was maintained due to no appreciable improvements in the training or validation models' accuracy or false negative rate. A summary of the adjusted cut-points effects on model accuracies and false negative rates in presented in Table 23.

Table 23: Effects of Adjusted Cut-points on Model Accuracy and False Negative Rates

Model Accuracies Seeking to Minimize False Negative Classifications in Training & Validation Models								
			Training Data Set			Validation Data Set		
Region	Model	Cut Point	Accuracy	False-Positive Decrease	Effects on Model Accuracy	Accuracy	False-Positive Decrease	Effects on Model Accuracy
Arab & North African States	In Conflict	0.30	94.2%	-100.0%	0.0%	86.0%	-50.0%	11.6%
	Not in Conflict	0.15	81.7%	-33.3%	-11.6%	66.7%	-25.0%	6.7%
Eastern Europe & Central Asia	In Conflict	0.34	92.5%	-100.0%	0.0%	86.2%	-50.0%	3.4%
	Not in Conflict	0.33	87.1%	-84.6%	1.1%	76.7%	0.0%	0.0%
Latin America	In Conflict	0.45	83.7%	-66.7%	6.7%	90.0%	-66.7%	6.7%
	Not in Conflict	0.40	90.9%	-28.6%	0.0%	88.1%	0.0%	0.0%
OECD	In Conflict	0.50	88.7%	0.0%	0.0%	86.4%	0.0%	0.0%
	Not in Conflict	0.30	96.0%	-25.0%	0.0%	94.1%	0.0%	-0.9%
South & East Asia	In Conflict	0.50	87.3%	0.0%	0.0%	84.1%	0.0%	0.0%
	Not in Conflict	0.42	84.8%	0.0%	-3.1%	84.0%	-50.0%	-4.0%
Sub-Saharan Africa	In Conflict	0.30	85.1%	-87.5%	-1.3%	87.8%	-83.3%	5.4%
	Not in Conflict	0.30	85.2%	-16.7%	0.0%	86.3%	0.0%	0.0%
Combined World Results	In Conflict	0.40	87.3%	-53.3%	0.5%	86.8%	-29.2%	2.5%
	Not in Conflict	0.32	84.5%	-34.5%	-0.3%	84.9%	-3.6%	-1.0%

Adjusted cut point values were tailored to each conditional model and ranged from 0.15 to 0.50, with the average cut-point set to 0.40 and 0.32 for the world level aggregate "In Conflict" and "Not in Conflict" models. These average cut points have negligible adverse impacts on overall and rare-event accuracies, and in many cases offer modest improvements at the regional level. Subsequently, the adjusted cut-points result in an overall decrease in the conditional model false negative rates at the aggregate world level for both the "In Conflict" and "Not in Conflict" models.

Analysis of Hosmer-Lemeshow Goodness of Fit Tests

The Hosmer-Lemeshow Goodness of Fit test provides the third and final method to assess the overall suitability of the logistic regression models. The Hosmer-Lemeshow test assesses the fit transition probabilities, $\pi(x_i)$, generated for each model instance, as they relate to the observed transition state. Model subgroupings were tailored to the individual models based on number of occurrences in the training model set and their corresponding transition probabilities. The design objective is to construct 10 equally sized sub-groups, providing a corresponding test statistic of $\chi^2_{(0.05,\ 8)} = 15.507$. However smaller numbers of sub-grouping were employed in 5 of the 12 tests. Via Equation 16, we are able to develop the Hosmer-Lemeshow Statistic ($\hat{C}$) and compare it to its corresponding Chi-square test statistic for each model. Assessed fit of a particular model is considered satisfactory if $\hat{C} < \chi^2_{(0.05,\ g-2)}$. The results of this analysis are summarized in Table 24.

Table 24: Hosmer-Lemeshow Goodness of Fit Test Results

Hosmer-Lemeshow Goodness of Fit Results given α = 0.05					
Region	**Model**	**H-L Statistic (C)**	**Test Statistic**	**P{T.S. > C}**	**Assessment**
Arab & North African States	**In Conflict**	1.550	5.991	0.461	**Model Apears to fit the data well.**
	Not in Conflict	6.390	15.507	0.604	**Model Apears to fit the data well.**
Eastern Europe & Central Asia	**In Conflict**	0.414	5.991	0.813	**Model Apears to fit the data well.**
	Not in Conflict	18.392	15.507	0.018	**Model Does Not Fit Data Well**
Latin America	**In Conflict**	1.425	5.991	0.490	**Model Apears to fit the data well.**
	Not in Conflict	4.812	15.507	0.777	**Model Apears to fit the data well.**
OECD	**In Conflict**	0.236	3.841	0.627	**Model Apears to fit the data well.**
	Not in Conflict	0.347	7.815	0.951	**Model Apears to fit the data well.**
South & East Asia	**In Conflict**	856.726	15.507	0.000	**Model Does Not Fit Data Well**
	Not in Conflict	37.342	15.507	0.000	**Model Does Not Fit Data Well**
Sub-Saharan Africa	**In Conflict**	6.440	15.507	0.598	**Model Apears to fit the data well.**
	Not in Conflict	48.543	15.507	0.000	**Model Does Not Fit Data Well**

Initial results indicate that 8 of the 12 conditional models appear to provide satisfactory fits with the exceptions being: Eastern Europe – Not in Conflict, both South

& East Asia models, and the Sub-Saharan Africa – Not in Conflict model. Analysis of these four models identified the set of outliers, provided in Table 25, that significantly contribute to the adverse test results.

Table 25: Hosmer-Lemeshow Test Significant Outliers

Nation	Year	Transition/Remain in Conflict (0, 1)	Probability	Sub Group
Easter Europe & Central Asia (Not in Conflict)				
Belarus	2008-2009	0	0.974	10
South & East Asia (In Conflict)				
Maldives	2004-2005	0	0.965	2
Bangladesh	2007-2008	0	0.973	2
Cambodia	2004-2005	0	0.977	2
Korea, North	2010-2011	0	0.987	2
Timor-Leste	2008-2009	0	0.995	3
China	2004-2005	0	0.999	4
Sri Lanka	2009-2010	0	1.000	7
South & East Asia (Not inConflict)				
Samoa	2011-2012	1	0.005	3
Sub-Saharan Africa (Not in Conflict)				
Congo, Republic of the	2006-2007	1	0.014	4
Comoros	2006-2007	1	0.015	4
Comoros	2009-2010	1	0.015	4
Mali	2005-2006	1	0.024	5
Mauritania	2007-2008	1	0.028	5
Sierra Leone	2010-2011	1	0.041	6

While the Hosmer-Lemeshow test assesses the overall fit of the model to the data, the overarching objective of this analysis is to identify and assess the existence of any significant model defects; this is achieved through outlier analysis. For the purposes of this study, significant outliers are misclassified observations with assigned transition probabilities less than 0.10 for “In Conflict” and greater than 0.90 for “Not in Conflict” models. In two of the four models: Eastern Europe – Not in Conflict and South & East

Asia – Not in Conflict, the presence of a single outlier indicates a possible issue in model fit, a highly dubious result given that a single outlier represents approximately 1% of the total instances for each model. This result is due in part to the method used to calculate the Hosmer-Lemeshow statistic ($\hat{C}$), which exponentially penalizes differences in the number of observed (o_{ik}) and expected (e_{ik}) occurrences (per bin), when the number of expected occurrences is small (i.e., $e_{ik} < 0.15$). While a single significant outlier does not elicit concern in the overall suitability of a particular model, the presence of multiple outliers may indicate the presence of model defects that require further investigation.

The seven significant outliers present in the South & East Asia-In Conflict model represent misclassifications of nations predicted to remain in conflict but which transitioned to a non-conflict status in the following year. Similarly, the six instances in the Sub-Saharan Africa-Not in Conflict model represent occurrences of nations predicted to remain out of conflict but which transitioned to a conflict status in the subsequent year. Given the demonstrated difficulty of correctly classifying conflict transitions, an audit of the individual outliers was conducted to determine if the assigned conflict transition probabilities were appropriate for the nation and region. For the South & East Asia – In Conflict model, the audit revealed that the assigned probabilities were appropriate in five of the seven instances, the exceptions being Maldives (2004 – 2005) and North Korea (2010 – 2011), given average probability of remaining in conflict and the number of years the nations were in a state of violent conflict between 2004 and 2014. The audit of the Sub-Saharan Africa-Not in Conflict model determined that the assigned probabilities were appropriate for three of the five nations, with only Mali and Mauritania, tending to be in a state of conflict, due to above average political instability, over the same 11-year period. The results of this audit are provided in Table 26.

Table 26: Audit of Significant Outliers

Nation	Year	Transition/Remain in Conflict (0, 1)	Probability	Average Probabity	Number Years in Conflict Status (2004 -2014)
South & East Asia (In Conflict)					
Maldives	2004-2005	0	0.965	0.521	3
Bangladesh	2007-2008	0	0.973	0.984	10
Cambodia	2004-2005	0	0.977	0.992	8
Korea, North	2010-2011	0	0.987	0.551	2
Timor-Leste	2008-2009	0	0.995	0.620	4
China	2004-2005	0	0.999	0.999	10
Sri Lanka	2009-2010	0	1.000	0.884	9
Sub-Saharan Africa (Not In Conflict)					
Congo, Republic of the	2006-2007	1	0.014	0.123	3
Comoros	2006-2007	1	0.015	0.312	3
Comoros	2009-2010	1	0.015	0.312	3
Mali	2005-2006	1	0.024	0.606	9
Mauritania	2007-2008	1	0.028	0.452	6
Sierra Leone	2010-2011	1	0.041	0.206	2

Overall Assessment of Logistic Regression Models

Given the results of this analysis, each of the 12 conditional logistic regression models are considered satisfactory and valid for the purposes of this study. Each of the logistic regression models exhibit excellent to superior levels of discrimination for the training data sets and adequate discrimination for the validation data sets. Model accuracies exceeded pre-established benchmarks (80% accuracy) in all 12 training models and 10 of 12 validation models, with overall model accuracies averaging 86.0% and 88.5% for the "In Conflict" and "Not in Conflict" models respectively. Assessment of model fit initially determined that only 8 of 12 models appeared to fit the data, however further analysis determined that the transition probabilities assigned to the "significant outliers" were suitable and acceptable given historical data. Table 27 provides a summary of the results of the various analysis conducted on the conditional logistic regression models. Given our metrics we assess as superior the Latin America – Not in Conflict and OECD – Not in Conflict models due to their overall AUC, accuracy and model fit. Additionally six models are assessed as excellent models, while four models as assessed as satisfactory due to their overall fit of the data.

Table 27: Overall Assessment of Conditional Logistic Regression Models

Overall Assessmnet of Conditional Logistic Regression Models					
Region	**Model**	**Training Data Set AUC**	**Overall Model Accuracy**	**Hosmer-Lemeshow Goodness of Fit Results**	**Overall Model Assessment**
Arab & North African States	In Conflict	0.962	85.3%	Model Apears to fit the data well.	Model is Excellent
	Not in Conflict	0.930	86.7%	Model Apears to fit the data well.	Model is Excellent
Eastern Europe & Central Asia	In Conflict	0.972	89.6%	Model Apears to fit the data well.	Model is Excellent
	Not in Conflict	0.946	84.1%	Model Does Not Fit Data Well	Model is Satisfactory
Latin America	In Conflict	0.878	82.1%	Model Apears to fit the data well.	Model is Excellent
	Not in Conflict	0.952	90.2%	Model Apears to fit the data well.	Model is Superior
OECD	In Conflict	0.914	88.0%	Model Apears to fit the data well.	Model is Excellent
	Not in Conflict	0.974	95.6%	Model Apears to fit the data well.	Model is Superior
South & East Asia	In Conflict	0.938	86.2%	Model Does Not Fit Data Well	Model is Satisfactory
	Not in Conflict	0.932	87.9%	Model Does Not Fit Data Well	Model is Satisfactory
Sub-Saharan Africa	In Conflict	0.889	85.1%	Model Apears to fit the data well.	Model is Excellent
	Not in Conflict	0.874	85.5%	Model Does Not Fit Data Well	Model is Satisfactory

4.3 Analysis of Significant Conflict Transition Variables

While there is significant benefit in accurate prediction of nation-state violent conflicts, many of these benefits are rendered operationally irrelevant without an understanding of the underlying correlation and effects of the significant predictor variables. This analysis seeks to assess the relative importance, based upon p-value, of the specific predictor variables within a model and determine how those variables are correlated with a transition into conflict. Figure 17 provides the basic mapping scheme for covariate correlation based upon correlation type (positive or negative) and magnitude (Dark Green – highly negatively correlated; Dark Red – highly positively correlated).

Additionally in all subsequent analyses, the predictor variables are listed from left to right in terms of statistical significance, based on their p-value, within the model.

	Transition to Conflict (Y)	Negative Correlation to Conflict	Minor Correlation to Conflict	Positive Correlation to Conflict
Transition to Conflict (Y)	1	-0.500	0.000	0.500

Figure 17: Covariate Correlation to Dependent Variable

The operational relevance of this analysis is predicated on identifying variables that can be either monitored or affected in some manner with the goal controlling a nation's transition into or out of conflict. While correlation does not imply causation, this analysis seeks to enable the influencing of the behavior of these large scale regional dynamic systems in a manner beneficial to United States strategic objectives.

Arab & North African States – In Conflict

Of the four statistical variables employed in the Arab and North African States – In Conflict model, ethnic diversity and democratic governments are statistically the most influential variables associated with conflict transitions for Arab nations currently in conflict. As seen in Figure 18, ethnic diversity is negatively correlated with transitions into conflict, implying that increasing a nation's ethnic diversity score (i.e., the percentage of the population made up by the dominant ethnic group) reduces the probability that an Arab nation currently in conflict will remain in conflict. Conversely, the presence of democratic governments is positively correlated to a nation's probability of remaining in a state of violent conflict. While previous studies have suggested that

that risk of conflict is highest among emerging democracies (Goldstone, et al., 2005), the significance of this variable is heavily influenced by the conflicts in Algeria, Lebanon, and Tunisia, the only nations with fully democratic governments within the region during this time period. In all instances, these nations are classified as being in a state of violent conflict, with no observed transitions out of that state.

Analysis of Effects		
Source	Chi-Sq	Prob > Chi-Sq
Ethnic Diversity	9.634	0.002
Regime Type (Democratic)	9.231	0.002
3 Yr Freedom Trend	2.744	0.098
Regime Type (Emerging)	1.751	0.186

	Transition to Conflict (Y)	Ethnic Diversity	Regime Type (Democratic)	3 Yr Freedom Trend	Regime Type (Emerging)
Transition to Conflict (Y)	1	-0.1768	0.2785	-0.0271	-0.0417

Figure 18: Arab & North African States (In Conflict) Covariate Effects

A more accurate appraisal of the effects of regime type within this region can be obtained by comparing the ratio of instances violent conflict by government type. The Arab & North African data set contains 187 total instances, with 58.3%, or 109 observations, of those instances classified as being in state of violent conflict. From this data set, 91 nation-year instances are classified as having Autocratic governments, with the remaining 96 instances classified as having one of the five alternative regime types. Overall the rate of violent conflict in autocratic regimes was 29.7%, 27 total instances, significantly lower than the regional average. However, nations listed as having some other regime type experienced conflict in 85.4% or 82 instances over the 11-year period. The significance of this finding is the correlation between Arab autocratic governments lower probabilities of conflict. Goldstone found similar results in the CIA-funded study, where he found that the risk of instability was lowest in full autocracies (Goldstone, et al., 2005).

Arab & North African States – Not in Conflict

As in the Arab & North African "In Conflict" model, ethnic diversity is identified as the most significant of the six variables employed in this model. However, for nations currently not in conflict, higher ethnic diversity scores are correlated with an increased likelihood of that such a nation will transition into conflict in the following year. Since many of the same nations are present in both the "In Conflict" and "Not in Conflict" data sets, such a finding implies that an imbalance exists in the region's ethnic diversity, exacerbating the overall instability of the region. In addition to ethnic diversity, increased religious diversity scores (% of the population comprised by largest religious group), death rates, and youth populations are correlated with transitions into conflict. Additionally these variables are also positively correlated with each other, indicating likely interdependencies between these predictor variables. On the other hand, greater average life expectancies are correlated to lower incidences of transitions into conflict, though this result may be a function that life expectancies should logically be greater when violent conflicts are not taking place. The summary of variable effects and correlations is provided in Figure 19.

Analysis of Effects		
Source	Chi-Sq	Prob > Chi-Sq
Ethnic Diversity	16.895	<.0001
Religious Diversity	16.471	<.0001
Life Expectancy	16.262	<.0001
Death Rate	14.099	0.0002
Youth Bulge	12.768	0.0004
Military Expend (% GDP)	7.605	0.0058

	Transition to Conflict (Y)	Ethnic Diversity	Religious Diversity	Life Expectancy	Death Rate	Youth Bulge	Military Expend (% GDP)
Transition to Conflict (Y)	1	0.290	0.202	-0.363	0.284	0.258	0.047

Figure 19: Arab & North African States (Not in Conflict) Covariate Effects

Eastern Europe & Central Asia – In Conflict

Analysis of the variables associated with conflict transitions of eastern European and central Asian nations currently identified as being in a state of conflicts identifies a nations international trade level, as a percentage of it gross domestic product (GDP) as the most significant with the model. Trade is identified as being negatively correlated with a state remaining in conflict, an expected result given that stable and less violent nations should have higher levels of international trade. Similar to other regional models, population statistics (specifically those correlated with increased youth populations high densities) are correlated with increased incidences of transitions into conflict. As seen in Figure 20, fertility rates, infant mortality rates, and population density are all positively correlated with transitions into conflict, and with each other. This finding indicates a reduction in one of the variables, such as “Fertility Rate”, may result, over time, in subsequent decreases in a nation’s infant mortality rate, population density or both, with a corresponding decrease in the probability of violent conflict.

Analysis of Effects		
Source	Chi-Sq	Prob > Chi-Sq
Trade (% GDP)	40.106	<.0001
Fertility Rate	24.026	<.0001
Infant Mortality rate	11.693	0.0006
Population density	11.361	0.0008
Freedom Score	6.395	0.0114

	Transition to Conflict (Y)	Trade (% GDP)	Fertility Rate	Infant Mortality rate	Population density	Freedom Score
Transition to Conflict (Y)	1	-0.501	0.152	0.171	0.166	0.033

Figure 20: Eastern Europe & Central Asia (In Conflict) Covariate Effects

Eastern Europe & Central Asia – Not in Conflict

A total of nine variables were identified as significant for eastern European and central Asian nations current in a state of non-conflict. Of note is the significance associated with regime type, specifically those governments identified as either emerging democracies, or experiencing foreign interruption of their political processes, which is given in Figure 21. As noted earlier, the existence of transitional or emerging governments is highly correlated with violent conflict, which makes logical sense due to the loss of government function and continuity. As was the case for the Arab nation models, this finding is only part of story. For the period of 2004 to 2014, there are 308 total instances in the Eastern Europe & Central Asian data set; of these 125 instances (40.6%) are identified as being in a state of conflict. However, unlike the Arab and North African models, democratic nations, within the region are less likely to be in state of violent conflict. Of this subset, only 45 (25.6%) of the 176 instances involving democratic governments were identified as being in a state of conflict. Further analysis revealed that of the 16 nations identified as having democratic governments, only Pakistan is located outside of Eastern Europe, indicating that government type may not provide the operational fidelity required for conflict prediction and forecasting within this region.

	Transition to Conflict (Y)	Government Type (Foreign Interruption)	3 Yr Freedom Trend	Government Type (Emerging)	Improved Water	GDP Per Capitia	2 Yr Conflict Intensity Trend	Religious Diversity	Arable Land	2 Yr Freedom Trend
Transition to Conflict (Y)	**1**	**0.254**	**0.167**	**0.447**	**-0.218**	**-0.422**	**-0.182**	**0.162**	**0.095**	**0.077**

Analysis of Effects		
Source	**Chi-Sq**	**Prob > Chi-Sq**
Government Type (Foreign Interruption)	20.532	<.0001
3 Yr Freedom Trend	18.373	<.0001
Government Type (Emerging)	16.741	<.0001
Improved Water	15.752	<.0001
GDP Per Capitia	15.654	<.0001
2 Yr Conflict Intensity Trend	11.392	0.0007
Religious Diversity	10.846	0.001
Arable Land	9.767	0.0018
2 Yr Freedom Trend	7.241	0.0071
Government Type (Democratic)	0.235	0.6282

Figure 21: Eastern Europe & Central Asia (Not in Conflict) Covariate Effects

Of the remaining variables, access to improved water sources and the GDP per Capita were both highly significant and negatively correlated with transitions into conflict. Within this model, these variables represent likely candidates that can be monitored, manipulated and improved through the judicious application of the diplomatic, information, military, and economic elements of national power, resulting in a possible reduction in the total number of future transitions into conflict.

Latin America – In Conflict

Figure 22 provides the covariates for this model. Non-autocratic functioning governments are highly correlated with increased levels of violence in Latin American nations, with 95% of the conflict incidences occurring in these nations. Fully democratic nations account for 21 of the 27 nations within the Latin American data set and subsequently account for a majority of the conflict transitions that occur within the region. However, nations identified as having emerging democratic governments, such as Ecuador, Suriname, or Venezuela are nearly twice as likely to remain in conflict as their fully democratic neighbors. Increased religious diversity and freedom scores are correlated with transitions out of violent conflict, indicating that increasing the percentage of the population made up by the religious majority or increasing individual liberties may result in increased incidences of transitions to a non-conflict state. The CIA-funded study yielded similar results showing that increased factionalism due to ethnic and religious diffrences was positively correlated with political instability (Goldstone, et al., 2005).

Analysis of Effects		
Source	**Chi-Sq**	**Prob > Chi-Sq**
Government Type (Democratic)	12.796	0.000
Government Type (Emerging)	11.637	0.001
Government Type (Anarchy)	10.305	0.001
Religious Diversity	9.480	0.002
Freedom Score	1.775	0.183

	Transition to Conflict (Y)	Government Type (Democratic)	Government Type (Emerging)	Government Type (Anarchy)	Religious Diversity	Freedom Score
Transition to Conflict (Y)	1	0.249	0.162	0.190	-0.356	-0.021

Figure 22: Latin America (In Conflict) Covariate Effects

Latin America – Not in Conflict

As shown in Figure 23, nations currently not in conflict with higher ethnic diversity scores tend to experience few transitions into conflict than nations with more diverse populations. However, ethnic diversity is positively correlated with religious diversity, which is shown to have a moderate destabilizing effect for countries not in conflict. Similar to the Arab and North African nations, there appears to be an imbalance with regards to the region's ethnic and religious demographics that may aggravate regional discord. On the other hand, access to improved water sources appears to be positively correlated to fewer transitions into violent conflict. However, this finding may also be the result of a more permissive environment allowing for improved access to fresh water.

Analysis of Effects		
Source	**Chi-Sq**	**Prob > Chi-Sq**
Ethnic Diversity	29.473	<.0001
Improved Water	23.991	<.0001
Religious Diversity	22.436	<.0001
Regime Type (Democratic)	18.764	<.0001
Infant Mortality rate	13.266	0.0003
Fertility Rate	7.781	0.0053
Birth Rate	5.701	0.017

	Transition to Conflict (Y)	Ethnic Diversity	Improved Water	Religious Diversity	Regime Type (Democratic)	Infant Mortality rate	Fertility Rate	Birth Rate
Transition to Conflict (Y)	1	-0.179	-0.369	0.288	0.132	0.274	0.375	0.359

Figure 23: Latin America (Not in Conflict) Covariate Effects

OECD – In Conflict

Nations belonging to the Organization for Economic Cooperation and Development (first world nations) experience violent conflict rates 50% below the world average. However, like other regions, increased youth populations within OECD nations are correlated with increased levels of violence and the tendency for nations to transition or remain in a state of conflict. While not identified as a significant variable within the final "In Conflict" model, population migrations represented by the two "Refugee" variables are correlated with transitions into conflict as well as increased youth populations and military expenditures within OECD nations. With regard to population migrations, historically refugees are 2.2 times more likely to seek asylum in an OECD nation than originate from one. According the 2014 HIIK Conflict Barometer, conflicts arising from population migrations have resulted in, or contributed to, many of the violent conflicts experienced by OECD nations, with noted examples being the ongoing immigration and border conflict between the United States and Mexico, violence associated with Refugee and immigrant populations in France, and the ongoing Refugee crisis along Turkey's southern borders with Iraq and Syria. The summary of the covariate for both OECD models are given in Figures 24 and 25.

Analysis of Effects		
Source	Chi-Sq	Prob > Chi-Sq
Youth Bulge	9.366	0.002
Military Expend (% Gov Spending)	6.600	0.010
Caloric Intake	5.733	0.017

	Transition to Conflict (Y)	Youth Bulge	Military Expend (% Gov Spending)	Caloric Intake
Transition to Conflict (Y)	1	0.289	-0.159	0.335

Figure 24: OECD (In Conflict) Covariate Effects

OECD – Not in Conflict

As in the "In Conflict" model, defense expenditures and youth populations are considered significant predictors of conflict transitions for nations currently not in a state of conflict. Again, population migrations are highly correlated to many of the significant variables within this model, underpinning the importance of this emerging global trend in national and regional stability and security. Common to all regions, improvements in the overall quality of life, measured through proxy variables such as death rates and average life expectancy are correlated with decreased levels of violence, even if such predictor variables are not identified as significant within the final model(s).

Analysis of Effects		
Source	Chi-Sq	Prob > Chi-Sq
Military Expend (% GDP)	21.345	<.0001
Infant Mortality rate	14.897	0.000
Caloric Intake	13.821	0.000
Youth Bulge	8.893	0.003
Death Rate	8.375	0.004
Birth Rate	4.445	0.035

	Transition to Conflict (Y)	Military Expend (% GDP)	Infant Mortality rate	Caloric Intake	Youth Bulge	Death Rate	Birth Rate
Transition to Conflict (Y)	1	0.387	0.382	-0.145	0.231	-0.214	0.173

Figure 25: OECD (Not in Conflict) Covariate Effects

South & East Asia – In Conflict

As shown in Figure 26, increased levels of population growth are correlated with transition out of conflict. This relatively counterintuitive finding is correlated with improvements in overall quality of life and influenced by many of the island nations within the Pacific that have higher population growth percentages and decreased levels of violence than many of the mainland and coastal Asian nations. Government type is considered a highly significant predictor variable within this region, with democratic governments experiencing rates of conflict above regional averages. However, unlike other regions, fully autocratic governments do not offer significant improvements to out-of-conflict transition rates, and they seem as likely to perpetuate ongoing conflicts as any other government type. Finally, as seen in other regional models, increasing trade levels is correlated with decreased levels of violence, and it is positively correlated with military expenditures which may also bring about transitions out of conflict.

Analysis of Effects		
Source	Chi-Sq	Prob > Chi-Sq
Population Growth	23.784	<.0001
Government Type (Democratic)	14.765	0.000
Ethnic Diversity	14.703	0.000
Military Expend (% GDP)	8.329	0.004
Trade (% GDP)	6.971	0.008
Government Type (Emerging)	4.468	0.035

	Transition to Conflict (Y)	Population Growth	Government Type (Democratic)	Ethnic Diversity	Military Expend (% GDP)	Trade (% GDP)	Government Type (Emerging)
Transition to Conflict (Y)	1	-0.232	0.233	-0.130	-0.164	-0.257	0.027

Figure 26: South & East Asia (In Conflict) Covariate Effects

South & East Asia – Not in Conflict

As shown in the "In Conflict" model, increases in military expenditures are affiliated with transitions in non-conflict statuses for all nations within South and East Asia. This variable which is positively correlated with a nation's trading ability may result in improvements to internal security apparatuses within many of these nations resulting in decreased levels of violence. However, the ten-year trend within the region has shown a general increase in military spending, for all nations, which may indicate developing arms race, with the potential of increased cross border conflicts. Previous studies, notably the Boekestein study, have also identified the significance of trade, caloric intake, and refugee migrations as conflict predictor variables within South and East Asia. Additionally improvements in overall quality of life, measured through proxy variables such as death rates and life expectancy, are positively correlated with improvements and access to food supplies and potable water. The covariate correlations for this conditional model are provided in Figure 27.

Analysis of Effects		
Source	**Chi-Sq**	**Prob > Chi-Sq**
Fresh Water per Capita	9.052	0.003
Military Expend (% GDP)	8.910	0.003
Caloric Intake	7.030	0.008
Arable Land	6.697	0.010
Refugee (Origin)	5.165	0.023

	Transition to Conflict (Y)	Fresh Water per Capita	Military Expend (% GDP)	Caloric Intake	Arable Land	Refugee (Origin)
Transition to Conflict (Y)	1	-0.160	-0.202	-0.297	-0.036	0.369

Figure 27: South & East Asia (Not in Conflict) Covariate Effects

Sub-Saharan Africa – In Conflict

Population demographics positively correlated to increases in population density, such as increases in youth populations, birth rates, and refugees appear to exacerbate and prolong existing conflicts in Sub-Saharan African nations. It also appears that populations increased diversity, due to predominately tribal cultures found in these nations, are more at risk for violent conflict than those nations with higher ethnic diversity scores. Again, improvements in quality of life statistics, in this case available fresh water and life expectancy, are correlated with out of conflict transitions. Over the 11-year period Sub-Saharan Africa experience conflict in 253 (47%) of the 539 observed instances. Government type was identified as being significant with this conditional model. Predominantly, Sub-Saharan African governments are categorized as either emerging democracies (23 nations) or full democracies (22 nations), with only Eritrea and Swaziland identified as having fully autocratic governments as of 2014. Within this region, emerging democracies are twice as likely to experience violent and sustained conflicts as fully democratic nations, most likely associated with the inherent instability of their governments. Figure 28 provides the covariate effects for the Sub-Saharan Africa-In Conflict model.

Analysis of Effects		
Source	Chi-Sq	Prob > Chi-Sq
Refugee (Origin)	17.041	<.0001
Youth Bulge	15.891	<.0001
Fresh Water per Capita	13.408	0.000
Ethnic Diversity	13.113	0.000
Life Expectancy	10.855	0.001
Birth Rate	9.641	0.002
Government Type (Democratic)	9.151	0.003
Government Type (Emerging)	6.478	0.011
Military Expend (% Gov Spending)	3.753	0.053
Government Type (Anarchy)	2.340	0.126
Government Type (Transition)	1.047	0.306

	Transition to Conflict (Y)	Refugee (Origin)	Youth Bulge	Fresh Water per Capita	Ethnic Diversity	Life Expectancy	Birth Rate	Government Type (Democratic)	Government Type (Emerging)	Military Expend (% Gov Spending)
Transition to Conflict (Y)	1	0.200	0.201	-0.297	-0.142	-0.127	0.172	-0.017	0.042	0.096

Figure 28: Sub-Saharan Africa (In Conflict) Covariate Effects

Sub-Saharan Africa – Not in Conflict

As shown in Figure 29, and other regions, arable land appears to be a source of instability and confounding factor for conflict transitions. Historical and recent conflicts over arable land have generally arisen due to either actual or perceived scarcity of the resource, with the general conclusion being that limited availability and access to arable land leads to conflict (Black, 2010). However, as is other regional models, arable land is identified as being positively correlated to violent conflict, implying that increasing the supply of this resource will lead to increased levels of violence, which is contradictory to previous studies. Analysis of this and other regional models has shown that arable land is also positively correlated with such statistics as increased population densities, youth populations, and increased number of refugees seeking asylum, all of which have demonstrated a positive correlation to instance of violent conflict across the globe. Essentially, it appears that nations in Sub-Saharan Africa with increased food production capacities are at a moderately greater risk of violent conflict associated with a corresponding increase in their populations due to procreation and migration.

Analysis of Effects		
Source	Chi-Sq	Prob > Chi-Sq
Freedom Score	16.268	<.0001
Arable Land	15.462	<.0001
Trade (% GDP)	12.283	0.001
Birth Rate	10.673	0.001
Infant Mortality rate	9.290	0.002
Youth Bulge	7.406	0.007
Refugee (Asylum)	5.137	0.023

	Transition to Conflict (Y)	Freedom Score	Arable Land	Trade (% GDP)	Birth Rate	Infant Mortality rate	Youth Bulge	Refugee (Asylum)
Transition to Conflict (Y)	1	-0.207	0.279	-0.201	0.121	0.189	0.107	0.095

Figure 29: Sub-Saharan Africa (Not in Conflict) Covariate Effects

Summary

A total of 30 variables, including the different levels of the Government and Regime type variables, were employed in the construction of the 12 conditional logistic regression models. Table 28 provides the ranking of variables in terms of statistical significance for each conditional model, with variables listed in terms of overall world view significance. Ethnic diversity, youth bulge, military expenditure by percentage of GDP, infant mortality rate, and religious diversity were identified as the five ordinally most significant variables at the combined world level, based upon their weighted average rankings. Studies conducted by the Peace Research Institute of Oslo (PRIO), also found similar variables highly significant, lending credence to this finding (Urdal, 2002). Ethnic diversity, which is significant in 5 of the 12 logistic regression models, and is the single most significant variable in Arab and North African states, is negatively correlated to nations transitioning into or remaining in conflict. Additionally, increased youth populations which are also significant in five models, are positively correlated to increased levels of violence. Of note, the various levels of the government and regime type variables were identified as statistically significant in 6 of the 12 models. "In Conflict" models tend to employ government or regime type variables more frequently than "Not in Conflict" models.

Table 28: Ranking of Variables in terms of Model Statistical Significance

Varaible / Region	Arab & North Africa		Eastern Europe & Central Asia		Latin America		OECD		South & East Asia		Sub-Saharan Africa	
	In Conflict	Not In Conflict	In Conflict	Not In Conflict	In Conflict	Not In Conflict	In Conflict	Not In Conflict	In Conflict	Not In Conflict	In Conflict	Not In Conflict
Ethnic Diversity	1	1				1			3		4	
Youth Bulge		5					1	4			2	6
Military Expend (% GDP)		6						1	4	2		
Infant Mortality rate			3			5		2				5
Religious Diversity		2		7	4	3						
Trade (% GDP)			1						5			3
Caloric Intake							3	3		3		
Fresh Water per Capita										1	3	
Government Type (Foreign Interruption)				1								
Population Growth									1			
Government Type (Democratic)					1				2		7	
Government Type (Emerging)				3	2				6		8	
Freedom Score			5		5							1
3 Yr Freedom Trend	3			2								
Birth Rate						7		6			6	4
Regime Type (Democratic)	2					4						
Improved Water				4		2						
Refugee (Origin)										5	1	
Arable Land				8						4		2
Life Expectancy		3									5	
Fertility Rate			2			6						
Death Rate		4						5				
Military Expend (% Gov Spending)							2				9	
Government Type (Anarchy)					3							
Regime Type (Emerging)	4											
Population density			4									
GDP Per Capitia				5								
2 Yr Conflict Intensity Trend				6								
Refugee (Asylum)												7
2 Yr Freedom Trend				9								

4.4 Analysis of Nation Specific Markov Models

Overview

As stated in Chapter 3, the use of Markov models is intended as an operationally relevant forecasting model of future conflict trends conditioned on whether a nation is or is not currently in a state of violent conflict. Conditional probabilities for each nation are calculated using both the "In Conflict" and "Not in Conflict" models on the 2014 data set, which is the base year for all Markov models. A Visual Basic (VBA) based Markov model tool, operating in the Microsoft Excel environment, was developed to generate the required outputs and aid in the analysis of future conflict trends. In addition to calculating the conditional probabilities for any future year, this tool also calculates the sojourn times, mean conflict recurrence times, and long-run conflict probabilities specific to each of the 182 nations included in this study.

Model Validation

Analysis of Expected Number of States in Conflict for 2014

As part of a higher level analysis and validation of the conditional conflict probabilities, this study compared the global and regional incidence of violent conflict observed by HIIK, with the expected number of nations in conflict determined using the conditional probabilities calculated by the logistic regression models using 2014 conflict data, and a 0.50 cut point. This analysis does not seek to specifically identify which nations are in conflict for a particular region, but rather provide the expected incidence level by region that can be compared to current global trends. A summary of this comparison, by region, is provided in Table 29.

Table 29: Comparison of HIIK Observed and Expected Incidences of Conflict using a 0.50 Cut Point for 2014.

Region	HIIK Observed States Not in Conflict	Expected Number of States not in Conflict	HIIK Observed States in Conflict	Expected Number of States in Conflict
Arab & North African States	3	4.44	14	12.56
Easter Eurpoe & Central Asia	17	16.57	11	11.43
Latin America	13	14.00	14	13.00
OECD	26	25.99	7	7.01
South & East Asia	15	13.57	13	14.43
Sub-Saharan Africa	24	27.64	25	21.36
World View	98	102.21	84	79.79

The 2014 conditional conflict probabilities that are subsequently used to develop the nation specific Markov models for this study predict approximately 80 nations experiencing violent conflict and 102 nations remaining out of conflict, and an expected conflict incidence rate of 43.8%. The observed incidence of conflict for 2014 had 84 nations experiencing some level of violent conflict, with 98 nations remaining in a state of no conflict, resulting in an observed conflict incidence rate of 46.2%. At the regional level, the absolute difference in the observed and expected incidence of violent conflict was less than 1.45 in five of the six regions, and 3.64 in the Sub-Saharan Africa region. The Arab and North African States, followed by Latin America, and South and East Asia can be expected to experience conflict rates of 50% or greater. Conversely, the conflict incidence rates for OECD nations are less than half the world average at 21.3%. Overall, the conditional models provide a very accurate prediction of the 2014 conflict incidence rates of each region, and the world as a whole.

Forecasting Validation

The validation of the nation specific Markov models presented an interesting challenge due to the inability to foresee all future events with 100 percent certainty. As a result, we looked to the past to develop a validation set to compare against the Markov models using conditional probabilities calculated using 2014 conflict data. To validate our 2014 Markov models, we construct another set of Markov models having conditional probabilities calculated using 2011 conflict data; these model are subsequently known as the 2011 Markov Models. This set of 2011 Markov models subsequently forecasts the 2014 conflict probabilities, which are then compared to the conflict probabilities calculated using 2014 conflict data to assess the level of deviation between the two models. The 2011 year set was selected for validation purposes due to it containing nearly all the data used in the construction of the logistic regression models and its relatively recent timeframe that more closely resembles conditions present in the 2014 operational environment. The purpose of this validation is to ascertain the fidelity of the 2014 conditional conflict probabilities by comparing their deviations from the 2014 probabilities predicted using 2011 conflict data. The deviation is calculated using Equation 27.

$$\text{Deviation (Not in Conflict)} = \left| {}_{2011}P_{0,0}^{3} - {}_{2014}P_{0,0}^{0} \right|$$
$$\text{Deviation (In Conflict)} = \left| {}_{2011}P_{1,1}^{3} - {}_{2014}P_{1,1}^{0} \right|$$
$$\text{Average Deviation} = \frac{\left[\text{\% Deviation (Not in Conflict)} + \text{\% Deviation (In Conflict)}\right]}{2}$$

Equation 27: Markov Validation

These equations were applied to the 2011 and 2014 Markov models for all 182 nations considered in this study. The validation process then analyzed to statistics for the entire set of models, which are provided in Table 30. On average, the difference between the 2014 Markov models and the 2011 Models predicting 2014 was 0.12 with a variance of 0.016. Additionally, a total of 152 of the 182 models had average difference less than 0.25. Only the Ukrainian model experiences deviations greater than 0.50 for both the "Not in Conflict" and "In Conflict" conditional probabilities; this result is attributed to the ongoing conflicts in Crimea that significantly escalated in intensity in late 2013 and early 2014, and is considered reasonable.

Table 30: Markov Model Validation Statistics

Category	Average Difference Between Models	Variance	Number ≤ 0.05	Number ≤ 0.10	Number ≤ 0.25	Number ≤ 0.50	Number > 0.50
Non-Conflict Deviation	0.1227	0.0369	100	22	32	16	12
Conflict Deviaition	0.1211	0.0297	100	20	27	26	9
Average Model Deviation	0.1219	0.0158	69	32	51	29	1

Additionally, Markov model accuracy was assessed by comparing the 2014 conflict forecasts created by the Markov models developed using 2011 data. In total the 2011 Markov models correctly classified the conflict status of 155 of the 182 nations, for a total forecast accuracy of 85.16%. Given the high number of nation models that experience average deviations less than 25%, and the number of significant events that have occurred since 2011 (The Arab Spring, the Rise of the Islamic State, Crimean conflict, etc.), the 2014 Markov models appear as valid representations of current conflict transition probabilities.

Analysis of HIIK Conflict Intensity Levels and Conflict Probability

As part of the model validation process, this study analyzed the conditional conflict probabilities as they relate to the HIIK levels of violence. The theory behind this analysis is that there should exist a strongly positive correlation between a nation's conditional probability of conflict and its level of violence in 2014. As part of this analysis, the HIIK levels of violence were mapped to the corresponding ranges of probabilities shown in Table 31, with the assumption that the HIIK levels of violence are linear and well scaled.

Table 31: HIIK Intensity Bin Assignments

HIIK Intesity Level Bin Assignments		
HIIK Bin	> Lower Bound	≤ Upper Bound
0	0.000	0.167
1	0.167	0.333
2	0.333	0.500
3	0.500	0.667
4	0.667	0.833
5	0.833	1.000

Nations are then assigned to a HIIK bin based upon their assigned conditional conflict probability. The average HIIK score, based upon the nations' actual conflict intensity for 2014, is then calculated for each bin as shown in Figure 30.

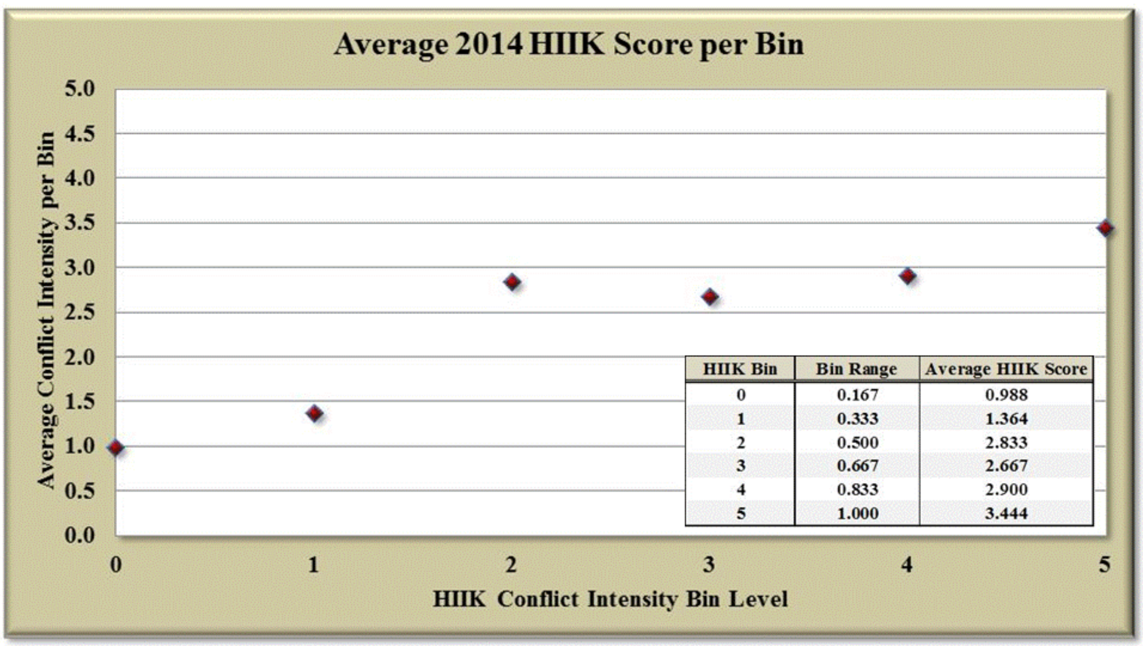


Figure 30: Average HIIK Conflict Intensity Levels by Bin

As can be seen, the average HIIK score is positively correlated with its bin assignment; with a calculated correlation of *0.731*. However the average HIIK score does not strictly increase over the entire bin range, noted by the decrease from Bin 2 to

Bin 3. The decrease in the average HIIK score for Bin 3 may be the result of an outlier(s) that may significantly decrease the average score within the bin. Such points would be significant false-positives or false-negatives; nations assigned either a very low or very high conditional conflict probability in respect to its actual level of violence. Figure 31 provides a visual depiction of bin assignments versus conflict intensity for 2014.

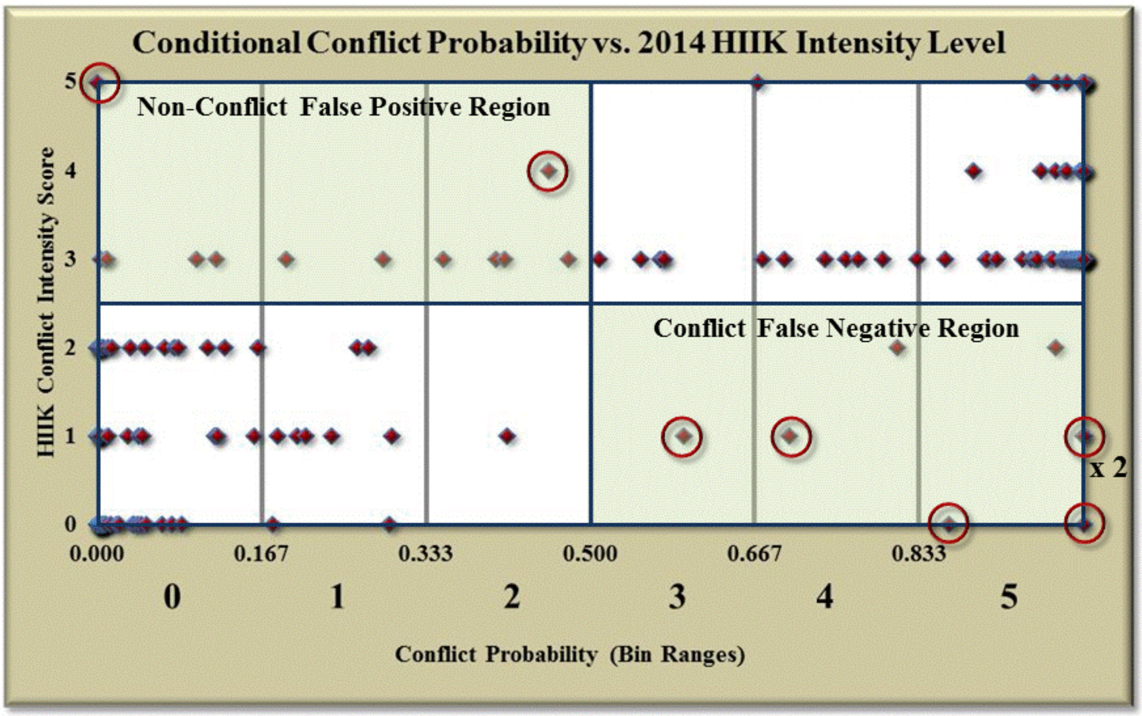


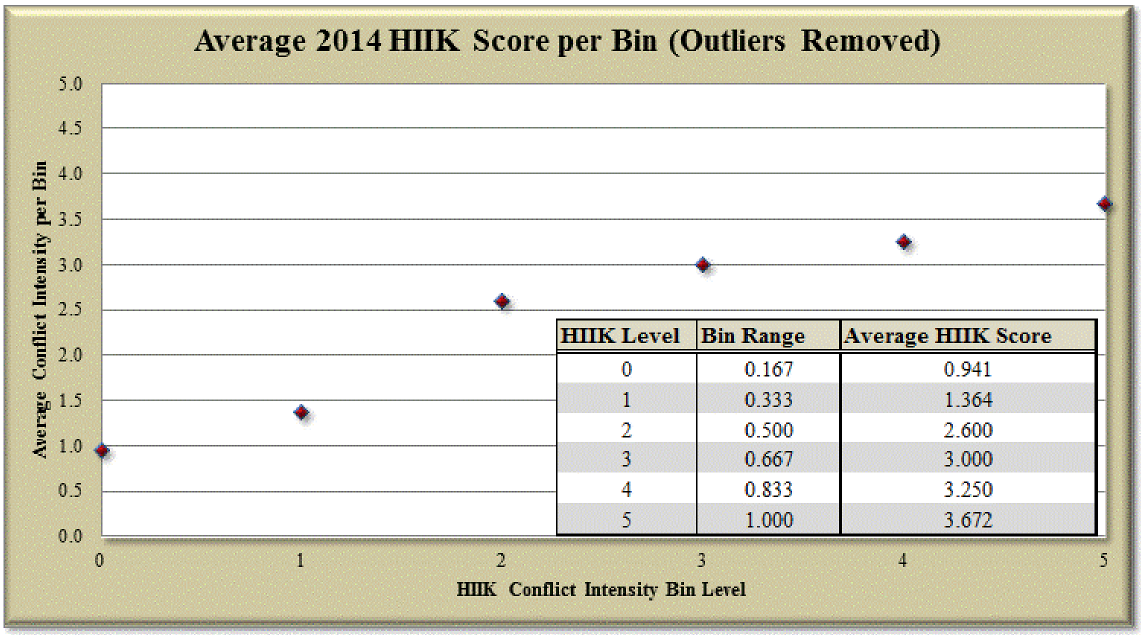
Figure 31: Identification of Significant Outliers by HIIK Bin

A total of eight possible significant outliers were initially identified, based upon having an absolute deviation in the HIIK conflict intensities and assigned bins greater than or equal to two (with the exception points having a HIIK level of 3); these point are marked by the circles in Figure 31. The identified outliers consist of: Libya (Bin 0), Egypt (Bin 2), Cameroon (Bin 3), Panama (Bin 4), Kiribati (Bin 5), Qatar (Bin 5), Oman (Bin 5), and the United Arab Emirates (Bin 5). A formal outlier analysis was conducted to verify these outliers, and identify other potential outliers within the data. Outliers were identified through examination of the scaled "R-studentized" residuals, a method commonly used in linear regression to identify extreme points that are considerably different from a majority of the data (Montgomery, Peck, & Vinning, 2012). This process identified the nine nations listed in Table 32 as being possible significant outliers in their respective models if they fail to transition from their 2014 conflict status by 2015.

Table 32: Significant Outliers, 2014

Nation	HIIK Bin	Conflict Probability	HIIK Intensity Level
Gabon	0	0.004	3
Kyrgyzstan	0	0.008	3
Tunisia	0	0.009	3
Vietnam	0	0.100	3
Libya	0	0.000	5
Kiribati	5	0.863	0
Qatar	5	1.000	0
Oman	5	1.000	1
United Arab Emirates	5	1.000	1

A majority of the these outlier nations are from the Arab & North African States region, with Tunisia and Libya identified as possible significant false-negative classifications; and Qatar, Oman, and the United Arab Emirates (UAE) identified as possible significant false-positive classifications. This result further highlights the extreme instability within the region and the effects of the Arab Spring, hindering conflict transition analysis. Removal of these outliers results in the plot provided in Figure 32. Comparison of this plot with that shown in Figure 30, shows an improvement in the overall linearity of the plot, and the expected positive correlation associated the average conflict intensity and HIIK bin level. Such a result indicates that we have identified all significant outlier nations for 2014. Ultimately the objective of this analysis is the identification, as opposed to the removal, of possible significant misclassified nations for the purpose of monitoring both the Markov model outputs and future conflict status for consistency and accuracy.



HIIK Level	Bin Range	Average HIIK Score
0	0.167	0.941
1	0.333	1.364
2	0.500	2.600
3	0.667	3.000
4	0.833	3.250
5	1.000	3.672

Figure 32: Average HIIK Conflict Intensity Levels with Outliers Removed

Overall Assessment of Markov Models

Analysis of overall suitability and validity of the nation-specific Markov models has demonstrated that the tool functions properly and provides accurate calculations based on the logistic regression model inputs. The validity of the 2014 logistic regression model inputs was verified through comparison of 2011 conditional probabilities predicting the 2014 conflict transition probabilities. This comparison yielded satisfactory results with 83.5% of all nations experiencing absolute deviations in respective conflict probabilities less than 0.25. The conditional probabilities were then compared with the

HIIK conflict intensity levels, identifying a positive correlation associated with conflict probability and a nation's observed conflict intensity for 2014. As part of this comparison, analysis of significantly misclassified outliers identified only nine nations whose models may produce faulty or inaccurate forecasts, based on 2014 predictions, and may require further refinement in future studies. Finally, comparison of the expected and observed regional incidences of conflict indicated a high level of accuracy in the models' ability predict regional levels of violence, further substantiating the suitability of the models for forecasting future conflict trends.

Key Markov Model Outputs

The objective of the nation-specific Markov models is to provide operationally relevant insights on future conflict trends. In addition to conflict forecasts, which will be discussed at length in the following section, this study also seeks to determine the sojourn times, long-run conflict probabilities, and mean conflict recurrence times for each nation. It should be understood that these calculations are predicated on the assumption that current conditions regarding the 2014 independent conflict variable remain unchanged within each region, and that the forecasted trends may be altered through the application of national power, Black Swan events (Taleb, 2010), or both. The complete table of Markov model outputs is provided in Appendix C.

Sojourn times $E[R_j]$ are simply time expected for a nation's n^{th} conflict transition. For this study, we examine the first and second sojourn times, and their respective variances, for each nation beginning in the base year 2014. An example of first and second Sojourn times, as well as the 2014 Markov model, is provided in Figure 33.

Conflict Tranistion Probability Markov Chain Tool					Sojourn Times (R_i)	
	Number of Years into Future =		5			
4	Country	Year	2014		Current Status:	No Conflict
	Angola				Number of Years to 1st transition:	4.21
			No Conflict	Conflict	Variance	13.50
Status:	No Conflict	No Conflict	0.7623733	0.237627	Number years to 2nd Transition	21.70
		Conflict	0.0571556	0.942844	Variance	302.12

Figure 33: Example of First and Second Sojourn Times

In this example, Angola's Markov 2014 model indicates that the nation is more likely to be in a state of conflict, due to the highly likelihood (94%) that once Angola enters into a state of conflict, it will remain in that state the following year. This tendency is subsequently reflected in Angola's sojourn times. Given that Angola was in a state of non-conflict in 2014, it is calculated that Angola will experience its first transition into conflict in approximately 4.21 years with a standard deviation of approximately 4 years. It is therefore likely that Angola will transition into conflict within the next 8 years. However, as stated previously, once Angola enters into a state of conflict it is predicted to remain in that state for approximately 18 years. The second sojourn time, in this case the time for Angola to transition back into a state of non-conflict, is simply the sum of its first sojourn time and its expected time to remain in conflict, and is calculated to be approximately 21.7 years from 2014, with a standard deviation in the expected second sojourn time of approximately 17 years. The increased variance associated with this standard deviation can subsequently be equated to an higher levels of risk, in terms of model accuracy, due to a prolonged prediction horizon.

While Angola’s sojourn times are representative of many of the nations included in this study, numerous nations have predicted sojourn times that span hundreds if not thousands of years. This phenomena, is due to nations having an overwhelming tendency, as of 2014, to remain in one state or another. Figure 34 provides an example of significantly long sojourn times for Canada, which is primarily in a state of non-conflict, and the Central African Republic, which is predicted to spend an vast amount of time in a state of conflict.

Conflict Tranistion Probability Markov Chain Tool					Sojourn Times (R_i)	
	Number of Years into Future =	5				
31	Country	Year	2014		Current Status	No Conflict
	Canada				Number of Years to 1st transition:	3.39E+05
			No Conflict	Conflict	Variance	1.15E+11
Status:	No Conflict	No Conflict	0.9999971	2.95E-06	Number years to 2nd Transition	3.39E+05
		Conflict	0.0507637	0.949236	Variance	1.15E+11
32	Country	Year	2014		Current Status:	Conflict
	Central African Republic				Number of Years to 1st transition:	4.55E+11
			No Conflict	Conflict	Variance	2.07E+23
Status:	Conflict	No Conflict	0.0285233	0.971477	Number years to 2nd Transition	4.55E+11
		Conflict	2.198E-12	1	Variance	2.07E+23

Figure 34: Example of Significantly Long Sojourn Times

As can be seen, the expected time for Canada to transition into a state of conflict is approximately 339,000 years, indicating a significant preference towards non conflict. Similarly the Central African Republic shows an even greater predilection to remain in state of conflict based on the 2014 model. The significantly large variances, for these and similar nations, are functions of the extreme time horizons associated with their sojourn times and indicate that a transition can occur any time within the forecast window. The operational relevance of these significantly long sojourn times is the insight that certain nations are not expected to experience a conflict transition within the foreseeable future if 2014 conditions remain unchanged.

The long-run proportion (π_j) of time a nation spends either in a state of conflict or non-conflict is an indicator of the transience, the tendency to transition in or out of conflict, of a nation. Again the operational relevance of this statistic is the identification of nations that either tend to be in one state or the other, as well as nations that have the experience frequent conflict transitions. In total, 95 nations have long run probabilities that indicate a tendency for violent conflict, while 87 nations have long run probabilities that indicate a predisposition for non-conflict. An example of these three categories of long-run conflict probabilities is presented in Figure 35.

Conflict Tranistion Probability Markov Chain Tool					**Long Run Conflict Probabilites (π_j)**	
	Number of Years into Future =	**5**				
35	**Country**	Year	2014			
	China					
			No Conflict	Conflict	Probability Not in Conflict	Probability in Conflict
Status:	**Conflict**	No Conflict	0.9112038	0.088796	0.002596693	0.997403307
		Conflict	0.0002312	0.999769		
36	**Country**	Year	2014			
	Colombia					
			No Conflict	Conflict	Probability Not in Conflict	Probability in Conflict
Status:	**Conflict**	No Conflict	0.9001821	0.099818	0.528059099	0.471940901
		Conflict	0.1116872	0.888313		
37	**Country**	Year	2014			
	Comoros					
			No Conflict	Conflict	Probability Not in Conflict	Probability in Conflict
Status:	**No Conflict**	No Conflict	0.9808778	0.019122	0.859574436	0.140425564
		Conflict	0.1170507	0.882949		

Figure 35: Long Run Conflict Probabilities

As can be seen, China's long-run probability indicates that China is expected to be in conflict 99.7% of the time, a nearly permanent state of conflict, that is reinforced by its violent history and ongoing internal conflicts. To a lesser extent, Comoros is expected to be in a state of non-conflict nearly 86% of the time, indicating that the nation has some region-specific conflict risk factors but possesses a level of stability that limits the overall incidence of violent conflict. Columbia, however, has long-run conflict probabilities that predict the nation will spend nearly equal amounts of time in and out of conflict, equating to a high conflict transience rate. Transience, which will be discussed in depth in the following section, may be an indicator of a nation's susceptibility to both internal and/or external forces resulting in a nation's transition from one conflict status to the other.

Long-run conflict probabilities can be translated into the mean conflict status recurrence (m_j), or the average number of steps a nation requires to return to its current state. As shown in Equation 24, the mean recurrence time is calculated by simply taking the inverse of the long-run conflict probability. While similar to sojourn time, the mean recurrence is the long run average of conflict transition steps, and it represents the predicted number of steps a nation can expect to experience in order to return to either a state of conflict or non-conflict. Figure 36 provides the mean recurrence steps that correspond to the long-run probabilities for China, Columbia and Comoros.

Conflict Tranistion Probability Markov Chain Tool						Mean Conflict Recurrence (M_i)	
	Number of Years into Future =	5					
35	Country	Year		2014			
	China						
				No Conflict	Conflict	Expected Steps to Transition from No Conflict to No Con	385.11
Status:	Conflict		No Conflict	0.9112038	0.088796	Expected Steps to Transition from Conflict to Conflict	1.00
			Conflict	0.0002312	0.999769		
36	Country	Year		2014			
	Colombia						
				No Conflict	Conflict	Expected Steps to Transition from No Conflict to No Co	1.89
Status:	Conflict		No Conflict	0.9001821	0.099818	Expected Steps to Transition from Conflict to Conflict	2.12
			Conflict	0.1116872	0.888313		
37	Country	Year		2014			
	Comoros						
				No Conflict	Conflict	Expected Steps to Transition from No Conflict to No Con	1.16
Status:	No Conflict		No Conflict	0.9808778	0.019122	Expected Steps to Transition from Conflict to Conflict	7.12
			Conflict	0.1170507	0.882949		

Figure 36: Example of Mean Recurrence Times

Corresponding to an overwhelming probability of remaining in conflict, the predicted non-conflict recurrence (m_0), in China is approximately 385 steps, equating an extremely low transience rate. However, when China does enter into a state of non-conflict, it is expected that the nation will transition back into conflict within a year. As stated earlier, Columbia is predicted to spend nearly equal amounts of time, over the long-run, in states of conflict and non-conflict. This transient tendency equates to recurrence rates, for both conflict and non-conflict, of approximately two steps. Given, this prediction, Columbia could theoretically experience up to 2.5 conflict recurrences every 10 model steps, possibly resulting in severe and recurrent instability within the region.

4.5 Forecasting Global Conflict trends

Overview

The use of nation specific Markov models enables forecasting of conflict for time horizons far greater than those possible with logistic regression alone. For this study we examine the predicted incidences of violent conflict for 2016, 2019, and 2024, identifying which nations are predicted to experience significant changes in their conflict probabilities. As part of this analysis, we will then examine the predicted individual transience of each nation over this ten year forecasting period, identifying which nations are predicted to experience a conflict transition rate above regional and world averages. It should be remembered that the forecasts provided in this study are predicated on the assumption that regional factors germane to violent conflict remain unchanged from current conditions throughout the forecast period.

Two, Five, and Ten Year Conflict Forecasts

World Overview

The two-, five- and ten- year conflict forecasts for each nation were calculated by raising their specific Markov models, using 2014 conflict transition probabilities, to the 2^{nd}, 5^{th}, and 10^{th} powers. The analysis focused on determining the incidence of conflict at the regional and world levels by identify which states had a probability of greater than or equal to 0.50. Additionally, the analysis also identified the ten-year conflict trends for each nation by calculating the difference in the 2014 and 2024 conflict probabilities. The analysis sought to identify which nations experienced significant, moderate, or slight changes in the probability of conflict; Table 33 provides the assessment of the change in conflict over the range of probabilities. Negative changes in conflict probability equate to a predicted decrease in the level of violence over ten year period, while positive equate to an increase in violence over the same time span. Additionally, nations that experience an absolute change in conflict probability less than or equal to 0.05 are assessed as having no significant change in their conflict status over the ten year time horizon.

Table 33: Forecasting Assessment Matrix

Change in P[Conflict]: 2014 to 2024	Assessment of Change in Conflict
$\Delta P \leq -0.50$	Significantly Less Conflict
$\Delta P < -0.25$	Moderately Less Conflict
$\Delta P < -0.05$	Slightly Less Conflict
$-0.05 \leq \Delta P \leq 0.05$	No Change
$\Delta P > 0.05$	Slightly More Conflict
$\Delta P > 0.25$	Moderately More Conflict
$\Delta P \geq 0.50$	Significantly More Conflict

A total of 17 nations were identified as having significant changes in their probabilities of conflict over the ten-year forecast period, and are presented in Table 34. Twelve of these nations are projected to experience significantly more conflict by 2024, while only five nations are expected to realize significant decreases in their levels of violence over the same time frame. In total, 40 of the 182 nations considered in this study are predicted to experience increases in conflict over the ten-year forecast period. Additionally, 30 nations are expected to realize net decreases in conflict, with 112 nations experience no significant change in their current conflict levels over the same period.

Table 34: Significant Changes in Conflict Probability Over 10 Year Period

Country	Region	10 Year Trend
Libya	Arab Countries	Significantly More Conflict
Tunisia	Arab Countries	Significantly More Conflict
Kazakhstan	Eastern Europe and Central Asia	Significantly More Conflict
Romania	Eastern Europe and Central Asia	Significantly More Conflict
Trinidad and Tobago	Latin America	Significantly More Conflict
Korea, North	South and East Asia	Significantly More Conflict
Micronesia, Federated States of	South and East Asia	Significantly More Conflict
Mongolia	South and East Asia	Significantly More Conflict
Nepal	South and East Asia	Significantly More Conflict
Timor-Leste	South and East Asia	Significantly More Conflict
Angola	Sub Saharan Africa	Significantly More Conflict
Sierra Leone	Sub Saharan Africa	Significantly More Conflict
Honduras	Latin America	Significantly Less Conflict
Paraguay	Latin America	Significantly Less Conflict
Greece	OECD	Significantly Less Conflict
Cambodia	South and East Asia	Significantly Less Conflict
Burundi	Sub Saharan Africa	Significantly Less Conflict

Table 35 provides a summary of the global incidence of conflict and ten year conflict trends. As of 2014, 84 of the 182 nations considered in the study were observed to be in violent conflict, and it is predicted that this global incidence of conflict will remain constant over the 10 year period. However, over the same time frame, it is predicted that 40 of nations will experience increased probabilities of conflict, while only 30 nations will realize decreases in their respective conflict probabilities. However, it should be noted that changes in conflict probabilities do not necessarily equate to conflict transitions but instead identify nations that are expected to experience a measurable change in their current levels of violence. Analysis of the long term conflict probabilities, based on 2014 data, indicates that the global incidence of violence is expected to increase, with a projected 95 (52%) of the 182 nations existing in a state of violent conflict. The complete two-, five-, and ten-year forecasts for each nation are provided in Appendix D.

Table 35: Summary of Conflict Forecasts: World View

World View Conflict Trend Statistics		
Statistic	**Count**	**Percentage**
Total Nations Considered:	**182**	**100%**
Total Nations in Conflict 2014	**84**	**46%**
Projections		
Number Projected in Conflict 2016	**79**	**43%**
Number Projected in Conflict 2019	**80**	**44%**
Number Projected in Conflict 2024	**84**	**46%**
Likelihood Trends 2014 - 2024		
Number trending towards conflict	**40**	**22%**
Number trending towards non-conflict	**30**	**16%**
Number experiencing no change	**112**	**62%**

Arab & North African States

The Arab and North African States currently experience the highest rates of violent conflict among the six geographic regions, with 14 of 17 states experiencing violent conflict as of 2014, as shown in Table 36. These levels of violence are project to increase over the ten-year forecast, with a projected regional violent conflict rate of 100% by year 2024, given no change in current conditions. These regional conflict rates are expected to continue indefinitely past the ten- year forecast horizon. It should be noted that the conflict rates within this region are predicted to cycle between 14 and 17 nations during the forecast period. This cycling is the result of predicted state transitions by Egypt, Jordan, Libya, Morocco, and Tunisia during the forecast period. Over the long run, it is projected that this cycling will cease, and that all 17 nations within the region will be in a state of conflict given no change to current conditions.

Table 36: Arab & North African States Conflict Forecast Summary

Arab & North African States Conflict Trend Statistics		
Statistic	**Count**	**Percentage**
Total Nations in Region	**17**	**100%**
Total Nations in Conflict 2014	**14**	**82%**
Projections		
Number Projected in Conflict 2016	**17**	**100%**
Number Projected in Conflict 2019	**14**	**82%**
Number Projected in Conflict 2024	**17**	**100%**
Likelihood Trends 2014 - 2024		
Number trending towards conflict	**6**	**35%**
Number trending towards non-conflict	**0**	**0%**
Number experiencing no change	**11**	**65%**

Eastern Europe & Central Asia

The Eastern Europe and Central Asia region is projected to steady growth in its rate of violent conflict over the ten year forecasting period, with a projected conflict incidence rate of 46% by year 2024. Long run conflict rates are expected to peak at 61%, with 17 of the 28 existing in a state of conflict. Within the region violent conflict is expected to cluster in the Caucasus and the states bordering Afghanistan, while many of the eastern European and Baltic nations are predicted to remain out of conflict over the same period. Internecine violence within Russia and Ukraine is predicted to continue unabated over the next decade, and may lead to increased instability within the surrounding former Soviet states. The regional summary is provided in Table 37.

Table 37: Eastern Europe & Central Asia Conflict Forecast Summary

Eastern Europe & Central Asia Conflict Trend Statistics		
Statistic	**Count**	**Percentage**
Total Nations in Region	**28**	**100%**
Total Nations in Conflict 2014	**11**	**39%**
Projections		
Number Projected in Conflict 2016	**9**	**32%**
Number Projected in Conflict 2019	**11**	**39%**
Number Projected in Conflict 2024	**13**	**46%**
Likelihood Trends 2014 - 2024		
Number trending towards conflict	**8**	**29%**
Number trending towards non-conflict	**3**	**11%**
Number experiencing no change	**17**	**61%**

Latin America

Latin American violent conflict rates are predicted to remain constant over the forecasting period with 13 of the 27 nations predicted experience some level of violent conflict. Violent conflict is predicted to cluster in South and Central American nations, while only two Caribbean nations (Jamaica and Trinidad) are projected to be in state of conflict by 2024. The forecast also predicts that Brazil, Columbia, and Venezuela will remain in conflict with levels of violence remaining constant in Brazil and Venezuela. Columbia, on the other hand, is projected to experience a moderate decrease in it conflict probability by 2024, given current conditions persist. Over the long run, conflict rates are expected to increase to approximately 70%, with 19 of the 27 nations predicted to be in a state of violent conflict, a majority of which are located in Central America and norther South America. The regional summary is provided in Table 38.

Table 38: Latin America Conflict Forecast Summary

Latin America Conflict Trend Statistics		
Statistic	**Count**	**Percentage**
Total Nations in Region	**27**	**100%**
Total Nations in Conflict 2014	**14**	**52%**
Projections		
Number Projected in Conflict 2016	**13**	**48%**
Number Projected in Conflict 2019	**13**	**48%**
Number Projected in Conflict 2024	**13**	**48%**
Likelihood Trends 2014 - 2024		
Number trending towards conflict	**6**	**22%**
Number trending towards non-conflict	**4**	**15%**
Number experiencing no change	**17**	**63%**

OECD

Currently the OECD region experiences the lowest rates of violent conflict among the six geographic regions, a trend that is currently in a state of equilibrium, and is expected to continue over the forecast period. Of the six OECD nations predicted to be in conflict in 2024, only South Korea is predicted to experience a transition into conflict, while Chile, Israel, Mexico, Turkey, and the United Kingdom are project to remain in conflict for the foreseeable future. It is also predicted that only Poland and the United States are predicted to experience slight increases in their respective conflict probabilities, while all other nations will realize either a decrease or no significant change in the conflict probabilities over the next decade. Long run incidences of conflict are expected to drop to 15%, with the nations of Chile, Israel, Mexico, South Korea, and Turkey remaining in a state of conflict. The regional summary is provided in Table 39.

Table 39: OECD Conflict Forecast Summary

OECD Conflict Trend Statistics		
Statistic	**Count**	**Percentage**
Total Nations in Region	**33**	**100%**
Total Nations in Conflict 2014	**7**	**21%**
Projections		
Number Projected in Conflict 2016	**6**	**18%**
Number Projected in Conflict 2019	**6**	**18%**
Number Projected in Conflict 2024	**6**	**18%**
Likelihood Trends 2014 - 2024		
Number trending towards conflict	**2**	**6%**
Number trending towards non-conflict	**4**	**12%**
Number experiencing no change	**27**	**82%**

South & East Asia

Rates of violent conflict in the South and East Asian region are projected to eclipse those of the both Latin America and Sub-Saharan Africa regions, with 17 of the 28 regional nations predicted to be in a state of conflict by 2024. Over the forecast period six nations (Laos, Micronesia, Mongolia, Nepal, North Korea, and Timor-Leste) are predicted to experience transitions into conflict, while only two nations (Cambodia and Vietnam) are predicted to transition out of conflict. Both China and India are predicted to remain in conflict over the next decade, with their respective conflict probabilities remaining nearly constant over the same period. As shown in Table 40, conflict cycles over the course of the forecast period due to the transitions discussed previously. Ultimately the incidence rate of violent conflict is predicted to stabilize at 64% with 18 of the 28 nations experiencing some level of violent conflict.

Table 40: South & East Asia Conflict Forecast Summary

South & East Asia Conflict Trend Statistics		
Statistic	**Count**	**Percentage**
Total Nations in Region	**28**	**100%**
Total Nations in Conflict 2014	**13**	**46%**
Projections		
Number Projected in Conflict 2016	**14**	**50%**
Number Projected in Conflict 2019	**18**	**64%**
Number Projected in Conflict 2024	**17**	**61%**
Likelihood Trends 2014 - 2024		
Number trending towards conflict	**8**	**29%**
Number trending towards non-conflict	**4**	**14%**
Number experiencing no change	**16**	**57%**

Sub-Saharan Africa

While the Sub-Saharan Africa region has the most states currently and predicted to b in violent conflict, it is the only region, other than the OECD, that is projected to experience a decrease in its regional rate of conflict over the ten year forecast. Of the 18 nations predicted to be in conflict in 2024, only Angola, Cameroon, and Sierra Leone are predicted to transition into conflict; additionally 10 nations are projected to transition out of conflict over the same period. Similar to the OECD region, average conflict probabilities are projected to decrease in Sub-Saharan Africa over the next decade with 15 nations projected to have lower probabilities of conflict, while only 10 nations are predicted to have increased conflict probabilities given current conditions. Ove the long run, conflict rates in Sub-Saharan Africa are predicted to stabilize at 39% with 19 of the 49 nations existing in a state of violent conflict. The regional summary is provided in Table 41.

Table 41: Sub-Saharan Africa Conflict Forecast Summary

Sub-Sahara Africa Conflict Trend Statistics		
Statistic	**Count**	**Percentage**
Total Nations in Region	**49**	**100%**
Total Nations in Conflict 2014	**25**	**51%**
Projections		
Number Projected in Conflict 2016	**20**	**41%**
Number Projected in Conflict 2019	**18**	**37%**
Number Projected in Conflict 2024	**18**	**37%**
Likelihood Trends 2014 - 2024		
Number trending towards conflict	**10**	**20%**
Number trending towards non-conflict	**15**	**31%**
Number experiencing no change	**24**	**49%**

Analysis of Conflict Transience in Nations

Conflict transience describes a nation's tendency to transition into and out of conflict frequently. Highly transient nations, such as Columbia, Morocco, or the United States, are identified as those having long-run conflict probabilities (π_j) approaching 0.50. Such conflict probabilities indicate that a nation spends nearly equal amounts of time in states of conflict and non-conflict, resulting in relatively frequent conflict transitions. A nation's transience score is based on the sum of the mean recurrence steps (M_0, M_1) provided in Equation 24. The expected number of conflict transitions over a given time period (T) is given by Equation 28. For the purposes of this study, T is set to 10 years to coincide with the 10 year forecast discussed in the previous section.

Equation 28: Expected Number of Conflict Recurrences for Given Time Period

$$E[Recurrences] = \frac{T}{\sum_{j=0}^{1} M_j}$$

Where:

T = Time period of interest (years)
M_j = Mean Recurrence (number of steps)
j = Conflict State {0, 1}

For a hypothetical nation exhibiting a long-run conflict probability of 0.50, the sum of the mean recurrence times for both conflict and non-conflict is 4 years, resulting in 2.5 expected recurrences over a 10 year period. This hypothetical nation is used as the transience benchmark, against which all nations are compared, yielding the Transience Score provided in Equation 29.

Equation 29: Nation Specific Transience Score

$$Transience\ Score = \frac{E[Recurrences]}{2.5}$$

A nation's transience score is utilized to identify nations that are identified as predisposed to conflict transitions. Transience Scores approaching one indicate highly transient nations, while scores approaching zero identify nations that tend to remain in one state over the other. The Transience Scores for each nation listed in the regional tables provided in Appendix E.

Table 42 provides the top 25 most transient nations identified in this study, 12 of which were identified as being in conflict in 2014. Libya and Tunisia, which have experienced relatively few conflict transitions over the past 20 years, were identified as the most transient nations within the study with respective scores of approximately one. This finding is attributed to the dynamic changes resulting from the Arab Spring that first began in Tunisia and quickly spread to Libya and other Arab nations. Similarly, Morocco, Jordan, and Egypt are also identified as being highly transient based on 2014 data. The United States is also identified as being highly transient, which concurs with its recent history of experiencing five conflict transitions between 2004 and 2014. The transience of the United States is credited in part to it ongoing worldwide military engagements, instability due to a highly polarized political process, and ongoing conflicts along its southern border with Mexico resulting from population migration and an increasingly violent illicit narcotics trade. Seven nations from the Sub-Saharan Africa Region are identified as being highly transient. Similar to the United States, Cameroon and Cote d'Ivoire also have a history of multiple conflict transitions, experiencing four and two recurrences respectively between 2004 and 2014.

Table 42: Top 25 Most Transient Nations

Nation	2014 Conflict Status	Region	M_0	M_1	Expected Number Recurrences per 10 Years	Transience Score	Rank
Libya	Conflict	Arab Countries	2.00	2.00	2.5	1.00000	1
Tunisia	Conflict	Arab Countries	2.01	1.99	2.5	0.99998	2
United States	Conflict	OECD	1.94	2.07	2.5	0.99898	3
Bosnia and Herzegovina	No Conflict	Eastern Europe and Central Asia	1.93	2.07	2.5	0.99884	4
Cameroon	No Conflict	Sub Saharan Africa	2.10	1.91	2.5	0.99774	5
Colombia	Conflict	Latin America	1.89	2.12	2.5	0.99685	6
Kiribati	No Conflict	South and East Asia	1.89	2.12	2.5	0.99677	7
Botswana	No Conflict	Sub Saharan Africa	1.89	2.13	2.5	0.99637	8
Malawi	No Conflict	Sub Saharan Africa	1.84	2.19	2.5	0.99228	9
Montenegro	No Conflict	Eastern Europe and Central Asia	1.84	2.20	2.5	0.99199	10
Benin	No Conflict	Sub Saharan Africa	2.20	1.83	2.5	0.99191	11
Korea, South	No Conflict	OECD	2.21	1.83	2.5	0.99112	12
Morocco	Conflict	Arab Countries	2.24	1.81	2.5	0.98884	13
Jordan	Conflict	Arab Countries	2.54	1.65	2.4	0.95477	14
Georgia	No Conflict	Eastern Europe and Central Asia	2.58	1.63	2.4	0.94998	15
Vietnam	Conflict	South and East Asia	1.62	2.63	2.4	0.94325	16
Bahamas	No Conflict	Latin America	2.63	1.61	2.4	0.94241	17
Laos	No Conflict	South and East Asia	2.71	1.58	2.3	0.93130	18
Cote d'Ivoire	Conflict	Sub Saharan Africa	2.76	1.57	2.3	0.92453	19
Lesotho	Conflict	Sub Saharan Africa	1.56	2.80	2.3	0.91823	20
Ecuador	Conflict	Latin America	2.80	1.55	2.3	0.91809	21
Egypt	Conflict	Arab Countries	2.84	1.54	2.3	0.91189	22
Mozambique	Conflict	Sub Saharan Africa	2.87	1.54	2.3	0.90836	23
Uzbekistan	No Conflict	Eastern Europe and Central Asia	2.95	1.51	2.2	0.89568	24
Dominican Republic	No Conflict	Latin America	1.50	3.00	2.2	0.88888	25

On the other end of the spectrum are the nations with exceedingly low Transience Scores that are projected to remain in their current conflict states for the foreseeable future. Table 43 provides the listing of the top 25 least transient nations, with the Caribbean nations of Cuba and Antigua identified least transient nations within the Study. Latin American and OECD nations account for 12 of the 25 nations and show a propensity to remain in a state of non-conflict. Of this group, only Suriname and Trinidad, both classified as not in conflict in 2014, are identified as having a predisposition for long term conflict. Five Arab nations are also identified within this group, all of which showing proclivity towards remaining in a state of conflict, supporting the results of the conflict forecast provided in the previous section.

Table 43: Top 25 Least Transient Nations

Nation	2014 Conflict Status	Region	M_0	M_1	Expected Number Recurrences per 10 Years	Transience Score	Rank
Cuba	No Conflict	Latin America	1.00	1.00E+36	0.0	0.00000	1
Antigua and Barbuda	No Conflict	Latin America	1.00	1.00E+36	0.0	0.00000	2
Kuwait	Conflict	Arab Countries	1.56E+13	1.00	0.0	0.00000	3
Malta	No Conflict	Eastern Europe and Central Asia	1.32E+13	1.00	0.0	0.00000	4
Central African Republic	Conflict	Sub Saharan Africa	4.42E+11	1.00	0.0	0.00000	5
Qatar	No Conflict	Arab Countries	1.43E+11	1.00	0.0	0.00000	6
United Arab Emirates	No Conflict	Arab Countries	1.42E+11	1.00	0.0	0.00000	7
Bahrain	Conflict	Arab Countries	1.42E+11	1.00	0.0	0.00000	8
South Sudan	Conflict	Sub Saharan Africa	2.65E+10	1.00	0.0	0.00000	9
Pakistan	Conflict	Eastern Europe and Central Asia	3.93E+09	1.00	0.0	0.00000	10
Iceland	No Conflict	OECD	1.00	2.21E+09	0.0	0.00000	11
Norway	No Conflict	OECD	1.00	6.22E+08	0.0	0.00000	12
Ireland	No Conflict	OECD	1.00	1.43E+08	0.0	0.00000	13
Suriname	No Conflict	Latin America	7.21E+07	1.00	0.0	0.00000	14
Trinidad and Tobago	No Conflict	Latin America	4.29E+07	1.00	0.0	0.00000	15
Denmark	No Conflict	OECD	1.00	9.38E+06	0.0	0.00000	16
Sweden	No Conflict	OECD	1.00	8.85E+06	0.0	0.00000	17
Finland	No Conflict	OECD	1.00	5.08E+06	0.0	0.00000	18
Iraq	Conflict	Arab Countries	2.06E+06	1.00	0.0	0.00000	19
Nepal	No Conflict	South and East Asia	1.26E+06	1.00	0.0	0.00000	20
Indonesia	Conflict	South and East Asia	8.14E+05	1.00	0.0	0.00000	21
Belgium	No Conflict	OECD	1.00	7.16E+05	0.0	0.00001	22
Micronesia, Federated States of	No Conflict	South and East Asia	4.88E+05	1.00	0.0	0.00001	23
Netherlands	No Conflict	OECD	1.00	3.82E+05	0.0	0.00001	24
Tajikistan	Conflict	Eastern Europe and Central Asia	2.70E+05	1.00	0.0	0.00001	25

Analysis of Table 41 identified seven nations: Malta, Qatar, United Arab Emirates, Suriname, Trinidad and Tobago, Nepal, and Micronesia, classified as not in conflict in 2014 that show an inclination towards long-term uninterrupted conflict. A subsequent comparison with each of these nations' respective ten-year forecast shows that five nations are predicted to be in conflict in 2024, and only Malta and Suriname are forecasted to remain in the current state over the same time period. Subsequently, Micronesia, Nepal, Qatar, Trinidad and Tobago, and the United Arab Emirates are identified as being at risk for near term transitions into conflict.

The distribution of Transience Scores is presented in Figure 37. As can be seen, 113 (62%) of the 182 nations have Transience Score less than or equal to 0.200. Additionally only 37 (20%) nations have moderate transience scores between 0.20 and 0.80, while 32 (18%) nations are classified as being highly transient with scores greater than 0.80. This result demonstrates the typical finding that nations tend to remain in either a state of conflict or non-conflict, and that in general national-level conflict transitions are rare events.

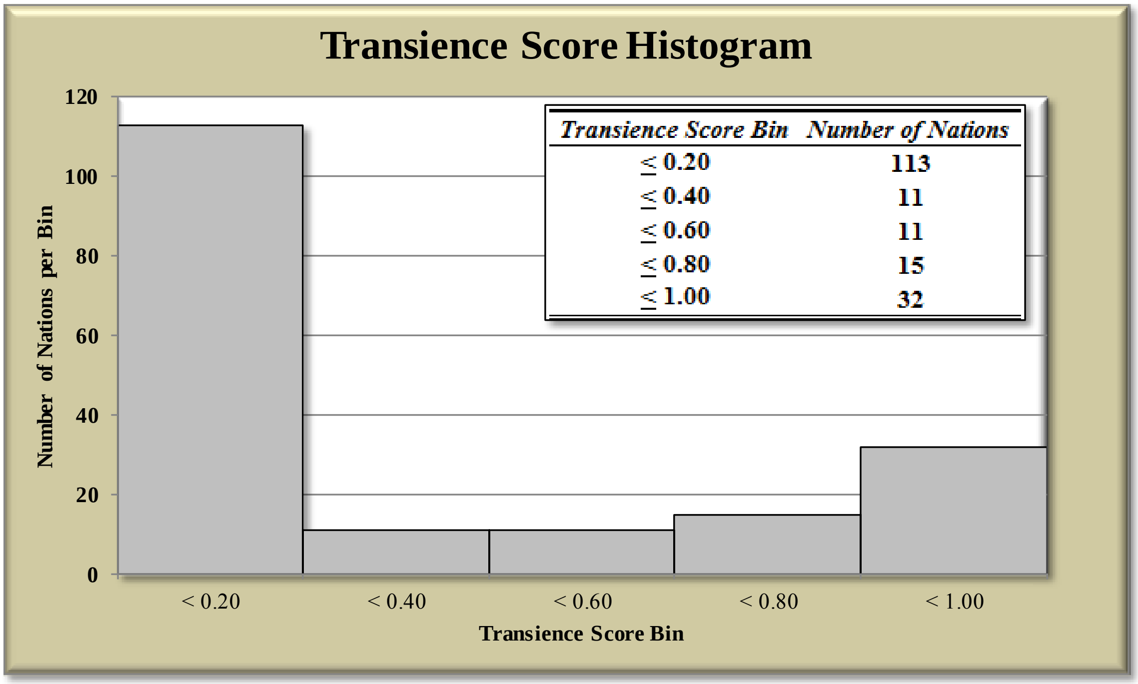


Figure 37: Transience Score Histogram

V. Conclusions and Recommendations

"Be prepared to re-examine your reasoning"

Robert S. McNamara, In Respect: The Tragedy and Lessons of Vietnam

5.1 Chapter Overview

This chapter provides the research conclusions derived from developing a set of region specific conditional logistic regression and Markov models for the prediction and forecasting of conflict transition in nations. In Section 5.2 we provide a summary of the study's problem statement, research questions, and methodology. Next, Section 5.3 discusses the significance of the research and its applicability in operational and strategic level planning. Finally, in Section 5.3 we discuss possible future research concerning the prediction and spread of violent conflict in nations.

5.2 Conclusions of Research

This study considered 30 statistical and trend variables in the development of models to predict future incidences of conflict transitions. Relying on logistic regression, Markov models, and methodologies proven in previous studies, this research reconfirmed the validity of using geographic sub-regions to develop conditional logistic regression models for the 182 nations considered in this study. These models subsequently developed the conflict transition probabilities utilized in the set of Markov models enabling long range forecasts of regional and global incidences of conflict seldom seen in previous analytical efforts. Ultimately the models developed for this study and subsequent analysis answered the five research questions posed in Chapter 1.

Question 1: How accurately can statistical models predict conflict transitions for individual nations?

A total of 12 conditional logistic regression models were developed for six geographic regions for this study. These models achieved weighted predictive accuracies, at the world level, of 88.76% on the training data set, and 84.67% on the validation data set. Regional weighted predictive accuracies exceeded 90% in the Arab and North African States model (93.72%) and the OECD model (93.84%) on the training set data as well as the validation OECD model which achieved a predictive accuracy of 93.46%, far exceeding the pre-established bench mark of 80% predictive accuracy. In addition to their overall classification accuracy, the logistic regression models correctly classify 43.6% of transitions out of conflict, and 52.2% of transitions into conflict; a metric concerning rare events, that has never been examined in previous conflict prediction studies.

The overall model accuracies of this study significantly exceed those generated by the Goldstone (Goldstone, et al., 2005), Hegre (Hegre et al., 2011), and CAA studies (Reed, 2013). Additionally, the regional models compare favorably with the recent Boekestein (Ahner, Boekestein, & Deckro, 2015) study, ultimately generating higher validation data set accuracies for all six regions. With this result in mind, it is recommended that the Eastern Europe and Central Asian region be revaluated with a possible reassignment of some or all of the central Asian nations to the Arab & North African region, which shares similar ethnic, political, and geographic features. A key insight gained from the validation of the logistic regression models is the finding that conflict transitions are generally easier to predict in nations not in conflict, compared to those currently in conflict. This finding may be due impart to the quality and accuracy of the statistical data collected in these nations, resulting from the presence of a more permissive and less violent environment.

Question 2: What factors are the significant predictors of conflict transitions?

Thirty independent variables were required to construct the 12 conditional logistic regression models. Amongst the six geographic regions, statistics such as ethnic diversity, burgeoning youth populations, national military expenditures, religious diversity, and the type of government emerged as the most common and significant factors pertaining to conflict transitions. While the conditional models within a specific region were always considerably different from each other, in many cases they share common variables. However, in certain instances, such as the case with religious diversity in Latin American nations, the current status of a nation affects how the variable will increase or decrease to probability of a conflict transition. As is the case in the Latin American models increases to a nation's religious diversity score, will result in an increased likelihood that the nation will transition out of conflict in the following year. However, increasing the same religious diversity score in nations currently not in conflict corresponds to a subsequent increase in the likelihood that these nations will transition into conflict in the next year. Due to this phenomenon, care must be taken when analyzing how a particular nation and region will react to the application of national power or other external forces over an extended period of time.

Question 3: How is the number of global conflicts predicted to change by 2024 and beyond?

Two-state Markov models were developed using the probabilities generated by the 12 conditional logistic regression models, for each of the 182 nations considered in this study, enabling the forecasting and trend analysis of regional and global conflict. Over the next decade the global incidence of conflict is predicted to remain at 2014 levels with 84 (46%) nations experiencing some level of violent conflict. However, given 2014 conflict data, the global incidence of conflict is expected to increase to 95 (52%) nations in long run. Regionally, conflict levels are predicted to increase by 2024 in Arab and North African states, Eastern Europe and Central Asia, and in South and East Asia; these trends are predicted to continue past the forecast horizon. Conflict levels within Central and South America are projected to remain constant over the next decade with 48% of the regions nations experiencing violent conflict. However, conflict rates within this region are predicted to increase to approximately 70%, with 19 of the 27 nations predicted to be in a state of violent conflict in the long run. Conversely, conflict levels are predicted to decrease in OECD and Sub-Saharan African nations over the next decade and in the long run, with regional conflict incidence rates of 18% and 37% respectively.

Question 4: What nations are susceptible to conflict transitions; which nations appear invulnerable to conflict transitions?

Identification of nations susceptible or invulnerable to conflict transitions is predicated on a nation's Transience Score on a continuous scale from 0 to 1. Nations with transience scores approaching 1 are said to be susceptible to frequent conflict transitions, while nations with low scores tend to experience infrequent conflict transitions. Of the 182 nations considered in this study, 32 nations were identified as being susceptible to frequent conflict transitions. Out of this sub-group, Libya, Tunisia, and the United States are projected to experience repeated conflict transitions over the next decade. This study also identified 113 nations as being relatively invulnerable to conflict transitions, with these nations remaining either in a state of conflict or non-conflict over the long run. The nations of Cuba, Antigua and Barbuda, as well as Kuwait respectively had the three lowest transience scores with the study, and thus are not projected to experience a conflict transition from their current state in the foreseeable future. As a region, Sub-Saharan Africa followed by Latin America are the most susceptible to conflict transitions, while the OECD and South and East Asian regions are projected to be the least susceptible to such events.

Question 5: Which nations, currently not in conflict, are identified as near-term risks for transitions into violent conflict?

Analysis of long-run conflict probabilities and transience scores sought to identify nations as not being in a state of conflict in 2014 that show a tendency towards existing in a state of violent conflict. This analysis identified the nations of Micronesia, Nepal, Qatar, Trinidad and Tobago, and the United Arab Emirates as at risk for near-term transitions into violent conflict. This assessment is based on the geographic location of these nations, current regional political climates, and their proclivity towards long term internal conflicts.

5.3 Significance of Research

This study accurately predicts the conflict status of 182 nations and provides senior leadership with insight into future conflict trends for both nations and regions, allowing for both near-term planning and long-range strategy development. The research provided herein enables the identification of nations susceptible to conflict transitions, along with relevant factors that may possibly aid or prevent such a transition from occurring. Such revelations are vital in the development and implementation of operational plans and national strategies, as they enable the identification of possible indicators of impending conflict transitions, and enable the informed allocation of resources to support our operational and strategic end states. At the same time, it should be evident that national strategies cannot simply apply "one size fits all" policies to geographic regions or assume that they will achieve the desired effects within each nation. Care must be taken to truly ascertain the conflict status of nations of interest and precisely apply the elements of national power in order to achieve strategic end states. Additionally, administrations must balance the risks and benefits as well as the second and third order effects of such international policies which, given the regional and global interconnectedness of the 21st century, will undoubtedly have far reaching and possibly global ramifications.

5.4 Recommendations for Future Research

As part of continuing research, this study recommends six areas that may yield significant analytical insights into nation-state conflict.

Relaxed Forecasting Assumptions

To enable the forecasting of violent conflict, this study assumes that any variable identified as significant within the model will remain relevant from year-to-year, and for the duration of the conflict forecasting period. Essentially, this study assumes that conditions present in 2014 will remain unchanged for the foreseeable future. It is recommended that future analyses look at relaxing this assumption. Similar to the Hegre conflict model (Hegre et al., 2011), these studies would use existing or develop internal projections of relevant independent conflict variables to develop forecasting and prediction models of regional and global conflict.

Analysis of Alternate Geographic Regions

As noted above, it is recommended that the geographic regions used in this study be reanalyzed and adjusted to improve regional commonality among the nations. In this regard, a possible alternative is to model the regions as the six geographic Unified Combatant Commands (UCC). The databases constructed for this study are currently set up to develop logistic regression models based off either the geographic regions used in this study or the current areas of responsibility of the combatant commands. Such a study may yield insights regarding the predicted incidences of conflict within each UCC, and potential realignment of their respective areas of responsibility. Such an analysis has immediate operational and strategic relevance following Chairman of the Joint Chiefs of Staff, General Joseph Dunford's directive to revamp combat commands for the "fight of the future" (Scarborough, 2015).

Analysis of Significant Conflict Predictor Variables

As part of continuing research it is recommended that an in depth analysis be conducted into correlation between significant covariates and transitions into conflict. This analysis would seek to ascertain if a causal relationship does in fact exist between the independent covariates and the dependent variable. Such an analysis would also seek to ascertain how manipulating such variables at both the national and regional levels affect transitions into conflict.

Development and Implementation of the Border Conflict Score Variable

The variable *Border Conflict Score* was identified as significant in many of the preliminary logistic regression models and, despite its absence in the final models, it is believed that this variable may be a significant predictor of the spread of violent conflict. With that being said, the current methodology used to develop this variable fails to properly account for island nations or nations having large coastlines and few land borders. As a result, this variable does not effectively model nations such as Australia or the Philippines, or other that do not share a land border with any other nation, resulting in a *Border Conflict Score* of zero. An island and coastal nation analog to this variable (e.g., shared fisheries, number of international deep water ports, number of disputed claims to islands, or some other metric that may be used as a vector to model the spread of conflict) must be developed and implemented for use in future studies.

Dynamic Border Conflict Variables in Forecasting Models

Following the use of the *Border Conflict Score* variable within logistic regression models, subsequent methodologies may wish explore forecasting future incidences of nation-state conflict using a dynamic border conflict score within Markov models. Such a methodology would allow the nation specific *Border Conflict Score* variable to change by recalculating the conditional probabilities at each transition, followed by Monte Carlo simulations that obtain average outcomes. Such an analysis would be able to derive insights into how conflict begins, terminates, and/or spreads based upon interactions at international boundaries.

Conflict Spread through Interconnected Regional Networks

The final recommendation for future research explores modeling the spread of violent conflict through interconnected regional networks. Such an analysis may seek to employ a methodology similar to the Spears study that explored viral epidemiology in connected networks (Spears, 2001). Such an analysis would require contributions from multiple disciplines including logistic regression, Markov and stochastic modeling, dynamic programming, and network analysis. Due to the complexity of this problem, it is recommended that such a study focus on a specific region, such as South West Asia or Sub-Saharan Africa, as opposed to a global model. This analysis would explore causes and develop insights into the spread of violent conflict across international borders due to such factors as trade, population migrations, or climatic and economic conditions.

Appendix A: Regional Assignments of Nations

Table 44: Regional Assignments of Nations

Number per Region	Sub-Saharan Africa	South and East Asia	Eastern Europe and Central Asia	Arab & North African States	Latin America	OECD
1	Angola	Bangladesh	Afghanistan	Algeria	Antigua and Barbuda	Australia
2	Benin	Bhutan	Albania	Bahrain	Argentina	Austria
3	Botswana	Brunei Darussalam	Armenia	Egypt	Bahamas	Belgium
4	Burkina Faso	Cambodia	Azerbaijan	Iraq	Barbados	Canada
5	Burundi	China	Belarus	Jordan	Belize	Chile
6	Cabo Verde	Korea, North	Bosnia and Herzegovina	Kuwait	Bolivia	Czech Republic
7	Cameroon	Fiji	Bulgaria	Lebanon	Brazil	Denmark
8	Central African Republic	India	Croatia	Libya	Colombia	Estonia
9	Chad	Indonesia	Cyprus	Morocco	Costa Rica	Finland
10	Comoros	Kiribati	Georgia	Oman	Cuba	France
11	Congo, Republic of the	Laos	Iran	Qatar	Dominican Republic	Germany
12	Cote d'Ivoire	Malaysia	Kazakhstan	Saudi Arabia	Ecuador	Greece
13	Congo, Democratic Republic of the	Maldives	Kyrgyzstan	Syria	El Salvador	Hungary
14	Djibouti	Micronesia, Federated States of	Latvia	Tunisia	Grenada	Iceland
15	Equatorial Guinea	Mongolia	Lithuania	United Arab Emirates	Guatemala	Ireland
16	Eritrea	Myanmar	Malta	West Bank	Guyana	Israel
17	Ethiopia	Nepal	Montenegro	Yemen	Haiti	Italy
18	Gabon	Papua New Guinea	Pakistan		Honduras	Japan
19	Gambia	Philippines	Moldova		Jamaica	Luxembourg
20	Ghana	Samoa	Romania		Nicaragua	Mexico
21	Guinea	Singapore	Russia		Panama	Netherlands
22	Guinea-Bissau	Solomon Islands	Serbia		Paraguay	New Zealand
23	Kenya	Sri Lanka	Slovakia		Peru	Norway
24	Lesotho	Thailand	Tajikistan		Suriname	Poland
25	Liberia	Timor-Leste	Macedonia		Trinidad and Tobago	Portugal
26	Madagascar	Tonga	Turkmenistan		Uruguay	Korea, South
27	Malawi	Vanuatu	Ukraine		Venezuela	Slovenia
28	Mali	Vietnam	Uzbekistan			Spain
29	Mauritania					Sweden
30	Mauritius					Switzerland
31	Mozambique					Turkey
32	Namibia					United Kingdom
33	Niger					United States
34	Nigeria					
35	Rwanda					
36	Sao Tome and Principe					
37	Senegal					
38	Seychelles					
39	Sierra Leone					
40	Somalia					
41	South Africa					
42	South Sudan					
43	Sudan					
44	Swaziland					
45	Togo					
46	Uganda					
47	Tanzania					
48	Zambia					
49	Zimbabwe					

Appendix B: Region Specific Conditional Logistic Regression Models

Table 45: Summary of Arab and North African State Models

Arab State Models

Given "Conflict"

Term	Estimate	Prob > Chi-Sq
Intercept	43.828	0.0265
3 Yr Freedom Trend	22.435	0.0781
Ethnic Diversity	-45.350	0.0271
Regime Type (Emerging)	-3.107	0.2338
Regime Type (Democratic)	4.444	0.0161

Model Significance (Prob > Chi-Sq):	0.006
Area Under the Curve (Training):	0.962

Given "Non-Conflict"

Term	Estimate	Prob > Chi-Sq
Intercept	728.886	0.009
Death Rate	-16.257	0.0104
Life Expectancy	-7.545	0.0096
Youth Bulge	-2.279	0.0092
Ethnic Diversity	168.499	0.0091
Religious Diversity	-211.666	0.0083
Military Expend (% GDP)	1.317	0.0209

Model Significance (Prob > Chi-Sq):	0.000
Area Under the Curve (Training):	0.930

Table 46: Summary of Eastern Europe & Central Asian Models

Eastern Europe & Central Asia Models

Given "Conflict"

Term	Estimate	Prob > Chi-Sq
Fertility Rate	8.958	0.0051
Infant Mortality rate	-0.337	0.0086
Population density	0.078	0.0159
Trade (% GDP)	-0.381	0.007
Freedom Score	11.598	0.0311

Model Significance (Prob > Chi-Sq):	0.000
Area Under the Curve (Training):	0.972

Given "Non-Conflict"

Term	Estimate	Prob > Chi-Sq
Intercept	-28.959	0.0011
Arable Land	2.801	0.0037
GDP Per Capitia	0.000	0.0047
Improved Water	0.208	0.0024
2 Yr Freedom Trend	-44.417	0.0177
3 Yr Freedom Trend	58.898	0.0012
Religious Diversity	10.689	0.0052
2 Yr Conflict Intensity Trend	-10.734	0.0018
Government Type (Emerging)	4.796	0.0028
Government Type (Democratic)	-0.342	0.6275
Government Type (Foreign Interruption)	6.652	0.0004

Model Significance (Prob > Chi-Sq):	0.000
Area Under the Curve (Training):	0.946

Table 47: Summary of Latina American Models

Latin American Asia Models

Given "Conflict"		
Term	**Estimate**	**Prob > Chi-Sq**
Religious Diversity	-33.471	0.0361
Government Type (Emerging)	36.483	0.0281
Government Type (Democratic)	37.238	0.0233
Government Type (Anarchy)	30.524	0.0287
Freedom Score	-7.958	0.1997

Model Significance (Prob > Chi-Sq):	0.010
Area Under the Curve (Training):	0.878

Given "Non-Conflict"		
Term	**Estimate**	**Prob > Chi-Sq**
Birth Rate	-1.094	0.0186
Fertility Rate	9.990	0.009
Improved Water	-0.644	0.0031
Infant Mortality rate	-0.535	0.0039
Ethnic Diversity	-21.442	0.0002
Religious Diversity	11.832	0.0005
Regime Type (Democratic)	72.267	0.004
$(\text{Religious Diversity}-0.62476)^2$	-18.746	0.1017

Model Significance (Prob > Chi-Sq):	0.000
Area Under the Curve (Training):	0.952

Table 48: Summary of OECD Models

OECD Nation Models

Given "Conflict"		
Term	**Estimate**	**Prob > Chi-Sq**
Intercept	-19.386	0.0181
Military Expend (% Gov Spending)	-0.355	0.0255
Youth Bulge	0.489	0.0371
Caloric Intake	0.005	0.027

Model Significance (Prob > Chi-Sq):	0.001
Area Under the Curve (Training):	0.914

Given "Non-Conflict"		
Term	**Estimate**	**Prob > Chi-Sq**
Intercept	74.164	0.0173
Birth Rate	1.594	0.0709
Death Rate	-1.611	0.0262
Military Expend (% GDP)	5.874	0.0145
Infant Mortality rate	3.015	0.0157
Youth Bulge	-3.148	0.0249
Caloric Intake	-0.017	0.016

Model Significance (Prob > Chi-Sq):	0.000
Area Under the Curve (Training):	0.974

Table 49: Summary of South & East Asian Models

Given "Conflict"		
Term	**Estimate**	**Prob > Chi-Sq**
Intercept	25.048	0.0014
Military Expend (% GDP)	-0.347	0.0272
Population Growth	-6.212	0.002
Trade (% GDP)	-0.109	0.0059
Ethnic Diversity	-12.087	0.0066
Government Type (Emerging)	2.916	0.0346
Government Type (Democratic)	4.402	0.004

Model Significance (Prob > Chi-Sq):	0.000
Area Under the Curve (Training):	0.938

Given "Non-Conflict"		
Term	**Estimate**	**Prob > Chi-Sq**
Intercept	17.886	0.0126
Arable Land	-20.358	0.0325
Military Expend (% GDP)	-2.259	0.0175
Refugee (Origin)	9.60E-06	0.0437
Fresh Water per Capita	-5.59E-05	0.0162
Caloric Intake	-0.005	0.0241

Model Significance (Prob > Chi-Sq):	0.000
Area Under the Curve (Training):	0.932

Table 50: Summary of Sun-Saharan African Models

Sub-Saharan Africa Nation Models

Given "Conflict"		
Term	**Estimate**	**Prob > Chi-Sq**
Birth Rate	-0.400	0.0047
Life Expectancy	-0.190	0.003
Military Expend (% Gov Spending)	-0.158	0.0454
Youth Bulge	0.672	0.0006
Refugee (Origin)	1.17E-05	0.0044
Fresh Water per Capita	-7.40423E-05	0.0017
Ethnic Diversity	-4.421	0.0011
Government Type (Emerging)	2.797	0.0118
Government Type (Democratic)	3.779	0.0034
Government Type (Anarchy)	26.327	0.9999
Government Type (Transition)	-1.530	0.3111

Model Significance (Prob > Chi-Sq):	0.000
Area Under the Curve (Training):	0.938

Given "Non-Conflict"		
Term	**Estimate**	**Prob > Chi-Sq**
Arable Land	7.801	0.0002
Birth Rate	-0.474	0.0031
Infant Mortality rate	0.053	0.0041
Youth Bulge	0.346	0.011
Refugee (Asylum)	5.91E-06	0.0153
Trade (% GDP)	-0.052	0.0051
Freedom Score	-5.637	0.0004

Model Significance (Prob > Chi-Sq):	0.000
Area Under the Curve (Training):	0.932

Appendix C: Markov Model Outputs

Table 51: Markov Model results by Nation

Country	Status 2014	Sojourn Times (Ri)				Mean Conflict Recurrence Steps (M_j)		Long Run Conflict Probabilites (πj)	
		Number of Years to 1st transition:	Variance	Number Years to 2nd Transition:	Variance	E[Steps] to Transition from No Conflict to No Conflict	E[Steps] to Transition from Conflict to Conflict	Probability Not in Conflict	Probability in Conflict
Afghanistan	Conflict	65274	4E+09	65299	4E+09	2585	1.00	0.0004	0.9996
Albania	No Conflict	13	162	511	246990	38.59	1.03	0.0259	0.9741
Algeria	Conflict	35	1276	36	1276	37.22	1.03	0.0269	0.9731
Angola	No Conflict	4	14	22	302	5.16	1.24	0.1939	0.8061
Antigua and Barbuda	No Conflict	6.42E+35	4E+71	6.42E+35	4E+71	1.00	1.00E+12	1.0000	0.0000
Argentina	No Conflict	309	95175	311	95178	1.01	146.94	0.9932	0.0068
Armenia	Conflict	111	12514	112	12514	112.48	1.01	0.0089	0.9911
Australia	No Conflict	8.82E+06	8E+13	8.82E+06	8E+13	1.00	95902	1.0000	0.0000
Austria	No Conflict	1.75E+06	3E+12	1.75E+06	3E+12	1.00	1.32E+05	1.0000	0.0000
Azerbaijan	Conflict	15789	2E+08	15792	2E+08	5894	1.00	0.0002	0.9998
Bahamas	No Conflict	307	94047	808	3E+05	2.63	1.61	0.3800	0.6200
Bahrain	Conflict	1.42E+11	2E+22	1.42E+11	2E+22	1.42E+11	1.00	0.0000	1.0000
Bangladesh	Conflict	223	50149	224	50149	198	1.01	0.0050	0.9950
Barbados	No Conflict	371	137537	1652	1777180	4.45	1.29	0.2247	0.7753
Belarus	No Conflict	2	3.39	3.44	3.42	1.43	3.34	0.7005	0.2995
Belgium	No Conflict	2.33E+07	5E+14	2.33E+07	5E+14	1.00	7.16E+05	1.0000	0.0000
Belize	Conflict	2.72E+07	7E+14	2.72E+07	7E+14	1.35E+05	1.00	0.0000	1.0000
Benin	No Conflict	28	772	62	1885	2.20	1.83	0.4550	0.5450
Bhutan	No Conflict	4928	2.43E+07	5026	2.43E+07	1.02	51.06	0.9804	0.0196
Bolivia	Conflict	1.36	2.11	2.37	2.12	3.01	1.50	0.3326	0.6674
Bosnia and Herzegovina	No Conflict	1.23	0.29	2.39	0.46	1.93	2.07	0.5170	0.4830
Botswana	No Conflict	22	457	41	813	2	2.13	0.5301	0.4699
Brazil	Conflict	9408	8.85E+07	9410	8.85E+07	5727.78	1.00	0.0002	0.9998
Brunei Darussalam	Conflict	0.96	344	359	128360	1.05	20	0.9495	0.0505
Bulgaria	No Conflict	4104	1.68E+07	4105	1.68E+07	1.00	3464	0.9997	0.0003
Burkina Faso	Conflict	14	398	23	478	3.16	1.46	0.3164	0.6836
Burundi	Conflict	0.08	11	189	35571	1.02	51	0.9802	0.0198
Cabo Verde	No Conflict	184	33514	185	33514	1.01	172.58	0.9942	0.0058
Cambodia	Conflict	1.56	93	58	3204	1.18	6.53	0.8468	0.1532
Cameroon	No Conflict	1.68	1.15	3.53	2.72	2.10	1.91	0.4762	0.5238
Canada	No Conflict	3.39E+05	1E+11	3.39E+05	1E+11	1.00	17227	0.9999	0.0001

Table 51 Continued

Country	Status 2014	Sojourn Times (Ri)				Mean Conflict Recurrence Steps (M_i)		Long Run Conflict Probabilites (πj)	
		Number of Years to 1st transition:	Variance	Number Years to 2nd Transition:	Variance	E[Steps] to Transition from No Conflict to No Conflict	E[Steps] to Transition from Conflict to Conflict	Probability Not in Conflict	Probability in Conflict
Central African Republic	Conflict	4.55E+11	2E+23	4.55E+11	2E+23	4.42E+11	1.00	0.0000	1.0000
Chad	Conflict	46	2146	47	2146	41	1.02	0.0242	0.9758
Chile	Conflict	3.35	13	4.35	13	5.12	1.24	0.1952	0.8048
China	Conflict	4314	2E+07	4326	2E+07	385	1.00	0.0026	0.9974
Colombia	Conflict	4.23	71	14	162	1.89	2.12	0.5281	0.4719
Comoros	No Conflict	52	2683	61	2747	1.16	7.12	0.8596	0.1404
Congo, Republic of the	No Conflict	193	37168	194	37168	1.01	184	0.9946	0.0054
Costa Rica	No Conflict	24	537	115	8793	4.86	1.26	0.2058	0.7942
Cote d'Ivoire	Conflict	21	1061	40	1397	2.76	1.57	0.3626	0.6374
Croatia	No Conflict	33	1065	111942	1E+10	3379	1.00	0.0003	0.9997
Cuba	No Conflict	4.53E+27	2E+55	4.53E+27	2E+55	1.00	1.00E+12	1.0000	0.0000
Cyprus	No Conflict	7475	6E+07	3.11E+05	9E+10	42	1.02	0.0240	0.9760
Czech Republic	No Conflict	2.93E+05	9E+10	2.93E+05	9E+10	1.00	5.00E+04	1.0000	0.0000
Korea, North	No Conflict	7.78	53	32	609	4.10	1.32	0.2440	0.7560
Congo, Democratic Republic of the	Conflict	5403	3E+07	5415	3E+07	461	1.00	0.0022	0.9978
Denmark	No Conflict	3.78E+08	1E+17	3.78E+08	1E+17	1.00	9.38E+06	1.0000	0.0000
Djibouti	Conflict	0.12	0.16	10	80	1.12	9.32	0.8927	0.1073
Dominican Republic	No Conflict	3.35	7.85	5.02	8.98	1.50	3.00	0.6667	0.3333
Ecuador	Conflict	1.23	1.75	2.29	1.82	2.80	1.55	0.3569	0.6431
Egypt	Conflict	1.20	1.56	2.20	1.56	2.84	1.54	0.3516	0.6484
El Salvador	Conflict	1.39E+05	2E+10	1.41E+05	2E+10	67	1.02	0.0149	0.9851
Equatorial Guinea	No Conflict	8.38	62	9.41	62	1.12	9.11	0.8903	0.1097
Eritrea	Conflict	1.77	29	16	206	1.43	3.34	0.7002	0.2998
Estonia	No Conflict	1.64E+05	3E+10	1.64E+05	3E+10	1.00	51983	1.0000	0.0000
Ethiopia	Conflict	20	442	22	442	21	1.05	0.0479	0.9521
Fiji	No Conflict	92	8373	8449	7E+07	92	1.01	0.0109	0.9891
Finland	No Conflict	1.08E+08	1E+16	1.08E+08	1E+16	1.00	5.08E+06	1.0000	0.0000
France	No Conflict	51421	3E+09	51459	3E+09	1.00	1357	0.9993	0.0007
Gabon	Conflict	0.11	0.00	8.18	57	1.12	9.04	0.8894	0.1106
Gambia	No Conflict	27	712	29	713	1.06	18.57	0.9462	0.0538
Georgia	No Conflict	13	153	33	545	2.58	1.63	0.3882	0.6118

Table 51 Continued

Country	Status 2014	Sojourn Times (Ri)				Mean Conflict Recurrence Steps (M_j)		Long Run Conflict Probabilites (πj)	
		Number of Years to 1st transition:	Variance	Number Years to 2nd Transition:	Variance	E[Steps] to Transition from No Conflict to No Conflict	E[Steps] to Transition from Conflict to Conflict	*Probability Not in Conflict*	*Probability in Conflict*
Germany	No Conflict	68683	5E+09	68687	5E+09	1.00	18509	0.9999	0.0001
Ghana	No Conflict	128	16260	148	16653	1.16	7.30	0.8629	0.1371
Greece	Conflict	0.14	15	130	16643	1.03	31	0.9673	0.0327
Grenada	No Conflict	913	8E+05	2.29E+05	5E+10	250	1.00	0.0040	0.9960
Guatemala	Conflict	151	23535	154	23541	52	1.02	0.0190	0.9810
Guinea	No Conflict	15	199	17	201	1.14	7.93	0.8740	0.1260
Guinea-Bissau	No Conflict	6.29	33	7.84	34	1.25	5.05	0.8020	0.1980
Guyana	No Conflict	2.25E+05	5E+10	1.65E+09	3E+18	7339	1.00	0.0001	0.9999
Haiti	Conflict	0.72	3.16	5.99	26	1.45	3.25	0.6919	0.3081
Honduras	Conflict	0.16	6.33	55	3013	1.06	19.05	0.9475	0.0525
Hungary	No Conflict	5636	3.18E+07	5640	3.18E+07	1.00	1333	0.9992	0.0008
Iceland	No Conflict	1.46E+12	2E+24	1.46E+12	2E+24	1.00	2.21E+09	1.0000	0.0000
India	Conflict	14123	2E+08	14135	2E+08	1096	1.00	0.0009	0.9991
Indonesia	Conflict	1.76E+06	3E+12	1.76E+06	3E+12	8.14E+05	1.00	0.0000	1.0000
Iran	Conflict	60	3650	61	3650	45.23	1.02	0.0221	0.9779
Iraq	Conflict	2.06E+06	4E+12	2.06E+06	4E+12	2.06E+06	1.00	0.0000	1.0000
Ireland	No Conflict	8.83E+10	8E+21	8.83E+10	8E+21	1.00	1.43E+08	1.0000	0.0000
Israel	Conflict	53	3366	59	3398	10.58	1.10	0.0945	0.9055
Italy	No Conflict	2406	6E+06	2409	6E+06	1.00	844.94	0.9988	0.0012
Jamaica	Conflict	13541	2E+08	13543	2E+08	7092	1.00	0.0001	0.9999
Japan	No Conflict	1.43E+05	2E+10	1.43E+05	2E+10	1.00	68705	1.0000	0.0000
Jordan	Conflict	0.93	0.83	1.93	0.83	2.54	1.65	0.3937	0.6063
Kazakhstan	No Conflict	8.21	59	4568	2E+07	556.35	1.00	0.0018	0.9982
Kenya	Conflict	57	3549	60	3555	21.41	1.05	0.0467	0.9533
Kiribati	No Conflict	1.16	0.18	2.19	0.22	1.89	2.12	0.5284	0.4716
Kuwait	Conflict	1.56E+13	2E+26	1.56E+13	2E+26	1.56E+13	1.00	0.0000	1.0000
Kyrgyzstan	Conflict	0.00	0.01	780	607095	1.00	774	0.9987	0.0013
Laos	No Conflict	3.80	11	10	46	2.71	1.58	0.3689	0.6311
Latvia	No Conflict	22805	5E+08	22806	5E+08	1.00	14045	0.9999	0.0001
Lebanon	Conflict	314	98949	315	98949	316	1.00	0.0032	0.9968

Table 51 Continued

Country	Status 2014	Sojourn Times (Ri)				Mean Conflict Recurrence Steps (M_j)		Long Run Conflict Probabilites (πj)	
		Number of Years to 1st transition:	Variance	Number Years to 2nd Transition:	Variance	E[Steps] to Transition from No Conflict to No Conflict	E[Steps] to Transition from Conflict to Conflict	Probability Not in Conflict	Probability in Conflict
Lesotho	Conflict	3.62	93	22	408	1.56	2.80	0.6430	0.3570
Liberia	No Conflict	271	73010	282	73116	1.04	26	0.9616	0.0384
Libya	Conflict	0.50	0.00	1.50	0.00	2.00	2.00	0.5000	0.5000
Lithuania	No Conflict	97	9313	98	9313	1.01	95.88	0.9896	0.0104
Luxembourg	No Conflict	3.32E+06	1E+13	3.32E+06	1E+13	1.00	1.02E+05	1.0000	0.0000
Madagascar	No Conflict	67	4380	68	4381	1.03	38	0.9737	0.0263
Malawi	No Conflict	215	46152	396	78570	1.84	2.19	0.5439	0.4561
Malaysia	Conflict	49	3070	57	3124	8.12	1.14	0.1232	0.8768
Maldives	No Conflict	4.72	18	6.20	18	1.31	4.20	0.7620	0.2380
Mali	Conflict	19	514	24	536	5.43	1.23	0.1843	0.8157
Malta	No Conflict	76	5705	1.00E+15	1E+30	1.32E+13	1.00	0.0000	1.0000
Mauritania	No Conflict	21	415	24	424	1.16	7.11	0.8593	0.1407
Mauritius	No Conflict	207	42665	208	42665	1.00	207	0.9952	0.0048
Mexico	Conflict	1347	2E+06	1368	2E+06	64	1.02	0.0155	0.9845
Micronesia, Federated States of	No Conflict	5.62	26	2.74E+06	8E+12	4.88E+05	1.00	0.0000	1.0000
Mongolia	No Conflict	3.38	8.02	70	4387	21	1.05	0.0482	0.9518
Montenegro	No Conflict	15	221	28	373	1.84	2.20	0.5448	0.4552
Morocco	Conflict	0.68	0.29	1.68	0.29	2.24	1.81	0.4472	0.5528
Mozambique	Conflict	30	2110	55	2703	2.87	1.54	0.3486	0.6514
Myanmar	Conflict	2183	5E+06	2241	5E+06	40	1.03	0.0251	0.9749
Namibia	No Conflict	23	491	24	491	1.06	18	0.9436	0.0564
Nepal	No Conflict	3.63	10	4.56E+06	2E+13	1.26E+06	1.00	0.0000	1.0000
Netherlands	No Conflict	1.28E+07	2E+14	1.28E+07	2E+14	1.00	3.82E+05	1.0000	0.0000
New Zealand	No Conflict	32719	1E+09	32795	1E+09	1.00	432	0.9977	0.0023
Nicaragua	Conflict	6.91E+05	5E+11	6.91E+05	5E+11	90663	1.00	0.0000	1.0000
Niger	Conflict	4.99	43	8.00	49	3.36	1.42	0.2978	0.7022
Nigeria	Conflict	30	1370	39	1452	4.93	1.25	0.2030	0.7970
Norway	No Conflict	1.00E+11	1E+22	1.00E+11	1E+22	1.00	6.22E+08	1.0000	0.0000
Oman	No Conflict	1.00	0.00	54658	3E+09	54658	1.00	0.0000	1.0000
Pakistan	Conflict	6.37E+09	4E+19	6.37E+09	4E+19	3.93E+09	1.00	0.0000	1.0000

Table 51 Continued

Country	Status 2014	Sojourn Times (Ri)				Mean Conflict Recurrence Steps (M_{ij})		Long Run Conflict Probabilites (πj)	
		Number of Years to 1st transition:	Variance	Number Years to 2nd Transition:	Variance	E[Steps] to Transition from No Conflict to No Conflict	E[Steps] to Transition from Conflict to Conflict	Probability Not in Conflict	Probability in Conflict
Panama	No Conflict	1.43	0.61	11	89	7.96	1.14	0.1257	0.8743
Papua New Guinea	Conflict	6349	4.03E+07	6352	4.03E+07	2468	1.00	0.0004	0.9996
Paraguay	Conflict	0.04	118	3364	1.13E+07	1.00	297	0.9966	0.0034
Peru	Conflict	75	5756	76	5756	77	1.01	0.0129	0.9871
Philippines	Conflict	1.05E+05	1E+10	1.05E+05	1E+10	74588	1.00	0.0000	1.0000
Poland	No Conflict	5.49	25	8.11	29	1.48	3.10	0.6776	0.3224
Portugal	No Conflict	12349	1.52E+08	12353	1.52E+08	1.00	3417	0.9997	0.0003
Qatar	No Conflict	1.00	7E-15	1.43E+11	2E+22	1.43E+11	1.00	0.0000	1.0000
Korea, South	No Conflict	1.03	0.03	2.27	0.33	2.21	1.83	0.4529	0.5471
Moldova	No Conflict	30	865	31	865	1.03	31	0.9673	0.0327
Romania	No Conflict	6.13	31	79162	6E+09	12908	1.00	0.0001	0.9999
Russia	Conflict	4940	2.55E+07	5056	2.55E+07	44.53	1.02	0.0225	0.9775
Rwanda	Conflict	0.84	7.59	11	92	1.34	3.95	0.7466	0.2534
Samoa	No Conflict	178	31372	1934	3115603	11	1.10	0.0918	0.9082
Sao Tome and Principe	No Conflict	553	305496	607	308349	1.10	11	0.9112	0.0888
Saudi Arabia	Conflict	30	903	31	903	32	1.03	0.0317	0.9683
Senegal	No Conflict	263	68867	279	69100	1.06	18	0.9434	0.0566
Serbia	Conflict	486	2.40E+05	491	2.40E+05	101	1.01	0.0099	0.9901
Seychelles	No Conflict	7368	5.43E+07	7369	5.43E+07	1.00	7362	0.9999	0.0001
Sierra Leone	No Conflict	8.44	63	71	3901	8.40	1.14	0.1191	0.8809
Singapore	No Conflict	630	3.97E+05	631	3.97E+05	1.00	631	0.9984	0.0016
Slovakia	No Conflict	5160	2.66E+07	5161	2.66E+07	1.00	5071	0.9998	0.0002
Slovenia	No Conflict	1.09E+05	1E+10	1.09E+05	1E+10	1.00	18163	0.9999	0.0001
Solomon Islands	No Conflict	45	1945	46	1946	1.03	35	0.9712	0.0288
Somalia	Conflict	1.03E+06	1E+12	1.03E+06	1E+12	2.14E+05	1.00	0.0000	1.0000
South Africa	Conflict	0.50	3.11	9.04	67	1.27	4.66	0.7855	0.2145
South Sudan	Conflict	4.28E+10	2E+21	4.28E+10	2E+21	2.65E+10	1.00	0.0000	1.0000
Spain	No Conflict	2.33E+05	5E+10	2.33E+05	5E+10	1.00	21248	1.0000	0.0000
Sri Lanka	Conflict	24677	6.09E+08	24680	6.09E+08	8325	1.00	0.0001	0.9999
Sudan	Conflict	1613	2.60E+06	1614	2.60E+06	1590	1.00	0.0006	0.9994

Table 51 Continued

Country	Status 2014	Sojourn Times (Ri)				Mean Conflict Recurrence Steps (M_j)		Long Run Conflict Probabilites (πj)	
		Number of Years to 1st transition:	Variance	Number Years to 2nd Transition:	Variance	E[Steps] to Transition from No Conflict to No Conflict	E[Steps] to Transition from Conflict to Conflict	Probability Not in Conflict	Probability in Conflict
Suriname	No Conflict	20	397	1.47E+09	2E+18	7.21E+07	1.00	0.0000	1.0000
Swaziland	No Conflict	12	139	13	139	1.09	11.73	0.9147	0.0853
Sweden	No Conflict	3.15E+08	1E+17	3.15E+08	1E+17	1.00	8.85E+06	1.0000	0.0000
Switzerland	No Conflict	179	31772	185	31800	1.03	31	0.9682	0.0318
Syria	Conflict	2.26	6.12	3.28	6.14	3.94	1.34	0.2536	0.7464
Tajikistan	Conflict	3.00E+05	9E+10	3.00E+05	9E+10	2.70E+05	1.00	0.0000	1.0000
Thailand	Conflict	560	4.44E+05	687	4.59E+05	6.26	1.19	0.1596	0.8404
Macedonia	Conflict	0.37	3.10	13	139	1.19	6.23	0.8394	0.1606
Timor-Leste	No Conflict	4.95	20	692	4.71E+05	140	1.01	0.0072	0.9928
Togo	Conflict	0.03	1.14	103	10447	1.02	62	0.9839	0.0161
Tonga	No Conflict	899	8.08E+05	949	8.11E+05	1.06	19	0.9474	0.0526
Trinidad and Tobago	No Conflict	12	123	4.99E+08	2E+17	4.29E+07	1.00	0.0000	1.0000
Tunisia	Conflict	0.51	0.01	1.51	0.01	2.01	1.99	0.4977	0.5023
Turkey	Conflict	94	8972	95	8972	96	1.01	0.0104	0.9896
Turkmenistan	No Conflict	1991	3.96E+06	1992	3.96E+06	1.00	1992	0.9995	0.0005
Uganda	Conflict	45	2949	57	3095	5.35	1.23	0.1870	0.8130
Ukraine	Conflict	19	362	20	362	20	1.05	0.0489	0.9511
United Arab Emirates	No Conflict	1.00	0	1.42E+11	2E+22	1.42E+11	1.00	0.0000	1.0000
United Kingdom	Conflict	0.04	241	5995	3.59E+07	1.00	375	0.9973	0.0027
Tanzania	Conflict	0.15	2.74	32	973	1.07	15	0.9342	0.0658
United States	Conflict	0.82	1.20	2.64	2.68	1.94	2.07	0.5160	0.4840
Uruguay	No Conflict	359	1.29E+05	845576	7E+11	2353	1.00	0.0004	0.9996
Uzbekistan	No Conflict	8.93	71	26	358	2.95	1.51	0.3385	0.6615
Vanuatu	No Conflict	25	590	26	590	1.05	22	0.9548	0.0452
Venezuela	Conflict	4.03	19	5.12	19	5.51	1.22	0.1816	0.8184
Vietnam	Conflict	0.42	0.12	2.23	1.58	1.62	2.63	0.6191	0.3809
West Bank	Conflict	18	349	19	349	20	1.05	0.0495	0.9505
Yemen	Conflict	1238	1.53E+06	1239	1.53E+06	1240	1.00	0.0008	0.9992
Zambia	No Conflict	140	19499	188	21787	1.34	3.90	0.7436	0.2564
Zimbabwe	Conflict	0.39	0.58	4.09	11	1.38	3.63	0.7242	0.2758

Appendix D: Regional Conflict Forecasts for 2016, 2019, and 2024

Table 52: Arab & North African States 2, 5, and 10 Year Forecasts

Country	Status 2014	Conditional Conflict Transition Probabilities given 2014 Status								10 Year Trend
		2014		2016		2019		2024		
		P(No Conflict)	P(Conflict)	P(No Conflict)	P(Conflict)	P(No Conflict)	P(Conflict)	P(No Conflict)	P(Conflict)	
Algeria	Conflict	0.028	0.972	0.027	0.973	0.027	0.973	0.027	0.973	No Change
Bahrain	Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
Egypt	Conflict	0.542	0.458	0.248	0.752	0.368	0.632	0.351	0.649	Slightly More Conflict
Iraq	Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
Jordan	Conflict	0.649	0.351	0.228	0.772	0.439	0.561	0.388	0.612	Moderately More Conflict
Kuwait	Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
Lebanon	Conflict	0.003	0.997	0.003	0.997	0.003	0.997	0.003	0.997	No Change
Libya	Conflict	1.000	0.000	0.000	1.000	1.000	0.000	0.000	1.000	Significantly More Conflict
Morocco	Conflict	0.809	0.191	0.155	0.845	0.602	0.398	0.394	0.606	Moderately More Conflict
Oman	No Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
Qatar	No Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
Saudi Arabia	Conflict	0.033	0.967	0.032	0.968	0.032	0.968	0.032	0.968	No Change
Syria	Conflict	0.331	0.669	0.230	0.770	0.254	0.746	0.254	0.746	Slightly More Conflict
Tunisia	Conflict	0.991	0.009	0.009	0.991	0.973	0.027	0.044	0.956	Significantly More Conflict
United Arab Emirates	No Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
West Bank	Conflict	0.052	0.948	0.049	0.951	0.050	0.950	0.050	0.950	No Change
Yemen	Conflict	0.001	0.999	0.001	0.999	0.001	0.999	0.001	0.999	No Change

Table 53: Eastern Europe & Central Asia 2, 5, and 10 Year Forecasts

Country	Status 2014	Conditional Conflict Transition Probabilities given 2014 Status								10 Year Trend
		2014		2016		2019		2024		
		P(No Conflict)	P(Conflict)	P(No Conflict)	P(Conflict)	P(No Conflict)	P(Conflict)	P(No Conflict)	P(Conflict)	
Afghanistan	Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
Albania	No Conflict	0.924	0.076	0.855	0.145	0.676	0.324	0.460	0.540	Moderately More Conflict
Armenia	Conflict	0.009	0.991	0.009	0.991	0.009	0.991	0.009	0.991	No Change
Azerbaijan	Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
Belarus	No Conflict	0.585	0.415	0.745	0.255	0.698	0.302	0.701	0.299	Slightly Less Conflict
Bosnia and Herzegovina	No Conflict	0.189	0.811	0.739	0.261	0.448	0.552	0.527	0.473	Moderately Less Conflict
Bulgaria	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Croatia	No Conflict	0.970	0.030	0.941	0.059	0.858	0.142	0.736	0.264	Slightly More Conflict
Cyprus	No Conflict	1.000	0.000	1.000	0.000	0.999	0.001	0.999	0.001	No Change
Georgia	No Conflict	0.922	0.078	0.855	0.145	0.699	0.301	0.546	0.454	Moderately More Conflict
Iran	Conflict	0.016	0.984	0.021	0.979	0.022	0.978	0.022	0.978	No Change
Kazakhstan	No Conflict	0.878	0.122	0.771	0.229	0.523	0.477	0.274	0.726	Significantly More Conflict
Kyrgyzstan	Conflict	0.992	0.008	0.999	0.001	0.999	0.001	0.999	0.001	No Change
Latvia	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Lithuania	No Conflict	0.990	0.010	0.990	0.010	0.990	0.010	0.990	0.010	No Change
Malta	No Conflict	0.987	0.013	0.974	0.026	0.936	0.064	0.876	0.124	Slightly More Conflict
Montenegro	No Conflict	0.935	0.065	0.879	0.121	0.755	0.245	0.642	0.358	Moderately More Conflict
Pakistan	Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
Moldova	No Conflict	0.967	0.033	0.967	0.033	0.967	0.033	0.967	0.033	No Change
Romania	No Conflict	0.837	0.163	0.700	0.300	0.411	0.589	0.169	0.831	Significantly More Conflict
Russia	Conflict	0.000	1.000	0.000	1.000	0.001	0.999	0.002	0.998	No Change
Serbia	Conflict	0.002	0.998	0.004	0.996	0.007	0.993	0.009	0.991	No Change
Slovakia	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Tajikistan	Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
Macedonia	Conflict	0.429	0.571	0.639	0.361	0.816	0.184	0.839	0.161	Moderately Less Conflict
Turkmenistan	No Conflict	0.999	0.001	0.999	0.001	0.999	0.001	0.999	0.001	No Change
Ukraine	Conflict	0.051	0.949	0.049	0.951	0.049	0.951	0.049	0.951	No Change
Uzbekistan	No Conflict	0.888	0.112	0.795	0.205	0.600	0.400	0.442	0.558	Moderately More Conflict

Table 54: Latin America 2, 5, and 10 Year Forecasts

Country	Status 2014	Conditional Conflict Transition Probabilities given 2014 Status								10 Year Trend
		2014		2016		2019		2024		
		P(No Conflict)	P(Conflict)	P(No Conflict)	P(Conflict)	P(No Conflict)	P(Conflict)	P(No Conflict)	P(Conflict)	
Antigua and Barbuda	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Argentina	No Conflict	0.997	0.003	0.995	0.005	0.993	0.007	0.993	0.007	No Change
Bahamas	No Conflict	0.997	0.003	0.994	0.006	0.984	0.016	0.968	0.032	No Change
Barbados	No Conflict	0.997	0.003	0.995	0.005	0.987	0.013	0.973	0.027	No Change
Belize	Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
Bolivia	Conflict	0.491	0.509	0.257	0.743	0.341	0.659	0.332	0.668	Slightly More Conflict
Brazil	Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
Colombia	Conflict	0.112	0.888	0.200	0.800	0.367	0.633	0.479	0.521	Moderately Less Conflict
Costa Rica	No Conflict	0.958	0.042	0.918	0.082	0.810	0.190	0.666	0.334	Moderately More Conflict
Cuba	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Dominican Republic	No Conflict	0.701	0.299	0.670	0.330	0.667	0.333	0.667	0.333	No Change
Ecuador	Conflict	0.522	0.478	0.280	0.720	0.365	0.635	0.357	0.643	Slightly More Conflict
El Salvador	Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
Grenada	No Conflict	0.999	0.001	0.998	0.002	0.995	0.005	0.989	0.011	No Change
Guatemala	Conflict	0.006	0.994	0.011	0.989	0.017	0.983	0.019	0.981	No Change
Guyana	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Haiti	Conflict	0.426	0.574	0.590	0.410	0.686	0.314	0.692	0.308	Moderately Less Conflict
Honduras	Conflict	0.326	0.674	0.540	0.460	0.833	0.167	0.934	0.066	Significantly Less Conflict
Jamaica	Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
Nicaragua	Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
Panama	No Conflict	0.299	0.701	0.160	0.840	0.126	0.874	0.126	0.874	Slightly More Conflict
Paraguay	Conflict	0.088	0.912	0.168	0.832	0.368	0.632	0.601	0.399	Significantly Less Conflict
Peru	Conflict	0.013	0.987	0.013	0.987	0.013	0.987	0.013	0.987	No Change
Suriname	No Conflict	0.951	0.049	0.905	0.095	0.778	0.222	0.605	0.395	Moderately More Conflict
Trinidad and Tobago	No Conflict	0.914	0.086	0.835	0.165	0.638	0.362	0.407	0.593	Significantly More Conflict
Uruguay	No Conflict	0.997	0.003	0.994	0.006	0.986	0.014	0.973	0.027	No Change
Venezuela	Conflict	0.203	0.797	0.179	0.821	0.182	0.818	0.182	0.818	No Change

Table 55: OECD 2, 5, and 10 Year Forecasts

Country	Status 2014	Conditional Conflict Transition Probabilities given 2014 Status								10 Year Trend
		2014		2016		2019		2024		
		P(No Conflict)	P(Conflict)	P(No Conflict)	P(Conflict)	P(No Conflict)	P(Conflict)	P(No Conflict)	P(Conflict)	
Australia	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Austria	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Belgium	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Canada	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Chile	Conflict	0.241	0.759	0.185	0.815	0.195	0.805	0.195	0.805	No Change
Czech Republic	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Denmark	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Estonia	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Finland	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
France	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Germany	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Greece	Conflict	0.228	0.772	0.403	0.597	0.716	0.284	0.902	0.098	Significantly Less Conflict
Hungary	No Conflict	1.000	0.000	1.000	0.000	0.999	0.001	0.999	0.001	No Change
Iceland	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Ireland	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Israel	Conflict	0.017	0.983	0.031	0.969	0.060	0.940	0.082	0.918	Slightly Less Conflict
Italy	No Conflict	1.000	0.000	0.999	0.001	0.999	0.001	0.999	0.001	No Change
Japan	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Luxembourg	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Mexico	Conflict	0.001	0.999	0.001	0.999	0.003	0.997	0.006	0.994	No Change
Netherlands	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
New Zealand	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Norway	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Poland	No Conflict	0.818	0.182	0.739	0.261	0.683	0.317	0.678	0.322	Slightly More Conflict
Portugal	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Korea, South	No Conflict	0.028	0.972	0.782	0.218	0.299	0.701	0.496	0.504	Moderately Less Conflict
Slovenia	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Spain	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Sweden	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Switzerland	No Conflict	0.994	0.006	0.990	0.010	0.980	0.020	0.973	0.027	No Change
Turkey	Conflict	0.011	0.989	0.010	0.990	0.010	0.990	0.010	0.990	No Change
United Kingdom	Conflict	0.062	0.938	0.121	0.879	0.275	0.725	0.475	0.525	Moderately Less Conflict
United States	Conflict	0.587	0.413	0.506	0.494	0.516	0.484	0.516	0.484	Slightly More Conflict

Table 56: South & East Asia 2, 5, and 10 Year Forecasts

Country	Status 2014	Conditional Conflict Transition Probabilities given 2014 Status								10 Year Trend
		2014		2016		2019		2024		
		P(No Conflict)	P(Conflict)	P(No Conflict)	P(Conflict)	P(No Conflict)	P(Conflict)	P(No Conflict)	P(Conflict)	
Bangladesh	Conflict	0.004	0.996	0.005	0.995	0.005	0.995	0.005	0.995	No Change
Bhutan	No Conflict	1.000	0.000	1.000	0.000	0.999	0.001	0.998	0.002	No Change
Brunei Darussalam	Conflict	0.052	0.948	0.102	0.898	0.235	0.765	0.412	0.588	Moderately Less Conflict
Cambodia	Conflict	0.098	0.902	0.185	0.815	0.390	0.610	0.600	0.400	Significantly Less Conflict
China	Conflict	0.000	1.000	0.000	1.000	0.001	0.999	0.002	0.998	No Change
Korea, North	No Conflict	0.871	0.129	0.765	0.235	0.542	0.458	0.361	0.639	Significantly More Conflict
Fiji	No Conflict	0.989	0.011	0.978	0.022	0.947	0.053	0.897	0.103	Slightly More Conflict
India	Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.001	0.999	No Change
Indonesia	Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
Kiribati	No Conflict	0.137	0.863	0.854	0.146	0.342	0.658	0.602	0.398	Moderately Less Conflict
Laos	No Conflict	0.737	0.263	0.583	0.417	0.411	0.589	0.372	0.628	Moderately More Conflict
Malaysia	Conflict	0.018	0.982	0.033	0.967	0.067	0.933	0.098	0.902	Slightly Less Conflict
Maldives	No Conflict	0.788	0.212	0.765	0.235	0.762	0.238	0.762	0.238	No Change
Micronesia, Federated States of	No Conflict	0.822	0.178	0.676	0.324	0.375	0.625	0.141	0.859	Significantly More Conflict
Mongolia	No Conflict	0.704	0.296	0.500	0.500	0.196	0.804	0.071	0.929	Significantly More Conflict
Myanmar	Conflict	0.000	1.000	0.001	0.999	0.002	0.998	0.004	0.996	No Change
Nepal	No Conflict	0.725	0.275	0.525	0.475	0.200	0.800	0.040	0.960	Significantly More Conflict
Papua New Guinea	Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
Philippines	Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
Samoa	No Conflict	0.994	0.006	0.989	0.011	0.972	0.028	0.945	0.055	No Change
Singapore	No Conflict	0.998	0.002	0.998	0.002	0.998	0.002	0.998	0.002	No Change
Solomon Islands	No Conflict	0.978	0.022	0.973	0.027	0.971	0.029	0.971	0.029	No Change
Sri Lanka	Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
Thailand	Conflict	0.002	0.998	0.003	0.997	0.007	0.993	0.014	0.986	No Change
Timor-Leste	No Conflict	0.798	0.202	0.637	0.363	0.325	0.675	0.109	0.891	Significantly More Conflict
Tonga	No Conflict	0.999	0.001	0.998	0.002	0.995	0.005	0.990	0.010	No Change
Vanuatu	No Conflict	0.960	0.040	0.955	0.045	0.955	0.045	0.955	0.045	No Change
Vietnam	Conflict	0.900	0.100	0.492	0.508	0.631	0.369	0.619	0.381	Moderately More Conflict

Table 57: Sub-Saharan Africa 2, 5, and 10 Year Forecasts

Country	Status 2014	Conditional Conflict Transition Probabilities given 2014 Status								10 Year Trend
		2014		2016		2019		2024		
		P(No Conflict)	P(Conflict)	P(No Conflict)	P(Conflict)	P(No Conflict)	P(Conflict)	P(No Conflict)	P(Conflict)	
Angola	No Conflict	0.762	0.238	0.595	0.405	0.334	0.666	0.218	0.782	Significantly More Conflict
Benin	No Conflict	0.965	0.035	0.932	0.068	0.845	0.155	0.734	0.266	Slightly More Conflict
Botswana	No Conflict	0.954	0.046	0.913	0.087	0.812	0.188	0.699	0.301	Moderately More Conflict
Burkina Faso	Conflict	0.049	0.951	0.090	0.910	0.180	0.820	0.257	0.743	Slightly Less Conflict
Burundi	Conflict	0.262	0.738	0.454	0.546	0.773	0.227	0.936	0.064	Significantly Less Conflict
Cabo Verde	No Conflict	0.995	0.005	0.994	0.006	0.994	0.006	0.994	0.006	No Change
Cameroon	No Conflict	0.406	0.594	0.486	0.514	0.476	0.524	0.476	0.524	Slightly Less Conflict
Central African Republic	Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
Chad	Conflict	0.021	0.979	0.024	0.976	0.024	0.976	0.024	0.976	No Change
Comoros	No Conflict	0.981	0.019	0.964	0.036	0.927	0.073	0.892	0.108	Slightly More Conflict
Congo, Republic of the	No Conflict	0.995	0.005	0.995	0.005	0.995	0.005	0.995	0.005	No Change
Cote d'Ivoire	Conflict	0.030	0.970	0.058	0.942	0.128	0.872	0.211	0.789	Slightly Less Conflict
Congo, Democratic Republic of the	Conflict	0.000	1.000	0.000	1.000	0.001	0.999	0.001	0.999	No Change
Djibouti	Conflict	0.879	0.121	0.893	0.107	0.893	0.107	0.893	0.107	No Change
Equatorial Guinea	No Conflict	0.881	0.119	0.891	0.109	0.890	0.110	0.890	0.110	No Change
Eritrea	Conflict	0.169	0.831	0.297	0.703	0.524	0.476	0.656	0.344	Moderately Less Conflict
Ethiopia	Conflict	0.046	0.954	0.048	0.952	0.048	0.952	0.048	0.952	No Change
Gabon	Conflict	0.996	0.004	0.877	0.123	0.889	0.111	0.889	0.111	Slightly More Conflict
Gambia	No Conflict	0.963	0.037	0.952	0.048	0.946	0.054	0.946	0.054	No Change
Ghana	No Conflict	0.992	0.008	0.985	0.015	0.965	0.035	0.939	0.061	Slightly More Conflict
Guinea	No Conflict	0.931	0.069	0.900	0.100	0.876	0.124	0.874	0.126	Slightly More Conflict
Guinea-Bissau	No Conflict	0.841	0.159	0.810	0.190	0.802	0.198	0.802	0.198	No Change
Kenya	Conflict	0.017	0.983	0.027	0.973	0.042	0.958	0.046	0.954	No Change
Lesotho	Conflict	0.099	0.901	0.182	0.818	0.363	0.637	0.521	0.479	Moderately Less Conflict
Liberia	No Conflict	0.996	0.004	0.993	0.007	0.985	0.015	0.976	0.024	No Change

Table 57 Continued

Country	Status 2014	Conditional Conflict Transition Probabilities given 2014 Status								10 Year Trend
		2014		2016		2019		2024		
		P(No Conflict)	P(Conflict)	P(No Conflict)	P(Conflict)	P(No Conflict)	P(Conflict)	P(No Conflict)	P(Conflict)	
Madagascar	No Conflict	0.985	0.015	0.979	0.021	0.974	0.026	0.974	0.026	No Change
Malawi	No Conflict	0.995	0.005	0.991	0.009	0.977	0.023	0.956	0.044	No Change
Mali	Conflict	0.043	0.957	0.076	0.924	0.136	0.864	0.172	0.828	Slightly Less Conflict
Mauritania	No Conflict	0.952	0.048	0.921	0.079	0.877	0.123	0.862	0.138	Slightly More Conflict
Mauritius	No Conflict	0.995	0.005	0.995	0.005	0.995	0.005	0.995	0.005	No Change
Mozambique	Conflict	0.022	0.978	0.042	0.958	0.095	0.905	0.164	0.836	Slightly Less Conflict
Namibia	No Conflict	0.956	0.044	0.946	0.054	0.944	0.056	0.944	0.056	No Change
Niger	Conflict	0.141	0.859	0.215	0.785	0.286	0.714	0.297	0.703	Slightly Less Conflict
Nigeria	Conflict	0.027	0.973	0.050	0.950	0.103	0.897	0.153	0.847	Slightly Less Conflict
Rwanda	Conflict	0.303	0.697	0.483	0.517	0.691	0.309	0.743	0.257	Moderately Less Conflict
Sao Tome and Principe	No Conflict	0.998	0.002	0.996	0.004	0.991	0.009	0.983	0.017	No Change
Senegal	No Conflict	0.996	0.004	0.993	0.007	0.983	0.017	0.972	0.028	No Change
Seychelles	No Conflict	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	No Change
Sierra Leone	No Conflict	0.882	0.118	0.779	0.221	0.547	0.453	0.327	0.673	Significantly More Conflict
Somalia	Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
South Africa	Conflict	0.429	0.571	0.624	0.376	0.770	0.230	0.785	0.215	Moderately Less Conflict
South Sudan	Conflict	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	No Change
Sudan	Conflict	0.001	0.999	0.001	0.999	0.001	0.999	0.001	0.999	No Change
Swaziland	No Conflict	0.919	0.081	0.915	0.085	0.915	0.085	0.915	0.085	No Change
Togo	Conflict	0.595	0.405	0.831	0.169	0.974	0.026	0.984	0.016	Moderately Less Conflict
Uganda	Conflict	0.018	0.982	0.035	0.965	0.075	0.925	0.120	0.880	Slightly Less Conflict
Tanzania	Conflict	0.448	0.552	0.682	0.318	0.899	0.101	0.933	0.067	Moderately Less Conflict
Zambia	No Conflict	0.993	0.007	0.986	0.014	0.966	0.034	0.937	0.063	Slightly More Conflict
Zimbabwe	Conflict	0.710	0.290	0.724	0.276	0.724	0.276	0.724	0.276	No Change

Appendix E: Transience Scores by Region

Table 58: Arab & North African States Transience Scores

Nation	2014 Conflict Status	Probability Not in Conflict	Probability in Conflict	Mi	Mj	Expected Number Recurrences per 10 Years	Transience Score	Rank
Libya	Conflict	0.50	0.50	2.00	2.00	2.5	1.00000	1
Tunisia	Conflict	0.50	0.50	2.01	1.99	2.5	0.99998	2
Morocco	Conflict	0.45	0.55	2.24	1.81	2.5	0.98884	3
Jordan	Conflict	0.39	0.61	2.54	1.65	2.4	0.95477	4
Egypt	Conflict	0.35	0.65	2.84	1.54	2.3	0.91189	5
Syria	Conflict	0.25	0.75	3.94	1.34	1.9	0.75724	6
West Bank	Conflict	0.05	0.95	20.19	1.05	0.5	0.18829	7
Saudi Arabia	Conflict	0.03	0.97	31.55	1.03	0.3	0.12275	8
Algeria	Conflict	0.03	0.97	37.22	1.03	0.3	0.10458	9
Lebanon	Conflict	0.00	1.00	316.06	1.00	0.0	0.01262	10
Yemen	Conflict	0.00	1.00	1239.96	1.00	0.0	0.00322	11
Oman	No Conflict	0.00	1.00	54657.53	1.00	0.0	0.00007	12
Iraq	Conflict	0.00	1.00	2.06E+06	1.00	0.0	0.00000	13
Bahrain	Conflict	0.00	1.00	1.42E+11	1.00	0.0	0.00000	14
United Arab Emirates	No Conflict	0.00	1.00	1.42E+11	1.00	0.0	0.00000	15
Qatar	No Conflict	0.00	1.00	1.43E+11	1.00	0.0	0.00000	16
Kuwait	Conflict	0.00	1.00	1.56E+13	1.00	0.0	0.00000	17

Table 59: Eastern Europe & Central Asia Transience Scores

Nation	2014 Conflict Status	Probability Not in Conflict	Probability in Conflict	Mi	Mj	Expected Number Recurrences per 10 Years	Transience Score	Rank
Bosnia and Herzegovina	No Conflict	0.52	0.48	1.93	2.07	2.5	0.99884	1
Montenegro	No Conflict	0.54	0.46	1.84	2.20	2.5	0.99199	2
Georgia	No Conflict	0.39	0.61	2.58	1.63	2.4	0.94998	3
Uzbekistan	No Conflict	0.34	0.66	2.95	1.51	2.2	0.89568	4
Belarus	No Conflict	0.70	0.30	1.43	3.34	2.1	0.83922	5
Macedonia	Conflict	0.84	0.16	1.19	6.23	1.3	0.53914	6
Ukraine	Conflict	0.05	0.95	20.44	1.05	0.5	0.18611	7
Moldova	No Conflict	0.97	0.03	1.03	30.54	0.3	0.12667	8
Albania	No Conflict	0.03	0.97	38.59	1.03	0.3	0.10096	9
Cyprus	No Conflict	0.02	0.98	41.64	1.02	0.2	0.09375	10
Russia	Conflict	0.02	0.98	44.53	1.02	0.2	0.08781	11
Iran	Conflict	0.02	0.98	45.23	1.02	0.2	0.08648	12
Lithuania	No Conflict	0.99	0.01	1.01	95.88	0.1	0.04128	13
Serbia	Conflict	0.01	0.99	100.87	1.01	0.1	0.03926	14
Armenia	Conflict	0.01	0.99	112.48	1.01	0.1	0.03524	15
Kazakhstan	No Conflict	0.00	1.00	556.35	1.00	0.0	0.00718	16
Kyrgyzstan	Conflict	1.00	0.00	1.00	774.33	0.0	0.00516	17
Turkmenistan	No Conflict	1.00	0.00	1.00	1992.11	0.0	0.00201	18
Afghanistan	Conflict	0.00	1.00	2584.87	1.00	0.0	0.00155	19
Croatia	No Conflict	0.00	1.00	3378.72	1.00	0.0	0.00118	20
Bulgaria	No Conflict	1.00	0.00	1.00	3463.90	0.0	0.00115	21
Slovakia	No Conflict	1.00	0.00	1.00	5070.88	0.0	0.00079	22
Azerbaijan	Conflict	0.00	1.00	5893.80	1.00	0.0	0.00068	23
Romania	No Conflict	0.00	1.00	12908.27	1.00	0.0	0.00031	24
Latvia	No Conflict	1.00	0.00	1.00	14044.51	0.0	0.00028	25
Tajikistan	Conflict	0.00	1.00	2.70E+05	1.00	0.0	0.00001	26
Pakistan	Conflict	0.00	1.00	3.93E+09	1.00	0.0	0.00000	27
Malta	No Conflict	0.00	1.00	1.32E+13	1.00	0.0	0.00000	28

Table 60: Latin American Transience Scores

Nation	2014 Conflict Status	Probability Not in Conflict	Probability in Conflict	Mi	Mj	Expected Number Recurrences per 10 Years	Transience Score	Rank
Colombia	Conflict	0.53	0.47	1.89	2.12	2.5	0.99685	1
Bahamas	No Conflict	0.38	0.62	2.63	1.61	2.4	0.94241	2
Ecuador	Conflict	0.36	0.64	2.80	1.55	2.3	0.91809	3
Dominican Republic	No Conflict	0.67	0.33	1.50	3.00	2.2	0.88888	4
Bolivia	Conflict	0.33	0.67	3.01	1.50	2.2	0.88788	5
Haiti	Conflict	0.69	0.31	1.45	3.25	2.1	0.85275	6
Barbados	No Conflict	0.22	0.78	4.45	1.29	1.7	0.69694	7
Costa Rica	No Conflict	0.21	0.79	4.86	1.26	1.6	0.65382	8
Venezuela	Conflict	0.18	0.82	5.51	1.22	1.5	0.59461	9
Panama	No Conflict	0.13	0.87	7.96	1.14	1.1	0.43951	10
Honduras	Conflict	0.95	0.05	1.06	19.05	0.5	0.19895	11
Guatemala	Conflict	0.02	0.98	52.50	1.02	0.2	0.07474	12
El Salvador	Conflict	0.01	0.99	67.25	1.02	0.1	0.05859	13
Peru	Conflict	0.01	0.99	77.37	1.01	0.1	0.05103	14
Argentina	No Conflict	0.99	0.01	1.01	146.94	0.1	0.02704	15
Grenada	No Conflict	0.00	1.00	250.43	1.00	0.0	0.01591	16
Paraguay	Conflict	1.00	0.00	1.00	296.51	0.0	0.01344	17
Uruguay	No Conflict	0.00	1.00	2352.97	1.00	0.0	0.00170	18
Brazil	Conflict	0.00	1.00	5727.78	1.00	0.0	0.00070	19
Jamaica	Conflict	0.00	1.00	7092.41	1.00	0.0	0.00056	20
Guyana	No Conflict	0.00	1.00	7338.81	1.00	0.0	0.00054	21
Nicaragua	Conflict	0.00	1.00	9.07E+04	1.00	0.0	0.00004	22
Belize	Conflict	0.00	1.00	1.35E+05	1.00	0.0	0.00003	23
Trinidad and Tobago	No Conflict	0.00	1.00	4.29E+07	1.00	0.0	0.00000	24
Suriname	No Conflict	0.00	1.00	7.21E+07	1.00	0.0	0.00000	25
Antigua and Barbuda	No Conflict	1.00	0.00	1.00	1.00E+36	0.0	0.00000	26
Cuba	No Conflict	1.00	0.00	1.00	1.00E+36	0.0	0.00000	27

Table 61: OECD Transience Scores

Nation	2014 Conflict Status	Probability Not in Conflict	Probability in Conflict	Mi	Mj	Expected Number Recurrences per 10 Years	Transience Score	Rank
United States	Conflict	0.52	0.48	1.94	2.07	2.5	0.99898	1
Korea, South	No Conflict	0.45	0.55	2.21	1.83	2.5	0.99112	2
Poland	No Conflict	0.68	0.32	1.48	3.10	2.2	0.87389	3
Chile	Conflict	0.20	0.80	5.12	1.24	1.6	0.62828	4
Israel	Conflict	0.09	0.91	10.58	1.10	0.9	0.34225	5
Greece	Conflict	0.97	0.03	1.03	30.56	0.3	0.12660	6
Switzerland	No Conflict	0.97	0.03	1.03	31.45	0.3	0.12313	7
Mexico	Conflict	0.02	0.98	64.32	1.02	0.2	0.06122	8
Turkey	Conflict	0.01	0.99	96.22	1.01	0.1	0.04114	9
United Kingdom	Conflict	1.00	0.00	1.00	375.30	0.0	0.01063	10
New Zealand	No Conflict	1.00	0.00	1.00	431.57	0.0	0.00925	11
Italy	No Conflict	1.00	0.00	1.00	844.94	0.0	0.00473	12
Hungary	No Conflict	1.00	0.00	1.00	1332.65	0.0	0.00300	13
France	No Conflict	1.00	0.00	1.00	1357.00	0.0	0.00295	14
Portugal	No Conflict	1.00	0.00	1.00	3416.88	0.0	0.00117	15
Canada	No Conflict	1.00	0.00	1.00	17226.97	0.0	0.00023	16
Slovenia	No Conflict	1.00	0.00	1.00	18162.87	0.0	0.00022	17
Germany	No Conflict	1.00	0.00	1.00	18508.67	0.0	0.00022	18
Spain	No Conflict	1.00	0.00	1.00	21247.61	0.0	0.00019	19
Czech Republic	No Conflict	1.00	0.00	1.00	49986.28	0.0	0.00008	20
Estonia	No Conflict	1.00	0.00	1.00	5.20E+04	0.0	0.00008	21
Japan	No Conflict	1.00	0.00	1.00	6.87E+04	0.0	0.00006	22
Australia	No Conflict	1.00	0.00	1.00	9.59E+04	0.0	0.00004	23
Luxembourg	No Conflict	1.00	0.00	1.00	1.02E+05	0.0	0.00004	24
Austria	No Conflict	1.00	0.00	1.00	1.32E+05	0.0	0.00003	25
Netherlands	No Conflict	1.00	0.00	1.00	3.82E+05	0.0	0.00001	26
Belgium	No Conflict	1.00	0.00	1.00	7.16E+05	0.0	0.00001	27
Finland	No Conflict	1.00	0.00	1.00	5.08E+06	0.0	0.00000	28
Sweden	No Conflict	1.00	0.00	1.00	8.85E+06	0.0	0.00000	29
Denmark	No Conflict	1.00	0.00	1.00	9.38E+06	0.0	0.00000	30
Ireland	No Conflict	1.00	0.00	1.00	1.43E+08	0.0	0.00000	31
Norway	No Conflict	1.00	0.00	1.00	6.22E+08	0.0	0.00000	32
Iceland	No Conflict	1.00	0.00	1.00	2.21E+09	0.0	0.00000	33

Table 62: South & East Asia Transience Scores

Nation	2014 Conflict Status	Probability Not in Conflict	Probability in Conflict	Mi	Mj	Expected Number Recurrences per 10 Years	Transience Score	Rank
Kiribati	No Conflict	0.53	0.47	1.89	2.12	2.5	0.99677	1
Vietnam	Conflict	0.62	0.38	1.62	2.63	2.4	0.94325	2
Laos	No Conflict	0.37	0.63	2.71	1.58	2.3	0.93130	3
Korea, North	No Conflict	0.24	0.76	4.10	1.32	1.8	0.73792	4
Maldives	No Conflict	0.76	0.24	1.31	4.20	1.8	0.72537	5
Thailand	Conflict	0.16	0.84	6.26	1.19	1.3	0.53664	6
Cambodia	Conflict	0.85	0.15	1.18	6.53	1.3	0.51893	7
Malaysia	Conflict	0.12	0.88	8.12	1.14	1.1	0.43199	8
Samoa	No Conflict	0.09	0.91	10.89	1.10	0.8	0.33358	9
Tonga	No Conflict	0.95	0.05	1.06	19.02	0.5	0.19927	10
Brunei Darussalam	Conflict	0.95	0.05	1.05	19.81	0.5	0.19174	11
Mongolia	No Conflict	0.05	0.95	20.75	1.05	0.5	0.18345	12
Vanuatu	No Conflict	0.95	0.05	1.05	22.10	0.4	0.17279	13
Solomon Islands	No Conflict	0.97	0.03	1.03	34.74	0.3	0.11182	14
Myanmar	Conflict	0.03	0.97	39.84	1.03	0.2	0.09789	15
Bhutan	No Conflict	0.98	0.02	1.02	51.06	0.2	0.07681	16
Fiji	No Conflict	0.01	0.99	91.83	1.01	0.1	0.04308	17
Timor-Leste	No Conflict	0.01	0.99	139.76	1.01	0.1	0.02841	18
Bangladesh	Conflict	0.01	0.99	198.29	1.01	0.1	0.02007	19
China	Conflict	0.00	1.00	385.11	1.00	0.0	0.01036	20
Singapore	No Conflict	1.00	0.00	1.00	630.98	0.0	0.00633	21
India	Conflict	0.00	1.00	1096.45	1.00	0.0	0.00364	22
Papua New Guinea	Conflict	0.00	1.00	2468.48	1.00	0.0	0.00162	23
Sri Lanka	Conflict	0.00	1.00	8324.62	1.00	0.0	0.00048	24
Philippines	Conflict	0.00	1.00	74588.47	1.00	0.0	0.00005	25
Micronesia, Federated States of	No Conflict	0.00	1.00	4.88E+05	1.00	0.0	0.00001	26
Indonesia	Conflict	0.00	1.00	8.14E+05	1.00	0.0	0.00000	27
Nepal	No Conflict	0.00	1.00	1.26E+06	1.00	0.0	0.00000	28

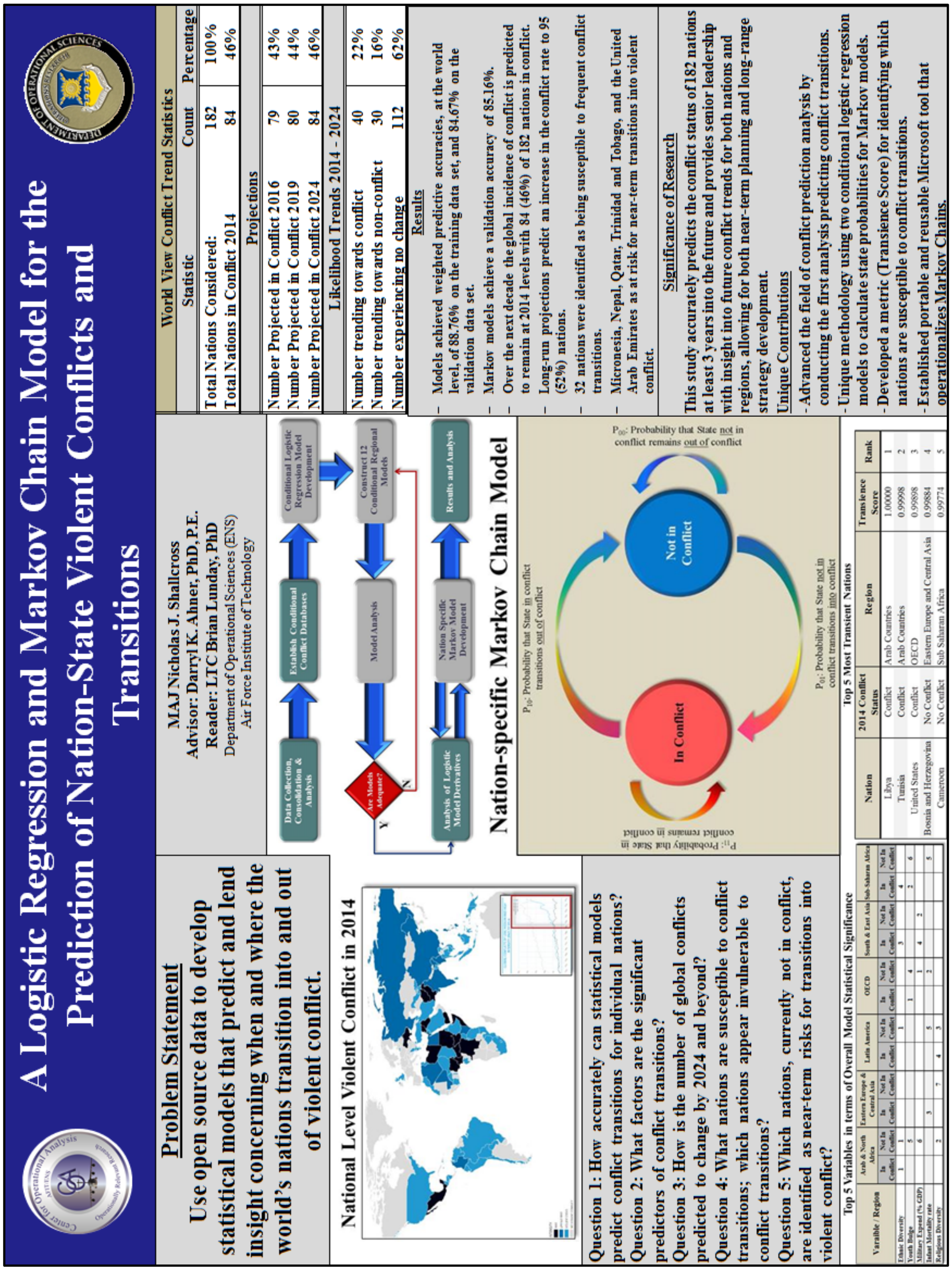
Table 63: Sub-Saharan Africa Transience Scores

Nation	2014 Conflict Status	Probability Not in Conflict	Probability in Conflict	Mi	Mj	Expected Number Recurrences per 10 Years	Transience Score	Rank
Cameroon	No Conflict	0.48	0.52	2.10	1.91	2.5	0.99774	1
Botswana	No Conflict	0.53	0.47	1.89	2.13	2.5	0.99637	2
Malawi	No Conflict	0.54	0.46	1.84	2.19	2.5	0.99228	3
Benin	No Conflict	0.46	0.54	2.20	1.83	2.5	0.99191	4
Cote d'Ivoire	Conflict	0.36	0.64	2.76	1.57	2.3	0.92453	5
Lesotho	Conflict	0.64	0.36	1.56	2.80	2.3	0.91823	6
Mozambique	Conflict	0.35	0.65	2.87	1.54	2.3	0.90836	7
Burkina Faso	Conflict	0.32	0.68	3.16	1.46	2.2	0.86516	8
Eritrea	Conflict	0.70	0.30	1.43	3.34	2.1	0.83968	9
Niger	Conflict	0.30	0.70	3.36	1.42	2.1	0.83652	10
Zimbabwe	Conflict	0.72	0.28	1.38	3.63	2.0	0.79892	11
Zambia	No Conflict	0.74	0.26	1.34	3.90	1.9	0.76270	12
Rwanda	Conflict	0.75	0.25	1.34	3.95	1.9	0.75667	13
South Africa	Conflict	0.79	0.21	1.27	4.66	1.7	0.67403	14
Nigeria	Conflict	0.20	0.80	4.93	1.25	1.6	0.64706	15
Guinea-Bissau	No Conflict	0.80	0.20	1.25	5.05	1.6	0.63529	16
Angola	No Conflict	0.19	0.81	5.16	1.24	1.6	0.62519	17
Uganda	Conflict	0.19	0.81	5.35	1.23	1.5	0.60809	18
Mali	Conflict	0.18	0.82	5.43	1.23	1.5	0.60132	19
Mauritania	No Conflict	0.86	0.14	1.16	7.11	1.2	0.48355	20
Comoros	No Conflict	0.86	0.14	1.16	7.12	1.2	0.48282	21
Ghana	No Conflict	0.86	0.14	1.16	7.30	1.2	0.47308	22
Guinea	No Conflict	0.87	0.13	1.14	7.93	1.1	0.44059	23
Sierra Leone	No Conflict	0.12	0.88	8.40	1.14	1.0	0.41962	24
Gabon	Conflict	0.89	0.11	1.12	9.04	1.0	0.39351	25

Table 63 Continued

Nation	2014 Conflict Status	Probability Not in Conflict	Probability in Conflict	Mi	Mj	Expected Number Recurrences per 10 Years	Transience Score	Rank
Equatorial Guinea	No Conflict	0.89	0.11	1.12	9.11	1.0	0.39074	26
Djibouti	Conflict	0.89	0.11	1.12	9.32	1.0	0.38306	27
Sao Tome and Principe	No Conflict	0.91	0.09	1.10	11.26	0.8	0.32367	28
Swaziland	No Conflict	0.91	0.09	1.09	11.73	0.8	0.31195	29
Tanzania	Conflict	0.93	0.07	1.07	15.19	0.6	0.24595	30
Senegal	No Conflict	0.94	0.06	1.06	17.68	0.5	0.21347	31
Namibia	No Conflict	0.94	0.06	1.06	17.74	0.5	0.21280	32
Gambia	No Conflict	0.95	0.05	1.06	18.57	0.5	0.20378	33
Ethiopia	Conflict	0.05	0.95	20.88	1.05	0.5	0.18242	34
Kenya	Conflict	0.05	0.95	21.41	1.05	0.4	0.17811	35
Liberia	No Conflict	0.96	0.04	1.04	26.04	0.4	0.14773	36
Madagascar	No Conflict	0.97	0.03	1.03	38.01	0.3	0.10247	37
Chad	Conflict	0.02	0.98	41.31	1.02	0.2	0.09448	38
Burundi	Conflict	0.98	0.02	1.02	50.51	0.2	0.07763	39
Togo	Conflict	0.98	0.02	1.02	62.15	0.2	0.06332	40
Cabo Verde	No Conflict	0.99	0.01	1.01	172.58	0.1	0.02304	41
Congo, Republic of the	No Conflict	0.99	0.01	1.01	184.18	0.1	0.02160	42
Mauritius	No Conflict	1.00	0.00	1.00	206.70	0.0	0.01926	43
Congo, Democratic Republic of the	Conflict	0.00	1.00	461.03	1.00	0.0	0.00866	44
Sudan	Conflict	0.00	1.00	1590.39	1.00	0.0	0.00251	45
Seychelles	No Conflict	1.00	0.00	1.00	7362.49	0.0	0.00054	46
Somalia	Conflict	0.00	1.00	214348.28	1.00	0.0	0.00002	47
South Sudan	Conflict	0.00	1.00	2.65E+10	1.00	0.0	0.00000	48
Central African Republic	Conflict	0.00	1.00	4.42E+11	1.00	0.0	0.00000	49

Appendix F: Story Board



Bibliography

Ahner, D. K., Boekestein, B. C., & Deckro, R. F. (2015). *A Predictive Model of World Conflict using Open Source Data*. Wright-Patterson AFB, Ohio: Air Force Institute of Technology.

Allison, P. (2013, March 5). Why I Don't Trust the Hosmer-Lemeshow Test for Logistic Regression. Retrieved December 15, 2015, from http://statisticalhorizons.com/hosmer-lemeshow

Black, N. (2010). Change We Can Fight Over: The Relationship between Arable Land Supply and Substate Conflict. *Strategic Insights, 9*((1) Spring/Summer).

Boekestein, B. C. (2015). *A Predictive Logistic Regression Model of World Conflict Using Open Source Data*. Wright-Patterson AFB, Ohio: Air Force Institute of Technology MS Thesis.

Box, G. E. (1979). *Robustness in the Strategy of Scientific Model Building.* New York: Academic Press.

Buhag, H. (2005). *The Geography of Armed Conflict*. Norway: Norwegian University of Science and Technology, Doctoral Thesis.

Center for Systemic Peace. (2015). *Integrated Network for Societal Conflict Research Data Page*. Retrieved November 11, 2015, from Center for Systemic Peace: http://www.systemicpeace.org/inscrdata.html

Central Intelligence Agency. (2015). *The World Fact Book*. Retrieved September 20, 2015, from https://www.cia.gov/library/publications/the-world-factbook/

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research, 16*, 321-357.

Collier, P., & Hoeffler, A. (2002). On the Incidence of Civil War in Africa. *Journal of Conflict Resolution, 46*(1), 13-28.

Cooper, B., & Lipsitch, M. (2004). The analysis of hospital infection data using hidden Markov models. *Biostatistics, 5*(2), 223-227.

Department of Defense. (2014). *Joint Publication 1-02; Department of Defense Dictionary of Military and Associated Terms.*

Fearon, J. D., & Laitin, D. D. (2001). *Ethnicity, Insurgency, and Civil War*. Stanford, CA: Stanford University.

Food and Agriculture Organization of the United Nations. (2015). *Databases*. Retrieved November 10, 2015, from Food and Agriculture Organization of the United Nations: http://www.fao.org/statistics/databases/en/

Freedom House. (2015). *Methodology*. Retrieved November 10, 2015, from Freedom House: http://www.freedomhouse.org/report-types/freedom-world

Goldstone, J. A. (2002). Population and Security: How Demographic Change Can Lead to Violent Conflict. *Journal of International Affairs, 56*(1), 3 - 22.

Goldstone, J. A., Bates, R. H., Gurr, T. R., Lustik, M., Marshall, M. G., Ulfelder, J., & Woodward, M. (2005). *A Global Forecasting Model of Political Instability.* Political Instability Task Force.

Hegre, H., Karlsen, J., Nygard, H. M., Strand, H., & Urdal, H. (2011). *Predicting Armed Conflict, 2010-2050*. Oslo, Norway: University of Oslo: Center for the Study of Civil War, PRIO.

Heidelberg Institute for International Conflict Research. (2014). *Conflcit Barometer 2014.* Heidelberg: Heidelberg Institute for International Conflict Research.

Hinrichs, C., & Boiler, C. (2010). JMP Essentials: An Illustrated Step-by-Step Guide for New Users. Cary, North Carolina: SAS Institute Inc.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression, Third Edition.* Hoboken: John Wiley & Sons, Inc.

Johnson, R. A. (2014). Predicting Future War. *The U.S. Army War College Parameters, 44*(Spring 2014), 65-76.

Kilcullen, D. (2013). *Out of the Mountains: The Coming Age of the Urban Guerilla.* New York: Oxford University Press.

Kisielewski, M., Rosa, J. C., & Asher, J. (2010). *Statistical Approaches to Developing Indicators of Armed Violence*. Geneva: Technical paper prepared by StratAid for Small Arms survey.

Lunardon, N., Menardi, G., & Torelli, N. (2014). ROSE: A Package for Binary Imbalanced Learning. *The R Journal, 6*(1), 79-89.

Marshall, M. G., Gurr, T. R., & Jaggers, K. (2014). *Polity IV Project: Political Regime Charactersitics and Transitions, 1800-2013.* Center for Systemic Peace.

Martinez-Beneito, M. A., Conesa, D., Lopez-Quilez, A., & Lopez-Maside, A. (2007). *Bayesian Markov switching models for the early detection of influenza pandemics*. Valencia, Spain: University of Valencia.

MathWorks. (2015). *SMOTE (Synthetic Minority Over-Sampling Technique)*. Retrieved October 1, 2015, from MATLAB Central: http://www.mathworks.com/matlabcentral/fileexchange/38830-smote--synthetic-minority-over-sampling-technique-

McCarthy, C. (1992). *Blood Meridian.* New York: Vintage.

McNamara, R. S. (1996). *In Retrospect: the Tragedy and Lessons of Vietnam.* New York: Vintage Books.

Montgomery, D. C., Peck, E. A., & Vinning, G. G. (2012). *Introduction to Linear Regression Analysis, Fifth Edition.* Hoboken, New Jersey: John Wiley & Sons, Inc.

Qiao, L., & Wang, X. (1999). *Unrestricted Warfare.* (FBIS, Ed.) Beijing: PLA Literature and Arts Publishing House.

Reed, J. (2013, March). *Forcast and Analysis of Complex Threats IV (FACT IV) Project Overview*. Fort Belvoir, Virginia: Center for Army Analysis.

Ross, S. M. (2014). *Introduction to Probability Models, 11th Edition.* Oxford: Academic Press.

SAS Institute. (2015). *Multivariate platform options.* Retrieved November 17, 2015, from Statistical Discovery from SAS: http://www.jmp.com/support/help/Multivariate_Platform_Options.shtml#106307

Scarborough, R. (2015, December 14). Gen. Dunford: U.S. commands, war plans outdated. *Washington Times*. Retrieved January 21, 2015, from http://www.washingtontimes.com/news/2015/dec/14/joseph-dunford-to-update-us-militarys-combat-comma/?page=all

Shearer, R., & Marvin, B. (2010). Recognizing Patterns of Nation-State Instability that Lead to Conflict. *Military Operations Research, 15*(3), 17-29.

Spears, W. M. (2001). *Modeling Viral Epidemiology in Connected Networks*. Washington, D.C.: Naval Research Laboratory.

Taleb, N. N. (2010). *The Black Swan: the Impact of the Highly Improbable.* (2nd ed.). New York: Random House Trade Paperbacks.

Tzu, S. (2007). *The Art of War*. Minneapolis: Filiquarian.

United Nations. (2015). *Member States*. Retrieved Novermber 9, 2015, from United Nations: http://www.un.org/en/members

Uppsala University. (2015). *Conflict Encyclopedia*. Retrieved May 19, 2015, from Uppsala Conflict Data Program: http://www.ucdp.uu.se/gpdatabase/search.php

Urdal, H. (2002). *The Devilin the Demographics: How Neo-Malthusian Population Pressure and Youth Bulges Influence the Risk of Domestic Armed Conflict*. Oslo, Norway: University of Oslo, International Peace Research Institute (PRIO) Thesis.

Ward, M. D., Greenhill, B. D., & Bakke, K. M. (2010). The perils of policy by p-value: Predicting civil conflicts. *Journal of peace Research, 47*(4), 364-375.

World Bank. (2015). *Data*. Retrieved November 11, 2015, from http://data.worldbank.org/

REPORT DOCUMENTATION PAGE			*Form Approved* *OMB No. 074-0188*
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to an penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**			
1. REPORT DATE *(DD-MM-YYYY)* 24-03-2016	**2. REPORT TYPE** Master's Thesis		**3. DATES COVERED** *(From – To)* October 2014 – March 2016
TITLE AND SUBTITLE A Logistic Regression and Markov Chain Model for the Prediction of Nation-state Violent Conflicts and Transitions			**5a. CONTRACT NUMBER**
			5b. GRANT NUMBER
			5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Shallcross, Nicholas J., MAJ, USA			**5d. PROJECT NUMBER**
			5e. TASK NUMBER
			5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way, Building 640 WPAFB OH 45433-8865			**8. PERFORMING ORGANIZATION REPORT NUMBER** AFIT-ENS-MS-16-M-128
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) CENTER FOR ARMY ANALYSIS 6001 GOETHALS ROAD, Fort Belvoir, VA 22060 Phone: 703-806-5505 Email: usarmy.belvoir.hqda-dcs-g-8.mail.caa-svc-cmdgrp@mail.mil ATTN: Dave Reynolds			**10. SPONSOR/MONITOR'S ACRONYM(S)** CAA
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **DISTRUBTION STATEMENT A.** APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.			
13. SUPPLEMENTARY NOTES This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.			
14. ABSTRACT Using open source data, this research formulates and constructs a suite of statistical models that predict future transitions into and out of violent conflict and forecasts the regional and global incidences of violent conflict over a ten-year time horizon. A total of thirty predictor variables are tested and evaluated for inclusion in twelve conditional logistic regression models, which calculate the probability that a nation will transition from its current conflict state, either "In Conflict" or "Not in Conflict", to a new state in the following year. These probabilities are then used to construct a series of nation-specific Markov chain models that forecast violent conflict, as well as yield insights into regional conflict trends out to year 2024 and beyond. The logistic regression models proposed in this study achieve training dataset accuracies of 88.76%, and validation dataset accuracies of 84.67%. Additionally, the Markov models achieve three year forecast accuracies of 85.16% during model validation. This study predicts that global violent conflict rates remain constant through year 2024, but are projected to increase beyond that timeframe with 95 of the 182 considered nations projected to be in a state of violent conflict from the current 84 nations in conflict.			
15. SUBJECT TERMS Conflict Transitions, Logistic Regression, Markov Models			

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Dr. D. K. Ahner, AFIT/ENS
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)*
U	U	U	UU	197	(937) 785-3636, x4708 Darryl.Ahner@afit.edu

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18