

Technical Political Forecasting: Opportunities and Challenges

Philip A. Schrodt

Parus Analytical Systems
schrodt735@gmail.com

Institute for Security Studies Innovation Retreat
Pretoria, South Africa
20 May 2014

Outline

- ▶ Why the technical environment is very different than ten years ago
- ▶ Technical political forecasting
- ▶ Event data
- ▶ Collaborative prediction; new social media
- ▶ What can we forecast?
- ▶ Some questions for ISS

Why 2014 isn't 2004

- ▶ Jeremy Rifkin [NYT March 2014]: The most disruptive technologies in the current environment combine network effects with zero marginal cost
- ▶ Key: zero *marginal* costs: open source software is only “free-as-in-puppy” and specialized infrastructure may still need to be built
- ▶ Examples
 - ▶ Operating systems: Linux
 - ▶ Statistical software: R
 - ▶ Encyclopedia: Wikipedia
 - ▶ Commercial examples: Google, Facebook, Twitter, YouTube, iTunes, iPhone and Android apps and eBay create an infrastructure and then have users provide the content for free

Data!





Minorities at Risk



UPPSALA
UNIVERSITET

PENNSTATE



UCDP
GEOREFERENCED EVENT DATASET

CIRI Human Rights Data Project
www.humanrightdata.org



An Open-Source Application for
Publishing, Citing and Discovering Research Data

Universität
Konstanz



Polity

GTD
Global Terrorism Database

Geolocation: Uppsala Conflict Data Project Geospatial

Welcome to the UCDP GED - Uppsala Conflict Data Program's Georeferenced Event Dataset

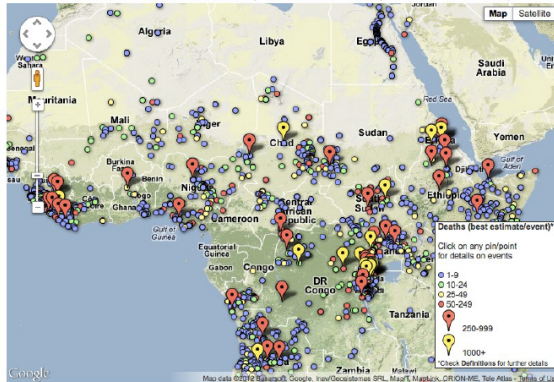
The interactive map below covers all the locations geocoded as part of GED. Drag and move the red and grey thumbs of the slider below to filter for particular intervals, years, or even particular months or days. Use the filters to select only one of the three UCDP types of violence. Clicking on each dot or pin will display information regarding events taking place in those respective locations. [Download data here.](#)

1989

2010

Filter for:

☒ All violence ☐ State-based conflicts only ☐ Non-state conflicts only ☐ One-sided conflicts only



The GED is the product of two and a half years of work at the **Department of Peace and Conflict Research, Uppsala University**. The UCDP GED contains conflict data disaggregated spatially and temporally down to the level of the individual incidents of violence. For more details please see the [About UCDP](#) link above.

Computing Power

Control Data Corporation 3600
(ca.1965)
32 K (48-bit) RAM memory
1 processor
~1-million operations per second
Output: line printer



Penn State High Performance Computing Facility
15 cluster computers
100 to 2000 2.66 Ghz processors in each cluster
~50 Gb RAM accessible to each processor
130 Tb disk space
4 interactive visualization rooms



Motorola Razr
16 Gb RAM memory
Dual processor
~500-million operations per sec
540 x 860 color display

Open Source Software



Open source, open access, open collaboration

- ▶ There is a strong if incomplete norm towards open sharing of data and methods
 - ▶ Big Data analytical methods can be very sensitive to assumptions and if you can't see all of the code, you probably don't really know what it is doing. “Black box” solutions are increasingly unacceptable.
- ▶ By all available evidence, US government forecasting projects are using similar methodologies to those available in open sources; in fact they are probably lagging somewhat behind this
- ▶ We now have significant NGO and academic work, and an international “epistemic community” has developed around the topic.

Technical Political Forecasting: The Debate



► **ARGUMENT**

PRINT | TEXT SIZE | EMAIL | SINGLE PAGE

Why the World Can't Have a Nate Silver

The quants are riding high after Team Data crushed Team Gut in the U.S. election forecasts. But predicting the Electoral College vote is child's play next to some of these hard targets.

BY JAY ULFELDER | NOVEMBER 8, 2012

Vs.



► **ARGUMENT**

PRINT | TEXT SIZE | EMAIL | SINGLE PAGE

Predicting the Future Is Easier Than It Looks

Nate Silver was just the beginning. Some of the same statistical techniques used by America's forecaster-in-chief are about to revolutionize world politics.

BY MICHAEL D. WARD, NILS METTERNICH | NOVEMBER 16, 2012

Factors encouraging technical political forecasting

- ▶ Conspicuous failures of existing methods: end of Cold War, post-invasion Iraq, Arab spring
- ▶ Success of forecasting models in other behavioral domains
 - ▶ Macroeconomic forecasting [maybe...]
 - ▶ Elections: Nate Silver effect
 - ▶ Demographic and epidemiological forecasting
 - ▶ Famine forecasting: USAID FEWS model
 - ▶ Example: statistical models for mortgage repayment were quite accurate
- ▶ Technological imperatives
 - ▶ Increased processing capacity
 - ▶ Information available on the web
- ▶ Decision-makers now expect visual displays of analytical information, which in turn requires systematic measurement
 - ▶ “They won’t read things any more”

This must be important: it's in *The Economist*!

The science of civil war

What makes heroic strife

Computer models that can predict the outbreak and spread of civil conflict are being developed

Apr 21st 2012 | from the print edition



95



40



Large Scale Conflict Forecasting Projects

- ▶ State Failures Project 1994-2001
- ▶ Joint Warfare Analysis Center 1997
- ▶ FEWER [Davies and Gurr 1998]
- ▶ Center for Army Analysis 2002-2005
- ▶ Swiss Peace Foundation FAST 2000-2008
- ▶ Political Instability Task Force 2002-present
- ▶ DARPA ICEWS 2007-present
- ▶ IARPA ACE and OSI
- ▶ Peace Research Center Oslo (PRIO) and Uppsala University UCDP models
- ▶ US Holocaust Memorial Museum Prediction Poll

Convergent Results

- ▶ Most models require only a [very] small number of variables
- ▶ Indirect indicators—famously, infant mortality rate as an indicator of development—are very useful
- ▶ Temporal autoregressive effects are huge: the challenge is predicting onsets and cessations, not continuations
- ▶ Spatial autoregressive effects—“bad neighborhoods”—are also huge
- ▶ Multiple modeling approaches generally converge to similar accuracy
- ▶ 80% accuracy in the 6 to 24 month forecasting window occurs with remarkable consistency: few if any replicable models exceed this, and models below that level can usually be improved
- ▶ Measurement error on many of the dependent variables—for example casualties, coup attempts—is still very large
- ▶ Forecast accuracy does not decline very rapidly with increased forecast windows, suggesting long term structural factors rather than short-term “triggers” are dominant. “Trigger models” more generally do poorly except as *post hoc* explanations.

What are event data?

WEIS Primary Categories

01	Yield	11	Reject
02	Comment	12	Accuse
03	Consult	13	Protest
04	Approve	14	Deny
05	Promise	15	Demand
06	Grant	16	Warn
07	Reward	17	Threaten
08	Agree	18	Demonstrate
09	Request	19	Reduce Relationship
10	Propose	20	Expel
		21	Seize
		22	Force

News Story Example: Example: 18 December 2007

BAGHDAD. Iraqi leaders criticized Turkey on Monday for bombing Kurdish militants in northern Iraq with airstrikes that they said had left at least one woman dead.

The Turkish attacks in Dohuk Province on Sunday—involving dozens of warplanes and artillery—were the largest known cross-border attack since 2003. They occurred with at least tacit approval from American officials. The Iraqi government, however, said it had not been consulted or informed about the attacks.

Massoud Barzani, leader of the autonomous Kurdish region in the north, condemned the assaults as a violation of Iraqi sovereignty that had undermined months of diplomacy. “These attacks hinder the political efforts exerted to find a peaceful solution based on mutual respect.”

New York Times, 18 December 2007

http://www.nytimes.com/2007/12/18/world/middleeast/18iraq.html?_r=1&ref=world&oref=slogin (Accessed 18 December 2007)

TABARI Coding: Lead sentence

BAGHDAD. Iraqi leaders criticized Turkey on Monday for bombing Kurdish militants in northern Iraq with airstrikes that they said had left at least one woman dead.

Event Code: 111

Source: IRQ GOV

Target: TUR

Event Code: 223

Source: TUR

Target: IRQKRD REB

TABARI Coding: First event

BAGHDAD. Iraqi leaders criticized Turkey on Monday for bombing Kurdish militants in northern Iraq with airstrikes that they said had left at least one woman dead.

Event Code: 111

Source: IRQ GOV

Target: TUR

Event Code: 223

Source: TUR

Target: IRQKRD REB

TABARI Coding: Actors

BAGHDAD. Iraqi leaders criticized Turkey on Monday for bombing Kurdish militants in northern Iraq with airstrikes that they said had left at least one woman dead.

Event Code: 111

Source: IRQ GOV

Target: TUR

Event Code: 223

Source: TUR

Target: IRQKRD REB

TABARI Coding: Agent

BAGHDAD. Iraqi leaders criticized Turkey on Monday for bombing Kurdish militants in northern Iraq with airstrikes that they said had left at least one woman dead.

Event Code: 111

Source: IRQ GOV

Target: TUR

Event Code: 223

Source: TUR

Target: IRQKRD REB

TABARI Coding: Second event

BAGHDAD. Iraqi leaders criticized Turkey on Monday for bombing Kurdish militants in northern Iraq with airstrikes that they said had left at least one woman dead.

Event Code: 111

Source: IRQ GOV

Target: TUR

Event Code: 223

Source: TUR

Target: IRQKRD REB

TABARI Coding: Second event target

BAGHDAD. Iraqi leaders criticized Turkey on Monday for bombing Kurdish militants in northern Iraq with airstrikes that they said had left at least one woman dead.

Event Code: 111

Source: IRQ GOV

Target: TUR

Event Code: 223

Source: TUR

Target: IRQKRD REB

TABARI Coding: Agent

BAGHDAD. Iraqi leaders criticized Turkey on Monday for bombing Kurdish militants in northern Iraq with airstrikes that they said had left at least one woman dead.

Event Code: 111

Source: IRQ GOV

Target: TUR

Event Code: 223

Source: TUR

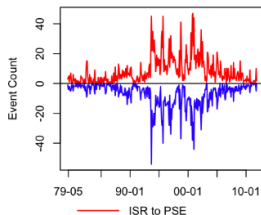
Target: IRQKRD REB

Quad Counts

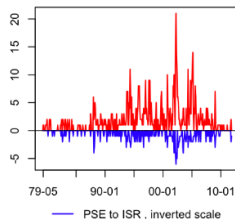
- ▶ Verbal Cooperation (VERCP): The occurrence of dialogue-based meetings (i.e. negotiations, peace talks), statements that express a desire to cooperate or appeal for assistance (other than material aid) from other actors. CAMEO categories 01 to 05.
- ▶ Material Cooperation (MATCP): Physical acts of collaboration or assistance, including receiving or sending aid, reducing bans and sentencing, etc. CAMEO categories 06 to 09.
- ▶ Verbal Conflict (VERCF): A spoken criticism, threat, or accusation, often related to past or future potential acts of material conflict. CAMEO categories 10 to 14.
- ▶ Material Conflict (MATCF): Physical acts of a conflictual nature, including armed attacks, destruction of property, assassination, etc. CAMEO categories 15 to 20.

KEDS Project Levant Data, 1979-2010

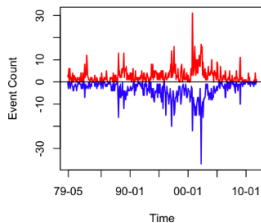
Verbal Cooperation



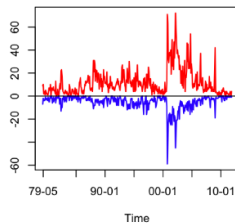
Material Cooperation



Verbal Conflict

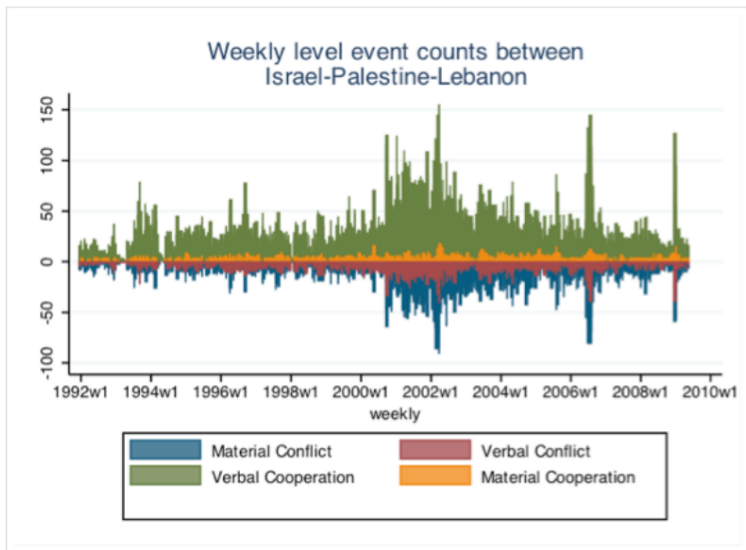


Material Conflict



KEDS Project Levant Data, 1992-2010

Visualization by Jay Yonamine



Key observation from contemporary event data

The combination of fully automated coding and the increasing number of reports on the web means that we now have an inexpensive “instrument” for systematically monitoring global political behavior in real time.

To date, however, this is largely done using the major international news services: Reuters, Agence France Press, Associated Press, BBC World Monitor, Xinhua,

Why event data are well suited for predicting political change

- ▶ Structural indicators such as GDP, infant mortality, regime type, past or adjacent conflict change too slowly
 - ▶ They nonetheless affect the overall probability
- ▶ Social media indicators change too quickly for long range forecasts
 - ▶ This is also a very new type of data
 - ▶ Though it may be possible to use aggregate measures
- ▶ Newsworthy events are “just right”
 - ▶ As existing models have demonstrated
 - ▶ Which is why they are “newsworthy”
 - ▶ Structural indicators either are reflected in the patterns of events, or can be additional covariates

Phoenix Data System

- ▶ Open source, open access, open collaboration
- ▶ Hosted on GitHub:
<https://openeventdata.github.io/>
- ▶ Fully modular open-source pipeline to produce daily event data from web sources which can be implemented on inexpensive “cloud” server systems
- ▶ Python-language event coder based on the Stanford CoreNLP English-language parser
- ▶ Geolocation: Penn State “GeoVista” project coder, UT/Dallas coder
- ▶ Open Event Data Alliance: membership organization to provide at least one source of daily updates with 24/7/365 data reliability. Ideally, multiple such data sets rather than “one data set to rule them all”

Extending the event ontologies

Existing event ontologies were derived from earlier Cold War event frameworks and consequently miss substantial amounts of political behavior that is currently relevant.

- ▶ natural disaster
- ▶ disease
- ▶ criminal activity
- ▶ financial activity
- ▶ refugees and related humanitarian issues
- ▶ human rights violations
- ▶ electoral and parliamentary activity

Collaborative Prediction

Philip Tetlock. *Expert Political Judgment*

The human political forecasting is only slightly more accurate than chance: “dart-throwing chimp.”

Forecasts in the popular media—“pundits”—are actually *worse* than chance because forecasters in the popular media are encouraged to make sensational forecasts because fundamentally they are entertainers, not analysts.

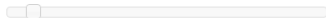
“Good Judgment Project”: A small number of “super forecasters” perform significantly better than chance over long periods of time and large numbers of questions. Furthermore these individuals exhibit common characteristics and strategies, and to a limited extent, these can be taught. *Economist: The Year 2014*

Holocaust Memorial Museum Forecasting Project

Before 1 January 2015, will an episode of state-led mass killing occur in Mali?

[» more info](#)

Move the slider to change the chances this will happen:



Your Forecast: % chance

SAVE FORECAST

Set an alert about this question

QUESTION ENDS

December 31, 2014 @ 12:00pm EST

CURRENT CONSENSUS

16%

CHANGE TODAY

no change today (0.00%)

CATEGORY

Sub-Saharan Africa

TAGS

Mali masskilling State

CREATED BY

[Jay](#) (all their questions)



FORECASTS MADE

18 (Most: weltonchang)

BACKGROUND INFORMATION

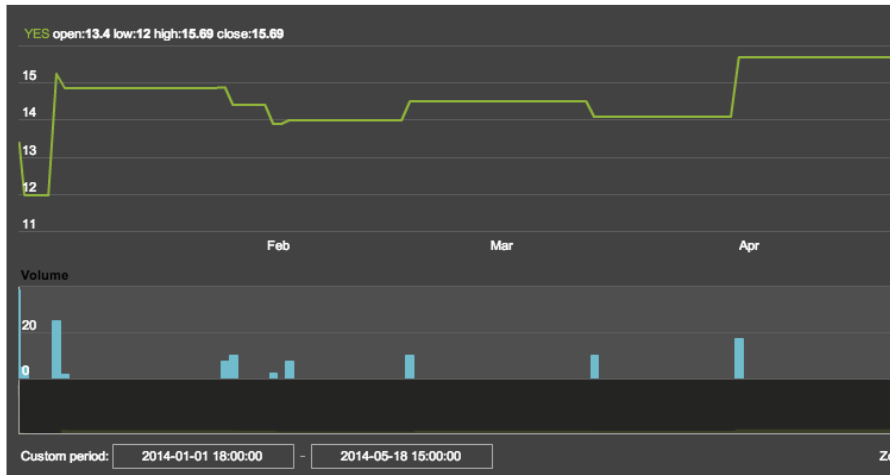
Our 2013 statistical risk assessments identified Mali as one of the 30 countries worldwide most susceptible to an onset of state-led mass killing, and Mali ranked third in our [wiki survey on the same topic for 2014](#).

We consider a state-led mass killing to have occurred if the deliberate actions of state agents or other groups acting at their behest result in the deaths of at least 1,000 noncombatant civilians over a period of one year or less.

- **State-led** refers to cases in which the relevant violence is carried out by uniformed troops, police, or other agents of state security, or by other groups acting at the behest of government officials. In cases where the state's role is ambiguous, we look for evidence of government encouragement of violence or coordination with state policies or military operations.
- A **noncombatant civilian** is any person who is not a current member of a formal or irregular military organization and who does not apparently pose an immediate threat to the life, physical safety, or property of other people.
- The reference to **deliberate actions** distinguishes mass killing from deaths caused by natural disasters, infectious diseases, the accidental killing of civilians during war, or the unanticipated consequences of other government policies. Fatalities should be considered intentional if they result from actions designed to

Holocaust Memorial Museum Forecasting Project

CONSENSUS TREND



Categories of ACE Questions

- ▶ Leadership Turnover and Elections in Stable Democracies
- ▶ Leadership Turnover and Social Change in Authoritarian Regimes
- ▶ Economic and Diplomatic Decisions by International Organizations
- ▶ Negotiation Processes
- ▶ Macro-economic Indicators and Financial Markets
- ▶ Military Actions, Casualty Counts, and Refugee Flows
- ▶ Legal Proceedings Within State Boundaries

What about new social media?



Mainstream media

What about new social media?



Mainstream media



Internet and
new social media

New social media

The good

- ▶ Widely available to elites
- ▶ More or less uncensored in open many societies
- ▶ Should provide early information on changing sentiment prior to observing actual collective action

The bad

- ▶ No filters and mostly politically irrelevant:
- ▶ “Wanna getta pizza? ;)”
- ▶ Easily manipulated by anyone—business, government, NGOs—who wants to go to the trouble of doing so

The ugly

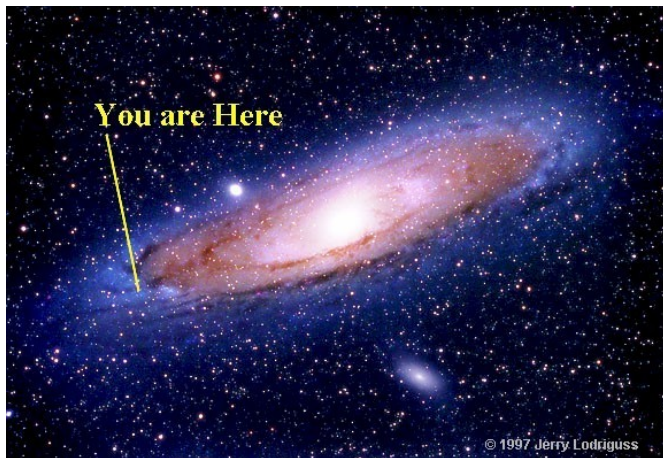
- ▶ No standardization of content

Utility in prediction

- ▶ Multiple studies show this seems to work in the 6 to 48 hour range

Challenge: Black swans

Ideal forecasting targets are neither too common nor too frequent



Good Judgment Project: look for events with a 10% probability

Black swan: Irene Country Lodge, 19 May 2014



Challenge: distinguishing black swans from rare events

Black swan: an event that has a low probability even conditional on other variables

Rare event: an event that occurs infrequently, but conditional on an appropriate set of variables, does not have a low probability

Medical analogy: certain rare forms of cancer appear to be highly correlated with specific rare genetic mutations. Conditioned on those mutations, they are not black swans.

Another important category: high probability events which are ignored. The “sub-prime mortgage crisis” was the result of the failure of a large number of mortgage which models had completely accurately identified as “sub-prime” and thus likely to fail. This was not a low probability event.

Upton Sinclair: It is hard to persuade someone to believe something when a great deal of money can be made not believing it.

Irreducible sources of uncertainty

- ▶ Specification error: no model of a complex, open system can contain all of the relevant variables;

Consulting
Group
© Knowledge Limited 2009



Irreducible sources of uncertainty

- ▶ Specification error: no model of a complex, open system can contain all of the relevant variables;
- ▶ Measurement error: with very few exceptions, variables will contain some measurement error
 - ▶ presupposing there is even agreement on what the “correct” measurement is in an ideal setting;
 - ▶ Predictive accuracy is limited by the square root of measurement error: in a bivariate model if your reliability is 80%, your accuracy can't be more than 90%
- ▶ Free will
 - ▶ Rule-of-thumb from our rat-running colleagues:
“A genetically standardized experimental animal, subjected to carefully controlled stimuli in a laboratory setting, will do whatever it wants.”
- ▶ Quasi-random structural error: Complex and chaotic deterministic systems behave as if they were random under at least some parameter combinations

Balancing factors which make behavior predictable

- ▶ Individual preferences and expectations, which tend to change very slowly
- ▶ Organizational and bureaucratic rules and norms
- ▶ Structural constraints:
the Maldives will not respond to climate-induced sea level rise by building a naval fleet to conquer Singapore.
- ▶ Choices and strategies at Nash equilibrium points

Paradox of political prediction

Political behaviors are generally highly incremental and vary little from day to day, or even century to century (Putnam).

Nonetheless, we *perceive* politics as very unpredictable because we focus on the unexpected (Kahneman).

Consequently the only “interesting” forecasts are those which are least characteristic of the system as a whole. However, only some of those changes are actually predictable.

Ethical concerns

- ▶ Thus far, we've generally had the luxury of no one paying attention to any of our predictions : what if governments do start paying attention?
 - ▶ “Policy relevant forecast interval” is around 6 to 24 months
 - ▶ USAID/FAO famine forecasting model
 - ▶ It is *possible* that our models could become less accurate because crises are being averted, but I don't see that happening any time soon.
- ▶ Difficulties in getting *anyone*, including experts (see Kahneman, Tetlock), to correctly interpret probabilistic forecasts
- ▶ Possible impact on sources
 - ▶ Local collaborators
 - ▶ Journalists (cf. Mexico)
 - ▶ NGOs to the extent we are using their information

Some ideas to consider

- ▶ Where are probabilistic forecasts—the likelihood something will occur, but not the exact time and place—useful?
- ▶ Assuming that Phoenix and other systems can provide event data from the major international sources, how much additional information could be added by
 - ▶ Local sources: if useful, what additional languages are needed?
 - ▶ New Social Media: Twitter, Facebook and whatever is next
- ▶ Are there event patterns that would allow either the creation or recalibration of structural indicators, either providing measures for structural characteristics which aren't readily available (corruption, organized crime), or identifying situations where the structural measures are not consistent with observed events based on comparable cases?
- ▶ Could customized NGO reports add to the existing media reports? Could this be done securely? Could it be sustained over time?

Thank you

Email: schrodt735@gmail.com

Slides:

<http://eventdata.parusanalytics.com/presentations.h>

Software: <https://openeventdata.github.io/>

Papers:

<http://eventdata.parusanalytics.com/papers.html>