Mihai Croicu

Forecasting battles

New machine learning methods
for predicting armed conflict

UPPSALA
UNIVERSITET

Dissertation presented at Uppsala University to be publicly examined in Brusewitzsalen, Gamla Torget 6, Uppsala, Friday, 21 March 2025 at 13:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Thomas Chadefaux (Trinity College, Dublin).

Abstract

Croicu, M. 2025. Forecasting battles. New machine learning methods for predicting armed conflict. *Report / Department of Peace and Conflict Research* 132. 62 pp. Uppsala: Uppsala University. ISBN 978-91-506-3086-2.

Over the past decade, the field of conflict forecasting has undergone a remarkable metamorphosis, transforming from a series of isolated efforts with low predictive power into large, globe-spanning projects with impressive performance. However, despite this evolution, many challenges still remain. First, while we are good at predicting absolute risks, we are poor at predicting conflict dynamics (onsets, escalations, de-escalations and terminations). Second, we are over-reliant on spatio-temporal features and mechanistic models due to the nature of the event-data we use, thus excluding actor agency. Third, we do not handle either data or model uncertainty. Fourth, we are lagging behind the state-of-the-art in machine-learning. This dissertation attempts to resolve some of these salient difficulties, by contributing to six core elements of current-generation forecasting systems. First, **time**, by looking at the substantive effects and uncertainties of the temporal distance between data and forecast horizons. Second, **space**, by looking at the inherent uncertainties of high-resolution geospatial data and proposing a statistical method to address this. Third, **feature space**, by tackling the extreme feature sparsity in event-data and proposing a novel, deep active learning approach to mine features from existing large conflict-related text corpora. Fourth, **substantive knowledge**, by combining findings from the previous papers to take a fresh look at the microdynamics of conflict escalation. Fifth, **the forecasting process** itself, by building models that directly forecast from text, eliminating the intermediate step of manual data curation. Finally, **the frontier of event-data**, by looking at whether the news-media heavy way we collect violent fatal events can be extended to the collection of non-violent events. Methodologically, the dissertation introduces state-of-the art methods to the field, including the use of large language models, Gaussian processes, active learning and deep time series modelling. The six papers in the dissertation exhibit significant performance improvement, especially in forecasting dynamics.

Keywords: conflict forecasting, predictive methodology, event data, battle events, spatial forecasting, machine learning, large language models, computational linguistics, civil war, armed conflict

Mihai Croicu, Department of Peace and Conflict Research, Box 514, Uppsala University, SE-75120 Uppsala, Sweden.

ISSN 0566-8808
ISBN 978-91-506-3086-2
URN urn:nbn:se:uu:diva-545176 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-545176)

To Andra

List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I Hegre, H., Croicu, M., Eck, K. and Högbladh, S. (2020). Introducing the UCDP Candidate Events Dataset. Published in *Research and Politics* 7(3):1–8.

II Croicu, M. (2024). Enhancing geospatial precision in conflict data: A stochastic approach to addressing known geographically imprecise observations in conflict event data. A previous version of this paper was presented at the *64th Annual Convention of the International Studies Association, Montréal*, 15–18 March 2023.

III Croicu, M. (2024). Deep Active Learning for Data Mining from Conflict Text Corpora. A previous version of this paper was presented at the *Using LLMs and Text-as-Data in Political Science Research* workshop, Barcelona, 31 January 2024.

IV Croicu, M., Kreutz, J. (2024). Provocation by Design? Holy Places, Public Transport, and Civil Conflict Escalation. Under review at *International Organization*.

V Croicu, M., von der Maase, S. P. (2024). From Newswire to Nexus: Using Text-Based Actor Embeddings and Transformer Networks to Forecast Conflict Dynamics. A previous version of this paper was presented at the *120th American Political Science Association Annual Meeting*, Philadelphia, 5–8 August 2024.

VI Croicu, M., Eck, K. (2022). Reporting of non-fatal conflict events. Published in *International Interactions* 48(3):450–470.

Contents

1. Acknowledgements

This dissertation feels like it has had a longer and much more circuitous path than most. Its story, in fact, goes way back, to one of my first conscious memories. To a cold, dark winter night, almost exactly 35 years ago to the day I'm writing this. And, as it is with most early childhood memories, it's blurry, and to some extent pieced together and reconstructed after-the-fact. It was a memory of a room in pitch black darkness, with three people in it: me and my parents. Outside, weird, loud, repeating, banging noises, and bright flashes of light, in what to a four-year old me felt like a game. Me being told to keep quiet, not make a sound. Then slowly crawling to the only room with no windows. Me sensing a kind of fear I've never seen before, not in me – I distinctly remember not being afraid, but in my parents. Followed by loud thuds, almost like thunder strikes. Then it cuts out for a while. My next memory must have been the next day, or perhaps it was next week? I can't tell. A bright, lit winter, morning. With happy people, a happiness I've never, in fact, ever felt or seen ever after. My parents and neighbors giving out food (I want to say cabbage rolls, but, again, memories of a four-year old are feeble and uncertain) to unknown people, all dressed up in weird and nearly identical green clothes, carrying with them a big flag, with a big round hole in the middle. And celebrating, using weird words, of which I don't remember any, but I now know were such like revolution, freedom, victory, relief, democracy, joy. I now know what those sounds, sights and feelings were: they were the sounds of war, etched forever as some of the first conscious memories I have. For the days I remember were December days, and the year was 1989, and I was in a ground floor apartment in Bucharest, Romania, in a high-rise strategically wedged between two ministries and a military hospital. I later came to know those sounds came from a firefight, the flashes and bangs being those of 7.62 caliber military rifle rounds being discharged. The loud noises were those of glass shattering, mostly in the parking lot outside. With some making it to the inside of our apartment (I do believe my dad has kept at least one of those rounds that punched a hole through our apartment windows and lodged itself somewhere inside our living room to this day; I also understand the replacement of the enormous pane of window glass that now had holes through it took a while). This early experience, memory, trauma, whatever you may call it set in motion a curiosity that became a lifelong interest, with all the questions that come with it. Questions I discovered, many years later, nobody actually had an answer to – we don't yet even have a conclusive answer to the fundamental questions of why we fight, why we go to war, and why it is a fundamental part of human nature. Questions that, slowly, set me on the path to what became this dissertation.

As you can expect, there were many, many people that walked besides me on this long and twisting path taking me from that December night to today, shaping my intellectual development and curiosity, and guiding me along the path that eventually resulted in this dissertation. Closest to today, and most important for the content in the book you are reading right now, they are, of course, my three supervisors: Håvard Hegre, Joakim Kreutz and Nils Weidmann. Håvard is one of, if not the most brilliant person I have had the pleasure and opportunity to meet, a true intellectual force. And not only that, but perhaps even more importantly, he has always had an unwavering belief in me, in my abilities, and in my intellectual capacity, even at those dark times I have deeply doubted myself – to the point where he took a last minute early flight from some important conference in one of my many hours of need to simply have a coffee and a chat with me. I deeply appreciate it all and will never forget your trust, kindness, honesty, friendship, mentorship and appreciation. And thank you for seeing something in me that I could not see myself. Joakim is a great friend, always there, always supporting me, always open to me, since the day we've met over a decade ago. Without all the discussions on anything and everything, from football to rebellion, from methods to coffee and from family to life in general, this thesis, and the Master's thesis that preceded this dissertation would not have ever happened. Thank you for the uncountably many happy lunches, dinners and beers, sometimes to the early hours in the morning. Thank you for your steadfast support, enormous trust and immensely cherished friendship. And thank you for guiding my first steps into the world of research, with our joint little idea more than eleven years ago becoming my first peer-reviewed paper on conflict I've ever written. And let there be many, many more joyful occasions in the future! Last but not least, Nils, the one that brought order into what was otherwise a chaotic process – the person that saw the big picture, stood up for my interests above all else, and saw the end of the tunnel even when it felt like I was spinning in circles. Thank you, also, for guiding my steps towards machine-learning and towards computational linguistics – without our joint project ten years ago, my life would have, in all likelihood, taken a vastly different (and in all likelihood much less interesting) path.

But this team of giants on whose shoulders I stand is the latest act in a long list of people that have provided intellectual inspiration, mentorship and support. In somewhat reverse chronological order – Simon Polichinel von der Maase, for being such a great and trusted friend, and amazing co-author. Simon, thanks for it all, and here's to many future great adventures! I will always remember our serendipitous meeting, when you were visiting and we suddenly discovered we were doing the exact same thing (to the third digit after the decimal point!). Kristine Eck, for pushing me to actually finish my long-delayed Master's thesis, without which no PhD would have been possible. And, then, for being such an amazing co-author, with really cool ideas and unparalleled energy to see it through even in the worst Covid times. Hanne Fjelde, Nina von Uexküll and Halvard Buhaug, for teaching me what my real ambition level should be, and for letting me take part in your moon-shot (that ended up working out!). And, before that, Peter Wallensteen and Erik Melander, as well as the UCDP-trio, Lotta Themner, Stina Högbladh and Therese Pettersson for taking me on board UCDP, and then entrusting me with extremely challenging technical responsibilities. And Magnus Öberg, for the long list of discussions and ideas and suggestions and help that made everything so much better, beginning with that first time you've read one of my papers, at a seminar in 2008. And Margareta Sollenberg, for all her support and kind words and knowledge and standing up for me throughout the years. And then there are Niklas Karlén and Vanessa Maier, it was a pleasure co-authoring with you! And, to top this list off, we must go back to the very beginning of my academic trip – none of this would have ever been possible without Mats Hammarström, a true legend, of whom I am the last student in an illustrious lineage, whose open door and encouragement almost fifteen years ago started my journey in the department, giving help at one of those many times of hopelessness encountered along the academic path.

And then there are my many friends that have helped me in countless ways all along the twisting path, becoming my academic family. Paola Vesco, for all the soul-seaching and philosophical pondering that went in our many conversations, as well as all the nearly infinite kindness and encouragement you've given me. Jim Dale, for being a good friend, taking so many loads off my shoulders and helping out with everything, as well as for the many beers, coffees and chats on everything from Daleks to rabbits and from galaxies to whiskey! Espen Geelmuyden Rød and Sabine Otto, for the wonderful friendship, insights, mentorship (and sometimes wine), helping me to avoid so many pitfalls and dead-ends that you've gone through and made sure I didn't. Then, there's Maxine Leis whom I thank for all the intellectual discussions as well as some really powerful encouragement when I needed it the most, Thorsten Rogall, for going above and beyond to help with so much great feedback on my papers (oh, and for the jokes!), Sofia Nordenving, for all the kind words and Malika Rakhmankulova for all the chats about shared experiences in lands and times far-far away. And to all of you, thank you for all the many wonderful days with our shared love of good lunches, fancy coffee and jokes. Then, there's David Randahl, with whom I started the PhD at the same time, and who was the one person that shard my weird interest at the meeting point of computer science, conflict research and statistics, and could actually provide with deep feedback and interesting ideas. Of course, this list needs to be completed by the rest of the VIEWS team, now and past, without which this amazing project would never have been: Angelica Lindqvist-McGowan, Alexa Timlick, Marina Matić, Chandler Williams, Xiaolong Sun, Garrett Benz, Frederick Hoyles and Remco Jansen. Then, there is my extended cohort of PhD students with which I have shared the trials and tribulations – besides David, Kate Lonergan, Kristina Petrova, Ida Rudolfsen, Gulla Olafsdottir and Juan Diego Duque Salazar – with which many beers as well as many happy and sad days were shared, in Uppsala, at conferences and courses. And then again Stefan Döring and Sebastian van Baalen, always there with their experience, always ready to help.

And, there is the support group that is the *Wednesday Group* (I am old enough to remember it as the *Monday Group*!). Some mentioned above, but also Amanda Lanigan, Hüsne Akgol, Jenniina Kotajoki, Giacomo Cassano, Tim Gåsste, thank you all for the insightful comments and support.

And, obviously, the many friends from my UCDP times that have provided help, support, kindness and friendship - particularly Sam Taub, for all the kindness, warmth and friendship, Ralph Sundberg, for seeding the idea that a PhD is achievable, and Johan Brosché, as well as Gökhan Çifikli, Chris Shay and David Gebre-Medhin – you three were right, I did change my mind in the end and did do a PhD! Then, of course, there's Marie Allansson, who made my life easier in a thousand different ways over a thousand of different occasions and Susanne Schaftenaar – it's 16 years we go back by now, incredible! And, then, there are the friends in the Department: Helena Grusell and our shared love of Indian food, Eric Skoog, for his selflessness and heartfelt desire to help, Liana Lopes and her stories about the warm beauty of the tropics even in the darkest of days (also, thanks with all the help with bureaucracy), Sofia Hatz and Allan Dafoe, for all the comments to my work and for all the many and interesting AI-related discussions, and Christofer Hägg and his teaching of the intricate dark arts of keeping track of prolongation hours! And then there are the people that looked out after us and our wellbeing while doing all this: Desirée Nilsson, Lisa Hultman, Kristine Höglund, Isak Svensson, Helena Milroth, Susanne Olsson, Tereza Heiligová and all the rest of the DPCR staff that made this journey possible and worthwhile.

Then, there are those at PRIO, the home away from home, and the Center for Advanced Studies at the Norwegian Academy of Sciences, that has been my third home for a semester: Jonas Vestby, Andreas Tollefsen, Sebastian Schutte, Håvard Strand, Céline Cunen and Gudmund Horn Hermansen. Special mention goes to Nils Hjort, who helped, after a short, maybe ten-minute conversation, solve a riddle that plagued Paper 2 for over three years.

In fact, this long path means that some of these thanks go to people that I've met long before my PhD, and even before my time in Uppsala. I thank Liviu Cotfas, for not only being a good friend, but also for being responsible for me writing my first peer-reviewed academic research (I am astounded that thing has 12 citations!). I also thank Florin Feşnic for encouraging me to do research, as a bachelor student, and showing us that research can be fun. And I thank Simona Popa, for teaching me how to program a computer in what must have been four years of patience and dedication. And, of course, I have to thank the ones that are responsible for starting me on my quest for knowledge

– my four grandparents. Three of them are no longer with us: my grandpa, Col. Vasile Gârjău – for giving me his thirst for knowledge (and sharing his knowledge of all stuff military and railway related), my grandpa Gheorghe Croicu – for being open to discuss politics (and even political theory) with a child, and my grandma Marilena Gârjău, for teaching me (mostly against my will) formal grammar and linguistics, which proved to be useful many years after passing my school exams. And, to my grandma Elena, who will read this, I shall write in a language that she can read and understand (albeit in a font too small for her 94-year-old eyes): *Mamaie, mulțumesc pentru bunătate, căldură, calm si speranță. Mulțumesc pentru promisiunea ținută şi pentru nucuşoare!*.

Finally, in the honor spot, comes the most important person in my life, and the most important person for this dissertation. And this wonderful person is Andra, my soulmate, my love, my partner during the past 14 years. Andra is the only reason this dissertation could ever come into existence, and the only reason that you are reading it today. She was and is my steadfast rock and lighthouse in what proved to be a very stormy sea, she was the one that picked up the pieces when I was in pieces, she was the one that wiped my tears (sometimes, and not even rarely, literally), and the one that helped me push through what was, at times, a dark abyss. Not once, not twice, but countless times. Throughout this process of frequent chaos and sometimes darkness and misery, she brought me up from the deepest and darkest circles of hell, and lifted me up from what felt like utter despair. She encouraged me to press on, and gave me the energy, spirit and light to carry on. And stood by my side with the steadfastness and determination of a dragon, even when I did not deserve it. And, on many occasions, when I was stuck, she helped with incredibly smart ideas, suggestions and avenues I did not even imagine existed (it does help that her skills and knowledge parallel and complement mine). I will forever, forever, forever, be grateful to her for all this, for everything, and I do promise not to do another PhD ever again. *Te iubesc!*

Mihai Croicu
12 December 2024

1.1 Funding Receipts

This dissertation was only possible because of the following research grants that provided generous funding, access to high performance computation and travel:

- The European Research Council (ERC) Advanced Grants 694640 – ViEWS and 101055176 – ANTICIPATE.
- Riksbankens Jubileumsfond (RJ) Program Grant M21 –0002 – Societies at Risk.
- the Research Council of Norway through the Uncertainty of Forecasting Fatalities in Armed Conflict (UFFAC) Grant 334977.
- The National Academic Infrastructure for Supercomputing in Sweden (NAISS) grant 2023/5–551 for extensive high performance cluster and GPU access.
- Anna Maria Lundins Stipendiefond at Smålands Nation grant numbers AMh2018–0134, AMh2018–0084, AMh2021–0090, AMv2022–0159 and AMh2023–0083 for the generous travel funds.

2. Introduction

Over the past decade, conflict forecasting has undergone a remarkable transformation, marked by significant advances in both the scope of research and predictive performance. For years, the field was characterized by small-scale, isolated efforts with low and static predictive performance (Schrodt, 2006; Hegre et al., 2017; Halterman et al., 2023). However, substantial efforts such as the Violence and Impacts Early-Warning System (VIEWS), of which this dissertation is part, CoupCast, ACLED Volatility Risk Index/CAST, Conflict Forecast etc.(Halterman et al., 2023; Rød et al., 2024; Hegre et al., 2021b, 2017, 2022b, 2024; Mueller et al., 2024) have led to a revolution in both the scope of forecasting in conflict research and in terms of overall predictive performance.

The success has been recognized by practitioners from various United Nations (UN) agencies (Blocher et al., 2022), the European Union (Halkia et al., 2017) etc. to attempt to replicate them, and develop their own, practitioner-oriented, operational systems. This represents one of the rare instances where academic innovations have been replicated and implemented in the policy sphere.

These advances have been made possible by leveraging synergies across three key developments in the construction of large, tightly integrated forecasting systems. First and foremost, *a revolution in data* has transformed the field – moving from the traditional resolution of country-year to that of highly disaggregated, global collections of individual *battle events*, geo-referenced down to village level and day level. Notable examples include datasets such as the Uppsala Conflict Data Program's Geo-referenced Events Dataset (UCDP GED) (Sundberg and Melander, 2013; Croicu and Sundberg, 2018) and the Armed Conflict Event and Location Dataset (ACLED) (Raleigh et al., 2010). Second, there was *a revolution in methods*, away from the classical reliance on poorly performing and highly constraining generalized linear methods. Initially, this lead to the adoption of classical (shallow) machine-learning methods such as Random Forests (Hegre et al., 2019), Gradient Boosted Trees (Hegre et al., 2021a) or Dynamic Time Warping (Chadefaux, 2023). This was followed later by the adoption of deep neural networks as exemplified by von der Maase (2022a), Malone (2022) and D'Orazio and Lin (2022). Third, a paradigm shift has occurred in *the integration of domain knowledge and theory into forecasting*. Debates such as those between Beger et al. (2021) and Blair and Sambanis (2021) have clarified the role of theoretical insights in prediction, while works such as Colaresi and Mahmood (2017) have developed explicit protocols for including theory and domain knowledge in forecasting systems.

An archetype of such a forecasting system has emerged where event data is collected by a separate project (e.g., UCDP or ACLED). This data is aggregated to subnational or national level, and used in conjunction with static, long-term data (such as economic or social indicators) to create a forecasting ensemble that predicts targets derived from event data sources (Hegre et al., 2021a; Rød et al., 2024; Hegre et al., 2021b).

While these advances have been enormously successful, significant challenges remain. First, while models excel at predicting absolute, static risks of conflict and capturing the spatio-temporal patterns and diffusion of conflict once it has erupted, they remain limited in their ability to predict conflict dynamics – onsets, escalations, de-escalations, and terminations (Mueller and Rauh, 2022a; Hegre et al., 2021b; Vesco et al., 2022).

Second, all forecasting efforts rely heavily on *secondary sources*, typically the above-mentioned event data. Producing these datasets requires hundreds of years of human annotation work for each year of data, yet they yield only a limited set of features. Since these events are almost exclusively mined from extensive text-based news corpora, this process effectively serves as an extraordinarily costly form of feature engineering. Moreover, since these datasets usually precede the forecasting tasks they inform, this approach is inherently inefficient.

Third, apart from some very limited cutting-edge exceptions (Hegre et al., 2024), all forecasting efforts are incapable of handling uncertainty, in both incoming data streams and predictions.

Fourth, to some extent, forecasting methodologies have begun to lag behind the state-of-the art in machine-learning, even in cutting-edge areas like the VIEWS prediction competitions (Vesco et al., 2022; Hegre et al., 2024).

Finally, forecasting has had a relatively limited impact on theory-building and theory-evaluation, though this is beginning to change. Notable exceptions can be found in the rapidly evolving climate and conflict literature (Buhaug et al., 2021; Vesco et al., 2021).

In total, the six papers in this dissertation aim to address the limitations mentioned above and advance the state of the art in the following four fundamental directions:

1. First, by contributing to our methodological understanding of conflict event data at a fundamental level. The dissertation explores the hard limits that these kinds of data inherently have in terms of both spatio-temporal resolution and feature-space richness. The dissertation then proposes methodological advances to address and mitigate some of the most salient of these inherent limitations.
2. Second, the dissertation explores the interplay between event data and conflict forecasting. Beyond the established understanding that conflict event data accounts for up to 90% of the predictive power

in state-of-the-art forecasting models, the dissertation investigates the bidirectional relationship between these elements, exploring how event data influences predictive performance while also leveraging forecasting techniques to improve our understanding of event data itself.

3. Third, the dissertation revisits some of the fundamentals of event data collection, specifically the newswire corpora that underpin such efforts. It explores how to apply modern natural language processing (NLP) methods to enrich event data break and to reduce the dependency of conflict forecasting models on event data altogether.
4. Finally, at a more domain-specific level, the dissertation explores what we know about micro-dynamics of the use of violence, extending our understanding of escalatory and de-escalatory patterns of conflict within civil wars.

2.1 Organization of the thesis

2.1.1 Research questions

This thesis explores the four interlinked themes described above in a comprehensive summary followed by six papers.

The comprehensive summary presents the state of the art in conflict forecasting, conflict event data, and applied natural language processing (NLP) as used in conflict research. The summary dedicates one theoretical chapter to each of these core concepts, in which, I define, introduce and discuss the main theoretical concepts underlying the thesis. This provides the reader with a solid foundation on which to base the understanding of the six papers. The summary is then followed by a set of conclusions, and, finally by the papers themselves.

Paper I deals with **time** and **temporal uncertainties**. It asks two rather simple questions: *How do delays in data availability affect forecasting?* and *What are the substantive effects of collecting conflict event data closer to the real time?* Paper I additionally provides the reader with a comprehensive technical background on forecasting in conflict research.

Paper II deals with **space** and **spatial uncertainties**. The question it asks is: *How can we mitigate the inherent spatial uncertainty and fuzziness in high-resolution event data to improve our forecasts?* To address this issue, paper II formalizes an inherent spatial problem in event data – the "known geographic imprecision" (KGI) problem, i.e. the inconsistent spatial resolution at which the data resolves, and develops a multiple-imputation derived statistical tool to address this.

Paper III deals with **feature space**. It asks, given the extreme sparsity of features inherent to conflict events, *How can we use modern natural language processing in order to mine additional information for event data?* Large, annotated text corpora are inherent "by-products" of event data collections, but are extremely difficult to use to mine new information. The paper takes advantage of two hand-annotated datasets (one on electoral and one on religious violence) alongside the corpora used by one of the largest event datasets (UCDP GED) to propose and test a low-human intensity yet computationally accessible active deep learning approach to mine new features of interest from the data.

Paper IV deals with **substantive knowledge**. It combines the findings from the previous three papers and shows how such newly attained richness can be leveraged to revisit a substantive question: *What explains conflict escalation during civil war?* Using data mined in Paper III it looks at provocative attacks against high symbolic value targets (religious institutions and transportation infrastructure) and their relationship to sudden surges in violence.

Paper V deals with **the process of forecasting**. It asks two questions: *Can we build a model that directly forecasts from text as input data, without the intermediate step of (manual) data collection, extraction, and curation?* and *Can we develop a model operating at the actor level with predictive power at that level, thus proving that we can extract actor-level features?* This paper thus challenges the prevailing paradigm that event data is a prerequisite for forecasting. Instead, it introduces and evaluates a machine learning system that fine-tunes large language models to directly predict conflict escalation and dynamics from the raw corpora underlying event data collection.

Finally, **Paper VI** deals with **the frontier of event data**. This paper asks whether *collecting event data on conflict processes can be extended from violent processes to non-violent processes*, and evaluates the trade-offs, pitfalls, and limitations of doing so. The paper identifies a key frontier in event data: the specialization of methods for extracting data specific to lethal interactions. It leverages two ground-truth, gold-standard datasets to ascertain whether these methods can be applied to less-than-lethal events.

2.1.2 Contributions made by the thesis

Contributions are made to the four interlinked themes in different ways:

In terms of *forecasting* – **Paper V** is a pure forecasting paper, proposing a new predictive model for escalation and de-escalation processes within ongoing armed conflicts. **Papers II** and **III** use forecasting techniques together with Monte-Carlo simulation as tools to validate the applicability of recently introduced or adopted methodological approaches. **Paper I** evaluates forecasting system improvements in terms of a proposed data collection methodology. Further, **Paper I** has become the basis of the approach underpinning the connection between conflict data and VIEWS – the largest conflict forecasting system in production globally. **Paper IV** uses forecasting to further test the external validity of theoretical claims explored through classical inferential identification strategies. Forecasting methods vary across a full spectrum, evolving as the field has evolved, from direct multi-step forecasting with random forests in **Paper I**, through boosted tree approaches on ablated and simulated data in **Paper II**, to encoder-based large language models-based approaches in **Paper V** and neural network recursive dynamic time-series methods in **Paper IV**.

Methodologically, **Paper II** introduces a novel spatial multiple-imputation framework to improve the spatial resolution of data. **Paper III** adapts and tests a large language model-based active learning approach from other domains to conflict data. **Paper IV** explores and pioneers the use of a state-of-the-art sequence-to-sequence forecasting method – the multivariate extension of the Neural Basis Expansion Analysis for Interpretable Time Series Forecasting (N-BEATS). **Paper V** makes use of state-of-the-art large language models (such as DeBERTa and Mistral) coupled with modern retrieval-augmented generation techniques (RAG) for the first time in the field of conflict research.

From a *conflict data* perspective, the papers not only enhance our understanding of event data itself, but also contribute significantly to data collection by introducing novel datasets: **Papers III and V** introduce a global, human curated, event dataset of attacks against religious targets (buildings and officials) as well as a machine-collected, human-in-the-loop, active-learning based dataset of attacks on public transport infrastructure (airports, aircraft, train stations, trains, buses, bus stops, etc.). **Paper VI** provides and validates a novel, media-based dataset of kidnappings in Nepal. **Paper II** provides a dataset of multiple spatial imputations of the UCDP GED, with corrections for the Known Geographic Imprecision (KGI) problem.

Finally, in terms of *theoretically substantive contributions*, all the papers advance the frontier of our theoretical knowledge of event-data and its limitations. Beyond this, underlying **Papers IV** and **V** delve into the empirically underexplored yet often theorized phenomenon of conflict dynamics – specifically, the escalatory and de-escalatory phases within civil conflict. These studies aim to answer questions about how the strategic selection of targets influences the dynamics of civil conflict. **Papers I** and **VI** explore media bias and media coverage during periods of fog of war.

Next follow some theoretical considerations that form the basis of the dissertation and are required for understanding these contributions in context.

2.2 Forecasting armed conflict

Forecasting – in the sense of using statistical and computational methods to infer potential future behavior – has roots that extend to the dawn of quantitative inquiry in the study of armed conflict and war. Pioneering work, such as Richardson (1960), even predate the existence of meaningful computational resources required for such inquiry, being contemporary with the earliest efforts at systematic conflict data collection. Indeed, as early as 1973, forecasting has been described as the most important task in the study of armed conflict (Singer, 1973).

Even from these early pioneering days, forecasting has been viewed as having a dual role. On one hand, it is a fundamental tool for informing policy decisions and determining risks and needs for intervention; on the other hand, it is an essential and unique litmus test to be used when evaluating our understanding of conflict processes (Singer, 1973; Rummel, 1969; Bueno de Mesquita, 1981, 1984; Bueno de Mesquita et al., 1985).

This duality has drawn strong parallels to the fields of meteorology and climatology, with comparisons made from as early as 1969 (Rummel, 1969) to as recently as 2024 (Hegre et al., 2024). Like the atmosphere and its climate, conflict is an enormously complex system that resists full causal understanding, characterized by intricate feedback loops, multiplicative interactions, and multiple emergent properties at the system level (Gallo, 2013). Similar to meteorology, this complexity, where the whole is not the sum of its parts, makes it difficult to use much of our scientific toolkit (Gallo, 2013). This is not surprising, since much of our inferential toolkit, whether observational or experimental, relies on a divide-and-conquer strategy – isolating individual components of the system to identify cause-effect relationships while eliminating and controlling for confounding factors (Lyons, 2006).

This makes forecasting an essential scientific tool. Instead of isolating conflict into its small components and attempting to recreate them in an experimental, lab-like environment (which is impossible), we can leverage theoretical models and empirical data to build a systematic model of conflict. This model can then be used to forecast future events. The predictive power of such models – and the improvements made over time – serve as evidence that our theoretical understanding aligns with real-world systems, thus signifying genuine knowledge (Colaresi and Mahmood, 2017).

Examples of this approach, where forecasting is used to validate theory, can be found in Hegre et al. (2017), which evaluates theories on conflict traps – complex mechanisms underlying armed conflicts – and in Buhaug et al. (2021), which explores the intricate theoretical relationships between local economic conditions, climate-induced shocks, and violent civil conflict. Furthermore, forecasting as knowledge discovery is not unique to climate science (Alexander, 2016), but can be found in other fields dealing with high complexity systems, such as theoretical machine-learning and AI research (Bengio et al., 2017).

2.2.1 A short history - forecasting as a blend of theory, data and method

Forecasting has evolved in lock-step with our theoretical understanding of conflict processes, our computational power and our ability to collect and organize systematic data.

Early approaches were based on theories of systemic capabilities (mostly at country level, such as the production of steel, wheat, or arms) as determinants of conflict. From a data perspective, these approaches relied on coarse, country-level indicators of capabilities, military positioning, and power (Singer, 1972; Richardson, 1960). Methodologically, given the primitive computational resources of the time, even when considering the super-computers of the time, simulation techniques were employed, primarily built around simple partial differential equations, a hallmark of the war-gaming approaches used during that era (Abt and Gorden, 1969; Singer, 1973).

Innovations in our theoretical understanding of the micro-foundations of conflict during the 1970s and 1980s, with seminal works such as Gurr (1970) or Tilly (1978), have led to methodological improvements in forecasting models. These include game-theoretical approaches, with some disaggregation of e.g. leadership-level preferences and payoffs in the forecasting models (Bueno de Mesquita, 1981, 1984; Bueno de Mesquita et al., 1985).

By the mid-1990s, significant improvements took place in the methodological and data collection domains: work such as Gurr (1995); Gurr et al. (1999); Goldstone et al. (2000, 2010) have permitted, through projects such as the State Failure Task Force / Political Instability Task Force (PITF) and Minorities at Risk (MaR), the collection of disaggregated microdata on various aspects of state collapse, ethno-political instability, and risk factors. These efforts were further complemented by work such as Schrodt et al. (1994); Schrodt and Hall (2006)'s KEDS – an early natural-language processing (NLP) attempt to extract information out of textual data without human assistance. Computational improvements have permitted more complex statistical models, with new methodological avenues being attempted, some ahead of their time, such as neural networks Schrodt (1991); King and Zeng (2001), tree-based methods (Schrodt, 1990) and vector autoregressive-based methods (Brandt et al., 2014). However, due to relatively poor performance (King and Zeng, 2001; Halterman et al., 2023), this generation of conflict forecasts did not result in major breakthroughs, to the point where the entire endeavor of forecasting was sidelined for classical observational inferential work using simple statistical models (such as logistic regression) and restricted sets of indicators.

2.2.2 Briefly on the state-of-the-art

The renaissance of forecasting originates in the crisis of traditional observational inferential methods – with many foundational findings generated through null hypothesis testing being found spurious when subjected to a more rigorous testing routines such as out-of-sample evaluations and sensitivity analyses (Ward et al., 2010; Hegre and Sambanis, 2006; Rød et al., 2020).

This has led to a reassessment of forecasting as a tool for furthering knowledge; a reassessment that has been helped by, and in terms led to, significant improvement in three domains: data, modelling and theory/specification.

The most significant advancement lies in the realm of data: the event data revolution – the availability, for the first time, of highly disaggregated sets of microdata on armed conflict and associated phenomena. Datasets such as the Uppsala Conflict Data Program's Georeferenced Event Dataset (GED) (Sundberg and Melander, 2013; Croicu and Sundberg, 2018), Armed Conflict Location and Event Dataset (ACLED) (Raleigh et al., 2010), the Mass Movement in Autocracies Database (MMAD) (Weidmann and Rød, 2019), Social Conflict in Africa Dataset (SCAD) (Salehyan et al., 2012) and Lockheed Martin's Integrated Crisis Early Warning System (ICEWS) (Shilliday et al., 2012) provide millions of data-points, coded at extremely high resolution – typically individual battles coded for individual days and individual geo-referenced villages and towns, on a global scale, with very long time-series. These datasets have enabled the field to move beyond the extremely rudimentary, high-level country-year datasets that previously dominated, providing a significant advance through meticulous manual data collection and curation. This progress has also addressed the substantial noise issues of early NLP-based machine-extracted data, such as the previously discussed KEDS. Today, these highly detailed datasets form the foundation for most of the field's prediction targets (Vesco et al., 2022; Hegre et al., 2021a, 2019), as well as constitute the main feature sets for forecasting models. In fact, evaluations indicate that of conflict history features derived from event data account for up to 90% of the predictive power in most contemporary models (Hegre et al., 2024).

The second advance has been methodological. Classical statistical approaches, usually centered around logistic regression, with its highly constraining functional form, have been replaced by machine-learning algorithms. These broadly paralleled the evolution of machine-learning in general, transitioning from very simple models to extremely complex architectures. As a result, a wide array of higher-performance approaches emerged: tree-based ensembles such as random forests (Hegre et al., 2019), modern gradient boosting trees (Hegre et al., 2021a), and Markov models (Randahl and Vegelius, 2022). More recently, this has expanded to include various neural network architectures (such as long short-term memory, convolutional and graph neural-networks) as well as advanced geometric approaches (Vesco et al., 2022; Malone, 2022; D'Orazio and Lin, 2022; Brandt et al., 2022; von der Maase, 2022a; Chadefaux, 2023). Ensembling – forming mixtures of topical and methodological "experts" (in terms of models) – became generalized, sometimes using very advanced algorithms, such as genetic optimization (Hegre et al., 2021a). This paralleled the developments in computer science, where mixtures of experts (MoE) have become the state-of-the-art.

Finally, advances in our understanding of theory and model specification have also shaped the current generation of conflict prediction models. Colaresi and Mahmood (2017) introduced a principled method of handling domain knowledge in model specification through an iterative, Bayesian approach, with a goal of avoiding overfitting and biased prediction. Fariss and Jones (2018)'s theoretically informed functional forms and regularization added to this work. The debate between Beger et al. (2021) and Blair and Sambanis (2021) added further to this advancement, by defining "hard" and "soft" theory and showing mechanisms by which both have a role in informing prediction.

The result of these advances have lead to an unprecedented improvement in predictive performance, as well as the mainstreaming of the domain as a (if not *the*) major endeavor in the field of conflict research.

2.2.3 The frontier

Methodologically, while the field has roughly followed advances in applied machine-learning, it has, surprisingly, lagged behind in one of the most obvious directions of advancement suggested by that literature. Almost all current-day forecasting approaches in conflict research are *shallow learners* (Bengio et al., 2017). Even the most advanced approaches in the field continue to depend almost entirely on manual feature engineering, reflecting an underutilization of more sophisticated methodologies (Vesco et al., 2022).

Even when *deep learning* approaches are used, they are either adapted to work with these engineered features – event data or spatio-temporal rasterized aggregations of them (von der Maase, 2022a; Häffner et al., 2023; D'Orazio and Lin, 2022; Brandt et al., 2022).[1]

This is even more surprising given the origins of the conflict history data that provides most of these models' predictive performance. Conflict history data is derived from battle event data, which itself is derived from manually curated textual corpora. These corpora are usually compiled from a combination of newswire, newspaper, and boots-on-the-ground sources.

However, until now, very little effort has been made to leverage the richer signals from these non-curated corpora of textual data directly for forecasting armed conflict within a deep learning framework. Indeed, all previous attempts at a fully automatic pipeline have attempted to extract features automatically, either as events or dictionaries of events (Häffner et al., 2023; Skorupa Parolin et al., 2020, 2021) or as unsupervised (classical) topic models (Mueller and Rauh, 2022b; Mueller et al., 2024; Chadefaux, 2017, 2023) resulting in features such as event counts, proportions of topics or topic fre-

[1] The aggregation is most commonly done to PRIO-GRID-month, a 0.5 decimal degree squared geographic grid (Tollefsen et al., 2012).

quencies for shallow learners (such as random forests or logistic classification models).

Some potential explanations for the low prevalence of deep learning approaches can be found. First and foremost, much of the methodological improvements in the field of conflict research, especially before the forecasting renaissance, came from causal inference[2] approaches (Fariss and Jones, 2018) and from spatial econometrics (Tollefsen et al., 2012). Both of these approaches rely on relatively simple parametric statistics coupled with complex pre- and post-estimation manipulations such as, e.g., survey experiments, randomization, matching etc. This resulted in a significant skill gap throughout the field, that only now begins to be addressed. Second, there has been significant difficulty in obtaining and licensing the raw text data for conflict-related machine-learning: access has only become feasible in the past couple of years, and only on a relatively small scale, at extremely high cost. Third, there has been a focus on model parsimony and interpretability, based on approaches common in observational research and game-theoretical research. This is, in many respects, the polar opposite of the approach taken by most modern machine-learning approaches, which rely on e.g. automatic feature extraction and feature engineering (Bengio et al., 2017).

What is clear is that interpretability is not a direct hinder to adoption. Tools for model-agnostic interpretation of results and estimation of (pseudo-)causal effects have been successfully used in the field with otherwise normally uninterpretable shallow-learning models, showing that complex, noninterpretable models are not a hinder in themselves. Indeed, examples of such model-agnostic interpretation include the use of discrimination plots, ablation tests, and causal (generalized) random forests (Colaresi and Mahmood, 2017; Dietrich and Eck, 2020; Hegre et al., 2019). These provide the field with similar tools to those that can be used to interpret deep learning models, thus showing that deep learning models per se are not problematic.

However, irrespective of the cause for this gap, it still exists, with the reliance on shallow machine-learning reducing the overall predictive power of models due to limited feature availability and low model expressivity.

The dissertation addresses this gap in **Paper V**, by exploring, adjusting, and fine-tuning modern, Transformer-based, large language models on the annotated text corpora that underlie the event data to directly forecast escalatory and deescalatory patterns of conflict. Further, **Paper V** makes additional methodological contributions, using *retrieval augmented generation* techniques to improve forecasting beyond classical training/fine-tuning (Lewis et al., 2020). The dissertation further addresses the above-mentioned issue, differently, in **paper III**, this time by preserving the text $\longrightarrow$ feature $\longrightarrow$ forecast paradigm, but instead introducing a human-in-the-loop active-learning

[2] Also known as "inference by design"

approach to use these corpora intelligently to mine rich features orthogonal to event data for escalatory forecasting purposes.

Turning to forecasting methodologies, while multiple innovative approaches have been explored in the past – such as dynamic recursive simulation (Hegre et al., 2017), the field has settled on an expensive yet rather simple approach: direct multistep forecasting (known in the field as step-shifting). In this approach, one model is trained for each prediction step, i.e. temporal distance between data (training) horizon and forecasting target (Hegre et al., 2021a). Although this approach is still modern and highly relevant, the last few years have been one of extreme advancement in the field of machine-learning for time-series forecasting approaches, especially in financial applications (Jeffrey et al., 2023). **Paper IV** explores these methods, and implements a modern forecasting approach based on a sequence-to-sequence multi-variate neural network, an adapted version of Neural Basis Expansion Analysis for Interpretable Time Series forecasting (N-BEATS) (Oreshkin et al., 2019).

Now turning to substantive advancements, the field has been extremely successful at predicting the spatio-temporal mechanics of conflict (Hegre et al., 2017, 2021a; Vesco et al., 2022). In many respects, this success has addressed the critiques of Cederman and Weidmann (2017), especially regarding conflict traps and spillover effects. However, the field has been extremely poor at predicting the dynamics of conflict – escalatory and de-escalatory patterns, as well as onsets and terminations of conflict (Mueller and Rauh, 2022a; Halterman et al., 2023; Vesco et al., 2022; Hegre et al., 2022b). **Papers IV** and **V** tackle these challenges directly, by explicitly modelling conflict dynamics – change – instead of merely modelling incidence or prevalence of conflict. Further, **paper IV** takes a theory-testing approach, examining shifts in conflict dynamics before and after attacks on symbolically important locations (religious sites and transportation hubs). **Paper V** extends this work by proposing a novel approach to measuring escalation and de-escalation, using Gaussian process smoothing to mitigate measurement challenges and enhance predictive performance.

However, what merits the most attention when attempting to advance the frontier may be a look back in the rearview mirror – revisiting *event data*, the most predictive feature set we rely on in the field. These datasets were created during a period of minimal forecasting activity and were designed with completely different goals and purposes, mainly to study the local determinants of conflict (Sundberg and Melander, 2013; Raleigh et al., 2010)). Revisiting this foundational data and exploring ways to improve and adapt it specifically for forecasting purposes could represent the most promising avenue for advancing the field.

The next chapter takes a deep dive into that world of data, systematizing knowledge, identifying the frontiers where event data and forecasting meet unresolved challenges and exploring improvements at those frontiers.

2.3 Disentangling conflict event data

2.3.1 Attempting a general definition of event data

To improve our efforts at forecasting armed conflict, a prime avenue of inquiry is, therefore, to revisit and take a more in-depth look at disaggregated conflict event data, as it is the main input that underlies our forecasting efforts. Indeed, most large scale forecasting efforts, such as VIEWS (Hegre et al., 2019, 2021a), ACLED CAST (Raleigh et al., 2010) or ICEWS (Shilliday et al., 2012) rely on such data twice – *first*, to define and measure the target outcome and *second*, as the most predictive feature set in the form of past conflict history (Hegre et al., 2024).

An in-depth exploration of the data generation process (DGP) of these datasets will thus shed light on both the limitations and biases of our forecasting efforts, as well as provide us with an understanding and the tools to mitigate these issues, allowing for better forecasting performance.

While there is no unified definition of what event data is, a tentative one based on the efforts of the multitude of projects in the field is that **event data is any kind of microdata collection[3] that codifies information on actions of political contention carried out by a political group (usually referred to as an actor) at a temporal resolution that is more granular than a calendar week and at a spatial resolution of at least that of a town/village** (Davies et al., 2024; Steinert-Threlkeld, 2019; Donnay et al., 2019; Weidmann and Schutte, 2017; Croicu and Sundberg, 2018; Weidmann, 2016, 2015; Sundberg and Melander, 2013; Raleigh et al., 2010, 2023). Usually ranging in the hundreds of thousands of events, these projects offer microdata covering decades of conflictual and politically contentious activity, frequently with a global scope. Examples include the UCDP Geo-referenced Event Dataset (GED) (Croicu and Sundberg, 2018), the Armed Conflict Location and Event Dataset (ACLED) (Raleigh et al., 2010), the Global Terrorism Database (GTD) (LaFree and Dugan, 2007), the Social Conflict in Africa Database (SCAD) (Salehyan et al., 2012), the Mass Movements in Autocracies Dataset (MMAD) (Weidmann and Rød, 2019), as well as Phoenix/Petrarch (Norris et al., 2017; Beieler, 2016), ICEWS/JABARI (Shilliday et al., 2012) or the Kansas Event Data System/TABARI (Schrodt and Van Brackle, 2012; Schrodt, 2012).[4]

Beyond this imprecise definition, one thing shared by all event data projects is the underlying source material used to collect and curate these large event

[3] in the meaning of Winkelmann and Boes (2006).

[4] The purview of this dissertation, at the intersection with forecasting, unfortunately does not allow for a detailed history of the origin and development of event data, and cannot even hope to do justice to what is one of the main fundamental revolutions in the scientific study of conflict. Refer to Öberg and Sollenberg (2011) and Schrodt (2012) for an in depth treatment of the birth and history of event data, as presented from by the two main ontological "camps" described below.

datasets.[5] These source materials are always large text corpora, usually of a very similar composition. Such corpora are comprised of a combination of newswire and newspaper articles sourced from global news agencies (*Reuters, Agence France-Presse, Associated Press, Xinhua, Al-Jazeera, TASS*), local news agencies (such as *Tolo News* in Afghanistan, or *Suspilne* in Ukraine), large newspapers of record, whether global, national or sub-national (e.g. *New York Times, The Hindu, Al-Ahram, Dawn* etc.), and an assortment of local news sources (local newspapers, local TV, and radio, e.g. *Radio Okapi* in the Democratic Republic of the Congo) (Croicu and Sundberg, 2018; Croicu and Weidmann, 2015; Davies et al., 2024; Öberg and Sollenberg, 2011; Raleigh et al., 2010). For consistency and convenience, these are frequently collected through aggregators such as LexisNexis or Factiva. They are often supplemented by data extracted from more local sources, such as reports produced by organizations on the ground (the International Crisis Group, UNOCHA, etc.) or triangulated reports presented by warring parties, and can extend to social media accounts, etc. (Croicu and Sundberg, 2018; Öberg and Sollenberg, 2011; Raleigh et al., 2010). They result in vast and highly structured text corpora, typically in the hundreds of thousands or millions of articles, that are then mined for structured information. News-wire text, however, dominates all collections, from around 60–70% for UCDP GED and ACLED to 100% for datasets like ICEWS.

Beyond text, very few inroads leading to practical results have been made to date that yield practical, production-grade results. Steinert-Threlkeld (2019); Steinert-Threlkeld et al. (2022); Lu and Pan (2022) and von der Maase (2022c) are exploring and developing methodologies for event data extracted from images and videos. However, these approaches are in their infancy, at the stage of early, exploratory and methodological work. Further, except for von der Maase (2022c), most focus on protest events rather than conflict events. As a result, even when these approaches mature, adapting them to a conflict context will require significant transfer learning and piloting. Given these limitations, non-text signals (images, videos, etc.) will not be considered in this discussion.

[5] A note on terminology: Most event data collection efforts will use the term "coder" for the person(s) and "coding" for activity of collecting and curating information from source material into event data collections. This term is highly specific to the field, and does not seem to be used broadly. To avoid terminological confusion with the more common meaning of coder as a software programmer, this dissertation and the articles therein will employ the term annotator for this person and annotation for the activity. This is the same terminology that is used in computer science, natural language processing, machine-learning, information retrieval – fields that this dissertation is interacting with significantly.

2.3.2 Two worlds of event data

Beyond the definition and underlying news-heavy corpora, conflict event datasets vary significantly in their structure, scope, and methodologies. A key distinction lies in the ontology of the data, which is determined by the resolution at which conflict events are recorded (Croicu and Sundberg, 2018).[6] Broadly, this results in two types of datasets, based on what resolution the event itself is coded at:

- the *"incident"* level dataset. These define the event as a real-life chain of actions done by a group that results in a given outcome, such as a battle, a protest, an attack, a riot etc. Events are characterized by a location (e.g. a city or a street), actor(s) and date(Croicu and Sundberg, 2018). Datasets such as UCDP GED, ACLED, MMAD, SCAD, GTD etc. fall in this category.[7] Incident-level datasets are generally manually collected and manually curated, typically by combining an assortment of local and global sources, including news media, NGO and IGO reporting etc., with limited automation in the process (Öberg and Sollenberg, 2011; Croicu and Weidmann, 2015).
- the *"Atomic"* or "machine-coded" (Miller et al., 2022) event datasets, examples of these including Multi-COPED (Skorupa Parolin et al., 2022a), Pheonix, TABARI, KEDS, etc. These define the event as being a unique, single, atomized action carried out by an actor at a point in time. These actions are usually defined as a noun-verb pair, with both present in relatively concise dictionaries – the most common being CAMEO (e.g. "protesters demonstrate", "the United States President declares", "rebels attack", "forces retreat") (Gerner et al., 2002). Most of these datasets are collected in an unsupervised machine-learning approach. This is done either by using traditional NLP methods – generally shallow parsing based on simple parts-of-speech tagging (Schrodt, 2012; Beieler, 2016) or via (frequently unsupervised) classification based on various contemporary neural network architectures (such as e.g.Skorupa Parolin et al. (2020, 2021, 2022b)), sometimes using a specialized pre-trained language model such as ConfliBERT (Hu et al., 2022). These datasets are always automatically collected (machine-coded) – with the most recent generations even relying on automated dictionary generation, essentially removing the last vestige of human curation from the loop (Häffner et al., 2023; Skorupa Parolin et al., 2021).

Essentially, *incident* level event datasets can be seen, at a conceptual level, as an aggregation, distillation and filtering of the *atomic* level datasets, in the

[6] By ontology, I mean the way these datasets define their subject of data collection and thus build a world of events.

[7] As a note, while MMAD in its default configuration is an incident-level dataset, it also aims to provide a more disaggregated approach, where the observation is the event-report, i.e. a single piece of news pertaining to the event.

same way as the traditional *country-year* or *conflict-year* datasets such as the UCDP/PRIO Armed Conflict dataset are a distillation of *incident* datasets. A synthetic example helps illustrate this distinction: Consider a battle between two groups, A and B, occurring along a specific segment of the frontline (locality X). During this battle, Group A uses a roadside bomb to lure Group B into an ambush involving small-arms fire. Group B immediately responds by calling in air support, which bombs Group A's immediate rear positions in X. Following this escalation, Group A threatens further action, compelling Group B to retreat. An incident-level dataset would record this entire sequence as a single event (assuming it took place within the same locality and timeframe). In contrast, an atomic-level dataset would record multiple distinct events. For example, using CAMEO, the engagement might be recorded as six separate events[8], depending on the capabilities of the automatic annotators. In contrast, the incident-level dataset will have a much richer feature-set, whereas the atomic dataset will rarely include more than the pair of actors, the verb, and sometimes a rough geolocation estimate drawn from an automatic geocoding API targeting an automatic gazetteer.

Despite this stark distinction in resolution, both ontologies are frequently treated as a unified event-data class, despite their conceptual and architectural incompatibilities. This conflation likely stems from the almost complete lack of interaction between the two literatures, especially at the stages of data generation and production, despite both traditions having developed independently over the past three decades.[9] Further, when these two literatures do interact, as they do in forecasting (e.g. Brandt et al. (2022) using "atomic" event data to forecast "incident" level events), confusion and lack of common understanding occurs, both in terminology and substantive meaning.

These issues can usually be resolved with careful clarification – except in instances where the very literature introducing the confusion is tasked with benchmarking, evaluating, and proposing improvements to event data itself. In this case, pooling the two ontologies together and failing to explicitly distinguish between these two families of event data leads to flawed conclusions and problematic research agendas. A recent example is provided by Miller et al. (2022), drawing a conclusion of "inflation-bias" (the presence of spurious copies of one event) as one of the three major problems with event data collections. However, this conclusion was based on an incorrect comparison

[8] *A – 190 attack using military force – B, A – 1833 bomb (roadside) – B, A – 1933 fight with small arms – B, B – 195 aerial bomb – A, B – 1384 threaten conventional attack – A, A – 0874 retreat.*

[9] This is even more clear when considering nobody would group *incident* and *conflict-year* datasets into one bin – despite the distance between these being similar to the distance between the "atomic" and "incident" datasets. Incident and classical conflict-year datasets, however, have evolved in close connection, frequently by the same data providers, unlike atomic and incident level data.

of an atomic dataset with an incident-level dataset, failing to account for their fundamental differences.[10].

To avoid such potential confusion, especially since methodological advances that underpin both ontologies are used throughout this dissertation, both this introduction and the essays will clearly denote the distinction between "atomic" and "incident" level data where appropriate. For brevity, however, when "event data" is used hereon, the "incident level" ontology should be assumed. However, these two worlds of event-data also provide unexplored avenues for synergies. One of these is the methodological divide between manual curation and annotation versus machine-based information extraction. **Papers III** and **IV** explore this divide, proposing a way to bridge it using active learning techniques specifically adapted to what was previously the domain of manual data collection.

Another promising avenue is the application of natural language processing and text-oriented machine-learning, which have been confined to a narrow niche in the atomic event data community.[11] **Paper III** explores such techniques for a novel purpose: augmenting and identifying new features of interest in UCDP GED, a classical incident-level dataset. It does this by combining, in an active learning framework, the best of both worlds, automatic (supervised) machine-learning and human curation, a novelty in the field.

Paper V leverages techniques inspired by and originally designed for atomic event-data generation to directly forecast conflict escalation and de-escalation. It operates under the hypothesis that escalatory signals are obscured when data is aggregated into mainly spatio-temporal features, as is common in incident-level datasets. The results confirm this hypothesis, demonstrating that the raw textual data, combined with a fine-tuned large language model, excels at extracting signals indicative of future shifts in conflict dynamics.

2.3.3 A look at the data itself – feature sparsity and data augmentation

While event data is extremely fine-grained and high-resolution, it is remarkably sparse in its feature space. In other words, the information encoded with

[10] Another similar confusion is present in Demarest and Langer (2022). The authors draw conclusions about biases and uncertainties in event data by conflating various bits-and-pieces of the (distinct and incompatible) data generation process from both families of event data. The resulting ensemble process does not reflect the actual data generating process of either approaches, and some of the evaluations done do not travel at all across the two domains. Examples of this conflation include discussing recall and precision of a pre-filtering stage for MMAD, an incident level dataset in the context of atomic, machine-collected data, and using the results from MMAD to evaluate atomic data performances.

[11] There have been very few and very tentative explorations such as Olsson et al. (2020) outside this domain; such explorations were quickly abandoned.

each event is severely limited. Most incident-level event datasets only contain information related to four aspects:

1. A date or a date range.
2. A location (usually geo-referenced as a point, sometimes with additional information on that point's administrative divisions).
3. An actor or set of actors involved in the incident, usually drawn from a fixed dictionary of actors.
4. A highly limited set of event-level features. The most common is a severity indicator: e.g. fatality estimates in the case of UCDP GED (Croicu and Sundberg, 2018), number of protesters in the case of MMAD (Weidmann and Rød, 2019) and SCAD (Salehyan et al., 2012). Other features frequently included are a set of simple taxonomies that categorize events into a few buckets based on an a priori defined labeling scheme. Examples include the three-category event type used by UCDP,[12] the six event types and twenty-four sub-event types used by ACLED (Raleigh et al., 2010),[13] and the ten-level protest-type codes used by SCAD (Salehyan et al., 2012). These datasets typically allow only coarse filtering of events, constrained to a limited set of dimensions deemed most significant by their creators.

This scarcity of features contrasts sharply with the extensive theoretical understanding of the micro-dynamics of conflict processes. Major elements that explain and forecast conflict dynamics are notably absent from event datasets. It is not surprising that such data lacks comprehensive static information, such as the historical relations between conflict-affected actors and communities (Kalyvas, 2003) or the ethnic dimensions of actors that contribute to conflict (Cederman et al., 2011; Cederman and Vogt, 2017), as this information is static at the resolution at which these datasets operate. What is more striking, however, is the absence of dynamic information – data that captures rapidly evolving and volatile processes. Examples include tactical decisions, such as whether to engage in skirmishes or traditional warfare with territorial control, whether to employ violence against civilians versus positive interaction through rebel governance, or whether to adopt terrorism-like tactics (Kalyvas and Kocher, 2007; Polo and Gleditsch, 2016; Bara et al., 2021; Wood, 2014). Similarly, information about shifts in these tactics as responses to battlefield dynamics is not captured (Wood, 2014). Furthermore, critical information on the presence and strength of rebel governance or government control (Bara et al., 2021; Wood, 2014), as well as localized recruitment processes (Eck, 2014) remains uncollected in current datasets.

A simple explanation for this limitation lies in the immense resource demands of coding event data. Collecting even the four dimensions described

[12] state-based, non-state, and one-sided conflict

[13] The six event types are Battles, Protests, Riots, Strategic developments, Violence against civilians, Explosions/Remote violence.

above requires infrastructure-level investments, often with costs reaching millions of dollars. This challenge is particularly pronounced for incident-level datasets, which rely heavily on labor-intensive manual data curation.[14] One consequence of this state of affairs is for conflict forecasting systems and models, where event data constitutes the vast majority of the dynamic signal used for training. Since this signal is concentrated in three dimensions – space (with events geo-referenced down to village level), time (with events having exact timestamps) and intensity (with fatality figures), it is not surprising that mechanistic spatio-temporal models such as *conflict traps* (Walter, 2004; Collier and Sambanis, 2002; Beck et al., 1998; Collier et al., 2003; Hegre et al., 2017, 2021b) and *conflict diffusion* (Buhaug and Gleditsch, 2008; O'Loughlin et al., 2010; Schutte and Weidmann, 2011; Crost et al., 2015; Bara, 2018) perform best.

One proposed solution to alleviate this almost universal sparsity in terms of available data features is *event data integration*: combining multiple relatively "feature-poor" sources of event data into a unified dataset. This results in a richer feature set (Donnay et al., 2019) by gluing together bits and pieces of the features of interest from multiple sources. However, this is a problematic process, since each of the constituent datasets are built around a different set of definitions for the various conflict phenomena, contain a different set of sources and have different inclusion criteria. Further, event data integration puts significant strain on the researcher in determining whether the data are compatible in any way, requiring significant expertise with each of the dataset. Further, it is always unclear how to evaluate what the result of such integration realistically returns. If using an inclusionary integration strategy (akin to a database full outer join), due to the definitional differences across datasets, it quickly becomes unclear what events are missed, and what features are partially or totally obscured due to relative differences in coding choices (Dawkins, 2021). Conversely, if taking a very strict inclusion criteria, it is unclear what the intersection of these datasets produces in definitional terms. For example, let's assume we are integrating a dataset that includes only violence between ethnic groups with one containing only rebel-government dynamics. In such a case is it reasonable to expect that the resulting dataset will contain a good representation of ethnic civil war? These problems are clearly not unsurmountable, despite some skepticism (Miller et al., 2022),[15] but require significant deep expertise with both the constituent datasets and the integra-

[14]Through their structure, atomic, machine-collected datasets are not able to capture such complex processes through their simple dictionaries of noun-verb phrases. In text corpora, information on recruitment or tactics or strategy would be scattered and present in context and subtext more than in text itself, lending itself difficult to be extracted.

[15]Miller et al. (2022) calls this approach "franken-datasets" even while not assessing either the method or the result in detail.

tion approach employed. Further, even when these are present, such efforts demand significant effort and conscious decisions on top of this expertise.[16]

Another proposed approach, sometimes used in conjunction with the first one, is *spatio-temporal aggregation* of event datasets to a common, coarser, level where definitions and inclusion criteria are less problematic. Typical spatial aggregations result in either a sub-national administrative unit dataset or an arbitrary raster-based spatio-temporal unit dataset. The most typical of these is a spatial fishnet of .5 decimal degrees (55x55 km) such as PRIO-GRID (Tollefsen et al., 2012). Typical temporal aggregations are the week, month, episode or year. Spatio-temporal aggregation has become the standard solution in conflict forecasting efforts, with leading systems like VIEWS and CAST employing it for both prediction targets and historical conflict features (Hegre et al., 2019; Zhukov et al., 2019). While the loss of resolution is not a significant problem, as discussed above,[17] the transformation of event data into a pure spatio-temporal raster can pose major challenges. Chief among these is the loss of non- spatiotemporal features that are essential for understanding conflict diffusion processes, such as actor behaviors, group-level characteristics, and dyadic interactions (Kim et al., 2023). Moreover, event-level uncertainty indicators, which might reveal spatio-temporal errors exceeding those of the aggregate dataset, are also omitted. This omission can compound spatio-temporal biases in the resulting dataset (Cook and Weidmann, 2019, 2022).

A third approach is to use existing event datasets to mine and extract further features of interest. Usually, this is done by resorting to the original text corpora underlying the event data collection efforts, and employing a new set of annotators to extract a new set of features of interest. This is usually extremely resource intensive, often requiring time measured in annotator-years. As a result, successful efforts are rare, and their implementation times can span several years. Examples include the Peacemakers at Risk dataset (Lindberg Bromley, 2018), the Deadly Electoral Conflict dataset (DECO) (Fjelde and Höglund, 2022) and the Cities and Armed Conflict Event Dataset (CACE) (Elfversson, 2021). While this approach provides new features to an event dataset, with perfect compatibility with the original source, the cost and effort required make this approach prohibitively expensive.[18]

[16] Moreover, since the resulting datasets are by definition customs mashups of definitions and concepts, their suitability is dependent on the particular research question that is investigated. This means the process needs to be started over for each research question.

[17] Especially given the nearly identical resolution of PRIO-GRID to the maximum useful resolution identified in Weidmann (2015, 2016).

[18] There is also a fourth approach – carrying out disaggregated subnational analyses on a single case – country or conflict. This approach trades generalizability over time and space for the ability to obtain non-event based local data, usually from surveys, to combine with event data. Such an approach has been very successful in recent times for quite a lot of inferential work. Examples include studying attitudes and values in regard to policing in Afghanistan (Deglow

This dissertation contributes to the debate, offering ways to alleviate feature sparsity in event data. I do this in two different ways: First, **Paper III**, takes inspiration from automatic 'atomic' event data methodologies, active learning and modern advances in NLP, in order to partially automate the third approach while maintaining human curation. This allows for feature extraction and augmentation from existing conflict corpora at a fraction of the traditional cost and time. **Paper IV** demonstrates a direct application of this method – by using the extracted data to investigate the impact of targeted attacks on public spaces during civil conflict on conflict dynamics. Second, **Paper II** introduces a methodology to quantify, measure and mitigate event resolution uncertainty in the spatio-temporal domain. This approach closes the gap on a missing yet essential step when aggregating event data to coarser spatio-temporal resolutions.

These efforts additionally result in three datasets – one on attacks on religious sites and personnel (**Papers III and IV**), one on attacks on transportation infrastructures (**Papers IV**) and one of spatio-temporal imputations of uncertainty (**Paper II**).

2.3.4 Biases and the frontier: beyond battle events

Conflict data is remotely collected observational data, derived from mining text corpora. As such, it inherently reflects the biases and limitations of newswire reporting (Öberg and Sollenberg, 2011; Gohdes and Price, 2013; Croicu and Weidmann, 2015; Croicu and Kreutz, 2017; Weidmann, 2015, 2016; Weidmann and Schutte, 2017). This essentially makes them convenience samples, shaped by what global and local media are able and willing to report. Factors such as access, media interest (e.g., media fatigue), the danger faced by reporters, and inherent biases of news organizations all play a major role in determining what gets captured (Croicu and Kreutz, 2017; Gohdes and Price, 2013; Gohdes and Carey, 2017).

To address these biases, conflict event data collections have been partly designed to mitigate their effects. A key strategy has been to focus on only capturing high-notability incidents – events that would be reported regardless of these biases. This typically includes fatal armed conflicts, large-scale protests, and riots (Öberg and Sollenberg, 2011; Salehyan et al., 2012; Gohdes and Price, 2013; Landman and Gohdes, 2013; Croicu and Sundberg, 2018). Indeed, the underlying design of these collections is guided by assumptions

and Sundberg, 2021), conflict-related migration in Bangladesh (Petrova, 2021), local effects of peacekeeping on conflict resolution in Côte D'Ivoire (Smidt, 2020), trust after terrorist attacks in Nigeria (Harding and Nwokolo, 2024) etc. However, this approach is evidently designed around causal inference research, and not appropriate for any kind of predictive analytics where comparability over time and space is paramount. The dissertation provides, however, some tools to this "single-case study large-N" community as well, e.g. **Paper II** provides an approach for the quantification of uncertainties at local level and estimation of e.g. local urban biases.

and theories suggesting that high-intensity, high-visibility events, such battles and riots, are more reliably covered by the media (Earl et al., 2004; Franzosi, 1987). Consequently, the mediated nature of the data generation process (actual incident $\rightarrow$ news media $\rightarrow$ fatality) has shaped event data design, with an emphasis on high notability events to minimize biases introduced by media reporting (Öberg and Sollenberg, 2011; Gohdes and Price, 2013).

After decades of data collection, it is evident that there is significant under-reporting as well as reporting bias even with these high-notability events – and even with fatal events (Radford, 2022; Gohdes and Price, 2013; Landman and Gohdes, 2013; Weidmann, 2016; Croicu and Kreutz, 2017). This is especially concerning because media bias and under-reporting are not randomly distributed. These issues are most pronounced in areas with low media interest, limited media accessibility, high conflict intensity, and significant repression – precisely the regions that are often of greatest interest in conflict research (Galtung, 1969; Gohdes and Price, 2013; Croicu and Kreutz, 2017; Weidmann, 2015; Weidmann and Schutte, 2017; Weidmann, 2016).

Moreover, biases, patchy coverage, and under-reporting are likely to worsen as traditional news media decline in production and consumption (Franklin, 2008; O'Sullivan et al., 2017; Croicu and Kreutz, 2017). This reduction in international news-gathering resources can lead to the collapse of news monitoring of certain areas or situations of interest, or worse, can enable co-opted, agenda-driven coverage due to ownership changes or takeovers by conflict-aligned actors(Radford, 2022; Baum and Zhukov, 2015, 2019; Zhukov and Baum, 2016). A case in point is BBC Monitoring, the world's largest translator, aggregator and curator of locally produced news articles, and on which most event data collections rely. The service has been defunded significantly in the 2010s, resulting in multiple staff cuts and output reductions(Johnson, 2019). Social media and citizen journalism, once hailed as solutions to these problems, have proven even more problematic. Platforms have become conduits for disinformation, fake news, "active measures", targeted propaganda, and deception, with algorithms being exploited on a wide scale by conflict actors to amplify disinformation, misinformation, bias, and uncertainty (Tansini and Ben-Haim, 2021; Stanescu, 2022; Piazza, 2022). Indeed, after a short period of enthusiasm more than a decade ago (Zeitzoff, 2011), conflict event data projects have largely avoided integrating social media as a data source (Steinert-Threlkeld, 2018). To date, no major academic event dataset contains substantial social-media-derived data. The only exception is ACLED, but only on a very limited scale. Further, Twitter, once the main avenue for such inquiries, is now actively avoided, due to significant biases and platform uncertainties.

Despite growing awareness of biases, under-reporting, and the diminishing quality of underlying corpora, the field has simultaneously expanded its scope to capture events conceptually distant from the original focus on high-notability, high-intensity incidents. These include non-fatal events within armed conflicts, such as sexual violence or kidnappings, and broader "strategic developments". In some cases, attempts have been made to capture events deliberately obscured by combatants, such as troop movements or the establishment of rebel bases, for example in ACLED (Raleigh et al., 2010). This has not been accompanied by a change in methodology or assumptions, leading to a knowledge gap.

The temptation to use such data for forecasting and predictive analytics is strong, since it provides conceptual richness to otherwise feature-poor datasets, especially when moving beyond spatio-temporal domains. For example, small riots can predict larger riots, which in turn predict broader conflict (Rød and Weidmann, 2023), while troop movements, control of roads, and local industrial activity can predict fighting in specific areas (Zhukov, 2012; Zhukov et al., 2019). However, for such data to be useful in large-scale forecasting efforts, it must be consistent and comparable across time, space and context.[19]

Indeed, there are significant claims in the literature, e.g. Raleigh et al. (2023) and Miller et al. (2022) arguing that the definitional clarity and restrictive inclusion criteria used to reduce news media biases may actually hinder our understanding of underlying conflict processes. These authors suggest that current generation event data, for both predictive and inferential analysis, should cast a much wider net, ignoring such mitigation approaches. However, these claims remain untested empirically, with some arguments in favor of such approaches lacking even face validity. For example, simply adding more dots on a map does not necessarily provide a more accurate representation of underlying conflict processes; in fact, it may result in a skewed portrayal, especially if those dots are disproportionately biased towards urban areas. [20]

[19] In the case of a system such as VIEWS, the minimum context window for comparability required is as long as 108 months – since the training, calibration and test partitions are 36 months each; if training data is inconsistent with true forecast data, the whole window will be affected.

[20] Raleigh et al. (2023) counters this point by claiming ACLED uses local informants to drive data collection. This is rarely the case, however, with the vast majority of ACLED event data, even in the most recent times, being drawn from traditional news-media, usually the same collections of national newspapers and news-agencies that UCDP or SCAD or MMAD use, and on which the entire media bias argument is made. 76.68% of ACLED events between December 2023 and July 2024 rely on news media either totally or partially, i.e. 407,050 of 530,788. Over half of ACLED is in fact based on the national newspapers of record, which, for all intent and purposes, are identical with news-wire data obtained from aggregators (as those aggregators will contain the newspapers of record globally). This proportion is actually higher than UCDPs reliance on news media, contrary to the claims made in Raleigh et al. (2023), indicating a severe misunderstanding in how news-media is used by non-ACLED sources. Local informants ("local partners") stand for 6.01% of the total number of events, either in full or in part, and are usually concentrated to small geographical areas and time periods (such as Yemen, Brazil and Benin). In fact, ACLED relies less on local partners than they do on propaganda disseminated by combatants – when measured on ACLED's own metric of preference – event counts: there are more events sourced to just one official news-source belonging to a warring party – the Ministry of Defense of Ukraine – than there are to all local partners and informants globally taken together - 40,855 against 31,357. All numbers are based on ACLED's own classification

Paper VI is the first paper in the field to test such claims, by evaluating the performance of media-based datasets against gold-standard, boots-on-the-ground observations by eyewitnesses, where such observation were never before accessible to event data coders. However, gold-standard data is hard to obtain, each instance being in effect akin to a natural experiment, further requiring extremely laborious work to correctly process (Demarest and Langer, 2022)[21] By applying this tried and tested, but extremely difficult method of employing gold-standard data in a comparative approach, we show that the traditional event data collection methodologies can work well in high-newsworthiness, high-coercion scenarios where bias and misrepresentation are difficult, such as abductions. However, news-driven data collection does not work at all and quickly breaks down in cases of voluntary actions such as most "strategic developments" (troop movements, base establishments etc.), essentially showing many of the claims and recommendations in Raleigh et al. (2023) are at best misguided and at worst will lead to severe biases in forecasting work.

2.3.5 The fog of (disaggregated) war: the omnipresent problems of resolution

The resolution of event-data is usually described in terms of spatio-temporal attributes. For instance, for GED, event are labeled to individual spatial coordinates at town/village level and individual dates (Croicu and Sundberg, 2018). However, as a consequence of biases and under-reporting, the resolution of the collected data is heterogeneous (Cook and Weidmann, 2022; Weidmann, 2016, 2015).

This also manifests itself in the spatio-temporal domain, the main analytical focus of these datasets. One notable example is what I define as the known geographic imprecision (KGI) problem, where the exact location of an event is unknown and is instead assigned to a large administrative unit (e.g., a region, or even an entire country). A representative point within this unit – often a major urban area or spatial centroid – is chosen, leading to biased analyses influenced by these arbitrary assignments. Temporal resolution also presents challenges. Capturing events in near-real time, during or shortly after a conflict, is fraught with uncertainties, since limitations in reporting and "fog-of-

scheme for sources; reliance on news-media is probably an undercount – many of the listed local partners are themselves aggregators of news-media in their own right such as *Liveuamap*, frequently relying on news-agency reporting for their own efforts.

[21] Gold-standard based comparative studies or similar methodologies has been shown capable of evaluating spatio-temporal and omission biases in the coding of fatal events in various contexts such as Afghanistan (Weidmann, 2016), Bosnia (Dulic, 2018), Kosovo (Gromes, 2023) and South Sudan (Dawkins, 2021), but never before used in the context of non-violent or less-than-lethal events.

war" tactics obscure accurate data collection (Burchell, 2020; Radford et al., 2023).

Paper II formalizes KGI, reframing it as an uncertainty/missing data problem in the spatial domain. **Paper II** then proposes a novel spatial multiple-imputation method using Gaussian processes – a geo-statistical method taking inspiration from literature on natural resource prospecting. It then tests the approach through an extensive Monte-Carlo simulation battery, showing substantial predictive performance improvement, as well as through a partial replication study, showing inferential improvement.

Paper I evaluates the practical implications of the temporal "fog of war" in near-real-time reporting, evaluating the amount of uncertainty by comparing data collected close to "near-real time" with data collected traditionally (with between 12 and 18 month delay). The findings reveal that the problem is more extensive than previously anticipated. Indeed, more than 40% of these events differ entirely between the two datasets – appearing in real-time collections but absent in later data revisions, or vice versa. Moreover, almost half of the remaining 60% experience significant revisions in spatial, temporal, intensity, or actor-related attributes. To address these challenges, **Paper I** introduces methodologies for principled near-real-time data collection and post-factum revision. These methodologies, now adopted by the Uppsala Conflict Data Program (UCDP) for its GED-candidate datasets, are shown to produce data with substantially higher forecasting accuracy than the classical (fully revised, non-real-time) datasets. This is achieved in two ways: first, by having access to near-real-time data, and second, by including uncertainty measurements during the collection process.

2.4 Natural language processing and conflict forecasting

2.4.1 NLP and event data: three generations of language models

Natural Language Processing (NLP) has a long history in the field, with early applications dating back to the 1990s, such as the use of part-of-speech tagging and simple text parsers (Schrodt et al., 1994). These efforts, however, have been historically focused on two main directions:

1. *Atomic event data collection* efforts, with the goal of extracting dictionary based pairs of noun-verbs. More recently, automatically generating dictionaries has been proposed as an extension of these methods.
2. *Data filtering for further human annotation*: NLP tools have been used to streamline the annotation process, usually at its early stages. For example, such tools have been applied to filter out irrelevant articles from a corpus (analogous to spam filters for e-mails) or to use unsupervised topic models for initial data exploration when developing annotation schemes or codebooks.

Multiple generations of efforts in using NLP in the field can be distinguished. The first generation is best exemplified by automatic event coders such TABARI (Gerner et al., 2002; Skorupa Parolin et al., 2019), developed in the late 1990s. This generation implemented a very simple and very shallow rule- and dictionary- based dependency parsing approach (Schrodt and Van Brackle, 2012; Schrodt, 2020). The goal of such first generation tools was to extract semantic and morphological relationships, with a goal of labelling patterns such as "subject actor (initiator) – verb action – object-actor (target)". However, these first-generation tools faced criticism for their extremely low signal-to-noise ratios, which rendered them ineffective at identifying even basic actual conflict trends (Hammond and Weidmann, 2014).

In response to this criticism, the field transitioned to a second generation of NLP tools that utilized more advanced methodologies. Primitive, custom-built parsers were replaced with state-of-the-art dependency parsing approaches, such as Universal Dependencies (McDonald et al., 2013). These tools, exemplified by systems like PETRARCH (Salam et al., 2018), allowed for more complex and much richer representations of language, with better performance at information extraction and much less noise. Additionally, they facilitated multilingual text analysis and minimized reliance on the arbitrary generation and implementation of parsing rules. This new generation of NLP tools has underpinned the development of event data collections like Phoenix (Salam et al., 2018; Skorupa Parolin et al., 2019; Beieler, 2016), marking a substantial evolution.

Finally, most recently, a third generation of models has emerged, leveraging the rapid advancements in NLP. These approaches integrate state-of-the art techniques from computational linguistics and computer science to enhance the extraction of atomic event data (Skorupa Parolin et al., 2020, 2022b, 2021; Halterman et al., 2023). These efforts have resulted in the development of a foundational language model tailored specifically for conflict-related texts, such as ConfliBERT (Hu et al., 2022). ConfliBERT is based on the BERT (Devlin, 2018) architecture, a small Transformer-based encoder model pretrained on conflict-focused corpora dominated by newswire texts. Despite the advanced capabilities of such models, their application has largely remained focused on a single, straightforward goal: extracting atomic events, i.e. simple noun-verb-noun phrases denoting actor-action-actor patterns.

2.4.2 NLP in the world of atomic event data

Despite significant advancements, even the most cutting-edge approaches, such as HANKE (Skorupa Parolin et al., 2020), Come-KE (Skorupa Parolin et al., 2021) and PLOVER / POLECAT (Halterman et al., 2023) still follow the paradigm established by the 25-year-old CAMEO (Skorupa Parolin et al., 2019) noun-verb dictionaries, with their proposed systems only providing limited extensions to its actor-action-actor ontology. Further, while CAMEO is sometimes described as a standard (Skorupa Parolin et al., 2020, 2021), this type of dictionary and dictionary-driven "atomic" event-data have had very limited forays into substantive conflict research. This was mostly due to continued and persistently high noise-to-signal ratios – much higher than in manually curated datasets such as UCDP GED and ACLED. Indeed, while no direct head-to-head forecasting-oriented benchmark was ever done between the two approaches[22], evidence from Beger et al. (2021) and Brandt et al. (2022) indicate "atomic" event data are underperforming their manually curated alternatives such as UCDP GED substantially, especially in VIEWS-like frameworks (Hegre et al., 2022a; Vesco et al., 2022). This performance gap persists even when atomic event data are paired with highly sophisticated forecasting methodologies, where results remain inferior to those achieved with manually curated datasets (Brandt et al., 2022). Indeed, all leading current forecasting systems such as VIEWS, ACLED CAST, and CoupCast (Hegre et al., 2021a, 2019; Raleigh et al., 2023) do not make use of "atomic" event data, instead making use of manually curated approaches such as UCDP GED and ACLED, with the last major CAMEO-based approach – ICEWS (Shilliday et al., 2012) – stopping making their forecasts and data available in 2023.[23] Furthermore, this lack of predictive performance coupled with the high signal-to-noise ration explains, at least in part, the use of very simple techniques, such as e.g. topic modelling with Latent Dirichlet Allocation (LDA), whenever text features are employed by conflict forecasting systems (Chadefaux, 2014, 2023; Mueller and Rauh, 2016; Mueller et al., 2024).

Even when communities outside CAMEO have adopted similar methodologies, they have mainly retained the basic paradigm of dictionary-based event extraction. For example, Häffner et al. (2023) use a neural network-based approach to create a novel dictionary for event extraction, completely distinct from CAMEO, using this dictionary in a shallow-learning forecasting approach further downstream. This is somewhat surprising, given that modern representation-based learners, such as those used in their study, do not require the intermediary step of defining coarse "noun-verb" event representations for forecasting. Instead, these models are capable of directly learning complex, less interpretable, yet more predictive representations tailored to a forecasting objective function.[24]

[22] mostly because the two kinds of event data are so incompatible

[23] No large-scale forecasting effort based on PLOVER/POLECAT, the open-source successor of ICEWS seems to exist at the time of writing. Brandt et al. (2022)'s prototype performed worse when compared to other competitors with similar architectures using manually curated data in the VIEWS 2020 prediction competition. Chadefaux (2023)'s approach makes limited use of ICEWS-based data as a complement to other approaches.

[24] Note, however, that Häffner et al. (2023)'s goal was interpretability and reusability of the dictionary, not forecasting; forecasting was just a tool to validate the approach, per the reasons given in Chapter 2 of this dissertation.

2.4.3 NLP and manual data curation of incident-level data

Beyond atomic event data, NLP techniques are mostly employed as pre-filters to exclude irrelevant articles before subsequent human curation. Generally, this relies on large, human annotated corpora of labeled data, using traditional, supervised machine-learning models trained to recognize patterns identified by human annotators.

Examples include Croicu and Weidmann (2015), who applied shallow learning approaches (bag-of-words with support-vector machines) to extract news articles about protests and riots; Zhukov (2012), who used shallow learning to filter NGO reports in the Caucasus for insurgent attack locations; and Zhang and Pan (2019), who utilized a two-stage convolutional neural network (CNN) to extract news articles about collective action in China. Beyond pre-filtering, efforts at information extraction have been more limited. Notable examples include Zhukov (2023), who applied a long short-term memory (LSTM) approach to extract and classify combat events in Ukraine, and Olsson et al. (2020), who achieved modest success using word embeddings generated by a bidirectional LSTM neural network in an attempt to replicate some human-designed features from the UCDP GED.

A common challenge with such approaches is their reliance on large annotated text corpora. This dependency is especially pronounced in scenarios involving rare events or highly imbalanced contexts, such as military tactics or electoral violence. In these cases, the subject of interest is very rare, perhaps appearing in 1–2% of all the texts, significantly limiting the applicability of the method. Addressing such problems requires excessively large training corpora, with tens of thousands of labeled instances (Zhang and Pan, 2019) or even hundreds of thousands (Zhang and Pan, 2019). These in turn impose substantial limitations on the practical use of these methods.

A common feature with such approaches is that they require a large corpus of annotated text to work. This is especially prominent in the case of rare events and highly imbalanced contexts such as e.g. military tactics, electoral violence or religious violence, where the subject of interest is very rare, perhaps appearing in 1–2% of all the texts, limiting the applicability domain of the method. For such problems, one needs excessively large training corpora, frequently in tens of thousands (Zhang and Pan, 2019) or even hundreds of thousands of labeled instances (Croicu and Weidmann, 2015). These in turn limit severely what these methods can be used for. One notable exception is Zhukov (2023), who does do some collection of rare events based on limited training. However, it is unclear how well this classifier performs, as he uses accuracy as the evaluation metric of interest – a highly sensitive and highly problematic metric to use with rare events.

This dissertation makes several contributions to advancing the use of NLP in the field of conflict research. **Paper III** introduces active learning to conflict information mining, using a human-in-the-loop approach to mine rare incident types such as electoral conflict, attacks on religious targets, and attacks on transportation infrastructure. This is achieved by mining UCDP corpora alongside fine-tuning ConfliBERT (Hu et al., 2022), the foundational large language model developed by the atomic event community. The paper also provides a comprehensive methodological framework, enabling further adaptation and application to other domains. **Paper IV** applies the mined data to address substantive questions on conflict escalation and the dynamics of targeted attacks – questions previously intractable given the data scarcity. This highlights the practical utility of the proposed approach in advancing conflict research.

Moving beyond the event data paradigm altogether, **Paper V** explores the capacity of deep learning methods to extract their own representations, i.e. generate a model-internal set of variables/features that are generally not human-interpretable, from noisy, complex signals. Specifically, it integrates actor-level data (e.g., rebel-group attributes) with UCDP corpora and retrieval-augmented generation (RAG) techniques (Lewis et al., 2020) to fine-tune a series of large language models – DeBERTa (He et al., 2020) and Mistral (Jiang et al., 2023). These models are trained to forecast the escalation and de-escalation of armed conflict, which is a novel forecasting target, achieving very strong predictive performance.

In summary, the articles in this thesis leverage state-of-the-art applied NLP techniques wherever text forms a part of the modelling pipeline. Notably, several contributions are entirely novel to the field, including the use of custom-trained large language models for direct text-to-forecast predictions and the innovative integration of active learning with deep learning methodologies.

2.5 Conclusions

While this dissertation addresses several key challenges in forecasting armed conflict, it also explores multiple novel approaches and raises new questions aimed at improving forecasting systems. These contributions pave the way for further advancements with the potential to achieve even greater predictive performance.

Figure 2.1 illustrates the contributions of this thesis in the context of a state-of-the-art forecasting system, such as the current generation of VIEWS. These are categorized into four key areas: data (both raw and processed datasets); feature engineering, which involves extracting, adapting, augmenting, clarifying, and processing datasets into forecasting-ready features; modelling, which refers to the forecasting techniques and processes employed; and predictions, representing the final output. The thesis also includes data contributions (new datasets) shared with the broader research community, as well as a summary of the theoretical insights offered by the papers.

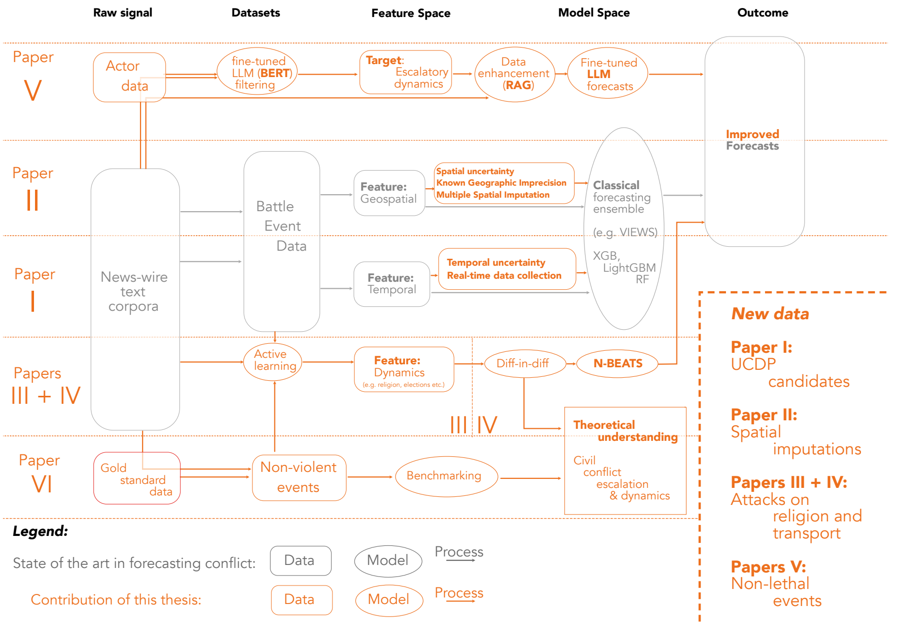

Figure 2.1. A schematic diagram of the thesis compared to the state-of-the-art in forecasting. Components of existing systems are labeled in black. Improvements made by the six papers in the thesis are labeled in orange.

Papers I and II underscore the critical role of uncertainty in prediction and propose methods to address it within the spatio-temporal domain, particularly on the data front. **Paper I** clearly demonstrates the improvements enabled by incorporating noisy but real-time battle event data. It quantifies the trade-offs between data quality and timeliness that arise from shifting from traditional annual annotation methods to near-real time updates. Additionally, it introduces a framework for identifying and labeling uncertainties in data quality caused by real-time annotation under conditions of the fog of war. These approaches have since become standard in the field, particularly for UCDP GED.

Paper II addresses the Known Geographic Imprecision (KGI) problem inherent in high-resolution geocoded conflict data. It introduces a novel, multiple-imputation framework that is both computationally efficient and theoretically robust for managing such uncertainty. Through nearly 200,000 simulations, the paper demonstrates that this method significantly improves predictive accuracy by mitigating the unavoidable spatial fuzziness of battle event data. Avenues for future extensions of these contributions are readily apparent, especially given the growing emphasis on multiple imputation in forecasting methodologies. Moreover, it is encouraging to observe that advancements in data handling are matched by parallel progress in modeling, with uncertainty now becoming a central focus in cutting-edge conflict forecasting experiments (Hegre et al., 2024). **Paper II** also revisits fundamental questions in conflict research, such as the interplay between urbanization, terrain accessibility, and violence, which warrant renewed exploration in light of the newly available data.

The approach introduced in **Paper III** offers a computationally inexpensive way to mine new features from existing conflict corpora. This innovation enables the extensive exploration of research avenues previously hindered by data unavailability and scarcity. **Paper IV** exemplifies this potential, investigating novel conflict dynamics within conflict episodes by examining escalatory patterns in response to targeted attacks on religious sites and transportation infrastructure. It proposes a new mechanism to explain these dynamics and highlights a path forward for improving the validity of findings derived from observational data. This advancement is particularly valuable in areas where experimental methods are not feasible. Notably, the generalizability of this approach stands out: conflict corpora are global in scope and have long time-series, whereas traditional approaches like quantitative case studies were limited to a single case with inherent constraints.

Papers I, II and **VI** revisit the limitations of event data, specifically addressing what can and cannot be collected using current approaches. These findings underscore the need to reassess recent assumptions about event data, such as in Raleigh et al. (2023). Some claims in that study are based on assumptions directly contradicted by the empirical evidence in this dissertation and prior theoretical studies – most notably, the feasibility of using predominantly news media-driven data and methodologies to collect reliable information on non-violent, non-lethal, low-intensity contention.

Papers II and **V** emphasize the crucial role of actors (i.e., governments and rebel groups) in conflict forecasting. Although this dissertation does not explicitly collect actor-level data, it incorporates them implicitly – as sources of data, levels of analysis, and latent scaffolds within the proposed methods. This should come as no surprise, since these actors are the primary drivers of armed conflict, war, and other types of contentious and violent politics, rather than the spatio-temporal units (such as grid cells and countries) typically used in forecasting efforts. Unrestricted access to actor-level data, along with the relational networks and interaction dynamics these actors engage in, would likely lead to a significant improvement in forecasting performance, bypassing the limitations imposed by spatio-temporal approximations and aggregations.

Paper V demonstrates that such an avenue is feasible, though it is clearly constrained by the availability of explicit actor-level data. This limitation is corroborated by other research, such as Kim et al. (2023), which encounters similar challenges in accessing high-quality actor-level data. It is encouraging that efforts are underway to address this gap, particularly in data collection (Braithwaite and Cunningham, 2020; Acosta, 2019). Indeed, while actor-level data collection has made progress, as seen in work by the author of this dissertation (Meier et al., 2023), it remains in its earliest stages. No existing methodology yet provides the real-time actor-level (in particular, rebel-level) data necessary for forecasting at the required scale.

Papers II, III, IV and V present significant methodological advances in applying state-of-the-art machine-learning approaches, both in conjunction with text and with tabular data. Looking ahead, one key avenue for further exploration is the use of language models for tasks far beyond information extraction, an area where a small but growing community is becoming increasingly active (Ollion et al., 2023). More ambitious applications, such as forecasting, as demonstrated in **Paper V**, also hold great promise. Given the rapid advancements in natural language processing, it is paramount – especially in the context of **Paper V** – that new methodologies including cutting-edge RAG techniques, innovative language models, and emerging foundational models are experimented with, trained, fine-tuned and applied to conflict forecasting.

Beyond these immediate advancements, and with a longer and more uncertain time horizon in mind, the integration of other media sources into forecasting models – beyond news-heavy text corpora – holds significant potential. While social media is problematic due to echo chambers and bias amplification (Tansini and Ben-Haim, 2021; Stanescu, 2022; Piazza, 2022), image corpora, which are slowly becoming more systematically available (von der Maase, 2022b; Steinert-Threlkeld, 2019), may offer a valuable new direction for forecasting. This is especially relevant given the declining quality and quantity of news reporting, which currently underpins much of our forecasting efforts (Johnson, 2019).

3. Abstracts of the essays

Paper I

Title:
Introducing the UCDP Candidate Events Dataset

Authorship:
Co-authored with Håvard Hegre, Kristine Eck and Stina Högbladh

Article:
This article presents a new, monthly updated dataset on organized violence—the Uppsala Conflict Data Program Candidate Events Dataset. It contains recent observations of candidate events, a majority of which are eventually included in the Uppsala Conflict Data Program Georeferenced Event Dataset as part of its annual update after a careful vetting process. We describe the definitions, sources and procedures employed to code the candidate events, and a set of issues that emerge when coding data on organized violence in near-real time. Together, the Uppsala Conflict Data Program Candidate and Georeferenced Event Datasets minimize an inherent trade-off between update speed and quality control. Having monthly updated conflict data is advantageous for users needing near-real time monitoring of violent situations and aiming to anticipate future developments. To demonstrate this, we show that including them in a conflict forecasting system yields distinct improvements in terms of predictive performance: Average precision increases by 20–40% relative to using the Uppsala Conflict Data Program Georeferenced Event Dataset only.

Paper II

Title:
Enhancing geospatial precision in conflict data: A stochastic approach to addressing known geographically imprecise observations in conflict event data.

Authorship:
Single-authored

Abstract:
The proliferation of large-scale, geographically disaggregated data on armed conflicts, protests, and similar events has opened new avenues of research, but has also introduced significant data quality challenges. A notable yet often overlooked issue involves observations with "known geographic imprecision" (KGI), where event locations are unknown and instead arbitrarily assigned by dataset authors. Although this issue is widely recognized and accounts for up to a quarter of observations in datasets like UCDP GED, it is rarely addressed by users. This paper presents a stochastic method derived from the multiple-imputation literature, employing spatio-temporal Gaussian processes and leveraging latent actor-conflict features in the data to enhance location accuracy. Extensive Monte-Carlo simulations demonstrate that this approach substantially enhances the accuracy of these observations and improves predictive performance beyond the state-of-the-art when applied out-of-sample. Additionally, an adapted version of the UCDP GED dataset that employs this new procedure is provided, showcasing the practical application and benefits of the methodology.

Paper III

Title:
Data mining from conflict text corpora with deep active learning

Authorship:
Single-authored

Abstract:
High-resolution event data on armed conflict and related processes have revolutionized the study of political contention. However, most datasets of this type only collect spatio-temporal and conflict intensity data at that level of detail. Information on dynamics, such as targets, tactics, and purposes, is rarely collected due to the substantial effort of collecting data. This study proposes an inexpensive, high-performance approach to increase the feature richness of such datasets by leveraging active learning – an iterative process of improving a machine learning model based on guided human input at each step of the learning process. Active learning is employed to then fine-tune (train in steps) a large, encoder-only language model fitted to the rich corpus of textual data underlying such datasets. This allows for the extraction of features related to conflict dynamics, such as electoral violence and attacks on religious targets. The approach achieves a performance comparable to the human (gold-standard) coding, while reducing the necessary human annotation by as much as 99 percent.

Paper IV

Title:
Provocation by design? Targeting of religious spaces and civil war dynamics

Authorship:
Co-authored with Joakim Kreutz

Abstract:
What explains conflict escalation during civil war? This article explores whether provocative attacks on religious sites and public transport constitute a precursor to a surge of violence. One argument pertains that the symbolic and doctrinal importance of places of worship means that attacks on these will affect individuals and the community emotionally and thereby increase the risk of escalation. However, it can also be suggested that the everyday societal importance of a public space is similar for religious sites and public transport hubs. We test these arguments using novel new global event data on these forms of selective targeting for 1989 – 2015, and find that the risk of conflict escalation increase in the aftermath of either attacks on places of worship or public transport, suggesting that community behavior is more affected to disruptions of societal everyday life than to the importance of symbols.

Paper V

Title:
From newswire to nexus. Using text-based actor embeddings and transformer networks to forecast conflict escalation

Authorship:
Co-authored with Simon Polichinel von der Maase

Abstract:
This study advances the field of conflict forecasting by using text-based actor embeddings with transformer models to predict dynamic changes in violent conflict patterns at the actor level. More specifically, we combine newswire texts with structured conflict event data and leverage recent advances in Natural Language Processing (NLP) techniques to forecast escalations and de-escalations among conflicting actors, such as governments, militias, separatist movements, and terrorists. This new approach accurately and promptly captures the inherently volatile patterns of violent conflicts, which existing methods have not been able to achieve.

To create this framework, we began by curating and annotating a vast international newswire corpus, leveraging hand-labeled event data from the Uppsala Conflict Data Program. By using this hybrid dataset, our models can incorporate the textual context of news sources along with the precision and detail of structured event data. This combination enables us to make both dynamic and granular predictions about conflict developments.

We validate our approach through rigorous back-testing against historical events, demonstrating superior out-of-sample predictive power. We find that our approach is quite effective in identifying and predicting phases of conflict escalation and de-escalation, surpassing the capabilities of traditional models. By focusing on actor interactions, our explicit goal is to provide actionable insights to policymakers, humanitarian organizations, and peacekeeping operations in order to enable targeted and effective intervention strategies.

Paper VI

Title:
Reporting of non-fatal conflict events

Authorship:
Co-authored with Kristine Eck

Abstract:
Temporally and spatial disaggregated datasets are commonly used to study political violence. Researchers are increasingly studying the data generation process itself to understand the selection processes by which conflict events are included in conflict datasets. This work has focused on conflict fatalities. In this paper, we explore how non-fatal conflict events are reported upon and enter into datasets of armed conflict. To do so, we compare reported non-fatal conflict events with the population of events in two direct observation datasets, collected using a boots-on-the-ground strategy: mass abductions in Nepal (1996–2006) and troop movements in Darfur. We show that at the appropriate level of aggregation media reporting on abductions in Nepal largely mirrors the "true" population of abductions, but at more disaggregated levels of temporal or spatial analysis, the match is poor. We also show that there is no overlap between a media-driven conflict dataset and directly-observed data on troop movements in Sudan. These empirics indicate that non-fatal data can suffer from serious underreporting and that this is particularly the case for events lacking elements of coercion. These findings are indicative of selection problems in regards to the reporting on non-fatal conflict events.

The cover illustration is an abstract, symbolic representation of the thesis created by `DALL-E` when prompted (via `GPT-4o`) with the dissertation abstract and these six abstracts concatenated together. It's a kind of techno-uroboros, a large machine-learning model looking in the mirror at a version of itself: a text-to-image large language model imagining the symbolism behind a thesis that constructs, works with and uses large language models to predict conflict.

4. Publication list

Due to the scope of the dissertation, not all publications by the author have been included in the dissertation, and conversely, not all papers in the dissertation are published. Following is a publication list of all peer-reviewed work by the author. Articles in bold face form an integral part of this dissertation. They are listed with the corresponding paper number as in the Table of Contents.

1. Croicu, M. and Weidmann, N. B. (2015). Improving the selection of news reports for event coding using ensemble classification. *Research & Politics*, 2(4):1–8.
2. von Uexkull, N., Croicu, M., Fjelde, H., and Buhaug, H. (2016). Civil conflict sensitivity to growing-season drought. *Proceedings of the National Academy of Sciences*, 113(44):12391–12396.
3. Croicu, M. and Kreutz, J. (2017). Communication technology and reports on political violence: Cross-national evidence using African events data. *Political Research Quarterly*, 70(1):19–31.
4. Hegre, H., Allansson, M., Basedau, M., Colaresi, M., Croicu, M., Fjelde, H., Hoyles, F., Hultman, L., Högbladh, S., Jansen, R., et al. (2019). ViEWS: A political violence early-warning system. *Journal of Peace Research*, 56(2):155–174.
5. **Hegre, H., Croicu, M., Eck, K., and Högbladh, S. (2020). Introducing the UCDP Candidate Events Dataset. *Research & Politics*, 7(3):1–8 - paper I**.
6. Buhaug, H., Croicu, M., Fjelde, H., and Von Uexkull, N. (2021). A conditional model of local income shock and civil conflict. *The Journal of Politics*, 83(1):354–366.
7. Vesco, P., Kovacic, M., Mistry, M., and Croicu, M. (2021). Climate variability, crop and conflict: Exploring the impacts of spatial concentration in agricultural production. *Journal of Peace Research*, 58(1):98–113.
8. Hegre, H., Bell, C., Colaresi, M., Croicu, M., Hoyles, F., Jansen, R., Leis, M. R., Lindqvist-McGowan, A., Randahl, D., Rød, E. G., et al. (2021a). ViEWS2020: Revising and evaluating the ViEWS political violence early-warning system. *Journal of Peace Research*, 58(3):599–611.
9. **Croicu, M. and Eck, K. (2022). Reporting of non-fatal conflict events. *International Interactions*, 48(3):450–470 - paper VI**.
10. Meier, V., Karlén, N., Pettersson, T., and Croicu, M. (2023). External support in armed conflicts: Introducing the UCDP External Support Dataset (ESD), 1975–2017. *Journal of Peace Research*, 60(3):545–554.

References

Abt, C. and Gorden, M. (1969). Report on project TEMPER. In Pruitt, D. and Snyder, R., editors, *Theory and Research on the Causes of War*. Prentice-Hall.

Acosta, B. (2019). Reconceptualizing resistance organizations and outcomes: introducing the Revolutionary and Militant Organizations dataset (REVMOD). *Journal of Peace Research*, 56(5):724–734.

Alexander, L. V. (2016). Global observed long-term changes in temperature and precipitation extremes: A review of progress and limitations in IPCC assessments and beyond. *Weather and Climate Extremes*, 11:4–16.

Bara, C. (2018). Legacies of violence: Conflict-specific capital and the postconflict diffusion of civil war. *Journal of Conflict Resolution*, 62(9):1991–2016.

Bara, C., Deglow, A., and van Baalen, S. (2021). Civil war recurrence and postwar violence: Toward an integrated research agenda. *European Journal of International Relations*, 27(3):913–935.

Baum, M. A. and Zhukov, Y. M. (2015). Filtering revolution: Reporting bias in international newspaper coverage of the Libyan civil war. *Journal of Peace Research*, 52(3):384–400.

Baum, M. A. and Zhukov, Y. M. (2019). Media ownership and news coverage of international conflict. *Political Communication*, 36(1):36–63.

Beck, N., Katz, J. N., and Tucker, R. (1998). Taking time seriously: Time-series-cross-section analysis with a binary dependent variable. *American Journal of Political Science*, 42(4):1260–1288.

Beger, A., Morgan, R. K., and Ward, M. D. (2021). Reassessing the role of theory and machine learning in forecasting civil conflict. *Journal of Conflict Resolution*, 65(7-8):1405–1426.

Beieler, J. (2016). Generating politically-relevant event data. In *Proceedings of the First Workshop on NLP and Computational Social Science*, pages 37–42.

Bengio, Y., Goodfellow, I., and Courville, A. (2017). *Deep learning*, volume 1. MIT press Cambridge, MA, USA.

Blair, R. A. and Sambanis, N. (2021). Is theory useful for conflict prediction? a response to Beger, Morgan, and Ward. *Journal of Conflict Resolution*, 65(7-8):1427–1453.

Blocher, J., Destrijcker, L., Fischer, B., Gleixner, S., Gornott, C., Hegre, H., Jansen, L., Jones, B., Kjærum, A., and Lindqvist-McGowan, A. e. a. (2022). *Moving from Reaction to Action-Anticipating Vulnerability Hotspots in the Sahel: A synthesis report from the Sahel Predictive Analytics project in support of the United Nations Integrated Strategy for the Sahel (UNISS).* UNHCR OSCDS, `https://uu.diva-portal.org/smash/get/diva2:1711326/FULLTEXT01.pdf`.

Braithwaite, J. M. and Cunningham, K. G. (2020). When organizations rebel: introducing the Foundations of Rebel Group Emergence (FORGE) dataset. *International Studies Quarterly*, 64(1):183–193.

Brandt, P. T., D'Orazio, V., Khan, L., Li, Y.-F., Osorio, J., and Sianan, M. (2022). Conflict forecasting with event data and spatio-temporal graph convolutional networks. *International Interactions*, 48(4):800–822.

Brandt, P. T., Freeman, J. R., and Schrodt, P. A. (2014). Evaluating forecasts of political conflict dynamics. *International Journal of Forecasting*, 30(4):944–962.

Bueno de Mesquita, B. (1981). Risk, power distributions, and the likelihood of war. *International Studies Quarterly*, 25(4):541–568.

Bueno de Mesquita, B. (1984). Forecasting policy decisions: an expected utility approach to post-Khomeini Iran. *PS: Political Science & Politics*, 17(2):226–236.

Bueno de Mesquita, B., Newman, D., and Rabushka, A. (1985). *Forecasting political events: The future of Hong Kong*. Yale University Press.

Buhaug, H., Croicu, M., Fjelde, H., and Von Uexkull, N. (2021). A conditional model of local income shock and civil conflict. *The Journal of Politics*, 83(1):354–366.

Buhaug, H. and Gleditsch, K. S. (2008). Contagion or confusion? Why conflicts cluster in space. *International Studies Quarterly*, 52(2):215–233.

Burchell, K. (2020). Practicing media—mediating practicel reporting, uncertainty, and the orchestrated fog of war: A practice-based lens for understanding global media events. *International Journal of Communication*, 14:23.

Cederman, L.-E. and Vogt, M. (2017). Dynamics and logics of civil war. *Journal of Conflict Resolution*, 61(9):1992–2016.

Cederman, L.-E. and Weidmann, N. B. (2017). Predicting armed conflict: Time to adjust our expectations? *Science*, 355(6324):474–476.

Cederman, L.-E., Weidmann, N. B., and Gleditsch, K. S. (2011). Horizontal inequalities and ethnonationalist civil war: A global comparison. *American Political Science Review*, 105(3):478–495.

Chadefaux, T. (2014). Early warning signals for war in the news. *Journal of Peace Research*, 51(1):5–18.

Chadefaux, T. (2017). Conflict forecasting and its limits. *Data Science*, 1(1-2):7–17.

Chadefaux, T. (2023). An automated pattern recognition system for conflict. *Journal of Computational Science*, 72:102074.

Colaresi, M. and Mahmood, Z. (2017). Do the robot: Lessons from machine learning to improve conflict forecasting. *Journal of Peace Research*, 54(2):193–214.

Collier, P. et al. (2003). *Breaking the Conflict Trap: Civil War and Development Policy*. World Bank Publications.

Collier, P. and Sambanis, N. (2002). Understanding civil war: A new agenda. *Journal of Conflict Resolution*, 46(1):3–12.

Cook, S. J. and Weidmann, N. B. (2019). Lost in aggregation: Improving event analysis with report-level data. *American Journal of Political Science*, 63(1):250–264.

Cook, S. J. and Weidmann, N. B. (2022). Race to the bottom: Spatial aggregation and event data. *International Interactions*, 48(3):471–491.

Croicu, M. and Eck, K. (2022). Reporting of non-fatal conflict events. *International Interactions*, 48(3):450–470.

Croicu, M. and Kreutz, J. (2017). Communication technology and reports on political violence: Cross-national evidence using African events data. *Political Research Quarterly*, 70(1):19–31.

Croicu, M. and Sundberg, R. (2018). UCDP GED codebook version 18.1. https://ucdp.uu.se/downloads/ged/ged181.pdf.

Croicu, M. and Weidmann, N. B. (2015). Improving the selection of news reports for event coding using ensemble classification. *Research & Politics*, 2(4):1–8.

Crost, B., Felter, J., et al. (2015). Is conflict contagious? Evidence from a natural experiment. Technical report, Households in Conflict Network, https://hicn.org/working-paper/is-conflict-contagious-evidence-from-a-natural-experiment/.

Davies, S., Engström, G., Pettersson, T., and Öberg, M. (2024). Organized violence 1989–2023, and the prevalence of organized crime groups. *Journal of Peace Research*, page 00223433241262912.

Dawkins, S. (2021). The problem of the missing dead. *Journal of Peace Research*, 58(5):1098–1116.

Deglow, A. and Sundberg, R. (2021). Local conflict intensity and public perceptions of the police: Evidence from Afghanistan. *The Journal of Politics*, 83(4):1337–1352.

Demarest, L. and Langer, A. (2022). How events enter (or not) data sets: the pitfalls and guidelines of using newspapers in the study of conflict. *Sociological Methods & Research*, 51(2):632–666.

Devlin, J. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, https://arxiv.org/abs/1810.04805.

Dietrich, N. and Eck, K. (2020). Known unknowns: media bias in the reporting of political violence. *International Interactions*, 46(6):1043–1060.

Donnay, K., Dunford, E. T., McGrath, E. C., Backer, D., and Cunningham, D. E. (2019). Integrating conflict event data. *Journal of Conflict Resolution*, 63(5):1337–1364.

Dulic, T. (2018). The patterns of violence in Bosnia and Herzegovina: Security, geography and the killing of civilians during the war of the 1990s. *Political Geography*, 63:148–158.

D'Orazio, V. and Lin, Y. (2022). Forecasting conflict in Africa with automated machine learning systems. *International Interactions*, 48(4):714–738.

Earl, J., Martin, A., McCarthy, J. D., and Soule, S. A. (2004). The use of newspaper data in the study of collective action. *Annual Review of Sociology*, 30(1):65–80.

Eck, K. (2014). Coercion in rebel recruitment. *Security Studies*, 23(2):364–398.

Elfversson, E. (2021). Cities and armed conflict: A systematic urban-rural coding of UCDP conflict events data. *Data in brief*, 39:107554.

Fariss, C. J. and Jones, Z. M. (2018). Enhancing validity in observational settings when replication is not possible. *Political Science Research and Methods*, 6(2):365–380.

Fjelde, H. and Höglund, K. (2022). Introducing the Deadly Electoral Conflict Dataset (DECO). *Journal of Conflict Resolution*, 66(1):162–185.

Franklin, B. (2008). The future of newspapers. *Journalism Practice*, 2(3):306–317.

Franzosi, R. (1987). The press as a source of socio-historical data: issues in the methodology of data collection from newspapers. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 20(1):5–16.

Gallo, G. (2013). Conflict theory, complexity and systems approach. *Systems Research and Behavioral Science*, 30(2):156–175.

Galtung, J. (1969). Violence, peace, and peace research. *Journal of Peace Research*, 6(3):167–191.

Gerner, D. J., Schrodt, P. A., Yilmaz, O., and Abu-Jabr, R. (2002). The creation of CAMEO (Conflict and Mediation Event Observations): An event data framework for a post Cold War world. In *Political Research Online: Proceedings of the Annual meeting of the American Political Science Association*, volume 98. `https://parusanalytics.com/eventdata/papers.dir/Gerner.APSA.02.pdf`.

Gohdes, A. and Price, M. (2013). First things first: Assessing data quality before model quality. *Journal of Conflict Resolution*, 57(6):1090–1108.

Gohdes, A. R. and Carey, S. C. (2017). Canaries in a coal-mine? What the killings of journalists tell us about future repression. *Journal of Peace Research*, 54(2):157–174.

Goldstone, J. A., Bates, R. H., Epstein, D. L., Gurr, T. R., Lustik, M. B., Marshall, M. G., Ulfelder, J., and Woodward, M. (2010). A global model for forecasting political instability. *American Journal of Political Science*, 54(1):190–208.

Goldstone, J. A., Gurr, T. R., Harff, B., Levy, M. A., Marshall, M. G., Bates, R. H., Epstein, D. L., Kahl, C. H., Surko, P. T., Ulfelder, J. C., et al. (2000). State failure task force report: Phase III findings. *McLean, VA: Science Applications International Corporation*, 30.

Gromes, T. (2023). Reporting of conflict fatalities in the UCDP Georeferenced Event Dataset: insights from Kosovo. *Zeitschrift für Vergleichende Politikwissenschaft*, 17(2):189–206.

Gurr, T. R. (1970). *Why men rebel*. Princeton University Press.

Gurr, T. R. (1995). *Minorities at Risk – A Global View of Ethnopolitical Conflicts*. United States Institute of Peace Press, Arlington, Virginia.

Gurr, T. R., Harff, B., Levy, M., Dabelko, G. D., Surko, P. T., and Unger, A. N. (1999). State failure task force report: Phase II findings. *Environmental Change & Security Project Report*, 5:49–72.

Häffner, S., Hofer, M., Nagl, M., and Walterskirchen, J. (2023). Introducing an interpretable deep learning approach to domain-specific dictionary creation: A use case for conflict prediction. *Political Analysis*, 31(4):481–499.

Halkia, S., Ferri, S., Joubert-Boitat, I., Saporiti, F., Kauffmann, M., et al. (2017). The Global Conflict Risk Index (GCRI) regression model: data ingestion, processing, and output methods. *Luxembourg: Publications Office of the European Union*, pages 1–12, DOI: `10.2760/303651`.

Halterman, A., Bagozzi, B., Beger, A., Schrodt, P., and Scraborough, G. (2023). PLOVER and POLECAT: A new political event ontology and dataset. In *International Studies Association Conference Paper*. DOI: `10.31235/osf.io/rm5dw`.

Hammond, J. and Weidmann, N. B. (2014). Using machine-coded event data for the micro-level study of political violence. *Research & Politics*, 1(2):1–7.

Harding, R. and Nwokolo, A. (2024). Terrorism, trust, and identity: Evidence from a natural experiment in nigeria. *American Journal of Political Science*, 68(3):942–957.

He, P., Liu, X., Gao, J., and Chen, W. (2020). DeBERTa: Decoding-enhanced BERT with disentangled attention. *arXiv preprint arXiv:2006.03654*, `https://arxiv.org/abs/2006.03654`.

Hegre, H., Akbari, F., Croicu, M., Dale, J., Gåsste, T., Jansen, R., Landsverk, P., Leis, M., Lindqvist-McGowan, A., Mueller, H., et al. (2022a). Forecasting fatalities. VIEWS Working Paper Series, Uppsala University, `https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-476476`.

Hegre, H., Allansson, M., Basedau, M., Colaresi, M., Croicu, M., Fjelde, H., Hoyles, F., Hultman, L., Högbladh, S., Jansen, R., et al. (2019). ViEWS: A political violence early-warning system. *Journal of Peace Research*, 56(2):155–174.

Hegre, H., Bell, C., Colaresi, M., Croicu, M., Hoyles, F., Jansen, R., Leis, M. R., Lindqvist-McGowan, A., Randahl, D., Rød, E. G., et al. (2021a). ViEWS2020: Revising and evaluating the ViEWS political violence early-warning system. *Journal of Peace Research*, 58(3):599–611.

Hegre, H., Croicu, M., Eck, K., and Högbladh, S. (2020). Introducing the UCDP Candidate Events Dataset. *Research & Politics*, 7(3):1–8.

Hegre, H., Nygård, H. M., and Landsverk, P. (2021b). Can we predict armed conflict? How the first 9 years of published forecasts stand up to reality. *International Studies Quarterly*, 65(3):660–668.

Hegre, H., Nygård, H. M., and Ræder, R. F. (2017). Evaluating the scope and intensity of the conflict trap: A dynamic simulation approach. *Journal of Peace Research*, 54(2):243–261.

Hegre, H. and Sambanis, N. (2006). Sensitivity analysis of empirical results on civil war onset. *Journal of Conflict Resolution*, 50(4):508–535.

Hegre, H., Vesco, P., and Colaresi, M. (2022b). Lessons from an escalation prediction competition. *International Interactions*, 48(4):521–554.

Hegre, H., Vesco, P., Colaresi, M., Vestby, J., Timlick, A., Kazmi, N. S., Becker, F., Binetti, M., Bodentien, T., Bohne, T., Brandt, P. T., Chadefaux, T., Drauz, S., Dworschak, C., D'Orazio, V., Fritz, C., Frank, H., Gleditsch, K. S., Häffner, S., Hofer, M., Klebe, F. L., Macis, L., Malaga, A., Mehrl, M., Metternich, N. W., Mittermaier, D., Muchlinski, D., Mueller, H., Oswald, C., Pisano, P., Randahl, D., Rauh, C., Rüter, L., Schincariol, T., Seimon, B., Siletti, E., Tagliapietra, M., Thornhill, C., Vegelius, J., and Walterskirchen, J. (2024). The 2023/24 views prediction challenge: Predicting the number of fatalities in armed conflict, with uncertainty. *Journal of Peace Research*, in press, DOI: `10.1177/00223433241300862`.

Hu, Y., Hosseini, M. S., Skorupa Parolin, E., Osorio, J., Khan, L., Brandt, P., and D'Orazio, V. (2022). ConfliBERT: A pre-trained language model for political conflict and violence. In *Proceedings of the Association for Computational Linguistics*. Association for Computational Linguistics, DOI: `10.18653/v1/2022.naacl-main.400`.

Jeffrey, N., Gunawan, A. A. S., and Kurniawan, A. (2023). Development of multivariate stock prediction system using N-HITS and N-BEATS. In *Proceedings of the Computational Methods in Systems and Software*, pages 50–63. Springer.

Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., Casas, D. d. l., Bressand, F., Lengyel, G., Lample, G., Saulnier, L., et al. (2023). Mistral 7b. *arXiv preprint arXiv:2310.06825*, `https://arxiv.org/abs/2310.06825`.

Johnson, L. (2019). Translation and open-source intelligence: BBC Monitoring. *The Palgrave Handbook of Languages and Conflict*, pages 251–271.

Kalyvas, S. N. (2003). The ontology of “political violence”: action and identity in civil wars. *Perspectives on Politics*, 1(03):475–494.

Kalyvas, S. N. and Kocher, M. A. (2007). How “free” is free riding in civil wars?: Violence, insurgency, and the collective action problem. *World Politics*, 59(2):177–216.

Kim, S., Liu, H., and Desmarais, B. (2023). Spatial modeling of dyadic geopolitical interactions between moving actors. *Political Science Research and Methods*, 11(3):633–644.

King, G. and Zeng, L. (2001). Improving forecasts of state failure. *World Politics*, 53(4):623–658.

LaFree, G. and Dugan, L. (2007). Introducing the Global Terrorism Database. *Terrorism and Political Violence*, 19(2):181–204.

Landman, T. and Gohdes, A. (2013). A matter of convenience. In Seybolt, T. B., Aronson, J. D., and Fischhoff, B., editors, *Counting civilian casualties: An introduction to recording and estimating nonmilitary deaths in conflict*. Oxford University Press, New York, NY.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474.

Lindberg Bromley, S. (2018). Introducing the UCDP Peacemakers at Risk dataset, sub-Saharan Africa, 1989–2009. *Journal of Peace Research*, 55(1):122–131.

Lu, Y. and Pan, J. (2022). The pervasive presence of Chinese government content on Douyin trending videos. *Computational Communication Research*, 4(1):68–97.

Lyons, T. D. (2006). Scientific realism and the stratagema de divide et impera. *The British Journal for the Philosophy of Science*, 57(3):537–560.

Malone, I. (2022). Recurrent neural networks for conflict forecasting. *International Interactions*, 48(4):614–632.

McDonald, R., Nivre, J., Quirmbach-Brundage, Y., Goldberg, Y., Das, D., Ganchev, K., Hall, K., Petrov, S., Zhang, H., Täckström, O., et al. (2013). Universal dependency annotation for multilingual parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 92–97.

Meier, V., Karlén, N., Pettersson, T., and Croicu, M. (2023). External support in armed conflicts: Introducing the UCDP External Support Dataset (ESD), 1975–2017. *Journal of Peace Research*, 60(3):545–554.

Miller, E., Kishi, R., Raleigh, C., and Dowd, C. (2022). An agenda for addressing bias in conflict data. *Scientific Data*, 9(1):593.

Mueller, H. and Rauh, C. (2022a). The hard problem of prediction for conflict prevention. *Journal of the European Economic Association*, 20(6):2440–2467.

Mueller, H. and Rauh, C. (2022b). Using past violence and current news to predict changes in violence. *International Interactions*, 48(4):579–596.

Mueller, H., Rauh, C., and Seimon, B. (2024). Introducing a global dataset on conflict forecasts and news topics. *Data & Policy*, 6:e17.

Mueller, H. F. and Rauh, C. (2016). Reading between the lines: Prediction of political violence using newspaper text. *American Political Science Review*, 2(112):358–375.

Norris, C., Schrodt, P. A., and Beieler, J. (2017). Petrarch2: Another event coding program. *Journal of Open Source Software*, 2(9):133.

Öberg, M. and Sollenberg, M. (2011). Gathering conflict information using news resources. In Höglund, K. and Öberg, M., editors, *Understanding peace research*, pages 47–73. Routledge.

Ollion, E., Shen, R., Macanovic, A., and Chatelain, A. (2023). ChatGPT for text annotation? mind the hype. *SocArXiv preprint*, DOI: `10.31235/osf.io/x58kn`.

O'Loughlin, J., Witmer, F. D., and Linke, A. M. (2010). The Afghanistan-Pakistan wars, 2008-2009: Micro-geographies, conflict diffusion, and clusters of violence. *Eurasian Geography and Economics*, 51(4):437–471.

Olsson, F., Sahlgren, M., Abdesslem, F. B., Ekgren, A., and Eck, K. (2020). Text categorization for conflict event annotation. In *Proceedings of the Workshop on Automated Extraction of Socio-political Events from News 2020*, pages 19–25.

Oreshkin, B. N., Carpov, D., Chapados, N., and Bengio, Y. (2019). N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437*, `https://arxiv.org/abs/1905.10437`.

O'Sullivan, J., Fortunati, L., Taipale, S., and Barnhurst, K. (2017). Innovators and innovated: Newspapers and the postdigital future beyond the "death of print". *The Information Society*, 33(2):86–95.

Petrova, K. (2021). Natural hazards, internal migration and protests in Bangladesh. *Journal of Peace Research*, 58(1):33–49.

Piazza, J. A. (2022). Fake news: The effects of social media disinformation on domestic terrorism. *Dynamics of Asymmetric Conflict*, 15(1):55–77.

Polo, S. M. and Gleditsch, K. S. (2016). Twisting arms and sending messages: Terrorist tactics in civil war. *Journal of Peace Research*, 53(6):815–829.

Radford, B. J. (2022). High resolution conflict forecasting with spatial convolutions and long short-term memory. *International Interactions*, 48(4):739–758.

Radford, B. J., Dai, Y., Stoehr, N., Schein, A., Fernandez, M., and Sajid, H. (2023). Estimating conflict losses and reporting biases. *Proceedings of the National Academy of Sciences*, 120(34):e2307372120.

Raleigh, C., Kishi, R., and Linke, A. (2023). Political instability patterns are obscured by conflict dataset scope conditions, sources, and coding choices. *Humanities and Social Sciences Communications*, 10(1):1–17.

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.

Randahl, D. and Vegelius, J. (2022). Predicting escalating and de-escalating violence in Africa using Markov models. *International Interactions*, 48(4):597–613.

Richardson, L. (1960). *Arms and Insecurity: A Mathematical Study of the Causes and Origins of War*. Boxwood Press.

Rød, E. G., Gåsste, T., and Hegre, H. (2024). A review and comparison of conflict early warning systems. *International Journal of Forecasting*, 40(1):96–112.

Rød, E. G., Knutsen, C. H., and Hegre, H. (2020). The determinants of democracy: a sensitivity analysis. *Public Choice*, 185(1):87–111.

Rød, E. G. and Weidmann, N. B. (2023). From bad to worse? how protest can foster armed conflict in autocracies. *Political Geography*, 103:102891.
Rummel, R. J. (1969). Forecasting international relations: A proposed investigation of three-mode factor analysis. *Technological Forecasting*, 1(2):197–216.
Salam, S., Brandty, P., Holmesy, J., and Khan, L. (2018). Distributed framework for political event coding in real-time. In *2018 2nd European Conference on Electrical Engineering and Computer Science (EECS)*, pages 266–273. DOI: `10.1109/EECS.2018.00057`.
Salehyan, I., Hendrix, C. S., Hamner, J., Case, C., Linebarger, C., Stull, E., and Williams, J. (2012). Social Conflict in Africa: A new database. *International Interactions*, 38(4):503–511.
Schrodt, P. (2020). Technical forecasting of political conflict. In *Technical Notes*. Indiana University Workshop in Methods.
Schrodt, P. A. (1990). Predicting interstate conflict outcomes using a bootstrapped ID3 algorithm. *Political Analysis*, 2:31–56.
Schrodt, P. A. (1991). Prediction of interstate conflict outcomes using a neural network. *Social Science Computer Review*, 9(3):359–380.
Schrodt, P. A. (2006). Forecasting conflict in the Balkans using hidden Markov models. In *Programming for peace: Computer-aided methods for international conflict resolution and prevention*, pages 161–184. Springer.
Schrodt, P. A. (2012). Precedents, progress, and prospects in political event data. *International Interactions*, 38(4):546–569.
Schrodt, P. A., Davis, S. G., and Weddle, J. L. (1994). Political science: KEDS – a program for the machine coding of event data. *Social Science Computer Review*, 12(4):561–587.
Schrodt, P. A. and Hall, B. (2006). Twenty years of the Kansas event data system project. *The political methodologist*, 14(1):2–8.
Schrodt, P. A. and Van Brackle, D. (2012). Automated coding of political event data. In *Handbook of computational approaches to counterterrorism*, pages 23–49. Springer.
Schutte, S. and Weidmann, N. B. (2011). Diffusion patterns of violence in civil wars. *Political Geography*, 30(3):143–152.
Shilliday, A., Lautenschlager, J., et al. (2012). Data for a worldwide ICEWS and ongoing research. *Advances in Design for Cross-Cultural Activities*, 455.
Singer, J. D. (1972). The Correlates of War project: Interim report and rationale. *World Politics*, 24(2):243–270.
Singer, J. D. (1973). The peace researcher and foreign policy prediction. *Peace Science Society (International)*, 21:1–13.
Skorupa Parolin, E., Hosseini, M., Hu, Y., Khan, L., Brandt, P. T., Osorio, J., and D'Orazio, V. (2022a). Multi-COPED: A multilingual multi-task approach for coding political event data on conflict and mediation domain. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, pages 700–711.
Skorupa Parolin, E., Hu, Y., Khan, L., Brandt, P. T., Osorio, J., and D'Orazio, V. (2022b). Confli-T5: An autoprompt pipeline for conflict related text augmentation. In *2022 IEEE International Conference on Big Data*, pages 1906–1913. DOI: `10.1109/BigData55660.2022.10020509`.

Skorupa Parolin, E., Hu, Y., Khan, L., Osorio, J., Brandt, P. T., and D'Orazio, V. (2021). COME-KE: A new transformers based approach for knowledge extraction in conflict and mediation domain. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 1449–1459. DOI: `10.1109/BigData52589.2021.9672080`.

Skorupa Parolin, E., Khan, L., Osorio, J., D'Orazio, V., Brandt, P. T., and Holmes, J. (2020). HANKE: Hierarchical attention networks for knowledge extraction in political science domain. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 410–419. DOI: `10.1109/DSAA49011.2020.00055`.

Skorupa Parolin, E., Salam, S., Khan, L., Brandt, P., and Holmes, J. (2019). Automated verbal-pattern extraction from political news articles using CAMEO Event Coding Ontology. In *2019 IEEE Intl Conference on Intelligent Data and Security (IDS)*, volume 1, pages 258–266. DOI: `10.1109/BigDataSecurity-HPSC-IDS.2019.00056`.

Smidt, H. M. (2020). United nations peacekeeping locally: enabling conflict resolution, reducing communal violence. *Journal of Conflict Resolution*, 64(2-3):344–372.

Stanescu, G. (2022). Ukraine conflict: the challenge of informational war. *Social sciences and education research review*, 9(1):146–148.

Steinert-Threlkeld, Z. C. (2018). *Twitter as data*. Cambridge University Press.

Steinert-Threlkeld, Z. C. (2019). The future of event data is images. *Sociological Methodology*, 49(1):68–75.

Steinert-Threlkeld, Z. C., Chan, A. M., and Joo, J. (2022). How state and protester violence affect protest dynamics. *The Journal of Politics*, 84(2):798–813.

Sundberg, R. and Melander, E. (2013). Introducing the UCDP Georeferenced Event Dataset. *Journal of Peace Research*, 50(4):523–532.

Tansini, F. and Ben-Haim, Y. (2021). Strategies for communicating information and disinformation in war: Managing and exploiting uncertainty in social media. In *The Conduct of War in the 21st Century*, pages 58–71. Routledge.

Tilly, C. (1978). *From Mobilization to Revolution*. McGraw-Hill.

Tollefsen, A. F., Strand, H., and Buhaug, H. (2012). PRIO–GRID: A unified spatial data structure. *Journal of Peace Research*, 49(2):363–374.

Vesco, P., Hegre, H., Colaresi, M., Jansen, R. B., Lo, A., Reisch, G., and Weidmann, N. B. (2022). United they stand: Findings from an escalation prediction competition. *International Interactions*, 48(4):860–896.

Vesco, P., Kovacic, M., Mistry, M., and Croicu, M. (2021). Climate variability, crop and conflict: Exploring the impacts of spatial concentration in agricultural production. *Journal of Peace Research*, 58(1):98–113.

von der Maase, S. P. (2022a). Conflictnet 1.0: A probabilistic recurrent u-net for globe conflict forecasting. *University of Copenhagen (UCPH), Department of Political Science (IFS) and the Center for Social Data Science (SODAS)*.

von der Maase, S. P. (2022b). From front-line to front-page. *University of Copenhagen (UCPH), Department of Political Science (IFS) and the Center for Social Data Science (SODAS)*.

von der Maase, S. P. (2022c). Introducing the Bodies as Battleground dataset. *University of Copenhagen (UCPH), Department of Political Science (IFS) and the*

Center for Social Data Science (SODAS).
von Uexkull, N., Croicu, M., Fjelde, H., and Buhaug, H. (2016). Civil conflict sensitivity to growing-season drought. *Proceedings of the National Academy of Sciences*, 113(44):12391–12396.
Walter, B. F. (2004). Does conflict beget conflict? Explaining recurring civil war. *Journal of Peace Research*, 41(3):371–388.
Ward, M. D., Greenhill, B. D., and Bakke, K. M. (2010). The perils of policy by p-value: Predicting civil conflicts. *Journal of Peace Research*, 47(4):363–375.
Weidmann, N. B. (2015). On the accuracy of media-based conflict event data. *Journal of Conflict Resolution*, 59(6):1129–1149.
Weidmann, N. B. (2016). A closer look at reporting bias in conflict event data. *American Journal of Political Science*, 60(1):206–218.
Weidmann, N. B. and Rød, E. G. (2019). *The Internet and Political Protest in Autocracies*. Oxford University Press.
Weidmann, N. B. and Schutte, S. (2017). Using night light emissions for the prediction of local wealth. *Journal of Peace Research*, 54(2):125–140.
Winkelmann, R. and Boes, S. (2006). *Analysis of microdata*, volume 313. Springer.
Wood, R. M. (2014). From loss to looting? battlefield costs and rebel incentives for violence. *International Organization*, 68(4):979–999.
Zeitzoff, T. (2011). Using social media to measure conflict dynamics: An application to the 2008–2009 Gaza conflict. *Journal of Conflict Resolution*, 55(6):938–969.
Zhang, H. and Pan, J. (2019). CASM: A deep-learning approach for identifying collective action events with text and image data from social media. *Sociological Methodology*, 49(1):1–57.
Zhukov, Y. M. (2012). Roads and the diffusion of insurgent violence: The logistics of conflict in Russia's North Caucasus. *Political Geography*, 31(3):144–156.
Zhukov, Y. M. (2023). Near-real time analysis of war and economic activity during Russia's invasion of Ukraine. *Journal of Comparative Economics*, 51(4):1232–1243.
Zhukov, Y. M. and Baum, M. A. (2016). How selective reporting shapes inferences about conflict. *Unpublished working paper*, `https://scholar.harvard.edu/files/zhukov/files/2016_zhukovbaum_unpublished.pdf`.
Zhukov, Y. M., Davenport, C., and Kostyuk, N. (2019). Introducing xSub: A new portal for cross-national data on subnational violence. *Journal of Peace Research*, 56(4):604–614.