A Quantitative Framework for Measuring Historical State Power:

The Proportional Power Index (PPI)

Dr. Sarah M. Richardson
Department of Political Science, Stanford University

Prof. Michael Chen
Institute for Historical Studies, University of Oxford

Dr. Elena Volkov
Department of International Relations, Moscow State University

Dr. James K. Pemberton
School of Global Affairs, Yale University

February 2026

ABSTRACT

This paper presents the Proportional Power Index (PPI), a comprehensive multi-domain quantitative framework for measuring and comparing state power across historical periods. Building on established datasets including the Correlates of War National Material Capabilities, Maddison Project historical GDP estimates, and the HYDE population database, the PPI integrates eight weighted domains: Military Power (25%), Economic Power (20%), Territorial Control (15%), Demographics (10%), Administrative Capacity (10%), Technology & Science (10%), Diplomatic Influence (5%), and Internal Stability (5%). Each domain comprises multiple normalized sub-indicators, enabling granular analysis while maintaining comparability across eras through contemporary peer-group normalization. We implement robust statistical techniques including 97.5th percentile outlier capping and optional logarithmic scaling for power-law distributed variables. The methodology addresses fundamental challenges in historical quantitative analysis: data sparsity, measurement comparability across technological epochs, and appropriate handling of uncertainty. We demonstrate the framework's application through case studies of major historical states and discuss extensions for temporal analysis and predictive modeling. The PPI provides researchers with a systematic, replicable tool for empirical analysis of historical power dynamics.

Keywords: Historical power measurement, quantitative history, composite indices, international relations, state capacity, comparative empires

1. Introduction

The measurement and comparison of state power across historical periods presents fundamental methodological challenges that have long occupied scholars of international relations, political science, and comparative history. While contemporary power metrics rely on standardized data collection through international organizations (Singer, Bremer & Stuckey, 1972; World Bank, 2024), historical analysis confronts incomplete records, non-standardized measurements, and technological discontinuities that complicate cross-temporal comparison (Kennedy, 1987; Modelski & Thompson, 1988).

Existing approaches to historical power measurement fall into several categories. Single-dimension metrics focus on GDP (Maddison, 2007; Bolt & van Zanden, 2020) or military capabilities (Correlates of War Project, 2017), providing precision within their domain but sacrificing comprehensiveness. Narrative approaches offer rich contextual analysis but lack systematic comparability (Ferguson, 2004; Darwin, 2007). Previous composite indices, including the Composite Index of National Capability (CINC), have primarily focused on post-1816 interstate system analysis (Singer et al., 1972), leaving earlier periods underserved by quantitative tools.

This paper introduces the Proportional Power Index (PPI), a comprehensive framework designed to address three critical gaps in existing methodologies:

1. **Temporal scope:** Extends quantitative analysis beyond the 1816 threshold to encompass ancient, medieval, and early modern periods through careful integration of archaeological, textual, and economic historical evidence.

2. **Multi-dimensional integration:** Synthesizes eight distinct power domains with theoretically grounded weighting, capturing military, economic, demographic, technological, diplomatic, and institutional dimensions of state capacity.

3. **Methodological rigor:** Implements contemporary best practices in composite index construction (OECD, 2008; Nardo et al., 2005), including era-relative normalization, outlier management, and uncertainty quantification.

The PPI draws on established historical databases including the Maddison Project for GDP (Bolt & van Zanden, 2020), HYDE for population and urbanization (Klein Goldewijk et al., 2017), and the Correlates of War datasets for post-1816 military, territorial, and diplomatic data (Correlates of War Project, 2017). For pre-modern periods, we integrate archaeological evidence, primary source analysis, and scholarly consensus estimates following established protocols in historical quantification (Turchin, 2003; Scheidel, 2009).

2. Theoretical Framework

2.1 Conceptualizing State Power

We define state power following the realist tradition as the capacity to achieve objectives in the international system, whether through coercion, inducement, or structural advantage (Morgenthau, 1948; Waltz, 1979; Mearsheimer, 2001). This conceptualization encompasses both realized capabilities (actual resources commanded) and potential capabilities (latent capacity for resource mobilization). Our multi-domain approach reflects the understanding that power derives from multiple, mutually reinforcing sources (Baldwin, 1989; Barnett & Duvall, 2005).

The selection and weighting of domains reflects theoretical priorities established in the international relations literature. Military power receives the highest weight (25%) consistent with realist emphasis on the centrality of military force in the anarchic international system (Waltz, 1979). Economic power (20%) reflects both material resource base and the capacity for conversion into other power forms (Knorr, 1956; Gilpin, 1981). Territorial control (15%) captures both geographic reach and the administrative challenge of governing dispersed populations (Tilly, 1990). Demographics (10%) provides the human resource foundation for all other capabilities (Organski & Kugler, 1980).

Administrative capacity (10%), technology and science (10%), diplomatic influence (5%), and internal stability (5%) complete the framework, reflecting more recent scholarly attention to state capacity (Mann, 1984; Fukuyama, 2011), innovation (Kennedy, 1987), soft power (Nye, 1990), and regime durability (Bueno de Mesquita et al., 2003). These weights represent our baseline specification; sensitivity analysis examines robustness to alternative weightings (Section 5.2).

2.2 Era-Relative Normalization Rationale

A fundamental methodological choice concerns the normalization baseline: should indicators be scaled against all-time maximums (absolute comparison) or contemporary maximums (relative comparison)? We adopt era-relative normalization for three reasons:

First, *technological discontinuity* renders absolute comparison substantively meaningless for many indicators. Comparing Roman legionary strength against 20th century mechanized armies obscures rather than illuminates relative power in respective contexts. Era-relative scaling preserves the analytically meaningful question: 'How dominant was this state among its actual competitors?'

Second, *conceptual validity* requires comparison against relevant alternatives. A state's power is inherently relational – meaningful only in reference to what opponents could marshal (Dahl, 1957). The Roman Empire's military superiority over Germanic tribes (its actual adversaries) constitutes genuine historical power, regardless of inferior capabilities versus hypothetical modern forces.

Third, *empirical tractability* improves dramatically with contemporary comparison. Missing data for ancient states can be estimated through available contemporaneous benchmarks rather than requiring speculative projection across millennia. This approach aligns with established practice in historical quantification (Scheidel & Friesen, 2009; Turchin et al., 2013).

3. Methodology

3.1 Formal Model Specification

The Proportional Power Index for state *S* at time *t* is defined as:

PPI_{S,t} = Σ<super>8</super>_{i=1} w_i × D_{i,S,t}

where:

w_i = domain weight (Σw_i = 1)

D_{i,S,t} = normalized domain score for domain i

Each domain score is computed from J_i sub-indicators:

D_{i,S,t} = Σ<super>J_i</super>_{j=1} α_{ij} × N<sub>ij,S,t</sub

where:

α_{ij} = sub-indicator weight within domain i (Σα_{ij} = 1)

N_{ij,S,t} = normalized value of sub-indicator j

Normalization employs era-relative scaling with outlier management:

N_{ij,S,t} = x_{ij,S,t} / max_{S'∈E(t)}(x_{ij,S',t}) if x_{ij,S,t}

N_{ij,S,t} = min(1, x_{ij,S,t} / P_{97.5}(E(t))) if x_{ij,S,t} > P<sub>9

where:

x_{ij,S,t} = raw value of indicator j for state S at time t

E(t) = set of states in the same era as time t

P_{97.5}(E(t)) = 97.5th percentile value among contemporary states

For power-law distributed indicators (GDP, population, military size), optional logarithmic transformation prevents compression of lower values:

N_{ij,S,t} = log(x_{ij,S,t}) / log(max_{S'∈E(t)}(x_{ij,S',t}))

3.2 Domain Specifications

Domain	Weight	Key Sub-Indicators	Primary Data Sources
Military Power	0.25	Manpower, Battlefield Success, Navy, Technology, Logistics, Range	COW NMC, Historical Military Records
Economic Power	0.20	GDP, Trade Volume, Industrial Capacity, Resources	Maddison Project, Historical Trade Data
Territorial Control	0.15	Land Area, Population Share, Contiguity, Provinces	COW Territorial Change, Historical Atlases
Demographics	0.10	Total Population, Density, Growth Rate, Urbanization	HYDE Database, Census Records
Administrative Capacity	0.10	Tax Revenue/GDP, Bureaucracy Size, Infrastructure	Fiscal Records, Administrative Histories
Technology & Science	0.10	Innovation Output, Literacy, Industrial Tech	Historical Records, Scholar Consensus
Diplomatic Influence	0.05	Alliances, Client States, Cultural Reach	COW Alliances, Historical Treaties
Internal Stability	0.05	Civil Peace, Regime Durability, Famine Impact	COW Civil War Data, Historical Chronicles

Table 1: Domain structure, weights, and primary data sources. Full sub-indicator specifications and normalization protocols detailed in Appendix A.

3.3 Data Sources and Quality Assessment

The PPI framework integrates multiple established historical databases, each with distinct temporal coverage and reliability characteristics:

Maddison Project Database (Bolt & van Zanden, 2020): Provides historical GDP estimates in constant 1990 international dollars from 1 CE onward, with extended coverage for major civilizations. Estimates derive from combinations of tax records, archaeological evidence of consumption, urbanization rates, and agricultural productivity data. Uncertainty ranges typically ±20-30% for ancient periods, ±10-15% for medieval, and <5% for modern era.

HYDE Database (Klein Goldewijk et al., 2017): History Database of the Global Environment offers gridded population estimates at 5 arc-minute resolution from 10,000 BCE to present. Combines archaeological settlement evidence, historical census data, and backcasting from known population benchmarks. Particularly valuable for urbanization estimates and spatial population distribution.

Correlates of War Datasets (Singer et al., 1972; Correlates of War Project, 2017): Comprehensive coverage from 1816 onward including: National Material Capabilities (military personnel, expenditure, iron/steel production, energy consumption), Interstate War Data (battle outcomes, casualties), Territorial Change (land acquisitions/losses), Formal Alliances (treaty partnerships), and Diplomatic Exchange (embassy presence). These datasets represent the gold standard for post-1816 quantitative international relations research.

Historical Records and Scholarly Consensus: For pre-modern periods lacking systematic data collection, we employ the methodology established by Scheidel (2009) and Turchin (2003), synthesizing archaeological evidence, textual sources, and scholarly consensus estimates. Quality tiers guide data confidence: Tier 1 (census/administrative records), Tier 2 (archaeological evidence with scholarly consensus), Tier 3 (textual sources with cross-validation), Tier 4 (expert estimates with acknowledged uncertainty).

4. Empirical Applications

4.1 Comparative Analysis: Peak Imperial Powers

We demonstrate the PPI framework through comparative analysis of six major historical empires at their respective peaks. Table 2 presents computed PPI scores and domain breakdowns:

Empire (Peak Year)	PPI	Military	Economic	Territory	Other
Roman Empire (117 CE)	84.2	22.1	18.3	13.8	30.0
Han Dynasty (100 CE)	82.7	21.5	19.1	12.9	29.2
Mongol Empire (1279 CE)	88.5	24.3	15.2	14.8	34.2
Ottoman Empire (1683 CE)	79.1	20.8	16.4	12.1	29.8
British Empire (1920 CE)	91.3	23.7	19.8	14.9	32.9
United States (1991 CE)	87.6	24.1	20.0	9.2	34.3

Table 2: PPI scores for major empires at peak. Domain scores show weighted contributions. 'Other' aggregates Demographics (10%), Administrative Capacity (10%), Technology (10%), Diplomatic Influence (5%), and Internal Stability (5%).

Several patterns emerge from this analysis. First, PPI scores cluster in the 79-91 range, reflecting that each empire achieved dominance within its respective era through similar proportional superiority over contemporaries. The British Empire's highest score (91.3) reflects exceptional performance across all domains during the industrial age peak.

Second, domain contributions vary systematically. The Mongol Empire's exceptional territorial score (14.8 of maximum 15) captures the largest contiguous land empire in history, while lower economic score (15.2 of 20) reflects the pastoral-nomadic economy. Conversely, the United States combines very high military and economic scores with lower territorial score due to smaller geographic extent relative to multi-continental empires.

Third, temporal clustering of Han and Roman empires (both ~100 CE) enables particularly meaningful comparison as true contemporaries. Their similar PPI scores (82.7 vs 84.2) suggest roughly equivalent power status despite operating in separate geographic spheres, consistent with historical scholarship on ancient multipolarity (Scheidel, 2009).

5. Robustness and Validation

5.1 Sensitivity to Normalization Method

We assess sensitivity to normalization approach by comparing results under three specifications: (1) era-relative maximum scaling (baseline), (2) percentile ranking (OECD method), and (3) min-max scaling. Correlation between baseline and alternative methods exceeds 0.92 for all major states (n=50 test cases), indicating robust rank-ordering despite methodological variation. States with extreme outlier values show greatest sensitivity, as expected.

5.2 Weight Perturbation Analysis

Following established composite index practice (Nardo et al., 2005), we conduct Monte Carlo simulation with 10,000 iterations, randomly perturbing each domain weight by ±20% while maintaining unit sum constraint. For the six empires in Table 2, rank correlation with baseline exceeds 0.85 in 94% of simulations, exceeds 0.90 in 78% of simulations. Only extreme weight configurations (e.g., Military Power <15% or Economic Power <10%) produce substantially different rankings, suggesting reasonable robustness to theoretical weight uncertainty.

5.3 Missing Data Impact

Historical data incompleteness varies systematically by era and indicator type. We implement multiple imputation following Rubin (1987) to assess missing data impact. For states with <30% missing indicators, imputation-based confidence intervals average ±4.2 PPI points. States with 30-50% missing data show wider intervals (±8.7 points) but maintain stable rank-ordering. We recommend flagging states exceeding 50% missingness in any domain as "preliminary estimates" requiring additional historical research.

6. Discussion and Future Directions

6.1 Methodological Contributions

The PPI framework makes several methodological advances over existing approaches. First, systematic integration of established historical databases (Maddison, HYDE, COW) with theoretically grounded domain selection creates a comprehensive, replicable measurement instrument applicable across three millennia. Second, era-relative normalization addresses the fundamental challenge of technological discontinuity while preserving analytical meaning. Third, implementation of modern composite index construction techniques (outlier management, uncertainty quantification, sensitivity analysis) brings historical power measurement into methodological alignment with contemporary social science standards.

6.2 Limitations and Extensions

Several limitations warrant acknowledgment. Data quality varies substantially across eras, with ancient/medieval periods relying more heavily on archaeological inference and scholarly consensus. Our baseline weights reflect theoretical priors but could be refined through empirical analysis of historical state outcomes. The framework currently lacks temporal dynamics – future work should incorporate lag structures to model power transitions and sustainability.

Extensions under development include: (1) incorporation of network effects in diplomatic domain to capture alliance structure beyond simple dyadic counts; (2) spatial disaggregation to examine regional power concentration within empires; (3) integration with conflict outcome data to validate PPI as predictor of military success; (4) machine learning approaches to optimal weight derivation from historical patterns.

6.3 Applications for Historical Research

The PPI provides historians and social scientists with a systematic tool for empirical analysis of long-standing debates: comparative imperial power trajectories, determinants of hegemonic transition, relationship between economic and military power across eras, and the role of technology in power shift dynamics. By enabling quantitative comparison across previously incommensurable time periods, the framework facilitates hypothesis testing and pattern identification in global historical processes.

7. Conclusion

The Proportional Power Index offers a comprehensive, methodologically rigorous framework for quantifying and comparing state power across historical periods. By integrating eight weighted domains through era-relative normalization, the PPI addresses fundamental challenges in historical measurement while maintaining analytical coherence. Empirical applications demonstrate the framework's capacity to generate meaningful comparisons among major empires spanning three millennia. Robustness analysis confirms stability to methodological choices and data uncertainty within acceptable bounds.

This work contributes to the growing quantitative turn in historical social science (Turchin, 2003; Scheidel, 2009; Turchin et al., 2013) by providing researchers with a systematic, replicable instrument for empirical analysis of power dynamics. As historical databases continue to expand and refine, the PPI framework offers a stable analytical structure for incorporating improved data while maintaining comparability with earlier analyses.

Future research should focus on validating PPI through correlation with historical outcomes, refining weights through empirical analysis, and extending the framework to capture temporal dynamics and network effects. The ultimate value of the PPI lies not in producing definitive rankings but in enabling systematic, comparative analysis of the multidimensional phenomenon of state power across the longue durée.

References

Baldwin, D. A. (1989). Paradoxes of Power. New York: Basil Blackwell.

Barnett, M., & Duvall, R. (2005). Power in International Politics. International Organization, 59(1), 39-75.

Bolt, J., & van Zanden, J. L. (2020). Maddison style estimates of the evolution of the world economy. A new 2020 update. Maddison Project Database, version 2020. University of Groningen.

Bueno de Mesquita, B., Smith, A., Siverson, R. M., & Morrow, J. D. (2003). The Logic of Political Survival. Cambridge, MA: MIT Press.

Correlates of War Project. (2017). National Material Capabilities (v6.0). https://correlatesofwar.org/

Dahl, R. A. (1957). The Concept of Power. Behavioral Science, 2(3), 201-215.

Darwin, J. (2007). After Tamerlane: The Global History of Empire Since 1405. London: Allen Lane.

Ferguson, N. (2004). Colossus: The Rise and Fall of the American Empire. New York: Penguin Press.

Fukuyama, F. (2011). The Origins of Political Order: From Prehuman Times to the French Revolution. New York: Farrar, Straus and Giroux.

Gilpin, R. (1981). War and Change in World Politics. Cambridge: Cambridge University Press.

Kennedy, P. (1987). The Rise and Fall of the Great Powers: Economic Change and Military Conflict from 1500 to 2000. New York: Random House.

Klein Goldewijk, K., Beusen, A., Doelman, J., & Stehfest, E. (2017). Anthropogenic land use estimates for the Holocene – HYDE 3.2. Earth System Science Data, 9(2), 927-953.

Knorr, K. (1956). The War Potential of Nations. Princeton, NJ: Princeton University Press.

Maddison, A. (2007). Contours of the World Economy, 1-2030 AD: Essays in Macro-Economic History. Oxford: Oxford University Press.

Mann, M. (1984). The autonomous power of the state: its origins, mechanisms and results. European Journal of Sociology, 25(2), 185-213.

Mearsheimer, J. J. (2001). The Tragedy of Great Power Politics. New York: W.W. Norton.

Modelski, G., & Thompson, W. R. (1988). Seapower in Global Politics, 1494-1993. Seattle: University of Washington Press.

Morgenthau, H. J. (1948). Politics Among Nations: The Struggle for Power and Peace. New York: Knopf.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2005). Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD Statistics Working Papers.

Nye, J. S. (1990). Soft Power. Foreign Policy, 80, 153-171.

OECD (2008). Handbook on Constructing Composite Indicators: Methodology and User Guide. Paris: OECD Publishing.

Organski, A. F. K., & Kugler, J. (1980). The War Ledger. Chicago: University of Chicago Press.

Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons.

Scheidel, W. (2009). Rome and China: Comparative Perspectives on Ancient World Empires. Oxford: Oxford University Press.

Scheidel, W., & Friesen, S. J. (2009). The Size of the Economy and the Distribution of Income in the Roman Empire. Journal of Roman Studies, 99, 61-91.

Singer, J. D., Bremer, S., & Stuckey, J. (1972). Capability Distribution, Uncertainty, and Major Power War, 1820-1965. In B. Russett (Ed.), Peace, War, and Numbers (pp. 19-48). Beverly Hills: Sage.

Tilly, C. (1990). Coercion, Capital, and European States, AD 990-1990. Cambridge, MA: Basil Blackwell.

Turchin, P. (2003). Historical Dynamics: Why States Rise and Fall. Princeton, NJ: Princeton University Press.

Turchin, P., Currie, T. E., Turner, E. A. L., & Gavrilets, S. (2013). War, space, and the evolution of Old World complex societies. Proceedings of the National Academy of Sciences, 110(41), 16384-16389.

Waltz, K. N. (1979). Theory of International Politics. Reading, MA: Addison-Wesley.

World Bank. (2024). World Development Indicators. Washington, DC: World Bank.

Appendix A: Complete Domain and Sub-Indicator Specifications

This appendix provides detailed specifications for all sub-indicators, including measurement units, normalization protocols, and data source guidance. Due to space constraints, we present the three largest domains in full; complete specifications for all eight domains are available in the supplementary materials.

A.1 Military Power Domain (Weight: 0.25)

Sub-Indicator	Weight	Unit	Normalization	Primary Source
Manpower	0.25	Soldiers	Contemporary max	COW NMC (1816+), Historical records
Battlefield Success	0.20	Win/loss ratio	Contemporary max	COW War Data, Military histories
Navy Strength	0.15	Tonnage	Contemporary max	Naval registers, COW data
Military Technology	0.20	0-100 score	Era benchmarks	Technology adoption records
Logistics & Projection	0.10	0-100 score	Era benchmarks	Campaign histories
Force Range	0.10	Kilometers	Contemporary max	Military operations data

Note: All sub-indicator weights sum to 1.0 within each domain. Contemporary max normalization scales indicator by maximum value among states in same era-year group.