Correlates of War: National Material Capabilities Analysis

Mouhamed Mbengue

University of Rochester, Data Mining (Fall 2025)

Abstract— Analyzing the National Material Capabilities (NMC) dataset of the Correlates of War Project is the aim of the project. For the Composite Index of National Capability (CINC), the study looks at correlations, variable distributions, similarities of countries, and maps differences of capabilities across the world. The results confirm shifting power dynamics, structural correlations, and the dominance of a few strong states.

I. INTRODUCTION

A. Q1: Overview of COW Project

The academic work of the Correlates of War project seeks to understand the characteristics and triggers of war by meticulously gathering data. Started in 1963 by political scientist J. David Singer, then at the University of Michigan, COW, has assisted in the creation of more analytical methodologies for the study of international relations. COW's data collection on state interactions, armed forces, and conflicts, starting in 1816, is unique. Few databases allow such detailed analysis of international relations in the context of war, peace, and power.

Unlike commercial resources, this is an interdisciplinary project maintained by scholars and institutions freely available for academic and research purposes.

II. DATA

A. Q2: Dataset Documentation (v5.0)

The Correlates of War (COW) National Material Capabilities dataset has multiple versions. Version 5.0, used here, extends coverage through 2012 with improved documentation. Version 6.0 continues to 2016 with more extended data.

Key variables are:

- **milex**: Military expenditure
- **milper**: Military personnel
- **irst**: Iron and steel production
- **pec**: Primary energy consumption
- **tpop**: Total population
- **upop**: Urban population
- **cinc**: Composite Index of National Capability

Sources include the UN, World Bank, and SIPRI. Limitations involve missing data, reporting biases, and aggregation challenges, which complicate causal inference.

III. ANALYSIS

A. Q3: Top 10 Countries in 2012

According to CINC's 2012 report, China, the USA, India, Russia, Japan, Brazil, South Korea, Germany, Iran, and the UK were the 10 most powerful nations. Figure 1 shows the line graph depicting the change in power over time. China and India's power rose quickly from the late 20th century, the USA's power peaked in the mid-20th century and slowly declined thereafter, Russia declined drastically after 1991 and is slowly recovering, while the rest of Europe, lost power in the long run.
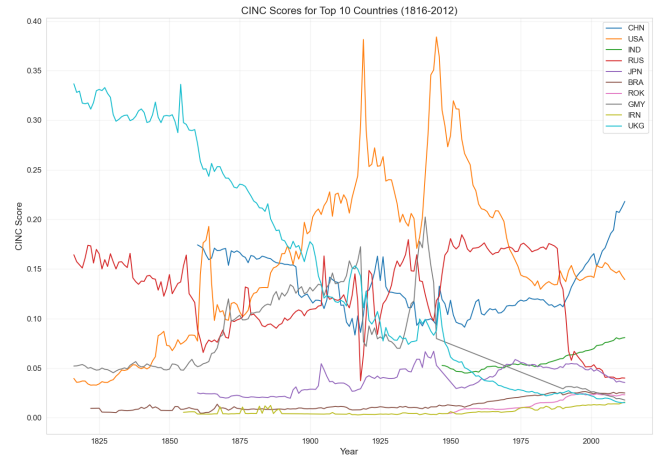

Fig. 1. Top 10 most capable countries in 2012 and their historical trajectories.

B. Q4: Ridgeline Plot of Distributions

The ridgeline graph displaying CINC scores (1816–2012) for the top ten powers indicates abnormality for most graphs due to historical developments. The United States and China show wide, flat patterns which indicate long periods of growth, whereas flat patterns of the CINC scores peaked due to Soviet dominance with the subsequent decline after 1991. In contrast, India, South Korea, Japan, and Brazil show more concentrated peaks. In summary, smaller or emerging states display more symmetric patterns, while the great powers exhibit abnormal patterns due to shaped geopolitical disruptions.

C. Q5: Correlation Analysis

The 2012 correlation matrix shows strong structural relationships among variables with implications for modeling.

Variables Correlated by Nature.

- Energy consumption and CINC ($r = 0.988$) – energy is a core CINC component.
- Total and urban population ($r = 0.937$) – urban population is a subset of total population.
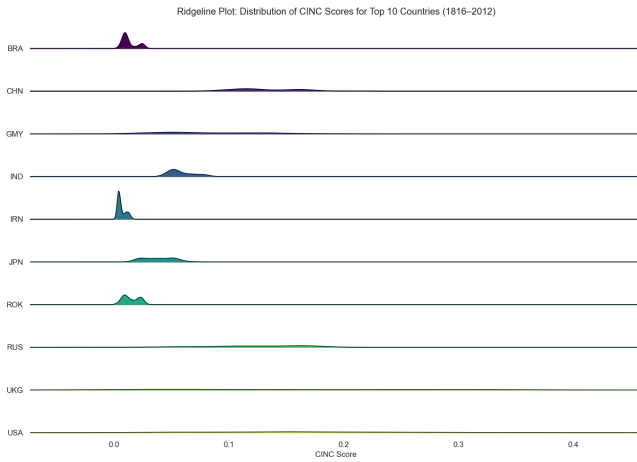- Energy consumption and industrial production ($r = 0.902$) – industry requires energy.

Fig. 2. Ridgeline plot of CINC distributions for top 10 countries.

- Military personnel and population ($r > 0.8$) – larger populations field larger militaries.

Effects of Positive Correlations. Correlations above 0.9 create multicollinearity, inflating errors and destabilizing regression coefficients. For example, including both energy consumption and CINC would obscure each variable's contribution. In practice, one of each highly correlated pair must be excluded or reduced through dimensionality techniques.

Effects of Negative Correlations. No negative relationships appear; all variables are positively correlated ($r = 0.28$–0.99). This is expected since they all measure aspects of national power. Their complementary nature simplifies interpretation because all move in the same direction when predicting outcomes.
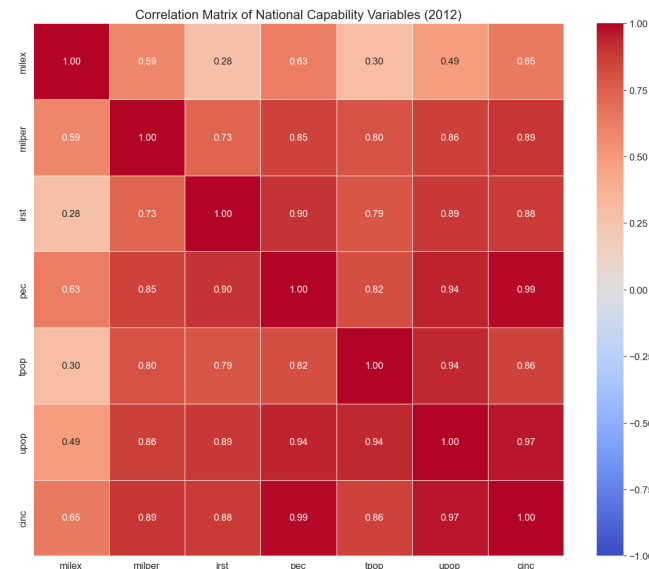

Fig. 3. Correlation heatmap of NMC variables for 2012.

D. Q6: Most Similar Systems Design

After undertaking log transformation and standardization, the most similar countries by pairwise calculation came out to be the microstates and small island nations. European microstates (Monaco, Liechtenstein, San Marino) formed the closest cluster, whereas Caribbean islands (Grenada, St. Vincent, St. Kitts-Nevis) and Pacific islands (Tuvalu, Nauru, Kiribati, Tonga, Samoa) also clustered together with Samoa. The most similar pairs did not include any Great Powers. The log transformation captured variability well for the smaller states and microstates and islands being the most similar in material capabilities was a consistent finding for both Manhattan and Euclidean measures.

E. Q7: Global Mapping

The 2012 CINC analysis highlights sharp regional disparities. As of 2012, China controls East Asia, India has greater dominance than Pakistan in South Asia, Russia is in charge of the ex-Soviet states, and the U.S. is in charge of North America. Western Europe is quite balanced, while the Middle East and Southeast Asia are viewed as competitive multipolar markets. South China Sea, the Russian western periphery, the Korean Peninsula, and the Middle East are all conflict prone. These conflicts have been mirrored by Russia's invasion of Ukraine, the tensions with China, the instability in the Middle East, and the Korean tensions since 2012. While material resources explain much, alliances, interdependence, and nuclear deterrence also shape outcomes.
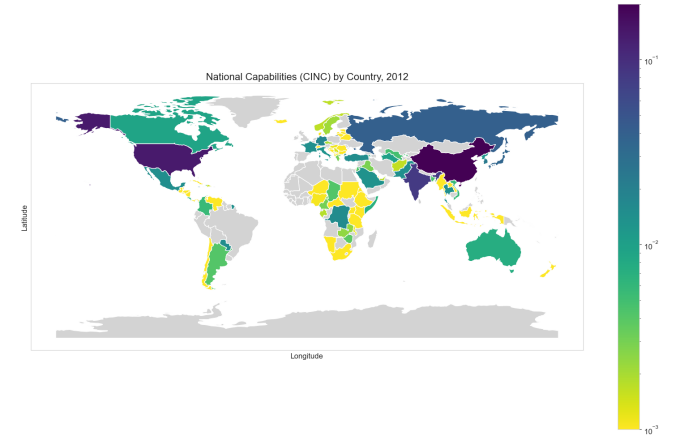

Fig. 4. Choropleth world map of CINC values (2012).

IV. CONCLUSION

The NMC dataset reveals concentrated material capabilities among a few powers, with China's rise, India's growth, Russia's decline, and the USA's relative decline justifying long-term tendencies. Correlation analysis detects multicollinearity, while similarity measures detect clustering among microstates rather than great powers. Useful for quantitative comparison, the CINC measure ignores qualitative factors such as technology and alliances, and causal inference has to be applied with caution.