

# SAND REPORT

SAND2004-5973  
Unlimited Release  
Printed December 2004

REFERENCE COPY

## Climate Change Effects on International Stability: A White Paper



SANDIA NATIONAL  
LABORATORIES  
TECHNICAL LIBRARY

Mark Boslough, James Sprigg, George Backus, Mark Taylor, Laura McNamara, Joy Fujii, Kathryn Murphy, Leonard Malczynski, and Rhonda Reinert

Prepared by  
Sandia National Laboratories  
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,  
a Lockheed Martin Company, for the United States Department of  
Energy under Contract DE-AC04-94AL85000

Approved for public release; further dissemination unlimited.



**Sandia National Laboratories**

LIBRARY DOCUMENT  
DO NOT DESTROY  
RETURN TO  
LIBRARY VAULT

TOTAL PAGES: 81  
COPY \_\_\_\_\_

ELECTRONIC

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

**NOTICE:** This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from

U.S. Department of Energy  
Office of Scientific and Technical Information  
P.O. Box 62  
Oak Ridge, TN 37831

Telephone: (865)576-8401  
Facsimile: (865)576-5728  
E-Mail: [reports@adonis.osti.gov](mailto:reports@adonis.osti.gov)  
Online ordering: <http://www.osti.gov/bridge>

Available to the public from

U.S. Department of Commerce  
National Technical Information Service  
5285 Port Royal Rd  
Springfield, VA 22161

Telephone: (800)553-6847  
Facsimile: (703)605-6900  
E-Mail: [orders@ntis.fedworld.gov](mailto:orders@ntis.fedworld.gov)  
Online order: <http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online>



SAND2004-5973  
Unlimited Release  
Printed December 2004

LIBRARY DOCUMENT  
DO NOT DESTROY  
RETURN TO  
LIBRARY VAULT

# Climate Change Effects on International Stability: A White Paper

Mark Boslough, James Sprigg, George Backus, Mark Taylor, Laura McNamara, Joy Fujii, Kathryn Murphy, Leonard Malczynski, and Rhonda Reinert

Sandia National Laboratories  
P. O. Box 5800  
Albuquerque, NM 87185-1110

## Abstract

This white paper represents a summary of work intended to lay the foundation for development of a climatological/agent model of climate-induced conflict. The paper combines several loosely-coupled efforts and is the final report for a four-month late-start Laboratory Directed Research and Development (LDRD) project funded by the Advanced Concepts Group (ACG). The project involved contributions by many participants having diverse areas of expertise, with the common goal of learning how to tie together the physical and human causes and consequences of climate change. We performed a review of relevant literature on conflict arising from environmental scarcity. Rather than simply reviewing the previous work, we actively collected data from the referenced sources, reproduced some of the work, and explored alternative models. We used the unfolding crisis in Darfur (western Sudan) as a case study of conflict related to or triggered by climate change, and as an exercise for developing a preliminary concept map. We also outlined a plan for implementing agents in a climate model and defined a logical progression toward the ultimate goal of running both types of models simultaneously in a two-way feedback mode, where the behavior of agents influences the climate and climate change affects the agents. Finally, we offer some "lessons learned" in attempting to keep a diverse and geographically dispersed group working together by using Web-based collaborative tools.

## **Acknowledgments**

We wish to thank Dr. Sean O'Brien of the Center for Army Analysis (CAA) for assisting us with model runs using the Fuzzy Analysis of Statistical Evidence (FASE) code and for directing us to the Emergency Events Database. Rich Colbaugh of the National Security Agency provided helpful advice and consultation related to national security implications. Tim Trucano of Sandia National Laboratories (Sandia) provided useful information and insight regarding verification and validation issues. We also thank Andrew Scholand of Sandia for letting us use his Wiki.

This work was funded by the ACG (Advanced Concepts Group) under LDRD project 78307.

# Contents

Nomenclature.....	8
1 Introduction.....	9
1.1 Background.....	9
1.2 Technical Approach.....	10
1.3 Document Overview.....	11
1.4 References.....	11
2 Examining Darfur: A Candidate Case Study of Conflict Related to Climate Change.....	12
2.1 General Description.....	12
2.1.1 About Darfur.....	12
2.1.2 The Current Crisis.....	13
2.1.3 Origins of the Current Crisis.....	13
2.1.4 Structural Conditions.....	14
2.1.4.1 Environmental Factors.....	14
2.1.4.2 Economic Factors.....	14
2.1.4.3 Social Factors.....	15
2.1.4.4 Political Factors.....	15
2.1.4.5 International Response.....	15
2.2 Systems Dynamics Modeling of the Conflict.....	16
2.3 References.....	18
3 Causal Factors of Conflict.....	21
3.1 Overview of Conflict Forecasting.....	21
3.1.1 Background.....	21
3.1.2 Forecasting and Causal Relationships.....	22
3.2 FASE Analysis of Instability.....	22
3.2.1 The KOSIMO Database.....	22
3.2.2 Reproduction.....	24
3.2.3 Natural Disasters.....	27
3.3 Logistic Analysis of Instability.....	27
3.3.1 Method and Statistical Model.....	28
3.3.1.1 Dependent Variable: Instability.....	30
3.3.1.2 Social Groups.....	30
3.3.1.3 Prior Instability.....	31
3.3.1.4 Selection of Variables.....	32
3.3.2 Results for Regression Models.....	32
3.3.2.1 Interpreting the Logistic Results.....	34
3.3.2.2 Social Groups.....	36
3.3.3 Intensity of Instability: Model and Results.....	36
3.3.4 Natural Disasters.....	37
3.3.5 Extrapolation and Forecasting.....	39
3.3.6 Forecasting Impacts of Caloric Reductions.....	41
3.3.6.1 Calorie-Instability Relationship.....	42
3.3.7 Remarks.....	44
3.4 Orthogonal Types of Conflict.....	44

	3.4.1 Civil War Versus Local Conflict.....	45
	3.4.2 Logistic Model of Civil War.....	45
	3.4.3 Logistic Model of Local War.....	46
	3.5 Remarks .....	48
	3.6 References.....	48
4	Using Recent Advances in Behavioral Analysis to Address Climatic Impact on Conflict.....	50
	4.1 Introduction.....	50
	4.1.1 Reviewed Methods.....	50
	4.1.2 Section Overview.....	51
	4.2 Qualitative Choice Theory.....	51
	4.2.1 Probabilistic Decisions under Imperfect Conditions.....	52
	4.2.2 QCT Mathematical Basics .....	53
	4.2.3 Testing QCT for Climate-Conflict Impacts .....	55
	4.3 Cointegration and Granger Causality.....	59
	4.3.1 Cointegration Mathematical Basics .....	60
	4.3.1.1 Stationarity.....	60
	4.3.1.2 Unit Roots.....	60
	4.3.2 Granger Causality Mathematical Basics .....	62
	4.3.2.1 Testing Cointegration for Climate-Conflict Impacts.....	63
	4.4 Bayesian Causal Discovery.....	64
	4.4.1 BCD Mathematical Basics .....	65
	4.5 Summary.....	67
	4.6 References.....	67
5	Agent and Climate Model Coupling .....	70
	5.1 The Community Climate System Model .....	70
	5.2 Development of a Coupled Model.....	71
	5.3 References.....	71
6	Using Wiki in a Collaborative Environment.....	73
	6.1 General Platform Features .....	73
	6.2 Why We Chose to Use a Wiki.....	74
	6.3 User Impressions of the Wiki .....	74
	6.4 References.....	75
7	Summary of Findings and Lessons Learned .....	76
8	Future Plans and Applications.....	78
	8.1 Ongoing Project Development.....	78
	8.2 Applications.....	79

## Figures

2-1	Map of Sudan. <i>Source:</i> University of Texas Libraries. Perry-Castaneda Library Map Collection.....	12
2-2	Concept map of the Darfur conflict.....	17
3-1	Conditional probabilities of conflict. ....	25
3-2	Conditional probabilities of conflict versus disaster index. ....	27
3-3	Impact of ethnic homogeneity on risk of crisis. ....	31

3-4	Likelihood of instability versus trade openness. ....	35
3-5	Likelihood of instability versus average daily calories. ....	35
3-6	Marginal likelihood of conflict versus group size. ....	36
3-7	Logistic response to caloric reductions. ....	42
3-8	Forecasted probabilities of conflict resulting from caloric reductions. ....	43
3-9	Fitted relationship of conflict versus caloric reductions. ....	43
3-10	Orthogonal conflict space. ....	44
4-1	Illustrative choice distribution. ....	52
4-2	Illustrative market-share response. ....	53
4-3	Illustrative climate impacts on conflict. ....	59
4-4	Simple DAG. ....	65
4-5	Multipath DAG. ....	65

## Tables

3-1	Partial Conflicts in Afghanistan. ....	23
3-2	Overview of Regression Models. ....	29
3-3	Regression Results. ....	33
3-4	Regression Results for Intensity. ....	37
3-5	Logistic Estimation of the Occurrence of Conflict Including Variables Representing Number Killed and Injured in Disasters. ....	38
3-6	Statistical Impact of Prior-Conflict Variable. ....	40
3-7	Forecast Accuracy Excluding the Prior-Conflict Variable. ....	40
3-8	Forecast Accuracy Including the Prior-Conflict Variable. ....	40
3-9	Logistic Estimation of the Occurrence of Conflict. ....	41
3-10	Logistic Estimation of Occurrences of Civil War. ....	45
3-11	Logistic Estimation of Occurrences of Local War. ....	47
4-1	Variables Affecting the Probability of Conflict. ....	57
4-2	Five-Lag Granger Causality Test. ....	63
4-3	Single-Lag Granger Test. ....	64

# Nomenclature

AAAS	American Association for the Advancement of Science
ACG	Advanced Concepts Group
ADF	augmented Dickey-Fuller
AU	African Union
BBC	British Broadcasting Company
BCD	Bayesian causal discovery
CAA	Center for Army Analysis
CCSM	Community Climate System Model
CFSD	Catalina Foothills School District
CO <sub>2</sub>	carbon dioxide
COW	Correlates of War
CRED	Centre for Research on the Epidemiology of Disasters
CSM	<i>Christian Science Monitor</i>
DAG	directed acyclic graph
DGP	data-generating process
DW	Durbin-Watson
ECM	error-correction mechanism
EM-DAT	Emergency Events Database
FASE	Fuzzy Analysis of Statistical Evidence
HRF	Human Rights First
HRW	Human Rights Watch
HTML	HyperText Markup Language
ICG	International Crisis Group
IPCC	Intergovernmental Panel on Climate Change
JEM	Justice Equality Movement
KOSIMO	Conflict Simulation Model
LANL	Los Alamos National Laboratory
LDRD	Laboratory Directed Research and Development
MLE	maximum-likelihood estimation
NATO	North Atlantic Treaty Organization
NP	nondeterministic polynomial
PBS	Public Broadcasting System
QCT	qualitative choice theory
RUM	random utility maximization
Sandia	Sandia National Laboratories
SFTF	State Failure Task Force
SGS	Spirtes, Glymour, and Scheines
SLA	Sudan Liberation Army
SLM	Sudan Liberation Movement
SPLM	Sudan Peoples Liberation Movement
UN	United Nations
UNCCD	United Nations Convention to Combat Desertification
V&V	verification and validation
WHO	World Health Organization



# Climate Change Effects on International Stability: A White Paper

## 1 Introduction

The security of the United States is increasingly vulnerable to challenges by a new threat. Changes in the global climate, caused by a combination of natural and human-induced influences, are already occurring. The paleoclimate record suggests that very abrupt climate changes are possible. As the world's climate changes, human societies will respond by adaptive behaviors that could lead to changing global alliances, civil unrest, and war as populations shift and vie for territory and natural resources.

Strong and accelerating global trends are reducing the availability of vital resources such as water, food, soil, territory, and stable climate, potentially heightening tensions between or within nations. Foreign policy experts have compiled evidence that such scarcities can precipitate conflict. As Homer-Dixon (1993) writes, “. . . researchers have gathered enough information to reach a disturbing conclusion: environmental scarcities are already contributing to violent conflicts in many parts of the developing world. Moreover, these conflicts may be the early signs of an upsurge in violence in the coming decades—especially in poor countries—that is caused or aggravated by environmental change.”

The ability of policy makers to anticipate and help to mitigate the potential conflicts in society that are caused by changes in global climate requires a better understanding of the interaction between social and physical Earth systems. To further this understanding, we are developing methods to incorporate historical climatology into an established macrostructural analysis as a necessary first step toward developing a coupled climatological/agent forecast model of climate-induced conflict. In the initial phase of this project, which was funded by Sandia National Laboratories' (Sandia's) Advanced Concepts Group (ACG) under the Laboratory Directed Research and Development (LDRD) Program, we worked toward developing a proof-of-principle methodology with the goal of laying the groundwork for coupling agent-based social models to physics-based climate models.

### 1.1 Background

Although there have been a number of efforts within the defense and intelligence communities to develop models capable of forecasting areas of potential conflict on time scales of 10 to 20 years, notably missing from these models is an explicit dependence on weather and climate. The Pentagon's Office of Net Assessment, led by Andrew Marshall, recently acknowledged this deficiency (Schwartz and Randall 2003) and has issued a series of recommendations for accelerating research on the connection between climate change and conflict.

Much of the research work that has been done in analyzing conflict and environmental change follows a qualitative approach. For example, Homer Dixon (1993) has written extensively about the relationships among environmental degradation, resource scarcity, and conflict. Similarly, the Global Business Network recently used scenario analysis (Schwartz and Randall 2003) to hypothesize social responses to various scenarios of abrupt climate change. In 2003, Sandia's ACG sponsored a workshop on global climate change and international security (Karas 2003) in which a panel of experts developed three scenarios to illustrate the types of conflicts that could be precipitated by changes in the global climate. These scenarios included escalating conflicts between destabilized nuclear powers, the loss of nuclear weapons from the arsenal of a destabilized country, and the decision of an impacted nonnuclear nation to develop a nuclear deterrence.

## 1.2 Technical Approach

To address the limitations of current conflict-forecasting models described above, Sandia pursued a multipronged approach to analyze the relationships between climate and conflict. This approach involved the following activities:

- We monitored several emerging conflicts in the world, including Bangladesh and the Sudanese region of Darfur. We selected the Darfur conflict as a candidate case study for further analysis.
- We quantitatively explored causal factors related to the study of state conflict using the Fuzzy Analysis of Statistical Evidence (FASE) method, which was developed by the Center for Army Analysis (Chen 2000). We also applied traditional statistical regression methods. The FASE pattern-classification algorithm has already provided conditional probabilities of conflict based on historical data for such independent variables as youth bulge, infant mortality rate, and trade openness (O'Brien 2002). We extended O'Brien's data set to incorporate the occurrence of natural disasters.
- We reviewed and tested other statistical methods from the behavioral and social sciences to further probe the causal relationships between climate and conflict. These methods are qualitative choice theory (QCT), cointegration, Granger causality, and Bayesian causal discovery (BCD).
- We identified the process that would be required to couple an agent-based model with a climate model.

In addition, to implement this approach across our multidisciplinary team, we took advantage of an internal Web collaborative environment, the Wiki, through which we could post and share the results of our work and exchange information and ideas in a timely fashion.

## 1.3 Document Overview

This white paper is organized as a set of individually referenced topical sections, closely tracking our technical approach. Section 1 has introduced the need for including the variable of climate as part of forecasting conflict and has outlined our preliminary technical approach to addressing this need. Section 2 presents a brief overview of the Darfur crisis and includes a preliminary visual model that identifies interacting factors and conditions that we believe contribute to that conflict. In Section 3, we briefly introduce the field of conflict forecasting and then describe the analyses of conflict performed with the FASE algorithm and with traditional statistical regression models, including interpretation of the results of these analyses. Section 4 contains an overview of selected statistical approaches from the behavioral and social sciences and provides examples of how these approaches could be applied in conflict analyses. Section 5 proposes a process by which we can couple well-established climate models with agent-based models. For those unfamiliar with collaborative platforms, Section 6 highlights key features of the Wiki and reports on our experience with it. A summary of key findings and lessons learned from all of the project's activities is provided in Section 7. The report concludes with Section 8, where we discuss future development of a coupled climate/agent-based model and possible applications for its use in related and different fields.

## 1.4 References

- Chen, Y.Y. 2000. "Fuzzy Analysis of Statistical Evidence." *IEEE Transactions on Fuzzy Systems* 8, no. 6: 796–99.
- Homer-Dixon, T. 1993. *Environmental Scarcity and Global Security* (Headline Series no. 300). New York: Foreign Policy Association.
- Karas, T. H. 2003. *Global Climate Change and International Security*. SAND2003-4114. Albuquerque, NM: Sandia National Laboratories.
- O'Brien, S. P. 2002. "Anticipating the Good, the Bad, and the Ugly: an Early Warning Approach to Conflict and Instability Analysis." *Journal of Conflict Resolution* 46, no. 6: 791–811.
- Schwartz, P., and D. Randall. 2003. *An Abrupt Climate Change Scenario and Its Implications for United States National Security*. Available at [http://www.ems.org/climate/pentagon\\_climate\\_change.html#report](http://www.ems.org/climate/pentagon_climate_change.html#report) (accessed 30 October 2004).

## 2 Examining Darfur: A Candidate Case Study of Conflict Related to Climate Change

As part of this LDRD project, we began investigating a world conflict that involves changes in climate. We chose the Sudanese region of Darfur as a candidate case study for two reasons. First, the conflict is occurring now. Thus, there is a great deal of information about it in the media, including news, opinion, and propaganda. Second, we believe that this conflict demonstrates the connection between climate-change-induced environmental scarcity and conflict. It is a connection that is not simple, but is rather one that has many contributing factors that are exacerbated by the changes in climate.

For several months we used Google News to monitor the conflict and posted this information on our internal Web site (see Section 6). As we learned about the complex interacting factors and conditions, we also developed a relational diagram of the conflict, using tools from systems dynamics. In this section, we provide a general description of the crisis in Darfur and show our initial attempt to pictorially represent this conflict. Importantly, the work described in this section constitutes a framework for developing case studies of responses to climate change and environmental scarcity in follow-on work for this project.

Because the crisis is ongoing and taking place in a remote part of the world, accurate and up-to-date information is not available from peer-reviewed sources. Consequently, we relied heavily on Web-based international media for information. The references in Section 2.3 reflect the transient nature of the information associated with this report, which is intended as a guide to future work but not as definitive research.

### 2.1 General Description

#### 2.1.1 About Darfur

Darfur is located in the northwestern part of Sudan. To the west, as shown in Figure 2-1, Darfur borders Libya, Chad, and the Central African Republic (HRW 2004). Darfur is approximately 80% the size of Texas (U.S. Senate 2004) and covers 150,000 square miles (Ryle 2004). The population of Darfur is estimated at around 6 million (U.S. Senate 2004).

Geographically, Darfur extends from desert in the north to a central area, which includes the Jebel Marra volcanic plateau, to savannah in the south (Verney 2004). The central area has more rainfall and more fertile soil than the other areas and is thus the richest agriculturally (HRW 2004; Verney 2004).



**Figure 2-1.** Map of Sudan. *Source:* University of Texas Libraries. Perry-Castaneda Library Map Collection.

## 2.1.2 The Current Crisis

The United Nations (UN) has called the current crisis in Darfur the worst humanitarian disaster in the world today (CSM 2004). The crisis erupted in February 2003 when two loosely allied rebel groups, the Sudan Liberation Army (SLA) and the Justice Equality Movement (JEM), launched attacks on government posts in Darfur. The rebels had a few victories in the initial months of the conflict, but then the government “turned loose” the Janjaweed militias, backed by its regular forces, on civilians who were thought to support the insurgency (ICG 2004). The Janjaweed militias are made up of Arab nomadic shepherds and have been described as a civilian terrorist force (Age 2004). The Janjaweed have been accused of major human rights violations, mass killings of civilians, rape and other forms of sexual violence, forced displacement, and burning of villages (Reeves 2004a; HRF 2004a). The Janjaweed have also been accused of intentionally destroying irrigation systems and food stores so that the civilian populations do not return to the burned-out villages (ICG 2004).

In April 2004, a cease fire that was mediated by Chad was signed by the Sudan government and the two rebel groups, with help from the African Union (AU), the UN, and several Western states (ICG 2004). However, as of July 2004, serious fighting continued despite the ceasefire (*Washington Times* 2004).

The crisis has taken a huge toll on the civilian population of Darfur. The number of persons displaced by the conflict has been estimated at approximately 1.45 million within Darfur (HRF 2004a); most are in refugee camps within Darfur (ICG 2004), with approximately 200,000 in camps in Chad (HRF 2004a). The statistic on the number of deaths, however, appears to be questionable. In early October 2004, the British Broadcasting Company (BBC 2004a) reported that an estimated 50,000 people had died as a result of the conflict since it began in 2003. By late October, the BBC reported that the number of deaths had risen to 70,000 (BBC 2004b). However, according to Professor Eric Reeves (2004b) of Smith College, the “50,000” figure refers to the number of people who have died from disease in the refugee camps since April 2004, as reported by the World Health Organization (WHO). Reeves estimates that as many as 300,000 people have died from violence, disease, and malnutrition since February 2003 when the Darfur crisis erupted.

During late October 2004, peace talks between the Sudanese government and the rebels resumed in Nigeria (*Cape Times* 2004), but these talks were subsequently stalled because of concerns by the Sudan Liberation Movement (SLM) about security (BBC 2004b). Note that the SLM is another name for the SLA (Sudan Liberation Army).

## 2.1.3 Origins of the Current Crisis

The International Crisis Group (ICG 2004) states that there were multiple causes for the insurgency in Darfur in 2003, including economic and political marginalization, underdevelopment, and the government’s policy (longstanding) of supporting the Janjaweed militias against the primarily African farming communities. According to several sources, the roots of the current violence can be traced to traditional clashes

between nomadic (pastoral) Arab herders and sedentary African farmers (HRW 2004; U.S. Senate 2004; Powell 2004). Such clashes occurred as Arab herders from the north migrated south in the dry season in search of water sources and grazing for their cattle (HRW 2004; Verney 2004) and the cattle and camels of the herders trampled the fields of the African farmers (HRW 2004). Some sources say the conflicts in Darfur have been going on for several decades, while others say centuries (HRW 2004; Verney 2004; El-Leithy 2004). Traditionally, such conflicts were resolved by negotiation, but the conflicts intensified during the 1980s and 1990s because of drought and also the government's policy of arming the Arab herders and removing the weapons of the farmers (Verney 2004).

## **2.1.4 Structural Conditions**

The conflict is fueled by a number of interacting factors, including environmental, economic, social, and political factors, as highlighted below. International concerns and actions toward the Darfur crisis are also briefly addressed.

### **2.1.4.1 Environmental Factors**

Climate change since the 1970s has accelerated the pace of desertification, putting pressure on those who live in the northern part of Darfur to move southward (Verney 2004) and thus contributing to the historic struggle for land between the herders and the farmers (Mulugeta 2004). Desertification is defined as "land degradation in arid, semi-arid and dry subhumid areas resulting from various factors, including climatic variations and human activities" (McCarthy et al. 2001). Examples of such human activities are overcultivation, deforestation, and poor irrigation practices, which reduce the amount of arable land (UNCCD 2004).

Though the rainy season came to Darfur in 2004, civilians have had to flee the land to escape the conflict and thus were unable to plant their crops, contributing to a shortage of food in the region. The rains also spur flash floods, which make the roads impassable, restrict the delivery of assistance, and increase the risk of disease (Benn 2004). In addition, there has been concern that locusts currently threatening northern Africa would swarm to Darfur, where locust-control efforts would be impossible (AAAS 2004; U.S. Senate 2004). Locusts eat their weight in food every day (AAAS 2004).

### **2.1.4.2 Economic Factors**

The farmers in Darfur grow crops such as sorghum, millet, groundnuts, and tomatoes; the nomadic pastoralists raise camels in the north and cattle in the south. Livestock is Darfur's main export (Verney 2004).

The groups in the conflict have different styles of living, or ways of being in the world, i.e., nomadic versus sedentary, which directly relate to the competition between the pastoralists and the farmers for land and water as climate change affects the region.

### **2.1.4.3 Social Factors**

Generally, the sedentary farmers of the central area of Darfur are composed of non-Arab or African ethnic groups, such as the Fur, Masaalit, Tunjur, Tama, Bergid, and Berti. The pastoralists, on the other hand, are predominantly of Arab descent. The herding tribes of the northern area include Arab ethnic groups such as the northern Rizeigat, Mahariya, and the African Zaghawa. Cattle-herding Arab tribes, such as the southern Rizeigat and Habbaniya, inhabit the southern area of Darfur (HRW 2004). In the current conflict, the Fur, Masaalit and Zaghawa African ethnic groups view the attacks on their communities as racially and ethnically motivated (BBC 2004a).

Regarding religious orientation, Darfur is uniformly a Muslim region (ICG 2004).

Lord (2004), commenting on the contribution of social factors to the Darfur conflict, noted:

The root cause of the Darfur conflict is actually ecological, with prolonged droughts and rapid desertification driving poor pastoral 'Arabs' to take over the lands of even poorer settled 'black' farmers. With extensive damage to the ecology throughout the region, what we see as ethnic conflict is really resource conflict at root, with religion even further down the list of factors.

### **2.1.4.4 Political Factors**

The present Arab Islamic government of Sudan came to power in 1989 in a military coup (ICG 2004). For its participation in the Darfur crisis, the government has been denounced by the international community (see Section 2.1.4.5).

Sudan itself has had an ongoing civil war for 21 years between the government and southern rebels, the Sudan People's Liberation Movement (SPLM). In May 2004, a preliminary peace agreement was signed in Naivasha, Kenya (ICG 2004).

### **2.1.4.5 International Response**

The crisis in Darfur has become a subject of international concern and involvement, as noted by recent developments. In September 2004, U.S. Secretary of State Colin Powell concluded that genocide has been committed in Darfur (U.S. Senate 2004; HRF 2004a). Also in September 2004, the UN Security Council adopted a resolution threatening oil sanctions against Sudan if the government did not rein in the militias (the Janjaweed) (BBC 2004a). In October 2004, the UN announced the establishment of a commission of inquiry to investigate violations of international humanitarian law and human rights law in Darfur and to determine whether laws of genocide have occurred (HRF 2004b).

A number of relief organizations have been working in Darfur. In July, the TransAfrica Forum (2004) published a list of these organizations: Catholic Relief Services, International Committee of the Red Cross, Inter Action, Medair, Medicines Sans Frontiers/Doctors Without Borders, Relief Web, UN Children's Fund, UN World Food Programme, U.S. Agency for International Development, and the WHO.

Financial support for diplomatic and humanitarian relief efforts has been provided in great part by the United States, the United Kingdom, and the European Community (Benn 2004).

The AU has dispatched several hundred cease-fire monitors to Darfur and announced, in October 2004, that it will increase this force tenfold. As of the end of October 2004, only about one-half of the 2.25 million people in Darfur who need food assistance were being reached (PBS 2004).

## **2.2 Systems Dynamics Modeling of the Conflict**

Systems dynamics modeling is an approach used to describe interrelated systems. The method provides a set of tools that enables practitioners to construct qualitative and quantitative diagrams about the behavior of the selected systems (CFSD 2003).

During the course of this project, we began to develop a concept map for understanding the Darfur crisis. In this task, we combined several of the graphical components from the set of systems dynamics tools to show the interaction of important elements of the conflict. These elements were taken from our review of the current literature and international media sources on the current crisis, as summarized in Section 2.1. Figure 2-2 illustrates a preliminary concept map that we developed to help us understand the Darfur crisis and its relationship to environmental change.



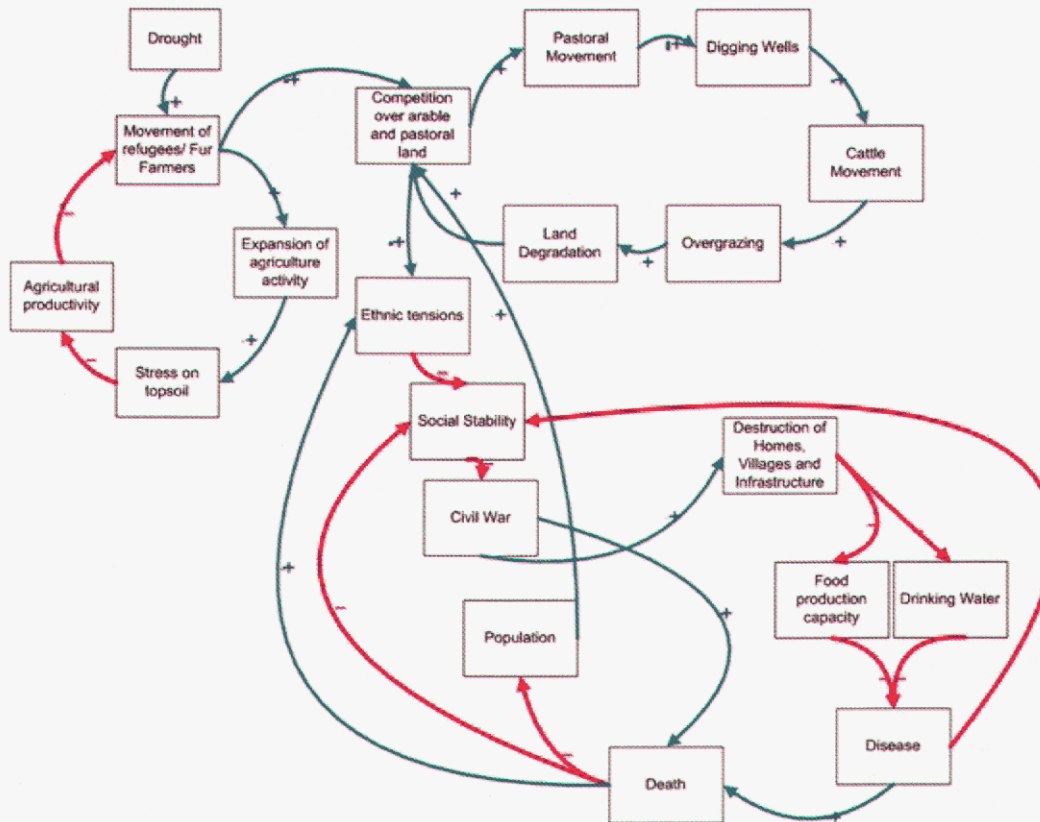


Figure 2-2. Concept map of the Darfur conflict.

In the above figure, we have used several symbols:

- A box represents a quantitative element (i.e., factor or condition) that has a causal relationship with another element and that can increase or decrease over time.
- An arrow denotes that one element is affecting another element. The “+” and “-” symbols that are associated with an arrow indicate the effect of the influence of one element on the other element.
  - In general, a “+” means that both elements move in the same direction, i.e., an increase in the first element is expected to cause an increase in the second element, or a decrease in the first element is expected to cause a decrease in the second element.
  - In general, a “-” means that both elements move in the opposite direction, i.e., an increase in the first element is expected to cause a decrease in the second element, or a decrease in the first element is expected to cause an increase in the second element.

Taking a small piece of the map, we can explain how the elements interact and influence each other. In the upper left, we have the element of drought, which is related

to human land-use patterns as well as climate change. Drought affects the movement of refugees (the Fur farmers). There is a + symbol on the arrow connecting the drought to the movement of refugees. This indicates that as the drought increases, the movement of refugees is expected to increase (or conversely, if the drought decreases, the refugees will move less). The movement of refugees has a similar relationship to the expansion of agricultural activity and competition over arable and pastoral land. For example, as the movement of refugees increases, agricultural activity will increase as the refugees find new areas to farm and also there will be competition for these new areas to farm.

## 2.3 References

AAAS (American Association for the Advancement of Science). 2004. "Random Samples" *Science* 305, no. 5682: 337c. 16 July.

Age. 2004. "Sudan Threatens to be the Next Rwanda." 29 May. Available at <http://www.theage.com.au/articles/2004/05/28/1085641708758.html?from=storylhs> (accessed 4 October 2004).

BBC (British Broadcasting Company). 2004a. "UN to Investigate Darfur Genocide." *BBC News*, 8 October. Available at <http://news.bbc.co.uk/2/hi/africa/3725770.stm> (accessed 30 October 2004).

———. 2004b. "Darfur Talks Falter Over Security." 27 October. Available at <http://news.bbc.co.uk/2/hi/africa/3729275.stm> (accessed 27 October 2004).

Benn, H. 2004. "Race Against Time to Save Darfur." *Scotsman*, 12 June. Available at <http://thescotsman.scotsman.com/index.cfm?id=666922004> (accessed 1 November 2004).

*Cape Times*. 2004. "Delayed Darfur Peace Talks Finally Get Under Way." 26 October. Available at <http://www.capetimes.co.za/index.php?fSectionId=272&fArticleId=2274468> (accessed 27 October 2004).

CFSD (Catalina Foothills School District). 2003. "Tips for Using System Dynamics Tools." Available at <http://www.clexchange.org/ftp/documents/Implementation/IM2003-12TipsUsingSDTools.pdf> (accessed 3 November 2004).

CSM (*Christian Science Monitor*). 2004. "Darfur Beyond Description." 27 July. Available at <http://www.csmonitor.com/2004/0727/p08s02-comv.html> (accessed 1 November 2004).

El-Leithy, I. 2004. "Khartoum Plans to Bring Back Fleeing Darfuris," *IslamOnline.net*, 7 July. Available at <http://www.islam-online.net/English/News/2004-07/07/article02.shtml> (accessed 1 November 2004).

- HRF (Human Rights First). 2004a. "Sudan: The Human Rights Emergency in Darfur." October. Available at [http://www.humanrightsfirst.org/international\\_justice/regions/sudan/sudan.htm](http://www.humanrightsfirst.org/international_justice/regions/sudan/sudan.htm) (accessed 30 October 2004).
- . 2004b. "Newly Established U.N. Commission of Inquiry on Genocide in Darfur Needs Full Cooperation and Support." 8 October. Available at [http://www.humanrightsfirst.org/media/2004\\_alerts/1008.htm](http://www.humanrightsfirst.org/media/2004_alerts/1008.htm) (accessed 30 October 2004).
- HRW (Human Rights Watch). 2004. "Darfur in Flames: Atrocities in Western Sudan: Background." April. Available at [http://www.hrw.org/reports/2004/sudan0404/3.htm#\\_Toc68525370](http://www.hrw.org/reports/2004/sudan0404/3.htm#_Toc68525370) (accessed 4 October 2004).
- ICG (International Crisis Group). 2004. *Sudan: Now or Never in Africa*. ICG Africa Report N°80. Nairobi/Brussels: ICG. Available at [http://www.icg.org/library/documents/africa/horn\\_of\\_africa/080\\_sudan\\_now\\_or\\_never\\_in\\_darfur.pdf](http://www.icg.org/library/documents/africa/horn_of_africa/080_sudan_now_or_never_in_darfur.pdf) (accessed 3 November 2004).
- Lord, C. 2004. "One Hellhole, Under God." *Freezerbox Magazine*, 2 August. Available at <http://www.freezerbox.com/archive/article.asp?id=296> (accessed 1 November 2004).
- McCarthy, J. J., O. F. Canziani, N. A. Leary, D. J. Dokken, and K. S. White, eds. 2001. *Climate Change 2001: Impacts, Adaptation, and Vulnerability*. Cambridge, UK: Cambridge University Press. Available at [http://www.grida.no/climate/ipcc\\_tar/wg2](http://www.grida.no/climate/ipcc_tar/wg2) (accessed 3 November 2004).
- Mulugeta, S. 2004. "Humanitarian Disaster in Sudan." *CounterCurrents.org*, 29 July. Available at <http://www.countercurrents.org/sudan-mulugeta290704.htm> (accessed 1 November 2004).
- PBS (Public Broadcasting System). 2004. "Crisis in Sudan." *Online News Hour*, 20 October. Available at [http://www.pbs.org/newshour/bb/africa/july-dec04/sudan\\_10-20.html](http://www.pbs.org/newshour/bb/africa/july-dec04/sudan_10-20.html) (accessed 30 October 2004).
- Powell, C. L. 2004. "The Crisis in Darfur." Written remarks of Secretary Colin Powell before the Senate Foreign Relations Committee (September 9). Available at <http://www.state.gov/secretary/rm/36032.htm> (accessed 10 September 2004).
- Reeves, E. 2004a. "Genocide in Sudan." *In These Times*, 6 May. Available at [http://www.inthesetimes.com/site/main/article/genocide\\_in\\_sudan](http://www.inthesetimes.com/site/main/article/genocide_in_sudan) (accessed 28 October 2004).
- . 2004b. "Darfur Mortality Update: Current Data for Total Mortality from Violence, Malnutrition, and Disease." *Genocide Watch*, 8 October. Available at

<http://www.genocidewatch.org/SudanDARFUrMORTALITYUPDATE8oct2004.htm>  
(accessed 30 October 2004).

Ryle, J. 2004. "Disaster in Darfur." *The New York Review of Books* 51, no. 13 (12 August). Available at <http://www.nybooks.com/articles/17326> (accessed 8 September 2004).

TransAfrica Forum. 2004. "Relief Organizations Working in Darfur." July. Available at [http://www.transafricaforum.org/documents/DarfurReliefOrgsJuly04\\_001.pdf](http://www.transafricaforum.org/documents/DarfurReliefOrgsJuly04_001.pdf)  
(accessed 3 November 2004).

UNCCD (United Nations Convention to Combat Desertification). 2004. "Ten Years On: UN Marks World Day to Combat Desertification." 1 June. Available at <http://www.unccd.int/publicinfo/menu.php> (accessed 2 November 2004).

University of Texas Libraries. 2004. Map of Sudan. From the Perry-Castaneda Library Map Collection. Available at <http://www.lib.utexas.edu/maps/africa/sudan.jpg>  
(accessed 16 November 2004).

U.S. Senate. 2004. Committee on Foreign Relations. *The Current Situation in Sudan and the Prospects for Peace: Hearing before the Committee on Foreign Relations*. 108th Cong., 2d sess., 9 September.

Verney, P. 2004. "Sudan: Darfur's 'Human-Made' Disaster." *Green Left Weekly*, 28 July. Available at <http://www.greenleft.org.au/back/2004/591/591p16.htm> (accessed 1 November 2004).

*Washington Times*. 2004. "U.S. Warns Against Traveling to Sudan." 1 July. Available at <http://washingtontimes.com/upi-breaking/20040701-100718-8948r.htm> (accessed 1 November 2004).

## 3 Causal Factors of Conflict

In this section, we explore the factors and data related to the study of state conflict as preparation for forthcoming efforts to apply new methods, such as agent-based simulation, in the design of new models of conflict. We revisit recent work by O'Brien (2002) in which he forecasts the likelihood of country instability. Our purpose is to identify measurable factors that are understood or believed to influence the likelihood of state conflict. This effort will help us in formulating new models of conflict, particularly in gaining understanding about how climatic changes might affect the social or political stability in different regions or countries.

Rather than simply reviewing the work that has already been presented, critiqued, and expanded in our references, we have actively collected data from the referenced sources, reproduced some of the work, and explored alternative models. We produce and present our findings primarily to provide empirical descriptions of some of the factors and principles that relate to state conflict. We do not present our work as being complete or without error. We provide neither detailed literature reviews nor descriptions of data and empirical methods, but rather leave the reader to acquaint herself with such information from our references.

We begin our exploration of the factors and data related to the study of state conflict with a brief overview in Section 3.1 of conflict forecasting. Section 3.2 discusses how the Conflict Simulation Model (KOSIMO) database is used in conflict analysis, explains how we applied the FASE (Fuzzy Analysis of Statistical Evidence) pattern classification algorithm to reproduce some of O'Brien's results, and shows how we extended O'Brien's analysis to incorporate the occurrences of natural disasters. Section 3.3 describes our use of traditional statistical methods to analyze O'Brien's model of instability and provides an interpretation of the results. In Section 3.4, we depart from the KOSIMO database, which projects orthogonal forms of conflict into a single subjective index, and explore more precise and objective measures of conflict in relation to the independent variables.

### 3.1 Overview of Conflict Forecasting

#### 3.1.1 Background

In 1994, Vice President Al Gore initiated the State Failure Task Force (SFTF) to identify "early warnings" of conflict. The SFTF primarily used methods such as logistic regression, neural networks, and genetic algorithms to identify data patterns that might serve as early warnings.<sup>1</sup>

In response to the SFTF, King and Zeng (2001) offer corrections and extensions to the SFTF findings. A thorough discussion of the background, methods, and issues related to these efforts is documented by King and Zeng and by O'Brien (2002).

---

<sup>1</sup> From O'Brien 2002.

O'Brien (2002) extends this line of work primarily by introducing the use of FASE developed by Chen (1995, 2000). The FASE method is intended to classify states according to their likelihood of conflict. O'Brien introduces this method to overcome some of the problems that plague multinomial logistic regression, such as multicollinearity and incomplete records.

All of the efforts above involve forecasting models, which use historical data to extrapolate the likelihood of state conflict under potential future conditions. Indeed, O'Brien *validates* his approach on the criteria of forecasting performance metrics.

### 3.1.2 Forecasting and Causal Relationships

Although we explore forecasting models throughout Section 3, we must clarify that this exercise is not intended to improve upon the existing forecasting methods, but rather to help us understand and explain causal relationships for incorporation into fundamentally new methodologies. Regardless of our intent, we explore forecasting models because there is a link between forecasting and causal structure. King and Zeng (2001) provide an illuminating discussion of this link.

## 3.2 FASE Analysis of Instability

O'Brien analyzes "instability," which includes actual conflict as well as nonviolent crises and "war-in-sight." He uses the analogy of forecasting the "oiliness of the oily rags," whether or not a spark occurs to set them ablaze. That is, O'Brien seeks to forecast the instability of a country, regardless of whether an event pushes the country into conflict. He pursues this objective by introducing a dependent variable derived from the KOSIMO database.

### 3.2.1 The KOSIMO Database

The KOSIMO database (see Pfetsch and Rohloff 2000) is maintained by the Heidelberg Institute for International Conflict Research. It was constructed under the assumption that all violent conflicts evolve from nonviolent crises for the purpose of quantifying relative regional instabilities. The database allows O'Brien to project all forms of conflict, regardless of the nature of the dispute, into a single index of intensity. This is a broad-stroke approach, which categorizes the instability of various countries annually into one of four levels of conflict intensity:

1. Latent conflict; completely nonviolent
2. Crisis; mostly nonviolent
3. Severe crisis; sporadic, irregular use of force, "war-in-sight" crisis
4. War; systematic, collective use of force by regular troops

The KOSIMO Manual provides the operational definition of conflict,<sup>2</sup> which reveals the motivation for using this database to forecast "instability":

---

<sup>2</sup> See <http://www.hiik.de/en/manual.htm>.

The term ‘conflict’ is defined as the clashing of overlapping interests (positional differences) around national values and issues (independence, self-determination, borders and territory, access to or distribution of domestic or international power); the conflict has to be of some duration and magnitude of at least two parties (states, groups of states, organizations or organized groups) that are determined to pursue their interests and win their case. At least one party is the organized state. Possible instruments used in the course of a conflict are negotiations, authoritative decisions, threat, pressure, passive or active withdrawals, or the use of physical violence and war.

Note: In contrast to a purely quantitative criterion, i.e., 1,000 battle deaths for a ‘war,’ KOSIMO uses a qualitative definition: ‘. . . some duration and magnitude’. This definition was chosen to allow for non-violent conflicts that have not (yet) led to battle deaths, but—in the eyes of the participants—have the potential to escalate into a violent conflict. Also, this definition excludes all non-national, constitutional, criminal and economic conflicts. The dividing line between a political conflict in the sense of our operational definition and any other conflict is drawn after the study of each individual conflict. (Heidelberg Institute for International Conflict Research 1998–99)

To better understand the KOSIMO data, we consider the case of Afghanistan. KOSIMO identifies a basic (underlying) conflict in Afghanistan and separates the basic conflict in Afghanistan into five partial conflicts, as listed in Table 3-1.

**Table 3-1. Partial Conflicts in Afghanistan**

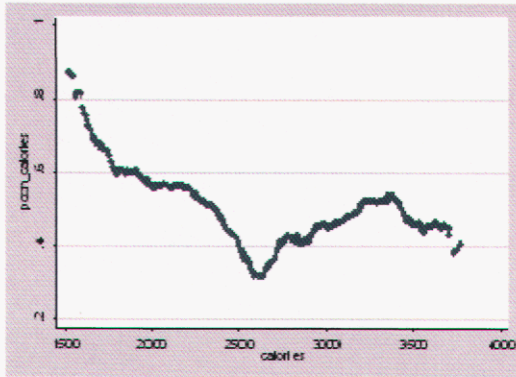
Basic Conflict	Name
Afghanistan	Afghanistan I (civil war I)
Afghanistan	Afghanistan II (Soviet intervention)
Afghanistan	Afghanistan III (civil war II)
Afghanistan	Afghanistan IV (civil war III)
Afghanistan	Afghanistan V (civil war IV)

Further inspection of the KOSIMO database (results not shown here) reveals that KOSIMO lists Afghanistan as being in a state of war (level 4) from 1979–99. However, the Correlates of War (COW) database, which identifies explicit and objectively defined occurrences of conflict, lists only the Mujahedin war from 1978–92 and classifies this conflict as a civil war for control of the central government. In the case of Afghanistan, KOSIMO identifies a state of instability from 1992–99, whereas COW does not identify measurable or well-defined occurrences of conflict during that time period.

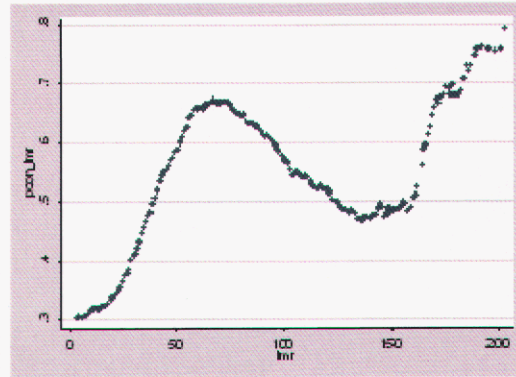
### **3.2.2 Reproduction**

We obtained data used by O'Brien from the O'Brien Web Site, which includes a full description of the data and sources (see Section 3.6. for Web address). We were able to identically reproduce O'Brien's conditional-probability plots, as shown in Figure 3-1. These are plots of the FASE conditional probability of conflict across the range of values for each of the independent variables.

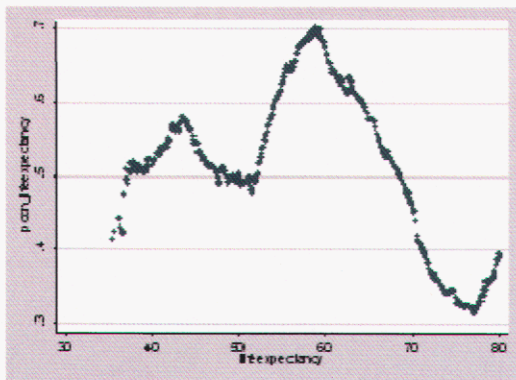




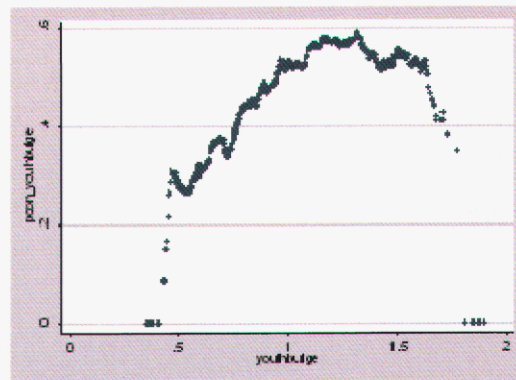
Average caloric consumption



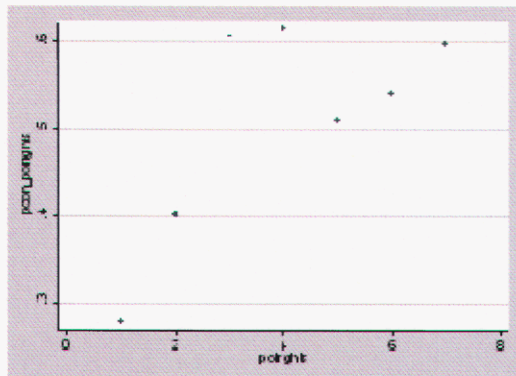
Infant mortality rate



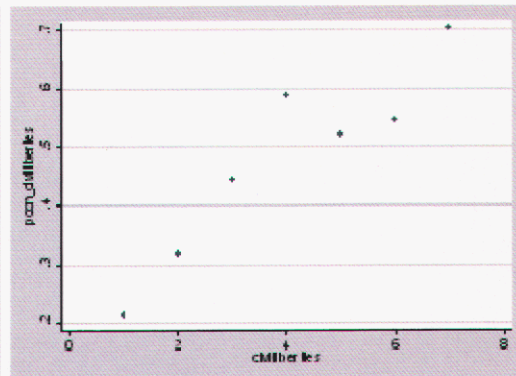
Average life expectancy



Ratio of young-to-old (youth bulge)

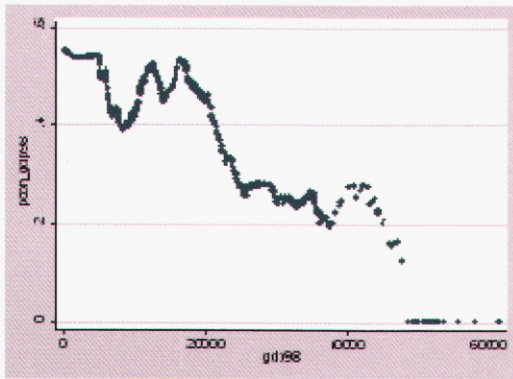


Political rights

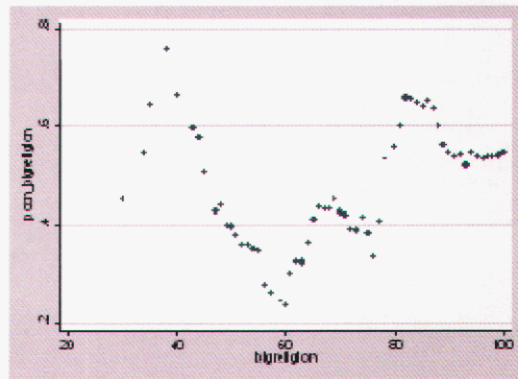


Civil liberties

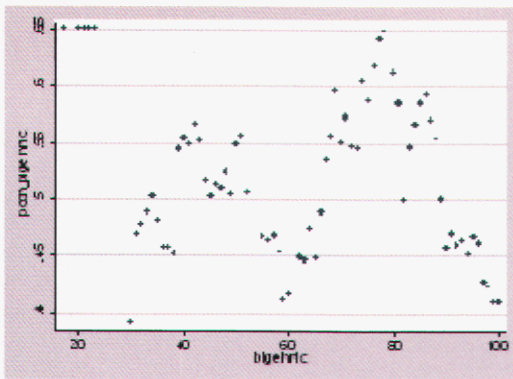
Figure 3-1. Conditional probabilities of conflict.



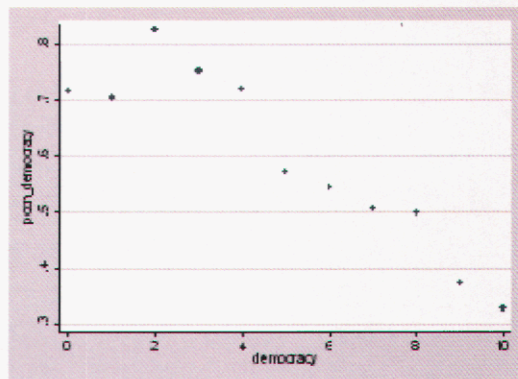
**GDP per capita**



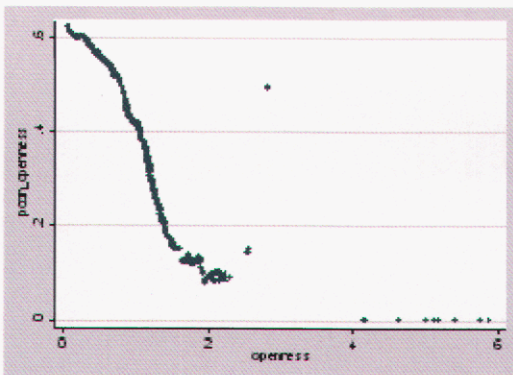
**% largest religious group**



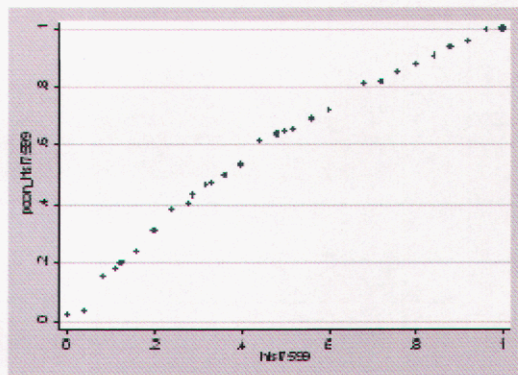
**% largest ethnic group**



**Level of democracy**



**Trade openness**

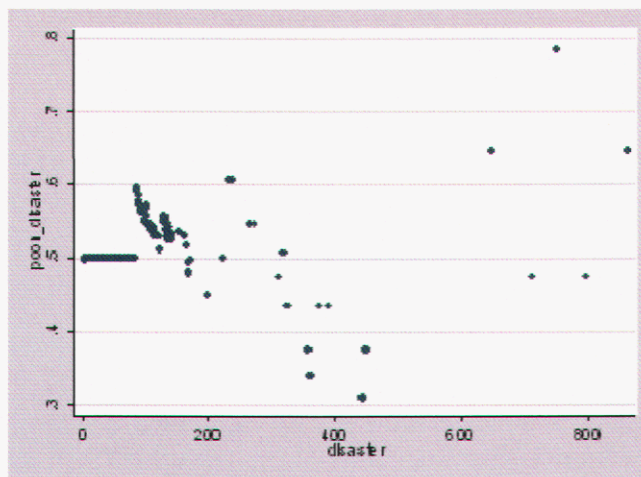


**% history in conflict**

**Figure 3-1. Conditional probabilities of conflict (Continued).**

### 3.2.3 Natural Disasters

In his discussion, O'Brien (2002) describes the independent variables as “oily rags” that require a spark to set the country ablaze. Our preliminary analysis was aimed at determining whether natural disasters might contribute to the oiliness of the rags, provide the spark, or some combination of the two. As a preliminary analysis, we created and incorporated a natural disasters variable into the FASE analysis, resulting in the plot of the conditional probability shown in Figure 3-2. The “disaster index” is a normalized linear combination of number of persons killed, injured, made homeless, and otherwise affected by the disaster, along with a dollar cost estimate from the database. The coefficients of the components were arbitrary for this first-look analysis, which we intend to revisit with more rigor.



**Figure 3-2.** Conditional probabilities of conflict versus disaster index.

Taken at face value, this plot suggests that below-median disasters have no discernible effect on the likelihood of conflict. It is notable that the largest disasters in the database are associated with the highest likelihood of conflict, but data are too sparse for any definitive statement.

Our incorporation of disasters into the FASE analysis is suspect because this approach deviates from the premise of O'Brien's model. He uses macrostructural variables to forecast *instability* irrespective of any initiating event leading to actual conflict, whereas we propose disasters as *an initiating event*. Thus, we should estimate the *interactive* effects of disasters in conjunction with other factors. We will discuss this issue further in Section 3.3.4.

## 3.3 Logistic Analysis of Instability

Here, we revisit O'Brien's model of instability, but we substitute statistical regression methods in place of FASE for analyzing the data. Statistical regression

methods have been the most common empirical methods for this line of research. Our goal is to explore second-order effects of the independent variables on instability to more accurately and usefully analyze causal relationships.

Logistic regression models estimate odds ratios for each independent variable. The odds ratios represent the marginal change in the likelihood of the event (in this case, instability) occurring in response to a unit increase in the independent variable. Each odds ratio for each parameter has an associated standard-error estimate, which we use to compute confidence intervals. When two or more independent variables are highly correlated, the standard errors can be large, leading to very wide confidence intervals, which means that we cannot say anything meaningful about the impact of those independent variables on the dependent variable.

O'Brien sought to avoid this problem of multicollinearity by introducing FASE. However, we find that we obtain reasonably stable estimates with reasonably narrow confidence intervals using logistic regression. Further, we can better parameterize relationships that change across the range of some of the independent variables. Some of our statistical results correspond to O'Brien's conclusions, but others contradict O'Brien's conclusions and correspond more closely to the results and conclusions presented by Paul Collier (2000) at the World Bank.

### **3.3.1 Method and Statistical Model**

We explore the relationship between conflict and macrostructural factors using a series of logistic regression models. We explore three classes of models, which vary based on how we define the independent variable and which country-year observations we include in the data sample. For each class of model, we explore three specifications, which vary based on how we structure the independent variables. The result is a set of nine statistical regression models for broad comparison, which demonstrate the stability of our statistical representation of the underlying relationships. We identify the nine regression models in the first column of Table 3-2 using a two-part naming convention. Each model is identified by a number corresponding to one of three classes: crisis, conflict, or intensity. The identification also includes a letter corresponding to the three specifications. Thus, for example, model class 1 identifies the crisis class and has three specifications, namely 1-A, 1-B, and 1-C. Note that both the classes and specifications are defined in more detail throughout this discussion.

**Table 3-2. Overview of Regression Models**

Model ID	Measure of Instability (independent variable)	Regression method	Social homogeneity (# regimes)	Prior conflict lag indicator (included)
1-A	<b>Crisis:</b> (excludes war; armed conflict with troops)	Logistic	2	No
1-B			3	No
1-C			3	Yes
2-A	<b>Conflict</b> (includes war; armed conflict with troops)	Logistic	2	No
2-B			3	No
2-C			3	Yes
3-A	<b>Intensity</b> (maps KOSIMO level to a unit interval)	Linear	2	No
3-B			3	No
3-C			3	Yes

Recall that regression analysis requires the analyst to define a dependent variable as a mathematical function of one or more independent variables. Based on the explicit form of the function selected by the analyst, the regression uses a best-fit technique to identify the parameters that quantify the relationship between the dependent variable and each of the independent variables. The estimated relationships will vary based on the specified functional form, the choice of dependent and independent variables, and the selection of sample data. We vary these choices to obtain the nine variations listed in Table 3-2. The remainder of Section 3.3.1 is devoted to describing the method and reasoning behind the nine regression models described in Table 3-2. Section 3.3.1.1 describes the measures of instability listed in the second column of the table. These measures directly correspond to the regression methods listed in the third column of the table.

Section 3.3.1.2 describes the potential effect on instability of dominant social groups. We estimate this effect based on the proportion of the population belonging to the largest social group, and allow this effect to differ when the proportion falls into one of two or three ranges. The largest social group falls into the first range when it does not constitute a substantial majority of the population, it falls into the second range when it does constitute a substantial majority, and it falls into the third range when it constitutes an overwhelming majority encompassing nearly the entire population. The fourth column in Table 3-2 indicates whether we allow for two ranges (first and second) or three ranges for estimating the marginal impact of an expanding largest social group on instability.

Section 3.3.1.3 describes the role of prior instability as a factor for current instability, corresponding to the lag indicator listed in the fifth column of the table. Section 3.3.1.4 lists the independent variables obtained from O'Brien for use in this analysis.

The regression results for model class 1 (crisis) and model class 2 (conflict) are presented and interpreted in Section 3.3.2, followed by the results for model class 3 (intensity) in Section 3.3.4.

### 3.3.1.1 Dependent Variable: Instability

The three classes of models listed in Table 3-2 differ, among other ways, by the choice of dependent variable. Equation (3.1) defines a binomial dependent variable for model class 1. The variable equals 0 when there are no crises (KOSIMO level 1), it equals 1 in the occurrence of crisis (KOSIMO level 2) or severe crisis (KOSIMO level 3), and it excludes all country-year observations corresponding to armed conflict (KOSIMO level 4).

$$crisis = \begin{cases} 0, & \text{KOSIMO} = 0 \\ 1, & \text{KOSIMO} = 2, 3 \\ -, & \text{KOSIMO} = 4 \end{cases}. \quad (3.1)$$

In the second model, we define instability as a binomial occurrence of any form of conflict. Equation (3.2) defines a binomial dependent variable for model class 2. The variable equals 0 when there are no crises (KOSIMO level 1), and it equals 1 in the occurrence of crisis (KOSIMO level 2), severe crisis (KOSIMO level 3), or armed conflict (KOSIMO level 4).

$$conflict = \begin{cases} 0, & \text{KOSIMO} = 1 \\ 1, & \text{KOSIMO} = 2, 3, 4 \end{cases}. \quad (3.2)$$

We excluded observations corresponding to armed conflict in model class 1 and included such observations in model class 2 to determine whether such differences in specification lead to vastly different estimates of causal relationships. Under O'Brien's supposition that various forms of conflict are linked and can be mapped to a linear representation of instability, we would not expect the underlying causal relationships to vary between the two models. However, if different forms of conflict are orthogonal with respect to underlying causal factors, then the results should vary between these two models.

To further examine O'Brien's supposition, we explored a very crude third model (model class 3) in which we define *intensity* by projecting instability into a unit interval:

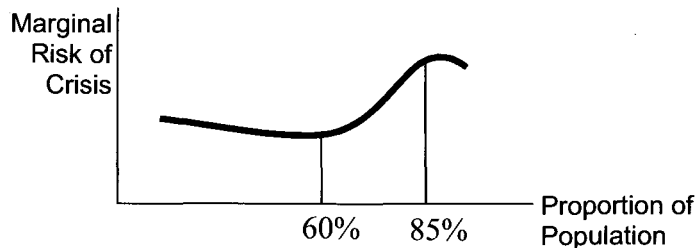
$$intensity = \begin{cases} 0, & \text{KOSIMO} = 0 \\ 1/3, & \text{KOSIMO} = 2 \\ 2/3, & \text{KOSIMO} = 3 \\ 1, & \text{KOSIMO} = 4 \end{cases}. \quad (3.3)$$

### 3.3.1.2 Social Groups

O'Brien and others, like Collier at the World Bank, suggest that conflict can arise when one social group, such as an ethnic group or a religious group, encompasses a

majority of the population and can therefore oppress minority groups. Indeed, the FASE plots in Figure 3-1 corresponding to the percent of largest ethnic group and the percent of largest religious group appear to demonstrate an increase in instability (conditional probability of conflict) as the percentage of those groups increases from 60% to the range of 75–90%.

The suggested causal effect of social-group size, which is supported by the FASE plots, implies that the impact of social-group size on instability varies with the level of homogeneity. In a heterogeneous society, there is no majority to oppress the rest of the population. In a society that is almost completely homogeneous, there are not sufficient minorities to respond to oppression in a significant manner. In the intermediate range, however, when the largest social group constitutes a substantial majority, we expect to see greater instability. Figure 3-3 illustrates this expected varying homogeneity-instability relationship, with a rising risk of crisis in the intermediate range as the largest social group expands from 60% to 85% of the population.



**Figure 3-3.** Impact of ethnic homogeneity on risk of crisis.

We use linear spline functions to allow for such nonlinear relationships, as depicted in Figure 3-3, in our regression models. We estimate different relationships between instability and the proportion of the population belonging to the largest social (ethnic or religious) group, depending on whether the largest group constitutes a majority, and whether a significant proportion of the population belongs to minority groups.

In our models, we (somewhat arbitrarily) asserted that the largest social group constitutes a majority if it exceeds 60% of the population. We also (somewhat arbitrarily) asserted that minority groups were *significant* if the largest social group did not exceed 85%. Of course, there are many more rigorous statistical methods for determining such thresholds. A better approach might be to search for the most [statistically] efficient change point. We also attempted to use nonlinear models of tipping points, such as second-order and cross-order variables. However, our approach was expedient, seemed reasonable given the anecdotal discussions in the literature, and yielded confirmatory, albeit preliminary, results.

### 3.3.1.3 Prior Instability

O'Brien incorporates a variable calculated as the percentage of history spent in conflict; the corresponding FASE conditional-probability plot is shown in the last graph of Figure 3-1 (labeled "% history in conflict"). O'Brien describes this variable as a cross-

country, and therefore culturally dependent, proxy for the propensity of a country to resolve conflict through violence. Collier (2000) also looks at the propensity of conflict, but from an intertemporal, rather than a cross-sectional, perspective. He finds that immediately after the end of hostilities associated with civil war there is a 40% chance of further conflict.

We created an autoregressive indicator variable, similar to that used by Collier, to denote whether a country was in a state of KOSIMO instability in the previous year. We included this variable in three of nine regressions, as specified in the fifth column of Table 3-2; the corresponding models are 1-C, 2-C, and 3-C.

#### **3.3.1.4 Selection of Variables**

We model instability as a function of several independent variables from O'Brien's data, as follows:

1. index of civil liberty
2. index of political rights
3. index of democracy
4. trade openness
5. youth bulge
6. life span
7. caloric intake
8. proportion of largest ethnic group
9. proportion of largest religious group
10. prior instability

We use stepwise procedures to help determine which independent variables to include in the nine regression models. This process removes variables from the model based on the standard error of their influence on instability. We specified a critical P-value of 0.15 for removing variables from the model. Removal of a variable implies that the variable does not significantly influence instability in a consistent fashion.

#### **3.3.2 Results for Regression Models**

Table 3-3 shows the results for all nine regression models. Each independent variable listed in the leftmost column has up to nine corresponding parameters for the nine regression models, with P-values displayed underneath in parentheses. An explanation of how to interpret the results follows the table.



**Table 3-3. Regression Results**

Dependent variable:	(1) Crisis			(2) Conflict			(3) Intensity		
N	1423			1567					
No. countries	109			111					
Years per country	[2,25] $\mu=13$			[4,25] $\mu=14$					
Model	1-A	1-B	1-C	2-A	2-B	2-C	3-A	3-B	3-C
Linkage	logistic						linear		
R <sup>2</sup>							.31	.31	.19
Pseudo R <sup>2</sup>	0.19	0.21	0.75	0.20	0.22	0.75			
Civil Liberty	2.093 (.000)	2.187 (.000)	1.539 (.021)	2.319 (.000)	2.399 (.000)	1.726 (.001)	0.108 (.000)	0.104 (.000)	0.022 (.000)
Political Rights	---	---	---	---	---	---	---	---	---
Democracy	1.154 (.000)	1.155 (.000)	1.090 (.021)	1.142 (.000)	1.157 (.000)	1.091 (.011)	0.015 (.000)	0.016 (.000)	0.003 (.020)
Trade openness	0.148 (.000)	0.186 (.000)	0.448 (.073)	0.136 (.000)	0.163 (.000)	0.358 (.014)	-0.16 (.000)	-0.125 (.000)	---
Life expectancy	0.973 (.013)	---	---	---	---	---	---	0.004 (.002)	---
Youth bulge	2.755 (.004)	5.451 (.000)	6.492 (.011)	1.793 (.060)	3.058 (.001)	4.503 (.031)	---	0.132 (.001)	0.039 (.060)
Colonies	1.14 (.000)	1.12 (.000)	1.14 (.005)	1.10 (.000)	1.08 (.000)	1.09 (.025)	0.004 (.036)	---	---
Prior conflict	424.9 (.000)			353.4 (.000)			0.572 (.000)		
Intercept							0.194 (.225)	0.105 (.542)	0.005 (.934)
<b>% largest ethnic group</b>									
Slope	0.945 (.000)	0.938 (.000)	0.971 (.014)	0.958 (.000)	0.950 (.000)	0.965 (.049)	---	-0.002 (.085)	0.001 (.002)
Indicator	0.088 (.010)	<.001 (.000)	---	---	0.002 (.001)	---	0.413 (.000)	-0.841 (.000)	---
Interaction	1.060 (.000)	1.140 (.000)	---	1.022 (.000)	1.116 (.000)	1.018 (.066)	-0.005 (.000)	0.013 (.000)	---
Indicator	691 (.041)		4.107 (.009)	1060 (.019)		---	1.132 (.002)		0.304 (.092)
Interaction	0.921 (.028)		---	0.912 (.008)		---	-0.016 (.000)		-0.004 (.023)
<b>% largest religious group</b>									
Slope	0.821 (.000)	0.819 (.000)	0.850 (.006)	0.835 (.000)	0.827 (.000)	0.845 (.000)	-0.009 (.001)	-0.011 (.000)	-0.003 (.002)
Indicator	<.001 (.000)	<.001 (.000)	<.001 (.000)	<.001 (.000)	<.001 (.000)	<.001 (.000)	-0.636 (.000)	-1.06 (.000)	---
Interaction	1.266 (.000)	1.367 (.000)	1.314 (.000)	1.235 (.000)	1.312 (.000)	1.264 (.000)	0.012 (.000)	0.019 (.000)	0.001 (.001)
Indicator	22K (.000)		14M (.004)	24K (.000)		437K (.010)	1.178 (.000)		0.058 (.003)
Interaction	0.885 (.000)		0.824 (.002)	0.890 (.000)		0.863 (.010)	-0.013 (.000)		---

For model classes 1 and 2, the parameters listed in Table 3-3 are called *odds ratios*. To interpret the results, first recall that *odds* in standard logistic regressions of a dichotomous variable represent the likelihood of an occurrence relative to a nonoccurrence. The odds ratio represents the change in the odds due to a unit increase in the independent variable. Odds ratios greater than one imply that the odds, and therefore probability, are increasing; odds less than one imply the odds are decreasing. In this analysis, in which the dependent variable has four levels of intensity, the “odds” are the likelihood that instability will escalate to the next level of intensity. For example, the estimated odds ratio for life expectancy is 0.97, or 97%, which implies that a unit increase in life expectancy will be associated with a decrease in the odds of conflict by 3%. In this case, a unit increase in life expectancy is represented by *one year*. So, a one-year increase in life expectancy will decrease the odds of conflict by 3%.

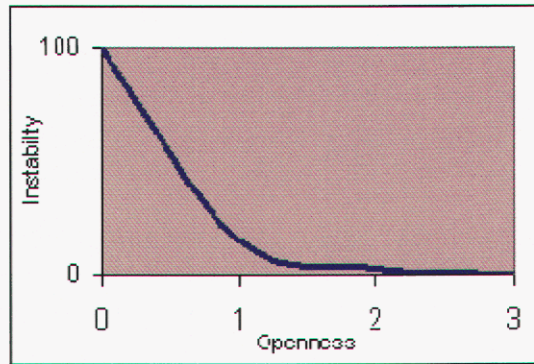
Empty cells in Table 3-3 denote that an independent variable was omitted from the full model by design. Cells containing “--” denote that the independent variable was removed from the model by the stepwise procedure. For model classes 1 and 2 (logistic regression), the listed parameters represent odds ratios; for model class 3 (linear regression), the listed parameters are coefficients.

The stepwise procedure removed the *political rights* variable from every regression model, probably because this variable is 90% correlated with both the *civil liberties* variable and the *democracy* variable. The stepwise procedure also removed the *life expectancy* variable from most of the regressions. This variable is most highly correlated with the *calories* variable, but only at 75%. Since we have no strong theory or anecdote to suggest that life expectancy is strongly related to instability, we are not surprised that it was dropped.

### 3.3.2.1 Interpreting the Logistic Results

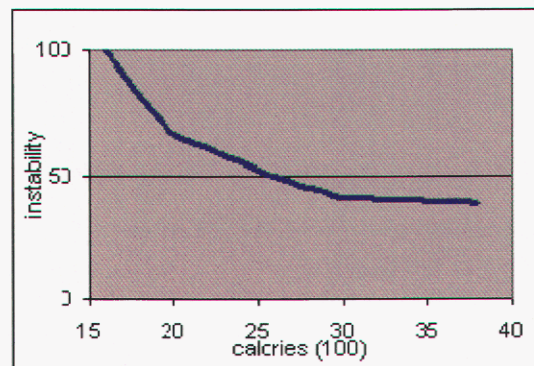
The odds-ratio estimates are relatively stable across the six logistic regressions for model classes 1 (crisis) and 2 (conflict). We interpret the results as follows:

- The *civil liberties* index has seven possible values. We find that an incremental decrease in the civil liberties index roughly doubles the odds of instability. We find this result in Table 3-3 by examining the row named “civil liberty,” for which the first six columns (models 1 and 2) show odds ratios near 2. In the case of model 1-C, the odds ratio is only 1.54, which implies that the odds of instability increase by 54% in response to an incremental decrease in civil liberties.
- The *democracy* index has 21 possible values. We find that an incremental increase in the democracy index increases the odds of instability by roughly 10–15%.
- The *trade openness* variable ranges from 0.09 to 2.82 in our sample. We find that a unit increase in trade openness, say, from 1.0 to 2.0, decreases the odds of instability by 85%. Figure 3-4 illustrates that most of the marginal risk is removed by progressing halfway up the trade openness scale.



**Figure 3-4.** Likelihood of instability versus trade openness.

- The *youth bulge* ranges from 0.46 to 1.9 in our sample. We find that increasing the youth bulge from the bottom to the top of this range roughly triples the odds of instability.
- The *caloric* intake ranges from 1,602 to 3,756 in our sample. The results imply that a 100-calorie increase actually increases the marginal likelihood of conflict by 8–14%. To better capture the true relationship between calories and instability, we ran a subsequent regression model (not shown) that estimated the relationship between calories and instability for three segments of caloric intake: (1) less than 2,000 calories, (2) from 2,000 to 3,000 calories, and (3) above 3,000 calories. This modification had negligible impacts to our other conclusions, but estimated that marginal instability declines over the lower range of caloric intake and stabilizes as a nation approaches the median caloric intake of 2,500 calories, as illustrated in Figure 3-5. The findings from this modification suggest that increases in caloric intake decrease the marginal likelihood of conflict.



**Figure 3-5.** Likelihood of instability versus average daily calories.

### 3.3.2.2 Social Groups

We analyzed the marginal likelihood of instability with respect to social-group size for three ranges of group size. In range 1, the largest social group encompasses less than 60% of the population; in range 2, the largest social group encompasses more than 60% but less than 85% of the population; and in range 3, the largest social group encompasses more than 85% of the population.

We estimated the marginal instability with respect to both *ethnic* and *religious* groupings. As expected, the marginal instability in both cases increases for majority group size in the range of 60% to 85%, then decreases for majority group size above 85%.

The plots in Figure 3-6 show the estimated relationships between group size and instability. We are particularly interested in the red plots obtained from model 2-B, which support our proposition that instability will increase with respect to a group majority until the minority becomes too small to respond to oppression in a significant manner. For comparison, we included the blue plots obtained from model 2-A, which demonstrate the potential errors that can result if one fails to sufficiently allow for nonlinear relationships.

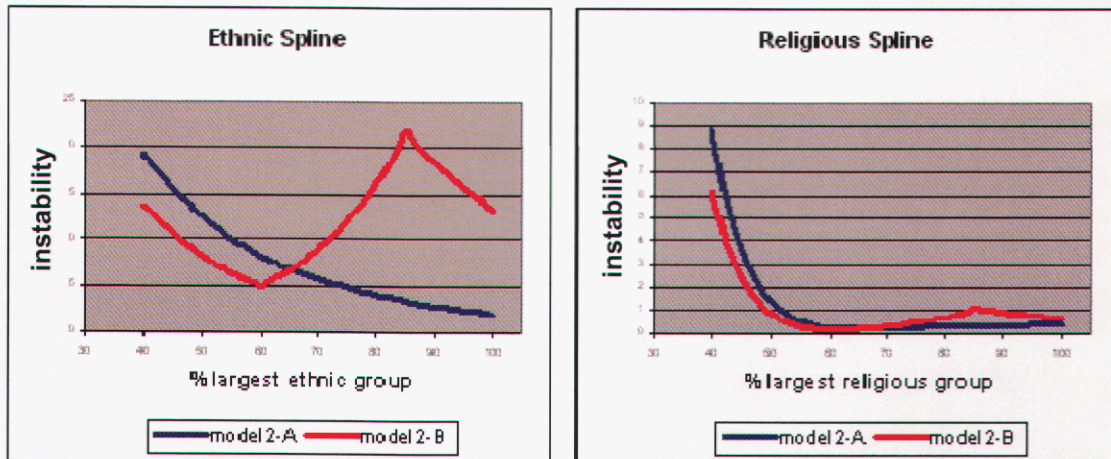


Figure 3-6. Marginal likelihood of conflict versus group size.

### 3.3.3 Intensity of Instability: Model and Results

We ran a very crude linear regression model of intensity of instability by transforming the maximum KOSIMO value into a new variable *intensity*  $\in [0, 1]$ . The stepwise procedure omitted *democracy*, *youth bulge*, *political rights*, and *trade openness* from the model. The estimated coefficients for all independent variables are significant at the 95% confidence level. The  $R^2$  of 70% indicates that the model explains 70% of the variation in intensity across countries and years.

**Table 3-4. Regression Results for Intensity**

Source	SS	df	MS	Number of obs = 1567		
Model	131.564478	9	14.6182754	F( 9, 1557) = 400.66		
Residual	56.808213	1557	.036485686	Prob > F = 0.0000		
				R-squared = 0.6984		
				Adj R-squared = 0.6967		
				Root MSE = .19101		
Total	188.372691	1566	.120289075			

intensity	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
priorconfl~t	.5828934	.0107285	54.33	0.000	.5618496	.6039373
libert	.0152409	.003883	3.93	0.000	.0076245	.0228573
calor	-.0000373	.0000153	-2.43	0.015	-.0000674	-7.21e-06
life	.0015866	.0007498	2.12	0.034	.0001159	.0030573
ethnic	.0015435	.000767	2.01	0.044	.0000391	.0030479
ethnicx	-.0050837	.0009852	-5.16	0.000	-.0070161	-.0031513
ethnic1	.3566995	.0665896	5.36	0.000	.2260848	.4873143
relign	-.0015189	.0007663	-1.98	0.048	-.0030219	-.0000159
relignx	.001002	.000368	2.72	0.007	.0002801	.0017238
constant	-.0310817	.0610057	-0.51	0.610	-.1507437	.0885802

This model accurately predicted the KOSIMO intensity level in 78% of the cases in the sample.

Note that in Table 3-4 and in other similar tables presented subsequently in this discussion, we list variables beginning with “ethnic” and appended with a number, an alphabetic character, or both. The definitions of these variables are as follows:

- ethnic1 = 0 when ethnic < 60%
- ethnic1 = 1 when ethnic > 60%
- ethnic2 = 0 when ethnic < 85%
- ethnic2 = 1 when ethnic > 85%
- ethnicx = ethnic \* ethnic1
- ethnicx2 = ethnic \* ethnic2

Those variables appended with a single number (ethnic1 and ethnic2) are Bernoulli variables, which are “yes/no” or “0/1” variables. Such variables thus have two possible outcomes.

We use the same definitional approach described above for ethnic-related variables for those variables listed in the tables that are associated with religion, e.g., relign, relign1, relignx, etc.

### 3.3.4 Natural Disasters

We explore the analogy of oily rags waiting for a spark to set them ablaze by introducing natural disasters into our analysis. We suggest that natural disasters escalate instability when a significant group of survivors is impacted by death, injury, homelessness, or similar effects. Specifically, we introduce four new variables representing the three-year average of the three previous years of the number of people (1) killed, (2) injured, (3) left homeless, and (4) otherwise affected by a natural disaster.

We extracted data from the Emergency Events Database (EM-DAT), which is made available on the Worldwide Web by the Centre for Research on the Epidemiology of Disasters (CRED) of the WHO (World Health Organization). These data are available in spreadsheet form (see CRED 2004).

When previous years were unavailable, we used a one- or two-year average instead of a three-year average. We introduced these variables into the conflict model identified as model 2-B. Table 3-5 shows the regression results.

**Table 3-5. Logistic Estimation of the Occurrence of Conflict Including Variables Representing Number Killed and Injured in Disasters**

Logit estimates						Number of obs	=	1567
Log likelihood = -759.82368						Pseudo R2	=	0.2513
conflict	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]			
libert	2.451068	.2137089	10.28	0.000	2.066042	2.907849		
demo	1.150886	.0217512	7.44	0.000	1.109034	1.194317		
open	.2418304	.0566301	-6.06	0.000	.1528209	.3826828		
youth	2.949534	1.012244	3.15	0.002	1.505329	5.779303		
calor	.8888329	.0390172	-2.68	0.007	.8155579	.9686913		
calor2	1.053206	.0158496	3.44	0.001	1.022595	1.084734		
calor3	1.051157	.0101168	5.18	0.000	1.031514	1.071174		
ethnic	.9389448	.0101647	-5.82	0.000	.9192323	.95908		
ethnic1	.0006801	.0013363	-3.71	0.000	.0000145	.0319914		
ethnicx	1.137359	.0310091	4.72	0.000	1.078177	1.199789		
ethnic2	1588.092	4943.214	2.37	0.018	3.55928	708580.5		
ethnicx2	.9082237	.0328139	-2.66	0.008	.846134	.9748697		
relign	.8201978	.0263593	-6.17	0.000	.770128	.8735229		
relign1	1.16e-06	2.28e-06	-6.96	0.000	2.49e-08	.0000545		
relignx	1.298315	.0465447	7.28	0.000	1.21022	1.392822		
relign2	10168.69	25473.69	3.68	0.000	74.97423	1379171		
relignx2	.9017984	.0256697	-3.63	0.000	.8528644	.9535401		
killed	1.024814	.0105918	2.37	0.018	1.004263	1.045785		
injured	1.046087	.0115067	4.10	0.000	1.023775	1.068884		

The impacts of a natural disaster are generally local or regional, and their likelihood to escalate instability depends on the particulars of the region rather than on the size of the population residing within the national border. Therefore, we do not normalize with respect to population.<sup>3</sup>

We estimated a regression model using both linear and log-linear transformations of the disaster variables. In both cases, the stepwise procedure found that the variables for *homeless* and *otherwise affected* were not significant, and the variables for *killed* and *injured* had an increasing effect on instability. The insignificance of the variables for homeless and otherwise affected was clearest in the model of log-linear transformations, where the stepwise procedure discarded those two variables at 64% and 70% significance levels, respectively. There are several reasons why these variables might have such high

<sup>3</sup> The previous analysis in Section 3.2.3 was based on a composite variable that was normalized with respect to population.

standard errors, including the likelihood that the data for these two variables are not reliable and consistent across countries. Regardless of why those two variables were discarded, we ultimately decided that the model presented above, resulting from log-linear transformations of the data, was the more conservative model. The variables for killed and injured were constructed as follows: If  $n_{it}$  represents the number of deaths in the  $i^{\text{th}}$  country in year  $t$ , then our regression variable for deaths is  $X_{it} = \ln((n_{it} + n_{it-1} + n_{it-2})/3)$ .

As before, the stepwise procedure discards *political rights* and *life expectancy* from the model. Additionally, the procedure discards *homeless* and *otherwise-affected* casualties. The relationships estimated in the previous model (2-B) hold. Additionally, we find that a percentage increase in the number killed and injured increases the odds of instability by 2% and 4%, respectively.

### 3.3.5 Extrapolation and Forecasting

O'Brien presents his model as a tool for forecasting future instability from prior historic factors. He uses split samples whereby he *trains* the model using data from an initial time period, extrapolates (forecasts) to a subsequent time period, and compares his forecasts with actual KOSIMO instability values. We conducted a similar experiment using our logistic model.

In preparation for a split-sample forecasting experiment, we first explored the model's ability to predict within-sample, and to extrapolate cross-sectionally to countries and years that were excluded from the regression sample due to missing data. We ran two regression models for this preliminary step.

First, we ran the model from Section 3.3.4 with 1567 complete country-year observations, denoted as model A in Table 3-6. Then we extrapolated the probability of instability in the 1938 country years that had been omitted from the regression by populating missing data with the global mean of the respective independent variable. This model accurately forecasted instability in 77% of the country years in the regression sample and in 70% of the country years that were not in the regression sample.

Second, we added an additional variable to account for prior conflict. O'Brien and Collier both include prior conflict as a factor in their model, arguing that prior conflict propagates instability. O'Brien addresses the issue by including a variable to represent the proportion of sample years that a country has spent in instability. We excluded that variable from our model because although statistically significant, the variable obtained an odds ratio approximately equal to one and added very little to the  $R^2$ . Rather, we address the issue here by introducing an indicator variable to indicate whether the country was unstable in the previous year (see model B in Table 3-6). This variable increased the  $R^2$  from 25% to 75%, increased the forecast accuracy within-sample from 77% to 95%, and increased the forecast accuracy out-of-sample from 70% to 89%.

**Table 3-6. Statistical Impact of Prior-Conflict Variable**

Model	Includes prior-conflict variable	R <sup>2</sup>	Country years	Sample	Extrapolated	Total
			# records	1567	1938	3505
A	No	25%	% correct	77%	70%	73%
B	Yes	75%	% correct	95%	89%	92%

Having explored our baseline, we now conduct a variation of one of O'Brien's experiments in which he estimated based on a sample from 1975–89 and forecasted for the period 1990–94. Table 3-7 shows the accuracy of our forecasts using model A, which excludes a variable denoting prior instability. Although our sample is much smaller than O'Brien's, we obtain a forecast accuracy of 74% compared to O'Brien's accuracy of 77%. The number of country-year observations is listed in parentheses below the forecast-accuracy percentage.

**Table 3-7. Forecast Accuracy Excluding the Prior-Conflict Variable**

Model A	Sample years	Forecasted years	O'Brien's forecast
Sample countries	78% (552)	74% (350)	77%
Extrapolated countries	65% (950)	63% (390)	

Table 3-8 shows that accuracy increases substantially when we include our prior-conflict variable, increasing accuracy from the 60–80% range in model A to the mid-90% range in model B.

**Table 3-8. Forecast Accuracy Including the Prior-Conflict Variable**

Model B	Sample years	Forecasted years	O'Brien's forecast
Sample countries	94% (552)	95% (350)	77%
Extrapolated countries	90% (950)	91% (390)	

We find that including a variable to represent prior instability can greatly increase the forecast accuracy of a model, even when using a much smaller sample of countries and even when extrapolating to other countries outside of the original sample.



### 3.3.6 Forecasting Impacts of Caloric Reductions

We are ultimately interested in understanding the impacts of climate change on social instability. One example of such a climate change might be desertification, leading to food reduction. As a crude experiment, we compare instability forecasts under sample conditions against the instability forecasts when daily calories per capita are arbitrarily reduced.

For this experiment, we exclude the autoregressive variable called *prior conflict* for a more *honest* representation of the underlying factors of instability. We further explore the role of consumption more deeply by introducing the following interaction terms: *calories·civil liberty*, *calories·political rights*, *calories·democracy*, *calories·youth*, and *calories·life*.

We estimated the model using data from 1975 through 1989. The results are shown in Table 3-9. The stepwise procedure excluded all terms containing *youth bulge* and *life expectancy*, as well as the first-order terms *democracy* and *civil liberty*. As before, we find that calories reduced the odds of instability at a decreasing rate, but this model also indicates that calories reduce the odds of instability faster in those countries with lesser degrees of democracy, civil liberty, and political rights.

**Table 3-9. Logistic Estimation of the Occurrence of Conflict**

Logit estimates		Number of obs =		552	
Log likelihood = -237.4921		Pseudo R2 =		0.3112	
conflict	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
calor	.4680991	.0562648	-6.32	0.000	.3698492 .592449
calor2	1.105046	.0324306	3.40	0.001	1.043277 1.170472
calor3	1.119827	.0246717	5.14	0.000	1.0725 1.169242
cal_demo	1.005867	.0018278	3.22	0.001	1.002291 1.009456
cal_libert	1.036536	.0082882	4.49	0.000	1.020418 1.052908
cal_rights	1.057908	.0234404	2.54	0.011	1.012949 1.104863
rights	.1923973	.0907347	-3.49	0.000	.0763435 .4848708
open	.0765448	.0309054	-6.36	0.000	.0346927 .1688859
ethnic	.8949474	.0173601	-5.72	0.000	.861561 .9296277
ethnicx	1.191992	.0477947	4.38	0.000	1.101903 1.289447
ethnic1	.0000856	.0002363	-3.39	0.001	3.82e-07 .0191915
ethnic2	.1309745	.0992778	-2.68	0.007	.0296471 .5786175
relign	.7846355	.0468963	-4.06	0.000	.6978999 .8821506
relignx	1.461669	.0959263	5.78	0.000	1.285247 1.662309
relign1	8.14e-10	2.87e-09	-5.93	0.000	8.07e-13 8.20e-07
relignx2	.978393	.0071788	-2.98	0.003	.9644236 .9925647

To compare the forecasted impacts of caloric reduction on instability, we first obtained a set of fitted instability values for the years 1990 through 1994 called  $\hat{y}_{base}$ . We then reduced the calories for all country years by 10% and obtained a new set of fitted instability values called  $\hat{y}_{drop}$ . We then sorted the observations according to  $\hat{y}_{base}$ . Figure

3-7 compares the forecasted probability of conflict under actual calorie levels (blue) against the forecasts when calories are reduced by 10% (red). The horizontal axis is simply the rank order of the observations when sorted according to  $\hat{y}_{base}$ . We find that the caloric reduction increases the estimated odds of instability in every country year.

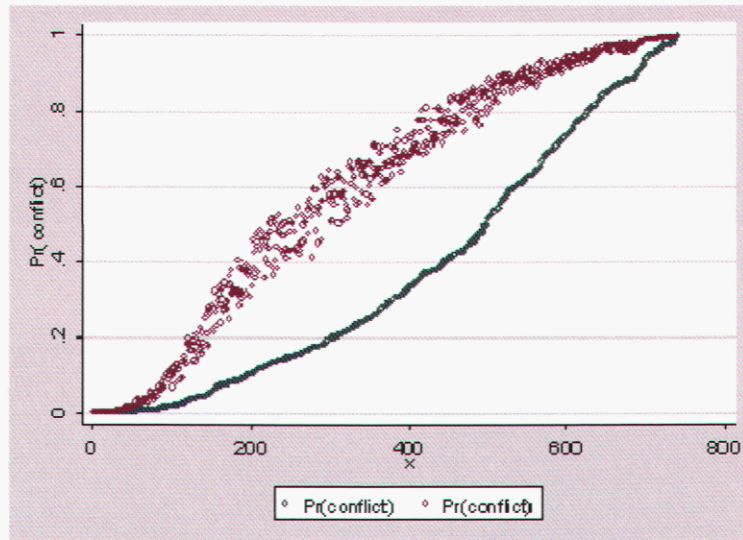
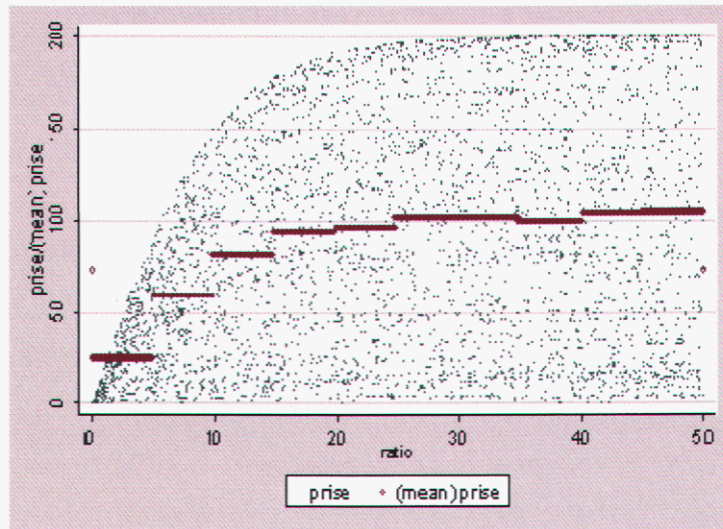


Figure 3-7. Logistic response to caloric reductions.

### 3.3.6.1 Calorie-Instability Relationship

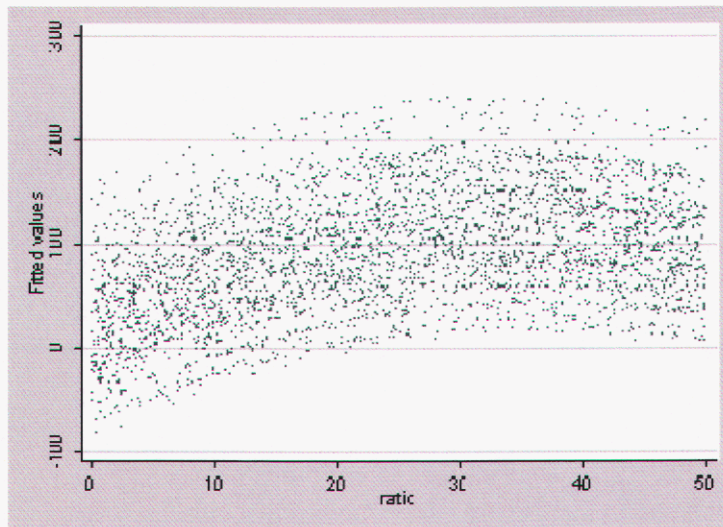
We made the previous comparison by assuming a 10% across-the-board reduction in calories. We will now estimate a relationship between stability and calories. We do so by appending multiple fitted values of the odds of instability for each country year for different-sized caloric reductions.

Figure 3-8 shows the forecasted percent increase in the odds of instability, called *prise*, in response to various percent reductions (0% to 50%) in daily calories per capita. We find that the odds of instability generally double for the first 20% reduction in calories. To better visualize the impacts of random caloric reductions on expected instability, we divide the percent reductions into ten 5% intervals; for example, the first interval includes caloric reductions from 0% to 5%, the second interval includes caloric reductions from 5% to 10%, etc. We compute the mean percent increase in the odds of instability for each interval, called *mean(prise)*, and plot these in the odds of instability by about 75%, but the average impact of larger-percentage caloric reductions converges at about a 100% increase in the odds of instability



**Figure 3-8.** Forecasted probabilities of conflict resulting from caloric reductions.

We complete this crude extrapolation exercise by fitting a regression (not shown) of the percent rise in  $\text{Pr}(\text{instability})$  with respect to several variables including the percent reduction in calories. Figure 3-9 shows the marginal quadratic relationship between these; the vertical axis represents the percent rise in  $\text{Pr}(\text{instability})$ , and the horizontal axis refers to the percent reduction in calories. This chart shows that caloric reductions cause the forecasted likelihood of conflict to increase at a decreasing rate with respect to the percent reduction in calories.



**Figure 3-9.** Fitted relationship of conflict versus caloric reductions.

### 3.3.7 Remarks

These exercises demonstrate that relatively simple statistical models can generate forecasts in and out of sample, and give insight into the underlying causal structure of instability with respect to measurable macrostructural variables.

Despite this series of statistical exercises, we are skeptical of purely correlative methods for analyzing conflict factors. We suggest that fitted and actual values of instability measures should be examined in the context of case studies of the various countries in our samples.

Further, we are skeptical of these statistical models because the definition of instability provided by KOSIMO is so broad and could possibly comprise a variety of conflicting relationships with the independent variables. Therefore, although we could and should investigate the design of these models more carefully, it seems more prudent to explore the definitions of conflict.

## 3.4 Orthogonal Types of Conflict

In Sections 3.1 through 3.3, we analyzed KOSIMO instability as a function of several underlying independent variables. Although it was a useful scoping exercise, we are skeptical of the reliability and value of analyzing a single all-encompassing index of instability.

We suggest that instances of conflict fall along one or more orthogonal axes of instability (as illustrated in Figure 3-10), and that meaningful empirical study should take this into account. Collier (2000) presents a model that is consistent with our suggestion by limiting his analysis to a single axis: "civil war."

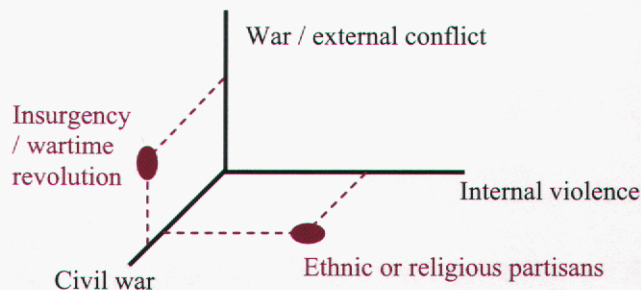


Figure 3-10. Orthogonal conflict space.

The COW (Correlates of War) database contains more objective definitions of conflict, which are clearer for our purposes than those provided by KOSIMO. However our quick perusal seems to indicate that the data are incomplete. The COW database compiles a list of records corresponding to conflicts, but this list does not examine the ongoing conflict-related characteristics of a region or country, outside of an instance of

documented conflict. Therefore, there are no defined steps for documenting and measuring the completeness of the list.

### 3.4.1 Civil War Versus Local Conflict

The COW database contains intrastate war data for 1816–1997 (see COW 2004). COW classifies intrastate wars into three categories: (1) civil war for control of the central government, (2) conflict over local issues, and (3) intercommunal conflict.

We sought to determine whether the determinants are different for each type of conflict. Thus, we merged the COW data with the O’Brien data set. Once merged, the data contained incidences of civil war and local-issues conflict, but no incidences of intercommunal conflict for the country years in the data set. We created two new Bernoulli variables (“yes/no” or “0/1” variables, which have two possible outcomes) for each country year in the O’Brien data. The first variable lists whether the state was engaged in civil war during each year, and the second variable lists whether the state was engaged in conflict over local issues. Based on several runs, we decided to exclude *political rights*, *life span*, and *youth bulge* from the option set of variables.

### 3.4.2 Logistic Model of Civil War

The statistical results are contained in Table 3-10. The estimated odds ratios for most independent variables are significant at the 95% confidence level. The pseudo  $R^2$  (a property of logistic regression that is analogous to the  $R^2$  of linear regressions) indicates that the model explains 32% of the variation in the occurrence of civil war across countries and years. This model accurately predicted civil war in about 95% of the cases in the sample.

**Table 3-10. Logistic Estimation of Occurrences of Civil War**

Logit estimates		Number of obs = 1501		Pseudo R2 = 0.32	
Log likelihood = -192.34577					
civilwar	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
libert	1.343757	.1396492	2.84	0.004	1.096125 1.647333
relign1	.0000218	.0000074	-3.17	0.002	2.85e-08 .0167479
open	.0166382	.0115927	-5.88	0.000	.0042465 .0651905
calor	.9971128	.0004565	-6.32	0.000	.9962185 .9980079
ethnic	1.033127	.0218709	1.54	0.124	.9911376 1.076895
ethnicx	.9299716	.0242554	-2.78	0.005	.8836266 .9787474
ethnic1	412.6841	697.0897	3.57	0.000	15.0592 11309.25
relign	.9444142	.0574102	-0.94	0.347	.8383371 1.063914
relignx	1.156064	.073745	2.27	0.023	1.020197 1.310025

The following provides an interpretation of the results:

- We found that an incremental move down the *civil liberties* index increased the risk of civil war by 34%.

- The stepwise procedure removed *democracy* from the list of variables.
- We found that a unit increase in *trade openness* decreases the risk of conflict by 98%.
- Unlike the previous models using KOSIMO measures of conflict, we now find that a unit increase in calories *decreases* the risk of conflict by 0.3%.
- We find that a 1% increase in the size of the largest religious group decreases the odds of conflict by nearly 6% when under the 60% threshold, which we derive by noting that the odds ratio for the religious-group variable is roughly 94%. That is, the odds of conflict after a 1% increase in the size of the largest religious group are 94% as high as (or 6% lower than) the odds before the increase in the group size. However, an equal 1% increase in group size increases the odds by 10% when above the 60% threshold, which we derive by taking the product of odds ratios for the variables named “reign” and “reignx” (which appear in several of the tables in Section 3) because reign gives the underlying odds ratio, and reignx gives the change in the odds ratio when reign > 60%. This product of these odds ratio, in this case, is roughly 110%, which implies that the odds of conflict are 10% higher after a 1% increase in group size than before the increase in group size.
- The opposite path occurs for ethnic homogeneity. We find that a 1% increase in the size of the largest ethnic group increases the odds of conflict by 3% when under the 60% threshold, which we obtain by noting than the odds ratio for “ethnic” is 103%. However, a 1% increase in the size of the largest ethnic group decreases the odds of conflict by 4% when above the 60% threshold, which we obtain from the product of “ethnic” and “ethnicx”.

### 3.4.3 Logistic Model of Local War

The statistical results for local war, identifying conflict over local issues, are contained in Table 3-11. The estimated odds ratios for all independent variables are significant at the 95% confidence level. The pseudo  $R^2$  indicates that the model explains 41% of the variation in the occurrence of local war across countries and years. This model accurately predicted local war in about 97% of the cases in the sample.

**Table 3-11. Logistic Estimation of Occurrences of Local War**

Logit estimates		Number of obs =		1812	
Log likelihood = -122.82354		Pseudo R2 =		0.417	
localwar	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
interwar	8.930313	7.123913	2.74	0.006	1.87001 42.6471
libert	2.548555	.3963734	6.02	0.000	1.878918 3.456848
relignl	1.34e-06	5.61e-06	-3.24	0.001	3.80e-10 .00476
calor	.9990791	.0004284	-2.15	0.032	.9982399 .9999191
ethnic	.778308	.0300449	-6.49	0.000	.7215936 .8394799
ethnidx	1.182376	.0310884	6.37	0.000	1.122988 1.244906
relign	.8659086	.0607887	-2.05	0.040	.7545983 .9936382
relignx	1.24813	.0938289	2.95	0.003	1.077136 1.446271

In this model, we found a strong inverse relationship between trade openness and local war. Indeed, there was a fully determined correspondence between local war and trade openness for 437 observations, which were therefore omitted from the sample by the statistics package due to lack of variation. In response, we took this relationship between local war and trade openness as given, and removed the variable from the model to estimate on a larger sample. This step of removing a significant variable from the model violates certain assumptions of the model but did not drastically affect the results.

The following provides an interpretation of the results:

- We found that an incremental 17% move down the *civil liberties* index more than doubles the risk of local war.
- The stepwise procedure removed *democracy* from the list of variables.
- Similar to the model of civil war, we now find that a unit increase in calories *decreases* the risk of conflict by 0.1%. Although statistically significant, this result implies that the effects on the likelihood of conflict due to moderate increases in caloric intake will be of negligible magnitude.
- As in the model of local war, we find that a 1% increase in the size of the largest religious group decreases the risk of conflict by 14% when under the 60% threshold, but increases the risk by 10% when above the 60% threshold.
- Unlike the model of civil war, ethnic homogeneity follows the same path as religious homogeneity. We find that a 1% increase in the size of the largest ethnic group decreases the risk of conflict by 22% when under the 60% threshold, and continues to decrease the risk by 4% when above the 60% threshold.
- We included an additional variable in the model of local war: the existence of international war. We find that local wars are nine times more likely to occur when the national government is at war with another country.

### 3.5 Remarks

We found significant relationships between various macrostructural factors and the likelihood of instability. We examined various proxies for instability. These proxies for instability include the subjective KOSIMO index, which captures a broad spectrum of type of conflict and includes nonviolent conflicts, and objective measures of the occurrences of specifically defined types of conflict, such as civil and local wars.

We find that there is a rich body of empirical work that will support subsequent efforts to model social responses to changes in economic and political environments. Although we did not significantly represent or extend the state of this art, we do find that the current conflict research faces common limitations due to a reliance on conventional statistical methods. Such limitations arise from lack of data, an inability to properly account for very complex relationships, and an assumption that history is representative of the future. We hold that agent-based and system-based modeling approaches will draw from the existing research and extend our ability to model and analyze prospective future scenarios stemming from climate change.

We derived some crude empirical findings and conducted some basic forecasting experiments to support our overarching goal to understand climate effects on instability. Empirically, the data suggest that stability can degrade in response to natural disasters that cause deaths and injuries, and in response to decline in average daily calories. For the case of caloric reductions, we conducted simple experiments in which we reduced the average daily calories and forecasted the likelihood of conflict in different countries. Although further refinements and testing are necessary, these exercises demonstrate our ability to incorporate social, political, and economic data into a framework to forecast the effects of certain climate-related changes. Ultimately, such exercises might help parameterize or verify subsequent agent-based models, and provide a baseline for identifying and measuring the greater realism afforded by agent-based models.

### 3.6 References

Chen, Y. Y. 1995. "Statistical Inference Based on the Possibility and Belief Measures." *Transactions of the American Mathematical Society* 347, no. 5 (May): 1855–63.

———. 2000. "Fuzzy Analysis of Statistical Evidence." *IEEE Transactions on Fuzzy Systems* 8, no. 6: 796–99.

Collier, P. 2000. "Economic Causes of Civil Conflict and Their Implications for Policy." Working Paper of the World Bank Research Program to Study the Economics of Civil War, Crime and Violence.

———. 2001. "Implications of Ethnic Diversity." *Economic Policy* 16 (April): 127–66.

COW (Correlates of War). 2004. "Correlates of War 2." Available at <http://cow2.la.psu.edu/> (accessed 17 November 2004).



- CRED (Centre for Research on the Epidemiology of Disasters). 2004. "EM-DAT: Country Profiles." Spreadsheet data are available at <http://www.em-dat.net/disasters/country.php> (accessed 17 November 2004).
- Heidelberg Institute for International Conflict Research. 1998–99. *KOSIMO Manual*. Available at [www.hiik.de/en/manual.htm](http://www.hiik.de/en/manual.htm) (accessed 22 October 2004).
- King, G., and L. Zeng. 2001. "Improving Forecasts of State Failure." *World Politics* 53 (July): 623–58.
- O'Brien, S. P. 2002. "Anticipating the Good, the Bad, and the Ugly." *Journal of Conflict Resolution* 46, no. 6 (December): 791–811.
- O'Brien, S. P. (Web Site). "Anticipating the Good, the Bad, and the Ugly: An Early Warning Approach to Conflict and Instability Analysis." Available at <http://www.geocities.com/seanob88/> (accessed 31 October 2004).
- Pfetsch, F. R., and C. Rohloff. 2000. "KOSIMO: A Databank on Political Conflict." *Journal of Political Conflict* 37, no. 3:379–90.

## **4 Using Recent Advances in Behavioral Analysis to Address Climatic Impact on Conflict**

### **4.1 Introduction**

This exploratory research reviews those methods and approaches that are potentially most appropriate for describing the impacts of climate change on societal stress and conflict. While the complete simulation of such impacts would presumably use sophisticated agent-based simulation, this exploratory effort examines advances in behavioral and statistical analysis for guidance in those future efforts.

#### **4.1.1 Reviewed Methods**

The recent Nobel Prize–winning work of Clive Granger, Robert Engel, and Daniel McFadden provides new and powerful approaches for understanding historical and statistical data. The Granger/Engel work focuses on cointegration. Cointegration separates the phenomena that remain functioning during dramatic change from those that break down and cause new (possibly counterproductive) responses. Further, cointegration is based on causality, in a temporal sense: the future and the present cannot affect the past. Granger causality can determine the direction and legitimacy of implied causal relationships. The causal logic correctly implies stocks, or levels, of physical and informational quantities that dominate the human decision-making process. Granger causality can invalidate assumed relationships among phenomena and data, but it cannot truly prove causality. Although only introduced within this work, Bayesian causal discovery (BCD) appears to have the ability to actually determine the true causal relationships and, thereby, (with cointegration techniques) increases the validity of simulating impacts to levels previously unimaginable. Lastly, qualitative choice theory (QCT) developed by Daniel McFadden has a long and successful history of simulating human decision making. QCT explicitly includes the uncertainty associated with decisions, responses, and the consequences of both. The theory notes the limitations of humans to comprehend information, conditions, and stimuli, and can readily include irrational behavior and internal preferences and beliefs in simulating human responses.

Specifically, this effort builds on the work described in Section 3 and focuses on conditions that potentially lead to conflict. QCT is a probabilistic method and, therefore, directly applicable to portraying the probability of conflict. Similarly, conflict tends to be associated with a building up of conditions that generate conflict. Thus, the concepts of cointegration and Granger causality would appear to be generally applicable. A literature review indicates that cointegration and its associated methods have only been peripherally utilized for conflict analysis (Shellman 2004a, 2004b). There is yet no reported use of QCT. QCT and its linkage to cointegration represent new and promising ground for the analysis of conflict causality and evolution.

Like the causal analysis described in Section 3, this work uses the O'Brien (2002) data set augmented by data from the CRED (Centre for Research on the Epidemiology of Disasters [Catholic University of Louvain, Belgium]). Although there are a large number of conflict-related data sets, we used only the O'Brien data set to evaluate approaches and methodologies because of the limited time frame for conducting these exploratory analyses.

The results of the exploratory analyses presented in this section do indicate the need for dynamic considerations where pressures build up, over time. The results also strongly suggest the casual relationship between conflict and natural disaster is quite dynamic and more complex than aggregate econometric methods can encapsulate. An agent-based system can recognize the delayed impact of *changes* in extreme weather that affect societies and their ability to cope with them, given the surrounding physical, economic, cultural, and geopolitical constraints.

#### **4.1.2 Section Overview**

Our discussion of the reviewed methods begins with an introduction to QCT and provides an illustrative example using the O'Brien data set. In general, QCT is the prime candidate to causally model all human decisions. Next, we introduce cointegration and Granger causality with an example of both concepts using the O'Brien data set. Cointegration shows that the build-up of historical conditions affects how current situations trigger conflict. Cointegration implies the existence of interacting paths that need to be addressed with a systems perspective. This type of knowledge is critical in establishing principles to limit the potential for conflict escalation. A brief diversion on BCD (Bayesian causal discovery) is included to indicate how a more quantified data set could be used to ensure agent-based simulation correctly reflects true causal relationships. Without a causal understanding of the system, it is impossible to determine the validity of any proposed interventions.

### **4.2 Qualitative Choice Theory**

QCT has a long history in psychology. This theory has only been fully developed for economic and behavioral use through the work of the 2000 Nobel Prize-winner Daniel McFadden (McFadden 1982). Independently of whether an individual is rational, irrational, profit maximizing, or satisficing, QCT applies to the decision-making process. In an all-encompassing economic sense, we all compare the "value" of one choice to others. QCT simply says that individuals make a choice based on their perception of utility regarding those choices. QCT causes any and all information (preferences, tastes, culture, costs, etc.) utilized by the individual to define a valid (or at least functional) representation of choice behavior. Like BCD (Bayesian causal discovery), QCT starts with the data reflecting the conditional probability of a choice given possibly interacting, conflicting, and limited information.

## 4.2.1 Probabilistic Decisions under Imperfect Conditions

At the individual level, QCT represents the probability that a particular decision will be made. It is thus directly applicable to a multiple-agent perspective (for example, the Los Alamos National Laboratory [LANL] Transim model). At a societal (or even tribal) level, the probability translates to the fraction of the population making a particular decision. Theoretically, any form of the probability distribution can act as the basis for the analysis. In practice, the Weibul distribution has the greatest numerical ease-of-use and has shown itself to be empirically the most likely shape of the actual distribution. The Weibul distribution is skewed to the left with a broad tail to the right. This implies that while individuals consider higher “cost” or lower “value” options, they tend to focus on the lower “cost” and higher “value” options.

People do not have perfect information. A sampling of the population shows different perceptions of actual costs and personal preferences. The choice made is called random utility maximization, or RUM (McFadden 1986). Figure 4-1 below shows an example distribution of perceived price for three technologies (choices). Preferences are not included to simplify the example.

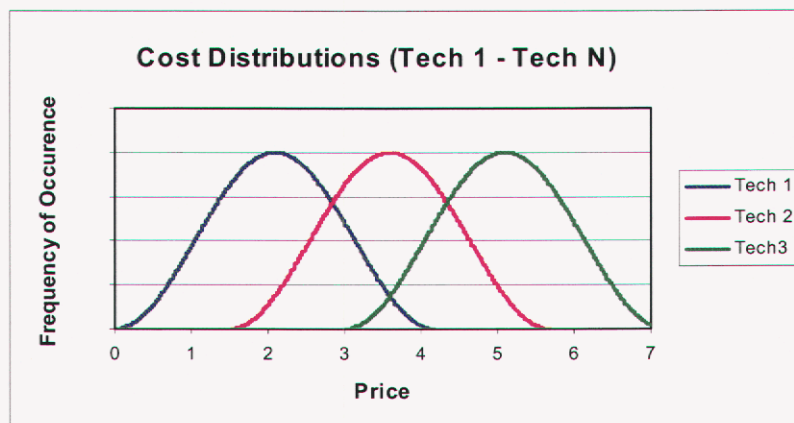


Figure 4-1. Illustrative choice distribution.

Maximum-likelihood estimation (MLE) methods determine the shape of the distribution as a function of costs and preferences in the model (McFadden 1986). The actual market share is determined by mathematical integration over the distributions (McFadden 1974). Nonetheless, the physical process can be understood intuitively. Most individuals will perceive technology 1 as less expensive and select it. However, several individuals will perceive technology 2 competitive with technology 1 and select technology 2. Finally, a small few will perceive technology three as the least expense and select it.

The market share of technology 1 would be as shown in Figure 4-2, as its price varies relative to the prices of the other choices. The price ratio depicts the weighted price of the other alternatives divided by the price of technology 1. As the price of technology 1 becomes small compared to the other choices, the market share of technology 1 would

go to unity. If the uncertainty is large (as in a residential decision), the slope is gradual. If there is significant effort to reduce costs (have less uncertainty), the curve is steeper, as shown for industrial choices. If there is perfect information, as assumed in an unconstrained linear programming (L-P) framework, then the market share would jump from 0.0 to 1.0 with the smallest of price differentials.

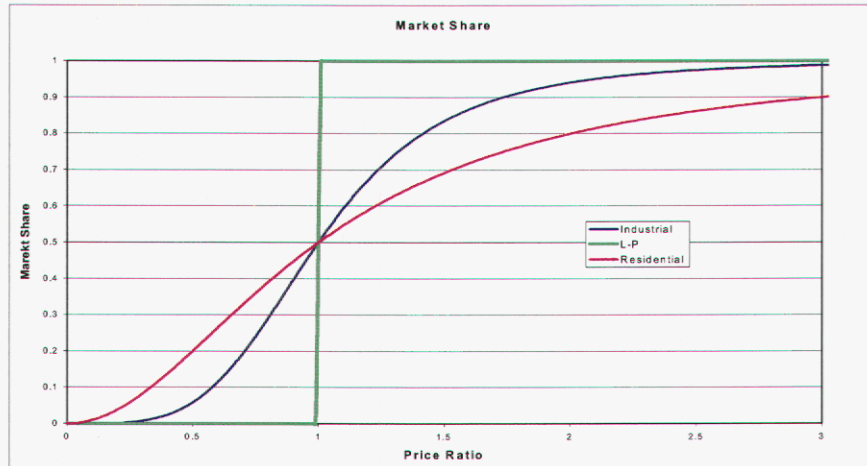


Figure 4-2. Illustrative market-share response.

## 4.2.2 QCT Mathematical Basics

The integration of Figure 4-1 produces the probability of the choice, or in the aggregate, the market share, of the  $i$ 'th choice ( $MS_i$ ) per Figure 4-2. For a Weibul distribution, this integral has a closed-form solution:

$$MS_i = \frac{e^{U_i}}{\sum_{j=1}^N e^{U_j}}, \quad (4.1)$$

where  $U_i$  is the utility of choice  $i$ , and  $e$  is the base of the natural logarithm. The utility function is often written, for example, as a simple linear function of price ( $P_i$ ) with the constant (nonprice) term noted by Train (1986):

$$U_i = A_i + B * P_i. \quad (4.2)$$

In this case,  $A$  would be (assumed constant) nonprice factors of taste and preference for the  $i$ 'th choice.  $A$  can also capture the ability to make the choice (e.g., the limitation of physician selection in an insurance plan) or the availability of the choice (e.g., the availability of corn in the Sahel). Note that  $B$  does not have a subscript. It is directly related to the uncertainty of the choice—how well the information of the choice set is known and understood. The uncertainty of the decision process is the same for all choices

in a set because it is an ordinal, and not a cardinal, process that compares all options at once.

There can be a hierarchy of choice, like a binary tree, but the logic is called nesting because the decision process is represented as a nested hierarchy of decisions. Each level is a choice among all the options of that level (e.g., choosing the flavor of ice cream to eat occurs after choosing which place to go for the snack after the decision to go for a snack.) Each decision level is self-contained but can be conditional on the level below it.

The derivation of the theory of QCT requires that all choices at any level are mutually exclusive (e.g., the decision to live in Kashmir or migrate to India). Empirically this limitation is nonbinding. A classic example is the addition of travel choice by painting half of all the buses green and the remaining buses blue. There really has been no change in the choices—taking the green bus is no different from taking the blue bus. The  $A$  of Equation (4.2) can capture this fallacy by simply multiplying the blue-bus and green-bus choices, in this example, by 0.5. The same process can often allow the complicated nested equations to be reduced to a single layer called a “comb” that requires only the single use (and estimation) of Equation (4.1).

Reducing the uncertainty, increasing the understanding of the choices, and making better decisions (as contained in the  $B$  term of Equation [4.2]), takes time and effort. The benefit may not be worth the effort. When buying a house, a purchaser may want to know the price within 1% or less. For a candy bar, a 200% variance in uncertainty is tolerable. The consequences of purchasing a house are much more momentous than those of purchasing a candy bar. The magnitude of  $B$  appears to vary directly with the importance of the decision. That importance is the cost of the decision compared to the value of the entire output (a labor-year of income for a person and the revenue for a company).

Data indicate the linear function of Equation (4.2) works well for small variations of the input variables, but the actual underlying function is logarithmic. Equation (4.3) is a simple logarithmic enhancement of Equation (4.2):

$$U_i = A_i + B \cdot \ln(P_i). \quad (4.3)$$

The use of the logarithm indicates that people can determine relative proportionality but not absolute differences in price (or other components of utility). This implication is consistent with the previous discussion that  $B$  is proportional to the percentage impact it has on total outcome.

If Equation (4.3) is substituted into Equation (4.1) and  $m$  is defined as

$$m_i = \exp(A_i), \quad (4.4)$$

then Equation (4.2) becomes

$$MS_{(i)} = \frac{m_i P_i^B}{\sum_{j=1}^N m_j P_j^B}. \quad (4.5)$$

Equation (4.5) is consistent with the engineering assessment of options according to the distribution of (estimated) cost versus (estimated) performance. The uncertainty of the estimate ( $B$ ) is also a function of the importance of accuracy. This is the only example known where engineering/scientific theory and economic theory agree.

While MLE is required for the unbiased estimation of Equation (4.5), within a feedback system, ordinary least-square estimation often produces adequate parameterization to generate accurate forecasts.

Note that because the decision process is always ordinal, there is no absolute concept of preference. Therefore, one of the  $m_i$  must be arbitrarily selected as the numeraire and set to unity.

The use of QCT seems to force a rigor and a method for defining the implicit or explicit decisions associated with a simulation hypothesis. Experience indicates that QCT forces a self-consistency of thought and theory that always has a causal description consistent with empirical data.

### 4.2.3 Testing QCT for Climate-Conflict Impacts

Building on the work described in Section 3, QCT can simulate the conflict as the decision (choice) to actively confront an apparent threat or opportunity based on the perceived conditions society or individuals may be using for that choice. The set of information available by year and country from O'Brien and CRED data includes the following:

- implied level of conflict from KOSIMO
- prior conflict
- political rights metric
- civil liberty metric
- democracy metric
- open society metric
- life expectancy
- youth bulge
- ethnic majority dominance
- religious majority dominance
- calories per capita
- number killed in disasters
- number injured in disasters
- number made homeless in disasters
- number otherwise affected by disasters

Realistically, only a fraction of this information contributes to the actual decision to engage in conflicts. Some of the variables listed above are surrogates for the actual perceived conditions. For example, both the ethnic and religious majority may reflect a tension between a majority wanting to protect or enforce its favored institutions and a minority that threatens or feels threatened by the majority. Both BCD and cointegration provide methods to determine the “correct” causal hypotheses.

Until statistical analysis shows otherwise, the utility of conflict ( $U_c$ ) is assumed to be a function of all the information ( $X_i$ ) noted above:

$$U_c = A + \sum B_i \ln(X_i) , \quad (4.6)$$

where  $i$  designates which piece of information is used. Both  $A$  and  $B_i$  are estimated. If  $A$  came out to be 0.0, it would imply that the  $X_i$  reflect all the relevant causal drivers of the conflict.

For this illustrative analysis, it is assumed that individuals only use current information to make decisions. The later Granger causality discussion will consider the importance of historical (remembered) information for formulating current decisions. Again, the logarithmic form is used to reflect the affine nature of decision. Humans can only recognize proportional (relative) changes. There is no absolute measure of perceptions or conditions.

The utility of no conflict ( $U_n$ ) is considered constant and arbitrarily assigned a unity value. Because the probability of a choice is based on the ordinal rather than the (nonexistent) cardinal comparison of the utilities, the choice of the  $U_n$  has no impact on results.

Given that this analysis assumes a simplified binary choice (no conflict or conflict), the probability of conflict ( $P_c$ ) is then

$$P_c = \text{Exp}(U_c) / (\text{exp}(U_n) + \text{exp}(U_c)). \quad (4.7)$$

For this analysis, the KOSIMO index of conflict is normalized as a continuum covering a range of no conflict (0.0) to complete conflict (1.0). For illustrative purposes, the maximum KOSIMO value (4) is assumed to reflect a position halfway to “total conflict” and given a value of 0.5. All other values are scaled accordingly to give a conflict ratio. At a societal level, the interpretation is that if the conflict ratio is 0.5, then 50% of the population is involved. At an individual level, this means there is a 50% chance the individual has chosen to enter into the conflict. Thus, the conflict ratio equals  $P_c$  at the individual level.



**Table 4-1. Variables Affecting the Probability of Conflict**

Variable	Coefficient	Std. Error
disaster affected	0.0032	0.0016
disaster homeless	0.0006	0.0021
disaster injured	0.0027	0.0018
disaster deaths	-0.0037	0.0014
calories per capita	-0.1097	0.0821
religious majority	0.0474	0.0424
ethnic majority	-0.0691	0.0264
youth bulge	0.1010	0.0461
life expectancy	0.1976	0.0799
open society	-0.0432	0.0171
democracy	0.0036	0.0021
liberty	0.0753	0.0384
political rights	0.0135	0.0360
prior conflict	0.0895	0.0015
constant	-1.0833	0.3220
R <sup>2</sup>	0.740544	
Durbin-Watson (DW)	316.4106	

Note that the standard error on the coefficients is high. This is not a good regression, but the purpose here is more to evaluate methods than to provide usable quantifications. A positive sign for a coefficient indicates that a large value for the independent variable leads to a higher probability of conflict. Many of the qualitative results make good causal sense.

A better standard of living (calories per capita) reduces the probability of conflict. A large number of restless, fighting-age youths provide ammunition to pursue conflicts. That prior conflict has a relatively large coefficient indicates that past information may be as relevant as current information. Survivors of disasters may need to take actions that lead to conflicts, such as the recent riots from hurricanes in Haiti. Deaths actually relieve the pressure on resources and could actually have the negative affect shown. The relatively large value of the constant term indicates that the use of only current-valued variables is too limiting, that there are many other important variables that have been omitted, or both.

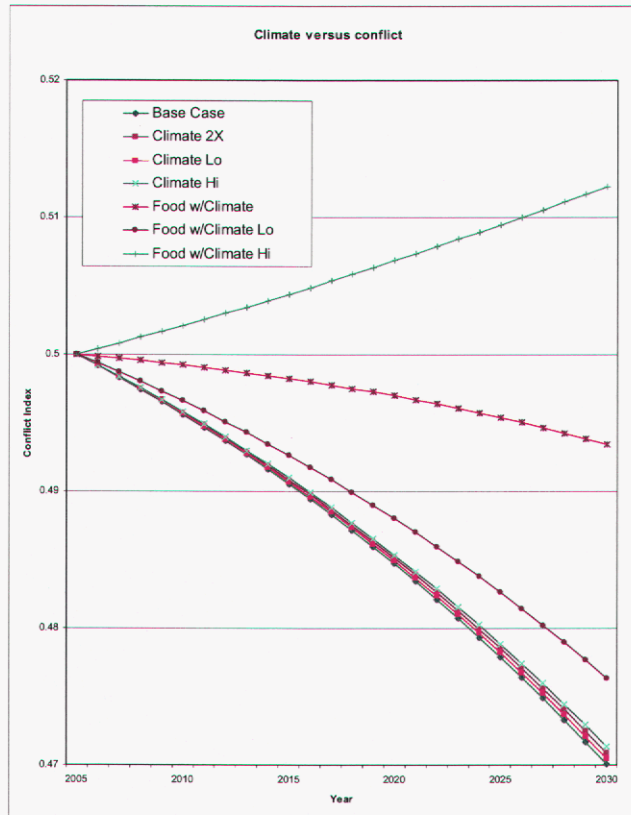
To avoid confusion, we should emphasize the fact that civil liberties, political rights, and democracy indexes are defined in terms of rank. The civil liberties and political rights indexes are listed in *decreasing* rank order, ranging from “1” to “7”. That is, “1” denotes the most civil liberties, and “7” denotes the least civil liberties. Similarly, “1” denotes the most political rights, and “7” denotes the least political rights. The democracy index, on the other hand, is in *increasing* rank order ranging from -10 to 10, where “-10” denotes the least democratic and “10” denotes the most democratic. In this analysis, greater democracy is associated with a lower likelihood of conflict, as we would

expect, and has been proposed through the state-strength literature. However, higher civil liberties indexes (less freedom) correlate with a lower likelihood of conflict. The fact that civil liberties and political rights indexes seem to be positively correlated with conflict could imply that the conflicts may need to be disaggregated to include purpose. North Atlantic Treaty Organization (NATO) countries, allies of the United States, and the United States alone regularly engaged in "policing" conflicts. Only more advanced nations (presumably associated with greater civil liberty and political rights) would pursue policing functions.

The regression indicates that the fewer the civil liberties, the less the probability of conflict. This result could be an artificial consequence of not separating the impact of civil liberties and other terms into regimes. A discussion in Section 3 shows a reversal of the impact of civil liberties when the analysis separates out regimes of ethnic and religious concentration. (See Table 3-9.) In a totalitarian society, strict control of the population limits the possibility of internal conflict. As civil liberties first become restricted, however, one would expect higher possibilities of resistance in the form of conflict. The focus of this effort is to determine the ability to recognize and delineate these issues, rather than the rigorous specification of the phenomena.

The opposing signs on both ethnic and religious majorities might imply that they represent complex concepts with many interacting layers. At best, the results imply that neither variable can be used as an aggregate surrogate for more-specific societal conditions that are associated with ethnic or religious affiliations. Note that the analysis of Section 3 also encountered this situation. (See Table 3-10.)

The value of a QCT dynamic framework is illustrated in Figure 4-3, which shows the potential impact of climate change on conflict. An economy is initialized with a 50/50 chance of conflict. With economic growth (assuming the calorie-per-capita coefficient as a surrogate for the impact of economic improvement), the probability of conflict declines as the economy grows (here assumed to be 3% per year). If climate change causes the incidence of destructive storms to double by 2030, there is only a slight increase in the probability of conflict. Three lines are shown to include the impact with the mean value of the estimated impact and to bound it within the standard error of parameter estimates (hi/lo). The "model assumptions," however, only catch the temporary impact of disgruntled survivors and do not capture any economic impacts that could lead to large shortages, unemployment, and social unrest. If the climate change is also assumed to impact food production (in a dominantly agrarian society), the impact is much larger.



**Figure 4-3.** Illustrative climate impacts on conflict.

As shown in the top of Figure 4-3, in the mean-value case, the climate almost negates the impact of increased economic growth. In the worst case, the climate almost reverses the benefits of economic growth. With better data sets and a more comprehensive use of QCT, it is possible to accurately quantify the impact of climate events and mitigation efforts on conflict incidence and intensity.

In summary, even at this cursory level, QCT shows itself as a valuable tool to understand conflict behavior.

### 4.3 Cointegration and Granger Causality

Cointegration was first conceived by Clive Granger (1981), but the development of the method was not achieved until 1987 (Engle 1987). Cointegration is now a widely accepted and used technique (Hamilton 1994; Engel 1991; Maddala 1992; Hendry 1993,1995). It focuses on determining the dynamics of no memory, short-term memory, and long-term memory within a data-generating process (DGP). The concept of “no memory” corresponds to a set of algebraic equations. Short- and long-term memories are naturally associated with state variables (integrated levels). Short-term memory conveniently corresponds to the reinforcing feedback dynamics and long-term memory to negative feedback dynamics.

### 4.3.1 Cointegration Mathematical Basics

Cointegration discusses dynamics in terms of variables being jointly integrated (or differenced). To statisticians, the effort is to find stationarity in the residual error term. That is, they want the variables to stay related without the error term growing over time.

#### 4.3.1.1 Stationarity

Differencing of a series with serial correlation will always result in stationarity if differenced enough times. An undifferenced equation is designated  $I(0)$ , a first difference  $I(1)$ , etc.

A general introductory cointegrated equation would be

$$\Delta Y_t = B0 + \sum B_i * \Delta X_{i,t} + Bn * (Y_{t-1} - F(X_{i,t-1})) + u_t, \quad (4.8)$$

where  $\Delta$  is the difference operator:  $\Delta X = (X_t - X_{t-1})$ , and  $u_t$  is the error term. A problem with econometrics is that the time interval always coincides with the data collection interval. In the “delta  $t$ ” limit, the difference equations are differential equations. The distinction between a “true” difference (discrete) equation and the “true” differential/integral equation will result in an anomalous  $\Delta$  term that is “without physical interpretation” in the cointegrated equation. The  $F(X_{i,t-1})$  is the asymptotic value of  $Y$  when  $X$  is held constant. A nonzero  $B0$  distorts this definition and is often restricted to a 0.0 value. The  $Bn * (Y_{t-1} - F(X_{i,t-1}))$  term is called the error-correction mechanism (ECM). An example of  $F(X_{i,t-1})$  could be

$$F(X_{i,t-1}) = A0 + A1 * X1 + A2 * X2 \dots \quad (4.9)$$

The use of a nonzero  $B0$  means that the  $A0$  in Equation (4.9) becomes  $A0 = A0 + B0 / Bn$ . As noted above,  $u_t$  is the classical “error term.” Cointegration ensures that this term is never serially correlated.

#### 4.3.1.2 Unit Roots

The determination of cointegration is based on the concept of a unit root. Equation (4.10) is a simple autoregressive equation that highlights the cointegration logic.

$$Y_t = \rho Y_{t-1} + \varepsilon_t. \quad (4.10)$$

In Equation (4.10),  $\varepsilon$  is the error term but, unlike  $u$  in Equation (4.8), it might be serially correlated. If  $\rho$  is greater than unity, there is a positive feedback situation. If it is less than unity, there is a negative feedback situation. This makes sense by rewriting the equation to look a bit more like systems dynamics:

$$Y_t = Y_{t-1} + dt * (\rho - 1) * Y_{t-1}, \quad (4.11)$$

$$\Delta Y = dY = (\rho - 1) * Y = \alpha Y . \quad (4.12)$$

The sign of  $\alpha$  determines the feedback-loop polarity. The polarity depends on the value of  $\rho$  compared to unity. If  $\alpha$  is thought of as a positive growth rate, as in population growth, then  $\rho$  is greater than 1.0. The equation is not cointegrated. It only has short-term memory. The level changes slowly as other possible inputs affect it. A simple system-dynamics delay, as shown in Equation (4.13), with its long-term, cointegrating memory that parrots human-memory dynamics will clarify the unit-root significance (Sterman 2000)

$$Y_t = Y_{t-1} + dt * (S_t - Y_{t-1}) / T \quad (4.13)$$

or

$$\Delta Y_t = (S_t - Y_{t-1}) / T , \quad (4.14)$$

where  $S$  is the input variable to be smoothed and  $T$  is the averaging time.

In general cointegration terms,

$$\Delta Y_t = B0 + B1 * \Delta S_t + B2 * (S_{t-1} - Y_{t-1}) . \quad (4.15)$$

Equation (4.15) looks like the original cointegration equation (Equation [4.8]) above. By comparing Equations (4.11) and (4.13) to Equation (4.15),  $B1$  and  $B2$  have the definitions below.

$$B2 = 1/T , \quad (4.16)$$

$$B1 = (\rho - 1) . \quad (4.17)$$

The  $(\rho - 1)$  term is comparable to its use in Equation (4.12) above.

Note the minor issue of the time-subscript change on  $S$  between the difference equation (4.15) and the implied differential equation (4.11). This is not statistically significant, but it does change the causal interpretation of estimation results.

The regression of Equation (4.15) corresponds to Equation (4.8) or (4.10) only if  $\rho$  is unity—the unit root. The unit root indicates the reinforcing loop limit. The value of  $\rho$  just needs to be unity from a statistical perspective. A value of less than unity will do as well in most cases. There is a problem if  $\rho$  is much above unity. The test that  $\rho$  is statistically unity is then not the conventional  $t$  statistic for  $(\rho - 1)$  being nonzero, but rather a modified distribution that is heavily skewed toward values below zero. The verification of the unit root is called the augmented Dickey-Fuller (ADF) test, in cointegration jargon. A delay (memory) function is perfectly cointegrated.

A population-growth equation is not cointegrated. The serial correlation of the error term can be removed by simply assuming a growth rate. A growth-rate equation has the  $B_n$  of Equation (4.8) equal to 0.0 and the  $B_i$  not all equal to 0.0.

### 4.3.2 Granger Causality Mathematical Basics

The explicit use of lagged values determines “causality.” In cointegration, the test (Granger causality) is not to prove causality, but to verify when there is not causality. If  $Y_t$  is a well-correlated function of  $X_{i,t-1}$ , the  $X_i$  could be causing  $Y$ ; but if  $Y_t$  is more correlated with a function of  $X_{i,t+1}$  (note the “+”), then the  $X_t$  does not Granger-cause  $Y$ . Another perspective on Granger causality is to say that  $Y$  is explained better by order- $n$   $[l(n)]$  lags of  $X$  than by lags of  $Y$  alone.

The test of whether  $Y_t$  is a function of  $X_{i,t}$  occurs in the first pass of the two-stage cointegration regression process. The first stage estimates the long-term (asymptotic) solution, and the second stage estimates the dynamic  $\Delta X$  contribution. Note that higher-order  $\Delta Y [I(n)]$  components can also be added to Equation (4.8).

Granger causality seeks to falsify the  $X$ -causality by testing if all the  $\varepsilon_i$  are 0.0. This process requires comparison to the autocorrelation Equation (4.19) with the inclusive Equation (4.18):

$$Y_t = c + \sum_{i=1}^p \alpha_i * Y_{t-i} + \sum_{i=1}^p \beta_i * X_{t-i} + \mu_t, \quad (4.18)$$

$$Y_t = c + \sum_{i=1}^p \gamma_i * Y_{t-i} + \varepsilon_t. \quad (4.19)$$

Let  $R1$  be the sum of squared residuals for Equation (4.18), and let  $R2$  be it for Equation (4.19):

$$R1 = \sum_1^N \mu_i^2, \quad (4.20)$$

$$R2 = \sum_1^N \varepsilon_i^2. \quad (4.21)$$

If  $N$  is the number of observations, the test is then

$$N * (R2 / R1 - 1) = \chi^2(p). \quad (4.22)$$

Cointegration regularly verifies that assumptions about simultaneous relationships/interactions, such as the price being a function of current supply and demand (Hendry 2001, 2001), or that current weather drives current conflict, are not valid. Thus cointegration supports the agent-based–simulation view that interactions are caused by previous conditions or by long-term assets/perceptions associated with

previous conditions. The historical relevance, then, further implies that feedback must dominate the process. There must be state variables. Therefore, cointegration and agent-based modeling are integrally tied together. The implied existence of cointegration theory implies the existence of “memory.” If the process is balancing (negative feedback), the process must be cointegrated.

#### 4.3.2.1 Testing Cointegration for Climate-Conflict Impacts

Both cointegration and Granger causality were applied to the O’Brien data set to determine any relationship between disasters and conflict. The first test used a lag across five time periods. As shown in Table 4-2, the test produces possibly acceptable statistical results, but from a cointegration perspective it fails. The  $\chi^2$  (chi-square) indicates there is approximately a 70% chance that disasters have no impact on conflict, that is, a 70% chance there is no causality.

**Table 4-2. Five-Lag Granger Causality Test**

Variable	Coefficient
L1KOSIMO	2.45E-05
L2KOSIMO	1.56E-05
L3KOSIMO	-3.1E-05
L4KOSIMO	2.65E-05
L5KOSIMO	3.11E-05
L1Disaster	-0.04729
L2Disaster	0.150822
L3Disaster	-0.05167
L4Disaster	0.02955
L5Disaster	0.848405
Constant	0.125666
R <sup>2</sup>	0.861049
DW	505.656
$\chi^2$	0.686

Granger causality is very sensitive to the choice of lag length. A hypothesis could state that farmers can handle a year or so of bad weather, but then the accumulation of financial damage or the dwindling of food reserves sets up the impact of the final “tipping point” disaster. Table 4-3 used a single five-year lag. It implies the hypothesis is valid and indicates that historical disasters have only a 0.3% chance of *not* affecting conflict.

**Table 4-3. Single-Lag Granger Test**

Variable	Coefficient
L5KOSIMO	0.00019
L5Disaster	0.701222
Constant	0.541153
R <sup>2</sup>	0.530242
DW	465.047
$\chi^2$	0.0033

Lastly, the cursory cointegration test also indicates that conflict does indeed reflect a unit-root process ( $\rho = 1$ ). Therefore, the mechanism describing conflict should be robust over a wide range of future and unforeseen conditions. A simulation including those discovered mechanisms, thereby, supports a simulation platform whose results have a high degree of confidence.

Cointegration produces incredibly accurate forecasts in timing and magnitude. The equations are often found by using all the information available and include higher-order differencing (all the combinations). This is called a-theoretical modeling. Some parameters can be construed to have an economic interpretation. Those that have none are still deemed legitimate by analysts because they add to the accuracy of the forecast and intervention impacts. It appears that cointegration can help with system definition and its verification. Cointegration does help determine the implied equilibrium goal of societal systems. It can determine whether positive or negative feedback mechanisms are missing. Cointegration tends to show that historical data *do not* have structural breaks. Humans make decisions as they always have, despite wars and stock-market bubbles. Cointegration indicates that the model equations should produce the “break” response from the changing input data (or endogenous model dynamics) and not from the changing model structure. These qualities would appear very useful for addressing the impact of climate change on conflict or for ensuring that agent-based models produce valid results.

#### 4.4 Bayesian Causal Discovery

BCD (Bayesian causal discovery) is a final complementary method to both QCT and cointegration. BCD is relatively new and remains controversial (Kim 1997). The development of the technique depended in part on the greater computational capabilities available today for solving nondeterministic polynomial (NP)-hard problems via exhaustive search. The original (Spirtes, Glymour and Scheines [SGS] algorithm) approach was developed by Spirtes, Glymour and Scheines (1993). It goes beyond Granger casualty to actually claim knowledge of the causal relationship within a specified degree of confidence (Pearl 1998; Salmon 2001).

In general, BCD assumes only nonexperimental data—the only data generally available in socioeconomic modeling. It combines several concepts and assumptions that



may, at first, sound overly restrictive, but they are actual broadly applicable to most real-world situations.

#### 4.4.1 BCD Mathematical Basics

The basic concept of BCD is based on the assumptions that actual causal relationships exist and that the data available are usable to support a position on the confidence in a causal assertion. All potential causal relationships (even those that will later prove to be independent) are collected. These are assembled into all possible directed acyclic graphs (DAGs). “Acyclic” indicates that no feedback loops are modeled. This does not mean that there are no feedback loops. It means that a feedback loop contains a state variable (an integrated level). The time-subscript change occurring at the level allows an unfolding of the loop for statistical analysis purposes, such that the  $t - 1$  is at one end of the DAG and the  $t$  term is at the other end of the DAG. (Note again that this logic validates the approach associated with agent-based simulation).

If we hypothesize that  $X$  causes  $Y$  and that  $Y$  causes  $Z$ , the DAG would simply be as shown in Figure 4-4.

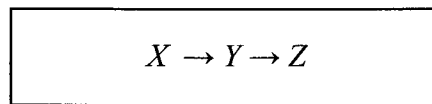


Figure 4-4. Simple DAG.

Just like a flow diagram, the arrow direction indicates (assumed) causality. The causality can be designated *a priori* with a positive or negative as a restriction to the statistical analysis of the data that determines the correlations and rejects many DAGs.

There are generally many variables and thus many potential combinations with multiple connections, such as shown in Figure 4-5.

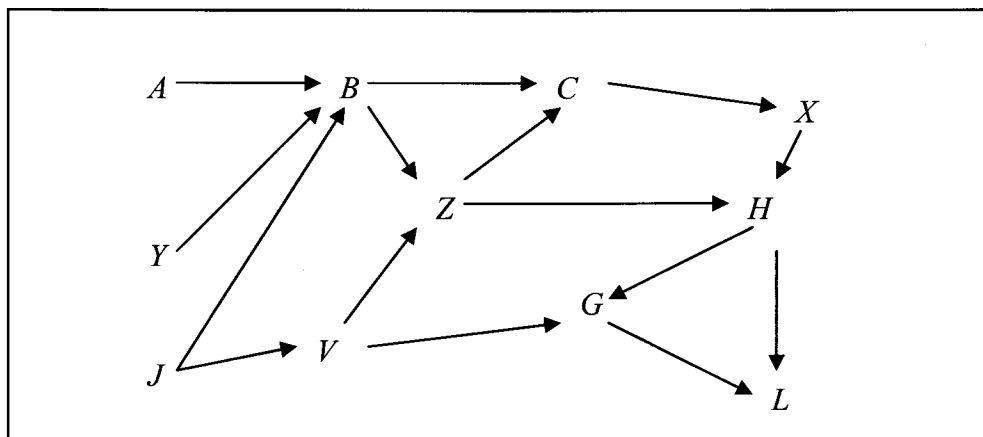


Figure 4-5. Multipath DAG.

The SGS algorithm would look at all possible combinations of connections of the variables in Figure 4-5 as a possible cause of any other variables. The SGS algorithm looks for “d-separation” points in the DAGs it generates. In the case of  $L$  in Figure 4.5, conditional on  $V$ ,  $Z$ , and  $X$  above, only  $G$  and  $H$  affect  $L$ . In Figure 4-4,  $Y$  d-separates  $X$  and  $Z$ . Given  $X$ ,  $Z$  only depends causally on  $Y$ .

The d-separated sets in any DAG are then tested to determine the correlation (or lack thereof) among the variables in a given direction of causality. If two variables are not d-separated, they are conditionally dependent. All the conditionally dependent variables (d-connected) affecting another variable in a DAG constitute the potential casual relationship with that variable.

In any BCD analysis, the data must be “faithful” to any causal assertion. The “causal faithfulness” is a key concept in that it not only implies that the representation is valid, but that there is acceptable confidence (faith) that the relationship is truly causal. All the relationships are represented as Bayesian networks where the data are used to determine the conditional probability of one variable, given its d-connected variables and conditional on the values affecting the d-separated variables (Howson 1993). The independence of a variable is provided by the causal Markov condition. To have a causal Markov condition, a variable  $X$  must be independent of (not causally related to) every other variable, conditional on all of its direct causes. That is, the variable  $X$  must be causally faithful. Causal Markov conditions are the basis for the assumed causality found via the statistical search (Suppes 1990).

In any BCD analysis, those DAGs that cannot be causal due to time-lag considerations ( $Y_{t+1}$  can never cause  $Z_t$ ), are thrown out. Any *a priori* considerations that exclude some combinations are also used to reduce the possibilities that must be tested.

Equation (4.23) illustrates the more general logic. Other prior considerations ( $K$ ) are added as appropriate (or justified) and combined with the data ( $O$ ) to produce the posterior probability that, for example,  $Y$  causes  $Z$  given observations  $O$  and prior knowledge  $K$ . Let  $S$  be the set of all possible combinations of DAG available related to  $Y$  causes  $Z$ . The probability that  $Y$  causes  $Z$  conditional on  $O$  and  $K$  is then

$$P(Y \rightarrow Z | O, K) = \sum_{S: \{Y \rightarrow Z\} \text{ element of } S} P(S, O | K). \quad (4.23)$$

Glymour provides the details of the process (Glymour 1999). Additional statistics quantify the probability that the probability of causality is meaningful. These probabilities can be ranked. Usually, only one DAG stands out as most probable. Remember that, by definition, all causal possibilities are present; so the result is definitionally determining “THE Causality.” It is possible that there will be a small group of causal inferences with statistically comparable probabilities. Frequently, this means that all are part of a more elaborate single causal process that includes latent (unmeasured) variables not known or not available in the data. To an econometrician, this phenomena is called *encompassing*. It is implied that the latent variable has interactions

with all the “competing causality” DAGs. The inclusion of a meaningful latent variable can also be used to tie the set together for statistical parameterization of the equations.

Sometimes the data are simply not available to determine the causality. This is truer for aggregate country data than, for example, industry data. A latent variable that represents a critical level tightly tied in feedback violates all the premises required for causal discovery. Luckily, such a hidden level is uncommon. The technique only avoids controversy when it is used for verification of assumed causality or to sort through competing causal hypotheses during model construction.

Causal discovery techniques can add credibility to contentious causal assumptions. These techniques can sort through contradictory casual assumptions. They can find (potential) causal relationships when the set of relationships used in the model seems inconsistent with historical data. Conversely, these techniques can show where there is variable independence despite claims by “reviewers” of cooperative or coordinated activities. The use of causal discovery is then to “prove” that there is not causality. In this situation, only the DAG assumed by the reviewer is needed to show the lack of correlation, i.e., the lack of a causal Markov condition.

## 4.5 Summary

This work suggests that cointegration, Granger causality, QCT (qualitative choice theory), and BCD (Bayesian causal discovery) can be useful to the analysis of climate change. Further, even cursory results will help guide the direction of future work. For example, the buildup of societal pressures from previous disasters is key to understanding how disasters affect conflict. The apparent unit-root nature of the process indicates the simulation will be robust and valid. Even with extremely limited data, QCT provides insights and gives confidence in its use to simulate human behavior.

## 4.6 References

- Engle, R. F., and C. W. J. Granger. 1987. “Co-integration and Error Correction Representation, Estimation, and Testing.” *Econometric* 55:251–76.
- . 1991. *Long-Run Economic Relationships: Readings in Cointegration*. Oxford, UK: Oxford University Press.
- Glymour, C., and G. Cooper, eds. 1999. *Computation, Causation, and Discovery*. Cambridge, MA: MIT Press.
- Granger, C. W. J. 1981. “Some Properties of Time Series Data and Their Use in Econometric Model Specification.” *Journal of Econometrics* 16:121–30.
- Hamilton, J. D. 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press.
- Hendry, D. F. 1993. *Econometrics: Alchemy or Science*. Cambridge, UK: Blackwell Publishers.

- . 1995. *Dynamic Econometrics*. Oxford, UK: Oxford University Press.
- . 2000. “Explaining Cointegration Analysis: Part 1.” *The Energy Journal* 21, no 1:1–42 [International Association of Energy Economists, Cleveland, OH].
- . 2001. “Explaining Cointegration Analysis: Part 2.” *The Energy Journal* 22, no. 1: 75–120 [International Association of Energy Economists, Cleveland, OH].
- Howson, C., and P. Urbach. 1993. *Scientific Reasoning: a Bayesian Approach*. Chicago: Open Court.
- Maddala, G. S. 1992. *Introduction to Econometrics*. New York: Macmillan Publishing Company.
- McFadden, D. 1974. “Conditional Logit Analysis of Qualitative Choice Behavior.” In *Frontiers in Econometrics*, edited by P. Zarembka. New York: Academic Press.
- . 1982. “Qualitative Response Models.” In *Advances in Econometrics*, edited by Werner Hildenbrand. New York: Cambridge University Press.
- . 1986. “Econometric Model of Probabilistic Choice.” In *Structural Analysis of Discrete Data with Econometric Applications*, edited by C. F. Manski and D. McFadden. Cambridge, MA: MIT Press.
- McKim, V. R., and S. Turned, eds. 1997. *Causality in Crisis?: Statistical Methods and the Search for Casual Knowledge in the Social Sciences*. Notre Dame, IN: University of Notre Dame Press.
- O’Brien, S. P. 2002. “Anticipating the Good, the Bad, and the Ugly.” *Journal of Conflict Resolution* 46, no. 6 (December): 791–811.
- Pearl, J. 1998. *Causality: Model, Reasoning, and Inference*. Cambridge, UK: Cambridge University Press.
- Salmon, W. 2001. *Causality and Explanation*. Oxford, UK: Oxford University Press.
- Shellman, S. M. 2004a. “Measuring the Intensity of Intra-national Political Interactions Events Data: Two Interval-Level Scales.” *International Interactions* 30, no. 2 (April–June 2004).
- . 2004b. “Time Series Intervals and Statistical Inference: The Affects of Temporal Aggregation on Event Data Analysis.” *Political Analysis* 12: 97–104.
- Spirtes, P., C. Glymour, and R. Scheines. 1993. *Causation, Prediction, and Search*. New York: Springer Verlag.
- Sterman, J. D. 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Irwin, NY: McGraw-Hill.

Suppes, P. 1970. "A Probabilistic Theory of Causality:" In *Acto Philisophica Fennica* XXIV Amsterdam: North Holland Publishing Co.

Train, K. 1986. *Qualitative Choice Analysis*, Cambridge, MA: MIT Press.

## 5 Agent and Climate Model Coupling

In this section, we discuss how to couple an agent-based social model with a physics-based climate model. The discussion begins with an overview of the climate model. Then, we define the steps necessary to fully couple the two models.

### 5.1 The Community Climate System Model

For the climate model, we focus on the Community Climate System Model (CCSM), the United States' largest research effort (CCSM 2004). The CCSM is a fully-coupled, global climate model that provides state-of-the-art computer simulations of the Earth's past, present, and future climate states. It is a large (\$9 million per year) interagency collaboration that is funded by the National Science Foundation, the Department of Energy, the National Aeronautics and Space Administration, and the National Oceanic and Atmospheric Administration. The CCSM is headquartered at the National Center for Atmospheric Research, but major components are developed at other organizations such as Oak Ridge and Los Alamos National Laboratories. The CCSM has over 100 active participants with expertise in a wide range of disciplines.

The CCSM is used for studying the Earth's climate and the climate's sensitivities to a range of natural and man-made changes to our environment. The CCSM has four main components: an atmospheric model, an ocean model, a land surface model and a sea ice model. The individual models do not talk directly to each other, but instead communicate via a separate software component known as the flux coupler. The CCSM is freely available to any interested researcher. The model components are all designed to run reasonably well on both cache-based and vector processors and make use of industry-standard parallel-programming environments to run on distributed-memory systems. The CCSM is quite portable and runs on a wide variety of supercomputer platforms. It includes built-in test facilities suitable for validating installation and verifying some types of model changes.

To document and validate the CCSM, various multicentury control runs are carried out. All output data from these control runs are available to the public. The Earth System Grid (ESG) is the primary method for distributing these output data (ESG 2004). Another series of CCSM experiments is under way that will provide data for an upcoming Intergovernmental Panel on Climate Change (IPCC) Assessment Report. The IPCC study will result in 100-year data sets for a variety of future scenarios, mostly related to different levels of increase in atmospheric carbon dioxide (CO<sub>2</sub>) concentrations. Data from these experiments will be made available in late 2004. The data take the form of monthly means for a vast array of quantities predicted and diagnosed by the model. Examples include monthly mean temperatures and precipitation, vegetation growth, and soil conditions. The data are stored using the NetCDF format, and are easily accessible via software libraries for FORTRAN, C, and C++.

## 5.2 Development of a Coupled Model

The ultimate goal in an agent-based coupled CCSM model would be full two-way coupling, where the agents can respond to the wide range of climate data produced by the CCSM and the CCSM can respond to the effects predicted by the agent model. For example, consider desertification brought on by a long-lasting drought. The initial desertification may trigger political upheaval resulting in major changes to farming and other land use. These changes could greatly accelerate desertification as compared to what would have occurred naturally, and this accelerated desertification cannot be predicted by a model that did not take into account human reactions to climate change.

There is a logical progression of steps towards the full two-way coupling of agent-based social models with the CCSM. We first assume that a robust agent-based social model has been developed that can be used to study the sensitivity of social, political, and economic situations to climate change. Such a model would rely on climate data for some of its input. For example, a model of Darfur would require knowledge of the rate of desertification predicted by the CCSM's land surface model. The first step towards two-way coupling is one-way coupling, where the agent model reacts to the climate data produced by the CCSM but without affecting the CCSM. Such a one-way coupling can easily be implemented in an "off-line" fashion, where the agent model simply uses precomputed CCSM data sets as input for its climate data. Because of the tremendous amount of CCSM data available for a wide range of future climate scenarios, we believe one-way off-line coupling will be adequate for the vast majority of problems.

Using an off-line coupling approach means that the agent model can only access the data that the CCSM model archived while it was being run. It is possible that the type of data archived is not sufficient for some agent models. For example, studies may show that important effects are sensitive to daily fluctuations in certain quantities as opposed to the monthly means available in the CCSM data sets. In this case, one would have to implement an "on-line" coupling model in which the agents are incorporated directly into the CCSM model. Implementing on-line coupling requires some additional software engineering, but this task is made straightforward by the CCSM design. While the CCSM model is being run on a parallel computer, any component (including a possible agent-based component) can request the precise data it needs from the flux coupler, using the flux coupler's well-documented interfaces.

Once the development of an on-line one-way coupled model is complete, the final step of extending this to a two-way coupled model is straightforward. For example, any human activity predicted by the agent-based model that results in changes in CO<sub>2</sub> production can be incorporated back into the CCSM via the flux coupler.

## 5.3 References

CCSM (Community Climate System Model). 2004. Available at <http://www.ucar.edu> (accessed 4 November 2004).

ESG (Earth System Grid). 2004. Available at <http://www.earthsystemgrid.org> (accessed 1 November 2004).



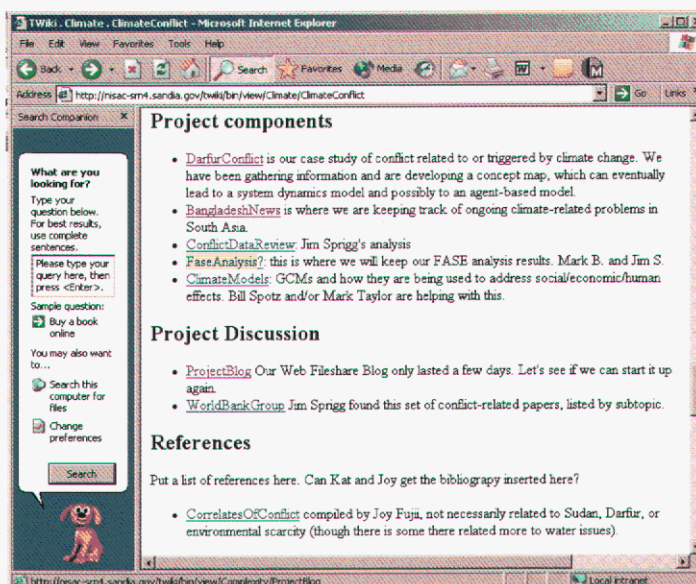
## 6 Using Wiki in a Collaborative Environment

To facilitate communication across our climate project team, we created a Wiki site on Sandia's internal Web. According to the developer of Wiki, Ward Cunningham, Wiki is "the simplest online database that could possibly work" (Cunningham and Leuf 2002). More specifically, Wiki is a component of server software that allows users to create and edit the content of Web pages using any Web browser. In this section, we describe the general features of the Wiki platform that we used, our reasons for choosing the Wiki mode of interactive communication, and impressions of team members regarding the value of the Wiki platform.

### 6.1 General Platform Features

We created a Wiki site for the climate project by making use of TWiki, a Web-based collaborative platform that is used by many companies for project development and planning activities as well as for document management (Matias 2003). TWiki is open-source software that can be downloaded free (Thoeny 2004). Our Wiki is available to members of our project team, as well as to users of Sandia's internal Web, a restricted network. All users on Sandia's internal Web can access our Wiki through a Web address and can view information on the site. However, to change information on the site, users are required to have their own user names and associated passwords on this system. Thus, access to Sandia's internal Web, also an authenticated system, provides limited access into our Wiki. Members of our climate project have full access to the platform features, once registered.

As shown in the inset on the right, a Web site on Wiki has a similar feel to a Web site on the Internet. For example, Web pages are written in HyperText Markup Language (HTML); however, TWiki offers some very powerful additional features. There is no Webmaster. Using HTML, any registered user can create a page on the Web site and edit any content on any page of the Web site. Any registered user can upload and download any file as an attachment to any Web page. The inset shows a portion of a page that one of our team members created. This page contains links to other Web site pages, as shown in the colored hypertext entries. We used this page as an index to the site. Our site also had



a blog, where we could comment on what we were doing at the time and communicate with everyone about our work.

## **6.2 Why We Chose to Use a Wiki**

Our climate project is composed of members who come from several different disciplines and who are physically located in different places, i.e., different buildings at Sandia, offsite from Sandia, and different states as well. The project also had a very short time frame, which pointed to the need for some type of centralized medium through which all would be connected.

The features of the Wiki supported our needs by allowing all team members to be informed about the work in progress at the same time. Members can access documents, change them, and post them (as edited) back to the Wiki at any time. Thus, two or more people working on different topics can make changes to the same document at the same time. In the more traditional communications environment, documents in review are sent to individuals one at a time. Any changes a reviewer makes are generally not known to other members of the team except, perhaps, to the person responsible for updating the document and anyone else the reviewer chooses to send those updates to.

## **6.3 User Impressions of the Wiki**

The time frame of this project was only a few months, so it is too early to evaluate the effectiveness of the Wiki in facilitating communications for our team. Implementing the Wiki was an experiment, and we expected that this type of media may take a longer period for users to get familiar with its capabilities. Nonetheless, we wanted to get a pulse on the prospect of continuing to use the Wiki in the next phase of the climate project and thus queried team members about their impressions of using the platform.

In general, responses to using the Wiki were quite favorable, though most of the team members took advantage of only a few of the features. Several members commented on the value of having documents that are in review accessible to everyone, whether they are posted in HTML on the Web pages or downloadable from the Web pages as attached files. Another team member commented, "It's like a virtual library of our project. I can see what everyone else is doing, which saves all of us time because I don't have to call or email them to find out what is going on." Several members said that the downloading feature was better than email because you didn't have to wonder if you were getting the latest version of a document.

As with any new technology, there was a bit of resistance in getting started. And for some there was a learning curve to take advantage of certain features such as creating and editing Web pages. These actions could require writing in HTML, and not everyone had done this before. The Wiki offers a tutorial on how to generate HTML using simplified formatting symbols, and several team members used that tutorial to get started.

According to the principal investigator on the climate project, there is a certain point where usage of new forms of media is optimal. Given the responses to the user impressions about the Wiki, as well as observations on the site, our team is not quite at that point. However, team members indicated that with more time available, they would likely use the Wiki more often and try out more of its features. Regarding the use of the Wiki for the next phase of the project, one member ventured that the platform could serve the team very well as a central repository for source code and from which to run the new models during the development period.

## 6.4 References

Cunningham, W., and B. Leuf. 2002. "What Is Wiki." 27 June. Available at <http://wiki.org/wiki.cgi?WhatIsWiki> (accessed 24 October 2004).

Matias, N. 2003. "What is a Wiki?" 3 November. Available at <http://www.sitepoint.com/article/what-is-a-wiki/4> (accessed 24 October 2004).

Thoeny, P. 2004. "TWiki<sup>TM</sup> – A Web-Based Collaboration Platform." 12 October. Available at <http://www.twiki.org/> (accessed 24 October 2004).

## 7 Summary of Findings and Lessons Learned

This section summarizes the key findings from the various activities in which we engaged during the initial phase of the climate project.

**Darfur Candidate Case Study.** Monitoring the current events and reviewing the historical literature about the Darfur conflict provided rich insight into the complex and interrelated factors and conditions that can lead to war. The exercise in developing a concept map prompted us to identify the interlocking systemic elements of the conflict and to carefully consider how and in what ways each element can influence and be influenced by other elements.

**Causal Factors of Conflict Study.** Conducting the quantitative analyses to explore the causal factors related to the study of state conflict helped us in a number of ways. First, we identified and summarized representative data that are used in various conflict models, including the COW database and the KOSIMO database. Second, we found through traditional statistical analyses of the data that traditional methods are useful but limited for understanding how to model the effects of macrostructural changes, such as wealth, food, income, and social diversity, on conflict. Traditional methods cannot extrapolate into extreme situations that have occurred in history. Most of the work performed to understand the effects of macrostructural changes on conflict has been empirical, by which we mean that the predictive nature of the analyses is based on identifying patterns in the historical data. Third, we note that traditional statistical methods have inherent limitations, such as measurement error, latent variables, and an inability to model many complex interrelationships.

In our causal factors study, we also found that there are significant relationships between various conditions of a region and its likelihood to engage in conflict. With regard to climate change, we found that food consumption and casualties from natural disasters can affect conflict. For example, climate change could decrease crop yields or increase the likelihood of disasters like flooding and tidal waves, and thereby increase the instability and likelihood of conflict.

**Other Statistical Methods Study.** Our review and analysis of other models include the use of three relatively recent entrants to knowledge analysis for the behavioral and social sciences: QCT (qualitative choice theory), Granger causality, BCD (Bayesian causal discovery), and cointegration. Preliminary assessment of these methods revealed the following:

- (1) Rudimentary cointegration analysis indicates there does appear to be resilient mechanisms that remain functional during conflict. The simulation of these mechanisms can provide robust indicators of conflict dynamics.
- (2) Cursory analysis indicates that accumulating circumstances rather than immediate events dominate the incidence of conflict. Absent external pressures, internal pressures appear to act to restore conditions to a status quo situation.

- (3) QCT does seem able to robustly capture the human (behavioral) process of assimilating information (internal and external circumstances and pressures) that leads to a decision to enter into conflict. QCT robustly simulates both rational and irrational behaviors.
- (4) The cointegration and QCT work further indicates a causality that could be verified by BCD. The determination of causality remains an open issue in conflict analysis that BCD might be able to address.

The use of the preliminary data indicates that environmental change could easily reverse the progress made to-date in stabilizing economies (nations) and, with a lower probability, send the economies into a condition where the probability of conflict increases significantly.

**Development of Coupling Process for Climate and Agent-Based Models.** We found that coupling an agent-based model with the U.S. flagship climate model, the CCSM, requires only straightforward software engineering due to the design of the CCSM. This coupling can be done in a series of small steps leading to the development of a fully coupled agent/climate model. Each step can be tested and verified before moving to the next step.

**Wiki Platform Usage.** As expected, we found that introducing a new form of media requires a sufficient time period in which users can get comfortable and experiment with the new technology. Encouraging members to communicate their work progress through the Wiki does have a payoff, as the Web we created contains contributions by all team members.

## 8 Future Plans and Applications

In this section, we briefly outline our approach to developing a coupled climate/agent-based model and possible applications for its use in related and different fields.

### 8.1 Ongoing Project Development

For the next phase of this LDRD project, computational modeling offers a promising means to develop and test hypotheses about linkages between environmental change and its impact on critical resources and potential subsequent impact on preventing or contributing to conflict. Our multidisciplinary team will develop a suite of computational tools to understand the emergence of unforeseen system thresholds (i.e., “tipping points”) caused by environmental stresses, especially those associated with abrupt or gradual climate change.

We will initially focus on developing a set of simple agent-based models that will take advantage of the work and the results we observed during the initial project start-up period, as documented in this white paper. We will develop actual, not candidate, case studies from the literature, focusing on the loss of arable land due to desertification or rising sea levels, leading to migration and population pressure. Darfur is one area where such processes have well-documented impacts. Other possible areas for review include Egypt, Bangladesh, and China.

Agent behaviors will be further improved by running genetic algorithms on Sandia’s high-performance computers. These algorithms allow us to automatically explore a vast range of potential agent rules using a survival-of-the-fittest methodology to find rules that best identify the key feedbacks that lead to conflict and stability. This will provide insight into local responses leading to emergent phenomena associated with conflict and stability and to seek mitigation strategies for highly idealized cases. We will also incorporate CCSM (Community Climate System Model) climate data, including 100-year simulations using various CO<sub>2</sub> scenarios that are being run for the upcoming IPCC Fourth Assessment Report. Data for this report will be available in late 2004. The highest resolution available will be approximately 150 kilometers, which will be sufficient to determine at a regional level (including substate or transboundary) how changes in rainfall, temperature, and growing season might impact communities and lead to conflict.

We also hope to add formal uncertainty quantification and validation components to the initial activities described above. A mixed climate/agent simulation tool poses significant problems for verification and validation (V&V), which we consider to be essential elements in any decision process that might incorporate the results of this type of simulation. Verification centers on the accumulation of evidence that the feedbacks in our overall model are correctly implemented and provide mathematically accurate answers. Validation is the somewhat more difficult question of whether the associated simulations (or the conceptual models embedded in them) are correct representations of the required phenomena.

Subsequent development will involve combining and migrating the simple agent models to a more complex cellular model in which agents act within cells representing multiple regions. These boundaries will be generated from case studies and may include political, ecological, economic, or culturally defined borders. Agents may attempt to move from one cell to another, based upon factors associated with competition over resources. Climate changes will be introduced as impacting the availability and accessibility of local resources. The model will incorporate representations of key variables related to conflict.

We will also develop a software capability of coupling the CCSM to our agent model to allow the bidirectional feedbacks necessary to model desertification caused by land-use responses. Model parameters will be varied to generate hypotheses about climate-change-induced loss of arable land. Of importance is the capability to change time scales, so we can explore how responses vary under conditions of gradual versus more rapid climate change. We will apply the V&V path forward, identified above, to our application of the CCSM and will begin to develop validation procedures for the agent-based models used in the project. Our ultimate goal is to have a useful agent model and the capability of coupling it to the CCSM. We intend to be running coupled CCSM/agent simulations at 150-kilometer resolution, allowing us to model the feedback loop between the agents and physical climate systems.

## **8.2 Applications**

High-performance computing, coupled with a convergence of capabilities at Sandia, puts this area of research in position for rapid advancement. The results of our ongoing work should be useful in multiple program areas at Sandia, including intelligence analysis and systems approaches to environmental problems. This work supports Sandia's mission to develop technical solutions for complex problems that threaten national security. Currently, analysis of responses to climate change consists of expert judgment and narrative analysis. We expect that coupling agent behaviors to physical processes represents a technical innovation in computational modeling and will lay the groundwork for the development of interdisciplinary technologies capable of analyzing ecological and critical-resource problems.

## Distribution

### External Distribution

1	Sean P. O'Brien Center for Army Analysis 6001 Goethals Road Ft. Belvoir, VA 22060-5230	CAA
1	Richard Colbaugh Department of Management New Mexico Institute of Mining and Technology Socorro, NM 87801	NSA
1	Karen Walker Science and Technology Directorate U.S. Department of Homeland Security Washington, DC 20520	DHS
1	Kristin Glass National Center for Genome Resources 2935 Rodeo Park Drive East Santa Fe, NM 87505	NCGR

### Sandia Internal Distribution

1	0161	Patent & Licensing Office	11500
1	0619	Review & Approval Desk	09216 for DOE/OSTI
2	0899	Technical Library	09616
1	9018	Central Technical Files	08945-1
1	0321	W. J. Camp	09200
1	1110	D. E. Womble	09210
5	0318	J. E. Nelson	09216
20	0318	S. G. Wagner	09209
20	0318	M. B. E. Boslough	09216
1	0196	J. A. Sprigg	09216
1	1109	R. J. Pryor	09216
1	0321	R. W. Leland	09220
1	0749	L. A. Malczynski	06010
1	0839	L. A. McNamara	06924
1	0196	G. A. Backus	09216



1	0318	R. K. Reinert	09216
1	0318	C. R. Jorgensen	09216
1	0196	M. D. Peters	09216
1	0318	M. A. Taylor	09216
1	0316	S. S. Dosanjh	09233
1	0316	W. F. Spatz	09233
1	0839	G. Yonas	16000
1	0839	T. H. Karas	16000
1	0839	C. M. Johnson	16000
1	0123	H. R. Westrich	01011
1	1221	R. L. Rigby	15130
1	0703	W. B. Cox	06218
1	0839	P. S. Gingrich	16000
1	0735	K. M. Murphy	16001
1	0370	T. G. Trucano	09211
1	1138	A. J. Scholand	06221
1	1206	M. J. Carr	05932
1	0839	W. B. Jones	16000
1	0839	N. K. Hayden	05925
1	0839	J. H. Moore	16000
1	1138	M. A. Ehlen	06221
1	0776	J. S. Stein	06852
1	0735	T. S. Lowry	06115
1	0735	H. D. Passell	06115
1	0735	R. E. Finley	06115
1	0741	M. L. Tatro	06200
1	0701	P. B. Davies	06100
1	1375	D. E. Ellis	06900
1	0711	C. Shaw	06100
1	1219	T. P. Wright	05923
1	0720	R. C. Pate	06115
1	0720	A. A. M. Hasan	06143
1	0720	P. I. Pohl	06143
1	0703	M. P. Ross	06218
1	1363	M. T. Olascoaga	06920
1	1373	K. L. Biringer	06924
1	1373	A. L. Pregoner	06920
1	1374	J. H. Saloio Jr.	06927
1	1371	R. M. Huelskamp	06926
1	1379	W. G. Rhodes III	06952
1	1377	M. J. McDonald	06956
1	0779	H-N. Jow	06874
1	1373	A. H. Mohagheghi	06924

**LIBRARY DOCUMENT  
DO NOT DESTROY  
RETURN TO  
LIBRARY VAULT**