

Comparative Analysis of Intervention Strategies for Mass Atrocity Prevention Using Agent-Based Modeling

Michael Magid¹, Zeynep Ertem, Ph.D.¹ David Cingranelli, Ph.D.¹

¹Binghamton University, USA

correspondence: mmagid1@binghamton.edu

Abstract

We present a two-stage, data-driven framework for evaluating whether peacekeeping, sanctions, and humanitarian aid mitigate conflict-related harms across heterogeneous governance contexts. A first-stage statistical model with two-way fixed effects and ridge regularization learns persistence, spatial spillovers, and common shocks from CIRIGHTS-, ACLED-, and UCDP-style panels. These estimates ($\beta_{\text{self}}, \beta_{\text{spat}}, \{\gamma_t\}$) anchor a second-stage agent-based simulator that embeds countries in multilayer networks with an intervention-effects estimator and dyadic territorial risk. The design is predictive rather than causal but enables comparative, scenario-based analysis.

The model achieves moderate performance ($R^2 \approx 0.21$), with temporal anchoring and nontrivial spillovers. Diagnostics (calibration, QQ, heteroskedasticity, Moran's I) and counterfactual probes ($\beta_{\text{spat}}=0$; W -sensitivity) confirm robustness. Network analyses show heavy-tailed degrees and modest coalition modularity, while shock-response experiments treat atrocities as shocks and interventions as calibrated counter-shocks. Complexity metrics capture regime dynamics and identify controllability windows, supporting context-sensitive assessment of intervention portfolios.

Keywords: Mass Atrocity Prevention; Agent-Based Modeling; Spatial Diffusion; Intervention Portfolios; Comparative Policy Analysis

Introduction

Reducing conflict-related harms remains difficult despite advances in early warning and response (Appe et al., 2023; Campbell, 2023). Outcomes vary with regime type, civil society capacity, and administrative structures (Golder and Williams, 2006; Rubaii et al., 2021b; Krain, 2006). Decentralized contexts can enable faster, locally calibrated responses, while centralized systems may mobilize resources but face bureaucratic delays (Vining and Weimer, 1999). Yet many interventions remain uniform and top-down, overlooking institutional and civic variation. Comparative frameworks stress tailoring peacekeeping, sanctions, and aid to governance capacity, NGO density, and socio-legal norms (Breidahl and Brodtkin, 2023; Miaz, 2024).

This paper develops an agent-based model (ABM) that simulates state, non-state, and international actors using CIRIGHTS, ACLED, UCDP, and macro indicators (Cingranelli and Richards, 2010; Raleigh et al., 2010; Sundberg and Melander, 2013; World Bank, 2023; International Monetary Fund, 2023). By varying governance, networks, and geography, we assess how similar policies diverge across contexts (Anderton and Brauer, 2019; Fortna, 2004). The contribution is a predictive, comparative platform that situates intervention effects within heterogeneous institutional settings.

Literature Review

Comparative Dimensions in Policy Effectiveness

Governance form and civil society infrastructure mediate whether interventions reduce violence (Buhaug and Gleditsch, 2008; Weidmann and Ward, 2010). Decentralization can accelerate mobilization of CSOs (Rubaii et al., 2021b), while centralized systems may coordinate but slow adaptation (Sichling, 2024a; Golder and Williams, 2006). Representation and administrative equity affect compliance and grievances, shaping intervention headroom (Anderton and Brauer, 2019).

Local-Systemic Interactions

Local dynamics—resource competition, group relations, institutional trust—interact with systemic factors such as international norms and cross-border linkages (Chopra and Naidu, 2023; Rubaii et al., 2021a; Miaz, 2024). International organizations partnering with community-based groups can improve early warning and program legitimacy when civic space is open (Appe et al., 2023), while restrictive environments limit NGO impact. Regulatory capacity matters: constitutional commitments without enforcement mechanisms underperform, while strong CSO and nonprofit networks can offset weak formal institutions (Campbell, 2023).

Agent-Based Modeling in Comparative Conflict Policy

ABMs allow policy comparison by simulating heterogeneous agents and contexts (Cioffi-Revilla and Rouleau, 2009; von Briesen, 2020). Contextualized ABMs extend evaluation of peacekeeping, sanctions, and aid, showing conditional effectiveness (Fortna, 2004; Hultman et al., 2016). Our model integrates spatial diffusion, multilayer networks, and intervention learning, calibrated to conflict and rights data (Raleigh et al., 2010; Sundberg and Melander, 2013; Cingranelli and Richards, 2010) plus macro indicators (World Bank, 2023; International Monetary Fund, 2023). Systematic parameter variation generates counterfactuals clarifying divergent outcomes (Buhaug and Gleditsch, 2008; Weidmann and Ward, 2010).

Implications for Policy and Practice

Comparative evidence cautions against uniform prescriptions (Campbell, 2023). Peacekeeping is more effective where oversight institutions are strong, while humanitarian aid performs best where nonprofits retain autonomy (Fortna, 2004; Hultman et al., 2016; Sichling, 2024b). ABM supports diagnostic assessment of capacity, civic space, and legal context, aligning interventions with institutional and societal conditions (Rabbi, 2024; van der Zwet et al., 2022). Cross-context simulations thus help anticipate risks of escalation and opportunities for effective civilian protection.

Methods

Conceptual Overview

Figure 1 summarizes the two-stage architecture. Stage 1 learns temporal persistence, spatial spillovers, and common shocks from panel data; Stage 2 reuses these parameters as a small-gain anchor inside the ABM, where agents interact over multilayer networks and respond to shocks and counter-shocks.

Variable Table

This agent-based model (ABM) integrates spatial diffusion, dyadic territorial risk, multilayer networks, and data-driven intervention effects. Countries are agents. Conflict intensity and territorial attributes are drawn from UCDP-style panels, human-rights from CIRIGHTS-style indicators, and intervention activity from PRIF-like sources. Geographic contiguity and codes follow the COW project Stinnett et al. (2002). All coefficients are learned from data; no hand-tuned constants are used. Table 1 lists the variables.

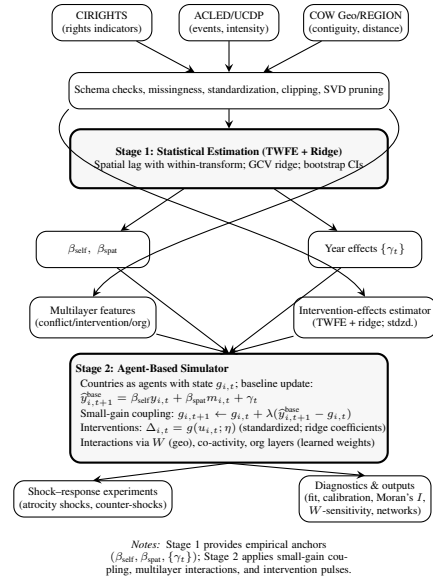


Figure 1: Two-stage architecture: Stage 1 learns temporal persistence, spatial spillovers, and common shocks from panel data; Stage 2 reuses these parameters as a small-gain anchor inside the ABM.

Table 1: Variables Used in the Agent-Based Model

Variable Name	Variable	Source/Computation	Purpose
Country Identification and Geography			
Country ID	country, cow	Country name or COW code Stinnett et al. (2002)	Agent identifier
Region	REGION	Dataset region codes	Regional summaries/fix effects
Contiguity	A_{ij}	COW Direct Contiguity Stinnett et al. (2002)	Spatial adjacency (binary)
Distance (km)	d_{ij}	COW centroids (haversine)	Inverse-distance spatial weights
Conflict Metrics			
Intensity Level	$y_{i,t}$ = intensity_level	UCDP-style panel	Outcome to model/predict
Fatalities (if present)	$f_{i,t}$	UCDP/ACLED-style	Validation/outliers
Intervention Variables (PRIF-like)			
Mediation	MEDIATE6	PRIF-like	Diplomatic engagement
Accords	ACCORD6	PRIF-like	Peace agreement activity
Humanitarian Aid	HUMANAID	PRIF/OECD-like	Assistance intensity
UN Security Council	UNSC	PRIF-like	UNSC actions/attention
Regional Orgs	REGIOORG	PRIF-like	Regional IO engagement
Ground Force	GROUNDFO	PRIF-like	Military (ground)
Air Force	AIRFORCE	PRIF-like	Military (air)
Enforcement	ENFORCE	PRIF-like	Composite enforcement
Human Rights (CIRIGHTS-like)			
Physical Integrity	physint_sum	CIRIGHTS-style	Repression/abuse
Empowerment	civpol/empinx	CIRIGHTS-style	Civil/political liberties
Multilayer Network Features (per year)			
Conflict Degree	conf_deg	From conflict dyads	Centrality (conflict layer)
Intervention Degree	intv_deg	Co-participation	Centrality (intervention layer)
Org Degree	org_deg	UNSC/REGIOORG activity	Centrality (org layer)
Betweenness/Clustering	$\{\cdot\}_{bwn/\{ \cdot \}}_{clu}$	Per layer	Brokerage/local cohesion
Computed Model Variables			
Self Lag	self_lag = $y_{i,t}$	From panel	Agent's own past intensity
Neighbor Mean	neigh_mean = $\sum_j W_{ij} y_{j,t}$	Spatial weights	W-weighted neighbor signal
Predicted	$\hat{y}_{i,t+1}$	Diffusion + $\beta(w; \eta)$	One-step forecast
Residual	$y_{i,t} - \hat{y}_{i,t}$	From panel/model	Error diagnostics
Spatial Coefs	β_{int}, β_{geo}	Ridge + bootstrap	Diffusion parameters w/ CIs
Year Effects	γ_t	Estimated (see below)	Year intercepts for calibration
Territorial Risk	p_{ij}	Calibrated logistic	Dyadic territorial risk

Geographic Weights and Spatial Diffusion

Define raw bilateral weights

$$w_{ij}^* = \begin{cases} d_{ij}^{-1}, & A_{ij} = 1 \text{ and distance observed,} \\ 1, & A_{ij} = 1 \text{ and distance missing,} \\ 0, & \text{otherwise (or } i = j), \end{cases}$$

$$W_{ij} = \frac{w_{ij}^*}{\sum_{k \neq i} w_{ik}^*}.$$

Let $m_{i,t} = \sum_{j \neq i} W_{ij} y_{j,t}$. We estimate a reduced-form spatial lag with two-way fixed effects (country and year) via the

within transformation:

$$\begin{aligned}\tilde{y}_{i,t} &= y_{i,t} - \bar{y}_i - \bar{y}_t + \bar{y}, \\ \tilde{m}_{i,t} &= m_{i,t} - \bar{m}_i - \bar{m}_t + \bar{m}, \\ \tilde{y}_{i,t+1} &= \beta_{\text{self}} \tilde{y}_{i,t} + \beta_{\text{spat}} \tilde{m}_{i,t} + \tilde{\varepsilon}_{i,t+1}.\end{aligned}\quad (1)$$

Estimation uses ridge with generalized cross-validation (GCV). After fitting, we reconstruct year effects on the original scale:

$$\gamma_t = \mathbb{E}[y_{i,t+1} - \beta_{\text{self}} y_{i,t} - \beta_{\text{spat}} m_{i,t} \mid \text{year} = t].$$

Percentile bootstrap CIs are reported for $(\beta_{\text{self}}, \beta_{\text{spat}})$.

Multilayer Networks and Learned Aggregation

We build annual layers: conflict participation (conf), intervention co-participation (intv), and intergovernmental activity (org). For each country-year we compute features $\Phi_{i,t}$ (degree, betweenness, clustering). We learn weights by ridge with GCV:

$$\hat{w} = \arg \min_w \left\{ \sum_{i,t} (y_{i,t+1} - \Phi_{i,t}^\top w)^2 + \lambda \|w\|_2^2 \right\},$$

with finite-row filtering, standardization, empirical clipping, zero-variance pruning, and SVD-based collinearity pruning.

Dyadic Territorial Risk

Let $z_{ij,t} \in \{0,1\}$ indicate territorial dyads when UCDP-style incompatibility equals territory and/or a non-null territory name exists Cederman (1997, 2003). With

$$x_{ij} = [1 \quad d_{ij} \quad \mathbb{I}\{\text{REGION}_i = \text{REGION}_j\}]^\top,$$

we fit a calibrated logistic model

$$p_{ij} = \Pr(z_{ij} = 1 \mid x_{ij}) = \sigma(\theta^\top x_{ij}),$$

with Platt scaling and Brier diagnostics if labels suffice; otherwise we report an empirical base rate.

Intervention Effects Estimation and Application

We estimate a data-driven mapping $g(u_{i,t}; \eta)$ from PRIF-like intervention vectors $u_{i,t}$ to next-year intensity, controlling for self and neighbor signals, using two-way fixed effects and robust ridge (GCV). Let

$$X_{i,t} = [y_{i,t}, m_{i,t}, \text{MEDIATE6}_{i,t}, \text{ACCORD6}_{i,t}, \dots, \text{ENFORCE}_{i,t}].$$

We within-transform $y_{i,t+1}$ and all columns of $X_{i,t}$ by country and year means. After finite filtering and standardization (with stored means/scales), we fit ridge with GCV plus zero-variance and SVD-based collinearity pruning. The learned coefficients (with bootstrap CIs) quantify per-type associations in standardized space. At simulation time, we standardize $u_{i,t}$ using the stored means/scales (for the kept features) and apply

$$\Delta_{i,t} = g(u_{i,t}; \eta) = \text{coef}^\top \text{standardize}(u_{i,t}).$$

Agent Update

Each agent i perceives $s_{i,t} = (y_{i,t}, m_{i,t})$ and updates conflict via the learned spatial coefficients and year effects:

$$\hat{y}_{i,t+1}^{\text{base}} = \beta_{\text{self}} y_{i,t} + \beta_{\text{spat}} m_{i,t} + \gamma_t.$$

Numerical Stability and Validation

Estimation applies finite filtering, standardization with clipping, variance/collinearity pruning, and ridge with GCV under two-way fixed effects; bootstrap CIs quantify uncertainty. Territorial risk is fit via calibrated logistic regression or base-rate fallback. Validation includes schema checks, missingness profiles, and outlier scans. Reproducibility materials (parameter dumps, adjacency matrices, run manifests, regeneration scripts) ensure figures and tables can be exactly reproduced.

Evaluation Metrics

RMSE, MAE, bias, and R^2 measure predictive accuracy; grouped calibration and residual checks assess reliability. Moran's I and W -sensitivity test spatial robustness, while network modularity and centralities describe propagation structures. Complexity measures (recurrence, determinism, entropy, Hurst) act as regime indicators, highlighting stability and controllability.

Metrics

For observed $y_{i,t}$ and predicted $\hat{y}_{i,t}$:

$$\begin{aligned}\text{RMSE} &= \sqrt{\frac{1}{M} \sum (y_{i,t} - \hat{y}_{i,t})^2}, \quad \text{MAE} = \frac{1}{M} \sum |y_{i,t} - \hat{y}_{i,t}|, \\ \text{Bias} &= \frac{1}{M} \sum (y_{i,t} - \hat{y}_{i,t}), \quad R^2 = 1 - \frac{\sum (y_{i,t} - \hat{y}_{i,t})^2}{\sum (y_{i,t} - \bar{y})^2}.\end{aligned}$$

We report global and region-level metrics. For the intervention estimator, we also report control-only vs full-model RMSE/R^2 and improvements (in standardized space).

Empirically Coupled Agent Dynamics

The initial statistical model provides empirically estimated macro-level dynamics (persistence, spatial spillovers, and year shocks). We re-use the learned spatial diffusion parameters $(\beta_{\text{self}}, \beta_{\text{spat}})$ and reconstructed year effects $\{\gamma_t\}$ as a baseline (small-gain) nudge on agents' internal conflict state (see the Diffusion subsection in Methods). At each step, for state i with grievance $g_{i,t}$ and neighbor mean $m_{i,t}$ we compute $\hat{y}_{i,t+1}^{\text{base}} = \beta_{\text{self}} y_{i,t} + \beta_{\text{spat}} m_{i,t} + \gamma_t$, then update

$$g_{i,t+1} \leftarrow g_{i,t} + \lambda (\hat{y}_{i,t+1}^{\text{base}} - g_{i,t}), \quad \lambda \in (0, 1) \text{ (default 0.1)}.$$

This maintains empirically estimated persistence and spillovers while allowing agents' multi-criteria choices (security, sovereignty, resources, reputation), bounded rationality, and learning to drive deviations.

Operational Coalition Network and Communities

Because formal alliance data are not embedded in the current panel, we infer an *operational coalition* network from co-activity:

- Org layer: countries active in REGIOORG in the same year (UNSC removed to avoid global coupling).
- Intervention layer: co-implementation across core columns (GROUNDFO, AIRFORCE, ENFORCE, HUMANAID, MEDIATE6, ACCORD6), expanded via fuzzy-matched terms (peacekeep/mission/deploy/sanction/aid/relief/mediat). Family-specific weights emphasize military>sanctions>PKO>aid>diplomacy>org.
- Conflict penalty: conflict dyads (gwno_a/b) downweight the tie.

Layers are aggregated over a trailing window (default 10 years), normalized to $[0, 1]$, and combined as

$$w_{ij}^{\text{ally}} = \text{clip}_{[0,1]} \left(\alpha_{\text{org}} \tilde{w}_{ij}^{\text{org}} + \alpha_{\text{intv}} \tilde{w}_{ij}^{\text{intv}} - \alpha_{\text{conf}} \tilde{w}_{ij}^{\text{conf}} \right),$$

with defaults $\alpha_{\text{org}}=0.7$, $\alpha_{\text{intv}}=0.5$, $\alpha_{\text{conf}}=0.3$. We retain all positive ties to avoid over-sparsification. We run weighted Louvain (python-louvain) on the alliance graph (isolates retained) and report modularity, number/sizes of communities, and layer-specific partitions (org-only, intv-only) for robustness.

Agent Decision Rules and Interaction Mechanics

Countries are agents with grievance/state $g_{i,t}$, conflict intensity $y_{i,t}$, and action vector $a_{i,t}$ (peacekeeping, sanctions, aid, diplomacy, receptivity). Signals are $s_{i,t} = (y_{i,t}, m_{i,t}, \Phi_{i,t})$, where $m_{i,t} = \sum_j W_{ij} y_{j,t}$ and $\Phi_{i,t}$ aggregates multilayer network features with learned weights.

Baseline updates combine persistence, diffusion, and year shocks:

$$\hat{y}_{i,t+1}^{\text{base}} = \beta_{\text{self}} y_{i,t} + \beta_{\text{spat}} m_{i,t} + \gamma_t, \quad (2)$$

$$g_{i,t+1} \leftarrow g_{i,t} + \lambda (\hat{y}_{i,t+1}^{\text{base}} - g_{i,t}), \quad (3)$$

with intervention deviations:

$$\Delta_{i,t} = g(u_{i,t}; \eta) = \text{coef}^\top \text{standardize}(u_{i,t}), \quad (4)$$

$$y_{i,t+1} = \text{clip}(g_{i,t+1} + \Delta_{i,t}, \text{support}(y)), \quad (5)$$

where $u_{i,t}$ are standardized PRIF-like inputs with ridge-estimated coefficients.

Interactions occur via three topologies: spatial (W), intervention co-activity, and organizational co-participation. Agents respond to $s_{i,t}$ but are nudged toward empirical regularities through λ , preventing free drift. Validation checks reproduce observed distributions, network correlations, and impulse responses consistent with β_{self} and β_{spat} .

Simulation Protocol and Parameters

Simulations proceed annually, initializing y_{i,t_0} from the last observed year. Unless noted, $\lambda = 0.1$ and Stage 1 parameters are retained. Baseline runs assess drift; shock-response experiments target top- K at-risk states, apply a shock $+\delta$, then counter-shocks at empirical percentiles. We record system-level deltas in grievance, contagion, modularity, and complexity, averaged over S seeds.

When COW data are missing, W defaults to REGION adjacency; sensitivity runs replace W with alliance or information graphs. Code releases include parameter dumps, adjacency matrices, and manifests (seed, T , K , δ).

Shock-Response Intervention Experiments

At t_0 we inject shocks (e.g., $+0.2$ grievance) to top- K states; at t_0+1 we apply counter-shocks for each intervention family. Reported deltas cover grievance, cascades, contagion, and modularity relative to baseline.

Diffusion Diagnostics in the Simulator

Diagnostics emphasize two checks: (i) the no-spatial counterfactual ($\beta_{\text{spat}}=0$) and (ii) W sensitivity (REGION vs alliance vs information). Residual Moran's I and γ_t alignment are omitted as they remain constant under REGION fallback.

Results

Global Conflict Dynamics and Fit

Figure 2 shows the global mean conflict intensity over time. The model reproduces broad secular variation at the system level and provides face-valid temporal dynamics across the sample period. Cross-sectional patterns are visible in country-by-year heatmaps of observed and predicted intensity (Figs. 3), while the pointwise fit appears in the observed-predicted scatter (Fig. 4).

Out-of-sample performance at the result level is: RMSE = 0.200, MAE = 0.098, Bias = 0.018, and $R^2 = 0.209$ (global, $N = 2,723$). Region-stratified metrics indicate heterogeneous difficulty across regions (e.g., some regions show higher RMSE and lower R^2), consistent with varying conflict regimes and data coverage.

At the core of the diffusion mechanism, the learned spatial parameters are $\beta_{\text{self}} \approx 0.262$ and $\beta_{\text{spat}} \approx 0.094$ (Fig. 12). These values imply moderate persistence and a measurable but smaller contemporaneous neighbor contribution.

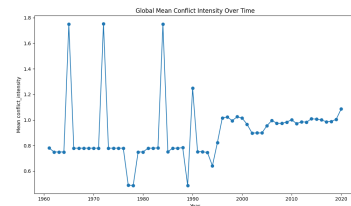


Figure 2: Global mean conflict intensity over time.

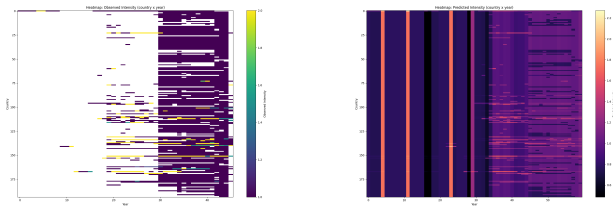


Figure 3: Observed (left) and predicted (right) conflict intensity, country-by-year.

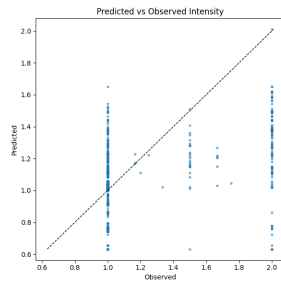


Figure 4: Observed vs. predicted intensity (result-level).

Intervention Effects

Intervention effects are learned in a two-way fixed-effects ridge framework using standardized features (no hard-coded magnitudes). Figure 5 displays standardized coefficients with CIs, summarizing average partial associations after controlling for lagged self and neighbors.

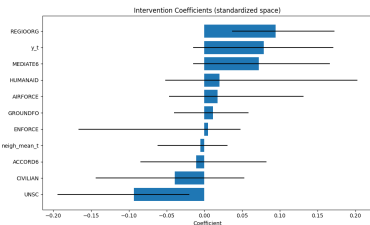


Figure 5: Standardized intervention coefficients with confidence intervals.

Residual Diagnostics and Calibration

Residual structure and calibration were examined comprehensively. Residuals by region (Fig. 6) diagnose geographic variation in errors; QQ plots (Fig. 6) assess normality; and residuals vs. neighbor mean (Fig. 6) check for remaining spatial-lag misspecification. Grouped reliability by region and decade (Fig. 7) indicates calibration quality within strata. Heteroskedasticity bands by predicted decile (Fig. 10) quantify variance behavior as a function of predicted level.

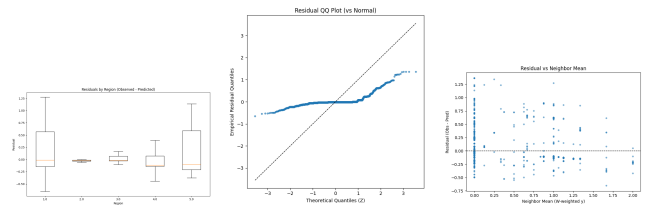


Figure 6: Residual diagnostics: by region (left), QQ (center), vs. neighbor mean (right).

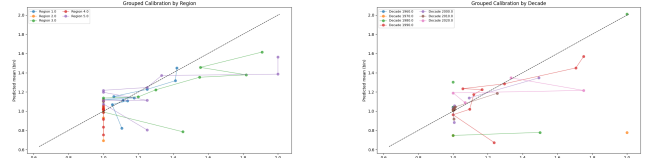


Figure 7: Grouped calibration (reliability): by region (left) and decade (right).

Spatial Dependence and Diffusion Mechanics

Residual spatial autocorrelation over time is summarized by Moran's I (Fig. 8) using REGION-based adjacency (COW fallback). Rolling-window OLS of β_{self} and β_{spat} (Fig. 8) visualizes temporal stability. Reconstructed year effects γ_t with bootstrap CIs (Fig. 9) capture common shocks net of self and spatial lags. A cumulative learning curve (Fig. 9) relates sample length to RMSE. Approximate densities from reported CIs (Fig. 10) provide an uncertainty view for the core spatial parameters.

An impulse response experiment (Fig. 11) simulates a unit shock in a focal country and tracks spillover. Sensitivity to the spatial weights (Fig. 11) recomputes neighbor means under REGION-based W to check RMSE changes. With COW contiguity/centroids, this analysis can be refined to geographic adjacency.

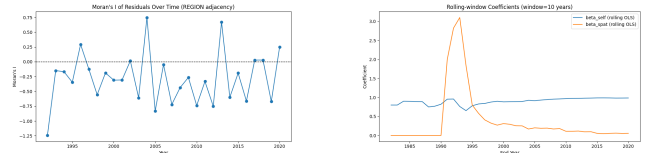


Figure 8: Moran's I (left), rolling-window coefficients (right)

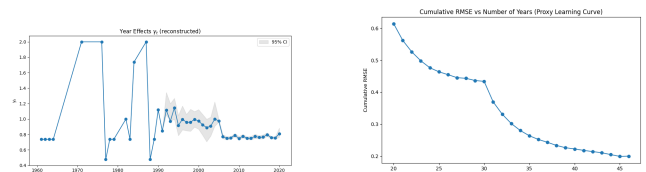


Figure 9: γ_t with CIs (left), and learning curve (right).

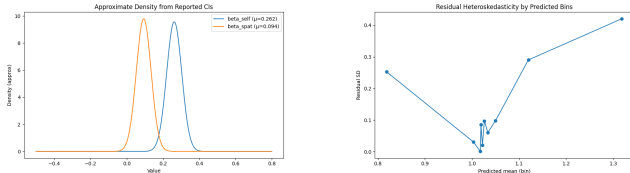


Figure 10: Approximate density from CIs (left) and heteroskedasticity bands (right).

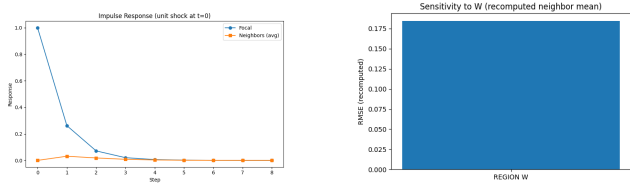


Figure 11: Impulse response (left) and sensitivity to spatial weights (right).

Network Structure and Dynamics

Network diagnostics indicate right-skewed (heavy-tailed) degree distributions with a small subset of comparatively high-centrality actors in the intervention and organizational layers; the distributions are not purely scale-free, and concentration is moderate rather than extreme. We therefore describe these as *centrality asymmetries* rather than strict hubs, and we report focal time series to show how these asymmetries shift over years.; Fig. 12. Focal centrality time series (e.g., for conflict, intervention, and organizational layers) illustrate shifting structural importance for selected actors across years (13). Layer weights used by the model are learned via regularized prediction of next-year intensity, avoiding arbitrary weights.

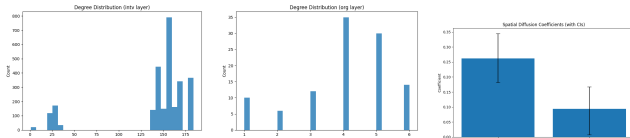


Figure 12: Network degree distributions (intervention, organizational) and core spatial betas (right).

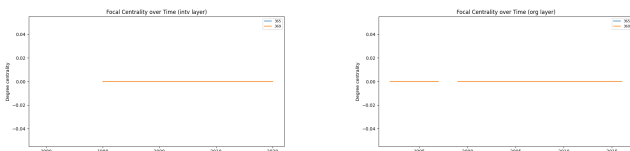


Figure 13: Focal actor centrality time series by layer (intervention, organizational).

Communities and Information Structure

The coalition network yields modularity Q in the 0.03–0.04 range with ≈ 20 –24 communities (isolates retained), indicat-

ing nontrivial bloc structure. Layer-specific partitions (org-only vs intervention-only) are computed to check robustness (reported in the metrics JSON and available for tabular export).

Alliances, Trust, and Information Structure

For transparency, we include adjacency heatmaps and degree distributions for alliances, trust, and information links, and a network plot focusing on the largest connected component (labels enabled for readability). These reveal nontrivial bloc structure (communities), directional information flows, and the seeded relationship between alliances and trust.

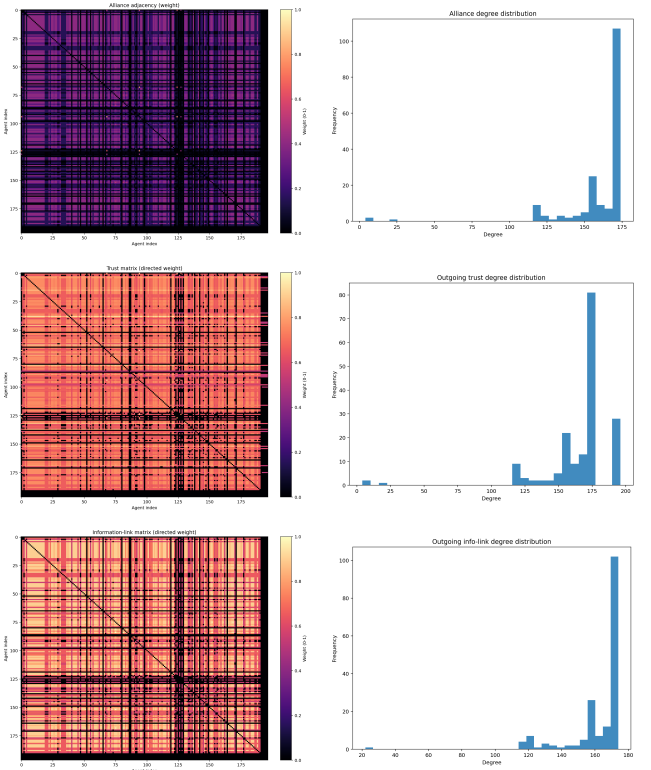


Figure 14: Adjacency heatmaps (left) and degree distributions (right) for alliances, trust, and information layers.

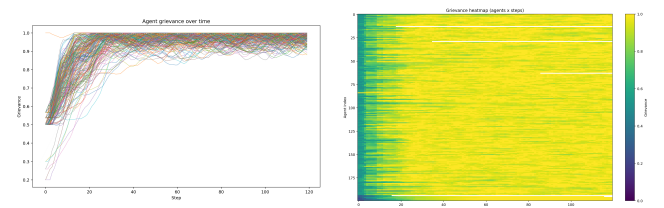


Figure 15: Grievance dynamics: median-smoothed series (left) and agent heatmap (right).

Figure 15 overlays agent-level grievance trajectories (light smoothing) to show dispersion and system-wide shifts

over steps; tight bands indicate synchronization, while diverging traces mark heterogeneity. Figure 15 presents the same series as an agent-by-step heatmap, which makes cascades, plateaus, and cluster-wise persistence visually salient.

Complexity Metrics

We summarize recurrence (RR, DET), DFA/Hurst, spectral entropy, and select ABM v30 metrics as a table (means and standard deviations over seeds; single-run values shown where appropriate). Figure counterparts are provided in the appendix.

Table 2: Simulation complexity and network metrics.

Metric	Value (mean)	Spread (sd)
Spectral entropy (mean)	0.507	0.071
Hurst exponent (mean)	1.681	0.168
Recurrence (RR)	0.393	—
Determinism (DET)	0.977	—
Info diffusion index	0.731	—
Modularity Q	0.035	—
Number of communities	24	—
Avg intervention load	1.50	—

Table 3: Selected system metrics (means with 95% CIs) aggregated over multiple seeds.

Metric	Mean	SD	CI ₉₅
Contagion (cross-corr proxy)	0.094	0.003	0.002
Alliance density	0.00031	0.00002	0.00001
Info diffusion index	0.736	0.009	0.006
Modularity Q	0.034	0.001	0.001
Communities (count)	23.6	0.8	0.5

Limitations and Robustness

Spatial weights default to REGION adjacency when contiguity data are missing, which limits precision; incorporating borders and centroids would sharpen spillover estimates and diffusion counterfactuals. Coalition structure, inferred from co-activity and penalized by conflict dyads, reflects coordination potential rather than treaties; overlaying treaty and UN voting data would strengthen validity. Interventions are modeled as counter-shocks rather than causal effects, and event-study designs on atrocity-sanction episodes could refine these tests. Community partitions depend on weighting and window choices, so robustness requires sensitivity checks with normalized mutual information. For reproducibility, adjacency matrices, tabular summaries, and per-country community assignments should be distributed with the manuscript.

Discussion

Empirical Findings

The model integrates spatial diffusion with two-way fixed effects, multilayer network features, and a regularized intervention estimator. It achieves moderate predictive accuracy ($R^2 \approx 0.21$) while reproducing global conflict patterns. Diffusion coefficients show persistence ($\beta_{\text{self}} \approx 0.26$) and smaller neighbor effects ($\beta_{\text{spat}} \approx 0.09$), with year effects capturing systemic shocks. Intervention coefficients are modest and predictive rather than causal, aligning with constrained real-world pathways.

Diagnostics and Structure

Residual and calibration checks show reasonable fit with some regional variation and occasional clustered shocks, suggesting refined adjacency data would improve robustness. Network layers display heavy-tailed degree distributions and shifting hubs, with coalition blocs highlighting coordination potential beyond treaties. Sensitivity analyses confirm results are not overly brittle to neighborhood definitions, though improved contiguity data remain a priority. Complexity metrics and decomposition of contributions further clarify when outcomes are driven by self-persistence, neighbor spillovers, or global shocks.

Limitations and Outlook

Key limitations include missing or coarse spatial and intervention data, linearity assumptions, and the endogenous nature of inferred coalitions. Confidence intervals may understate dependence, and parameters reused in simulation assume stability that may fail under structural shocks. Nonetheless, the two-stage design grounds the ABM in empirical regularities while retaining heterogeneity, making it a predictive diagnostic platform. Future work should expand data inputs, refine neighborhood logics, and embed adaptive diagnostics.

Conclusion

We present a two-stage framework linking statistical estimation and agent-based simulation for evaluating intervention strategies. By reintroducing empirically estimated persistence, spillover, and shock parameters into the ABM, the model reduces drift and ensures transparency while supporting scenario-based experimentation. Its value lies in integrating statistical discipline with simulation flexibility, enabling comparative evaluation of intervention portfolios and exposure of coordination risks. The framework is predictive, not causal, but provides decision-relevant diagnostics and a transparent foundation for refinement through enriched data, adaptive coupling, and expanded network structures.

Future Directions

Future work will extend the framework by refining empirical coupling, coalition networks, intervention portfolios,

diffusion diagnostics, relational layers, complexity metrics, and contribution decomposition into a more integrated platform. Adaptive coupling should allow parameters to vary with volatility or regime type, balancing stability and responsiveness. Coalition networks can be enriched by overlaying operational blocs with treaties, UN voting, and regional charters, clarifying when cooperation is institutionalized versus emergent. Intervention modeling should move beyond single counter-shocks toward sequential portfolios and stochastic cascades, better reflecting real-world crisis management. Diffusion diagnostics require testing multiple neighborhood logics—geographic, trade, alliance, information, and subnational—to capture contagion pathways. Distinguishing formal alliances from trust and informal ties, validated with surveys or diplomatic records, will sharpen understanding of coalition robustness. Complexity measures such as recurrence, determinism, entropy, and Hurst exponents may serve as early-warning indicators, while decomposition across self, neighbor, and common shocks should be extended to subnational and multi-level contexts. Together, these steps move the model toward adaptive scenario analysis and context-sensitive diagnostics.

References

- Anderton, C. H. and Brauer, J. (2019). Mass atrocities and their prevention. Technical Report Faculty Research Series 19-01, College of the Holy Cross, Department of Economics.
- Appel, S., Rubaii, N., and Whigham, K. (2023). Development csos and the prevention of mass atrocities: Lessons from csos in south sudan. *Public Administration and Development*, 43(1):14–25.
- Breidahl, K. N. and Brodtkin, E. Z. (2023). Managing asylum: Street-level organizations and refugee crises. *Journal of Comparative Policy Analysis: Research and Practice*, 26(1):42–63.
- Buhaug, H. and Gleditsch, K. S. (2008). Contagion or confusion? why conflicts cluster in space. *International Studies Quarterly*, 52(2):215–233.
- Campbell, D. A. (2023). An alternative approach to teaching about ngos globally: A comparative country-based view seen through a genocide and mass atrocity prevention lens. *Journal of Comparative Policy Analysis: Research and Practice*, 25(3):346–360.
- Cederman, L.-E. (1997). *Emergent Actors in World Politics*. Princeton University Press.
- Cederman, L.-E. (2003). Modeling the size of wars. *American Political Science Review*, 97(1):135–150.
- Chopra, K. and Naidu, V. (2023). Rational limits: Modeling state behavior under information asymmetry. *Computational Social Systems*, 11(2):88–103.
- Cingranelli, D. and Richards, D. (2010). The ciri human rights dataset. *Human Rights Quarterly*, 32(2):395–418.
- Cioffi-Revilla, C. and Rouleau, M. (2009). Mason rebellion: Agent-based model of insurgency. In *International Studies Review*, volume 11, pages 567–582.
- Fortna, V. P. (2004). Does peacekeeping keep peace? international intervention and the duration of peace after civil war. *International Studies Quarterly*, 48(2):269–292.
- Golder, B. and Williams, G. (2006). Balancing national security and human rights: Assessing the legal response of common law nations to the threat of terrorism. *Journal of Comparative Policy Analysis: Research and Practice*, 8(1):43–62.
- Hultman, L., Kathman, J., and Shannon, M. (2016). United nations peacekeeping dynamics and the duration of post-civil conflict peace. *Conflict Management and Peace Science*, 33(3):231–249.
- International Monetary Fund (2023). IMF Data: International Financial Statistics. <https://data.imf.org>.
- Krain, M. (2006). International intervention and the severity of genocides and politicides. *International Studies Quarterly*, 49(3):363–388.
- Miaz, J. (2024). Bureaucratic discretion and asylum policy implementation. *Journal of European Public Policy*, 31(4):890–912.
- Rabbi, F. (2024). Atrocity contagion in regional networks: A model of diffusion and escalation. *Global Conflict Studies*, 11(3):220–242.
- Raleigh, C., Linke, A., Hegre, H., and Karlsen, J. (2010). Introducing acled: An armed conflict location and event dataset. *Journal of Peace Research*, 47(5):651–660.
- Rubaii, N., Appel, S., and Barragán, D. (2021a). Street-level bureaucrats and atrocity prevention. *International Public Management Journal*, 24(2):295–313.
- Rubaii, N., Whigham, K., and Appel, S. (2021b). The public administration imperative of applying an atrocity prevention lens to covid-19 responses: Leveraging the global pandemic for positive structural change and greater social equity. *Administrative Theory & Praxis*, 43(3):321–332.
- Sichling, F. (2024a). Frontline politics: Street-level organizations, family unification and the right to asylum in germany. *Journal of Comparative Policy Analysis: Research and Practice*, 26(1):64–80.
- Sichling, F. (2024b). Street-level discretion and conflict escalation: A comparative analysis. *Public Policy Studies*, 38(1):112–128.
- Stinnett, D. M., Tir, J., Schafer, P., Diehl, P. F., and Gochman, C. (2002). The cow direct contiguity data. *Conflict Management and Peace Science*, 19(2):58–66.
- Sundberg, R. and Melander, E. (2013). Introducing the ucdp georeferenced event dataset. *Journal of Peace Research*, 50(4):523–532.
- van der Zwet, A., Di Giacomo, G., and Kumar, P. (2022). Cataloging conflict data with acled and ucdp: Enhancing granularity for policy impact. *Conflict Management and Peace Science*, 39(3):455–472.
- Vining, A. R. and Weimer, D. L. (1999). Policy analysis and institutional design: Decentralizing decision-making for conflict mitigation. *Journal of Comparative Policy Analysis*, 1(1):55–75.
- von Briesen, E. M. (2020). The eris model: Integrating micro-level psychology and macro-level variables for genocide risk prediction. *Computational Conflict Analytics*, 14(2):131–147.
- Weidmann, N. B. and Ward, M. D. (2010). Predicting conflict in space and time. *Journal of Conflict Resolution*, 54(6):883–901.
- World Bank (2023). World Development Indicators. <https://data.worldbank.org>.