

Predicting Armed Conflict Using Machine Learning

Graig R. Klein, Binghamton University

Nicholas P. Tatonetti, Columbia University

Our Goal...

- Empirical Political Science typically = Regression Analysis
- Movements toward Big Data and Machine Learning are not new, but comparatively young & less used than in other sciences
- Attempt to use a newer methodology to help forecast the onset of armed conflict
 - We focus exclusively on Asia, Middle East & North African and The Americas (excluding the U.S.)
 - The data we use are from 1990-2013. We build an algorithm based on data from 1990-2010; then test our algorithm on data from 2011-2013. And, we present forecasts for future armed conflict.

State Response (Chenoweth & Stephan 2008, 2011)
Political Salience of Cultural/Ethnic Differences (Posner 2004; Cederman et al. 2012)
Ethnic Fragmentation/Ethnicity (Tilly 2005)
Economic & Social Motivations (Starr 1994)
Internet Communication & Foreign Exchange (Golder 1995; Collier 2011)
Population Pressure/Farmers' Irrigation (Gardner 1999; Grafton-Cardiff 2003)
Private benefits & Patronage (Hendy & Feiler 1966)
Inequality & Uncertainty of Response (Gurr 1970; Fearon & Laitin 1999; Davenport
(Chenoweth & Stephan 2008, 2011)
State Sanctioned Coercion/Regime Violence (Gurr 1970; Fearon & Laitin 1999; Davenport
(Gurr & Lichbach 1981) Inverted-U (Lichbach 1981)
Policy (Muller & Soberg)

$$Y = x_1 + x_2 + x_3 + x_4 \dots + e$$

A Recent Forecast Model

- Bell, et al. (2013) → Coercion, Capacity & Coordination
 - Build of Gurr & Moore (1997) & O'Brien (2002)
 - Build of the opportunity & willingness arguments (Most & Starr, 1989)
- Coercion = violations of physical integrity → citizens' acceptance of political violence
 - Increases willingness
 - Torture, Extrajudicial Killing, Disappearances, Political Imprisonment
- Capacity = ability of state to protect its power & decrease opportunity for mobilization
 - Decreases Opportunity
 - GDP per Capita, Electric Power Consumption, Military Personnel
- Coordination = availability & ease of cooperation, organization, mobilization
 - Increases Opportunity
 - Freedom of Association, Mobile Phones per 100, Internet Users per 100, Non-violent protest

Our Approach to Forecasting

- We start with a similar theoretical foundation
- But instead of selecting specific measures to test theoretical expectations, we combine multiple theoretical foundations into one model by writing an algorithm to assess patterns in 'Big Data'
- We can now include multiple theories into one statistical model

Theory Foundations & Our Measures

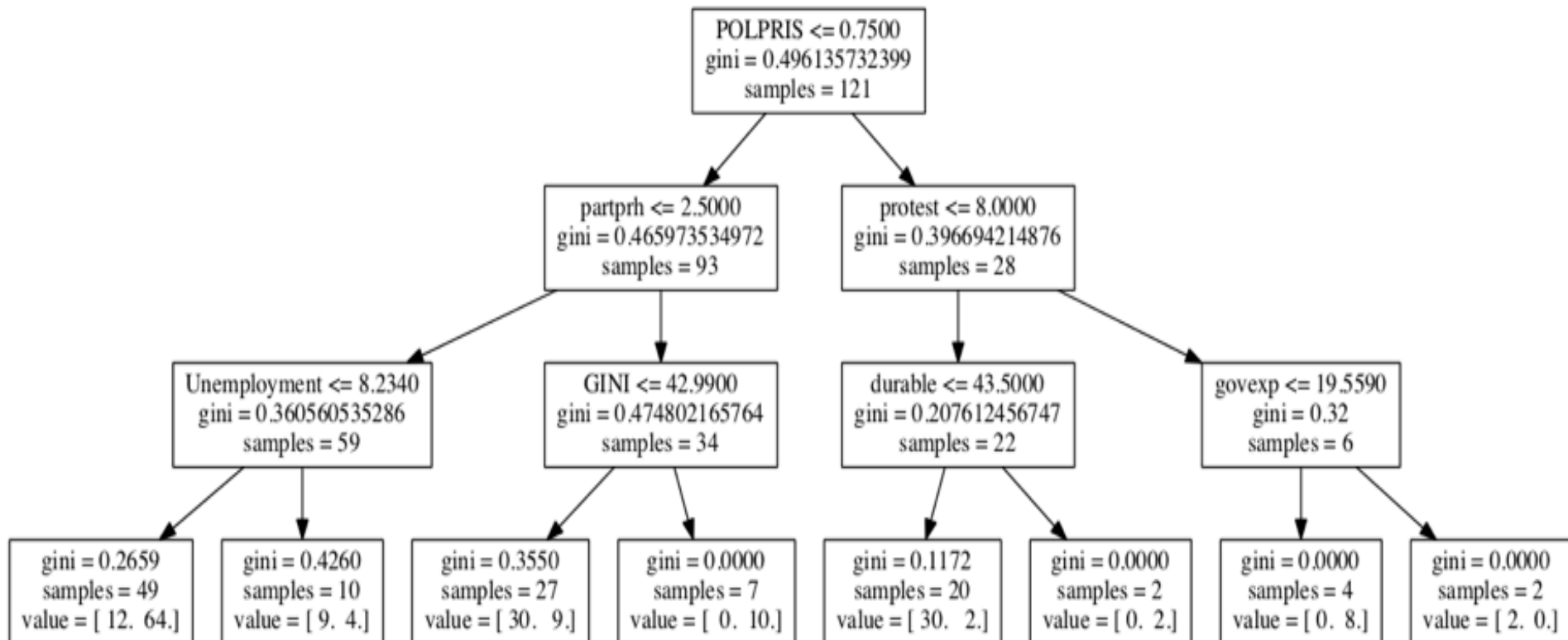
- **Mobilization**
 - Protest Activity
 - Protest Size (# of participants)
 - Total Population
 - Urban Population & Urban Population Growth
 - Freedom of Association & Freedom of Speech
 - Number of Ethnopolitically Relevant Groups
 - Size of Excluded Population Relative to Total Population
 - Size of Powerless Population (%)
 - Size of Largest Excluded Group (%)
 - Mobile Phones per 100 people
 - Internet Users per 100 people
- **Action-Reaction**
 - State Response to Protests
 - Protester Violence
- **Geographic Conditions**
 - Mountainous Terrain
 - Forest Area
- **Resource Acquisition & Exploitation**
 - Oil Rents
 - National Income
- **History**
 - Ethnic Fractionalization
 - Time Under Colonial or Imperial Rule
 - Past Armed War & Ongoing Armed Conflict
 - Past Armed Conflict & Positive Change in Polity Score
- **Relative Deprivation, Want Formation & Want Formation**
 - Net National Debt
 - Protester Demands
 - Government Expenditure
 - Unemployment Rate
 - GDP per Capita & Change in GDP per Capita
 - Food Deficit
 - Male Youth Unemployment
 - GINI
- **Political Institutions**
 - Polity Score & Change in Polity Score
 - Military Expenditure
 - Human Rights (Torture, Killings, Disappearances, Political Imprisonment)
 - Women's Political Rights
 - FDI Net Inflow (BoP)
 - Competitiveness of Executive Recruitment
 - Political Competition & Prohibited Political Parties
 - Constitution Provision for Integration of Ethnic Groups
 - Constitutional Event & Type
 - Constitutional Right to Form Political Parties
 - Constitutional Means for Handling Crimes by Previous Regime
 - Number of Consecutive Regime Periods
 - Duration of Current Regime
 - Regime Type & Previous Regime Type
 - Size of Largest Party in Legislature

Machine Learning

- Derives a statistical model from the observed data
 - E.g. a logistic regression model of the incidence of conflict as a function of a country's attributes
- Big data sets introduce challenges
 - A large number of independent variables can produce meaningless models
- Ensemble methods correct for this issue by combining many models, each is only built on a small subset of the data

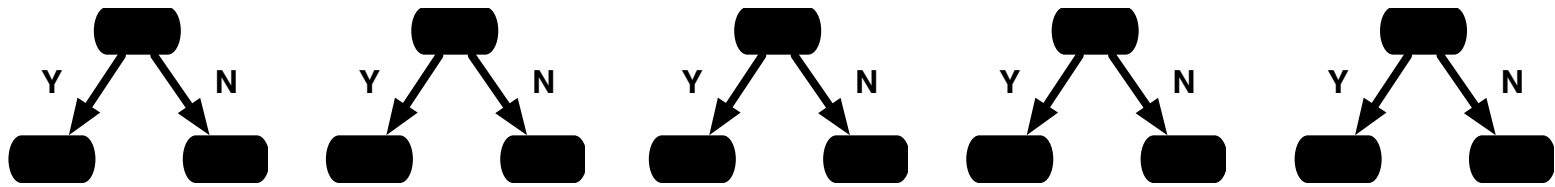
Random Forest

- An ensemble machine learning method
- Combines many Decision Trees together to form an unbiased predictor
- A single decision tree:



Random Forest

- Any particular tree may not be robust, but the *ensemble* of trees has been shown to be a robust and unbiased predictor

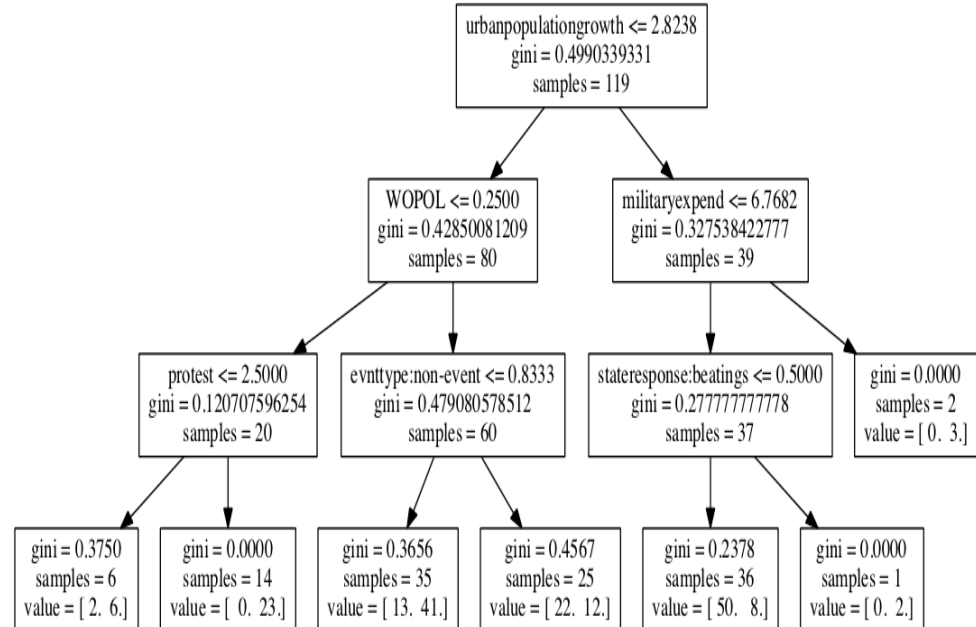
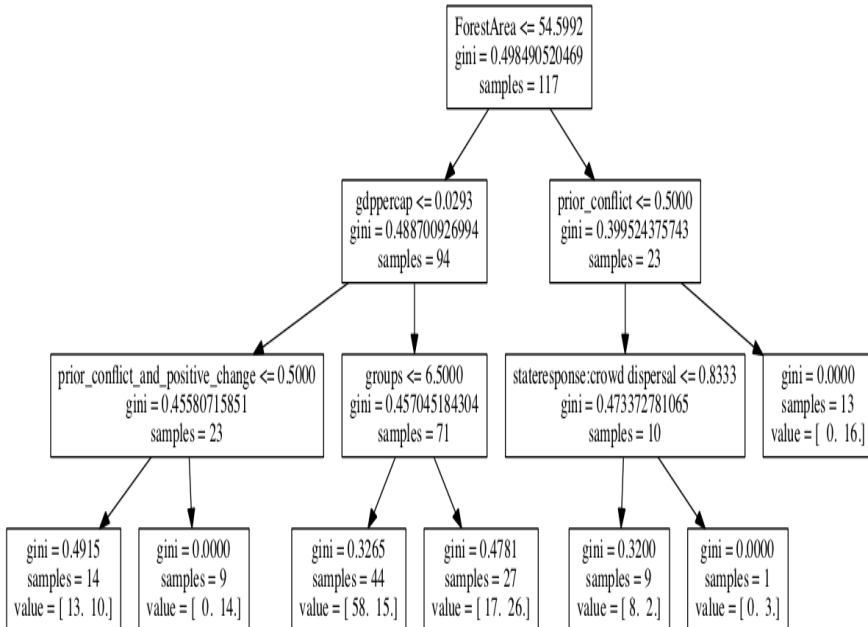
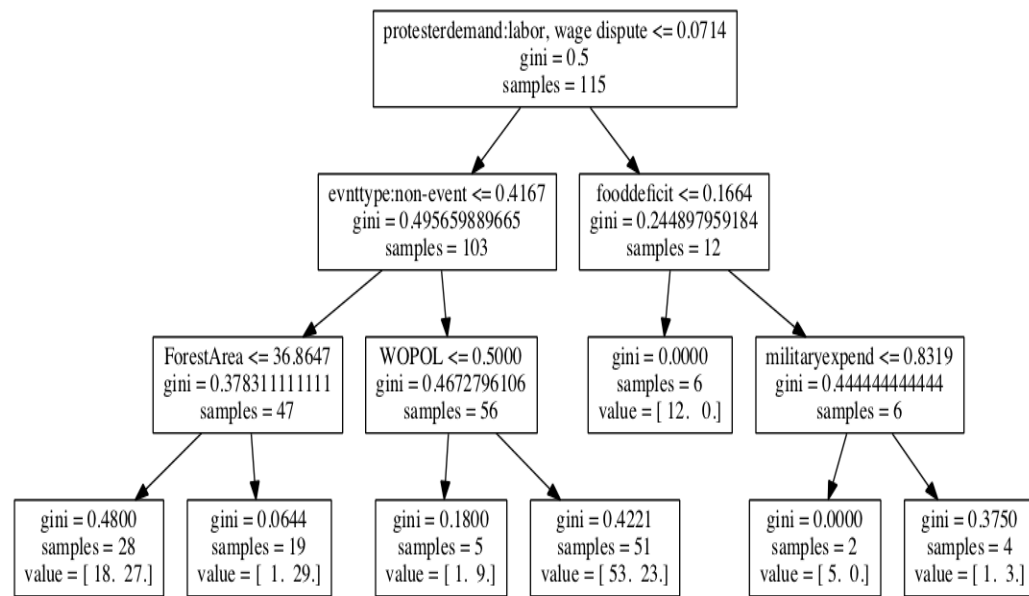
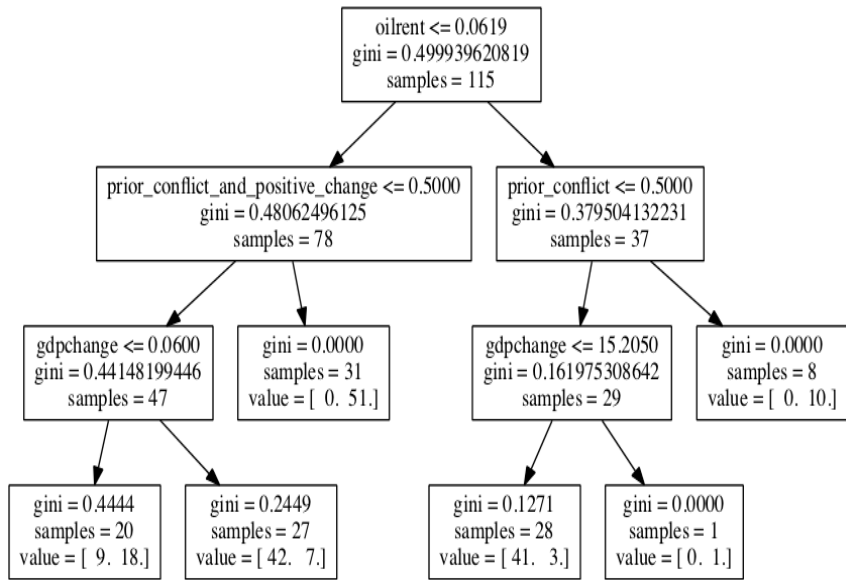


- Each tree gets a vote and the decision with the greatest number of votes wins

Trees in the Random Forest

- Aggregating the trees provides observation of general patterns in the data and conditions leading to the onset of armed conflict

Examples of Trees in Our Forest



Diagnostics

	Accuracy	Recall	Precision
O'Brien's (2010) Ideals	80%	80%	70%
Bell et al. (2013)	68%	67%	71%
Klein & Tatonetti	86%	78%	93%

- Accuracy = # of correct predictions / # of predictions made
- Recall = # of correctly predicted increases / # of increases occurred
- Precision = # of correctly predicted increases / # of increases predicted to occur

Forecasting Conflict

- We move from our algorithm & patterns of conflict from 1990-2010 to forecast the onset of armed conflict from 2011-2013
- In the countries in our sample, there were 11 onsets of armed conflict from 2011-2013
- 2011-2013 is our test set
 - The algorithm never saw these incidents until testing predictions
 - Also, never exposed to data beyond 2010
 - We keep ourselves & our predictions completely naïve

Predicted Probabilities of Conflict 2011-2013

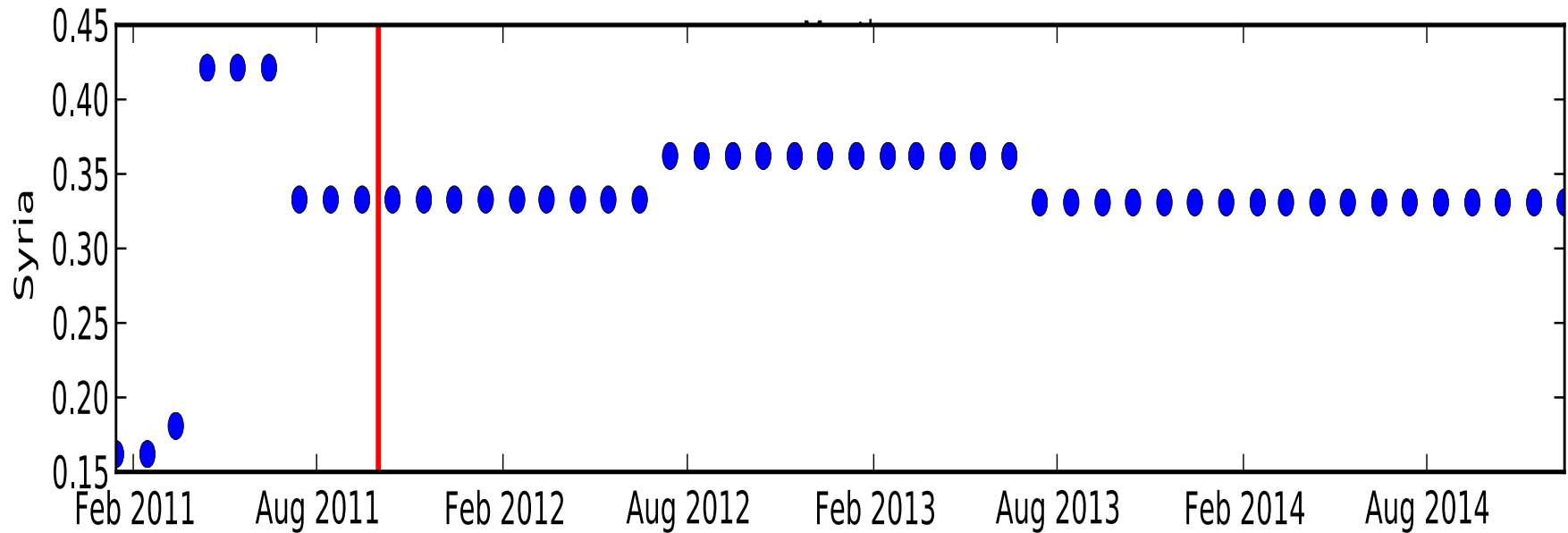
*Baseline probability of armed conflict during the time period is .05

➤ 11 total onsets in 228 country years

Country	Predicted Probability of Armed Conflict	Date	Armed Conflict	Other Political Events
Burma	.75	6-27-2011	1	
India	.74	12-11-2013	1	
Burma	.69	2-22-2013	1	
Burma	.66	4-29-2013	1	
Pakistan	.63	7-5-2011	1	
Azerbaijan	.57	9-11-2012	1	
Libya	.53	12-31-2012	0	Widespread tribal clashes erupt in 2013
Sri Lanka	.52	12-31-2013	0	
China	.51	12-31-2013	0	Dramatic increase in Uyghur violence in 2014
Sudan	.48	12-31-2012	0	Tribal conflict erupts in Darfur in Jan. 2013

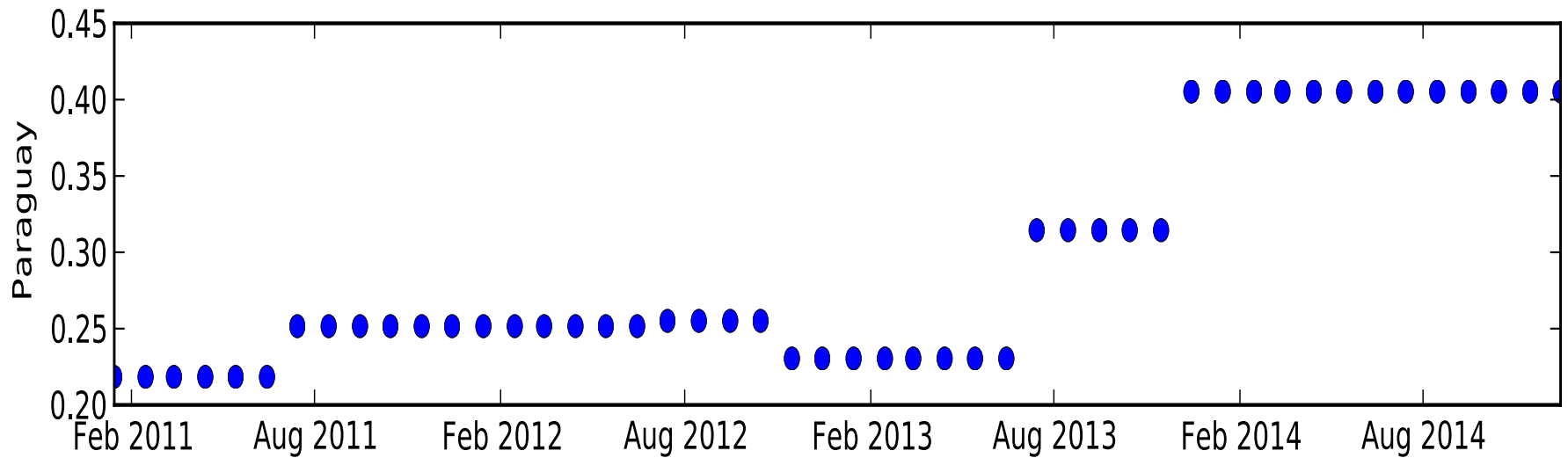
Predicted Probabilities Across Time

- Syria



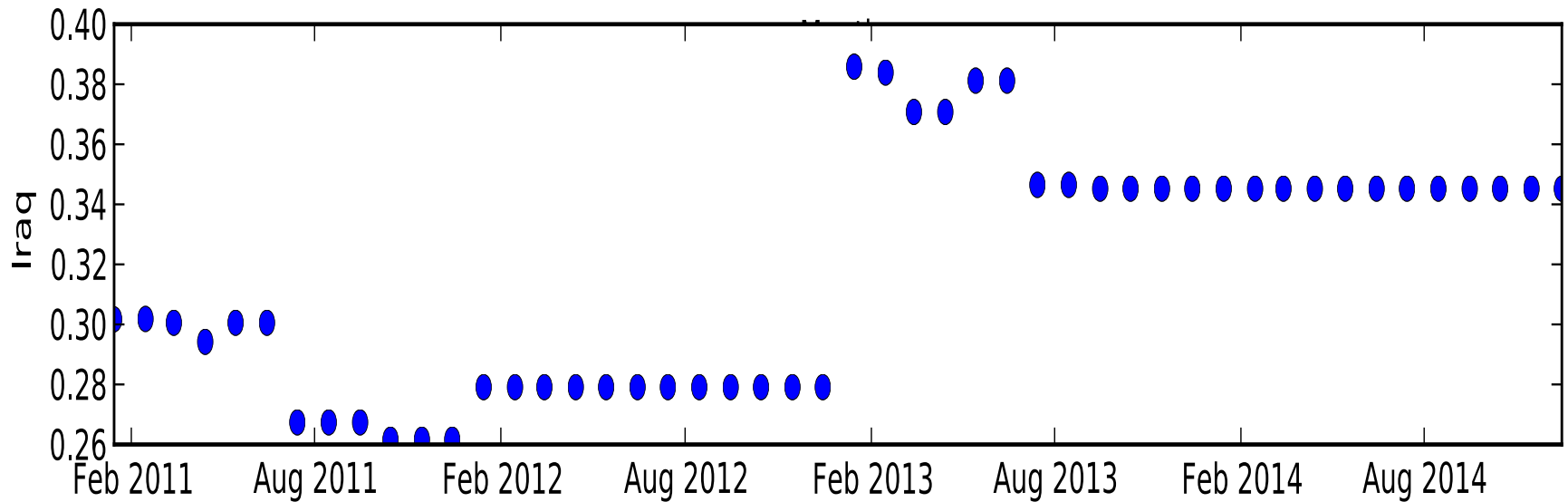
In Today's Forecast

- Paraguay



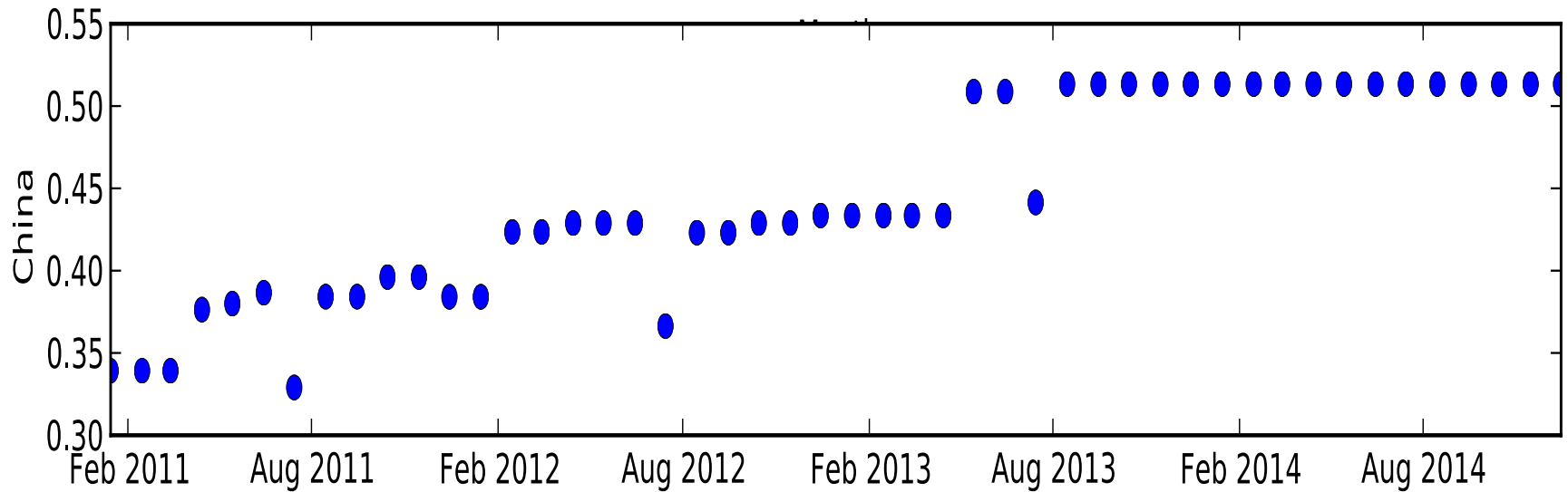
In Today's Forecast

- Iraq



In Today's Forecast

- China



Next Steps

- Expand the dataset to include more countries
- Look to forecast beyond 2014
- Develop a method of splitting very large countries into geographic regions
 - “inherited variables” as there’s no regional specification
 - i.e. Polity Score, Regime Type, etc.
 - Geographic specific variables
 - i.e. Protest location, Terrain, Ethnic Populations