Modeling Ukraine Military Events

Seth Spire, Ethan Meidinger, Maggie Crowner, Natalie Seah, Reed Baumgardner, and Max Ellingsen

Data Background

- Data is recorded by ACLED (Armed Conflict Location and Event Data), backed by United Nations
  - Slightly deviates from given data set, emphasizes all events across entire country with any political relation
  - County-equivalent (raion) analysis rather than state-equivalent (oblast) level
- Pertinent data includes date, time, location, description, casualties, and actors involved within the event, though additional information is available
- 21 days were missing from the provided data set
  - Reason for missing data was not recorded
  - Assumption of zero conflict was made to ensure the best possible model
- Data spans from 1/1/2020 – 10/23/2024

What is Being Modeled?

Daily Dominant Location of Events

Events analyzed:

- Armed clash
- Air/drone strike
- Shelling/artillery/missile attack
- Disrupted weapons use

Daily Fatality Severity

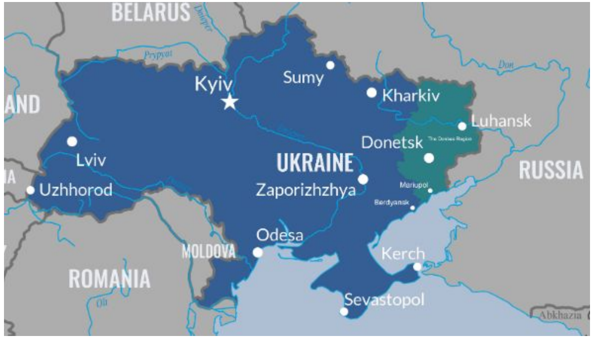
Separated into bins:

- 0
- 1-100
- 101-200
- 201-300
- 301+

Main Features of the Data

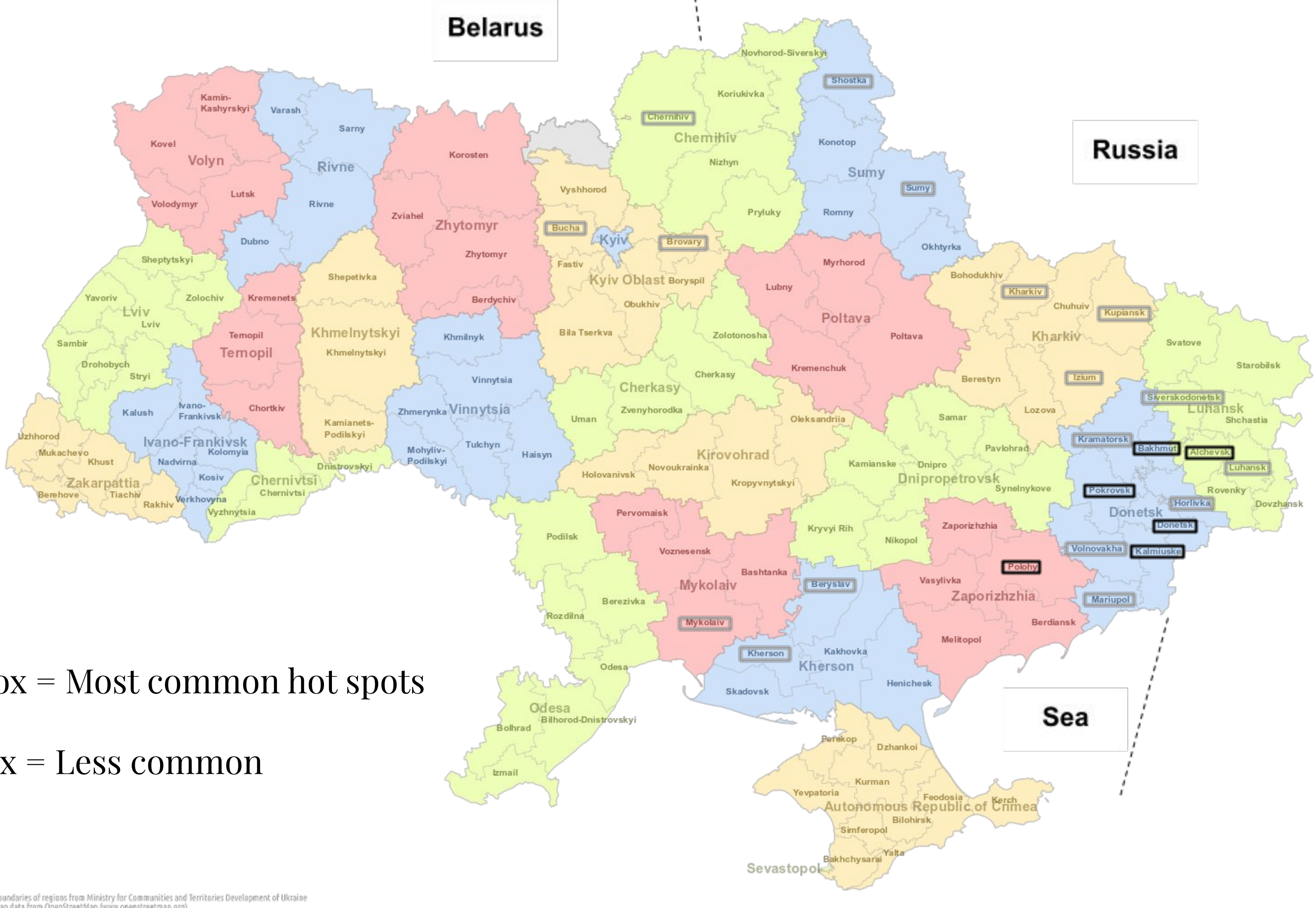
We found an uneven geographic distribution of events

- Donetskyi 417 (23.7%)
- Pokrovskyi 365 (20.8%
- Bakhmutskyi 335 (19.1%)
- Kalmiuskyi 125 (7.1%)
- Alchevskyi 125 (7.1%)
- Polohivskyi 116 (6.6%)
- All other regions combined for about 15.6%



Fatality data skews towards low-severity days

- 699 days recording 0 fatalities,
- 552 days between 1 and 100,
- 306 days between 101 and 200,
- 105 days between 201 and 300, and
- 75 days exceeding 301 fatalities.



Black Box = Most common hot spots

Gray Box = Less common

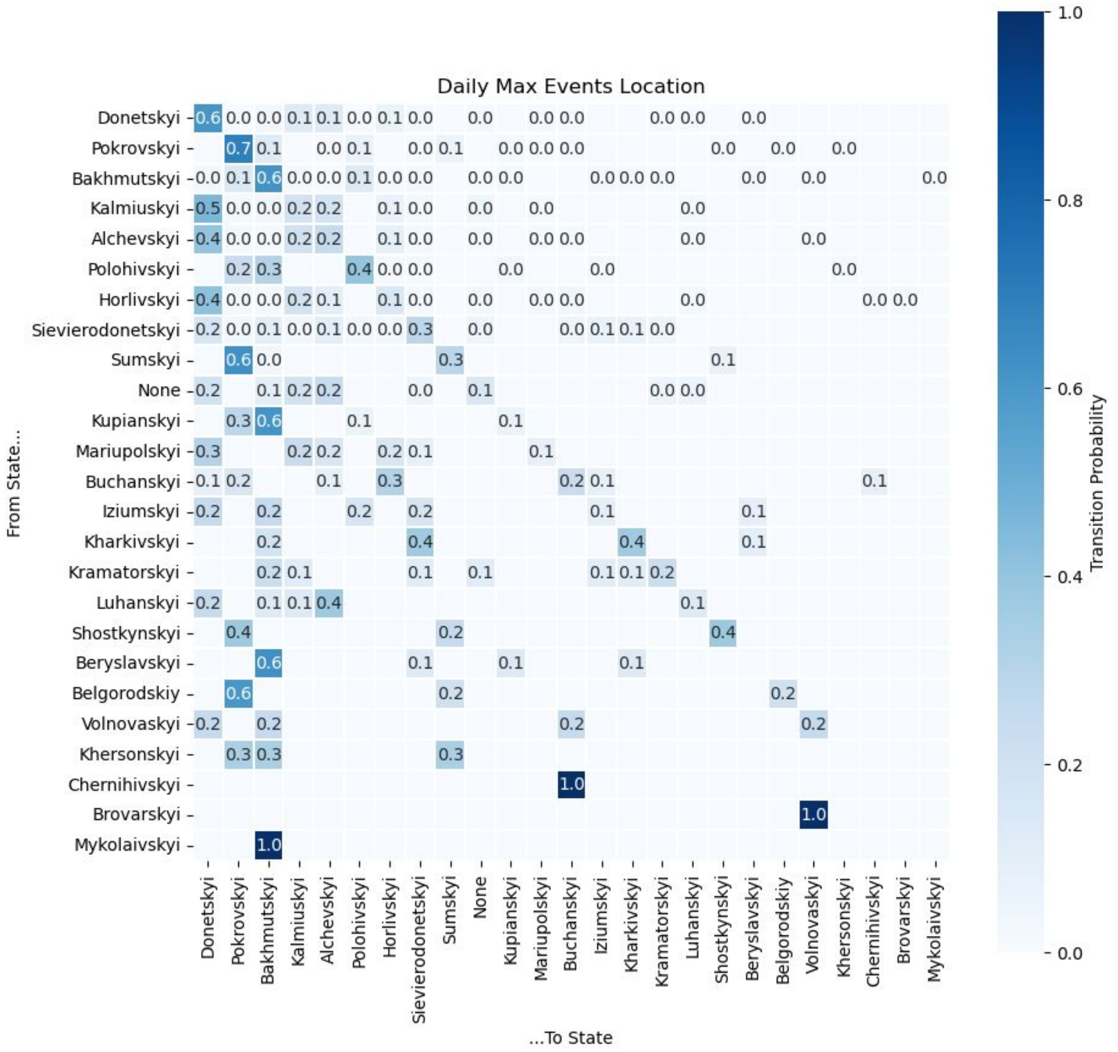
Daily Location of Most Events Markov Chain

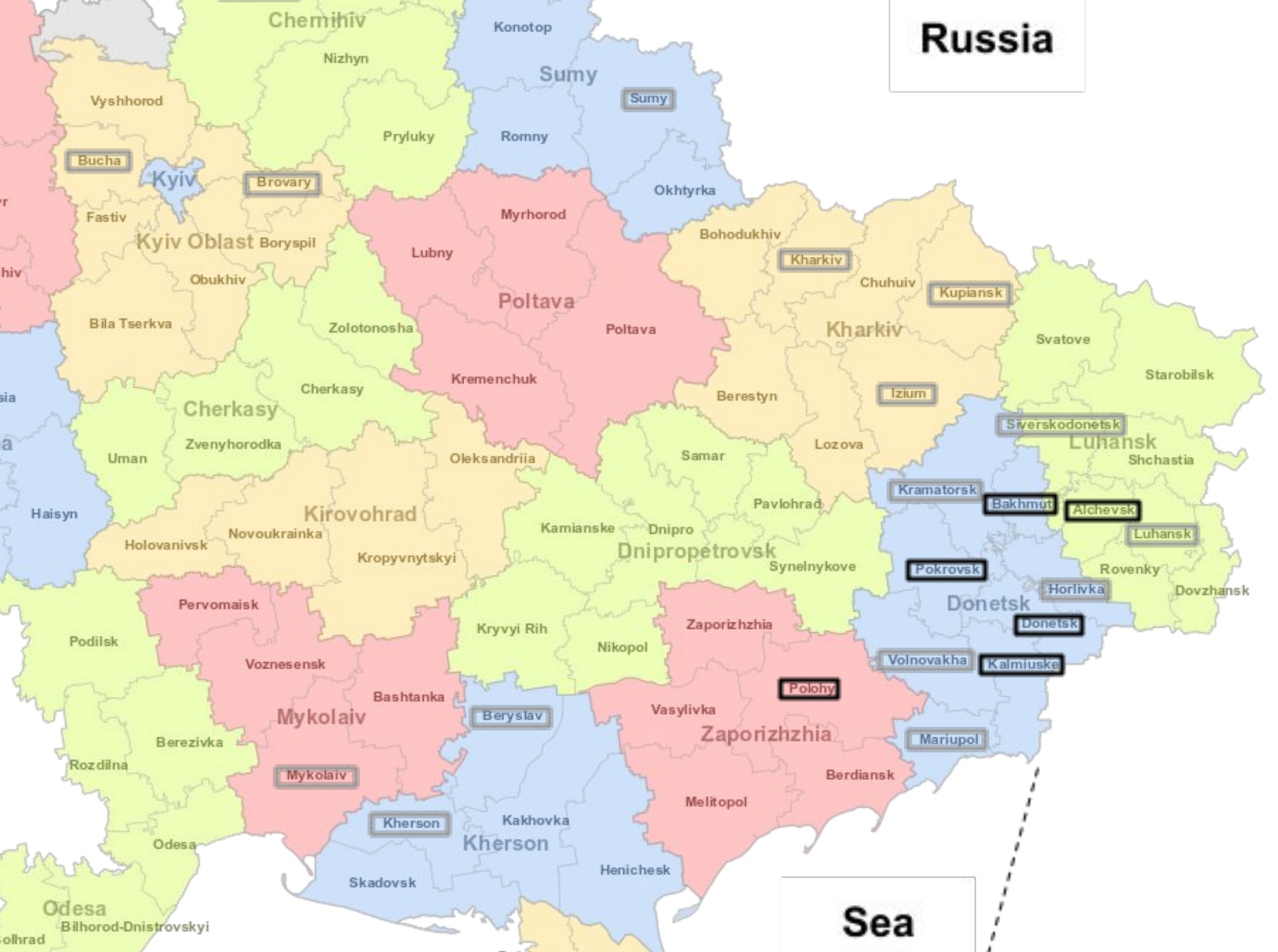
Event hotspots include:

- -Donetskyi
- -Bakhmutskyi
- -Pokrovskyi
- -Polohivskyi
- -Kalmiuskyi
- -Alchevskyi

Two noticeable cycles:

- -Pokrovskyi + Bakhmutskyi + Polohivskyi (2022 - Pres)
- -Donetskyi + Kalmiuskyi + Alchevskyi (2020 - 2022)




Russia
Chernihiv
Nizhyn
Konotop
Sumy
Sumy
Romny
Okhtyrka
Vyshhorod
Pryluky
Bucha
Kyiv
Brovary
Fastiv
Kyiv Oblast
Boryspil
Obukhiv
Bila Tserkva
Lubny
Myrhorod
Poltava
Poltava
Zolotonosha
Kremenchuk
Bohodukhiv
Kharkiv
Chuhuiv
Kupiansk
Kharkiv
Berestyn
Izium
Svatove
Starobilsk
Cherkasy
Cherkasy
Zvenyhorodka
Uman
Haisyn
Oleksandriia
Samar
Lozova
Siverskodonetsk
Luhansk
Shchastia
Kramatorsk
Bakhmut
Alchevsk
Luhansk
Kirovohrad
Holovanivsk
Novoukrainka
Kropyvnytskyi
Kamianske
Dnipro
Dnipropetrovsk
Pavlohrad
Synelnykove
Pokrovsk
Horlivka
Rovenky
Dovzhansk
Donetsk
Donetsk
Pervomaisk
Podilsk
Kryvyi Rih
Nikopol
Zaporizhzhia
Volnovakha
Kalmiuske
Polohy
Voznesensk
Bashtanka
Beryslav
Vasylivka
Zaporizhzhia
Mariupol
Mykolaiv
Berezivka
Rozdilna
Mykolaiv
Berdiansk
Melitopol
Kherson
Kakhovka
Kherson
Odesa
Henichesk
Skadovsk
Odesa
Bilhorod-Dnistrovskyi
Sea

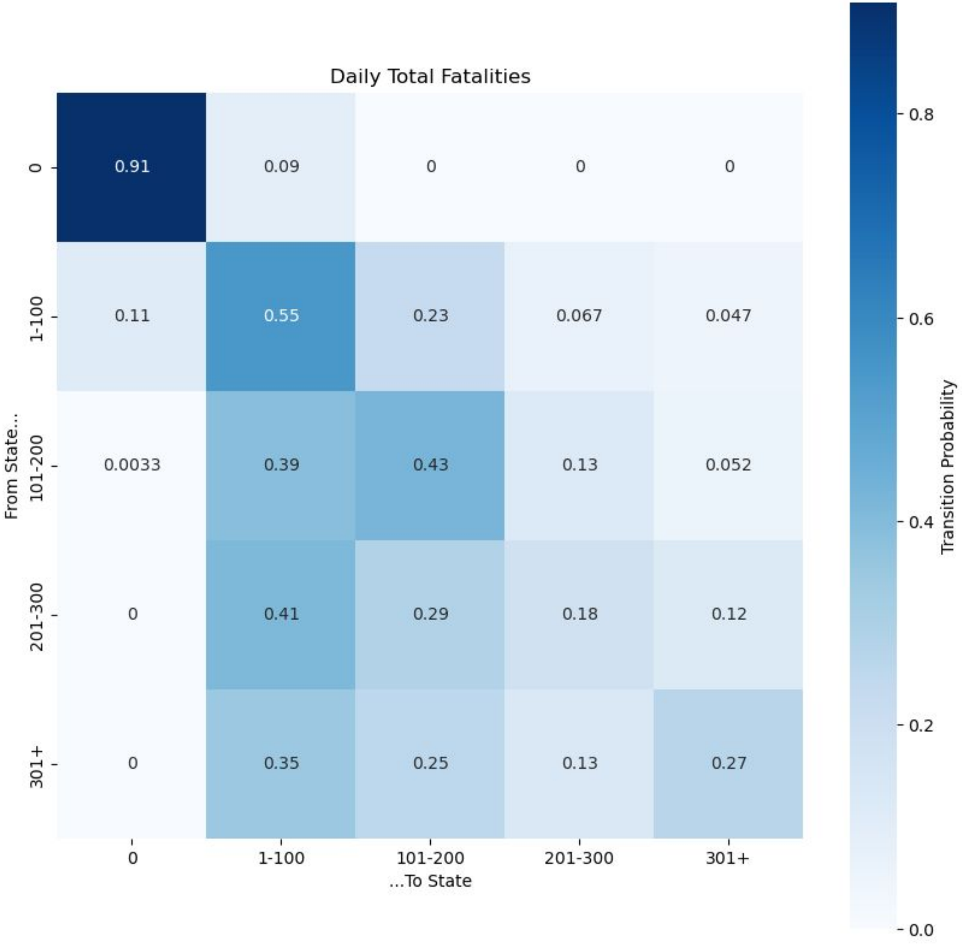
Daily Location of Most Events by year

- Pre 2022 Invasion mostly in Donetskyi, Kalmiuskyi, and Alchevskyi which forms cycle in Markov Chain
- Post Invasion mostly in Pokrovskyi, Bakhmutskyi, and Polohivskyi which also has cycle
- But this year-to-year progression not truely modeled

admin2	2020	2021	2022	2023	2024
Donetskyi	234.0	159.0	24.0	0.0	0.0
Pokrovskyi	0.0	2.0	13.0	119.0	231.0
Bakhmutskyi	8.0	8.0	164.0	146.0	9.0
Kalmiuskyi	38.0	80.0	7.0	0.0	0.0
Alchevskyi	38.0	68.0	19.0	0.0	0.0
Polohivskyi	0.0	0.0	19.0	87.0	10.0

Daily Total Fatalities Markov Chain

- No single-day jumps from 0 fatalities to 301+ fatalities
  - Increase in fatalities happens gradually, no sudden spikes
- No single-day jumps from 301+ fatalities to 0 fatalities
  - Decrease in fatalities happens at a quicker rate than the increase
- Daily total fatalities is heavily right skewed



Daily Total Fatalities by Year

- Pre 2022 Invasion mostly at zero fatalities
- Post invasion is much higher
- Markov Chain partially models this (stays at zero, or stays higher) but can jump back and forth

fatality_level	2020	2021	2022	2023	2024
0	321.0	330.0	45.0	3.0	0.0
1-100	26.0	34.0	194.0	221.0	77.0
101-200	0.0	0.0	81.0	107.0	118.0
201-300	0.0	0.0	26.0	22.0	57.0
301+	0.0	0.0	18.0	12.0	45.0

Generating New Sequences with Markov Chains

- Used trained Markov chain models to simulate plausible future daily sequences
  - 1000 days of simulated location states
    - Mariupolskyi → Alchevskyi → Donetskyi → Donetskyi → Kalmiuskyi → Alchevskyi → Donetskyi → ...
  - 1000 days of simulated fatality levels
    - 201–300 → 1–100 → 101–200 → 1–100 → 1–100 → 301+ → 1–100 → 201–300 → ...

Outcome of Markov Chain Compared to Train

Table 1: Training vs Simulated Fatality Frequencies

Fatalities	Train	Sim
0	0.402418	0.404
1–100	0.317789	0.311
101–200	0.176166	0.186
201–300	0.060449	0.063
301+	0.043178	0.036

Table 2: Training vs Simulated Location Frequencies

Location	Train	Sim
Donetskyi	0.237201	0.241
Pokrovskyi	0.207622	0.211
Bakhmutskyi	0.190557	0.211
Kalmiuskyi	0.071104	0.062
Alchevskyi	0.071104	0.070
Polohivskyi	0.065984	0.061
Horlivskyi	0.035836	0.043
Sievierodonetskyi	0.024460	0.017
Sumskyi	0.018203	0.012
None	0.011945	0.008
Kupianskyi	0.010239	0.012
Mariupolskyi	0.007395	0.011
Buchanskyi	0.007395	0.005
Iziumskyi	0.006826	0.001
Kharkivskyi	0.006257	0.004
Kramatorskyi	0.005119	0.005
Luhanskyi	0.004551	0.003
Shostkynskyi	0.004551	0.005
Beryslavskyi	0.004551	0.005
Belgorodskiy	0.002844	–
Volnovaskyi	0.002275	0.006
Khersonskyi	0.001706	0.003
Chernihivskyi	0.001138	0.002
Brovarskyi	0.000569	0.002
Mykolaivskyi	0.000569	–

Good for Modeling Fatality
Good for Modeling Location

Why is this happening?

- Single order Markov Chains are limited
  - Not able to predict the longer sequences that may have led to other minority locations
  - Certain locations were not predicted
- All fatality classes were predicted
  - Less classes with more even distribution (compared to location) led to every class being predicted
- Still cannot fully model year-to-year progressions completely
  - Real war looks different before and after 2022 invasion, but simulations will go between those pre and post invasion cycles

Conclusion

- Fatalities rise and fall gradually, with conflict concentrated in recurring hotspot cycles
- Simulations effectively modeled short-term changes in fatalities and geographic spread
- Future work:
  - Higher order Markov models
  - Incorporate more data such as geographic proximity or event characteristics
  - Handle year-to-year changes