A CONDITIONAL LOGISTIC REGRESSION PREDICTIVE MODEL OF WORLD CONFLICT CONSIDERING NEIGHBORING CONFLICT AND ENVIRONMENTAL SECURITY

THESIS

Benjamin D. Leiby, Captain, USAF

AFIT-ENS-MS-17-M-140

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-MS-17-M-140

A CONDITIONAL LOGISTIC REGRESSION PREDICTIVE MODEL OF WORLD CONFLICT CONSIDERING NEIGHBORING CONFLICT AND ENVIRONMENTAL SECURITY

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

Benjamin D. Leiby, BS, MBA

Captain, USAF

March 2017

AFIT-ENS-MS-17-M-140

A CONDITIONAL LOGISTIC REGRESSION PREDICTIVE MODEL OF WORLD CONFLICT CONSIDERING NEIGHBORING CONFLICT AND ENVIRONMENTAL SECURITY

THESIS

Benjamin D. Leiby, BS, MBA

Captain, USAF

Committee Membership:

Dr. Darryl K. Ahner
Chair

Dr. Raymond Hill
Member

AFIT-ENS-MS-17-M-140

Abstract

Forecasts of conflict are of utmost importance for assisting combatant commanders in developing strategic and operational campaign and country plans that consider the dynamic changes that evolve within their area of responsibility. This research formulates and constructs five suites of statistical models to better understand the collinearity of environmental factors affecting conflict and compares the classification accuracy between forcing these factors into logistic regression models. A total of thirty-nine predictor variables are tested and evaluated for inclusion in a six region, two conflict state combination suite. The five suites of twelve models calculate the probability of whether a country will transition to either an “In Conflict” or “Not In Conflict” state for the following year. Handpicking the best models proposed in this study from each suite achieves modeling classification accuracies of 92.0% with 82.6% prediction accuracies. Through exploring new variables and selection methods, the models demonstrate that leveraging the collinearity of environmental factors help provide strategic insight in developing Department of Defense Theater Campaign Plans to effect the stability of national security.

KEYWORDS: Conflict Transitions, Environmental Factors, Logistic Regression

To my wife--your unfailing love, tireless devotion, and unwavering support to me during this study and throughout my career makes everything better.

Acknowledgements

I would like to thank my research advisor, Dr. Darryl Ahner, for the guidance he provided during this process. His guidance in teaching me logistic regression and passion for the topic inspired me to keep searching for better results.

I would also like to thank Dr. Raymond Hill for accepting to be my reader. His insights and suggestions have made me a better writer.

Finally, I would like to thank Dr. Thomas Dube for this mentorship of instilling in me the idea that all results are worth documenting even if they don't appear significant at the time.

Benjamin D. Leiby

Table of Contents

List of Equations

List of Figures

Page

List of Tables

Page

A CONDITIONAL LOGISTIC REGRESSION PREDICTIVE MODEL OF WORLD CONFLICT CONSIDERING NEIGHBORING CONFLICT AND ENVIRONMENTAL SECURITY

I. Introduction

1.1 General Issue

Davenport and Harris wrote in their book, "Competing on Analytics", how the challenge to analytics is not simply to identify various applications of analytics, but to find some clear strategic or competitive edge (Davenport & Harris, 2007). This research focuses on finding factors that simplify model inputs or increase accuracy in predicting nations in conflict through exploring new variables and new variable selection methods. In predictive analysis, correlation is leveraged to associate one behavior with evidence from another behavior. A pitfall, however, of solely relying on correlation calculations is that correlation does not provide justification of causality. Understanding the issues pertaining to the problem being modeled, along with statistical calculations is paramount in ensuring model variables are both relevant and adequate. Additionally, Ward and Bakke illustrated in their civil conflict research that sole reliance on statistical significance for variable selection may limit improvements in predictive accuracy (Ward & Bakke, 2010). Expert knowledge may provide insight into unmasking significant factors due to collinearity that ultimately improve predictive results; namely, leveraging environmental factors in predicting nation conflict transitions. The conjecture, considered here, is that the availability of water per capita combined with the security of country borders is relevant to constructing the best predictive models.

The importance of predicting conflict states aligns directly with the Department of Defense's Theater Campaign Planning strategic goals. Combatant commanders are charged with assessing changes to the strategic or operational environment that may lend insight into disruptive effects on campaign and country plans (Department of Defense, 2012). Their assessments, at a minimum, should include the emergence of both new and significant threats and opportunities, along with changes in the balance of military power in the region and security relationships between regional neighbors (Department of Defense, 2012). The ability to model country conflict states and quantitatively identify significant factors to those states would tremendously help commanders in achieving robust assessments. Through the models, they gain insight into whether conflict will continue or not, if their resources to affect conflict transition are appropriately allocated or not, and how sensitive the various influences are to change. Our goal is to provide additional insight into developing mature models to assist these commanders.

1.2 Problem Statement

Use open source data to construct models that provide insight into which influences most effect an accurate prediction of classifying when a nation will transition into or out of conflict.

1.3 Research Objective and Focus

The objective of this study is to identify the effects that environmental factors, specifically water data and neighboring country conflict status data, have on prediction accuracy of models that predict nation conflicts. This study uses conditional logistic regression models to predict transitions of nation conflicts with open source data from the years 2004 to 2014.

1.4 Study Assumptions and Limitations

This study utilizes the same database featured in the Shallcross study (Shallcross, 2016) with the addition of two independent variables: the number of bordering counties in conflict around the observed country as a percentage and a similarly derived binary variable that signifies if at least one neighboring country is in conflict or not. This brings the total number of independent variables up to thirty-nine (see Appendix F) and one binary dependent variable (transition to conflict). The study assumes that the variables identified in each model retain a high level of significance for the years assessed (labeled as training data sets) and continue to maintain a high level of significance for subsequent years (labeled as validation data sets). Notion such as training year sets is consistent with Shallcross for describing modeling generation. Likewise, validation year sets describe prediction results. Training and validation usually reference topics related to data, while modeling and prediction reference modeling attributes. For this study, the terms can be used interchangeably as training year data is used to generate models and validation year data is used to generate predictions. Finally, this study also assumes that the data sets in each geographic region share some common elements that reasonably warrant being modeled together.

II. Literature Review

2.1 Previous Research

This research continues to investigate the advantages of logistic regression to predict conflict in nation states. Starting in 1994, the Central Intelligence Agency investigated several methods to explain political instability and state failures to include logistic regression, neural networks, and Markov models (Shearer & Marvin, 2010). However, the logistic regression model has been the method that produces the most accurate predictions with over 80% accuracy (Goldstone, et al., 2005). Two main works prelude this research, which provide the foundation for this study: the Boekestein study (Boekestein, 2015) and the Shallcross study (Shallcross, 2016). Both studies along with this one consider 182 nations for conflict prediction classification. Although it was desired to develop one whole-world model for prediction purposes, the nations were grouped into six regions based upon insight from renowned statistician Hans Rosling (Boekestein, 2015).

The Boekestein study focused on using the Heidelberg Institute for International Conflict Research (HIIK) “Levels of Violence” as the dependent variable in logistic regression (Boekestein, 2015). Although the HIIK proposes six conflict intensity levels, these levels are mapped into one of two groups: “not violent conflict” and “violent conflict”. The first three (0-no conflict, 1-dispute, 2-non-violent conflict) are categorized as “not violent conflict”, while the later three (3-violent crisis, 4-limited war, 5-war) are categorized as “violent conflict”. Boekestein then developed his models choosing independent variables through a combination of variance inflation factor screening and correlation testing. Ultimately, he achieved regional predictive accuracies of greater than 78% with cut off parameters at 0.50 and greater than 80% accuracy with cut off parameters adjusted to 0.28, which are comparable to the CIA study (Boekestein, 2015).

The Shallcross study continued the Boekestein research and augmented it with Markov modeling. The two-state Markov modeling provided the probabilities of transition to conflict for the following year, given the current conflict status of the nation in question (Shallcross, 2016). Instead of using the HIIK-map dependent variable that Boekestein used, Shallcross looked into the transition of the HIIK variable from the previous year as the dependent variable. Additionally, Shallcross doubled the number of models by segregating the dependent variables into two preceding conflict camps: "prior year not violent conflict" and "prior year violent conflict". With a cut point parameter set at 0.50, Shallcross was able to achieve greater than 80% classification accuracy in all twelve training models averaging 88.76% and nine of the validation data sets averaging 84.67% (Shallcross, 2016).

2.2 Environmental Variables

Homer-Dixon's study on acute conflict concluded "environmental change may contribute to conflicts as diverse as war, terrorism, or diplomatic and trade disputes" (Homer-Dixon, 1991). He studied various environmental effects that could cause social tendencies that could lead to war. A story board could be developed to illustrate these tendencies, such as droughts leading to migration, and migration to the overuse and polluting of neighboring water supplies, which ultimately end in ethnic clashes (Homer-Dixon, 1991). The United Nation's study on conflict in Sudan over a decade later acknowledged that even though environmental factors may not be the sole cause of conflict in the area, environmental factors could be a quantifiable contributing source (United Nations Environment Programme, 2007). The study noted a few common conclusions which include, namely, the increase in population growth creates a greater demand for natural resources. Population growth may correlate to predicting conflict; however, a theory of causation could conclude that a constraint on fresh water supply per capita could be the true correlation driver for a conflict transition. Furthermore, theorizing that migration of refugees from neighboring countries in conflict may be the true reason for the decrease in availability of fresh water per capita inciting conflict in the country. The refugee theory gains further credibility as the Sudan study noted that the second most significant effect of conflict found was people fleeing conflict zones (United Nations Environment Programme, 2007). Direct impacts of conflict included the targeting of natural resources for destruction, while indirect impacts included environmental impacts due to population displacement and looting of natural resources (United Nations Environment Programme, 2007). Both studies concluded that while not primary factors in predicting conflict, both the availability of fresh water to the populace and the state of conflict in neighboring nations may be secondary factors with untapped importance.

2.3 Logistic Regression Theory

Logistic regression methods are a data analysis approach to describe the relationship between a dichotomous dependent response and one or more independent variables also known as covariates. The response is bound to only two possibilities such as “In Conflict” or “Not In Conflict”. One state is numerically quantified as 0, while the other is numerically quantified as 1. The relationship between the covariates and the binary response is quantified as the conditional mean, or the expected value of the response given the average of the covariates. This relationship is similar to linear regression with the exception that logistic regression bounds the conditional mean between 0 and 1 graphically forming an "S" shaped curve. The distribution that is then fit to the conditional mean is the logistic distribution, which coincides with the regression name. Our goal is to build the smallest possible model of covariates to ascertain predictive truth about the dichotomous response with little to no misclassification errors.

When building a model, it is necessary to test the overall significance of the fitted coefficients. This requires an iterative approach in order to maximize a log-likelihood function. We start with a generalized linear model similar to Equation 1 and solve for the beta coefficients transforming it into the logit, g($\mathbf{x}$). The betas are the selected covariates' scalar multiplier and the x's are the data points associated with the covariates selected, from the first selected to the nth selected.

$$g(\boldsymbol{x}) = \beta_0 + \beta_1 \boldsymbol{x} + \cdots + \beta_n \boldsymbol{x}$$

Equation 1: Logit Transformation

The logit transformation is then expressed as the conditional mean of the dependent variable, π($\mathbf{x}$), as depicted in Equation 2.

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

Equation 2: Conditional Mean of the Dependent Variable

Once in the conditional mean form, the dependent variable, coded as either a 1 or 0, is placed in Equation 3 to assist in computing its contribution to the likelihood function. Dependent variables coded as 1 are denoted as a conditional probability of

$\pi(\mathbf{x})$, while dependent variables coded as 0 are denoted as conditional probability of 1-$\pi(\mathbf{x})$.

$$\pi(\mathbf{x}|\mathrm{DV}) = \pi(x_i)^{y_i}[1 - \pi(x_i)]^{1-y_i}$$

Equation 3: Coding of Dependent Variable

where:

i = each row of instantiated data for 1 to m independent variables

y = dichotomous dependent response

The product of these terms then provides the form for the likelihood function, $l(\beta)$, as denoted in Equation 4.

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

Equation 4: Likelihood Function

Once the likelihood function is created, an iterative optimization approach is used to find the maximum log-likelihood estimate by zeroing in on the optimal beta coefficients for each selected covariate. There are many off-the-shelf software programs that use this iterative approach including JMP, SAS, and Excel's solver. The log-likelihood of the model can then be tested against the log-likelihood of a saturated model, or intercept-only model. This log-likelihood ratio is multiplied by minus two to obtain a quantity whose distribution is known for hypothesis testing, the chi-squared distribution (Hosmer, Lemeshow, & Sturdivant, 2013). The quantity is known as the deviance, D, as shown in Equation 5.

$$D = -2\ln\frac{likelihood\ of\ the\ fitted\ model}{likelihood\ of\ the\ saturated\ model}$$

Equation 5: Deviance

When the dependent outcome is either 0 or 1, the likelihood of the saturated model is identically equal to 1 so that the deviance state is just the minus two log-likelihood of the fitted model. The context for testing the significance of a fitted model can be thought of in the same way that the residual sum-of-squares is used in linear regression (Hosmer, Lemeshow, & Sturdivant, 2013). To test for the significance of the model, the deviance is compared with and without the variables, known as the G statistic, as shown in Equation 6.

$$G = D(model\ without\ variable) - D(model\ with\ the\ variable)$$

Equation 6: G Statistic

The G statistic plays the same role in logistic regression that the numerator of the partial F-test does in linear regression (Hosmer, Lemeshow, & Sturdivant, 2013). Within the test for significance, the null hypothesis (H_0) questions whether the saturated model provides greater explanatory power than the fitted model. A p-value utilizing the chi-squared distribution and the appropriate number of degrees of freedom corresponding to the number of covariates used relates the p-value with H_0. A low p-value (ie. p-value < 0.05) indicates that the H_0 can be rejected because the model developed provides greater explanatory power with the fitted variables. However, there may be reasons to consider p-values as high as 0.20 when considering individual variables, as explained later in this study.

There are multiple methods for deciding which variables to bring into the model. Boekestein uses the correlation method, whereas, Shallcross uses the purposeful selection method. The method used to fit the model is unimportant as long as the model results in relevant variables producing stable estimates. The traditional approach seeks to build the most parsimonious model that accurately reflects the true outcome experience of the data (Hosmer, Lemeshow, & Sturdivant, 2013).

The stepwise selection method is a statistically driven approach that iteratively brings in a new variable with the most statistically significant G statistic with the option to remove a variable, due to multicollinearity, should its statistical significance decrease below a specified threshold. Each variable is individually tested in the model to calculate its own G statistic for that iteration. The G statistic is hypothesis tested not against the saturated model, but the model with only one additional variable; therefore, the degrees of freedom are set to 2. The variable with the highest G statistic is then considered for inclusion into the model and another iteration of tests is observed. Throughout the iterations, it is possible for a model variable to decrease in its significance for the model. A model variable with a G statistic that becomes too small is then dropped from the model. It is important to set the threshold to drop a variable lower than the value to include a variable to preclude adding and dropping the same variable multiple times.

Once a model is built with statistically relevant variables, the fit of the model is assessed to assist in keeping the most parsimonious model. The G statistic is the preliminary test; however, two other tests can be conducted: discrimination and calibration. Dreiseitl and Ohno-Machado describe discrimination as the goodness of measuring the classification of data, while calibration is the accuracy of prediction probability (Dreiseitl & Ohno-Machado, 2002). Discrimination commonly encompasses the measures of classification accuracy and the area under the receiver operating characteristic (ROC) curve. Calibration measures how close the predictions of a given model are to the real underlying probability, such as what the Hosmer-Lemeshow $\hat{C}$ statistic attempts to quantify.

The accuracy of the model can be observed though the use of classification tables. The classification table should not be used to measure the sole fit of the model as they are highly dependent on the distribution of estimated probability and the assumed cut point (Hosmer, Lemeshow, & Sturdivant, 2013). However, they are useful in providing additional insight into the analysis where classification is the stated goal, such as this study. Classification tables map the estimated logistic probability of a data point as either a 1 or a 0 based upon an assumed cut point, commonly set to 0.5. These data points are then tallied in a cross-matrixed table between the classified value and observed value. Data points correctly classified as a 1 with an observed value of 1 are said to be true positives, while data points classified as a 0 with an observed value of 0 are said to be true negatives. The ratio of true positives to false negatives measure the sensitivity of the model and the ratio of true negatives to false positives measure the specificity of the model. The accuracy of the model is then calculated by the sum of true positives and true negative divided by the total number of data points assessed.

Similar to classification tables, the ROC curve provides insight into the discrimination of a model. The ROC curve measures the sensitivity and specificity of the model, but unlike the classification table, which depend on a single cut point, the ROC curve measures across an entire range of cut points to provide a better and more complete description of classification accuracy (Hosmer, Lemeshow, & Sturdivant, 2013). It plots the probability of detecting true positives (sensitivity) against the detection of false negatives (1-specificity). The area under the curve is calculated providing a measure of the model's ability to discriminate between its predictive accuracy. In general, Hosmer, Lemeshow and Sturdivant suggest the use of Table 1 as a guideline for assessing the ROC curve's "goodness" of discrimination. The primary difference between the ROC curve and the classification tables is that the ROC curve quantifies accuracy across a range of cut points whereas the classification table provides a point estimate.

Table 1: Area Under the ROC Curve Assessment

Area Under the ROC Curve assessment	
$ROC = 0.5$	No discrimination, "coin toss".
$0.5 < ROC < 0.7$	Poor discrimination
$0.7 \leq ROC < 0.8$	Acceptable discrimination
$0.8 \leq ROC < 0.9$	Excellent discrimination
$0.9 \leq ROC$	Outstanding discrimination

(Hosmer, Lemeshow, & Sturdivant, 2013)

Calibration of a model can be assessed by the Hosmer-Lemeshow $\hat{C}$ statistic. Although Dreiseitl and Ohno-Machado do not mention this statistic by name, they mention a description similar to the $\hat{C}$ statistic as a more refined way to measuring calibration rather than taking the difference between the two different estimates of probability (Dreiseitl & Ohno-Machado, 2002). Hosmer, Lemeshow and Sturdivant suggest two different methods for obtaining this $\hat{C}$ statistic: percentiles of estimated probabilities and fixed values of probabilities (Hosmer, Lemeshow, & Sturdivant, 2013). The percentile (binning) method observes the number of estimated probabilities and then bins them into equal numbers according to calculated cut points, which are estimated probabilities sorted from smallest to largest. The number of rows of data can be iteratively tested to find an integer within a specified range that would leave zero remainder. The quotient would then become the number of data points within each bin. Hosmer, Lemeshow and Sturdivant use ten bins; however, iterative approaches for this research ranged from six to fifteen, with precedence on a greater number of bins should multiple binning options provide a zero remainder. Ahner and Spainhour note that the Hosmer-Lemeshow test biases toward insignificance when the number of events in the bins are small (<5 per bin) (Ahner & Spainhour, 2015). The bias, however, can be mitigated by simultaneously considering the fixed values method or ensuring a large sample size. The cut points are the estimated probabilities when the bins are filled. Each bin is assessed for the number of observed data points against the estimated number within each bin for its associated cut point. As seen in Equation 7, the observations versus the expected outcomes of all bins are then summed together for its $\hat{C}$ statistic value. The fixed values (tenths) method set the cut points at specified values; in this research, every tenth is a bin from zero to one for a total of ten bins.

Through extensive simulation, the distribution of the $\hat{C}$ statistic was well approximated by the chi-square distribution with g-2 degrees of freedom (Hosmer, Lemeshow, & Sturdivant, 2013). Therefore, the $\hat{C}$ statistic can be tested under the null hypothesis that the model is fit, where the $\hat{C}$ statistic less than the chi-square distribution could not prove inadequacy in explanatory power of the fitted model. It is noted that Hosmer, Lemeshow and Klar prefer the binning method to the tenths method as his research has shown a sense of better adherence to the chi-square distribution with g-2

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

Equation 7: Hosmer-Lemeshow Test Ĉ Statistic

where:

g = number of bins

h = (total number of estimated probabilities) / (number of bins)

k = specified bin number from 1 to g

c_k = number of patterns in k^{th} bin (y = 1 or y = 0)

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

$o_{0k} = \sum_{j=1}^{c_k} y_j$, if y = 0

$\hat{e}_{1k} = \sum_{j=1}^{c_k} h\hat{\pi}_j$, if y = 1

$\hat{e}_{0k} = \sum_{j=1}^{c_k} h\left(1 - \hat{\pi}_j\right)$, if y = 0

degrees of freedom; however, he has not dismissed the goodness that the tenths method provides because of the potential risks the binning method proposes by possibly fracturing the bins (Hosmer, Lemeshow, & Klar, 1988).

One of the main concerns about developing models is overfitting. Overfitting usually develops when too many variables are used to explain a model. Dreseitl and Ohno-Machado suggest that in logistic regression, overfitting can be avoided by restricting model complexity by using a p-value for statistical testing of 0.05 and using few or no interaction terms in the model (Dreiseitl & Ohno-Machado, 2002). However, Hosmer, Lemeshow and Sturdivant concluded that a p-value of 0.05 may be too stringent, often excluding important variables and suggest opening the inclusion range to as high as 0.20 (Hosmer, Lemeshow, & Sturdivant, 2013).

III. Methodology

3.1 Data

The data in this research is consistent with the Shallcross data in both the dependent and independent variables with the addition of two different metrics for border conflict scoring. The water data is already in the Shallcross data set; however, this research considers its significance to modeling differently. The same 182 nations (see Appendix E) were studied according to data collected from various sources to include the Heidelberg Institute for International Conflict Research (HIIK), The World Bank, Central Intelligence Agency World Fact Book, Freedom House, the Center for Systemic Peace, and the Food & Agriculture Organization of the United Nations (Shallcross, 2016).

The data for the environmental focus of this research revolves around the study of water data in the form of fresh water per capita and border conflict scores viewed as both a binary factor and a number of bordering nations factor. The binary border conflict factor looks at each nation per year and categorizes a nation's border conflict score with a "1" if at least one bordering nation was considered to be in conflict for that particular year or a "0" if no bordering nations were in conflict. The number of bordering nations factor considers the number of bordering nations in conflict for each given year and receives a percentage score based on the number of bordering nations in conflict compared to the total number of bordering nations.

The dependent variable remains consistent with the Shallcross research as transition to conflict status per given year. Each nation was given a binary code based on their associated HIIK conflict intensity level. HIIK conflict intensity levels of 0, 1, or 2 where considered being not in conflict while HIIK conflict intensity levels of 3, 4, or 5 were considered to be in conflict. These binary scores were then compared to the following year and mapped as either a 1 if they either remained or transitioned to conflict or a 0 if they either remained or transitioned to not conflict.

3.2 Logistic Regression Modeling

Developing logistic regression models, involved exploring the nature of the relationship between the dichotomous dependent variable (transition to conflict year status) and the covariates. The utilization of a conditional mean displayed a clearer picture than the use of a scatterplot to obtain a quick look of any relationship that may exist. A univariable logistic regression plot with conditional means was used to assist in verifying the assumption that any relationship is fairly linear. The plots displayed both a direct line plot and a third-polynomial trend line of the conditional mean. Nonlinear behaviors that displayed large deviations from linear were investigated for possible transformation. The only covariate that displayed non-linear behavior in some modeling groups was the population growth variable. An absolute value transformation was applied to the data and both the raw data covariate and the transformed data covariate were assessed for inclusion in every model studied.

The modeling method used in this study was the bidirectional step-wise logistic regression selection method. The model started with only the intercept, similar to forward step-wise selection, and added variables whose inclusion gave the most statistically significant fit improvement as described by the highest G statistic. Should a modeled covariate's G statistic fall below the exclusion threshold, that variable was removed from the model and never again considered for inclusion. All the modeling calculations were programmed into Microsoft Excel using visual basic for application (VBA) code to expedite the modeling verification process. The VBA program written for this study was preferred above JMP because JMP did not provide an adequate method to facilitate the desired step-wise technique in investigating environmental factors. Excel's GRG non-linear solver was used to iterate the maximum likelihood function in estimating the beta parameters; however, if the iterative approach exceeded Excel's calculation ability, JMP was substituted to obtain the beta coefficients. An error that would occur during the calculation of the likelihood function was when the combination of the coefficient estimate and product of terms exceed Excel's number limit (-2.2E-308 to 9.9E+307) to find the maximum likelihood. One mitigation technique was to rescale some of the variables in terms of thousands to limit the number of digits used to calculate products, which are annotated in the results. All final models were reassessed in JMP to verify the calculations in Excel, and the JMP ROC curve values were annotated as the final ROC curve results to overcome any round-off error in Excel.

During step-wise selection, the following criteria were used to stop entering variables into the model; all conditions needed to be met.

1. Include all covariates whose associated G statistic value exceeds the chi-squared distribution with fifteen percent probability and two degrees of freedom $\left(\chi^2_{(0.15,2)} = 3.794\right)$.
2. If further variables exist in the pool of non-model variables with a G statistic value greater than the chi-squared distribution with twenty-five percent probability, assess on their merit of fit to ensure one of the

associated $\hat{C}$ statistics continues to remain less than their associated chi-squared distribution.

3. Either the accuracy of the model or the ROC curve value has to increase by at least one whole percent.
4. During models where a specific variable is assumed important to unmask correlations, as per Hosmer, Lemeshow and Sturdivant's guidance, that variable is entered into the model first before proceeding with stepwise selection and may never leave the model even if the it's G statistic value falls below the removal threshold (Hosmer, Lemeshow, & Sturdivant, 2013).

Consistent with the Shallcross modeling, each model was developed according to data filtered on geographical region and the current year conflict state. The training and verification years used in the models were also consistent with Shallcross as seen in Table 2.

Table 2: Model Training and Validation Year Sets

Regional Models	Arab & North Africa		Eastern Europe & Central Asia		Latin America	
Year Sets	In Conflict	Not in Conflict	In Conflict	Not in Conflict	In Conflict	Not in Conflict
Training Year Set	2004-2010	2004-2009	2006-2011	2004-2010	2008-2011	2004-2010
Validation Year Set	2011-2013	2010-2013	2012-2013	2011-2013	2012-2013	2011-2013

Regional Models	OECD		South & East Asia		Sub-Saharan Africa	
Year Sets	In Conflict	Not in Conflict	In Conflict	Not in Conflict	In Conflict	Not in Conflict
Training Year Set	2004-2010	2005-2009	2004-2010	2008-2011	2004-2010	2004-2010
Validation Year Set	2011-2013	2010-2013	2011-2013	2012-2013	2011-2013	2011-2013

(Shallcross, 2016)

Sixty models were generated for this research according to the combination of geographic region, conditional conflict state parameter, and modeling focus. Modeling focus is defined by specific variable selection techniques, which separates the models into suites. The five suites are (1) Most Significant, (2) Forced Water Variable, (3) Forced Binary Border Conflict variable, (4) Forced Number of Border Conflicts variable, and the (5) Combined Effects of Forced Water variable with Number of Border Conflicts variable. The (1) Most Significant suite applied step-wise selection according to the standard G statistic analysis. The other four suites also applied step-wise selection after first selecting the forced variable: Water per Capita for the water variable; Border Conflict, Binary for the binary indicator that at least one bordering nation is in conflict variable; and Border Conflict, Number for the percentage of bordering nations in conflict variable. Unlike all other variables in the step-wise selection technique, the forced variables were not allowed to leave the model even if their significance fell below the threshold for modeling. The results of all sixty models are in Appendix A.

3.3 Step-Wise Implementation

Bidirectional step-wise logistic regression selection method requires a plethora of laborious calculations to ascertain the independent variables with the most significance. Each independent variable is independently added to the model in order to determine its G statistic to quantify its power to the model. After all variables have undergone this calculation, a determination can then be made as to which variable will be the most beneficial to the model. This procedure iterates until a stopping indicator is found. A VBA program was developed to help expedite calculating the G statistic for all thirty-nine independent variables during each step iteration. Additionally, the VBA program provided outputs of the model p-value, ROC curve value, accuracy at 0.50 cut-off value, both methods for the $\hat{C}$ statistic, and the prediction accuracy for a validation data set at

0.50 cut-off value as seen in Figure 1. The outputs allow the analyst to quickly make decisions for model development.

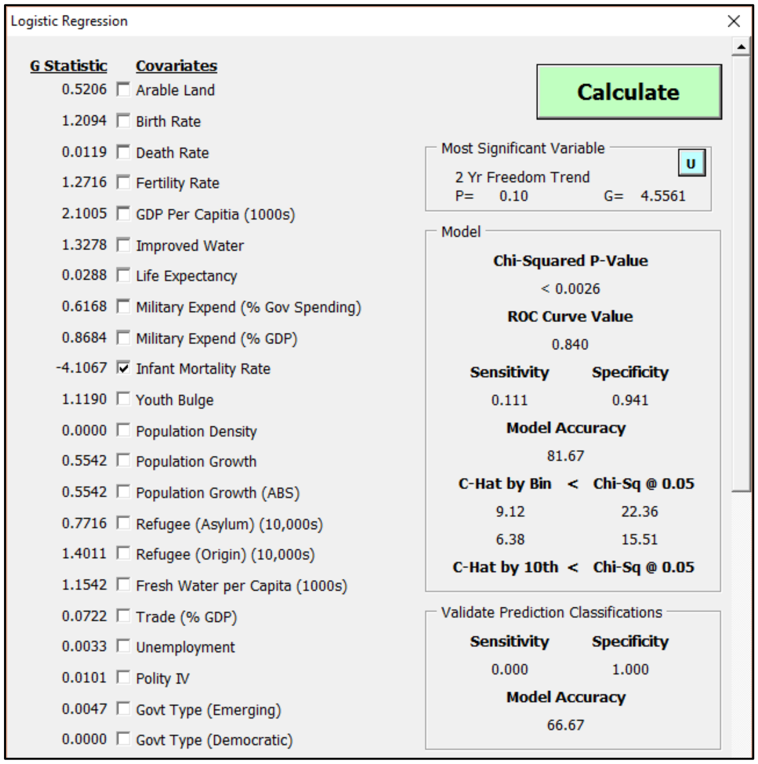


Figure 1: VBA Step-Wise Logistic Regression Program

The following walkthrough of the Organization for Economic Cooperation and Development (OECD) “Not In Conflict” data is used as an example of the step-wise technique. The first iteration of G statistics with the “intercept only” model identifies “Freedom Score” as the most beneficial variable to be included into the model with a p-value less than 0.0001 and G statistic of 18.10. This variable is added to the model first for the (1) Most Significant focus area model. However, for the (5) Combined focus area, both “Fresh Water per Capita” and “Border Conflict Score, Number” are added to the model despite their less powerful G statistic of 0.566 and 1.038 respectively. The second iteration for the (5) Combined focus area model identifies “Freedom Score” as the most beneficial variable included into the model with a p-value of 0.0013 and G statistic of 17.82. Notably, the two environmental factors both have G statistics that would warrant removal from the model under the modeling threshold rules developed; however, because these two variables are of modeling interest, they are not allowed to leave the model. The ROC curve value is 0.844 and model accuracy at 0.50 cut-off is 94.21%. The $\hat{C}$ statistic by bins is out of range with a value of 238 when its chi-squared test should be under 16.92; however, the $\hat{C}$ statistic by tenths is within range with a value of 6.76 with corresponding chi-square test cap of 15.51. A snapshot of the Excel output for the (5) Combined second iteration can be seen in Figure 2. The third iteration identifies "Military Expend (% Gov Spending)" with p-value 0.013 and G statistic 10.33. The

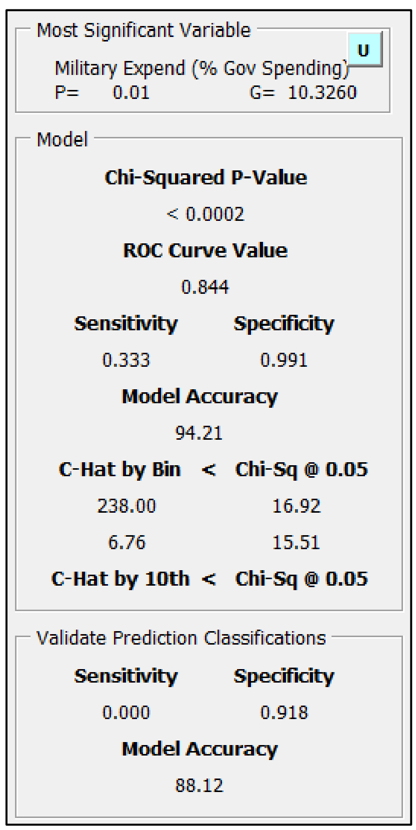


Figure 2: Second Iteration Excel Output - OECD "Not In Conflict" (5) Combined

ROC curve value of 0.935 demonstrates an increase of over one percent, although the modeling accuracy has remained that same. Both $\hat{C}$ statistics fall in range with 13.34 by binning and 14.47 by tenths. The fourth iteration identifies “Population Growth” with p-value 0.143 and G statistic 2.15. As expected, the ROC curve demonstrates an increase to 0.958 although the accuracy remains at 94.21%. The variable, however, is not added to the model and the iterations stop due to both $\hat{C}$ statistics going out of range with a value of 54.61 by binning and 23.28 by tenths. The model is double-checked in JMP and recorded with a p-value less than 0.0001 using variables Fresh Water per Capita (p-value 0.4762), Border Conflict Score, Number (p-value 0.9322), Freedom Score (p-value 0.0054), Military Expend (% Gov Spending) (p-value 0.0016). The model is then recorded with coefficients and other metrics as seen in Figure 3. This procedure is used to develop the remaining fifty-nine models.

Region: **Organization for Economic Cooperation and Development**
State: Not In Conflict
Focus: *(5) Force Water per Capita and Border Conflict, Number Variable*

Variables	Coefficients
Intercept	30.3497
Military Expend (% Gov Spending)	0.2718
Fresh Water per Capita (1000s)	0.0177
Freedom Score	-37.7722
Border Conflict Score, Number	0.1230

Model Significance, $p > \chi^2$: 0.0000

Model — Years: 2005-2009

Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	5	3
	as NC	4	109
		9	**112**

Accuracy	Sensitivity	Specificity
94.21	0.556	0.973

AUC: 0.935 N: 121

$\hat{C}$ by Bins: 13.34 $\hat{C}$ by 10ths: 14.47

$\chi^2_{(0.05,9)}$: 16.92 $\chi^2_{(0.05,8)}$: 15.51

Validation — Years: 2010-2013

Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	1	3
	as NC	3	90
		4	**93**

Accuracy	Sensitivity	Specificity
93.81	0.250	0.968

N: 97

Figure 3: Final Output - OECD “Not In Conflict” (5) Combined

IV. Analysis and Results

4.1 Model Parsimony

William of Ockham is famously noted for saying “plurality should not be posited without necessity” (Occam's Razor, 2015). This is interpreted to mean when faced with multiple choices, the simplest is likely the best. One of our goals was to generate models that were more parsimonious that our predecessor’s use of purposeful selection without sacrificing accuracy. The strategy employed observing accuracy while adding variables in the model. For example, when developing the “Not In Conflict” Latin America model, variable section stopped after including “Military Expenditure (% Government Spending)” even though the following variable, “Border Conflict”, had a p-value of 0.16 because the inclusion of “Border Conflict” failed to raise accuracy by at least one percent. Statistically, a p-value of 0.16 would indicate adequate goodness of fit, yet by simultaneously observing accuracy, it is conjectured that it does not provide any additional explanatory power. By employing this technique, increased parsimony was observed in seven of the twelve regional models compared to the Shallcross models while maintaining training accuracy as seen in Table 3. The largest difference in training accuracy is seen in “In Conflict” Latin America where the new technique classified better by 10.8%. The largest different where the new technique did not train better was only 1.6% (“Not In Conflict” Arab and North Africa). In both the “In Conflict” Arab and North Africa model and the “In Conflict” OECD model, the (1) Most Significant suite ended variable selection because the threshold of p-value was met before observing stalled accuracy. Validation accuracies varied more individually, however, the difference in weighted average was small at only 0.3% in favor of Shallcross. The difference in weighted average for the training accuracies was slightly larger at 1.6% in favor of the (1) Most Significant suite. Overall, the technique of observing accuracy while practicing step-wise selection provided positive parsimonious effects.

Table 3: Parsimony Comparison by Technique

(1) Most Significant Suite vs Shallcross

Conflict State	Region	Number of Variables		Training Accuracy		Validation Accuracy	
		Most Significant	Shallcross	Most Significant	Shallcross	Most Significant	Shallcross
In Conflict	Arab and North Africa	6	**5**	96.2%	94.2%	93.0%	74.4%
	Eastern Europe and Central Asia	6	6	95.5%	92.5%	79.3%	82.8%
	Latin America	**5**	6	91.9%	81.1%	73.3%	83.3%
	OECD	6	**4**	90.6%	88.7%	81.8%	86.4%
	South and East Asia	7	7	93.7%	87.3%	68.2%	84.1%
	Sub-Saharan Africa	**4**	12	85.7%	86.4%	86.5%	82.4%
Not In Conflict	Arab and North Africa	**6**	7	91.7%	93.3%	66.7%	60.0%
	Eastern Europe and Central Asia	**3**	11	89.4%	86.0%	83.7%	76.7%
	Latin America	**4**	9	90.2%	90.9%	85.7%	88.1%
	OECD	**4**	7	96.0%	96.0%	94.1%	95.0%
	South and East Asia	6	6	89.4%	87.9%	92.0%	88.0%
	Sub-Saharan Africa	**7**	8	85.7%	85.2%	82.2%	86.3%
	Average (Weighted Acc. Avg)	**5.33**	**7.33**	**90.3%**	**88.8%**	**84.4%**	**84.7%**

When comparing the parsimony across the five suites, there appeared to be statistical equivalency between the average number of variables as seen in Table 4. The inclusion of the "Water per Capita" variable increased the standard deviation for the average number of variables per model, however, the combined effects of "Water per Capita" and "Border Conflict, Number" in the (5) Combined suite experienced a tighter standard deviation than the (1) Most Significant suite. Not all forced variables nor the

Table 4: Parsimony Comparison by Suite

Suite	Number of Variables		Number of p > 0.25 Variables		Training Accuracy		Validation Accuracy	
	Average	Std Dev	Average	Std Dev	Average	Std Dev	Average	Std Dev
(1) Most Significant	5.3	1.30	0.0	0.00	90.3	3.58	84.5	9.10
(2) Forced Water per Capita	5.9	2.15	0.4	0.51	90.0	3.59	84.8	9.28
(3) Forced Border Conflict, Number	5.6	1.24	0.8	0.58	89.5	3.33	85.2	9.62
(4) Forced Border Conflict, Binary	5.1	1.44	0.8	0.72	89.6	4.05	82.8	9.94
(5) Combined Water & Border Conflict	6.0	1.04	1.1	0.79	89.7	4.09	82.6	7.05

intercept in the models remained under the desired p-value threshold in every instance. However, regardless of the forced variable p-value, the average training accuracy of the models remained statistically equivalent. This phenomenon, however, warranted further investigation.

Observationally, it appeared that if the “Water per Capita” variable, the “Border Conflict Score, Number” variable, or the intercept on the forced one-variable focus area models (also known as suites 2 and 3) exceeded the desired p-value threshold, then the variable would also exceed the desired p-value threshold when modeled in the (5) Combined suite models. There were four exceptions to this observation.

1. For the “In Conflict” Latin America regional models, the intercept exceeded the desired p-value in (3) Forced Border Conflict, Number suite, yet all variables in the (5) Combined suite were significant.
2. For “Not In Conflict” Latin America regional models, “Fresh Water per Capita” in the (2) Forced Water per Capita suite and both the intercept and “Border Conflict Score, Number” variable in the (3) Forced Border Conflict, Number suite exceeded the desired p-value, yet all variables in the (5) Combined suite were significant.
3. For the “In Conflict” OECD regional models, the “Border Conflict Score, Number” variable exceeded the desired p-value in (3) Forced Border Conflict, Number suite, yet in the (5) Combined suite the variable achieved an acceptable p-value while the “Water per Capita” variable increased to an undesirable p-value unlike in the (2) Forced Water per Capita suite.

4. For the “In Conflict” South and East Asia regional models, acceptable p-values for all modeled variables in the forced one-variable models (suites 2 and 3), yet experienced undesirable p-values for the intercept and “Fresh Water per Capita” variable in the (5) Combined suite.

Ultimately, there was no conclusive evidence that environmental factors were the driving force for the increased parsimony over the Shallcross (2016) models. The results from Table 4 lend to the hypothesis that the step-wise regression technique employed affected the parsimony of the models more than the variables themselves. An added benefit to this conclusion, however, is that multi-collinearities in the data may support multiple variations of models with equivalent accuracies.

4.2 Model Accuracy

Strategic advantage in predictive analytics requires a measure of high fidelity. The technique of observing accuracy as variables are added to models already ensures a level of retained accuracy to previous work as seen in Table 3. However, the goal of investigating environmental factors sought to provide insight into achieving increased accuracy.

Accuracy taken at the 0.50 cut point did not demonstrate favorability toward the (5) Combined environmental factors suite compared to the (1) Most Significant suite as seen in Table 4. The (5) Combined suite only provided greater than one percent improvement over the other models in only two of the twelve regional models for the training set data (“In Conflict” Eastern Europe and Central Asia: +1.49 and “In Conflict” South and East Asia: +5.66). Comparing the models according to the area under the

ROC curve, the (5) Combined environmental factors models never showed an improvement above one percent compared to the other suite models. Despite the (5) Combined suite's inability to demonstrate significant improvement, it also did not regress significantly in predictive accuracy.

When comparing the models based upon their cut point accuracy from training data to validation data, there appeared to be a trend between all models when comparing accuracy variation. For example, as seen in Table 5, variation trended well for "Not In Conflict" OECD with all models demonstrating approximately 95% training accuracy and maintaining approximately 94% validation accuracy. Conversely, all models trended poorly for "Not In Conflict" Arab and North Africa demonstrating approximately 90% training accuracy and declining to approximately 67% validation accuracy. Transitions of conflict state are fairly rare occurrences averaging only 15% of all data points used in this study. Parsed by region and conflict state, some combinations were even scarcer with levels as low as 0.6%. The (5) Combined suite's classification table accuracies at 0.50 cut point were compared to the number of transitions in the data set for correlation. The hypothesis is that the modeling trends of training to validation accuracy are correlated to the disparity in the percent of transitions in each regional data set. A

Table 5: Suite Comparisons for Training and Validation Trends

Suite	"Not In Conflict" OECD		"Not In Conflict" Arab & North Africa	
	Training	Validation	Training	Validation
(1) Most Significant	96.0%	94.1%	91.7%	66.7%
(2) Force Water per Capita Variable	94.2%	93.8%	96.7%	66.7%
(3) Force Border Conflict, Number Variable	94.4%	94.1%	88.3%	66.7%
(4) Force Border Conflict, Binary Variable	95.2%	94.1%	85.0%	66.7%
(5) Combined Water & Border Conflict Variable	94.2%	93.8%	88.3%	66.7%
Average	**94.8%**	**94.0%**	**90.0%**	**66.7%**
Standard Deviation	**0.79%**	**0.13%**	**4.41%**	**0.00%**

positive transition change means that more transition data points were observed in the training set while negative transition change means more transition data points were observed in the validation set. The "In Conflict" regions trended to decrease in the percent of transitions from 2004 to 2013, whereas "Not In Conflict" regions trended the opposite with four out of the six regions showing an increase in transitions for the validation years. The argument, as supported by Table 6, would support models such as "In Conflict" Sub-Saharan Africa, which although it had a slight improvement in accuracy from training to validation, the transition in raw data had only a slight decrease in transitions from training set to validation set resulting in a difference of only -1.2%. The negative sign indicates that the model accuracy had a smaller percent change than the percent transitioned change. Similarly, "Not In Conflict" Arab and North Africa had a large discrepancy in accuracy (21.66%), which might be attributed to the discrepancy in transitions between the training and validation set (-18.33%) for a small difference of 3.3%. Although the signs are opposite, the hypothesis test assumes that large differences

Table 6: Cut Point Accuracy to Transition Percent Correlation

(5) Combined Water per Capita & Border Conflict, Number Modeling Suite

Conflict State	Region	Accuracy		Percent Transitioned		Percent Change		Delta
		Training	Validation	Training	Validation	Accuracy	Transitioned	
In Conflict	Arab & North Africa	94.23	81.40	9.62	2.33	12.83	7.29	5.5
	Eastern Europe & Central Asia	97.01	79.31	19.40	20.69	17.70	-1.29	16.4
	Latin America	91.89	73.33	18.92	6.67	18.56	12.25	6.3
	OECD	96.23	77.27	15.09	13.64	18.96	1.45	17.5
	South & East Asia	92.41	72.73	16.46	13.64	19.68	2.82	16.9
	Sub-Saharan Africa	85.71	86.49	16.88	14.86	-0.78	2.02	-1.2
Not In Conflict	Arab & North Africa	88.33	66.67	15.00	33.33	21.66	-18.33	3.3
	Eastern Europe & Central Asia	90.24	79.07	13.01	16.28	11.17	-3.27	7.9
	Latin America	85.61	83.33	13.64	19.05	2.28	-5.41	-3.1
	OECD	94.21	93.81	7.14	3.96	0.40	3.18	-2.8
	South & East Asia	87.88	84.00	18.18	8.00	3.88	10.18	-6.3
	Sub-Saharan Africa	86.24	80.82	15.87	16.44	5.42	-0.57	4.9
	Weighted Average	89.67	82.49	14.62	12.45	8.19	1.54	**6.7**

between the training and validation transition percent would have an equally large change in accuracy of the models. Therefore, absolute values were used to calculate the difference between the accuracy change in model accuracy and percent of transitioning data to observe correlation. Positive values mean the classification table accuracies had larger discrepancies than the transitions in the data while negative values mean the transitions in the data varied more than the classification accuracies. It is assumed that the more negative the number, the better the model was at overcoming the variation of transition differences between the training data and the validation data. Regardless, large delta values indicate that correlation may not exist between accuracy and maintaining similar transition percentages in the data set. For the most part, Table 6 shows that with the majority of models we may not be able to rule out possible correlation between accuracy degradation and variation of transitioning dependent variables from training to validation data sets with a weighted average of only 6.7%. However, three of the "In Conflict" models have a large enough disparity to assume that correlation is not the significant factor in the poor predictive accuracy of the validation sets. The volatile state of already being in conflict may confound predictive achievement in modeling status transitions for these three models.

The implications of this finding may conclude that the study assumption of years used for the models may include a cultural shift from normality. A commonly known example would be the Arab Spring of 2011 and placing the training and validation modeling delineation at the same time frame. The Arab Spring (or more appropriately, the Arab Inferno because of the event is a description of substantial increase in anti-government protests, uprisings, and armed rebellions) may be a cultural shift where including data points prior to 2011 in the modeling may not be characteristic of future conflict predictions.

Furthering this investigation, the accuracy of the transitioning variable was observed. Even though the accuracy of the model may be high, the disparity between staying in a given state and transitioning out of a state could bias the weighted average of the model accuracy. This result, as seen in Figure 4, can be most clearly observed in the "Not In Conflict" Latin America model where its model training accuracy is 85.61%, but the sensitivity of the classification table is only 16.7% (3 out of 18 known transitions). The model is only 16.7% accurate at correctly identifying a transitioned country as having transitioned. If the goal of the model is to provide useful insight, not only does it have to be accurate in identifying the correct state, but also not overly biased as to be mostly incorrect at identifying these rare occurrences of interest. The VBA code was reconfigured to also display both sensitivity and specificity when analyzing each possible covariate for inclusion into the model. The hypothesis is that there might be alternate variables to bring into the model that may not have the largest G statistic for a particular

Region: **Latin America**
State: Not In Conflict
Focus: *(5) Force Water per Capita and Border Conflict, Number Variable*

Variables	Coefficients
Intercept	-5.5905
Fertility Rate	1.5510
Population Density	-0.0079
Fresh Water per Capita (1000s)	-0.0132
Border Conflict Score, Number	1.7083

Model Significance, $p > \chi^2$: 0.0000

Model — Years: 2004-2010

Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	3	4
	as NC	15	110
		18	**114**

Accuracy	Sensitivity	Specificity
85.61	0.167	0.965

AUC: 0.856 N: 132

Ĉ by Bins: 10.49 Ĉ by 10ths: 8.97

$\chi^2_{(0.05,10)}$: 18.31 $\chi^2_{(0.05,8)}$: 15.51

Validation — Years: 2011-2013

Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	1	0
	as NC	7	34
		8	**34**

Accuracy	Sensitivity	Specificity
83.33	0.125	1.000

N: 42

Figure 4: Final Output - Latin America "Not In Conflict" (5) Combined

step, but may have a G statistic that is significant (above 3.78) and have a higher transitioning variable accuracy. Transitioning variable accuracy meaning specificity for "In Conflict" models and sensitivity for "Not In Conflict" models. "In Conflict" Eastern Europe and Central Asia and "In Conflict" Arab and North Africa were selected for a cursory look. At every step in the modeling selection, the highest G statistic also had the highest transitioning variable accuracy (specificity for these two models). There were other variables that shared the highest transitioning variable accuracy with a G statistic above 3.78, but selecting them did not provide any prediction benefit.

The implications of this finding support the conclusion that the categorization of countries by region may be less optimal than another type of categorization. The locations of false classifications were investigated to see if the misclassifications were concentrated along the borders to the different regions. If they were, then maybe the borders could be redrawn. Figure 5 illustrates the countries that demonstrated a transition change during either the training or validation data set years. Green countries had

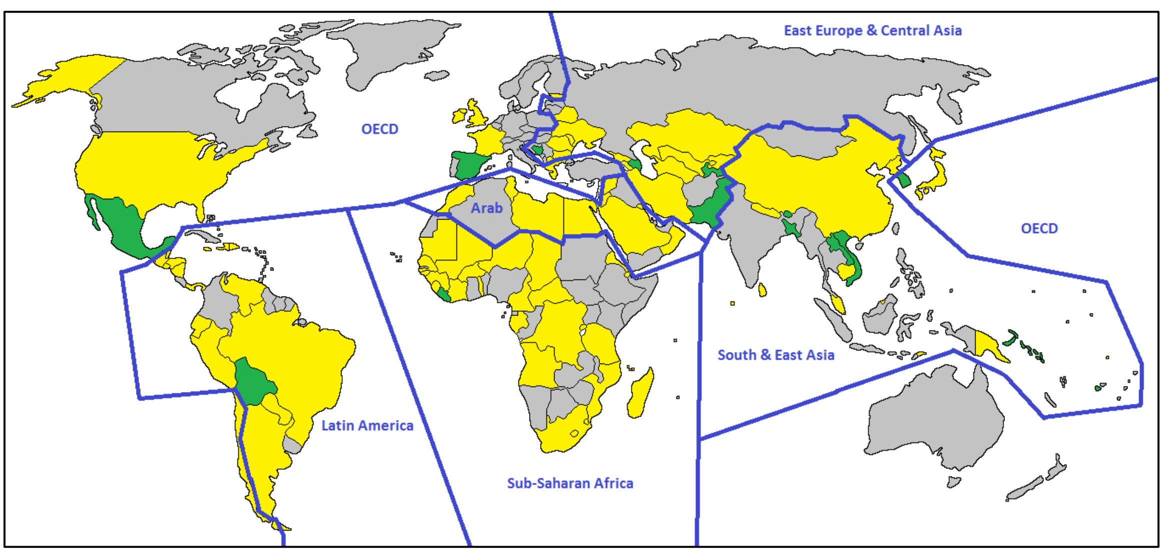


Figure 5: Geographic Visual of Classification Accuracy by Region

conflict transitions that were correctly predicted by the (5) Combined suite with a cut point of 0.5, while yellow countries had at least one conflict transition that was incorrectly predicted with the same cut point. Grey countries had no transitions. Even with model accuracies above 80%, the majority of transitioning countries are misclassified at least once. Additionally, there were no indications showing that alternative borders would increase accuracy results.

Investigating a different categorization schema, four whole world models were developed using data between the years 2007 and 2013 as seen in Appendix B. The first two models were modeled based upon their state. As seen in Figure 6, the non-transitioning variables in both the "In Conflict" model (sensitivity) and the "Not In Conflict" model (specificity) were above 94%; however, the transitioning variables classified at best only 12.5%. Overall, the whole world models only classified 10 of the

"In Conflict"

Years: 2007-2010

Training			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	237	39
	as NC	2	5
		239	**44**
Accuracy		Sensitivity	Specificity
85.5%		99.2%	11.4%

Years: 2011-2013

Validation			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	219	28
	as NC	12	4
		231	**32**
Accuracy		Sensitivity	Specificity
84.8%		94.8%	12.5%

"Not In Conflict"

Training			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	1	6
	as NC	56	382
		57	**388**
Accuracy		Sensitivity	Specificity
86.1%		1.8%	98.5%

Validation			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	0	4
	as NC	36	243
		36	**247**
Accuracy		Sensitivity	Specificity
85.9%		0.0%	98.4%

Figure 6: Whole World Model Classification Tables

169 transitions correctly even though the model accuracy was above 84% at a 0.50 cut point. Although the classification by regions demonstrates less than optimal results for the transitioning variables, their classification is still better than using the whole world model.

The other set of whole world models investigated modeling only the countries that had at least one transition between 2007 and 2013. The thought behind this modeling scheme was to decrease the difference in the ratio of non-transitioning data points to transitioning data points. The models were closer to 2:1 as opposed to the 6:1 of the region specific models. After the models were generated, all countries (including those that did not have a transition) were assessed by classification table as seen in Figure 7. The "In Conflict" model experienced positive effects in classifying the transitioned variable, but at the expense of misclassifying the non-transitioned variable. The "Not In

"In Conflict"

Years: 2007-2010

Training			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	64	27
	as NC	7	17
		71	**44**
Accuracy		Sensitivity	Specificity
70.4%		90.1%	38.6%

Years: 2011-2013

Validation			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	40	18
	as NC	27	14
		67	**32**
Accuracy		Sensitivity	Specificity
54.6%		59.7%	43.8%

"Not In Conflict"

Training			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	3	2
	as NC	54	128
		57	**130**
Accuracy		Sensitivity	Specificity
70.1%		5.3%	98.5%

Validation			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	0	0
	as NC	36	48
		36	**48**
Accuracy		Sensitivity	Specificity
57.1%		0.0%	100.0%

Figure 7: Modified Whole World Model Classification Tables

Conflict" model did not see any benefit to overcome modeling bias of the non-transitioned variable. Overall, this technique did not achieve superior results above classifying by regions.

4.3 Variable Manipulation

Outside of accuracy, alternative modeling may provide strategic advantage by ensuring the inclusion of a cost effective variable. In the case of predicting conflict, the cost effective variable may be more malleable by either an ally or an adversary, such as environmental factors. For example, targeting fresh water sources may be easier or more time sensitive to manipulate than adjusting the ratio of ethnic diversity.

Conflict state appears to be an important consideration when developing assumptions based upon the sign of a model coefficient. Considering only models with variable p-values less than 0.25 (annotated in black) for the (5) Combined suite, Table 7 clearly shows trends in three of the state/variable combinations. Although the trend is weak with only one regional data point for "In Conflict" and only two regional data points for "Not In Conflict".

The assumption is that as "Water per Capita" increases, the likelihood that a nation will transition or remain "Not In Conflict" should increase. Likewise, the assumption for "Border Conflict" is that as the number of "In Conflict" neighboring counties increases, the likelihood that a nation will remain or transition "In Conflict" should increase. In other words, Table 7 assumes that the "Water per Capita" column should show all negative signs, while the "Border Conflict" column should show all positive signs. The discrepancies are highlighted in yellow.

Table 7: Environmental Variables Sign and Coefficient Data

Conflict State	Region	Water per Capita		Border Conflict	
		Sign	Value	Sign	Value
In Conflict	Arab & North Africa	Positive	85.8	Negative	-9.42
	Eastern Europe & Central Asia	Positive	0.21	Positive	3.82
	Latin America	Positive	0.086	Negative	-6.21
	OECD	Negative	-0.03	Positive	9.99
	South & East Asia	Negative	-0.02	Negative	-4.91
	Sub-Saharan Africa	Negative	-0.07	Positive	1.14
Not In Conflict	Arab & North Africa	Negative	-2.86	Positive	0.08
	Eastern Europe & Central Asia	Negative	-0.08	Positive	0.28
	Latin America	Negative	-0.01	Positive	1.71
	OECD	Positive	0.017	Positive	0.12
	South & East Asia	Negative	-0.02	Positive	1.26
	Sub-Saharan Africa	Negative	-0.003	Positive	0.27

The reason for these discrepancies traces back to correlation with other variables in the models. The sum product of the logit transformation produces confounding averages for the variable coefficients as shown in Equation 1. The true effect of each variable can only be seen in a single variable logistic regression that excludes multicollinearity. Environmental factors were plotted for the twelve conflict state and region combinations as seen in Appendix D. The assumption is that the plots would show trends such as illustrated in Figure 8 by the Whole World plots: "Fresh Water per

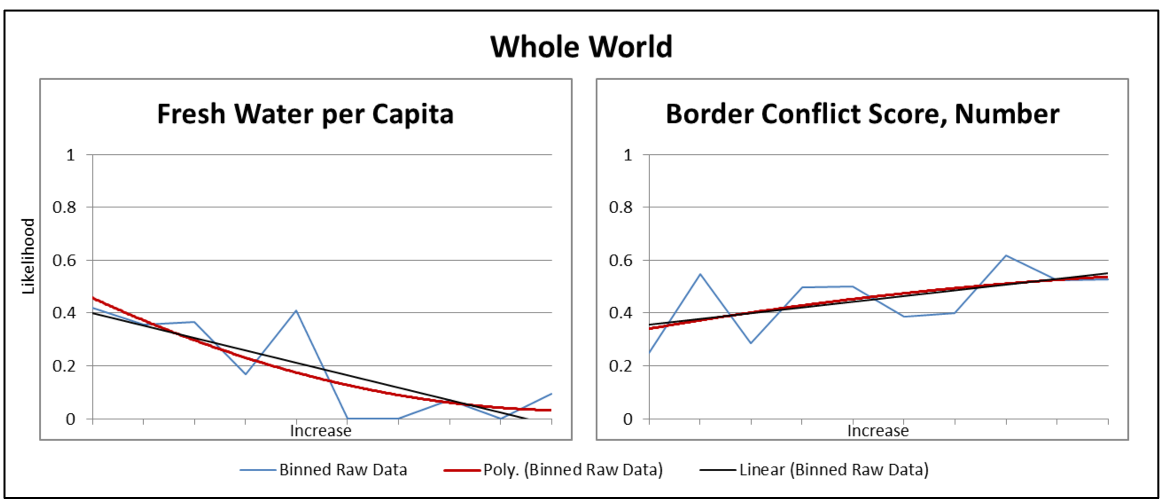


Figure 8: Whole World Univariable Logistic Regression Plots

Capita" tendencies should have decreasing slopes while "Border Conflict Score, Number" tendencies should have increasing slopes. The raw data was binned into ten segments and averaged as demonstrated by the blue line. Both a linear and a second-order polynomial trend line were fit to ascertain trends, as seen by the black and red lines respectively. Fresh Water per Capita shows that as fresh water increases, the likelihood that a nation will transition to "In Conflict" decreases. The border conflict score shows that as the number of surrounding nations in conflict increase, the likelihood that a nation will transition to "In Conflict" increases. Despite the signs in Table 7, the plots in Appendix D support the assumptions made about "Fresh Water per Capita" and "Border Conflict Score, Number" with the exception of "In Conflict" Latin America.

Further in-depth investigation is outside the scope of this study; however, "In Conflict" Arab and North Africa was considered to demonstrate the effects of forcing variable coefficients to obey the environmental factors assumptions. The assumption being that "Fresh Water per Capita" requires a negative coefficient because of its negative slope, while "Border Conflict Score, Number" requires a positive coefficient because of its positive slope. "Fresh Water per Capita" was forced to retain a negative coefficient of 0.5 or less and "Border Conflict Score, Number" was forced to retain a positive coefficient of 0.5 or more. Figure 9 shows the results of the modified model. The optimization of the coefficients tried to force the water and border conflict coefficients to zero. The classification results did not change, as listed in Figure 9, with the two coefficients set to zero. The classification accuracy is not as good as the (1) Most Significant suite, which attained 96% training accuracy and 93% validation accuracy. Additionally, the training accuracy of 90.38% is lower than the (5) Combined suite

Region: **Arab and North Africa**
State: In Conflict
Method: *(5) Combined Suite holding Water coefficient negative and Border Conflict coefficient positive*

Variables	Coefficients
Intercept	29.7296
Arable Land	3.4939
Refugee (Asylum) (10,000s)	-0.0143
Fresh Water per Capita (1000s)	-0.5000
Govt Type (Autocratic)	-4.2610
Ethnic Diversity	-27.2174
Border Conflict Score, Number	0.5000

Model Significance, $p > \chi^2$: 0.0638

Years: 2004-2010

Model			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	45	3
	as NC	2	2
		47	**5**

Accuracy	Sensitivity	Specificity
90.38	0.957	0.400

AUC: 0.927 N: 52

Ĉ by Bins: 8.63 Ĉ by 10ths: 1.89

$\chi^2_{(0.05,11)}$: 19.68 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	37	1
	as NC	5	0
		42	**1**

Accuracy	Sensitivity	Specificity
86.05	0.881	0.000

N: 43

Figure 9: Modified Model Forcing Coefficients to Obey Assumptions

training accuracy. However, the validation accuracy of the model at 86.05% was higher than the 81% validation of the (5) Combined suite. Despite the differences in accuracy, all accuracies are higher than the 80% benchmark of the CIA study, and therefore maintain plausibility that models could be forced to obey environmental factor assumptions in order to consider the effects of variable manipulation.

4.4 Conflict Prediction Optimization

Although no one specific suite demonstrated superior results, with the exception of forcing the binary border conflict variable (had no superior results), each suite did provide at least one model to construct a conflict prediction optimization portfolio. The optimization portfolio was constructed by choosing the focus area model that demonstrated the superior classification accuracy within each region and conflict state combination. Furthermore, sensitivity analysis was implemented on the cut-off value for the classification table to ensure the best possible classification accuracy.

Sensitivity analysis provides a range of values from which one point could be selected for the generation of the classification tables. Each model was analyzed according to both the overall classification accuracy and the component percentages of incorrect classifications using a range of cut-off values from zero to one, stepping every hundredth. In some cases, multiple suites would share a highest overall classification accuracy value. The tie breaker for the optimal model followed four simple steps.

1. If the optimal training accuracy is the same, choose the suite that has superior validation accuracy at the optimal training accuracy cut-off.
2. If the validation accuracy is the same, choose the suite where the transitioning variable "fault" is minimized. The transitioning variable is the rare occurrence for the conflict state. When the region conflict state combination is "In Conflict", then the transitioning variable "fault" to be minimized is the False "Conflict" variable. In other words, classifying a country as "In Conflict" when it has been observed to be "Not In Conflict" would be False "Conflict". Conversely, when the region conflict state combination is "Not In Conflict", the transitioning variable "fault" to be minimized is False "Not Conflict": classifying a country as "Not In Conflict" when it has been observed to be "In Conflict".
3. The training data set "faults" would be considered before the validation data set "faults".
4. If the transitioning variable "fault" percentage at the optimal training accuracy cut-off value is the same, then the suite with the least variables that have a p-value over 0.25 is chosen for the portfolio.

Table 8 provides a break-down of the models selected for the optimized portfolio. The majority of models were selected based upon superior training accuracy. South East Asia "In Conflict" was selected based upon validation accuracy. Eastern Europe and Central Asia "Not In Conflict" and Latin America "In Conflict" were selected based upon the transitioning variable "fault" percentage. Finally, Arab and North Africa "In Conflict" and Sub-Saharan Africa "In Conflict" were selected based upon minimizing variables that had p-values over 0.25.

Table 8: Optimized Portfolio Quantitative Results

Region	State	Focus	Cut-Off	Training Accuracy	ROC	H-L Test	Validation Accuracy
Arab & North Africa	In Conflict	Most Significant	0.50	96.15%	0.940	Both	93.02%
	Not In Conflict	Forced Water	0.40	98.33%	0.993	Both	66.67%
Eastern Europe & Central Asia	In Conflict	Forced Border Conflict, Number	0.50	97.01%	0.996	Tenths	72.41%
	Not In Conflict	Forced Water	0.52	91.06%	0.916	Both	70.07%
Latin America	In Conflict	Combined Water & Border Conflict	0.50	91.89%	0.890	Both	73.33%
	Not In Conflict	Forced Border Conflict, Number	0.60	91.67%	0.908	Tenths	83.33%
OECD	In Conflict	Combined Water & Border Conflict	0.51	98.11%	0.956	Tenths	77.27%
	Not In Conflict	Most Significant	0.50	96.03%	0.963	Binning	94.06%
South & East Asia	In Conflict	Combined Water & Border Conflict	0.40	94.94%	0.948	Both	75.00%
	Not In Conflict	Forced Water	0.50	90.91%	0.903	Both	84.00%
Sub-Saharan Africa	In Conflict	Most Significant	0.50	85.71%	0.769	Both	86.49%
	Not In Conflict	Most Significant	0.40	87.83%	0.812	Tenths	80.82%
Weighted Average				**91.99%**			**82.56%**

Although the sensitivity analysis provides a range of cut-off values that could be used for optimizing accuracy, a single cut-off was selected for the table. For example, the Arab and North Africa "In Conflict" model could have used any cut-off between 0.38 and 0.51 and still maintain an accuracy of 96.15%. If the range included 0.50, then 0.50 was selected. Otherwise, the selection used either a multiple of ten within the optimal range or in the case of Eastern Europe and Central Asia "Not In Conflict" and OECD "In Conflict", the point optima of 0.52 and 0.51 respectively.

Through the use of different modeling foci, classification accuracies were achieved above 85% with some models achieving results as high as 98%. The ROC curve value provides an assessment of how well the model discriminates over the range of data points similar to a sensitivity analysis. Nine of the models are categorized as outstanding discriminators, while Latin America "In Conflict" and Sub-Saharan Africa "Not In Conflict" are classified as excellent and Sub-Saharan Africa "In Conflict" as acceptable. The Hosmer-Lemeshow test provides an assessment of how well the model is calibrated to predict the real underlying probability, also known as model fit. All twelve models passed at least one of the Hosmer-Lemeshow tests with seven of the models passing both. The seven that passed both could be considered as being doubly verified on its calibration of model fit. Table 9 provides a qualitative assessment of the optimized portfolio. Green identifies either high classification accuracy (greater than 90%) for logistic regression models, outstanding or excellent classification discrimination through ROC curve interpretation, or one to both acceptable calibration results of model

Table 9: Optimized Portfolio Qualitative Assessment

Region	State	Suite	Modeling Accuracy	Discrimination	Calibration	Predicting Accuracy
Arab & North Africa	In Conflict	Most Significant	96%	Oustanding	Both	93%
	Not In Conflict	Forced Water	98%	Oustanding	Both	67%
Eastern Europe & Central Asia	In Conflict	Forced Border Conflict	97%	Oustanding	One	72%
	Not In Conflict	Forced Water	91%	Oustanding	Both	70%
Latin America	In Conflict	Combined	92%	Excellent	Both	73%
	Not In Conflict	Forced Border Conflict	92%	Oustanding	One	83%
OECD	In Conflict	Combined	98%	Oustanding	One	77%
	Not In Conflict	Most Significant	96%	Oustanding	One	94%
South & East Asia	In Conflict	Combined	95%	Oustanding	Both	75%
	Not In Conflict	Forced Water	91%	Oustanding	Both	84%
Sub-Saharan Africa	In Conflict	Most Significant	86%	Acceptable	Both	86%
	Not In Conflict	Most Significant	88%	Excellent	One	81%
Weighted Average			**92%**			**83%**

fit through the Hosmer-Lemeshow test interpretation. Yellow identifies adequate classification accuracy or acceptable discrimination. Red (no models met this criteria) identifies less than the desired 80% classification accuracy, poor to no model discrimination, or the failure of both Hosmer-Lemeshow tests for model calibration.

A weighted average was used to combine the conflict states of each region to provide a comparison to both the Boekestein (2015) and Shallcross (2016) models. As seen in Table 10, the new portfolio demonstrates improvements in training the data sets to predict conflict, although the quantified prediction does not show improvements. It is assumed that superior training of the models would lead to superior prediction results unless the models were over fit. However, over fit is dismissed due to the models achieving superior calibration and maintaining parsimony. The degradation in prediction results as shown in all three models is then assumed to be a product of the starting assumptions, such as the model variables will remain significant through subsequent prediction years.

Table 10: Comparison of Model Accuracies with Predecessors

	Accuracy of Training Regional Models			Accuracy of Predicting Regional Conflict		
Region	**Boekestein Model**	**Shallcross Model**	**Optimized Portfolio**	**Boekestein Model**	**Shallcross Model**	**Optimized Portfolio**
Arab & North Africa	84.31%	93.72%	**97.32%**	70.59%	70.68%	**78.90%**
Eastern Europe & Central Asia	77.38%	87.83%	**93.16%**	75.00%	**79.16%**	70.90%
Latin America	90.12%	88.75%	**91.72%**	77.78%	**86.10%**	81.14%
OECD	95.96%	93.84%	**96.65%**	92.42%	**93.46%**	89.09%
South & East Asia	90.48%	87.57%	**93.10%**	76.79%	**85.51%**	79.10%
Sub-Saharan Africa	82.31%	85.74%	**86.88%**	74.49%	**84.34%**	83.37%
Combined World Results	86.63%	88.76%	**91.99%**	78.30%	**84.67%**	82.56%

In defense of the lower prediction accuracies to Shallcross (2016), it is observed that the overall prediction accuracy of the (1) Most Significant suite was 84.4%, which would be very close to the Shallcross prediction accuracy. However, because validation data would not be normally known in real-world applications, we did not use validation data as the primary driver for developing the optimized portfolio. The second rule used to develop the optimized portfolio (which refers to using validation data) was only invoked on one of the twelve models. Additionally, the one model (South East Asia "In Conflict") selected shared the exact same validation accuracy as Shallcross (2016). The difference shown in Table 10 is an aggregate of both "In Conflict" and "Not In Conflict" prediction accuracies.

Finally, the optimized portfolio is compared to a naïve prediction based on the logic that transitions are very rare occurrences. Therefore, the assumption is that a country will remain in its current state for the next couple years. Table 11 demonstrates the logic with a 1, 2, and 3 year outlook along with a three year cumulative result. Each year shows how many transitions occurred contrary to a country's state identified in the last year of the training set for each region. For example, the training data year set for "Not In Conflict" South and East Asia ended in 2011. In 2011, Maldives was in a "Not

Table 11: Optimized Portfolio vs Naïve Prediction

Region	State	Naïve Per Year Accuracy			Cumulative	Optimized Portfolio
		Year 1	Year 2	Year 3	3-Year	
Arab & North Africa	In Conflict	100.0%	100.0%	100.0%	100.0%	93.0%
	Not In Conflict	77.8%	22.2%	33.3%	44.4%	66.7%
Eastern Europe & Central Asia	In Conflict	83.3%	83.3%	66.7%	77.8%	72.4%
	Not In Conflict	87.5%	87.5%	62.5%	79.2%	70.1%
Latin America	In Conflict	88.9%	100.0%	88.9%	92.6%	73.3%
	Not In Conflict	88.2%	70.6%	64.7%	74.5%	83.3%
OECD	In Conflict	66.7%	77.8%	55.6%	66.7%	77.3%
	Not In Conflict	95.8%	95.8%	95.8%	95.8%	94.1%
South & East Asia	In Conflict	81.8%	90.9%	90.9%	87.9%	75.0%
	Not In Conflict	73.3%	80.0%	86.7%	80.0%	84.0%
Sub-Saharan Africa	In Conflict	87.5%	91.7%	79.2%	86.1%	86.5%
	Not In Conflict	88.0%	80.0%	84.0%	84.0%	80.8%
Weighted Average		**86.2%**	**83.4%**	**77.9%**	**82.5%**	**82.6%**
Standard Deviation		**9.18%**	**20.83%**	**19.49%**	**14.76%**	**8.77%**

In Conflict" state, therefore, it is assumed that it would remain "Not In Conflict" for the next couples years. However, in 2012, it transitioned into conflict and remained "In Conflict" until 2013. In 2014, it transitions out of conflict. Therefore, the Year 1 prediction result would be 0%, Year 2 prediction result would be 0%, and Year 3 prediction result would be 100%. The three year cumulative result would be 33%. Surprisingly, the naïve predictions does surprisingly well in Year 1, however, it then starts to degrade toward Year 3. Overall, its weighted average for the three year cumulative prediction compares with the Optimized Portfolio, except the standard deviation between regions is almost twice as large. If data were available, projecting further than three years would suggest through the trends that the optimized portfolio would outperform the naïve prediction method.

V. Conclusions and Recommendations

5.1 Conclusions of Research

This study used step-wise logistic regression observing thirty-nine independent variables to predict the transition propensity of 182 countries. It used six geographic regions and two conflict status indices to generate twelve models for five different suites. The suites were compared to ascertain improvements in modeling for strategic advantage. The conjecture was that by forcing environmental factors into the models, collinearities in the independent variables would unmask and vastly improve both model parsimony and predictive accuracy. Although some gains were found in this study, improvements in parsimony and predictive accuracy can only marginally be attributed to environmental factors. However, by allowing a mixture of methods to develop an optimal suite of models, the training accuracies achieved a new high of 92% accurate classification, although the validation accuracies did not see similar results.

All models developed in this study demonstrated statistical equivalency as whole suites concerning the number of significant variables required for predictive purposes. Although the environmental factors did not produce smaller models, they did not increase the number of variables required either. However, the step-wise technique used in this study through the use of the VBA program developed averaged two less required variables than the purposeful selection technique used in the previous study. The credit for this achievement is based on observing model discrimination characteristics as well as variable significance while building the models. By assessing discrimination characteristics of the model while building, parsimony can be advanced, which otherwise is secondary due to the less stringent p-value threshold needed for logistic regression in this area. Although a variable may demonstrate significance in the p-value range of 0.15 to 0.25, if the variable does not improve the model in the area of cut point accuracy or ROC score, the variable is of little use for prediction purposes.

Unfortunately, environmental factors did not demonstrate improvements in the area of predictive accuracy, as seen in Table 4, garnering an overall 82.6% validation accuracy: two percent less than (1) Most Significant suite. Although, the variance in validation accuracy was tighter by three percent standard deviation over the (1) Most Significant suite. For the most part, all environmentally-forced models were statistically equivalent to the models generated based on significance alone.

The benefit of leveraging the collinearities of environmental factors in model generation is in ensuring certain variables are in a model due to their easier manipulation. Although the models did not all follow the environmental factors assumptions, it was demonstrated that models could be forced to follow the assumptions with little degradation in predictive accuracy.

5.2 Significance of Research

The driving purpose of this research was to assist commanders in developing Theater Campaign Plans by providing insights into the future posture of countries around the world. Through logistic regression, the models developed can provide decision makers with the best accuracy for predictive assessments. The added insight from this study demonstrated that through correlation in the plethora of open source data available, alternative variables may be substituted into models and still retain statistically equivalent accuracies. This is important because not all countries show variation in the same variables within their respected regional models. If intelligence analysts question the results of a country within a model, an alternative model may be developed with different yet correlated variables and the country can be reassessed as an alternative option with little to no degradation in accuracy. This implication is quite astounding and worth repeating, alternative model options are not necessarily limited to only environmental factors, but available to many of the correlated factors in the open source data. Alternatively, sensitivity analysis could be conducted on key variables to open the discussion of emerging threats or opportunities.

5.3 Recommendations for Future Research

Throughout this study, possible biases were uncovered that may be limiting predictive accuracy through logistic regression. The first example was a possible correlation between model accuracy and the ratio of transitions in the data sets. The second example was the finding of low transitioning variable accuracy while maintaining high overall model accuracy. These findings warrant further investigation into methods to improve modeling for prediction purposes; methods that begin this different assumptions.

Future research could investigate alternative ways to classify grouping countries together for model generation rather than using the assumption that geographic regions share some common element that proposes the best grouping. Any one of the independent variables could be used as a discriminator for possible groupings.

If the regions are maintained, additional research could also implement a systematic approach for deciding when a cultural shift has happened in a grouping of countries where previous year data may be deemed irrelevant for model generation. This study used the same year groupings as the previous study without regard to investigating the method used to choose the year groupings.

The other major assumption was that variables would retain significance throughout future prediction years. Two issues could be happening with this study that causes the decline in predictive power. First, variables are declining in statistical significance so much that other variables are now driving the influences of conflict transition that are not captured in the model. Alternatively, the variable is retaining significance, but its representative coefficient is significantly skewed to properly represent future predictions, therefore the coefficients need to vary from year to year. Both of these issues could be difficult to capture, but modeling may be developed to minimize the effects.

One other way to possibly minimize the effects of good modeling to poorer prediction trends is to derive a new dependent variable. The current dependent variable is a one year transition change. The modeling did not provide better one year predictive power over the naïve approach (86.2% naïve to 82.6% modeling). However, the modeling did provide better three year predictive power over the naïve approach (77.9% naïve to 82.6% modeling) along with better standard deviation (19.49% naïve to 8.77% modeling). The better dependent variable may be a three year transition mode to overcome unexplained transitions not supported by the significant variables. Some of the incorrect classifications noted frequent dependent variable changes with little to no change in the significant variables. Alternatively, this idea could also be applied to the independent variables instead to identify lags in prediction outcomes.

In line with both the national cultural shift theory and the frequent dependent variable changes, another variable could be developed for modeling consideration: malleability of a country to transition states. The more a country transitions from one state to the other, the more malleable it may be to transitions in the future. This would be similar to the idea that bordering countries in conflict may prompt a country to also transition into conflict. Additionally, in this study, the border conflict variables were calculated as a static measure. Using Monte Carlo simulations, nation specific dynamic border conflict scores could be generated for inclusion into model generation.

Finally, simulation could be used on the developed models using distributions on the coefficients to investigate the effects that varying coefficients may have on predictive accuracy. Sensitivity analysis can help ascertain the shape and range of the distributions to use. Alternative models may be desired to include forcing environmental factors assumptions even though the modeling accuracies appeared to decline in the one example demonstrated in Section 4.3 on Arab and North Africa. The distributions may ultimately help the models predict better.

Appendix A: Region Specific Conditional Logistic Regression Models

Region: **Arab and North Africa**
State: In Conflict

Method: *Most Significant*

Variables	Coefficients
Intercept	97.6174
Birth Rate	-0.5283
Death Rate	1.3634
Govt Type (Autocratic)	-7.0265
3 Yr Freedom Trend	50.4594
Ethnic Diversity	-90.1850

Model Significance, $p > \chi^2$: 0.0023

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	46	1
	as NC	1	4
		47	**5**

Accuracy	Sensitivity	Specificity
96.15	0.979	0.800

AUC: 0.940 N: 52

Ĉ by Bins: 7.76 Ĉ by 10ths: 1.27

$\chi^2_{(0.05,11)}$: 19.68 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	40	1
	as NC	2	0
		42	**1**

Accuracy	Sensitivity	Specificity
93.02	0.952	0.000

N: 43

Method: *Force Water per Capita Variable*

Variables	Coefficients
Intercept	124.4976
Birth Rate	-0.5684
Fresh Water per Capita (1000s)	8.6859
Govt Type (Autocratic)	-11.9355
Ethnic Diversity	-110.1565

Model Significance, $p > \chi^2$: 0.0117

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	46	4
	as NC	1	1
		47	**5**

Accuracy	Sensitivity	Specificity
90.38	0.979	0.200

AUC: 0.919 N: 52

Ĉ by Bins: 1.25 Ĉ by 10ths: 5.80

$\chi^2_{(0.05,11)}$: 19.68 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	42	1
	as NC	0	0
		42	**1**

Accuracy	Sensitivity	Specificity
97.67	1.000	0.000

N: 43

Method: *Force Border Conflict, Number Variable*

Variables	Coefficients
Intercept	22.8725
Govt Type (Autocratic)	-2.8334
Ethnic Diversity	-20.9552
Border Conflict Score, Number	-0.0039

Model Significance, $p > \chi^2$: 0.0340

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	46	4
	as NC	1	1
		47	**5**

Accuracy	Sensitivity	Specificity
90.38	0.979	0.200

AUC: 0.924 N: 52

Ĉ by Bins: 7.09 Ĉ by 10ths: 0.45

$\chi^2_{(0.05,11)}$: 19.68 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	42	1
	as NC	0	0
		42	**1**

Accuracy	Sensitivity	Specificity
97.67	1.000	0.000

N: 43

Figure 10: "In Conflict" Arab and North Africa Models

Region: **Arab and North Africa**
State: In Conflict

Method: *Force Border Conflict, Binary Variable*

Variables	Coefficients
Intercept	24.0225
Govt Type (Autocratic)	-3.0193
Ethnic Diversity	-22.7888
Border Conflict Score, Binary	0.8307

Model Significance, $p > \chi^2$: 0.0289

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	46	4
	as NC	1	1
		47	**5**

Accuracy	Sensitivity	Specificity
90.38	0.979	0.200

AUC: 0.909 N: 52

Ĉ by Bins: 5.10 Ĉ by 10ths: 4.81

$\chi^2_{(0.05,11)}$: 19.68 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	42	1
	as NC	0	0
		42	**1**

Accuracy	Sensitivity	Specificity
97.67	1.000	0.000

N: 43

Method: *Force Water per Capita and Border Conflict, Number Variable*

Variables	Coefficients
Intercept	638.0824
Arable Land	-145.1089
Refugee (Asylum) (10,000s)	-0.1238
Fresh Water per Capita (1000s)	85.8006
Govt Type (Autocratic)	-51.5482
Ethnic Diversity	-625.8244
Border Conflict Score, Number	-9.4254

Model Significance, $p > \chi^2$: 0.0014

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	46	2
	as NC	1	3
		47	**5**

Accuracy	Sensitivity	Specificity
94.23	0.979	0.600

AUC: 0.974 N: 52

Ĉ by Bins: 3.20 Ĉ by 10ths: 6.43

$\chi^2_{(0.05,11)}$: 19.68 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	35	1
	as NC	7	0
		42	**1**

Accuracy	Sensitivity	Specificity
81.40	0.833	0.000

N: 43

Figure 10: "In Conflict" Arab and North Africa Models Cont.

Region: **Arab and North Africa**

State: Not In Conflict

Method: *Most Significant*

Variables	Coefficients
Intercept	-4.2497
GDP Per Capitia (1000s)	-0.0902
Infant Mortality Rate	0.1013
Population Density	0.0029
2 Yr Freedom Trend	92.9065
5 Yr Freedom Trend	-81.5652

Model Significance, $p > \chi^2$: 0.0003

Years: 2004-2009

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	5	1
	as NC	4	50
		9	**51**

Accuracy	Sensitivity	Specificity
91.67	0.556	0.980

AUC: 0.917 N: 60

Ĉ by Bins: 12.80 Ĉ by 10ths: 14.31

$\chi^2_{(0.05,13)}$: 22.36 $\chi^2_{(0.05,8)}$: 15.51

Years: 2010-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	0	0
	as NC	5	10
		5	**10**

Accuracy	Sensitivity	Specificity
66.67	0.000	1.000

N: 15

Method: *Force Water per Capita Variable*

Variables	Coefficients
Intercept	178.9987
Arable Land	-81.7842
GDP Per Capitia (1000s)	-1.1670
Military Expend (% Gov Spending)	-0.4891
Infant Mortality rate	4.0372
Fresh Water per Capita (1000s)	-69.2920
Govt Type (Emerging)	-48.8530
Caloric Intake (1000s)	-78.8817
2 Yr Freedom Trend	682.0007
5 Yr Freedom Trend	-983.4996
Regime Type (Central)	23.2102

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2009

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	8	1
	as NC	1	50
		9	**51**

Accuracy	Sensitivity	Specificity
96.67	0.889	0.980

AUC: 0.993 N: 60

Ĉ by Bins: 11.12 Ĉ by 10ths: 10.92

$\chi^2_{(0.05,13)}$: 22.36 $\chi^2_{(0.05,8)}$: 15.51

Years: 2010-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	0	0
	as NC	5	10
		5	**10**

Accuracy	Sensitivity	Specificity
66.67	0.000	1.000

N: 15

Method: *Force Border Conflict, Number Variable*

Variables	Coefficients
Intercept	10.0836
GDP Per Capitia (1000s)	-0.1178
Infant Mortality rate	0.1385
2 Yr Freedom Trend	94.3971
5 Yr Freedom Trend	-82.6127
Religious Diversity	-15.4508
Border Conflict Score, Number	-0.5603

Model Significance, $p > \chi^2$: 0.0011

Years: 2004-2009

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	4	2
	as NC	5	49
		9	**51**

Accuracy	Sensitivity	Specificity
88.33	0.444	0.961

AUC: 0.928 N: 60

Ĉ by Bins: 23.79 Ĉ by 10ths: 5.62

$\chi^2_{(0.05,13)}$: 22.36 $\chi^2_{(0.05,8)}$: 15.51

Years: 2010-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	0	0
	as NC	5	10
		5	**10**

Accuracy	Sensitivity	Specificity
66.67	0.000	1.000

N: 15

Figure 11: "Not In Conflict" Arab and North Africa Models

Region: **Arab and North Africa**
State: Not In Conflict

Method: *Force Border Conflict, Binary Variable*

Variables	Coefficients
Intercept	-3.0837
Infant Mortality rate	0.1469
Border Conflict Score, Binary	-1.6929

Model Significance, $p > \chi^2$: 0.0042

Years: 2004-2009

Model			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	1	1
	as NC	8	50
		9	**51**

Accuracy	Sensitivity	Specificity
85.00	0.111	0.980

AUC: 0.800 N: 60

Ĉ by Bins: 10.65 Ĉ by 10ths: 6.45

$\chi^2_{(0.05,13)}$: 22.36 $\chi^2_{(0.05,8)}$: 15.51

Years: 2010-2013

Validation			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	0	0
	as NC	5	10
		5	**10**

Accuracy	Sensitivity	Specificity
66.67	0.000	1.000

N: 15

Method: *Force Water per Capita and Border Conflict, Number Variable*

Variables	Coefficients
Intercept	-5.8973
Infant Mortality Rate	0.2224
Fresh Water per Capita (1000s)	-2.8646
2 Yr Freedom Trend	82.3974
5 Yr Freedom Trend	-61.0751
Border Conflict Score, Number	0.0843

Model Significance, $p > \chi^2$: 0.0022

Years: 2004-2009

Model			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	4	2
	as NC	5	49
		9	**51**

Accuracy	Sensitivity	Specificity
88.33	0.444	0.961

AUC: 0.882 N: 60

Ĉ by Bins: 23.01 Ĉ by 10ths: 6.67

$\chi^2_{(0.05,13)}$: 22.36 $\chi^2_{(0.05,8)}$: 15.51

Years: 2010-2013

Validation			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	0	0
	as NC	5	10
		5	**10**

Accuracy	Sensitivity	Specificity
66.67	0.000	1.000

N: 15

Figure 11: "Not In Conflict" Arab and North Africa Models Cont.

Region: **Eastern Europe and Central Asia**
State: In Conflict

Method: *Most Significant*

Variables	Coefficients
Intercept	32.3602
Trade (% GDP)	-0.7133
Govt Type (Autocratic)	-23.2371
Ethnic Diversity	-22.4941
Religious Diversity	37.4880
2 Yr Conflict Intensity Trend	15.0450

Model Significance, $p > \chi^2$: 0.0000

Years: 2006-2011 — Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	53	2
	as NC	1	11
		54	**13**

Accuracy	Sensitivity	Specificity
95.52	0.981	0.846

AUC: 0.989 N: 67

Ĉ by Bins: 5.94 Ĉ by 10ths: 9.76

$\chi^2_{(0.05,9)}$: 16.92 $\chi^2_{(0.05,8)}$: 15.51

Years: 2012-2013 — Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	20	3
	as NC	3	3
		23	**6**

Accuracy	Sensitivity	Specificity
79.31	0.870	0.500

N: 29

Method: *Force Water per Capita Variable*

Variables	Coefficients
Intercept	13.0103
Fertility Rate	1.0033
Fresh Water per Capita (1000s)	0.1737
Trade (% GDP)	-0.4169
Govt Type (Autocratic)	-12.1684
Religious Diversity	13.7481

Model Significance, $p > \chi^2$: 0.0000

Years: 2006-2011 — Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	53	3
	as NC	1	10
		54	**13**

Accuracy	Sensitivity	Specificity
94.03	0.981	0.769

AUC: 0.973 N: 67

Ĉ by Bins: 2.03 Ĉ by 10ths: 3.20

$\chi^2_{(0.05,9)}$: 16.92 $\chi^2_{(0.05,8)}$: 15.51

Years: 2012-2013 — Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	20	3
	as NC	3	3
		23	**6**

Accuracy	Sensitivity	Specificity
79.31	0.870	0.500

N: 29

Method: *Force Border Conflict, Number Variable*

Variables	Coefficients
Intercept	173.3314
Birth Rate	-0.7613
Improved Water	-0.8975
Refugee (Asylum) (10,000s)	0.6445
Trade (% GDP)	-1.5186
Govt Type (Autocratic)	-51.5249
Border Conflict Score, Number	21.6462

Model Significance, $p > \chi^2$: 0.0000

Years: 2006-2011 — Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	53	1
	as NC	1	12
		54	**13**

Accuracy	Sensitivity	Specificity
97.01	0.981	0.923

AUC: 0.996 N: 67

Ĉ by Bins: 17.34 Ĉ by 10ths: 2.89

$\chi^2_{(0.05,9)}$: 16.92 $\chi^2_{(0.05,8)}$: 15.51

Years: 2012-2013 — Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	18	3
	as NC	5	3
		23	**6**

Accuracy	Sensitivity	Specificity
72.41	0.783	0.500

N: 29

Figure 12: "In Conflict" Eastern Europe and Central Asia Models

Region: **Eastern Europe and Central Asia**
State: In Conflict

Method: *Force Border Conflict, Binary Variable*

Variables	Coefficients
Intercept	7021.3510
Refugee (Origin) (10,000s)	-32.7947
Trade (% GDP)	-0.3841
Govt Type (Autocratic)	-12.7005
Religious Diversity	17.9020
Border Conflict Score, Binary	-1.7072

Model Significance, $p > \chi^2$: 0.0000

Years: 2006-2011

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	52	0
	as NC	2	13
		54	**13**

Accuracy	Sensitivity	Specificity
97.01	0.963	1.000

AUC: 1.000 N: 67

Ĉ by Bins: 1.62 Ĉ by 10ths: 4.35

$\chi^2_{(0.05,9)}$: 16.92 $\chi^2_{(0.05,8)}$: 15.51

Years: 2012-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	17	3
	as NC	6	3
		23	**6**

Accuracy	Sensitivity	Specificity
68.97	0.739	0.500

N: 29

Method: *Force Water per Capita and Border Conflict, Number Variable*

Variables	Coefficients
Intercept	25.5861
Fresh Water per Capita (1000s)	0.2192
Trade (% GDP)	-0.6175
Govt Type (Autocratic)	-19.9635
Ethnic Diversity	-18.0347
Religious Diversity	31.2693
Border Conflict Score, Number	3.8213

Model Significance, $p > \chi^2$: 0.0000

Years: 2006-2011

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	53	1
	as NC	1	12
		54	**13**

Accuracy	Sensitivity	Specificity
97.01	0.981	0.923

AUC: 0.984 N: 67

Ĉ by Bins: 8.30 Ĉ by 10ths: 4.71

$\chi^2_{(0.05,9)}$: 16.92 $\chi^2_{(0.05,8)}$: 15.51

Years: 2012-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	20	3
	as NC	3	3
		23	**6**

Accuracy	Sensitivity	Specificity
79.31	0.870	0.500

N: 29

Figure 12: "In Conflict" Eastern Europe and Central Asia Models Cont.

Region: **Eastern Europe and Central Asia**
State: Not In Conflict

Method: *Most Significant*

Variables	Coefficients
Intercept	-0.9651
Govt Type (Democratic)	-4.4188
Regime Type (Democratic)	1.6582

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	6	3
	as NC	10	104
		16	**107**

Accuracy	Sensitivity	Specificity
89.43	0.375	0.972

AUC: 0.868 N: 123

Ĉ by Bins: 5.89 Ĉ by 10ths: 5.65

$\chi^2_{(0.05,9)}$: 16.92 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	0	0
	as NC	7	36
		7	**36**

Accuracy	Sensitivity	Specificity
83.72	0.000	1.000

N: 43

Method: *Force Water per Capita Variable*

Variables	Coefficients
Intercept	2.1063
Fresh Water per Capita (1000s)	-0.0911
Govt Type (Democratic)	-5.5198
5 Yr Freedom Trend	18.2396
Regime Type (Central)	-3.6212

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	8	4
	as NC	8	103
		16	**107**

Accuracy	Sensitivity	Specificity
90.24	0.500	0.963

AUC: 0.916 N: 123

Ĉ by Bins: 15.59 Ĉ by 10ths: 15.14

$\chi^2_{(0.05,9)}$: 16.92 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	0	2
	as NC	7	34
		7	**36**

Accuracy	Sensitivity	Specificity
79.07	0.000	0.944

N: 43

Method: *Force Border Conflict, Number Variable*

Variables	Coefficients
Intercept	-0.8254
Govt Type (Democratic)	-4.9586
3 Yr Freedom Trend	13.6371
Regime Type (Democratic)	2.2636
Border Conflict Score, Number	-0.5673

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	7	2
	as NC	9	105
		16	**107**

Accuracy	Sensitivity	Specificity
91.06	0.438	0.981

AUC: 0.876 N: 123

Ĉ by Bins: 6.95 Ĉ by 10ths: 18.99

$\chi^2_{(0.05,9)}$: 16.92 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	0	0
	as NC	7	36
		7	**36**

Accuracy	Sensitivity	Specificity
83.72	0.000	1.000

N: 43

Figure 13: "Not In Conflict" Eastern Europe and Central Asia Models

Region: **Eastern Europe and Central Asia**
State: Not In Conflict

Method: *Force Border Conflict, Binary Variable*

Variables	Coefficients
Intercept	1.5456
GDP Per Capitia (1000s)	-0.2474
Govt Type (Democratic)	-2.5771
Border Conflict Score, Binary	-1.4321

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	2	1
	as NC	14	106
		16	**107**

Accuracy	Sensitivity	Specificity
87.80	0.125	0.991

AUC: 0.895 N: 123

Ĉ by Bins: 4.26 Ĉ by 10ths: 9.60

$\chi^2_{(0.05,9)}$: 16.92 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	0	0
	as NC	7	36
		7	**36**

Accuracy	Sensitivity	Specificity
83.72	0.000	1.000

N: 43

Method: *Force Water per Capita and Border Conflict, Number Variable*

Variables	Coefficients
Intercept	1.9347
Fresh Water per Capita (1000s)	-0.0848
Govt Type (Democratic)	-5.4762
5 Yr Freedom Trend	18.6893
Regime Type (Central)	-3.6136
Border Conflict Score, Number	0.2839

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	7	3
	as NC	9	104
		16	**107**

Accuracy	Sensitivity	Specificity
90.24	0.438	0.972

AUC: 0.911 N: 123

Ĉ by Bins: 20.09 Ĉ by 10ths: 11.74

$\chi^2_{(0.05,9)}$: 16.92 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	0	2
	as NC	7	34
		7	**36**

Accuracy	Sensitivity	Specificity
79.07	0.000	0.944

N: 43

Figure 13: "Not In Conflict" Eastern Europe and Central Asia Models Cont.

Region: **Latin America**
State: In Conflict

Method: *Most Significant*

Variables	Coefficients
Intercept	61.3387
Population Density	-0.0164
Refugee (Asylum) (10,000s)	-0.1674
Trade (% GDP)	-0.1417
Religious Diversity	-57.5274

Model Significance, $p > \chi^2$: 0.0010

Years: 2008-2011

Model			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	29	2
	as NC	1	5
		30	**7**

Accuracy	Sensitivity	Specificity
91.89	0.967	0.714

AUC: 0.938 N: 37

Ĉ by Bins: 5.47 Ĉ by 10ths: 10.86

$\chi^2_{(0.05,10)}$: 18.31 $\chi^2_{(0.05,8)}$: 15.51

Years: 2012-2013

Validation			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	22	2
	as NC	6	0
		28	**2**

Accuracy	Sensitivity	Specificity
73.33	0.786	0.000

N: 30

Method: *Force Water per Capita Variable*

Variables	Coefficients
Intercept	17.6928
Fresh Water per Capita (1000s)	0.0606
Religious Diversity	-20.0089

Model Significance, $p > \chi^2$: 0.0070

Years: 2008-2011

Model			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	30	5
	as NC	0	2
		30	**7**

Accuracy	Sensitivity	Specificity
86.49	1.000	0.286

AUC: 0.836 N: 37

Ĉ by Bins: 9.51 Ĉ by 10ths: 2.38

$\chi^2_{(0.05,10)}$: 18.31 $\chi^2_{(0.05,8)}$: 15.51

Years: 2012-2013

Validation			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	26	1
	as NC	2	1
		28	**2**

Accuracy	Sensitivity	Specificity
90.00	0.929	0.500

N: 30

Method: *Force Border Conflict, Number Variable*

Variables	Coefficients
Intercept	-4.6445
Population Growth	-7.1681
Refugee (Asylum) (10,000s)	-1.3934
Govt Type (Emerging)	17.0671
Caloric Intake (1000s)	10.4985
Border Conflict Score, Number	-10.1512

Model Significance, $p > \chi^2$: 0.0008

Years: 2008-2011

Model			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	29	3
	as NC	1	4
		30	**7**

Accuracy	Sensitivity	Specificity
89.19	0.967	0.571

AUC: 0.957 N: 37

Ĉ by Bins: 6.09 Ĉ by 10ths: 4.78

$\chi^2_{(0.05,10)}$: 18.31 $\chi^2_{(0.05,8)}$: 15.51

Years: 2012-2013

Validation			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	21	1
	as NC	7	1
		28	**2**

Accuracy	Sensitivity	Specificity
73.33	0.750	0.500

N: 30

Figure 14: "In Conflict" Latin America Models

Region: **Latin America**
State: In Conflict

Method: *Force Border Conflict, Binary Variable*

Variables	Coefficients
Intercept	22.4638
Birth Rate	-0.3485
Refugee (Asylum) (10,000s)	-0.2183
Border Conflict Score, Binary	-12.0105

Model Significance, $p > \chi^2$: 0.0082

Years: 2008-2011

Model			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	29	4
	as NC	1	3
		30	**7**

Accuracy	Sensitivity	Specificity
86.49	0.967	0.429

AUC: 0.876 N: 37

Ĉ by Bins: 5.72 Ĉ by 10ths: 4.34

$\chi^2_{(0.05,10)}$: 18.31 $\chi^2_{(0.05,8)}$: 15.51

Years: 2012-2013

Validation			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	25	2
	as NC	3	0
		28	**2**

Accuracy	Sensitivity	Specificity
83.33	0.893	0.000

N: 30

Method: *Force Water per Capita and Border Conflict, Number Variable*

Variables	Coefficients
Intercept	3.1458
Refugee (Asylum) (10,000s)	-0.5097
Fresh Water per Capita (1000s)	0.0859
Govt Type (Emerging)	6.4719
Border Conflict Score, Number	-6.2121

Model Significance, $p > \chi^2$: 0.0026

Years: 2008-2011

Model			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	29	2
	as NC	1	5
		30	**7**

Accuracy	Sensitivity	Specificity
91.89	0.967	0.714

AUC: 0.890 N: 37

Ĉ by Bins: 10.15 Ĉ by 10ths: 1.55

$\chi^2_{(0.05,10)}$: 18.31 $\chi^2_{(0.05,8)}$: 15.51

Years: 2012-2013

Validation			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	20	0
	as NC	8	2
		28	**2**

Accuracy	Sensitivity	Specificity
73.33	0.714	1.000

N: 30

Figure 14: "In Conflict" Latin America Models Cont.

Region: **Latin America**
State: Not In Conflict

Method: *Most Significant*

Variables	Coefficients
Intercept	13.6172
Improved Water	-0.1675
Population Growth	1.5560
Trade (% GDP)	-0.0526

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010 — Model

Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	8	3
	as NC	10	111
		18	**114**

Accuracy	Sensitivity	Specificity
90.15	0.444	0.974

AUC: 0.863 N: 132

Ĉ by Bins: 6.64 Ĉ by 10ths: 11.54

$\chi^2_{(0.05,10)}$: 18.31 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013 — Validation

Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	2	0
	as NC	6	34
		8	**34**

Accuracy	Sensitivity	Specificity
85.71	0.250	1.000

N: 42

Method: *Force Water per Capita Variable*

Variables	Coefficients
Intercept	15.3431
Improved Water	-0.1575
Population Density	-0.0106
Fresh Water per Capita (1000s)	-0.0127
Ethnic Diversity	-7.8153
Religious Diversity	5.2837

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010 — Model

Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	5	0
	as NC	13	114
		18	**114**

Accuracy	Sensitivity	Specificity
90.15	0.278	1.000

AUC: 0.869 N: 132

Ĉ by Bins: 179.11 Ĉ by 10ths: 10.39

$\chi^2_{(0.05,10)}$: 18.31 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013 — Validation

Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	2	0
	as NC	6	34
		8	**34**

Accuracy	Sensitivity	Specificity
85.71	0.250	1.000

N: 42

Method: *Force Border Conflict, Number Variable*

Variables	Coefficients
Intercept	4.3686
Death Rate	-0.9243
Fertility Rate	0.6227
5 Yr Freedom Trend	-18.0551
Ethnic Diversity	-11.6969
Religious Diversity	8.3132
Border Conflict Score, Number	-0.8659

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010 — Model

Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	8	2
	as NC	10	112
		18	**114**

Accuracy	Sensitivity	Specificity
90.91	0.444	0.982

AUC: 0.908 N: 132

Ĉ by Bins: 34.26 Ĉ by 10ths: 10.52

$\chi^2_{(0.05,10)}$: 18.31 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013 — Validation

Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	2	0
	as NC	6	34
		8	**34**

Accuracy	Sensitivity	Specificity
85.71	0.250	1.000

N: 42

Figure 15: "Not In Conflict" Latin America Models

Region: **Latin America**
State: Not In Conflict

Method: *Force Border Conflict, Binary Variable*

Variables	Coefficients
Intercept	3.4247
Death Rate	-0.8796
Fertility Rate	0.7163
5 Yr Freedom Trend	-16.8656
Ethnic Diversity	-10.6658
Religious Diversity	7.6784
Border Conflict Score, Binary	-0.2525

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	9	2
	as NC	9	112
		18	**114**

Accuracy	Sensitivity	Specificity
91.67	0.500	0.982

AUC: 0.907 N: 132

Ĉ by Bins: 23.30 Ĉ by 10ths: 9.58

$\chi^2_{(0.05,10)}$: 18.31 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	2	0
	as NC	6	34
		8	**34**

Accuracy	Sensitivity	Specificity
85.71	0.250	1.000

N: 42

Method: *Force Water per Capita and Border Conflict, Number Variable*

Variables	Coefficients
Intercept	-5.5905
Fertility Rate	1.5510
Population Density	-0.0079
Fresh Water per Capita (1000s)	-0.0132
Border Conflict Score, Number	1.7083

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	3	4
	as NC	15	110
		18	**114**

Accuracy	Sensitivity	Specificity
85.61	0.167	0.965

AUC: 0.856 N: 132

Ĉ by Bins: 10.49 Ĉ by 10ths: 8.97

$\chi^2_{(0.05,10)}$: 18.31 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	1	0
	as NC	7	34
		8	**34**

Accuracy	Sensitivity	Specificity
83.33	0.125	1.000

N: 42

Figure 15: "Not In Conflict" Latin America Models Cont.

Region: **Organization for Economic Cooperation and Development**

State: In Conflict

Method: *Most Significant*

Variables	Coefficients
Intercept	-26.4845
Population Growth (ABS)	3.1556
Trade (% GDP)	0.1233
Unemployment	-0.4582
Caloric Intake (1000s)	5.0356
Religious Diversity	12.9842

Model Significance, $p > \chi^2$: 0.0002

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	43	3
	as NC	2	5
		45	**8**

Accuracy	Sensitivity	Specificity
90.57	0.956	0.625

AUC: 0.964 N: 53

Ĉ by Bins: 2.61 Ĉ by 10ths: 6.81

$\chi^2_{(0.05,11)}$: 19.68 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	17	2
	as NC	2	1
		19	**3**

Accuracy	Sensitivity	Specificity
81.82	0.895	0.333

N: 22

Method: *Force Water per Capita Variable*

Variables	Coefficients
Intercept	-8.5423
Birth Rate	0.6265
Fresh Water per Capita (1000s)	-0.0566
Unemployment	-0.3053
Religious Diversity	7.5955

Model Significance, $p > \chi^2$: 0.0006

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	44	4
	as NC	1	4
		45	**8**

Accuracy	Sensitivity	Specificity
90.57	0.978	0.500

AUC: 0.925 N: 53

Ĉ by Bins: 6.26 Ĉ by 10ths: 7.54

$\chi^2_{(0.05,11)}$: 19.68 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	17	3
	as NC	2	0
		19	**3**

Accuracy	Sensitivity	Specificity
77.27	0.895	0.000

N: 22

Method: *Force Border Conflict, Number Variable*

Variables	Coefficients
Intercept	-2.9112
Population Growth (ABS)	3.5094
Religious Diversity	4.2014
Border Conflict Score, Number	-1.6698

Model Significance, $p > \chi^2$: 0.0010

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	44	5
	as NC	1	3
		45	**8**

Accuracy	Sensitivity	Specificity
88.68	0.978	0.375

AUC: 0.903 N: 53

Ĉ by Bins: 4.63 Ĉ by 10ths: 1.97

$\chi^2_{(0.05,11)}$: 19.68 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	19	2
	as NC	0	1
		19	**3**

Accuracy	Sensitivity	Specificity
90.91	1.000	0.333

N: 22

Figure 16: "In Conflict" OECD Models

Region: **Organization for Economic Cooperation and Development**
State: In Conflict

Method: *Force Border Conflict, Binary Variable*

Variables	Coefficients
Intercept	-14.7184
Military Expend (% GDP)	1.1095
Unemployment	-0.6952
Religious Diversity	18.2488
Border Conflict Score, Binary	6.8951

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	44	3
	as NC	1	5
		45	**8**

Accuracy	Sensitivity	Specificity
92.45	0.978	0.625

AUC: 0.950 N: 53

Ĉ by Bins: 9.54 Ĉ by 10ths: 8.22

$\chi^2_{(0.05,11)}$: 19.68 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	15	1
	as NC	4	2
		19	**3**

Accuracy	Sensitivity	Specificity
77.27	0.789	0.667

N: 22

Method: *Force Water per Capita and Border Conflict, Number Variable*

Variables	Coefficients
Intercept	-24.5627
Military Expend (% GDP)	3.0667
Fresh Water per Capita (1000s)	-0.0308
Unemployment	-1.0610
Religious Diversity	31.4487
Border Conflict Score, Number	9.9899

Model Significance, $p > \chi^2$: 0.0003

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	45	2
	as NC	0	6
		45	**8**

Accuracy	Sensitivity	Specificity
96.23	1.000	0.750

AUC: 0.956 N: 53

Ĉ by Bins: 21.45 Ĉ by 10ths: 5.03

$\chi^2_{(0.05,11)}$: 19.68 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	15	1
	as NC	4	2
		19	**3**

Accuracy	Sensitivity	Specificity
77.27	0.789	0.667

N: 22

Figure 16: "In Conflict" OECD Models Cont.

Region: **Organization for Economic Cooperation and Development**

State: Not In Conflict

Method: *Most Significant*

Variables	Coefficients
Intercept	30.8965
Military Expend (% Gov Spending)	0.2360
Freedom Score	-39.2211
Border Conflict Score	0.7046

Model Significance, $p > \chi^2$: 0.0000

Years: 2005-2009

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	5	1
	as NC	4	116
		9	**117**

Accuracy	Sensitivity	Specificity
96.03	0.556	0.991

AUC: 0.963 N: 126

Ĉ by Bins: 10.04 Ĉ by 10ths: 17.91

$\chi^2_{(0.05,12)}$: 21.03 $\chi^2_{(0.05,8)}$: 15.51

Years: 2010-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	1	3
	as NC	3	94
		4	**97**

Accuracy	Sensitivity	Specificity
94.06	0.250	0.969

N: 101

Method: *Force Water per Capita Variable*

Variables	Coefficients
Intercept	30.5187
Military Expend (% Gov Spending)	0.2732
Fresh Water per Capita (1000s)	0.0176
Freedom Score	-37.9194

Model Significance, $p > \chi^2$: 0.0000

Years: 2005-2009

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	5	3
	as NC	4	109
		9	**112**

Accuracy	Sensitivity	Specificity
94.21	0.556	0.973

AUC: 0.934 N: 121

Ĉ by Bins: 13.44 Ĉ by 10ths: 14.47

$\chi^2_{(0.05,9)}$: 16.92 $\chi^2_{(0.05,8)}$: 15.51

Years: 2010-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	1	3
	as NC	3	90
		4	**93**

Accuracy	Sensitivity	Specificity
93.81	0.250	0.968

N: 97

* Iceland excluded

Method: *Force Border Conflict, Number Variable*

Variables	Coefficients
Intercept	27.0952
Military Expend (% Gov Spending)	0.2711
Freedom Score	-33.9650
Border Conflict Score, Number	0.0780

Model Significance, $p > \chi^2$: 0.0000

Years: 2005-2009

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	5	3
	as NC	4	114
		9	**117**

Accuracy	Sensitivity	Specificity
94.44	0.556	0.974

AUC: 0.937 N: 126

Ĉ by Bins: 8.38 Ĉ by 10ths: 16.46

$\chi^2_{(0.05,12)}$: 21.03 $\chi^2_{(0.05,8)}$: 15.51

Years: 2010-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	1	3
	as NC	3	94
		4	**97**

Accuracy	Sensitivity	Specificity
94.06	0.250	0.969

N: 101

Figure 17: "Not In Conflict" OECD Models

Region: **Organization for Economic Cooperation and Development**

State: Not In Conflict

Method: *Force Border Conflict, Binary Variable*

Variables	Coefficients
Intercept	25.7955
Military Expend (% Gov Spending)	0.2571
Freedom Score	-33.3832
Border Conflict Score, Binary	1.2583

Model Significance, $p > \chi^2$: 0.0000

Years: 2005-2009

Model

Classification Table, Cut Point @ 0.50		Observed as C	Observed as NC
Classified	as C	5	2
	as NC	4	115
		9	**117**

Accuracy	Sensitivity	Specificity
95.24	0.556	0.983

AUC: 0.947 N: 126

Ĉ by Bins: 7.99 Ĉ by 10ths: 14.72

$\chi^2_{(0.05,12)}$: 21.03 $\chi^2_{(0.05,8)}$: 15.51

Years: 2010-2013

Validation

Classification Table, Cut Point @ 0.50		Observed as C	Observed as NC
Classified	as C	1	3
	as NC	3	94
		4	**97**

Accuracy	Sensitivity	Specificity
94.06	0.250	0.969

N: 101

Method: *Force Water per Capita and Border Conflict, Number Variable*

Variables	Coefficients
Intercept	30.3497
Military Expend (% Gov Spending)	0.2718
Fresh Water per Capita (1000s)	0.0177
Freedom Score	-37.7722
Border Conflict Score, Number	0.1230

Model Significance, $p > \chi^2$: 0.0000

Years: 2005-2009

Model

Classification Table, Cut Point @ 0.50		Observed as C	Observed as NC
Classified	as C	5	3
	as NC	4	109
		9	**112**

Accuracy	Sensitivity	Specificity
94.21	0.556	0.973

AUC: 0.928 N: 121

Ĉ by Bins: 13.34 Ĉ by 10ths: 14.47

$\chi^2_{(0.05,9)}$: 16.92 $\chi^2_{(0.05,8)}$: 15.51

Years: 2010-2013

Validation

Classification Table, Cut Point @ 0.50		Observed as C	Observed as NC
Classified	as C	1	3
	as NC	3	90
		4	**93**

Accuracy	Sensitivity	Specificity
93.81	0.250	0.968

N: 97

* Iceland excluded

Figure 17: "Not In Conflict" OECD Models Cont.

Region: **South and East Asia**
State: In Conflict

Method: *Most Significant*

Variables	Coefficients
Intercept	-41.6586
Death Rate	4.9345
Military Expend (% Gov Spending)	-0.7810
Fresh Water per Capita (1000s)	-0.1636
Polity IV	0.5270
Caloric Intake (1000s)	9.3287
5 Yr Freedom Trend	-25.3828

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	65	4
	as NC	1	9
		66	**13**

Accuracy	Sensitivity	Specificity
93.67	0.985	0.692

AUC: 0.936 N: 79

Ĉ by Bins: 13.96 Ĉ by 10ths: 11.77
$\chi^2_{(0.05,11)}$: 19.68 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	29	5
	as NC	9	1
		38	**6**

Accuracy	Sensitivity	Specificity
68.18	0.763	0.167

N: 44

Method: *Force Water per Capita Variable*

Variables	Coefficients
Intercept	-41.6586
Death Rate	4.9345
Military Expend (% Gov Spending)	-0.7810
Fresh Water per Capita (1000s)	-0.1636
Polity IV	0.5270
Caloric Intake (1000s)	9.3287
5 Yr Freedom Trend	-25.3828

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	65	4
	as NC	1	9
		66	**13**

Accuracy	Sensitivity	Specificity
93.67	0.985	0.692

AUC: 0.936 N: 79

Ĉ by Bins: 13.96 Ĉ by 10ths: 11.77
$\chi^2_{(0.05,11)}$: 19.68 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	29	5
	as NC	9	1
		38	**6**

Accuracy	Sensitivity	Specificity
68.18	0.763	0.167

N: 44

Method: *Force Border Conflict, Number Variable*

Variables	Coefficients
Intercept	4.6618
Death Rate	1.9212
Military Expend (% Gov Spending)	-0.6242
Population Growth	-5.8052
Polity IV	0.4849
5 Yr Freedom Trend	-22.4267
Border Conflict Score, Number	-0.4732

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	63	5
	as NC	3	8
		66	**13**

Accuracy	Sensitivity	Specificity
89.87	0.955	0.615

AUC: 0.920 N: 79

Ĉ by Bins: 12.87 Ĉ by 10ths: 9.76
$\chi^2_{(0.05,11)}$: 19.68 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation			
Classification Table, Cut Point @ 0.50			
		Observed	
		as C	as NC
Classified	as C	32	4
	as NC	6	2
		38	**6**

Accuracy	Sensitivity	Specificity
77.27	0.842	0.333

N: 44

Figure 18: "In Conflict" South and East Asia Models

Region: **South and East Asia**
State: In Conflict

Method: *Force Border Conflict, Binary Variable*

Variables	Coefficients
Intercept	-45.5357
Death Rate	5.4174
Military Expend (% Gov Spending)	-1.0509
Fresh Water per Capita (1000s)	-0.2044
Polity IV	0.6345
Caloric Intake (1000s)	12.0018
5 Yr Freedom Trend	-34.4032
Border Conflict Score, Binary	-1.8218

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	65	4
	as NC	1	9
		66	**13**

Accuracy	Sensitivity	Specificity
93.67	0.985	0.692

AUC: 0.942 N: 79

Ĉ by Bins: 8.15 Ĉ by 10ths: 10.61

$\chi^2_{(0.05,11)}$: 19.68 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	29	6
	as NC	9	0
		38	**6**

Accuracy	Sensitivity	Specificity
65.91	0.763	0.000

N: 44

Method: *Force Water per Capita and Border Conflict, Number Variable*

Variables	Coefficients
Intercept	7.3780
Death Rate	1.7978
Military Expend (% GDP)	-3.3378
Population Growth	-6.2622
Fresh Water per Capita (1000s)	-0.0231
Polity IV	0.6886
5 Yr Freedom Trend	-19.7886
Border Conflict Score, Number	-4.9130

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	64	4
	as NC	2	9
		66	**13**

Accuracy	Sensitivity	Specificity
92.41	0.970	0.692

AUC: 0.948 N: 79

Ĉ by Bins: 13.99 Ĉ by 10ths: 10.73

$\chi^2_{(0.05,11)}$: 19.68 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	30	4
	as NC	8	2
		38	**6**

Accuracy	Sensitivity	Specificity
72.73	0.789	0.333

N: 44

Figure 18: "In Conflict" South and East Asia Models Cont.

Region: **South and East Asia**
State: Not In Conflict

Method: *Most Significant*

Variables	Coefficients
Intercept	5.2966
Arable Land	-12.5697
GDP Per Capitia (1000s)	-0.1811
Fresh Water per Capita (1000s)	-0.0354
Unemployment	-0.3378
Freedom Score	-2.9531

Model Significance, $p > \chi^2$: 0.0003

Years: 2008-2011

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	6	1
	as NC	6	53
		12	**54**

Accuracy	Sensitivity	Specificity
89.39	0.500	0.981

AUC: 0.906 N: 66

Ĉ by Bins: 7.28 Ĉ by 10ths: 8.48

$\chi^2_{(0.05,9)}$: 16.92 $\chi^2_{(0.05,8)}$: 15.51

Years: 2012-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	1	1
	as NC	1	22
		2	**23**

Accuracy	Sensitivity	Specificity
92.00	0.500	0.957

N: 25

Method: *Force Water per Capita Variable*

Variables	Coefficients
Intercept	15.4872
Arable Land	-21.4787
GDP Per Capitia (1000s)	-0.1052
Military Expend (% GDP)	-1.1856
Fresh Water per Capita (1000s)	-0.0651
Unemployment	-0.3814
Caloric Intake (1000s)	-3.1659

Model Significance, $p > \chi^2$: 0.0001

Years: 2008-2011

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	7	1
	as NC	5	53
		12	**54**

Accuracy	Sensitivity	Specificity
90.91	0.583	0.981

AUC: 0.903 N: 66

Ĉ by Bins: 14.55 Ĉ by 10ths: 7.43

$\chi^2_{(0.05,9)}$: 16.92 $\chi^2_{(0.05,8)}$: 15.51

Years: 2012-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	1	3
	as NC	1	20
		2	**23**

Accuracy	Sensitivity	Specificity
84.00	0.500	0.870

N: 25

Method: *Force Border Conflict, Number Variable*

Variables	Coefficients
Intercept	-9.4977
Life Expectancy	0.3937
Military Expend (% GDP)	-0.9691
Caloric Intake (1000s)	-7.2191
Border Conflict Score, Number	1.4607

Model Significance, $p > \chi^2$: 0.0014

Years: 2008-2011

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	4	1
	as NC	8	53
		12	**54**

Accuracy	Sensitivity	Specificity
86.36	0.333	0.981

AUC: 0.855 N: 66

Ĉ by Bins: 23.57 Ĉ by 10ths: 6.09

$\chi^2_{(0.05,9)}$: 16.92 $\chi^2_{(0.05,8)}$: 15.51

Years: 2012-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	1	1
	as NC	1	22
		2	**23**

Accuracy	Sensitivity	Specificity
92.00	0.500	0.957

N: 25

Figure 19: "Not In Conflict" South and East Asia Models

Region: **South and East Asia**
State: Not In Conflict

Method: *Force Border Conflict, Binary Variable*

Variables	Coefficients
Intercept	-20.9158
Life Expectancy	0.4474
Infant Mortality Rate	0.0702
Caloric Intake (1000s)	-5.7867
Border Conflict Score, Binary	1.5589

Model Significance, $p > \chi^2$: 0.0017

Years: 2008-2011

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	7	3
	as NC	5	51
		12	**54**

Accuracy	Sensitivity	Specificity
87.88	0.583	0.944

AUC: 0.836 N: 66

Ĉ by Bins: 12.51 Ĉ by 10ths: 3.91

$\chi^2_{(0.05,9)}$: 16.92 $\chi^2_{(0.05,8)}$: 15.51

Years: 2012-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	0	2
	as NC	2	21
		2	**23**

Accuracy	Sensitivity	Specificity
84.00	0.000	0.913

N: 25

Method: *Force Water per Capita and Border Conflict, Number Variable*

Variables	Coefficients
Intercept	8.4497
Fresh Water per Capita (1000s)	-0.0186
Caloric Intake (1000s)	-3.3483
Freedom Score	-2.5726
Border Conflict Score, Number	1.2571

Model Significance, $p > \chi^2$: 0.0031

Years: 2008-2011

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	6	2
	as NC	6	52
		12	**54**

Accuracy	Sensitivity	Specificity
87.88	0.500	0.963

AUC: 0.816 N: 66

Ĉ by Bins: 10.79 Ĉ by 10ths: 6.46

$\chi^2_{(0.05,9)}$: 16.92 $\chi^2_{(0.05,8)}$: 15.51

Years: 2012-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	0	2
	as NC	2	21
		2	**23**

Accuracy	Sensitivity	Specificity
84.00	0.000	0.913

N: 25

Figure 19: "Not In Conflict" South and East Asia Models Cont.

Region:	**Sub-Saharan Africa**
State:	In Conflict

Method: *Most Significant*

Variables	Coefficients
Intercept	1.4951
Death Rate	0.1821
Fresh Water per Capita (1000s)	-0.0769
Ethnic Diversity	-2.8332

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	126	20
	as NC	2	6
		128	**26**

Accuracy	Sensitivity	Specificity
85.71	0.984	0.231

AUC: 0.769 N: 154

Ĉ by Bins: 20.59 Ĉ by 10ths: 4.14

$\chi^2_{(0.05,12)}$: 21.03 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	63	10
	as NC	0	1
		63	**11**

Accuracy	Sensitivity	Specificity
86.49	1.000	0.091

N: 74

Method: *Force Water per Capita Variable*

Variables	Coefficients
Intercept	1.4951
Death Rate	0.1821
Fresh Water per Capita (1000s)	-0.0769
Ethnic Diversity	-2.8332

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	126	20
	as NC	2	6
		128	**26**

Accuracy	Sensitivity	Specificity
85.71	0.984	0.231

AUC: 0.769 N: 154

Ĉ by Bins: 20.59 Ĉ by 10ths: 4.14

$\chi^2_{(0.05,12)}$: 21.03 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	63	10
	as NC	0	1
		63	**11**

Accuracy	Sensitivity	Specificity
86.49	1.000	0.091

N: 74

Method: *Force Border Conflict, Number Variable*

Variables	Coefficients
Intercept	1.2237
Death Rate	0.1599
Fresh Water per Capita (1000s)	-0.0765
Ethnic Diversity	-2.6451
Border Conflict Score, Number	1.1443

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	126	20
	as NC	2	6
		128	**26**

Accuracy	Sensitivity	Specificity
85.71	0.984	0.231

AUC: 0.792 N: 154

Ĉ by Bins: 15.11 Ĉ by 10ths: 5.50

$\chi^2_{(0.05,12)}$: 21.03 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	63	10
	as NC	0	1
		63	**11**

Accuracy	Sensitivity	Specificity
86.49	1.000	0.091

N: 74

Figure 20: "In Conflict" Sub-Saharan Africa Models

Region: **Sub-Saharan Africa**
State: In Conflict

Method: *Force Border Conflict, Binary Variable*

Variables	Coefficients
Intercept	0.7474
Death Rate	0.1626
Fresh Water per Capita (1000s)	-0.0728
Ethnic Diversity	-2.4464
Border Conflict Score, Binary	0.9459

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	125	20
	as NC	3	6
		128	**26**

Accuracy	Sensitivity	Specificity
85.06	0.977	0.231

AUC: 0.795 N: 154

Ĉ by Bins: 14.50 Ĉ by 10ths: 6.84

$\chi^2_{(0.05,12)}$: 21.03 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	58	10
	as NC	5	1
		63	**11**

Accuracy	Sensitivity	Specificity
79.73	0.921	0.091

N: 74

Method: *Force Water per Capita and Border Conflict, Number Variable*

Variables	Coefficients
Intercept	1.2237
Death Rate	0.1599
Fresh Water per Capita (1000s)	-0.0765
Ethnic Diversity	-2.6451
Border Conflict Score, Number	1.1443

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	126	20
	as NC	2	6
		128	**26**

Accuracy	Sensitivity	Specificity
85.71	0.984	0.231

AUC: 0.792 N: 154

Ĉ by Bins: 15.11 Ĉ by 10ths: 5.50

$\chi^2_{(0.05,12)}$: 21.03 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	63	10
	as NC	0	1
		63	**11**

Accuracy	Sensitivity	Specificity
86.49	1.000	0.091

N: 74

Figure 20: "In Conflict" Sub-Saharan Africa Models Cont.

Region: **Sub-Saharan Africa**
State: Not In Conflict

Method: *Most Significant*

Variables	Coefficients
Intercept	4.3821
Arable Land	2.9255
Population Growth (ABS)	-0.8120
Trade (% GDP)	-0.0513
Govt Type (Democratic)	1.3220
Freedom Score	-6.5956
Regime Type (Central)	3.2142

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	8	5
	as NC	22	154
		30	**159**

Accuracy	Sensitivity	Specificity
85.71	0.267	0.969

AUC: 0.812 N: 189

Ĉ by Bins: 21.89 Ĉ by 10ths: 13.58

$\chi^2_{(0.05,7)}$: 14.07 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	1	2
	as NC	11	59
		12	**61**

Accuracy	Sensitivity	Specificity
82.19	0.083	0.967

N: 73

Method: *Force Water per Capita Variable*

Variables	Coefficients
Intercept	4.6568
Arable Land	2.9332
Population Growth (ABS)	-0.8459
Fresh Water per Capita (1000s)	-0.0059
Trade (% GDP)	-0.0522
Govt Type (Democratic)	1.2769
Freedom Score	-6.7164
Regime Type (Central)	3.1620

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	8	6
	as NC	22	153
		30	**159**

Accuracy	Sensitivity	Specificity
85.19	0.267	0.962

AUC: 0.812 N: 189

Ĉ by Bins: 19.72 Ĉ by 10ths: 12.58

$\chi^2_{(0.05,7)}$: 14.07 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	1	2
	as NC	11	59
		12	**61**

Accuracy	Sensitivity	Specificity
82.19	0.083	0.967

N: 73

Method: *Force Border Conflict, Number Variable*

Variables	Coefficients
Intercept	2.2684
Trade (% GDP)	-0.0481
Govt Type (Democratic)	1.3481
Freedom Score	-5.3373
Regime Type (Central)	3.3462
Border Conflict Score, Number	0.2650

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	7	3
	as NC	23	156
		30	**159**

Accuracy	Sensitivity	Specificity
86.24	0.233	0.981

AUC: 0.793 N: 189

Ĉ by Bins: 11.13 Ĉ by 10ths: 9.12

$\chi^2_{(0.05,7)}$: 14.07 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	0	3
	as NC	12	58
		12	**61**

Accuracy	Sensitivity	Specificity
79.45	0.000	0.951

N: 73

Figure 21: "Not In Conflict" Sub-Saharan Africa Models

Region: **Sub-Saharan Africa**
State: Not In Conflict

Method: *Force Border Conflict, Binary Variable*

Variables	Coefficients
Intercept	-3.0625
GDP Per Capitia (1000s)	-0.2258
Caloric Intake (1000s)	2.0882
Freedom Score	-6.0008
Regime Type (Emerging)	-1.9817
Border Conflict Score, Binary	0.0560

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	7	2
	as NC	23	157
		30	**159**

Accuracy	Sensitivity	Specificity
86.77	0.233	0.987

AUC: 0.779 N: 189

Ĉ by Bins: 19.65 Ĉ by 10ths: 7.96

$\chi^2_{(0.05,7)}$: 14.07 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	3	6
	as NC	9	55
		12	**61**

Accuracy	Sensitivity	Specificity
79.45	0.250	0.902

N: 73

Method: *Force Water per Capita and Border Conflict, Number Variable*

Variables	Coefficients
Intercept	2.3443
Fresh Water per Capita (1000s)	-0.0033
Trade (% GDP)	-0.0481
Govt Type (Democratic)	1.3217
Freedom Score	-5.3795
Regime Type (Central)	3.2987
Border Conflict Score, Number	0.2724

Model Significance, $p > \chi^2$: 0.0000

Years: 2004-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	7	3
	as NC	23	156
		30	**159**

Accuracy	Sensitivity	Specificity
86.24	0.233	0.981

AUC: 0.792 N: 189

Ĉ by Bins: 11.11 Ĉ by 10ths: 10.04

$\chi^2_{(0.05,7)}$: 14.07 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	1	3
	as NC	11	58
		12	**61**

Accuracy	Sensitivity	Specificity
80.82	0.083	0.951

N: 73

Figure 21: "Not In Conflict" Sub-Saharan Africa Models Cont.

Note: Variables in red have a significance p-value greater than 0.25.

Appendix B: Whole World Conditional Logistic Regression Models

Region: **Whole World**
State: In Conflict

Method: *Whole World Forcing Water & Border Conflict, Number*

Variables	Coefficients
Intercept	3.4755
Fresh Water per Capita (1000s)	-0.0354
Trade (% GDP)	-0.0295
Govt Type (Autocratic)	-1.2470
Border Conflict Score, Number	0.0186

Model Significance, $p > \chi^2$: 0.0002

Model — Years: 2007-2010

Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	237	39
	as NC	2	5
		239	**44**

Accuracy	Sensitivity	Specificity
85.51	0.992	0.114

AUC: 0.700 N: 283

Ĉ by Bins: 5.98 Ĉ by 10ths: 8.16

$\chi^2_{(0.05,4)}$: 9.49 $\chi^2_{(0.05,8)}$: 15.51

Validation — Years: 2011-2013

Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	219	28
	as NC	12	4
		231	**32**

Accuracy	Sensitivity	Specificity
84.79	0.948	0.125

N: 263

Method: *Transitioned Countries Only Forcing Water & Border Conflict, Number*

Variables	Coefficients
Intercept	3.7321
Improved Water	-0.0247
Military Expend (% GDP)	0.2652
Fresh Water per Capita (1000s)	-0.0253
Trade (% GDP)	-0.0244
Govt Type (Autocratic)	-2.0003
Border Conflict Score, Number	-0.6172

Model Significance, $p > \chi^2$: 0.0040

Model — Years: 2007-2010

Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	64	27
	as NC	7	17
		71	**44**

Accuracy	Sensitivity	Specificity
70.43	0.901	0.386

AUC: 0.737 N: 115

Ĉ by Bins: 11.55 Ĉ by 10ths: 5.17

$\chi^2_{(0.05,4)}$: 9.49 $\chi^2_{(0.05,8)}$: 15.51

Validation — Years: 2011-2013

Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	40	18
	as NC	27	14
		67	**32**

Accuracy	Sensitivity	Specificity
54.55	0.597	0.438

N: 99

Figure 22: "In Conflict" Whole World Models

Region: **Whole World**
State: Not In Conflict

Method: *Whole World Forcing Water & Border Conflict, Number*

Variables	Coefficients
Intercept	1.3377
Death Rate	-0.0936
GDP Per Capitia (1000s)	-0.0374
Fresh Water per Capita (1000s)	-0.0056
Trade (% GDP)	-0.0158
Freedom Score	-2.0386
Border Conflict Score, Number	0.0321

Model Significance, $p > \chi^2$: 0.0000

Years: 2007-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	1	6
	as NC	56	382
		57	**388**

Accuracy	Sensitivity	Specificity
86.07	0.018	0.985

AUC: 0.780 N: 445

Ĉ by Bins: 10.27 Ĉ by 10ths: 24.62

$\chi^2_{(0.05,10)}$: 18.31 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	0	4
	as NC	36	243
		36	**247**

Accuracy	Sensitivity	Specificity
85.87	0.000	0.984

N: 283

Method: *Transitioned Countries Only Forcing Water & Border Conflict, Number*

Variables	Coefficients
Intercept	-0.1151
Arable Land	-1.4283
Refugee (Asylum) (10,000s)	0.0088
Fresh Water per Capita (1000s)	0.0029
Trade (% GDP)	-0.0125
Border Conflict Score, Number	0.2631

Model Significance, $p > \chi^2$: 0.0779

Years: 2007-2010

Model — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	3	2
	as NC	54	128
		57	**130**

Accuracy	Sensitivity	Specificity
70.05	0.053	0.985

AUC: 0.639 N: 187

Ĉ by Bins: 13.58 Ĉ by 10ths: 6.26

$\chi^2_{(0.05,9)}$: 16.92 $\chi^2_{(0.05,8)}$: 15.51

Years: 2011-2013

Validation — Classification Table, Cut Point @ 0.50

		Observed as C	Observed as NC
Classified	as C	0	0
	as NC	36	48
		36	**48**

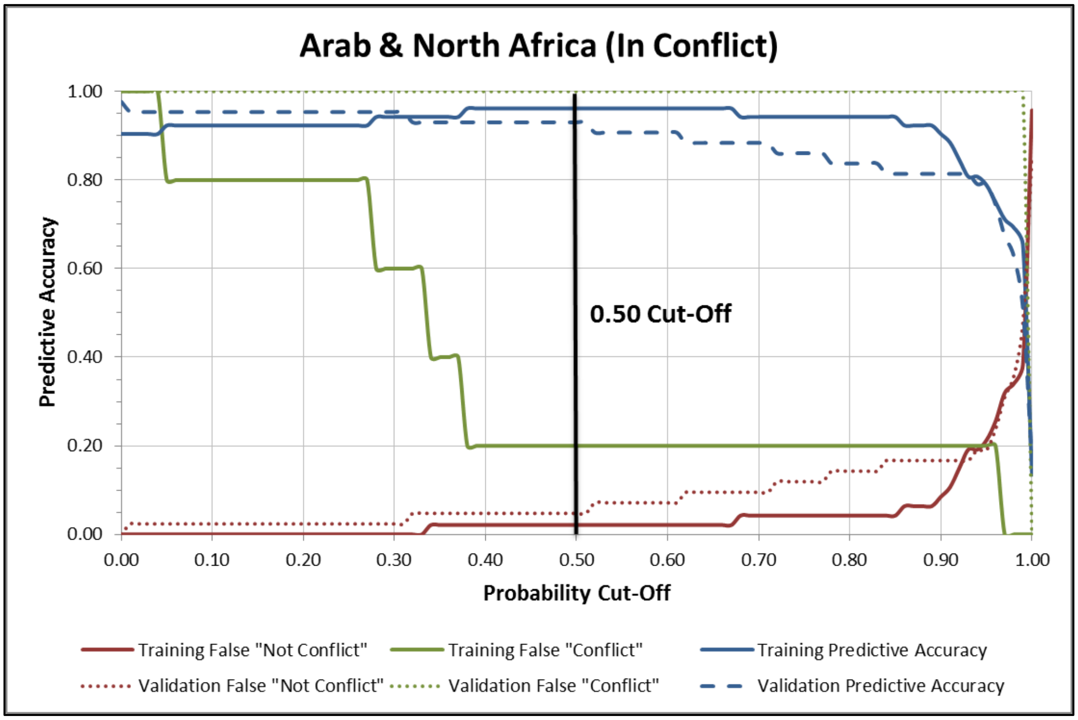
Accuracy	Sensitivity	Specificity
57.14	0.000	1.000

N: 84

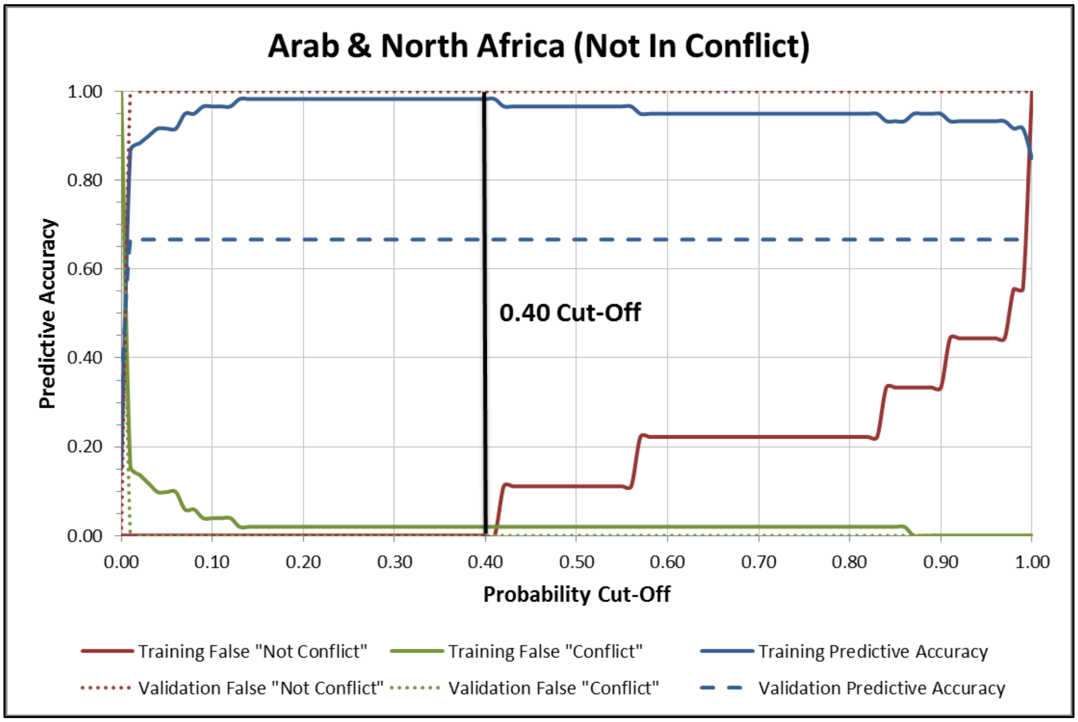
Figure 23: "Not In Conflict" Whole World Models

Note: Variables in red have a significance p-value greater than 0.25.

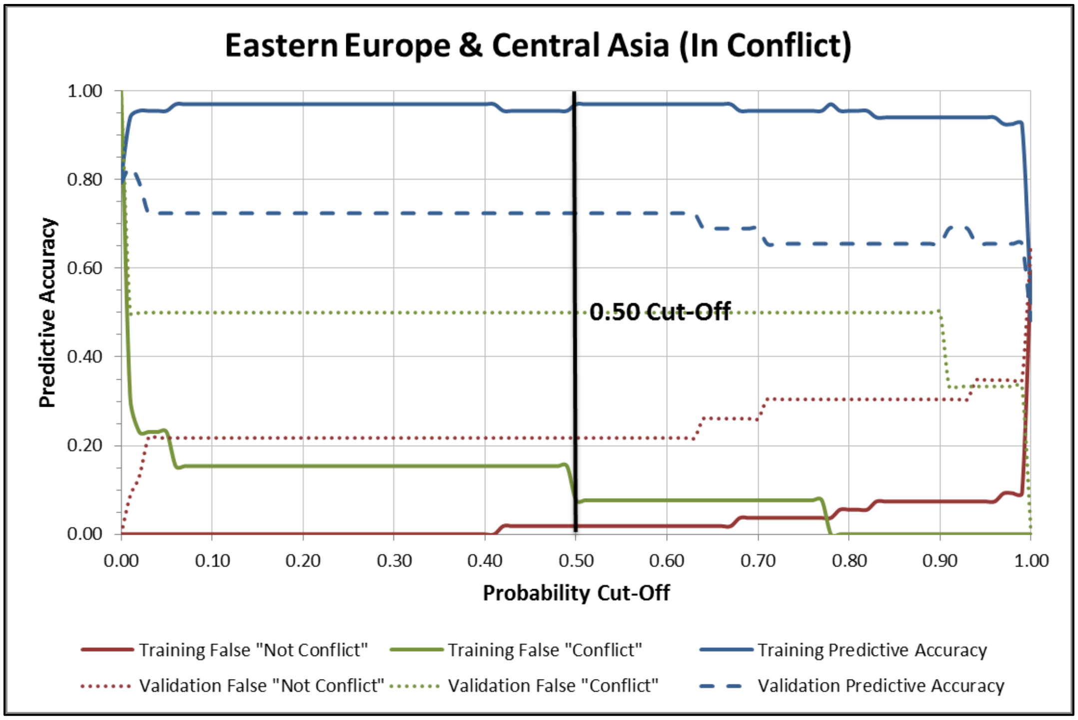
Appendix C: Sensitivity Analysis of Predictive Accuracy vs Probability Cut-Off



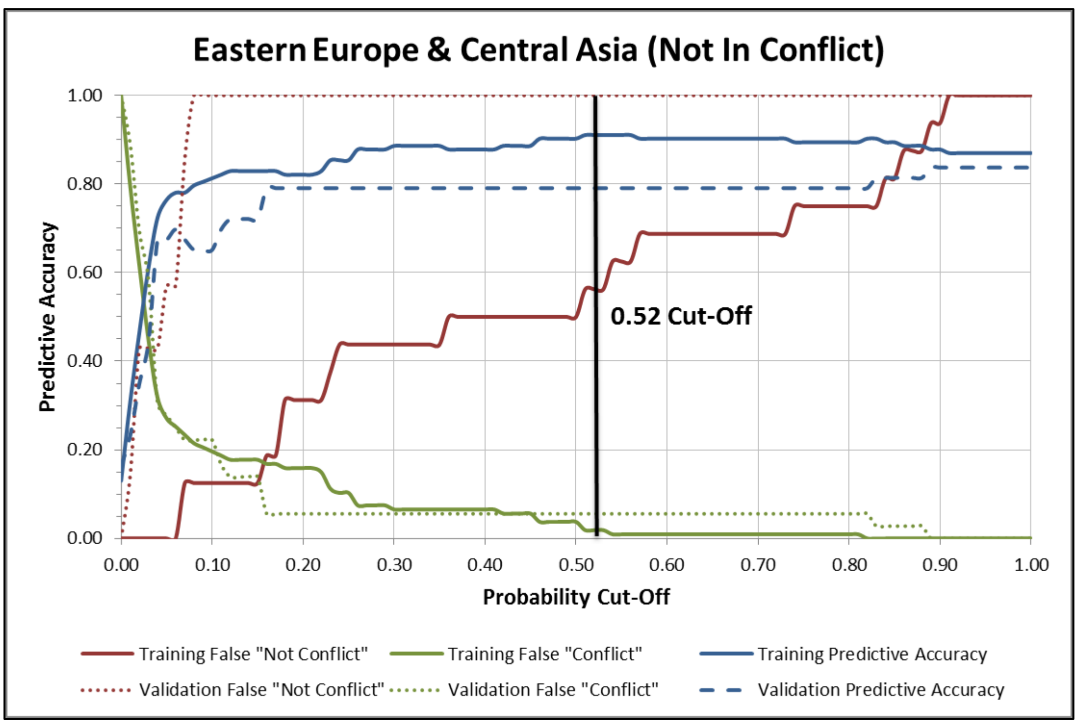
Suite: Most Significant



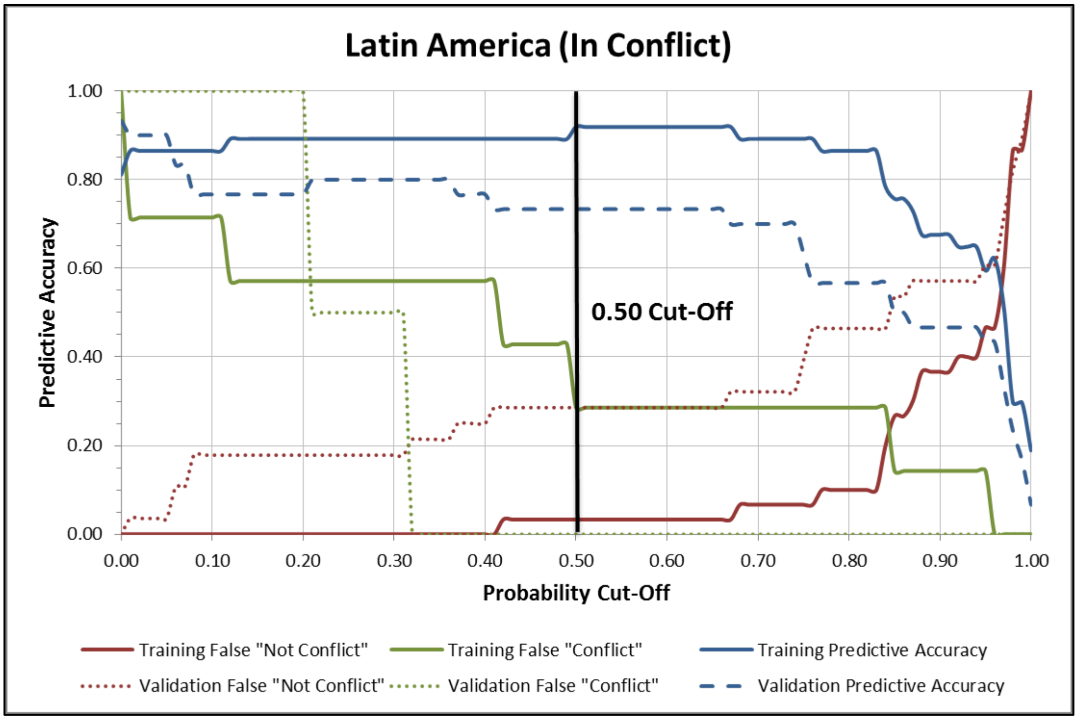
Suite: Forced Water per Capita Variable



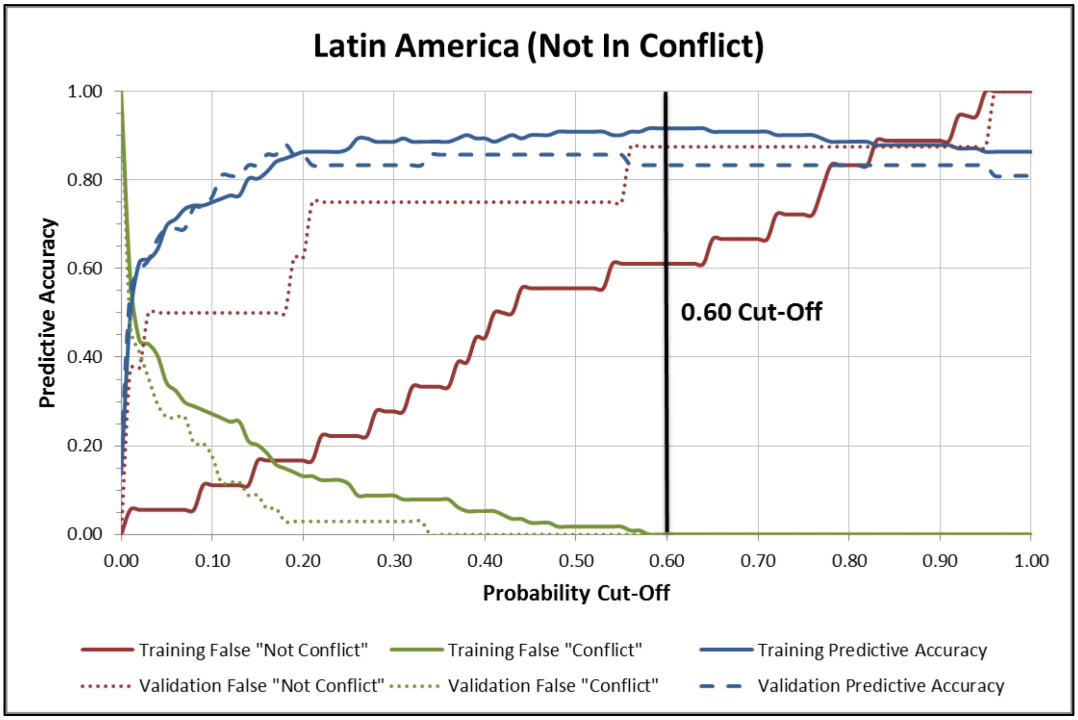
Suite: Forced Border Conflict, Number Variable



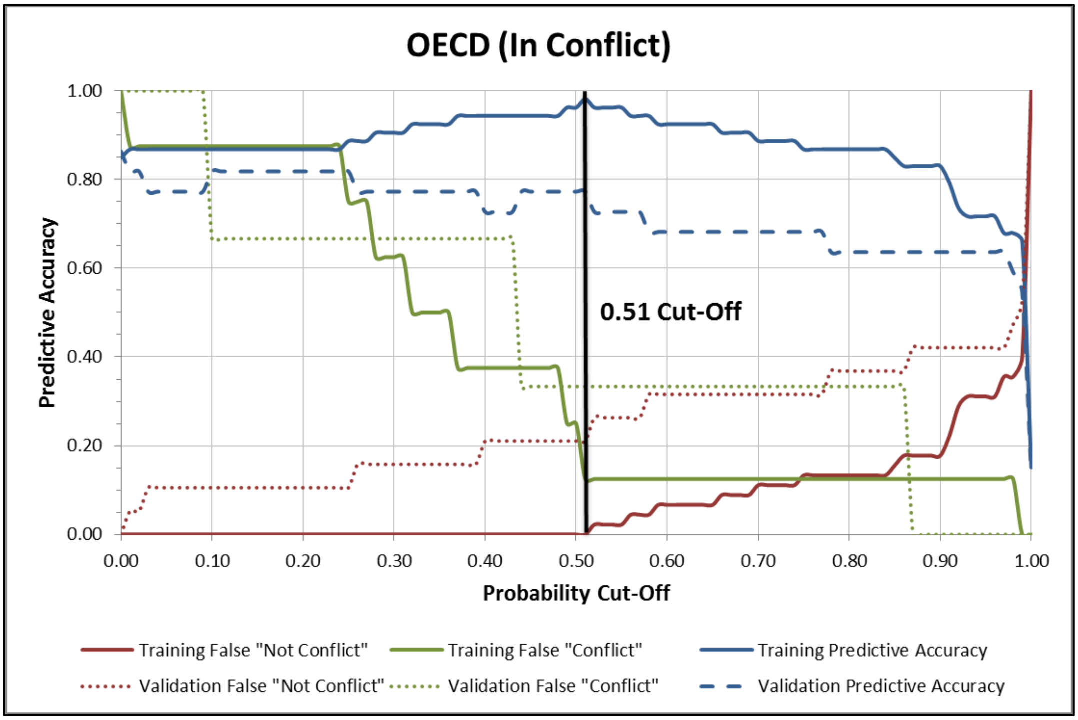
Suite: Forced Water per Capita Variable



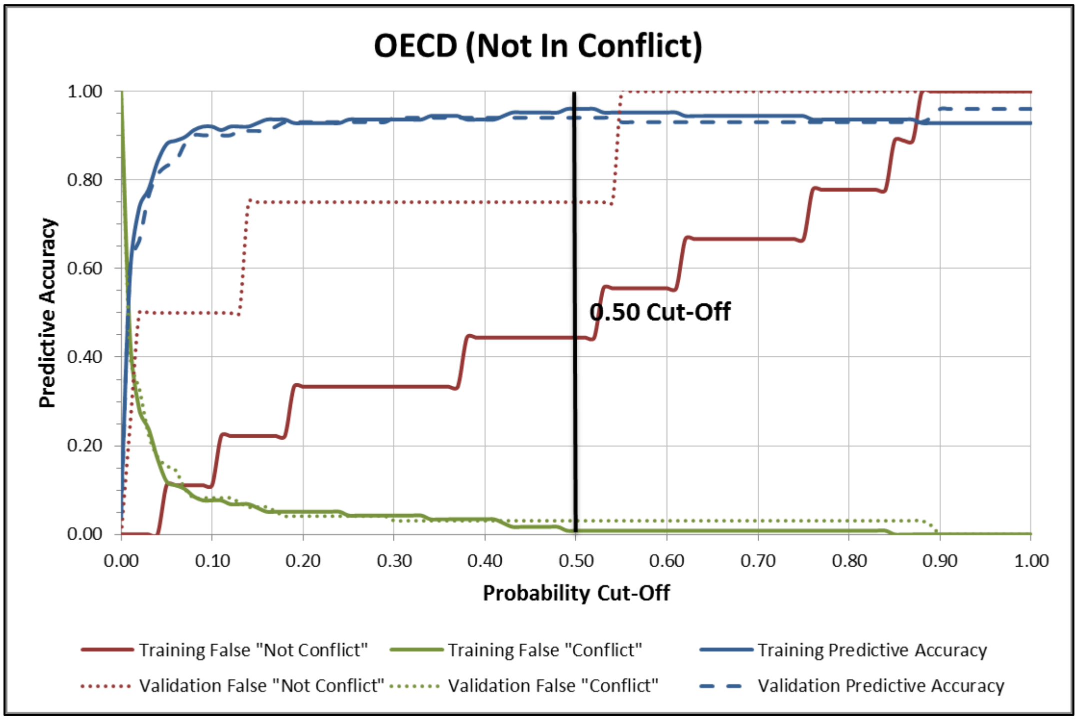
Suite: (Combined) Forced Water Per Capita and Border Conflict, Number Variable



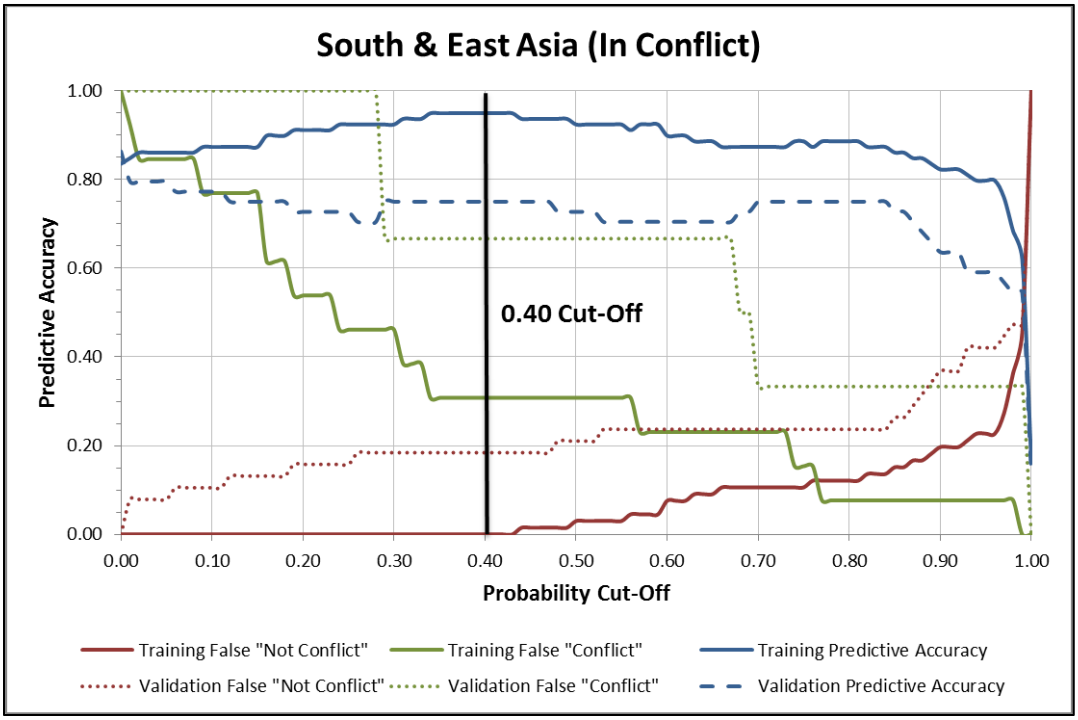
Suite: Forced Border Conflict, Number Variable



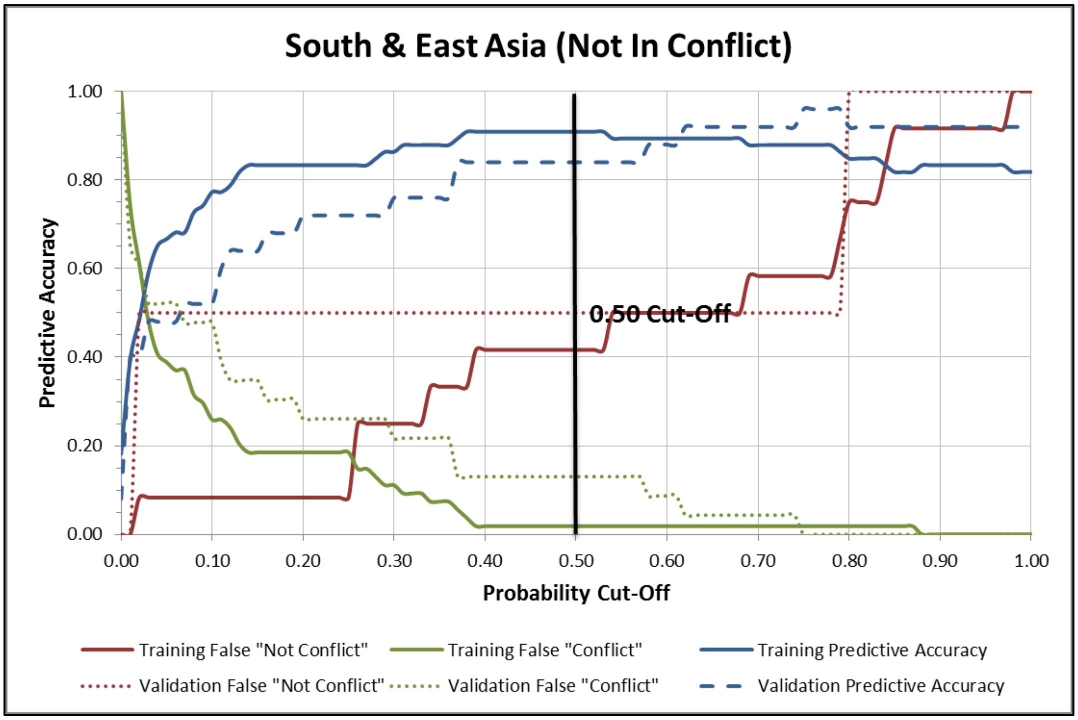
Suite: (Combined) Forced Water Per Capita and Border Conflict, Number Variable



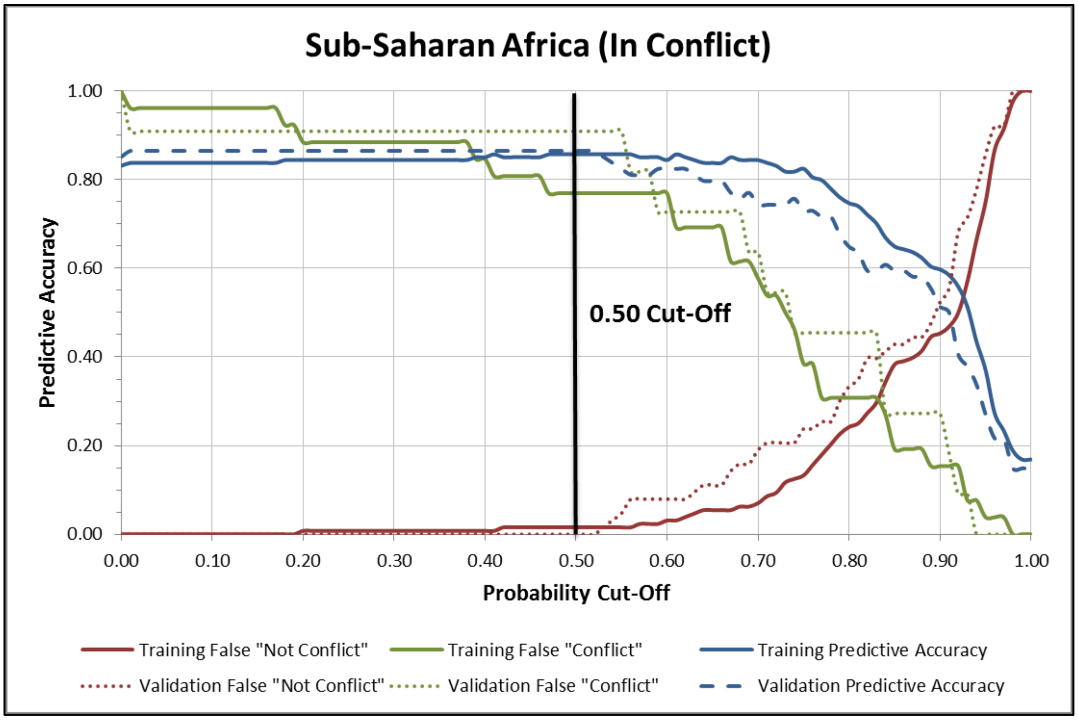
Suite: Most Significant



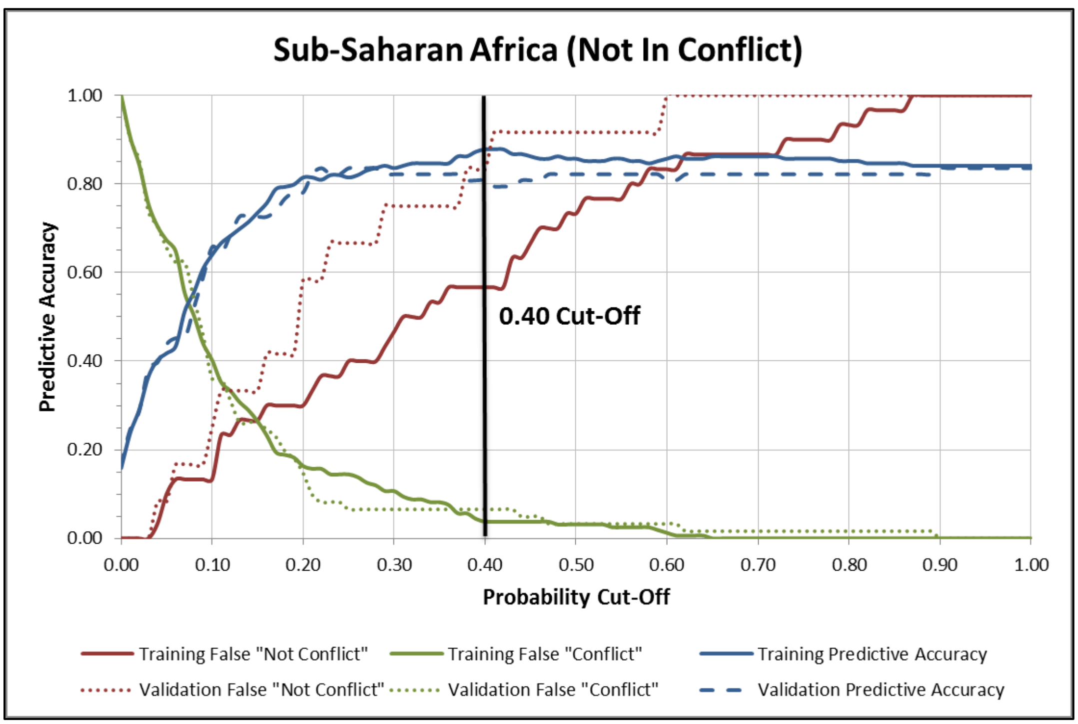
Suite: (Combined) Forced Water Per Capita and Border Conflict, Number Variable



Suite: Forced Water per Capita Variable



Suite: Most Significant



Suite: Most Significant

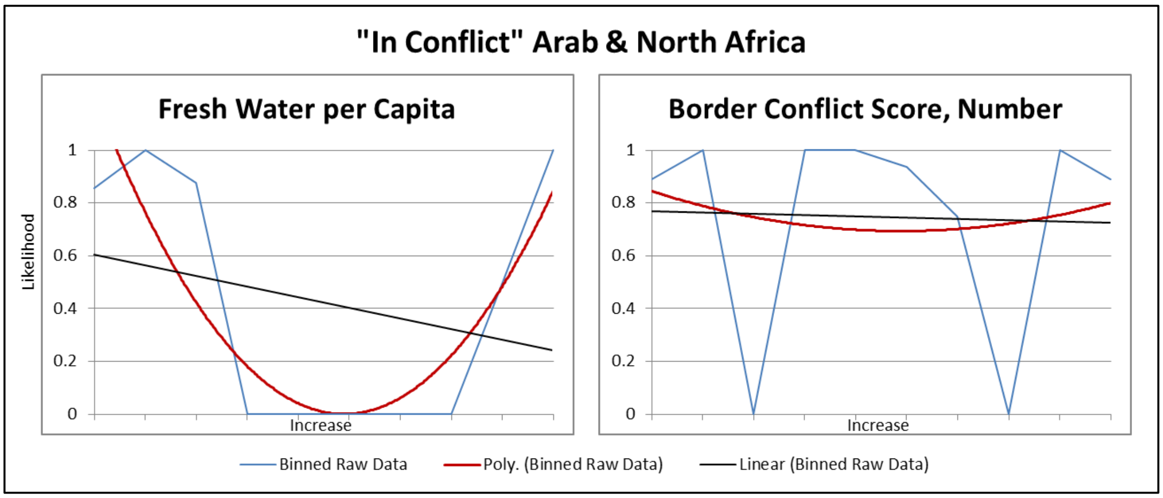
Appendix D: Environmental Factors Univariable Logistic Regression Plots

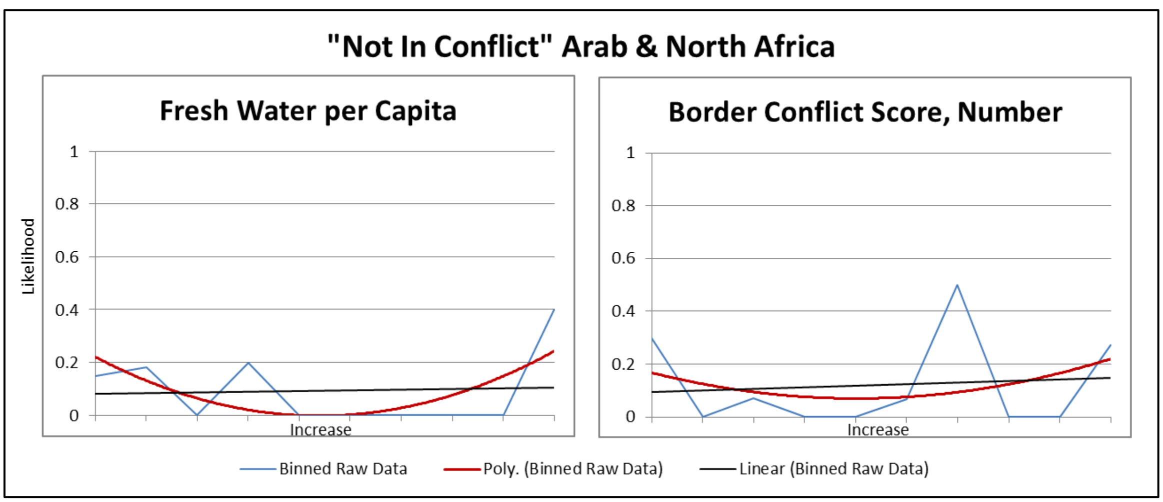
Vertical Axis:

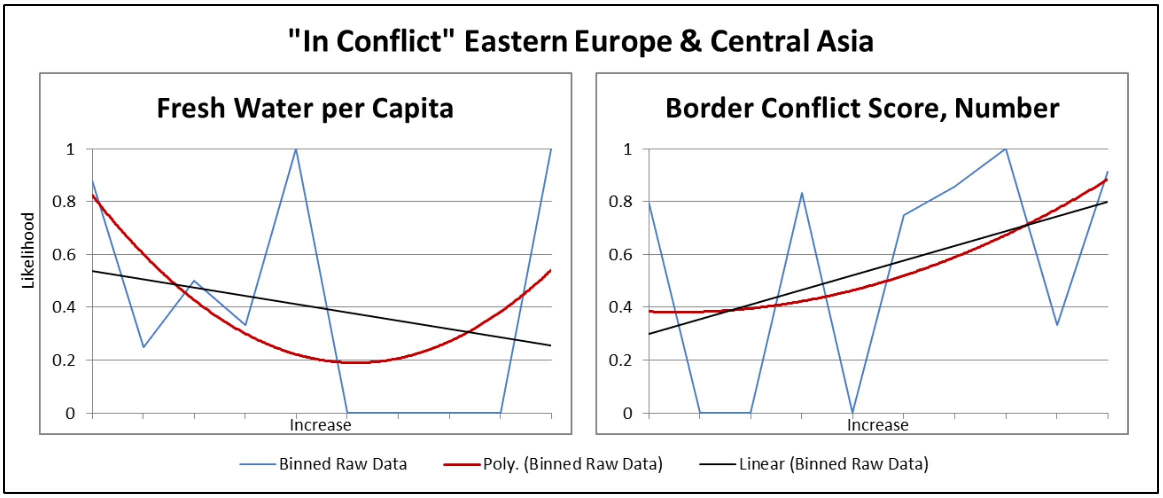
- Likelihood of 0, classified as “Not In Conflict”, and increase to 1, classified as “In Conflict”.

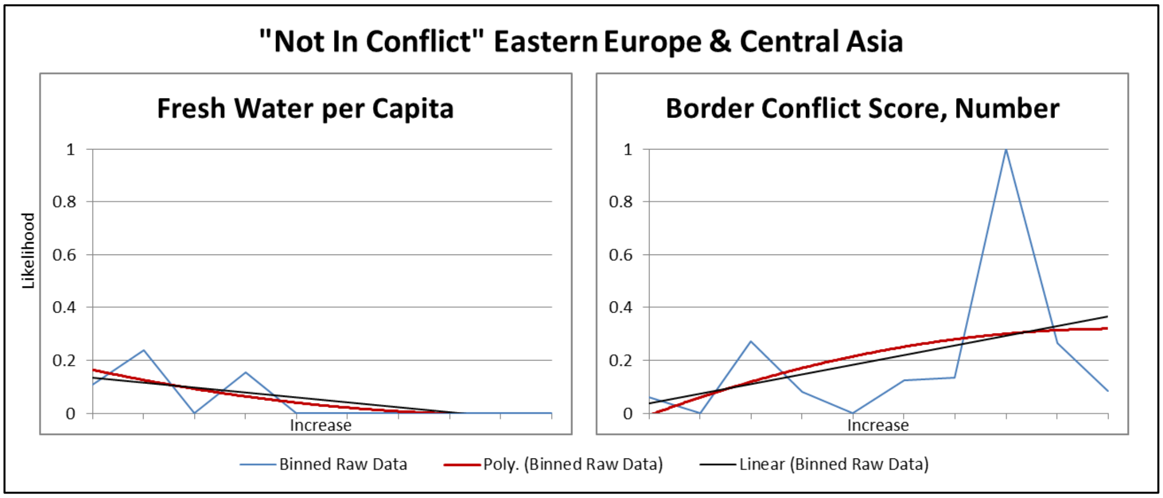
Horizontal Axis:

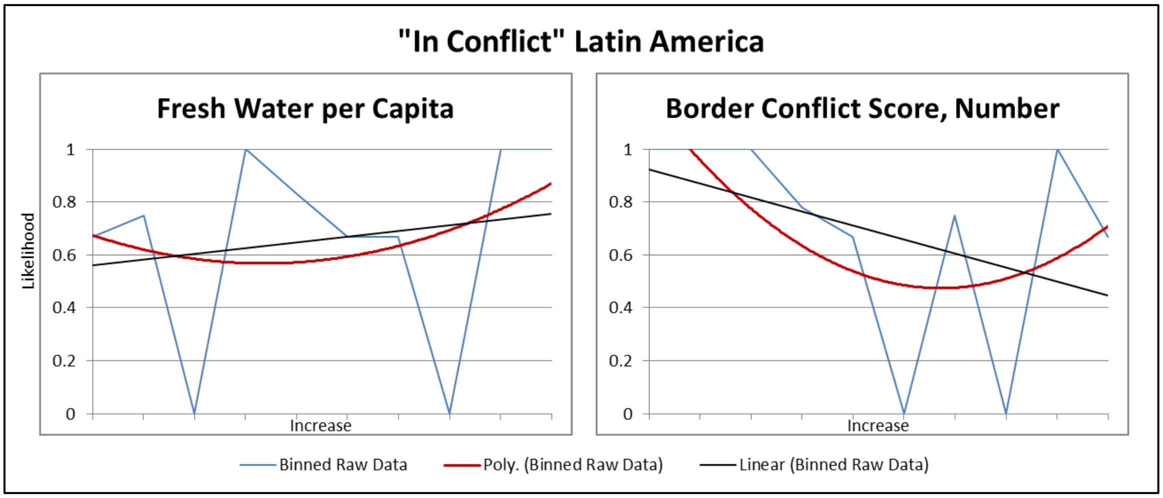
- Fresh Water per Capita increases left to right from the region’s lowest value to its highest
- Border Conflict increases left to right from the region’s lowest value to its highest



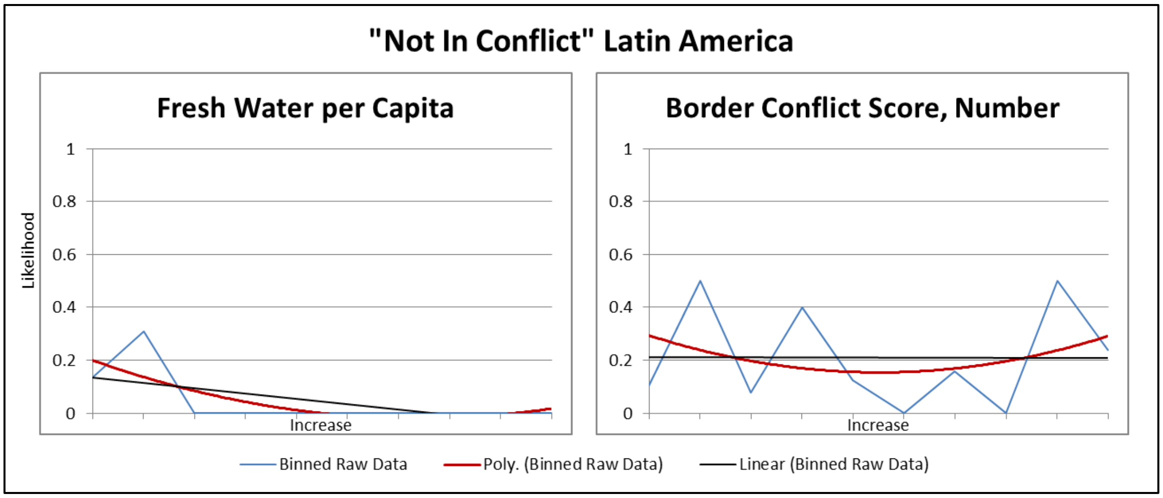


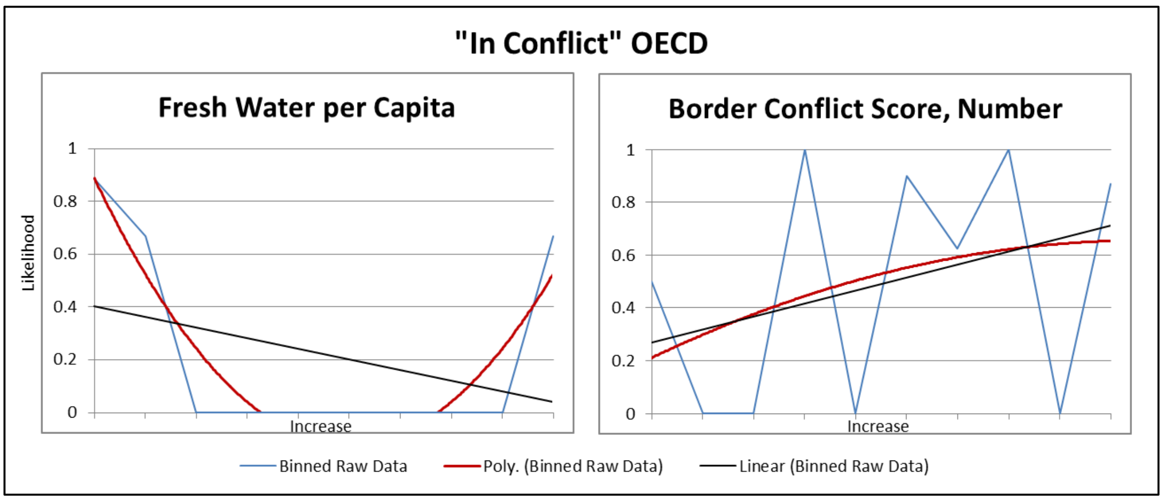


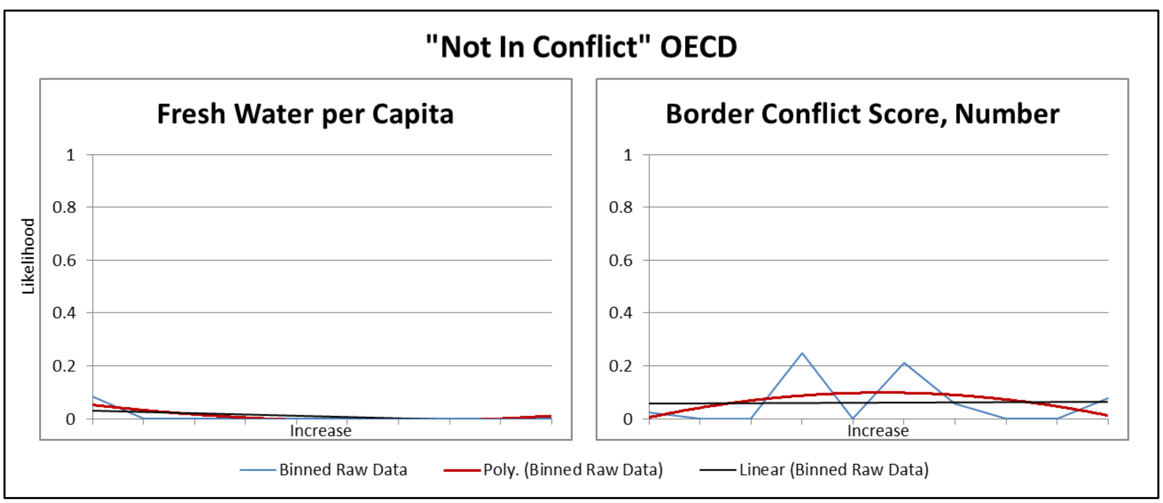


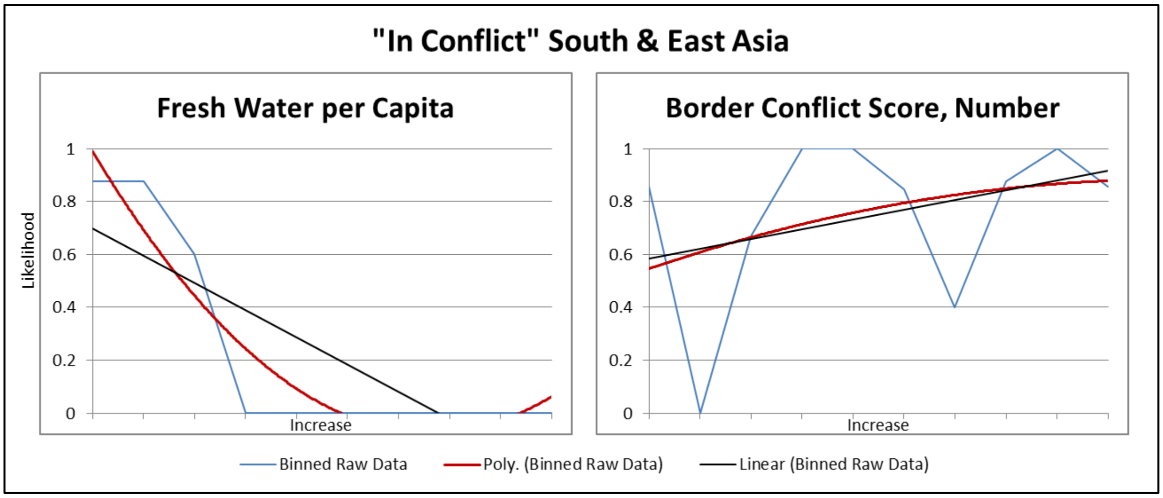


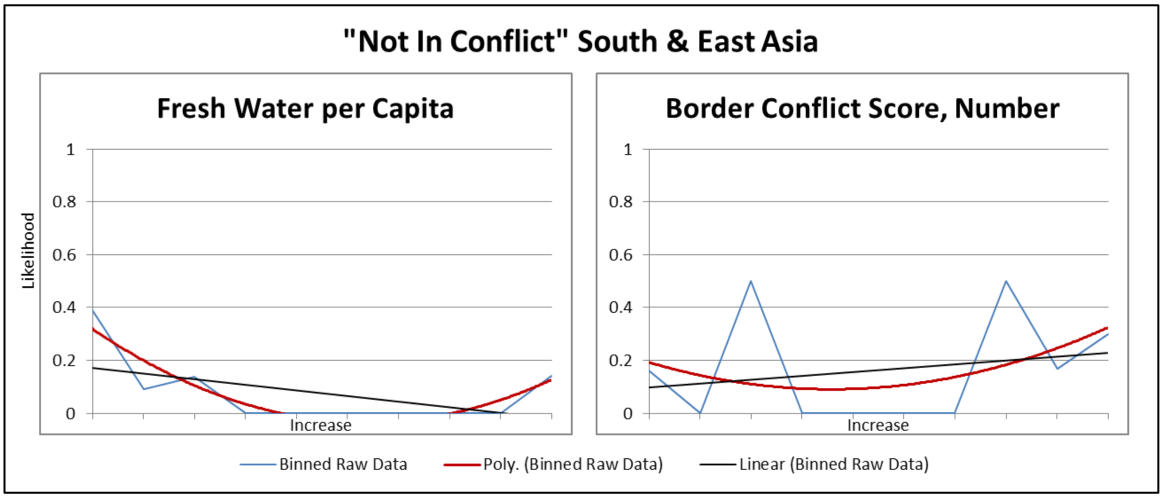
* Note: Only model with strong trends opposite of assumptions.

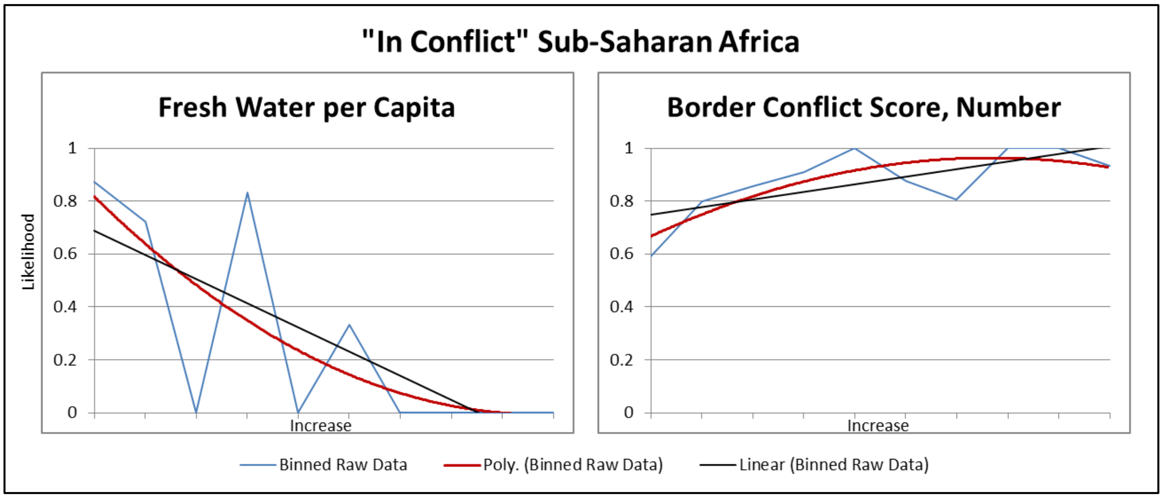

"Not In Conflict" Latin America
Fresh Water per Capita
Border Conflict Score, Number
Likelihood
1
0.8
0.6
0.4
0.2
0
Increase
Binned Raw Data
Poly. (Binned Raw Data)
Linear (Binned Raw Data)

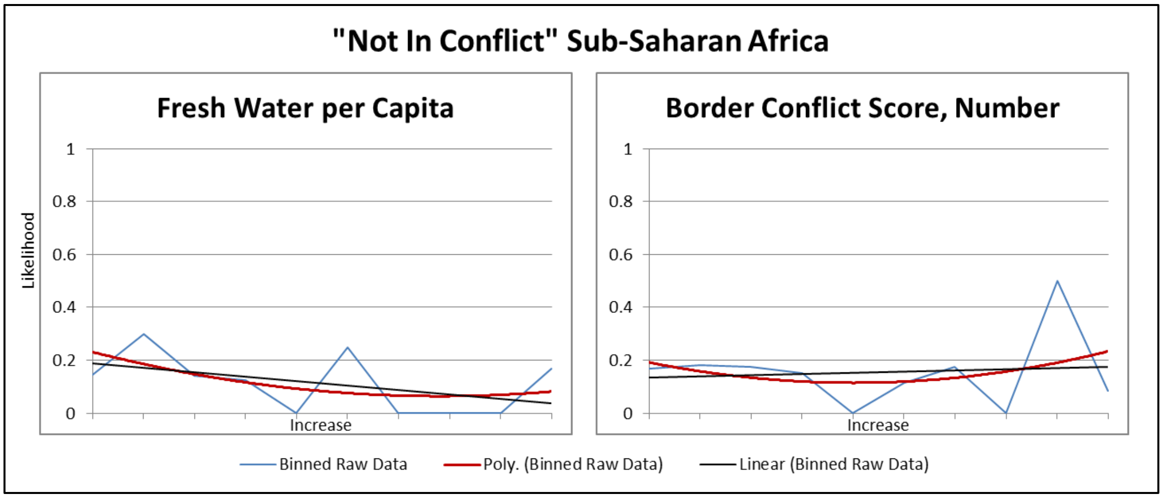

"In Conflict" OECD
Fresh Water per Capita
Border Conflict Score, Number
Likelihood
1
0.8
0.6
0.4
0.2
0
Increase
Binned Raw Data
Poly. (Binned Raw Data)
Linear (Binned Raw Data)


"Not In Conflict" OECD
Fresh Water per Capita
Border Conflict Score, Number
Likelihood
1
0.8
0.6
0.4
0.2
0
Increase
Binned Raw Data
Poly. (Binned Raw Data)
Linear (Binned Raw Data)


"In Conflict" South & East Asia
Fresh Water per Capita
Border Conflict Score, Number
Likelihood
1
0.8
0.6
0.4
0.2
0
Increase
Binned Raw Data
Poly. (Binned Raw Data)
Linear (Binned Raw Data)


"Not In Conflict" South & East Asia
Fresh Water per Capita
Border Conflict Score, Number
Likelihood
1
0.8
0.6
0.4
0.2
0
Increase
Binned Raw Data
Poly. (Binned Raw Data)
Linear (Binned Raw Data)


"In Conflict" Sub-Saharan Africa
Fresh Water per Capita
Border Conflict Score, Number
Likelihood
1
0.8
0.6
0.4
0.2
0
Increase
Binned Raw Data
Poly. (Binned Raw Data)
Linear (Binned Raw Data)


"Not In Conflict" Sub-Saharan Africa
Fresh Water per Capita
Border Conflict Score, Number
Likelihood
1
0.8
0.6
0.4
0.2
0
Increase
Binned Raw Data
Poly. (Binned Raw Data)
Linear (Binned Raw Data)

Appendix E: Regional Assignment of Nations

Table 12: Regional Assignment of Nations

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

Appendix F: Independent Variables and their Sources

Table 13: Independent Variables and their Sources

#	Variables	Units	Notes
		World Bank	
1	Arable Land	hectares per person	
2	Birth Rate	per 1000 people	
3	Death Rate	per 1000 people	
4	Fertility Rate	births per woman	
5	GDP per Capita	current USD	
	GDP per Capita (1000s)	current USD per 1000	Adjusted Units
6	Improved Water Source	% pop with access	
7	Life Expectancy	years	
8	Military Expend (% Gov Spending)	% Gov Spending	
9	Military Expend (% GDP)	% GDP	
10	Infant Mortality Rate	per 1000 live births	
11	Youth Bulge (Population Age 0-14)	% total population	
12	Population Density	people per sq kilometer	
13	Population Growth	annual %	
	Population Growth (ABS)	positive annual %	Derived absolute value of Population Growth
14	Refugee Population by country of Asylum	% population	
	Refugee (Asylum) (10,000s)	% population per 10000	Adjusted Units
15	Refugee Population by country of Origin	% population	
	Refugee (Origin) (10,000s)	% population per 10000	Adjusted Units
16	Renewable Fresh Water per Capita	avg cubic meters	
	Fresh Water per Capita (1000s)	avg cubic meters per 1000	Adjusted Units
17	Trade (% GDP)	% GDP	
18	Unemployment	total % of labor force	
		Center for Systemic Peace	
19	Polity IV	political behavior score	
20	Government Type (Emerging)	binary classification	Derived Classification from Polity IV Score
21	Government Type (Democratic)	binary classification	Derived Classification from Polity IV Score
22	Government Type (Foreign Interruption)	binary classification	Derived Classification from Polity IV Score
23	Government Type (Anarchy)	binary classification	Derived Classification from Polity IV Score
24	Government Type (Autocratic)	binary classification	Derived Classification from Polity IV Score
25	Government Type (Transition)	binary classification	Derived Classification from Polity IV Score
		Food & Agriculture Organization of the United Nations	
26	Caloric Intake	avg per person	
	Caloric Intake (1000s)	avg 1000s per person	Adjusted Units
		Freedom House	
27	Freedom Score	scored 0 to 1	
28	2 Yr Freedom Trend	scored -1 to 1	Derived from Freedom Score
29	3 Yr Freedom Trend	scored -1 to 1	Derived from Freedom Score
30	5 Yr Freedom Trend	scored -1 to 1	Derived from Freedom Score
		CIA World Fact Book	
31	Regime Type (Central)	binary classification	
32	Regime Type (Democractic)	binary classification	
33	Regime Type (Emerging)	binary classification	
34	Ethnic Diversity	% of dominate ethnic group	
35	Religious Diversity	% of dominate ethnic group	
36	Border Conflict Score	varied 0 to 5	
		Heidelberg Institute for International Conflict Research	
37	2 Yr Conflict Intensity Trend	varied 0 to 1	Derived from HIIK
38	Border Conflict Score, Number	% number of neighbors	Derived from HIIK
39	Border Conflict Score, Binary	binary classification	Derived from HIIK

(Shallcross, 2016)

Bibliography

Ahner, D. K., & Spainhour, R. (2015). Development of Analytical Models of Blue Force Interactions with Improvised Explosive Device Incidents. *Military Operational Research Society Journal, 20*(2).

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.

Davenport, T., & Harris, J. (2007). *Competing on Analytics.* Boston: Harvard Business School Press.

Department of Defense. (2012, February). *Theater Campaign Planning: Planners' Handbook*.

Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic Regression and Artificial Neural Network Classification Models: A Methodology Review. *Journal of Biomedical Informatics*(35), 352-359.

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.

Homer-Dixon, T. F. (1991). On the Threshold: Environmental Changes as Causes to Acute Conflict. *International Security*, *16*(2), 76-116.

Hosmer, D. W., Lemeshow, S., & Klar, J. (1988). Goodness-of-Fit Testing for Multiple Logistic Regression Analysis When the Estimated Probabilities are Small. *Biometrical Journal*(30), 911-924.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3rd ed.). Hoboken: John Wiley & Sons, Inc.

Occam's Razor. (2015). *In Encyclopaedia Britannica Online*. Retrieved from http://www.britannica.com

Shallcross, N. (2016). A Logistic Regression and Markov Chain Model for the Prediction of Nation-State Violent Conflicts and Transitions. Wright-Patterson AFB, Ohio: Air Force Institute of Technology MS Thesis.

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

United Nations Environment Programme. (2007). *Sudan, Post-Conflict Environmental Assessment.* Retrieved October 26, 2016, from http://www.unep.org/sudan/post-conflict/PDF/UNEP_Sudan.pdf

Ward, M., & Bakke, K. (2010). The Perils of Policy by P-Value: Predicting Civil Conflicts. *Journal of Peace Research, 47*(4), 1-13.

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
23-03-2017	Master's Thesis	Oct 2015 - Mar 2017

4. TITLE AND SUBTITLE

A Conditional Logistic Regression Predictive Model of World Conflict considering Neighboring Conflict and Environmental Security

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Leiby, Benjamin D., Capt, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way, Bldg 641
Wright-Patterson AFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENS-MS-17-M-140

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

JS J7 Joint Lessons Learned Division
116 B Lake View Parkway
Suffolk, VA 23435-2697
(757) 203-7589

10. SPONSOR/MONITOR'S ACRONYM(S)

JS J7

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution Statement A. Approved For Public Release; Distribution Unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Forecasts of conflict are of utmost importance for assisting combatant commanders in developing strategic and operational campaign and country plans that consider the dynamic changes that evolve within their area of responsibility. This research formulates and constructs five suites of statistical models to better understand the collinearity of environmental factors affecting conflict and compares the classification accuracy between forcing these factors into logistic regression models. A total of thirty-nine predictor variables are tested and evaluated for inclusion in a six region, two conflict state combination suite. The five suites of twelve models calculate the probability of whether a country will transition to either an “In Conflict” or “Not In Conflict” state for the following year. Handpicking the best models proposed in this study from each suite achieves modeling classification accuracies of 92.0% with 82.6% prediction accuracies. Through exploring new variables and selection methods, the models demonstrate that leveraging the collinearity of environmental factors help provide strategic insight in developing Department of Defense Theater Campaign Plans to effect the stability of national security.

15. SUBJECT TERMS

Conflict Transitions, Environmental Factors, Logistic Regression

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					Dr. Darryl K. Ahner, AFIT/ENS
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UU	104	**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U			(937) 255-6565 x4708 darryl.ahner@afit.edu

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